

T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

**TP53 VE MEFV GENLERİNİN ILLUMINA VE MGI
TEKNOLOJİLERİ İLE DİZİLENMESİNDE DATA
KALİTELERİNİN KARŞILAŞTIRILMASI**

Ceren Aleyna KEÇECİ

YÜKSEK LİSANS TEZİ

Moleküler Biyoloji ve Genetik Anabilim Dalı

Moleküler Biyoloji ve Genetik Programı

Danışman

Prof. Dr. Muhammed Hamza MÜSLÜMANOĞLU

Ağustos, 2023

T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

**TP53 VE MEFV GENLERİNİN GENLERİNİN ILLUMINA VE MGI
TEKNOLOJİLERİ İLE DİZİLENMESİNDE DATA
KALİTELERİNİN KARŞILAŞTIRILMASI**

Ceren Aleyna KEÇECİ tarafından hazırlanan tez çalışması 24.08.2023 tarihinde aşağıdaki jüri tarafından Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Moleküler Biyoloji ve Genetik Anabilim Dalı, Moleküler Biyoloji ve Genetik Programı **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Prof. Dr. Muhammed Hamza
MÜSLÜMANOĞLU
Yıldız Teknik Üniversitesi
Danışman

Jüri Üyeleri

Prof. Dr. Muhammed Hamza MÜSLÜMANOĞLU, Danışman

Yıldız Teknik Üniversitesi

Prof. Dr. Semiha ERİŞEN, Üye

Yıldız Teknik Üniversitesi

Prof. Dr. Emre TEPELİ, Üye

Atlas Üniversitesi

Danışmanım Prof. Dr. Muhammed Hamza MÜSLÜMANOĞLU sorumluluğunda tarafımda hazırlanan *TP53* ve *MEFV* genlerinin Illumina ve MGI teknolojileri ile dizilenmesinde data kalitelerinin karşılaştırılması başlıklı çalışmada veri toplama ve veri kullanımında gerekli yasal izinleri aldığımı, diğer kaynaklardan aldığım bilgileri ana metin ve referanslarda eksiksiz gösterdiğimi, araştırma verilerine ve sonuçlarına ilişkin çarpıtma ve/veya sahtecilik yapmadığımı, çalışmam süresince bilimsel araştırma ve etik ilkelerine uygun davrandığımı beyan ederim. Beyanımın aksinin ispatı halinde her türlü yasal sonucu kabul ederim.

Ceren Aleyna KEÇECİ

İmza



Aileme

TEŞEKKÜR

Her konuda arkamda olan, bana yol gösteren ve motive eden sevgili anneme, ablalarım ve ailemin geri kalan diğer üyelerine çok teşekkür ederim.

Kan bağımlı olmasa da ailem gibi olan yine en büyük destekçilerimden biri Hamza Vedat Cengiz'e, her motivasyonum düştüğünde zor anlarımda yanımda olan çocukluk arkadaşlarım Çiçek Kına ve Fatma Orak'a, vazgeçmek istediğimde bir kere daha dene diyen arkadaşım Sena Doğan'a ve en önemlisi mezun olamayacağımı düşündüğümde son anda bana konusunu teklif edip yol gösteren, tezi yazmamda yardımcı olan Tuğçe Acar'a ve yeni konumla beraber son dönemimde danışmanım olan Hamza Müslümanoğlu hocama, tezimin biyoinformatik alanındaki çalışmalarında desteğini esirgemeyen Oğuzhan Kalyon'a teşekkürlerimi borç bilirim.

Eski danışmanım olup istediğim konuyu çalışmama izin veren ama bazı imkansızlıklardan dolayı bırakmak zorunda olduğum, bana proje yazma fırsatı verip yeni şeyler öğrenmeme vesile olan hocam Şenay Vural Korkut'a, uzmanlığımı almam konusunda katkısı olan herkese sonsuz teşekkürler.

Ceren Aleyna KEÇECİ

İÇİNDEKİLER

SİMGE LİSTESİ	vi
KISALTMA LİSTESİ	vii
ŞEKİL LİSTESİ	viii
TABLO LİSTESİ	ix
ÖZET	xi
ABSTRACT	xiii
1 GİRİŞ	1
1.1 Literatür Özeti	1
1.2 Tezin Amacı	4
1.3 Hipotez	4
2 YENİ NESİL DİZİLEME	5
2.1 Birinci Nesil DNA Dizileme Yöntemleri	5
2.2 Yeni Nesil Dizileme	8
2.3 Tüm Ekzom Sekanslama	12
2.4 Data Kalite Parametreleri	15
3 HAZIRLIKLAR	23
3.1 Tüm Ekzom Sekanslama (WES)	23
4 SONUÇ	38
4.1 Bulgular	38
4.2 İstatistiksel Hesaplamalar	46
4.3 Tartışma	47
4.4 Sonuç	51
4.5 Öneriler	51
KAYNAKÇA	52
TEZDEN ÜRETİLMİŞ YAYINLAR	54

SİMGE LİSTESİ

A	Adenine
AT	Adenine Thymine
C	Cytosine
dk	Dakika
EtOH	Ethanol
P	Fosfat
G1	Growth 1 (faz)
G	Guanine
GC	Guanine Cytosine
μ l	Mikrolitre
mL	Mililitre
ng	Nanogram
sn	Saniye
$^{\circ}$ C	Santigrat derece
NaOH	Sodyum Hidroksit
H ₂ O	Su
S	Synthesis (sentez fazı)
T	Thymine
%	Yüzde

KISALTMA LİSTESİ

ATP	Adenosine Triphosphate
ACMG	American College of Medical Genetics and Genomics
ASCII	American Standard Code for Information Interchange
BAC	Bacterial Artificial Chromosome
BÇ	Baz çifti
BAI	Binary Alignment Index
BAM	Binary Alignment Map
CNV	Copy Number Variation
DNA	Deoxyribonucleic Acid
DNTP	Deoxynucleotide Triphosphates
DNB	DNA Nanoball
FMF	Familial Mediterranean Fever
hg19	Human Genome 19
Indel	Insertion/ Deletion
IGV	Integrative Genomics Viewer
LINE	Long Interspersed Nuclear Element
NGS	Next Generation Sequencing
pH	Potential or Power of Hydrogen
Q20- Q30	Quality Score of 20-30
RNA	Ribonucleic Acid
SOLID	Sequencing by Oligonucleotide Ligation and Detection
SNP	Single Nucleotide Polymorphism
SSDNA	Single Strand Deoxyribonucleic Acid
PZR	Polimeraz Zincir Reaksiyonu
TAG	Tagmentation
TSB	Tagmentation Stop Buffer
VCF	Variant Call Format
YAC	Yeast artificial chromosome
YND	Yeni Nesil Dizileme
WES	Whole Exome Sequencing

ŞEKİL LİSTESİ

Şekil 2.1 DNA sentezi sırasında açığa çıkan fosfat grupları	9
Şekil 2.2 Oligo ligasyonu saptama yoluyla sekanslama (SOLID)	10
Şekil 2.3 <i>TP53</i> geninin kodladığı p53 proteinin rol aldığı mekanizmalar.....	14
Şekil 2.4 <i>MEFV</i> geni hücresel işlevleri	15
Şekil 2.5 Ham fastq datasının dönüştürülmesi	19
Şekil 2.6 Örnek fastq dizisi.....	19
Şekil 2.7 Örnek IGV görüntüsü.....	21
Şekil 2.8 Saptanan bir varyantın IGV görüntüsü.....	22
Şekil 4.1 Illumina ve MGI WES datalarının ortalama derinlik değerleri.....	39
Şekil 4.2 Illumina ve MGI cihazlarında çalışılan WES datalarından kesilmiş <i>TP53</i> ve <i>MEFV</i> genlerini içeren bölgelerin okumalarının ortalama derinliği.....	40
Şekil 4.3 Illumina ve MGI cihazlarında çalışılan <i>TP53</i> ve <i>MEFV</i> genlerinin toplam okuma sayıları.....	42
Şekil 4.4 Illumina ve MGI cihazlarıyla çalışılan WES datalarının Q20 değerleri	43
Şekil 4.5 Illumina ve MGI datalarının <i>TP53</i> ve <i>MEFV</i> genleri için Q20 değerleri.....	44
Şekil 4.6 Illumina ve MGI cihazlarından çıkan WES datalarının GC içeriği	45
Şekil 4.7 Illumina ve MGI cihazlarının <i>TP53</i> ve <i>MEFV</i> genleri için GC içerikleri.....	46

TABLO LİSTESİ

Tablo 2.1 Maxam-Gilbert kimyasalları.....	6
Tablo 2.2 Shotgun dizileme ile elde edilen sekansların okunması örneği	6
Tablo 3.1 Enzimatik fragmentasyon karışımı	23
Tablo 3.2 Enzimatik fragmentasyon reaksiyon koşulları.....	24
Tablo 3.3 Ligasyon Karışımı.....	24
Tablo 3.4 Yakalama öncesi PZR programı	25
Tablo 3.5 Yakalama öncesi PZR karışımı.....	25
Tablo 3.6 Hibridizasyon karışımı.....	27
Tablo 3.7 Hibridizasyon reaksiyon koşulları	27
Tablo 3.8 Yakalama sonrası PZR reaksiyon koşulları	28
Tablo 3.9 Yakalama sonrası PZR reaksiyon karışımı	28
Tablo 3.10 Denatürasyon reaksiyonu koşulları.....	29
Tablo 3.11 Tek zincir sirkülerizasyonu karışımı.....	29
Tablo 3.12 Tek zincir sirkülerizasyonu reaksiyon koşulları	30
Tablo 3.13 Enzimatik parçalama karışımı	30
Tablo 3.14 Enzimatik parçalama reaksiyonu koşulları	30
Tablo 3.15 DNB oluşturma reaksiyonu 1	31
Tablo 3.16 DNB oluşturma reaksiyonu 1 koşulları	31
Tablo 3.17 DNB oluşturma reaksiyonu 2	32
Tablo 3.18 DNB oluşturma reaksiyonu 2 koşulları	32
Tablo 3.19 DNB yükleme karışımı	32
Tablo 3.20 Sekans kartuşu hazırlanışı.....	33
Tablo 3.21 TAG termal döngü programı	33
Tablo 3.22 TAG master karışımı	34
Tablo 3.23 PTC (Etiketleme sonrası temizlik) termal döngü programı.....	34
Tablo 3.24 PZR master karışımı	35
Tablo 3.25 Kullanılması gereken adaptör hacim hesaplamaları	35
Tablo 3.26 Etiketli DNA'nın termal döngü programı.....	35
Tablo 3.27 Önerilen yükleme konsantrasyonları	36
Tablo 3.28 Önerilen hacim.....	37
Tablo 4.1 Illumina ve MGI WES datalarının ortalama derinlikleri	39
Tablo 4.2 Illumina ve MGI datalarının <i>TP53</i> ve <i>MEFV</i> genleri için ortalama derinlikleri.....	40

Tablo 4.3 Illumina ve MGI datalarının <i>TP53</i> ve <i>MEFV</i> genlerinin toplam okuma sayıları.....	41
Tablo 4.4 Illumina ve MGI cihazlarından çıkan WES datalarının Q20 değerleri	42
Tablo 4.5 Illumina ve MGI cihazlarının <i>TP53</i> ve <i>MEFV</i> genleri için Q20 değerleri	43
Tablo 4.6 Illumina ve MGI cihazlarından çıkan WES datalarının GC içeriği.....	45
Tablo 4.7 Illumina ve MGI cihazlarının <i>TP53</i> ve <i>MEFV</i> genleri için GC içerikleri	46
Tablo 4.8 T testi uygulanarak bulunan p değerleri	47



TP53 ve MEFV Genlerinin Illumina ve MGI Teknolojileri ile Dizilenmesinde Data Kalitelerinin Karşılaştırılması

Ceren Aleyna KEÇECİ

Moleküler Biyoloji ve Genetik Anabilim Dalı

Yüksek Lisans Tezi

Danışman: Prof. Dr. Muhammed Hamza MÜSLÜMANOĞLU

1990 yılında başlayan ve 2003'te biten İnsan Genom Projesi ile insan genomunun dizilenmesiyle beraber, genomda meydana gelen ve birçok hastalığa neden olan mutasyonlar saptanabilmiştir. Bu projeye beraber, beklenen gen sayısından ne kadar az sayıda gene sahip olduğumuz, genomların aynı tür ve hatta türler arası benzerlik oranının ne kadar fazla olduğu gibi bilgilere de sahip olmuş olduk. İnsan genomunun dizilenmesi tedavi, tanı ve tarama için birçok alanda gereklidir ve kullanılmaktadır. İnsan genomunu dizileme konusunda hızlı, güvenilir ve daha az maliyetli olan günümüzde kullanılan YND yönteminin kullanıldığı birçok dizileme teknolojisi ve cihazı mevcuttur. Bu çalışmada; Illumina NovaSeq 6000 ve MGI DNB Seq-G400 cihazlarından çıkan tüm ekzom sekanslama (WES) datalarını, GC içeriği, ortalama derinlik ve Q20 değerleri gibi kalite parametrelerine göre kıyaslayarak amaca uygun olarak hangisinin seçilebileceğini araştırdık. Örneklem grubunu 12 Illumina ve 12 MGI cihazında çalışılan WES dataları olarak dizayn ettik. Bunun yanında iki gen seçerek (*TP53-MEFV*), gen bazında okumaların istatistiklerini de kıyasladık. Elde edilen sonuçlar, python ve excel ile t testi uygulanarak incelendi. MGI ve Illumina teknolojisi kullanılarak sekanslanan hem WES datalarının hem de kesilip iki gene özel bakılan dataların kalite parametrelerinin arasındaki farkın istatistiki olarak anlamlı farklar elde ettik. GC içerikleri Illumina WES

datalarında %51-52'den farklı deęerler vermezken, MGI'dan elde edilen WES datalarında bu deęer %48-49 ve kesilmiş datalarda ise Illumina'da 56-57% ve MGI'da %55'tir. Bunun yanı sıra ortalama derinlik hesaplandığında genel olarak hem kesilmiş (Illumina-55,4 ve MGI-29,9) hem WES datalarında (Illumina-58,85 ve MGI-48,52) Illumina datalarının daha yüksekken, MGI datalarında göreceli olarak düşük kalmıştır. Q20 deęerlerinde ortalama deęer olarak sırasıyla WES ve kesilmiş Illumina datalarının (98,80-98,87) MGI datalarından (97,47-97,29) daha yüksek olduęu görülmüştür. Sonuç olarak; bu çalışmada kıyaslanan data kalite parametreleri dışında başka parametreler olduęu iki teknoloji arasında yapılacak tercihlerde araştırma amaçları, örneklerin özellikleri ve imkanlar gibi birçok faktör göz önünde bulundurulmalıdır.

Anahtar Kelimeler: ILLUMINA, MGI, YND, GC içeriği, ortalama derinlik

Comparison of Data Quality in Sequencing of TP53 and MEFV Genes with Illumina and MGI Technologies

Ceren Aleyna KEÇECİ

Department of Molecular Biology and Genetics

Master Thesis

Supervisor: Prof. Dr. Muhammed Hamza MUSLMANOGLU

With the sequencing of the human genome with the Human Genome Project, which started in 1990 and ended in 2003, mutations in the genome that cause many diseases have been identified. With this project, we have gained information such as how few genes we have than the expected number of genes, and how high the similarity ratio of genomes is between the same species and even between species. Sequencing of the human genome is necessary and used in many fields for treatment, diagnosis and screening. There are many sequencing technologies and devices that use the NGS method used today, which is fast, reliable and less costly for sequencing the human genome. In this study; We compared all exome sequencing (WES) data from Illumina NovaSeq 6000 and MGI DNB Seq-G400 devices according to quality parameters such as GC content, average depth and Q20 values, and investigated which one could be selected for the purpose. We designed the sample group as WES data studied on 12 Illumina and 12 MGI devices. We also compared the statistics of reads by gene by selecting two genes (TP53-MEFV). Obtained results were analyzed by applying t test with python and excel. We obtained statistically significant differences between the quality parameters of both the WES data sequenced using MGI and Illumina technology and the data that was cut and looked at specifically for two genes. While GC contents do not give values different than

51-52% in Illumina WES data, this value is 48-49% in WES data obtained from MGI and 56-57% in Illumina and 55% in MGI in truncated data. In addition, when the average depth is calculated, it was higher in Illumina data in both cut (Illumina-55.4 and MGI-29.9) and WES data (Illumina -58.85 and MGI-48.52), while it was relatively low in MGI data. . In Q20 values, the mean value of WES and truncated Illumina data (98.80-98.87), respectively, was higher than MGI data (97.47-97.29). In conclusion; many factors such as research objectives, characteristics of the samples and possibilities should be considered in the choices to be made between the two technologies, where there are other parameters other than the data quality parameters compared in this study.

Keywords: ILLUMINA, MGI, NGS, GC content, average dept



1.1 Literatür Özeti

İnsan genomunun gizeminin çözülmesi, bu genomun içinde gizli diziyi sekanslamamız ve proteine kadar olan yolculuğunu incelememizle mümkün olmuştur. İnsan genomu dört temel bazdan (A,T,G,C) oluşan, bu bazların çeşitli kombinasyonlarıyla oluşan sekansa sahip bütün bir dizidir. Bu dizinin insanlar arasında insan genom projesiyle beraber %99.9 benzer olduğu saptansa da tek nükleotid polimorfizmleri, kopya sayısı değişiklikleri, küçük insersiyon ve delesyonlar gibi faktörlerle birbirimizden farklı olduğumuz bulunmuştur. Dizideki bu farklılıklar saç rengimiz, göz rengimiz gibi fenotipik farklılıkları oluşturduğu gibi bazen olumsuz sonuçlar da doğurabilir. Mutasyonlar nedeniyle insan genom sekansında meydana gelen değişiklikler hastalıklar ve ilaç direnciyle sonuçlanabilir. Bu mutasyonlar tümör baskılayıcı genlerde, proto-onkogenlerde, belirli bir fenotiple ilişkili olduğu bilinen genlerde, hotspot bölgelerde, çerçeve kaymasına neden olacak yerlerde, proteinin erken sonlanmasına neden olup stop kodunu oluşturacakları yerlerde daha ciddi sonuçlar doğurabilmektedir. Yüzyıllardan beri, insan genomunu dizileyebilmek için çeşitli yöntemler kullanılmıştır. Bu yöntemlere örnek olarak tarihsel sırayla; Maxam- Gilbert yöntemi, Sanger sekanslama, ikinci nesil dizileme ve üçüncü nesil dizileme yöntemleri örnek verilebilir. Maxam-Gilbert yöntemi kullanılan kimyasallardan dolayı ve sekanslanan dizinin kısıtlı boyutundan dolayı çok verimli ve tercih edilir bir yöntem değildir. Sanger sekanslama hala daha YND datalarının doğrulanması için kullanılan ama yine boyut dolayısıyla kısıtlayıcı özelliğe sahip bir yöntemdir. İkinci nesil dizileme olan, çalışmamızda kullandığımız yöntemdir. Yüzlerce örneğin eş zamanlı olarak daha az maliyetle sekanslandığı bu yöntem çok daha kısa bir sürede, zararlı kimyasallar kullanmadan sekanslayabilmeyi mümkün kılmıştır. YND genetik araştırmalar için devrim sayılabilecek bir buluştur. Illumina ve MGI birçok YND teknolojisine ve birbirinden farklı özellikleri olan cihazlara sahip firmalardır.

Illumina 1998 yılında Kaliforniya’da kurulmuş genetik çalışmalar, analizleri ve genomik teknolojilerde öncü olan şirketlerden biridir. Yüksek verimlilik ve yüksek hassasiyetle çeşitli genom dizileme platformları mevcuttur. Genom dizileme teknolojisinde sentez

tabanlı yaklaşımı kullanır. DNA ilk olarak fragmente edilir, ardından adaptörlerle işlenir. Bu adaptörler DNA parçalarının amplifikasyonu ve dizilenmesinde kullanılan kısa DNA dizileridir. Adaptörlerin; DNA parçalarından kütüphane oluşturulması işleminde DNA'nın çift sarmal yapısına dönüştürülmesinde, DNA amplifikasyonunda adaptörler üzerine bulunan DNA bölgesinin hedeflenerek çoğaltılmasında, flow cecele kütüphanenin bağlanmasında önemli rolleri vardır. Adaptörlerle işlenen fragmente edilmiş DNA parçaları, flow cecele bağlanır. DNA sentez enzimi ile amplifikasyonu gerçekleştirilir. Sentez tabanlı sekanslamada, 4 nükleotid de (A,T,G,C) sırayla eklenir ve her eklenme sırasında floresan ışımaya kaydedilir. Bu şekilde ışımalar takip edilerek baz ekleme sırasına göre bir sekans çıkarılır [1].

MGI; 2016 yılında kurulan Çin'de olan BGI'nin yan kuruluşudur. Genomik sekanslama, kitler, cihazlar ve biyoteknolojik çözümler sunan bir şirket olmakla birlikte, Illumina gibi öncü firmalardan biridir. Çeşitli platformları bulunmakla birlikte, öne çıkan DNBSEQ platformları mevcuttur. "DNA Nanoball Sequencing" in kısaltması olan DNBSEQ mevcut dizileme teknolojileri ve platformlardan farklı bir yöntem kullanarak yüksek verimlilik ve eş zamanlı yüzlerce örneğin senkronize dizilenmesini sağlayan bir teknolojidir. DNB oluşturma aşaması için spesifik adaptörler kullanılır ve işlenir. Her bir DNB yoğun sayıda kopya içeren ve dizilenmeye hazır DNA formunu temsil eder. İlk olarak DNA fragmentleri tek iplikli forma getirilir sonra DNB oluşturulur ve flow celle yüklenir. DNB hassas bir form olduğu için yapım aşaması zor olsa da diğer dizileme teknolojilerinden farklı olarak düşük maliyetlidir, yüksek doğruluk ve düşük hata oranı içerir [2].

YND teknolojisi göreceli olarak diğer teknolojilere oranla daha yeni bir alan olduğu için Türkiye'de de çok fazla kullanım yerine sahip değildir. Birçok laboratuvarında, araştırma merkezlerinde yeni yeni oturmuş bir düzende kullanılmaya başlanmış olsa da araştırma ve vaka çalışmaları için kanser gibi köklü araştırma alanları kadar yayına sahip değildir. Özellikle YND teknolojisinden elde edilen dataların kıyaslandığı, iki farklı teknolojinin kullanıldığı çalışmalar sınırlı sayıdadır. Literatür taramasında, Ion Torrent, Illumina, Roche ve MGI gibi firmaların YND platformlarının kullanılarak data kalite parametrelerinin karşılaştırıldığı çalışmalar mevcut olmakla birlikte; örnek sayısının kısıtlılığı, kullanılan kitlerin farklılığı, kıyaslanan data kalite parametrelerinin çeşitli olmaması gibi belli başlı dezavantajlar mevcuttur. Yapılan çalışmalar varyant üzerine yoğunlaşmış ve genellikle genetik varyasyon analizi yapmak için kullanılan yazılımların

(GATK gibi) kıyaslanması gibi alanlara katkı sağlamıştır. Elde edilen ham datanın kıyaslanması ile platformların amaca uygun kıyaslanmasına dair az sayıda çalışma mevcuttur. Data kalite parametresi olarak literatür çalışmaları; Q20,Q30 değerleri, ortalama derinlik, GC içeriği, duplikasyon seviyesi, okunamayan baz yüzdesi, ortalama uzunluk gibi parametreleri içermektedir. Q20 değeri 10^{-2} hata oranına karşılık gelen dizilemedeki hata oranı olasılığını veren bir değerdir. Daha yüksek Q20 değerleri, daha az hata oranı dolayısıyla daha güvenilir dizileme anlamına gelir. Q30 değeri bu olasılığın 10^{-3} değeridir. Ortalama derinlik sekanslamak istenen bölgenin ortalama okuma sayısıdır, bir bölge ne kadar çok okunursa o kadar güvenilir data vereceğinden yüksek ortalama derinlik güvenilir data demektir. GC içeriği, sekanslanan dizideki G ve C'nin yüzdesidir ve primer etkinliği ve genomik bölgenin yapısal özelliğini etkileyebileceğinden önemlidir. Duplikasyon seviyesi, birbirinin aynısı olan okumaların ya da tekrar eden okumaların yüzdesidir. Yüksek duplikasyon seviyesi, amplifikasyon hatalarını gösterebilir. Bu yüzden tekrar eden okumaların düşük olması, daha iyi veri kalitesi ve daha güvenilir sonuçlar olarak yorumlanır. Okunamayan baz yüzdesi ise, belirlenemeyen veya hatalı baz çiftlerinin yüzdesidir. Bu yüzde baz kalitesinin düşüklüğü, dizileme hatası veya veri işleme hatalarının belirteci olarak yorumlanabilir. Bu yüzden daha yüksek bir okunamayan baz yüzdesi genomik verilerin güvenilirliğini etkileyebileceğinden tercih edilmez [3].

Bu çalışmamızda ilk olarak; yeni nesil dizileme bölümümüzde geçmiş dönemden günümüze gelen önemli genom sekanslama yöntemlerinden ve ardından YND teknolojilerinden bahsettik. Biz çalışmamızda iki farklı öncü firmanın iki farklı sekanslama teknolojisini (Illumina Novaseq 6000 ve MGI DNBSseq G400) kullanarak dizilediği tüm ekzom sekanslama datasının belirli data kalite parametrelerine göre kıyaslanmasını gerçekleştirdik. Bunu yaparken geniş bir alandansa seçilmiş iki genin gen bazında okumasının kıyaslanmasını yapabilmek için *TP53* ve *MEFV* genlerini datadan keserek özel olarak inceledik. YND teknolojisinin ardından, tüm ekzom sekanslamanın ne olduğu/ neden yapıldığı ve seçilen iki genin rollerinden bahsettikten sonra, seçtiğimiz data kalite parametreleri tanıttık. Bu data kalite parametrelerini kıyaslarken kullandığımız yazılımları (Python) ve uygulamaları (IGV) tanımladık. Hazırlık kısmında WES datalarını elde edebilmek için yürütülen laboratuvar çalışmalarımızı derledik. Ardından sonuç ve öneriler kısmında elde ettiğimiz 12 Illumina ve 12 MGI datalarının GC içeriği,

ortalama derinlik ve Q20 deęerleri ve bu deęerlerin t testi sonucunda istatistiki olarak deęerlendirilmesini sunduk.

1.2 Tezin Amacı

Nasıl gözükeceęimiz, hangi hastalıklara sahip olacaęımız, hangi hastalıklara yatkın ve taşıyıcı olduęumuz, derimizin renginin ne olacaęı gibi birçok bilgi DNA'mızda saklıdır. DNA'mızın şifresinin çözümlenmesi tıp alanında hem tanı hem de tedavi için önemli rol oynamış ve hayatımızı kolaylaştırmıştır. İnsan genom projesiyle beraber ilk nesil dizileme ile başlayan insan genomunun dizilenmesi, şu an ikinci ve hatta üçüncü nesil dizileme teknolojileriyle devam etmektedir. Teknoloji ilerledikçe başlangıç DNA miktarının azalması, tekrar bölgeleri ve yapısal deęişiklikler gibi dizilenmesi zor olan bölgelerin dizilenmesindeki doğruluğun artması, maliyetin azalması, dizileme hızının artması ve aynı anda dizilenen örnek miktarının artmasıyla beraber DNA dizilenmesi daha kolay ve yapılabilir hale gelmiştir.

Bu tezde; farklı firmalara ve farklı teknolojilere sahip iki cihazın (Illumina NovaSeq 6000 ve MGI DNB Seq-G400) tüm ekzom sekanslama datasındaki farklılıklarını görmeyi amaçladık. Data kalitesi parametreleri olarak fastqc raporlarını kıyaslayarak hem genel bir kıyaslama yapmayı hem de bam dosyasından *MEFV* ve *TP53* genlerini çekerek gen bazındaki okuma kalitesini karşılaştırarak hangi teknolojinin kıyaslanan kalite parametrelerinde daha iyi sonuç vereceęini belirlemeyi amaçladık.

1.3 Hipotez

Illumina NovaSeq 6000 ile yapılan tüm ekzom sekanslama datasının kalitesi MGI DNB Seq-G400 cihazından alınan datadan, araştırdıęımız ortalama derinlik, Q20 deęeri ve GC içerięi gibi parametrelerde göreceli olarak daha iyi data kalitesine sahiptir.

İnsan Genom Projesi, 1990 yılında başlayıp 2003 yılında biten en büyük bilimsel başarılarından biri olarak düşünülen ve bilimde birçok alanda çığır açılmasına yol açan bir projedir. İnsan genomu dizilenmeden önce; çok yüksek sayıda gen olduğu, kromozomlardaki gen sayılarının homojen olduğu ve aynı türdeki / türler arası farklılığın fazla olduğu düşünülmekteydi. Projeden sonra yaklaşık 20 bin gen olduğu, protein kodlayan bölgelerin tüm genomun yalnızca %2'sini oluşturduğu, genlerin kromozomlarda eşit olarak dağılmadığı ve bireyler arasında genomun %99.9 aynı olduğu farklılığın SNP (single nucleotide polymorphisms)'ler ve CNV (copy number variants)'lerin oluşturduğu ortaya çıkmıştır. Toplamda 3.1 milyar nükleotid saptanmasına rağmen bunun sadece %2'sinin protein kodladığını saptandıktan sonra yeni bir olgu olan alternatif splicing ortaya çıkmıştır. Alternatif splicing sayesinde; tek bir genden birden fazla mRNA molekülü oluşturulmaktadır. Gende yer alan ekzon bölgeleri farklı şekillerde birleşerek farklı mRNA molekülleri oluşturabilirler. Farklı mRNA'lar farklı proteinler kodladığından, az sayıda genden nasıl bu kadar çok sayıda protein meydana getirildiği açıklanmış oldu. Bu projeye beraber, insan genomunun dinamik olduğu da öğrenilmiş oldu. Genomun %50'sinin LINE ve Alu sekansları gibi transposable elementler ya da diğer tekrar DNA sekansları olduğu belirlendi [4].

İnsan genom projesi Maxam-Gilbert yöntemi, Shotgun dizileme ve Sanger dizileme gibi yöntemlerin beraber kullanılmasıyla tamamlanmıştır [5].

2.1 Birinci Nesil DNA Dizileme Yöntemleri

2.1.1 Maxam-Gilbert Yöntemi

Maxam-Gilbert yöntemi adını yöntemi bulanlardan alan bir DNA dizileme yöntemidir. Allan Maxam ve Walter Gilbert tarafından geliştirilen bu yöntemde; DNA belirli kimyasallarla modifiye edilir ardından spesifik bazlarla kesilerek, dizisi belirlenir.

DNA'da bulunan pürinler (Adenin ve Guanin) dimetil sülfat, primidinler (Sitozin, Timin ve Urasil) Hidrazin ile kırılırlar. Hangi bazın kırılacağı ortamın pH durumuna göre belirlenir. Pürinler dimetil sülfat ile kırılırken; eğer ortam bazik ise Guanin bazından,

asidik ise Adenin bazından kırılır. Aynı şekilde primidinler hidrazin ile kırılırken yüksek tuz derişimi ve bazik ortamda sitozinden kırılır ve zinciri kırmak için kullanılan kimyasal ise Piperidin'dir. DNA çift iplikçığı ayrılır ve dizisi saptanacak olan DNA, 5'ucundan P ile etiketlenir ya da floresan bir boya ile boyanır. Toplamda 4 tüp hazırlanır, her tüpte farklı bazın kırılması için tepkimeler gerçekleşir. Ardından jel elektroforezinde yürütülerek, kısa olan fragmentten yani alttan başlayarak baz diziliş saptanır [6]. Tablo 2.1.'de kullanılan kimyasallar ve bu kimyasalların hangi bazın kesimi için kullanıldığı verilmiştir.

Tablo 2.1 Maxam-Gilbert Kimyasalları

Baz	Kimyasal modifikasyon için kullanılan kimyasal
Guanin	Dimetil sülfat
Adenin+ Guanin	Asit
Sitozin + Timin	Hidrazin
Sitozin	Hidrazin+Baz

Maxam-Gilbert yöntemi; nispeten karmaşık tekniğe sahip olması, tehlikeli kimyasalların kapsamlı kullanımının gerekliliği, 500'den fazla baz çiftinin analiz edilmesi için kullanılamaması gibi dezavantajlara sahiptir [7].

2.1.2 Shotgun Dizileme

Bu yöntem Sangerle ya da diğer yöntemlerle dizilenemeyen uzun DNA dizileri için kullanılan, insan genom projesinde de büyük rol oynayan yöntemdir. Büyük boyutlardaki DNA parçaları BAC (Bacterial Artificial Chromosomes) ve YAC (Yeast Artificial Chromosomes) gibi vektörler kullanılarak dizilenir. İlk olarak hücrelerden DNA izolasyonu gerçekleştirilir, ardından DNA random olarak restriksiyon enzimleriyle fragmente edilir (300-800 bp). Bakteri ve maya gibi vektörlere insert edilip uygun konakçılara klonlanır. Ardından elde edilen diziler çoğaltılarak örtüşen dizilerle, dizilenmek istenen orijinal dizi elde edilir [8]. Tablo 2.2'de elde edilen kısa sekansların birleştirilerek nasıl tek sekans edilebileceği verilmiştir.

Tablo 2.2 Shotgun dizileme ile elde edilen sekansların okunması örneği

Orijinal Dizi	ACTGGCATTGCAAGCT
Elde edilen ilk sekans	ACTGGCATTGC.....AAGCT
Elde edilen ikinci sekans	ACTGGCA.....ATTGCT
Birleştirilmiş sonucu	ACTGGCATTGCAAGCT

Shotgun dizileme yöntemi daha uzun DNA'ların fragmente edilerek dizilenmesi imkanı sağladığından avantaja sahip olsa da, referans genomu gerektirmesi, tekrarlayan dizileri bir araya getirmede eksik kalması, bilgisayar işlem gücüne fazla ihtiyaç duyması ve tekrarlayan bölgeler yüzünden montaj aşamasından hataya açık olması gibi dezavantajlara sahiptir [9].

2.1.3 Sanger Sekanslama

Sanger Sekanslama ismini Frederick Sanger'den alan uzun yıllar boyunca kullanılan en yaygın DNA sekanslama yöntemidir. Normal PZR reaksiyonuna dNTP'lerin yanı sıra ddNTP'lerin eklenmesiyle gerçekleştirilir. DdNTP'ler (di-deoxynucleotide triphosphates) dNTP'lerin modifiye edilmesiyle elde edilir. Normal dNTP'lerden farkı, 3'-OH gruplarının olmaması ve floresan (veya radyoaktif) işaretli olmalarıdır. Zincir sonunda 3'-OH grubu olmamasından dolayı iki nükleotit arasında fosfodiester bazı oluşmaz ve yanına başka baz gelememesinden dolayı sekanslama sona erer. Bu nedenle, Sanger sekanslama zincir sonlandırma yöntemi olarak da bilinir. Maxam-Gilbert yöntemindeki gibi yine dört tüp hazırlanır, her tüpe dNTP'ler, primerler, buffer, DNA polimeraz koyulur, ardından her tüpe bir tanesi denk gelecek şekilde ddNTP'ler eklenir. Sekanslamanın devam edebilmesi ve değişik boyutlarda sekans verisi elde edebilmek için eklenen ddNTP'lerin miktarının dNTP'lerden daha az olması gerekir. Reaksiyon tamamlandıktan sonra, kapiller jel elektroforezi aşaması gerçekleştirilir. Jeldeki görüntü yine shotgun dizilemedeki gibi aşağıdan yukarı (kısa fragmentten uzun fragmente) doğru sıralanır ve dizileme sağlanır.

Eski teknikte kullanılan radyoaktif işaretleme yönteminin taşıdığı risk üzerine teknik geliştirilerek floresan işaretleme de Sanger yöntemi belirli kısıtlamalar ve dezavantajlara sahiptir.

- Sanger dizileme 300-1000 bp aralığında sadece kısa fragmentleri dizileyebilir.
- Dizi kalitesi ilk 10-40 bazda iyi değildir çünkü primerin denk geldiği bölgedir, düşük kaliteli bölgeler Phred gibi kalite tahmini yazılımı kullanılması gerekir ve okuma kalitesi açısından şüphelidir.
- Shotgun gibi klonlama yönteminin kullanıldığı sekanslamadan sonra Sanger ile dizinin belirlenmesi istenirse, dizileme vektörün parçalarını da içerebilir.
- Sekans kalitesi 700-900 baz okumasından sonra düşer [10].

İnsan genom projesinde bir arada kullanılan Maxam-Gilbert yöntemi, Shotgun Sekanslama ve Sanger Sekanslama gibi birinci nesil dizileme (first generation sequencing) olarak adlandırılan yöntemler günümüzde hala bazı alanlarda nadiren ya da belirli bir değişimi doğrulama için kullanılmalarına rağmen; uzun zaman alması, maddi olarak pahalıya mal olması ve tekrar bölgeleri gibi sekanslaması zor olan bölgelerdeki başarısızlıkları, tek başına insan genomu gibi uzun ve kompleks sekansları dizilememeleri gibi zorlukları nedeniyle yerini yeni nesil dizileme (ikinci/üçüncü nesil dizileme) olarak adlandırılan yeni teknolojilere bırakmıştır.

2.2 Yeni Nesil Dizileme

İlk bölümde anlatılan, birinci nesil dizileme yöntemlerinin (Maxam-Gilbert, Shotgun ve Sanger dizileme) insan genomu gibi uzun ve tekrarlayan bölgelerin, tek nükleotid polimorfizmlerin ve kopya sayısı değişimlerinin varlığı nedeniyle karmaşıklığından dolayı dizilenmesi zor türlerde kullanılamaması, belirli boyutta DNA'nın sekanslanması, kullanılan kimyasalların riski, zaman ve nakit kaybı olması gibi dezavantajlarından sonra hızlıca gelişen ve büyük etki uyandıran yeni nesil dizileme teknolojileriyle beraber insan genomu gibi kompleks DNA'ların dahi dizilenmesi günler sürmeden daha uygun fiyata dizilenebilir hale gelmiştir. Yeni nesil dizileme teknolojisi PZR yöntemi emülsiyon PZR'ı ve köprü PZR'ı olarak ikiye ayrılırken, dizileme olarak pirodizileme (sentez yoluyla dizileme), ligasyon yoluyla dizileme, ion semikonduktör (proton deteksiyon) dizileme olarak üç sınıfa ayrılır [8]. PacBio dizileme ve Nanopore dizileme gibi yöntemler de üçüncü nesil dizileme olarak sınıflandırılır.

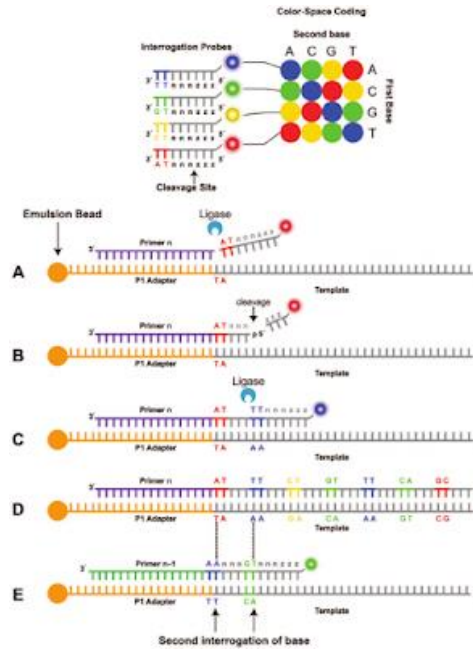
2.2.1 YND Çalışma Prensipleri

YND teknolojisi dizileme teknolojisi olarak cihazlara göre farklılık gösterse de temel olarak 3 aşamada gerçekleştirilir; örneklerin hazırlanması, nükleik asit dizilenmesi ve data analizi. Örneklerin hazırlanması aşamasında; platformun farklılığına göre dizilemek istenen başlangıç materyali DNA veya RNA örnekleri izole edilir. Ardından elde edilen örnekler fragmente edilerek platformun uygun olduğu şablon boyutuna uygun olarak ayarlanır. Adaptör ligasyonu gerçekleştirilerek mikroçip, mikrobead ya da nanowell gibi katı yüzeylere tutunması kolaylaştırılır ya da sirküler yapı haline getirilir. Ardından dizilemek için yeterli kopyaya sahip olmak adına her şablonun amplifikasyonu gerçekleştirilir. Kütüphaneler yalnızca bir uçtan (tek uçlu okuma) ya da iki uçtan okuma (eşli uçlu okuma) ile dizilebilir.

Lusiferin oksitlendiğinde ışığa meydana gelir ve bu ışığa kaydedildiğinde, sekanslama işlemi sırasında eklenen bazın diziyeye eklendiği dolayısıyla fosfat grubunun ayrıldığı ve sekans sırasının o bazda olduğu tespit edilir [11]. Tek seferde tek baz eklendiğinden bazın tespiti de ışığa ile yapıldığından karışması olası değil, güvenilirdir. Sanger sekanslamadan sonra tek bir döngüde uzun okuma verdiğinden tercih edilir olsa da bu yöntemde hata okuma ve bulma sistemleri olmadığından güvenilir data elde edilememektedir. Sentez yoluyla dizileme metodunu en çok ILLUMINA cihazlarında kullanır. Thermo Fisher ve Pacific Biosciences gibi firmalarının da platformları mevcuttur.

2.2.4 Ligasyon ile Dizileme

Bu yöntemde floresan ile işaretlenmiş oligonükleotid dizilerinin prob olarak kullanımı gerçekleştirilir. Bu problemler farklı farklı uzunluktadır. DNA ligaz enzimi kullanılır ve problemler primerle ligasyona uğrar. Floresan ışığa ligasyon meydana geldiğinde olur ve ışığa alındığında hangi probun ışığa yaptığı kaydedilerek sekans belirlenir. Bu ışığı kayıt edebilmek için farklı problemler belirli sırayla eklenir. Ligasyondan sonra probtaki 3 baz ve floresan ışığa yapan boya içeren kısım atılır ve ışığa yapmaz [11]. Dolayısıyla bir sonraki ışığa engellenmemiş olur. Oxford Nanopore Technologies ligasyon ile dizileme prensibini kullanır ve GridION, MinION gibi platformları mevcuttur. Şekil 2.2’de ligasyon yoluyla dizileme yöntemi resmedilmiştir.



Şekil 2.2 Oligo ligasyonu saptama yoluyla sekanslama (SOLID) [12].

2.2.5 İyon semikondüktör ile Dizileme (Ion Torrent)

DNA sekansı sırasında, nükleotid eklendiğinde açığa çıkan H^+ iyonlarının serbest kalmasıyla pH değişir. Bu pH değişimiyle H^+ iyonunun açığa çıktığı dolayısıyla sekansa yeni nükleotid eklendiği saptanmış olur. Nükleotidler sırayla eklenir ve pH değişimi takip edilir [13]. Bu prensibi Thermo Fisher Scientific kullanır ve Ion Proton, Ion PGM gibi platformları vardır.

2.2.6 Üçüncü Nesil Dizileme (Tek Molekül Dizileme)

Yeni nesil dizileme teknolojisi, Sanger sekanslama, Shot-gun dizileme ve Maxam-Gilbert sekanslama gibi birinci nesil dizileme yöntemlerinin zahmetli hazırlıkları, kısa dizileme alanı, uzun süreli sekanslama gibi problemlerine çözüm getirmiş olsa da kısa okuma sekanslaması yaptığı için hala birtakım dezavantajlara sahiptir. Kısa okumalar meydana geldiği için, tekrar bölgeleri ve yapısal değişiklikler gibi sekanslanan alanların analizleri için birçok biyoinformatik araçlar ve karmaşık işlemlerin ardından bir sekans çıkarılır. Ve bu da cihazlardan çıkan datanın analizini zorlaştırır ve süresini uzatır. Büyük genomik bölgeler sekanslanırken parça parça sekanslanıp sıralama için doğruluğu da sorgulanabilir. İkinci nesil dizilemenin birtakım dezavantajlarına üçüncü nesil dizileme ile çözüm bulunmuştur. Bunlardan bazıları; tekrar bölgeleri ve yapısal değişiklikler gibi dizilenmesi ve align edilmesi zor bölgelerin parça parça dizilenip sıralama aşaması için çeşitli biyoinformatik yöntemlerle çalışılmasından tek parça halinde okunarak sıralama yapılması adımının atlanması üçüncü nesil dizilemenin bir avantajıdır. Ve bu avantaj aynı zamanda doğruluğu artıran da bir avantajdır. Üçüncü nesil dizilemelerin hata oranları daha azdır ve yüksek doğruluk seviyelerine sahiptir, genetik varyasyon ve hastalıkların daha hassas tespitine olanak sağlar. Yine aynı şekilde, ikinci nesil dizilemelerde sekanslama bittikten sonra analiz edilebilip datalara erişilebilirken, üçüncü nesil dizilemelerde gerçek zamanlı dizileme yapılır. Dolayısıyla datalara erişim için runun bitmesinin beklenmesine gerek yoktur ve bu da analiz için zaman kaybını azaltmış olur. Üçüncü nesil dizileme yöntemleri, metilasyon gibi epigenetik değişikliklerin belirlenmesinde de daha iyidir. Bu da epigenetiğin rol oynadığı hastalıkların teşhisi için daha güvenilir bir olanak sağlar. Tüm bunlarla birlikte, üçüncü nesil dizileme daha yüksek doğruluk, daha güvenilir data, daha yüksek veri eldesi ve epigenetik gibi alanların dahli ile daha kapsamlı analiz imkanı sunar. Üçüncü nesil dizileme metodu da sentez ile dizileme, nanopora dayalı ve tek moleküllü gerçek zamanlı dizileme ve ligasyon bazlı dizileme olmak üzere çeşitli teknolojileri kullanır. İkinci nesil dizileme teknolojilerinden

farklı olan nanopore dizileme; DNA, RNA ve protein dizilemede kullanılır. Nanopore bir proteindir ve nano boyutta olmasından dolayı ve içinde por şeklinde bir boşluk olmasından dolayı bu isme sahiptir. DNA molekülündeki dört nükleotid de birbirinden farklı olduğu için pordan geçerken oluşturduğu akım farklı olur ve por içerisindeki geçiş sırasında bu voltaj farkıyla tespit sağlanır [14]. Oxford Nanopore Teknolojileri, Pacific Biosciences ve Illumina TruSeq platformları bu teknolojileri kullanan firmalardır.

2.3 Tüm Ekzom Sekanslama

Tüm ekzom sekanslama; organizmada tüm genlerde ekzon olarak adlandırılan protein kodlayan bölgelerin sekanslandığı bir testtir. Genler ekzon ve intron olmak üzere 2 segmentten oluşur. Ekzonik bölgeler protein kodlanan bölgeler olmakla birlikte, intronik bölgeler düzenleyici rolleri üstlenir. Protein kodlayan bölgelerde meydana gelen mutasyonlar proteinin kodlanmasını etkileyeceğinden ekzonik bölgede meydana gelen mutasyonlar, delesyon ya da insersiyonlar daha hayattır. Tüm ekzom sekanslamanın amacı, proteinlerin oluşmasında rol oynayan bu bölgelerin sekanslanıp herhangi bir mutasyonun meydana gelip gelmediğini dolayısıyla o gen tarafından kodlanan proteinin vücut içinde üretiminde mutasyon nedeniyle herhangi bir sorun olup olmadığını tespit etmektir. Bazı genlerin ekzonik bölgelerinde meydana gelen mutasyonların hangi sendrom veya hastalığa sebebiyet verdiği bilinmektedir. Örneğin; *FBNI* geninde meydana gelen mutasyonların Marfan Sendromuna, *MEFV* geninde meydana gelen mutasyonların ailesel akdeniz ateşine ve *BTD* geninde meydana gelen mutasyonların biotinidaz eksikliğine neden olduğu bilinmektedir. Bazen bireylerin şikayetleri ve fenotiplerine göre bazı tanılardan şüphelenilebilir, ancak bunun için genotiplerine yani genlerinin sekanslanılmasına da ihtiyaç duyulur. Örneğin; kalp, göğüs ve iskelet sorunları olan bir bireyde, fenotipiyle yorumlandığında marfan sendromundan şüphelenilip *FBNI* gen analizi yapılabilir. Ancak birçok kas ve sendromun fenotipik özellikleri de bu özellikleri karşılayabilir. Dolayısıyla tek sendroma, tek hastalığa ilişkin genlerin sekanslanmasındansa tüm genlerin ekzonik bölgelerinin sekanslanması daha avantajlıdır. Bu tanı veya tedavi için gözden bir şey kaçırılmamasına olanak tanır. Ya da hastanın şu an değilse bile ileride olabileceği hastalıklar, ailesinde hastalık öyküsü olduğu için yatkınlık durumu ya da evlendiğinde veya çocuk sahibi olmak istediğinde risk faktörü olabilecek durumların öğrenilebilmesi için tüm ekzom sekanslama testleri önerilmektedir. Tüm ekzom sekanslama testleri, tüm ekzonik bölgeleri sekanslarken,

yapılan birçok çalışma önemli olan alanların sadece ekzonik bölgeler olmadığını göstermiştir. Dolayısıyla bazı genlerde intronik bölgelerde meydana gelen mutasyonların da gerek düzenleyici rollerinden gerek önemli RNA bazlı bileşenlerin alt yapısını oluşturmalarından önemli oldukları saptanmıştır. Bu bulgular intronik bölge değişimlerinin (insersiyon/delesyon veya mutasyonların) tespitini de önemli kılmıştır. Bu bulgular sonucunda da tüm ekzon sekanslamasının yanı sıra yeni bir yaklaşım olan tüm genom sekanslama teknolojisi yaygınlaşmaya başlamıştır.

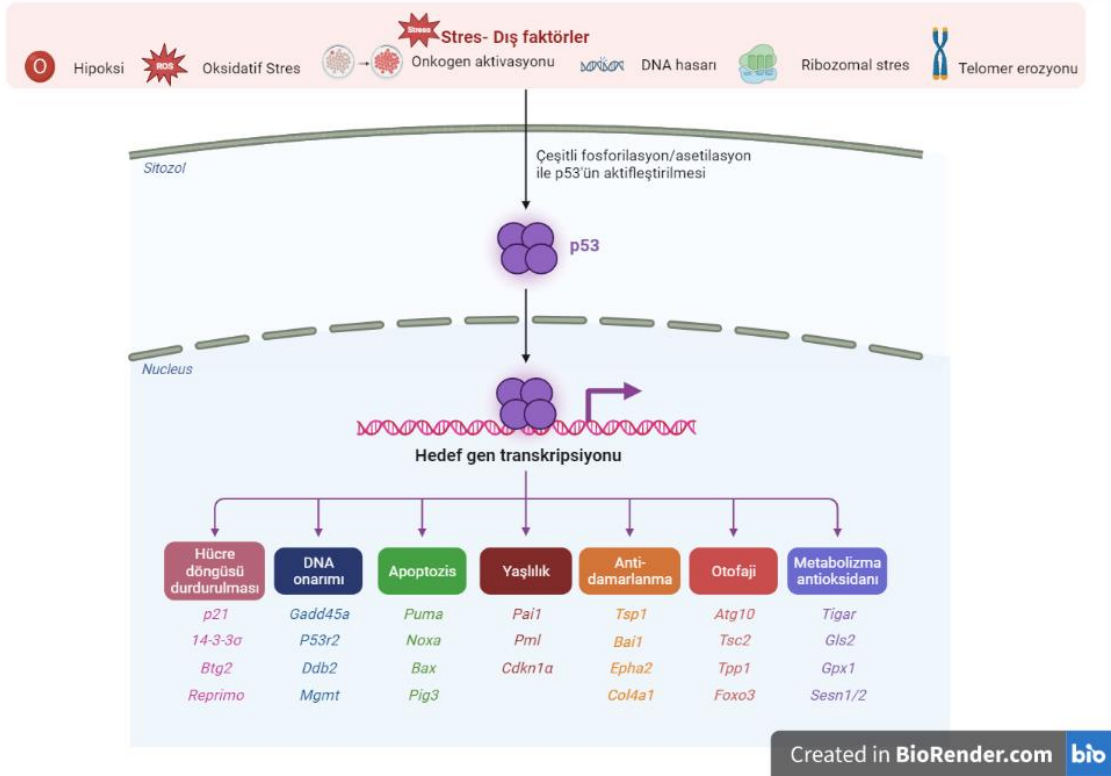
Tüm genom sekanslama teknolojisinde, tüm ekzom sekanslama teknolojisinin yanı sıra intronik bölgeler de amplifiye edilmektedir.

2.3.1 TP53 Geni.

TP53 (Tumor Protein P53) geni, tümör baskılayıcı olarak rol oynayan p53 proteinini kodlayan gendir. *TP53* insan tümörlerinde en sık değişen ve genomun bekçisi olarak isimlendirilen gendir. İnsan genomunda 17.kromozomda 17p13.1 bölgesinde bulunur. 15 kadar p53 protein izoformu kodlar. P53 proteini bir transkripsiyon faktörü ve tümör baskılayıcıdır. Rol aldığı birçok hücresel süreç vardır. Bunlardan bazıları; DNA tamiri, apoptozis, hücre döngüsünün düzenlenmesi, senesans düzenlenmesi ve angiogenez düzenlenmesidir. Kansere oluşumunun başlangıcından sonuna kadar rollere sahiptir. Tümör baskılayıcı olduğundan, iki kopyasının da mutasyona uğraması sonucunda işlev göremez hale gelir ve birçok hücresel işlev duraksama ve aksama noktasına gelerek kanser oluşumu gibi ciddi sonuçlar doğurabilir.

P53 proteini, hücre döngüsünün G1 evresinde rol oynar. G1 fazından S (sentez) aşamasına geçişi düzenleyerek hazır olmayan ya da bozuk olan DNA'nın sentezlenmemesi, metabolik hazırlığın tam olduğunda geçişin sağlanması gibi rolleri vardır. Eğer DNA'da bir hasar mevcutsa, mutasyona uğramamış TP53 geninin proteini p53, G1 fazında hücre döngüsünün ilerlemesini durdurarak, S fazına geçişi duraklatır. DNA hasarı olması durumunda p53 DNA tamir mekanizmalarını etkinleştirerek bu hasarın onarılmasını ve bozuk olmayan DNA'nın sentezlenmesine olanak verir. Tüm bu süreçler başarıyla sonlanmadıysa ve DNA tamir edilemiyorsa, yine mutasyona uğramamış bir p53 proteini bu sefer hücreyi apoptoza yönlendirir. Ciddi stres durumlarıyla ve DNA hasarlarında hücreyi planlı hücre ölümüne girmesi için uyararak kontrollü ölümü sağlar. Bu da bozuk DNA'lı yoğun stresli bir hücrenin bölünmesinin durdurulmasına ve potansiyel kanserleşmenin önlenmesine olanak tanımış olur [15].

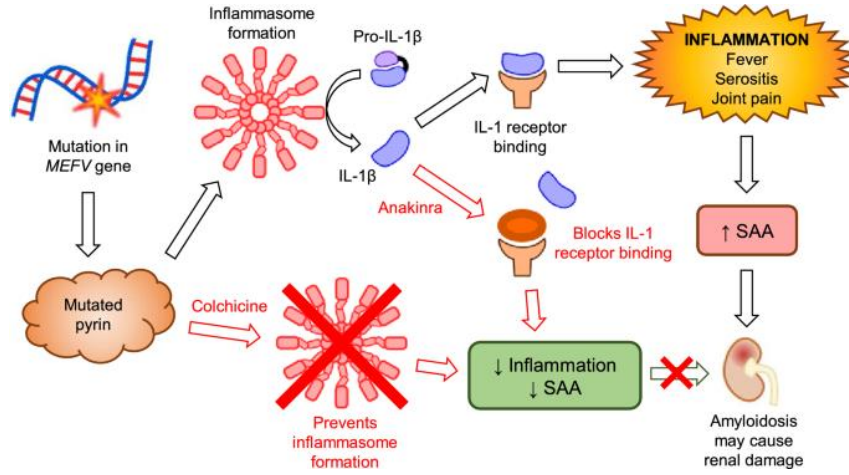
Şekil 2.3'te p53 proteininin rol aldığı yollar ve etkileşime girdiği diğer genler/proteinler verilmiştir.



Şekil 2.3 TP53 geninin kodladığı p53 proteininin rol aldığı mekanizmalar.

2.3.2 MEFV Geni.

MEFV (Mediterranean Fever) geni, pyrin veya geni ile aynı isimde de anılan *MEFV* proteinini kodlar. *MEFV* geni 16.kromozomda yer alır ve bizim ülkemizde de yaygın görülen Akdeniz ateşi olarak bilinen ve FMF (Familial Mediterranean Fever) olarak anılan bir hastalık için resesif kalıtım gösteren bir genidir. İsminden de anlaşılacağı gibi ateş, karın, eklem ve göğüs ağrısı gibi semptomlara neden olur. *MEFV* geni bilinen bir fenotiple ilişkili ve protein kodlayan bir gen olduğundan, *MEFV* geninde meydana gelen mutasyonlar FMF hastalığının temelinde yatan mekanizmayla doğrudan ilişkilidir. Başta enfeksiyonla savaşan nötrofiller olmak üzere bazı immün sistem hücrelerinde transkribe edilir. Doku hasarı ve bu hasara tepki inflamasyonun düzenlenmesi, nötrofillerin inflamatuvar yanıt için doğru hareket edip yanıtı düzenlemesi gibi rolleri mevcuttur. Ancak *MEFV* geninde bir mutasyon meydana gelirse, *MEFV* proteini de işlevini yerine getiremez ve immün hastalıklar, akut karın ağrısı, ateş gibi semptomlarla mutasyonlu bireyler kendini gösterebilir [16].



Şekil 2.4 MEFV geni hücresel işlevleri. (BAĞIŞ, 2020)

2.4 Data Kalite Parametreleri

DNA sekans datalarının kalite kontrolü laboratuvarında kütüphane hazırlama aşaması, sekanslama aşaması ve data analizi aşaması olmak üzere 3 basamakta gerçekleşmektedir. Data kalitesini ilk olarak etkileyen başlangıç materyali olan DNA/RNA'nın kalitesidir. DNA'nın kalitesini etkileyen birçok faktör bulunmaktadır. Hastadan alınan kanın konulduğu EDTA'lı tüpün pH'ı, inkübasyon sıcaklığı ve zamanı, EDTA'nın miktarı, hastanın kanının alınma şekli ve laboratuvara ulaşırken ki taşıma koşulları dahi giriş materyali olan DNA'nın kalitesini etkileyebilmektedir. Bazı teknolojiler giriş materyali olan DNA'yı degrade olarak çalışsa da Illumina NovaSeq 6000 platformu 350-600 bç aralığında DNA örneklerini tercih ederken, MGI DNBSeg 6400 platformu ise 300 bç ile 1 kb aralığında DNA örneklerini tercih eder. Ancak kütüphane hazırlığı sırasında tercih edilen protokol nedeniyle bu boyut farklı optimize edilebilmektedir.

Kütüphane hazırlığı yapılırken; giriş materyali kontrolü, fragmentasyon ve uç-yenileme işleminden sonraki boyut kontrolü, adaptör ligasyonu yapıldıktan sonraki boyut kontrolü ve boyut seçilimi, kütüphane normalizasyonu ve kuantifikasyonu aşamalarında yapılan kontrol, YND datasının kalite kontrolü için kritik adımlardandır.

Kütüphane hazırlığı sırasında YND için gerekli olan minimum miktardaki DNA değerinin mevcut olup olmadığı DNA izolasyonu aşamasından sonra florometrik (Quibit) ya da spektrofotometrik (Nanodrop) kullanılarak ya da başlangıç materyali için beklenen boyutta olup olmadığı ya da DNA'nın degrade halde olup olmadığı kapiller jel elektroforezi yardımıyla kontrol edilmelidir. Fragmentasyon ve adaptör ligasyonu aşamalarından sonra elde edilen fragmanların istenen boyutta olup olmadığı, adaptörlerin

bağlanıp bağlanmadığı gibi kritik aşamalar kapiller jel elektroforezi ile teyit edilmelidir. Sekanslamadan önce Bioanalyzer ile YND kütüphaneleri, giriş boyutunun beklenen gibi olup olmadığı ve adaptör dimerleri ile kontaminasyonun olup olmadığı kontrol edilmelidir.

Sekansa başlamadan önceki kütüphane hazırlığı kontrollü ilerledikten ve uygun sekans programı seçilip cihaz ayarlandıktan sonra elde edilen datanın iyi olup olmadığı ise çeşitli analiz programları kullanılarak teyit edilebilmektedir. Bu programlara; FASTX-Toolkit, NGS QC Toolkit, TagDust, SIMPLEX, PRINSEQ örnek verilebilir⁸. Datalarda kontrol edilen baz başı sekans kalitesi, kalite skorları, GC kontenti, bilinmeyen (N) baz okunma oranı, sekans uzunluğu dağılımı, duplikasyon seviyesi, overekspres edilmiş sekanslar, adaptör kontenti ve kmer kontenti gibi parametrelerin datanın kalitesinin teyidi için kontrol edilmesi gerekmektedir.

2.4.1 GC İçeriği

GC içeriği DNA'yı daha kararlı yapan, GC oranı arttıkça DNA'nın erime noktasını artıran bir parametre olmakla birlikte, GC içeriği yüksek dizilerin dizilenmesi de bir o kadar güçlük çıkarmaktadır. GC zengin bölgelerin stabilitesini artıran şey aradaki hidrojen bağından çok, baz istiflemesi ile olan etkileşimlerdir. PZR işlemi için tasarlanan primerlerde de GC içeriğine dikkat edilir ve GC içeriği yüksek olan primerler tasarlanmamaya çalışılır. Çünkü ne kadar GC içeriği yüksek olursa, DNA'yı denatüre etmek için gereken ısı o kadar yüksek olmalıdır. GC stabiliteyi artırdığından, ekstrem sıcaklıklarda yaşayan canlıların genomunda GC kontentinin daha yüksek olduğu görülmüştür. GC içeriği zengin bölgeler, ikincil yapılar, self-dimer, cross-dimer ve hairpin loops yapılarını oluşturma eğilimi yüksek olan bölgeler olduğundan bu şekilde denatürasyonları zor olup uzun süre birikebilirler. GC içeriği bu gibi özelliklerinden dolayı, sekanslama datasının kapsamını etkileyen bir faktördür. Çok yüksek ya da çok düşük GC içeriğine sahip olan bölgelerde kapsama oranı önemli derece etkilenip azalırken, ortalama bir GC içeriğine sahip bölgelerin ise kapsamı iyidir. GC içeriğinin ortalama %50-60 olması beklenir. %30-40 GC içeriği az iken, %70-80 ise yüksek olarak değerlendirilir. Dolayısıyla bu oran AT zengin bölgelerde GC içeriği arttığında kapsamın artmasıyla sonuçlanırken, GC zengin bölgelerde GC içeriği arttığında kapsamın azalmasıyla sonuçlanır. Bu kapsamın azalması engelini aşmak için PZR adımlarının atlanması gibi bir seçenek sunulabilir, bunun için de daha büyük DNA girişi gerekmektedir [18]. Sekanslanan dataların GC içeriğini incelemek için, fastQC (fast

quality control) kullanıldı. FastQC; fastq datasının analizini yapan bir yazılımdır. FastQC raporunda datanın kalitesi hakkında baz kalitesi, nükleotid dağılımı, adaptör kirliliği, kopya sayısı, GC içeriği gibi birçok bilgiye sahip olabiliriz.

2.4.2 Ortalama Derinlik

Ortalama derinlik, dataların güvenilirliğini ve doğruluğunu artıran bir kalite parametresidir. Okunan her bir bazın okunma sayısını yani o nükleotidin dizilenme sayısını verir [19]. Dolayısıyla daha yüksek bir ortalama derinlik, o bazın daha çok okunduğu anlamına gelir. Referans genom ile kıyaslayarak data analizi yapılırken tespit edilebilecek olan tek nükleotid varyantları (SNV), insersiyon ve delesyonlar gibi değişimlerin doğruluğu o değişim bölgesinin fazla okunmuş olmasıyla daha yüksek bir doğruluk kazanır. Sekanslanan alanların tamamındaki bu okuma sayısının ortalaması ortalama derinliği verir. Tüm ekzom sekanslama datalarında elde edilmek istenen optimum ortalama derinlik araştırma amacına göre değişiklik göstermektedir. American College of Medical Genetics and Genomics (ACMG) gibi kuruluşların kılavuzlarına göre en az 30x kapsama önerilir [20]. Bu da her bir bazın ortalama olarak en az 30 kez okunması anlamına gelmekle birlikte, varyantlar okunurken okuma kalitesi kötülüğünden varyant kaybını önleyerek daha güvenilir bir data sağlar. Bununla birlikte, somatik mutasyonlar gibi okuması germline mutasyonlardan daha az olan mutasyonlar da saptanmak istenebilir. Somatik mutasyonlar, germline mutasyonlar gibi her hücrede olmadığı sonradan çevresel faktörlerle meydana geldiği için hücrelerdeki miktarı germline kadar olmadığı için, okuma sayısı da ona oranla az olur. Bunun için normal bir datada istenen kapsam 30x iken, somatik mutasyon isteniyorsa 50x veya daha yukarı çıkarılabilir. Çalışmanın amacına, ulaşılmak istenen dataya göre runda ayrılan alanın artırılması ile okuma sayısı artırılabilir ve ortalama derinlik artırılarak daha güvenli bir data elde edilebilir.

Bununla birlikte dataların ortalama derinliği, kapsam ile karıştırılmaktadır. Bir datanın ortalama derinliği, sekanslanan bölgedeki her bazın okunma sayısının ortalamasıdır. Bir datanın kapsamı ise, o bölgenin yüzdesini temsil eder. Yani; sekanslanan bölgenin ne kadar derinliğe kadar sekanslandığı anlamına gelir. Eğer bir datanın kapsamı %80 ise, sekanslanmak istenen bölgenin %80'inin sekanslandığı anlamına gelir. İki kalite parametresi de birbiriyle ilişkili olmakla birlikte farklı kavramlardır ve karıştırılmamalıdır.

2.4.3 Q20 Değeri

Q değeri hatalı baz okuma olasılığıyla hesaplanan, “Phred Quality Score” olarak da adlandırılan bir data kalite parametresidir. Q değeri $-10\log_{10}(e)$ formülüyle hesaplanırken, ‘e’ baz okunurken tahmini yanlış okunma olasılığıdır. Genel olarak data kalite parametresi olarak kullanılan Q20 ile Q30 değeridir. Q20 değeri; her 100 bazda 1 hata oranını gösterirken; Q30 değeri her 1000 bazda 1 hata oranını gösterir. Bu yüzden de Q30 değeri YND için data kalite parametreleri arasında kabul edilen bir ölçüttür. Biz bu çalışmada Q20 değerini hesaplayarak 100 bazda 1 hata oranını hesapladık. Daha küçük alanda 1 baz yanlışlık oranı hesaplandığından hassasiyeti Q30’a göre daha az olmakla ve daha az kaliteyi belirtmekle beraber; Q20 değeri de birçok çalışmada kullanılan bir kalite parametresidir. Birçok araştırma ve uygulama için Q20 değeri de geçerli olup, düşük derinlikli data setlerinde kullanılabilir.

Q20 değeri logaritmik bir hesaplama ile bulunduğundan; hata oranının düşüklüğü Q20 değerinin yüksekliği anlamına gelmektedir. Yani daha yüksek bir Q20 değeri daha düşük bir hata oranına dolayısıyla daha güvenilir bir data setine işaret eder. Düşük bir Q20 değeri ise datanın güvenilirliğini düşüren bir etmendir. Düşük Q20 değeri, yanlış pozitif sonuçlara da yol açabilen sıralama kalitesinin düşük olduğunu gösteren bir belirteçtir. Yüksek bir Q20 değeri; 100 bazda 1 okunan hata oranının düşük olduğunu gösterdiğinden, tercih sebebidir [21].

2.4.4 IGV (Integrative Genomics Viewer)

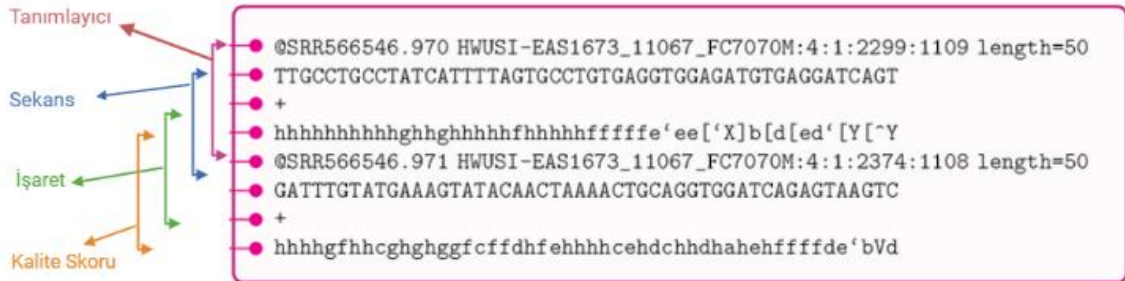
Data kalite parametrelerinin yanı sıra; sekanslanan datanın gözle görülür şekilde kaç kere okunduğu; o lokasyonda hangi bazın okunduğu, varyant var ise heterozigot/homozigot olma durumu gibi birçok konu IGV adı verilen bir program yardımıyla görselleştirilebilmektedir.

MGI ya da Illumina gibi YND cihazlarından çıkan dataların ilk formatı Fastq formatıdır. Şekil 2.5’te verildiği gibi fastq formatı sırasıyla BAM ve VCF dosyalarına dönüştürülür.



Şekil 2.5 Ham data fastq formatının dönüştürülmesi

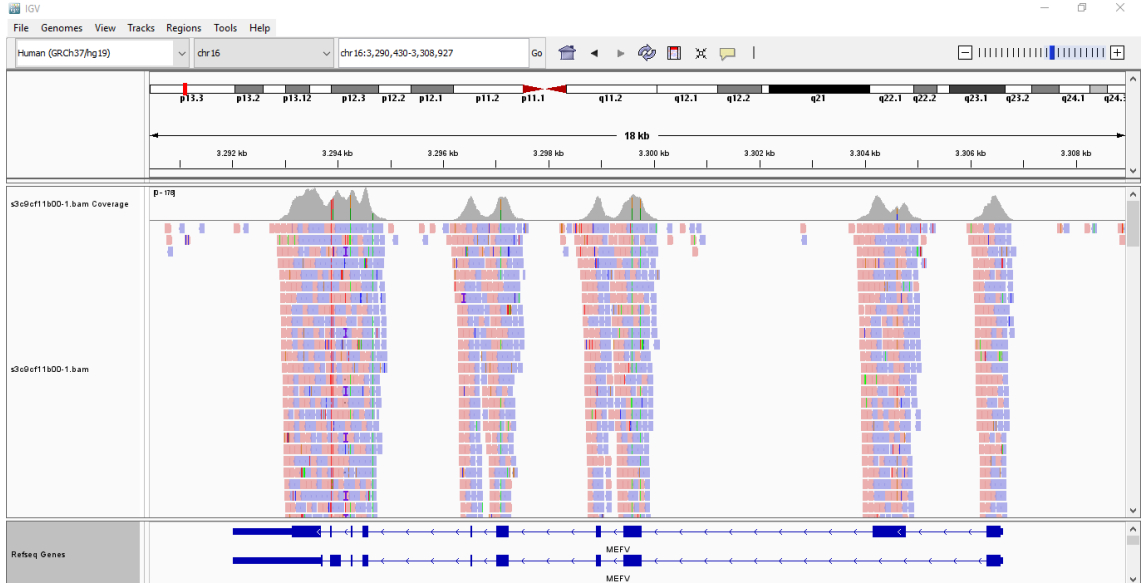
Fastq dosyası her biri dört satırdan oluşan bir formattır. Bu dört satır, tanımlayıcı, DNA sekansı, işaret ve kalite skorundan oluşur. İlk satırda yer alan tanımlayıcı; her kayıtn başında yer alır ve benzersizdir. Her sekansın hangi örneğe ve hangi dizileme aleti ile yapıldığı hakkında bilgi içerir. Genellikle @ ile başlar, okumanın hangi örneğe ait olduğunu belirlemek için bir kod veya dize içerebilir. Tanımlayıcılar enstrüman adı, runın adı, flow cell ID'si, flowcelldeki lane numarası, x ve y koordinatları, indeks sekansları gibi spesifik bilgileri içerir. Sekans, okunan DNA dizisini yani nükleotid dizilişini ifade eder. YND dataları genellikle birden fazla örneği içerir. Üçüncü satırda yer alan işaret; sekansın bitip kalite skorlarının başladığını belirtmek için ikinci satırdaki sekans dizisi ile dördüncü satırdaki kalite skorları arasında yer alır.



Şekil 2.6: Örnek fastq dizisi

Fastq dosyalarının dördüncü satırında yer alan kalite skorları her bir nükleotidin doğruluğunu belirtmek için ASCII (American Standard Code for Information Interchange) karakterleri ile yazılır. ASCII karakterlerinde daha yüksek kalite skorları, daha güvenilir nükleotidleri temsil eder [22]. Kalite skorları Phred skorları hesaplanarak yazılır. Phred skoru $-10\log_{10}(P)$ formülü ile hesaplanır. Yanlış eşleşme olasılığı 0-60 arası hesaplanır, P yanlış eşleşme olasılığını verir ve $P = 10^{(-Q/10)}$ ile hesaplanır. Q ise kalite skoru veya hata olasılığıdır. Hata olasılığının düşük olması beklenir çünkü kalite skoru artar ve Phred skoru yüksek çıkar [23].

Ham data olan fastq verileri daha sonra BAM dosyasına çevrilir. BAM dosyaları Fastq dosyalarına oranla daha az yer kaplayan sıkıştırılmış bir depolama dosya formatıdır. Fastq datalarının referans genomuyla hizalanması ve sıralanmasıyla oluşturulan biçimdir. DNA verilerinin sıralanması ve hizalanması işlemi (Fastq dosyasından BAM dosyasına çevrilme işlemi), varyant analizi gibi ileri analizler için olanak sağlayan bir işlemdir. BAI (Binary Alignment Index) dosyası BAM dosyası oluştururken oluşturulması gereken, BAM dosyasının hızlı erişimini sağlamak için indekslenmesini yapan dosya formatıdır. BAI dosyası BAM dosyasının kromozom bölümlerini indeksler. Genetik konumlarının yerini, BAM dosyasının yerine göre indeksleyerek o konuma hızlı şekilde erişimi sağlar. Özetle, BAI dosyası BAM dosyasını belirli kromozomal lokasyonlara göre indeksleyerek, incelenmek istenen lokasyona erişimi ve taramayı hızlandıran sürümdür. Dolayısıyla BAM dosyası incelenirken BAI dosyasının da oluşturulması gerekir. BAM dosyasında datalar IGV adı verilen bir yazılımla görselleştirilerek analiz edilebilir. “Integrative Genomics Viewer”ın kısaltması olan IGV data analizi yaparken kullanılan bir araçtır. IGV programında insan genomunun referansının olmasıyla birlikte, C.elegans, fare, tavşan, domates gibi sayısız bugüne kadar dizilenmiş olan canlıların da referans genomları bulunmaktadır. Referans genomları indirme imkanı da sağlayan bu yazılım, dizilenen dataların bam ve bai dosyalarının yüklenmesiyle görselleştirerek çalışır. Referans genoma göre incelenen genin ekzon ve intronları aşağıdaki fotoğraftaki gibi gözükür. Koyu lacivert kutucuklar ekzonik bölgeleri, lacivert ince çizgiler de intronik bölgeleri ifade eder. RefSeq Genes yazan kısımda iki satır bulunmasının sebebi o gene ait iki farklı transkript bulunmasından kaynaklanır. Alternatif splicing dolayısıyla genlerin birden fazla transkriptleri mevcut olabilir. YND cihazları kullanılan kitlere göre, belirli transkriptleri okuyup, onlara göre dizileme yaparlar. Dolayısıyla bir genin bir transkriptindeki ekzon sayısı, diğer transkriptine göre farklı olabilir.



Şekil 2.7 Örnek bir IGV görüntüsü

IGV’de analiz yapmak istediğimiz de ise, ilgilenilen bölge, ekzon ve nükleotid yaklaştırılarak incelenebilir. RefSeq Genes yazan kısımda, referans genomu göre o lokasyonda olması gereken protein kodları ve nükleotidler yazmaktadır. Mavi ve pembe okumalar iki yönden okumayı temsil eder, biri forward okuma iken diğeri reverse okumadır. Eğer dizilenen data referans genomu ile aynı olup, herhangi bir mutasyonu yoksa aşağıdaki gibi pembe ve mavi okumalar sade şekilde gözükür. Ama eğer dizilenen datada referans genomdan farklı olarak mutasyon mevcutsa, referans genomuyla aynı olan okumalar gözükmezken, farklı okunan baz o baza özgü renk ile okumaların üzerinde gösterilir. Aşağıdaki fotoğrafta belirtilen lokasyonda gösterilen mutasyonda, o nükleotid bölümü toplamda 120 kere okunmuş iken, 58 kere Adenin, 62 kere referans genomuyla aynı olan Guanin bazı okunmuştur. Hasta o nükleotid bölgesinde Guanin’den Adenin değişimi açısından heterozigot olduğundan okumaların %48’i A olarak gösterilmektedir. IGV yazılımı, hastanın sekans okumasının bu şekilde görselleştirmesini sağlarken, analizi için de bir olanak sağlamış olur.



Şekil 2.8 Saptanan bir varyantın IGV görüntüsü

Bam dosyasının ardından datalar, VCF (Variant Call Format) adı verilen varyant analizine olanak sağlayan bir dosya formatına dönüştürülür. VCF formatında, referans genomdan farklı olarak tespit edilen nükleotid değişimleri, insersiyon ve delesyonlar görülür. Varyant calling adı verilen süreçle BAM dosyası referans genomuyla hizalanarak değişimler tespit edilir ve VCF dosyası oluşturulur. VCF dosyası, genetik varyasyonların yanında polimorfizmleri de verebilir. Bu değişimlerin sıklığı, kalitesi, kalıtım modları, fenotipik özellikleri gibi bilgileri kapsar. Genomik varyantların araştırılması, tespiti ve karşılaştırılması için kullanılan dosya formatı olan VCF, belirli yazılımsal adımlardan sonra R programı gibi başka farklı programlar yardımıyla da excele çevirebilir.

3.1 Tüm Ekzom Sekanslama (WES)

3.1.1 Örnek toplama ve DNA izolasyon

Bu çalışma Next Genetic Center (NGC) Genetik Hastalıkları Değerlendirme Merkezi'nde tüm etik kurallar göz önünde bulundurularak yapıldı. Kan örnekleri farklı yaş gruplarına sahip, farklı şehirlerden gelen hastalardan alındı. 24 adet tüm ekzom sekanslama işlemi için hasta seçildi. 12'si Illumina protokolü ile çalışılırken, 12'si MGI protokolüne göre çalışıldı.

Hastaların kanı mor kapaklı EDTA'lı tüplere alındıktan sonra, Promega cihazında Maxwell® RSC Blood DNA Kit kullanılarak izole edildi. İzole edilen DNA, Qubit® dsDNA HS Test Kiti ile başlangıç konsantrasyonu belirlendi (50-400 ng arası olmalıdır). ve %1 agaroz jel elektroforezi ile boyut kontrolleri yapıldı.

3.1.2 MGI DNB Seq-G400 Protokolü

İzole edilen DNA'ların her biri 50 ng/mL olacak şekilde 0.2'lik tüplere alınarak toplam hacim 10 µL'e distile su ile tamamlanır.

3.1.3 Fragmantasyon

Tablo 3.1'de belirtilen hacimlerde buz üzerinde 1.5 mL mikrosentrifüj tüpüne enzimatik fragmantasyon karışımı hazırlandı.

Tablo 3.1 Enzimatik Fragmantasyon Karışımı

Bileşenler	1 Reaksiyon için Miktar
Su	25 µL
10x Fragmantasyon Tamponu	5 µL
5x Fragmantasyon Enzimi	10 µL
Toplam	40 µL

40 µL enzimatik fragmantasyon karışımı her 10 µL genomik DNA tüpünün üzerine eklendi ve kibarca pipetaj yapıldı. Tablo 3.2'deki termal döngü koşulları kuruldu.

Tablo 3.2 Enzimatik Fragmentasyon Reaksiyon Koşulları

Adımlar	Sıcaklık	Zaman
Soğutma	4 °C	
Fragmentasyon	32 °C	22 dk
İnaktivasyon	65 °C	30 dk
Tutma	4 °C	Sonsuz

Tüp alındıktan sonra buz üzerine konuldu.

3.1.4 Adaptör Ligasyonu

5.5 µL MGI Adaptörü tüm örnek tüplerine eklenip, buz içerisinde pipetaj yapıldı. Tablo 3.3'deki gibi buz üzerinde 1.5 ml mikrosentrifüj tüpüne ligasyon karışımı hazırlandı ve kibarca pipetaj yapıldı. Ligasyon karışımı çok yoğun viskoz kıvamında olduğu için birkaç saniye vortekslendi.

Tablo 3.3 Ligasyon Karışımı

Bileşenler	Miktar
H ₂ O	14.5 µL
DNA Ligasyon Tamponu	20 µL
DNA Ligasyon Karışımı	10 µL
MGI Adaptör	5.5 µL
Toplam	50 µL

44.5 µL ligasyon karışımı tüpe eklendi ve kibarca pipetaj yapıldı ve spin atılarak sıvının altta toplanmasına dikkat edildi. Termal döngü 20°C'de 15 dk ayarlandı ve kapak ısıtmasının kapalı olmasına ya da minimum sıcaklıkta olmasına dikkat edildi.

3.1.5 Adaptör Ligasyonu Ürünlerinin Purifikasyonu

DNA purifikasyon boncukları kullanmadan 30 dk kadar önce oda sıcaklığına çıkarıp kullanmadan önce vortekslendi. 80 µL DNA temiz boncukları adaptör ligasyonu yapılmış PZR tüpüne eklendi, 10 kere pipetaj yapıldı ve oda sıcaklığında 5 dk inkübe edildi. Kısaca sentrifüjlendikten sonra manyetik standı konulup sıvı berraklaşana kadar 2-5 dk beklenildi. Süpernatant dikkatli şekilde boncuklara değmeden atıldı. 200 µL yeni hazırlanmış 80% etanol yıkama için eklendi. 30 sn kadar beklendikten sonra dikkatlice süpernatant atıldı. Bu adım bir kere daha tekrarlandıktan sonra, tüm sıvının atıldığına emin olunup oda sıcaklığında tüpün kapağı kalan sıvının buharlaşması için açık bekletildi. Pelletin aşırı kurumamasına dikkat ettikten sonra manyetik standı tüpler alınarak 17 µL su pipetajla eklendi ve 5 dk bekletildi. Tüpler tekrardan manyetik standı alınarak sıvı

berraklaşınca kadar 2-5 dk bekletildi. 15 µL berrak süpernatant yeni 0.2 mL tüplere pelletin dağılmadığına emin olunarak alındı.

3.1.6 Yakalama öncesi PZR Amplifikasyonu

Termal döngü programı tablo 3.4'teki gibi ayarlandı.

Tablo 3.4 Yakalama öncesi PZR Programı

Adımlar	Sıcaklık	Zaman	Döngü Sayısı
Başlatma	98 °C	45 sn	1
Denatürasyon	98 °C	15 sn	10
Bağlanma	60 °C	30 sn	
Uzama	72 °C	30 sn	
Son Uzama	72 °C	1 dk	1
Tutma	4 °C	Sonsuz	-

PZR karışımı tablo 3.5'teki gibi 1.5 mL mikrosentrifüje buz üzerinde hazırlandı.

Tablo 3.5 Yakalama öncesi PZR Karışımı

Bileşenler	Miktar (µL) /Numune Başına
KAPA HiFi HotStart ReadyMix	25 µL
MGI Primer Mix	10 µL

35 µL PZR karışımı, pürifikasyonu yapılmış adaptör ligasyonlu 15 µL örneklerin üzerine kibarca pipetajla eklendi. Ardından kısaca spin atılarak hemen kurulmuş termal döngüye konuldu.

3.1.7 Yakalama Öncesi PZR Ürünlerinin Pürifikasyonu

Termal döngü programı bittikten sonra, çalışmadan 30 dk önce çıkarılan DNA pürifikasyon boncukları vortekslenerek homojen hale getirildi. 50 µL (1x) homojenize edilmiş boncuklar, PZR örneklerinin üzerine eklendi ve iyice karıştırıldı. 5 dk oda sıcaklığında inkübe edildikten sonra, manyetik standı yerleştirilerek tamamen berrak olana kadar bekletildi. Süpernatant dikkatli şekilde boncuklara değmeden atıldı. 200 µL yeni hazırlanmış 80% etanol yıkama için eklendi. 1 dk kadar beklendikten sonra dikkatlice süpernatant atıldı. Bu adım bir kere daha tekrarlandıktan sonra, tüm sıvının atıldığına emin olunup oda sıcaklığında tüpün kapağı kalan sıvının buharlaşması için açık bekletildi. Pelletin aşırı kurumamasına dikkat ettikten sonra manyetik standtan tüpler alınarak 22 µL su pipetajla eklendi ve 2 dk oda sıcaklığında bekletildi. Tüpler tekrardan manyetik standı alınarak sıvı berraklaşınca kadar 3 dk bekletildi. 20 µL berrak süpernatant yeni 0.2 mL tüplere pelletin dağılmadığına emin olunarak alındı. Qubit®

dsDNA Broad Range Quantitation Assay (ThermoFisher) kullanılarak ölçüm gerçekleştirildi. Örneklerin konsantrasyonunun capture aşamasından sonra kütüphane çeşitliliğinin az olmaması için ≥ 80 ng/ μL olması gerekmektedir. Tüm örneklerimiz ölçüme uyduğu için devam edildi.

3.1.8 Pool Kütüphaneleri

Her hibridizasyon reaksiyonu toplam 1500 ng indexlenmiş 8 ayrı kütüphaneden eşit olarak pool oluşturulmuş kütüphane gerektirir. 187,5'i ng/ μL cinsinden ölçülen konsantrasyon değerlerine bölerek her amplifiye edilmiş indeksli kütüphanenin 187,5 ng'si için gereken hacmi μL cinsinden hesaplandı. Her pool için indexlenmiş örnek sayısı=8, her pool için indeksli örnek miktarı= 187.5 ng, her pool için toplam ağırlık 1500 ng. Yani 8 örnek bir pool yapıldı, her örneğin konsantrasyonunun 187.5 ng olmasını amaçladık. Toplam 8×187.5 'ten 1500 ng olacak şekilde ayarlandı. Hesaplama her pooldaki indeksli örnek sayısı 1 ise 500 ng, 2 ise 1000 ng, 4 örnek için 375 ng ve 8 örnek için 187,5 ng olarak toplam ağırlık tüm poollarda maksimum 1500 ng olacak şekilde ayarlandı. Hesaplanan miktarlardaki tüm kütüphaneler 0.2 mL PZR tüpüne aktarıldı. Pool hacmi 25 μL 'den küçük olanlara dH₂O eklenirken, 25 μL 'den büyük olanlara direkt olarak kullanıldı. Pool hacminin 1.5 katı kadar DNA pürifikasyon boncukları kullanıldı (25 μL kütüphane hacmi ise, 37.5 μL DNA pürifikasyon beadleri eklenir.) ve oda sıcaklığında 5 dk inkübe edildi. Ardından manyetik stantta 1 dk kadar bekletilerek, sıvı berraklaşana kadar bekletildi ve süpernatant atıldı. Ardından 200 μL yeni hazırlanmış 80% ethanol eklendi. 1 dk kadar beklendikten sonra dikkatlice süpernatant atıldı. Bu adım bir kere daha tekrarlandıktan sonra, tüm sıvının atıldığına emin olunup oda sıcaklığında tüpün kapağı kalan sıvının buharlaşması için açık bekletildi. Pelletin aşırı kurumamasına dikkat ettikten sonra, başlamadan dışarı çıkarılan 65°C'de ısıtılan hibridizasyon karışımı oda sıcaklığına çıkarılıp yavaşça pipetaj yapılarak homojen hale getirildi. Manyetik stanttan alınan tüplerden dikkatlice süpernant çekilerek yeni 0.2 mL Strip tüpe 20 μL Hibridizasyon karışımı eklendi. Hibridizasyon karışımı çok viskoz olduğundan yavaşça pipetleme önerilir.

Tablo 3.6 Hibridizasyon Karışımı

Bileşenler	Miktar
Hibridizasyon Karışımı	20 µL
Twist Spesific ya da Universal Blockers- MGI	2 µL
Comprehensive Exome Probeları v1.3	4 µL
Blocker Solüsyonu	5 µL
Su (40 µL'e tamamlayacak kadar)	10 µL
Toplam	41 µL

Termal döngü programı Tablo 3.7'deki gibi ayarlandı ve çalıştırıldı. Kapak sıcaklığı 85 °C olarak ayarlandı. Minimum 16 saat 70°C'de tutuldu. 16 saatte örnekler uçmasını diye 30 µL hibridizasyon enhancer eklendi.

Tablo 3.7 Hibridizasyon Reaksiyon Koşulları

Adımlar	Sıcaklık	Zaman	Döngü Sayısı
1	95 °C	5 dk	1
2	70 °C	Min 16 saat tutulur	

3.1.9 Boncukların Hazırlanması

Çalışma başlangıcından 30 dk önce oda sıcaklığına çıkarılmış Streptavidin bağlı boncuklar 3 kere vortekslendi. 100 µL Streptavidin bağlı boncuklar yeni 2 mL mikrosentrifüj tüpüne alındı. 200 µL bağlama tamponu tüm tüplere pipetaj yapılarak eklendi. Tüpler manyetik standı alınarak 1 dk bekletildi. Supernatant, pellete zarar vermemeye özen göstererek atıldı. Bu işlem 2 kere daha tekrarlanarak toplamda 3 kez yapıldı. 3. Yıkamanın ardından, son 200 µL bağlama tamponu eklendi ve homojen olana kadar vortekslenerek resuspend edildi. 16 saatlik hibridizasyondan sonra, tüplerdeki örneğin tamamı yıkanan streptavidin bağlı boncukların içine hızlıca eklendi.

3.1.10 Hedefe Bağlanma

Hibridizasyon Reaksiyonu tüpleri ile Streptavidin bağlı boncuklar 30 dk oda sıcaklığında shaker üzerinde sabit hızda homojenize olacak şekilde karıştırıldı. Tüpler 30 dk sonra alınıp spin atılarak, sıvı tüpün altına toplandı. Manyetik standı alınarak 1 dk beklenildi. Supernatant pelleti bozmadan atıldı. Manyetik standtan alınarak 200 µL yıkama tamponu 1 karışımı pipetajla karıştırıldı. Total karışım (~200 µL) yeni 1.5 mL mikrosentrifüj tüpüne alındı ve manyetik standı 1 dk konuldu. Supernatant yine dikkatli bir şekilde

pellete dokunmadan çıkarıldı. Manyetik stanttan alınan tüpe 200 µL 48°C önceden ısıtılmış yıkama tamponu 2 pipetaj yapılarak eklendi. 48°C’de 5 dk inkübe edildi. 1 dk manyetik standda konularak süpernatant atıldı. Yıkama tamponu 2 ile yıkama aşaması 2 kere daha tekrarlanarak, 3 kez yapıldı. Son yıkama da tüm süpernatant atılırken, boncukların kurumamasına dikkat edildi. Tüpleri manyetik stanttan çıkarıldı ve 45 µl su eklendi. Homojenize olana kadar pipetleme yaparak karıştırıldıktan sonra bu solüsyon buz üzerinde inkübe edildi. 22,5 µl’si PZR aşaması için kullanılırken, 22,5 µl’si saklandı.

3.1.11 Yakalama Öncesi PZR Amplifikasyonu ve Pürifikasyonu

Tablo 3.8’deki gibi yakalama sonrası PZR reaksiyonu kuruldu. Kapak sıcaklığı 105°C olarak ayarlandı.

Tablo 3.8 Yakalama sonrası PZR Reaksiyon Koşulları

Adımlar	Sıcaklık	Zaman	Döngü Sayısı
Başlangıç	98 °C	45 sn	1
Denatürasyon	98 °C	15 sn	8
Bağlanma	60 °C	30 sn	
Uzama	72 °C	30 sn	
Son Uzama	72 °C	1 dk	1
Tutma	4°C	Sonsuz	-

22.5 µL Streptavidin bağlı boncukların olduğu solüsyon 0.2 mL yeni tüpe aktarılarak buz üzerinde bekletildi. Kalan 22.5 µL solüsyon -20°C’de sonra kullanmak için saklanabilir. PZR karışımı Tablo 3.9’deki gibi pipetaj yapılarak hazırlandı.

Tablo 3.9 Yakalama sonrası PZR Karışımı

Bileşenler	Miktar
Streptavidin bağlı boncuk solüsyonu	22.5 µL
KAPA HiFi HotStart Ready karışımı	25 µL
MGI Primer karışımı	2.5 µL
Toplam	50 µL

Reaksiyon karışımı hazırlandıktan sonra spin atılarak sıvı tüpün aşağısında toplandı ve PZR programı başlatıldı.

DNA pürifikasyon boncukları çalışma başlangıcından 30 dk önce oda sıcaklığına çıkarılıp homojenize hale getirmek için vortekslendi. PZR programından sonra, 50 µL DNA pürifikasyon boncukları PZR tüplerine eklenerek vortekslendi ve oda sıcaklığında 5 dk inkübe edildi. Ardından manyetik stantta 1 dk kadar bekletilerek, sıvı berraklaşana kadar bekletildi ve süpernatant atıldı. Ardından 200 µL yeni hazırlanmış 80% etanol eklendi.

1 dk kadar beklendikten sonra dikkatlice süpernatant atıldı. Bu adım bir kere daha tekrarlandıktan sonra, tüm sıvının atıldığına emin olunup oda sıcaklığında tüpün kapağı kalan sıvının buharlaşması için açık bekletildi. Pelletin aşırı kurumamasına dikkat ettikten sonra manyetik standtan tüpler alınarak 32 µL su pipetajla eklendi ve 2 dk oda sıcaklığında bekletildi. Tüpler tekrardan manyetik standda alınarak sıvı berraklaşınca kadar 3 dk bekletildi. 30 µL berrak süpernatant yeni 0.2 mL tüplere pelletin dağılmadığına emin olunarak alındı. Qubit® dsDNA Broad Range Quantitation Assay (ThermoFisher) kullanılarak ölçüm gerçekleştirildi. Örneklerin konsantrasyonunun 10-15 ng/ µL olması gerekmektedir. Tüm örneklerimiz ölçüme uyduğu için devam edildi.

3.1.12 Denatürasyon

PZR ürün boyutuna göre, 1 pmol (240,24 ng) PZR ürünü yeni 0.2 mL PZR tüpüne alındı. DNA fragman boyunun 1000'e bölümünün 660 ile çarpılmasıyla 1 pmol PZR ürününe karşılık gelen DNA miktarı hesaplandı. Formülde 364 bç dsDNA fragman boyu yerine konulduğunda, 240,24 ng PZR ürünüyle reaksiyona devam edileceği bulundu. TE Tamponu toplam hacim 48 µL olacak şekilde eklendi. PZR tüpleri Tablo 3.10'daki koşullar ayarlanarak termal döngüye konuldu.

Tablo 3.10 Denatürasyon Reaksiyonu Koşulları

Sıcaklık	Zaman
Kapak ısıtması	Açık
95 °C	3 dk

Reaksiyon tamamlandıktan sonra, tüpler hızlıca 2 dk buzda bekletildi, ardından kısaca sentrifüj edildi.

3.1.13 Tek Zincir Sirkülerizasyonu

Buz üzerinde tek zincir sirkülerizasyon karışımı Tablo 3.11'deki gibi hazırlandı.

Tablo 3.11 Tek Zincir Sirkülerizasyonu Karışımı

Bileşenler	Miktar
Splint Tamponu	11.6 µL
DNA Rapid Ligaz	0.5 µL
Toplam	12.1 µL

12.1 µL karışım, 0.2 mL PZR tüplerinin üzerine ilave edildi. 3 kere 3 sn vortekslelendikten sonra kısaca sentrifüjlenerek solüsyon tüpün altında toplandı. Ardından Tablo 3.12'deki termal döngü koşulları kurularak çalıştırıldı.

Tablo 3.12 Tek Zincir Sirkülerizasyon Reaksiyon Koşulları

Sıcaklık	Zaman
Kapak ısıtması	Açık
37 °C	30 dk
4 °C	Sonsuz

Reaksiyon biter bitmez, PZR tüpleri buzun üzerine alındı.

3.1.14 Enzimatik Parçalama

Tablo 3.13'teki enzimatik parçalama karışımı buz üzerinde hazırlandı.

Tablo 3.13 Enzimatik Parçalama Karışımı

Bileşenler	Miktar
Parçalama Tamponu	1.4 µL
Parçalama Enzimi	2.6 µL
Toplam	4 µL

4 µL enzimatik parçalama karışımı, buz üzerinde bekletilen PZR tüplerine eklendi, 3 kere 3 sn vortekslendi, kısa bir sentrifüjle solüsyon tüpün altında toplandı. Ardından termal döngü tablo 3.14'teki gibi ayarlanarak çalıştırıldı.

Tablo 3.14 Enzimatik Parçalama Reaksiyon Koşulları

Sıcaklık	Zaman
Kapak ısıtması	Açık
37 °C	30 dk
4 °C	Sonsuz

Termal döngüden alınan tüpler kısaca sentrifüjlendi. 7.5 µL parçalama durdurma tamponu tüplere eklendi. 3 kere 3 sn vortekslendi. Ardından tüm solüsyon yeni 1.5 mL sentrifüj tüpüne aktarıldı.

3.1.15 Enzimatik Parçalama Ürünlerinin Temizlenmesi

DNA temizleme boncukları çalışmaya başlamadan 30 dk önce oda sıcaklığına çıkarılıp vortekslendi. 170 µL DNA temizleme boncukları enzimatik parçalama yapılmış ürünlerin üzerine 10 kere pipetaj yapılarak eklendi ve 10 dk inkübe edildi. Kısaca sentrifüj edilerek, 1.5 mL tüp manyetik ayrıştırma standına 2-5 dk sıvı berrak hale gelene kadar bırakıldı. Pipet kullanarak süpernatant çıkarıldı. 500 µL yeni hazırlanmış 80% etanol eklendi. 30 sn kadar beklendikten sonra dikkatlice süpernatant atıldı. Bu adım bir kere daha tekrarlandıktan sonra, tüm sıvının atıldığına emin olunup oda sıcaklığında tüpün kapağı

kalan sıvının buharlaşması için açık bekletildi. Pelletin aşırı kurumamasına dikkat ettikten sonra manyetik standtan tüpler alınarak 32 µL TE tamponu pipetajla eklendi ve 10 dk oda sıcaklığında bekletildi. Tüpler tekrardan manyetik standda alınarak sıvı berraklaşmaya kadar 2-5 dk bekletildi. 30 µL berrak süpernatant yeni 1.5 mL tüplere pelletin dağılmadığına emin olunarak alındı. Qubit® dsDNA Broad Range Quantitation Assay (ThermoFisher) kullanılarak ölçüm gerçekleştirildi. Ölçüm sonucunun 0.8 ng/ µL'den yüksek olması gerekir, koşullar sağlandığı için DNB oluşturma aşamasına geçildi.

3.1.16 DNB Oluşturma

0.2 mL PZR tüpleri alındı ve Tablo 3.15'deki karışım hazırlandı.

Tablo 3.15 DNB Oluşturma Reaksiyonu 1

Bileşenler	Miktarı (µL)
ssDNA kütüphaneleri	X
Düşük TE Tamponu	20-X
DNB Oluşturma Tamponu	20
Toplam	40

Tablo 3.29'daki X hesaplanırken kullanılan formül: 364 (fragman boyutu) x 330 (sabit sayı) x a (fmol) / ss qubit x 106 . Formüldeki a verilecek runa göre hesaplanan lane başına düşen alanı temsil etmektedir. (PE100 runı için 1 lane 40 fmol alan kaplıyor ve 80 GB'a denk geliyor.) Kibarca vorteksleyerek karıştırıldı ve 5 sn boyunca mini sentrifüjle spin-down yapıldı. PZR makinesine hazırlanan karışım Tablo 3.16'daki program ayarlanarak başlatıldı.

Tablo 3.16 DNB Oluşturma Reaksiyonu 1 Koşulları

Sıcaklık	Zaman
Kapak ısıtması (105°C)	Açık
95 °C	1 dk
65 °C	1 dk
40 °C	1 dk
4 °C	Sonsuz

Oda sıcaklığında ve buz üzerinde uzun süre tutulmamasına özen gösterilmesi gereken DNB oluşturma enzim karışımı 2 (LC) çıkarılıp buz üzerine yerleştirildi. PZR makinesi 4°C'ye geldikten sonra PZR tüpleri alındı ve 5 sn kadar kısaca sentrifüjlendi. Buz üzerine tüpler bırakıldıktan sonra, DNB oluşturma reaksiyonu 2 karışımı Tablo 3.17'deki gibi hazırlandı.

Tablo 3.17 DNB Oluşturma Reaksiyonu 2

Bileşenler	Miktar (µL)
DNB Oluşturma Enzim Karışımı 1	40 µL
DNB Oluşturma Enzim Karışımı 2 (LC)	4 µL

Tüm DNB oluşturma reaksiyonu 2 karışımı, DNB oluşturma reaksiyonu 1 karışımına eklendi. Kibarca vortekslenerek karıştırıldı, 5 sn minisentrifüjle sentrifüjlendi, ardından tablo 3.18'deki koşullarda termal döngüde çalıştırıldı.

Tablo 3.18 DNB Oluşturma Reaksiyonu 2 Koşulları

Sıcaklık	Zaman
Kapak ısıtması (35°C)	Açık
30 °C	25 dk
4 °C	Sonsuz

Termal döngü 4°C'e gelir gelmez 20 µL DNSB oluşturma durdurma tamponu geniş delikli pipetle damla damla eklendi. Sekans aşamasından önce DNB'ler 4°C'de saklanabilir. Qubit® dsDNA Broad Range Quantitation Assay (ThermoFisher) kullanılarak ölçüm gerçekleştirildi. DNB viskoz olduğu için 2 µL alınması önerilir. Ölçüm sonucunun 8 ng/ µL'den yüksek olması gerekir, eğer düşük olursa yeniden DNB oluşturma aşamasının yapılması gerekmektedir, 8-40 aralığından yüksekse 20 ng/ µL'ye DNB yükleme tamponu 1 ile dilue edilmelidir.

3.1.17 DNB Yükleme

Yeni PZR tüplerine Tablo 3.19'daki DNB yükleme karışımı hazırlandı.

Tablo 3.19 DNB Yükleme Karışımı

Bileşenler	Miktar (µL)
DNB Yükleme Tamponu 2	8 µl
DNB Oluşturma Enzim Karışımı 2 (LC)	0,25 µl
DNB	25 µl

DNB yükleme karışımı, taşınabilir DNB yükleyici üzerinde, geniş uçlu pipet ucu ile flow celle 30 µl yüklendi ve flow cell oda sıcaklığında 30 dk inkübe edildi. Sekans kartuşu Tablo 3.20'deki miktarlara göre hazırlandı.

Tablo 3.20 Sekans Kartuşu Hazırlanışı

Kuyu Numarası	Eklene Reaktif	Miktar
1	dNTP Mix	1800 µl
	Sequencing Enzyme Mix	1800 µl
2	dNTP Mix II	1500 µl
	Sequencing Enzyme Mix	1500 µl
15	MDA Reagent	X µl
	MDA Enzyme Mix	500 µl

30 dk sonunda, oda sıcaklığında inkübe edilen flow cell ve hazırlanan sekans kartuşu cihazdaki yerlerine yerleştirildi ve cihaza PE100 okuması için gerekli bilgiler girilerek run başlatıldı.

3.1.18 Illumina NovaSeq 6000 Protokolü

DNA örneklerinin hazırlanması 3 güne yayılarak hazırlandı. İlk gün genomik DNA'nın etiketlenmesi, etiketli DNA'nın temizlenmesi, PZR, PZR ürünlerinin temizlenmesi, ilk hibridizasyon yapıldı. İkinci gün ilk yıkama ve ikinci hibridizasyon yapıldı. Üçüncü gün ikinci yıkama ve PZR ardından kütüphane doğrulaması yapıldı.

3.1.19 DNA Etiketlenmesi

Tagment DNA Buffer, Tagment DNA Enzim ve genomik DNA dondurudan çıkarılıp buz üzerine alındı. Çözöldükten sonra tüpler 3-5 kez hafifçe ters düz edilip mikrosantrifüjde kısa süre döndürülerek karıştırıldığından emin olundu. TAG termal döngü programı Tablo 3.21'deki gibi hazırlandı.

Tablo 3.21 TAG termal döngü programı

Sıcaklık	Zaman
Kapak sıcaklığı (100°C)	Açık
55 °C	15 dk
10 °C	Tutma

Total giriş miktarınının 100-500 ng olması için 2-30 µl DNA 96 kuyulu plate eklendi. DNA hacimleri 30 µl'den düşük ise nuclease-free su eklenerek tamamlandı. TAG karışımı Tablo 3.22'deki gibi hazırlandı ve vortekslendi.

Tablo 3.22 TAG master karışımı

Bileşenler	Miktar (μL)
BLT (Bead-Linked Transposomes)	11 μL
TB1(Tagmentation Buffer 1)	11 μL

Karışım eşit miktarda kuyulara bölündü. Mikroşal ile kapatılarak PZR programına konuldu.

3.1.20 Etiketleme Sonrası Temizlik

PZR'dan önce adaptör etiketli DNA'ların temizlenme aşamasına geçildi. 10 μL Tagment durdurucu buffer (TSB) tagmentasyon reaksiyonuna eklendi. 10 kere beadleri resuspend etmek için yavaşça pipetlendi. Mikroşal ile kapatılarak Tablo 3.23'teki termal döngü programına konuldu.

Tablo 3.23 PTC (Post Tagmentation Cleanup) termal döngü programı

Sıcaklık	Zaman
Kapak sıcaklığı (100°C)	Açık
37 °C	15 dk
10 °C	Tutma

Manyetik standın üzerine konarak oda sıcaklığında 3 dk kadar bekletildi ve süpernatant atıldı. 100 μL tagment yıkama bufferı doğrudan beadlerin üzerine köpürme yapmaması için yavaşça pipetleme yapılarak eklendi. Plate manyetik standın üzerine yerleştirilerek 3 dk oda sıcaklığında bekletildi. Supernatant atıldı ve bu yıkama işlemleri 2 kez tekrarlandı. Ardından 100 μL tagment yıkama bufferı doğrudan beadlerin üzerine köpürme yapmaması için yavaşça pipetleme yapılarak bir kez daha eklendi. Plate yine manyetik stand üzerinde bir sonraki aşamaya geçene kadar ve beadlerin aşırı kurumasını önlemek için tagment yıkama bufferı içinde tutuldu.

3.1.21 Etiketli DNA'nın Çoğaltılması

PZR bileşenleri Tablo 3.24'deki gibi hazırlandı ve vortekslendi. 280 * g 10 sn boyunca sentrifüjlendi. Plate manyetik stand üzerindeyken süpernant atıldı. Manyetik standtan alınarak her kuyucuğa 40 μL PZR master mix beadlerin üzerine eklendi. Boncuklar tamamen resuspend edilene kadar pipetaj yapıldı, 1600 rpm'de 1 dk karıştırıcıda bekletildi. Plate kapatılıp 3 saniye 280*g'de sentrifüj edildi.

Tablo 3.24 PZR master karışımı

Bileşenler	Miktar (μL)
Enhanced PZR karışımı (EPM)	22 μL
Nuclease-free su	22 μL

Uygun index adaptörler her örneğe eklendi ve 10 kere pipetaj yapıldı. Karıştırıcıya konularak 1600 rpm 1 dk bekletildi.

Tablo 3.25 Kullanılması gereken adaptör hacim hesaplamaları

İndex Kit Tipi	Kitin Kullanım Alanı	Her örnek için Hacim
24 plex (ikili dizin)	Tek tüpler	5 μL i7 adaptörü
		5 μL i5 adaptörü
96 plex (ikili dizin)	96 kuyulu plate	10 μL i7 ve i5 index adaptör çifti

Mikroseal ile plate kapatılarak, 30 sn 280*g'de sentrifüjlendi. Ardından PZR programı Tablo 3.26'daki gibi ayarlandı.

Tablo 3.26 Etiketli DNA'nın termal döngü programı

Sıcaklık	Zaman	
Kapak sıcaklığı (100°C)	Açık	
68 °C	3 dk	
98 °C	3 dk	
98 °C	45 sn	X döngü
62 °C	30 sn	
98 °C	2 dk	
68 °C	1 dk	
10 °C	Tutma	

Döngü sayısı total DNA'ya göre ayarlanır. 1-9 ng ise 12 PZR döngüsü, 10-24 ng 8 döngü, 25-49 ng 6 döngü, 50-99 ng 5 döngü, 100-500 ng 5 döngü şeklinde ayarlanır.

3.1.22 Kütüphanelerin Temizlenmesi

Tüpler bileşenler tüpün sonunda toplansın diye 280*g'de 1 dk sentrifüjlendi. Manyetik standı sıvı berraklaşınca kadar 5 dk kadar konuldu. 45 μL süpernatant her tüpten alınarak yeni bir orta boy plate aktarıldı. Örnek purifikasyonlar beadleri birkaç defa vorteksenerek resuspend edildi. Süpernatantın eklendiği her kuyuya 40 μL nuclease-free su eklendi. 45 μL örnek purifikasyon beadleri eklendi ve karıştırmak için 10 kere pipetaj

yapıldı. Oda sıcaklığında 5 dk inkübe edildikten sonra, manyetik standta alınarak sıvı berraklaşınca kadar (~5 dk) bekletildi. Bekletilme süresince yeni midi plate alınarak içine 15 µL örnek purifikasyon beadleri (dilue edilmemiş) eklendi. 125 µL süpernatant, 15 µL örnek purifikasyon eklenen yeni tüplere eklendi ve pipetaj yapıldı. 5 dk oda sıcaklığına ardından 5 dk manyetik standta inkübe edildi. Sıvı berraklaşınca kadar bekletildikten sonra supernant atıldı. Manyetik standta iken 200 µL 80% taze hazırlanmış EtOH eklendi. 30 sn inkübe edildi ve beadlere zarar vermeden supernant atıldı. Bu yıkama işlemi 2 kere tekrar edildi. Tüplerin içinde EtOH kalmamasına özen göstererek yeni pipet ucuyla kalanlar da çekildi, beadlere zarar verilmedi ve tüplerin ağzı açık bekletilerek kurumaya bırakıldı. Beadler çok kurumadan sadece EtOH buharlaştırıldıktan sonra 32 µL resuspenzyon bufferi pipetajla eklendi ve 2 dk inkübe edildi. Manyetik standta alınarak sıvı berraklaşınca kadar bekletildi ve yeni 96 kuyulu plate 30 µL transfer edildi.

3.1.23 Pool Kütüphaneleri

Giriş DNA 100-500 ng arasında olduğu için, 1.7 ml mikrosentrifüj tüpüne her kütüphaneden (384 kütüphaneye kadar) 5 µL eklendi, vortekslendi ve sentrifüjlendi. Quibit ile pool kütüphanesi ölçüldü.

3.1.24 Başlangıç Konsantrasyonu için Kütüphanelerin Seyreltilmesi

Ortalama kitaplık boyunun, 600 baz çifti kullanıldı. Molarite değerini kullanarak, kitaplıkların başlangıç konsantrasyonuna seyreltilmek için gereken TSB'ı ve hacmi hesaplandı. NovaSeq 6000 için 2 nM başlangıç konsantrasyonu kullanıldı. Okuma sayısı ve flow celle uygun olarak insan genomu, ekzom, transkriptom için belirlenen önerilen kütüphane sayısına baktıktan sonra; 4 kuyulu bir flow cell için ekzom datalarında ~250 kütüphane her lane için önerilir.

Tablo 3.27 Önerilen Yükleme Konsantrasyonları

Final Yükleme Konsantrasyonu (pM)	Pool Kütüphane Konsantrasyonu (nM)
150	0.75
200	1
250	1.25
300	1.50
350	1.75
400	2
450	2.25
500	2.50

10 mM Tris-HCl (pH 8,5) kullanılarak kütüphaneler, istenen pool kütüphanesi konsantrasyonlarına normalleştirildi. Optimum DNA yükleme konsantrasyonu, kütüphane tipine ve insert boyutuna bağlıdır. ≤ 450 bp'lik uç boyutlarına sahip kütüphaneler için (Illumina DNA PZR-free kütüphane poolları) 400-600² final yükleme konsantrasyonu ve 2-3² pool edilmiş yükleme konsantrasyonları önerilir. NovaSeq 6000 için 300 pm son yükleme konsantrasyonu önerilir.

3.1.25 Poolların Normalizasyonu ve PhiX Kontrol Eklenmesi

Her normalize edilmiş kütüphane yeni mikrosentrifüj tüpüne eklendi. 4 kuyulu flow cell için 310 μ L final hacmi önerilir.

Tablo 3.28 Önerilen Hacim

Flow Cell	0.2 N NaOH	Denatüre Edilmemiş Kütüphane Poolları (μ L)	Son Hacim (μ L)
SP/S1	25	100	125 ya da PhiX ile 125.6
S2	37	150	187 ya da PhiX ile 187.9
S4	77	310	387 ya da PhiX ile 388.9

0.2 N NaOH denatüre edilmemiş kütüphane poollarına eklendi ve kısaca vortekslendi. 280*g'de 1 dk sentrifüjlendi. Oda sıcaklığında 8 dk denatüre etmek için inkübe edildi. 400 mM Tris-HCl (pH 8.0) nötralize etmek için eklendi. (tek kuyu için 25, iki kuyu için 38, dört kuyulu flow cell için 78 μ L önerilir). Toplam hacim PhiX kontrol ile 466.9 μ L oldu ve kısaca vortekslendi. 280*g'de 1 dk sentrifüjlendi. Kütüphane tüplerine tüm hacim aktarıldı. Hemen kitaplık tüpünü, cluster kartuşuna yüklemeye ve run verilmeye özen gösterildi. Eğer 30 dk içinde yapılamıyorsa -25°C'de saklanabilir.

3.1.26 Sekans

SBS ve cluster reaktif kartuşları çözdürüldü. Kütüphane tüpü çözülmüş cluster kartuşuna yüklendi. Tek kütüphane poolü kütüphane tüpüne yüklendi ve ExAmp bileşeni ile karıştırılarak cluster ve sekans için flow celle yüklendi. Yazılım arayüzünden "sekans" seçildi ve tek flow cell run seçildi. Mevcut çalışma için yeni sarf malzemeler yüklendi. Run özellikleri seçilerek son bir kontrol yapılarak, başlatıldı.

4.1 Bulgular

Cihazdan ilk çıkan dosya formatı olan fastq formatını ilk olarak bam dosyasına çevirdik. Samtools bir yazılım paketidir ve YND datalarının dosya formatlarını birbirine çevirme işleminde kullanılabilir. Fastq dosyalarını bam dosyasına çevirirken; “samtools view -bS input.sam -o output.bam” kodunu kullanarak input.sam kısmına bizim fastq dosyamızın adı output.bam kısmına da oluşturmak istediğimiz bam dosyasının adı yazıldı. Bunun sebebi tüm ekzom sekanslama datalarını incelerken daha spesifik ve küçük bir alandaki okumaları kıyaslamak için 2 gen seçerek kesmek istememizdi. 2 geni içerecek şekilde bir bed dosyası tasarlandı. İnsan genomu hg37 (hg19) referans alınarak *MEFV* ve *TP53* genleri (*MEFV* için chr16:3292020-3306900 ve *TP53* için chr17:7571500-7590900) kesildi. Bam dosyasını keserken yine samtools kullanılarak “samtools view -b -h -L TP53BTD.bed bamdosyasınınadı.bam > kesilmişbamdosyasınınadı.bam” koduyla bed dosyasına göre bam kesilirken “samtools index kesilmişbamdosyasınınadı.bam” koduyla da bai dosyası oluşturuldu. Bu bam dosyaları ortalama derinlik ve Q20 değerini araştırmak için kullanılırken, fastq dosyaları ise GC içeriğini hesaplamak için kullanıldı.

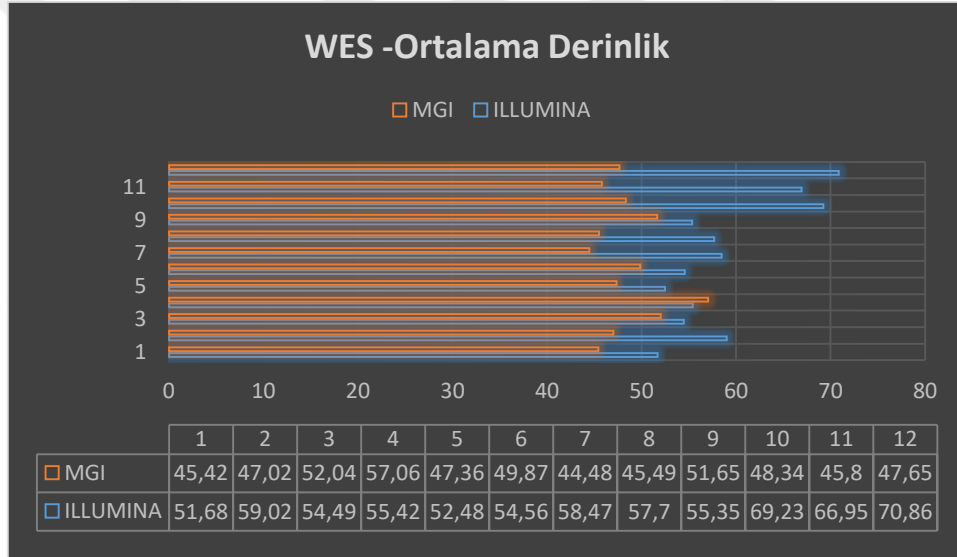
Kesilmiş bam dosyaları ardından tekrar fastq dosyalarına yine samtools kullanılarak “samtools bam2fq bamdosyasınınadı.bam > bamdosyasınınadı.fastq” kodu ile çevrildi. Illumina datalarının her birinin adı 9ETK ile başlarken, MGI datalarının her biri barkod ve lane ile ifade edildi. Hesaplamalar yapılırken orijinal isimleri kullanıldığı için sonuç tablosunda da aynı isimlerle yazıldı.

4.1.1 Ortalama Derinlik

Topladığımız 12 Illumina ve 12 MGI WES ve kesilmiş dataların ortalama derinlikleri pysam kütüphanesi kurulduktan sonra python kullanılarak yazılan bir script koduyla hesaplandı. Hem WES datalarının hem de kesilmiş dataların hepsi sırasıyla hesaplandı.

Tablo 4.1 Illumina ve MGI WES dalarının ortalama derinlikleri

ILLUMINA-WES	Sonuç	MGI-WES	Sonuç
9ETK1513	51,68	44899_L02_14	45,42
9ETK1526	59,02	47218_L04_124	47,02
9ETK1527	54,49	47218_L04_121	52,04
9ETK1528	55,42	47250_L01_43	57,06
9ETK1529	52,48	47218_L04_127	47,36
9ETK1531	54,56	44899_L02_77	49,87
9ETK1602	58,47	47250_L02_71	44,48
9ETK1603	57,7	44899_L02_80	45,49
9ETK1605	55,35	47218_L04_125	51,65
9ETK1606	69,23	44899_L02_79	48,34
9ETKS1075	66,95	53083_L01_42	45,8
9ETKS1077	70,86	47250_L02_14	47,65
Toplam	706,21	Toplam	582,18
Ortalama	58,85	Ortalama	48,515

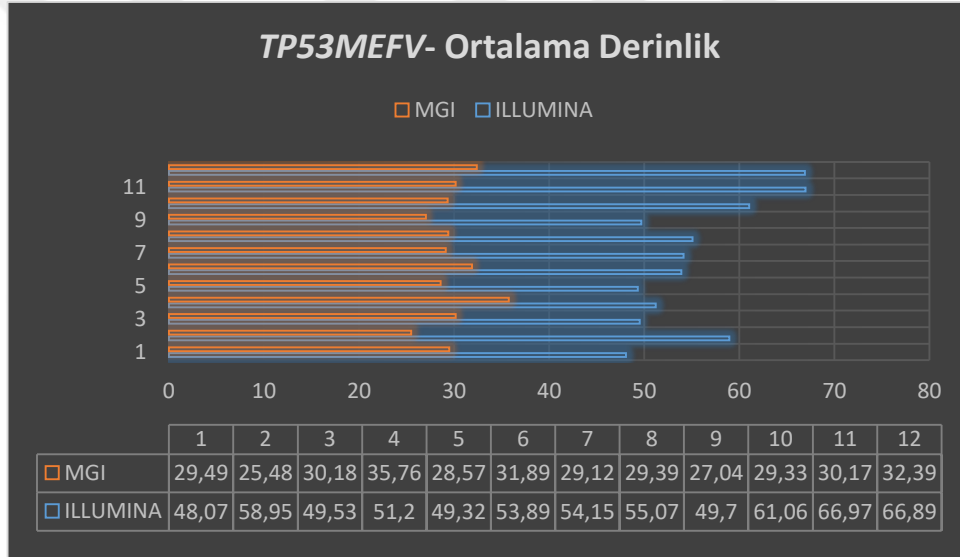


Şekil 4.1 Illumina ve MGI WES dalarının ortalama derinlik değeri

Tüm ekzom sekanslama dalarının ortalama derinlikleri kıyaslandığında, Illumina-WES dalarının ortalama derinlikleri ortalama olarak 58,85 iken, MGI-WES dalarının ise 48,52 olarak hesaplanmıştır. Ortalama derinlik ne kadar yüksekse, data doğruluğunun o kadar yüksek olacağı istenen bölgenin o kadar çok okunacağını düşünüldüğünden Illumina cihazında çalışılan WES dalarının göreceli olarak daha fazla okuma sayısına, daha iyi ortalama derinliğe dolayısıyla da daha güvenilir dataya sahip olduğu söylenebilir.

Tablo 4.2 Illumina ve MGI datalarının *TP53-MEFV* genleri için ortalama derinlikleri

ILLUMINA-WES	Sonuç	MGI-WES	Sonuç
9ETK1513	48,07	44899_L02_14	29,49
9ETK1526	58,95	47218_L04_124	25,48
9ETK1527	49,53	47218_L04_121	30,18
9ETK1528	51,2	47250_L01_43	35,76
9ETK1529	49,32	47218_L04_127	28,57
9ETK1531	53,89	44899_L02_77	31,89
9ETK1602	54,15	47250_L02_71	29,12
9ETK1603	55,07	44899_L02_80	29,39
9ETK1605	49,7	47218_L04_125	27,04
9ETK1606	61,06	44899_L02_79	29,33
9ETKS1075	66,97	53083_L01_42	30,17
9ETKS1077	66,89	47250_L02_14	32,39
Toplam	664,8	Toplam	358,81
Ortalama	55,4	Ortalama	29,90



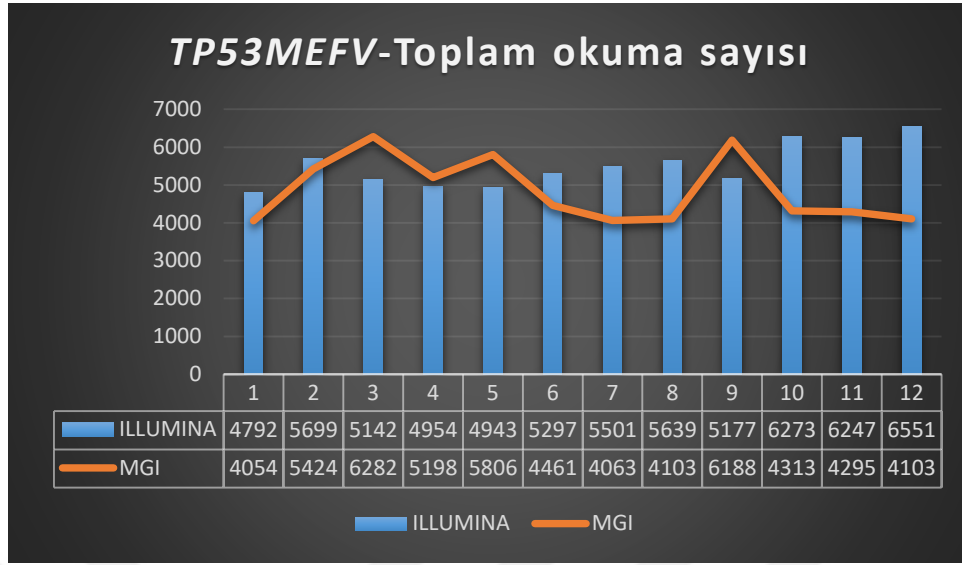
Şekil 4.2 Illumina ve MGI cihazlarında çalışılan WES datalarından kesilmiş TP53 ve MEFV genlerini içeren bölgelerin okumalarının ortalama derinliği

Tüm ekzom sekanslama datalarından seçtiğimiz iki geni kesmemizin nedeni ortalama derinlik ve diğer data kalite parametrelerini daha spesifik ve küçük bölgelerde de kıyaslayabilmektir. Yani tüm ekzom sekanslama işleminde bazı genler daha az okunabiliyorken (pseudogeni olan genler gibi) bazı genlerin okuması daha iyi olabilir. Bu nedenle pseudogeni olmadığı bilinen ve genellikle okumasında sorun olmayan ve iyi okuma alınan iki önemli geni seçtik. Illumina kesilmiş datalarının ortalama derinlikleri ortalama olarak 55,4 iken, MGI kesilmiş datalarının ise 29,90 olarak hesaplanmıştır. Kesilmiş daha spesifik alanlar araştırıldığında da Illumina cihazından elde edilen dataların göreceli olarak daha fazla ortalama derinliğine sahip olduğu görülmüştür.

Ortalama derinliğe ek olarak, sadece *TP53* ve *MEFV* genlerinin okumalarını kıyaslayabilmek için toplam okuma sayılarını da araştırdık. Bu kalite parametresini tüm ekzom sekanslama datalarında yapmak kodu çalıştırırken çok uzun süreceğinden ve çok büyük sayılar çıkacağından yapılmadı. *TP53* ve *MEFV* genlerinin toplam okuma sayıları Illumina cihazından çıkan datalarda ortalama 5517,92 iken MGI cihazından çıkan datalarda ise ortalama 4857,5 olarak hesaplandı. Okuma sayısının fazla olması data güvenilirliğini artıran bir husus olduğu için yine Illumina cihazından çıkan dataların güvenilirliği göreceli olarak daha iyi diye yorumlandı.

Tablo 4.3 Illumina ve MGI cihazlarında *TP53* ve *MEFV* genlerinin toplam okuma sayıları

ILLUMINA-TP53/MEFV	Sonuç	MGI-TP53/MEFV	Sonuç
9ETK1513	4792	44899_L02_14	4054
9ETK1526	5699	47218_L04_124	5424
9ETK1527	5142	47218_L04_121	6282
9ETK1528	4954	47250_L01_43	5198
9ETK1529	4943	47218_L04_127	5806
9ETK1531	5297	44899_L02_77	4461
9ETK1602	5501	47250_L02_71	4063
9ETK1603	5639	44899_L02_80	4103
9ETK1605	5177	47218_L04_125	6188
9ETK1606	6273	44899_L02_79	4313
9ETKS1075	6247	53083_L01_42	4295
9ETKS1077	6551	47250_L02_14	4103
Toplam	66215	Toplam	58290
Ortalama	5517,92	Ortalama	4857,5



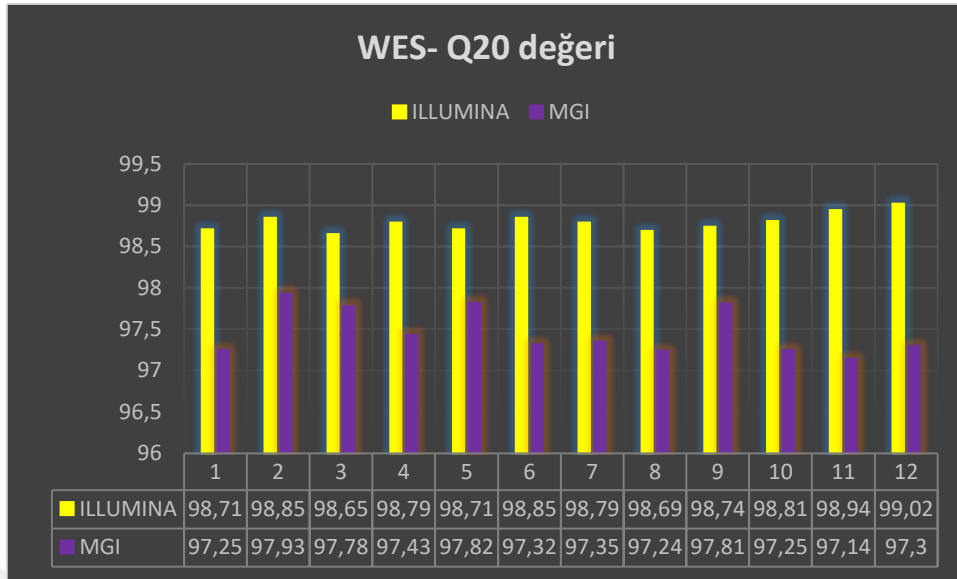
Şekil 4.3 Illumina ve MGI cihazlarında TP53 ve MEFV genlerinin toplam okuma sayıları

4.1.2 Q20 değeri

Q20 değeri her 100 bazda bir hata oranını veren bir data kalite parametresi olmakla birlikte, hata oranını verdiğinden en önemli parametrelerden biri olarak geçmektedir. Q20 değerini yine python kullanarak tüm datalar için hesapladık.

Tablo 4.4 Illumina ve MGI cihazlarından çıkan WES datalarının Q20 değerleri

ILLUMINA-WES	Sonuç	MGI-WES	Sonuç
9ETK1513	98,71	44899_L02_14	97,25
9ETK1526	98,85	47218_L04_124	97,93
9ETK1527	98,65	47218_L04_121	97,78
9ETK1528	98,79	47250_L01_43	97,43
9ETK1529	98,71	47218_L04_127	97,82
9ETK1531	98,85	44899_L02_77	97,32
9ETK1602	98,79	47250_L02_71	97,35
9ETK1603	98,69	44899_L02_80	97,24
9ETK1605	98,74	47218_L04_125	97,81
9ETK1606	98,81	44899_L02_79	97,25
9ETKS1075	98,94	53083_L01_42	97,14
9ETKS1077	99,02	47250_L02_14	97,3
Toplam	1185,55	Toplam	1169,62
Ortalama	98,79	Ortalama	97,47

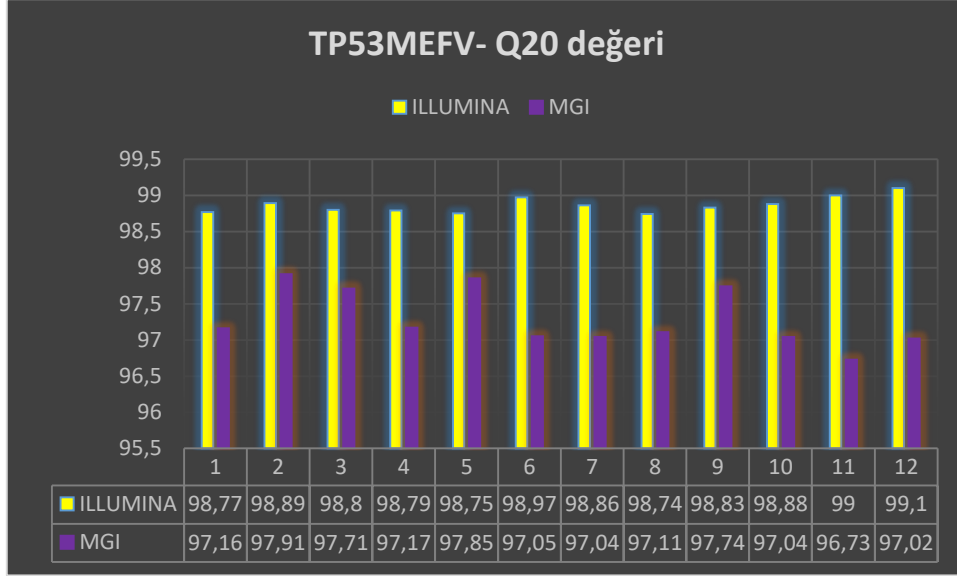


Şekil 4.4 Illumina ve MGI cihazlarıyla çalışılan WES datalarının Q20 değerleri

Tüm ekzom sekanslama datalarının Q20 değerleri kıyaslandığında; Illumina için ortalama 98,80 iken, MGI için ortalama 97,47’dir. Q20 değeri 100 bazda 1 hata oranını gösterdiği için bu değer ne kadar büyük olursa o kadar az hata oranı olduğunu dolayısıyla daha güvenli dataya sahip olduğumuzu göstermektedir. Dolayısıyla Illumina datalarının ortalama Q20 değeri MGI datalarından daha yüksek olduğundan göreceli olarak hata oranı daha az ve daha güvenilir dataya sahiptir.

Tablo 4.5 Illumina ve MGI cihazlarının *TP53* ve *MEFV* genleri için Q20 değerleri

ILLUMINA-TP53/MEFV	Sonuç	MGI-TP53/MEFV	Sonuç
9ETK1513	98,77	44899_L02_14	97,16
9ETK1526	98,89	47218_L04_124	97,91
9ETK1527	98,8	47218_L04_121	97,71
9ETK1528	98,79	47250_L01_43	97,17
9ETK1529	98,75	47218_L04_127	97,85
9ETK1531	98,97	44899_L02_77	97,05
9ETK1602	98,86	47250_L02_71	97,04
9ETK1603	98,74	44899_L02_80	97,11
9ETK1605	98,83	47218_L04_125	97,74
9ETK1606	98,88	44899_L02_79	97,04
9ETKS1075	99	53083_L01_42	96,73
9ETKS1077	99,1	47250_L02_14	97,02
Toplam	1186,38	Toplam	1167,53
Ortalama	98,87	Ortalama	97,29



Şekil 4.5 Illumina ve MGI datalarının TP53 ve MEFV için Q20 değerleri

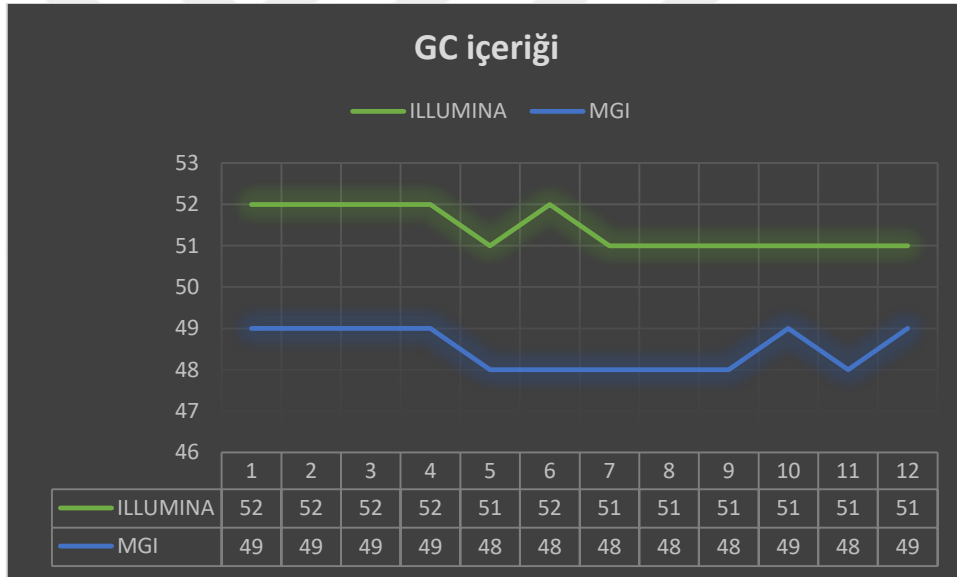
TP53 ve *MEFV* genlerinin kesildiği dataların Q20 değerleri kıyaslandığında; Illumina için ortalama 98,80 iken, MGI için ortalama 97,47'dir. Q20 değeri 100 bazda 1 hata oranını gösterdiği için bu değer ne kadar büyük olursa o kadar az hata oranı olduğunu dolayısıyla daha güvenli dataya sahip olduğumuzu göstermektedir. Dolayısıyla Illumina datalarının ortalama Q20 değeri MGI datalarından daha yüksek olduğundan göreceli olarak hata oranı daha az ve daha güvenilir dataya sahiptir.

4.1.3 GC İçeriği

Hem WES dataların hem de kesilmiş dataların GC içeriğini hesaplarırken fastq dosyalarını kullandık. FastQC adı verilen bir yazılımla; tüm dataların toplam sekanslarını, sekans uzunluklarını, GC içeriklerini, her baz sekans kalitesini, sekans içeriğini, okunamayan baz yüzdesini, duplikasyon seviyesini, üst üste okunan diziler ve adaptör içeriklerini inceledik. Bu parametrelerden sadece GC yüzdelerini kıyaslama yapmak için kullandık.

Tablo 4.6 Illumina ve MGI cihazlarından çıkan WES datalarının GC İçerikleri

ILLUMINA-WES	Sonuç	MGI-WES	Sonuç
9ETK1513	52	44899_L02_14	49
9ETK1526	52	47218_L04_124	49
9ETK1527	52	47218_L04_121	49
9ETK1528	52	47250_L01_43	49
9ETK1529	51	47218_L04_127	48
9ETK1531	52	44899_L02_77	48
9ETK1602	51	47250_L02_71	48
9ETK1603	51	44899_L02_80	48
9ETK1605	51	47218_L04_125	48
9ETK1606	51	44899_L02_79	49
9ETKS1075	51	53083_L01_42	48
9ETKS1077	51	47250_L02_14	49
Toplam	617	Toplam	582
Ortalama	51,42	Ortalama	48,5

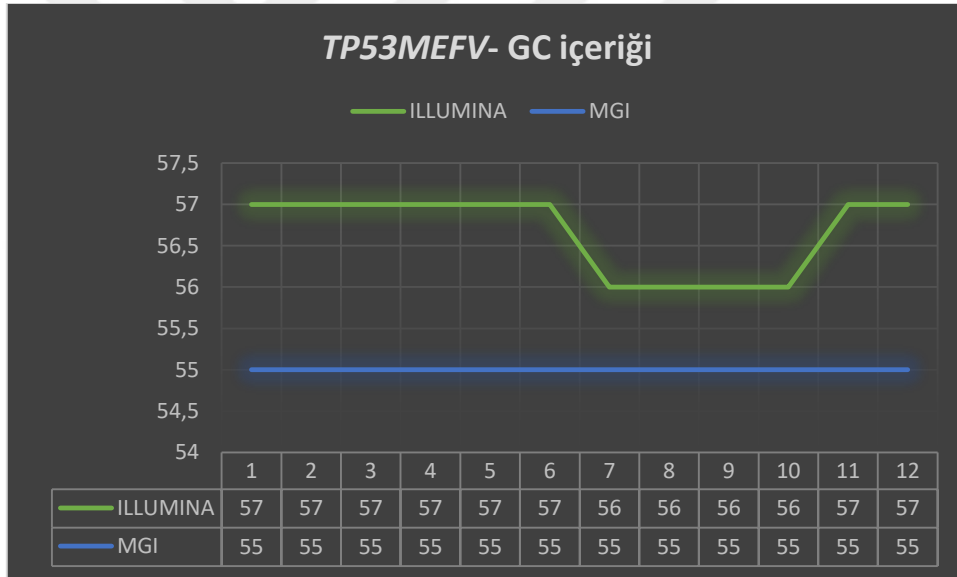


Şekil 4.6 Illumina ve MGI cihazlarından çıkan WES datalarının GC İçerikleri

GC içeriği hakkında net bir yüzdeler belirtilmemekle birlikte, çok düşük (30-40%) ile çok yüksek (60-70%) GC içerikleri güvenilir bulunmamaktadır. İdeal olan GC içeriği 45-50% olarak belirlenmiştir. Dolayısıyla hem Illumina hem de MGI cihazlarıyla çalışılan WES datalarının GC içeriği ideal aralıktadır.

Tablo 4.7 Illumina ve MGI cihazlarının *TP53* ve *MEFV* genleri için GC içerikleri

ILLUMINA-TP53/MEFV	Sonuç	MGI-TP53/MEFV	Sonuç
9ETK1513	57	44899_L02_14	55
9ETK1526	57	47218_L04_124	55
9ETK1527	57	47218_L04_121	55
9ETK1528	57	47250_L01_43	55
9ETK1529	57	47218_L04_127	55
9ETK1531	57	44899_L02_77	55
9ETK1602	56	47250_L02_71	55
9ETK1603	56	44899_L02_80	55
9ETK1605	56	47218_L04_125	55
9ETK1606	56	44899_L02_79	55
9ETKS1075	57	53083_L01_42	55
9ETKS1077	57	47250_L02_14	55
Toplam	680	Toplam	660
Ortalama	56,67	Ortalama	55



Şekil 4.7 Illumina ve MGI cihazlarının *TP53* ve *MEFV* genleri için GC içerikleri

Hem WES dataları hem de kesilmiş datalar için GC içeriği ideal aralıktadır.

4.2 İstatistiksel Hesaplamalar

Illumina ve MGI cihazlarından çıkan dataların ortalama derinlik, Q20 değeri ve GC içeriklerinin kıyaslanabilmesi için, iki bağımsız değişkene sahip olduğumuz ve belli bir hipotezimiz olmadığı için 't testi' uygulandı. Hem excel hem de Python kullanılarak hesaplamalar gerçekleştirildi ve ikisinde de aynı sonuçlar bulundu.

Tablo 4.8 T testi uygulanarak bulunan p değerleri

Parametreler	P Değeri (WES)	P Değeri (MEFV-TP53)
Ortalama Derinlik	0,000089166	0,000000000023858
Q20 değeri	0,000000000000349	0,000000000005453
GC İçeriği	0,000000000002687	0,0000000000618662
Toplam Okuma Sayısı	-	0,039535548
Data Boyutu	0,0000000064131	-

P değeri kıyaslanmak istenen iki farklı parametrenin arasındaki ilişkinin anlamlı olup olmadığını gösteren bir değerdir. Bu değer eğer <0.05 ise iki değer arasındaki farkın istatistiki olarak anlamlı olduğu kabul edilir. P değeri eğer >0.05 ise iki değer arasında istatistiki olarak anlamlı bir fark olmadığı kabul edilir. P değeri ne kadar düşükse elde edilen sonuçlar o kadar daha az rastgele olur ve anlamlılık artar. İncelediğimiz tüm parametrelerde p değerimiz <0.05 olduğu için Illumina NovaSeq 6000 platformu ile MGI DNBSseq 6400 platformu kıyaslandığında istatistiki olarak anlamlı farklar elde ettik. Özetle; ortalama derinlik, GC içeriği, Q20 değeri, ortalama okuma sayısı, data boyutu gibi data kalite parametreleri tüm ekzom sekanslama ve *TP53-MEFV* genleri için kıyaslandığında Illumina ve MGI teknolojilerinin arasındaki fark istatistiki olarak anlamlıdır.

4.3 Tartışma

YND datalarının kalitesini kontrol ve kıyaslama yapmak için göz önünde bulundurulmuş birçok kalite parametresi bulunmaktadır. Biz bu çalışmamızda GC içeriği, ortalama derinlik, Q20 değeri, toplam okuma sayısı ve data boyutu gibi parametreleri inceledik. Ama bununla birlikte birçok çalışma Q30 değerini inceler. 1000 bazda 1 hata payını verdiği için hassasiyeti daha fazla olmakla birlikte daha çok kullanılan bir kontrol parametresidir. Fastq dosyalarından fastQC adı verilen yazılımla elde edilen kontrol parametreleri: temel istatistikler, her baz başına sekans kalitesi, her sekansın kalite skoru, GC içeriği, okunamayan baz içeriği, sekans uzunluğu dağılımı, sekans duplikasyon seviyesi, fazla okunan sekanslar, adaptör ve kimer içeriğidir. Burada elde edilen tüm kalite parametreleri iki farklı cihazdan elde edilen dataları kıyaslamak için kullanılabilir. Temel istatistikler dosya adı, dosyanın türü, sekans uzunluğu ve %GC içeriği gibi özet bilgilerin olduğu tablodur. Her baz başına sekans kalitesi; her bazın kalite değerlerini grafik olarak verir. Yüksek kalite değerleri iyi data olarak değerlendirilir. GC içeriği; 4 bazdan G ve C'nin toplam dağılımının yüzde olan ifade edilmiştir. Okunamayan baz içeriği, okunamayan bazlar N yani nükleotid olarak ifade edilir, normal bir baz

okumasının yeterli seviyede güven vermediği durumlarda okunur. Sekans uzunluğu dağılımı; tüm dizilerde ortalama sekans uzunluğunu verir ve bu ölçüm dizileme platformuna bağlı olarak değişir. Sekans duplikasyon seviyesi; her sekans için yineleme derecesini sayar ve bir grafik oluşturur. Fazla okunan (aşırı temsil edilen) diziler; önemli bir parametredir. Çünkü aslında çalışmanın kalitesini de belirler. Kütüphane hazırlığı aşamasında bir kirliliğin olup olmadığını ya da çeşitli olmadığını belirtir. Adaptör ve kimer içerikleri de adaptör dizisine sahip dizilerin ve kimer ($k=7$, k-mer) oluşturan dizilerin oranlarına bakılır [3]. Literatürdeki çalışmalarda bu parametreler de incelenmiştir.

Kumar ve arkadaşlarının yaptığı bir çalışmada [24], YND teknoloji enstrümanlarının karşılaştırılması yapılmıştır. Bu karşılaştırma Illumina, Ion Torrent, BGI, PacBio ve Nanopore firmalarının çeşitli cihazları karşılaştırılarak yapılmıştır. Bu cihazları karşılaştırırken kullandıkları parametreler ise; run başına maksimum okuma sayısı, maksimum data çıkış aralığı, okuma uzunluğu, run süresi (saat olarak) ve uygulama alanlarıdır. İkinci ve üçüncü nesil dizileme teknolojilerinin ayrı ayrı ve hibrit kullanımları kıyaslanmıştır. Her yöntemin diğerlerine göre avantajlarına değinilerek, araştırılmak istenen amaca göre cihaz seçimi için uygun bir yayın çıkarılmıştır. Kullanım alanları ve istenen okuma uzunluğu, süre ve okuma sayısı gibi parametrelerle cihaz seçimini kolaylaştıran bir yayın oluşturulmuştur.

Aynı şekilde Clark ve arkadaşlarının yaptığı bir çalışmada [25], Agilent, Illumina ve Nimblegen cihazlarıyla tüm ekzom sekanslama dataları karşılaştırılmıştır. Bu platformlar hedef zenginleştirme verimi, off-target zenginleştirme, GC içerikleri, SNV tayini ve küçük indellerin tayini gibi parametrelere göre kıyaslandılar. İlk olarak DNA örnekleri Stanford Üniversitesi Hastanesi'nden ve Qiagen'in genomik DNA izolasyon kitiyle izole edilmiştir. Ardından kütüphane hazırlığı, ekzom zenginleştirme yapılmıştır. Elde ettikleri WES datalarını WGS ile kıyaslamak için WGS datası da elde edilmiştir. Ardından GATK (Genome Analysis Toolkit) kullanarak varyant çağırma işlemi gerçekleştirilmiştir. SNV tespitinde, WES dataları WGS datalarından genel olarak daha yüksek Phred skoruna sahiptir. Nimblegen için ekzom datasında 573, genom datasında 320, Agilent için ekzom datasında 428 genom datasında 192 ve Illumina için ekzom datasında 341 genom datasında 380 olarak tespit edilmiştir. Sadece Illumina dataları için ekzom ve genom dataları arasındaki fark az tespit edilmiştir. Off-target bölgelerin okunması oranında ise Agilent için 12.8%, Nimblegen için 9.3% ve Illumina için 35.6% tespit edilmiştir. Üç

platform için de yüksek orandaki off-target zenginleşme oranı tekrar eden ögelerde ve segmental duplikasyonlarda saptanmıştır. Bu cihazlar tasarımı gereği, Agilent RNA'ları DNA'dan daha çok kullanan ve Ensembl genlerini daha fazla kapsayan bir cihaz iken, Illumina 27.73 Mb UTR bölgelerini de hedefleyen cihazdır. Nimblegen ise daha fazla miRNA'yı okuyabilen cihaz olarak sınıflandırılmıştır. Bununla birlikte elde ettikleri WES datalarında align edilen okuma sayısı Agilent için 99.3%, Nimblegen (99.2%) ve Illumina için 98.3% olarak hesaplanmıştır.

Yapılan bir başka çalışmada da Korostin ve arkadaşları tarafından [26] gerçekleştirilmiş ve tüm ekzom sekanslama dataları MGISEQ-2000 ve Illumina HiSeq 2500 cihazları kullanılarak kıyaslanmıştır. Varyant arama parametreleri Samtools mpileup, GATK, Strelka2 ve Sentieon kullanarak ayrı ayrı gerçekleştirmişlerdir. Etik izinler alındıktan sonra DNA örneği (25 yaşında yumurta bağışlama kriterlerini karşılayan bir kadından) toplanmış ve izole edilmiştir. DNA kütüphaneleri oluşturulmuş ve cihazlarda ayrı ayrı sekanslanmıştır. FastQC raporları kullanılarak okuma, baz, GC içeriği, Q20 ve Q30 değerleri araştırılmıştır. MGISEQ'te okuma ($*10^6$) 780 iken Illumina'da 726, baz MGI'da 117 iken Illumina'da 108.9 bulunmuştur. GC içerikleri aynı ve %40 saptanmıştır. Bununla birlikte Q20 ve Q30 değerleri MGI için sırasıyla 99.92% ve 95.03% iken Illumina için sırasıyla 99.99% ve 97.18% olarak hesaplanmıştır. Varyant çağırma adımlarını tek tek dört farklı pipeline kullanarak yaptıklarında hem MGI hem Illumina için benzer sonuçlar elde etmişler. Bununla birlikte MGISEQ-2000 ile Illumina HiSeq 2500 karşılaştırıldığında, Illumina tarafından üretilen veriler için ek olarak %3-4 oranında duplikasyon ile sonuçlanmıştır. Sonuç olarak iki platform için de genel olarak yakın veya benzer sonuçlar elde etmişken, aralarındaki farkı varyant arama ve indel deteksiyonu gibi adımlarda bulmuşlardır. Varyant çağırma işlemi için 4 farklı pipeline kullanıp teyit etmişlerdir. Ve bu elde ettikleri farkında istatistiki olarak önemsiz olduğu sonucuna varmışlardır.

Bir başka çalışma da Zhang ve arkadaşları tarafından yapılmış [27] ve bu çalışmada Ion TargetSeq ve HiSeq 2000 cihazları kullanılarak iki farklı ekzom capture kitleri kullanılarak varyant aramadaki farklılıklar kıyaslanmıştır. 4 farklı kişiden DNA'ları toplayıp Qiagen'in kitiyle izole edip kütüphane hazırlığı yaptıktan sonra iki farklı ekzom kiti kullanarak sekanslamışlar ve de varyant araması yaptıktan sonra Sanger yapmışlardır. Sanger doğrulama yaparak iki adımlı gittikleri çalışmada HiSeq ile elde edilen datanın göreceli olarak daha yüksek doğruluğa sahip olduğu ve özellikle İnDel tespiti konusunda

daha iyi olduğu sonucu varmışlardır. 4 birey için de ayrı ayrı TargetSeq-Proton ve SureSelect-HiSeq platformları kullanılarak yapılan ekzom sekanslama işlemlerinde 20x'teki kapsamlar sırasıyla; (önce TargetSeq ardından HiSeq yazılacaktır) 79-71, 86-74, 84-63 ve 84-66 tespit edilmiştir. Ve ortalama kapsama derinlikleri dört birey için yine aynı sırayla; 39.8-40, 51.9-45.3, 50.3-33.4 ve 49.8-36 olarak saptanmıştır. Daha yüksek kapsama derinliği daha iyi sonuçlar anlamına gelir. Aynı şekilde SNP ve küçük indeller için yaptıkları araştırmada iki platform için önemli farklılıklar tespit edilmiştir. Küçük indeller HiSeq'te 55.3% saptanırken, TargetSeq'te 78.4% olarak saptanmıştır. SNP'lerin doğrulanması için Sanger sekanslama yapılmış ve HiSeq için bu oran 88.3% iken TargetSeq için 60.0% olarak bulunmuştur. Dolayısıyla 4 farklı tüm ekzom sekanslama datasının kıyasında; SNP ve küçük indellerin tespitinde SureSelect HiSeq, TargetSeq-Proton'a oranla daha yüksek doğruluk oranına sahip varyant arama özelliğine sahip olduğu bulunmuştur.

Son olarak Lang ve arkadaşlarının yaptığı [28] ve bizim çalışmamıza en çok benzeyen bizim de tez için yararlandığımız çalışma olan Illumina NextSeq500 ve MGISEQ-2000 cihazlarının karşılaştırılması çalışması mevcuttur. 272 adet 29-91 yaş grubu arasında örnek toplanmıştır. 79'u plasma örneği, 21'i beyaz kan hücresi ve 172'si formalin ile sabitlenmiş ya da parafin blok olarak seçilmiştir. Bu çalışmada datalar Q20 oranı, GC içeriği, ortalama derinlik ve Capture verimi parametrelerine göre kıyaslanmıştır. Bunu yaparken MGI ile dizilenen datalarda 101-133 bç fragment kaybının olduğunu saptamışlardır. Ama kütüphane hazırlığını Illumina kitlerine göre yaptıklarını için MGISEQ protokolünü kullanmadıkları için bu farkın sebebini kit farklılığından dolayı diye yorumlamışlardır. MGI cihazları DNB oluşturma teknolojisini kullanırken Illumina köprü PZR'ı yöntemini kullanmaktadır, bu da aradaki parça boyutu kaybını açıklayabilir. Bu aradaki parça boyutu kaybından dolayı, hemolize olmuş plasma örnekleri ya da parafin bloğa uzun süredir sabitlenmiş parafin blok gibi kısa boyutlu numunelerin Illumina ile dizilenmesinin daha mantıklı olduğu kararına varmışlardır. Kalite kontrol parametreleri arasındaki farkı değerlendirmek için t testi uygulamışlardır ve tüm parametrelerde aradaki farkın istatistiki olarak anlamlı olduğu sonucuna varmışlardır.

Biz de bu çalışmamızda 24 farklı hastadan alınan DNA örnekleriyle yapılan tüm ekzom sekanslama datalarında GC içeriği, ortalama derinlik ve Q20 değerlerini kıyaslayarak iki yönlü hipotez için t testi uygulayarak NovaSeq ve DNBSek arasında istatistiki olarak anlamlı farklar elde ettik.

4.4 Sonuç

Illumina NovaSeq 6000 ve MGI DNB Seq-G400 cihazlarını kullanarak yaptığımız tüm ekzom sekanslama ve gen bazlı sekanslama datalarının kıyaslanmasında, NovaSeq cihazı ortalama derinlik ve Q20 değeri açısından DNB Seq-G400 cihazına oranla daha yüksek değerler verdiğiinden ve aralarındaki farkın istatistiki olarak anlamlı olmasından dolayı yüksek okuma oranı ve daha güvenilir data istenen çalışmalarda tercih edilebileceği sonucuna varılmıştır. Bununla birlikte GC içeriklerinin iki cihaz içinde optimum değer aralığında olması, okumaların kirli olmaması, elde edilen data boyutlarının yeterli olması, okuma sayısının yeterli olması ve Q20 değerlerinin ikisi için de istenen değerde olup güvenilir datalar vermesi gibi parametreleri her iki teknoloji de karşıladığından hem gen bazında hem de tüm ekzom sekanslama çalışmaları için ikisinin de kullanılabilmesine karar verilmiştir.

4.5 Öneriler

Bizim çalışmamızda birbirinden tamamen farklı kurumlardan gelen 24 farklı bireyin tüm ekzom sekanslama çalışması yapılmıştır. İki cihazın kıyaslanması yapılırken başka farklı parametre olmaması adına, birbiriyle aynı örnekler kıyaslanarak sadece cihaz kalitesi kıyaslanabilirdi. Genlerin okumaları kıyaslanırken yeni bir çalışma ile değil tüm ekzom sekanslama datasından kesilerek yapıldı. Dolayısıyla, yeni yapılacak bir çalışma için değişken olan koşulun sadece iki cihaz ve onların teknolojileri olması, gen bazında bir araştırma yapılacaksa eğer sadece o genlerin primerlerini içeren ayrı bir çalışma konularak sadece istenen genlerin okumalarının kıyaslanması ve daha fazla data kalite parametresi kullanılarak daha kapsamlı bir kıyaslama ile sonuçlandırılması önerilir.

- [1] Illumina (2023). About Us, <https://emea.illumina.com/company/about-us.html> (Erişim Tarihi 04.03.2023).
- [2] MGI (2021). About MGI- History, <https://en.mgi-tech.com/about/> (Erişim Tarihi 04.03.2023).
- [3] Babraham Bioinformatics (2015). FastQC, <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (Erişim Tarihi 05.03.2023).
- [4] Collins, Francis S. ve Fink Leslie (1995). “The Human Genome Project”, Alcohol Health Research World, Cilt 19, No 3, s. 190-195.
- [5] Samiksha, S (2014). “Techniques Used in Human Genome Project”, <https://www.yourarticlelibrary.com/biology/gene-sequencing/techniques-used-in-human-genome-project/29586> (Erişim Tarihi 05.03.2023)
- [6] DNA Dizi Analiz Yöntemleri (2000). Hacettepe Üniversitesi, Ankara
- [7] Aryal, Sagar (2022). “DNA Sequencing-Definition, Principle, Steps, Types, Uses”, <https://microbenotes.com/dna-sequencing-maxam-gilbert-and-sanger-dideoxy-method/> (Erişim Tarihi: 05.03.2023)
- [8] Üstek, Duran et al. (2009). “Yeni Nesil DNA Dizileme”, Deneysel Tıp Araştırma Enstitüsü Dergisi, Cilt 1, No 1, s. 11-18.
- [9] Dr. Shaffer, Catherine (2019). “Shotgun Sequencing”, <https://www.news-medical.net/life-sciences/Shotgun-Sequencing.aspx> (Erişim Tarihi: 07.03.2023)
- [10] Ari, Şule ve Arıkan, Muzaffer (2016). “Next-Generation Sequencing: Advantages, Disadvantages, and Future”, Hakeem, Khalid Rehman et al. (der.), Plant Omics: Trends and Applications (s.109-135). Newyork, Springer.
- [11] Polat, Gencay Kaan (2020). “DNA Dizileme Yöntemleri: Genler Nasıl Dizilenir? Yeni Nesil DNA Sekanslama Yöntemleri Nelerdir?”, <https://evrimagaci.org/dna-dizileme-yontemleri-genler-nasil-dizilenir-yeni-nesil-dna-sekanslama-yontemleri-nelerdir-9770> (Erişim Tarihi: 09.03.2023)
- [12] DNA Dizileme Yöntemleri (2019). Biyoinformatik Dünyası, <http://biyoinformatikdunyasi.blogspot.com/2019/07/dna-dizileme-yontemleri-2.html> (Erişim Tarihi: 05.03.2023)
- [13] Dönmez Dicle et al. (2015). “Yeni Nesil DNA Dizileme Teknolojileri ve Bitkilerde Kullanımı”, Türk Bilimsel Derlemeler Dergisi, Cilt 8, No 1, s. 30-37.
- [14] Bölükbaşı, Ekrem ve Aras Sumer, Emine (2015). “Third Generation DNA Sequencing Technologies”, International Journal of Scientific and Technological Research, Cilt 1, No 1, s 14-21.
- [15] Ozaki, Toshinori ve Nakagawara, Akira (2011). “Role of p53 in Cell Death and Human Cancers”, Cancers, Cilt 3, No 1, s. 994-1013.
- [16] Varan, Okan et al. (2019). “Ailevi Akdeniz Ateşi Hastalarında MEFV Varyantları ve Klinik Korelasyonları”, Ankara Eğitim ve Araştırma Hastanesi Dergisi, Cilt 52, No 2, s. 162-166.

- [17] Prof. Dr. Bağış, Haydar (2020). “Ailevi Akdeniz Ateşi (FMF) Genetik Geçişli Hastalıktır”, <https://www.haydarbagis.com/tr/news/desc/4449/ailevi-akdeniz-atesi-fmf-genetik-gecisli-hastaliktir.html> (Erişim Tarihi: 09.03.2023)
- [18] Amr, Sami S. ve Funke, Birgit (2015). “Targeted Hybrid Capture for Inherited Disease Panels”, *Clinical Genomics Chapter 16, USA, Academic Press.*
- [19] Ajay, Subramanian S. et al. (2011). “Accurate and comprehensive sequencing of personal genomes”, *Genome Research, Cilt 21, No 9, s. 1498-1505.*
- [20] Rehm, Heidi L. et al. (2013). “ACMG clinical laboratory standards for next-generation sequencing”, *Genetics in Medicine, Cilt 15, No 9, s. 733-747.*
- [21] *Quality Scores for Next-Generation Sequencing (2012). Illumina, Amerika Birleşik Devletleri.*
- [22] Cock, Peter J. et al. (2010). “The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants”, *Nucleic Acid Research, Cilt 38, No 6, s. 1767- 1771.*
- [23] *The fastq sequence format (2022). OBITools, Almanya*
- [24] Kumar, Kishore R. et al. (2019). “Next-Generation Sequencing and Emerging Technologies”, *Seminars in Thrombosis and Hemostasis Medical Publishers, Cilt 45, No 7, s. 661-673.*
- [25] Clark, Michael J. (2011). “Performance comparison of exome DNA sequencing Technologies”, *Nat Biotechnology, Cilt 29, No 10, s. 908-914.*
- [26] Korostin, Dmitriy et al. (2020). “Comparative analysis of novel MGISEQ-2000 sequencing platform vs Illumina HiSeq 2500 for whole-genome sequencing”, *PLoS One, Cilt 15, No 3*
- [27] Zhang, Guoqiang et al. (2015). “Comparison and evaluation of two exome capture kits and sequencing platforms for variant calling”, *BMC Genomics, Cilt 16, No 5, s. 581-589.*
- [28] Lang, Jidong et al. (2021). “Evaluation of the MGISEQ-2000 Sequencing Platform for Illumina Target Capture Sequencing Libraries”, *Frontiers in Genetics, Cilt 12, No 2.*

TEZDEN ÜRETİLMİŞ YAYINLAR

Konferans Bildirileri

1. C.A.Keçeci, M.H.Müslümanoğlu (2023) “*TP53* ve *MEFV* Genlerinin ILLUMINA ve MGI Teknolojileri ile Dizilenmesinde Data Kalitelerinin Karşılaştırılması”, 10. Uluslararası GAP Zirvesi Bilimsel Araştırmalar Kongresi, Sözel Özet Sunumu, (s. 171)

