

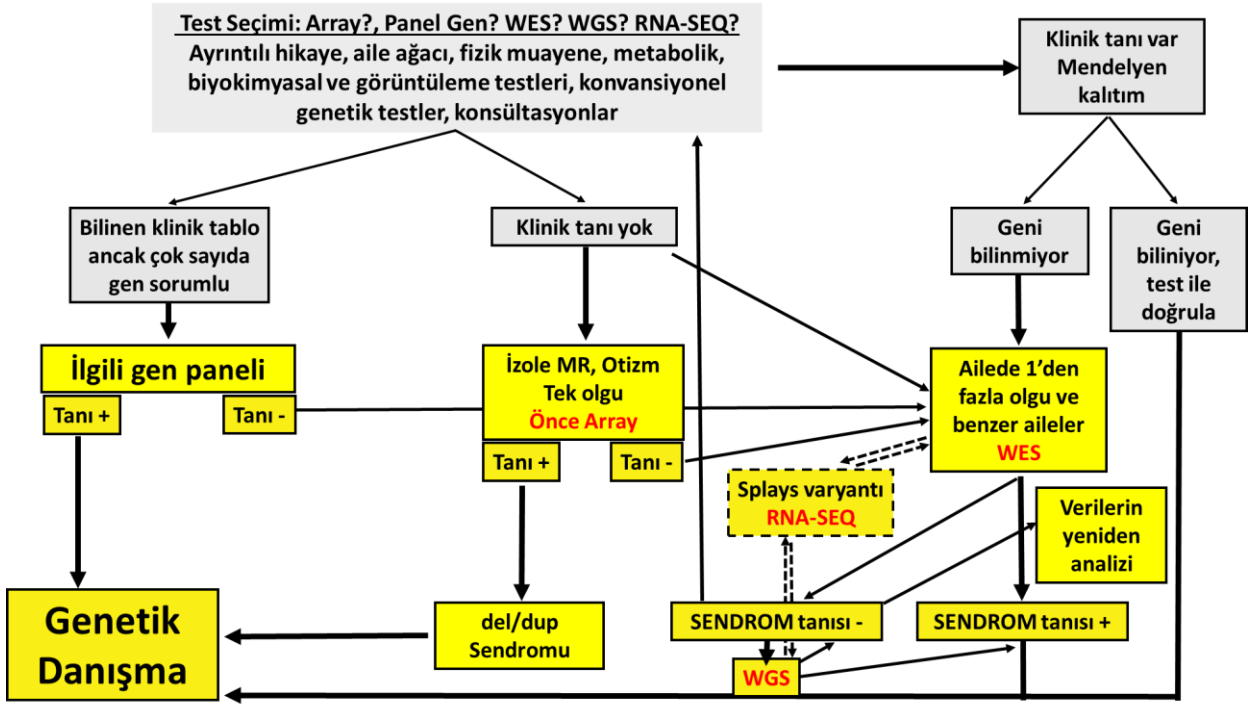
YENİ NESİL DİZİ ANALİZİNDE ALGORİTMALAR ÇALIŞMA GRUBU
YENİ NESİL DİZİ ANALİZİNDE ALGORİTMALAR
ÖZET

Yeni nesil dizilemede algoritmalar çalışma grubu aşağıdaki konularda görüş bildirmektedir:

- 1- Yeni nesil dizileme test yönetimi (test öncesi danışma, testin istenmesi, yorumlanması, test sonrası danışma) genetik uzmanlarınca yapılmalıdır. Böylece hem halen pahalı olan bu teknolojilerin gereksiz kullanımının önüne geçilebilir hem de teknolojinin ve/veya genetik heterojenitenin doğası gereği eksik olduğu kısımlar için diğer genetik testlerin kullanılarak hastaya tanı koyma imkanı ortaya konabilir.
- 2- Yeni nesil dizileme test istemi öncesi ve sonrasında, test uygulanacak kişiye genetik danışma verilmelidir. Bu danışmada testin ortaya koyabileceği bulgular, testin sınırlılıkları, ek teste ihtiyaç olup olmayacağı, ikincil bulguların tespit edilmesi durumundaki stratejiler ve test sonucunda testin endikasyonuna yönelik tanısall amaçlı bir varyant bulunamaz ise raporlamanın nasıl olacağı, elde edilen ham verinin paylaşımı, eğer gerekli görülürse belirli bir periyod sonrasında (ideali 1-2 yıl olabilir) verinin yeniden analizinin istenip istenmediği konuları açığa kavuşturulmalıdır.
- 3- Yeni nesil dizileme testlerinin rutin klinik tanıda kullanılabilmesi için hem laboratuvar alanı, hem minimum ihtiyaç duyulan cihazlar ve hem de biyoinformatik uzmanı gibi ihtiyaç duyulan personel listesi Tıbbi Genetik Derneği'nce belirlenerek Genetik Hastalık Tanı Merkezi Yönetmeliğine ilave edilmelidir.
- 4- Hastalara ikinci sayfadaki şekilde verilen algoritma uygulanarak öncelikle fenotipleme sonra hastanın fenotipine göre konvansiyonel yöntemler ve en son yeni nesil dizileme teknolojileri kullanılmalı ve bu teknolojinin panel, ekzom veya genom düzeyinde yapılmasına karar verilmelidir. Bu kararı verecek kişi yine Tıbbi Genetik uzmanı olmalıdır.
- 5- Yeni nesil dizileme teknolojisini uygulamak isteyen laboratuvarlar için temel bilgiler içerikte daha detaylı verilmiştir. Her laboratuvar kendi ihtiyacına göre yakalama ve dizileme platformunu seçmelidir.
- 6- Yeni nesil dizileme uygulanacak her bir materyal için, laboratuvarlar nükleik asit izolasyonu, kullanılacak platformlar ve biyoinformatik araçları valide ettikten sonra tanısall amaçlı rutin kullanıma sokmalıdır.
- 7- Multipleksleme kullanılacak ise valide olmuş indekslerin kullanılması önerilmektedir.
- 8- Biyoinformatik analizlerin her bir basamağından (FASTQ, BAM ve VCF dosyalarının üretimi) sonra verilerin kalite değerlendirmeleri yapılmalı ve yeterli kaliteye ulaşmayan örnekler sonraki değerlendirme basamaklarına geçmemelidir.
- 9- Varyantların önceliklenmesinde temel olarak, sinonim varyantların filtrelenmesi, popülasyon bilgi bankalarında minör alel sıklığının %1'in üstünde (hastalık insidansı ile korele olarak) olan varyantların filtrelenmesi ve in silico programlardan faydalanılarak önceliklendirilen hastalık ile ilişkili olabilecek varyantların Sanger dizileme ile segregasyonunun analizi ve segregasyon olan varyantların OMIM, Pubmed, COSMIC, ClinVar v.b. bilgi bankalarında rapor edilip edilmediğinin tespitinin yapılması basamaklarını içerir. Varyantların önceliklendirilmesi sürecinde hiçbir zaman in silico tahminlere göre varyantın patojenliği hakkında karar verilmemelidir. Veri analizinde fenotipleme ve pedigr bilgilerine göre varyant önceliklendirilmesi yapılabilir. Örneğin resesif kalıtmadan şüpheleniliyor ise öncelikle homozigot ve birleşik heterozigot varyantların değerlendirilmesi yapılabilir.
- 10- Yeni nesil dizileme teknolojilerine ait rapor oluşturulurken, moleküler genetik raporu formatı temel olarak aynı şekilde kalırken testin kapsamına göre (tek gen, multigen panel, ekzom, genom v.b.) modifikasyonlar yapılmalıdır. Çalışma grubumuz tüm ekzom dizileme ve tüm genom dizileme raporlamasında en azından testin yapıldığı toplumda sık görülen hastalıklar için hastanın taşıyıcılık durumunun belirtilmesinin,

ve nadir hastalıklardan sorumlu olan gen mutasyonu varlığında raporla hastanın hekiminin uyarılmasının gerekli olduğunu düşünmektedir. Detaylar ilgili bölümde bulunabilir.

- 11- RNA dizileme tüm genom dizilemenin daha fazla günlük kullanıma girmesiyle, varyantların etkilerinin değerlendirilmesi noktasında daha fazla yer bulacaktır. Rutin kullanımdaki yeri hakkında ilgili bölümde detaylı bilgi bulunabilir.
- 12- Yeni nesil dizileme ile birlikte üretilen büyük verilerin, saklanması, aktarılması ve analizleri ile ilgili ulusal bir politika belirlenmelidir.
- 13- Yeni nesil dizileme teknolojileri etik konuları da gündeme getirmektedir. Teste özgü onam formu alınmalı, test öncesi ve sonrası genetik danışma verilmeli, ikincil bulgular ve verilerin yeniden analizleri ile ilgili politikalar belirlenmelidir.



Yeni Nesil Dizileme Testlerinin Kullanımında Önerilen Algoritma

İÇERİK

Başlık	Sayfa No:
Özet	1-2
İçerik	3
1. Yeni Nesil Dizileme: Giriş	4-7
2. Fenotipleme	8-16
3. Genotipleme	17-20
4. Islak Laboratuvar (WET-LAB)	20-22
5. Kuru Laboratuvar (DRY-LAB)	22-51
5.1. Primer Analiz (Baz çağrılması ve demultiplexing)	25-30
5.2. Sekonder Analiz (Haritalama)	30-44
5.3. Tersiyer Analiz (Varyantların Annote Edilmesi)	45-50
5.4. Teknik Zorluklar	50-51
6. Verilerin İnterpretasyonu/Yorumlanması	52-55
7. Yeni Nesil Dizileme Testlerinde Raporlama	55-58
8. RNA Dizilemenin rutin tanıdaki yeri	58-61
9. Diğer tip dizilemeler ve yeni nesil dizilemenin diğer kullanım alanları	61
10. Veri saklama ve hasta raporlarının takip edilebilirliği	61-62
11. Ek Bilgiler	63-64
12. Referanslar	65-74
13. Şekil, Tablo ve Bilgi Kutuları Sayfa No	75

GİRİŞ

Yeni Nesil DNA dizi analizinde algoritmaların oluşturulması için Ahmet Okay Çağlayan'ın (MD) moderatörlüğünde biraraya gelen çalışma grubunda biyoinformatik uzmanları Mahmut Şamil Sağıroğlu (PhD), Tunahan Çakır (PhD), Saliha Durmuş (PhD), klinik ve araştırma laboratuvarı yöneticileri ile rutin hasta yönetiminde YND kullanma tecrübesine sahip sorumlu uzmanlar, Beyhan Tüysüz (MD), Zehra Oya Uyguner (PhD), Ahmet Okay Çağlayan (MD), Ayça Aykut (MD, PhD), Kaya Bilguvar (MD, PhD), Yavuz Oktay (PhD) ve Zafer Yüksel (MD) yer aldı.

Yeni nesil dizileme (YND) alanı hızla gelişmekte olduğundan, yeni biyoinformatik algoritmalar ve teknolojiler ortaya çıktıkça burada sunulan bazı bilgilerin de revizyona ihtiyacı olacaktır. Bu kılavuz hazırlanırken daha önce yayımlanmış kılavuzlar esas alınmış, bazı kısımlarda birebir çevirisi yapılmıştır. Ayrıca belirli teknolojiler detaylı şekilde ele alınmamış olup, her bir test uygulamasına ve platforma özgün karakteristiklerin ayrıca değerlendirilmesi gerekebilir. Örneğin somatik varyantların ya da dolaşımdaki serbest fetal DNA'nın tespiti gibi farklı tanı gereksinimlerine yaklaşım için ek bilgilere ihtiyaç duyulabilir. Bu rehber ve öneriler, YND'yi rutin genetik tanıda uygulamak isteyen merkezlere genel bilgi sağlamak amacıyla oluşturulmuştur. Pek çok klinik laboratuvar, klinik genetik biyoinformatik uzmanına sahip değildir ve temel bir rehber ihtiyacı duyacaktır.

YND ile ilgili olan ve bilinmesi gereken bazı terimler aşağıda, Illumina tüm genom ve tüm ekzom analizlerinde sıkça kullanılan yaklaşımlarla uyumlu terminolojiler üzerinden örneklendirilerek anlatılmaktadır:

1. **Insert (Fragment)**—Dizileme için kullanılan DNA fragmanı
2. **Read (Okuma)** Dizilenen *insert*'in bir bölümü veya tamamı
3. **Single Read (Tekli Okuma)** —*Insert*'i sadece tek bir uçtan dizilemek için kullanılan dizileme prosedürü
4. **Paired-End (Eşli Uç)** —*Insert*'i her iki uçtan birden dizilemek için kullanılan dizileme prosedürü
5. **Flowcell**—Üzerinde DNA fragmanlarının koyulup dizildiği küçük bir cam parçası. *Flowcell*, DNA fragmanlarına bağlanmış olan adaptörlerin hibridizasyonuna olanak tanıyan problemlerle kaplıdır.
6. **Lane**—*Flowcell*, adına *lane* denilen fiziksel olarak birbirinden ayrılmış kanallardan meydana gelir. Dizileme işlemi, tüm *lane*'ler üzerinde eş zamanlı olarak yapılır.
7. **Multiplexing/De-multiplexing**—Aynı lane üzerinde birkaç tane örneğin dizilmesine *multiplexing* denir. Tek bir *lane* üzerinde dizilenen okumaların ayrılmasına da *de-multiplexing* denir ve her bir okumanın dizinini tanıyan ve bunu her bir örneğin bilinen dizinleriyle karşılaştıran bir indeks ile yapılır.
8. **Pipeline**—Algoritması olan ardışık bilgisayar işlemleri.
9. **Sequence Coverage (Dizi Kapsamı)**— Kapsam (*coverage*), genomda ortalama her bir bazı kapsayan okuma sayısına karşılık gelir. Kapsam şu şekilde hesaplanabilir:

$$\text{Ortalama kapsam} = \frac{\text{okuma uzunluğu} \times \text{okuma sayısı}}{\text{genom büyüklüğü}}$$

Yukarıdaki hesaplama sadece haritalanan okuma sayısı dâhil edilir. Genom varyantlarını belirlemek için 30 kapsam genelde minimum kabul edilirken, *de novo* birleştirme sıklıkla daha yüksek bir kapsamı gerektirir. Bunun yanı sıra ihtiyaç

duyulan kapsam, deney tasarımına da bağlıdır. Örneğin eğer tekrar dizileme bir popülasyonda yapılıyorsa ve örnek heterojen bir genom havuzunu içeriyorsa, kapsam nadir varyantları sağlıklı tespit etmek için kapsamı daha yüksek tutmak gerekebilir²³.

1. Yeni Nesil Dizileme: Giriş

1.1. Nedir?

Yeni nesil dizileme (YND), birbirine paralel milyonlarca kısa DNA fragman dizisini tespit etme yöntemidir. YND ve özellikle tüm ekzom dizileme (TED), uygun maliyetli ve yüksek çıktılı bir teknoloji olup, insan genomunda bilinen 20.000'den fazla genin kodlama bölgelerini eşzamanlı analiz etme olanağı tanır^{24; 25}. Günümüzde TED, insan genomunun protein kodlayan kısmını incelemek için en yaygın kullanılan teknolojik yaklaşımdır. TED sadece insan genomunun %1-1,5'ini²⁶ kapsamakla birlikte, bu küçük genom kısmı bile hastalığa neden olan bilinen mutasyonların yaklaşık %85'ini barındırmaktadır. Ancak olguların % 25-40'ında hastalıkların genetik kökenini aydınlatılmaktadır^{28; 29}. Bu oran, karyotip ve kromozal mikrodizi (%15-20) gibi daha klasik yöntemlerle elde edilenden yüksektir^{30; 31}.

YND, klinik ortamda yaygın şekilde kullanılmaktadır. Çok sayıda laboratuvar "tanısal belirsizliği" azaltmak amacıyla klinik TED veya diğer YND tabanlı testleri uygulamakta veya hizmet satın almak suretiyle kullanmaktadır.³³ Yakın tarihli yayınlar, ilaç seçimi ve dozlamaya ilgili kararlarda yardımcı olması amacıyla, nadir görülen hastalıkların ve kanserin teşhisinde YND'nin kullanımını örneklerle göstermektedir^{34; 35}. Klinik laboratuvarlar YND'yi aynı zamanda insan lökosit antijeni tiplemesi, farmakogenetik, bulaşıcı ve kronik hastalık testi için de kullanmaktadır³⁶⁻⁴¹. Örneğin nadir görülen, klinik ve genetik açıdan heterojen olan ve belirsiz veya atipik bir görünüm sergileyen hastalıklardan etkilenmiş hastalara tanı koymak zordur^{42; 43}. Bir tanı konulmadan önce hastalar, psikososyal olarak yıpratıcı, zaman alan ve maliyetli olabilen kapsamlı değerlendirmelerden geçmek zorundadır³⁰. YND teknolojisi bir bireyin DNA'sında bir hastalıkla potansiyel şekilde bağlantılı olabilecek varyantları tespit ettiği kadar henüz anlamlandırılmayan bir çok başka varyantın da gösterilmesine neden olur. Teknik bazı sınırlamalar hastalara tanı konma sürecini etkileyebilir. Örneğin testin hassasiyeti (hedef dizilerin kapsamı) %100 değildir ve ilişkili herhangi bir genin tamamının veya bir kısmının kapsamaması nedeniyle yanlış-negatif sonuçlara ulaşılabilir²⁵. Aynı şekilde dizi derinliğinin yeterli olmaması da tespit edilen varyantların güvenilirliğini değiştirmektedir. TED, tanı için güçlü bir tanısal araç olmakla birlikte, tüm klinik endikasyonlar için en iyi yaklaşım olmadığını bilmek gerekir ve klinik bulgular ve ortaya çıkan fenotip varyantların tespit edilmesi için gerekli ilişkilerin kurulmasında en önemli basamaktır⁷.

Sonuç olarak YND teknolojisinin klinik laboratuvar ortamında rutin olarak kullanımı için önemli bir teknolojik altyapı ile birlikte klinik, bilimsel ve biyoinformatikte özgünleşmiş deneyim ve derin tecrübeleri olan bireylere ihtiyaç duyulur.

1.2. YND'nin Sanger Dizilemeden Farkı Nedir?

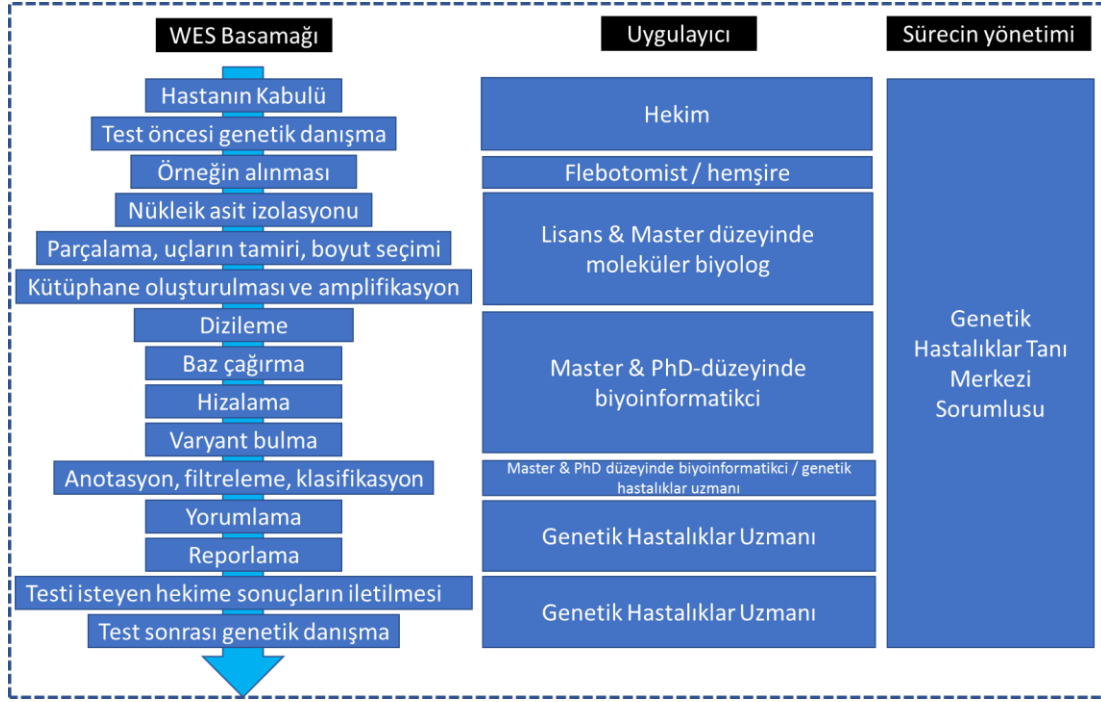
YND, hem yöntem hem de elde edilen bulguların tarifi açısından Sanger dizilemeden temelde farklılık gösterir⁴⁹. Kapiller elektroforez tabanlı yarı otomatik Sanger dizileme⁵⁰⁻⁵² günümüzde konsensus olarak DNA dizileme için altın standart olarak kabul edilmektedir. Ancak son zamanlarda bu yöntemin artık altın standart olmayabileceği ile ilgili yayınlar da bulunmaktadır⁴⁹. Sanger dizileme, yüksek güvenilir sonuçlarla birlikte uzun okuma uzunlukları üretir, ancak daha uzun genom alanları hedeflendiğinde maliyeti çok yüksek olur ve işlemler zaman alıcı olabilir. Bu nedenle Sanger dizi analizi, büyük genlerin veya çok sayıda genin incelenmesi gereken durumlarda rutin klinik hizmet için pratik bir yaklaşım sunamaz. YND ise gen panellerinin daha düşük bir maliyetle dizilenmesine olanak tanır ve hızlı, geniş ölçekli ekzom ve tüm genom dizi analizi gerçekleştirme kabiliyeti sunar⁵⁴⁻⁵⁶.

YND, Sanger yöntemine göre her ne kadar bazı avantajlar sunsa da şu anda mevcut olan enstrümanların büyük bir çoğunluğu kısa okuma uzunlukları üretir ve bunlardan raporlanabilir dizi analizi sonuçları elde edebilmek (YND ile üretilen verilerin yüksek miktarı; verilerin analizi, yönetimi ve saklanması) için biyoinformatik yazılımların kullanılarak komplike hizalama veya birleştirme prosedürlerinin uygulanması gerekir⁵⁷. YND tabanlı test kullanmayı düşünen klinik laboratuvarlar, verilerin yorumlanmasındaki zorlukları, maliyeti ve hızı göz önünde bulundurmalıdır. YND testleri, ham veri dosyalarını bir referans dizisi ile doğru şekilde hizalayabilen, dizi varyantları için anotasyonlar ekleyebilen, hangi varyantların klinik anlamlılığa sahip olduğunu tayin edebilen ve hangi varyantların doğrulayıcı teste ihtiyacı olduğunu belirleyebilen informatik düzen gerektirir. Bir hasta örneğinde tespit edilen varyantların sayısı, hedeflenen genom bölgesinin büyüklüğüyle orantılıdır ve varyant sınıflamasının altından kalkılamayacak sayıda veri üretebilmektedir. Girdi olarak varyant dosyaları kullanılarak klinik karar desteği için *downstream pipeline* ve algoritmalar geliştirilebilir^{48; 58} ve bunlar ClinSeq⁵⁹ gibi projelerde test edilmektedirler, ancak bu analizler halen otomatizasyondan uzak olup tüm basamakların sonunda deneyimli bir genetikçi tarafından verilerin değerlendirilmesine ve klinik raporların bu şekilde oluşturulmasına ihtiyaç olunur. **Çalışma grubumuz tüm basamakların sonunda deneyimli bir genetikçi tarafından verilerin değerlendirilmesi ve klinik raporların ancak bu değerlendirmeden sonra klinik genetikçi ve gerektiğinde hastanın klinisyeni ile konsulte edilerek oluşturulmasını önermektedir.**

1.3. Önerilen teste özgü fiziksel alan, cihaz ve personel ihtiyacı listesi

YND testlerini uygulamak isteyen laboratuvarların hem personel hem de fiziksel alan açısından minimum şartları sağlamış olması gerekmektedir. Bu şartlar tavsiye niteliğinde olup, nihai karara konudaki uzmanların ortak fikir birliği ile ulaşılmalıdır.

YND'nin teknik ve yorumlama karmaşıklığını dikkate aldığımızda, klinik YND tabanlı testlerin uygulama, değerlendirme, yorumlama ve raporlandırma aşamalarında konuda uzmanlığını eğitim ve belge ile kanıtlayan (Uluslararası ve ulusal genetik dernekleri; Türkiye'de Tıbbi Genetik Derneği tarafından onaylı olması gibi) uzman kişiler/kadrolar tarafından yapılmasını tavsiye ederiz. TED veya TGD hizmeti sunan laboratuvarların, genler, varyasyon ve hastalık fenotipleri arasındaki ilişkileri değerlendirmek için bu konuda deneyimi olan klinik genetik uzmanlarına ihtiyaçları vardır (**Şekil 1**).



Şekil 1. Yeni nesil dizileme basamakları ve personel görev listesi. Hum Genet (2016) 135:655–673'den⁵ değiştirilerek alınmıştır.

1.4. Testin ortalama sonuçlanma süresi

Klinik bir YND testinin yapmak için gereken zaman büyük oranda değişiklik gösterir. Süre, araştırılan genomun (panel, ekzom veya genom gibi) kapsamına, otomasyon seviyesine, kütüphane hazırlama için kullanılan prosedürlere, dizileme yöntemine, hizalama, varyant çağırma için kullanılan biyoinformatik işlem akışına, filtrelemeye ve klinik değerlendirmeye bağlıdır. Deneyimli genom varyant analiz uzmanlarının sayısı ve laboratuvarın bilgisayar kapasitesi de bir etmendir. Klinik değerlendirme, sürecin en çok zaman alan bileşeni olabilir. Çoğu hasta literatürde hastalıkla ilişkili olarak önemi iyi tanımlanmamış birtakım varyantlar içerir, ancak bu varyantlar hastanın fenotipiyle ilişkili olmayabilir. Bu nedenle bu varyantların klinik değerlendirmesi için gereken süre, sınıflandırma için manuel göz gezdirme gerektiren her bir varyant için yaklaşık 30 dakika ile birkaç saat arasında değişir. Kalıtsal durumlar için fenotip bilgilerinin toplanması ve aile bireylerinin ilave testten geçirilmesi bazen gerekebilir. 2017 yılı itibarıyla bir YND testini tüm çıktıları analiz edildikten sonra raporlamak için gereken süre hastalık, gen ve kalıtım modeli ile uyumlu bilinen mutasyon saptanırsa, birkaç saatten birkaç haftaya kadar değişiklik gösterebilir. Bilinen hastalıklarda bilinen genlerde yeni varyantlar kalıtım modeli ve varyantın tipine göre çoğu zaman aile çalışması da gerektireceğinden çok daha uzun bir zaman alabilir. Ayrıca düşük kapsamlı bölgelerde patojenik varyant şüphesi varlığında Sanger dizileme ile doğrulanması/dışlanması gerekir. Ek olarak hastalık ilişkili bir varyant saptandığında ailenin diğer fertlerinde de test yapılması gerekirse bu aşama için Sanger dizi testi için tasarım ve uygulama yapılması gerekir. Yeni fenotipler ya da bilinen hastalıklarda herhangi bir mutasyon saptanmadığı durumlarda ise yeni ilişkili gen araştırılması gerekir böyle durumlarda testin sonuçlanması aylar hatta yıllar dahi sürebilir. Bu durumda ailelere ara rapor vermek ve bilinen genlerde olası patojenik varyantların dışlandığını (limitasyonlarını da göz önüne alarak) bildiren bir rapor sunulması

uygundur.

2. Fenotipleme

YND verilerinin yorumlanmasında en kritik nokta, olmaz ise olmazı klinik yönlendirmedir. Klinik tanısı/bulgusu olmayan bir olguda analiz yapılmaz ve test için numune kabul edilmez. Bu konuda bir kolaylaştırıcı bir yaklaşım olarak Human Phenotype Ontology Projesi⁶⁰ ve mevcut terminoloji atlası [<http://compbio.charite.de/hpoweb/showterm?id=HP:0000001> (Son erişim tarihi 21.05.2017)] kullanılarak standart fenotip veri girişi sağlanacak biçimde tanı testi istekleri oluşturulabilir. Bu konuda ulusal bir kayıt sistemi oluşturulabilirse (Tıbbi Genetik Derneği internet sitesi gibi) klinik kalıplar ve terminoloji için ülkemize özgün bir veritabanı da oluşmuş olur. Böylelikle klinik tanı sürecinde ve YND test istemi öncesinde yararlı bir araç [Phenomizer⁶⁰ benzeri] gelişiminin de önü açılmış olur.

2.1. Hasta Seçimi

2.1.1. Klinik ekzom ve genom dizilemeyi (KEGD) kimler istemeli?

Hastalık tanısını koymak amacıyla genetik testi isteyen hekimlerin rolü, klinik özelliklere, aile öyküsüne ve fiziksel muayeneye dayanarak bir ayırıcı tanı koymaktır. Genetik test(ler) için istek, bir tanıyı teyit etmek veya dışlamak için verilebilir. Genetik testlerin birçok çeşidi vardır ancak burada tanısız amaçlı testlerden bahsedilmektedir. Öneriler tavsiye niteliğinde olup, hedefi akılcı test kullanımında YND'nin yerini belirlemektir.

Amerikan Tıbbi Genetik ve Genom Koleji (ACMG), KEGD için genetik uzmanlarına danışılmasını ve yeterli genetik danışmanlık alınmasını tavsiye etmektedir⁶¹. Ancak halihazırda genetik alanındaki uzman sayısının azlığı nedeniyle klinik uzmanlarından danışmanlık istenmesi de bazı şartların sağlanması suretiyle önerilmiştir ¹. **Ancak çalışma grubumuz YND testinin sadece genetik hastalıkları uzmanlarınca istenebilmesini önermektedir.** Test öncesi aşağıda belirtilen aşamaların mutlaka yapılabilir olması gerekir:

1. Genetik odaklı tıbbi bilgi, detaylı aile öyküsü ve bu öyküye dayalı iyice sorgulanarak çizilmiş aile ağacı fiziksel muayeneyi içine alan temel klinik genetik muayene, biyokimyasal ve çeşitli görüntüleme tetkiklerinin sonuçları, gerekirse ek konsültasyonlar yapılmalı, düşünülen hastalık, kalıtım modeli genetik ilişkisini anlatan bir epikriz yazılmalıdır.
2. Diğer mevcut testler (genomik ve metabolik tarama gibi genom dışı testler), hastanın yaşını ve bilişsel yeteneğini de göz önünde bulundurarak KEGD'nin spesifik klinik endikasyon için gereken test seçeneği olup olmadığı belirlenmelidir.
3. Birincil ve ikincil bulgular için bilgilendirilmiş onam dâhil test öncesi yeterli danışmanlık sağlanmalıdır.
4. KEGD'nin sonuçlarını yorumlamak ve test sonrası yeterli danışmanlık sunmak veya bu adım için bir genetik uzmanından yardım istemek gerekir. Bilgiler mutlaka arşivlenmeli, olgu ve ailesi düzenli takibe alınmalıdır.

2.1.2. KEG endikasyonları nelerdir? (ACMG önerisi¹)

1. Bir hastanın fenotipine göre tahmin edilen genetik durum için daha optimize genetik bir panelin mevcut olduğu biliniyorsa, hedeflenen gen panelinin daha hassas olduğu varsayılarak bu gen paneline öncelik verilmelidir (örneğin Noonan spektrum bozuklukları gibi)
2. Fenotip, bilinen bir genetik duruma benzemediğinde (MKA/MR) ve aile öyküsü bilinen bir Mendel modeline işaret etmediğinde, kromozomal mikroarray testi KEGD'den önce göz önünde bulundurulmalıdır. Bu yaklaşım, KEGD'nin kopya sayısı varyantlarını tespit hassasiyeti, kromozomal mikroarray platformlarının hassasiyetine eşit olana kadar önerilecektir. Fakat belirtmek gerekir ki tüm genom dizilemenin kopya sayısı varyasyonlarını, kromozomal mikroarray ile aynı hassasiyette tespit etme kabiliyeti dikkate alındığında bu yaklaşım, genom dizileme (GD) platformlarının rutin klinikte çok daha fazla kullanıma başlamasına güncelliğini yitirebilecektir.
3. Anne-baba arasında akrabalık bulunan hastalarda KEGD, kromozom mikrodizilemeden önce göz önünde bulundurulmalıdır. **Çalışma grubu bu öneriyi ülkemizdeki akraba evliliği sıklığını gözönünde bulundurarak desteklememekte öncelikli olarak mikrodizi analizini önermektedir.**
4. Hedef gen paneli tanısal bir sonuç vermediğinde KEGD, zorunlu bir sonraki adım olarak düşünülmemelidir. Bununla birlikte bilgisayar ortamındaki KEGD analizi, ilk başta belirli bir gen alt kümesine odaklanabilir ve sonrasında eğer ilk analizle tanı konulamaz ise analiz tüm veri setine kadar genişletilebilir (örneğin, <http://www.genedx.com/test-catalog/xomedxslice/>). KEGD maliyetleri düştükçe ve tanı laboratuvarları kendi iş akışlarını modifiye/modernize edip, "tek platformlu" genetik testler yaptıkça bu yaklaşım daha yaygın hale gelebilir. Bu gerçekleşene kadar ise klinikte uzmanlar ancak şu koşullarda tanının konulmadığı bir gen panelinden sonra KEGD yapılmasını isteyebilir:
 - (i) Genetik bir neden için güçlü şüphe mevcutsa ve uygulanan gen panelinin içeriğindeki genler güncel değilse,
 - (ii) Hastanın fenotipi değişmiş ise veya negatif panel için atipik ise ya da
 - (iii) Risk altında olabilecek aile bireyleri için tanı ve/veya derhal klinik kullanıma ilişkin klinik bir acil durum söz konusu ise.

2.1.3. Hasta ve ailelere test öncesi danışma

Bir KEGD testi istemeden önce hastalar ve aileler için özetle aşağıdaki konularda temel bilgilerin verilmesi önerilmektedir (Tablo 1.)¹

Eğitim Konusu	Anahtar Elementler
Temel Bilgiler	Temel genetik bilgiler
	KEGD hakkında temel bilgilerin verilmesi
	Testin fiyatı ve sigorta tarafından ödemesinin yapılıp yapılmadığının tespiti
Spesifik Klinik Genom / Ekzom Dizileme Bilgisi	Klinik genom/ekzom dizilemenin daha önce uygulanmış genetik testlerden farklarının anlatılması
	Kime test uygulanacak (örneğin, trio, sadece proband)
	Ortalama sonuç verme süresi
Psikososyal Etkiler	Ne tür sonuçların çıkabileceğinin tartışılması (örneğin, primer ve sekonder bulgular v.s.)
	Belirsizlik potansiyeli (örneğin, önemi bilinmeyen varyantlar)
	Kimin sonuçları raporlanacak (proband veya trio)
Test sonrası konular	Birey, çekirdek aile, ve genişletilmiş aile üyeleri
	Diğer aile üyelerinin test edilmelerinin gerekli olunabileceği
	Elde edilen sonuçlara göre kişilerin sigorta konusunda yaşayabilecekleri sıkıntılar
	Örneklerin saklanması (hangi koşullarda ne kadar süre v.b.)
	NGS verilerinin saklanması, yeniden analiz, ve tekrar kondağa geçilmesi
	Araştırma fırsatları

Tablo 1. KEGD yapmadan önce danışmada verilmesi gereken bilgiler. Genet Med. (2016) 18:1075-1084'den¹ değiştirilerek alınmıştır

Hastalar: KEGD'ye ilişkin hasta eğitimi, hedef kitle açısından özellikli bir biçimde tasarlanmalı ve bilgilendirilmiş onam belgesindeki yazılı bilgilerin ötesine geçmelidir. Bazı yayınlanmış makaleler ve rehberler, eğitimi mümkün olan sonuç türlerinin üzerine odaklamanın en önemli konu olduğu sonucuna varmıştır.

Hizmet Sağlayıcılar. KEGD testi istemeden önce klinik uzmanları kendi kabiliyet ve eğitim ihtiyaçlarını değerlendirmelidir. Ana başlıklar arasında şunlar yer almalıdır: temel klinik genetik bir değerlendirme ile aile öyküsü neyi içermelidir, en az üç kuşak pedigrî çizimi (doğum yerlerini ve bu yerlerin birbirlerine yakınlıklarını belirtir şekilde), önceki değerlendirmelerin kritik bir özetini vermek için genetik testlerin farklı türlerinin gözden geçirilmesi, klinik uzmanların bu testin spesifik klinik endikasyon için en iyi test olup olmadığını belirlemelerine yardımcı olmak için KEGD'ye genel bir bakış (endikasyonlar, güçlü yönler ve sınırlamalar), test öncesi danışmanlık neyi içerir, sonuçların yorumlanması, test sonrası danışmanlık ve genetik bilim uzmanından desteğin ne zaman istenmesi gerektiği.

Sigorta ve devlet paydaşları: Sigorta sağlayıcılar ve devlet kurumlarının, testin kapsamı ve masrafların ödenmesi konularında bilgiye dayalı kararlar alması için KEGD'nin anlaşılması zorunludur. Ana eğitim başlıkları arasında klinik geçerlilik bulunur. Bu basamak için Tıbbi Genetik Derneğinin görevlendirdiği kişiler NGS ile ilgili bilgileri ilgili kişilere anlatabilir veya Tıbbi Genetik Derneği yazılı rapor olarak direkt ilgili kişilere sunabilir.

Eğitim nasıl verilir? Klinik uzman-hasta etkileşimi, yeterli hasta eğitimi için kilit görev görür, bununla birlikte hasta eğitimi materyalleri, hasta-klinik uzman etkileşiminin tamamlayıcısı olarak verilmelidir. Hastaların KEGD hakkında eğitimi için sayısız online kaynak mevcuttur ve bunlar ücretsizdir. Bu kaynakların formatları arasında video, interaktif modüller ve farklı eğitim ihtiyaçlarını karşılamak için tasarlanan uzman anlatımlı test açıklamaları yer alır. Bunun yanında kaynakların doğru olarak belirlenmesine ihtiyaç vardır, zira bir çok bilgi kirliliği yaratan ve doğru olmayan bilgileri içeren online web siteleri de bulunmaktadır.

Genetik bilimciler şu anda bir KEGD testi istemek, uygulamak ve yorumlamak için gerekli olan pek çok alanda son derece yetkindirler, ancak yine de KEGD'nin klinik yorumu için gereken belirli becerilerin edinilmesi, hasta bakımı açısından kritik bir unsur olarak durmaktadır. Uzmanlığa özgü eğitim fırsatlarının yaratılması teşvik edilmelidir, örneğin KEGD ile ilgili olduğu için varyant yorumlama üzerine odaklanan Tıbbi Genetik Derneği'nin finanse ettiği konferanslar gibi. Diğer kısa kurslar ve sürekli tıbbi eğitim (CME) modülleri Avrupa, Amerika ve dünya çapında mevcuttur. Detaylar ise meslek topluluklarının web sitelerinde bulunabilir ve bunlar arasında Amerikan İnsan Genetik Derneği (ASHG) (http://www.ashg.org/education/Health_Professionals.shtml), Amerikan Tıbbi Genetik Koleji (<http://www.acmg.net/ACMG/Education/>), Avrupa İnsan Genetiği Derneği (<http://www.eshg.org/courses.0.html>) yer alır.

2.2. Test Seçimi

Klinik değerlendirme sonrası mutasyonların sık rastlandığı küçük-orta boyda tek bir gen dizilemesi (<20 ekzon; en az 1-2 ekzon büyüklüğü ~1200 bç) gerektiriyorsa Sanger dizileme yaklaşımı kullanılabilir. Birden fazla ilişkili büyük genler için ise YND tabanlı testler endikedir. Fenotip ne kadar belirli tek bir hastalığa ya da hastalık grubuna indirgenebiliyorsa kapsama durumu gözetilerek ilk olarak gen panellerinden daha sonra da tüm ekzom ve tüm genom dizilemesine doğru bir yol gözetilmelidir (Şekil 2).

Kimi durumlarda maliyet/etkinlik durumu ya da hasta/hastalık bazında düşünülerek karar verilebilmektedir. Testin tanı koymak ya da dışlamak amacıyla istenip istenmediği belirlenmelidir. Bazı laboratuvarlar, çok sayıda Sanger diziyi (PCR ve saflaştırma sonrası) YND teknolojisi ile yürütebilir, bu yöntem optimize edilmesi durumunda laboratuara özgün bir yaklaşım olarak kullanılabilir.

YND, DNA [multi gen panellerinin hedefli tekrar dizilenmesini, tüm ekzom dizileme (TED) ve tüm genom dizilemeyi (TGD)] ve RNA kaynaklı testleri içeren çeşitli seviyelerde kullanılabilir (Tablo 2). Herbirinin kendine göre avantajları ve dezavantajları vardır (Tablo 2, Tablo 3).

	Tüm genler, tüm ekzonlar, tüm kodlamayan bölgeler	Keşif	Fiyat; Derinlik; Düşük allel fraksiyonu için düşük sensitivite
	Tüm genler, tüm ekzonlar	Klinik araştırma; Panel-negatif tanısal test olarak; neo-epitop tahmini	Fiyat; Derinlik; Düşük allel fraksiyonu için orta derecede sensitivite
	300–600 gen	Tanısal; Klinik denemeler; Klinik araştırma	Kapsam
	<100 gen	Tanısal; Hastalık progresyonunun monitörizasyonu	Kapsam
	50–80 genes, spesifik ekzonlar, varyantlar	Tanısal	Kapsam
	mRNA	Varyant validasyonu; neo-epitop ifadesi; Füzyon tespiti	Hedef dokunun her zaman ulaşılabilir olmaması
	Füzyon genleri	Füzyon tespiti	Hedef dokunun her zaman ulaşılabilir olmaması; Kapsam; Varyant validasyon kapasitesi hedef bölgeye sınırlı

Tablo 2. Yeni Nesil Dizileme Uygulamaları. Genome Medicine (2016) 8:112'den⁶ değiştirilerek alınmıştır.

Varyant tipi	Kodlayan / Kodlamayan	Yıkıcı	Ekzomla Bulunma	Genomla Bulunma
5'/3'UTR varyantları	Kodlamayan	Hayır	Evet	Evet
Sinonim varyantlar	Kodlayan	Hayır	Evet	Evet
Sinonim olmayan varyantlar	Kodlayan	Evet	Evet	Evet
Çerçeve insersiyon/delesyonlar	Kodlayan	Evet ⁽¹⁾	Evet	Evet
Stop kodon kaybı/kazanımları	Kodlayan	Evet ⁽¹⁾	Evet	Evet
Çerçeve insersiyon/delesyonlar	Kodlayan	Evet	Evet	Evet
Splice donör/acceptor varyantları	Kodlayan	Evet ⁽¹⁾	Evet	Evet
Intronik varyantlar	Kodlamayan	Hayır	Bazen	Evet
Upstream promoter varyantları	Kodlamayan	Hayır	Bazen	Evet
Proximal downstream varyantlar	Kodlamayan	Hayır	Bazen	Evet
Distal enhancer/repressor varyantları	Kodlamayan	Hayır	Hayır	Evet
microRNA varyantları	Kodlamayan	Hayır	Hayır	Evet
cis-Regulatory element (CRE) varyantları	Kodlamayan	Hayır	Hayır	Evet
Büyük yapısal varyantlar	Kodlayan / Kodlamayan	Evet	Bazen	Evet

(1): Kanonik fonksiyon kaybettirici mutasyonlar

Tablo 3. Yaygın genetik varyasyon çeşitleri ve ekzom veya genom dizileme ile bulunabilme durumları. Hum Hered (2016) 81:78–87'den⁴ değiştirilerek alınmıştır.

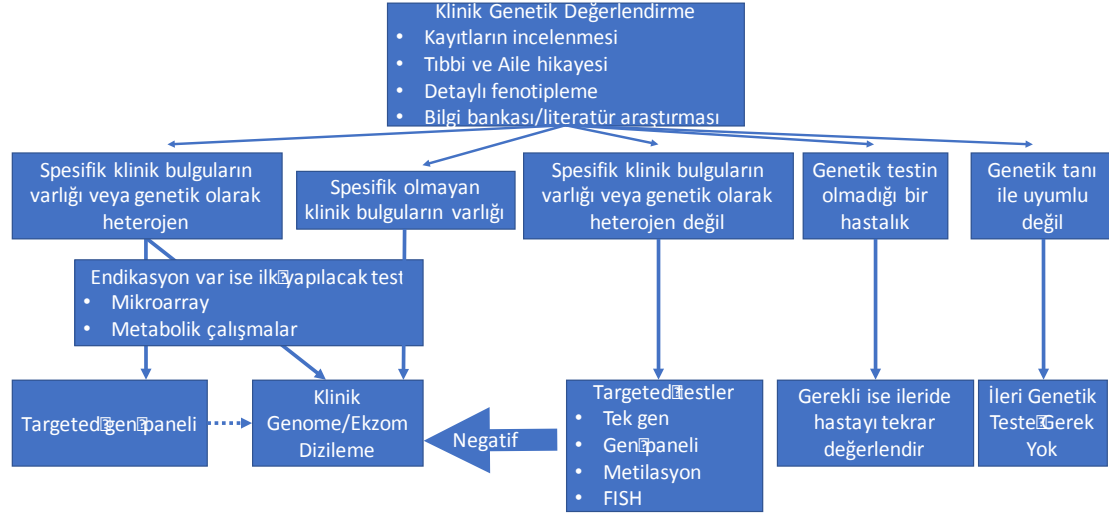
Hedefli tekrar dizileme ve TED, ilgilenilen genom bölgelerinin dizilemeden önce zenginleştirilmesini gerektirir. Hedef zenginleştirme birkaç strateji kullanılarak gerçekleştirilebilir, bunlar arasında PCR tabanlı yakalama (*capture*), moleküler ters çevirme (inversiyon) prob tabanlı yakalama ve hibrit yakalama yöntemleri yer alır⁶². Çok geniş bir gen paneline eşdeğer olan tüm ekzom dizileme, seçici zenginleştirmeyi ve insan genomunun protein kodladığı bilinen bölgelerinin çoğunun dizilenmesini içerir^{63,56}. Tüm genom dizileme hedeflenmiş dizilemeden ve TED'den farklılık gösterir çünkü hedef zenginleştirme gerektirmez ve hem protein kodlayan hem de protein kodlamayan genom bölgelerinin sorgulanmasına izin verir. Ancak, kullanılan yaklaşım her ne olursa olsun, YND testini isteyen hekimlerin, diğer tür genetik veya laboratuvar testlerinde olduğu gibi, detaylı fenotip bilgilerini sağlayarak, laboratuvara test sonuçlarının analizi ve yorumlanmasında yardımcı olmaları gerekir. Bu adım, kullanılacak uygun filtreleme stratejilerini geliştirmek ve önceliklendirme sırasını saptamak, incelemeye dahil edilecek genleri belirlemek, alt-fenotiplere doğru ilerlemek, gibi bir çok önemli basamağın gerçekleşmesinde olmaz ise olmaz bir basamaktır.

Tanısal yaklaşım seçimi, büyük ölçüde hastanın sağlık durumuyla ilişkilidir. Tek gen testleri (klasik yöntemler); klinik tablo ve diğer test sonuçları, belirli bir hastalığa karşılık geldiğinde ve spesifik bir genle ilişkilendirildiğinde tercih edilmelidir; akondroplazi örneğinde *FGFR3* geninde tek bir nükleotiddeki değişime bakmak gerektiğinde olduğu gibi³³. Klinik ve genetik heterojenite söz konusu olduğunda ise gen paneli yaklaşımı çok daha uygundur. Mutasyonun tipine göre büyük delesyon/duplikasyon dışlaması (MLPA, Mikroarray) panel gen testinden önce uygulanabilir ve bir kısım hastanın tanısı YND yapılmadan konulmuş olur. Mevcut durumda çok genli panellerin hedefli tekrar dizilenmesi, klinik laboratuvar ortamında sunulan YND'nin en çok kullanılan uygulamasıdır, çünkü hem emek, hem süre hem de maliyet açısından çok daha avantajlıdır^{55; 64-68}. Bu heterojen hastalar aynı zamanda ve özellikle aşırı klinik değişkenlik varlığında TED veya maliyetlerin makul seviyelere düşmesi durumunda gelecekte TGD kullanılarak da incelenebilir. Bu yaklaşım klinik özelliklerin özgün olmadığı ya da farklı gen mutasyonlarına bağlı kliniklerin aynı hastada olabileceğinin düşünülmesi durumunda tanıya ulaşmada kolaylık sağlayabilir. Bu yaklaşım genotipten fenotipe gidilen "reverse genetik" olarak tanımlanır.

TED ve TGD testleri sadece hastanın klinik endikasyonuna yönelik genlerin testleri ile sınırlandırılabilir. Eğer bu yapılırsa, laboratuvarın ilgili parametreleri (örneğin spesifik genlerin kapsamı) hastanın raporunda belirtmelidir. Çünkü bu durum hekimin, TED/TGD testinin performansını mevcut bir hastalığa özgü panel testiyle karşılaştırmasına izin vermiş olacaktır. Veriler analiz edilirken, resesif hastalıklar için taşıyıcı durumunu içerebilecek rastlantısal/ikincil bulguların potansiyel etkisine de dikkat edilmeli bu konuda hasta ve ailesine bu ek bilgileri vermek için gerekli onamın önceden alınmış olmasına dikkat edilmelidir.

Laboratuvarın sorumlusu veya testi uygulayan hekim, sundukları testin avantaj ve sınırlarını hastanın anlayacağı bir dilde mutlaka anlatmalıdır. Belirli bir genetik heterojenliğe sahip hastalıkların hedeflenmiş TED/TGD stratejileri çok daha verimli olabilir, ancak böyle bir test spesifik bir hastalık endikasyonu için sunulmuş ise gen ekleme ve kapsamdaki sınırlar açık şekilde not edilmiş olmalıdır. Hastalıkla ilişkili varyantların, ekzonik bölgelerin dışında olması bekleniyorsa (örneğin düzenleyici

bölge, kopya sayısı farklılıkları ve yapısal varyasyonlar gibi), test tasarımı stratejisi bu tip varyasyonları kapsayacak biçimde düzenlenmelidir. Laboratuvarlar, hastalıkla ilişkili varyantların tüm türlerini tespit edebilmek amacıyla bir TED testini tamamlayıcı testlerle takviye etmeyi düşünebileceği gibi her varyantın saptanamayacağını net şekilde ifade etmelidir. Bu varyantların eklenmesi, testin hassasiyetini artırmak için standart teste ilave yapılmasını veya destekleyici teknolojilerin kullanılmasını gerektirebilir (örneğin ekzonik olmayan bölgeler için bir TED testine yeni yakalayıcı (*capture*) problemlerinin eklenmesi, düşük kapsamlı bölgeleri dolduran Sanger dizileme, kopya sayısı farklılıkları ve yapısal varyantları tespit eden moleküler veya sitogenetik yöntemler gibi).



Şekil 2. Fenotipleme ve test seçimi ile ilgili önerilmiş bir algoritma örneği. Genet Med (2016) 18:1075-1084'den¹ değiştirilerek alınmıştır.

2.2.1. Tek gen

Özellikle küçük bir veya birkaç gende ($\sim < 8$ ekzon, $\sim < 1200$ bç/ekzon) ve orta büyüklükteki bir gende ($\sim 8-20$ ekzon, $\sim < 1200$ bç/ekzon) hastalığın genetik kökenini desteklemek için moleküler genetik tanı isteniyorsa, beklenen mutasyonların tipine bağlı olarak çeşitli klasik tanı yöntemleri kullanılır. Örneğin, küçük mutasyonlar için Sanger dizi analizi, ürün incelemesi ise örneğin multipleks sistemde PCR ile SRY varlığı/yokluğu, tek nokta incelemesi ise "PCR+Restriksiyon enzim kesimi", büyük yapısal mutasyonlar için MLPA, gerçek zamanlı PCR, Southern blot, microarray testleri, dinamik mutasyonlar ise Southern blot, fragman analizi, bağlantı analizi ise STR (simple tandem repeats; monozigotik, di-zigotik fetus ayrımı, fetal dokuda maternal hücre kontaminasyonunun dışlanması) gibi. Büyük genlerle ilişkili hastalıklarda ise (Duchenne/Becker Müsküler distrofisinde [DMD/BMD] 79 ekzonlu *DMD* testi, nörofibromatozide 57 ekzonlu *NF1* testi, kistik fibrozda 27 ekzonlu *CFTR* testi) mutasyonun frekansına bağlı olarak YND teknolojileri kullanılarak tanı algoritması uygulanır. Örneğin DMD/BMD'de olduğu gibi yapısal patolojiler ön planda ise önce MLPA ya da mikroarray, mutasyon saptanmayanlarda YND uygulaması gibi. Ancak her durumda, uygulanan yöntemin kapasitesinin farkında olunması ve ek testlerin yapılması gerekip gerekmediğine karar verilmesi gerekir.

2.2.2. Multigen panel

Gen panelleri, hipotez odaklı bir yaklaşımdır ve ilgilenilen hastalıkla ilgili genleri hedefler. Adayların sayısı sınırlı olduğu için paneller, özenli bir şekilde tasarlanarak hedeflerin tutarlı şekilde kapsanması (%99'a kadar) sağlanabilir⁶⁹. Bu yaklaşım takip edilerek YND ile tespit edilmeye karşı refrakter olan geriye kalan birkaç bölge Sanger dizileme kullanılarak kolayca dizilenebilir. Gen panellerinde testin hedefi (örneğin otozomal resesif işitme kaybı), test edilen genlerin isimleri, rapor edilebilir aralıkları, duyarlılık ve özgünlükleri belirtilmelidir. Yapılan testin tanısal yakalama oranı (*diagnostic yield*) mutlaka test yapılan laboratuvar tarafından önerilmeden önce belirlenmeli ve daha sonra sunulması gerekir. Bu noktada Matthijs ve ark.⁷⁰ önerdiği üçlü sınıflandırma sistemi testi önerecek laboratuvarlar tarafından takip edilip uygulanabilir:

A Tipi Test: Laboratuvar, kodlayan bölge ve yanlarında uzanan intronik dizilerin, >%99 oranında kapsadığını ve açıkta kalan diğer yerlerin Sanger dizilemesi (ya da başka dizileme yöntemi ile), hangi platformun kullanıldığına bağlı olarak tamamladığını beyan eder (Örneğin homopolimer bölgeleri).

B Tipi Test: Laboratuvar tam olarak hangi bölgelerin >%99 oranında kapsadığını beyan eder ve açıkta kalan tarafın bir kısmını Sanger veya başka bir dizileme ile tamamlar.

C Tipi Test: Bu tarz test YND dizilemenin kalitesine göredir ve ek bir Sanger dizileme ya da başka bir dizileme önerilmez.

Hastalığa hedeflenmiş gen panelleri, hastalıkla ilişkili olduğu bilinen genleri sorgular. Sınırlı bir gen seti üzerine odaklanmak, analitik sensitivite ve spesifitede artış için daha fazla kapsam (*coverage*) derinliğine olanak tanır. Daha fazla kapsam derinliği ise heterozigot varyantlara olan güveni ve mitokondri veya onkoloji uygulamalarında mozaisizm veya düşük seviyeli heterojenliği/heteroplazmikliği tespit etme olasılığını artırır. Bunun yanı sıra hedef hastalıkta sadece rolü belirlenen genler dizildiği için, bulguları klinik bir bağlamda yorumlama becerisi daha fazladır. Düşük kapsam gösteren bölgelerde (örneğin GC açısından zengin veya tekrar bölgeleri) YND verilerindeki boşlukları doldurmak için Sanger dizileme veya alternatif bir teknoloji kullanılabilir ve bu da testin klinik hassasiyetini artırır. Daha az sayıda gen hedeflenirse çalışma daha küçük kapasiteli cihazlarda yapılabilir ve TED/TGD ile kıyaslandığında her bir testte (barkotlama ve havuzlama) daha yüksek sayıda hasta örneği ile çalışılabilir. Ayrıca veri miktarı ve depolama gereksinimleri de daha yönetilebilir olur. Hedeflenen YND panelleri, genetik laboratuvarlarına moleküler tanısal testler için yapılan geri ödeme modellerine de daha kolay uyum sağlayabilir. Bu nedenle TED/TGD'nin teknik ve yorumlama niteliği hastalığa hedeflenmiş testlerin niteliğine ulaşana kadar, TED veya TGD yaklaşımlarına geçmeden önce, testi hastalığa hedeflenmiş panellerle başlatmak daha uygun olabilir. Diğer taraftan yeni hastalık genleri büyük bir hızla tanımlanmaktadır (<http://omim.org/statistics/update>), bu durum gen panellerinin periyodik olarak güncellenmesini gerektirir. Hastalığa hedeflenmiş bir panel, birden fazla örtüşen fenotip için gen içeriyorsa laboratuvarlar, analizi, alt-fenotiple ilişkili genlerin bir alt-paneliyle sınırlandırma seçeneğini hekimlere sağlamayı düşünmelidir (örneğin, geniş bir kardiyomiyopati gen paneli içindeki hipertrofik kardiyomiyopati genleri gibi), bu şekilde tespit edilmiş önemi bilinmeyen varyantların sayısı azaltılabilir. Laboratuvarlar aynı zamanda tespit edilen varyantların beklenen sayısını ve değerlendirilmeleri için gereken zaman ve

uzmanlığın hesabını da göz önünde bulundurmalıdır. Bir rehber olarak potansiyel klinik anlamlılığa sahip varyantların sayısı, analiz edilen hedef bölgenin büyüklüğüyle yaklaşık olarak orantılı olacaktır.

2.2.3. Birden fazla hastalığa yönelik genel paneller

Multigen panel ile aynı niteliktedir. Farklı olarak birden fazla hastalıkla ilişkili genleri içerir. Bununla birlikte kapsamı TED'ye göre dardır. Örnek olarak bilinen tüm OMIM genlerini içeren paneller verilebilir.

2.2.4. TED

TED, insan genomunda RNA kodlayan ekzonları ve tüm bilinen genlerin splice bölgelerini test eden bir yöntemdir. Ekzomun, genomun yaklaşık %1-2'sinden oluştuğu tahmin edilmekle birlikte, hastalığa neden olduğu bilinen mutasyonların yaklaşık %85'ini içermektedir⁷¹. TED'de ekzomu hedefleme, genelde ticari olarak satılan çeşitli reaktif kitleri kullanarak ekzomik kısmın yakalanmasına dayanır.⁶³ Laboratuvarlar ticari olarak temin edilen kitler arasındaki farkları göz önünde bulundurmalı ve her bir tasarımdaki tespit edilmeye dirençli bölgelerin olduğunun farkında olunmalıdır. Ekzom yakalama yöntemlerinin hiçbiri tam anlamıyla verimli değildir, bu nedenle laboratuvarlar kullanılan yöntemi ve kendi validasyon çalışmalarına dayanarak beklenen yakalama (*capture*) verimliliğini tanımlamalıdır. Hedeflenen kodlama bölgeleri için örneği zenginleştirmek amacıyla dizileme öncesi örnek hazırlığı yapılması gerekir. YND yoluyla ekzomun kapsamına ait güncel tahminler %90 ile %95 arasındadır⁷². Her ne kadar sayılar sürekli artıyor olsa da mevcut dizileme derinliklerinin kullanılmasıyla hedeflerin %10 kadarının iyi bir şekilde kapsanmadığı ($\leq 20x$)^{33; 73} tahmin edilmektedir. Bunun yanında herhangi bir belirli fenotiple ilişkilendirilen hemen hemen tüm genleri sorgulama yeteneğine ek olarak TED, hastanın durumundaki değişikliklere ve genetik bilgilerde kaydedilen ilerlemelere göre verileri, farklı zaman noktalarında tekrar analiz etme olanağını sunar⁷⁴.

Geçen birkaç sene içinde TED'in özellikle SNV (*single nucleotide variations*, tek nükleotidlik varyantlar) ve küçük INDEL'leri (küçük yapısal varyantlar, insersiyon ve delesyonlar) tespit etmede güçlü bir araç olduğu kanıtlanmıştır. Ancak bu varyantların hastalık bağlamında yorumlanması ve raporlanması, klinik laboratuvarlar için kolay olmayabilir. Halka açık veri tabanlarının ve literatürün yardımıyla bile varyantların büyük bir çoğunluğu, önemi bilinmeyen varyantlar (VUS, *variant of unknown significance*) olarak sınıflandırılır⁷³. VUS (Nonsense mutasyonların, çerçeve kayması mutasyonların ve splice bölgesi mutasyonların aksine), net bir patojenliği olmayan varyantlara karşılık gelir ve klinik öneminin bildirildiği varyant veri tabanlarında (ClinVar gibi) raporlanmamışlardır (Ya da VUS olarak bildirilmişlerdir). Diğer aile bireylerinin dizilenmesi, örneğin *de novo* mutasyonu beklendiğinde ebeveynlerin dizilenmesi VUS'un yorumlanmasına yardımcı olabilir. Yine de bir hastalığa neden olan varyant ile nadir görülen bir polimorfizm arasındaki ayırım belirsizliğini korumaktadır. Klinik laboratuvarların çoğunda VUS rapora eklenir ama klinik uzmanların büyük çoğunluğu klinik ilişki bariz olmadığı için herhangi bir işlem yapmaz. Bu VUS'ların bazıları, yeni literatür oluştuğunda ya da aynı VUS'a sahip ilave vakalar teşhis edildikçe tekrar sınıflandırılabilir.

Tanı laboratuvarlarının çoğu, varyantları şu anda Sanger dizileme ile doğrulamakta ve patojenlik hakkında daha fazla kanıt sağlamak ve genetik danışma için mümkün olduğunca segregasyon analizi yapmaktadırlar⁶⁹. Teknoloji ilerlemeye devam ettikçe

ikinci bir yöntem yoluyla sistematik validasyon, teşhis edilen her varyant için gerekli olmayabilir. Önceki çalışmalar, yüksek kalite skoru gibi belli kriterler uygulandığında Sanger ile YND arasında %100 uyuşma olduğunu göstermiştir^{73; 75}. Literatürde birçok kez raporlanan ve fonksiyonel çalışmaların patojenliğini net şekilde gösterdiği varyantlar da bu kategori içine girebilir. Özellikle bulgular sadece birkaç vakada bildirilmişse ve hasta yönetimi TED sonuçlarına göre yapılacaksa, Sanger dizileme teyidi, düşük kapsamla dizilenmiş bölgelerde patojenik şüphesi olan varyantlar için yapılmalıdır^{73; 76}. INDEL'lerin teşhisi halen zorlu bir iştir ve sonuçların da yüksek oranda yanlış-negatif çıkma olasılığı vardır. Bu noktada bu varyantlar raporlanmadan önce Sanger dizileme ile mutlaka doğrulanmalıdır. Böylece INDEL'lerde gerek varyantın gerçek olup olmadığı gerekse de raporda kullanılacak isimlendirmede netlik sağlanmış olur. TED, hem hastalıkla ilişkili olduğu bilinen genlerdeki varyantları tespit etmek için hem de yeni gen-hastalık ilişkilerini keşfetmek için kullanılır. TED'nin kapsamı ve maliyeti, hedeflenen gen panelleri ile TGD arasında olmaktadır. Bazı klinik laboratuvarların uyguladığı bir strateji, TED'yi yapmak ama sadece hastalıkla ilişkili olduğu bilinen genlerin yorumlamak ve bu genler için rapor hazırlamak şeklindedir. Hastanın semptomlarını açıklayabilen hiçbir mutasyon tespit edilemez ise, yeni hastalık-gen ilişkilerini tespit etmek amacıyla geriye kalan ekzom için veriler tekrar analiz edilebilir. **Çalışma grubumuz TED/TGD raporlamasında en azından testin yapıldığı toplumda sık görülen hastalıklar için hastanın taşıyıcılık durumunun belirtilmesi, ve nadir hastalıklardan sorumlu olan gen mutasyonu varlığında raporla hastanın hekiminin uyarılmasının gerekli olduğunu düşünmektedir.** Bir ekzom verisi içerisinde kapsama derinliği (*coverage depth*) tekdüze olmadığından, TED'nin analitik hassasiyeti, hedeflenen gen panellerinin hassasiyetinden daha düşük olabilir. Hastalığa hedeflenmiş test panellerinde eksik içeriği doldurmak için yaygın şekilde kullanılan Sanger dizileme, TED'de özellikle belli genler için istenmedikçe yapılmaz, testi çok pahalı ve zaman alıcı kılar ve nadiren kullanılır.

2.3.4 TGD

TGD, haploid bir insan genomuna ait üç milyar bazı dizilemeyi hedefler. Genomun hem kodlayan hem de kodlamayan bölgelerini içerir. TGD'nin bir avantajı, dizileme öncesi örnek hazırlığının kolay olmasıdır, yani PCR veya hedeflenen bölgeler için zenginleştirme stratejileri gerektirmez böylelikle TED'de görülen zenginleştirme kaynaklı hatalar yoktur. Yöntem her ne kadar yakalama sınırlamalarına (*capture limitations*) bağlı olmasa da, insan DNA'sının tüm bölgeleri mevcut yöntemlerle doğru şekilde dizilenebilir değildir (örneğin tekrarlayan gen bölgeleri gibi), bu nedenle trinükleotit tekrar hastalıkları gibi hastalıkla ilişkili olduğu bilinen bölgenin değerlendirilmesine ilişkin sınırlamalar günümüzde bu yöntemde de geçerlidir.

Kodlama yapmayan bölgelerdeki varyantların yorumlanmasındaki kısıtlılıklar nedeniyle kodlayan bölgeler genelde en önce analiz edilir. Eğer neden olan mutasyonlar bulunamazsa, bu durumda veriler tekrar analiz edilerek hastalıkla ilişkili genlerin ifadesini etkileyebilen kodlama yapmayan bölgelerde düzenleyici bölge varyantları (*enhancer*, *promoter*, *intron*daki düzenleyici bölgeler gibi) taranabilir. Veriler aynı zamanda kodlama bölgeleri dışında da olabilen kopya sayısı farklılıkları (*CNV*, *copy number variations*) veya yapısal varyantlar için de incelenebilir ya da artan kantitatif doğruluk nedeniyle TGD'de daha kolay tespit edilebilirler. Her ne kadar bahsedilen sınırlamaların gelecekte azalma olasılığı olsa da, TGD en az ortalama kapsam derinliği ile şu anda en pahalı teknolojidir.

3. Genotipleme

3.1. Yakalama (Capture) Platformu Seçimi

Bir dizi analizi cihazı seçerken, dizilenen bölgenin büyüklüğünü, gereken kapsam derinliğini, öngörülen örnek hacmini, geri dönüş zamanını ve maliyeti titizlikle göz önünde bulundurmak gerekmektedir. Bu konular masaüstü bir dizileme cihazına mı yoksa bağımsız bir dizileyiciye mi yatırım yapılması konusunda verilecek karar için önemlidir, zira bu iki dizileyici arasında dizileme kapasitesi açısından belirgin farklar vardır. Örneğin uzun okuma uzunlukları, yüksek homoloji gösteren bölgelerin analizi gibi belli uygulamalar için kullanışlı olabilir. Kısa dizi okuma teknolojileri diğer uygulamalar için yeterli olabilir. Maksimum okuma uzunluğu platforma bağlı olduğu için, test tasarımı, platform seçimi ve okuma uzunluğunun seçimi tespit edilmesi gereken varyantın tipine dayalı olmalıdır.

Genel dizileme metodları arasında tek uçlu dizileme (Single-end; genomik DNA okumaları sadece tek bir uçtan dizilenir) ve çift/eşleştirilmiş uçlu dizileme (*Paired-end*; her iki uç dizilenir ve Mate-pair end; eşleştirilmiş uç dizilenir) yer alır. *Paired-end* (kısa insertler için; 150-600 bp) dizileme, özellikle tekrar eden bölgelerde okumaları haritalandırma yeteneğini artırır ve her bir DNA fragmanının iki yönlü dizilemesi yapıldığı için testin kapsamını (*coverage*) artırmada ilave bir avantaja sahiptir. *Mate-pair* (uzun insertler için; 2-40 kb için) özellikle yapısal varyant tespiti için kullanışlıdır. Prenatal testler ve tedavi amaçlı karar alma için yapılan hedeflenmiş NGS gen panelleri, zamana karşı daha az duyarlı olan diğer uygulamalarla karşılaştırıldığında daha hızlı geri dönüş sağlar. Yüksek çıktılı enstrümanlar, ancak yeterli miktarda test istenirse örnek başına düşen maliyeti azaltabilir. Germline heterozigositesi altındaki varyantların tespit edilmesini gerektiren bozukluklar için (örneğin tümörlerde somatik mutasyon testi ve heteroplazmik mitokondri varyantlarının tespiti gibi), daha yüksek kapsama elde eden yaklaşımlar tercih edilmelidir.

3.1.1. Genetik coverage (örneğin kaç gen kapsanıyor)

Kullanılan platforma ve yakalama yöntemine bağlı olarak bir testin (panel, TED, TGD) hangi genleri kapsadığı testi öneren laboratuvar tarafından genel bilgilendirme amacıyla ve testin istenmesi esnasında daha sağlıklı karar verebilmek amacıyla belirtilebilir. Ancak burada unutulmaması gereken genomik mimarinin doğası gereği kapsanamayacak bölgelerin (tekrar bölgeleri gibi) ve testin sınırlılığı dolayısı ile tespit edilemeyecek (kopya sayısı değişikliği gibi) varyantların testi isteyen hekim tarafından iyi anlaşılması ve testin isteneceği kişi ya da ailelere iyi aktarılması gerekmektedir.

3.1.2. Matematik coverage (örneğin, kaç baz kapsanıyor)

İstatistiksel kapsama her bir YND kullanılarak yapılan (özellikle TED ve TGD) analizlere ait raporların ayrılmaz bir parçasıdır. Amplikonların yüzde kaçının en az kaç (x) (0, 4, 8, 10, 20, 50 gibi ayrı değerlerle birlikte ortalama değer) derinlikte kapsandığının belirtilmesi önerilmektedir. Böylelikle testin başlangıç endikasyonunu yerine getirip getirmediğinin ve ek bir genetik teste ihtiyaç olup olmadığının anlaşılmasına yardımcı olmaktadır.

TED'e ve TGD'ye yönelik tavsiye edilen kapsama oranlarını Tablo 4'de görebilirsiniz.

Kategori	Tespit veya uygulama	Tavsiye Edilen Kapsama (x)	Referans
Tüm genom dizilemesi	Homozigot SNV'ler	15x	Bentley et al., 2008 ⁷⁷
	Heterozigot SNV'ler	33x	Bentley et al., 2008 ⁷⁷
	INDEL'ler	60x	Feng et al., 2014 ⁷⁸
	Genotip çağrılar	35x	Ajay et al., 2011 ⁷⁹
	CNV	1-8x	Xie et al., 2009 ⁸⁰ ; Medvedev et al., 2010 ⁸¹
Tüm ekzom dizilemesi	Homozigot SNV'ler	100x (3x local derinlik)	Clark et al., 2011 ⁶³ ; Meynert et al., 2013 ⁸²
	Heterozigot SNV'ler	100x (13x local derinlik)	Clark et al., 2011 ⁶³ ; Meynert et al., 2013 ⁸²
	INDEL'ler	-	Feng et al., 2014 ⁷⁸

Tablo 4. TED'de ve TGD'de Kapsama ve Okuma Tavsiyeleri

Gen Panellerinde panelin tasarım aşamasının bir parçası olması gereken bu basamak testi isteyen hekimlere önceden bilgilendirme şeklinde sunulmalıdır. Eğer mümkünse yukarıda belirtilen 3'lü gruplandırma sisteminden hangisine dahil olduğunun tespit edilip paylaşılmasında yarar vardır. TED ve TGD analizlerinde ise test endikasyonu ile ilişkili ya da ilişkisiz ilgi duyulan genlerin hangi oranda kapsandığına dair bilgi test öncesinde testi isteyen hekim tarafından testi yapan laboratuvar ile paylaşıldığı takdirde bilgilendirme amacıyla verilebilir.

3.1.3. Even coverage pattern (eşit kapsam düzeni)

Bir nükleotidin genin hangi bölgesinde yer almasına bağlı olmaksızın (ekzon, intron vb.) kapsama derinliğinin eşit düzeyde seyretmesidir. Böylelikle özellikle kopya sayısı değişikliklerine dair bilgi edinme şansı ortaya konabilmektedir. Tasarımına bağlı olarak Gen Panelleri (genom verisi üzerinden panel odaklanması gibi) ve TGD analizi eşit kapsam düzeyini sunabilen testlerdendir.

3.2. Dizileme Platformu Seçimi

Günümüzde birkaç adet ticari YND platformu mevcuttur ancak teknoloji gelişmeye devam etmektedir. Klinik laboratuvar ortamlarına entegre edilen ilk nesil teknolojiler, bir *flowcell* üzerine çok büyük ölçekte paralel şekilde dizilenen klonal olarak amplifiye edilmiş DNA fragmanlarından yararlanır^{55; 56; 83-86}. Bu teknolojiler sentez yoluyla dizileme (*sequencing by synthesis*), ve bağlama yoluyla dizilemeyi (*sequencing by ligation*) içeren çeşitli kimyasal olaylardan faydalandıkları için, platformlar benzer işlemlerden geçirme adımlarını paylaşır. Önce DNA parçalanır ve platforma özgü oligonükleotid adaptörleri onarılan uçlara eklenerek bir fragman kütüphanesi oluşturulur. Daha sonra örneğin Illumina'da fragmanlar klonal olarak amplifiye edilir ve sonra bir *flowcell* üzerinde dizilenir, elde edilen luminesan veya floresan görüntülerin algoritmik olarak işlenmesiyle dizi okumaları elde edilir⁵⁵. Ion Torrent sisteminde ise hedef PZR, primerlerin kısmi kesilmesi, barkod ve adaptör bağlanması ve saflaştırması ardından emülsiyon PCR ile klonal çoğaltma, yüklü partiküller üzerinde zenginleştirme ardından senteze dayalı dizileme yapar ve nükleotidlerin uzayan DNA zincirine eklenmesi sonucu açığa çıkan H⁺ iyonlarının neden olduğu pH değişimlerini algılayan bir sisteme sahiptir. Dolayısıyla dizileme için floresan işaretli nükleotidlere, kamera ya da lazer sistemlerine ihtiyaç duymaz⁸⁶. Diğer YND platformları da bazı hizmet sağlayıcılardan tedarik edilebilir⁸⁷. Mate-pair dizileme olarak bilinen tamamlayıcı bir yaklaşım, daha uzun DNA fragmanlarının YND ile analizine izin veren çift uçlu dizileme stratejisinin modifiye edilmiş bir şeklidir, bu şekilde yapısal yeniden düzenlemeler daha iyi açıklanır. Eş çiftler, bilinen büyüklüğe sahip DNA fragmanlarının ortak bir bağlayıcıya daireselleştirilmesiyle elde edilir, okumalar bilinen bir mesafede bağlayıcıdan ikiye ayrılır ve sonrasında çift uçlu strateji kullanılarak dizilenir^{84; 88}. Her bir platform laboratuvarla ilgili özgül parametrelere ve enstrüman büyüklüğünü, enstrüman maliyetini, işlem süresini, okuma uzunluğunu ve numune başı maliyeti içeren test gerekliliklerine sahiptir.

3.2.1. Toplam dizileme kapasitesi

Toplam dizileme kapasitesi platformun üretebileceği dizi ürünü boyutunu ifade etmektedir. Bu değer platforma göre değişmekle birlikte 7.5 Gb - 6000 Gb arasında ifade edilmektedir.

3.2.2. Dizi okuma uzunluğu; amplikon boyu

Platformun bir seferde okuyabildiği dizi büyüklüğünü ifade etmektedir. Örneğin Illumina Platformları bu okumaları 150-300 bp ile yapmaktadır. Bu değer genellikle 25 milyon – 20 milyar arasında ifade edilmektedir.

3.2.3. Dizileme işlemi süresi

Platformun dizileme işlemi ne kadar sürede tamamladığını ifade etmektedir. Bu değer genellikle 4 saat – 6 gün arasında ifade edilmektedir. **Tablo 5'de** Illumina'ya ait farklı platformların toplam dizileme kapasitesinin, dizi okuma uzunluğunun ve dizileme işlemi sürelerinin yer aldığı tabloyu görebilirsiniz.

Illumina Platformları	MiniSeq	MySeq	NextSeq	HiSeq	HiSeq X	NovaSeq
Dizileme İşlemi Süresi	4–24 saat	4–55 saat	12–30 saat	< 1–3.5 gün (HiSeq 3000/HiSeq 4000) 7 saat–6 gün (HiSeq 2500)	< 3 gün	19–40 saat
Toplam Dizileme Kapasitesi	7.5 Gb	15 Gb	120 Gb	1500 Gb	1800 Gb	6000 Gb
Dizi Okuma Uzunluğu	25 Milyon	25 Milyon	400 Milyon	5 Milyar	6 milyar	20 milyar
Maksimum Okuma Uzunluğu	2x150 bç	2x300 bç	2 × 150 bç	2 × 150 bç	2 × 150 bç	2 × 150 bç

Tablo 5. Illumina platformlarına ait bazı bilgiler

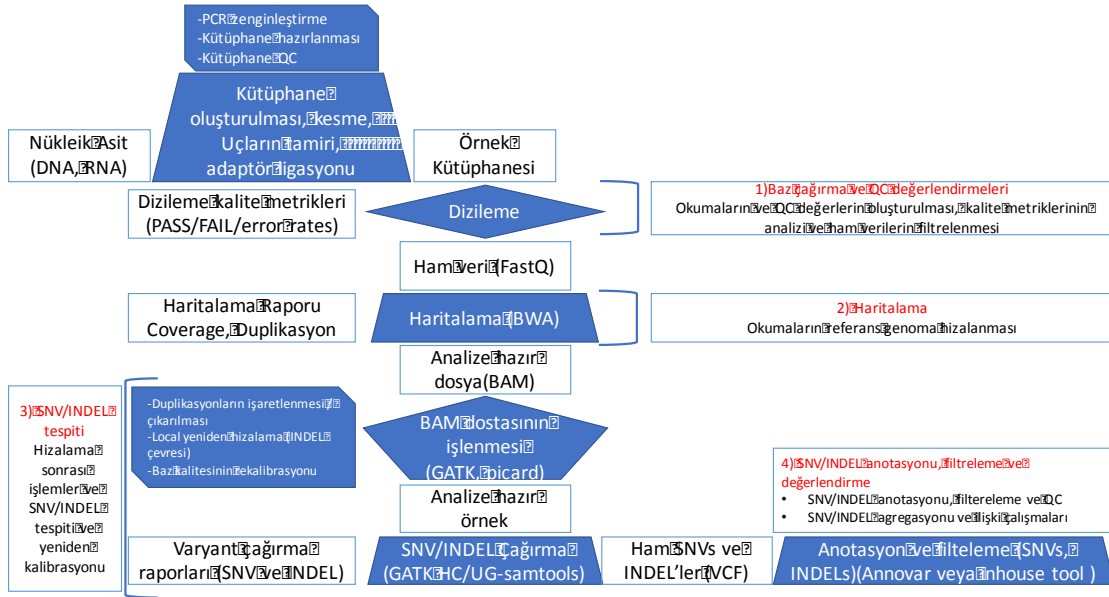
3.2.4. Son kalite ve doğruluk ölçümü

4. WET-LAB (Islak laboratuvar çalışmaları)

Klinik laboratuvarında ıslak tezgah (*wet bench*) süreçlerinin detaylı bir dokümantasyonu, kalite değerlendirmesinin kritik bir parçasıdır. DNA/RNA izolasyonu ve küçük parçalara kesilmesi, kütüphanenin hazırlanması, barkodlama, örnek havuzu ve dizi üretimine ait tüm standart çalışma protokolleri kayıt altına alınmalıdır, böylece her bir adım ve sonraki manipülasyonlar takip edilebilir. Bu ise tüm yöntemlerin ve kullanılan malzemelerin dokümantasyonunu ve ıslak tezgah işlemlerinin başından sonuna kadar kullanılan enstrümanları, enstrüman yazılımını ve sürümleri içerir. Ayrıca kullanılan kontrollerin de açıklanması gerekir (**Şekil 3**).

4.1. YND uygulamaları için ne tür örnek gerekir?

YND, nükleik asitlerin izole edilebildiği her örnekten gerçekleştirilebilir (örneğin periferik kan, deri biyopsisi, fibroblast kültür hücreleri, taze veya donmuş dokular, parafine gömülmüş dokular, koryonik villüs direkt/kültür, amnio direkt/kültür, kordon kanı, kardiyosentez numuneleri gibi). Her bir numune tipi için standart operasyon prosedürleri (SOP) geliştirilmelidir. Laboratuvarların numune tiplerini belirlemesi ve NGS analizleri için gereken minimum miktarı tespit etmesi gerekir. DNA'nın kalitesi ve varyant tespiti gereksinimleri, bazı numune tipleri için olasılıkla farklılık gösterecektir, bu nedenle laboratuvarın kendi validasyon verilerine dayanarak her bir numune tipi için kabul edilebilir parametreler tayin etmesi gerekecektir. Bu tayini de her bir numune şekli için validasyon işlemleri ile doğrulaması gerekmektedir.



Şekil 3. Wet-lab ve dry lab ana işlemlerinin şematizasyonu. Expert Rev Mol Diagn (2015) 15,749-760'dan⁷ değiştirilerek alınmıştır.

4.2. Hasta örneklerinin alımı, nakli, muhafazası

Genetik hastalıklar tanı merkezlerinin, DNA bazlı testler için gerekli olan örnek alımı, nakli ve muhafazası ile ilgili takip edilen algoritmalar yeni nesil dizileme testleri için de geçerlidir.

4.3. Nükleik asit izolasyonu

4.3.1. DNA gereksinimleri ve işlenmesi

Laboratuvar testi gerçekleştirmek için, minimum DNA gereksinimlerini belirlemelidir. Dikkat edilecek konular arasında, testin her bir numune için mi yapıldığı ve teyit ve takip işlemleri için ne kadar DNA'ya gereksinim duyulacağı yer alır. Yeterli bir DNA kalitesi ve miktarı elde edebilmek için laboratuvar kendi prosedür kılavuzunda ve/veya kalite yönetim programında DNA izolasyonu ve miktar tayini için yazılı protokollere sahip olmalıdır (fluorometri, spektrometre gibi). Laboratuvar bu parametrelerin dokümantasyonunu her bir hasta kaydında bulundurmalıdır.

4.3.2. Numune kabul şartlarına uymayan örnekler

Bir numune laboratuvarın numune kabul şartlarını karşılamıyorsa ve standart altı kabul ediliyorsa, önerilen numunenin reddedilmesi ve yeni bir numune istenmesidir. Yeni bir numune elde edilmesi mümkün değilse, laboratuvar bu teknikte deneyimli ise tüm genom amplifikasyonu veya saflığı iyi olmayan örnekler için yeniden çöktürerek izolasyon yapabilir. Bu durumda teknik içinde kendiliğinden var olan olası hatalara yatkınlıklar (tüm genomun eşit olmayan veya eksik amplifikasyonu gibi) raporda detaylandırılarak hekimin ve hastanın bu tekniğin sınırlamalarından haberdar olması sağlanmalıdır. Tüm genom amplifikasyon işleminin nasıl ve ne zaman yapılacağını belirten yazılı standartlar, laboratuvar kılavuzunun içine dâhil edilmelidir.

4.4. Ekzom kütüphanesinin hazırlanması

Kütüphanenin oluşturulması, belli büyüklük aralığına sahip ve her iki uca adaptör dizileri içeren DNA fragmanlarının rastgele oluşturulması sürecidir. Adaptörler,

platforma özgü PZR'nin ve dizileme primerlerinin tamamlayıcısıdır. Genomik DNA'nın parçalara ayrılması birçok yöntemle yapılabilir, bu yöntemlerin her biri zayıf ve güçlü yönleri sahiptir. Uygulamaların/platformların çoğu için kütüphanenin PZR amplifikasyonu, dizilemeden önce gereklidir ve aşağıdaki basamakları kapsar:

4.4.1. Parçalara ayırma (fragmentasyon) ve uçların tamiri

4.4.2. Adaptor ligasyonu ve kesik uç tamiri

4.4.3. Boyut seçme

4.4.4. Kütüphane normalizasyonu ve kuantifikasyonu

4.5. TED ve Gen Panelleri için Hedeflerin Zenginleştirilmesi

TGD yapılmadıkça ilgilenilen genler veya bölgeler dizileme öncesi yakalanıp zenginleştirilmelidir. Hedefler görece küçük sayıda genlerden (belirli bir hastalıkla ilişkili tüm genler gibi) tüm ekzoma (protein kodlayan tüm bilinen ekzomlar) kadar değişiklik gösterebilir. Hedef zenginleştirme yaklaşımları ikiye bölünebilir: multipleks PZR tabanlı yöntemler (tek veya multipleks PZR veya damlacık PZR) ve katı ya da solüsyon içi oligonükleotid hibridizasyon bazlı yöntemler. Hibridizasyon bazlı yakalama TED için kullanılabilir ancak PZR tabanlı yaklaşımlar henüz bu büyüklüğe ölçeklenememektedir.

4.5.1. Barkodlama

Barkodlama numunelerin özgün dizi tabanlı kodlarla moleküler etiketlenmesi anlamındadır, tipik olarak üç veya daha fazla baz çiftinden oluşur. Bu işlem hasta numunelerinin birbirleri ile karışmadan havuzlanmasını sağlar ve işlem maliyetini azaltır. Havuzlanabilecek numunelerin sayısı, dizilenecek bölgenin büyüklüğüne bağlı olacaktır. Barkodlar adaptörün bir parçası olabileceği gibi, çoğu protokol içinde yer alan bir PZR zenginleştirme adımının bir parçası olarak da eklenebilir.

4.6. Ekzom Dizileme

4.7. Kontrol Gereklilikleri

4.7.1. Kalıbı olmayan controller (NTC; No template control)

Tüm amplifikasyon basamaklarında

4.7.2. Negatif kontrol

İlk validasyon çalışmalarında ve takibinde periyodik olarak

4.7.3. Pozitif / Sensitivite kontrolü

Her dizileme işleminde

4.8. Dizileme

5. DRY-LAB (Kuru laboratuvar çalışmaları)

YND teknolojileri, veri saklama, analiz ve yorumlama açısından önemli bilgisayar altyapısı gerektiren veriler üretir. Bu hızlı şekilde gelişen teknolojiler, gen panellerini, ekzomları ve tüm genomu çeşitli genetik bozukluklarda analiz etmek için sayısı günden güne artan laboratuvarlarca klinik ortamda kullanılmaktadır. YND teknolojileri ile üretilen veriler bir dizi adımda analiz edilir (**Şekil 4**). İlk başta dizileme platformu ile milyon-milyar arası değişen dizi okumaları üretilir (birincil analiz). YND veri analizi araçlarının bir araya getirilmesi (analiz planı:*informatik pipeline*), dizileme enstrümanı ile üretilen ham verileri işleyip analiz ederek bir rapor oluşturur. 2017 tarihi itibarıyla pek çok laboratuvar, kendi çalışma plan ve aşamalarını geliştirmek amacıyla, kurum içi geliştirilen analitik prosedürlere ek olarak ticari olan veya genel kullanıma açık olan araçların isteğe uyarlanmış kombinasyonlarını kullanmaktadır.

Veri analizinin bir sonraki adımında (ikincil analiz) YND okumaları, dizi varyantlarının nerede olduğunu tespit etmek için gereken bir referansa hizalanır. Dizileme

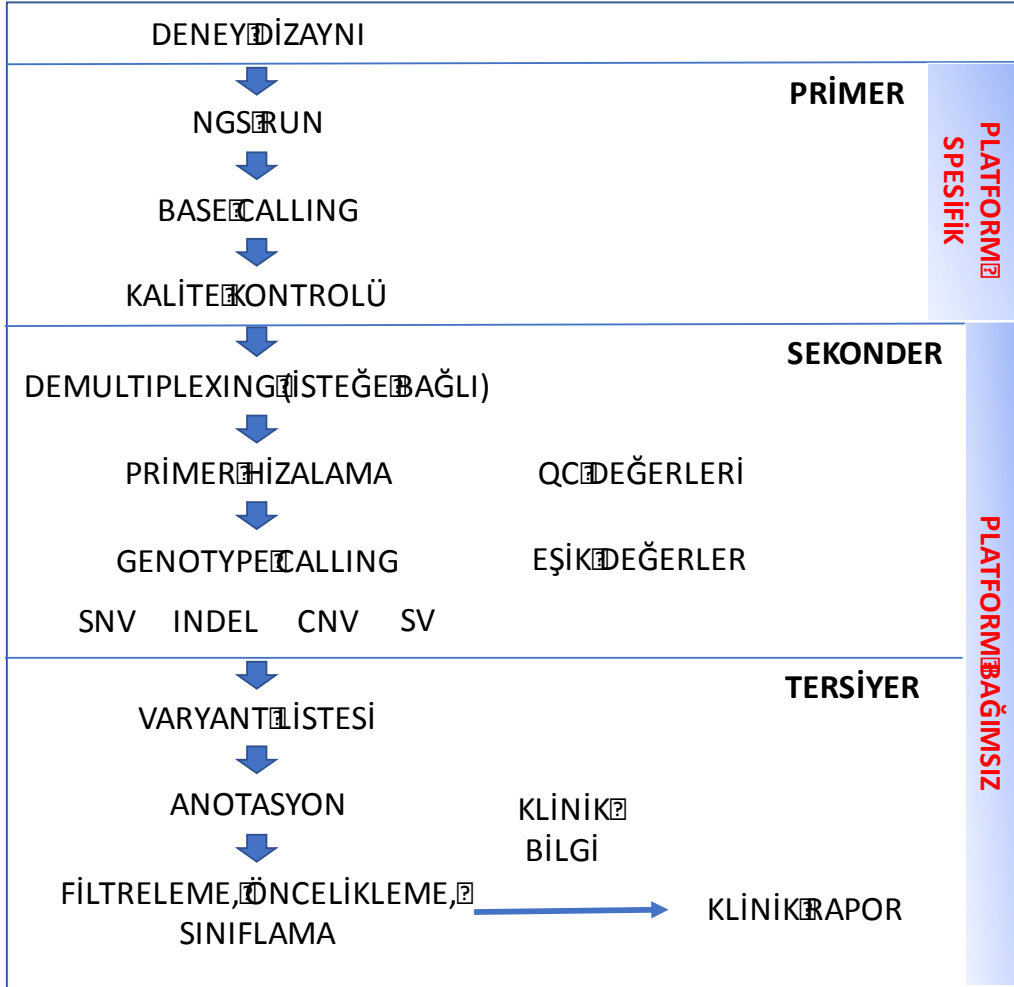
öncesinde birden fazla hasta numunesi havuzlanırsa, her bir hasta ile ilişkilendirilen okumalar, referans genomu hizalanmadan önce birbirinden ayrılır (*de-multiplexing*) ve bağımsız şekilde analiz edilir. Sonrasında okuma hizalama sistematik şekilde incelenir ve sıklıkla genomun lokal bölgelerinde yeniden hizalanarak yapay üretilen veriler uzaklaştırılır ve doğru genotip çağruları sağlanır. Sonuç olarak bu hizalar kullanılarak, adına varyant çağırma denilen bir süreçte hasta dizilimi ile referans genom arasındaki farklılıklar tespit edilir. Bu "ikincil analiz" ürünü, varyant çağrı dosyasında saklanan dizi varyantlarının tanımıdır. İkincil analiz sırasında yazılım araçları her bir okumayı haritalamak ve insan referans genomuna hizalamak için kullanılır, referans genom ise Genom Referans Konsorsiyumu tarafından sunulur (GRCh38; <http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/>). Bu ağ düzenli olarak güncellenir. Klinik laboratuvarların erişim tarihinin ve her bir hizalama için kullanılan veri nüshasının dokümanite edilmesi tavsiye edilmektedir. Bu şekilde varyant konumu bir referansa kadar geriye doğru takip edilebilir. Tüm insan genomu dizildiğinde ve referans bir diziyeye hizalandığında tipik olarak 3 ile 4 milyon arası dizi varyantı tespit edilir. Bunun içinde tek nükleotid varyantları (SNV), küçük insersiyonlar, delesyonlar ve bunların bir kombinasyonu (indel) yer alır. Kopya sayısı varyasyonlarını, karmaşık varyantları ve yapısal yeniden düzenlemeleri içeren diğer varyantların tespiti sorunlu olabilir ve sıklıkla özel bir analiz gerektirir.

Bulunan varyant için ilave analiz yapılarak hastanın klinik durumuyla ilgili olanlar tespit edilmeye çalışılır (üçüncül analiz). Öncelikle öngörülen moleküler sonuçlar (örneğin anlamsız bir varyant veya çerçeve kayması varyantı ile prematüre stop kodonunun oluşması veya yanlış anlam varyantının saptanması gibi) varyantların isimlendirilmesi (*annotation*) yapılabilir. Bu işlem klinik açıdan önemli olması beklenmeyenlerin filtrelenmesine izin verir. Daha sonra ilave tanımlar eklenir ve bunlar hastalık ilişkisi hakkında bilinenleri, popülasyon prevalansını ve diğer faktörleri içerir. Hastanın tıbbi durumuyla ilgili varyantları tespit etmek için önceki bölümlerde anlatılmış olan ayrıntılı bir klinik değerlendirmenin yapılmış olması gereklidir. İnsan nüfusunda var olan patojenik varyantların spektrumu sınırlı şekilde anlaşılmaktadır, ancak klinik bilgiler mevcut olduğunda hastalıkla ilişkili olması muhtemel varyantları öncelik sırasına koymak için analiz sırasında kullanılmaktadır.

Son adım bir test sonucu raporunun geliştirilmesidir, bu rapor dizileme analizinden gelen bulguları hastanın klinik verileriyle birleştirerek, tespit edilen varyantların herhangi birisinin veya bir kombinasyonunun kişinin hastalığını açıklayıp açıklamadığına bakılır. Birincil, ikincil ve üçüncül analizlerin çoğu önemli miktarda otomatik informatik bileşenler içerir ki bu da pek çok klinik moleküler test laboratuvarları için faaliyetlerinde önemli bir değişikliktir.

YND platformlarınca üretilen dizilim verilerinin devasa miktarı dikkate alındığında, doğru ve verimli veri işleme ve analiz için işlem aşamalarının geliştirilmesi elzemdir. Bu ise kapsamlı biyoinformatik destek ve donanım altyapısı gerektirir. Verilerin analizi için ıslak laboratuvar adımlarından bağımsız olan araçların kullanılması ve laboratuvarın bu veri analiz araçlarını kendi ihtiyaçlarına göre uyarlama olasılığı dikkate alındığında, analitik pipelinelerin ilk test geliştirme sırasında ayrı ayrı validasyonu tavsiye edilir. Piyasada geliştirilmiş bir yazılım kullanılıyorsa, laboratuvar satıcı tarafından verilen tüm validasyon verilerini mutlaka kayıt altına almalıdır, ancak laboratuvar aynı zamanda aracın bağımsız bir validasyonunu da kendi laboratuvar şartlarında yapmalıdır. Laboratuvar, toplam dizileme işleminin yeterli

kaliteye sahip olup olmadığını ve başarılı kabul edilip edilmediğini belirlemek için gerekli olan parametreleri ve eşik değerleri de tayin etmelidir. Bu kapsamda dizileme işlemi sırasında ve tamamlandığında ara noktalardaki analizler yer alabilir (örneğin gerçek zamanlı hata oranı, yakalanan hedef yüzdesi, hizalanan okumaların yüzdesi, ortalama kapsama derinliği, insert büyüklüğünün aralığı gibi). Bunun yanı sıra laboratuvar kapsama için eşik değerleri belirleyip takip etmelidir, böylece varyant çağırma için yeterli kapsamanın ve analitik hassasiyet ve özgüllüğü etkileyen allelik fraksiyonun elde edildiğinden emin olmalıdır. YND varyant çağırma araçlarının varsayılan eşik değerler uyguladığını unutulmamalıdır. Bu değerlerin analitik performansı artırmak amacıyla optimize edilmesi gerekebilir. Laboratuvar ayrıca, özellikle barkotlama kullanıldıysa analitik pipeline'ın numune kimliğini doğru şekilde takip edebildiğinden emin olmalıdır.



Şekil 4. Yeni nesil dizilemede bir analiz temel basamakları. Nat Biotechnol (2015) 33:689-93'den⁸⁹ değiştirilerek alınmıştır.

Biyoinformatik iş akışı (pipeline) için bir kalite güvence programı geliştirilerek NGS verilerinin analizi, yorumlanması ve raporlanması desteklenmelidir. Kalite güvence programı ayrıca, test işlemi sırasında geliştirilen pipeline sapmalarını raporlamak ve çözmek için laboratuvar tarafından yürürlüğe konan düzeltici önlemleri de belgelendirmelidir.

Laboratuvar, YND verilerinin analizinde ve her bir hasta testinin analizinde kullanılan

aşamaların her bileşenine ait spesifik versiyonu yakalamada kullandığı biyoinformatik işlemleri ve aşamalarını dokümanete etmelidir. Laboratuvarın yazılım sürümlerini, her sürümün içerdiği belirli değişiklikleri ve yeni sürümün klinik numuneler üzerinde uygulanma tarihini takip etmesini sağlayan bir sistem geliştirilmelidir. Sürecin her bir adımı için girdi ve çıktı dosyalarının tanımını, metrikleri ve optimum performans için Q/C (kalite kontrol) parametrelerini içerecek şekilde bir kalite güvence programı geliştirilmelidir. YND'nin biyoinformatik işlem aşamaları aşağıdaki gibi özetlenmeye çalışılmıştır:

5.1. Örnek bilgilerinin (*sample sheet*) oluşturulması

Basit bir virgüllerle ayrılmış dosya olup (CSV; *comma separated values*) içerisinde kütüphane için kullanılan kimya, örnek isimleri ve her bir örnek için kullanılan indeksler ile run'ı tanımlayan diğer metrikleri içerir.

5.2. Primer Analiz (Sekansların oluşturulması, Baz çağırılması ve *de-multiplexing*):

Kullanılan YND platformunun tedarikçisinin sağladığı yazılım ile yapılan baz çağırma ve *de-multiplexing* işlemidir. Bu işlem sonucunda çıktı olarak FASTQ (*Fast Adaptive Shrinkage Thresholding Algorithm and Quality*) dosyası oluşmaktadır (**Şekil 5**) (**Bilgi Kutusu 1**). Bu dosya nükleotid dizisini ve okuduğu her bir nükleotidin kalite değerlerini bir arada bulunduran bir yazılım dosyasıdır. Bu dosya kullanılan cihaz üzerinde veya yerel/bulut bazlı sunucular üzerinde oluşturulur (**Şekil 5**) ve tipik olarak baz çağırma içerir, bu şekilde dizi okumalarını (*sequence reads*) içeren dosya ve ilgili baz için kalite skorları oluşturulmuş olur. İki adet dosya formatı sıklıkla kullanılmaktadır: FASTQ baz başına kalite skorlarıyla birlikte okumaları ve BAM (*Binary Alignment/Map*) ikili hizalama format dosyasında ⁸ haritalanmış ve/veya haritalanmamış okumaları içerir. SAMTool, BAM dosyasını sıkıştırıp küçülterek dizi hizalama/harita (SAM: *Sequence Alignment/Map*) dosyasının içindeki bilgilerin okunabilir hale dönüşmesini sağlayan bir yazılım aracıdır. Veriler bir kalite kontrol adımından geçer, burada okumalar filtrelenir ve referans genomuna hizalanmadan önce sağlayıcı ya da laboratuvar tarafından belirlenen kalite kontrol kriterlerini karşılamayan baz çağırma içerirler uzaklaştırılır. Bu süreç sırasında okumaların 5' ve/veya 3' uçları dizileme aracı yazılımı tarafından otomatik olarak kesilir veya manuel ayarlama ile belirlenen kriterlere göre ayarlanır çünkü baz kalite skorları genelde her bir okumanın ucunda daha düşüktür. Dizileme cihazı yazılımı PCR primerlerini rutin olarak uzaklaştırmaz. Hedeflenmiş dizileme için primer diziler çıkarılmalı veya yazılım tarafından kırılmalıdır (yani primer dizi hizalamada kullanılması için tutulur ama varyant çağırma sırasında gizlenir). Bu şekilde primer bölgelerdeki SNP'lerin doğru varyant alel frekansı ile çağırılması sağlanır. Birincil analiz için ilave tavsiyeler ve standartlar için ilgili kaynaklara başvurulabilir ^{13; 90}.

Bilgi Kutusu 1: FASTQ formatı

Nükleotid sekanslarının ve onların kalite değerlerinin saklandığı düz yazı tabanlı dosya biçimidir (https://en.wikipedia.org/wiki/FASTQ_format). Her sekans 4 satırdan oluşmaktadır. İlk satır '@' işareti ile başlar ve sekansın kimliğini belirtir. İkinci satırda karakterle gösterilmiş ham sekans verisi vardır. Üçüncü satır + ile başlar ve isteğe bağlı boş olabilir ya da birinci satırdaki sekans kimliği yazılır. Dördüncü satırda ise sekansın kalite değeri ASCII formatında gösterilir. Kalite değeri sekansın okuma esnasında yanlış okuma olasılığı gösterilir. Büyük değerler, okumanın daha az olasılıkta yanlış okunduğu anlamına gelmektedir. $-10 \cdot \log_{10}(p)$ formülü ile kalite değeri olasılığa çevrilir.

```
@SRR001666.1 071112_SLXA-EAS1_s_7:5:1:817:345 length=36
GGGTGATGGCCGCTGCCGATGGCGTCAAATCCCACC
+SRR001666.1 071112_SLXA-EAS1_s_7:5:1:817:345 length=36
IIIIIIIIIIIIIIIIIIIIIIIIIIIIII9IG9IC
```

Şekil 5. NCBI sitesinde indirilen SRR001666.1 fastq dosyasının içindeki bir sekansın gösterimidir.

5.1.1 Sinyal analizi

Tüm Illumina tabanlı YND cihazlarında ilk ham veri BCL (*base calling*) dosyaları olup, ikili (*binary*) baz çağırma ve primer dizileme girdilerini gösterir. Bir *binary* dosyası olarak her siklustaki (sekans adımı) her sütunun (*tile*) hem baz çağırma hem de bu çağrılan bazların kalite bilgilerini içerir. FASTQ'dan farkı baz çağırmanın (*base calling*) direkt cihaz tarafından yapılmış olmasıdır. FASTQ dosya formatında ise baz çağırma tüm dizi okuması bittikten sonra yapılmaktadır.

5.1.2 Bazların çağırılması (*base calling*)

YND tabanlı varyant tespitinin verimliliği, baz çağırmanın doğruluğuna kritik şekilde bağlıdır. Baz çağırma, tek bir dizileme okuması içindeki her konumda bulunan spesifik nükleotidin teşhis edilmesidir; bu ise sürecin teknolojiye özgü doğası nedeniyle tipik olarak cihaz yazılımı içine entegre edilir. Her YND platformu, veri üretme süreci sırasında yapılan hataların türü ve oranını etkileyen belirli dizileme yargılarına (*bias*) sahiptir. Bu kapsam içine okuma üzerinde sinyal-yoğunluk (*intensite*) bozulumu, homopolimerik bölgelerde hatalı insersiyon ve delesyonlar girer⁹¹. Teknolojiye özgü yargıları (*bias*) hesaba katan baz çağırma yazılımı, platforma özgü sorunların ele alınmasında yardımcı olabilir. En iyi uygulama ise platforma özgü hataları azaltmak için tasarlanan bir baz çağırma paketi kullanmaktır. Genelde, uygun ve platforma özgü bir baz çağırma algoritması dizileme cihazı içinde olur. Her bir baz çağırma, çağırmanın kesinliğini değerlendiren bir kalite ölçüsüyle ilişkilendirilir. Bu genelde **Phred** benzeri skor olarak raporlanır (ancak bazı yazılım paketleri farklı bir kalite ölçüsü kullanır ve biraz farklı değişkenleri ölçer). Dizileme kimyasındaki süregelen gelişmeler ve yüksek performanslı baz çağırma algoritmalarının kullanılması sayesinde mevcut YND teknolojileri gelişerek %99.5'ten daha fazla bir baz çağırma doğruluğuna ulaşmıştır.

5.1.3 Baz kalite skorlaması

Baz çağırma hataları, varyant çağırma ve sonuçların biyolojik yorumlanmasını etkileyebileceği için, daha fazla analize geçmeden önce ham verilerin değerlendirilip sonrasında düzeltilmesi gerekir. Bu kalite kontrolü (QC) değerlendirmesi iki aşamalı bir süreçtir. Birinci aşama (hizalama öncesi), ham verilerin kendi kalitesini dizileme platformuyla üretilen metrikler kullanılarak değerlendirmeyi içerir (% hata, baz başına kalite skorları ve % duplikasyon). Yetersiz kaliteye sahip okumalar dışlanır ya da kırılabilir. Duplike okumalar genelde ekarte edilerek PCR tabanlı hatalara karşı önlem alınmış olur. Ardından, ikinci aşama yani haritalandırma gerçekleşir (aşağıya bakınız) ve hizalama profillerinden gelen metriklerin hesaplanmasını içerir. FASTQC gibi araçlar [Babraham biyo-informatik: www.bioinformatics.babraham.ac.uk/projects/fastqc/] ve Picard araç kiti [Picard: <http://picard.sourceforge.net/command-line-overview.shtml>] QC değerlendirmesinde kullanılırlar. Çok sayıda metrik üretirler, bunların bazıları şu şekildedir: referans genomuna hizalanan okumaların oranı, ortalama kapsama derinliği, aşırı temsil edilmiş diziler, hata oranı ve duplikasyon oranı, bunlar dizilemenin başarısını ölçmek için kullanılabilir. Kabul edilebilir olmayan nihai kaliteye sahip dizileme işlemi tekrar edilebilir ve ilave verilerle takviye edilebilir.

5.1.4 Okumaların üretilmesi (FASTQ dosyasının oluşturulması)

5.1.5 FASTQ verilerinin kalite değerlendirmeleri

Bu noktada ham veriler henüz cihazdan çıkmıştır. Fastq formatında verilerdir. Bu dosyalar nükleotid dizileri ve herbir bazın okumasının kalite değerlerinden (*Phred Score*) oluşur.

a. Ortalama derinlik (*coverage deep*)

Toplam okunan nükleotidsayısının okunan canlı türünün genom boyuna bölünmesiyle bulunan değerdir. TGD için beklenen *de facto* standard 40X derinlik, TED için 100X ve hedefli dizilemeler için >300X kullanılmaktadır.

b. Okumaların Q30 değerinin üzerinde olması

Herbir nükleotid okumasının Phred skorları üzerinden Q30 değerini geçip geçmediği kontrol edilir. Geçenlerin tüm nükleotid sayısına oranı bir değer üretir. Bu değer genel olarak %90 üzerinde olması beklenir. Bu ortalama okumanın iyi olduğuna işaret edebilir. Platform bağımlı olarak bu değer farklılaşabilmektedir.

c. Okunan dizilerin okuma pozisyonuna bağlı olarak Phred skorlarının dağılımı

Dizileme teknolojilerinin çoğunluğu "cycle by cycle" okuma yapmaktadır. Bazı durumlarda cihazın herhangi bir sürecinde kanallarda baloncuk oluşabilmektedir. Bu durum genel Q30 kalitesine etki etmeyebilmektedir. Ancak bu döngüde okunan tüm nükleotidler hatalı olabilmektedir. Bu durumun anlaşılabilmesi için herbir döngü için beklenen okuma kalitesinden çok sapılmadığının belirlenmesinde fayda vardır. Bu durumda veri yine de kullanılabilir gibi bu konudan haberdar olarak gerekli hassasiyet gösterilmesi gerekmektedir.

Ayrıca her bir pozisyondaki ortalama kalite değerine bakılarak analiz sürecinde kırma (*Trim*) işleminin gerekip gerekmediğine karar verilebilir. Örneğin baştan 3 nükleotid okumasını atmak gibi. Diğer bir taraftan uzun okuma yapan teknolojilerde 10000 baz okuma sonrası atılabilir. Bu son iki değerler örnek olarak verilmiş olup genel bir kriter olarak kullanılmamalıdır.

d. Okumaların ortalama Phred değerlerinin dağılımı

Okumaların ortalama Phred değerleri bulunur ve herbir Phred değeri kaç adet okuma içerdiği hesaplanır. Yine dizileme platformuna bağlı olarak okuma hücreleri (*cell*) üzerinde bazı bölgeler hatalı okunmuş olabilmektedir. Bu oran verilerin içerisinde ne kadar istenmeyen okuma olduğu yönünde bilgi vermektedir. Ayrıca bu dağılıma bakarak hangi standardı sağlayan okumaların hesaplama sürecine sokulacağı yönünde karar verilebilir. Bu hesaplamanın c maddesinde kırpma işlemine karar verilmiş ise bu süreçten sonra uygulanmasında fayda vardır.

Yukarıda tanımlanan kriterlerde gözlenecek herhangi bir anormalliğin laboratuvara geri besleme olarak verilmesinde fayda vardır.

a. Kırpma (*Trimming*):

i. **Primer:** Amplikon temelli dizilemede primerler okumalardan ayrıştırılmalıdır. Bu amaçla CutAdapt, BWA (haritalama esnasında *soft clipping*) yazılımları kullanılmaktadır.

ii. **Adaptör (opsiyonel):** Dizi adaptörleri ekleme boyutunun okuma uzunluğundan küçük olduğu durumlarda ayrıştırılmalıdır. Aksi takdirde haritalamada ve varyant çağrılmasında yanlış pozitif ve yanlış negative varyantlara yol açabilir. CutAdapt, BWA (*Burrows wheeler Aligner*; haritalama esnasında *soft clipping*), Trimmomatic, SeqPrep gibi yazılımlar kullanılmaktadır.

ii. **Düşük-kalite (opsiyonel):** Düşük kalitedeki bazlar haritalamada ve varyant çağrılmasında yanlış sonuçlara yol açabilir. Dolayısı ile okumaların sonundan ve başından ayrıştırılmaları gerekmektedir. Bu amaçla, CutAdapt, BWA, Trimmomatic, SeqPrep gibi yazılımlar kullanılmaktadır.

Sekonder analizden önce tamamlanması gereken bir basamak da *de-multiplexing* işlemidir. *Multipleksleme*, tek bir YND reaksiyonunda birden fazla hasta numunesinin fiziksel havuzlanması ve eşzamanlı dizilenmesidir. Dizileme analizi için multiplekslenebilecek numunelerin sayısı, analiz türüne (gen paneli, ekzom veya genom gibi), gereken dizilemenin derinliğine, multiplekslenen numunelerin zamanında ve doğru şekilde hazırlanması için manuel ve otomatik prosedürlerin teknik kısıtlılıklarına ve dizileme platformunun iş/zaman oranına bağlıdır. Şu anda klinik laboratuvarlar hat (*lane*) başına ortalama 6 hasta numunesini havuzlayabilmektedir. Bu rakam zaman içinde muhtemelen değişecek ve uygulamaya, platforma ve laboratuvarın birden fazla numuneyi işleme kapasitesine sıkıca bağlı olacaktır^{92; 93}.

Numuneler, bir okuma (parçalanmış genomik DNA'ya eklenen kısa bir dizi olan ve ayrıca barkot olarak ya da multipleksleme kimlik tanıtıcı olarak da bilinen) bir indeks olarak "etiketlenerek" karılır, bu şekilde havuzlama öncesinde dizilemenin hangi hastadan elde edildiği bilinir. Multiplekslemenin çözülmesi, dizileme tamamlandıktan sonra bazı bilgisayar araçları kullanılarak her bir dizi okumasının ayrılarak, barkod yardımıyla doğru hasta numunesi ile ilişkilendirilmesi anlamına gelir. Demultipleksleme sırasında okumaların bir hasta numunesine yanlış şekilde atanmasının önlenmesi önemlidir. Multiplekslemenin güvenilirliği, indeks tasarımıyla ve dizileme analizi öncesinde okumalara eklenmeyi içeren süreçle ilişkili birçok faktörden büyük oranda etkilenir. Multipleksleme & De-multipleksleme optimizasyonu ve validasyonunda şunlar dikkate alınır:

- 1) İndeks tasarımı (baz bileşiminin uzunluğu ve çeşitliliği)
- 2) İndekslerin dizilenecek her bir okumaya eklenme süreci

3) De-multipleksleme için yazılım araçlarının seçimi ve kullanılması

Bu indeksleri analiz etmek için ticari olarak tedarik edilebilen indeksler ve yazılım programları, genelde klinik laboratuvarlarda kullanılır. Kütüphanenin oluşturulması sırasında bir indeksi hastanın numunesine eklemek için kullanılabilen çeşitli yöntemler vardır. Örneğin, bağlama tabanlı (*ligation-based*) yöntem sıralı (*in-line*)-yöntem, indeksin fragmente numune DNA'sına bağlanmış kütüphane adaptörlerine eklenmesi. İndeksler tipik olarak PCR yoluyla bir veya iki tane forward veya reverse dizileme kütüphane adaptörü içine, veya adaptörler ve PCR primerlerinden oluşan bir kombinasyon ile gömülür. Platform tedarikçileri, bu prosedürleri kendi cihazlarıyla kullanmak için uyarlayıp optimize etmişlerdir, ayrıca son kullanıcılara indeks ve protokol de temin ederler. Çalışma grubu, laboratuvarların eğer düşünülen klinik uygulama için optimize edilebilecek ve validasyonu yapılabilecekse, piyasada bulunan ve platform üreticileri tarafından önerilen indeksleri ve protokolleri kullanmasını tavsiye eder.

İlgili uzmanlığa sahip olmayan klinik laboratuvarların kendi indekslerini tasarlamaları çok önerilen bir durum değildir çünkü isteğe göre geliştirilen indeksler, multipleksing/demultipleksing esnasında birbirlerinden ayırt edilebilmelerini garantileyen kapsamlı validasyon süreci gerektirirler. Gerek ticari gerekse isteğe göre uyarlanmış olsun, indeksler klinik kullanımdan önce valide edilmelidir.

İndeks tasarlayan klinik laboratuvarlar, herhangi iki indeks dizisi arasındaki "edit mesafesini" (bir indeks dizisini bir diğer indeks dizisine dönüştürmek için gereken ikamelerin, insersiyon, delesyonların sayısı) göz önünde bulundurması gerekir. Edit mesafesi optimize edilerek dizi kimliği güvenceye alınmalı ve farklı hasta numunelerinden gelen okumaların hatalı atanmasına neden olabilecek hatalar önlenmelidir (buna bazen çapraz kontaminasyon denir). Bu hatalardan kaçınmak için çalışma grubu laboratuvarların aynı reaksiyon/lane içinde birden fazla bazla farklılık gösteren indeksler kullanmalarını tavsiye etmiştir. İndeks baz bileşiminin dengesi (her bir baz konumundaki A,C,G,T karışımı) optimize edilerek bazı teknolojiler açısından kümelerin tespit edilmesine yardımcı olunmalıdır. Dikkat edilmesi gereken bir diğer özellik indeksin uzunluğudur. Altı veya daha fazla baz uzunluğundaki indeksler, daha doğru numune atamasına izin verirken, daha kısa diziler (4 bazlı indeksler gibi) kullanıldığında numune kimliğiyle ilgili bazı problemler rapor edilmiştir. Kütüphane oluşturma sırasında PCR kullanıldığında, daha kısa indeksler daha yüksek replikasyon hatası oluşturma olasılığına sahiptir. PCR sırasında yanlış işlem ayrıca bir indeksin bir diğerine dönüştürülmesine yol açabilir. Bu durum okumaların yanlış hasta numunesi ile ilişkilendirilmesine yol açabilir. Multipleksing/De-multipleksing hataları ise bir hasta numunesinde beklenmedik şekilde (yanlış eşleşmiş) indekslenmiş okumaların düşük seviyeli varlığıyla sonuçlanabilir ve bunlar tipik olarak düşük bir alelik fraksiyonda temsil edilir. Düşük seviyeli yanlış eşleşmiş indeksler genelde germ-hattı otozomal hastalıklar için önemli bir sorun teşkil etmezken, diğer uygulamalar için önemli bir sorun olabilir (örneğin somatik varyasyon, mosaisizm ve mitokondrial heteroplazm testi gibi). Optimizasyon süreci sırasında yanlış eşleşmelere izin vermek, dizileme sırasındaki diğer hata türlerinin ve replikasyonun bir sonucu olarak indeksler arasında ayırımın kaybedilmesi olasılığının anlaşılması açısından bilgilendirici olabilir.

Klinik laboratuvarların yanlış eşleşmiş indekslere sahip okumaları ekarte etmeleri tavsiye edilmiştir. Hiçbir yanlış eşleşmeye izin verilmediğinde kapsamaya katkıda

bulunan okumaların yüzdesinde bir azalma gözlenebilir ancak bu durum kapsamının derinliği artırılarak giderilebilir.

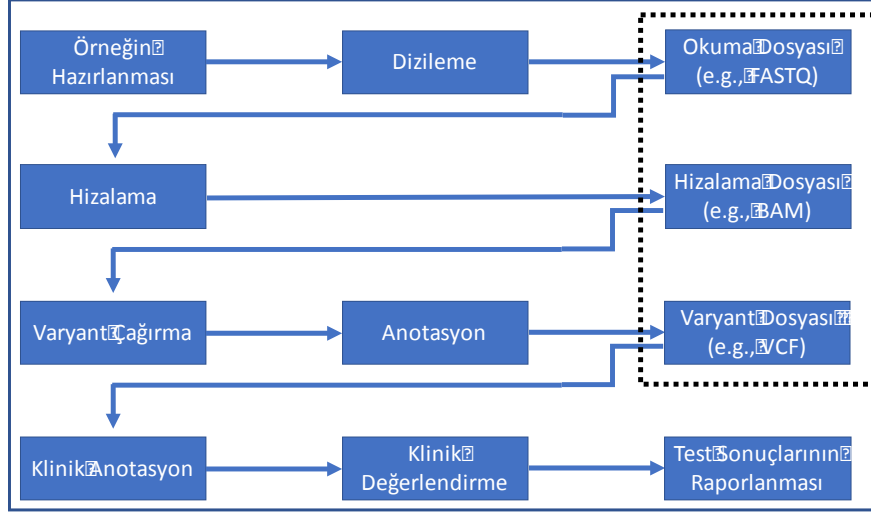
Multiplexing/De-multiplexing'in uygunluk derecesi değerlendirilerek doğrulanmalı, böylece dizi okumalarının kendi hasta numunelerine doğru atanması sağlanmalıdır. Bu işlem bir "uyum/uyuşma ve kontaminasyon" ekranı kullanılarak yapılabilir, bu ekranda hasta numunesi ikiye ayrılır ve hem SNP dizi analizi (veya ortogonal yöntem) hem de multiplexing YND gerçekleştirilir. İki analiz yöntemi arasındaki uyuşma, dizi okumaları indekslenen numuneye doğru atandığında meydana gelir. Uyuşma olmamasının nedeni genelde numune karışıklığı veya kötü tasarlanmış indekstir. Çapraz kontaminasyon hataları da dizileme sırasında meydana gelebilir (nükleotidler yanlış eklenebilir, bir nükleotid yanlış okunabilir) ve dizi okumalarının yanlış atanmasına neden olabilir. Uyuşma ve kontaminasyon ekranı, ikinci bir prosedür olmasına bağlıdır, bu prosedür ayrı numuneden alınan sonuçların karşılaştırılmasını sağlar. Alternatif yöntemler de (numunelerin havuzlar karşısında tek tek dizilmesi gibi) kullanılarak, numune havuzlamanın ve demultiplexlemenin uygunluk derecesi tayin edilebilir. Laboratuvarlar algılanan barkotların sadece teste dâhil edilen barkotlardan geldiğinden emin olmalıdır. Bu yaklaşım sadece hataları ölçmekle kalmaz, ayrıca kullanılan her bir indeks için barkotlama verimliliğini de izler, aynı zamanda barkotlanmamış artefakt dizilerin uzaklaştırılması için kullanılabilir. Beklenen alelik fraksiyondaki veya mitokondrial heteroplazmdaki sapmalar, multiplexing/demultiplexing veya dizi analizinin diğer adımlarındaki hataları işaret edebilir. Kontrol prosedürleri, her hasta numunesi için hata oranının değerlendirilmesini sağlar.

5.3. Sekonder Analiz (Haritalama / Sekans hizalama dahil Sekanların İşlenmesi):

5.3.1. Elde Edilen Okumaların Genoma Hizalanması (BAM file)

Bir hastanın ekzom veya genomunu tekrar inşa etmenin en kolay yolu, üretilen dizi okumalarının bir "referans" genomuna hizalanmasıdır. Bu sürecin ilk kısıtlaması, mevcut insan referans genomlarının halen eksik olması ve hangi optimal referansın (veya birden fazla referansın) kullanılması gerektiği hakkında herhangi bir fikir birliği olmamasıdır. Milyonlarca kısa okumayı verimli şekilde işlemekten geçirmek için birçok hizalama programı geliştirilmiştir. Hizalayıcılar genelde tek baz yanlış eşleştirme varlığında iyi performans gösterir, ancak insersiyon ve delesyonlar için daha fazla uğraş gerekebilir. Hizalama aşamasının sonunda elimizde yüksek bir derinlikte birlikte dizilenen hastadan elde edilmiş ekzon (*exomic*) bölgeleri kataloğu olur. Tipik bir deneyde her bir ekzonik baz, yaklaşık 100 adet bağımsız dizileme reaksiyonuyla temsil edilebilir. Bunun anlamı 100 kat (veya 100x) kapsamadır ve bu seviyedeki bir derinlik hemen hemen tüm hedeflenen bölgelerin iyi kapsamasını garanti etmek ve her bir okumada meydana gelebilecek rastgele dizileme hatalarına karşı koruma sağlamak için gereklidir. Okuma hizalama, kısa DNA dizi okumalarının (genelde 50-400 baz çifti) genom boyunca referans dizisine göre doğru şekilde konumlandırılmasını içerir. Okumaları hizalamak için doğruluk ve işleme hızında farklılık gösteren çeşitli algoritmalar geliştirilmiştir. Beklenen varyasyonların tipine bağlı olarak laboratuvar, verilere uygulanacak bir veya daha fazla okuma hizalama aracı tercih etmelidir. Okuma hizalama için ticari olarak temin edilen veya kullanıma açık kaynak araçları kullanılabilir. Bu araçlar çeşitli hizalama algoritmaları kullanır ve belli tipte veriler için daha verimli olabilir. İnsan referans genomu, *Genom Referans Konsorsiyumu* (GRC) tarafından oluşturulmuştur ve bu referans genomunun direkt

GenBank FTP sitesinden alınması önerilir. Referans genom için tek bir kaynak kullanılması, değişik anotasyonlar veya koordinat sistemleri kullanabilen farklı referans genomlarına göre varyant tespit ederken, verilerin paylaşılmasıyla veya sonuçların iletilmesiyle ilgili problemlerin en aza indirilmesine yardımcı olur.



Şekil 6. Analizler sonrası üretilen dosya çeşitleri. J Mol Diagn (2017) 19:417e426'dan⁹⁴ değiştirilerek alınmıştır.

Özgüllüğü iyileştirmek amacıyla panel, ekzom veya genom dizilemeden gelen okumaların hedeflenen diziler yerine komple insan genom referansına hizalanmasını tavsiye edilmektedir. Bu uygulama homolog bölgelerin varlığı nedeniyle okumanın yanlış haritalandırılma olasılığını azaltır (ama her zaman ortadan kaldırmaz) (örneğin psödogenler, paraloglar ve duplikasyonlar gibi). Bazı durumlarda diziler referans genomdan eksiktir ve bu durum hizalanmayan ya da yanlış hizalanan okumalara yol açabilir. Diziler referans genomdan birkaç neden yüzünden eksik olabilir, numune hazırlığıyla ve dizilemeyle ilgili teknik konular ve kromozom referansında alternatif alellerin olmaması gibi. Yüksek allelik çeşitliliğe sahip bölgelerde (örneğin doku uyumu kompleksi gibi) birden fazla alel vardır. Bunlar referansta alternatifler olarak gösterilir ve insan çeşitliliğini modellemek açısından kullanışlıdır. GRCh37 her ne kadar en az bir adet alternatif haplotiple birlikte üç bölgeye sahipse de, GRCh38 >170 alternatif alel içerir. Alternatif alellerin eklenmesi, birden fazla haplotipin daha iyi gösterilmesine olanak verir, ancak bu aleller, alternatif haplotipleri segmental duplikasyon gibi biyolojik olaylardan ayırt edemeyen modern analiz işlemleri için bir güçlük teşkil eder. Bu ise nihayetinde yeni ve güncellenmiş yazılım araçlarına olan ihtiyacı ortaya çıkarır.

Bu aşamada RNA ve DNA sekansları için ayrı hizalama araçları kullanılmaktadır. Sıklıkla BWA, Noalign, Stampy, SOAP2, LifeScope, Bowtie gibi platformlar kullanılmaktadır. Bu işlem sonucunda çıktı olarak BAM dosyası oluşmaktadır (**Şekil 6, Şekil 7, Şekil 8**). (Bilgi Kutusu 2)

Transkriptler içinde konum atama için genelde HGVS kuralları kullanılır. HGVS, koordinatları transkriptin 5' → 3' yönüne göre atar (5' ile 3' yönü DNA'nın şeker omurgasındaki karbon sayılarını gösterir). Bu durum genomun yönünden bağımsızdır (genomun + veya - ipliği). HGVS, bir insersiyon veya duplikasyon için transkript konumunun en 3' (sağ) baz konumu olması gerektiğini belirtir⁹⁵⁻⁹⁷. HGVS gelişen bir

standart olmaya devam etmektedir ve deęişen kurallar nedeniyle net olmayan tanımlar ortaya çıkabilmektedir. Adlandırmanın yanlış uygulanmasına ait raporlar da mevcuttur, bu durum ilave karmaşıklıęa neden olmuştur.^{95; 98} Varyantların, yayınlanan kuralları takip eden HGVS tanımları kullanılarak tanımlanmasını ve kısaltılmış HGVS tanımlarının tam HGVS tanımıyla bağlantılı olması gerektięi tavsiye edilmektedir ^{13; 89; 95; 99}. Pozisyon karmaşıklıęı olasılıęını en aza indirmek için diziler, net genom koordinatlarıyla bağlantılı şekilde raporlanmalıdır.

Birçok haritalama ve hizalama algoritması hassasiyet, özgüllük ve hız arasında bir denge kurmaya çalışır. Bu nedenle pek çok hizalayıcı bir okumayı referans içinde doğru bölgeye yerleştirebilir, ancak standart altı hizalama da oluşacaktır. Bu durum lokal yeniden hizalama ile düzeltilebilir, bu özellik pek çok informatik işleyişte ortak olan bir özelliktir. Okuma haritalandırmanın uygunluk derecesini tanımlayan en önemli metrik, haritalandırma kalite skorudur (*mapping quality score*). Bu skor, bir okumanın yanlış yerleştirilme olasılıęını tahmin eder ve birçok parametreye dayanır. Bu haritalandırma skoru genelde bir BAM dosyasında saklanır ve varyant çağırma algoritmaları tarafından tanınır. Farklı algoritmalarından gelen haritalama skorları birbirine benzer deęildir ve yazılım ise düşük kaliteli haritalama skorlarının nasıl ele alındıęına baęlı olarak deęişiklik gösterir. Örneęin eęer bir okuma referans genom içinde birden fazla lokasyona eęit derecede iyi haritalanıyorsa, bazı eşleştirciler bunu atacak, bazıları rastgele yerleştirecek, bazıları birden fazla konuma yerleştirecek ve bazıları ise en yüksek haritalama kalite skoruyla konuma haritalayacaktır. Bu davranış daha sonra yapılacak olan varyant çağırmaı belirgin şekilde etkileyebilir ve testin optimizasyonu ile klinik validasyonu sırasında tipik olarak hesaba katılır. Hizalamadaki hatalar ayrıca varyant çağırılarının doğruluęunun deęerlendirilmesiyle de ölçülebilir. Haritalama kalitesi bilgilerini kullanan varyant çağırıcıların kullanılması faydalı olabilir: örneęin, düşük haritalama skorlarına sahip okumaların altında bir eşięe sahip varyant çağırıcılar varyant çağırmaı kullanılmaz. Çeşitli haritalama ve hizalama algoritmaları geliştirilerek NGS yazılım paketleri içine eklenmiştir. Bu hizalama araçları, farklı varyant türlerini tespit etmek için farklı hassasiyet seviyelerine ayarlanabilir. İlgili algoritmalar ve yazılım paketleri önceki bir çok yayında anlatılmış ve karşılaştırılmıştır ¹⁰⁰⁻¹⁰⁵. Bugün sıklıkla kullanılan araçlar için iki temel hizalama teknięi benimsenmiştir: “anahtarlı tablo tabanlı uygulamalar” (*hash table-based implementations*) ve “Burrows–Wheeler dönüşüm (BWT) tabanlı yöntemler” ¹⁰⁴. “Anahtarlı tablo tabanlı algoritmalar dizi verilerini indeksleyip tarayarak okumaların referans genom dizisinde hızlı aranmasını ve yerleştirilmesini kolaylaştırır” ¹⁰⁴. Bu araçlar bir veri yapısı inşa ederek çalışır (veya anahtarlı/hash tablo) ve genelde ya okumalarda ya da referans genom dizisinde bulunan kısa oligomerlerin bir indeksidir (ayrıca kaynak da ya da *seed* de denir). (e.g. MAQ ¹⁰⁶). Bu tablo, kısa kaynak dizilerini paylaşan referansta ve hizalanacak okumada lokasyon bularak aday haritalama konumlarını tespit eder. Aday haritalama konumları sonrasında deęerlendirilerek kesin hizalama belirlenir. Örneęin BFAST ¹⁰⁷, NovoAlign (<http://www.novocraft.com>, accessed August 18, 2014), MOSAIK (<https://wiki.gacrc.uga.edu/wiki/MOSAIK>, accessed August 18, 2014), ve Isaac (<http://bioinformatics.oxfordjournals.org/content/early/2013/06/04/bioinformatics.btt314>, accessed August 18, 2014) gibi yazılım araçları, anahtarlı tablo tabanlı algoritmalar kullanır ve şunlara göre farklılık gösterirler: kaynaęın uzunluęu, ilk

haritalamada izin verilen yanlış eşleşmelerin sayısı, kaynak uzantısının türü, bellek gereksinimleri, hız ve doğruluk.

Okumaları hizalamak için kısa oligomerlerden oluşan bir tablo yerine *Burrows–Wheeler Dönüşüm* (BWT) tabanlı yöntemler, dizi eşleştirme yaklaşımı kullanarak, referans genoma ait bir indeks oluşturarak hızlı arama yapmayı kolaylaştırır ^{104; 106}. Bu yöntem genom konumlarını okuma için iyi bir eş olarak hızlı şekilde tespit eder, sonrasında anahtarlı tablo tabanlı yöntemle benzer şekilde bu adayları tam anlamıyla değerlendirerek okumaları spesifik konumlarına yerleştirir. BWT tabanlı yöntemlerin uygulanması daha az zaman alır ve bu yöntemler, anahtarlı tablolara dayanan pek çok yöntemle göre bellek açısından daha verimlidir ^{104; 106}. Kısa okuma hizalama programlarına örnek olarak Bowtie 2 ¹⁰⁸, BWA ¹⁰⁹, TMap ¹⁰⁵ ve SOAP2 ¹¹⁰ verilebilir.

Hizalama algoritmasının ve stratejisinin seçilmesinde YND uygulaması (tespit edilecek varyantların sınıfı [kısa eklenme, SV, SNV vb.], kullanılan yeni nesil platform, analizin tüm genomu mu yoksa hedeflenen bölgeleri mi kapsadığı gibi) ve laboratuvarın bilgisayarla hesaplama kapasiteleri (yüksek performanslı küme (*cluster*) ortamının olup olmadığı gibi) dikkate alınmalıdır. Her bir hizalayıcı, farklı sınıftan varyantların optimal tespiti için özgün ayarlar gerektirecektir. Bazı hizalayıcılar belli bir platformla birlikte kullanılmaları için optimize edilirler, örneğin TMAP ¹⁰⁵, Ion Torrent verilerini referans bir genoma haritalamak için özel tasarlanmıştır. BLASR ¹¹¹ gibi diğer hizalayıcılar, platforma özgü hataları değerlendirme kapasitesine sahiptir, örneğin INDEL dizileme hatalarında meydana gelebilecek bir artış gibi. Belli hizalayıcılar optimize edilerek özel tipte dizi varyantlarını tespit ederler, örneğin kısa insersiyonlar veya delesyonlar, SNV veya CNV gibi, ya da daha kesin lokal hizalama sağlamakla birlikte, iş akışında ilk adım olarak okuma yerleştirme için çok fazla kaynak yoğunlukludurlar. Bir informatik akışta farklı algılama kapasitelerine sahip hizalayıcıları (paralel veya seri şekilde) bir araya getirerek laboratuvarlar, tek bir hizalayıcı kullanılarak elde edilenden çok daha fazla ve çeşitli varyantı tespit edebilen testler tasarlayabilirler ¹¹². Buna göre çalışma grubu klinik laboratuvarların bir hizalayıcı kombinasyonunu veya aynı hizalayıcıyı farklı ayarlarla deneyip, hedeflenen varyant türlerini etkin şekilde tespit edip etmediklerini değerlendirmelerini tavsiye etmiştir (örneğin SNV ve CNV tespiti için). Alternatif olarak kullanıcılar, belli paneller ve uygulamalar için önceden optimize edilmiş komple iş akışı içeren yazılım paketlerini de tercih edebilir. Örneğin, hizalama için ve indellerin tespit edilmesinde hizalama için kullanılabilecek eşige kıyasla SNV'leri çağırma amacıyla kabul edilebilir haritalama kalite skorlarıyla ilgili daha sıkı bir kalite eşığı kullanılabilir. Pek çok aracın altta aynı algoritmayı kullandığını anlamak önemlidir, bu nedenle farklı yazılım paketleri aynı temel zayıf ve güçlü yönler sahip olabilir^{103; 112; 113}. Laboratuvarlar testleri optimize etmeli ve bu sayede yanlış-negatif çağrı sayısını en aza indirirken, aşırı sayıda yanlış pozitif çağrıları da engellemelidirler. Bu ise çok çeşitli varyant tipleri içeren karakterize referans malzemeler kullanılarak deneylerin geliştirilerek optimize edilmesi yoluyla başarılabilir (örneğin SNV, büyük ve küçük indeller gibi).

Laboratuvarların başlangıçta yazılımın varsayılan ayarlarını kullanmaları ve sadece kendi klinik uygulamaları için uygun olduğunda bu ayarları valide ederek değiştirmeleri tavsiye edilmektedir. Değiştirilebilecek yazılım ayarlarının örnekleri şunlardır: düşük kaliteli dizi kırpması (*low quality sequence trimming*), izin verilen hatalı eşleşme sayısı (*number of allowed mismatches*), izin verilen boşluk açıklığı ve boşluk uzantısı, okumalar için minimum haritalanabilme (*mappability for reads*).

İstenen sonuçların elde edilmesini ve hizalama sürecinin diğer öğelerinin riske girmemesini garantilemek için yapılan değişiklikler tekrar validasyonu gerektirir. Bu nedenle varsayılan ayarlarda yapılan değişikliklerin ve sonraki değerlendirmenin gereken bilgiye sahip bir informatik uzmanıyla birlikte yapılması tavsiye edilmektedir. Bu adımlar test geçirme (validasyonu) öncesinde gerçekleştirilir ve hasta testlerinde kullanılacak yazılımın son ayarlarının dokümanede edilmesini sağlar.

Hizalama yazılımı seçilirken ve optimize edilirken, hizalamanın kalitesini garantilemek için birkaç faktör göz önünde bulundurulmalıdır. Bunlar arasında çeşitli YND platformlarının farklı hata profilleri, genel ve varyant tipine özgü hata oranı (örneğin yanlış eşleştirmelere vs. boşluklar, tekrar eden bölge hataları) ve cihazın ortalama okuma uzunluğu yer alır. Hizalama hataları, haritalama kalite skorunu, sonraki varyant çağrılarının transisyon/transversiyon oranını (Ti/Tv) ve sonraki varyant çağrılarında sinonim – sinonim olmayan değişikliklerin oranını içeren kalite metriklerinin değerlendirilmesiyle tespit edilebilir. Her bir hizalama algoritması, özgün bir haritalama kalite skoru yaratır (ki bu skor algoritmalar arasında karşılaştırılabilir değildir) ve okumaları yukarıda açıklandığı gibi farklı şekillerde atar (haritalama yapmak gibi). Ti/Tv oranı, genel kalite göstergesi görevi görür çünkü belirli hedef bir bölge için (gen paneli, ekzom veya genom gibi) yaklaşık olarak sabittir¹⁷. Ayrıca hataların bilgisayara bağlı olmayan kaynakları da vardır. Örneğin numunenin kalitesi ve gen paneli için kullanılan hedef zenginleştirme sürecinin uygunluk derecesi ve ekzom dizileme, üretilen dizinin kalitesini ve sonraki hizalamayı etkileyebilir, bu durumlar ise okumaların yerleştirilmesini (*read mapping*) etkileyebilir. Bunun nedeni, hizalama için uygun toplam dizi yüzdesini değiştiren düşük kaliteli baz çağrılarının olması ve ayrıca hedef üzeri ve hedef dışı hizalamaların oranıdır. Sonuç olarak hedef üzeri –on-target- / hedef dışı oranı-off target-, aynı deneyin farklı bölümlerindeki yakalama kalitesinin bir ölçüsüdür. Yüksek hedef üzeri oranlar direk nihai dizi çağrılarının uygunluk derecesiyle ilgilidir.

Homolog diziler, özellikle eğer YND tarafından dizi okumasının ortalama uzunluğundan daha uzun olursa, optimum hizalama açısından zorluk yaratırlar¹⁷. Çift uçlu dizileme veya uzun okumalar yardımcı olmakla birlikte, homolog bölge kütüphane fragman büyüklüğünden çok daha büyükse artık etkili olmazlar. Hedef bölgedeki kapsama, ilgilenilen bölgeye %100 uyan uzun diziler genomun başka bir yerinde mevcut olduğunda azalacaktır. Bu durum varyantların kaçırılmasına neden olabilir. Hizalayıcıların okumaları net olmayan şekilde haritalayamadığı durumda, %100'den daha az özdeş olan ama bir homolog seviyesini aşan diziler, yanlış-pozitif varyant çağrılarını vermeye meyilli olacaktır (yani, varyant çağrısı ilgilenilen gendeki fiili bir değişiklikten çok gen ile psödogen arasındaki farkı yansıtır). Ekzom ve genom dizileme ile çoğu gen paneli için, genomun hedeflenen bölgesine optimum hizalamayı sağlayan yöntemler geliştirilmedikçe, hedef dışı olasılığını ve zorla oluşan hizalamaları en aza indirmek için okumaların sadece hedef bölgeye değil referans genoma hizalanmaları tavsiye edilmektedir. Bu durum, PCR yakalama için hizalamanın uygunluk derecesinin optimize edilmesini ve geçerli kılınmasını şart koşan yöntemler kullanıldığında bazı gen panelleri ve alt-ekzom/genom analizleri için gerekli olmayabilir. Bazı rahatsızlıklar için kapsamlı testler, başka bir bölgeye yüksek dizi homoloğu olan belli genleri içermek zorundadır. Buna verilecek bir örnek nonsendromik işitme kaybındaki en önemli genlerden olan stereosilin (STRC) genidir¹¹⁴. NGS potansiyel olarak evrensel bir teknoloji platformu olup, pek çok farklı

varyant türünü sorgulamakta ve tanısal testleri bir araya getirmekte kullanılabilir. Homolog genler öngörülebilir bir gelecek boyunca karmaşık bir konu olarak kalmaya devam edecektir. Bazı durumlarda Sanger gibi alternatif bir yöntem kullanmak gerekli olabilir. Sanger, lokasyona özel primerler veya problemler kullanarak sadece hedef bölgenin amplifikasyonunu veya dizilenmesini sağlar.

5.3.1.1. Haritalama Sonrası Okuma Kalite Kriterleri

Okumaların kalite kriterlerine göre elenmesi ve trimlenmesi sonucu mapping işlemine alınır. Bu aşamadan sonra Readlerin büyük bir kısmının referans genom üzerindeki lokasyonları belirlenmiş olur. Bu pozisyon değerleri üzerine bakılması gereken kriterler aşağıda kısaca açıklanmıştır.

a. Haritalama oranı (*Mapped*) oranı

Mapping algoritmasına bağlı olarak readler map edildiği halde bazı readler unmapped olarak kalırlar. Map edilenlerin oranı önemli bir parametre olabilmektedir. Bu sayının insan genomu dizilemesi için %98 ve üzeri olması iyi bir sonuçtur. Bu sayının çok düşük olması laboratuvar'da potansiyel bir kontaminasyonu işaret edebilmektedir. Bu konuda lab.'a geri bildirim yapılması gerekebilir. Bu sorunun kaynağının araştırılmasında fayda vardır.

b. Hedeflenen bölgenin okuma derinliği

TGD dışında (TED ve targeted seq. gibi) dizileme çalışmalarında probe'ların hedeflediği bir bölge vardır. Laboratuvarda kullanılan yakalama kitinin hedeflediği bölgeye düşen okumalar esas alınarak ortalama derinlik yeniden hesaplanır. Bu değer laboratuvarın işlemi ne kadar iyi ve efektif yaptığını göstergesidir. Eğer bu işlemde istenilenden daha çok okuma kaybı yaşıyorsa laboratuvara geri bildirim yapılmalıdır. Bu koşul altında örneğin TED için ortalama derinlik *de facto* standart olarak 70x üzerinde olmasında fayda vardır.

c. Hedeflenen bölgedeki derinliklerin dağılımı

Hedef bölgedeki tüm pozisyonların derinliklerin dağılımı çıkarılmalıdır. Bu dağılımlara bakılarak bu dağılımın mümkün olduğunca eşit olması beklenmektedir. Bu kriterin sebebi yakalama aşamasında bazı durumlarda hedef bölgenin belirli bölgesine bir ön yargı (*capture bias*) oluşmaktadır. Örneğin bazı bölgeler 1000X dizileme yapılmış ancak bazı bölgeler ise 4X'in altında dizilenmiş ise birçok varyant/mutasyonun tespiti özellikle heterozigot olanlarının tespiti oldukça güç olacaktır.

d. 10 derinlik üzerinde okunan hedef bölgelerin yüzdesi

Bu kriter bir önceki kriterin özel ve kolay kullanılabilir bir halidir. Hedeflenen bölge içerisinde 10X derinliği üzerinde dizileme yapılmış bölgelerin tüm hedef bölgeye oranıdır. TGD için %90 olması ve WES için %95 ve hedefli dizileme için %99 olması tercih edilmektedir.

e. Ortalama mapping quality değeri

Mapping algoritmasının ürettiği mapping quality değeridir. Bu ortalama değer ilgili algoritmanın tanımladığı standartların en az 2 olmasında fayda vardır.

f. Kromozom uzunlukları veri üzerinden hesaplanan oranı

Hedeflenen bölgeler esas alınarak kromozom uzunluklarının tüm referans genoma oranları bulunur. Aynı oran teorik olarak da bulunur. Bu değerlerin birbirinden sapmasının oldukça küçük olması beklenir. Aksi durumda bir kontaminasyon veya daha büyük bir tıbbi anomaliden şüphelenilebilir. Bu durumun sebepleri belirlenmelidir.

g. Fragment (insert) boylarının dağılımı

Özellikle pair-end ve nispeten kısa dizileme yapan teknolojilerde fragment boylarının dağılımının çıkarılması gerekmektedir. Özellikle uzun INDEL ve yapısal varyasyon'ların tespit edilebilmesi için bu dağılımın standart sapmasının oldukça düşük olması gerekmektedir. Örneğin ortalama 350 uzunluğunda fragman uzunluğu için 50 veya daha düşük değerde standart sapma olması yapısal varyantların biyoinformatik algoritmalar kullanılarak bulunması açısından başarıyı artıran bir durumdur. Tekrarlamak gerekirse uzun ve tek-uçlu okumalarda bu kriterin bir önemi yoktur. Ayrıca amplikon sekanslamada için bu durumun söz konusu olmayacağını belirtmek gerekir.

5.3.2. Hizalama sonrası işlemler

Ek işlemler genelde varyant çağrılmadan önce yapılarak, ilk haritalama ve hizalamadaki hatalar azaltılır ve hizalama optimize edilir. Bu adımlar arasında lokal tekrar hizalama ve bazı varyant çağırıcılar ile zenginleştirme yöntemleri için polimeraz zincir reaksiyonu (PCR) duplikasyonlarının uzaklaştırılması ve baz çağrısı skorlarının tekrar kalibrasyonu yer alır ¹¹⁵⁻¹¹⁷.

5.3.2.1 Duplike okumaların filtrelenmesi/uzaklaştırılması

Shotgun (kısa parçalı) dizilemede DNA'nın rastgele kesilmesi sebebiyle bir kaç duplikasyon beklenmekte ancak duplikasyonlar PCR esnasında görüntülemenin artefaktı olarak da ortaya çıkabilmektedir. Amplikon temelli dizilemede duplikasyonlar beklenmektedir ve uzaklaştırılmamalıdır. Bu aşama için Picard MarkDuplicates yazılımı kullanılmaktadır.

Bilgi Kutusu 2: Sequence alignment/map (sam/bam) format

Referans genom tarafından hizalanmış ham sekans verisinin (örn. fastq) saklandığı sekme-ayraç- düz-yazı dosya formatıdır (<https://samtools.github.io/hts-specs/SAMv1.pdf>). Hizalama yazılımları tarafından üretilir. 8 sütundan oluşan bir tablo şeklindedir. Sıkıştırılmış (binary) SAM dosyaları, BAM olarak saklanır.

Şekil 7. Örnek bir SAM dosya formatı

```
@HD VN:1.5 SO:coordinate
@SQ SN:ref LN:45
r001 99 ref 7 30 8M2I4M1D3M = 37 39 TTAGATAAAGGATACTG *
r002 0 ref 9 30 3S6M1P1I4M * 0 0 AAAAGATAAGGATA *
r003 0 ref 9 30 5S6M * 0 0 GCCTAAGCTAA * SA:Z:ref,29,-,6H5M,17,0;
r004 0 ref 16 30 6M14N5M * 0 0 ATAGCTTCAGC *
r003 2064 ref 29 17 6H5M * 0 0 TAGGC * SA:Z:ref,9,+,5S6M,30,1;
r001 147 ref 37 30 9M = 7 -39 CAGCGGCAT * NM:i:1
```

Şekil 8. SAM dosyasının sütunlarının açıklamaları

Col	Field	Type	Regex/Range	Brief description
1	QNAME	String	[!-?A-Z]{1,254}	Query template NAME
2	FLAG	Int	[0,2 ¹⁶ -1]	bitwise FLAG
3	RNAME	String	\ [!-()+-<>-~][!-~]*	Reference sequence NAME
4	POS	Int	[0,2 ³¹ -1]	1-based leftmost mapping POSITION
5	MAPQ	Int	[0,2 ⁸ -1]	MAPping Quality
6	CIGAR	String	\ [0-9]+[MIDNSHPX=] \+	CIGAR string
7	RNEXT	String	\ = [!-()+-<>-~][!-~]*	Ref. name of the mate/next read
8	PNEXT	Int	[0,2 ³¹ -1]	Position of the mate/next read
9	TLEN	Int	[-2 ³¹ +1,2 ³¹ -1]	observed Template LENgth
10	SEQ	String	\ [A-Za-z=.\+]	segment SEQUENCE

5.3.2.2 INDEL'lerin Hizalanması

Dizilenen örnekte Indel varlığında bu bölgenin etrafında genellikle çoklu baz yanlış eşleşmesi meydana gelmektedir (Özellikle de okumaların başına ve sonuna denk geldiği durumlarda). Bu artefaktların azaltılması için local yeniden hizalama metodları ile bu çağrılardaki yanlış pozitiflik oranı azalmaktadır. Bu amaçla GATK Realigner Target Creator & Indel Realigner ve SRMA yazılımları kullanılmaktadır.

5.3.2.3 Baz Kalite Skoru Yeniden Kalibrasyonu

Okumaların kalite skoru referans genoma haritalama aşamasından sonra tekrar kalibre edildiğinde yanlış baz çağrılmaları saptanabilir. Çoğu algoritmada yanlış baz çağruları, basit baz çağrılması veya bilinen polimorfizm veritabanları kullanılarak gerçek varyantlardan ayrılabilir. Bu aşama için GATK Realigner Target Creator & Indel Realigner ve ReQON gibi yazılımlar kullanılmaktadır.

Bilgi Kutusu 3

İkincil analiz, okumaları referans dizilemeye hizalama ve varyant arama/çağruları oluşturma sürecinden oluşur. Birden fazla hasta numunesi havuzlanmışsa (pooled) (her bir numune kimlik teşhisi için ayrı ayrı etiketlenir) veya dizileme işlemi sırasında multipleks'lenmiş ise, ortaya çıkan veriler için de-multipleksleme yapılmalı ve analiz öncesi hasta dizisi ayrılmalıdır. Bunun ardından her bir okumanın haritalanması ve referans genomuna hizalanması gelir. Okumalar pek çok nedenden dolayı referans genomuna özgün şekilde hizalanamayabilir. Örneğin okumalar genomun son derece homoloji gösteren bölgelerinden elde edilmişse birden fazla lokasyona hizalanabilir. Referans genom içinde temsil edilmeyen bölgelerden (loci) elde edilen okumalar (örneğin belli HLA alelleri gibi) HLA bölgesine haritalanabilir ancak hizalanamazlar. Alternatif olarak sonuçta yanlış-pozitif olabilecek bir veya daha fazla hedef dışı lokasyona hizalanabilirler. Eğer okuma referans genomda kullanılan diziden farklı bir haplotipten elde edilmişse, aşırı alelik çeşitlilik gösteren bölgeler de (HLA gibi) hizalanmayı güçleştirebilir. Bu nedenle GRC (<http://genomereference.org>), verinin yeterli olduğu durumda aşırı çeşitlilik içeren bölgeler için referansta alternatif dizi bilgileri içerir. Bu yaklaşım alelik çeşitliliği beraberinde getirir ve şu anda en sık kullanılan analiz pipeline'ların çoğu, alelik duplikasyonu paralog duplikasyondan ayırt edememektedir. Bununla birlikte daha eksiksiz bir referans genomunun okuma hizalamasını geliştirebileceği net olarak görülmektedir. Küçük insersiyon ve delesyon varyantlarının (indeller) sağlıklı şekilde tespiti bir hayli zordur. Tutarsız hizalamalar yanlış SNV çağrularına veya ilk okuma hizalaması içinde indelin yanlış çağrılmasına neden olabilir. Genom Analiz Araç Kiti (GATK) gibi yaygın şekilde kullanılan pek çok araç, potansiyel indelleri içeren bölgelerde lokal tekrar hizalama yapmak için işlem sonrası adımlar gerçekleştirir. Ayrıca bir hastanın okumalarının de-novo birleştirilmesi, standart hizalama protokolleri büyük insersiyon/delesyonları tespit edemediğinde ya da indeller mevcut hiçbir referansı temsil etmediğinde yardımcı olabilir⁹. Bu kopya sayısı değişikliklerini tespit etmek zordur ve referans genom dizisinin tek kopyalı bölgelerine karşı normalizasyon gerektirir.

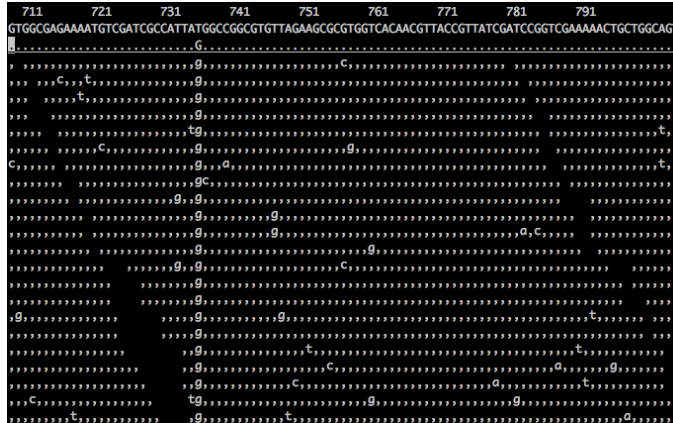
5.3.3. Varyant Arama (Variant Calling- VCF file):

Referans Genomdan farklılıkların tespit edildiği ve genotiplemenin yapıldığı aşamadır (**Bilgi Kutusu 3**). Varyant çağırma süreci ekzom başına 20,000 ile 100,000 arasında varyant ve tüm genom dizileme için yaklaşık 3-4 milyon varyant üretir. Bu varyantları en az %90'ı tipik olarak SNV'dir^{118; 119}. Genotip çağırıcılar, sonuçları bir varyant çağrı dosyasına gönderir¹²⁰. Bu dosyalar, tespit edilen dizi varyantları hakkındaki bilgileri ve anotasyonları kaydeder (örneğin SNV, indel, vb.). Bu dosyaların çoğu varyantın yapısal ve işlevsel sonuçları gibi ilave bilgileri eklemeyi ve referans diziyi veya *no-call*'ları kaydetme kapasitesini destekler. Varyant tespiti için pek çok yazılım aracı kullanılmakta olup başka yayınlarda anlatılmış ve karşılaştırılmıştır¹¹⁵⁻¹¹⁷ (**Şekil 9**). (**Bilgi Kutusu 4, Bilgi Kutusu 5**) YND biyo-informatik pipeline'ları, SNV'leri, küçük insersiyon ve delesyonları ve ayrıca daha büyük yapısal varyasyonları çağırabilir (daha büyük insersiyonlar, kromozomlar arası translokasyon ve kopya sayısı varyantları gibi). 2015'te hiçbir yazılım aracı bu varyant sınıflarını eşit bir doğrulukla teşhis edebilmiş değildir .

Bilgi Kutusu 4: Samtools v1.3.1

İçerisinde birçok analiz aracı bulunduran bir araç setidir. Variant calling için mpileup aracı kullanılır (<http://samtools.sourceforge.net/mpileup.shtml>). Analiz çıktısı VCF formatında saklanır. Tview aracı ile hizalanmış veri ve referans genom arasındaki farklılıklar görselleştirilebilir. Ek olarak, VCF formatındaki verinin görselleştirilmesi için IGV (Integrative Genomics Viewer, Broad Institute) gibi araçlar da kullanılabilir.

Şekil 9. Samtools tutorial'ında (http://biobits.org/samtools_primer.html) yer alan hizalanmış ve referans genom arasındaki variant'ların tview aracı ile görselleştirilmiş halidir.



Bilgi Kutusu 5: Varyant Çağırma (Varyantların tespit edilmesi)

Ham verinin kalite çalışmalarını geçen ve referans genoma hizalanan kısmından, referans genoma göre farklı olan varyantlar tespit edilir. Bu varyasyonlar tek nükleotid varyantları (SNVs), çoklu-nükleotid varyantları, küçük insersiyon veya delesyonlar (InDels, genellikle 50bp'den küçüktürler) ², büyük yapısal varyantlar (SVs) ve kopya sayısı farklılıkları olabilir. Günümüzde Bayes istatistik modelleri temelli olasılık metodları, örneğin SAM- tools ⁸ ve Genome Analyzer Tool Kit ¹⁶ deneysel yaklaşımlara göre, dizileme hatalarına daha doğru istatistiksel modelleme sağlamaları nedeniyle çok daha sık olarak kullanılmaktadır ¹⁷. Bununla birlikte söylemek gerekir ki bu araçların doğruluğu, dizilemede elde edilen ortalama derinliğe ve ilgilenilen varyant tipine (SNV'ler InDel'lere göre daha doğru tespit edilebilir) bağlıdır ve oldukça değişkendir. Özellikle, sensitivite (ekzomdaki tüm doğru varyantların tespit edilmesi) dizileme coverage'ına kullanılan yazılımdan çok daha bağımlıdır. 100-150 ortalama ekzom coverage'ı, bilinen tüm kodlayan ekzomların %95'inin en az 20 kez veya daha fazla derinlikte okunduğunu göstermektedir ki bu da kabaca ekzom seviyesinde sensitiviteyi %95 yapmaktadır. Selektivite veya bulunan varyantın gerçek bir varyant olup olmadığının olasılığı daha çok kullanılan yazılıma bağlıdır Genome Analyzer Tool Kit ¹⁹. Değerlendirme metoduna bağlı olarak, SNV'lerin tespit edilmesinde ekzom dizilemenin selektivitesi 99 ve 99.7% olarak tespit edilmiştir. InDel'ler için selektivite daha düşüktür, ortalama %80 civarındadır ¹⁹. Hernekadar analiz metodları gelişmiş olsa da yanlış pozitif olan varyantların tespit edilmesinde verinin deneyimli kişilerce Integrated Genomics Viewer ²¹ gibi araçların kullanılarak incelenmesi önerilmektedir.

Sonuç olarak, düşünülen klinik uygulamanın hedeflediği varyantların spektrumunu tespit edebilecek ayarlar kombinasyonu ve/veya yazılımları belirlemek amacıyla birden fazla varyant çağırıcının değerlendirilmesi gerektiği, bir varyant çağırıcının platform için uygunluğunun göz önünde bulundurulması gerektiği tavsiye edilmektedir ^{19; 101; 121}. Örneğin bazı laboratuvarlar verileri, SNV tespiti için özelleştirilmiş varyant çağırma yazılımı kullanarak analiz eder ve bunu indelleri çağırma için optimize edilmiş yazılım kullanılarak yapılan ayrı bir analiz takip eder. Bu ayrı analizlerden elde edilen sonuçlar sonrasında bir araya getirilebilir. Ancak birleştirilmiş pipelineler ile birlikte gelen yüksek biyo-informatik/IT genel masrafları ortaya çıkar, çoklu yaklaşımlar arasındaki uyumsuzluk ise çözülmek zorundadır ¹²². Bazı durumlarda farklı analizler arasındaki uyumsuzlukları gidermek için farklı ama yerleşik teknolojilerden faydalanan alternatif yöntemler (kantitatif PCR veya Sanger dizileme gibi) gerekebilir.

Hatalar hem hizalama hem de varyant çağırma sırasında ortaya çıkabilir. Bu iki süreç birbiriyle sıkı sıkıya bağlantılıdır; bu nedenle varyant çağırma yazılımı hizalama stratejisine uygun şekilde optimize edilmelidir. Belirli uygulamalar için optimize haritalama, hizalama ve varyant çağırma sunan yazılım paketleri uygun olabilir. Klinik laboratuvarlarda bu değerlendirme, araştırılmakta olan varyant türleri için (SNV, indeller, vb.) varyant çağırmanın hassasiyeti ve özgüllüğü üzerine odaklanmalıdır.

Çeşitli sistematik hatalar dizileme, haritalama ve hizalama sırasında oluşabilir, dolayısıyla pek çok varyant çağırıcı bu süreçlerle ilgili metrikleri değerlendirir ve belirlenen kriterleri karşılamayan varyantları filtreler¹²³. Bu filtreler genelde yüksek güven derecesine sahip varyantları teşhis etmek için diğer parametrelere ek olarak baz çağırma kalite skorunu, haritalama kalite skorunu, kapsamayı, strand bias tahminini ve alelik yüzdesini (alelik fraksiyonu veya zigotluğu) kullanır. Bu metriklerin çoğuna ait kabul kriterleri genelde uygulamaya özgü olmakla birlikte genel ilkeler geçerlidir. Örneğin heterozigot bir çağrı için alelik fraksiyonun %50 civarında ortalanması beklenir, dolayısıyla kabul kriterleri makul bir aralığı içerecek şekilde belirlenebilir (bazı laboratuvarlar %30 ile %70 arasında değişen eşik değerleri kullanır). Farklı varyant türleri için (SNV ve indeller gibi) farklı ayarlar gerekli olabilir. Eğer ayarlar ve metrikler çok sıkı ise, veriler kaybedilebilir veya filtrelenebilir, bu ise yanlış-negatiflere yol açar. Eğer daha az sıkı ise, bu defa çok fazla yanlış-pozitif ortaya çıkar. Ayrıca varyant tespiti düşük kapsamlı bir genom bölgesinde yapılırsa, çok sayıda dizileme hatası varyant olarak ortaya çıkabilir. Örneğin, sadece 5 veya 6 okuma ile kapsanan bir bölgede heterozigot varyant çağrısını destekleyen tek bir okumanın, bir dizi hatası ile haritalama hatası arasındaki ayrımı yapması güçtür. Bazı laboratuvarlar varyantları çağırma için gereken belli sayıda kalite okuması için katı bir eşik uygular. Başka laboratuvarlarda katı eşikler yoktur ama bunun yerine hataları ortadan kaldırmak için bu varyantların ortogonal bir yöntemle doğrulanmasını ya da aile bireylerinden alınan verilerle karşılaştırma isterler. Bununla birlikte kapsamının çok büyük olması da bir hata göstergesi olabilir çünkü düşük karmaşıklığa sahip bölgelerdeki okumalar, hatalı şekilde tek bir bölgeye haritalanabilir, sonuçta beklenenden daha yüksek bir kapsama ortaya çıkar. Beklenenden daha yüksek kapsama, EGFR geninde olabileceği gibi aşırı amplifikasyonun neden olabileceği haritalama hatalarının bir neticesi olabilir.

Bu aşamada GATK Unified Genotyper, GATK Haplotype Caller, samtools ve Playpus gibi yazılımlar kullanılmaktadır. Örnek olarak samtools verilebilir (**Bilgi kutusu 4**). En sık kullanılan birkaç tane varyant dosya formatı vardır, bunlar arasında VCF, genomeVCF ve genom varyasyon formatı yer alır^{120; 124}. 1000 Genom Projesi tarafından geliştirilen VCF, DNA dizileme varyantlarının depolanması için jenerik bir formatı temsil eder, bu kapsama ilgili anotasyonlarıyla birlikte tek nükleotid varyantları (SNV), indeller ve yapısal varyantlar girer¹²⁰. Bu format bir başlık ve veri (dizi varyantları) bölümü içerir. Başlık, standardize bilgilerin dizi verilerine uygun hale getirilmesine izin verir. VCF formatı aşağı yönlü analizleri (varyant filtrasyonu gibi) desteklemek için klinik ortamda yaygın şekilde kullanılır. Günümüzde varyant çağırma formatı (VCF) dosya spesifikasyonu, klinik toplum tarafından en çok kullanılan olmuştur, ancak uygulanması ve içerdiği anotasyon bilgileri laboratuvarlar arasında değişiklik gösterir. GenomeVCF, hem varyant olan hem de varyant olmayan pozisyonların gösterilmesini sağlar. Genom varyasyon formatı başka bir spesifikasyondur ve farklı organizmalar arasında gen anotasyonlarının karşılaştırılmasına izin vermek için ilk geliştirilen araç olan GFF3 formatına dayanır¹²⁴. Genom varyasyon formatının gelişimi, bir genom için detaylandırılmış anotasyonlar sağlar ve dizi değişimlerini ve gen ürünü üzerine beklenen etkilerini tanımlamak için ontolojilerden alınan terimleri kullanır¹²⁴. Belli bir dosya spesifikasyonu (VCF gibi) için dizi hizalama ve varyant çağırma için farklı yöntemlerin kullanılması gibi içerik gösteriminin veya dosya içindeki varyantı ve tipini tanımlamak için kullanılan yöntemlerin laboratuvarlar arasında belirgin şekilde

değişiklik gösterebileceği unutulmamalıdır. Standardize VCF gerekliliklerinin olmaması laboratuvarlar arasında varyant çağırma dosyalarının alışveriş edilmesinde ve karşılaştırılmasında büyük bir engel teşkil eder.

Bu dosya formatlarına ait spesifikasyonlar öncelikle klinik uygulamalar yerine araştırmaları desteklemek için oluşturulmuş veya popülasyon varyasyonlarını kataloglamak (VCF gibi) ya da tek bir şahsi genomun (genom varyasyon formatı gibi) derin anotasyonlarını yakalamak için tasarlanmıştır. Bu formatlar, veri öğelerine istinaden bir tutarlılık seviyesi sağlamak üzere tasarlanmıştır ve ayrıca çok geniş kapsamlı araştırma uygulamalarına uyum sağlayacak şekilde içerik sunumunda esnekliğe izin verirler. Varyant dosyaları genelde anotasyonlu dizi bilgilerini ve meta verileri içerir. Anotasyonlu dizi bilgileri, varyant tipinin pozisyonu gibi özel içeriğin anlaşılması için önemli olan ilgili verileri ve diziyi tanımlar. Meta veriler tipik olarak komple veri setiyle ilgili bilgileri verir.

VCF gibi spesifikasyonların gelişmeye devam edeceği kabul edilerek tek dosya formatının kullanılması önerilmemektedir. Var olan yazılımlar dosya spesifikasyonları arasında dönüşüm yapmaya olanak tanır. Ancak 2017 itibarıyla VCF spesifikasyonunun uzun geçmişinin bir sonucu olarak klinik uygulama içinde en çok kullanılan, gelişim gösteren spesifikasyon olduğu ve bu tipte dosyaları manipüle etmek için zengin araç setlerinin kullanılabilirliği kabul edilmektedir. Pratikte kullanılan varyant dosyası spesifikasyonundan bağımsız olarak spesifikasyon gelişim kaydettikçe bazı değişiklikler yapılmıştır. Dolayısıyla, varyant dosyalarının hem spesifikasyonun hem de kullanılan versiyonun bir tanımının içermesi gerektiği önerir. İdeal olarak en güncel versiyon kabul edilecektir, ancak pratikte kullanılan versiyonların varyasyonu, öngörülebilir gelecekte de var olmaya devam edecektir. Bu nedenle Varyant Call (arama) işlemi sonucunda çıktı olarak **VCF dosyası (Şekil 10, Şekil 11) (Bilgi Kutusu 6)** oluşmaktadır diye bahsedeceğiz.

5.3.3.1. Varyantların Belirlenmesi Sonrası Değerlendirilmesi Gereken Kalite Parametreleri:

Bu aşamada map edilen readler üzerinden varyasyon ve mutasyonlar tespit edilir. Elde edilen herhangi bir varyasyonun uygunluğu açısından güvenilirliklerinin değerlendirilmesi gerekmektedir. Bu kriterler aşağıda kısaca tanımlanmıştır.

a. Heterozigot dağılımlarının oranlarının tutarlılığı Kontaminasyon

Bu genel dğerlendirmeye yönelik bir parametredir. Henüz bu değer üzerinden kalite skoru veren bir araç olmamakla birlikte bakılmasında fayda olabilecek bir parametredir. Heterozigot varyasyonların normal koşullarda allellik dengesi %50-%50 olmalıdır. Ancak kontaminasyon ve farklı anomaliliklerde bu denge bozulmaktadır. Bu durumun tespit edilmesi durumunda lab. ile değerlendirme yapılarak problemin sebebi tespit edilmelidir.

b. Varyasyon/Mutant karar verilen lokasyonun okuma derinliği

Söz konusu varyasyonun bulunduğu dizileme derinliği bu noktada verilen kararın güvenilirliğini ortaya koymaktadır. Bu değer 12X'in üzerinde olması het./hom. kararının verilebilmesi için önem arz etmektedir.

c. Herbir okumanın ortalama kalite score'u

Bu varyasyonun tespitini yapan biyoformatik aracının ürettiği kalite score'udur. Bu skor ilgili aracın tanımladığı yeterlilik skorunun üzerinde olması gerekmektedir. Herbir varyantın doğruluğu bu değerle ortaya konabilir.

Bilgi Kutusu 6: Variant Call Format (VCF/BCF)

DNA sekans analizinde hizalanmış veriye variant calling yapıldıktan sonra oluşan dosya formatıdır. SNP ve INDEL variant yapılarının tutulduğu yazı tabanlı dosyadır. 8 tane sütundan oluşmaktadır. VCF dosyasının sıkıştırılmış hali BCF olarak adlandırılır.

Şekil 10. VCF dosyasında her sütunun ne anlama geldiği bilgisidir (<http://samtools.sourceforge.net/mpileup.shtml>).

Col	Field	Description
1	CHROM	Chromosome name
2	POS	1-based position. For an indel, this is the position preceding the indel.
3	ID	Variant identifier. Usually the dbSNP rsID.
4	REF	Reference sequence at POS involved in the variant. For a SNP, it is a single base.
5	ALT	Comma delimited list of alternative sequence(s).
6	QUAL	Phred-scaled probability of all samples being homozygous reference.
7	FILTER	Semicolon delimited list of filters that the variant fails to pass.
8	INFO	Semicolon delimited list of variant information.
9	FORMAT	Colon delimited list of the format of individual genotypes in the following fields.
10+	Sample(s)	Individual genotype information defined by FORMAT.

Şekil 11. Örnek bir VCF dosyası içeriğidir

(https://en.wikipedia.org/wiki/Variant_Call_Format).

```
##fileformat=VCFv4.0
##fileDate=20110705
##reference=1000GenomesPilot-NCBI37
##phasing=partial
##INFO=<ID=NS,Number=1,Type=Integer,Description="Number of Samples With Data">
#CHROM POS ID REF ALT QUAL FILTER
chr1 45796269 . G C
chr1 45797505 . C G 0
chr1 45798555 . T C SS
chr1 45798901 . C T
chr1 45805566 . G C SS
chr2 47703379 . C T
chr2 48010488 . G A
chr2 48030838 . A T
chr2 48032875 . CTAT -
chr2 48032937 . T C
```

d. Transition/transversion oranı

Toplumsal genom dizileme projelerinden elde edilen bilgilere göre, herhangi bir bireyin tüm genom dizi verisinden elde edilecek nokta mutasyonların transition/transversion oranı 2.0 ila 2.1 değerlerinde olmalıdır. Bu oran ekzom dizisinde 2.8 ila 3.0 olabilmektedir. Targeted ve amplicon dizilemelerinde bu oran dizilenen bölgeye bağlı olarak çok daha farklı değerler alabilir.

e. Varyantın duplikasyon/Homopolimer/Tandem tekrar bölgesinde olup olmaması durumu

Tespit edilen varyantlar duplikasyon bölgesinde olup olmadığına bakılması gerekmektedir. Bu değerler ilgili veri bankalarından anotasyon aşamasında çekilmektedir. Özellikle ilgili mutasyonun heterozigot ve homozigot olduğu konusunda kararın güvenilirliğine etki etmektedir. Duplikasyon bölgesinde heterozigot olarak tanımlanmış bir mutasyonun esasında iki ayrı bölgeden aynı yere haritalanmış edilmiş homozigot mutasyonlar olabilir. Bu noktaya dikkat edilmesi

gerekmektedir. Dizileme platformuna bağılı olarak homopolimer bölgelerde tespit edilen varyasyonların güvenilirlikleri düşük olabilmektedir. Aynı durum tandem repeatler için de geçerli olduđu durumlar ortaya çıkmaktadır.

f. Sürekli yanlış varyant tespit edildiđi bölgelerde olması durumu

Genomun bazı bölgelerinin dizi içeriğinden dolayı dizileme cihazları çok hatalı okuma yapmaktadır. Bu bölgeler cihazdan cihaza farklılıklar göstermektedir. Karar vericinin deęerlendirmeye aldıđı mutasyonun bu şekilde bir mutasyon olup olmadıđını deęerlendirmesi gerekmektedir. Bu durum daha önce aynı platform ve aynı lab ile dizileme yapılmıř verilere bakılarak deęerlendirilebilir. Söz konusu mutasyon eski birkaç dizilme ile birlikte görüntüleme araçları ile açılır ve bu durum kontrol edilerek karara bağlanır.

g. Varyantın MAF deęerleri

İlgili varyantın yerel veya uluslararası veri tabanlarındaki görülme sıklığı önemlidir. Zaten çok görülen bir varyantın ilgili numunede de görülmüş olması okumanın potansiyel olarak daha güvenilir olabileceğini işaret edebilir. Hiç görülmemiş veya nadir bir mutasyon için kalite skorları ve okuma kalitesi daha dikkatli incelenmelidir. Gerekirse Sanger validasyonu faydalı olabilmektedir.

h. Aile veya Trio dizilemesi ise tutarlılığı

Eđer dizileme aile bireyleri ile birlikte (kısaca *trio* dizilemesi) yapılmıř ise tanımlanan varyantın bulunduđu lokasyonun aile segregasyonu ile tutarlı olması gerekmektedir. Bu durumu sağlamaması durumunda varyantın tespitinin güvenilirliğinden řüphelenilmesi gerekmektedir. Mutasyonun *de novo* olduđundan řüpheleniliyor ise *trio*'da dođal olarak segregasyon söz konusu olmayacaktır.

i. Okumanın hastalık veritabanlarında yer alması

Eđer ilgili varyant veya mutasyon herhangi bir hastalık veritabanında yer alıyor ve hastalığın semptomları ile örtüşüyorsa ve yukarıdaki kalite parametrelerini sağlıyorsa nispeten güvenilirliği yüksek bir mutasyon olarak tanımlanabilir. Bu noktada deęerlendirilmesi gereken en önemli husus hastalık veritabanında yer alan mutasyonun ne kadar güvenilir olduđudur. Bazı veritabanlarındaki hastalık mutasyon ilişkilendirilmeleri çok geçmiş tarihli yayınlara veya tutarsız yaklaşımlara dayandırılabilir.

j. Homozigot olduđu ifade edilen bir nükleotidin gerçekten de homozigot olma deęerlendirmesi

Araçların homozigot olduđunu ifade ettiđi bir varyantın gerçekte homozigot olup olmadıđı önemli bir husustur. Bireydeki varyant gerçekte homozigot ise tespiti dođal olarak dođru bir tespit olacaktır. Güvenilirliği de dođrudan okuma derinliği veya okuma kalitesi ile ilişkili olacaktır. Ancak bireydeki gerçek mutasyon heterozigot olduđu halde homozigot okuma olasılığı da söz konusu olabilir. Bu duruma nispeten sıklıkla rastlanmaktadır. Bunun sebebi yukarıda birçok başlık altında bahsedilen okuma derinliğine bağılıdır. İlgili varyantın tespit edildiđi lokasyonun 12X altında okunması durumunda heterozigot bir lokasyonun yanlışlıkla homozigot olarak deęerlendirilmesi olasılığı oldukça yükselmektedir. Bu durum Sanger ile validasyona gönderilerek çözülebilir. Ayrıca örneğin 12X ile okunmuş bir lokasyonda bir alel 11 kere diđerisi ise 1 kere okunduğunda bu hatalı okuma olarak atlanarak homozigot rapor edilebilir. Bu notadaki asıl dramatik etki ise 11 kere okunan allelin referans ile aynı olması durumudur. Bu durumda potansiyel varyant hiç rapor edilmeyecektir.

k. Heterozigot olanın heterozigot olma olasılığı

Heterozigot olan bir varyasyonun heterozigot olarak tespit edilmesi doğru bir tespittir. Gerçekte homozigot olan varyasyonun heterozigot olarak rapor edilmesi nadiren ortaya çıkmaktadır. Bunun en temel nedeni az derinlikte okunmuş bir lokasyonda cihazın hatalı okuması ve rapor etmesi durumudur. Yukarıdaki kriterleri sağlayan bir varyasyonda nadiren bu durum gerçekleşir. Diğer bir etki ise yine yukarıda bahsedildiği gibi nükleotid dizisine veya cihaza *bias* (yanlılık) bir lokasyonda sürekli hatalı okuma yatkınlığından kaynaklanabilir. Burada dikkat edilmesi gereken en önemli husus her bir allelin hangi derinlikte okunmuş olduğudur. 6X kere okunmuş bir alel asgari doğrulanmış sayılabilir.

l. Hastalıkla ilişkili hiçbir mutasyonun olmadığı kararının verilmesi

Bu oldukça karmaşık ve mevcut hesaplama akışlarının desteklemediği bir durumdur. Bilimsel araştırmalardaki ana motivasyon fenotiple ilgili olabilecek bir mutasyon aranması şeklindedir. Ancak klinik uygulamalarda iki temel soru vardır. Birincisi semptomlarla ilişkili bir mutasyonun varlığının tespit edilmesidir. Bu durum bilimsel araştırmalar ile benzerlik taşıdığı için yukarıda açıklanan kriterler esas alınarak bu durum çalışılabilir.

İkincisi ve daha karmaşık olanı ise genel manada herhangi patojenik bir mutasyonun olmadığı belirlenmesidir. Bu durumun mutlak manada garanti edilmesi uygulamada mümkün değildir. Yeterli derinlikle okunamayan bölgelerde hangi mutasyonun olup olmadığı bilinemeyeceği için bu soruyu daha somut hale getirmek gerekmektedir.

Soruyu daha kolaylaştıracak olursak hastanın sahip olduğu semptomlara dayanarak şüphelenilen hastalıkla ilgili bir mutasyonun olmadığı tespit edilmesi gerekebilmektedir. Bu durum pratikte mümkün olmakla birlikte mevcut biyoinformatik akışlar bilimsel araştırmalara göre yapıldığı için söz konusu duruma cevap verememektedirler. Başka bir ifade ile söz konusu hastalıkla ilgili bir gen veya birkaç gen olsun ve hastalıkla ilişkilendirilmiş örneğin 20 mutasyon noktası olsun. Biyoinformatik akışlar varyasyon listelerini sadece bulunan mutasyonların yer alacağı şekilde raporladıkları için okunmayan bölgelerde potansiyel olarak var olan mutasyon yok olarak varsayılmaktadır. Örneğe dönecek olursak 20 mutasyon noktasının sadece 19'unun yeterli derinlikte okunduğu ve ilgili bir mutasyonun olmadığı söylenebilir. Ancak okunmamış olan mutasyon da yok varsayıldığı için orjinal soruya dönecek olursak bu hastalığın olmadığı söylenememiş olmaktadır. Bu durum aslında yeni programlar yazılarak düzenlenebilir ancak her biyoinformatik akış bunu desteklememektedir.

5.4 Tersiyer Analiz (Varyantların tanımlanması)

Üçüncül analiznin amacı, hangi dizi varyantının hastanın klinik tablosuyla ilgili olduğunu belirlemektir (**Bilgi Kutusu 7**). Çalışma grubu, üçüncül analiznin tasarımı aşamasında laboratuvarların varyantların tanımı veya etiketlenmesi için hastanın klinik tablosuyla ilgili olmayan varyantları çıkarmak veya “filtrelemek” için kullanılabilir özelliklere sahip yöntemler seçmelerini tavsiye etmektedir. Anotasyonlar şu gibi bilgileri içerir: popülasyondaki minör allel sıklığı, protein fonksiyonu üzerine varyant etkilerinin tahmini ve hastalık-varyant veri tabanlarının kullanımı. Bir genin fonksiyonuyla ilgili diğer anotasyonlar, herkese açık veri setlerinden elde edilebilir, örneğin NCBI (<http://www.ncbi.nlm.nih.gov>) tarafından ya da Avrupa Moleküler Biyoloji Laboratuvarı– Avrupa Biyo-İnformatik Enstitüsü/Sanger Merkezi (Hinxton, İngiltere) (<http://useast.ensembl.org/index.html>) tarafından geliştirilenler gibi. Bu anotasyonlar, ikincil analizler sırasında üretilen ve varyantın konumunu, varyantın türünü, zigotluğunu ve seçilen metrikleri (kapsamanın derinliği gibi) tanımlayan diğer anotasyonları tamamlar nitelikte olur. Laboratuvarların üçüncül analizlere farklı yaklaşımları olduğunu ve bunların da tanı hedefi uygulamasıyla değişiklik göstereceğini bilen çalışma grubu, üçüncül analizler için minimum anotasyon seti tanımlamamayı tercih etmiştir (**Bilgi Kutusu 8**). Durum böyle olsa bile çalışma grubu, anotasyonların varyantın konumunu ve türünü, alel sıklığını, öngörülen yapısal ve/veya fonksiyonel sonuçları ve bilinen hastalık ilişkilerini içermeleri gerektiğini önermiştir. Üçüncül analizdeki son adım, hangi varyantların raporlanması ve laboratuvar raporu içinde nasıl tanımlanmaları gerektiğini belirlemek için klinik bir değerlendirmeyi içerir. Pek çok analiz için özellikle ekzom ve genom testinde hastanın klinik tablosuyla ilgileri açısından çok sayıda varyant değerlendirilmek zorundadır. Bu değerlendirme adımının tasarlanması sırasında laboratuvarlar bir algoritma seçer ve klinik açıdan hastalıkla ilişkili genler içinde yer alan patojenik varyantları tahmin etmek için ayarlarını optimize eder. Başka algoritmalar kullanılarak seçilen varyantlar filtrelenebilir. Çalışma grubu, laboratuvarların üç önemli soruyu anotasyonların ve önceliklendirme (ing. prioritization) yöntemlerinin tasarımı ve optimizasyonu sırasında göz önünde bulundurmalarını tavsiye etmiştir:

1. Varyant, genin normal fonksiyonunu hastalık mekanizmasının anlaşılmasıyla tutarlı şekilde bozuyor mu ya da değiştiriyor mu?
2. Bu bozulma hastayı hastalığa veya insan sağlığıyla ilgili diğer sonuçlara götürüyor mu ya da hastayı bunlara yatkınlaştırıyor mu?
3. Bu sağlık sonucunun, hastanın klinik tablosuyla ve YND testi endikasyonu ile bir ilgisi var mı?

Laboratuvarlar varyantları genelde öngörülen fonksiyonel etkiye göre kategorize eder (örneğin sıklıkla benin, muhtemelen benin, anlamı bilinmeyen, muhtemelen patojenik, veya patojenik,¹⁰ ancak bunlar laboratuvarlar arasında bazen değişim gösterir). Amerikan Patolog Koleji (Northfield, IL), Amerikan Tıbbi Genetik ve Genomik Koleji (Bethesda, MD) ve Moleküler Patoloji Birliği (Bethesda, MD) varyant sınıflaması için güncellenmiş bir rehber geliştirmiştir⁹⁶. Çalışma grubu YND platformlarının, referans dizilerin, analiz ve yorumlama yazılımının (örneğin okuma haritalandırma ve varyant çağırma veya anotasyonu gibi) ve ilgili referans veri tabanlarının (örneğin mutasyon veri tabanları ve alel-frekans depoları gibi) sık sık güncellenmesini önermiş ve bu güncellemelerin bir testin performans

spesifikasyonunu etkileyebileceğini belirtmiştir. Bu tip değişiklikler genelde bir testin tamamının veya bir kısmının tekrar validasyonu gerektirir. Bir test sırasında erişim yapılan pek çok web tabanlı araç da sık sık güncellenmektedir. Çalışma grubu, bunun klinik laboratuvarlar için bir zorluk teşkil ettiğini görmüştür. Bu konuda şu tavsiyede bulunmuştur: Bu araçlar online elde edilip sürüm tayini yapılamıyorsa, klinik laboratuvarlar yazılım araçlarını kurum içine taşımalıdır, böylece her bir hasta testi için modifikasyonlar versiyonlanabilir, dokümanite edilebilir ve başvuru yapılabilir. Tıbbi uygulamalar için uygun yüksek kalitede veri sağlayabilen bir varyant veri tabanının olmaması sorun teşkil eder. HGMD (Human Genome Mutation Database) gibi var olan veri tabanları, varyantların klinik ilgisini değerlendirmede değerli araçlardır, ancak çalışma grubu pek çok girişin yeterince iyi yapılmamasından dolayı dikkat edilmesi gerektiğini ve ya yanlış-pozitif bir varyant-hastalık ilişkisini ya da sadece küçük bir grup bireye atfedilebilecek popülasyona özgü riskli alleli temsil edebilecekleri konusunu vurgulamıştır. Bu tip hatalar, yanlış filtrelemeye veya bir varyantın hastanın klinik tablosuyla ilişkisine göre yanlış atanmasına neden olabilir. NCBI ClinVar veri tabanı, Klinik Genom Kaynağı ve İnsan Variome Projesi gibi önemli çabalar sarf edilerek, sağlığı etkileyen insan genetik varyantları ve allelleri toplanmaya çalışılmaktadır. Bu çabaların ve diğer çabaların YND'nin tıbbi uygulamada kullanımını ilerletmesi beklenmektedir.

Varyantlar, laboratuvarlar arasında sonuçların karşılaştırmasını güçleştirerek farklı şekillerde kodlanabilmektedir¹²⁴. Bu sorunu ele almak için, veri paylaşımını kolaylaştırmak açısından "klinik nitelikli" varyant dosya formatının oluşturulması önerilmiştir.

Bilgi Kutusu 7

Üçüncül analizde ikincil analizin ürünü kullanılarak hastanın klinik durumuyla ilgili varyantlar filtrelenir, öncelik verilir ve sınıflandırılır. Bu süreç varyant anotasyonu ile başlar, bu işlem belirli bir varyant hakkında mevcut tüm bilgilerin toplanması ve ilişkilendirilmesidir. Anotasyonu yapılan varyantlar sonrasında standart kurallar kullanılarak sınıflandırılır, filtrelenir ve önceliklendirilir, bu şekilde test endikasyonu ile ilgili varyantlar belirlenir. Bazı anotasyonlar direkt ikincil analiz safhasından gelse de (varyant kalitesi ve kapsamının derinliği gibi), bunların çoğunluğu, elde edilen diziden anlaşılabilir ve ilgili veri tabanlarından elde edilir (örneğin bir varyantın bilinen veya öngörülen fonksiyonel sonuçları, popülasyon frekansı ve/veya bir genin hastalık veya fenotiple ilişkisi gibi). Bu anotasyonlar ticari veya kullanıma ücretsiz olarak açık araçlar kullanılarak otomatik süreçlerle elde edilebilir. Varsayılan genomik varyantların önceliklendirilmiş bir listesi, ilişkilendirilen genleri ve tahmin edilen fonksiyonel etkiler klinik bir değerlendirmeye tabi tutularak laboratuvar sonuç raporu içine eklenecek ilgili varyantlar tespit edilir. Bu adımlar genelde aday varyantların bir uzman tarafından manuel gözden geçirilmesini gerektirir. Bu uzman varyantların hasta ile ilişkisini mesleki olarak değerlendirecek tecrübeye sahip, hangi varyantların raporlanacağına kesin şekilde karar verebilen, varyantların çıkarımları ve testin sınırlamaları ile yorumlanması hakkında bilgiye sahip bir uzman olmak zorundadır. Aday varyantları gözden geçirmek için gereken süre belirgin şekilde değişiklik gösterebilir. Literatürde iyi karakterize edilmiş bir varyant hastanın fenotipi bağlamında kolayca yorumlanabilir. Literatürde sınırlı temsil edilmiş ama klinik ilişki ile tutarlı verilere sahip varyant/lar (örneğin segregasyon verileri, yapısal/fonksiyonel korelasyonlar veya hastalıkla ilişkili bir gende bulunma gibi), yorumlama ve değerlendirme için çok daha fazla zaman gerektirebilir. Varyant filtreleme ve önceliklendirme, makine öğrenme ve diğer algoritmik yaklaşımlar yoluyla gelecekte daha otomatik hale getirilebilir, ancak varyantların ve genlerin derinlemesine incelenmesi muhtemelen uzunca bir süre manuel olarak devam edecektir.

Hastalıkla ilişkili olduğu belirlenmiş genlerdeki varyantlar, tahmin edilen patojenliğe bağlı olarak ve hastanın klinik durumuna göre manuel değerlendirme ve sınıflandırma sırasında önceliklendirilir. Pek çok klinik laboratuvar varyantları beş ayrı kategoride sınıflandırılır (benin, muhtemelen benin, önemi bilinmeyen varyant, muhtemelen patojenik veya patojenik gibi)^{10; 11}. Klinik sınıflandırmanın filtreleme ve önceliklendirme adımlarından tam olarak ayrı olmadığı bilinmelidir, çünkü bu daima doğrusal bir süreç değildir. Örneğin bir dizi varyant, söz konusu tıbbi durumla ilişkisi olmadığı bilinen bir genle ilişkisi nedeniyle ilk analiz aşamalarında filtrelenebilir.

Analizdeki son adım klinik sonuçların raporlanmasıdır. YND için klinik sonuç raporları, test sonucunun ve sınırlamalarının anlaşılması için yeterli bilgi ihtiva etmelidir. Moleküler genetik test sonuçlarının raporları için önemli rehberler ve yorumlar sunulmuştur^{10; 13-15}. YND için sonuçların raporlanmasındaki zorluk, bir klinik uzmanın gerek duyduğu faydalı ve gerekli bilgileri uygun seviyede detaylarla dengelemektir, böylece analizin kullanımı ve sınırlamaları net şekilde anlaşılabilir ve uygun kullanıma, kullanıcılara ve diğer yardımcılarına yönlendirilebilir.



Şekil 12. Tüm analiz basamaklarının özeti. Cancer Inform (2014) 13:67–82'den değiştirilerek alınmıştır.

Varyant değerlendirilmesi detaylı isimlendirmeye dayanmaktadır. Anotasyonlar, ilgisiz genlerin filtrelenerek analizden çıkarılması için ve test endikasyonu ile ilişkili genin veya/veya varyantın olasılığına göre önceliklendirilmiş (veya derecelendirilmiş) bir liste oluşturmak için kullanılır. **En temel anotasyonlar gen adı, varyantın gende yer aldığı bölge (ekzonik, splice, intronik, genler arası vb.) ve varyant değişim bilgisidir.** Ek olarak bilinen polimorfizmler için MAF (minor allel frekansı), patojenisite ve korunmuşluk ve klinik veritabanları kullanılmaktadır. Bu amaçla Annovar, SnpEff, Cartagenia Bench Lab NGS, dbSNP, 1000 Genomes, ExAc, ESP 6500, SIFT, PhyloP, Mutation Taster, COSMIC, OMIM, ClinVar, HGMD gibi platformlar kullanılmaktadır. Bu işlem sonucunda çıktı olarak CSV, TSV, TXT, Excel dosyaları veya web tabanlı arayüzleri oluşturulmaktadır.

Bilgi Kutusu 8: Varyant anotasyonu

Anotasyon, elde edilen DNA varyantlarının fonksiyonel bilgiyi eklemek suretiyle anlamlandırılması işlemidir. Bu işlem izole edilen varyantlarla hastanın fenotipi arasındaki bağlantıyı sağlamayı amaçlamaktadır.

İlk basamak bilinen gen anotasyonlarını, bulunan varyantları fonksiyonel kategorilerine (örn. Sinonim, yanlış anlamlı, stop kodon ya da kırılma mutasyonları gibi) göre sınıflandırmak için kullanılmaktadır.

İkinci temel basamak ise halka açık veri bankalarında varyantın varlığı ve sıklığı bilgisidir ki sıklığı normal popülasyonlarda %5'den daha fazla görülen varyantlar polimorfizm olup nadir genetik hastalıklara sebep olmaları mümkün değildir. Birçok yeni nesil dizileme çalışması varyant anotasyonunu Minör Allel Sıklığını (MAF) kullanarak major popülasyon veri bankalarından alır. Örneğin NHLBI Exome Sequencing Project ³, dbSNP ¹², The 1000 Genomes Project Consortium (1000G), ExAC varyant bankası [ExAC Browser (Beta)]. <http://exac.broad.institute.org>. Ek bilgi, ilgili varyantların evrimsel korunmuşluğunun ölçülmesi (birtakım araçlar kullanılarak örneğin PhastCons ¹⁸, GERP++ ²⁰, PhyloP ²² ve varyantın protein üzerindeki patojenik etkisinin tahmin edilmesi [SIFT ²⁷ veya Poly-Phen ³² gibi in silico araçlar kullanılarak] ile elde edilebilir.

Üçüncü basamakta da genotip ile fenotip arasındaki bağlantının kurulmasında hem gen hem de varyant seviyesinde hastalığa özgü spesifik veri bankalarından elde edilen bilgiler kullanılabilir. Bu tür veri bankalarına örnek olarak İnsan Gen Mutasyon Bilgi Bankası (HGMD) ⁴⁴, Gen2Phen ⁴⁵, ClinVar ⁴⁶ ve OMIM ⁴⁷ verilebilir. Günümüzde üç yaygın olarak kullanılan anotasyon yazılımı bulunmaktadır: ANNOVAR ⁴⁸, SnpEff ⁵³ ve VEP [The Bioinformatics Knowledgeblog. Available from: <http://bioinformatics.knowledgeblog.org>].

Standardize minimum bir anotasyon seti henüz oluşturulmuş değildir. Klinik laboratuvarların kullandığı anotasyonların türü ve sayısı değişiklik gösterir. Bu anotasyon alanlarının bazıları birbiriyle örtüşmekte, diğerleri ise kendilerine özgü kalmaktadır. Örtüşen alanlarda bile kısıtlanmamış dizilerin kullanılması sıklıkla alan verilerinin uyumsuz kodlanmasına veya ontoloji ve sözdiziminde farklılıklara yol açar.

Tipik bir anotasyon dosyasında aşağıdaki verilerin olması önerilir:

- Hangi referans genom versiyonunun kullanıldığı ve genom dizisi koordinatları
- Gen sembolü, gen ismi
- Transkript ID'si nükleotid ve amino asit konumu
- Varyant türleri (örneğin sinonim, sinonim olmayan, yanlış anlamlı, anlamsız varyantlar gibi fonksiyonların öngörülen kaybı, çerçeve kayması varyantları, bir proteini çerçeve modifikasyonunu okuyarak zamanından önce kırılma ihtimali olan varyantlardır). Bunlar genin belirli bir transkriptine göre tanımlanmalıdır.
- Genelde laboratuvar tarafından seçilen tek kanonikal transkript.

- Protein işlevi anotasyonu/ patojenlik tahmin programlarına dayanarak protein üzerine öngörülen etki (hasar verici, etkisiz, vb.)
- Nükleotid koruma skorları veren predikسیونlar
- Hata tespit anotasyonları
 - Örneğin kalite skorları, haritalama kalitesi, kapsama derinliği ve varyant çağrısını destekleyen diğer kanıtlar
- Popülasyon veri setlerinden alel frekansları
- İnsan Geni Mutasyon Veri Tabanı veya ClinVar gibi bir varyant veri tabanında varyantın olup olmadığına ve önceden klinik açıdan anlamlı olarak etiketlenip etiketlenmediğine ilişkin yapılandırılmış bilgiler de eklenir. ClinVar, bir diziye geri haritalanabilecek bir format şeklinde tüm OMIM varyantlarını içerir, bu ise diğer veri tabanlarına göre dikkate değer bir avantajdır. OMIM içinde bulunan pek çok varyant buna izin verecek şekilde biçimlendirilmiş değildir.

Anotasyon araçlarının seçiminde, klinik testle tespit edilecek olan dizi varyantlarının türünü ve ayrıca belirli varyant türlerini tespit edecek yazılımın güçlü yönleri ve kısıtlamalarını dayanak alınmalıdır. Geçerli kılınması gereken sorunsuz bir otomasyon elde etmek için anotasyon ve filtreleme araçlarının, tüm informatik pipeline'ına entegre edilmesi gerekir.

Çalışma grubu, güvenilir ve tıbbi olarak şekil verilmiş veri tabanları mevcut olana kadar, klinik bir değerlendirme amacıyla varyant anotasyonunda kullanılan verilerin dikkatli şekilde değerlendirilerek yeterli kanıtla desteklendiklerinden emin olunmasını tavsiye eder.

5.4. Teknik Zorluklar:

5.4.1. Varyant Tipleri

Çok sayıdaki üçlü tekrarlar, büyük delesyonlar-duplikasyonlar, yapısal mutasyonlar örneğin translokasyonlar ve/veya inversiyonlar, füzyon gen tespiti, uniparental dizomi gibi durumlar tüm ekzom dizileme yöntemi ile tespit edilememektedir. Çeşitli varyantların yeni nesil dizileme metodları ile tespit edilebilme güçleri **Tablo 3**'de gösterilmiştir.

5.4.2. Regülatuvar ve Ekstranükleer DNA

Mitokondrial DNA, regülatuvar bölgeler, rutin olarak tüm ekzom dizilemede kullanılan bölgeler değildir.

5.4.3. Homoloji Bölgeleri

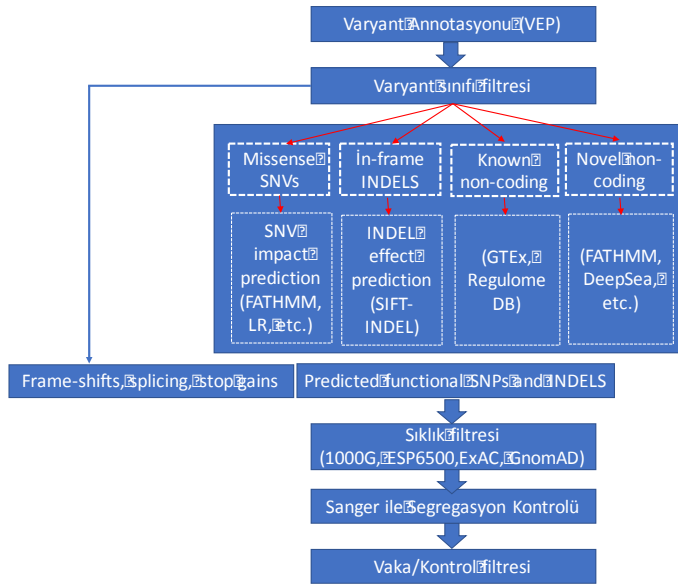
Daha önceki bölümlerde bahsedilmiştir.

5.4.5. Kapsanmayan bölgeler

Genelde kapsama ne kadar artarsa ve TED/TGD için ne kadar fazla aile bireyi dizilenirse varyant tespiti o kadar güvenilir olur. Düşük kapsama, özellikle amplifikasyon yanlılığı (ing. bias) varlığında varyantların kaçırılma (yanlış negatif) ve allelik durumların yanlış belirlenmesi (zigotluk) risklerini artırır, ayrıca dizileme artefaktlarının etkin şekilde filtrelenmesini azaltır ki bu durum yanlış pozitif varyantlara neden olur. Laboratuvarlar, kendi tanısal yaklaşımlarına göre varyantları tespit etmek için gereken minimum kapsama eşliğini belirlemeli (TED/TGD için sadece indeks vaka dizileme, indeks vaka artı ebeveyn dizileme gibi) ve test için garanti edilen minimum eşikle ilgili analitik performansı raporlamalıdır. Germ hattı heterozigot varyantların tespit edilmesi ve hedeflenen bir panelin tüm bazlarını

kapsamak için bazı laboratuvarlar minimum 10-20X kullanırlar. TED ve TGD testleri için minimum ortalama kapsamının ve mutlak minimum eşığe ulaşan bazların yüzdesinin takip edilmesi daha faydalı olabilir. Örneğin bir laboratuvar indeks vaka için TED'in minimum 100x'lik ortalama kapsama değeriyle laboratuvarın tanımladığı hedefin %90-95'inin en az 10x kapsamaya ulaştığından emin olabilir. Trio dizilemede en düşük eşik değer olarak 70x kullanılabilir. TGD için bir laboratuvar testin 30x'lik minimum ortalama kapsamaya ulaştığından emin olunabilir. Daha yüksek kapsama ve ilave varyant çağırma parametreleri, miks veya mozaik numunelerden varyantların tespit edilmesi için gereklidir (düşük yüzdeli tümör hücrelerine sahip somatik tümör örnekleri, mitokondri heteroplazmisi, germ hattı mosaisizmi gibi). Minimum kapsamının, platformun pek çok özelliğine ve deneye büyük oranda bağlı olduğu hatırlanmalıdır. Bu kapsama baz çağırma hata oranları, duolike okumalara karşılık kaç tane okumanın bağımsız olduğu gibi kalite parametreleri ve analitik pipeline performansı gibi diğer faktörler girer. Bu nedenle kapsama için belirli bir minimum eşik önerilmesi mümkün değildir, dolayısıyla laboratuvarlar analitik validasyon için toplam metriklere uygun olan minimum kapsama eşliğini seçmek zorundadır. Genlerin özellikle ilk ekzonları, ya da GC içeriğinin çok olduğu, kapsanamayan zor ekzonik bölgeler, ekzonik bölgeler dışındaki bölgeler (intronlar, promotor bölgeleri, vb) tüm ekzom dizileme yöntemi ile analiz edilememektedir.

5.4.6. Düşük Düzeydeki Mozaisizmler



Şekil 13. Varyant anotasyonunda kullanılan filtreler. BioTechniques (2017) 62:18-30'dan¹²⁵ değiştirilerek alınmıştır.

6. Verilerin Yorumlanması:

Tespit edilen mutasyon(lar) ile araştırılan hastalık hakkında ilişkiyi ya da sebep-sonuç ilişkisini kurmak üzere genomik bulguları hastanın klinik fenotipiyle ilişkilendirme safhasıdır (**Bilgi Kutusu 9**).

Tipik varyant filtreleri düşük kaliteli varyantları, intronik/intergenic varyantları, sessiz tek nükleotid değişimlerini veya toplumdaki düşük frekanslı bilinen polimorfizmleri dışlamaktadır. Ek olarak laboratuvarın kendi veri bankasında (*in-house*) derlenmiş varyant frekans bilgileri de kullanılmaktadır. Böylelikle kullanılan platform ve analiz pipeline kaynaklı yanlış pozitif varyantların anlaşılabilmesi de kolaylaşmaktadır. Cartagenia Bench Lab NGS, SnpSift, Ingenuity gibi platformlar bu aşama için kullanılmaktadır. Bu aşamada kullanılacak ve biyoinformatik aşamalara adapte edilebilecek çok çeşitli *in silico* araçlar yer almaktadır. Bunlarla ilgili olarak Niroula A. ve ark.¹²⁶ yaptığı derlemeye başvurulabilir.

6.1. Varyant Filtreleme/Önceliklendirme

Bu basamak, her laboratuvarın ve/veya analistin kendi tecrübesine göre şekillenebilen bir basamaktır ve tek bir öneri yoktur. Sadece diğer basamaklardaki gibi öncelikle ilgili basamağın kalite kontrolü yapıldıktan sonra, hastalık sebebi olan varyantın tespit edilmesine yönelik değişik önceliklendirmeler yapılabilir. Tüm farklı önceliklendirme yöntemleri sonunda hepsi aynı hastalıkla ilişkili varyantları bulacaktır (**Şekil 14, Şekil 15**). Aşağıda genel yaklaşım basamakları verilmiştir.

6.1.1. Primer filtreleme

6.1.1.1. Kalite kontrol

6.1.1.2. Sinonimlerin dışlanması

6.1.1.3. Minör allel frekansına göre filtreleme (Topluma özgün, laboratuvarın kendi veri bankası)

6.1.1.4. Kalıtım modeline göre filtreleme

6.1.2. Sekonder filtreleme

6.1.2.1. In silico tahmin programları

Önceki bölümlerde bahsedildiği üzere klinik laboratuvarların geliştirdiği informatik pipeline'ların çoğu, patojenlik tahmin programları kullanarak varyantın protein yapısı veya işlevi üzerine olası etkisini değerlendirir. Patojenlik tahmin programları, gen yapısını ya da ortaya çıkan protein ürününü bozma olasılığı daha fazla olan varyantların tespit edilmesine yardımcı olur; ancak hassasiyetleri ve özgüllükleri düşüktür¹²⁷⁻¹³⁰. Her bir patojenlik tahmin programı tarafından kullanılan algoritmanın anlaşılması önemlidir. Araçlardan oluşan bir kombinasyon kullanmak bilgilendirici olabilir, ancak aynı algoritmayı kullanan araçlardan elde edilen sonuçlar, bağımsız bir kanıt olarak düşünülmemelidir. Örneğin, pek çok laboratuvar benzer algoritmaları kullanıyor olsa bile hem SIFT hem de PolyPhen2 kullanarak veri setlerini analiz eder, böylece benzer etki kestirimleri elde edilir. Mümkün olan yerde laboratuvar, her biri patojenliği tahmin etmede farklı bir yaklaşım kullanan birden fazla tahmin programı kullanmayı düşünmelidir. Tahminler birden fazla kritere dayanır, örneğin varyantın biyokimyasal doğası, patojenlik bilgileri ve olasılıkla yapısal bilgiler. Bu programlar, yapısal veya işlevsel bir değişikliği artırabilen ya da azaltabilen diğer komşu varyantların etkisini entegre etme kapasitesinden yoksun olma gibi sınırlamalara sahiptirler. Örneğin bir çalışmada, PolyPhen2 ve SIFT gibi bilgisayar ortamında kullanılan tahmin programlarının, incelenen vakaların yaklaşık %71'inde bir varyantın benin mi yoksa patojenik mi olduğunu doğru şekilde

tahmin ettikleri belirtilmiştir ¹³¹. Bu nedenle tahmin programlarından alınan sonuçların, diğer destekleyici kanıtların yokluğunda varyantları filtrelemek veya sınıflandırmak için tek kaynak olarak kullanılmaması önerilir. Ayrıca bir protein X-ray kristalografik yapısı varsa, yanlış anlamlı (missense) varyantın patojenik etkisini ortaya çıkarmak için moleküler yapısal modelleme yardımcı olabilir ¹³² ve çalışma grubu biyoinformatik olarak imkan var ise önceden analiz için yeterli bilgi bulunan genler için tüm varyantların bu yöntemle de değerlendirilmesini önermektedir. Aşağıda sık kullanılan in silico tahmin programlarına örnekler verilmiştir.

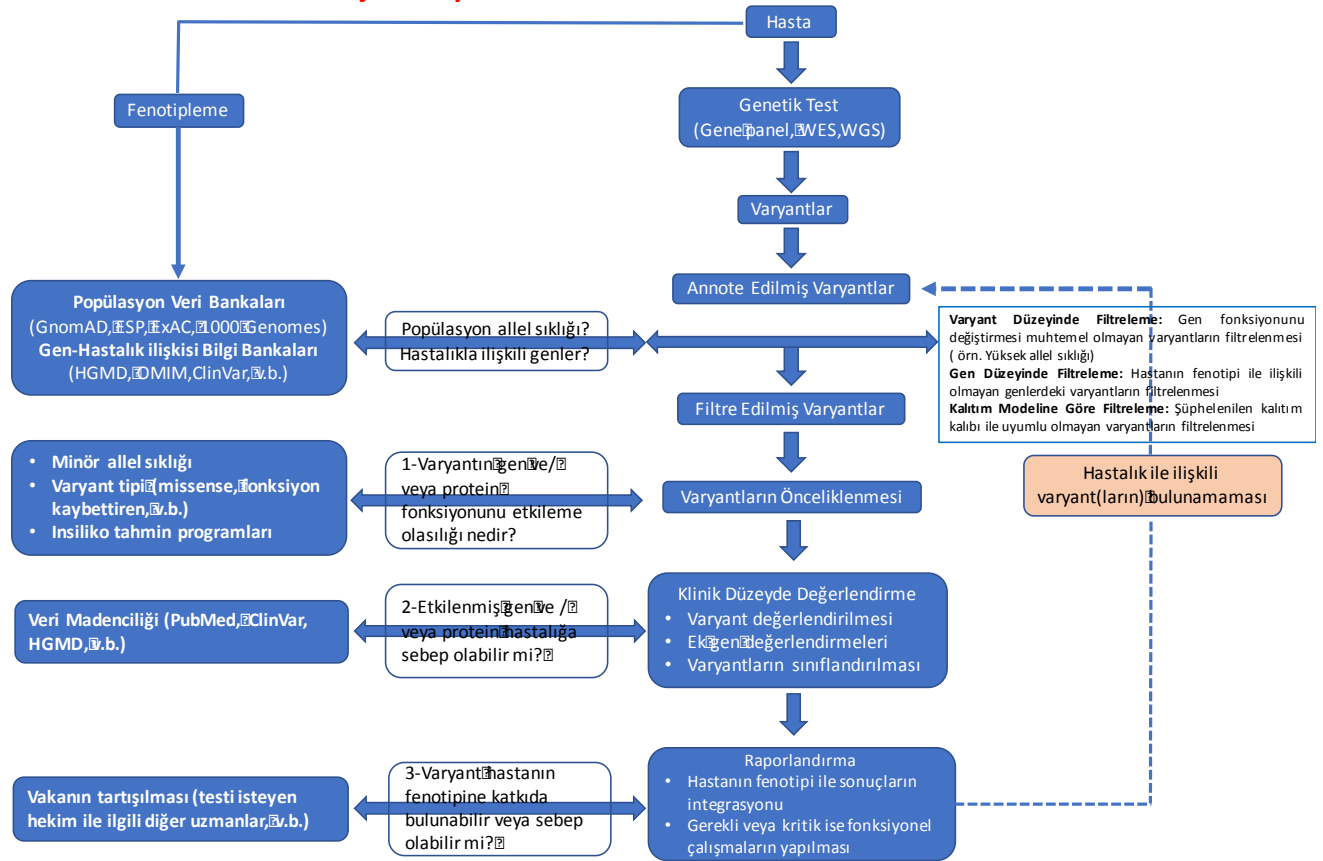
Non-sinonim varyantlar için tahmin programları: SIFT, Polyphen, Mutation Taster, Grantham

Sinonim varyantlar için tahmin programları: FATHMM-MKL, GWAVA, CADD

Kırılma noktalarındaki varyantlar için tahmin programı: MaxEnt, GeneSplicer, Human Splicing Finder, MutPred Splice

Korunmuşluk için tahmin programları: PhyloP, GERP

6.1.2.2. Patojen varyant tahmini



Şekil 14. Fenotiplemeden raporlamaya algoritma önerisi-1. Nat Biotechnol (2015) 33:689-93'den⁸⁹ değiştirilerek alınmıştır.

Genom ve ekzom analizleri yapılırken, hastadan elde edilen dizinin analizinin "klinik açıdan" ilgili genlerle sınırlandırmak standart bir uygulamadır. "Tıbbi ekzom" terimi, klinik açıdan ilgili olduğu bilinen genlerin ekzonik kısmına verilen addır ^{46; 133}. Geriye baktığımızda bu genler OMIM veri tabanından veya HGMD'den elde edilmekteydi. Bu ve bunun gibi veri tabanları gerekli kaynaklardır, ancak bunların pek çoğu yeterli klinik özenle düzenlenmiş değildir. Örneğin, patojenik olarak listelenen pek çok varyantın aslında benin olduğu ve bu varyantlardan bazılarının da gen-hastalık

ilişkinin atanması için bir taban oluşturabileceği şu anda iyi bilinen bir gerçektir. Birçok gen için hastalıkla ilişkili genler hakkında yayınlanan literatür, kesin bir ilişki veya olası bir ilişki bile kurmak açısından yeterli kanıt içermez. NCBI ile işbirliği içinde yürütülen kapsamlı çabalar ClinVar ⁴⁶, ve ClinGen: Clinical Genome Resource Program (<http://www.iccg.org/about-the-iccg/clingen/>, accessed August 18, 2014) gibi bu konuları ele almak ve klinik nitelikte veri tabanları geliştirmek üzerine yoğunlaşmıştır. **Hastalıkla ilgili varyantların filtreleme olasılığını en aza indirmek için filtreleme algoritmalarının bildirilen patojenik varyantları içeren (HGMD gibi) veri tabanlarını kullanması gerektiği tavsiye edilmektedir.** Bazı programlar kullanıcıya bir gen veya varyantın literatürde ne zaman bildirildiğini söyleyecektir (HGMD ve OMIM gibi), bu durumda kullanıcı verileri değerlendirip hastanın fenotipine uyup uymadığına karar verebilir (**Şekil 12, Şekil 13, Şekil 14, Şekil 15**).

6.1.2.3. Kopya Sayısı Varyantları

NGS Panellerinde, ekzom ve genom verilerinde kopya sayısı değişiklikleri tespit edilebilmektedir. Varyant tespit edildiği takdirde doğrulanması tercihen ikincil bir

Bilgi Kutusu 9: Varyant filtrasyonu ve değerlendirilmesi

Kalite kontrol sistemlerinden geçerek varyant filtrasyonu aşamasına ulaşmış bir ekzom verisi ile insan referans genomunun tipik olarak karşılaştırılması sonucunda yaklaşık 100,000 genomik pozisyonda farklılık tespit edilmektedir. Bu varyantların çoğu fonksiyonel olarak nötral olup tüm bu varyantların manuel olarak değerlendirilmesi oldukça uzun zaman almaktadır. Filtrasyon işleminin amacı, bu onbinlerce variant listesinden birkaç tane, hastalığa sebep olabilecek varyantı tespit edebilmek ve böylece genetikçi veya araştırmacının bu varyantların ilgili hastalıkla ilişkisini belirleyebilmesine olanak sağlamaktır. Genellikle, kodlamayan varyantlar, en azından analiz ilk aşamasında, filtre edilir. İkinci olarak, halka açık veya in-house veri tabanlarında yüksek minor allel sıklığına sahip olan varyantlar filtre edilmelidir. Nadir resesif hastalıklar için, minor allel sıklığı cut-off değeri %1 civarında olabilir. Dominant veya X'e bağlı hastalıklar için cut-off değeri sifıra yakın bir değer olmalıdır. Elde edilen varyantların Sanger dizileme ile segregasyon analizlerinin yapılması ile primer filtreleme tamamlanmış olur. Varyantın fonksiyonel etkisinin tahmin edildiği yazılımlar (SIFT, PolyPhen2, CADD, Human Splicing Finder, v.b.) veya evrimsel korunmuşluk tahmin yazılımları (PhyloP, GERP skorları) varyantların önceliklendirilmesi için kullanılabilir. Ancak sadece tahmin yazılımlarına göre varyantlar filtre edilmemelidir. Resesif kalıtım düşünüldüğünde sadece nadir homozigot veya multiple heterozigot varyantlar (Birleşik heterozigot olanları tespit edebilmek için) değerlendirilecektir. Tipik olarak ekzom verisinde yaklaşık 5-10 arası aday gen kalacaktır, ancak bu sayı akraba evliliği ürünü olan hastalarda akrabalık derecesi arttıkça artmak suretiyle daha fazla olacaktır. Dominant kalıtım kalıbı düşünüldüğünde final varyant listesinde ekzom başına yaklaşık 50-100 varyant kalacaktır. Bu sayıyı azaltmak için diğer aile bireylerden de dizileme yapılması faydalı olmaktadır. Benzer şekilde de novo mutasyonları ararken de anne-baba ve çocuk'tan oluşan trio dizileme yapılması önerilmektedir. Tipik bir ekzom 1-2 *de novo* kodlayan varyant içermektedir, ancak bu sayıya ulaşmak, anne-baba ekzom verisi olmadan mümkün olmamaktadır. Aynı aileden birden fazla bireye ekzom yapılması, hastalık sebebi olan varyantların ayırt edilmesi için özellikle de filtrasyon basamağında oldukça faydalıdır. Ancak indeks vakaya ek aile bireylerinin dizilenmesi sıklıkla dizilemenin fiyatı ile ilgilidir. Örneğin trio dizileme solo'ya göre çok daha ucuz olmalıdır. Bu basamakların hepsi yarı-otomatize sistemlerle yapılmaktadır ve şimdiye kadar geniş gruplarca kabul edilmiş kullanıcı dostu ve esnek kullanımı olan yazılım bulunmamaktadır. Son basamak her zaman filtre edilmiş aday genlerin deneyimli bir genetikçi tarafından manuel olarak değerlendirilmesidir ki ilgili varyantların ham veriden kontrol edilmesini, detaylı anotasyonların görsel inspeksiyonunu ve ilgili literatürün gözden geçirilmesini ve final raporun

metodla yapılarak raporlanabilir. Bu amaçla tüm ekzom verisinden (XHMM, CoNIFER,

ExomeDepth, and CONTRA) tüm genom verisinden de (CANVAS, Manta) gibi yazılımlar kullanılmaktadır. Kapsamanın tekilliği bakımından genom panelleri ve tüm genom dizilemesi CNV tespiti için en uygun yöntem olarak öne çıkmaktadır.

6.2. Verileri yorumlamadaki zorluklar

6.2.1. Toplum Veritabanı (GnomAD, ExAC, 1000 Genomes, in-house DB)

Her popülasyona özgü varyantlar vardır. Ülkemiz popülasyonu her ne kadar uniform olmasa da verilerin analiz edilebilmesi için popülasyona göre sapmaların çıkarılması için ülkemizde yapılan tüm çalışmalarda ham verilerin ilgili kurum çatısında (Sağlık Bakanlığı, TÜBİTAK, TÜSEB, v.b.) toplanması ve araştırmacıların kullanımına açılması, gerekli etik düzenlemeler yapılarak ivedilikle kullanıma geçirilmelidir.

6.2.2. Klinik Varyant Veritabanları

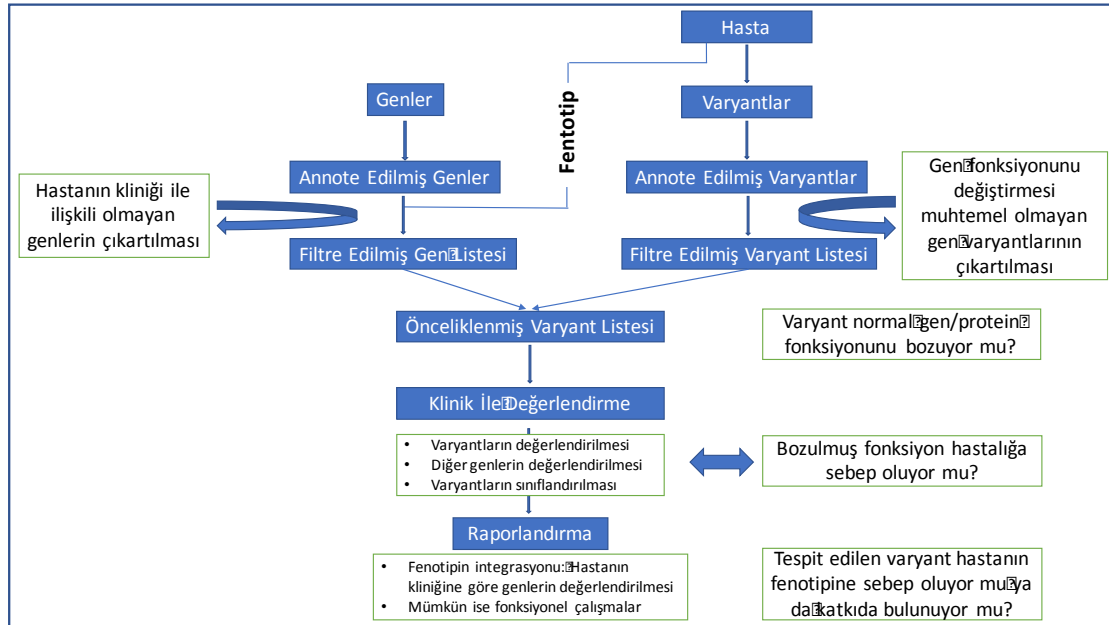
Günümüzde birçok klinik varyant veri tabanı olmakla birlikte en sık kullanılanlara örnek olarak ClinVar ve OMIM verilebilir.

6.2.3. Rehber

Konu ile ilgili her bir basamak için A'dan Z'ye Tıbbi Genetik Derneği çatısı altında algoritmalar hazırlanmalıdır.

6.2.4. Klinik ve Araştırma Ağları (GeneMatcher, MyGene2 etc)

Özellikle Mendelian kalıtmı nadir hastalıklarda bulunan aday genlerin karşılaştırılması veya diğer bir deyişle aynı aday gende ikinci bir hit bulunması ve sonuç olarak ilgili hastalıklara sebep olan hastalık yapıcı genlerin keşfi için araştırmacıları bir araya getiren ağlardan faydalanılabilir.



Şekil 15. Fenotiplemeden raporlamaya algoritma önerisi-2

7. Yeni Nesil Dizileme Testlerinde Raporlama

7.1. Giriş

YND testleri, hastalığın genetik kökenini klinik bulgulardan yola çıkarak açıklamak için yapılır. Klinik bulguların bilinmeyen bir fenotipe işaret etmesi durumunda genetik bulgular da kesin klinik tanıya yol gösterici olabilir (reverse genetics). Bu testin son ürünü, yukarıda bahsedilen anotasyonlar da dahil olmak üzere belli bir varyanta ait mevcut tüm bilgilerin bir raporudur. Yayınlanan literatürün değerlendirilmesi kritik

önem taşımakta olup, hastalık sebebi olan bir varyantın tespit edilmesini veya ilgili gen hakkında yapılmış fonksiyonel testler gibi klinik açıdan önemli bilgiler verir. In silico tahminler; önceki bölümlerde değinildiği gibi genetik bilimindeki uzmanlar tarafından manuel yapılan anotasyonlara katkı sağlamakla birlikte, sadece destekleyici kanıt olarak kullanılır, çünkü pek çok aracın doğruluğu standart altıdır. Anotasyon, filtreleme ve önceliklendirme sonrasında geriye kalan varyantların sayısı, teste bağlı olarak birkaç tane ile birkaç yüz arasında değişir. Hastalıkla ilişkisi olduğu bilinen iyi karakterize varyantlar için çok az ilave değerlendirme gerekir. Diğerleri için ise yukarıda açıklandığı gibi, hangilerinin klinik açıdan ilgili olduğunu tespit etmek için manuel bir gözden geçirme gerekli olur. **Hastalık ilişkisi hakkında klinik değerlendirmenin ilgili klinik uzmanlığa sahip personel tarafından yapılması önerilir. Bazı durumlarda bu çalışma genetik hastalıklar uzmanı ile testi isteyen hekimler ve biyoinformatik uzmanlarının işbirliğiyle yürütülebilir.** Klinik değerlendirmenin ana amacı, laboratuvar raporu içine neyin dahil edileceğini belirlemektir.

7.2. Sonuç Verme Süresi

Yeni nesil dizileme testini uygulayacak laboratuvar, test endikasyonlarına göre testlerin önceliklendirilmesi algoritmaları ve ortalama sonuç verme sürelerini yazılı olarak belirlemelidir. Ortalama sonuç verme süreleri klinik açıdan makul olmalıdır.

7.3. Yeni Nesil Dizileme Klinik Raporlarında Bulunması gereken Alanlar

Raporlama temel olarak beş ana bölümden oluşmalıdır:

1-Hasta Bilgileri

Adı soyadı, doğum tarihi, test tarihi, alınan test örneği, varsa aile no, varsa diğer aile bireyleri, endikasyon (varsa ayırıcı tanı) ve kullanılan yöntem/yöntemler (fenotiple ilişkili varyantların Sanger ile doğrulanmasının yapılması, özgün genetik danışmaya yardımcı olması, taşıyıcılık testi ve prenatal tanı amacı ile gerekeceğinden yapılması önerilir ve yapılmış ise yöntemde belirtilmelidir).

2-Sonuçlar

Sadece fenotiple ilişkili olan veya ilişkili olduğu öngörülen (*in silico*, model organizma, laboratuvarın yayınlanmamış önceki hastalara ait veri bankası bilgileri/segregasyon ve ek testler ile desteklenmesinde tanı lehine yararlı/gerekli görülen) varyantlar raporda HGVS nomenklatürüne uygun yazılarak bildirilmelidir (www.hgvs.org). Ek olarak, mutasyonun ilk tanımlandığı tarihten itibaren alışlagelmiş yaygın kullanılan tanımı mevcut ise mutlaka alt not ya da parantez içinde verilmelidir. Mutasyonun saptandığı genin adı, kısaltılmış adı, kromozomdaki lokasyonu, genom asamblesi, nomenklatürün alındığı transkriptin numarası, ekzon, intron, promoter (3', 5') bölgesi, zigositesi belirtilmeli, hem cDNA ve varsa hem de protein düzeyinde tanımlama yapılmalı, mutasyonun tipi de (anlamsız, yanlış anlamlı, çerçeve kayması gibi) mutlaka yazılmalıdır. Rapor öncesi saptanan varyantın/mutasyonun BAM verisi görsel olarak da kontrol edilmelidir (Alamut visual, IGV, Sanger doğrulaması yapıldı ise elektroferogramı gibi). Testte, daha önce fenotiple ilişkisi kurulmamış olmasına karşın, protein fonksiyonuna etki edecek tüm varyantlar da ayrıca incelenmiş olmalı ve gerekli görülürse raporda yer verilmelidir. Raporlanan her varyantın Sanger dizileme ile teyit edilmesi (indel varyantlar ve düşük kapsamla dizilenmiş varyantlar hariç), her bir örnek için her bir basamağın valide edildiği, in-house standartların belirlendiği ve raporda bu durum ile ilgili şeffaf ve açık bilgi paylaşıldığı takdirde gerekli olmayabilir.

3- Yorumlar

Her bir hastalığın senaryosu farklı olduğundan, bu bölümde mümkün olduğu kadar net ve yalın bir yorum yapılmalıdır. Dışlanan tüm genlerin listesi ve yöntemin saptamada yetersiz kaldığı mutasyon tipi bu bölümde yazılmalıdır. Algoritmik olarak başka bir yöntem ile tanı testine devam edilmesi gerekirse bu bölümde yazılmalıdır.

TED hizmeti, kullanılan tekniğin doğal sınırı nedeniyle bütün genlerin tüm kodlayan bölgelerini tam olarak kapsamayı taahhüt edemez. Ancak gerekli görülmesi durumunda, fenotiple ilişkili görülen genlerin dizisinde eksiklik olan kodlayan bölgeler ayrı bir hizmet altında sunulabilmelidir.

4- İkincil Bulgular

Ekzom ve genom dizileme klinik açıdan raporlanabilir, rastlantısal veya ikincil bulgular ortaya çıkarabilir. **Bu nedenle bilgilendirilmiş onam formu alınırken mutlaka bu konu ile ilgili bilgi verme imkanları sunulmalı ve uygun yanıt analiz başlamadan kayıt altına alınmalıdır.** En son önerilen ismiyle İkincil bulgular konusu ilk defa yakın tarihli Amerikan Tıbbi Genetik Kolejinin rehberinde ele alınmış olup, önemli tartışmaların da konusunu teşkil etmektedir¹³⁴⁻¹³⁶. **Her bir laboratuvarın bu konuyu ele alan ülkelere özgü bir politika geliştirmeleri tavsiye edilmektedir. Örneğin ikincil bulgulara, toplumumuza yönelik sık görülen hastalıklar için toplum sağlığı adına, patojenik bir varyant görüldüğü zaman raporlanması önerilen bir gen listesi oluşturulabilir (örneğin MEFV, G6PD, HBB gibi genler). Ancak ana amacın bu olmadığı da hastaya mutlaka genetik danışma seansında anlatılmalıdır.** Bir test yapılması istenmeden önce, ilgili hekime dizileme için genomun hangi bölgelerinin hedeflendiğinin, test sınırlamalarının ve test sonucu raporunda ne verilmesi gerektiğinin açıklamasını yapmak iyi bir laboratuvar uygulaması olarak önerilmektedir.

Bir diğer nokta da ikincil/tesadüfî varyantlar ölmüş olguların analizleri için yapılan incelemelerde raporlandırılmamalı ancak yaşayan olgular için verilmelidir.

5- Varyant sınıflandırma bilgisi ve testin limitleri

Güncel bilimsel gelişmelere göre varyant sınıflandırmaları değişse de güncel olan geçerli yaklaşım: 'sınıf 1-patojenik'; 'sınıf 2-muhtemelen patojenik'; 'sınıf 3- önemi bilinmeyen varyant /VUS'; 'sınıf 4-muhtemelen benin'; 'sınıf 5-benin'; 'sınıf 6; hastalık ilişkili varyant' şeklindedir.

Tüm Ekzom Analizinde, panel gen testinde genlerin kodlayan ekzonlarının hangi oranda kapsandığı bildirilmelidir.

Biyoinformatik analiz, sadece tıbbi öykü, klinik ve laboratuvar bulguları ile ilişkisi bilinen genlerdeki varyantlar için yapılmalıdır. Nadir polimorfizmler yanlış pozitifliğe/yanlış negatifliğe, yanlış, yetersiz ve doğru olmayan yorumlara neden olabilir. Sonuçların klinik bulgularla örtüşmediğinin düşünülmesi durumunda ek biyoinformatik analizlerin yapılması istenebilir. Bu test, tekrar dizi bölgelerini ve bu bölgelerdeki artışları doğru olarak gösteremeyebilir.

Patojenik varyantların ve önemi bilinmeyen varyantların kalıtsal durumlarda raporlanması gerektiği önerilmektedir. Benin varyantların raporlanması önerilmemektedir. Ayrıca yeni veriler buldukça bulguların analizini bildirmek amacıyla, laboratuvarın varyantların tekrar sınıflandırılması veya izlenmesi için stratejilere sahip olmaları da önerilmiştir.

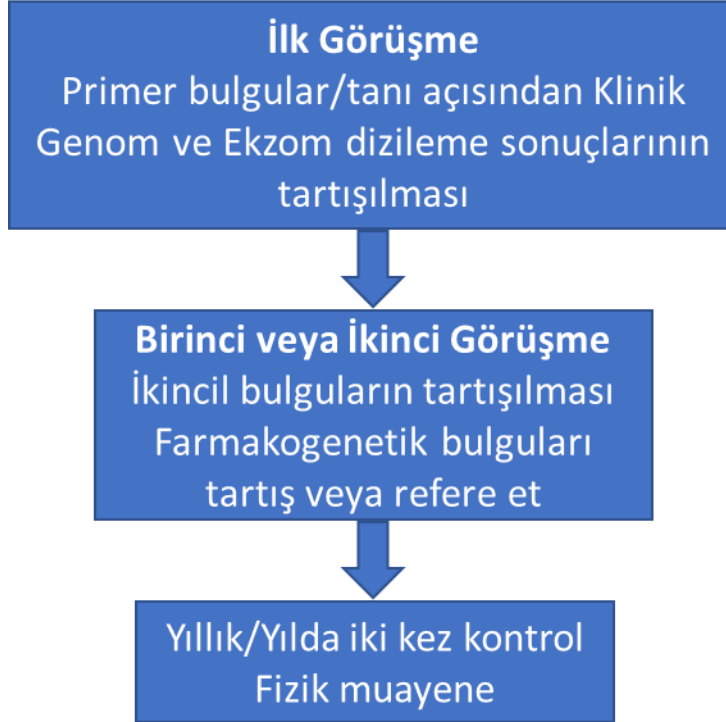
Yeni Nesil Dizi Analizinin kolaylaşabilmesi, yaygınlaşabilmesi ve ucuzlaşabilmesi için SNV, CNV öncelikli ancak kompleks düzenlenmeler, sitogenetik anomaliler, üçlü

nükleotid tekrarları gibi varyantlara/polimorfizmlere dair ulusal veritabanlarının kurulmasına öncülük edilmesi gereklidir. Dolayısı ile rapor edilen bütün varyantlar öncelikle ulusal daha sonra uluslararası veritabanlarına bildirilmesi tercih edilmelidir.

7.4. Verinin Yeniden Analizi

Yeni nesil dizileme ile sayıları binlerden milyonlara kadar varan sayıda varyant elde edilebilmektedir. Elde edilen varyantlar ile ilgili bilgiler zamanla değişebilme ve dolayısıyla güncel bilgiler ile yeniden raporlama gerekebilmektedir. Ek olarak biyoinformatik araçlar her geçen gün gelişmekte ve daha önce tespit edilemeyen varyantlar tespit edilebilmektedir. **Laboratuvarlar genetik testin yeniden analizi ve bu analiz için ek ücret alınıp alınmayacağı konusundaki politikasını belirlemelidir. Henüz hangi sıklıkla verilerin yeniden değerlendirilmesi gerektiği ile ilgili fikir birliği bulunmamaktadır.** Bu konuda özet kısmındaki tavsiyeye göre hareket edilebilir.

Primer test endikasyonu ile ilgili genlerden bulunan önemi bilinmeyen varyantlar, özellikle muhtemelen patojenik olarak sınıflandırılmışsa hastanın klinisyeni ile kontakta kalıp hastanın fenotipinin takip edilmesi önerilmektedir. ACMG kılavuzunda hastanın primer hekimleri ile yeniden irtibata geçilmesi konusunda daha detaylı bilgi bulunabilir (**Şekil 16**).



Şekil 16. Raporlama sonrası takip önerisi. Genet Med (2016) 18:1075-1084'den¹ değiştirilerek alınmıştır.

8. RNA Dizilemenin rutin tanıdaki yeri

YND'nin başlangıç materyalinin RNA olduğu durumlar için genel olarak kullanılan terim RNA-dizileme (RNA-seq)'dir. RNA-seq'in rutin tanıda kendine yer bulmaya başlamasının en temel nedeni TED ve TGD ile dahi, Mendelian hastalıklarda patojenik varyantı belirlemedeki başarı oranlarının %50'leri geçememesi, pek çok bozukluk için bu oranın daha da düşük olmasıdır. Özellikle protein-kodlayan dizilerin dışında kalan varyantların anlamlandırılması, henüz tanı koymada kullanılacak güvenilirlik seviyesinde yapılamamaktadır; TED'in kapsadığı dizilerden olan splay bölgelerinde yer alan varyantlar da bunlara dahildir. TGD'nin artık \$1000'in altında yapılmaya

başlanması sonucunda bu dizilere promotor, enhancer ve diğer düzenleyici dizilerdeki varyantlar da eklenmiştir. Özetle, protein-kodlayan dizilerin dışında kalan varyantların anlamlandırılmasına yönelik giderek artan bir ihtiyaç vardır. Bu ihtiyaca yönelik elimizde olan belki de en etkili yanıt hasta örneklerinden elde edilecek RNA'nın dizilemesidir; böylece varyantların olası etkilerini doğrudan ölçmek mümkün olmaktadır. Özellikle splays bölgelerini bozan ya da yeni splays bölgeleri oluşturan mutasyonların etkilerinin anlaşılması için RNA-seq yöntemi etkili biçimde kullanılabilir. Bu yaklaşımı başarıyla uygulayan çok merkezli iki çalışmanın sonuçları 2017 yılında yayımlanmıştır^{137; 138}.

RNA-seq'in bir diğer kullanım alanı transkriptom analizi ile ifadesi farklılık gösteren genlerin, gen ağlarının ve yolların belirlenmesidir. 2000'li yıllarda transkriptomun yüksek çıktılı analizi için altın standard olan mikrodizin (ya da mikroarray) teknolojisinden farklı olarak, RNA-seq yaklaşımında öncül transcript bilgisine gerek yoktur. RNA-seq ile elde edilen veriler referans genomu eşlenerek ilgili transkript hakkında farklı seviyede bilgi elde edilir: i) Dizi varyantları, mutasyonlar, ii) alternatif splaying ve diğer izoformlar, iii) ifade seviyesi, iv) alelik ifade farklılığı. Transkriptler hakkında öncül bilgiye gereksinim duyulmaması sayesinde, daha önce tanımlanmamış transkriptlerin ve izoformların ilk defa tanımlanması bu yöntem ile mümkündür. Nitekim, Kremer ve ark.'nın mitokondriyopatili hasta fibroblastları, Cummings ve ark.'nın ise nadir kas hastalıklı hastaların kas biyopsilerinde yaptıkları RNA-seq analizleri, bu yöntemin uygulanabilirliğini ortaya koymuştur: TED/TGD'nin başarısız olduğu hastaların yaklaşık %10-20'sinde, TED/TGD sonucu güçlü aday varyantların belirlendiği hastaların ise %65'in üzerindeki bir bölümünde patojenik varyantın belirlenmesi RNA-seq ile mümkün olmuştur^{137; 138}.

RNA-seq'in diğer YND yaklaşımlarına göre klinikte uygulanmasının önündeki en önemli zorluklar ise patojenik varyantın etkisinin incelenen doku örneklerinde göstermesi gerekliliğine bağlı olarak uygun dokunun elde edilmesi ve RNA'nın bozunmasını engelleyecek koşullarda saklanmasıdır. İdeal olan hastalıklı dokudan alınan biyopsinin analizi¹³⁷ olsa da pek çok durumda bu mümkün olamayabilir. Alternatif olarak deri biyopsisinden elde edilmiş fibroblastlar¹³⁸, kan hücreleri¹³⁹, mutasyon tanıtılmış hücre hatları da¹⁴⁰ kullanılabildiği bildirilmiştir; hastalık ve varyantı barındıran genin farklı dokularda ifadesine bağlı olarak bu alternatiflerin kullanılabilme potansiyeli farklılık gösterir. Ancak, tüm dokular için "normal" doku sorunu bulunmakta olup, uygun kontrollerin bulunmadığı durumlarda GTEx (<https://www.gtexportal.org/home/>) gibi veritabanlarından faydalanılabilir.

Özellikle biyopsi örneğinin kullanılacağı durumlarda dikkat edilmesi gereken önemli bir nokta RNA'nın bütünlüğünü bozmadan dokunun saklanmasıdır. Bu amaçla geleneksel olarak kullanılan yöntem, dokunun alınır alınmaz sıvı nitrojende dondurulması ("snap-freeze") ve yine sıvı nitrojende, mümkün değilse de -80 °C'de saklanmasıdır. Ancak pratik uygulamadaki zorluklar nedeni ile bu yöntem alternatif olarak geliştirilen yöntemde koruyucu bir sıvı (örn. RNAlater) içinde doku -20 °C'de uzun süre saklanabilmektedir. Saklanan dokudan izole edilen RNA'nın degradasyonunun minimum olması ise, RNA-seq'den sağlıklı sonuçlar elde edilmesi için kritiktir.

Yukarıda da belirtildiği gibi RNA-seq terimi başlangıç materyali RNA olan YND yaklaşımları için ortak olarak kullanılan bir terim olup analiz edilen RNA türlerinin çeşidine göre farklı isimler alabilmektedir. En sık kullanılan yöntemde sadece

mRNA'lar hedeflenir ve bu nedenle mRNA-seq (ya da RNA-seq) olarak adlandırılır. Bu tip dizilemede kütüphane hazırlanırken mRNA'nın ayrıştırılması için oligo-dT eklenmiş manyetik boncuklar kullanılır. Böylece toplam RNA'nın yaklaşık %80-85'ini oluşturan ribosomal RNA (rRNA) ve yaklaşık %10'unu oluşturan transfer RNA (tRNA)'lardan kurtulmuş olunur. Fragmentasyon sonrası ilk ve ikinci iplik cDNA sentezi yapılır. Daha sonra, bir sonraki aşamada küt uçlu fragmanların birbirlerine eklenmesini engellemek için her bir fragmanın 3' ucuna tek bir adenin nükleotidi eklenir (adenilasyon ya da A-kuyruklanması). Adaptörlerin ligasyonla eklenmesi ise sonraki aşama olan dizilemede cDNA fragmanlarının "flowcell"e tutunmalarını sağlar. Her iki ucuna da adaptör eklenmiş cDNA'lerin zenginleştirilmesi için düşük döngü sayılı (yaklaşık 15) PCR uygulanır. Hazırlanan kütüphanenin qPCR ya da benzer bir yöntemle kantifikasyonu ise dizileme verilerinin kalitesi açısından önemlidir.

mRNA-seq'te kullanılan poly-A zenginleştirmeden farklı olarak "Total RNA-seq" yaklaşımında ribozomal RNA'yı (rRNA) tanıyan ve biyotinlenmiş problemler kullanılır. Böylece sadece mRNA değil aynı zamanda poly-A içermeyen farklı kodlamayan-RNA çeşitleri (örn. lincRNA, snRNA, snoRNA vd.) de dizilenir. Her iki yöntem de yakın zaman önce geliştirilen yenilik yardımı sayesinde her okumanın kaynaklandığı DNA iplikçliğini belirleyebilmektedir, bu şekilde yapılan RNA-seq için "ipliğe özgü" (*stranded ya da strand-spesifik*) terimi kullanılmaktadır. Bu yöntemde ikinci cDNA iplik sentezi sırasında dTTP yerine dUTP kullanılır.

Bir diğer RNA-seq uygulaması ise küçük-RNA-seq'tir (small RNA-seq). Bu yöntemde miRNA'lar başta olmak üzere diğer küçük RNA moleküllerinin dizilemesi gerçekleştirilir. Bunun için öncelikle, doku ya da hücrelerden RNA izole edilirken küçük RNA'ları da kapsayacak bir yöntem kullanılmalıdır. Küçük RNA-seq için kullanılan kütüphane hazırlama teknolojisi, total RNA-seq ve mRNA-seq'ten farklı olarak RNA'nın fragmentasyona gerek duymaz. Dahası, miRNA ve benzer küçük RNA'ların uçlarında bulunan 5' fosfat ve 3' hidroksil grupları sayesinde uygun adaptörler seçici olarak bu tip RNA moleküllerine kolaylıkla eklenebilir. Elde edilen kütüphanedeki 140-150 baz çifti uzunluğundaki moleküller miRNA'ları dizilemek için genellikle 1x50 bp okuma uzunluğu tercih edilir.

Bazı durumlarda tüm transkriptom yerine belli bir gen setinin ürettiği RNA'ları dizilemek yeterli olabilir. Örneğin düşük RNA miktarı ve/veya kalitesine sahip örnekler için hedefli RNA-seq idealdir. Çeşitli hastalık ve hücre içi sinyal yollarına özgü farklı RNA-seq panelleri mevcut olup alternatif olarak, istenilen hedef genlere ait RNA'lar da PCR ile zenginleştirilip dizilemede kullanılabilir.

DNA dizilemede okuma derinliği kritik bir parametre iken, RNA dizilemede okuma derinliği yerine *okuma sayısı* ölçülür. İfadesi yüksek genlerin iki örnek arasındaki ifade farklılıklarının belirlenmesi için 10 milyon okuma yeterli olabilirken, daha düşük seviyede ifade edilen genlerin karşılaştırılması, alternatif izoformların ve füzyon transkriptlerinin belirlenmesi ve kantifikasyonu, dizi varyantlarının belirlenmesi vb. amaçlar için ise en az 30-50 milyon okuma gerekir. Dizileme maliyetlerinin düşmesi ile 50-100 milyon okuma seviyesi günümüzde sıklıkla kullanılır olmuştur. Küçük RNA'lar için ise çok daha az sayıda okumaya ihtiyaç vardır; 5-10 milyon okuma ile düşük seviyede ifade edilen küçük RNA'lar dahi saptanabilmektedir.

Okuma sayısının yanında, bir diğer önemli parametre okumaların tek-uçlu (SE, single-end ya da single-read) ya da eş-uçlu (PE, paired-end) olmasıdır. Salt kantifikasyon amacı güdülen durumlarda tek-uçlu okumalar yeterli olurken, dizi varyantlarının

saptanması başta olmak üzere diğer uygulamalarda eş-uçlu okumalar tercih edilir. Bunun nedeni, aynı molekülün iki farklı uçtan okunması sonucunda hata oranının çok daha düşük olmasıdır.

RNA-seq verilerinin analizi pek çok yönden TED ve diğer DNA dizileme verilerinin analizi ile benzerlik taşır, ancak doğal olarak farklı biyoinformatik araçlara ihtiyaç duyulur. Farklı RNA-seq uygulamalarından elde edilen verilerin analizi için tek bir ideal pipeline yoktur ancak aynı basamakları izlerler: ham verilerin işlenmesi ve kalite kontrolü, işlenmiş okumaların genom haritalanması ve kalite kontrolü, varyantların analizi ve anotasyonu. En sık kullanılan araçlar arasında:

Haritalama ve splays kavşaklarının belirlenmesi: STAR, HISAT2, TopHat2, GSNAP

Gen ve izoform kantifikasyonu: RSEM, HTSeq-counts, featureCounts, Cufflinks, eXpress, kallisto, Sailfish, MISO, rMATS

Örnekler arası ifade farklılıklarının belirlenmesi: DESeq2, TMM, Cuffdiff2, EdgeR sayılabilir. Gen ifadesinin kantifikasyonunda salt okuma sayısı yerine çeşitli normalizasyon işlemleri uygulanarak rapor edilir: RPKM (bir milyon okuma başına ekzon modelinin her bin bazlık bölümündeki okuma sayısı), FPKM (bir milyon okuma başına ekzon modelinin her bin bazlık bölümündeki fragman sayısı) ve FPKM'e benzer biçimde hesaplanan ve diğer iki yöntemle göre daha sık tercih edilen TPM (transkript sayısı / milyon).

Verilerin analizi sırasında çeşitli kalite kontrol basamakları bulunur. Bu basamaklar ve sık kullanılan *in silico* araçlar (parantez içinde): PCR sonucu duplike okumaların ve rRNA/tRNA okumalarının eliminasyonu, adaptörlerin kesilmesi (FASTQC); haritalanmış okumaların kalite kontrolü, intronik ve intergenic okumaların oranının belirlenmesi, degradesyona bağlı 3' biası, gen üzerinde eşit dağılım (RSeQC, Picard, QoRTs); kantifikasyon, normalizasyon ve ifade farklılıklarının belirlenmesi (betimleyici plotlar).

9. Diğer tip dizilemeler ve yeni nesil dizilemenin diğer kullanım alanları

Kanserlerde uygulanan yeni nesil dizileme çalışmaları ile somatik mutasyonların tespit edilmesi, ya da anne kanında serbest dolaşan fetal DNA'ya uygulanan yeni nesil dizileme çalışmaları ve benzeri gibi özel durumların kendilerine ait endikasyon, laboratuvar, analiz ve kalite değerlendirme kriterleri olup bu konuların her biri ayrı çalışma gerektirmektedir ve bu çalışmada değinilmemiştir.

10. Veri saklama ve hasta raporlarının takip edilebilirliği

NGS, devasa boyutta veri üretir ve laboratuvarlar bu verileri ya kurum içi ya da kurum dışında saklamayı tercih edebilir. Bulut teknolojisi veri analizi ve saklanmasında yaygınlaşan bir teknoloji olmaktadır. Ancak pek çok bulut bilgisayar ortamı Sağlık Sigortası Taşınabilirlik ve Sorumluluk Yasasına uygun değildir, bu nedenle laboratuvarlar veri saklama yönteminin bu Yasaya uygun olduğundan emin olmalıdır. YND analizlerinin çok adımlı doğası nedeniyle birbirinden farklı bilgiler içeren ve farklı büyüklükte olan dosyalar oluşacaktır. YND dizileme görüntü dosyaları birkaç terabayt büyüklüğünde olduğu için genelde saklanmaz. Laboratuvarlar heterojen dizi hizalama ve varyant çağırma algoritmaları kullanabilir, dolayısıyla YND sürecinde üretilen dosyaların türü laboratuvarlar arasında büyük oranda farklılık gösterecektir. Laboratuvarlar kendi politikalarında hangi dosya türlerini muhafaza ettiklerini ve ne kadar süreyle muhafaza edeceklerini açıkça belirtmelidir, bir veri muhafaza politikası geliştirilmeli ve yerel, devlet ve federal şartlara uygun olmalıdır. CLIA yönetmeliği analitik sistem kayıtlarının ve test raporlarının en az 2 sene

saklanması şart koşar. Ülkemizde de benzer bir uygulama bulunmaktadır. YND teknolojilerine uygunluk açısından laboratuvarların ham veri dosyalarını en az 2 yıl süreyle saklamasını öneririz. Böylece ilk sonuçlar istendiğinde tekrar üretilebilir ve sonraki analizlerde analitik pipeline bilgilerine başvurulabilir (bam veya fastq dosyaları tüm okumaları saklar). Ayrıca laboratuvarlar hem VCF'yi hem de klinik açıdan ilgili varyantların alt kümelerini yorumlayan nihai klinik test raporunu mümkün olan süre boyunca muhafaza etmeyi göz önünde bulundurmalıdır.

10.1. Laboratuvarın Biyo-İnformatik Pipeline ile Üretilen Girdi Dosyaları, Ara Dosyalar ve Nihai Veri Dosyalarıyla İlgili Bir Politikası Olmalıdır

Laboratuvarlar biyo-informatik pipeline'ı ile üretilen veri dosyalarının saklanması için bir prosedür geliştirip takip etmelidir. YND ve ilgili veri analizleri ile büyük veri dosyaları ortaya çıkar, örneğin: Flowcell görüntüleme dosyaları, baz değerlerini ve kalite skorlarını içeren dizi okuma dosyaları, sonraki analiz adımlarıyla oluşan diğer ara dosyalar ve varyant metin dosyaları (VCF) gibi. Tüm bu dosyaları uzun bir süre muhafaza etmek genelde pek pratik ve ekonomik değildir, bu nedenle bu kontrol listesi, laboratuvarların veri dosyası muhafaza sürelerini ve nihai rapor düzenlendikten sonra hangi dosyaların muhafaza edileceğini belirten bir politika geliştirmelerini ister. **Laboratuvarların karşılık gelen kalite skorlarıyla birlikte dizileme dosyalarını (FASTQ dosyaları gibi) muhafaza etmeleri veya bu dosyaların tekrar üretilebileceği bir arşiv formatında (BAM dosyaları gibi) muhafaza etmeleri tavsiye edilmektedir.** Ayrıca varyantların frekanslarına yönelik bir veri tabanının olması sonraki tanılarda varyantların önceliklendirilmesinde kullanılabileceği için önemlidir. Bu formatlar daha sonraki bir tarihte tekrar analize olanak tanıyacaktır. Genom veya büyük ölçekli dizileme verilerinin, FASTQ dosyalarının veya standart arşiv formatlarının uzun sürelerle saklanması mevcut saklama teknolojileriyle maliyetli olabilir, ancak daha yeni sıkıştırılabilen formatlar yakın zamanda bir çözüm sayılabilecektir. Bu dosyaların ne kadar saklanacağı daha karmaşık bir karardır ve sayısız konuya bağlıdır: veri setinin büyüklüğü, laboratuvarın saklama kapasitesi, tıbbi ve hukuki konular, diğer kurumsal, yerel veya ulusal veri saklama gereklilikleri. Laboratuvarın veri saklama ve dosya muhafaza etme süresi politikasının kurumsal, yerel veya ulusal veri saklama gerekliliklerine uygun olması gerektiğini vurgulamak isteriz.

10.2. YND Veri Aktarımı Gizlilik Politikası

Laboratuvar, dizileme verilerinin dahili ve harici saklanması ve aktarılmasının hasta gizliliğini ve güvenliğini koruduğunu garanti eden süreçleri açıklayan bir politika ve prosedüre sahip olmalıdır—

Yeni nesil dizileme, özellikle gen dizileriyle ilgili olmak üzere çok büyük miktarda veri ve şu bilgileri üretir: isim, doğum tarihi, tıbbi kayıt numarası, koruma altındaki sağlık bilgilerinin diğer öğeleri. Bu bilgiler her hastayı teşhis etmekte kullanılabilir. Laboratuvarlar bu bilgilerin gizliliğini korumak için katı prosedürler oluşturmalıdır. Laboratuvarlar genom bilgilerini başka sağlık kurumlarına ve üçüncü taraf tedarikçilere (bulut tabanlı hesaplama kaynakları sağlayanlar gibi) aktarmak için sıkı politikalara sahip olmalıdır. Gizliliği garanti eden prosedürler arasında veri şifreleme, güvenli veri aktarımı, kontrollü erişim ile kullanıcı kimliği doğrulama, koruma altındaki sağlık bilgileri ve verilerin iletilmesini ve kurumlardan ve/veya kullanıcılardan gelen bilgileri izleyen denetimler olmalıdır.

Ek Bilgiler

Kullanılan Analiz Araçları

1. Fastx Tool Kit

DNA sekans analizlerinde kalite kontrol için kullanılan analiz aracı setidir. Bu analiz aracının binary formatı şu linkten indirilebilir: http://hannonlab.cshl.edu/fastx_toolkit/download.html

Fastx Tool Kit içinde yer alan analiz araçlarından en önemlileri şunlardır: fastq_quality_converter, fastq_quality_filter, fastq_quality_trimmer, fastx_clipper, fastx_trimmer, fastx clipper

2. FastQC

Fastq sekans dosyalarını analiz edip kalite kontrol bilgileri sunan bir analiz aracıdır. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> adresinden ilgili işletim sistemine uygun sürümü indirilip kullanılabilir.

3. Picard Tools

İçerisinde birçok analiz aracı bulunduran analiz aracı setidir. Bu setin SortSam isimli hizalanmış sekans sonucu oluşan bam dosyasını sıralamak amacıyla kullanılan analiz aracı <http://broadinstitute.github.io/picard/> adresinden indirilebilir.

Format Dönüştürme

Değişik formatlarda kaydedilen sekans verileri ve onların çeşitli özellikleri ile ilgili dosyaların bazı analiz araçlarına girdi olması için değişik formatlara dönüştürülmesi gerekmektedir. Bunun için birçok farklı yazılım kullanılmıştır.

1. Fastx-ToolKit v0.0.13

Bu araç paketi içerisindeki fastq_to_fasta aracı ile fastq formatındaki sekans verileri fasta formatında dönüştürülmektedir.

2. Samtools v1.3.1

İçerisinde birçok analiz aracı içeren yaygın olarak birçok yazılım tarafından kullanılan programdır (<http://samtools.sourceforge.net/>). samtools view aracı, SAM formatındaki veri ile onun binary formatı olan BAM dosyası arasında format dönüşümü yapmaktadır.

3. Picard v2.5.0

SAM/BAM/VCF formatındaki dosyalar için kullanılan içerisinde birçok araç bulunan bir yazılımdır (<https://broadinstitute.github.io/picard/index.html>).

o SortSam aracı ile SAM/BAM dönüşümleri yapılmakta hem de veri sorgu ismine göre ya da koordinata göre sıralanabilmektedir.

o SamtoFastq aracı ile SAM formatındaki dosya fastq formatında dönüştürülmektedir. Aynı işlemin tersi FastqtoSam aracı ile hizalanmadan yapılmaktadır. o VcfFormatConverter ile VCF-BCF dönüşümleri yapılmaktadır.

Kalite Kontrol, Kırpma, Önişleme

Girdi ham verisinden okuma kalitesi düşük olan sekans parçacıkları çıkartılır. Bu amaçla Fastx toolkit kalite kontrol yazılım araçları kullanılabilir. Alternatif olarak araştırılan ve test edilmiş yazılımlar şu şekildedir:

1. Fastx-ToolKit v0.0.13

DNA-sekans analizinde kalite kontrol için kullanılır. İçerisinde birçok analiz aracı barındıran araç setidir (http://hannonlab.cshl.edu/fastx_toolkit/). Diğer alternatif araçlara göre avantajı dokümantasyonunun güzel hazırlanmış olması ve yaygın kullanımında dolayı kolay çalıştırılabilir olmasıdır. Dezavantaj olarak çift-uçlu okuma verisi (paired-end) üzerinde çalışmamaktadır.

1.1 Fastx Clipper

Okunan sekanslarda kullanıcının belirlediği adaptör sekansının çıkarılması ya da sadece adaptör sekansı içeren okumaların saklanması sağlayan araçtır.

1.2 Fastx Trimmer

Sekanslarının sonundan ve başından istenildiği miktarda kırpma işlemi yapmak için kullanılır.

1.3 Fastq Quality Filter

Sekansların kalite değerlerine göre filtreleme işlemi yapar. Belirlenen eşik değeri üzerindeki sekansları saklar diğerlerini siler.

2. Skewer v0.2.2

Kalite kontrol için kullanılan güncel bir analiz aracıdır (<https://github.com/relipmoc/skewer>).

o Adaptör sekans eleme, sekans uzunluğu kırpma, kaliteye göre kırpma işlemleri yapmaktadır.

o Fastx-ToolKit'den avantajı hem tek-sonlu hem çift-sonlu veride üzerinde çalışabilmektedir.

o Piyasadaki en hızlı kalite kontrol aracı olduğu söylenmektedir (<http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-15-182>).

3. Cutadapt v1.10

Sekans analizinde kaliteye göre filtreleme, adaptör eleme, uzunluk kırpma işlemleri yapmaktadır (<http://cutadapt.readthedocs.io/en/stable/index.html>).

o Tek-uçlu ve çift-uçlu veri üzerinde çalışabilmektedir.

o Diğer analiz araçlarından farklı olarak farklı tiplerde adaptor eleme işlemi yapabilmektedir. Bu tipler; 3', 5' ve linked adapter'dir.

4. Trimmomatic v0.32

Illumina sekans verileri üzerinde adaptör sekans eleme, düşük kalite-ortalama kalite sekans kırpma, belirli sekans uzunluğunda kırpma işlemleri yapmaktadır (<http://www.usadellab.org/cms/?page=trimmomatic>). Diğer analiz araçları göre daha hızlıdır fakat fazla hafıza tüketmektedir. (Ortalama 2.5 GB)

5. Fastq-MCF v1.04.636

Adaptör sekans eleme, sekans sonundan kırpma, kaliteye göre kırpma işlemleri yapmaktadır (<https://code.google.com/archive/p/ea-utils/wikis/FastqMcf.wiki>).

o Tek-uçlu ve çift-uçlu veri üzerinde çalışabilmektedir.

6. FastQC v0.11.5

YND ile elde edilen sekans verisi için kalite kontrol aracıdır (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>).

o SAM-BAM-FASTQ formatındaki verileri işleyerek görselleştirme yapabilmektedir.

o Sekans verisinde *overrepresented sequence* (en çok tekrar eden) bulabilmektedir. Böylelikle diğer kümeleme analizleri için adaptör sekans belirlenmektedir.

o Her veri okuma parçası için ortalama kalite değerleri, tüm sekansların ortalama kalite değerleri, sekans uzunluk dağılımı grafikleri üretmektedir.

Referanslar

1. Bowdin, S., Gilbert, A., Bedoukian, E., Carew, C., Adam, M.P., Belmont, J., Bernhardt, B., Biesecker, L., Bjornsson, H.T., Blitzer, M., et al. (2016). Recommendations for the integration of genomics into clinical practice. *Genet Med* 18, 1075-1084.
2. Kadalayil, L., Rafiq, S., Rose-Zerilli, M.J., Pengelly, R.J., Parker, H., Oscier, D., Strefford, J.C., Tapper, W.J., Gibson, J., Ennis, S., et al. (2015). Exome sequence read depth methods for identifying copy number changes. *Brief Bioinform* 16, 380-392.
3. Fu, W., O'Connor, T.D., Jun, G., Kang, H.M., Abecasis, G., Leal, S.M., Gabriel, S., Rieder, M.J., Altshuler, D., Shendure, J., et al. (2013). Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* 493, 216-220.
4. Zappala, Z., and Montgomery, S.B. (2016). Non-Coding Loss-of-Function Variation in Human Genomes. *Hum Hered* 81, 78-87.
5. Lapin, V., Mighion, L.C., da Silva, C.P., Cuperus, Y., Bean, L.J., and Hegde, M.R. (2016). Regulating whole exome sequencing as a diagnostic test. *Hum Genet* 135, 655-673.
6. Lennon, N.J., Adalsteinsson, V.A., and Gabriel, S.B. (2016). Technological considerations for genome-guided diagnosis and management of cancer. *Genome Med* 8, 112.
7. Tetreault, M., Bareke, E., Nadaf, J., Alirezaie, N., and Majewski, J. (2015). Whole-exome sequencing as a diagnostic tool: current challenges and future opportunities. *Expert Rev Mol Diagn* 15, 749-760.
8. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078-2079.
9. (!!! INVALID CITATION !!! 90).
10. Richards, C.S., Bale, S., Bellissimo, D.B., Das, S., Grody, W.W., Hegde, M.R., Lyon, E., Ward, B.E., and Molecular Subcommittee of the, A.L.Q.A.C. (2008). ACMG recommendations for standards for interpretation and reporting of sequence variations: Revisions 2007. *Genet Med* 10, 294-300.
11. Kearney, H.M., Thorland, E.C., Brown, K.K., Quintero-Rivera, F., South, S.T., and Working Group of the American College of Medical Genetics Laboratory Quality Assurance, C. (2011). American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants. *Genet Med* 13, 680-685.
12. Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M., and Sirotkin, K. (2001). dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 29, 308-311.
13. Rehm, H.L., Bale, S.J., Bayrak-Toydemir, P., Berg, J.S., Brown, K.K., Deignan, J.L., Friez, M.J., Funke, B.H., Hegde, M.R., Lyon, E., et al. (2013). ACMG clinical laboratory standards for next-generation sequencing. *Genet Med* 15, 733-747.
14. Lubin, I.M., McGovern, M.M., Gibson, Z., Gross, S.J., Lyon, E., Pagon, R.A., Pratt, V.M., Rashid, J., Shaw, C., Stoddard, L., et al. (2009). Clinician perspectives

- about molecular genetic testing for heritable conditions and development of a clinician-friendly laboratory report. *J Mol Diagn* 11, 162-171.
15. Chen, B., Gagnon, M., Shahangian, S., Anderson, N.L., Howerton, D.A., Boone, J.D., Centers for Disease, C., and Prevention. (2009). Good laboratory practices for molecular genetic testing for heritable diseases and conditions. *MMWR Recomm Rep* 58, 1-37; quiz CE-31-34.
 16. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20, 1297-1303.
 17. Genomes Project, C., Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature* 491, 56-65.
 18. Margulies, E.H., Blanchette, M., Program, N.C.S., Haussler, D., and Green, E.D. (2003). Identification and characterization of multi-species conserved sequences. *Genome Res* 13, 2507-2518.
 19. O'Rawe, J., Jiang, T., Sun, G., Wu, Y., Wang, W., Hu, J., Bodily, P., Tian, L., Hakonarson, H., Johnson, W.E., et al. (2013). Low concordance of multiple variant-calling pipelines: practical implications for exome and genome sequencing. *Genome Med* 5, 28.
 20. Davydov, E.V., Goode, D.L., Sirota, M., Cooper, G.M., Sidow, A., and Batzoglou, S. (2010). Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput Biol* 6, e1001025.
 21. Thorvaldsdottir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* 14, 178-192.
 22. Pollard, K.S., Hubisz, M.J., Rosenbloom, K.R., and Siepel, A. (2010). Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res* 20, 110-121.
 23. Shomron, N. In *METHODS IN MOLECULAR BIOLOGY*. (Humana Press.), pp 2,3.
 24. Yang, Y., Muzny, D.M., Xia, F., Niu, Z., Person, R., Ding, Y., Ward, P., Braxton, A., Wang, M., Buhay, C., et al. (2014). Molecular findings among patients referred for clinical whole-exome sequencing. *JAMA* 312, 1870-1879.
 25. Harding, K.E., and Robertson, N.P. (2014). Applications of next-generation whole exome sequencing. *J Neurol* 261, 1244-1246.
 26. Ng, S.B., Turner, E.H., Robertson, P.D., Flygare, S.D., Bigham, A.W., Lee, C., Shaffer, T., Wong, M., Bhattacharjee, A., Eichler, E.E., et al. (2009). Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 461, 272-276.
 27. Ng, P.C., and Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res* 31, 3812-3814.
 28. Rosenfeld, J.A., Mason, C.E., and Smith, T.M. (2012). Limitations of the human reference genome for personalized genomics. *PLoS One* 7, e40294.
 29. Sawyer, S.L., Hartley, T., Dymont, D.A., Beaulieu, C.L., Schwartzentruber, J., Smith, A., Bedford, H.M., Bernard, G., Bernier, F.P., Brais, B., et al. (2016).

- Utility of whole-exome sequencing for those near the end of the diagnostic odyssey: time to address gaps in care. *Clin Genet* 89, 275-284.
30. Srivastava, S., Cohen, J.S., Vernon, H., Baranano, K., McClellan, R., Jamal, L., Naidu, S., and Fatemi, A. (2014). Clinical whole exome sequencing in child neurology practice. *Ann Neurol* 76, 473-483.
 31. Levenson, D. (2014). Whole-exome sequencing emerges as clinical diagnostic tool: testing method proves useful for diagnosing wide range of genetic disorders. *Am J Med Genet A* 164A, ix-x.
 32. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. *Nat Methods* 7, 248-249.
 33. Williams, E.S., and Hegde, M. (2013). Implementing genomic medicine in pathology. *Adv Anat Pathol* 20, 238-244.
 34. Boycott, K.M., Vanstone, M.R., Bulman, D.E., and MacKenzie, A.E. (2013). Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nat Rev Genet* 14, 681-691.
 35. Chang, F., and Li, M.M. (2013). Clinical application of amplicon-based next-generation sequencing in cancer. *Cancer Genet* 206, 413-419.
 36. Liu, C., Yang, X., Duffy, B., Mohanakumar, T., Mitra, R.D., Zody, M.C., and Pfeifer, J.D. (2013). ATHLATES: accurate typing of human leukocyte antigen through exome sequencing. *Nucleic Acids Res* 41, e142.
 37. Mori, A., Deola, S., Xumerle, L., Mijatovic, V., Malerba, G., and Monsurro, V. (2013). Next generation sequencing: new tools in immunology and hematology. *Blood Res* 48, 242-249.
 38. Gillis, N.K., Patel, J.N., and Innocenti, F. (2014). Clinical implementation of germ line cancer pharmacogenetic variants during the next-generation sequencing era. *Clin Pharmacol Ther* 95, 269-280.
 39. Bertelli, C., and Greub, G. (2013). Rapid bacterial genome sequencing: methods and applications in clinical microbiology. *Clin Microbiol Infect* 19, 803-813.
 40. Barzon, L., Lavezzo, E., Costanzi, G., Franchin, E., Toppo, S., and Palu, G. (2013). Next-generation sequencing technologies in diagnostic virology. *J Clin Virol* 58, 346-350.
 41. Churko, J.M., Mantalas, G.L., Snyder, M.P., and Wu, J.C. (2013). Overview of high throughput sequencing technologies to elucidate molecular pathways in cardiovascular diseases. *Circ Res* 112, 1613-1623.
 42. Pinxten, W., and Howard, H.C. (2014). Ethical issues raised by whole genome sequencing. *Best Pract Res Clin Gastroenterol* 28, 269-279.
 43. Soden, S.E., Saunders, C.J., Willig, L.K., Farrow, E.G., Smith, L.D., Petrikin, J.E., LePichon, J.B., Miller, N.A., Thiffault, I., Dinwiddie, D.L., et al. (2014). Effectiveness of exome and genome sequencing guided by acuity of illness for diagnosis of neurodevelopmental disorders. *Sci Transl Med* 6, 265ra168.
 44. Stenson, P.D., Mort, M., Ball, E.V., Shaw, K., Phillips, A., and Cooper, D.N. (2014). The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. *Hum Genet* 133, 1-9.

45. Webb, A.J., Thorisson, G.A., Brookes, A.J., and Consortium, G.P. (2011). An informatics project and online "Knowledge Centre" supporting modern genotype-to-phenotype research. *Hum Mutat* 32, 543-550.
46. Landrum, M.J., Lee, J.M., Riley, G.R., Jang, W., Rubinstein, W.S., Church, D.M., and Maglott, D.R. (2014). ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res* 42, D980-985.
47. Amberger, J.S., Bocchini, C.A., Schiettecatte, F., Scott, A.F., and Hamosh, A. (2015). OMIM.org: Online Mendelian Inheritance in Man (OMIM(R)), an online catalog of human genes and genetic disorders. *Nucleic Acids Res* 43, D789-798.
48. Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 38, e164.
49. Beck, T.F., Mullikin, J.C., Program, N.C.S., and Biesecker, L.G. (2016). Systematic Evaluation of Sanger Validation of Next-Generation Sequencing Variants. *Clin Chem* 62, 647-654.
50. Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 74, 5463-5467.
51. Swerdlow, H., Wu, S.L., Harke, H., and Dovichi, N.J. (1990). Capillary gel electrophoresis for DNA sequencing. Laser-induced fluorescence detection with the sheath flow cuvette. *J Chromatogr* 516, 61-67.
52. Hunkapiller, T., Kaiser, R.J., Koop, B.F., and Hood, L. (1991). Large-scale and automated DNA sequence determination. *Science* 254, 59-67.
53. Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* 6, 80-92.
54. Tucker, T., Marra, M., and Friedman, J.M. (2009). Massively parallel sequencing: the next big thing in genetic medicine. *Am J Hum Genet* 85, 142-154.
55. Voelkerding, K.V., Dames, S., and Durtschi, J.D. (2010). Next generation sequencing for clinical diagnostics-principles and application to targeted resequencing for hypertrophic cardiomyopathy: a paper from the 2009 William Beaumont Hospital Symposium on Molecular Pathology. *J Mol Diagn* 12, 539-551.
56. Su, Z., Ning, B., Fang, H., Hong, H., Perkins, R., Tong, W., and Shi, L. (2011). Next-generation sequencing and its applications in molecular diagnostics. *Expert Rev Mol Diagn* 11, 333-343.
57. Sboner, A., Mu, X.J., Greenbaum, D., Auerbach, R.K., and Gerstein, M.B. (2011). The real cost of sequencing: higher than you think! *Genome Biol* 12, 125.
58. Yandell, M., Huff, C., Hu, H., Singleton, M., Moore, B., Xing, J., Jorde, L.B., and Reese, M.G. (2011). A probabilistic disease-gene finder for personal genomes. *Genome Res* 21, 1529-1542.
59. Biesecker, L.G., Mullikin, J.C., Facio, F.M., Turner, C., Cherukuri, P.F., Blakesley, R.W., Bouffard, G.G., Chines, P.S., Cruz, P., Hansen, N.F., et al. (2009). The ClinSeq Project: piloting large-scale genome sequencing for research in genomic medicine. *Genome Res* 19, 1665-1674.

60. Kohler, S., Vasilevsky, N.A., Engelstad, M., Foster, E., McMurry, J., Ayme, S., Baynam, G., Bello, S.M., Boerkoel, C.F., Boycott, K.M., et al. (2017). The Human Phenotype Ontology in 2017. *Nucleic Acids Res* 45, D865-D876.
61. Directors, A.B.o. (2012). Points to consider in the clinical application of genomic sequencing. *Genet Med* 14, 759-761.
62. Mamanova, L., Coffey, A.J., Scott, C.E., Kozarewa, I., Turner, E.H., Kumar, A., Howard, E., Shendure, J., and Turner, D.J. (2010). Target-enrichment strategies for next-generation sequencing. *Nat Methods* 7, 111-118.
63. Clark, M.J., Chen, R., Lam, H.Y., Karczewski, K.J., Chen, R., Euskirchen, G., Butte, A.J., and Snyder, M. (2011). Performance comparison of exome DNA sequencing technologies. *Nat Biotechnol* 29, 908-914.
64. Jones, M.A., Bhide, S., Chin, E., Ng, B.G., Rhodenizer, D., Zhang, V.W., Sun, J.J., Tanner, A., Freeze, H.H., and Hegde, M.R. (2011). Targeted polymerase chain reaction-based enrichment and next generation sequencing for diagnostic testing of congenital disorders of glycosylation. *Genet Med* 13, 921-932.
65. Anderson, M.W., and Schrijver, I. (2010). Next generation DNA sequencing and the future of genomic medicine. *Genes (Basel)* 1, 38-69.
66. Gowrisankar, S., Lerner-Ellis, J.P., Cox, S., White, E.T., Manion, M., LeVan, K., Liu, J., Farwell, L.M., Iartchouk, O., Rehm, H.L., et al. (2010). Evaluation of second-generation sequencing of 19 dilated cardiomyopathy genes for clinical applications. *J Mol Diagn* 12, 818-827.
67. Bell, C.J., Dinwiddie, D.L., Miller, N.A., Hateley, S.L., Ganusova, E.E., Mudge, J., Langley, R.J., Zhang, L., Lee, C.C., Schilkey, F.D., et al. (2011). Carrier testing for severe childhood recessive diseases by next-generation sequencing. *Sci Transl Med* 3, 65ra64.
68. Zhang, J., Chiodini, R., Badr, A., and Zhang, G. (2011). The impact of next-generation sequencing on genomics. *J Genet Genomics* 38, 95-109.
69. Frebourg, T. (2014). The challenge for the next generation of medical geneticists. *Hum Mutat* 35, 909-911.
70. Matthijs, G., Souche, E., Alders, M., Corveleyn, A., Eck, S., Feenstra, I., Race, V., Sisternans, E., Sturm, M., Weiss, M., et al. (2016). Guidelines for diagnostic next-generation sequencing. *Eur J Hum Genet* 24, 1515.
71. Majewski, J., Schwartzenuber, J., Lalonde, E., Montpetit, A., and Jabado, N. (2011). What can exome sequencing do for you? *J Med Genet* 48, 580-589.
72. Cirulli, E.T., Singh, A., Shianna, K.V., Ge, D., Smith, J.P., Maia, J.M., Heinzen, E.L., Goedert, J.J., Goldstein, D.B., and Center for, H.I.V.A.V.I. (2010). Screening the human exome: a comparison of whole genome and whole transcriptome sequencing. *Genome Biol* 11, R57.
73. Xue, Y., Ankala, A., Wilcox, W.R., and Hegde, M.R. (2015). Solving the molecular diagnostic testing conundrum for Mendelian disorders in the era of next-generation sequencing: single-gene, gene panel, or exome/genome sequencing. *Genet Med* 17, 444-451.
74. Yang, Y., Muzny, D.M., Reid, J.G., Bainbridge, M.N., Willis, A., Ward, P.A., Braxton, A., Beuten, J., Xia, F., Niu, Z., et al. (2013). Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N Engl J Med* 369, 1502-1511.

75. Strom, S.P., Lee, H., Das, K., Vilain, E., Nelson, S.F., Grody, W.W., and Deignan, J.L. (2014). Assessing the necessity of confirmatory testing for exome-sequencing results in a clinical molecular diagnostic laboratory. *Genet Med* 16, 510-515.
76. Biesecker, L.G., and Green, R.C. (2014). Diagnostic clinical genome and exome sequencing. *N Engl J Med* 371, 1170.
77. Bentley, D.R., Balasubramanian, S., Swerdlow, H.P., Smith, G.P., Milton, J., Brown, C.G., Hall, K.P., Evers, D.J., Barnes, C.L., Bignell, H.R., et al. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456, 53-59.
78. Fang, H., Wu, Y., Narzisi, G., O'Rawe, J.A., Barron, L.T., Rosenbaum, J., Ronemus, M., Iossifov, I., Schatz, M.C., and Lyon, G.J. (2014). Reducing INDEL calling errors in whole genome and exome sequencing data. *Genome Med* 6, 89.
79. Ajay, S.S., Parker, S.C., Abaan, H.O., Fajardo, K.V., and Margulies, E.H. (2011). Accurate and comprehensive sequencing of personal genomes. *Genome Res* 21, 1498-1505.
80. Xie, C., and Tammi, M.T. (2009). CNV-seq, a new method to detect copy number variation using high-throughput sequencing. *BMC Bioinformatics* 10, 80.
81. Medvedev, P., Fiume, M., Dzamba, M., Smith, T., and Brudno, M. (2010). Detecting copy number variation with mated short reads. *Genome Res* 20, 1613-1622.
82. Meynert, A.M., Bicknell, L.S., Hurler, M.E., Jackson, A.P., and Taylor, M.S. (2013). Quantifying single nucleotide variant detection sensitivity in exome sequencing. *BMC Bioinformatics* 14, 195.
83. Mardis, E.R. (2008). Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet* 9, 387-402.
84. ten Bosch, J.R., and Grody, W.W. (2008). Keeping up with the next generation: massively parallel sequencing in clinical diagnostics. *J Mol Diagn* 10, 484-492.
85. Shendure, J., and Ji, H. (2008). Next-generation DNA sequencing. *Nat Biotechnol* 26, 1135-1145.
86. Rothberg, J.M., Hinz, W., Rearick, T.M., Schultz, J., Mileski, W., Davey, M., Leamon, J.H., Johnson, K., Milgrew, M.J., Edwards, M., et al. (2011). An integrated semiconductor device enabling non-optical genome sequencing. *Nature* 475, 348-352.
87. Drmanac, R., Sparks, A.B., Callow, M.J., Halpern, A.L., Burns, N.L., Kermani, B.G., Carnevali, P., Nazarenko, I., Nilsen, G.B., Yeung, G., et al. (2010). Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* 327, 78-81.
88. Metzker, M.L. (2010). Sequencing technologies - the next generation. *Nat Rev Genet* 11, 31-46.
89. Gargis, A.S., Kalman, L., Bick, D.P., da Silva, C., Dimmock, D.P., Funke, B.H., Gowrisankar, S., Hegde, M.R., Kulkarni, S., Mason, C.E., et al. (2015). Good laboratory practice for clinical next-generation sequencing informatics pipelines. *Nat Biotechnol* 33, 689-693.
90. Gargis, A.S., Kalman, L., Berry, M.W., Bick, D.P., Dimmock, D.P., Hambuch, T., Lu, F., Lyon, E., Voelkerding, K.V., Zehnbaue, B.A., et al. (2012). Assuring the quality of next-generation sequencing in clinical laboratory practice. *Nat Biotechnol* 30, 1033-1036.

91. Ledergerber, C., and Dessimoz, C. (2011). Base-calling for next-generation sequencing platforms. *Brief Bioinform* 12, 489-497.
92. Chin, E.L., da Silva, C., and Hegde, M. (2013). Assessment of clinical analytical sensitivity and specificity of next-generation sequencing for detection of simple and complex mutations. *BMC Genet* 14, 6.
93. Valencia, C.A., Ankala, A., Rhodenizer, D., Bhide, S., Littlejohn, M.R., Keong, L.M., Rutkowski, A., Sparks, S., Bonnemann, C., and Hegde, M. (2013). Comprehensive mutation analysis for congenital muscular dystrophy: a clinical PCR-based enrichment and next-generation sequencing panel. *PLoS One* 8, e53083.
94. Lubin, I.M., Aziz, N., Babb, L.J., Ballinger, D., Bisht, H., Church, D.M., Cordes, S., Eilbeck, K., Hyland, F., Kalman, L., et al. (2017). Principles and Recommendations for Standardizing the Use of the Next-Generation Sequencing Variant File in Clinical Settings. *J Mol Diagn* 19, 417-426.
95. Berwouts, S., Morris, M.A., Girodon, E., Schwarz, M., Stuhmann, M., and Dequeker, E. (2011). Mutation nomenclature in practice: findings and recommendations from the cystic fibrosis external quality assessment scheme. *Hum Mutat* 32, 1197-1203.
96. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., et al. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* 17, 405-424.
97. den Dunnen, J.T., Dalgleish, R., Maglott, D.R., Hart, R.K., Greenblatt, M.S., McGowan-Jordan, J., Roux, A.F., Smith, T., Antonarakis, S.E., and Taschner, P.E. (2016). HGVS Recommendations for the Description of Sequence Variants: 2016 Update. *Hum Mutat* 37, 564-569.
98. Tack, V., Deans, Z.C., Wolstenholme, N., Patton, S., and Dequeker, E.M. (2016). What's in a Name? A Coordinated Approach toward the Correct Use of a Uniform Nomenclature to Improve Patient Reports and Databases. *Hum Mutat* 37, 570-575.
99. Kalman, L.V., Agundez, J., Appell, M.L., Black, J.L., Bell, G.C., Boukouvala, S., Bruckner, C., Bruford, E., Caudle, K., Coulthard, S.A., et al. (2016). Pharmacogenetic allele nomenclature: International workgroup recommendations for test result reporting. *Clin Pharmacol Ther* 99, 172-185.
100. Church, D.M., Schneider, V.A., Graves, T., Auger, K., Cunningham, F., Bouk, N., Chen, H.C., Agarwala, R., McLaren, W.M., Ritchie, G.R., et al. (2011). Modernizing reference genome assemblies. *PLoS Biol* 9, e1001091.
101. Pabinger, S., Dander, A., Fischer, M., Snajder, R., Sperk, M., Efremova, M., Krabichler, B., Speicher, M.R., Zschocke, J., and Trajanoski, Z. (2014). A survey of tools for variant analysis of next-generation genome sequencing data. *Brief Bioinform* 15, 256-278.
102. Ruffalo, M., LaFramboise, T., and Koyuturk, M. (2011). Comparative analysis of algorithms for next-generation sequencing read alignment. *Bioinformatics* 27, 2790-2796.
103. Li, H., and Homer, N. (2010). A survey of sequence alignment algorithms for next-generation sequencing. *Brief Bioinform* 11, 473-483.

104. Flicek, P., and Birney, E. (2009). Sense from sequence reads: methods for alignment and assembly. *Nat Methods* 6, S6-S12.
105. Merriman, B., Ion Torrent, R., Team, D., and Rothberg, J.M. (2012). Progress in ion torrent semiconductor chip based sequencing. *Electrophoresis* 33, 3397-3417.
106. Li, H., Ruan, J., and Durbin, R. (2008). Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res* 18, 1851-1858.
107. Homer, N., Merriman, B., and Nelson, S.F. (2009). BFAST: an alignment tool for large scale genome resequencing. *PLoS One* 4, e7767.
108. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9, 357-359.
109. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754-1760.
110. Li, R., Yu, C., Li, Y., Lam, T.W., Yiu, S.M., Kristiansen, K., and Wang, J. (2009). SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* 25, 1966-1967.
111. Chaisson, M.J., and Tesler, G. (2012). Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. *BMC Bioinformatics* 13, 238.
112. Oliver, G.R. (2012). Considerations for clinical read alignment and mutational profiling using next-generation sequencing. *F1000Res* 1, 2.
113. Yu, X., Guda, K., Willis, J., Veigl, M., Wang, Z., Markowitz, S., Adams, M.D., and Sun, S. (2012). How do alignment programs perform on sequencing data with varying qualities and from repetitive regions? *BioData Min* 5, 6.
114. Francey, L.J., Conlin, L.K., Kadesch, H.E., Clark, D., Berrodin, D., Sun, Y., Glessner, J., Hakonarson, H., Jalas, C., Landau, C., et al. (2012). Genome-wide SNP genotyping identifies the Stereocilin (STRC) gene as a major contributor to pediatric bilateral sensorineural hearing impairment. *Am J Med Genet A* 158A, 298-308.
115. Coonrod, E.M., Durtschi, J.D., Margraf, R.L., and Voelkerding, K.V. (2013). Developing genome and exome sequencing for candidate gene identification in inherited disorders: an integrated technical and bioinformatics approach. *Arch Pathol Lab Med* 137, 415-433.
116. Nielsen, R., Paul, J.S., Albrechtsen, A., and Song, Y.S. (2011). Genotype and SNP calling from next-generation sequencing data. *Nat Rev Genet* 12, 443-451.
117. Neuman, J.A., Isakov, O., and Shomron, N. (2013). Analysis of insertion-deletion from deep-sequencing data: software evaluation for optimal detection. *Brief Bioinform* 14, 46-55.
118. Gilissen, C., Hoischen, A., Brunner, H.G., and Veltman, J.A. (2012). Disease gene identification strategies for exome sequencing. *Eur J Hum Genet* 20, 490-497.
119. Stitzel, N.O., Kiezun, A., and Sunyaev, S. (2011). Computational and statistical approaches to analyzing variants identified by exome sequencing. *Genome Biol* 12, 227.
120. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156-2158.

121. Altmann, A., Weber, P., Bader, D., Preuss, M., Binder, E.B., and Muller-Myhsok, B. (2012). A beginners guide to SNP calling from high-throughput DNA-sequencing data. *Hum Genet* 131, 1541-1554.
122. Lyon, G.J., and Wang, K. (2012). Identifying disease mutations in genomic medicine settings: current challenges and how to accelerate progress. *Genome Med* 4, 58.
123. Fuentes Fajardo, K.V., Adams, D., Program, N.C.S., Mason, C.E., Sincan, M., Tifft, C., Toro, C., Boerkoel, C.F., Gahl, W., and Markello, T. (2012). Detecting false-positive signals in exome sequencing. *Hum Mutat* 33, 609-613.
124. Reese, M.G., Moore, B., Batchelor, C., Salas, F., Cunningham, F., Marth, G.T., Stein, L., Flicek, P., Yandell, M., and Eilbeck, K. (2010). A standard variation file format for human genome sequences. *Genome Biol* 11, R88.
125. Jalali Sefid Dashti, M., and Gamielien, J. (2017). A practical guide to filtering and prioritizing genetic variants. *Biotechniques* 62, 18-30.
126. Niroula, A., and Vihinen, M. (2016). Variation Interpretation Predictors: Principles, Types, Performance, and Choice. *Hum Mutat* 37, 579-597.
127. Flanagan, S.E., Patch, A.M., and Ellard, S. (2010). Using SIFT and PolyPhen to predict loss-of-function and gain-of-function mutations. *Genet Test Mol Biomarkers* 14, 533-537.
128. Ohanian, M., Otway, R., and Fatkin, D. (2012). Heuristic methods for finding pathogenic variants in gene coding sequences. *J Am Heart Assoc* 1, e002642.
129. Castellana, S., and Mazza, T. (2013). Congruency in the prediction of pathogenic missense mutations: state-of-the-art web-based tools. *Brief Bioinform* 14, 448-459.
130. Vihinen, M. (2012). How to evaluate performance of prediction methods? Measures and their interpretation in variation effect analysis. *BMC Genomics* 13 Suppl 4, S2.
131. Thusberg, J., Olatubosun, A., and Vihinen, M. (2011). Performance of mutation pathogenicity prediction methods on missense variants. *Hum Mutat* 32, 358-368.
132. Jun, G., Flickinger, M., Hetrick, K.N., Romm, J.M., Doheny, K.F., Abecasis, G.R., Boehnke, M., and Kang, H.M. (2012). Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. *Am J Hum Genet* 91, 839-848.
133. Berg, J.S., Amendola, L.M., Eng, C., Van Allen, E., Gray, S.W., Wagle, N., Rehm, H.L., DeChene, E.T., Dulik, M.C., Hisama, F.M., et al. (2013). Processes and preliminary outputs for identification of actionable genes as incidental findings in genomic sequence data in the Clinical Sequencing Exploratory Research Consortium. *Genet Med* 15, 860-867.
134. Holtzman, N.A. (2013). ACMG recommendations on incidental findings are flawed scientifically and ethically. *Genet Med* 15, 750-751.
135. McGuire, A.L., Joffe, S., Koenig, B.A., Biesecker, B.B., McCullough, L.B., Blumenthal-Barby, J.S., Caulfield, T., Terry, S.F., and Green, R.C. (2013). Point-counterpoint. Ethics and genomic incidental findings. *Science* 340, 1047-1048.

136. Allyse, M., and Michie, M. (2013). Not-so-incidental findings: the ACMG recommendations on the reporting of incidental findings in clinical whole genome and whole exome sequencing. *Trends Biotechnol* 31, 439-441.
137. Cummings, B.B., Marshall, J.L., Tukiainen, T., Lek, M., Donkervoort, S., Foley, A.R., Bolduc, V., Waddell, L.B., Sandaradura, S.A., O'Grady, G.L., et al. (2017). Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. *Sci Transl Med* 9.
138. Kremer, L.S., Bader, D.M., Mertes, C., Kopajtich, R., Pichler, G., Iuso, A., Haack, T.B., Graf, E., Schwarzmayr, T., Terrile, C., et al. (2017). Genetic diagnosis of Mendelian disorders via RNA sequencing. *Nat Commun* 8, 15824.
139. Codina-Sola, M., Rodriguez-Santiago, B., Homs, A., Santoyo, J., Rigau, M., Aznar-Lain, G., Del Campo, M., Gener, B., Gabau, E., Botella, M.P., et al. (2015). Integrated analysis of whole-exome sequencing and transcriptome profiling in males with autism spectrum disorders. *Mol Autism* 6, 21.
140. Tawamie, H., Martianov, I., Wohlfahrt, N., Buchert, R., Mengus, G., Uebe, S., Janiri, L., Hirsch, F.W., Schumacher, J., Ferrazzi, F., et al. (2017). Hypomorphic Pathogenic Variants in TAF13 Are Associated with Autosomal-Recessive Intellectual Disability and Microcephaly. *Am J Hum Genet* 100, 555-561.

Şekil No	Sayfa No
Şekil 1. Yeni nesil dizileme basamakları ve personel görev listesi	7
Şekil 2. Fenotipleme ve test seçimi ile ilgili önerilmiş bir algoritma örneği	13
Şekil 3. Wet-lab ve dry lab ana işlemlerinin şematizasyonu	21
Şekil 4. Yeni nesil dizilemede bir analizin temel basamakları	24
Şekil 5. NCBI sitesinde indirilen SRR001666.1 fastq dosyasının içerisindeki bir sekansın gösterimidir.	26
Şekil 6. Analizler sonrası üretilen dosya çeşitleri.	31
Şekil 7. Örnek bir SAM dosya formatı	36
Şekil 8. SAM dosyasının sütunlarının açıklamaları	36
Şekil 9. Samtools tutorial'ında yer alan hizalanmış ve referans genom arasındaki variant'ların tview aracı ile görselleştirilmiş halidir.	38
Şekil 10. VCF dosyasında her sütunun ne anlama geldiği bilgisidir	42
Şekil 11. Örnek bir VCF dosyası içeriğidir	42
Şekil 12. Tüm analiz basamaklarının özeti.	48
Şekil 13. Varyant anotasyonunda kullanılan filtreler.	51
Şekil 14. Fenotiplemeden raporlamaya algoritma önerisi	53
Şekil 15. Fenotiplemeden raporlamaya algoritma önerisi-2	55
Şekil 16. Raporlama sonrası takip önerisi	58

Tablo No

Tablo 1. KEGD yapmadan önce danışmada verilmesi gereken bilgiler	9
Tablo 2. Yeni Nesil Dizileme Uygulamaları.	11
Tablo 3. Yaygın genetik varyasyon çeşitleri ve ekzom veya genom dizileme ile bulunabilme durumları.	11
Tablo 4. TED'de ve TGD'de Kapsama ve Okuma Tavsiyeleri	18
Tablo 5. Illumina platformlarına ait bazı bilgiler	20

Bilgi Kutusu No

Bilgi Kutusu 1: FASTQ formatı	26
Bilgi Kutusu 2: Sequence alignment/map (sam/bam) format	36
Bilgi Kutusu 3: İkincil analizler	37
Bilgi Kutusu 4: Samtools v1.3.1	38
Bilgi Kutusu 5: Varyant Çağırma (Varyantların tespit edilmesi)	39
Bilgi Kutusu 6: Variant Call Format (VCF/BCF)	42
Bilgi Kutusu 7: Üçüncül analizler	47
Bilgi Kutusu 8: Varyant anotasyonu	49
Bilgi Kutusu 9: Varyant filtrasyonu ve değerlendirilmesi	54