

MEDICAL GENETICS

UpToDate TOOLS FOR GENETICS



2015

Index

- Genetic counseling and testing
- Genomic disorders
- Tools for genetics and genomics
- Chromosomal translocations, deletions, and inversions
- Use of chromosomal microarray in obstetrics
- Polymerase chain reaction
- Incidental findings from genetic testing
- Clinical applications of next-generation DNA sequencing
- Gene expression profiling
- Proteomics, and microRNA profiling in clinical oncology
- Overview of pharmacogenomics
- Personalized medicine
- Genetic association studies

1

Derleme
Dr. Yavuz ŞAHİN

Genetic counseling and testing

Authors

Benjamin A Raby, MD, MPH
Wendy Kohlmann, MS, CGC
Vickie Venne, MS, CGC

Section Editor

Anne Slavotinek, MBBS, PhD

Deputy Editor

Jennifer S Tirnauer, MD

All topics are updated as new evidence becomes available and our [peer review process](#) is complete.

Literature review current through: Feb 2015. | **This topic last updated:** Mar 10, 2015.

INTRODUCTION — The inherited nature of disease has long been recognized. The field of medical genetics initially sought to explain the cause of rare diseases with a single gene Mendelian inheritance pattern, such as cystic fibrosis or Huntington disease. Recent discoveries have led to an exponentially increased understanding of the causes of many common diseases, in which disease risk often is determined by the combined effects of genetic and non-genetic factors.

Gene alterations that confer increased risk for adult onset conditions such as cancer, diabetes, cardiovascular disease, and other common diseases have been identified [1]. The hope is that identification of heightened genetic risk will allow specifically tailored interventions (ie, “personalized” or “precision” medicine), encourage lifestyle changes, and ultimately prevent disease. As an example, in the case of women identified with a *BRCA1* or *BRCA2* gene mutation, a prospective multicenter study found that women who underwent risk-reducing salpingo-oophorectomy, compared to those who did not, had lower all-cause mortality (HR 0.40, 95% CI 0.26-76) and a lower risk of cancer incidence and cancer-specific mortality [2]. There were no breast cancers in those women with *BRCA* mutations who underwent risk-reducing mastectomy, compared to breast cancer occurrence in 98 of 1372 women who did not have a mastectomy [2].

Targeting appropriate patients for genetic counseling and testing is key to the successful transfer of genetics research to improvement of health and quality of life. The primary care clinician works at the front line where advances in genetics research can be directly applied to patient care.

This review will focus on providing general guidelines related to genetic assessment for common, adult-onset conditions and will address obtaining a health history, genetic counseling and testing, and the primary care clinician's role in this process. The genetics of specific disorders are addressed separately.

Additional discussions are provided on the following topics:

- Genetics terminology — (See ["Glossary of genetic terms"](#).)
- Pharmacogenomics — (See ["Overview of pharmacogenomics"](#).)
- Personalized medicine, including direct-to-consumer testing — (See ["Personalized medicine"](#).)
- Disclosure of incidental findings from genetic testing — (See ["Incidental findings from genetic testing"](#).)

ROLE OF THE PRIMARY CARE OR REFERRING CLINICIAN — Primary care clinicians have multiple roles in assessing the genetic background of their patients [3]:

- Obtaining the family history
- Assessing risk information
- Identifying and referring patients for whom a genetic consultation would be beneficial
- Coordinating long-term management for individuals who have hereditary syndromes

When a genetic referral is indicated, the referring provider should set patient expectations for the referral. Following the genetic consultation, when risk and genetic information from the consultant are available, the primary care clinician is generally responsible for long-term follow-up. This may include overseeing behavioral modifications, interventions for disease prevention, or intensified screening for disease surveillance.

Tools for the assessment of genetic risk include instruments to aid in obtaining a family history, medical records of family members, phenotypic measurements (eg, serum iron when hemochromatosis is a potential concern), and genetic testing.

FAMILY HISTORY — The initial step in assessing inherited risk for many chronic conditions is collecting data related to the family history. The history is then reviewed for patterns consistent with various modes of inheritance.

A 2009 panel sponsored by the National Institutes of Health evaluated evidence regarding the accuracy and effectiveness of obtaining a family history in the primary care setting [4]. The panel found that there is limited scientific evidence regarding how best to collect the history, the key elements of the history, the accuracy of reported history, or the impact of a positive family history on patient outcomes. While it is clear that more research in this field is indicated, and that the family history will not detect all families with inherited disorders, it remains a starting point for identifying appropriate candidates who could benefit from a genetic consultation.

Although the family history is not highly predictive for common disease, it can help determine patients who are at average, moderate, or high risk for specific conditions, and can help to personalize health lifestyle messages [5,6]. Among 200 pedigrees from patients seen in a prenatal clinic, 5 to 15 percent were classified as moderate risk for at least one common adult condition and 1 to 10 percent were classified as high risk [7]. Similar prevalence of high risk patients has been reported in other studies [8]. Although moderate- and high-risk patients account for only a small portion of the patient population, these families represent a significant proportion of the disease burden. As an example, a study in Utah found that while only 14 percent of the population had a positive family history for coronary heart disease, these families accounted for 48 percent of all the coronary heart disease in the state, and 72 percent of all early onset disease [9].

Despite the potential significance of the family history, one study in community family medicine practices found that family histories were only discussed in 51 percent of new patient visits and 22 percent of follow-up visits [10]. A study that surveyed community health center patients concluded that those who were more likely to talk with doctors as well as family members about their family health history were more likely to report a history of cancer or heart disease, opening the opportunity for clinicians to instigate a more comprehensive family history collection [11]. However, even when the family history is discussed, the level of detail and documentation of the history are often insufficient for making adequate risk assessments [8,12,13].

Collecting the information — The goal of obtaining a family history in the primary care setting is to identify diseases that are familial and patients who are appropriate for referral to a genetic service for more thorough evaluation [3]. It is not generally necessary to collect information beyond second-degree relatives (eg, grandparents, aunts, uncles, grandchildren); more complete pedigrees can be obtained by a genetic counselor when indicated. The patient should be aware that, for these purposes, concerns relate only to blood relatives, and not relatives through adoption or marriage, although including non-biological relatives in the pedigree may be helpful for understanding family relationships and for reproductive counseling.

Although obtaining a detailed family history is often considered time consuming, a set of focused questions can yield important screening information. The American Academy of Family Physicians has

developed the SCREEN mnemonic for obtaining a family history [14]:

- **Some Concerns:** Do you have any (some) concerns about diseases or conditions that seem to run in your family?
- **Reproduction:** Have there been any problems with pregnancy, infertility or birth defects in your family?
- **Early disease, death or disability:** Have any members of your family been diagnosed with a chronic disease at an early age or have members of your family died at an early age?
- **Ethnicity:** How would you describe your ethnicity? Or what country did your ancestors come from?
- **Non-genetic conditions:** Are you aware of any non-medical conditions or risk factors, like smoking or problem drinking that are present in your family?

Additional questions related to family size (eg, "How many brothers and sisters do you have?"; "Do you come from a large family?") and consanguinity (eg, "Were your parents, or either set of grandparents, related to each other prior to marriage?") are often helpful in interpreting the responses to the screening questions.

Patients may be unaware of precise medical diagnoses in family members, but can provide revealing information regarding symptoms and medication use. As well, patients may neglect to mention the medical conditions of siblings who have died (often the most informative information). Direct questions addressing these issues are also helpful.

Several tools are also available to aid in the collection of family history information (table 1). There is no evidence for the best way to obtain a family history in terms of instrument (using a form or interview) or format (open or closed questioning). Several institutions are developing and testing a variety of tools, especially those that integrate decision support for medical management and can be incorporated into an electronic medical record. While there are advantages to using a family health history, such as lower costs and easy acceptability, there are myriad challenges, and the utility as a public health tool is still to be determined [15-18].

When possible, patients should be encouraged to complete written or online family history questionnaires prior to their initial visit. Collecting the family history in advance of the visit allows patients to contact relatives and verify information. Clinicians can then efficiently review the information at the time of the visit. Both the maternal and the paternal history should be obtained.

The family history should be clearly and completely documented in the medical record, whether collected during the session or as part of a patient-completed intake form. Appropriate documentation should include the following components:

- Condition(s) reported
- Relationship of affected individual to the patient
- Age of relative at onset of condition

A useful method for documentation is a medical pedigree or family tree. The tree should be centered on the patient, and extend upward to include parents, aunts, uncles, and grandparents; laterally to include siblings and cousins; and downward to include children and grandchildren. Individuals are denoted with boxes (males), circles (females) or diamonds (unknown gender). A deceased individual is denoted by a strikethrough, and members who are affected by a specific condition are shown by shading part or all of their pedigree symbol. Current age (or age at death) and medical conditions are listed below each symbol. In families where more than one major medical condition is segregating, it is important to assure that shading is consistent for each condition and a legend is created to explain the shade marks. Alternately, it may be worthwhile to generate more than one pedigree (eg, one for cancer-related conditions, one for heart disease, etc.). An example of a four-generation pedigree is provided (figure 1).

Pedigree representation can be downloaded from resources such as [My Family Health Portrait](#) and included in the medical record. Pedigrees typical of different Mendelian inheritance patterns are shown in several figures ([figure 2](#) and [figure 3](#) and [figure 4](#)).

The family history is dynamic and should be updated periodically and as needed as part of a diagnostic work-up. Follow-up visits should include questioning about recent deaths, births, or new diagnoses of significant conditions in family members.

History obtained by a genetic counselor — The family history obtained by a genetic specialist is considerably more detailed than is feasible in the primary care setting. Estimates from medical genetic centers indicate that several hours may be spent collecting and verifying family history information [\[10,19\]](#), although this is not typical. This level of detail and accuracy is sometimes necessary for making precise risk estimates, estimating the likelihood of genetic mutation, and making management recommendations.

A family history taken by a genetic service is typically targeted to the condition of concern and goes back (and forward) at least three generations: the patient's parents, siblings and children; grandparents, grandchildren, aunts and uncles; nieces and nephews, grandchildren and first cousins. Information to be collected includes current age, health status, age at death and cause of death, and medical diagnoses with related environmental exposures.

When a pattern of disease is noted, more focused questions are asked to assess for features of possible syndromes and direct the assessment. Medical record review of family members can confirm diagnoses when there is uncertainty or resolve inaccuracies if family members provide differing information [\[3,20\]](#).

Interpreting the history — Information in the family history should be evaluated for predictors of inherited risk. When conditions are common in the general population, such as breast cancer (1 in 8 lifetime risk for women), or diabetes (1 in 6 individuals over the age of 60), clinicians will often encounter families with multiple affected relatives.

Individuals can be classified into three risk categories (average, moderate, or high) for most common adult onset conditions. Criteria based on the family history have been proposed for categorizing an individual's risk for common diseases ([table 2](#)) [\[7\]](#).

Key factors that suggest that a genetic disorder is present include:

- Multiple affected individuals in multiple generations from either side of the family
- Occurrence of the disease at an earlier age than usual
- Close degree of relatedness (ie, first degree or second degree relative) of affected relatives
- Presence of associated conditions in the family (eg, the association between breast and ovarian cancer or colorectal and endometrial cancer)
- Unusual presentations of common conditions (eg, bilateral disease in breast cancer, breast cancer in men)
- History of consanguinity (diseases caused by rare recessive mutations are more common in families with matings of related individuals). Consanguinity is generally more relevant in the pediatric evaluations than for conditions with adult-onset.

Additionally, the American Academy of Family Physicians has developed a table, based on the mnemonic GENES, for identifying red flags for hereditary conditions ([table 3](#)) [\[21\]](#):

- **G**roup of congenital anomalies
- **E**xceptional presentation
- **N**eurodevelopmental delay
- **E**xtrême pathology

- **Surprising laboratory values**

This is most relevant for pediatric evaluations, but may be useful in adult onset conditions as well.

Patterns of inheritance — A family history may display patterns consistent with a strong genetic influence. (See "[Overview of Mendelian inheritance](#)".)

Evaluation of the family history, referred to by geneticists as pedigree analysis, can suggest the mode of inheritance:

- Autosomal dominant disorders are associated with multiple generations of affected relatives, including both males and females ([figure 2](#)). Male to male transmission may be observed.
- Autosomal recessive disorders show a pattern in which an individual's siblings are more likely to be affected than one's parents ([figure 3](#)).
- X-linked recessive disorders typically affect only males related to each other through their mothers. Male to male transmission is not observed ([figure 4](#)).

Limitations — There are several limitations to relying on family history to identify high risk families. These include:

- Small family size or gender underrepresentation — These factors may limit the ability to recognize patterns in the pedigree. As an example, small families or families with few women may fail to exhibit a strong pattern of breast and ovarian cancer even if a genetic mutation in the *BRCA1* or *BRCA2* genes is present [[22](#)].
- Uncertainty in parentage — Pedigrees will be inaccurate if parentage is incorrectly identified, or incomplete if parentage is not known. Population-based studies have found that up to 15 percent of subjects may misidentify parentage [[23](#)].
- Disease prevention — Preventive practices (eg, removing premalignant polyps at colonoscopy, prophylactic oophorectomy, or treatment for high cholesterol levels) may prevent at-risk relatives from presenting with the illness.
- De novo mutations and incomplete penetrance — Past generations will not display traits of de novo mutations. Incomplete penetrance can obscure the genetic pattern represented in pedigrees.

These limitations should be kept in mind when a patient presents with an unusual diagnosis, such as colon cancer at age 40, with a negative family history, and a genetic basis should not be ruled out [[22](#)].

INDICATIONS FOR GENETIC SERVICES — Genetic counseling is not necessarily tied to genetic testing, although genetic testing may be offered as part of the evaluation. Once an initial family history is collected, the individual's family history should be categorized as average, moderate, or high risk ([table 2](#)). Individuals presenting with strong family histories for conditions with a known genetic basis should be offered genetic counseling. If genetic testing (ideally of an affected family member) identifies a mutation, other at-risk family members can be tested subsequently. For many conditions, specific screening and risk reduction strategies can be implemented for individuals at high risk. (See "[Genetics of Alzheimer disease](#)", section on 'Genetic testing'.)

In contrast to the varied beneficial diagnostic, prognostic, and (often) therapeutic implications of genetic testing in well-characterized Mendelian diseases, there is currently little value of genetic testing for more common polygenic diseases (like asthma or Alzheimer's disease). Though conditions have a heritable component, and often susceptibility variants have been mapped, the predictive utility of these variants is very low, as disease alleles rarely confer more than 5 to 10 percent increased susceptibility. Moreover, with rare exception, knowledge of a patient's genotype at these loci does not lead to changes in

management, limiting the utility of these variants in clinical practice. In contrast, common variants predictive of pharmacological response (pharmacogenetic variants) are being introduced clinically in some settings.

However, genetic testing for monogenic diseases may still be worthwhile even if therapeutic interventions are not available. The lack of complete understanding of the genetics of a condition does not preclude patients from having questions or concerns. For anxious families or individuals, a genetic consultation can provide education and counseling to help people understand and cope with risk information and make management decisions based on available options.

Potential benefits to genetic counseling where genetic risk is uncertain include:

- Individuals often have mistaken understanding of their risk of developing a disease; genetic counseling can help patients understand a more appropriate risk number and how to manage this information [24].
- Research or registry opportunities may exist that will allow early access to new genetic information and medical management or prevention opportunities for identified individuals or families [25].
- Some families may choose to bank DNA for the possibility of future genetic testing even if a research registry is not available.

Moderate risk — Individuals falling into the moderate risk category by family history ([table 2](#)) often have a two- to three-fold increased risk for disease. The increased risk can be influenced by a combination of lifestyle, environmental and low penetrance genetic factors that are shared within families.

In general, a genetic evaluation is not necessary for individuals who fall into the moderate risk range, unless the family is particularly anxious. However, it remains important for the clinician to take a moderate family history into account when planning care. For example, a patient whose father had colon cancer at age 55 would fall into the moderate risk category, with an estimated two-fold increased risk for colon cancer [26,27]. Colonoscopy for this patient would be recommended at age 40 years rather than the usual 50 years of age. (See "[Screening for colorectal cancer in patients with a family history of colorectal cancer](#)".)

High risk — The genetic contribution to risk for individuals in the high-risk category is strong, and these individuals may have a 50 percent or greater lifetime risk for developing disease. For inheritable conditions that are subject to population-based screening (eg, colorectal cancer or breast cancer), onset of the condition in high risk families often occurs at an earlier age than screening onset recommended for the general population.

Patients classified as high risk are appropriate candidates for a full genetic evaluation. However, even when high risk families are identified, referrals for genetic services are often not made. A study of 387 consecutive patients seen in a gastroenterology practice at an academic medical center found that only 17 percent of those reporting a family history meeting guidelines for evaluation for Lynch syndrome were referred for genetic services [12].

Barriers to referral may include [28]:

- Perceived lack of benefit from genetic information
- Uncertainty about how to incorporate genetic information into clinical management
- Lack of clinician awareness that a cluster of diverse cancers may relate to a genetic condition
- Lack of awareness of the availability of genetic tests or genetic counselors
- Patient reluctance to undergo genetic assessment
- Fear that something harmful will be found

- Cost and reimbursement issues

Uncertainty about genetic basis — Some clinicians may be uncertain whether there is sufficient information about the genetic etiology of a condition for a referral to be useful. The rapid rate of gene discovery following the mapping of the genome and the advent of genome-wide association studies and next-generation sequencing make it challenging to keep informed.

Access to genetic test information — The [Genetic Testing Registry \(GTR\) website](http://www.ncbi.nlm.nih.gov/gtr/) is a resource for medical genetic information that lists available genetic tests for thousands of conditions, with a directory of certified clinical laboratories specializing in genetic testing (choose “tests” from the homepage <http://www.ncbi.nlm.nih.gov/gtr/>). As of June 2013, the site lists over 2700 diseases for which a genetic basis has been established and for which clinical genetic testing is available. This number represents a doubling of tests that were available in 2002. This rapid rate of expansion of clinical genetic information is anticipated to continue for the foreseeable future.

CONTENT OF GENETIC COUNSELING — As defined by the National Society of Genetic Counselors, genetic counseling is the process of helping people understand and adapt to the medical, psychological and familial implications of genetic contributions to disease [29]. This process integrates the following:

- Interpretation of family and medical histories to assess the chance of disease occurrence or recurrence.
- Education about inheritance, testing, management, prevention, resources and research.
- Counseling to promote informed choices and adaptation to the risk or condition.

The process of genetic counseling is about sharing information. It involves obtaining the family history and medical information from the patient and referring clinician, obtaining and reviewing medical records (of the patient and family members) to resolve uncertainties, understanding the patient's perception of risk of developing the condition, and his or her perspective on obtaining more certain information. The genetic counselor then formulates and presents a risk assessment that includes:

- What genetic testing is available and what test results may mean for the patient
- How to modify that risk, if possible
- How to deal with the information as it pertains to medical management, emotional response, and family dynamics

The majority of genetic consultations take place over one to three visits. Most initial genetic consultations are provided in person. Some centers may provide genetic test results by telephone or other service delivery models, such as televideo.

The initial visit focuses on reviewing information and providing risk assessment. Many centers also obtain information about the psychosocial impact of the condition on the family, with particular regard to experiences in caregiver roles or early deaths. When appropriate, the option of genetic testing is discussed.

If genetic testing is pursued, there will usually be a follow-up visit or telephone call to discuss results and management implications. Additional follow-up visits may be scheduled to assist the patient in better understanding the implications of the result and help to make management decisions. For patients who are not appropriate candidates for genetic testing, or who elect not to be tested, risk estimates and management recommendations will be made on the basis of personal and family history.

Most genetic services are in the form of consultations to the primary care provider, and typically patients will be referred back to their physicians with recommendations for their follow-up care. However, some specialist clinics will continue to follow patients at regular intervals.

Preparation for referral — To alleviate patient anxiety and better prepare the patient (for example, if additional family health history is useful), the primary care clinician needs to discuss with the patient that the referral is being made, describe what features in the family prompted the referral, and provide a description of what the visit will entail [30]. Many genetics services send questionnaires to the family prior to their initial visit.

Where to refer — The [Genetic Testing Registry website \(http://www.ncbi.nlm.nih.gov/gtr/\)](http://www.ncbi.nlm.nih.gov/gtr/) provides an international listing of clinics (by state, for the United States), as well as labs in Argentina, Australia, Austria, Belgium, Brazil, and Canada for counseling and/or diagnostic testing services for adult genetics, biochemical genetics, cancer genetic risk assessment, pediatric genetics, preimplantation diagnosis, prenatal diagnosis, and telemedicine.

Risk assessment — Before the first session, if sufficient information is available, a risk assessment is performed. Several models, some based on empiric data and others based on computer algorithms, can be used to determine risk of developing the condition or risk of having a particular mutation. These models are relatively well developed for many cancer syndromes (including breast, ovarian, and colorectal) and are being evaluated for cardiovascular disease, asthma, diabetes and some neurological diseases. (See ["Risk prediction models for breast cancer screening"](#) and ["Screening for colorectal cancer in patients with a family history of colorectal cancer"](#) and ["Risk factors for prostate cancer"](#) and ["Estimation of cardiovascular risk in an individual patient without known cardiovascular disease"](#) and ["Risk factors for type 2 diabetes mellitus", section on 'Prediction models'.](#))

Once an objective risk figure has been determined, it is important to correlate this with the patient's belief of his or her risk. The patient's personal experience with affected family members may significantly impact his or her perception of personal risk. When the patient's beliefs are not concordant with the calculated risk, the primary care provider and genetic counselor need to provide ongoing communication and counseling to help the patient understand the implications of the risk figure.

The initial risk figure is subject to modification. Risk may be modified by additional family history information, lifestyle changes or medications that increase decrease risk for particular conditions. (See ["Cancer prevention"](#).)

Screening is a means of early disease detection and does not modify the risk, but can modify treatment options and the course of the disease if the condition is identified at an early stage. More aggressive screening (earlier, more frequent, other modalities) may be indicated for targeted conditions. As an example, for patients with familial adenomatous polyposis, screening colonoscopy is recommended beginning at age 10 to 12. (See ["Familial adenomatous polyposis: Screening and management of patients and families"](#).)

Information to send with the referral — Once a decision is made to refer a patient to a genetic counselor, the referring clinician should provide the following information to the consultant:

- Specific questions to be addressed by the referral. As an example, "This patient is at high risk for breast cancer due to her family history that includes both thyroid cancer and early onset breast cancer. What testing is recommended for her?"
- A summary of available family history information, either in a table or pedigree format, including affected family members up to two generations removed from patient, indicating their age at diagnosis and current age.
- A summary of the patient's pertinent medical history, including laboratory test results or biopsy reports, where relevant.

PRINCIPLES OF GENETIC TESTING — Genetic testing for adult-onset conditions serves two principal purposes: to confirm or establish a diagnosis, or to assess risk for asymptomatic patients.

Knowledge of an increased risk for a particular disease may lead to changes in patient behavior or clinical interventions to decrease the risk. Management strategies to decrease risk range from increased sun protection, to enhanced screening, to risk-reducing surgery or chemoprophylaxis, with varying levels of effectiveness in improving outcomes. Examples include the following:

- (See "[Cancer prevention](#)".)
- (See "[Chemoprevention strategies in prostate cancer](#)".)
- (See "[Selective estrogen receptor modulators and aromatase inhibitors for breast cancer prevention](#)".)
- (See "[Management of hereditary breast and ovarian cancer syndrome patients with BRCA mutations](#)".)
- (See "[Primary prevention of melanoma](#)".)
- (See "[Familial adenomatous polyposis: Screening and management of patients and families](#)".)
- (See "[Lynch syndrome \(hereditary nonpolyposis colorectal cancer\): Screening and management](#)".)
- (See "[Peutz-Jeghers syndrome and juvenile polyposis: Screening and management of patients and families](#)".)

Genetic testing for germline mutations can be conducted on virtually any tissue. Most laboratories prefer blood specimens, although cheek (buccal) swabs and saliva samples also may be an option for certain types of genetic testing.

Purpose of genetic testing — Genetic testing can be done for diagnostic or predictive purposes ([table 4](#)). Additionally, testing can be performed at different times during a patient's life. Examples include preimplantation testing; prenatal screening and testing; newborn screening; and testing at some point after birth to establish a diagnosis, determine carrier status (for autosomal recessive conditions), or for predictive purposes.

- **Preimplantation testing** — Preimplantation testing is performed on embryos resulting from in vitro fertilization for high-risk couples. Identification of chromosomal abnormalities or mutations allows selection of unaffected embryos to implant. (See "[Preimplantation genetic diagnosis](#)".)
- **Prenatal testing** — Prenatal testing identifies conditions in which a mutation has been identified in a parent. Typically, the gene being analyzed is inherited in an autosomal dominant manner, so each fetus has an a priori 50 percent risk of inheriting the mutation. Specimens for analysis are obtained either by chorionic villus sampling or amniocentesis.

The goals of prenatal counseling and testing are to provide reassurance or guide and support decisions regarding whether to continue or terminate a pregnancy that may result in an affected individual. Prenatal testing for a condition such as Huntington disease, which is fully penetrant, carries different implications than testing for a condition that confers an increased risk of developing the disease, such as colon cancer. (See "[Basic principles of genetic counseling for the obstetrical provider](#)".)

- **Diagnostic testing** — Diagnostic genetic testing is used to support a specific diagnosis in a patient with an existing condition. There are over 2700 conditions for which genetic testing can be offered. Most of these conditions are relatively rare and manifest shortly after birth or in childhood. Some, such as Huntington disease, do not manifest until later in life, at which time genetic testing would also be appropriate.

Another example of diagnostic testing is offering full sequence analysis of the *BRCA1* and *BRCA2* genes to a 32-year-old woman recently diagnosed with breast cancer who wants to make surgical decisions. Diagnostic testing of an individual with multiple polyps can distinguish between Lynch

syndrome and attenuated familial adenomatous polyposis, and therefore provide information regarding risks of extracolonic manifestations.

- **Predictive testing** — Predictive testing is a method of risk assessment for unaffected individuals who are at risk for developing conditions with a heritable component. The value of a gene test is influenced by the penetrance of the gene (whether the presence of a gene makes it highly likely that the person will develop the related phenotype) and whether there are effective prevention or early treatment strategies to impact the course of the condition. (See ['Penetrance, ascertainment bias, and variable expression'](#) below and ['Validity and utility'](#) below.)
- **Carrier testing** — Carrier testing is used to identify asymptomatic individuals who are heterozygous (ie, carry one copy of an altered gene) for a mutation that causes a genetic disorder in homozygotes (individuals who carry two altered copies of a gene). Carrier testing can be performed in families with a family history for a specific genetic condition, or in particular ethnic groups or populations in which risk mutations are at relatively high prevalence (eg, Tay Sachs screening among Ashkenazi Jews). (See ["Prenatal screening for genetic disease in the Ashkenazi Jewish population"](#).)

Penetrance, ascertainment bias, and variable expression — The predictive value of a positive genetic test can be limited by the lack of population-based data on penetrance (likelihood of developing disease given inheritance of a disease-causing mutation) and variable expressivity (phenotypic variations in the way the disease is expressed).

Genetic mutations are initially studied in high-risk families identified because of their high expression of disease (ie, high penetrance). This practice overestimates the penetrance that later is found in more population-based studies. As an example, while high-risk families with a genetic mutation in *BRCA1* show an 85 percent lifetime risk of developing breast cancer, population-based studies report a 57 percent lifetime risk of developing cancer in individuals with this mutation [31,32]. Cancer risk in some individuals with a *BRCA1* mutation may be further decreased as they seek more aggressive interventions, such as a risk-reducing surgery [2,31-38]. (See ["Genetic testing for hereditary breast and ovarian cancer syndrome"](#).)

Ascertainment bias is reflected not only in disease penetrance, but also in variability in the expressivity of a condition. Initial discovery of mutations often derives from high risk families with prominent impact of the disease. For example, in the original research studies, families eligible to be considered for Lynch syndrome were defined by the Amsterdam criteria that required three or more affected individuals, including two who were closely related, and one with colon cancer under the age of 50. When a population based sample of colon cancer patients was evaluated for Lynch syndrome, however, only 13 percent of mutation carriers met these criteria [39,40].

Variable expressivity of the phenotype will result in a wide range of severity, age of onset, or progression of disease. Some people with two mutations in the gene responsible for causing hemochromatosis (HFE), for example, do not go on to develop the complications of iron overload, even in the absence of treatment [41]. (See ["Genetics of hereditary hemochromatosis"](#).)

Estimates of risk based upon high-risk populations are greater than the risk estimated in those identified by population-based screening. When providing risk information, the literature should be reviewed to find the studies conducted with participants that most closely match a patient's situation. In many cases, patients will need to be presented with a potential range of risk estimates and an explanation of the reason for this range.

Screening and management options for identified mutations have often been developed for high-risk families. The efficacy of these therapies for individuals at lower risk may be unproven [38]. Genetic consultation can play an important role to assure that highest-risk individuals benefit from appropriate

management, while lower risk individuals not receive unnecessary intervention in the absence of applicable evidence for improved outcomes.

Validity and utility — Clinical validity refers to the ability of a genetic test to predict a disease phenotype. For some tests, such as mutational analysis in families with familial adenomatous polyposis, the clinical validity is very high. Virtually all individuals who test positive for a mutation in the APC gene will develop features of the disease [42]. In contrast, individuals who test positive for the APOE e4/e4 genotype have approximately a 30 percent chance of developing Alzheimer's disease, and most individuals with this genotype will not develop the disease [43].

The clinical utility of a genetic test refers to the impact of the testing information on clinical care. While the genetic test for Huntington has high clinical validity, the clinical utility is limited because of few treatment or intervention options available for those who test positive. However, family members at risk for this disease may choose to undergo gene testing to make reproductive plans or for the psychological benefit in relieving uncertainty and facilitating future planning [44].

Determining the clinical utility involves considering the severity of the disease, the options for management, and the likelihood that genetic testing will change the course of treatment. The CDC provides an [online resource](#) with a list of questions to ask in order to assess the clinical validity and utility of a genetic testing ([table 5](#)) [45].

Ethical, legal, and psychosocial issues — Clinicians must consider the benefits and risks of genetic testing including the possibility of false positive and false negative results. Genetic information may have unique risks based upon the meaning attributed to a disorder being genetic and the value of the information for family members [46].

Psychosocial issues — Because of the familial nature of genetics, the psychosocial issues require not just an understanding of the patient, but of the dynamics of the patient's family. Many genetic counselors use the family history intake to assess the family dynamics and more recently, to incorporate outside influences that impact the social support of the patient [47,48].

Patient autonomy is a central tenet in genetic counseling, and assuring that the patient has sufficient support to make decisions is vital. This is encouraged by providing accurate and appropriate information in an empathic environment. Understanding the potential responses to difficult or unexpected news and being able to respond in a productive manner is essential to long-term acceptance of genetic information and management options.

In the absence of proven beneficial interventions, the value of genetic testing for an individual may be the relief derived when a test is negative in the face of a known mutation in an affected family member. On the other hand, some patients receiving negative results experience "survivor guilt" if other family members are found to have the mutation. Family dynamics can be affected. As an example, some family members found not to have the gene mutation or Huntington's disease may be rejected by affected family members due to the loss of a common risk status [49,50].

Genetic discrimination — A common concern about genetic testing is the potential for inadequate protection of privacy of genetic information, and the potentially negative impact of genetic testing on employment and insurance coverage. The concern about genetic discrimination affects not only the individual undergoing testing, but also can extend to their unaffected family members [51,52]. Fear of genetic discrimination is a common reason for declining genetic testing [53].

In the United States, legal protection against the use of genetic information by group health insurance plans and employers is available at the federal level through the Genetic Information Nondiscrimination Act (GINA) [54]. Enacted in 2008, GINA prohibits discrimination by health insurers and employers on the basis of genetic information. GINA prohibits the use of genetic information to determine insurance

eligibility, coverage, underwriting, or premium charges. GINA also prohibits health insurers and employers from asking or requiring a person to take a genetic test [55,56].

GINA has the following limitations [57]:

- It does not address the use of genetic information in other insurance markets (eg, long-term care, life, and disability insurance) [58]. The regulation of insurance markets outside of health insurance is determined at the state level [59].
- It does not apply to employers with fewer than 15 employees, the US military (and the TRICARE military health system), the Indian Health Service, the Veterans Health Administration, or the Federal Employees Health Benefits Program [58].

Since its enactment in 2008, there have been few challenges to GINA. However, the majority of individuals remain unaware of the legislation, and those who are aware of it have several misconceptions about the specific protections and limitations associated with it [57]. An online survey of 1479 people conducted in June 2014 revealed that 79 percent of respondents were not familiar with GINA and of those who claimed to be familiar with it, only a minority could correctly describe the protections afforded by the legislation [57].

Individuals who are disabled by a genetic disorder are protected from discrimination by the Americans with Disabilities Act (ADA). (See "[Disability assessment and determination in the United States](#)".)

Most states in the United States also have additional legislation protecting against the use of genetic information by health insurers and employers [46,60].

The United States Affordable Care Act (ACA) also prohibits variations in health insurance premiums based on health status and genetic information [61].

Legislation and policies regarding the use of genetic information in insurance and employment outside the United States varies considerably by country [62,63].

- Although the potential for genetic discrimination in healthcare is considerably less in nations with government-sponsored healthcare systems, the British government has issued a moratorium prohibiting insurers from using genetic test results to set premiums for certain life insurance, long-term care, and income protection policies that is in effect until 2017 [64,65].
- In 1997 the Council of Europe established the European Convention on Human Rights and Biomedicine (also known as the Oviedo Convention), which prohibits discrimination based on an individual's genetic background.
- Over 30 European countries, Australia, and Canada have also implemented some form of anti-discrimination legislation or moratorium, or signed on to abide by the standards of the Oviedo Convention [66].
- Legal protection against genetic discrimination is lacking in other areas of the world, and fear of stigmatization in these countries may significantly hinder access to genetic services and research [67].

Disclosure to family members — Genetic risk, whether assessed by genetic testing or family history, raises the potential ethical or legal obligation of a physician to disclose genetic information to other at risk family members. Most patients are willing to provide consent for such disclosure, or readily do it themselves [68,69]. For some, their express purpose for genetic testing is to make information available to family members.

However, a patient's refusal to voluntarily disclose raises difficult conflict between two competing ethical

obligations: the duty to protect patient privacy and autonomy versus the duty to disclose for the purpose of preventing future harm. Legal opinions vary regarding the physician's duty to warn at-risk family members about a familial genetic condition. In the limited number of cases involving genetic information, the courts have affirmed a physician's legal responsibility to warn family members [70]. Despite numerous guidelines and policy papers, there is currently no clear consensus or set of legal guidelines available in most countries. Until such time as clear legal obligations are defined, physicians should consult with medical geneticists, genetic counselors, or bioethicists in the event that patients do not provide consent to share information.

Gene test uptake — Geneticists refer to the percentage of patients who obtain a test when it is offered as uptake, especially when offering a genetic test to an otherwise healthy individual. The above factors (penetrance, utility, ethical issues) are some of the reasons why there is varied use of DNA predictive testing for different condition by patients. Other reasons include a sense that the test is not urgent or the risk perception is not significant enough to warrant testing. Finally, finances play a role in predictive genetic testing uptake, as insurance companies vary in their reimbursement policies.

PRACTICAL ISSUES IN GENETIC TESTING — Genetic testing is currently available for over 2700 conditions [71]. An overview of genetic testing and specifics of available genetic tests is available online at [Genetic Testing Registry \(http://www.ncbi.nlm.nih.gov/gtr/\)](http://www.ncbi.nlm.nih.gov/gtr/). (See "[Tools for genetics and genomics: Cytogenetics and molecular genetics](#)".)

Whom to test — When possible, molecular genetic testing should initially be performed on an affected family member. For ethical reasons, genetic testing for conditions that most commonly manifest in adults is generally not done in children and adolescents, unless a preventive intervention is effective if done during those time periods.

Genetic testing of an affected individual is usually important to facilitate interpretation of testing in other family members and is usually the first test performed. Once a particular mutation has been identified, testing other family members is technically straightforward. While it is possible to begin the genetic testing process in an unaffected individual, there is a greater chance that these results will be inconclusive. As an example, if an unaffected person does not have a mutation in a particular gene, is it because that gene is not mutated in the family, or because the mutation does exist in the family and the unaffected person did not inherit it?

In some cases the affected family members are deceased or unable or unwilling to participate. Genetic counseling in these situations should evaluate the optimal testing approach in the family, facilitate family communication, and explain the limitations of genetic testing when it cannot be performed on the ideal candidate.

Informed consent for testing — Informed consent, a tenet of patient-centered medicine, is a foundation for the voluntary nature of genetic testing. Informed consent means that the patient understands and agrees to the procedure. When offering genetic testing, especially to unaffected individuals, most clinicians follow language from the Department of Health, Education, and Welfare, which encourages the patient to voluntarily exercise free power of choice, without any element of force, fraud, deceit, duress, over-reaching, or other ulterior form of constraint or coercion [72]. (See "[Informed consent](#)".)

Consent, therefore, involves discussion, and not just the signing of a document. Discussion should include the procedure; the risks, benefits and limitations; the potential results; and the management options that can be available based on the result.

Two exceptions exist to the general suggestion to obtain informed consent for genetic testing. Newborn screening in the United States does not involve active consent, but parents who do not want their child tested can sign a form that indicates their refusal. Additionally, in some clinical pediatric circumstances,

genetic testing is considered part of the diagnostic evaluation, so less of an effort may be made to obtain fully informed consent.

Samples for testing — There are many techniques for genetic testing, and the best approach depends on the gene to be evaluated. Testing may be DNA-based, cytogenetic, or metabolic, depending upon its purpose ([table 6](#)). The methods for obtaining the specimen (cheek swab, blood or tissue sample), the costs, and the waiting time for results (ranging from days to weeks or months) differ according to the test involved.

For some conditions, such as some thrombophilias or hemochromatosis, testing may involve analysis of clotting factors or serum iron studies, rather than direct DNA analysis. (See ["Screening for inherited thrombophilia in asymptomatic individuals"](#) and ["Evaluating patients with established venous thromboembolism for acquired and inherited risk factors"](#) and ["Screening for hereditary hemochromatosis"](#).)

Insurance reimbursement — In the United States, insurance reimbursement is becoming more common for genetic testing, as testing becomes a more routine part of clinical care. Coverage of genetic testing by insurance will vary depending on the type of test, whether it is being done to make a diagnosis in an affected person or to predict risk in an unaffected individual, and on the patient's specific type of insurance coverage. A study that reported on findings of whole-exome sequencing for 250 patients (mostly children with neurologic phenotypes, all of whom had undergone prior genetic testing) found that insurance coverage for the costs was similar to other genetic testing (eg, reimbursement for the majority of tests) [[73](#)].

Single-gene genetic testing is commonplace in many countries with nationalized public healthcare systems, as are perinatal mutation screening programs. The United Kingdom is investing in strategies to implement whole exome sequencing into clinical practice [[74](#)].

Where to test — Some genetic tests for common conditions, such as Factor V Leiden, are offered through many laboratories, while other genetic tests are more specialized and may only be available from one or two clinical laboratories.

In some cases, testing is only available in a research setting. Most research laboratories are not Clinical Laboratory Improvement Amendments (CLIA) certified and not authorized to provide results back for the purpose of clinical decision making. Therefore, when pursuing testing through a research study, it is important to clarify with the investigator the exact testing that will be done on the sample, if and how results might become available, and the expected turnaround time.

For conditions in which genes have not yet been identified or there is a lack of evidence of an effective intervention, preservation of genetic material from the affected individual by DNA banking for future genetic testing is a reasonable option. This involves the storage of DNA material (usually blood but occasionally tissue samples) for future testing should it become desired. Local laboratory specific information on DNA banking is available through [Genetic Testing Registry](#) [[75](#)]. In addition, there are several laboratories that offer this service direct to the consumer.

Outcomes of testing — There are typically three outcomes from genetic testing: the causative mutation may be identified; no mutation is identified, or a variant of uncertain significance is identified.

Finding a causative mutation can confirm a diagnosis in the affected person. Clarification of the genetic etiology of the proband's disease may aid in planning treatment or in predicting the risks for other health problems. Identification of a mutation also allows for genetic testing to be offered to at-risk family members.

Once a mutation has been identified, other family members can usually be tested for the specific genetic alteration with close to 100 percent accuracy. The cost of genetic testing is often reduced once the

specific genetic mutation has been identified. Those family members who have inherited the genetic mutation may be candidates for enhanced screening or risk reduction strategies, while those who have not inherited the alteration can be spared unnecessary procedures.

An inherited risk is not ruled out when a genetic mutation is not identified. A negative result may mean that the genetic cause cannot be identified with current technologies. The sensitivity of genetic testing varies, with tests for Huntington's disease and factor V Leiden having a sensitivity greater than 99 percent, while tests for other syndromes such as Marfan syndrome have only a 70 percent sensitivity [76-78]. For many conditions, only some of the contributing genetic factors have been identified. Mutations in the CDKN2A gene, for example, account for 20 to 40 percent of families with hereditary melanoma, suggesting that there are likely other loci that are contributing to risk [79].

Determining whether a particular variant disrupts the function of the gene and is related to risk for disease, or is a polymorphism in the gene with no clinical implications, can be difficult. When variants of uncertain significance are identified, risk estimates should be made on the basis of the family history, taking into account the laboratory report. An inherited risk should neither be ruled out nor in on the basis of this uncertain finding. It is also important to let the family know that most laboratories continue to research variants of uncertain significance, and in some cases, they can be reclassified. Having a plan in place to communicate and follow-up with the individual or family can be important.

GENOME SEQUENCING AND GENOMIC TESTING

Uses of genome sequencing — Whole genome sequencing, exome sequencing, and genotyping are beginning to change the paradigm of genetic testing. (See "[Principles and clinical applications of next-generation DNA sequencing](#)" and "[Genetic association studies: Principles and applications](#)".)

These technologies provide sequence information about genetic variation across an individual's entire genome rather than a specific disease gene, and they can provide information about the risks for many different diseases simultaneously. Genomic testing can also involve looking for single-nucleotide polymorphisms (SNPs) that have been found to be associated with a risk for disease. (See "[Glossary of genetic terms](#)" and "[Overview of genetic variation](#)", section on 'Single nucleotide polymorphisms (SNPs)').

Whole genome and exome sequencing strategies have been used successfully on a research basis to identify the genetic basis of several rare diseases [80,81]. However, incorporating whole genome or exome data into clinical care is just beginning. Motivations for incorporating genomic testing into clinical care include a desire to learn more about the genetic factors that may be contributing to one's risk for disease and to help forward research [82]. Research suggests that patients are interested in receiving information about their genomes. The ClinSeq project, sponsored by the United States National Institutes of Health (NIH), is the first large-scale study using whole genome sequencing and returning results to participants [83].

Consent for genome sequencing — Additional research is needed to determine how to consent individuals appropriately for this type of testing, how to return information in a meaningful way, and how to store and analyze genetic data. Whole genome or exome sequencing requires substantial informatics resources and expertise to analyze the many nucleotides to identify genetic variants that will have clinical relevance. Also many genetic variants identified through this technology will be unclassified, so their effect on gene function is unknown. (See "[Principles and clinical applications of next-generation DNA sequencing](#)", section on 'Interpretation'.)

Genome sequencing has the potential for a much wider range of findings than traditional genetic testing, including carrier status for rare recessive disease that may impact reproductive decision making, genetic predisposition to diseases for which there are actionable steps for reducing risk, genetic predisposition for disease for which there are no known approaches for modifying risk, and potential pharmacogenetic

interactions (eg, altered metabolism of a drug that increases or decreases its efficacy).

Further, the process for consenting patients for this type of testing will differ from testing targeted for a known condition. Patients will need to be informed about the variety of information that may come from this type of testing, and novel counseling approaches may need to be developed to convey this vast range of information to patients effectively and efficiently.

Disclosure of incidental findings from genome sequencing — The most controversial aspect of genomic testing involves the amount of information that should be returned to the patient, especially if the clinical validity and utility are not fully understood. This subject is discussed in detail separately. (See ["Incidental findings from genetic testing"](#).)

MANAGEMENT AND FOLLOW-UP — The long-term management of patients who have inherited an increased risk for disease may involve more aggressive screening initiated at an earlier age than in the general population, counseling for lifestyle modifications, initiation of pharmacologic or surgical interventions, and management of related psychosocial issues.

The primary care clinician will play a crucial role in helping patients adhere to guidelines over the long-term and in coordinating care between multiple specialists. Many genetic conditions are associated with multiple risks. For example, individuals with Lynch syndrome have an increased risk for multiple cancers and will need appropriate screening for cancers not present in the proband family member [84]. (See ["Lynch syndrome \(hereditary nonpolyposis colorectal cancer\): Clinical manifestations and diagnosis"](#) and ["Lynch syndrome \(hereditary nonpolyposis colorectal cancer\): Screening and management"](#) and ["Endometrial and ovarian cancer screening and prevention in women with Lynch syndrome \(hereditary nonpolyposis colorectal cancer\)"](#).)

While informing a patient about the risks to one's relatives may be sufficient to satisfy the physician's duty for family disclosure, this may not be sufficient to prepare the patient to take appropriate action. Genetic services often provide patients with written materials to facilitate the sharing of information. Primary care physicians can help play a role in checking in with patients to see how the information is flowing through the family and identifying possible communication barriers [3].

ADDITIONAL RESOURCES — The field of genetics is rapidly evolving. A number of online resources are available to help practitioners obtain up-to-date information:

- Genetic Testing Registry (<http://www.ncbi.nlm.nih.gov/gtr/>) — Directories of genetic clinics and laboratories.
- GeneReviews (<http://www.ncbi.nlm.nih.gov/sites/GeneTests/review>) — Overviews and disease-specific discussions of counseling, diagnosis and management for individuals and families with inherited disorders.
- National Cancer Institute (www.cancer.gov/cancerinfo) — Genetic components of cancer including evidence based overview to clinical genetics (PDQ).
- The Human Genome Epidemiology Network (HuGE Net) (www.cdc.gov/genomics/hugenet/default.htm) — Epidemiology and public health aspects of genetics.
- National Society of Genetic Counselors (www.nsgc.org) — The website will provide information regarding how to find a counselor as well as consumer information regarding developing a family history, and what to expect from a genetic consultation.
- Online Mendelian Inheritance in Man (www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM) — Catalog of human genetic disorders.
- Kansas University Medical Center (www.kumc.edu/gec/geneinfo.html) — Educational and clinical

information on genetics; links to many helpful web sites on genetics including those listed here.

- Genetic Science Learning Center (<http://learn.genetics.utah.edu/>) — Educational and clinical information about genetics and genetic conditions; links to many helpful web sites about genetics; exercises for teaching genetics from classroom to the clinic.

SUMMARY

- Primary care clinicians work on the front line where advances in genetic research can be directly applied to patient care. Their role, in terms of genetic counseling, is to obtain the family history, identify high-risk individuals, refer appropriate patients for genetic services, and coordinate long-term management. (See ['Role of the primary care or referring clinician'](#) above.)
- Although the family history is not a highly-predictive measure for many common diseases, it can help determine which patients are at average, moderate, or high risk for specific conditions. Key questions to ask in obtaining a family history relate to patient concerns about disease, problems with pregnancy or birth defects in the family, early deaths or disease onset, ethnicity, and determination of nongenetic risk factors for disease. Patients should be encouraged to complete written or online family history questionnaires prior to their initial visit. (See ['Family history'](#) above.)
- Patients classified as high risk are appropriate candidates for a full genetic evaluation. Individuals falling into the moderate risk category by family history generally may not require formal genetic evaluation. They often have a two- to threefold increased risk for disease related to a combination of lifestyle, environmental, and low-penetrance genetic factors that are shared within families. (See ['Indications for genetic services'](#) above.)
- The Genetic Testing Registry website (<http://www.ncbi.nlm.nih.gov/gtr/>) lists available genetic tests and relies on submitters to provide information that is accurate and free of bias. NIH makes no endorsements of tests or laboratories listed in GTR. (See ['Indications for genetic services'](#) above.)
- Genetic testing can be done for diagnostic or predictive purposes. Additionally, testing can be done to determine carrier states, as well as for prenatal testing, preimplantation testing, and newborn screening ([table 4](#)). The predictive value of a positive genetic test is limited by variable penetrance and expression. The clinical utility of a genetic test refers to the impact of the testing information on clinical care. (See ['Principles of genetic testing'](#) above.)
- Ethical, legal, and psychosocial impacts of genetic testing must be taken into consideration. The 2008 Genetic Information Nondiscrimination Act (GINA) prohibits health insurance and employment discrimination of unaffected individuals based on genetic information in the United States. (See ['Ethical, legal, and psychosocial issues'](#) above.)
- Genetic testing of an affected individual is usually important to facilitate interpretation of testing in other family members. Once a particular mutation has been identified, testing other family members is technically straightforward. (See ['Practical issues in genetic testing'](#) above.)
- Whole genome sequencing, exome sequencing, and genotyping provide sequence information about genetic variation across an individual's entire genome rather than a specific disease gene. Additional research is needed to determine how to consent individuals appropriately for this type of testing and how to return information in a meaningful way. The most controversial aspect of genomic testing involves the amount of information that should be returned to the patient. (See ['Genome sequencing and genomic testing'](#) above and ['Incidental findings from genetic testing'](#).)
- The long-term management of patients who have an inherited risk for disease may involve more aggressive screening initiated at an earlier age than in the general population, counseling for lifestyle modifications, initiation of pharmacologic or interventional preventive measures, and

management of related psychosocial issues. (See '[Management and follow-up](#)' above.)

Use of UpToDate is subject to the [Subscription and License Agreement](#).

REFERENCES

1. Scheuner MT, Sieverding P, Shekelle PG. Delivery of genomic medicine for common chronic adult diseases: a systematic review. *JAMA* 2008; 299:1320.
2. Domchek SM, Friebel TM, Singer CF, et al. Association of risk-reducing surgery in BRCA1 or BRCA2 mutation carriers with cancer risk and mortality. *JAMA* 2010; 304:967.
3. Guttmacher AE, Porteous ME, McInerney JD. Educating health-care professionals about genetics and genomics. *Nat Rev Genet* 2007; 8:151.
4. Berg AO, Baird MA, Botkin JR, et al. National Institutes of Health State-of-the-Science Conference Statement: Family History and Improving Health. *Ann Intern Med* 2009; 151:872.
5. Wilson BJ, Qureshi N, Santaguida P, et al. Systematic review: family history in risk assessment for common diseases. *Ann Intern Med* 2009; 151:878.
6. Claassen L, Henneman L, Janssens AC, et al. Using family history information to promote healthy lifestyles and prevent diseases; a discussion of the evidence. *BMC Public Health* 2010; 10:248.
7. Scheuner MT, Wang SJ, Raffel LJ, et al. Family history: a comprehensive genetic risk assessment method for the chronic conditions of adulthood. *Am J Med Genet* 1997; 71:315.
8. Frezzo TM, Rubinstein WS, Dunham D, Ormond KE. The genetic family history as a risk assessment tool in internal medicine. *Genet Med* 2003; 5:84.
9. Hunt SC, Gwinn M, Adams TD. Family history assessment: strategies for prevention of cardiovascular disease. *Am J Prev Med* 2003; 24:136.
10. Medalie JH, Zyzanski SJ, Langa D, Stange KC. The family in family practice: is it a reality? *J Fam Pract* 1998; 46:390.
11. Kaphingst KA, Goodman M, Pandya C, et al. Factors affecting frequency of communication about family health history with family members and doctors in a medically underserved population. *Patient Educ Couns* 2012; 88:291.
12. Grover S, Stoffel EM, Bussone L, et al. Physician assessment of family cancer history and referral for genetic evaluation in colorectal cancer patients. *Clin Gastroenterol Hepatol* 2004; 2:813.
13. Tyler CV Jr, Snyder CW. Cancer risk assessment: examining the family physician's role. *J Am Board Fam Med* 2006; 19:468.
14. SCREEN: Taking a Family for Familial Disease. Available at: <http://www.aafp.org> (Accessed on July 30, 2007).
15. Hulse NC, Ranade-Kharkar P, Post H, et al. Development and early usage patterns of a consumer-facing family health history tool. *AMIA Annu Symp Proc* 2011; 2011:578.
16. Valdez R, Yoon PW, Qureshi N, et al. Family history in public health practice: a genomic tool for disease prevention and health promotion. *Annu Rev Public Health* 2010; 31:69.
17. Scheuner MT, de Vries H, Kim B, et al. Are electronic health records ready for genomic medicine? *Genet Med* 2009; 11:510.
18. Feero WG, Bigley MB, Brinner KM, Family Health History Multi-Stakeholder Workgroup of the American Health Information Community. New standards and enhanced utility for family health history information in the electronic health record: an update from the American Health Information Community's Family Health History Multi-Stakeholder Workgroup. *J Am Med Inform Assoc* 2008; 15:723.

19. Bernhardt BA, Pyeritz RE. The economics of clinical genetics services. III. Cognitive genetics services are not self-supporting. *Am J Hum Genet* 1989; 44:288.
20. Mouchawar J, Laurion S, Ritzwoller DP, et al. Assessing controversial direct-to-consumer advertising for hereditary breast cancer testing: reactions from women and their physicians in a managed care organization. *Am J Manag Care* 2005; 11:601.
21. Whelan AJ, Ball S, Best L, et al. Genetic red flags: clues to thinking genetically in primary care practice. *Prim Care* 2004; 31:497.
22. Weitzel JN, Lagos VI, Cullinane CA, et al. Limited family structure and BRCA gene mutation status in single cases of breast cancer. *JAMA* 2007; 297:2587.
23. Bellis MA, Hughes K, Hughes S, Ashton JR. Measuring paternal discrepancy and its public health consequences. *J Epidemiol Community Health* 2005; 59:749.
24. Pilarski R. Risk perception among women at risk for hereditary breast and ovarian cancer. *J Genet Couns* 2009; 18:303.
25. Nee LE. Genetic counseling and presenilin-1 Alzheimer's disease: "Research Family" members share some thoughts. *Am J Alzheimers Dis Other Dement* 2007; 22:99.
26. Fuchs CS, Giovannucci EL, Colditz GA, et al. A prospective study of family history and the risk of colorectal cancer. *N Engl J Med* 1994; 331:1669.
27. Burt RW, DiSario JA, Cannon-Albright L. Genetics of colon cancer: impact of inheritance on colon cancer risk. *Annu Rev Med* 1995; 46:371.
28. Suther S, Goodson P. Barriers to the provision of genetic services by primary care physicians: a systematic review of the literature. *Genet Med* 2003; 5:70.
29. National Society of Genetic Counselors Website. Available at: www.nsgc.org (Accessed on November 16, 2009).
30. Metcalfe A, Werrett J, Burgess L, Clifford C. Psychosocial impact of the lack of information given at referral about familial risk for cancer. *Psychooncology* 2007; 16:458.
31. Ford D, Easton DF, Stratton M, et al. Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. The Breast Cancer Linkage Consortium. *Am J Hum Genet* 1998; 62:676.
32. Struwing JP, Hartge P, Wacholder S, et al. The risk of cancer associated with specific mutations of BRCA1 and BRCA2 among Ashkenazi Jews. *N Engl J Med* 1997; 336:1401.
33. Burke W, Press N, Pinsky L. BRCA1 and BRCA2: a small part of the puzzle. *J Natl Cancer Inst* 1999; 91:904.
34. Easton DF, Ford D, Bishop DT. Breast and ovarian cancer incidence in BRCA1-mutation carriers. Breast Cancer Linkage Consortium. *Am J Hum Genet* 1995; 56:265.
35. Thorlacius S, Struwing JP, Hartge P, et al. Population-based study of risk of breast cancer in carriers of BRCA2 mutation. *Lancet* 1998; 352:1337.
36. Peto J, Collins N, Barfoot R, et al. Prevalence of BRCA1 and BRCA2 gene mutations in patients with early-onset breast cancer. *J Natl Cancer Inst* 1999; 91:943.
37. Fodor FH, Weston A, Bleiweiss IJ, et al. Frequency and carrier risk associated with common BRCA1 and BRCA2 mutations in Ashkenazi Jewish breast cancer patients. *Am J Hum Genet* 1998; 63:45.
38. Welch HG, Burke W. Uncertainties in genetic testing for chronic disease. *JAMA* 1998; 280:1525.
39. Hampel H, Frankel WL, Martin E, et al. Screening for the Lynch syndrome (hereditary nonpolyposis colorectal cancer). *N Engl J Med* 2005; 352:1851.
40. Hampel H, Stephens JA, Pukkala E, et al. Cancer risk in hereditary nonpolyposis colorectal cancer

syndrome: later age of onset. *Gastroenterology* 2005; 129:415.

41. Cogswell ME, McDonnell SM, Khoury MJ, et al. Iron overload, public health, and genetics: evaluating the evidence for hemochromatosis screening. *Ann Intern Med* 1998; 129:971.
42. Galiatsatos P, Foulkes WD. Familial adenomatous polyposis. *Am J Gastroenterol* 2006; 101:385.
43. Breitner JC, Wyse BW, Anthony JC, et al. APOE-epsilon4 count predicts age when prevalence of AD increases, then declines: the Cache County Study. *Neurology* 1999; 53:321.
44. Codori AM, Brandt J. Psychological costs and benefits of predictive testing for Huntington's disease. *Am J Med Genet* 1994; 54:174.
45. ACCE: A CDC-sponsored project carried out by the Foundation of Blood Research. Available at: www.cdc.gov/genomics/gtesting/ACCE/acce_proj.htm (Accessed on November 16, 2009).
46. Clayton EW. Ethical, legal, and social implications of genomic medicine. *N Engl J Med* 2003; 349:562.
47. Weil J. *Psychosocial Genetic Counseling*, Oxford University Press, New York 2000.
48. Peters JA, Kenen R, Giusti R, et al. Exploratory study of the feasibility and utility of the colored eco-genetic relationship map (CEGRM) in women at high genetic risk of developing breast cancer. *Am J Med Genet A* 2004; 130A:258.
49. Marteau TM, Croyle RT. The new genetics. Psychological responses to genetic testing. *BMJ* 1998; 316:693.
50. Green RM, Thomas AM. Whose gene is it? A case discussion about familial conflict over genetic testing for breast cancer. *J Genet Couns* 1997; 6:245.
51. Rothenberg KH. Genetic information and health insurance: state legislative approaches. *J Law Med Ethics* 1995; 23:312.
52. Pokorski RJ. A test for the insurance industry. *Nature* 1998; 391:835.
53. Allain DC, Friedman S, Senter L. Consumer awareness and attitudes about insurance discrimination post enactment of the Genetic Information Nondiscrimination Act. *Fam Cancer* 2012; 11:637.
54. Genetic Information Nondiscrimination Act (GINA). Public Law 110-233, 122 Stat. 881 (final passage May 21, 2008).
55. Hudson KL, Holohan MK, Collins FS. Keeping pace with the times--the Genetic Information Nondiscrimination Act of 2008. *N Engl J Med* 2008; 358:2661.
56. Korobkin R, Rajkumar R. The Genetic Information Nondiscrimination Act--a half-step toward risk sharing. *N Engl J Med* 2008; 359:335.
57. Green RC, Lautenbach D, McGuire AL. GINA, genetic discrimination, and genomic medicine. *N Engl J Med* 2015; 372:397.
58. <http://www.genome.gov/10002077> (Accessed on October 11, 2013).
59. <https://www.disabilityrightslegalcenter.org/about/documents/GeneticsStateSurveyFinal.pdf> (Accessed on October 11, 2013).
60. Hall MA, Rich SS. Patients' fear of genetic discrimination by health insurers: the impact of legal protections. *Genet Med* 2000; 2:214.
61. Hudson KL. Genomics, health care, and society. *N Engl J Med* 2011; 365:1033.
62. Van Hoyweghen I, Horstman K. European practices of genetic information and insurance: lessons for the Genetic Information Nondiscrimination Act. *JAMA* 2008; 300:326.
63. Godard B, Raeburn S, Pembrey M, et al. Genetic information and testing in insurance and employment: technical, social and ethical issues. *Eur J Hum Genet* 2003; 11 Suppl 2:S123.
64. Lerman C, Shields AE. Genetic testing for cancer susceptibility: the promise and the pitfalls. *Nat*

Rev Cancer 2004; 4:235.

65. Morrison PJ. Insurance, unfair discrimination, and genetic testing. *Lancet* 2005; 366:877.
66. Otlowski M, Taylor S, Bombard Y. Genetic discrimination: international perspectives. *Annu Rev Genomics Hum Genet* 2012; 13:433.
67. Murashige N, Tanimoto T, Kusumi E. Fear of genetic discrimination in Japan. *Lancet* 2012; 380:730.
68. Patenaude AF, Dorval M, DiGianni LS, et al. Sharing BRCA1/2 test results with first-degree relatives: factors predicting who women tell. *J Clin Oncol* 2006; 24:700.
69. MacDonald DJ, Sarna L, van Servellen G, et al. Selection of family members for communication of cancer risk and barriers to this communication before and after genetic cancer risk assessment. *Genet Med* 2007; 9:275.
70. Godard B, Hurlimann T, Letendre M, et al. Guidelines for disclosing genetic information to family members: from development to use. *Fam Cancer* 2006; 5:103.
71. GeneTests: Medical Genetics Information Resource (database online). Available at: www.ncbi.nlm.nih.gov/sites/GeneTests/?db=GeneTests (Accessed on November 09, 2009).
72. US Department of Health, Education, and Welfare. The Institutional Guide to DHEW Policy on Protection of Human Subjects. GPO; US Department of Health, Education, and Welfare, Washington, DC 1971.
73. Yang Y, Muzny DM, Reid JG, et al. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N Engl J Med* 2013; 369:1502.
74. Cressey D. UK pushes whole genome sequencing into clinical practice. Newsblog at Nature.com, December 2012. Available at: <http://blogs.nature.com/news/2012/12/uk-pushes-whole-genome-sequencing-into-clinical-practice.html> (Accessed on November 04, 2013).
75. www.ncbi.nlm.nih.gov/projects/GeneTests/static/concepts/primer/primerordertest.shtml#choosing (Accessed on November 16, 2009).
76. Warby, SC, Graham, RK, Hayden, MR. Huntington Disease. In: GeneReviews at GeneTests: Medical Genetics Information Resource (database online). Available at: www.ncbi.nlm.nih.gov/bookshelf/br.fcgi?book=gene&part=huntington#huntington (Accessed on November 17, 2009).
77. Dietz, HC. (Updated June 30, 2009). Marfan Syndrome. In: GeneReviews at GeneTests: Available at: www.ncbi.nlm.nih.gov/bookshelf/br.fcgi?book=gene&part=marfan#marfan (Accessed on November 17, 2009).
78. Kujovic JL. Factor V Leiden Thrombophilia. In: GeneReviews at GeneTests. Available at: www.ncbi.nlm.nih.gov/bookshelf/br.fcgi?book=gene&part=factor-v-leiden#factor-v-leiden (Accessed on November 17, 2009).
79. Kefford RF, Newton Bishop JA, Bergman W, Tucker MA. Counseling and DNA testing for individuals perceived to be genetically predisposed to melanoma: A consensus statement of the Melanoma Genetics Consortium. *J Clin Oncol* 1999; 17:3245.
80. Roach JC, Glusman G, Smit AF, et al. Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science* 2010; 328:636.
81. Ku CS, Naidoo N, Pawitan Y. Revisiting Mendelian disorders through exome sequencing. *Hum Genet* 2011; 129:351.
82. Facio FM, Brooks S, Loewenstein J, et al. Motivators for participation in a whole-genome sequencing study: implications for translational genomics research. *Eur J Hum Genet* 2011; 19:1213.
83. Biesecker LG, Mullikin JC, Facio FM, et al. The ClinSeq Project: piloting large-scale genome

sequencing for research in genomic medicine. *Genome Res* 2009; 19:1665.

84. Levin B, Barthel JS, Burt RW, et al. Colorectal Cancer Screening Clinical Practice Guidelines. *J Natl Compr Canc Netw* 2006; 4:384.

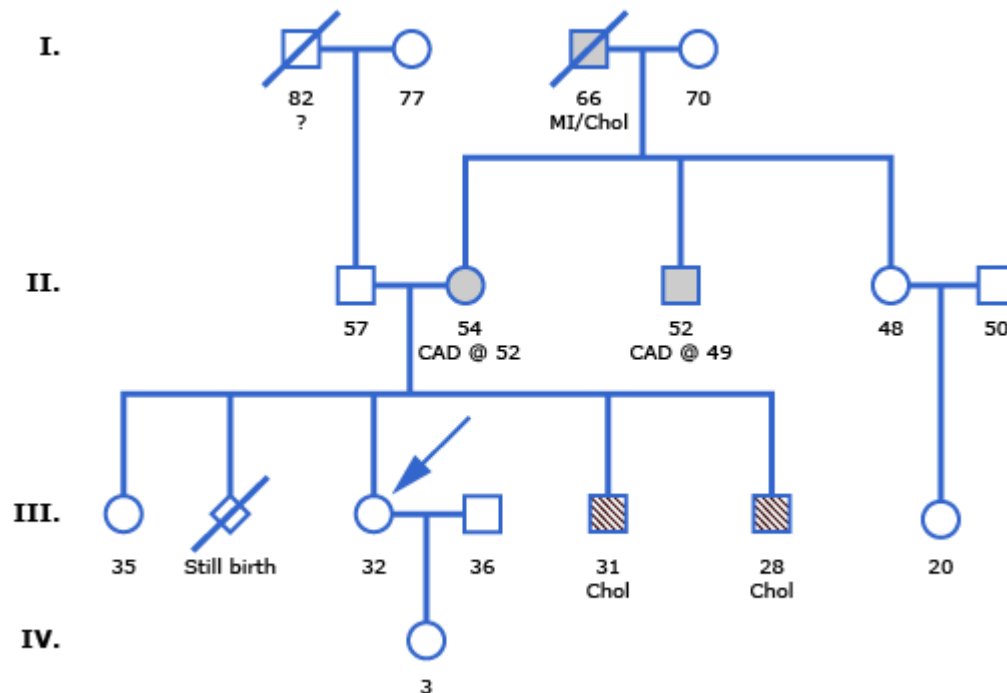
Topic 2901 Version 23.0

GRAPHICS**Resources for collecting family history information**

Resource	URL	Content
US Surgeon General: My Family Health Portrait	https://familyhistory.hhs.gov/ https://familyhistory.hhs.gov/fhh-web/home.action?request_locale=es	Tool for patients to create their own family tree (Available in English and Spanish).
National Society of Genetic Counselors: Your Family History - Your Future	http://www.nsgc.org/consumer/familytree/index.cfm	Guidelines on how to gather and record family history information.
American Medical Association: Family History Tools	http://www.ama-assn.org/go/familyhistory	Tools for collecting family history in prenatal, pediatric, and adult settings.
American Academy of Family Physicians: Family History Program	http://www.aafp.org	Guidance on efficient approaches for collecting and evaluating family history information.
Utah Department of Health	http://health.utah.gov/genomics/familyhistory/toolkit.html	Resource to help patients collect their own family histories.
American Heart Association	http://www.americanheart.org/presenter.jhtml?identifier=3044889	Resource for patients to collect family history information pertinent to heart disease and diabetes.
Ohio State University Comprehensive Cancer Center - James Cancer Hospital	http://www.jamesline.com/patientsandvisitors/prevention/	Online resource which provides cancer risk assessment based on family history.
March of Dimes	http://www.marchofdimes.com/pnhec/4439_1109.asp	Printable preconception/prenatal family and medical history questionnaire.

Graphic 62528 Version 1.0

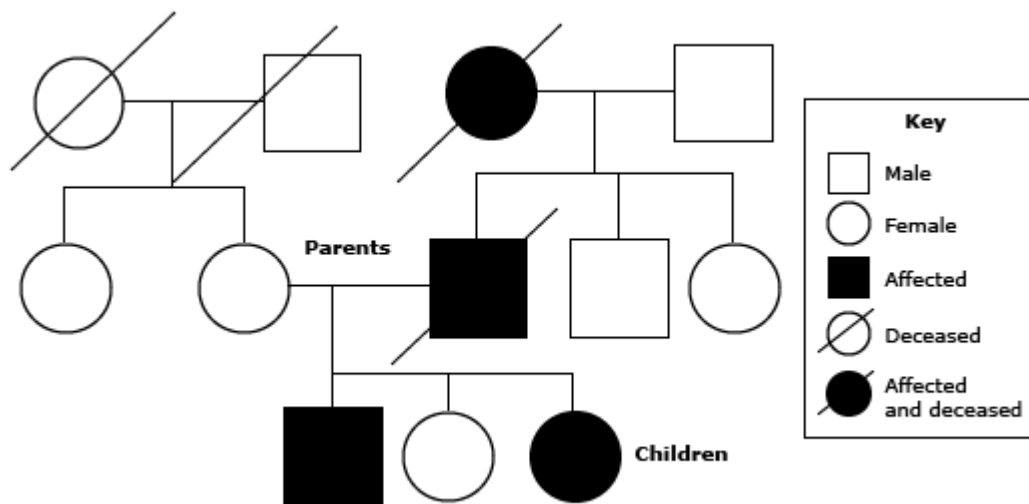
An example of a 4-generation medical pedigree



Patient denoted with arrow. From top, left to right: I. 1 = Paternal grandfather, died age 82 of unknown cause; I-2 paternal grandmother, alive age 77; I-3 maternal grandfather, died age 66, history of MI and hyperlipidemia; I-4 maternal grandmother, alive age 70. II-1 Father age 57; II-2 Mother, alive age 54 with coronary disease onset at age 52; II-3 Maternal uncle, alive age 52, with coronary disease onset at age 49; II-4 Maternal aunt, age 48; II-5 husband of II-4, age 50. III-1 sister, age 35; III-2 a sibling (gender unknown) who died at birth; III-3 patient, age 32; III-4 patient's husband, age 36; III-5 brother, age 31 with hyperlipidemia; III-6 brother, age 28 with hyperlipidemia; III-7 female cousin, age 20. IV-1 daughter, age 3.

Graphic 55908 Version 2.0

Pedigree of autosomal dominant inheritance

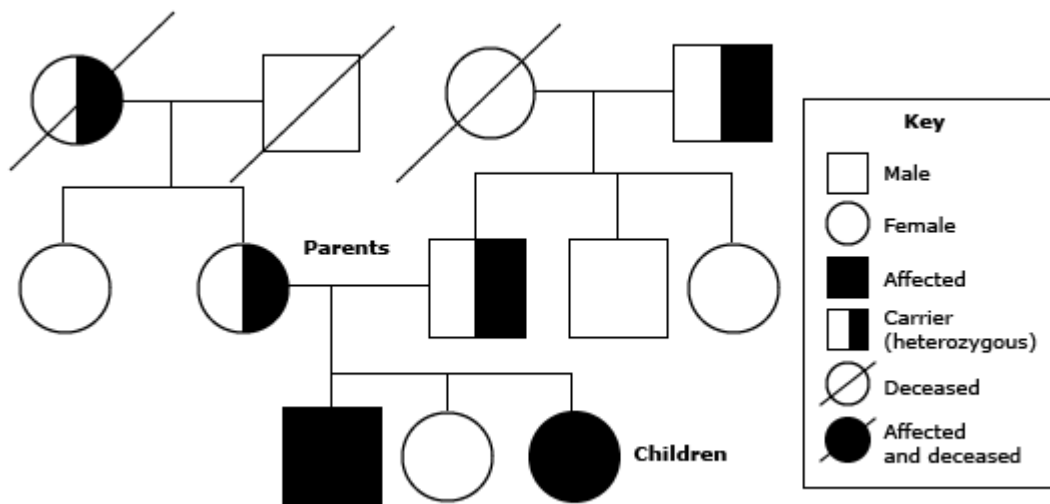


Multiple generations are affected.

Courtesy of Linda Pinsky, MD.

Graphic 75203 Version 2.0

Pedigree of autosomal recessive inheritance

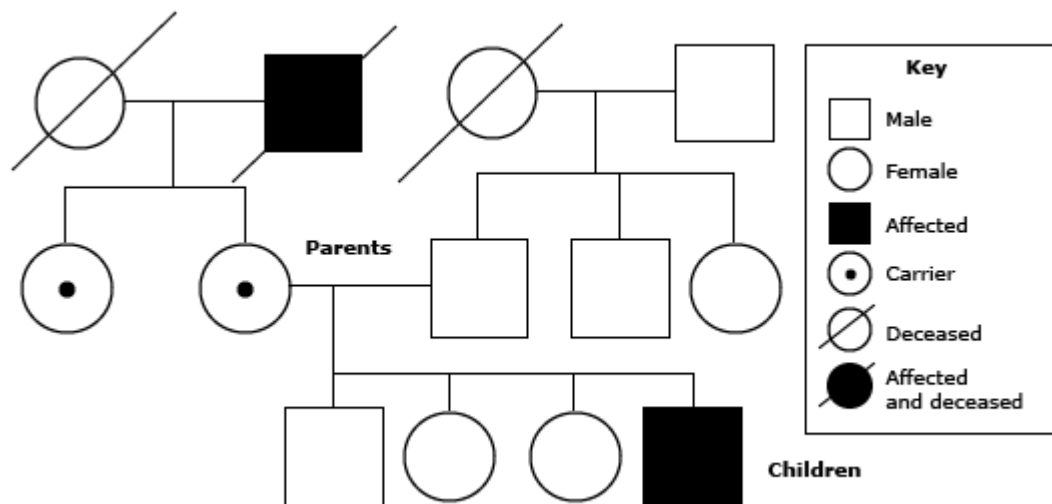


An individual's siblings are more likely to be affected than parents.

Courtesy of Linda Pinsky, MD.

Graphic 65919 Version 2.0

Pedigree of x-linked recessive inheritance



Females are carriers and only males are affected.

Courtesy of Linda Pinsky, MD.

Graphic 56888 Version 2.0

Risks for common diseases

Average risk	Moderate risk	High risk
<ul style="list-style-type: none"> • No known family history, OR • Only one second or more distantly related relative 	<ul style="list-style-type: none"> • One first degree relative (FDR) with onset of disease at an average age, OR • Two second degree relatives (SDR) on the same side of the family* 	<ul style="list-style-type: none"> • Premature disease or unusual presentation in an FDR • 2 \geq affected FDRs* • 2 \geq SDRs, with at least one having premature onset* • 3 \geq affected relatives* • Moderate risk status on both sides of the family

* Relatives must be on the same side of the family.

Graphic 71096 Version 1.0

American Academy of Family Physicians Family GENES mnemonic for identifying red flags for hereditary conditions

Features	Examples
G roup of congenital anomalies	While anatomic variation is normal, the presence of multiple variations may be indicative of a genetic syndrome.
E xtrême or exceptional presentation of common conditions	Examples of this include the onset of disease at a younger than average age, unusually severe presentation of illness, multiple primary cancers, recurrent miscarriages.
N eurodevelopmental delay or degeneration	Developmental delay or regression in children and early onset dementia in adults may be a sign of a genetic condition.
E xtrême or exceptional pathology	Some tumor pathologies may be suggestive of an inherited condition. Examples include medullary thyroid cancer, pheochromocytoma, plexiform neurofibromas, and multiple colon polyps.
S urprising laboratory values	Certain laboratory values may indicate an inherited defect in metabolic or clotting pathways. Examples include elevated fasting transferrin-iron saturation and cholesterol level >500 mg/dL.

Data from Rich, EC, Burke, W, Heaton, CJ, et al. Reconsidering the family history in primary care. *Journal of General Internal Medicine* 2004; 19:273-280 and SCREEN: Taking a Family for Familial Disease. At American Academy of Family Physicians <http://www.aafp.org>. (Accessed July 30, 2007).

Graphic 61908 Version 2.0

Genetic terms

Term	Definition
Genotype	The genetic blueprint
Phenotype	The translation of the genotype into the physical or biochemical manifestation in an individual
Penetrance	The frequency with which the genotype results in any phenotypic changes
Variable expressivity	The range of phenotypic changes (age of onset, severity of disease, different signs and symptoms, etc) that may result from aspecific genotype
Mutation	A change in the sequence of DNA that results in a phenotypic change; a mutation may be beneficial, neutral or harmful
Polymorphism	A variation in DNA sequence that occurs in greater than 1 percent of the population
Presymptomatic	Term used in predictive genetic testing to identify unaffected persons at increased risk for genetic disease
Predispositional	Term used in predictive genetic testing; refers to detecting a genotype with less than 100 percent penetrance in an asymptomatic individual - ie, the person may or may not develop the condition
Types of chromosomes	
Autosomal chromosomes	22 paired sets that are not sex chromosomes
Allele	Each of a pair of genes on homologous chromosomes
Homozygous	When both of the paired alleles have the mutation
Heterozygous	When the mutation occurs on only one of a pair of alleles
Compound heterozygote	Genotype formed by each allele having a different mutation
Sex-linked chromosomes	The X and Y chromosomes - XX in females, XY in males
Modes of inheritance	
Multifactorial	A constellation of genetic and environmental factors that interact to produce a specific condition
Mendelian inheritance	Autosomal dominant and recessive inheritance, and X-linked inheritance
Autosomal dominant inheritance	When a disorder can result if one of a specific pair of genes has a mutation. Child of an affected individual has a 50 percent chance of inheriting the gene involved.
Autosomal recessive inheritance	When a disorder can result only if both genes in a specific pair have a mutation. Child of an affected individual is an obligate carrier; parents of an affected individual have a 25 percent

	chance of having an affected child in each pregnancy.
X-linked inheritance	Mutation exists on X-chromosome: Most X-linked conditions are recessive - two copies of a specific gene on the paired X chromosome must be abnormal for a female to be affected.
Non-Mendelian inheritance	Pattern of inheritance that is not autosomal dominant, autosomal recessive or X-linked
Polygenic	When more than one gene is involved in producing a condition
Mitochondrial	Condition arising from a mutation in the DNA of the mitochondria, inherited from the mother only
Types of genetic testing	
Diagnostic genetic testing	Assessment of DNA sequence or structure for defined genetic risk factors for the purpose of supporting a specific disease diagnosis
Carrier screening	Assessment of DNA sequence or structure for specific disease-causing alleles in at-risk populations
Prenatal testing	Assessment of fetal DNA sequence or structure for specific disease-causing alleles or karyotypes
Preimplantation testing	Assessment of sequence or structure of blastocyst-derived DNA with the purpose of selecting embryos for implantation that are free of pathogenic genotypes
Newborn screening	A variety of genetic and metabolic tests performed in the newborn period for the purposes of early (ie, preclinical) identification of children with disease-causing genotypes

Graphic 72267 Version 3.0

The ACCE Model's list of targeted questions aimed at a comprehensive review of genetic testing

Element	Component	Specific question
Disorder/setting		1. What is the specific clinical disorder to be studied?
		2. What are the clinical findings defining this disorder?
		3. What is the clinical setting in which the test is to be performed?
		4. What DNA test(s) are associated with this disorder?
		5. Are preliminary screening questions employed?
		6. Is it a stand-alone test or is it one of a series of tests?
		7. If it is part of a series of screening tests, are all tests performed in all instances (parallel) or are only some tests performed on the basis of other results (series)?
Analytic validity		8. Is the test qualitative or quantitative?
	Sensitivity	9. How often is the test positive when a mutation is present?
	Specificity	10. How often is the test negative when a mutation is not present?
		11. Is an internal QC program defined and externally monitored?
		12. Have repeated measurements been made on specimens?
		13. What is the within- and between-laboratory precision?
		14. If appropriate, how is confirmatory testing performed to resolve false positive results in a timely manner?
		15. What range of patient specimens have been tested?
		16. How often does the test fail to give a useable result?
		17. How similar are results obtained in multiple laboratories using the same, or different, technology?
Clinical validity	Sensitivity	18. How often is the test positive when the disorder is present?

Clinical utility	Specificity	19. How often is the test negative when a disorder is not present?
		20. Are there methods to resolve clinical false positive results in a timely manner?
	Prevalence	21. What is the prevalence of the disorder in this setting?
		22. Has the test been adequately validated on all populations to which it may be offered?
		23. What are the positive and negative predictive values?
		24. What are the genotype/phenotype relationships?
		25. What are the genetic, environmental or other modifiers?
	Intervention	26. What is the natural history of the disorder?
	Intervention	27. What is the impact of a positive (or negative) test on patient care?
	Intervention	28. If applicable, are diagnostic tests available?
	Intervention	29. Is there an effective remedy, acceptable action, or other measurable benefit?
	Intervention	30. Is there general access to that remedy or action?
		31. Is the test being offered to a socially vulnerable population?
	Quality assurance	32. What quality assurance measures are in place?
	Pilot trials	33. What are the results of pilot trials?
	Health risks	34. What health risks can be identified for follow-up testing and/or intervention?
		35. What are the financial costs associated with testing?
	Economic	36. What are the economic benefits associated with actions resulting from testing?
	Facilities	37. What facilities/personnel are available or easily put in place?
	Education	38. What educational materials have been developed and validated and which of these are available?
		39. Are there informed consent requirements?
	Monitoring	40. What methods exist for long-term monitoring?

ELSI		41. What guidelines have been developed for evaluating program performance?
	Impediments	42. What is known about stigmatization, discrimination, privacy/confidentiality and personal/family social issues?
		43. Are there legal issues regarding consent, ownership of data and/or samples, patents, licensing, proprietary testing, obligation to disclose, or reporting requirements?
	Safeguards	44. What safeguards have been described and are these safeguards in place and effective?

Data from: Haddow JE, Palomaki GE. ACCE: A Model Process for Evaluating Data on Emerging Genetic Tests. In: Human Genome Epidemiology: A Scientific Foundation for Using Genetic Information to Improve Health and Prevent Disease. Khoury, M, Little, J, Burke, W (Eds.), Oxford University Press 2003, p. 217.

Reproduced from Centers for Disease Control. Available at http://www.cdc.gov/genomics/gtesting/ACCE/acce_proj.htm#T1. Accessed on 2/24/2012.

Graphic 68534 Version 2.0

Methods of genetic testing

DNA-based
Direct DNA Testing
Looks for known mutation that causes a condition; requires that the indicated mutation can be identified by testing.
Linkage Testing
When the conditions for direct DNA testing are not met, linkage testing can infer the presence of a mutation through the study of several family members to identify DNA markers that are co-inherited with the gene of interest.
Methylation Studies
Analyzes the attachment of methyl groups to the DNA molecule of genes in disorders in which methylation patterns influence disease.
Protein Truncation Test (PTT)
Identifies shortened protein product when the mutation that acts by this method, resulting in reduced or absent protein function.
Uniparental Disomy (UPD)
Detection of two copies of the same chromosome pair from one parent and none from the other parent that can result in genetic disease.
X Inactivation Studies
In general, one of the pair of X-chromosomes is randomly inactivated in each cell and does not reactivate.
When in certain condition, X-inactivation is non-random, it can be measured and carrier status of women with a history of these disorders can be predicted on this basis.
Cytogenetic
Fluorescence In Situ Hybridization (FISH)
Viewing of a fluorescence-labeled chromosomal probe introduced into the cell nucleus where it attaches to its match within the cell; used to test for missing or additional chromosome material.
Biochemical
Analyte
Quantitative analysis of a substance in the body when increased or decreased amounts may be indicative of a genetic disorder.
Enzyme Assay
An enzyme assay that quantifies enzyme activity in order to assess genetic disease or carrier status of a specific disease related to this reaction.
Protein Analysis
Analysis of protein structure looking for alteration that may result in a disease state.

Array Genomic Hybridization (aGH)
Quantitative analysis of labeled total genomic DNA to a microarray (DNA chip) to detect submicroscopic deletions and duplications.

Reproduced with permission from GeneTests (www.genetests.org).

<http://www.genetests.org>

Graphic 63290 Version 1.0

Disclosures

Disclosures: **Benjamin A Raby, MD, MPH** Other Financial Interest (Spouse): Parexel International [Contract research organization (Double-blind randomized clinical trials)]. **Wendy Kohlmann, MS, CGC** Consultant/Advisory Boards: Myriad [genetic testing (hereditary cancer testing)]. **Vickie Venne, MS, CGC** Nothing to disclose. **Anne Slavotinek, MBBS, PhD** Nothing to disclose. **Jennifer S Tirnauer, MD** Nothing to disclose.

Contributor disclosures are reviewed for conflicts of interest by the editorial group. When found, these are addressed by vetting through a multi-level review process, and through requirements for references to be provided to support the content. Appropriately referenced content is required of all authors and must conform to UpToDate standards of evidence.

Conflict of interest policy

Genomic disorders: An overview

Author

Carlos A Bacino, MD, FACMG

Section Editors

Helen V Firth, DM, FRCP, DCH
Benjamin A Raby, MD, MPH

Deputy Editor

Jennifer S Tirnauer, MD

All topics are updated as new evidence becomes available and our [peer review process](#) is complete.

Literature review current through: Feb 2015. | **This topic last updated:** Mar 13, 2015.

INTRODUCTION — Genomic disorders are diseases that result from the loss or gain of chromosomal/DNA material. The most common and better-delineated genomic disorders are divided in two main categories: those resulting from copy number losses (deletion syndromes) and copy number gains (duplication syndromes).

An overview of genomic disorders is presented here. Specific syndromic disorders are reviewed separately.

COPY NUMBER VARIATIONS — Structural genetic variation refers to a class of sequence alterations spanning more than 1000 bases (one kilobase or kb) [1]. This class includes quantitative variations such as copy number variations (CNVs), sequence rearrangements (such as those observed among immunoglobulins), and other less common variations, including chromosomal rearrangements that may or may not alter the genome contents and in some cases result in disease.

CNVs, the most prevalent type of structural variation, are DNA segments spanning thousands to millions of bases whose copy number varies between different individuals [2,3]. These submicroscopic genomic differences in the number of copies of one or more sections of DNA result in DNA gains or losses. Copy number gains can be the result of duplications, triplications, or even multiple copy number gains. Most deletions are one copy loss (heterozygous), but in some instances the loss can affect both copies (homozygous).

CNVs, which are most commonly inherited but can occur *de novo*, were initially thought to be rare events resulting from sporadic mutation and correlated with specific Mendelian diseases [4,5]. These misperceptions about their rarity and absolute disease linkage were primarily due to technical limitations precluding genome-wide assessments in large cohorts. Advances in technology have shown that deviation from the diploid state is widespread and contributes substantially to genetic diversity. Some studies have suggested that CNV differences in the human genome are as extensive as 20 percent, although this may be an overestimation [6,7]. It is estimated conservatively that most individuals carry an average of three large-scale CNVs [3]. The number of known CNVs that contribute to disease pathogenesis continues to increase.

The physical distribution of CNVs appears to be nonrandom, with both CNV hot and cold spots reported [6,8]. CNV frequency is greatest in regions of segmental duplication (a 4- to 10-fold enrichment for CNVs), consistent with nonallelic homologous recombination as a primary mechanism for CNV mutation [2,9-12]. (See "[Genetic and environmental causes of birth defects](#)", section on 'Nonallelic homologous recombination'.)

CNVs are more commonly observed in gene-rich regions. CNVs appear to be enriched in specific gene families, including immune and inflammatory response genes, cell signaling and cell adhesion molecules, structural proteins, and olfactory receptors [6]. Most of these differences probably represent benign CNVs that reflect normal variation with no apparent clinical consequence [13].

CNVs can be pathogenic if they involve a dose-sensitive gene(s) or if they influence genomic regions through regulatory elements [14,15]. Some pathogenic CNVs cause syndromic disorders with consistent phenotypic features (eg, deletions of elastin in Williams syndrome, duplications of PMP22 in Charcot-Marie-Tooth disease type 1A [CMT1A]), while others are associated with disease susceptibility or resistance (eg, cancer, human immunodeficiency virus [HIV] infection, autoimmune disorders, autism).

CNVs can be responsible for Mendelian diseases associated with large gains and losses of genetic material or even small losses or gains at the exonic level, as well as syndromes with more complex combinations of genetic and environmental factors. Examples include:

- Contiguous gene deletions/duplications as seen for example in Williams-Beuren syndrome, 22q11 deletion-syndrome, Smith-Magenis syndrome, and Potocki-Lupski syndrome. These recurrent CNVs are mediated by nonallelic homologous recombination (NAHR) at sites of low-copy repeats. (See ["Williams-Beuren syndrome"](#) and ["Microdeletion syndromes \(chromosomes 12 to 22\)"](#), section on '17p11.2 deletion syndrome (Smith-Magenis syndrome)' and ["Microduplication syndromes"](#), section on '17p11.2 duplication syndrome (Potocki-Lupski syndrome)').
- Deletions of genes or portions of genes (exons) leading to many Mendelian-inherited genetic disorders, including disorders that are autosomal dominant (eg, Rubinstein-Taybi syndrome) and X-linked recessive (eg, Duchenne muscular dystrophy) [16]. (See ["Microdeletion syndromes \(chromosomes 12 to 22\)"](#), section on '16p13.3 deletion syndrome (Rubinstein-Taybi syndrome)' and ["Clinical features and diagnosis of Duchenne and Becker muscular dystrophy"](#), section on 'Genetics and pathogenesis'.)
- A deletion at the complement factor 4 (C4) locus that confers a 1.6- to 5.3-fold risk for systemic lupus erythematosus [17]. (See ["Epidemiology and pathogenesis of systemic lupus erythematosus"](#).)
- A deletion in FCGR3B associated with granulomatosis with polyangiitis (Wegener's) [18]. (See ["Pathogenesis of granulomatosis with polyangiitis and related vasculitides"](#).)
- A deletion of defensin-beta 4 (DEFB4) associated with increased risk of colonic Crohn's disease [19]. (See ["Genetic factors in inflammatory bowel disease"](#).)
- Increased frequency of *de novo* germ-line CNVs in patients with autism spectrum disorder (ASD) and schizophrenia [20,21]. CNVs in multiple areas of the human genome potentially involved in autism pathogenesis have been described. (See ["Autism spectrum disorder: Terminology, epidemiology, and pathogenesis"](#).)

This line of research has also identified specific genes and pathways involved in ASD and related syndromes (eg, duplications of the 15q11-q13 region, deletions and duplications of 16p11.2, 1q21 duplications) [22-24]. (See ["Microdeletion syndromes \(chromosomes 1 to 11\)"](#) and ["Microdeletion syndromes \(chromosomes 12 to 22\)"](#).)

In addition, some conditions are associated with multiple CNVs, which may explain their variable phenotypes [25]. As an example, in a retrospective study, array comparative genomic hybridization (CGH) was used to evaluate CNVs in 2312 children with developmental disabilities who already had one predefined CNV and in 8329 children without developmental disabilities [26]. This study found that, in comparison to controls without developmental disabilities, individuals with developmental disabilities had an increased number of second site CNVs. This increase in CNVs may have played a causative role in the disabilities (eg, by causing disruption of a new gene or altering gene dosage), or may be an indirect marker of susceptibility to genomic damage. (See ["Array comparative genomic hybridization"](#) below.)

Causes of CNVs — Low copy repeats are stretches of repetitive DNA sequences (segmental

duplications) approximately 10 to 300 kilobases in size that share ≥ 95 percent homology. Erroneous pairing of these highly homologous regions can cause misalignment and unequal recombination during meiosis. This can lead to duplication and deletion of chromosomal material resulting in CNVs. This process is known as nonallelic homologous recombination (NAHR) ([figure 1](#)), the most common mechanism for the formation of genomic rearrangements [[27,28](#)].

NAHR can result in either deletions or duplication via the same mechanisms and due to low copy repeats mediated nonallelic homologous recombination. A classic example is the case of Charcot-Marie-Tooth Type I, a peripheral neuropathy caused by duplications of the PMP22 gene on chromosome 17p11.2. The same region when deleted leads to a different neuropathy known as Tomaculous neuropathy also known as hereditary neuropathy with liability to pressure palsies (HNPP). (See "[Microduplication syndromes](#)".)

Other mechanisms include nonhomologous end joining and microhomology-mediated break-induced replication, although discussion of these is beyond the scope of this chapter [[29](#)].

Despite a great deal of knowledge about the structural details of how CNVs occur, we do not know what predisposes certain individuals to develop these changes more than other individuals.

Of interest, in a large study of patients with developmental disabilities, parental data provided information about whether CNVs were inherited or arose *de novo* [[26](#)]. This study suggested that CNVs were more likely to arise *de novo* in the syndromic disorders (eg, Williams-Beuren syndrome), whereas CNVs were more likely to be inherited in the disorders with variable phenotype (eg, intellectual disability). A potential explanation may be that reproductive fitness is reduced in individuals with the more severe syndromic disorders. (See "[Microdeletion syndromes \(chromosomes 1 to 11\)](#)", [section on '7q11.23 deletion syndrome \(Williams syndrome\)'](#).)

Interpreting CNVs — The interpretation of CNVs has steadily improved due to use of large control databases that allow a direct comparison with apparently normal controls. Examples of these databases include the Database of Genomic Variants (DGV) or large sequencing projects such as the 1000 Genomes Project [[30,31](#)]. Other databases that include phenotypic information, as in the case of DECIPHER (**D**atabas**E** of **C**hromosomal **I**mbalance and **P**henotype in **H**umans using **E**nsembl **R**esources), are becoming valuable resources for findings detected by array-comparative genomic hybridization [[32](#)]. [DECIPHER](#) is a web-based resource that contains array and clinical data deposited by over 200 member centers around the world.

Interpreting the pathogenicity of CNVs can be quite challenging in the presence of multiple CNVs in one individual. Some studies have shown that multiple rare CNVs, either inherited or *de novo*, can compound the clinical severity [[26,33](#)]. It is also important to highlight the presence of common CNVs in the general population. It is unclear whether many of these common polymorphisms may play a role in common disorders [[34](#)].

DISEASE MECHANISMS — There are different potential mechanisms that can lead to disease in genomic disorders secondary to deletions and duplications. The main mechanism is related to changes in dosage sensitive genes. "Haploinsufficiency" ("haplo" = half) defines the concept where loss or gain of one allele of a gene leads to abnormal protein production or function, thereby causing disease [[35](#)].

Deletions can interfere with the required gene product dose, resulting in disease. One example is Williams-Beuren Syndrome, which is caused by a microdeletion in chromosome 7q11.23 involving multiple genes, including the gene for elastin [[36,37](#)]. Having only half of the normal amount (dose) of elastin is enough to cause disruption of the arteries leading to aortic narrowing and multiple other arterial abnormalities. The deletions in Williams-Beuren syndrome, similar to what happens in many other genomic disorders, can be of different sizes. Typically the loss encompasses 1.55 Mb, but in some cases the deletion can be more extensive or even smaller. The size difference is due to the different LCRs

around the critical region that can be involved in mediating the rearrangements. (See ["Chromosomal translocations, deletions, and inversions"](#), section on 'Deletions' and ["Microdeletion syndromes \(chromosomes 1 to 11\)"](#), section on '7q11.23 deletion syndrome (Williams syndrome)' and ["Supravalvular aortic stenosis"](#).)

The importance of gene dosage in determining the effect of CNVs was illustrated in the sex differences from a large study of CNVs in patients with developmental disabilities [26]. When compared to females, males had more of the "variable phenotype genomic disorders" (eg, intellectual disability) but not syndromic disorders (eg, autism spectrum disorders). Females may be protected from these more genetically multifactorial disorders due to sex chromosome bias (ie, the protection of females from weakly deleterious mutations on one X chromosome by the normal corresponding genes on the other X chromosome).

Duplications may also disrupt a gene and alter the amount of protein produced by interfering with protein synthesis or assembly. Another mechanism includes unmasking recessive alleles [38-40]. When a recessive allele is deleted, it could potentially uncover a pathogenic mutation in the remaining copy. This would lead to disease, because there is no working copy for the affected gene.

Other disease mechanisms include interference with (i) imprinted genes like in the case of paternal duplications of 11p15 that lead to Beckwith-Wiedemann syndrome or (ii) with regulatory elements outside genes like in brachydactyly type A2 and duplications outside the BMP2 gene [41,42]. (See ["Beckwith-Wiedemann syndrome"](#), section on 'Genetics and pathogenesis'.)

Arrays that contain single nucleotide polymorphisms (SNPs) can further aid in the identification of imprinting disorders caused by uniparental disomy. One such example is Angelman syndrome (AS) and uniparental disomy (UPD) caused by isodisomy. Isodisomies result from either nondisjunction in meiosis II or postzygotic duplication (monosomy rescue). A small number of AS cases are the result of UPD. In those cases there is absence of the maternal contribution for a region of chromosome 15 (15q11-q13). These cases are typically associated with monosomy rescue (duplication of a chromosome from a monosomic zygote), where there are two identical copies of the paternal chromosome 15 and no maternal contribution. Other examples of UPD can be seen in patients with Prader-Willi syndrome. (See ["Epidemiology and genetics of Prader-Willi syndrome"](#).)

Contiguous gene syndromes — Contiguous gene syndromes can occur when large CNVs affect several contiguous genes [43,44]. For example, Williams-Beuren Syndrome is caused by a 1.5-1.8 Mb deletion on chromosome 7q11 that typically encompasses nine genes. Occasionally, molecular-phenotypic correlation is possible. For example, in WAGR syndrome, a genomic disorder consisting of **W**ilms tumor, **A**niridia, **G**enitourinary anomalies, and mental **R**etardation, clinical features are attributable to the loss of individual genes by a large deletion: deletions of *WT1* are responsible for Wilms tumor, while *PAX6* deletions are responsible for the aniridia findings. Both genes are contiguously located within the short arm of chromosome 11. (See ["Microdeletion syndromes \(chromosomes 1 to 11\)"](#), section on '11p13 deletion syndrome (WAGR syndrome)'.)

DETECTION OF GENOMIC DISORDERS — Genomic disorders are typically detected by array comparative genomic hybridization (array CGH). Most laboratories confirm gains or losses detected on an array with an independent method such as fluorescence *in situ* hybridization (FISH), multiple ligation dependent probe amplification (MLPA), or quantitative PCR (Q-PCR). Parental FISH testing, and in selected cases array CGH, is warranted in all cases of detected genomic abnormalities as many are inherited and relevant for future pregnancies [45]. Whereas the genomic position of a loss is clear from an array, gains may be either tandem duplications or insertions. If the latter arise as a consequence of a parental insertional translocation (IT), this may have important implications for future pregnancies [46].

Array comparative genomic hybridization — Array comparative genomic hybridization (array CGH),

also known as chromosome microarray or microarray-based comparative genomic hybridization, is the gold standard laboratory test for the detection of CNVs that cause genomic disorders ([figure 2](#)). Array CGH allows detection of small losses or gains of genomic material down to several kilobases (kb) and even the exon level. It is widely used in the evaluation of patients with intellectual disabilities and/or congenital malformations [47-52]. (See "[Tools for genetics and genomics: Cytogenetics and molecular genetics](#)", section on 'Array comparative genomic hybridization'.)

The two main platforms currently used for CNV detection are oligonucleotide arrays (oligonucleotides are stretches of DNA ranging from 25 to 60 base pairs) ([figure 2](#)), and single nucleotide polymorphism arrays (SNP arrays) [53]. There are approximately 10 million polymorphic SNPs throughout the human genome. Both SNPs and oligonucleotide arrays can detect copy number variations, but SNP arrays can be used in addition to determine absence of homozygosity (AOH) as seen in cases of consanguinity [54] and in cases of uniparental disomy [55] when there is inheritance of regions or entire chromosomes from one single parent instead of the normal biparental contribution. SNPs can also detect loss of heterozygosity (LOH) typically seen in somatic cancer cell changes. SNPs can also be very useful for the detection of somatic mosaicism, a situation where two or more cell lines can be present in a single individual, and triploidies, a rare instance when there could be a total of 69 chromosomes (3 haploid sets) that can be detected in the prenatal setting. Some current platforms are combining the use of array CGH and SNPs integrated in one single platform.

Other molecular diagnostic techniques — Other molecular techniques used for the detection of genomic disorders include fluorescence in situ hybridization (FISH), PCR-based studies like Q-PCR (quantitative PCR), and MLPA (Multiple Ligation Dependent Probe Amplification). Protocols using sequencing of the whole genomes or exome (all regions of the genome that encode proteins) will likely become viable options for detecting CNV variations [56,57]. (See "[Tools for genetics and genomics: Cytogenetics and molecular genetics](#)".)

FISH uses larger stretches of DNA (approximately 50 to 100 kilobase probes) labeled with fluorescence reagents to target specific genome regions. The use of FISH, however, requires knowledge of what specific area is being targeted and is dependent upon a clinical diagnosis [58,59]. MLPA uses a cocktail of multiple probes available in kits and targets specific chromosomes or disease regions. A single reaction allows simultaneous hybridization of multiple probes to multiple regions or even multiple exons within a gene.

The advantage of MLPA over FISH is the lower cost and more comprehensive coverage for copy number gains/losses detection due to the use of multiple probes. However, MLPA does not provide location. FISH studies, on the other hand, can determine location if visualized in metaphase chromosome spreads. A copy number gain can be the result of chromosome duplication immediately adjacent to the area of interest, part of a marker chromosome (a structurally abnormal chromosome that makes up a partial trisomy), or the result of an insertion or translocation. Duplications that are very close to each other may be challenging to detect by metaphase FISH and are best detected by interphase FISH.

Sequencing of the whole genome or exome can also be used to detect genomic disorders. This was demonstrated in a study that used sequencing from 849 individuals to identify areas of CNVs and their role in gene dosage [56]. Genome sequencing has also been used to map breakpoints in the genome responsible for duplications of chromosomal regions [60]. Sequencing has the potential to provide improved resolution, but the sensitivity and specificity of sequencing for detecting CNVs is variable. Information about sequencing methods, collectively referred to as next-generation sequencing (NGS), is presented separately. (See "[Principles and clinical applications of next-generation DNA sequencing](#)".)

SUMMARY

- Genomic disorders are diseases that result from the loss or gain of chromosomal/DNA material. The most common and better-delineated genomic disorders are divided in two main categories: those resulting from copy number losses (deletion syndromes) and copy number gains (duplication syndromes). (See ['Introduction'](#) above and ["Congenital cytogenetic abnormalities"](#).)
- Copy number variations (CNVs) are submicroscopic genomic differences in the number of copies of one or more sections of DNA that result in DNA gains or losses ([figure 1](#)). Some pathogenic CNVs cause syndromic disorders with consistent phenotypic features. Other CNVs are associated with disease susceptibility or resistance, and the same CNV can be associated with several diverse disorders. (See ['Copy number variations'](#) above.)
- The main mechanism that leads to disease in genomic disorders secondary to deletions and duplications is changes in dosage-sensitive genes. Other disease mechanisms include interference with imprinted genes and with regulatory elements outside genes. (See ['Disease mechanisms'](#) above.)
- Genomic disorders are typically detected by array comparative genomic hybridization (array CGH). Most laboratories confirm gains or losses detected on an array with an independent method such as fluorescence *in situ* hybridization (FISH), multiple ligation dependent probe amplification (MLPA), or quantitative PCR (Q-PCR). (See ['Detection of genomic disorders'](#) above.)

Use of UpToDate is subject to the [Subscription and License Agreement](#).

REFERENCES

1. Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. *Nat Rev Genet* 2006; 7:85.
2. Redon R, Ishikawa S, Fitch KR, et al. Global variation in copy number in the human genome. *Nature* 2006; 444:444.
3. Tuzun E, Sharp AJ, Bailey JA, et al. Fine-scale structural variation of the human genome. *Nat Genet* 2005; 37:727.
4. Ewart AK, Morris CA, Atkinson D, et al. Hemizygosity at the elastin locus in a developmental disorder, Williams syndrome. *Nat Genet* 1993; 5:11.
5. Magenis RE, Toth-Fejel S, Allen LJ, et al. Comparison of the 15q deletions in Prader-Willi and Angelman syndromes: specific regions, extent of deletions, parental origin, and clinical consequences. *Am J Med Genet* 1990; 35:333.
6. Cooper GM, Nickerson DA, Eichler EE. Mutational and selective effects on copy-number variants in the human genome. *Nat Genet* 2007; 39:S22.
7. Sharp AJ. Emerging themes and new challenges in defining the role of structural variation in human disease. *Hum Mutat* 2009; 30:135.
8. Perry GH, Tchinda J, McGrath SD, et al. Hotspots for copy number variation in chimpanzees and humans. *Proc Natl Acad Sci U S A* 2006; 103:8006.
9. Sharp AJ, Cheng Z, Eichler EE. Structural variation of the human genome. *Annu Rev Genomics Hum Genet* 2006; 7:407.
10. Sharp AJ, Locke DP, McGrath SD, et al. Segmental duplications and copy-number variation in the human genome. *Am J Hum Genet* 2005; 77:78.
11. Sebat J, Lakshmi B, Troge J, et al. Large-scale copy number polymorphism in the human genome. *Science* 2004; 305:525.
12. Wong KK, deLeeuw RJ, Dosanjh NS, et al. A comprehensive analysis of common copy-number

variations in the human genome. *Am J Hum Genet* 2007; 80:91.

13. Wellcome Trust Case Control Consortium, Craddock N, Hurles ME, et al. Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature* 2010; 464:713.
14. Girirajan S, Eichler EE. Phenotypic variability and genetic susceptibility to genomic disorders. *Hum Mol Genet* 2010; 19:R176.
15. Fanciulli M, Petretto E, Aitman TJ. Gene copy number variation and common human disease. *Clin Genet* 2010; 77:201.
16. Boone PM, Bacino CA, Shaw CA, et al. Detection of clinically relevant exonic copy-number changes by array CGH. *Hum Mutat* 2010; 31:1326.
17. Yang Y, Chung EK, Wu YL, et al. Gene copy-number variation and associated polymorphisms of complement component C4 in human systemic lupus erythematosus (SLE): low copy number is a risk factor for and high copy number is a protective factor against SLE susceptibility in European Americans. *Am J Hum Genet* 2007; 80:1037.
18. Fanciulli M, Norsworthy PJ, Petretto E, et al. FCGR3B copy number variation is associated with susceptibility to systemic, but not organ-specific, autoimmunity. *Nat Genet* 2007; 39:721.
19. Fellermann K, Stange DE, Schaeffeler E, et al. A chromosome 8 gene-cluster polymorphism with low human beta-defensin 2 gene copy number predisposes to Crohn disease of the colon. *Am J Hum Genet* 2006; 79:439.
20. Sebat J, Lakshmi B, Malhotra D, et al. Strong association of de novo copy number mutations with autism. *Science* 2007; 316:445.
21. Stefansson H, Rujescu D, Cichon S, et al. Large recurrent microdeletions associated with schizophrenia. *Nature* 2008; 455:232.
22. Girirajan S, Dennis MY, Baker C, et al. Refinement and discovery of new hotspots of copy-number variation associated with autism spectrum disorder. *Am J Hum Genet* 2013; 92:221.
23. Barber JC, Hall V, Maloney VK, et al. 16p11.2-p12.2 duplication syndrome; a genomic condition differentiated from euchromatic variation of 16p11.2. *Eur J Hum Genet* 2013; 21:182.
24. Hogart A, Wu D, LaSalle JM, Schanen NC. The comorbidity of autism with the genomic disorders of chromosome 15q11.2-q13. *Neurobiol Dis* 2010; 38:181.
25. Sanders SJ, Ercan-Sencicek AG, Hus V, et al. Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron* 2011; 70:863.
26. Girirajan S, Rosenfeld JA, Coe BP, et al. Phenotypic heterogeneity of genomic disorders and rare copy-number variants. *N Engl J Med* 2012; 367:1321.
27. Stankiewicz P, Inoue K, Bi W, et al. Genomic disorders: genome architecture results in susceptibility to DNA rearrangements causing common human traits. *Cold Spring Harb Symp Quant Biol* 2003; 68:445.
28. Stankiewicz P, Shaw CJ, Dapper JD, et al. Genome architecture catalyzes nonrecurrent chromosomal rearrangements. *Am J Hum Genet* 2003; 72:1101.
29. Hastings PJ, Lupski JR, Rosenberg SM, Ira G. Mechanisms of change in gene copy number. *Nat Rev Genet* 2009; 10:551.
30. 1000 Genomes Project Consortium, Abecasis GR, Altshuler D, et al. A map of human genome variation from population-scale sequencing. *Nature* 2010; 467:1061.
31. Zhang J, Feuk L, Duggan GE, et al. Development of bioinformatics resources for display and analysis of copy number and other structural variants in the human genome. *Cytogenet Genome Res* 2006; 115:205.

32. Firth HV, Richards SM, Bevan AP, et al. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *Am J Hum Genet* 2009; 84:524.
33. Girirajan S, Brkanac Z, Coe BP, et al. Relative burden of large CNVs on a range of neurodevelopmental phenotypes. *PLoS Genet* 2011; 7:e1002334.
34. Gijsbers AC, Schoumans J, Ruivenkamp CA. Interpretation of array comparative genome hybridization data: a major challenge. *Cytogenet Genome Res* 2011; 135:222.
35. Seidman JG, Seidman C. Transcription factor haploinsufficiency: when half a loaf is not enough. *J Clin Invest* 2002; 109:451.
36. Wu YQ, Sutton VR, Nickerson E, et al. Delineation of the common critical region in Williams syndrome and clinical correlation of growth, heart defects, ethnicity, and parental origin. *Am J Med Genet* 1998; 78:82.
37. Schubert C. The genomic basis of the Williams-Beuren syndrome. *Cell Mol Life Sci* 2009; 66:1178.
38. Ropers HH. Genetics of intellectual disability. *Curr Opin Genet Dev* 2008; 18:241.
39. Wong LJ, Dimmock D, Geraghty MT, et al. Utility of oligonucleotide array-based comparative genomic hybridization for detection of target gene deletions. *Clin Chem* 2008; 54:1141.
40. Kurth I, Klopocki E, Stricker S, et al. Duplications of noncoding elements 5' of SOX9 are associated with brachydactyly-anonychia. *Nat Genet* 2009; 41:862.
41. Russo S, Finelli P, Recalcati MP, et al. Molecular and genomic characterisation of cryptic chromosomal alterations leading to paternal duplication of the 11p15.5 Beckwith-Wiedemann region. *J Med Genet* 2006; 43:e39.
42. Dathe K, Kjaer KW, Brehm A, et al. Duplications involving a conserved regulatory element downstream of BMP2 are associated with brachydactyly type A2. *Am J Hum Genet* 2009; 84:483.
43. Ballabio A. Contiguous deletion syndromes. *Curr Opin Genet Dev* 1991; 1:25.
44. Punnett HH, Zakai EH. Old syndromes and new cytogenetics. *Dev Med Child Neurol* 1990; 32:824.
45. Kang SH, Shaw C, Ou Z, et al. Insertional translocation detected using FISH confirmation of array-comparative genomic hybridization (aCGH) results. *Am J Med Genet A* 2010; 152A:1111.
46. Nowakowska BA, de Leeuw N, Ruivenkamp CA, et al. Parental insertional balanced translocations are an important cause of apparently de novo CNVs in patients with developmental anomalies. *Eur J Hum Genet* 2012; 20:166.
47. Bejjani BA, Theisen AP, Ballif BC, Shaffer LG. Array-based comparative genomic hybridization in clinical diagnosis. *Expert Rev Mol Diagn* 2005; 5:421.
48. Cheung SW, Shaw CA, Yu W, et al. Development and validation of a CGH microarray for clinical cytogenetic diagnosis. *Genet Med* 2005; 7:422.
49. Lu X, Shaw CA, Patel A, et al. Clinical implementation of chromosomal microarray analysis: summary of 2513 postnatal cases. *PLoS One* 2007; 2:e327.
50. Lu XY, Phung MT, Shaw CA, et al. Genomic imbalances in neonates with birth defects: high detection rates by using chromosomal microarray analysis. *Pediatrics* 2008; 122:1310.
51. Miller DT, Adam MP, Aradhya S, et al. Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *Am J Hum Genet* 2010; 86:749.
52. Stankiewicz P, Beaudet AL. Use of array CGH in the evaluation of dysmorphology, malformations, developmental delay, and idiopathic mental retardation. *Curr Opin Genet Dev* 2007; 17:182.
53. Cooper GM, Mefford HC. Detection of copy number variation using SNP genotyping. *Methods Mol Biol* 2011; 767:243.
54. Schaaf CP, Scott DA, Wyszniowska J, Beaudet AL. Identification of incestuous parental

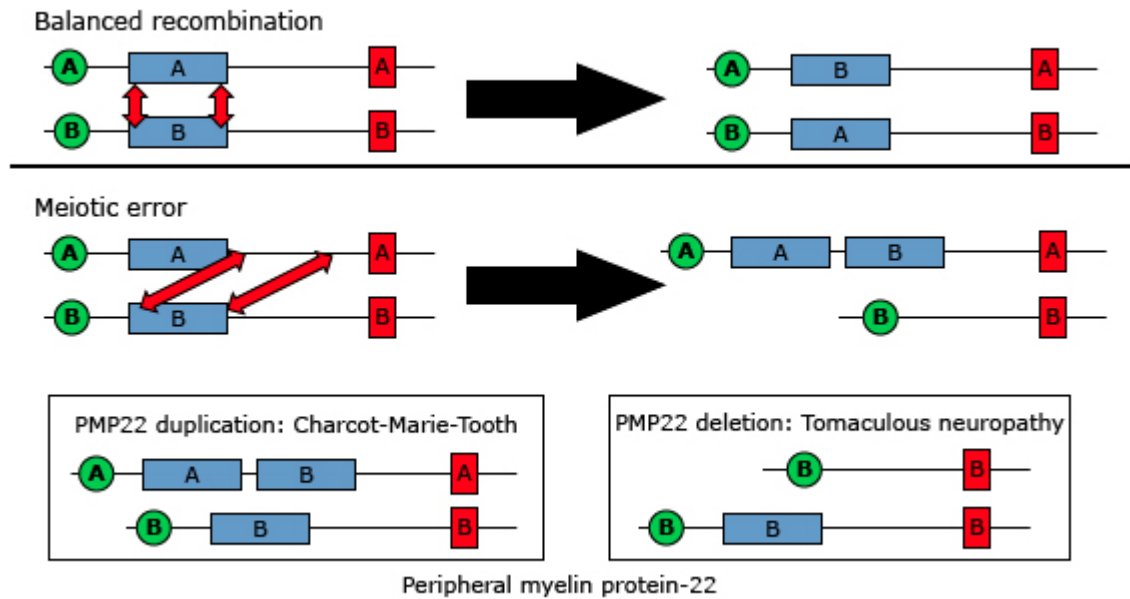
relationships by SNP-based DNA microarrays. *Lancet* 2011; 377:555.

55. Papenhausen P, Schwartz S, Risheg H, et al. UPD detection using homozygosity profiling with a SNP genotyping microarray. *Am J Med Genet A* 2011; 155A:757.
56. Handsaker RE, Van Doren V, Berman JR, et al. Large multiallelic copy number variations in humans. *Nat Genet* 2015; 47:296.
57. de Ligt J, Boone PM, Pfundt R, et al. Detection of clinically relevant copy number variants with whole-exome sequencing. *Hum Mutat* 2013; 34:1439.
58. Shaffer LG, Kennedy GM, Spikes AS, Lupski JR. Diagnosis of CMT1A duplications and HNPP deletions by interphase FISH: implications for testing in the cytogenetics laboratory. *Am J Med Genet* 1997; 69:325.
59. Eijk-Van Os PG, Schouten JP. Multiplex Ligation-dependent Probe Amplification (MLPA®) for the detection of copy number variation in genomic sequences. *Methods Mol Biol* 2011; 688:97.
60. Newman S, Hermetz KE, Weckselblatt B, Rudd MK. Next-Generation Sequencing of Duplication CNVs Reveals that Most Are Tandem and Some Create Fusion Genes at Breakpoints. *Am J Hum Genet* 2015; 96:208.

Topic 16647 Version 8.0

GRAPHICS

Nonhomologous recombination resulting in copy number variation



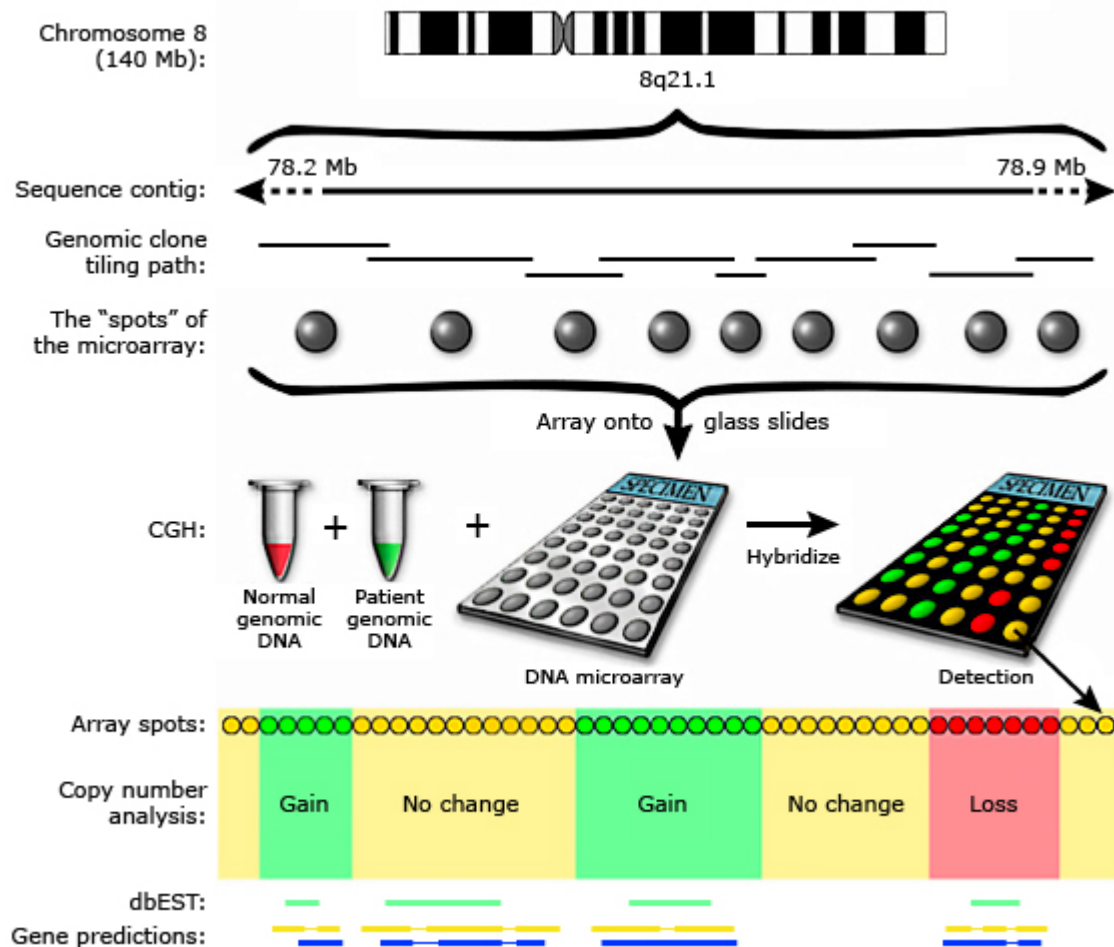
(Upper panel) Normal recombination of homologous region.

(Lower panel) Aberrant recombination, resulting in imbalanced swap of DNA between chromosomes.

PMP22: peripheral myelin protein 22.

Graphic 67644 Version 3.0

Schematic CGH for microarray setup



Schematic representation of the array CGH technique for a focused analysis of copy number imbalances along a region of interest (eg, 8q21.1). A tiling path of genomic clones (eg, BACs, PACs, PIs, cosmids) is generated to cover the region. After extraction and purification, these genomic DNA targets are arrayed onto glass slides.

Array CGH is performed by hybridizing labeled normal (Cy3) and tumor (Cy5) genomic DNA into the microarray and detected using a microarray scanner.

Each array spot, realigned *in silico* as a single contiguous map to correspond with the tiling path, can be analyzed by fluorescence ratio to identify the regions of copy number changes. These results may be correlated with *in silico* techniques to identify candidate genes of interest.

CGH: comparative genomic hybridization; BAC: bacterial artificial chromosome; PAC: P1 bacteriophage artificial chromosome; P1: P1 bacteriophage; Cy3: cyanine dye with green fluorescence; Cy5: cyanine dye with red fluorescence.

Reproduced from: Beheshti B, Park PC, Braude I, Squire JA. Micorarray CGH. In: Molecular Cytogenetics: Protocols and Applications. Humana Press, 2002, with kind permission from Springer Science + Business Media B.V. Copyright © 2002.

Graphic 58888 Version 10.0

Disclosures

Disclosures: **Carlos A Bacino, MD, FACMG** Nothing to disclose. **Helen V Firth, DM, FRCP, DCH** Nothing to disclose. **Benjamin A Raby, MD, MPH** Other Financial Interest (Spouse): Parexel International [Contract research organization (Double-blind randomized clinical trials)]. **Jennifer S Tirnauer, MD** Nothing to disclose.

Contributor disclosures are reviewed for conflicts of interest by the editorial group. When found, these are addressed by vetting through a multi-level review process, and through requirements for references to be provided to support the content. Appropriately referenced content is required of all authors and must conform to UpToDate standards of evidence.

Conflict of interest policy

Tools for genetics and genomics: Cytogenetics and molecular genetics**Authors**

Iris Schrijver, MD
James L Zehnder, MD
Athena M Cherry, PhD

Section Editor

Benjamin A Raby, MD, MPH

Deputy Editor

Jennifer S Tirnauer, MD

All topics are updated as new evidence becomes available and our [peer review process](#) is complete.

Literature review current through: Feb 2015. | **This topic last updated:** Jun 27, 2014.

INTRODUCTION — Intense genetic research has exponentially increased our knowledge of the genetic code of humans and other organisms, leading to the development of numerous methods that facilitate our understanding of normal and abnormal genetic processes. Many of these methods and related techniques are now routinely used in the molecular diagnosis of both inherited disorders and diseases that result from somatic mutations, such as the hematologic malignancies. Molecular genetic and cytogenetic diagnostics are invaluable additions to laboratory testing and clinical evaluation, providing diagnostic, therapeutic, and prognostic information. (See "[Genetic abnormalities in hematologic and lymphoid malignancies](#)".)

Although the advent of improved and faster molecular methods has transformed the traditional diagnostic process, keeping current with the most recent advances is daunting [1]. A conceptual approach to a selection of the most common standard and novel diagnostic tools will be applied to this review. An outline of the advantages and limitations of each of the techniques is included, as well as some examples of their applications. A brief introduction to the terms required to properly understand these applications is provided separately. (See "[Principles of molecular genetics](#)".)

Three general categories of testing can be distinguished.

- **Mutation detection** of known sequence changes can be performed. This type of testing is targeted and typically limited to a predefined number of sequence changes, selected in advance. Selection is generally based on association with clinical phenotypes. Sequence changes may be located within a single gene or across multiple genes. Depending on the testing method used, the number of included sequence changes can range from a single mutation to thousands of mutations.
- **Cytogenetic studies** of large structural variants are typically performed when the phenotype does not seem limited to point mutations and relatively small deletions and duplications. Such studies are helpful in syndromic phenotypes and for constellations of symptoms typically associated with abnormalities on the scale of chromosomes rather than single exons or genes.
- **Genotyping methods** can be applied to identify mutations not selected in advance. These methods aim to discover mutations, and can target a gene with a known heterogeneous distribution of mutations, or can target larger segments of the genome to identify known or novel variations.

DETECTING KNOWN MUTATIONS — There are many different approaches for the identification of selected, known mutations. Typically, these start with the polymerase chain reaction, after which additional assay steps are performed. The following section lists examples of some of the frequently used techniques, together with their advantages and disadvantages. General mutation detection methods such as DNA sequencing can also be applied to the identification of known mutations. Although the entire sequence in an amplified fragment would be read, mutations within that segment would be identified readily by this more comprehensive method.

Polymerase chain reaction — The automated polymerase chain reaction (PCR) is commonly the first

step in the vast majority of DNA analyses because it increases the amount of DNA available for analysis. This important diagnostic tool is discussed in detail separately. (See ["Tools for genetics and genomics: Polymerase chain reaction"](#).)

Restriction enzyme digestion — Restriction enzyme digestion can be used to detect mutations that create or destroy a restriction enzyme site. Specific restriction enzymes, usually isolated from bacteria, recognize unique short sequences within a DNA fragment. These enzymes can cleave the DNA strands at that exact site. If a mutation either changes the DNA code to a sequence that creates a new restriction enzyme site or obliterates an existing restriction enzyme site, the mutation can be detected based upon the presence or absence of the site ([figure 1](#)).

One example of such a test is a restriction digestion assay for Muenke syndrome, a type of craniosynostosis syndrome defined by a single mutation (c.749C>G, p.Pro250Arg) in the FGFR3 gene that creates a new restriction site. In this assay, exon 7 is amplified by PCR, and the sample is subjected to a restriction digest with Ban I, the enzyme that recognizes the new site. DNA from patients who carry the mutation is cleaved by the enzyme, while control DNA lacking the mutation remains uncleaved.

Advantages of this technique include the following:

- It is technically easy and can be performed within one day.
- Restriction enzyme analysis detects specific mutations and can be applied to many samples concurrently. These samples are then run side-by-side on a single gel.

Disadvantages are the following:

- This method is impractical for disorders caused by a large number of different mutations and for the detection of mutations associated with nucleotide sequences that require the use of expensive restriction enzymes.
- Only a small fraction of existing point mutations actually create or remove a restriction site. In some instances, this problem can be circumvented by the introduction of an artificial restriction site during PCR amplification [\[2\]](#).
- Incomplete digestion can produce erroneous results. This problem can be overcome by using adequate controls.

Amplification refractory mutation system — The amplification refractory mutation system (ARMS) can be used to detect known point mutations. This technique requires a multiplex PCR reaction [\[3\]](#). In this reaction, two primer pairs are added to a single PCR tube and two separate sequences from one piece of DNA are amplified in the same reaction. One reaction (using the control primer pair) is an internal control to demonstrate that the PCR reaction itself has worked. The other reaction (using a primer pair specific for the mutation under study) will amplify the target sequence depending upon the presence or absence of a specific point mutation. A second tube contains DNA from the same patient, and includes the control primer pair and a primer pair that amplifies only the normal sequence. The only difference between the primers for the normal and mutant sequence is complementarity of one primer at the 3' end, where one is identical to the normal and one to the variant sequence. The other primer is the same for both reactions, and the product size will be the same. When run side by side on a gel, these samples will exhibit homozygosity for the normal sequence, homozygosity for the mutation, or heterozygosity ([figure 2](#)).

This assay is an actual modification of the PCR itself, rather than an additional step after the original amplification. ARMS can be performed for just one, or a small number of simultaneously tested mutations. An example is an ARMS test for BRAF mutation c.1799T>A (p.V600E) which has been identified in colorectal carcinoma, melanoma, and papillary thyroid carcinoma. Because the mutation is a

point mutation, changing one nucleotide to another, primers can be designed such that either the wild-type or the mutant sequence perfectly matches the very 3 prime end of the primer. In an optimized assay, a PCR product would then only be expected with the primer pair that entirely matches.

Advantages of this technique include the following:

- It is easy to perform and complete within one day.
- It can detect specific point mutations and assess many samples concurrently. Modification of the method allows analysis of several mutations in one test tube.

Disadvantages are the following:

- It is impractical for disorders caused by a large number of mutations.
- Primer pairs must be designed for all reactions. Amplification of these reactions under the same conditions is a prerequisite. If the conditions are suboptimal for one of the primer pairs, weak amplification or nonspecific amplification may result in ambiguous results.
- Every patient sample requires multiple PCR tubes.

Allele specific oligonucleotide hybridization — Allele-specific oligonucleotide (ASO) hybridization involves the placing ("spotting") of denatured PCR-amplified DNA onto a membrane and subsequent hybridization with short allele-specific, labeled probes. Under optimal hybridization and washing conditions, hybridization will only occur if the probe sequence is perfectly complementary to the single-stranded sample DNA.

Typically, PCR products from one patient sample are fixed onto two identical membranes (a "dot-blot"), one of which is hybridized with a probe that contains the normal sequence, while the other is hybridized with a probe for the mutant sequence. The two probes should differ by just one nucleotide, corresponding to the point mutation under investigation. After exposure to an autoradiographic film in the case of radioactive probe labeling, or after chemical treatment in the case of biotinylated oligomers, positive signals are scored and heterozygosity or homozygosity for the normal or mutant sequence can be determined ([figure 3](#)).

ASO hybridization can be modified to analyze a panel of mutations for a single patient. In "reverse" allele-specific hybridization, for example, sequence specific probes are spotted onto the membrane and only one membrane is used per patient. Reverse ASO is less cost-effective than regular ASO, but can decrease the turnaround time per sample. This method is frequently applied in molecular testing for cystic fibrosis. (See ["Cystic fibrosis: Clinical manifestations and diagnosis"](#), section on 'Molecular diagnosis'.)

Advantages of ASO hybridization include the following:

- It is suitable for analysis of specific mutations or polymorphisms in numerous samples [\[4\]](#).
- It is highly sensitive and specific if properly optimized.
- Adaptations for multiplex PCR analysis or automated microarray (DNA chip) analysis are possible. (See ["Amplification refractory mutation system"](#) above and ["Genotyping microarrays"](#) below.)

Disadvantages include:

- Each ASO probe can only detect one specific sequence.
- ASO hybridization is amenable to small DNA mutations only.
- There is potential non-specificity if the hybridization and/or washing conditions are not fully

optimized.

Genotyping microarrays — Genotyping microarrays are available on a variety of different molecular platforms. They can all interrogate a flexible number of mutations at the same time, which makes them attractive for high throughput analyses in one or multiple genes, for one or many different patients at the same time. These assays are typically automated, and in a high-throughput setting (ie, automated analysis of multiple samples simultaneously), do not require hands-on work after the initiation of the assay run.

Advantages of this method include the following:

- It is suitable for high-throughput analysis of specific mutations or polymorphisms in numerous samples.
- Relatively less hands-on work is required per sample.
- Interpretation of the data can also be highly automated.

Disadvantages include:

- The microarray analysis equipment as well as the individual arrays can be expensive.
- Often this method is not suitable for low-volume testing, in particular when the microarrays can be used only one time, regardless of whether only one or many samples are tested.

DETECTING CYTOGENETIC ABNORMALITIES — Cytogenetic abnormalities are genetic defects that involve large regions of chromosomes rather than small pieces of DNA (ie translocations, large deletions, or aneuploidies). These defects can be detected by at least three methods – chromosomal (karyotypic) analysis, fluorescence in situ hybridization (FISH) using specific DNA probes on either metaphase chromosomes or interphase nuclei, and array comparative genomic hybridization (aCGH). The latter two methods are considered “molecular cytogenetics,” because they can detect anomalies which are below the resolution of chromosomal analysis.

Chromosomal analysis — Chromosomal analysis (also called chromosome banding) is used to detect changes in large regions of chromosomes (translocations, large deletions, or aneuploidies).

To perform chromosome analysis, lymphocytes, usually obtained from the peripheral blood, are cultured in vitro and stimulated to divide under the influence of mitogens. Other cell types, such as amniocytes, bone marrow cells, fibroblasts, and tumor cells can also be analyzed, often without a mitogenic stimulus. Once the cells divide readily, a chemical is added to arrest mitotic division in metaphase. In this phase of the cell cycle, the chromosomes are maximally contracted and hence their banding patterns are easier to recognize ([figure 4](#)).

Chromosomal banding is then used to identify each individual chromosome, for assessment of whether the correct number of each chromosome is present (two of each autosome plus the sex chromosomes) and whether there are structural abnormalities. This technique can be performed using various enzymes and dyes. The most frequently used banding technique is GTG (G-banding with trypsin and Giemsa-banding) ([figure 5](#)). An identical banding pattern is seen in Q-banding, in which the chromosomes are stained with a fluorescent dye and viewed under ultraviolet illumination.

The typical resolution of chromosome banding is about 400 bands in a haploid set of 23 chromosomes. A single chromosome band may contain 6 megabases (Mb) of DNA and approximately 150 genes. If higher banding resolution is desired to study relatively small chromosomal rearrangements, the cell cycles can be synchronized and cells arrested in prometaphase (550 bands) or even prophase (approximately 800 bands). The latter technique is called high resolution banding.

Examples of the use of this method include evaluation of patients with acute myeloid leukemia (AML). (See ["Cytogenetics in acute myeloid leukemia"](#).)

In many cases, chromosome banding has been replaced with alternative methods such as FISH and microarray analysis. (See ["Fluorescence in situ hybridization"](#) below and ["Array comparative genomic hybridization"](#) below.)

Advantages of chromosomal banding include:

- In contrast to molecular genetic studies, chromosome banding techniques show the entire genome at one time.
- This method is suitable in diagnostic situations where a specific anomaly is suspected (eg, the Philadelphia chromosome in chronic myeloid leukemia, CML). It may also be useful to monitor disease; for example, to detect additional chromosomal abnormalities commonly seen in disease progression of CML. (See ["Cellular and molecular biology of chronic myeloid leukemia", section on 'Progression to acute phase CML'](#).)
- In disorders having deletions of varying size within a specific chromosome, such as multiple myeloma [5], karyotypes from many patients can be compared, in order to find the critical disease-associated region ([figure 6](#)).

Disadvantages are:

- Most chromosome banding techniques can only detect major structural abnormalities and will not detect smaller regions of DNA gain or loss.
- Interpretation is labor intensive and highly dependent upon operator experience and skill.

Fluorescence in situ hybridization — Fluorescence in situ hybridization (FISH) is a technique that allows counting of the number and location of large pieces of chromosomes. This technique has greatly increased the sensitivity, specificity, and resolution of chromosome analyses [6]. FISH can be performed on metaphase chromosomes or interphase nuclei; interphase FISH can be done on paraffin embedded tissue.

Metaphase FISH allows identification of large chromosomal abnormalities, including deletions, duplications, and translocations, as well as smaller chromosomal microdeletions and duplications. For metaphase FISH, cells are arrested in mitosis as for chromosomal banding. They are then fixed using a mixture of [acetic acid](#) and methanol and then "dropped" on a glass microscope slide where they are affixed. DNA probes of a few hundred kilobases (kb) in length are used that match regions the chromosomes containing the DNA sequence in question. These probes are directly hybridized with the chromosomes on the slide (hence, the term "in situ" hybridization); immediate detection of the fluorescence signal is possible via fluorescence microscopy. Less commonly used isotopic and non-isotopic chemical labeling methods are also available ([image 1](#)).

FISH probes produce a fluorescent dot on the chromosome to which they hybridize. Thus, every pair of chromosomes (or chromosome regions) produces two dots. These double dots sometimes fuse to form one signal. Cells that are monosomic for the chromosomal region in question would show only a single dot per nucleus, while trisomic cells would show three dots.

FISH can also be modified for analysis of interphase nuclei. (See ["Interphase FISH"](#) below.)

Advantages of this technique are:

- The resolution of FISH is much better than traditional chromosome banding (FISH can resolve 2 megabases (Mb) in length, compared to 6 Mb for chromosomal banding).

- FISH can be applied to both dividing (metaphase) and non-dividing (interphase) cells.
- The protocol is technically fairly straightforward.
- Hybridization with multiple probes enables detection of translocation products. An example is the BCR-ABL1 fusion of t(9;22) in chronic myeloid leukemia ([figure 7](#)).
- FISH can identify a range of structural abnormalities including deletions, duplications, aneuploidy and the presence of derivative (structurally rearranged) chromosomes.
- FISH may be used to monitor recurrent or residual disease in bone marrow transplant patients. (See "[General aspects of cytogenetic analysis in hematologic malignancies](#)", [section on 'FISH'](#).)

Disadvantages of the technique are:

- Small mutations, including small deletions and insertions as well as point mutations, cannot be identified.
- Uniparental disomy (inheritance of both copies of a chromosome from the same parent) will be missed because the probe merely detects the presence or absence of a locus or specific portion of a chromosome and not its source ([figure 8](#)).
- Chromosomal inversions will be missed since a probe can only detect the presence of a specific sequence but not its precise location within the chromosome.
- Probes are not yet commercially available for all chromosomal regions.
- The clinician has to choose the correct FISH probe in order to make an accurate diagnosis.

Interphase FISH — Interphase FISH is used when dividing cells are not available (eg, in fully differentiated cells or in tissues that have been fixed and embedded in paraffin). It also improves the resolution of FISH probes. For interphase FISH, cells are tested with FISH probes without cell cycle synchronization. When using intact nuclei (eg, not from paraffin embedded tissue), the cells are harvested using a hypotonic solution, fixed, and again “dropped” on the slides. For FISH on paraffin embedded tissue, thin slices of pathological specimens are cut and affixed to the slides.

Because chromosomes are only minimally condensed in interphase, this modification of FISH analysis provides the opportunity to hybridize probes at a high resolution (well under 1 Mb, compared to 2 Mb for metaphase FISH). In the interphase nucleus, chromosomal structures cannot be discerned and only the hybridized probe will light up ([figure 9](#)). Hybridization with two different color probes that cross the breakpoint regions of genes involved in a translocation results in the expected signals from the normal chromosome as well as a fusion signal from the derived chromosome(s) on which the genes are juxtaposed by translocation [7].

Advantages of interphase FISH include the higher resolution than metaphase FISH; the ability to perform the test immediately without culturing the cells, which makes it faster; and the applicability to paraffin embedded sections. Interphase FISH is of special value in prenatal diagnosis of various aneuploidy states, such as trisomy 18 or trisomy 21, in which the ability to obtain results rapidly aids in decision making. (See "[Diagnostic amniocentesis](#)", [section on 'Fluorescence in situ hybridization'](#).)

The major disadvantage of interphase compared to metaphase FISH is that in interphase FISH the chromosomes themselves cannot be visualized. Thus, information cannot be provided regarding overall chromosome number and composition.

Spectral karyotyping — Spectral karyotyping (SKY, also called multicolor FISH) is a FISH technique that accurately identifies the chromosomal origin of all elements in a karyogram (complete chromosome set) using multiple wavelengths of light to generate signals of many colors. A combination of five

fluorochromes can be used as probes to "paint" all 22 autosomes, as well as the X and the Y chromosomes, in different colors ([figure 10](#)). The karyogram is analyzed by a computerized spectral imager and the chromosomes are classified based on their particular emission spectra [[8,9](#)].

Advantages of this technique are:

- This method allows complete karyotyping using automated analysis.
- The origin of marker chromosomes, small insertions, and complex rearrangements can be inferred through the presence of color-coded chromosomal segments [[9,10](#)].

Disadvantages include:

- The equipment required may be prohibitively expensive for small diagnostic laboratories.
- Even though part of the analysis is computerized, the overall technique is labor intensive.
- Structural rearrangements within a single chromosome will not be detected.
- The resolution of SKY is relatively low (up to 15 Mb).
- SKY may be used when a specific abnormality is suspected, but is not applicable as a screening method.

Array comparative genomic hybridization — Comparative genomic hybridization (CGH) allows detection of amplifications and deletions of smaller regions of DNA along the lengths of all of the chromosomes. The technique works by comparing the genomic content (or DNA) of a patient (or target) with a normal control individual (or individuals) ([figure 11](#)). The resolution of "classical" or metaphase CGH is relatively low (approximately 15 Mb of DNA) [[11](#)].

Array CGH is a modification of CGH in which the comparator DNA, RNA, or tissue is arrayed on a glass slide or glass beads [[12](#)]. There are three basic types of array: bacterial artificial chromosomes (BAC) arrays, oligonucleotide arrays (typically 60 base pairs in length), or single nucleotide polymorphism (SNP) arrays (typically a few nucleotides). Most SNP-based arrays now also include single locus probes as well as SNPs.

There are at least two different types of arrays: "targeted" arrays and whole genomic arrays.

- Targeted arrays "target" known microdeletion/microduplication syndromes, as well as other known loci of inherited Mendelian disorders (eg, tuberous sclerosis). The first targeted arrays were arrays of around 500 to 600 BACs.
- Whole genomic arrays cover the entire genome at varying levels of resolution. The first whole genomic arrays were BAC arrays with around 2600 BACs spaced at about 1 Mb throughout the genome (ie, with about 1 Mb resolution). Oligonucleotide and SNP arrays have supplanted the use of BACs. Most laboratories use either oligonucleotide or SNP arrays with an average resolution of about 35 kb throughout the genome.

Both SNP and oligonucleotide arrays can detect copy number variations, but only SNP arrays can be used to determine whether there is absence or loss of heterozygosity (AOH or LOH) for different regions, or even entire chromosomes in the presence of normal copy number [[13](#)]. AOH refers to the inheritance of either paternal or maternal alleles alone ([figure 8](#)). The absence of biparental inheritance can be seen in uniparental disomy (UPD) or in some cancer tissue samples. UPD can be the result of heterodisomy, in which two different homologous chromosomes from the same parent (either maternal or paternal) are present, instead of the normal biparental contribution (one chromosome from each parent). Alternatively, UPD can occur when there are two identical copies of a single parental chromosome (isodisomy). SNP arrays can only detect UPDs secondary to isodisomies [[13](#)].

Interpretation of array CGH can also be complicated by the presence of copy number variants (CNVs) that can be benign, pathogenic or unknown. (See ["Genomic disorders: An overview", section on 'Copy number variations'](#).)

Advantages of array CGH are:

- Typically, dividing cells and tissue culture are not necessary; the technique requires only good, high quality DNA.
- Resolution of the array is dependent upon the type of array used and the average spacing of the "probes" on the array.

Disadvantages of this technique are:

- Balanced structural rearrangements (ie balanced translocations, inversions, insertions) will not be detected, because there is no change in copy number.
- Levels of mosaicism (ie copy number changes in some but not all cells) of 20 percent or less will not be detected.
- Interpretation of copy number changes which have not been previously reported can be challenging, particularly if a phenotypically normal parent carries the same change.
- The technology is still relatively expensive, even though costs have declined.

GENOTYPING NEW MUTATIONS — In diseases with allelic heterogeneity, finding the disease-causing mutation is not straightforward. In these disorders, genes need to be scanned for mutations before a specific disease-causing mutation may be identified. After a new sequence variant has been detected and characterized, its pathogenicity must be established. This is commonly accomplished by screening a large number of normal controls, who are expected not to carry the same allelic variant. In addition, the DNA sequence of affected and unaffected individuals within the same family should be compared.

Many methods of mutation screening are currently available, but the techniques described below are most widely applied. However, even the most common screening methods have limited application in clinical diagnostic laboratories, since the search for individual disease-causing mutations is a laborious and expensive process.

Heteroduplex analysis — Heteroduplex analysis is used to detect point mutations on one strand of a DNA helix. The technique uses denaturation and reannealing of the double stranded target DNA [14]. If complementary single strands re-anneal, they form a perfectly aligned homoduplex. On the other hand, single strands that are not completely complementary because of the presence of a point-mutation in one strand form a heteroduplex. The failure to anneal at all base positions results in the formation of a "bubble" in the newly formed double strand. As compared to normal DNA, DNA with a bubble migrates more slowly during electrophoresis. ([figure 12](#)).

This method is relatively easy to perform, and requires little optimization. Some of the available gel matrices are less toxic than polyacrylamide gels. A disadvantage is that this scanning method does not identify the location of the mutation within the analyzed fragment.

Single strand conformation analysis — Single strand conformation analysis (SSCA) is based on the observation that single-stranded DNA fragments assume unique conformations that depend on their sequence composition when run in a non-denaturing polyacrylamide gel. Migration through the gel thus depends on chain length as well as strand conformation [15]. Before interpretations can be made, the specific pattern of migration in wild-type control samples must be known. This can be achieved by the addition of normal controls in the run.

Once optimized, this method is amenable to screening a fairly large number of samples at one time. However, this scanning method does not identify the location of the mutation within the analyzed fragment. SSCA also requires optimization, and reproducibility of the conditions in all analyses of the same DNA sequence is crucial.

Automated sequencing — The most direct approach to mutation detection is automated sequencing. A DNA sequence of interest is amplified in the presence of a primer and dideoxynucleotide chain terminators. Either the primer or the chain terminators are fluorescently labeled. As the sequence undergoes electrophoresis in an automated sequencer, a laser beam excites the fluorescent nucleotides, producing a signal that can be compiled into a DNA sequence ([figure 13](#)).

Advantages of this method include the precise identification of all DNA variations in the target sequence. Thus, the method can be applied to the detection of known mutations as well as to the identification of unknown mutations. For a large number of inherited conditions, new mutations continue to be discovered and add to the understanding of mutation spectra and genotype-phenotype correlations. Clinical testing is enhanced with the more comprehensive, inclusive testing approach.

Disadvantages of the method include the financial cost, time, and the potential for poor quality data due to poor quality starting material. It should be noted that the finding of a previously uncharacterized point mutation does not mean the mutation has clinical significance. Clinical correlation of so-called “variants of unknown significance” (VUSs) can be challenging.

Whole genome sequencing — Whole genome sequencing allows identification of mutations in the entire genome, without the investigator having to choose a gene or chromosome region of interest. Improvements in the technology for high throughput sequencing continue to reduce the cost of this method. This method is discussed in more detail elsewhere. (See ["Personalized medicine", section on 'Exome and genome sequencing'](#) and ["Principles and clinical applications of next-generation DNA sequencing"](#).)

As genome sequencing becomes available, more and more variants of unknown significance (VUSs) are expected to be identified. Clinical classification of these variants is likely to be a major challenge. Patients must be informed prior to testing that such VUSs will be identified, and though some may be clinically important (ie, confer disease risk), the current limits of medical knowledge preclude making definitive statements regarding their relevance to the patient's health. (See ["Incidental findings from genetic testing"](#).)

Southern and Northern blotting — Southern blotting can be used to detect small mutations as well as large deletions, duplications and gene rearrangements that alter restriction enzyme cleavage sites or the sizes of the resulting pieces of DNA [16]. In this technique, genomic DNA is digested with one or more restriction endonucleases and transferred to a membrane. This membrane is hybridized with a single-stranded radioactively labeled probe, under conditions that facilitate double-strand formation between the probe and those fragments on the gel containing the complementary sequence. When an autoradiographic film is exposed to the membrane and developed, the hybridized sequences become visible as bands. The location of each of these bands corresponds to the size of the fragment to which the probe is bound ([figure 14](#)).

A similar technique can be applied when RNA is the primary material. With this method, dubbed "Northern blotting", gene transcript size, abundance and expression patterns in different tissues can be analyzed.

Advantages of these methods are the potential to detect a wide range of mutations and large structural rearrangements. Southern blotting can also be used to detect changes in gene methylation status (which affects sensitivity to restriction nucleotide digestion).

Disadvantages include the requirement for larger amounts of DNA than the other methods mentioned herein, and the labor and time requirements. Generally, this assay can take a week to perform. The use of radioactive materials is both expensive and hazardous. However, nonradioactive methods, such as chemiluminescence, can be substituted.

SUMMARY

- Cytogenetic and molecular diagnostic tools are applied for three major purposes in clinical genetics: detecting specific mutations, studying large chromosomal structural variants, and genotyping to find mutations that have not been previously identified. (See ['Introduction'](#) above.)
- Most tools to identify selected mutations involve polymerase chain reaction (PCR) technology. Techniques involving specific restriction enzymes can be used to find mutations that might affect the enzyme's target cleavage site, although only a small fraction of point mutations are amenable to such techniques. The amplification refractory mutation system (ARMS) involves a multiplex PCR reaction and can detect specific point mutations. Other techniques involve oligonucleotide hybridization and genotyping microarrays. Microarrays can allow for high throughput. (See ["Tools for genetics and genomics: Polymerase chain reaction"](#) and ["Detecting known mutations"](#) above.)
- Cytogenetic analysis to identify large structural variation may involve chromosomal (karyotype) analysis by chromosomal banding; fluorescence in situ hybridization (FISH) on metaphase or interphase nuclei; or array comparative genomic hybridization (CGH). The resolution increases, meaning smaller and smaller genetic defects can be detected, from karyotyping to interphase FISH to array CGH. (See ["Detecting cytogenetic abnormalities"](#) above.)
- For diseases with allelic heterogeneity, new sequence variants continue to be discovered. Methods that allow analysis of entire genes or genomes improve this discovery. Most methods for mutation screening have limited applicability in clinical diagnostic laboratories. The most direct approach to mutation detection is automated sequencing of the target DNA. Whole genome sequencing will likely become available in the near future at reasonable cost. (See ["Genotyping new mutations"](#) above.)

Use of UpToDate is subject to the [Subscription and License Agreement](#).

REFERENCES

1. Zamicnikova A. Genetic testing methods in myelodysplastic syndromes and leukemias: A review. *Clin Leukemia* 2007; 1:331.
2. Haliassos A, Chomel JC, Grandjouan S, et al. Detection of minority point mutations by modified PCR technique: a new approach for a sensitive diagnosis of tumor-progression markers. *Nucleic Acids Res* 1989; 17:8093.
3. Newton CR, Graham A, Heptinstall LE, et al. Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Res* 1989; 17:2503.
4. Schrijver I, Liu W, Francke U. The pathogenicity of the Pro1148Ala substitution in the FBN1 gene: causing or predisposing to Marfan syndrome and aortic aneurysm, or clinically innocent? *Hum Genet* 1997; 99:607.
5. Shaughnessy J, Tian E, Sawyer J, et al. High incidence of chromosome 13 deletion in multiple myeloma detected by multiprobe interphase FISH. *Blood* 2000; 96:1505.
6. van Ommen GJ, Breuning MH, Raap AK. FISH in genome research and molecular diagnostics. *Curr Opin Genet Dev* 1995; 5:304.
7. Haralambieva E, Banham AH, Bastard C, et al. Detection by the fluorescence in situ hybridization

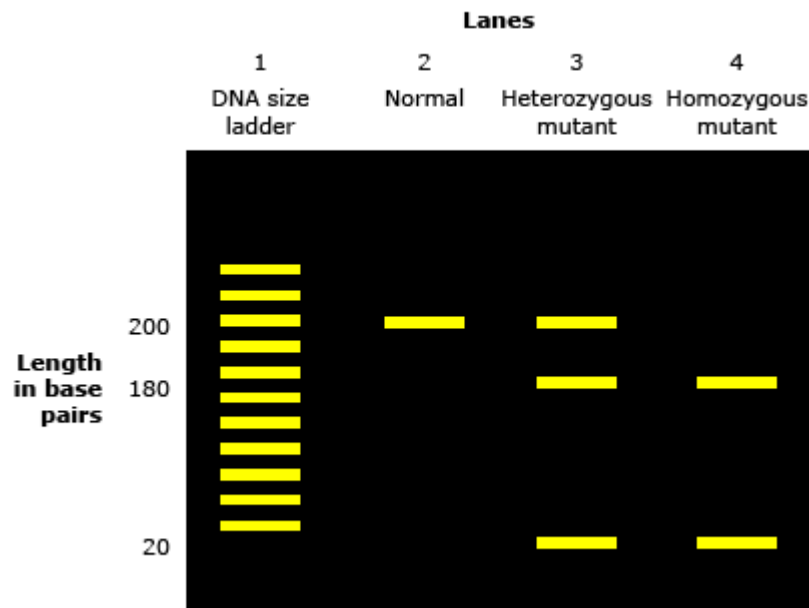
technique of MYC translocations in paraffin-embedded lymphoma biopsy samples. *Br J Haematol* 2003; 121:49.

8. Schröck E, Veldman T, Padilla-Nash H, et al. Spectral karyotyping refines cytogenetic diagnostics of constitutional chromosomal abnormalities. *Hum Genet* 1997; 101:255.
9. Jalal SM, Law ME, Stamberg J, et al. Detection of diagnostically critical, often hidden, anomalies in complex karyotypes of haematological disorders using multicolour fluorescence in situ hybridization. *Br J Haematol* 2001; 112:975.
10. Hilgenfeld E, Padilla-Nash H, McNeil N, et al. Spectral karyotyping and fluorescence in situ hybridization detect novel chromosomal aberrations, a recurring involvement of chromosome 21 and amplification of the MYC oncogene in acute myeloid leukaemia M2. *Br J Haematol* 2001; 113:305.
11. Kallioniemi A, Kallioniemi OP, Sudar D, et al. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science* 1992; 258:818.
12. Aradhya S, Cherry AM. Array-based comparative genomic hybridization: clinical contexts for targeted and whole-genome designs. *Genet Med* 2007; 9:553.
13. Papenhausen P, Schwartz S, Risheg H, et al. UPD detection using homozygosity profiling with a SNP genotyping microarray. *Am J Med Genet A* 2011; 155A:757.
14. Nagamine CM, Chan K, Lau YF. A PCR artifact: generation of heteroduplexes. *Am J Hum Genet* 1989; 45:337.
15. Hayashi K. PCR-SSCP: a method for detection of mutations. *Genet Anal Tech Appl* 1992; 9:73.
16. Southern EM. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* 1975; 98:503.

Topic 2893 Version 17.0

GRAPHICS

Restriction enzyme digestion products



This representation of an agarose gel shows how restriction enzyme digestion can be used to distinguish between an individual who is homozygous for a normal version of a gene versus individuals homozygous or heterozygous for a mutation in that gene. In this case the mutation creates a new restriction enzyme site. The normal piece of DNA is 200 base pairs long and lacks the restriction enzyme site, so it is not cut by the enzyme. The DNA with the mutation has the restriction enzyme site, so it is cut into two pieces of 180 and 20 base pairs.

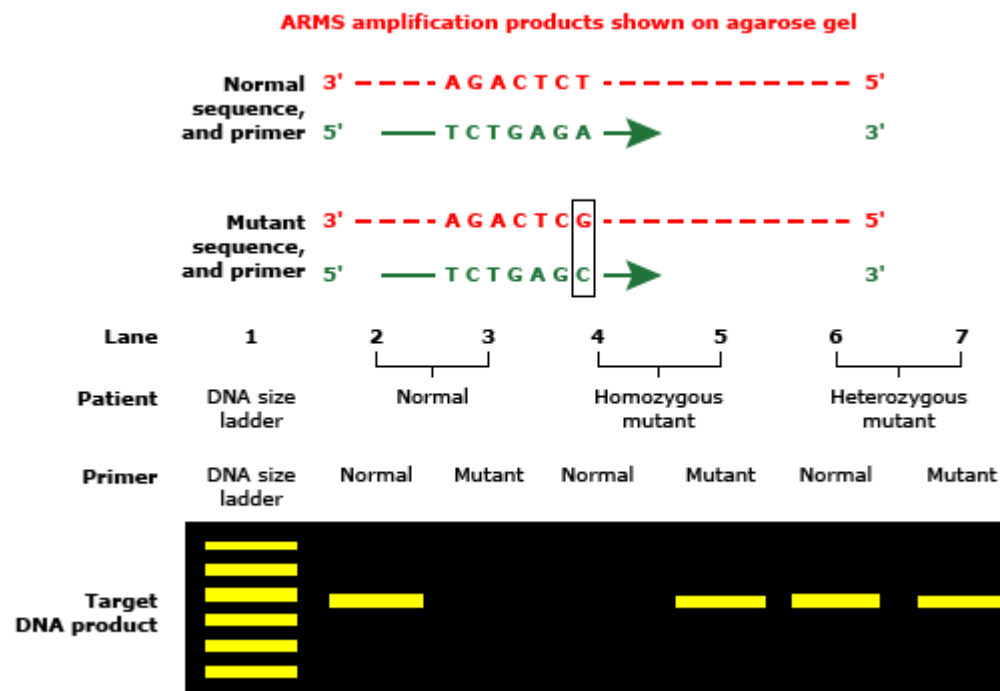
Lane 2: Individual with the normal gene; only the 200 base pair DNA is seen.

Lane 3: Individual heterozygous for the mutation; the normal 200 base pair DNA and the 180 and 20 base pair pieces are seen.

Lane 4: Individual heterozygous for the mutation; only the 180 and 20 base pair pieces are seen.

Graphic 81760 Version 2.0

Amplification refractory mutation system



This example of amplification refractory mutation system (ARMS) PCR shows how this technique can be used to distinguish between an individual who is homozygous for a normal gene and individuals homozygous or heterozygous for mutation in that gene. In this case separate primers are designed to match the normal gene sequence or the mutant gene sequence. The DNA sequences in red represent the patient DNA; the sequences in green are the primers. The box shows the mutant base and the corresponding altered base on the mutant primer.

An agarose gel is run with the PCR products, and the presence of a band shows whether the normal or mutant DNA sequence was present.

Lanes 2 and 3: Individual with the normal gene; only the normal primer gives a PCR product.

Lanes 4 and 5: Individual homozygous for the mutation; only the mutant primer gives a PCR product.

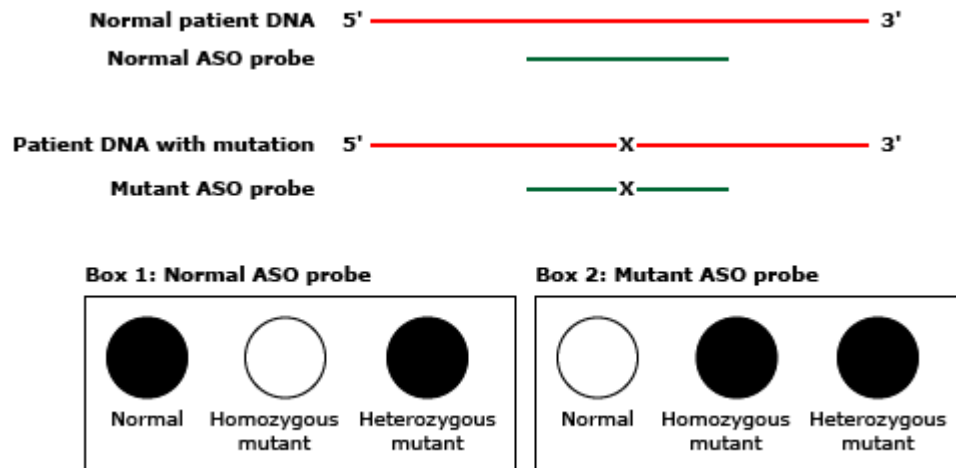
Lanes 6 and 7: Individual heterozygous for the mutation; both the normal and mutant primers give PCR products.

Note: A control DNA product was present in all lanes, verifying that the PCR reaction worked.

ARMS: amplification refractory mutation system; PCR: polymerase chain reaction.

Graphic 55076 Version 2.0

Allele specific oligonucleotide (ASO) hybridization and dot-blot analysis



This representation of ASO hybridization with dot blot analysis shows how this technique can be used to distinguish between an individual who is homozygous for a normal gene from individuals homozygous or heterozygous for a mutation in this gene. Patient DNA is red, and probes are green. In this case a normal gene is detected by a probe with the normal sequence, and a mutant gene is detected by a probe with a mutant sequence. The point mutation is denoted by an "X".

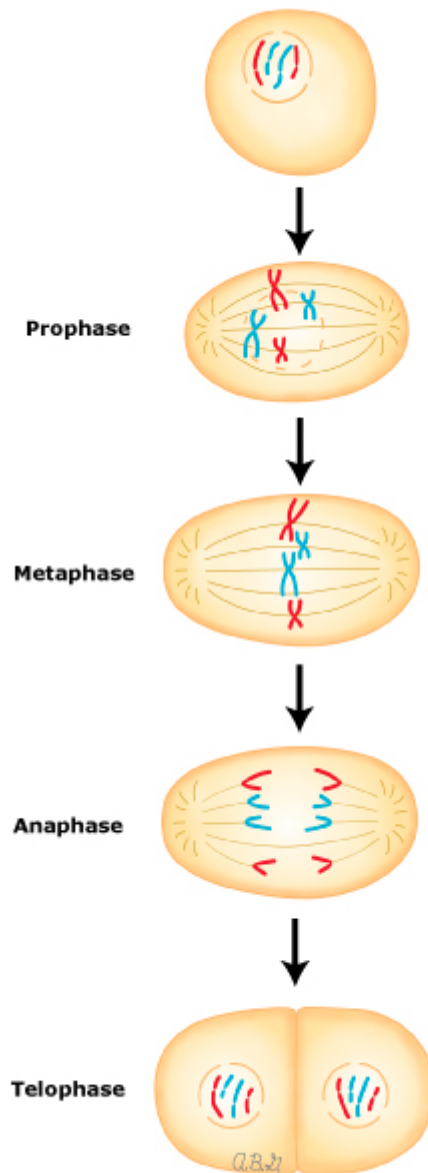
Box 1: the normal probe is hybridized to spots of DNA "dotted" from a normal individual, an individual homozygous for the mutation, and an individual heterozygous for the mutation. The probe gives a signal for the normal and heterozygous individuals, but not the homozygous individual (who lacks the normal gene).

Box 2: the mutant probe is hybridized to spots of DNA "dotted" from a normal individual, an individual homozygous for the mutation, and an individual heterozygous for the mutation. The probe gives a signal for the homozygous and heterozygous individuals but not the normal individual (who lacks the mutant gene).

ASO: allele specific oligonucleotide.

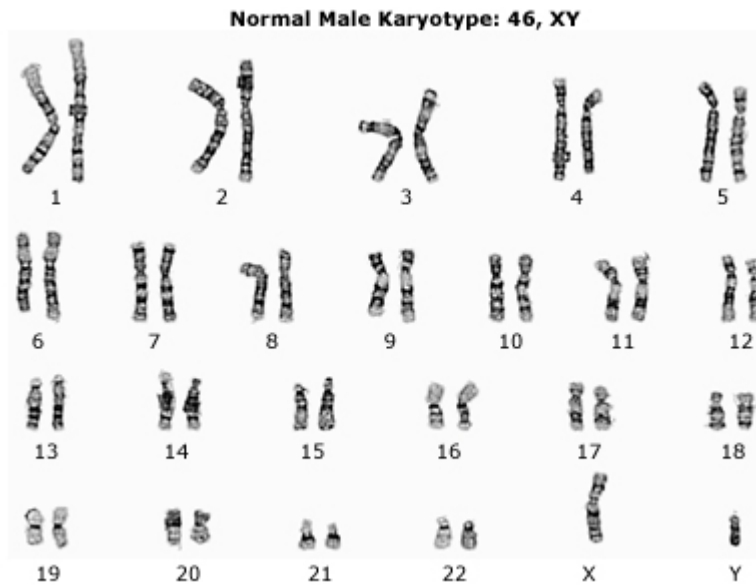
Graphic 80013 Version 2.0

Mitosis



Graphic 52419 Version 2.0

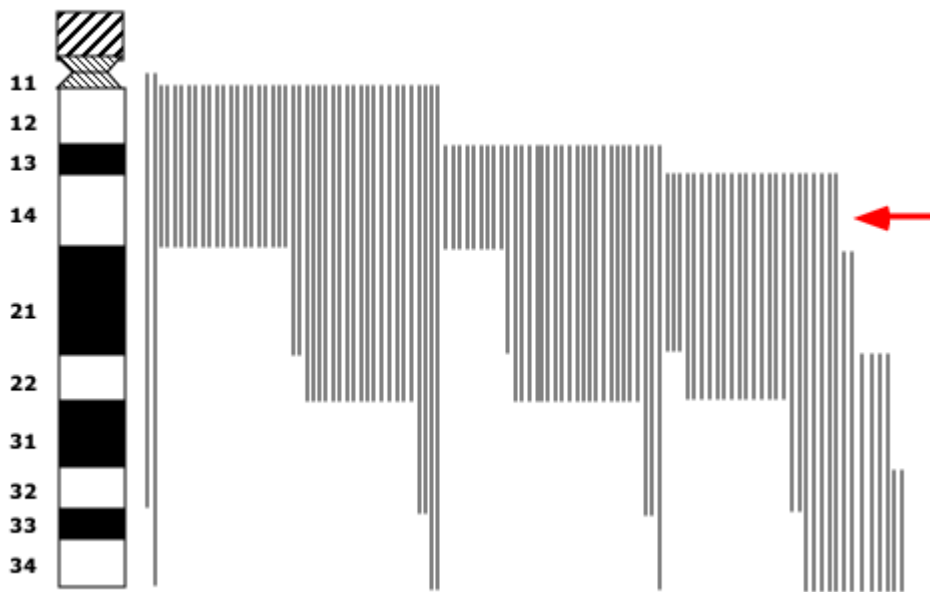
Chromosomal banding



GTG (G-banding with trypsin and Giemsa)-banding technique, of a male with a normal karyotype, 46, XY.

Graphic 81862 Version 1.0

Chromosome 13 deletions in multiple myeloma

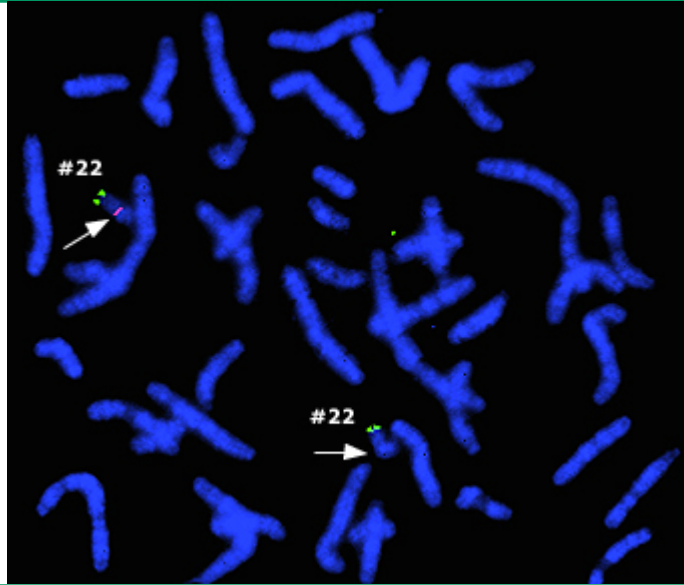


Conventional cytogenetic analysis with G banding was performed in 106 patients with multiple myeloma. Each vertical line to the right of the figure (the long arm of chromosome 13) indicates the part of chromosome 13 that was found to be deleted in one patient. With the exception of eight patients, the minimal region of deletion overlap appears to be in the 13q14 region (red arrow).

Reprinted with permission from Shaughnessy, J, et al. Blood 2000; 96:1505. Copyright 2000, American Society of Hematology.

Graphic 64638 Version 1.0

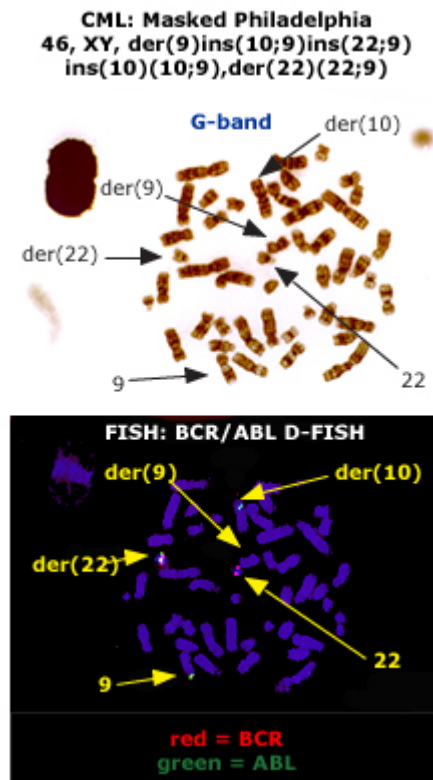
Microdeletion FISH



Fluorescence in situ hybridization (FISH) of a patient with a deletion of chromosome 22q, the DiGeorge/VCF syndrome. As shown, the VCF probe only hybridizes with one chromosome 22 (red line and arrows). By comparison, the fluorescent green probe is a control probe that hybridizes to both chromosomes.

Graphic 67440 Version 2.0

Complex rearrangement in chronic myeloid leukemia

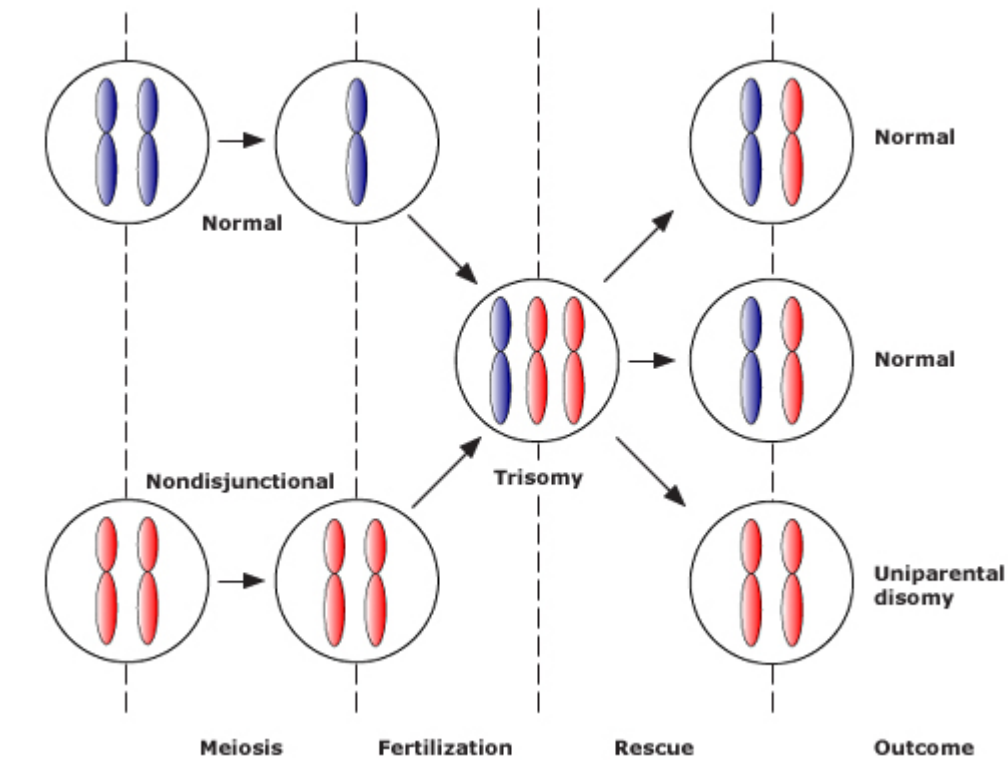


Sequential g-band (top) and FISH (bottom) images of metaphase chromosomes from a complex variant Philadelphia rearrangement without the typical Philadelphia chromosome. The FISH probe is a dual-color combination of probes that span the ABL (green) and BCR (red) breakpoint regions. Juxtaposition of the ABL and BCR signals on the derivative chromosome 22 results in a yellow fusion signal. This complex rearrangement involves submicroscopic insertion of most of the ABL gene into chromosome 22.

Photo courtesy of Athena Cherry, PhD.

Graphic 74102 Version 2.0

Postulated mechanism of uniparental disomy

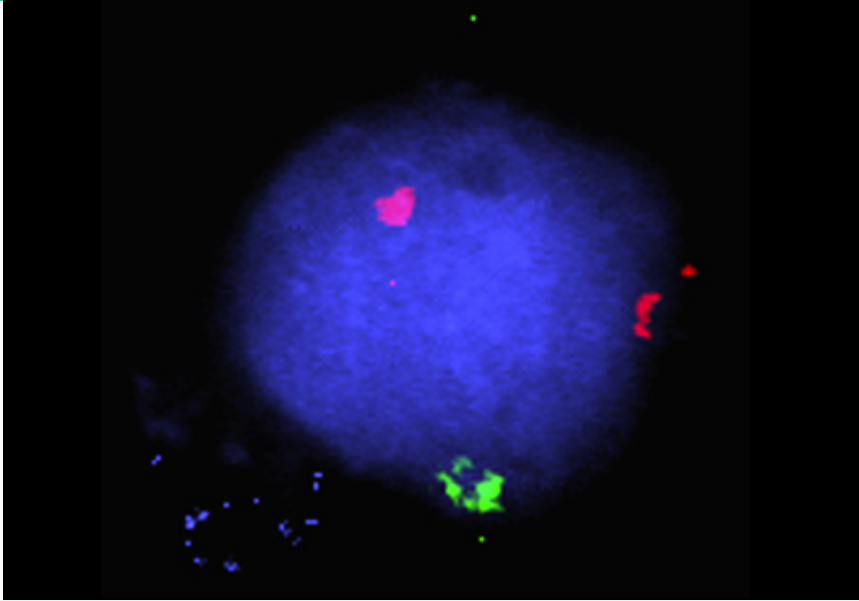


Chromosomal nondisjunction during meiosis leads, after fertilization, to trisomy. Subsequent loss of one chromosome (rescue) could lead to the formation of cells with a chromosome from each parent or cells in which both chromosomes were from the disomic gamete.

Reprinted from Preece, MA, Moore, GE. Genomic Imprinting, Uniparental Disomy and Fetal Growth. Trends in Endocrinology and Metabolism 2000; 11:270. Copyright 2000, with permission from Elsevier Science.

Graphic 72197 Version 1.0

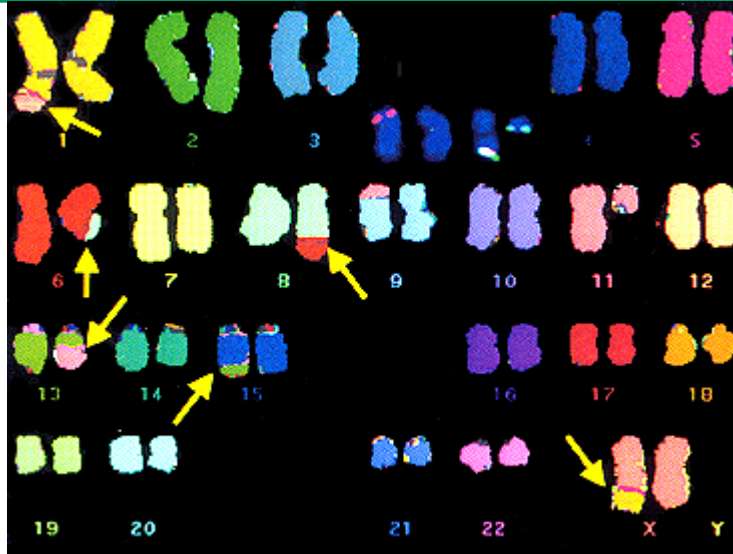
Interphase FISH



Interphase fluorescence in situ hybridization (FISH) is a modification of FISH analysis that allows high resolution analysis because probes are hybridized to minimally condensed interphase chromosomes. In the interphase nucleus, chromosomal structures cannot be discerned and only the hybridized probe will light up. The cell nucleus in this figure was hybridized with a probe for a region on the X chromosome (red) and one for an area on the Y chromosome (green). The presence of three separate signals indicates that this individual has two X chromosomes (red) and one Y chromosome (green). This chromosomal abnormality causes Klinefelter syndrome.

Graphic 60540 Version 1.0

Spectral karyotyping

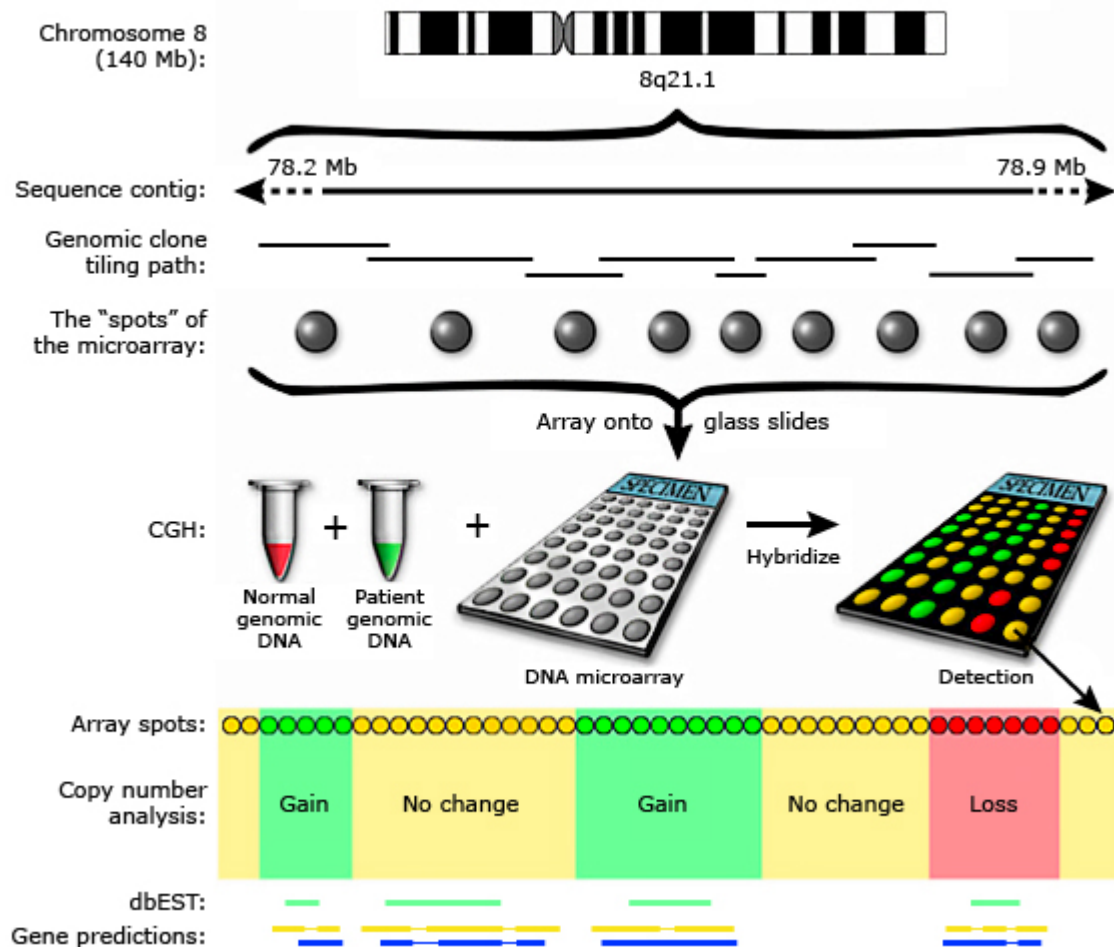


This is an example of a multi-color fluorescence in situ hybridization analysis (M-FISH, spectral karyotyping) of a complex karyotype. In this technique, each chromosome is given a unique color; translocations are easily seen (yellow arrows).

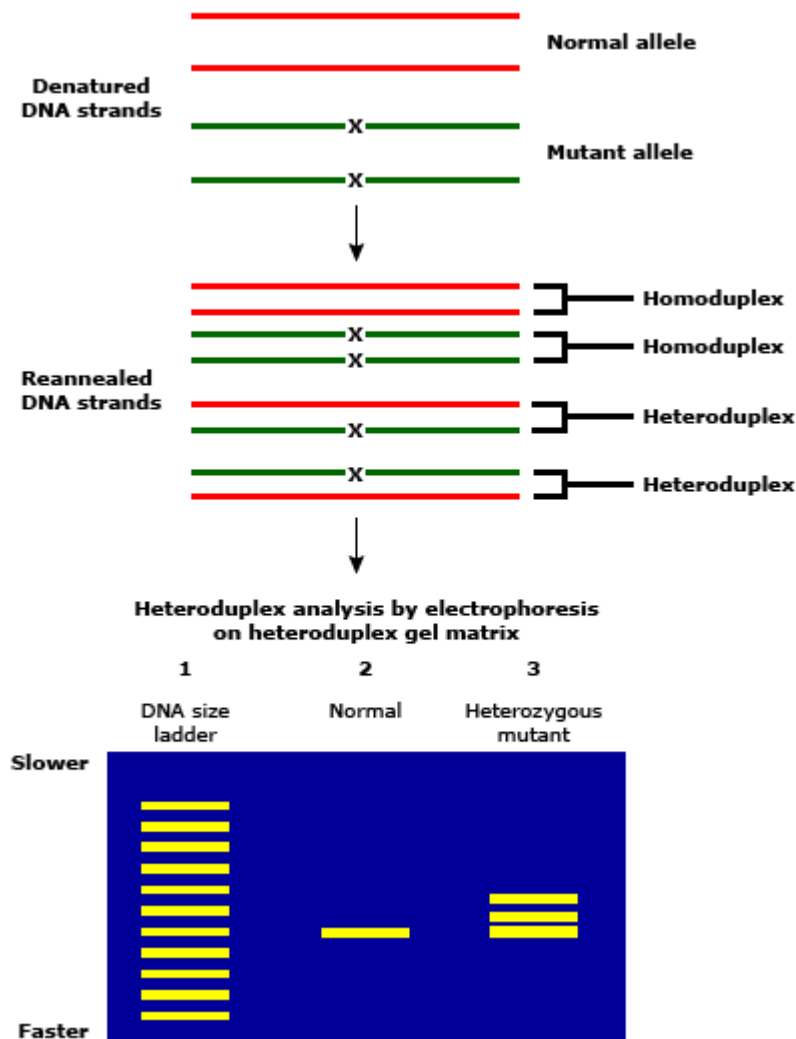
Reproduced with permission from: Jalal, SM, et al. Br J Haematol 2001; 112:975. Copyright © 2001 Blackwell Science.

Graphic 78864 Version 1.0

Schematic CGH for microarray setup



Heteroduplex analysis



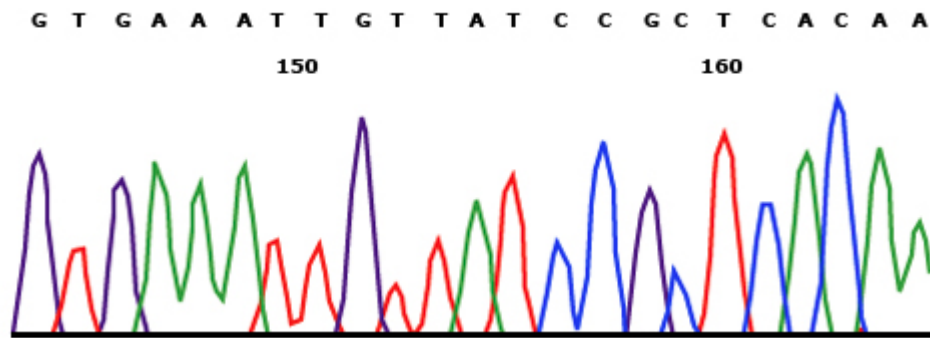
This representation of heteroduplex analysis shows how this technique can be used to distinguish between an individual with normal DNA and heterozygous for a gene mutation. An individual's DNA is amplified by polymerase chain reaction (PCR), and the PCR products from both alleles are mixed and denatured (top section). The PCR products are then reannealed, allowing all combinations of DNA pairing. Homoduplexes are generated by the pairing of normal with normal and mutant with mutant DNA; heteroduplexes are generated from the pairing of normal with mutant DNA. (middle section). The mutation is denoted by an "X".

Lane 2: DNA from an individual homozygous for the normal gene forms only homoduplexes, which run faster through the gel.

Lane 3: DNA from an individual heterozygous for the gene mutation forms heteroduplexes, which contain bubbles. This makes the DNA run more slowly through the gel.

Graphic 55618 Version 3.0

Automated DNA sequencing



This representation of automated DNA sequencing shows how this technique can be used to detect mutations and other changes in DNA sequence. DNA is amplified in the presence of primers and chain terminators. Either the primers or the terminators are fluorescently labeled, with a different color fluorescence for each base. The automated sequencer uses a laser to excite the DNA containing these fluorescently labeled nucleotides and produces a color-coded intensity profile. In this example, guanine (G) is purple, adenine (A) is green, thymine (T) is red, and cytosine (C) is blue. The sequence can be read from the color of the peaks.

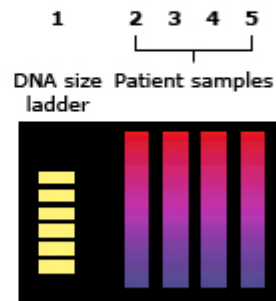
Graphic 77662 Version 2.0

Southern blotting

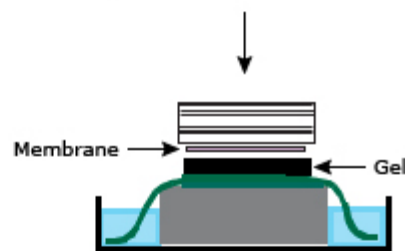
Step 1: Genomic DNA is digested by a restriction enzyme



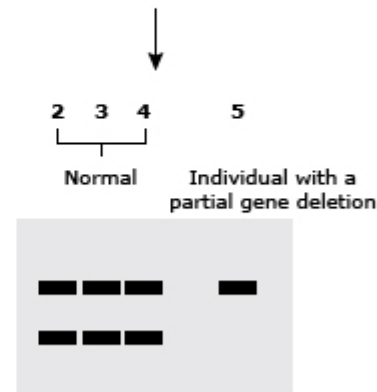
Step 2: DNA is run on a gel



Step 3: DNA is transferred from the gel to a membrane



Step 4: A radioactive probe is hybridized to the membrane and the membrane is exposed to x-ray film



This representation shows the steps in Southern blotting.

Step 1: Genomic DNA is digested by a restriction enzyme.

Step 2: DNA is run on an agarose gel, giving a smeared appearance.

Step 3: DNA is transferred from the gel to a membrane.

Step 4: A radioactive probe is hybridized to the membrane and the membrane is exposed to X-ray film. The bands show the places where the probe hybridized to the patient DNA. If the DNA has a gene deletion or a mutation that eliminates a restriction enzyme site, then one or more bands may be missing. This step can also be performed using a non-radioactive probe.

In the final image, normal individuals (Lanes 2 to 4) show the expected pattern, while an individual with a partial gene deletion (Lane 5) is missing one of the bands.

Graphic 78900 Version 2.0

Disclosures

Disclosures: **Iris Schrijver, MD** Nothing to disclose. **James L Zehnder, MD** Grant/Research Support: GSK [determinants of thrombopoietin agonist response (eltrombopag)]. **Athena M Cherry, PhD** Nothing to disclose. **Benjamin A Raby, MD, MPH** Other Financial Interest (Spouse): Parexel International [Contract research organization (Double-blind randomized clinical trials)]. **Jennifer S Tirnauer, MD** Nothing to disclose.

Contributor disclosures are reviewed for conflicts of interest by the editorial group. When found, these are addressed by vetting through a multi-level review process, and through requirements for references to be provided to support the content. Appropriately referenced content is required of all authors and must conform to UpToDate standards of evidence.

Conflict of interest policy

Chromosomal translocations, deletions, and inversions

Authors

Iris Schrijver, MD
James L Zehnder, MD
Athena M Cherry, PhD

Section Editor

Richard A Larson, MD

Deputy Editor

Jennifer S Tirnauer, MD

All topics are updated as new evidence becomes available and our [peer review process](#) is complete.

Literature review current through: Feb 2015. | **This topic last updated:** Aug 11, 2014.

INTRODUCTION — In this review, we will describe the most common structural chromosomal anomalies, give examples of disease processes resulting from these germline or somatic alterations, and discuss the mechanisms underlying these disorders. Due to the important role of chromosomal abnormalities in the pathogenesis of many hematological disorders, a general review of these specific aberrations is presented separately, as is a review of cytogenetic and molecular genetic tools used to characterize these abnormalities. (See "[General aspects of cytogenetic analysis in hematologic malignancies](#)" and "[Tools for genetics and genomics: Cytogenetics and molecular genetics](#)".)

OVERVIEW — Chromosomal aberrations are due to either numerical abnormalities or structural defects. The normal diploid number of chromosomes in humans is 46. There are 23 pairs of chromosomes with 22 pairs of autosomes and two sex chromosomes, the X and the Y. Human females have two X chromosomes (46,XX), while males have one X and one Y chromosome (46,XY).

Abnormality of chromosome number — A numerical abnormality in a cell's chromosomal endowment may be caused by the gain of one or more complete haploid chromosome sets (polyploid karyotype). An example is the triploid chromosomal number (eg, 69,XXY) in a partial hydatiform mole. (See "[Gestational trophoblastic disease: Pathology](#)", [section on 'Genetics'](#)".)

More commonly, there is a selective gain or loss of an individual chromosome (aneuploidy). As an example, trisomy 21, which causes Down syndrome, is characterized by the gain of one additional copy of chromosome 21. All mutations that change the total number of chromosomes are considered to be **genome mutations**.

Abnormality of chromosomal structure — Structural chromosomal anomalies, also called **chromosomal mutations**, comprise those changes that are due to one or more breaks in a chromosome. Following a break, the separated fragments are likely to participate in chromosomal rearrangements. Structural chromosomal changes can result in a displacement of chromosomal regions without any loss or duplication of genetic material (ie, balanced rearrangements) or they may be unbalanced:

- Balanced rearrangements are frequently inherited and are not commonly associated with phenotypic abnormalities. Two exceptions to this rule include a breakpoint that directly disrupts a gene, and the displacement of chromosomal material from an X chromosome to an autosome or vice versa.
- Unbalanced rearrangements result in a partial trisomy or monosomy and are a frequent cause of congenital abnormalities or developmental delay.

On a submicroscopic level, **gene mutations** such as single base-pair substitutions, insertions, deletions, duplications, nucleotide repeat expansions and inversions are usually too small to be identified by standard karyotypic analysis, and require DNA analysis for their detection. These submicroscopic mutations can be as small as a change in a single base pair, or as large as megabase stretches of DNA

sequence in very large genes. (See ["Overview of genetic variation"](#).)

MECHANISMS OF CHROMOSOMAL BREAKAGE — Chromosomal breakage is caused by double strand DNA breaks of endogenous or exogenous origin. Endogenous breaks are associated with DNA replication, recombination, transcription, and repair processes, all of which require temporary double strand DNA breaks. Permanent endogenous breaks occur in programmed cell death due to aging of the cell, cell damage, or regulation of cell numbers. Exogenous causes include exposure to irradiation and certain chemical substances, as well as other types of DNA damage.

The repair of DNA breaks occurs with high fidelity as long as the damaged area can be removed and replaced by a DNA sequence that is copied from an undamaged template. Double strand DNA breaks, however, need repair across the damaged area on both strands and are more susceptible to errors. By altering gene structure or expression patterns, such mutations can have pathogenic consequences, including malignant transformation.

Chromosomal breakage is also a requirement for subsequent translocations, deletions, inversions, and other chromosomal rearrangements [1]. These will be discussed in the following sections.

TRANSLOCATIONS — Most individuals with a balanced translocation are phenotypically normal. However, early studies demonstrated a significantly higher incidence of "balanced" rearrangements in institutionalized individuals [2]. It is theorized that, in such instances, the translocation directly disrupts a gene (by "position effect") or elicits the loss of a small part of the chromosome.

There are three types of chromosomal translocations: reciprocal, Robertsonian, and insertional (non-reciprocal).

Reciprocal translocations — Congenital reciprocal translocations have an incidence of 0.2 percent and originate from the breakage of two non-homologous chromosomes that interchange their separated parts. Rarely, more than two chromosomes are involved in a reciprocal exchange of chromosome segments. When this occurs, the translocation is considered to be complex.

Normally, genetic material is not lost in a reciprocal translocation. During meiosis, the normal and translocated chromosomes pair by forming a quadrivalent structure, from which one normal, one balanced, and four unbalanced products could be generated ([figure 1](#)). The unbalanced products can be subdivided into two products with correct and two products with incorrect centromere segregation. The latter situation, in which the gamete contains two identical centromeres, is very rare and less compatible with a viable outcome. While the theoretical expectation is that 50 percent of the gametes produced are abnormal or unbalanced, the empirical general risk of having abnormal liveborn offspring rarely exceeds 15 percent.

Reciprocal translocations also frequently occur as a somatic mutation in hematologic malignancies and other neoplasms. The reciprocal translocation in one cell either juxtaposes two genes and induces upregulation of one of these genes, or causes oncogene activation by creation of a novel fusion gene. In both scenarios, clonal expansion of the affected cell directly leads to tumorigenesis. (See ["Genetic abnormalities in hematologic and lymphoid malignancies"](#).)

The first situation is exemplified by the translocation of an oncogene to a gene that either encodes an immunoglobulin heavy or light chain, or a T-cell receptor. These genes naturally undergo extensive rearrangements to ensure the necessary diversity in cell-mediated or antibody responses. This process, however, increases the risk of erroneous interchromosomal rearrangements. An example is t(14;18), which is present in nearly 80 percent of patients with follicular B-cell lymphoma. In this translocation, the BCL2 gene on chromosome 18 is adjoined to the immunoglobulin heavy chain gene on chromosome 14, which is constitutively expressed and causes overexpression of BCL2 [3].

In the second circumstance, the breakpoints occur within two genes and a hybrid gene is created.

Expression of the chimeric protein can change cell signaling pathways due to the altered activity of transcription factors. One of many examples is t(9;22) which causes fusion of the BCR gene on chromosome 22 and the ABL gene on chromosome 9 ([figure 2](#)). This hybrid BCR-ABL1 product is present in all patients with chronic myeloid leukemia [4]. (See "[Genetic abnormalities in hematologic and lymphoid malignancies](#)" and "[Molecular genetics of chronic myeloid leukemia](#)".)

Although it is currently unknown which cellular processes determine that chromosome breaks will result in a reciprocal translocation, some pathogenic mechanisms of juxtaposed and chimeric genes have been elucidated:

- The BCL2 gene encodes proteins that are oncogenic by prevention of programmed cell death rather than by promoting cell proliferation. B-cells with t(14;18) escape apoptosis and are thus predisposed to the accumulation of additional mutations.
- The PRAD1 gene on chromosome 11, which is juxtaposed to the immunoglobulin heavy chain locus on chromosome 14 in many patients with mantle cell lymphoma, encodes the cell cycle regulator cyclin D1. The t(11;14) directly interferes with normal cell cycle events [5].

These examples illustrate the functional diversity of genes that undergo translocation to the immunoglobulin or T-cell receptor genes.

However, the genes most frequently involved in oncogenic translocations are transcription factors. Their dysregulation gives rise to inappropriate target gene activation in one or more cellular pathways [6]. An example is the widely studied BCR-ABL1 fusion gene, which encodes an oncoprotein that is involved in the activation of multiple signal transduction cascades by constitutive expression of the ABL1 oncogene. (See "[Cellular and molecular biology of chronic myeloid leukemia](#)".)

While cell proliferation and survival become independent of the cytokines that normally regulate the differentiation as well as proliferation of the hematopoietic cells, phosphorylation of several substrates activates signal transduction pathways including RAS, JUN kinase, MYC and STAT [7]. The induction of additional transcription factors may upregulate or reduce the expression of a potentially large number of genes. Finally, the BCR-ABL1 fusion protein may also influence the cell cycle directly [8].

Robertsonian translocations — The acrocentric human chromosomes 13, 14, 15, 21 and 22 have very short p-arms that contain only chromosomal satellites and the genetic code for ribosomal RNAs. Their centromeres, therefore, are very close to one end of the chromosome. When the long arms of two acrocentric chromosomes merge by translocation, their short arms are lost and, depending upon the location of the breakpoints, a dicentric or monocentric fusion-chromosome is created.

Loss of the short arms of acrocentric chromosomes does not have phenotypic consequences, because the lost sections do not contain unique genetic sequences. The only consequence is a balanced karyotype with 45 instead of 46 chromosomes. This is identified in approximately 1 in 1000 individuals.

However, unbalanced gametes of heterozygous carriers are common and give rise to a monosomic or trisomic fetus. Most monosomies and trisomies are lethal and spontaneously abort early in the pregnancy. Surviving fetuses with trisomy 21, however, can be viable and affected with Down syndrome [9], although about 85 percent of trisomy 21 conceptuses will not make it to term. When this syndrome is caused by a parental Robertsonian translocation, the carrier parent typically has a fusion of the long arms of chromosomes 14 and 21 ([figure 3](#)). One of the three viable gametes will be balanced, one will be normal, and one will contain the fused chromosome [der(14;21)] as well as the unaffected chromosome 21. Normal fertilization of this gamete results in a fetus with trisomy 21. The empiric risk for a child with Down syndrome is 10 to 15 percent for a carrier mother, but only 2 percent if the father carries the Robertsonian translocation.

Insertional translocations — Insertional translocations, by definition, are not reciprocal and are caused

by integration of a segment from a "donor" chromosome that has incurred two breakpoints into a recipient chromosome that contains only one breakpoint. The resulting zygotes may be normal, balanced, partially trisomic, or partially monosomic. The phenotypic outcome of this anomaly, once considered rare, strongly depends upon the location of the breakpoints and on the nature of the translocated sequence.

With the advent of array comparative genomic hybridization (aCGH, also known as chromosomal microarray assay), it appears that insertional translocations (IT) are not as rare as previously believed. In 2000, the estimated incidence of microscopically visible ITs was 1/80,000, whereas more recent studies have shown them to be more common with one estimate of 1/500 [10-14].

DELETIONS — Deletion of chromosomal material may or may not be detected by standard chromosomal banding techniques (ie, macrodeletions and microdeletions, respectively). (See ["Tools for genetics and genomics: Cytogenetics and molecular genetics"](#), section on 'Genotyping new mutations'.)

Macrodeletions — To microscopically see a chromosomal deletion by chromosome banding, the deletion must span at least three to five megabases (Mb). Considering that the average gene density approximates one gene per 50 kilobases (kb), it is not surprising that the loss of such a large region is likely to lead to phenotypic manifestations. Perhaps the most well-known example is Cri-du-chat syndrome, which is due to a terminal deletion of the short arm of chromosome 5 ([figure 4](#)). Congenital abnormalities in this disorder include cat-like neonatal crying, microcephaly with dysmorphic facial features, cardiac defects, hypotonia, and severe mental retardation. Even though the size of the deleted fragment differs from patient to patient, chromosomal bands 5p15.2 and 5p15.3 are always included. These bands, therefore, encompass the critical region for this phenotype [15].

Another example of a variable size of the deleted fragment has been described in patients with multiple myeloma. In 106 patients, all but eight had a deletion of the long arm of chromosome 13 which included the 13q14 region ([figure 5](#)) [16].

Microdeletions — Cytogenetic developments, such as high resolution banding techniques and fluorescence in situ hybridization (FISH), have enabled the detection of relatively small ("micro") deletions. If the FISH probe is designed to localize to the region that is deleted on one allele, only one double signal will be present in the (pro)metaphase chromosomal evaluation. (See ["Tools for genetics and genomics: Cytogenetics and molecular genetics"](#).)

One example of a microdeletion syndrome is Smith-Magenis syndrome, with an incidence of 1 in 25,000 and characterized by a microdeletion of chromosome band 17p11.2. Phenotypic features include intellectual disability, a deep and hoarse voice, self-abusive behavior, insomnia, short stature and brachydactyly. While most deletions span about 5 Mb and are cytogenetically discernible near the centromere, the remainder must be diagnosed with more refined diagnostic methods, such as FISH ([image 1](#)) [17]. The interstitial deletion in this syndrome is facilitated by the presence of regions of low-copy repeats that flank the common breakpoint regions [18]. These segments are highly homologous, which makes them prone to unequal crossing over in meiosis I. During unequal crossing over, two non-sister chromatids recombine at an incorrect location due to inappropriate alignment of highly homologous DNA sequences. Thus, one of the chromatids acquires a duplication of the intervening sequence, while the corresponding segment is deleted on the other chromatid ([figure 6](#)). This is called non-allelic homologous recombination (NAHR), and is the basis of most microdeletion/microduplication syndromes. (See ["Microdeletion syndromes \(chromosomes 1 to 11\)"](#) and ["Microdeletion syndromes \(chromosomes 12 to 22\)"](#).)

Another example of a microdeletion syndrome, which is the consequence of NAHR, is Williams syndrome (or Williams-Beuren syndrome, WBS). Phenotypic features include cardiovascular disease (most frequently supravalvular aortic stenosis), distinctive faces with full lips and periorbital fullness, and

an unusual and intellectual disability with relative cognitive strength in verbal abilities. Over 99 percent of individuals with Williams syndrome have a contiguous gene deletion which encompasses the elastin (ELN) gene at 7q11.23. Approximately 95 percent of individuals affected with WBS have a 1.55 Mb microdeletion, while the other 5 percent have a larger 1.84 Mb microdeletion. (See "[Microdeletion syndromes \(chromosomes 1 to 11\)](#)".)

Homologous recombination during mitosis is both rare and abnormal. Nevertheless, unequal crossing over can also occur in mitotic division and leads to unequal sister chromatid exchange. The basic mechanism of misalignment between repeat sequences is the same as in meiosis. This abnormal event in mitosis, however, results in a somatic mutation and can predispose to cancer if the deleted sequence contains (or is part of) a tumor suppressor gene.

A subset of patients with hypereosinophilic syndrome have a novel FIP1L1-PDGFR α fusion due to a large interstitial deletion on chromosome 4q12. This results in a fusion protein with constitutive tyrosine kinase activity, eosinophilic proliferation and end-organ damage. Patients with this fusion frequently respond to the tyrosine kinase inhibitor, [imatinib](#) mesylate. Diagnosis of this fusion protein can be made by demonstrating the loss of the CHIC2 gene within the deleted region by FISH, or by PCR demonstrating the presence of a fusion transcript [19,20]. (See "[Clinical manifestations, pathophysiology, and diagnosis of the hypereosinophilic syndromes](#)", section on 'Myeloproliferative HES variants'.)

INVERSIONS — Chromosomal inversions are characterized by two breaks on the same chromosome, rotation of the intervening segment by 180 degrees, followed by chromosomal reintegration of the intervening segment in an "upside down" position. Inversions can be subclassified based on the location of their breaks. Breakpoints on just one chromosome arm define paracentric inversions, whereas pericentric inversions result from breaks on both sides of the centromere. Inversion carriers are typically phenotypically normal. However, inversion carriers may have reproductive issues. At meiosis the inverted chromosome must pair with the non-inverted, normal chromosome forming a loop. If crossing-over occurs, unbalanced or abnormal gametes may result.

Paracentric inversions — Paracentric inversions ([figure 7](#)) are relatively rare, but have been observed in all 23 autosomes and both sex chromosomes [21]. The chromosomes in individuals with a paracentric inversion are more likely to be ascertained because of infertility or repeated miscarriages by the carrier, than because of phenotypic abnormalities in the offspring. The risk of an affected child is generally very small, because of one of the following mechanisms:

- The transmitted chromosome is normal
- The transmitted chromosome is inverted and compatible with life, as in the parent
- After recombination in the inverted region, the transmitted chromosome becomes acentric or dicentric.

Gametes with acentric fragments or dicentric chromosomes are not typically viable. In general, inherited paracentric inversions are innocuous. However, there are rare reported cases of abnormal offspring in individuals with paracentric inversions.

Pericentric inversions — A pericentric inversion ([figure 8](#)) not only causes changes in the banding pattern of the affected chromosome, but may change the position of the centromere as well. Telomeric to the inverted segment, recombinant chromosomes will be partially duplicated or deleted [22]. Larger segments are more susceptible to crossing over in meiosis I, since generally at least one cross-over takes place per chromosome arm. Hence, larger pericentric inversions are associated with an increased risk of phenotypic anomalies in offspring because the duplicated or deleted segments will be smaller and more likely compatible with life.

In cells with malignant transformation due to somatic mutation, reciprocal translocations and deletions

are much more commonly identified than inversions. As a class, inversions are encountered in only 2 percent of all chromosome aberrations seen in cancer patients who have one or multiple chromosome anomalies.

SUMMARY

- Chromosomal aberrations are due to either numerical abnormalities or structural defects. Polyploidy is the gain of one or more complete set of haploid chromosomes (eg, 69,XXY in a partial hydatiform mole). Aneuploidy, gain or loss of an individual chromosome, is more common. Trisomy 21, which causes Down syndrome, is characterized by the gain of one additional copy of chromosome 21. Polyploidy and aneuploidy represent **genome mutations**. (See '[Abnormality of chromosome number](#)' above.)
- Structural chromosomal anomalies, or **chromosomal mutations**, are due to one or more breaks in a chromosome. Structural chromosomal changes can be balanced (without any loss or gain of genetic material, and typically phenotypically normal) or unbalanced (with loss and gain of genetic material, and typically phenotypically abnormal). (See '[Abnormality of chromosomal structure](#)' above.)
- There are three types of chromosomal translocations: reciprocal, Robertsonian, and insertional (non-reciprocal). Congenital reciprocal translocations can result in gametes with unbalanced genetic material, although most viable offspring carry normal or balanced chromosomes. Reciprocal translocations frequently occur as a somatic mutation in hematologic and other malignancies. Transcription factor genes are most frequently involved in oncogenic translocation. (See '[Reciprocal translocations](#)' above.)
- Robertsonian translocations involve acrocentric chromosome with short p-arms that can be lost in translocation of the long arms, resulting in a balanced karyotype with 45, rather than 46, chromosomes and a normal phenotype. However, unbalanced gametes of heterozygous carriers are common and give rise to a monosomic or trisomic fetus. Trisomy 21 can be caused by a parent with a Robertsonian translocation. (See '[Robertsonian translocations](#)' above.)
- Insertional translocations are caused by integration of a segment from a "donor" chromosome that has incurred two breakpoints into a recipient chromosome that contains only one breakpoint. The resulting zygotes may be normal, balanced, partially trisomic, or partially monosomic. (See '[Insertional translocations](#)' above.)
- Macrodeletions, which can be seen microscopically, must span at least three to five megabases (Mb) and therefore are likely to have phenotypic manifestations. Examples are found in patients with Cri-du-chat syndrome and some patients with multiple myeloma. (See '[Macrodeletions](#)' above.)
- Smaller deletions can be detected by high-resolution banding techniques and fluorescence in situ hybridization (FISH), as well as by array comparative genomic hybridization (aCGH). Most microdeletion or microduplication syndromes result from non-allelic homologous recombination (NAHR), in which highly homologous repeat DNA sequences lead to unequal crossing over and erroneous recombination in meiosis I. Similarly, unequal crossing over in mitotic division can result in a somatic mutation which can predispose to cancer if the deleted sequence contains a tumor suppressor gene. (See '[Microdeletions](#)' above.)
- Inversions, in which a segment of the chromosome is rotated 180 degrees, can be paracentric, involving one chromosome arm, or pericentric, involving both arms of the chromosome. Chromosomes in individuals with inversions are likely to be recognized at the time of evaluation for infertility or repeated miscarriages by the carrier, or found incidentally during prenatal diagnosis. A pericentric inversion causes changes in the banding pattern of the affected chromosome, and may

change the position of the centromere. Larger pericentric inversions are associated with an increased risk of offspring with phenotypic anomalies because the duplicated or deleted segments will be smaller and more likely compatible with life. (See '[Inversions](#)' above.)

Use of UpToDate is subject to the [Subscription and License Agreement](#).

REFERENCES

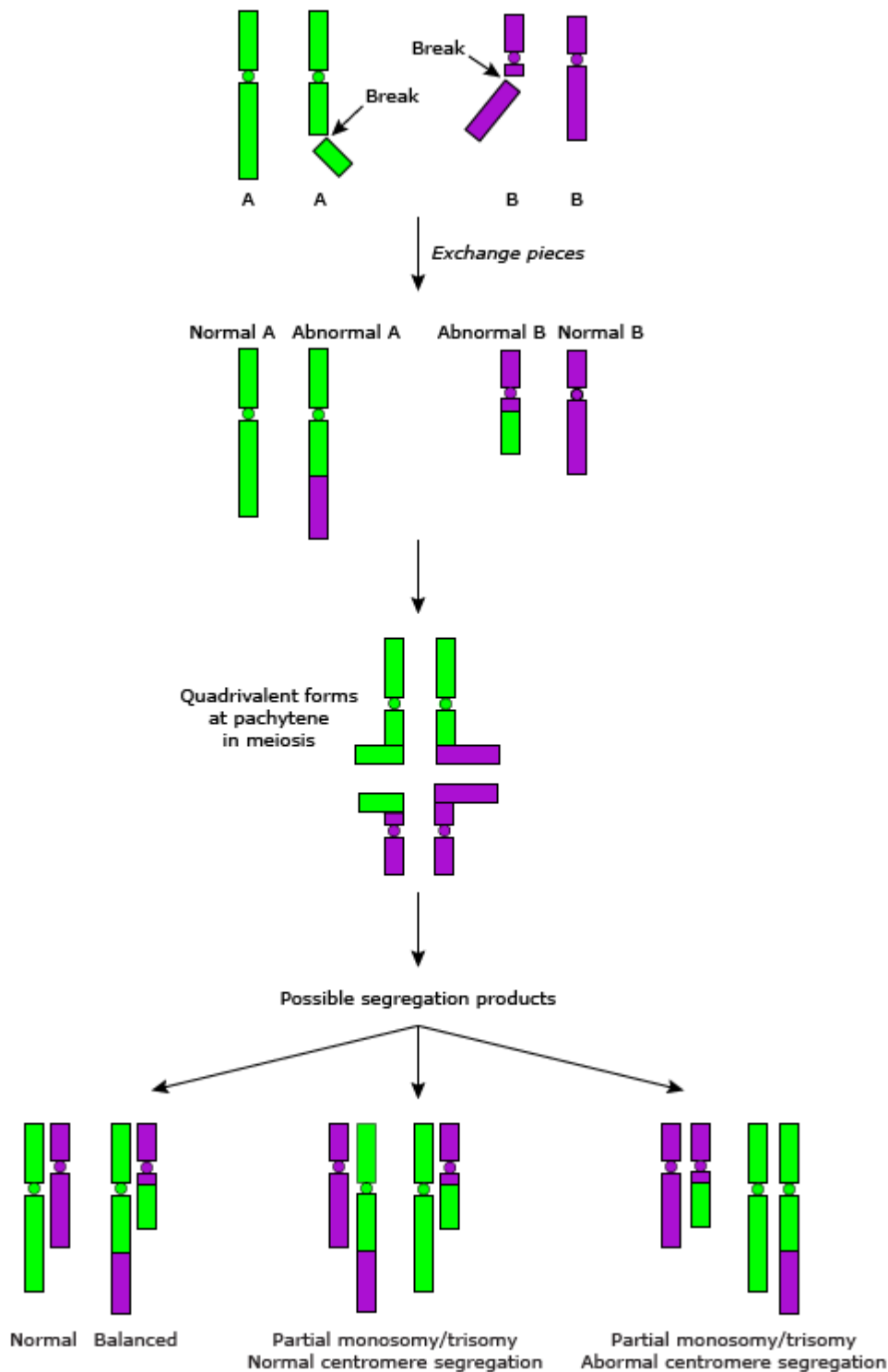
1. Vamvakas S, Vock EH, Lutz WK. On the role of DNA double-strand breaks in toxicity and carcinogenesis. *Crit Rev Toxicol* 1997; 27:155.
2. Warburton, D. De novo structural rearrangements: implications in prenatal diagnosis. In: *Clinical Genetics: Problems in Diagnosis and Counseling*, Wiley, AM, Carter, TP, Kelly, S, Porter (Eds), Academic Press, New York 1982. p.63.
3. Rezuze WN, Abernathy EC, Tsongalis GJ. Molecular diagnosis of B- and T-cell lymphomas: fundamental principles and clinical applications. *Clin Chem* 1997; 43:1814.
4. Melo JV. BCR-ABL gene variants. *Baillieres Clin Haematol* 1997; 10:203.
5. Griesser H. Applied molecular genetics in the diagnosis of malignant non-Hodgkin's lymphoma. *Diagn Mol Pathol* 1993; 2:177.
6. Mitelman F, Johansson B, Mertens F. The impact of translocations and gene fusions on cancer causation. *Nat Rev Cancer* 2007; 7:233.
7. Sawyers CL. Chronic myeloid leukemia. *N Engl J Med* 1999; 340:1330.
8. Warmuth M, Danhauser-Riedl S, Hallek M. Molecular pathogenesis of chronic myeloid leukemia: implications for new therapeutic strategies. *Ann Hematol* 1999; 78:49.
9. Mutton D, Alberman E, Hook EB. Cytogenetic and epidemiological findings in Down syndrome, England and Wales 1989 to 1993. National Down Syndrome Cytogenetic Register and the Association of Clinical Cytogeneticists. *J Med Genet* 1996; 33:387.
10. Kang SH, Shaw C, Ou Z, et al. Insertional translocation detected using FISH confirmation of array-comparative genomic hybridization (aCGH) results. *Am J Med Genet A* 2010; 152A:1111.
11. de Ravel T, Aerssens P, Vermeesch JR, Fryns JP. Trisomy of chromosome 16p13.3 due to an unbalanced insertional translocation into chromosome 22p13. *Eur J Med Genet* 2005; 48:355.
12. Shinawi M, Cheung SW. The array CGH and its clinical applications. *Drug Discov Today* 2008; 13:760.
13. Bi W, Sapir T, Shchelochkov OA, et al. Increased LIS1 expression affects human and mouse brain development. *Nat Genet* 2009; 41:168.
14. Carvalho CM, Zhang F, Liu P, et al. Complex rearrangements in patients with duplications of MECP2 can occur by fork stalling and template switching. *Hum Mol Genet* 2009; 18:2188.
15. Marinescu RC, Johnson EI, Dykens EM, et al. No relationship between the size of the deletion and the level of developmental delay in cri-du-chat syndrome. *Am J Med Genet* 1999; 86:66.
16. Shaughnessy J, Tian E, Sawyer J, et al. High incidence of chromosome 13 deletion in multiple myeloma detected by multiprobe interphase FISH. *Blood* 2000; 96:1505.
17. Juyal RC, Greenberg F, Mengden GA, et al. Smith-Magenis syndrome deletion: a case with equivocal cytogenetic findings resolved by fluorescence in situ hybridization. *Am J Med Genet* 1995; 58:286.
18. Chen KS, Manian P, Koeuth T, et al. Homologous recombination of a flanking repeat gene cluster is a mechanism for a common contiguous gene deletion syndrome. *Nat Genet* 1997; 17:154.
19. Cools J, DeAngelo DJ, Gotlib J, et al. A tyrosine kinase created by fusion of the PDGFRA and

- FIP1L1 genes as a therapeutic target of imatinib in idiopathic hypereosinophilic syndrome. *N Engl J Med* 2003; 348:1201.
20. Gotlib J. Eosinophilic myeloid disorders: new classification and novel therapeutic strategies. *Curr Opin Hematol* 2010; 17:117.
 21. Pettenati MJ, Rao PN, Phelan MC, et al. Paracentric inversions in humans: a review of 446 paracentric inversions with presentation of 120 new cases. *Am J Med Genet* 1995; 55:171.
 22. Kaiser P. Pericentric inversions. Problems and significance for clinical genetics. *Hum Genet* 1984; 68:1.

Topic 4490 Version 13.0

GRAPHICS

Mechanism of reciprocal translocations

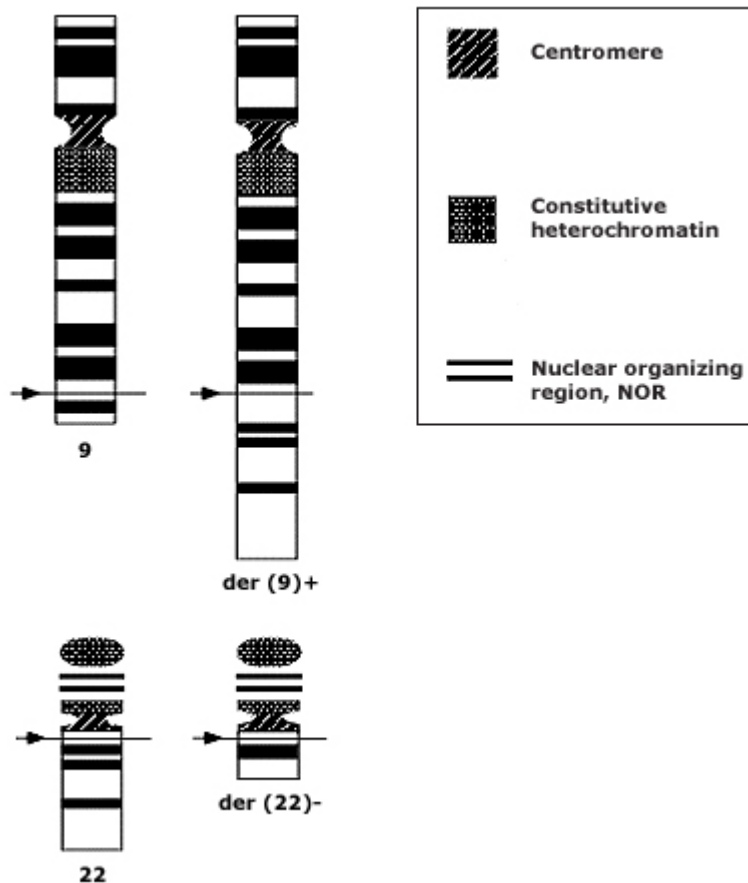


Nonhomologous chromosomes exchange parts without loss of genetic content. When meiotic division occurs, the normal and translocated chromosomes can only pair by forming a quadrivalent structure. The resulting segregation products include a normal chromosome complement, a balanced rearrangement, and four unbalanced products. The unbalanced products can be subdivided into two

products with correct and two products with incorrect centromere segregation.
The latter event is rare and associated with decreased viability.

Graphic 58692 Version 9.0

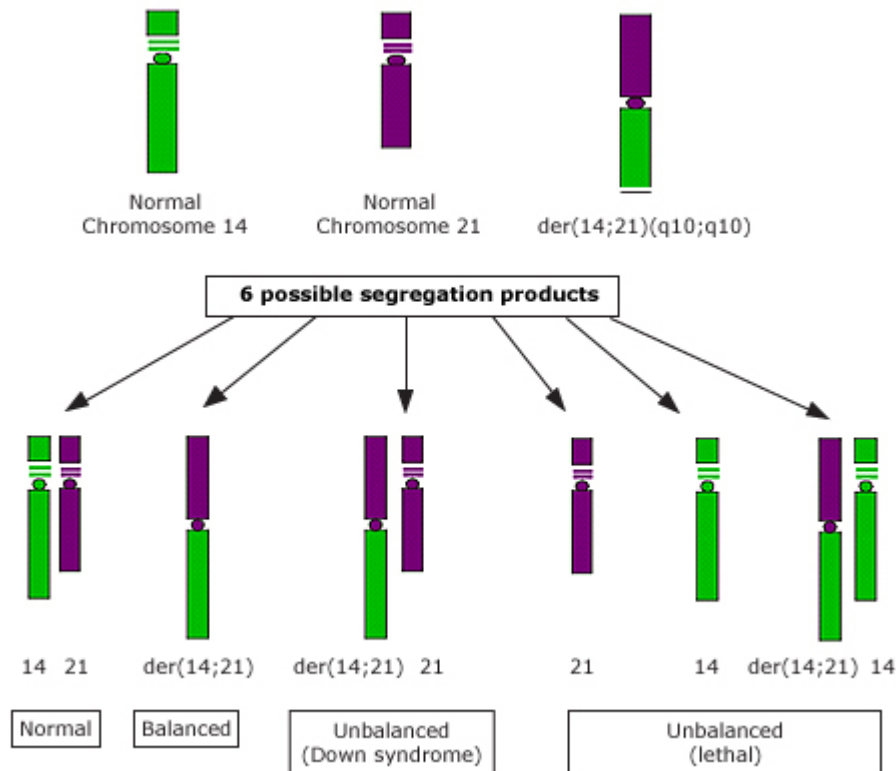
Reciprocal translocation in CML



Example of the reciprocal translocation $t(9;22)$, which is a characteristic finding in chronic myelogenous leukemia (CML). A small, distal piece of chromosome 9 (indicated by the arrow in the left upper chromosome) is broken off and translocated to the distal chromosome 22, as illustrated by the derivative chromosome $der(22)-$. The chromosomal segment that was originally present in the normal chromosome 22 (below the arrow in the left lower chromosome) is translocated to the shortened chromosome 9 to form a derivative, large chromosome $der(9)+$.

Graphic 72417 Version 1.0

Robertsonian translocation

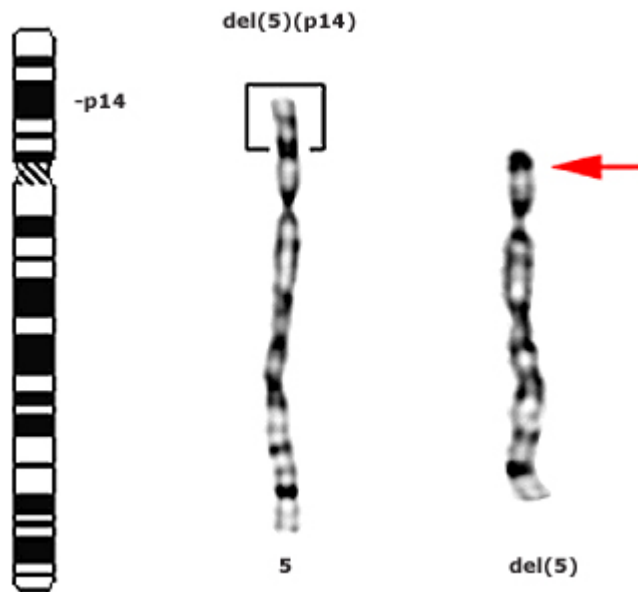


Example of a Robertsonian translocation in which the long arms of one chromosome 14 and one chromosome 21 are fused in the carrier parent (upper panel) is a cause of Down syndrome in the offspring. One of the three viable gametes will be normal, one will have a balanced rearrangement, and one will contain the fused chromosome [der(14;21)] as well as the unaffected chromosome 21. Normal fertilization of this gamete results in a fetus with trisomy 21. Other possible segregation products are gametes lacking a chromosome 21, gametes lacking a chromosome 14, and gametes with one chromosome 14 and a derivative chromosome der(14;21), all of which are not viable.

Courtesy of Iris Schrijver, MD.

Graphic 81746 Version 1.0

Cri du chat syndrome

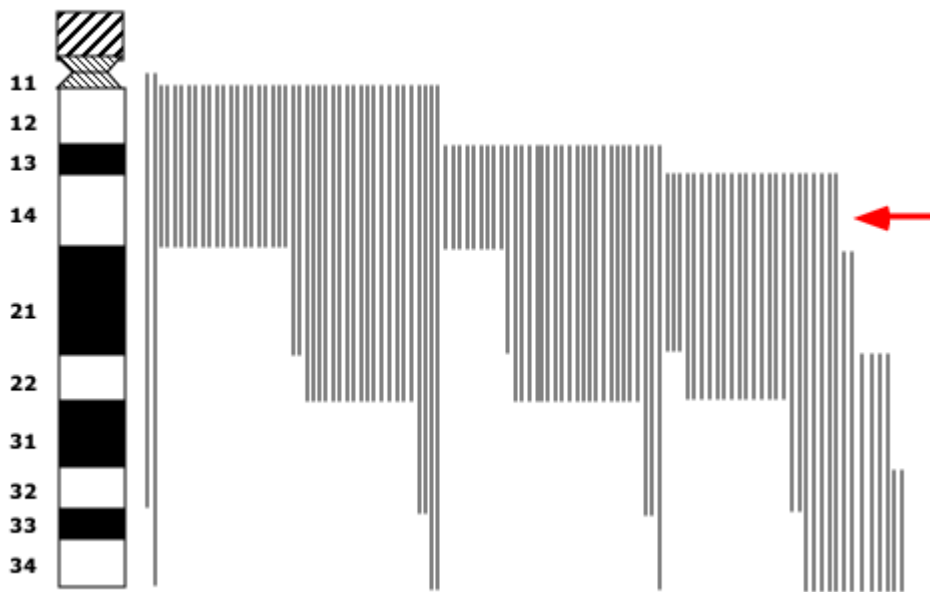


The cri du chat syndrome is due to a large deletion of the terminal short arm of chromosome 5. The size of the deletion may be different in each patient, but it always encompasses chromosomal bands 5p15.2 and 5p15.3. The representation of chromosome 5 (on the left) indicates the deletion breakpoint that was identified in one patient. The shortened p-arm of a chromosome 5, stained by GTG (G-banding with trypsin and Giemsa)-banding, is shown on the right. For comparison, a chromosome 5 from a normal individual is placed in the middle. The boxed fragment indicates the extent of the deletion in the affected chromosome.

Courtesy of Athena Cherry, Stanford Hospital and Clinics.

Graphic 66560 Version 3.0

Chromosome 13 deletions in multiple myeloma

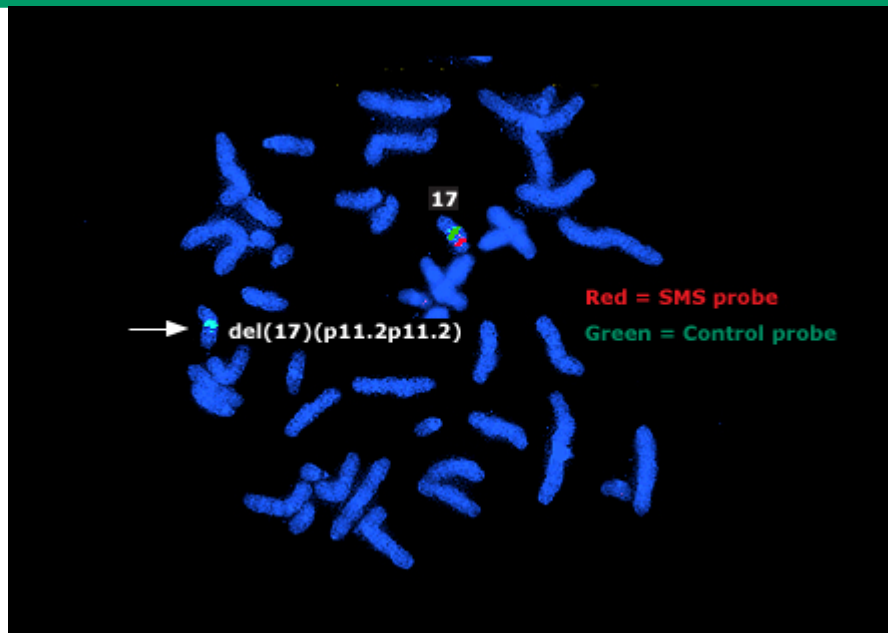


Conventional cytogenetic analysis with G banding was performed in 106 patients with multiple myeloma. Each vertical line to the right of the figure (the long arm of chromosome 13) indicates the part of chromosome 13 that was found to be deleted in one patient. With the exception of eight patients, the minimal region of deletion overlap appears to be in the 13q14 region (red arrow).

Reprinted with permission from Shaughnessy, J, et al. Blood 2000; 96:1505. Copyright 2000, American Society of Hematology.

Graphic 64638 Version 1.0

Smith-Magenis syndrome

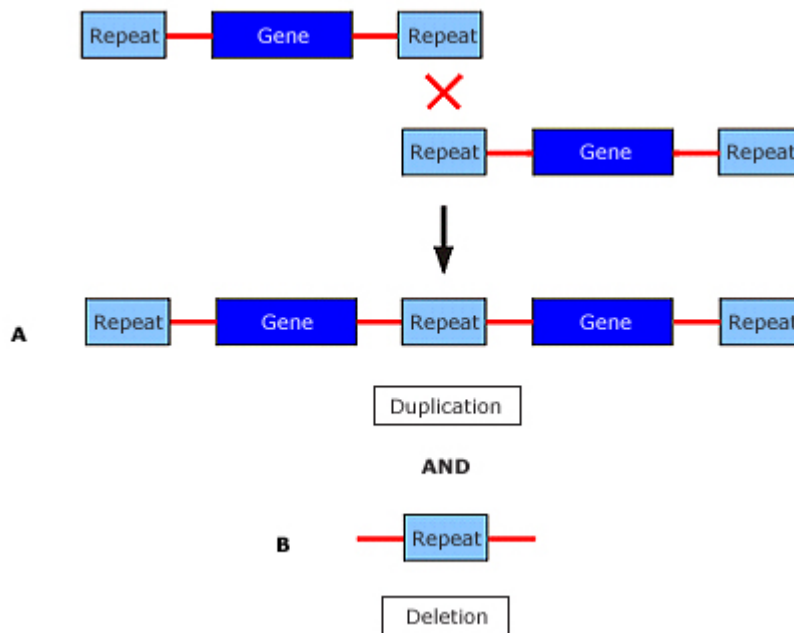


Smith-Magenis syndrome (SMS) is typified by a microdeletion of chromosome band 17p11.2. Fluorescence in situ hybridization (FISH) analysis of a patient's chromosomes is shown. The green control probe shows that an unrelated region of chromosome 17 is present in both chromosomes 17. The red probe, which detects the 17p11.2 region commonly deleted in SMS, only fluoresces on a single chromosome. This finding strongly supports the diagnosis of SMS.

Courtesy of Athena Cherry, Stanford Hospital and Clinics.

Graphic 67417 Version 3.0

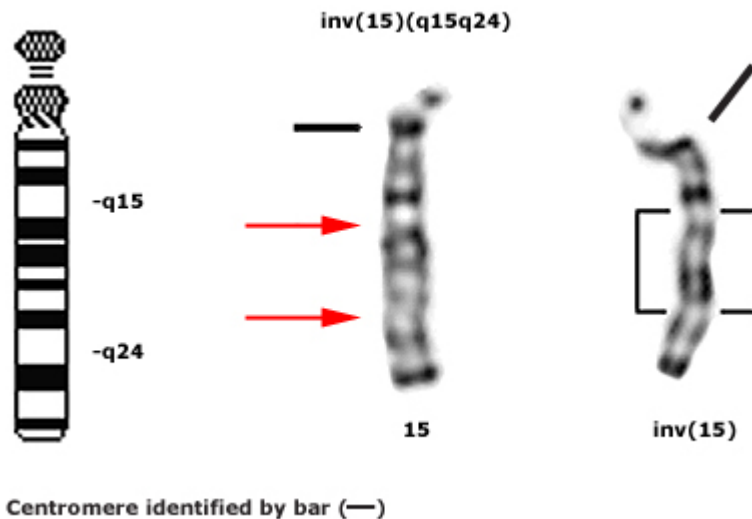
Molecular mechanism of unequal crossing over



A common mechanism of interstitial deletions is unequal crossing over in meiosis I. When highly homologous repeats ("Repeat") flank a DNA segment ("Gene"), two non-sister chromatids may recombine at an incorrect location due to inappropriate alignment. Thus, one of the chromatids (labeled "A") acquires a duplication of the intervening sequence, while the corresponding segment is deleted on the other chromatid and only one repeat remains (labeled "B").

Graphic 63004 Version 1.0

Paracentric inversion

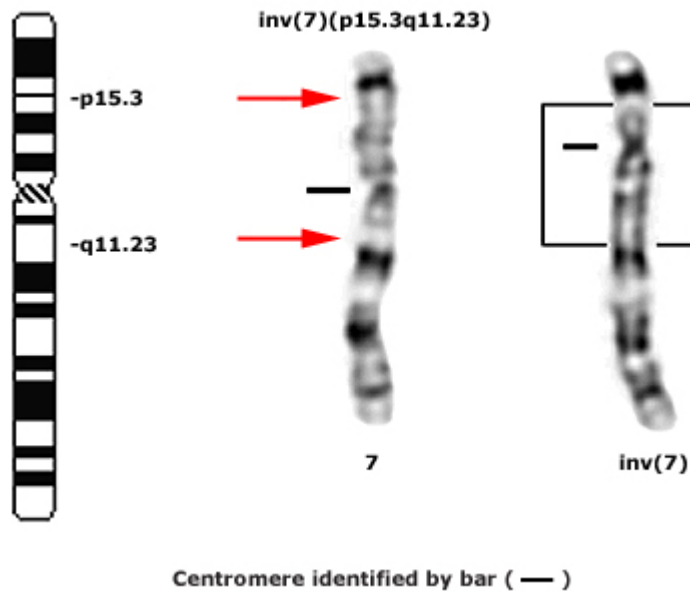


In a paracentric inversion, the inverted chromosome segment is located on one side of the centromere (ie, on one chromosome arm). A representation of normal chromosome 15 is shown on the left. An example of a GTG stained chromosome 15 from a patient with a paracentric inversion of the long arm of chromosome 15 between bands q15 and q24 is shown on the right. A normal GTG stained chromosome 15 is shown in the center. The red arrows show the locations on the normal chromosome where breaks occurred in the abnormal chromosome (corresponding to the boxed area).

GTG: G-banding with trypsin and Giemsa

Graphic 63995 Version 2.0

Pericentric inversion of chromosome 7



In a pericentric inversion, the inverted chromosome segment involves part of both chromosome arms and spans the centromere. This patient (GTG stained chromosome on the right) has a pericentric inversion of chromosome 7. The boxed area corresponds to the area between the two red arrows in the normal chromosome 7 (center) and to the indicated bands on the representation of chromosome 17 (left). The black bar to the left of both GTG banded chromosomes shows the location of the centromere. This pericentric inversion changed the position of the centromere.

GTG: G-banding with trypsin and Giemsa

Graphic 51808 Version 2.0

Disclosures

Disclosures: **Iris Schrijver, MD** Nothing to disclose. **James L Zehnder, MD** Grant/Research Support: GSK [determinants of thrombopoietin agonist response (eltrombopag)]. **Athena M Cherry, PhD** Nothing to disclose. **Richard A Larson, MD** Clinical Trial Support: Amgen [leukemia (Blinatumomab)]; Astellas [leukemia (ASP2215)]; Ariad [leukemia (Ponatinib)]; Erytech [leukemia (ERYASP)]; Ambit Bioscience [leukemia (quizartinib)]. Consultant/Advisory Boards: Ariad [leukemia (Ponatinib)]; Amgen [leukemia (Blinatumomab)]; Astellas [leukemia (ASP2215)]; Bayer [leukemia (Copanlisib)]; BristolMyers [leukemia (Dasatinib)]; Celgene [leukemia (Azacitidine)]; CVS/Caremark [leukemia (multiple drugs for hematologic malignancies)]; Incyte [leukemia]; Novartis [leukemia (Nilotinib)]; Pfizer [leukemia (Gemtuzumab ozogamicin)]. **Jennifer S Tirnauer, MD** Nothing to disclose.

Contributor disclosures are reviewed for conflicts of interest by the editorial group. When found, these are addressed by vetting through a multi-level review process, and through requirements for references to be provided to support the content. Appropriately referenced content is required of all authors and must conform to UpToDate standards of evidence.

Conflict of interest policy

Use of chromosomal microarray in obstetrics

Author

David T Miller, MD, PhD

Section Editor

Louise Wilkins-Haug, MD, PhD

Deputy Editor

Vanessa A Barss, MD, FACOG

All topics are updated as new evidence becomes available and our [peer review process](#) is complete.

Literature review current through: Feb 2015. | **This topic last updated:** Feb 10, 2015.

INTRODUCTION — Prenatal diagnosis and postnatal evaluation of pregnancy loss often involves cytogenetic analysis of amniocytes, chorionic villi, or fetal cells. Although conventional Giemsa(G)-banding of metaphase chromosomes detects aneuploidies and large structural changes, this approach also has limitations: it does not consistently identify microscopic gene defects (<3 to 10 million base pairs per G-band) and requires cell culture, which takes a minimum of seven days to obtain an adequate number of dividing cells. The use of fluorescent in situ hybridization (FISH) reduces the time to obtain a result and is less labor intensive, but can only detect a limited number of prespecified targets.

Chromosomal microarray (CMA) is an array-based molecular cytogenic technique that can overcome these limitations. This technique compares the genomic content (DNA) of a patient (target) with that of a normal control individual (or individuals) and detects not only aneuploidies and large structural changes, but also submicroscopic gains, losses, and unbalanced rearrangements in genes ([figure 1](#)).

Different laboratories perform CMA using different technology platforms and with different array design and content [[1-6](#)].

This topic will discuss the use of CMA in obstetrics. Basic principles of genetics are reviewed separately, and include:

- (See "[Basic principles of genetic disease](#)".)
- (See "[Principles of molecular genetics](#)".)
- (See "[Tools for genetics and genomics: Gene expression profiling](#)".)
- (See "[Tools for genetics and genomics: Cytogenetics and molecular genetics](#)".)
- (See "[Genomic disorders: An overview](#)", section on 'Array comparative genomic hybridization'.)
- (See "[Tools for genetics and genomics: Gene expression profiling](#)", section on 'Microarray analysis and interpretation'.)

BENEFITS AND LIMITATIONS OF CMA — CMA has advantages and disadvantages compared with conventional G-banding.

Benefits

Higher diagnostic yield — Although the resolution of the array is dependent upon the type of array used and the average spacing of the probes on the array, the minimum resolution is 50,000 to 200,000 base pairs compared with 3 to 10 million base pairs for G-banding. The ability of CMA to detect extremely small changes in the genome results in a higher rate of detection of abnormalities compared with conventional G-banding [[7-11](#)].

- A 2013 systematic review of four large studies calculated the utility of prenatal microarrays when the conventional G-band karyotype was normal and found copy number changes of clinical significance were noted in [[12](#)]:
 - 2.4 percent (295/12,362) of cases from all prenatal ascertainment groups reported in four large studies

- 6.5 percent (201/3090) of cases with an abnormal ultrasound examination in these four studies
- 1.0 percent (50/5108) of cases performed because of advanced maternal age
- 1.1 percent (44/4164) of cases performed because of parental anxiety, history of chromosome abnormality, abnormal serum screening result, and “other”
- In a 2014 systematic review that combined data from 18 studies (n = 2220 fetuses), 3.1 to 7.9 percent of fetuses with a structural ultrasound anomaly restricted to one anatomical system and a normal karyotype had a submicroscopic copy number variant that explained its phenotype and provided information for fetal prognosis [13].

The higher diagnostic yield of microarray is particularly noteworthy in analysis of stillborns:

- In a prospective study of 532 stillborn fetuses that compared diagnostic yield with CMA versus traditional karyotype, CMA resulted in a positive finding: 42 percent more often than karyotype in all stillbirths, 35 percent more often in antepartum stillbirths, and 54 percent more often in stillbirths with anomalies [8]. The actual yields of positive results for CMA and karyotype were: all stillbirths 8.3 versus 5.8 percent (P = 0.007), antepartum stillbirths 8.8 versus 6.5 percent (P = 0.02), and stillbirths with congenital anomalies 29.9 versus 19.4 percent, (P = 0.008). Importantly, CMA was successful (from a technical standpoint) more often than karyotype (87.4 versus 70.5 percent, P<0.001), since karyotype requires cell culture, which is often unsuccessful in this setting.

Faster turnaround time — G-banding requires cell culture to obtain sufficient DNA from dividing cells, which results in a turnaround time of one to two weeks. CMA can be performed directly on high quality DNA extracted from isolated cells; results can be available in as little as one day since time for culture is not required. CMA is also more amenable to automation in the laboratory, which provides objective interpretation and decreases labor costs.

Limitations/disadvantages

Inability to detect balanced structural rearrangements — In contrast to G-banding, CMA is unable to detect balanced structural rearrangements because there is no change in copy number. Although this is a disadvantage of the technique, it is not always of clinical significance because:

- The likelihood that a truly balanced rearrangement would interrupt a gene is low. Even those that have breakpoints within a gene are not necessarily pathogenic.
- When the parent and fetus have the same balanced rearrangement, it is generally considered benign if the parent is unaffected.
- Although there is greater concern that a de novo balanced rearrangement may be clinically significant (ie, associated with an abnormal phenotype), de novo balanced rearrangements are rare and represent a small proportion of the abnormalities detected by G-banding.

In one study including 269,371 analyses, the rate of prenatally diagnosed de novo balanced chromosome rearrangements in amniocytes, chorionic villus samples, and fetal blood samples was 0.09, 0.08 and 0.05 percent, respectively [14]. In another study of 377,357 amniocenteses, de novo reciprocal translocations were detected in approximately 0.05 percent and a balanced Robertsonian translocation in 0.01 percent [15].

De novo cytogenetic rearrangements that appear balanced by G-banding may have submicroscopic genomic differences in the number of copies of one or more sections of DNA. These copy number variants may be known to be benign or associated with phenotypic consequences, or their significance may be unknown. Although most apparently balanced reciprocal translocations in phenotypically normal

individuals do not contain genomic imbalance and are benign [16], there is a 6 percent probability that an apparently balanced translocation identified on prenatal G-banding has phenotypic consequences [17]. As a result, the overall ability to predict abnormalities (eg, developmental delay, malformation) is more strongly correlated with findings on array CGH than G-banding [18-20].

Poor ability to detect triploidy — Standard CMA does not recognize triploidy because the relative gene content is balanced. The poor ability of standard CMA to detect most triploidies does not represent a significant limitation as a clinical test because triploidy is rare, the diagnosis is often suspected by prenatal ultrasound examination after the first trimester, and most cases result in fetal demise or neonatal death rather than long-term disability. Although fetal 69,XXX may be missed, fetal 69,XXY will be detected because of the sex chromosome imbalance in the latter case.

Some laboratories incorporate specific algorithms in their processing to better detect triploidy, either through the use of single nucleotide polymorphisms (SNP) arrays [21] or through an initial assessment of a limited number of cells by fluorescence in situ hybridization (FISH).

Variable ability to detect low level mosaicism — Low levels of mosaicism are difficult to detect, but generally result in a milder phenotypic abnormality than the same but non-mosaic genomic imbalance. G-banding performs somewhat better than most clinical CMA platforms for detection of low level mosaicism: a typical 20-cell G-banded karyotype detects mosaicism at approximately 14 percent of cells or more, while most clinical CMA platforms typically detect mosaicism at 20 to 30 percent of cells or more. Laboratories that use SNP array platforms for CMA can detect lower levels of mosaicism (~10 percent). If there is suspicion of lower level mosaicism on G-banding or postnatally, a cytogenetics laboratory can be asked to count additional cells to improve the detection of mosaicism below 10 percent of cells.

Detection of variants of uncertain significance — Copy number variants of uncertain significance (VUS) are copy number changes that have not been reported and thus have an unknown phenotype. These small changes are not detectable by G-banding, but are identified by CMA in 1 to 2 percent of cases [7,8,21,22]. The additional information provided by CMA can be challenging to interpret, particularly if a phenotypically normal parent carries the same change. In the absence of clear prognostic information, parents may find it difficult to make a decision about continuing the pregnancy. (See ["Overview of genetic variation", section on 'Copy number variations \(CNVs\)'](#).)

In a large prospective study performed from 2008 to mid 2011 [7], CMA detected VUS in 3.4 percent (130 of 3822) of all cases that were normal by karyotype. Although this degree of uncertainty is concerning, the authors pointed out that collective knowledge in this field continues to improve, thereby decreasing the frequency of VUS. If the VUS identified in this study were reanalyzed in 2012, the overall VUS rate would be approximately 1.5 percent (56 of 3822).

Detection of VUS has raised the question of whether it is ethically justifiable to withhold these test results from the patient [23]. VUS should become less problematic in the future as the International Standards for Cytogenomic Arrays (ISCA) is creating a database of array results and associated phenotypes for the National Institutes of Health (NIH).

Expense — CMA is relatively expensive, like many genetic tests, although costs are declining.

Variable clinical sensitivity — Clinical sensitivity refers to the proportion of patients who have the clinical symptoms of a disorder and are detectable by the testing method. Some genetic conditions are primarily caused by a change in genomic copy number (eg, a microdeletion or microduplication syndrome). An example of a relatively common genomic disorder is velocardiofacial syndrome, also known as 22q11.2 deletion syndrome. Among individuals with signs/symptoms of this condition, a microdeletion will be detected by chromosomal microarray (CMA) in over 90 percent. When the microdeletion is not detected, the patient usually does not have all of the symptoms and may have

another diagnosis that is similar. In contrast, neurofibromatosis type-1 (NF1) can be caused by a genomic deletion that includes the NF1 gene, but less than 5 percent of NF1 patients have a deletion; most of these patients have a sequence change in the gene that is undetectable by CMA. If a patient has a genetic condition in which a subset of cases are caused by sequence changes, then other testing should be considered either in place of, or in addition to, CMA.

PROCEDURES

Specimen — Prenatal CMA is performed on DNA from uncultured amniocytes (amniotic fluid), chorionic villus cells, or cord blood [24]. It is not performed on cell free DNA in maternal blood during noninvasive prenatal testing because currently there is no amplification step so the signal from the fetus would be “drowned out” by the maternal DNA.

Probe — A typical microarray contains thousands of different probes representing targeted areas of the genome. High resolution oligonucleotide probes are now standard for clinical laboratories, as compared to earlier versions using bacterial artificial chromosome (BAC) clones. The oligonucleotide probes are smaller and can better define the breakpoints compared to BAC probes, but their ability to detect smaller copy number variations (CNVs) makes them more likely to detect variants of uncertain clinical significance.

Targeted versus whole genome — Prenatal CMA may assess the whole genome or may be targeted to evaluate regions of the genome known or highly suspected to cause specific phenotypes when deleted or duplicated.

Compared to whole genome CMA, targeted CMA lowers the chance of identifying variants of uncertain significance since the targets are associated with known phenotypes. However, targeted CMA also potentially results in a lower diagnostic yield as unique chromosomal imbalances that could be clinically significant may be missed.

Whole genome CMA leads to challenging counseling issues because copy number variants of unknown significance are detected. In addition, copy number variants related to late-onset diseases may be identified in genomic regions other than those specifically targeted, thus the additional information provided by whole genome CMA can have implications for both the parent and offspring. For example, a prenatal diagnostic procedure to detect Down syndrome may exclude Down syndrome, but identify an inherited BRCA 1 or 2 mutation or the gene for Huntington disease. Thus, although CMA may be used as a diagnostic test to identify a specific disorder in a patient at risk for that disorder, it can also act as a screening test for a broad range of other genetic abnormalities, blurring the distinction between prenatal diagnosis and prenatal screening.

CLINICAL INDICATIONS FOR CMA IN OBSTETRICS — Clinical guidelines for use of CMA are evolving as more data on the benefits and limitations of this technique are published [25-28].

The American College of Obstetrics and Gynecology (ACOG) and Society for Maternal-Fetal Medicine (SMFM) recommend use of CMA when genetic analysis is desired because of [28]:

- Fetal congenital anomalies
- Fetal demise

They concluded that either G-banding or CMA is reasonable for patients undergoing invasive prenatal diagnosis who have a structurally normal, viable fetus, and made no practice recommendation regarding cytogenetic analysis of first and early second trimester pregnancy losses. Some clinicians have advocated microarray as a first-line test whenever fetal chromosomal analysis is planned [29].

In the postnatal setting, American College of Medical Genetics (ACMG) guidelines recommend CMA as the first-tier test for patients with [30]:

- Unexplained developmental delay/intellectual disability (see ["Intellectual disability \(mental retardation\) in children: Evaluation for a cause"](#))
- Autism spectrum disorders (see ["Autism spectrum disorder: Diagnosis"](#))
- Multiple congenital anomalies not explained by a specific syndrome (see ["Approach to congenital malformations"](#))

Prenatal ultrasound showing fetal structural abnormalities — Fetal structural abnormalities detectable by ultrasound should raise suspicion of a chromosomal abnormality in the fetus. In a study that karyotyped 2086 fetuses after ultrasonographic examination revealed fetal malformations, growth restriction, or both, chromosomal abnormalities by G-banding were detected in 301 cases (14 percent), and were more common among fetuses with multisystem malformations (29 percent) than among those with isolated defects (2 percent) [31].

Among fetuses presenting with an ultrasound structural abnormality and a normal G-banded karyotype, a 2013 meta-analysis reported that CMA was abnormal in 7.2 percent (95% CI 5.4-9.8; 7 large cohort studies published 2011 to 2013) when the referral indication was structural malformation on ultrasound [32]. This number is higher than the 1 to 3 percent previously thought to have a genetic imbalance (eg, microdeletion, microduplication) on CMA [33-35], but the lower prevalence was based on earlier studies. The increased yield reported by the meta-analysis included data from recent studies and was likely due to increased coverage of CMA probes since the detection of abnormalities on the array increases as the density of the array increases, as well as when there are multiple anomalies present. Postnatal studies of neonates with structural anatomical abnormalities confirm these findings [36,37].

In fetuses with congenital heart disease, a systematic review concluded performing CMA had value even if conventional G-banding and 22q11 microdeletion analysis are normal [38]. In cases of isolated congenital heart disease with normal karyotype and 22q11 microdeletion analysis by FISH, the yield of additional pathogenic copy number variants detected by CMA was approximately 3 percent. The incremental yield was 7 percent in fetuses with both cardiac and noncardiac anomalies.

Fetal demise — CMA is useful in evaluation of stillbirth (pregnancy loss ≥ 20 weeks of gestation) because both chromosomal abnormalities and culture failure are common in these cases. Chromosomal abnormalities identifiable by G-banded karyotype are identified in about 5 percent of stillborn fetuses in the absence of an anatomical malformation and 35 to 40 percent of stillborns when structural abnormalities are present [17,39]. Culture failure is common when the fetus has died, and thus prevents the accurate diagnosis of a karyotypic abnormality in these cases, CMA can overcome this difficulty.

As discussed above, in a study of samples from 532 stillbirths, SNP oligonucleotide microarray analysis was more likely to yield results than karyotype analysis (87.4 versus 70.5 percent, $P < 0.001$), provided better detection of genetic abnormalities (aneuploidy or pathogenic copy number variants, 8.3 versus 5.8 percent; $P = 0.007$), and identified more genetic abnormalities among the 443 antepartum stillbirths (8.8 versus 6.5 percent, $P = 0.02$) and the 67 stillbirths with congenital anomalies (29.9 versus 19.4 percent, $P = 0.008$) [8,40]. A copy-number VUS was reported in 2.6 percent (12 of 465) of the samples.

CMA may also have some advantages over G-banded karyotyping for diagnostic evaluation of miscarriages (pregnancy loss before 20 weeks). In a 2013 systematic review, CMA and karyotyping results agreed in 86.0 percent of cases (95% CI 77.0-96.0), CMA detected 13 percent (95% CI 8.0-21.0) additional chromosome abnormalities over conventional full karyotyping but conventional full karyotyping detected 3 percent (95% CI 1.0-10.0) additional abnormalities over CMA [41]. The incidence of a variant of unknown significance was 2 percent (95% CI 1.0-10.0). However, the clinical value of CMA over G-banding in these cases is less clear than in the evaluation of stillbirth because CMA may miss polyploidy, which is the cause of 10 to 20 percent of miscarriages, and there is a lack of information about potential relationships between small deletions, duplications, and other variants detected by CMA.

and risk of miscarriage. A previously unreported CNV is found in 30 to 50 percent of miscarriages, requiring further parental evaluation; only 6 percent of miscarriages reveal structural chromosomal abnormalities requiring parental follow-up [42].

Marker chromosome — A marker chromosome refers to small chromosome fragments that may be observed with a G-banded karyotype. Because of the small size of the marker, the chromosome of origin cannot always be determined. CMA is often able to identify the chromosomal origin of a marker chromosome and thus allow prognostic counseling based on its origin and gene content. (See ["Congenital cytogenetic abnormalities", section on 'Supernumerary marker chromosome'.](#))

POSSIBLE INDICATIONS FOR PRENATAL CMA — Prenatal CMA may be useful after a normal G-band karyotype or instead of conventional karyotyping in the following situations, but data are limited and the diagnostic yield of CMA in these settings is not known:

- Intrauterine growth restriction (IUGR)
- Increased nuchal lucency (See ["First trimester cystic hygroma and increased nuchal translucency", section on 'Obstetrical management'.](#))
- Abnormal maternal serum screening showing an increased risk for Down syndrome (See ["Down syndrome: Prenatal screening overview", section on 'Screen-positive test'.](#))
- One or more soft markers for aneuploidy on ultrasound (see ["Sonographic findings associated with fetal aneuploidy"](#))
- Oligohydramnios or polyhydramnios

Noninvasive prenatal screening (NIPS) by next-generation sequencing of cell-free DNA in maternal plasma is used for primary or secondary screening for common aneuploidies (trisomies 21, 18, 13, and sex chromosome aneuploidies). Commercial companies have begun to offer expanded panels including screening for common microdeletion syndromes such as 22q11.2 deletion (DiGeorge syndrome) without reporting the genomic coordinates or whether the deletion is maternal or fetal [43]. If performed, diagnostic prenatal or postnatal testing with chromosomal microarray and appropriate parental studies to determine precise genomic coordinates and inheritance should follow a positive microdeletion NIPS result.

PATIENT COUNSELING — Couples considering CMA testing should receive pre-test and post-test genetic counseling that is nondirective and clearly explains that testing is voluntary. Nondirective information enables the parents to balance the risks, benefits, and limitations of this approach to prenatal diagnosis. Genetic counseling should involve a provider with specific expertise in the area of prenatal genetic testing, usually a genetic counselor, medical geneticist, or provider with relevant experience, given the range of potential results from CMA and the variable clinical effects of particular genomic imbalance disorders.

Pre-test counseling should provide the following information [44]:

- Risks associated with prenatal diagnosis (eg, procedure-related fetal loss)
- Potential psychological implications of prenatal diagnosis (eg, uncertainty, anxiety, need for consideration of pregnancy termination or other intervention). In particular, the patient should understand that the severity of disease cannot be predicted by prenatal genetic studies, genetic studies may detect the gene for a disease that never manifests clinically, and genes for adult onset diseases in themselves may be identified. This can be particularly problematic for copy number changes associated with risk of neurocognitive or neuropsychiatric symptoms where the degree of symptom severity can be quite variable, even within the same family. Depending on the type of microarray, additional concerns regarding the identification of consanguinity and non-paternity

exist.

- Implications of having a child with a genomic disorder
- Advantages and disadvantages of prenatal CMA
 - Detection rate for chromosomal abnormalities is higher than with G-banding
 - Possibility and clinical implications of an undetected balanced rearrangement
 - Possibility and clinical implications of detecting a variant of unknown significance
 - Possibility of detecting a carrier state for genetic disease within the parent
 - Possibility of identifying non-paternity/consanguinity
- Information about the length of time necessary to obtain CMA results
 - Results may be available one week after biopsy with direct cell preparation (eg, chorionic villus biopsy, biopsy of fetal tissue, amniocytes)
 - If amniocytes are cultured initially for G-banding and then CMA is performed, results may be available three weeks after amniocentesis

Post-test counseling is best provided through consultation with a genetic counselor or a medical geneticist, particularly after a diagnosis of a genomic imbalance disorder, to provide the following information:

- Normal/negative results do not rule out the presence of a genetic disorder, because many disorders are not caused by microdeletions or microduplications. Other genetic testing may be warranted based on the clinical presentation.
- Abnormal/positive results allow a better understanding of the etiology of clinical features present in the fetus, but not necessarily an ability to predict the medical or developmental outcome of a child born with that genomic disorder. The couple should be told about:
 - The spectrum of medical and intellectual issues associated with the identified genomic disorder
 - The difficulty in predicting phenotype due to variable penetrance and expressivity
 - How mutations arise (de novo versus inherited)
 - Recurrence risks for future pregnancies
 - Information about pregnancy termination. (See ["Surgical termination of pregnancy: First trimester"](#) and ["Second trimester pregnancy termination: Overview and surgical termination"](#).)

SUMMARY AND RECOMMENDATIONS

- Advantages of chromosomal microarray (CMA) over conventional G-banding in the prenatal setting include higher diagnostic yield and faster turnaround time. (See ["Higher diagnostic yield"](#) above.)
- Certain types of chromosomal changes are not detectable or are poorly detectable by CMA, such as balanced translocations, triploidy, and low level mosaicism; however, these represent a smaller number of cases as compared to those with genomic imbalance. (See ["Limitations/disadvantages"](#) above.)
- CMA detects variants of unknown significance in about 1 to 2 percent of cases. Counseling patients with these findings can be difficult. (See ["Detection of variants of uncertain significance"](#) above.)

above.)

- In the obstetrical setting, CMA is a reasonable option for further evaluation of fetuses with structural abnormalities, after fetal demise (particularly when chromosomal analysis is desired but G-banding is not possible due to failure of cell culture), and when a marker chromosome is identified. It is equivalent to G-banding when the primary goal is to detect aneuploidy, as in Down syndrome screening. Some clinicians have advocated microarray as a first-line test whenever fetal chromosomal analysis is planned. (See '[Clinical indications for cma in obstetrics](#)' above.)
- Genetic counseling by a qualified provider should always be offered before and after prenatal CMA. Pre-test counseling should include a discussion of the medical and psychological risks and the advantages and disadvantages of this approach compared with conventional G-banding. The genetic principles of variants of uncertain significance and variable expressivity and penetrance should be explained. Post-test counseling should include interpretation of the findings and an explanation of possible follow-up studies. (See '[Patient counseling](#)' above.)

Use of UpToDate is subject to the [Subscription and License Agreement](#).

REFERENCES

1. Rauch A, Rüschendorf F, Huang J, et al. Molecular karyotyping using an SNP array for genomewide genotyping. *J Med Genet* 2004; 41:916.
2. de Vries BB, Pfundt R, Leisink M, et al. Diagnostic genome profiling in mental retardation. *Am J Hum Genet* 2005; 77:606.
3. Hochstenbach R, Ploos van Amstel HK, Poot M. Microarray-based genome investigation: molecular karyotyping or segmental aneuploidy profiling? *Eur J Hum Genet* 2006; 14:262.
4. Vermeesch JR, Rauch A. Reply to Hochstenbach et al. 'Molecular karyotyping'. *Eur J Hum Genet* 2006; 14:1063.
5. Vermeesch JR, Fiegler H, de Leeuw N, et al. Guidelines for molecular karyotyping in constitutional genetic diagnosis. *Eur J Hum Genet* 2007; 15:1105.
6. Hoyer J, Dreweke A, Becker C, et al. Molecular karyotyping in patients with mental retardation using 100K single-nucleotide polymorphism arrays. *J Med Genet* 2007; 44:629.
7. Wapner RJ, Martin CL, Levy B, et al. Chromosomal microarray versus karyotyping for prenatal diagnosis. *N Engl J Med* 2012; 367:2175.
8. Reddy UM, Page GP, Saade GR, et al. Karyotype versus microarray testing for genetic abnormalities after stillbirth. *N Engl J Med* 2012; 367:2185.
9. Shaffer LG, Dabell MP, Fisher AJ, et al. Experience with microarray-based comparative genomic hybridization for prenatal diagnosis in over 5000 pregnancies. *Prenat Diagn* 2012; 32:976.
10. Shaffer LG, Rosenfeld JA, Dabell MP, et al. Detection rates of clinically significant genomic alterations by microarray analysis for specific anomalies detected by ultrasound. *Prenat Diagn* 2012; 32:986.
11. Scott F, Murphy K, Carey L, et al. Prenatal diagnosis using combined quantitative fluorescent polymerase chain reaction and array comparative genomic hybridization analysis as a first-line test: results from over 1000 consecutive cases. *Ultrasound Obstet Gynecol* 2013; 41:500.
12. Callaway JL, Shaffer LG, Chitty LS, et al. The clinical utility of microarray technologies applied to prenatal cytogenetics in the presence of a normal conventional karyotype: a review of the literature. *Prenat Diagn* 2013; 33:1119.
13. de Wit MC, Srebniak MI, Govaerts LC, et al. Additional value of prenatal genomic array testing in

fetuses with isolated structural ultrasound abnormalities and a normal karyotype: a systematic review of the literature. *Ultrasound Obstet Gynecol* 2014; 43:139.

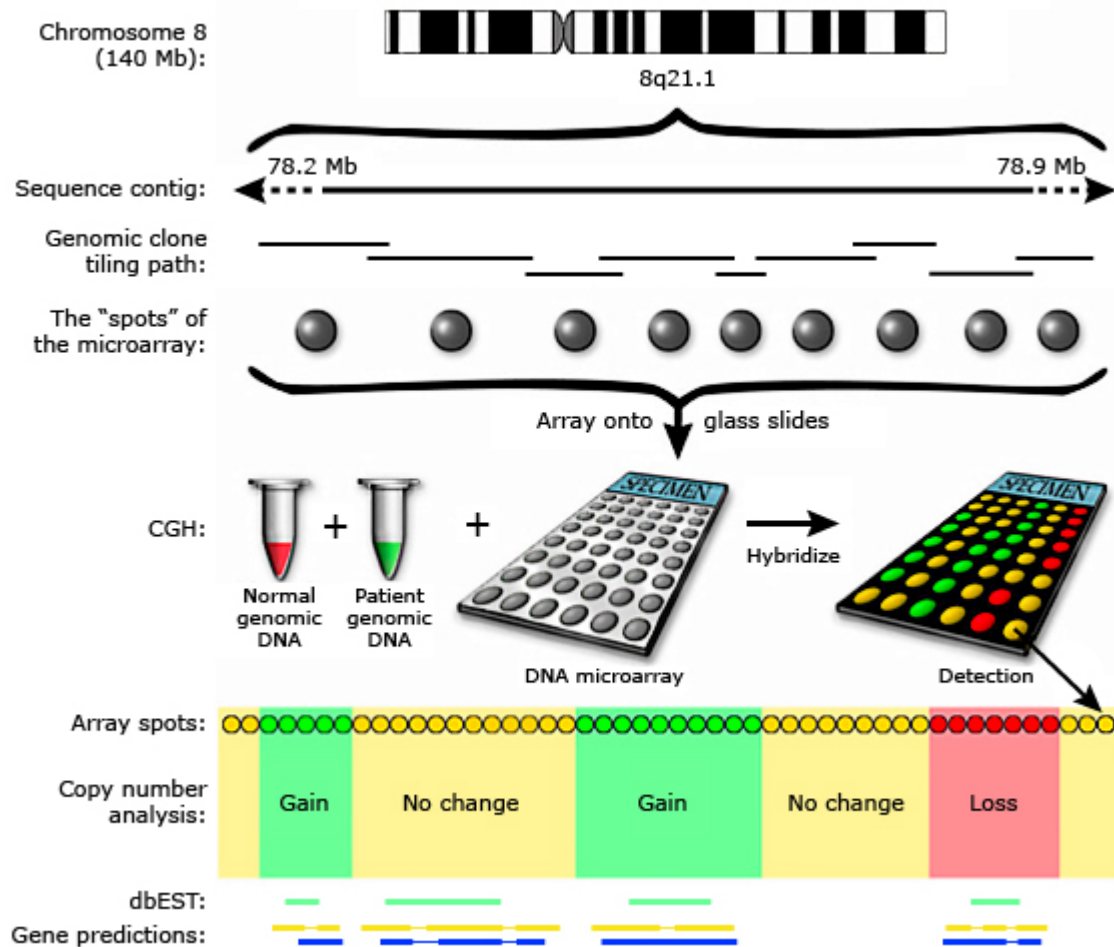
14. Giardino D, Corti C, Ballarati L, et al. De novo balanced chromosome rearrangements in prenatal diagnosis. *Prenat Diagn* 2009; 29:257.
15. Warburton D. De novo balanced chromosome rearrangements and extra marker chromosomes identified at prenatal diagnosis: clinical significance and distribution of breakpoints. *Am J Hum Genet* 1991; 49:995.
16. Baptista J, Prigmore E, Gribble SM, et al. Molecular cytogenetic analyses of breakpoints in apparently balanced reciprocal translocations carried by phenotypically normal individuals. *Eur J Hum Genet* 2005; 13:1205.
17. Savage MS, Mourad MJ, Wapner RJ. Evolving applications of microarray analysis in prenatal diagnosis. *Curr Opin Obstet Gynecol* 2011; 23:103.
18. Gribble SM, Prigmore E, Burford DC, et al. The complex nature of constitutional de novo apparently balanced translocations in patients presenting with abnormal phenotypes. *J Med Genet* 2005; 42:8.
19. De Gregori M, Ciccone R, Magini P, et al. Cryptic deletions are a common finding in "balanced" reciprocal and complex chromosome rearrangements: a study of 59 patients. *J Med Genet* 2007; 44:750.
20. Feenstra I, Hanemaaijer N, Sikkema-Raddatz B, et al. Balanced into array: genome-wide array analysis in 54 patients with an apparently balanced de novo chromosome rearrangement and a meta-analysis. *Eur J Hum Genet* 2011; 19:1152.
21. Breman A, Pursley AN, Hixson P, et al. Prenatal chromosomal microarray analysis in a diagnostic laboratory; experience with >1000 cases and review of the literature. *Prenat Diagn* 2012; 32:351.
22. Hillman SC, Pretlove S, Coomarasamy A, et al. Additional information from array comparative genomic hybridization technology over conventional karyotyping in prenatal diagnosis: a systematic review and meta-analysis. *Ultrasound Obstet Gynecol* 2011; 37:6.
23. McGillivray G, Rosenfeld JA, McKinlay Gardner RJ, Gillam LH. Genetic counselling and ethical issues with chromosome microarray analysis in prenatal testing. *Prenat Diagn* 2012; 32:389.
24. Brady PD, Vermeesch JR. Genomic microarrays: a technology overview. *Prenat Diagn* 2012; 32:336.
25. Lee CN, Lin SY, Lin CH, et al. Clinical utility of array comparative genomic hybridisation for prenatal diagnosis: a cohort study of 3171 pregnancies. *BJOG* 2012; 119:614.
26. Duncan A, Langlois S, SOGC Genetics Committee, CCMG Prenatal Diagnosis Committee. Use of array genomic hybridization technology in prenatal diagnosis in Canada. *J Obstet Gynaecol Can* 2011; 33:1256.
27. Novelli A, Grati FR, Ballarati L, et al. Microarray application in prenatal diagnosis: a position statement from the cytogenetics working group of the Italian Society of Human Genetics (SIGU), November 2011. *Ultrasound Obstet Gynecol* 2012; 39:384.
28. American College of Obstetricians and Gynecologists Committee on Genetics. Committee Opinion No. 581: the use of chromosomal microarray analysis in prenatal diagnosis. *Obstet Gynecol* 2013; 122:1374.
29. Fiorentino F, Napoletano S, Caiazzo F, et al. Chromosomal microarray analysis as a first-line test in pregnancies with a priori low risk for the detection of submicroscopic chromosomal abnormalities. *Eur J Hum Genet* 2013; 21:725.
30. Manning M, Hudgins L, Professional Practice and Guidelines Committee. Array-based technology and recommendations for utilization in medical genetics practice for detection of chromosomal abnormalities. *Genet Med* 2010; 12:742.

31. Nicolaides KH, Snijders RJ, Gosden CM, et al. Ultrasonographically detectable markers of fetal chromosomal abnormalities. *Lancet* 1992; 340:704.
32. Hillman SC, McMullan DJ, Hall G, et al. Use of prenatal chromosomal microarray: prospective cohort study and systematic review and meta-analysis. *Ultrasound Obstet Gynecol* 2013; 41:610.
33. Sahoo T, Cheung SW, Ward P, et al. Prenatal diagnosis of chromosomal abnormalities using array-based comparative genomic hybridization. *Genet Med* 2006; 8:719.
34. Shaffer LG, Coppinger J, Alliman S, et al. Comparison of microarray-based detection rates for cytogenetic abnormalities in prenatal and neonatal specimens. *Prenat Diagn* 2008; 28:789.
35. Kleeman L, Bianchi DW, Shaffer LG, et al. Use of array comparative genomic hybridization for prenatal diagnosis of fetuses with sonographic anomalies and normal metaphase karyotype. *Prenat Diagn* 2009; 29:1213.
36. Lu XY, Phung MT, Shaw CA, et al. Genomic imbalances in neonates with birth defects: high detection rates by using chromosomal microarray analysis. *Pediatrics* 2008; 122:1310.
37. Goemaere N, Douben H, Van Opstal D, et al. The use of comparative genomic hybridization and fluorescent in situ hybridization in postmortem pathology investigation of congenital malformations. *Pediatr Dev Pathol* 2010; 13:85.
38. Jansen FA, Blumenfeld YJ, Fisher A, et al. Array comparative genomic hybridization and fetal congenital heart defects: a systematic review and meta-analysis. *Ultrasound Obstet Gynecol* 2015; 45:27.
39. Warburton D, Susser M, Stein Z, Kline J. Genetic and epidemiologic investigation of spontaneous abortion: relevance to clinical practice. *Birth Defects Orig Artic Ser* 1979; 15:127.
40. Reddy UM, Page GP, Saade GR. The role of DNA microarrays in the evaluation of fetal death. *Prenat Diagn* 2012; 32:371.
41. Dhillon RK, Hillman SC, Morris RK, et al. Additional information from chromosomal microarray analysis (CMA) over conventional karyotyping when diagnosing chromosomal abnormalities in miscarriage: a systematic review and meta-analysis. *BJOG* 2014; 121:11.
42. Rajcan-Separovic E. Chromosome microarrays in human reproduction. *Hum Reprod Update* 2012; 18:555.
43. Yatsenko SA, Peters DG, Saller DN, et al. Maternal cell-free DNA-based screening for fetal microdeletion and the importance of careful diagnostic follow-up. *Genet Med* 2015.
44. Wapner RJ, Driscoll DA, Simpson JL. Integration of microarray technology into prenatal diagnosis: counselling issues generated during the NICHD clinical trial. *Prenat Diagn* 2012; 32:396.

Topic 16683 Version 15.0

GRAPHICS

Schematic CGH for microarray setup



Schematic representation of the array CGH technique for a focused analysis of copy number imbalances along a region of interest (eg, 8q21.1). A tiling path of genomic clones (eg, BACs, PACs, PIs, cosmids) is generated to cover the region. After extraction and purification, these genomic DNA targets are arrayed onto glass slides.

Array CGH is performed by hybridizing labeled normal (Cy3) and tumor (Cy5) genomic DNA into the microarray and detected using a microarray scanner.

Each array spot, realigned *in silico* as a single contiguous map to correspond with the tiling path, can be analyzed by fluorescence ratio to identify the regions of copy number changes. These results may be correlated with *in silico* techniques to identify candidate genes of interest.

CGH: comparative genomic hybridization; BAC: bacterial artificial chromosome; PAC: P1 bacteriophage artificial chromosome; P1: P1 bacteriophage; Cy3: cyanine dye with green fluorescence; Cy5: cyanine dye with red fluorescence.

Reproduced from: Beheshti B, Park PC, Braude I, Squire JA. Micorarray CGH. In: *Molecular Cytogenetics: Protocols and Applications*. Humana Press, 2002, with kind permission from Springer Science + Business Media B.V. Copyright © 2002.

Graphic 58888 Version 10.0

Disclosures

Disclosures: **David T Miller, MD, PhD** Consultant/Advisory Boards: Claritas Genomics (non-equity professional service agreement) [Chromosomal Microarray Testing (Pediatric post-natal patients only) (provider of diagnostic services)]. **Louise Wilkins-Haug, MD, PhD** Nothing to disclose. **Vanessa A Barss, MD, FACOG** Nothing to disclose.

Contributor disclosures are reviewed for conflicts of interest by the editorial group. When found, these are addressed by vetting through a multi-level review process, and through requirements for references to be provided to support the content. Appropriately referenced content is required of all authors and must conform to UpToDate standards of evidence.

Conflict of interest policy

Tools for genetics and genomics: Polymerase chain reaction

Author

Benjamin A Raby, MD, MPH

Section Editor

Anne Slavotinek, MBBS, PhD

Deputy Editor

Jennifer S Tirnauer, MD

All topics are updated as new evidence becomes available and our [peer review process](#) is complete.

Literature review current through: Feb 2015. | **This topic last updated:** Aug 07, 2014.

INTRODUCTION — The polymerase chain reaction (PCR) is the basis of many modern molecular biology and molecular genetics techniques. In just a few hours, PCR can amplify a single DNA molecule a million-fold [1]. The greatly amplified target DNA is subsequently analyzed via other techniques.

Since its first publication in 1985 [2], the impact of PCR on biomedical research has been immense. This technology allows large quantities of rare sequences to be synthesized, cloned, and analyzed with high reliability and minimum effort. The award of the 1993 Nobel Prize in Chemistry to Kary B. Mullis for inventing the technique recognized the importance of PCR-based methods [3].

PCR allows for rapid and highly-specific amplification of DNA fragments. The method is relatively inexpensive and commonly performed in most molecular laboratories. Pieces of DNA from about 50 base pairs (bp) to over 10 kilobases (kb) can be amplified, even from vanishingly small amounts of starting genomic DNA. The two most important principles underlying PCR are:

- Complementarity-driven binding of DNA to form a duplex
- Template-driven, semi-conservative synthesis of DNA, by DNA polymerases

These two principles are discussed separately in detail. (See "[Principles of molecular genetics](#)".)

This topic presents a brief overview of PCR as well as a discussion of some applications of PCR and quantitative PCR. Other molecular genetic tools are discussed separately.

- (See "[Tools for genetics and genomics: Cytogenetics and molecular genetics](#)".)
- (See "[Tools for genetics and genomics: Gene expression profiling](#)".)
- (See "[Tools for genetics and genomics: Specially bred and genetically engineered mice](#)".)
- (See "[Tools for genetics and genomics: Model systems](#)".)

THE PCR PROCESS — Two technical components underlie the polymerase chain reaction (PCR) process: the existence of thermostable DNA polymerases purified or cloned from microorganisms living in hot springs and the ability to readily synthesize specific oligonucleotide primers of 20 to 30 residues.

PCR is a single tube reaction. Each reaction is customized to amplify a specific genomic sequence. This specificity is dictated by the pair of oligonucleotide primers (short DNA sequences) designed to hybridize to either strand of a target sequence.

The standard PCR reaction mixture consists of a large excess of oligonucleotide primer pairs, a template DNA (typically genomic DNA), free deoxynucleotide triphosphates (DNA bases), reaction buffer, and thermostable DNA polymerase (the enzyme that drives the PCR reaction).

Thermostable DNA polymerases can withstand heating to 95°C with minimal loss of activity and function optimally near 70°C [1]. The standard enzyme currently used in most PCR is derived from the bacterium *Thermus* (also *Thermophilus*) *aquaticus* (Taq). The error rate (per nucleotide per cycle) for standard Taq is about 1 in 10⁵ nucleotides. High-fidelity polymerases with in vitro proofreading ability, such as those derived from the archaeobacterium *Pyrococcus furiosus* (Pfu), have error rates that are approximately 10

times lower than Taq and are used for applications that mandate more stringent DNA synthesis [4].

The PCR consists of three stages that are repeated 30 to 40 times ([figure 1](#)):

- **Denaturing** — The mixture is heated to 95°C to allow the double-stranded template DNA to denature into single strands.
- **Annealing** — The mixture is cooled to a temperature just below the predicted primer pair melting temperatures, resulting in primer binding to the single-stranded DNA templates, followed by binding of DNA polymerase to the 3' end of the primers.
- **Elongation** — The temperature is then raised to a polymerase activation temperature (~70 to 72°C) to initiate chain elongation. The polymerase catalyzes the addition of free nucleotides to the 3' ends of the primers. Bases are added to complement the template sequence (A with T, C with G). Elongation continues for about one minute and is terminated by cycling back to step one of the reaction: heating the temperature to 95°C, resulting in strand separation.

The number of potential targets for primer annealing has been doubled after one cycle of PCR, as both the original substrate DNA and the newly-synthesized strands are available for primer binding. With each subsequent cycle, the number of amplified targets (defined at both ends by the reaction primers) is doubled, resulting in exponential amplification of the target sequence. It is possible to start with as little as one template DNA molecule and generate tens of millions of copies. If the reaction starts with two copies, then N cycles will theoretically yield 2^N copies.

To amplify the sequence of interest, single-stranded oligonucleotide primers (which are typically 20 nucleotides in length) must be carefully designed and synthesized. Highly-specific amplification can be achieved if the sequence combination of the primer pairs is present only once in the genome. Computer programs that facilitate primer design and calculate annealing temperatures are freely distributed online [5].

After the reaction is completed, the amplified product (called an amplicon) can be visualized by gel electrophoresis, during which the negatively charged DNA fragments migrate toward the positive electrode through an agarose gel. The pores of the gel depend upon its agarose percentage and can be compared to a (molecular) sieve. Thus, migration speed of the DNA fragments through the gel depends on both DNA fragment size and the applied voltage. After electrophoresis and addition of an ethidium bromide stain, which forms a complex with nucleic acids, exposure of the gel to ultraviolet light demonstrates the amplified fragment as a single fluorescing band. The gel can be photographed for a permanent record of the amplification process.

The advantages of the PCR technique are the following:

- **Rapidity** — The amplification reaction takes approximately three hours. Additional time (several hours) is required to first extract DNA from cells and then perform gel electrophoresis after completion of PCR.
- **Extreme sensitivity** — PCR can amplify the DNA from a single cell and can be used in preimplantation genetic diagnosis in embryos.
- **Robust** — Degraded DNA can frequently be successfully amplified.
- **Specificity** — The technique also permits the detection of small nucleotide mutations (see below), as well as trinucleotide repeat expansions.
- **Inexpensive and simple**

There are a number of disadvantages to the PCR technique:

- Because of its extreme sensitivity, contamination of the sample with small amounts of extraneous DNA may result in a spurious or false-positive finding. To help decrease the incidence of this problem, diagnostic PCR laboratories are often subdivided into separate designated pre- and post-amplification areas in different rooms.
- To design appropriate PCR primers, the nucleotide sequence information of the region of interest must be known.
- The ideal length of a sequence to be amplified (PCR amplicon) is several hundred nucleotides. Although longer length PCR methods are available, the analysis of very large regions is more difficult and requires different techniques. Similarly, large mutations, such as duplications or deletions that encompass the amplified region, cannot be detected because only the unaffected allele will be amplified.

UNIQUE GENOMIC SEQUENCES — Although the human genome sequence is very long (3 billion bases), relatively short DNA sequences suffice to specify a unique sequence in the genome. The expected count of a specified sequence of length N bases can be expressed by the following formula, based on probability theory:

$$(3 \times 10^9) \times 1/4^N$$

where $1/4^N$ is the probability of a given sequence of length N bases and the constant 3×10^9 is the number of bases in the human genome.

Thus, a PCR primer 20 nucleotides in length would be estimated to occur by chance at a frequency of 0.003 per genome, and a primer 25 nucleotides in length would occur at a frequency of 0.000003 per genome. Thus, these estimates demonstrate that a specific 20-residue sequence is likely to be unique in the human genome. Furthermore, PCR requires two primers to colocalize with a specific orientation to one another in a relatively small genomic region. Therefore, the probability of amplifying more than one region is exceedingly small.

Sequence alignment algorithms can be useful to align designed primer pairs to the known genome sequence, to verify if a primer pair is unique in the genome, to check if any polymorphisms underlie the primers, and to calculate the exact length of the expected amplicon [6].

PCR COMPARED TO OTHER AMPLIFICATION METHODS — The exponential amplification of polymerase chain reactions (PCR) can be contrasted with other types of nucleic acid amplification that are widely used in molecular biology. Two examples of other amplification methods are whole genome amplification and RNA amplification.

Whole genome amplification — Whole genome amplification (WGA) non-specifically amplifies the entire genome and is used when sample quantities are too small to meet the desired application's requirements. Random, very short primers are used, along with Phi29 polymerase, a highly-active polymerase capable of incorporating over 70,000 nucleotides per priming event [7]. In order to limit amplification bias (uneven amplification of different genomic sequences or alleles), it is best to amplify the template DNA only a few 1000-fold.

WGA is distinct from standard PCR in that it is non-specific (all genomic sequences are targeted, not only one site), and it generates fragments that are tens of thousands of bases in length, compared to 250 to 1000 bases that are typical for PCR.

RNA amplification — Linear amplification of RNA is the precursor to most gene expression microarray experiments. Linear amplification is preferred over exponential amplification (ie, classical PCR) in this setting because occasional imbalances in sequence amplification rates (that result from preferential reaction kinetics and favor amplification of some regions over others) can bias subsequent quantification

[8]. The general method for global amplification of mRNA is based upon the Eberwine protocol [9], whereby single-stranded RNA is the template, the reaction is primed by a poly-T oligonucleotide primer that binds the 3' poly-A tail common to most mRNA sequences, and the reaction is catalyzed by a reverse transcriptase which generates complementary DNA strands. Following second strand synthesis with DNA polymerase, and subsequent sample purification, T7 polymerase is used to drive in vitro transcription (IVT), producing multiple copies of amplified RNA (aRNA) from each dsDNA. By incorporating labeled nucleotides in the in vitro transcription step, amplification products are also prepared to be used as probes in expression analysis. Amplification achieved is typically in the range of 10- to 100-fold, but can vary depending on the amount of input RNA.

CLINICAL APPLICATIONS OF PCR — Though the applications of polymerase chain reactions (PCR) are innumerable, its clinical utility is greatest in the context of genotyping and sequencing for diagnostic and predictive testing.

Genotyping and sequencing — Genotyping is the process of characterizing an individual's genotype (the combination of alleles) at a particular genomic locus (location). Most types of genetic variation can be characterized using PCR techniques, including microsatellite and single nucleotide polymorphism (SNP) markers, insertion/deletion variants (indels), and some structural variants, such as copy-number variants. (See "[Overview of genetic variation](#)".)

SNP genotyping — PCR is the basis of the most commonly used single-nucleotide polymorphism (SNP) genotyping methods, including probe-hybridization based assays (such as the Taqman SNP assay) [10], allele-specific PCR [11,12], and minisequencing assays, such as those implemented in genome-wide genotyping microarrays [13]. These and other methods have been adopted for clinical use in CLIA-certified laboratories.

DNA sequencing is the process of characterizing the nucleotide (base-pair) sequence of a specific DNA sequence. Clinically, sequencing is used to identify the pathogenic mutations in individuals with genetic disorders caused by rare variants at a specific locus. An example is resequencing of the *BRCA1* locus in individuals at risk for familial breast cancer. Current clinical sequencing is PCR-based, with each assay focused on a specific gene region. Next-generation sequencing platforms that are essentially PCR-free enable simultaneous high-throughput characterization of billions of nucleotides simultaneously. (See "[Principles and clinical applications of next-generation DNA sequencing](#)".)

Detection of rare sequences — PCR technology also allows detection of rare sequences in a population of DNA or RNA molecules. This application is particularly useful in searching for DNA rearrangements in the setting of neoplasia. An example is the discovery that a Herpes simplex–related virus is involved in the pathogenesis of Kaposi's sarcoma [14-16].

Rare sequence detection by PCR is used routinely in obstetrics for preimplantation testing and non-invasive prenatal diagnosis (NIPD). Preimplantation testing (whereby established embryos are screened for suspected genetic mutations prior to uterine implantation) is demanding because PCR must be performed from a single cell, removed from the very early embryo. The major concerns in single-cell PCR are preventing allele dropout and contamination with maternal or other outside DNA [17]. NIPD relies on detection of free fetal DNA (ffDNA) in maternal blood; since ffDNA makes up only a small percentage of the circulating DNA, pains must be taken to differentiate fetal from maternal sequences [18]. (See "[Noninvasive prenatal testing using cell-free nucleic acids in maternal blood](#)".)

Quantifying the amount of a nucleic acid sequence — Real-time quantitative PCR (qPCR) is a highly-sensitive method for quantifying the absolute or relative amount of a specific nucleic acid sequence in which the accumulation of PCR products over time is measured directly, without post-PCR modifications. Common applications of qPCR include gene expression profiling, quantification of viral load, and copy number variation genotyping.

qPCR is performed by quantifying the amount of amplified product with each round of PCR cycling. In addition to the two primers that are necessary for successful PCR amplification, this method applies a fluorescent, non-extendible probe that hybridizes to the target sequence between the primer pair.

This probe contains a fluorogenic reporter dye at its 5' end and a quencher at the 3' end. The quencher blocks fluorescence emission as long as it is in close proximity to the reporter. As the primer extends downstream during amplification, however, the exonuclease activity of the DNA polymerase cleaves away the hybridized probe and removes the quencher. An increase in the fluorescence signal ensues ([figure 2](#)).

Due to the exponential nature of the PCR, the fluorescence signal increases proportionally to the amount of generated PCR product until a plateau is reached. Quantitation is accomplished by comparing the cycle number at which the patient sample reaches a predetermined level of fluorescence to a standardized curve of a control sample, thus deriving copy number at the start of the reaction ([figure 3](#)). This method is rapidly becoming the method of choice for monitoring residual disease in patients receiving chemotherapy and/or hematopoietic cell transplantation for hematologic malignancies [[19,20](#)]. (See "[Genetic abnormalities in hematologic and lymphoid malignancies](#)".)

Advantages of this method include the following:

- Extreme sensitivity and a wide dynamic range
- Time-efficient, as the method requires no post-PCR processing of samples
- Automated high-throughput instruments (such as the ABI PRISM 7700 Sequence Detection System) are available with interpretation-supporting software

Gene expression profiles — Specific gene transcripts can be measured by qPCR from RNA samples that have been converted to cDNA by reverse transcription. Typically, relative quantification is used to calculate fold-change among a set of samples. For example, inflammation-related genes were assayed by Taqman qPCR, comparing normal tissue with colon cancer samples [[21](#)]. qPCR expression profiling is used to validate measurements of specific candidate genes identified through genome-wide expression microarray studies.

Clinical measurement of viral RNA or DNA — PCR of RNA isolated from blood is a standard tool in monitoring the viral load in HIV-infected patients [[22,23](#)]. Several available commercial real-time PCR assays provide absolute quantitation of HIV-1 RNA copies per mL of plasma [[24](#)]. Key considerations for HIV RNA testing are throughput, automation, accuracy, and dynamic range. PCR quantification has also been applied to other RNA viruses causing chronic infection, such as hepatitis C virus [[25,26](#)]. (See "[Techniques and interpretation of HIV-1 RNA quantitation](#)" and "[Diagnosis and evaluation of chronic hepatitis C virus infection](#)".)

The 2009 outbreak of swine-origin influenza A (H1N1) prompted rapid development of diagnostic real-time, RT-PCR assays [[27](#)]. The assays developed by the Centers for Disease Control consist of four sets of primers and Taqman probes for universal detection of type A influenza, universal detection of swine influenza A, specific detection of swine H1 influenza, and the positive reference control gene, RnaseP [[28](#)].

Another area in which qPCR is becoming useful is in the quantitation of Epstein-Barr viral load in the diagnosis and monitoring of patients with post-transplant lymphoproliferative disease [[29](#)]. (See "[Epidemiology, clinical manifestations, and diagnosis of post-transplant lymphoproliferative disorders](#)", [section on 'Measurement of Epstein-Barr viral load'](#)".)

Monitoring of hepatitis B treatment response by quantifying viral DNA in serum, with real-time PCR, is another important clinical application [[30](#)]. (See "[Diagnosis of hepatitis B virus infection](#)", [section on](#)

'Serum HBV DNA assays'.)

Copy number variants — Copy number variants (CNVs) are large (>1 kb) genomic regions showing differential copy number among individuals [31]. Real-time qPCR is a robust and relatively high-throughput method for targeted CNV genotyping. (See ["Overview of genetic variation"](#).)

CRITICAL EVALUATION OF DATA — The ease of performing polymerase chain reactions (PCR) has led to wide dissemination of the methodology. However, reports of PCR analyses should be reviewed for validity of the results.

Every PCR experiment requires a minimum of two technical controls (positive and negative controls), in addition to biological controls specific to the research question being addressed. The technical controls should demonstrate that amplification occurs when it should (positive control) and that it does not occur when it should not (negative control). The positive control template should be prepared in the same manner as the test samples, with the positive control DNA typically extracted using the same extraction protocol and starting material as the sample. The negative control (also referred to as a no-template control) involves using all components of the PCR mixture except that water is added in lieu of the DNA template.

The quality of the test sample DNA is critical, as highly-fragmented DNA could result in allelic dropout (selective amplification of only one allele at a heterozygous locus) or false-negative results. Impurities in the template DNA sample could inhibit amplification. This is of particular concern in qPCR, since inhibitors will lower reaction efficiency.

SUMMARY

- Polymerase chain reactions (PCR) allows for rapid and highly-specific amplification of DNA fragments. Two technical components underlie the PCR process: the existence of thermostable DNA polymerases purified or cloned from microorganisms living in hot springs and the ability to readily synthesize specific oligonucleotide primers of 20 to 30 residues. The PCR consists of three stages, multiply repeated: denaturing, annealing, and elongation. After the reaction is completed, the amplified product (called an amplicon) can be visualized by gel electrophoresis. Relatively short DNA sequences suffice to specify a unique sequence in the genome. (See ["The PCR process"](#) above.)
- The exponential amplification of PCR can be contrasted with other types of nucleic acid amplification that are widely used in molecular biology. Whole genome amplification (WGA) non-specifically amplifies the entire genome and is used when sample quantities are too small for PCR to meet the desired application's requirements. Linear amplification of RNA is the precursor to most gene expression microarray experiments. (See ["PCR compared to other amplification methods"](#) above.)
- Most types of genetic variation can be characterized using PCR techniques, including microsatellite and single nucleotide polymorphism (SNP) markers, insertion/deletion variants (indels), and some structural variants such as copy-number variants. (See ["SNP genotyping"](#) above.)
- PCR technology allows detection of rare sequences in a population of DNA or RNA molecules. This application is particularly useful in searching for DNA rearrangements in the setting of neoplasia. Rare sequence detection by PCR also is used routinely in obstetrics for preimplantation testing and non-invasive prenatal diagnosis (NIPD). (See ["Detection of rare sequences"](#) above.)
- Real-time quantitative PCR (qPCR) is a highly-sensitive method for quantifying the absolute or relative amount of a specific nucleic acid sequence. This method is often used for monitoring residual disease in patients receiving chemotherapy and/or hematopoietic cell transplantation for hematologic malignancies. (See ["Quantifying the amount of a nucleic acid sequence"](#) above.)

- PCR of RNA isolated from blood is a standard tool in monitoring the viral load in HIV-infected patients. qPCR also has been used for assays of H1N1, Epstein-Barr viral load monitoring post-transplant patients, and monitoring response to therapy for hepatitis B. (See '[Clinical measurement of viral RNA or DNA](#)' above.)

Use of UpToDate is subject to the [Subscription and License Agreement](#).

REFERENCES

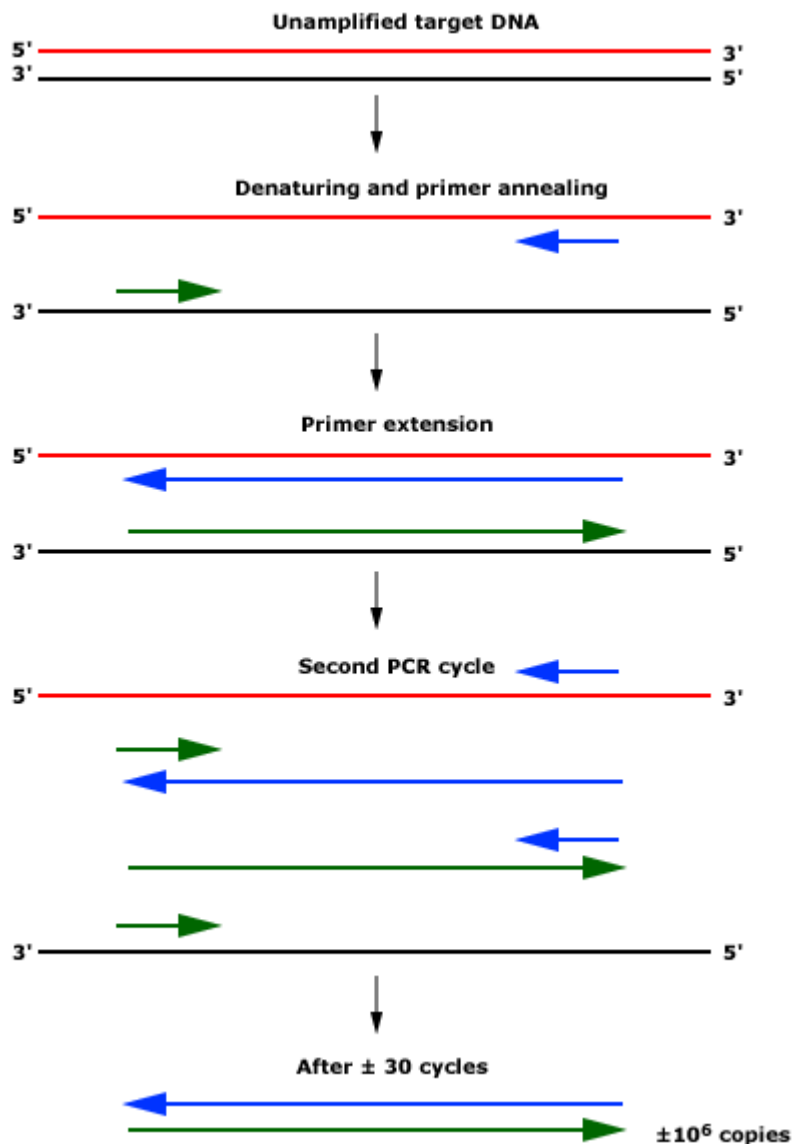
1. Templeton NS. The polymerase chain reaction. History, methods, and applications. *Diagn Mol Pathol* 1992; 1:58.
2. Saiki RK, Scharf S, Faloona F, et al. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 1985; 230:1350.
3. Rabinow P. Making PCR: a story of biotechnology, University of Chicago Press, Chicago 1996.
4. Eckert KA, Kunkel TA. DNA polymerase fidelity and the polymerase chain reaction. *PCR Methods Appl* 1991; 1:17.
5. Rozen S, Skaletsky HJ. Primer3 on the WWW for general users and for biologist programmers. In: *Bioinformatics Methods and Protocols: Methods in Molecular Biology*, Krawetz S, Misener S (Eds), Humana Press, Totowa, NJ 2000. p.365.
6. Zweig AS, Karolchik D, Kuhn RM, et al. UCSC genome browser tutorial. *Genomics* 2008; 92:75.
7. Lasken RS. Genomic DNA amplification by the multiple displacement amplification (MDA) method. *Biochem Soc Trans* 2009; 37:450.
8. Ginsberg SD. RNA amplification strategies for small sample populations. *Methods* 2005; 37:229.
9. Nygaard V, Hovig E. Options available for profiling small samples: a review of sample amplification technology when combined with microarray profiling. *Nucleic Acids Res* 2006; 34:996.
10. Livak KJ, Flood SJ, Marmaro J, et al. Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. *PCR Methods Appl* 1995; 4:357.
11. Papp AC, Pinsonneault JK, Cooke G, Sadée W. Single nucleotide polymorphism genotyping using allele-specific PCR and fluorescence melting curves. *Biotechniques* 2003; 34:1068.
12. Hurd CM, Cavanagh G, Schuh A, et al. Genotyping for platelet-specific antigens: techniques for the detection of single nucleotide polymorphisms. *Vox Sang* 2002; 83:1.
13. Matsuzaki H, Dong S, Loi H, et al. Genotyping over 100,000 SNPs on a pair of oligonucleotide arrays. *Nat Methods* 2004; 1:109.
14. Chang Y, Cesarman E, Pessin MS, et al. Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. *Science* 1994; 266:1865.
15. Moore PS, Chang Y. Detection of herpesvirus-like DNA sequences in Kaposi's sarcoma in patients with and without HIV infection. *N Engl J Med* 1995; 332:1181.
16. Mesri EA, Cesarman E, Arvanitakis L, et al. Human herpesvirus-8/Kaposi's sarcoma-associated herpesvirus is a new transmissible virus that infects B cells. *J Exp Med* 1996; 183:2385.
17. Spits C, Sermon K. PGD for monogenic disorders: aspects of molecular biology. *Prenat Diagn* 2009; 29:50.
18. Maddocks DG, Alberry MS, Attilakos G, et al. The SAFE project: towards non-invasive prenatal diagnosis. *Biochem Soc Trans* 2009; 37:460.
19. Olsson K, Gerard CJ, Zehnder J, et al. Real-time t(11;14) and t(14;18) PCR assays provide

- sensitive and quantitative assessments of minimal residual disease (MRD). *Leukemia* 1999; 13:1833.
20. Martin S, Fischer C, Free M, et al. LightCycler-based quantitative real-time PCR monitoring of patients with follicular lymphoma receiving rituximab in combination with conventional or high-dose cytotoxic chemotherapy. *Eur J Haematol* 2005; 74:282.
 21. Erreni M, Bianchi P, Laghi L, et al. Expression of chemokines and chemokine receptors in human colon cancer. *Methods Enzymol* 2009; 460:105.
 22. Clarke JR, McClure MO. HIV-1 viral load testing. *J Infect* 1999; 38:141.
 23. Churchill D, Weber J. Measurement of plasma HIV viral load—clinical applications. *J Infect* 1999; 38:147.
 24. Scott LE, Noble LD, Moloi J, et al. Evaluation of the Abbott m2000 RealTime human immunodeficiency virus type 1 (HIV-1) assay for HIV load monitoring in South Africa compared to the Roche Cobas AmpliPrep-Cobas Amplicor, Roche Cobas AmpliPrep-Cobas TaqMan HIV-1, and BioMerieux NucliSENS EasyQ HIV-1 assays. *J Clin Microbiol* 2009; 47:2209.
 25. Takeuchi T, Katsume A, Tanaka T, et al. Real-time detection system for quantification of hepatitis C virus genome. *Gastroenterology* 1999; 116:636.
 26. Martell M, Gómez J, Esteban JI, et al. High-throughput real-time reverse transcription-PCR quantitation of hepatitis C virus RNA. *J Clin Microbiol* 1999; 37:327.
 27. Perez-Padilla R, de la Rosa-Zamboni D, Ponce de Leon S, et al. Pneumonia and respiratory failure from swine-origin influenza A (H1N1) in Mexico. *N Engl J Med* 2009; 361:680.
 28. CDC protocol of realtime RTPCR for influenza A (H1N1). Geneva: World Health Organization, April 2009. Available at: www.who.int/csr/resources/publications/swineflu/CDCRealtimeRTPCR_SwineH1Assay-2009_20090430.pdf (Accessed on June 23, 2009).
 29. Wagner HJ, Wessel M, Jabs W, et al. Patients at risk for development of posttransplant lymphoproliferative disorder: plasma versus peripheral blood mononuclear cells as material for quantification of Epstein-Barr viral load by using real-time quantitative polymerase chain reaction. *Transplantation* 2001; 72:1012.
 30. Andersson KL, Chung RT. Monitoring during and after antiviral therapy for hepatitis B. *Hepatology* 2009; 49:S166.
 31. McCarroll SA. Extending genome-wide association studies to copy-number variation. *Hum Mol Genet* 2008; 17:R135.

Topic 2896 Version 14.0

GRAPHICS

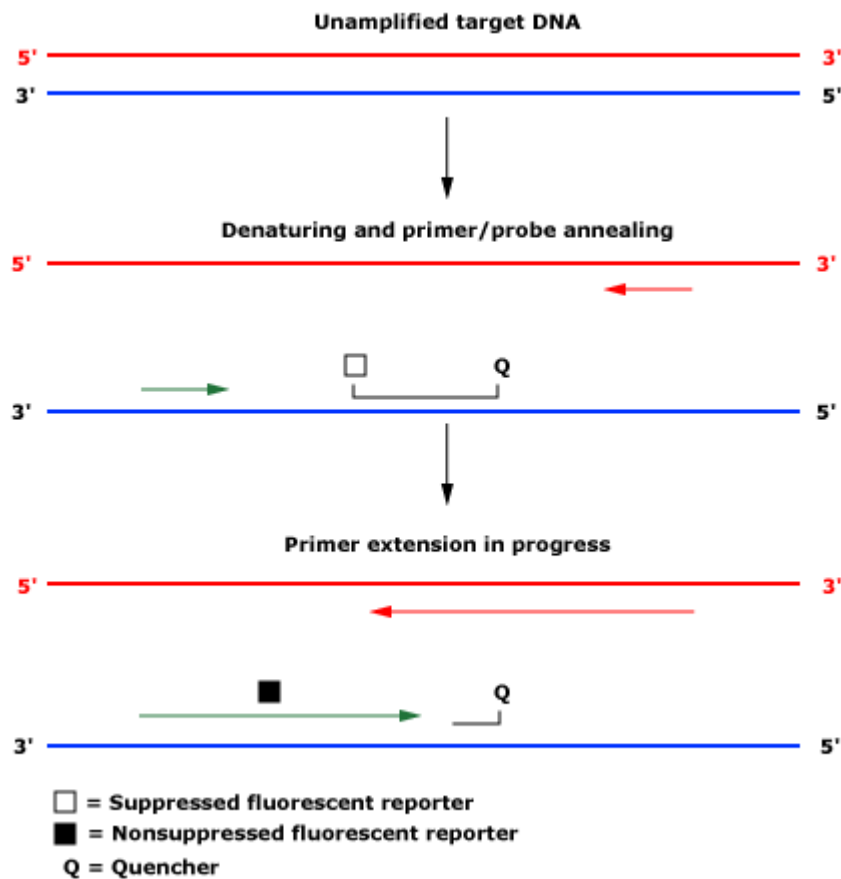
Polymerase chain reaction (PCR)



In the polymerase chain reaction (PCR), unamplified target DNA (red and black strands at top) is separated into two single strands by heating. Next, primers (green and blue arrows) anneal to the opposite DNA strands of the desired target sequence. The complementary strand is then synthesized (longer blue and green arrows). Subsequent cycles of strand separation (denaturing), primer hybridization (annealing), and strand synthesis (extension) ensure exponential amplification of the target sequence by using the amplified sequences of the previous cycle as a new template (as shown in the Second PCR cycle). The target PCR product, which consists of approximately 10^6 copies after 30 cycles, is shown at the bottom of the figure.

Graphic 64383 Version 1.0

Quantitative real-time polymerase chain reaction (PCR)

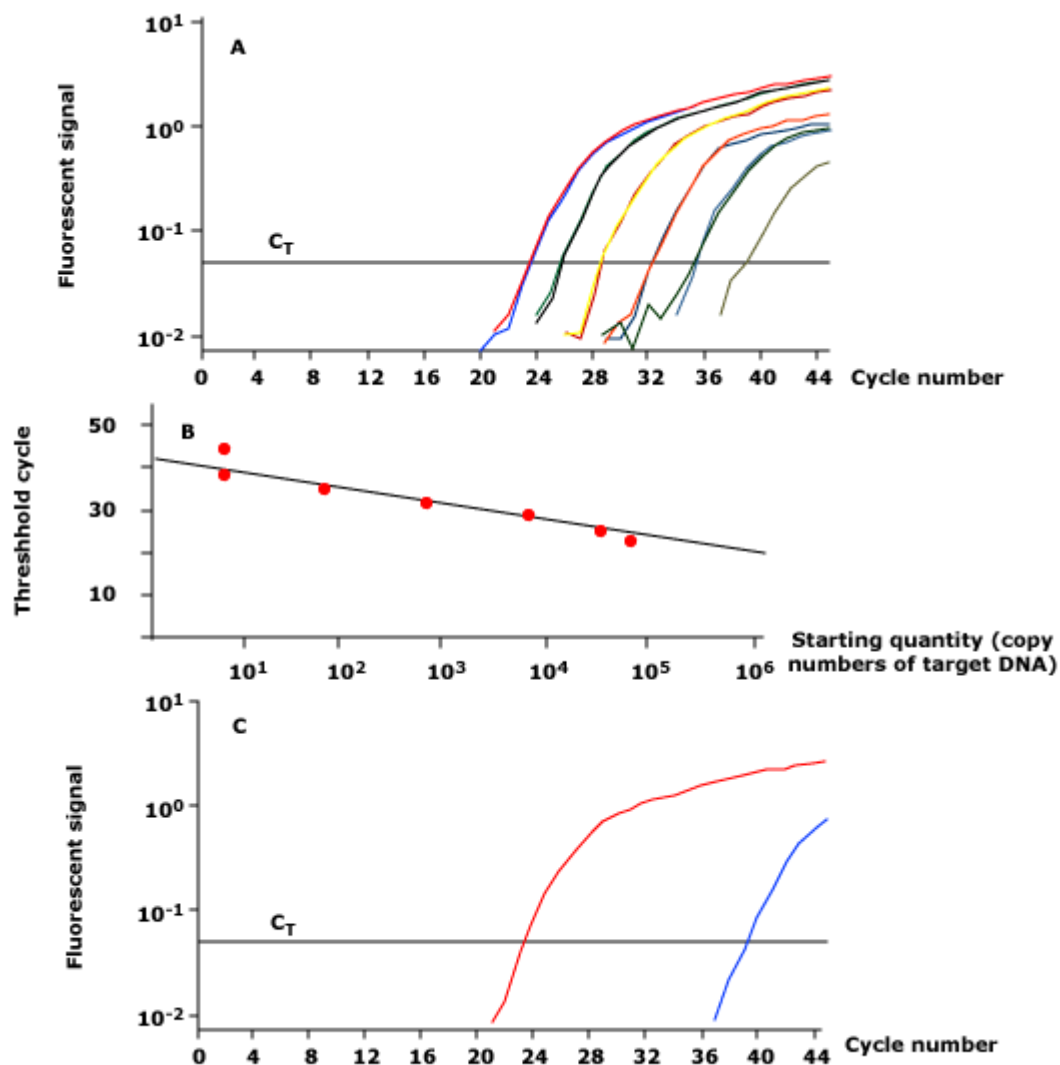


The mechanism of quantitative real-time PCR, which is based upon a probe that contains a fluorogenic reporter dye at its 5' end and a quencher at the 3' end (shown as a black line with a white box and a Q at its 5' and 3' end, respectively). With this technique, the tagged probe, which is designed to hybridize to one strand of the target DNA sequence (blue line), is added to a normal PCR. The quencher blocks fluorescence emission as long as it is in close proximity to the reporter. As the primer extends downstream during amplification (extended green arrow), the exonuclease activity of the DNA polymerase cleaves away the hybridized probe and removes the connection between the fluorescent dye (black box) and the quencher (Q). An increase in fluorescence signal ensues as more amplicons are generated.

Q: quencher.

Graphic 81663 Version 3.0

Example of quantitative real-time polymerase chain reaction



The polymerase chain reaction (PCR) can be quantitated in its early exponential phase. Panel A: If there are many targets in the starting sample (eg, curves on the left), the generated fluorescent signal will cross a predetermined threshold value (C_T) early in the PCR cycling process. If there are few targets (eg, curves on the right), the signal will rise more slowly and will cross C_T at a later PCR cycle. Panel B: When C_T is plotted versus the starting quantity of target DNA, the relationship is linear over more than 8 logarithms. To establish the amount of targets in a sample, one can simply determine the cycle number at which C_T is crossed and use the standard curve to infer the number of targets present in the sample. Panel C: In an example of a cancer patient with t(14;18), the initially large number of targets present (red curve) is markedly reduced after treatment (blue curve).

Graphic 71655 Version 1.0

Disclosures

Disclosures: **Benjamin A Raby, MD, MPH** Other Financial Interest (Spouse): Parexel International [Contract research organization (Double-blind randomized clinical trials)]. **Anne Slavotinek, MBBS, PhD** Nothing to disclose. **Jennifer S Tirnauer, MD** Nothing to disclose.

Contributor disclosures are reviewed for conflicts of interest by the editorial group. When found, these are addressed by vetting through a multi-level review process, and through requirements for references to be provided to support the content. Appropriately referenced content is required of all authors and must conform to UpToDate standards of evidence.

Conflict of interest policy

Incidental findings from genetic testing**Authors**

Kurt Christensen, MPH, PhD
Sarah S Kalia, ScM, CGC
Robert Green, MD, MPH

Section Editor

Benjamin A Raby, MD, MPH

Deputy Editor

Jennifer S Tirnauer, MD

All topics are updated as new evidence becomes available and our [peer review process](#) is complete.

Literature review current through: Feb 2015. | **This topic last updated:** Feb 25, 2015.

INTRODUCTION — The increasing use and capabilities of genomic tools such as whole genome sequencing and exome sequencing raise important questions about how to handle health-related information that may inform prevention or treatment strategies, but are unrelated to the reasons testing was ordered. The questions of whether – and how – to disclose incidental findings (IFs) from genetic testing have generated much debate, and the importance of how these questions are answered is expected to grow as laboratories and physicians transition to whole genome and whole exome sequencing rather than targeted gene panels.

This topic review discusses an approach to the disclosure of IFs from genetic testing.

An overview of genomic disorders, tools for genomic testing, issues related to genetic counseling, and a glossary of genetic terms are presented separately:

- (See ["Genomic disorders: An overview"](#).)
- (See ["Principles and clinical applications of next-generation DNA sequencing"](#).)
- (See ["Genetic counseling and testing"](#).)
- (See ["Glossary of genetic terms"](#).)

DEFINITIONS AND CLASSIFICATION OF VARIANTS — The term incidental finding (IF; also called “unexpected result” or “secondary variant”) has been defined in several different ways [1,2]. Narrowly defined, IFs are genetic variants that were unsolicited by the patient, physician, or laboratory, and were identified unintentionally during genomic analysis. Here, we use a broader definition of IFs that includes health-informative variants that are unrelated to the specific purposes of testing, but may have been identified through an intentional search, including an untargeted scan for any genetic finding of interest [3].

IFs can be identified by any genetic test, including single gene tests and microarray analysis. However, IFs are likely to be increasingly identified from genome sequencing methods such as whole genome sequencing or whole exome sequencing, because molecular genetics laboratories are increasingly incorporating these types of testing, and whole genome/whole exome methods have a greater potential to identify IFs than methods that assay only one or a few genes [4]. (See ["Principles and clinical applications of next-generation DNA sequencing", section on "Whole genome, exome, or gene panel"](#).)

When reported, IFs are classified by the laboratory performing the testing according to their likely pathogenicity, based on an interpretation of data from the literature that correlates variants with clinical disease in reference populations. (See ["Decisions made by the laboratory"](#) below.)

Most laboratories classify variants in categories such as the following:

- **Pathogenic/known deleterious** – Pathogenic/known deleterious variants are variants previously reported in patients with disease and/or are strongly suspected of being pathogenic based on preclinical studies.

- **Likely Pathogenic/presumed deleterious** – Likely pathogenic/presumed deleterious variants are those with sequence features that are likely to be implicated in disease pathogenesis but for which conclusive evidence of pathogenicity is not available.
- **Variant of unknown significance** – Variants of unknown significance (VUS; also called “variant of uncertain significance” or “finding of unknown clinical significance”) are variants that have some features suggestive of possible functional consequence, but for which there is insufficient evidence for either a pathogenic or benign role.
- **Likely benign/presumed benign** – Likely/presumed benign variants are those for which weak data supporting pathogenicity may be available, but for which the majority of evidence suggests the effect of the variant is benign.
- **Benign/known benign** – Benign/known benign findings are genetic variants not predicted to alter gene expression or function.

Decisions about whether to report a variant as an IF and how to classify the variant are made by individual laboratories, sometimes without input from ordering physicians or patients [5,6].

An often-cited list of 56 genes published by the American College of Medical Genetics and Genomics (ACMG) represents a minimal list of genes to be examined for IFs based on criteria that emphasize consensus-based clinical validity (ie, the accuracy with which the finding predicts disease risk or presence) and utility (ie, the ability to use IFs to impact clinical management, and risks and benefits resulting from test use) ([table 1](#)) [5]. The ACMG plans to update its recommended list of genes that laboratories should proactively examine during sequencing [7]. Other groups have generated local criteria for reportable findings that would alter clinical management, such as glucose 6 phosphate dehydrogenase deficiency (G6PD) and Fabry disease [8]. In addition, the Presidential Commission for the Study of Bioethical Issues has recommended that professional organizations deliberate and provide specific guidance for different fields and contexts, rather than endorsing a uniform obligation across varied contexts [3].

Of note, recommendations such as the ACMG list provide a list of genes to be examined for variants, rather than specific variants within those genes that should be reported, and specialists often disagree whether specific variants warrant reporting as IFs [9]. Moreover, even when a gene’s association with disease risk is well established, penetrance estimates may have been developed from high-risk populations, and these estimates may generalize poorly to different groups such as healthy individuals in the general population. (See ["Overview of Mendelian inheritance", section on 'Causes of non-Mendelian inheritance and effects that confound recognition of Mendelian inheritance'](#) and ["Genetic counseling and testing", section on 'Penetrance, ascertainment bias, and variable expression'](#).)

LIKELIHOOD OF DETECTING AN INCIDENTAL FINDING — The prevalence of IFs depends on the criteria used for identification and reporting.

- All individuals have pharmacogenomic variants that affect medication response, as well as hundreds of variants that provide information about disease susceptibility [10,11].
- Between 25 and 50 percent of individuals are carriers for at least one severe, recessive childhood disorder, although estimates as high as 2.8 carrier variants per person have been noted [10,12,13].

The frequencies of IFs that might directly lead to a change in clinical management are illustrated by the following reports:

- An analysis of 1000 individuals randomly selected from the National Heart, Lung, and Blood Institute’s Exome Sequencing Project, which screened for variants in 114 genes associated with highly penetrant, ‘medically-actionable conditions,’ identified IFs in 3.4 percent of individuals of

European descent and 1.2 percent of individuals of African descent [14]. The lower rate of IF identification in African Americans is believed to be due to the smaller database of available information for these individuals rather than a true lower rate of variants.

- A more restrictive analysis was conducted among 572 ClinSeq Study participants ages 45 to 65, most of whom were generally healthy or had a personal history of cardiac disease. Analysis of 37 cancer-related genes identified seven individuals (1.2 percent) with deleterious BRCA1/2 variants (including three who did not meet family history-based criteria for a genetics referral) and one individual with a deleterious mutation associated with paragangliomas [15].
- Preliminary data from an evaluation of 462 participants in the Framingham Heart Study who were unselected for family history revealed that 1.7 percent had pathogenic variants in a subset of disease-associated, dominantly-inherited Mendelian genes [16].

INFORMED CONSENT — Consensus exists that patients should be informed about the potential for identifying IFs prior to undergoing genetic testing, including what kinds of IFs could be detected, which IFs would be disclosed, and how the information would be communicated [17-19].

Discussions between patients and providers about genomic IFs should culminate in joint decisions that account for the patient's clinical status, values, and stated preferences [3,20]. In the cases where genomic sequencing is used, the American College of Medical Genetics and Genomics (ACMG) recommends counseling by a medical geneticist or genetic counselor prior to testing; this would include notification that IFs may be identified and clarification about what kinds of IFs would be disclosed [21].

Patients should also be provided an "opt out" option in which laboratories would restrict bioinformatics analysis to genes relevant to the reasons testing was ordered. Individuals may also opt out of disclosure of IF results after analysis has occurred. The advantages and disadvantages of opting out of genetic analysis versus disclosure of results are complex and raise a number of ethical issues. Many experts believe that an individual who wishes to opt out of receiving certain IFs is better served by opting out of genomic analysis, in order to preserve patient autonomy and privacy. In two large studies that used exome sequencing for diagnosis in suspected genomic disorders, 1 to 8 percent of individuals opted out of receiving information on IFs; with the higher percentage of opt outs for information about recessive disorders and pharmacogenomic variants [8,22]. Regardless of whether a patient has chosen to opt out of gene analysis or disclosure, ordering physicians may inform patients that their preferences for receiving IFs will be re-assessed at the time when primary findings are disclosed.

The National Society of Genetic Counselors (NSGC) has released a statement emphasizing the importance of pretest counseling to ensure patients make informed decisions about receiving IFs, particularly from genomic sequencing [23]. Suggested components of pretest counseling include:

- Explain the information that may be returned, including unexpected and unsolicited information about risks of cancer and cardiac disease.
- Gather the patient's family history and discuss risks that arise in the family history.
- Discuss the limitations of analysis for IFs.

Importantly, genomic sequencing does not substitute for family history analysis in assessing disease risk. This was illustrated in a study that compared family history-based risk assessment with single nucleotide polymorphism (SNP)-based testing for breast, prostate, and colon cancer risk [24]. Family history-based assessment classified 22 of 88 participants as high-risk, while SNP-based testing only identified one high-risk patient.

Informed consent for sequencing and the accompanying IF analysis should also include discussions of the likelihood of detecting an IF, the types of findings that will and will not be disclosed, the patient's preferences for learning these types of information, and a plan for disclosing the results [21].

Potential harms of disclosure of IF include psychological distress from learning of disease risk, financial and personal costs of additional testing that may be indicated, and adverse consequences of therapeutic interventions for which evidence of benefit is lacking in patients diagnosed incidentally. As with all genetic test results, federal legislation protects patients from employment or health insurance discrimination, but similar protections are lacking for other types of insurance (eg, life, disability, and long-term care insurance) [25]. There may be additional psychological harms from learning about a risk of disease for which there is no effective intervention (eg, Huntington disease), or risks to privacy [1,2]. In addition, understanding about genetics is poor among many physicians, raising risks that IF reports will be misinterpreted [26-28], particularly if the IF relates to a rare condition. Conversely, a report of no IFs detected may falsely reassure an individual that s/he is not at risk of disease and cause the individual to ignore personal symptoms or a strong family history of disease. Of note, the evidence threshold laboratories use for reporting a variant to physicians and patients tends to be much higher in the context of incidental findings compared with variants being analyzed for diagnostic purposes. (See ['Interpretation of "negative" results'](#) below.)

A final consideration for the consenting process is what to do with incidental findings after death. Patients and research participants often express wishes that any genetic risk information about them be shared with relatives after their death [29]. However, health privacy regulations in the Health Insurance Portability and Accountability Act (HIPAA) prohibit such disclosure without prior authorization from the decedent or a legal representative [30]. Addressing this issue during the informed consent process may help patients initiate discussions with relatives about whether they would want to receive postmortem disclosure of IFs, along with any other medical information that might be critical to interpreting the clinical significance of an IF. (See ['Postmortem disclosure'](#) below.)

Additional ethical and psychosocial issues are discussed separately. (See ["Genetic counseling and testing", section on 'Ethical, legal, and psychosocial issues'](#).)

DISCLOSURE OF INCIDENTAL FINDINGS TO PATIENTS

Decisions made by the laboratory — Each laboratory performing genomic testing makes decisions regarding which results are reported to the clinician. Unlike more straightforward clinical laboratory testing (eg, measurement of a serum protein concentration), genomic testing measures sequence variation at the nucleotide level in a very large portion of the genome; laboratories determining genetic sequence information are responsible for providing the interpretation of the clinical relevance of the variants identified.

- Each laboratory typically has its own policies for identifying, analyzing, and reporting IFs. When genomic sequencing is used, a laboratory may have a pre-specified panel of genes it will examine for IFs and report variants for unless instructed to do otherwise by the ordering physician; in some cases, laboratories may have a policy in place to allow clinicians or patients to opt out of receiving IFs, or to allow them to receive an expanded panel of results upon request (eg, carrier status variants, pharmacogenomic variants).

Laboratories are also encouraged to establish criteria and thresholds for reporting findings associated with consanguinity, such as multiple regions of extended homozygosity, because these findings can increase risks for autosomal recessive conditions [31-33]. In circumstances where genetic testing results suggest possible conception between first- or second-degree relatives, laboratories are encouraged to discuss their findings with the clinician who ordered the test, because the clinician may have legal obligations for reporting potential rape or sexual assault [33].

- Each laboratory decides what risk assessment to assign to each variant within the gene; this is generally done according to an accepted classification scheme based on the likely clinical significance of each variant (see ['Definitions and classification of variants'](#) above). Interpretations of

risk are generally made by reviewing published and/or proprietary epidemiologic, genetic, and biologic data, and using algorithms to predict the functional significance of certain variants. Patient information (eg, laboratory results, personal and family history of disease) may also inform how variants are classified.

Recommendations for conditions and genes to query for IFs usually consider the existence of validated screening procedures to confirm the presence or absence of disease [5]. The American College of Medical Genetics and Genomics (ACMG) has identified a list of 56 genes associated with 24 conditions that it considers a minimal set for which known or likely pathogenic variants should be reported ([table 1](#)) [5]; there are plans for this list to be updated [7,34,35].

There may be relatively broad consensus in reporting a known pathogenic variant in a known cancer gene (eg, pathogenic BRCA1 mutation). However, laboratory assessments regarding the significance of a variant may diverge when the pathogenicity of the variant has not been established previously. This divergence in assessments was illustrated in a study that compared results from two genomics laboratories using a genotype-panel to test DNA from five individuals [36]. The information reporting the genetic sequence was concordant for more than 99 percent of markers tested, but the risk prediction was concordant across all five subjects for only 4 of 13 medical conditions. Of note, this study evaluated susceptibility variants for common conditions as well as Mendelian diseases, which may have contributed to the high degree of disparity in risk predictions.

Most laboratories will report a genomic IF if the sequence variant is “previously reported and is a recognized cause of the disorder” (ie, pathogenic/known deleterious), and many will report a variant that is “previously unreported and is of the type which is expected to cause the disorder” (ie, likely pathogenic/presumed deleterious) [37]. However, estimates of the predictive value of testing, penetrance of specific variants, spectrum of possible phenotypes, and efficacy of interventions may have been developed from research on sick patients and may not generalize well to asymptomatic populations. Laboratories should consider the patient’s baseline risk and ethnic background when they assess the pathogenicity of a particular variant [38].

Experts have called for the development of standards for the identification and reporting of genomic IFs, and recommendations continue to evolve [39]. In the interim, individual laboratories are encouraged to develop clear policies based on their own assessment of existing evidence [5,7]. In all situations laboratories are strongly encouraged to set high thresholds for defining pathogenicity for patients who have no prior personal or family history suggesting a condition with a strong genetic component [40]. Physicians will need to take these considerations into account when they decide whether and how to disclose IFs to their patients. (See '[Review patient report](#)' below.).

Review patient report — Prior to meeting with the patient, the clinician should do the following:

- Review patient preferences communicated at the time of consent for genomic testing (eg, did the patient opt out of receiving certain results?).
- Weigh the potential harms and benefits of reporting the IF.
- Understand the clinical implications of the IF for the patient.
- Reevaluate the patient’s personal medical history, family history, and physical examination in light of the IF, if appropriate.

Importantly, patient preferences for information and the clinical significance of an IF may change over time. The decision about whether and when to disclose IFs will thus depend on the clinical context and the judgment of clinicians involved in the patient’s care. Additionally, ethical obligations to respect patients’ autonomy and right not to know information must be weighed against fiduciary duties to inform patients of potentially life-saving information [41-47]. A Presidential Commission has urged professional

organizations to develop guidelines for responding to genomic IFs that address these ethical and legal issues [3]. (See ["Informed consent"](#), [section on 'The duty to inform'](#) and ["Liability concerns"](#) below.)

In addition, physicians will need to consider the IF in the context of the patient's primary reason for testing. As an example, patients for whom testing was done for a more immediate problem may be unreceptive to learning about IFs or poorly able to process the information until the primary indication for genomic testing has been addressed.

Whether to disclose results indicating consanguinity is a question that may warrant special consideration, given varying norms about marriage within families. Clinicians will need to consider how disclosure affects the physician-patient relationship, the psychosocial impact on patients, and/or any ethical or legal obligations for reporting that may exist [48]. It may be helpful to disclose results to families with a team that includes a geneticist, genetic counselor, and patient advocate to minimize the potential for distress and to ensure families understand the implications for disease [48]. If either parent of a suspected consanguineous conception was a minor at the time of conception, a child-protection team should be consulted as necessary to understand legal obligations for reporting potential rape or sexual assault [33].

The clinical relevance of IFs for the patient should always be interpreted in the context of the patient's personal medical history and family history of disease, as well as physical examination and other relevant information (eg, overall health, behavioral risks, reproductive issues). IFs may help to explain prior diagnoses of the patients or their relatives. In some cases, repeating the physical examination for subtle findings associated with disease characteristics may be appropriate.

The risk of disease is also influenced by the disease penetrance and expressivity, which varies by patient population:

- Variants with well-established pathogenicity may have reduced penetrance, and an individual with such a variant may never develop disease. (See ["Basic principles of genetic disease"](#), [section on 'Penetrance'](#).)
- Variants may have variable expressivity; identification of a variant may not predict the severity of the phenotype or the specific disease manifestations that may develop. (See ["Non-Mendelian inheritance patterns of monogenic diseases"](#), [section on 'Penetrance and variable expressivity'](#).)
- The predictive value of testing, spectrum of phenotypes, and efficacy of interventions in asymptomatic populations may not be well-established, because clinical information about genomic variants is often drawn from epidemiological research on individuals with classic disease phenotypes and/or families with a strong history of disease [41]. As a consequence, penetrance estimates for genetic variants are overstated in some databases and publications [4,20,49,50]. Databases and resources that provide information more applicable to the general population are being developed. (See ["Online information and resources"](#) below.)
- Information regarding variant interpretation may also be less relevant for members of ethnicities that are underrepresented in reference populations. (See ["Underrepresented ethnicities"](#) below.)

An example of a decision matrix regarding which IFs should be disclosed is presented in the table ([table 2](#)). This approach classifies variants into 'bins' based on their potential clinical utility to the patient, clinical validity (accuracy of predicting disease for that patient), and risks that disclosure may pose (eg, psychological distress, familial discord). It proposes routinely disclosing IFs of established utility and clear or presumed pathogenicity. IFs with clear validity but questionable utility to the patient could be disclosed depending on the potential harms of disclosure and patient preferences. Finally, IFs without established validity and utility would be withheld altogether [51]. The use of such a decision matrix may make genomic information less overwhelming [52]. However, such algorithms may be difficult to

implement consistently, given that consensus about how to classify variants is often lacking and patient preferences may change over time [9].

Disclose IFs to patient — Disclosure of IFs involves informing the patient that the IF(s) are present, counseling regarding the clinical implications of the IF(s), and discussing whether any additional intervention is needed. Physicians who receive IF reports must document whether and how they have acted upon the findings and their rationale, regardless of whether or not the patient has chosen to receive the information.

Importantly, detection of a pathogenic or likely pathogenic IF provides information about risk, not diagnosis of a disease, and the identification of an IF is not a substitute for further diagnostic testing, if indicated. Conversely, “negative” reports from genomic testing may also require counseling to discuss the interpretation and further diagnostic steps. (See '[Additional evaluation](#)' below and '[Referrals](#)' below and '[Interpretation of “negative” results](#)' below.)

Thus, there are a variety of potential responses to an IF:

- More intensive screening may be indicated. As an example, colonoscopy screening for colon cancer may be initiated at an earlier age than recommended for the general population in a patient with a gene mutation associated with colon cancer predisposition. (See '[Additional evaluation](#)' below.)
- Prophylactic surgery may be considered. As an example, prophylactic hysterectomy and bilateral oophorectomy may be performed in a woman with a BRCA1 mutation who has completed childbearing.
- Carrier screening for a partner may be performed for a patient of reproductive age. As an example, if a patient has a recessive mutation for a disorder with a high carrier frequency in the population (eg, sickle cell disease in African Americans, cystic fibrosis in Caucasians, or Gaucher disease Type 1 in Ashkenazi Jewish individuals), the partner may be screened and results incorporated into preconception counseling discussions.
- The patient may undertake lifestyle modification for risk reduction. As an example, a patient with a mutation associated with increased risk for melanoma may have increased adherence to screening recommendations for skin cancer.
- Genetic counseling may be appropriate. Examples include addressing uncertainty about how to interpret an IF and the impact of the variant on clinical management; assisting a patient in choosing among multiple options for which there is clinical equipoise; and helping a patient in planning discussions regarding IFs with relatives who may share the genetic risk.

One of the most challenging issues is the management of a patient with a known pathogenic variant for a potentially life-threatening disease (eg, BRCA1 mutation and risk of breast or ovarian cancer) for which the most effective intervention is major surgery. Evidence to guide management may exist for patients with a personal or family history of the disease, but comparable evidence is lacking for unselected populations [41]. Clinicians may need to make critical decisions about how to respond to IFs reported by the laboratory despite a lack of consensus about how to interpret these findings [49]. (See '[Informed consent](#)' above.)

The amount of time required to review IFs will depend on the context that warranted genetic testing in the first place and the content of the IF report. When testing was initiated for diagnostic or treatment purposes, it may be appropriate to postpone disclosure of nonurgent IFs until primary concerns have been addressed and the patient is better able to understand and react to the information [53].

Importantly, the nature of IFs means that the physician disclosing the result may have little or no prior knowledge about condition. Physicians may need to allocate time to self-educate in preparation for IF

disclosure.

A final consideration that providers will need to consider is whether patients will communicate information to other family members. Genetic services often provide patients with written materials to facilitate sharing information with others. Primary care physicians can encourage patients to communicate with family members by checking in with patients to see how the information is flowing through the family and identifying possible barriers to communication [26]. In some cases, patients may designate family members to receive their genomic information upon their death. (See ['Electronic data storage'](#) below and ['Postmortem disclosure'](#) below.)

Interpretation of “negative” results — A “negative” genomic sequencing IF report does **not** imply that an individual is not at increased risk of developing disease(s). Reasons for this include the following:

- Most individuals carry dozens, if not hundreds, of variants that have not been characterized due to lack of evidence, but which may predispose them to disease. Such variants may be readily detected through genomic sequencing, but laboratories are discouraged from reporting variants that have limited evidence for pathogenicity [38].
- Genome sequencing has a number of technical limitations. It reliably detects single nucleotide variants and small insertions or deletions, but can be less reliable for detecting moderate-sized insertions or deletions. Sequencing, particularly whole exome sequencing, may fail to detect larger insertions, deletions, repeat expansions, and rearrangements [4,49,54]. (See ["Principles and clinical applications of next-generation DNA sequencing"](#), section on ['Technical considerations'](#).)
- The search for IFs during genomic sequencing usually is not as complete as a targeted gene query that may use Sanger sequencing to detect single nucleotide changes and insertions/deletions, or as array comparative genomic hybridization (CGH) to detect larger structural variation. Other differences between genomic sequencing and Sanger sequencing are discussed separately. (See ["Principles and clinical applications of next-generation DNA sequencing"](#), section on ['Terminology and evolution of technologies'](#).)

Thus, if a patient has, or develops, a personal or family history that is consistent with a specific genetic condition, referral to a genetics specialist is recommended; and targeted gene analysis may be appropriate. (See ["Genetic counseling and testing"](#), section on ['Indications for genetic services'](#) and ['Referrals'](#) below.)

FOLLOW-UP CARE AND RECORD KEEPING — Genomic data may be useful to the patient for current as well as future management, and the clinician should ensure that the patient understands how genomic IFs might be used in the present and the future. As an example, a pharmacogenomic IF may not have immediate relevance but might affect future medication management. An ongoing challenge is how to integrate genomic data throughout the patient's life.

Additional evaluation — To minimize the potential for false-positive IFs, clinicians should address with the laboratory whether reported IFs had been confirmed through an orthogonal testing strategy (eg, Sanger sequencing if the IF had been identified during whole exome sequencing) and should consider additional genetic testing such confirmatory evaluation has not been performed. Some patients may also require additional evaluation after identification of genomic IFs. Examples include electrocardiogram (ECG) and echocardiography after detecting an IF associated with hypertrophic cardiomyopathy, or imaging studies and blood pressure monitoring after detecting an IF associated with hereditary paraganglioma-pheochromocytoma syndrome.

Searchable directories of laboratories that provide disease-specific testing are provided on the Genetic Testing Registry and GeneTests websites.

- **Genetic Testing Registry** – www.ncbi.nlm.nih.gov/gtr/

- **GeneTests** – www.genetests.org/

In contrast, some information such as pharmacogenomic IFs do not require additional testing, but should be recorded in the medical record and communicated to the patient, so that appropriate dosing changes or closer monitoring can be instituted if implicated medications are considered in the future. (See "[Overview of pharmacogenomics](#)" and "[Personalized medicine](#)".)

Referrals — Depending on which IFs are reported and the physician's familiarity with interpreting and acting upon them, patients may benefit from referrals to specialists experienced with the condition(s) in question [54].

Clinicians should also be prepared to help patients locate a genetics professional to manage follow-up of IFs [54]. Genetic counselors are master's-level healthcare professionals who work with patients and physicians through all aspects of the genetic testing process, from informed consent and test ordering through results return and medical management. Web-based tools for locating a genetic counselor are provided by the [American College of Medical Genetics and Genomics](#) (ACMG) and the [National Society of Genetic Counselors](#) (NSGC) [55,56].

Electronic data storage — Most health systems have limited capabilities for storing and retrieving genomic information. Efforts are underway to develop real-time decision support systems to notify providers of the availability of genomic information that would inform a specific treatment decision [57]. The National Institutes of Health (NIH) has created networks of institutions to facilitate the development of medical record systems better suited to storing and using genomic sequencing information [57]. Barriers to the integration of genomic findings into electronic medical records and accompanying decision support include a lack of standardized nomenclature for genetic variants, and a lack of consensus about what information to store [58]. A more extensive discussion of the integration of genomic data into electronic medical records was addressed in a special issue of *Genetics in Medicine* in 2013 [59].

More generally, the knowledge base for genomics continues to rapidly expand. Patients may expect to be informed about updates regarding the effects of genomic variants on their health and well-being, and laboratories have the capability to re-examine data files, which are stored in a widely used formats such as binary version of sequence alignment/map (BAM) or variant call format (VCF), if additional analyses are needed in the future. Few health systems have the capability to provide automated updates to genomic test results, although software such as the GeneInsight Suite has been developed to store genomic sequencing information and automatically provide updated reports to clinicians as they relate to primary test findings [60,61]. It is important for physicians to be clear with patients about whether and how they would be informed about any updated interpretations of IFs. Of note, patients often appreciate sharing responsibility for tracking scientific developments and monitoring how they might affect interpretation of genetic testing results [45].

Postmortem disclosure — Clinicians have a commitment to uphold the privacy and autonomy of their patients, and direct disclosure by a clinician to a patient's relatives after the patient's death is generally avoided unless the patient or a legally authorized representative had explicitly given permission.

When incidental genomic results are available for a deceased patient, Health Insurance Portability and Accountability Act (HIPAA) regulations allow disclosure of the deceased patient's clinical results to a health care provider for purposes of treating a relative of the decedent. However, when a deceased patient's clinical results would not be directly applied to a relative's clinical care, HIPAA guidelines prohibit disclosure of the decedent's results directly to relatives unless an authorization had been signed by the decedent or his/her legally designated administrator or executor [30].

Liability concerns — No federal or state statutes directly address a clinician's duty to disclose IFs to patients [3]. However, liability for failure to disclose IFs is a common concern among physicians, and

some argue that statements such as the American College of Medical Genetics and Genomics (ACMG) recommendations for IFs create fiduciary obligations for disclosure as such professional standards may help define a legal standard of care [3,46].

SPECIAL POPULATIONS — Additional ethical and practical considerations apply to certain populations. Examples include underrepresented ethnicities, children, and participants in research studies.

Underrepresented ethnicities — Classifying a genetic variant for individuals of some ethnicities can be challenging, because most informatics algorithms incorporate the prevalence of the variant in a reference population (eg, 1000 Genomes cohort), in which the patient's ethnicity may be poorly represented [49]. This issue can be particularly challenging for individuals with African ancestry, given their greater genetic heterogeneity. Additionally, penetrance estimates for pathogenic variants have often been established from epidemiological research on primarily Caucasians and may not apply to other ethnic groups.

In addition, ethnic groups often vary in their beliefs about genetics and interpretation of genetic risk information [62]. Physicians should be sensitive to the way ethnicity may affect both the technical interpretation of genetic data and the way patients interpret and respond to genetic test results.

Children — Disclosure of IFs in children may be particularly challenging, because the initial testing of a child may be performed for the purpose of determining the genetic basis of a presumed genetic syndrome that has major impacts on health, and the family may be focused on that indication rather than secondary results. Additional ethical issues related to consent from minors are also of concern. Clinicians need to weigh the potential costs of disclosure, including harms from diagnostic testing, long-term surveillance, familial testing, psychosocial disruptions, and financial expenses, against the potential health benefits of disclosing IFs to the child and family.

Policy statements related to pediatric genetic testing often discourage the disclosure of adult-onset conditions in children until they have reached the age of majority and can decide for themselves whether or not to pursue testing [46,63,64]. However, the American College of Medical Genetics and Genomics (ACMG) Incidental Findings Working Group recommended seeking and reporting a minimum list of IFs to ordering physicians regardless of the patient's age, because another opportunity to identify these "hidden" risks may never arise in the future, and the information might impact the health of the parent carrying the same variant [7]. The ACMG recommends that the same list of 56 genes be analyzed for incidental findings for all patients undergoing whole genome or whole exome sequencing, regardless of age (table 1) [5]. (See '[Definitions and classification of variants](#)' above.)

Critics note that the ACMG's stance conflicts with previous recommendations against testing children for adult-onset conditions and that the recommendations undermine a child's right to decide whether to learn, in adulthood, whether they have inherited a particular genetic risk variant [44,46]. However, without a phenotype or family history suggestive of a disease, the patient's risk would not otherwise have come to clinical attention. In contrast to predictive genetic testing for a known familial risk, an incidental findings report may be the only chance for the patient and relatives to learn of their at-risk status, a scenario that the ACMG categorizes as "opportunistic screening." This scenario presents a different calculus than those in which a patient is aware of his/her at-risk status due to a family history of the disease, and can choose in adulthood to pursue genetic testing for that condition [5,7,39].

Proponents note that an incidental finding discovered in a child who undergoes genome sequencing for another clinical indication may reveal a risk with immediate medical relevance to one of the child's parents (eg, a mutation in the BRCA1 gene). For families without a known family history of the disease, the parent may be unaware of the increased risk. Early knowledge of, and screening for, a known genetic risk could improve the parent's prognosis and even save his or her life, which could indirectly benefit a child. For parents in their childbearing years, incidental risk information about devastating

childhood onset conditions could also be useful for reproductive planning [5,7,39].

Research subjects — Some individuals undergo genetic testing as part of a research study that may be unrelated to their personal or family health care. Further, this testing is often performed in a research laboratory, rather than a Clinical Laboratory Improvement Amendments (CLIA)-certified laboratory, and thus may not have the accompanying analytic validity provided by clinical testing. These differences raise additional ethical issues for the research subject. Obligations for disclosure are unclear, with some noting that researchers are obligated to disclose actionable, high-impact IFs from any stage of research, and others preferring disclosure only in limited circumstances [65-68]. Complicating the matter, patient consent for receiving research IFs may have been solicited years before the research has been conducted or outside of a medical context.

Increasingly, investigators are disclosing individual results of genomic testing from research studies if these results may affect participant healthcare [19]. In general, researchers should consider the same principles for responding to IFs from genomic research and IFs from clinical genomic testing. Consent for the research should emphasize that genetic testing is being conducted for research rather than clinical purposes [21]. As translational genomics research advances, the boundaries between clinical and research testing may become increasingly blurred. Researchers and healthcare systems are in the process of developing policies, systems, and infrastructure to accommodate the responsible return of individual research results as IFs [47].

ONLINE INFORMATION AND RESOURCES — Many professional organizations for physicians have developed educational materials about genomics and/or guidelines for specific genetic conditions.

- **ACMG** – The American College of Medical Genetics and Genomics (ACMG) website (www.acmg.net) provides resources such as policy statements, practice guidelines, and educational information.

The ACMG is also creating clinical decision support tools, known as “ACT Sheets,” specific to genomic sequencing for physicians with limited experience in genomic medicine; the ACMG plans to communicate updates to its IF recommendations through its newsletter and in its peer-reviewed monthly journal, *Genetics in Medicine* [7].

- **NIH** – The National Human Genome Research Institute (NHGRI) at the National Institutes of Health (NIH) hosts an Online Health and Support Resources portal (www.genome.gov/11510197) and articles summarizing support and educational materials [69].

A number of additional resources are available to provide information about genetic conditions. Sites such as **ClinVar** are preferred to some of the existing public databases of genomic variants, such as the database of Genotypes and Phenotypes (dbGaP) and the Human Gene Mutation Database (HGMD), which frequently exaggerate the association between genetic variants and disease [70]. The following resources may be helpful:

- **MedGen** – MedGen (www.ncbi.nlm.nih.gov/medgen/) is the NCBI's portal to information about human disorders and other phenotypes having a genetic component, including professional guidelines.
- **GeneReviews** – GeneReviews (www.ncbi.nlm.nih.gov/books/NBK1116/) provides overviews and disease-specific discussions of counseling, diagnosis, and management for individuals and families with inherited disorders.
- **ClinVar** – ClinVar (www.ncbi.nlm.nih.gov/clinvar/) is a freely accessible, public archive of reports of the relationships among human variations and phenotypes, with supporting evidence.
- **OMIM** – Online Mendelian Inheritance in Man (OMIM, www.ncbi.nlm.nih.gov/omim) provides a

compendium of human genes and genetic phenotypes.

- **NCI** – The National Cancer Institute (NCI) website (www.cancer.gov/cancertopics/pdq/cancerdatabase) presents physician data query (PDQ) summaries of information about specific cancers, cancer genetics, and an overview of clinical genetics based on clinical evidence.
- **PharmGKB** – Pharmacogenomics Knowledge Base (PharmGKB, www.pharmgkb.org/) is a comprehensive resource of curated knowledge about the impact of genetic variation on drug response.

SUMMARY AND RECOMMENDATIONS

- Incidental findings (IFs) are health-informative genetic variants that are identified unintentionally during genomic analysis or are identified through a scan for any genetic findings of interest. Detection of IFs is expected to increase with the expanded use of whole genome sequencing and whole exome sequencing. (See ['Definitions and classification of variants'](#) above.)
- The likelihood of detecting IFs depends on the type of genetic testing performed and the criteria used for identification and reporting. Actionable IFs (ie, IFs that might lead to a change in management) have been found in approximately 1 to 3 percent of individuals participating in genome sequencing studies. (See ['Likelihood of detecting an incidental finding'](#) above.)
- Patients undergoing genomic testing should be informed about the potential for IF information, including what kinds of IFs would be disclosed and how they would be communicated; this should occur as part of the consent process before genetic testing is initiated. Patients should also be provided an option to forego IF analysis of their genome (ie, an “opt out” option). The informed consent process may also involve a discussion of what to do with IFs after the patient’s death, which may help patients initiate discussions with relatives about whether they would want to receive postmortem disclosure of IFs, along with any other medical information that might be critical to their interpretation. (See ['Informed consent'](#) above and ['Postmortem disclosure'](#) above.)
- Each laboratory performing genomic testing makes decisions regarding which genes will be analyzed for IFs, what risk assessment to assign to each variant within the gene, and which results are reported to the clinician. The American College of Medical Genetics and Genomics (ACMG) has identified a list of 56 genes associated with 24 conditions that it considers a minimal set for which known or likely pathogenic variants should be reported ([table 1](#)). Numerous commissions have called for the refinement of professional standards for identifying and responding to genomic IFs. (See ['Decisions made by the laboratory'](#) above.)
- The decision about whether and when to disclose IFs will depend on the clinical context and the judgment of clinicians involved in the patient’s care. IFs should always be interpreted in the context of the patient’s medical and family histories of disease, physical examination, and other relevant information. Ethical obligations to respect patients’ autonomy and right not to know information must be weighed against fiduciary duties to inform patients of potentially life-saving information. A potential decision matrix regarding which IFs should be disclosed is presented in the table ([table 2](#)). (See ['Review patient report'](#) above.)
- Disclosure of IFs involves informing the patient that the IF(s) are present, counseling regarding the clinical implications of the IF(s), and discussing whether any additional intervention is needed. Importantly, the presence of a pathogenic IF is a marker of increased risk, **not** a diagnosis of disease. Conversely, a “negative” IF report does not imply that the patient is not at increased risk of developing disease(s). (See ['Disclose IFs to patient'](#) above and ['Interpretation of “negative” results'](#) above.)
- Technical limitations apply to identification and interpretation of IFs in individuals of

underrepresented ethnicities for whom data on disease prevalence and penetrance are lacking. (See '[Underrepresented ethnicities](#)' above.)

- Special ethical considerations apply to disclosure of incidental findings from genomic sequencing of children and research subjects. (See '[Children](#)' above and '[Research subjects](#)' above.)
- A variety of resources are available to assist clinicians, including policy statements, practice guidelines, and educational information. (See '[Online information and resources](#)' above.)

Use of UpToDate is subject to the [Subscription and License Agreement](#).

REFERENCES

1. Christenhusz GM, Devriendt K, Dierickx K. Disclosing incidental findings in genetics contexts: a review of the empirical ethical research. *Eur J Med Genet* 2013; 56:529.
2. Shkedi-Rafid S, Dheensa S, Crawford G, et al. Defining and managing incidental findings in genetic and genomic practice. *J Med Genet* 2014; 51:715.
3. Presidential Commission for the Study of Bioethical Issues. Anticipate and communicate: Ethical management of incidental and secondary findings in the clinical, research, and direct-to-consumer contexts, U.S. Department of Health & Human Services 2013. www.bioethics.gov/sites/default/files/FINALAnticipateCommunicate_PCSBI_0.pdf (Accessed on September 24, 2014).
4. Green RC, Rehm HL, Kohane IS. Clinical genome sequencing. In: *Genomic and Personalized Medicine*, 2nd Ed, Ginsburg GS, Willard HF (Eds) (Eds), Academic Press, San Diego 2013. Vol 1, p.102.
5. Green RC, Berg JS, Grody WW, et al. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet Med* 2013; 15:565.
6. Green RC, Lupski JR, Biesecker LG. Reporting genomic sequencing results to ordering clinicians: incidental, but not exceptional. *JAMA* 2013; 310:365.
7. American College of Medical Genetics and Genomics. Incidental findings in clinical genomics: a clarification. *Genet Med* 2013; 15:664.
8. Yang Y, Muzny DM, Xia F, et al. Molecular findings among patients referred for clinical whole-exome sequencing. *JAMA* 2014; 312:1870.
9. Green RC, Berg JS, Berry GT, et al. Exploring concordance and discordance for return of incidental findings from clinical sequencing. *Genet Med* 2012; 14:405.
10. Tabor HK, Auer PL, Jamal SM, et al. Pathogenic variants for Mendelian and complex traits in exomes of 6,517 European and African Americans: implications for the return of incidental results. *Am J Hum Genet* 2014; 95:183.
11. Xue Y, Chen Y, Ayub Q, et al. Deleterious- and disease-allele prevalence in healthy individuals: insights from current predictions, mutation databases, and population-scale resequencing. *Am J Hum Genet* 2012; 91:1022.
12. Lizarin GA, Haque IS, Nazareth S, et al. An empirical estimate of carrier frequencies for 400+ causal Mendelian variants: results from an ethnically diverse clinical sample of 23,453 individuals. *Genet Med* 2013; 15:178.
13. Bell CJ, Dinwiddie DL, Miller NA, et al. Carrier testing for severe childhood recessive diseases by next-generation sequencing. *Sci Transl Med* 2011; 3:65ra4.
14. Dorschner MO, Amendola LM, Turner EH, et al. Actionable, pathogenic incidental findings in 1,000 participants' exomes. *Am J Hum Genet* 2013; 93:631.
15. Johnston JJ, Rubinstein WS, Facio FM, et al. Secondary variants in individuals undergoing exome

sequencing: screening of 572 individuals identifies high-penetrance mutations in cancer-susceptibility genes. *Am J Hum Genet* 2012; 91:97.

16. Gold NB, Bick AG, McLaughlin HM, et al. Penetrance of actionable incidental genomic findings in exomes from the Framingham Heart Study. American Society of Human Genetics Annual Meeting. San Diego, 2014.
17. Reiff M, Ross K, Mulchandani S, et al. Physicians' perspectives on the uncertainties and implications of chromosomal microarray testing of children and families. *Clin Genet* 2013; 83:23.
18. Middleton A, Patch C, Wiggins J, et al. Position statement on opportunistic genomic screening from the Association of Genetic Nurses and Counsellors (UK and Ireland). *Eur J Hum Genet* 2014; 22:955.
19. Wolf SM. The past, present, and future of the debate over return of research results and incidental findings. *Genet Med* 2012; 14:355.
20. Ross LF, Rothstein MA, Clayton EW. Mandatory extended searches in all genome sequencing: "incidental findings," patient autonomy, and shared decision making. *JAMA* 2013; 310:367.
21. ACMG Board of Directors. Points to consider for informed consent for genome/exome sequencing. *Genet Med* 2013; 15:748.
22. Lee H, Deignan JL, Dorrani N, et al. Clinical exome sequencing for genetic identification of rare Mendelian disorders. *JAMA* 2014; 312:1880.
23. <http://nsgc.org/p/bl/et/blogid=45&blogaid=18> (Accessed on October 02, 2014).
24. Heald B, Edelman E, Eng C. Prospective comparison of family medical history with personal genome screening for risk assessment of common cancers. *Eur J Hum Genet* 2012; 20:547.
25. Hudson KL. Prohibiting genetic discrimination. *N Engl J Med* 2007; 356:2021.
26. Guttmacher AE, Porteous ME, McInerney JD. Educating health-care professionals about genetics and genomics. *Nat Rev Genet* 2007; 8:151.
27. Shields AE, Blumenthal D, Weiss KB, et al. Barriers to translating emerging genetic research on smoking into clinical practice. Perspectives of primary care physicians. *J Gen Intern Med* 2005; 20:131.
28. Secretary's Advisory Committee on Genetics, Health, and Society. Genetics Education and Training. February 2011. <http://www.genome.gov/Pages/Careers/HealthProfessionalEducation/SACGHS-EducationReport2011.pdf> (Accessed on October 30, 2014).
29. Allen NL, Karlson EW, Malspeis S, et al. Biobank participants' preferences for disclosure of genetic research results: perspectives from the OurGenes, OurHealth, OurCommunity project. *Mayo Clin Proc* 2014; 89:738.
30. Rothstein MA. Disclosing decedents' research results to relatives violates the HIPAA Privacy Rule. *Am J Bioeth* 2012; 12:16.
31. Hamamy H, Antonarakis SE, Cavalli-Sforza LL, et al. Consanguineous marriages, pearls and perils: Geneva International Consanguinity Workshop Report. *Genet Med* 2011; 13:841.
32. Sund KL, Zimmerman SL, Thomas C, et al. Regions of homozygosity identified by SNP microarray analysis aid in the diagnosis of autosomal recessive disease and incidentally detect parental blood relationships. *Genet Med* 2013; 15:70.
33. Rehder CW, David KL, Hirsch B, et al. American College of Medical Genetics and Genomics: standards and guidelines for documenting suspected consanguinity as an incidental finding of genomic testing. *Genet Med* 2013; 15:150.
34. https://www.acmg.net/docs/Incidental_Findings_in_Clinical_Genomics_A_Clarification.pdf (Accessed on October 03, 2014).
35. https://www.acmg.net/docs/ACMG_Releases_Highly-

Anticipated_Recommendations_on_Incidental_Findings_in_Clinical_Exome_and_Genome_Sequencing.pdf (Accessed on October 03, 2014).

36. Ng PC, Murray SS, Levy S, Venter JC. An agenda for personalized medicine. *Nature* 2009; 461:724.
37. Richards CS, Bale S, Bellissimo DB, et al. ACMG recommendations for standards for interpretation and reporting of sequence variations: Revisions 2007. *Genet Med* 2008; 10:294.
38. Rehm HL, Bale SJ, Bayrak-Toydemir P, et al. ACMG clinical laboratory standards for next-generation sequencing. *Genet Med* 2013; 15:733.
39. McGuire AL, Joffe S, Koenig BA, et al. Point-counterpoint. Ethics and genomic incidental findings. *Science* 2013; 340:1047.
40. ACMG Board of Directors. Points to consider in the clinical application of genomic sequencing. *Genet Med* 2012; 14:759.
41. Burke W, Antommaria AH, Bennett R, et al. Recommendations for returning genomic incidental findings? We need to talk! *Genet Med* 2013; 15:854.
42. Daack-Hirsch S, Driessnack M, Hanish A, et al. 'Information is information': a public perspective on incidental findings in clinical and research genome-based testing. *Clin Genet* 2013; 84:11.
43. Dal-Ré R, Katsanis N, Katsanis S, et al. Managing incidental genomic findings in clinical trials: fulfillment of the principle of justice. *PLoS Med* 2014; 11:e1001584.
44. Wolf SM, Annas GJ, Elias S. Point-counterpoint. Patient autonomy and incidental findings in clinical genomics. *Science* 2013; 340:1049.
45. Townsend A, Adam S, Birch PH, et al. "I want to know what's in Pandora's Box": comparing stakeholder perspectives on incidental findings in clinical whole genomic sequencing. *Am J Med Genet A* 2012; 158A:2519.
46. Ross LF, Rothstein MA, Clayton EW. Premature guidance about whole-genome sequencing. *Per Med* 2013; 10.
47. Wolf SM. Return of individual research results and incidental findings: facing the challenges of translational science. *Annu Rev Genomics Hum Genet* 2013; 14:557.
48. Delgado F, Tabor HK, Chow PM, et al. Single-nucleotide polymorphism arrays and unexpected consanguinity: considerations for clinicians when returning results to families. *Genet Med* 2014.
49. Krier JB, Green RC. Management of incidental findings in clinical genomic sequencing. *Curr Protoc Hum Genet* 2013; Chapter 9:Unit9.23.
50. Berg JS, Adams M, Nassar N, et al. An informatics approach to analyzing the incidentalome. *Genet Med* 2013; 15:36.
51. Berg JS, Khoury MJ, Evans JP. Deploying whole genome sequencing in clinical practice and public health: meeting the challenge one bin at a time. *Genet Med* 2011; 13:499.
52. McCormick JB, Sharp RR, Farrugia G, et al. Genomic medicine and incidental findings: balancing actionability and patient autonomy. *Mayo Clin Proc* 2014; 89:718.
53. Petty RE, Cacioppo JT. The elaboration likelihood model of persuasion. In: *Advances in Experimental Social Psychology*, Berkowitz L. (Ed), Academic Press, Orlando 1986. Vol 19, p.124.
54. Biesecker LG, Green RC. Diagnostic clinical genome and exome sequencing. *N Engl J Med* 2014; 370:2418.
55. https://www.acmg.net/ACMG/Terms_and_Conditions/ACMG/Terms_and_Conditions.aspx?redirect=https://www.acmg.net/ACMG/Find_Genetic_Services/ACMG/ISGweb/FindaGeneticService.aspx?hkey=720856ab-a827-42fb-a788-b618b15079f9 (Accessed on October 06, 2014).
56. <http://nsgc.org/p/cm/ld/fid=164> (Accessed on October 06, 2014).

57. Gottesman O, Kuivaniemi H, Tromp G, et al. The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future. *Genet Med* 2013; 15:761.
58. Kullo IJ, Haddad R, Prows CA, et al. Return of results in the genomic medicine projects of the eMERGE network. *Front Genet* 2014; 5:50.
59. Kannry JL, Williams MS. Integration of genomics into the electronic health record: mapping terra incognita. *Genet Med* 2013; 15:757.
60. Aronson SJ, Clark EH, Babb LJ, et al. The GeneInsight Suite: a platform to support laboratory and provider use of DNA-based genetic testing. *Hum Mutat* 2011; 32:532.
61. Wilcox AR, Neri PM, Volk LA, et al. A novel clinician interface to improve clinician access to up-to-date genetic results. *J Am Med Inform Assoc* 2014; 21:e117.
62. Singer E, Antonucci T, Van Hoewyk J. Racial and ethnic variations in knowledge and attitudes about genetic testing. *Genet Test* 2004; 8:31.
63. Guidelines for genetic testing of healthy children. *Paediatr Child Health* 2003; 8:42.
64. European Society of Human Genetics. Genetic testing in asymptomatic minors: Recommendations of the European Society of Human Genetics. *Eur J Hum Genet* 2009; 17:720.
65. Shalowitz DI, Miller FG. Disclosing individual results of clinical research: implications of respect for participants. *JAMA* 2005; 294:737.
66. Partridge AH, Winer EP. Sharing study results with trial participants: time for action. *J Clin Oncol* 2009; 27:838.
67. National Bioethics Advisory Commission. Research involving human biological materials: ethical Issues and policy guidance. Rockville, MD 1999. Vol 1. bioethicsarchive.georgetown.edu/nbac/hbm.pdf (Accessed on September 24, 2014).
68. Clayton EW, Ross LF. Implications of disclosing individual results of clinical research. *JAMA* 2006; 295:37; author reply 37.
69. Uhlmann WR, Guttmacher AE. Key Internet genetics resources for the clinician. *JAMA* 2008; 299:1356.
70. Riggs ER, Wain KE, Riethmaier D, et al. Towards a Universal Clinical Genomics Database: the 2012 International Standards for Cytogenomic Arrays Consortium Meeting. *Hum Mutat* 2013; 34:915.

Topic 96539 Version 3.0

GRAPHICS**American College of Medical Genetics list of genes for which incidental findings should be disclosed**

Disease category	Syndrome	Gene(s)
Cancer	Breast/ovarian cancer	BRCA1, BRCA2
	Li-Fraumeni syndrome, Peutz-Jeghers syndrome; PTEN hamartoma syndrome	TP53, STK11, PTEN
	Lynch syndrome, familial adenomatous polyposis, MYH-associated polyposis	MLH1, MSH2, MSH6, PMS2, APC, MUTYH
	Von Hippel Lindau syndrome; retinoblastoma, tuberous sclerosis, Wilms tumor	VHL, RB1, TSC1, TSC2, WT1
	Multiple endocrine neoplasia 1 or 2; familial medullary thyroid cancer	MEN1, RET
	Hereditary paraganglionoma-pheochromocytoma syndrome, neurofibromatosis type 2	SDHD, SDHAF2, SDHC, SDHB, NF2
Cardiovascular disease	Hypertrophic or dilated cardiomyopathy	MYBPC3, MYH7, TNNT2, TNNI3, TPM1, MYL3, ACTC1, PRKAG2, GLA, MYL2, LMNA
	Catecholaminergic polymorphic ventricular tachycardia, arrhythmogenic right ventricular cardiomyopathy, Romano-Ward Long QT syndromes, Brugada syndrome	RYR2, PKP2, DSP, DSC2, TMEM43, DSG2, KCNQ1, KCNH2, SCN5A
	Familial hypercholesterolemia	LDLR, APOB, PCSK9
Connective tissue or vascular integrity	Ehlers Danlos syndrome	COL3A1
	Marfan syndrome, Loeys-Dietz syndrome, familial thoracic aortic aneurysms and dissections	FBN1, TGFB1, TGFB2, SMAD3, ACTA2, MYLK, MYH11
Malignant hyperthermia sensitivity		RYR1, CACNA1S

This list includes genes identified by the American College of Medical Genetics (ACMG) as clinically actionable when known pathogenic (and in some cases, expected pathogenic) variants are identified by whole genome or exome sequencing. Refer to UpToDate topics on incidental findings, genomic testing, and genetic counseling for further details.

Data from: Green RC, Berg JS, Grody WW, et al. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. Genet Med 2013; 15:565.

Graphic 97842 Version 2.0

Proposed system for "binning" of incidental findings from genome sequencing

	Criteria:	Clinical utility	Clinical validity			Unknown clinical implications
Genes	Bins:	Bin 1 Medically actionable incidental information	Bin 2A Low risk incidental information	Bin 2B Medium risk incidental information	Bin 2C High risk incidental information	Bin 3 All other loci
	Examples:	<i>BRCA1/2</i> <i>MLH1, MSH2</i> <i>FBN1</i> <i>NF1</i>	PGx variants and common risk SNPs	<i>APOE</i> Carrier status for recessive Mendelian disorders	Huntington disease Familial Creutzfeldt-Jakob disease ALS (SOD1)	
	Estimated number of genes/loci:	10s	10s (eventually 100s-1000s)	1000s	10s	~20,000
Alleles that would be reportable (YES) or not reportable (NO) in a clinical context						
Variants	Known deleterious	YES	YES/NO*	YES/NO*	YES/NO*	N/A [¶]
	Presumed deleterious	YES	N/A ^Δ	YES/NO*	YES/NO*	NO [◇]
	VUS	NO	N/A ^Δ	NO	NO	NO [◇]
	Presumed benign	NO	N/A ^Δ	NO	NO	NO
	Known benign	NO	NO	NO	NO	NO

Refer to UpToDate topics on genomic sequencing and disclosure of incidental findings for further details.

N/A: not applicable; VUS: variant of uncertain significance.

* Reporting through decision making with an appropriate provider if elected by the patient.

¶ By definition, variants in genes with unknown implications could not be considered deleterious.

Δ By definition, SNPs or PGx variants will either be present or absent.

◇ Variants in genes with unknown clinical implications would not be reported; however, they may serve as an important substrate for research, potentially uncovering new disease genes.

Reprinted by permission from Macmillan Publishers Ltd: *Genetics in Medicine*. Berg JS, Khoury MJ, Evans JP. Deploying whole genome sequencing in clinical practice and public health:

Meeting the challenge one bin at a time. Genet Med 2011; 13:499. Copyright © 2011.

Graphic 97881 Version 1.0

Disclosures

Disclosures: Kurt Christensen, MPH, PhD Nothing to disclose. Sarah S Kalia, ScM, CGC Nothing to disclose. Robert Green, MD, MPH Nothing to disclose. Benjamin A Raby, MD, MPH Other Financial Interest (Spouse): Parexel International [Contract research organization (Double-blind randomized clinical trials)]. Jennifer S Tirnauer, MD Nothing to disclose.

Contributor disclosures are reviewed for conflicts of interest by the editorial group. When found, these are addressed by vetting through a multi-level review process, and through requirements for references to be provided to support the content.

Appropriately referenced content is required of all authors and must conform to UpToDate standards of evidence.

Conflict of interest policy

Principles and clinical applications of next-generation DNA sequencing

Author

Joseph V Thakuria, MD, MMSc

Section Editor

Benjamin A Raby, MD, MPH

Deputy Editor

Jennifer S Tirnauer, MD

All topics are updated as new evidence becomes available and our [peer review process](#) is complete.

Literature review current through: Feb 2015. | **This topic last updated:** Feb 25, 2015.

INTRODUCTION — Technologies for sequencing DNA have improved dramatically, to the point that it has become practical to sequence an individual's entire genome. Next-generation sequencing (NGS) is a type of DNA sequencing technology that uses parallel sequencing of multiple small fragments of DNA to determine sequence. This “high-throughput” technology has allowed a dramatic increase in the speed (and a decrease in the cost) at which an individual's genome can be sequenced.

The ability to sequence an entire genome raises several challenging questions for the clinician, including the following:

- When should NGS be considered clinically?
- What is the best choice among several types of genetic testing available?
- What is the clinical significance of findings from sequencing of an entire genome?
- Which findings should be acted upon and/or conveyed to the patient?

This topic will review basic concepts and clinical uses of next-generation DNA sequencing (NGS). Genetic testing, counseling, and reporting of incidental findings from genome sequencing are discussed separately. (See ["Genetic counseling and testing"](#) and ["Incidental findings from genetic testing"](#).)

Alternative methods for evaluating genetic and genomic disorders such as conventional sequencing, polymerase chain reaction, chromosomal analysis, and gene expression profiling are also presented separately. (See ["Overview of pharmacogenomics"](#) and ["Tools for genetics and genomics: Cytogenetics and molecular genetics"](#) and ["Tools for genetics and genomics: Gene expression profiling"](#) and ["Personalized medicine"](#).)

TERMINOLOGY AND EVOLUTION OF TECHNOLOGIES — A large investment has been made in improving DNA sequencing technologies, to make them cheaper, faster, and more accurate. The following terms are used to distinguish sequencing methodologies:

- **Sanger sequencing** — Manual or automated sequencing using the methods developed by Sanger, Maxam, and Gilbert is considered the “first” generation of DNA sequencing methods [1]. These types of sequencing are also commonly referred to as “conventional” or “traditional” sequencing. Sanger sequencing determines the sequence of large DNA fragments (up to approximately 500 to 900 bases), by collecting and aligning a series of different length products polymerized along the DNA template. The original Sanger method used radioactive markers for each nucleotide, and later adaptations have used fluorescently tagged versions.

Sanger sequencing is used clinically when the sequence of a specific gene is being tested. As an example, conventional sequencing could be used to identify a mutation in factor IX in a patient with suspected hemophilia B, without examining the rest of the individual's genome. Sanger sequencing is preferred in this setting because mutations in the gene encoding factor IX are already known to cause the disease; gene mutations can be correlated with laboratory markers of the disease (eg, activated partial thromboplastin time [aPTT]); and mutations in other genes are exceedingly unlikely to cause hemophilia B. (See ["Genetics of the hemophilias", section on 'Hemophilia B genes'](#).)

In contrast, Sanger sequencing often cannot provide information about large portions of the genome (eg, multiple genes) at a practical cost and within a reasonable timeframe. One estimate predicted that sequencing an entire human genome using Sanger sequencing would take 60 years [2].

- **Next-generation sequencing** — Next-generation sequencing (next-gen sequencing; NGS) uses sequencing of multiple DNA fragments, performed in parallel. This technology is also referred to by other terms (eg, high-throughput sequencing, deep sequencing, second-generation sequencing). In contrast to Sanger sequencing, the speed of sequencing and amounts of DNA sequence data generated with NGS are exponentially greater, and are produced at significantly reduced costs.

Several “platforms” (ie, sequencing instruments and associated reagents) for NGS have been developed. Across NGS platforms, there is typically a sample preparation or “library preparation” step in which the patient’s DNA, which serves as the template, is purified, amplified and fragmented, followed by physical isolation of DNA fragments by attachment to solid surfaces or small beads. Sequence data are generated on these small fragments, and the electronic results are computationally aligned against a “reference” genome or sequence (ie, a previously sequenced genome designated as a “normal” reference). An updated version of a commonly used human reference genome, GRCh38, was released by the National Center for Biotechnology Innovation (NCBI) on December 24, 2013.

NGS is reserved for clinical scenarios in which it is considered useful to determine the sequences of multiple genes. (See ['Clinical use of next-generation sequencing'](#) below.)

- **Third-generation sequencing** — Third-generation sequencing uses parallel sequencing similar to NGS, but unlike NGS, third-generation sequencing uses single DNA molecules rather than amplified DNA as a template. Thus, third-generation sequencing potentially eliminates errors in DNA sequence introduced in the laboratory during the DNA amplification process.

Third-generation methods are under development and generally are not clinically available.

Additional terminology used to describe a variety of genetic and genomic concepts is presented separately. (See ["Glossary of genetic terms"](#).)

In addition to evolving technical innovations, methods to optimize data interpretation continue to be studied, as large amounts of sequence data are becoming available before the clinical implications are known. (See ['Interpretation'](#) below.)

TECHNICAL CONSIDERATIONS

Source of DNA — The starting material that provides a template for clinical next-generation sequencing (NGS) is double-stranded nuclear DNA. This can be obtained from a variety of cell types. In some cases, the DNA is further modified in the laboratory (eg, to remove non-coding regions when exome sequencing is performed). (See ['Whole genome, exome, or gene panel'](#) below.)

DNA extracted from leukocytes in whole blood is a sterile source of DNA used for most clinical testing. When DNA is derived from non-sterile sources (eg, buccal swab, saliva sample), there is a potential risk for DNA sequence from bacterial flora to be confused for host sequence. However, non-sterile sources of DNA have the advantage of not requiring a blood draw, and less-invasive means for collecting DNA are increasingly used in large-scale genomic research studies. If needed, DNA can also be extracted from fixed tissues (eg, from formalin-fixed, paraffin-embedded pathology samples).

For patients who have undergone hematopoietic cell transplantation, leukocyte DNA from a blood sample is likely to contain donor rather than recipient genetic sequences. In this setting, DNA from a buccal swab, saliva, or hair follicles can be used [3,4]. Hair follicles are likely to be the most reliable

source, because buccal and saliva samples have been reported to show chimerism for donor and recipient sequences [5].

Starting material DNA must be amplified prior to NGS in order to run the sequencing reactions. Amplification involves exponential in vitro replication of patient DNA using PCR-based methods. While the DNA polymerase rarely makes mistakes (eg, substituting the wrong nucleotide), the amplification step has the potential to introduce errors in sequence not present in the original DNA from the patient.

NGS methods can also be used to sequence RNA and mitochondrial DNA; however, the majority of known human disease genes reside in nuclear DNA; hence these sources of material are generally reserved for research studies or diagnosing specific rare disorders.

For research purposes, it may be informative to compare genome sequences using normal and diseased tissue (eg, a tumor) from the same individual. Such comparisons can help determine the timing during which mutations are acquired during cancer development, allowing an estimation of their relative contributions to carcinogenesis.

Whole genome, exome, or gene panel — NGS can be used to sequence every nucleotide in an individual's DNA (ie, the whole genome), or limited to smaller portions of the genome such as the exome or a preselected subset of genes.

- **Whole genome sequencing** — Whole genome sequencing is costlier than more limited sequencing, because the whole genome is equivalent to approximately 3.3×10^9 bases (3.3 gigabases [Gb]). Whole genome sequencing may become preferable to exome sequencing as cost decreases and more information about the role of non-coding DNA in human disease becomes available.
- **Exome sequencing** — The exome contains the portions of genes that encode proteins; it represents only 1.5 to 2.0 percent of the genome (ie, about 30 megabases [Mb]). The remaining (non-exomic) DNA consists of introns and regulatory regions that control other aspects of gene function such as splicing and gene expression levels.

Exome sequencing is a reasonable approach for some clinical situations, because over 85 percent of known disease-causing mutations are found in exons. This approach substantially reduces cost and data storage requirements compared with whole genome sequencing. Exome sequencing also simplifies clinical reporting, because the significance of variants in exons is easier to interpret in most cases.

The main disadvantage of exome sequencing over whole genome sequencing is that exome sequencing could potentially miss a pathogenic variant(s) in a non-coding region of the genome. Thus, whole genome sequencing may be used in selected cases if initial exome sequencing is not diagnostic.

- **Targeted gene panels** — Gene panels provide sequence data for a limited subset of genes (typically 10 to 100 genes). Targeted gene panels are used in settings in which it would be appropriate to sequence many genes to make a diagnosis (eg, a Mendelian disorder for which the number of candidate genes is too large for traditional Sanger sequencing). As an example, targeted gene panels can be used to determine the genetic basis for hereditary early-onset and/or familial nonsyndromic hearing loss. Potential etiologies include mutations in over 60 genes; and screening for all of these individually would be impractical using traditional sequencing methods. The number of available panels continues to increase, with validated panels available for hereditary cardiomyopathies, lung diseases, ciliopathies, and other disorders in which molecular diagnosis is better facilitated by sequencing multiple genes known to be causative in the majority of cases.

Targeted gene panels may be preferable to exome sequencing due to the considerable cost advantage, and the lower likelihood of identifying variants of unknown significance that are unrelated to the disease being evaluated.

The utility of targeted gene panels over Sanger sequencing is also increasing as costs decline.

Accuracy — In general, one can expect usable data for over 92 to 95 percent of NGS sequence for whole genome and exome sequencing. Accuracy for targeted NGS gene panels is higher, since sequencing a smaller region of the genome allows for a greater degree of probe-template overlap (also called “probe tiling”). Sanger sequencing remains the “gold standard” for diagnosis based on gene sequencing, with >99.99 percent accuracy reported for most genes sequenced.

Clinical laboratories (ie, those with Clinical Laboratory Improvement Amendments [CLIA] certification) generally perform Sanger sequencing to confirm any variant reported back to the ordering clinician as pathogenic, because of the greater accuracy of Sanger sequencing [6,7]. (See '[Interpretation](#)' below.)

Additional methods are used to limit the number of variants prioritized for validation and reporting, including the use of stringent bioinformatic filters and targeted sequencing in affected family members.

Interpretation — Clinical interpretation of data from NGS is more challenging than interpretation of traditional sequencing, because NGS provides data regarding “variants” in multiple genes, many of which are unexpected, outside the gene(s) of interest, and of unknown prognostic significance.

Results of NGS are generally reported as one of the following:

- “Pathogenic” — Pathogenic variants are variants previously reported in patients with disease and/or are strongly suspected of being pathogenic based on preclinical studies.
- “Likely Pathogenic” — Likely pathogenic variants are those with sequence features that are likely to be implicated in disease pathogenesis but for which conclusive evidence of pathogenicity is not available.
- “Likely benign” — Likely benign variants are those for which weak data in the medical literature supporting pathogenicity may be available, but for which the majority of evidence suggests the effect of the variant is benign.
- “Benign” — Benign findings are genetic variants not predicted to alter gene expression or function.
- “Unknown clinical significance” — Variants of unknown significance (VUS; also called “variant of uncertain significance” or “finding of unknown clinical significance”) are variants that have some features suggestive of possible functional consequence, but for which there is insufficient evidence for either a pathogenic or benign role.

The clinician and patient must be aware that the likelihood of receiving a result of “uncertain significance” is reasonably high, because the clinical implications of many variants are unknown. As an example, a review on exome sequencing noted that an average sequence could identify over three million variants, of which the clinical implications might be uncharacterized for approximately 15 to 20 percent [8]. Thus clinicians ordering NGS testing and providing results of NGS to patients should receive proper training in how to discuss these issues. (See '[Genetic counseling and testing](#)', section on '[Content of genetic counseling](#)' and '[Incidental findings from genetic testing](#)'.)

Clinical NGS laboratories generally will be in agreement on their interpretation of a variant as “pathogenic” or “benign”. In contrast, significant laboratory-to-laboratory variability exists for the “likely pathogenic,” “likely benign,” and “variant of unknown significance” categories. Consistency across laboratories continues to improve as more data (especially clinical outcomes data) become available for any given variant.

The challenges in interpreting NGS data are illustrated by following reports:

- A whole genome sequence analysis of an otherwise healthy individual with Charcot-Marie-Tooth (CMT) disease, in addition to facilitating the identification of the CMT-causing mutations, identified 159 mutations with known disease (or trait) associations, including 33 variants associated with cancers, 48 associated with complex diseases, and 21 associated with Mendelian (monogenic) diseases [9]. In the absence of significant family history for most of these conditions, the interpretation and management of these variants are unclear. However, longitudinal follow-up of the patient may be helpful as new data become available.
- A whole genome sequence analysis in a healthy 40-year-old man with a family history of coronary artery disease and sudden death in one of 27 relatives identified numerous variants associated with both rare and common conditions, including three mutations previously associated with sudden cardiac death [10]. Predictive models were applied for 32 conditions, with probability estimates suggesting substantially increased risk for the development of obesity, myocardial infarction (probability increased from 2.0 percent pre-testing to 8.9 percent post-testing), type 2 diabetes (probability increased from 27 to 54 percent), prostate cancer (probability increased from 16 to 23 percent), and depression, and a reduced risk for Alzheimer's disease (probability reduced from 9 to 1 percent). Based upon these results, in conjunction with the patient's strong family history of atherosclerosis, the patient's physician advised initiating lipid-lowering therapies despite a lipid profile for which treatment would not otherwise be recommended by established guidelines. Given the anecdotal nature of this report and the lack of prospective data, it is unclear whether any gains were achieved from the availability of the sequence data.

These cases illustrate the importance of discussing the likelihood of identifying unsuspected variants or variants of unknown significance prior to ordering testing, and the importance of having genetic counselors and other experienced subspecialists available to guide the patient once such variants are discovered. (See "[Genetic counseling and testing](#)", [section on 'Consent for genome sequencing'](#) and "[Incidental findings from genetic testing](#)".)

CLINICAL USE OF NEXT-GENERATION SEQUENCING — Next-generation sequencing (NGS) is not always the most appropriate clinical genetic test. It is expensive, time-consuming, and often unnecessary for diagnosing genetic conditions for which the clinical evaluation has limited potential candidates to one or a few genes amenable to Sanger sequencing or other more traditional methods of detecting genetic defects. However, it is appropriate to consider exome sequencing or targeted NGS gene panels when a large number of pathogenic genes need to be screened [11]. Similarly, exome sequencing or whole genome sequencing should be considered when a condition demonstrates high heritability in a family or is suspected to have a genetic basis, but the number of potential candidate genes is large, or responsible gene(s) are unknown.

Indications for next-generation sequencing — Consideration of NGS as a clinical tool (eg, for genetic diagnosis) is appropriate in individuals for whom sequencing of a single gene is unlikely to provide a diagnosis.

Examples include suspected genetic disorders in the following settings:

- One of many potential genes may be responsible, and/or the clinician does not know which gene(s) to test because many different genes cause the same phenotype (eg, due to genetic heterogeneity).
- Obvious candidate genes have been tested and were found to be normal. This is especially applicable when the percentage of disease attributed to these candidate genes is low, and other potentially causative genes for the disorder are thought to exist but have not yet been identified. Such analyses are often aided by comparison of NGS results from affected and unaffected family

members.

- It would be less costly and more efficient to sequence the entire genome, exome, or gene panel than to sequence individual candidate genes sequentially.

One of the most common medical indications for whole genome sequencing or whole exome sequencing is evaluation of severe intellectual disability or developmental delay believed to have a genetic etiology in a patient with a negative initial evaluation. In some cases, evaluation of an affected child and both parents (“trio sequencing”) is performed – especially when the inheritance pattern is dominant and a de novo mutation is suspected [12,13]. The value of NGS in this setting has been illustrated in several studies, in which the likelihood of reaching a molecular diagnosis is on the order of 25 percent [12-16].

As examples:

- Exome sequencing was performed for 2000 individuals born with severe neurologic deficits unexplained by prior clinical evaluation; slightly fewer than half were younger than 5 years of age [14,15]. Molecular diagnosis was established in 504 of the patients (25 percent). Diagnosis was more likely in those with neurologic findings rather than anomalies of other organ systems; and the implicated genetic variant was more likely to be previously undescribed rather than a known variant. Almost one-third were due to genetic causes of disease discovered within the prior one to two years.
- Exome sequencing was performed in 814 patients (half younger than 5 years of age) with a variety of unexplained syndromes, the most common of which were developmental delay in children and ataxia in adults [12]. Molecular diagnosis was established in 213 (26 percent). Diagnosis was more likely in patients who had trio sequencing rather than proband sequencing alone; patients younger than 5 years of age; and patients with retinal disorders (for which a larger fraction of disease genes may be known). Examples were presented in which identification of a variant previously not associated with disease coincided with publication of a case report demonstrating the association.
- Whole genome sequencing was used to evaluate 50 children with severe intellectual disability who did not have a diagnosis after extensive genetic testing that included exome sequencing; a genetic diagnosis was made in 20 of these children (40 percent) [16]. This cohort was likely to have been enriched for de novo cases, since none of the children had a positive family history for intellectual disability.

Consequently, when an extensive evaluation including chromosomal microarray is negative for developmental delay with a suspected genetic etiology, NGS may be appropriate.

NGS is also being incorporated into the National Institutes of Health “[Undiagnosed diseases program](#)”, which evaluates patients who have a longstanding medical condition that eludes diagnosis [17]. In some cases, molecular diagnoses may have been challenging prior to the availability of NGS if the phenotype was nonspecific (ie, intellectual disability without distinguishing syndromic features) and the candidate gene list too large for traditional Sanger sequencing to be practical. NGS can be used to diagnose genetic diseases if the mode of Mendelian inheritance is known and familial samples are also sequenced for comparison.

In addition to diagnosing genetic disorders, NGS might be helpful in identifying an infectious pathogen when usual microbial or serologic testing is unrevealing. This was demonstrated in a case report in which a 14-year-old boy developed unexplained fever and meningoencephalitis following travel to Puerto Rico; the boy had an underlying immunodeficiency syndrome [18]. An extensive infectious disease evaluation was unrevealing, and the patient’s clinical status deteriorated. NGS identified a species of *Leptospira*, a pathogenic spirochetal organism acquired by exposure to contaminated water or soil, in the patient’s cerebrospinal fluid (CSF) but not serum. Antibiotic therapy was initiated based on the NGS

results, with clinical improvement. Subsequent testing confirmed the diagnosis.

NGS is also being offered to healthy individuals to determine increased disease risks, pharmacogenomic variants, and nonmedical information (eg, ancestry). However, this practice is limited due to lack of prospective clinical data and prohibitive costs of universally offering genomic screening. Research efforts are underway to determine the number of healthy people needed to screen with NGS before clinical benefit is seen.

In contrast to clinical indications, potential uses of NGS as a research tool are extensive. Examples include identification of new genes involved in unexplained syndromes, and determination of genetic changes during the development of acute myeloid leukemia (AML) and other cancers [19]. These findings may ultimately lead to breakthroughs in understanding disease pathogenesis, and to the development of additional diagnostic testing and/or management strategies for these diseases. (See ["Molecular genetics of acute myeloid leukemia"](#).)

Genetic discrimination — A common concern about genetic testing is the potential for inadequate protection of privacy of genetic information and associated impacts on employment and insurance coverage. This issue is discussed in detail separately. (See ["Genetic counseling and testing"](#), section on ["Genetic discrimination"](#).)

Disclosure of findings from genome sequencing — Genome sequencing may lead to incidental findings with potential clinical importance. Issues of whether and how to disclose such findings, and which findings to report to the individual being tested, as well as recommendations from the American College of Medical Genetics (ACMG) regarding which results should be disclosed, are discussed in detail separately. (See ["Incidental findings from genetic testing"](#).)

Historically, individual genomic reports provided directly from companies to consumers often included many variants unlikely to be clinically pathogenic. Because a genotyping chip rather than NGS platform was generally used, most variants were located in non-coding regions of DNA and occurred at high frequency in the general population; these variants were unlikely to have any significant predictive value. Often such variants were reported at higher frequencies than the lifetime disease risk observed in the population, further discounting their likelihood of clinical significance. Issues related to direct-to-consumer genetic testing are discussed in detail separately. (See ["Personalized medicine"](#), section on ["Direct-to-consumer testing"](#).)

PRACTICAL ISSUES

Where to order — Several university and commercial laboratories in the United States offer next-generation sequencing (NGS). Most clinical NGS testing uses targeted NGS gene panels and exome sequencing. An exception is clinical whole genome sequencing offered through Illumina.

Cost and turnaround time — The typical cost of clinical NGS is several thousand dollars. Factors that affect cost include the type of genome analyzed (eg, somatic tumor plus germline); whether a whole genome, exome, or gene panel is requested; the requested turnaround time (standard versus expedited); whether a single affected proband or additional family member(s) are tested; and the extent of sequence interpretation.

Standard turnaround time for NGS is generally 8 to 12 weeks, but may be longer for more involved, complex cases.

Insurance reimbursement — Insurance reimbursement for NGS is a complex issue. In a study that reported on findings of whole-exome sequencing for 250 patients (mostly children) with neurologic phenotypes found that insurance coverage was similar to other genetic testing (eg, reimbursement for the majority of tests) [14]. Of note, the individuals in this study had relatively major neurologic phenotypes, and all had undergone prior genetic testing, often extensive.

SUMMARY

- Next-generation sequencing (NGS) is a type of DNA sequencing technology that uses parallel sequencing of multiple small fragments of DNA to determine sequence. This technology has allowed a dramatic increase in the speed (and a decrease in the cost) at which an individual's genome can be sequenced. (See ['Terminology and evolution of technologies'](#) above.)
- NGS can be performed on double stranded DNA from a variety of sources. DNA extracted from leukocytes in whole blood is used for most clinical testing. (See ['Source of DNA'](#) above.)
- NGS can be used to sequence an individual's whole genome or smaller portions of the genome such as the exome or a preselected subset of genes. (See ['Whole genome, exome, or gene panel'](#) above.)
- Sanger sequencing is used to confirm the presence of specific mutations identified by NGS in clinical settings, because of the greater accuracy of traditional sequencing methods. The types of variants confirmed (eg, known pathogenic versus variant of unknown significance) differ by laboratory. In clinical settings, only variants that have been validated by Sanger sequencing are reported. (See ['Accuracy'](#) above and ['Interpretation'](#) above.)
- NGS may be appropriate for diagnosing suspected genetic disorders when sequencing of a single gene is unlikely to provide a diagnosis. Examples include the following settings (see ['Indications for next-generation sequencing'](#) above):
 - One of several potential genes may be responsible
 - Obvious candidate genes have been tested and found to be normal
 - The cost of NGS would be less than that of sequencing individual candidate genes sequentially
- Concerns regarding consent for NGS testing, genetic discrimination, and disclosure of incidental findings from NGS are discussed separately. (See ["Genetic counseling and testing"](#) and ["Incidental findings from genetic testing"](#).)
- Several university and commercial laboratories in the United States offer NGS. The cost is several thousand dollars and depends on the type and amount of the genome analyzed. Standard turnaround time is generally 8 to 12 weeks. Insurance reimbursement may be similar to other genetic testing in individuals with major clinical phenotypes for whom initial testing is unrevealing. (See ['Practical issues'](#) above.)

Use of UpToDate is subject to the [Subscription and License Agreement](#).

REFERENCES

1. Rizzo JM, Buck MJ. Key principles and clinical applications of "next-generation" DNA sequencing. *Cancer Prev Res (Phila)* 2012; 5:887.
2. Bennett ST, Barnes C, Cox A, et al. Toward the 1,000 dollars human genome. *Pharmacogenomics* 2005; 6:373.
3. Hong YC, Liu HM, Chen PS, et al. Hair follicle: a reliable source of recipient origin after allogeneic hematopoietic stem cell transplantation. *Bone Marrow Transplant* 2007; 40:871.
4. Thiede C, Prange-Krex G, Freiberg-Richter J, et al. Buccal swabs but not mouthwash samples can be used to obtain pretransplant DNA fingerprints from recipients of allogeneic bone marrow transplants. *Bone Marrow Transplant* 2000; 25:575.

5. Rennert H, Leonard DG, Cushing M, et al. Avoiding pitfalls in bone marrow engraftment analysis: a case study highlighting the weakness of using buccal cells for determining a patient's constitutional genotype after hematopoietic stem cell transplantation. *Cytotherapy* 2013; 15:391.
6. Rehm HL, Bale SJ, Bayrak-Toydemir P, et al. ACMG clinical laboratory standards for next-generation sequencing. *Genet Med* 2013; 15:733.
7. Strom SP, Lee H, Das K, et al. Assessing the necessity of confirmatory testing for exome-sequencing results in a clinical molecular diagnostic laboratory. *Genet Med* 2014; 16:510.
8. Bamshad MJ, Ng SB, Bigham AW, et al. Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet* 2011; 12:745.
9. Lupski JR, Reid JG, Gonzaga-Jauregui C, et al. Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. *N Engl J Med* 2010; 362:1181.
10. Ashley EA, Butte AJ, Wheeler MT, et al. Clinical assessment incorporating a personal genome. *Lancet* 2010; 375:1525.
11. Biesecker LG, Green RC. Diagnostic clinical genome and exome sequencing. *N Engl J Med* 2014; 370:2418.
12. Lee H, Deignan JL, Dorrani N, et al. Clinical exome sequencing for genetic identification of rare Mendelian disorders. *JAMA* 2014; 312:1880.
13. Wright CF, Fitzgerald TW, Jones WD, et al. Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. *Lancet* 2014.
14. Yang Y, Muzny DM, Reid JG, et al. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N Engl J Med* 2013; 369:1502.
15. Yang Y, Muzny DM, Xia F, et al. Molecular findings among patients referred for clinical whole-exome sequencing. *JAMA* 2014; 312:1870.
16. Gilissen C, Hehir-Kwa JY, Thung DT, et al. Genome sequencing identifies major causes of severe intellectual disability. *Nature* 2014; 511:344.
17. Kuehn BM. NIH's Undiagnosed Diseases Program expands: 6 new sites offer potential answers to more patients. *JAMA* 2014; 312:587.
18. Wilson MR, Naccache SN, Samayoa E, et al. Actionable diagnosis of neuroleptospirosis by next-generation sequencing. *N Engl J Med* 2014; 370:2408.
19. Waddell N, Pajic M, Patch AM, et al. Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature* 2015; 518:495.

Topic 17039 Version 9.0

Disclosures

Disclosures: **Joseph V Thakuria, MD, MMSc** Nothing to disclose. **Benjamin A Raby, MD, MPH** Other Financial Interest (Spouse): Parexel International [Contract research organization (Double-blind randomized clinical trials)]. **Jennifer S Tirnauer, MD** Nothing to disclose.

Contributor disclosures are reviewed for conflicts of interest by the editorial group. When found, these are addressed by vetting through a multi-level review process, and through requirements for references to be provided to support the content. Appropriately referenced content is required of all authors and must conform to UpToDate standards of evidence.

Conflict of interest policy

Tools for genetics and genomics: Gene expression profiling**Authors**

Avrum Spira, MD, MSc
Katrina Steiling, MD, MSc

Section Editor

Benjamin A Raby, MD, MPH

Deputy Editor

Jennifer S Tirnauer, MD

All topics are updated as new evidence becomes available and our [peer review process](#) is complete.

Literature review current through: Feb 2015. | **This topic last updated:** Nov 13, 2013.

INTRODUCTION — The genetic basis for disease is determined by the inheritance of genes containing specific sequences of DNA. The phenotypic expression of these genes, through the synthesis of specific proteins, involves interaction with environmental signals that trigger activation of particular genes.

According to the central dogma of biology, RNA is transcribed from a DNA template; messenger RNA (mRNA) is then translated into protein ([figure 1](#)). Transcription and translation underlie gene expression.

Approximately 3 to 5 percent of genes are active in a particular cell, even though all cells have the same information contained in their DNA. Most of the genome is selectively repressed, a property that is governed by the regulation of gene expression, mostly at the level of transcription (ie, the production of messenger RNA from the DNA). In response to a cellular perturbation, changes in gene expression take place that result in the expression of hundreds of gene products and the suppression of others. This molecular heterogeneity can affect when and how a disease presents clinically in an individual with genetic predisposition to a condition and how individuals with a given disease will respond to specific treatments.

Analyses of gene expression can be clinically useful for disease classification, diagnosis, prognosis, and tailoring treatment to underlying genetic determinants of pharmacologic response.

This topic will focus on the role of mRNA in the cell, platforms for profiling mRNA expression, the challenges in interpreting the data from these analyses, and the emerging clinical applications of gene expression measurements. An overview of molecular genetics and the role of gene expression profiling in clinical oncology are presented separately. (See "[Principles of molecular genetics](#)" and "[Overview of gene expression profiling, proteomics, and microRNA profiling in clinical oncology](#)".)

RNA IN CELL FUNCTION — There are several classes of RNA. Messenger RNA (mRNA) is the RNA that is translated into protein. Noncoding RNAs are not translated into protein and serve other functions in the cell. Classes of noncoding RNAs include transfer RNAs (tRNAs), ribosomal RNAs (rRNAs), small nuclear RNAs (snRNAs), small nucleolar RNAs (snoRNAs), microRNAs, Piwi-interacting RNAs (piRNAs), and large intergenic noncoding RNAs (lincRNAs) [[1-5](#)].

Messenger RNA accounts for about 1 percent of the total RNA in a cell [[6](#)] and is transcribed from approximately 20,000 to 25,000 protein-coding genes in the human genome [[7](#)]. After mRNA is transcribed from DNA, it typically undergoes further modifications including the addition of a methyl-guanosine cap (5' cap), the addition of a series of adenines to the 3' end of RNA (poly-A tail), and the splicing out of introns ([figure 2](#)) [[1](#)]. mRNA is then transported from the nucleus to the cytoplasm where it is translated into protein. Messenger RNA serves as a transient intermediate between DNA and protein, and is degraded in minutes to hours [[1](#)].

MEASURING GENE EXPRESSION — Since mRNA represents the functional bridge between DNA and protein, alterations in mRNA may serve as markers for the activation or inhibition of a particular gene.

The challenge in measuring RNA relates to the susceptibility of RNAs to degradation by ribonucleases

(RNases). Methods of RNA detection take advantage of the single stranded structure of RNA and its complementarity to the DNA from which it was transcribed.

Expression profiling of single genes or small gene panels — Prior to the development of microarray technology, available methods of measuring gene expression included:

- Northern blot (see '[Northern blot](#)' below)
- Ribonuclease protection assay (see '[Ribonuclease protection assay](#)' below)
- In-situ hybridization (see '[In-situ hybridization](#)' below)
- Reverse-transcription quantitative polymerase chain reaction (RT-PCR) (see '[Real-time reverse transcription polymerase chain reaction](#)' below)
- Spotted cDNA arrays (see '[Spotted cDNA arrays](#)' below)

A comparison of these methods is summarized in a table ([table 1](#)).

Northern blot — Northern blots allow the determination of both the presence of an RNA molecule and its size [8]. RNA molecules from a patient sample are first separated based on size using gel electrophoresis. The size-separated RNA molecules are then transferred and cross-linked to a nylon membrane. The RNA of interest is detected by incubating the membrane with a labeled single-stranded DNA probe that is complimentary to this RNA. Probe bound to the RNA of interest can then be detected using autoradiography or chemiluminescence.

Ribonuclease protection assay — Whereas the Northern blot uses complimentary DNA probes, the ribonuclease protection assay (RPA) uses antisense RNA probes [6]. These so-called riboprobes are single-stranded radiolabeled RNA molecules complimentary to the RNA of interest. The riboprobe is incubated with the RNA from the sample and binds to complementary RNA to form double-stranded RNA complexes. Incubating the mixture with ribonucleases degrades unbound single-stranded RNA from both the sample and excess probe. The remaining double stranded RNA complexes are size-separated by electrophoresis, and detected by autoradiography.

In-situ hybridization — In-situ hybridization (ISH) uses a nucleic acid probe to detect any other nucleic acids in a tissue section. ISH can localize the RNA of interest at the anatomic or cellular level. The tissue section is fixed to preserve tissue morphology and nucleic acid integrity [9,10]. The sample is then treated with proteases to eliminate proteins bound to the RNA of interest [9,10]. A labeled probe is hybridized to the sample and detected using autoradiography or chemiluminescence [10]. In-situ hybridization using a chemiluminescence-labeled probe is also called fluorescence in situ hybridization (FISH). The use of FISH to detect gene mutations is discussed separately. (See "[Tools for genetics and genomics: Cytogenetics and molecular genetics](#)", section on '[Fluorescence in situ hybridization](#)'.)

Real-time reverse transcription polymerase chain reaction — Real-time reverse transcription polymerase chain reaction (RT-PCR) is a relatively simple approach that can be used to assay small or large numbers of genes from a single sample [11]. After isolating RNA from a sample, complementary DNAs (cDNAs) are synthesized by reverse transcription with an RNA-dependent DNA polymerase. This cDNA mixture is then combined with a DNA-dependent DNA polymerase and fluorescently-labeled oligonucleotide primers [12]. These primers are short sequences of nucleotides complementary to a portion of the cDNA and allow amplification. Fluorescence increases as the cDNA of interest is amplified with PCR. The fluorescence intensity is monitored and the total number of PCR cycles is counted [6].

The point at which the PCR cycle can distinguish fluorescence related to gene amplification from background is the cycle threshold, and this number can be used to estimate the relative starting quantity of the RNA of interest [11]. Careful primer selection is required to prevent amplification of related genes [6]. (See "[Tools for genetics and genomics: Polymerase chain reaction](#)".)

Spotted cDNA arrays — Unlike Northern blotting, RPA, or ISH, spotted cDNA arrays are capable of testing the relative expression levels, between two conditions, of several hundred genes. With increased knowledge of which sequences are expressed from the genome, it became possible to create cDNA probes targeting the expressed DNA sequences from which RNA is transcribed.

cDNA probes are amplified using PCR and spotted onto a glass slide [13]. RNA is then isolated from two samples representing different conditions. mRNA from each sample is isolated, and labeled with one of two fluorescent dyes (green or red) [14]. The samples are then mixed together and co-hybridized to the cDNA probes on the glass slide [13]. This approach directly compares gene expression in the first condition to the second condition, and allows the detection of as many genes as there are probes on the array. However, reproducibility is limited across arrays because of the need to manually spot probes on slides.

Genome-wide gene expression profiling — Platforms for profiling gene expression take advantage of increased knowledge of the sequence of the human genome and require smaller quantities of starting RNA. Current platforms for profiling gene expression include:

- Oligonucleotide arrays (microarrays)
- Transcriptome sequencing

A comparison of these methods is summarized in a table ([table 1](#)) .

Oligonucleotide arrays (microarrays) — Oligonucleotide arrays operate on a similar principle to spotted cDNA arrays, but differ in how they are produced. Rather than spotting probes onto a glass slide, the short probes are synthesized directly on the slide [15,16]. Depending on the commercial manufacturer, probes vary from ~20 to 60 base pairs in length. Several types of arrays are commercially available.

Sample preparation begins with the isolation of RNA from the tissue of interest, resulting in an extraction that contains all of the genes transcribed in the tissue at the time the RNA is isolated. The RNA is then reverse-transcribed into cDNA and amplified using polymerase chain reaction (PCR) technique. Finally, a [biotin](#) label is incorporated through an in vitro transcription process, which converts cDNA into labeled cRNA.

A single sample of the labeled cRNA is applied to each array. Hybridization occurs between the labeled cRNA from the sample and complementary probes on the array. This is followed by binding to an avidin-conjugated fluorophore, and a washing step that removes any unbound material. The fluorophore is excited by a laser scanner coupled to a computer that captures the image fluorophores linked to hybridized target molecules on the array, thus enabling the detection of the expression of thousands of genes simultaneously.

In general, the greater the amount of mRNA from a particular gene (ie, the higher the expression of that gene), the more fluorescently-labeled material corresponding to that gene will bind to complementary probes on the array. Background fluorescence or nonspecific binding may limit detection of lowly expressed transcripts. Probe-based detection for gene expression limits analysis to genes that are known.

Transcriptome sequencing — An alternative for measuring gene expression is the direct sequencing and quantification of RNA molecules. This method of measuring gene expression has also been referred to as “RNA-seq”, “massively parallel sequencing”, “next generation sequencing”, or “deep sequencing”, and several commercial platforms are available. The details of each system vary. In general, the sample is prepared so that many sequencing reactions occur simultaneously on a glass slide, and yields millions of RNA sequence reads obtained by laser scanning [17].

This technique allows improved detection of low abundance transcripts, as well as detection of novel

transcripts and polymorphisms within a transcript's sequence.

Microarray analysis and interpretation — Microarrays can assay large numbers of genes with relatively high throughput. They are currently the platform most often used for gene expression profiling of clinical specimens. Typically, investigators involved with microarray experiments are interested in comparing gene expression across different conditions [18]. While there are many approaches to the data analysis in order to accomplish this goal, there are generally several analytical steps that must first be taken (figure 3).

There are four general considerations in approaching microarray data analysis and interpretation:

- Normalization, quality assessment, and preprocessing
- Data storage and analysis
- Multiple comparison problem
- Biological interpretation

Normalization, quality assessment, and preprocessing — Normalization adjusts measured fluorescent intensities so that they are comparable across different experiments. Quality assessment eliminates low-quality microarrays or measurements. Preprocessing prepares the data for statistical analysis.

Normalization — Each microarray can be considered a separate experiment that contains slightly different amounts of starting RNA and different labeling efficiencies [18]. Data normalization adjusts the fluorescent intensities representing the amount of RNA bound to each probe so that these intensities are comparable across different arrays. There are several methods for normalizing microarray data, including:

- Scaling — Adjusts intensities by a constant factor so that the average expression level across microarrays is similar.
- Quantile normalization — Adjusts the distribution of intensities across microarrays. This is accomplished by ranking the probe intensities from highest to lowest for each array. A numerical value is assigned to represent this intensity on an individual array based on the behavior across all arrays and the rank of that probe on the individual array.
- Locally weighted scatterplot smoothing (LOWESS) — Adjusts the brightness or darkness of different fluorescent labels for two-color array experiments.

Quality assessment — Quality assessment occurs both before and after data normalization.

- Pre-normalization quality assessment evaluates the microarray itself to ensure there are no bubbles, scratches, or other artifacts on the array. Some commercial arrays also contain controls inserted during sample processing ("spike-in" controls) to ensure that all steps leading to the hybridization were successful.
- Post-normalization quality assessment evaluates a microarray relative to all others in the experiment. This helps to identify outlier samples or differences in batches of microarrays. Microarrays identified as significantly different from others can be adjusted statistically or excluded from the analysis.

Preprocessing — Many statistical procedures assume a normal distribution of data. Gene expression levels from microarrays can be mathematically transformed, often using a logarithmic scale, so that they become normally distributed. Preprocessing can also include filtering out low quality probe sets or genes with low variability across all samples in the experiment.

Data storage and analysis — Microarray experiments require computational tools to store raw data, analyze gene expression, and ensure uniformity across different laboratories.

Data storage — The fluorescent intensities generated by scanning an oligonucleotide array with a laser scanner results in an image file. Most scientific journals specify that raw data be made publicly available as a requirement for publication [19]. A typical microarray raw data file is 0.1-1GB per array, generating a large amount of data that must be stored [17]. Beyond the storage of raw data, data files also include the clinical variables associated with each sample and the normalized, quality assessed preprocessed expression levels for each array. This is often accomplished with the use of a database capable of storing and integrating both gene expression data and clinical variables.

Data analysis — There are several possible levels of data analysis, ranging from simple statistical tests that can be performed with commercial software packages, to advanced analyses and the development of novel algorithms. Advanced analyses and novel algorithms are implemented with a variety of programming languages, such as Perl and Python, and computational software, such as R [20] and Matlab [21]. The flexibility to write, modify and share algorithms using these tools makes them particularly well suited for microarray data analysis.

- **Differential expression** — One of the most common analyses performed on microarray data is to determine which genes are altered in one condition as compared to another. This can be accomplished by performing a t-test, ANOVA, or linear model.
- **Class prediction** — In this type of analysis, samples from two conditions are split into a training set and a test set. A list of genes that distinguishes the two conditions is derived from the training set of samples, and the accuracy of this gene expression signature is assessed on the test set of samples.
- **Class discovery** — Microarray data can be used to explore novel molecular phenotypes. By taking the most variable genes across all samples regardless of their clinical phenotype, it can be determined which samples are most closely correlated with each other based on gene expression alone. Samples that share similar patterns of gene expression may represent previously unrecognized subtypes of the disease.
- **Network analysis** — The number of genes assayed by microarrays allows the entire dataset to be harnessed to make new predictions about how genes might interact. One approach, the Context Likelihood Relatedness (CLR) algorithm, uses a mathematical representation of the amount of information that one gene's expression gives about any other gene's expression [22]. This can then be represented as a network of predicted interactions between genes.

The multiple comparison problem — Statistical analyses of several thousand genes pose unique problems in the interpretation of the statistical results because of the large number of tests performed.

If a large number of statistical tests are performed, it is more likely that an individual test represents a false positive finding. In general, statistical tests ask whether two groups are similar (null hypothesis) or different (alternative hypothesis). Statistical tests estimate the certainty about choosing the alternative hypothesis over the null hypothesis. By incorrectly concluding that there is a difference between the groups where there is no actual difference, a Type I error has occurred and the result is falsely positive. The probability of making this error is estimated by the alpha level, or p-value. For a p value of 0.05, there is a 1 in 20 chance of making a Type I error. If one statistical test is performed for each of 10,000 genes from a microarray using a p-value of 0.05, there will be 500 Type I errors. In most circumstances, this is an unacceptably high number of false positives.

Several methods exist to manage the multiple comparison problem.

- **Bonferroni correction** — This strict correction lowers the probability of making a Type I error. The

desired alpha level is divided by the total number of tests to yield the p-value required for an individual test to reach statistical significance. For a family of 10,000 tests and a desired Type I error rate of 0.05, the adjusted p-value for an individual test to reach significance is $0.05/10,000$, or 5×10^{-6} . While this method limits the number of false positives, it also makes real effects more difficult to detect.

- **Benjamini-Hochberg false discovery rate** — An alternative approach is to control the expected number of false positives [23]. The false discovery rate (FDR) is the rate at which genes called significant are actually not significant [24]. If 5000 genes are significant at certain p-value, and 500 false positives are expected at this p-value, the FDR is 0.1.
- **Q-value** — The third approach is an extension of the FDR approach [24]. The q-value measures significance based upon the FDR [24]. Whereas each gene's p-value indicates a certainty of a true difference expression, each gene's q-value indicates the fraction of false positives at this level. Hence, a q-value of 0.05 denotes that among genes with a q-value ≤ 0.05 , all but 5 percent are likely true positive results. In this way, one can tolerate a small proportion of false positive results without having to ignore the much larger proportion (in this case, 95 percent) of true positive results.

Biological interpretation — The final step in microarray data analysis is to interpret the results in a biologically meaningful context. Making biological sense of a microarray-derived gene list is one of the more challenging aspects of the analysis. While there are many strategies for accomplishing this goal, two broad approaches are discussed below. Additional studies are often required to validate biological predictions that are made from the microarray data.

Comparison with other microarray datasets — In one study in 2005, RNA was distributed to four laboratories for analysis by 12 different microarray platforms [25]. Good reproducibility was found within a laboratory, but there was decreased reproducibility across laboratories [25]. Reproducibility improved with the use of standardized protocols for RNA labeling, hybridization, microarray processing, data acquisition and data normalization.

Several tools exist for comparing gene expression datasets, including large databases containing gene expression data to look for a shared gene expression signature [26], alternative gene probes [27], and analytic tools that incorporate phenotypic associations [28-30]

Enrichment ranking — Gene set enrichment analysis (GSEA) is a method by which gene expression data is ranked by association with phenotypes, and is used as a means to identify biologically-relevant pathways [28,29]. Other techniques provide other mechanisms to enrich pathways or functional categories [31] or to visualize previously published interactions between genes of interest [32]. Data visualization using heat maps, which organize samples by columns and genes by rows according to similarity in gene expression, are also useful for determining which groups of genes or samples share similar patterns of expression.

OVERVIEW OF CLINICAL APPLICATIONS — Gene expression profiling within clinical specimens has the potential to be used for disease screening, diagnosis, prognostication, and optimizing treatment regimens. As the platforms for measuring gene expression continue to evolve, personalized approaches to the diagnosis and treatment of complex human disease will increasingly find their place in routine clinical practice.

The promise of using gene expression profiling to identify subjects at risk for disease has yet to be fully reached. However, in certain circumstances this tool has been incorporated into routine clinical evaluation. As an example, the Oncotype Dx assay is used to guide the evaluation and management of subsets of patients with breast cancer. (See "[Prognostic and predictive factors in early, non-metastatic breast cancer](#)", section on 'Gene expression profiles' and "[Prognostic and predictive factors in metastatic](#)

[breast cancer", section on 'Next generation sequencing of breast tumors'.\)](#)

The use of gene expression profiling in clinical oncology is discussed in detail separately. (See ["Overview of gene expression profiling, proteomics, and microRNA profiling in clinical oncology".\)](#)

Diagnosis — Two examples of the use of gene expression profiling to target selected patients for more invasive testing follow:

- Current protocol calls for routine endomyocardial biopsy surveillance for cardiac transplant patients to detect the presence of acute cellular rejection. A study of gene expression profiling of RNA from peripheral blood mononuclear cells in heart transplant recipients revealed that an 11 gene expression profile could predict the presence of rejection [33]. Use of this gene expression profile has the potential to decrease the number of transplant patients who would need to undergo invasive myocardial biopsy to confirm the diagnosis of rejection. (See ["Acute cardiac allograft rejection: Diagnosis", section on 'Gene expression profiling'.\)](#)
- Smokers suspected of having lung cancer on the basis of an abnormal chest CT scan often need to undergo invasive diagnostic procedures, beyond bronchoscopy, to achieve a final diagnosis. An 80 gene expression signature resulting from gene expression profiling of histologically normal airway epithelium obtained during bronchoscopy was capable of distinguishing between smokers with and without lung cancer [34]. Combining results of this gene expression biomarker with clinical variables, in an integrated clinico-genomic model, improved the discriminatory potential to predict lung cancer [34]. Gene expression profiling may help to stratify patients with an abnormal chest CT scan who should undergo invasive diagnostic testing for a potential lung cancer and those for whom less intensive monitoring would be appropriate. (See ["Screening for lung cancer".\)](#)

Prognosis — The most advanced application of gene expression profiling is in predicting outcome from disease. The risk of certain therapies might be outweighed by the potential benefit for patients at high risk for relapse or with a poor prognosis, whereas the risks might outweigh the benefits for patients with a relatively good prognosis. Gene expression profiling has been helpful in targeting appropriate therapy for patients with acute leukemia, prostate cancer, colon cancer, breast cancer, lung carcinoma, and lymphoma. (See ["Overview of gene expression profiling, proteomics, and microRNA profiling in clinical oncology", section on 'Clinical applications of DNA microarrays'.\)](#)

SUMMARY

- There are a variety of methods for measuring RNA to evaluate gene expression. These methods differ in their requirements for the amount of starting RNA, their sensitivity to detect the RNA of interest, and the computational requirements for data analysis. (See ["Measuring gene expression"](#) above.)
- While oligonucleotide array profiling is the commonly used biomarker discovery platform, there are numerous challenges and pitfalls in the analysis and interpretation of the large volume of data generated. (See ["Microarray analysis and interpretation"](#) above.)
- Gene expression profiling is rapidly emerging as a novel approach for the diagnosis and prognosis of complex human disease. However, a number of important barriers remain, including validation of these biomarkers in prospective multi-center studies to demonstrate their reproducibility and accuracy across multiple sites and operators. (See ["Overview of clinical applications"](#) above.)

Use of UpToDate is subject to the [Subscription and License Agreement](#).

REFERENCES

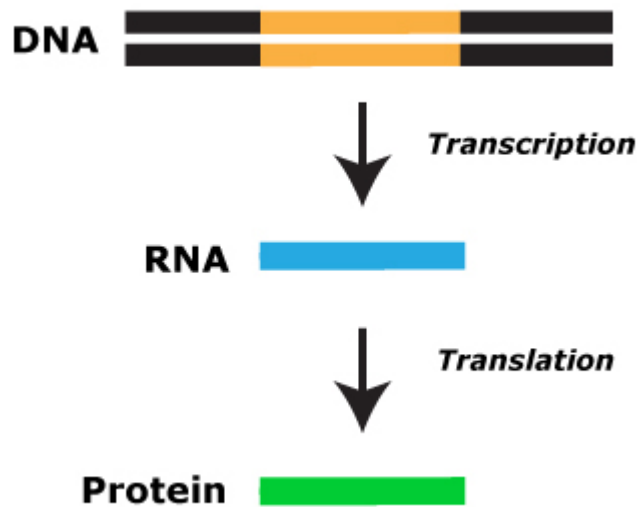
1. Brown TA. Genomes 3, 3rd ed, Garland Science, 2007.
2. Lee RC, Feinbaum RL, Ambros V. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* 1993; 75:843.
3. Girard A, Sachidanandam R, Hannon GJ, Carmell MA. A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* 2006; 442:199.
4. Aravin A, Gaidatzis D, Pfeffer S, et al. A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* 2006; 442:203.
5. Khalil AM, Guttman M, Huarte M, et al. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci U S A* 2009; 106:11667.
6. Dvorák Z, Pascussi JM, Modrianský M. Approaches to messenger RNA detection - comparison of methods. *Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub* 2003; 147:131.
7. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* 2004; 431:931.
8. Alwine JC, Kemp DJ, Stark GR. Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. *Proc Natl Acad Sci U S A* 1977; 74:5350.
9. Gall JG, Pardue ML. Formation and detection of RNA-DNA hybrid molecules in cytological preparations. *Proc Natl Acad Sci U S A* 1969; 63:378.
10. Jin L, Lloyd RV. In situ hybridization: methods and applications. *J Clin Lab Anal* 1997; 11:2.
11. Nolan T, Hands RE, Bustin SA. Quantification of mRNA using real-time RT-PCR. *Nat Protoc* 2006; 1:1559.
12. VanGuilder HD, Vrana KE, Freeman WM. Twenty-five years of quantitative PCR for gene expression analysis. *Biotechniques* 2008; 44:619.
13. Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 1995; 270:467.
14. Churchill GA. Fundamentals of experimental design for cDNA microarrays. *Nat Genet* 2002; 32 Suppl:490.
15. Pease AC, Solas D, Sullivan EJ, et al. Light-generated oligonucleotide arrays for rapid DNA sequence analysis. *Proc Natl Acad Sci U S A* 1994; 91:5022.
16. Nuwaysir EF, Huang W, Albert TJ, et al. Gene expression analysis using oligonucleotide arrays produced by maskless photolithography. *Genome Res* 2002; 12:1749.
17. Wilhelm BT, Landry JR. RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing. *Methods* 2009; 48:249.
18. Quackenbush J. Microarray data normalization and transformation. *Nat Genet* 2002; 32 Suppl:496.
19. Brazma A, Hingamp P, Quackenbush J, et al. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet* 2001; 29:365.
20. R Development Core Team. R: A Language and Environment for Statistical Computing 2009. R Foundation for Statistical Computing. Available at: www.R-project.org (Accessed on December 14, 2009).
21. The Mathworks I. The MathWorks - MATLAB and Simulink for Technical Computing 2009. Available at: www.mathworks.com (Accessed on December 14, 2009).
22. Faith JJ, Hayete B, Thaden JT, et al. Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles. *PLoS Biol* 2007; 5:e8.
23. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to

- multiple testing. *J R Statist Soc* 1995; 57:289.
24. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* 2003; 100:9440.
 25. Bammler T, Beyer RP, Bhattacharya S, et al. Standardizing global gene expression analysis between laboratories and across platforms. *Nat Methods* 2005; 2:351.
 26. Barrett T, Troup DB, Wilhite SE, et al. NCBI GEO: archive for high-throughput functional genomic data. *Nucleic Acids Res* 2009; 37:D885.
 27. Dai M, Wang P, Boyd AD, et al. Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. *Nucleic Acids Res* 2005; 33:e175.
 28. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005; 102:15545.
 29. Mootha VK, Lindgren CM, Eriksson KF, et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* 2003; 34:267.
 30. Lamb J, Crawford ED, Peck D, et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 2006; 313:1929.
 31. Dennis G Jr, Sherman BT, Hosack DA, et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* 2003; 4:P3.
 32. Ingenuity Systems. Ingenuity Pathway Analysis Software 2009. Available at: www.ingenuity.com (Accessed on December 14, 2009).
 33. Deng MC, Eisen HJ, Mehra MR, et al. Noninvasive discrimination of rejection in cardiac allograft recipients using gene expression profiling. *Am J Transplant* 2006; 6:150.
 34. Beane J, Sebastiani P, Whitfield TH, et al. A prediction model for lung cancer diagnosis that integrates genomic and clinical features. *Cancer Prev Res (Phila)* 2008; 1:56.

Topic 14602 Version 10.0

GRAPHICS

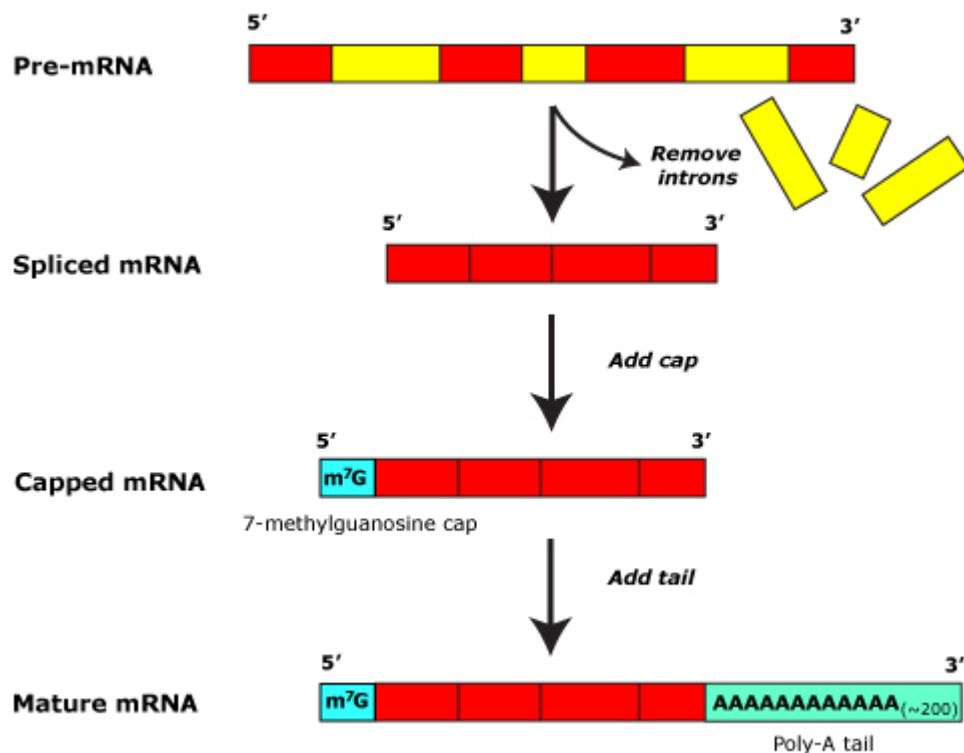
Central dogma of molecular biology



DNA codes for protein through the intermediary of RNA. For the longest time, it was believed that this flow was unidirectional. It is now clear that this is not entirely true. Nevertheless, this model frames the basics of our understanding of molecular biology, and while the study of genetics focuses on variation that originates in DNA, we must always be mindful that this variation has its action by eventually interfering with protein function.

Graphic 54914 Version 1.0

Post-transcription modification of mRNA



Pre-mRNA transcripts are unsuitable for protein translation. Introns must be removed through a process called splicing to align the exons for their subsequent translation. The spliced RNA is an unstable molecule, with highly reactive moieties at both the 5' and 3' ends. Addition of the 5' cap and the 3' poly-A tail are critical for stabilization of RNA and for their subsequent metabolism.

Graphic 70605 Version 1.0

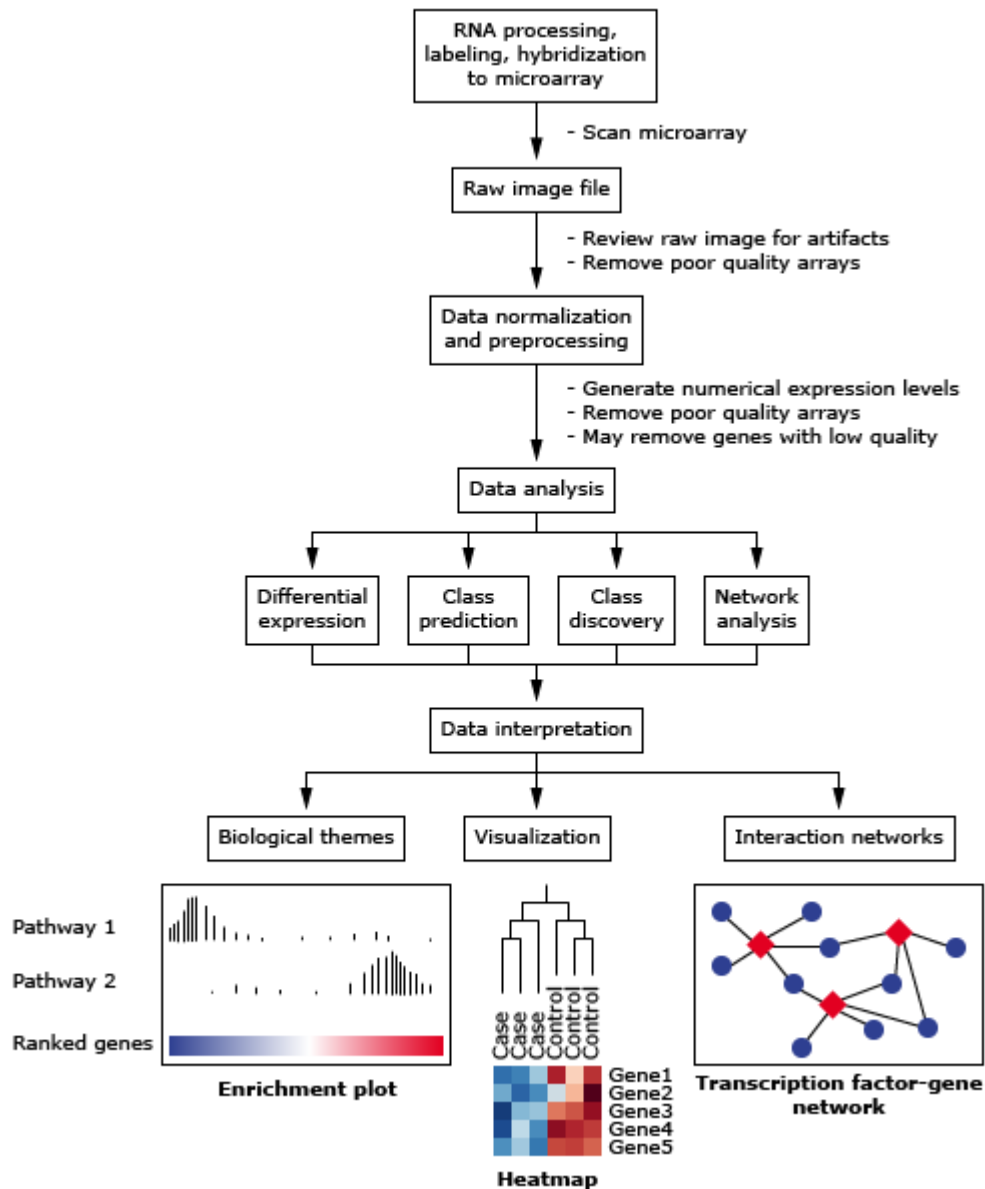
Comparison of methods to quantify gene expression

	Amount of starting RNA required	Sensitivity of test	Analytic requirements	Information obtained from the test	Number of targets
Northern blot	High	Low	Low	Presence and size of RNA of interest	Few
Ribonuclease protection assay	High	Moderate	Low	Presence and size of RNA of interest	Few
In-situ hybridization	High	Low to moderate	Low	Presence and spatial localization of RNA of interest	Few
Spotted cDNA arrays	Moderate	Moderate to high	Moderate to high	Relative gene expression between two conditions	Several hundred
Oligonucleotide arrays	Low to moderate	Moderate to high	Moderate to high	Relative gene expression	Thousands
RT-PCR	Low	High	Low to moderate	Relative gene expression	Few to several hundred
Transcriptome sequencing	Low to moderate	High	Very high	Digital counts of each RNA	Unlimited

molecule (as
measure of
gene
expression) with
excellent
dynamic range

Graphic 76106 Version 2.0

Analysis strategy for oligonucleotide arrays (microarrays)



Graphic 71360 Version 3.0

Disclosures

Disclosures: **Avrum Spira, MD, MSc** Consultant/Advisory Boards: Veracyte [gene expression diagnostics (diagnostic test)]. Equity Ownership/Stock Options: Veracyte [gene expression diagnostics (diagnostic test)]. **Katrina Steiling, MD, MSc** Patent Holder: Boston University [COPD gene expression (co-inventor on patent regarding airway gene expression in COPD)]. **Benjamin A Raby, MD, MPH** Other Financial Interest (Spouse): Parexel International [Contract research organization (Double-blind randomized clinical trials)]. **Jennifer S Tirnauer, MD** Nothing to disclose.

Contributor disclosures are reviewed for conflicts of interest by the editorial group. When found, these are addressed by vetting through a multi-level review process, and through requirements for references to be provided to support the content. Appropriately referenced content is required of all authors and must conform to UpToDate standards of evidence.

Conflict of interest policy

Overview of gene expression profiling, proteomics, and microRNA profiling in clinical oncology

Author

Patrick C Ma, MD, MSc

Section Editors

Benjamin A Raby, MD, MPH
Reed E Drews, MD

Deputy Editor

Diane MF Savarese, MD

All topics are updated as new evidence becomes available and our [peer review process](#) is complete.

Literature review current through: Feb 2015. | **This topic last updated:** Mar 18, 2015.

INTRODUCTION — Important goals of cancer research include the discovery of novel cellular targets to exploit for novel targeted treatments, new biomarkers for early cancer detection, and to provide a better classification of cancers for prognostication and treatment selection.

Toward this end, a significant effort has been devoted to understanding the molecular basis of carcinogenesis and the biologic behavior of human cancers. Progress has been slow, reflecting the complexity of the molecular alterations that characterize tumor cells. Carcinogenesis is a multistep process involving genetic and epigenetic events that result in altered expression of numerous genes [1]. Confounding this complexity, many of the so-called oncogenes and tumor suppressor genes are signaling molecules, which control the expression of a subset of downstream genes.

Cells respond to environmental signals by modulating the expression of genes contained within the nucleus. When genes are activated, they are transcribed to generate messenger RNA (mRNA), which is transported from the nucleus to the cytoplasm and translated into protein by the ribosomes.

Approximately 3 to 5 percent of genes are active in a particular cell, even though all cells have the same information contained in their DNA. Most of the genome is selectively repressed, a property that is governed by the regulation of gene expression, mostly at the level of transcription (ie, the production of messenger RNA from the DNA). In response to a cellular perturbation, changes in gene expression take place that result in the expression of hundreds of gene products and the suppression of others. This molecular heterogeneity is thought to underlie, at least in part, the variability in outcome and response to therapy that characterizes tumors of different histology.

Significant variability also exists for tumors of a specific histologic type. In general, clinical management decisions and prognostic estimates are based solely upon histopathologic analysis of tumor tissue. However, tumor behavior cannot be adequately understood through the analysis of one or a small numbers of genes, particularly for the common solid tumors.

The examination of multiple expressed genes and/or proteins provides more useful information for both classification and prognostication of individual tumors. The development of microarray methodology, which permits the expression of thousands of genes to be assayed simultaneously, represents a powerful technique to read the “molecular signature” of an individual patient's tumor. This process is termed gene expression profiling (GEP). Analyzing gene expression patterns across individual patients with the “same” disease may reveal molecular differences. Such classification may allow better treatment selection and prognostication. Proteomics is a slightly different but related approach, in which the expressed proteins of a tissue or cell type are assayed rather than the genes themselves.

A description of the methods used for GEP and proteomics, as well as a brief overview of some potential clinical applications, will be presented here. A brief description of the new development of microRNA (miRNA) profiling in cancer diagnosis and prognosis will also be included. More detailed discussion of the applications of these techniques to specific tumor types is discussed in the appropriate tumor topics.

GEP AND DNA MICROARRAY ANALYSIS — The traditional method of measuring the level of expression of a single gene is by assaying RNA by Northern blot analysis. DNA microarray analysis uses the same principles to simultaneously measure the expression level of thousands of genes on a platform called a microarray. (See ["Tools for genetics and genomics: Cytogenetics and molecular genetics".](#))

There are several stages to performing a microarray analysis [2]:

- Preparation of the microarray
- Generation of fluorescent targets from the RNA of the samples
- Hybridization to the probes
- Data acquisition: scanning of the signal intensity emanating from the hybridized labeled probes
- Data analysis: the extraction of biologically useful information from the vast quantity of data that is generated from microarray analysis. This aspect is often the most challenging component of GEP.

Preparation of the microarray — The DNA array consists of an orderly arrangement of DNA spots on a glass slide or chip. In its simplest form, a few dozen complementary DNAs (cDNAs) or oligonucleotides corresponding to particular genes are immobilized onto the substrate in a known order within the grid.

In highly-sophisticated microarray "chips", up to hundreds of thousands of unique oligonucleotide probes, representing thousands of known genes or expressed sequence tags (ESTs), are synthesized in a microgrid on a glass substrate about the size of a thumbnail. ESTs are segments of expressed genes that have been sequenced, but do not correspond to known genes.

Each oligonucleotide probe, which is specific for a particular gene, is located on a precise place within the microgrid; this is the probe cell. Each probe cell is very small, about 24 microns by 24 microns, and contains millions of copies of each specific oligonucleotide. A particular gene (eg, the gene encoding thymidylate synthetase) may be represented on the microgrid by 20 or more probe cells, called a probe set. The oligonucleotide probes (usually 15 to 25 nucleotides in length) in each probe cell of the probe set may differ from each other, some corresponding to the 5' end of the mRNA sequence, some the middle, and some the 3' end. This gives the sample RNA a broad range of sequences with which to hybridize.

Various sources of DNA may be used as probes:

- In a cDNA microarray the DNA spots are composed of DNA sequences (cDNA clones, or ESTs) that are extracted from plasmid-containing bacterial cultures; they correspond in sequence to specific genes derived from the UniGene public database. These polymerase chain reaction (PCR)-amplified ESTs or whole cDNA clones, usually 500 to 2000 base pairs (bp) in length, are deposited on the microarray surface using high-speed robotics.
- For an oligonucleotide (oligo) array, oligonucleotide sequences, typically 15 to 25 bp in length, that are unique to the genes of interest, are synthesized in situ on the chip itself (one popular example is the Affymetrix GeneChip). As with cDNA arrays, the information needed to construct the oligo probes is derived from published databases.

Although the basic principles underlying cDNA and oligo microarrays are similar, there are two significant differences. One is the amount of RNA that is required from patient samples for each experiment. In general, cDNA microarray experiments require more than 50 micrograms of total RNA from target tissues, while 5 micrograms may be sufficient for an oligo array experiment.

The second difference has to do with the way control or reference samples are handled. In a cDNA

microarray, a control sample (eg, normal as compared to diseased tissue) is simultaneously analyzed along with the test sample on the same chip. Gene expression is estimated by comparing the RNA content of the test and control samples, and the results are then reported as a ratio. By contrast, in the oligo-based approach, a test sample on one oligo chip is compared to a reference or control sample on another chip, and interpretation requires comparing the levels from the test sample to those of the control or reference sample.

Target preparation — To obtain the test samples, messenger RNA (mRNA) is extracted from the patient's tissue. This mRNA represents all the genes expressed in that tissue at that time. The mRNA is then converted to cDNA using reverse transcriptase and deoxyribonucleotides and is labeled with fluorescent dyes. The labeled nucleic acids (in some cases cDNA and in some cases RNA) derived from the sample are called targets.

The process of generating the targets can be illustrated by the methods used to compare the GEP of tumor and normal cells for a cDNA microarray experiment. Initially, total RNA or mRNA is extracted from the tumor tissue, and reverse transcribed into cDNA using reverse transcriptase in the presence of fluorescently labeled deoxynucleotides (typically using the fluorochrome Cy3 [rhodamine] which emits red light upon excitation at a wavelength of 565 nm). This results in the generation of a cDNA into which the fluorescently labeled deoxynucleotides are incorporated.

Upon completion of the reaction, all the RNA in the sample will have been converted into fluorescently labeled cDNA. The greater the abundance of a particular mRNA sequence, the greater the amount of labeled cDNA synthesized during reverse transcription.

In a parallel reaction, an equivalent amount of RNA isolated from normal cells is also reverse transcribed, but differentially labeled using a different fluorochrome (often Cy5 [[fluorescein](#)] which emits green light upon excitation at a wavelength of 670 nm).

Differential labeling is not necessary for oligo arrays. Instead, the sample mRNA is first reverse transcribed into cDNA in the absence of radiolabeled nucleotides. This single-stranded DNA is then converted to double-stranded cDNA, which is in turn transcribed into cRNA in the presence of biotinylated nucleotides. This additional step amplifies the target mRNA approximately 50-fold. The fluorescent tag is added after hybridization by the use of an avidin-conjugated fluorophore (see below).

Hybridization — The labeled sample or target is then placed onto the array and hybridization based on complementary base pairing is allowed to take place. The more mRNAs in the original sample (ie, the higher the expression of a particular gene), the more fluorescently-labeled material corresponding to that gene is in the target and the more labeled material will be hybridized at that gene's location on the array.

For the cDNA array, the differentially labeled test and reference cDNA samples are mixed in equal amounts, and incubated with the microarray slide until hybridization is complete. The labeled sequences of the target mixture hybridize to the homologous sequence of the gene (ie, the probe) on the array. The amount of RNA in the original sample dictates the extent of binding to a specific probe. Because the test and normal samples are differentially labeled, the relative amount of binding in the test sample relative to control is determined by detecting the degree and type of fluorescence (ie, red for the tumor tissue and green for the control).

For oligo chip array experiments, the biotinylated cRNA is hybridized to the oligo probes on the chip, and followed by binding to an avidin-conjugated fluorophore.

In the final microarray (both cDNA and oligo), each cell or spot on the microarray chip (sometimes termed a "feature") represents the quantification of gene expression for a given gene in a given sample. On the microarray, the individual samples are shown in one direction, while the specific genes are shown in the other. As an example, a microarray is shown in the figure in which each row (ie, from left to right)

represents a patient sample, while each column (ie, top to bottom) represents a specific gene ([figure 1](#)).

Data acquisition — Using fluorescent detectors coupled to computers equipped with specialized software, the amount of gene expression for thousands of genes and ESTs can be estimated simultaneously, generating a gene expression profile (GEP) for that sample.

Once hybridization is complete, the arrays are washed, and a laser beam is used to excite the fluorophores that are linked to the hybridized target molecules.

cDNA microarrays — For cDNA microarrays, the degree and type of fluorescence is assessed using a laser scanner that is capable of detecting and quantifying the amount of light emitted in the green and the red wavelength, then converting this to a digital image.

Four possible scenarios are outlined below:

- If a gene is underexpressed in tumor, many more RNA copies of this gene will be present in normal compared to tumor tissue, and the spot that corresponds to this gene on the microarray will appear green.
- In contrast, if the gene is expressed at a higher level in tumor tissue, more mRNA will be present in the tumor compared to normal tissue, and the spot on the microarray that corresponds to this gene will appear red.
- If genes are similarly expressed in normal and tumor tissue, competition between the two probes will result in an equal amount of hybridization with red and green-labeled probes, and the spot will appear yellow on the microarray.
- If a gene is expressed in neither tumor nor normal tissue, there will be no hybridization to this spot on the array, and the spot will appear black.

This information can be quantified by calculating the ratio of both signal intensities. Thus, if the level of gene expression is similar in tumor and normal tissue, the ratio is 1. In contrast, genes that are more highly expressed in the tumor (upregulated) have a ratio greater than 1, while downregulated genes have a ratio that is less than 1. The absolute level of gene expression in test as compared to control samples is of importance in selecting genes of interest. Genes that are overexpressed or underexpressed to a level threefold or greater are more likely to be of interest than those whose expression differs by a lesser amount.

Expression levels of each gene can be compared across many tissue or tumor samples. In this way, groups or clusters of genes can be identified whose expression varies in similar ways. The most common mathematical algorithm used to compare GEPs is aggregative hierarchical clustering, which is described in detail below.

Oligo arrays — For oligo arrays, the abundance of a target molecule on a certain spot is estimated by scanning the intensity of emitted fluorescence of the linked reported molecules. Raw data from two microarrays (one from diseased tissue and another from normal or control tissue) must then be compared to identify genes that are differentially expressed.

Data analysis — The purpose of analysis of microarray data is to extract “biologic order” based upon similarities and differences in gene expression. Data normalization is necessary as the first step for both cDNA and oligo arrays. The signal intensity must be adjusted to a common standard to correct for differences in background noise or the efficiency of data acquisition.

After normalization, the raw GEP levels are presented as expression ratios (test versus control sample) that are typically log-transformed to simplify the presentation of differences (eg, threefold increase in expression, twofold decrease in expression of the test sample relative to control).

Defining the purpose of the GEP experiment is critical. In oncology, microarray experiments have focused on the following objectives:

- **Gene discovery** — Identification of candidate genes that are differentially expressed in normal as compared to tumor tissue, or between different tumor samples is a major goal. The identification of these genes may help to define molecular pathogenesis, pathways of drug resistance, or new potential targets for therapy (see below).

Discovery of such candidate genes may be facilitated by comparative genomic hybridization techniques, in which chromosomal regions, which may harbor tumor suppressors and/or oncogenes, may be scanned for the presence of significant genetic deletions or amplifications [3].

- **Molecular classification** — The use of GEP profiles of known disease classes to accurately predict the biologic group, diagnostic category, or prognostic classification of an unknown tumor sample is termed **class prediction**. In contrast, **class discovery** refers to the identification of new molecular disease classes that may be relevant to future disease classification, prognostication, and/or treatment selection. Because disease classes are not predefined, these studies aim to determine whether discrete, previously unrecognized molecular subsets of a disease entity can be identified based solely upon the GEP.

Selecting genes of interest — When comparing the GEP of samples to identify genes that are differentially expressed, candidate genes may be selected by several methods. An arbitrary, fold-based difference in expression (eg, >2-, 5-, or 10-fold difference in spot intensity), or a statistical test, such as the t-test (for two samples) or F-test (for three or more samples) can be used to select those genes that are significantly differentially expressed among patients or experiments [4,5].

These genes then undergo further analysis to confirm overexpression, typically by real-time polymerase chain reaction (PCR) techniques [2]. The next step is to group or cluster the genes of interest based upon similarity of expression within a given tissue or tumor in order to identify genes that may be coregulated, and therefore, functionally related [6].

The clustering of similarly expressed genes may generate a pattern or profile that separates tumors of distinct phenotypes, classes or disease stage. These clusters are identified by mathematical algorithms.

The organization of the sample according to clusters can be performed in one of two ways:

- **Unsupervised**, ie, not based upon any prior information regarding the analyzed objects. Unsupervised analysis utilizes information provided by all genes that are expressed in the tissue.
- **Supervised**, ie, based upon prior known information, and a preselected group of genes.

Both approaches can be applied to the same data set for different purposes. However, the supervised method is a better method to prospectively identify tumors that belong to a predefined clinically important subgroup (ie, class prediction).

Unsupervised clustering — In unsupervised cluster analysis, the statistical algorithm is not trained to recognize a specific gene expression pattern from any previously known class or grouping of patients [2]. Unsupervised clustering is often hierarchical (ie, a classification with nested classes resembling a phylogenetic tree), or it may be nonhierarchical (classification into clusters without specifying the relationship between individual members of a class). In turn, hierarchical clustering may be agglomerative (starting from single members and their relationship with one another, and growing into bigger classes) or divisive (a technique that starts by grouping all members into one class, and subsequently breaking the class into smaller groups).

Hierarchical agglomerative clustering is the most commonly used method in gene expression analysis. It begins by identifying a single gene pair with the most similar expression across patients or experiments.

These genes are then "fused" (subsequently represented by the mean of their two expression levels); this value is then considered a single element in further analyses. The process of "fusing" is repeated until all elements are included in the analysis, and a dendrogram is assembled ([figure 2](#)). The dendrogram resembles an evolutionary tree, where terminal branches depict closely related specimens. Within this single hierarchical tree, the intercluster distance (branch length of the dendrogram) directly correlates with dissimilarity in gene expression. Similarly expressed genes are ordered next to each other, and visualized as color coded rows (red for overexpressed, green for underexpressed).

In this way, genes are organized into functional categories, which may reveal clinically important and previously unknown molecular classes within one single disease group (ie, class discovery). A commonly used strategy for outcome prediction is to select cases with known outcome (eg, tumors that did and did not respond to a particular therapy).

Validation is an important component of unsupervised clustering. Cases should be randomly divided into two groups. The first group, the training set, is used to discover a genomic predictive test, and the second, referred to as the validation set, is used to confirm the results and estimate its predictive accuracy in independent cases [\[7\]](#).

Supervised clustering — In contrast to unsupervised cluster analysis, supervised algorithms predict classes based on previously known information. The procedure begins with training a machine, using a machine-learning algorithm, to recognize specific gene expression patterns that are associated with known patient groups or clinical outcomes (ie, normal versus tumor tissue, specific disease types, responders versus non-responders, relapse versus no relapse).

Several algorithms are available, including support vector machines [\[8\]](#), weighted voting methods [\[9,10\]](#), nearest neighbor classification [\[11,12\]](#), and artificial neural networks [\[13\]](#). The goal is for the machine to learn to distinguish between members and non-members of a specific class based upon expression data, and construct a classifier or discriminator that will be used to accurately assign an unknown sample to one of these predefined classes. Such a classifier may then be used to assign certain genes to functional classes, and diseases into predefined categories that might predict diagnosis, prognosis, stage, or the response to therapy [\[2,14,15\]](#).

The unsupervised method is less biased and better suited to reveal previously unrecognized tumor subtypes within morphologically similar cancers based upon similarities in global GEP (ie, class discovery). Although molecular subgroups may or may not reflect biologically or clinically important distinctions, GEPs are strikingly accurate in separating tumors along known classification lines in many cases. As examples, unsupervised clustering of breast cancer specimens consistently separates tumors into estrogen receptor (ER)-positive and ER-negative clusters, indicating that ER status has a major impact on the GEP of breast cancer [\[16-18\]](#). In a similar way, acute leukemia can be separated into myeloid (AML) and lymphoblastic (ALL) subtypes on the basis of GEP alone ([figure 1](#)) [\[14\]](#).

CLINICAL APPLICATIONS OF DNA MICROARRAYS — GEP is a powerful tool to identify genes and pathways that are aberrantly expressed during carcinogenesis. While these discoveries enhance our understanding of molecular pathogenesis, they can also suggest novel therapeutic targets, provide information about drug resistance pathways, and refine prognostic classifications.

A major problem in clinical oncology is the heterogeneous response of histologically similar tumors to treatments such as cytotoxic chemotherapy. Increasingly, specific molecules are being identified that are clinically useful predictors of response to certain agents, including the estrogen/progesterone receptor (hormone therapy for breast cancer), HER2 genomic amplification/overexpression/mutation ([trastuzumab](#) therapy for breast and gastric cancer), EGFR kinase domain mutations (especially exon 19 deletions and exon 21 L858R substitution) in lung cancer (EGFR targeted inhibitors [gefitinib](#) and [erlotinib](#)), K-ras mutations in colorectal cancer (lack of response to EGFR-targeted antagonists), KIT

overexpression (tyrosine kinase therapy in gastrointestinal stromal tumors), BCR/ABL gene translocation ([imatinib](#) in chronic myelogenous leukemia), chromosomal translocated anaplastic lymphoma kinase (ALK) fusion gene ([crizotinib](#) in lymphomas and lung cancers), and mutated BRAF V600E ([vemurafenib](#) in advanced melanoma). However, these advances pertain to only a minority of malignancies.

Encouraging emerging data suggest that prediction of response to chemotherapy or biologically targeted agents may be possible by analyzing gene expression profiles (GEPs). With rapid advances in the DNA microarray technologies and more sophisticated studies, microarray analysis has started to make ways into clinical trials and practices in oncology.

Several examples of the potential application of GEPs in clinical oncology are described here to illustrate the utility of this technology in common solid tumors and hematologic malignancies.

Acute leukemia — The first report to show that GEPs could be used to classify tumors analyzed a group of acute leukemias ([figure 1](#)) [14]. Based upon gene expression patterns, acute myeloid leukemia (AML) could be distinguished from acute lymphoblastic leukemia (ALL) without standard histology information. Similarly, B-cell versus T-cell ALL could be separated based on GEP. This study served as proof of principle that clinically useful classifications could be made simply by gene expression patterns. In another report, a case with equivocal histology by standard criteria was accurately classified by gene analysis, demonstrating the potential utility of GEP beyond standard histologic and immunocytochemical methods [19]. (See "[Clinical manifestations, pathologic features, and diagnosis of acute myeloid leukemia](#)", section on 'Myeloid sarcoma'.)

Others have shown that GEP can distinguish among prognostically important subgroups of children with ALL and adults with AML, in some cases identifying those who eventually fail therapy [20-22]. If these findings are confirmed by others, the logical next step is to apply a more intense initial treatment strategy to such patients, selected on the basis of their GEP. (See "[Clinical manifestations, pathologic features, and diagnosis of acute myeloid leukemia](#)" and "[Overview of the presentation and diagnosis of acute lymphoblastic leukemia in children and adolescents](#)", section on 'Cytogenetics'.)

It may also be possible to screen for agents capable of inducing leukemic cell differentiation through changes in their GEP [23]. A major caveat is that gene expression profiles of clinical samples may differ significantly from those seen in cell lines representing the corresponding leukemia [24].

Prostate cancer — A potential application of GEP in men at risk for prostate cancer is the identification of biomarkers that can help select men with a borderline elevation in serum prostate specific antigen (PSA) for biopsy. In addition, GEP might be used to identify men whose early stage tumors are destined to recur and thus would benefit from more aggressive therapy.

GEP has been used to identify several genes (eg, hepsin and pim-1) that are upregulated in prostate cancer compared to benign prostatic hyperplasia and normal prostate tissue, and some are highly correlated with clinical outcome [25-30]. Investigators have found a ≥ 3 -fold difference in expression in over 3000 genes when nonrecurrent prostate cancers were compared to metastatic tumors [29].

Colon cancer — The serine phosphatase PRL-3 is consistently upregulated in metastatic as compared to non-metastatic colorectal cancers [31]. The finding that metastatic potential appears to be encoded in the primary has challenged the notion that metastases arise from rare cells that have acquired the ability to metastasize [32].

GEP is under study as a way to improve prognostication, and perhaps, individualize adjuvant therapy recommendations [33]. An area of intense study is the use of GEP to predict which patients with node-negative resected colon cancer are at a relatively higher risk of relapse, and thus, might benefit from adjuvant chemotherapy, as is typically recommended for patients with node-positive disease. (See "[Adjuvant therapy for resected stage III \(node-positive\) colon cancer](#)".)

Breast cancer — A molecular classification for breast cancer has been proposed based upon GEP [16-18,34-36]. Luminal (mainly estrogen receptor [ER] positive), basal-like (mostly ER-negative), normal-like, and erbB2+ (mostly HER-2 overexpressing, ER-negative) subgroups have been identified, and have different prognoses.

A major area of investigation is the use of such molecular profiling to predict response to therapy [37-42]. The GEPs of breast cancers that respond best to neoadjuvant (preoperative) chemotherapy (ie, basal-like, erbB2+) differ from those of nonresponding or resistant tumors [37-39,42-44].

Analysis of GEP can also distinguish sporadic breast cancers from those associated with BRCA mutations [45]. Perhaps more importantly, GEP can also permit stratification of defined subgroups (ie, those with axillary lymph node-negative breast cancer or grade 2 tumors) into prognostically separate categories [46-52]. In at least some reports, outcome prediction by GEP outperforms existing prognostic classifications. This topic is discussed in detail elsewhere. (See "[Prognostic and predictive factors in early, non-metastatic breast cancer](#)".)

GEP analysis by DNA microarray is available in patients with breast cancer (the 21-gene recurrence score assay, like Oncotype DX™) to quantify the likelihood of a breast cancer recurrence in women with newly diagnosed, node-negative hormone receptor-positive early stage breast cancer. The assay is designed to identify those women whose risk of recurrence is low enough to justify the omission of chemotherapy and use of [tamoxifen](#) alone as systemic adjuvant therapy.

Although commercially available, the benefit of using 21-gene recurrence score assays (eg, Oncotype DX™) to select the adjuvant therapy strategy has not been tested in a prospective trial. Such an approach is being evaluated in the phase III Trial Assigning IndividuaLized Options for Treatment (Rx) (the TAILORx clinical trial), sponsored by the National Cancer Institute and led by the Eastern Cooperative Oncology Group (ECOG). This topic is discussed separately. (See "[Prognostic and predictive factors in early, non-metastatic breast cancer](#)", [section on 'Recurrence Score'](#).)

Lung carcinoma — GEP can reliably separate lung tumors into their morphologic categories of squamous, large cell, small cell, and adenocarcinoma ([figure 2](#)) [53-55]. GEP may have even more utility in the assessment of patients with non-small cell lung cancers (NSCLCs) and similar histology.

Several investigators have attempted to subclassify these tumors by correlating GEP patterns with clinicopathologic variables [53,56,57]:

- A series that included 41 lung adenocarcinomas identified three prognostically separate subgroups [53]. The genes involved in this classification included thyroid transcription factor, hepsin, cathepsin L, vascular endothelial growth factor C (VEGF-C), and the intercellular adhesion molecule-1 (ICAM-1).
- In another report of 139 lung adenocarcinomas defined four distinct subclasses [57]. Tumors expressing neuroendocrine-type genes had a significantly less favorable survival than those lacking such characteristics [57]. The genes that defined the neuroendocrine cluster adenocarcinomas included dopa decarboxylase, achaete-scute homolog 1, and the serine protease kallikrein 11.
- Others used GEP to predict outcome from surgery in 67 patients with resected stage I adenocarcinoma [56]. A specific group of genes distinguished high-risk from lower risk groups, with significantly different survival [56]. Among the 50 genes comprising the risk index were erbB2, VEGF, S100P, cytokeratin 7 and 18, and fas-associated death domain protein.
- In another series of 125 patients from Taiwan with surgically resected NSCLC, 16 genes were identified that correlated with increased or decreased survival. Further RT-PCR validation assay confirmed the microarray findings and showed that survival was significantly associated with five of the 16 genes (DUSP6, MMD, STAT1, ERBB3, and LCK). The five-gene signature was further

validated in microarray data from patients of Western population and was an independent predictor of recurrence and overall survival for patients with surgical resection of NSCLC without any adjuvant therapy. This GEP profile is being used to select high risk patients for adjuvant chemotherapy in prospective clinical trials [58,59].

Lymphoma — Gene expression profiling (GEP) by means of DNA microarrays is an evolving approach to classification, diagnosis, and prognostication of NHL [14,60-67].

As an example, diffuse large B-cell lymphoma (DLBCL) is a clinically heterogeneous disease in which approximately 40 percent of patients with advanced stage disease respond well to combination chemotherapy and are long-term survivors. Using GEP, DLBCL has been subclassified into three distinct molecular subgroups, germinal center B-cell-like (GCB), activated B-cell-like (ABC), and other (type 3), that appear to be derived from different stages of B-cell differentiation, utilize different oncogenic mechanisms, and differ clinically in their ability to be cured by multiagent chemotherapy [64,66].

Patients whose tumors express genes characteristic of germinal center B cells (GCB) have a significantly better outcome from chemotherapy than those whose gene expression is more typical of activated B cells (ABC) (figure 3) [64,66]. In one series for example, a clustering algorithm applied to 58 patients with DLBCL receiving cyclophosphamide, doxorubicin, vincristine and prednisone (CHOP) chemotherapy separated patients into two groups with very different five-year overall survival rates (70 versus 12 percent) [66]. (See "[Prognosis of diffuse large B cell lymphoma](#)", section on 'Overview'.)

Although most of the early studies used fresh frozen tissue sections, similar results have been reported with GEP performed on formalin-fixed, paraffin-embedded material [60,61]. No formal head-to-head comparisons of GEP from fresh versus archived materials have yet been performed.

GEP has also been used to develop a more precise molecular diagnosis of primary mediastinal B-cell lymphoma, a clinically unfavorable entity that cannot be reliably distinguished from other types of diffuse large B-cell lymphoma [68]. These tumors do poorly with CHOP chemotherapy alone and may need more aggressive therapy than used for standard DLBCL. (See "[Epidemiology, clinical manifestations, pathologic features, and diagnosis of diffuse large B cell lymphoma](#)", section on 'Primary mediastinal large B cell lymphoma'.)

Finally, GEP has the potential to reveal new therapeutic molecular targets [62]. As an example, the ABC subtype of DLBCL is characterized by constitutive activation of the nuclear factor kappaB (NF-kappaB) signaling pathway; interference with this pathway selectively kills these lymphoma cells [65]. The ubiquitin-proteasomal pathway and the NF-kappaB axis are intimately involved in the control of apoptosis [69,70]. Inhibitors of this pathway (eg, proteasome inhibitors) can induce apoptosis in human leukemia cells that ectopically express the antiapoptotic protein Bcl-2. One such agent, the synthetic dipeptide boronic acid bortezomib, is a potent promoter of apoptosis in several human tumor cell types. (See "[Treatment of relapsed or refractory multiple myeloma](#)", section on 'Bortezomib'.)

Summary — The rapidly evolving field of DNA microarray analysis and gene expression profiling has wide-ranging implications for the molecular classification of tumors, refinement of prognostic estimates, and prediction of response to therapy. Despite its exciting potential and significant recent advances, this field remains relatively new, and it is premature to conclude that microarray data can be used as a sole means of classifying cancers or predicting outcomes of treatment [71].

Among the specific challenges that must be met are the need for larger studies with appropriate validation [72-75], standardization of methods and establishment of guidelines for the conduct and reporting of studies (as are being worked on by the Microarray Gene Expression Data Society, www.mged.org) [75], and the formation of repositories and registries where research institutions may deposit data for comparison with independent works involving the same malignant disorder [76]. Finally, DNA microarray-based tests must demonstrate utility in prospectively designed clinical trials before this

technology is considered a routine part of clinical evaluation. These studies may eventually establish a new treatment paradigm in personalized cancer therapy in the future.

PROTEOMICS — Despite significant progress in the field of DNA microarrays and GEP, there are some limitations to using a genomic (nucleic acid) approach to molecular tumor classification:

- Proteins rather than nucleic acids represent the functional output of the cell. RNA expression is often not tightly correlated with protein expression [77,78].
- Genomic analysis does not detect posttranslational changes such as proteolytic processing, phosphorylation, or glycosylation, which can alter the activity of a protein. These posttranslational modifications are often critical in various signaling pathways involved in tumorigenesis [77,79,80].

Compared to the study of DNA or RNA expression patterns, analysis of protein expression in tissues, serum, and other biologic samples (called proteomics) may provide a more accurate understanding of the molecular complexities of human tumors [81].

Protein separation and quantification: 2-dimensional gel electrophoresis — The central tool for proteomics analysis has been two-dimensional polyacrylamide gel electrophoresis (2D-PAGE or 2-DE), which separates solubilized proteins on the basis of charge (using a pH gradient within the gel) in the first dimension, and molecular size in the second dimension (hence the term 2-DE) (figure 4). Following separation, the protein spots on the gel are then visualized by a photochemical silver-based staining technique. This analytical technique can separate and display approximately 1000 proteins from complex mixtures on a single gel [82-84].

Initial reports of the analysis of protein expression profiles in cancer patients using 2-DE described only differences in patterns of protein expression rather than specific proteins [85]. Separated proteins could only be identified by comigration, specific antibody labeling, or chemical microsequencing, all of which were time-consuming and technically demanding. Because there are no simple methods for protein amplification, highly sensitive methodologies needed to be developed to identify low-abundance proteins that are important in the malignant process.

Although 2-DE is a powerful tool for the study of differences in protein expression that characterize any two cell types or tissues, it is difficult to apply this technique to large numbers of clinical samples [83,86]. The number of unique proteins contained in a tumor is higher than the number of spots detectable with 2-DE. The human genome contains as many as 40,000 genes. Even if only 3 to 5 percent are expressed at any one time (1200 to 2000 genes), the number of distinct proteins is much higher because of mRNA splicing, posttranslational modification, and proteolysis.

In addition, many early investigations studied lysates from cell lines and human tissues to identify “tumor-specific” changes in protein expression. This approach is limited by the heterogeneity of human tumors, and the inclusion of stroma, normal cells, and premalignant lesions along with tumor in many samples.

The modern era of proteomics as a method of molecular “fingerprinting” of human tumors was made possible by several technological advances:

- Genomic sequencing (ie, completion of the first phase of the Human Genome Project) and the availability of publicly available databases that catalog known protein and nucleic acid sequences
- Highly sensitive quantitative techniques to detect small amounts of protein in biologic samples
- Laser capture microdissection (LCM) technology, which permits the isolation of selected cells out of a tissue section
- Technical advances in mass spectrometry (MS), allowing highly sensitive and specific protein

identification both in the solubilized (liquid) state and in the solid state (ie, directly from tissue samples or sections)

Despite these advances, 2-DE remains the foundation of most proteomics studies (particularly investigations of biomarkers in serum), because of its ability to separate relatively large proteins from biologic materials. With this simple technique and subsequent analysis of gel images with specialized software [87], differences in protein expression (eg, presence versus absence, or intensity of different proteins or isoforms) between healthy and diseased tissues can be revealed. The proteins of interest can then be identified, typically by MS-based strategies.

Protein identification: mass spectrometry — A mass spectrometer is a highly accurate instrument that can separate individual proteins of very similar molecular weight or mass. MS requires a smaller amount of material and has a higher throughput than other sequencing methods for protein identification. The goal of applying MS in clinical proteomics is to generate protein profiles (mass:charge [m/z] ratio versus signal intensity) from body fluids (eg, serum, saliva, urine) or tissue samples to detect quantitative and qualitative changes in protein expression levels that correlate with the disease states.

MS operates in the gas phase, and peptides thus must first be ionized. The development of newer, more gentle ionization methods has led to breakthroughs in the application of MS to biotechnology and clinical proteomics. Ionization can be accomplished by electrospray ionization for solubilized (liquid state) samples, or by laser desorption techniques for samples in the solid state (see below).

The mass of the ions can then be accurately measured by various coupled analyzers, such as a “time of flight” analyzer. In simple terms, the MS instrument accelerates the ionized proteins across a fixed voltage potential in a high vacuum, measuring the amount of time (the “time of flight” or TOF) required for the ion to hit a detector. This allows the molecular weight or mass of a single positively charged molecule to be accurately measured and displayed according to their mass:charge ratio (m/z).

Proteins that are separated on 2-DE gels can be identified by cutting individual spots from the gels. The excised protein spot is digested with trypsin, cleaving proteins at specific amino acid sequences and breaking them down into a mixture of peptides. The peptide fragments produced by trypsin digestion of any protein have a unique mass that can be measured accurately with MS to produce a peptide mass “fingerprint.”

This protein signature or fingerprint is then compared with the peptide masses predicted from theoretical digestion of protein sequences contained within published databases such as SWISS-PROT [88], thus allowing identification of the protein from which the fragments originated [89,90].

Primary fingerprinting with MS — Analyzing cellular proteins with 2-D gels is a technically challenging, time-intensive process that is difficult to automate and is not readily applicable to assessing protein alterations from hundreds of samples or the multitude of proteins that characterize a heterogeneous tumor. 2-DE fails to resolve most proteins that are larger than approximately 100 kDa or smaller than 5 kDa. Furthermore, 2-DE cannot routinely resolve complex protein mixtures into more than 1000 spots.

Complementary technologies have been developed which apply the sensitivity and specificity of MS to rapid and direct mapping and imaging of biomolecules that are present in body fluids or tissue sections. The available techniques, SELDI-TOF MS (surface-enhanced laser desorption ionization time of flight mass spectrometry) and MALDI-TOF MS (matrix-assisted laser desorption ionization time of flight mass spectrometry), use a laser to desorb and ionize proteins from a solid surface and thus are called “laser desorption” techniques [91,92].

SELDI-TOF MS — SELDI-TOF MS differs from MALDI-TOF MS mainly in the functionalization of the chips (the MALDI target) to allow selective fractionation and enrichment of peptides/proteins in the clinical sample analytes. With SELDI-TOF MS, a solubilized sample is adsorbed directly onto one of

several proprietary chromatographic biochip arrays made of stainless steel or aluminum-based solid surfaces or chips (ProteinChip) that have physicochemical characteristics (eg, charge or hydrophobicity) that permit the capture and fractionation of proteins from complex mixtures prior to laser desorption for MS. After a wash step, proteins and peptides of selective affinity are retained on the chip, and analyzed by a linear TOF MS technology similar to MALDI-TOF MS. Using this technique, peptide and protein profiles can be easily obtained within minutes and hundreds of proteins can be profiled simultaneously from samples of patients or a small number of cells when coupled with LCM technology [93].

The SELDI-TOF MS platform offers several advantages over 2-DE. It is able to profile proteins regardless of their intrinsic hydrophobicity (a limitation of 2-DE), and it has a high sensitivity for the detection of proteins in the 15 kDa range, a size that is problematic for 2-DE resolution. In addition, SELDI-TOF MS requires a smaller amount of sample and has a higher throughput than 2-DE.

In early studies, SELDI MS analysis of serum proteins was capable of defining patterns that could distinguish patients with cancer from normal controls and from those with premalignant lesions [93]. Subsequent attempts have focused on using SELDI-TOF MS to identify discriminatory "proteomics signatures" of spectral peaks between normal versus malignant tissues, as well as to identify single disease-related tumor biomarkers [94-100].

A potential problem is that routine identification of the protein peaks using this approach is not straightforward due to the requirement for front-end fractionation of the protein samples, and subsequent downstream purification methods. However, direct on-chip protein identification is technically possible [101].

In addition, the clinical application of proteomic patterns requires that results be reproducible in different laboratories and on multiple samples from the same patient [102]. As an example, the initial positive results from sera of women with ovarian cancer and a control group could not be readily reproduced when the datasets were generated using two different protein chips [97,103].

MALDI-TOF MS — In contrast to SELDI-TOF MS, MALDI-TOF MS does not require solubilization or prepreparation of tissue samples. Using this technique, high resolution mass spectrograms can be obtained directly from solid phase samples (eg, in tissue sections) [91,92,104,105].

In a typical procedure, a 12 micron frozen tissue section is freeze-dried onto a conductive MALDI plate, and coated with a solution of a matrix, usually a low molecular weight crystalline compound such as a cinnamic acid analog (figure 5). When dry, the sample is introduced into the vacuum inlet of a mass spectrometer, and then irradiated with a finely focused laser beam from a 377 nm N2 laser. The beam, which is approximately circular with a diameter of 25 microns, is directed at regions of interest in the tumor, which are selected by examining adjacent tissue sections that are processed for conventional light microscopy (hematoxylin and eosin, H&E).

Each consecutive spot produces a mass spectrum that reflects the molecules that are present within the irradiated area. For most tissue sections, over 200 protein and peptide peaks are found in the mass spectrum from each spot ablated by the laser. Application of the laser, acquisition of data, download to a computer, data processing, and repositioning of the sample stage are all automated, and the procedure is rapid.

Samples can be obtained from well-defined regions (eg, normal tissue, malignant tissue, necrotic areas, the leading edge of a tumor), many of which have unique subsets of proteins and peptides. Images can then be created of different patterns of protein expression by integrating the peak areas and plotting the relative values using a color scale (figure 6). Proteins can also be identified through extraction, fractionation, proteolysis, MS sequencing of one or more of the fragments, and protein database searching.

The ClinProt system is another integrated method for serum profiling that utilizes magnetic beads (MB) and AnchorChip technology for front-end sample fractionation and preparation prior to MS analysis. MALDI-TOF MS is used to generate protein profiles in both linear and reflector modes, with subsequent MALDI-TOF MS to identify low molecular weight serum peptides in the protein profile with high spectral resolution [106].

A particular advantage of this platform is its ability to allow profiling to discovery of proteomic fingerprint and peptide/protein identification all in a single, integrated system with high throughput. MB-based sample fractionation can be easily automated and scaled up for subsequent biomarker purification. High performance ultraflex MALDI-TOF/TOF can be used to identify peptides directly from serum profiles using the same sample spot and/or re-spot the same sample.

Clinical applications — The result of a proteomics analysis is the correlation of changes in protein expression with a given phenotype. Once a protein or biomarker "signature profile" has been identified, results must be validated in independent, large study sets. This requires high throughput technology, permitting the analysis of hundreds of cancer tissue samples at a time.

The identified proteins may represent important biomarkers for early diagnosis, yield valuable information about the disease process and thus identify targets for novel therapeutic intervention, or provide surrogate markers for therapeutic efficacy or treatment-related toxicity.

The following illustrate clinical areas where proteomics may have potential clinical utility.

Non-small cell lung cancer — Preliminary data suggest that MALDI-TOF MS analysis of protein expression profiles from frozen sections of human surgically resected lung tumors permits the separation of tumors according to histologic type, providing potentially useful prognostic information in patients with resected non-small cell lung cancer (NSCLC) [91,107,108].

In one report, protein spectra were obtained from 79 lung tumors and 14 normal lung tissues [107]. A model was established using 42 lung tumors and eight normal samples, which was then validated on a blinded test cohort of 37 lung tumors and six normal lung samples. Using this approach over 1000 distinct protein species could be analyzed from less than one nanogram of each tissue sample.

Representative examples of MS spectra from adenocarcinoma, squamous cell, and normal lung are shown in the following figure (figure 7). The protein spectra of lung cancers and normal lung are different; peaks are higher in the tumor samples than in normal lung, and clearly distinguishable from background noise. A discriminatory pattern of protein expression was also evident when the adenocarcinoma and squamous cell carcinoma were compared (figure 7), and when lung primary tumors were compared to pulmonary metastasis.

These proteomics results may have prognostic value as well. When the protein spectra of 66 patients with resected NSCLC were analyzed, a single pattern comprised of 15 distinct MS peaks could distinguish between patients with good or poor prognoses (median survival 33 versus 6 months, respectively) [107]. This association was present even after adjustment for stage and histologic grade.

Other data suggest that serum protein profiling may be useful in the early diagnosis of aerodigestive tumors (lung, head and neck cancer) in subjects who are at high risk (eg, smokers) [109]. The spectral profile derived from the sera of patients with cancer was significantly different from the sera of control subjects. If these results are confirmed by others, a proteomics-based analysis may be useful as a screening test for these cancers.

There have been considerable efforts in recent years in developing serum-based proteomics classifiers to predict survival and treatment benefits in NSCLC. A commercially available MALDI-TOF based classifier (VeriStrat) has been developed and validated as a tool for stratifying outcomes and predicting survival after treatment with [erlotinib](#) versus chemotherapy in patients with recurrent advanced NSCLC.

These data are discussed elsewhere. (See "[Advanced non-small cell lung cancer: Subsequent therapies for previously treated patients](#)", section on '[Proteomic signature assay](#)'.)

Prostate cancer — The lack of specificity of serum PSA for discriminating between prostate cancer and benign prostatic hyperplasia (BPH) limits its utility for early diagnosis. (See "[Screening for prostate cancer](#)", section on '[Measuring PSA](#)'.)

Several reports suggest that serum proteomics may provide a better biomarker for early detection of prostate cancer [[94,95,100,110](#)].

- The first series used a serum SELDI MS to analyze serum proteins and differentiate prostate cancer from BPH [[94](#)]. Serum protein profiles from 326 men with prostate cancer, BPH, or unaffected controls were used to develop a decision tree classification algorithm, which correctly classified 96 percent of the samples. In a subsequent blinded validation set, this approach had a sensitivity of 83 percent and a specificity of 97 percent.
- Using a different approach, serum was analyzed for the presence of autoantibodies directed against tumor-specific antigens in patients with prostatic cancer [[110](#)]. Serum samples were analyzed from 39 patients and 21 controls to establish a panel of 22 markers. An algorithm was then developed in a training phase involving 129 samples, and validated in samples from an additional 128 patients. Results were superior to those with PSA in this population.

If the results from approaches such as these are reproducible in larger studies and if they are superior to PSA in a screening population as opposed to patients with biopsy proven disease, serum proteomics may be useful either as a screening tool or in deciding whether to perform a biopsy in a man with an elevated serum PSA level.

Adult acute lymphoblastic leukemia — In a study of peripheral blood from 57 adults with a confirmed diagnosis of acute lymphoblastic leukemia, use of a decision tree based on proteomics provided positive prediction of recurrence 84 to 92 percent of the time [[111](#)]. Recurrence prediction was independent of cytogenetics, marrow blast count, serum level of lactate dehydrogenase, beta-2 microglobulin levels, or surface marker studies.

MICRORNA PROFILING — MicroRNAs (miRNAs) are a novel class of endogenous small (18 to 24 nucleotides long) noncoding single-stranded RNAs that regulate gene expression at the posttranscriptional level. Evolutionarily conserved, miRNAs bind to the 3' untranslated regions of messenger RNAs (mRNAs), and induce degradation or inhibition of protein translation.

MiRNAs possess many critical regulatory functions in a wide range of biological processes such as cell proliferation, differentiation, survival and apoptosis, and stress response. The miRBase miRNA registry now includes over 1000 human miRNAs [[112](#)]. Any one particular miRNA has the potential to modulate the expression and functions of hundreds of downstream target genes [[113](#)]. In addition, the existence of feedback regulation mechanisms between miRNA, its targets, and their products allows for amplification or inhibition of a specific signal. Hence, alteration of even a handful of miRNAs can possibly result in dramatic deregulation of physiologic cellular functions.

Human miRNA genes are frequently located at fragile sites and genomic regions involved in cancer [[114](#)]. MiRNA expression is deregulated widely in solid tumors and hematologic malignancies, and miRNAs have been implicated both in the initiation and progression of human cancer [[115](#)]. MiRNAs appear to have a dual role in carcinogenesis by serving both as oncogenes (termed oncomirs [[116](#)]) and tumor suppressors [[117](#)].

The following sections will provide a few examples of early data on miRNA profiling for cancer diagnosis and prognosis.

Leukemia/lymphoma — Among patients diagnosed with B-cell chronic lymphocytic leukemia (CLL), investigators have found deletions or downregulation of two clusters of miRNA genes, miR-15a and miR-16-1 [118]. These results provided the first evidence of a miRNA serving as a tumor suppressor in a human cancer. Both miR-15a and miR-16-1 have been shown to target the anti-apoptotic protein BCL-2 [119], and its deregulation (mutation or deletion) appears relevant to the development of indolent rather than aggressive CLL. Other miRNAs may be involved as well. In one large study, a 13-miRNA signature was identified that could distinguish between indolent and aggressive CLL [120]. (See ["Pathophysiology and genetic features of chronic lymphocytic leukemia", section on 'MicroRNA'.](#))

MiRNA profiles with a unique association to acute myelogenous leukemia (AML) subgroups have been reported, which also appear to have prognostic significance [121]. The impact of miRNAs on AML pathogenesis is not yet understood [122]. (See ["Molecular genetics of acute myeloid leukemia", section on 'Mutations affecting DNA methylation'](#) and ["Prognosis of acute myeloid leukemia", section on 'MicroRNA expression profiling'.](#))

MiRNAs may also be important in lymphomas. MiR-155 is upregulated in diffuse large B-cell lymphoma, Burkitt's lymphoma, and Hodgkin's lymphomas, suggesting it may function as an oncogene. In a transgenic mouse model, a B-cell malignancy ultimately developed in mice with constitutive miR-155 overexpression in B-lymphocytes [123].

Lung cancer — Functional studies on one of these miRNAs, let-7, indicate that it regulates the oncogene RAS and the HMGA2 (high mobility group at-hook 2) gene directly in patients with lung cancer [124,125]. These data suggest a potential pathogenetic role in the disease.

MiRNA profiling for classification and prognostication in lung cancer is an active area of investigation [126-130]:

- MiRNA expression profiles correlate with survival of lung adenocarcinomas, including those classified as stage I [126,131]. A potential prognostic role for miRNA expression was also suggested in a study that evaluated the expression of a panel of 752 miRNAs in 77 cases of stage I non-small cell lung cancer (NSCLC), 37 of which recurred and 40 did not [132]. Differential expression between the two groups was present for 49 percent of the miRNAs studied.
- MiRNA studies have also been applied to the diagnostic distinction between squamous and non-squamous NSCLC subtypes [133]. This work has potential clinical relevance given that recent advances in treatment of lung cancer require greater accuracy in the subclassification of NSCLC (eg, use of [bevacizumab](#) in combination with chemotherapy for patients with non-squamous histology NSCLC). (See ["Systemic therapy for the initial management of advanced non-small cell lung cancer without a driver mutation", section on 'Bevacizumab'.](#))

Breast cancer — The relevance of miRNA activity in breast cancer biology and behavior is supported by the evidence that miRNAs participate in the modulation of two of the most important molecules in breast cancer, estrogen receptor and HER2 [134-138].

Several studies have described an association between miRNA profiles and pathologic and clinical features of breast cancer, as well as prognosis [134]:

- MiRNAs are differentially expressed in specific molecular subgroups of breast cancer (eg, luminal A versus basal-like [139]) and in certain histologic subtypes (eg, inflammatory versus noninflammatory breast cancer [140]). (See ["Prognostic and predictive factors in early, non-metastatic breast cancer", section on 'Genomic profiles'](#) and ["Inflammatory breast cancer: Pathology and molecular pathogenesis".](#))
- In two studies, miR-21 was found overexpressed with a key oncogenic role in promoting tumor cell survival and proliferation; overexpression correlated with advanced stage disease and poor

prognosis [141,142]. In addition, a regulatory role for miRNA-10b has been proposed for mediating breast cancer invasion and metastasis [143,144].

Role in prognostication — An increasing amount of data support a role for microRNAs in clinical outcome prediction for a wide variety of cancers. A systematic review and meta-analysis of 43 studies pertaining to 20 different tumor types concluded that the miRNAs most frequently associated with poor outcomes after accounting for differences in miRNA assessment due to platform type were let-7 (decreased expression) and miRNA 21 (increased expression) [145]. The analysis adjusted for how many times each miRNA was assayed, as not all known miRNAs were assessed upfront in all included studies.

Several miR classifiers were associated with prognosis above and beyond traditional clinical and prognostic metrics for a diverse group of human malignancies. Moreover, the finding of specific miRNA species (ie, let-7 and miR-21) that were associated with survival outcome and consistently identified across diverse studies in different cancers suggests that some miR-coordinated regulatory pathways are common to many cancers.

Role in cancer therapeutics — The discovery of miRNAs functioning as potential oncogenes and tumor suppressor genes in cancer has generated interest in their use as targets for cancer therapy [146,147]. In addition, miRNA might also have treatment modulating effects in human cancer, and potential as anticancer therapeutic agents [147-149].

Role in cancer screening and early detection — Some miRNAs (miRNA-421 and miR-106a) are highly expressed in gastric cancers and are detectable in peripheral blood and gastric aspirates, suggesting their potential utility as biomarkers. However, a role for miRNAs in screening for gastric cancer has not been established. (See ["Screening and prevention of gastric cancer", section on 'Screening methods'.](#))

Similarly, others have shown that miRNA-21 serum levels are higher in patients with colorectal cancer than in healthy controls and that levels are higher with increasing disease stages [150-152]. While these data suggest potential utility for miRNA-21 as a diagnostic biomarker for colorectal cancer, prospective validation in independent studies is needed.

Summary — Emerging evidence suggests a role for deregulation of microRNA (miRNA) expression in tumorigenesis, with potential applications in cancer diagnostics, classification, prognostication, and possibly treatment. This field is in its infancy and further work will be needed to validate any clinical applications for the use of miRNA in human cancer care.

SUMMARY

- For patients with cancer, clinical management decisions and prognostic estimates are often based solely upon histopathologic analysis of tumor tissue. However, tumor behavior is often inadequately explained by histology alone. The examination of multiple expressed genes and/or proteins may provide more useful information for both classification and prognostication of individual tumors. (See ["Introduction"](#) above.)
- DNA microarray analysis can simultaneously measure the expression level of thousands of genes. Expression levels of each gene can then be compared across many tissue or tumor samples. In this way, groups or clusters of genes can be identified whose expression varies in similar ways. Gene expression profiling (GEP) is a powerful tool to identify genes and pathways that are aberrantly expressed during carcinogenesis. (See ["GEP and DNA microarray analysis"](#) above.)
- The rapidly evolving field of DNA microarray analysis and GEP has wide-ranging implications for the molecular classification of tumors, refinement of prognostic estimates, and prediction of response to therapy. GEP can also suggest novel therapeutic targets. However, despite its exciting

potential and significant recent advances, this field remains relatively new, and it is premature to conclude that microarray data can be used as a sole means of classifying cancers or predicting outcomes of treatment. (See '[Clinical applications of DNA microarrays](#)' above.)

- The application of clinical proteomics to clinical oncology has accelerated tremendously with rapid advances in mass spectrometry (MS) technology. (See '[Proteomics](#)' above.)
- These advances may be useful in several dimensions (see '[Clinical applications](#)' above):
 - Proteomics research will enhance our understanding of tumor biology, particularly the aberrant cellular signaling that characterizes malignant disease.
 - Rapid advances in MS platforms coupled with high throughput and automation capacity will facilitate identification of potential novel biomarkers.
 - Identification of complex disease-specific biomarker patterns may allow both better sensitivity and specificity when compared to individual single biomarker analysis, as is used currently for several cancers.
- Despite the exciting potential for clinical proteomics, substantial challenges remain, particularly with regard to standardization, validation, and reproducibility of results. Predictive proteomic biomarkers ultimately need to be tested in prospective therapeutics clinical trials to be fully validated.
- MicroRNAs (miRNAs) are a novel class of small (approximately 22 nucleotides long) noncoding single-stranded RNAs that regulate gene expression at the posttranscriptional level. Emerging evidence suggests a role for deregulation of miRNA expression in tumorigenesis, with potential applications in cancer diagnostics, classification, prognostication, and possibly treatment. This field is in its infancy and further work will be needed to validate any clinical applications for the use of miRNA in human cancer care. (See '[MicroRNA profiling](#)' above.)

Use of UpToDate is subject to the [Subscription and License Agreement](#).

REFERENCES

1. Mariadason JM, Augenlicht LH, Arango D. Microarray analysis in the clinical management of cancer. *Hematol Oncol Clin North Am* 2003; 17:377.
2. Tefferi A, Bolander ME, Ansell SM, et al. Primer on medical genomics. Part III: Microarray experiments and data analysis. *Mayo Clin Proc* 2002; 77:927.
3. Garnis C, Lockwood WW, Vucic E, et al. High resolution analysis of non-small cell lung cancer cell lines by whole genome tiling path array CGH. *Int J Cancer* 2006; 118:1556.
4. Wittes J, Friedman HP. Searching for evidence of altered gene expression: a comment on statistical analysis of microarray data. *J Natl Cancer Inst* 1999; 91:400.
5. Baggerly KA, Coombes KR, Hess KR, et al. Identifying differentially expressed genes in cDNA microarray experiments. *J Comput Biol* 2001; 8:639.
6. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* 1998; 95:14863.
7. Simon R. Diagnostic and prognostic prediction using gene expression profiles in high-dimensional microarray data. *Br J Cancer* 2003; 89:1599.
8. Ramaswamy S, Tamayo P, Rifkin R, et al. Multiclass cancer diagnosis using tumor gene expression signatures. *Proc Natl Acad Sci U S A* 2001; 98:15149.
9. Hwang D, Schmitt WA, Stephanopoulos G, Stephanopoulos G. Determination of minimum sample

size and discriminatory expression patterns in microarray data. *Bioinformatics* 2002; 18:1184.

10. Mukherjee S, Tamayo P, Rogers S, et al. Estimating dataset size requirements for classifying DNA microarray data. *J Comput Biol* 2003; 10:119.
11. Ben-Dor A, Bruhn L, Friedman N, et al. Tissue classification with gene expression profiles. *J Comput Biol* 2000; 7:559.
12. Tibshirani R, Hastie T, Narasimhan B, Chu G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc Natl Acad Sci U S A* 2002; 99:6567.
13. Khan J, Wei JS, Ringnér M, et al. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nat Med* 2001; 7:673.
14. Golub TR, Slonim DK, Tamayo P, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 1999; 286:531.
15. Brown MP, Grundy WN, Lin D, et al. Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proc Natl Acad Sci U S A* 2000; 97:262.
16. Pusztai L, Ayers M, Stec J, et al. Gene expression profiles obtained from fine-needle aspirations of breast cancer reliably identify routine prognostic markers and reveal large-scale molecular differences between estrogen-negative and estrogen-positive tumors. *Clin Cancer Res* 2003; 9:2406.
17. Gruvberger S, Ringnér M, Chen Y, et al. Estrogen receptor status in breast cancer is associated with remarkably distinct gene expression patterns. *Cancer Res* 2001; 61:5979.
18. Perou CM, Sørlie T, Eisen MB, et al. Molecular portraits of human breast tumours. *Nature* 2000; 406:747.
19. Haeflrich T, Kohlmann A, Bacher U, et al. Gene expression profiling for the diagnosis of acute leukaemia. *Br J Cancer* 2007; 96:535.
20. Yeoh EJ, Ross ME, Shurtleff SA, et al. Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. *Cancer Cell* 2002; 1:133.
21. Bullinger L, Döhner K, Bair E, et al. Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia. *N Engl J Med* 2004; 350:1605.
22. Valk PJ, Verhaak RG, Beijnen MA, et al. Prognostically useful gene-expression profiles in acute myeloid leukemia. *N Engl J Med* 2004; 350:1617.
23. Stegmaier K, Ross KN, Colavito SA, et al. Gene expression-based high-throughput screening(GE-HTS) and application to leukemia differentiation. *Nat Genet* 2004; 36:257.
24. Leupin N, Kuhn A, Hügli B, et al. Gene expression profiling reveals consistent differences between clinical samples of human leukaemias and their model cell lines. *Br J Haematol* 2006; 135:520.
25. Dhanasekaran SM, Barrette TR, Ghosh D, et al. Delineation of prognostic biomarkers in prostate cancer. *Nature* 2001; 412:822.
26. Latil A, Bièche I, Chêne L, et al. Gene expression profiling in clinically localized prostate cancer: a four-gene expression model predicts clinical behavior. *Clin Cancer Res* 2003; 9:5477.
27. Stephan C, Yousef GM, Scorilas A, et al. Hepsin is highly over expressed in and a new candidate for a prognostic indicator in prostate cancer. *J Urol* 2004; 171:187.
28. Lapointe J, Li C, Higgins JP, et al. Gene expression profiling identifies clinically relevant subtypes of prostate cancer. *Proc Natl Acad Sci U S A* 2004; 101:811.
29. LaTulippe E, Satagopan J, Smith A, et al. Comprehensive gene expression analysis of prostate cancer reveals distinct transcriptional programs associated with metastatic disease. *Cancer Res* 2002; 62:4499.
30. Yu YP, Landsittel D, Jing L, et al. Gene expression alterations in prostate cancer predicting tumor aggression and preceding development of malignancy. *J Clin Oncol* 2004; 22:2790.

31. Saha S, Bardelli A, Buckhaults P, et al. A phosphatase associated with metastasis of colorectal cancer. *Science* 2001; 294:1343.
32. Ramaswamy S, Ross KN, Lander ES, Golub TR. A molecular signature of metastasis in primary solid tumors. *Nat Genet* 2003; 33:49.
33. Watanabe T, Kobunai T, Sakamoto E, et al. Gene expression signature for recurrence in stage III colorectal cancers. *Cancer* 2009; 115:283.
34. Sotiriou C, Neo SY, McShane LM, et al. Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proc Natl Acad Sci U S A* 2003; 100:10393.
35. Sørlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* 2001; 98:10869.
36. Sørlie T, Tibshirani R, Parker J, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A* 2003; 100:8418.
37. Thuerigen O, Schneeweiss A, Toedt G, et al. Gene expression signature predicting pathologic complete response with gemcitabine, epirubicin, and docetaxel in primary breast cancer. *J Clin Oncol* 2006; 24:1839.
38. Ayers M, Symmans WF, Stec J, et al. Gene expression profiles predict complete pathologic response to neoadjuvant paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide chemotherapy in breast cancer. *J Clin Oncol* 2004; 22:2284.
39. Iwao-Koizumi K, Matoba R, Ueno N, et al. Prediction of docetaxel response in human breast cancer by gene expression profiling. *J Clin Oncol* 2005; 23:422.
40. Jansen MP, Foekens JA, van Staveren IL, et al. Molecular classification of tamoxifen-resistant breast carcinomas by gene expression profiling. *J Clin Oncol* 2005; 23:732.
41. Hannemann J, Oosterkamp HM, Bosch CA, et al. Changes in gene expression associated with response to neoadjuvant chemotherapy in breast cancer. *J Clin Oncol* 2005; 23:3331.
42. Rouzier R, Perou CM, Symmans WF, et al. Breast cancer molecular subtypes respond differently to preoperative chemotherapy. *Clin Cancer Res* 2005; 11:5678.
43. Chang JC, Wooten EC, Tsimelzon A, et al. Patterns of resistance and incomplete response to docetaxel by gene expression profiling in breast cancer patients. *J Clin Oncol* 2005; 23:1169.
44. Hess KR, Anderson K, Symmans WF, et al. Pharmacogenomic predictor of sensitivity to preoperative chemotherapy with paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide in breast cancer. *J Clin Oncol* 2006; 24:4236.
45. Hedenfalk I, Duggan D, Chen Y, et al. Gene-expression profiles in hereditary breast cancer. *N Engl J Med* 2001; 344:539.
46. Pusztai L, Ayers M, Stec J, Hortobágyi GN. Clinical application of cDNA microarrays in oncology. *Oncologist* 2003; 8:252.
47. Glinsky GV, Higashiyama T, Glinskii AB. Classification of human breast cancer using gene expression profiling as a component of the survival predictor algorithm. *Clin Cancer Res* 2004; 10:2272.
48. van 't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002; 415:530.
49. van de Vijver MJ, He YD, van't Veer LJ, et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 2002; 347:1999.
50. Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 2004; 351:2817.
51. Wang Y, Klijn JG, Zhang Y, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* 2005; 365:671.

52. Sotiriou C, Wirapati P, Loi S, et al. Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *J Natl Cancer Inst* 2006; 98:262.
53. Garber ME, Troyanskaya OG, Schluens K, et al. Diversity of gene expression in adenocarcinoma of the lung. *Proc Natl Acad Sci U S A* 2001; 98:13784.
54. Yamagata N, Shyr Y, Yanagisawa K, et al. A training-testing approach to the molecular classification of resected non-small cell lung cancer. *Clin Cancer Res* 2003; 9:4695.
55. Jones MH, Virtanen C, Honjoh D, et al. Two prognostically significant subtypes of high-grade lung neuroendocrine tumours independent of small-cell and large-cell neuroendocrine carcinomas identified by gene expression profiles. *Lancet* 2004; 363:775.
56. Beer DG, Kardia SL, Huang CC, et al. Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nat Med* 2002; 8:816.
57. Bhattacharjee A, Richards WG, Staunton J, et al. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc Natl Acad Sci U S A* 2001; 98:13790.
58. Chen HY, Yu SL, Chen CH, et al. A five-gene signature and clinical outcome in non-small-cell lung cancer. *N Engl J Med* 2007; 356:11.
59. Potti A, Mukherjee S, Petersen R, et al. A genomic strategy to refine prognosis in early-stage non-small-cell lung cancer. *N Engl J Med* 2006; 355:570.
60. Rimsza LM, Leblanc ML, Unger JM, et al. Gene expression predicts overall survival in paraffin-embedded tissues of diffuse large B-cell lymphoma treated with R-CHOP. *Blood* 2008; 112:3425.
61. Malumbres R, Chen J, Tibshirani R, et al. Paraffin-based 6-gene model predicts outcome in diffuse large B-cell lymphoma patients treated with R-CHOP. *Blood* 2008; 111:5509.
62. Staudt LM. Molecular diagnosis of the hematologic cancers. *N Engl J Med* 2003; 348:1777.
63. Thieblemont C, Nasser V, Felman P, et al. Small lymphocytic lymphoma, marginal zone B-cell lymphoma, and mantle cell lymphoma exhibit distinct gene-expression profiles allowing molecular diagnosis. *Blood* 2004; 103:2727.
64. Alizadeh AA, Eisen MB, Davis RE, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 2000; 403:503.
65. Wiestner A, Staudt LM. Towards molecular diagnosis and targeted therapy of lymphoid malignancies. *Semin Hematol* 2003; 40:296.
66. Shipp MA, Ross KN, Tamayo P, et al. Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. *Nat Med* 2002; 8:68.
67. Rosenwald A, Wright G, Chan WC, et al. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N Engl J Med* 2002; 346:1937.
68. Rosenwald A, Wright G, Leroy K, et al. Molecular diagnosis of primary mediastinal B cell lymphoma identifies a clinically favorable subgroup of diffuse large B cell lymphoma related to Hodgkin lymphoma. *J Exp Med* 2003; 198:851.
69. Goldberg AL, Rock K. Not just research tools--proteasome inhibitors offer therapeutic promise. *Nat Med* 2002; 8:338.
70. Voorhees PM, Dees EC, O'Neil B, Orlowski RZ. The proteasome as a target for cancer therapy. *Clin Cancer Res* 2003; 9:6316.
71. Winegarden N. Microarrays in cancer: moving from hype to clinical reality. *Lancet* 2003; 362:1428.
72. Ntzani EE, Ioannidis JP. Predictive ability of DNA microarrays for cancer outcomes and correlates: an empirical assessment. *Lancet* 2003; 362:1439.
73. Dobbin KK, Beer DG, Meyerson M, et al. Interlaboratory comparability study of cancer gene

- expression analysis using oligonucleotide microarrays. *Clin Cancer Res* 2005; 11:565.
74. Michiels S, Koscielny S, Hill C. Prediction of cancer outcome with microarrays: a multiple random validation strategy. *Lancet* 2005; 365:488.
 75. Dupuy A, Simon RM. Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting. *J Natl Cancer Inst* 2007; 99:147.
 76. Stoeckert CJ Jr, Causton HC, Ball CA. Microarray databases: standards and ontologies. *Nat Genet* 2002; 32 Suppl:469.
 77. Anderson L, Seilhamer J. A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* 1997; 18:533.
 78. Tew KD, Monks A, Barone L, et al. Glutathione-associated enzymes in the human cell lines of the National Cancer Institute Drug Screening Program. *Mol Pharmacol* 1996; 50:149.
 79. Wilkins MR, Sanchez JC, Williams KL, Hochstrasser DF. Current challenges and future applications for protein maps and post-translational vector maps in proteome projects. *Electrophoresis* 1996; 17:830.
 80. Chen G, Gharib TG, Huang CC, et al. Discordant protein and mRNA expression in lung adenocarcinomas. *Mol Cell Proteomics* 2002; 1:304.
 81. Bichsel VE, Liotta LA, Petricoin EF 3rd. Cancer proteomics: from biomarker discovery to signal pathway profiling. *Cancer J* 2001; 7:69.
 82. Corthals GL, Wasinger VC, Hochstrasser DF, Sanchez JC. The dynamic range of protein expression: a challenge for proteomic research. *Electrophoresis* 2000; 21:1104.
 83. Banks RE, Dunn MJ, Hochstrasser DF, et al. Proteomics: new perspectives, new biomedical opportunities. *Lancet* 2000; 356:1749.
 84. Cristea IM, Gaskell SJ, Whetton AD. Proteomics techniques and their application to hematology. *Blood* 2004; 103:3624.
 85. Wright GL Jr. Two-dimensional acrylamide gel electrophoresis of cancer-patient serum proteins. *Ann Clin Lab Sci* 1974; 4:281.
 86. Pandey A, Mann M. Proteomics to study genes and genomes. *Nature* 2000; 405:837.
 87. Appel RD, Palagi PM, Walther D, et al. Melanie II--a third-generation software package for analysis of two-dimensional electrophoresis images: I. Features and user interface. *Electrophoresis* 1997; 18:2724.
 88. Bairoch A, Apweiler R. The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 1998. *Nucleic Acids Res* 1998; 26:38.
 89. Wilm M, Shevchenko A, Houthaeve T, et al. Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature* 1996; 379:466.
 90. Gygi SP, Rist B, Gerber SA, et al. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* 1999; 17:994.
 91. Caprioli RM, Farmer TB, Gile J. Molecular imaging of biological samples: localization of peptides and proteins using MALDI-TOF MS. *Anal Chem* 1997; 69:4751.
 92. Stoeckli M, Chaurand P, Hallahan DE, Caprioli RM. Imaging mass spectrometry: a new technology for the analysis of protein expression in mammalian tissues. *Nat Med* 2001; 7:493.
 93. Paweletz CP, Gillespie JW, Ornstein DK, et al. Rapid protein display profiling of cancer progression directly from human tissue using a protein biochip. *Drug Dev Res* 2000; 49:34.
 94. Adam BL, Qu Y, Davis JW, et al. Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men. *Cancer Res* 2002; 62:3609.

95. Petricoin EF 3rd, Ornstein DK, Paweletz CP, et al. Serum proteomic patterns for detection of prostate cancer. *J Natl Cancer Inst* 2002; 94:1576.
96. Mathelin C, Cromer A, Wendling C, et al. Serum biomarkers for detection of breast cancers: A prospective study. *Breast Cancer Res Treat* 2006; 96:83.
97. Petricoin EF, Ardekani AM, Hitt BA, et al. Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* 2002; 359:572.
98. Poon TC, Yip TT, Chan AT, et al. Comprehensive proteomic profiling identifies serum proteomic signatures for detection of hepatocellular carcinoma and its subtypes. *Clin Chem* 2003; 49:752.
99. Koopmann J, Zhang Z, White N, et al. Serum diagnosis of pancreatic adenocarcinoma using surface-enhanced laser desorption and ionization mass spectrometry. *Clin Cancer Res* 2004; 10:860.
100. Ornstein DK, Rayford W, Fusaro VA, et al. Serum proteomic profiling can discriminate prostate cancer from benign prostates in men with total prostate specific antigen levels between 2.5 and 15.0 ng/ml. *J Urol* 2004; 172:1302.
101. Caputo E, Moharram R, Martin BM. Methods for on-chip protein analysis. *Anal Biochem* 2003; 321:116.
102. Robbins RJ, Villanueva J, Tempst P. Distilling cancer biomarkers from the serum peptidome: high technology reading of tea leaves or an insight to clinical systems biology? *J Clin Oncol* 2005; 23:4835.
103. Baggerly KA, Morris JS, Coombes KR. Reproducibility of SELDI-TOF protein patterns in serum: comparing datasets from different experiments. *Bioinformatics* 2004; 20:777.
104. Chaurand P, DaGue BB, Pearsall RS, et al. Profiling proteins from azoxymethane-induced colon tumors at the molecular level by matrix-assisted laser desorption/ionization mass spectrometry. *Proteomics* 2001; 1:1320.
105. Chaurand P, Schwartz SA, Caprioli RM. Imaging mass spectrometry: a new tool to investigate the spatial organization of peptides and proteins in mammalian tissue sections. *Curr Opin Chem Biol* 2002; 6:676.
106. Pusch W, Flocco MT, Leung SM, et al. Mass spectrometry-based clinical proteomics. *Pharmacogenomics* 2003; 4:463.
107. Yanagisawa K, Shyr Y, Xu BJ, et al. Proteomic patterns of tumour subsets in non-small-cell lung cancer. *Lancet* 2003; 362:433.
108. Chen G, Gharib TG, Wang H, et al. Protein profiles associated with survival in lung adenocarcinoma. *Proc Natl Acad Sci U S A* 2003; 100:13537.
109. Sidransky D, Irizarry R, Califano JA, et al. Serum protein MALDI profiling to distinguish upper aerodigestive tract cancer patients from control subjects. *J Natl Cancer Inst* 2003; 95:1711.
110. Wang X, Yu J, Sreekumar A, et al. Autoantibody signatures in prostate cancer. *N Engl J Med* 2005; 353:1224.
111. Albitar M, Potts SJ, Giles FJ, et al. Proteomic-based prediction of clinical behavior in adult acute lymphoblastic leukemia. *Cancer* 2006; 106:1587.
112. MicroRNA database available online at <http://www.mirbase.org/> (Accessed on October 12, 2011).
113. Lim LP, Lau NC, Garrett-Engele P, et al. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* 2005; 433:769.
114. Calin GA, Sevignani C, Dumitru CD, et al. Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proc Natl Acad Sci U S A* 2004; 101:2999.
115. Lovat F, Valeri N, Croce CM. MicroRNAs in the pathogenesis of cancer. *Semin Oncol* 2011; 38:724.

116. Esquela-Kerscher A, Slack FJ. Oncomirs - microRNAs with a role in cancer. *Nat Rev Cancer* 2006; 6:259.
117. Cheng CJ, Slack FJ. The Duality of OncomiR Addiction in the Maintenance and Treatment of Cancer. *Cancer J* 2012; 18:232.
118. Calin GA, Dumitru CD, Shimizu M, et al. Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc Natl Acad Sci U S A* 2002; 99:15524.
119. Cimmino A, Calin GA, Fabbri M, et al. miR-15 and miR-16 induce apoptosis by targeting BCL2. *Proc Natl Acad Sci U S A* 2005; 102:13944.
120. Calin GA, Ferracin M, Cimmino A, et al. A MicroRNA signature associated with prognosis and progression in chronic lymphocytic leukemia. *N Engl J Med* 2005; 353:1793.
121. Garzon R, Volinia S, Liu CG, et al. MicroRNA signatures associated with cytogenetics and prognosis in acute myeloid leukemia. *Blood* 2008; 111:3183.
122. Garzon R, Heaphy CE, Havelange V, et al. MicroRNA 29b functions in acute myeloid leukemia. *Blood* 2009; 114:5331.
123. Costinean S, Zanesi N, Pekarsky Y, et al. Pre-B cell proliferation and lymphoblastic leukemia/high-grade lymphoma in E(mu)-miR155 transgenic mice. *Proc Natl Acad Sci U S A* 2006; 103:7024.
124. Johnson SM, Grosshans H, Shingara J, et al. RAS is regulated by the let-7 microRNA family. *Cell* 2005; 120:635.
125. Lee YS, Dutta A. The tumor suppressor microRNA let-7 represses the HMGA2 oncogene. *Genes Dev* 2007; 21:1025.
126. Yanaihara N, Caplen N, Bowman E, et al. Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. *Cancer Cell* 2006; 9:189.
127. Yu SL, Chen HY, Chang GC, et al. MicroRNA signature predicts survival and relapse in lung cancer. *Cancer Cell* 2008; 13:48.
128. Raponi M, Dossey L, Jatke T, et al. MicroRNA classifiers for predicting prognosis of squamous cell lung cancer. *Cancer Res* 2009; 69:5776.
129. Fanini F, Vannini I, Amadori D, Fabbri M. Clinical implications of microRNAs in lung cancer. *Semin Oncol* 2011; 38:776.
130. Boeri M, Pastorino U, Sozzi G. Role of microRNAs in lung cancer: microRNA signatures in cancer prognosis. *Cancer J* 2012; 18:268.
131. Takamizawa J, Konishi H, Yanagisawa K, et al. Reduced expression of the let-7 microRNAs in human lung cancers in association with shortened postoperative survival. *Cancer Res* 2004; 64:3753.
132. Patnaik SK, Kannisto E, Knudsen S, Yendamuri S. Evaluation of microRNA expression profiles that may predict recurrence of localized stage I non-small cell lung cancer after surgical resection. *Cancer Res* 2010; 70:36.
133. Lebanony D, Benjamin H, Gilad S, et al. Diagnostic assay based on hsa-miR-205 expression distinguishes squamous from nonsquamous non-small-cell lung carcinoma. *J Clin Oncol* 2009; 27:2030.
134. Castañeda CA, Agullo-Ortuño MT, Fresno Vara JA, et al. Implication of miRNA in the diagnosis and treatment of breast cancer. *Expert Rev Anticancer Ther* 2011; 11:1265.
135. Scott GK, Goga A, Bhaumik D, et al. Coordinate suppression of ERBB2 and ERBB3 by enforced expression of micro-RNA miR-125a or miR-125b. *J Biol Chem* 2007; 282:1479.
136. Adams BD, Furneaux H, White BA. The micro-ribonucleic acid (miRNA) miR-206 targets the

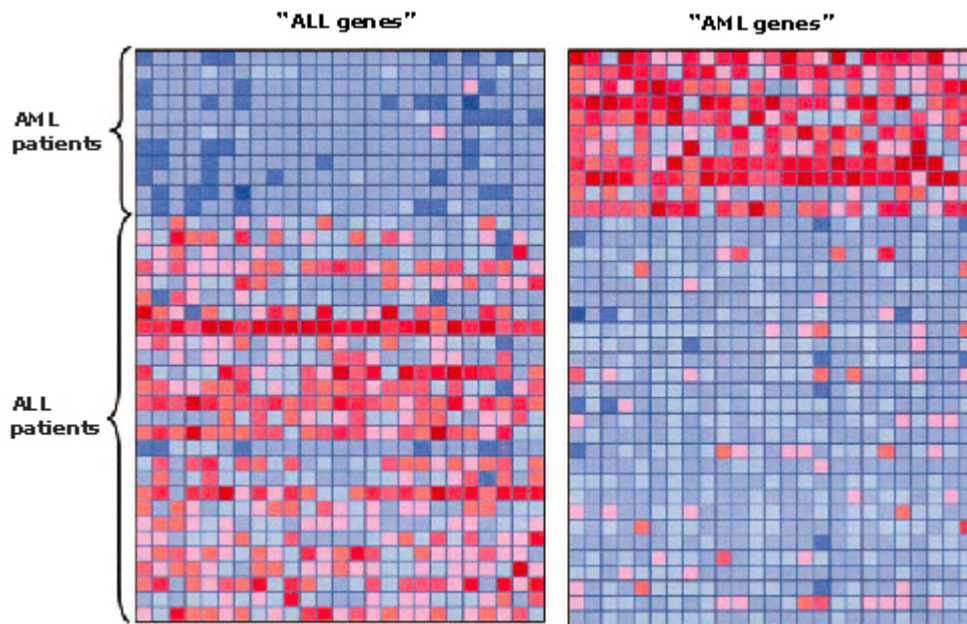
human estrogen receptor-alpha (ERalpha) and represses ERalpha messenger RNA and protein expression in breast cancer cell lines. *Mol Endocrinol* 2007; 21:1132.

137. Ferracin M, Querzoli P, Calin GA, Negrini M. MicroRNAs: toward the clinic for breast cancer patients. *Semin Oncol* 2011; 38:764.
138. Lowery AJ, Miller N, Devaney A, et al. MicroRNA signatures predict oestrogen receptor, progesterone receptor and HER2/neu receptor status in breast cancer. *Breast Cancer Res* 2009; 11:R27.
139. Blenkiron C, Goldstein LD, Thorne NP, et al. MicroRNA expression profiling of human breast cancer identifies new markers of tumor subtype. *Genome Biol* 2007; 8:R214.
140. Van der Auwera I, Limame R, van Dam P, et al. Integrated miRNA and mRNA expression profiling of the inflammatory breast cancer subtype. *Br J Cancer* 2010; 103:532.
141. Yan LX, Huang XF, Shao Q, et al. MicroRNA miR-21 overexpression in human breast cancer is associated with advanced clinical stage, lymph node metastasis and patient poor prognosis. *RNA* 2008; 14:2348.
142. Qian B, Katsaros D, Lu L, et al. High miR-21 expression in breast cancer associated with poor disease-free survival in early stage disease and high TGF-beta1. *Breast Cancer Res Treat* 2009; 117:131.
143. Steeg PS. Cancer: micromanagement of metastasis. *Nature* 2007; 449:671.
144. Ma L, Teruya-Feldstein J, Weinberg RA. Tumour invasion and metastasis initiated by microRNA-10b in breast cancer. *Nature* 2007; 449:682.
145. Nair VS, Maeda LS, Ioannidis JP. Clinical outcome prediction by microRNAs in human cancer: a systematic review. *J Natl Cancer Inst* 2012; 104:528.
146. Sandhu S, Garzon R. Potential applications of microRNAs in cancer diagnosis, prognosis, and treatment. *Semin Oncol* 2011; 38:781.
147. Thorsen SB, Obad S, Jensen NF, et al. The therapeutic potential of microRNAs in cancer. *Cancer J* 2012; 18:275.
148. Weidhaas JB, Babar I, Nallur SM, et al. MicroRNAs as potential agents to alter resistance to cytotoxic anticancer therapy. *Cancer Res* 2007; 67:11111.
149. Trang P, Medina PP, Wiggins JF, et al. Regression of murine lung tumors by the let-7 microRNA. *Oncogene* 2010; 29:1580.
150. Toiyama Y, Takahashi M, Hur K, et al. Serum miR-21 as a diagnostic and prognostic biomarker in colorectal cancer. *J Natl Cancer Inst* 2013; 105:849.
151. Kanaan Z, Rai SN, Eichenberger MR, et al. Plasma miR-21: a potential diagnostic marker of colorectal cancer. *Ann Surg* 2012; 256:544.
152. Gall TM, Frampton AE, Krell J, et al. Blood-based miRNAs as noninvasive diagnostic and surrogate biomarkers in colorectal cancer. *Expert Rev Mol Diagn* 2013; 13:141.

Topic 2821 Version 17.0

GRAPHICS

Gene expression profiling in acute leukemia



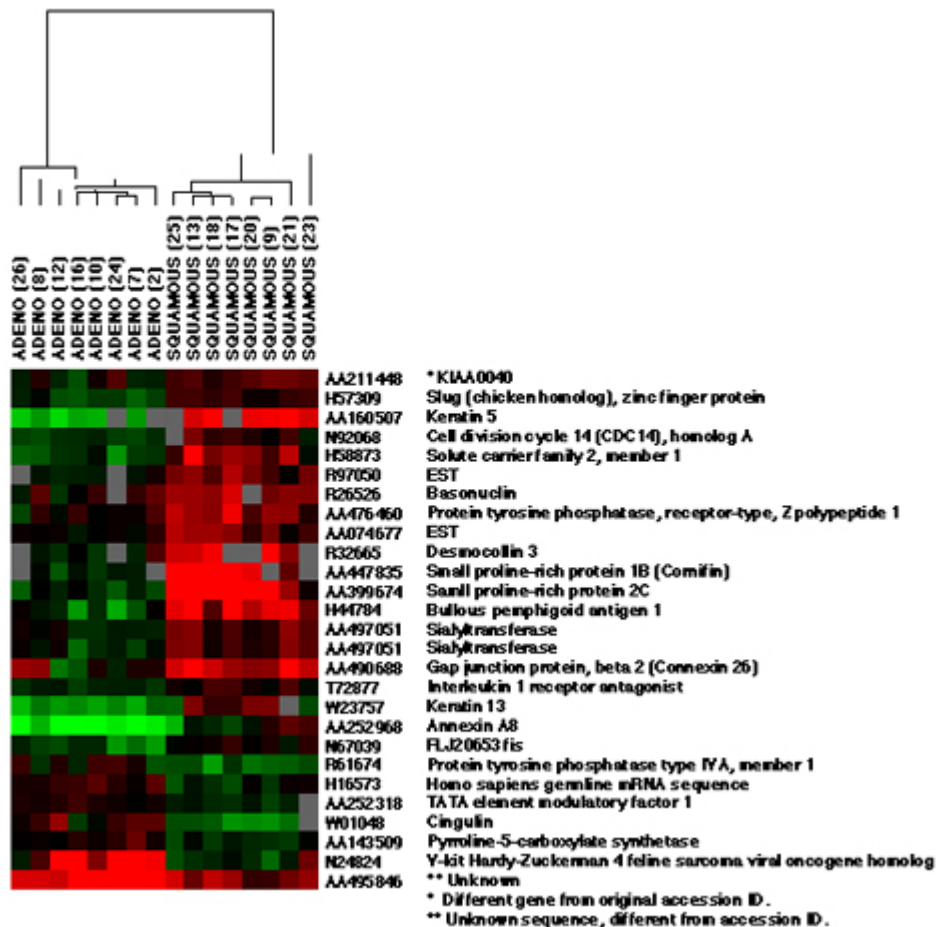
This figure illustrates the use of gene expression arrays (or profiles) for distinguishing between acute myeloid leukemia (AML) and acute lymphoblastic leukemia (ALL). Genes most highly expressed in ALL and AML are shown in the left-hand and right-hand panels, respectively (ie, "ALL genes" and "AML genes", respectively). Expression levels of each of the individual genes greater than the mean are shown in red; those genes expressed below the mean are shown in blue. Each column (ie, top to bottom) corresponds to a specific gene, while each row (ie, left to right) corresponds to expression levels in a single sample. The top 11 samples were from patients with AML, while the remainder were from patients with ALL. Note that, while the genes as a group appear correlated with either AML or ALL, no single gene is uniformly expressed in either type of leukemia.

From: Golub TR, Slonim DK, Tamayo P, et al. Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. Science 1999; 286:531. Reprinted with permission from AAAS.

<http://www.sciencemag.org>

Graphic 74048 Version 9.0

Gene expression profiling in NSCLC: adenocarcinoma vs. squamous carcinoma

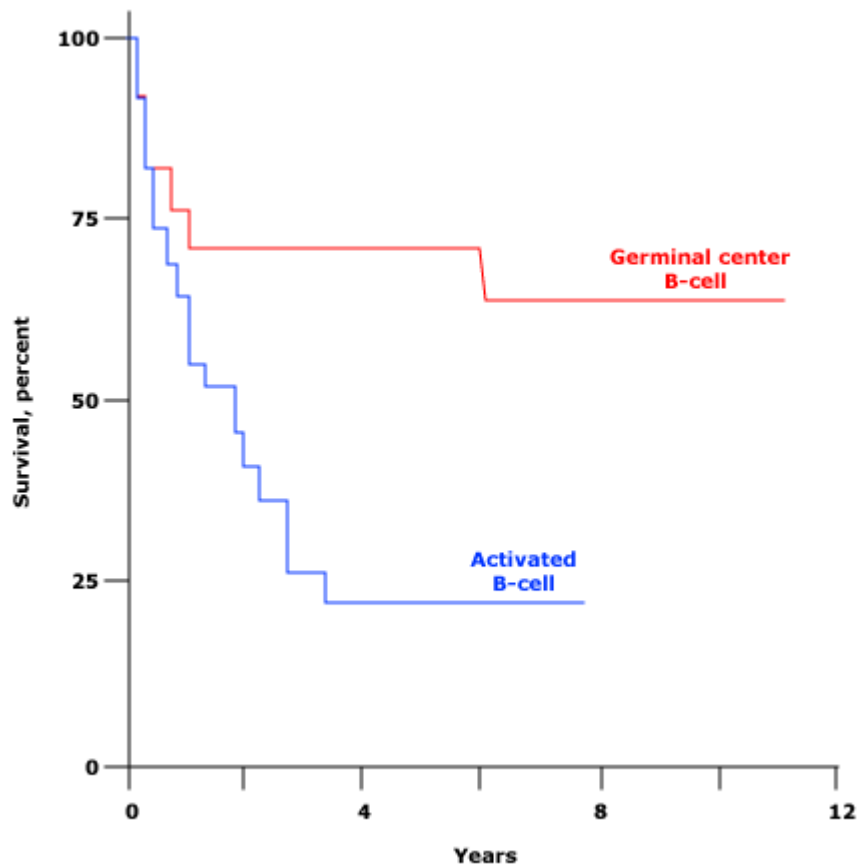


Gene expression profiling, non-small cell lung cancer (NSCLC). This is an example of hierarchical agglomerative clustering analysis. The experiment began by spotting samples of resected untreated NSCLC onto a custom 5000 gene glass slide array. The analysis begins with identifying a single gene pair with the most similar expression across patients or experiments. These genes are then "fused" (subsequently represented by the mean of their two expression levels); this value is then considered a single element in further analyses. The process of "fusing" is repeated until all elements are included in the analysis, and a dendrogram is assembled. This dendrogram resembles an evolutionary tree, where terminal branches depict closely related specimens. Within this single hierarchical tree, the intercluster distance (branch length of the dendrogram) directly correlates with dissimilarity in gene expression. Similarly expressed genes are then ordered next to each other, and visualized as color coded rows (red for overexpressed, green for underexpressed). In this case, the gene expression profiling has been used to separate tumors into adenocarcinoma versus squamous cell histology.

Courtesy of David Carbone, MD, PhD.

Graphic 81774 Version 2.0

Survival in diffuse large B-cell lymphoma according to cell of origin

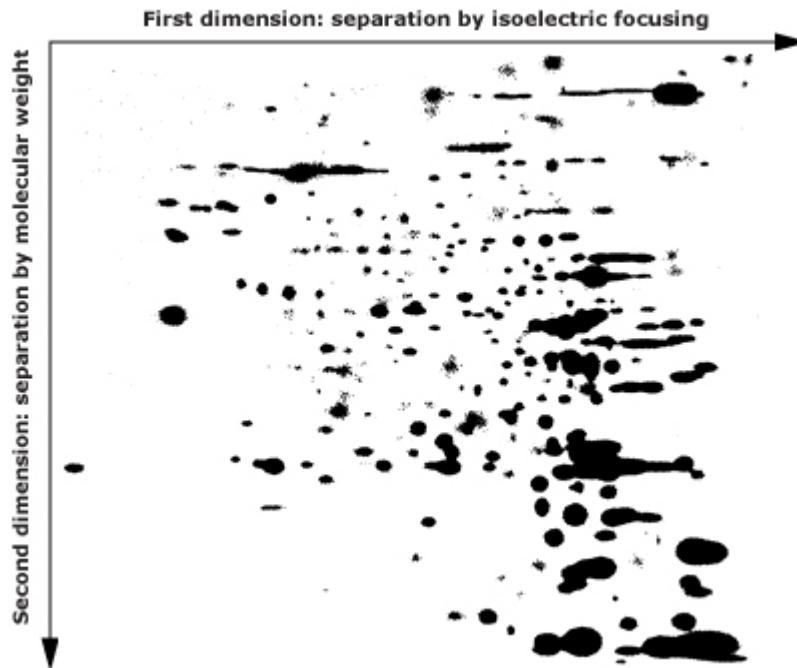


This graph illustrates overall survival in a group of patients with diffuse large B-cell lymphoma (DLBCL) whose cell of origin was determined by means of gene expression profiling (see text). Survival of patients with DLBCL whose malignant cells were thought to arise from a germinal center B-cell (red) was significantly better than that of patients whose cells arose from an activated B-cell (blue).

Data from Alizadeh A, et al. Nature 2000; 403:503.

Graphic 63638 Version 2.0

Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE)



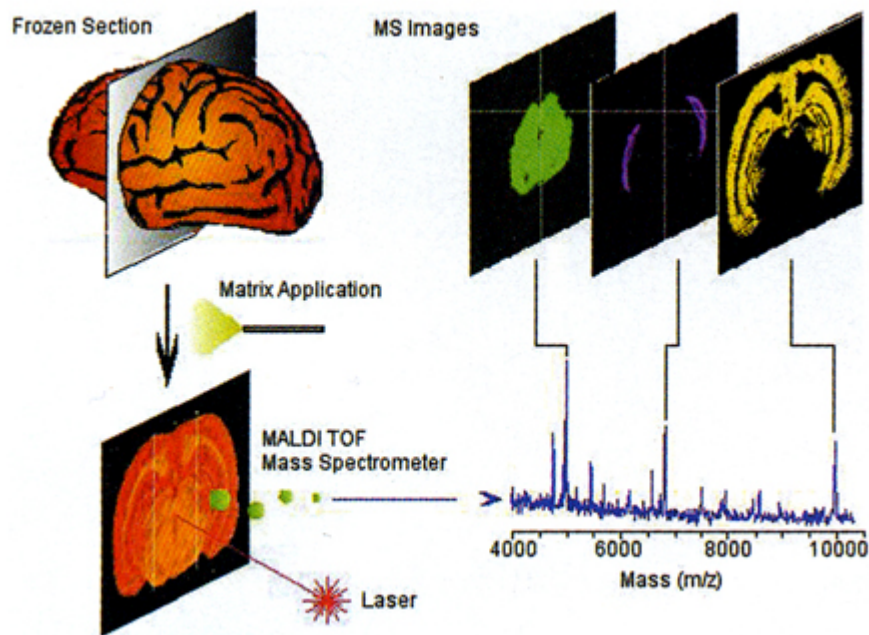
Reference gel showing proteins that are separated using two-dimensional electrophoresis. The proteins are separated first according to their isoelectric point and then secondarily then according to size. The proteins were visualized by staining the gel with a silver stain.

Reproduced with permission from: Dwek MY, Alaiya AA. Proteome analysis enables separate clustering of normal breast, benign breast, and breast cancer tissues. Br J Cancer 2003; 89:305. Copyright © 2003 Nature Publishing Group.

<http://www.nature.com>

Graphic 76724 Version 2.0

Methodology developed for the spatial analysis of tissue by MALDI mass spectrometry



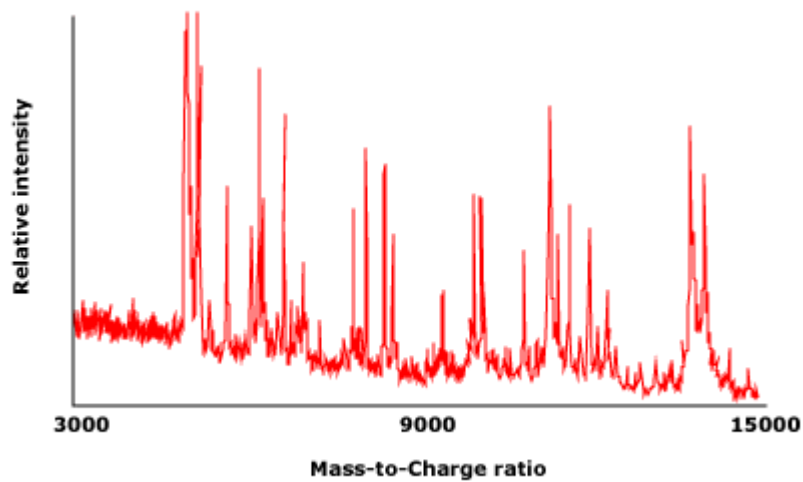
Frozen sections are mounted on a metal plate, coated with a UV-absorbing matrix and placed in the mass spectrometer. A pulsed UV laser desorbs and ionizes analytes from the tissue and their m/z values are determined using a time-of-flight (TOF) analyzer. From a raster over the tissue and measurement of the peak intensities over thousands of spots, mass spectrometric images are generated at specific molecular weight values.

Reproduced with permission from: Stoeckli M, Chaurans P, Hallahan D, Caprioli R. Imaging mass spectrometry: A new technology for the analysis of protein expression in mammalian tissues. Nature Med 2001; 7:493. Copyright © 2001 Nature Publishing Group.

<http://www.nature.com>

Graphic 70974 Version 5.0

Proteomic spectrum of human lung tumor tissue

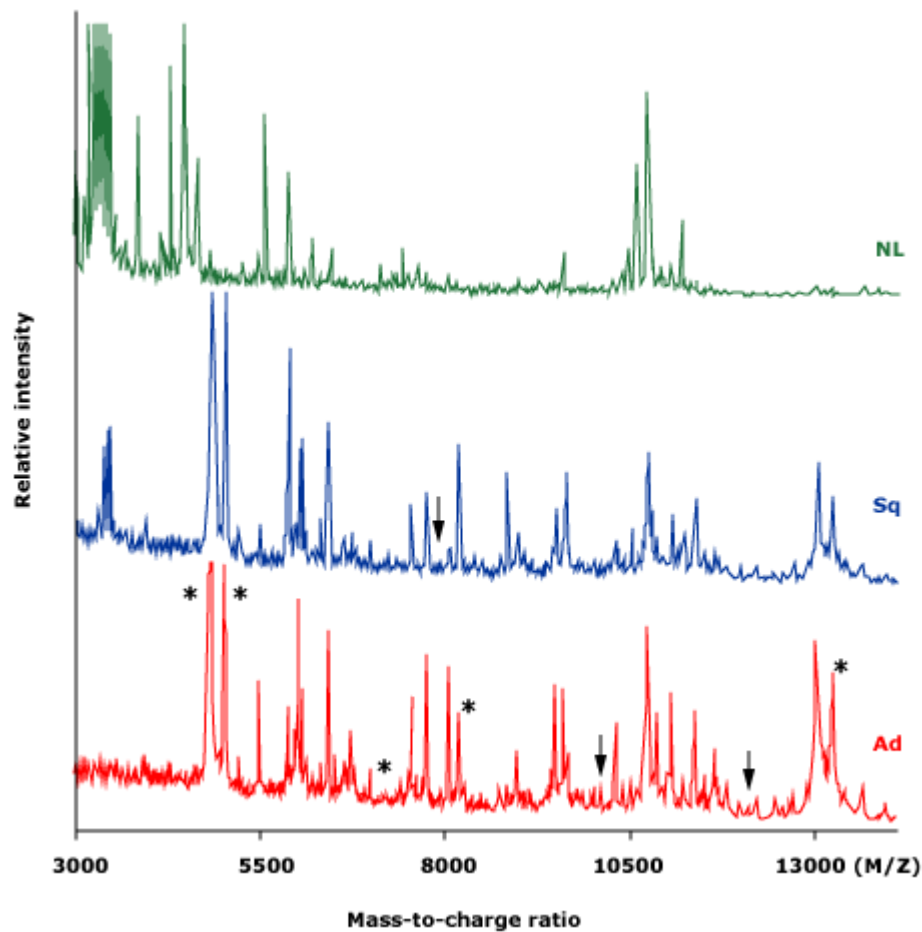


Representative matrix-assisted laser desorption ionization time-of flight mass spectrography (MALDI-TOF MS) protein spectrum obtained directly from human lung tumor tissue.

Courtesy of D Carbone, MD, PhD.

Graphic 79121 Version 1.0

Proteomic spectrum of normal human lung, and tumor tissue from squamous cell and adenocarcinomas



Representative matrix-assisted laser desorption ionization time-of flight mass spectrography (MALDI-TOF MS) protein spectra obtained directly from human lung tissue, demonstrating differences in protein spectra from normal lung (top panel), squamous cell lung cancer (middle panel) and lung adenocarcinoma (bottom panel). Representative spectra obtained from both tumor and normal lung tissue samples are shown with the molecular weight calculation (mass/charge [m/z] values). Examples of the MS peaks identified as discriminatory between normal and tumoral tissue are indicated by asterisks. Differentially expressed peaks between adenocarcinoma and squamous cell carcinoma are indicated by arrowheads.

Courtesy of D. Carbone, MD, PhD.

Graphic 62162 Version 1.0

Disclosures

Disclosures: **Patrick C Ma, MD, MSc** Nothing to disclose. **Benjamin A Raby, MD, MPH** Other Financial Interest (Spouse): Parexel International [Contract research organization (Double-blind randomized clinical trials)]. **Reed E Drews, MD** Nothing to disclose. **Diane MF Savarese, MD** Nothing to disclose.

Contributor disclosures are reviewed for conflicts of interest by the editorial group. When found, these are addressed by vetting through a multi-level review process, and through requirements for references to be provided to support the content.

Appropriately referenced content is required of all authors and must conform to UpToDate standards of evidence.

Conflict of interest policy

Overview of pharmacogenomics

Authors

Kelan Tantisira, MD, MPH
Scott T Weiss, MD, MS

Section Editor

Benjamin A Raby, MD, MPH

Deputy Editors

Diane MF Savarese, MD
Jennifer S Tirnauer, MD

All topics are updated as new evidence becomes available and our [peer review process](#) is complete.

Literature review current through: Feb 2015. | **This topic last updated:** Feb 06, 2015.

INTRODUCTION — There is tremendous interindividual variability in the response to pharmacologic agents [1]. Plasma drug levels can vary more than 1000-fold when the same drug dose is administered to two individuals having approximately the same weight [2]. Drug-drug interactions, drug-food interactions, sex, age, disease state (ie, renal and hepatic function) and pregnancy can all influence variability in drug responses between patients. However, genetic factors are also likely to play a major role, since the individual response to a given pharmacologic agent is highly reproducible [3,4].

Pharmacogenomics is the study of the role of inherited and acquired genetic variation on drug response. It is distinguished from pharmacogenetics, which focuses on individual candidate genes (identified by approaches such as genome-wide association studies (GWAS), genome-wide expression profiling, or methylation studies) to identify markers across the genome that affect drug metabolism, distribution, receptor targets, and biologic effect [5]. However, the two terms are frequently used interchangeably.

In theory, the identification of genetic factors that influence drug absorption, metabolism, and action at the receptor level should allow for individualized therapy; this could optimize drug efficacy and minimize toxicity profiles in a given population [1,5-8]. The potential for cost savings (through increased drug efficacy) and for decreased morbidity and mortality (through increased drug safety and fewer adverse drug reactions [ADRs]) is immense [9-13]. Although many ADRs are preventable and attributed in many cases to human error, others appear idiosyncratic, and potentially influenced by genetic factors. In one study of 2227 ADRs identified in a large teaching hospital, fewer than 50 percent had readily identified causes, and thus, might have been due to pharmacogenetic variability [14].

This topic provides a practical overview of the field of pharmacogenomics, including the historical context, the different ways in which pharmacogenomic differences can affect the response to drug therapy, an overview of US Food and Drug Administration (FDA) approved labeling of medications for which genetic testing is recommended, and the challenges that must be overcome before routine pharmacogenetic testing (genotyping) is fully integrated into mainstream clinical medicine. Throughout, examples of relevant pharmacogenetic associations will be cited.

While the details of all known specific drug-gene interactions are beyond the scope of this overview, current information regarding most of the known associations can be obtained through the National Institute of Health Pharmacogenetics Research Network's PharmGKB: The Pharmacogenetics and the Pharmacogenomics Knowledge Base [15].

HISTORICAL CONTEXT — The roots of pharmacogenomics may date back as early as the 6th century BC, when Pythagoras was said to have recognized that eating fava beans caused illness in some, but not all, individuals [16]. In the 1940s, the immunochemist William Boyd noted that in contrast to Mediterranean populations, native Britons almost never developed hemolytic anemia after ingestion of fava beans; he suggested a genetic difference as the probable explanation [16]. It is now known that the hemolytic anemia associated with ingestion of fava beans, which may also occur with a variety of pharmacologic agents, is due to X-linked glucose-6-phosphate dehydrogenase deficiency (G6PD) [16]. (See "[Genetics and pathophysiology of glucose-6-phosphate dehydrogenase deficiency](#)".)

The following pioneering discoveries ushered in the modern field of pharmacogenetics:

- In the early 1900s, Garrod recognized a subset of psychiatric patients who developed porphyria upon administration of sulphonal [17].
- In 1932, Snyder attributed the 'phenylthiourea non-taster' phenotype to an autosomal recessive trait whose inheritance varied according to ethnicity [18].
- In the 1950s, patients with an inherited deficiency of plasma cholinesterase were noted to have a prolonged paralytic response to suxamethonium [19].
- In 1957, Motulsky first suggested that interindividual differences in drug efficacy as well as ADRs were at least partially attributable to genetic differences [20]. The term "pharmacogenetics" was first coined by Frederick Vogel, a German geneticist, in 1959 [21].
- The role of genetics in the interindividual variation in drug response was reinforced by a series of twin studies conducted in the early 1960s [22-26]. Plasma half-lives of many drugs, including [warfarin](#), were found to be similar in monozygotic twins, but varied significantly among dizygotic twins, siblings, and the general population.
- In the late 1970s, two independent groups associated a deficiency in drug metabolism that was inherited in an autosomal recessive pattern with a markedly higher frequency of side effects in certain volunteers taking the antihypertensive debrisoquine [27] and the antiarrhythmic sparteine [28]. Approximately ten years later, the metabolism of both drugs was shown to result from allelic variation at the cytochrome P450 2D6 (CYP2D6) locus [29]. This was the first example of applied genotyping in the field of pharmacogenetics.

Since then, advances in human genetic technology and an increased understanding of the molecular basis of drug metabolism, transport, and action at the level of its target (often a receptor) has led to the characterization of dozens of mutations and functional pharmacogenetic polymorphisms that influence the response to drug therapy.

WAYS IN WHICH GENETIC FACTORS INFLUENCE DRUG RESPONSE — The heritability of some drug responses was formally established via genetic studies, including the twin studies of Vessel and Page described above [22-26].

However, most drugs have not been formally studied to assess the potential for response variability based upon hereditary characteristics, primarily because it is neither safe nor ethical to administer most medications to individuals in whom they are not indicated. Barring formal genetic studies, the distribution of responses to a certain drug in populations can strongly suggest a genetic component.

The interindividual response distribution to a drug may be either multimodal or unimodal. A multimodal distribution suggests distinct population subgroups of responders, and therefore plausibility of a pharmacogenetic basis for response heterogeneity [3]. A unimodal drug response distribution also supports the possibility of pharmacogenetic variation in response, with subjects demonstrating either a "good" or a "poor" response.

In particular, the combination of significant interindividual variability in level and/or response to a drug plus high repeatability within individual patients (the likelihood of a subsequent beneficial/poor response to a drug given a prior response) strongly suggests the potential for pharmacogenetic effects [3].

Pharmacogenomic influences on drug response have traditionally been divided into four categories based upon the impact of genetic variability on the pharmacologic properties of a drug [30]:

- Effect on drug pharmacokinetics; an example is a genetic variant that alters drug metabolism, affecting plasma concentration.

- Effects on pharmacodynamics; an example is a genetic variation that reduces binding of the drug to its receptor, thereby decreasing therapeutic efficacy.
- Effects on idiosyncratic reactions, such as the likelihood of a hypersensitivity reaction to a certain drug.
- Effects on disease pathogenesis or severity and response to specific therapies; these include specific molecular defects related to the pathogenesis of certain malignancies for which specific targeted therapies have been developed.

Specific examples of medications and specific genetic variants relevant to each of these pharmacogenetic categories are listed in the table ([table 1](#)). The following sections will expand upon these four categories, and provide clinically relevant examples illustrating how pharmacogenetic variability might potentially influence drug response.

Altered pharmacokinetics — Pharmacokinetics refers to the study of the transport and metabolism of administered drugs, including absorption, distribution, tissue localization, biotransformation, and excretion [[31](#)]. The vast majority of pharmacogenetic investigations reported in the literature have been in this area, with the major areas of study being genetic differences in drug metabolizing enzymes and drug transporters.

In humans, there are two phases of xenobiotic (ie, substances foreign to the body) metabolism that are controlled by several hundred drug metabolizing enzymes ([figure 1](#)), and their corresponding genes:

- During phase I metabolism, polar groups are added to lipophilic molecules by oxidation, reduction, or hydrolysis to facilitate water-solubility. This group of reactions is catalyzed predominantly by the cytochrome P450 superfamily of mixed function oxidases (CYPs), which represent a classic example of polymorphic (ie, occurring in different types) drug metabolizing enzymes. (See "[Drugs and the liver: Metabolism and mechanisms of injury](#)", section on 'Phase I reactions'.)
- Following phase I metabolism, most compounds are still insufficiently hydrophilic for excretion, and require further processing (phase II reactions result) to form readily excretable, nontoxic substances. An example of a polymorphic phase II drug metabolizing enzyme is thiopurine-S-methyltransferase (TPMT), which is involved in the metabolism of [azathioprine](#) and [6-mercaptopurine](#) (6-MP) [[32](#)]. (See "[Drugs and the liver: Metabolism and mechanisms of injury](#)", section on 'Phase II reactions'.)

CYP isoenzymes and drug metabolism — The cytochrome P450s (CYPs) are members of a superfamily of oxidative enzymes, which represent the major system for oxidative metabolism of therapeutic substances. Sequencing of the human genome has revealed 58 different human CYP genes (information on human CYP P450 genes available online at <http://drnelson.uthsc.edu/cytochromeP450.html>) [[33](#)].

The majority of the CYP genes encoding for enzymes that are active in xenobiotic metabolism are polymorphic, and polymorphisms that affect drug metabolism are seen in a significant portion of the population [[27,34-36](#)]. In many cases, the frequency varies according to ethnic background. As an example, expression of CYP2D6 is functionally absent in 7 percent of Caucasians and African-Americans, while deficiency is rare among Asians [[37](#)]. Due to the importance of these genes, a website has been created that is continuously updated for newly identified CYP polymorphisms [[38](#)].

Polymorphisms in CYP genes may contribute to either diminished or absent metabolism, or excessive metabolism of a compound [[39](#)]. In general, genotypic CYP variations result in three metabolic phenotypes: ultra-rapid metabolizers, extensive (normal) metabolizers, and poor metabolizers.

The clinically most important polymorphic variations in hepatic CYPs are seen in the CYP2C9,

CYP2C19, CYP2D6, and CYP3A4 genes, which code for enzymes that are responsible for the phase I metabolism of approximately 60 to 70 percent of all therapeutic medications used in humans.

Clopidogrel and CYP2C19 variants — The adverse in vivo effects of reduced function CYP2C19 alleles on benefit from the anti-platelet agent [clopidogrel](#) have been demonstrated in a number of studies and a meta-analysis [40]. In March 2010, a boxed warning from the US Food and Drug Administration was issued to alert clinicians that genetic testing is available to identify individuals with poor metabolizer variants of CYP2C19 who may not receive the full benefits of clopidogrel and require dose adjustment or use of a different drug. Two subsequent meta-analyses, however, which included more recently published reports and had more stringent definitions of adverse events, did not show a correlation between CYP2C19 genotype and adverse cardiovascular events in patients treated with clopidogrel [41,42]. These data have not led to a change in the US Food and Drug Administration (FDA) recommendation. Although guidelines for CYP2C19 genotype-directed antiplatelet therapy are available from the [Clinical Pharmacogenetics Implementation Consortium](#) [43], many experts do not recommend **routine** testing of patients for "clopidogrel resistance" by genetic testing for CYP2C19 poor metabolizers. This subject is discussed in detail elsewhere. (See "[Clopidogrel resistance and clopidogrel treatment failure](#)", section on '[Variation in clopidogrel metabolism](#)'.)

CYP2D6 variants — Genetic polymorphisms have been particularly well studied in the CYP2D6 subfamily, which is responsible for the metabolism of a wide variety of drugs such as [codeine](#), [nortriptyline](#), [metoprolol](#), the selective serotonin reuptake inhibitors (SSRIs), [simvastatin](#), and [tamoxifen](#) [44,45]. (See "[Serotonin-norepinephrine reuptake inhibitors \(SNRIs\): Pharmacology, administration, and side effects](#)" and "[Adjuvant endocrine therapy for non-metastatic, hormone receptor-positive breast cancer](#)".)

CYP2D6 is highly polymorphic, with over 90 known allelic variants [15]. The following example illustrates the markedly different impact that the same genetic variant in CYP2D6 can have on the pharmacokinetic profile of two drugs, [codeine](#) and [nortriptyline](#).

Both [codeine](#) and [nortriptyline](#) are metabolized by CYP2D6; however, codeine is a prodrug and its analgesic properties are not manifest until it is metabolized by CYP2D6, primarily to [morphine](#) and codeine-6-glucuronide. In contrast, nortriptyline is the active moiety and its metabolism results in inactivation of the drug to its primary metabolite.

At conventional doses, subjects who are poor metabolizers based upon CYP2D6 genotype will derive no therapeutic benefit from [codeine](#) (because it will not be converted to its active moiety), but they may be "overdosed" with [nortriptyline](#), and at increased risk of side effects ([figure 2](#)).

Conversely, at conventional doses of [codeine](#), subjects who are ultra-rapid metabolizers based upon CYP2D6 genotype have higher than expected [morphine](#) levels (an initial "overdose"), with more side effects and a shorter than expected duration of pain control [46]. On the other hand, these patients may derive no significant therapeutic benefit from [nortriptyline](#) because of excessive metabolism of the drug.

Given the evolving knowledge of pharmacogenomics and its impact on pharmacokinetics, preliminary dosing recommendations for several drugs have been made based upon known genetic variations in drug metabolizing enzymes ([figure 3](#)) [47,48]. However, these have not been validated by prospective studies. Although guidelines for [codeine](#) therapy based upon CYP2D6 phenotype are available from the [Clinical Pharmacogenetics Implementation Consortium](#) [49], routine assay to identify specific CYP2D6 genotypes as a means of choosing a dose for any of these compounds is not yet considered standard practice.

Thiopurine methyltransferase — Polymorphisms in the CYP isoenzymes represent the most common genetic variant affecting pharmacokinetics, but there are others. As an example, thiopurine S-methyltransferase (TPMT, OMIM #187680) is responsible for the metabolism of the class of

therapeutic compounds called thiopurines (eg, [azathioprine](#), 6-[mercaptopurine](#) [6-MP]). Polymorphisms in the TPMT gene can result in functional inactivation or markedly decreased activity of the enzyme, and an increased risk of treatment-related leukopenia [50-53]. Over 24 low-functioning genetic variants have been identified, but the two most common (TPMT*2 and *3) account for more than 95 percent of defective TPMT activity in patients [32,54,55]. Approximately 10 percent of people have reduced TPMT activity and 0.3 percent (1 in 300) have no detectable levels [56,57].

TPMT testing is not specifically recommended by the US Food and Drug Administration prior to treatment with a thiopurine ([azathioprine](#), [mercaptopurine](#), [thioguanine](#)). In the case of [azathioprine](#), they recommend that "consideration be given to either genotype or phenotype patients for TPMT." For individuals with low or absent TPMT activity, dose reductions of up to 90 percent may be needed, based upon experience in pediatric acute lymphocytic leukemia [58-60]. Specific dosing recommendations for thiopurines according to TPMT phenotype were made by the [Clinical Pharmacogenetics Implementation Consortium](#) in 2011 [61] and reiterated in their 2013 update [62].

Prospective testing of TPMT genotype for dose adjustments has been shown to be effective at reducing toxicity of 6-MP without compromising efficacy [63]. However, TPMT genotyping has not been universally adopted, and the cost effectiveness and optimal clinical circumstances in which to perform routine testing are not well defined; the available data on cost-effectiveness are conflicting [64-66]. Furthermore, a systematic review of the use of TPMT genotyping in the treatment of patients with chronic inflammatory disease found imprecision in estimates of genotyping sensitivity and insufficient evidence of outcomes effectiveness [67].

Expert opinions differ regarding the role of TPMT genotyping prior to the administration of thiopurines for treatment of inflammatory and autoimmune disorders; some advocate routine testing while others, citing the low frequency of homozygous variants among Caucasians (only about 1 in 300) and the fact that the majority of patients who develop myelosuppression while taking [azathioprine](#) do not have detectable TPMT gene mutations, disagree with this approach. Some clinicians, in particular those treating acute leukemia with 6-MP, only perform TPMT genotyping if there is unexpectedly severe or prolonged myelosuppression. (See "[6-mercaptopurine \(6-MP\) metabolite monitoring and TPMT testing in the treatment of inflammatory bowel disease with 6-MP or azathioprine](#)", section on 'TPMT genotyping' and "[Pharmacology and side effects of azathioprine when used in rheumatic diseases](#)", section on '[Pharmacogenetics and TPMT testing](#)' and "[Chronic immunomodulating therapies for myasthenia gravis](#)", section on '[Azathioprine](#)' and "[Post-remission therapy for Philadelphia chromosome negative acute lymphoblastic leukemia in adults](#)", section on '[Maintenance therapy](#)'.)

Drug transport — Inherited variations in drug transport proteins make up the second major area of pharmacogenetic investigation in the area of pharmacokinetics. The potential clinical relevance of inherited differences in membrane drug transport proteins can be illustrated by the following examples.

- **Efficacy of non-nucleoside reverse transcriptase inhibitors** — Members of the adenosine triphosphate (ATP)-binding cassette (ABC) family of membrane transporters are among the most extensively studied in terms of drug disposition and modulating drug action at the cellular level [4]. One member, the P-glycoprotein, is encoded by the human ABCB1 gene (also called the multidrug resistant 1 or MDR1 gene, OMIM 171050). P-glycoprotein enhances the energy-dependent cellular efflux of many substrates, such as bilirubin, and several medications, including many antiretroviral agents [4]. Several studies have examined whether certain single nucleotide polymorphisms (SNPs) within the ABCB1 gene may predict either plasma concentrations of drug, virologic efficacy, or drug toxicity in subsets of HIV-infected patients [68-71]. While each of these studies has demonstrated differences related to one particular variant (MDR1 3435C to T), the mechanistic basis for these associations remains to be clarified.

Some data also suggest that genotypic differences in drug transport across membranes may influence the ability to achieve high cerebrospinal fluid (CSF) drug levels, although the clinical significance of this finding (ie, whether rates of HIV-related dementia are lower in patients with these genetic variants) is unclear [68]. (See "[Pharmacology of non-nucleoside reverse transcriptase inhibitors](#)".)

- **Sunitinib-related toxicity** — In a study of 219 patients treated with sunitinib, a multitargeted tyrosine kinase inhibitor used for the treatment of advanced renal cell cancer, the prevalence and severity of several side effects, including mucosal inflammation and hand-foot syndrome, were significantly higher in patients who had inherited specific haplotypes of the ABCB1 and ABCB2 efflux transporter genes [72].
- **Solute carrier organic anion transporter family polymorphisms and high-dose [methotrexate](#)** — Plasma concentrations are quite variable following administration of high-dose methotrexate (HD-MTX) for treatment of leukemia, lymphoma, and osteosarcoma. A variety of inherited genetic polymorphisms affect MTX pharmacokinetics, toxicity, and efficacy [73]. As an example, the SLCO1B1 (solute carrier organic anion transporter family, member 1B1, OMIM 604843) gene encodes organic anion transporting polypeptide 1B1 (OATP1B1). This peptide is localized at the membrane of hepatocytes and mediates uptake of substrates from sinusoidal blood, resulting in their net excretion from the blood, likely via biliary excretion. Inherited polymorphisms in the SLCO1B1 gene and OATP1B1 levels appear to be important determinants of plasma methotrexate (MTX) levels and toxicity following administration of high-dose MTX [73-76]. (See "[Therapeutic use and toxicity of high-dose methotrexate](#)".)

Variants in SLCO2B1 (solute carrier organic anion transporter family, member 2B1, OMIM 604988) have also been associated with variation in [montelukast](#) levels and with asthmatic response to montelukast [77]. (See "[Agents affecting the 5-lipoxygenase pathway in the treatment of asthma](#)", section on '[Heterogeneity of response to controller therapies](#)'.)

Altered pharmacodynamics — Pharmacodynamics is the study of the biochemical and physiologic consequences of drug administration (ie, the effect of a drug at its therapeutic target and at other nontarget sites) [31]. Genetic variation may lead to interindividual differences in therapeutic response despite the presence of appropriate concentrations of the drug at its intended target. These variations can modulate drug response by affecting the drug target itself or one of the downstream components in the target's mechanistic pathway.

While studies of pharmacogenetic predictors of efficacy at therapeutic target sites may eventually become the primary basis for "individualized therapy," there are relatively few clinical examples of replicable pharmacogenomic associations related to pharmacodynamics, compared to pharmacokinetics [78]. There are at least three reasons for the prominence of pharmacokinetic over pharmacodynamic pharmacogenetic studies:

- The biology of drug metabolism and drug transport is relatively straightforward, and each compound typically has one principal enzyme responsible for its metabolism. In contrast, the physiology of most drug target pathways is fairly complex, providing multiple venues that may require investigation prior to the discovery of an interaction between the drug and genetic differences within the target pathway.
- Variability in drug metabolism related to genetic factors (eg, the difference in drug levels seen in ultra-rapid versus slow metabolizers because of CYP variants) can be as high as a 10,000-fold difference, whereas differences in target binding related to genetics are generally less than 20-fold [79,80]. The resultant ability (power) to detect differences in drug metabolism is clearly far greater than the ability to detect variance in drug targets.

- The pattern of inheritance permits easier identification of pharmacokinetic variants. Typically, the inheritance pattern underlying the pharmacogenetic effects related to differences in drug metabolism or transport is Mendelian (eg X-linked inheritance underlying G6PD deficiency). (See ["Overview of Mendelian inheritance"](#).)

In contrast, the inheritance patterns of pharmacodynamic pharmacogenetic relationships usually demonstrate relationships consistent with the complex traits underlying the purpose of the drug. Thus, these target responses tend to be determined by polygenic or gene-environment interactions, both of which are much more difficult to identify.

Warfarin and VKORC1 polymorphisms — An example of a genetic variant that influences pharmacodynamics is the effect of polymorphisms in the gene encoding vitamin K epoxide reductase complex (VKORC1) on response to the anticoagulant [warfarin](#), a drug with a narrow therapeutic index and a wide variability in individual dosing. VKORC1 variants account for approximately 25 percent of the phenotypic variability in warfarin dosing [81]. (See ["Lack of cost-effectiveness analyses"](#) below.)

VKORC1 is responsible for the conversion of Vitamin K-epoxide to Vitamin K, which is the rate-limiting step in the physiological process of Vitamin K recycling and the primary therapeutic target for [warfarin](#). Common polymorphisms within the VKORC1 gene appear to modulate the mean daily dose of warfarin required to achieve the targeted intensity of anticoagulation. In a study on the effect of VKORC1 polymorphisms in 297 patients undergoing warfarin anticoagulation, a low-dose haplotype group (group A) and a high-dose haplotype group (non-A or group B) were identified, with the following mean warfarin maintenance doses [81]:

- Group A/A — 2.7 ± 0.2 mg/day
- Group A/B — 4.9 ± 0.2 mg/day
- Group B/B — 6.2 ± 0.3 mg/day

Genetic testing for detecting variants of the VKORC1 genes is available to help clinicians assess whether a patient may be especially sensitive to [warfarin](#), and require a lower starting dose; they also test for genetic variants in CYP2C9 that influence warfarin metabolism. However, routine genotyping of patients prior to starting warfarin is not widely accepted (or recommended in guidelines from the American College of Chest Physicians [82]). This position is supported by the results of three randomized trials, which found that pharmacogenetic-based as compared to clinically-guided dosing had either no or at best marginal usefulness in improving clinical outcomes (time within a therapeutic range of the International Normalized Ratio [INR] or excess bleeding). (See ["Therapeutic use of warfarin and other vitamin K antagonists"](#), section on 'Initial dose' and ["Therapeutic use of warfarin and other vitamin K antagonists"](#), section on 'Use of genotyping' and ["Vitamin K and the synthesis and function of gamma carboxyglutamic acid"](#), section on 'Mutations and polymorphisms of the VKOR complex'.)

Effect on idiosyncratic reactions — An idiosyncratic reaction is an adverse drug reaction (ADRs) that cannot be anticipated based upon the known drug target. Three examples in which an idiosyncratic reaction to a drug varies according to genetic factors are [abacavir](#) hypersensitivity and HLA-B*5701, TCL1A variants and musculoskeletal side effects from aromatase inhibitors, and HLA-A*3101 or B*1502 and [carbamazepine](#) hypersensitivity.

- [Abacavir](#) is a nucleoside analog with potent activity against HIV; however, a proportion of patients develop severe hypersensitivity reactions to this drug. Abacavir hypersensitivity is associated with carriage of the major histocompatibility complex (MHC) class I allele HLA-B*5701. It is theorized that abacavir is metabolized to an aldehyde-reactive metabolite, which undergoes classical MHC class 1 processing. Presentation of the peptide-HLA complex on an antigen-presenting cell to the receptor of an abacavir-specific CD8+ T-cell activates the release inflammatory cytokines, resulting in the clinical syndrome of acute hypersensitivity syndrome.

Several trials have demonstrated the improved safety of drug administration with prior immunogenetic screening for HLA-B*5701, and guidelines from expert groups (including the [Clinical Pharmacogenetics Implementation Consortium, CPIC \[83\]](#)) endorse the use of abacavir only in patients who have tested negative for HLA-B*5701. (See "[Abacavir hypersensitivity reaction](#)".)

- Inherited variants in the T-cell leukemia 1A (TCL1A) gene have been associated with the risk of musculoskeletal side effects in women receiving an aromatase inhibitor (AI) for treatment of breast cancer [84]. However, the magnitude of the excess risk is small, and only a small fraction (approximately 11 percent) of the clinically significant musculoskeletal adverse events from AIs may be attributable to this particular genetic variant [85]. (See "[Adjuvant endocrine therapy for non-metastatic, hormone receptor-positive breast cancer](#)".)
- [Carbamazepine](#), an anticonvulsant and mood stabilizing drug, causes Stevens-Johnson syndrome and the related toxic epidermal necrolysis, idiosyncratic reactions that may be reduced in frequency by avoiding the drug in patients carrying one of two HLA alleles, B*1502 (in Asian populations) and A*3101 (in Europeans). Testing of Asian populations for the B*1502 allele before using the related drug oxcarbamazepine has also been suggested by US FDA ([table 1](#)). (See "[Antiepileptic drugs: Mechanism of action, pharmacology, and adverse effects](#)", section on 'Role of HLA testing' and "[Antiepileptic drugs: Mechanism of action, pharmacology, and adverse effects](#)", section on 'Oxcarbazepine'.)

Disease pathogenesis — The final category of drug response that may be influenced by genetics is that certain genetic variations can influence disease pathogenesis, the underlying severity of a disease, and response to specific therapies. As an example, cystic fibrosis (CF) is a multisystem disorder that is caused by mutations in the cystic fibrosis transmembrane conductance regulator (CFTR) gene, which regulates chloride and water transport in the lungs, digestive tract, and elsewhere. The G551D mutation, which is present in about 5 percent of CF patients, interferes with activation of the CFTR chloride channel.

[Ivacaftor](#), an orally available CFTR modulator that is specific for the G551D mutation, restores functioning of the CFTR protein in patients who harbor this mutation. In a clinical trial leading to its approval in the United States, subjects with CF and a G551D mutation in at least one of their CFTR genes experienced improved FEV₁, less frequent pulmonary exacerbations, and greater weight gain with ivacaftor compared with those receiving placebo [86]. All individuals with CF should undergo genotyping to determine whether they carry the G551D mutation. (See "[Cystic fibrosis: Overview of the treatment of lung disease](#)", section on 'CFTR modulators'.)

US FOOD AND DRUG ADMINISTRATION AND USE OF GENOMIC BIOMARKERS TO GUIDE

THERAPY — The US Food and Drug Administration (FDA) has had a longstanding interest in the field of pharmacogenomics. In August 2008, as part of its Critical Path Initiative, the scientific process through which a medical product is transformed from discovery to development, the US FDA released a table listing genomic biomarkers that have established roles in determining drug response. A full listing of valid genomic biomarkers in the context of drug labeling in the United States is available online [87], and a synopsis of selected drugs is provided in the table ([table 1](#)).

Pharmacogenomic information is currently contained in about 10 percent of labels for drugs approved by the US FDA. Although most of the US FDA labeling recommendations have involved variants that influence drug pharmacokinetics, there are examples within each of the pharmacogenetic categories described above of medications for which the FDA recommends testing prior to administration ([table 1](#)). (See '[Altered pharmacodynamics](#)' above and "[Systemic chemotherapy for metastatic colorectal cancer: Completed clinical trials](#)", section on 'Pharmacokinetic variability'.)

Although the US FDA has officially approved a few tests, the major contribution of the agency in the field of pharmacogenetics has been in the updating of drug labels to contain information on pharmacogenomic issues that are applicable to a given pharmacologic agent. [Warfarin](#) holds a unique place in the current recommendations, as it is the only pharmacologic agent for which testing for two independent genetic variants (in the CYP2C9 and VKORC1 genes) is recommended. CYP2C9 is the primary enzyme involved in the metabolism of warfarin, while polymorphisms within the VKORC1 gene appear to modulate the mean daily dose of warfarin required to acquire target anticoagulation intensity. Thus, the warfarin recommendations incorporate testing of variants involved in both the pharmacokinetics and pharmacodynamics of warfarin. (See '[Warfarin and VKORC1 polymorphisms](#)' above.)

CHALLENGES TO WIDESPREAD USE OF GENOTYPING — Despite the labeling changes implemented by the US Food and Drug Administration (FDA) and the increasing number of pharmacogenetic studies in the published literature, integration of pharmacogenetic testing into clinical care has been slow, and many of the tests recommended by the US FDA for individual drugs are not in routine use. Among the reasons for lack of implementation are:

- Limitations in the design of published pharmacogenetic studies (in particular, the lack of prospective randomized trials demonstrating improved clinical outcomes when drug therapy or specific dose is selected on the basis of genotype)
- Regulatory and ethical concerns
- Lack of cost effectiveness analyses
- Limitations in the number of available pharmacogenetic tests and lack of guidelines for test implementation
- A lack of education on the benefits of pharmacogenetic testing, both for patients and providers

Study design limitations — Several problems in study design have limited the translation of pharmacogenetics into the clinical sphere.

The initially reported genetic and pharmacogenetic associations often have not been reproducible (ie, many of the initial reports represented "false-positive" findings) [88,89]. In many cases, the original pharmacogenetic studies were "spin-offs" from clinical trials and, thus, were underpowered for genetic association studies given the small sample sizes allocated to a given treatment arm. Replication of results from genome-wide association studies requires identification of a large and appropriate sample, and this can be both difficult and expensive [5].

Another contributor to the failure to replicate initial findings is heterogeneity among studies, which can be phenotypic (ie, the selective use of different endpoints, times of assessment, types of interventions, and genetic groups across studies) or genotypic (ie, evaluation of disparate markers within the same gene or varying genetic allele frequencies due to ethnicity, which may be overt or subtle).

The problems of phenotypic and genotypic heterogeneity can be illustrated by an evaluation of the pharmacogenetic associations of the beta2-adrenergic receptor (ADRB2, OMIM# 109690) gene with the response to beta2-agonist therapy in asthma [90]. In a review of 21 studies that focused on the two most common ADRB2 coding variants (Arg16Gly and Gln27Glu), the total number of statistically significant probed and reported associations was 487 when the multiple endpoints and types of comparisons presented by the various studies were considered [90]. However, 465 (95 percent) were probed only once; six associations were probed twice, and only two of the associations were probed five times for the same endpoint, time of assessment, type of intervention, and genetic group. When the data were reanalyzed, adjusting for the number of comparisons, only 38 of the original 487 associations (8 percent) were statistically significant for the comparison between all available genetic groups. Heterogeneity in

the available data precludes a definitive conclusion regarding the utility of ADRB2 variants to guide asthma therapy. (See ["Beta-2 adrenergic receptor dysfunction and polymorphism in asthma", section on 'Receptor polymorphisms'.](#))

In addition to the study design problems of small sample size and phenotypic/genotypic heterogeneity, the identified variants themselves may not be amenable to clinical application. Most pharmacogenetic variants identified to date are either rare traits with relatively strong phenotypic effects, or common traits that have relatively weak phenotypic effects.

Two clinically relevant examples illustrate these issues:

- Thiopurine s-methyltransferase (TPMT) is responsible for the metabolism of the thiopurines [azathioprine](#) and [mercaptopurine](#) (6-MP). Genetic variations in the TPMT gene can result in functional inactivation of the enzyme, and a markedly increased risk of life-threatening treatment-associated myelosuppression. (See ['Thiopurine methyltransferase'](#) above.)
- A common haplotype in the corticotropin releasing hormone receptor 1 (CRHR1, OMIM# 122561) gene has been associated with an improved lung function response to inhaled corticosteroids in patients with asthma [91]. Following eight weeks of inhaled corticosteroid therapy, adults and children who were homozygous for the haplotype (27 percent of the population) demonstrated two to three times the FEV1 improvement than those who were homozygous for other haplotypes ([figure 4](#)).

However, despite the high frequency of the haplotype, the overall proportion of the variability in FEV1 response that was explained by the haplotype was under 5 percent. Therefore, genotyping of CRHR1 in isolation would be insufficient as a predictor of outcomes with inhaled corticosteroid usage.

Regulatory and ethical concerns — Although pharmacogenetic information is now included on about 10 percent of labels for drugs approved by the US FDA, questions remain regarding the regulation of genotyping tests, and the extent to which pharmacogenetic analyses should be incorporated into each phase of new drug development [92,93]. From an ethics perspective, a longstanding concern has been whether the identification of pharmacogenetic variants (in particular those associated with a poor treatment response or prognosis) in an individual could result in stigmatization (eg, denial of insurance) [92].

A major step in the protection of individuals' rights came with the passage of H.R. 493, the Genetic Information Nondiscrimination Act of 2008 [94]. This law protects Americans against discrimination based on their genetic information in matters related to health insurance and employment and should translate into increased acceptance of pharmacogenetic testing by the public in the future.

Lack of cost-effectiveness analyses — Circumstances that favor cost effectiveness of a pharmacogenetic test include a high prevalence of the genetic variant of interest in the target population, a good correlation between phenotype and genotype, satisfactory diagnostic test criteria, a disease that is associated with significant morbidity or mortality if left untreated, and a significant reduction in adverse drug reactions resulting from testing [95].

Despite the multitude of pharmacogenetic association studies in the literature, relatively few cost-effective analyses have been performed. These studies are crucial to determine reimbursement for routine pharmacogenetic testing, but their performance is problematic for two reasons [95]:

- There are only limited data addressing how often pharmacogenetic testing actually prevents **clinically significant** adverse drug reactions [96-98].
- The price of pharmacogenetic tests is likely to drop continuously over the next few years.

The first issue can be illustrated by pharmacogenetic testing for CYP2C9 and VKORC1 variants in patients being started on the anticoagulant [warfarin](#) (see '[Warfarin and VKORC1 polymorphisms](#)' above). VKORC1 variants account for approximately 25 percent of the phenotypic variability in warfarin dosing, with CYP2C9 variants adding an additional 6 to 10 percent [81]. The available data suggest that combining knowledge of genetic variants with clinical characteristics can explain between 55 and 71 percent of the variability in warfarin dosing [99,100].

A 2006 working paper from the American Enterprise Institute-Brookings Joint Center for Regulatory Studies (and partially published in 2008) estimated that formally integrating genetic testing into routine [warfarin](#) therapy could allow American warfarin users to avoid 85,000 serious bleeding events and 17,000 strokes annually, resulting in estimated savings of \$1.1 billion annually, with a range of about \$100 million to \$2 billion [96,97]. Nonetheless, routine genotyping is not widely accepted (or recommended in guidelines from the American College of Chest Physicians [82]) because of the lack of evidence from prospective randomized trials that pharmacogenomic-based individualized dosing actually improves clinical outcomes [101,102].

Furthermore, a cost-effectiveness study of using pharmacogenomic information for [warfarin](#) dosing concluded that routine genotyping before warfarin dosing was **unlikely** to be cost-effective for typical patients with nonvalvular atrial fibrillation (ie, with the estimated cost-effectiveness of testing exceeding \$170,000 per quality-adjusted life-year gained) (QALY) [103]. The authors concluded that on the basis of available data and cost of testing (about \$400 in 2007), there was only a 10 percent chance that genotype-guided dosing is likely to be cost-effective (ie, <\$50,000 per QALY).

Lack of tests and guidelines for test implementation — When compared to other clinical tests, pharmacogenetic testing is relatively precise because of the stability of DNA in the laboratory, the fact that germline DNA does not change in an individual's lifetime, and the robust nature of the available genetic tests [102]. Nevertheless, relatively few pharmacogenetic tests are available for use in clinical practice, and even when tests are available, there has been slow adoption of these tests to inform clinical decision making in practice [5].

Limitations cited in the development and implementation of new tests include the difficulty of identifying and incorporating multiple (interacting) variants in the same test, the translation of a test developed based upon average population responses to the case of a specific individual, the need to develop tests that can be interpreted within a clinical context, the need for a test associated with biologically meaningful differences, the need for tests to undergo formal regulatory approval (eg, Clinical Laboratory Improvement Amendment (CLIA) in the United States) [102], and the lack of clear, peer-reviewed evidence-based guidelines that translate laboratory test results into actionable prescribing decisions for specific drugs.

Need to educate patients and providers — Pharmacogenomics and the promise of personalized medicine is frequently mentioned in the popular lay press. It is therefore incumbent upon both the developers of a given pharmacogenetic test as well as the health care professionals responsible for ordering the test to be cognizant of the test characteristics and interpretation, and to be able to effectively disseminate that information to patients. As noted above, a crucial part of testing efforts should be directed toward the meaning of the Genetic Information Nondiscrimination Act of 2008, providing reassurance regarding the safety and privacy of testing. (See '[Personalized medicine](#)' and '[Regulatory and ethical concerns](#)' above.)

There is also a need for educating health care providers in both the broad array of potential predictive tools and the strengths and weaknesses of these tools. An expanding array of markers will be increasingly available, including genetic variants, genomic, proteomic, and other molecular biomarkers. The context for these educational efforts is not yet clear.

The use of biomarkers in clinical practice may lead a paradigm shift in patient management. Providers will potentially need to perform a diagnostic test prior to prescribing a drug to determine which drug is best suited to each individual patient. Of paramount importance, providers will need to be reassured and cognizant that performing these tests will not and cannot replace sound clinical judgment.

CLINICAL PHARMACOGENETICS IMPLEMENTATION CONSORTIUM — Adoption of pharmacogenomic testing in clinical practice has been slow for a variety of reasons, including those discussed above. However, some argue that the available evidence linking routine use of pharmacogenomic analysis to reduced disease risk is sufficient to warrant clinical implementation now [104]. In addition to the enhanced coverage of the US Food and Drug Administration (FDA) in pharmacogenomic labeling of drugs, which now includes boxed warnings for several drugs (table 1), and several cost-effectiveness and “real world” efficacy studies have begun to overcome the limitations cited above, suggesting that implementation of pharmacogenomics testing will likely be incorporated into clinical practice in the future [87,98,105]. On the other hand, others disagree, stating that convincing data to link biomarker testing to patient outcomes do not exist in the majority of cases, and that inclusion of biomarker testing recommendations in drug labels is premature [106].

Leading pharmacogenetic investigators, from the NIH-sponsored Pharmacogenetics Research Network and elsewhere, have begun to publish a series of articles on clinical implementation ([Clinical Pharmacogenetics Implementation Consortium](#) [CPIC] Guidelines) of pharmacogenomics testing [49,61,107-114]. However, clinical guidelines from this group largely focus on HOW pharmacogenomic testing can be integrated into clinical practice rather than WHETHER they should or should not be used to make treatment decisions in individual patients. Recommendations on whether testing is justified in clinical practice still emanate predominantly from subspecialty-based expert clinical groups (eg, the American College of Chest Physicians).

SUMMARY — Genetic polymorphisms for many drug metabolizing enzymes and drug targets (eg, receptors) have been identified and probably contribute to interpatient variability in drug response. There are four general mechanisms by which genetic factors can influence the response to pharmacologic agents:

- Effect on drug pharmacokinetics
- Effects on pharmacodynamics
- Effect on idiosyncratic reactions
- Effect on disease pathogenesis, the underlying severity of a disease, or the ability to respond to a specific therapy

Pharmacogenetic testing is available in some areas in conjunction with certain drug classes, and may enable physicians to understand why patients respond differently to various drugs and to make better decisions about therapy. However, the goal of “individualized therapy” based upon pharmacogenetic testing has yet to be realized. Despite the promise of a growing body of research relating to pharmacogenetics and its impact on drug response, and US Food and Drug Administration (FDA) guidelines as to the use of genetic markers to guide therapy for a variety of agents (table 1), use of these tests is not widespread with a few notable exceptions:

- Some molecularly targeted agents used for treatment of specific cancers are restricted to those tumors that display certain genetic features (eg, use of the anti HER2 monoclonal antibody [trastuzumab](#) is restricted to breast cancers that overexpress HER2). (See ['Ways in which genetic factors influence drug response'](#) above and ["HER2 and predicting response to therapy in breast cancer"](#).)
- Some molecularly targeted agents for cystic fibrosis are restricted to patients who harbor specific mutations in the cystic fibrosis transmembrane regulator gene (eg, [ivacaftor](#) for patients with the

G551D mutation). (See '[Disease pathogenesis](#)' above.)

- Guidelines from expert groups endorse the use of the anti-HIV drug [abacavir](#) only in patients who have tested negative for HLA-B*5701 (see '[Effect on idiosyncratic reactions](#)' above).
- Before starting treatment with [azathioprine](#) or 6-[mercaptopurine](#), many (but not all) clinicians advocate screening patients for mutations in the thiopurine methyltransferase (TPMT) gene that cause TPMT deficiency. (See '[Thiopurine methyltransferase](#)' above and '[6-mercaptopurine \(6-MP\) metabolite monitoring and TPMT testing in the treatment of inflammatory bowel disease with 6-MP or azathioprine](#)'.)
- In March 2010, a new boxed warning from the US Food and Drug Administration for the antiplatelet agent [clopidogrel](#) was issued to alert clinicians that genetic testing (using the Roche AmpliChip Cytochrome P450 Genotyping test) is available to identify individuals with poor metabolizer variants of CYP2C19 who may not receive the full benefits of the drug. However, two separate meta-analyses have come to opposite conclusions regarding the influence of CYP2C19 genotype on adverse cardiovascular events in patients treated with clopidogrel. It remains to be seen whether genetic testing will be implemented in clinical practice. (See '[Clopidogrel and CYP2C19 variants](#)' above and '[Clopidogrel resistance and clopidogrel treatment failure](#)', section on '[Variation in clopidogrel metabolism](#)'.)

Numerous barriers exist to the direct application of pharmacogenomics advances in knowledge to drug therapy in the context of clinical care, which will need to be overcome before personalized drug therapy becomes a routine component of mainstream medicine.

Use of UpToDate is subject to the [Subscription and License Agreement](#).

REFERENCES

1. Leufkens HG. The interface between pharmacoepidemiology and pharmacogenetics. *Eur J Pharmacol* 2000; 410:121.
2. Ingelman-Sundberg M. Pharmacogenetics: an opportunity for a safer and more efficient pharmacotherapy. *J Intern Med* 2001; 250:186.
3. Drazen JM, Silverman EK, Lee TH. Heterogeneity of therapeutic responses in asthma. *Br Med Bull* 2000; 56:1054.
4. Evans WE, McLeod HL. Pharmacogenomics--drug disposition, drug targets, and side effects. *N Engl J Med* 2003; 348:538.
5. Wang L, McLeod HL, Weinshilboum RM. Genomics and drug response. *N Engl J Med* 2011; 364:1144.
6. Vesell ES. Therapeutic lessons from pharmacogenetics. *Ann Intern Med* 1997; 126:653.
7. Evans WE, Relling MV. Pharmacogenomics: translating functional genomics into rational therapeutics. *Science* 1999; 286:487.
8. Mancinelli L, Cronin M, Sadée W. Pharmacogenomics: the promise of personalized medicine. *AAPS PharmSci* 2000; 2:E4.
9. Lazarou J, Pomeranz BH, Corey PN. Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. *JAMA* 1998; 279:1200.
10. Kongkaew C, Noyce PR, Ashcroft DM. Hospital admissions associated with adverse drug reactions: a systematic review of prospective observational studies. *Ann Pharmacother* 2008; 42:1017.

11. Ernst FR, Grizzle AJ. Drug-related morbidity and mortality: updating the cost-of-illness model. *J Am Pharm Assoc (Wash)* 2001; 41:192.
12. Johnson JA, Bootman JL. Drug-related morbidity and mortality. A cost-of-illness model. *Arch Intern Med* 1995; 155:1949.
13. Meisel C, Gerloff T, Kirchheiner J, et al. Implications of pharmacogenetics for individualizing drug treatment and for study design. *J Mol Med (Berl)* 2003; 81:154.
14. Classen DC, Pestotnik SL, Evans RS, et al. Adverse drug events in hospitalized patients. Excess length of stay, extra costs, and attributable mortality. *JAMA* 1997; 277:301.
15. National Institute of Health Pharmacogenetics Research Network's PharmGKB: The Pharmacogenetics and the Pharmacogenomics Knowledge Base. www.pharmgkb.org (Accessed on July 13, 2011).
16. Davidson RG. Pharmacogenetics: Pharmacogenomics, PDQ Medical Genetics (Ed), Elsevier, Oxford 2002. p.232.
17. Garrod AE. The incidence of alkaptonuria: a study in chemical individuality. *Lancet* 1902; 160:1616.
18. Snyder LH. Studies in human inheritance. IX. The inheritance of taste deficiency in man. *Ohio J Sci* 1932; 32:436.
19. Kalow W. Familial incidence of low pseudocholinesterase level. *Lancet* 1956; 271:576.
20. MOTULSKY AG. Drug reactions enzymes, and biochemical genetics. *J Am Med Assoc* 1957; 165:835.
21. Vogel F. Moderne Probleme der Humangenetik. *Ergeb Inn Med Kinderheilkd* 1959; 12:52.
22. Vesell ES, Page JG. Genetic control of drug levels in man: phenylbutazone. *Science* 1968; 159:1479.
23. Vesell ES, Page JG. Genetic control of drug levels in man: antipyrine. *Science* 1968; 161:72.
24. Vesell ES, Page JG. Genetic control of dicumarol levels in man. *J Clin Invest* 1968; 47:2657.
25. Vesell ES, Page JG. Genetic control of the phenobarbital-induced shortening of plasma antipyrine half-lives in man. *J Clin Invest* 1969; 48:2202.
26. Vesell ES, Passananti GT, Greene FE, Page JG. Genetic control of drug levels and of the induction of drug-metabolizing enzymes in man: individual variability in the extent of allopurinol and nortriptyline inhibition of drug metabolism. *Ann N Y Acad Sci* 1971; 179:752.
27. Mahgoub A, Idle JR, Dring LG, et al. Polymorphic hydroxylation of Debrisoquine in man. *Lancet* 1977; 2:584.
28. Eichelbaum M, Spannbrucker N, Steincke B, Dengler HJ. Defective N-oxidation of sparteine in man: a new pharmacogenetic defect. *Eur J Clin Pharmacol* 1979; 16:183.
29. Gonzalez FJ, Skoda RC, Kimura S, et al. Characterization of the common genetic defect in humans deficient in debrisoquine metabolism. *Nature* 1988; 331:442.
30. Silverman ES, Hjoberg J, Palmer LJ, et al. Application of Pharmacogenetics to the Therapeutics of Asthma. In: *Therapeutic Targets of Airway Inflammation*, Eissa NT, Huston D (Eds), Marcel Dekker, New York 2003. Vol 177, p.1000.
31. Steimer W, Potter JM. Pharmacogenetic screening and therapeutic drugs. *Clin Chim Acta* 2002; 315:137.
32. Yates CR, Krynetski EY, Loennechen T, et al. Molecular diagnosis of thiopurine S-methyltransferase deficiency: genetic basis for azathioprine and mercaptopurine intolerance. *Ann Intern Med* 1997; 126:608.
33. Nelson DR. The cytochrome p450 homepage. *Hum Genomics* 2009; 4:59.

34. Tanaka E. Update: genetic polymorphism of drug metabolizing enzymes in humans. *J Clin Pharm Ther* 1999; 24:323.
35. Hasler JA. Pharmacogenetics of cytochromes P450. *Mol Aspects Med* 1999; 20:12.
36. Dorne JL, Walton K, Renwick AG. Human variability in xenobiotic metabolism and pathway-related uncertainty factors for chemical risk assessment: a review. *Food Chem Toxicol* 2005; 43:203.
37. Mizutani T. PM frequencies of major CYPs in Asians and Caucasians. *Drug Metab Rev* 2003; 35:99.
38. Website for CYP alleles available online at www.cypalleles.ki.se (Accessed on July 13, 2011).
39. Ueshima Y, Tsutsumi M, Takase S, et al. Acetaminophen metabolism in patients with different cytochrome P-4502E1 genotypes. *Alcohol Clin Exp Res* 1996; 20:25A.
40. Mega JL, Simon T, Collet JP, et al. Reduced-function CYP2C19 genotype and risk of adverse clinical outcomes among patients treated with clopidogrel predominantly for PCI: a meta-analysis. *JAMA* 2010; 304:1821.
41. Bauer T, Bouman HJ, van Werkum JW, et al. Impact of CYP2C19 variant genotypes on clinical efficacy of antiplatelet treatment with clopidogrel: systematic review and meta-analysis. *BMJ* 2011; 343:d4588.
42. Holmes MV, Perel P, Shah T, et al. CYP2C19 genotype, clopidogrel metabolism, platelet function, and cardiovascular events: a systematic review and meta-analysis. *JAMA* 2011; 306:2704.
43. Scott SA, Sangkuhl K, Gardner EE, et al. Clinical Pharmacogenetics Implementation Consortium guidelines for cytochrome P450-2C19 (CYP2C19) genotype and clopidogrel therapy. *Clin Pharmacol Ther* 2011; 90:328.
44. Lee WM. Drug-induced hepatotoxicity. *N Engl J Med* 1995; 333:1118.
45. Marez D, Legrand M, Sabbagh N, et al. Polymorphism of the cytochrome P450 CYP2D6 gene in a European population: characterization of 48 mutations and 53 alleles, their frequencies and evolution. *Pharmacogenetics* 1997; 7:193.
46. Kelly LE, Rieder M, van den Anker J, et al. More codeine fatalities after tonsillectomy in North American children. *Pediatrics* 2012; 129:e1343.
47. Brockmöller J, Kirchheiner J, Meisel C, Roots I. Pharmacogenetic diagnostics of cytochrome P450 polymorphisms in clinical drug development and in drug treatment. *Pharmacogenomics* 2000; 1:125.
48. Kirchheiner J, Brøsen K, Dahl ML, et al. CYP2D6 and CYP2C19 genotype-based dose recommendations for antidepressants: a first step towards subpopulation-specific dosages. *Acta Psychiatr Scand* 2001; 104:173.
49. Crews KR, Gaedigk A, Dunnenberger HM, et al. Clinical Pharmacogenetics Implementation Consortium guidelines for cytochrome P450 2D6 genotype and codeine therapy: 2014 update. *Clin Pharmacol Ther* 2014; 95:376.
50. Evans WE, Horner M, Chu YQ, et al. Altered mercaptopurine metabolism, toxic effects, and dosage requirement in a thiopurine methyltransferase-deficient child with acute lymphocytic leukemia. *J Pediatr* 1991; 119:985.
51. Lennard L, Van Loon JA, Weinshilboum RM. Pharmacogenetics of acute azathioprine toxicity: relationship to thiopurine methyltransferase genetic polymorphism. *Clin Pharmacol Ther* 1989; 46:149.
52. Zelinkova Z, Derijks LJ, Stokkers PC, et al. Inosine triphosphate pyrophosphatase and thiopurine S-methyltransferase genotypes relationship to azathioprine-induced myelosuppression. *Clin Gastroenterol Hepatol* 2006; 4:44.
53. Evans WE, Hon YY, Bomgaars L, et al. Preponderance of thiopurine S-methyltransferase

- deficiency and heterozygosity among patients intolerant to mercaptopurine or azathioprine. *J Clin Oncol* 2001; 19:2293.
54. Schaeffeler E, Fischer C, Brockmeier D, et al. Comprehensive analysis of thiopurine S-methyltransferase phenotype-genotype correlation in a large population of German-Caucasians and identification of novel TPMT variants. *Pharmacogenetics* 2004; 14:407.
 55. Coulthard SA, Howell C, Robson J, Hall AG. The relationship between thiopurine methyltransferase activity and genotype in blasts from patients with acute leukemia. *Blood* 1998; 92:2856.
 56. Krynetski EY, Tai HL, Yates CR, et al. Genetic polymorphism of thiopurine S-methyltransferase: clinical importance and molecular mechanisms. *Pharmacogenetics* 1996; 6:279.
 57. McLeod HL, Lin JS, Scott EP, et al. Thiopurine methyltransferase activity in American white subjects and black subjects. *Clin Pharmacol Ther* 1994; 55:15.
 58. Andersen JB, Szumlanski C, Weinshilboum RM, Schmiegelow K. Pharmacokinetics, dose adjustments, and 6-mercaptopurine/methotrexate drug interactions in two patients with thiopurine methyltransferase deficiency. *Acta Paediatr* 1998; 87:108.
 59. Lennard L, Gibson BE, Nicole T, Lilleyman JS. Congenital thiopurine methyltransferase deficiency and 6-mercaptopurine toxicity during treatment for acute lymphoblastic leukaemia. *Arch Dis Child* 1993; 69:577.
 60. Evans WE, Johnson JA. Pharmacogenomics: the inherited basis for interindividual differences in drug response. *Annu Rev Genomics Hum Genet* 2001; 2:9.
 61. Relling MV, Gardner EE, Sandborn WJ, et al. Clinical Pharmacogenetics Implementation Consortium guidelines for thiopurine methyltransferase genotype and thiopurine dosing. *Clin Pharmacol Ther* 2011; 89:387.
 62. Relling MV, Gardner EE, Sandborn WJ, et al. Clinical pharmacogenetics implementation consortium guidelines for thiopurine methyltransferase genotype and thiopurine dosing: 2013 update. *Clin Pharmacol Ther* 2013; 93:324.
 63. Relling MV, Pui CH, Cheng C, Evans WE. Thiopurine methyltransferase in acute lymphoblastic leukemia. *Blood* 2006; 107:843.
 64. Payne K, Newman W, Fargher E, et al. TPMT testing in rheumatology: any better than routine monitoring? *Rheumatology (Oxford)* 2007; 46:727.
 65. Donnan JR, Ungar WJ, Mathews M, et al. A cost effectiveness analysis of thiopurine methyltransferase testing for guiding 6-mercaptopurine dosing in children with acute lymphoblastic leukemia. *Pediatr Blood Cancer* 2011; 57:231.
 66. Veenstra DL, Higashi MK, Phillips KA. Assessing the cost-effectiveness of pharmacogenomics. *AAPS PharmSci* 2000; 2:E29.
 67. Booth RA, Ansari MT, Loit E, et al. Assessment of thiopurine S-methyltransferase activity in patients prescribed thiopurines: a systematic review. *Ann Intern Med* 2011; 154:814.
 68. Fellay J, Marzolini C, Meaden ER, et al. Response to antiretroviral treatment in HIV-1-infected individuals with allelic variants of the multidrug resistance transporter 1: a pharmacogenetics study. *Lancet* 2002; 359:30.
 69. Haas DW, Bartlett JA, Andersen JW, et al. Pharmacogenetics of nevirapine-associated hepatotoxicity: an Adult AIDS Clinical Trials Group collaboration. *Clin Infect Dis* 2006; 43:783.
 70. Haas DW, Smeaton LM, Shafer RW, et al. Pharmacogenetics of long-term responses to antiretroviral regimens containing Efavirenz and/or Nelfinavir: an Adult Aids Clinical Trials Group Study. *J Infect Dis* 2005; 192:1931.
 71. Tozzi V, Libertone R, Liuzzi G. HIV pharmacogenetics in clinical practice: recent achievements and future challenges. *Curr HIV Res* 2008; 6:544.

72. van Erp NP, Eechoute K, van der Veldt AA, et al. Pharmacogenetic pathway analysis for determination of sunitinib-induced toxicity. *J Clin Oncol* 2009; 27:4406.
73. Radtke S, Zolk O, Renner B, et al. Germline genetic variations in methotrexate candidate genes are associated with pharmacokinetics, toxicity, and outcome in childhood acute lymphoblastic leukemia. *Blood* 2013; 121:5145.
74. Treviño LR, Shimasaki N, Yang W, et al. Germline genetic variation in an organic anion transporter polypeptide associated with methotrexate pharmacokinetics and clinical effects. *J Clin Oncol* 2009; 27:5972.
75. Lopez-Lopez E, Martin-Guerrero I, Ballesteros J, et al. Polymorphisms of the SLCO1B1 gene predict methotrexate-related toxicity in childhood acute lymphoblastic leukemia. *Pediatr Blood Cancer* 2011; 57:612.
76. Zhang HN, He XL, Wang C, et al. Impact of SLCO1B1 521T > C variant on leucovorin rescue and risk of relapse in childhood acute lymphoblastic leukemia treated with high-dose methotrexate. *Pediatr Blood Cancer* 2014; 61:2203.
77. Mougey EB, Feng H, Castro M, et al. Absorption of montelukast is transporter mediated: a common variant of OATP2B1 is associated with reduced plasma concentrations and poor response. *Pharmacogenet Genomics* 2009; 19:129.
78. Johnson JA, Lima JJ. Drug receptor/effector polymorphisms and pharmacogenetics: current status and challenges. *Pharmacogenetics* 2003; 13:525.
79. Meyer UA. Pharmacogenetics: the slow, the rapid, and the ultrarapid. *Proc Natl Acad Sci U S A* 1994; 91:1983.
80. Nebert DW, Ingelman-Sundberg M, Daly AK. Genetic epidemiology of environmental toxicity and cancer susceptibility: human allelic polymorphisms in drug-metabolizing enzyme genes, their functional importance, and nomenclature issues. *Drug Metab Rev* 1999; 31:467.
81. Rieder MJ, Reiner AP, Gage BF, et al. Effect of VKORC1 haplotypes on transcriptional regulation and warfarin dose. *N Engl J Med* 2005; 352:2285.
82. Ansell J, Hirsh J, Hylek E, et al. Pharmacology and management of the vitamin K antagonists: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition). *Chest* 2008; 133:160S.
83. Martin MA, Klein TE, Dong BJ, et al. Clinical pharmacogenetics implementation consortium guidelines for HLA-B genotype and abacavir dosing. *Clin Pharmacol Ther* 2012; 91:734.
84. Ingle JN, Schaid DJ, Goss PE, et al. Genome-wide associations and functional genomic studies of musculoskeletal adverse events in women receiving aromatase inhibitors. *J Clin Oncol* 2010; 28:4674.
85. Offit K, Robson ME. New pharmacogenomic paradigm in breast cancer treatment. *J Clin Oncol* 2010; 28:4665.
86. Ramsey BW, Davies J, McElvaney NG, et al. A CFTR potentiator in patients with cystic fibrosis and the G551D mutation. *N Engl J Med* 2011; 365:1663.
87. US Department of Health and Human Services FaDA, Table of valid genomic biomarkers in the context of approved drug labels. Available online at <http://www.fda.gov/Drugs/ScienceResearch/ResearchAreas/Pharmacogenetics/ucm083378.htm> (Accessed on July 13, 2011).
88. Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K. A comprehensive review of genetic association studies. *Genet Med* 2002; 4:45.
89. Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG. Replication validity of genetic association studies. *Nat Genet* 2001; 29:306.
90. Contopoulos-Ioannidis DG, Alexiou GA, Gouvas TC, Ioannidis JP. An empirical evaluation of

multifarious outcomes in pharmacogenetics: beta-2 adrenoceptor gene polymorphisms in asthma treatment. *Pharmacogenet Genomics* 2006; 16:705.

91. Tantisira KG, Lake S, Silverman ES, et al. Corticosteroid pharmacogenetics: association of sequence variants in CRHR1 with improved lung function in asthmatics treated with inhaled corticosteroids. *Hum Mol Genet* 2004; 13:1353.
92. Roden DM, Altman RB, Benowitz NL, et al. Pharmacogenomics: challenges and opportunities. *Ann Intern Med* 2006; 145:749.
93. Kirchheiner J, Fuhr U, Brockmöller J. Pharmacogenetics-based therapeutic recommendations--ready for clinical practice? *Nat Rev Drug Discov* 2005; 4:639.
94. Hudson KL, Holohan MK, Collins FS. Keeping pace with the times--the Genetic Information Nondiscrimination Act of 2008. *N Engl J Med* 2008; 358:2661.
95. Swen JJ, Huizinga TW, Gelderblom H, et al. Translating pharmacogenomics: challenges on the road to the clinic. *PLoS Med* 2007; 4:e209.
96. McWilliam A, Lutter R, Nardinelli C. Healthcare impact of personalized medicine using genetic testing: an exploratory analysis for warfarin. *Personalized Medicine* 2008; 5:279.
97. www.healthanddna.com/warfarinsavings.pdf?pid=1127 (Accessed on July 13, 2011).
98. Reese ES, Daniel Mullins C, Beitelshes AL, Onukwugha E. Cost-effectiveness of cytochrome P450 2C19 genotype screening for selection of antiplatelet therapy with clopidogrel or prasugrel. *Pharmacotherapy* 2012; 32:323.
99. Sconce EA, Khan TI, Wynne HA, et al. The impact of CYP2C9 and VKORC1 genetic polymorphism and patient characteristics upon warfarin dose requirements: proposal for a new dosing regimen. *Blood* 2005; 106:2329.
100. Wadelius M, Pirmohamed M. Pharmacogenetics of warfarin: current status and future challenges. *Pharmacogenomics J* 2007; 7:99.
101. Garcia DA. Warfarin and pharmacogenomic testing: the case for restraint. *Clin Pharmacol Ther* 2008; 84:303.
102. Weiss ST, McLeod HL, Flockhart DA, et al. Creating and evaluating genetic tests predictive of drug response. *Nat Rev Drug Discov* 2008; 7:568.
103. Eckman MH, Rosand J, Greenberg SM, Gage BF. Cost-effectiveness of using pharmacogenetic information in warfarin dosing for patients with nonvalvular atrial fibrillation. *Ann Intern Med* 2009; 150:73.
104. Altman RB. Pharmacogenomics: "noninferiority" is sufficient for initial implementation. *Clin Pharmacol Ther* 2011; 89:348.
105. Epstein RS, Moyer TP, Aubert RE, et al. Warfarin genotyping reduces hospitalization rates results from the MM-WES (Medco-Mayo Warfarin Effectiveness study). *J Am Coll Cardiol* 2010; 55:2804.
106. Wang B, Canestaro WJ, Choudhry NK. Clinical evidence supporting pharmacogenomic biomarker testing provided in US Food and Drug Administration drug labels. *JAMA Intern Med* 2014; 174:1938.
107. Wilke RA, Ramsey LB, Johnson SG, et al. The clinical pharmacogenomics implementation consortium: CPIC guideline for SLCO1B1 and simvastatin-induced myopathy. *Clin Pharmacol Ther* 2012; 92:112.
108. Hershfield MS, Callaghan JT, Tassaneeyakul W, et al. Clinical Pharmacogenetics Implementation Consortium guidelines for human leukocyte antigen-B genotype and allopurinol dosing. *Clin Pharmacol Ther* 2013; 93:153.
109. Leckband SG, Kelsoe JR, Dunnenberger HM, et al. Clinical Pharmacogenetics Implementation Consortium guidelines for HLA-B genotype and carbamazepine dosing. *Clin Pharmacol Ther* 2013;

94:324.

110. Scott SA, Sangkuhl K, Stein CM, et al. Clinical Pharmacogenetics Implementation Consortium guidelines for CYP2C19 genotype and clopidogrel therapy: 2013 update. *Clin Pharmacol Ther* 2013; 94:317.
111. Caudle KE, Thorn CF, Klein TE, et al. Clinical Pharmacogenetics Implementation Consortium guidelines for dihydropyrimidine dehydrogenase genotype and fluoropyrimidine dosing. *Clin Pharmacol Ther* 2013; 94:640.
112. Martin MA, Hoffman JM, Freimuth RR, et al. Clinical Pharmacogenetics Implementation Consortium Guidelines for HLA-B Genotype and Abacavir Dosing: 2014 update. *Clin Pharmacol Ther* 2014; 95:499.
113. Ramsey LB, Johnson SG, Caudle KE, et al. The clinical pharmacogenetics implementation consortium guideline for SLCO1B1 and simvastatin-induced myopathy: 2014 update. *Clin Pharmacol Ther* 2014; 96:423.
114. Clancy JP, Johnson SG, Yee SW, et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) guidelines for ivacaftor therapy in the context of CFTR genotype. *Clin Pharmacol Ther* 2014; 95:592.

Topic 2904 Version 32.0

GRAPHICS**Various genomic biomarkers in the context of US Food and Drug Administration (FDA)-approved drug labels**

Biomarker	Drug	Synopsis of drug label information	Label recommends testing prior to instituting therapy
ALK (anaplastic lymphoma kinase) gene rearrangements	Crizotinib	Crizotinib selectively inhibits ALK tyrosine kinase. ALK gene rearrangements, which include the echinoderm microtubule-associated protein-like 4-AKL (EML4-ALK) fusion oncogene, may result in expression of oncogenic fusion proteins and increased tumor growth. Detection of ALK-positive NSCLC is necessary for selection of patients for treatment with crizotinib.	Yes
BRAF ^{V600} , BRAF ^{V600E} mutations	Vemurafenib	Vemurafenib inhibits BRAF kinase, which inhibits tumor growth in melanomas that have specific types of mutated BRAF at the V600 site. Confirmation of a V600 BRAF mutation is required for selection of patients for treatment of melanoma with vemurafenib.	Yes
c-KIT expression	Imatinib	Imatinib inhibits proliferation and induces apoptosis in gastrointestinal	Yes

		stromal tumors (GIST) cells that express an activating c-KIT mutation.	
CCR5 (Chemokine C-C motif receptor 5)	Maraviroc	Maraviroc is indicated for treatment-experienced adult patients infected with only CCR5-tropic strains of HIV-1 that are resistant to multiple antiretroviral agents.	Yes
Cystic fibrosis transmembrane conductance regulator (CFTR) gene, G551D mutation	Ivacaftor	Ivacaftor is indicated for treatment of patients with cystic fibrosis who harbor a G551D mutation in at least one of their CFTR genes.	Yes
Cytochrome P450 (CYP) 2C19 variants	Clopidogrel	Patients with genetically reduced CYP2C19 function have lower systemic exposure to the active metabolite of clopidogrel, diminished antiplatelet response, and higher cardiovascular event rates following myocardial infarction than do patients with normal CYP2C19 function.	Yes
	Voriconazole	CYP2C19 is significantly involved in the metabolism of voriconazole; poor metabolizers have on average four-fold higher voriconazole exposure than their homozygous counterparts.	No
	Citalopram	Maximum recommended daily dose of 20 mg (rather than 40 mg) for	No

CYP2C9 variants		CYP2C19 poor metabolizers.	
	Warfarin For additional information on warfarin see below	Increased bleeding risk for patients carrying either the CYP2C9*2 or CYP2C9*3 allele.	Yes
	Celecoxib	CYP2C9 poor metabolizers may have abnormally high plasma levels due to reduced metabolic clearance.	No
CYP2D6 variants	Aripiprazole	CYP2D6 poor metabolizers should receive a reduced dose initially followed by titration according to clinical response.	No
	Atomoxetine	Patients with reduced activity in this pathway (poor metabolizers) have higher plasma concentrations of atomoxetine compared to people with normal activity (extensive metabolizers).	No
	Fluoxetine	CYP2D polymorphisms of the poor metabolizer phenotype do not affect the metabolism of fluoxetine because of compensatory mechanisms. However, fluoxetine inhibits the activity of the 2D6 isoenzyme and it may affect the metabolism of other drugs that are metabolized with this enzyme, potentially making normal metabolizers resemble poor metabolizers.	No

		Drugs that are predominantly metabolized by CYP2D6 that have a relatively narrow therapeutic index (example tricyclic antidepressants) should be initiated at the low end of the dose range if concomitant fluoxetine is being used.	
	Codeine	Some individuals may be ultra-rapid metabolizers due to a specific CYP2D6* genotype; these individuals convert codeine into its active metabolite, morphine, more rapidly and completely than other people. This rapid conversion results in higher than expected serum morphine levels.	No
	Tetrabenazine	Patients without expression of CYP2D6 drug metabolizing enzyme likely to have increased exposure.	Yes, if daily doses >50 mg
Deletion of chromosome 5q(del(5q))	Lenalidomide	Indicated for the treatment of myelodysplastic syndrome in patients with deletion 5q cytogenetic abnormality with transfusion dependent anemia.	Yes
Dihydropyrimidine dehydrogenase (DPD) deficiency	Fluorouracil, capecitabine	Unexpected, severe toxicity (eg, stomatitis, diarrhea, neutropenia, and neurotoxicity) associated with these drugs has been	No

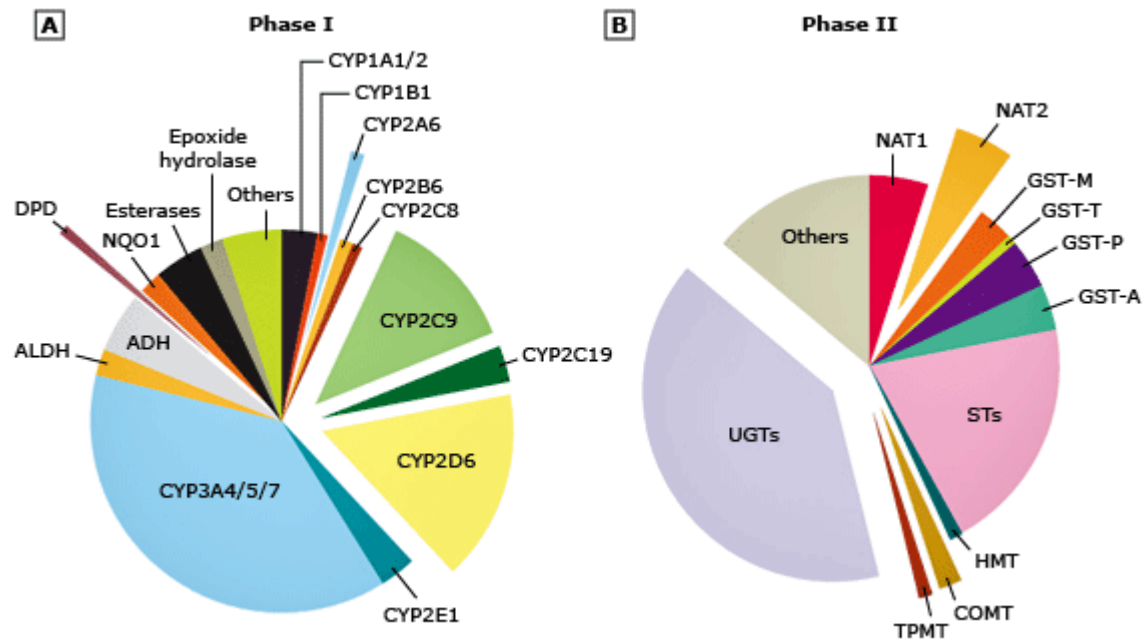
		attributed to a deficiency of DPD activity.	
Familial hypercholesterolemia	Atorvastatin	Dose adjustment is needed in pediatric patients (age 10 to 17) with homozygous familial hypercholesterolemia.	Yes
Glucose-6-phosphate dehydrogenase (G6PD) deficiency	Rasburicase, dapsone, primaquine, chloroquine	These drugs administered to patients with glucose-phosphate dehydrogenase (G6PD) deficiency can cause severe hemolysis.	Yes
HER2 overexpression	Trastuzumab, lapatinib	Over-expression of HER2/neu necessary for selection of patients who are appropriate for use of these agents.	Yes
HLA-B*1502 allele presence	Carbamazepine, oxcarbazepine	Serious dermatologic reactions can result in individuals with the HLA-B*1502 allele.	Yes (in patients of Asian descent)
HLA-B*5701 allele presence	Abacavir	Hypersensitivity reactions, lactic acidosis, and severe hepatomegaly can occur in patients who carry the HLA-B*5701 allele at high risk for such reactions to abacavir; screening for this allele is recommended prior to initiating treatment with abacavir.	Yes
KRAS mutation	Panitumumab, cetuximab	Use of these drugs is not recommended for the treatment of patients with colorectal cancer who have mutated KRAS in their tumors.	Yes

Arylamine N-acetyltransferase (NAT) variants	Rifampin, isoniazid, pyrazinamide, isosorbide dinitrate, hydralazine hydrochloride	Slow acetylation may lead to higher blood levels of these drugs and thus, an increase in toxic reactions.	No
Philadelphia chromosome positivity	Busulfan	This drug is clearly less effective in patients with chronic myelogenous leukemia (CML) who lack the Philadelphia (Ph1) chromosome.	No
	Dasatinib	Indicated for the treatment of adults with Philadelphia chromosome-positive acute lymphoblastic leukemia (ALL) with resistance or intolerance to prior therapy.	Yes
PML/RAR alpha gene expression (retinoic acid receptor) expression	Tretinoin, arsenic trioxide	Indicated for treatment of acute promyelocytic leukemia (APL) characterized by a t(15,17) translocation and/or the presence of PML/RAR alpha gene. The response rate of other APML subtypes has not been demonstrated and therefore, patients who lack this genetic marker should be considered for alternative treatment.	Yes, if no t(15,17)
Protein C deficiency (hereditary or acquired)	Warfarin For additional information on warfarin see above and below	Hereditary or acquired deficiencies of protein C or its cofactor protein S have been associated with tissue necrosis following drug administration.	No
Thiopurine methyltransferase	Azathioprine, 6-mercaptopurine,	Individuals with TPMT deficiency or lower	No Azathioprine

(TPMT) variants	thioguanine	activity due to mutation are at increased risk for myelotoxicity.	label recommends consideration of TPMT testing
Uridine diphospho-glucuronosyltransferase 1A1 (UGT 1A1) variants	Irinotecan	Patients exposed to this drug who have the UGT1A1*28 allele may be more susceptible to toxicity.	Yes
	Nilotinib	Can increase bilirubin levels; the (TA)7/(TA)7 genotype (UGT1A1*28 allele) is associated with a statistically significant increase in the risk of hyperbilirubinemia.	No
Urea cycle disorder (UCD) deficiency	Valproic acid	Use is contraindicated in individuals with urea cycle disorder. Evaluation for UCD prior to therapy should be considered for patients in certain high risk groups including those with family history of UCD.	Yes for patients with history of unexplained encephalopathy or coma, unexplained mental retardation, cyclical vomiting and lethargy, history of elevated ammonia or glutamine, family history of UCD, or unexplained infant deaths.
Vitamin K epoxide reductase (VKORC1) variants	Warfarin For additional information on warfarin see above	Certain single nucleotide polymorphisms in the VKORC1 gene (especially the -1639G>A allele) have been associated with lower dose requirements for warfarin.	Yes

Graphic 65875 Version 13.0

Drug-metabolizing enzymes exhibiting clinically relevant genetic polymorphisms



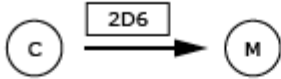


Most drug-metabolizing enzymes exhibit clinically relevant genetic polymorphisms. Essentially all of the major human enzymes responsible for modification of functional groups [classified as phase I reactions (Panel A)] or conjugation with endogenous substituents [classified as phase II reactions (Panel B)] exhibit common polymorphisms at the genomic level; those enzyme polymorphisms that have already been associated with changes in drug effects are separated from the corresponding pie charts. The percentage of phase I and phase II metabolism of drugs that each enzyme contributes is estimated by the relative size of each section of the corresponding chart.

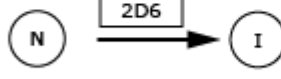
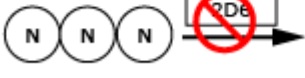

ADH: alcohol dehydrogenase; ALDH: aldehyde dehydrogenase; CYP: cytochrome P450; DPD: dihydropyrimidine dehydrogenase; NQO1: NADPH:quinone oxidoreductase or DT diaphorase; COMT: catechol O-methyltransferase; GST: glutathione S-transferase; HMT: histamine methyltransferase; NAT: N-acetyltransferase; STs: sulfotransferases; TPMT: thiopurine methyltransferase; UGTs: uridine 5'-triphosphate glucuronosyltransferases.

From: Evans WE, Relling MV. Pharmacogenetics: Translating functional geneomics into rational therapeutics. Science 1999; 286:487. Reprinted with permission from AAAS. Copyright © 1999. Readers may view, browse, and/or download material for temporary copying purposes only, provided these uses are for noncommercial personal purposes. Except as provided by law, this material may not be further reproduced, distributed, transmitted, modified, adapted, performed, displayed, published, or sold in whole or in part, without prior written permission from the publisher.

Graphic 77320 Version 2.0

Effect of variation in CYP2D6 metabolism with clinical response to two therapeutic agents

Codeine		Pain relief	Adverse effects
EM		Yes	No
PM		No	Yes
UM		Yes	Yes

Nortriptyline		Relief of depression	Adverse effects
EM		Yes	No
PM		Yes	Yes
UM		No	No

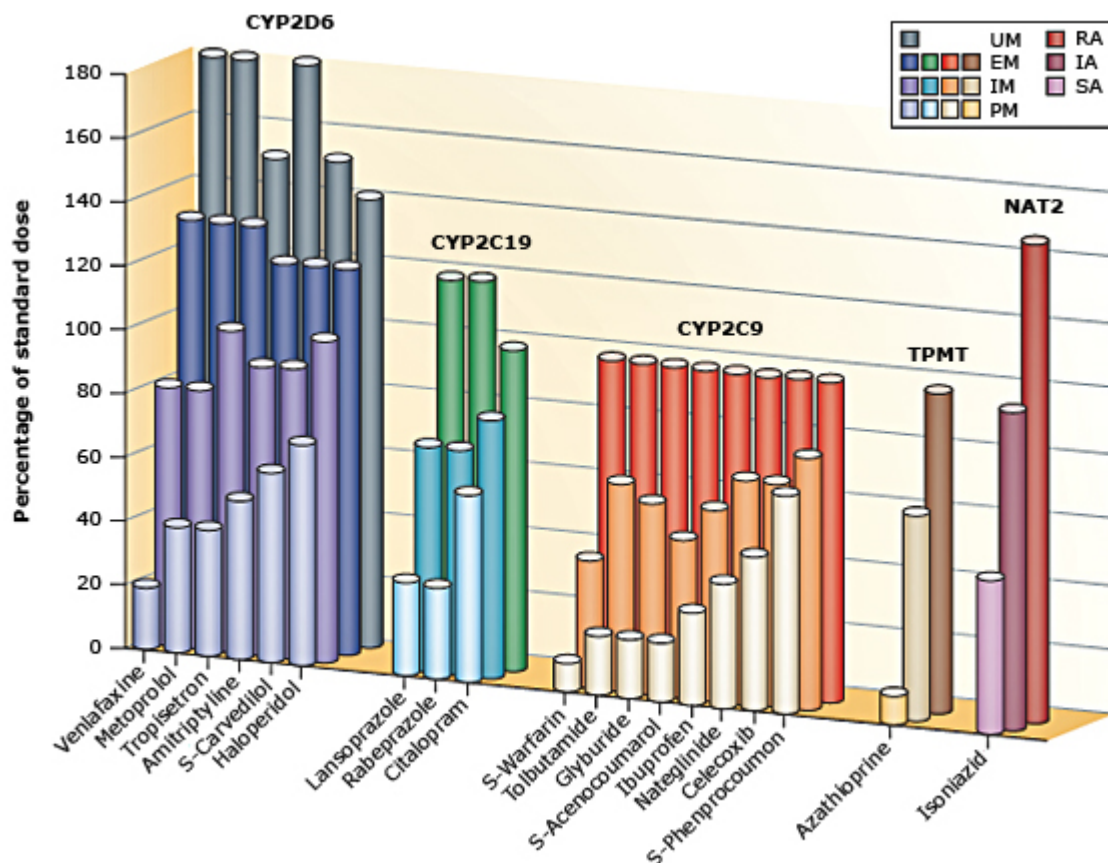
Based upon genetic make-up and the resultant ability to metabolize therapeutic agents, individuals can be classified as extensive (normal) metabolizers (EM), poor metabolizers (PM), or ultrarapid metabolizers (UM). Since codeine is metabolized into an active agent (morphine), poor metabolizers may require increased dosing for a given therapeutic effect, while ultrarapid metabolizers may build up excessive levels of morphine, leading to adverse sequelae. Conversely, nortriptyline is the active therapeutic agent and is metabolized to an inactive form. In this case poor metabolism leads to adequate therapy but an increased incidence of side effects, while extensive metabolizers may require increased dosing for therapeutic effect.

C: codeine; M: morphine; N: nortriptyline; I: inactive nortriptyline; EM: extensive metabolizers; PM: poor metabolizers; UM: ultrarapid metabolizers.

Reproduced with permission from PharmGKB. Copyright © 2009.

Graphic 55020 Version 1.0

Examples of dose adjustments based on PGDx



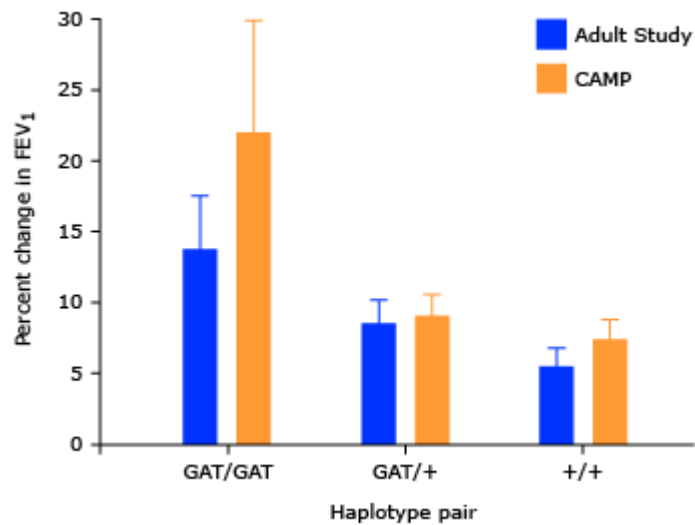
The influence of genetic polymorphisms in cytochrome P450 enzymes CYP2D6, CYP2C19 and CYP2C9, thiopurine S-methyltransferase (TPMT) and N-acetyltransferase type 2 (NAT2) is expressed as subpopulation-specific dosages, according to the difference in pharmacokinetic parameters from clinical studies. The dose adjustments illustrated by the bars in this graph are based on differences in dose-related pharmacokinetic parameters (clearance, AUC, STEADY STATE CONCENTRATION) caused by particular genotypes and are calculated using the methods described earlier. Substantial adjustments need to be made to drug dose to achieve the same level of drug exposure in individuals with different genotypes.

AUC: area under the curve; EM: extensive metabolizer; IA: intermediate acetylator; IM: intermediate metabolizer; PM: poor metabolizer; RA: rapid acetylator; SA: slow acetylator; UM: ultra-rapid metabolizer.

Reprinted by permission from Macmillan Publishers LTD: Kirchheiner J, Fuhr U, Borckmoller J. Pharmacogenetics-based therapeutic recommendations--ready for clinical practice? *Nat Rev Drug Discov* 2005; 4:639. Copyright © 2005.

Graphic 78144 Version 2.0

Eight week response to inhaled corticosteroids, stratified by CRHR1 GAT haplotype status in the Adult Study and CAMP



Utilizing the htSNPs rs1876828, rs242939 and rs242941, the mean FEV₁ improvement in those adults imputed with the GAT/GAT homozygous haplotype was 13.7 percent, while in those homozygous for two non-GAT haplotypes it was 5.5 percent. In CAMP, those imputed for the GAT/GAT haplotype demonstrated a 21.8 percent improvement in FEV₁ versus 7.4 percent for those with no GAT haplotype. Improvement in those heterozygous for the GAT haplotype was intermediate between the two groups, suggesting an additive effect. Mean values \pm SEM are shown.

Reproduced with permission from: Tantisira KG, Lake S, Silverman ES, et al. Corticosteroid Pharmacogenetics: Association of sequence variants in CRHR1 with improved lung function in asthmatics treated with inhaled corticosteroids. Hum Mol Genet 2004; 13:1353. Copyright © 2004 Oxford University Press.

Graphic 60913 Version 11.0

Disclosures

Disclosures: **Kelan Tantisira, MD, MPH** Nothing to disclose. **Scott T Weiss, MD, MS** Nothing to disclose. **Benjamin A Raby, MD, MPH** Other Financial Interest (Spouse): Parexel International [Contract research organization (Double-blind randomized clinical trials)]. **Diane MF Savarese, MD** Nothing to disclose. **Jennifer S Tirnauer, MD** Nothing to disclose.

Contributor disclosures are reviewed for conflicts of interest by the editorial group. When found, these are addressed by vetting through a multi-level review process, and through requirements for references to be provided to support the content.

Appropriately referenced content is required of all authors and must conform to UpToDate standards of evidence.

Conflict of interest policy

Personalized medicine

Author

Benjamin A Raby, MD, MPH

Section Editor

Anne Slavotinek, MBBS, PhD

Deputy Editor

Jennifer S Tirnauer, MD

All topics are updated as new evidence becomes available and our [peer review process](#) is complete.

Literature review current through: Feb 2015. | **This topic last updated:** Mar 09, 2015.

INTRODUCTION — With the rapid pace of progress in the fields of biotechnology, genetics, and genomics, molecular genetic profiling may soon become an integral tool for clinicians to guide individualized management of many medical conditions. Personalized medicine (also termed personalized genomics, genomic medicine, or precision medicine) refers to the application of patient-specific profiles, incorporating genetic and genomic data as well as clinical and environmental factors, to assess individual risks and tailor prevention and disease-management strategies.

This topic will review the current state of personalized medicine and the challenges and obstacles confronted in the broad implementation of genomics in the clinic.

The principles and techniques that underlie these tests, and counseling for genetic testing, are discussed separately.

- (See ["Tools for genetics and genomics: Gene expression profiling"](#).)
- (See ["Principles and clinical applications of next-generation DNA sequencing"](#).)
- (See ["Tools for genetics and genomics: Cytogenetics and molecular genetics"](#).)
- (See ["Genetic association studies: Principles and applications"](#).)
- (See ["Overview of pharmacogenomics"](#).)
- (See ["Genetic counseling and testing"](#).)

For discussions of specific tests, readers are referred to disease-specific topics.

DEFINITION — Personalized medicine involves the use of an individual's genetic profile to guide decisions made in regards to the prevention, diagnosis, and treatment of disease [1]. The definition encompasses a broad range of current clinical practices in which genomics results are used to guide patient care. Examples include dose modification of medicines depending on single nucleotide polymorphisms (SNPs) that determine drug metabolism speed and individualized screening for effective therapies in cancer [2,3].

BENEFITS AND LIMITATIONS — Advocates for personalized medicine cite its potential to yield significant health and economic benefits for patients, practitioners, and society, including:

- Improved medical decision making
- Delivery of appropriate therapies that are tailored to a patient's sequence variants rather than the general population
- Optimized disease prevention strategies, including lifestyle and behavioral modification as well as pharmacoprevention
- Reduced exposure to medications of lower efficacy
- Reduced exposure to medications that have the potential for greater toxicity, with resulting lower incidence of treatment-related complications
- Reduced healthcare costs
- Enhanced patient satisfaction with the treatment process, improved tolerance of therapy, and better

compliance

Several studies have demonstrated that the availability of genetic information enhances patient compliance with behavior modification and other disease prevention strategies [4-8].

- In a study of 199 smokers, individuals who were homozygous for a severely deficient, alpha-1-antitrypsin null allele were significantly more likely to attempt to quit smoking (59 percent) than either mutation carriers (heterozygotes; 34 percent) or individuals with a normal genotype (26 percent) [4].
- In a study of 261 smokers who were asked to consider one of two randomly assigned hypothetical scenarios, subjects who were informed of having genetic risk factors for heart disease were more likely to report an inclination to quit smoking than individuals in a high-risk (but non-genetic) group [7]. A substantial proportion of respondents reported that their decisions were motivated by the belief that smoking cessation would translate to reductions in heart disease, suggesting that genetic determinism would not negatively influence the ability to modify behavior and improve risk factor avoidance.
- In a study of 781 individuals who were carriers of mutations that cause familial hypercholesterolemia (ascertained through an affected relative), cholesterol-lowering medication use increased from 51 to 81 percent two years after learning genotype status [8]. Significant reductions in low-density lipoprotein C levels were noted, although normalized levels were not achieved in most subjects.

What is not clear is whether genetic information for individuals who have no known increased risk prior to genetic testing will lead to change in lifestyle behaviors. In a sample of over 2000 patients who completed genomewide testing through a direct-to-consumer vendor, testing did not result in any short-term change in diet or exercise behavior [9].

Skeptics also argue that, while there are selected examples where specific biomarkers or genetic tests can help guide medical decision making, more wide-scale profiling remains largely unrealistic in clinical practice. Factors cited to support this viewpoint include the high costs of testing, the lack of reliable predictive biomarkers for most conditions, the lack of clear therapeutic alternatives (based on genetic differences) for many conditions, and the lack of knowledge and expertise among most clinicians regarding genetics, risk prediction, and genetic counseling [10,11]. These issues are discussed in more detail below. (See '[Obstacles for implementation](#)' below.)

Nonetheless, momentum for the implementation of personalized medicine in clinical practice is increasing. The price of exome sequencing and whole genome sequencing is continuing to fall, and the catalog of disease-associated genetic copy number variants or sequence variants, produced from genome-wide association studies, array comparative genomic hybridization (array CGH), and next-generation sequencing (including exome sequencing and whole genome sequencing), is rapidly increasing [6]. Despite obstacles, there is strong support for the development of personal genomics among healthcare policy makers and research funding agencies [7,8].

PERSONALIZED MEDICINE INITIATIVES — Available biomarker assays are impacting the practice of several medical specialties, most notably in oncology.

The use of a personalized medicine approach has also been advanced by some federal initiatives. As examples:

- In 2007, the Department of Health and Human Services (HHS) in the United States launched the Personalized Health Care Initiative (PHCI) [12]. This initiative proposed a set of goals “for achieving gene-based medical care combined with health information technology.” The PHCI aims to accelerate the development of personalized treatment strategies and transform the practice of

medicine towards individualized patient care.

Components of the program include translational research initiatives to develop high-throughput sequencing technologies (ie, next-generation sequencing), enhanced mapping of the genetic determinants of disease and drug responsiveness, development of an informatics infrastructure to promote electronic medical records, and broad implementation of genomic data.

- In 2015, a precision medicine initiative was proposed in the United States, which would promote enhanced use of existing and new databases of genomic information to improve diagnosis and therapy [13]. In the near-term, the initiative will focus on cancer; a longer-term aim is to generate knowledge related to a broader range of diseases.

TYPES OF GENETIC TESTING — In general terms, genetic testing has focused on three types of genetic variation: (i) relatively rare, disease-causing sequence variants (mutations) that are associated with Mendelian diseases (eg, conditions with autosomal dominant, autosomal recessive, and X-linked inheritance patterns); (ii) more common variants, often single nucleotide polymorphisms (SNPs) that are associated with an increased risk of complex diseases (eg, conditions in which the etiology is multifactorial in terms of more than one gene involved or gene-environment interactions); (iii) and sequence variants, often SNPs, that modulate drug response (pharmacogenetic variants). Broadly, three types of genetic tests are available:

- **Specific single gene tests** — Most gene-specific tests are performed as part of a focused risk evaluation for heritable disease or for diagnostic considerations. Examples include *BRCA1* and *BRCA2* gene sequencing for carrier identification in at-risk individuals with a strong family history of breast cancer, or pharmacogenetic testing to guide dosing of chemotherapeutic agents in individuals with sequence variants that affect drug metabolism. (See "[Genetic counseling and testing](#)" and "[Overview of pharmacogenomics](#)".)
- **Specific gene panels** — Panels that provide sequence for multiple genes have been developed for clinical testing to determine the etiology of conditions with high genetic heterogeneity, in which mutations in multiple genes can cause the same phenotype. Examples include panels to test for genetic causes of intellectual disability, epilepsy, autism, and hereditary deafness.
- **Genotyping panels of selected susceptibility variants** — These panels are often bundled as direct-to-consumer (DTC) genetic tests and include SNPs that have been associated with common, complex diseases such as type 2 diabetes, autoimmune disease, and metabolic traits. Due to the inherent genetic complexities of these disorders, the genetic effects conferred by each of the SNPs are often relatively small (odds ratios of <1.5 for risk variants) and inconsistent across different ethnic populations. The predictive accuracy of these tests therefore is highly variable and most of the panels have not been comprehensively validated. (See "[Direct-to-consumer testing](#)" below.)
- **Whole genome or exome sequencing** — High-throughput sequencing platforms (eg, next-generation sequencing) enable sequencing of the coding regions of the entire human exome (ie, all regions of the genome that encode proteins) or genome (ie, regions of the genome that encode proteins as well as regulatory elements). The use of whole genome sequencing remains in the realm of research testing for most conditions, due to the difficulties in interpreting the large amounts of data that result from this test. (See "[Principles and clinical applications of next-generation DNA sequencing](#)".)

Pharmacogenetic testing — The earliest clinical implementations of genetic profiling have been in the area of pharmacogenetics, also referred to as pharmacogenomics. Pharmacogenetics is the study of variability in drug response due to genetic factors, and includes the prediction of a patient's response to a specific therapy and susceptibility to toxicity and adverse events. Pharmacogenetic data may inform both the selection of a particular treatment and the individualized dose and dosing schedule for that

treatment. This subject is described in more detail separately. (See "[Overview of pharmacogenomics](#)".)

Drug labels for numerous drugs include information regarding pharmacogenetic biomarkers that can be tested. Though most notably impacting dosing of medications used to treat hematologic malignancies and solid tumors, pharmacogenetic markers are also available for medications used in the treatment of infectious, cardiac, rheumatologic, and pulmonary diseases [14]. Label content includes warnings regarding genotype-specific contraindications or toxicities, dosing recommendations, or information regarding the availability of genetic tests (without specific recommendations for testing). An updated listing of pharmacogenetic biomarkers cited in drug labels in the United States is available on the [US Food and Drug Administration](#) website.

The most extensively studied pharmacogenetic variants are those of the cytochrome P450 drug metabolizing liver enzymes (CYPs). Fifty-eight CYPs have been characterized in humans, and functional SNPs that alter functional activity have been identified for many CYPs. The Amplichip *CYP450* test is clinically available in the United States as a DNA oligonucleotide genotyping microarray for simultaneous characterization of 29 SNPs in the genes *CYP2D6* and *CYP2C19* [15]. These variants influence the metabolism of a wide range of commonly prescribed medications, including 33 with pharmacogenetic biomarker labels. However, this test has not been widely adopted for clinical use, since the clinical value of this array has not been validated in prospective studies and insurance reimbursement is infrequently available.

Genotyping panels — Genotyping refers to testing for the allelic status of an individual at polymorphic loci that are often associated with disease. In contrast to sequencing, genotyping assays are designed from previous knowledge of the existence of a particular genetic variant, commonly a SNP. Numerous Clinical Laboratory Improvement Amendments (CLIA) certified genotyping technologies have been developed for simultaneous characterization of dozens to hundreds of loci. (See "[Genetic counseling and testing](#)".)

Publication of hundreds of genome-wide association studies (GWAS) have identified thousands of common variants that confer only a small to modest alteration in disease risk, most commonly for complex diseases. These findings have led to the development of genotyping panels for testing for many of these variants. The SNP content and diseases covered vary from panel to panel, although there is overlap among the most commonly studied variants. The predictive accuracy of these panels is highly variable, and the clinical utility of these tests has not been established and remains unclear at present. (See "[Genetic association studies: Principles and applications](#)".)

Exome and genome sequencing — Genome or exome sequencing differs from specific gene tests or genotyping panels by determining the sequence of every exon or gene in an individual's genome using massive, parallel sequencing. The technology used to obtain the sequence is referred to as next-generation sequencing (NGS). Exome sequencing is more common; whole genome sequencing is largely undertaken in a research setting.

Advantages of NGS are the ability to provide information regarding all sequence variants, including alterations in copy number if whole genome sequencing is performed, rather than only interrogating a small fraction of known genes. As more knowledge is gained regarding the spectrum of disease-causing variants and catalogs of such variation are assembled, complete sequence information will provide greater predictive accuracy than genotyping panels. The NGS technology and its clinical indications are discussed in detail separately. (See "[Principles and clinical applications of next-generation DNA sequencing](#)".)

Because NGS interrogates the entire exome or genome, it may also reveal unexpected findings related to previously unknown genetic risks, loosely termed "incidental" or "secondary" findings. Some of these variants may have a clear association with disease, whereas others may be of uncertain significance.

Decisions regarding which findings to disclose to the patient are an area of active research. Consent for exome sequencing typically provides patients and families with the choice as to whether they would like to receive incidental findings (eg, “opt-out” option). (See ["Incidental findings from genetic testing"](#) and ["Genetic counseling and testing"](#).)

Fetal personalized medicine — With the recognition that a sufficient quantity of fetal DNA is present in the maternal circulation for clinical study, it is becoming increasingly easier and safer to accurately assess the structural integrity and sequence variation of fetal genomes. Consequently, prenatal genetic testing is playing an increasing role in obstetrical care, with a move towards implementation of personalized approaches in fetal medicine [16]. (See ["Noninvasive prenatal testing using cell-free nucleic acids in maternal blood"](#).)

An example is the implementation of a non-invasive prenatal test (NIPT) of cell-free fetal DNA (cf-DNA), also called free fetal DNA (ff-DNA) sequences from maternal blood samples for the Rhesus D blood group (RhD) allele. In about 10 percent of non-Hispanic white pregnancies, seronegative RhD mothers carry an RhD-positive fetus. These mothers are at risk of RhD sensitization during delivery, with consequential risk to subsequent RhD-positive pregnancies for life-threatening hemolytic crises. Maternal sensitization is preventable with timely administration of anti-D antibodies, and anti-D prophylaxis of RhD-negative mothers is standard in prenatal care. However, about 40 percent of RhD-negative mothers (those carrying RhD-negative fetuses) receive anti-RhD (a blood product) unnecessarily. (See ["Management of pregnancy complicated by Rhesus \(Rh\) alloimmunization"](#).)

Non-invasive genotyping assays have been developed that type the RhD allele in ff-DNA circulating in maternal blood. The assay has a diagnostic accuracy approaching 100 percent [17], with better accuracies noted with newer assays that include rigorous positive controls that confirm sufficient quantities of circulating fetal DNA [18]. However, the test performs less well in non-white populations, due to the presence of alternative sequence variants. Non-invasive screening of RhD-negative pregnancies using this method could spare about 40 percent of mothers from unnecessary exposure to anti-RhD therapy and avoid more invasive testing (ie, amniocentesis) in previously sensitized mothers during subsequent pregnancies. Many European countries of predominantly white ancestry have implemented routine ff-DNA screening for RhD-sensitized mothers. Fewer support the more generalized screening in all RhD-negative mothers, largely due to the high costs of mass screening. For example, a cost-effectiveness analysis in the United Kingdom concluded that at current test and immunization costs, routine ff-DNA screening of all RhD-negative mothers would not result in cost-savings or appreciable health benefits [19]. However, strategies to lower the costs of testing, including bundling of this test with other testing (eg, other blood groups, structural genetic aberrations), may result in broader implementation.

Other applications of cf-DNA in testing (eg, for trisomy 21) are presented separately. (See ["Noninvasive prenatal testing using cell-free nucleic acids in maternal blood", section on 'Prenatal applications'](#).)

DIRECT-TO-CONSUMER TESTING — Genetic testing for disease susceptibility variants was previously accessible to the general public through direct-to-consumer (DTC) testing [20]. However, the DTC genetic testing industry is rapidly evolving, and various legislative and regulatory bodies are actively developing policies regarding these services.

In late 2013, the US Food and Drug Administration (FDA) requested that the company 23andMe discontinue marketing of their health-related personal genome service (PGS; ie, disease-risk prediction) in the United States, because the clinical validity of the service had not been demonstrated by the company [21]. The FDA considered the PGS kit to be a device and hence under their regulatory purview. In their notification to the company, the FDA expressed concerns regarding the potential misuse of reported genetic information leading to inappropriate management. As an example, patients might make treatment decisions regarding prophylactic mastectomy, chemoprevention, or aggressive surveillance

based on false-positive or false-negative *BRCA1* genotypes. An editorialist suggested that the FDA may not have intervened if the genome service had been available only to an ordering physician, rather than marketed directly to consumers [22].

23andMe suspended its clinical services in the United States, with the exception that raw genetic data and ancestry testing continue to be provided. The company began marketing personal genome testing directly to the general public in the United Kingdom in late 2014 [23].

A list of available DTC companies and genetic tests is available from the [Genetics and Public Policy Center](#).

Given the heterogeneity among DTC vendors in the markers tested and the predictive models used to develop risk estimates, questions have been raised regarding the reliability and reproducibility of these services. The accuracy of identifying mutations or disease-associated SNPs is one concern; the accuracy of using the resulting information to predict risk of disease is a separate concern of equal or possibly greater importance.

Two studies have evaluated these issues; both have found high concordance between companies in determining genotype, but significant variability in interpretation of disease risk based on genotype:

- One study compared genotyping panels from two DTC companies: 23andMe and Navigenics [24]. DNA from five individuals was sent to each company for testing, and risk predictions for 13 diseases were compared. Genotype calls (ie, the genotypes reported at each locus based on SNP analysis) were in excellent agreement (>99 percent concordance). In contrast, risk prediction for diseases was discordant in about a third of cases. Disease risk predictions from the two companies **disagreed** more than half of the time for seven conditions (eg, systemic lupus erythematosus, heart disease, Crohn disease, type 2 diabetes); risk predictions were concordant in four conditions. Conditions that demonstrated strong agreement were those where an identical SNP with a very high genetic effect was tested by both companies (for example, the SNP for celiac disease that confers a sevenfold increase in risk). SNPs with such strong effects are uncommonly observed in common complex diseases.
- A second study evaluated the concordance across three companies (23andMe, Navigenics, and deCODE), although only one sample was tested [25]. Genotype concordance was high (99.6 percent; ie, analytical validity of the available tests was high [26]), but significant variability in risk prediction was noted with undetermined clinical validity and utility. The variability in risk prediction was influenced by the SNPs that were genotyped for each condition and the reference population used [25]. Diseases for which genotyped SNPs have strong predictive value were more likely to receive similar risk estimates from different DTC companies [26]. This study also pointed out the need for risk data that is based upon client ethnicity [25], as the majority of genome-wide association studies have focused on Caucasians, with uncertain applications to other ethnicities, including Asian, African, and Hispanic [26].

These reports suggest that risk prediction remains unreliable, and interpretation of results for most SNPs should be approached with caution. In addition, patients should be made aware that:

- For all but a limited number of SNPs of large risk, the majority of SNPs tested provide only incremental changes in a patient's risk profile.
- Prospective studies of the predictive accuracy of these products have not been performed, thus precluding the provision of effective counseling or reliable decision-making for most results.
- There is substantial inter-company variability in the estimates reported [25,26].
- The influence of non-genetic factors, including race and lifestyle factors, on the interpretation of

these results is unknown for many of the SNPs tested.

OTHER PERSONALIZED MEDICINE PLATFORMS — The early and intuitive focus for personalized medicine has been the development of genetic-based tests. Other “omic” approaches are being developed that will provide a more complete characterization of risk that includes variation between individuals in gene regulation, epigenetics, and cellular metabolism [27]. Such approaches, which are under development as part of research and rarely provided clinically, include:

- Gene expression profiling (also referred to as transcriptomics) — Analysis of mRNA (of either individual genes or panels of gene targets), representing gene expression patterns; often uses microarray technology although whole transcriptome analysis (also known as RNA-Seq) is also possible. Gene expression is dynamic and influenced by a range of cellular, genetic and environmental factors, which makes gene expression a particularly attractive target for profiling malignant cells. (See ["Tools for genetics and genomics: Gene expression profiling"](#).)
- Proteomics — Qualitative and quantitative analysis of the collection of protein constituents in a biological sample. Typically performed using modification of polyacrylamide gel electrophoresis (PAGE) or matrix-assisted laser desorption/ionization (MALDI) approaches, these methods provide measures of the types and abundance of proteins in a biological sample. (See ["Overview of gene expression profiling, proteomics, and microRNA profiling in clinical oncology"](#), section on 'Proteomics'.)
- Metabolomics — The characterization of metabolic profiles; typically consists of a collection of assays that characterize panels of metabolites related to specific pathways. These studies can be static (cross-sectional profiling at a given time-point) or dynamic (assessing the change in profile patterns following a specific metabolic challenge) [28]. In combination with separation methods such as high-performance liquid chromatography or gas chromatography, metabolites are typically characterized either by mass spectrometry (MS) or nuclear magnetic resonance spectroscopy (NMR).
- Lipidomics — Characterization of the complete collection of lipids. Lipid structures, like metabolites, can be differentiated by MS or NMR [29,30]. These methods are being applied towards the development of diagnostic tests that assess the lipid composition of cell membranes [31].
- Epigenomics — Profile of the modifications to DNA (often, methylation) that control gene expression. Unlike genomic changes, epigenomic changes are affected by the environment and may change with age, stress, or exposures to the individual or earlier generations.
- Exposomics — The sum of exposures an individual incurs over a period of time. These may include nutrients, foods, toxins, stresses, exercise, vaccinations, medications, and other exposures. The exposome is highly dynamic and malleable over an individual's life.
- Microbiomics — Characterization of the microbes (typically, bacteria) that reside in or on an individual. A common example is the gut microbiome, which might influence adiposity and/or immunity.

Few of these applications are being used routinely in clinical practice, although development of reliable, clinically adaptable assays for these platforms is actively being pursued. One important exception is in the field of oncology, where gene expression profiling of malignant cells or tissue is becoming an important diagnostic and prognostic tool. Some of the earliest and most successful implementations of gene expression profiling in oncology focused on single genes, such as determining estrogen receptor expression status in breast cancer for informing prognosis and chemotherapeutic options [32]. Subsequently, whole transcriptome expression profiling technologies have been applied to tumor samples, with notable success [33,34]. (See ["Overview of gene expression profiling, proteomics, and](#)

[microRNA profiling in clinical oncology".](#)).

OBSTACLES FOR IMPLEMENTATION — Despite early successes in the clinical introduction of a limited number of pharmacogenetic assays, multiple barriers preclude the widespread implementation of personalized medicine as standard clinical practice across all medical fields. The development of validated biomarkers and genetic assays represents an important bottleneck as does the analysis of the masses of data that can result from next-generation sequencing technologies, although it is likely that the prospect of affordable exome and whole genome sequencing will remove this obstacle in the future. Other ethical and data storage challenges will remain.

- **Limited predictive value of most tests** — There are few examples where genetic testing provides substantial gains in guiding therapeutic recommendations, although diagnostic genetic testing may be very useful for patient management and recurrence risk estimation. However, the number of genetic variants associated with disease susceptibility and pharmacogenetic response is increasing.
- **Lack of physician knowledge** — The lack of general knowledge of genetics among most medical practitioners is one of the most pressing challenge preventing broad implementation of personalized medicine. Many medical practitioners do not feel adequately prepared to provide counseling for genetic testing [35] and the number of genetic counselors and clinical geneticists in North America has been hypothesized to be lower than could meet expected demands. However, medical schools are actively developing curricula content in the areas of personalized medicine to prepare the next generation of physicians, including several programs that have developed case-studies of whole genome sequence data [36,37]. However, there is little infrastructure in place to facilitate the education of clinicians already in practice.

To bridge the knowledge gaps, reporting is structured with clinical annotation to provide detailed interpretations, including detailed summaries of the significance of sequence variants and estimated disease risks, advisories based on the types of identified pharmacogenetic variants. Laboratory reports may also include novel variations of unclear clinical significance. (See ["Incidental findings from genetic testing".](#))

- **Inadequate informatics infrastructure** — The archiving of clinical laboratory data resides with the health care providers (and their affiliated institutions) who originally requested the tests. This model is inadequate to accommodate next-generation sequence data, and newer approaches that include an easily accessible electronic medical record (EMR) are under development. Although adoption of EMRs has increased, the majority of those in place would require substantial enhancement to accommodate the data storage and processing needs anticipated for genomic data. The Personalized Health Care Initiative has made the nationwide development of EMRs a top priority [12], and the Office of National Coordination for Health Information Technology (ONC) was established in 2004 by executive order to facilitate the establishment of an electronic health record (EHR) across the United States that would coordinate information across multiple EMRs.
- **Information providence and patient privacy issues** — Many questions arise when considering how personal genomic information should be managed.
 - Who would be responsible for ordering a pharmacogenetic genotyping panel (ie, the major CYP loci) that profiles responsiveness for a wide array of drugs?
 - How would such information be transmitted to other clinicians and pharmacists?
 - When should genomic profiling be ordered? If personal genomic profiling is performed as part of a neonatal screening program, who will be the purveyors of such data?

- Would pediatricians be responsible for interpreting data and facilitating the development of patient management plans related to conditions of adult onset?

Many geneticists do not recommend testing for adult onset disorders in childhood unless there are diagnostic or management implications relevant to the child at the age of testing. (See "[Genetic counseling and testing](#)", section on 'Ethical, legal, and psychosocial issues' and "[Genetic counseling and testing](#)", section on 'Whom to test' and "[Genetic counseling and testing](#)", section on 'Informed consent for testing'.)

- At what age would children be informed of their specific liabilities?
- Should information be provided for conditions for which there is no known cure or treatment?
- **Inconsistent standardization and oversight of testing** — As illustrated by the preliminary concordance studies of direct-to-consumer (DTC) genetic tests (see '[Direct-to-consumer testing](#)' above), there is considerable inconsistency in the predicted genetic risk [\[24,25\]](#). These inconsistencies are largely due to differences in the risk-estimation models used and in the single nucleotide polymorphism (SNP) content on the array platforms. Standardization of these models and the SNP content will help improve this problem. The US Food and Drug Administration (FDA) is developing policies regarding standardization and oversight of these tests, and policies have been discussed in Australia and the United Kingdom; some have advocated for international certification of quality standards [\[38\]](#).
- **Reimbursement issues** — Changes in the reimbursement policies will be needed to promote a personalized medicine initiative. Several attempts have been made to modify the Medicare clinical laboratory fee schedule to accommodate the reimbursement for complex genetic tests.
- **Societal issues and misconceptions** — Numerous societal challenges must be addressed prior to widespread implementation of genomic medicine [\[39,40\]](#).
 - Acceptance of genetically-based treatment recommendations may be low, particularly among minority groups where the issues of genetics and race can be intimately interrelated [\[33\]](#).
 - Concern that misconceptions of genetic determinism could result in a medicalization of society, whereby healthy individuals become preoccupied with disease prevention based on their profile or undergo unnecessary tests or procedures. One study of individuals who underwent direct-to-consumer genome profiling found no measurable evidence for increased anxiety, comparing baseline and post-testing anxiety scores [\[9\]](#).
 - Genome sequencing will likely identify a variety of “high-risk” alleles in all individuals. With the limited available understanding of risk concepts or penetrance of mutations and risk alleles, the identification of risk alleles or recessive traits could inappropriately influence reproductive decisions.
 - Individuals may be adversely impacted when advised of increased risk for sensitive issues such as psychiatric disorders or behavioral traits.
 - The desirability of learning if an individual has an increased risk for conditions for which there is no available treatment or prevention is questionable. In one study, more than 80 percent of at-risk individuals who had no symptoms chose not to pursue testing for Huntington disease [\[41\]](#).
 - Family dynamics might be affected by disclosure, or failure to disclose, genetic information.

SUMMARY

- Personalized medicine involves the use of an individual's genetic profile to guide decisions made in regards to the prevention, diagnosis, and treatment of disease. (See ['Definition'](#) above.)
- Potential benefits of personalized medicine include customized treatment plans that could encompass targeted pharmacotherapy to improve drug response and reduce toxicity and expense, targeted recommendations for lifestyle modifications and other disease-prevention strategies, and enhanced patient satisfaction with healthcare. Genetic information can enhance patient compliance with behavior-modification recommendations. (See ['Personalized medicine initiatives'](#) above.)
- The earliest implementation of genetic profiling into clinical practice has been in the area of pharmacogenetic testing. Pharmacogenetic biomarkers are available for over 70 drugs; many of these markers involve polymorphisms in cytochrome drug metabolizing activity. (See ['Pharmacogenetic testing'](#) above and ["Overview of pharmacogenomics"](#).)
- Genotyping refers to testing for the allelic status of an individual at polymorphic loci that are often associated with disease. In contrast to sequencing, genotyping assays are designed from previous knowledge of the existence of a particular genetic variant, commonly a single nucleotide polymorphism (SNP). (See ['Genotyping panels'](#) above.)
- Next-generation sequencing (NGS) refers to a type of sequencing technology in which the sequences of multiple genome fragments are determined in parallel, allowing an exponential increase in the amount of sequence data generated. Exome or whole genome sequencing can provide information regarding a large number of genetic variants (sequence variants) in a single test. This technology and its clinical indications are discussed in detail separately. (See ["Principles and clinical applications of next-generation DNA sequencing"](#) and ["Incidental findings from genetic testing"](#).)
- Genetic testing for disease susceptibility variants was previously accessible to the general public through direct-to-consumer (DTC) testing; however, in late 2013, the US Food and Drug Administration (FDA) requested that the company 23andMe discontinue marketing of their health-related personal genome service in the United States because the clinical validity of the service had not been demonstrated by the company. For DTC genetic analyses that were largely based on genotyping platforms, biomarker identification appears accurate, but there is wide and discrepant variation in the interpretation of the clinical relevance of the identified markers. (See ['Direct-to-consumer testing'](#) above.)
- Analyses based upon gene expression profiling, proteomics, metabolomics, or lipidomics may enhance the predictive value of testing used for personalized medicine. Gene expression profiling is used for some tumors, reflecting cellular, genetic, and environmental factors. (See ['Other personalized medicine platforms'](#) above.)
- Multiple issues will need to be addressed to successfully implement personalized medicine. Beyond the development of validated biomarkers, areas of concern include physicians' knowledge base of genetics and risk interpretation; infrastructure to store and confidentially retrieve an individual's genetic data; timing of testing; issues of providence in terms of clinicians responsible for ordering and transmitting information; test standardization and quality; reimbursement; and overcoming public misperceptions related to genetic data. (See ['Obstacles for implementation'](#) above.)

Use of UpToDate is subject to the [Subscription and License Agreement](#).

REFERENCES

1. Genetics Home Reference <http://ghr.nlm.nih.gov/glossary=personalizedmedicine> (Accessed on February 17, 2015).
2. Rettie AE, Tai G. The pharmacogenomics of warfarin: closing in on personalized medicine. *Mol Interv* 2006; 6:223.
3. Crystal AS, Shaw AT, Sequist LV, et al. Patient-derived models of acquired resistance can identify effective drug combinations for cancer. *Science* 2014; 346:1480.
4. Carpenter MJ, Strange C, Jones Y, et al. Does genetic testing result in behavioral health change? Changes in smoking behavior following testing for alpha-1 antitrypsin deficiency. *Ann Behav Med* 2007; 33:22.
5. Collins RE, Wright AJ, Marteau TM. Impact of communicating personalized genetic risk information on perceived control over the risk: a systematic review. *Genet Med* 2011; 13:273.
6. Wright AJ, French DP, Weinman J, Marteau TM. Can genetic risk information enhance motivation for smoking cessation? An analogue study. *Health Psychol* 2006; 25:740.
7. Sanderson SC, Michie S. Genetic testing for heart disease susceptibility: potential impact on motivation to quit smoking. *Clin Genet* 2007; 71:501.
8. Huijgen R, Kindt I, Verhoeven SB, et al. Two years after molecular diagnosis of familial hypercholesterolemia: majority on cholesterol-lowering treatment but a minority reaches treatment goal. *PLoS One* 2010; 5:e9220.
9. Bloss CS, Schork NJ, Topol EJ. Effect of direct-to-consumer genomewide profiling to assess disease risk. *N Engl J Med* 2011; 364:524.
10. Li C. Personalized medicine - the promised land: are we there yet? *Clin Genet* 2011; 79:403.
11. Vogenberg FR, Barash CI, Pursel M. Personalized medicine: part 3: challenges facing health care plans in implementing coverage policies for pharmacogenomic and genetic testing. *P T* 2010; 35:670.
12. US Department of Health and Human Services. www.hhs.gov. (Accessed on May 03, 2011).
13. Collins FS, Varmus H. A new initiative on precision medicine. *N Engl J Med* 2015; 372:793.
14. US Food and Drug Administration. <http://www.fda.gov> (Accessed on May 03, 2011).
15. Jain KK. Applications of AmpliChip CYP450. *Mol Diagn* 2005; 9:119.
16. Bianchi DW. From prenatal genomic diagnosis to fetal personalized medicine: progress and challenges. *Nat Med* 2012; 18:1041.
17. Minon JM, Gerard C, Senterre JM, et al. Routine fetal RHD genotyping with maternal plasma: a four-year experience in Belgium. *Transfusion* 2008; 48:373.
18. Scheffer PG, van der Schoot CE, Page-Christiaens GC, de Haas M. Noninvasive fetal blood group genotyping of rhesus D, c, E and of K in alloimmunised pregnant women: evaluation of a 7-year clinical experience. *BJOG* 2011; 118:1340.
19. Szczepura A, Osipenko L, Freeman K. A new fetal RHD genotyping test: costs and benefits of mass testing to target antenatal anti-D prophylaxis in England and Wales. *BMC Pregnancy Childbirth* 2011; 11:5.
20. Wright CF, Hall A, Zimmern RL. Regulating direct-to-consumer genetic tests: what is all the fuss about? *Genet Med* 2011; 13:295.
21. <http://www.fda.gov/ICECI/EnforcementActions/WarningLetters/2013/ucm376296.htm> (Accessed on December 02, 2013).
22. Zettler PJ, Sherkow JS, Greely HT. 23andMe, the Food and Drug Administration, and the future of genetic testing. *JAMA Intern Med* 2014; 174:493.
23. Wise J. Personal genome service launches in UK. *BMJ* 2014; 349:g7435.

24. Ng PC, Murray SS, Levy S, Venter JC. An agenda for personalized medicine. *Nature* 2009; 461:724.
25. Imai K, Kricka LJ, Fortina P. Concordance study of 3 direct-to-consumer genetic-testing services. *Clin Chem* 2011; 57:518.
26. Zehnbauser B. Direct-to-consumer genetics testing--fair comparisons? *Clin Chem* 2011; 57:369.
27. Coughlin SS. Toward a road map for global -omics: a primer on -omic technologies. *Am J Epidemiol* 2014; 180:1188.
28. Holmes E, Wilson ID, Nicholson JK. Metabolic phenotyping in health and disease. *Cell* 2008; 134:714.
29. Gross RW, Han X. Lipidomics at the interface of structure and function in systems biology. *Chem Biol* 2011; 18:284.
30. Puri R, Duong M, Uno K, et al. The emerging role of plasma lipidomics in cardiovascular drug discovery. *Expert Opin Drug Discov* 2012; 7:63.
31. Quinn PJ, Rainteau D, Wolf C. Lipidomics of the red cell in diagnosis of human disorders. *Methods Mol Biol* 2009; 579:127.
32. Paik S, Tang G, Shak S, et al. Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *J Clin Oncol* 2006; 24:3726.
33. Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 2004; 351:2817.
34. Friedman DR, Weinberg JB, Barry WT, et al. A genomic approach to improve prognosis and predict therapeutic response in chronic lymphocytic leukemia. *Clin Cancer Res* 2009; 15:6947.
35. Scheuner MT, Sieverding P, Shekelle PG. Delivery of genomic medicine for common chronic adult diseases: a systematic review. *JAMA* 2008; 299:1320.
36. Walt DR, Kuhlik A, Epstein SK, et al. Lessons learned from the introduction of personalized genotyping into a medical school curriculum. *Genet Med* 2011; 13:63.
37. http://deansnewsletter.stanford.edu/archive/09_28_09.html (Accessed on September 28, 2009).
38. Hauskeller C. Direct to consumer genetic testing. *BMJ* 2011; 342:d2317.
39. Ormond KE, Wheeler MT, Hudgins L, et al. Challenges in the clinical application of whole-genome sequencing. *Lancet* 2010; 375:1749.
40. Samani NJ, Tomaszewski M, Schunkert H. The personal genome--the future of personalised medicine? *Lancet* 2010; 375:1497.
41. Creighton S, Almqvist EW, MacGregor D, et al. Predictive, pre-natal and diagnostic genetic testing for Huntington's disease: the experience in Canada from 1987 to 2000. *Clin Genet* 2003; 63:462.

Topic 14599 Version 18.0

Disclosures

Disclosures: Benjamin A Raby, MD, MPH Other Financial Interest (Spouse): Parexel International [Contract research organization (Double-blind randomized clinical trials)]. Anne Slavotinek, MBBS, PhD Nothing to disclose. Jennifer S Tirnauer, MD Nothing to disclose.

Contributor disclosures are reviewed for conflicts of interest by the editorial group. When found, these are addressed by vetting through a multi-level review process, and through requirements for references to be provided to support the content. Appropriately referenced content is required of all authors and must conform to UpToDate standards of evidence.

Conflict of interest policy

Genetic association studies: Principles and applications

Authors

John Attia, MD, PhD, FRCPC,
FRACP
Gordon Guyatt, MD

Section Editor

Benjamin A Raby, MD, MPH

Deputy Editor

Jennifer S Tirnauer, MD

All topics are updated as new evidence becomes available and our [peer review process](#) is complete.

Literature review current through: Feb 2015. | **This topic last updated:** Apr 28, 2014.

INTRODUCTION — Interest in the genetic determinants of disease originated with Gregor Mendel's observations on the genetics of the pea in the 1860s. Subsequent studies have identified many of the genes responsible for "Mendelian" diseases, conditions that follow a clear familial pattern. However, diseases inherited in a Mendelian fashion (eg, Huntington disease and cystic fibrosis) are rare.

More recently, the Human Genome Project has generated growing interest in genetic contribution to "complex" diseases. Such diseases combine some familial predisposition with a large environmental contribution. Examples include cardiovascular disease, diabetes, asthma, cancer, and obesity.

This topic will discuss the principles and clinical applications of genetic association studies in the elucidation of the genetic basis for common diseases with complex genetic components. Additional discussions of modes of inheritance and a glossary of genetic terms are presented separately. (See ["Overview of Mendelian inheritance"](#) and ["Principles of complex trait genetics"](#) and ["Non-Mendelian inheritance patterns of monogenic diseases"](#) and ["Glossary of genetic terms"](#).)

TERMINOLOGY AND STUDY DESIGN

Genetic association studies — Genetic association studies are analogous to traditional epidemiologic association studies. Instead of seeking association between traditional risk variables (eg, hypertension) and disease outcomes (eg, stroke), a genetic association study looks for an association between a genetic variable and a specified condition.

Single nucleotide polymorphisms — The genetic variable most commonly studied is a single nucleotide polymorphism (SNP, pronounced "snip"). A SNP is a base pair change that occurs in at least 1 percent of the population. SNPs are not clearly deleterious, in contrast to mutations, which occur less frequently and generally have an adverse effect on protein function. (See ["Overview of genetic variation"](#).)

Study design — The usual study design for association studies is case-control or nested case-control, where controls are selected from the general population. Other designs, using family-based data (eg, pedigrees, parent-offspring trios, or sibling pairs) are becoming less popular due to difficulties in obtaining sufficient sample size and the need for specialized statistical methods for analysis [1].

Candidate gene approach — Earlier gene association studies employed a "candidate" gene approach in which a genetic variant of interest was selected on the basis of the known biology of a disease. The genetic variant may be of interest because of its presumed biologic function, or association with a disease in previous studies. In this approach, one or a small number of known variants are genotyped, usually by PCR methods, in a number of cases and controls.

Genome-wide approach (GWAS) — Genome-wide association studies (GWAS, pronounced "gee-wass") test hundreds of thousands of genes simultaneously ([figure 1](#)). GWAS use microarray technology, which "arrays" a large series of test sequences on a solid surface. Technological advances

with microarrays have enabled researchers to genotype an individual for between 500,000 and 2 million SNPs on a platform no bigger than a microscope slide.

The principle behind microarrays is that DNA from an individual is hybridized against an array of short oligonucleotide probes (ie, DNA sequences) that are immobilized on a surface (the “array”). Sequences on the array are chosen to assay the regions of the genome with the most variation. If the individual being tested has the DNA sequence that is complementary to a probe on the array, the DNA binds and is detected ([figure 2](#)).

Most GWAS are case-control studies, where a group of cases and a group of separate controls are gathered, DNA is isolated, and microarray data is obtained on each individual. The case-control design facilitates large sample sizes and therefore power, which is particularly important for detecting potentially small genetic effects. GWAS have also been performed using longitudinal cohorts, with cases and controls “nested” within the cohort.

At the most basic level, the distribution of the three genotypes for each SNP (AA, Aa, aa) is compared between cases and controls using a Chi-square test. This is repeated over and over for each of the 500,000 or more SNPs on the array. As human genetic variation is increasingly catalogued, for example in databases such as “Hap Map” and “1000 Genomes”, this information can be used to impute genetic variants that were not directly measured by the microarray. In other words, the 500,000 to 2 million SNPs that are directly assayed are used to make a best guess on the other SNPs that were not measured, often yielding up to 6 million SNPs or more to test for association.

These association studies have exploded in number in the medical literature. Between 2007 and 2009, GWAS identified nearly 1000 SNPs associated with a range of human traits and common diseases [2]. By March 2011, that number had risen to almost 4000 SNPs for over 200 conditions [3]. GWAS represented 3 percent of all genetic association studies reported in 2003 (24 studies); this number rose to over 12 percent in 2010 (1010 studies) and has continued to rise, with 9949 GWAS studies catalogued by early 2014.

Despite the issue of “missing heritability” (see [‘Missing heritability’](#) below), GWAS have had some notable successes identifying genes of large effect. An example is the relationship between a complement factor H (CFH) polymorphism and age-related macular degeneration (AMD) (see [“Age-related macular degeneration: Clinical presentation, etiology, and diagnosis”, section on ‘Genetic factors’](#)). This polymorphism is involved in regulation of the alternate complement pathway that results in increased inflammation. Individuals with one allele coding for a histidine substitution for tyrosine in position 402 of the CFH protein (CFH Y402H) are at increased risk for AMD. Meta-analysis indicates a possible multiplicative model in that each copy of the C allele at this locus (coding for histidine) increases the risk of AMD by approximately 2.5-fold compared to the T allele (ie, the TC heterozygote group has an odds ratio [OR] for AMD of 2.5 and the CC homozygote has an odds ratio of approximately 6 relative to the TT genotype group) [4].

GENETIC DETERMINANTS OF COMPLEX TRAITS

Common disease-common variant hypothesis — The basic assumption regarding genetic determinants of complex disease is that common variants in many genes will each lead to a small rise (or fall) in the risk of disease, and that the overall risk of disease is determined by the combination of multiple variants and environmental exposures [5]. The single nucleotide polymorphisms (SNPs) chosen for inclusion on microarrays are usually the more common variants in the population.

A competing theory is “common disease-rare variant,” in which it is anticipated that most common diseases will be due to multiple rare variants (<1 percent), perhaps even unique mutations. If this is the case, GWAS would be less useful, because they are designed to reveal associations that occur commonly; and one would need to move to whole genome sequencing to detect rare variants associated

with a disease. The technology for whole genome sequencing is advancing quickly, and it is anticipated that this technology will become competitively priced, and hence more common, in the near future. (See ["Principles and clinical applications of next-generation DNA sequencing".](#))

In reality, the dichotomy between common and rare variants is not absolute, and results from one type of association will likely complement the other. As an example, sequencing of the cystic fibrosis gene revealed that one major allele, delta F508, accounts for 67 percent of disease in Europe, but the remainder is due to hundreds of rarer disease alleles/mutations [6].

In the last few years, GWAS have led to the successful identification of many genes involved in complex disease, such as coronary artery disease, type 2 diabetes, stroke, multiple sclerosis, breast cancer, bipolar disorder, rheumatoid arthritis, and Alzheimer's disease [7-15]. These studies largely support the hypothesis that multiple genes, each conferring small relative risks (ie, 1.2-1.8), collectively contribute to increased disease risk. Other GWAS results have shed light on the pathogenesis of a disease process (eg, the role of melatonin in type 2 diabetes) [16,17].

Missing heritability — Despite many successes, GWAS have not explained as much of the variance in complex traits as was originally expected. Heritability studies look at the variance in a trait and use family relationships (eg, twins and siblings) to partition how much of the variance is due to genetic factors and how much is due to environmental factors. As an example, studies have estimated that height is 80 to 90 percent "heritable," but GWAS and meta-analyses of GWAS on height have turned up more than 40 genes, accounting for less than 5 percent of the variance [18,19]. This discrepancy between the numerator (amount of heritability explained from GWAS SNPs) and denominator (amount of total heritability from family/population data) has been termed "missing heritability" [2,20].

There are several hypotheses about what may account for the "missing heritability":

- Some of the variation may be explained by rare alleles that will only be detected by whole genome sequencing, rather than SNPs.
- Some of the variation may be explained by copy number variants (CNVs). CNVs are polymorphisms that consist of deletions or duplications of whole segments of DNA, spanning at least 1000 bases [21]. Although CNVs can show association for some diseases where SNPs have not yielded any signal (eg, autism) [22], it seems unlikely that common CNVs will explain most of the unaccounted heritability for the majority of complex traits [23]. An investigation of the association of CNVs with eight complex diseases found that most CNV signals were in areas that had already been highlighted by SNP associations, and most CNVs were "tagged" by SNPs [24]. (See ["Genomic disorders: An overview", section on 'Copy number variations'.](#))
- Current methodology may not be adequate to detect gene/gene interactions (epistasis). This is particularly true if each genetic variant alone shows no effect but shows an effect in the presence of the other variant. Modelling indicates that gene/gene and higher order interactions could account for missing heritability, but would not have been detected due to insufficient sample size in the studies to date [25].
- Heritability may also be accounted for by epigenetic changes (eg, methylation of DNA, modification of histone proteins). There is increasing evidence that these non-DNA changes can also be passed on from generation to generation [26].

Others, however, maintain that there is no "missing heritability," arguing that estimates of heritability are complex and may have been interpreted simplistically. High heritability does not necessarily mean genetic determination or genes of large effect, but may reflect a large component of shared environment as well [27]. Overestimates of total heritability may have created "phantom heritability" (ie, presumed genetic contributions that could not be identified even if all causative genetic variants were identified)

[25]. This line of argument is supported by ability to explain much of the variance in a trait as long as enough SNPs are evaluated. For example, one model using just under 30,000 SNPs could account for 45 percent of the variance in height [28]. Fitting all genotyped SNPs explains a substantial portion of the heritability for such complex phenotypes as height, body mass index (BMI), von Willebrand factor levels, and QT interval [29].

JUDGING STUDY VALIDITY — It is increasingly necessary for physicians to become familiar with GWAS data and how to gauge validity of these studies.

Potential biases — There are multiple potential biases for case-control studies in traditional epidemiology. The potential for bias is less, though not absent, in genetic studies. In contrast to most environmental exposures, the genetic "exposure" is not chosen by the participant, doesn't vary with age or calendar year, is not subject to recall bias, and is not influenced by the disease or treatment.

One potential bias that cannot be avoided, however, is survivor bias, in that those who are included in a case-control study are those who have survived the initial "insult" (figure 3). As an example, in a study of stroke, results may differ between a case-control and cohort GWAS if certain genes are associated with initial stroke severity, causing early mortality before recruitment into a case-control study.

Validity framework — Many groups have published guidelines for critical appraisal of genetic association studies [30,31]. One framework, from the "Users' Guides to the Medical Literature" series, asks the following questions [32,33]:

- Was the disease **phenotype** properly defined and accurately recorded by someone blind to the genetic information?
- Have any potential differences between disease and non-disease groups, particularly **ancestry**, been properly addressed?
- Was measurement of the **genetic variants** (ie, genotyping) unbiased and accurate?
- Do the genotype proportions observe **Hardy-Weinberg equilibrium**?
- Have the investigators adjusted for **multiple comparisons**?
- Are the results **consistent with other studies**?

The issues raised in these questions are as follows:

Phenotyping — What may seem like a unique disease (eg, ischemic stroke), may be a heterogeneous group of diseases (eg, large vessel stroke, lacunar stroke, and embolic stroke), or may be variously defined (eg, clinical signs only, or CT/MRI imaging confirmation). Investigators may selectively report only the disease definition or subgroup that yields a significant association [34].

Alternatively, an apparent single disease entity may represent genetically separate but clinically similar diseases ("genetic heterogeneity"), as seen with epilepsy [35]. Including diseases with different genetic backgrounds may obscure a true genetic association.

To avoid bias, those who are doing the phenotyping should be blind to the results of the genotyping and vice versa. The possibility of bias is greater for candidate gene studies where single nucleotide polymorphisms (SNPs) are typed manually in contrast to GWAS in which the genotype identification is automated.

Ancestry and comorbidities — Although genetic association studies are less prone to traditional confounding than other case control studies, there are two potential sources of bias:

- Ethnicity or racial mix — This particular form of confounding, referred to as "population

stratification," occurs when disease and non-disease populations include a different ethnic/racial mix. If the likelihood of developing the condition of interest varies with ancestry, then any SNP with a different allele frequency between ancestral groups will appear to be linked (spuriously) to the disease ([figure 4](#)).

As with traditional confounders, the way to correct for this is to measure the confounder and adjust for it in the analysis or to stratify the analysis. Ethnic subgroups may be identified by self-report or by statistical analysis of SNPs themselves [[36,37](#)]. As an example, a spurious association between the CYP3A4-V polymorphism and prostate cancer in African-Americans disappeared when results were adjusted for additional genetic markers associated with ancestry in the population studied [[38](#)].

- **Comorbidities/associated phenotypes** — An apparent association between a genotype and condition of interest may represent a direct association between the genotype and another condition, with the disease of interest being linked to this other condition. As an example, two GWAS showed an association between type 2 diabetes and a SNP in the FTO (fat mass and obesity associated) gene [[8,39](#)]. These studies selected diabetic patients and controls irrespective of their body mass index (BMI); another study that matched diabetic patients and controls on BMI showed no association. Thus, obesity is associated with diabetes and the initial SNP was thought to be linked with diabetes when in fact it was linked with the associated obesity.

Accuracy of genotyping — Genotyping error is a threat to the validity of genetic association studies. Genotyping errors may arise if there is a problem with the DNA samples or with the technology that is employed to identify alleles [[40](#)].

DNA samples may differ between diseased and non-diseased participants in ways that lead to inaccuracies in genotyping. As an example, in a GWAS for type 2 diabetes, blood used for the control population was stored from a cohort in 1958 while samples from patients with known diabetes were from the present day. The older blood resulted in genotyping errors that led to false positive SNP associations [[41](#)].

Genotyping error rates vary widely, from <1 to 30 percent [[42](#)]; rates up to a few percent are common in even the best studies [[43](#)]. It is common practice for investigators to cull SNPs that do not have a high call rate (eg, 95 to 98 percent). Call rate refers to the percent of DNA samples that can be genotyped. This increases the validity of the data but does not ensure that the genotype information is correct. The genotype may be misidentified as, for example, when one of the alleles in a heterozygote is harder to identify than the other, resulting in a heterozygote being mislabeled a homozygote. Checking for Hardy-Weinberg equilibrium may be one way of detecting genotyping problems. (See '[Hardy-Weinberg equilibrium](#)' below.)

Hardy-Weinberg equilibrium — In the same way that most continuous variables in medicine observe a normal distribution, most allele distributions observe what is called Hardy-Weinberg equilibrium (HWE). This describes the steady state where there are no selective forces (eg, mutation, inbreeding, selective survival) acting on a particular locus or gene. The Hardy-Weinberg law states that if there are two alleles (named A and a) at a particular locus, with frequency p and q respectively, then after one generation of random mating, the genotype frequencies of the AA, Aa, and aa groups in the population will be p^2 , $2pq$, and q^2 , respectively. Given that there are only two alleles possible, A or a, then:

- $p + q = 1$,

and

- $p^2 + 2pq + q^2 = 1$.

HWE is commonly used as a quality measure. It has become general practice, in a genetic association study, to check whether the allele frequencies at a particular SNP observe HWE proportions. Results are considered to be consistent with HWE when the p-value is <0.05 . HWE calculators are available online [44,45]. For a cohort study, HWE should be tested in the whole study population, whereas for a case-control study, it should be tested in the controls, since these are supposedly representative of the general population.

However, HWE is non-specific and may be insensitive [46]. Genotyping error and population stratification are two of many factors that may upset HWE proportions. Other factors that may affect HWE include new mutations, inbreeding, or a selective advantage of one allele over another.

Multiple comparisons — The usual p-value for judging statistical significance in traditional epidemiological studies (0.05) is intended for a single comparison. The .05 threshold means that a result as extreme or more extreme than the one seen will occur by chance once in 20 times; this is taken as low enough to not be chance and to indicate a significant association. However, if one is looking at 100 SNPs, then by chance alone one might expect 5 SNPs to reach this threshold; in fact there is over a 99 percent chance that at least one will reach this threshold.

In GWAS in which over 500,000 SNPs are tested simultaneously, the possibilities for false-positive are countless. The current consensus is that for such large scale studies, a p-value in the range of 5×10^{-8} (in contrast to the usual 5×10^{-2}) should be considered the threshold [47].

Replication — It is essential that studies replicate their results, given the large potential for false positive signals. This is commonly done by repeating the study in different populations.

There is a growing movement to combine GWAS results using meta-analysis. Most of the genetic associations between SNPs and complex diseases are small in magnitude, and therefore even sizable studies may fail to detect underlying associations [48]. Meta-analysis improves the identification of replicable associations and increases precision by increasing power. This is illustrated by studies in Crohn's disease [49] and in asthma [50].

The HuGE Net website lists many of the meta-analyses performed to date and also hosts the HuGE Navigator, where one can find out what single studies, genome-wide association studies, meta-analyses, and synopses are available [51,52].

Replication can also take the form of functional studies in vitro or in animal models. Demonstration by cell culture or mouse mutagenesis that genetic variants can lead to difference in protein level or function can provide powerful support for the effect of genetic variants. Follow-up studies can also provide evidence to support an initial GWAS result, including:

- Candidate gene association studies that focus on a few genetic variants that are near the site of the initial GWAS locus
- Fine mapping studies, involving a denser set of SNPs around the initial GWAS locus or involving resequencing of the entire locus to pick up rare variants

INTERPRETING GWAS RESULTS — Investigators usually report the magnitude of a genetic association using traditional measures of association: relative risks (RRs) in cohort studies, odds ratios (ORs) in case-control studies, and hazard ratios (HRs) in survival analyses that take account of the timing of events.

Understanding the magnitude of the risk depends on the genetic model involved. For dominant variant alleles (producing a protein isoform that dominates function), the presence of even one copy (ie, heterozygosity) will result in maximal increase in risk. For recessive variant alleles, both alleles must be present to result in an increase in risk (ie, heterozygotes will not show an increase in risk). In both cases,

a single RR, OR or HR describes the magnitude of the association.

If the effect of a variant allele is additive, then there is a "dose-response" effect: its presence in one gene will lead to an increase in risk, while its presence in both genes will lead to a further increase. There are two possible ways to calculate this further increase: one is to take the square of the risk (variably called the log-additive, per-allele, or multiplicative risk model), and the other is to take two times the risk (called the linear additive model). Recent work in diabetes indicates that most associations seem to follow the log-additive model [53].

It is important to understand that the magnitude of the effect does not necessarily translate into a causal relationship between the single nucleotide polymorphism (SNP) and the disease. It is possible that the SNP is only a marker for another SNP nearby that is linked (termed "linkage disequilibrium") and that is the true causal variant. While linkage disequilibrium does not detract from the potential use of the SNP as a clinical marker, it limits the potential to draw pathophysiologic conclusions from the identified SNP.

CLINICAL APPLICATION — A framework for translating GWAS results to clinical application has been developed and is presented here [32,33]:

- Does the genetic association improve predictive power beyond easily measured clinical variables?
- What are the absolute versus relative genetic effects?
- Is the risk-associated allele likely to be present in my patient?
- Is the patient likely better off knowing the genetic information?

These issues need to be addressed thoughtfully and quickly, particularly given the "direct-to-consumer" availability of genetic testing [54,55].

Predictive power — The effect of single SNPs in complex disease to date has been small (ie, odds ratios in the 1.1-1.6 range). Therefore, there has been interest in combining the effect of many genes into a genetic "profile" for greater clinical utility.

As an example, in one study, investigators created a profile of 5 SNPs associated with prostate cancer [56]; they found an OR of 1.6 for those who were homozygous or heterozygous for the risk allele at one SNP, and up to 4.5 for those who were homozygous or heterozygous for the risk allele at 4 SNPs. The greater the number of SNPs in each profile, however, the lower the number of people with that combination; hence the increased magnitude of effect is offset by the small number of people to whom that effect is relevant [57].

There are rare examples, particularly in the area of pharmacogenomics, where genetic associations may have clinical applications because of their large magnitude. As an example, a SNP in the thiopurine methyltransferase (TPMT) gene identifies individuals who are at increased risk of life-threatening hematologic toxicity from the chemotherapeutic agent 6-mercaptopurine [58]. Genotyping this SNP can avoid substantial harm by additional monitoring or substituting an alternative chemotherapeutic agent in those with the high-risk genotype. (See ["6-mercaptopurine \(6-MP\) metabolite monitoring and TPMT testing in the treatment of inflammatory bowel disease with 6-MP or azathioprine"](#) and ["Overview of the treatment of acute lymphoblastic leukemia in children and adolescents"](#).)

The most likely short-term clinical application of a genetic association is to provide prognostic information in the context of other known predictors. To be of value, the genetic marker must provide independent predictive power beyond traditional clinical predictive variables (eg, age, sex, smoking status) and beyond easily obtained surrogate measures of familial aggregation, such as family history. This is often not the case, particularly if the genetic polymorphism exerts its effect through a readily measured variable (eg, a gene controlling lipids exerts its effect through increases in LDL).

For dichotomous (eg, yes/no) outcomes, there are a number of statistical tools to quantitate how much predictive power the genetic information adds to existing data. One is to calculate the area under the Receiver Operating Characteristic (ROC) curve, an approach often used for diagnostic tests [59]. An ROC curve plots the true positive rate (sensitivity) on the y-axis against the false positive rate (1-specificity) on the x-axis (figure 5). An ROC curve with no greater predictive ability than chance would approximate a straight diagonal line from the origin (0, 0) to the upper right hand corner (1.0, 1.0). The area under the curve (AUC) would be 0.5. The visual representation of a perfectly predictive test would be a line that goes straight up the y-axis to 1.0 and then straight across the x-axis to 1.0 and would have an AUC of 1.

As an example, the addition of genetic data to a model for prediction of type 2 diabetes did not improve the model fit and resulted in appropriate reclassification of at most 4 percent of people [60].

However, a high AUC or even an increment in AUC with the addition of genetic variables does not guarantee that the genetic information will be clinically useful. Points along the ROC curve may still correspond to sensitivities and specificities that are too low to be used for screening or prognosis. Furthermore, even if the sensitivity and specificity are high, the usefulness of the genetic variant for classification of risk may be limited due to low allele frequency (ie, only a small proportion of the risk group will be detected or stratified correctly) [61].

In summary:

- A very small p-value does not necessarily translate into a useful clinical marker because effect size (RR or OR) may be small.
- A high effect size does not necessarily translate into a clinical useful marker because the AUC may be low.
- A high AUC does not necessarily translate into a clinical useful marker because the combined sensitivity and specificity may be insufficiently high for screening or prognostication.
- High sensitivity and specificity may not necessarily translate into a clinical useful marker because the allele frequency may be low.

These considerations show the difficulty in translating a statistically significant genetic predictor into a clinical useful and accurate clinical classifier.

Absolute versus relative effects — If the patient's risk of disease is low in the absence of a variant allele, even a 5- or 10-fold increase in risk in the presence of the allele may represent a small absolute increase in risk. Conversely, if the baseline risk is high, a modest increase in relative risk could impact clinical decision making.

As an example, the Factor V Leiden mutation increases the risk of venous thrombosis by about sixfold [62]. However, the baseline risk of thrombosis in the general population is sufficiently low (about 0.2 percent) that one would not use genotyping as a population screening test [63]. However, the prevalence of Factor V Leiden in patients with venous thrombosis is 12 to 20 percent, and thus testing some patients with established venous thrombosis may be appropriate [64,65]. (See ["Evaluating patients with established venous thromboembolism for acquired and inherited risk factors", section on 'Evaluation for hypercoagulable disorders'.](#))

Allele frequency in patient's population — In applying the results, clinicians must consider the likelihood that the particular allele is present in a particular patient. As an example, while factor V Leiden is relatively common in Caucasians (about 1 in 20 individuals are heterozygotes), it is virtually non-existent in Chinese populations. Hence, genotyping for factor V Leiden is unnecessary in a Chinese individual presenting with idiopathic deep venous thrombosis. (See ["Factor V Leiden and activated](#)

[protein C resistance: Clinical manifestations and diagnosis", section on 'Factor V genotypes'.\)](#)

Allele frequencies for various genes and populations of interest are available in the **Allele frequency database** (ALFRED) or at the HapMap website [66,67]. Some gene-disease associations may be restricted to a very select subgroup. As an example, BRCA1 was identified in patients with early onset breast cancer who had a strong family history [68]. This group, however, only accounts for approximately 5 percent of all breast cancers. Hence, this genetic association is not worth testing for in those who present with the usual, late-onset breast cancer without a strong family history. However, in certain ancestry groups such as Ashkenazi Jews who have a high prevalence of BRCA1 mutations, testing may be appropriate in women with breast or ovarian cancer (www.nccn.org). (See "[Genetic testing for hereditary breast and ovarian cancer syndrome](#)".)

Patient impact — Even if the magnitude of the genetic effect is not sufficiently high to be clinically useful in prognostication, it may still be useful clinically in terms of changing risk behavior [69]. Presenting personal genetic information may take advantage of the layperson's perception of DNA as their "life code" to prompt behavior change.

As an example, early evidence suggests that providing information about glutathione-S-transferase (GST) genotypes, which affect nicotine metabolism, may influence smoking cessation rates [70]. This potential benefit should be balanced against the potential worry of knowing that one's risk of health problems years in the future is increased, and the potential for increased insurance premiums or life insurance/disability denial.

BIOLOGICAL IMPLICATIONS — Apart from potential clinical uses, GWAS results can also shed light on new pathophysiological mechanisms. Principles have been proposed to move beyond identifying single nucleotide polymorphisms (SNPs) to defining their possible functional significance, which might include targeted resequencing, studying the SNP in other populations to further define and narrow linkage disequilibrium, exploring gene transcription functions and epigenetic regulation, and using model systems and cell models to further evaluate the proposed causal variants [3]. The challenge of dissecting the correlation between genotype and phenotype will require rigorous evaluation.

SUMMARY

- Genetic association studies seek to identify the genetic component to risk of non-Mendelian complex disorders. The basic assumption regarding genetic determinants of complex disease is that the overall risk of disease is determined by the combination of multiple common genetic variants, each with small effect, and environmental exposures. The candidate gene approach selects a genetic variant of potential interest and looks for this gene in cases and controls. The genome-wide approach looks for an association between millions of variants and the condition of interest. (See "[Terminology and study design](#)" above.)
- Genome-wide association studies (GWAS) have identified some significant single nucleotide polymorphisms (SNPs) associated with disease, but there is still debate about the amount of variance explained by these, (ie, whether there is "missing heritability"). (See "[Genetic determinants of complex traits](#)" above.)
- GWAS should be evaluated for validity, considering potential errors related to phenotyping, genotyping, ancestral differences between cases and controls, and multiple comparators. Hardy-Weinberg allele distribution and consistency between studies are other important considerations. (See "[Judging study validity](#)" above.)
- Questions in applying results of genetic association studies to clinical practice include the effect on predictive power, the magnitude of absolute versus relative effects, generalizability to the given patient, and whether the genetic information will be beneficial to the patient. (See "[Interpreting](#)

GWAS results' above.)

Use of UpToDate is subject to the [Subscription and License Agreement](#).

REFERENCES

1. Dawn Teare M, Barrett JH. Genetic linkage studies. *Lancet* 2005; 366:1036.
2. Visscher PM, Montgomery GW. Genome-wide association studies and human disease: from trickle to flood. *JAMA* 2009; 302:2028.
3. Freedman ML, Monteiro AN, Gayther SA, et al. Principles for the post-GWAS functional characterization of cancer risk loci. *Nat Genet* 2011; 43:513.
4. Thakkinstian A, Han P, McEvoy M, et al. Systematic review and meta-analysis of the association between complement factor H Y402H polymorphisms and age-related macular degeneration. *Hum Mol Genet* 2006; 15:2784.
5. Iyengar SK, Elston RC. The genetic basis of complex traits: rare variants or "common gene, common disease"? *Methods Mol Biol* 2007; 376:71.
6. Estivill X, Bancells C, Ramos C. Geographic distribution and regional origin of 272 cystic fibrosis mutations in European populations. The Biomed CF Mutation Analysis Consortium. *Hum Mutat* 1997; 10:135.
7. Samani NJ, Erdmann J, Hall AS, et al. Genomewide association analysis of coronary artery disease. *N Engl J Med* 2007; 357:443.
8. Zeggini E, Weedon MN, Lindgren CM, et al. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* 2007; 316:1336.
9. Matarín M, Brown WM, Scholz S, et al. A genome-wide genotyping study in patients with ischaemic stroke: initial analysis and data release. *Lancet Neurol* 2007; 6:414.
10. International Multiple Sclerosis Genetics Consortium, Hafler DA, Compston A, et al. Risk alleles for multiple sclerosis identified by a genomewide study. *N Engl J Med* 2007; 357:851.
11. Easton DF, Pooley KA, Dunning AM, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* 2007; 447:1087.
12. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007; 447:661.
13. Plenge RM, Seielstad M, Padyukov L, et al. TRAF1-C5 as a risk locus for rheumatoid arthritis--a genomewide study. *N Engl J Med* 2007; 357:1199.
14. Coon KD, Myers AJ, Craig DW, et al. A high-density whole-genome association study reveals that APOE is the major susceptibility gene for sporadic late-onset Alzheimer's disease. *J Clin Psychiatry* 2007; 68:613.
15. Li H, Wetten S, Li L, et al. Candidate single-nucleotide polymorphisms from a genomewide association study of Alzheimer disease. *Arch Neurol* 2008; 65:45.
16. Prokopenko I, Langenberg C, Florez JC, et al. Variants in MTNR1B influence fasting glucose levels. *Nat Genet* 2009; 41:77.
17. Bouatia-Naji N, Bonnefond A, Cavalcanti-Proença C, et al. A variant near MTNR1B is associated with increased fasting plasma glucose levels and type 2 diabetes risk. *Nat Genet* 2009; 41:89.
18. Weedon MN, Lango H, Lindgren CM, et al. Genome-wide association analysis identifies 20 loci that influence adult height. *Nat Genet* 2008; 40:575.
19. Lettre G, Jackson AU, Gieger C, et al. Identification of ten loci associated with height highlights new biological pathways in human growth. *Nat Genet* 2008; 40:584.

20. Maher B. Personal genomes: The case of the missing heritability. *Nature* 2008; 456:18.
21. Wain LV, Armour JA, Tobin MD. Genomic copy number variation, human health, and disease. *Lancet* 2009; 374:340.
22. Weiss LA, Shen Y, Korn JM, et al. Association between microdeletion and microduplication at 16p11.2 and autism. *N Engl J Med* 2008; 358:667.
23. Conrad DF, Pinto D, Redon R, et al. Origins and functional impact of copy number variation in the human genome. *Nature* 2010; 464:704.
24. Wellcome Trust Case Control Consortium, Craddock N, Hurles ME, et al. Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature* 2010; 464:713.
25. Zuk O, Hechter E, Sunyaev SR, Lander ES. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc Natl Acad Sci U S A* 2012; 109:1193.
26. Skinner MK, Manikkam M, Guerrero-Bosagna C. Epigenetic transgenerational actions of environmental factors in disease etiology. *Trends Endocrinol Metab* 2010; 21:214.
27. Visscher PM, Hill WG, Wray NR. Heritability in the genomics era--concepts and misconceptions. *Nat Rev Genet* 2008; 9:255.
28. Yang J, Benyamin B, McEvoy BP, et al. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* 2010; 42:565.
29. Yang J, Manolio TA, Pasquale LR, et al. Genome partitioning of genetic variation for complex traits using common SNPs. *Nat Genet* 2011; 43:519.
30. Little J, Bradley L, Bray MS, et al. Reporting, appraising, and integrating data on genotype prevalence and gene-disease associations. *Am J Epidemiol* 2002; 156:300.
31. Little J, Higgins JP, Ioannidis JP, et al. STrengthening the REporting of Genetic Association studies (STREGA): an extension of the STROBE Statement. *Ann Intern Med* 2009; 150:206.
32. Attia J, Ioannidis JP, Thakkinstian A, et al. How to use an article about genetic association: B: Are the results of the study valid? *JAMA* 2009; 301:191.
33. Attia J, Ioannidis JP, Thakkinstian A, et al. How to use an article about genetic association: C: What are the results and will they help me in caring for my patients? *JAMA* 2009; 301:304.
34. Contopoulos-Ioannidis DG, Alexiou GA, Gouvias TC, Ioannidis JP. An empirical evaluation of multifarious outcomes in pharmacogenetics: beta-2 adrenoceptor gene polymorphisms in asthma treatment. *Pharmacogenet Genomics* 2006; 16:705.
35. Berkovic SF, Scheffer IE. Genetics of the epilepsies. *Epilepsia* 2001; 42 Suppl 5:16.
36. Barnholtz-Sloan JS, McEvoy B, Shriver MD, Rebbeck TR. Ancestry estimation and correction for population stratification in molecular epidemiologic association studies. *Cancer Epidemiol Biomarkers Prev* 2008; 17:471.
37. Pritchard JK, Rosenberg NA. Use of unlinked genetic markers to detect population stratification in association studies. *Am J Hum Genet* 1999; 65:220.
38. Kittles RA, Chen W, Panguluri RK, et al. CYP3A4-V and prostate cancer in African Americans: causal or confounding association because of population stratification? *Hum Genet* 2002; 110:553.
39. Frayling TM, Timpson NJ, Weedon MN, et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* 2007; 316:889.
40. Leek JT, Scharpf RB, Bravo HC, et al. Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat Rev Genet* 2010; 11:733.
41. Clayton DG, Walker NM, Smyth DJ, et al. Population structure, differential bias and genomic control in a large-scale, case-control association study. *Nat Genet* 2005; 37:1243.

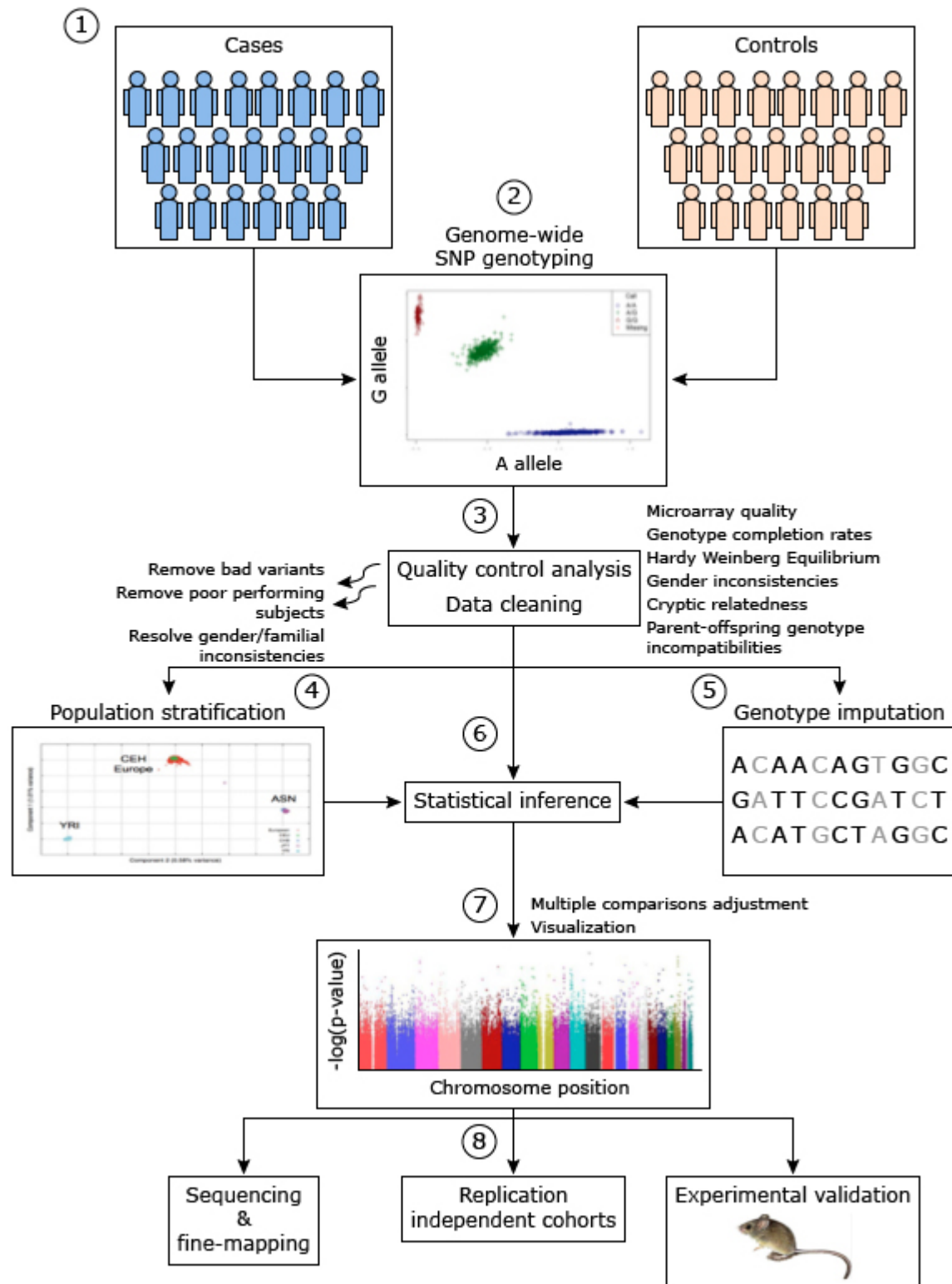
42. Akey JM, Zhang K, Xiong M, et al. The effect that genotyping errors have on the robustness of common linkage-disequilibrium measures. *Am J Hum Genet* 2001; 68:1447.
43. Pompanon F, Bonin A, Bellemain E, Taberlet P. Genotyping errors: causes, consequences and solutions. *Nat Rev Genet* 2005; 6:847.
44. www.tufts.edu/~mcourt01/Documents/Court%20lab%20-%20HW%20calculator.xls (Accessed on March 05, 2014).
45. www.oege.org/software/hwe-mr-calc.shtml (Accessed on March 05, 2014).
46. Cox DG, Kraft P. Quantification of the power of Hardy-Weinberg equilibrium testing to detect genotyping error. *Hum Hered* 2006; 61:10.
47. McCarthy MI, Abecasis GR, Cardon LR, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* 2008; 9:356.
48. Moonesinghe R, Khoury MJ, Liu T, Ioannidis JP. Required sample size and nonreplicability thresholds for heterogeneous genetic associations. *Proc Natl Acad Sci U S A* 2008; 105:617.
49. Franke A, McGovern DP, Barrett JC, et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet* 2010; 42:1118.
50. Moffatt MF, Gut IG, Demenais F, et al. A large-scale, consortium-based genomewide association study of asthma. *N Engl J Med* 2010; 363:1211.
51. Yu W, Wulf A, Yesupriya A, et al. HuGE Watch: tracking trends and patterns of published studies of genetic association and human genome epidemiology in near-real time. *Eur J Hum Genet* 2008; 16:1155.
52. www.hugenavigator.net (Accessed on March 05, 2014).
53. Salanti G, Southam L, Altshuler D, et al. Underlying genetic models of inheritance in established type 2 diabetes associations. *Am J Epidemiol* 2009; 170:537.
54. Hunter DJ, Khoury MJ, Drazen JM. Letting the genome out of the bottle--will we get our wish? *N Engl J Med* 2008; 358:105.
55. Janssens AC, Gwinn M, Bradley LA, et al. A critical appraisal of the scientific basis of commercial genomic profiles used to assess health risks and personalize health interventions. *Am J Hum Genet* 2008; 82:593.
56. Zheng SL, Sun J, Wiklund F, et al. Cumulative association of five genetic variants with prostate cancer. *N Engl J Med* 2008; 358:910.
57. Janssens AC, van Duijn CM. Genome-based prediction of common diseases: advances and prospects. *Hum Mol Genet* 2008; 17:R166.
58. McLeod HL, Krynetski EY, Relling MV, Evans WE. Genetic polymorphism of thiopurine methyltransferase and its clinical relevance for childhood acute lymphoblastic leukemia. *Leukemia* 2000; 14:567.
59. Irwig L, Bossuyt P, Glasziou P, et al. Designing studies to ensure that estimates of test accuracy are transferable. *BMJ* 2002; 324:669.
60. Meigs JB, Shrader P, Sullivan LM, et al. Genotype score in addition to common risk factors for prediction of type 2 diabetes. *N Engl J Med* 2008; 359:2208.
61. Jakobsdottir J, Gorin MB, Conley YP, et al. Interpretation of genetic association studies: markers with replicated highly significant odds ratios may be poor classifiers. *PLoS Genet* 2009; 5:e1000337.
62. Emmerich J, Rosendaal FR, Cattaneo M, et al. Combined effect of factor V Leiden and prothrombin 20210A on the risk of venous thromboembolism--pooled analysis of 8 case-control studies including 2310 cases and 3204 controls. Study Group for Pooled-Analysis in Venous Thromboembolism. *Thromb Haemost* 2001; 86:809.

63. Cushman M, Tsai AW, White RH, et al. Deep vein thrombosis and pulmonary embolism in two cohorts: the longitudinal investigation of thromboembolism etiology. *Am J Med* 2004; 117:19.
64. Ridker PM, Hennekens CH, Lindpaintner K, et al. Mutation in the gene coding for coagulation factor V and the risk of myocardial infarction, stroke, and venous thrombosis in apparently healthy men. *N Engl J Med* 1995; 332:912.
65. Koster T, Rosendaal FR, de Ronde H, et al. Venous thrombosis due to poor anticoagulant response to activated protein C: Leiden Thrombophilia Study. *Lancet* 1993; 342:1503.
66. The ALlele FREquency Database. Available at: <http://alfred.med.yale.edu/alfred/index.asp> (Accessed on March 05, 2014).
67. International HapMap Project. Available at: www.hapmap.org (Accessed on March 05, 2014).
68. Miki Y, Swensen J, Shattuck-Eidens D, et al. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* 1994; 266:66.
69. Marteau TM, Lerman C. Genetic risk and behavioural change. *BMJ* 2001; 322:1056.
70. Hamajima N, Suzuki K, Ito Y, Kondo T. Genotype announcement to Japanese smokers who attended a health checkup examination. *J Epidemiol* 2006; 16:45.

Topic 2902 Version 16.0

GRAPHICS

Stages of a genome-wide association study



(1, 2) Well-characterized cases and controls of similar ancestry are genotyped using oligonucleotide microarrays, typically consisting of assays for 500,000 to 1,000,000 SNPs. A representative cartesian plot of results for one such assay is presented. Individuals

homozygous for the major allele are plotted in blue, homozygotes for the minor allele in red, and heterozygotes in green.

(3) Next, extensive quality control analysis is conducted, including assessments at the level of the individual samples (microarray quality, subject genotype completion rates, gender inconsistencies) and individual assays (assay quality scores, genotype completion rates, Hardy Weinberg equilibrium, parent-child genotype inconsistencies). Screens for cryptic sample relatedness are performed using Identity By State statistics. Data are then filtered, with removal of results for problematic subjects or assays.

(4) Principal components analysis (shown) enables screening for evidence of population stratification. If detected, outlier subjects can be removed from consideration, and the eigenvectors generated can be used as covariates in the subsequent association testing.

(5) Using reference genotype panels, additional SNP genotypes for untyped variants (grey SNPs) can be imputed from the typed variants (in black) and considered for association testing.

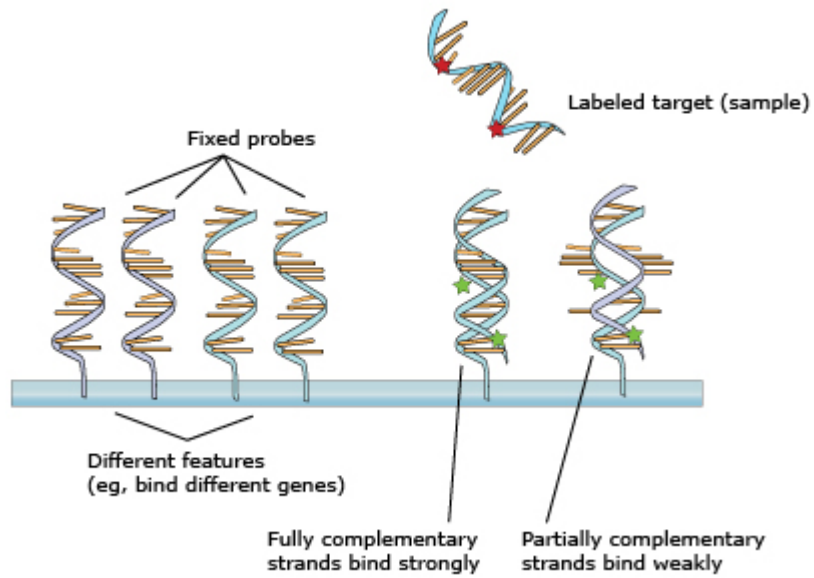
(6) Statistical inference is performed by testing individual variants for association by logistic regression for case-control studies or linear regression for quantitative traits, with adjustment for relevant covariates, including eigenvectors. To accommodate the large number of hypotheses tested, p-values must be adjusted, with more stringent cutoffs p-values designated in advance.

(7) Results can be visualized using so-called "Manhattan Plots," where the individual $-\log_{10}$ p-values are plotted as a function of physical position.

(8) Significant associations can be passed for additional validation, including replication studies in independent cohorts, sequencing and fine-mapping studies to identify causal variants, and experimental validation to confirm biological relevance and function. Additional modeling, including gene-by-gene interaction and gene-by-environment modeling, can also be conducted.

Graphic 54220 Version 2.0

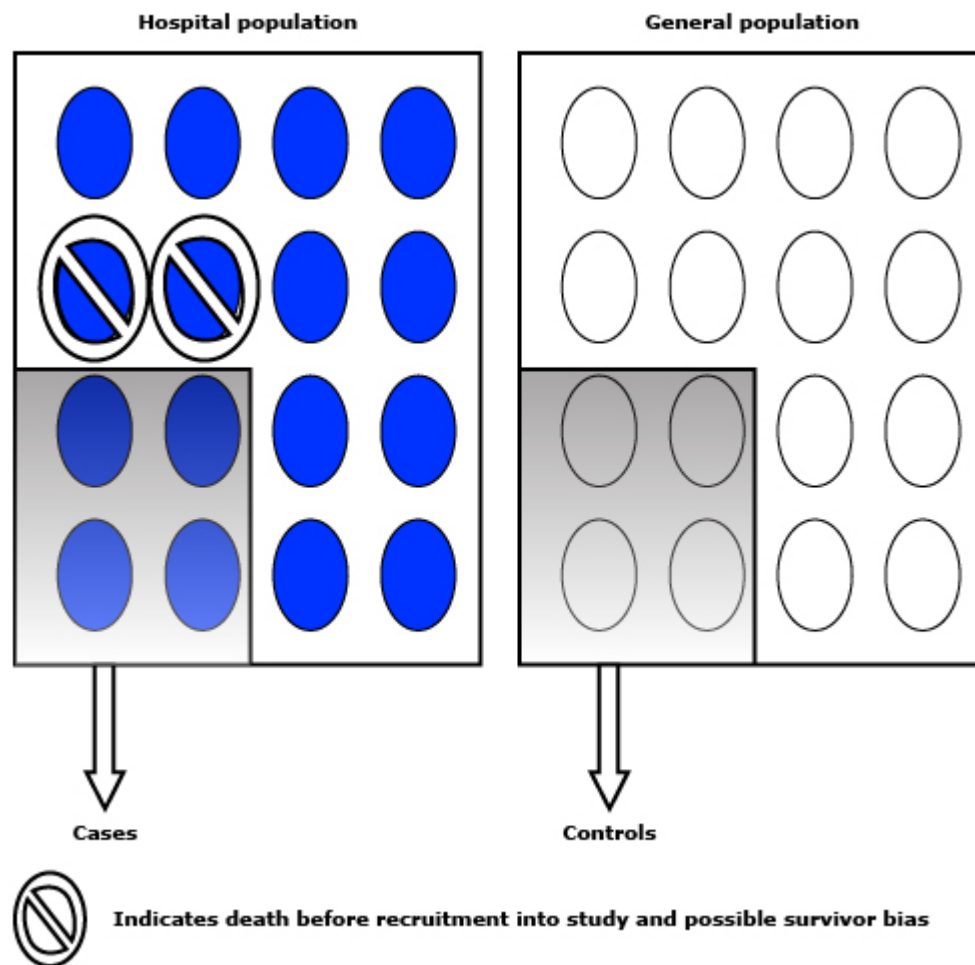
Hybridization



Reproduced from Wikipedia. Available at: http://en.wikipedia.org/wiki/File:NA_hybrid.svg.

Graphic 52164 Version 1.0

Survivor bias

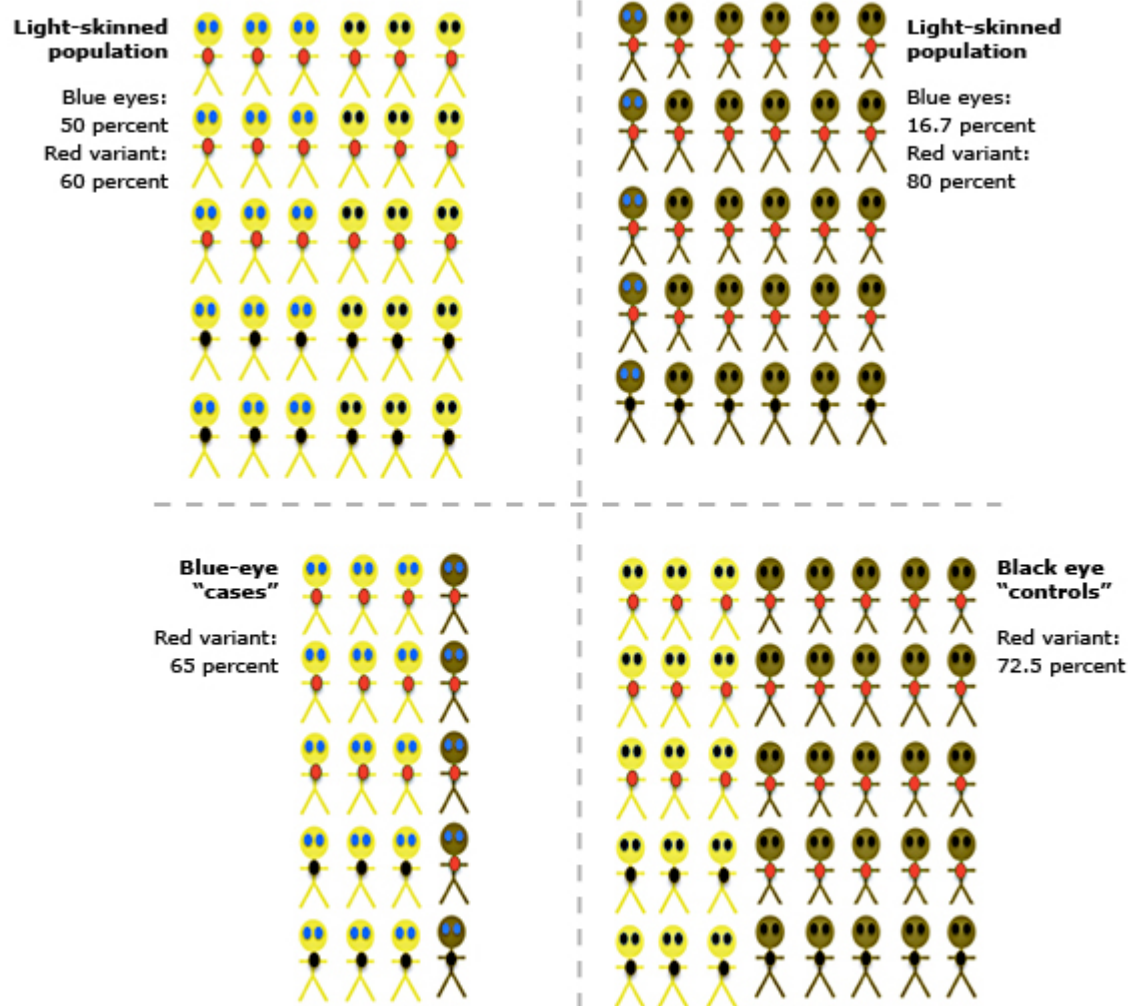


In this case-control study, cases are drawn from a hospital population. Out of all those with disease in the hospital (16 blue circles), 4 cases agree to be part of the study (shaded square). These 4 may represent a biased sample in that those with severe disease or those who die before being approached for participation are not represented. Out of all those with no disease in the general population, 4 controls agree to be part of the study. By comparing these 4 cases and 4 controls, any genes found to be selectively enriched (or depleted) in the cases may be linked to occurrence of the disease but may also be linked to severity or survival.

In a prospective study, blood is taken at baseline so that even if someone develops disease and dies before recruitment into a case-control study could happen they are still represented in the cohort study, ie, the cases are more representative.

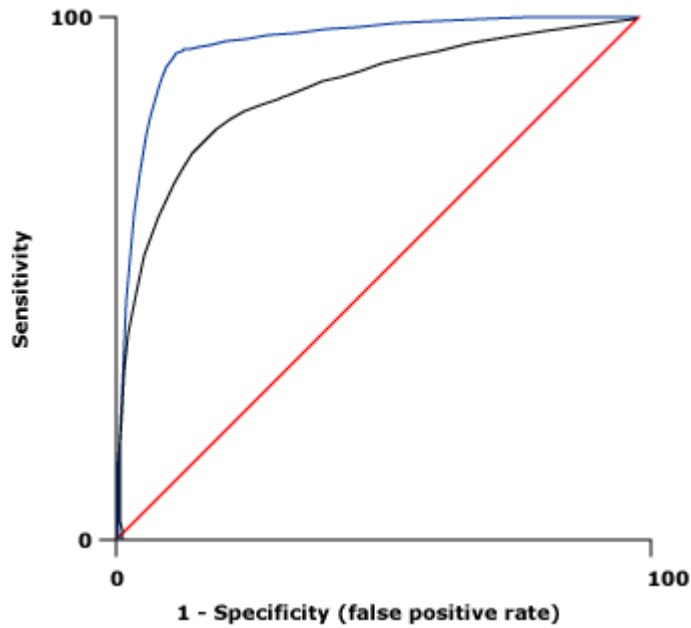
Graphic 50171 Version 1.0

Population stratification



Graphic 78339 Version 1.0

Receiver operating characteristic curve



The receiver operating characteristic curve (ROC) plots sensitivity against 1- specificity (ie, the false positive rate). Cutoff values for the diagnostic test vary along the length of the black curve, reflecting the interchange between sensitivity and specificity. The area under the ROC curve represents the accuracy of the test. A test that performs no better than chance would be represented by a straight line (red line) with an area under the ROC of 0.5. A near perfect test would have a rectangular configuration (blue line) with an area under the ROC approaching 1.00.

Graphic 82175 Version 1.0

Disclosures

Disclosures: **John Attia, MD, PhD, FRCPC, FRACP** Nothing to disclose. **Gordon Guyatt, MD** Nothing to disclose. **Benjamin A Raby, MD, MPH** Other Financial Interest (Spouse): Parexel International [Contract research organization (Double-blind randomized clinical trials)]. **Jennifer S Tirnauer, MD** Nothing to disclose.

Contributor disclosures are reviewed for conflicts of interest by the editorial group. When found, these are addressed by vetting through a multi-level review process, and through requirements for references to be provided to support the content.

Appropriately referenced content is required of all authors and must conform to UpToDate standards of evidence.

Conflict of interest policy