



T.C.

BİRÜNİ ÜNİVERSİTESİ

LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

MOLEKÜLER VE TIBBİ GENETİK ANABİLİM DALI

DOKTORA PROGRAMI

KİSTİK FİBROZİS TEŞHİS VE TEDAVİSİNDE
KULLANILACAK RNA SEKANSINDA
İLERİ VERİ ANALİZLERİ

Selvi ERGİN, MSc., MBA

Danışman

Doç. Dr. Meryem ALAGÖZ

Eş Danışman

Prof. Dr. Osman Uğur SEZERMAN

İstanbul, Nisan, 2022

TEZ ONAYI



BEYAN

Bu tezin bana ait olduğunu, tüm aşamalarında etik dışı davranışımın olmadığını, içinde yer alan bütün bilgileri akademik ve etik kurallar içinde elde ettiğimi, kullanmış olduğum bütün bilgilere kaynak gösterdiğimi ve bu kaynakları da kaynaklar listesine aldığımı, yine bu tezin yürütülmesi ve yazımı sırasında patent ve telif haklarını ihlal edici bir davranışımın olmadığını beyan ederim.

Selvi ERGİN



TEŞEKKÜR

Her şeyden önce danışmanım Doç. Dr. Meryem Alagöz'e tavsiyeleri, sürekli desteği ve doktora çalışmam sırasında gösterdiği sabır için sonsuz teşekkürlerimi sunmak isterim.

Akademik araştırmalarımın her döneminde beni cesaretlendiren engin bilgi birikimi ve deneyimlerinden dolayı eş danışmanım Prof. Dr. Osman Uğur Sezerman'a teşekkür ederim.

Ayrıca eğitimim süresince bilgi ve tecrübelerini benden esirgemeyen değerli bölüm hocalarıma,

Çalışmamdaki teknik destekleri için Mehmet Akif Bıçakcı, Emine Ravza Gür, Berk Gürdamar ve Acıbadem Üniversitesi Biyoistatistik ve Tıp Bilişimi AD'na,

Doktora tez çalışmam boyunca her zaman yanımda olan canım anneme ve babama teşekkürlerimi sunarım.

İÇİNDEKİLER

İçindekiler	Sayfa No
İÇ KAPAK.....	-
TEZ ONAYI.....	-
BEYAN.....	iii
TEŞEKKÜR	iv
İÇİNDEKİLER	v
SEMBOL VE KISALTMALAR LİSTESİ	vii
TABLO LİSTESİ	ix
ŞEKİL LİSTESİ.....	xi
TÜRKÇE ÖZET	xiv
İNGİLİZCE ÖZET	xv
1. Giriş ve Amaç	1
2. Genel Bilgiler	4
2.1. Mendel Kalıtım	4
2.2. Nadir Hastalıklar	5
2.2.1. Kistik Fibrozis.....	6
2.2.1.1 Tarihsel Bakış	6
2.2.1.2 Klinik Sunum ve Tanı	7
2.2.1.3 KFTR Geni ve Proteini	8
2.2.1.4 KFTR Disfonksiyonu ve Hastalığa Neden Olan Mutasyonlar.....	11
2.2.1.5 Hücredeki KFTR.....	15
2.2.1.6 KFTR'nin Fonksiyonları.....	20
2.2.1.7 Temel Defektin Düzeltilmesine Yönelik Tedaviler	21
2.3. Genetik Varyasyonlar.....	25

2.4. Genetik Tanı Testleri	29
2.4.1. Yeni Nesil Dizileme	29
2.4.1.1. Hedef Sekanslama: Targeted Sekanslama-TS:	33
2.4.1.2. Tüm Ekzom Sekanslama-TES	33
2.4.1.3. Tüm Genom Sekanslama-TGS	34
2.5. RNA Sekanslama	35
2.6. Yeni Nesil Dizileme Data Analizi.....	47
2.6.1. RNASeq Data Analizi	52
3. Materyal ve Metot	60
3.1. Veriler	60
3.1.1 Veri Kaynağı 1	60
3.1.2. Veri Kaynağı 2	63
3.2. Analiz Hattı	66
3.3 Analiz İçin Uygun Ortam Oluşturma.....	67
3.3.1. Kalite Kontrol: FastQC	67
3.3.2. FastQC → Varyant Çağırma.....	69
3.3.2.1 Varyantların Değerlendirilmesi.....	71
3.3.3 Transkriptom Analizi	71
4. Sonuç	73
4.1. Veri kaynağı 1	73
4.2. Veri Kaynağı 2	112
4.3. Transkriptom Analiz sonuçları.....	130
5. Tartışma	140
6. Öneriler	165
7. Kaynaklar	167
8. Özgeçmiş	213
9. İntihal Raporu	214

SEMBOL VE KISALTMALAR LİSTESİ

2D	İki Dimension (İki Boyutlu)
3D	Üç Dimension (Üç Boyutlu)
AA	Amino Asit
A-CGH	Array Comparative Genomic Hybridisation (Dizinle Karşılaştırmalı Genomik Hibridizasyon)
AID	Activation-Induced Cytidine Deaminase (Aktivasyonla İndüklenen Sitidin Deaminaz)
ATP	Adenozin Trifosfat
BAM	Binary Alignment/Map (İkili Hizalama/Harita)
BCL	Base Call
bç	baz çifti
CAT	Connective Airway Tissue (Bağ Airway Doku)
cDNA	komplementer Deoksiribonükleik Asit
CNV	Copy Number Variation (Kopya Sayısı Değişimi)
CO ₂	Karbon Dioksit
CPM	Count Per Million (Milyon Başına Sayım)
DHS	DNase I Hypersensitive Site (DNase I Aşırı Duyarlı Bölge)
DMD	Duchenne Musküler Distrofi
DNA	Deoksiribonükleik Asit
dNTP	Deoksinükleotid Trifosfat
ECL4	Fourth Extracellular Loop (Dördüncü Ekstrasellüler Döngü)
ECM	Ekstrasellüler Matriks
FDR	False Discovery Rate (Yanlış Keşif Oranı)
GLCM	Gray-Level Co-Occurrence (Gri Düzey Eş Oluşum Matrisi)
gRNA	guide Ribonükleik Asit
Kb	Kilo baz
KD	Kilo Dalton

KF	Kistik Fibrozis
KFTR	Kistik Fibroz Transmembran Regulator
Mb	Mega baz
MEM	Minimum Essential Medium (Minimum Temel Ortam)
MLPA	Multiplex Ligation-dependent Probe Amplification (Çoğul Ligasyon Bağımlı Arttırımı)
MW	Molecular Weight (Moleküler Ağırlık)
NGS	Next Generation Sequencing (Yeni Nesil Dizileme)
NH	Nadir Hastalıklar
OD	Otozomal Dominant
OMIM	Online Mendelian Inheritance in Man
OR	Otozomal Resesif
PCR	Polimeraz Zincir Reaksiyonu
qRT-PCR	Real-Time Quantitative Reverse Transcription PCR (Gerçek Zamanlı Kantitatif Ters Transkripsiyon PCR)
RNA	Ribonükleik Asit
RNASeq	RNA Sekanslama
SAM	Sequence Alignment/Map (Dizi Hizalama/Harita)
SHG	Second Harmonic Generation (İkinci Harmonik Nesil)
SMS	Single Molecule Sequencing (Tek Molekül Dizilimi)
SNP	Single Nucleotide Polymorphism (Tek Nükleotit Polimorfizmi)
SNV	Single Nucleotide Variant (Tek Nükleotid Varyant)
SRP	Signal Recognition Particle (Sinyal Tanıma Partikülü)
ssDNA	single stranded DNA (Tek Zincirli DNA)
SV	Structural Variant (Yapısal Varyant)
UTR	Untranslated Region (Kodlanmayan Bölge)
VCF	Variant Call Format (Varyant Çağırma Formatı)
VST	Variance Stabilizing Transformation (Varyans Sabitleyici Dönüşüm)
VUS	Variant of Uncertain Significance (Önemi Bilinmeyen Varyant)
WES	Whole Exome Sequencing (Tüm Ekzom Sekanslama)
WGS	Whole Genome Sequencing (Tüm Genom Sekanslama)

TABLO LİSTESİ

Tablo 2.1 Mutasyon türüne göre istatistikler	12
Tablo 2.2 FDA onaylı KFTR modülatörleri ve KF bireylerde kullanım endikasyonları	22
Tablo 2.3 Mevcut klinik kullanımda olanlar dışındaki KFTR modülatörleri	23
Tablo 2.4 Genetik varyantların isimlendirilmesi için kullanılan kısaltmalar.....	28
Tablo 3.1 Orta ve Ağır KF Hastaları ayrıntılı klinik özellikleri	63
Tablo 3.2 qRT-PCR validasyonu için kullanılan primerler	65
Tablo 3.3 Biyoinformatik analizde kullanılan araçlar ve ilgili web kaynakları.....	67
Tablo 3.4 GATK VCF sonuçlarına ait örnek tablo	70
Tablo 3.5 Samtools VCF sonuçlarına ait örnek bir tablo.....	71
Tablo 4.1 Veri Kaynağı 1 örnek listesi ve özellikleri	73
Tablo 4.2 Korelasyon uzunluğu analizi	76
Tablo 4.3 3D veri setinde up regüle olmuş genler	78
Tablo 4.4 2D ve 3D örneklerinde varyant çağırma toplu sonuçları.....	80
Tablo 4.5 2D ve 3D örneklerinde varyant sayıları ve paket karşılaştırılması.....	82
Tablo 4.6 2D ve 3D örneklerinde sadece patojenlerin listesi.....	83
Tablo 4.7 Varyans dispersion etkisi	90
Tablo 4.8 2D örneklerinde Padj değerine göre sıralanmış ilk 20 up ve down regüle DEG	94
Tablo 4.9 2D örneklerinde log2FoldChange değerine göre sıralanmış en iyi 20 up regüle DEG	95
Tablo 4.10 2D örneklerinde log2FoldChange değerine göre sıralanmış en iyi 20 down regüle DEG	96
Tablo 4.11 2D örneklerinde log2FoldChange değerine göre en az down regüle olan genden itibaren tüm genler.....	97
Tablo 4.12 2D örneklerinde log2FoldChange değerine göre en az up regüle olan genden itibaren tüm genler.....	98
Tablo 4.13 3D örneklerinde Padj değerine göre sıralanmış ilk 20 up ve down regüle DEG	104

Tablo 4.14 3D örneklerinde log2FolChange değerine göre sıralanmış en iyi 20 up regüle DEG	105
Tablo 4.15 3D örneklerinde log2FoldChange değerine göre sıralanmış en iyi 20 down regüle DEG	106
Tablo 4.16 3D örneklerinde log2FoldChange değerine göre en az down regüle olan genden itibaren tüm genler.....	107
Tablo 4.17 3D örneklerinde log2FoldChange değerine göre en az down regüle olan genden tüm genler.....	108
Tablo 4.18 2D ve 3D örneklerinde Padj değerine göre sıralanmış ilk 20 up ve down regüle DEG karşılaştırılması.....	110
Tablo 4.19 : 2D ve 3D örneklerinde log2FoldChange değerine göre sıralanmış ilk 20 down regüle DEG karşılaştırılması	110
Tablo 4.20 Veri Kaynağı 2 örnek listesi ve özellikleri	112
Tablo 4.21 Orta ve Ağır KF örneklerinde varyant çağırma toplu sonuçları	120
Tablo 4.22 Orta ve Ağır KF örneklerinde varyant sayıları ve paket karşılaştırılması	128
Tablo 4.23 Orta ve Ağır KF örneklerinde sadece patojenlerin listesi.....	128
Tablo 4.24 Orta ve Ağır KF örneklerinin Padj değerine göre sıralanmış ilk 20 up ve down regüle DEG.....	135
Tablo 4.25 Orta ve Ağır KF örneklerinde log2FoldChange değerine göre sıralanmış en iyi 20 up regüle DEG	136
Tablo 4.26 Orta ve Ağır KF örneklerinde örneklerinde log2FoldChange değerine göre sıralanmış en iyi 20 down regüle DEG	137

ŞEKİL LİSTESİ

Şekil 2.1 Otozomal Resesif Kalıtım.....	5
Şekil 2.2 Genden proteine KFTR geni ve proteini.....	9
Şekil 2.3 KFTR proteini ve Domainleri.....	11
Şekil 2.4 Gen mutasyonlarının sınıfları ve ilgili terapötik stratejiler.....	14
Şekil 2.5 KFTR PQQ etkileşim Yolağı.....	19
Şekil 2.6 Çeşitli KFTR mutasyon sınıfları olan KF bireylerde KFTR fonksiyonunu geri yüklemek için farmakolojik stratejiler	22
Şekil 2.7 Genetik Varyasyon Türleri	26
Şekil 2.8 Varyant oluşum mekanizmaları	27
Şekil 2.9 Klinik tanı amaçlı YND teknolojisinin uygulamaları.....	35
Şekil 2.10 Standart bir RNASeq temel adımları	37
Şekil 2. 11 Tanı amaçlı RNASeq analizinin şematik gösterimi.....	38
Şekil 2.12 RNAseq kütüphane hazırlığı.....	41
Şekil 2.13 Ekzom, genom tabanlı entegre genomik ve transkriptomik yaklaşımların uygulanması üzerine tanısal verim.....	43
Şekil 2.14 Okuma hizalama algoritmasına genel bakış	48
Şekil 2.15 Bir gen boyunca derinlik ve kapsam için TGS ve TES'in karşılaştırılması	49
Şekil 2.16 Genomik projelerde haritalama, hizalama, yığın ve	50
Şekil 2.17 ENCODE projesinden dört özofagus skuamöz mukoza ve dört özofagus muskularis epitel toplam RNASeq örneğini gösteren PCA	57
Şekil 2.18 Tipik bir RNASeq pipeline örneği.....	59
Şekil 3.1 Verilerin yeniden değerlendirilmesi için analiz hattı.....	66
Şekil 3.2 FastQC tool kullanılarak oluşturulan N.HLF_2D_Rep1'a ait örnek rapor görüntüsü.....	68
Şekil 3.3 TRIMMING işlemi sonrasında Örnek rapor görüntüsü	69

Şekil 4.1 Gri renkte fibroblastlar tarafından üretilen endojen kolajen ve mavi renkte hücre çekirdekleri.....	74
Şekil 4.2 KF bağ hava yolu dokusunu (CAT) geliştirmek için biyo-mühendislik süreci	74
Şekil 4.3 Hem NA- μ TP'ler hem de CF- μ TP'ler zamanla boyutlarını artıran görüntü	75
Şekil 4.4 Hücre kültür sonuçlarının toplu görüntüsü	75
Şekil 4.5 μ TPs morfolojik karakterizasyonu.....	77
Şekil 4.6 2D ve 3D veri kümeleri arasında karşılaştırma.....	79
Şekil 4.7 2D örneklerinde eşlenen okuma sayısı	84
Şekil 4.8 EDASeq ile 2D örneklerinde gen seviyeleri gösterimi.....	85
Şekil 4.9 EDASeq ile 2D örneklerinde GC içeriğinde gen seviyesi	85
Şekil 4.10 DESeq ile 2D örneğinde ham count sonucu	86
Şekil 4.11 2D KF örneklerinde şeritler arasındaki örnek varyansı	87
Şekil 4.12 2D KF ve Normal tekrar örneklerinin kümelenmesi	89
Şekil 4.13 2D KF ve Normal örnekleri PCA grafiği.....	90
Şekil 4.14 2D örneklerinde Dispersion grafiği	91
Şekil 4.15 2D örneklerinde test edilen tüm genler için log2 kat değişikliklerine karşı normalize olan count verilerinin ortalaması.....	93
Şekil 4.16 2D örneklerinde total okumalara istinaden LCF genel sonuç.....	94
Şekil 4.17 2D örneklerinde up ve down regüle ilk 20 genin heatmap gösterimi	95
Şekil 4.18 2D örneklerinde en az down regüle olan 20 genin heatmap gösterimi.....	97
Şekil 4.19 2D örneklerinde volcano plot gösterimi	99
Şekil 4.20 2D örneklerinde Diferansiyel olarak ifade edilen ilk 20 gen.....	99
Şekil 4.21 EDASeq ile 3D örneklerinde eşlenen okuma sayısı	100
Şekil 4.22 EDASeq ile 3D örneklerinde gen seviyelerinin gösterimi.....	100
Şekil 4.23 EDASeq ile 3D Örneklerinde GC içeriğinde gen seviyesi	101
Şekil 4.24 DESeq ile 3D örneğine ait ham count sonucuna bir örnek.....	101
Şekil 4.25 3D KF örneklerinde şeritler arasındaki örnek varyansı	102
Şekil 4.26 3D KF ve Normal tekrar örneklerinin kümelenmesi	102
Şekil 4.27 3D KF ve Normal örnekleri PCA grafiği.....	103
Şekil 4.28 3D örneklerinde Dispersion grafiği	103
Şekil 4.29 3D örneklerinde test edilen tüm genler için log2 kat değişikliklerine karşı normalize olan count verilerinin ortalamasının gösterimi	104
Şekil 4.30 3D örneklerinde total okumalara istinaden LCF genel sonuç.....	104

Şekil 4.31 3D örneklerinde up ve down regüle ilk 20 genin heatmap gösterimi	105
Şekil 4.32 3D örneklerinde en az down regüle olan 20 genin heatmap gösterimi...	107
Şekil 4.33 3D örneklerinde volcano plot gösterimi	109
Şekil 4.34 3D örneklerinde Diferansiyel olarak ifade edilen ilk 20 gen.....	109
Şekil 4.35 2D ve 3D örneklerinde down regüle genlere ait Venn-Şeması.....	111
Şekil 4.36 2D ve 3D örneklerinde up regüle genlere ait Venn-Şeması	111
Şekil 4.37 Veri kaynağı 2 PCA analizi	113
Şekil 4.38 Veri kaynağı 2 Diferansiyel olarak ifade edilen genlerde heatmap.....	114
Şekil 4.39 Veri kaynağı 2 RNASeq volcano plot	114
Şekil 4.40 Veri kaynağı 2 RNASeq sonuçlarının qPCR validasyonu.....	115
Şekil 4.41 Veri kaynağı 2 RNASeq ve qPCR'ın korelasyon analizi.....	116
Şekil 4.42 Veri kaynağı 2 qPCR ve RNASeq ile IL-8 ekspresyonu.....	117
Şekil 4.43 Veri kaynağı 2 IL-8 ekspresyonu.....	117
Şekil 4.44 Veri Kaynağı 2 RNASeq fonksiyonel zenginleştirme analizi	118
Şekil 4.45 Veri kaynağı 2 RNASeq ISMARA analizi	119
Şekil 4.46 Veri kaynağı 2 RNASeq ağ analizi.....	120
Şekil 4.47 Orta ve Ağır KF örneklerinde eşlenen okuma sayısı.....	130
Şekil 4.48 Orta ve Ağır KF örneklerinde gen seviyeleri.....	131
Şekil 4.49 Orta ve Ağır KF örneklerinde GC içeriğinde gen seviyesi.....	131
Şekil 4.50 Ağır KF örneklerinde örneklerde şeritler arasındaki örnek varyansı ..	132
Şekil 4.51 Orta ve Ağır KF örneklerinde korelasyon grafiği.....	133
Şekil 4.52 Orta ve Ağır KF örneklerinde PCA grafiği.....	133
Şekil 4.53 Orta ve Ağır KF örneklerinde Dispersion grafiği.....	134
Şekil 4.54 Orta ve Ağır KF örneklerinde test edilen tüm genler için log2 kat değişikliklerine karşı normalize olan count verilerinin ortalamasını gösterimi.....	134
Şekil 4.55 Orta ve Ağır KF örneklerinde total okumalara istinaden LCF genel sonuç	135
Şekil 4.56 Orta ve Ağır KF örneklerinde up ve down regüle ilk 20 genin heatmap gösterimi.....	136
Şekil 4.57 Orta ve Ağır KF örneklerinde en az down regüle olan 20 genin heatmap gösterimi.....	138
Şekil 4.58 Orta ve Ağır KF örneklerinde volcano plot gösterimi	138
Şekil 4.59 Orta ve Ağır KF örneklerinde Diferansiyel olarak ifade edilen ilk 20 gen	139

TÜRKÇE ÖZET

ERGİN, S. (2022). KİSTİK FİBROZİS TEŞHİS VE TEDAVİSİNDE KULLANILACAK RNA SEKANSINDA İLERİ VERİ ANALİZLERİ. Doktora Tezi, Biruni Üniversitesi Lisansüstü Eğitim Enstitüsü, İstanbul.

KF çok çeşitli klinik ve genetik varyantları olan, yaşamı sınırlayan, OR geçişli bir genetik hastalıktır. Epitelyal hücre zarı boyunca anormal klorür ve sodyum taşınmasına katkıda bulunan kromozom 7q'da'ki KF transmembran iletkenlik düzenleyicisindeki gen mutasyonlarından kaynaklanmaktadır. YND platformu tabanlı RNASeq tekniklerinin geliştirilmesi sonucunda son on yılda büyük ilerlemelerin kaydedilmesiyle bu teknikler sayesinde birçok biyolojik alanda araştırmalar hızlanmıştır. Bu çalışmada, RNASeq transkriptomik analizi yapılan, 2 ayrı veri kaynağından elde edilen toplamda 44 KF örneklerinde kendi belirlediğimiz pipeline ile birlikte RNASeq datasında mutasyon taraması ve ayrıntılı bir transkriptom yanıtın varlığının ortaya çıkarılması amaçlanmıştır. Birinci veri kaynağımızda epitelyal disfonksiyonu incelemek için birkaç *in vitro* KF modeli geliştirilerek stromal değişikliklere odaklanılarak morfolojik ve transkriptomik özellikler araştırılmıştır. İkinci veri kaynağımızda 'modifiye edici genlerin' olası etkisi nedeniyle, benzer KFTR mutasyonlarına sahip hastalar arasında bile hastalık şiddeti değişken olduğu için bu genleri belirlemek için, orta ve ağır akciğer fenotipi olan KF hastalarında analizler uygulanmıştır. Çalışmamızda kullandığımız 2 veri kaynağından varyant arama için kullandığımız GATK aracı ile daha fazla varyant elde ettik. Bu da literatür ile uyumludur. Her iki veri kaynağımızda yaptığımız varyant çağırma sonrasında, KF ile RNASeq datalarda anlamlı bir sonuç bulunamamıştır. Transkriptom analizimizde ise ilk veri kaynağında toplamda 426 transkript 2D, 352 transkript de 3D durumunda, 2. veri kaynağımızda da 140 transkriptin yeniden düzenlendiği gösterilmiştir. İlk veri kaynağıyla uyumlu olarak WNT2 ve TBX4 genlerinin up regüle olduğu bulunmuştur. Diğer veri kaynağıyla uyumlu olarak çalışmada EGR1 ve CXCL8 genlerinin down regüle olduğu saptanmıştır. Literatür ile uyumlu olarak ZNF683 geninin ağır KF örneklerinde aşırı ekspresyona uğradığını gösterdik.

Anahtar kelimeler: Kistik fibrozis, RNASeq, Transkriptomik

İNGİLİZCE ÖZET

ERGİN, S. (2022). ADVANCED ANALYSIS OF DATA ON RNA SEQUENCE TO BE USED IN THE DIAGNOSIS AND TREATMENT OF CYSTIC FIBROSIS. PhD Thesis, Biruni University Graduate Education Institute, Istanbul.

CF is an AR genetic illness with a wide range of clinical and genetic variables. It is caused by mutations in the CF transmembrane conductivity regulator gene on chromosome 7q, which causes abnormal chloride and salt transport across the epithelial cell membrane. Thanks to NGS platform-based RNASeq techniques, great progress has been made in the last ten years, and research in many biological fields have accelerated. In this study, it aimed to reveal the presence of a detailed transcriptome response and mutation screening in the RNASeq data. We determined the pipeline for 44 CF samples obtained from two different data sources. Several *in vitro* CF models were development to investigate epithelial dysfunction, focusing on stromal alterations and analyzing morphological and transcriptome aspects in the first data source. Due to the potential impact of 'modifying genes' in the second data source, studies were carried out in CF patients with moderate and severe lung phenotypes to identify these genes, as disease severity varies even among patients with similar CFTR mutations. From the two data sources we used in our study, we obtained more variants with the GATK tool we used for variant search. This is compatible with the literature. No meaningful results were found in CF by RNASeq datas after variant calling in both of our data sources. In our transcriptome analysis, in the first data source 426 transcripts were rearranged in 2D, 352 transcripts in 3D and in our second data source 140 transcripts were rearranged. Consistent with the first data source, the WNT2 and TBX4 were found up regulated and in the other source, EGR1 and CXCL8 down regulated. Moreover, we showed that the ZNF683 gene was overexpressed in severe CF.

Keywords: Cystic fibrosis, RNA Seq, Transcriptomics

1. Giriş ve Amaç

Kistik fibrozis (KF), beyaz popülasyonda yaygın olarak görülen ve yıllık insidansı yaklaşık 3.500 canlı doğumda 1 olan otozomal resesif (OR) bir hastalıktır. Bu çoklu sistem hastalığı, kromozom 7 üzerindeki KF transmembran regülatör (KFTR) genindeki, transmembran klorür reabsorpsiyonunun düzenlenmesi için gerekli bir proteini şifreleyen genetik mutasyonlar ile karakterize edilmektedir. Bu aktivite esas olarak KF ve karaciğer hastalığı olanların tanısına, değerlendirilmesine ve yönetimine odaklanır.

Yeni Nesil Dizileme (YND) dâhil olmak üzere şu anda farklı moleküler testler mevcuttur (Vigliar et al., 2015a). YND teknolojilerinden biri olan "sentez yoluyla sekanslama" yaklaşımıyla, farklı örneklerde farklı biyobelirteçlerin aynı anda analizi sağlanmaktadır (Vigliar et al., 2015b). Genel kural olarak, YND platformları üç sekanslama yaklaşımından sentez, hibridizasyon ve ligasyon işlemlerini benimseyebilmektedirler (Reuter ve ark., 2015). Kısaca, farklı ticari platformlar mevcut olmasına rağmen, bir YND'ya ait iş akışı sırasıyla 1) Kütüphane hazırlığı 2) Klonal olarak fragmanların amplifikasyonu 3) Masif paralel dizileme 4) Veri analizi olmak üzere aynı adımlar takip edilmektedir (Vigliar et al., 2015a).

DNA temelli yaklaşım, nokta mutasyonlarının, insersiyonların ve delesyonların tespiti için kullanılmaktadır. Ancak bu yaklaşım yeniden gen düzenlemelerinde yer alan intronik bölgelerin varlığından dolayı gen füzyonları düşünüldüğünde sınırlı kalabilmektedir (De Maglio et al., 2022). Bu sınırlama da bir RNA sekanslama (RNASeq) yaklaşımıyla aşılmaktadır. Diğer taraftan da RNA'nın DNA'dan daha az kararlı olması nedeniyle analitik analiz öncesi sorunların ortaya çıkma ihtimali bulunmaktadır (Bruno and Fontanini, 2020).

Tüm canlılık mekanizması, organizmanın özelliklerini tanımlayıp hücrelerin ve organizmanın biyolojik işlevlerini bir bütün olarak sürdürmektedir. Bu nedenle, genomik süreçleri ve hastalıkların kökenini anlamak için RNA analizi esastır. Toplu olarak transkriptomlar olarak bilinen RNA'lar, kodlayan ve kodlamayan bölgelere sahip karmaşık genomik yapılar olup genler ve proteinler arasında aracılık görevleri bulunmaktadır. Bu nedenle, transkriptom üzerinde ayrıntılı çalışmalar yapmak genomik işlevi anlamak ve hücrelerin moleküler bileşimlerini tanımlamak için

gereklidir. Ek olarak, transkriptom hakkında daha kapsamlı bilgi elde etmek hastalıkların gelişiminin yanı sıra nedeninin de anlaşılmasına yardımcı olmaktadır. Tüm hücre gruplarından tek hücreli RNASeq'ına kadar bu teknolojinin tanıtılması, transkriptomun işlenmesine ve haritalanmasına fırsatı vermiştir.

Laboratuvar tabanlı iş akışının ardından verinin işlenmesi ve analizi çeşitli hesaplama araçları kullanılarak yapılmaktadır. Verinin işlenmesi, referans genomları olan ve olmayan organizmalar için gerçekleştirilmektedir. Referans genomu olan organizmalarda kısa RNASeq okumaları referans genom kullanılarak haritalanmaktadır. Referans genomu olmayan organizmalar için de *de novo* transkriptom montajı uygulanmaktadır (Rosenberg et al., 2018; Engström et al., 2013).

RNASeq, gen ekspresyonu, splicing olayları, maternal veya paternal alel ekspresyonunun nicel ölçümü yoluyla transkriptomun ayrıntılı bir şekilde anlaşılmasını sağlayarak diğer moleküler yöntemlere kıyasla avantajlar sunmaktadır. Bu avantajlarla birlikte hastalıkların nedeninin daha düşük maliyetle verimli bir şekilde yorumlamasına da yardımcı olmaktadır. Örneğin, RNASeq, DNA Sekanslama (DNaseq) gibi, ifade edilen bölgelerdeki genomik varyasyonları da tespit edebilmektedir (Piskol et al., 2013). Üretilmesi ve saklanması maliyetli olan DNaseq verilerine alternatif olarak hiçbir ekzom yakalama aracının geliştirilmediği wild veya evcil çiftlik hayvan türlerinde kullanılabilmektedir (Jehl et al., 2021).

Yukarıda bahsedilen avantajlara rağmen, RNASeq henüz kodlama bölgelerinde SNP tespiti için sıklıkla kullanılmamaktadır. Aslında, RNASeq tarafından SNP tespiti ve genotip çağrısında üç ana zorluk bulunmaktadır. Öncelikle, transkriptom olgun transkriptlerden oluştuğu için ekzon-ekzon bağlantılarıyla örtüşen RNASeq okumalarının haritalanması DNaseq okuma hizalamasına kıyasla daha zordur (Pan et al., 2008). Bununla birlikte, ekzon-ekzon bağlantılarına yakın tespit edilen SNP'ler için dikkatli olunması önemli olsa da RNASeq haritalama yöntemleri son yıllarda iyi bir şekilde geliştirilmiştir (Peng et al., 2012; Lagarrigue et al., 2013b).

İkinci zorluk, RNA düzenlemesinin esas olarak DNA seviyesinde bulunmayan RNA seviyesindeki varyasyonları tespiti ile ortaya çıkmaktadır. Ancak yapılan çalışmalar, RNA düzenlemenin ayırt edici özellikleri nedeniyle, standart koşullarda güvenilir RNASeq tabanlı varyasyon saptanmasının yalnızca biraz engellediğini

göstermiştir. Üçüncü zorluk da, genlerin, genom boyunca oldukça homojen bir okuma derinliği sunan DNaseSeq'in aksine, birkaç okumadan milyonlarca okumaya kadar değişen okuma derinliklerine yol açan oldukça değişken ifade seviyeleri sergilenmesi olarak sayılmaktadır. Aslında, kodlayan ve kodlamayan transkriptler, farklı hücre tiplerinde ve gelişimsel veya fizyolojik aşamalarda, hücre başına birkaç kopyadan milyonlarca kopyaya kadar çok farklı seviyelerde ifade edilmektedirler (Sims et al., 2014).

Bu tez çalışmasında, nadir hastalıklar arasında yer alan KF örneklerinde RNASeq analizi ile hem transkriptomik yaklaşımlar değerlendirildi hem de mutasyon taraması yapıldı. Kendi hücre dışı matriksine gömülü primer fibroblastlardan oluşan yeni bir KF 3D ve 2D stromal akciğer modeli gösterilip, morfolojik ve transkriptomik özellikleri araştırılan bir çalışmaya ait RNASeq datası ve Genetik modifiye edicileri tanımlamak için, Orta ve Ağır akciğer fenotipi olan KF hastalarında transkriptomik analizler uygulanan bir diğer çalışmaya ait RNASeq dataları alınıp kendi belirlediğimiz yeni pipeline ile yeniden değerlendirdik. Bu sayede yeni tanı ve tedavi yöntemlerinin geliştirilmesini destekleyecek KF'de rolü olan yeni genler keşfedilmeye çalışılıp bunlarla ilgili yollar ve mutasyonlar araştırıldı. Bu analiz ile birlikte ayrıntılı bir transkriptom yanıtının varlığının ortaya çıkarılması ve RNASeq datasında mutasyon taramasının yapılması amaçlandı.

2. Genel Bilgiler

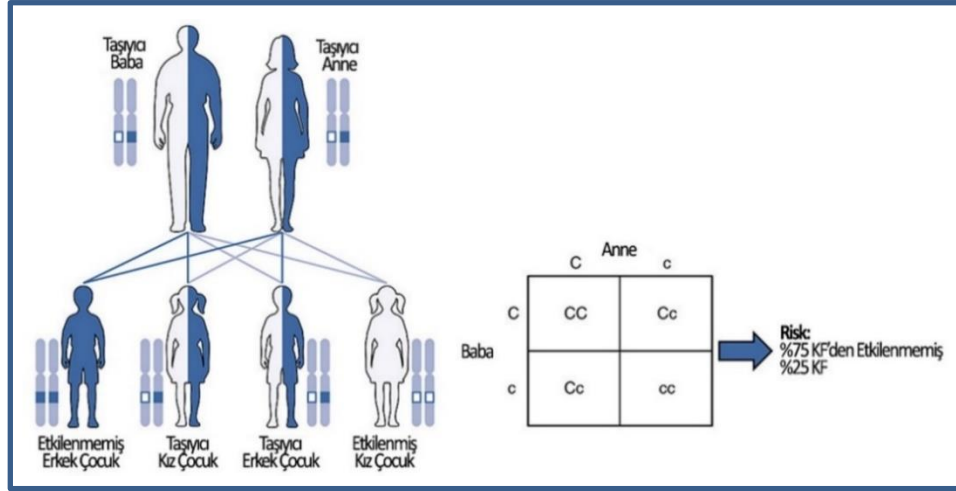
2.1. Mendel Kalıtım

Tıbbi genetik alanı geleneksel olarak kromozom anomalilerine ve Mendel hastalıklarına odaklanmaktadır (Brigatti, 2017).

Mendel grubu, gen fonksiyonu üzerinde önemli bir etki sergileyen ve durumun fenotipinden sorumlu olan tek gen mutasyonlarının neden olduğu grubu oluşturmaktadır. Bu mutasyonlar, genellikle sağlığı etkilediği ve potansiyel olarak üreme yeteneğini azalttığı için genel popülasyonda nadir olarak görülmektedir (Rahit and Tarailo-Graovac, 2020).

Bugüne kadar tanımlanan Mendel bozukluklarının %80'inden fazlasının pediatrik başlangıçlı bir fenotipi bulunmaktadır. Ancak yetişkinlikte ortaya çıkan Mendel bozuklukları da klinikte giderek daha fazla dikkat çekmeye başlamıştır. Bazen pediatrik ve erişkin başlangıçlı bir gen varyantı, alel/mutasyonun şiddeti ve bunun zaman içindeki ortaya çıkışı ile ilgili olabilmektedir (Baltimore, OMIM; Wright et al., 2018). Tıbbi genetikte, olası bir Mendel bozukluğunun kalıtım biçimini netleştirmeye yardımcı olan ana araçlardan biri, üç kuşak bir soyağacıdır. Soyağacı, çeşitli semboller kullanılarak, indeks hasta (veya proband), hem yaşayan hem de ölen yakın ve uzak akrabalar, infertilite, düşük öyküsü, evlat edinme ve etnik köken hakkında bilgileri içermektedir.

Bir soyağacının incelenmesi, bir ailedeki genetik bir bozukluğun kalıtım modelini belirlemede ilk adımı oluşturmaktadır. Genel olarak, otozomal koşullar erkekleri ve kadınları eşit olarak etkiler, ancak bazıları diğer genetik ve genetik olmayan faktörlerden dolayı cinsiyetle sınırlı olabilmektedirler. Otozomal dominant (OD) kalıtım, etkilenen bireylerin her nesilde ortaya çıktığı “dikey geçiş” olarak adlandırılmaktadır. OR kalıtımda fenotip, tipik olarak sadece probandın kardeşlerinde görülerek sadece “yatay geçiş”i göstermektedir. OR fenotip yalnızca hem anne hem de baba alellerinde mutasyonları, yani bialelik patojenik varyasyon barındıran bireylerde ifade edilmektedir (Zhu et al., 2016). Taşıyıcı ebeveynlerin çocuklarının etkilenme oranı %25'tir. Bu değer genetikte kullanılan Punnet kare modeli kullanılarak elde edilmektedir (Şekil 2.1).



Şekil 2.1 Otozomal Resesif Kalıtım

(https://commons.wikimedia.org/wiki/File:2926_Autosomal_Recessive_Inheritance-new.jpg#/media/File:2926_Autosomal_Recessive_Inheritance-new.jpg, erişim tarihi 01 Kasım 2021)

2.2. Nadir Hastalıklar

Nadir hastalıklar (NH) için şimdiye kadar çeşitli tanımlar yapılmıştır. Avrupa Birliği'nde, prevalansı $<1:2.000$ olan, yaşamı tehdit eden veya kronik olarak güçten düşüren bir hastalık olarak tanımlanmaktadır. 6.000-8.000 farklı NH tahminen 30 milyon insanı etkilemektedir. (https://ec.europa.eu/info/research-and-innovation/research-area/health-research-and-innovation/rare-diseases_en, erişim tarihi, 11 Ekim 2021). ABD'de ise bir hastalık 200.000'den az kişiyi etkilediğinde, yani 1.540 kişi başına yaklaşık 1 hastaya eşdeğer bir hastalık nadir olarak tanımlanmaktadır (Schieppati et al., 2008). Tanınmış NH'den sadece birkaçının prevalansı $1/2000$ civarındadır. Örneğin, Duchenne Musküler Distrofi (DMD) gibi çoğu çok daha nadirdir ve tahmini prevalansı erkekler için $1/20.000$ 'dir (Mah et al., 2014). Hemofilinin tahmini prevalansı ise 6.000'de 1 ile 17.000 erkekte 1 arasında değişmektedir (Stonebraker et al., 2004). Diğer hastalıklar o kadar nadirdir ki, rapor edilen prevalansı 8 milyon doğumda bir olan Progeria gibi ultra nadir olarak adlandırılmaktadırlar. Dünya çapında herhangi bir zamanda bu hastalıkla yaşayan tahminen $1/200-1:250$ kişi olduğu bilinmektedir (Pollex and Hegele, 2004; Schieppati et al., 2008)

Belirli bir NH'a sahip hasta sayısı nispeten az olsa da, toplu olarak, herhangi bir NH'a sahip insan sayısı fazladır. 5000 ile 7000 arasında NH olduğu tahmin edilmektedir. Avrupa Nadir Hastalıklar Topluluğu (EURORDIS), nadir bir hastalığı olan bireylerin oranının AB'deki nüfusun %6-8'i arasında olduğunu tahmin etmektedir (Rode, 2005; Angelis et al., 2015). NH'ların çoğu, mevcut tedavilerin eksikliği ve bu tür tedavileri keşfetmek ve geliştirmek için göreceli araştırma eksikliği ile ilişkilidir. Ayrıca, NH'lar, yüksek sosyo-ekonomik maliyetlerin yanı sıra önemli ölçüde bireysel psikolojik yük ile de ilişkilidir (Angelis et al., 2014). 2000 yılında AB, Yetim İlaçlara ilişkin Avrupa Yönetmeliğini yayınlayarak, "Nadir durumlardan etkilenen hastaların diğer hastalarla aynı kalitede tedavi görme hakkına sahip olması gerektiğini" ifade etmiştir (32000R0141). Bu yönetmeliğin amacı, o zamana kadar ilaç endüstrisi tarafından çoğunlukla ihmal edilen nadir, yetim hastalıklara yönelik ilaçların geliştirilmesini teşvik etmektir (Hall and Carlson 2014). Bu düzenleme, ABD (1983), Japonya (1993) ve Avustralya'da (1998)'daki benzer eylemleri izlemiştir. AB'de, araştırma enstitüleri veya ilaç şirketleri, genellikle pazarlama onayına ve klinik geliştirmeye yönelik ilk adım olan yetim atama için başvurabilmektedirler. Yetim atamayla gelen faydalar arasında düzenleyici ücretlerin düşürülmesi, ek düzenleyici rehberlik ve ürünün pazar onayından sonra AB'de on yıllık bir pazarlama süresi yer almaktadır (Heemstra et al., 2009). 2000 yılından bu yana pazara giren yetim ilaç sayısı büyük oranda artmıştır. Ağustos 2017'ye kadar AB, NH'lar için 137 yetim ilaca pazar izni vermiştir.

2.2.1. Kistik Fibrozis

KF, KFTR genindeki mutasyonların neden olduğu OR bir hastalıktır (Valampampil and Gupte, 2021)

2.2.1.1 Tarihsel Bakış

KF semptomlarına ait ilk tıbbi raporlarda mekonyum ileus tanımlarıyla 17.yüzyıla kadar uzansa da, daha önce *Alnından öpüldüğünde tadı tuzlu olan çocuk büyülüydü ve yakında ölmesi gerekirdi* referansıyla ortaçağa dayanmaktadır (Farinha, 2017). 1938'de Dr. Dorothy Andersen, hastalığı ilk kez yetersiz beslenme ile ilişkili olan "pankreasın kistik fibrozu" olarak tanımlamıştır (Cooney et al., 2018).

80'li yıllarda ter kanalı hücrelerinde klorür (Cl⁻) geçirgenliği üzerine yapılan ilgili çalışma sonucunda KF epitel hücrelerinin apikal membranındaki majör defektin

azalmış Cl⁻ ve artan sodyum (Na⁺) taşınımı ile ilgili olduğu tanımlanmıştır (Knowles et al., 1983; Quinton and Bijman 1983; Boucher et al., 1986). Daha sonra, 1989 yılında farklı laboratuvarların ortak çalışmaları sonucunda KF'den sorumlu genin klonlanması ve amino asit (AA) dizisinden elde edilen yapıya dayalı olarak KFTR olarak adlandırılan proteinin tanımlanması sağlanmıştır (Riordan et al., 1989; Rommens et al., 1989; Kerem ve ark. 1989). Konumsal klonlama ile tanımlanan diğer genlerde olanlara benzer şekilde, KF'den sorumlu genin bu olduğunun kanıtı, hastalardaki mutasyonların tanımlanmasıyla gerçekleşmiştir. Bu bireyler ve etkilenmemiş olanlar arasındaki fenotipik farklılık hesaba katılmıştır. Aynı zamanda, KF hastalarındaki alellerin yaklaşık %70'inde bir mutasyon tanımlanmıştır. Bu ekzon 10'da (NBD1) Phe-508'in delesyonu ile sonuçlanan 3bç'lik bir delesyon olarak bulunmuştur. Bu mutasyona DF508 (Riordan et al., 1989) adı verilmiştir ve daha sonra Antonarakis (1998) tarafından önerilen kılavuzlara göre F508del olarak değiştirilmiştir. Bunun KF'den sorumlu gen olduğuna dair son kanıt, KF hastalarından elde edilen hücrelerle yapılan deneylerden ve KFTR cDNA ile transfeksiyon üzerine Cl⁻ transportunun restorasyonundan elde edilmiştir (Drumm et al., 1990; Rich et al., 1990).

2.2.1.2 Klinik Sunum ve Tanı

Önceleri “pankreasın kistik fibrozu” olarak bilinen bu durum, giderek daha fazla basitçe “kistik fibrozis” olarak adlandırılmaya başlanmış. Belirtiler sadece pankreasın ekzokrin fonksiyonunun bozulmasıyla değil, aynı zamanda morbidite ve mortalitenin ana nedeni olan ilerleyici akciğer hastalığına da bağlıdır. Ek komplikasyonlar arasında bağırsak bezleri (mekonyum ileus), Biliyer siroz ve ter bezleri bulunmaktadır. Kısırlık hemen hemen her zaman erkeklerde görülmesine rağmen kadınlarda da görülme ihtimalinin göz önünde bulundurulması gerekmektedir (Zielenski and Tsui, 1995).

Daha spesifik tanımlamak gerekirse, KF (CF, MIM #219700), solunum sıkıntısı, kronik pankreatit ve Vaz deferenslerin bilateral konjenital yokluğuna (CBAVD, MIM #22180) bağlı erkek kısırlığı gibi çeşitli klinik fenotiplerle ilişkili KFTR genindeki (veya ABCG2 MIM #602421) mutasyonlardan kaynaklanan, epitel hücrelerinde Cl⁻ iletkenliğinin otozomal genetik bir bozukluğudur (Rosenstein and Cutting, 1998; Castellani et al., 2008). KF, ortalama 1/2500 canlı doğum prevalansla, 1/25 taşıyıcı frekansı ile monojenik bir hastalıktır (Castellani et al., 2009). KFTR geni,

kromozom 7q31.3 üzerinde yaklaşık 240kb'lik bir bölgeye yayılarak 27 ekzondan oluşmaktadır (Zielenski et al., 1991). Gen, 6.5kb'lik olgun mRNA'ya transkibe olmaktadır (5t678). Kodlanmış protein, 1480 AA içerir ve 168KD'lik hesaplanmış bir moleküler kütleyle sahiptir. KFTR proteini, entegre bir membran glikoproteini oluşturmaktadır ve ABC (ATP-Binding Cassette) protein ailesi üyelerinden birisidir (Zielenski et al., 1991; Rosenstein and Cutting, 1998; Welsh et al., 2001; Castellani et al., 2008; Castellani et al., 2009).

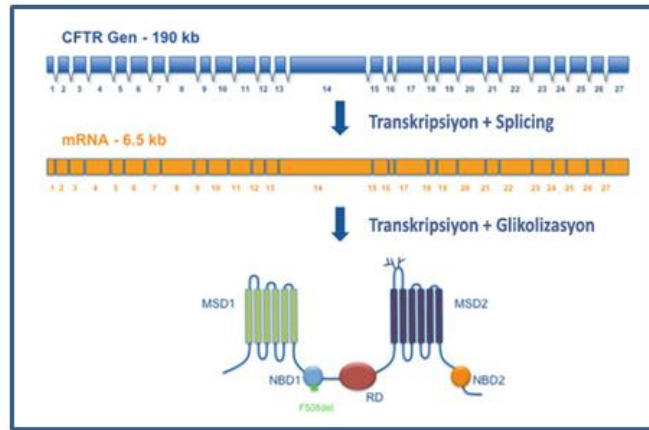
Büyük klinik değişkenlik ve semptomların çeşitliliği, bebeklerde KF tanısını karmaşık bir konu haline getirir. Önceden etkilenmiş bir çocuğu veya akrabası olan çiftlere sunulan, gönüllü gebelik sonlandırmayı takip eden Prenetal tanı, toplumda KF'yi kontrol etmek için alternatif bir stratejidir (Ryley et al. 1992). DNA'nın moleküler analizi, genin 1989'da klonlanmasından birkaç yıl önce başlamıştır. Bir fetüsün KF'den etkilenip etkilenmeyeceğini belirlemek için, daha sonra KFTR geni ile güçlü bağlantı dengesizliği içinde olduğu gösterilen birkaç belirteç kullanılmıştır. Genin 1989'da klonlanmasından bu yana (Kerem ve ark. 1989; Riordan et al. 1989; Rommens et al., 1989), bağlantı analizi, PCR bazlı DNA amplifikasyonunu takiben doğrudan mutasyon tespiti ile neredeyse tamamen değiştirilmiştir. Mutasyon tespiti için çeşitli teknikler, özellikle karmaşık klinik vakalarda, genetik danışmanlık, prenatal tanı ve taşıyıcı taramasında çok faydalıdır. KFTR geninde tanımlanan mutasyonların sayısının artmasının nedeniyle bu analiz genellikle her laboratuvarında bölgedeki en sık mutasyonlarla sınırlı kalmaktadır.

2.2.1.3 KFTR Geni ve Proteini

Gen: KF ile ilgili yıllardır fizyologlar ve biyokimyacılar çalışmalar yürütmektedirler. KF'de epitel hücrelerinin zarındaki Cl⁻ akışında sorun olduğu gösterilmiştir (Quinton, 1990). Ancak bu patolojide kusurlu protein ürününün doğrudan tanımlanması için bu yeterli olmamıştır. KF lokusu 1986'da keşfedilerek tanı amacıyla RFLP'lerin kullanımına yol açmıştır (Ferrie et al., 1992). Tanımlanan polimorfik belirteçlerden XV-2c ve Km-19'un KF geni ile güçlü bir bağlantı dengesizliği içinde olduğu bulunmuştur. Toronto'da Lap-Chee Tsui, Ann Arbor'da Francis Collins ve Londra'da Robert Williamson tarafından yönetilen üç laboratuvarın ortak çabalarının sonucunda 1989'da KF geni klonlanmıştır. Bu da doğru DNA belirteçlerinin bulunmasıyla başlayan, lokusun tespit edilmesi ve genin kendisinin

izolasyonu ile son bulan 5 yıllık yoğun bir çalışma ile sonuçlanmıştır (Kerem ve ark. 1989; Riordan et al.,1989; Rommens et al. 1989).

KF geni oldukça uzundur ve yaklaşık 189Kb'lik (Ellsworth et al., 2000) bir genomik bölgeyi kapsamaktadır (Şekil 4.3). İlk olarak 1'den 24'e kadar numaralandırılarak genin 24 ekzona sahip olduğu açıklanmıştır. Kısa bir süre sonra bu ekzonlardan üçünün (6, 14 ve 17) küçük intronlar içerdiği bulunmuştur. Bu nedenle toplam ekzon sayısı 27'dir. Bir süre için, ekzonlar 6a ve 6b, 14a ve 14b, 17a ve 17b dâhil olmak üzere 1'den 24'e kadar numaralandırılmıştı ve en son 1-27 olarak numaralandırması kabul edildi. Ekzonların boyutu, 723bç'lik ekzon13 hariç, 50-250bç arasında değişmektedir. En büyüğü yaklaşık 28Kb olmak üzere intronların boyutu da değişkendir (Ellsworth et al., 2000). KF mRNA 6.2Kb'ye sahiptir ve 1480 AA kalıntısı ile KFTR proteinini kodlar (Riordan et al., 1989) (Şekil 2.2).



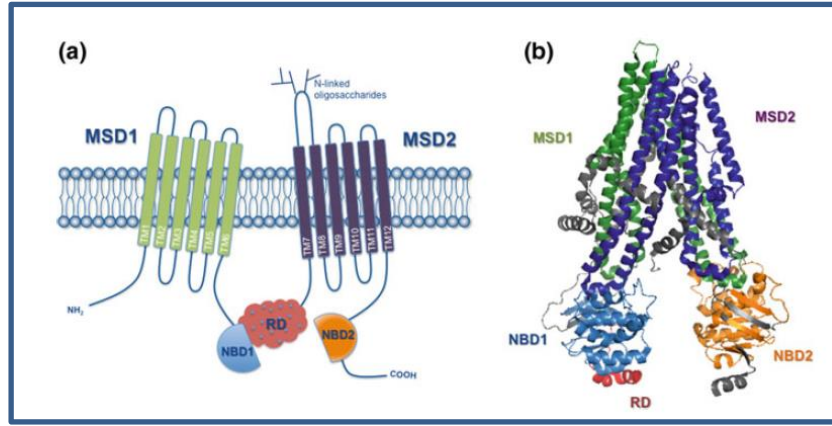
Şekil 2.2 Genden proteine KFTR geni ve proteini

Gen 27 ekzon içerir ve kromozom 7'nin uzun kolunda yaklaşık 190kb'lik bir bölgeyi kapsar. mRNA, UTR'ler dâhil 6.1kb uzunluğundadır. Protein, iki membranı kapsayan alan (MSD1 ve MSD2), iki nükleotid bağlayıcı alan (NBD1 ve NBD2) ve bir düzenleyici alan (RD) dâhil olmak üzere ABC taşıyıcı süper ailesine aittir. Hastalığa neden olan en yaygın mutasyon F508del'in (508 konumunda fenilalanin delesyonu) yeri Şekil 2.2'de gösterilmiştir (Collins, 1992).

Minimal KFTR promotörü, transkripsiyon başlangıç bölgesinin yaklaşık 226bç yukarısında bulunur ve tipik bir house-keeping promoter yapısına sahip gibi görünmektedir (Yoshimura et al., 1991). CpG açısından zengindir, TATA kutusu içermez, birden çok transkripsiyon başlangıç bölgesine sahiptir.

Genellikle bu tip promoterlarda olduğu gibi promoter zayıftır ve belirgin bir doku özgüllüğü olmadığından distal düzenleyici elemanların KFTR ekspresyonunun kontrolünde bir rolü olabileceğini düşündürmektedir. Aslında, KFTR lokusunu kapsayan genomik bölge, kodlama bölgesinin hem upstream (-20.9Kb'de)'de hem de down-stream (+5.4, +6.8, +7.0, +7.4 ve +15.6 Kb'de)'de ve çeşitli intronlarda birkaç DNase I aşırı duyarlı bölgesi (DHS) içerir (Phylactides et al., 2002; Smith et al., 2000). Son çalışmalar, hücre tipine özgü transkripsiyonu aktive etmek için, intronik enhancer elementlerin KFTR promotörünün yakınına getirildiği bir KFTR transkripsiyonel merkezinin varlığını önermektedir (Blackledge et al., 2009; Ott et al., 2009a). Oluşan döngü, KFTR eksprese olan hücrelerde saptanırken genin inaktif olduğu hücrelerde yoktur. Bu yapı, proksimal promotör, gen içindeki diziler, ancak promotörün 100Kb down-stream ve lokusa 3' uzaklıkta bulunan diğerleri arasında yakın bir etkileşime izin verir ve tümü ilgili DHS'ye karşılık gelir. Böylece kromatin yapısını değiştiren proteinlerin alımında yer alırlar (Ott et al., 2009a). Gen ifadesinin zamansal ve mekânsal düzenlenmesi aynı zamanda çoklu transkripsiyon başlangıç bölgelerinin (Yoshimura et al. 1991; Koh et al., 1993) kullanımına ve alternatif upstream ekzonlarının güçlenmesinin de dayanmaktadır (Broackes-Carter et al., 2002; Mouchel et al., 2003; Lewandowska et al., 2010).

Protein—Domainler ve Yapı: KFTR geni, 1480 AA uzunluğunda politopik bir membran proteini olan KFTR proteinini kodlar. Bu protein, her biri altı membranı kapsayan bölgeyi içeren iki membran kapsayan domain (MSD1 ve MSD2) veya transmembran segmentleri (TM1 ila TM12), iki ATP bağlayıcı alan (NBD1 ve NBD2) ve birkaç fosforilasyon bölgesine sahip benzersiz bir yüksek polar R domainden (RD) oluşur (Şekil 2.3).



Şekil 2.3 KFTR proteini ve Domainleri

Şekil 2.3'de KFTR proteini ve Domainlerini ile ilgili olarak; **a)** KFTR proteini beş alandan oluşur: MSD1 ve MSD2, her biri kanal pore oluşturan altı transmembran segmentten (TM1-6 ve 7-12), iki sitozolik NBD1 ve NBD2 ve düzenleyici RD'den oluşur. KFTR, dördüncü hücre dışı döngüdeki asparagin kalıntılarına bağlı iki N-bağlı glikana ve birkaç fosforilasyon bölgesine sahiptir (özellikle RD'de). **b)** Beş alanın montajı insan KFTR'sinin yayınlanmış kriyo-EM yapısının PDB koordinatlarına göre gösterilen, doğal KFTR'nin kompakt yapısına yol açan spesifik molekül içi etkileşimleri içerir (Liu et al., 2017). Açık kaynak kodlu PyMOL programı kullanılarak hazırlanan şekil (<https://www.pymol.org/>)

2.2.1.4 KFTR Disfonksiyonu ve Hastalığa Neden Olan Mutasyonlar

KFTR mutasyon veri tabanında listelenen 2106 mutasyon bulunmaktadır (KFTR mutasyon veri tabanı 2021). Spesifik bir DNA değişikliğinin gen ekspresyonu üzerindeki etkisine göre, mutasyonlar Yanlış Anlamlı (Missense) Mutasyon, anlamsız/stop-kodon (Nonsense) Mutasyon, çerçeve kayması (frameshift), küçük delesyonlar, insersiyonlar veya duplikasyonlar olarak sınıflandırılmaktadırlar.

Missense mutasyon: DNA dizisindeki üçlü (triplet) kodonda bir nokta mutasyon sonrasında kodondan sentezlenen AA değişimi,

Nonsense Mutasyon: Üçlü (triplet) kodonda yani mRNA dizisinde erken bir durdurma kodonunun girişine neden olup STOP kodon oluşumu,

Frameshift: Açık okuma çerçevesinde değişikliğe neden olan yaygın delesyonlar veya insersiyonlar,

mRNA prosesini etkileyen gen düzeyinde görülen mutasyonlar da vardır. Örneğin, Splice bölge mutasyonu Splicing mekanizmasını etkiler ve bunun sonucunda translasyon gerçekleşmez. Bu mutasyonlar aynı zamanda protein ürünü üzerinde yukarıda açıklanan etkilerden birine de neden olmaktadır. Belirli bir DNA mutasyonunun ilgili protein ürünü üzerindeki kesin etkisini bilmek için hastadan elde edilen materyallerde ve/veya heterolog sistemlerde ekspresyon çalışmaları yapılmalıdır (Rubenstein and Zeitlin 1998). Mutasyon türüne göre istatistikler Tablo 2.1’de gösterilmektedir (Alibakhsh et al., 2021).

Tablo 2.1 Mutasyon türüne göre istatistikler

Mutasyon Tipi	Miktar	% Frekansı
Missense	816	38.75
Frameshift	342	16.24
Splicing	230	10.92
Nonsense	177	8.40
In frame in/del	43	2.04
Large in/del	59	2.80
Promoter	17	0.81
Sequence variation	269	12.77
Bilinmeyen	153	7.26

KFTR geninde tanımlanan hastalığa neden olan mutasyonlar, genin eksprese edildiği epitelde Cl⁻ salgısının azalmasına neden olsa da farklı hücrel fenotipler meydana gelmektedir. Mutasyon etkisi ile KF hastalığı arasında doğrudan ilişki gösterilen mutasyonların sayısı giderek artmaktadır (Sosnay et al., 2013). Bu yalnızca hastalık mekanizmalarını anlamak için değil aynı zamanda temel kusuru düzeltmeyi amaçlayan yeni protein yönelimli tedaviler önermek için de oldukça önemlidir. Genin klonlanmasından beri çoğunlukla protein seviyesindeki fonksiyonel sonuçlara dayalı olarak KFTR mutasyonları için çeşitli sınıflandırmalar önerilmiştir (Welsh and Smith 1993; Wilschanski et al., 1995; Welsh et al., 2001; Rowe et al., 2005; De Boeck and Amaral, 2016). Tüm mutasyonlar KFTR protein ekspresyonu, yapısı ve fonksiyonları üzerindeki etkilerine göre 7 alt gruba ayrılmıştır (Şekil 4.6) (Boeck and Amaral, 2016).

Sınıf I: Protein üretiminin olmaması; Bu sınıfta esas olarak G542X, R553X, Q637X veya W1282X gibi ciddi klinik fenotiplerle ilişkili nonsense mutasyonlar, 394delTT veya 3905insT olarak çerçeve kayması mutasyonları (erken durdurma kodonlarına yol açar) ve 621+1G > T olarak splicing mutasyonları vardır.

Sınıf II: Plazma zarında bulunan protein miktarının yokluğuna veya azalmasına neden olan yetersiz hücre içi trafik; Örneğin Phe 508 del, Asn 1303 Lys, Ile 507 del, Arg 560 Thr'ye translokasyondan önce bozunmaya uğrar. En yaygın mutasyon, 3bç çifti delesyonuna bağlı olarak Phe508 del'dir (C15211523 del CTT). Bu mutasyon 508 konumunda dışlanmaya yol açar. Özellikle bu mutasyon tüm KFTR alellerinin %70'inden sorumludur. Phe508 del, endoplazmik retikulumla (ER) ilişkili degradasyona (ERAD) ve plazma membranında minimum protein ekspresyonuna maruz kalan KFTR proteininin yanlış katlanmasına katkıda bulunur.

Sınıf III: Membranda bulunan ancak cAMP stimülasyonuna kusurlu yanıt veren protein (regülasyon defekti); KFTR proteinini hücre zarına transloke olmuş ancak aktive olmamıştır. Bu sınıftaki mutasyonlar çoğunlukla, G551D, G551S, G1244E, S1255P veya G1349D gibi missense mutasyonlardır. Ayrıca genellikle ciddi bir klinik fenotiple ilişkilidirler.

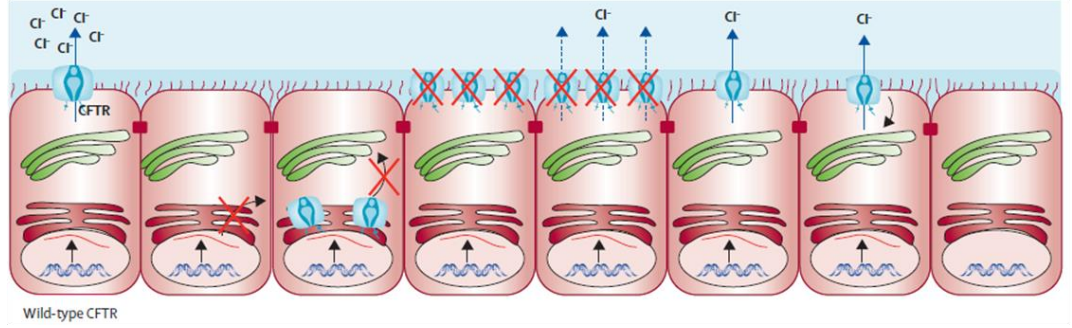
Sınıf IV: Membranda bulunan ve cAMP'ye yanıt veren ancak iletkenliği azaltılmış protein; Bazı KFTR işlevleri korunarak hafif bir KF biçimine yol açmaktadır. Bu sınıfın yaygın mutasyonları arasında Arg 117 His, Arg 347 Pro, Arg 117 Cys, Arg 334 Trp vb. bulunmaktadır.

Sınıf V: Protein üretiminde azalma; bu sınıf, anormal ve normal KFTR'nin, alternatif olarak splicing nedeniyle, eşzamanlı varlığına yol açan mutasyonları içermektedir. Bu vakalarda, bu nedenle normal KFTR üretimi bulunmaktadır. Ancak 3849-10kbCT veya 3272-26A>G insersiyon mutasyonları gibi düşük seviyelerde oldukları için bunlar genel olarak hafif bir fenotip ile ilişkilidir.

Sınıf VI: Plazma zarında proteinin azaltılmış stabilitesi; bu sınıf, KFTR'nin katlandığı, trafiğe girdiği ve neredeyse düzgün çalıştığı ancak zarı yırtılma ömrünün azaldığı mutasyonları içermektedir. Bu hızlandırılmış iş hacmi, endositozdaki bir artıştan veya geri dönüşümdeki bir azalmadan kaynaklanabilir. 120del23, KFTR'nin N-terminalinden yoksun olduğu bir mutasyon (Ramalho et al., 2009) ve F508del bu defekti gösterir.

Sınıf VII: Bu sınıf, büyük delesyonlar gibi KFTR mRNA üretiminin bile olmadığı mutasyonların gösterilmesi için tanıtılmıştır. Bu mutasyonlar sınıf I'e dâhil edilebilmesine rağmen, yeni bir sınıfın tanımı, herhangi bir farmakolojik yaklaşımla

kurtarılamaz oldukları gerçeğini açıklamaktadır (De Boeck and Amaral 2016). dele2, 3 (21kb) ve 1717-1G>A gibi mutasyonlar bu sınıfa dâhildir.



	Sınıf I	Sınıf II	Sınıf III	Sınıf IV	Sınıf V	Sınıf VI	Sınıf VII
CFTR defekti	Protein yok	Trafik yok	Geçit defekti	Azaltılmış iletkenlik	Az protein	Daha stabilite	mRNA YOK
Örnek Mutasyon	GLy542X, Trp1282X	Phe508del, Asn1303Lys, Ala561Glu	Gly551Asp, Ser549Arg, Gly1349Asp	Arg117His, Arg334Trp, Ala455Glu	Ala455Glu, 3272-26A→G, 3849+10 kg C→T	c. 120del23, rPhe508del	dele2,3(21 kb), 1717-1G→A
Düzeltilici tedavi	Kurtarma sentezi	Kurtarma trafiği	Kanal etkinliğini düzenleme	Kanal etkinliğini düzenleme	Doğru splicing	Promote stabilitesi	Ulaşılamaz
İlaç (onaylı)	Okuma bileşikleri (hayır)	Düzeltiliciler (evet)	Güçlendiriciler (evet)	Güçlendiriciler (hayır)	Antisens Oligonükleotidler, düzeltiliciler, güçlendiriciler? (hayır)	Stabilizatörler (hayır)	Bypass tedavileri (hayır)

Şekil 2.4 Gen mutasyonlarının sınıfları ve ilgili terapötik stratejiler

Şekil 2.4’de de gösterildiği gibi KFTR mutasyonları, aynı modülatörlerin bir sınıftaki tüm kusurlara uygulanabileceği beklentisiyle yedi fonksiyonel sınıfa ayrılır. Bununla birlikte, sınıf VII mutasyonlarının herhangi bir modülatör tarafından kurtarılabılır olması beklenmemektedir. Sınıf VII mutasyonları için bir terapötik strateji, alternatif Cl⁻ kanallarının uyarılmasını içerebilir (baypas tedavileri). Phe508del=rescued Phe508del.

Aslında bir mutasyonu sınıflandırmak görüldüğü gibi basit değildir (Welsh et al., 2001), çünkü bazen heterolog ekspresyon sistemlerindeki çalışmalar doğal dokulardaki gözlemlerle çelişkili olabilmektedir. En çarpıcı örneklerden biri en yaygın mutasyon F508del'dir. Ter bezlerinin epitel hücrelerinde WT- ve F508del-KFTR'nin immünohistokimyasal lokalizasyonu, bu mutant proteinin apikal membrana

ulaşmadığını, hücre içinde lokalize kaldığı gösterilmiştir (Dalemans et al., 1992). Bu anormal lokalizasyon, homozigot F508del hastalarından alınan primer hava yolu epitel hücrelerinde doğrulanıp sınıf II'ye dahil edilmesine yol açmıştır (Kartner et al., 1992). Ancak, kısa bir süre sonra, doğal dokularda apikal membranda F508del-KFTR saptanmıştır (Kalın et al., 1999). Daha sonra, aslında F508del-KFTR'nin, hücrelerin sadece küçük bir bölümünün apikal bölgesinde olduğu gösterilmiştir (Penque et al., 2000). Belirli bir mutasyon, farklı düzeylerde işlev bozukluğuna yol açabildiğinden birden fazla sınıfa dâhil edilmektedir. Bu nedenle, yanlış yerleştirme F508del ile ilişkili ana işlev bozukluğu olarak tanımlanabilse de ek kusurlar aşağıdakiler gibi tanımlanarak F508del-KFTR sıklıkla sınıf III ve VI ile ilişkilendirilmiştir (Farinha and Matos 2016).

- Azaltılmış fonksiyon (Drumm et al., 1991),
- Artan membran instabilitesi (Lukacs et al., 1993),
- Subapikal membranlardan yetersiz alım (Bradbury et al., 1992),
- Azaltılmış transkript seviyeleri (Ramalho et al., 2002)

Her bir mutasyonu sınıflandırmanın zorluğu varyantların tedavilerde gruplandırıldığı alternatif bir sınıflandırma yaklaşımına yol açmaktadır. Defektler, KFTR miktarını (transkripsiyon, splicing, katlama, trafik ve stabilizeyi kapsamaktadır) ve KFTR işlevini etkileyenler (iletkenlik ve aktiviteyi kapsayan) olarak ikiye ayrılmaktadırlar (Cutting, 2015). Her mutasyonun, temel kusurunun uyacağı tüm farklı kategorileri yansıtan bir sayı kombinasyonu ile sınıflandırılacağına göre daha karmaşık sistemler de önerilmiştir (Veit et al., 2016).

2.2.1.5 Hücredeki KFTR

KFTR'nin Biyosentezi, İşlenmesi ve Degrasyon: Gen ekspresyonundaki son adım fonksiyonel bir proteinin biyosentezidir. Bir zar proteini olan KFTR, salgı yolunu takip etmektedir. Bu, proteinlerin ER'dan Golgi'ye ve oradan salgı kesecikleri içinde hücre lizozomlarındaki, hücre dışı ortamdaki veya plazma zarındaki nihai hedeflerine exportu için genel bir yol olarak işlev görür.

KFTR biyogenezi ER zarına eş-translasyonel insersiyon ile başlar (Lu et al., 1998). ER zarına eş-translasyonel yoluyla eklenen proteinler bu yerleştirmeden sorumlu yaklaşık 20 AA kalıntısından oluşan bir sinyal dizisine sahiptir. Bu dizi, sinyal tanıma partikülü (SRP) tarafından tanınarak ER zarına ribozom eşleşmesini desteklemektedir. SRP, ribozomu, en önemli Sec61 kompleksi olan birkaç alt birimden oluşan translocon adı verilen bir kanala bağlayan ER zarındaki spesifik bir reseptöre bağlanır. Sinyal dizisi kanala girer ve yeni oluşan polipeptidin ER membranına birlikte translasyonel olarak yerleştirilmesine izin verilir (Lu et al., 1998). ER zarına yerleştirildikten sonra, yeni sentezlenen wt veya mutant (örneğin F508del) KFTR, dördüncü hücre dışı döngüde (ECL4) bulunan asparagin kalıntıları 894 ve 900'de glikosile edilmektedir. N-glikosilasyon, 14-osidik tortulu bir glikonjugat ilavesiyle meydana gelmektedir. Bu oligosakarit, iki N-asetilglukosamin (GlcNAc), üç glikoz (Gluc) ve dokuz mannoz (Man) tortusunu içerir ve konsensüs dizileri Asn-X-Ser/Thr içinde meydana geldikleri şekilde bu asparagin tortularına eklenirler. Bu yapı ER içinde sentezlenir ve özellikli bir oligosakkaril transferaz tarafından katalize edilen bir reaksiyonda membran lipidik öncü dolikol fosfattan proteine aktarılır (Farinha, 2017). F508del mutasyonunun KFTR işlevini nasıl tehlikeye attığına dair ilk kanıt F508del-KFTR'nin transfekte edilmiş COS hücrelerinde biyosentezi üzerine yapılan çalışmalardan elde edilmiştir. Bu kanıt bu mutantın büyük ölçüde ER içinde tutulduğunu düşündürmektedir (Cheng et al., 1990).

KF, mutant veya anormal proteinin fonksiyonel konformasyonunu gerçekleştirilememesinden kaynaklandığı bilinen, genellikle protein katlanması veya konformasyonel bozukluklar olarak adlandırılan çok çeşitli kalıtsal hastalıklara aittir (Taubes, 1996). KFTR'nin sadece izole edilmiş NBD1'ini kullanan *in vitro* çalışmalarda, F508del-NBD1 katlamanın etkinliğinin, wt-NBD1'inkine kıyasla ciddi şekilde bozulduğu gösterilmiştir (Strickland et al., 1997).

ER hücre yüzeyine sekretuar yol boyunca trafik sağlayan diğer birçok hücresel protein gibi, KFTR de protein kalite kontrol (PQK) faktörleri tarafından sürekli olarak izlenmektedir. KFTR ile ilgili kalite kontrol "kararları" genellikle hücrede iki kilit noktada verilmektedir. Bu kararlar ya KFTR sentezlenirken ER'de ve/veya KFTR katlanıp ER'den alındıktan sonra ER sonrası bölmelerde alınmaktadır (Estabrooks and Brodsky, 2020). Farklı KFTR varyantlarının bu konumların her birinde PQK ile

nasıl etkileşime girdiği önemlidir. POK mekanizması, hücresele proteinlerin bozulmasına veya katlanmasına yol açan yollar ve yanlış katlanmış proteinler biriktiğinde tetiklenen stres tepkileri topluca protein homeostazı veya “proteostaz” olarak adlandırılmıştır. KPTR POK, kanalı taban tabana zıt iki hücresele kaderden birine doğru yönlendiren faktörleri gerektirir. Bazı faktörler katlanma ve olgunlaşmayı desteklemek için KPTR'ye bağlanırken, diğer faktörler katlanmamış KPTR'yi tanır ve bunun yerine bu yanlış katlanmış, potansiyel olarak toksik kanalları bozunmaya yönlendirir. Bazı faktörler, ER'deki proteinlerin biyogenezi sırasında hem pro-matüratif hem de pro-degradatif aktivitelere sahip olsa da, tipik olarak bu iki aktiviteden birinin daha büyük bir net etkisi bulunmaktadır (Brodsky et al., 2007). Geniş anlamda, etkileşim ortakları üç gruptan birine ayrılmaktadırlar. Bunlar şaperonlar, yardımcı şaperonlar ve E3 ubiquitin ligazlarıdır. Moleküler şaperonlar (ısı şoku proteinleri veya Hsps olarak da bilinir), hücresele POK mekanizmasının KPTR ile etkileşime giren ilk bileşenleri arasındadır ve sıklıkla ko-translasyonel olarak hareket ederler. Şaperonlar, KPTR etki alanı derlemesini teşvik etmede ve glikosilasyonunun insan yapısını nasıl etkilediğini kontrol etmede temel oyuncularlardır. Bunlar Hsp90'ı ve sırasıyla Hsc70 ve Hsp70 olarak adlandırılan Hsp70'in hem yapısal olarak ifade edilen hem de stresle indüklenebilir izoformlarını içerirler (Loo et al., 1998; Meacham et al., 1999; Scott-Ward et al., 2009; Matsumura et al., 2011; Bagdany et al., 2017). Hsc70, Hsp70 ve Hsp90, KPTR'ye bağlanıp ATP hidroliz döngüleri boyunca katlamayı desteklerler. Spesifik olarak, hücre kültürü model sistemlerinde bunların ablasyonu, yeni oluşan kanalların terminal olarak yanlış katlanmasına ve bozulmasına neden olur. Bu şaperonların her biri nominal olarak KPTR katlanmasını desteklese de, Hsp70 ve Hsp90, çoğu kanal katlanmasını ve olgunlaşmasını antagonize olan çok sayıda başka POK bileşenini işe alır. Örneğin, Hsp70/Hsp90 düzenleyici protein (HOP), KPTR'ye bağlı Hsp90'ı Hsc70 veya Hsp70 ile etkileşime sokarak ve karşılığında kanalı bozunma için yönlendiren bir E3 ubiquitin ligazı olan CHIP ile etkileşime sokarak KPTR üzerinde genel bir pro-degradatif etkiye sahip olur (Şekil 4.7). Bu olaylar, KPTR'yi pro-katlanma yolundan pro-degradatif bir yola sokmaktadır. Beklendiği gibi, HOP'nin inhibisyonu KPTR olgunlaşmasını desteklemektedir (Marozkina et al., 2010; Zaman et al., 2016). Bunun tersine, Hsc70 ve Hsp70 için bir nükleotid değişim faktörü olan HspBP1, Hsc70 veya Hsp70'e bağlı CHIP'yi doğrudan inhibe edip yeni oluşan

KFTR'nin pro-katlanma yolunda kalmasına yardımcı olmaktadır (Kabani et al., 2002; Alberti et al., 2004).

Diğer Hsc70 yardımcı şaperonları, başlıca Hsp40 yardımcı şaperonları (J-proteinleri olarak da bilinirler), KFTR'nin erken olgunlaşması sırasında etkileri bakımından büyük farklılıklar gösterirler. ER-tümleşik Hsp40 Hdj2 (DNAJA1), KFTR'yi bağladığı ve üretken katlamayı desteklediği tanınan Hsp40 ailesinin ilk üyesidir (Meacham et al., 1999). Hsc70'in başlangıçta bağımsız olarak NBD1 katlanmasını *in vitro* olarak kolaylaştırdığı gözlemlenirken (Strickland et al., 1997) bu şaperonun hem Hdj2'yi hem de ikinci bir Hsp40, Hdj1'i (DNAJB1) ER tarafından tutulan wt etkin bir şekilde kurtarmak için kullanırken F508del KFTR'yi bozunmadan kullanmadığı bulunmuştur (Farinha et al., 2002). Ancak verimli KFTR katlanması için Hdj2 veya Hdj1 gibi Hsp40'lar gerekli olsa da bu yardımcı şaperonlar ayrıca parçalayıcı POK mekanizmasının unsurları olarak da işlev görebilirler. Örneğin, KFTR'yi katlanmadaki rolüne rağmen, Hdj2, KFTR alt alanlarının incelenmesi üzerine kanıtlandığı gibi, Hsc70/CHIP komplekslerinin ubiquitin ligaz aktivitesini de keskin bir şekilde artırabilir (Younger et al., 2004). Başka bir Hsp40, sistein dizisi proteini (Csp, DNAJC5), CHIP'yi bağımsız olarak alma yeteneğinden dolayı KFTR'nin bozulmasını kolaylaştırmaktadır (Zhang et al., 2002; Zhang et al., 2006; Schmidt et al., 2009). Benzer şekilde, DNAJB12 hem olgunlaşmamış wt formlarının hem de F508del KFTR'nin bozunmasını uyarır. Ancak bunu başka bir E3 ubiquitin ligazı, RMA1'i Hsc70'e dâhil ederek yapmaktadır (Yamamoto et al., 2010; Grove et al., 2011). Hsp40'ların bozunma sırasındaki rolü, model sistemlerde yapılan çalışmalarda da kanıtlanmıştır. Örneğin, maya ER lokalize Hsp40 homologları Ydj1 ve Hlj1 ektopik olarak eksprese edilen KFTR'nin bozunmasına katkıda bulunmak için fazladan işlev görürler (Youker et al., 2004). Hsp70'in yıkıcı etkilerinin ve onun daha geniş ko-şaperoni ağının (Yang et al., 1993) aksine, Hsp90, maya modelindeki delesyonu veya insan hücre kültüründe küçük moleküller kullanılarak inhibisyonu yoluyla işlevinin bozulması, KFTR sitozolik alanlarının doğru bir şekilde birleştirilmesini engellediği ve yeni oluşan kanalların bozulmasına neden olduğu için önemli bir pro-folding şaperondur (Loo et al., 1998; Youker et al., 2004). Buna karşılık, küçük ısı şoku proteinleri (sHsps), wt ve F508del KFTR biyogenezi sırasında daha karmaşık bir rol oynarlar. Bu şaperonlar başlangıçta KFTR'nin POK'sine dâhil edilmiş çünkü maya sHsps, Hsp26 ve Hsp42'nin delesyonu, ubiquitin ER tarafından tutulan kanala

2.2.1.6 KFTR'nin Fonksiyonları

Klorür Kanalı Olarak KFTR: KFTR, birkaç organın dışa bakan yüzeylerini kaplayan hücrelerin apikal zarlarında bulunan ATP kapılı bir klorür kanalıdır (Meng et al., 2019; Riordan, 2008). KFTR'nin yüksek düzeyde yüklü düzenleyici bölgesinin (R-bölgesi) protein kinaz A veya C tarafından fosforilasyonu, ATP'nin mevcut olması koşuluyla, daha sonra açılabilen ve klorür iyonu akışına izin veren kanalı aktive etmektedir (Chappe et al., 2003). Cl⁻ iyonları, konsantrasyon gradyanları boyunca proteinden (pasif olarak) geçmektedir. Tipik olarak bu, KFTR kanal açıklığının hücreden klorür iyonlarının kaybı ve hücre dışı ortamda birikmesi ile ilişkili olduğu anlamına gelmektedir. Klorürün hareketinin neden olduğu yük dengesizliği, aynı yöndeki sodyum iyonlarının eş zamanlı akışıyla düzeltilir ve ortaya çıkan ozmotik dengesizlik hücre dışı boşluğa su akışı tarafından düzeltilir. Aşırı KFTR aktivitesi ve dolayısıyla su akışı intestinal yolda diyareye neden olabilir. Normalin altında KFTR aktivitesi, bağırsakta kabızlığa ve ayrıca sağlıklı akciğerlerin korunması için hayati önem taşıyan mukosiliyer temizleme mekanizmasının durmasına yol açabilir (Donaldson et al., 2006). Bu son sonuç KFTR geninin iki mutasyona uğramış ve kusurlu kopyasının kalıtımı nedeniyle KFTR aktivitesini önemli ölçüde azaltan KF hastaları için en ciddi semptomdur (Riordan et al., 1989). KFTR'de kanal açılması, ATP-bağlama alanlarının dimerize olduğu dışa dönük durumla ilişkilidir (Vergani et al., 2005; Cai et al., 2011; Yeh et al., 2017). KFTR kanal işlevi de fosforilasyon tarafından aktive edilerek kontrol edilir ve burada yeni yapısal veriler fosforilasyonun KFTR aktivitesini nasıl etkileyebileceğine dair farklı modeller sağlamaktadırlar (Zhang et al., 2017; Fay et al., 2018). Avrupa kökenli hastalarda ağırlıklı olarak gözlenen mutasyon, proteinin AA dizisinde 508 konumunda bir fenilalanin kalıntısının kaybıyla sonuçlanan üç bazın silinmesi olmasına rağmen, KFTR genindeki mutasyonlar çeşitlidir (Riordan et al., 1989; Veit et al., 2016).

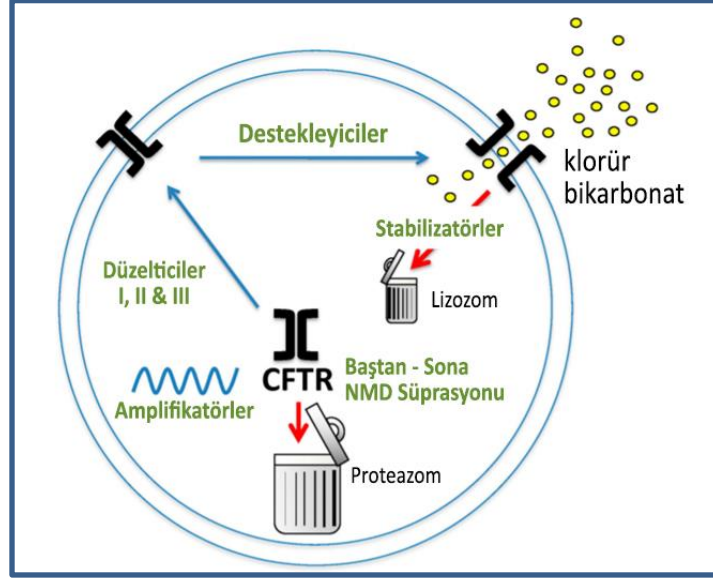
Diğer Fonksiyonlar: KFTR tarafından diğer kanalların ve epitelyal iyon taşınmasının düzenlenmesi, hücre zarında bir "bağlama platformu" rolüne dayanmaktadır. KFTR'ye sabitlenmiş özel "mikro domainler", sinyal molekülleri, kinazlar, taşıma proteinleri, PDZ alanı içeren proteinler, miyozin motorları, Rab GTPazlar ve SNARE'leri içeren moleküler anahtarlar tarafından dinamik olarak düzenlenen bir dizi proteini birlikte gruplayabilirler (Guggino and Stanton, 2006).

KFTR'nin ayrıca hücre içi pH'ı, yani endozomlarda düzenlemesi önerilmiştir. Endozomlarda KFTR bulunduğunda, plazma zarına taşınması sırasında, organel zarında bir Cl⁻ kanalı olarak işlev görmektedir. Bu nedenle organelin iç pH'ı, KFTR ile H⁺'ye bağlı ATPaz (proton pompası) faaliyetleri arasındaki dengeden kaynaklanacaktır (Barasch et al., 1991; Lukacs et al., 1992). Ek olarak endozom asitlenmesinin KF hücrelerinde kusurlu olduğu gösterilmiştir (Barasch et al., 1991). Bu pH değişikliği, endozomun uygun asitleştirilmesi sağlanmadığında aktif olmayan spesifik glikosilasyonlardan sorumlu enzimlerin, yani sialiltransferazların aktivitesinde bir değişikliğe yol açmaktadır. KF hastalarında, membran glikoproteinleri (KFTR dışında), sialilasyon seviyelerinde bir azalma ve sülfat gruplarının mevcudiyetinde artış göstermektedir. Bu da muhtemelen *Pseudomonas* gibi patojenik ajanların yapışması ile ilgili sonuçlara sahiptir (Barasch et al., 1991). Son olarak, KFTR'nin hücresel farklılaşma, epitel rejenerasyonu ve organizasyonu gibi süreçlerde de rol oynadığı gösterilmiştir.

2.2.1.7 Temel Defektin Düzeltilmesine Yönelik Tedaviler

KF nedeni bilinmeyen ölümcül bir hastalıktan moleküler hassasiyetle anladığımız bir duruma evrimleşmiştir (Collins, 2019; Bergeron and Cantin, 2021). KFTR işlevini geri kazanmayı amaçlayan umut verici yeni tedaviler, KF'li çoğu bireyin yaşamını hızla değiştirmektedir (Bardin et al., 2021).

Farklı mutasyonlardan kaynaklanan anormal KFTR proteininin işlevini düzeltmek için umut verici stratejiler Şekil 4.8'te özetlenmiştir. KFTR modüle edici ajan sınıfları arasında güçlendiriciler, düzelticiler, amplifikatörler, okuma ajanları, NMD baskılayıcılar ve stabilizatörler bulunmaktadır.



Şekil 2.6 Çeşitli KFTR mutasyon sınıfları olan KF bireylerde KFTR fonksiyonunu geri yüklemek için farmakolojik stratejiler

Şekil 2.6’da, Yeşil renkle gösterilenler, KFTR’nin (güçlendiriciler) açık olasılığını artıran, yanlış katlanmış proteinin ERAD’dan kaçışını kolaylaştıran ajanlardır (sınıf I düzelticiler NBD1, sınıf II üzerinde NDBII ve sınıf III düzelticiler, sınıf I düzelticinin varlığında ek düzeltici etkisine sahiptir). KFTR mRNA (amplifikatörler) miktarının artırımı PTC mutasyonları ile mRNA’nın uygun translasyonunu arttırımı (okuma ajanları), NMD’yi (NMD baskılayıcı) azaltımı ve plazma zarına yerleştirilmiş KFTR proteininin (stabilizörler) bozulmasının önlenmesidir. FDA onaylı KFTR modülatörleri ve KF bireylerde kullanım için endikasyonlar Tablo 2.2’te özetlenmiştir (Bergeron and Cantin, 2021).

Tablo 2.2 FDA onaylı KFTR modülatörleri ve KF bireylerde kullanım endikasyonları

İlaç (Şirket Adı)	Aksiyon Durumu	Yaş	Mutasyon Sınıfı	Alleller
Ivacaftor (Kalydeko)	Güçlendirici	2 yaş <	III, IV	G551D, G1244E, G1349D, G178R, G551S, S1251N, S1255P, S549N, S549R, R117H, diğer nadir mutasyolar
Ivacaftor and lumacaftor (Orkambi)	Güçlendirici ve düzeltici	12 yaş <	II	F508del homozigot
Ivacaftor and tezacaftor (Symdeko)	Güçlendirici ve ikinci nesil düzeltici	12 yaş <	II/II, veya II/RF*	F508del homozigot veya F508del/RF*

Ivacaftor, tezacaftor and elexacaftor (Trikafta)	Güçlendirici ve ikinci nesil düzeltici ve yeni nesil düzeltici	12 yaş <	II/II, II/diğer	KF teşhisi ve en az bir F508del veya başka bir KFTR'ye yanıt veren mutasyon**
--	--	----------	-----------------	---

*RF; Residüel fonksiyon KFTR mutasyonu şunları içermektedir E56K, P67L, R74W, D110E, D110H, R117C, E193K, L206W, R347H, R352Q, A455E, D579G, 711 + 3A→G, E831X, S945L, S977F, F1052V, K1060T, A1067T, R1074L, 12H, D 2789 + 5G→A, 3272–26A→G, 3849 + 10kbC→T. **Mutant KFTR'yi ekspres eden Fisher sıçan tiroid hücrelerinde KFTR klorür taşıma analizine dayalı olarak yanıt veren KFTR geninde bir mutasyon.

Diğer KFTR güçlendiriciler ve düzelticiler, son çalışmalarda özetlendiği ve Tablo 2.3'da listelendiği gibi, şu anda pre-klinik veya klinik faz geliştirme aşamalarında (Bergeron and Cantin, 2021).

Tablo 2.3 Mevcut klinik kullanımda olanlar dışındaki KFTR modülatörleri

MOLEKÜL ID	ŞİRKET	MEKANİZMA	KLİNİK DENEME NO
VX-561 deutivacaftor	Vertex İlaç	güçlendirici	NCT03911713
ABBV-974	AbbVie	güçlendirici	NCT02707562 NCT02690519
ABBV-2451	AbbVie	güçlendirici	NCT03540524
ABBC-3067	AbbVie	güçlendirici	NCT03969888
QBW251 icentacaftor	Novartis	güçlendirici	NCT02190604
VX-121	Vertex İlaç	düzeltilici	NCT03912233 NCT03768089
VX-440 olacaftor	Vertex İlaç	düzeltilici	NCT02951182
VX-659 bamocaftor	Vertex İlaç	düzeltilici	NCT03447249
ABBV-2222 galicaftor	AbbVie	düzeltilici	NCT03969888
ELX-02	Eloxx İlaç	Read-through	NCT04135495 NCT04126473

KF'li kişiler için yeni tedavilerin klinik geliştirme programları, tedavilere verilen bireysel tepkilerde belirgin farklılıklar ortaya çıkarmış ve bu nedenle bazı kişiler için en iyi tedaviyi seçmeyi zorlaştırmıştır. KFTR genotipi, terapötik yanıtları tanımlamada önemli bir faktör olmasına rağmen, KFTR alellerinden bağımsız genetik ve çevresel faktörler, modülatör tedavisine verilen yanıtları belirgin şekilde etkilemektedir. Özellikle nadir KFTR mutasyonlarının varlığında her bir KF bireyi için modülatörlere verilen tedavi yanıtlarını daha iyi tahmin etmek için tanısal testlere (teranostik) açık bir ihtiyaç vardır. Bireysel tepkileri izlemek için, sınıf III mutasyonlar (Durmowicz et al., 2018), ter klorür konsantrasyonu (Zemanick et al., 2018), uyarılmış pluripotent hücreler (Eckford et al., 2019) ve bağırsak organoidleri (Dekkers et al., 2013; Dekkers et al., 2016; Ramalho et al., 2021) için Fisher sıçan tiroid hücrelerinde modülatörlere verilen nadir mutasyon tepkisinin *ex vivo* testi dahil olmak üzere çeşitli

öngörü sonuç ölçütleri önerilmiştir. KF için önerilen birkaç öngörücü sonuç ölçütü, Avrupa Kistik Fibrozis Derneği stratejik planlama görev gücü tarafından "KF için yeni ilaçlara erişimin hızlandırılması" konusunda gözden geçirilmiştir (Amaral et al.,). Bu çalışmalar arasında, bağırsak organoidleri, yeni tedavileri bireysel ihtiyaçlara uyarlamak için özel bir umut vaat etmektedir. KFTR modülatör tedavisine KFTR fonksiyonel yanıtı, uzun vadeli klinik yanıtlarla yakından ilişkili olduğu gösterilen bağırsak organoid FIS'nin *ex vivo* ölçümleriyle yansıtılmaktadır. Bu da bunu nadir KFTR mutasyonları taşıyan hastalarda modülatörlere bireysel yanıtları tahmin etmek için çekici bir test haline gelmektedir (Berkers et al., 2019; Groot et al., 2020; Silva et al., 2020).

KFTR mutasyon sınıfından bağımsız yeni bir terapötik strateji mRNA'yı amplifiye etmek ve proteinin translasyonunun artmasına izin vermektir. KFTR proteini daha sonra, KFTR protein yoğunluğunu artırmak ve hücrenin apikal yüzeyinde işlev görmek için düzelticiler ve güçlendiricilerle birlikte kurtarılabilirler (Molinski et al., 2017). Amplifikatörler mRNA translasyonunu artırır ve yeni oluşan polipeptitlerin translokon olarak bilinen bir protein kompleksi yoluyla membrana entegrasyon için ER'a geçişini kolaylaştırarak mRNA bozulmasını azaltırlar.

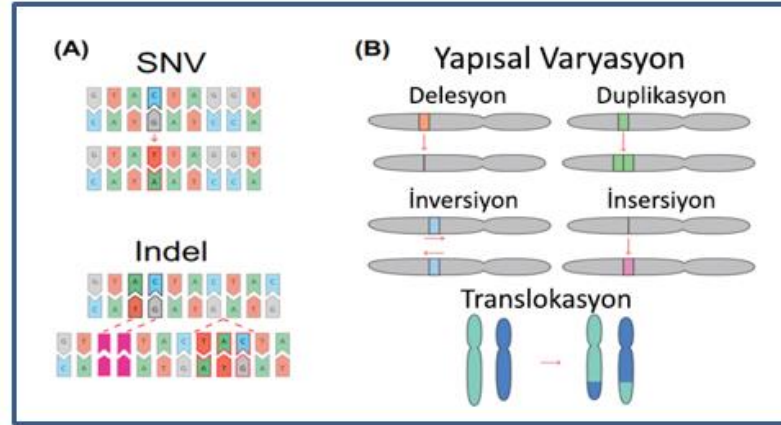
mRNA'nın insanlara verilmesini amaçlayan teknolojiler, terapötik ve aşı alanlarında hızla ilgi kazanmıştır. İyonize edilebilir lipid, kolesterol, fosfolipid ve etilen glikolden oluşan nanopartiküller, konakçının bağışıklık savunmasından kaçınırken membran destabilizasyonu ve endozomal kaçış kullanarak mRNA'yı iletmek üzere tasarlanmıştır. mRNA'nın başarılı bir şekilde verilmesi ve fonksiyonel KFTR proteininin üretimi, KFTR mRNA'yı taşıyan nanopartiküllerin aerosolizasyonunun, KF hava yolu hücrelerinden bunun %55'ine kadar iyodür akışını geri yüklemek için yeterli KFTR proteininin ekspresyonunu indüklediği bir murin KF modelinde gösterilmiştir. Bu durum sağlıklı farelerde transfeksiyondan 3 gün sonra gözlenmiştir (Robinson et al., 2018).

Fonksiyonel mRNA ve proteini solunum yoluna iletmek için alternatif bir yaklaşım gen tedavisidir. Otuz yıl önce KFTR geninin keşfedilmesinden bu yana umut verici bir tedavi olarak kabul edilmesine rağmen, vektörlerin yeterli taşıma kapasitesine sahip olması, insan hava yolu için tropizmlı bir kapsid ifade etmesi ve

düşük immünojenisiteye sahip olması gerektiğinden, KF gen terapisi geliştirmenin oldukça zor olduğu kanıtlanmıştır (Bergeron and Cantin, 2021).

2.3. Genetik Varyasyonlar

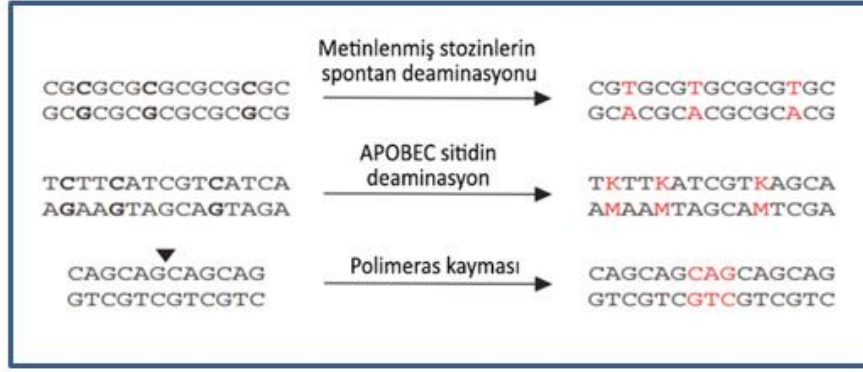
DNA dizi varyasyonu, hem germ hem de somatik hücrelerin sabit bir özelliğidir ve tek DNA nükleotid değişikliklerinden tüm kromozomların delesyonu veya duplikasyonuna kadar değişen bir ölçekte meydana gelebilir. Genomik varyantlar büyük ölçüde büyüklükleri ve oluşum mekanizmaları ile tanımlanır. Tek nükleotid varyantları (SNV), DNA'da, transisyon (yani, pürin ile pürinin veya bir pirimidin ile pirimidin değişimi) veya transversiyonlar (bir pürinden pirimidine geçiş veya tam tersi) olarak da sınıflandırılan, bir baz çiftinde meydana gelen Süstitüsyon olarak bilinirler (Nesta et al., 2020). İki kişi arasındaki farkın ~%80'ini oluşturan insan genomundaki en çok sayıda varyantlardır ve her nesil için baz çifti başına $\sim 1 \times 10^{-8} - 1 \times 10^{-9}$ mutasyon oranında meydana gelmektedirler (Shendure and Akey, 2015; Sasani et al., 2019). SNV mutasyonel hotspot, CpG açısından zengin lokuslardaki (Coulondre et al., 1978; Shendure and Akey, 2015;) germline varyantlarını ve aktivasyonla indüklenen sitidin deaminaz (AID) veya apolipoprotein B mRNA düzenleme enzimi, katalitik polipeptit (APOBEC) aracılı deaminasyonun neden olduğu somatik mutasyonları içermektedir (Muramatsu et al., 2000; Burns et al., 2013). Bir insersiyon veya delesyon (indel), <50 baz çiftini kapsayan bir veya daha fazla nükleotidin kazanımı veya kaybı olarak tanımlanmaktadır.



Şekil 2.7 Genetik Varyasyon Türleri

Genetik Varyasyon Türlerinden SNV'ler ve indeller, tek bir olayda bir ile 50bp etkileyen değişikliklerdir. **(A)** Bir C:T SNV örneği ve bir iki bp delesyon ve bir üç bp insersiyon indel. **(B)** SV'leri oluşturan ≥ 50 bp üzerindeki olayların örnekleri; bu olaylar, bu temel varyant türlerinin delesyonlar, duplikasyonlar, inversiyonlar, insersiyonlar, translokasyonları içerir. Şekil 2.7'de temsil edilen kromozomlardan üstteki kromozom referanstır. Alt bulunan kromozom ise varyasyonu vurgulamaktadır (Nesta et al., 2020).

İndellerin yaklaşık %45'i genomun ~%4'ünde yoğunlaşmıştır ve çoğu polimeraz kayması ile açıklanmaktadır (Şekil 2.8). Homopolimerik çalışmalar ve tandem tekrar bölgeleri gibi polimeraz kaymasına eğilimli diziler, indellerin %70'ini içermektedirler (Montgomery et al., 2013). Kodlama bölgesi indeksleri, SNV'lerden daha güçlü negatif seçim altındadır. dbSNP veri tabanına dayalı olarak, kodlama bölgelerindeki SNV'lerin sayısı kısa indellerden yedi kat fazlayken, SNV'lerin ve indellerin yoğunluğu intronik bölgelerde hemen hemen aynıdır (Jin et al., 2018).



Şekil 2.8 Varyant oluşum mekanizmaları

Varyant oluşumunun mekanizmaları, ortaya çıkan varyantın yanı sıra dizi bağlamından da çıkarılabilmektedir. Örneğin, metillenmiş sitozinlerin kendiliğinden deaminasyonu, GC açısından zengin lokuslarda C:T geçişleriyle sonuçlanır. APOBEC sitidin deaminazlar, tCw motifleri üzerinde hareket ederek onları tTw veya tGw'ye değiştirir (IUPAC kodu K). Bu model kataegiste gözlenir. Son olarak, tekrarlayan DNA bölgeleri, polimeraz şablondan ayrıldığında ve yeniden birleşme üzerine yanlış bölge ile eşleştiğinde, ekspansiyon veya kontraksiyon ile sonuçlandığında indellere maruz kalırlar.

Yapısal varyantlar (SV), delesyonlar ve duplikasyonlar (CNV), inversiyonlar, insersiyonlar ve translokasyonlar olmak üzere genomun büyük ölçekli değişiklikleridir (>50bp). Tahmini SV mutasyon oranı, nesil başına ~0,29'dur (Collins et al., 2020). SV'ler, SNV'lerden ve indellerden daha az sıklıkta ortaya çıksa da herhangi iki insan genomunu karşılaştırırken daha fazla sayıda nükleotitteki varyasyonu dikkate almaktadırlar (Conrad et al., 2010).

İnsan genetiği ve genomikte varyasyonlar çeşitli semboller ile isimlendirilirler. İnsan Varyasyonu Genom Topluluğu (www.hgvs.org), dizi varyasyonlarının ve gen adlarının tanımlanması için tutarlı isimlendirme standartlarını korumaktadır. İnsan genleri, İnsan Geni Adlandırma Komitesi (HGNC) tarafından belirlenen semboller kullanılarak adlandırılır ve genellikle büyük harfle ve italik olarak yazılırlar. Örneğin, *SMN1* geni survival of motor neuron 1 kısaltmasıdır. Aynı genin protein ürünü için ise italik olmayan sembol kullanır. Örneğin: SMN1 protein. Referans dizisindeki belirlenmiş varyantlara veya değişikliklere ve bunların farklı moleküller üzerindeki etkilerine atıfta bulunmak için çeşitli semboller ve kısaltmalar kullanılmaktadır. Belirli

moleküllere yapılan referanslar genellikle Ulusal Biyoteknoloji Bilgi Merkezi tarafından sağlanan RefSeq veri tabanını kullanır. Tablo 2.4'de yaygın kısaltmalar ve isimlendirme kuralları gösterilmiştir.

Tablo 2.4 Genetik varyantların isimlendirilmesi için kullanılan kısaltmalar

Kısaltma	Yorumlama	Örnek
“g.”	Bir genomik referans dizisini ifade eder. Bunu, belirli bir nükleotit veya varyantın kromozomu ve konumu dâhil olmak üzere genomik koordinatlar izler. Birden fazla derleme varsa, referans derlemenin sürümünü dâhil edilmelidir.	NC_000009.12: g.114195977G.C
“c.”	Kodlayıcı bir DNA referans dizisini ifade eder. Genellikle belirli bir transkript izoformuna atıfta bulunularak verilir. Bunu, kodlama dizisi ve nükleotid veya varyant içindeki konum izler.	NM_032888 (COL27A1): c.2089G.C
“p.”	Bir protein referans dizisini ifade eder. Genellikle belirli bir protein izoformuna atıfta bulunularak verilir. 3 harfli kod kullanılarak protein dizisi içindeki pozisyon ve amino asit değişimi takip eder.	NP_116277 (COL27A1):p.Gly697Arg

İnsan genetiğinde yaygın olarak kullanılan bir kaynak, Online Mendelian Inheritance in Man (OMIM) veri tabanıdır (www.omim.org). OMIM, varsayılan bir genetik temele sahip insan genleri ve fenotiplerinin sürekli güncellenen kapsamlı bir özetidir. Fenotip ve genotip arasındaki ilişkilere ve literatür taraması ve küratörlüğüne dayalı olarak Mendel bozuklukları olarak adlandırılan gen-hastalık ilişkilerini kuran belgelere odaklanır. Bu kitap boyunca birçok farklı genetik bozukluğa isimleriyle ve ayrıca kısaltmalarıyla atıfta bulunularak belirlenmiş altı haneli kimlik numaralarını veya MIM numaraları verilmektedir Okuyucu daha sonra klinik özellikleri ve ilgili bilgiler hakkında daha fazla bilgi edinmek için bu benzersiz tanımlayıcıları kullanarak OMIM'de bu tür ilgi hastalıklara bakabilmektedir. Tümünde atıfta bulunulan genetik bozuklukların MIM sayılarından önceki # sembolü, bu bozuklukta etkilenen moleküler temelin veya genin bilimsel literatürde tanımlandığını ve belgelendiğini göstermektedir.

Bireysel varyantların hastalığa neden olan (patojenik) veya hastalık ekspresyonu ile sonuçlanmayan (iyi huylu) varyantlar için belirleme kriterleri için bir çerçeve, öngörülen fonksiyonel etki, popülasyon sıklığı, evrimsel koruma, ilgili bireyler arasında alel ayrımı veya laboratuvar çalışmaları gibi faktörlere dayalı olarak ACMG tarafından oluşturulmuştur (Richards et al., 2007). Bu faktörlerin tümü

bilinmeyebileceğinden veya iddialarıyla çelişebileceğinden, varyantlar "olası patojenik", "muhtemelen iyi huylu" veya basitçe "önemi bilinmeyen bir değişken" olarak da sınıflandırılabilirler. Bu tanımlamalar laboratuvarlar arasında farklılık gösterebilir ve genellikle bir iddiayı diğerine karşı desteklemek için yeni kanıtlara dayalı olarak değişebilirler. Açık erişimli ClinVar veri tabanı, hasta örneklerinde bulunan varyantlardan, klinik önemine ilişkin iddialardan, test isteği yapan hakkında bilgilerden ve bu atamada kullanılan destekleyici verilerden bu yorumları toplayarak korumaktadır.

2.4. Genetik Tanı Testleri

Genomik testler, birçok genetik hastalığın tanı ve tedavisinde rutin olarak kullanılmaktadır. Tek bir testte, yalnızca hastanın durumunun etiyolojisini açıklamakla kalmayıp aynı zamanda tedavi kontrolünü bilgilendirebilecek DNA değişiklikleri için yüz binlerce gen de incelenmektedir. Klinik genomik testler, on yıldan daha kısa bir süredir klinik uygulamada kullanılmaktadır ve bu kısa süre nedeniyle genomik testlerin uygun klinik kullanımı ve yorumlanması hala gelişmektedir (Thodeson and Park, 2019).

2.4.1. Yeni Nesil Dizileme

DNaseq, bir DNA molekülündeki nükleotidlerin dizi ve sırasını belirleme işlemidir. 1953 yılında Watson ve Crick tarafından X-ışını kristalografi görüntüleri ve Raymond Gosling ve Rosalind Franklin tarafından elde edilen veriler kullanılarak DNA'nın çift sarmal yapısının belirlenmesinden sonra (Attar, 2013), birçok bilim insanı DNA'nın sekanslanmasına izin verecek moleküler biyoloji tekniklerinin geliştirilmesi ile ilgili çalışmalar yapmıştır. Aslında protein ve RNAseq teknikleri DNaseq tekniklerinden önce geliştirilmiştir (Gonzaga-Jauregui and Mendoz, 2021). Ancak, DNA'nın baz tamamlayıcılığının doğal özelliğinin anlaşılması DNaseq teknolojilerinin geliştirilmesine aracı olmuştur. 1977'de Frederick Sanger, zincir sonlandırıcı nükleotidleri kullanarak DNaseq yöntemini yayınlamışlardır (Sanger et al., 1977). Bu yöntem bazı modern ve en çok kullanılan sıralama teknolojilerinin temeli haline gelmiştir.

Sekanslama teknolojileri, DNA zincirindeki nükleotid dizisini belirlemek için, enzimatik kimya ve kullanılan mekanizmadan yararlanılarak kategorize edilmektedir. Genel olarak bunlar, bugüne kadarki dizileme teknolojilerinin çoğunu kapsayan

sentezleme ile dizileme (SBS) (Mitra et al., 2003) ve ligasyon ile dizileme (SBL) (Antoine et al., 2011) olarak ikiye ayrılmaktadırlar. Adından da anlaşılacağı gibi SBS teknolojileri, "reading" adı da verilen, nükleotidleri dâhil ederek DNA dizisini belirlemek için bazıları birer birer dâhil etmektedir. Bu sayede yeni bir tamamlayıcı zinciri sentezlemek için bir şablon DNA tek zincirini kullanmaktadır. SBS teknolojileri, işleme enzimi olarak bir DNA polimerazı kullandığı için bunların çoğunun termostabil olacak şekilde tasarlanmasına ve yüksek işlenebilirlik, özgüllük ve aslına uygunluk oranlarına sahip olmalarına rağmen, kullanılan sentetazların doğal hata oranına bağlıdır. Tersine SBL, şablon DNA zincirindeki diziye tamamlayıcı olan önceden belirlenmiş dizilere sahip kısa oligonükleotidler birleştiren bir DNA ligaz enziminin kullanımını içermektedir. Her durumda, dizileme reaksiyonunu ve yeni nükleotitlerin dâhil edilmesini başlatmak için şablon zincirin bir ucunu tamamlayıcı bir kısa oligonükleotit primeri kullanılır.

Sekanslama teknolojileri genel olarak 3 gruba ayrılır:

1) **Sanger Dizileme:** Sanger dizileme en yaygın kullanılan birinci nesil dizileme teknolojisidir. İnsan genomunun sekanslanması aslında bütünüyle Sanger Dizilimi kullanılarak tamamlanmıştır (IHGSC, 2001). Sanger dizilimi, DNA zincirinin uzaması için bir sonraki nükleotidin 5' ucu ile kovalent fosfodiester bağı oluşturmada önemli olan 3-OH doğal dNTP grubundan yoksun modifiye ddNTP'ler kullanan bir SBS teknolojisi.

2) **İkinci Nesil Dizileme Teknolojisi:** YND olarak da bilinen ikinci nesil sekanslama teknolojilerinin ilki Paralel multiple kısa okuma sekanslamadır. Bu strateji, orijinal bir DNA örneğinden türetilen ve *in vitro* olarak PCR kullanılarak çoğaltılan binlerce kısa DNA fragmanın (multipleks-çoklu) şablonundan oluşan bir kitaplığın hazırlanmasına ve ardından dahil edilen bç'lerinin kimliklerini saptamak için tüm fragmanların aynı anda (paralel) dizilenmesine dayanmaktadır. Sanger dizileme ile karşılaştırıldığında, bu yaklaşım dizileme sürecini önemli ölçüde hızlandırırken, maliyeti de önemli ölçüde azaltmıştır (Yohe and Thyagarajan, 2017). 2005 ve 2012 yılları arasında, çeşitli ticari şirketler tarafından entegre sekanslama platformları olarak çeşitli YND teknolojileri geliştirilmiştir. Bir istisna dışında YND teknolojilerinin çoğu, reaksiyonun sentez fazı sırasında nükleotid eklenmesi için

okuma olarak kullanılan molekülü saptamak için kullanılan kimya ile birbirlerinden farklılık gösteren SBS teknolojileriydi.

Tersinir terminator (Reversible terminators): En yaygın olarak kullanılan Illumina (Solexa) teknolojisi, değiştirilmiş floresan etiketli tersinir sonlandırıcılar ddNTP'ler ile Sanger dizileme ile aynı prensibi kullanmaktadır. Tersinir sonlandırıcılar kullanan SBS, bir flowcell slaytında solid fazlı köprü amplifikasyonu tarafından oluşturulan kümelere (cluster) dayanmaktadır. Sekans döngüsü, dört floresan etiketli 30 modifiye nükleotid sonlandırıcı ve bu modifiye edilmiş nükleotitleri dâhil edebilen tasarlanmış bir DNA polimeraz karışımının yıkanmasıyla başlar. Nükleotid, hazırlanmış şablondaki bir sonraki baz için tamamlayıcı ise, polimeraz tarafından eklenecektir. Uzatma bloke edilecek ve floresanların her birinin lazer uyarımı ile elde edilen floresan sinyali, yüksek çözünürlüklü bir kamera tarafından tespit edilecektir. Görüntüledikten sonra, modifiye edilmiş nükleotitlerin sonlandırıcı grubu, bir sonraki spesifik nükleotidin eklenmesi ve yeni bir döngünün başlatılması için 3'-OH'nin rejenerasyonuna izin veren floresan ile birlikte bölünmektedir. Bu sonlandırıcının varlığı, aynı döngüde hiçbir ek veya spesifik olmayan nükleotidin eklenmemesini sağlayarak, her döngüde şablon başına yalnızca bir nükleotidin görüntülenmesine izin verdiğinden, bu teknolojinin kimyası için anahtardır. Gerçek nükleotid dizisini elde etmek için dört renkli görüntülerin tümü işlenmektedir (Ju et al., 2006; Bentley et al., 2008).

Pyrosekans: Roche'un 454 dizileme teknolojisi ticarileştirildikten sonra, Pyrosekans, yalnızca kuyu başına bir amplifiye DNA molekülü içeren tek bir bead tutacak kadar büyük ve derin milyonlarca mikrokuyu içeren bir picotiter plakasına (PTP) dizilmiş beادلere eklenmiş amplifiye DNA moleküllerini kullanır. Sekans işlemi sırasında, sülfürlaz ve lusiferaz enzimlerinin bağlı olduğu daha küçük beadler yıkanır ve hedef beadleri kaplamak için kuyulara yayılmasına izin verilir. DNA sentezi sırasında, DNA polimeraz bir nükleotid içerdiğinde, bir pirofosfat grubu salınır ve bu reaksiyon, lusiferazı lusiferine fosforile etmek için bu pirofosfatı bir sülfürlaz enzimi tarafından ATP'ye dönüştürmek için pyrosekanslama teknolojisi tarafından kullanılır. Bu döngüde bir nükleotidin spesifik olarak dahil edilmesi nedeniyle lusiferin tarafından üretilen ışık, cihazdaki bir dijital kamera tarafından kaydedilir ve akış

grafiği veya pirogram olarak bilinen bir çıktı üretilir (Ronaghi et al., 1996; Ronaghi et al., 1998; Margulies et al., 2005; Wheeler et al., 2008).

Yarı iletken iyon sekanslama: Yarı iletken sekanslama, DNA sentezi reaksiyonu sırasında salınan hidrojen iyonunun çok hassas bir pH metre (bir mikroçip sensörü) tarafından tespit edilmesine dayanan bir SBS teknolojisidir. Emülsiyon PCR ile şablon amplifikasyonundan sonra, şablona bağlı akrilamid boncukları yarı iletken çipin kuyularına yüklenir. Nükleotidlerin çipten her seferinde bir çevrim akmasına izin verilir. DNA polimeraz bir sonraki tamamlayıcı dNTP'yi içerdiğinde, reaksiyon, katılan nükleotidin trifosfatının hidrolizinden dolayı pirofosfat ve hidrojen üretir. Serbest bırakılan hidrojen iyonu (H⁺), artan proton hassasiyeti sağlayan bir tantal oksit kaplı sensör tarafından tespit edilebilen, belirli nükleotid akış döngüsüne dâhil edilen dNTP'lerin sayısı ile orantılı bir pH değişikliği üretir. Dâhil edilen nükleotit başına 0.02 pH değişimi sensör tarafından kaydedilir ve sonunda bir dizi çıkışına dijitalleştirilen bir voltaj sinyaline dönüştürülür (Rothberg et al., 2011; Merriman, 2012).

Oligonükleotid ligasyonu ile sekanlama ve deteksiyon: Applied Biosystems tarafından ticarileştirildikten sonra, Oligonükleotid Ligasyon ve Tespitiyle Sekanslama (SOLiD), SBS olmayan tek YND teknolojisidir. Adından da anlaşılacağı gibi, SOLiD, sekanslama işlemi sırasında bir DNA ligazı kullanan bir SBL teknolojisidir (Shendure et al., 2005; McKernan et al., 2009).

Kullanılan kimyadaki temel farklılıklara ek olarak, çeşitli YND teknolojileri, onları araştırmacılar ve klinik laboratuvarlar için daha az veya daha çekici kılan birkaç başka açıdan da farklıdır. Örneğin, okuma uzunlukları veya dizileyebildikleri parçaların boyutu, düzinelerce bp'den (SOLiD) başlayıp yüzlerceye (454) kadar değişmektedir. Ek olarak bahsedildiği gibi, YND teknolojilerinin çeşitli derecelerde sekans hatası oranları, YND çeşitlerinden ve ayrıca sinyal algılama eşiklerinden muzdarip olduğu göz önüne alındığında, Sanger bugüne kadar en doğru sıralama teknolojisi olmaya devam etmektedir (Goodwin et al., 2016; Shendure et al., 2017; McCombie et al., 2019). Bununla birlikte, bu dezavantajlar kısa süre içinde aynı genomik bölgenin daha yüksek DNA dizisi kapsamı ile dizilemesi ve aşırı örneklemenin birkaç kat fazlalığına izin vererek, yüksek verim ve bp başına düşen maliyet ile telafi edilmektedir. Sonuç olarak, yanlış pozitif oranı azaltmak için çoğu

bölgede yeterli bir okuma derinliği ve kapsamı elde etmek için YND teknolojilerini kullanımı çok önemlidir. Hedef bölge başına yüksek bir kapsam alanı, Illumina NovaSeq 6000, S4 flowcell gibi mevcut en yeni cihazlarda run başına 20 milyara kadar sıralı okuma ile herhangi bir spesifik bölgede sinyali artırarak hata oranının düşürülmesi amaçlanır. Bir zamanlar geliştirilen ve mevcut olan YND teknolojilerinin çeşitliliğine rağmen, ölçeklenebilir çıktı ve azalan bç sıralama maliyetleri, Illumina'yı araştırma ve klinik ortamlarda en popüler ve yaygın olarak kullanılan YND teknolojisi olarak konumlandırmıştır.

3) Üçüncü Nesil Dizileme Teknolojisi: Üçüncü nesil dizileme teknolojileri, YND teknolojilerinin amplifikasyon aşamasını kaldırıp DNA moleküllerini doğrudan dizilerken uzun okumalar üretmeye odaklanan tek molekül dizileme (SMS) teknolojileri ile karakterize edilmektedirler. En popüler uzun okunan SMS platformları arasında Pacific Biosciences (PacBio) ve Oxford Nanopore bulunmaktadır. SMS ya bir polimeraz (PacBio) eklenmesinden ya da DNA'nın nanometre ölçekli bir gözenek (Oxford Nanopore) yoluyla translokasyonu ile üretilen iyonik sinyaller yoluyla, daha uzun DNA moleküllerindeki tek tek nükleotidleri tanımlar (Goodwin et al., 2016; Shendure et al., 2017; McCombie et al., 2019).

YND teknolojisi, insan DNA'sının çeşitli kısımlarını analiz etmek için uygulanabilmektedir. Hareket bozuklukları ve diğer kalıtsal insan hastalıkları olan hastalarda germ hattı varyantları için klinik tanı amaçlı olarak, üç yaklaşım yaygın olarak kullanılmaktadır (Şekil 2.9) (Rehm et al., 2013; Adams et al., 2018).

2.4.1.1. Hedef Sekanslama: Targeted Sekanslama-TS:

"WET LAB" analizleri aşamasında yalnızca belirli bir hastalık için yaklaşık 20-200 genden oluşan bir panel sekanslama çalışmalarıdır. Bu en uygun maliyetli uygulamadır ve dar bir seçimle genellikle mükemmel okuma derinlikleri elde edilmektedir. Klinik diagnostik kullanımı için önemli dezavantajı, daha sonra tanımlanabilecek olası ilgili genlerin eklenmesine izin vermemesidir.

2.4.1.2. Tüm Ekzom Sekanslama-TES

TES'un tanıtılması ile nadir görülen hastalıkların araştırılması ve tanısında bir paradigma kayması yaratılmıştır (Boycott et al., 2013). Başlangıçta neredeyse yalnızca yeni gen keşfi için kullanılsa da (Beaulieu et al., 2014; Chong et al., 2015), TES artık

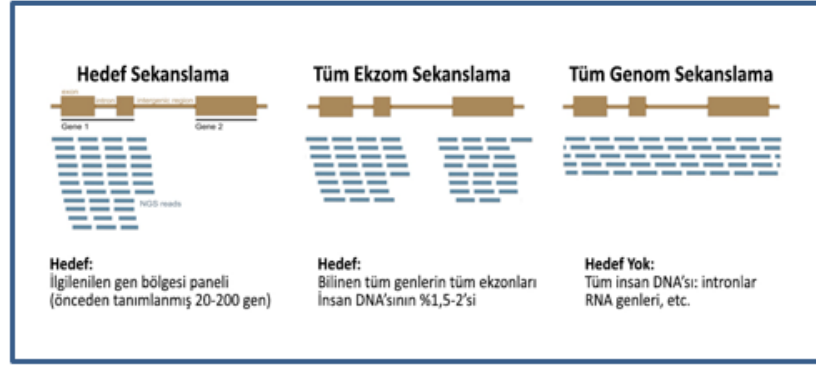
tıbbi genetikçilerin klinik olarak en zorlu NH'a sahip olan hastası için rutinde tercih edilen yaklaşım olarak kabul edilmektedir (Yang et al., 2014; Lee et al., 2014; Farwell et al., 2015; Yavarna et al., 2015; Posey et al., 2016; Retterer et al., 2016).

Tüm insan nükleer DNA'sının (genom) %2'sinden daha azını protein kodlayan ekzonlar temsil eder ve tüm ekzonların toplamına "ekzom" denir. Bunlar şu anda bilinen hastalığa neden olan varyantların yaklaşık %85'ini içermektedirler (Choi et al., 2009; Caspar et al., 2018; Eldomery et al., 2017). Moleküler tanı alan kişiler için sadece tanı serüveni sona ermekle kalmaz, aynı zamanda aile üyeleri için de uygun genetik danışmanlık da sağlanır. Bazı durumlarda hastalığın yönetimi hastalığın doğal seyri bilgisine dayalı olarak şekillendirilir (Boycott et al., 2013; Sawyer et al., 2016).

TES, TGS'den daha az maliyetli olduğu için genetik tanı ve gen keşfi için daha yaygın olarak kullanılmaktadır. Ancak, TES'in başarılı ve faydalı uygulamalara sahip olmasına rağmen genomun yaklaşık ~%2'sinde nükleotit varyasyonu değerlendirilmesi ile sınırlı kalmaktadır (Gonzaga-Jauregui et al., 2012).

2.4.1.3. Tüm Genom Sekanslama-TGS

Hem TS de hem de TES'de, yalnızca TS panelindeki genlerin veya yalnızca ekzonların "yakalandığı: capture" okunan fragmanların dizilenmesinden önce ek bir zenginleştirme adımı gerekmektedir. TGS ise herhangi bir ön seçim (capture) adımı olmaksızın tüm insan DNA'sının %95-98'ini analiz etmektedir (Gorcenco et al., 2020). Dolayısıyla bu kapsamlı dizileme Friedreich ataksisi gibi iyi bilinen hareket bozukluklarına neden olan mutasyonların bulunduğu ve hareket bozuklukları için artan sayıda nedenin tanımlandığı ekzon dışı alanları da kapsamaktadır. Yapılan son çalışmalar ile genel olarak erişkin başlangıçlı ataksinin nispeten yaygın bir nedeni olabilen nöropati ve vestibüler arefleksi sendromu (CANVAS) ile serebellar atakside intronik tekrar genişlemeleri RFC1'i ve resesif erken başlangıçlı serebellar atakside kodlamayan RNA geni RNU12'deki bir varyant içerdiği gösterilmiştir (Lindquist et al., 2013; Elsaid et al., 2017; Tankard et al., 2018; Rafehi et al., 2019). Bu bilgiler TS veya TES ile elde edilemez. Diğer taraftan da TGS, şu anda bilginizin sınırlı olduğu ekzonların dışındaki genetik varyasyonlar hakkında çok fazla bilgi üreten en pahalı ve yoğun veri işleme seçeneğidir.



Şekil 2.9 Klinik tanı amaçlı YND teknolojisinin uygulamaları

Şekil 2.9’da da gösterildiği gibi YND teknolojileri, genomun belirli bölgelerine odaklanabilir. Örneğin belirli bir bozuklukla ilgili olduğu bilinen bir dizi gen (TS) veya bilinen yaklaşık 20.500 genin tümünde protein kodlayan tüm genomik segmentlerin (ekzonlar) toplamı (TES). İnsan hastalığında rol oynadığı bilinen yaklaşık 4.000 genin ekzonlarının seçimi (klinik ekzom) bir ara form olarak kullanılmaktadır (Di Fonzo et al., 2018). Okuma derinlikleri, analiz edilecek bir DNA segmentinin başında ve sonunda incelenmektedir. Buna karşılık TGS, ilgi alanlarını önceden belirlemez. Ekzonlar, intronlar ve örneğin RNA genleri ve biyolojik işlevler için önemi giderek daha fazla tanınan diğer kodlamayan diziler dâhil olmak üzere intergenik bölgeler hakkında veri üretmektedir. Kapsanan genomik alan ne kadar büyük olursa, cihaz ve bilgisayar çalışma süresi ve kaynakları o kadar fazla tüketilir. Ancak teknolojinin hızlı gelişimi, TGS yaklaşımlarını giderek daha uygun maliyetli hale getirmektedir.

2.5. RNA Sekanslama

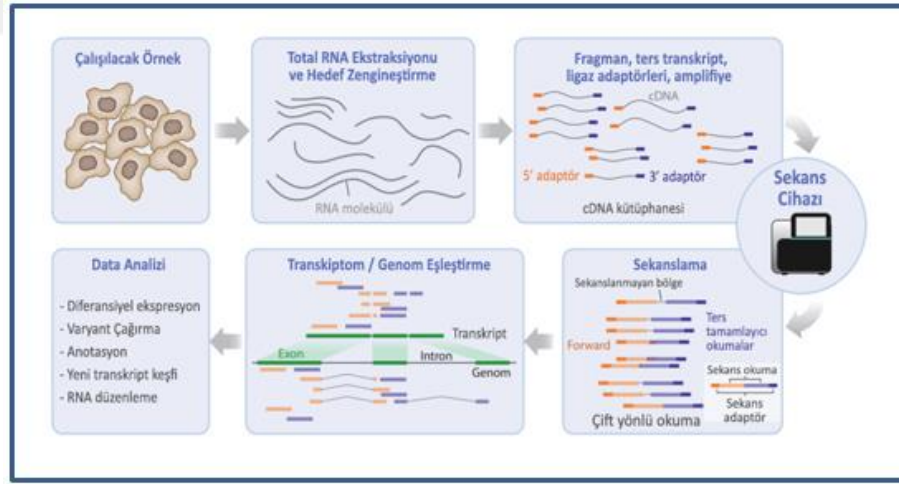
"Transkriptom" terimi ilk olarak 1990'larda, kısa ömürlü ara RNA moleküllerinin incelenmesi yoluyla genomun yapısı ile biyoloji arasındaki boşluğu kapatmak amacıyla tanıtılmıştır (Pietu et al., 1999). Bütünüyle, transkriptom bir hücrede, dokuda veya bireyde hem kodlayan hem de kodlamayan tüm RNA moleküllerinin eksiksiz setini içerir (Morozova et al., 2009). Özellikle kodlayıcı RNA molekülleri, protein sentezlemek için DNA düzeyinde kodlanmış talimatları haberci RNA (mRNA) molekülleri biçiminde ileten ve transkriptomun alt kümesini oluşturan bilgi taşıyıcılarıdır. mRNA moleküler tanısal fayda ve hastalık ilişkisi için daha fazla dikkat çekmiştir. Diğer taraftan, kodlamayan RNA'lar, esas olarak, hastalık

nedenselliği ve tanısında daha az vurgulanan gen transkripsiyonu ve translasyonundaki düzenleyici rollerle ilişkilidir.

İlk transkriptomik analizler, gen keşfi ve gen dizisi belirleme aracı olarak kısa cDNA dizilerini içeren eksprese edilmiş sekans etiket (EST)'ler (Adams et al., 1991) ve transkriptleri benzersiz bir şekilde tanımlamak için birleştirilmiş rastgele transkript fragmanları boyunca kısa dizi etiketleri kullanan Gen İfadesinin Seri Analizi (SAGE) gibi yöntemleri kapsamaktadır (Velculescu et al., 1995). SAGE her bir etiketin tespit edilme sayısının, analiz edilen biyo-örnek/materyaldeki transkriptin ifade seviyelerine karşılık geldiği göz önüne alındığında, ifade seviyesi ölçümüne de izin vermiştir. Transkriptomik analizlerin bu metodolojik öncüleri, tam transkriptlerin splicing olaylarının ve diğer yapısal anlayışların değerlendirilmesine izin vermeyen kısa dizilerin incelenmesiyle engellenmiştir (Wang et al., 2009). Çağdaş transkriptomik tekniklerin geliştirilmesiyle bu sınırlamaların önüne geçilmiştir.

Günümüzde, transkriptom çalışmaları için 2 ana teknoloji olan ekspresyon mikroarray ve RNASeq kullanılmaktadır. Her 2 metotta da diğer hücresel bileşenlerden RNA ekstraksiyonu, ekstrakte edilen moleküllerin zenginleştirilmesi, cDNA üretmek için ters transkripsiyon ve ardından sırasıyla hibridizasyon veya dizileme yoluyla analizler kullanılmaktadır (Bryant and Manning, 1998). Ekspresyon mikroarray ilk olarak 1990'ların ortalarında tanıtılmıştır. Bu teknikde ilgili gen dizisine tamamlayıcı olan "prob" olarak bilinen kısa oligonükleotit oligomerlerin dizildiği katı cam yüzeyler bulunmaktadır (Schena et al., 2005). İlgili numunelerden ve kontrollerden floresan olarak etiketlenmiş cDNA, daha sonra dizinin yüzeyine uygulanarak problemlara karşılık bir şekilde hibridizasyona izin verilir. Hibridize olan molekülleri görselleştirmek, transkript bolluğu ve ekspresyon seviyeleri ile ilişkili floresan yoğunluğu ile test edilen koşullar boyunca transkript bolluğu oluşturmak için bir lazer taraması kullanılmaktadır. Mikroarray, problemlerin tasarımı için ilgili organizmanın genomik dizisi ve eksprese edilmiş genleri hakkında önceden bilgi sahibi olmayı gerektirir. Teknolojik ilerlemeler, neredeyse tüm bilinen genlerin değerlendirilmesine ve mikroarrayler kullanılarak geliştirilmiş floresan algılama yoluyla düşük bolluklu transkriptlerin bile saptanmasına olanak sağlasa da son on yılda neredeyse tamamen YND teknikleri bu metotun yerini almıştır.

RNASeq, transkriptomun analizi için günümüzde kullanılan etkin bir tekniktir. Bu teknikte yüksek verimli YND ve hesaplamalı veri işleme yöntemlerindeki ilerlemelerden yararlanılmaktadır. RNASeq'in ilk aşamaları, RNA ekstraksiyonu ve kalite kontrolünü içeren mevcut farklı metodolojik varyasyonlar arasında tipik olarak benzerdir. Analiz edilmesi istenen RNA türü sınıfının zenginleştirilmesi, tipik olarak belirli bir örnekten toplam izole edilmiş RNA havuzunun sadece %3'ünü temsil eden mRNA molekülleri ile gerekli bir adımdır. Kütüphane yapısı, mRNA türlerinin daha kararlı cDNA'ya ters transkripsiyonu ve ardından kısa cDNA molekülleri üretmek için parçalanma veya daha karmaşık transkriptomik sonuçları bulmak için yakın zamanda tanıtılan uzun okuma (long-read) RNASeq platformlarının kullanımı ile ortaya çıkmıştır (Soneson et al., 2019). RNASeq ile üretilen dizilerin aralığı kullanılan yöntemle bağlı olarak 30bç'den binlerce bç'ne kadar değişebilir. Elde edilen okumalar ya tek yönlü ya da çift yönlü (paired-end) bir şekilde dizilir veya referans genomu hizalanır. Bunlar dışında üretilen okumaların birbiri üzerine eşlenmesiyle bir *de novo* transkriptom düzeneğinde kullanılır (Wang et al., 2009). Standart bir RNASeq deneyindeki temel adımlar genel olarak Şekil 2.10.'da gösterilmiştir.

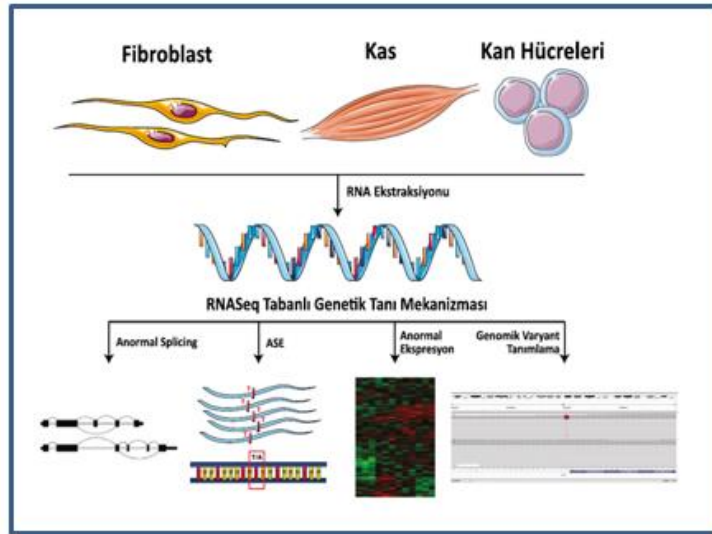


Şekil 2.10 Standart bir RNASeq temel adımları

Şekil 2.10'da gösterildiği gibi standart bir RNASeq çalışmasındaki temel adımlarda cDNA kütüphanesi, izole edilmiş RNA hedeflerinden üretilir ve ardından sekanslanır. Elde edilen okumalar, bir referans genom veya transkriptom ile eşlenir. Down-stream veri analizi, çalışmanın amacına bağlı olarak değerlendirilmektedir (Berge et al., 2019).

Tüm transkriptomu yüksek bir sekanslama derinliğinde sorgulama fırsatına ek olarak, mikroarray için gereken mikrogramların aksine nanogramlar gibi çok daha düşük miktarlarda başlangıç materyali bu yöntem için yeterlidir. Bu da tekniğin tek hücre de bile uygulanmasına izin vermektedir (Hashimshony et al., 2012). Ayrıca sekanslama yaklaşımı hem array teknolojilerinden hem de SAGE'den daha "kantitatif"dir.

RNASeq, belirli bir zaman noktasında, çalışma yapılan hücrelerde ve/veya dokularda bir bireyin tüm transkriptomuna açık platformda ayrıntılı bir görünüm sağlayarak hem RNA türlerini niceleştirme hem de nadir ve yeni RNA transkript varyantlarını saptama fırsatı sunar (Mortazavi et al., 2008). Sonuç olarak, RNASeq, SNV'ler, indeller ve ekspresyon seviyesi dengesizlikleri (eQTL), monoalelik veya alele spesifik ekspresyon (ASE), diferansiyel izoform ekspresyonu ve alternatif splicing gibi anormal transkript düzenlemesine yol açan olayları içeren lezyonların saptanmasını sağlamaktadır (Şekil 2.11).



Şekil 2. 11 Tanı amaçlı RNASeq analizinin şematik gösterimi

Şekil 2.11’de tanı amaçlı RNASeq analizinin şematik gösterimine baktığımızda geleneksel olarak örneklenen üç erişilebilir doku, fibroblastları, kasları ve kanı içermektedir. RNA izolasyonu sonrası NH’da tanısal sonuçlar veren analizler, tipik olarak, splicing sapmalarının, ASE vakalarının, ekspresyon seviyesi dengesizliklerinin ve/veya doğrudan RNASeq yoluyla veya RNA’nın DNA analizleriyle entegrasyonu yoluyla varyant çağırmanın tanımlanmasını içermektedir (Kousi, 2021).

Ökaryotik transkriptom, doku ve hücre tipine özgü ekspresyon, alternatif mRNA izoformları, örtüşen transkriptler ve antisens transkriptlerini içeren dinamik doğası nedeniyle karmaşıktır (Lee et al., 2013). Ökaryotik organizmalarda izoform yapılarının doğru tanımlanması, büyüklükleri birkaç 100bp'den 500.000bp'e kadar değişebilen intronların varlığı nedeniyle zordur (Sakharkar et al., 2004).

Bir gen, genom dizisinden kolayca anlam çıkarılamayan birden fazla izoforma sahip olabileceğinden alternatif splicing bunu karıştırabilir. Bir organizmanın genomunun kopyalanmış kısmını dizilemeye yönelik en eski yöntemler, mRNA'ların çoğunda bulunan poli(A) kuyruklarından yararlanmıştır. Anahtar adımlar, bir RNA transkriptinin 3'-poli(A) kuyruğuna annealing edip cDNA dizisini sentezleyecek poli(T) primerleri ile ters transkripsiyon içeriyordu (Sim et al., 1979; Sutcliffe et al., 1982; Putney et al., 1983; Adams et al., 1991; Parkinson et al., 2009).

Bir RNASeq kütüphane hazırlama aşaması dört adımdan oluşmaktadır.

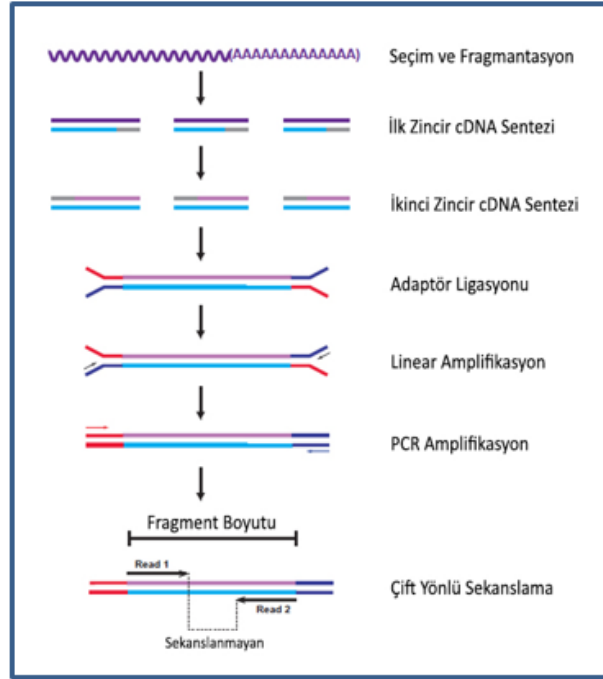
- 1) RNA pürifikasyonu ve boyut seçimi,
- 2) Rastgele fragmentasyon,
- 3) cDNA sentezi,
- 4) Sekans adaptör ligasyonu.

Bu adımlardan bazıları birleştirilebilir, adımların sırası değiştirilerek protokol modifiye edilebilir ve ek RNA seçimi ve deplesyon adımları kullanılabilir. Kütüphane hazırlığı ayrıca her fragment için iplik bilgisini korumak üzere değiştirilebilir (Levin et al., 2010). Kütüphaneyi oluşturmak için kullanılan yöntem, çalışmanın deneysel hedeflerine göre uyarlanmalı ve sıralama verilerinin down-stream analizi için dikkate alınmalıdır.

Bir RNASeq Kütüphanesi hazırlarken ilk adım, poli(A) seçilmiş, poli(A)-depleted veya toplam RNA'nın sıralanacağını belirlemektir (Peng et al., 2012; van Dijk et al., 2014). Poly(A) seçiminde, poli(A) kuyruklu RNA'ları seçmek için bir Poly(T) kolonu veya manyetik boncuklar kullanılmaktadır. Bu da, rRNA dâhil olmak üzere çoğu nonpoli (A) RNA'yı ortadan kaldırmaktadır. Ancak, A açısından zengin bölgelere sahip bazı nonpoly (A) transkriptler de seçilebilmektedir. Poli(A)-depletion, poli(A)-seçiminden sonra hibritlenmemiş RNA'nın toplanmasından oluşurken toplam RNA

herhangi bir zenginleştirme yöntemini kullanmaz. Hem poli(A)-depleted hem de toplam RNA numuneleri için rRNA'lar son derece boldur ve tükenmeleri gerekmektedir. Yaygın yöntemler, dupleks-spesifik nükleaz tedavisi, subtraktive hibridizasyon ve manyetik boncuk bazlı rRNA deplesyon içermektedir (Mortazavi et al., 2008; He et al., 2010; Yi et al., 2011). Bir sonraki adım, protokole bağlı olarak rastgele parçalanma veya cDNA sentezi olabilir. Çift sarmallı cDNA fragmanları oluşturulduktan sonra, dizileme adaptörleri fragmanlara bağlanır, amplifiye edilir ve standart Illumina çift uçlu DNA dizileme protokolü izlenerek sekanslanır. RNA için zincir bilgileri, mevcut protokollerde yapılan değişikliklerle korunabilir ancak ek bir karmaşıklık katmanı gerektirir. Bu da ek bir hata ve bias kaynağı yaratmaktadır (Levin et al., 2010). Tüm kütüphane adımlarını basitleştirilmiş bir protokole yerleştirmek için geliştirilen çeşitli ticari "kitler" bulunmaktadır. Bu kitler, tek hücrelerden doku örneklerine, dizi bilgilerinin korunmasına ve poli(A)-seçimi veya toplam RNASeq'e kadar farklı RNA giriş seviyeleri için optimize edilmiştir.

En yaygın olarak kullanılan kütüphane hazırlama yöntemi Illumina TruSeq protokolüdür. TruSeq protokolü, iki tur poli(A) seçimi ve ardından RNA'yı parçalamak için iki değerlikli katyonlarla 94°C'de inkübasyon ve onu tek iplikli cDNA sentezine hazırlamak için rastgele heksamerler kullanmaktadır. İlk cDNA zincirini oluşturmak için ters transkriptaz kullanılır. İkinci sarmal cDNA, RNA şablonunu parçalamak için RNaz H ile DNA polimeraz kullanılarak sentezlenir. Parçalar daha sonra köreltilerek tek bir 3'-A ile kuyruklanır. Ardından Illumina sekanslama Y adaptörleri, fragmanlar üzerine bağlanır. Parçalar daha sonra Illumina'nın lineer bir PCR amplifikasyon adımı içeren eşleştirilmiş uç dizileme protokolü izlenerek dizilenir.



Şekil 2.12 RNAseq kütüphane hazırlığı

Illumina platformuyla ilişkili sekanslama sorunlarına ek olarak, RNASeq kütüphanelerinin hazırlanmasına özgü, down-stream analizini karıştırabilen ve biyolojik olarak ilgili sinyalleri gizleyen bir dizi sorun bulunmaktadır (van Dijk et al., 2014). Bu sorunlar öncelikle kütüphane zenginleştirme, fragmentasyon ve reverse-transkripsiyon sırasındaki bistan kaynaklanmaktadır (Hansen et al., 2010; van Gurp et al., 2013; van Dijk et al., 2014). Bu önyargılardan bazıları hesaplamalı olarak ele alınabilirken, diğerlerinin düzeltilmesi daha zordur ve herhangi bir biyolojik sonuç çıkarmadan önce düşünülmesi gerekmektedir.

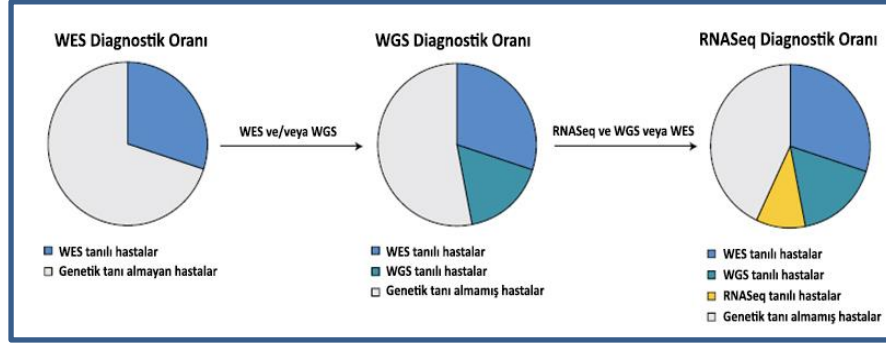
Poli(A)-seçimi, transkriptlerin 3'-ucuna doğru bir parça yanlılığına ve transkriptlerin 5'-uçlarında daha düşük okuma kapsamına yol açabilir (Wang et al., 2009). End-bias, gen ekspresyonu kantifikasyonunu, varyant aramayı ve splice yeri tanımlamasını etkileyebilir (Roberts et al., 2011). Bu, özellikle düşük ifadeli transkriptler için problem yaratmaktadır. Biyolojik gürültü, derin sekanslanmış toplam RNA ve poli(A) depleted örneklerde daha yaygındır ve hatalı olarak ifade edilen veya eklenmiş ek transkriptler içerebilir (Ameur et al., 2011; Haas et al., 2012). Bu kütüphaneler ayrıca, tam olarak eklenmemiş pre-mRNA'lar ve transpozonlar gibi tekrar elementleri nedeniyle daha fazla sayıda intronik diziye sahiptirler. Bu ek

sekanslar, deney ile ilgili olmayabileceğinden düşük ekspresyon seviyelerine sahip transkriptler için duyarlılığı azaltabilirler. Kütüphane seçiminde, kütüphane uygun şekilde DNaz ile muamele edilmemişse, kontamine edici genomik DNA içerebilir. Bu da normalde ifade edilmeyen bölgelerin yanlış pozitif kapsamına girmesine neden olabilir (Haas et al., 2012).

Daha uzun bir transkript, daha küçük bir transkriptten daha fazla potansiyel fragmana sahip olacağından, kütüphane fragmentasyonu daha uzun okumalara yönelik bir okuma yanlışlığına yol açmaktadır (Hansen et al., 2010; Roberts et al., 2011). Bu da benzer ekspresyon seviyelerine sahip daha küçük bir transkriptten daha fazla dizili çiftte sahip uzun bir transkript ile sonuçlanmaktadır. Ters transkriptaz, hata düzeltme değildir ve yüksek kaliteli DNA polimerazlardan daha yüksek bir hata oranına sahiptir (Menéndez-Arias, 2002).

Yukarıda bahsedilenler sekanslama reaksiyonu sırasında spektral örtüşmeden kaynaklanmayan gerçek uyumsuzluk hataları olduklarından yüksek baz niteliklerine sahip olabilirler. Bu da DNA kütüphanelerine kıyasla yanlış pozitif uyumsuzluk çağrılarının artmasına neden olabilmektedir. Sonuç olarak, özellikle RNA düzenlenmesinin ve varyantlarının bi-identifikasyonu için sorun oluşmaktadır. RT, kimerik transkriptlerin üretilmesine yol açan yapılandırılmış bölgelerin ters transkripsiyonu sırasında moleküller arası veya moleküller arası olarak şablonları değiştirebilir (Cocquet, 2006; Houseley and Tollervey, 2010). Bu transkriptler, yanlış pozitif ekleme olaylarının ve gen füzyonlarının tanımlanmasına neden olabilirler. Bunlara ek olarak rastgele heksamer birinci iplik sentezi hazırlama, dizileme okumalarının başlangıcında nükleotit bileşiminde yanlışlıklara yol açabilirler (Hansen et al., 2010; van Gurp et al., 2013; Toungh et al., 2014).

RNASeq'in, TES ve TGS için tamamlayıcı bir araç olarak kullanıldığında, RNA transkript ekspresyonu üzerinde VUS'ların doğrudan işlevsel bir değerlendirmesine izin verdiği için, nadir görülen genetik bozuklukların moleküler tanısını kolaylaştırdığı bildirilmiştir (Cummings et al., 2017; Kremer et al., 2017; Gonorazky et al., 2019). (Şekil 2.13) (Kousi, 2021).



Şekil 2.13 Ekzom, genom tabanlı entegre genomik ve transkriptomik yaklaşımların uygulanması üzerine tanısal verim

Genel olarak, genomik ve transkriptomik yaklaşımların tamamlayıcı kullanımının, ilgili doku erişilebilirliği tarafından ve buna bağlı olarak sınırlı tanısal çözünürlük oranıyla, farklı bozuklukların bir spektrumunda %10-35 oranında artan moleküler tanı verimi ile sonuçlandığı bildirilmiştir (McKean et al., 2016; Cummings et al., 2017; Kremer et al., 2017; Gonorazky et al., 2019).

RNASeq'na dayalı genetik tanının altında yatan mekanizmalara baktığımızda, TES ve/veya TGS'ye dayalı olarak tanımlanamayan gerçek değişiklikleri olan hasta kohortlarında moleküler tanıları sağlamak için RNASeq kullanan çalışmalar ile önceden yorumlanamayan varyasyon veya anlaşılması zor düzenleyici etkilerin ortaya çıkan fenotiple sonuçlandığı dört ana mekanizma bildirilmiştir. Bu mekanizmalar hem bilinen hem de potansiyel olarak yeni aday hastalık genlerini etkileyerek, hastalık genomiği ve hastalığa katkıda bulunan varyant türleri ve alelleri hakkındaki anlayışımızı genişletmiştir.

1. Anormal Splicing: Pre-mRNA eklemesindeki alterasyonlar, uzun zamandır Mendel Hastalıklarının ana nedeni olarak kabul edilmiştir (Tazi et al., 2009; Singh and Cooper, 2012). Bu mekanizmanın hastalıkla ilgisi, beta-globin (HBB) genindeki mutasyonlardan kaynaklanan ve β -globin ekspresyonunun tam (β^0) veya kısmi (β^1) kaybına neden olan genetik olarak heterojen bir hastalık grubu olan hemoglobinopatilerdeki seminal çalışma ile vurgulanmıştır (Weatherall and Clegg, 2008). Bu çalışmalar, heterolog sistemlerde mutant transkriptlerin ifadesine dayandığı için bu tür analizlerin gerçekleştirilebileceği ölçek sınırlıydı. Ancak hasta ve kontrol kohortlarının karşılaştırmalı analizlerine dayanan YND ve hesaplama araçlarının ortaya çıkmasıyla yüksek verimli analizler mümkün hale gelmiştir. Bu gibi

durumlarda, anormal şekilde eklenmiş transkriptlerin tanımlanması, hasta numunesinin/örneklerinin Genotip-Doku Ekspresyonu (GTEx) Konsorsiyum Projesi tarafından açıklananlar gibi bir dizi kontrol bireyi veya veri tabanı ile karşılaştırılmasına bağlıdır (GTEx Consortium, 2015). Bazı tahminlere göre, SNV'lere neden olan hastalıkların toplam %62'si, bunların %25'i görünüşte iyi huylu değişiklikler olmakla birlikte normal splicing'i bozabilmektedirler (Lo'pez-Bigas et al., 2005; Pagani et al., 2005). Bununla birlikte, bu VUS'lar, splicing bağlantılarının konsensüs bağlantı yeri donör GT ve alıcı AG dizilerinin ötesinde uzandığında, bunların splicing üzerindeki etkileri şu anda mevcut biyoinformatik araçlarla güvenle tahmin edilemez. Eşanlımlı değişiklikleri temsil eden bu varyantların birçoğu, derin intronik dizilerde bulunurlar veya etkilerini diğer karmaşık cis düzenleyici unsurlar aracılığıyla uygularlar. Tanısal splicing defektleri belirlemede RNASeq'in faydası nadir kas bozuklukları olan 50 kişiden oluşan bir kohort çalışmasında, splicing-defekt mutasyonların tanı oranını %35 artırdığı şeklinde vurgulanmıştır (Cummings et al., 2017). LoF ve GoF alellerinin normal transkript eklemesini etkilediği bugüne kadar tanımlanan ana mekanizmalar, Intronik Gain, Ekzonik uzantı, Ekzon atlama, Tekrarlama bölgelerinde splicing kesintileridir. Distrofin genindeki derin intronik varyantların, DMD olan bir hastada mevcut olduğunda erken sonlandırmaya yol açan bir psödo-ekzon oluşumuyla sonuçlandığı gösterilmiştir.

2. Alel Spesifik ekspresyon (ASE): Alelik dengesizlik veya monoalelik ekspresyon olarak da adlandırılan ASE, belirli bir genetik lokusta iki ayrı alel veya haplotipin ekspresyon farkının ölçülmesinden sonra kurulmaktadır (Lappalainen et al., 2013). Alel dengesizliği, bir alelden transkripsiyon seçici olarak tercih edildiğinde veya baskılandığında veya transkriptler NMD gibi mekanizmalar yoluyla translasyon sonrası bozulmaya uğradığında meydana gelir. Monoalelik ekspresyonun en belirgin iki örneği, dişilerde X bağlantılı genlerin çift dozajının, iki X kromozomundan birinin rastgele seçilmesi ve susturulması ile açıklandığı X kromozomu inaktivasyonu (Lyon, 1986) ve genlerin/alellerin ebeveyninden türetilmiş, yani menşe ebeveyni, seçici bir şekilde ifade edildiği genomik imprintingdir (Lee and Bartolomei 2013; Peters, 2014). Bu daha iyi karakterize edilmiş mekanizmalara ek olarak, ASE ayrıca otozomal genler için CNV'lerin, cis düzenleyici elementlerin ve eQTL'ların (Fu et al., 2015) nadir değişikliklerinin, long ncRNA'ların (Lee and Bartolomei 2013) ve alel spesifik

kromatin işaretlerinin varlığını içeren ancak bunlarla sınırlı olmayan bir dizi mekanizma yoluyla rapor edilmiştir (Chotalia et al., 2009).

3. Anormal Ekspresyon: Ters nedensellik yaklaşımında, RNASeq, vaka örneklerinde aşırı ekspresyon değişikliklerinin tanımlanmasını kolaylaştırabilir. Böyle bir senaryo altında, ekspresyonu beklenenden önemli ölçüde yüksek veya düşük olan ve dolayısıyla potansiyel olarak hastalık adaylarını temsil eden genleri incelemek için hastaların gen ekspresyon seviyeleri GTEx'te bulunanlar gibi kontrol bireylerinininkiyle karşılaştırılır. Bu adaylar, ifade aykırı değerler olarak adlandırılır ve genellikle ifade üzerindeki etkilerinin tespit edilmesi zor olan nadir kodlamayan varyantlarla ilişkilendirilir. Bugüne kadar, bir dizi rapor, nadir görülen kodlamayan değişiklikler ile tek dokulardaki aşırı ekspresyon değişiklikleri arasındaki bu nedensel bağlantıyı örneklemiştir (Montgomery et al., 2011; Zhao et al., 2016). Daha karmaşık durumlarda, ifade aykırı değerlere yol açan kesin varyantlar tanımlanmadan kalır ve anormal ifade tarafından ortaya çıkarılan hastalık nedenselliğini doğrulamak için sıklıkla proteomik veya metabolomik gibi diğer omik yöntemleriyle kesişmeyi içeren diğer yaklaşımların uygulanmasını gerektirir. Bu TES analizinin makul aday varyantları ortaya koymadığı, kas hipotonisi, gelişme geriliği ve nörolojik gerileme olan üç hastanın raporuyla gösterilmiştir (Kremer et al., 2017). RNASeq, bir hastada MGST1'in ve kalan ikisinde de TIMMDC1'in azalmış ekspresyonunu ortaya çıkarmıştır. Bu da sırasıyla MGST1 ekspresyonunun yalnızca %2'sini gösteren ve saptanabilir TIMMDC1 izi olmayan nicel proteomikler kullanılarak daha da doğrulanarak bu bireyler için genetik tanıyı ortaya koymuştur. Ayrıca aday terapötik hedefler için ipuçları sağlamıştır (Kremer et al., 2017).

4. Genomik varyantların tanımlanması: Belirtildiği gibi, RNASeq tipik olarak TES ve/veya TGS yoluyla genomik DNaseq için tamamlayıcı bir araç olarak kullanılır. Uygulamada RNAseq, genotipi kombinatoriyal olarak deşifre etmek ve aynı zamanda ekspresyon seviyeleri ve splicing hakkında bilgi sağlamak için kullanılabilir. Bu yaklaşımın dezavantajı, DNA dizilemesinden farklı olarak, transkriptomik analizler tipik olarak düzensiz kapsama, değişken transkript ifadesi, alternatif izoformların tespiti yoluyla kapsamlı olmayan ekzonik inklüzyonlar ve birkaç hususu belirtmek için RNA düzenleme olayları tarafından engellenmesidir

(Byron et al., 2016). Bir hasta kohortunda RNASeq ve TGS arasında sistematik bir karşılaştırılmada tek başına RNASeq'in TGS aracılığıyla tanımlanan varyantların yalnızca %90'ını kurtarabildiği ortaya konulmuştur. Bu sonuç bize tespit edilemeyen varyantların ya filtrelendiğini ya da kapsamı zayıf olan/yok olan ya da eksprese edilen izoformlarda ekzon bulunmayan alanlarda bulunduğunu göstermektedir (Gonorazky et al., 2019). Cornelia de Lange sendromu klinik tanısı ile başvuran hastalarda yapılan bir çalışmada TES için benzer sonuçlar bildirilmiştir (Rentas et al., 2020). TGS ile tanımlanan VUS miktarının aksine, RNASeq, önceden mRNA'ya çevrilmemiş bölgelerin (UTR) dâhil edilmesi yoluyla TES ile karşılaştırıldığında, dikkate alınacak aday varyasyonunu yalnızca biraz artırmaktadır. Yapılan bir çalışmada RNASeq aracılığıyla tespit edilen varyantların %2'sinin TGS'de görülmediğinin bildirilmesi dikkat çekicidir. Bu değişiklikler ya yanlış keşifleri ya da genetik bozukluklarda olası bir nedensel rolü olan olası somatik mozaik varyasyonu temsil ederek daha karmaşık mekanizmaları incelememize izin veren yeni dizileme teknolojilerinin geliştirilmesiyle giderek daha fazla artmaktadır (Rentas et al., 2020).

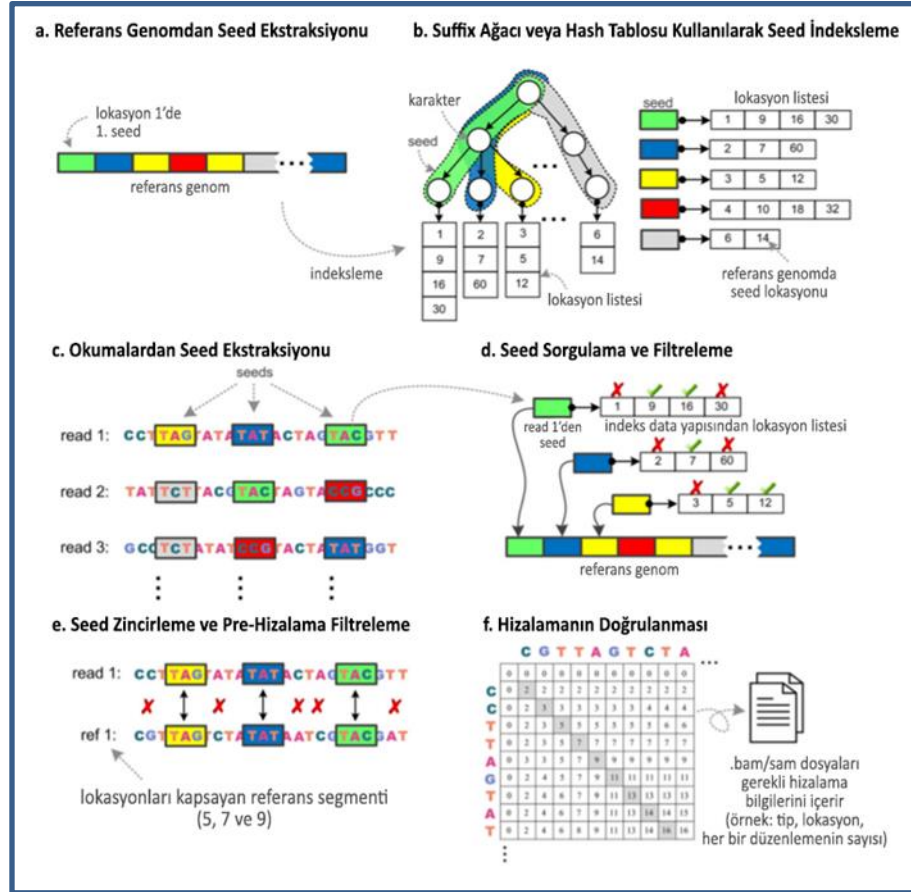
RNASeq'in tanısal faydasına ek olarak, raporlar, hastalık modifiye edicileri bulmak için çıkarılabilen primer lokus disfonksiyonuna yanıt olarak meydana gelen genom çapında ekspresyon değişikliklerinin ortaya çıkarılmasında RNASeq kullanımının altı çizilmiştir. RNASeq'in bu özelliğini örnekleyen bir çalışma alışılmadık derecede hafif bir fenotip sergileyen Golden Retriever'larda kendiliğinden oluşan DMD soyağacından iki "escaper" köpeğin transkriptomik analizini içermektedir (Kornegay et al., 1988; Vieira et al., 2015;). Özellikle "escaper" köpekler, tipik olarak etkilenen köpeklerde erken ilerleyici kas dejenerasyonu, fibroz, kas atrofisi, kontraktürler ve 2 yaşına kadar erken ölümle sonuçlanan distrofinin tamamen yokluğuna rağmen fonksiyonel kas ve normal yaşam süresi sergilemişlerdir (Vieira et al., 2015). Wt, etkilenmiş ve "escaper" köpekler arasında, aday transkripsiyonel dengesizliklere yol açan varyantların tanımlanması için genomik analizle bütünleşmiş karşılaştırmalı transkriptomik analiz, Jagged 1'i iki "escaper"da daha hafif fenotipin altında yatan bir aday değiştirici gen olarak tanımlayarak yeni aday terapötik yaklaşımlar için bir öngörü sağlamıştır (Vieira et al., 2015). Huntington hastalığı (HD, MIM #143100) olan 4000 hastanın transkriptomik analizini bildiren yeni bir çalışma, 48 dokuda, DNA onarımı ile ilişkili PMS2 geni gibi hastalık modifiye edicileri ve bununla birlikte bir yolak içindeki birkaç gen arasında ekspresyonun kombinatoryal

düzensizliğini ortaya çıkarmıştır. Bu çalışmada bildirilen böyle bir ortak ifade modülü, bilişsel gerileme ve klinik başlangıç ile ilişkili bir kortikal gen ifade setinin tanımlanmasını içerdiğinden, hastalık modülasyonu ve tedavisini hedefleyen gelecekteki yaklaşımlar için aday bir hedef yol sunmaktadır (Wright et al., 2020).

2.6. Yeni Nesil Dizileme Data Analizi

Sekanslama sonrasında, kullanılan platform tarafından tespit edilen sinyaller, sekans cihazı tarafından kullanılan kimyasala bağlı olarak işlenir ve down-stream analizleri için bir dizi A, C, G ve T gibi uygun bir dizi temsiline dönüştürülür. Sekanslama sonrası veri analizinin ana adımları aşağıdaki gibidir:

- 1. Haritalama ve hizalama (Mapping ve alignment):** Farklı teknolojiler kullanılarak üretilen veriler, sekanslanan (dizi okumaları olarak isimlendirilir) ayrı ayrı DNA parçalarının her birinin dizisini ve her bir okuma için bazı kodlanmış kalite bilgilerini içeren dosyalara aktarılır. Bir sonraki adım genellikle, bu dizi okumalarının her birinin, dizi okumalarının ait olduğu genomdaki kesin ve benzersiz konumları bulmak için bir şablon olarak kullanılan insan genomu referans dizisine eşlenmesini ve hizalanmasını içermektedir. Bir referans olmadan kısa okuma dizilerinden bir genomun tamamını bir araya getirememesi nedeniyle, kullanılan referans genomla eşleşmeyen tüm bireysel dizileme okumaları genellikle bu adımda çift ve düşük kaliteli okumalarla birlikte atılır. Uzun okunan SMS teknolojilerinde, bu adım atlanabilir ve *de novo* assemble için diğer biyoinformatik yöntemlerle subzisyon edilebilir (Wee et al., 2019). Sekans okumalarını referans genoma eşleştirmek için kullanılabilecek yayınlanmış farklı hizalama araçları mevcuttur. Bu hizalama araçları, okuma hizalamasının doğruluğunu ve hızını artırmak için çeşitli algoritmalar kullanırlar (Şekil 2.14).

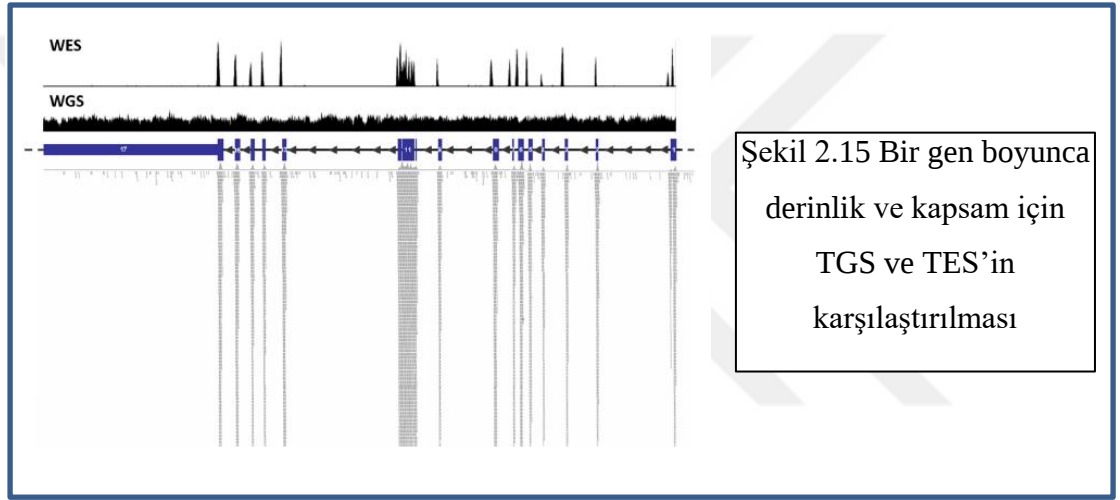


Şekil 2.14 Okuma hizalama algoritmasına genel bakış

Okuma hizalama algoritmasına genel olarak bakıldığında. **a.** Referans genom dizisinden seed çıkarılır. **b.** Her ekstrakte edilmiş seed ve referans genomdaki tüm oluşum lokasyonları, tercih edilen veri yapısı kullanılarak saklanır (suffix ağacı ve hash tablosu örnek olarak sunulmuştur). Seeds ortak önekleri, ek ağacının dallarında bir kez saklanır, hash tablosunda ise her bir seed ayrı ayrı saklanır. **c.** Her okuma dizisinden seeds çıkarılır. **d.** Referans genomda ekstrakte edilen her seed oluşumu indeks veri tabanı sorgulanarak belirlenir. Bu örnekte, 1. okumadaki üç seed, referans genomda 5, 7 ve 9 konumlarında bitişik görünürler. Aynı seedsden ikisi, başka iki yerde de bitişik görünmektedir (12 ve 16). Bitişik olmayan diğer konumlar, ilk okumayla iyi bir eşleşme sağlayamayacağından filtrelendir (X ile işaretlenir). **e.** Bitişik seeds, boşluklar arasındaki uyumsuzluklar incelenerek daha uzun bir seed zinciri oluşturmak için birbirine bağlanırlar. Hesaplama açısından pahalı DP hesaplamasının gerekli olup olmadığına hızlı bir şekilde karar vermek için pre-alignment filtreler de uygulanabilir. **f.** pre-alignment filtresi, referans genomdaki bir okuma ile bir bölge

arasındaki hizalamayı kabul ettiğinde, daha sonra DP tabanlı (veya DP tabanlı olmayan) doğrulama algoritmaları, farklılıkların tam sayısı, her bir farklılığın konumu ve türleri gibi hizalama bilgilerini içeren hizalama dosyasını (BAM veya SAM formatlarında) oluşturmak için kullanılırlar (Alser et al., 2021).

TGS ve TES için haritalama ve hizalamayı karşılaştırırken, her iki dizileme yaklaşımının kapsamı ve derinliğindeki farklılıkları açıkça görebiliriz (Şekil 2.15). TGS için sekans okumaları, kodlayan ve kodlamayan diziler de dâhil olmak üzere ve bunlar arasında daha eşit kapsama sağlayarak genom boyunca eşlenir.

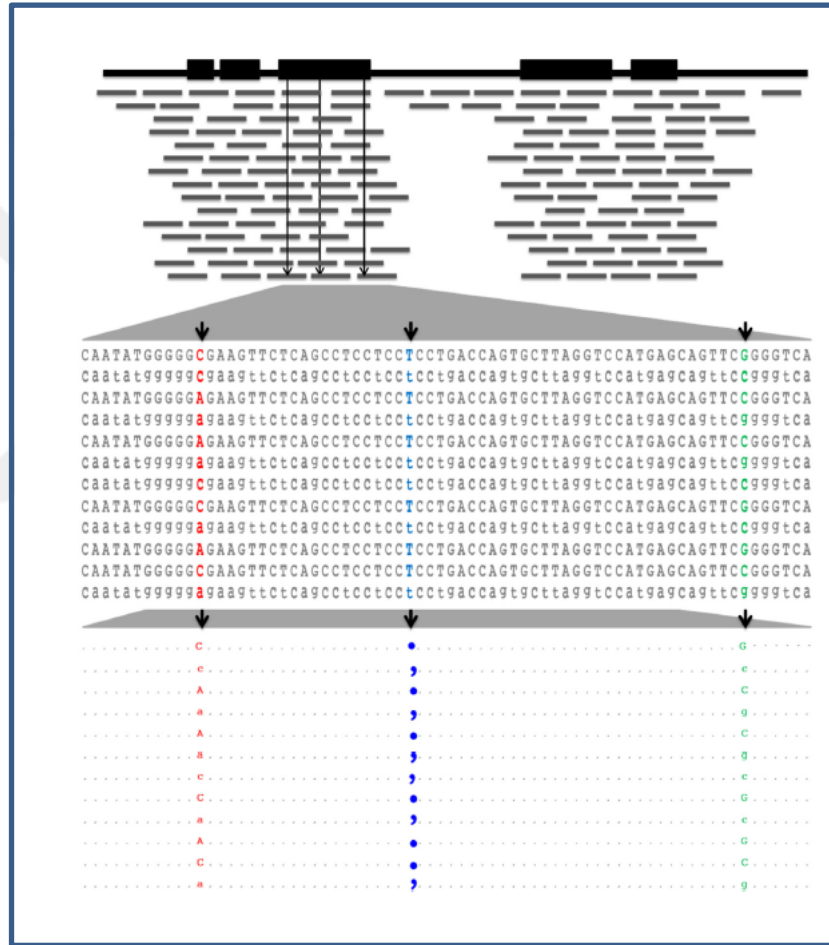


Şekil 2.15’de gösterildiği gibi bir gen boyunca derinlik ve kapsam için TGS ve TES farklıdır. Kapsam, TGS’de genom genelinde daha eşit olsa da okuma derinliği genellikle daha azdır (yaklaşık 30X). TES içinse kısa sıralı okumalar (alt paneldeki gri çizgiler) yalnızca hedeflenen bölgeler, özellikle ekzonlar üzerinde çok daha fazla okuma derinliğinde (yaklaşık 100X) birikmektedir (Gonzaga-Jauregui and Mendoza, 2021).

TES'in hedef doğası gereği, hedeflenen capture bölgesi için kullanılan reaktifte dayalı olarak yalnızca ekzonlara ve yakalanan genomik bölgelere okunan dizinin haritalanması ve hizalanması ile sekans sonuçlanacaktır. Kapsam, TGS'deki genom boyunca daha eşit olduğundan, maliyeti telafi etmek için derinlik hesaba katılır. Bu nedenle, TGS genellikle ortalama 30X okuma derinliğinde yapılacaktır. Yani ortalama olarak genomdaki her baz 30 kez okunmuş olacak veya 30 farklı kısa okuma ile

kapsanacaktır. TES ise daha küçük bir genomik alana odaklandığı için yaklaşık 100X veya daha yüksek ortalama okuma derinliğinde gerçekleştirilmektedir.

2) **Varyant ve genotip çağırma:** Hizalama dosyasından, genomdaki her bir baz hakkında bilgi çıkarılabilir ve bir yığın genomik dosyasına rapor edilebilir. Bu dosya, genomdaki her pozisyon için, o belirli pozisyondaki tüm okumaların hizalanması veya “yığılması: piling up” ile gözlemlenen nükleotidi belgelemektedir (Şekil 2.16).



Şekil 2.16 Genomik projelerde haritalama, hizalama, yığın ve varyant çağırma işlemleri

Şekil 2.16’da gösterildiği gibi genomik projelerde haritalama, hizalama, yığın ve varyant çağırma işlemlerinde bireysel okumalar referans insan genom dizisiyle eşleştirildikten sonra, bu okumaların bir yığını oluşturulur ve genomdaki her hizalanmış pozisyonda her baz rapor edilir. Bu verilerin ve oluşturulan dosyaların işlenmesini kolaylaştırmak amacıyla, semboller, forward-strand (.) bildirilen bir

referans tabanını, reverse strand (,) bir referans tabanını temsil etmek için ve sırasıyla forward veya reverse zincirde belirli varyant çağrılarını temsil etmek için büyük ve küçük harfler atanmıştır (Gonzaga-Jauregui and Mendoza, 2021).

Genomdaki nükleotitlerin çoğu referansla aynı olacaktır. Genomik analizler için, hastaların veya test numunelerinin referans diziden farklı olacağı ile ilgili önemli bir varsayım bulunmaktadır. Bu nedenle, down-stream analizlerinin ana odak noktası, değişken olan pozisyonlar veya bölgelerdir. "Yığın" elde edildikten sonra, değişken pozisyonlar başka bir dosyaya çıkarılır ve buna varyant çağırma denir. Bu adımda varyantın genotipi de değerlendirilir. YND verilerinden varyant ve genotip çağırısı yapan birkaç algoritma bulunmaktadır (Shen et al., 2010; McKenna et al., 2010; Nielsen et al., 2011). Bu "caller" tarafından oluşturulan ve referans dizisinden farklı konumları içeren dosya Varyant Çağrı Formatı (VCF: Variant Calling Format) dosyası olarak bilinmektedir. Varyantlar "çağrıldığında", çoğunluğun gerçek pozitifler olduğundan emin olmak için, dizileme hatalarını ve düşük kaliteli aramaları filtreleyerek kalitelerini, konum genelinde okuma sayısını (kapsam), varyant olarak adlandırılan okuma sayısını, zincir yanlılığını (strand bias), genotip kalitesini ve gerçek bir varyant olma olasılığını değerlendirmek önemlidir.

3) Varyant anotasyonu: Klinik bir ortamda ekzom veya TGS kullanılırken ortaya çıkan bir zorluk, bu teknolojiler tarafından keşfedilen çok sayıda varyantın olmasıdır. Örneğin, ortalama her ekzomda 20.000'den fazla varyant bulunmaktadır. Tipik olarak bu varyantların 10.000-12.000'i, bir proteinin AA dizilimini değiştirerek eş anlamlı olmayacaktır. 120-tanesi protein kırpma olarak (truncating) ve varyantların 54'ü de daha önce patojenik olarak rapor edilmiş olacaktır. Bu nedenle, daha yönetilebilir bir liste oluşturmak için bu değişken kümesini filtrelemek ve önceliklendirmek çok önemlidir (Consortium 1000 Genomes Project 2015; Lek et al., 2016). Varyant anotasyonu, farklı pipeline ve farklı tanı laboratuvarları arasında çok daha fazla değişkenliğin olduğu bir alandır. Şu anda, varyant açıklamasını gerçekleştirmek için kullanılan iki ana araç olan Annovar (Wang et al., 2013) ve VEP (McLaren et al., 2016) bulunmaktadır. Varyant anotasyonunun ana hedeflerinden biri varyantın genomik bağlamını belirlemektir. Anotasyon için

kullanılan araç, her varyant için bilinen bir transkript veya bir promotör veya ekleme bölgesi gibi diğer genomik özelliklerle örtüşüp örtüşmediğini bildirir.

4) Varyant veri yorumlaması: Genomik veri analizinin son ve muhtemelen en önemli ve zaman alıcı adımı, varyantın yorumlamasıdır. Bu adımda, dizileme ve açıklama yoluyla toplanan genomik varyantlar ve bilgiler, genellikle ek biyoinformatik algoritmalar ve çalışma türüne özel analiz iş akışları tarafından desteklenen, belirli bir varyantın veya varyantların uygunluğunu değerlendirecek olan uzman bir genomist tarafından, incelenen fenotip ile ilişkili olarak, analiz edilir ve yorumlanır. Bu süreç, tipik olarak, belirli bir varyantın patojenik olduğuna dair yeterli kanıt olup olmadığını belirlemek için ACMG yönergelerinin (Richards et al., 2015) veya diğer laboratuvara özgü stratejilerin uygulanmasını içermektedir. Örneğin, varyantın protein fonksiyonunu etkileme olasılığının olup olmadığını, popülasyonda nadir olup olmadığını ve hastanın fenotipiyle tutarlı bir gende olup olmadığını göz önünde bulundurularak analizler gerçekleştirilir.

2.6.1. RNASeq Data Analizi

RNASeq teknolojisi ile biyolojik bir örnekten mRNA'yı alınıp bir referans genomla eşlenir (Wang et al., 2009). Ortaya çıkan veriler, dizileri her numunedeki her bir gene eşlenen bir dizi mRNA fragmanından oluşan “reads”dir. Bu okuma sayıları, her bir biyolojik numunedeki genom genlerinin ekspresyon seviyelerinin ölçümlerini temsil eder.

RNA, DNA'ya benzer şekilde sekanslanabilir. mRNA hücreden alınır ve sekanslanan cDNA'yı (tamamlayıcı DNA) oluşturmak için bir şablon iplikçik olarak kullanılır. RNASeq büyük bir kısmını aynı anda birçok hücreden alınan RNA oluşturmaktadır (bulk RNASeq). Dokular ayrılarak RNASeq genellikle dokuya özgü ifadeyi incelemek için kullanılır. Yapılan son gelişmeler ile aynı zamanda RNA'nın tek bir hücreden de sekanslanması da mümkün hale gelmiştir. İnsanlarda genler intronlar ve ekzonlar içerdiğinden ve intronlar mRNA oluşturmak için eklendiğinden, sıralı mRNA'yı insan genomuna geri hizalarken mutlaka tek parça halinde hizalanmayacaktır. RNASeq yalnızca ekzonik bölgelere hizalanır ve ekzonlar arasındaki birleştirme sınırını kapsayan okumalar, aralarında bir intron olacak şekilde, bir ekzonun yarısına, bir sonrakinin başlangıcının yarısına, bölünmüş bir şekilde

sürekli olmayan bir şekilde hizalanacaktır. Bunun için, STAR (Dobin et al., 2013), TopHat2 (Kim et al., 2013) ve HISAT2 (Kim et al., 2019) dâhil olmak üzere RNAseq verileri için özelleştirilmiş splice hizalayıcılar kullanılmaktadır. Bu hizalamaların ekzonlara bölünmesinin "iyi" hizalamalar olması gerekmektedir. Bu hizalayıcılar, okuma bölünmelerinin beklendiği yerler için ek bir ön bilgi sağlamak için referans genom üzerindeki gen anotasyon bilgilerini alabilirken aynı zamanda ekzonların anote olmadığı yeni splits de bulabilirler (Sherman, 2021).

Ökaryotik transkriptom içindeki gen yapısı ve fonksiyon ilişkilerinin anlaşılmasındaki zorluklardan biri, RNA düzenleme, alternatif splicing ve poliadenilasyonu içeren yüksek düzeyde transkript işlenmesidir (Thomson, 2006; Halbeisen et al., 2007; Shi, 2012). Bir örnek içinde hangi RNA'ların bulunduğunu ve bunların transkripsiyon sonrası değişikliklerinin küresel ölçekte kapsamını belirlemeye ve ölçmeye ihtiyaç vardır.

Genel olarak ideal RNASeq çalışması:

- 1) Bir genin genel ifadesini kantifiye eder,
- 2) Gen içinde kodlanan RNA izoformlarından anlam çıkarır,
- 3) Her bir izoformun ifadesini ölçer,
- 4) Eksprese edilmiş varyantları, RNA editing ve diğer transkripsiyon

sonrası değişiklikleri tanımlar.

Sekanslama teknolojisi, RNA'nın tüm yönlerini belirlemek için uygun bir yöntemdir. Bir hücrenin toplam RNA içeriğini nicel olarak dizileyerek gen ekspresyonunu, izoform yapılarını ve izoform ekspresyonunu hesaplamak mümkündür (Wang et al., 2008; Trapnell et al., 2010). Dizileme inovasyonunun çoğu, DNaseq üzerine odaklanmıştır. Ancak 1970'de RNA'yı cDNA'ya reverse etmek için hareket eden RT keşfi ile DNA dizileme teknolojisini RNA'ya uygulamak için bir araç sağlanmış olundu (Temin and Mizutani 1970: Sim et al., 1979).

Genel olarak analiz aşamasında, işlenen verilerin yanı sıra ham dizilerin kalite kontrolü (QC), herhangi bir çalışmanın bütünlüğünü sağlamak için esastır. İdeal olarak, sekanslama verilerinin QC'si, sekanslama merkezinin kendisi tarafından gerçekleştirilir. Böylece orijinal RNA örneği, kütüphane hazırlığı veya sekanslama ile ilgili herhangi bir düzensizlik hızlı bir şekilde tespit edilip düzeltilebilir. Sekans

sonrası hemen kontrol edilecekler standart sıralama QC metrikleri, okuma uzunluğu, okumaların baz kalite dağılımı ve nükleotid bileşimi, PCR kopyalarının yüzdesi, ortalama GC içeriği, okuma sekans hatalarının yüzdesi ve aşırı temsil edilen sekansların tespiti. Bu metriklerin çoğu, orijinal okumaların temsili bir örneğini alarak popüler FastQC tool (Ewels et al., 2016) kullanılarak hesaplanabilir. Belirlenen QC kriterlerini karşılayan çalıştırmalar sonraki işlemler için kullanılmalıdır.

Sıralama verilerinin QC'sinden sonraki ilk adım, kısa okumaların kaynaklandığı genomik lokalizasyonu bulmaktır. İncelenen türün yüksek kaliteli bir referans genomu varsa RNASeq okumaları bu referansa eşlenebilir. Ökaryot genomları için poli-A ile zenginleştirilmiş RNA moleküllerinin çoğu, zaten eklemeye geçirmiş olan olgun transkriptlere karşılık gelmektedirler. Sekanslamadan sonra bazı okumalar tamamen bir ekzon içine düşerken diğerleri ekzon-ekzon sınırlarını kapsayacaktır. Öyleyse intronik dizilere sahip bir genome karşı haritalama, boşluklu hizalamaları işleyebilen bir haritalama aracı gerektirir. Bu boşluklar, araştırılan türün intronlarının uzunluğunu kapsayacaktır (örneğin, insan intronları için 5bp ile 1Mb). Çoğu RNASeq eşleyicisi (Langmead et al., 2009), önce kısa okumaları bitişik olarak referans genome hizalamaya çalışan ve ardından eşlenmemiş okumaları bir ek bağlantı veri tabanına karşı eşlemek için rastgele daha küçük bölümlere ayıran DNA tabanlı hizalayıcıların uzantıdır. Ancak STAR (Dobin et al., 2013) bitişik olmayan dizileri doğrudan referans genomuna hizalamak için özel olarak tasarlanmıştır. Bu tez kapsamında STAR tool kullanılmamıştır. Hız ve doğruluk açısından STAR, Tophat (Kim et al., 2013) gibi diğer RNASeq eşleyicilerinden daha iyi performans gösterir ancak büyük miktarda bellek (28 GB) gerektirir. Diğer taraftan da DNA ve RNA mapper HISAT2 (Kim et al., 2015) daha az duyarlılığa sahip olma pahasına, STAR'dan çok daha hızlı ve daha hassastır. Ayrıca daha düşük bellek gereksinimlerine (yaklaşık 4 GB) sahiptir. HISAT2, günümüzde Tophat3'ün bir sonraki sürümü için temel motor hizalaması tool olarak kullanılmaktadır.

Tekrar belirtmek gerekirse sıklıkla RNASeq ile genlerin ekspresyon seviyeleri ölçülmektedir. Bunun için haritalamadan sonra bir kantifikasyon aracı kullanılmalıdır. Genlere eşlenen kısa okumaların sayısını saymanın birden çok yolu bulunmaktadır. Bunun için en basit yaklaşım fiili standart bir araç kullanmaktır. Örneğin HTSeq-count bu araçlardan biridir (Anders et al., 2015). HTSeq aslında birçok izoformun

ekzonlarını paylaştığı için izoform miktar tayininde uygundur. Esas amaç diferansiyel ekspresyon gerçekleştirmekse yeterli olan yaklaşık gen kantifikasyonu için yeterli bir araçtır.

Kantifikasyon aracının Cufflinks tarafından kullanılan rpkm/fpkm gibi diğer normalleştirilmiş ölçüler yerine 'counts' vermesi önemlidir (Trapnell et al., 2012). En popüler RNASeq diferansiyel ifade programları (Robinson et al., 2009; McCarthy et al., 2012; Law et al., 2014; Ritchie et al., 2015; Chen et al., 2016) girdi olarak bir sayım matrisine ihtiyaç duyarlar. Normalleştirme bu yazılımlar tarafından dâhili olarak yapılır ve sayım matrisinin önceden dönüştürülmesi bu araçların kullanılmasını engeller.

RNASeq, farklı biyolojik koşullar altında gen bolluklarındaki farklılıkları incelemek için sık kullanılan bir teknoloji haline gelmiştir. RNASeq kullanarak diferansiyel ifadenin tespiti için hesaplamalı ve istatistiksel yöntemlerin sürekli gelişimine rağmen, uygulanacak en iyi yaklaşımın hangisi olduğu konusunda hala büyük bir tartışma var. Kritik adımlardan biri, ham veri sayılarını örnekler arasında karşılaştırılabilir gen ifadelerine çevirmek için uygun bir normalleştirme yöntemine sahip olmaktır. RNASeq normalleştirme yaklaşımları varsayımlarında farklılık göstermektedir. Down-stream analizin ciddi şekilde etkilenmemesi ve bazen hatalı sonuçlara yol açmaması için varsayımların sürdürüldüğü doğru normalleştirme yöntemini seçmek önemlidir.

Gen ifadesi, bir gene düşen okuma sayısı ile orantılıdır, ancak birkaç faktörün dikkate alınması gerekir. Uzun genler, aynı ifade düzeyinde kısa genlerden daha fazla okuma biriktirme eğiliminde olduğundan, örnek gen karşılaştırmaları için “raw count”ın gen uzunluğuna göre normalleştirilmesi gerekir. GC içeriği, GC içeriği amplifikasyon bias’ları nedeniyle bir örnek içindeki ifade seviyelerini karşılaştırırken dikkate alınması gereken başka bir faktördür. Diğer taraftan, örnekler arası karşılaştırmalar için, kütüphane boyutu ve RNA kompozisyon önyargılarının da dikkate alınması gerekir. RNASeq örneklerinin sekanslama derinliğindeki farklılıklar kaçınılmazdır. Bu farklılık normalizasyon adımı sırasında düzeltilmelidir. Eğer bu dikkate alınmazsa daha fazla sekans derinliğine sahip örnek yanlış bir şekilde up regüle genlere sahip olacaktır. Bunlara ek olarak RNASeq, bir gene eşlenen okuma sayısının diğer genlerin ifadesine bağlı olacağı özelliğine sahiptir. Bu nedenle, okumaların

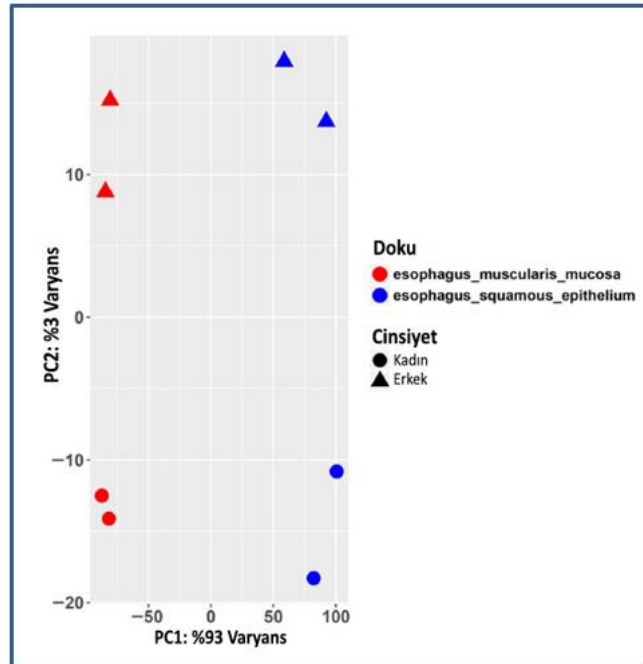
genler arasında nasıl dağıldığına ilişkin farklılıklar da büyük önem taşımaktadır. RNASeq kütüphanesinde, sekanslama efforunun çoğunu tüketen az sayıda yüksek düzeyde eksprese edilmiş gen varsa, daha az ifade edilen diğer genler için daha az okuma olacaktır. Bu peak ifadeleri olmayan başka bir örnekle karşılaştırıldığında, uygun düzeltme uygulanmazsa, yetersiz örneklenmiş genler down regüle olmuş olarak görünecektir.

Dizileme derinliklerinde değişkenlik ve bileşimsel sapma ile farklı senaryolarda daha iyi performans gösterdiği görülen normalleştirme yöntemleri, edgeR geliştiricileri ve DESeq yöntemi tarafından önerilen M değerlerinin (TMM: trimmed mean of M) kırılmış ortalamasıdır. Ancak bu normalleştirmeler, teknik etkilerin diferansiyel olarak eksprese edilen ve non-diferansiyel şekilde eksprese edilen genlerin aynı şekilde etkilediği ve koşullar arasında simetrik diferansiyel ekspresyon olduğu ortak varsayımına sahiptirler. Bu da up regüle olan genlerin sayısının down regüle genlerin sayısına benzer olması gerektiği anlamına gelir. Normalizasyonun ana amacı, örnekler arasında DE olmayan genlerin ekspresyonunu dengelemek ve daha sonra genlerin geri kalanını DE olmayan genlerle aynı normalleştirme faktörü ile normalize etmektir. Popüler RPKM/FPKM normalizasyon adımlarının yalnızca, her hücredeki toplam mRNA miktarının deneysel koşullar arasında aynı olması durumunda iyi sonuç alınacağını belirtmek önemlidir. Buna ek olarak birkaç yayın da, kütüphane kompozisyon biaslarının ele alınamaması nedeniyle, örnekler arası karşılaştırmalar için bu normalizasyonun kullanımı konusunda özellikle kritik olduğu vurgulanmıştır (Bullard et al., 2010; Dillies et al., 2013).

Örnekler içi karşılaştırmalara müdahale eden ek faktörler, örnekler veya koşullar arasında transkript uzunluğundaki değişiklikleri, transkript boyunca kapsamadaki konumsal bias, ortalama parça boyutunu ve genlerin GC içeriğini içerirler. GC düzenlemeleri için EDASeq paketi kullanılmaktadır (Steijger et al., 2013).

Ekspresyon heterojenliği biliniyorsa, DE analizinde modellenen ve ayarlanabilen biyolojik (örneğin yaş, cinsiyet, genotip veya doku) ve teknik faktörler (deneysel seriler) tarafından yönlendirilirken bilinmiyorsa ve açıklanamazsa analizi saptırabilir. Örneğin, İlgilenilen faktör ile gerçek bağlantı maskelenebilir. Bu gizli faktörler tamamen rastgele ve bloke edilmiş deneysel tasarımlarda ve popüler

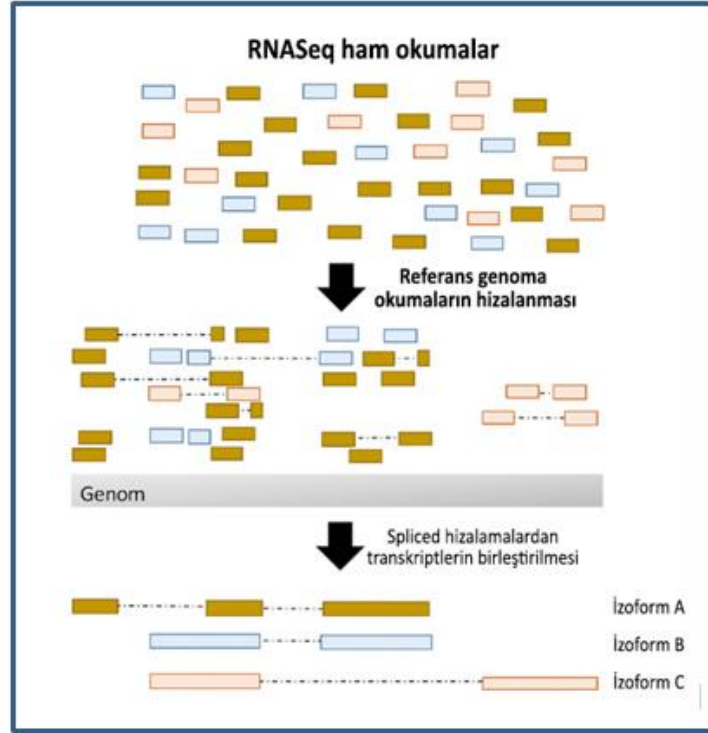
yöntemlerle küresel normalizasyondan sonra bile devam edebilirler. Gen ekspresyonu üzerinde etkisi olan bazı ölçülmemiş değişkenlere sahip olmak genellikle kaçınılmazdır. Sonuç olarak, herhangi bir analize başlamadan önce ifade verilerini bir boyutluluk azaltma yöntemiyle araştırmak şiddetle tavsiye edilmektedir. Örneğin, örneklerin gruplandırılmasını görsel olarak incelemek için ana bileşen analizi (PCA) ile örnekler grafik üzerinde çizilebilirler. Her bir temel bileşen, baskın gen ekspresyon modellerini temsil eder ve daha yüksek bileşenler, varyansın yüksek oranını açıklarlar. İlk temel bileşenler ilgilenilen değişkeni yansıtmıyorsa (örnekler beklendiği gibi kümelenmez) bu sinyali maskeleyen başka istenmeyen faktörlerin olduğunun bir göstergesidir. Özofagus RNASeq örneklerinin iyi bir şekilde ayrılması Şekil 2.17'da gösterilmiştir. Varyansın %93'ünü açıklayan birinci temel bileşen (PC1), muskularis mukoza ve skuamöz epitel özofagus örneklerini birbirinden ayırır. Varyansın %3'ünü açıklayan ikinci bileşen (PC2) örnekleri cinsiyete göre ayırır. Bu gözlemle, farklı özofagus doku parçaları arasında cinsiyetler arasında olduğundan daha fazla ifade farklılığı beklendiği öne sürülebilir.



Şekil 2.17 ENCODE projesinden dört özofagus skuamöz mukoza ve dört özofagus muskularis epiteli toplam RNASeq örneğini gösteren PCA

RNASeq count verileri bir Poisson dağılımıyla daha iyi modellenir ancak biyolojik kopyalara sahip olmak, verilerin aşırı dağılmasına neden olur. Bu nedenle, RNASeq DE için en popüler yazılım paketleri negatif bir binom kullanırlar (McCarthy et al., 2012; Love et al., 2014). Küçük numune boyutları nedeniyle her bir gen için grup içi varyansı kesin olarak tahmin etmek zordur. Bu zorluk edgeR (Robinson et al., 2009) ve DESeq2 (Love et al., 2014) DE yaklaşımları kullanılarak, gen dağılımının ortalama gen ekspresyonuna bağlı olduğunu varsayarak ve bilgiyi genler arasında bir havuzda toplayarak bu sınırlama iyileştirilebilir. Örnekler arasındaki benzerlikleri incelemek ve potansiyel aykırı değerleri belirlemek için DE analizine başlamadan önce en üstteki değişken genlerle bir PCA yapılması şiddetle tavsiye edilmektedir. Uzaktaki örnekler, örneklerin geri kalanından çok uzakta eşleniyorsa, bunların analizden çıkarılması gerekmektedir. Bununla birlikte hem DESeq2 hem de edgeR örnek/gen aykırı değerlerine karşı dayanıklıdır. Yeterli biyolojik kopya ile aykırı değer numunesi olan ve olmayan DE analizinin yapılması ve sonuçların nasıl etkilendiğinin kontrol edilmesi önerilir.

Önceden anlamlılığa göre filtrelenmiş bir gen listesinde (genellikle $FDR < 0.05$) fonksiyonel zenginleştirmeyi arayan, gen seti testleri, herhangi bir cutoff uygulamaya gerek kalmadan sıralı gen listelerinde çalışırlar. RNASeq çalışmaları için derece metriksi, DESeq2 tarafından rapor edilen Walt testi istatistiği olabilir. Bu standart hataya bölünen kat değişikliğinin ($\log_2FC$) $\log_2$ 'sine karşılık gelmektedir.



Şekil 2.18 Tipik bir RNASeq pipeline örneği

Şekil 2.18'e baktığımız zaman RNAseq okumaları, spliced bir hizalayıcı ile referans genoma hizalanır. Spliced hizalamalar kesikli çizgilerle gösterilmektedir. Ardından bir transkriptome birleştirici, hizalamaları izoformlara çözmeye çalışır. Burada, üç izoform vardır. Bir okumanın kaynaklandığı izoformun 'gerçeği' renkle belirtilir. Ancak aralarında spliced bağlantı olmadığından bir assemble bunu çözemeyebilir. Örneğin pembe okumalara sahip iki ekzon onlara hizalanır (Sa et al., 2018).

3. Materyal ve Metot

Bu tez çalışmasında 44 örnekte belirlenen Biyoinformatik analiz hattı ile RNASeq ile elde edilen veriler kullanılarak analiz yapıldı.

Çalışmalar Biruni Üniversitesi Genom Merkezi ve Acıbadem Üniversitesi Tıp Fakültesi Temel Tıp Bilimleri Biyoistatistik ve Tıp Bilişimi Anabilim Dalı'nda ortak olarak yürütüldü.

Biyoinformatik analiz öncesi ücretsiz bir biyomedikal veri tabanı olan PUBMED'de Kistik fibrozis, RNASeq, RNASeq analiz hattı gibi anahtar kelimeler kullanılarak literatür taraması yapıldı. “Transcriptome profile of Primary Human Lung Fibroblasts (Cystic Fibrosis vs. Non-Cystic Fibrosis)” ve “Transcriptomic Profile of Cystic Fibrosis Patients identifies Type I Interferon response and Ribosomal stalk proteins as potential modifiers of disease severity” başlıklı güncel çalışmalar tez çalışmamızın amacına uygun olduğu için seçildi.

Tez çalışması kapsamında kullanılan 14 adet RNASeq verisi “Transcriptome profile of Primary Human Lung Fibroblasts (Cystic Fibrosis vs. Non-Cystic Fibrosis)” başlıklı çalışmaya ait olup GSE141537 numarasıyla Gene Expression Omnibus (GEO) deposundan, Kalan 30 adet, PRJNA393803 proje numaralı ve “Transcriptomic Profile of Cystic Fibrosis Patients identifies Type I Interferon response and Ribosomal stalk proteins as potential modifiers of disease severity” başlıklı çalışmaya ait olan veri ise Avrupa Biyoinformatik Enstitüsü (EBI) veri tabanından alındı. Biyoinformatik analiz için çalışmanın yapılacağı uygun format ve yapıda bilgisayar kullanıldı.

3.1. Veriler

3.1.1 Veri Kaynağı 1

1) Transcriptome profile of Primary Human Lung Fibroblasts (Cystic Fibrosis vs. Non-Cystic Fibrosis)

- 4 örnek: 47 yaşında, Erkek, Sağlıklı, akciğer fibroblast 2D kültür

- 4 örnek: 45 yaşında, Kadın, Kistik Fibroz, akciğer fibroblast 2D kültür
- 3 örnek: 47 yaşında, Erkek, Sağlıklı, akciğer fibroblast, *In vitro* 3D konnektif doku
- 3 örnek: 45 yaşında, Kadın, Kistik Fibroz, akciğer fibroblast, *In vitro* 3D konnektif doku

Bu çalışmanın alt başlıkları aşağıdaki gibidir:

-GSE141535 numaralı *Transcriptome profile of Primary Human Lung Fibroblasts (Cystic Fibrosis vs. Non-Cystic Fibrosis) [2D]*,

-GSE141536 numaralı *Transcriptome profile of Primary Human Lung Fibroblasts (Cystic Fibrosis vs. Non-Cystic Fibrosis) arranged in a 3D model of the connective airways tissue*

Çalışmalardan biri PMID: 32492951 numarasıyla *Intrinsic Abnormalities of Cystic Fibrosis Airway Connective Tissue Revealed by an In Vitro 3D Stromal Model* başlığıyla yayımlanmıştır.

Çalışmada kullanılan hücreler Normal/kistik fibrozlu insan akciğer fibroblastı olup Lonza tarafından satın alınmıştır (N-HLF CC-2512 ve CF-HLF CC-194843; EuroClone S.p.A. Milan, İtalya). Hem CF-HLF hem de N-HLF, Kafkas popülasyonuna ait donörlerin akciğerlerinden çıkarılmıştır. Özellikle, N-HLF CC-2512 kodu için sırasıyla 47 ve 52 yaşındaki erkek hastalardan gelen 0000669504 ve 0000548315 lotlar ve CF-HLF CC-194843 kodu için KF'den etkilenen 45 yaşındaki bir kadın hastadan gelen 9F3124 lotu kullanılmıştır. Lonza şirketi, von Willebrand Faktör İfadesi/Faktör VIII, sitokeratin 18 ve 19 için negatif olduklarını garanti etmektedir. Alfa düz kas aktin durumunda, hücrelerin %5'e kadar pozitif olmasına izin verilir. Bu değerden daha fazlasını ifade eden hücre toplulukları, kalite kontrollerden geçemez ve bu nedenle satın alınamaz. Çalışmada KF donörünün KFTR'sinin özellikli mutasyonu ile ilgili bilgi şirkette mevcut olmadığı belirtilmiştir. Hücreler, zenginleştirilmiş MEM'de kültürlenmiştir (Sigma Aldrich S.r.l. Milan, Italy) supplemented with 20%FBS (Sigma Aldrich), 2% of Non-Essential Amino acids (EuroClone), 1% of L-Glutamine (Microtech, Microgem Naples, Italy) and %1% of penicillin/streptomycin (Sigma Aldrich), until passage 7). Çalışmada belirtilen

protokollere göre elde edilen 3D örnekler toplanarak yıkanmıştır ve paslanmaz çelik bir boncuk kullanılarak bir girdap üzerinde Trizol (Thermo Fisher Scientific S.p.A. Rodano (MI), İtalya, Avrupa) içinde muamele edilip ardından solüsyon, şırınga iğnesine geçirilmiştir. Homojenizasyondan sonra, numunelere 200uL kloroform eklenip, vortekslenmiştir ve 12,000rcf'de 10 dakika santrifüjlenmiştir. Üst faza 500uL izopropanol eklenmiştir. Örnekler, oda sıcaklığında 10 dakika süreyle inkübe edilip 12,000rcf'de 10 dakika süreyle santrifüjlenmiştir. Süpernatantın çıkarılmasından sonra pelletler, 1ml EtOH %75 ile yıkanıp, 10 dakika 7600rcf'de santrifüjlenip ardından 100uL RNaz içermeyen su içinde yeniden süspanse edilmiştir. RNA'nın çökmesini gerçekleştirmek için numuneye 100uL asit fenol-kloroform ilave edilmiştir. Maksimum hızda 15 dakikalık bir santrifüjlemeden sonra, üst faza 9uL NaAC (3M) ve 250uL EtOH %100 ilave edilmiştir. Örnekler vortekslenip -80 °C'de 1 saat inkübe edilmiştir. Daha sonra tüpler maksimum hızda 4°C'de 15 dakika santrifüj edilmiştir. Süpernatant atıldıktan sonra pellet RNaz içermeyen su içinde yeniden süspanse edilmiştir. RNA örneğini kolajen artıklarından daha fazla temizlemek için çözelti, maksimum hızda santrifüjlenip, Yüksek Saf RNA Doku Kitinin (Roche) kolonuna aktarılmıştır. RNA 20uL RNaz içermeyen su içinde ayrıştırılmıştır. Benzer şekilde, doku homojenizasyon fazı hariç, aynı protokol kullanılarak 2D numunelerden de RNA izolasyonu yapılmıştır. Kütüphane hazırlığı, üreticinin talimatlarına göre QuantSeq 3'mRNASeq Kütüphane hazırlama kiti (Lexogen, Viyana, Avusturya) kullanılarak her numuneden toplam 100ng RNA ile gerçekleştirilmiştir. Toplam RNA, Qubit 2.0 florimetrik Test (Thermo Fisher Scientific) kullanılarak ölçülmüştür. Kütüphanelerin kalitesi D1000 ScreenTape sistem (Agilent Technologies) kullanılarak değerlendirilmiştir. Hazırlanan kütüphaneler, S1 100cycle flowcell (Illumina Inc.) kullanılarak NovaSeq6000 sekans cihazında sekanslanmıştır. Boyut olarak 300bp'lik amplifiye edilmiş parçalanmış cDNA, 100bp'lik bir okuma uzunluğu ile tek uçlu (single-end) sekanslanmıştır. Illumina NovaSeq temel çağrı (BCL) dosyaları, bcl2fastq aracılığıyla fastq dosyalarına dönüştürülmüştür. Ardından QuantSeq 3'mRNA Dizileme Veri İşleme ve Analizi yapılmıştır. Analiz için, adaptör dizilerini, A2M-A kuyruklarını ve düşük kaliteli uç tabanları kaldırmak için (ortalama kalitesi 6'nın altında olan bölgeler) sekans okumaları BBduk yazılımı kullanılarak kırılmıştır. Hizalama, cellRanger web sitesinden elde edilen hg38 referans kullanılarak STAR 2.6.0a3 (Dobin et al., 2013) ile gerçekleştirilmiştir. Genlerin ekspresyon seviyeleri,

Gencode/Ensembl gen modeli (Frankish et al., 2019) kullanılarak htseq count (Anders et al., 2015) ile belirlenmiştir. DE analizi, çok faktörlü deneyler için uygun, genelleştirilmiş doğrusal modellere dayanan istatistiksel bir paket olan edgeR kullanılarak gerçekleştirilmiştir (Robinson et al., 2010). İstatistiksel anlamlılık için seçilen eşik Yanlış Keşif Oranı (FDR) <0.05'na ayrıntılı olarak bakılmıştır.

3.1.2. Veri Kaynağı 2

2) Transcriptomic Profile of Cystic Fibrosis Patients identifies Type I Interferon response and Ribosomal stalk proteins as potential modifiers of disease severity

Almanya Tuebingen Üniversitesi Çocuk Hastanesi ve İsviçre Bern Üniversitesi Pediatrik Solunum Tıbbi Bölümü'nden çalışma için toplam 32 F508del homozigot KF hastasında bu çalışma gerçekleştirilmiştir. RNA kalitesini sağlamak ve RNA bozulmasını önemli ölçüde azaltmak için tüm kan örnekleri PaxGene tüplerinde toplanmıştır. Katılımcılar, KF'nin klinik fenotipini ölçmek için bir standart olan FEV1 değerlerine göre Ağır KF hastalar (n=16) ve orta KF hastalar (n=16) olarak sınıflandırılmıştır. Hastaların ayrıntılı klinik özellikleri Tablo 3.1 'de açıklanmıştır.

Tablo 3.1 Orta ve Ağır KF Hastaları ayrıntılı klinik özellikleri

Klinik Parametreler	Orta Şiddet-KF Hasta	Ağır Şiddet-KF Hasta	P value
Sayı	16	16	-
Yaş: mean (SD)	25 (12)	25 (9)	-
Kadın:Erkek	10:6	8:8	-
KFTR Mutasyon	F508del/ F508del	F508del/ F508del	-
FEV1 (tahmin %), mean (SD)	92 (16)	36 (10)	< 0.0001***
Enfeksiyonlar	PSA, SA, CA, SM	PSA, SA, CA, SM, AF	-

*** Mann-Whitney U test.; PSA, Pseudomonas aeruginosa; SA, Staphylococcus aureus; CA, Candida albicans; SM, Stenotrophomonas maltophilia; AF, Aspergillus fumigatus

Klinik parametrelerin istatistiksel analizleri için veriler, Mann-Whitney U testi kullanılarak analiz edilip, SD olarak sunulmuştur. Etik onay, Almanya Tuebingen Üniversitesi Çocuk Hastanesi Kurumsal inceleme kurulundan alınmıştır. Tüm katılımcılardan bilgilendirilmiş yazılı onam; çocuk olanlar için, ilgili ebeveynlerden veya vasilerden onay alınmıştır. RNA Kütüphane hazırlığı için toplam RNA, standart protokollere (www.qiagen.com) sahip bir Qiacube robotu kullanılarak PAXgene tüplerinden izole edilmiştir. Fazla globin mRNA'sı GLOBIN clear kit ile çıkarılmıştır. Dizileme kütüphanesi üreticinin (www.illumina.com) talimatları izlenerek TruSeq™

Total RNA Sample Prep Kit ile hazırlanmıştır. Her bir dizileme kütüphanesi, bir dizinin hangi örnekten türetildiğini tanımlayan 6 nükleotid uzunluğunda bir barkodla etiketlenmiştir. Boyut seçiminden sonra, tüm dizileme kütüphanesi bir Qubit1 florometresinde ölçülüp, eşmolar miktarlarda 16 kütüphane birleştirilerek iki dizileme pool oluşturulmuştur. Bunlar, Illumina GAIIx tek okumalı flowcell'in 8 şeridine ve iki MiSeq flowcell'e yüklenmiştir. Bağlı moleküller, bir cBot cihazında klonal olarak amplifiye edilmiştir. Daha sonra, her fragmentten ilk 50 nükleotid sekanslanıp ardından adaptördeki barkod dizisini yorumlamak için 7 nükleotid dizileme çalışması yapılmıştır (www.illumina.com). RNASeq veri analizi için, ham data kalitesi FastQC kalite kontrol aracı kullanılarak değerlendirilip, yüksek kaliteli okumalar, STAR RNASeq hizalama aracı (Dobin et al., 2013) kullanılarak Homo sapiens referans genomuna (hg19) göre hizalanmıştır. STAR hizalayıcılar, hizalamadan önce okuma dizilerinden adaptör dizilerini otomatik olarak algılayıp kaldırır. Daha sonra, eşlenen okumaların toplam sayısı ve her örnek için benzersiz şekilde eşlenen okumaların sayısı hesaplanmıştır. Bu çalışmada, ek bir kalite kontrol katmanı olarak, düşük eşlenmiş okuma sayısı ve benzersiz şekilde eşlenmiş okumaların küçük bir yüzdesine sahip örnekler atılmıştır. Bu nedenle çalışmadaki total örnek sayısı 30'dur. Ek kalite kontrolünden geçen örnekler için HTSeq (Anders et al., 2015) kullanılarak gen ekspresyon seviyesi CPM olarak tahmin edilmiştir. Orta KF ve Ağır KF arasında farklı şekilde ifade edilen genleri (DEG'ler) tanımlamak için EdgeR R paketi kullanılmıştır (Robinson et al., 2010). Her bir KF grubundaki numuneler biyolojik kopyalar olarak muamele edilip sadece anlamlılık sınırını geçen genler DEG olarak muamele edilmiştir. DEG'leri belirlemek için anlamlılık kesme değeri olarak ifadede -0.5'ten küçük veya 0.5'ten büyük ve P değeri (FDR düzeltilmiş P değeri) 0.05'ten küçük log kat değişimi kullanılmıştır. RNASeq verilerinin global düzenleyici etkileşimini ISMARA çevrimiçi aracı (www.ismara.Unibas.ch) kullanarak tahmin etmek için Motif aktivite yanıt analizi (MARA) gerçekleştirilmiştir. MARA, transkripsiyon faktörü (TF) bağlanma bölgelerini belirli bir algoritma ile tahmin edip belirli bir örnekte gen ekspresyonu açısından spesifik TF'nin etkisini hesaplar. Çalışma grupları arasında motif aktivitesindeki farklılıkları analiz etmek için Mann-Whitney U sıra toplamı testleri uygulanmıştır. Tüm büyük DEG'ler, Açık Hedef Platform (www.targetvalidation.org) kullanılarak KF ile ilişkileri açısından kontrol edilmiştir.

qRT-PCR validasyonu, RNASeq için kullanılan aynı örnekleri kullanarak DEG'lerin 16'sı için gerçekleştirilmiştir. Primerler (Tablo 3.2), Primer3 çevrimiçi aracı (www.primer3.ut.ee) ile dizayn edilmiştir. Reaksiyon sonrası seçilen DEG'lerin nispi ekspresyon seviyeleri 18S rRNA'ya normalize edilmiştir. Orta KF ve Ağır KF arasındaki mRNA ekspresyonundaki farklılıklar, REST 2009 yazılımı ile ikili sabit yeniden tahsis randomizasyon testleri ile analiz edilmiştir (Pfaffl et al., 2002). Ağır KF hastaları grubu kontrol grubu olarak kullanılmıştır ve istatistiksel anlamlılık randomizasyon testleri ile belirlenmiştir. Ayrıca her bir genin ifadesi $2^{-\Delta\Delta C_t}$ yöntemi kullanılarak her hasta için belirlenip çalışma grupları arasında ayrı gen analizleri için Mann-Whitney U testleri uygulanmıştır. Anlamlılık seviyesi, <0,05'lik bir P-değerine ayarlanmıştır.

Tablo 3.2 qRT-PCR validasyonu için kullanılan primerler

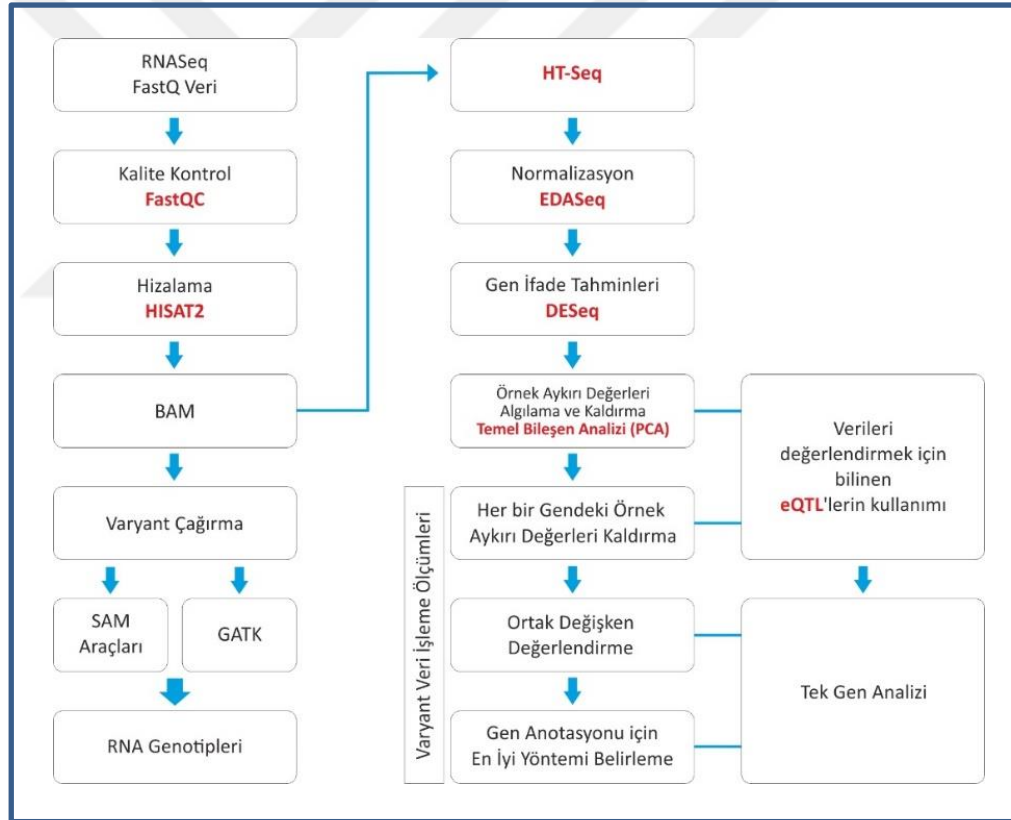
Gen	Primer sekans (5'-3')	Amplicon Boyut (bp)	Annealing Sıcaklığı (°C)
HERC5	F:CCAGCTTGCTTGCCAACAG	157	58
	R:CGGCCAGTAAACCCTCTTCT		
IFIT1	F:TGGACCCTGAAAACCCTGAA	243	54
	R:TCTGTGAGGACATGTTGGCT		
IFIT2	F:GCGAAACAACCTGCTCCATCT	205	55
	R:CCAAGACATGCAAAGCCTCA		
RSAD2	F:AAGAGGAGGAAGAGGACCCT	250	55
	R: CAGAACCTCACCAACTTGCC		
IFI44L	F: GAGCAACTGGTGTGTCGTTT	213	56
	R: CCTATTTCTGTGCTCTCTGGC		
GOS2	F:GGAATGGAGAGACAGAGGGG	239	59
	R: AGTGCAAAATGGTAGACGCA		
CXCL10	F: TGGATGTTCTGACCCTGCTT	201	56
	R: AAAGAATTTGGGCCCCCTTG		
FOSB	F:TCTGTCTTCGGTGGACTCCTTC	209	57
	R:GCAAACCGTAGATGCTCAGGG		
OAS3	F: GCTTCACAGAGCTACAACGG	167	58
	R: CTCCCAGGCATACACAGTCA		
IFI6	F: AGCAGCGTCGTCATAGGTAA	213	56
	R: TGCACTCTAGCCTGGACAAT		
MX1	F: CATCCAGCCACCATTCCAAG	168	57
	R: AGAATCGCTTGAACCTGGGA		
EGR1	F: AGCTGGAGGAGATGATGCTG	257	54
	R: CCAGCACCTTCTCGTTGTTC		
IL8	F: TCTTGGCAGCCTTCTGATT	211	56
	R:TCCAGACAGAGCTCTCTCCATC		
LOC644172	F: CCGACGTCCATTTCTCCAAG	184	58
	R: TCATCCACTTCCAGCTCAGG		
ZFN683	F: AGCCTTGCCTTACCCGCTGAAA	129	60

	R: AATGGACGCTCTCCACTGTGCA		
EPB41L4B	F: ACCCACTTCCTTGACAGAGT	233	55
	R: CGCAAGTTAGCAGCACCAAT		
18s rRNA	F: GTAACCCGTTGAACCCATT	151	54
	R: CCATCCAATCGGTAGTAGCG		

Bu tez çalışması kapsamında bütçe kaynağı olmadığı için qRT-PCR ile validasyon yapılmamıştır

3.2. Analiz Hattı

Verilerin yeniden değerlendirilmesi için bizim kullandığımız analiz hattı aşağıdaki gibidir;



Şekil 3.1 Verilerin yeniden değerlendirilmesi için analiz hattı

Analizde kullanılan araçlar ve ilgili web kaynakları ise aşağıdaki gibidir;

Tablo 3.3 Biyoinformatik analizde kullanılan araçlar ve ilgili web kaynakları

Program/Veri Tabanı Adı	Kullanılan URL
Fastq files	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE141537 https://www.ebi.ac.uk/ena/browser/view/PRJNA393803
FastQC	http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
Conda (v4.8.3)	Paketlerin ve çevrenin kurulumu için
HISAT2	https://anaconda.org/bioconda/hisat2
Samtools	http://www.htslib.org
GATK	https://gatk.broadinstitute.org/hc/en-us
HTSeq	https://pypi.python.org/pypi/HTSeq
EDASeq	http://www.bioconductor.org
Rstudio	https://www.rstudio.com/
DESeq2	http://bioconductor.org/packages/release/bioc/html/DESeq2.html
VariantFiltration	https://gatk.broadinstitute.org/hc/en-us/articles/360050815032-VariantFiltration

3.3 Analiz İçin Uygun Ortam Oluşturma

Biyoinformatik Analiz için Ubuntu Terminalinde uygun çevreler oluşturuldu. Bunun için öncelikle Anaconda indirildi. Anaconda indirildikten sonra analiz öncesi Ubuntu terminal üzerinden bioconda aracılığıyla aşağıdaki programlar indirilip bilgisayara kuruldu:

- FastQC programı
- Trimming için trim_galore
- Alignment için Hisat2
- Genome Reference Consortium Human Build 38
- Samtools
- bcftools
- GATK

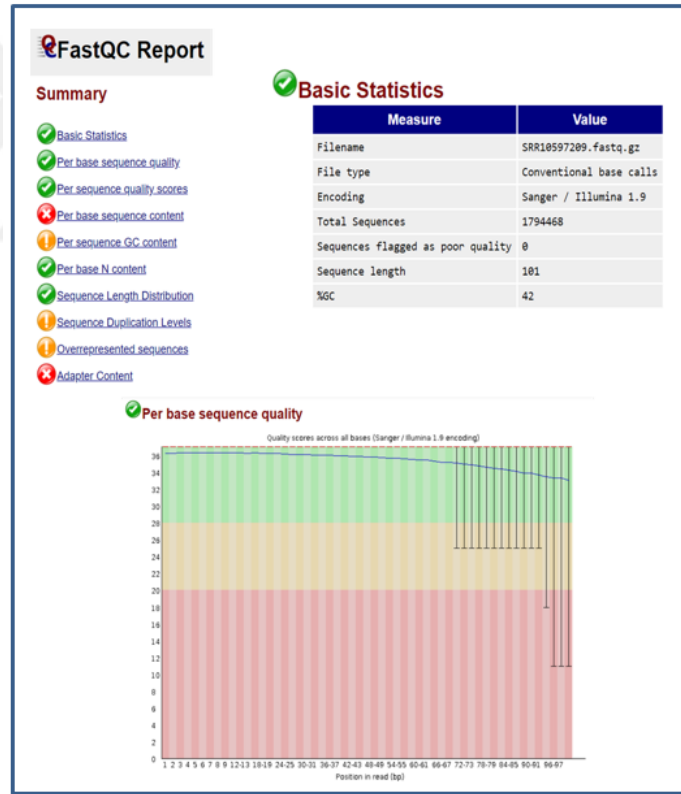
3.3.1. Kalite Kontrol: FastQC

Örneklerin öncelikle FastQC ile Kalite Kontrol analizi yapıldı. FastQC, elde edilen verilerde ileri sekans analizleri yapılmadan önce bilinmesi gereken herhangi bir sorun olup olmadığına dair hızlı bir izlenim almak için kullanılan modüler bir analiz setidir. FastQC tool kullanılarak oluşturulan raporlar değerlendirildi.

N.HLF_2D_Rep1'a ait örnek rapor görüntüsü Şekil 3.2'de gibidir. (Basis Statics ve Per Base Sequence Quality eklenmiştir).

Sekans Dosyasının Açılması:

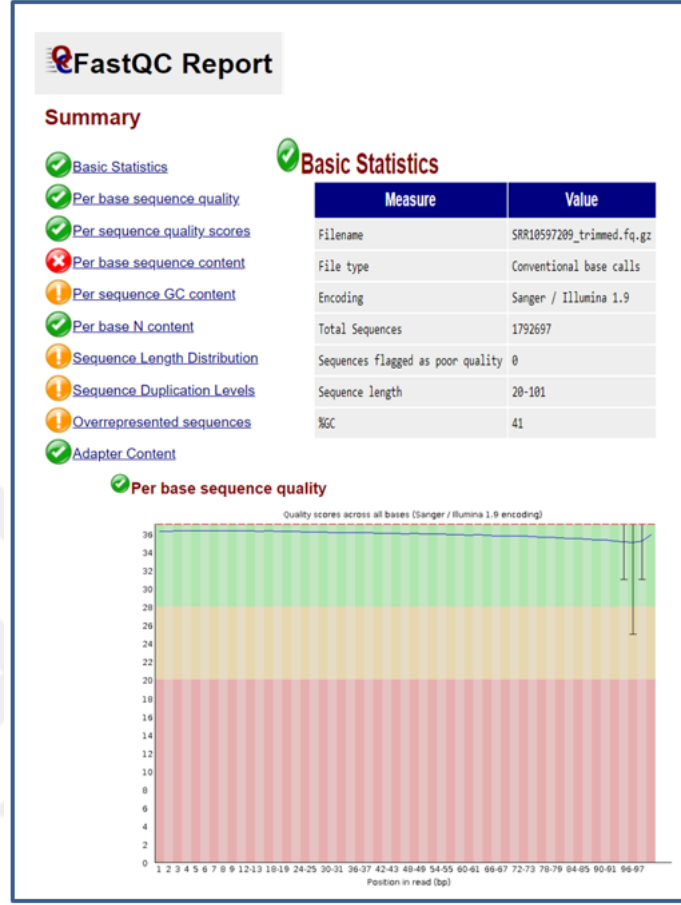
1. Bir veya daha fazla Sekans dosyasını etkileşimli olarak açmak için programı çalıştırılıp ardından Dosya> Aç seçilir.
2. Daha sonra analiz etmek istenilen dosyalar seçilir.
3. Yeni açılan dosyalar, ekranın üst kısmındaki sekmeler grubunda hemen görünecektir.
4. Dosyaların boyutuna bağlı olarak onları açmak birkaç dakika sürebilir.
5. FastQC, bir seferde yalnızca bir dosyanın açıldığı ve yeni dosyaların mevcut dosyalar işlenene kadar bekleyeceği bir kuyruk sistemi çalıştırır.



Şekil 3.2 FastQC tool kullanılarak oluşturulan N.HLF_2D_Rep1'a ait örnek rapor görüntüsü

FastQC analiz değerlendirilmesi sonrasında TRIMMING işlemi uygulanmıştır. Bunun için de *trim_galore* kullanıldı. Trimming sonrası tekrar FASTQC analizi

yapıldı. Örnek rapor görüntüsü (Şekil 3.3) aşağıdaki gibidir. (Basis Statics ve Per Base Sequence Quality eklenmiştir).



Şekil 3.3 TRIMMING işlemi sonrasında Örnek rapor görüntüsü

3.3.2. FastQC → Varyant Çağırma

Çalışmada öncelikle alignment için HISAT2 kullanıldı. HISAT2, YND okumalarını (hem DNA hem de RNA) bir insan genom popülasyonuna hizalamak için hızlı ve hassas bir programdır. Alignment için daha önceden bilgisayarımıza indirdiğimiz Genome Reference Consortium Human Build 38 kullanıldı ve örneklerimiz SAM uzantılı dosyaya, ardından da SAM'ın Binary gösterimi olan BAM uzantılı dosyaya çevrildi. SAM dosyaları ve BAM dosyaları farklı bir formatlarda aynı bilgileri içerirler. BAM, öncelikle referans kimliğine ve ardından en soldaki koordinata göre sıralanmalıdır. Ardından da İndeksleme yapılmalıdır. Burada amaç tüm hizalamalardan geçmeden belirli bir bölgeyle örtüşen hizalamaların hızlı bir şekilde alınmasıdır. HaplotypeCaller kullanılarak BAM dosyasından VCF dosyası

oluşturuldu. Böylelikle GATK VCF elde edilmiş oldu. Elde edilen çıktıya ait örnek Tablo 3.4’da gösterilmektedir.

Tablo 3.4 GATK VCF sonuçlarına ait örnek tablo

Chr	Start	End	Ref	Alt	Func.refGene	Gene.refGene	GeneDetail.refGene	ExonicFunc.refGene	AAChange.refGene	cytoBand
chr1	631010	631010	T	C	intergenic	LOC101928626;MIR12136	dist=2001;dist=1605	.	.	1p36.33
chr1	634113	634113	A	-	intergenic	MIR12136;OR4F3	dist=1428;dist=51603	.	.	1p36.33
chr1	634244	634244	T	C	intergenic	MIR12136;OR4F3	dist=1559;dist=51472	.	.	1p36.33
chr1	680540	680540	A	C	intergenic	MIR12136;OR4F3	dist=47855;dist=5176	.	.	1p36.33
chr1	809967	809967	T	C	intergenic	LOC100288069;FAM87B	dist=31279;dist=7404	.	.	1p36.33
chr1	809992	809992	A	-	intergenic	LOC100288069;FAM87B	dist=31304;dist=7379	.	.	1p36.33
chr1	944296	944296	G	A	UTR3	NOC2L;SAMD11	15658:c.*398C>T;NM_152486:c.*1	.	.	1p36.33
chr1	944307	944307	T	C	UTR3	NOC2L;SAMD11	15658:c.*387A>G;NM_152486:c.*1	.	.	1p36.33
chr1	965592	965592	T	G	UTR3	KLHL17	NM_198317:c.*401T>G	.	.	1p36.33
chr1	965643	965643	T	C	UTR3	KLHL17	NM_198317:c.*452T>C	.	.	1p36.33
chr1	1014274	1014274	A	G	exonic	ISG15	.	synonymous SNV	M_005101:exon2:c.A294G	1p36.33
chr1	1292467	1292467	C	T	UTR3	ACAP3	NM_030649:c.*1097G>A	.	.	1p36.33
chr1	1352965	1352965	A	G	UTR3	MXRA8	6T>C;NM_001282582:c.*639T>C;N	.	.	1p36.33
chr1	1353091	1353091	C	T	UTR3	MXRA8	10G>A;NM_001282582:c.*513G>A;N	.	.	1p36.33
chr1	1373729	1373729	G	T	downstream	AURKAIP1	dist=7	.	.	1p36.33
chr1	1386672	1386672	A	G	UTR3	CCNL2	9T>C;NM_001320153:c.*559T>C;N	.	.	1p36.33
chr1	1496122	1496122	G	A	UTR3	ATAD3B	921:c.*305G>A;NM_001317238:c.*	.	.	1p36.33
chr1	1580045	1580045	C	T	intergenic	SSU72;FNDC10	dist=5182;dist=17967	.	.	1p36.33
chr1	1716966	1716966	A	G	intronic	CDK11A	.	.	.	1p36.33
chr1	1719368	1719368	T	C	exonic	CDK11A	.	synonymous SNV	A315G;p.R105R;CDK11A:R	1p36.33
chr1	1719406	1719406	G	A	exonic	CDK11A	.	nonsynonymous SNV	c.C277T;p.R93W;CDK11A:R	1p36.33
chr1	1727506	1727506	T	C	UTR3	SLC35E2A	NM_182838:c.*4949A>G	.	.	1p36.33
chr1	1810691	1810691	A	T	intronic	GNB1	.	.	.	1p36.33

Ayrıca Samtools ile de VCF (Tablo 3.5) dosyası oluşturuldu. Samtools, SAM, BAM ve CRAM biçimlerindeki hizalamaları işleyen bir dizi yardımcı programdır. Biçimler arasında dönüşüm sağlar, sıralama, birleştirme ve indeksleme yaparak herhangi bir bölgedeki okumaları hızlı bir şekilde alabilir. Öncelikle BAM dosyasından BCF oluşturuldu. Bunun için VCF ve onun ikili karşılığı BCF'deki varyant çağrılarını işleyen bir yardımcı program olan BCFtools kullanıldı. BCFtools, samtools mpileup komutunun çıktısından varyant çağrısı yapar. İndekslenmiş VCF ve BCF her durumda çalışacaktır. Genel olarak, birden fazla VCF aynı anda okunduğunda indekslenmeleri ve dolayısıyla sıkıştırılmaları gerekir. İndekslenme sonrası RAW VCF oluşturuldu. VCF bir metin dosyası formatıdır. Meta bilgi satırları, bir başlık satırı ve ardından her biri genomdaki bir konum hakkında bilgi içeren veri satırları içerir. SNP'ler veya insersiyonlar/delesyonlar gibi genetik varyasyon çağrılarını depolamak için kullanılır.

Tablo 3.5 Samtools VCF sonuçlarına ait örnek bir tablo

Chr	Start	End	Ref	Alt	Func.refGene	Gene.refGene	GeneDetail.refGene	ExonicFunc.refGene	AAChange.refGene	CytoBand
chr1	634113	634113	A	-	intergenic	MIR12136;OR4F3	dist=1428;dist=51603	-	-	1p36.33
chr1	634244	634244	T	C	intergenic	MIR12136;OR4F3	dist=1559;dist=51472	-	-	1p36.33
chr1	809967	809967	T	C	intergenic	LOC100288069;FAM87B	dist=31279;dist=7404	-	-	1p36.33
chr1	809991	809991	A	-	intergenic	LOC100288069;FAM87B	dist=31303;dist=7380	-	-	1p36.33
chr1	944296	944296	G	A	UTR3	NOC2L;SAMD11	5658:c.*398C>T;NM_152486:c.*	-	-	1p36.33
chr1	944307	944307	T	C	UTR3	NOC2L;SAMD11	5658:c.*387A>G;NM_152486:c.*	-	-	1p36.33
chr1	965592	965592	T	G	UTR3	KLHL17	NM_198317:c.*401T>G	-	-	1p36.33
chr1	965643	965643	T	C	UTR3	KLHL17	NM_198317:c.*452T>C	-	-	1p36.33
chr1	1014274	1014274	A	G	exonic	ISG15	-	synonymous SNV	005101:exon2:c.A29	1p36.33
chr1	1234950	1234950	T	G	UTR3	B3GALT6	NM_080605:c.*1682T>G	-	-	1p36.33
chr1	1255143	1255143	C	T	UTR3	UBE2J2	94315:c.*60G>A;NM_194457:c.*	-	-	1p36.33
chr1	1292467	1292467	C	T	UTR3	ACAP3	NM_030649:c.*1097G>A	-	-	1p36.33
chr1	1311788	1311788	T	C	UTR3	INTS11	A>G;NM_001256462:c.*71A>G;N	-	-	1p36.33
chr1	1352965	1352965	A	G	UTR3	MXRA8	T>C;NM_001282582:c.*639T>C;N	-	-	1p36.33
chr1	1353014	1353014	G	C	UTR3	MXRA8	C>G;NM_001282582:c.*590C>G;N	-	-	1p36.33
chr1	1353091	1353091	C	T	UTR3	MXRA8	G>A;NM_001282582:c.*513G>A;N	-	-	1p36.33
chr1	1386672	1386672	A	G	UTR3	CCNL2	T>C;NM_001320153:c.*559T>C;N	-	-	1p36.33
chr1	1496122	1496122	G	A	UTR3	ATAD3B	921:c.*305G>A;NM_001317238:c.	-	-	1p36.33
chr1	1580045	1580045	C	T	intergenic	SSU72;FNDCC10	dist=5182;dist=17967	-	-	1p36.33
chr1	1716966	1716966	A	G	intronic	CDK11A	-	-	-	1p36.33
chr1	1727506	1727506	T	C	UTR3	SLC35E2A	NM_182838:c.*4949A>G	-	-	1p36.33
chr1	1732912	1732912	T	C	intronic	SLC35E2A	-	-	-	1p36.33
chr1	1783270	1783270	C	T	intergenic	NADK;GNB1	dist=3239;dist=2015	-	-	1p36.33
chr1	1785563	1785563	T	C	UTR3	GNB1	b>G;NM_001282538:c.*1500A>G	-	-	1p36.33

GATK ve Samtools ile elde edilen raw-VCF'e, varyant filtrasyonu uygulandı. FastQC analizinden elde edilen sonuca göre Varyant kalitesi 20'ye göre filtrelendi. Bu kalite phred ölçeklidir. Phred kalite puanı otomatik DNaseq ile üretilen bazların tanımlama kalitesinin bir ölçüsüdür. Nitelendirdiği her şeyin %1 hata şansı ile %99 doğru olduğu anlamına geldiğinden 20 veya üzeri bir Phred Puanı kabul edilebilir. raw-VCF'ten → Annovar-VCF oluşturmak için öncelikle, aşağıdaki link kullanılarak bazı veritabanları indirildi.

https://www.openbioinformatics.org/annovar/annovar_download_form.php

3.3.2.1 Varyantların Değerlendirilmesi

RNA Genotiplendirme için GATK ve Samtools ile elde edilen filtrelenmiş VCF'ler birleştirilip RNA genotiplendirme için hazırlandı.

3.3.3 Transkriptom Analizi

3.3.3.1 HTSeq

Tez çalışması kapsamında kullanılan analiz hattının 2. kısmın transkriptom analizidir. Ubuntu terminal ile HTSeq oluşturuldu. Bunun için hg38.refGene.gtf.gz indirildi. HTSeq count yapıldıktan sonra tüm veriler R Studio programında EDASeq aşamasına geçmek için birleştirildi.

3.3.3.2 EDASeq

EDASeq verilerin normalizasyonu için kullanılmaktadır. EDASeq’de iki normalleştirme adımı vardır. Bunlardan ilki gene özgü ve numuneye özgü etkileri ayarlayan numune içi normalleştirme adımıdır. Diğeri ise numune arasındaki dağılım farklılıklarını düzelteren numuneler arası normalleştirme adımıdır.

3.3.3.3 DESeq

DESeq2, Cooks mesafesini kullanarak aykırı değerleri otomatik olarak algılar ve bu genleri analizden çıkarır. Ayrıca, normalleştirilmiş sayıların ortalaması bir optimizasyon prosedürü tarafından belirlenen bir eşğin altında olan genleri otomatik olarak kaldırır. Bu araç girdi olarak bir raw-count tablosu alır. Count tablosu, deney gruplarını tanımlayan bir metadata dosyası ile ilişkilendirilmelidir. DESeq2, tüm numunelerde her gen için geometrik ortalamanın hesaplandığı bir dâhili normalleştirme gerçekleştirir. Her numunedeki bir genin sayıları daha sonra bu ortalamaya bölünür. Bir numunedeki bu oranların medyanı, o numune için boyut faktörüdür. Bu prosedür örneğin bir deney koşulunda yalnızca az sayıda gen çok yüksek düzeyde eksprese edildiğinde ancak diğerinde ifade edilmediğinde ortaya çıkabilecek kütüphane boyutu ve RNA bileşimi bias düzeltir. Sonuç olarak, DESeq2, Bioconductor aracılığıyla sağlanan popüler bir diferansiyel ifade analiz paketidir. Diferansiyel ifade testleri, negatif bir binom genelleştirilmiş doğrusal modele dayanmaktadır.

4. Sonuç

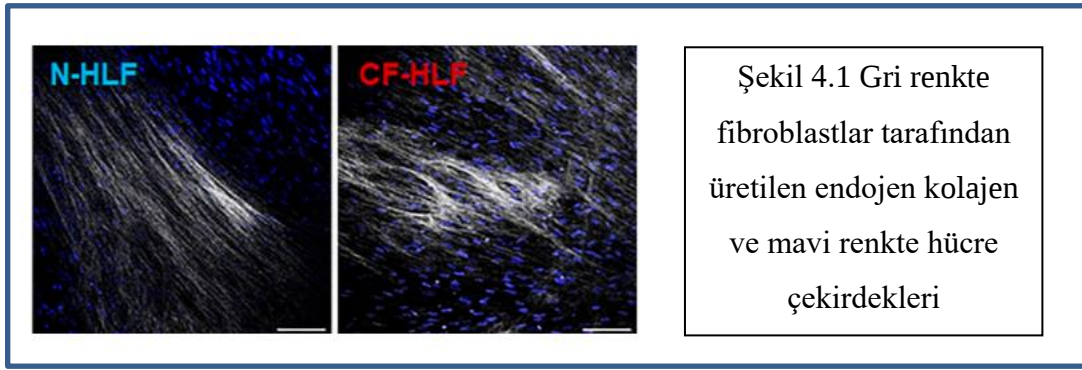
4.1. Veri kaynağı 1

Transcriptome profile of Primary Human Lung Fibroblasts (Cystic Fibrosis vs. Non-Cystic Fibrosis). Bu çalışmada 14 örnek kullanılmıştır.

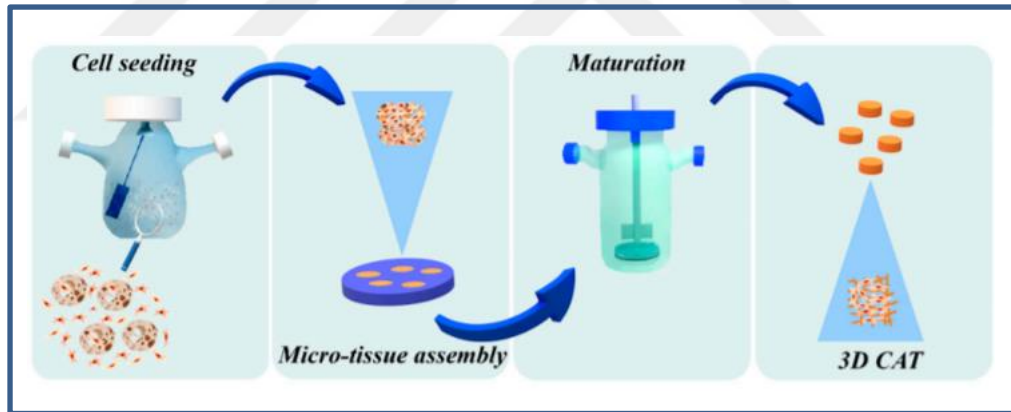
Tablo 4.1 Veri Kaynağı 1 örnek listesi ve özellikleri

ÖRNEKLER	Baz	Büyüklik	Hücre Tipi	Donör Yaşı	Cinsiyet
N.HLF_2D_Rep1	181.24M	56.30Mb	Normal akciğer fibroblast	47	Erkek
N.HLF_2D_Rep2	481.13M	148.55Mb	Normal akciğer fibroblast	47	Erkek
N.HLF_2D_Rep3	255.98M	77.94Mb	Normal akciğer fibroblast	47	Erkek
N.HLF_2D_Rep4	177.29M	53.85Mb	Normal akciğer fibroblast	47	Erkek
CF.HLF_2D_Rep1	220.47M	67.81Mb	KF akciğer fibroblast	45	Kadın
CF.HLF_2D_Rep2	224.65M	68.76Mb	KF akciğer fibroblast	45	Kadın
CF.HLF_2D_Rep3	201.35M	61.05Mb	KF akciğer fibroblast	45	Kadın
CF.HLF_2D_Rep4	265.48M	81.10Mb	KF akciğer fibroblast	45	Kadın
N.HLF_3D_Rep1	611.93M	199.42 Mb	Normal akciğer fibroblast	47	Erkek
N.HLF_3D_Rep2	572.71M	181.17 Mb	Normal akciğer fibroblast	47	Erkek
N.HLF_3D_Rep3	701.25M	222.36 Mb	Normal akciğer fibroblast	47	Erkek
CF.HLF_3D_Rep1	663.28M	210.96 Mb	KF akciğer fibroblast	45	Kadın
CF.HLF_3D_Rep2	814.52M	259.21 Mb	KF akciğer fibroblast	45	Kadın
CF.HLF_3D_Rep3	934.04M	292.06 Mb	KF akciğer fibroblast	45	Kadın

Bu çalışmaya göre, Biyomühendislik Bottom-up yaklaşımı bağlayıcı hava yolu doku eşdeğeri üretimine yol açmaktadır. Çalışmada ilk olarak, hem N-HLF hem de CF-HLF 1 hafta boyunca askorbik asit ile %100 birleşik hücre tabakasını uyararak *in vitro* endojen kollajen üretme kapasitesi değerlendirilmiştir. Şekil 4.1’de (Mazio et al., 2020). endojen kollajen liflerinin İkinci Harmonik Nesil (SHG) sinyali sayesinde *de novo* sentezlenmiş matrisi gri olarak gösterilmiştir.



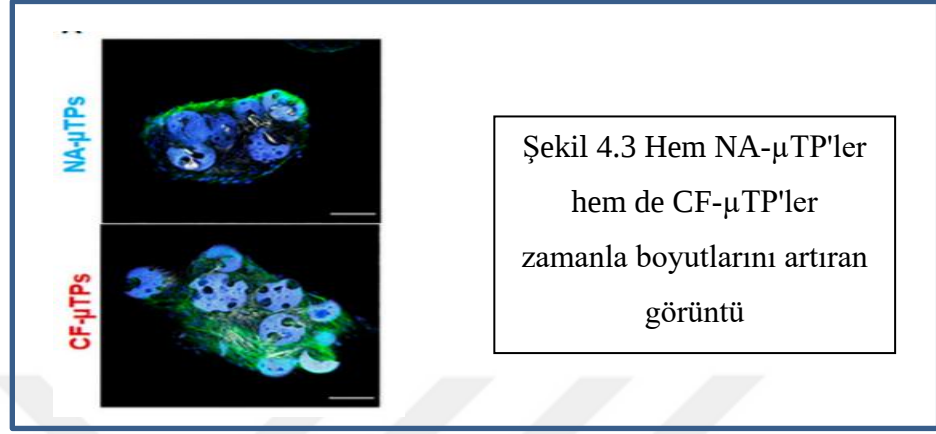
Ardından, Şekil 4.2 (Mazio et al., 2020)'de gösterildiği gibi önceden oluşturulmuş bir doku mühendisliği bottom-up yaklaşımını (Imparato et al., 2013) hafifçe değiştirerek akciğer bağ hava yolu dokusu (CAT) elde edilmiştir. Hem N-HLF hem de CF-HLF dinamik kültür koşullarında Hücre Dışı Matrisin (ECM) elemanlarını çoğaltmak ve sentezlemek için gözenekli jelatin mikro yapı iskelelerine yapışabilmiştir.



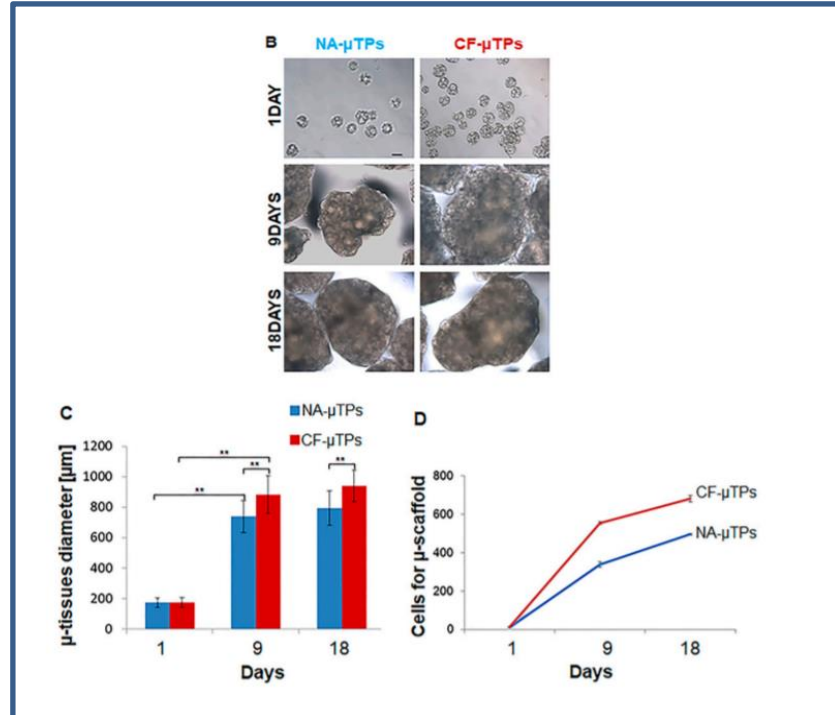
Şekil 4.2 KF bağ hava yolu dokusunu (CAT) geliştirmek için biyo-mühendislik süreci

KF bağ hava yolu dokusunu (CAT) geliştirmek için CF-HLF, gözenekli jelatin mikro-iskeleler üzerinde, döner flask biyoreaktörüne ekilmiş. Fibroblastların tohumlanmasından 18 gün sonra KF mikro dokuları (CF-uTP: cystic fibrosis micro-tissues) alınmış ve olgunlaşma odalarına birleştirilmiş. 3 hafta sonra kendi matrislerine gömülü fibroblastların varlığı ile öne çıkan nihai CAT elde edilmiş. Aynı süreç N-HLF'den başlayarak, normal bağ hava yolu dokusunu (NA-CAT) deneysel kontrol olarak elde etmek için kullanılmış. Zamanla hücre-hücre ve hücre-matris etkileşimleri, Normal ve KF mikro dokularının (NA- μ TP'ler ve CF- μ TP'ler) oluşumu sağlanmış

oldu. Her iki mikro doku da, Şekil 4.3'de (Mazio ve ark.,2020) gösterildiği gibi, endojen kollajen matrisine gömülü hücrelerin varlığına sahipti.



Ancak CF-μTP'ler her zaman noktasında NA-μTP'lerden daha büyüktü. μ-scaffold başına hücre sayısı ile KF akciğer fibroblastlarının normal olanlardan daha hızlı çoğaldığı gösterilmiştir.



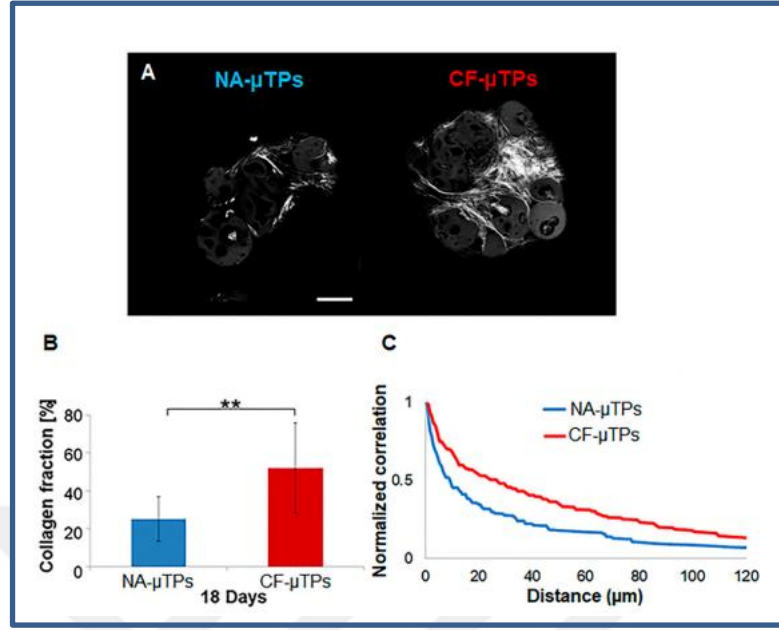
Şekil 4.4 Hücre kültür sonuçlarının toplu görüntüsü

Şekil 4.4'de sonuçlara toplu olarak baktığımızda **B.** NA- μ TP'lerin ve CF- μ TP'lerin optik mikroskop görüntüleri, mikro-iskelelerde fibroblastların tohumlanmasından 1, 9 ve 18 gün sonra, ölçek çubuğu 100 μ m. **C.** Zamanla NA- μ TP'lerin (mavi) ve CF- μ TP'lerin çapının (kırmızı) histogramı (** p değeri <0.01), veriler Ortalama $\pm$ Standart Sapma olarak gösterilir. **D.** NA- μ TP'lerde (mavi) ve CF- μ TP'lerde (kırmızı) zaman içinde mikro- scaffold başına hücre sayısını gösteren grafik.

Ayrıca, KF fibroblastlarının, SHG görüntülerinde (Şekil 4.5A) ve kolajen fraksiyonunun analizinde görüldüğü gibi daha yüksek miktarda kolajen ürettiği gösterilmiştir (Şekil 4.5B) (Mazio et al., 2020). Artan hücre sayısı ve kolajen miktarının, NA- μ TP'lere kıyasla CF- μ TP'lerin boyutundaki artışla uyumlu olduğu gösterilmiştir. Kolajen ağ morfometrisi daha sonra SHG görüntüleri Gri Düzey Birlikte Oluşum matrisinin (GLCM) korelasyon özelliği ile analiz edilerek araştırılmıştır (Wu et al., 2016; Chen et al., 2012; Imparato et al., 2016). CF- μ TP'lerin korelasyon eğrisi (Şekil 4.5B) (Mazio et al., 2020), NA- μ TP'lerden daha yavaş bozulmuş, bu da hastalık modelinde daha düzensiz bir kollajen ağı olduğunu göstermiştir (Hristu et al., 2018). Bunlara ek olarak, CF- μ TP'lerin korelasyon uzunluğu, NA- μ TP'lerden önemli ölçüde daha yüksek olduğundan (Tablo 4.2) (Mazio et al., 2020), hastalık modelinde daha kalın kolajen liflerinin varlığını ortaya koymuştur (Cicchi et al., 2010).

Tablo 4.2 Korelasyon uzunluğu analizi

ÖRNEK	Korelasyon Uzunluğu (λ)
NA- μ TPs	62.6254 $\pm$ 16.23
CF- μ TPs	135.88 $\pm$ 43.31



Şekil 4.5 μTPs morfolojik karakterizasyonu

Şekil 4.5’da (Mazio et al., 2020) μTPs morfolojik karakterizasyonları gösterilmiştir **A.** Normal ve KF mikro dokularındaki kolajen liflerinin (parlak gri) SHG görüntüleri (NA-μTP’ler ve CF-μTP’ler). **B.** NA-μTP’lerde (mavi) ve CF-μTP’lerde (kırmızı) kolajen fraksiyonu (%) diyagramı, veriler Ortalama ± Standart Sapma olarak gösterilir. **C.** NA-μTP’lerin (mavi) ve CF-μTP’lerin (kırmızı) SHG görüntülerini analiz ederek elde edilen normalleştirilmiş korelasyon grafiği

Transkriptomik analiz epitel morfogenezinde ve inflamatuvar tepkide yer alan genlerin up regüle edilmesinde 3D Stromal ortamın temel bir rolünü ortaya çıkarmaktadır. Bu amaçla, GSE141537 numsrslı çalışmada, CF-HLF ve N-HLF arasında gözlemlenen morfolojik farklılıkların altında yatan moleküler imzanın etkisini daha fazla araştırmak için, hem 2D hem de 3D (CAT) konfigürasyonlarında düzenlenen hücrelerin tarafsız bir RNAseq analizi yapılmıştır. Çalışma ile Transkriptomik analizin de benzer şekilde CF-HLF ve N-HLF arasında gözlemlenen farklılıklar da doğrulanmıştır. Aynı zamanda 3D ortamın doku fizyo/patolojisini *in vitro* olarak kopyalamadaki temel rolü de vurgulanmıştır. Bunlara ek olarak, hem CF-HLF hem de N-HLF KFTR’yi ifade etmediğinden, CF-HLF ve N-HLF arasında

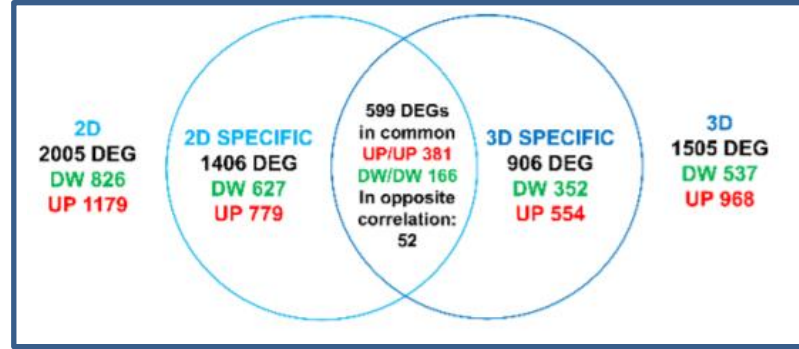
gözlenen farklılıklar *in vitro* KFTR disfonksiyonu ile ilgili değil muhtemelen, birincil hücrelerin çıkarıldığı biyopsi anında hastaya özgü *in vivo* akciğer durumundan türetilen diğer faktörlerle ilişkili olduğu vurgulanmıştır. Hem 2D hem de 3D koşullar, CF-HLF'nin transkriptomunun kontrol olarak kullanılan N-HLF'nin transkriptomuyla karşılaştırılmasıyla analiz edilmiştir. Her iki koşulda da transkriptomda önemli bir değişiklik gözlemlenmiştir. Ayrıntılı olarak, sırasıyla 2D ve 3D transkriptomlarında 2005 ve 1505 DEG bulunmuştur. 2D veya 3D koşulunun gen ekspresyon profilleri üzerindeki etkisini araştırmak ve en düzensiz yolları belirlemek için iki veri seti üzerinde gen ontoloji zenginleştirme analizi (GOEA) gerçekleştirilmiştir (Huang et al., 2009). Bu tez çalışmamızda bu analiz gerçekleştirilmedi.

HAS-2'yi içeren CF-CAT'de CF-HLF tarafından aşırı eksprese edile bir dizi gen gözlemlenmiştir. Bunlar arasında, gen ekspresyonu ve immüno Floresan ile gözlemlenen hyalüronik asit miktarı arasında pozitif bir korelasyonu gösteren TWIST1 ve TGF β 3 yer almıştır. Bunlara ek olarak periostin sentezinin pozitif düzenleyicileri ve WNT yolağındaki WNT2, SFRP2 ve RSPO2 de bulunmaktadır. Son çalışmalar akciğer fibrozu sırasında gözlenene benzer şekilde, KF'li hastaların akciğer fibroblastlarında WNT yolağının anormal bir aktivasyonunu önermektedir (Königshoff et al., 2008; Munguía-Reyes et al., 2018). Akciğer fibrozunda olduğu gibi CF-HLF'de aktive olan diğer genler ise TBX4 (Xie et al., 2016) ve ITGA8'dir. (Hung et al., 2018). HHIP ve CHI3L1'nin, hem KF'de hem de KOAH ve astımda aşırı eksprese edildiği çalışmalarda gösterilmiştir (Hector et al., 2011; Van der Plaat et al., 2017). (Tablo 4.2) (Mazio ve ark., 2020).

Tablo 4.3 3D veri setinde up regüle olmuş genler

GEN	logFC_CF-HLF vs. N-HLF
TGF β 3	1.755
HAS-2	2.062
TWIST1	2.425
WNT2	3.052
SFRP2	3.695
RSPO2	4.331
TBX4	4.593
ITGA8	4.874
CHI3L1	5.027
HHIP	5.09

Venn şemasında gösterildiği gibi, her biri kendi kontrolünde normalleştirilen iki veri seti (2D ve 3D) karşılaştırılmıştır (Şekil 4.6) (Mazio et al., 2020).



Şekil 4.6 2D ve 3D veri kümeleri arasında karşılaştırma

Bu analiz, regülasyonu iki farklı kültür koşuluna bağlı veya bağımsız olan genlerin izole edilmesine ve böylece spesifik olarak 2D (779 up regüle ve 627 down regüle olmak üzere toplam 1406)'de ve 3D (554 up regüle ve 352 down regüle olmak üzere toplam 906)'de düzenlenen transkriptlerin izole edilmesine izin vermiştir. Bunlara ek olarak, hem 2D hem de 3D olarak yaygın olarak düzenlenen 559 DEG izole edilmiştir. Her iki veri setinde de sırasıyla 381 gen up regüle ve 166 gen down regüle olurken, 52'si ters şekilde düzenlenmiştir. Bu sonuç CF-HLF'nin hem 2D hem de 3D'de *in vitro* olarak aktif bir durumu koruduğunu kuvvetle göstermiştir. Bunun yerine 3D modeldeki hücreler, özellikle örneğin iç bölgelerinde bulunanlar, ECM'nin kendisi tarafından yapılan modifikasyonların bir sonucu olarak farklı bir ortam algılayabilirler. CF-HLF tarafından up regüle genler arasında TGFB3, SFRP2, RSPO2, ITGA8 ve CHI3L1 yer almaktadır. Sadece 3D kültür koşulunda spesifik olarak up regüle olan HAS-2 aşırı ekspresyona uğramıştır. Bunun tam tersi olarak da bu gen 2D olarak down regüle olmuştur.

Bu tez kapsamında kullanılan pipeline ile ettiğimiz sonuçlar Varyant çağırma ve Transkriptome olarak ayrılmıştır.

A-Varyant Çağırma: Toplu olarak sonuçlar Tablo 4.4'de paylaşılmıştır.

Tablo 4.4 2D ve 3D örneklerinde varyant çağırma toplu sonuçları

		Bening		*Benign/LB		**CIOP	
N.HLF_2D_Rep1	GATK	24	159	7		rs56302025	
	Samtools	18				0	
N.HLF_2D_Rep2	GATK	37	284	2	7	rs35118288, rs56302025	
	Samtools	30		0		0	
N.HLF_2D_Rep3	GATK	27	251	1	5	.	rs191743300, .
	Samtools	39		0		rs746952638	
N.HLF_2D_Rep4	GATK	25	196	0	6	rs779413093, rs35118288, rs56302025	
	Samtools	36		3		0	
CF.HLF_2D_Rep1	GATK	20	197	0	9		
	Samtools	23		1			
CF.HLF_2D_Rep2	GATK	35	249	1	8	rs56302025	
	Samtools	40		1		0	
CF.HLF_2D_Rep3	GATK	27	264	1	12	rs191743300, rs56302025	
	Samtools	52		2		0	
CF.HLF_2D_Rep4	GATK	28	260	11	11	rs35118288	
	Samtools	55		1		rs145662304	
N.HLF_3D_Rep1	GATK	23	161	6	6	rs56302025	
	Samtools	27		1		0	
N.HLF_3D_Rep2	GATK	28	219	0	10	1	
	Samtools	36		1		0	
N.HLF_3D_Rep3	GATK	34	244	1	9	rs56302025	
	Samtools	46		0		rs753928772, rs200916070	
CF.HLF_3D_Rep1	GATK	38	262	1	10		
	Samtools	39		10			
CF.HLF_3D_Rep2	GATK	37	250	8		rs200075817, rs56302025	
	Samtools	57				0	
CF.HLF_3D_Rep3	GATK	48	299	0	12	rs56302025	
	Samtools	65		3		rs764235096	

* Benign/Likely_benign **Conflicting_interpretations_of_pathogenicity

		Drug Response		Muhtemel benign	
N.HLF_2D_Rep1	GATK	0	rs7294 rs1056892	3	6
	Samtools	rs1695		0	
N.HLF_2D_Rep2	GATK	0	rs7294 rs1056892	9	7
	Samtools	rs1695		0	
N.HLF_2D_Rep3	GATK	0	rs7294	4	rs559542 rs1058467
	Samtools	rs1695 rs1751034		3	
N.HLF_2D_Rep4	GATK	rs1695 rs7294		6	8
	Samtools	rs1056892		1	
CF.HLF_2D_Rep1	GATK	rs7294. rs1056892		5	8
	Samtools			1	
CF.HLF_2D_Rep2	GATK	rs7294. rs1056892		2	9
	Samtools			2	
CF.HLF_2D_Rep3	GATK	0	rs7294, rs1056892	2	3
	Samtools	rs4961		rs11376	

CF.HLF_2D_Rep4	GATK	rs7294, rs1056892	6	8
	Samtools	rs4961, rs1056892	3	
N.HLF_3D_Rep1	GATK	rs7294, rs1056892	5	4
	Samtools	rs7294, rs1751034, rs2214102	4	
N.HLF_3D_Rep2	GATK	0	6	5
	Samtools	rs1695	2	
N.HLF_3D_Rep3	GATK	0	4	10
	Samtools	rs1695, rs11615	1	
CF.HLF_3D_Rep1	GATK	0	9	10
	Samtools	rs4961	0	
CF.HLF_3D_Rep2	GATK	rs7294, rs1056892	8	11
	Samtools	rs7294, rs1056892, rs4961	0	
CF.HLF_3D_Rep3	GATK	0	14	12
	Samtools	2	0	

*** Pathogenic/Likely_pathogenic,_other

		Muhtemel patojenik	Patojenik		***P/LP_O
N.HLF_2D_Rep1	GATK	rs139428292			
	Samtools				
N.HLF_2D_Rep2	GATK	rs797046100	0		rs139428292
	Samtools	0	rs151339003		
N.HLF_2D_Rep3	GATK				
	Samtools				
N.HLF_2D_Rep4	GATK	-	rs118204046, rs551348450, rs543007243		rs139428292
	Samtools	rs267607277	rs118204046		
CF.HLF_2D_Rep1	GATK		rs751610886		
	Samtools		0		
CF.HLF_2D_Rep2	GATK		., rs764261443		
	Samtools		0		
CF.HLF_2D_Rep3	GATK		Intellectual_disability, KMT2A		
	Samtools		0		
CF.HLF_2D_Rep4	GATK				
	Samtools				
N.HLF_3D_Rep1	GATK	1, KMT2A, exonic	Intellectual disability, autosomal dominant 49, TRIP12	rs60639710, Acute myeloid leukemia with maturation	rs139428292
	Samtools	0	0		0
N.HLF_3D_Rep2	GATK	rs773840580, SACS, Charlevoix-Saguenay_spastic_ataxia	6	rs281864780	rs139428292
	Samtools	Marfan syndrome Familial thoracic aortic aneurysm and aortic dissection Cardiovascular phenotype	rs2073194, rs766233063		0
N.HLF_3D_Rep3	GATK	rs751568153	PAX6, 11p13, TRIP12, frameshift insertion, 2q36.3	rs121918461	rs139428292
	Samtools				
CF.HLF_3D_Rep1	GATK		rs771203308		
	Samtools		rs80358198		
CF.HLF_3D_Rep2	GATK	0	rs764261443		0
	Samtools	rs786204591	0		rs62635763

CF.HLF_3D_Rep3	GATK	CREBBP, Rubinstein-Taybi_syndrome_1, ACTB	rs777766787, Polycystic_liver_disease_2	
	Samtools	FBN1, Marfan syndrome[Familial thoracic aortic aneurysm and aortic dissection, ACTB	0	

		Risk Faktör	****U_S	
N.HLF_2D_Rep1	GATK	rs2234767	7	4
	Samtools		0	
N.HLF_2D_Rep2	GATK		11	12
	Samtools		1	
N.HLF_2D_Rep3	GATK	rs2273535	9	12
	Samtools	0	5	
N.HLF_2D_Rep4	GATK	rs2234767	10	10
	Samtools		4	
CF.HLF_2D_Rep1	GATK		4	6
	Samtools		4	
CF.HLF_2D_Rep2	GATK		8	4
	Samtools		1	
CF.HLF_2D_Rep3	GATK	rs2273535	9	4
	Samtools		6	
CF.HLF_2D_Rep4	GATK	rs2273535	7	11
	Samtools		1	
N.HLF_3D_Rep1	GATK		9	5
	Samtools		rs6991, rs62241769	
N.HLF_3D_Rep2	GATK		10	5
	Samtools		4	
N.HLF_3D_Rep3	GATK		8	11
	Samtools		3	
CF.HLF_3D_Rep1	GATK		7	8
	Samtools		3	
CF.HLF_3D_Rep2	GATK		1	4
	Samtools		3	
CF.HLF_3D_Rep3	GATK		15	7
	Samtools		7	

****Uncertain_significance, CF.HLF_3D_Rep1: Ortak: Benign_risk_factor, rs759330, rs669, A2M: CF.HLF_3D_Rep3 Benign_risk_factor: Samtools: rs669

2D ve 3D örneklerinde GATK ve Samtools arasındaki variant sayıları farkı Tablo 4.5'te gösterilmiştir.

Tablo 4.5 2D ve 3D örneklerinde varyant sayıları ve paket karşılaştırılması

ÖRNEKLER	SAMTOOLS	ORTAK VARYANT SAYISI	GATK
N.HLF_2D_Rep1	1729	7867	5685
N.HLF_2D_Rep2	4031	19036	13407
N.HLF_2D_Rep3	3913	18445	9745
N.HLF_2D_Rep4	2798	13833	7503

CF.HLF_2D_Rep1	2247	10377	7595
CF.HLF_2D_Rep2	2412	11756	7348
CF.HLF_2D_Rep3	2122	10570	6053
CF.HLF_2D_Rep4	3232	16220	9657
N.HLF_3D_Rep1	4274	16160	16218
N.HLF_3D_Rep2	5250	19611	19185
N.HLF_3D_Rep3	6583	25990	25822
CF.HLF_3D_Rep1	6056	23398	25723
CF.HLF_3D_Rep2	6522	26155	23617
CF.HLF_3D_Rep3	9747	44353	34970

Tablo 4.6 2D ve 3D örneklerinde sadece patojenlerin listesi

		PATOJENİK	
N.HLF_2D_Rep1	GATK		
	Samtools		
N.HLF_2D_Rep2	GATK	0	
	Samtools	rs151339003	
N.HLF_2D_Rep3	GATK		
	Samtools		
N.HLF_2D_Rep4	GATK	rs118204046, rs551348450, rs543007243	
	Samtools	rs118204046	
CF.HLF_2D_Rep1	GATK	rs751610886	
	Samtools	0	
CF.HLF_2D_Rep2	GATK	., rs764261443	
	Samtools	0	
CF.HLF_2D_Rep3	GATK	Intellectual_disability, KMT2A	
	Samtools	0	
CF.HLF_2D_Rep4	GATK		
	Samtools		
N.HLF_3D_Rep1	GATK	Intellectual disability, autosomal dominant 49, TRIP12	rs60639710, Acute myeloid leukemia with maturation
	Samtools	0	
N.HLF_3D_Rep2	GATK	6	rs281864780
	Samtools	rs2073194, rs766233063	
N.HLF_3D_Rep3	GATK	PAX6, 11p13, TRIP12, frameshift insertion, 2q36.3	rs121918461
	Samtools		
CF.HLF_3D_Rep1	GATK	rs771203308	
	Samtools	rs80358198	
CF.HLF_3D_Rep2	GATK	rs764261443	
	Samtools	0	
CF.HLF_3D_Rep3	GATK	rs777766787, Polycystic_liver_disease_2	
	Samtools	0	

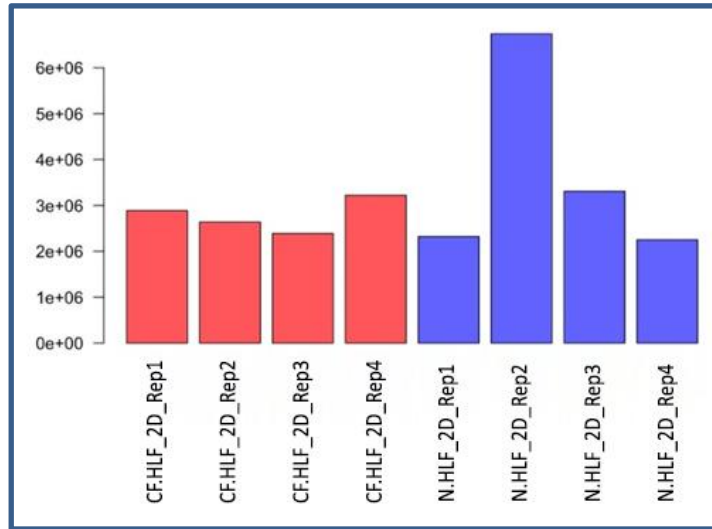
B- Transkriptom Analiz sonuçları:

- GSE141535 numaralı *Transcriptome profile of Primary Human Lung Fibroblasts (Cystic Fibrosis vs. Non-Cystic Fibrosis) [2D]*

EDASeq, GC içerik sapması için şerit içi normalleştirme yöntemlerinin ve şeritler arası normalleştirme yöntemlerinin uygulamalarıyla birlikte, sıralama derinliğini ve muhtemelen dağıtımdaki diğer farklılıkları ayarlamak için verilerin bazı temel özelliklerinin sayısal özetleri ve grafiksel gösterimleri için kullanılır.

Kalite kontrol için önemli olanlar, her şeritte üretilen toplam okuma sayısı, bir referans genomla eşlenen okumaların sayısı ve yüzdesidir. Toplam okuma sayısının düşük olması, girdi RNA'sının düşük kalitesinin bir belirtisi olabilirken, düşük eşleme yüzdesi okumaların kalitesiz olduğunu, referans genomla ilgili sorunları veya yanlış etiketlenmiş şeritleri gösterebilir.

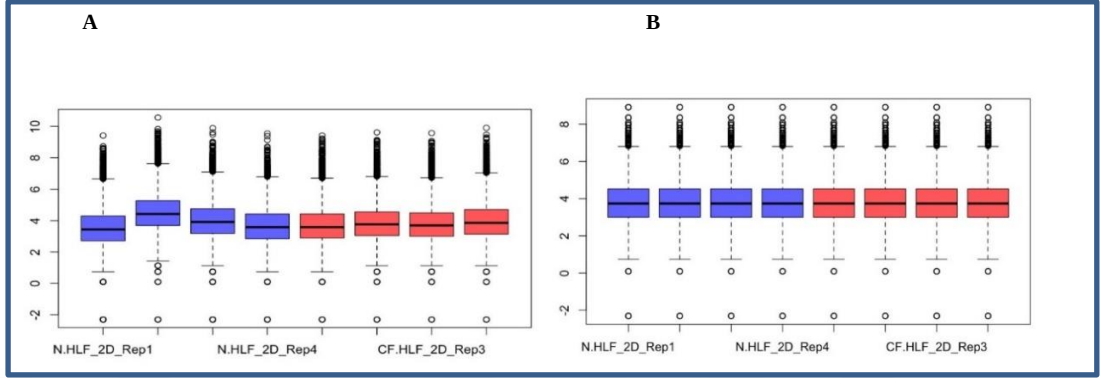
EDASeq analizi için BAM dosyaları kullanılarak aşağıdaki count grafikleri oluşturuldu. *GSE141535* numaralı çalışma'ya ait count grafiği şekil 4.7'de gösterilmektedir.



Şekil 4.7 2D örneklerinde eşlenen okuma sayısı

Gen seviyesi sayımlarının lane arası dağılımı boxplot grafik ile gösterilmektedir. Gen seviyesinde count ile çalışırken göz önünde bulundurulması gereken ana hususlardan biri laneeler arasındaki count dağılımlarındaki farktır. Boxplot

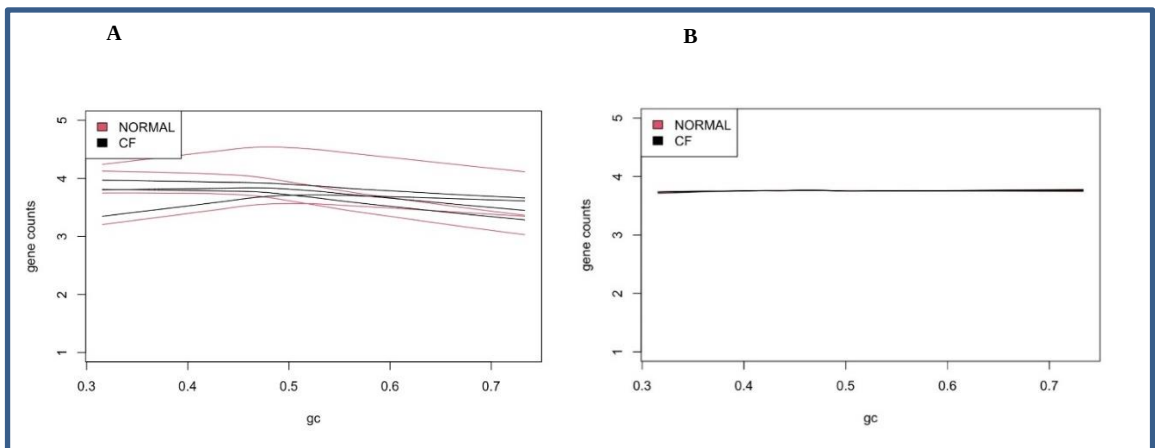
yöntemi, her şeritteki gen sayımlarının logaritmalarının gösterilmesinde kolay bir yoldur. 2D Örneklerle ait normalizasyon öncesi ve sonrası boxplot grafikleri Şekil 4.8’de gösterilmektedir.



Şekil 4.8 EDASeq ile 2D örneklerinde gen seviyeleri gösterimi

Şekil 4.8’de 2D örneklerinde gen seviyeleri ile ilgili olarak **A)** Normalizasyon öncesi gen seviyesi sayımlarının lane arası dağılımı **B)** Normalize gen seviyesi sayımlarının (log) şeritler arası dağılımı.

Okuma sayıları üzerinde gene özgü etkilere bakıldığında araştırmacılar, gen uzunluğu, GC içeriği ve haritalanabilirlik gibi sekans özellikleriyle ilgili seçim bias bildirmiştir. BiasPlot kullanılarak elde edilen gen seviyesi sayımlarının GC içeriğine bağımlılığı göstermektedir (Şekil 4.9).



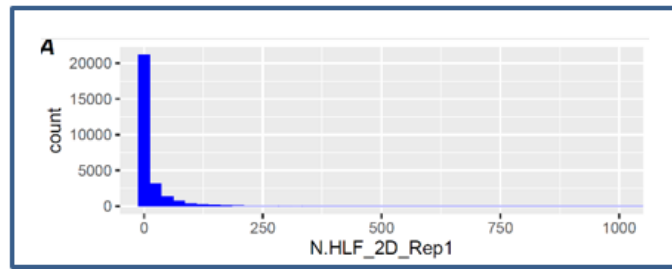
Şekil 4.9 EDASeq ile 2D örneklerinde GC içeriğinde gen seviyesi

2D Örneklerine ait GC içeriğinde gen seviyesi. **A)** normalizasyon öncesi, her şerit için GC içeriği üzerinde gen seviyesi sayımlarının (log) düşük regresyonu ve **B)** Her şerit için GC içeriğinde normalleştirilmiş gen seviyesi sayımlarının (log) düşük regresyonu.

EDASeq paketi kullanılarak elde edilen normalize olmuş sayımlar, diferansiyel olarak eksprese edilmiş genleri bulmak için edgeR veya DESeq gibi paketlere uygulanır. DESeq'in kendi içindeki paketler kullanılarak normalizasyona dayalı count verilerinden DESeq ile diferansiyel ifade analizi yapılabilir (Risso, 2013) Bu tez kapsamında da EDASeq sonrası R Studio'da DESeq uygulaması yapıldı. DESeq ile özet olarak differanssiyel ekspresyon analizini gerçekleştirerek normal ve KF örnekleri arasında ekspresyon açısından hangi genlerin önemli farklılıklara sahip olduğu gösterilmeye çalışıldı. DESeq paketi ile normalize edilen count verileri kullanıldı.

Normalizasyonda sekanslama derinliği gibi farklılıklar ve varyans modellemesinde birkaç replika sayıları ve geniş dinamik ekspresyon aralıkları dikkate alınır. Anlamlılığı test etmek için bu normalizasyon ve varyans modellemesini doğru şekilde uygulayan doğru bir istatistiksel modelle ihtiyaç duyulur. Uygun istatistiksel modeli belirlemek de sayıların dağılımı hakkında bilgiye ihtiyaç duyulur (https://github.com/hbctraining/DGE_workshop_salmon/blob/master/lessons/01_DGE_setup_and_overview.md, 10 Ocak 2022).

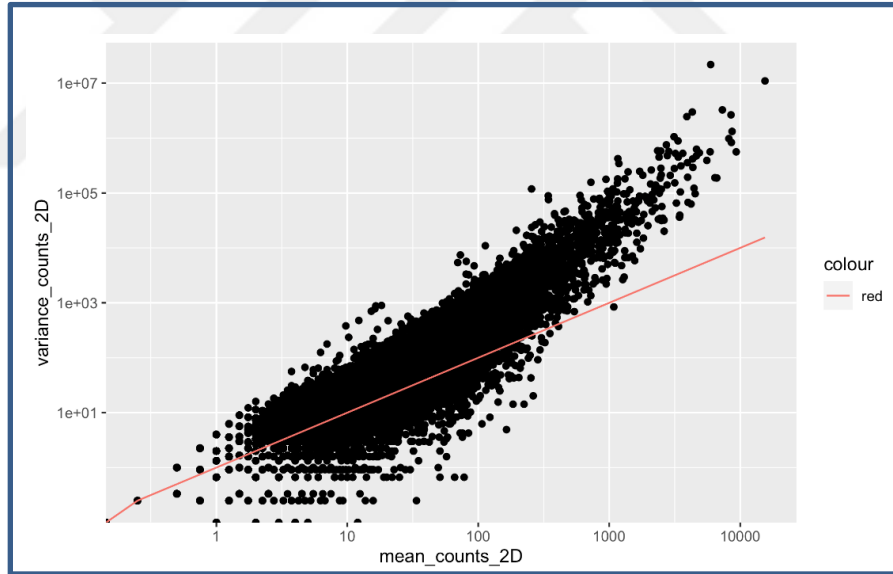
RNAseq ham veri sayımının nasıl dağıldığı hakkında bir fikir edinmek için 2D örneklerinden sadece birine ait grafik DESeq paketi kullanılarak Şekil 4.10'de gösterilmektedir. Grafikte ekspresyon için herhangi bir üst sınırın olmaması nedeniyle uzun bir sağ kuyruk görülmektedir.



Şekil 4.10 DESeq ile 2D örneğinde ham count sonucu

Diferansiyel olarak eksprese edilen genleri belirlemek için, gruplar içindeki varyasyon göz önüne alındığında, gruplar arasında önemli ölçüde farklı ortalama ifadeye sahip olan genlerin tanımlanması gerekmektedir. Grup içindeki (tekrarlar arasındaki) varyasyonun, şekil 4.11’de gösterildiği gibi, ortalama ifade ile varyansın arttığı gerçeğinin hesaba katılmalıdır (her siyah nokta bir gendir). DE genlerini doğru bir şekilde tanımlamak için DESeq2’nin varyans ve ortalama arasındaki ilişkiyi hesaplamaktadır. Düşük ifade edilen genler için varyans daha düşük olacağından tüm DE genlerinin düşük sayımlı genler olması istenmez (https://github.com/hbctraining/DGE_workshop_salmon/blob/master/lessons/04_DE_DESeq2_analysis.md, 10 Ocak 2022).

Sadece 2D KF örneklerine ait ortalama varyans grafiği şekil 4.11’de gösterilmektedir.



Şekil 4.11 2D KF örneklerinde şeritler arasındaki örnek varyansı

Şekil 4.11’de her bir gen için (siyah noktalar) ortalamanın bir fonksiyonu olarak çizilen KF örnekleri verilerinde şeritler arasındaki örnek varyansını göstermekteyiz. Data verilerinin dağılımı ortalamamamanın, varyansa eşit olmadığını göstermektedir. Ortalama ifadesi yüksek olan genler için, replikalar arasındaki varyans ortalamadan daha büyük olma eğilimindedir (dağılım kırmızı çizginin üzerindedir). Not: Kırmızı çizgi $x = y$ ’yi temsil eder.

Bir RNASeq analizinde faydalı bir ilk adım, genellikle numuneler arasındaki genel benzerliği değerlendirmektir. Bunun için sorulacak doğru sorular aşağıdaki gibidir:

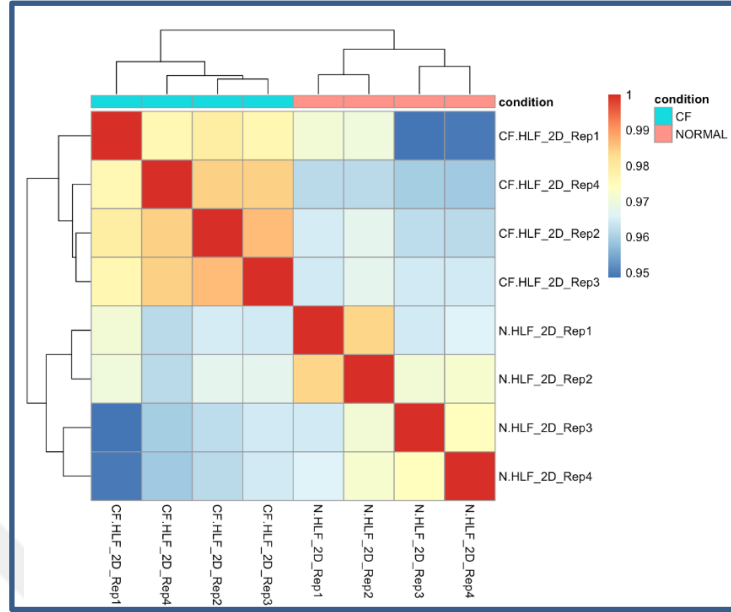
- 1- Hangi örnekler birbirine benzer, hangileri farklıdır?
- 2- Bu çalışanın tasarımından beklenene uymakta mıdır?
- 3- Veri setindeki başlıca varyasyon kaynakları nelerdir?

Örneklerin benzerliğini keşfetmek için PCA ve hiyerarşik kümeleme yöntemleri kullanılarak örnek düzeyinde kalite kontrol gerçekleştirilir. Aynı zamanda biyolojik tekrarların birbirine benzer olduğu hakkında bir fikir edinmek ve veri kümesinde bulunan aykırı değerleri ve ana varyasyon kaynaklarını belirlemek için de bu uygulamalar bize bilgi vermektedir

(https://github.com/hbctraining/DGE_workshop_salmon/blob/master/lessons/03_DGE_QC_analysis.md, 10 Ocak 2022).

Öncelikle Kütüphane boyutu için normalleştirilmiş count verisi ile farklı örnekler arasındaki sayıları karşılaştırdık. Bu şekilde her hangi bir RNASeq çalışmasının kalitesini değerlendirmek için örneklerin gen ifadesi açısından birbirine ne kadar benzer olduğu keşfedilmektedir. Denetimsiz kümeleme (unsupervised clustering) yöntemlerini kullanırken, normalize olmuş sayıların log2 transformasyonu, görselleştirme için mesafeleri/kümelemeyi iyileştirmektedir. RNASeq verileri için DESeq2 tarafından varyans stabilize edici dönüşüm (VST) kullanılır (https://github.com/hbctraining/DGE_workshop_salmon/blob/master/lessons/03_DGE_QC_analysis.md, 10 Ocak 2022).

VST'nin, yaygın olarak kullanılan bir log2 dönüşümü alma yaklaşımıyla karşılaştırıldığında iyi performans gösterdiği gösterilmiştir (Dunning et al., 2008). Özetle, VST ile, kütüphane boyutuna veya diğer normalleştirme faktörlerine göre normalleştirilmiş log2 ölçeğinde dönüştürülmüş veriler üretilir. Çalışmamızda DESeq2 nesnesi üzerinde DESeq2 vst() işlevini kullanarak normalize edilmiş count verisini dönüştürdük. Korelasyon değerlerini oluşturduktan sonra 2D örnekleri için heatmap oluşturduk (Şekil 4.12).

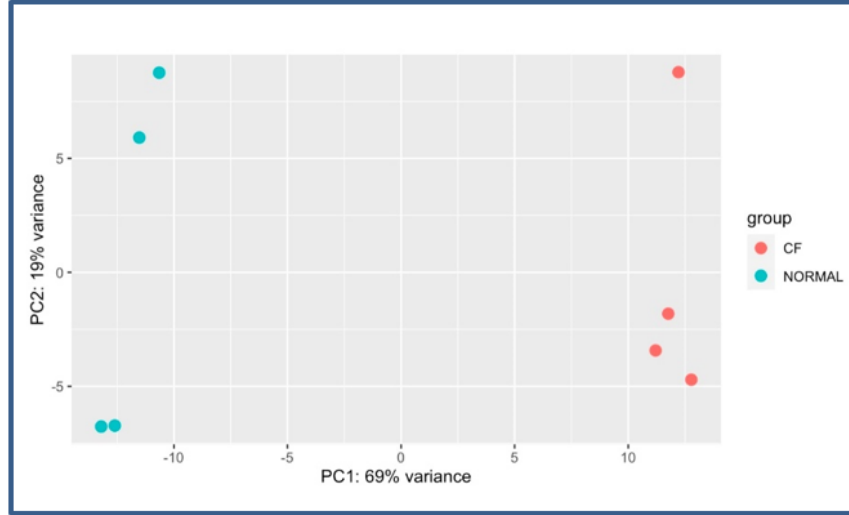


Şekil 4.12 2D KF ve Normal tekrar örneklerinin kümelenmesi

Hiyerarşik kümeleme, bir veri kümesindeki güçlü kalıpları ve olası aykırı değerleri belirlemek için tamamlayıcı yöntemdir. Heatmap, veri kümesindeki tüm ikili örnek kombinasyonları için gen ifadesinin korelasyonunu gösterir. Genlerin çoğu diferansiyel olarak eksprese olmadığından, örnekler genellikle birbirleriyle yüksek korelasyona sahiptir (0.80'den yüksek değerler). 0.80'in altındaki örnekler, verilerde aykırı bir değer ve/veya örnek kontaminasyonu olduğunu gösterebilir (https://github.com/hbctraining/DGE_workshop_salmon/blob/master/lessons/03_DGE_QC_analysis.md, Erişim tarihi: 10 Ocak 2022).

Heatmap'den elde edilen sonuç, normal olan replikaların bir arada kümelendiğini göstermektedir. Bu sonuç bize oluşturduğumuz koşullar arasında farklı şekilde ifade edilen genlerin grup oluşturduğunu ispatlamaktadır.

Daha önce de belirtildiği gibi veri setinde bulunan varyasyonu vurgulamak için PCA kullanılmaktadır. Tez çalışmamızdaki verilerimizde en büyük varyansı temsil eden ilk ana bileşen yani PC1 ile 2D KF ve Normal örneklerin kendi aralarında grup oluşturduğu gözlemlenmiştir (Şekil 4.13).



Şekil 4.13 2D KF ve Normal örnekleri PCA grafiği

İki örnek, PC1 tarafından temsil edilen varyasyona önemli ölçüde katkıda bulunan genler için benzer ekspresyon seviyelerine sahipse, bunlar PC1 ekseninde birbirine yakın olarak ortaya çıkar.

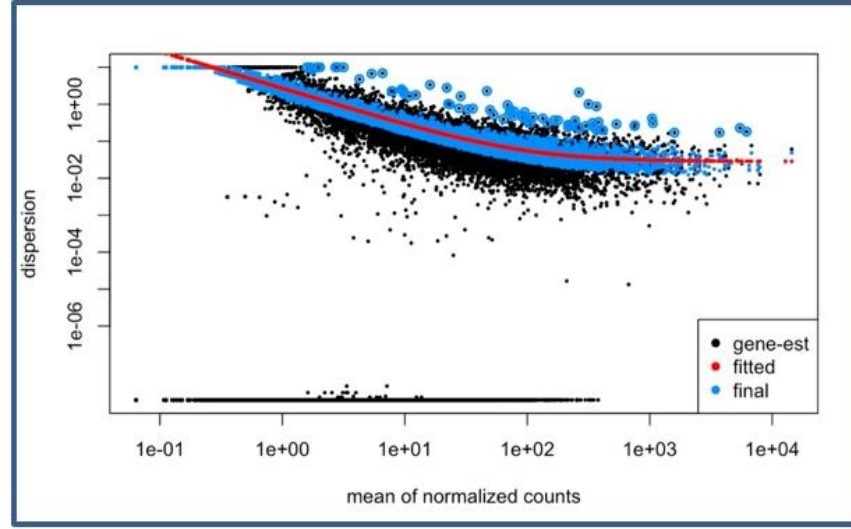
Varyans, gen ekspresyon seviyesi ile ilişkili olduğundan, varyansı, verilerdeki varyasyon ölçüsü olarak kullanmak yerine, bir genin varyansını ve ortalama ekspresyon seviyesini açıklayan *dispersion* adı verilen bir varyasyon ölçüsü kullanır. Dispersion, $Var = \mu + \alpha * \mu^2$ ile hesaplanır. α Dispersionu, $Var =$ varyans ve $\mu =$ ortalama temsil eder. Bunun sonucunda da elde edilen etki Tablo 4.7'de gösterilmiştir.

Tablo 4.7 Varyans dispersion etkisi

	Dağılım üzerindeki etkisi
Varyans artar	Dispersion artar
Ortalama ekspresyon artar	Dispersion azalır

Orta ile yüksek sayım değerlerine sahip genler için, dağılımın karekökü, varyasyon katsayısına (Var/μ) eşit olacaktır. Dolayısıyla 0,01 dağılım, biyolojik kopyalar arasında beklenen ortalama etrafında %10'luk bir değişim anlamına gelmektedir. Aynı ortalamaya sahip genler için dağılım tahminleri, yalnızca varyanslarına göre farklılık gösterecektir. Bu nedenle, dağılım tahminleri, belirli bir ortalama değer için gen ifadesindeki varyansı yansıtır. Aşağıdaki şekilde siyah noktalar olarak temsil edilmektedirler.

2D örneklerine ait artan ortalama ile dağılım düşmesi beklenmektedir. Bu şekil 4.14’de gösterilmektedir.



Şekil 4.14 2D örneklerinde Dispersion grafiği

RNAseq tarafından üretilen count verileri aşırı dağılım (varyans > ortalama) sergiler ve sayıları modellemek için kullanılan istatistiksel dağılımın bu aşırı dağılımı hesaba katması gerekmektedir. DESeq2, aşağıdaki denklem ile RNASeq sayılarını modellemek için negatif bir binom dağılımını kullanır.

raw count for gene i , sample j

The mean is taken as "normalized counts" scaled by a normalization factor

one dispersion per gene

$$K_{ij} \sim \text{NB}(s_{ij}q_{ij}, \alpha_i)$$

Modelleme, bir dizi parametre verildiğinde verilerin nasıl davrandığını tahmin etmenin matematiksel olarak resmileştirilmiş bir yoludur. DESeq2, bu formülü her gen için modelimiz olarak kullanacak ve normalleştirilmiş sayım verilerini buna uyduracaktır. Model uygun hale getirildikten sonra, aşağıdaki formül kullanılarak standart hatalarıyla birlikte her bir örnek grubu için katsayılar tahmin edilmektedir.

$$\log_2 q_{ij} = \sum_r x_{jr} \beta_{ir}$$

Katsayılar, her bir numune grubu için log2 kat değışikliklerinin tahminleridir. Bununla birlikte, bu tahminler, düşük okuma sayılarıyla gözlemlediğimiz büyük dağılımı hesaba katmaz. Bunu önlemek için, model tarafından hesaplanan log2 kat değışikliklerinin ayarlanması gerekir.

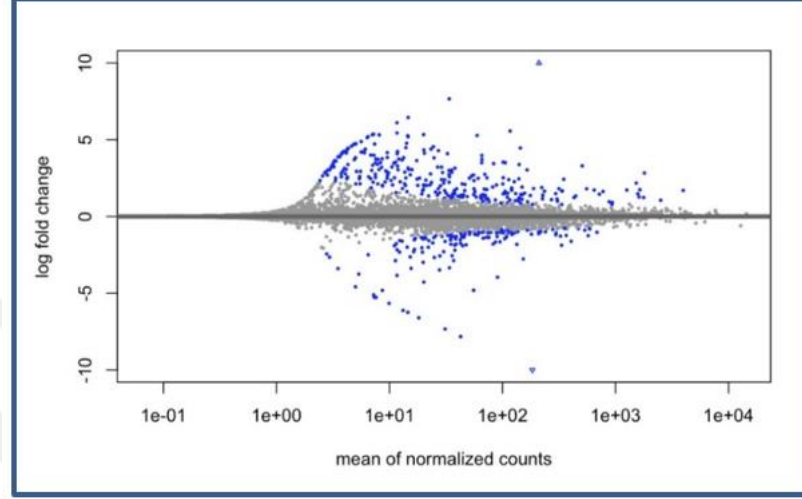
DESeq2, daha doğru log2 kat değışim tahminleri oluşturmak için, bir gen için bilgi düşük olduğunda, LFC (log2 foldchanges) tahminlerinin sıfıra doğru küçülmesine izin verir. Bunlar Düşük count ve yüksek dispersiyon değerleri içerebilirler. LFC shrinkage, daha doğru tahminler oluşturmak için tüm genlerden gelen bilgileri kullanır. Aynı zamanda LFC shrinkage etkisini göstermek için MA plot kullanılmaktadır (Şekil 4.15).

DESeq2 ile Wald testi, iki grubu karşılaştırırken hipotez testi için yaygın olarak kullanılmaktadır. Bir Wald test istatistiğı, en az gözlemlenen değer kadar uç bir test istatistiğinin rastgele seçilmesi olasılığı ile birlikte hesaplanmaktadır. Bu olasılığa testin p değeri denir. Eğer p değeri küçükse, boş hipotezi reddederiz ve sıfıra karşı kanıt olduğu belirtilir yani gen diferansiyel olarak ifade edilir ([https://hbctraining.github.io/DGE_workshop/lessons/05_DGE_DESeq2_analysis2.h](https://hbctraining.github.io/DGE_workshop/lessons/05_DGE_DESeq2_analysis2.html)tml, 10 Ocak 2022).

P-değeri doğrudan Wald testinden $p < 0.05$ anlamlılık sınırı ile kullanılırsa bunun anlamı %5 yanlış pozitif olma ihtimalidir. Her p değeri, tek bir testin (tek gen) sonucudur. Ne kadar çok gen test edilirse yanlış pozitif oranı o kadar fazla artacaktır. Bu bir çoklu test problemidir. Örneğın, 20.000 gen DE için test edilirse, $p < 0.05$ 'te tesadüfen 1.000 gen bulunması beklenir. Toplamda farklı şekilde ifade edilen 3000 gen bulunursa, genlerin yaklaşık üçte biri yanlış pozitif olacaktır ([https://hbctraining.github.io/DGE_workshop/lessons/05_DGE_DESeq2_analysis2.h](https://hbctraining.github.io/DGE_workshop/lessons/05_DGE_DESeq2_analysis2.html)tml, 10 Ocak 2022)

Ayarlanmış p-değeri < 0.05 'e göre, 1070 transkript 2D durumunda özel olarak düzenlenmiştir. Bunlardan 622 tanesi up regüle iken 448 tanesi down regüle olmuştur.

Transkriptler için $lfcThreshold = 0.32$ eşiği ile tekrar analiz edildiğinde 426 transkript 2D durumunda özel olarak yeniden düzenlenmiş oldu. Bunlardan 292 tanesi up regüle iken 134 tanesi down regüle olmuştur. Bu verilerin görselleştirdiği grafik Şekil 4.15’da gösterilmektedir.



Şekil 4.15 2D örneklerinde test edilen tüm genler için log2 kat değişikliklerine karşı normalize olan count verilerinin ortalaması

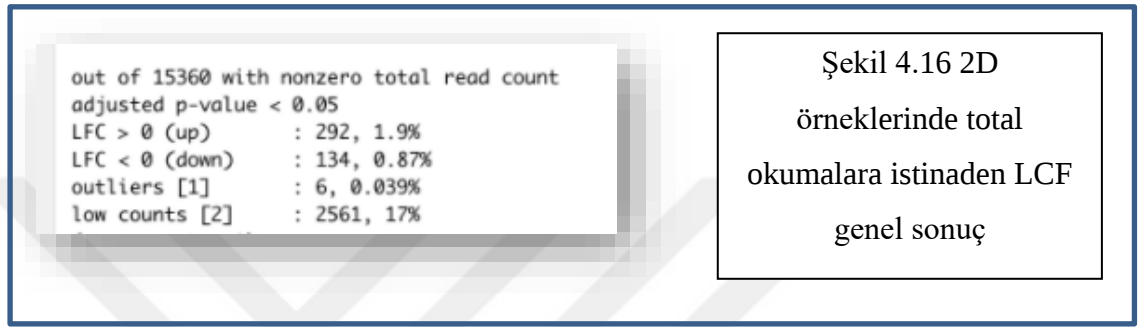
Şekil 4.15’da önemli ölçüde DE olan genler, kolayca tanımlanabilecek şekilde mavi renkli olarak gösterilmektedir. Bu bize kat değişikliklerinin büyüklüğünü ve bunların ortalama ifadeye göre nasıl dağıldığını değerlendirmemizi sağlar. Genel olarak, tüm ekspresyon seviyeleri aralığında önemli genler görmeyi bekleriz.

Farkları görmek adına sonuçlar için ek $lfcThreshold=1$ olarak da değerlendirildi. Ayarlanmış $p\text{-değeri} < 0.05$ 'e göre, $lfcThreshold=1$ için 139 transkript 2D durumunda özel olarak düzenlenmiştir. 111 up regüle iken 28 tanesi down regüle olmuştur.

Tez çalışmamızda sonuçlar, ayarlanmış $p\text{-değeri} < 0.05$ 'e göre, $lfcThreshold=0.32$ eşiği ile değerlendirilmiştir. Elde edilen sonuç Şekil 4.16 ve padj değerine göre önemli ilk 20 gen Tablo 4.8’de gösterilmiştir. Sonuçların en önemli sütunu olan padj sütunu, çoklu test için ayarlanan p değerini temsil etmektedir. Tipik olarak, $padj < 0.05$ gibi bir threshold, önemli genleri tanımlamak için iyi bir başlangıç noktasıdır

(https://github.com/hbctraining/DGE_workshop_salmon_online/blob/master/lessons/05b_wald_test_results.md, 10 Ocak 2022).

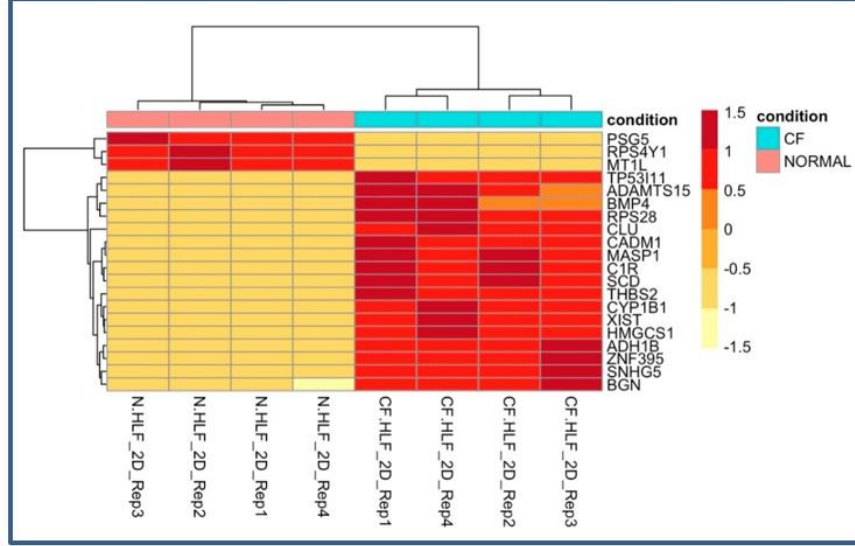
DESeq paketinde bulunan summary fonksiyonu ile örnekler, log2 kat değişiklikleri, standart hatalar, test istatistikleri, p değerleri ve ortalama p değerleri arasında temel ortalamalar veren DESeq analizinden bir sonuç tablosu çıkartmaktadır (Şekil 4.17).



Tablo 4.8 2D örneklerinde Padj değerine göre sıralanmış ilk 20 up ve down regüle DEG

	Genler	baseMean	log2FoldChange	lfcSE	pvalue	padj
1	SNHG5	507.356595	3.297333947	0.196097598	1.35E-52	1.73E-48
2	CLU	1803.30363	2.8280244	0.176776954	2.10E-46	1.34E-42
3	C1R	348.562314	2.421247183	0.154622466	2.01E-43	8.56E-40
4	TP53I11	143.851133	4.459833574	0.32614953	8.64E-38	2.76E-34
5	BMP4	116.764609	5.564339142	0.442046194	8.65E-34	1.85E-30
6	SCD	140.354825	3.385583511	0.255160447	8.31E-34	1.85E-30
7	CYP1B1	105.867816	3.501693331	0.277349857	4.87E-31	8.90E-28
8	CADM1	101.372543	3.540961549	0.286489208	6.56E-30	1.05E-26
9	ADH1B	83.5007805	3.43620725	0.284581222	1.78E-28	2.53E-25
10	PSG5	90.3676649	-3.957677795	0.342239099	4.37E-27	5.59E-24
11	XIST	209.902078	10.57526921	1.104132214	3.32E-26	3.86E-23
12	ZNF395	65.757834	3.725302013	0.326972663	4.97E-26	5.30E-23
13	MASP1	59.5890423	5.275562296	0.486380145	1.37E-25	1.35E-22
14	THBS2	65.4388585	3.502691098	0.310470227	3.03E-25	2.76E-22
15	RPS4Y1	183.997505	-10.16417851	1.10914911	1.99E-24	1.70E-21
16	BGN	165.031827	3.037563113	0.273503318	6.43E-24	5.14E-21
17	ADAMTS15	65.5448748	3.981008637	0.372774234	1.78E-23	1.34E-20
18	MT1L	333.383377	-1.871112323	0.166776083	1.91E-22	1.36E-19
19	HMGCS1	113.646858	2.400024946	0.221626439	2.13E-22	1.40E-19
20	RPS28	1645.00497	1.747484242	0.153088571	2.20E-22	1.40E-19

Tüm önemli genlerin normalleştirilmiş değerleri çıkartılarak pheatmap() ile ekspresyonlarının heatmap'ını oluşturduk. 426 gen arasında ilk 20 gen için padj değerine göre elde edilen heatmap grafiği şekil 4.17'de gösterilmiştir.



Şekil 4.17 2D örneklerinde up ve down regüle ilk 20 genin heatmap gösterimi

2D örneklerinde log2FoldChange değerine göre sıralanmış en iyi 20 up regüle DEG Tablo 4.9'de gösterilmiştir. En düşükten en yükseğe göre sıralanmıştır.

Tablo 4.9 2D örneklerinde log2FoldChange değerine göre sıralanmış en iyi 20 up regüle DEG

	Genler	baseMean	log2FoldChange	lfcSE	pvalue	padj
1	CCDC144A	4.84203483	4.708670339	1.26543938	8.08E-06	0.00056766
2	VAT1L	7.38403569	4.718302131	1.16364007	3.86E-06	0.00030285
3	DES	5.04234588	4.734700359	1.28488114	8.46E-06	0.00059122
4	OLFML1	14.1645379	4.770621706	0.90773931	6.26E-08	7.28E-06
5	CLIC6	5.6193582	4.879329038	1.26847882	5.30E-06	0.00039215
6	LINC00839	6.35615213	5.137765539	1.18274525	1.17E-06	0.00010311
7	SYT1	14.6120455	5.193280318	1.00667244	5.65E-08	6.69E-06
8	COL14A1	6.45338029	5.198229223	1.1610188	7.27E-07	6.74E-05
9	APCDD1	14.5064085	5.255016287	1.00880511	4.19E-08	5.21E-06
10	MASP1	59.5890423	5.275562296	0.48638015	1.37E-25	1.35E-22
11	PCDHGB7	8.10744484	5.316471891	1.2712844	1.49E-06	0.00013074
12	RBP1	7.05518242	5.323969746	1.15194406	3.88E-07	3.82E-05
13	GPR37	20.0339708	5.334325016	0.88690571	7.37E-10	1.33E-07
14	BACH2	7.2338433	5.363434112	1.14357728	2.91E-07	3.02E-05
15	GFRA1	11.6658947	5.439113152	1.06509111	5.10E-08	6.16E-06
16	BMP4	116.764609	5.564339142	0.44204619	8.65E-34	1.85E-30

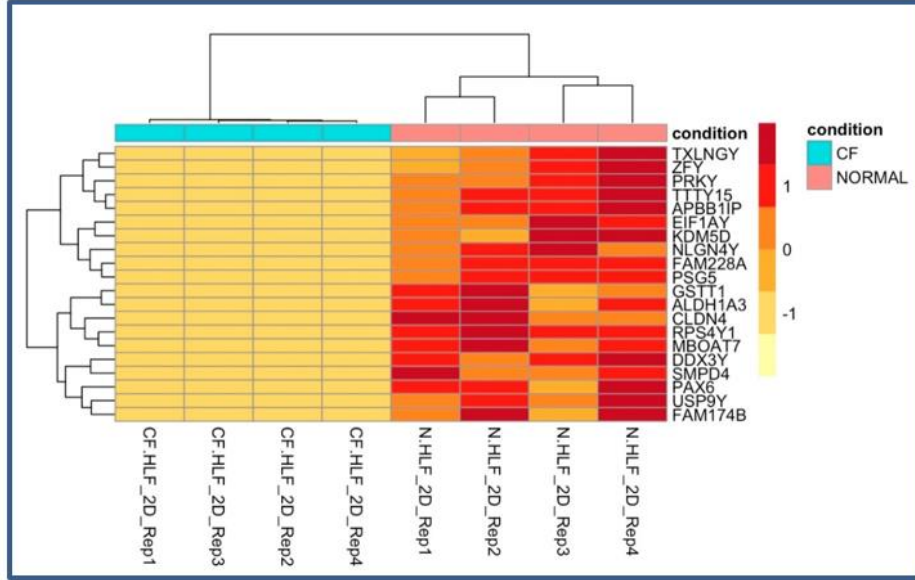
17	SCN3A	11.6556597	6.106358167	1.08091785	1.49E-09	2.51E-07
18	HSPB3	14.634169	6.448399436	1.06607823	1.03E-10	2.23E-08
19	MXRA5	33.7544978	7.66353896	1.06461435	7.17E-15	2.70E-12
20	XIST	209.902078	10.57526921	1.10413221	3.32E-26	3.86E-23

2D örneklerinde log2oldChange değerine göre sıralanmış en iyi 20 down regüle DEG Tablo 4.10'de gösterilmiştir.

Tablo 4.10 2D örneklerinde log2FoldChange değerine göre sıralanmış en iyi 20 down regüle DEG

	Genler	baseMean	log2FoldChange	lfcSE	pvalue	padj
1	RPS4Y1	183.997505	-10.16417851	1.10914911	1.99E-24	1.70E-21
2	EIF1AY	42.6240884	-7.822464086	1.0755437	2.83E-15	1.13E-12
3	GSTT1	31.0847275	-7.329978999	1.08577544	2.85E-13	8.49E-11
4	USP9Y	18.1310771	-6.602191204	1.05922152	2.81E-11	6.91E-09
5	DDX3Y	14.5465612	-6.241684283	1.07658982	5.59E-10	1.05E-07
6	TTY15	13.17723	-6.122971563	1.07785916	1.24E-09	2.14E-07
7	TXLNGY	9.91488336	-5.658854526	1.12190898	4.90E-08	5.97E-06
8	PRKY	7.60121385	-5.271404223	1.13998173	4.10E-07	3.98E-05
9	FAM228A	7.31795621	-5.258408651	1.13094416	3.82E-07	3.79E-05
10	ZFY	7.24034425	-5.107441382	1.21640879	1.81E-06	0.000153434
11	MBOAT7	8.6475083	-4.819661127	1.13893459	2.23E-06	0.000184275
12	SMPD4	55.4558363	-4.812280241	0.49705257	1.50E-20	9.16E-18
13	NLGN4Y	5.00054031	-4.583945528	1.29908144	1.30E-05	0.000845027
14	PAX6	20.1031255	-4.277382754	0.72253218	4.99E-09	7.75E-07
15	PSG5	90.3676649	-3.957677795	0.3422391	4.37E-27	5.59E-24
16	ALDH1A3	11.653549	-3.841972958	0.97215778	9.24E-06	0.000639193
17	FAM174B	5.35068085	-3.75404292	1.5162325	0.00011969	0.005608585
18	APBB1IP	27.7097755	-3.483930548	0.51543895	1.44E-10	2.96E-08
19	KDM5D	3.50396449	-3.390839967	1.790131	0.000305	0.012466126
20	CLDN4	15.5233694	-3.37695892	0.66339941	3.35E-07	3.38E-05

426 gen arasında log2FoldChange değerine göre en az down regüle ilk 20 gen için elde edilen heatmap grafiği Şekil 4.18'da gösterilmektedir.



Şekil 4.18 2D örneklerinde en az down regüle olan 20 genin heatmap gösterimi

2D örneklerinde log2FoldChange değerine göre en az down regüle olan tüm genler soldan sağa sıralanarak Tablo 4.11’de gösterilmiştir.

Tablo 4.11 2D örneklerinde log2FoldChange değerine göre en az down regüle olan genden itibaren tüm genler

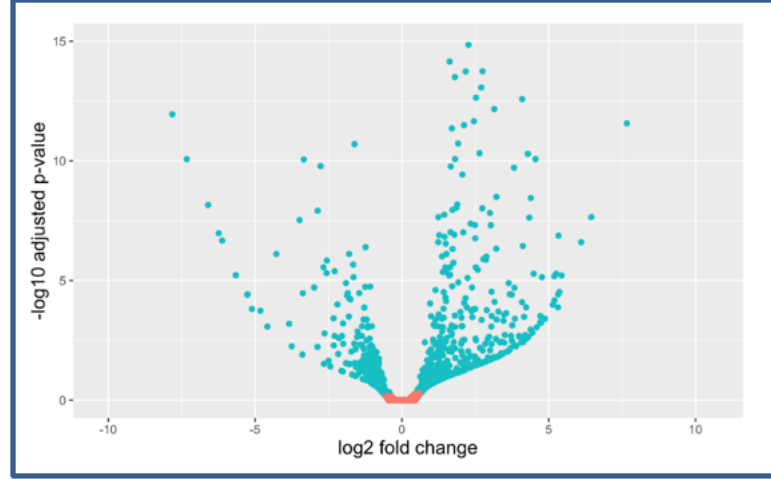
RPS4Y1	EIF1AY	GSTT1	USP9Y	DDX3Y	TTY15	TXLNGY	PRKY	FAM228A	ZFY
MBOAT7	SMPD4	NLGN4Y	PAX6	PSG5	ALDH1A3	FAM174B	APBB1P	KDM5D	CLDN4
VAMP8	CCK	MMP3	CHST15	PLAU	TCF21	RNF5	DDIT4L	SNAP25	DHRS3
GRAMD1B	HSD17B2	PRKG2	THOC3	HXA3	WNK4	PITX1	AMPH	ADAP2	C1QTNF2
RPS6KA6	CDCA7	POLR3G	CCND2	MT1L	ID3	DNMBP	RPS17	ID1	PSG4
LRRN4CL	WNT16	INSYN2B	LCLAT1	PLAT	GPRC5A	PKMYT1	S100A4	PCOTH	FAM162B
TIPIN	RDH10	MACIR	LIMCH1	DNASE1L1	CRLF1	TSC22D2	MELK	PLPP2	CCL2
EIF5A2	PCNA	EPB41L3	KNSTRN	EPHB2	MFSD1	PLAUR	IMPA1	TYMS	SNHG17
DKK1	SOAT1	HAUS6	FAM107B	STK32C	CPED1	DCBLD2	BUB3	CACNA1A	WASF3
CDH6	TGM2	TNFRSF12A	STC1	WNT5B	ZWINT	TOX2	AOX1	CDK2	AK5
BUB1	PRDX3	FLNC	SLC2A1	ANXA7	BLOC1S2	GNB2	F2RL2	STAMBPL1	BNIP3
SDC1	DNAJC9	HELLS	RAI14	PAK1	FGF5	NIPA2	GTPBP4	NMT2	CDCP1
HNRNPH3	MPHOSPH6	ARHGAP22	REEP3	LSM4	GSTO1	GNG11	TRPA1	HNRNPF	DRAP1
ANXA5	NAV1	EMP1	RRM1						

2D örneklerinde log2FoldChange değerine göre en az up regüle olan tüm genler soldan sağa sıralanarak Tablo 4.12’de gösterilmektedir.

Tablo 4.12 2D örneklerinde log2FoldChange değerine göre en az up regüle olan genden itibaren tüm genler

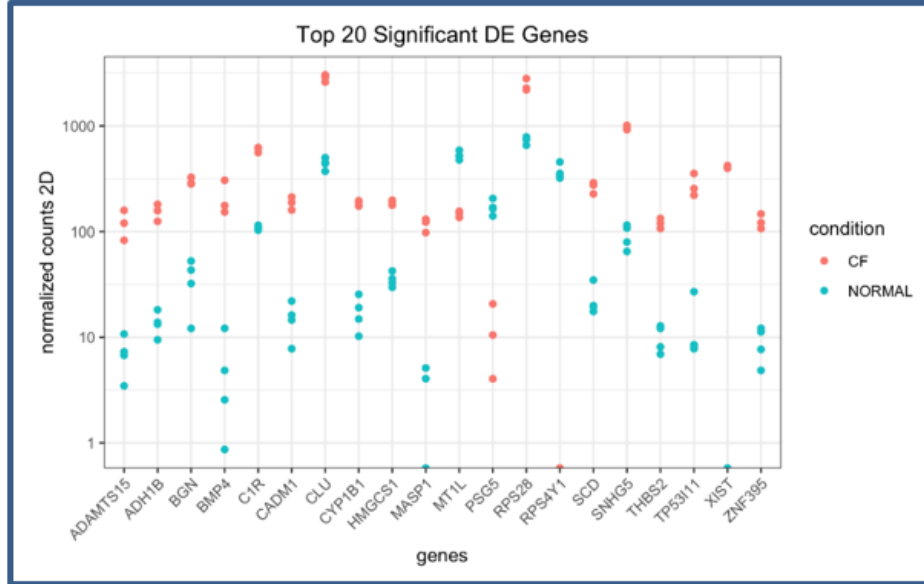
COL8A1	FST	ARL4C	MYL6B	PRDM8	FDFT1	IGFBP4	COL6A3	TGFB1I1	RNASEH2C
UBE2H	NTM	RGS4	PDE5A	ARHGAP29	RPS4X	SV2A	PDGFRB	MMP2	NUPR1
IRF2BPL	P3H4	ZNF503	LDLR	C14orf132	NFIL3	C1QTNF1	C8orf82	PTPN13	PYCR1
CEBPD	COL1A2	DPYSL3	SOD2	ADAM12	GSTM4	PYCARD	MYLK	EPB41L1	FCGRT
IGF1R	LMOD1	SREBF2	ARSD	AKAP12	TRIB2	LOC730101	MSMO1	IGFBP7	FADS1
AEBP1	FBXO32	MEGF8	NNMT	NREP	SPOCK1	RGS2	SQLE	KCNQ1OT1	EVL
C1S	NEAT1	HRH1	SDC2	PDE1C	OLFML3	DCN	ANGPTL2	CXCL12	PARD3B
GYPC	CXXC5	PRELP	SEMA5A	RNASET2	HR	DEPP1	RAC2	CCN3	THY1
EHD3	AQP1	SOX4	COL3A1	DHCR24	QPRT	SERPING1	TCF7L1	EFHC1	POSTN
COL1A1	INSIG1	SCARA3	COL6A1	RPS28	H6PD	HLX	RUNX1T1	NEDD4L	PCOLCE
GDF5	PTN	NACAD	MGARP	SYNJ2	CEMIP	SLC25A23	COL12A1	COL5A1	HLA-C
SESN3	CD4	CARMIL1	POU2F2	ARHGDIB	JUP	CCDC8	MN1	CPM	PIM1
CA12	PTGFR	GJA1	HLA-A	SELENOP	GREM2	ADGRL2	SHANK2	ZNF219	SOBP
IGFBP5	FYN	MMP1	FAM89A	L1CAM	COL11A1	KRT19	ADM2	TBX4	SUSD3
SLC22A17	EFNA5	EPHB3	SPECC1	HMGCS1	C1R	DTNA	PTX3	ELFN2	DPP4
EPB41	ZNF385D	SSC5D	DIO2	GSTM3	GPR1	RELN	SERPINE2	ANK2	HTRA1
PTPRU	WNT2	LINC00484	SCN7A	FADS2	VAMP1	PLXDC2	SYNPO2	MAOA	ENOX1
NPTX1	AKR1C2	PCDH7	LUM	NBEA	CSRP2	GALNT16	BMF	CLU	HHIP-AS1
MFAP4	FBN2	SEMA3B	GBP2	SEMA6A	PLXNA4	SLC9A3-AS1	SYT7	LPIN3	TMOD1
ITGB8	ZNF608	OAS3	TSPAN2	BAALC	LPAR4	BGN	FRG1BP	NTN4	SLITRK6
NOVA1	DOK6	DGKI	VCAM1	TRO	FGF9	RHOJ	SULF1	GSTM5	CCN5
ZMAT1	MIAT	SNHG5	PTPN3	GUCY1B1	SCD	FOXO1	PODN	RBPM52	ADH1B
HTATSF1P2	MEST	ZNF853	CYP1B1	THBS2	SOX6	CADM1	SEMA3D	ARHGAP20	CGNL1
IL32	KIT	SOX11	PROS1	SETBP1	ZNF395	MAOB	IL21R	RAI2	MIR4458HG
TRIM55	HOXC8	LRRC32	SNCAIP	ITGA11	ADAMTS15	NLGN1	PDGFRL	LOC101929130	FOXE1
MAP2	MGP	LINC02550	MACROH2A2	TRIM58	ZDBF2	USP32P1	ACSS3	HTRA3	PRUNE2
PSD4	FRG1CP	GRIN2A	SNED1	GAP43	ROR1	LYN	RORB	TP53I11	LINC01133
CSNK2B	JAM2	CCDC144A	VAT1L	DES	OLFML1	CLIC6	LINC00839	SYT1	COL14A1
APCDD1	MASP1	PCDHGB7	RBP1	GPR37	BACH2	GFRA1	BMP4	SCN3A	HSPB3
MXRA5	XIST								

Yukarıdaki şekillere ek olarak daha fazla global bir görünüm için Volcano Plot kullandık. 426 gen için volcano plot Şekil 4.19’de gösterilmektedir.



Şekil 4.19 2D örneklerinde volcano plot gösterimi

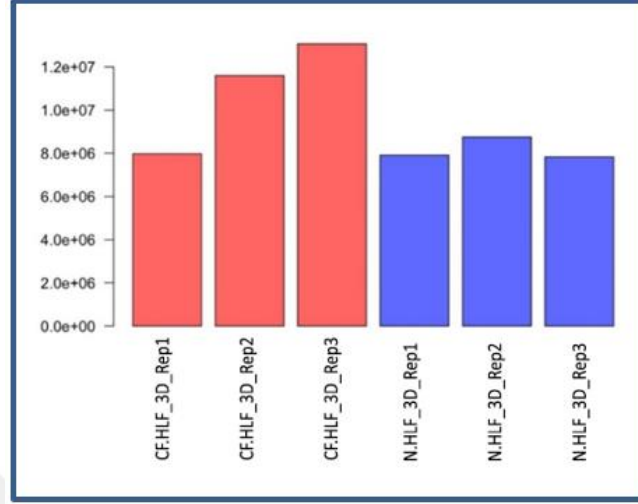
Çoğu zaman, aynı anda birden fazla ilgili genin ifadesini kontrol etmek önemlidir. Diferansiyel olarak ifade edilen ilk 20 gen için (padj değerleriyle) normalleştirilmiş count verilerinin grafiği Şekil 4.20’de gösterilmektedir. Bunun için, önce sonuçlarımızı sıralayarak ve ilk 20-genimizin gen isimlerini belirledik (padj değerlerine göre).



Şekil 4.20 2D örneklerinde Diferansiyel olarak ifade edilen ilk 20 gen

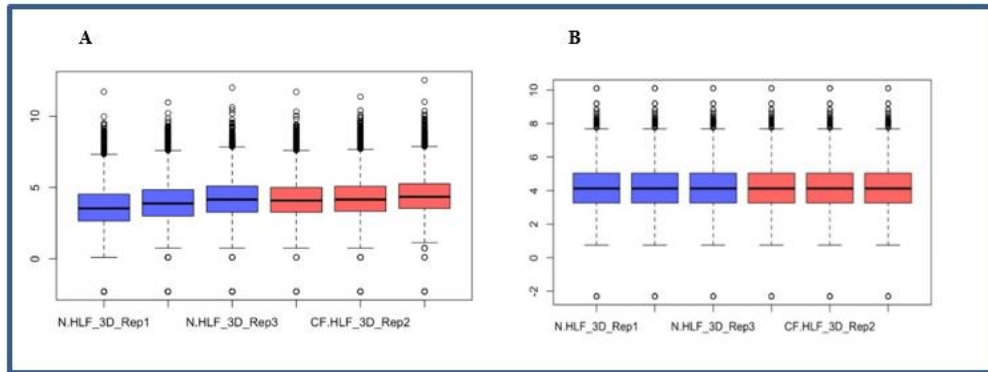
- GSE141536 numaralı *Transcriptome profile of Primary Human Lung Fibroblasts (Cystic Fibrosis vs. Non-Cystic Fibrosis) arranged in a 3D model of the

connective airways tissue* EDASeq analizi için BAM dosyaları kullanılarak Şekil 4.21’de gösterilen count grafikleri oluşturuldu.



Şekil 4.21 EDASeq ile 3D örneklerinde eşlenen okuma sayısı

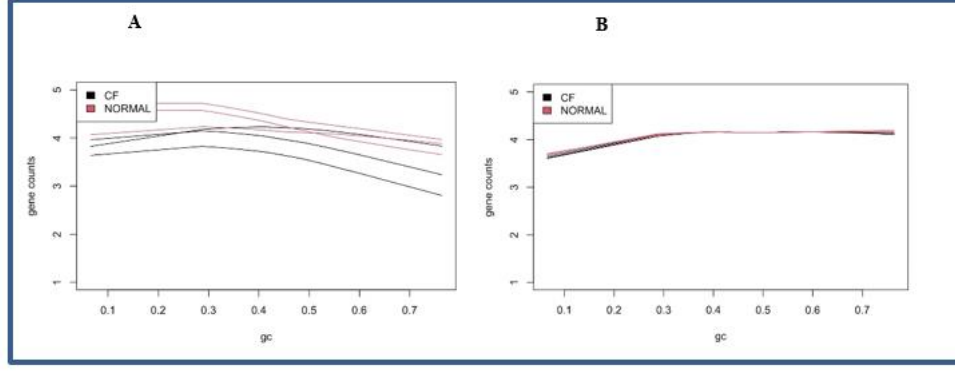
Örneklere ait normalizasyon öncesi ve sonrası boxplt grafikleri Şekil 4.22’te gösterilmektedir.



Şekil 4.22 EDASeq ile 3D örneklerinde gen seviyelerinin gösterimi

3D örneklerinde **A)** Normalizasyon öncesi gen seviyesi sayımlarının lane arası dağılımı **B)** Normalize gen seviyesi sayımlarının (log) şeritler arası dağılımı.

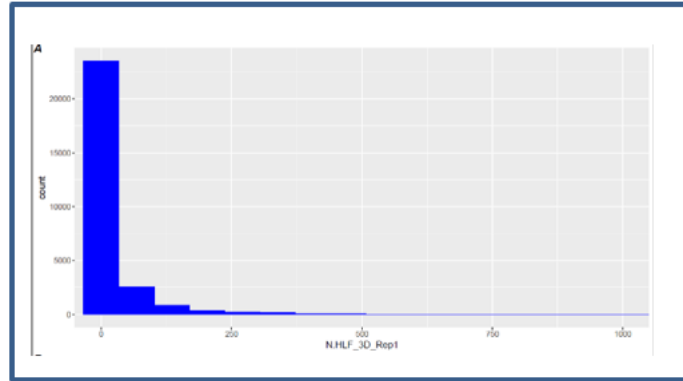
3D örneklerine ait normalizasyon öncesi ve sonrası biasPlot grafikleri Şekil 4.23’da gösterilmektedir.



Şekil 4.23 EDASeq ile 3D Örneklerinde GC içeriğinde gen seviyesi

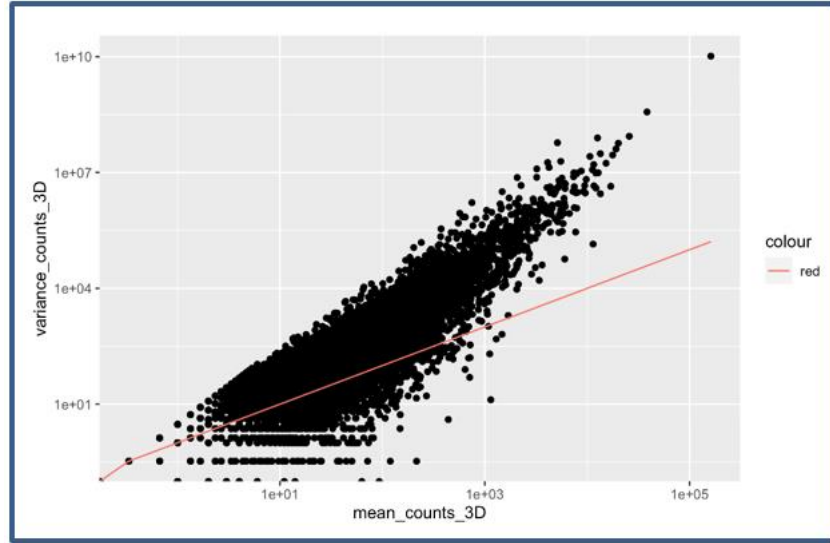
3D Örneklerine ait **A)** Normalizasyon öncesi, her şerit için GC içeriği üzerinde gen seviyesi sayımlarının (log) düşük regresyonu ve **B)** Her şerit için GC içeriğinde normalleştirilmiş gen seviyesi sayımlarının (log) düşük regresyonu.

DESeq paketi kullanılarak bir adet 3D örneğine ait ham veri count dağılımı şekil 4.24’da gösterilmektedir.



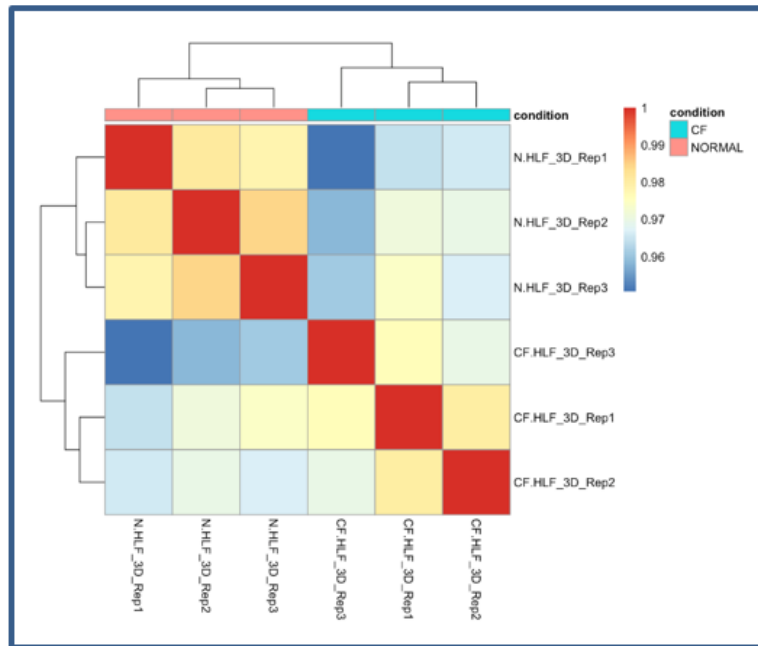
Şekil 4.24 DESeq ile 3D örneğine ait ham count sonucuna bir örnek

Sadece 3D KF örneklerine ait mean variance grafiği şekil 4.25’da gösterilmektedir. Burada, her bir gen için (siyah noktalar) ortalamanın bir fonksiyonu olarak çizilen 3D KF örnekleri verilerinde şeritler arasındaki örnek varyansını gösteriyoruz. Data verilerinin dağılımı ortalamamamanın, varyansa eşit olmadığını gösteriyor. Ortalama ifadesi yüksek olan genler için, replikalar arasındaki varyans ortalamadan daha büyük olma eğilimindedir (dağılım kırmızı çizginin üzerindedir). Not: Kırmızı çizgi $x = y$ ’yi temsil etmektedir.



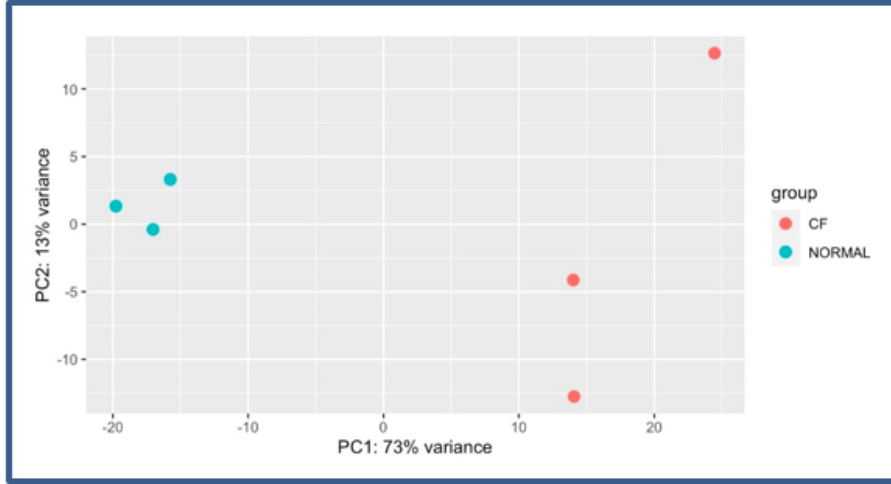
Şekil 4.25 3D KF örneklerinde şeritler arasındaki örnek varyansı

3D örnekler için, korelasyon değerlerini oluşturduktan sonra, elde edilen şekil 4.26'de gösterilmektedir.



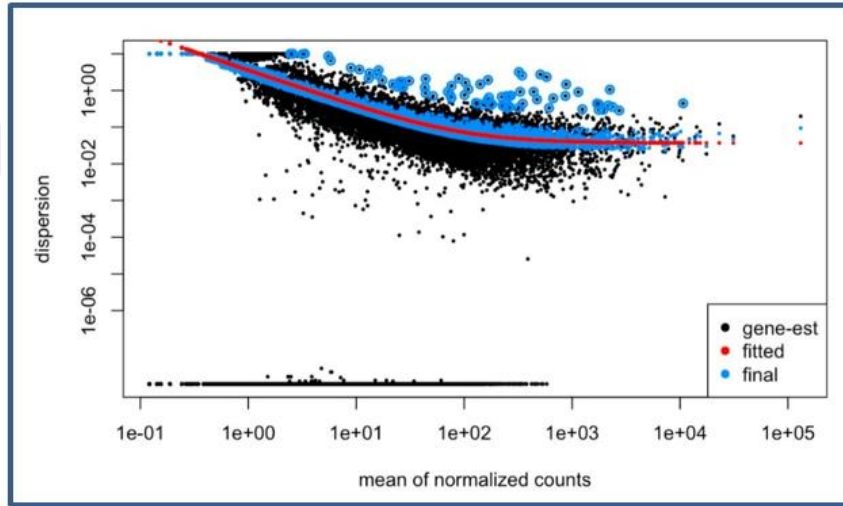
Şekil 4.26 3D KF ve Normal tekrar örneklerinin kümelenmesi

3D KF ve Normal örneklerini kendi aralarında oluşturduğu grubaşmaya ait PCA grafiğı şekil 4.27'de gösterilmektedir.



Şekil 4.27 3D KF ve Normal örnekleri PCA grafiği

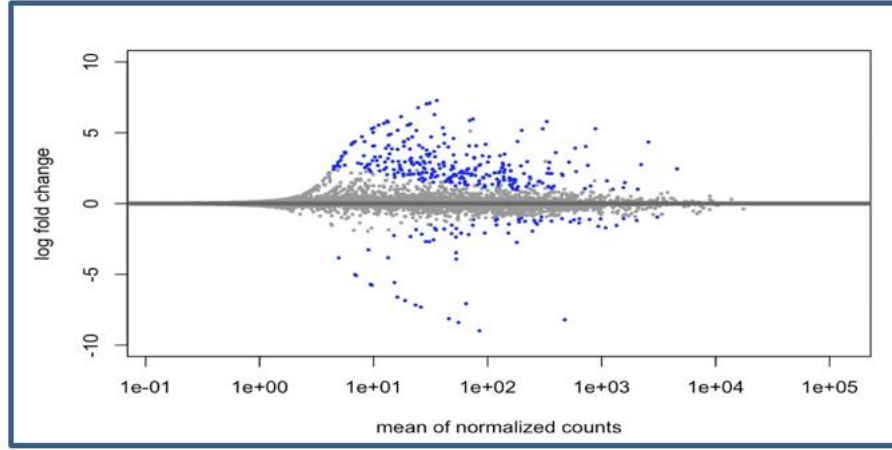
Artan ortalama ile dağılım düşmesi beklenmektedir. 3D örneklerine ait ayırım grafik şekil 4.28’da gösterilmektedir.



Şekil 4.28 3D örneklerinde Dispersion grafiği

Ayarlanmış p-değeri < 0.05 'e göre, 665 transkript 3D durumunda özel olarak düzenlenmiştir. Bunlardan 458 tanesi up regüle iken 207 tanesi down regüle olmuştur.

Transkriptler için Ayarlanmış p-değeri < 0.05 lfcThreshold=0.32 eşiği ile tekrar analiz edildiğinde 352 transkript 3D durumunda özel olarak yeniden düzenlenmiş oldu. Bunlardan 282 tanesi up-regüle iken 70 tanesi down regüle olmuştur. Bu verilerin görselleştirdiği grafik şekil 4.29’da gösterilmektedir.



Şekil 4.29 3D örneklerinde test edilen tüm genler için log2 kat değişikliklerine karşı normalize olan count verilerinin ortalamasının gösterimi

Şekil 4.29’da önemli ölçüde DE olan genler, kolayca tanımlanabilecek şekilde mavi renkli olarak gösterilmektedir.

DESeq paketinde bulunan summary fonksiyonu ile 3D örneklerine ait genel sonuç şekil 4.30’de gösterilmektedir.

```

out of 15817 with nonzero total read count
adjusted p-value < 0.05
LFC > 0 (up)      : 282, 1.8%
LFC < 0 (down)    : 70, 0.44%
outliers [1]      : 13, 0.082%
low counts [2]    : 3821, 24%
(mean count < 2)
[1] see 'cooksCutoff' argument of ?results
[2] see 'independentFiltering' argument of ?results

```

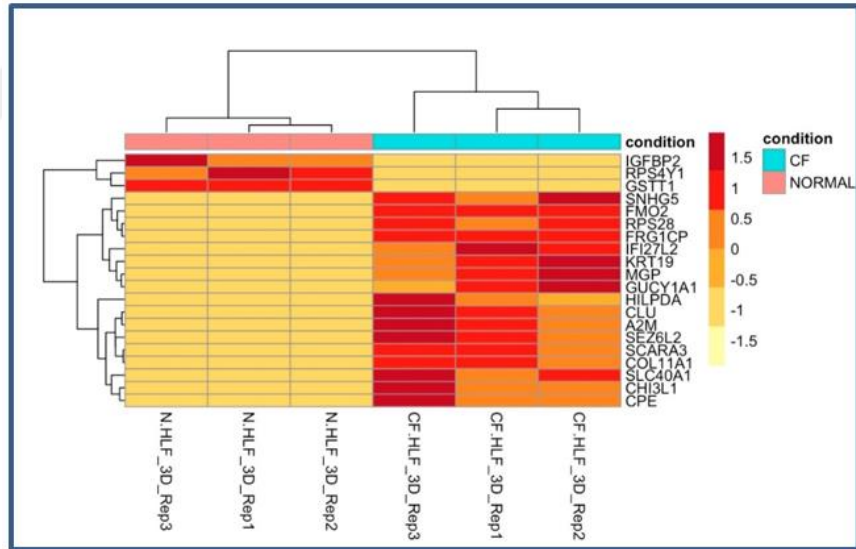
Şekil 4.30 3D örneklerinde total okumalara istinaden LCF genel sonuç

Tablo 4.13 3D örneklerinde Padj değerine göre sıralanmış ilk 20 up ve down regüle DEG

	Genler	baseMean	log2FoldChange	lfcSE	pvalue	padj
1	CLU	2572.80368	4.338926544	0.25927065	2.56E-55	3.06E-51
2	RPS4Y1	475.490466	-8.207654394	0.52445149	6.11E-53	3.66E-49
3	CHI3L1	304.673482	5.278654929	0.41827148	1.53E-33	6.11E-30
4	IFI27L2	199.209286	5.161546835	0.41560405	1.78E-32	5.33E-29
5	SCARA3	395.352066	3.600932102	0.28569716	9.66E-32	2.31E-28
6	MGP	726.513711	4.031261754	0.33470305	9.19E-30	1.83E-26
7	CPE	183.335423	4.382761194	0.38553834	4.07E-27	6.96E-24

8	RPS28	4584.742	2.4471787	0.21349283	3.86E-24	5.79E-21
9	SNHG5	757.373569	2.693549977	0.25450703	1.38E-21	1.83E-18
10	HILPDA	585.932083	3.906536753	0.39228633	3.71E-21	4.44E-18
11	A2M	2220.16822	2.755215662	0.2704289	2.49E-20	2.71E-17
12	FMO2	342.84724	2.484388922	0.24805435	4.38E-19	4.38E-16
13	KRT19	96.1450454	4.179999174	0.44994847	6.07E-19	5.59E-16
14	FRG1CP	69.6710255	5.866016146	0.71369681	4.02E-16	3.44E-13
15	GSTT1	84.7990869	-8.98226273	1.17129195	9.54E-16	7.62E-13
16	GUCY1A1	79.4020803	4.080240859	0.50509702	5.83E-15	4.36E-12
17	SLC40A1	154.061288	2.721871149	0.32114575	8.98E-15	6.33E-12
18	COL11A1	68.9557071	4.071256032	0.51243871	1.47E-14	9.81E-12
19	IGFBP2	180.380009	-2.74589356	0.33562424	5.69E-14	3.59E-11
20	SEZ6L2	90.3999354	3.37645568	0.42909653	6.93E-14	4.15E-11

Tüm önemli genlerin normalleştirilmiş değerleri çıkartılarak pheatmap() ile ekspresyonlarının heatmapını çizdik. 352 gen arasında ilk 20 gen için padj değerine göre elde edilen heatmap grafiği şekil 4.31’de gösterilmektedir.



Şekil 4.31 3D örneklerinde up ve down regüle ilk 20 genin heatmap gösterimi

3D örneklerinde log2FoldChange değerine göre sıralanmış en iyi 20 up regüle DEG Tablo 4.14’de gösterilmektedir. Genler en düşüğe en yükseğe göre sıralanmıştır.

Tablo 4.14 3D örneklerinde log2FolChange değerine göre sıralanmış en iyi 20 up regüle DEG

	Genler	baseMean	log2FoldChange	lfcSE	pvalue	padj
1	CHI3L1	304.673482	5.278654929	0.41827148	1.53E-33	6.11E-30
2	ZNF385D	10.0585624	5.348428628	1.43105087	5.57E-06	0.00045724
3	SLITRK6	40.5334996	5.352398593	0.839319	1.03E-10	2.68E-08
4	SGCA	19.7891562	5.526436956	1.2921745	1.21E-06	0.00012608
5	HSPB3	11.069108	5.53404133	1.37821598	2.52E-06	0.00023598
6	CES1	20.4723337	5.57028056	1.28737265	1.01E-06	0.00010863
7	EGFL6	21.0316463	5.62274743	1.28064333	7.92E-07	9.22E-05
8	PRUNE2	12.3029149	5.654350509	1.39080687	2.03E-06	0.00019779
9	SYT1	13.3626641	5.745193289	1.39317321	1.62E-06	0.00016085
10	XIST	330.16304	5.787348463	1.17035479	9.02E-08	1.24E-05
11	HOXC6	13.1317178	5.827563791	1.32021197	6.37E-07	7.48E-05
12	FRG1CP	69.6710255	5.866016146	0.71369681	4.02E-16	3.44E-13
13	EGLN3	74.3394658	5.958175446	1.30830535	3.60E-07	4.59E-05
14	MCHR1	17.4119034	6.133773748	1.36184042	3.76E-07	4.70E-05
15	APLN	34.2138343	6.278320322	1.27255689	6.59E-08	9.30E-06
16	TRIM55	24.6011826	6.76546955	1.24674874	4.89E-09	8.87E-07
17	H19	29.3325493	7.027382123	1.23235745	9.93E-10	2.05E-07
18	AOC3	29.1823848	7.0427215	1.21810647	6.51E-10	1.42E-07
19	RORB	31.126218	7.088883428	1.2413187	9.00E-10	1.89E-07
20	GAP43	35.8041522	7.274542427	1.23935617	3.33E-10	7.68E-08

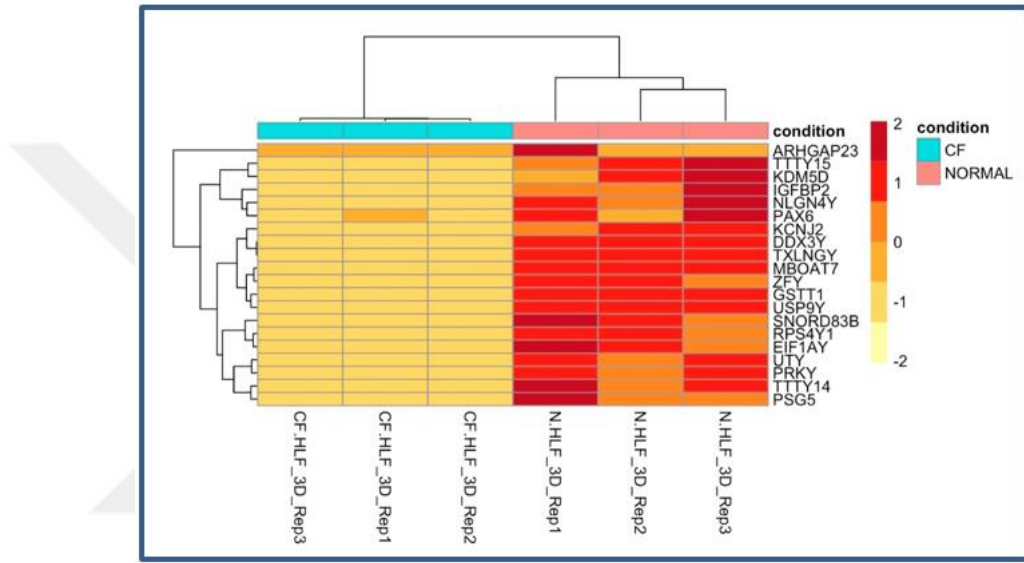
3D örneklerinde log2FoldChange değerine göre sıralanmış en iyi 20 down regüle DEG tablo 4.15’de gösterilmektedir.

Tablo 4.15 3D örneklerinde log2FoldChange değerine göre sıralanmış en iyi 20 down regüle DEG

	Genler	baseMean	log2FoldChange	lfcSE	pvalue	padj
1	GSTT1	84.7990869	-8.98226273	1.17129195	9.54E-16	7.62E-13
2	USP9Y	55.5213659	-8.400490721	1.18256203	7.74E-14	4.42E-11
3	RPS4Y1	475.490466	-8.207654394	0.52445149	6.11E-53	3.66E-49
4	DDX3Y	45.7036333	-8.130244399	1.19054994	5.80E-13	2.58E-10
5	TTY15	25.9787782	-7.313529147	1.23290545	2.31E-10	5.64E-08
6	TXLNGY	23.3783849	-7.169760405	1.22976546	4.51E-10	1.00E-07
7	EIF1AY	64.8199844	-7.070889375	1.05493883	3.45E-12	1.25E-09
8	MBOAT7	18.886166	-6.847487048	1.24968124	3.53E-09	6.82E-07
9	ZFY	16.1762923	-6.608625478	1.26898276	1.54E-08	2.68E-06
10	UTY	9.67148369	-5.772328319	1.38433771	1.40E-06	0.00014221
11	NLGN4Y	9.42081821	-5.707548753	1.41080462	2.06E-06	0.00019794
12	PRKY	15.229138	-5.577130036	1.31933158	1.36E-06	0.00013907
13	KDM5D	7.03112919	-5.079026048	1.63729585	2.16E-05	0.00145521
14	TTY14	6.87062405	-5.024818428	1.63628028	2.36E-05	0.00155379
15	KCNJ2	53.0901164	-3.923222024	0.51846552	2.14E-13	1.07E-10
16	SNORD83B	4.95650601	-3.83768236	2.10836024	0.00025341	0.01193291

17	PAX6	13.4121645	-3.827226907	1.08770009	2.35E-05	0.00155353
18	PSG5	52.8403101	-3.472642203	0.48435332	4.76E-12	1.63E-09
19	ARHGAP23	8.98892476	-3.267470028	2.24095565	0.0005358	0.02206336
20	IGFBP2	180.380009	-2.74589356	0.33562424	5.69E-14	3.59E-11

352 gen arasında log2FoldChange’e göre en az down regüle ilk 20 gen için elde edilen heatmap grafiği Şekil 4.32’de gösterilmektedir.



Şekil 4.32 3D örneklerinde en az down regüle olan 20 genin heatmap gösterimi

3D örneklerinde log2FoldChange değerine göre en az down regüle olan tüm genler soldan sağa sıralanarak Tablo 4.16’da gösterilmiştir.

Tablo 4.16 3D örneklerinde log2FoldChange değerine göre en az down regüle olan genden itibaren tüm genler

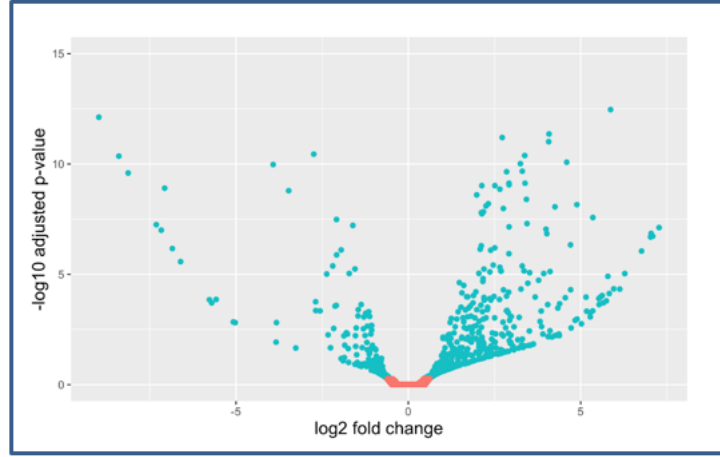
GSTT1	USP9Y	RPS4Y	DDX3Y	TTY15	TXLNGY	EIF1AY	MBOAT7	ZFY	UTY
NLGN4Y	PRKY	KDM5D	TTY14	KCNJ2	SNORD83B	PAX6	PSG5	ARHGAP23	IGFBP2
SMPD4	SYNRG	CAMK2B	CLDN11	RIMS1	TENT5C	KCNMA1	CHST15	VAMP8	RASSF2
PITX1	EDN1	CRLF1	SNAP25	SNHG17	LRRC2	CACNA1A	DKK1	LMO7	CCND2
RAB3B	F3	STK32C	EPHB2	SEC11C	SLC14A1	BDNF	SMURF2	SLC7A11	SVIL
TNFRSF11B	ABCC4	GCLM	RDH10	RAI14	ALPK2	SCN9A	TXNRD1	ATP8B1	PALLD
PVR	LOC728392	SERPINE1	FGF2	NMT2	CDKN1A	SLC4A7	CYB5A	UACA	STAT1

3D örneklerinde log2FoldChange değerine göre en az up regüle olan tüm genler soldan sağa sıralanarak Tablo 4.17’de gösterilmektedir.

Tablo 4.17 3D örneklerinde log2FoldChange değerine göre en az down regüle olan genden tüm genler

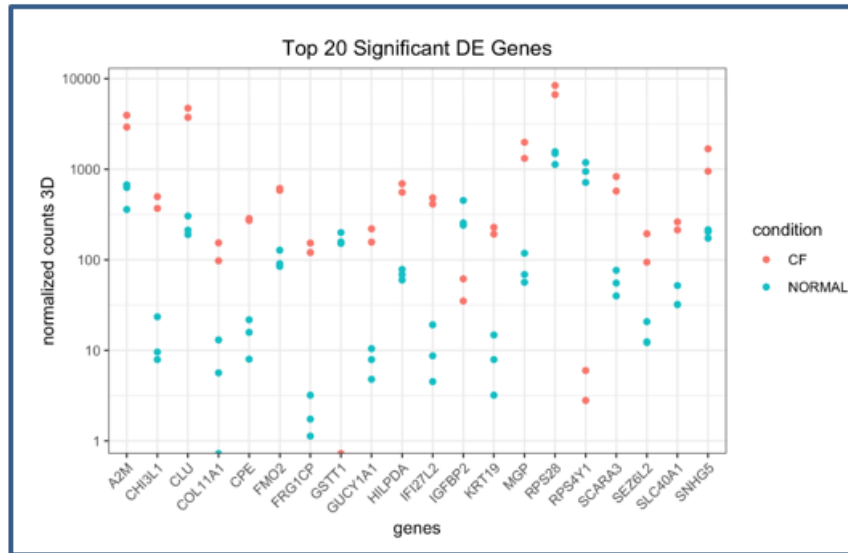
RUNX1T1	BMP4	THY1	SGCD	ILIR1	BGN	FGL2	FBXO17	C1R	HCFC1R1
SNORD13	CFH	SSPN	SELENOP	ADH1B	CEBPD	IL32	LDHA	LIMCH1	ARSD
IFI6	GSTM3	SNHG7	QPR1	PCDH7	SEMA3C	HAS2	TBX5	H6PD	ERO1A
CXCL12	CENPF	COL3A1	EDNRA	LDLR	EDIL3	ENTPD1	PPP1R3B	GULP1	BAALC
EMILIN1	SETBP1	PLA2R1	PKIA	PLTP	THBS2	SOCS3	SLC9A3R2	RARRES1	SLC43A3
GPNMB	MOXD1	NFIL3	NET1	GALNT6	SCN1B	TM4SF1	SOBP	AEBP1	MOCS1
PDK1	GUCY1B1	PTGIS	CHST2	SPON1	SLIT2	SSC5D	CNIH3	TOP2A	PRXL2A
TENM4	MAF	BTN3A2	FGF10	TMEM91	PTPRD	LINC00839	HTRA1	CCN4	SCG2
MYBL1	CD4	PTGFR	SYNE2	GNG2	BOC	DAAM2	RNASET2	ROR1	SEMA4B
SMPDL3A	LUM	SCRIP2	MASP1	ARL4C	S100A10	DTNA	AK4	PCOLCE	COL14A1
HMCN1	SERPINF1	STON2	PTGDS	MFAP4	SYNPO2	HSPB7	ACSS3	GTF2IP1	NCAM2
CPM	ALDOC	LOC100233156	G0S2	IGF1	CFI	MME	GDF10	ENO2	PTPRU
MTSS1	LRRC17	ADAMTS15	ADAMDE C1	C11orf96	SOD3	SMIM1	GSTM5	XG	THRB
PLCE1	KCNE4	ABIEBP	RPS28	TWIST1	HTATSFP1P2	ANK2	FMO2	SCARNA22	CHL1
PLSCR4	NDRG2	LYN	HF1A-AS2	COL10A1	CADPS2	LCE2A	GJA5	GPC4	FILIP1
APOE	DES	MATN2	SLC6A6	EFS	MRPS6	SNHG5	LAMC3	FER1L4	LGALS9
SLC40A1	RNF150	MBP	A2M	SPRY1	COL6A5	NEDD4L	SPAG4	KIAA0040	TREM1
MCOLN2	PDE4D	MXRA5	MLC1	CRABP2	CDKN2C	PRELP	GPRC5D-AS1	PALMD	PRDM8
ZNF395	AQP1	WNT2	GLDN	ARHGAP28	SLC5A3	SLIT3	YPEL4	ARHGAP28	MEDAG
LINC00484	EPB41	SNX10	CCNB1	PLN	CCN5	COLEC12	PLXNA4	HOXC8	RSPO2
MAOA	MMP1	FRG1BP	WFDC3	LPIN3	ZDBF2	ADAM23	SEZ6L2	CA12	PCSK2
APCDD1L	TNFSF13B	LEF	PCOLCE2	ITPR1	HOXC9	SCARA3	LOC100287846	SLC7A2	PDE4C
ESM1	PDE1A	AKR1C2	RBP1	SMOC2	HILPDA	OLFML1	C3	IGSF10	MGP
PTEGER1	PAK3	COL11A1	MAOB	GUCY1A1	CCDC152	MIR4458HG	CGNL1	KRT19	RPRML
LPAR4	CCDC144A	CCDC144B	CLU	GPC3	MT3	CPE	SEMA4D	FGF14	TBX4
RCAN2	CSNK2B	HIF3A	MTUS1	RSPO3	PEX5L	INHBB	CHRD1	IFI27L2	FGF11
COL21A1	B2M	CHI3L1	ZNF385D	SLITRK6	SGCA	HSPB3	CES1	EGFL6	PRUNE2
SYT1	XIST	HOXC6	FRG1CP	EGLN3	MCHR1	APLN	TRIM55	H19	AOC3
RORB	GAP43								

Daha fazla global bir görünüm için 3D örneklerde 352 gene ait Volcano Plot şekil 4.33'de gösterilmektedir.



Şekil 4 33 3D örneklerinde volcano plot gösterimi

3D örneklerde Diferansiyel olarak ifade edilen ilk 20 gen için (padj değerleriyle) normalleştirilmiş count verilerini grafiği şekil 4.34'de gösterilmektedir.



Şekil 4.34 3D örneklerinde Diferansiyel olarak ifade edilen ilk 20 gen

2D ve 3D örneklerde Padj değerine göre ilk 20 up ve down regüle genler karşılaştırıldığında 4 ortak gen olduğu görüldüğü. Bu genler Tablo 4.18'da koyu olarak gösterilmektedir.

Tablo 4.18 2D ve 3D örneklerinde Padj değerine göre sıralanmış ilk 20 up ve down regüle DEG karşılaştırılması

<u>2D</u>				<u>3D</u>		
	Genler	log2FoldChange	padj	Genler	log2FoldChange	padj
1	SNHG5	3.297333947	1.73E-48	CLU	4.338926544	3.06E-51
2	CLU	2.8280244	1.34E-42	RPS4Y1	-8.207654394	3.66E-49
3	C1R	2.421247183	8.56E-40	CHI3L1	5.278654929	6.11E-30
4	TP53I11	4.459833574	2.76E-34	IFI27L2	5.161546835	5.33E-29
5	BMP4	5.564339142	1.85E-30	SCARA3	3.600932102	2.31E-28
6	SCD	3.385583511	1.85E-30	MGP	4.031261754	1.83E-26
7	CYP1B1	3.501693331	8.90E-28	CPE	4.382761194	6.96E-24
8	CADM1	3.540961549	1.05E-26	RPS28	2.4471787	5.79E-21
9	ADH1B	3.43620725	2.53E-25	SNHG5	2.693549977	1.83E-18
10	PSG5	-3.957677795	5.59E-24	HILPDA	3.906536753	4.44E-18
11	XIST	10.57526921	3.86E-23	A2M	2.755215662	2.71E-17
12	ZNF395	3.725302013	5.30E-23	FMO2	2.484388922	4.38E-16
13	MASP1	5.275562296	1.35E-22	KRT19	4.179999174	5.59E-16
14	THBS2	3.502691098	2.76E-22	FRG1CP	5.866016146	3.44E-13
15	RPS4Y1	-10.16417851	1.70E-21	GSTT1	-8.98226273	7.62E-13
16	BGN	3.037563113	5.14E-21	GUCY1A1	4.080240859	4.36E-12
17	ADAMTS15	3.981008637	1.34E-20	SLC40A1	2.721871149	6.33E-12
18	MT1L	-1.871112323	1.36E-19	COL11A1	4.071256032	9.81E-12
19	HMGCS1	2.400024946	1.40E-19	IGFBP2	-2.74589356	3.59E-11
20	RPS28	1.747484242	1.40E-19	SEZ6L2	3.37645568	4.15E-11

2D ve 3D örneklerde log2FoldChange değerine göre ilk 20 down regüle genler karşılaştırıldığında 13 tane ortak gen olduğu görüldüğü. Bu genler Tablo 4.19’de koyu olarak gösterilmektedir.

Tablo 4.19 : 2D ve 3D örneklerinde log2FoldChange değerine göre sıralanmış ilk 20 down regüle DEG karşılaştırılması

<u>2D</u>			<u>3D</u>	
	Genler	log2FoldChange	Genler	log2FoldChange
1	RPS4Y1	-10.16417851	GSTT1	-8.98226273
2	EIF1AY	-7.822464086	USP9Y	-8.400490721
3	GSTT1	-7.329978999	RPS4Y1	-8.207654394
4	USP9Y	-6.602191204	DDX3Y	-8.130244399
5	DDX3Y	-6.241684283	TTY15	-7.313529147
6	TTY15	-6.122971563	TXLNGY	-7.169760405
7	TXLNGY	-5.658854526	EIF1AY	-7.070889375
8	PRKY	-5.271404223	MBOAT7	-6.847487048
9	FAM228A	-5.258408651	ZFY	-6.608625478
10	ZFY	-5.107441382	UTY	-5.772328319
11	MBOAT7	-4.819661127	NLGN4Y	-5.707548753

12	SMPD4	-4.812280241	PRKY	-5.577130036
13	NLGN4Y	-4.583945528	KDM5D	-5.079026048
14	PAX6	-4.277382754	TTY14	-5.024818428
15	PSG5	-3.957677795	KCNJ2	-3.923222024
16	ALDH1A3	-3.841972958	SNORD83B	-3.83768236
17	FAM174B	-3.75404292	PAX6	-3.827226907
18	APBB1IP	-3.483930548	PSG5	-3.472642203
19	KDM5D	-3.390839967	ARHGAP23	-3.267470028
20	CLDN4	-3.37695892	IGFBP2	-2.74589356

2D ve 3D örnekler için toplu sonuçlar Venn-şeması ile şekil 4.35 ve şekil 4.36'de gösterilmektedir.



Şekil 4.35 2D ve 3D örneklerinde down regüle genlere ait Venn-Şeması

2D ve 3D örneklerde up regüle genlere ait Venn-Şeması şekil 4.36'de gösterilmektedir.



Şekil 4.36 2D ve 3D örneklerinde up regüle genlere ait Venn-Şeması

4.2. Veri Kaynağı 2

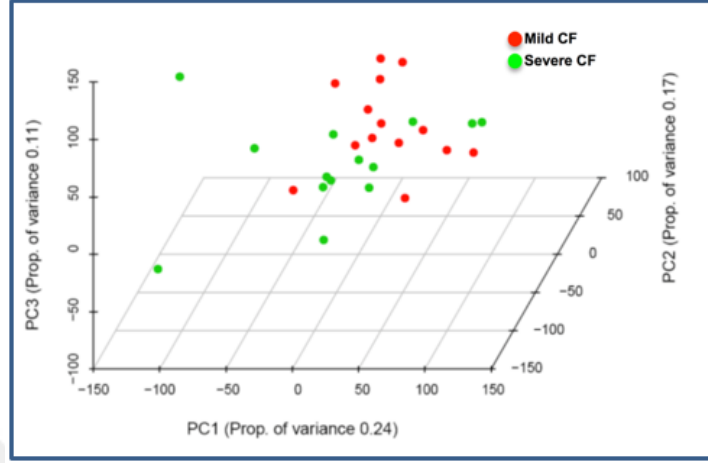
Transcriptomic Profile of Cystic Fibrosis Patients identifies Type I Interferon response and Ribosomal stalk proteins as potential modifiers of disease severity. Bu çalışmada 30 örnek kullanılmıştır.

Tablo 4.20 Veri Kaynağı 2 örnek listesi ve özellikleri

ÖRNEKLER	Baz	Boyut	Hastalık Evresi	Yaş	Cinsiyet
MCF_S01	278.18M	149.06Mb	Orta CF	19	Kadın
MCF_S02	204.59M	103.33Mb	Orta CF	20	Erkek
MCF_S03	184.32M	93.40Mb	Orta CF	19	Erkek
MCF_S04	215.75M	108.91Mb	Orta CF	21	Kadın
MCF_S05	250.79M	125.83Mb	Orta CF	9	Kadın
MCF_S06	272.37M	137.03Mb	Orta CF	25	Erkek
MCF_S07	264.05M	140.21Mb	Orta CF	11	Kadın
MCF_S08	297.88M	160.65Mb	Orta CF	28	Kadın
MCF_S09	310.23M	165.31Mb	Orta CF	35	Erkek
MCF_S10	310.82M	168.42Mb	Orta CF	25	Erkek
MCF_S11	270.92M	145.52Mb	Orta CF	18	Erkek
MCF_S12	361.62M	193.36Mb	Orta CF	40	Erkek
MCF_S13	259.79M	138.85Mb	Orta CF	32	Erkek
MCF_S14	245.31M	131.34Mb	Orta CF	42	Kadın
MCF_S16	284.25M	150.57Mb	Orta CF	15	Erkek
MCF_S17	338.75M	182.65Mb	Orta CF	23	Erkek
SCF_S01	258.44M	130.73Mb	Ağır CF	16	Kadın
SCF_S02	285.85M	144.73Mb	Ağır CF	17	Kadın
SCF_S03	193.87M	97.40Mb	Ağır CF	14	Kadın
SCF_S04	261.51M	131.49Mb	Ağır CF	20	Erkek
SCF_S05	295.95M	160.23Mb	Ağır CF	NA	NA
SCF_S06	240.98M	128.73Mb	Ağır CF	NA	NA
SCF_S07	318.61M	160.51Mb	Ağır CF	NA	NA
SCF_S08	285.28M	154.75Mb	Ağır CF	NA	NA
SCF_S09	274.40M	139.87Mb	Ağır CF	16	Erkek
SCF_S10	242.62M	123.70Mb	Ağır CF	14	Kadın
SCF_S11	304.69M	162.81Mb	Ağır CF	NA	NA
SCF_S12	305.01M	163.71Mb	Ağır CF	NA	NA
SCF_S13	264.37M	135.63Mb	Ağır CF	NA	NA
SCF_S14	255.49M	129.68Mb	Ağır CF	NA	NA

Çalışmada daha önce de belirtildiği gibi RNASeq kullanılarak Orta ve Ağır KF hastalarından alınan periferik kan lökositlerinin transkriptomik profili araştırılmıştır.

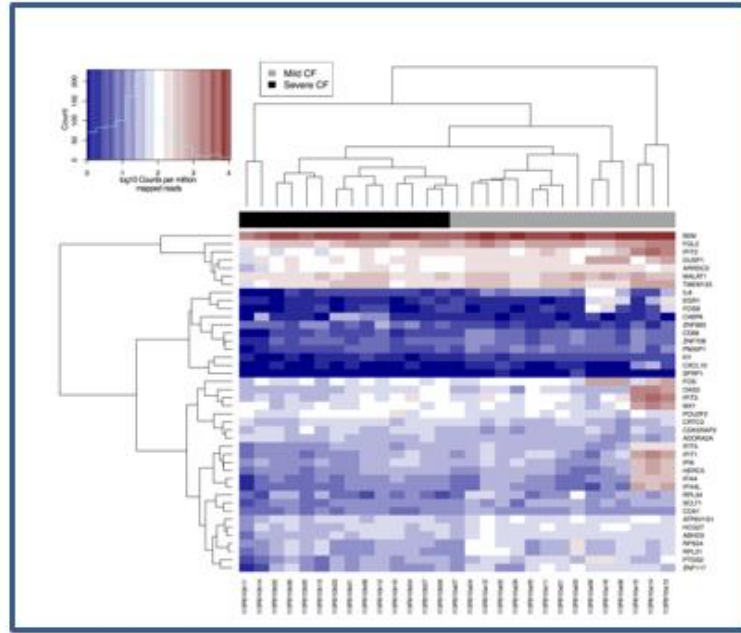
PCA analizi, incelenen iki grubun ifade profillerindeki farklılıkların nispeten küçük olduğunu ortaya çıkarmıştır (Şekil 4.37).



Şekil 4.37 Veri kaynağı 2 PCA analizi

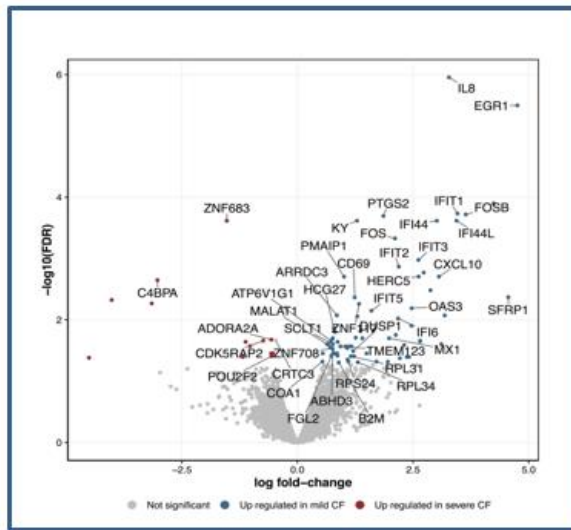
Şekil 4.37’ de Orta ve Ağır KF hastalarının PCA analiz grafiği gösterilmiştir. Filtrelenmiş CPM’nin PCA’sı, araştırılan iki grubun ifade profillerindeki farklılıkları değerlendirmek için verileri saymaktadır. Her nokta bir örneği temsil etmektedir. İki nokta ne kadar yakınsa, ifade profilleri o kadar benzerdir (Kormann et al., 2017)

Orta KF ve Ağır KF arasında farklı olarak ifade edilen genlere bakıldığında DE analizi ile Orta ve Ağır KF hasta grubu arasında 88 genin farklı şekilde eksprese edildiği gösterilmiştir. Bunlardan 74 (%84) gen Orta KF hastalarında 14 (%16) gen de Ağır KF hastalarında daha yüksek ekspresyon seviyesi göstermiştir. Orta ve Ağır KF hastaları arasındaki DEG’ler karşılaştırıldığında seçilen DEG’lerin heatmap ile araştırılan gruplar için iki farklı küme ortaya çıkmıştır (Şekil 4.38). Her sütun ayrı bir hastayı, her yatay çizgi de ayrı bir geni temsil etmektedir. Kümelenmiş örneklerin ve genlerin dendrogramında Orta KF ve Ağır KF hastalar ekspresyon benzerliklerine göre kümelenir. Ekspresyon profilleri milyon okuma başına sayımlarla ölçülmektedir (CPM-Model) (Kormann et al., 2017).



Şekil 4.38 Veri kaynağı 2 Diferansiyel olarak ifade edilen genlerde heatmap

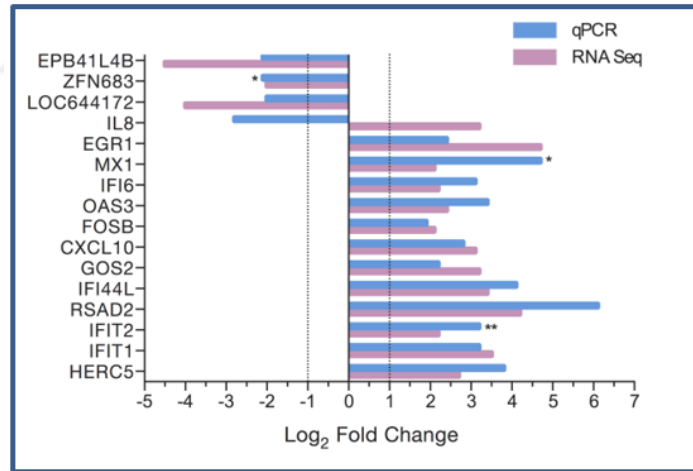
Gruplar arasındaki global gen seviyesi ekspresyon analizi, ilgili gen adı, ekspresyon seviyesi ve RNASeq kullanılarak tespitin doğruluğu ile birlikte Volcano-plot ile gösterilmektedir (Şekil 4.39) (Kormann et al., 2017).



Şekil 4.39 Veri kaynağı 2
RNASeq volcano plot

EGR1, SFRP1, RSAD2 ve FOSB, Ağır KF grubuna kıyasla Orta KF grubunda önemli ölçüde aşırı ifade edilen (>3.5) ana DEG'ler olarak gösterilmiştir. EGR1'in artan ekspresyonunun hafif akciğer hastalığı ile önemli ölçüde ilişkili olduğu bulunmuştur (kat değişimi=4.5; FDR $P=3.1 \times 10^{-6}$). Buna paralel olarak, diğer EGR ailesi genleri, EGR2 (fold change=2.4; FDR $P=0.04$) ve EGR3 (fold change=2.6; FDR $P=0.02$) Orta KF grubunda daha yüksek düzeyde ekspresyon göstermiştir. Benzer şekilde, EPB41L4B, LOC644172, C4BPA ve ZNF683 Ağır KF grubunda önemli ölçüde aşırı eksprese edilen (>1.5) ana DEG'ler olarak rapor edilmiştir. Bu genler arasında ZNF683 KF akciğer hastalığının şiddetli formu ile daha iyi ilişki sergilemiştir (fold change = 1.5; FDR $P = 0.0002$).

Aynı örnekleri kullanarak qPCR ile 16 DEG için RNASeq sonuçları doğrulanarak, RNASeq ve qPCR arasında yüksek uyumlu sonuçlar gözlemlenmiştir (%94; 16 DEG arasından 15'i doğrulanmıştır).

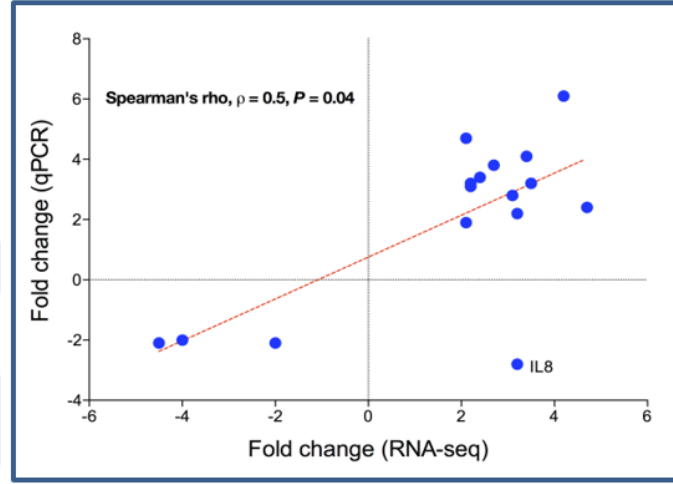


Şekil 4.40 Veri kaynağı 2 RNASeq sonuçlarının qPCR validasyonu

Şekil 4.40'de RNASeq verilerinin ekspresyon profili ve seçilen genler için qPCR verileri aynı numuneler kullanılarak karşılaştırılmıştır. Kontrol grubu olarak Ağır KF örnekleri kullanılmış ve ifade düzeyi 1 veya -1 olarak ayarlanmıştır. Ekspresyon, referans gen olarak 18s kullanılarak normalleştirilmiştir. Sonuçlar ortalama değerleri temsil etmekte ve kat değişiminin Log2 değerleri olarak eksprese edilmektedirler. Orta KF ve Ağır KF hastaları arasında qPCR sonuçlarında gözlenen

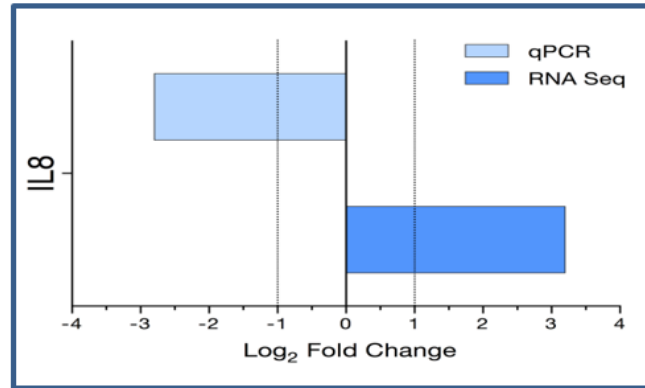
önemli fark (**P değeri<0.05; ***P değeri<0.001). Önem düzeyi, RNASeq verileri için <0.0001 P-değerine ayarlanmış (Kormann et al., 2017).

RNA sekansı ile qPCR arasındaki DEG ekspresyon seviyelerinin karşılaştırılması ile daha iyi bir korelasyon ortaya çıkarılmıştır (Spearman's rho, $\rho=0.53$, $P=0.04$) (Şekil 4.41) (Kormann et al., 2017).



Şekil 4.41 Veri kaynağı 2 RNASeq ve qPCR'ın korelasyon analizi

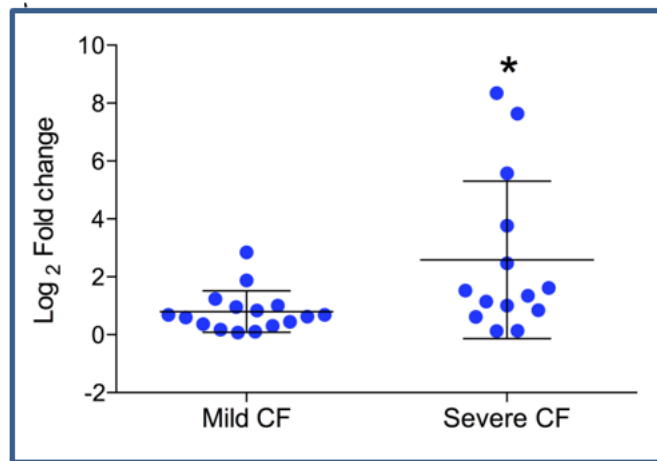
RNASeq ve qRT-PCR verileri karşılaştırılabilir olduğundan 15 genin benzer mRNA seviyelerine sahip olduğu gösterilmiş. Buna ek olarak, IFIT2 (fold change=3.2, $P<0.001$), MX1 (fold change=3.2, $P<0.05$) ve ZFN683 (fold change=-2.1, $P<0.05$) genlere sahip Orta KF grubu için qPCR sonuçlarında ekspresyon seviyelerinde önemli fark gözlenmiş. Ağır KF grubunun birçok örneği amplifikasyonda başarısız olduğundan EGR1 için anlamlı bir fark elde edilmemiş. qPCR tarafından doğrulanamayan tek DEG ise IL-8. IL-8 KF için çok önemli bir proinflatuar sitokin olduğundan, her iki platform için IL-8 verilerini daha da incelenmiş. RNAseq verilerinde, IL-8'in Orta KF grubunda aşırı eksprese edildiği gösterilmiş (Şekil 4.40 ve Şekil 4.42) (Kormann et al., 2017).



Şekil 4.42 Veri kaynağı 2 qPCR ve RNASeq ile IL-8 ekspresyonu

Şekil 4.42’de gösterilen grafikte platformlar arasında farklı sonuçları anlamak için IL-8 için Şekil 4.40'deki aynı veriler kullanılmış (Kormann et al., 2017).

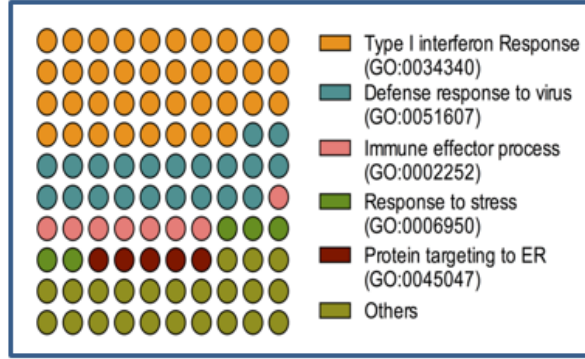
Diğer taraftan, IL-8 qPCR analizine göre Ağır KF grubunda önemli ölçüde aşırı eksprese olmuştur (P=0.01; Şekil 4.43).



Şekil 4.43 Veri kaynağı 2 IL-8 ekspresyonu

Şekil 4.43’de ki Nokta grafiği ile Ağır ve Orta KF hastalarında IL-8 ekspresyonu gösterilmiştir. Ekspresyon, referans gen olarak 18s kullanılarak normalleştirilmiştir. Sonuçlar ortalama değerleri temsil etmekte ve kat değişiminin Log2 değerleri olarak ifade edilmektedir (Kormann et al., 2017).

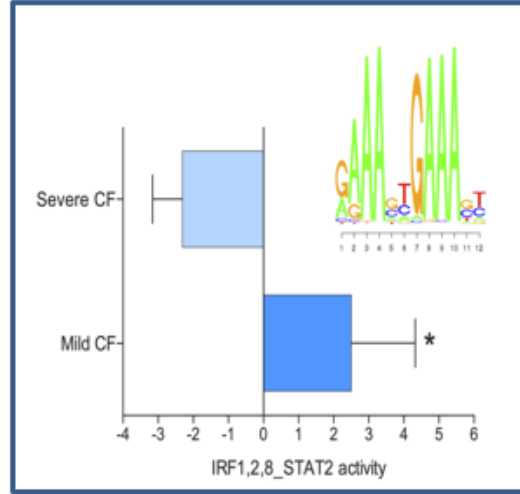
Orta ve Ağır KF grubu arasında farklı şekilde değiştirilebilecek biyolojik süreçleri anlamak için, Orta ve Ağır KF'de up regüle olan DEG'ler için bir GO zenginleştirme analizi gerçekleştirilmiştir. Orta KF grubu durumunda, tip I interferon yanıtına ait genler seçici olarak zenginleştirilmiş (%38, $P=1.45E-09$) (Şekil 4.44).



Şekil 4.44 Veri Kaynağı 2 RNASeq fonksiyonel zenginleştirme analizi

Temsili Gen Ontolojisi (GO), biyolojik süreçler için farklı şekilde ifade edilen genler arasındaki zenginleştirmeyi ifade eder. Tip I interferon tepkisine dâhil olan genler ve ER taşınmasından ve protein sentezinden sorumlu ribozomal proteinler aşırı temsil edilmiştir (Kormann et al., 2017).

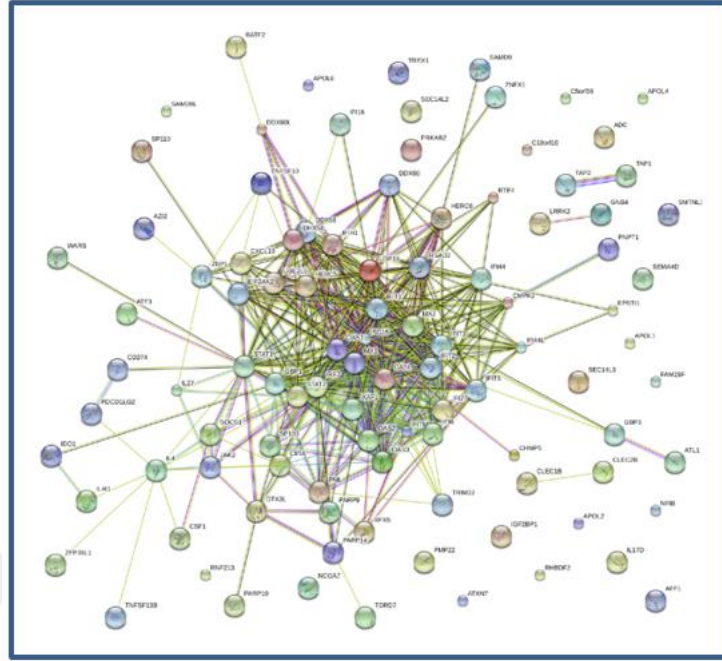
Bu yolak'a dâhil olan toplam on gen (EGR1, MX1, IFIT1, IFIT2, IFIT3, OAS3, OASL, IFI6, ISG15 ve RSAD2), Orta KF grubunda yüksek ekspresyona uğramış. Ek olarak, anti-viral yanıt yolları ve bağışıklık efektör süreci, önemli ölçüde zenginleşme göstermiş. Zenginleştirme analizi üç ribozomal sap protein genlerinden RPL31, RPL34 ve RPS242'nin ER hedefleyen protein yolağında zenginleştiğini belirlemiştir. Ağır KF grubunun on dört up regüle olan geni için hiçbir spesifik yolak zenginleştirilmemiştir. Belirli TF'lerin ve DEG'lerin düzenleyici ağlarının katkısını araştırmak için MARA analizini RNASeq verilerine uygulanmıştır. Veriler, regüle genlerin yüksek transkript seviyeleri ile Orta KF grubunda IRF1, IRF2, IRF8 ve STAT2 transkripsiyon faktörü aktivitesinin önemli ölçüde arttığını ortaya koymuştur. Ortalama aktivasyon z-skoru → Orta KF için 2.5: Ağır KF için -2.3: $P < 0.05$ (Şekil 4.45) (Kormann et al., 2017).



Şekil 4.45 Veri kaynağı 2 RNASeq ISMARA analizi

ISMARA analizi, Orta ve Ağır KF hastaları arasında IRF1,2,8 ve STAT2 aktivitesinde önemli bir fark öngörmüştür (*P değeri < 0.05). ISMARA analizi ağır, IRF1,2,8 ve STAT2 hedef genlerinin olası etkileşimini tahmin etmektedir.

Transkriptik Faktör (TF) için en iyi hedef genlerin (Target score >10), tip I interferon yanıt yolağına ait olduğu gösterilmiş. En yüksek puan (Hedef puan >25) RSAD2, IFI44L, IFIT1, ISG15, IFIT3, OAS3, HERC5 ve IFI44 genleri için gözlemlenmiş. Böylelikle bu çalışma ile hem zenginleştirme hem de yolak analizinin Orta KF fenotipinde tip I interferon yanıtının olası katkısını belirlemede tutarlılığı gösterilmiş. IRF1,2,8 ve STAT2 transkripsiyon faktörlerinin ağ analizi (Şekil 4.46) ile tip I interferon yanıtında yukarıda belirtilen genlerin ve diğer ilgili genlerin daha yakın etkileşimlerini gösterilmiş.



Şekil 4.46 Veri kaynağı 2 RNASeq ağ analizi

Veri kaynağı 2’de yapılan qPCR validasyon çalışması, GO enrichment, Network ve MARA analizleri bu tez çalışmamızda uygulanmamıştır.

Bu tez kapsamında kullanılan pipeline ile ettiğimiz sonuçlar Varyant çağırma ve Transkriptome olarak ayrılmıştır.

A-Varyant Çağırma: Toplu olarak sonuçlar Tablo 4.21’de paylaşılmıştır.

Tablo 4.21 Orta ve Ağır KF örneklerinde varyant çağırma toplu sonuçları

		Association		Bening		*Benign_DR	**Benign/LB	
SCF_S01	GATK	1	4	41	571	rs780668	1	33
	Samtools	0		74			3	
SCF_S02	GATK	0	4	24	660		4	26
	Samtools	2		63			2	
MCF_S01	GATK	1	4	39	630		2	22
	Samtools	3		80			6	
MCF_S02	GATK	0	2	33	505		0	20
	Samtools	1		58			2	
MCF_S03	GATK	0	4	33	469		2	14
	Samtools	1		63			3	
MCF_S04	GATK	0	5	23	475		3	20
	Samtools	2		62			2	
MCF_S05	GATK	0	3	35	619		0	20
	Samtools	2		82			3	

MCF_S06	GATK	0	6	24	641		3	18
	Samtools	1		85			3	
SCF_S03	GATK	0	2	38	538		1	17
	Samtools	1		85			1	
SCF_S04	GATK	0	6	43	582	rs780668	1	21
	Samtools	1		82			0	
SCF_S05	GATK	0	2	28	568		3	28
	Samtools	0		62			0	
SCF_S06	GATK	0	4	41	649		0	24
	Samtools	0		80			2	
SCF_S07	GATK	0	6	37	789		0	24
	Samtools	0		104			6	
SCF_S08	GATK	1	2	37	614		2	16
	Samtools	0		61			2	
SCF_S09	GATK	0	3	36	598		2	25
	Samtools	0		95			1	
SCF_S10	GATK	1	3	38	570	1	0	18
	Samtools	1		65			4	
SCF_S11	GATK	1	7	50	764		2	31
	Samtools	1		88			2	
SCF_S12	GATK	0	7	45	741		0	31
	Samtools	1		100			3	
SCF_S13	GATK	0	4	30	550		1	18
	Samtools	2		70			3	
SCF_S14	GATK	0	3	30	552		3	17
	Samtools	0		41			41	
MCF_S07	GATK			30	436		2	16
	Samtools			36			1	
MCF_S08	GATK	0	1	25	326		0	11
	Samtools	0		33			5	
MCF_S09	GATK	0	3	36	706		2	22
	Samtools	0		72			0	
MCF_S10	GATK	0	3	42	604		2	24
	Samtools	1		70			3	
MCF_S11	GATK	0	4	34	661		3	24
	Samtools	2		75			3	
MCF_S12	GATK	1	3	41	554		1	24
	Samtools	1		60			2	
MCF_S13	GATK	0	4	33	642		3	25
	Samtools	1		69			3	
MCF_S14	GATK	0	3	29	624		1	28
	Samtools	1		79			3	
MCF_S15	GATK	0	6	79	737		3	16
	Samtools	1		36			2	
MCF_S16	GATK	1	3	45	712		3	19
	Samtools	0		84			4	

*Benign_drug_response, **Benign/Likely_benig

		CIOP		*CIOPO	Drug_response	
SCF_S01	GATK	1	4		rs1695	rs180127 4,
	Samtools	rs11558258			0	rs396991, rs4880
SCF_S02	GATK	rs35947132	rs13871, rs111529228		0	6
	Samtools	rs115049252			rs4680	
MCF_S01	GATK		rs6030, rs80084721		0	rs7294, rs4961,
	Samtools	rs41291058, rs11466045, rs17227403			rs4680	rs4880
MCF_S02	GATK	rs41265019, rs55970278, rs5985		rs1801131	6	
	Samtools					
MCF_S03	GATK	rs150420496, rs145272777, rs5985			0	rs180127 4,
	Samtools				rs4680	rs396991, rs4880
MCF_S04	GATK	2	4		rs7294	3
	Samtools	0			rs4680	
MCF_S05	GATK	5			4	
	Samtools					
MCF_S06	GATK	rs139366969, rs11453124	rs137870461 rs41273880		0	4
	Samtools	rs138139146, rs5985			1	
SCF_S03	GATK	rs779413093			5	
	Samtools					
SCF_S04	GATK	rs147134923	rs111369573 rs5982			
	Samtools	rs146393315, rs5985				
SCF_S05	GATK	rs5985, rs139626312, rs34132016			0	5
	Samtools				rs25487	
SCF_S06	GATK	rs398028512, rs34737051	rs6030, rs4524, rs147744729	rs2071421	3	
	Samtools	0				
SCF_S07	GATK	1	5		0	6
	Samtools	3			1	
SCF_S08	GATK	0	5		0	3
	Samtools	1			1	
SCF_S09	GATK	4			7	
	Samtools					
SCF_S10	GATK	0	3		6	
	Samtools	1				
SCF_S11	GATK	rs398028512, rs148565559, rs11453124	rs147370143 rs5985		1	9
	Samtools	rs11558258, rs5985			2	
SCF_S12	GATK	0	rs41549716, rs56148525, rs41288979		8	
	Samtools	rs80338869				
SCF_S13	GATK	rs775762131	rs140887831		4	
	Samtools	0				
SCF_S14	GATK	0	rs36027220, rs577140751	rs6151429	rs1801274, rs1695	
	Samtools	rs35366573				
MCF_S07	GATK	rs45517423	rs33998267		0	3
	Samtools	0			2	
MCF_S08	GATK	rs398028512	rs201199629		rs396991	rs180127

	Samtools	0	rs139334167		0	4
MCF_S09	GATK	2	8		0	6
	Samtools	0			1	
MCF_S10	GATK	rs4524	rs6030, rs35026927, rs5984, rs72557941	rs6151429	6	
	Samtools	0				
MCF_S11	GATK	rs41291058	5		5	
	Samtools	rs36033115, rs115602636				
MCF_S12	GATK	1	10		5	
	Samtools	4				
MCF_S13	GATK	1	4		0	6
	Samtools	0			1	
MCF_S14	GATK	0	3		1	3
	Samtools	1			1	
MCF_S15	GATK					
	Samtools					
MCF_S16	GATK	8		8	2	4
	Samtools				2	

Conflicting_interpretations_of_pathogenicity, * Conflicting_interpretations_of_pathogenicity, other

		Muhtemel Benign		Muhtemel Patojenik	Patojenik	
SCF_S01	GATK	1	22		rs137854889, rs780235686, .: Chronic_granulomatous_disease_X-linked not_provided	
	Samtools	2			rs368698783	
SCF_S02	GATK	1	22		rs137854889, .: Chronic_granulomatous_disease_X-linked not_provided	rs17261572
	Samtools	5			rs1566734	
MCF_S01	GATK	2	25		0	
	Samtools	1			rs368698783	
MCF_S02	GATK	1	19		Chronic_granulomatous_disease_X-linked not_provided, X-linked_agammaglobulinemia X-linked_agammaglobulinemia_with_growth_hormone_deficiency not_provided	
	Samtools	1		X-linked_- severe combined immunodeficiency not_provided	rs368698783	
MCF_S03	GATK	4	16			
	Samtools	3				
MCF_S04	GATK	1	18	0	rs17261572	
	Samtools	3		rs763440821		
MCF_S05	GATK	5	21		rs786203560, rs746973259, Chronic_granulomatous_disease_X-linked not_provided, rs864321665	
	Samtools	2			0	
MCF_S06	GATK	5	10		Chronic_granulomatous_disease_X-linked not_provided	
	Samtools	0			0	

SCF_S03	GATK	2	14		Chronic_granulomatous_disease,_X-linked not_provided	rs1566734
	Samtools	4			0	
SCF_S04	GATK	4	16	rs397517245		
	Samtools	2		0		
SCF_S05	GATK	2	14		Chronic_granulomatous_disease,_X-linked not_provided	0
	Samtools	2				
SCF_S06	GATK	4	3		Spinocerebellar_ataxia_43	0
	Samtools	3				
SCF_S07	GATK	3	38		rs281864999, rs312262845	rs1566734
	Samtools	5			0	
SCF_S08	GATK	2	18		.: CTCF, rs312262845, .:CYBB: Chronic_granulomatous_disease,_X-linked not_provided	0
	Samtools	2				
SCF_S09	GATK	3	26		rs532964185, .: Chronic_granulomatous_disease,_X-linked not_provided,	rs17261572
	Samtools	2				
SCF_S10	GATK	3	21		.:not_provided, .: Chronic_granulomatous_disease,_X-linked not_provided, rs864321665	0
	Samtools	3				
SCF_S11	GATK	4	30		rs746973259, .:Chronic_granulomatous_disease,_X-linked not_provided	0
	Samtools	5				
SCF_S12	GATK	4	27	Mucopolysaccharidosis_type_1	rs773570504, .: Chronic_granulomatous_disease,_X-linked not_provided	rs17261572
	Samtools	3		0	rs2236379	
SCF_S13	GATK	4	14	frameshift insertion	Chronic_granulomatous_disease,_X-linked not_provided	0
	Samtools	3		0		
SCF_S14	GATK	10	24	Immunodeficiency_40	rs779726987, rs767399685, rs748536322	0
	Samtools	1				
MCF_S07	GATK	1	12			
	Samtools	0				
MCF_S08	GATK	5	13			
	Samtools	0				
MCF_S09	GATK	1	19	Kaufman_oculocerebrofacial_syndrome	rs779726987, rs587783017, CAPN3:not_provided, Chronic_granulomatous_disease,_X-linked not_provided	rs1566734
	Samtools	4		0	0	

MCF_S10	GATK	4	21		rs397507215	
	Samtools	4				
MCF_S11	GATK	5	19	Neurodevelopmental_disorder	Chronic_granulomatous_disease,_X-linked not_provided	
	Samtools	5		0	0	
MCF_S12	GATK	0	12			
	Samtools	3				
MCF_S13	GATK	5	23		rs864321665	
	Samtools	6			rs368698783	
MCF_S14	GATK	2	13		Chronic_granulomatous_disease,_X-linked not_provided, X-linked_agammaglobulinemia X-linked_agammaglobulinemia_with_growth_hormone_deficiency not_provided	rs17261572
	Samtools	1			0	
MCF_S15	GATK				rs137854889, rs312262845, not_provided, Intellectual_disability,_autosomal_dominant_49	
	Samtools				0	
MCF_S16	GATK	5	27		rs17261572	
	Samtools	5				

****Pathogenic/Likely_pathogenic ****Uncertain_significance

		****Pathogenic/LP	Pathogenic_protective	Protective
SCF_S01	GATK			
	Samtools			
SCF_S02	GATK		rs1805015	
	Samtools			
MCF_S01	GATK		0	
	Samtools		rs1805015	
MCF_S02	GATK			
	Samtools			
MCF_S03	GATK			
	Samtools			
MCF_S04	GATK			
	Samtools			
MCF_S05	GATK		0	rs1805015
	Samtools		rs1805015	
MCF_S06	GATK			
	Samtools			
SCF_S03	GATK			
	Samtools			
SCF_S04	GATK		rs1805015	

	Samtools			
SCF_S05	GATK			
	Samtools			
SCF_S06	GATK		rs1805015	
	Samtools		0	
SCF_S07	GATK	rs200788251		
	Samtools			
SCF_S08	GATK			
	Samtools			
SCF_S09	GATK		rs1805015	rs1805015
	Samtools			
SCF_S10	GATK			
	Samtools			
SCF_S11	GATK			
	Samtools			
SCF_S12	GATK		rs1805015	rs8177374, rs1805015
	Samtools			
SCF_S13	GATK			
	Samtools			
SCF_S14	GATK			rs1063192
	Samtools			
MCF_S07	GATK	Zellweger_syndrome not_provided		
	Samtools	0		
MCF_S08	GATK	rs775699005		
	Samtools	0		
MCF_S09	GATK			rs1063192
	Samtools			
MCF_S10	GATK	rs75184679, rs113994168	rs1805015	
	Samtools			
MCF_S11	GATK			
	Samtools			
MCF_S12	GATK			1
	Samtools			
MCF_S13	GATK		rs1805015	
	Samtools			
MCF_S14	GATK			
	Samtools			
MCF_S15	GATK			
	Samtools			
MCF_S16	GATK			
	Samtools			

****Pathogenic/Likely_pathogenic

		Risk_factor		*****US	
SCF_S01	GATK	rs149973869	rs34911341,	5	6
	Samtools	0	rs4833095	2	
SCF_S02	GATK	4		9	8

	Samtools			1		
MCF_S01	GATK	0	5	2	8	
	Samtools	1		0		
MCF_S02	GATK	5		7	9	
	Samtools			0		
MCF_S03	GATK	rs3737787, rs1051931		4	7	
	Samtools			0		
MCF_S04	GATK	0	7	3	5	
	Samtools	1		2		
MCF_S05	GATK	3		4	7	
	Samtools			1		
MCF_S06	GATK	4		6	7	
	Samtools			1		
SCF_S03	GATK	0		6	10	
	Samtools	rs12150220		0		
SCF_S04	GATK	4		9	8	
	Samtools			2		
SCF_S05	GATK	2		8	9	
	Samtools			3		
SCF_S06	GATK	rs4833095		7	7	
	Samtools			1		
SCF_S07	GATK	7		10	11	
	Samtools			3		
SCF_S08	GATK	1	3	6	8	
	Samtools	0		1		
SCF_S09	GATK	5		7	7	
	Samtools			2		
SCF_S10	GATK	5		7	10	
	Samtools			3		
SCF_S11	GATK	4		3	10	
	Samtools			4		
SCF_S12	GATK	2	5	5	11	
	Samtools	0		3		
SCF_S13	GATK	3		6	8	
	Samtools			1		
SCF_S14	GATK	2		30	10	
	Samtools			3		
MCF_S07	GATK	rs3737787		8	9	
	Samtools			2		
MCF_S08	GATK	rs203462		5	3	
	Samtools			3		
MCF_S09	GATK	1	5	7	8	
	Samtools	0		3		
MCF_S10	GATK	1	3	7	10	
	Samtools	0		2		
MCF_S11	GATK	0	9	6	5	
	Samtools	1		1		
MCF_S12	GATK	6		11	11	
	Samtools			1		
MCF_S13	GATK	rs4833095	rs203462	3	6	
	Samtools	0		1		
MCF_S14	GATK	1	5	8	7	
	Samtools	0		1		
MCF_S15	GATK	4		6	6	
	Samtools			2		
MCF_S16	GATK	rs3737787		6	11	
	Samtools			0		

*****Uncertain_significance

Tablo 4.22’de Samtools paketi ile elde edilen varyantlar, GATK paketi ile elde edilen varyantlar ve her iki pakette ortak olan varyantların sayısı paylaşılmıştır.

Tablo 4.22 Orta ve Ağır KF örneklerinde varyant sayıları ve paket karşılaştırılması

ÖRNEKLER	SAMTOOLS	ORTAK VARYANT SAYISI	GATK
SCF_S01	863	7682	1376
SCF_S02	951	8467	1246
MCF_S01	1017	8616	1394
MCF_S02	879	7638	1494
MCF_S03	802	6729	1491
MCF_S04	729	6415	1066
MCF_S05	1000	8594	1262
MCF_S06	1015	9221	1454
SCF_S03	742	6572	960
SCF_S04	911	7923	1197
SCF_S05	916	8112	1772
SCF_S06	921	7761	1186
SCF_S07	1163	11016	1795
SCF_S08	901	8084	1698
SCF_S09	987	8273	1354
SCF_S10	852	7185	1116
SCF_S11	1143	10091	1648
SCF_S12	1167	10069	1462
SCF_S13	751	6778	1228
SCF_S14	840	8370	1949
MCF_S07	709	6878	1509
MCF_S08	674	5916	1596
MCF_S09	1124	9982	1755
MCF_S10	1085	9714	2070
MCF_S11	1007	9032	1821
MCF_S12	1012	8991	1687
MCF_S13	1012	9077	1683
MCF_S14	1084	8740	1818
MCF_S15	1153	10279	1572
MCF_S16	1284	11455	2186

Varyant çağırma sonucunda sadece patojenlerin listesi aşağıdaki tabloda gösterilmektedir.

Tablo 4.23 Orta ve Ağır KF örneklerinde sadece patojenlerin listesi

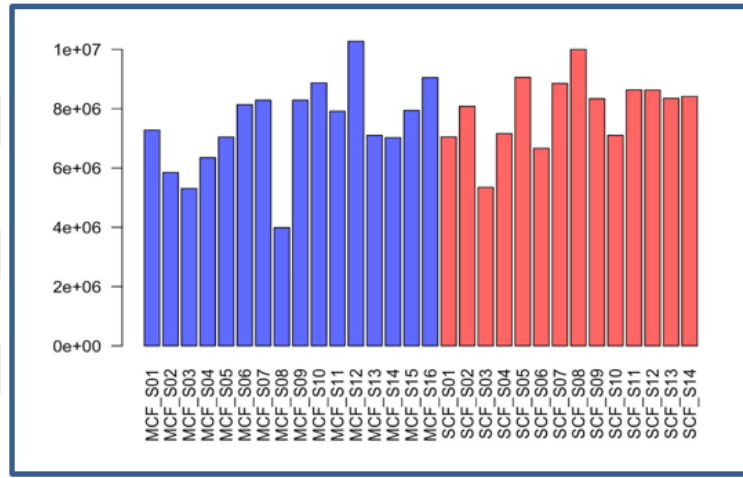
		Patojenik
SCF_S01	GATK	rs137854889, rs780235686, .: Chronic_granulomatous_disease,_X-linked not_provided
	Samtools	rs368698783
SCF_S02	GATK	rs137854889, .: Chronic_granulomatous_disease,_X-linked not_provided

	Samtools	rs1566734	
MCF_S01	GATK	0	
	Samtools	rs368698783	
MCF_S02	GATK	Chronic_granulomatous_disease,_X-linked not_provided, X-linked_agammaglobulinemia X-linked_agammaglobulinemia_with_growth_hormone_deficiency not_provided	
	Samtools	rs368698783	
MCF_S03	GATK		
	Samtools		
MCF_S04	GATK	rs17261572	
	Samtools		
MCF_S05	GATK	rs786203560, rs746973259, Chronic_granulomatous_disease,_X-linked not_provided, rs864321665	
	Samtools	0	
MCF_S06	GATK	Chronic_granulomatous_disease,_X-linked not_provided	
	Samtools	0	
SCF_S03	GATK	Chronic_granulomatous_disease,_X-linked not_provided	rs1566734
	Samtools	0	
SCF_S04	GATK		
	Samtools		
SCF_S05	GATK	Chronic_granulomatous_disease,_X-linked not_provided	
	Samtools	0	
SCF_S06	GATK	Spinocerebellar_ataxia_43	
	Samtools	0	
SCF_S07	GATK	rs281864999, rs312262845	rs1566734
	Samtools	0	
SCF_S08	GATK	:: CTCF, rs312262845, ::CYBB: Chronic_granulomatous_disease,_X-linked not_provided	
	Samtools	0	
SCF_S09	GATK	rs532964185, :: Chronic_granulomatous_disease,_X-linked not_provided,	rs17261572
	Samtools		
SCF_S10	GATK	::not_provided, :: Chronic_granulomatous_disease,_X-linked not_provided, rs864321665	
	Samtools	0	
SCF_S11	GATK	rs746973259, ::Chronic_granulomatous_disease,_X-linked not_provided	
	Samtools	0	
SCF_S12	GATK	rs773570504, :: Chronic_granulomatous_disease,_X-linked not_provided	rs17261572
	Samtools	rs2236379	
SCF_S13	GATK	Chronic_granulomatous_disease,_X-linked not_provided	
	Samtools	0	
SCF_S14	GATK	rs779726987, rs767399685, rs748536322	
	Samtools	0	
MCF_S07	GATK		
	Samtools		
MCF_S08	GATK		
	Samtools		
MCF_S09	GATK	rs779726987, rs587783017, CAPN3:not_provided, Chronic_granulomatous_disease,_X-linked not_provided	rs1566734
	Samtools	0	
MCF_S10	GATK	rs397507215	
	Samtools		
MCF_S11	GATK	Chronic_granulomatous_disease,_X-linked not_provided	
	Samtools	0	
MCF_S12	GATK		
	Samtools		
MCF_S13	GATK	rs864321665	
	Samtools	rs368698783	
MCF_S14	GATK	Chronic_granulomatous_disease,_X-linked not_provided, X-linked_agammaglobulinemia X-	rs17261572

		linked_agammaglobulinemia_with_growth_hormone_deficiency/not_provided	
	Samtools	0	
MCF_S15	GATK	rs137854889, rs312262845, not_provided, Intellectual_disability,_autosomal_dominant_49	
	Samtools	0	
MCF_S16	GATK	rs17261572	
	Samtools		

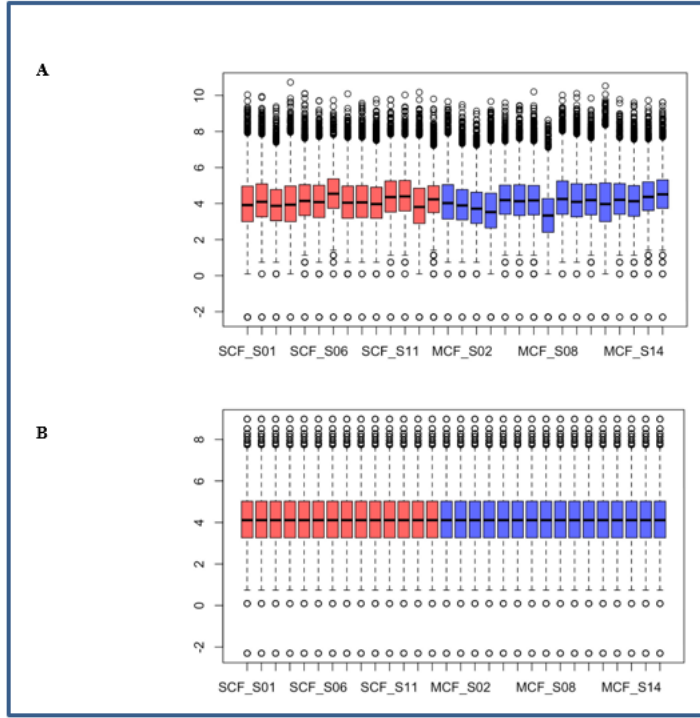
4.3. Transkriptom Analiz sonuçları

EDASeq analizi için BAM dosyaları kullanılarak aşağıdaki count grafikleri oluşturuldu. Örneklerle ait count grafiği Şekil 4.47’de gösterilmektedir.



Şekil 4.47 Orta ve Ağır KF örneklerinde eşlenen okuma sayısı

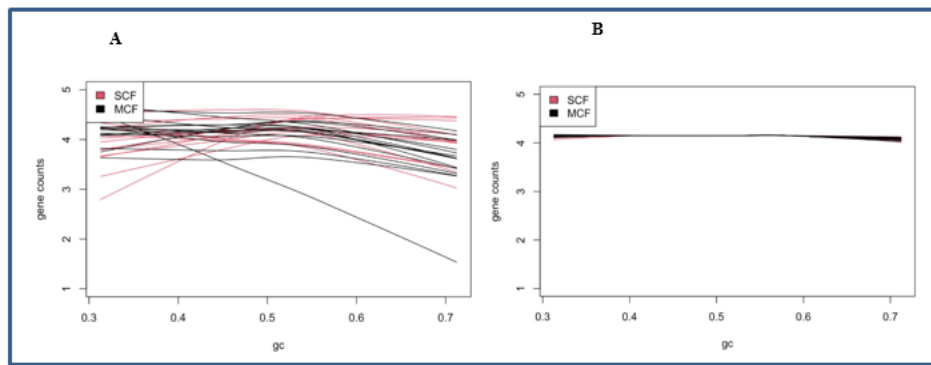
Örneklerle ait normalizasyon öncesi ve sonrası boxplot grafikler aşağıda paylaşılmıştır (Şekil 4.48).



Şekil 4.48 Orta ve Ağır KF örneklerinde gen seviyeleri

A) Normalizasyon öncesi gen seviyesi sayımlarının lane arası dağılımı B) Normalize gen seviyesi sayımlarının (log) şeritler arası dağılımı

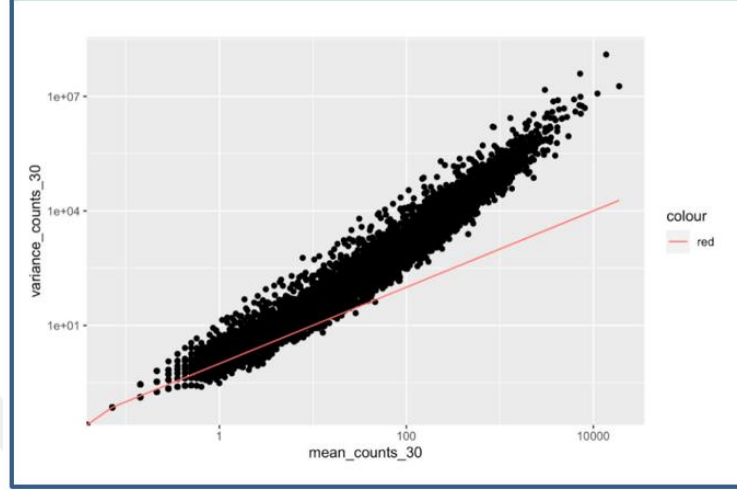
BiasPlot kullanılarak elde edilen gen seviyesi sayımlarının GC içeriğine bağımlılığı Şekil 4.49'de gösterilmektedir.



Şekil 4.49 Orta ve Ağır KF örneklerinde GC içeriğinde gen seviyesi

A) Normalizasyon öncesi, her şerit için GC içeriği üzerinde gen seviyesi sayımlarının (log) düşük regresyonu ve B) Her şerit için GC içeriğinde normalleştirilmiş gen seviyesi sayımlarının (log) düşük regresyonu

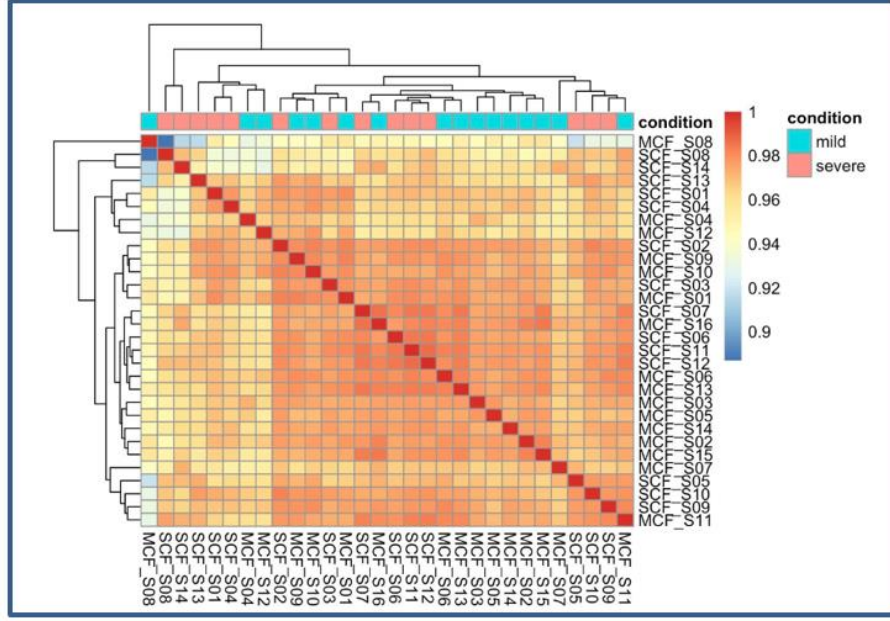
RNAseq ham veri count'un nasıl dağıldığı hakkında bir fikir edinmek için örneklere ait çizimler DESeq paketi kullanılarak yapılmıştır. Sadece Ağır KF örneklerine ait ortalama varyans grafiği şekil 4.50'de gösterilmektedir.



Şekil 4.50 Ağır KF örneklerinde örneklerde şeritler arasındaki örnek varyansı

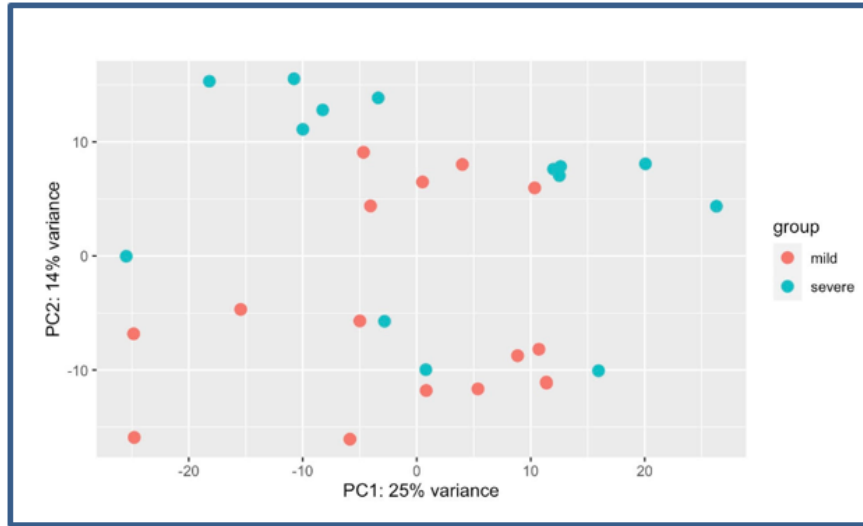
Şekil 4.50'de her bir gen için (siyah noktalar) ortalamanın bir fonksiyonu olarak çizilen Ağır KF örnekleri verilerinde şeritler arasındaki örnek varyansını gösterdik. Data verilerinin dağılımı ortalamamamanın varyansa eşit olmadığını gösteriyor. Ortalama ifadesi yüksek olan genler için, replikalar arasındaki varyans ortalamadan daha büyük olma eğilimindedir (dağılım kırmızı çizginin üzerindedir). Not: Kırmızı çizgi $x = y$ 'yi temsil eder.

Korelasyon değerlerini oluşturduktan sonra örnekler için heatmap oluşturduk (Şekil 4.51).



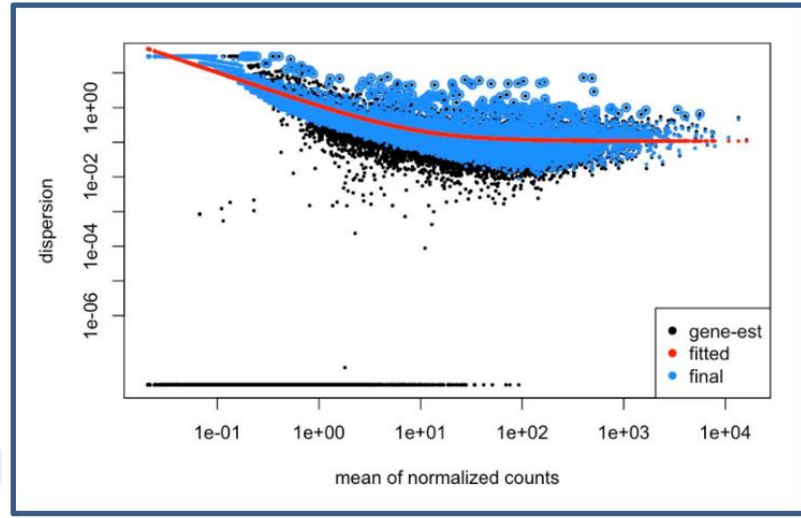
Şekil 4.51 Orta ve Ağır KF örneklerinde korelasyon grafiği

Daha önce de belirtildiği gibi veri setinde bulunan varyasyonu vurgulamak için PCA kullandık. Tez çalışmamızdaki verilerimizde Orta ve Ağır KF yakın ekspresyona sahip olduğu için tam ayrım gözlemlenmemiştir (Şekil 4.52).



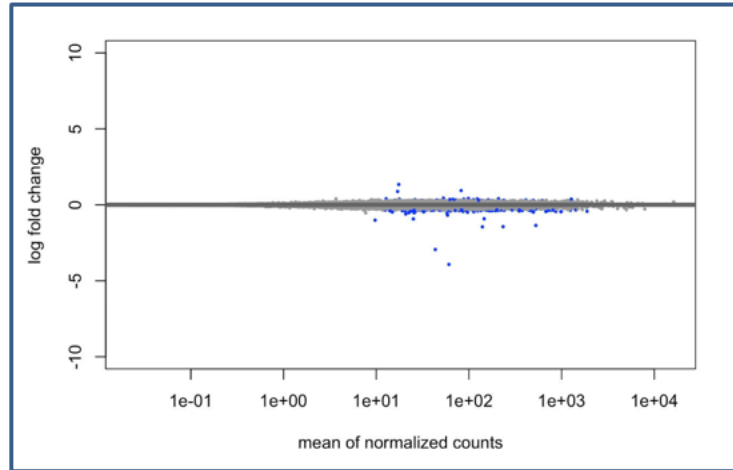
Şekil 4.52 Orta ve Ağır KF örneklerinde PCA grafiği

Örneklere ait artan ortalama ile dağılım düşmesi beklenmektedir. Bu dağılım grafiği şekil 4.53’de gösterilmektedir.



Şekil 4.53 Orta ve Ağır KF örneklerinde Dispersion grafiği

Ayarlanmış p-değeri < 0.05 'e göre, göre 140 transkript olarak düzenlenmiştir. Bunlardan 32 tanesi up regüle iken 108 tanesi down regüle olmuştur. Bu verilerin görselleştirdiği grafik şekil 4.54’de gösterilmektedir.



Şekil 4.54 Orta ve Ağır KF örneklerinde test edilen tüm genler için log2 kat değişikliklerine karşı normalize olan count verilerinin ortalamasını gösterimi

Şekil 4.54’de önemli ölçüde DE olan genler, kolayca tanımlanabilecek şekilde mavi renkli olarak gösterilmektedir.

DESeq paketinde bulunan summary fonksiyonu ile örneklere ait genel sonuç şekil 4.55’de gösterilmektedir.

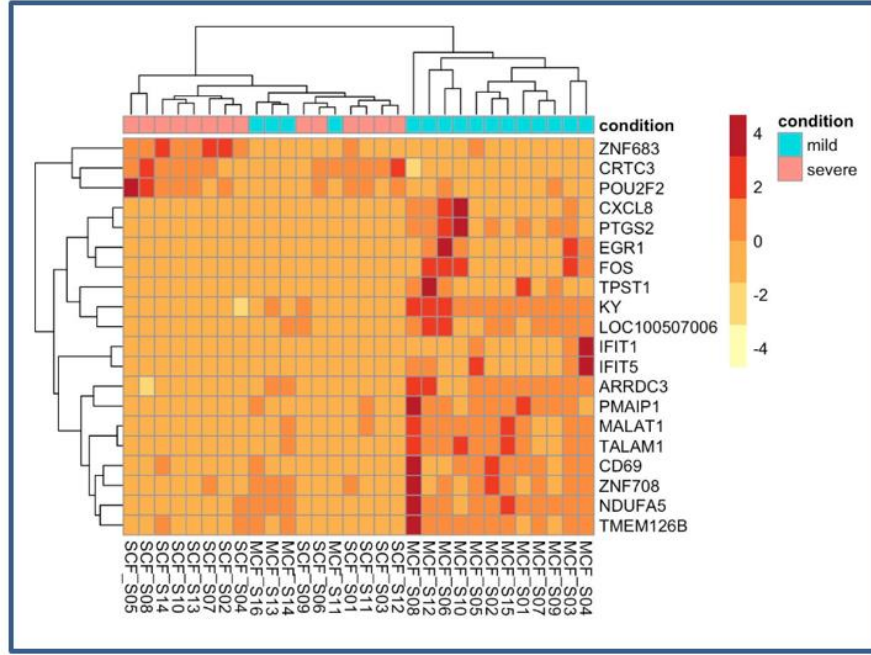
```
##
## out of 20038 with nonzero total read count
## adjusted p-value < 0.05
## LFC > 0 (up)      : 32, 0.16%
## LFC < 0 (down)    : 108, 0.54%
## outliers [1]      : 0, 0%
## low counts [2]    : 9884, 49%
```

Şekil 4.55 Orta ve Ağır KF örneklerinde total okumalara istinaden LCF genel sonuç

Tablo 4.24 Orta ve Ağır KF örneklerinin Padj değerine göre sıralanmış ilk 20 up ve down regüle DEG

	Genler	baseMean	log2FoldChange	lfcSE	pvalue	padj
1	CXCL8	43.393985	-2.939012556	0.53383611	1.26E-09	8.96E-06
2	EGR1	60.7656266	-3.915875372	0.76973648	1.76E-09	8.96E-06
3	KY	9.7253502	-1.012897261	0.44073849	5.87E-08	0.000198578
4	ZNF683	17.5102002	1.340775486	0.54800139	1.64E-07	0.000416079
5	LOC100507006	58.8286372	-0.68121359	0.33466689	3.06E-07	0.000527634
6	PTGS2	139.76164	-1.448157347	0.61409938	3.12E-07	0.000527634
7	NDUFA5	25.6281114	-0.52981992	0.22016994	1.54E-06	0.002236323
8	FOS	527.380221	-1.359178958	0.81125714	1.96E-06	0.002490791
9	ARRDC3	496.17347	-0.483041378	0.17747997	3.81E-06	0.003142531
10	CD69	20.8286886	-0.607844238	0.39369082	4.08E-06	0.003142531
11	CRTC3	123.344546	0.391881793	0.09741117	2.92E-06	0.003142531
12	IFIT1	233.692292	-1.435361325	1.04055736	4.64E-06	0.003142531
13	IFIT5	146.210971	-0.917138347	0.67859571	4.22E-06	0.003142531
14	MALAT1	990.149509	-0.445924175	0.12812546	4.47E-06	0.003142531
15	PMAIP1	22.421835	-0.52496549	0.26088467	3.65E-06	0.003142531
16	TALAM1	978.946642	-0.444002025	0.13028423	5.56E-06	0.003530552
17	TMEM126B	32.5062272	-0.460906789	0.16672151	7.57E-06	0.004272087
18	TPST1	25.1560054	-0.929456183	0.76302931	7.36E-06	0.004272087
19	POU2F2	208.488765	0.409511676	0.11740839	1.21E-05	0.006454321
20	ZNF708	23.2632592	-0.44559101	0.19513697	1.79E-05	0.009065553

Tüm önemli genlerin normalleştirilmiş değerleri çıkartılarak pheatmap() ile ekspresyonlarının heatmapını çizdik. 140 gen arasında ilk 20 gen için padj değerine göre elde edilen heatmap grafiği şekil 4.56’de gösterilmektedir.



Şekil 4.56 Orta ve Ağır KF örneklerinde up ve down regüle ilk 20 genin heatmap gösterimi

Orta ve Ağır KF örneklerinde log2FoldChange değerine göre sıralanmış en iyi 20 up regüle DEG Tablo 4.25’de gösterilmektedir. En düşüğe en yükseğe göre sıralanmıştır.

Tablo 4.25 Orta ve Ağır KF örneklerinde log2FoldChange değerine göre sıralanmış en iyi 20 up regüle DEG

	Genler	baseMean	log2FoldChange	lfcSE	pvalue	padj
1	VPS35L	62.9643021	0.325057233	0.10945634	0.00028945	0.03694489
2	FKBP4	72.4372694	0.326044714	0.11977327	0.00051111	0.045419272
3	CX3CR1	410.189535	0.330236298	0.13249367	0.00068345	0.049923216
4	AP1M1	161.613192	0.331905265	0.10572153	0.0001606	0.026739287
5	ADAM15	87.8537576	0.337725559	0.1326896	0.00052306	0.045419272
6	C11orf24	26.2863855	0.339540816	0.13700954	0.00053504	0.045419272
7	GALNT2	112.226331	0.341053582	0.13313094	0.00046783	0.04346072
8	DDX54	61.9161962	0.351780983	0.12937278	0.00028025	0.03694489
9	CD99L2	66.7944079	0.361965273	0.11627087	9.65E-05	0.020207754
10	ENO1	1272.07406	0.372933808	0.12369852	9.42E-05	0.020207754
11	SSBP3	88.196072	0.377154924	0.12403693	8.02E-05	0.018952142
12	AEBP1	17.7429437	0.377259562	0.21049322	0.00012567	0.023205501
13	CRTC3	123.344546	0.391881793	0.09741117	2.92E-06	0.003142531
14	DMWD	12.8107833	0.402434634	0.14678755	5.73E-05	0.016174696
15	POU2F2	208.488765	0.409511676	0.11740839	1.21E-05	0.006454321

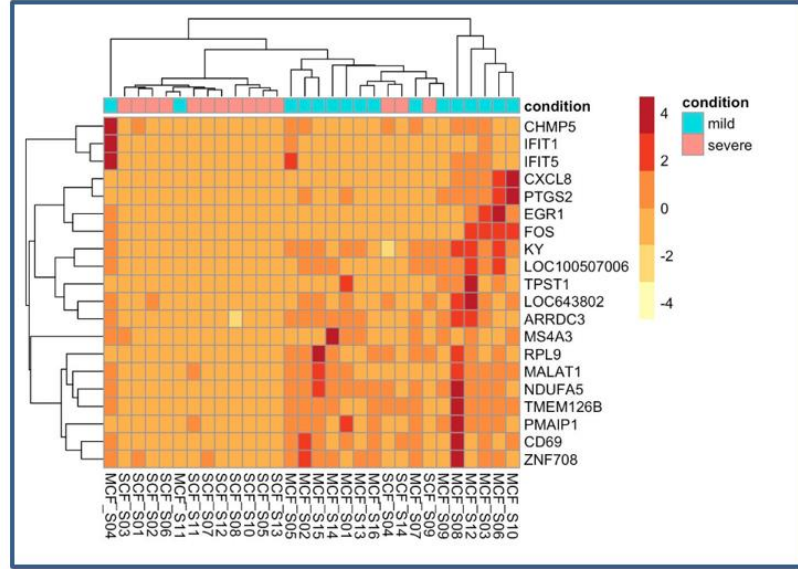
16	CDK5RAP2	98.619604	0.425299933	0.1565336	2.64E-05	0.010629026
17	HBG1	53.1163719	0.441600165	0.6926649	0.00012262	0.023205501
18	C4BPA	17.0129396	0.880191185	1.13225954	2.65E-05	0.010629026
19	HBA1	82.4778668	0.9402501	1.0448214	2.02E-05	0.009338838
20	ZNF683	17.5102002	1.340775486	0.54800139	1.64E-07	0.000416079

Orta ve Ağır KF örneklerinde log2FoldChange değerine göre sıralanmış en iyi 20 down regüle DEG Tablo 4.26’de gösterilmektedir.

Tablo 4.26 Orta ve Ağır KF örneklerinde örneklerinde log2FoldChange değerine göre sıralanmış en iyi 20 down regüle DEG

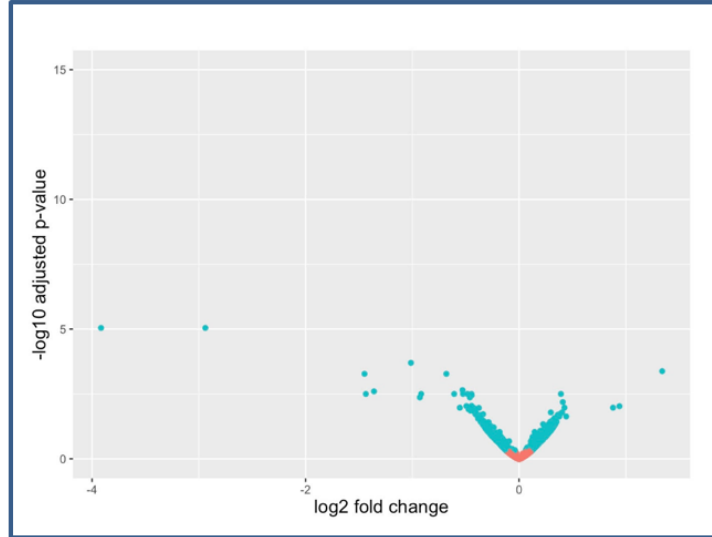
	Genler	baseMean	log2FoldChange	lfcSE	pvalue	padj
1	EGR1	60.7656266	-3.915875372	0.76973648	1.76E-09	8.96E-06
2	CXCL8	43.393985	-2.939012556	0.53383611	1.26E-09	8.96E-06
3	PTGS2	139.76164	-1.448157347	0.61409938	3.12E-07	0.000527634
4	IFIT1	233.692292	-1.435361325	1.04055736	4.64E-06	0.003142531
5	FOS	527.380221	-1.359178958	0.81125714	1.96E-06	0.002490791
6	KY	9.7253502	-1.012897261	0.44073849	5.87E-08	0.000198578
7	TPST1	25.1560054	-0.929456183	0.76302931	7.36E-06	0.004272087
8	IFIT5	146.210971	-0.917138347	0.67859571	4.22E-06	0.003142531
9	LOC100507006	58.8286372	-0.68121359	0.33466689	3.06E-07	0.000527634
10	CD69	20.8286886	-0.607844238	0.39369082	4.08E-06	0.003142531
11	LOC643802	57.9361924	-0.554853034	0.51623747	2.72E-05	0.010629026
12	NDUFA5	25.6281114	-0.52981992	0.22016994	1.54E-06	0.002236323
13	PMAIP1	22.421835	-0.52496549	0.26088467	3.65E-06	0.003142531
14	CHMP5	96.8039895	-0.492738144	0.34294703	1.90E-05	0.00919413
15	ARRDC3	496.17347	-0.483041378	0.17747997	3.81E-06	0.003142531
16	MS4A3	14.9191206	-0.473186858	0.39238772	3.74E-05	0.012241163
17	TMEM126B	32.5062272	-0.460906789	0.16672151	7.57E-06	0.004272087
18	RPL9	112.81447	-0.454734955	0.37851006	4.61E-05	0.013866338
19	MALAT1	990.149509	-0.445924175	0.12812546	4.47E-06	0.003142531
20	ZNF708	23.2632592	-0.44559101	0.19513697	1.79E-05	0.009065553

140 gen arasında log2FoldChange'e göre en az down regüle ilk 20 gen için elde edilen heatmap grafiği şekil 4.57'de gösterilmektedir.



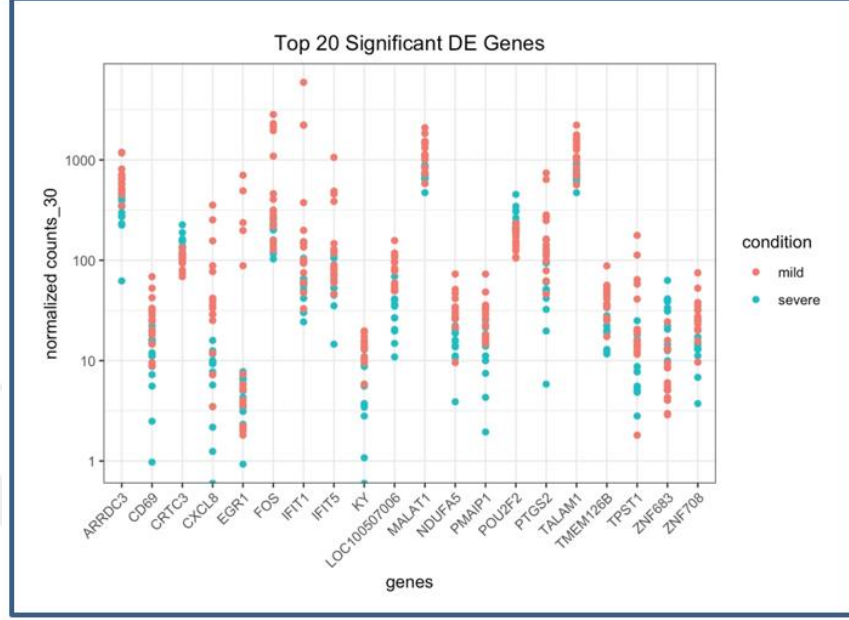
Şekil 4.57 Orta ve Ağır KF örneklerinde en az down regüle olan 20 genin heatmap gösterimi

Daha fazla global bir görünüm için Orta ve Ağır KF örneklerinde 140 gene ait Volcano Plot şekil 4.58'da gösterilmektedir.



Şekil 4.58 Orta ve Ağır KF örneklerinde volcano plot gösterimi

Orta ve Ağır KF örneklerinde Diferansiyel olarak ifade edilen ilk 20 gen için (padj değerleriyle) normalleştirilmiş count verilerinin grafiği şekil 4.59'da gösterilmektedir.



Şekil 4.59 Orta ve Ağır KF örneklerinde Diferansiyel olarak ifade edilen ilk 20 gen

5. Tartışma

RNASeq'deki ilerlemeler, son birkaç yılda transkriptomun incelenmesi ve yorumlanmasındaki en büyük devrimlerden biri olmuştur. Biyoinformatik sayesinde, veri analizi yazılımları ve platform teknolojilerindeki sürekli yenilikler ve gelişme ile maya gibi tek hücrelilerden memelilere kadar tüm organizmaların transkriptlerindeki çeşitli fizyolojik ve genetik anomaliler hakkındaki daha fazla soru cevap bulmaktadır (Ergin ve ark., 2022).

RNASeq, temel biyolojik araştırmalarda ve klinik uygulamalarda yaygın olarak kullanılan güçlü bir teknolojidir. Yüksek verimli RNASeq deneyleri, çeşitli kimyasal ve genetik bozulmaların transkriptomu nasıl etkilediğini anlamak için eşsiz bir fırsat sunmaktadır. Bu deneylerin gerçekleştirilmesindeki en büyük zorluk, pahalı olan ve yoğun emek gerektiren cDNA kütüphanesinin hazırlanmasıdır. Ayrıca çalışmada kullanılacak RNA'nın da miktarı önemlidir. RNA miktarları ile ilgili Naphade et al. (2022) tarafından yapılan bir çalışmada düşük girişli bir RNA kiti ile RNASeq çalışmaları için altın standart olarak kabul edilen TruSeq mRNA kitleri karşılaştırılmış. TruSeq, SMARTer ve SMARTer Ultra-Low RNASeq kitlerinin sistematik analizi (Palomares et al., 2019) ile ribodepletion kitleriyle birlikte kullanılan SMARTer kiti, TruSeq kitlerine kıyasla çoklu metriklerde 100 ve 10ng girişler için performansı düşürdüğü gösterilmiştir. Üstelik ribodepletion için RiboZero adımıyla ilişkilendirildiğinde, yöntemin otomasyon için uyumsuz olduğu ortaya çıkarılmıştır. Diğer taraftan, SMARTer Ultra-Low kitinin <1 ng ultra düşük başlangıç miktarı için nispeten iyi performansa sahip olduğu ve otomasyonla uyumlu olduğu gösterilmiştir. Strand özgüllüğü ve TruSeq kitiyle karşılaştırıldığında daha düşük başlangıç miktarı açısından değerlendirilen diğer RNASeq kütüphane hazırlama yöntemleri arasında, strand özgüllüğünü dikkate almayan Takara Bio SMART-Seq V4 Ultra Düşük Girişli RNA ve SMARTer Stranded Total RNASeq Kit v2-Pico Input (Pico) kitleri yer almaktadır (Sarantopoulou et al., 2019). Pico kiti, iplik özgüllüğünü muhafaza etmesine ve TruSeq kiti ile iyi bir uyuma sahip olmasına rağmen, Pico kiti kullanılarak oluşturulan kütüphanelerin %55 daha az diferansiyel olarak eksprese edilmiş gen

sayısı göstermiştir (Sarantopoulou et al., 2019). Illumina TruSeq kitleri ile önceki nesil Ovation RNASeq sisteminin karşılaştırmalı analizinde kitin 500 pg-10ng toplam RNA arasında değişen düşük girdiler için neredeyse eşit gen temsiliyle iyi bir performans gösterdiği ortaya çıkarılmasına rağmen, bu kütüphaneler platforma özel bir 5' veya 3' yanlılığı sergilemiştir (Tariq et al., 2011). Naphade et al. (2022) tarafından yapılan başka bir çalışmada kütüphaneleri geliştirmek için, iş akışının kolaylığına, otomasyon vaadine ve elde edilen biyoinformatik analizlerin doğruluğuna odaklanan üç yöntem Swift RNA, Swift Rapid RNA ve Illumina TruSeq stranded mRNA kitleri sistematik olarak karşılaştırılmış. Çalışmaya göre Swift RNA kütüphane kitinin daha geniş bir giriş aralığı barındırdığı ve çeşitli örnekler arasında tutarlı ve güvenilir sonuçlar sunduğu gösterilmiştir. Bu sıralı adaptör ligasyonuna izin veren ve girişler arasında adaptör titrasyonunun önlenmesine yardımcı olan patentli Adaptase teknolojisinin kullanılmasıyla elde edilmektedir. Bu da tüm girişlerde tek tip adaptör konsantrasyonunu korumaktadır. Bunlara ek olarak, Adaptaz, YND adaptörlerinin ssDNA'ya eşzamanlı olarak yüklenmesini ve bağlanmasını gerçekleştirmektedir. Adaptaz ayrıca tek bir hücre kadar düşük DNA girişlerine izin vermekte ve yüksek verimli şablondan bağımsız kimyası, minimum baz kompozisyon önyargısı ile geniş bir kapsam sağlamaktadır. Bu teknoloji, 10ng toplam RNA'ya kadar geniş bir giriş aralığını desteklemektedir. FFPE RNA gibi hasarlı örneklerden tutarlı performans da bildirilmiştir. Illumina ve Swift kitleri arasındaki iş akışı süresindeki farklılıklar, farklı kit kimyaları ile ilişkilendirilmektedir. Illumina TruSeq Stranded mRNA kitinde, iplik özgüllüğü, dTTP'nin dUTP ile değiştirilmesi ve ardından DNA Polimeraz I ve RNase H kullanılarak ikinci iplik cDNA sentezi ile mecburi kılınmaktadır. Bu yöntem, ikinci iplik üzerinde yeni eklenen urasil bazını çıkarmak için Urasil-DNA-Glikosilaz kullanarak çalışmaktadır. Böylece bu ipliğin PCR'de bir şablon olarak kullanılması engellenmektedir. Diğer taraftan, Swift kitleri için daha hızlı iş akışı, birinci zincir cDNA'dan kütüphane hazırlamayı destekleyen ve böylece ikinci zincir sentezi ihtiyacını ortadan kaldıran tescilli Adaptaz teknolojisi ile kolaylaştırılmıştır. Bunlara ek olarak, Swift RNA iş akışı, düşük kaliteli, düşük RIN (RNA Integrity Number) FFPE örnekleriyle de uyumludur. Aynı çalışmada incelenen tüm yöntemlerin, dizileme ve hizalama metrikleriyle ölçüldüğünde yüksek kaliteli kütüphaneler ürettiği de kanıtlanmıştır. Ayrıca, bu yöntemlerle üretilen kantitatif gen sayılarının da büyük ölçüde uyum içinde olduğu gösterilmiştir. Ancak, bu yöntemler, giriş RNA miktarına

duyarlılıklarında farklılık göstermektedirler. Üretici tarafından tavsiye edilen RNA girdi miktarı referans olarak alınarak, örneğin *in vitro* kök hücre modelleri tarafından makul bir şekilde üretilen bir yield olan 50ng'de farklı şekilde eksprese edilen genlerin doğruluğu yine aynı çalışmada değerlendirilmiştir. Swift RNA kitinin bu tür analizlerde kullanılan tipik önem seviyelerinde, yalnızca teknik hususlar nedeniyle farklı şekilde ifade edilen genler için en az “yanlış pozitif” oranına sahip olduğu bulunmuş. İlginç olarak da, genlerin ve yolakların farklı ekspresyonunun girdi miktarına bağımlılığı, ele alınan kitler arasında farklılık göstermiştir. Swift RNA kiti, giriş miktarına göre en az sayıda farklı şekilde ifade edilen KEGG yolağını üretmiş. Aynı çalışmada teknik artifaktlara dayalı sonuçlardan kaçınmak için araştırmacıların, yolakların ekspresyonunda en az %10'luk değişikliklere odaklanması gerektiği savunulmaktadır. Tez çalışmamızda kullandığımız 1. veri kaynağımızda başlangıç olarak kullanılan RNA miktarı 100ng ve kullanılan kütüphane hazırlık kiti ise QuantSeq 3'mRNA-Seq Library prep kit (Lexogen, Vienna, Austria)'tir. Diğer veri kaynağımızda TruSeq™ Total RNA Sample Prep Kit talimatlarına göre RNA kullanılmış.

Biz tez çalışmamızda iki farklı RNASeq ön işleme adımını kullanarak yüksek güvenilirlikli değişken çağrılarını yapmaya vurgu yaparak bir RNASeq varyant çağrı pipeline kullandık. Bunlar "GATK en iyi uygulamaları" (GATK olarak anılır) ve Samtools'tur. Ayrıca, yukarıdaki ön işleme yöntemlerinden birinin ve HaplotypeCaller (HC) varyant çağırma aracının bir kombinasyonu olarak tanımlanan bir varyant çağırma "yöntemi" ensemble kullandık. Bu pipeline aynı zamanda normal örneklerle bağlı kalmadan varyant çağrılarını yapmak için geliştirilmiştir. Örneğin kanserde varyant çağrı çalışmaları için sıklıkla kullanılan bu tür örnekler, KF veri seti için mevcut değildi. Yalnızca RNASeq verilerinden varyant çağrısı, hiçbir standardın olmadığı gelişmekte olan bir alan olarak kaldığından, yüksek güvenilir varyant çağrılarını yapmak, pipeline için katı bir filtreleme kriteri seti kullanılarak geliştirilmesi gerekmektedir. Bunun amacı gerçek pozitifler olarak adlandırılan değişkenlere en yüksek güveni sağlamak ve artefaktların sekanslanmasından, mevcut en iyi uygulamalarda bilinen eksikliklerden veya araç yanlışlıklarından kaynaklanan hatalardan kaçınmaktır. Tezimizde Referans tabanlı bir değişken çağırma yöntemi kullandık. Bu yöntem RNASeq'i bir referans genom ile hizalayıp veriler ve referans arasındaki varyantları tanımlanabilmesine izin vermektedir. Bu hizalama sonucunda,

her örnek için bir SAM dosyası elde ettik ve bu dosyadan mevcut bilgisayar kaynakları üzerindeki dosya verisini en aza indirmek için BAM dosyası oluşturduk. İnsan genomunun en son sürümü olan GRCh38 referansını kullandık. "Tam analiz seti" referansını kullandığımız için HISAT2 hizalama aracı seçtik. "Tam analiz" referansının popülasyonlar arasında var olan bilinen germline varyantlarını sağlar ve yanlış pozitifleri azaltması gibi faydaları bulunmaktadır. Referansa karar verdikten ve BAM dosyalarımızı oluşturulduktan sonra, iki bağımsız yöntem kullanılarak özel olarak RNASeq varyant çağrısı için önceden işlenmiş olan BAM dosyalarında varyantların bulunmasını istedik. Broad Institute, Genome Analysis Toolkit (GATK) kullanarak RNASeq verilerinden varyantları çağırmak için bir dizi "en iyi uygulama" geliştirmeye başlanmıştır. Bununla birlikte, halen geliştirilmekte olan metodolojiler yanlış pozitifleri içerebilen veya gerçek pozitifleri hariç tutabilen katı filtrelerin kullanımına dayanmaktadır.

Geleneksel olarak, KF akciğer hastalığının *in vitro* çalışması, KFTR mutasyonları taşıyan hava yolu epitel hücrelerinin çalışmasına odaklanmıştır. Mazio et al. (2020) tarafından yapılan, tezimizin de ilk veri kaynağı olan, çalışmada epiteli etkileyen çeşitli hastalıklarda stromal ortamın rolü ile tutarlı olarak, pulmoner KF stromasının, *in vivo* olarak akciğer hastalığının ilerlemesi üzerinde önemli bir etkisi olabileceği ve göz ardı edilemeyecek anormalliklere sahip olduğu gösterilmiştir. Aslında, bu çalışma sonuçları ile birincil CF-HLF ve N-HLF arasında var olan morfolojik ve transkriptomik farklılıklar ile bu özellikleri *in vitro* olarak vurgulamak için 3D ortamının önemi gösterilmiştir. Bu çalışmada KF bağ hava yolu dokusu (CF-CAT) 3D biyo-mühendislik modeli geliştirilmiş ve özellikleri normal dokusu (NA-CAT) ile karşılaştırılmıştır. Özellikle, CF-CAT, aşağıdan yukarıya bir yaklaşım kullanılarak *in vitro* olarak elde edilmiş ve kendi doğal ECM'lerine gömülü fibroblastların varlığı öne çıkarılmıştır. Bu çalışmanın araştırmacıları daha önce aynı yaklaşımı kullanarak, epitelyal ve stromal kompartmanlar arasındaki etkileşimin yanı sıra organa özgü ECM'nin bileşimi ve özellikleri açısından farklı dokuların fizyopatolojik ortamını *in vitro* özetleme olasılığını daha önce de göstermiştir (Casale et al., 2016; Brancato et al., 2017; De Gregorio et al., 2020). KF epitel hücrelerinde KFTR mutasyonundan kaynaklanmaktadır. Ancak bu patolojinin ciddi komplikasyonları tüm pulmoner dokuyu içermekte ve akciğer fonksiyonlarını önemli ölçüde tehlikeye atmaktadır. Gerçekten de mukus tıkanması, bakteriyel enfeksiyon

(Ratjen et al., 2015), inflamatuvar yanıt (Su et al., 2011; Nichols et al., 2015;) ve dokunun yeniden şekillenmesi (Gaggar et al., 2011; Mack et al., 2018;) tümü KF'nin merkezi özellikleridir ve varsayılan terapötik hedefleri temsil etmektedirler. Bu içerikte, bu çalışmaya ait model, KF sırasında meydana gelen akciğer dokusu yeniden şekillenmesini ve fibrozu benzersiz bir şekilde özetlemiştir. Aslında, KF'li hastalardan gelen akciğer fibroblastları, hücre proliferasyonu, ECM elemanlarının üretimi, kollajen lif kalınlığı ve ağın anormal organizasyonunda artış ile birlikte aktive ve fibrotik bir fenotip sergilemişlerdir. Çalışma ile ECM proteinleri arasında CF-HLF'nin daha yüksek miktarda kolajen I, hyaluronik asit ve periostin ürettiğini bulunmuştur. İkincisi, akciğer fibrozu ve inflamasyonun belirteci olarak kabul edilmiştir (O'Dwyer et al., 2017). Ayrıca, periostinin farklı patolojik koşullarda kolajen sentezinde (Norris et al., 2007), toplanmada ve organizasyonda rolü olduğu bildirilmiştir (González-González et al., 2018). Periostinin ayrıca ECM'nin diğer proteinlerini ve kemik morfojenik proteini 1'i (BMP-1) bağlayarak kollajen liflerinin çapraz bağlanmasını katalize ederek doku sertliğini de arttırmaktadır (Kudo et al., 2018). Aynı zamanda periostin, HAS-2 fonksiyonunu ve hyaluronik asit sentezini aktive etmektedir (Ghatak et al., 2014). CF-CAT'de bulunan yüksek miktarda hyalüronik asit ve HAS-2'nin aşırı ekspresyonu da şiddetli fibrozun belirteçleri olarak sunulmuştur (Li et al., 2011). KFTR ekspresyonu olmamasına rağmen CF-HLF davranışında önemli anormalliklerin varlığı, muhtemelen kronik patoloji sırasında hücrelerin *in vivo* aktivasyonunun bir sonucu olarak yorumlanmıştır. CF-HLF'nin *in vitro* olarak, KFTR fonksiyon kaybından başlayan ve *in vivo* şiddetli inflamasyona yol açan olaylar zincirinden kaynaklanan epigenetik değişiklikleri koruduğu görülmüştür. Bu hipotez, *in vitro* primer hücrelerin, *in vivo* hücre durumu ile ilgili bir epigenetik hafızayı koruyabildiği gerçeğine dayanmaktadır (Ivanov et al., 2016). Ek olarak, fizyolojik bir 3D ECM ortamının varlığı, hedeflenen epigenetik tedaviler için klinik olarak ilgili bir modeli temsil etmektedir (Ozes et al., 2013; Amatangelo et al., 2013). Aslında, KF, epitelde KFTR mutasyonunun neden olduğu monojenik bir hastalıktır. Bununla birlikte, epitelyal disfonksiyonun sonuçları, enfeksiyona, tekrarlayan streslere ve reaktif oksijen türlerinin aşırı üretimine karşı artan duyarlılıkla birlikte tüm akciğer dokusunu etkilemektedir. Bu olayların hastalık fenotipini ve şiddetini modüle eden benzersiz DNA metilasyon profillerinin oluşturulmasından sorumlu olduğu bilinmektedir. KF hastalarından alınan nazal epitel ve kan hücresi örneklerinde değiştirilmiş metilasyon

seviyeleri yapılan çalışmalar ile bulunmuştur (Scott ve ark., 2020; Magalhães ve ark., 2017). KF'deki değiştirilmiş akciğer ortamının, stromal fibroblastların DNA metilasyon profilini de etkileyebileceği varsayılmıştır. Bunun epitel bütünlüğünün kaybının ve yaşlanmayla ilişkili olayların epigenomik değişiklikleri ve patolojinin gelişmesine yol açan stokastik profibrotik metilasyon kaymasını tetiklediği akciğer fibrozu için gösterildiği gibi olabildiği öne sürülmüştür (Selman et al., 2014; Helling et al., 2015). Hem veri kaynağımızı aldığımız çalışmada hem de tezimizde hücrelerin epigenetik profili değerlendirilmedi ancak ileri çalışmalar için önerilmektedir. Veri kaynağımızda CF-HLF'nin transkriptomik profiline odaklanılmıştır. Çalışmanın sonuçları, ekstraselüler boşluk bölmesindeki elementlerin üretimi ve doku/organ morfogenez ve hücre çoğalmasının biyolojik süreçleri için CF-HLF up regüle genlerinin olduğunu göstermiştir. Aynı zamanda, akciğer fibrozunda rol oynadığı gösterilen WNT yolağı, TBX4 ve ITGA8 genleri aşırı eksprese olmuştur (Königshoff et al., 2008; Xie et al., 2016; Munguía-Reyes et al., 2018; Hung et al., 2018). Kendi belirlediğimiz analiz hattımızla RPS28 geninin hem 3D hem de 2D örneklerinde aşırı eksprese olduğunu bulduk. RPS28 aynı zamanda her iki örnek grubunda da up regüle genler arasında bulunmaktadır. RPS28 geni, muhtemelen ribozom biyogenezinde yer alan bir ribozomal proteini kodlar (Gripp et al., 2014). Yakın zamanda yapılan bir çalışmada RPS28 geninin akciğer dokularında aşırı eksprese olduğu gösterilmiştir (Na et al., 2019). Çalışma grubumuzun her ikisinde de aşırı eksprese olduğunu bulduğumuz bir diğer genimiz CLU genidir. Bu gen de hem 2D hem de 3D KF örneklerinde up regüle olmuştur. CLU, insan prostat ve meme kanserlerinde ve skuamöz hücreli karsinomlarda aşırı eksprese olmaktadır. CLU'nun baskılanması, bu hücreleri kemoterapötik ilaç aracılı apoptoza duyarlı hale getirmektedir (<https://www.omim.org/entry/185430>, Erişim tarihi 04 Şubat 2022).

Sadece 3D KF örneklerinde A2M geni aşırı eksprese olmuştur. Aynı zamanda up regüle gendir. Poller et al. (1989), çocukluktan beri serum A2M eksikliği ve kronik akciğer hastalığı olan bir hastada A2M geninde bir değişiklik tespit etmiştir. Çalışmada hastada A2M değişikliği için heterozigot olarak bulunmuş ve yazarlar bunun akciğer hastalığından sorumlu olduğunu öne sürmüştür. Yapılan başka bir çalışmada kronik akciğer hastalığı olan 30 hastanın l'inde tiyolester bölgesi içinde cys972'yi (TGT) tyr'e (TAT) değiştiren bir mutasyon bulunmuş. Tetramerik A2M molekülünün alt birimlerinin her birinde cys972 ve gln975 arasında oluşturulan dahili

tiyolesterin aktivasyonu, aktive edici proteinazın kovalent çapraz bağlanmasına dahil olduğundan, bu mutasyonun A2M fonksiyonuna müdahale ettiği tahmin edilmiştir (Poller et al., 1992). Biz de aynı zamanda Samtools ile yaptığımız varyant çağrısı ile 3D CF_HLF_Rep3 örneğimizde A2M geninde ekzonik bölgede T'den C'ye (start: 9079672 end:9079672; rs669) olan değişimi Benign/ risk faktör olarak bulduk. Bu nedenle tam olarak anlamlı bir sonuç olarak değerlendirmedik. Ancak bu sonuç ileride yapılacak çalışmalar için bir bulgu olarak kullanılacaktır.

BGN geni sadece 3D KF örneklerinde aşırı eksprese olan bir gen olarak bulunmuş. Bu gen aynı zamanda hem 3D hem de 2D örnekleri de up regüledir. BGN, lösin açısından zengin küçük bir proteoglikandır. Eklem kıkırdağının önemli bir yapısal bileşenidir ve tip VI kollajen ve büyük proteoglikan agregatları ile protein etkileşimlerinin oluşumu yoluyla kondrosit hücre dışı matrisinin montajına katılmaktadır. X kromozomunda bulunan bu gen ayrıca hücre sinyalleşmesinde de rol oynamaktadır (Iacob and Cs-Szabo, 2010). BGN geninde meydana gelen mutasyonlar Meester-Loeys sendromu ve Spondiloepifizyal displazi ile ilişkilidir. Morimoto et al., (2021) tarafından yapılan bir çalışmada BGN'nin akciğer kanseri için potansiyel bir biyomarker olduğu öne sürülmüştür. Data kaynağımız olan birinci veri setimizde PPS28, CLU, A2M, BGN genleri ile ilgili herhangi bir bilgiye rastlanmamıştır.

Ek olarak, akciğer fibroblastlarının KFTR eksprese olmadığını doğrulanmasna rağmen, CF-HLF'nin "İyon taşınmasının düzenlenmesi" (GO:0043269) ve "İyon taşınmasının olumsuz düzenlenmesi" (GO:0043271) terimlerinin up regüle olarak düzenlediği değerlendirilmiştir. Bu tür aşırı ekspresyonun, epitel hücrelerinde *in vivo* KFTR fonksiyonunun eksikliğinin bir sonucu olabileceği varsayılmış. Diğer taraftan, transkriptomik analiz, CF-HLF'nin EMT'yi aktive etme potansiyeline sahip olabileceğini düşündüren "Mezenkimal hücre farklılaşması" (GO:0048762), "Mezenkim gelişimi" (GO:0060485) ve "Mezenkimal hücre proliferasyonu" (GO:0010463) biyolojik süreçlerinde yer alan genlerin CF-HLF'yi up regüle ettiğini göstermiştir. Özellikle, EMT'ye dâhil olan genleri (FAM83D, WNT2, MSX2, SFRP2, HAS2, TMEM100, AXIN2, BMP7 ve TWIST1 gibi) içeren GO Terimi "Mezenkimal hücre farklılaşması"nın (GO:0048762) up regülasyonu, stromal akciğer fibroblastlarının *in vivo* olarak EMT'nin uyarılmasında bir rolü olabileceği varsayımına yol açmıştır. Bu fenomen diğer akciğer hastalıklarında iyi karakterize

edilmiştir ve son zamanlarda KF ile ilişkilendirilmiştir (Rout-pitt et al., 2018). İlginç olan ise EMT süreci her zaman akciğer fibrozunun itici mekanizması olarak kabul edilmiştir (Rout-pitt et al., 2018; Salton et al. 2019). Ancak burada fibroz sırasında akciğer fibroblastlarının aktivasyonunun sırayla EMT'yi uyurabileceği gösterilmiştir. Ayrıca yukarıda bahsedilen GO terimlerinin yukarı regülasyonu, modelimizde fibrotik durumların tipik bir örneği olarak fibroblast-miyofibroblast farklılaşmasını ve miyofibroblast biyogenezini desteklemiştir. Üstelik GO teriminin "Dallanan bir epitelin morfogenezi" (GO:0061138) 2D'de değil, yalnızca 3D kültür koşulunda zenginleştirildiği vurgulanmıştır. Daha önceki çalışmalarda, akciğerin *in vivo* gelişimi sırasında epitel dallanması için stromal bölmenin varlığının gerekli olduğu gösterilmiştir (Ribatti et al., 2014). Benzer şekilde, burada bağ dokusunun 3D *in vitro* ortamda epitel yapılarının morfogenezinde merkezi bir rolü olduğu gözlemlenmiştir. Dahası veriler, fibroblastlardan gelen artan bir uyaran nedeniyle KF'de dallanan epitel yapılarının oluşumunun değişebileceğini göstermiştir. Bu tür biyolojik sürecin, KF'de gözlenen submukozal bez morfolojisi ve fonksiyonunun düzensizliğine dâhil olabileceği varsayılmıştır. Aslında, bu yapılar *in vivo* bir dallanma morfogenezi (Anderson et al., 2017) süreci ile gelişir ve hastalık durumunda hiperplazik ve mukus tıkalıdır (Salinas et al., 2005; Hays et al., 2006). Ayrıca, submukozal bez değişikliği, KF fizyopatolojisine katkıda bulunan önemli bir faktör olarak kabul edilmiştir (Inglis et al., 2005). Submukozal bezler *in vivo* olarak epitelten alttaki bağ dokusuna dallanan epitelyal hücre süreciyle gelişmektedirler. Spesifik olarak, süreç epitelde plakod oluşumunun birinci aşamasını, ardından hücrelerin stromaya doğru uzaması, dallanması ve son olarak farklılaşma aşamasını içerir (Anderson et al., 2017). Önceki çalışmalarda, submukozal grandların morfolojisinin ve fonksiyonunun, büyük olasılıkla bu bezlerdeki yüksek oranda epitel hücrelerinin normal durumda KFTR'yi eksprese etmesi gerektiğinden, KF'de önemli ölçüde tehlikeye girdiği gösterilmiştir. KFTR defeksiyonu sonucunda submukozal bezler hiperplastik hale gelir ve KF'de mukus tıkanır (Salinas et al., 2005; Hays et al., 2006). Burada, CF-CAT'de "dallanan bir epitelin morfogenezi" için genlerin aşırı ekspresyonu, *in vivo* hastalık durumunda epitel yapılarının morfogenezinin düzenlenmesinde ECM'nin bir rolünün olduğunu düşündürmektedir. Diğer taraftan, 3D KF modelinde GO teriminin "dallanan epitelin morfogenezi"nin up regülasyonu, epitel hücreler olmadan kültürlenmiş olsa bile, fibroblastlar, böyle bir fizyolojik ilgili *in vitro* ortamda, *in vivo* meydana gelen

epitelyal stromal apraz karışmanın 'hafızasını' depolayabilmiştir. Bu yüzden, CF-CAT'in *in vitro* olarak submukozal bezlerin oluşumu için 'biyolojik toprak' olarak ve gelecekteki çalışmalarda KF'deki disfonksiyonlarının incelenmesi için kullanılabilir. Aynı zamanda, CF-HLF, "İnflamatuvar yanıt" (GO:0006954) ile ilgili genleri *in vitro* olarak da enflamatuvar uyaranlar olmaksızın aşırı eksprese olmuştur. Gerçekten de inflamasyon, yapısal akciğer dokusu değişikliği ve solunum yollarının kalıcı bakteri kolonizasyonu nedeniyle KF'nin bir özelliğidir (Nichols et al., 2015; Su et al., 2011). Bu çalışmanın sonuçları ile akciğer fibroblastlarının, KF hastalarında gözlemlenen güçlendirilmiş inflamatuvar yanıtta katkıda bulunduğu gösterilmiştir (Huaux et al., 2013). Enflamatuvar yanıt genleri arasında, CF-CAT'de ILR6'nın up regülasyonunu belirlenmiştir. Bu gen, immün yanıtta önemli bir rol oynayan güçlü bir pleiotropik sitokin olan interlökin 6 (IL6) reseptör kompleksinin bir alt birimini kodlamaktadır. IL6R, transmembran veya çözünür olabilir. IL6R'nin çözünür formunun, pro-inflamatuvar aktivitelere aracılık edebilen trans-sinyalleme mekanizması yoluyla IL6 tarafından hedeflenen hücrelerin türünü ve sayısını arttırdığı bilinmektedir. Bu nedenle IL6R, çeşitli inflamatuvar hastalıkların tedavisinde kullanılan bir monoklonal antikorun (tocilizumab) reseptörüdür (Rose-John et al., 2006; Scheller et al., 2011). Aktive fibroblastlar tarafından IL6R ekspresyonu, diğer inflamatuvar durumlarda zaten gösterilmiştir (Ebihara et al., 2011) ve veri kaynağımızda da KF akciğer stromasındaki katılımı rapor edilmiştir. Biz tez çalışmamızda IL6R up regüle genini gösteremedik. Veri kaynağımızda, transkriptomik analizin, stroma ve epitelyum, iltihaplanma ve fibroz arasında yer alan genlerin ifadesinde 3D mikro ortamın önemi vurgulanmıştır. Bütün bu süreçler, insan patolojisinin evriminde ima edilmektedir. Aslında, stromal/epitelyal karışımın hem doku morfogenezi hem de fonksiyon/disfonksiyon üzerinde sonuçları vardır ve KF hastalarında ciddi komplikasyonların ortaya çıkmasından inflamatuvar yanıt ve fibrozis sorumludur. Veri kaynağımız, bu patolojinin incelenmesi ve terapötik stratejilerin geliştirilmesi için KF'deki bağ hava yolu dokusundaki değişikliklere ışık tutmuştur. Veri kaynağımızın bulguları, CF-CAT'nin akciğer fibrozu ile ortak ilginç özellikler gösterdiğini kanıtlamıştır. Bu bağlamda, 3D *in vitro* stroma akciğer modelinin *in vivo* kronik patoloji sırasında doku fonksiyonunun kaybından sorumlu önemli değişiklikleri vurgulayabildiğine inanılmıştır. Özellikle, CF-CAT'nin insan patolojisinin geç aşamasını, enfeksiyon ve iltihaplanmalardan sonra klinikte tedavisi zor olan fibrozisi özetlediğine inanılmıştır. Bu veri kaynağımızın

içeriğinde CF-CAT'de daha yüksek miktarda kolajen tip-I, hyaluronik asit ve periostin ve CF-HLF tarafından çeşitli transkriptlerin up regülasyonu belirlenmiştir. Bunlar arasında TGF3, HAS-2, TWIST1, WNT2, SFRP2, RSPO2, CHI3L1, TBX4, ITGA8, HHIP yer almaktadır. Biz de tez çalışmamızda benzer olarak TWIST1, WNT2, RSPO2, CHI3L1, TBX4 genlerinin up regüle olduğunu gösterdik. Bu tür belirteçlerin, akciğer fibrozisi olan KF hastalarında akciğer fonksiyonunu kurtarmayı amaçlayan yeni terapötik stratejilerin hedefleri olabileceği de tartışılmıştır. Çok büyük ihtimalle, bağ dokusu değişiklikleri epitel hasardan kaynaklanır ve sırayla epitel fonksiyonunu bozmaktadır. Üstelik 3D CF-CAT modelinin, hastalığın hastaya özgü durumunu stroma düzeyinde kopyalamak için uygun olduğuna da inanılmıştır. Veri kaynağımız çalışmamızda, benzer yaştaki hastalardan gelen CF-HLF ve N-HLF'yi karşılaştırması seçilmiştir. Özellikle, yetişkin donörlerden alınan hücreler kullanılmıştır: CF-HLF için 45 ve N-HLF için 47/52 yaş. Genç hastalardan gelen hücrelerin genellikle yaşlılardan daha aktif olduğu gösterilmiştir (Kim et al., 2015).

TWIST1 geninin erişkin kalp ve iskelet kasında daha düşük ekspresyon, böbrek ve pankreasta zayıf ekspresyon saptanırken beyinde bir değişiklik bulunmamıştır. Endometriyal fibroblastlarda, peritoneal mezotelyal hücrelerde ve fetal akciğer fibroblastlarında ekspresyon saptanmış, ancak incelenen diğer hücre dizilerinde saptanmamıştır

(<https://www.omim.org/entry/601622?search=TWIST1&highlight=twist1>, Erişim tarihi: 01 Şubat 2022). Quaresma et al., (2020) tarafından yapılan bir çalışmada, TWIST1'in KF dokularında/hücrelerinde en tutarlı şekilde belirgin şekilde artan EMTa-TF (EMT) olduğu bulunmuştur. Bu da epitelyal belirteçlerin önemli ölçüde baskılanması olmadan mezenkimal belirteçlerin gözlenen up regülasyonu ile uyumlu olduğunu göstermektedir.

WNT2, normalde plasentada ve yetişkin dokularında esas olarak akciğerde eksprese edilen salgılanmış bir sinyal proteinini kodlamaktadır (Aavikko et al., 2021). Bağırsak kök hücre regülasyonu için çok önemli olan kanonik WNT sinyal yolağını aktive edebilmektedir. Kanonik WNT yolağının anormal aktivasyonu, kolorektal karsinogenezde önemli bir erken olay olarak kabul edilir ve özofagus, kolorektal ve mide kanserlerinde WNT2'nin aşırı ekspresyonu rapor edilmiştir (Fu et al., 2011; Kramer et al., 2017; Zhang et al., 2018). Wassink et al., (2001) WNT2'yi otizm için

aday bir gen olarak incelemiştir. Bunun nedenlerinden biri WNT gen ailesi, merkezi sinir sistemi de dahil olmak üzere çok sayıda organ ve sistemin gelişimini etkilemektedir. Diğer bir neden ise WNT2, otizmle bağlantılı 7q31-q33 bölgesinde bulunur ve otizmlili bir bireyde kromozomal bir kırılma noktasına bitişik olmasıdır. Başka bir neden de WNT yolağının işlevi için gerekli olan bir gen ailesinin bir üyesi olan DVL1 geninin bir fare nakavtı, öncelikle azalmış sosyal etkileşim ile karakterize edilen bir davranışsal fenotipi sergilemesidir.

Jackson et al., (2020) nötrofil homeostazının ve endotel bariyer fonksiyonunun düzenleyicisi olarak akciğerde RSPO2'nin beklenmedik ve biyolojik olarak önemli bir rolü olduğunu göstermişlerdir. Gelişmekte olan akciğerde yüksek RSPO2 ekspresyonu göz önüne alındığında, RSPO2 ayrıca doğumda hava solumaya geçiş sırasında meydana gelen mikrovasküler geçirgenliğin dinamik düzenlenmesinde rol oynayabilir (Miserocchi et al., 1994; Miserocchi et al., 2001). Zhang et al., (2020) tarafından yapılan başka bir çalışma RSPO3'ün insan fibrotik akciğerleri ve karaciğerlerinde oldukça yükseldiği ve RSPO3 düzeylerinin hastalık şiddeti ile önemli ölçüde ilişkili olduğu yönündeki bulguları, RSPO3'ün insanda akciğer ve karaciğer fibrozunda rol oynayabileceğine dair destekleyici kanıtlar sağlamıştır.

Coriati et al., (2020), tarafından yapılan bir çalışmada, (CHI3L1) YKL-40 proteininin KF'de hastalık şiddetinin bir biyobelirteç haline gelebileceğini ve sonucunda sağlık uzmanları tarafından ciddi hastalık durumu olan KF hastalarını tanımlamak için klinik bir araç olarak kullanılması önerilmektedir.

Karolak et al (2019), klinik ve histopatolojik olarak interstisyel neonatal akciğer bozuklukları tanısı almış, bunlardan 14 hastada asiner displazi, 2 hastada konjenital alveolar displazi ve 10 hastada diğer ölümcül akciğer hipoplazileri olan toplamda 26 ölen hastadan oluşan bir kohortu incelemişler. Bu çalışmada, 26 hastanın 16'sında (%61) TBX4 veya FGF10'u (602115) içeren nadir heterozigot kopya sayısı varyantları veya delesyonları tanımlanmıştır. Akciğer hipoplazisi olan bireyler, tahmin edilen akciğere özgü güçlendirici bölgede en az bir kodlamayan SNV barındırmıştır. Szafranski et al., (2016), tarafından yapılan başka bir çalışmada bir hastanın daha önce TBX4 geninde bir glu86-gln (E86Q; 601719.0005) mutasyonuna sahip olduğu gösterilmiştir. Bu değişken gnomAD veri tabanında bulunamamıştır. Başka bir

infantın (P022), TBX4'te aynı kodonda farklı bir deęişiklik (E86K) taşıdığı bildirilmiştir ancak hiçbir fonksiyonel çalışma bulunamamıştır.

İlk veri kaynağımıza ait çalışmamın başında 3 yaşındaki bir çocuktan N-HLF kullanılarak bazı deneyler yapılmış. Veri kaynağımızın sonuçları, genç fibroblastların, neredeyse KF yaşı bir hastayla karşılaştırılabilir, yüksek miktarda kolajen üretebildiğini göstermiştir. Maalesef, Lonza'dan genç donörlerden alınan hiçbir CF-HLF şisesi bulunmadığından bu çalışmaya devam edilememiştir. Normal genç hücrelerin, doku homeostazını ve fonksiyonunu koruyabilen ECM yeniden modelleme mekanizmalarının aktivasyonu yoluyla patolojik durumlara daha fizyolojik bir şekilde yanıt verebileceği varsayılmıştır (Cole et al., 2018). Ayrıca hastanın yaşı, farklı patolojik durumları ve stromanın farklı durumlarını yansıtabilir (örn. fibrotik veya fibrotik deęil). Bu hipotezi doğrulamak için, gelecekteki çalışmalarda, *in vitro* olarak daha geniş bir "yaş" bağ hava yolu dokuları aralığı elde etmek ve özelliklerini karşılık gelen birincil biyopsi ile karşılaştırmak amacıyla doğrudan insan biyopsilerinden fibroblastların çıkarılması planlanmıştır. En azından, CF-CAT'in gelecekte, benzersiz bir tam kalınlıkta KF modeli elde etmek ve KF'deki epitelyal stromal karışmanın sonuçlarını 3D güvenilir bir platformda araştırmak amacıyla epitel hücrelerinin tohumlanması için bir besleyici olarak kullanılabileceği savunulmuştur.

İkinci veri kaynağımızda KFTR lokusunda aynı kalıtsal mutasyonlara sahip KF hastaları, hastalık şiddeti ve akciğer fonksiyonunda önemli farklılıklar göstermektedir. KF'nin fenotipik heterojenliği büyük ölçüde deęiştirici genlerden etkilenmektedir (Gallati, 2014). KF'nin olası deęiştirici genlerini belirlemek için birkaç yaklaşım kullanılsa da, 2. veri kaynağımızdaki amaç için bir transkriptomik yaklaşım kullanılmıştır (Genin et al., 2008). Wright et al., (2006) yaptıkları bir çalışmada, mikroarray teknolojisini kullanan bir transkriptomik yaklaşımla burun solunum epitel hücrelerinde Orta ve Ağır KF hastaları arasındaki deęiştirici genleri belirlemeye çalışmışlardır. Bununla birlikte, daha önce de belirtildiği gibi, RNASeq kullanan global gen ekspresyon analizinin mikroarray'e göre birçok avantajı bulunmaktadır (Wang et al., 2014). Bu nedenle, 2. veri kaynağımıza ait çalışmada, F508del mutasyonu için homozigot olan KF hastalarında Orta ve Ağır akciğer fenotipinin deęiştirici gen(ler)ini belirlemek için öncelikle RNASeq teknolojisi kullanılmış (Kormann et al., 2017). 2. veri kaynağı çalışmamızda, öncelikle, bu hücre tiplerinde

olası farklılığı önlemek için periferik kan lökositlerinden elde edilen transkriptomik sonuçların solunum sisteminin doğal hücreleriyle (burun, trakeal ve bronş epitel hücreleri) karşılaştırılabilir olup olmadığı değerlendirilmiş. Veriler altı farklı çalışmayla karşılaştırılmış ve CXCL10, GOS2, PTGS2, IFIT1, IFIT3 ve ISG15 dahil olmak üzere genler için benzer ifade kalıpları gözlemlenmiş (Virella-Lowell et al., 2004; Zabner et al., 2005; Wright et al., 2006; Verhaeghe et al., 2007; Ogilvie et al., 2011; Clarke et al., 2013). Bu nedenle, up regüle olmuş gen profilleri, KF'deki kan lökositleri ve doğal solunum hücreleri arasında nispeten benzerdir. Tezimizde ilginç olarak PTGS2 ve IFIT1 genlerini down regüle olarak bulduk. Aynı şekilde, hem RNASeq hem de mikroarray platformları, up regüle edilmiş genlerle sonuçlanmıştır. 2. veri kaynağımıza ait verilerde, kan hücresi transkriptomunun KF'de hem üst hem de alt solunum yolu hücreleri için bir vekil olarak kullanılabileceği gösterilmiştir. Buna ek olarak, RNASeq verileri aynı örneklerle qPCR kullanılarak doğrulanmış ve yüksek bir tutarlılık skoru (%94) ile sonuçlanmıştır. Araştırılan iki grubun ekspresyon profillerindeki farklılıklar 88 DEGs ile sonuçlanmıştır. Bununla birlikte, KF'nin Orta ve Ağır pulmoner fenotipi arasında bu genler için çarpıcı farklılıklar da gözlemlenmiştir. 74 gen, Orta KF akciğer hastalığı olan hastalarda Ağır KF akciğer hastalığına kıyasla daha yüksek ekspresyon sergilemiştir. Toplam 14 gen, Orta KF olanlarla karşılaştırıldığında Ağır KF'li bireylerde önemli bir up regüle olmuştur. İlginç bir şekilde, zenginleştirme analizi, Orta KF grubunda tip I interferon yanıtında ve viral enfeksiyonlara karşı savunma yanıtında yer alan genlerin oldukça zengin olduğunu ortaya koymuştur. Önceki çalışmalar tip I interferon genlerinin (IFIT1, IFIT3 ve ISG15) KF ile ilişkisini bildirmiş olsa da (Clarke et al., 2013), 2. veri kaynağımıza ait çalışmada KF'yi modifiye etmede tip I IFN yolunun rolünü doğrulanmıştır. Bir sistem biyolojisi yaklaşımı kullanan yakın tarihli bir çalışmada, akciğer fonksiyonunu değiştiren ana KF değiştirici gen olarak bir tip I interferon geni IFI16 tanımlanmıştır (Trouve et al., 2017). Tip I IFN'ler, anti-viral yanıt için iyi bilinir ve viral enfeksiyonlar, KF hastalarında akciğer fonksiyonundaki düşüşe potansiyel katkıda bulunurlar (Frickmann et al., 2012). Bu nedenle, viral enfeksiyonların etkisinin gözlemlenen ekspresyon profiliyle dikkatlice değerlendirilmiştir. Bununla birlikte, çalışmaya katılan tüm katılımcılar, kan örneği alma sırasında soğuk algınlığı veya üst solunum yolu viral enfeksiyonları belirtileri göstermemiştir. Bu nedenle, viral enfeksiyonların Orta ve Ağır KF hastaları arasında gözlenen DEG'lere katkısının daha

az olası olduğu düşünülmüştür. Bununla birlikte, artan kanıtlar, Tip I IFN sinyali yolunun bakteri, parazit ve mantarlar dahil olmak üzere diğer patojenlere karşı konak savunmasında rol oynadığını göstermektedir (McNab et al., 2015). Örneğin, tip I interferon gen ekspresyonu ve polimorfizmleri, sistemik kandidiyaz duyarlılığı ile belirgin bir şekilde ilişkilendirilmiştir (Smeekens et al., 2013). Özellikle, daha önceki çalışmalar, tip I interferon sinyalleşmesinin normal akciğer epitel hücrelerinde *Pseudomonas aeruginosa* tarafından aktive edildiğini, ancak KF epitel hücrelerinde iptal edildiğini doğrulamıştır (Parker et al., 2012). *P. aeruginosa*, KF hastalarında pulmoner fonksiyondaki düşüşle doğrudan ilişkili iyi bilinen bir patojendir (Blanchard et al., 2017). Not olarak, her iki gruptaki çalışma hastalarında, gruplar arasındaki dağılımlarında önemli farklılıklar olmaksızın mantar ve bakteriyel enfeksiyonları mevcuttu. Bu nedenle, gelişmiş Tip I IFN sinyali nedeniyle Orta KF bireylerinin enfeksiyonlarını koruduğunu ve kontrolde akciğer fonksiyonunun azaldığını varsaymak mantıklı olmuştur. Bu hipotez, akciğer epitelindeki ve solunum yollarındaki tip I IFN yanıtının antibakteriyel etkisi ile daha da desteklenebilir (Parker et al., 2011). KF patogenezi değiştirilmediği interferonla ilgili başka bir gen olan IFRD1'in etkisi de iyi araştırılmıştır (Hector et al., 2013). Bu da, 2. veri kaynağımıza ait bulguları IFN'nin, özellikle de KF ile ilişkili akciğer fonksiyonunun değiştirilmesinde tip I IFN yanıtının katkısını desteklemiştir. Tip I IFN yanıtının up regüle edilmiş genleri arasında EGR1, orta KF hastalarında daha yüksek ekspresyon gösteren belirgin bir gen idi. Biz de tezimizde aynı şekilde orta KF örneklerinde EGR1 genini yüksek eksprese olduğunu gösterdik. Önceki bir transkriptomik çalışmada EGR1'in *P. aeruginosa* ile enfekte KF bronş epitel hücrelerinde down regüle edildiği belirlenmiştir (Balloy et al., 2015). Biz de tez çalışmamızda EGR1'in down regüle olduğunu bulduk. Ayrıca, *A. fumigatus* ve *Toxoplasma gondii* ile enfekte hücrelerde EGR1'in up regüle olduğu gözlenmiştir (Chen et al., 2015). Bu da enfeksiyonlar sırasında önemini göstermiştir. İlginç bir şekilde, diğer iki EGR ailesi üyesinin, EGR2 ve EGR3'ün de Orta KF grubunda up regüle düzenlendiğini bulunmuş. Biz tez çalışmamızda EGR2 ve EGR3 ile ilgili her hangi bir değişiklik bulamadık.

EGR'ler, DNA bağlayıcı çinko parmak proteinlerinin bir ailesidir ve transkripsiyonel düzenleyiciler olarak işlev görür. Ek olarak, EGR proteinlerinin Mx1, HERC5, RSAD ve diğerleri dâhil olmak üzere diğer tip I IFN genleriyle yakından etkileşime girdiği bulunmuştur. Bu nedenle, KF akciğer hastalığında potansiyel

değiştirici genler olarak EGR'lerin rolü, gelecekteki araştırmaları garanti etmektedir. Tip I IFN sinyalleşmesindeki diğer potansiyel değiştirici olan genler Mx1 ve IFIT2'nin *P. aeruginosa*'ya karşı baskın bir rol oynadıklarından ve ekspresyonları da qPCR ile doğrulanmıştır (Parker et al., 2011). MARA analizi, Orta KF hastalarında IRF1,2,8 ve STAT2 TF'nin daha yüksek transkripsiyonel aktivitesi ile sonuçlanmıştır. Bunlar Tip I IFN genlerinin daha yüksek ekspresyonundan sorumludur. Zenginleştirme analizi, Orta KF'li bireylerde yüksek oranda eksprese edilen üç ribozomal sap proteini RPL31, RPL34 ve RPS24'ü tanımlamıştır. Bu proteinlerin ER ve golgi taşınımına yönelik protein kaçakçılığına karıştığı gözlemlenmiştir. KF'de ribozomal sap proteinlerinin önemi iyi belgelenmiştir. Daha önce yapılan bir çalışmada Homozigot F508del mutasyonuna sahip birincil insan bronş epitel hücrelerinde, RPL12'nin susturulmasının KFTR iyon kanalı aktivitesi tarafından kurtarıldığı gösterilmiştir (Veit et al., 2016). Ek olarak, başka bir ribozomal sap proteini RPL27, klorüre bağımlı gen olarak rapor edilmiştir. Bu ekspresyon KFTR kanalının işleviyle ilgilidir (Valdivieso et al.,). F508del mutasyonu, protein kaçağı ve ER'de erken bozunma ile ilgili olduğundan, bu ribozomal sap proteinleri daha ileri araştırmalar için çekici bulunmaktadır.

KF akciğer ortamına nötrofiller ve yüksek proinflatuar kemokin IL-8 seviyeleri hakimdir (Jundi et al., 2015). Furlan et al., (2016) tarafından yapılan bir çalışmada, KF hastalarının FEV1 durumu ile IL-8 polimorfizmlerinin genetik bir ilişkisi sunulmuştur. İkinci veri kaynağımıza ait çalışmada, orta KF'li bireylerde IL-8 transkriptlerinin ekspresyonunun arttığı gözlemlenmiştir. Ancak, Ağır KF grubunda daha fazla IL-8 ekspresyonu bulunduğundan bu gözlem qPCR ile doğrulanamamıştır. Biz tez çalışmamızda IL-8 ile ilgili CXCL8 geninin orta KF'li bireylerde daha fazla eksprese olduğunu bulduk. Ayrıca bu gen çalışmamızda down regüle olarak bulundu. Yapılan bir çalışmada, Orta ve Ağır KF arasındaki nazal epitel hücrelerinde benzer bir sonuç gözlenmiş ve IL-8'in tahmin edilebilir şekilde hastalığın şiddeti ile ilişkili olmadığı savunulmuştur (Wright et al., 2006). Bununla birlikte, RNASeq ve qPCR arasında çelişkili bir sonuca dikkat çekildiği için, IL-8'in KF akciğer fonksiyonu değişkenliğinde rolünü araştırmak için bağımsız örneklerle daha fazla doğrulama gerekli olduğu ortaya atılmıştır. Yüksek örneklem sayısına sahip (n=6,365) GWAS meta-analiz çalışması, beş genomik bölgenin KF akciğer hastalığı şiddetiyle güçlü ilişkisini ortaya koymuştur (Corvol et al., 2015). 2. veri kaynağımıza ait çalışmada

gözlenen DEG'den herhangi birinin belirtilen genomik konumu paylaşp paylaşmadığı incelenmiştir. Dikkat çekici bir şekilde, kromozomal haritalamanın HCG27 genini 6p21.32'de rapor edilen HLA-DRA lokusuna çok yakın olan 6p21.33 pozisyonuna atadığı bulunmuştur. HCG27'nin bir pseudo-gen olduğu bildirilse de, Orta ve Ağır KF'li bireyler arasında ekspresyon seviyelerinde önemli bir fark gözlemlenmiştir. HCG27'yi içeren bölge aynı zamanda sedef hastalığı için önemli bir yatkınlık lokusudur (Nair et al., 2006). Bununla birlikte, RNASeq tabanlı bir transkriptomik yaklaşım kullanan 2. veri kaynağına ait çalışmada, bildirilen duyarlılık lokusları bulunabilmiş ve böylece daha büyük GWAS çalışması çoğaltabilmiştir. Daha önce açıklanan diğer genler, çalışmanın tanımlananları ile tam olarak eşleşme de, aynı biyolojik ailelere ait olduğu gösterilmiştir. Örneğin, KF'de hava yolu asitlenmesinden ATP12A'nın sorumlu olduğu bildirilmiş ve çalışmada aynı aileden bir gen olan ATP6V1G1'i tanımlanmıştır (Shah et al., 2016). Biz tez çalışmamızda bu geni tanımlayamadık. Genel olarak, 2. veri kaynağına ait çalışma, orta ve ağır KF akciğer hastalığı arasındaki karşılaştırmalı transkriptomik analizi KF pulmoner düşüşüne yeni bakış açıları sağlamıştır. Global gen ekspresyonu analizleri, tip I interferon tepkisi genlerinin ve ribozomal sap proteinlerinin olduğunu belirlemiş ve potansiyel KF modifiye edici genleri ortaya çıkarmıştır. Bu bulgular, KF tedavisi için bu gen listesinden yeni terapötik hedefler seçme potansiyeline sahip olduğunu göstermiştir.

İkinci veri kaynağımızla uyumlu olarak orta KF örneklerinde EGR1, RSAD2 ve FOS genlerinin yüksek ifade edildiğini bulduk. Bu veri kaynağımızda da SFRP1 genin de orta KF örneklerinde aşırı ekprese olduğu bulunmuş. Ancak biz tezimizde bu genle ilgili bir değişiklik saptayamadık. Ayrıca çalışmamızda Orta KF örneklerinde IFIT2, DUSP1 ve TRIM22 genlerinde yüksek ekspresyona uğramış olarak bulduk. Veri kaynağımızla uyumlu olarak Ağır KF örneklerinde C4BPA ve ZNF683 genleri yüksek ekspresyona sahipti. Veri kaynağımızda Ağır KF grubunda EPB41L4B ve LOC644172 genleri aşırı ekprese uğramış bulunurken biz bu genleri çalışmamızda saptayamadık. Veri kaynağımızdan farklı olarak hem orta hem de Ağır KF örneklerinde RESF1, TMEM123, SMCHD1, ENO1 genlerini aşırı ekprese olarak gösterdik.

Tez kapsamında kullandığımız pipeline ile 2. veri kaynağımızda sırasıyla CXCL8, EGR1, KY, PTGS2, NDUFA5, FOS, ARRDC3, CD69, IFIT1, IFIT5,

MALAT1, PMAIP1, TALAM1, TMEM126B, TPST1 ve ZNF708 padj'a göre ilk 20'de yer alan down regüle, ZNF683, CRT3, POU2F2 ise up regüle olan genlerdir.

ZNF683 geni sitogenetik lokasyonu 1p36.11'dir. Braun et al. (2015) yaptıkları çalışma ile HOBIT adını verdikleri ZNF683'ün BLIMP1 geninin genomik kopyalanmasından kaynaklandığını belirtmiştir. Yapılan çalışmalarda ZNF683 geninin akciğer NK hücrelerinde yüksek ekspresyona uğradığı gösterilmiştir (Marquardt et al., 2019; Brownlie et al., 2021). Veri kaynağımızda KF akciğer hastalığının ağır formu ile daha iyi ilişki sergilediğini savunmuştur. Bu çalışmalar sonucumuzun literatür ile uyumlu olduğunu göstermektedir.

Interlökin-8 (IL-8), CXCL8 geninde kodlanan bir nötrofil kemokindir. Normalde CXCL8 ekspresyonu, histon deasetilasyonu, promotöre oktamere-1 bağlanması ve nükleer faktör-B baskılayıcı faktörün (NRF) inhibe edici etkisi nedeniyle baskılanmaktadır. Çok sayıda nötrofile katkıda bulunan ana faktör, KF akciğerinde bulunan normal IL-8 düzeylerinden daha yüksek olmasıdır. Enfeksiyon ve inflamasyon, KF hava yolu hücrelerindeki içsel değişikliklerle birlikte anormal derecede yüksek intrapulmoner IL-8 seviyelerinden sorumludur. Anormal derecede yüksek CXCL8 ekspresyonunu inhibe etme stratejileri, KF akciğer hastalığı için terapötik potansiyele sahiptir (Jundi and Greene, 2015). Campbell et al., (2013) tarafından yapılan bir çalışmada hem MAPK hem de PI3K yollarının aynı anda hedeflendiği kombine bir yaklaşımının, malign hücrelerde sinyallemenin neden olduğu CXCL8 gen ekspresyonunu down regüle etmek için en başarılı yaklaşım olabileceği savunulmuştur. Yapılan başka bir çalışmada CXCL8'in sulindac'a yanıt olarak down regüle olduğu gösterilmiş ve kolorektal kanser hücre hattı modellerinde sulindac kaynaklı apoptozun bir aracısı olarak tanımlanırken, salisilat tedavisi makrofaj benzeri hücre hattı THP-1 CXCL8 ekspresyonunun down regülasyonu ile sonuçlanmıştır (Housby et al., 1999; Ji, 2009).

Prostaglandin endoperoksit sentaz-2 (PTGS2) veya siklooksijenaz-2 (COX2) geni, prostaglandin biyosentezi için anahtar enzimdir. Araşidonik asidi prostaglandin-E2'ye (PGE2) dönüştürür. Son yıllarda yapılan çalışmalardan elde edilen sonuçlar ile PTGS2 ve metabolitlerinin pulmoner fibroz üzerinde koruyucu bir etkiye aracılık ettiğini gösterilmiştir. Xu and Chen (2022) tarafından yapılan bir çalışmada

PTGS2'nin Sistemik Skleroz-İnterstisyel akciğer hastalığında önemli bir rol oynadığı gösterilmiştir.

Tez çalışmamızda da daha öncede belirtildiği gibi her replikat içindeki GC-içerik yanlılığını hesaba katmak için okuma sayısı normalleştirilmesi, tam transkriptomla eşlenen okuma sayıları üzerinde EDASeq'te gerçekleştirildi. EDASeq normalleştirilmiş değerleri daha sonra iki kopya arası normalleştirme için DESeq2 için kullanılacaktı ancak tüm normalleştirilmelerin DESeq2 ile yapılmasına karar verildi. Okuma sayıları, gen ekspresyonu değişikliklerinin yalnızca bir genin up veya down regülasyonu nedeniyle olup olmadığını veya nispi bolluktaki değişikliklerin ekspresyon seviyelerini etkileyip etkilemediğini araştırmak için veri setimize tam transkriptom kullanılarak normalleştirildi. DESeq2'nin, sayıların genin geometrik ortalamasına göre gen sayılarının medyan oranıyla hesaplanan numuneye özgü boyut faktörlerine bölüldüğü kendi dâhili normalleştirme yöntemi vardır. Bu yöntem özellikle numuneler arasındaki gen sayımı karşılaştırmaları, değişen dizileme derinlikleri için normalleştirme ve farklı ifade analizine izin verilmesi içindir. Oranların medyanı yöntemi, GC veya gen uzunluğu yanlılıklarını dikkate almaz. Batch efektleri dengeli deneysel tasarımlarda ortak değişkenleri hesaba katmak için diferansiyel gen ekspresyonu analizinden önce DESeqDataSet yapımı sırasında modellenabilir. Diferansiyel gen ekspresyon analizi daha sonra iki farklı koşul arasında önemli ölçüde farklı şekilde eksprese edilen genleri saptamak için kullanılır. DESeq2, sayım verilerini modellemek için negatif binom dağılımına sahip genelleştirilmiş bir doğrusal model kullanır. Diferansiyel ifade analizinin sonuçlarına, results() işlevi kullanılarak erişilir. Diferansiyel olarak eksprese edilen genler daha sonra ekstrakte edilebilir. Çalışmamızda, işlevi bilinen genler, ilk veri setimizde transkriptler için ayarlanmış p-değeri < 0.05 , lfcThreshold = 0.32 eşiği seçildi. Diğer veri setimizde ise p-değeri < 0.05 olarak dikkate alındı. Diferansiyel olarak eksprese edilen genler daha sonra denextend kullanılarak hiyerarşik olarak kümelendi, MaGE'den atanan fonksiyonla açıklama yapıldı ve pheatmap kullanılarak çizildi. Yapılan bir çalışmada Hem DESeq2 ile normalize edilmiş hem de GC ile normalize edilmiş veriler için diferansiyel gen ekspresyonu sonuçlarının karşılaştırılmış NAD'ye bağımlı format dehidrojenazın, GC ile normalize edilmiş verilerde oldukça farklı şekilde ifade edildiği bulunmuştur (Johnson et al., 2021). Akberdin et al. (2018) tarafından yapılan bir çalışmada metabolomik ve proteomik analizlerin bir

kombinasyonuna dayanarak aynı sonuca varılmış ancak DESeq2'nin medyan oran normalizasyonu ile overlook olmuştur.

Daha önce belirtildiği gibi RNASeq, transkriptomun karmaşıklığını anlamada etkilidir. Birkaç yöntem ile transkriptomda meydana gelen çeşitli yönleri anlamak için yaklaşımlar sağlansa da genomun fonksiyonel bölgelerindeki varyantları tanımlamaya yönelik uygulamasında çok az şey yapılmıştır. Bu amaçla, Adetunji et al. (2019) tarafından yapılan bir çalışmada, splice duyarlı RNASeq referans eşleme araçları, GATK kullanarak varyant tanımlama ve transkriptom dizilemesinden genomik varyantların doğru bir şekilde tanımlanmasına izin veren sonraki filtrelemenin bir kombinasyonunu kullanan çoklu hizalayıcı stratejisi olan VAP pipeline tasarlanmıştır. Bu çalışmaya ait sonuçlar, transkript bölgelerde meydana gelen SNP'lerle sınırlı olsa da, çok yüksek hassasiyet, duyarlılık ve özgüllük göstermiştir. RNASeq okumalarının haritalama aşamasının dikkate alınması, varyant çağırma için çok önemli bir adımdır. Bu nedenle Adetunji et al., (2019) yanlış pozitiflerin yaygınlığını azaltmak için üç RNASeq splice-aware aligner kullanarak bir referans haritalama stratejisi tasarlamıştır. Splice-aware hizalayıcının kullanımı, okuma eşlemesi için hem genom hem de transkriptom bilgisini aynı anda kullandığından, okumaların doğru bir şekilde birleştirilmesine olanak tanımaktadır. RNASeq'ten varyantları çağırma için çok sayıda uygulama bulunmaktadır. Genom sekanslama ile tespit edilen varyantların doğrulanmasını sağlamaktadır. Ayrıca gen regülasyonu için potansiyel transkripsiyon sonrası modifikasyonları ortaya çıkarır. İşlevsel olarak önemli olabilecek ancak DNA veya ekzom dizileme kullanılarak yakalanması zor olabilecek daha önce tanımlanmamış varyantların daha düşük maliyetle saptanmasına olanak tanır. Aynı çalışmada TGS verileri, RNASeq verilerine yol açan aynı örneklerden dizilenmemiş olsa bile, bu durum veri kümelerindeki zayıf örtüşmeyi açıklayabilmektedir. Örneğin, Ekzonlardaki RNASeq varyantlarının %87.5'i, dbSNP'de iyi karakterize edilmiş olmasına rağmen TGS'de bulunamamıştır. Buna rağmen, RNA varyantları, ilgilenilen özelliklerin genetik haritalanması için genetik belirteçlerin tanımlanmasında kullanılabilir. Böylece genotip ve fenotip arasındaki ilişkinin daha iyi anlaşılması sağlanır.

RNA SNP'leri yalnızca ifade edilen transkriptlerde saptanır. Kapsamlı coverage'dan bağımsız olarak, genomun bazı bölümlerinde varyant tespiti, potansiyel

ekspresyon eksikliği nedeniyle RNASeq tarafından garanti edilemez. Bu tez çalışmamızda 1. veri kaynağımızda, belirlediğimiz pipeline ile yaptığımız varyant çağırma sonrasında, KF ile ilgili örnekler RNASeq ile elde edilen datalardan anlamlı bir sonuç bulunamamıştır. Bu ilk veri kaynağına ait örneklerden CF.HLF_2D_Rep1, CF.HLF_2D_Rep2, CF.HLF_2D_Rep3, CF.HLF_2D_Rep4, CF.HLF_3D_Rep1, CF.HLF_3D_Rep2, CF.HLF_3D_Rep3 KF Lung Fibroblasts hücre tipleridir. Daha öncede belirtildiği gibi hücre hattının satın alındığı şirkette KF donörünün KFTR'sinin özellikli mutasyonu hakkında bir bilgi sağlanmamıştır. Normal örneklerde saptanan patojenik değişiklikler tezimizin konusunu oluşturmadığı için değerlendirmeye alınmamıştır. CF.HLF_2D_Rep3 de bulunan KMT2A (Lysine-Specific Methyltransferase 2A) (ENSG00000118058) 11q23.3'da bulunmaktadır. Yapılan çalışmada KMT2A (MLL) genindeki mutasyonların Wiedemann-Steiner sendromuna (WSS) (MIM#:605130) neden olduğu gösterilmiştir. WSS klinik özellikleri arasından Solunum yolu enfeksiyonları, kısa boy, hipertelorizm, gelişim geriliği, zihinsel yetersizlik yer almaktadır.

CF.HLF_3D_Rep1 örneğinde GATK ile rs771203308 bulunmuştur. 8q13.1-q13.2 bulunmaktadır ve buradaki gen CSPP1 (Centrosome Spindle Pole-Associated Protein 1)'de meydana gelen mutasyon Classical Joubert Syndrome'e neden olmaktadır (PMID: 24360807). Aynı örnekte Samtools ile rs80358198 bulunmuştur. İlgili gen ise FUCA1'dir. Bu gen tarafından kodlanan protein, fukoz içeren glikoproteinlerin ve glikolipidlerin parçalanmasında rol oynayan bir lizozomal enzimdir. Bu gendeki mutasyonlar, otozomal resesif geçişli bir lizozomal depo hastalığı olan fukozidoz (FUCA1D) ile ilişkilidir. Bu lokusun bir psödogeni 2. kromozomda mevcuttur (<https://www.ncbi.nlm.nih.gov/gene/2517>, Erişim tarihi: 03 Şubat 2022).

CF.HLF_2D_Rep2 ve CF.HLF_3D_Rep2 örneklerinde GATK ile bulunan rs764261443 ile ilgili veri tabanlarında ilgili bilgiler bulunamamıştır.

CF.HLF_3D_Rep3'de GATK ile rs777766787 anlamlı bir sonuç bulunamamıştır. Yine aynı örnekte Polikistik karaciğer hastalığı (PLD) sonucu elde edilmiştir. PLD, izole PLD'de bağımsız olarak veya OD polikistik böbrek hastalığı ve komplike mekanizmalara sahip OR polikistik böbrek hastalığının eşlik eden bir semptomu olarak bulunan nadir bir kalıtsal hastalıktır. PLD şu anda birleşik bir tanı

standardından yoksundur. PLD tanısı genellikle karaciğer kistlerinin sayısı 20'den fazla olduğunda konmaktadır (Zhang et al., 2020).

İkinci veri kaynağımızdaki örnekler ise orta ve ağır KF olarak ayrılmaktadır. Kullandığımız pipeline ile SCF_01, SCF_02 ve MCF_15 örneklerinde GATK ile rs137854889 saptanmıştır. İlişkili gen ise ZMPSTE24'dir. Kromozomal konumu 1p34.2, moleküler olarak 40,258,078-40,294,180 baz çiftleri arasında bulunmaktadır. ZMPSTE24 geni prelamin A'nın post-translasyonel modifikasyonunda rol alır. ZMPSTE24 genindeki bozukluklara sekonder laminopatiler denir. MADB (Mandibuloacral dysplasia with type B lipodystrophy) nadir görülen OR bir hastalıktır. Doğum sonrası büyüme geriliği, mandibular hipoplazi gibi kraniyofasiyal anomaliler, terminal falanks ve klavikülaların progresif osteolizi gibi iskelet anomalileri ve benekli hiperpigmentasyon ve atrofi gibi cilt değişiklikleri ile karakterizedir (<https://www.omim.org>, 01 Şubat 2022).

SCF_01, MCF_01, MCF_02, MCF_13 örneklerinde Samtools ile rs368698783 saptanmıştır. İlgili gen HBG1'dir. Kromozomal konumu 11p15.4'dir. Genomik koordinatları (GRCh38): 11:5,248,268-5,249,856 (NCBI). Fetal hemoglobin quantitative trait locus 1 fenotipi ile ilişkilidir.

SCF_02, SCF_03, SCF_07 ve MCF_09'de hem GATK'de hem de Samtools'da rs1566734 saptanmıştır. İlgili gen PTPRJ'dir. Kromozomal konumu 11p11.2'dir. Genomik koordinatları (GRCh38): 11:47,980,503-48,170,838 (NCBI). Kolon Kanseri fenotipi ile ilişkilidir.

MCF_05, SCF_11 örneklerinde GATK ile rs746973259 saptanmıştır. İlgili gen SETX'dir. Kromozomal konumu 9q34.13, Genomik koordinatları (GRCh38): 9:132326918-132326919 (NCBI).

SCF_07, SCF_08, MCF_15 örneklerinde rs312262845 saptanmıştır. İlgili gen OFD1'dir. Kromozomal konumu Xp22.3-p21.3'dir. Genomik koordinatları (GRCh38): X: 13746827-13746835 (NCBI).

SCF_02, MCF_04, SCF_09, SCF_12, MCF_14, MCF_16 hem GATK'de hem de samtools'da rs17261572 saptanmıştır. İlgili gen C1GALT1C1'dir. Kromozomal

konumu Xp22.3-p21.3'dir. Genomik koordinatları (GRCh38):X:120,625,673-120,630,053 (NCBI). İlgili fenotipi Tn polyagglutination sendromudur.

MCF_05, SCF_10, MCF_13 örneklerinde GATK ile rs864321665 saptanmıştır. İlgili gen BTK'dir. Kromozomal konumu Xq22.1, Genomik koordinatları (GRCh38):X: 101,349,449-101,390,795 (NCBI). İlgili fenotip Agammaglobulinemia, X-linked 1 ve Isolated growth hormone deficiency, type III, with agammaglobulinemia. Bu tez çalışmamızda 2. veri kaynağımızda, belirlediğimiz pipeline ile yaptığımız varyant çağırma sonrasında, Kistik fibrozis ile ilgili örneklerde RNASeq ile elde edilen datalardan anlamlı bir sonuç bulunamamıştır. Ayrıca doğrulamak amacıyla örneklerle IGV ile de analiz edilmiş ancak ilgili bir sonuç bulunamamıştır.

Bu bulgular bize, Alel-spesifik gen ekspresyonu veya doku-spesifik gen ekspresyonu, varyantı taşıyan alelin eksprese edilmeyebileceği veya toplanan dokuların ilgili genleri eksprese etmeyebileceği göz önüne alındığında, genomik varyantların keşfini engelleyebileceğini de göstermektedir. RNASeq verilerinden genomik varyantları çağırmanın sınırlamalarına rağmen, daha önce yapılan birçok çalışma ile RNASeq verilerinden SNP calling'in yüksek hassasiyet ve özgüllük gösterdiği ortaya koyulmuştur.

Piskol et al., (2013) tarafından yapılan başka bir çalışma ile splice-aware bir RNASeq read-mapping prosedürü kombinasyonu yoluyla transkriptom dizilemesinden genomik varyantların doğru tanımlanması sağlayan SNPiR'yi tasarlanmıştır. Her iki veri setinde de SNPiR ile yanlış pozitif çağrılar kaldırılarak genomik varyantlar yüksek hassasiyetle tespit edebilmiş. RNASeq verilerinin kullanılması ile işlevsel olarak önemli bölgelerdeki varyantlar için zenginleştirme ve eksprese edilmiş ekzonik bölgelerde varyant çağrısında yüksek hassasiyet elde edilmesine izin vermiştir. Bu yüksek kesinlik ve hassasiyet, düşük kapsamlı sekanslama verileri için de korunmuştur.

TES önceden tanımlanmış ilgi bölgeleri yani ağırlıklı olarak genomun protein kodlayan kısımlarını hedeflemek için özel olarak tasarlanmıştır. Bu nedenle, TES bu bölgelerdeki varyasyonları yüksek hassasiyetle tanımlayabilirken, genomun aynı bölümlerinde kapsamlı kapsama ve varyant tespiti, potansiyel ifade eksikliği nedeniyle

RNASeq tarafından garanti edilemez. Ek olarak, fenotiple ilgili dokuların toplanmasının zor olabileceği ve kolay erişilebilir dokuların hastalıkla ilgili varyantlara sahip genleri eksprese edemeyebileceği göz önüne alındığında, dokuya özgü gen ekspresyonu genomik varyantların keşfini engelleyebilmektedir. Ayrıca, nonsense varyantlar da, nonsense aracılı bozunmanın bir sonucu olarak gözden kaçabilmektedirler.

Bu çalışmalar bize RNASeq'ten SNP çağrısının, TGS/TES yaklaşımlarının yerini almayacağını, sadece her iki yaklaşıma da uygun bir alternatif sunarak TGS/TES'den tespit edilen SNP'leri tamamlayacağını veya doğrulamak için kullanılabilirliğini göstermektedir.

YND teknolojilerindeki gelişmeler, daha önce mümkün olmayan bir ölçekte ve çözünürlükte genomik dizi varyasyonunu karakterize etme yeteneklerini geliştirmiştir. Bu da genetik varyasyonun insan hastalığına nasıl katkıda bulunduğunu incelemek için yeni yollar açmıştır. Önümüze çıkan büyük zorluk, yeni teknolojiler tarafından üretilen bol miktarda verinin, down-stream analizler için yüksek kaliteli veriler elde etmek üzere nasıl işleneceğidir. Bu amaçla çeşitli hesaplama araçları geliştirilmiştir. Biz tez çalışmamızda aralarındaki farkı gösterebilmek adına hem GATK hem de SAMtools araçlarını kullanıp elde ettiğimiz sonuçları analiz ettik. Genel olarak baktığımızda GATK ile daha fazla çağrı yapıldığını bulduk. Örneğin N.HLF_2D_Rep1 örneğinde Samtools ile 1729 varyant elde edilirken GATK ile 5685 varyant edilmiştir. CF.HLF_3D_Rep3 örneğinde 9747 varyant Samtools ile 34970 varyant ise GATK ile elde edilmiştir. 2014 yılında Pirooznia et al. (2014) tarafından yapılan bir çalışmada, GATK'nın çok yüksek kaliteli değişken çağrı verileri sağladığını bulunmuştur. Yapılan diğer çalışmalarda (Li et al., 2009; DePristo et al., 2011), varsayılan insersiyon/delesyon (indels) etrafında eşlenen dizi okumalarının yeniden hizalanmasının ve varyant çağrısından önce temel kalite puanlarının yeniden kalibre edilmesinin bu performans için çok önemli olduğu gözlemlenmiştir. Yeniden hizalama/yeniden kalibrasyonlu ve kalibrasyonsuz GATK ve Samtools arasındaki bu karşılaştırmalar için, SNP genotip verisine sahip olmadığımız için, Sanger sekansı gibi yöntemlerle validasyon yapılamadı. Bu nedenle yanlış-negatif ve doğru-negatif çağrılar hakkında doğrulama yapılamadı. Pucker et al. (2019) tarafından yapılan bir çalışmada, varyant çağrı setlerinin kalitesi, daha önce bir kalite göstergesi olarak

tanımlanan transisyon/transversiyon oranı kullanılarak genel olarak, analiz edilen hemen hemen tüm çağrı setlerinin kalitesinin benzer olduğunu göstermiştir. Liu et al., (2013) tarafından yapılan bir karşılaştırma çalışmasında da Samtools ve GATK ile tüm pipeline için benzer transisyon/transversiyon oranları bildirilmiştir. Biz tez çalışmamızda transisyon/transversiyon oranı kullanarak bir karşılaştırma yapmadık. Bu karşılaştırma çalışmaları YND veri kümelerine uygulandığında farklı varyant çağrı hatlarının güçlü ve zayıf yönlerine ilişkin içgörüler sağlamaktadırlar. Değerlendirilen tüm araçların performansı, okunan veri kümelerinin özelliklerine ve genom dizisine bağlı olarak numuneler arasında farklılık gösterse de elde edilen bulgular yardımcı birer rehber niteliğindedir. Sonuç olarak, yapılan çağrıların doğrulanmasından elde edilen mevcut sonuçlara dayanarak, hizalama ve kalibrasyon ile hem Samtools hem GATK ile elde edilen sonuçlar değerlendirilmiştir.

RNASeq okumaları sıklıkla genomdaki birden çok yere veya birden fazla transkripti kodlayan genomik bölgelere hizalanır (Hita ve ark., 2022). Gen ekspresyonunun ölçülmesi veya *de novo* transkriptlerin birleştirilmesi RNASeq ile yaygın uygulamalar iken SNV'ler ve kısa indellerin tespiti de mümkündür. RNASeq verilerinden varyantları çağırmak daha zahmetli olduğundan, en popüler varyant çağrı ardışık düzenleri DNA için tasarlanmıştır. Zorluklardan biri, yanlış aramalara neden olan bağlantı yerlerinin yakınındaki yanlış hizalamalardır. Ek olarak, eklemeye duyarlı hizalayıcıların tümü indel tespiti için uygun değildir ve yanlış hizalama aracının kullanılması bunların gözden kaçırılmasına neden olabilir. TGS ile karşılaştırıldığında, RNASeq yalnızca yeterli okuma derinliğine sahip ifade edilmiş varyantları kurtaracaktır, bu da promotörler, enhansırlar, intronlar veya diğer düşük düzeyde ifade edilen bölgelerdeki varyantların tespit edilmesinin zor olacağı anlamına gelir. Bunlara ek olarak, diploid DNA dizisinde bulunan germline heterozigot varyantlarının aksine, her alelden eşit miktarlarda beklediğimiz yerde, RNA düzeyinde oran, alel spesifik ekspresyon nedeniyle 0,5'ten sapabilir. Daha fazlası, RNA-editing site, sekanslama veya haritalama hatalarıyla kolayca karıştırılabilir. Bu nedenle doğru DNA-RNA karşılaştırmaları yapılmalıdır. Mutasyona uğramış genler ifade edilmezse aynı kişiden kan RNASeq örneğine sahip olmak yardımcı olmaz.

Artık günümüzde, short-read RNASeq teknolojilerinin bazı sınırlamaları, hızla gelişen ultra uzun okuma dizileme teknolojileri yani PacBio ve Oxford Nanopore

Technologies tarafından ele alınmaya başlandı. Bu en son teknolojilerle, tam uzunluktaki transkriptomlar tüm türler için gerçek olabilmektedir. Özellikle, Nanopore tabanlı RNA dizileme teknolojileri, tüm bir RNA molekülünün bir protein nanopore yoluyla translokasyonu sırasında üretilen mevcut değişikliği gerçek zamanlı olarak ölçer ve tek bazlı çözünürlükte tanımlanmasına olanak tanırırlar. Doğrudan RNASeq'in bir avantajı, cDNA dönüştürme protokollerinin aksine, amplifikasyon bias ve kimerik artefaktları önlemesidir. Bu özelliğe ek olarak daha da önemlisi, RNA molekülünde mevcut epitranskriptomik değişiklikleri tespit etme kabiliyetine sahip olmasıdır. Tam transkriptleri kapsayan okumalarla, tekrarlayan bölgeler ve paylaşılan ekzonlar içeren transkriptleri tespit etmek ve doğru bir şekilde ölçmek kolay olacaktır. Örneğin, 4700'den fazla alel içeren polimorfik HLA-B lokusu Nanopore teknolojisi ile kapsamlı bir şekilde araştırılmıştır (Liu et al., 2017). Ancak diğer taraftan da, Nanopore sekansının hala bazı iyileştirmelere ihtiyacı vardır. Sıralama hata oranları, geleneksel Illumina yüksek verimli sıralamadan ($<0,1$) daha yüksektir (%5–%10) ve transkript ifadesinin doğru kantifikasyonu hala daha yüksek sekans çıktıları gerektirir. Buna rağmen örneğin MinION gibi cep boyutu cihazı onu laboratuvar odaları ile sınırlı olmayan taşınabilir bir teknoloji yapmaktadır. Aslında, Kuzey Kutbu (Edwards et al., 2016), Uluslararası Uzay İstasyonu (Castro-Wallace et al., 2017) ve Tanzanya'nın yağmur ormanları (Menegon et al., 2017) gibi uzak yerlerde gerçek zamanlı olarak çalıştığı test edilmiştir.

6. Öneriler

Genel olarak bakıldığında YND teknolojilerinden biri olan ve RNASeq olarak adlandırılan yöntem ile ilgili Biyoinformatik analizler sonrasında transkriptom ile ilgili bilgiler kolaylıkla ulaşılabilir olmuştur. RNASeq ile elde edilen verilerinden aynı zamanda genomik varyantların analiz de mümkün olduğundan bu yöntem oldukça değerlidir.

RNASeq ile varyant saptama etkinliğini vurgulamak için çalışmamızda Biyoinformatik araçları karşılaştırdık. Bu nedenle hem Samtools hem de GATK kullandık. Sonuç olarak GATK ile daha fazla varyant elde edebildik. Bu da çalışmamızın literatür ile uyumlu olduğunu göstermektedir. Çünkü yapılan çoğu çalışmada SNP tespiti için GATK'nin altın standart olduğu gösterilmiştir. Biz yine de gelecek çalışmalarda RNASeq çalışmasının varyant çağırma üzerindeki etkinliğinin tam olarak anlaşılabilmesi için aynı bireylerin aynı dokuları üzerinde hem RNASeq hem de DNaseq verisine sahip olup bunların sonuçlarının da karşılaştırılmasının önemli olduğunu vurgulamak isteriz.

Çalışmamız sonucunda elde edilen up regüle ve down regüle olan genler için ileride gen editleme yapılarak kesin mekanizma fonksiyonel deneylerle doğrulanabilir.

Bu tez çalışması, RNASeq'in KF örneklerindeki uygulamaları ve avantajları hakkında sınırlı, mevcut geçmiş çalışmaları ve başarıları kapsamaktadır. Çalışmanın okuyucularının toplanan bilgilerden yararlanacağı ve genetik olarak çeşitlendirilmiş insan hastalıklarının daha iyi anlaşılmasında RNA dizilemesinin gelecekteki uygulamalarına ışık tutacağı umut edilmektedir.

Çalışmamız KF örneklerinin RNASeq ile elde edilen verilerden varyant çağırma konusunda bölümümüzde yapılan ilk çalışmadır. Aynı zamanda ülkemizde de öncü çalışmalar arasında yer almaktadır. Tez yazımı ve çalışma boyunca hem RNASeq hem de Bionformatik çalışmalarında kullanılan yabancı terimlerin Türkçe karşılıklarında ortak bir terim bulunmadığından ifade karmaşıklığı yaşanmaktadır. Bu

nedenle bu tez çalışmamızın gelecekte yapılacak tez, araştırma ve rutin laboratuvar çalışmalarına yol göstermesi bakımından önemli bir kaynak olacağını düşünmekteyiz.



7. Kaynaklar

Aavikko, M., Kaasinen, E., Andersson, N., Pentimikko, N., Sulo, P., Donner, I., ... & Aaltonen, L. A. (2021). WNT2 activation through proximal germline deletion predisposes to small intestinal neuroendocrine tumors and intestinal adenocarcinomas. *Human molecular genetics*, 30(24), 2429-2440.

Adams D.R., C.M. Eng, (2018), Next-generation sequencing to diagnose suspected genetic disorders, *N. Engl. J. Med.* 379 1353–1362.

Adams, M. D., Kelley, J. M., Gocayne, J. D., Dubnick, M., Polymeropoulos, M. H., Xiao, H., ... & Venter, J. C. (1991). Complementary DNA sequencing: expressed sequence tags and human genome project. *Science*, 252(5013), 1651-1656.

Adetunji, M. O., Lamont, S. J., Abasht, B., & Schmidt, C. J. (2019). Variant analysis pipeline for accurate detection of genomic variants from transcriptome sequencing data. *PloS one*, 14(9), e0216838.

Ahner, A., Gong, X., Schmidt, B. Z., Peters, K. W., Rabeh, W. M., Thibodeau, P. H., ... & Frizzell, R. A. (2013). Small heat shock proteins target mutant cystic fibrosis transmembrane conductance regulator for degradation via a small ubiquitin-like modifier-dependent pathway. *Molecular biology of the cell*, 24(2), 74-84.

Ahner, A., Nakatsukasa, K., Zhang, H., Frizzell, R. A., & Brodsky, J. L. (2007). Small heat-shock proteins select $\Delta F508$ -CFTR for endoplasmic reticulum-associated degradation. *Molecular biology of the cell*, 18(3), 806-814.

Akberdin, I. R., Collins, D. A., Hamilton, R., Oshchepkov, D. Y., Shukla, A. K., Nicora, C. D., ... & Kalyuzhnaya, M. G. (2018). Rare earth elements alter redox balance in *Methylobacterium alcaliphilum* 20ZR. *Frontiers in microbiology*, 2735.

Alberti, S., Bohse, K., Arndt, V., Schmitz, A., & Hohfeld, J. (2004). The cochaperone HspBP1 inhibits the CHIP ubiquitin ligase and stimulates the maturation

of the cystic fibrosis transmembrane conductance regulator. *Molecular biology of the cell*, 15(9), 4003-4010.

Alibakhshi, R., Mohammadi, A., Khamooshian, S., Kazeminia, M., & Moradi, K. (2021). CFTR gene mutation spectrum among 735 Iranian patients with cystic fibrosis: A comprehensive systematic review. *Pediatric pulmonology*, 56(12), 3644-3656.

Alser, M., Rotman, J., Deshpande, D., Taraszka, K., Shi, H., Baykal, P. I., ... & Mangul, S. (2021). Technology dictates algorithms: recent developments in read alignment. *Genome biology*, 22(1), 1-34.

Amatangelo, M., Garipov, A., Li, H., Conejo-Garcia, J. R., Speicher, D., & Zhang, R. (2013). Three-dimensional culture sensitizes epithelial ovarian cancer cells to EZH2 methyltransferase inhibition. *Cell Cycle*, 12(13), 2113-2119.

Ameur, A., Zaghlool, A., Halvardson, J., Wetterbom, A., Gyllenstein, U., Cavelier, L., & Feuk, L. (2011). Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. *Nature structural & molecular biology*, 18(12), 1435-1440.

Anders, S., Pyl, P. T., & Huber, W. (20). February 2014. *HTSeq—a Python framework to work with high-throughput sequencing data*. *Bioinformatics* <http://dx.doi.org/10.1093/bioinformatics/btu638>.

Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *bioinformatics*, 31(2), 166-169.

Anderson, P. J., Lynch, T. J., & Engelhardt, J. F. (2017). Multipotent myoepithelial progenitor cells are born early during airway submucosal gland development. *American journal of respiratory cell and molecular biology*, 56(6), 716-726.

Angelis, A., Tordrup, D., & Kanavos, P. (2015). Socio-economic burden of rare diseases: a systematic review of cost of illness evidence. *Health Policy*, 119(7), 964-979.

Attar, N. (2013). Raymond Gosling: the man who crystallized genes. *Genome biology*, 14(4), 1-12.

Bagdany, M., Veit, G., Fukuda, R., Avramescu, R. G., Okiyoneda, T., Baaklini, I., ... & Lukacs, G. L. (2017). Chaperones rescue the energetic landscape of mutant CFTR at single molecule and in cell. *Nature communications*, 8(1), 1-15.

Bakr, A., Pak, O., Taye, A., Hamada, F., Hemeida, R., Janssen, W., ... & Sommer, N. (2013). Effects of Dimethylarginine Dimethylaminohydrolase-1 Overexpression on the Response of the Pulmonary Vasculature to Hypoxia. *American Journal of Respiratory Cell and Molecular Biology*, 49(3), 491-500.

Balloy, V., Varet, H., Dillies, M. A., Proux, C., Jagla, B., Coppée, J. Y., ... & Guillot, L. (2015). Normal and cystic fibrosis human bronchial epithelial cells infected with *Pseudomonas aeruginosa* exhibit distinct gene activation patterns. *PLoS One*, 10(10), e0140979.

Barasch, J., Kiss, B., Prince, A., Saiman, L., Gruenert, D., & Ai-Awqati, Q. (1991). Defective acidification of intracellular organelles in cystic fibrosis. *Nature*, 352(6330), 70-73.

Bardin, E., Pastor, A., Semeraro, M., Golec, A., Hayes, K., Chevalier, B., ... & Sermet-Gaudelus, I. (2021). Modulators of CFTR. Updates on clinical development and future directions. *European journal of medicinal chemistry*, 213, 113195.

Beaulieu, C. L., Majewski, J., Schwartzentruber, J., Samuels, M. E., Fernandez, B. A., Bernier, F. P., ... & FORGE Canada Consortium. (2014). FORGE Canada Consortium: outcomes of a 2-year national rare-disease gene-discovery project. *The American Journal of Human Genetics*, 94(6), 809-817.

Beck, T. F., & Mullikin, J. C. Program, NCS, & Biesecker, LG (2016). Systematic evaluation of sanger validation of next-generation sequencing variants. *Clinical Chemistry*, 62(4), 647-654.

Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., ... & Roe, P. M. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *nature*, 456(7218), 53-59.

Berge K.V., Hembach K.M., Sonesson C., Tiberi S., Clement L., Love M.I., Rob Patro, and Robinson M.D.,. Annual Review of Biomedical Data Science RNA Sequencing Data: Hitchhiker's Guide to Expression Analysis. *Annu. Rev. Biomed. Data Sci.* 2019. 2:139–73

Bergeron C., Cantin A.M., N Van den Berge, K., Hembach, K. M., Sonesson, C., Tiberi, S., Clement, L., Love, M. I., ... & Robinson, M. D. (2019). RNA sequencing data: Hitchhiker's guide to expression analysis. *Annual Review of Biomedical Data Science*, 2, 139-173.

Bergeron, C., & Cantin, A. M. (2021). New therapies to correct the cystic fibrosis basic defect. *International Journal of Molecular Sciences*, 22(12), 6193.

Berkers, G., van Mourik, P., Vonk, A. M., Kruisselbrink, E., Dekkers, J. F., de Winter-de Groot, K. M., ... & van der Ent, C. K. (2019). Rectal organoids enable personalized treatment of cystic fibrosis. *Cell reports*, 26(7), 1701-1708.

Blackledge, N. P., Ott, C. J., Gillen, A. E., & Harris, A. (2009). An insulator element 3' to the CFTR gene binds CTCF and reveals an active chromatin hub in primary cells. *Nucleic acids research*, 37(4), 1086-1094.

Blanchard, A. C., Horton, E., Stanojevic, S., Taylor, L., Waters, V., & Ratjen, F. (2017). Effectiveness of a stepwise *Pseudomonas aeruginosa* eradication protocol in children with cystic fibrosis. *Journal of Cystic Fibrosis*, 16(3), 395-400.

Boucher, R. C., Stutts, M. J., Knowles, M. R., Cantley, L., & Gatzky, J. T. (1986). Na⁺ transport in cystic fibrosis respiratory epithelia. Abnormal basal rate and response to adenylate cyclase activation. *The Journal of clinical investigation*, 78(5), 1245-1252.

Bournazos, A., Clarke, N., Cooper, S., Ghaoui, R., Lek, M., Oates, E., ... & Estrella, E. (2017). Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. American Association for the Advancement of Science.

Boycott, K. M., Vanstone, M. R., Bulman, D. E., & MacKenzie, A. E. (2013). Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nature Reviews Genetics*, 14(10), 681-691.

Bradbury NA, Jilling T, Berta G, Sorscher EJ, Bridges RJ, Kirk KL (1992) Regulation of plasma membrane recycling by CFTR. *Science* 256(5056):530–532

Brancato, V., Comunanza, V., Imparato, G., Corà, D., Urciuolo, F., Noghero, A., ... & Netti, P. A. (2017). Bioengineered tumoral microtissues recapitulate desmoplastic reaction of pancreatic cancer. *Acta Biomaterialia*, 49, 152-166.

Braun, J., Frentsch, M., Thiel, A. (2015). Hobit and human effector T-cell differentiation: the beginning of a long journey. *Europ. J. Immun.* 45: 2762-2765

Breunig, C. T., Durovic, T., Neuner, A. M., Baumann, V., Wiesbeck, M. F., Köferle, A., ... & Stricker, S. H. (2018). One step generation of customizable gRNA vectors for multiplex CRISPR approaches through string assembly gRNA cloning (STAgR). *PloS one*, 13(4), e0196015.

Broackes-Carter, F. C., Mouchel, N., Gill, D., Hyde, S., Bassett, J., & Harris, A. (2002). Temporal regulation of CFTR expression during ovine lung development: implications for CF gene therapy. *Human molecular genetics*, 11(2), 125-131.

Brodsky, J. L. (2007). The protective and destructive roles played by molecular chaperones during ERAD (endoplasmic-reticulum-associated degradation). *Biochemical Journal*, 404(3), 353-363.

Brownlie D, Scharenberg M, Mold JE, Hård J, Kekäläinen E, Buggert M, Nguyen S, Wilson JN, Al-Ameri M, Ljunggren HG, Marquardt N, Michaëlsson J. (2021). Expansions of adaptive-like NK cells with a tissue-resident phenotype in human lung and blood. *Proc Natl Acad Sci U S A*. 6;118(11):e2016580118. doi: 10.1073/pnas.2016580118.

Bruno, R., Fontanini, G., (2020). Next generation sequencing for gene fusion analysis in lung Cancer: a literature review. *Diagnostics (Basel)* 10, E521.

Bryant, S., & Manning, D. L. (1998). Isolation of messenger RNA. In *RNA Isolation and Characterization Protocols* (pp. 61-64). Humana Press.

Bullard, J. H., Purdom, E., Hansen, K. D., & Dudoit, S. (2010). Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC bioinformatics*, 11(1), 1-13.

Burns, M. B., Lackey, L., Carpenter, M. A., Rathore, A., Land, A. M., Leonard, B., ... & Harris, R. S. (2013). Erratum: APOBEC3B is an enzymatic source of mutation in breast cancer (Nature (2013) 494 (366-370). *Nature*, 502(7472).

Byron, S. A., Van Keuren-Jensen, K. R., Engelthaler, D. M., Carpten, J. D., & Craig, D. W. (2016). Translating RNA sequencing into clinical diagnostics: opportunities and challenges. *Nature Reviews Genetics*, 17(5), 257-271.

Cai, Z., Sohma, Y., Bompadre, S. G., Sheppard, D. N., & Hwang, T. C. (2011). Application of high-resolution single-channel recording to functional studies of cystic fibrosis mutants. In *Cystic Fibrosis* (pp. 419-441). Humana Press.

Campbell, L. M., Maxwell, P. J., & Waugh, D. J. (2013). Rationale and means to target pro-inflammatory interleukin-8 (CXCL8) signaling in cancer. *Pharmaceuticals*, 6(8), 929-959.

Casale, C., Imparato, G., Urciuolo, F., & Netti, P. A. (2016). Endogenous human skin equivalent promotes in vitro morphogenesis of follicle-like structures. *Biomaterials*, 101, 86-95.

Caspar, S. M., Dubacher, N., Kopps, A. M., Meienberg, J., Henggeler, C., & Matyas, G. (2018). Clinical sequencing: from raw data to diagnosis with lifetime value. *Clinical genetics*, 93(3), 508-519.

Castellani C, Piccoli L, Tamanini, A Giradi, P, Rizzotte P, et al. (2009) Association between carrier screening and incidence of cystic fibrosis. *JAMA* 302: 2573-2579.,

Castellani, C., Cuppens, H., Macek Jr, M., Cassiman, J. J., Kerem, E., Durie, P., ... & Elborn, J. (2008). Consensus on the use and interpretation of cystic fibrosis mutation analysis in clinical practice. *Journal of cystic fibrosis*, 7(3), 179-196.

Castro-Wallace, S. L., Chiu, C. Y., John, K. K., Stahl, S. E., Rubins, K. H., McIntyre, A. B., ... & Burton, A. S. (2017). Nanopore DNA sequencing and genome assembly on the International Space Station. *Scientific reports*, 7(1), 1-12.

Chappe, V., Hinkson, D. A., Zhu, T., Chang, X. B., Riordan, J. R., & Hanrahan, J. W. (2003). Phosphorylation of protein kinase C sites in NBD1 and the R domain control CFTR channel activation by PKA. *The Journal of physiology*, 548(1), 39-52.

Chen, F., Zhang, C., Jia, X., Wang, S., Wang, J., Chen, Y., ... & Han, L. (2015). Transcriptome profiles of human lung epithelial cells A549 interacting with *Aspergillus fumigatus* by RNA-Seq. *PloS one*, 10(8), e0135720.

Chen, X., Nadiarynkh, O., Plotnikov, S., & Campagnola, P. J. (2012). Second harmonic generation microscopy for quantitative analysis of collagen fibrillar structure. *Nature protocols*, 7(4), 654-669.

Cheng, S. H., Gregory, R. J., Marshall, J., Paul, S., Souza, D. W., White, G. A., ... & Smith, A. E. (1990). Defective intracellular transport and processing of CFTR is the molecular basis of most cystic fibrosis. *Cell*, 63(4), 827-834.

Choi, M., Scholl, U. I., Ji, W., Liu, T., Tikhonova, I. R., Zumbo, P., ... & Lifton, R. P. (2009). Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proceedings of the National Academy of Sciences*, 106(45), 19096-19101.

Chong, J. X., Buckingham, K. J., Jhangiani, S. N., Boehm, C., Sobreira, N., Smith, J. D., ... & Bamshad, M. J. (2015). The genetic basis of Mendelian phenotypes: discoveries, challenges, and opportunities. *The American Journal of Human Genetics*, 97(2), 199-215.

Chotalia, M., Smallwood, S. A., Ruf, N., Dawson, C., Lucifero, D., Frontera, M., ... & Kelsey, G. (2009). Transcription is required for establishment of germline methylation marks at imprinted genes. *Genes & development*, 23(1), 105-117.

Cicchi, R., Kapsokalyvas, D., De Giorgi, V., Maio, V., Van Wiechen, A., Massi, D., ... & Pavone, F. S. (2010). Scoring of collagen organization in healthy and diseased human dermis by multiphoton microscopy. *Journal of biophotonics*, 3(1-2), 34-43.

Cocquet, J., Chong, A., Zhang, G., & Veitia, R. A. (2006). Reverse transcriptase template switching and false alternative transcripts. *Genomics*, 88(1), 127-131.

Cole, M. A., Quan, T., Voorhees, J. J., & Fisher, G. J. (2018). Extracellular matrix regulation of fibroblast function: redefining our perspective on skin aging. *Journal of cell communication and signaling*, 12(1), 35-43.

Collins, F. S. (1992). Cystic fibrosis: Molecular Biology and Therapeutic Implications. *Science*, 256(5058), 774-779.

Collins, F. S. (2019). Realizing the dream of molecularly targeted therapies for cystic fibrosis. *New England Journal of Medicine*, 381(19), 1863-1865.

Collins, R. L., Brand, H., Karczewski, K. J., Zhao, X., Alföldi, J., Francioli, L. C., ... & Talkowski, M. E. (2020). A structural variation reference for medical and population genetics. *Nature*, 581(7809), 444-451.

Conrad, D. F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., ... & Hurles, M. E. (2010). Origins and functional impact of copy number variation in the human genome. *Nature*, 464(7289), 704-712.

Cooney, A. L., McCray, P. B., & Sinn, P. L. (2018). Cystic fibrosis gene therapy: looking back, looking forward. *Genes*, 9(11), 538.

Coriati, A., Bouvet, G. F., Massé, C., Ducruet, T., & Berthiaume, Y. (2021). Ykl-40 as a clinical biomarker in adult patients with cf: Implications of a chi3l1 single nucleotide polymorphism in disease severity. *Journal of Cystic Fibrosis*, 20(6), e93-e99.

Coulondre, C., Miller, J. H., Farabaugh, P. J., & Gilbert, W. (1978). Molecular basis of base substitution hotspots in *Escherichia coli*. *Nature*, 274(5673), 775-780.

Cutting, G. R. (2015). Cystic fibrosis genetics: from molecular understanding to clinical application. *Nature Reviews Genetics*, 16(1), 45-56.

Dalemans, W., Hinnrasky, J., Slos, P., Dreyer, D., Fuchey, C., Pavirani, A., & Puchelle, E. (1992). Immunocytochemical analysis reveals differences between the subcellular localization of normal and Δ Phe508 recombinant cystic fibrosis transmembrane conductance regulator. *Experimental cell research*, 201(1), 235-240.

Davies, J. C., Drevinek, P., Elborn, J. S., Kerem, E., Lee, T., Amaral, M. D., & de Boeck, K. (2019). Speeding up access to new drugs for CF: Considerations for clinical trial design and delivery. *Journal of Cystic Fibrosis*, 18(5), 677-684.

De Boeck, K., & Amaral, M. D. (2016). Progress in therapies for cystic fibrosis. *Lancet Respir Med* 4: 662–674. Link: <https://goo.gl/jbkzG4>.

De Boeck, K., Kent, L., Davies, J., Derichs, N., Amaral, M., Rowe, S. M., ... & Sermet-Gaudelus, I. (2013). CFTR biomarkers: time for promotion to surrogate endpoint. *European Respiratory Journal*, 41(1), 203-216.

De Gregorio, V., Corrado, B., Sbrescia, S., Sibilio, S., Urciuolo, F., Netti, P. A., & Imparato, G. (2020). Intestine-on-chip device increases ECM remodeling inducing faster epithelial cell differentiation. *Biotechnology and Bioengineering*, 117(2), 556-566.

De Maglio G., Pasello G, Dono M, Fiorentino M, Follador A, Sciortino M, Malapelle U, Tiseo M. (2022). The storm of NGS in NSCLC diagnostic-therapeutic pathway: How to sun the real clinical practice. *Crit Rev Oncol Hematol*. 169:103561. doi: 10.1016/j.critrevonc.2021.103561. Epub 2021 Nov 29.

de Winter–de Groot, K. M., Berkers, G., Marck–van der Wilt, R. E., van der Meer, R., Vonk, A., Dekkers, J. F., ... & Beekman, J. M. (2020). Forskolin-induced swelling of intestinal organoids correlates with disease severity in adults with cystic fibrosis and homozygous F508del mutations. *Journal of Cystic Fibrosis*, 19(4), 614-619.

Dekkers, J. F., Berkers, G., Kruisselbrink, E., Vonk, A., De Jonge, H. R., Janssens, H. M., ... & Beekman, J. M. (2016). Characterizing responses to CFTR-modulating drugs using rectal organoids derived from subjects with cystic fibrosis. *Science translational medicine*, 8(344), 344ra84-344ra84.

Dekkers, J. F., Wiegerinck, C. L., De Jonge, H. R., Bronsveld, I., Janssens, H. M., De Winter-de Groot, K. M., ... & Beekman, J. M. (2013). A functional CFTR assay using primary cystic fibrosis intestinal organoids. *Nature medicine*, 19(7), 939-945.

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., ... & Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature genetics*, 43(5), 491-498.

Di Fonzo, A., Monfrini, E., & Erro, R. (2018). Genetics of movement disorders and the practicing clinician; who and what to test for?. *Current Neurology and Neuroscience Reports*, 18(7), 1-11.

Dillies, M. A., Rau, A., Aubert, J., Hennequet-Antier, C., Jeanmougin, M., Servant, N., ... & Jaffrézic, F. (2013). A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Briefings in bioinformatics*, 14(6), 671-683.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., ... & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1), 15-21.

Donaldson, S. H., Bennett, W. D., Zeman, K. L., Knowles, M. R., Tarran, R., & Boucher, R. C. (2006). Mucus clearance and lung function in cystic fibrosis with hypertonic saline. *New England Journal of Medicine*, 354(3), 241-250.

Drumm, M. L., Pope, H. A., Cliff, W. H., Rommens, J. M., Marvin, S. A., Tsui, L. C., ... & Wilson, J. M. (1990). Correction of the cystic fibrosis defect in vitro by retrovirus-mediated gene transfer. *Cell*, 62(6), 1227-1233.

Drumm, M.L., Wilkinson, D.J., Smit, L.S., Worrell, R.T., Strong, T.V., Frizzell, R.A., Dawson, D.C., Collins, F.S. (1991) Chloride conductance expressed by DF508 and other mutant CFTRs in *Xenopus* oocytes. *Science* 254:1797–1799

Dunning, M. J., Ritchie, M. E., Barbosa-Morais, N. L., Tavaré, S., & Lynch, A. G. (2008). Spike-in validation of an Illumina-specific variance-stabilizing transformation. *BMC research notes*, 1(1), 1-6.

Durmowicz, A. G., Lim, R., Rogers, H., Rosebraugh, C. J., & Chowdhury, B. A. (2018). The US Food and Drug Administration's experience with ivacaftor in cystic fibrosis. Establishing efficacy using in vitro data in lieu of a clinical trial. *Annals of the American Thoracic Society*, 15(1), 1-2.

Ebihara, N., Matsuda, A., Nakamura, S., Matsuda, H., & Murakami, A. (2011). Role of the IL-6 classic-and trans-signaling pathways in corneal sterile inflammation and wound healing. *Investigative ophthalmology & visual science*, 52(12), 8549-8557.

Eckford, P. D., McCormack, J., Munsie, L., He, G., Stanojevic, S., Pereira, S. L., ... & Bear, C. E. (2019). The CF Canada-Sick Kids Program in individual CF therapy: A resource for the advancement of personalized medicine in CF. *Journal of Cystic Fibrosis*, 18(1), 35-43.

Edwards, A., Debbonaire, A. R., Nicholls, S. M., Rassner, S. M., Sattler, B., Cook, J. M., ... & Hodson, A. J. (2019). In-field metagenome and 16S rRNA gene amplicon nanopore sequencing robustly characterize glacier microbiota. *BioRxiv*, 073965.

Eldomery, M. K., Coban-Akdemir, Z., Harel, T., Rosenfeld, J. A., Gambin, T., Stray-Pedersen, A., ... & Lupski, J. R. (2017). Lessons learned from additional research analyses of unsolved clinical exome cases. *Genome medicine*, 9(1), 1-15.

Ellsworth, R. E., Jamison, D. C., Touchman, J. W., Chisoe, S. L., Maduro, V. V. B., Bouffard, G. G., ... & Green, E. D. (2000). Comparative genomic sequence analysis of the human and mouse cystic fibrosis transmembrane conductance regulator genes. *Proceedings of the National Academy of Sciences*, 97(3), 1172-1177.

Elsaid, M. F., Chalhoub, N., Ben-Omran, T., Kumar, P., Kamel, H., Ibrahim, K., ... & Aleem, A. A. (2017). Mutation in noncoding RNA RNU12 causes early onset cerebellar ataxia. *Annals of neurology*, 81(1), 68-78.

Engström PG., Steijger T, Sipos B, Grant GR, Kahles A, The RGASP Consortium, Räscher G, Goldman N, Hubbard TJ, Harrow J et al (2013) Systematic evaluation of spliced alignment programs for RNA-seq data. *Nat Methods* 10:1185–1191

Ergin, S., Kherad, N., & Alagoz, M. (2022). RNA sequencing and its applications in cancer and rare diseases. *Molecular Biology Reports*, 1-9.

Estabrooks, S., & Brodsky, J. L. (2020). Regulation of CFTR biogenesis by the proteostatic network and pharmacological modulators. *International Journal of Molecular Sciences*, 21(2), 452.

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, 32(19), 3047-3048.

Farinha, C. M. (2018). CFTR and cystic fibrosis. In *CFTR and Cystic Fibrosis* (pp. 1-56). Springer, Cham.

Farinha, C. M., Nogueira, P., Mendes, F., Penque, D., & Amaral, M. D. (2002). The human DnaJ homologue (Hdj)-1/heat-shock protein (Hsp) 40 co-chaperone is required for the in vivo stabilization of the cystic fibrosis transmembrane conductance regulator by Hsp70. *Biochemical Journal*, 366(3), 797-806.

Farinha, C. M., Swiatecka-Urban, A., Brautigan, D. L., & Jordan, P. (2016). Regulatory crosstalk by protein kinases on CFTR trafficking and activity. *Frontiers in chemistry*, 4, 1.

Farwell, K. D., Shahmirzadi, L., El-Khechen, D., Powis, Z., Chao, E. C., Tippin Davis, B., ... & Tang, S. (2015). Enhanced utility of family-centered diagnostic exome sequencing with inheritance model-based analysis: results from 500 unselected families with undiagnosed genetic conditions. *Genetics in Medicine*, 17(7), 578-586.

Ferrie RM, Schwarz MJ, Robertson NH, Vaudin S, Super M, Malone G, Little S (1992) Development, multiplexing, and application of ARMS tests for common mutations in the CFTR gene. *Am J Hum Genet* 51(2):251–262

Frankish, A., Diekhans, M., Ferreira, A. M., Johnson, R., Jungreis, I., Loveland, J., ... & Flicek, P. (2019). GENCODE reference annotation for the human and mouse genomes. *Nucleic acids research*, 47(D1), D766-D773.

Frickmann, H., Jungblut, S., Hirche, T. O., Groß, U., Kuhns, M., & Zautner, A. E. (2012). Spectrum of viral infections in patients with cystic fibrosis. *European Journal of Microbiology and Immunology*, 2(3), 161-175.

Fu, L., Zhang, C., Zhang, L. Y., Dong, S. S., Lu, L. H., Chen, J., ... & Guan, X. Y. (2011). Wnt2 secreted by tumour fibroblasts promotes tumour progression in oesophageal cancer by activation of the Wnt/ β -catenin signalling pathway. *Gut*, 60(12), 1635-1643.

Fu, X., Cheng, Y., Yuan, J., Huang, C., Cheng, H., & Zhou, R. (2015). Loss-of-function mutation in the X-linked TBX22 promoter disrupts an ETS-1 binding site and leads to cleft palate. *Human genetics*, 134(2), 147-158.

Furlan LL, Marson FA, Ribeiro JD, Bertuzzo CS, Salomao Junior JB, Souza DR. IL8 gene as modifier of cystic fibrosis: unraveling the factors which influence clinical variability. *Hum Genet*. 2016; 135(8):881–94. <https://doi.org/10.1007/s00439-016-1684-4> PMID: 27209008.

Gaggar, A., Hector, A., Bratcher, P. E., Mall, M. A., Griesse, M., & Hartl, D. (2011). The role of matrix metalloproteinases in cystic fibrosis lung disease. *European Respiratory Journal*, 38(3), 721-727.

Gallati, S. (2014). Disease-modifying genes and monogenic disorders: experience in cystic fibrosis. *The application of clinical genetics*, 7, 133.

Genin, E., Feingold, J., & Clerget-Darpoux, F. (2008). Identifying modifier genes of monogenic disease: strategies and difficulties. *Human genetics*, 124(4), 357-368.

Ghatak, S., Misra, S., Norris, R. A., Moreno-Rodriguez, R. A., Hoffman, S., Levine, R. A., ... & Markwald, R. R. (2014). Periostin induces intracellular cross-talk between kinases and hyaluronan in atrioventricular valvulogenesis. *Journal of Biological Chemistry*, 289(12), 8545-8561.

Gong, X., Ahner, A., Roldan, A., Lukacs, G. L., Thibodeau, P. H., & Frizzell, R. A. (2016). Non-native conformers of cystic fibrosis transmembrane conductance regulator NBD1 are recognized by Hsp27 and conjugated to SUMO-2 for degradation. *Journal of Biological Chemistry*, 291(4), 2004-2017.

Gonorazky, H. D., Naumenko, S., Ramani, A. K., Nelakuditi, V., Mashouri, P., Wang, P., ... & Dowling, J. J. (2019). Expanding the boundaries of RNA sequencing

as a diagnostic tool for rare Mendelian disease. *The American Journal of Human Genetics*, 104(3), 466-483.

Gonzaga-Jauregui, C., & Lupski, J. R. (Eds.). (2021). *Genomics of Rare Diseases: Understanding Disease Genetics Using Genomic Approaches*. Academic Press.

Gonzaga-Jauregui, C., & Mendoza, C. J. Z. (2021). Genomic sequencing of rare diseases. In *Genomics of Rare Diseases*. Academic Press. Page 76

Gonzaga-Jauregui, C., Lupski, J. R., & Gibbs, R. A. (2012). Human genome sequencing in health and disease. *Annual review of medicine*, 63, 35-61.

González-González, L., & Alonso, J. (2018). Periostin: a matricellular protein with multiple functions in cancer development and progression. *Frontiers in oncology*, 8, 225.

Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, 17(6), 333-351.

Gorcenco, S., Ilinca, A., Almasoudi, W., Kafantari, E., Lindgren, A. G., & Puschmann, A. (2020). New generation genetic testing entering the clinic. *Parkinsonism & related disorders*, 73, 72-84.

Gripp, K. W., Curry, C., Olney, A. H., Sandoval, C., Fisher, J., Chong, J. X.-L., UW Center for Mendelian Genomics, Pilchman, L., Sahraoui, R., Stabley, D. L., Sol-Church, K. (2014). Diamond-Blackfan anemia with mandibulofacial dystostosis (sic) is heterogeneous, including the novel DBA genes TSR2 and RPS28. *Am. J. Med. Genet.* 164A: 2240-2249

Grove, D. E., Fan, C. Y., Ren, H. Y., & Cyr, D. M. (2011). The endoplasmic reticulum-associated Hsp40 DNAJB12 and Hsc70 cooperate to facilitate RMA1 E3-dependent degradation of nascent CFTR Δ F508. *Molecular biology of the cell*, 22(3), 301-314.

GTEx Consortium. (2015). The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*, 348(6235), 648-660.

Guggino, W. B., & Stanton, B. A. (2006). Molecular switches regulating CFTR trafficking and activity. *Nat Rev Mol Cell Biol*, 7, 426-436.

Haas, B. J., Chin, M., Nusbaum, C., Birren, B. W., & Livny, J. (2012). How deep is deep enough for RNA-Seq profiling of bacterial transcriptomes?. *BMC genomics*, 13(1), 1-11.

Halbeisen, R. E., Galgano, A., Scherrer, T., & Gerber, A. (2008). Post-transcriptional gene regulation: from genome-wide studies to principles. *Cellular and molecular life sciences*, 65(5), 798-813.

Hall, A. K., & Carlson, M. R. (2014). The current status of orphan drug development in Europe and the US. *Intractable & rare diseases research*, 3(1), 1-7.

Hall, A. K., & Carlson, M. R. (2014). The current status of orphan drug development in Europe and the US. *Intractable & rare diseases research*, 3(1), 1-7.

Hansen, K. D., Brenner, S. E., & Dudoit, S. (2010). Biases in Illumina transcriptome sequencing caused by random hexamer priming. *Nucleic acids research*, 38(12), e131-e131.

Hashimshony, T., Wagner, F., Sher, N., & Yanai, I. (2012). CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell reports*, 2(3), 666-673.

Hays, S. R., & Fahy, J. V. (2006). Characterizing mucous cell remodeling in cystic fibrosis: relationship to neutrophils. *American journal of respiratory and critical care medicine*, 174(9), 1018-1024.

He, S., Wurtzel, O., Singh, K., Froula, J. L., Yilmaz, S., Tringe, S. G., ... & Hugenholtz, P. (2010). Validation of two ribosomal RNA removal methods for microbial metatranscriptomics. *Nature methods*, 7(10), 807-812.

Hector, A., Kormann, M. S., Mack, I., Latzin, P., Casaulta, C., Kieninger, E., ... & Hartl, D. (2011). The chitinase-like protein YKL-40 modulates cystic fibrosis lung disease. *PloS one*, 6(9), e24399.

Heemstra, H. E., van Weely, S., Büller, H. A., Leufkens, H. G., & de Vrieh, R. L. (2009). Translation of rare disease research into orphan drug development: disease matters. *Drug discovery today*, 14(23-24), 1166-1173.

Helling, B. A., & Yang, I. V. (2015). Epigenetics in lung fibrosis: from pathobiology to treatment perspective. *Current opinion in pulmonary medicine*, 21(5), 454.

Hita, A., Brocart, G., Fernandez, A., Rehmsmeier, M., Alemany, A., & Schvartzman, S. (2022). MGcount: a total RNA-seq quantification tool to address multi-mapping and multi-overlapping alignments ambiguity in non-coding transcripts. *BMC bioinformatics*, 23(1), 1-21.

Housby, J. N., Cahill, C. M., Chu, B., Prevelige, R., Bickford, K., Stevenson, M. A., & Calderwood, S. K. (1999). Non-steroidal anti-inflammatory drugs inhibit the expression of cytokines and induce HSP70 in human monocytes. *Cytokine*, 11(5), 347-358.

Houseley, J., & Tollervey, D. (2010). Apparent non-canonical trans-splicing is generated by reverse transcriptase in vitro. *PloS one*, 5(8), e12271.

Hristu, R., Eftimie, L. G., Stanciu, S. G., Tranca, D. E., Paun, B., Sajin, M., & Stanciu, G. A. (2018). Quantitative second harmonic generation microscopy for the structural characterization of capsular collagen in thyroid neoplasms. *Biomedical Optics Express*, 9(8), 3923-3936.

https://github.com/hbctraining/DGE_workshop_salmon/blob/master/lessons/01_DGE_setup_and_overview.md,

Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic acids research*, 37(1), 1-13.

Huaux, F., Noel, S., Dhooghe, B., Panin, N., Lo Re, S., Lison, D., ... & Leal, T. (2013). Dysregulated proinflammatory and fibrogenic phenotype of fibroblasts in cystic fibrosis. *PloS one*, 8(5), e64341.

Hung, C. F., Wilson, C. L., Chow, Y. H., & Schnapp, L. M. (2018). Role of integrin alpha8 in murine model of lung fibrosis. *PLoS One*, 13(5), e0197937.

Hwang TC, Kirk KL (2013). The CFTR ion channel: gating, regulation, and anion permeation. *Cold Spring Harb Perspect Med*. 1;3(1):a009498.

Imparato, G., Casale, C., Scamardella, S., Urciuolo, F., Bimonte, M., Apone, F., ... & Netti, P. A. (2017). A novel engineered dermis for in vitro photodamage research. *Journal of tissue engineering and regenerative medicine*, 11(8), 2276-2285.

Imparato, G., Urciuolo, F., Casale, C., & Netti, P. A. (2013). The role of micro scaffold properties in controlling the collagen assembly in 3D dermis equivalent using modular tissue engineering. *Biomaterials*, 34(32), 7851-7861.

Inglis, S. K., & Wilson, S. M. (2005). Cystic fibrosis and airway submucosal glands. *Pediatric pulmonology*, 40(4), 279-284.

Ivanov, N. A., Tao, R., Chenoweth, J. G., Brandtjen, A., Mighdoll, M. I., Genova, J. D., ... & Jaffe, A. E. (2016). Strong components of epigenetic memory in cultured human fibroblasts related to site of origin and donor age. *PLoS genetics*, 12(2), e1005819.

Jackson, S. R., Costa, M. F. D. M., Pastore, C. F., Zhao, G., Weiner, A. I., Adams, S., ... & Vaughan, A. E. (2020). R-spondin 2 mediates neutrophil egress into the alveolar space through increased lung permeability. *BMC Research Notes*, 13(1), 1-7.

Jehl F, Degalez F, Bernard M, Lecerf F, Lagoutte L, Désert C, Coulée M, Bouchez O, Leroux S, Abasht B, Tixier-Boichard M, Bed'hom B, Burlot T, Gourichon D, Bardou P, Acloque H, Foissac S, Djebali S, Giuffra E, Zerjal T, Pitel F, Klopp C, Lagarrigue S. (2021). RNA-Seq Data for Reliable SNP Detection and Genotype Calling: Interest for Coding Variant Characterization and Cis-Regulation Analysis by Allele-Specific Expression in Livestock Species. *Front Genet*. 28;12:655707. doi: 10.3389/fgene.2021.655707.

Ji, H., Greening, D. W., Kapp, E. A., Moritz, R. L., & Simpson, R. J. (2009). Secretome-based proteomics reveals sulindac-modulated proteins released from colon cancer cells. *PROTEOMICS–Clinical Applications*, 3(4), 433-451.

Jin, Y., Wang, J., Bachtar, M., Chong, S. S., & Lee, C. G. (2018). Architecture of polymorphisms in the human genome reveals functionally important and positively selected variants in immune response and drug transporter genes. *Human genomics*, 12(1), 1-13.

Johnson, Z. J., Krutkin, D. D., Bohutskyi, P., & Kalyuzhnaya, M. G. (2021). Metals and methylotrophy: Via global gene expression studies. In *Methods in Enzymology* (Vol. 650, pp. 185-213). Academic Press.

Ju, J., Kim, D. H., Bi, L., Meng, Q., Bai, X., Li, Z., ... & Turro, N. J. (2006). Four-color DNA sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators. *Proceedings of the National Academy of Sciences*, 103(52), 19635-19640.

Jundi, K., & Greene, C. M. (2015). Transcription of interleukin-8: how altered regulation can affect cystic fibrosis lung disease. *Biomolecules*, 5(3), 1386-1398.

Kabani, M., McLellan, C., Raynes, D. A., Guerriero, V., & Brodsky, J. L. (2002). HspBP1, a homologue of the yeast Fes1 and Sls1 proteins, is an Hsc70 nucleotide exchange factor. *FEBS letters*, 531(2), 339-342.

Kälin, N., Claaß, A., Sommer, M., Puchelle, E., & Tümmler, B. (1999). Δ F508 CFTR protein expression in tissues from patients with cystic fibrosis. *The Journal of clinical investigation*, 103(10), 1379-1389.

Karolak, J. A., Vincent, M., Deutsch, G., Gambin, T., Cogné, B., Pichon, O., ... & Stankiewicz, P. (2019). Complex compound inheritance of lethal lung developmental disorders due to disruption of the TBX-FGF pathway. *The American Journal of Human Genetics*, 104(2), 213-228.

Kartner, N., Augustinas, O., Jensen, T. J., Naismith, A. L., & Riordan, J. R. (1992). Mislocalization of Δ F508 CFTR in cystic fibrosis sweat gland. *Nature genetics*, 1(5), 321-327.

Kerem, B. S., Rommens, J. M., Buchanan, J. A., Markiewicz, D., Cox, T. K., Chakravarti, A., ... & Tsui, L. C. (1989). Identification of the cystic fibrosis gene: genetic analysis. *Science*, 245(4922), 1073-1080.

Kim, D., & Langmead, B. (2015). Salzberg SLHISAT. A fast spliced aligner with low memory requirements. *Nat Methods*, 12(4), 357-60.

Kim, D., Kim, S. Y., Mun, S. K., Rhee, S., & Kim, B. J. (2015). Epidermal growth factor improves the migration and contractility of aged fibroblasts cultured on 3D collagen matrices. *International Journal of Molecular Medicine*, 35(4), 1017-1025.

Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature biotechnology*, 37(8), 907-915.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., & Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome biology*, 14(4), 1-13.

Koh, J., Sferra, T. J., & Collins, F. S. (1993). Characterization of the cystic fibrosis transmembrane conductance regulator promoter region. Chromatin context and tissue-specificity. *Journal of Biological Chemistry*, 268(21), 15912-15921.

Königshoff, M., Balsara, N., Pfaff, E. M., Kramer, M., Chrobak, I., Seeger, W., & Eickelberg, O. (2008). Functional Wnt signaling is increased in idiopathic pulmonary fibrosis. *PloS one*, 3(5), e2142.

Kormann, M. S., Dewerth, A., Eichner, F., Baskaran, P., Hector, A., Regamey, N., ... & Antony, J. S. (2017). Transcriptomic profile of cystic fibrosis patients identifies type I interferon response and ribosomal stalk proteins as potential modifiers of disease severity. *PLoS One*, 12(8), e0183526.

Kornegay, J. N., Tuler, S. M., Miller, D. M., & Levesque, D. C. (1988). Muscular dystrophy in a litter of golden retriever dogs. *Muscle & Nerve: Official Journal of the American Association of Electrodiagnostic Medicine*, 11(10), 1056-1064.

Kousi, M. (2021), Transcriptomics in rare diseases. In: Genomics of Rare Diseases Understanding Disease Genetics Using Genomic Approaches. Gonzaga-Jauregui C, Lupski JR, (eds), UK: 2021 Elsevier Inc, s: 216.

Kramer, N., Schmöllerl, J., Unger, C., Nivarthi, H., Rudisch, A., Unterleuthner, D., ... & Dolznig, H. (2017). Autocrine WNT2 signaling in fibroblasts promotes colorectal cancer progression. *Oncogene*, 36(39), 5460-5472.

Kremer, L. S., Bader, D. M., Mertes, C., Kopajtich, R., Pichler, G., Iuso, A., ... & Prokisch, H. (2017). Genetic diagnosis of Mendelian disorders via RNA sequencing. *Nature communications*, 8(1), 1-11.

Kudo, A., & Kii, I. (2018). Periostin function in communication with extracellular matrices. *Journal of cell communication and signaling*, 12(1), 301-308.

Lagarrigue, S., Martin, L., Hormozdiari, F., Roux, P.-F., Pan, C., van Nas, A., et al. (2013b). Analysis of allele-specific expression in mouse liver by RNASeq: a comparison with Cis -eQTL identified using genetic linkage. *Genetics* 195, 1157–1166. doi: 10.1534/genetics.113.153882

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., ... & Frazier, M. (2001). C. International Human Genome Sequencing. *Initial sequencing and analysis of the human genome*, 860-921.

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology*, 10(3), 1-10.

Lappalainen, T., Sammeth, M., Friedländer, M. R., t Hoen, P. A., Monlong, J., Rivas, M. A., ... & Dermitzakis, E. T. (2013). Transcriptome and genome sequencing uncovers functional variation in humans. *Nature*, 501(7468), 506-511.

Law, C. W., Chen, Y., Shi, W., & Smyth, G. K. (2014). voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome biology*, 15(2), 1-17.

Lee, H., Deignan, J. L., Dorrani, N., Strom, S. P., Kantarci, S., Quintero-Rivera, F., ... & Nelson, S. F. (2014). Clinical exome sequencing for genetic identification of rare Mendelian disorders. *Jama*, 312(18), 1880-1887.

Lee, J. T., & Bartolomei, M. S. (2013). X-inactivation, imprinting, and long noncoding RNAs in health and disease. *Cell*, 152(6), 1308-1323.

Lee, T. I., & Young, R. A. (2013). Transcriptional regulation and its misregulation in disease. *Cell*, 152(6), 1237-1251.

Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., ... & MacArthur, D. G. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, 536(7616), 285-291.

Levin, J. Z., Yassour, M., Adiconis, X., Nusbaum, C., Thompson, D. A., Friedman, N., ... & Regev, A. (2010). Comprehensive comparative analysis of strand-specific RNA sequencing methods. *Nature methods*, 7(9), 709-715.

Lewandowska, M. A., Costa, F. F., Bischof, J. M., Williams, S. H., Soares, M. B., & Harris, A. (2010). Multiple mechanisms influence regulation of the cystic fibrosis transmembrane conductance regulator gene promoter. *American journal of respiratory cell and molecular biology*, 43(3), 334-341.

Li, R., Li, Y., Fang, X., Yang, H., Wang, J., Kristiansen, K., & Wang, J. (2009). SNP detection for massively parallel whole-genome resequencing. *Genome research*, 19(6), 1124-1132.

Li, Y., Jiang, D., Liang, J., Meltzer, E. B., Gray, A., Miura, R., ... & Noble, P. W. (2011). Severe lung fibrosis requires an invasive fibroblast phenotype regulated by hyaluronan and CD44. *Journal of Experimental Medicine*, 208(7), 1459-1471.

Lindquist, S. G., Duno, M., Batbayli, M., Puschmann, A., Braendgaard, H., Mardosiene, S., ... & Nielsen, J. E. (2013). Corticobasal and ataxia syndromes widen the spectrum of C9ORF72 hexanucleotide expansion disease. *Clinical genetics*, 83(3), 279-283.

Lister R et al (2008) Highly integrated single- base resolution maps of the epigenome in arabidopsis. *Cell* 133:523–536

Liu F, Zhang Z, Csanady L, Gadsby DC, Chen J (2017) Molecular structure of the human CFTR Ion channel. *Cell* 169(1):85–95 e88. doi:10.1016/j.cell.2017.02.024

Liu, C., Xiao, F., Hoisington-Lopez, J., Lang, K., Quenzel, P., Duffy, B., & Mitra, R. D. (2017). Accurate typing of class I human leukocyte antigen by Oxford nanopore sequencing. *BioRxiv*, 178590.

Liu, X., Han, S., Wang, Z., Gelernter, J., & Yang, B. Z. (2013). Variant callers for next-generation sequencing data: a comparison study. *PloS one*, 8(9), e75619.

Loo, M. A., Jensen, T. J., Cui, L., Hou, Y. X., Chang, X. B., & Riordan, J. R. (1998). Perturbation of Hsp90 interaction with nascent CFTR prevents its maturation and accelerates its degradation by the proteasome. *The EMBO journal*, 17(23), 6879-6887.

López-Bigas, N., Audit, B., Ouzounis, C., Parra, G., & Guigó, R. (2005). Are splicing mutations the most frequent cause of hereditary disease?. *FEBS letters*, 579(9), 1900-1903.

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*, 15(12), 1-21.

Lu, Y., Xiong, X., Helm, A., Kimani, K., Bragin, A., & Skach, W. R. (1998). Co-and posttranslational translocation mechanisms direct cystic fibrosis transmembrane conductance regulator N terminus transmembrane assembly. *Journal of Biological Chemistry*, 273(1), 568-576.

Lukacs, G. L., Chang, X. B., Bear, C., Kartner, N., Mohamed, A., Riordan, J. R., & Grinstein, S. (1993). The delta F508 mutation decreases the stability of cystic fibrosis transmembrane conductance regulator in the plasma membrane. Determination of functional half-lives on transfected cells. *Journal of Biological Chemistry*, 268(29), 21592-21598.

Lukacs, G. L., Chang, X. B., Kartner, N., Rotstein, O. D., Riordan, J. R., & Grinstein, S. (1992). The cystic fibrosis transmembrane regulator is present and

functional in endosomes. Role as a determinant of endosomal pH. *Journal of Biological Chemistry*, 267(21), 14568-14572.

Lyon, M. F. (1986). X chromosomes and dosage compensation. *Nature*, 320(6060), 313-313.

Mack, M. (2018). Inflammation and fibrosis. *Matrix Biology*, 68, 106-121.

Magalhães, M., Rivals, I., Claustres, M., Varilh, J., Thomasset, M., Bergougnoux, A., ... & De Sario, A. (2017). DNA methylation at modifier genes of lung disease severity is altered in cystic fibrosis. *Clinical epigenetics*, 9(1), 1-15.

Magoč, T., & Salzberg, S. L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, 27(21), 2957-2963.

Mah, J. K., Korngut, L., Dykeman, J., Day, L., Pringsheim, T., & Jette, N. (2014). A systematic review and meta-analysis on the epidemiology of Duchenne and Becker muscular dystrophy. *Neuromuscular Disorders*, 24(6), 482-491.

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bembien, L. A., ... & Rothberg, J. M. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 437(7057), 376-380.

Marozkina, N. V., Yemen, S., Borowitz, M., Liu, L., Plapp, M., Sun, F., ... & Zaman, K. (2010). Hsp 70/Hsp 90 organizing protein as a nitrosylation target in cystic fibrosis therapy. *Proceedings of the National Academy of Sciences*, 107(25), 11393-11398.

Marquardt N, Kekäläinen E, Chen P, Lourda M, Wilson JN, Scharenberg M, Bergman P, Al-Ameri M, Hård J, Mold JE, Ljunggren HG, Michaëlsson J. (2019). Unique transcriptional and protein-expression signature in human lung tissue-resident NK cells. *Nat Commun*. 26;10(1):3841. doi: 10.1038/s41467-019-11632-9.

Matsumura, Y., David, L. L., & Skach, W. R. (2011). Role of Hsc70 binding cycle in CFTR folding and endoplasmic reticulum-associated degradation. *Molecular biology of the cell*, 22(16), 2797-2809.

Mazio C., Scognamiglio, L. S., De Cegli, R., Galiotta, L. J., Di Bernardo, D., Casale, C., ... & Netti, P. A. (2020). Intrinsic abnormalities of cystic fibrosis airway connective tissue revealed by an in vitro 3D stromal model. *Cells*, 9(6), 1371..

McCarthy, D. J., Chen, Y., & Smyth, G. K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic acids research*, 40(10), 4288-4297...

McCombie, W. R., McPherson, J. D., & Mardis, E. R. (2019). Next-generation sequencing technologies. *Cold Spring Harbor perspectives in medicine*, 9(11), a036798.

McKean, D. M., Homsy, J., Wakimoto, H., Patel, N., Gorham, J., DePalma, S. R., ... & Seidman, J. G. (2016). Loss of RNA expression and allele-specific expression associated with congenital heart disease. *Nature communications*, 7(1), 1-9.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., ... & DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research*, 20(9), 1297-1303.

McKernan, K. J., Peckham, H. E., Costa, G. L., McLaughlin, S. F., Fu, Y., Tsung, E. F., ... & Blanchard, A. P. (2009). Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding. *Genome research*, 19(9), 1527-1541.

McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R., Thormann, A., ... & Cunningham, F. (2016). The ensembl variant effect predictor. *Genome biology*, 17(1), 1-14.

McNab, F., Mayer-Barber, K., Sher, A., Wack, A., & O'garra, A. (2015). Type I interferons in infectious disease. *Nature Reviews Immunology*, 15(2), 87-103.

Meacham, G. C., Lu, Z., King, S., Sorscher, E., Tousson, A., & Cyr, D. M. (1999). The Hdj-2/Hsc70 chaperone pair facilitates early steps in CFTR biogenesis. *The EMBO journal*, 18(6), 1492-1505.

Menegon, M., Cantaloni, C., Rodriguez-Prieto, A., Centomo, C., Abdelfattah, A., Rossato, M., ... & Delledonne, M. (2017). On site DNA barcoding by nanopore sequencing. *PLoS One*, 12(10), e0184741.

Menéndez-Arias, L. (2002). Molecular basis of fidelity of DNA synthesis and nucleotide specificity of retroviral reverse transcriptases.

Meng X, Clews J, Ciuta AD, Martin ER, Ford RC. (2019). CFTR structure, stability, function and regulation. *Biol Chem.* 25;400(10):1359-1370. doi: 10.1515/hsz-2018-0470.

Merriman, B., R&D Team, I. T., & Rothberg, J. M. (2012). Progress in ion torrent semiconductor chip based sequencing. *Electrophoresis*, 33(23), 3397-3417.

Miserocchi, G. I. U. S. E. P. P. E., Poskurica, B. H., & Del Fabbro, M. A. S. S. I. M. O. (1994). Pulmonary interstitial pressure in anesthetized paralyzed newborn rabbits. *Journal of Applied Physiology*, 77(5), 2260-2268.

Miserocchi, G., Passi, A., Negrini, D., Del Fabbro, M., & De Luca, G. (2001). Pulmonary interstitial pressure and tissue matrix structure in acute hypoxia. *American Journal of Physiology-Lung Cellular and Molecular Physiology*, 280(5), L881-L887.

Mitra, R. D., Shendure, J., Olejnik, J., & Church, G. M. (2003). Fluorescent in situ sequencing on polymerase colonies. *Analytical biochemistry*, 320(1), 55-65.

Mogk, A., Ruger-Herreros, C., & Bukau, B. (2019). Cellular functions and mechanisms of action of small heat shock proteins. *Annual review of microbiology*, 73, 89-110.

Molinski, S. V., Ahmadi, S., Ip, W., Ouyang, H., Villella, A., Miller, J. P., ... & Bear, C. E. (2017). O rkambi® and amplifier co-therapy improves function from a rare CFTR mutation in gene-edited cells and patient tissue. *EMBO molecular medicine*, 9(9), 1224-1243.

Montgomery, S. B., Lappalainen, T., Gutierrez-Arcelus, M., & Dermitzakis, E. T. (2011). Rare and common regulatory variation in population-scale sequenced human genomes. *PLoS genetics*, 7(7), e1002144.

Montgomery, S.B. et al. (2013) The origin, evolution, and functional impact of short insertion-deletion variants identified in 179 human genomes. *Genome Res.* 23, 749–761

Morin R et al (2008) Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. *Biotechniques* 45:81–94

Morozova, O., Hirst, M., & Marra, M. A. (2009). Applications of new sequencing technologies for transcriptome analysis. *Annual review of genomics and human genetics*, 10, 135-151.

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., & Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature methods*, 5(7), 621-628.

Moskowitz, S. M., Chmiel, J. F., Stern, D. L., Cheng, E., Gibson, R. L., Marshall, S. G., & Cutting, G. R. (2008). Clinical practice and genetic counseling for cystic fibrosis and CFTR-related disorders. *Genetics in Medicine*, 10(12), 851-868.

Mouchel, N., Brookes-Carter, F., & Harris, A. (2003). Alternative 5' exons of the CFTR gene show developmental regulation. *Human molecular genetics*, 12(7), 759-769.

Munguía-Reyes, A., Balderas-Martínez, Y.I., Becerril, C., Checa, M., Ramírez, R., Ortiz, B., Meléndez-Zajgla, J., Pardo, A., Selman, M. (2018). R-Spondin-2 Is Upregulated in Idiopathic Pulmonary Fibrosis and Affects Fibroblast Behavior. *Am J Respir Cell Mol Biol.* 59(1):65-76.

Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y., & Honjo, T. (2000). Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell*, 102(5), 553-563.

Na, C.H., et al. (2019). Integrated Transcriptomic and Proteomic Analysis of Human Eccrine Sweat Glands Identifies Missing and Novel Proteins. *Mol Cell Proteomics*. 18(7):1382-1395. doi: 10.1074/mcp.RA118.001101

Nagalakshmi U et al. (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* 320:1344–1349

Nair, R. P., Stuart, P. E., Nistor, I., Hiremagalore, R., Chia, N. V., Jenisch, S., ... & Elder, J. T. (2006). Sequence and haplotype analysis supports HLA-C as the psoriasis susceptibility 1 gene. *The American Journal of Human Genetics*, 78(5), 827-851.

Naphade, S., Bhatnagar, R., Hanson-Smith, V., Choi, I., & Zhang, A. (2022). Systematic comparative analysis of strand-specific RNA-seq library preparation methods for low input samples. *Scientific reports*, 12(1), 1-10.

Nature Publishing Group; 2015 [cited 2019 May 24];17:405–23. Available from: <http://www.nature.com/articles/gim201530>

Nesta, A. V., Tafur, D., & Beck, C. R. (2021). Hotspots of human mutation. *Trends in Genetics*, 37(8), 717-729.

Nichols, D. P., & Chmiel, J. F. (2015). Inflammation and its genesis in cystic fibrosis. *Pediatric pulmonology*, 50(S40), S39-S56.

Nielsen, R., Paul, J. S., Albrechtsen, A., & Song, Y. S. (2011). Genotype and SNP calling from next-generation sequencing data. *Nature Reviews Genetics*, 12(6), 443-451..

Norris, R. A., Damon, B., Mironov, V., Kasyanov, V., Ramamurthi, A., Moreno-Rodriguez, R., ... & Markwald, R. R. (2007). Periostin regulates collagen fibrillogenesis and the biomechanical properties of connective tissues. *Journal of cellular biochemistry*, 101(3), 695-711.

O'dwyer, D. N., & Moore, B. B. (2017). The role of periostin in lung fibrosis and airway remodeling. *Cellular and molecular life sciences*, 74(23), 4305-4314.

Ott, C. J., Blackledge, N. P., Kerschner, J. L., Leir, S. H., Crawford, G. E., Cotton, C. U., & Harris, A. (2009). Intronic enhancers coordinate epithelial-specific looping of the active CFTR locus. *Proceedings of the National Academy of Sciences*, 106(47), 19934-19939.

Ozes, A. R., & Nephew, K. P. (2013). 3D culture adds an extra dimension to targeted epigenetic therapies. *Cell Cycle*, 12(14), 2173-2174.

Pagani, F., Raponi, M., & Baralle, F. E. (2005). Synonymous mutations in CFTR exon 12 affect splicing and are not neutral in evolution. *Proceedings of the National Academy of Sciences*, 102(18), 6368-6372.

Palomares, M. A., Dalmasso, C., Bonnet, E., Derbois, C., Brohard-Julien, S., Ambroise, C., ... & Olasso, R. (2019). Systematic analysis of TruSeq, SMARTer and SMARTer Ultra-Low RNA-seq kits for standard, low and ultra-low quantity samples. *Scientific reports*, 9(1), 1-12.

Parker, D and Prince A. (2011). Type I interferon response to extracellular bacteria in the airway epithelium. *Trends Immunol.* 2011; 32(12):582–8.

Parker, D., Cohen, T. S., Alhede, M., Harfenist, B. S., Martin, F. J., & Prince, A. (2012). Induction of type I interferon signaling by *Pseudomonas aeruginosa* is diminished in cystic fibrosis epithelial cells. *American journal of respiratory cell and molecular biology*, 46(1), 6-13.

Parker, D., Martin, F. J., Soong, G., Harfenist, B. S., Aguilar, J. L., Ratner, A. J., ... & Prince, A. (2011). *Streptococcus pneumoniae* DNA initiates type I interferon signaling in the respiratory tract. *MBio*, 2(3), e00016-11.

Parkinson, J., & Blaxter, M. (2009). Expressed sequence tags: an overview. *Expressed sequence tags (ESTs)*, 1-12.

Parliament E. REGULATION (EC) No 141/2000 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 16 December 1999 on orphan medicinal products. 1999.

Peng, Z., Cheng, Y., Tan, B. C. M., Kang, L., Tian, Z., Zhu, Y., ... & Wang, J. (2012). Comprehensive analysis of RNA-Seq data reveals extensive RNA editing in a human transcriptome. *Nature biotechnology*, 30(3), 253-260.

Penque, D., Mendes, F., Beck, S., Farinha, C., Pacheco, P., Nogueira, P., ... & Amaral, M. D. (2000). Cystic fibrosis F508del patients have apically localized CFTR in a reduced number of airway cells. *Laboratory investigation*, 80(6), 857-868.

Peters, J. (2014). The role of genomic imprinting in biology and disease: an expanding view. *Nature Reviews Genetics*, 15(8), 517-530.

Phylactides, M., Rowntree, R., Nuthall, H., Ussery, D., Wheeler, A., & Harris, A. (2002). Evaluation of potential regulatory elements identified as DNase I hypersensitive sites in the CFTR gene. *European journal of biochemistry*, 269(2), 553-559.

Pietu G, Mariage-Samon R, Fayein NA, Matingou C, Eveno E, et al. (1999) The Genexpress IMAGE knowledge base of the human brain transcriptome: A prototype integrated resource for functional and computational genomics. *Genome Res* 9: 195–209

Piskol, R., Ramaswami, G., and Li, J. B. (2013). Reliable identification of genomic variants from RNA-Seq data. *Am. J. Hum. Genet.* 93, 641–651. doi: 10.1016/j.ajhg.2013.08.008

Poller, W., Barth, J., Voss, B. (1989) Detection of an alteration of the alpha-2-macroglobulin gene in a patient with chronic lung disease and serum alpha-2-macroglobulin deficiency. *Hum. Genet.* 83: 93-96, 1989.

Poller, W., Faber, J.-P., Klobeck, G., Olek, K. (1992). Cloning of the human alpha-2-macroglobulin gene and detection of mutations in two functional domains: the bait region and the thiolester site. *Hum. Genet.* 88: 313-319.

Pollex, R. L., & Hegele, R. A. (2004). Hutchinson–Gilford progeria syndrome. *Clinical genetics*, 66(5), 375-381.

Posey, J. E., Harel, T., Liu, P., Rosenfeld, J. A., James, R. A., Coban Akdemir, Z. H., ... & Lupski, J. R. (2017). Resolution of disease phenotypes resulting from multilocus genomic variation. *New England Journal of Medicine*, 376(1), 21-31.

Posey, J. E., Rosenfeld, J. A., James, R. A., Bainbridge, M., Niu, Z., Wang, X., ... & Plon, S. E. (2016). Molecular diagnostic experience of whole-exome sequencing in adult patients. *Genetics in Medicine*, 18(7), 678-685.

Pucker, B., Holtgräwe, D., Stadermann, K. B., Frey, K., Huettel, B., Reinhardt, R., & Weisshaar, B. (2019). A chromosome-level sequence assembly reveals the

structure of the *Arabidopsis thaliana* Nd-1 genome and its gene set. *PloS one*, 14(5), e0216233.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., ... & Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American journal of human genetics*, 81(3), 559-575.

Putney, S.D., Herlihy, W.C. and Schimmel, P. (1983) A new troponin T and cDNA clones for 13 different muscle proteins, found by shotgun sequencing. *Nature*, 302, 718–721.

Quaresma, M. C., Pankonien, I., Clarke, L. A., Sousa, L. S., Silva, I. A., Railean, V., ... & Amaral, M. D. (2020). Mutant CFTR Drives TWIST1 mediated epithelial–mesenchymal transition. *Cell death & disease*, 11(10), 1-18.

Quinton, P. M. (1990). Cystic fibrosis: a disease in electrolyte transport. *The FASEB Journal*, 4(10), 2709-2710.

Quinton, P. M., & Bijman, J. (1983). Higher bioelectric potentials due to decreased chloride absorption in the sweat glands of patients with cystic fibrosis. *New England Journal of Medicine*, 308(20), 1185-1189.

Radenbaugh, A. J., Ma, S., Ewing, A., Stuart, J. M., Collisson, E. A., Zhu, J., & Haussler, D. (2014). RADIA: RNA and DNA integrated analysis for somatic mutation detection. *PloS one*, 9(11), e111516.

Rafehi, H., Szmulewicz, D. J., Bennett, M. F., Sobreira, N. L., Pope, K., Smith, K. R., ... & Lockhart, P. J. (2019). Bioinformatics-based identification of expanded repeats: a non-reference intronic pentamer expansion in RFC1 causes CANVAS. *The American Journal of Human Genetics*, 105(1), 151-165.

Rahit, K. M., & Tarailo-Graovac, M. (2020). Genetic modifiers and rare mendelian disease. *Genes*, 11(3), 239.

Ramalho, A. S., Beck, S., Meyer, M., Penque, D., Cutting, G. R., & Amaral, M. D. (2002). Five percent of normal cystic fibrosis transmembrane conductance

regulator mRNA ameliorates the severity of pulmonary disease in cystic fibrosis. *American journal of respiratory cell and molecular biology*, 27(5), 619-627.

Ramalho, A. S., Fürstová, E., Vonk, A. M., Ferrante, M., Verfaillie, C., Dupont, L., ... & De Boeck, K. (2021). Correction of CFTR function in intestinal organoids to guide treatment of cystic fibrosis. *European Respiratory Journal*, 57(1).

Ramalho, A. S., Lewandowska, M. A., Farinha, C. M., Mendes, F., Gonçalves, J., Barreto, C., ... & Amaral, M. D. (2009). Deletion of CFTR translation start site reveals functional isoforms of the protein in CF patients. *Cellular Physiology and Biochemistry*, 24(5-6), 335-346.

Ratjen F.; Bell, S.C.; Rowe, S.M.; Goss, C.H.; Quittner, A.L.; Bush, A.(2015), Cystic fibrosis. *Nat. Rev. Dis. Prim.*, 1, 15010.

Rehm, H. L., Bale, S. J., Bayrak-Toydemir, P., Berg, J. S., Brown, K. K., Deignan, J. L., ... & Lyon, E. (2013). ACMG clinical laboratory standards for next-generation sequencing. *Genetics in medicine*, 15(9), 733-747.

Rentas, S., Rathi, K. S., Kaur, M., Raman, P., Krantz, I. D., Sarmady, M., & Abou Tayoun, A. (2020). Diagnosing Cornelia de Lange syndrome and related neurodevelopmental disorders using RNA sequencing. *Genetics in Medicine*, 22(5), 927-936.

Retterer, K., Juusola, J., Cho, M. T., Vitazka, P., Millan, F., Gibellini, F., ... & Bale, S. (2016). Clinical application of whole-exome sequencing across clinical indications. *Genetics in Medicine*, 18(7), 696-704.

Reuter, J.A., Spacek, D.V., Snyder, M.P., (2015). High-throughput sequencing technologies. *Mol. Cell* 58, 586–597.

Ribatti, D., & Santoiemma, M. (2014). Epithelial-mesenchymal interactions: a fundamental developmental biology mechanism. *International Journal of Developmental Biology*, 58(5), 303-306.

Rich, D. P., Anderson, M. P., Gregory, R. J., Cheng, S. H., Paul, S., Jefferson, D. M., ... & Welsh, M. J. (1990). Expression of cystic fibrosis transmembrane

conductance regulator corrects defective chloride channel regulation in cystic fibrosis airway epithelial cells. *Nature*, 347(6291), 358-363.

Richards, C. S., Bale, S., Bellissimo, D. B., Das, S., Grody, W. W., Hegde, M. R., ... & Ward, B. E. (2008). ACMG recommendations for standards for interpretation and reporting of sequence variations: Revisions 2007. *Genetics in Medicine*, 10(4), 294-300.

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., ... & Rehm, H. L. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in medicine*, 17(5), 405-423.

Riordan, J. R. (2008). CFTR function and prospects for therapy. *Annu. Rev. Biochem.*, 77, 701-726.

Riordan, J. R., Rommens, J. M., Kerem, B. S., Alon, N. O. A., Rozmahel, R., Grzelczak, Z., ... & Tsui, L. C. (1989). Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science*, 245(4922), 1066-1073.

Risoo, D. EDASeq: Exploratory Data Analysis and Normalization for RNA-Seq Modified: May 22, 2012. Compiled: April 4, 2013

Ritchie, M. E., Phipson, B., Wu, D. I., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research*, 43(7), e47-e47..

Roberts, A., Trapnell, C., Donaghey, J., Rinn, J. L. and Pachter, L. (2011) Improving RNA-Seq expression estimates by correcting for fragment bias. *Genome biology*, 12, R22.

Robinson, E., MacDonald, K. D., Slaughter, K., McKinney, M., Patel, S., Sun, C., & Sahay, G. (2018). Lipid nanoparticle-delivered chemically modified mRNA restores chloride secretion in cystic fibrosis. *Molecular Therapy*, 26(8), 2034-2046.

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), 139-140.

Rode, J. (2005). Rare diseases: understanding this public health priority. *Rare Dis*, 14, 3.

Rommens, J. M., Iannuzzi, M. C., Kerem, B. S., Drumm, M. L., Melmer, G., Dean, M., ... & Collins, F. S. (1989). Identification of the cystic fibrosis gene: chromosome walking and jumping. *Science*, 245(4922), 1059-1065.

Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlén, M., & Nyrén, P. (1996). Real-time DNA sequencing using detection of pyrophosphate release. *Analytical biochemistry*, 242(1), 84-89.

Ronaghi, M., Uhlén, M., & Nyrén, P. (1998). A sequencing method based on real-time pyrophosphate. *Science*, 281(5375), 363-365.

Rose-John, S., Scheller, J., Elson, G., & Jones, S. A. (2006). Interleukin-6 biology is coordinated by membrane-bound and soluble receptors: role in inflammation and cancer. *Journal of leukocyte biology*, 80(2), 227-236.

Rosenberg A., Roco C, Muscat R, Kuchina A, Sample P, Yao Z et al (2018) Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science* 360:176–182

Rosenstein, B. J., & Cutting, G. R. (1998). The diagnosis of cystic fibrosis: a consensus statement. *The Journal of pediatrics*, 132(4), 589-595.:

Rothberg, J. M., Hinz, W., Rearick, T. M., Schultz, J., Mileski, W., Davey, M., ... & Bustillo, J. (2011). An integrated semiconductor device enabling non-optical genome sequencing. *Nature*, 475(7356), 348-352.

Rout-Pitt, N., Farrow, N., Parsons, D., & Donnelley, M. (2018). Epithelial mesenchymal transition (EMT): a universal process in lung diseases with implications for cystic fibrosis pathophysiology. *Respiratory Research*, 19(1), 1-10.

Rowe, S. M. (2005). Miller S, Sorscher EJ. *Cystic fibrosis. N Engl J Med*, 352, 1992-2001.

Rubenstein, R. C., & Zeitlin, P. L. (1998). Use of protein repair therapy in the treatment of cystic fibrosis. *Current opinion in pediatrics*, 10(3), 250-255.

Ryley HC, Goodchild MC, Dodge JA (1992) Screening for cystic fibrosis. *Br Med Bull* 48 (4):805–822).

Sá, A. C. C., Sadee, W., & Johnson, J. A. (2018). Whole transcriptome profiling: An RNA-Seq primer and implications for pharmacogenomics research. *Clinical and translational science*, 11(2), 153.

Sakharkar, M. K., Chow, V. T., & Kanguane, P. (2004). Distributions of exons and introns in the human genome. *In silico biology*, 4(4), 387-393.

Salinas, D., Haggie, P. M., Thiagarajah, J. R., Song, Y., Rosbe, K., Finkbeiner, W. E., ... & Verkman, A. S. (2005). Submucosal gland dysfunction as a primary defect in cystic fibrosis. *The FASEB journal*, 19(3), 1-13.

Salton, F., Volpe, M. C., & Confalonieri, M. (2019). Epithelial–mesenchymal transition in the pathogenesis of idiopathic pulmonary fibrosis. *Medicina*, 55(4), 83.

Sanger, F. (1988). Sequences, sequences, and sequences. *Annual review of biochemistry*, 57(1), 1-29.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the national academy of sciences*, 74(12), 5463-5467.

Sarantopoulou, D., Tang, S. Y., Ricciotti, E., Lahens, N. F., Lekkas, D., Schug, J., ... & Grant, G. R. (2019). Comparative evaluation of RNA-Seq library preparation methods for strand-specificity and low input. *Scientific reports*, 9(1), 1-10.

Sasani, T. A., Pedersen, B. S., Gao, Z., Baird, L., Przeworski, M., Jorde, L. B., & Quinlan, A. R. (2019). Large, three-generation human families reveal post-zygotic mosaicism and variability in germline mutation accumulation. *Elife*, 8, e46922.

Sawyer, S. L., Hartley, T., Dymont, D. A., Beaulieu, C. L., Schwartzentruber, J., Smith, A., ... & Boycott, K. M. (2016). Utility of whole-exome sequencing for those near the end of the diagnostic odyssey: time to address gaps in care. *Clinical genetics*, 89(3), 275-284.

Schadt EE., Turner S, Kasarskis A (2010) A window into thirdgeneration sequencing. *Hum Mol Genet* 19:R227–R240

Scheller, J., Chalaris, A., Schmidt-Arras, D., & Rose-John, S. (2011). The pro- and anti-inflammatory properties of the cytokine interleukin-6. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, 1813(5), 878-888.

Schena, M., Shalon, D., Davis, R. W., & Brown, P. O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, 270(5235), 467-470.

Schieppati, A., Henter, J. I., Daina, E., & Aperia, A. (2008). Why rare diseases are an important medical and social issue. *The Lancet*, 371(9629), 2039-2041.

Schmidt, B. Z., Watts, R. J., Aridor, M., & Frizzell, R. A. (2009). Cysteine string protein promotes proteasomal degradation of the cystic fibrosis transmembrane conductance regulator (CFTR) by increasing its interaction with the C terminus of Hsp70-interacting protein and promoting CFTR ubiquitylation. *Journal of Biological Chemistry*, 284(7), 4168-4178.

Scott, M., & De Sario, A. (2020). DNA methylation changes in cystic fibrosis: Cause or consequence?. *Clinical Genetics*, 98(1), 3-9.

Scott-Ward, T.S.; Amaral, M.D. Deletion of Phe508 in the first nucleotide-binding domain of the cystic fibrosis transmembrane conductance regulator increases its affinity for the heat shock cognate 70 chaperone. *FEBS J.* 2009, 276, 7097–7109. [CrossRef] [PubMed],

Selman, M., & Pardo, A. (2014). Stochastic age-related epigenetic drift in the pathogenesis of idiopathic pulmonary fibrosis. *American journal of respiratory and critical care medicine*, 190(12), 1328-1330.

Shah, V. S., Meyerholz, D. K., Tang, X. X., Reznikov, L., Abou Alaiwa, M., Ernst, S. E., ... & Welsh, M. J. (2016). Airway acidification initiates host defense abnormalities in cystic fibrosis mice. *Science*, 351(6272), 503-507.

Shen, Y., Wan, Z., Coarfa, C., Drabek, R., Chen, L., Ostrowski, E. A., ... & Yu, F. (2010). A SNP discovery method to assess variant allele probability from next-generation resequencing data. *Genome research*, 20(2), 273-280.

Shendure, J. and Akey, J.M. (2015) The origins, determinants, and consequences of human mutations. *Science* 349, 1478–1483

Shendure, J., Balasubramanian, S., Church, G. M., Gilbert, W., Rogers, J., Schloss, J. A., & Waterston, R. H. (2017). DNA sequencing at 40: past, present and future. *Nature*, 550(7676), 345-353.

Shendure, J., Porreca, G. J., Reppas, N. B., Lin, X., McCutcheon, J. P., Rosenbaum, A. M., ... & Church, G. M. (2005). Accurate multiplex polony sequencing of an evolved bacterial genome. *Science*, 309(5741), 1728-1732.

Sherman, R. M. (2021). *Discovering novel human structural variation from diverse populations and disease patients: an exploration of what human genomics misses by relying on reference-based analyses* (Doctoral dissertation, Johns Hopkins University). Shi, Y. (2012) Alternative polyadenylation: new insights from global analyses. *RNA* (New York, N.Y.), 18, 2105–2117.

Silva, I. A., Doušová, T., Ramalho, S., Centeio, R., Clarke, L. A., Railean, V., ... & Amaral, M. D. (2020). Organoids as a personalized medicine tool for ultra-rare mutations in cystic fibrosis: The case of S955P and 1717-2A> G. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, 1866(11), 165905.

Sim, G.K., Kafatos, F.C., Jones, C.W., Koehler, M.D., Efstratiadis, A. and Maniatis, T. (1979) Use of a cDNA library for studies on evolution and developmental expression of the chorion multigene families. *Cell*, 18, 1303–1316.

Sims, D., Sudbery, I., Illott, N. E., Heger, A., and Ponting, C. P. (2014). Sequencing depth and coverage: key considerations in genomic analyses. *Nat. Rev. Genet.* 15, 121–132. doi: 10.1038/nrg3642

Singh, R. K., & Cooper, T. A. (2012). Pre-mRNA splicing in disease and therapeutics. *Trends in molecular medicine*, 18(8), 472-482.

Smeekeens, S. P., Ng, A., Kumar, V., Johnson, M. D., Plantinga, T. S., Van Diemen, C., ... & Xavier, R. J. (2013). Functional genomics identifies type I interferon pathway as central for host defense against *Candida albicans*. *Nature communications*, 4(1), 1-10.

Soneson, C., Yao, Y., Bratus-Neuenschwander, A., Patrignani, A., Robinson, M. D., & Hussain, S. (2019). A comprehensive examination of Nanopore native RNA sequencing for characterization of complex transcriptomes. *Nature communications*, 10(1), 1-14.

Sosnay, P. R., Siklosi, K. R., Van Goor, F., Kaniecki, K., Yu, H., Sharma, N., ... & Cutting, G. R. (2013). Defining the disease liability of variants in the cystic fibrosis transmembrane conductance regulator gene. *Nature genetics*, 45(10), 1160-1167.

Steijger, T., Abril, J. F., Engström, P. G., Kokocinski, F., Hubbard, T. J., Guigó, R., ... & Bertone, P. (2013). Assessment of transcript reconstruction methods for RNA-seq. *Nature methods*, 10(12), 1177-1184.

Stonebraker, J, Bolton-Maggs P, Michael Soucie J, Walker I, Brooker M. (2010). A study of variations in the reported haemophilia A prevalence around the world. *Haemophilia*. 16(1):20-32,

Strickland, E., Qu, B. H., Millen, L., & Thomas, P. J. (1997). The Molecular Chaperone Hsc70 Assists the in Vitro Folding of the N-terminal Nucleotide-binding Domain of the Cystic Fibrosis Transmembrane Conductance Regulator. *Journal of Biological Chemistry*, 272(41), 25421-25424.

Su, X., Looney, M. R., Su, H. E., Lee, J. W., Song, Y., & Matthay, M. A. (2011). Role of CFTR expressed by neutrophils in modulating acute lung inflammation and injury in mice. *Inflammation research*, 60(7), 619-632.

Sutcliffe, J. G., Milner, R. J., Bloom, F. E., & Lerner, R. A. (1982). Common 82-nucleotide sequence unique to brain RNA. *Proceedings of the National Academy of Sciences*, 79(16), 4942-4946.

Szafranski, P., Coban-Akdemir, Z. H., Rupps, R., Grazioli, S., Wensley, D., Jhangiani, S. N., ... & Stankiewicz, P. (2016). Phenotypic expansion of TBX4 mutations to include acinar dysplasia of the lungs. *American journal of medical genetics Part A*, 170(9), 2440-2444.

Tankard, R. M., Bennett, M. F., Degorski, P., Delatycki, M. B., Lockhart, P. J., & Bahlo, M. (2018). Detecting expansions of tandem repeats in cohorts sequenced with short-read sequencing data. *The American Journal of Human Genetics*, 103(6), 858-873.

Tanner, S. M., Laporte, J., Guiraud-Chaumeil, C., & Liechti-Gallati, S. (1998). Confirmation of prenatal diagnosis results of X-linked recessive myotubular myopathy by mutational screening, and description of three new mutations in the MTM1 gene. *Human mutation*, 11(1), 62-68.

Tariq, M. A., Kim, H. J., Jejelowo, O., & Pourmand, N. (2011). Whole-transcriptome RNAseq analysis from minute amount of total RNA. *Nucleic acids research*, 39(18), e120-e120.

Taubes, G. (1996). Misfolding the Way to Disease: Amyloid diseases such as Alzheimer's may result when proteins fold up incorrectly, causing them to aggregate and deposit abnormally in or around cells. *Science*, 271(5255), 1493-1495.

Tazi, J., Bakkour, N., & Stamm, S. (2009). Alternative splicing and disease. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, 1792(1), 14-26.

Temin, H.M. and Mizutani, S. (1970) RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature*, 226, 1211–1213.

Thodeson, D. M., & Park, J. Y. (2019). Genomic testing in pediatric epilepsy. *Molecular Case Studies*, 5(4), a004135..

Thomson, J. M., Newman, M., Parker, J. S., Morin-Kensicki, E. M., Wright, T., & Hammond, S. M. (2006). Extensive post-transcriptional regulation of microRNAs and its implications for cancer. *Genes&development*, 20(16), 2202-2207.

Toung, J. M., Lahens, N., Hogenesch, J. B., & Grant, G. (2014). Detection theory in identification of RNA-DNA sequence differences using RNA-sequencing. *PLoS One*, 9(11), e112040.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., ... & Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature protocols*, 7(3), 562-578.

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M. J., ... & Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature biotechnology*, 28(5), 511-515.

Trouvé, P., Génin, E., & Férec, C. (2017). In silico search for modifier genes associated with pancreatic and liver disease in Cystic Fibrosis. *PloS one*, 12(3), e0173822.

Valamparampil, J. J., & Gupte, G. L. (2021). Cystic fibrosis associated liver disease in children. *World Journal of Hepatology*, 13(11), 1727.

Valdivieso, Á. G., Clauzure, M., Massip-Copiz, M., & Santa-Coloma, T. A. (2016). The chloride anion acts as a second messenger in mammalian cells-modifying the expression of specific genes. *Cellular Physiology and Biochemistry*, 38(1), 49-64.

van der Plaat, D. A., de Jong, K., Lahousse, L., Faiz, A., Vonk, J. M., van Diemen, C. C., ... & Boezen, H. M. (2017). Genome-wide association study on the FEV1/FVC ratio in never-smokers identifies HHIP and FAM13A. *Journal of Allergy and Clinical Immunology*, 139(2), 533-540.

Van Dijk, E. L., Jaszczyszyn, Y., & Thermes, C. (2014). Library preparation methods for next-generation sequencing: tone down the bias. *Experimental cell research*, 322(1), 12-20.

van Dijk,E.L., Jaszczyszyn,Y. and Thermes,C. (2014) Library preparation methods for nextgeneration sequencing: tone down the bias. *Exp. Cell Res.*, 322, 12–20.

van Gurp,T.P., McIntyre,L.M. and Verhoeven,K.J.F. (2013) Consistent Errors in First Strand cDNA Due to Random Hexamer Mispriming. *PLoS ONE*, 8, e85583.

Veit, G., Avramescu, R. G., Chiang, A. N., Houck, S. A., Cai, Z., Peters, K. W., ... & Lukacs, G. L. (2016). From CFTR biology toward combinatorial pharmacotherapy: expanded classification of cystic fibrosis mutations. *Molecular biology of the cell*, 27(3), 424-433.

Veit, G., Oliver, K., Apaja, P. M., Perdomo, D., Bidaud-Meynard, A., Lin, S. T., ... & Lukacs, G. L. (2016). Ribosomal stalk protein silencing partially corrects the Δ F508-CFTR functional expression defect. *PLoS biology*, 14(5), e1002462.

Velculescu, V. E., Zhang, L., Vogelstein, B., & Kinzler, K. W. (1995). Serial analysis of gene expression. *Science*, 270(5235), 484-487.

Vergani, P., Lockless, S. W., Nairn, A. C., & Gadsby, D. C. (2005). CFTR channel opening by ATP-driven tight dimerization of its nucleotide-binding domains. *Nature*, 433(7028), 876-880.

Verhaeghe, C., Remouchamps, C., Hennuy, B., Vanderplasschen, A., Chariot, A., Tabruyn, S. P., ... & Bours, V. (2007). Role of IKK and ERK pathways in intrinsic inflammation of cystic fibrosis airways. *Biochemical pharmacology*, 73(12), 1982-1994.

Vieira, N. M., Elvers, I., Alexander, M. S., Moreira, Y. B., Eran, A., Gomes, J. P., ... & Zatz, M. (2015). Jagged 1 rescues the Duchenne muscular dystrophy phenotype. *Cell*, 163(5), 1204-1213.

Vigliar, E., Malapelle, U., de Luca, C., Bellevicine, C., Troncone, G., (2015a). Challenges and opportunities of next-generation sequencing: a cytopathologist's perspective. *Cytopathology* 26, 271–283.

Vigliar, E., Malapelle, U., de Luca, C., Bellevicine, C., Troncone, G., (2015b). Challenges and opportunities of next-generation sequencing: a cytopathologist's perspective. *Cytopathology* 26, 271–283.

Virella-Lowell, I., Herlihy, J. D., Liu, B., Lopez, C., Cruz, P., Muller, C., ... & Flotte, T. R. (2004). Effects of CFTR, interleukin-10, and *Pseudomonas aeruginosa* on gene expression profiles in a CF bronchial epithelial cell Line. *Molecular Therapy*, 10(3), 562-573.

Wang, C., Gong, B., Bushel, P. R., Thierry-Mieg, J., Thierry-Mieg, D., Xu, J., ... & Tong, W. (2014). A comprehensive study design reveals treatment-and transcript abundance-dependent concordance between rna-seq and microarray data. *Nature biotechnology*, 32(9), 926.

Wang, E. T., Sandberg, R., Luo, S., Khrebtkova, I., Zhang, L., Mayr, C., ... & Burge, C. B. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature*, 456(7221), 470-476.

Wang, K., Li, M., & Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic acids research*, 38(16), e164-e164.

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews genetics*, 10(1), 57-63.

Wassink, T. H., Piven, J., Vieland, V. J., Huang, J., Swiderski, R. E., Pietila, J., ... & Sheffield, V. C. (2001). Evidence supporting WNT2 as an autism susceptibility gene. *American journal of medical genetics*, 105(5), 406-413.

Watson, J. D., & Crick, F. H. (1953). Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature*, 171(4356), 737-738.

Weatherall, D. J., & Clegg, J. B. (2008). *The thalassaemia syndromes*. John Wiley & Sons.

Wee, Y., Bhyan, S. B., Liu, Y., Lu, J., Li, X., & Zhao, M. (2019). The bioinformatics tools for the genome assembly and analysis based on third-generation sequencing. *Briefings in functional genomics*, 18(1), 1-12.

Welsh MJ, Ramsey BW, Accurso F, Cutting GR, Scriver CL, Beaudet AL, Sly WS, Valle D (2001) Cystic Fibrosis. The molecular and metabolic basis of inherited disease. *McGraw-Hill, New York*, pp 5121–5188

Welsh, M. J., & Smith, A. E. (1993). Molecular mechanisms of CFTR chloride channel dysfunction in cystic fibrosis. *Cell*, 73(7), 1251-1254.

Wheeler, D. A., Srinivasan, M., Egholm, M., Shen, Y., Chen, L., McGuire, A., ... & Rothberg, J. M. (2008). The complete genome of an individual by massively parallel DNA sequencing. *nature*, 452(7189), 872-876.

Wilschanski, M., Zielenski, J., Markiewicz, D., Tsui, L. C., Corey, M., Levison, H., & Durie, P. R. (1995). Correlation of sweat chloride concentration with classes of the cystic fibrosis transmembrane conductance regulator gene mutations. *The Journal of pediatrics*, 127(5), 705-710.

Wright, C. F., FitzPatrick, D. R., & Firth, H. V. (2018). Paediatric genomics: diagnosing rare disease in children. *Nature Reviews Genetics*, 19(5), 253-268.

Wright, G. E., Caron, N. S., Ng, B., Casal, L., Casazza, W., Xu, X., ... & Hayden, M. R. (2020). Gene expression profiles complement the analysis of genomic modifiers of the clinical onset of Huntington disease. *Human molecular genetics*, 29(16), 2788-2802.

Wright, J. M., Merlo, C. A., Reynolds, J. B., Zeitlin, P. L., Garcia, J. G., Guggino, W. B., & Boyle, M. P. (2006). Respiratory epithelial gene expression in patients with mild and severe cystic fibrosis lung disease. *American journal of respiratory cell and molecular biology*, 35(3), 327-336.

Wu, S., Peng, Y., Hu, L., Zhang, X., & Li, H. (2016, January). Classification and recognition of texture collagen obtaining by multiphoton microscope with neural network analysis. In *Journal of Physics: Conference Series* (Vol. 680, No. 1, p. 012014). IOP Publishing.

Xie, T., Liang, J., Liu, N., Huan, C., Zhang, Y., Liu, W., ... & Noble, P. W. (2016). Transcription factor TBX4 regulates myofibroblast accumulation and lung fibrosis. *The Journal of clinical investigation*, 126(8), 3063-3079.

Xu, Z., & Chen, C. (2022). The downregulation of PTGS2 mediated ncRNAs by is tightly correlated with systemic sclerosis-interstitial lung disease. *Frontiers in Genetics*, 2797.

Y. Chen, D. Mccarthy, M. Ritchie, M. Robinson, G.K. Smyth, EdgeR: differential expression analysis of digital gene expression data user's guide,.

Yamamoto, Y. H., Kimura, T., Momohara, S., Takeuchi, M., Tani, T., Kimata, Y., ... & Kohno, K. (2010). A novel ER J-protein DNAJB12 accelerates ER-associated degradation of membrane proteins including CFTR. *Cell structure and function*, 35(2), 107-116.

Yang, Y., Janich, S., Cohn, J. A., & Wilson, J. M. (1993). The common variant of cystic fibrosis transmembrane conductance regulator is recognized by hsp70 and degraded in a pre-Golgi nonlysosomal compartment. *Proceedings of the National Academy of Sciences*, 90(20), 9480-9484.

Yang, Y., Muzny, D. M., Xia, F., Niu, Z., Person, R., Ding, Y., ... & Eng, C. M. (2014). Molecular findings among patients referred for clinical whole-exome sequencing. *Jama*, 312(18), 1870-1879.

Yavarna, T., Al-Dewik, N., Al-Mureikhi, M., Ali, R., Al-Mesaifri, F., Mahmoud, L., ... & Ben-Omran, T. (2015). High diagnostic yield of clinical exome sequencing in Middle Eastern patients with Mendelian disorders. *Human genetics*, 134(9), 967-980.

Yeh, H. I., Sohma, Y., Conrath, K., & Hwang, T. C. (2017). A common mechanism for CFTR potentiators. *Journal of General Physiology*, 149(12), 1105-1118.

Yi, H., Cho, Y. J., Won, S., Lee, J. E., Jin Yu, H., Kim, S., ... & Chun, J. (2011). Duplex-specific nuclease efficiently removes rRNA for prokaryotic RNA-seq. *Nucleic acids research*, 39(20), e140-e140.

Yohe, S., & Thyagarajan, B. (2017). Review of clinical next-generation sequencing. *Archives of pathology & laboratory medicine*, 141(11), 1544-1557.

Yoshimura, K., Nakamura, H., Trapnell, B. C., Dalemans, W., Pavirani, A., Lecocq, J. P., & Crystal, R. G. (1991). The cystic fibrosis gene has a “housekeeping”-type promoter and is expressed at low levels in cells of epithelial origin. *Journal of Biological Chemistry*, 266(14), 9140-9144.

Youker, R. T., Walsh, P., Beilharz, T., Lithgow, T., & Brodsky, J. L. (2004). Distinct roles for the Hsp40 and Hsp90 molecular chaperones during cystic fibrosis transmembrane conductance regulator degradation in yeast. *Molecular biology of the cell*, 15(11), 4787-4797.

Younger, J. M., Ren, H. Y., Chen, L., Fan, C. Y., Fields, A., Patterson, C., & Cyr, D. M. (2004). A foldable CFTR Δ F508 biogenic intermediate accumulates upon inhibition of the Hsc70–CHIP E3 ubiquitin ligase. *The Journal of cell biology*, 167(6), 1075-1085.

Zabner, J., Scheetz, T. E., Almabrazi, H. G., Casavant, T. L., Huang, J., Keshavjee, S., & McCray Jr, P. B. (2005). CFTR Δ F508 mutation has minimal effect on the gene expression profile of differentiated human airway epithelia. *American Journal of Physiology-Lung Cellular and Molecular Physiology*, 289(4), L545-L553.

Zaman, K., Sawczak, V., Zaidi, A., Butler, M., Bennett, D., Getsy, P., ... & Gaston, B. (2016). Augmentation of CFTR maturation by S-nitrosoglutathione reductase. *American Journal of Physiology-Lung Cellular and Molecular Physiology*, 310(3), L263-L270.

Zemanick, E. T., Konstan, M. W., VanDevanter, D. R., Rowe, S. M., Clancy, J. P., Odem-Davis, K., ... & Mayer-Hamblett, N. (2021). Measuring the impact of CFTR modulation on sweat chloride in cystic fibrosis: Rationale and design of the CHEC-SC study. *Journal of Cystic Fibrosis*, 20(6), 965-971.

Zhang Z., Wang, J. and Dong, X. (2018) Wnt2 contributes to the progression of gastric cancer by promoting cell migration and invasion. *Oncol. Lett.*, 16, 2857–2864.

Zhang ZY, Wang ZM, Huang Y. (2020). Polycystic liver disease: Classification, diagnosis, treatment process, and clinical management. *World J Hepatol*. 27;12(3):72-83. doi: 10.4254/wjh.v12.i3.72.

Zhang, H., Peters, K. W., Sun, F., Marino, C. R., Lang, J., Burgoyne, R. D., & Frizzell, R. A. (2002). Cysteine string protein interacts with and modulates the maturation of the cystic fibrosis transmembrane conductance regulator. *Journal of Biological Chemistry*, 277(32), 28948-28958.

Zhang, H., Schmidt, B. Z., Sun, F., Condliffe, S. B., Butterworth, M. B., Youker, R. T., ... & Frizzell, R. A. (2006). Cysteine string protein monitors late steps in cystic fibrosis transmembrane conductance regulator biogenesis. *Journal of Biological Chemistry*, 281(16), 11312-11321.

Zhang, M., Haughey, M., Wang, N. Y., Blease, K., Kapoun, A. M., Couto, S., ... & Xie, W. (2020). Targeting the Wnt signaling pathway through R-spondin 3 identifies an anti-fibrosis treatment strategy for multiple organs. *PloS one*, 15(3), e0229445.

Zhao, J., Akinsanmi, I., Arafat, D., Cradick, T. J., Lee, C. M., Banskota, S., ... & Gibson, G. (2016). A burden of rare variants associated with extremes of gene expression in human peripheral blood. *The American Journal of Human Genetics*, 98(2), 299-309.

Zhu, F., Wang, F., Yang, X., Zhang, J., Wu, H., Zhang, Z., ... & Cao, Y. (2016). Biallelic SUN5 mutations cause autosomal-recessive acephalic spermatozoa syndrome. *The American Journal of Human Genetics*, 99(4), 942-949.

Zielenski, J., & Tsui, L. C. (1995). Cystic fibrosis: genotypic and phenotypic variations. *Annual review of genetics*, 29(1), 777-807.

Zielenski, J., Rozmahel, R., Bozon, D., Kerem, B. S., Grzelczak, Z., Riordan, J. R., ... & Tsui, L. C. (1991). Genomic DNA sequence of the cystic fibrosis transmembrane conductance regulator (CFTR) gene. *Genomics*, 10(1), 214-228.

https://commons.wikimedia.org/wiki/File:2926_Autosomal_Recessive_Inheritance_new.jpg#/media/File:2926_Autosomal_Recessive_Inheritance-new.jpg,

https://ec.europa.eu/info/research-and-innovation/research-area/health-research-and-innovation/rare-diseases_en.

<https://www.pymol.org/>

https://github.com/hbctraining/DGE_workshop_salmon/blob/master/lessons/01_DGE_setup_and_overview.md

https://github.com/hbctraining/DGE_workshop_salmon/blob/master/lessons/04_DGE_DESeq2_analysis.md

https://github.com/hbctraining/DGE_workshop_salmon/blob/master/lessons/03_DGE_QC_analysis.md

https://hbctraining.github.io/DGE_workshop/lessons/05_DGE_DESeq2_analysis2.html

https://hbctraining.github.io/DGE_workshop/lessons/05_DGE_DESeq2_analysis2.html

https://github.com/hbctraining/DGE_workshop_salmon_online/blob/master/lessons/05b_wald_test_results.md

8. Özgeçmiş



9. İntihal Raporu

KİSTİK FİBROSİZ TEŞHİS VE TEDAVİSİNDE KULLANILACAK RNA SEKASINDA İLERİ VERİ ANALİZLERİ

ORJİNALLİK RAPORU			
%17	EN	%11	%14
BENZERLİK ENDEKSİ	İNTERNET KAYNAKLARI	YAYINLAR	ÖĞRENCİ ÖDEVLERİ
BİRİNCİL KAYNAKLAR			
1	journals.plos.org İnternet Kaynağı		%3
2	www.mdpi.com İnternet Kaynağı		%2
3	www.ncbi.nlm.nih.gov İnternet Kaynağı		%1
4	link.springer.com İnternet Kaynağı		%1
5	Anna Esteve-Codina. "RNA-Seq Data Analysis, Applications and Challenges", Elsevier BV, 2018 Yayın		%1
6	Claudia Gonzaga-Jauregui, Cinthya J. Zepeda Mendoza. "Genomic sequencing of rare diseases", Elsevier BV, 2021 Yayın		%1
7	d-scholarship.pitt.edu İnternet Kaynağı		%1