

T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

MEME KANSERİ YENİ NESİL DİZİLEME
VERİLERİNİ BİRLEŞTİREREK BİYOKİMYASAL
ETKİ HESAPLAMA YÖNTEMİ GELİŞTİRİLMESİ

Muhammet ÇELİK

DOKTORA TEZİ
Biyomühendislik Anabilim Dalı
Biyomühendislik Programı

Danışman
Doç. Dr. Alper YILMAZ

Mart, 2024

T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

**MEME KANSERİ YENİ NESİL DİZİLEME VERİLERİNİ
BİRLEŞTİREREK BİYOKİMYASAL ETKİ HESAPLAMA
YÖNTEMİ GELİŞTİRİLMESİ**

Muhammet ÇELİK tarafından hazırlanan tez çalışması 20.03.2024 tarihinde aşağıdaki jüri tarafından Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Biyomühendislik Anabilim Dalı Biyomühendislik Programı **DOKTORA TEZİ** olarak kabul edilmiştir.

Doç. Dr. Alper YILMAZ
Yıldız Teknik Üniversitesi
Danışman

Jüri Üyeleri

Doç. Dr. Alper YILMAZ, Danışman
Yıldız Teknik Üniversitesi

Prof. Dr. İsmail KOCAÇALIŞKAN, Üye
Biruni Üniversitesi

Prof. Dr. Barbaros NALBANTOĞLU, Üye
Yıldız Teknik Üniversitesi

Doç. Dr. Ahmet ÖZAYDIN, Üye
İstanbul Üniversitesi Cerrahpaşa

Dr. Öğr. Üyesi M. Erkan KARABEKMEZ, Üye
İstanbul Medeniyet Üniversitesi

Danışmanım Doç. Dr. Alper YILMAZ sorumluluğunda tarafımca hazırlanan Meme Kanseri Yeni Nesil Dizileme Verilerini Birleştirerek Biyokimyasal Etki Hesaplama Yöntemi Geliştirilmesi başlıklı çalışmada veri toplama ve veri kullanımında gerekli yasal izinleri aldığımı, diğer kaynaklardan aldığım bilgileri ana metin ve referanslarda eksiksiz gösterdiğimi, araştırma verilerine ve sonuçlarına ilişkin çarpıtma ve/veya sahtecilik yapmadığımı, çalışmam süresince bilimsel araştırma ve etik ilkelerine uygun davrandığımı beyan ederim. Beyanımın aksinin ispatı halinde her türlü yasal sonucu kabul ederim.

Muhammet ÇELİK

İmza

*Aileme
Eşime
ve Çocuklarıma*



TEŞEKKÜR

Bu tez çalışmamın tamamlanmasında, değerli danışmanım Doç. Dr. Alper YILMAZ'ın rehberliği ve katkıları büyük önem taşımıştır. Kendisinin sabrı, özverisi ve değerli önerileri için derin minnettarlığımı ifade etmek isterim.

Akademik yolculuğumun her aşamasında yanımda olan ve bana hem maddi hem de manevi destek sağlayan değerli hocam Prof. Dr. İsmail Kocaçalışkan'a ve doktora yeterlilik, tez izleme ve savunma süreçlerinde jürimde bulunan ve desteğini eksik etmeyen Doç. Dr. Ahmet Özaydın'a şükranlarımı sunarım.

Bölüm hocalarım Prof. Dr. Musa TÜRKER, Prof. Dr. Cem Bülent ÜSTÜNDAĞ, Prof. Dr. Mehmet Murat ÖZMEN ve tez savunma jüri üyelerim Prof. Dr. Barbaros NALBANTOĞLU ve Dr. Öğr. Üyesi Muhammed Erkan KARABEKMEZ'e, akademik rehberlikleri ve kişisel destekleri için özellikle teşekkürlerimi sunarım..

Üniversite yıllarımızdan bu yana devam eden dostluğumuz boyunca, hayatın her alanında birbirimize destek olduğumuz Dr. Öğr. Üyesi Hüseyin GÜMÜŞ'e ve tezimin son aşamalarında motivasyonumu yüksek tutmama yardımcı olan Dr. Hüsnü KARA'ya özel teşekkürlerimi sunarım.

Yakın dostlarım Atıfcan DEMİRCİOĞLU, Ufuk KINDAP, Mehmet Furkan ÖZDOĞAN, Murat AK ve özellikle Serhat DİKER ve Bilal MİTAF'a yazılım noktasındaki desteği ve birlikte geçirdiğimiz keyifli günler için teşekkür ederim.

Alper Yılmaz Araştırma Grub'undaki arkadaşlarım Oğuzhan KALYON, Gülce ÇELEN, Elif İrem KELEŞ ve özellikle de Selcen ARI-YUKA ve Nilay YÖNET'e, grup içindeki işbirliği ve destekleri için teşekkür ederim.

Kıymetli dostlarım İsmail AĞIR, Mustafa NİĞDE, Bilal ÖZTÜRK ve Mutlu Mustafa KUŞPINAR'a ve ismini tek tek sayamadığım ancak her biri bu yolculuğumda önemli bir yer tutan, iyi dilekleriyle beni güçlendiren tüm arkadaşlarıma, meslektaşlarıma ve bu süreçte yanımda olan herkese destekleri için özel bir teşekkür sunarım.

Eşime, doktoramın her aşamasında gösterdiği destek, sabır ve teşvik için özel bir

teşekkürümü sunmak isterim. Her adımda bana güç veren ve hayatımın her alanında beni tamamlayan varlığı için minnettarım.

Değerli evlatlarım Ayşe Hüma ve Meva Nur; varlıklarıyla hayatıma neşe katan, yorgunluğumu alan biricik varlıklarım, iyiki varsınız.

Son olarak, annem, babam, ablam ve kız kardeşime; hayatım boyunca bana sundukları sevgi ve destek için teşekkür ederim. Ailemin her bir ferdine, hayatımda sahip olduğum en değerli armağan oldukları için minnettarım.

Muhammet ÇELİK



İÇİNDEKİLER

KISALTMA LİSTESİ	ix
ŞEKİL LİSTESİ	xi
TABLO LİSTESİ	xiii
ÖZET	xiv
ABSTRACT	xvi
1 GİRİŞ	1
2 KANSER	4
2.1 Kanser Tanımı	4
2.2 Kanser Türlerinin Hücre Kökenine Göre Sınıflandırılması	6
2.3 Kanser Oluşumu	7
2.4 Kanser Metastazı	7
2.5 Kanser Genomiği	8
2.6 Kanserde Biyolojik Yolakların Etkileri	10
2.7 Kanserde Sistemik Bütünleştirici Yaklaşımlar	10
3 MEME KANSERİ	13
3.1 Meme Dokusu Anatomisi	13
3.2 Meme Kanseri Epidemiyolojisi	14
4 BİYOİNFORMATİK VE TRANSKRİPTOMİK	17
4.1 Biyoinformatik	17
4.2 Tıbbi Genetik ve Genomikte Biyoinformatik	18
4.3 Yeni Nesil Dizileme	18
4.3.1 Roche Genom Sekanslayıcı	21
4.3.2 İllumina	23
4.4 RNA-seq	25
4.5 Gen İfadesi Analizi İçin RNA-seq Veri Hazırlama	25
4.6 Diferansiyel İfade Analizi	26

4.7	LIMMA Metodu	27
4.8	Gene-Protein-Reaksiyon Kuralları	28
4.9	Kullanılan Veri Tabanları	30
4.10	R Yazılım Platformu	32
5	YÖNTEM	34
5.1	Veri Kaynakları	34
5.2	Reaksiyonlar ve Bileşiklerin Oluşturulması	34
5.2.1	Veri Kaynakları ve Tümör Profili Analizi	34
5.2.2	TCGA Veri Setinin İşlenmesi ve Analizi	34
5.2.3	Metabolik Reaksiyonların Ayrıştırılması ve Analizi	35
5.3	Hastalar Bazında HMR Değerlendirmesi	37
5.3.1	Gen-Protein-Reaksiyon İlişkilerinin İşlenmesi	37
5.3.2	Geliştirilen R Fonksiyonları ile GPR Kurallarının Analizi	38
5.3.3	Metodolojinin Uygulanması ve Birey Bazında HMR Hesaplamaları	39
5.4	LIMMA Analizi	40
5.4.1	Veri Gruplandırma ve Tasarım Matrisi	41
5.4.2	Voom Dönüşümü ve Lineer Model Uygulama	42
5.4.3	Diferansiyel İfade Analizi Sonuçları	42
5.4.4	Reaksiyon ve Metabolik Bileşiklerin Entegrasyonu	43
5.5	Metabolik Reaksiyon Analizi	43
5.5.1	Artan ve Azalan Bileşiklerin Belirlenmesi	43
5.5.2	Kenar ve Düğüm Verilerinin İşlenmesi	44
5.5.3	HMR Değişiklik Özeti	45
5.6	Metabolik Veri Setinin Filtrelenmesi ve İşlenmesi	46
5.6.1	Veri Setinin Hazırlanması ve Filtreleme İşlemleri	46
5.6.2	Gen-Protein-Reaksiyon Kurallarının İncelenmesi	46
5.6.3	Veri Setinin Daraltılması ve Metabolik Bileşiklerin Filtrelenmesi	46
5.7	Tümör ve Sağlıklı Örneklerin Reaksiyon İfadesi Analizi	47
5.7.1	Veri Seti ve Ön İşleme	47
5.7.2	Tümör ve Sağlıklı Örneklerdeki Reaksiyon İfade Seviyelerinin Karşılaştırmalı Analizi	48
5.7.3	Eşik Değeri Belirleme ve Filtreleme	48
5.7.4	Kontrol ve Sağlama İşlemleri	49
5.7.5	Değerlendirme ve Sonuçların Doğrulanması	50
5.8	Metabolik Reaksiyonların Frekans Analizi ve Metabolitlerin Belirlenmesi	52

5.8.1	Veri Setlerinin Bütünleştirilmesi ve Yeni Değişkenlerin Tanımlanması	52
5.8.2	Metabolik Ağ Analizi İçin Veri Setinin İncelenmesi	52
5.8.3	Frekans Analizi ve Veri Setinin Filtrelenmesi	53
6	SONUÇ	55
6.1	Metabolik Ağın Analizi	56
6.2	GPR İlişkilerinin Metabolik Ağ Analizlerindeki Rolü	58
6.3	LIMMA Analizi	60
6.4	Metabolik Değişikliklerin Analizi ve Özetlenmesi	64
6.5	Artan ve Azalan Reaksiyonların Kontrolü	70
6.6	Prolin Metabolizması Değişikliği	79
6.7	GDP-L-fukoz Metabolizması Değişikliği	80
6.8	Trans-4-Hidroksi-L-Prolin Metabolizması Değişikliği	81
6.9	Deoksiuridin Metabolizması Değişikliği	82
6.10	Skualen Metabolizması Değişikliği	84
	KAYNAKÇA	88
	TEZDEN ÜRETİLMİŞ YAYINLAR	104

KISALTMA LİSTESİ

BIGG	Biochemically, Genetically, and Genomically structured
BRCA	Breast Cancer
DESeq	Differential Expression Analysis Based on the Negative Binomial Distribution
DNA	Deoksiribonükleik Asit
dGMP	Deoksiguanozin Difosfat
dUMP	Deoksiüridin Monofosfat
EC	Enzim Komisyon
ECM	Ekstraselüler Matriks
edgeR	Empirical Analysis of Digital Gene Expression Data in R
EHMM	Edinburgh Human Metabolic Model
FPM	Fragments Per Million
GDC	Genomic Data Commons
GFUS	GDP-L-Fukoz Sentaz
GPR	Gen-Protein-Reaksiyon
HGNC	HUGO Gene Nomenclature Committee
HİF	Hipoksi ile İndüklenebilir Faktör
HMA	Human Metabolic Atlas
HMR	Human Metabolic Reaction
KEGG	Kyoto Encyclopedia of Genes and Genomes
LIMMA	Linear Models for Microarray and RNA-Seq Data
mRNA	Mitokondriyal Ribonükleik Asit
NCI	National Cancer Institute
PEP	Fosfoenolpirüvat

PYCR	Prolin-5-Karboksilat Reduktaz
RNA	Ribonükleik Asit
SQLE	Skualen Epoksidaz
TCGA	The Cancer Genome Atlas
TK1	Timidin Kinaz 1
TMM	Trimmed Mean of M-values
YND	Yeni Nesil Dizileme



ŞEKİL LİSTESİ

Şekil 2.1	2020’de Dünya genelinde her iki cinsten ve her yaşta tahmini kanser sayıları ve oranları [10]	5
Şekil 2.2	2020’de, Dünya genelinde, her iki cinsten ve her yaşta, tahmini ölüm sayıları ve oranları [10]	5
Şekil 2.3	Dünya genelinde 2020-2025 yılları için, her iki cinsten ve her yaşta tahmini yeni vaka sayısı [10]	6
Şekil 3.1	2020’de Meme anatomisi [50]	13
Şekil 3.2	2020’de Türkiye’de her iki cinsten ve her yaşta tahmini kanser sayıları ve oranları [10]	15
Şekil 3.3	2020’de, Türkiye’de, her iki cinsten ve her yaşta, tahmini ölüm sayıları ve oranları [10]	15
Şekil 4.1	DNA dizisinin megabaz(1 milyon baz) başına maliyeti [72]	20
Şekil 4.2	DNA dizisinin genom başına maliyeti[72]	20
Şekil 4.3	DNA örneklerinin kütüphane hazırlığı aşamaları ve pirosekanslama yöntemi	22
Şekil 4.4	Kütüphane hazırlığı ardından akış hücresi üzerinde dizileme	24
Şekil 4.5	GPR kurallarının görsel şeması	29
Şekil 5.1	İnsan metabolit ağı verilerinin işlenmesi: reaksiyonların ayrıştırılması ile yönünün belirlenmesi ve substrat ile ürünlerin ayrı sütunlara dağıtılması	35
Şekil 5.2	Substrat ve ürün veri setlerinin entegrasyonu ile metabolik bileşik ağının oluşturulması	37
Şekil 5.3	Reaksiyon ifade kurallarının semantik analizi ve sayısal dönüşüm için geliştirilen "gr_modifier" fonksiyonu	38
Şekil 5.4	Gen ifade kurallarının sayısal değerlendirmeye dönüştürülmesi ve kişiselleştirilmiş genetik aktivite sonuçlarının çıkarılması	39
Şekil 5.5	Meme kanseri RNA ifade verilerinin kişiselleştirilmiş HMR hesaplamaları ile entegrasyonu	40
Şekil 5.6	Lineer model ve eBayes fonksiyonları kullanılarak gruplar arası reaksiyon ifadesi farklılıklarının istatistiksel değerlendirilmesi	42
Şekil 5.7	LIMMA analizi sonuçları ile metabolit ağı entegrasyonunun Tidygraph Yapısı	43

Şekil 5.8	Artan ve azalan metabolit veri setlerinin oluşturulması ve reaksiyon ifade düzeylerine göre özetlenmesi	44
Şekil 5.9	Kenar bilgilerine "from" ve "to" indekslerine karşılık gelen metabolit isimlerinin eklenmesi	44
Şekil 5.10	Kenar verilerine "from" ve "to" düğüm id'lerinin karşılık gelen isimlerinin eklenmesi ve ifade düzeyindeki değişimlerin belirlenmesi	45
Şekil 5.11	Her HMR için metabolit sayılarının hesaplanması ve "from_name" ile "to_name" metabolitlerinin detaylandırılması . .	45
Şekil 5.12	Tümör ve sağlıklı örneklerin analiz için ayrıştırılması	47
Şekil 5.13	Reaksiyon ifade seviyelerindeki sapma değerlerine dayalı artış ve azalış oranlarının hesaplanması	48
Şekil 5.14	Yüksek değişim dösteren reaksiyonlarda en belirgin artış ve azalış gösteren metabolitler ile dengeli veri seti oluşturulması	49
Şekil 5.15	Analitik değerlendirme için öne çıkan reaksiyonların seçilmesi . .	49
Şekil 5.16	Seçilen reaksiyonların tümör ve sağlıklı örnek verileriyle ilişkilendirilerek 'new_df' için karşılaştırmalı ifade düzeyleri setinin oluşturulması	50
Şekil 5.17	Metabolik reaksiyon verilerinin entegrasyonu ve substrat-ürün dağılımının analizi	52
Şekil 5.18	Metabolit frekans dağılımlarının analizi ve taşıyıcı moleküllerin hariç tutularak son veri setinin oluşturulması	54
Şekil 6.1	Compound_network veri setinin özeti, 6008 Düğüm ve 85934 Kenar içerir	57
Şekil 6.2	Voom Grafiği	62
Şekil 6.3	Volkan Plot Grafiği	64
Şekil 6.4	Reaksiyonların ve GPR kurallarına göre hesaplanan ifade düzeylerinin kutu grafiği ile gösterimi	70

TABLO LİSTESİ

Tablo 5.1	Örnek substrat ve ilişkili reaksiyonlar veri seti	36
Tablo 5.2	Örnek ürün ve ilişkili reaksiyonlar veri seti	36
Tablo 6.1	4 hasta ve 2 sağlıklı bireye ait "tcga_brca" veri seti örneği	56
Tablo 6.2	HMR_result_final veri setinde bulunan iki meme kanseri hastasının metabolik reaksiyon kodları, GPR kuralları ve gen ifade sayımlarından örnekler	60
Tablo 6.3	İfadesi değişen reaksiyonların analiz sonuçları	62
Tablo 6.4	"limma_compound_network" veri setinin tablo gösterimi	65
Tablo 6.5	HMR_change_summary_unnested Veri Setinden Örnek	66
Tablo 6.6	HMR_change_summary_unnested_3 Veri Setinden Örnek	68
Tablo 6.7	Seçili reaksiyonların tümör örneklerinde sağlıklı kontrollere göre gösterdiği ifade artış ve azalış oranları	69
Tablo 6.8	Artış ve Azalış Gösteren Reaksiyonların Karşılaştırılması İçin Seçilen Reaksiyon Örnekleri: new_df Veri Seti Üzerinden Kontrol	71
Tablo 6.9	Artış ve azalış gösteren reaksiyonlarla karşılaştırılması için seçilen hasta örnekleri ve gen kurallarının örnekleri: saglama_data veri seti üzerinden kontrol	72
Tablo 6.10	Kontrol amaçlı normal_mean_2 veri seti tablosu	73
Tablo 6.11	Reaksiyon Değişiklikleri ve İlgili Bileşiklerin Örnekleri	74
Tablo 6.12	Metabolik bileşiklerin değişim frekansları	75
Tablo 6.13	Meme Kanseri İfadesi Değişen Reaksiyonlar Bağlı Olarak Artan ve Azalan Önemli Metabolitler	76

Meme Kanseri Yeni Nesil Dizileme Verilerini Birleştirek Biyokimyasal Etki Hesaplama Yöntemi Geliştirilmesi

Muhammet ÇELİK

Biyomühendislik Anabilim Dalı
Doktora Tezi

Danışman: Doç. Dr. Alper YILMAZ

Meme kanseri, dünya genelinde kadınlar arasında en yaygın görülen kanser türlerinden biridir ve her yıl yaklaşık 2,3 milyon yeni vaka teşhis edilmekte, bu durum da takriben 685.000 kişinin hayatını kaybetmesine neden olmaktadır. Bu çalışma, meme kanseri ile ilişkili yeni nesil dizileme verilerini kullanarak, hastalar ile sağlıklı bireylerin metabolizmalarındaki reaksiyon ifade farklılıklarını belirlemeye ve kanser sürecinde bu reaksiyonlara bağlı olarak değişen metabolitleri tespit etmeye odaklanmıştır. Bu bağlamda, İnsan Metabolik Atlası(HMA) ve Kanser Genom Atlası (TCGA) veri tabanlarından elde edilen transkriptomik veriler, R yazılımı kullanılarak işlenmiş ve gen-protein-reaksiyon kurallarına dayalı olarak her birey için reaksiyon ifade verileri hesaplanarak yeni bir veri seti oluşturulmuştur. Bu veri seti, LIMMA yazılım paketi kullanılarak farklılık gösteren reaksiyonların belirlenmesinde temel alınmıştır.

Analiz sonuçlarında bazı metabolitler dikkate alınmamıştır. Bunun nedeni, söz konusu metabolitlerin yaygın olması veya kanser metabolizmasındaki spesifik rollerinin öneminin sınırlı olması gibi faktörler olmuştur. Bu odaklanmış yaklaşım, araştırmamızı kanser hücrelerinin büyüme ve hayatta kalma stratejilerinde merkezi rol oynayan ve potansiyel terapötik müdahaleler için önem taşıyan metabolik yollara yönlendirmemizi sağlamıştır.

Araştırmamız, arginin ve prolin metabolizması, fruktoz ve mannoz metabolizması, nükleotid metabolizması ve kolesterol biyosentezi gibi metabolik yollardaki

anahtar reaksiyonları ele almıştır. Bu reaksiyonların kanser metabolizmasındaki rolleri, literatürle karşılaştırılarak tartışılmış ve literatürle yapılan karşılaştırmalar sonucunda bu bulguların uyumlu olduğu tespit edilmiştir. Bu uyum, uygulanan analiz yöntemlerinin doğruluğunu ve tutarlılığını teyit ederken, elde edilen sonuçların literatürle uyum göstermesi açısından, pratikte yeni bir analiz metodolojisi sunmuştur. Özellikle farklı ifade edilen reaksiyonların incelenmesi, kanser metabolizması ve metabolik bileşenlerin önemli olduğu diğer hastalıklar hakkında daha derinlemesine çalışmalar yapılması için yeni yollar açmaktadır.

Transkriptomik verilerin analizi üzerine kurulu bu çalışma, kanser mekanizmalarının daha iyi anlaşılmasına katkıda bulunarak, potansiyel olarak yan etkileri daha az olan ve doğrudan metabolizmadaki hedef metabolit ve enzimlere yönelik yeni terapilerin keşfedilmesi için sağlam bir zemin oluşturmuştur.

Anahtar Kelimeler: Meme kanseri, metabolomik, yeni nesil dizileme, biyobelirteç tanımlama, gen-protein-reaksiyon kuralları.

ABSTRACT

Developing Biochemical Impact Calculation Method By Integrating Next Generation Sequencing Data From Breast Cancer Samples

Muhammet ÇELİK

Department of Bioengineering
Doctor of Philosophy Thesis

Supervisor: Assoc. Prof. Dr. Alper YILMAZ

Breast cancer is among the most common types of cancer worldwide in women, with approximately 2.3 million new cases diagnosed annually, leading to roughly 685,000 deaths. This study has focused on using next-generation sequencing data associated with breast cancer to determine the differences in reaction expressions between patients and healthy individuals' metabolisms and to detect the metabolites that change according to these reactions during the cancer process. In this context, transcriptomic data obtained from the Human Metabolic Atlas (HMA) and The Cancer Genome Atlas (TCGA) databases have been processed using R software and reaction expression data for each individual have been calculated based on gene-protein-reaction rules, thereby creating a new dataset. This dataset has been fundamental in identifying differing reactions using the LIMMA software package.

Some metabolites have not been considered in the analysis results. The reason for this has been factors such as the prevalence of these metabolites or the limited significance of their specific roles in cancer metabolism. This focused approach has allowed our research to be directed towards metabolic pathways that play a central role in the growth and survival strategies of cancer cells and are significant for potential therapeutic interventions.

Our research has addressed key reactions in metabolic pathways such as arginine and proline metabolism, fructose and mannose metabolism, nucleotide metabolism, and cholesterol biosynthesis. The roles of these reactions in cancer metabolism have

been discussed by comparing them with the literature, and it has been determined that these findings are consistent with the literature. This consistency has confirmed the accuracy and reliability of the applied analysis methods, while the compatibility of the results with the literature has presented a new analysis methodology in practice. Specifically, the examination of differentially expressed reactions has opened new avenues for more in-depth studies on cancer metabolism and other diseases where metabolic components are significant.

Based on the analysis of transcriptomic data, this study has contributed to a better understanding of cancer mechanisms, potentially providing a solid foundation for discovering new therapies with fewer side effects that target specific metabolites and enzymes in the metabolism directly.

Keywords: Breast cancer, metabolomics, next-generation sequencing, biomarker identification, gene-protein-reaction rules.

1 GİRİŞ

Kanser, bir hücrenin DNA'sındaki (deoksiribonükleik asit) hasarın (mutasyon) sonucudur, böylece hücreler normal işlevselliğini kaybetmekte ve bunun sonucunda normal dokuların işlevleri bozulana kadar sınırsız olarak çoğalmaktadırlar. Kanser tedavisinde kişiselleştirilmiş yaklaşımlar geliştirmek amacıyla, araştırmacılar her bir kanser türünün biyolojik işaretleyicilerini belirlemeye ve bu işaretleyicileri ayrıntılı bir şekilde tanımlamaya çalışmaktadırlar. Bu sayede, hastalara en uygun ve doğru teşhis konulmasını hedeflemektedirler[1].

Meme kanseri (MK), kadınlarda en yaygın kanser olup, tüm kadın kanserlerinin % 23'ünü oluşturmaktadır. 2020 verilerine göre küresel olarak, yaklaşık 2,3 milyon kadın MK tanısı almakta ve 685 bin kişinin ölümüyle sonuçlanmaktadır. Erken tanı ve etkili tedaviler MK ölümlerini azaltmada büyük önem taşıdıklarından, araştırmacılar hastalığa işaret eden ek genetik biyobelirteçleri ve yeni ilaç hedeflerini belirlemeye çalışmaktadır [2].

Genetik biyobelirteçlerin temel yapıtaşları genlerdir. Gen, DNA'nın protein üretiminde gerekli bilgileri içeren bir segmenti olarak tanımlanmaktadır. Merkezi dogma çerçevesinde, genlerin RNA olarak adlandırılan ara molekülere transkripsiyonu gerçekleşmekte ve bu RNA molekülleri, hayati işlevleri yerine getiren proteinlere translasyon edilmektedir. Eğer bir gen, kanser baskılama işlevi gören bir proteini kodluyorsa ve bu gen yeterince aktif olmazsa (transkripsiyonu azalırsa), hücre kansere dönüşebilmektedir. Aynı şekilde, bir gen eğer kanseri destekleyen bir proteini kodluyorsa ve bu gen aşırı aktif hale gelirse (normalden fazla transkripsiyon yapılsa), bu da hücrenin kansere dönüşmesine yol açmaktadır [3].

Bu nedenle, kanser gelişimine katkı sağlayabilecek genlerin ve bunların reaksiyon ürünlerinin belirlenmesi, bu genlerin ve reaksiyonların aktivitelerinin gerektiğinde artırılması veya azaltılması, kanser türlerini sınıflandırmak için gereklidir. Bugüne kadar bu alanda yapılan tüm çalışmalara, harcanan zaman ve kaynaklara rağmen,

ilerleme yine de sınırlı kalmıştır. Gerçekte, vücut çeşitli iç ve dış faktörlerden etkilenebilir ki bu faktörler kanser oluşumuna katkıda bulunabilmektedir. Bu faktörlerin bireyden bireye gösterdiği değişkenlik nedeniyle, iki farklı kanser hastasından elde edilen örneklerin sonuçları birbirinden farklı olabilir. Dolayısıyla, daha etkili keşifler yapabilmek adına mevcut veri kaynaklarını daha ayrıntılı analiz edebilecek yeni tekniklerin geliştirilmesi gerekmektedir.

Gen ekspresyon verileri, bir genin belirli bir hastalık durumunda upregüle veya downregüle olduğu konusunda bilgi sağlayan deneysel veriler olarak kullanılmaktadır. Bu yaklaşım, kanserli hücrelerdeki reaksiyon ekspresyon seviyelerinin, sağlıklı hücrelere göre nasıl değiştiğini karşılaştırmak için kullanılmıştır. Geniş bir gen setinin laboratuvar ortamında analiz edilmesi hem maliyetli hem de pratikte zor olabilmektedir. Bu sebeple, makine öğrenimi ve veri madenciliği teknikleri, bu gen sayısını hastalıkla istatistiksel olarak ilişkili olabilecek daha yönetilebilir bir gen grubuna indirmek için kullanılabilmektedir. Bu yaklaşım sayesinde, araştırmacılar, laboratuvarında her bir gen ve enzimatik reaksiyonu potansiyel bir kanser biyobelirteci olarak test etmenin pratik olmayacağını göz önünde bulundurarak, sadece seçilen küçük grubuna odaklanabilmektedirler. Diğer bir ifadeyle, veri madenciliği tekniklerinin kanser araştırmalarında zaman ve maliyet tasarrufu sağlaması, araştırmacıların potansiyel kanser biyobelirteçlerine daha kolay ve etkili bir şekilde ulaşmalarına olanak tanımaktadır [4].

Gerçekleştirilen çalışmalar, farklı ifade gösteren genlerin (DEG) listesini belirlemeye ve bu genler üzerinde zenginleştirme analizi yapmaya yöneliktir. Ancak, bu süreçlerin sonunda, DEG listesindeki kritik öneme sahip genlerin saptanması her zaman mümkün olmamaktadır. Araştırmamızın odak noktası, bu gözden kaçabilen genler üzerinden ifadesi değişen reaksiyonları tespit etmek ve biyokimyasal etkilerini incelemek olmuştur. Bu amaçla, veritabanlarından alınan verilerde veri analizi ve işleme için istatistiksel yöntemler ve algoritmalar kullanılmıştır. Bu yaklaşımla, meme kanserinde artış veya azalış gösteren bazı bileşenler (compoundlar) tespit edilmiştir. Bu bulgular, literatürle hem uyumlu hem de uyumsuz sonuçlar içermekte olup, hastalıkların öngörülmesine yönelik yeni ve geniş bir perspektif sunmaktadır. Bu yaklaşım, araştırma alanına önemli katkılarda bulunmaktadır.

Çalışmamızda, bilgi sağlayıcı sonuçlar elde etmek amacıyla, biyolojik veritabanlarından alınan verileri, R programlama dilinde geliştirdiğimiz yeni kodlama ve hesaplama yöntemleriyle işleyerek ve filtreleyerek oluşturduğumuz yeni veri setleri kullanılmıştır. Elde edilen bu veriler, literatürden yapılan

doğrulamalarla birlikte, araştırmamızda sunulan kapsamlı bir yöntemin temelini oluşturacaktır.

Gen ekspresyon profillerinden elde edilen yeni biyobelirteçler, klinik uygulamalarda kritik bir rol oynayabilmektedir. Bu biyobelirteçler, klinik-patolojik özellikler arasındaki ilişkileri aydınlatarak, doktorların hastaya özel tedavi seçeneklerini belirlemelerine yardımcı olmaktadır. Etkili biyolojik belirteçler, erken kanser taramasına katkı sağlayabilir, teşhisin doğrulanmasına yardımcı, tedavi sonuçlarını öngörebilir ve terapötik yaklaşımları kişiselleştirebilmektedirler. Ayrıca, yeni tümör oluşum yollarını ortaya çıkararak, gelecekteki kanser araştırmaları için yol gösterici olmaktadır [5].

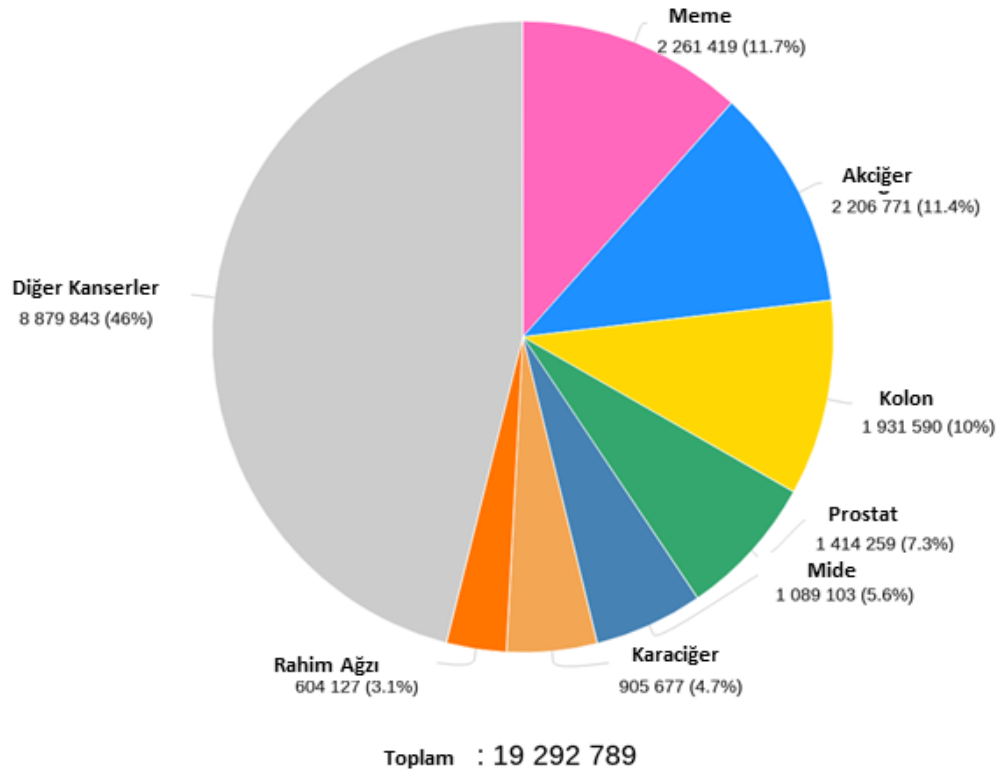


2.1 Kanserin Tanımı

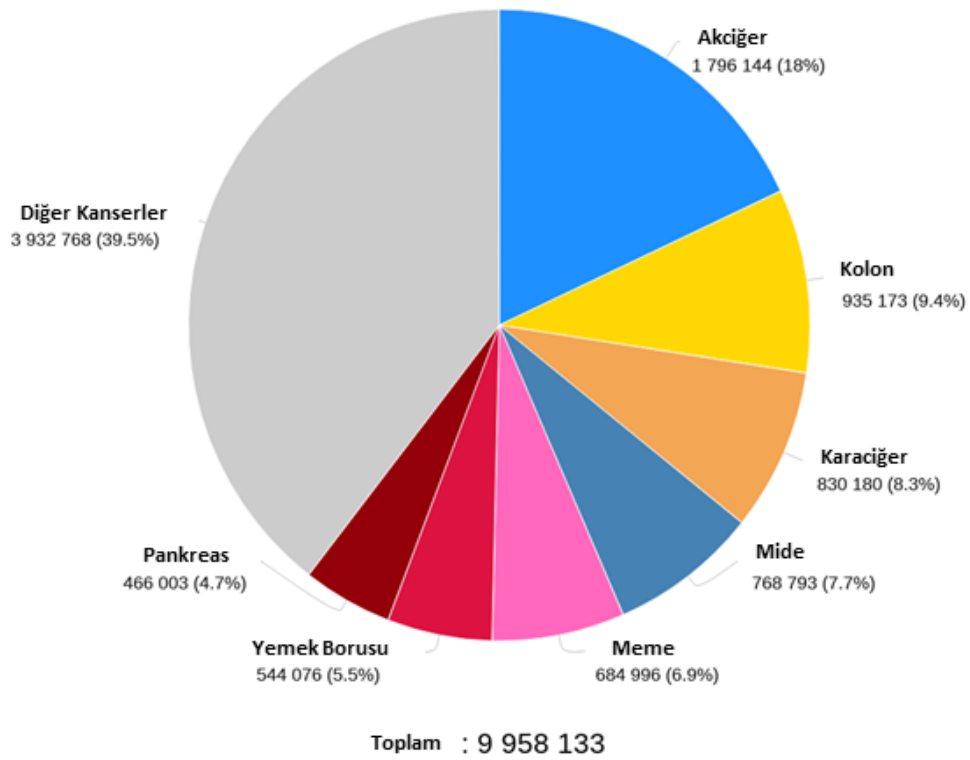
Kanser, hücrelerin kontrol dışı ve sınırsız biçimde çoğalması, büyüme sinyallerini kendi içinde üretme kapasitesi, dokulara sızma ve uzak organlara metastaz yapma yeteneği, ayrıca tümör hücrelerine kan akışını sağlayacak yeni damarlar geliştirme gibi özelliklerle tanımlanan, birden fazla faktöre bağlı bir hastalık olarak tanımlanmaktadır [6].

Kanser milyonları etkileyen en ölümcül hastalıklardan biridir ve Dünya Sağlık Örgütü(WHO), Uluslararası Kanser Araştırma Ajansı, 2020 kanser verilerine göre dünya çapında 19,3 milyon yeni vaka ve 9,9 milyon ölüm rapor edilmekte ve kanserli hasta sayısının 2025 yılına kadar 51 milyona çıkacağı tahmin edilmektedir. Şekil 2.1’de Dünya’da meme kanseri sayı ve oranları, Şekil 2.2’de kanserden ölümlerin sayı ve oranları ve Şekil 2.3’de ise 5 yıl içinde tahmin edilen yeni vaka değerleri görülmektedir [2]. Kanser yaşlanma hastalığı olarak bilinmektedir fakat tahmini tümör oluşumu, 20 yaş civarında başlar ve 50 yaş civarında kanser tespit edilmektedir [7]. Sigara içimi ve alkol tüketimi gibi sağlıksız yaşam tarzı seçimleri ile belirli çevresel faktörler, kansere yakalanma riskini artırmaktadır. [8].

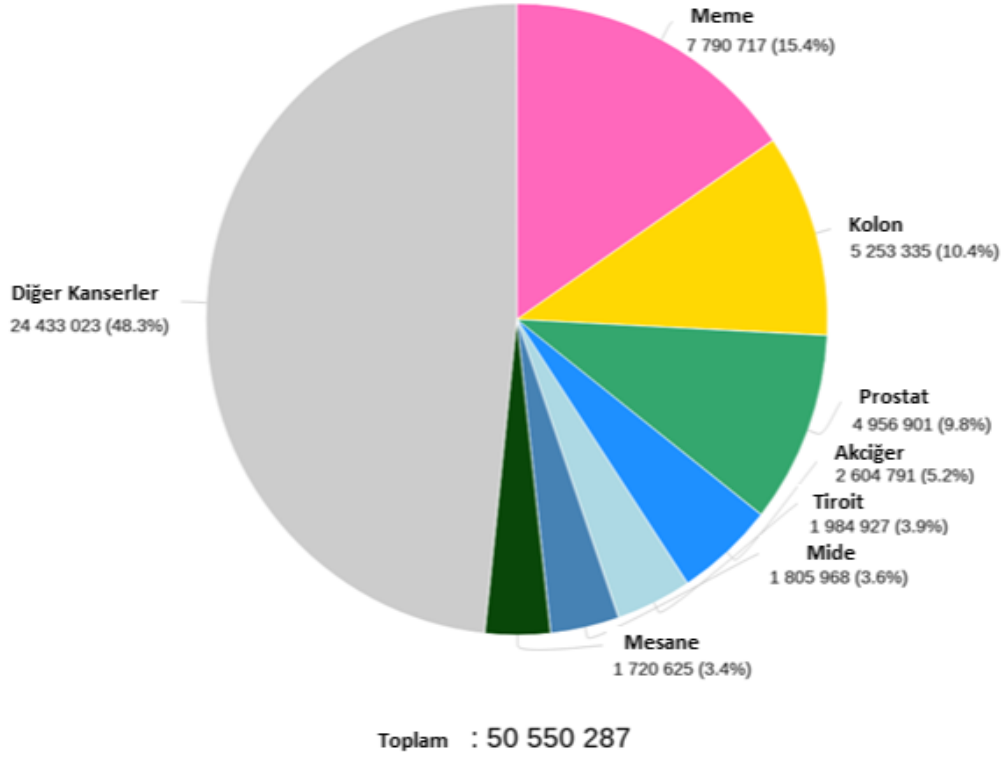
Kanserin temel işlevlerinin özünde, hücrelerde gerçekleşen ve ölümcül olmayan genetik hasar bulunmaktadır. Bu hasar öncelikle hücre çoğalmasını indükleyen genlerde aktivasyon, baskılayan genlerde ise inaktivasyon, hücrenin programlı ölümünü ilgilendiren genler ve DNA tamir genlerinde değişiklikler ile oluşabilmektedir. Kanser oluşumu, fenotipik düzeyde olduğu gibi genetik bakımdan da kompleks bir olaydır [9].



Şekil 2.1 2020’de Dünya genelinde her iki cinsten ve her yaşta tahmini kanser sayıları ve oranları [10]



Şekil 2.2 2020’de, Dünya genelinde, her iki cinsten ve her yaşta, tahmini ölüm sayıları ve oranları [10]



Şekil 2.3 Dünya genelinde 2020-2025 yılları için, her iki cinsten ve her yaşta tahmini yeni vaka sayısı [10]

2.2 Kanser Türlerinin Hücre Kökenine Göre Sınıflandırılması

Yüzden fazla kanser türü, etkilendikleri hücre tiplerine göre sınıflandırılabilir. Bu kanserler beş temel kategoriye ayrılmaktadır:

- **Karsinomlar:** Epitelyal hücrelerden kaynaklanan kanserlerdir ve vücut yüzeylerini kaplamaktadırlar.
- **Sarkomlar:** Sinirler, kan damarları, Kemikler, kaslar, tendonlar, kıkırdak, bağ ve yağ dokusu gibi dokulardan köken almaktadır.
- **Lenfoma ve Miyelom:** Lenf sistemi ve bağışıklık sistemi hücrelerinden gelişen kanser türleridir.
- **Lösemi:** Kemik iliğinden başlayarak anormal kan hücrelerinin üretimini içermektedir.
- **Merkezi Sinir Sistemi Kanseri:** Beyin ve merkezi sinir sistemi dokularından kaynaklanmaktadır [11].

2.3 Kanser Oluşumu

Kanserin oluşumunda kanserojen denilen zararlı alışkanlıklar, çevresel faktörler ve genetik faktörler etkilidir. DNA'ya zarar veren hem dışsal (ekzojen) hem de içsel (endojen) etkenler ve sonucunda oluşan genomik kararsızlıklar, DNA mutasyonlarının birikmesiyle neoplastik (kansereleşme) dönüşümü tetiklemektedir [12]. Yakın zamanda, epigenetik değişikliklerin de neoplastik dönüşüme katkıda bulunduğu düşünülmektedir. Kansereleşmeyi teşvik edebilecek genler onkogenler olarak adlandırılırken, kansereleşmeye karşı etkili olanlar tümör baskılayıcı genlerdir. Bu genlerdeki epigenetik değişiklikler veya mutasyonlar, bir onkogenin veya tümör baskılayıcı genin aşırı ya da az ifade edilmesine yol açarak kanserin oluşumuna katkı sağlamaktadır [13, 14]. Ek olarak, birçok mutasyon, belirli sinyalizasyon proteinlerinin aktivasyonunu veya inaktivasyonunu tetikler; bu da hücrelerin yayılmasına veya invazyonuna, aşırı çoğalmasına ve sonuç olarak kansereleşmeyi destekleyen sinyal yollarının aşırı aktivasyonuna neden olmaktadır [15].

2.4 Kanser Metastazi

Kanser hücreleri, başlangıçta henüz anlaşılabilen olaylar sonucu dolaşıma giren ve sonrasında bazı hücrelerin uzak organlarda yerleşip yayılma (metastaz) yapma kapasitesine sahip olduğu bir süreçle, kan veya lenfatik sistemi kullanarak vücudun diğer bölgelerine sıçramaktadır. Kanser hücrelerinin uzak organları işgal etme kabiliyeti tümöre özgüdür, ancak bazı durumlarda farklı tümör tipleri aynı organ bölgesine yerleşmektedirler. Lenfatik sistemi takip eden kanser hücreleri, ilerledikleri lenf düğümlerinde metastatik tümörlere dönüşmektedirler. Lenfatik sistem yoluyla vücudun yeni bölümlerine yayılmak için, kanser hücreleri orijinal tümörden ayrılmalı ve kendilerini bir lenf damarının dış duvarına tutturmalıdır. Daha sonra hücreler lenf ile yeni bir lenf noduna geçmektedirler [16]. Kan damarları yoluyla metastaz, hücrelerin damar duvarından geçip yeni bölgelere ulaşmasıyla gerçekleşir ve bu süreç birkaç aşamadan oluşmaktadır.

- **Lokal invazyon:** Kanser hücreleri, lokal olarak, çevredeki membranın etrafındaki bitişik dokuya girer.
- **İntravazasyon:** Tümör hücrelerinin damar duvarlarına geçiş yapması süreci, aynı zamanda "endotelial hücrelere göç" olarak da bilinir. Bu aşamada, kanser hücreleri, kan veya lenfatik damarların bazal membranını geçerek bu damarlara girerler.

- **Hematojen sađkalım ve translokasyon:** Kanser h creleri dolařım sistemine girdikten sonra orada hayatta kalmayı bařarabilir ve kan yoluyla v cudun uzak b lgelerindeki mikrodamarlara tařınabilirler. İnvazasyon ve hematojen sađkalım, kanser h crelerinin kan yoluyla yayılmasını sađlayan s re lerdir.
- **Ekstravazasyon:** Bu ařama, kanser h crelerinin kan dolařımından ayrılarak  vre dokulara ge iř yapmasıdır.
- **Kolonizasyon:** Kanser h creleri, v cudun uzak b lgelerine ulařtıktan sonra orada yerleřir ve adaptasyon sađlayarak  ođalmaya bařlarlar. Bu s re , uzak organlarda b y yen makroskopik ikincil t m rlerin oluřumuna yol a maktadır.

Kanser h crelerinin metastaza yeteneđini sađlayan molek ler s re ler tam anlamıyla belirlenmiř olmasa da, metastazın bařlaması ve geliřmesinde rol oynayan bazı genlerin iřlevsel aktiviteleri bilinmektedir. Metastaz bařlatma genleri, h cre motilitesinin, epitelden mezenkimal ge iřine(EMT), h cre dıřı matris bozunmasına ve damar oluřumuna, katkıda bulunmaktadır [17].

2.5 Kanseri Genomiđi

Kanser, genetik k kenli bir hastalıktır. Tahminen t m kanser vakalarının yalnızca %5 ila %10'u ailesel ge iř yoluyla edinilirken, b y k  ođunluđu sporadik, yani  vresel ve mutasyon etkileri ile ve beklenmedik bir řekilde ortaya  kmaktadır. Kalıtsal olmayan kanserler, genomik deđiřikliklerden kaynaklanan birikmiř genom deđiřkenliđinden kaynaklanmaktadır. Genomik deđiřiklikler; histon modifikasyonları ve DNA metilasyonu i eren epigenetik deđiřiklikler, DNA kopya sayısı deđiřiklikleri, gen ekspresyonu deđiřiklikleri ve mutasyonlar gibi genetik deđiřimlerden oluřmaktadır [18, 19].

Genetik arařtırmaları, sađlık ve hastalıkların incelenmesi dahil olmak  zere, k kl  bir řekilde deđiřtiren genomik teknolojiler, genoma ait t m genleri kapsayan sistematik ve objektif  alıřmalar yapılmasına imkan tanı mıřtır. Kanseri ilgili genomik arařtırmalar,  eřitli dokularda molek ler profiller  zerinden yaygın olarak ger ekleřtirilmiřtir [20]. Kullanım alanlarını geniřletmekte olan genom dizileme teknolojileri giderek daha fazla uygulanabilir hale gelmektedir. Bunun yanında genomik teknolojilerin sađlık sekt rindeki potansiyel uygulamalarının b y k bir kısmı hen z hayata ge irilmemiřtir [21].

Kanser genomları, özellikle tanı, prognoz ve tedavide çok sayıda fayda sağlamaktadır. Örneğin, tümör alt tipinin daha kesin tanısı veya daha az yan etki ile hedefe yönelik terapötikler de dahil olmak üzere moleküler tanıya dayanan özel tedavi ile hasta sonuçları ve yaşam kalitesi daha iyi olmuştur [22]. Şu ana kadar, genomiklerin klinik uygulamalara doğrudan katkısı kısıtlı olsa da, tespit edilen kanser genleri ve bunların tedavi amaçlı hedeflenmesi klinik ortamlarda kullanılmaya başlanmış ve bu uygulamalar hızla artmaktadır [23].

Kanserin ayırt edici özellikleri, genetik ve epigenetik değişimlerden, moleküler biyolojinin merkezi dogması üzerinden gerçekleşmektedir. Kanser sürecinde, onkogenlerin aşırı aktif olması veya tümör baskılayıcı genlerin yetersiz çalışması gibi genetik anomaliler, protein üretiminde bozulmalara ve bunun sonucunda ciddi etkilere sebep olmaktadır. DNA kopya sayısındaki değişiklikler, mRNA ekspresyonu ile doğrudan ilişkili olduğundan, gen ifadesinin düzeylerini etkilemektedir [24]. DNA'daki nokta mutasyonları da benzer şekilde, mRNA'nın üretilmesini düzenleyen transkripsiyon faktörlerinin veya miRNA'ların bağlanma bölgelerini bozarak gen ekspresyonunu etkileyebilir. Her iki DNA kopya sayısı değişikliği ve nokta mutasyonu, karşılık gelen proteinlerin ekspresyon değişikliklerine yol açabilir. Ancak, genetik değişimlerin her zaman gen ifadesini değiştirmede, bazen proteinlerin katlanması ve istikrarı yoluyla protein işlevlerini etkileyebileceği gözlemlenmiştir [25, 26].

Genetik değişikliklere ek olarak, epigenetik değişiklikler, kanser genlerini aktive ederek veya tümör baskılayıcı genleri inaktive ederek benzer şekilde protein ekspresyonunu da bozmaktadır. Metilasyon, epigenetik bir değişiklik formu olarak, tümör baskılayıcı genlerin transkripsiyonunu baskılayarak kanser gelişimine önemli bir katkıda bulunmuştur. Ek olarak, histon modifikasyonları (başka bir tipte epigenetik işaretleyici), DNA metilasyonu değişikliklerine ve kanserdeki gen aktivitesinin kontrolüne yüksek derecede bağlıdır [27, 28].

Proteinlerdeki değişiklikler ve bunların genetik ya da epigenetik değişiklikler yoluyla ifade edilmesi, protein-protein etkileşimlerini ve sonunda biyolojik ağların dinamiklerini etkileyen gen regülasyonunu etkiler. Buna ek olarak, biyolojik ağlardaki dengesizlikler, hücrelerin farklılaşması, çoğalması, göçü ve ölümü gibi temel yaşamsal işlevlerde bozulmalara yol açmaktadır. Bu durum göz önüne alındığında, kanser tek bir gene bağlı bir rahatsızlık yerine, biyolojik ağların genelinde meydana gelen değişimlerin bir sonucu olarak değerlendirilebilir [29]. Bu tez, genetik değişimlerin (özellikle transkriptom düzeyindeki değişikliklerin) ve bunların sonucunda yaşamsal faaliyetleri oluşturan biyokimyasal etkilerin incelenmesine odaklanmaktadır.

2.6 Kanserde Biyolojik Yolakların Etkileri

Kanser, genetik anlamda çeşitli formlara sahip olup, kanser hücrelerinde, normal hücrelerdeki genlerden farklı genel gen ekspresyonu değişiklikleri gözlemlenmiştir. Ancak, çoğu tümör örneğinde sınırlı sayıda anahtar değişiklik bulunmaktadır [6]. Yolak bazında yapılan genomik değişikliklerin analizi, farklı kanser örneklerinin ağ düzeyinde ortak bir dizi değişiklik paylaştığını göstermiştir [30].

Kanser sürecinde, birçok temel yolakta değişiklikler meydana gelmektedir. Meme ve yumurtalık kanseri ile diffüz büyük B-hücreli lenfomanın da içinde bulunduğu çeşitli kanser tiplerinde en yaygın etkilenen yollardan biri hücre döngüsüdür [31–33]. Hücre döngüsü, DNA replikasyonu ve hücre bölünmesini, döngüdeki kontrol noktaları aracılığıyla düzenlemektedir. Bu nedenle, hücre döngüsünün bozulması hücre sağkalımını etkileyerek kontrolsüz hücre çoğalmasına, yani kanserin en temel özelliklerinden birine neden olmaktadır [34]. Retinoblastoma (RB) geninin işlevsizliği, çoğu kanser türünde görülür ve RB'nin tümör baskılayıcı işlevi göz önüne alındığında, kanser gelişimi sırasında RB ile ilişkili yolların etkinliğinin azaltılması gereklidir. Kanser Genom Atlası (TCGA) tarafından yapılan over kanseri çalışması, incelenen tümör örneklerinin yaklaşık %67'sinde RB yolunun bozulduğunu ortaya koymuştur. Mitojenle aktive edilen protein kinaz (MAPK) yolları, hücre dışı sinyalleri entegre eden ve hücre büyümesi, çoğalma, hareket ve programlanmış hücre ölümü (apoptoz) gibi hayati biyolojik işlemleri düzenleyen bir sistemdir [31, 35].

2.7 Kanserde Sistemik Bütünleştirici Yaklaşımlar

İndirgenme olarak da bilinen böl ve ele geçir yaklaşımı, karmaşık biyolojik sistemleri daha küçük ve daha yönetilebilir parçalara bölmektedir. Böyle bir yaklaşım, son 40 yıldır bireysel genlerin veya proteinlerin kimyasal temelini ve işlevselliğini incelemekte başarılı olmuştur. Bunun yanı sıra, biyolojik sistemlerin karmaşıklığı, bireysel moleküler bileşenlerin seviyesinde her zaman açıkça görülemeyen sürpriz özellikler barındırır. Kanser fenotiplerinin, bireysel moleküler bileşenler aracılığıyla tam olarak açıklanabilmesi özellikle mümkün olmamaktadır.[29, 36].

Sistem biyolojisi, sistemde büyük resmi görebilmek için, indirgenme tarafından tespit edilemeyen, bütünsel ve kompozit özellikleri incelemek üzere tanıtılmıştır. Sistem biyolojisi yaklaşımları, tek bir bileşenin incelenmesinden ziyade, dinamik bir sistemin çeşitli faktörlerini farklı zaman dilimlerinde ve bağlamlarda eş zamanlı olarak değerlendirme imkanı sağlamıştır. Bu yaklaşımlar, daha indirgemeci

yaklaşımlara kapsamlı bir tamamlayıcı olarak gelişmektedir [37].

Genomik projeler, çeşitli organizmalara, fonksiyonel genetiğe ve bireyler ile sağlık veya hastalık riski arasındaki değişkenliğe odaklanan insan popülasyonlarına da uygulanmıştır [38].

Uluslararası projeler ve konsorsiyumlar, standart laboratuvar protokolleriyle toplanan verileri yayınlamaya başlamıştır. Örneğin, yüksek çözünürlüklü ve hızlı yüksek verimlilikli teknolojiler geliştirilmiş ve pek çok geniş çapta erişilebilir veri setleri oluşturulmuştur. TCGA, 33 farklı kanser türünden 20,000'den fazla primer kanser ve eşleşen normal örnekleri hakkında moleküler ve klinik bilgileri içeren bir veritabanıdır [29]. Ayrıca, Gen Ekspresyon Omnibusu (GEO) da, çoğunlukla mikroarray ve kısmen yeni nesil dizilim teknolojileri kullanarak toplanan bir milyondan fazla örnek içeren başka bir veritabanıdır [39].

Kanser genom projesi ise, insan genomunun dizilimi, kanser araştırma literatürü ve kanserle ilişkili gen dizilimleri kullanarak yapılan ilk kapsamlı genomik kanser araştırmalarından biridir [40, 41]. Kanser Genom Projesinin temel amacı, kanser ve normal doku örneklerini karşılaştırarak kanserlerde sıklıkla mutasyona uğramış olan “kanser genlerini” keşfetmektir. Bu durum, kanserde fonksiyonel olarak etkinleşen onkogenlerle birlikte inaktive olan tümör baskılayıcı genleri kapsamaktadır. Bu proje halen devam etmekte olup, kanser genleri üzerindeki somatik mutasyonları içeren Kanser Somatik Mutasyonlar Kataloğu (COSMIC) veri tabanının güncellenmesini sürdürmektedir. En son sürüm olan COSMIC v99 (Kasım 2023), 43.820 tam genom dahil olmak üzere 29.230 yayından 1.527.131 örnek içermektedir [42].

Geniş çaplı multi-omik veri setleri, kanser araştırmaları için önemli olanaklar sunmuş ve bu verilerin analizi ve entegrasyonu için sistemli metodların geliştirilmesini teşvik etmiştir. Kanser araştırmalarında multi-omik verilerin bütünleştirilmesi, hastalık alt türlerinin saptanması, prognostik biyobelirteçlerin belirlenmesi ve kanser mekanizmalarının daha iyi anlaşılmasında kritik rol oynamaktadır. Prognostik biyobelirteçler, hastaların sağ kalım sürelerinin tahmin edilmesinde kullanılırken, kanser alt tiplerinin tanımlanması, hastalara daha etkin tedavi alternatifleri sunma potansiyeline sahiptir. Ayrıca, kanser mekanizmalarının derinlemesine incelenmesi, kanser fenotiplerinin daha doğru yorumlanmasına olanak sağlamaktadır [43].

Birçok başarılı bütünleştirici multi-omik analiz uygulaması rapor edilmiştir. Özellikle, TCGA konsorsiyumu, çeşitli kanser türlerinde, klinik sonuçlarla ilişkili alt türleri belirlemiştir [44]. Bu alt türler, genetik, transkriptomik, proteomik ve

yolak verilerinin birleştirilmesi sonucunda tanımlanmıştır. Ayrıca, METABRIC grubunun yaptığı kapsamlı analizler sonucunda, 2,000 meme kanseri örneğinin genetik ve transkriptomik profilleri kullanılarak on yeni alt grup keşfedilmiştir [45]. Uri David Akavia ve ekip arkadaşları, kopya sayısı ve gen ekspresyonu verilerini bütünleştirerek kanser ilerlemesine katkıda bulunan değişiklikleri tespit etmek amacıyla bir hesaplama modeli geliştirmiş ve melanomda iki yeni sürücü gen keşfetmişlerdir. [46].

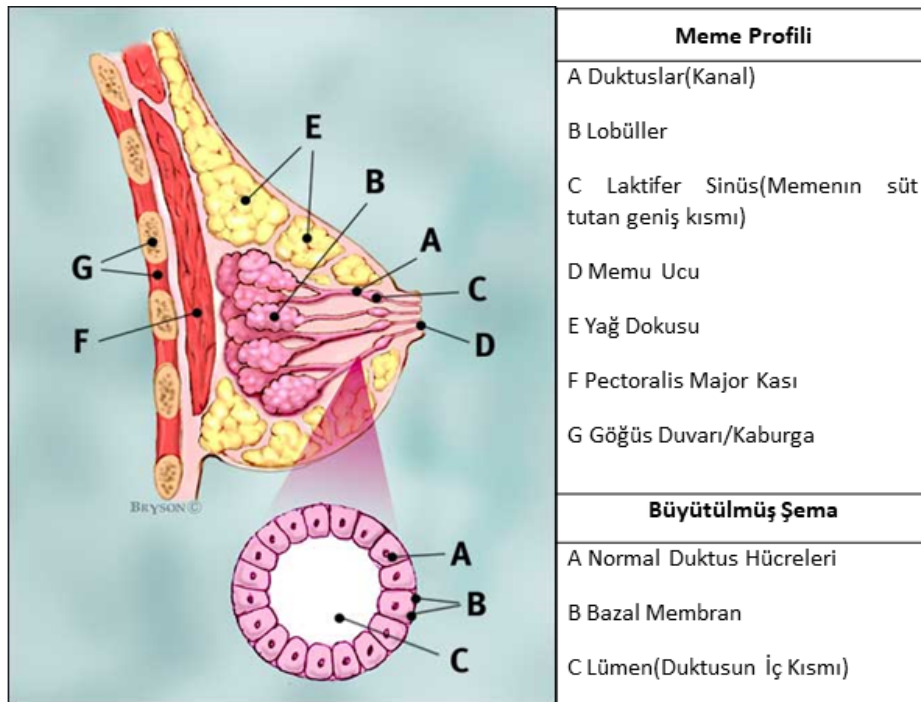
Klinik veriler, veri bütünleştirme sürecinde kritik bir işlev görür ve bu, sağkalım ilişkilendirme modelleriyle kanıtlanmıştır. Klinik bilgilerin entegrasyonu, sağkalım modelleri oluşturmanın en sık kullanılan yöntemlerinden biridir. Bu modellerde, biyobelirteçler (örneğin, genler, proteinler, yolaklar) ile yaşam süresi ve sağ kalım olayları arasındaki ilişki kurulmaktadır.. Tek bir moleküler veri kaynağı kullanarak oluşturulan sağkalım modellerinin hasta sağkalımı tahminlerinde sınırlı başarı sağladığı, ancak bütünleştirici yaklaşımlar kullanan modellerin daha güçlü tahmin yeteneklerine sahip olduğu bulunmuştur [47, 48].

3

MEME KANSERİ

3.1 Meme Dokusu Anatomisi

Meme deriye bağı bir çeşit apokrin bez olup süt üreten glandüler, lifli bir organdır. Göğüs ön bölgesinde ikinci ve altıncı interkostal aralıkta, aksillaya doğru uzantısı olan, pektoralis major kası ile birlikte asinüsler, duktuslar, parankim ve stroma gibi elemanlara sahiptir. Memenin salgı birimi olan asinüslerin iç kısımları silindirik veya küboid epitel ile döşeli, dış kısmı ise bağ dokusu, kan ve lenf damarları ile sarıdır. 10 ila 100 asinüs bir araya gelerek lobülleri, lobüllerde lobları oluşturur. Her asinüsün bir kanalı vardır. Bunlar birleşerek lobüllerin kanallarını, lobüllerin kanalları da lobların kanallarını meydana getirir. Memede 15 ila 20 adet bulunan loblar meme başına açılmadan önce laktifer sinüsler denilen genişlemeler yapmaktadır. Şekil3.1, biyolojik işlevlerini yerine getirmek için dallanmış bir duktal yapı oluşturan memenin anatomik yapısını göstermektedir [49]



Şekil 3.1 2020’de Meme anatomisi [50]

Östrojen, progesteron, adrenal kortikoidler, prolaktin, insülin ve büyüme hormonu gibi çok sayıda hormon meme gelişiminde rol oynamaktadır. Meme dokusu ergenlik, hamilelik ve emzirme sırasında ortaya çıkan gelişimsel değişikliklerle ilişkili bir dizi büyüme ve gerileme fazına maruz kalmaktadır [51, 52].

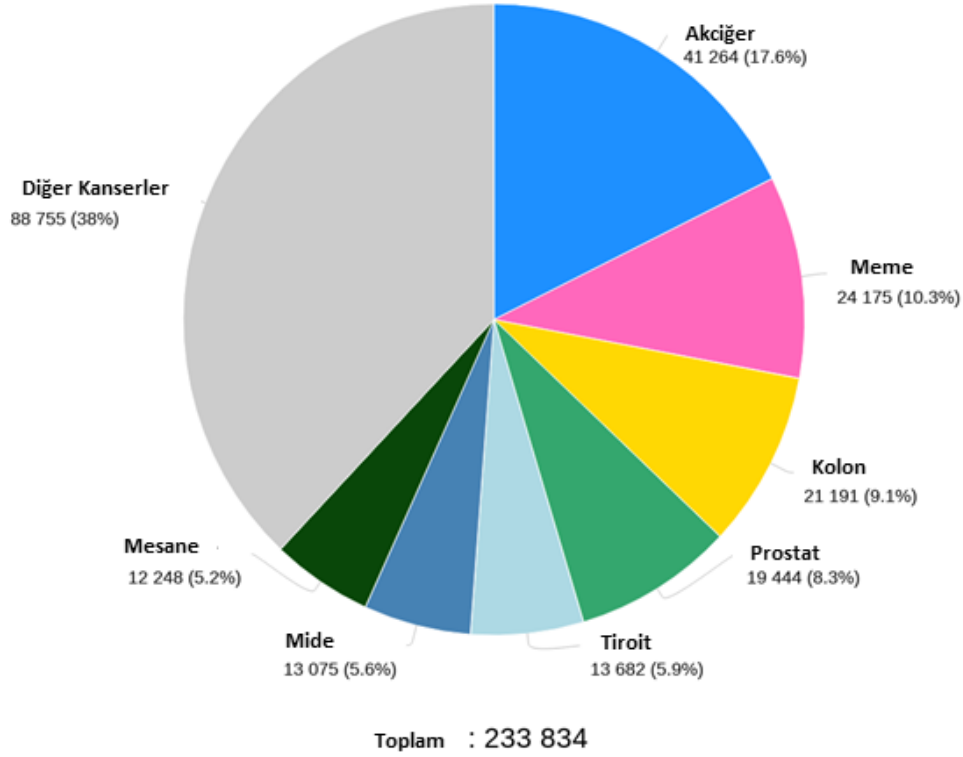
Duktal sistem memeli hayvanlardaki süt taşıma yapısını oluşturmaktadır. Bu sistem, süt üreten hücrelerin bulunduğu iç luminal hücreler ve bazal membrana bağlanan, oksitosine yanıt vererek kasılan dış tabaka hücrelerinden meydana gelmiştir. Dış tabaka hücrelerinin kasılma özelliği, onları düz kas hücrelerine benzer kılmaktadır. Bu benzerlikten ötürü, bu hücreler myoepithelium (epitel kası) olarak adlandırılmıştır. Myoepithelium hücrelerinin kasılması, sütün meme kanalları boyunca ilerlemesine ve memeden atılmasına yardımcı olmaktadır [53].

Meme kanserleri, metastazlarını genellikle lenf yolları ile yaptığından memenin lenf sistemi önemlidir. Meme lenf akımının büyük bir kısmı koltuk altı lenf nodlarına az bir kısmı ise iç meme lenf nodlarına olup memenin her bölgesinden hem koltuk altı hem de iç meme nodlarına lenf akımı olabilmektedir [54].

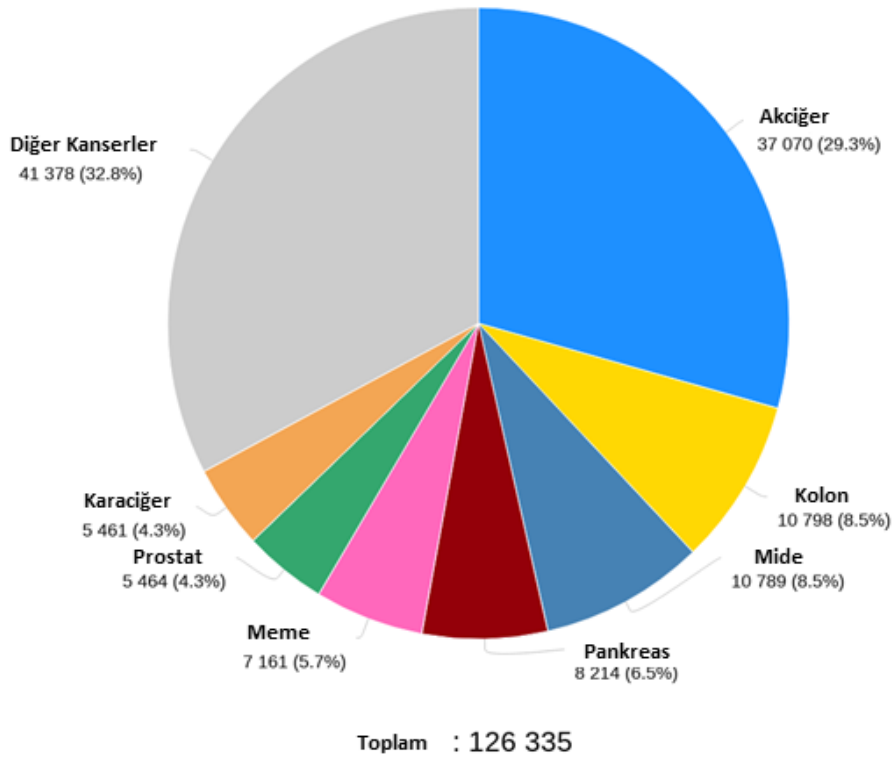
Meme bezleri, büyüme ve geri çekilme döngülerine maruz kaldığı için, sınırsız çoğalma kapasitesine sahip, henüz farklılaşmamış hücreler içermesi gerekmektedir. Olgunlaşmış meme bezi hücreleri, yani luminal ve miyoepitelyal hücreler, bu doku kök hücrelerinden hiyerarşik bir düzende türemektedir. Meme bezleri, miyoepitelyal, duktal ve süt üreten olmak üzere en az üç farklı hücre hattını barındıran pluripotent hücreleri içermektedir. Bu hücre hatlarının hepsinin tek bir köken hücre popülasyonundan geliştiği bilinmektedir. Bu süreç, hücre bölünmesi, büyüme ve gerileme evreleri boyunca, hücre hatlarına özgü spontan DNA mutasyonları sonucu kanser gelişim riskini artırmaktadır[51, 55].

3.2 Meme Kanseri Epidemiyolojisi

Meme kanseri, dünya çapında kadınlarda en yaygın kanser tipi olarak tanımlanan bir hastalıktır; ve 2020 yılında dünya çapında yaklaşık 2,3 milyon(%11,6) yeni vaka ve 685 bin ölümle kanserden ölümlerin önde gelen nedenidir. Türkiye’de kanser, ölüm sebepleri arasında, dolaşım sistemi hastalıklarının ardından ikinci sırada yer almaktadır. Uluslararası Kanser Araştırma Ajansı, kanser verilerine göre Türkiye’de 2020’de tahmini 233,834 kişi kansere, 24,175 kişi ise meme kanserine yakalanmıştır Şekil 3.2. Tüm kanserlerden ölümler 126,135 kişi ve meme kanseri nedenli ölümler ise 7161 kişi olarak rapor edilmiştir Şekil 3.3 [2].



Şekil 3.2 2020’de Türkiye’de her iki cinsten ve her yaşta tahmini kanser sayıları ve oranları [10]



Şekil 3.3 2020’de, Türkiye’de, her iki cinsten ve her yaşta, tahmini ölüm sayıları ve oranları [10]

Meme kanseri insidansı artmaktadır, bu da kısmen daha uzun yaşam beklentisi, daha az çocuk doğumu, ilk çocuğun geç doğumu, hormon replasman tedavisi, daha az fiziksel aktivite, fazla kilolu ve alkol ve sigara tüketimi gibi yaşam tarzı modellerinde meydana gelen değişikliklerle ilişkilidir. Sağkalım oranlarındaki artışa karşın, toplam meme kanseri kaynaklı ölüm sayıları ve genel ölüm oranları yakın zamanda yalnızca minimal bir azalmaya işaret etmektedir. Bu azalmanın büyük bir kısmı, iyileştirilmiş adjuvan tedavi yöntemlerine bağlı olarak değerlendirilirken, mamografi taramaları sayesinde erken evrede tespit edilebilen meme kanserlerinin tedavisindeki gelişmeler de daha düşük bir etkiye sahiptir. Bu duruma ek olarak, hayatta kalma oranlarındaki artış ile karşılaştırıldığında mortalite oranlarında görülen sınırlı düşüş, daha önceden tespit edilmemiş ve mortalite üzerinde önemli bir etkisi olmayan indolent kanserlerin fazla tanısına ve bunun yol açtığı endişelere neden olmuştur [56, 57].

Meme kanserinin iyileştirilmiş tanı ve tedavisi, ortalama sağkalımı ve relapstan kurtulma sürelerini arttırmıştır, böylece meme kanseri tanısı konan kadınların %85'inin önümüzdeki 5 yıl boyunca hayatta kalacağı tahmin edilmektedir [57]. Bununla birlikte, bir bireyin başarılı bir şekilde tedavi edilmesini engelleyen temel problem, halihazırda kullanılan terapötiklere çoklu ilaç direncinin (MDR) gelişmesidir. Tahminlere göre, hastaların yaklaşık %25-40'ında tedaviye direnç oluşması sonucunda, tedavi edilmesi zor metastatik kanserler ortaya çıkmaktadır. Bu durum, ileri tedavi, tanı ve tarama yöntemlerine rağmen, meme kanserinin hala başlıca ölüm sebeplerinden biri olmasını açıklamaktadır [58].

Meme kanseri tümörlerinin heterojenitesi ve tümör özelliklerinin korunmasında yer alan çok faktörlü sistemler, yeni tedavi yaklaşımlarının gelişmesini engelleyen başlıca nedenler olmaktadır. Çeşitli faktörler, efluks taşıyıcıları, Warburg veya ters Warburg etkisi yoluyla metabolizma değişiklikleri, progenitör kanser hücreleri, moleküler tıp, mRNA düzeyleri ve diğer unsurlar, MDR (çoklu ilaç direnci) fenotipinin oluşumuna katkı sağlayabilir. Geleneksel indirgemeci metodolojinin, meme kanseri tümörlerinin gelişiminde etkili olan moleküler ve biyokimyasal mekanizmaları tam olarak açıklamak için yetersiz kaldığı görülmektedir [59].

4.1 Biyoinformatik

Biyoinformatik, biyolojik sorunların çözümünde veri biliminin kullanılmasını amaçlayan bir disiplindir. Bilgisayar bilimi, istatistik, bilgi bilimi ve uygulamalı matematiğin benzersiz bir karışımı ile entegre edilerek, yaşamın dijital kodlarının veri odaklı metodlarla çözülmesi hedeflenmektedir. Algoritma geliştirilmesi, yazılım araçları oluşturulması ve veritabanları kurulması gibi işlemler aracılığıyla genetik ve genomik bilgilerin büyük miktarları işlenmekte ve yorumlanmaktadır [60].

Geniş çaplı genomik ve postgenomik veri kümeleri, yüksek kapasiteli teknolojiler aracılığıyla, daha önce hipotez testi sistemlerine dayalı olan veri odaklı biyolojik araştırmalara imkan sağlanmıştır. Hesaplama yöntemleri ile büyük veri kümeleri, hastalığa özgü mutasyon veritabanları, genotip-fenotip analizleri, istatistikler, ontolojiler ve daha fazlasıyla bütünleştirilerek biyolojinin dönüşümü gerçekleştirilmektedir. Biyolojik bilginin hesaplama yoluyla tanımlanması, düzenlenmesi ve erişilebilir hale getirilmesiyle biyolojik kavramlar daha katı ve test edilebilir hale gelmiş, biyoloji için yeni bir referans haritası sağlanmıştır. Gelecekte, biyolojik araştırmaların matematiksel, istatistiksel ve hesaplama yöntemleriyle yönlendirilmesi, biyolojiyi nicel bir bilim haline getirilmesi beklenmektedir [61].

Bu alan, gen ifadesi analizi, karşılaştırmalı genomik, moleküler evrim, ilaç keşfi ve kişiselleştirilmiş tıp gibi biyolojik araştırmanın çeşitli alanlarında kritik roller oynamaktadır. Özellikle, DNA/protein dizilim hizalamaları, genetik varyasyon tespiti, işlevsel tahminler, makromoleküler yapı ve işlev tahminleri, gen/biyomoleküler etkileşim ağı simülasyonları, çeşitli biyolojik veritabanlarının entegrasyonu, ilaç keşfi ve geniş çaplı genom çapında ilişkilendirme çalışmaları biyoinformatik yaklaşımları ile gerçekleştirilmektedir. Ayrıca, biyoinformatik araçlar ve platformlar, genetikten proteomik, metabolomik ve mikrobiyomik dahil olmak üzere çeşitli ölçeklerdeki ve çoklu omik verilerin entegrasyonu ve

analizini saęlamakta, bu sayede daha kapsamlı bir anlayıř sunmakta ve yeni tedavi yöntemleri ile terapilerin geliřtirilmesine olanak tanımaktadır [62].

4.2 Tıbbi Genetik ve Genomikte Biyoinformatik

Klinik laboratuvarlar, moleküler genetik testlerini geliřtirmek amacıyla Yeni Nesil Dizileme (YND) teknolojisini kullanmaya bařlamıřtır; bu da veri boyutu ve karmařıklıęında büyük bir artıřa sebep olmuřtur. Bu artıřa yanıt olarak, biyoinformatik; büyük veri kümelerini depolamak, yönetmek, işlemek, analiz etmek, entegre etmek ve yorumlamak için geliřmiř hesaplama yöntemleri, algoritmalar ve yazılımlar kullanılmaktadır [63].

Tıbbi genetikte önemli bir bileřen olarak, biyoinformatik, genetik verileri hasta bakımını iyileřtirecek bilgiye dönüřtürmektedir. Genetik verilerin analizi ve yorumlanması için çeřitli teknolojiler tarafından üretilen büyük miktarlardaki veriler, biyoinformatik yaklařımları kullanılarak ele alınmaktadır. Genetik verilerin iřlenmesi, varyantların tanımlanması, işlevsel etkilerinin tahmin edilmesi ve hastalık fenotipleriyle ilişkilendirilmesi için biyoinformatik araçlar ve algoritmaları kullanılır. Bu bilgiler, çeřitli genetik hastalıkların teşhis, prognoz ve tedavisini geliřtirmekte kullanılmaktadır. Örneęin, kanser genetięinde, biyoinformatik tümör hücrelerindeki genomik deęiřiklikleri analiz etmek için kullanılmakta ve terapötik hedeflerin belirlenmesine yardımcı olmaktadır [64].

Klinik genetik alanında ise, biyoinformatik, hastaların genetik verilerini analiz ederek kalıtsal hastalıkların teşhisine ve kiřiselleřtirilmiř tedavi planlarının geliřtirilmesine katkıda bulunmaktadır. Moleküler genetik testlerin hızını, verimlilięini ve maliyet etkinlięini artıran sürekli geliřtirilen biyoinformatik araçlar ve yaklařımlar sayesinde, geniř ve çeřitli genomik ve fenotipik bilgilerin entegrasyonu saęlanmakta, sonuçların doęruluęu artmakta ve hasta bakımı iyileřmektedir [65].

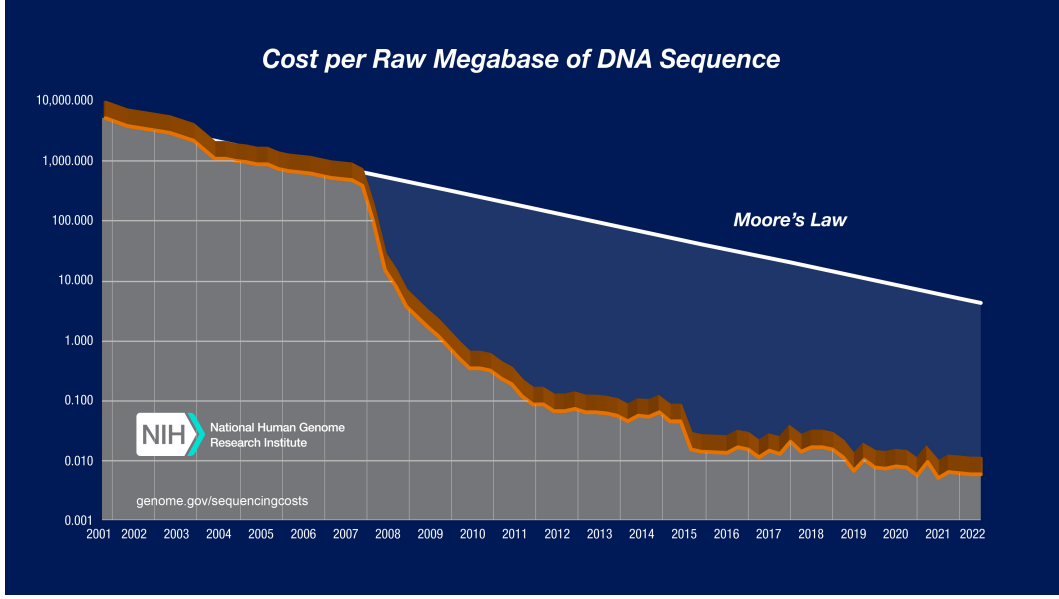
4.3 Yeni Nesil Dizileme

Genel olarak, dizilim DNA moleküllerinin baz dizilerinin belirlenmesi iřlemidir. Genom dizilimi kapasitesi, bilim insanlarına hücrenel mekanizmaları analiz etme ve önemli biyolojik keřifler yapma olanaęı saęlayarak, güncel biyolojik bilimin temel taşlarından biri haline gelmiřtir. İlk deneysel yöntemler geliřtirildięinden beri, DNA dizilimi sadece temel biyolojik arařtırmalarda deęil, aynı zamanda metagenomik, biyoteknoloji, tıbbi tanı ve kiřiye özgü tıp gibi bir dizi uygulamalı

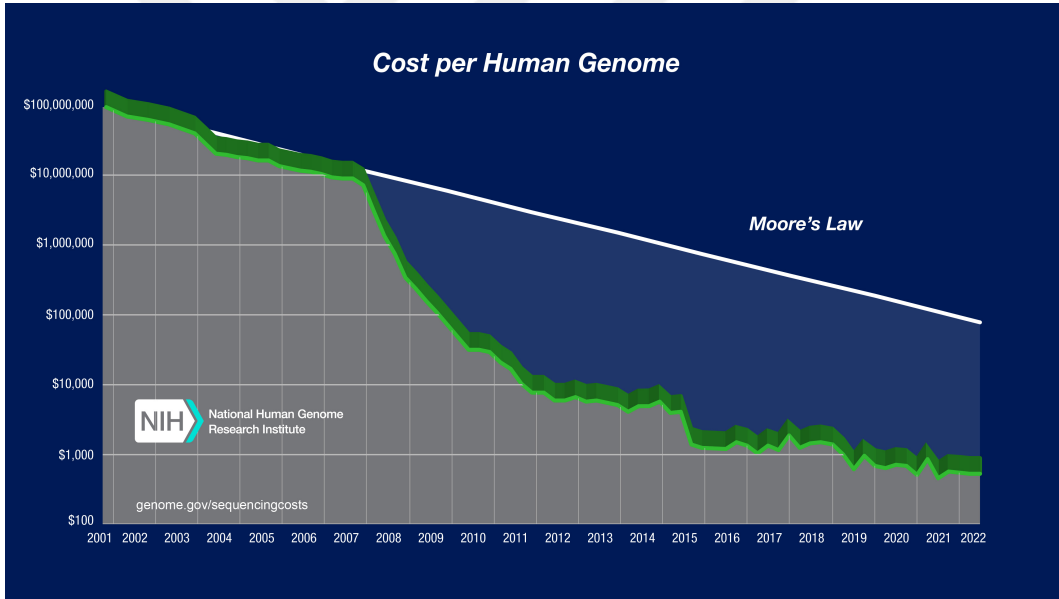
alanda da kritik bir teknolojiye dönüşmüştür [66].

Dizileme teknolojileri, 1977'de ilk bakteriyofaj genomunun kodunun çözülmesine izin veren ve ilk nesil dizileme olarak bilinen Sanger dizileme ile başlayarak, farklı türlerden binlerce genomun şimdi bilindiği noktaya kadar, üssel olarak ilerlemektedir [67]. DNA diziliminin önemli bir dönüm noktası, 2001 yılında ilk insan genomu taslağının tamamlanmasıyla belirlenmiştir. İnsan Genom Projesi kapsamında 1990 yılında başlayan insan genomunun dizilimi, 2001 yılında yayımlanmış ve yaklaşık 2.7 milyar dolarlık bir maliyete ulaşmıştır [68]. Daha sonra bu gelişme, 1000 Genomlar Projesi veya UK10K vb. diğer projeler tarafından yürütülmüştür [69, 70].

2003 yılına kadar, bir baz çiftinin sıralanması maliyetinin 1 dolar olduğu tahmin edilmiştir. 2012'ye kadar, yüksek verimli dizileme teknolojileri, maliyeti milyon bazda yaklaşık 4 cent'e indirmiştir. Son yıllarda, insan genomunun dizilim maliyeti yaklaşık 600 dolara kadar düşmüş ve ilerleyen dönemlerde bu maliyetin 100 dolar civarına daha da düşeceği öngörülmektedir. Bu dramatik maliyet azalışı, genom diziliminin sağlık hizmetleri dahil çeşitli alanlarda daha geniş uygulama bulmasını sağlamaktadır. Özellikle teşhis amaçlı ve kanser ve diyabet gibi karmaşık hastalıkları incelemek ve daha geniş ve detaylı genetik çalışmaların yapılmasını sağlamakiçin araştırmalarda giderek daha fazla kullanılmakta, bu da kişiselleştirilmiş tıp ve genetik hastalıkların anlaşılmasındaki ilerlemelere katkıda bulunmaktadır. Binlerce dolardan yüzlerce dolara düşen bu önemli maliyet azalımı, genomik alanda önemli bir kilometre taşını temsil etmektedir. Gelişmekte olan dizileme teknolojileri, yakın gelecekte maliyetlerin daha da azalacağını öngörmektedir. Bu azalma eğilimi, bilgisayar teknolojilerinde Moore Yasası'na dayanarak gözlemlenen süper üstel düşüşten çok daha hızlı bir şekilde gerçekleşmektedir Şekil 4.1 ve 4.2 [71–73].



Şekil 4.1 DNA dizisinin megabaz(1 milyon baz) başına maliyeti [72]



Şekil 4.2 DNA dizisinin genom başına maliyeti[72]

Dizileme teknolojilerinin geleceğini tahmin etmek zordur. Aynı zamanda, yeni teknolojiler tarafından getirilen iyileştirmeler, daha yüksek kaliteli ve daha uzun okumaların elde edilmesine yönelik hedeflenmiştir. Genomun karmaşık bölgelerini dizileyebilen uzun okumalara duyulan ihtiyaç (yüksek oranda tekrar eden, yüksek GC içeriği vb.) birçok şirket için öncelik haline gelmektedir. Bugün, çok çeşitli dizileme teknolojileri bulunmaktadır. Her bir teknoloji, kendine özgü niteliklere sahiptir ve avantajlarıyla beraber bazı kısıtlamaları da içeren belirli türde okumalar sağlar. Bu sebeple, söz konusu özelliklerin ve bunların uygulamalardaki rollerinin

kavranması önem arz etmektedir. Hedeflenen uygulamaya, bütçeye ve sonuçların istenen özelliklerine bağlı olarak, birkaç farklı teknolojik seçenek arasından seçim yapılabilmektedir [74]

Bu teknolojiler sıklıkla çeşitli adlar ve bazen anlaşılması zor kısaltmalar ile anılır. Derin dizileme, yüksek verimlilikle dizileme (HTS), geniş çapta paralel dizileme (MPS), ikinci, üçüncü, dördüncü veya ileri nesil dizileme (NGS), DNA dizileme ya da bazen sadece bu teknolojiyi ilk geliştiren şirketin adı ile (örneğin, SOLiD, Roche, MinION, İonTorrent, İllumina veya PacBio dizileme) adlandırılmaktadırlar. Bu doktora tezi, meme kanseri yeni nesil dizilimine odaklandığından, bundan böyle basitçe 'dizileme' terimi aksi belirtilmedikçe yeni nesil dizilemeyi(NGS) belirtmek için kullanılmıştır. Bu bölüm yeni nesil dizileme teknolojileri ve işlevlerine örnekler sunmaktadır.

İnceleme kapsamına alınan teknolojiler arasında iki uzun okuma dizisi teknolojisi bulunmaktadır; bunlar, Oxford Nanopore Teknolojileri tarafından üretilen MinION dizileme ve Pacific Biosciences of California, Inc. (PacBio) tarafından geliştirilen tek moleküllü gerçek zamanlı dizileme (SMRT) yöntemleridir. Ayrıca, ThermoFisher Scientific'in sahibi olduğu IonTorrent dizileme ve SOLiD dizileme teknikleri ile Illumina Inc. tarafından geliştirilen HiSeq dizileme gibi çeşitli kısa okuma dizileme teknolojileri de bulunmaktadır.

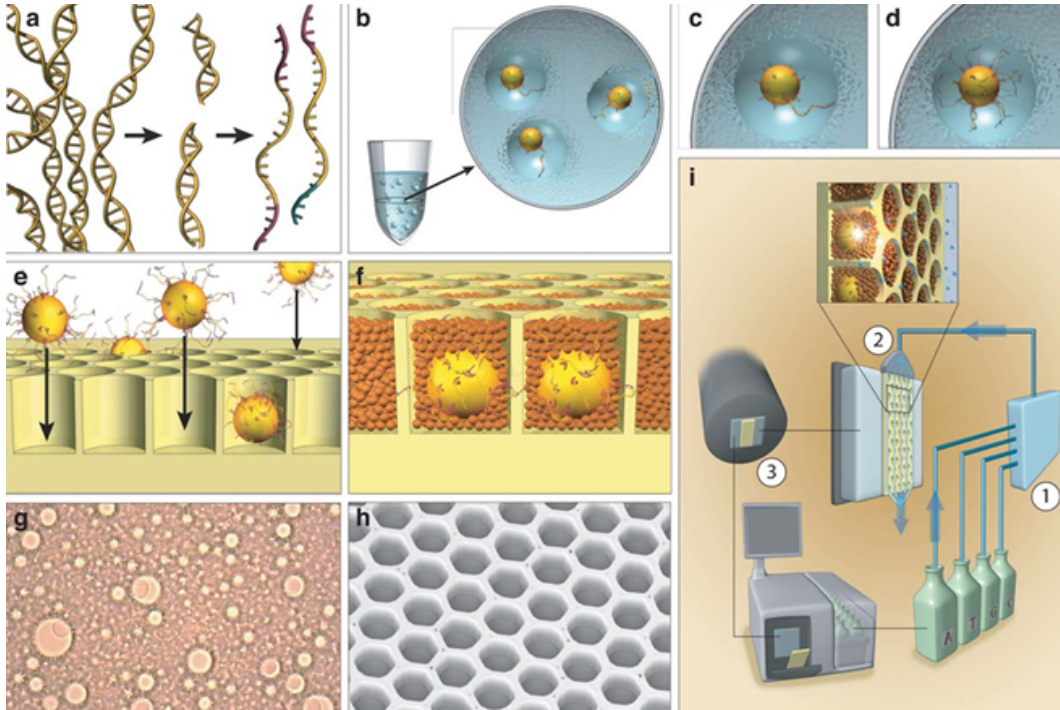
Tüm teknikler için genel dizileme süreci benzer bir akışa sahiptir. Başlangıçta, küçük DNA parçacıklarını içeren bir kütüphanenin hazırlanması zorunludur. Bu amaçla, çıkarılan DNA, küçük parçalara ayrılmaktadır. Daha sonra, dizileme cihazlarının bu parçacıkları kullanabilmesi için, fragmentlara adaptörler eklenmektedir. Çoğu dizileme ekipmanı, DNA'yı belirli bir uzunluk aralığında tutacak şekilde tasarlanmıştır, bu sebeple fragmentlar boyutlarına göre seçilir. Tek bir fragmentten elde edilen sinyal genellikle çok zayıftır. Bundan dolayı, sürecin bir sonraki aşaması PCR amplifikasyonunu içermektedir. Elde edilen son kütüphane, metin dosyaları üreten yeni nesil bir dizileme cihazında dizilmektedir. Bu dosyalar, okunan her bir dizinin girişlerini içermektedir [75].

Yeni nesil dizileme tekniklerinden iki öncü ve yaygın olarak kullanılan metodun özetlenmesi bu bölümde yer almaktadır: Roche ve Illumina teknolojileri.

4.3.1 Roche Genom Sekanslayıcı

Roche tarafından piyasaya sürülen ve 454 şirketi tarafından geliştirilen Genom Dizileyici Cihazı, piyasadaki ilk dizileme cihazlarından biri olmuştur. Her çalışma için 0.04 gigabaz (Gb) kapasiteli ve 500 baz çiftine kadar uzunlukta

okumalar sağlayabilen bu cihaz, mikro kuyucuk teknolojisini kullanarak, dizileme maliyetlerini %50 oranında düşürmüş ve hızı 500 kat artırmıştır, bu da dizileme teknolojilerinde yeni bir dönemin başlangıcı olmuştur. Roche 454 tarafından daha sonra geliştirilen Genome Sequencer FLX + ise, 700 baz çiftine kadar okuma uzunluğu ve çalışma başına 0.7 Gb veri üretme kapasitesi ile daha da gelişmiş bir performans sunmuştur. Bu teknolojiye kütüphane hazırlama ve pirosekanslama yöntemi Şekil 4.3’de detaylandırılmıştır. Bu cihazların sunduğu önemli bir avantaj, uzun okuma uzunlukları olurken, dezavantaj olarak homopolimer bölgelerde yüksek hata oranlarına sahip olmaları gösterilebilir [76].



Şekil 4.3 DNA örneklerinin kütüphane hazırlığı aşamaları ve pirosekanslama yöntemi: (a) Genomik DNA parçalanır, adaptör DNA dizilerine bağlanır ve tek sarmallara ayrılır. (b) DNA, çoğunlukla en fazla tek bir DNA parçası taşıyan PCR başlatıcısı ile kaplanmış mikro boncuklara immobilize edilir. (c) Boncuklar, bir PCR tepkimesinin bileşenleri ile birlikte termostabil bir emülsiyon içinde bölümlere ayrılır ve karışım, konvansiyonel PCR’de olduğu gibi termal döngüye sokulur, bu da başlangıç DNA parçasının on milyon kopyası ile kaplanmış boncuklara yol açar (d). Boncuklar emülsiyondan geri kazanılır, DNA iplikçikleri denatüre edilir ve boncuklar (başlangıç DNA parçasının milyonlarca tek sarmallı kopyasını taşıyarak) bir fiber-optik slaytın kuyucuklarına yerleştirilir (e). Pirosekanslama için gerekli enzimleri taşıyan daha küçük boncuklar her bir kuyucuğa yerleştirilir (f, g) Termostabil bir emülsiyonun temsili ve mikroskop fotoğrafı. (h) Boncuk yerleştirilmeden önce fiber-optik kaplama ve kuyucukları gösteren bir fiber-optik slaytın taramalı elektron mikrogramı. (i) Pirosekanslama cihazı, (1) bir sıvı montajı, (2) yerleştirilmiş boncukları içeren fiber-optik slaytı içeren bir akış odası ve (3) bir CCD kamera bazlı görüntüleme montajı ve bilgisayar işlem birimi içerir [77]

4.3.2 İllumina

İllumina dizileme cihazı (başlangıçta Solexa tarafından geliştirilen) HiSeq, en popüler dizileme cihazıdır. Versiyon 2500, bir parçanın her iki yanından (çift uçlu dizileme) 125 bazına kadar dizileme yapabilir ve maksimum 1000 Gb / run çıkışı oluşturabilir. İllumina ayrıca 2x150 baz ile eşleştirilmiş uçtan ve toplam 120 Gb çıkıştan üretebilen NextSeq'i geliştirdi. 15 Gb ve 2x300 bazları ile MiSeq, İllumina'nın en küçük cihazıdır [78]. İllumina tarafından 2014 yılından itibaren kullanıma sunulan HiSeq X dizileme platformu, her biri 150 baz çifti uzunluğunda olan çift yönlü okumalar aracılığıyla, 1.800 gigabaza kadar varan genişlikte yaklaşık 1.000 genomun dizilenmesine olanak tanımıştır. HiSeq serisi ve bu serinin devamı niteliğindeki cihazlar, yüksek hacimli okuma üretme konusunda üstün kabiliyetlere sahiptir, ancak ürettikleri okumaların uzunlukları oldukça kısıtlıdır. Ayrıca, bu cihazların kullandığı dizileme-sentez metodolojisi, GC içeriği konusunda bir sapmaya neden olmaktadır. Şekil 4.4, İllumina'nın dizileme iş akışını göstermektedir. Kütüphane hazırlama işleminin başlangıç aşamaları, Roche cihazlarının kullandığı prosedürlerle paralellik göstermektedir [79].

Örnek hazırlık sürecinde, DNA molekülü parçalara ayrılmakta ve bu parçalar boyutlarına göre sınıflandırılmaktadır (boyut ayrımı). Ardından, bu fragmanların her iki ucuna adaptörler eklenmekte ve daha sonra bir akış hücresine alınmaktadır. Akış hücresi, eklenen adaptörlere uygun tamamlayıcı probler içermektedir, bu da adaptörlerin akış hücresine bağlanmasını (hibritlenmesini) mümkün kılmaktadır. Parçaların serbest uçları, köprü şeklini oluşturan akış hücresindeki tamamlayıcı adaptörleriyle bağlanmaktadır.

Şablon DNA daha sonra, aynı dizilimden oluşan kümelerin akış hücresi içerisinde oluşumunu sağlayan köprü amplifikasyonu (PCR temelli bir teknik) ile çoğaltılır. Kümeler oluşturulduğunda, her seferinde bir nükleotidin okunmasını mümkün kılan geri dönüşümlü terminator kimyasalları kullanılır. Bu süreçte, DNA zincirlerinin her birine dört farklı türde geri dönüşümlü terminator baz (RT-baz) eklenerek zincirler genişletilir. Bir yüksek çözünürlüklü kamera, floresan işaretli nükleotidlerin görüntülerini çekmektedir. Sonrasında, boyalar kimyasal yöntemlerle uzaklaştırılmakta ve dizilim tamamen okununcaya kadar bu işlem tekrar edilmektedir. Bu işlem, milyonlarca kümede eş zamanlı olarak gerçekleşmekte, bu da teknolojinin saniyede birkaç milyon nükleotidi hızla dizilemesine olanak tanımaktadır. Sonuç olarak, elde edilen tüm okuma dizileri, dizileme cihazının gerçekleştirdiği toplam döngü sayısına eşit uzunluktadır.

Bu teknoloji, günümüzde mevcut olan en yüksek doğrulukta dizileme tekniklerinden biri olmasına rağmen, elde edilen okumalar ara sıra dizileme

a Illumina's library-preparation work flow

DNA fragments

Blunting by fill-in and exonuclease

P Phosphorylation

A Addition of A-overhang

Ligation to adapters

b

Grafted flow cell

Template hybridization

Initial extension

Denaturation

First cycle annealing

First cycle extension

Second cycle denaturation

Second cycle annealing

Second cycle extension

c

Emission

Excitation

Incorporate
Detect
Deblock
Cleave fluor

Fluor

Cleavage site

Block

OH free 3' end

Cluster amplification

P5 linearization

Block with ddNTPs

Denaturation and sequencing primer hybridization

24

4.4 RNA-seq

RNA-seq teknolojisi, transkribe edilen RNA'nın sıralanması, haritalanması ve miktarının belirlenmesi yoluyla tanımlanmakta ve geniş çapta transkriptom analizi için etkili bir metot sunmaktadır. Bu teknik, mRNA'lar, kodlamayan RNA'lar ve küçük RNA'lar dahil olmak üzere gen ekspresyonunu ölçme ve RNA türlerini tanımlama amacıyla kullanılmaktadır [83]. RNA-seq verileri, gen ifade modellerini ortaya koymakla kalmaz, aynı zamanda birçok yeni mRNA transkripti izoformunun ve alternatif dizilim mekanizmalarının keşfini de desteklemektedir [84, 85].

RNA-seq, hibridizasyon tabanlı tekniklere kıyasla birkaç önemli üstünlük sunmaktadır. İlk olarak, analiz edilecek genom hakkında önceden bilgi gerektirmez; bu özellik, dizisi henüz çıkarılmamış organizmaların transkriptomlarının sıralanmasını ve yeni transkriptlerin belirlenmesini mümkün kılmaktadır. İkinci olarak, düşük gürültü düzeyi ile yüksek tekrarlanabilirlik sunmaktadır. Ayrıca, geniş bir dinamik aralığa sahiptir ve gen ekspresyonunun nicel olarak ölçülmesinde etkin bir platform olduğu kanıtlanmıştır. Sonuç olarak, RNA-seq, tüm transkriptomun yüksek verimlilikle ve kapsamlı bir şekilde incelenmesine imkan tanıyan ilk dizileme yöntemi olarak kabul edilmektedir [86, 87].

4.5 Gen İfadesi Analizi İçin RNA-seq Veri Hazırlama

Transkriptom montajı, milyonlarca kısa okumanın dizileme platformları tarafından üretilerek bir referans genom üzerine yeniden yapılandırılması veya haritalanmasıyla gerçekleştirilmektedir; bu süreç, transkript ekspresyonunun incelenmesi veya gen kantifikasyonunun yapılması için gerekli olan tam bir transkriptom anotasyonu ve montajını gerektirmektedir [88, 89].

RNA-seq veri analizinin tüm adımları içinde merkezi bir yer tutan transkriptom montajı, referans tabanlı veya de novo montaj kullanılarak yapılabilmektedir. Referans tabanlı montajda, okumalar daha önce dizilmiş veya mevcut bir referans genom üzerine geri haritalanmaktadır; de novo montajda ise, referans genom kullanımına gerek kalmadan ifade edilen izoformların yeniden yapılandırılması için okumalar karşılaştırılmaktadır. Yüksek kaliteli bir referans genomunun bulunmadığı durumlarda, okumaların örtüşmesiyle transkriptler monte edilmektedir [90].

DeneySEL sonuçların biyolojik olarak yorumlanmasında fonksiyonel anotasyon büyük önem taşımaktadır. Büyük dizileme verileri çeşitli sınırlamalara sahip olduğu için birçok anotasyon aracı bulunmaktadır. Örneğin anotasyona dayalı

veri madenciliği işlevselliği entegre edilen ve aynı zamanda otomatik ve yüksek hacimli dizi anotasyonu sağlayan bir biyoenformatik araç olan Blast2GO (B2G), tasarlanmıştır. B2G, BLAST yapılması, haritalama, anotasyon, istatistiksel analiz ve görselleştirme olmak üzere, genellikle beş adımdan oluşan otomatik bir süreçten meydana gelmektedir [91].

4.6 Diferansiyel İfade Analizi

Gene ekspresyon analizi, belirli bir hücrede veya belirli koşullar altında genlerin transkripsiyon seviyesindeki motiflerinin incelenmesini kapsamaktadır. Yoğun paralel imza dizilimi, seri gen ekspresyon analizi, RNA dizilimi ve mikroarray ekspresyon profillemeye kullanılarak yapılan transkript bolluğu analizi, genomun işleyişi hakkında benzeri görülmemiş bir bakış açısı sunmuştur [92, 93]. Genlerden transkribe edilen RNA'ların göreceli bollukları, ilgili genlerin ifade seviyelerini, belirli bir gelişim aşaması veya fizyolojik durum için yansıtmaktadır. Transkripsiyon-çeviri sürecinin son ürünleri olmayan RNA'lar, gene ekspresyonu ve diferansiyel gen ekspresyonunun çalışılmasında, incelenen hücre durumları hakkında önemli bilgiler ortaya çıkarabilmektedir. Uygun biyoenformatik araçlarla birleştirildiğinde, RNA-seq, transkriptlerin ve izoformların ifadelerini tanımlamak ve nicelendirmek için harika bir yöntem sağlamaktadır. Eğer bir RNA-seq çalışmasının amacı diferansiyel ifade edilen genlerin tespitiyse ve iyi anotasyonlu bir transkriptom montajı mevcutsa, temel bir veri işleme boru hattı şu adımlardan oluşacaktır: okuma eşleme, okuma sayımı, sayımların normalleştirilmesi ve bu genlerin tespiti [94, 95].

Diferansiyel ifade analizi, normalleştirilmiş okuma sayımı verilerini alıp, deneysel koşullar arasında önemli ölçüde değişen genleri vurgulamak için istatistiksel analiz yapılmasını ifade etmektedir [96]. RNA-Seq verilerine dayalı olarak farklı tedavi grupları arasında farklılaşmış genlerin tanımlanması için birçok yazılım paketi geliştirilmiştir. En yaygın kullanılan yazılım paketleri şunlardır: edgeR, DESeq, LIMMA(Linear Models for Microarray and RNA-seq Data). LIMMA, gen ifadesi değişikliklerini tespit etmek için geliştirilmiş, mikro dizi ve RNA-seq verileri için önerilen istatistiksel yöntemler sunan etkili bir araçtır [97]. Bu çalışmada, diferansiyel reaksiyon ekspresyonu analizi yapmak amacıyla LIMMA kullanılmıştır.

4.7 LIMMA Metodu

LIMMA, mikro dizi verilerinde diferansiyel ifade analizini gerçekleştirmek için kullanılan istatistiksel yöntemleri içeren bir pakettir. RNA-seq sayım verilerini kullanmadan önce, bu veriler voom dönüşümü ile LIMMA için uygun hale getirilmektedir. LIMMA-Voom yöntemi (kalite ağırlıklı veya ağırlıksız), sayım verilerini normalleştirmekle işe başlamaktadır. Law ve arkadaşları milyonda sayım normalizasyonunu(FPM) önermişken, günümüzde yaygın olarak M-değerlerinin Budanmış Ortalaması(TMM) normalizasyonu tercih edilmektedir [98].

TMM normalizasyonu, farklı örnekler arasında okuma derinlikleri değişiklik gösterdiğinde bu değişiklikleri düzeltmek için kullanılmakta ve özellikle RNA-seq verilerindeki sistemik varyasyonları düzeltmek amacıyla tasarlanmıştır. Bu normalleştirme işlemi sonrasında, büyük sayımların daha yüksek varyans gösterdiği durumlarda, normal tabanlı istatistiksel analizlerde kullanılmasının karşılaştığı temel zorluk, değişkenliklerin eşit olmamasıdır. Log2 ile yapılan dönüşüm, bu sorunu dengelemeye yardımcı olmakta, ancak bazen düzeltmeyi aşırı yapabilmektedir. Büyük log dönüşümlü normalleştirilmiş sayımlar için varyans nispeten sabit kalırken, küçük ve orta büyüklükteki log dönüşümlü normalleştirilmiş sayımlar için yumuşak bir azalma eğilimi göstermektedir [99].

LIMMA hattına sunulmadan önce log dönüşümlü normalleştirilmiş sayımlar için ortalama ile varyans arasındaki ilişkinin belirlenmesi gerekmektedir. LIMMA-Voom, bu log dönüşümlü normalleştirilmiş sayımlarda gen seviyesindeki değişkenlik tahminlerini tüm gözlemlere genellemek yerine, her bir gözlem için ayrı ayrı modellemektedir. Bu eğilim, parametrik olmayan yöntemlerle hesaplanmaktadır. Ana zorluk, gözlemsel seviyede yeterli tekrarın bulunmaması nedeniyle varyansların doğru bir şekilde tahmin edilememesidir. Buna çözüm olarak, gen seviyesindeki ortalama-varyans ilişkisi tahmin edilmekte ve bu tahmin, bireysel gözlemlerin varyanslarını hesaplamak için interpolasyon yöntemiyle kullanılmaktadır. Log dönüşümlü sayımlardaki ortalama-varyans bağıntı gidermek için, işlem süresince ters varyanslar ağırlık olarak kullanılmaktadır. Kalite ağırlıklı LIMMA-Voom, bu işlemi örnek kalitesindeki farklılıkları dikkate alarak belirlenen örnek ağırlıklarıyla daha da genişletmiştir.

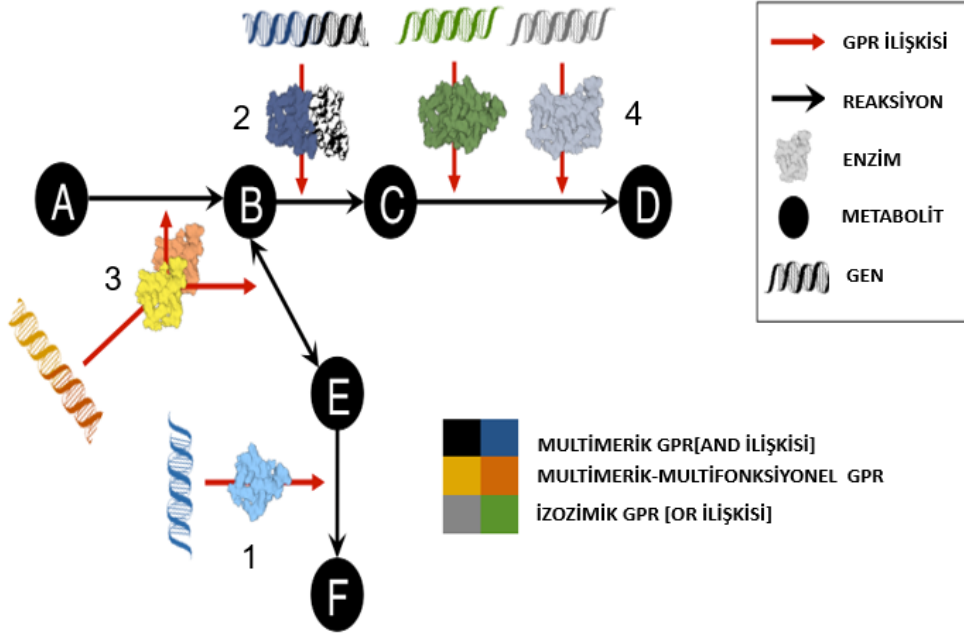
Nihai aşamada, log dönüştürülmüş normalleştirilmiş sayımlar ve bunlarla ilişkili ağırlıklar, diferansiyel ekspresyon analizi için geleneksel LIMMA sürecine entegre edilmektedir. Bir lineer model formüle edilmekte ve diferansiyel ifadeyi değerlendirmek için ampirik Bayes yöntemine dayalı bir t-testi kullanılmaktadır; bu yöntem, standart hatalar ve serbestlik dereceleri üzerinde ayarlamalar yapmaktadır. Standart hatalar üzerine ampirik Bayes yumuşatması uygulanarak, genler arası

bilgilerden yararlanılmaktadır. Ayrıca, serbestlik dereceleri, model için önceden belirlenen serbestlik derecesi miktarını ifade eden bir parametre ile ayarlanmaktadır. Elde edilen p-değerleri, Benjamini ve Hochberg yöntemi kullanılarak çoklu testler için düzeltilmektedir [100, 101].

4.8 Gene-Protein-Reaksiyon Kuralları

Gene-Protein-Reaksiyon(GPR) kuralları, genlerin ve proteinlerin biyokimyasal reaksiyonlarla nasıl ilişkilendirildiğini tanımlayan lojik kurallardır. Bu kurallar, özellikle metabolik ağlarda, genler ve reaksiyonlar arasında doğrudan bir ilişki olmaması nedeniyle önem taşımaktadır. mRNA bolluklarının biyokimyasal ağlara yerleştirilmesi sırasında, GPR kuralları hangi genlerin bir reaksiyonu katalizleyen enzim aktivitesinin ana belirleyicileri olduğunu tanımlamak için kullanılmaktadır. Bu ilişkilerin anlaşılması, özellikle sistem biyolojisi ve metabolik mühendislik bağlamında, organizmaların genetik içeriğine dayanarak metabolik davranışlarını yorumlamak ve tahmin etmek için hayati öneme sahiptir [102].

GPR ilişkileri, genlerin ve enzimlerin reaksiyonlara nasıl bağlandığını açıklamaktadır Şekil 4.5 Bu ilişkiler, proteinlerin yapısal ve işlevsel çeşitliliğini ve gen ifadesinin bu yapıları nasıl etkilediğini ortaya koymaktadır. Özellikle multimerik enzim kompleksleri için, en yaygın varsayım, enzim aktivitesini en düşük seviyede ifade eden genin belirlediğidir. İzoenzimler söz konusu olduğunda, aktivite tüm izoenzim genlerinin toplam ifadesine veya en yüksek ifade edilen izoenzim genine bağlı olmaktadır. Bu durum, gen ifadesi düzeylerinin enzim aktivitesi üzerindeki etkisini daha iyi anlamamıza olanak tanımıştır. Bu ilişkilerin anlaşılması, sistem biyolojisi ve metabolik mühendislik bağlamında, organizmaların genetik içeriğine dayanarak metabolik davranışlarını yorumlamak ve hastalık modelleme gibi alanlarda uygulamalı araştırmalar için hayati önem taşımaktadır. Böylece, spesifik gen değişikliklerinin metabolik yollar üzerindeki etkileri daha doğru bir şekilde anlaşılakta ve manipüle edilebilmektedir [103].



Şekil 4.5 GPR kurallarının görsel şeması: (1) Bir enzim bir reaksiyonu katalize etmektedir(açık mavi), (2) Çoklu protein alt birimleri tarafından oluşturulan bir enzim bir reaksiyonu katalize etmektedir(multimerik GPR, koyu mavi ve siyah), (3) Çoklu protein alt birimlerinden oluşan bir enzim birden fazla reaksiyonu katalize etmektedir(multimerik – çok fonksiyonlu GPR, sarı ve turuncu), (4) Birden fazla enzim bir reaksiyonu katalize etmektedir(izoenzimatik GPR, gri ve yeşil)[103]

GPR modellerinin oluşturulması sürecinde, genetik ve proteomik veriler KEGG, MetaCyc ve UniProt gibi veritabanları kullanılarak entegre edilmektedir. Bu modeller, metabolik ağların yeniden yapılandırılmasında merkezi bir rol oynar ve genom ölçekli metabolik modellerin (GEM'ler) temelini oluşturmaktadır. GPR kurallarının derinlemesine anlaşılması, sistem biyolojisi ve sentetik biyoloji uygulamaları için özellikle kritiktir. Örneğin, metabolik mühendislikte, belirli bir metabolit üretimini maksimize etmek için gerekli enzim yollarının modifikasyonu, GPR kuralları aracılığıyla yönlendirilebilmektedir. GPR kuralları sayesinde, belirli genlerin modifikasyonunun veya ifadesinin değiştirilmesinin genel metabolik akış üzerindeki etkileri belirlenebilmektedir. Böylece, hücre içi dinamiklerin daha etkin kontrolü ve istenmeyen yan ürünlerin oluşumunun önlenmesi sağlanmaktadır [104]

GPR kuralları, genetik bilgi ve protein işlevselliği arasındaki karmaşık ilişkileri çözümlmek için de kullanılmaktadır. Bu kurallar, hastalık durumlarında gen ifadesi değişikliklerinin patolojik süreçlere nasıl katkıda bulunduğunu anlamada kullanılmaktadır. Örneğin, kanser biyolojisinde, tümörlü dokularda gözlemlenen gen ekspresyon profilleri, GPR analizi ile değerlendirilerek, kanser ilerlemesinde rol oynayan kritik metabolik yollar belirlenmektedir. Bu yollar, potansiyel terapötik

hedefler olarak işaretlenmektedir[105].

Sonuç olarak, GPR ilişkileri, genetik bilginin metabolik ağ analizlerine entegre edildiği çerçeveyi oluşturmakta ve biyolojik sistemlerin anlaşılması ve manipüle edilmesinde güçlü bir araç sunmaktadır. Hesaplama ve deneysel metodolojiler ilerledikçe, gelişen bilimsel yöntemler ve deneysel veri setleri, bu modellerin doğruluğunu ve güvenilirliğini artırarak, onların biyolojik sistemler hakkında daha kesin bilgiler sağlamasına olanak tanımaktadır.

4.9 Kullanılan Veri Tabanları

Yüksek kapasiteli teknolojilerin gelişimi ve maliyetlerindeki düşüş; laboratuvarlarda çeşitli omik verilerin uygun maliyetle üretilmesine olanak tanımıştır. Son on ila on beş yıl içindeki süreçte, kanser teşhis, tedavi ve önleme konusunda artan bir ilgi gözlemlenmiştir; bu bağlamda, tümörlerdeki genetik değişimlerin derinlemesine incelenmesi hedeflenmiştir. Biyoinformatik ve genetik dizilimin teknolojik ilerlemeleri, kanser genomu üzerine yeni anlayışlar sağlamış ve dünya genelinde çeşitli amaçlarla başlatılan konsorsiyum projeleri, büyük örneklem büyüklükleri ve yüksek kalitede omik veriler üreterek bilim topluluğuna hizmet vermeye başlamıştır. Bu çerçevede, 2005 yılında, kanser genetiği üzerine genomik analiz teknolojilerinin kullanımını sağlayarak daha detaylı bir kavrayış sunmayı amaçlayan Kanser Genom Atlası (TCGA) gibi önemli projeler hayata geçirilmiştir. Bu proje, yeni kanser tedavi yöntemleri, tanı araçları ve önleyici stratejiler geliştirilmesine olanak tanımaktadır [106].

TCGA, Amerika Birleşik Devletleri hükümeti tarafından desteklenmektedir ve Kanser Genomik Merkezi ile birlikte Ulusal Kanser Enstitüsü ve İnsan Genomu Araştırma Ulusal Enstitüsü tarafından yürütülmektedir. Proje başlangıçta glioblastoma multiforme, akciğer ve yumurtalık kanseri üzerine yoğunlaşmış olup, zamanla 33 farklı kanser türünün karakterizasyonu ile ilgili DNA, RNA, protein ve epigenetik düzeylerde kapsamlı veriler sağlamıştır. TCGA verileri Genomik Veri Merkezi'nde (GDC) muhafaza edilmektedir [20, 107].

Metabolik modelleme ve yüksek kaliteli omiks verilerinin yaygın olarak sağlanması, prokaryotların yanı sıra insanlar ve diğer ökaryotik organizmalar için de metabolik modellerin geliştirilmesine olanak tanımıştır. İnsan metabolizmasının ilk genom ölçeğinde modeli olan Recon 1, 2007 yılında oluşturulmuştur. Bu model, 1496 gen, 2004 protein ve 2766 metaboliti kapsar ve toplamda 3742 reaksiyonu içermektedir; bu reaksiyonlar insan metabolik bilgi tabanının kapsamlı bir değerlendirmesi ile ve bibliyomik verilerle yeniden yapılandırılmıştır. Ayrıca,

model farklı hücre ve dokulardaki 288 bilinen metabolik işlevin simülasyonları ile doğrulanmıştır. Bu bilgiler BIGG veritabanında yer almaktadır [108, 109].

Recon 1'in ardından, insan metabolizmasının yeni bir yeniden yapılandırması gerçekleştirilip 2010 yılında yayımlanmıştır. Edinburgh İnsan Metabolik Modeli (EHMM), çeşitli veri tabanlarından alınan genomik anotasyonların bir araya getirilmesi ve literatürden elde edilen metabolik reaksiyon bilgileriyle manuel olarak oluşturulmuştur. Başlangıçta, yazarlar KEGG, UniProt ve HGNC veritabanlarından önemli bilgiler toplamışlar [110–112]. Ardından, Enzimler ve Metabolik Yollar veritabanları kullanılarak daha fazla veri elde edilmiştir. EHMM, yaklaşık 70 özgün insan metabolik ağını kapsayan, 2314 gen ve 2823 reaksiyon içeren yeniden düzenlenmiş bir yapıyı sunmaktadır. Model, metabolik reaksiyonların farklı hücresel organellere tahsis edilmesiyle ve metabolitlerin hücresel organeller arasında taşınmasını sağlayan taşıma reaksiyonlarının eklenmesi ile zenginleştirilmiştir [113, 114].

2012 yılında, iHuman1512 adıyla yeni bir metabolik model tanıtılmıştır; bu model, İnsan Metabolik Reaksiyon (HMR) veritabanı üzerine kuruludur. Bu veritabanı, daha önce geliştirilen iki modelin entegrasyonu ve KEGG ile HumanCyc'ten elde edilen bilgilerle zenginleştirilerek genişletilmiştir. Modelin oluşturulma sürecinde, veriler elle düzenlenmiş, dış veritabanlarında yer almayan metabolitler ve reaksiyonlar çıkarılmıştır. HMR, HepatoNet1, iAdipocytes1809, LIPID MAPS Lipidomik Geçidi ve REACTOME gibi spesifik insan metabolik modelleri ile entegre edilerek HMR 2.0 versiyonunu oluşturmuştur[115–119]. HMR 2.0, tamamlandığında 3675 gen, 6007 metabolit (3160'ı benzersiz) ve 8181 reaksiyon içermektedir. Bu modeller, İnsan Metabolik Atlas (HMA) veritabanında yer almıştır [120].

2013'te, Thiele ve diğerleri Recon 1 modelini gözden geçirme önerisinde bulunmuş ve bu model Recon 2 olarak adlandırılmıştır [121]. Recon 2, EHMM ve HepatoNet1 modellerini içermekte ve ayrıca yağ asidi oksidasyonu ve küçük bağırsak enterositlerinin yeniden yapılandırılmasına yönelik modüller eklemiştir [122, 123]. Recon 2'nin bir sonraki güncellemesi, 2016 yılında Recon 2.2 olarak yayımlanmıştır; bu güncelleme 5324 metabolit, 7785 reaksiyon ve 1675 gen içermektedir [124].

Son gelişmeler ışığında, insan metabolik modelleme alanında önemli ilerlemeler kaydedilmiştir. Recon 2.2'nin yayımlanmasının ardından, bu serinin en son güncellemesi olan Recon 3D, 10130 reaksiyon ve 5835 metabolit içerecek şekilde genişletilmiştir [125]. HMA veri tabanına model organizmaların eklenmesi ile

ismi yenilenen Metabolik Atlas'ın geliştirilmesi, hücresel metabolizmayı interaktif olarak keşfetmek ve omik verileri entegre etmek için çevrimiçi bir platform olarak önemli bir adım olarak kabul edilmektedir. Metabolik Atlas, insan dokuları ve tümörlerden *S. cerevisiae* gibi model organizmalara kadar 350'den fazla Genetik Mühendislik Modeli (GEM) içeren bir depo olarak hizmet vermektedir ve bu modeller organizma adı, durum, yayın yılı gibi çeşitli bilgilerle özetlenmektedir. Bu özellikler, Metabolik Atlas'ı, metabolizmayı incelemek için değerli bir kaynak ve sezgisel bir araç haline getirmekte ve insan sağlığı ve hastalıkları üzerine araştırmalar için değerli bir altyapı desteği sağlamaktadır [126].

4.10 R Yazılım Platformu

R yazılım platformu, özellikle büyük veri setlerinin analizi ve genom çapında ilişki çalışmaları (GWAS) gibi kompleks biyoistatistik projelerde sıklıkla tercih edilen bir araçtır. Açık kaynaklı ve ücretsiz bir yazılım olan R, kullanıcılarına geniş bir kütüphane sunmaktadır. Bu kütüphaneler, veri işleme, istatistiksel modelleme ve grafik çizme gibi temel işlevlerinin yanı sıra daha spesifik ihtiyaçlara yönelik fonksiyonları kolaylıkla uygulamalarına olanak tanımaktadır [127].

R'nin programlama dili, veri analizi ve istatistiksel hesaplamalar için optimize edilmiştir ve C, C++, Fortran ile entegre çalışabilir. Ayrıca, Python, SQL ve JavaScript gibi diğer popüler programlama dilleriyle de entegrasyon sağlanmaktadır. Bu entegrasyonlar, R'nin kullanım alanını genişletmekte ve farklı programlama ortamlarında çalışan kullanıcılar için esneklik sağlamaktadır [128].

Genetik ve genomik araştırmalarda, R'nin sunduğu olanaklar Bioconductor gibi özel platformlar aracılığıyla genişlemiştir. Bioconductor, genetik ve genomik verilerin analizi için entegre araçlar ve veri kümeleri sağlayan açık kaynaklı bir projedir. DESeq, edgeR, LIMMA gibi paketler, RNA dizilimi verilerinden gen ifadesi değişikliklerini saptamak için kullanılmakta ve normalizasyon, varyans stabilizasyonu gibi işlemleri içermektedir [129]. R ve Bioconductor'un entegre işlevselliği, genetik araştırmaların geleceğine yön veren yenilikçi bulguların ortaya çıkmasında kritik bir role sahiptir. Özellikle, genetik varyasyon analizi için 'VariantAnnotation' ve 'ggbio' gibi paketler geniş kullanım bulmakta, genetik verileri okumak, işlemek ve anlamlandırmak için araçlar sağlamaktadır [130].

R'nin esnekliği ve genişletilebilirliği, akademik çevrelerde ve çeşitli endüstrilerde tercih edilen bir araç haline gelmesine neden olmuştur. Farmakogenomik ve epidemiyolojik çalışmalar, R'nin güçlü istatistiksel analiz kabiliyetlerinden faydalanmaktadır. Bu alanlarda, R'nin kullanımı, karmaşık veri kümeleri üzerinde

doğru ve etkili analizler yapılmasını sağlayarak, daha doğru sağlık tahminleri ve tedavi stratejileri geliştirilmesine olanak tanımaktadır [131].

R'yi etkin bir şekilde kullanabilmek için RStudio gibi gelişmiş bir entegre geliştirme ortamı (IDE) kullanmak faydalıdır. RStudio, kod düzenleme, hata ayıklama ve veri görselleştirme işlemlerini kolaylaştırarak kullanıcı deneyimini iyileştirir [132]. Ayrıca, RMarkdown ve Shiny gibi paketler aracılığıyla analiz sonuçlarını raporlaştırma ve interaktif web uygulamaları oluşturma imkanı sunar. Son yıllarda, büyük veri, yapay zeka ve makine öğrenmesi gibi alanlarda da popüler bir araç haline gelmiştir. Özellikle, tidyverse paket koleksiyonu, veri manipülasyonu ve temizleme için güçlü araçlar sunmaktadır. ggplot2, veri görselleştirme için kullanılan başka bir temel R paketidir ve kullanıcıların karmaşık grafikleri kolayca oluşturmaya olanak tanımaktadır [133–136].

Ek olarak, R'nin geniş kullanıcı ve geliştirici topluluğu, sürekli olarak platformu geliştirmekte ve yeni paketler eklemektedir. Bu dinamik topluluk, R'nin çeşitli disiplinlerde ve sektörlerde uygulanabilirliğini artırarak, onu çok yönlü ve değerli bir araç yapmaktadır. Bu topluluk tarafından desteklenen sürekli gelişim, R'yi hem akademik hem de ticari kullanım için ideal bir seçenek haline getirmiştir [137].

5.1 Veri Kaynakları

Bu araştırma, Diener ve arkadaşlarının TCGA veritabanından elde ettikleri, 33 kanser türü içeren hastaların ve sağlıklı örneklerin verilerini analiz eden çalışmalarına dayanmaktadır [20, 138]. Ancak bu çalışmada sadece meme kanseri ile ilgili verileri ayrıntılı olarak incelenmiştir. Metabolik analizler, HMA veritabanından alınan verilerle gerçekleştirilmiştir [120]. Araştırmada kullanılan veriler, kanser genom analizlerinde yaygın olarak kullanılan standart veri setlerinden biridir ve bu analiz R programlama dili kullanılarak yapılmıştır. [128].

5.2 Reaksiyonlar ve Bileşiklerin Oluşturulması

5.2.1 Veri Kaynakları ve Tümör Profili Analizi

Metadata dosyasının oluşturulmasında, "sample_raw" dosyasından seçilen "file_id" (hasta dosya numarası) ve hastaların tümör durumunu lojik olarak ifade eden "tumor" kolonları seçilmiştir. Hastaların tümör profillerini daha ayrıntılı incelemek için kritik öneme sahip olan bu seçim, her bir hastanın tümör durumunun yanı sıra, özellikle genetik ve kanser araştırmalarında, bireysel hasta verilerinin detaylı analizini gerektiren durumlarda önem taşımaktadır. Bu veriler tidyverse, DT, ggraph, tidygraph ve knitr paketleri kullanılarak işlenmiştir [135, 139–142].

5.2.2 TCGA Veri Setinin İşlenmesi ve Analizi

TCGA veritabanından elde edilen BRCA dosyası, 60,483 gen ve 1222 hastayı kapsayan bir matris olarak bulunmaktadır. "brca_counts" adı verilen bu veri setindeki sütuna "file_id" adı verilerek, bu sütunun metadata dosyasındaki "file_id" sütunuyla eşleşmesi sağlanmıştır. Bu adımla, analiz sürecinde "inner join" komutu kullanılarak sadece bu iki sütunun eşleşen verileri üzerinde çalışılması mümkün hale gelmiştir. Yeni oluşturulan "tcga_brca" veri setinde, gelecekteki analizler için

önemli bir ayrım bulunmaktadır: 1109 hasta örneği ve 113 sağlıklı örnek. Bu ayrım, hasta ve sağlıklı bireyler arasındaki biyolojik ve genetik farklılıkları incelemek için kritik bir temel oluşturmaktadır.

5.2.3 Metabolik Reaksiyonların Ayrıştırılması ve Analizi

HMA web sitesinden indirilen "HumanModel" dosyasında bulunan gen ve reaksiyon verileri işlenerek veri setleri oluşturulmuştur. Bu işlemde, excel dosyasının ilk sayfasındaki veriler okunmuş ve "select_if" fonksiyonu ile boş değer içeren sütunlar veri setinden çıkarılmıştır. HMA'dan alınan bu veriler üzerinde şekil 5.1'de gözüktüğü gibi çeşitli veri ön işleme adımları gerçekleştirilmiştir. Bu işlemlere, reaksiyonların geri dönüşümlü ve geri dönüşümsüz olarak ayrılması ile başlanmıştır. Bu süreçte, çift yönlü reaksiyonlar, daha kolay işlenebilir olmaları için iki ayrı tek yönlü reaksiyona dönüştürülmüştür. Ardından, reaksiyon denklemleri, substrat ve ürünleri belirlemek üzere sol ve sağ bileşenler olmak üzere iki ayrı sütuna ayrılmıştır.

```
hm_rxns <- readxl::read_excel("rdskeyit//HumanModel.xlsx", 1) %>%
  select_if(~ any(!is.na(.)))

hm_rxn_tbl <- hm_rxns %>%
  dplyr::select(RXNID, EQUATION) %>%
  separate(EQUATION, c("Left", "Right"), sep="=>") %>%
  mutate(reversible = str_detect(Left, "<")) %>%
  mutate(Left = str_replace(Left, "<", ""))

hm_rxn_tbl2 <- hm_rxn_tbl %>%
  filter(reversible == TRUE) %>%
  select(RXNID, Right2=Left, Left2=Right) %>%
  select(RXNID, Left=Left2, Right=Right2) %>%
  bind_rows(hm_rxn_tbl)

hm_rxn_tbl2 %>%
  select(-reversible) %>%
  mutate(Left = strsplit(Left, "\\+ ")) %>%
  mutate(Right = strsplit(Right, "\\+ ")) %>%
  select(RXNID, Left) %>%
  unnest(Left) %>%
  select(from=Left, to= RXNID) %>%
  mutate(from = str_trim(from, side=c("both")), from = str_replace(from, regex("^([0-9]+|^[0-9].[0-9]+|^[0-9].[0-9]e-[0-9]+)"), "")) -> substrates

hm_rxn_tbl2 %>%
  select(-reversible) %>%
  mutate(Left = strsplit(Left, "\\+ ")) %>%
  mutate(Right = strsplit(Right, "\\+ ")) %>%
  select(RXNID, Right) %>%
  unnest(Right) %>%
  select(from=RXNID, to=Right) %>%
  mutate(to = str_trim(to, side=c("both")), to = str_replace(to, regex("^([0-9]+|^[0-9].[0-9]+|^[0-9].[0-9]e-[0-9]+)"), "")) -> products
```

Şekil 5.1 İnsan metabolit ağı verilerinin işlenmesi: reaksiyonların ayrıştırılması ile yönünün belirlenmesi ve substrat ile ürünlerin ayrı sütunlara dağıtılması

Tablo 5.1'de sol tarafta listelenen bileşenler, substratlar olarak tanımlanmıştır.

Bu tabloda, 'from' sütunu substratları, 'to' sütunu ise bu substratların katıldığı reaksiyonları (Human Metabolic Reaction database'deki kodlarla) göstermektedir. Örneğin, L-laktat[p]'nin HMR_4281 kodlu reaksiyona katıldığı belirtilmiştir. Bu bilgiler, belirli metabolik yolların incelenmesinde kullanılmıştır. Tablo 5.2'te ise sağ tarafta listelenenler ürünler olarak tanımlanmıştır. Bu tablo, 'from' sütununda belirtilen reaksiyon kodlarının ('to' sütununda) ürettiği son ürünleri göstermektedir. Örneğin, HMR_4281 kodlu reaksiyonun H⁺, NADH ve piruvat ürettiği belirtilmiştir.

Tablo 5.1 Örnek substrat ve ilişkili reaksiyonlar veri seti

from	to
L-lactate[p]	HMR_4281
NAD+[p]	HMR_4281
L-lactate[c]	HMR_4388
NAD+[c]	HMR_4388
H ₂ O[c]	HMR_4363
PEP[c]	HMR_4363

Tablo 5.2 Örnek ürün ve ilişkili reaksiyonlar veri seti

from	to
HMR_4281	H+[p]
HMR_4281	NADH[p]
HMR_4281	pyruvate[p]
HMR_4388	H+[c]
HMR_4388	NADH[c]
HMR_4388	pyruvate[c]

İki ayrı veri seti olarak oluşturulan substratlar ve ürünler veri setleri birleştirilerek, substratlardan reaksiyonlara ve reaksiyonlardan ürünlere uzanan uzun bir liste oluşturulmuştur. Bu liste, bileşik ağının (compound network) oluşturulmasında kullanılmıştır Şekil 5.2. "Trim" komutu ile listedeki boşluklar temizlenmiş, böylece verilerin daha düzenli ve işlenebilir hale getirilmesi sağlanmıştır. Bu işlem, veri setinin analiz edilmesi ve ilişkili bileşenler arasındaki bağlantıların anlaşılması için kritik öneme sahiptir ve metabolik reaksiyonların ayrıntılı bir şekilde incelenmesine olanak tanımıştır.

```

all_rxns <- bind_rows(substrates,products)

compound_network <-
  inner_join(substrates,products,by=c("to"="from")) %>%
  rename(RXNID=to) %>%
  dplyr::select(from,to=to.y,RXNID) %>%
  filter(from != to) %>%
  mutate(from=trimws(from),
         to=trimws(to)) %>%
  as_tbl_graph()%>%
  mutate(node_id=row_number())

```

Şekil 5.2 Substrat ve ürün veri setlerinin entegrasyonu ile metabolik bileşik ağının oluşturulması

5.3 Hastalar Bazında HMR Değerlendirmesi

5.3.1 Gen-Protein-Reaksiyon İlişkilerinin İşlenmesi

"gene_rule_data" veri seti, HMA'dan elde edilen gen-protein-reaksiyon ilişkilerini içermekte olup, gen ifade kurallarını ve reaksiyonların birleşimini temsil etmektedir. Gen ifade kurallarını işlemek ve analiz etmek için özel bir R fonksiyonu (gr_modifier) geliştirilmiştir Şekil 5.3. Bu fonksiyon, Boolean mantığıyla tanımlanmış gen ifade kurallarını (örneğin, (gene1 & gene2) | gene3) işleyerek, bu kuralları daha işlevsel istatistiksel ve matematiksel ifadelere dönüştürür.

Bu dönüşümde, 've' ('&') operatörleri "min" fonksiyonuyla, 'veya' ('|') operatörleri ise "sum" fonksiyonuyla değiştirilmiştir. Örneğin, (gene1 & gene2) | gene3 ifadesi sum(gene3, min(gene1, gene2)) şekline dönüştürülmüştür. Bu dönüşüm, gen ifade kurallarını kesin ve sayısal olarak işlenebilir bir formatta temsil etmeyi sağlamıştır, böylece bu kurallar ilgili metabolik ağ analizlerinde kullanılan algoritmalarla uyumlu hale gelmiştir.

"gr_modifier" fonksiyonu, verilen gen ifade kuralları üzerinde yinelemeli olarak çalışmakta ve bu kuralların her birini uygun matematiksel ifadelere çevirmektedir. Bu kısımda R'de işlevsel programlama tekniklerini kolaylaştıran ve liste işlemleri için güçlü araçlar sunan purr paketi ile purrr'ın işlevlerini gelecekteki değerler üzerinde paralel olarak çalışacak şekilde genişleten ve böylece veri analizi ve manipülasyon işlemlerini hızlandıran furr paketi kullanılmıştır[143, 144].

```

library(purrr)
library(furrr)

gr_modifier <- function (x) {
  if (is.atomic(x) || is.name(x)) {
    x
  } else if (is.call(x)) {
    if (identical(x[[1]], quote(`|`))) {

      x[[1]] <- quote(sum)
      x
    }
    if (identical(x[[1]], quote(`&`))) {

      x[[1]] <- quote(min)
      x
    }
    if (identical(x[[1]], quote(`^`))) {
      gr_modifier(x[[2]])
    }
    else {

      as.call(lapply(x, gr_modifier))
    }
  } else if (is.pairlist(x)) {
    as.pairlist(lapply(x, gr_modifier))
  } else {

    stop("Don't know how to handle type ", typeof(x),
        call. = FALSE)
  }
}

```

Şekil 5.3 Reaksiyon ifade kurallarının semantik analizi ve sayısal dönüşüm için geliştirilen "gr_modifier" fonksiyonu

5.3.2 Geliştirilen R Fonksiyonları ile GPR Kurallarının Analizi

Şekil 5.4'de gösterilen komutlar , "gene_rule_data" adlı veri setinden alınan GPR kurallarını işlemek ve değerlendirmek için tasarlanmıştır. Fonksiyon, her bir "gene_rule" ifadesini, genetik regülasyonun sayısal bir değerlendirmesine dönüştürmüş ve bu değerlendirmeleri kişiye özgü olarak çıkarmıştır. İlk adımda, "gene_rule" sütunundaki her bir ifade, mantıksal operatörleri temsil eden 'and' ve 'or' kelimelerinin yerine '&' ve '|' simgeleri ile değiştirilmiştir. Bu dönüşüm, GPR kurallarının daha karmaşık matematiksel işlemler için hazırlanmasını sağlamıştır.

Daha sonra, "gene_rule" ifadeleri, her bir reaksiyon için belirli bir sayısal değer atayabilmek amacıyla işlenmiştir. Bu işlem, "fixer" adlı yardımcı fonksiyon ve daha sonra "gr_modifier" fonksiyonu kullanılarak gerçekleştirilmiştir. "fixer" fonksiyonu, ifadeleri parçalayarak ve her bir gen adını "genesv2" vektöründeki ilgili sayısal değerlerle eşleştirerek, bu ifadelerin daha sonra işlenebilir hale getirilmesini

sağlamıştır. "gr_modifier" fonksiyonu ise, bu işlenmiş ifadeleri matematiksel bir değerlendirmeye dönüştürmüştür. Bu dönüştürme işlemi, GPR kurallarının mantıksal yapısını sayısal bir formata çevirerek, her bir kuralın genetik etkinliğinin değerlendirilmesini mümkün kılmaktadır.

Son aşamada, her bir "gene_rule" için hesaplanan sayısal değerler, fonksiyonun girdisi olarak alınan "col_name" parametresiyle isimlendirilmiş ve sonuçlar kişi bazında çıkarılmıştır. Sonuçlar, belirtilen dosya yoluna ".tsv" formatında kaydedilmiştir. Bu metodoloji, GPR kurallarının karmaşık yapısını anlaşılabilir ve işlenebilir bir formata dönüştürerek, reaksiyonların metabolik aktiviteleri üzerindeki etkilerini detaylı bir şekilde incelemeyi amaçlamaktadır.

```
extract_results_per_person <- function(named_vector, col_name){

  genesv2 <- named_vector
  col_name <- col_name

  fixer <- function(variable){
    variable <- str_replace_all(variable, c("\\(" = "\\( ", "\\)" = "\\) "))
    var1<- str_split(variable, " ", simplify = T)
    var2<- map_if(var1, str_detect(var1, "[^&|()]" ), ~
paste0("genesv2['",.,"'"])))
    var3<- paste(var2, collapse = " ")
    return(var3)
  }

  gene_rule_data %>%
  select(api_id,gene_rule) %>%
  mutate(ens_count=str_count(gene_rule,"ENSG")) %>%
  filter(ens_count < 250) %>%
  filter(gene_rule != "NULL") %>%
  mutate(gene_rule2 = str_replace_all(gene_rule, "and", "&"),
         gene_rule2 = str_replace_all(gene_rule2, "or", "|")) %>%
  mutate(gene_rule2 = map(gene_rule2, ~fixer(.))) %>%
  tidyr::unnest(gene_rule2) %>%
  mutate(gene_rule3= map(gene_rule2, ~gr_modifier(rlang::parse_expr(.)))) %>%
  mutate(result = map_dbl(gene_rule3, ~eval(.))) %>%
  select(api_id, gene_rule, !!col_name := result) %>%
  write_tsv(paste0("hmr_results/",col_name,".tsv"))
}
```

Şekil 5.4 Gen ifade kurallarının sayısal değerlendirmeye dönüştürülmesi ve kişiselleştirilmiş genetik aktivite sonuçlarının çıkarılması

5.3.3 Metodolojinin Uygulanması ve Birey Bazında HMR Hesaplamaları

Kişi bazında sonuçlar alındıktan sonra meme kanseri ile ilişkili RNA ifade verilerinin işlenmesi ve HMA'dan elde edilen reaksiyonlar(HMR) ile entegrasyonu gerçekleştirilmiştir. İlk olarak, 1222 bireyden (1109 hasta ve 113 sağlıklı) oluşan "brca_counts" veri seti, RNA ifade düzeylerini içeren bir matris (mat) olarak hazırlanmıştır Şekil 5.5. Ardından, mat matrisi, her bir bireye özgü alt veri setlerine

ayrılarak işlenmiştir. Bu işlem, her bir sütun için "extract_results_per_person" fonksiyonunu paralel olarak uygulamak üzere bölünmüştür. "future_walk2" kullanımı, bu işlemin paralel bir şekilde yapılmasını sağlamıştır, böylece çok sayıda hasta için işlem hızlıca gerçekleştirilmiş olup, her birey için özelleştirilmiş HMR hesaplamalarını içermiştir.

```
Mat <- readRDS("brca_counts.rds") %>%
  head() %>%
  knitr::kable()

mat %>%
  split(rep(colnames(mat), each = nrow(mat))) %>%
  map(set_names, rownames(mat)) %>%
  future_walk2(., names(.), ~extract_results_per_person(.x,.y))

get_hmr_from_file <- function(filename){
  read_tsv(filename, col_types = "ccd") %>%
  select(3)
}

test <- fs::dir_ls(path="hmr_results_initial/", glob = "*.tsv") %>%
  map_dfc(get_hmr_from_file)

single <- read_tsv("hmr_results_initial/8525c65e-bf81-46b3-b74e-
ae8260336dc4.tsv") %>%
  #select(1,2)/

final <- bind_cols(single, test)

saveRDS(final, "hmr_results_final.rds")

hmr_results_final <- readRDS("rdskayit/hmr_results_final.rds")
```

Şekil 5.5 Meme kanseri RNA ifade verilerinin kişiselleştirilmiş HMR hesaplamaları ile entegrasyonu

Hesaplanan HMR sonuçları, daha sonra "hmr_results" dizinine TSV formatında kaydedilmiştir. Sonraki aşamada, "get_hmr_from_file" fonksiyonu ile bu TSV dosyalarından elde edilen HMR sonuçları okunmuştur. Bu sonuçlar, test nesnesi aracılığıyla tek bir veri setinde birleştirilmiş ve spesifik bir TSV dosyasından (single) alınan bireysel verilerle entegre edilmiştir Şekil 5.5. Sonuç olarak elde edilen final veri seti, "hmr_results_final.rds" dosyasına kaydedilmiştir Tablo 6.2. Bu metodoloji ile, genetik verilerin metabolik reaksiyonlarla olan ilişkilerinin detaylı analizi ve anlaşılması hedeflenmektedir. Her birey için yapılan HMR hesaplamaları, genler ve metabolik verilerin kapsamlı bir şekilde birleştirilmesiyle sonuçlanmıştır.

5.4 LIMMA Analizi

Diferansiyel ifade analizi için iki temel Bioconductor paketi kullanılmıştır: LIMMA ve edgeR. Bu paketler, reaksiyon ifade verilerinin analizi ve diferansiyel ifade

analizleri için geliştirilmiştir. LIMMA, mikrodizi ve RNA-seq verileri için istatistiksel analizlerde kullanılan genel bir pakettir. Özellikle, reaksiyon ifadesi verilerinden anlamlı biyolojik bilgiler elde etmek için tasarlanmıştır. edgeR (Empirical Analysis of Digital Gene Expression Data in R), sayısal reaksiyon ifadesi verilerinin analizi için kullanılmakta ve özellikle RNA-seq verileri için geliştirilmiş diferansiyel ifade analizi yöntemleri sunmaktadır.

Bu paketlerin kurulumu, R paket yöneticisi BiocManager üzerinden gerçekleştirilmiştir. BiocManager, Bioconductor projesinin paketlerini ve bağımlılıklarını yönetmek için kullanılan bir araçtır [145, 146]. Paketlerin kurulumu şu adımları içermektedir:

1. BiocManager paketinin kurulu olup olmadığı kontrol edilmektedir. Eğer sisteme yüklü değilse, `install.packages("BiocManager")` komutu ile yüklenmektedir.
2. `BiocManager::install("limma")` ve `BiocManager::install("edgeR")` komutları ile sırasıyla LIMMA ve edgeR paketleri kurulmuştur. Ayrıca analiz sonucu oluşturulan grafikler için ggplot paketi de çalıştırılmıştır [136, 147, 148].

Çalışmamızda LIMMA paketi ile yapılan diferansiyel reaksiyon ifadesi analizi, reaksiyon ifadesindeki değişikliklerin hastalıkların anlaşılmasında ve tedavi yöntemlerinin geliştirilmesinde önemli rol oynamaktadır.

Çalışmamızda LIMMA paketi kullanılarak gerçekleştirilen diferansiyel reaksiyon ifade analizi, hastalıkların daha iyi anlaşılmasına ve tedavi yöntemlerinin geliştirilmesine katkıda bulunan, ekspresyondaki değişimleri belirlemede kritik bir öneme sahiptir.

5.4.1 Veri Gruplandırma ve Tasarım Matrisi

- "hmr_results_final" veri setinden ilk iki sütun çıkarılarak, hasta ID'leri içeren bir matris oluşturulmuştur. Bu işlem, LIMMA analizinde kullanılacak matrisin hazırlanması için gerekmektedir.
- Matrisin satır isimleri, hmr_results_final veri setindeki api_id değerleri ile etiketlenmiştir. Bu, analiz sırasında verilerin doğru şekilde eşleştirilmesini sağlamaktadır.
- hmr_matrix matrisindeki sütunlar, hasta gruplarına ait bilgileri içermektedir. Bu gruplar, tcga_brca veri seti ile birleştirilerek, her bir örneğin tümör durumuna göre etiketlenmiştir.

- LIMMA analizi için tasarım matrisi (design) oluşturulmuştur. Bu matris, grupların istatistiksel analizlerde kullanılması için gerekli yapıyı sağlamaktadır.

5.4.2 Voom Dönüşümü ve Lineer Model Uygulama

- Voom fonksiyonu, RNA-seq verileri için özel olarak geliştirilmiş ve LIMMA paketi tarafından sağlanan bir dönüşümdür. Bu dönüşüm, veri setindeki düşük sayımlı verilerin normalleştirilmesini sağlamıştır.
- lmFit fonksiyonu ile lineer model fit edilmiştir. Bu model, gruplar arasındaki gen ifadesi farklılıklarını belirlemek için kullanılmaktadır.

5.4.3 Diferansiyel İfade Analizi Sonuçları

lmFit fonksiyonu ile uygulanan lineer model, gruplar arasındaki reaksiyon ifadesi farklılıklarını belirlemek için kullanılmıştır. eBayes fonksiyonu ile gruplar arasındaki farkları değerlendiren istatistiksel bir model oluşturulmuş ve topTable fonksiyonu, modelden elde edilen en anlamlı reaksiyonları belirlemiştir Şekil 5.6. Bu sayede bu reaksiyonların ve ilgili istatistiksel bilgilerin kolayca incelenmesi sağlanmıştır.

```
fit <- lmFit(v, design)

colnames(design) <- c("Group1", "Group2")

fit2 <- eBayes(contrasts.fit(fit, contrasts = makeContrasts(Difference = "Group2 - Group1",
levels = design)))

res <- topTable(fit2, coef="Difference", n=Inf)
res <- res %>%
  as_tibble(rownames = "RXNID")

head(res) %>%
  knitr::kable()
```

Şekil 5.6 Lineer model ve eBayes fonksiyonları kullanılarak gruplar arası reaksiyon ifadesi farklılıklarının istatistiksel değerlendirilmesi

Reaksiyon ifade değişimlerinin istatistiksel anlamlılığını görselleştirmek için bir volkan grafiği oluşturulmuştur. "ggplot2" paketi kullanılarak hazırlanan bu grafik, her bir genin log2 katlanma değişimi (logFC) ile bu değişimin negatif log10 P-değerinin (-log10 P-Value) bir arada sunulmasını sağlamıştır. Bu görselleştirme, reaksiyon ifade seviyelerindeki önemli değişiklikleri ve bu değişikliklerin istatistiksel anlamlılığını vurgulamak için kullanılmıştır.

5.4.4 Reaksiyon ve Metabolik Bileşiklerin Entegrasyonu

LIMMA analizi sonucunda elde edilen diferansiyel ifade gösteren reaksiyonlar ve istatistik bilgilerini içeren "res" veri seti ile, daha önce oluşturulan ve sadece reaksiyonlar ile bileşikler içeren "compound_network" veri seti entegre edilmiştir. Bu entegrasyon, kanser durumunda hangi metabolik yolların etkilendiğini ve hangi metabolitlerin arttığını veya azaldığını ortaya koymuştur Şekil 5.7’te bu komutlar ve oluşan bu yeni veri setinin network yapısı verilmiştir. Ayrıca "as_tibble" komutu ile tablo formatına dönüştürülerek Tablo 6.4’da sunulmuştur.

```
limmacompoundnetwork <- readRDS("rds kayit/compound_network") %>%  
  activate(edges) %>%  
  inner_join(res, by="RXNID")  
  
limmacompoundnetwork  
  
|  
  
A tbl_graph: 6008 nodes and 60085 edges  
  
A directed multigraph with 1051 components  
  
A tibble: 60,085 × 9  
  from   to RXNID logFC AveExpr t P.Value adj.P.Val B  
  <int> <int> <chr>    <dbl> <dbl> <dbl>    <dbl>    <dbl> <dbl>  
1     1     517 HMR_4281 0.350  9.62  5.46 0.0000000581 0.0000000977 6.51  
2     1     537 HMR_4281 0.350  9.62  5.46 0.0000000581 0.0000000977 6.51  
3     1    1757 HMR_4281 0.350  9.62  5.46 0.0000000581 0.0000000977 6.51  
4     1       2 HMR_4281 0.350  9.62  5.46 0.0000000581 0.0000000977 6.51  
5     2     517 HMR_4281 0.350  9.62  5.46 0.0000000581 0.0000000977 6.51  
6     2     537 HMR_4281 0.350  9.62  5.46 0.0000000581 0.0000000977 6.51  
#> # A tibble: 60,079 more rows  
  
A tibble: 6,008 × 2  
  name      node_id  
  <chr>      <int>  
1 L-lactate[p]      1  
2 NAD+[p]           2  
3 L-lactate[c]      3  
#> # A tibble: 6,005 more rows
```

Şekil 5.7 LIMMA analizi sonuçları ile metabolit ağın entegrasyonunun Tidygraph Yapısı

5.5 Metabolik Reaksiyon Analizi

5.5.1 Artan ve Azalan Bileşiklerin Belirlenmesi

Düzeltilmiş p-değeri (adj.P.Val) 0.01'den küçük olan reaksiyonlar filtrelenerek, hasta ve kontrol grubu arasında ifadesi anlamlı derecede değişen reaksiyonlar tespit edilmiştir. Bu işlem ile çalışmamızın odak noktası olan değişen reaksiyonları daraltarak yeni bir alt küme oluşturulmuştur. adj.pvalue 0,01 olarak ayarlanınca hasta ve kontrol arasında değişen 4973 eşsiz reaksiyon (api_id) bulunmuştur.

Alt küme oluşturulduktan sonra artan ve azalan ifade düzeylerine göre reaksiyonlar iki ayrı kümeye ayrılmıştır. Bu iki küme, artan ve azalan metabolik bileşiklerin değerlendirilmesine olanak tanımıştır. LogFC değeri pozitif olan bileşikler "increased_compounds" olarak, LogFC değeri negatif olan bileşikler "decreased_compounds" olarak ayrılmıştır. Her iki durum için ilgili sütunlar (from, to, RXNID, logFC, adj.P.Val) seçilmiştir. Her RXNID için artan ve azalan bileşiklerin sayısı "HMR_summary" veri setinde özetlenmiştir Şekil 5.8.

```
increased_compounds <- subset_data %>%
  filter(logFC > 0) %>%
  select(from, to, RXNID, logFC, adj.P.Val)

decreased_compounds <- subset_data %>%
  filter(logFC < 0) %>%
  select(from, to, RXNID, logFC, adj.P.Val)

HMR_summary <- subset_data %>%
  group_by(RXNID) %>%
  summarise(
    num_increased_compounds = sum(logFC > 0),
    num_decreased_compounds = sum(logFC < 0)
  )
```

Şekil 5.8 Artan ve azalan metabolit veri setlerinin oluşturulması ve reaksiyon ifade düzeylerine göre özetlenmesi

5.5.2 Kenar ve Düğüm Verilerinin İşlenmesi

limma_compound_network" içindeki "from" ve "to" sütunlarındaki indeksler kullanılarak bu indekslere karşılık gelen metabolitlerin isimleri yeni sütunlar olarak kenar bilgilerine eklenmiştir Şekil 5.9.

```
nodes_info <- activate(limma_compound_network, "nodes")
edges_info <- activate(limma_compound_network, "edges")

edges_info <- edges_info %>%
  mutate(HMR_from = nodes_info$name[from],
         HMR_to = nodes_info$name[to])

limma_compound_network <- activate(limma_compound_network, "edges") %>%
  mutate(HMR_from = edges_info$HMR_from,
         HMR_to = edges_info$HMR_to)
```

Şekil 5.9 Kenar bilgilerine "from" ve "to" indekslerine karşılık gelen metabolit isimlerinin eklenmesi

Kenar ve düğüm verileri veri seti olarak çıkarılmış ve metabolit isimleri edge_df veri setine eklenmiştir. Bu işlemle, kenar verilerinde "from" ve "to" düğüm id'lerinin karşılık gelen isimleri "edge_df" veri setinde artan ve azalan olacak şekilde "change" adıyla yeni bir sütunda oluşturulmuştur. Şekil 5.10.

```

edge_df <- limma_compound_network %>% activate(edges) %>% as_tibble()
node_df <- limma_compound_network %>% activate(nodes) %>% as_tibble()

node_names <- node_df %>% select(name, node_id)

edge_df <- edge_df %>%
  left_join(node_names, by = c("from" = "node_id")) %>%
  rename(from_name = name) %>%
  left_join(node_names, by = c("to" = "node_id")) %>%
  rename(to_name = name)

edge_df <- edge_df %>%
  mutate(change = ifelse(logFC > 0, "increase", "decrease"))

```

Şekil 5.10 Kenar verilerine "from" ve "to" düğüm id'lerinin karşılık gelen isimlerinin eklenmesi ve ifade düzeyindeki değişimlerin belirlenmesi

5.5.3 HMR Değişiklik Özeti

Her HMR için "from_name" ve "to_name" sütunlarındaki özgün metabolit sayıları hesaplanmış ve sonuçlar "HMR_change_summary" veri setine kaydedilmiştir. Bu veri seti daha sonra liste halinde bulunan "from_compounds" ve "to_compounds" sütunları açılarak "HMR_change_summary_unnested" veri seti olarak genişletilmiştir Şekil 5.11. Tablo 6.5'de, artan ve azalan reaksiyonlar ve bu reaksiyonlarla ilişkili metabolitler örnek olarak sunulmuştur.

```

compounds_df <- limma_compound_network %>%
  activate(nodes) %>%
  as_tibble() %>%
  select(name, node_id)

HMR_change_summary <- edge_df %>%
  group_by(RXNID, change) %>%
  summarise(
    num_from_compounds = n_distinct(from_name),
    num_to_compounds = n_distinct(to_name),
    from_compounds = list(unique(from_name)),
    to_compounds = list(unique(to_name)),
    .groups = "drop" # This line suppresses a warning
  )

HMR_change_summary_unnested <- HMR_change_summary %>%
  unnest(cols = c(from_compounds)) %>%
  unnest(cols = c(to_compounds))

```

Şekil 5.11 Her HMR için metabolit sayılarının hesaplanması ve "from_name" ile "to_name" metabolitlerinin detaylandırılması

5.6 Metabolik Veri Setinin Filtrelenmesi ve İşlenmesi

5.6.1 Veri Setinin Hazırlanması ve Filtreleme İşlemleri

Metabolik verilerin spesifik özelliklerine odaklanarak belirli reaksiyonları izole etmek amacıyla , HMA veri tabanından elde edilen human_model dosyasının SUBSYSTEM sütunundaki "Pool reactions", "Transport reactions" ve "Isolated" etiketli reaksiyonlar elenmiştir. Bu aşamada R paketlerinden "readxl" paketi kullanılmıştır [149].

5.6.2 Gen-Protein-Reaksiyon Kurallarının İncelenmesi

Metabolizma alt tipleri belirlendikten sonra yine human_model dosyasındaki gen ifade kurallarının bulunduğu "GENE ASSOCIATION" sütununun ismi daha anlaşılabilir olması amacıyla "gene_rule" olarak değiştirilmiş ve bu sütunda "or" sayımı yapılarak, reaksiyon ifadeleri arasındaki kombinasyonel ilişkiler incelenmiştir. Bu sayım, reaksiyonların genetik regülasyon kompleksitesini gösterir ve yüksek "or" sayıları, bir reaksiyonun birden fazla gen tarafından kontrol edildiğini belirtir. Burada bir eşik değer olarak 17 belirlenmiş ve bu değerin altındaki reaksiyonlar, daha az genetik regülasyona sahip oldukları için analiz edilmiştir. 17 değerinin üzerindeki reaksiyonlar gen ifadesinin karmaşıklığını ve genetik kontrolün derecesini değerlendirmek

Çalışmamızda, reaksiyonların genetik regülasyon kompleksitesini incelemek amacıyla human_model dosyasındaki gen ifade kurallarının bulunduğu "GENE ASSOCIATION" sütununun adı daha anlaşılır olması için "gene_rule" olarak değiştirilmiştir. Bu sütundaki "or" sayıları, bir reaksiyonun kaç farklı gen tarafından kontrol edildiğini gösterir. Analizimizde, "or" sayısı 17 ve altında olan reaksiyonlar dahil edilmiş, 17'den yüksek "or" sayılarına sahip reaksiyonlar ise rekürsif (özyineleme) sorunları önlemek amacıyla analiz dışı bırakılmıştır. Bu sayede, analizimiz daha yönetilebilir hale getirilmiş ve veri setindeki gürültünün azaltılması sağlanmıştır. Bu yaklaşım, genetik regülasyonun daha az karmaşık olduğu reaksiyonların detaylı incelenmesine olanak tanımıştır.

5.6.3 Veri Setinin Daraltılması ve Metabolik Bileşiklerin Filtrelenmesi

Bir sonraki aşamada, bir önceki aşamada oluşturduğumuz "HMR_change_summary_unnested" veri seti, yukarıda belirtilen filtreye tabi tutulmuş olan "human_gem_filter_or_filtered" veri seti ile "anti_join" komutu kullanılarak birleştirilmiştir. Veri seti içinde, "from" ve "to" bileşikleri için belirlenen sayısal eşikler (5'ten küçük olanlar) ile daha fazla filtreleme yapılmış

ve böylece, daha az sayıda bileşik içeren metabolik reaksiyonlar ayrılmıştır. Bu seçim, metabolik ağ analizindeki odak noktasını, en aktif ve potansiyel olarak patolojik süreçlerle en yakından ilişkili olan metabolik reaksiyonlara yönlendirmek için yapılmıştır.

Filtreleme sonucunda elde edilen veri seti, artık daha yönetilebilir bir boyutta olup, metabolik ağdaki en önemli değişimleri temsil eden reaksiyon ve bileşikleri içermektedir. Bu yaklaşım, metabolik yolların ve reaksiyonların daha derinlemesine incelenmesi için sağlam bir temel oluştururken, aynı zamanda büyük veri setlerinin analizinde sıkça karşılaşılan karmaşıklığı ve gürültüyü de azaltmıştır.

Bu filtreleme süreci, metabolik ağ analizlerinde sinyal-gürültü oranını iyileştirmek, analizin odaklanacağı anahtar metabolik reaksiyonları belirlemek ve genel olarak metabolik ağın daha net bir haritasını elde etmek için kritik öneme sahiptir. Yapılan bu titiz veri işleme ve filtreleme işlemleri, metabolik ağın daha spesifik ve hastalığa özgü kısımlarına odaklanılmasını sağlamış ve hastalık mekanizmalarının daha iyi anlaşılmasına yardımcı olmuştur.

5.7 Tümör ve Sağlıklı Örneklerin Reaksiyon İfadesi Analizi

5.7.1 Veri Seti ve Ön İşleme

Bu bölümde analizimiz için "tcga_brca" veri setinden tümör örnekleri (tumor == TRUE) ve sağlıklı örnekler (tumor == FALSE) ayrıştırılarak ayrı vektörlerde saklanmıştır. Sağlıklı örnekler arasında, "Blood Derived Normal" (Kan kaynaklı) ve "Solid Tissue Normal" (Katı Doku kaynaklı) gibi örnek türleri bulunmaktadır. Bu örnekler, tümör dokusuna bitişik normal dokuları ve tümörden bağımsız sağlıklı dokuları temsil etmektedir. Reaksiyonlar ise "hmr_results_final" veri setinden elde edilmiştir. İlgili veri sütunları 'file_id' ve 'api_id' olarak tanımlanmıştır. Tümör ve sağlıklı örneklerin kesişim kümesi alınarak, analiz için kullanılacak örneklerin belirlenmesi sağlanmıştır Şekil 5.12.

```
hmr_results_final <- readRDS("rds kayit/hmr_results_final.rds")
tcga_brca <- readRDS("rds kayit/tcga_brca")

tumor_samples <- tcga_brca[tcga_brca$tumor == TRUE,]$file_id
api_ids <- hmr_results_final$api_id

healthy_samples <- tcga_brca[tcga_brca$tumor == FALSE,]$file_id
healthy_samples <- intersect(healthy_samples, colnames(hmr_results_final))
```

Şekil 5.12 Tümör ve sağlıklı örneklerin analiz için ayrıştırılması

5.7.2 Tümör ve Sağlıklı Örneklerdeki Reaksiyon İfade Seviyelerinin Karşılaştırmalı Analizi

Sağlıklı örnekler üzerinden reaksiyon ifade seviyelerinin aritmetik ortalaması hesaplanmış ve tümör örneklerinin reaksiyon ifade seviyeleriyle karşılaştırılarak her bir reaksiyon için tümör örneklerindeki sapma değerleri hesaplanmıştır. Her bir tümör örneği için hesaplanan reaksiyon ifade seviyelerindeki sapma değerleri, belirlenen kriterlere göre "artış" veya "azalış" olarak sınıflandırılmıştır. Bu sınıflandırma işlemi, reaksiyonlar için artış ve azalış frekanslarının hesaplanmasını ve her bir reaksiyonun genel örnekler içerisindeki yüzdesel dağılımının belirlenmesini mümkün kılmıştır. Hesaplanan bu frekanslar ve reaksiyonların kimlik bilgileri, "api_id" ile "increase" ve "decrease" etiketlerinin yüzdesel değerlerini içerecek şekilde "result_df" adı verilen veri setine kaydedilmiştir Şekil 5.13.

```
increase_decrease <- ifelse(deviation_from_mean_tumor > 0, "increase",
                             "decrease")

increase_counts <- rowSums(increase_decrease == "increase", na.rm = TRUE)
decrease_counts <- rowSums(increase_decrease == "decrease", na.rm = TRUE)

total_counts <- increase_counts + decrease_counts
increasecounts_2 <- (increase_counts / total_counts)
decrease_counts_2 <- (decrease_counts / total_counts)

result_df <- data.frame(api_id=api_ids, increase=increasecounts_2,
                        decrease=decrease_counts_2)
```

Şekil 5.13 Reaksiyon ifade seviyelerindeki sapma değerlerine dayalı artış ve azalış oranlarının hesaplanması

5.7.3 Eşik Değeri Belirleme ve Filtreleme

Filtreleme sürecinde, artış ve azalış oranları yüzde 90'ın üzerinde değişim gösteren reaksiyon ve metabolitler filtrelenmiş ve analiz edilmiştir. Bu seviyenin altında kalan değerler, elde edilen sonuçların sayısını aşırı yükselterek istatistiksel gürültüye neden olmuş ve reaksiyonların biyolojik duyarlılığını azaltabileceği gözlemlenmiştir. Ayrıca, yüzde 95 ve üzeri eşik değerleri ile yaptığımız filtreleme denemelerinde, metabolik bileşik sayısının önemli ölçüde azaldığı tespit edilmiştir. Bu nedenle, hem istatistiksel hem de biyolojik olarak anlamlı sonuçları elde etmek için yüzde 90 olarak belirlenen eşik değer, analizlerimizde kullanılmıştır. Yanlış pozitif sonuçları minimize etmek için ek olarak yüzde 10'luk bir filtre de uygulanmış ve bu sayede "count_increase_90" ve "count_decrease_90" olarak tanımlanan artış ve azalış gösteren reaksiyon sayıları belirlenmiştir Şekil 5.14. Sadece 83 reaksiyon artarken, 1158 reaksiyonun azaldığı gözlemlenmiş ve dengeli bir veri seti oluşturmak amacıyla, artan ve azalan gruplardan eşit miktarlarda seçim

yaparak, en büyük değışiklięi gösteren ilk 83 reaksiyon temel alınmıřtır. Bu seęim kriterleri, sonraki analizlerde kullanılmak üzere 'result_df_166' adlı yeni bir veri setinde birleřtirilmiřtir.

```
result_df_90 <- result_df %>%  
  filter((increase > 0.90 & decrease < 0.10) | (decrease > 0.90 & increase <  
0.10))  
count_increase_90 <- nrow(filter(result_df_90, increase > 0.90))  
  
count_decrease_90 <- nrow(filter(result_df_90, decrease > 0.90))  
  
count_increase_90  
## [1] 83  
  
count_decrease_90  
## [1] 1158  
  
top_increase_df <- result_df_90 %>% filter(increase > 0.90)%>% head(83)  
top_decrease_df <- result_df_90 %>% filter(decrease > 0.90)%>% head(83)  
  
result_df_166 <- bind_rows(top_increase_df, top_decrease_df)
```

řekil 5.14 Yüksek değışim dösteren reaksiyonlarda en belirgin artıř ve azalıř gösteren metabolitler ile dengeli veri seti oluřturulması

5.7.4 Kontrol ve Saęlama İřlemleri

İlk adımda, elde edilen "result_df_166" veri seti kullanılarak, artıř ve azalıř gösteren reaksiyonlar için belirli "api_id" değeri seęilmiřtir. Bu seęim, analizin ileri ařamalarında kullanılmak üzere, özellikle ilgi çekici olan reaksiyonları belirlemek amacıyla yapılmıřtır. Seęilen "api_id" değeri arasından 5'er tane artıř ve azalıř gösteren reaksiyonlar belirlenmiř, ve bu reaksiyonlar yeni bir veri seti olan "new_df" içinde birleřtirilmiřtir řekil 5.15. Bu iřlem, analizin doęruluęunu kontrol etmek için bir temel oluřturmuřtur.

```
top_increase_ids <- unique(result_df_166$api_id[result_df_166$increase > 0.90])  
top_decrease_ids <- unique(result_df_166$api_id[result_df_166$decrease > 0.90])  
  
random_increase_ids <- sample(top_increase_ids, 5)  
random_decrease_ids <- sample(top_decrease_ids, 5)  
  
new_df_increase <- result_df_166 %>% filter(api_id %in% random_increase_ids)  
new_df_decrease <- result_df_166 %>% filter(api_id %in% random_decrease_ids)  
  
new_df <- bind_rows(new_df_increase, new_df_decrease)
```

řekil 5.15 Analitik değeriendirme için öne çıkan reaksiyonların seęilmesi

Tümör örneklerindeki RNA ifade düzeyleriyle ilgili reaksiyonları detaylandırmak

amacıyla, "new_df" veri seti, "hmr_results_final_tumor_data_api_long" ile "semi_join" komutu kullanılarak birleştirilmiştir. Bu birleştirme sonucunda, gen kurallarına dayalı RNA ifade düzeylerinin yer aldığı "Sağlama_data" adında yeni bir veri seti oluşturulmuştur Şekil5.16.

Sağlıklı bireylerden elde edilen reaksiyon ifade seviyeleri, mean_healthy sütununda özetlenmiş ve her bir reaksiyonun ortalaması cbind fonksiyonu kullanılarak api_id değerleri ile birleştirilerek normal_mean veri seti oluşturulmuştur. Sonraki adımda, normal_mean veri seti semi_join fonksiyonu ile new_df veri setine birleştirilmiştir. Bu işlemle, new_df içerisindeki reaksiyonlar ile eşleşen mean_healthy değerleri korunmuş ve diğerleri çıkarılmıştır. Sonuç olarak, normal_mean içerisinde yalnızca new_df ile ilişkili reaksiyonların ortalamaları bulunan bir veri seti elde edilmiş ve normal_mean_2 olarak yeniden adlandırılmıştır. Bu süreç, sağlıklı örneklerden elde edilen veri ile analizde kullanılacak reaksiyon ifade düzeylerinin ilişkilendirilmesini sağlamıştır ve tablo olarak sonuç bölümünde sunuşmuştur.

```
hmr_results_final_tumor_data_api <- cbind(api_id = api_ids,
hmr_results_final_tumor_data)

hmr_results_final_tumor_data_api_long <-
gather(hmr_results_final_tumor_data_api,
key = "patient_id",
value = "gene_rule",
-api_id)

saglama_data <- hmr_results_final_tumor_data_api_long %>%
semi_join(new_df)

healty_meanwith_api_id<- cbind(api_id = api_ids, mean_healthy %>%
as_tibble())

normal_mean <- healty_meanwith_api_id

normal_mean_2 <- normal_mean %>%
semi_join(new_df) %>%
rename(ort = value)
```

Şekil 5.16 Seçilen reaksiyonların tümör ve sağlıklı örnek verileriyle ilişkilendirilerek 'new_df' için karşılaştırmalı ifade düzeyleri setinin oluşturulması

5.7.5 Değerlendirme ve Sonuçların Doğrulanması

Seçilen reaksiyonların, tümör ve sağlıklı örneklerdeki ifade düzeyleri arasındaki farklılıklar, "saglama_data" ve "normal_mean_2" veri setleri aracılığıyla incelenmiştir. Bu inceleme, analizin doğruluğunu ve yöntemlerin güvenilirliğini doğrulamak için önemlidir. Özellikle, "new_df" veri setinde belirlenen reaksiyonların, sağlıklı örneklerdeki reaksiyon ifade seviyelerinin ortalamasına göre artış veya azalış gösterip göstermediği değerlendirilmiştir.

Analiz sürecinin son aşamasında gerçekleştirilen sağlama işlemi, metodolojimizin doğruluğunu ve güvenilirliğini teyit etmek için kritik bir adımdır. Bu aşamada, "saglama_data" veri setindeki reaksiyon ifade düzeylerinin, sağlıklı örnekler için hesaplanan reaksiyon ifade seviyelerinin aritmetik ortalaması "(normal_mean_2)" ile karşılaştırılması yapılmıştır. Bu karşılaştırma, her bir seçilen reaksiyonun ifade seviyesinin, sağlıklı örneklerin ortalamasına göre artış veya azalış gösterip göstermediğini belirlemeyi amaçlamıştır.

"new_df" veri setinde belirlenen artış ve azalış oranları ile "saglama_data" veri setindeki gerçek ifade düzeyleri arasında tutarlılık gözlemlenmiştir. Bu tutarlılık, seçilen reaksiyonların ifade seviyelerindeki değişikliklerin, analiz sürecinde kullanılan yöntemler ve yapılan varsayımlar doğrultusunda doğru bir şekilde tespit edildiğini göstermektedir. Dolayısıyla, sağlama işlemi, analizimizin ve metodolojimizin doğruluğunu onaylayarak, çalışmamızın bilimsel dayanağını güçlendirmiştir.

Bu doğrulama, reaksiyon ifade düzeylerindeki değişikliklerin anlamlı ve biyolojik olarak önemli olduğunu gösteren kapsamlı istatistiksel analizlerle desteklenmektedir. Sonuç olarak, sağlama işlemi, tümör ve sağlıklı örnekler arasında ifade düzeylerindeki farklılıkların doğru bir şekilde belirlendiğini ve analiz sürecimizin güvenilir olduğunu kanıtlamaktadır. Bu, çalışmamızın meme kanseri gibi kompleks hastalıkların anlaşılması ve tedavisi üzerine potansiyel etkilerini daha da önemli kılmaktadır.

Son olarak analiz sürecinin doğruluğunu kontrol etmek amacıyla, seçilen reaksiyonların ifade seviyeleri üzerinde sağlıklı örneklerin ortalamaları ile karşılaştırmalı bir kutu grafiği oluşturulmuştur. Seçilen reaksiyonların ifade seviyelerinin tümör ve sağlıklı örneklerde nasıl değiştiğini görselleştirmek için kullanılmış bir ggplot grafiğini tanımlanmıştır. ggplot fonksiyonu ile bir dağılım grafiği (geom_jitter) ve kutu grafiği (geom_boxplot) bir arada oluşturulmuştur. "geom_jitter" kullanımı, veri noktalarının birbirinin üstüne binmesini önleyerek dağılımın daha net görülmesini sağlamaktadır.

Kutu grafiği, sağlıklı örneklerin reaksiyon ifade seviyelerinin ortalamasını (orange kutu olarak) göstermektedir ve bu, tümör örneklerindeki ifade seviyeleriyle karşılaştırılmaktadır. "scale_y_log10" kullanımı, reaksiyon ifade seviyelerinin logaritmik ölçekte sunulmasını sağlayarak, geniş değer aralıklarını daha anlaşılır hale getirmektedir. Bu grafik, reaksiyon ifade seviyelerinin tümör örneklerinde sağlıklı örneklerden anlamlı derecede farklılık gösterip göstermediğini görsel olarak doğrulamak için kullanılmış ve görseli sonuç bölümünde sunulmuştur.

5.8 Metabolik Reaksiyonların Frekans Analizi ve Metabolitlerin Belirlenmesi

5.8.1 Veri Setlerinin Bütünleştirilmesi ve Yeni Değişkenlerin Tanımlanması

HMA veri tabanından alınan veri seti(Human-GEM), insan metabolizmasının geniş kapsamlı analizi için kullanılmıştır. İlgili veri setlerinden elde edilen reaksiyonlar ve bu reaksiyonlarla ilişkili metabolik değişiklikler, reaksiyon ifadesinde gözlenen artış veya azalışı belirlemek için kullanılmıştır. İlk olarak, "result_df_166" veri setinde yer alan "api_id" sütunu "RXNID" olarak yeniden adlandırılmış ve "HMR_change_summary_unnested_3" veri seti ile birleştirilmiştir. Bu işlem, iki veri setindeki ortak reaksiyonları eşleştirmek için gerçekleştirilmiştir Şekil5.17. Daha sonra, artan ve azalan metabolik değişimleri tanımlamak için "HMR_change_summary_df" veri setinde "substrate" ve "product" sütunları oluşturulmuş ve ardından artık kullanılmayacak sütunlar çıkarılarak tablo halinde sonuç bölümünde sunulmuştur Tablo 6.11. Bu sütunlar, reaksiyonların substrat ve ürünlerinin metabolik ağdaki değişim yönünü tanımlamıştır.

```
result_df_166 <- readRDS("rds kayit/result_df_166")

result_df_166 <- result_df_166 %>%
  rename(RXNID = api_id)

HMR_change_summary_unnested_3 <- readRDS("rds kayit/HMR_change_summary_unnested_3")

HMR_change_summary_df <- result_df_166 %>%
  inner_join(HMR_change_summary_unnested_3)

# Yeni sütunlar oluştur
HMR_change_summary_df_2 <- HMR_change_summary_df %>%
  mutate(
    substrate = ifelse(change == "increase", "decrease", "increase"),
    product = ifelse(change == "increase", "increase", "decrease"))

HMR_change_summary_df_2 <- HMR_change_summary_df_2 %>%
  select(-num_from_compounds, -num_to_compounds)
HMR_change_summary_df_2 %>%
  sample_n(6)
knitr::kable()
```

Şekil 5.17 Metabolik reaksiyon verilerinin entegrasyonu ve substrat-ürün dağılımının analizi

5.8.2 Metabolik Ağ Analizi İçin Veri Setinin İncelenmesi

HMA veri tabanından elde edilen "human_gem" veri seti ile "HMR_change_summary_df_2" veri seti içerisinde RXNID üzerinden birleştirme işlemi yapılarak, metabolik reaksiyonların "Human-GEM" veri setindeki karşılıkları ile eşlenmiştir. Bu sayede, her bir metabolik reaksiyonun insan metabolik ağındaki alt sistemi de dahil olmak üzere daha kapsamlı bilgiler içeren

bir veri seti elde edilmiştir. Böylelikle, reaksiyonların metabolik yollardaki konumları ve potansiyel biyolojik rolleri hakkında daha detaylı çıkarımlarda bulunulabilmiştir. Yeni veri seti inceleme ve karşılaştırma yapabilmek için "son_veriler" excel dosyası olarak kaydedilmiştir. Bu amaçla "openxlsx" R paketi kullanılmıştır [150].

İlgili metabolik bileşiklerin (substratlar ve ürünler) analizi için, "HMR_change_summary_df_3"ten, "from_compounds" ve "to_compounds" sütunları seçilmiş ve "substrate" ile "product" adlı yeni sütunlar oluşturulmuştur. Sonraki adımda, "from_compounds" ve "to_compounds" sütunlarından elde edilen bu yeni sütunlar "bind_row" komutu ile birleştirilerek her metabolit ve değişim yönü "compound" ve "change_type" sütunlarında tanımlanmıştır.. Bu işlem, metabolik reaksiyonlarda artan veya azalan bileşiklerin belirlenmesi ve bunların reaksiyon yönlerine göre sınıflandırılması için yapılmıştır.

5.8.3 Frekans Analizi ve Veri Setinin Filtrelenmesi

Elde edilen konsolide veri seti üzerinden, her metabolitin ne sıklıkta arttığı veya azaldığı analiz edilmiştir. Bunun için "compound" ve "change_type" sütunları gruplandırılmış ve her grupta kaç tane reaksiyon olduğu sayılmıştır. Bu frekans tablosu, metabolik reaksiyonlar arasındaki değişimlerin genel bir özetini sağlamış ve potansiyel olarak önemli metabolik bileşikleri belirlemekte kullanılmıştır.

Elde edilen frekans tablosundan metabolik ağda sıkça yer alan ve reaksiyonlar arası taşıyıcı görevi gören ATP, NADH, H, H₂O vb. metabolitler hariç tutulmuştur Şekil 5.18. Bu işlem, analizde odaklanılacak metabolitleri daha da daraltmak ve metabolik değişimlerin daha spesifik örneklerine yoğunlaşmak için yapılmıştır.

```

frequency_table <- all_compounds_df %>%
  group_by(compound, change_type) %>%
  summarise(frequency = n()) %>%
  arrange(compound, change_type)

remove_extra_info <- function(compound) {
  compound <- gsub("\\[.*\\]", "", compound)
  gsub("[^a-zA-Z0-9]", "", compound)
}

frequency_table$compound <- sapply(frequency_table$compound, remove_extra_info)

frequency_table$compound <- sapply(frequency_table$compound, remove_extra_info)

excluded_compounds <- c("ATP", "NADH", "ADP", "H", "H2O", "NAD", "NADH", "NADP",
"NADPH")

filtered_frequency_table <- frequency_table %>%
  filter(!(compound %in% excluded_compounds))

filtered_frequency_table %>%
  filter(frequency>3) %>%
  ungroup() %>%
  sample_n(6)

filtered_frequency_table_final <- HMR_change_summary_df_3 %>%
  sinner_join(filtered_frequency_table, by = c("from_compounds" = "compound"))
|
filtered_frequency_table_final

```

Şekil 5.18 Metabolit frekans dağılımlarının analizi ve taşıyıcı moleküllerin hariç tutularak son veri setinin oluşturulması

Daha ileri analizler için, sıkça değişen metabolitlerin belirlenmesi amacıyla belirli bir frekans eşliğinin üzerindeki metabolitler seçilmiştir. Bu eşik değeri, analizin kapsamı ve amaçları doğrultusunda 3 olarak belirlenmiştir ve metabolik ağda önemli değişiklikleri gösteren bileşikler tanımlamak için kullanılmıştır. Örnek sonuçlar Tablo 6.12’de sunulmuştur. Elde edilen bu "filtered_frequency_table" veri seti "HMR_change_summary_df_3" veri seti ile birleştirilerek Tablo 6.13’de görülen daha spesifik reaksiyon ve metabolitler belirlenmiş ve bu metabolitler üzerinden literatür taraması gerçekleştirilerek sonuç bölümünde sunulmuştur.

Bu analizler sayesinde, kanser metabolizmasında önemli değişikliklere sebep olabilecek spesifik metabolik reaksiyonlar ve bileşiklerin artan veya azalan eğilimleri, kanser hücrelerinin metabolik adaptasyonları ve tedaviye yanıt mekanizmaları hakkında önemli bilgiler sağlayabilmektedir. Böylece, bu çalışma, metabolik ağ analizinin kanser araştırmalarında nasıl uygulanabileceğine ve potansiyel terapötik hedeflerin nasıl belirlenebileceğine dair değerli bir örnek teşkil etmektedir.

Bu araştırmanın temel amacı, meme kanserinin moleküler mekanizmalarını derinlemesine incelemek ve bu hastalığın patogeneze ışık tutacak biyomoleküler verileri ortaya koymak ve kanser gelişiminin metabolik reaksiyonlar üzerindeki etkilerini araştırmaktır. Araştırmada, HMA veri tabanından elde edilen veriler kullanılarak, gen ekspresyon kuralları ve RNA sayımları ile TCGA veritabanından alınan tümör verileri bütünleştirilmiştir. Kanserli ve normal hastaların metabolik profillerinin karşılaştırılmasıyla, LIMMA analizi kullanılarak kanser durumunda normale göre artan ve azalan reaksiyonlar belirlenmiştir. Bu yöntemle, kanserle ilişkili değişen metabolitlerin detaylı bir şekilde incelenmesi hedeflenmiştir. Bu süreçte elde edilen bulgular, kanser gelişiminde metabolik reaksiyonların nasıl değiştiğini ve bu değişikliklerin potansiyel biyolojik etkilerini ortaya koymaktadır. Bu entegrasyon sayesinde, ifadesi değişen reaksiyonlar ve bunların metabolik ağ içindeki değişimleri incelenmiş ve literatürdeki bulgularla karşılaştırılarak değerlendirilmiştir. İstatistiksel ve biyoinformatik yöntemlerin kapsamlı kullanımı ile elde edilen bulgular, kanser biyolojisi alanındaki mevcut anlayışımızı zenginleştirmektedir.

Sonuç bölümünde, yöntem bölümünde yapılan analizler ile elde edilen bulgular ve değerlendirmeleri sunulmuş ve bu bulguların sağlık ve hastalık biyolojisi alanındaki etkileri ve uygulamalı araştırmalar için potansiyel katkıları tartışılmıştır. Bu bağlamda, araştırmanın metodolojisi, kritik bulguları ve bunların bilimsel literatürdeki yerinin yanı sıra, gelecekteki araştırmalar için yön gösterici olabilecek potansiyel öneriler ve uygulamalı çalışmalar için öngörüler de ele alınmıştır.

İlk olarak, yöntem bölümünde tanımlanan TCGA veri tabanından elde edilen meme kanseri hastaları ve sağlıklı bireylere ait örneklerin analizleri ile başlanmıştır. Araştırmanın temelini oluşturan bu analizler, R programlama dili kullanılarak gerçekleştirilmiş ve genetik araştırmalarda veri işlemenin ve analizin önemi vurgulanmıştır.

TCGA veritabanından elde edilen "brca_counts" veri seti, 1109 meme kanseri hastası ve 113 sağlıklı bireyin verilerini içermekte olup, hasta ve sağlıklı örnekler arasındaki biyolojik ve genetik farklılıkları inceleyerek araştırmanın derinliğini ve kapsamını artırmakta ve sağlık durumlarına göre detaylı analizler yapılmasına imkan sağlamaktadır.

Analiz sürecinde, "metadata" veri setindeki "file_id" sütunu ile "BRCA" dosyasından elde edilen veri setindeki "file_id" sütunu eşleştirilerek "tcga_brca" veri seti oluşturulmuştur. Bu veri seti, analizler için temel alınmış ve ilk incelemeler, "tumor" sütunundaki lojik değerler ("TRUE" ve "FALSE") aracılığıyla, bireylerin hastalık ve sağlık durumlarının ayırt edilebilmesini sağlamıştır. Tablo 6.1, araştırmada incelenen kanser türünün genetik özellikleri ile tümör varlığının belirlenmesinde kritik bir rol oynamaktadır. Bu tabloda yapılan ayırım, veri setlerinin ve analiz metodolojilerinin, tümör varlığı ile kanser türünün genetik özelliklerinin nasıl belirlendiğini net bir şekilde göstermektedir.

Tablo 6.1 4 hasta ve 2 sağlıklı bireye ait "tcga_brca" veri seti örneği

file_id	tumor
83acd71c-c12a-4394-ba8d-05e9c5ad2cf1	TRUE
b5d2a287-939b-4bfd-aac9-147739a981b6	TRUE
0bf50119-afec-41f8-8b89-262dfb8c3a53	FALSE
6e6649cc-4847-47e5-8f24-c87daf8eca46	TRUE
38d8f158-ead0-405a-9d32-aa4f011dee70	FALSE
d4513df6-0b48-4bf3-b397-3c7b339a47d1	TRUE

6.1 Metabolik Ağın Analizi

HMA veri tabanından indirilen "HumanModel.xlsx" dosyası üzerinde gerçekleştirilen analizler, metabolik reaksiyonların ve bunların bileşenleri arasındaki ilişkilerin detaylı bir şekilde incelenmesine imkan tanımıştır. Dosyanın ilk sayfasındaki verilerin okunması ve boş değer içeren sütunların çıkarılması ile "hm_rxns" adlı veri seti oluşturulmuştur. Bu veri seti, reaksiyonları geri

dönüşümlü ve geri dönüşümsüz olarak ayırmak, reaksiyon denklemlerini sol ve sağ bileşenler olmak üzere iki ayrı sütuna ayırmak ve bu sürecin ardından substrat ve ürünleri içeren uzun bir listenin oluşturulması gibi işlemlere temel oluşturmuştur. Oluşturulan bu liste, metabolik reaksiyonların ve bileşenlerinin birbirleriyle olan ilişkilerini gösteren bir bileşik ağı (compound network) oluşturmak için kullanılmıştır. Bu ağ, "from" ve "to" sütunları aracılığıyla graf yapısındaki kenarları(edges) temsil ederken, metabolik reaksiyonlar ve bunların arasındaki ilişkiler açıkça ortaya konmuştur.

Analiz sonuçlarına göre, oluşturulan bileşik ağı, metabolik reaksiyonların ve bileşenlerinin karmaşık ilişkilerini göstermektedir. Bu ağda, bir reaksiyonun başlangıç ve bitiş noktalarını temsil eden sütunlar, reaksiyonların nasıl birbirleriyle bağlantılı olduğunu ve metabolik süreçlerin nasıl ilerlediğini anlamada kritik öneme sahiptir. Örneğin, "from" sütunu, bir reaksiyonun başlangıç noktasını (örneğin bir substrat veya reaktan) temsil ederken, "to" sütunu ise bitiş noktasını (örneğin bir ürün) temsil etmektedir. Rakamlar, her bir bileşiği veya düğümü (node) temsil eden benzersiz kimliklerdir. Örneğin, Şekil 6.1'teki "from" sütunundaki "1" numaralı "L-lactate[p]" bileşiğinden ve "to" sütunundaki "517" numaralı bir sonraki adımdaki bileşikten, "HMR_4281" reaksiyonuna kadar olan süreç, bu analizin derinliğini ve metabolik yolların anlaşılmasındaki önemini vurgulamıştır. Oluşturulan bileşik ağı, metabolik süreçlerin nasıl organize olduğunu ve belirli reaksiyonların nasıl gerçekleştiğini göstermektedir.

```
compound_network
# A tbl_graph: 6008 nodes and 85934 edges
#
# A directed multigraph with 6 components
#
# A tibble: 6,008 x 2
  name      node_id
  <chr>      <int>
1 L-lactate[p]      1
2 NAD+[p]           2
3 L-lactate[c]      3
4 NAD+[c]           4
5 H2O[c]            5
6 PEP[c]            6
# i 6,002 more rows
#
# A tibble: 85,934 x 3
  from    to RXNID
  <int> <int> <chr>
1     1   517 HMR_4281
2     1   537 HMR_4281
3     1  1757 HMR_4281
# i 85,931 more rows
```

Şekil 6.1 Compound_network veri setinin özeti, 6008 Düğüm ve 85934 Kenar içerir

6.2 GPR İlişkilerinin Metabolik Ağ Analizlerindeki Rolü

Biyolojik sistemlerde genler, proteinler ve metabolik reaksiyonlar arasındaki ilişkileri daha etkili bir şekilde anlayabilmek için gen ifade kuralları da denilen GPR (Gen-Protein-Reaksiyon) kuralları tanımlanmıştır. Bu kurallar, genlerin proteinleri nasıl kodladığını ve bu proteinlerin metabolik reaksiyonlarda nasıl işlev gördüğünü ifade etmektedir. GPR kuralları, özellikle metabolik ağ modellerinde yaygın olarak kullanılır ve gen ifadesi ile metabolik aktiviteler arasındaki ilişkiyi anlamak için temel bir araç olmuştur.

Bir GPR kuralı, genellikle bir veya birden fazla genin ürettiği proteinlerin, belirli bir metabolik reaksiyonu katalize etme yeteneğini ifade eder. Örneğin, bir kural “Gen1 ve Gen2 tarafından kodlanan proteinler, Reaksiyon A’yı katalize eder” şeklinde olabilir. Bu durumda, her iki genin ifadesi gerekli olduğundan, bu genlerin ifade düzeylerinin minimumu (yani her ikisinin de aktif olduğu durum) bu reaksiyonun gerçekleşmesi için gerekli olarak belirtmektedir. Alternatif olarak, “Gen1 veya Gen2’nin aktivasyonu, Reaksiyon B’nin katalizlenmesi için yeterli olabilir” şeklinde olabilir. Burada, genlerden herhangi birinin ifade edilmesi, söz konusu reaksiyonun aktivasyonu için yeterli kabul edilmektedir ve dolayısıyla genlerin ifade düzeylerinin toplamı esas alınmaktadır. Bu perspektif, metabolik ağ analizlerinde gen ifadesinin ve ilgili metabolik aktivitelerin kapsamlı bir şekilde değerlendirilmesi için kritik öneme sahiptir, zira genlerin kooperatif veya redüktif etkileşimleri metabolik süreçlerin dinamiklerini belirlemektedir [103].

GPR ilişkilerinin anlaşılmasını ve bu ilişkilerin metabolik yollar üzerindeki etkilerinin değerlendirilmesi amacıyla HMA veri tabanından elde edilen verilerin analizi, bu kuralların ve metabolik reaksiyonların nasıl birbirleriyle ilişkili olduğunu göstermiştir.

Çalışmamızın bu bölümü reaksiyon ifade kurallarının semantik olarak anlamlı ve analiz edilebilir bir formata dönüştürülmesine odaklanmıştır. “gr_modifier” fonksiyonu kullanılarak mantıksal ifadelerle tanımlanmış gen kurallarının istatistiksel ve matematiksel ifadelere dönüşümü, genetik regülasyonun karmaşık yapısını basitleştirmiştir. Bu basitleştirme, metabolik ağ analizlerinde kullanılan algoritmalar için gen kurallarının daha kesin ve sayısal olarak işlenebilir bir biçimde temsil edilmesini sağlamıştır. Bu yaklaşım, biyolojik sistemlerdeki genler, proteinler ve metabolik reaksiyonlar arasındaki ilişkilerin daha net bir şekilde anlaşılmasına olanak tanımıştır.

Ayrıca, her birey için yapılan özelleştirilmiş HMR hesaplamaları, metabolik reaksiyonların bireysel ifade düzeylerine göre nasıl değişebileceğini

göstermektedir. Elde edilen sonuçlar, genetik ve metabolik verilerin birleştirilmesiyle elde edilen kapsamlı bilgilerin, hastalıkların daha iyi anlaşılması ve tedavi yöntemlerinin geliştirilmesine katkıda bulunabileceğini göstermektedir.

Tablo 6.2 gen kuralları ile her bir hasta için hesaplanan RNA sayımlarını ve reaksiyonları bir araya getiren bir süreci özetlemiştir. Her satır, belirli bir metabolik reaksiyonu kodlayan genlerin birleşimini ve reaksiyonların ifade düzeylerinin nasıl işlendiğini göstermektedir. Tablonun her bir bileşenini detaylı olarak açıklayacak olursak:

api_id: Her bir metabolik reaksiyon için atanan benzersiz bir tanımlayıcıdır. Bu ID, reaksiyonun HMA veritabanındaki spesifik girişine karşılık gelir.

gene_rule: Reaksiyonun gerçekleşmesi için gerekli olan genlerin ifade kurallarını göstermektedir. Bu kurallar, genlerin proteinleri nasıl kodladığını ve bu proteinlerin metabolik reaksiyonlarda nasıl işlev gördüğünü ifade etmektedir. Örneğin:

- HMR_0001 reaksiyonu için ENSG00000074416 or ENSG00000100997 or ENSG00000115884 ifadesi, listelenen genlerden herhangi birinin aktif olmasının reaksiyon için yeterli olduğunu göstermiştir.
- HMR_0786 için ENSG00000115850 and ENSG00000163521 ifadesi ise bu genlerin hepsinin ifadesinin reaksiyonun gerçekleşmesi için gerekli olduğunu belirtmiştir.

RNA sayım değerleri: Her bir sütun, farklı hastalar için ölçülen RNA sayım değerlerini temsil etmektedir. Bu sayımlar, genlerin ifade düzeylerinin kantitatif bir ölçüsüdür. Sayımlar, gen kurallarına göre ya toplanmıştır (or mantığı ile) ya da minimum değer alınarak (and mantığı ile) hesaplanmıştır.

Tablo 6.2 HMR_result_final veri setinde bulunan iki meme kanseri hastasının metabolik reaksiyon kodları, GPR kuralları ve gen ifade sayımlarından örnekler

api_id	gene_rule	00156c51- 2686-4736- 88c1-c30c81ec0a12	001fe63b- 06c8-4ca1- 861a- 3d1dd0fdf5fc
HMR_0001	ENSG00000074416 or ENSG00000100997 or ENSG00000115884	47099	15221
HMR_0002	ENSG00000175445	5579	330
HMR_0786	ENSG00000115850 and ENSG00000163521	2	1

Detaylı olarak incelendiğinde, bu tablo, gen kurallarının ve bunların RNA sayım değerleri ile nasıl işlendiğinin anlaşılmasında kritik bir rol oynamıştır. Bu işlem, gen ifadesinin karmaşık düzenini ve metabolik reaksiyonlar üzerindeki etkisini anlamak için önemlidir. Her bir reaksiyonun, genlerin ifade düzeylerine göre nasıl etkilenebileceğini ve bu etkileşimin hastalık durumları veya metabolik durumlarla nasıl ilişkilendirilebileceğini göstermiştir. Sonuç olarak, bu tablo, genetik verilerin işlenmesi ve metabolik yolların analizi için bir metodoloji sağlamıştır ve biyomedikal araştırmalarda potansiyel uygulamalara işaret etmektedir.

6.3 LIMMA Analizi

Diferansiyel ifade analizi genellikle gen düzeyinde gerçekleştirilmektedir, ancak bu çalışmada metabolik reaksiyonların ifadesindeki değişiklikleri incelemek için kullanılmıştır. Metabolik reaksiyonların ifadesindeki değişimler, metabolik yolların ve ağların aktivitesindeki değişiklikleri yansıtarak, hastalık durumları hakkında önemli bilgiler sağlayabilmektedir. Metabolik reaksiyonların ifade düzeylerindeki değişiklikleri incelemek için LIMMA analizi uygulanmıştır. Bu analizle, sağlıklı ve hasta bireyler arasındaki metabolik farklılıkları anlamak amaçlanmıştır.

"hmr_results_final.rds" dosyasından "api_id" ve "gene_rule" sütunları çıkarılmış ve elde edilen matris LIMMA analizi için kullanılmıştır. Bu işlemle, metabolik reaksiyonların ifade düzeylerindeki değişikliklerin tespit edilmesi hedeflenmiştir. LIMMA analizi sonucunda elde edilen diferansiyel ifade gösteren reaksiyonlar,

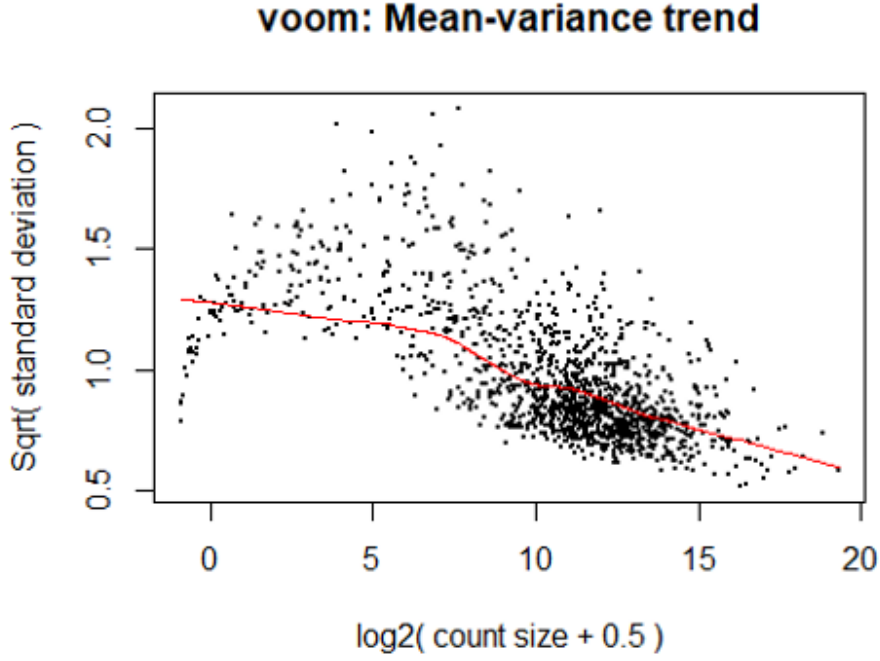
daha önce oluşturulan ve reaksiyonları ile bileşikleri içeren "compound_network" ile entegre edilmiştir. Bu entegrasyon, kanser durumunda hangi metabolik yolların etkilendiğini ve hangi metabolitlerin arttığını veya azaldığını ortaya koymak için yapılmıştır.

RNA-seq veri setimizin normalleştirilmesi ve analizi için voom dönüşümü kullanılmıştır. voom metodu, gen ifadesi sayımlarının ortalama değerleri ile bunların varyansları arasındaki ilişkiyi inceleyerek, sayım verilerinde gözlenen teknik varyansı azaltmayı amaçlamaktadır. Voom analizi, RNA-seq verilerinin ortalama-değişkenlik ilişkisini değerlendirmekte ve bu ilişkiyi gösteren bir trend çizgisi ile görselleştirmektedir. Grafik, reaksiyon ifade sayımlarının (count size) logaritmik dönüşümünü ($\log_2(\text{count size} + 0.5)$) x ekseninde, ve bu ifade sayımlarının karekök standart sapmalarını ($\sqrt{\text{standard deviation}}$) y ekseninde göstermiştir Şekil 6.2.

Grafikteki siyah noktalar, her bir reaksiyon ifadesi sayımının standart sapmasını temsil ederken, kırmızı eğri, bu standart sapmaların ifade sayımlarının logaritmik değerleriyle nasıl değiştiğini gösteren ortalama-değişkenlik trendidir. Voom analizinin gücü, RNA-seq verilerinin heteroskedastikliğini (yani, büyük sayımlarda daha büyük varyans) dikkate alarak veri setini normalleştirmesinden kaynaklanmaktadır

Bu trend çizgisi, düşük sayım değerlerinde yüksek varyans olduğunu ve sayım değeri arttıkça varyansın azaldığını göstermektedir. Bu durum, RNA-seq verilerinin genel bir özelliği olan 'ortalama ile varyans arasındaki bağımlılığı' yansıtmaktadır. voom ile elde edilen bu trend, sonraki diferansiyel ifade analizleri için veri setinin uygun şekilde normalleştirildiğini ve stabil olduğunu göstermiştir.

Sonuç olarak, bu grafik, analizimizin veri normalleştirme ve varyans stabilizasyonu aşamalarında RNA-seq veri setimizin kalitesinin ve güvenilirliğinin bir göstergesi olarak tezde yer almıştır. Bu normalleştirme süreci, sonraki diferansiyel ifade analizlerinin doğruluğunu ve güvenilirliğini artırmış, böylece elde edilen sonuçları biyolojik olarak daha anlamlı ve güvenilir hale getirmiştir.



Şekil 6.2 Voom Grafiği

Tablo 6.3, diferansiyel reaksiyon ifadesi analizinin sonuçlarını özetlemekte ve araştırmada belirlenen kritik reaksiyonların bir listesini sunmaktadır. Bu süreç, tümör ve normal dokular arasındaki reaksiyon ifade düzeyleri farklılıklarını belirlemeye odaklanmıştır. LIMMA paketinin kullanımı, RNA-seq verilerinin analizinde güçlü ve esnek bir yaklaşım sunmuştur. Voom dönüşümü, düşük sayımların neden olduğu varyans sorunlarını gidererek, verilerin daha doğru bir şekilde analiz edilmesini sağlamıştır. lmFit ve eBayes fonksiyonları aracılığıyla, anlamlı değişiklikler gösteren reaksiyonlar tespit edilmiştir. Bu reaksiyonlar, kanser biyolojisi ve tedavi stratejilerinin geliştirilmesinde potansiyel olarak önemli hedefler olabilmektedir.

Tablo 6.3 İfadesi değişen reaksiyonların analiz sonuçları

RXNID	logFC	AveExpr	t	P.Value	adj.P.Val	B
HMR_5519	2.4807226	5.912295	13.6465934	0.0000000	0.0000000	78.853806
HMR_1050	0.0388511	9.955634	0.7955768	0.4264322	0.4628686	-7.952515
HMR_6079	0.2565001	-6.417149	2.0047120	0.0452137	0.0525280	-5.100464
HMR_2152	0.3796068	9.406763	3.1456336	0.0016972	0.0023325	-3.261680
HMR_1057	-0.1502137	7.528969	-1.9945955	0.0463084	0.0534699	-6.061774
HMR_6185	-0.4029802	-6.067815	-2.4972368	0.0126470	0.0153414	-4.022539

Elde edilen sonuçlardaki sütunlar, gen ifadesi analizinde sıkça kullanılan

istatistiksel ölçütleri temsil etmektedir. Her bir sütunun anlamı şöyledir:

RXNID: Bu, her bir metabolik reaksiyonu tanımlayan benzersiz bir tanımlayıcıdır. Bu örnekte, “HMR_5519”, “HMR_1050” gibi tanımlayıcılar farklı metabolik reaksiyonları ifade etmektedir.

logFC (log Fold Change): Reaksiyon ifadesindeki değişimin logaritma oranıdır. Pozitif değerler, artan ifadeyi (upregulation); negatif değerler ise azalan ifadeyi (downregulation) göstermektedir. Örneğin, “2.4807226” değeri, ifadenin arttığını ve bu artmanın yaklaşık 302.5 kat logaritmik oranında olduğunu göstermiştir.

AveExpr (Average Expression): Bu sütun, incelenen reaksiyonun ortalama ifade düzeyini göstermektedir. Burada “5.912295” değeri, reaksiyonun ortalama ifade düzeyinin bu değer olduğunu belirtmiştir.

t: t-test istatistiği: Bu değer, gruplar arasındaki farkın istatistiksel olarak anlamlı olup olmadığını değerlendirmek için kullanılmaktadır. Büyük t değerleri, gruplar arasında daha büyük bir fark olduğunu göstermektedir. Örneğin, "13.6465934" değeri, gruplar arasında anlamlı bir fark olduğunu gösterirken, "2.0047120" değeri ise gruplar arasında çok küçük ve istatistiksel olarak önemsiz bir fark olduğunu işaret etmektedir.

P.Value: Bu, t-testinin p-değeri olup, bulunan farkın rastgele bir varyans sonucu olup olmadığını değerlendirmektedir. Düşük bir p-değeri (örneğin 0), gözlemlenen farkın tesadüfi olmadığını, yani istatistiksel olarak anlamlı olduğunu göstermektedir.

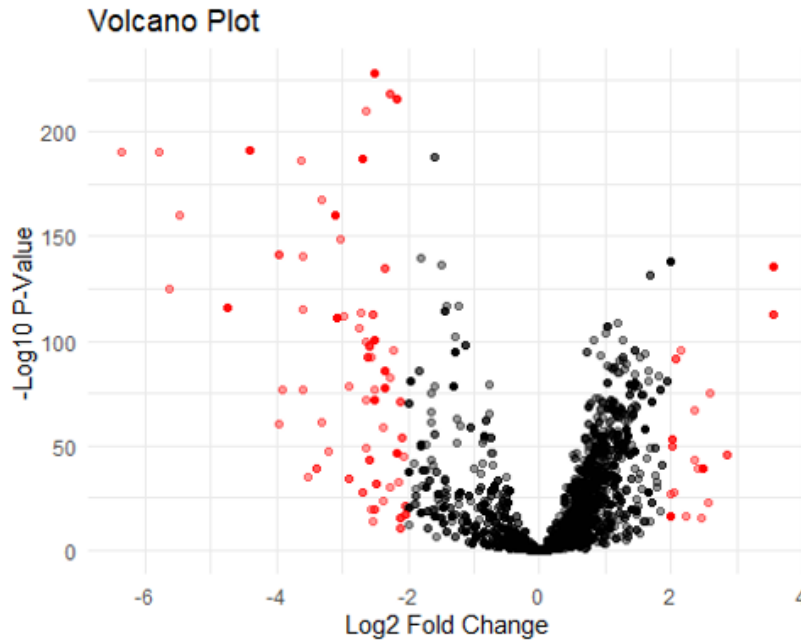
adj.P.Val (Adjusted P-Value): Bu değer, çoklu karşılaştırmaların düzeltilmesi sonucu elde edilen düzeltilmiş p-değeridir. Gen ifadesi analizlerinde, birçok gen aynı anda karşılaştırıldığı için, hata oranını kontrol etmek adına p-değerleri düzeltilmektedir. Örneğin Tablo 6.3’de, 1. reaksiyonumuz için düzeltilmiş p-değeri 0’dır, bu da sonucun yüksek derecede güvenilir olduğunu göstermiştir.

B: Bu sütun, reaksiyonun diferansiyel ifadesinin güvenilirliğini değerlendiren bir ölçüttür. Yüksek B değerleri, bulguların daha güvenilir olduğunu göstermektedir. Örneğin, “78.853806” oldukça yüksek bir B değeridir ve bu, sonucun oldukça güvenilir olduğunu göstermiştir.

Çalışmamızda, diferansiyel reaksiyon ifadesi analizinin bir parçası olarak "Volcano plot" kullanılmıştır. Bu grafik, reaksiyon ifade düzeylerindeki logaritmik kat değişikliği (logFC) ile istatistiksel anlamlılığı (-log₁₀ P-value) iki boyutlu olarak görselleştirmiştir. Volcano plot üzerinde, anlamlı derecede artan veya azalan

reaksiyon ifadeleri kırmızı ve siyah noktalarla vurgulanmıştır Şekil 6.3. Bu plotlar, logFC ile $-\log_{10}$ p-değerlerini birleştirerek, reaksiyon ifadesindeki değişimlerin hem büyüklüğünü hem de istatistiksel anlamlılığını ortaya koymaktadır.

Grafiğin genel dağılımı, veri setinin analize uygun olduğunu göstermektedir. Bu durum, diferansiyel olarak ifade edilen reaksiyonların listesinin, analiz için güvenilir olduğunu ve araştırmanın sonraki aşamalarında kullanılabileceğini belirtmektedir. Anlamlı olarak değişim gösteren reaksiyonların listesi, hastalığın moleküler mekanizmalarını anlamak ve tedavi stratejilerinin geliştirilmesi için potansiyel hedefler sunmuştur. Sonuç olarak, Volcano plot analizi, diferansiyel reaksiyon ifadesi analizindeki sonuçların doğruluğunu ve güvenilirliğini teyit etmiş ve araştırmamız için sağlam bir temel oluşturmuştur. Bu yaklaşım, araştırmanın ilerleyen aşamalarında, özellikle metabolik yol analizleri ve hastalık ilişkili metabolik değişikliklerin incelenmesinde kilit rol oynamıştır. Kırmızı ile işaretlenmiş reaksiyonlar, kanser biyolojisi ve tedavi stratejilerinin geliştirilmesi açısından özellikle incelenmiş ve böylece sağlıklı ve hasta bireyler arasındaki metabolik farklılıkların derinlemesine anlaşılmasına katkı sağlamıştır.



Şekil 6.3 Volkan Plot Grafiği

6.4 Metabolik Değişikliklerin Analizi ve Özetlenmesi

Sağlıklı ve hasta bireyler arasında metabolik reaksiyonlardaki farklılıkların incelenmesi ve kanserin metabolik etkilerinin anlaşılması amacıyla farklı ifade verileri ve metabolitlerin olduğu veri setleri birleştirilerek analize hazır hale getirilmiştir Tablo 6.4 .

Tablo 6.4 "limma_compound_network" veri setinin tablo gösterimi

from	to	RXNID	logFC	AveExpr	t	P.Value	adj.P.Val	B
29	353	HMR_5043	0.2266261	8.268468	5.3594205	0.0000001	0.0000002	6.1011097
2714	2605	HMR_5152	0.8639198	1.912168	3.7998213	0.0001519	0.0002205	-0.2476936
64	1684	HMR_6436	0.1221935	10.378699	4.5883643	0.0000049	0.0000075	2.1358633
265	66	HMR_1067	0.1221935	10.378699	4.5883643	0.0000049	0.0000075	2.1358633
55	3100	HMR_1249	-0.8338543	8.594537	-16.5661850	0.0000000	0.0000000	115.6098070
46	1011	HMR_6235	0.0334978	4.445973	0.3024465	0.7623631	0.8045871	-7.6980660

Kanser durumunda metabolik reaksiyonların değişimi, kanser hücrelerinin metabolik ihtiyaçlarının ve çevresel adaptasyonunun bir yansımasıdır. Artan ve azalan bileşiklerin belirlenmesi sayesinde, kanserli durumda metabolizmanın nasıl yeniden düzenlendiği gözlemlenmiştir. Bu yeniden düzenleme, kanser hücrelerinin proliferasyonu, büyümesi ve hayatta kalması için gerekli enerji ve yapısal bileşenleri sağlamak üzere metabolik yolların modifikasyonunu içermektedir. Bu bölümdeki analiz, artan ve azalan bileşiklerin belirlenmesi, metabolik reaksiyonların (RXNID) özetlenmesi, ve bu reaksiyonlarda yer alan bileşiklerin (from ve to) etkileşim ağlarının zenginleştirilmesi adımlarını içermektedir.

Analiz, logFC değerlerine göre artan ($\logFC > 0$) ve azalan ($\logFC < 0$) metabolik bileşikleri sınıflandırmıştır. Bu sınıflandırma, her bir reaksiyon için özgün bileşiklerin değişim yönlerini (artış ya da azalış) belirlemek için kullanılmıştır. Bu sayede, kanser gelişiminde önemli roller oynayabilecek metabolik yolların belirlenmesine yardımcı olacak temel bir veri seti oluşturulmuştur.

Her bir metabolik reaksiyon için artan ve azalan bileşiklerin sayıları özetlenmiş, bu özetler "HMR_summary" veri seti olarak kaydedilmiştir. Ayrıca, metabolik ağın kenar bilgileri, "from" ve "to" sütunlarında yer alan indeksler kullanılarak zenginleştirilmiştir. Bu işlem, metabolik ağdaki etkileşimlerin daha anlaşılır hale getirilmesine olanak tanımıştır.

Metabolik reaksiyonlardaki değişiklikler, "from_name" ve "to_name" sütunlarına göre özetlenmiş ve "HMR_change_summary" adında bir veri seti oluşturulmuştur. Daha sonra, bu veri seti detaylandırılarak "HMR_change_summary_unnested" adında yeni bir veri seti oluşturulmuş ve bu veri setinde her bir reaksiyon için özgün bileşiklerin sayıları ve değişiklik yönleri Tablo 6.5'te sunulmuştur.

Tablo 6.5 HMR_change_summary_unnested Veri Setinden Örnek

RXNID	change	from_compounds	to_compounds
HMR_5162	increase	L-glutaminyl-tRNA(gln)[c]	tRNA(thr)[c]
HMR_2056	decrease	estrone[l]	NADP+[l]
HMR_0245	decrease	9-heptadecylenic acid[c]	CoA[c]
HMR_1068	increase	CoA[c]	PPi[c]
HMR_4606	increase	histidine[c]	PPi[c]
HMR_6032	increase	propanoate[c]	propanoate[s]

Bu görselleştirme, kanser gelişimindeki metabolik reaksiyonların potansiyel rolünü anlamak için kritik bir araçtır. Örneğin, "HMR_5162" ve "HMR_2056" reaksiyonlarındaki değişiklikler sırasıyla L-glutaminyl-tRNA metabolitinin artışı ve estrone metabolitinin azalışını göstermektedir. Bu artış ve azalışlar, kanser metabolizmasının anlaşılmasında önemli ipuçları sunmuştur.

Bu çalışmanın ardından "Human-GEM" veri setinden yararlanılarak oluşturulan veri seti, sinyal-gürültü oranını iyileştirmek ve ana metabolik işlevlere odaklanmak için özel filtreleme işlemleriyle daraltılmıştır. Analizde, metabolik ağın temel akışını yansıtmayan 'Pool reactions', 'Transport reactions' ve 'Isolated' gibi özel reaksiyonlar filtrelenmiştir. Bu adım, analizin odaklanacağı anahtar metabolik reaksiyonların daha net bir şekilde belirlenmesini ve metabolik ağın temel işlevsel durumunun daha etkili bir şekilde gözlemlenmesini sağlamıştır.

Bu reaksiyonlar, genellikle metabolik ağın ana akışını temsil etmeyen özel durumlar olduğu için çıkarılmıştır. "Pool reactions"; metabolik ağın genel dengesini etkileyebilecek genel toplam reaksiyonları, "Transport reactions"; hücreler arası ve organeller arası taşıma işlemlerini, "Isolated"; metabolik yolların ana akışına doğrudan bağlı olmayan izole reaksiyonları temsil eder. Bu reaksiyonlar, metabolik ağ analizindeki sinyal-gürültü oranını iyileştirmek ve ağın ana işlevsel durumunu daha net gözlemlemek için çıkarılmıştır [151–153].

Gene association sütunundaki veriler üzerinde yapılan sayımsal analizle, reaksiyonların genetik regülasyon karmaşıklığı incelenmiştir. Belirlenen eşik değeri (17 "or" sayısı), analizin daha az genetik regülasyona sahip, dolayısıyla metabolik ağın temel işlevlerine daha yakından bağlı reaksiyonlara odaklanmasını sağlamıştır. Bu, çalışmanın odak noktasını, kanser metabolizmasındaki potansiyel değişikliklere yönlendirmeyi amaçlamıştır [154].

Metabolik ağ analizinde odak noktasını daraltmak için, filtrelenmiş veri setimiz

ile "HMR_change_summary_unnested" veri seti "anti_join" işlevi kullanılarak birleştirilmiştir. Ardından veri seti içinde, "from" ve "to" bileşikleri için belirlenen sayısal eşikler 5'ten küçük olacak şekilde uygulanmıştır. Bu işlemler, metabolik ağ analizimiz için spesifik olmayan veya ana metabolik akışa dahil olmayan reaksiyonları hariç tutarak, kanser metabolizması üzerinde potansiyel etkisi olan reaksiyonlara odaklanmamızı sağlamaktadır. Bu seçici filtreleme yaklaşımı, analizin karmaşıklığını azaltmış ve bilimsel araştırmamızda, hastalığın biyolojik mekanizmalarına ve terapötik müdahalelere daha doğrudan bir etki potansiyeli taşıyan metabolik reaksiyonlar üzerinde yoğunlaşmamıza imkan tanımıştır.

Eşik değeri olarak beşin seçilmesinin altında yatan sebep, genellikle metabolik ağ çalışmalarında gözlemlenen bir paternin sonucudur. Araştırmalar, çok sayıda bileşen içeren reaksiyonların genellikle metabolik ağın merkezi işleyişine daha az katkıda bulunduğunu ve daha az spesifik olduğunu göstermiştir. Öte yandan, daha az bileşen içeren reaksiyonlar, genellikle metabolik ağın kritik düğümleri olarak işlev görebilir ve hastalığın patofizyolojisinde önemli roller oynayabilir. Bu nedenle, az sayıda bileşik içeren reaksiyonlar, metabolik ağın en etkili elemanlarını ve hastalık ile doğrudan ilişkili potansiyel terapötik hedefleri ortaya çıkarmak için filtrelemede kullanılmıştır [155].

Tablo 6.6, filtrelenmiş veri setinden alınan örneklere dayalı olarak, kanser metabolizmasında değişen bazı kritik metabolik reaksiyonları özetlemektedir. Örneğin, "HMR_8062" reaksiyonunda, sülfotransferazlar (SULT1A3) tarafından katalize edilen bir süreç gözlemlenmektedir. Bu reaksiyon, NADPH, 3'-Phosphoadenosine-5'-phosphosulfate ve dehydroalanine kullanarak L-cysteate, NADP+ ve 3'-Phosphoadenosine-5'-phosphate üretir. Bu süreç, kanser hücrelerinin metabolik adaptasyonları ve enerji üretim mekanizmaları ile ilişkilendirilebilir, ancak bu durumda oksidatif fosforilasyonun artırılmasıyla doğrudan bir bağlantı kurmak yerine, kanser metabolizmasının daha geniş bir yelpazede incelenmesi gerekmektedir. [156].

Diğer yandan, "HMR_3750"'de isobutyryl-CoA'nın azalması, belirli amino asit metabolizması yollarındaki değişiklikleri göstermektedir. Bu azalma, kanser hücrelerinin belirli amino asitleri enerji üretimi veya diğer biyosentetik yollar için kullanma biçimindeki değişikliklerle ilişkili olabilmektedir [157].

"HMR_2837" ve "HMR_1927"'deki artışlar, sırasıyla karnitin ve oksijen metabolizmasındaki değişikliklere işaret etmektedir. Karnitin seviyelerindeki artış, yağ asitlerinin mitokondriye taşınmasında ve enerji üretiminde bir artışa bağlı olabilmektedir. Oksijen ve suyun artması, kanser hücrelerinin oksidatif

stres ile başa çıkma kapasitesinde veya metabolik düzenlemelerde bir değişikliği göstermiştir [158, 159].

Bu örnekler, metabolik reaksiyonlardaki değişikliklerin genel görünümünü sağlarken, kanser metabolizmasındaki adaptasyonları ve bu adaptasyonların kanser gelişimi ve ilerlemesi üzerindeki olası etkilerini anlamada kritik bileşen yolları keşfetme potansiyelini vurgulamıştır.

Tablo 6.6 HMR_change_summary_unnested_3 Veri Setinden Örnek

RXNID	change	num_from_compounds	num_to_compounds	from_compounds	to_compounds
HMR_5261	increase	3	2	ATP[l]	arginine[l]
HMR_8062	increase	4	3	dehydroalanine[c]	PAP[c]
HMR_2905	increase	4	4	L-carnitine[r]	docosa trienoyl carnitine[r]
HMR_3750	decrease	4	4	isobutyryl glycine[m]	isobutyryl-CoA[m]
HMR_2837	increase	4	4	eicosa noyl carnitine[r]	eicosa noyl carnitine[c]
HMR_1927	increase	4	3	O2[m]	H2O[m]

Yukarıdaki filtreleme işlemlerinin ardından tümör ve sağlıklı örneklerin ayrımını sağlamak ve başka analizler için yeni veri seti oluşturmak amacıyla "tcga_brca" ve "hmr_results_final" veri setleri kullanılarak tümör ve sağlıklı dokulara ait reaksiyon ifadesi profilleri oluşturulmuştur. Analiz sonucunda, tümör ve sağlıklı örnekler arasında reaksiyon ifadesi düzeylerinde önemli farklılıklar saptanmıştır. Ortalama hesaplama ve sapma analizi ile sağlıklı örnekler üzerinden elde edilen reaksiyon ifade seviyelerinin aritmetik ortalaması, tümör örneklerindeki reaksiyon ifade seviyeleri ile karşılaştırılmış ve her bir reaksiyon için sapma değerleri hesaplanmıştır. Bu sapma değerleri, tümör örneklerindeki reaksiyon ifade düzeylerindeki artış veya azalışı belirlemek için kullanılmıştır.

Frekans analizi, tümör örneklerindeki reaksiyon ifade seviyelerindeki değişimlerin "artış" veya "azalış" olarak sınıflandırılmasını sağlamış, ve bu sınıflandırma, her bir reaksiyonun örnekler içerisindeki yüzdesel dağılımının belirlenmesine olanak tanımıştır. Elde edilen sonuçlar, "api_id" ile birlikte artış ve azalış etiketlerinin yüzdesel değerlerini içeren "result_df" veri setinde kaydedilmiştir.

Eşik değeri belirleme ve filtreleme aşamasında, yüzde 90'ın üzerindeki değişimleri

gösteren reaksiyonların seçilmesi, analizlerin hem istatistiksel hem de biyolojik olarak anlamlı sonuçlar üretmesini sağlamıştır [160]. Bu süreç, analizdeki istatistiksel gürültüyü azaltmış ve yanlış pozitif sonuçların minimizasyonuna katkıda bulunmuştur.

Filtreleme sonrasında, artış gösteren gen sayısı 83, azalış gösteren gen sayısı ise 1158 olarak belirlenmiştir. Her iki gruptan eşit sayıda, en yüksek değerlerden itibaren ilk 83 genin seçilmesi, dengeli bir veri seti oluşturulmasını sağlamıştır.

Filtreleme sonrası elde edilen veriler, reaksiyon ifadelerindeki değişimlerin belirgin bir dengesizliği ortaya koymaktadır. Artış gösteren reaksiyonların sayısı sadece 83 iken, azalış gösteren reaksiyonların sayısı 1158 olarak gözlemlenmiştir. Bu, analizimizde karşılaşılan en çarpıcı sonuçlardan biridir ve kanser hücrelerindeki metabolik reaksiyonların çoğunluğunun düşüş gösterdiğini göstermektedir. Bu dengesiz dağılım, çalışmamızın ana çıktılarından biri olarak, kanser biyolojisi ve potansiyel tedavi hedeflerinin anlaşılmasında kritik öneme sahiptir. İstatistiksel olarak önemli olan bu reaksiyonlar, artan ve azalan gruplardan eşit sayıda seçilerek, en yüksek değişim gösteren ilk 83 reaksiyon üzerinden dengeli bir veri seti oluşturulmuştur. Artış ve azalış değerleri sırasıyla 0.90'ın üzerinde olan reaksiyonlar özel olarak incelenmiştir.

Tablo 6.7'de örneği sunulan seçili reaksiyonların ifadelerindeki artış ve azalış değerleri, tümör ve sağlıklı dokular arasındaki reaksiyon ifadesi farklılıklarını detaylı bir şekilde ortaya koymuştur. Örneğin, "HMR_4030" reaksiyonunun tümör örneklerinde sağlıklı örneklerle nazaran yüzde 93 artış gösterdiği, "HMR_0173" reaksiyonunun ise neredeyse tümör örneklerinin tamamında azalış (%99.72949) gösterdiği görülmektedir. Bu bulgular, söz konusu reaksiyonların tümör gelişiminde potansiyel olarak önemli roller oynadığını ve bu reaksiyonların ifadesindeki artış ve azalışların tümörle ilişkili olabileceğini göstermektedir.

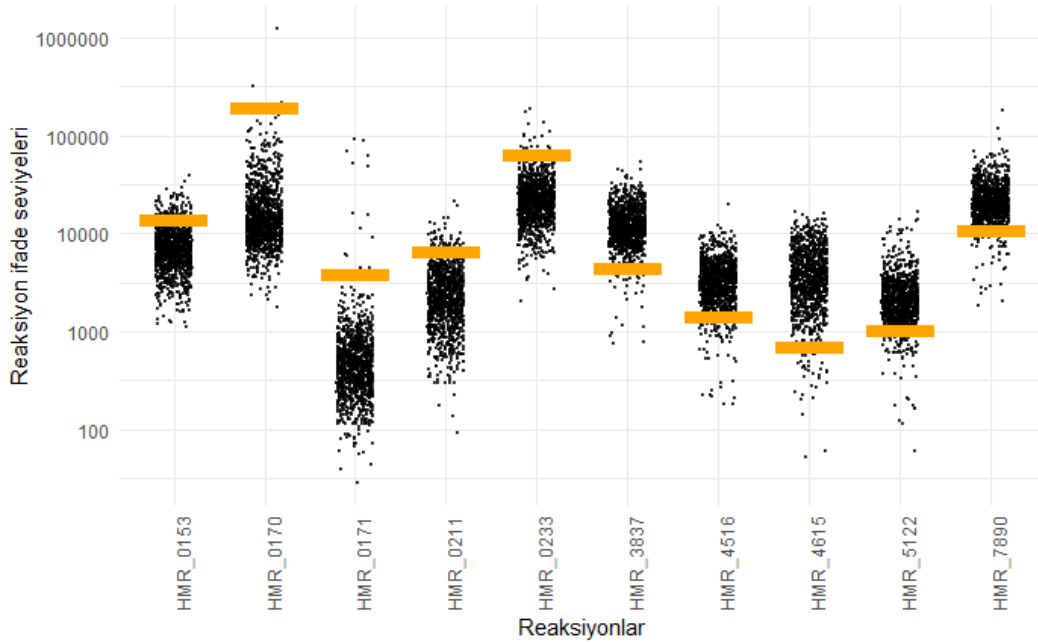
Tablo 6.7 Seçili reaksiyonların tümör örneklerinde sağlıklı kontrollere göre gösterdiği ifade artış ve azalış oranları

api_id	Increase	Decrease
HMR_0165	0.0144274	0.9855726
HMR_4030	0.9305681	0.0694319
HMR_4016	0.9305681	0.0694319
HMR_0209	0.0270514	0.9729486
HMR_6385	0.9522092	0.0477908
HMR_0173	0.0027051	0.9972949

Yukarıda anlatılan süreçler, metabolik ağ analizindeki derinlemesine inceleme ve filtreleme adımlarını yansıtmaktadır. Bu metodoloji, biyolojik verilerin karmaşıklığı ve büyüklüğü göz önünde bulundurularak, en alakalı ve anlamlı verilerin çıkarılmasını ve analiz edilmesini hedeflemektedir. Sonuç olarak, bu yöntemler, Kanserli ve normal hastalar arasındaki metabolik profil karşılaştırması, kanser gelişimiyle ilişkili metabolik yolların ve metabolitlerin tanımlanmasını sağlamıştır. Artan ve azalan metabolik bileşiklerin belirlenmesi, kanser hücrelerinin metabolik adaptasyonlarının ve bu adaptasyonların kanser gelişimi üzerindeki potansiyel etkilerinin daha iyi anlaşılmasına katkıda bulunmuştur.

6.5 Artan ve Azalan Reaksiyonların Kontrolü

Şekil 6.4'daki veriler, reaksiyon ifade düzeylerinin hastalarda nasıl değiştiğini gösteren bir kutu grafiği ve nokta grafiğinin kombinasyonudur. Her bir `api_id`, belirli bir reaksiyonu temsil etmekte ve kutu grafiği bu reaksiyonların ifade düzeylerindeki merkezi eğilim ve dağılımı gösterirken, nokta grafiği (jitter plot) ise bireysel örneklerin ifade düzeylerini detaylandırmıştır. Analiz sonuçlarının doğrulanması süreci, öncelikle reaksiyon ifade düzeylerindeki değişimlerin yüzde 90'dan fazla örnekte artış veya azalış gösterdiğini (increase/decrease) belirlemek için yapılan filtrelemenin doğruluğunu test etmektedir. "HMR_0032" API ID'sine sahip reaksiyon için, yüzde 90'dan fazla bir azalışın gerçek dünya verileriyle uyumlu olup olmadığını değerlendirmek için bir kontrol mekanizması işletilmiştir.



Şekil 6.4 Reaksiyonların ve GPR kurallarına göre hesaplanan ifade düzeylerinin kutu grafiği ile gösterimi

Grafikteki her bir kutunun ortasındaki yatay çizgi, medyan reaksiyon ifade düzeyini göstermektedir ve kutunun yüksekliği interkuartil aralığı (IQR) yani medyanın %25 altı ve %75 üstü verilerini kapsamaktadır. Noktalar ise bireysel örnekleri temsil etmekte ve reaksiyon ifade düzeylerinin bireysel dağılımını sergilemektedir. Örnek verecek olursak "HMR_0032" reaksiyonu için, kutu grafiğinde turuncu çizgi ile temsil edilen medyan değer 136,518.9204'tür. Bu değer "normal_mean_2" veri setinden türetilmiştir. Tablo 6.9'deki "saglama_data" veri setinde aynı reaksiyona karşılık gelen gen kuralı değeri 63911'dir, Bu değer medyanın yaklaşık %98 altında ve önemli ölçüde düşük olup, grafikte medyan çizgisinin altında yoğun bir kümeleme oluşturmuştur. Tersine, "HMR_3837" reaksiyonu için medyan değer 4373.1327 iken, "saglama_data" veri setinde "HMR_3837" reaksiyonunun gen kuralı değeri 7935'tir. Bu değer, medyanı yaklaşık %97 aşarak, grafikte medyan çizgisinin üstünde bir kümeleme oluşturmuştur. Tablo 6.8'deki "New_df" veri seti, "HMR_0032" için %98'lik bir azalma ve "HMR_3837" için %97'lik bir artış göstererek bu bulgularımızı doğrulamıştır.

Tablo 6.8 Artış ve Azalış Gösteren Reaksiyonların Karşılaştırılması İçin Seçilen Reaksiyon Örnekleri: new_df Veri Seti Üzerinden Kontrol

api_id	increase	decrease
HMR_3837	0.9711452	0.0288548
HMR_4028	0.9305681	0.0694319
HMR_4618	0.9657349	0.0342651
HMR_5465	0.9314698	0.0685302
HMR_6616	0.9305681	0.0694319
HMR_0032	0.0117223	0.9882777
HMR_0199	0.0027051	0.9972949
HMR_0211	0.0604148	0.9395852
HMR_0232	0.0027051	0.9972949
HMR_0258	0.0027051	0.9972949

Tablo 6.9 Artış ve azalış gösteren reaksiyonlarla karşılaştırılması için seçilen hasta örnekleri ve gen kurallarının örnekleri: saglama_data veri seti üzerinden kontrol

api_id	patient_id	gene_rule
HMR_0032	83acd71c-c12a-4394-ba8d-05e9c5ad2cf1	63911
HMR_0199	83acd71c-c12a-4394-ba8d-05e9c5ad2cf1	14033
HMR_0211	83acd71c-c12a-4394-ba8d-05e9c5ad2cf1	3070
HMR_0232	83acd71c-c12a-4394-ba8d-05e9c5ad2cf1	14033
HMR_0258	83acd71c-c12a-4394-ba8d-05e9c5ad2cf1	14033
HMR_3837	83acd71c-c12a-4394-ba8d-05e9c5ad2cf1	7935
HMR_4028	83acd71c-c12a-4394-ba8d-05e9c5ad2cf1	15174
HMR_4618	83acd71c-c12a-4394-ba8d-05e9c5ad2cf1	8001
HMR_5465	83acd71c-c12a-4394-ba8d-05e9c5ad2cf1	10514
HMR_6616	83acd71c-c12a-4394-ba8d-05e9c5ad2cf1	15174

Saglama_data çıktısında, "HMR_0032" reaksiyonunun tümör örneklerindeki ifade düzeyi 63911 olarak belirlenmiştir. Bu değer, aynı reaksiyon için Tablo 6.10'deki sağlıklı örnekler için hesaplanan ortalama ifade düzeyi olan 136518.92'ye kıyasla anlamlı bir düşüş göstermiştir. Bu durum, reaksiyon ifade düzeyinin tümör varlığında azaldığını ve bu azalmanın belirlenen yüzde 90'lık eşik değerine göre anlamlı olduğu istatistiksel olarak kutu grafiği ile doğrulanmıştır. Şekil 6.4. Aynı şekilde diğer 10 veri kontrol edilmiş ve 5 azalan ve 5 artan veri, grafik ile uyumlu bulunmuştur. Bu, reaksiyon ifade düzeyindeki artma veya azalmanın sadece sayısal bir analizle değil, aynı zamanda görsel bir temsille de desteklendiğini ortaya koymaktadır. Bu görsel ve analitik gözlemler, veri işlemlerinin doğruluğunu sağlamakta ve altta yatan veri setlerinin bütünlüğünü onaylamaktadır. Görsel temsil ile niceliksel veri analizi arasındaki uyum, çalışmadan çıkarılan sonuçların güvenilirliğini desteklemiş ve tez için yürütülen araştırmanın metodolojik sağlamlığını doğrulamıştır.

Tablo 6.10 Kontrol amaçlı normal_mean_2 veri seti tablosu

api_id	ort
HMR_0032	136518.9204
HMR_0199	190263.9381
HMR_0211	6332.8938
HMR_0232	190263.9381
HMR_0258	190263.9381
HMR_3837	4373.1327
HMR_4028	10159.4071
HMR_4618	666.7611
HMR_5465	5509.8319
HMR_6616	10159.4071

Bu tür bir bilimsel yorumlama, elde edilen analitik sonuçların biyolojik verilerle nasıl ilişkilendirileceğini ve bu ilişkilerin nasıl doğrulanabileceğini açıklamıştır. İfade düzeylerindeki bu değişiklikler, reaksiyonun biyolojik işlevi ve tümör biyolojisi bağlamında daha fazla değerlendirilebilir. Örneğin, eğer "HMR_0032", tümör baskılayıcı bir geni temsil ediyorsa, ifade düzeylerindeki azalma, tümör gelişimindeki olası bir düzenleyici rolü açıklayabilmektedir. Ayrıca, bu tür gözlemler ilgili genin tümör gelişimi ve ilerlemesi üzerindeki potansiyel etkilerini daha derinlemesine incelemek için daha kapsamlı çalışmaların yapılmasını teşvik etmektedir. Sonuç olarak, görselleştirme ve sayısal verilerin birleştirilmesi, biyoinformatik analizlerin gücünü ve reaksiyon ifade düzeylerinin anlaşılmasındaki önemini vurgulamaktadır. Bu yaklaşım, tümör biyolojisi ve reaksiyon ifadesi araştırmalarında önemli bir araç olarak kullanılmaktadır [161, 162].

Metabolik reaksiyonların frekans analizi ve metabolitlerin belirlenmesi, insan metabolizmasının karmaşık yapısını daha iyi anlamak için önemli bir adım teşkil etmektedir [163]. Yapılan analizler, belirli metabolik reaksiyonların ve metabolitlerin, kanser hücrelerinin metabolik adaptasyonları ve tedaviye yanıt mekanizmalarında önemli roller oynayabileceğini göstermiştir [164]. Özellikle, reaksiyon değişiklikleri ve ilgili metabolitlerin Tablo 6.11’de örnekleri sunulan "HMR_change_summary_df_2" veri seti üzerinden yapılan detaylı incelemesi, metabolik ağdaki spesifik değişimleri ve bu değişimlerin potansiyel biyolojik etkilerini ortaya koymuştur.

Tablo 6.11 Reaksiyon Değişiklikleri ve İlgili Bileşiklerin Örnekleri

RXNID	increase	decrease	change	from_compounds	to_compounds	substrate	product
HMR_6616	0.9305681	0.0694319	increase	dTDP[n]	dTTP[n]	decrease	increase
HMR_4529	0.9098287	0.0901713	increase	H ₂ O[c]	Pi[c]	decrease	increase
HMR_5416	0.9657349	0.0342651	increase	H ₂ O[m]	mitothioredoxin[m]	decrease	increase
HMR_6622	0.9657349	0.0342651	increase	thioredoxin[n]	oxidized thioredoxin[n]	decrease	increase
HMR_6623	0.9756537	0.0243463	increase	deoxyuridine[c]	dGDP[c]	decrease	increase
HMR_4779	0.9711452	0.0288548	increase	NAD+[m]	H+[m]	decrease	increase

Örneğin, "HMR_6622" reaksiyonundaki artış, thioredoxin ve oksitlenmiş thioredoxin gibi bileşiklerin metabolik ağda nasıl değiştiğini ve bu değişimlerin reaksiyonun ürünleri ve substratları üzerindeki etkilerini göstermektedir. Reaksiyonun artması, thioredoxin substratının azalmasına ve sonuç olarak oksitlenmiş thioredoxin ürününün artmasına yol açar. Bu durum, reaksiyonun çalışma hızının artmasıyla substratın tüketilmesi ve ürünün oluşumu arasındaki doğrudan ilişkiyi yansıtmaktadır.

Frekans analizi sonucunda elde edilen veriler, metabolik ağda sıkça yer alan ve önemli değişikliklere uğrayan bileşikleri belirlemiştir Tablo6.12. Bu bileşikler arasında, 1-pyrroline-5-carboxylate ve 2-deoxy-D-ribose-1-phosphate gibi azalan, GDP-L-fucose ve 5,10-methylene-THF gibi artan metabolitler bulunmaktadır. Bu metabolitlerin değişim frekansları, kanser metabolizmasında kritik yolları ve potansiyel terapötik hedefleri işaret etmektedir. Frekans eşiği 3'ün üzerindeki bileşiklerin belirlenmesi süreci, yaygın olarak bulunan ve iyi anlaşılan ATP ve NADH dışındaki metabolitleri hedef almıştır. Bu yaklaşım, metabolik ağlarda daha az bilinen ve nadiren karşılaşılan, ancak potansiyel olarak önemli biyolojik rolleri ve hastalıkla ilişkileri olan metabolitleri keşfetmeyi amaçlamaktadır. Örneğin, frekansı 16 olan 2-deoxy-D-ribose-1-phosphate gibi azalan ve frekansı 4 olan GDP-L-fucose gibi artan metabolitler, kanser metabolizmasında kritik yolları ve potansiyel terapötik hedefleri işaret edebilmektedir. Bu bileşiklerin değişim frekansları, onların metabolik ağdaki önemini ve potansiyel biyolojik rollerini vurgulamaktadır [1].

Bir frekans eşiği belirlemek ve bu eşiğin üzerindeki bileşikleri seçmek, analizin spesifitesini artırarak, gerçekten önemli olan metabolik değişikliklere odaklanmayı sağlamıştır. Bu yaklaşım, geniş veri setinden anlamlı bilgiler çıkarmayı ve metabolik ağ analizlerinde gürültüyü azaltmayı mümkün kılmıştır. Ayrıca, seçilen eşik değeri ve belirlenen bileşikler, metabolik ağdaki dinamikleri daha iyi anlamak ve hastalık mekanizmaları veya tedavi yanıtları ile ilişkili potansiyel hedefleri keşfetmek için temel oluşturmuştur.

Tablo 6.12 Metabolik bileşiklerin değişim frekansları

Compound	Change Type	Frequency
1pyrroline5carboxylate	decrease	4
2deoxyDribose1phosphate	decrease	16
GDP4dehydro6deoxyDmannose	decrease	4
GDPLfucose	increase	6
510methyleneTHF	increase	4
L1pyrroline3hydroxy5carboxylate	increase	4

Analizler, metabolik ağdaki değişimlerin genel bir özetini sağlamanın yanı sıra, metabolik reaksiyonlar arasındaki ilişkileri ve bu ilişkilerin biyolojik sistemler üzerindeki etkilerini anlamada önemli bir araç olduğunu göstermiştir. Özellikle, ATP, NADH, ADP vb. gibi taşıyıcı moleküllerin analiz dışı bırakılması, daha spesifik ve biyolojik olarak önemli metabolitlerin belirlenmesine olanak tanımıştır.

Metabolik reaksiyonların ve metabolitlerin kanser metabolizması üzerindeki etkilerini detaylı bir şekilde inceleyerek, kanser araştırmalarında metabolik analizlerin potansiyelini ortaya çıkarmak için elde edilen bulgular, kanser hücrelerinin metabolik adaptasyon mekanizmalarını ve tedaviye yanıtlarını daha iyi anlamak için temel oluşturmaktadır. Ayrıca, belirlenen spesifik metabolik değişiklikler ve metabolitler, yeni terapötik hedeflerin keşfi için önemli ipuçları sunmaktadır. Örneğin, 2-deoxy-D-ribose-1-phosphate ve GDP-L-fucose'un değişim frekansları, bu metabolitlerin kanser hücrelerinin enerji metabolizması ve glikozilasyon süreçlerinde oynadıkları rolleri vurgulamaktadır. Bu tür bulgular, kanser tedavisinde metabolik yolları hedef alan yeni stratejilerin geliştirilmesine katkıda bulunarak, bu alanda daha ileri araştırmalar için sağlam bir temel sunmaktadır [165].

Çalışmamızın bu bölümünde, araştırmamızda ele alınan ve ele alınmayan metabolitlerin seçim sürecine dair kritik kararların detaylarına ve bu süreçte yapılan tercihlerin bilimsel temellerine odaklanılmıştır. Bu seçimler, kanser metabolizması ve terapötik hedeflerin belirlenmesi bağlamında, mevcut bilimsel literatür ve araştırma bulgularının dikkatli bir değerlendirilmesi sonucunda gerçekleştirilmiştir. Öncelikle neden bazı metabolitleri çalışma kapsamımıza dahil etmediğimiz açıklanmıştır.

Çalışmamıza dahil etmediğimiz bu metabolitleri filtreleyerek oluşturulan meme kanseri hastaları ile sağlıklı bireyler arasında önemli derecede farklılık gösteren

Tablo 6.13 Meme Kanserinde İfadesi Değişen Reaksiyonlar Bağlı Olarak Artan ve Azalan Önemli Metabolitler

RXNID	Gene_rule	Substrate	Product	S_change	P_change	SUBSYSTEM
HMR_3833 HMR_3835	PYCR3 or PYCR2 or PYCR1	lpyrroline5carboxylate	proline	decrease	increase	Arginine and proline metabolism
HMR_3837 HMR_8609	PYCR3 or PYCR2 or PYCR1	lpyrroline5carboxylate	proline	decrease	increase	Arginine and proline metabolism
HMR_4400 HMR_8768	GFUS	GDP4dehydro6deoxyDmannose	GDP4fucose	decrease	increase	Fructose and mannose metabolism
HMR_4778 HMR_4780	PYCR3 or PYCR2 or PYCR1	trans4hydroxyLproline	L1pyrroline3hydroxy5carboxylate	decrease	increase	Arginine and proline metabolism
HMR_4779 HMR_4781	PYCR3 or PYCR2 or PYCR1	trans4hydroxyLproline	L1pyrroline3hydroxy5carboxylate	decrease	increase	Arginine and proline metabolism
HMR_4641 HMR_6623	TK1 TK1	deoxyuridine deoxyuridine	dUMP dUMP, dGDP	decrease decrease	increase increase	Nucleotide Pyrimidine
HMR_1470	SQLE	squalene	squalene 2,3-oxide	decrease	increase	Cholesterol biosynthesis 1 (Bloch pathway)

reaksiyonlara bağı metabolitlerin detaylı bir analizini içeren veri seti Tablo 6.13’de sunulmuştur. Bu analiz, araştırmamızın bulgularının, kanser metabolizmasının anlaşılması ve potansiyel terapötik müdahalelerin geliştirilmesi açısından mevcut bilimsel literatürle olan uyumunu ve katkısını ortaya koymuştur.

Aşağıda, çalışmamızda ele alınmayan bazı önemli metabolitler ve bu seçimin bilimsel gerekçeleri detaylandırılmıştır:

ATP, GTP, NAD, ve NADH vb gibi yaygın bileşikler, hücresel enerji dşuaöngüsü ve redoks reaksiyonları gibi temel metabolik süreçlerde merkezi roller oynamaktadırlar. Kanser metabolizmasında özgül bir değişiklik göstermeleri beklenmez, çünkü bunlar tüm hücre tiplerinde geniş bir yelpazede işlev görmektedirler [153, 166].

Deoksi-D-riboz-1-fosfat DNA onarımı ve sentezi ile ilişkilidir. Ancak, kanser hücrelerindeki metabolik değişikliklerin anlaşılmasında özgül bir gösterge olarak kullanılmasının önünde, bu metabolitin hücreler arası değişkenlik göstermesi ve kanser türlerine göre farklılık gösteren mekanizmaları nedeniyle zorluklar bulunmaktadır [167].

Metilen Tetra Hidrofolat DNA sentezi ve metilasyon reaksiyonlarında rol oynar. Kanser metabolizmasında önemli olmasına rağmen, bu metabolitin düzeylerindeki değişiklikler çok çeşitli faktörlere bağı olduğundan, tezimizin odak noktası olan spesifik metabolik yollarla doğrudan ilişkili olmamaktadır [168].

PEP (Fosfoenolpirüvat) Glikoliz ve glukoneogenezde kritik bir ara üründür. Kanser metabolizmasında genel bir değişiklik göstergesi olabilir, ancak bu çalışmada ele alınan spesifik metabolik değişikliklerle doğrudan bir ilişkisi bulunmamaktadır [169].

Timin, Adenin, Guanin, Hipoksantin nükleotid metabolizmasının yaygın temel bileşenleridir. Bu bazların metabolizması yeterince çalışılmıştır ve kanser hücrelerinde değişiklik gösterebilir; ancak, bu çalışmada odaklanılan metabolik yolların ve mekanizmaların özgül anlayışı açısından doğrudan ve spesifik bir bağlantı kurulmamıştır [170, 171].

Deoksinükleozidler (Deoksiadenozin, Deoksiguanozin, Deoksiinozin): DNA replikasyonu ve onarımı ile ilişkilidirler. Bu metabolitlerin düzeylerindeki değişiklikler, kanser hücrelerinde önemli olabilir; ancak, bu değişiklikler genellikle daha geniş DNA hasarı ve onarım mekanizmalarıyla ilişkilidir ve bu çalışmanın odak noktası olan spesifik metabolik yollarla doğrudan ilişkili bulunmamıştır [172,

173] .

Dihidrofolat ve Dolikol Fosfat Türevleri sırasıyla, folat metabolizması ve protein glikozilasyonunda rol oynamaktadırlar. bu çalışmada ele alınan spesifik metabolik değişikliklerle ve terapötik hedeflerle doğrudan bir ilişkileri bulunamamıştır [174, 175].

Tiyoredoksin ve Okside Tiyoredoksin: Hücrel redoks durumunun düzenlenmesinde önemli roller oynarlar. Kanseri hücrelerinde antioksidan savunma mekanizmalarının bir parçası olarak tüm hücrelerde çok yaygın olarak bulunmaktadır. Ancak, bu çalışmada, kanser ilerlemesi ve metabolik yeniden programlama arasındaki etkileşimi daha spesifik olarak ortaya koyan ve daha az yaygın metabolik yollar ve reaksiyonlar önceliklendirilmiştir [176, 177].

Şilomikron ve şilomikron havuzu, lipid metabolizmasında önemli bir rol oynamaktadırlar. Kanseri hücrelerinin lipid metabolizmasındaki değişiklikler üzerine çalışmalar bulunmakla birlikte, bu çalışmada lipid metabolizmasının bu genel metabolitleri yerine, daha spesifik metabolik değişiklikler önceliklendirilmiştir [178, 179].

Bu metabolitlerin her biri, kanser biyolojisi ve metabolizmasında önemli roller oynasa da, tezimizde ele alınan metabolik yollar ve mekanizmalarla doğrudan ilişkili olmayan geniş bir yelpazede işlev gördükleri için seçilmemiştir. Bu seçim, kanser metabolizmasının özgül yönlerine odaklanarak, bu alanda yeni terapötik yaklaşımlar geliştirme potansiyelini maksimize etme amacını taşımaktadır. Bu yaklaşım, mevcut bilimsel literatür ve araştırma bulgularının dikkatli bir değerlendirilmesiyle desteklenmektedir.

Tablo 6.13’de yöntemler kısmında kullanılan veri tabanlarından elde edilen verilerin işlenmesi ve filtrelemeler sonucunda, meme kanseri hastaları ile sağlıklı bireyler arasında önemli derecede farklılık gösteren reaksiyonlar, gen kuralları, ait oldukları alt sistemi ve bu reaksiyonlara bağlı olarak değişen substratlar ile ürünler sunulmuştur. Elde edilen sonuçlar, literatürle yapılan karşılaştırmalar neticesinde uyumlu olduğu tespit edilmiş, bu da uygulanan analiz yöntemlerinin doğruluğunu ve tutarlılığını teyit etmiştir. Aşağıda, bu literatür karşılaştırmalarına dayanarak elde edilen bulgular detaylı bir şekilde incelenmiştir.

6.6 Prolin Metabolizması Değişikliği

1-pirolin-5-karboksilattan prolin'e dönüşüm, arginin ve prolin metabolizma yolunun kritik bir adımıdır. Bu reaksiyon, protein sentezi, osmoregülasyon ve antioksidan savunma gibi önemli işlevlerde rol oynayan prolin amino asidinin sentezi için esastır. "HMR_3833", "HMR_3835", "HMR_3837", "HMR_8609" reaksiyonları 1-pirolin-5-karboksilatın proline indirgenmesini sağlamaktadır; bu işlem, substratın (1-pirolin-5-karboksilat) azalması ve ürünün (prolin) artışı ile sonuçlanmaktadır. Bu reaksiyon, PYCR1, PYCR2 ve PYCR3 enzimleri tarafından katalize edilmekte ve reaksiyon, EC 1.5.1.2 enzim komisyon numarası ile sınıflandırılmış olup, arginin ve prolin metabolizması alt sistemi içinde yer almaktadır.

Kanser bağlamında, prolin dahil olmak üzere amino asitlerin metabolizmasında gözlemlenen değişiklikler bulunmaktadır. Kanser hücreleri, protein sentezi ve diğer biyosentetik süreçler için amino asitlere olan artan talebi desteklemek üzere metabolik yolların yeniden programlanmasını sıklıkla sergilemektedir. Verilerimiz tarafından belirtilen kanserli durumlarda prolin seviyelerindeki artış, birkaç faktöre bağlanabilmektedir:

- Kanser hücreleri, artan protein sentezini desteklemek ve hipoksik koşullar veya oksidatif stres altında metabolik avantajlar sağlamak için reaksiyonların ifadesini artırmaktadır.
- Prolin, bir osmolit ve antioksidan olarak hareket etmektedir. Bu amino asidin kanser hücrelerinde birikimi, bu hücrelerin sıklıkla karşılaştığı osmotik ve oksidatif stresle başa çıkmada yardımcı olmaktadır.
- Prolin metabolizması, hücre proliferasyonu ve apoptozu düzenleyen sinyal yollarına bağlanmıştır. Prolin seviyelerindeki değişiklikler, kanser hücrelerinin hayatta kalması ve büyümesine katkıda bulunabilecek bu yolları etkileyebilmektedir.

PYCR1'in çeşitli kanser türlerinde aşırı ifade edildiğini ve inhibisyonunun tümör büyümesini baskılayabileceği gösterilmiştir, bu da prolin sentezinin kanser hücre metabolizmasında önemli bir rol oynadığını göstermektedir. Başka bir araştırma, prolin metabolizmasının kanser hücrelerinde hızlı büyümeyi desteklemek ve apoptoza karşı korumak üzere yeniden programlandığını göstermiştir, bu da kanserli dokularda gözlemlenen prolin seviyelerindeki artışın önemini daha da vurgulamaktadır.

Kanser hücrelerinde PYCR enzimlerinin aktivitesi yoluyla prolin üretimindeki artış, kanser hücrelerinin metabolik yeniden programlanması üzerine mevcut literatürle uyumlu bulunmuştur. Bu yeniden programlama, hızlı hücre büyümesi, stres yanıtı ve hayatta kalmayı desteklemek üzere amino asit metabolizmasındaki değişiklikleri içerir. PYCR1, PYCR2 ve PYCR3 enzimleri bu süreçte kilit bir rol oynar, bu da onları kanser tedavisi için potansiyel hedefler haline getirmektedir [180–183].

6.7 GDP-L-fukoz Metabolizması Değişikliği

GDP-4-dehidro-6-deoksi-D-mannoz'un GDP-L-fukoz'a dönüşümü, "HMR_4400" ve "HMR_8768" reaksiyonları ile EC 1.1.1.271 numarası ile sınıflandırılan GFUS enzimi tarafından katalize edilmektedir. Bu reaksiyon sitozolda gerçekleşmekte ve fruktoz ve mannoz metabolizma yolunun bir parçası olmaktadır. Süreç, substratta (GDP-4-dehidro-6-deoksi-D-mannoz) bir azalma ve üründe (GDP-L-fukoz) bir artışı içermektedir, bu da aktif bir fukozilasyon sürecini göstermektedir.

Fukozilasyon, proteinlere ve lipidlere fukoz (bir heksoz deoksi şeker) eklenmesini içeren bir glikozilasyon türüdür. Bu modifikasyon, hücre sinyalleme, bağışıklık yanıtı ve kanser gelişimi dahil olmak üzere çeşitli biyolojik süreçlerde kritik bir rol oynamaktadır. Fukozilasyon örüntülerindeki(hücre yüzeyine bağlanma şekilleri) değişiklikler kanser ilerlemesi, metastaz ve bağışıklıktan kaçınma ile ilişkilendirilmiştir.

Kanserde Artmış Fukozilasyon hücreler arası etkileşimleri, adezyonu ve migrasyonu etkilemektedir. Artmış fukozilasyon, birkaç kanser türünde artmış metastatik potansiyel ve kötü prognoz ile bağlantılı bulunmuştur. Kanser hücrelerindeki değişmiş fukozilasyon, bağışıklık yanıtını düzenlemekte, bağışıklıktan kaçınmaya katkıda bulunmaktadır. Örneğin, yüzey antijenlerinin fukozilasyonundaki değişiklikler, bağışıklık hücreleri tarafından tanınmayı bozmaktadır.

Kanser ilerlemesi ve bağışıklık düzenlemesindeki rolü nedeniyle, fukozilasyon örüntüleri ve fukoz metabolizmasında yer alan enzimler, örneğin GFUS, kanser teşhisi ve prognozu için potansiyel biyomarkerlar olarak kabul edilmektedir.

Yapılan bir araştırma kanserde değişmiş glikozilasyonun, fukozilasyon dahil, önemini vurgulamaktadır. Çalışma, kanser hücrelerinin tümör büyümesini, metastazı ve bağışıklıktan kaçınmayı teşvik etmek için bu modifikasyonları nasıl kullandığını tartışmıştır. Başka bir çalışma , kanserde GDP-fukozun rolüne odaklanmıştır, fukozilasyon süreçlerindeki sentezi ve kullanımının kanser

hücre biyolojisi için kritik olduğunu ve terapötik müdahaleler için nasıl hedef alınabileceğini açıklamıştır [184, 185].

Veri setimizde gözlemlenen GDP-L-fukoz üretimindeki artış, GFUS enzimi tarafından kolaylaştırıldığı için, kanser metabolizmasında fukozilasyonun kritik rolünü yansıtmaktadır. Kanser hücrelerindeki değişmiş fukozilasyon örüntüleri, tümör ilerlemesine, metastaza ve bağışıklıktan kaçınmaya katkıda bulunmaktadır, bu da fukozilasyon yolunu ve özellikle GFUS enzimini kanser terapisi için potansiyel hedefler haline getirmektedir. Fukozilasyon örüntülerindeki değişiklik, ayrıca fukozilasyonun kanser teşhisi ve prognozu için bir biyomarker olarak potansiyelini de vurgulamıştır [186, 187].

GDP-4-dehidro-6-deoksi-D-mannoz'dan GDP-L-fukoz'a dönüşümün bu detaylı analizi, özellikle glikozilasyon değişiklikleri bağlamında, kanser hücrelerinde meydana gelen karmaşık metabolik yeniden programlamaya içgörü sağlamıştır.

6.8 Trans-4-Hidroksi-L-Prolin Metabolizması Değişikliği

Bu reaksiyon, trans-4-hidroksi-L-prolinin L-1-pirolin-3-hidroksi-5-karboksilata dönüşümünü içermektedir ve PYCR1, PYCR2 ve PYCR3 enzimleri tarafından katalize edilmektedir. "HMR_4778", "HMR_4780", "HMR_4779" ve "HMR_4781" reaksiyonları ile hem sitozolik hem de mitokondriyal bölmelerde gerçekleşmektedir. EC 1.5.1.2 altında sınıflandırılan bu reaksiyonlar, arginin ve prolin metabolizma yolu olmaktadır ve substratta (trans-4-hidroksi-L-prolin) bir azalma ile üründe (L-1-pirolin-3-hidroksi-5-karboksilat) bir artışı içermektedir.

Prolin ve türevlerinin, hidroksprolin dahil, metabolizması, kolajen sentezi, hücre sinyalleme ve hipoksik koşullara yanıt gibi etkileri nedeniyle kanserde kritik önemi olmaktadır. Hidroksprolin, kolajenin önemli bir bileşenidir ve metabolizması, tümör ilerlemesi ve metastazda önemli bir rol oynayan ekstrasellüler matriks (ECM) yeniden yapılanmasının düzenlenmesiyle yakından ilişkili bulunmaktadır.

Hidroksprolinin metabolizması, doğrudan kolajen devir hızıyla ilişkili bulunmuştur. Kanser hücreleri sıklıkla, invazyonu ve metastazı kolaylaştırmak için çevrelerindeki ECM'yi değiştirmektedir. Hidroksprolin metabolizmasında yer alan enzimler, PYCR1, PYCR2 ve PYCR3, artan kolajen sentezi ve yeniden düzenlemesini desteklemek için yukarı düzenlenebilmektedir.

Hidroksprolin metabolizması aynı zamanda, tümör mikroçevresinde yaygın

bir durum olan hipoksiye hücreyel yanıtla da bağlantılıdır. Hidroksiprolini pirolin-3-hidroksi-5-karboksilata dönüştüren enzimler, düşük oksijen koşulları altında tümörün hayatta kalması ve büyümesine katkıda bulunan hipoksik yanıtta roller oynamaktadır.

Hidroksiprolin metabolizmasının metabolitleri, hücre büyümesi, farklılaşma ve apoptoz ile ilgili yolları etkileyen sinyal molekülleri olarak işlev görmektedir. Bu metabolik yollardaki değişiklikler, bu nedenle kanser hücre davranışını etkilemektedir.

Yapılan çalışmalar, hipoksiye duyarlı faktörlerin (HIF'ler) prolin ve hidroksiprolin metabolizmasında yer alan enzimlerin ifadesini düzenlediğini ve bu metabolik yolları tümörlerdeki hipoksik yanıtla bağlantılı hale getirdiğini göstermiştir. Başka bir araştırma ise kanserde kolajen ve hidroksilasyonun rolünü tartışmış ve prolin ve hidroksiprolin metabolizması aracılığıyla ECM'deki değişikliklerin kanser ilerlemesi ve metastaza nasıl katkıda bulunduğunu vurgulamıştır [188, 189].

Kanserde trans-4-hidroksi-L-prolinin L-1-pirolin-3-hidroksi-5-karboksilata dönüşümü, amino asit metabolizması, ECM yeniden düzenlenmesi ve hipoksik yanıt arasındaki karmaşık etkileşimi yansıtmaktadır. Bu dönüşümü kolaylaştıran PYCR1, PYCR2 ve PYCR3 enzimleri, artan kolajen devir hızı ve hipoksik koşullar altında hayatta kalma dahil olmak üzere kanser hücrelerinin adaptif mekanizmalarını desteklemektedir. Bu, kanser terapisinde hidroksiprolin metabolizmasını hedef almanın potansiyelini ve bu metabolik yolların tümör biyolojisindeki önemini vurgulamaktadır [183].

6.9 Deoksiuridin Metabolizması Değişikliği

"HMR_4641", "HMR_6623" reaksiyonları aracılığı ile, deoksiuridin'in deoksiuridin monofosfat'a (dUMP) ve bir diğer durumda deoksiganosin difosfat'a (dGDP) dönüşümünü içermektedir ve timidin kinaz 1 (TK1) enzimi tarafından katalize edilmektedir. EC 2.7.1.21 altında sınıflandırılmıştır ve sitozolda gerçekleşmektedir. Bu reaksiyon, özellikle pirimidin metabolizması bağlamında nükleotid metabolizma yolunun bir parçasıdır. Süreç, substratta (deoksiuridin) bir azalma ve üründe (dUMP ve bir durumda dGDP) bir artışı içermektedir, bu da aktif bir nükleotid kurtarma yolunu göstermektedir.

TK1'in DNA sentezindeki rolü ve kanser terapisinde hedef olarak potansiyeli tartışılmıştır. Enzimin nükleotid metabolizmasının kurtarma yolundaki önemini ve kanser tedavisi için olan etkilerini vurgulanmıştır. Diğer bir araştırma, kanser

hücrelerinde TK1 aktivitesinin düzenlenmesini incelemekte ve hücre çoğalması için bir biyomarker olarak önemini ve TK1 hedefli terapilerin geliştirilmesindeki potansiyelini vurgulamaktadır [190, 191].

Nükleotid metabolizması, özellikle kurtarma yolları, hızla çoğalan kanser hücrelerinde nükleotidlere olan yüksek talep nedeniyle kanserde kritik bir rol oynamaktadır. TK1 enzimi, birkaç nedenden dolayı kanser biyolojisinde özellikle ilgi çekici olmaktadır:

1. **Çoğalma İşareti:** TK1 aktivitesi, çoğalan hücrelerin artan DNA sentezi ve tamir ihtiyaçlarını yansıtarak kanser hücrelerinde sıklıkla yükselmektedir. TK1 seviyeleri, hücre çoğalması için sıkça kullanılan bir biyomarker olarak kullanılır ve tümörün agresifliği ve prognozu ile ilişkilendirilmiştir.
2. **Kemoterapi Hedefi:** Deoksiuridin'in dUMP'ye dönüşümünü içeren pirimidin nükleotidlerinin kurtarma yolu, kemoterapötik ajanlar için bir hedef olmuştur. TK1'i inhibe eden ilaçlar, kanser hücrelerinde DNA sentezini bozarak hücre döngüsü duraklamasına ve apoptoza yol açmaktadır.
3. **Metabolik Yeniden Programlama:** Kanser hücreleri sıklıkla, hızlı DNA ve RNA sentezi desteklemek için nükleotid metabolizması yollarının yeniden programlanmasını sergilemektedirler. Veri setimizde gözlemlenen dUMP (ve dGDP) üretimindeki artış, DNA replikasyonu ve tamiri için artırılmış nükleotid sentezini sağlayan bu tür metabolik adaptasyonları yansıtmaktadır.

Kanser hücrelerinde, TK1 tarafından kolaylaştırılan deoksiuridin'den dUMP'ye (ve dGDP'ye) dönüşüm, kanser hücrelerinin hızlı çoğalmasını destekleyen nükleotid kurtarma yollarının kritik rolünü vurgulamaktadır. Kanserde TK1 aktivitesinin yükselmesi, çoğalma için bir biyomarker ve kemoterapi için bir hedef olarak potansiyelini göstermiştir. Nükleotid metabolizma yolundaki bu reaksiyonun analizi, nükleotid sentezi taleplerini karşılamak için kanser hücrelerinin uyarlanabilir stratejilerini ortaya koymuş, bu metabolik adaptasyonları hedef alan potansiyel terapötik müdahalelere yeni perspektifler sunmuştur.

Nükleotid sentez yollarını hedef alan terapiler, kanser hücrelerinin büyüme ve çoğalma kapasitesini sınırlayarak, kanserin ilerlemesini durdurma veya yavaşlatma potansiyeline sahiptir. Bu bağlamda, nükleotid metabolizmasının hedef alınması, kanser tedavisindeki yenilikçi yaklaşımlar arasında önemli bir yer tutmaktadır [192–195].

6.10 Skualen Metabolizması Değişikliği

Bu reaksiyon, skualenin, skualen 2,3-oksit'e dönüşümünü içerir ve skualen epoksidaz (SQLE) enzimi tarafından katalize edilmektedir. EC 1.14.13.132 altında sınıflandırılmıştır ve "HMR_1470" reaksiyonuyla sitozolda gerçekleşmektedir. Bu reaksiyon, özellikle Bloch yolu bağlamında, kolesterol biyosentez yolunda kritik bir adımdır. Süreç, substratta (skualen) azalma ve üründe (skualen 2,3-oksit) artışı içermektedir, bu da aktif bir kolesterol sentez yolunu göstermiştir.

Kolesterol metabolizması, kanserde önemli bir rol oynamaktadır, çünkü kolesterol ve ara ürünleri hücre zarı bütünlüğü, steroid hormon sentezi ve hücre proliferasyonu ve hayatta kalımı teşvik eden sinyal yolları için hayati öneme sahiptir. Kolesterol biyosentezinin kritik bir adımını katalize eden SQLE enzimi, birkaç nedenle kanser biyolojisinde özellikle ilgi çekici olmuştur:

- **Hücre Proliferasyonu ve Zar Oluşumu:** Hızla çoğalan kanser hücreleri, zar biyosentezi için artan kolesterol gerektirmektedir. Skualenden skualen 2,3-oksit'e dönüşümü de içeren kolesterol biyosentez yollarının yukarı düzenlenmesi, kanser hücrelerinin büyüme taleplerini desteklemektedir.
- **Sinyal Yolları:** Kolesterol ve ara ürünleri, hücre proliferasyonu, apoptoz ve migrasyonu düzenleyen birkaç sinyal yolunda yer almaktadır. Bu nedenle, kolesterol metabolizmasındaki değişiklikler kanser ilerlemesi ve metastazı etkilemektedir.
- **Terapötik Hedef:** Kolesterol biyosentezindeki rolü göz önüne alındığında, SQLE potansiyel bir terapötik hedef olarak görülmektedir. SQLE inhibitörleri, kolesterol sentezini bozarak kanser hücrelerinin canlılığını ve çoğalmasını etkilemektedir.

SQLE'nin belirli kanser türlerinde aşırı ifade edildiği ve inhibisyonunun tümör büyümesini ve çoğalmasını azaltabileceği gösterilmiştir, bu da enzimin kanser terapisinde bir hedef olarak potansiyelini vurgulamaktadır. Kanser hücrelerinde, SQLE tarafından kolaylaştırılan skualenden skualen 2,3-oksit'e dönüşüm, kanser hücrelerinin hızlı büyümesini ve çoğalmasını desteklemede kolesterol biyosentezinin önemini vurgulamaktadır. Bu yolun kanserde yukarı düzenlenmesi, SQLE'yi kanser ilerlemesi için bir biyomarker ve terapötik müdahale için bir hedef olarak potansiyelini altını çizmiştir. Kolesterol biyosentez yolu içindeki bu analiz, kanser hücrelerinin artan kolesterol taleplerini karşılamak için meydana gelen karmaşık metabolik yeniden programlamayı ortaya

koymuřtur, kanser terapisinde kolesterol metabolizmasını hedef almanın potansiyel stratejilerine dair içgörüler sunmuř ve kolesterol biyosentez yollarını hedef almanın terapötik potansiyelini kapsamlı bir řekilde anlamamızı saęlamıřtır [196–200]

Bu kapsamlı analizde, TCGA ve HMA’dan elde edilen verilere dayanarak, kanser metabolizmasındaki çeřitli anahtar metabolik reaksiyonlar ve bunların etkileri incelenmiřtir. Bu reaksiyonlar, arginin ve prolin metabolizması, fruktoz ve mannoz metabolizması, nükleotid metabolizması ve kolesterol biyosentezi dahil olmak üzere çeřitli metabolik yolları kapsamaktadır.

Ana Bulguların Özeti:

- 1. Arginin ve Prolin Metabolizması:** 1-pirolin-5-karboksilattan proline ve trans-4-hidroksi-L-prolinden L-1-pirolin-3-hidroksi-5-karboksilata dönüşüm, PYCR1, PYCR2 ve PYCR3 tarafından katalize edilmekte ve kanserde prolin metabolizmasının önemini vurgulamaktadır. Bu reaksiyonlar, hızlı hücre büyümesi, stres yanıtı ve hipoksik koşullar altında hayatta kalma desteęi saęlamaktadır.
- 2. Fruktoz ve MannoZ Metabolizması:**GFUS tarafından GDP-4-dehidro-6-deoksi-D-mannoZdan GDP-L-fukoz’a dönüşüm, kanser ilerlemesi, metastaz ve baęıřıklıktan kaçınmada fukozilasyonun rolünü vurgulamaktadır.
- 3. Nükleotid Metabolizması:** TK1 tarafından deoksiuridinden dUMP’ye ve dGDP’ye dönüşüm, kanser hücrelerinin çoęalmasını destekleyen nükleotid kurtarma yollarının kritik rolünü yansıtmaktadır. TK1, potansiyel bir biyomarker ve terapötik hedef olarak hizmet etmektedir.
- 4. Kolesterol Biyosentezi:** SQLE tarafından skualenden skualen 2,3-oksit’e dönüşüm, kanser hücrelerinin çoęalması ve zar oluşumu açısından kolesterol metabolizmasının önemini vurgulamaktadır. SQLE, kanser hücrelerinin canlılığını bozma potansiyeline sahip potansiyel bir hedef olmaktadır.

İfadesi deęiřen reaksiyonlarla kanıtlanan kanser hücrelerindeki metabolik deęiřiklikler, metabolik yeniden programlama ve kanser ilerlemesi arasındaki karmařık etkileřimi vurgulamıřtır. Bu deęiřiklikler, sadece kanser hücrelerinin hızlı büyümesini ve hayatta kalmasını desteklemekle kalmayıp, aynı zamanda terapötik müdahale için potansiyel hedefler sunmuřtur. Bu reaksiyonları katalize eden enzimler—PYCR1, PYCR2, PYCR3, GFUS, TK1 ve SQLE—kansere

metabolizmasında kritik oyuncular olarak ortaya çıkmaktadır ve hedefe yönelik kanser terapilerinin geliştirilmesi için fırsatlar sunmaktadır.

Bu analiz, kanser metabolizmasının karmaşıklığını anlamanın önemini vurgulamış ve yeni terapötik stratejilerin belirlenmesine yönelik gelecekteki araştırmalar için bir temel sağlamıştır. Kanser hücrelerinin büyüme ve hayatta kalma için güvendiği metabolik yolları hedef alarak, kanser hastalarının prognozunu potansiyel olarak iyileştirebilecek daha etkili tedaviler geliştirebileceği gösterilmiştir.

Bu analizden elde edilen bulgular, mevcut literatürle uyumlu olup, metabolik yeniden programlamanın kanserin bir özelliği olduğunu pekiştirmiştir. Gelecekteki çalışmalar, kanserde bu metabolik yolların rollerini daha da aydınlatmaya ve bu yolları hedef almanın terapötik potansiyelini keşfetmeye odaklanmak olacaktır. Kanser metabolizmasını daha derinlemesine anlamamız derinleştikçe, bu karmaşık hastalıkla mücadele yeteneğimiz de artmaktadır.

Metabolik reaksiyonların ifade düzeylerindeki değişikliklerin incelenmesi, genetik düzeydeki diferansiyel ifade analizlerine benzer bir metodolojiyle yürütülmüştür. Ancak, bu yaklaşımın yeniliği, metabolik ağ düzeyinde diferansiyel analizlerin gerçekleştirilmesidir. Bu yöntem, literatürde daha az yaygın olarak karşılaşılan, metabolik düzeydeki değişiklikleri gen düzeyindeki değişikliklerle ilişkilendiren yenilikçi bir yaklaşım sunmaktadır. Bu metodoloji, metabolik ağ analizlerini geniş bir perspektiften ele alarak, kanser gibi karmaşık hastalıkların daha iyi anlaşılmasına katkıda bulunmayı hedeflemiştir. Elde edilen bulgular, kanser metabolizmasının detaylı haritalanması için temel oluşturmakta ve metabolik terapi stratejilerinin geliştirilmesine önemli katkılar sağlamaktadır. Bu süreç, gen ve metabolik ağ analizlerinin entegrasyonu konusunda öncü bir çalışma olarak değerlendirilmiştir.

Bu çalışma, kanser hücrelerinde gözlemlenen metabolik değişikliklerin geniş çaplı bir değerlendirmesini sağlamıştır; ancak, metodolojik sınırlılıklar ve potansiyel çıkarımlar göz önünde bulundurulduğunda, bazı kısıtlamalar bulunmaktadır. İlk olarak, bu analiz, meme kanserine odaklanmış olup, kullanılan biyoinformatik araçlar ve veri setleri, çeşitli kanser türleri ve doku tipleri için uygun olmayabilir. İkincil olarak, metabolik reaksiyonların ifade düzeylerindeki değişikliklerin analizi, kullanılan örneklerin kalitesine ve miktarına bağlı olarak değişkenlik gösterebilmektedir. Bu da sonuçların tekrarlanabilirliği ve güvenilirliği konusunda belirsizlik oluşturmaktadır. Ayrıca metabolik yolların işlevselliği üzerine yapılan çıkarımlar, deneysel olarak doğrulanmamış olup, bu yüzden bu yolların kanser biyolojisindeki gerçek rolü belirsiz olmaktadır. Son olarak bu analiz, çevresel ve

epigenetik faktörlerin metabolik profil üzerindeki etkilerini göz ardı etmektedir, bu da metabolik değişikliklerin nedenlerini tam olarak açıklamada eksikliklere yol açmaktadır.

Gelecekte yapılacak çalışmaların, çeşitli kanser türlerini kapsayacak şekilde genişletilmesi ve metabolik profillerin daha kapsamlı analizlerinin yapılması gerekmektedir; böylece mevcut sınırlılıkların üstesinden gelinmesi hedeflenmektedir. Bu, metabolik yolların kanser gelişimi ve ilerlemesinde oynadığı rollerin daha ayrıntılı olarak ortaya konulmasını sağlayabilir ve bu yolları hedef alan tedavilerin etkinliğinin artırılmasına katkıda bulunabilir. Ayrıca, metabolik reaksiyonlar üzerinde yapılan çıkarımların deneysel olarak doğrulanması, bu reaksiyonların kanser biyolojisindeki rollerinin daha iyi anlaşılmasını sağlayacaktır. Daha büyük ve çeşitli popülasyonlardan alınacak örneklerle yapılacak çalışmaların, bulguların genelleştirilebilirliğini ve uygulanabilirliğini artırması beklenmektedir. Çevresel ve epigenetik faktörlerin metabolik değişiklikler üzerindeki etkilerinin entegrasyonunun, hastalığın daha kapsamlı bir modelinin oluşturulmasına imkan tanıyacağı ve kişiselleştirilmiş tedavi yaklaşımlarının daha etkili olmasını sağlayacağı öngörülmektedir. Bu tür bir bütüncül yaklaşım, kanser tedavisinde yeni ve daha etkili stratejilerin keşfine önayak olabilir.

Son olarak, bu çalışmanın sonuçları, kanser tedavisinde kişiselleştirilmiş yaklaşımların geliştirilmesine yönelik gelecekteki araştırmalara yön vermektedir. Metabolik ve genetik profillemenin kombinasyonu, hastalığın özgül mekanizmalarını hedef alan daha etkili ve yan etkisi daha az olan terapilerin tasarlanmasına olanak tanımaktadır. Halihazırda var olan genomik profilleme ile kişiselleştirilmiş tedaviye ek olarak Metabolomik profilleme ile kişiselleştirilmiş tedaviler de eklenmektedir.

- [1] A. Batushansky, D. Toubiana, A. Fait, “Correlation-based network generation, visualization, and analysis as a powerful tool in biological studies: A case study in cancer cell metabolism,” *BioMed Research International*, vol. 2016, pp. 1–9, 2016. DOI: 10.1155/2016/8313272.
- [2] H. Sung *et al.*, “Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries,” *CA: A Cancer Journal for Clinicians*, vol. 71, no. 3, pp. 209–249, Feb. 2021. DOI: 10.3322/caac.21660.
- [3] J. Makoukji *et al.*, “Gene expression profiling of breast cancer in lebanese women,” *Scientific Reports*, vol. 6, no. 1, Nov. 2016. DOI: 10.1038/srep36639.
- [4] T. M. Tilli, C. d. S. Castro, J. A. Tuszynski, N. Carels, “A strategy to identify housekeeping genes suitable for analysis in breast cancer diseases,” *BMC Genomics*, vol. 17, no. 1, Aug. 2016. DOI: 10.1186/s12864-016-2946-1.
- [5] V. K. Sarhadi, G. Armengol, “Molecular biomarkers in cancer,” *Biomolecules*, vol. 12, no. 8, p. 1021, Jul. 2022. DOI: 10.3390/biom12081021.
- [6] D. Hanahan, R. A. Weinberg, “Hallmarks of cancer: The next generation,” *Cell*, vol. 144, no. 5, pp. 646–674, Mar. 2011. DOI: 10.1016/j.cell.2011.02.013.
- [7] P. Anand, C. Sundaram, S. Jhurani, A. B. Kunnumakkara, B. B. Aggarwal, “Curcumin and cancer: An “old-age” disease with an “age-old” solution,” *Cancer Letters*, vol. 267, no. 1, pp. 133–164, Aug. 2008. DOI: 10.1016/j.canlet.2008.03.025.
- [8] E. D. Pleasance *et al.*, “A small-cell lung cancer genome with complex signatures of tobacco exposure,” *Nature*, vol. 463, no. 7278, pp. 184–190, Dec. 2009. DOI: 10.1038/nature08629.
- [9] J. Lipsick, “A history of cancer research: Tumor suppressor genes,” *Cold Spring Harbor Perspectives in Biology*, vol. 12, no. 2, a035907, Feb. 2020. DOI: 10.1101/cshperspect.a035907.
- [10] J. Ferlay *et al.*, *Global cancer observatory: Cancer today*, International Agency for Research on Cancer, Available from: <https://gco.iarc.fr/today>, accessed 13 January 2024, 2021.
- [11] E. A. Rakha, G. M. Tse, C. M. Quinn, “An update on the pathological classification of breast cancer,” *Histopathology*, vol. 82, no. 1, pp. 5–16, Dec. 2022. DOI: 10.1111/his.14786.

- [12] T. R. Kasperek, T. C. Humphrey, “DNA double-strand break repair pathways, chromosomal rearrangements and cancer,” *Seminars in Cell & Developmental Biology*, vol. 22, no. 8, pp. 886–897, Oct. 2011. DOI: 10.1016/j.semcd.2011.10.007.
- [13] R. K. Thomas *et al.*, “High-throughput oncogene mutation profiling in human cancer,” *Nature Genetics*, vol. 39, no. 3, pp. 347–351, Feb. 2007. DOI: 10.1038/ng1975.
- [14] R. Nirmaladevi, “Epigenetic alterations in cancer,” *Frontiers in Bioscience*, vol. 25, no. 6, pp. 1058–1109, 2020. DOI: 10.2741/4847.
- [15] L. Yang *et al.*, “Targeting cancer stem cell pathways for cancer therapy,” *Signal Transduction and Targeted Therapy*, vol. 5, no. 1, Feb. 2020. DOI: 10.1038/s41392-020-0110-5.
- [16] J. Massagué, K. Ganesh, “Metastasis-initiating cells and ecosystems,” *Cancer Discovery*, vol. 11, no. 4, pp. 971–994, Apr. 2021. DOI: 10.1158/2159-8290.cd-21-0010.
- [17] J. Majidpoor, K. Mortezaee, “Steps in metastasis: An updated review,” *Medical Oncology*, vol. 38, no. 1, Jan. 2021. DOI: 10.1007/s12032-020-01447-w.
- [18] A. Castelló *et al.*, “Lower breast cancer risk among women following the world cancer research fund and american institute for cancer research lifestyle recommendations: Epigeicam case-control study,” *PLOS ONE*, vol. 10, no. 5, A. Sapino, Ed., e0126096, May 2015. DOI: 10.1371/journal.pone.0126096.
- [19] C. Tomasetti, B. Vogelstein, “Variation in cancer risk among tissues can be explained by the number of stem cell divisions,” *Science*, vol. 347, no. 6217, pp. 78–81, Jan. 2015. DOI: 10.1126/science.1260825.
- [20] J. N. Weinstein *et al.*, “The cancer genome atlas pan-cancer analysis project,” *Nature Genetics*, vol. 45, no. 10, pp. 1113–1120, Sep. 2013. DOI: 10.1038/ng.2764.
- [21] S. Goodwin, J. D. McPherson, W. R. McCombie, “Coming of age: Ten years of next-generation sequencing technologies,” *Nature Reviews Genetics*, vol. 17, no. 6, pp. 333–351, May 2016. DOI: 10.1038/nrg.2016.49.
- [22] S. Roychowdhury, A. M. Chinnaiyan, “Translating cancer genomes and transcriptomes for precision oncology,” *CA: A Cancer Journal for Clinicians*, vol. 66, no. 1, pp. 75–88, Nov. 2015. DOI: 10.3322/caac.21329.
- [23] N. J. O’Neil, M. L. Bailey, P. Hieter, “Synthetic lethality and cancer,” *Nature Reviews Genetics*, vol. 18, no. 10, pp. 613–623, Jun. 2017. DOI: 10.1038/nrg.2017.47.
- [24] R. Louhimo, S. Hautaniemi, “CNAmets: An R package for integrating copy number, methylation and expression data,” *Bioinformatics*, vol. 27, no. 6, pp. 887–888, Jan. 2011. DOI: 10.1093/bioinformatics/btr019.

- [25] M. D. Jansson, A. H. Lund, “MicroRNA and cancer,” *Molecular Oncology*, vol. 6, no. 6, pp. 590–610, Oct. 2012. DOI: 10.1016/j.molonc.2012.09.006.
- [26] B. Reva, Y. Antipin, C. Sander, “Predicting the functional impact of protein mutations: Application to cancer genomics,” *Nucleic Acids Research*, vol. 39, no. 17, e118–e118, Jul. 2011. DOI: 10.1093/nar/gkr407.
- [27] M. Ehrlich, “DNA hypomethylation in cancer cells,” *Epigenomics*, vol. 1, no. 2, pp. 239–259, Dec. 2009. DOI: 10.2217/epi.09.33.
- [28] L. Zhang, Q. Lu, C. Chang, “Epigenetics in health and disease,” in *Epigenetics in Allergy and Autoimmunity*. Springer Singapore, 2020, pp. 3–55, ISBN: 9789811534492. DOI: 10.1007/978-981-15-3449-2_1.
- [29] P. K. Kreeger, D. A. Lauffenburger, “Cancer systems biology: A network modeling perspective,” *Carcinogenesis*, vol. 31, no. 1, pp. 2–8, Oct. 2009. DOI: 10.1093/carcin/bgp261.
- [30] J. Chen, Y. Wang, B. Shen, D. Zhang, “Molecular signature of cancer at gene level or pathway level? case studies of colorectal cancer and prostate cancer microarray data,” *Computational and Mathematical Methods in Medicine*, vol. 2013, pp. 1–8, 2013. DOI: 10.1155/2013/909525.
- [31] M.-A. Lisio, L. Fu, A. Goyeneche, Z.-h. Gao, C. Telleria, “High-grade serous ovarian cancer: Basic sciences, clinical and therapeutic standpoints,” *International Journal of Molecular Sciences*, vol. 20, no. 4, p. 952, Feb. 2019. DOI: 10.3390/ijms20040952.
- [32] “Comprehensive molecular portraits of human breast tumours,” *Nature*, vol. 490, no. 7418, pp. 61–70, Sep. 2012. DOI: 10.1038/nature11412.
- [33] A. K. Pasanen *et al.*, “Cell cycle regulation score predicts relapse-free survival in non-germinal centre diffuse large B-cell lymphoma patients treated by means of immunochemotherapy,” *European Journal of Haematology*, vol. 91, no. 1, pp. 29–36, May 2013. DOI: 10.1111/ejh.12117.
- [34] H. K. Matthews, C. Bertoli, R. A. M. de Bruin, “Cell cycle control in cancer,” *Nature Reviews Molecular Cell Biology*, vol. 23, no. 1, pp. 74–88, Sep. 2021. DOI: 10.1038/s41580-021-00404-3.
- [35] Y.-J. Guo, W.-W. Pan, S.-B. Liu, Z.-F. Shen, Y. Xu, L.-L. Hu, “ERK/MAPK signalling pathway and tumorigenesis (review),” *Experimental and Therapeutic Medicine*, Jan. 2020. DOI: 10.3892/etm.2020.8454.
- [36] M. H. V. V. Regenmortel, “Reductionism and complexity in molecular biology: Scientists now have the tools to unravel biological complexity and overcome the limitations of reductionism,” *EMBO reports*, vol. 5, no. 11, pp. 1016–1020, Nov. 2004. DOI: 10.1038/sj.embor.7400284.
- [37] A. C. Ahn, M. Tewari, C.-S. Poon, R. S. Phillips, “The limits of reductionism in medicine: Could systems biology offer an alternative?” *PLoS Medicine*, vol. 3, no. 6, e208, May 2006. DOI: 10.1371/journal.pmed.0030208.

- [38] “A map of human genome variation from population-scale sequencing,” *Nature*, vol. 467, no. 7319, pp. 1061–1073, Oct. 2010. DOI: 10.1038/nature09534.
- [39] R. Edgar, “Gene Expression Omnibus: NCBI gene expression and hybridization array data repository,” *Nucleic Acids Research*, vol. 30, no. 1, pp. 207–210, Jan. 2002. DOI: 10.1093/nar/30.1.207.
- [40] F. S. Collins, A. D. Barker, “Mapping the cancer genome,” *Scientific American*, vol. 296, no. 3, pp. 50–57, Mar. 2007. DOI: 10.1038/scientificamerican0307-50.
- [41] E. S. Lander *et al.*, “Initial sequencing and analysis of the human genome,” *Nature*, vol. 409, no. 6822, pp. 860–921, Feb. 2001. DOI: 10.1038/35057062.
- [42] Z. Sondka *et al.*, “COSMIC: A curated database of somatic variants and clinical data for cancer,” *Nucleic Acids Research*, vol. 52, no. D1, pp. D1210–D1217, Nov. 2023. DOI: 10.1093/nar/gkad986.
- [43] Q. Al-Tashi *et al.*, “Machine learning models for the identification of prognostic and predictive cancer biomarkers: A systematic review,” *International Journal of Molecular Sciences*, vol. 24, no. 9, p. 7781, Apr. 2023. DOI: 10.3390/ijms24097781.
- [44] M. S. Cline *et al.*, “Exploring tcga pan-cancer data at the ucsc cancer genomics browser,” *Scientific Reports*, vol. 3, no. 1, Oct. 2013. DOI: 10.1038/srep02652.
- [45] C. Curtis *et al.*, “The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups,” *Nature*, vol. 486, no. 7403, pp. 346–352, Apr. 2012. DOI: 10.1038/nature10983.
- [46] U. D. Akavia *et al.*, “An integrated approach to uncover drivers of cancer,” *Cell*, vol. 143, no. 6, pp. 1005–1017, Dec. 2010. DOI: 10.1016/j.cell.2010.11.013.
- [47] P. Aylin, A. Bottle, A. Majeed, “Use of administrative data or clinical databases as predictors of risk of death in hospital: Comparison of models,” *BMJ*, vol. 334, no. 7602, p. 1044, Apr. 2007. DOI: 10.1136/bmj.39168.496366.55.
- [48] P. Chen *et al.*, “Identification of prognostic groups in high-grade serous ovarian cancer treated with platinum–taxane chemotherapy,” *Cancer Research*, vol. 75, no. 15, pp. 2987–2998, Aug. 2015. DOI: 10.1158/0008-5472.can-14-3242.
- [49] F. Hassiotou, D. Geddes, “Anatomy of the human mammary gland: Current status of knowledge,” *Clinical Anatomy*, vol. 26, no. 1, pp. 29–48, Sep. 2012. DOI: 10.1002/ca.22165.
- [50] A. J. Darlington, “Anatomy of the breast,” in *Digital Mammography*. Springer International Publishing, 2015, pp. 3–10. DOI: 10.1007/978-3-319-04831-4_1.

- [51] G. Dontu, M. Al-Hajj, W. M. Abdallah, M. F. Clarke, M. S. Wicha, “Stem cells in normal breast development and breast cancer,” *Cell Proliferation*, vol. 36, no. s1, pp. 59–72, Oct. 2003. DOI: 10.1046/j.1365-2184.36.s.1.6.x.
- [52] H. L. Olsson, M. L. Olsson, “The menstrual cycle and risk of breast cancer: A review,” *Frontiers in Oncology*, vol. 10, Jan. 2020. DOI: 10.3389/fonc.2020.00021.
- [53] B. Gusterson, “Do “basal-like” breast cancers really exist?” *Nature Reviews Cancer*, vol. 9, no. 2, pp. 128–134, Dec. 2008. DOI: 10.1038/nrc2571.
- [54] B. Paño *et al.*, “Pathways of lymphatic spread in gynecologic malignancies,” *RadioGraphics*, vol. 35, no. 3, pp. 916–945, May 2015. DOI: 10.1148/rg.2015140086.
- [55] A.-J. Twigger, W. T. Khaled, “Mammary gland development from a single cell ‘omics view,” *Seminars in Cell & Developmental Biology*, vol. 114, pp. 171–185, Jun. 2021. DOI: 10.1016/j.semcdb.2021.03.013.
- [56] L. A. Torre, R. L. Siegel, E. M. Ward, A. Jemal, “Global cancer incidence and mortality rates and trends—an update,” *Cancer Epidemiology, Biomarkers & Prevention*, vol. 25, no. 1, pp. 16–27, Jan. 2016. DOI: 10.1158/1055-9965.epi-15-0578.
- [57] R. L. Siegel, K. D. Miller, N. S. Wagle, A. Jemal, “Cancer statistics, 2023,” *CA: A Cancer Journal for Clinicians*, vol. 73, no. 1, pp. 17–48, Jan. 2023. DOI: 10.3322/caac.21763.
- [58] E. Değerli *et al.*, “Relationship between pathological response and molecular subtypes in locally advanced breast cancer patients receiving neoadjuvant chemotherapy,” *Journal of Chemotherapy*, vol. 35, no. 1, pp. 29–38, Feb. 2022. DOI: 10.1080/1120009x.2022.2043514.
- [59] S. Łukasiewicz, M. Czezelewski, A. Forma, J. Baj, R. Sitarz, A. Stanisławek, “Breast cancer—epidemiology, risk factors, classification, prognostic markers, and current treatment strategies—an updated review,” *Cancers*, vol. 13, no. 17, p. 4287, Aug. 2021. DOI: 10.3390/cancers13174287.
- [60] H. Li *et al.*, “Modern deep learning in bioinformatics,” *Journal of Molecular Cell Biology*, vol. 12, no. 11, pp. 823–827, Jun. 2020. DOI: 10.1093/jmcb/mjaa030.
- [61] F. Markowetz, “All biology is computational biology,” *PLOS Biology*, vol. 15, no. 3, e2002050, Mar. 2017. DOI: 10.1371/journal.pbio.2002050.
- [62] P. Hogeweg, “The roots of bioinformatics in theoretical biology,” *PLoS Computational Biology*, vol. 7, no. 3, D. B. Searls, Ed., e1002021, Mar. 2011. DOI: 10.1371/journal.pcbi.1002021.
- [63] J. Gauthier, A. T. Vincent, S. J. Charette, N. Derome, “A brief history of bioinformatics,” *Briefings in Bioinformatics*, vol. 20, no. 6, pp. 1981–1996, Aug. 2018. DOI: 10.1093/bib/bby063.

- [64] V. V. Klimontov, K. A. Koshechkin, N. G. Orlova, M. I. Sekacheva, Y. L. Orlov, "Medical genetics, genomics and bioinformatics—2022," *International Journal of Molecular Sciences*, vol. 24, no. 10, p. 8968, May 2023. DOI: 10.3390/ijms24108968.
- [65] G. R. Oliver, S. N. Hart, E. W. Klee, "Bioinformatics for clinical next generation sequencing," *Clinical Chemistry*, vol. 61, no. 1, pp. 124–135, Jan. 2015. DOI: 10.1373/clinchem.2014.224360.
- [66] H. Satam *et al.*, "Next-generation sequencing technology: Current trends and advancements," *Biology*, vol. 12, no. 7, p. 997, Jul. 2023. DOI: 10.3390/biology12070997.
- [67] F. Sanger, S. Nicklen, A. R. Coulson, "DNA sequencing with chain-terminating inhibitors," *Proceedings of the National Academy of Sciences*, vol. 74, no. 12, pp. 5463–5467, Dec. 1977. DOI: 10.1073/pnas.74.12.5463.
- [68] J. C. Venter *et al.*, "The sequence of the human genome," *Science*, vol. 291, no. 5507, pp. 1304–1351, Feb. 2001. DOI: 10.1126/science.1058040.
- [69] "An integrated map of genetic variation from 1, 092 human genomes," *Nature*, vol. 491, no. 7422, pp. 56–65, Oct. 2012. DOI: 10.1038/nature11632.
- [70] K. Walter *et al.*, "The uk10k project identifies rare variants in health and disease," *Nature*, vol. 526, no. 7571, pp. 82–90, Sep. 2015. DOI: 10.1038/nature14962.
- [71] D. Cox. "The \$100 genome: What breaking this accessibility barrier means for the future of genetic testing." Erişim tarihi: 17 Ocak 2024. (2023), [Online]. Available: <https://geneticliteracyproject.org/2023/01/30/the-100-genome-what-breaking-this-accessibility-barrier-means-for-the-future-of-genetic-testing/>.
- [72] K. A. Wetterstrand, *DNA sequencing costs: Data from the nhgri genome sequencing program (gsp)*, Erişim tarihi: 17 Ocak 2024, 2023. [Online]. Available: <https://www.genome.gov/about-genomics/fact-sheets/%7BDNA%7D-Sequencing-Costs-Data>.
- [73] J. Preston, A. VanZeeland, D. A. Peiffer, "Innovation at Illumina: The road to the \$600 human genome," *Nature*, 2021, Erişim tarihi: 17 Ocak 2024. [Online]. Available: <https://www.nature.com/articles/d42473-021-00030-9>.
- [74] E. R. Mardis, "A decade's perspective on DNA sequencing technology," *Nature*, vol. 470, no. 7333, Feb. 2011. DOI: 10.1038/nature09796.
- [75] J. Shendure *et al.*, "DNA sequencing at 40: Past, present and future," *Nature*, vol. 550, no. 7676, pp. 345–353, Oct. 2017. DOI: 10.1038/nature24286.

- [76] F. E. Dewey, S. Pan, M. T. Wheeler, S. R. Quake, E. A. Ashley, "DNA sequencing: Clinical applications of new DNA sequencing technologies," *Circulation*, vol. 125, no. 7, pp. 931–944, Feb. 2012. DOI: 10.1161/circulationaha.110.972828.
- [77] B. T. Kelly, J.-C. Baret, V. Taly, A. D. Griffiths, "Miniaturizing chemistry and biology in microdroplets," *Chemical Communications*, no. 18, p. 1773, 2007. DOI: 10.1039/b616252e.
- [78] M. Fedurco, "BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies," *Nucleic Acids Research*, vol. 34, no. 3, e22–e22, Feb. 2006. DOI: 10.1093/nar/gnj023.
- [79] D. R. Bentley *et al.*, "Accurate whole human genome sequencing using reversible terminator chemistry," *Nature*, vol. 456, no. 7218, pp. 53–59, Nov. 2008. DOI: 10.1038/nature07517.
- [80] J. C. Dohm, C. Lottaz, T. Borodina, H. Himmelbauer, "Substantial biases in ultra-short read data sets from high-throughput DNA sequencing," *Nucleic Acids Research*, vol. 36, no. 16, Jul. 2008. DOI: 10.1093/nar/gkn425.
- [81] S. L. Schwartz, M. L. Farman, "Systematic overrepresentation of DNA termini and underrepresentation of subterminal regions among sequencing templates prepared from hydrodynamically sheared linear DNA molecules," *BMC Genomics*, vol. 11, no. 1, p. 87, 2010. DOI: 10.1186/1471-2164-11-87.
- [82] E. R. Mardis, "Next-generation sequencing platforms," *Annual Review of Analytical Chemistry*, vol. 6, no. 1, pp. 287–303, Jun. 2013. DOI: 10.1146/annurev-anchem-062012-092628.
- [83] X.-S. Zhang, J.-J. Pei, L.-G. Zhao, F. Tang, X.-Y. Fang, "Rna-seq analysis and comparison of the enzymes involved in ionone synthesis of three cultivars of osmanthus," *Journal of Asian Natural Products Research*, vol. 20, no. 7, pp. 649–661, Apr. 2018. DOI: 10.1080/10286020.2018.1453503.
- [84] J. Chen, H. Zhang, M. Feng, D. Zuo, Y. Hu, T. Jiang, "Transcriptome analysis of woodland strawberry (*Fragaria vesca*) response to the infection by strawberry vein banding virus (SVBV)," *Virology Journal*, vol. 13, no. 1, Jul. 2016. DOI: 10.1186/s12985-016-0584-5.
- [85] Y. Wang *et al.*, "Analyses of long non-coding RNA and mRNA profiling using RNA sequencing during the pre-implantation phases in pig endometrium," *Scientific Reports*, vol. 6, no. 1, Jan. 2016. DOI: 10.1038/srep20238.
- [86] S. Zhao, W.-P. Fung-Leung, A. Bittner, K. Ngo, X. Liu, "Comparison of RNA-seq and microarray in transcriptome profiling of activated T cells," *PLoS ONE*, vol. 9, no. 1, S.-D. Zhang, Ed., e78644, Jan. 2014. DOI: 10.1371/journal.pone.0078644.
- [87] T. Ching, S. Huang, L. X. Garmire, "Power analysis and sample size estimation for RNA-seq differential expression," *RNA*, vol. 20, no. 11, pp. 1684–1696, Sep. 2014. DOI: 10.1261/rna.046011.114.

- [88] F. Finotello, B. Di Camillo, “Measuring differential gene expression with rna-seq: Challenges and strategies for data analysis,” *Briefings in Functional Genomics*, vol. 14, no. 2, pp. 130–142, Sep. 2014. DOI: 10.1093/bfgp/elu035.
- [89] V. Raghavan, L. Kraft, F. Mesny, L. Rigerte, “A simple guide to de novo transcriptome assembly and annotation,” *Briefings in Bioinformatics*, vol. 23, no. 2, Jan. 2022. DOI: 10.1093/bib/bbab563.
- [90] D. Deshpande *et al.*, “Rna-seq data science: From raw data to effective interpretation,” *Frontiers in Genetics*, vol. 14, Mar. 2023. DOI: 10.3389/fgene.2023.997383.
- [91] S. Gotz *et al.*, “High-throughput functional annotation and data mining with the blast2go suite,” *Nucleic Acids Research*, vol. 36, no. 10, pp. 3420–3435, Apr. 2008. DOI: 10.1093/nar/gkn176.
- [92] J. Lovén *et al.*, “Revisiting global gene expression analysis,” *Cell*, vol. 151, no. 3, pp. 476–482, Oct. 2012, ISSN: 0092-8674. DOI: 10.1016/j.cell.2012.10.012. [Online]. Available: <http://dx.doi.org/10.1016/j.cell.2012.10.012>.
- [93] K. R. Kumar, M. J. Cowley, R. L. Davis, “Next-generation sequencing and emerging technologies*,” *Seminars in Thrombosis and Hemostasis*, May 2024. DOI: 10.1055/s-0044-1786397.
- [94] D. Li, M. S. Zand, T. D. Dye, M. L. Goniewicz, I. Rahman, Z. Xie, “An evaluation of rna-seq differential analysis methods,” *PLOS ONE*, vol. 17, no. 9, D. J. Stekel, Ed., e0264246, Sep. 2022. DOI: 10.1371/journal.pone.0264246.
- [95] A. McDermaid, B. Monier, J. Zhao, B. Liu, Q. Ma, “Interpretation of differential gene expression results of rna-seq data: Review and integration,” *Briefings in Bioinformatics*, vol. 20, no. 6, pp. 2044–2054, Aug. 2018. DOI: 10.1093/bib/bby067.
- [96] S. Anders *et al.*, “Count-based differential expression analysis of rna sequencing data using r and bioconductor,” *Nature Protocols*, vol. 8, no. 9, pp. 1765–1786, Aug. 2013. DOI: 10.1038/nprot.2013.099.
- [97] S. Liu, Z. Wang, R. Zhu, F. Wang, Y. Cheng, Y. Liu, “Three differential expression analysis methods for rna sequencing: Limma, edger, deseq2,” *Journal of Visualized Experiments*, no. 175, Sep. 2021. DOI: 10.3791/62528.
- [98] C. W. Law, Y. Chen, W. Shi, G. K. Smyth, “Voom: Precision weights unlock linear model analysis tools for rna-seq read counts,” *Genome Biology*, vol. 15, no. 2, R29, 2014. DOI: 10.1186/gb-2014-15-2-r29.
- [99] R. Liu *et al.*, “Why weight? modelling sample and observational level variability improves power in rna-seq analyses,” *Nucleic Acids Research*, vol. 43, no. 15, e97–e97, Apr. 2015. DOI: 10.1093/nar/gkv412.
- [100] F. Rapaport *et al.*, “Comprehensive evaluation of differential gene expression analysis methods for rna-seq data,” *Genome Biology*, vol. 14, no. 9, R95, 2013. DOI: 10.1186/gb-2013-14-9-r95.

- [101] M. E. Ritchie *et al.*, “Limma powers differential expression analyses for rna-sequencing and microarray studies,” *Nucleic Acids Research*, vol. 43, no. 7, e47–e47, Jan. 2015. DOI: 10.1093/nar/gkv007.
- [102] M. Di Filippo, C. Damiani, D. Pescini, “Gpruler: Metabolic gene-protein-reaction rules automatic reconstruction,” *PLOS Computational Biology*, vol. 17, no. 11, C. D. Maranas, Ed., e1009550, Nov. 2021. DOI: 10.1371/journal.pcbi.1009550.
- [103] A. Richelle, C. Joshi, N. E. Lewis, “Assessing key decisions for transcriptomic data integration in biochemical networks,” *PLOS Computational Biology*, vol. 15, no. 7, V. Hatzimanikatis, Ed., e1007185, Jul. 2019. DOI: 10.1371/journal.pcbi.1007185.
- [104] D. Machado, M. J. Herrgård, I. Rocha, “Stoichiometric representation of gene–protein–reaction associations leverages constraint-based analysis from reaction to gene-level phenotype prediction,” *PLOS Computational Biology*, vol. 12, no. 10, K. R. Patil, Ed., e1005140, Oct. 2016. DOI: 10.1371/journal.pcbi.1005140.
- [105] M. Ponce-de-León, I. Apaolaza, A. Valencia, F. J. Planes, “On the inconsistent treatment of gene-protein-reaction rules in context-specific metabolic models,” *Bioinformatics*, vol. 36, no. 6, J. Kelso, Ed., pp. 1986–1988, Nov. 2019. DOI: 10.1093/bioinformatics/btz832.
- [106] C. Hutter, J. C. Zenklusen, “The cancer genome atlas: Creating lasting value beyond its data,” *Cell*, vol. 173, no. 2, pp. 283–285, Apr. 2018. DOI: 10.1016/j.cell.2018.03.042.
- [107] K. Tomczak, P. Czerwińska, M. Wiznerowicz, “Review the cancer genome atlas (tcga): An immeasurable source of knowledge,” *Współczesna Onkologia*, vol. 1A, pp. 68–77, 2015. DOI: 10.5114/wo.2014.47136.
- [108] N. C. Duarte *et al.*, “Global reconstruction of the human metabolic network based on genomic and bibliomic data,” *Proceedings of the National Academy of Sciences*, vol. 104, no. 6, pp. 1777–1782, Feb. 2007. DOI: 10.1073/pnas.0610772104.
- [109] J. Schellenberger, J. O. Park, T. M. Conrad, B. Ø. Palsson, “Bigg: A biochemical genetic and genomic knowledgebase of large scale metabolic reconstructions,” *BMC Bioinformatics*, vol. 11, no. 1, Apr. 2010. DOI: 10.1186/1471-2105-11-213.
- [110] M. Kanehisa, Y. Sato, M. Kawashima, M. Furumichi, M. Tanabe, “Kegg as a reference resource for gene and protein annotation,” *Nucleic Acids Research*, vol. 44, no. D1, pp. D457–D462, Oct. 2015. DOI: 10.1093/nar/gkv1070.
- [111] “Uniprot: A hub for protein information,” *Nucleic Acids Research*, vol. 43, no. D1, pp. D204–D212, Oct. 2014. DOI: 10.1093/nar/gku989.
- [112] S. Povey, R. Lovering, E. Bruford, M. Wright, M. Lush, H. Wain, “The hugo gene nomenclature committee (hgnc),” *Human Genetics*, vol. 109, no. 6, pp. 678–680, Oct. 2001. DOI: 10.1007/s00439-001-0615-0.

- [113] E. Selkov, “The metabolic pathway collection from emp: The enzymes and metabolic pathways database,” *Nucleic Acids Research*, vol. 24, no. 1, pp. 26–28, Jan. 1996. DOI: 10.1093/nar/24.1.26.
- [114] T. Hao, H.-W. Ma, X.-M. Zhao, I. Goryanin, “Compartmentalization of the edinburgh human metabolic network,” *BMC Bioinformatics*, vol. 11, no. 1, Jul. 2010. DOI: 10.1186/1471-2105-11-393.
- [115] P. Romero, J. Wagg, M. L. Green, D. Kaiser, M. Krummenacker, P. D. Karp, *Genome Biology*, vol. 6, no. 1, R2, 2004. DOI: 10.1186/gb-2004-6-1-r2.
- [116] C. Gille *et al.*, “Hepatonet1: A comprehensive metabolic reconstruction of the human hepatocyte for the analysis of liver physiology,” *Molecular Systems Biology*, vol. 6, no. 1, Jan. 2010. DOI: 10.1038/msb.2010.62.
- [117] A. Mardinoglu *et al.*, “Integration of clinical data with a genome-scale metabolic model of the human adipocyte,” *Molecular Systems Biology*, vol. 9, no. 1, Jan. 2013. DOI: 10.1038/msb.2013.5.
- [118] M. Sud, E. Fahy, D. Cotter, E. A. Dennis, S. Subramaniam, “Lipid maps-nature lipidomics gateway: An online resource for students and educators interested in lipids,” *Journal of Chemical Education*, vol. 89, no. 2, pp. 291–292, Dec. 2011. DOI: 10.1021/ed200088u.
- [119] D. Croft, “Building models using reactome pathways as templates,” in *In Silico Systems Biology*. Humana Press, 2013, pp. 273–283, ISBN: 9781627034500. DOI: 10.1007/978-1-62703-450-0_14.
- [120] N. Pornputtapong, I. Nookaew, J. Nielsen, “Human metabolic atlas: An online resource for human metabolism,” *Database*, vol. 2015, Jan. 2015. DOI: 10.1093/database/bav068.
- [121] I. Thiele *et al.*, “A community-driven global reconstruction of human metabolism,” *Nature Biotechnology*, vol. 31, no. 5, pp. 419–425, Mar. 2013. DOI: 10.1038/nbt.2488.
- [122] S. Sahoo, L. Franzson, J. J. Jonsson, I. Thiele, “A compendium of inborn errors of metabolism mapped onto the human metabolic network,” *Molecular BioSystems*, vol. 8, no. 10, p. 2545, 2012. DOI: 10.1039/c2mb25075f.
- [123] S. Sahoo, I. Thiele, “Predicting the impact of diet and enzymopathies on human small intestinal epithelial cells,” *Human Molecular Genetics*, vol. 22, no. 13, pp. 2705–2722, Mar. 2013. DOI: 10.1093/hmg/ddt119.
- [124] N. Swainston *et al.*, “Recon 2.2: From reconstruction to model of human metabolism,” *Metabolomics*, vol. 12, no. 7, Jun. 2016. DOI: 10.1007/s11306-016-1051-4.
- [125] E. Brunk *et al.*, “Recon3d enables a three-dimensional view of gene variation in human metabolism,” *Nature Biotechnology*, vol. 36, no. 3, pp. 272–281, Feb. 2018. DOI: 10.1038/nbt.4072.
- [126] J. L. Robinson *et al.*, “An atlas of human metabolism,” *Science Signaling*, vol. 13, no. 624, Mar. 2020. DOI: 10.1126/scisignal.aaz1482.

- [127] D. Schmidt, W.-C. Chen, M. A. Matheson, G. Ostrouchov, “Programming with big data in r: Scaling analytics from one to thousands of nodes,” *Big Data Research*, vol. 8, pp. 1–11, Jul. 2017. DOI: 10.1016/j.bdr.2016.10.002.
- [128] R Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2023. [Online]. Available: <https://www.R-project.org/>.
- [129] F. Seyednasrollah, A. Laiho, L. L. Elo, “Comparison of software packages for detecting differential expression in rna-seq studies,” *Briefings in Bioinformatics*, vol. 16, no. 1, pp. 59–70, Dec. 2013. DOI: 10.1093/bib/bbt086.
- [130] W. Huber *et al.*, “Orchestrating high-throughput genomic analysis with bioconductor,” *Nature Methods*, vol. 12, no. 2, pp. 115–121, Jan. 2015. DOI: 10.1038/nmeth.3252.
- [131] H. Jalal, P. Pechlivanoglou, E. Krijkamp, F. Alarid-Escudero, E. Enns, M. G. M. Hunink, “An overview of r in health decision sciences,” *Medical Decision Making*, vol. 37, no. 7, pp. 735–746, Jan. 2017. DOI: 10.1177/0272989x16686559.
- [132] Posit team, *Rstudio: Integrated development environment for r*, Posit Software, PBC, Boston, MA, 2023. [Online]. Available: <http://www.posit.co/>.
- [133] J. DOI, G. POTTER, J. WONG, I. ALCARAZ, P. CHI, “Web application teaching tools for statistics using r and shiny,” *Technology Innovations in Statistics Education*, vol. 9, no. 1, 2016, ISSN: 1933-4214. DOI: 10.5070/t591027492. [Online]. Available: <http://dx.doi.org/10.5070/T591027492>.
- [134] B. Baumer, D. Udwin, “R markdown,” *WIREs Computational Statistics*, vol. 7, no. 3, pp. 167–177, Feb. 2015. DOI: 10.1002/wics.1348.
- [135] H. Wickham *et al.*, “Welcome to the tidyverse,” *Journal of Open Source Software*, vol. 4, no. 43, p. 1686, Nov. 2019. DOI: 10.21105/joss.01686.
- [136] H. Wickham, *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. [Online]. Available: <https://ggplot2.tidyverse.org>.
- [137] T. L. Staples, “Expansion and evolution of the r programming language,” *Royal Society Open Science*, vol. 10, no. 4, Apr. 2023. DOI: 10.1098/rsos.221550.
- [138] C. Diener, O. Resendis-Antonio, “Personalized prediction of proliferation rates and metabolic liabilities in cancer biopsies,” *Frontiers in Physiology*, vol. 7, Dec. 2016. DOI: 10.3389/fphys.2016.00644.
- [139] Y. Xie, J. Cheng, X. Tan, *DT: A wrapper of the javascript library 'DataTables'*, R package version 0.28, 2023. [Online]. Available: <https://CRAN.R-project.org/package=DT>.

- [140] T. L. Pedersen, *GGraph: An implementation of grammar of graphics for graphs and networks*, R package version 2.1.0, 2022. [Online]. Available: <https://CRAN.R-project.org/package=ggraph>.
- [141] T. L. Pedersen, *Tidygraph: A tidy api for graph manipulation*, R package version 1.2.3, 2023. [Online]. Available: <https://CRAN.R-project.org/package=tidygraph>.
- [142] Y. Xie, “Knitr: A comprehensive tool for reproducible research in R,” in *Implementing Reproducible Research*. Chapman and Hall/CRC, Dec. 2018, pp. 3–31. DOI: 10.1201/9781315373461-1.
- [143] H. Wickham, L. Henry, *Purrr: Functional programming tools*, R package version 1.0.1, 2023. [Online]. Available: <https://CRAN.R-project.org/package=purrr>.
- [144] D. Vaughan, M. Dancho, *Furrr: Apply mapping functions in parallel using futures*, R package version 0.3.1, 2022. [Online]. Available: <https://CRAN.R-project.org/package=furrr>.
- [145] R. C. Gentleman *et al.*, “Bioconductor: Open software development for computational biology and bioinformatics,” *Genome Biology*, vol. 5, no. 10, R80, Sep. 2004. DOI: 10.1186/gb-2004-5-10-r80.
- [146] M. Morgan, M. Ramos, *Biocmanager: Access the bioconductor project package repository*, R package version 1.30.22, 2023. [Online]. Available: <https://CRAN.R-project.org/package=BiocManager>.
- [147] M. E. Ritchie *et al.*, “Limma powers differential expression analyses for RNA-sequencing and microarray studies,” *Nucleic Acids Research*, vol. 43, no. 7, e47–e47, Jan. 2015. DOI: 10.1093/nar/gkv007.
- [148] “edgeR: a Bioconductor package for differential expression analysis of digital gene expression data,”
- [149] H. Wickham, J. Bryan, *readxl: Read Excel Files*, R package version 1.4.2, 2023. [Online]. Available: <https://CRAN.R-project.org/package=readxl>.
- [150] P. Schauburger, A. Walker, *Openxlsx: Read, write and edit xlsx files*, R package version 4.2.5.2, 2023. [Online]. Available: <https://CRAN.R-project.org/package=openxlsx>.
- [151] J. Cooke *et al.*, “Genome scale metabolic network modelling for metabolic profile predictions,” Jul. 2023. DOI: 10.1101/2023.07.26.550622.
- [152] A. O. Zheng, A. Sher, D. Fridman, C. J. Musante, J. D. Young, “Pool size measurements improve precision of flux estimates but increase sensitivity to unmodeled reactions outside the core network in isotopically nonstationary metabolic flux analysis (inst-mfa),” *Biotechnology Journal*, vol. 17, no. 3, Feb. 2022. DOI: 10.1002/biot.202000427.
- [153] V. S. Martínez, P. A. Saa, J. Jooste, K. Tiwari, L.-E. Quek, L. K. Nielsen, “The topology of genome-scale metabolic reconstructions unravels independent modules and high network flexibility,” *PLOS Computational Biology*, vol. 18, no. 6, T. Cakir, Ed., e1010203, Jun. 2022. DOI: 10.1371/journal.pcbi.1010203.

- [154] L. van Steijn, F. J. Verbeek, H. P. Spaink, R. M. Merks, “Predicting metabolism from gene expression in an improved whole-genome metabolic network model of *Danio rerio*,” *Zebrafish*, vol. 16, no. 4, pp. 348–362, Aug. 2019. DOI: 10.1089/zeb.2018.1712.
- [155] S. Takano, J. C. C. Vila, R. Miyazaki, Á. Sánchez, D. Bajić, “The architecture of metabolic networks constrains the evolution of microbial resource hierarchies,” *Molecular Biology and Evolution*, vol. 40, no. 9, J. Zhang, Ed., Aug. 2023. DOI: 10.1093/molbev/msad187.
- [156] S. Kamiyama *et al.*, “Expression and the role of 3’-phosphoadenosine 5’-phosphosulfate transporters in human colorectal carcinoma,” *Glycobiology*, vol. 21, no. 2, pp. 235–246, Oct. 2010. DOI: 10.1093/glycob/cwq154.
- [157] J. W. Locasale, “Serine, glycine and one-carbon units: Cancer metabolism in full circle,” *Nature Reviews Cancer*, vol. 13, no. 8, pp. 572–583, Jul. 2013. DOI: 10.1038/nrc3557.
- [158] R. Flavin, S. Peluso, P. L. Nguyen, M. Loda, “Fatty acid synthase as a potential therapeutic target in cancer,” *Future Oncology*, vol. 6, no. 4, pp. 551–562, Apr. 2010. DOI: 10.2217/fon.10.11.
- [159] L. A. Sena, N. S. Chandel, “Physiological roles of mitochondrial reactive oxygen species,” *Molecular Cell*, vol. 48, no. 2, pp. 158–167, Oct. 2012. DOI: 10.1016/j.molcel.2012.09.025.
- [160] N. Sánchez-Baizán, L. Ribas, F. Piferrer, “Improved biomarker discovery through a plot twist in transcriptomic data analysis,” *BMC Biology*, vol. 20, no. 1, Sep. 2022. DOI: 10.1186/s12915-022-01398-w.
- [161] M. Hong *et al.*, “RNA sequencing: New technologies and applications in cancer research,” *Journal of Hematology & Oncology*, vol. 13, no. 1, Dec. 2020. DOI: 10.1186/s13045-020-01005-x.
- [162] M. Cieřlik, A. M. Chinnaiyan, “Cancer transcriptome profiling at the juncture of clinical translation,” *Nature Reviews Genetics*, vol. 19, no. 2, pp. 93–109, Dec. 2017. DOI: 10.1038/nrg.2017.96.
- [163] B. Moiz, A. Li, S. Padmanabhan, G. Sriram, A. M. Clyne, “Isotope-assisted metabolic flux analysis: A powerful technique to gain new insights into the human metabolome in health and disease,” *Metabolites*, vol. 12, no. 11, p. 1066, Nov. 2022. DOI: 10.3390/metabo12111066.
- [164] N. N. Pavlova, C. B. Thompson, “The emerging hallmarks of cancer metabolism,” *Cell Metabolism*, vol. 23, no. 1, pp. 27–47, Jan. 2016. DOI: 10.1016/j.cmet.2015.12.006.
- [165] J. Wang, C. Wang, H. Liu, H. Qi, H. Chen, J. Wen, “Metabolomics assisted metabolic network modeling and network wide analysis of metabolites in microbiology,” *Critical Reviews in Biotechnology*, vol. 38, no. 7, pp. 1106–1120, Apr. 2018. DOI: 10.1080/07388551.2018.1462141.
- [166] M. Huss, P. Holme, “Currency and commodity metabolites: Their identification and relation to the modularity of metabolic networks,” *IET Systems Biology*, vol. 1, no. 5, pp. 280–285, Sep. 2007. DOI: 10.1049/iet-syb:20060077.

- [167] S.-i. Akiyama *et al.*, “The role of thymidine phosphorylase, an angiogenic enzyme, in tumor progression,” *Cancer Science*, vol. 95, no. 11, pp. 851–857, Nov. 2004. DOI: 10.1111/j.1349-7006.2004.tb02193.x.
- [168] A. C. Newman, O. D. K. Maddocks, “One-carbon metabolism in cancer,” *British Journal of Cancer*, vol. 116, no. 12, pp. 1499–1504, May 2017. DOI: 10.1038/bjc.2017.118.
- [169] P.-C. Ho *et al.*, “Phosphoenolpyruvate is a metabolic checkpoint of anti-tumor T cell responses,” *Cell*, vol. 162, no. 6, pp. 1217–1228, Sep. 2015. DOI: 10.1016/j.cell.2015.08.012.
- [170] A. Siddiqui, P. Ceppi, “A non-proliferative role of pyrimidine metabolism in cancer,” *Molecular Metabolism*, vol. 35, p. 100962, May 2020. DOI: 10.1016/j.molmet.2020.02.005.
- [171] R. Garozzo, M. A. Sortino, C. Vancheri, D. F. Condorelli, “Antiproliferative effects induced by guanine-based purines require hypoxanthine-guanine phosphoribosyltransferase activity,” *Biological Chemistry*, vol. 391, no. 9, Sep. 2010. DOI: 10.1515/bc.2010.106.
- [172] R. Kohnken, K. M. Kodigepalli, L. Wu, “Regulation of deoxynucleotide metabolism in cancer: Novel mechanisms and therapeutic implications,” *Molecular Cancer*, vol. 14, no. 1, Sep. 2015. DOI: 10.1186/s12943-015-0446-6.
- [173] C. K. Mathews, “Deoxyribonucleotide metabolism, mutagenesis and cancer,” *Nature Reviews Cancer*, vol. 15, no. 9, pp. 528–539, Aug. 2015. DOI: 10.1038/nrc3981.
- [174] A. Singh, N. Deshpande, N. Pramanik, S. Jhunjhunwala, A. Rangarajan, H. S. Atreya, “Optimized peptide based inhibitors targeting the dihydrofolate reductase pathway in cancer,” *Scientific Reports*, vol. 8, no. 1, Feb. 2018. DOI: 10.1038/s41598-018-21435-5.
- [175] Y. Maeda, T. Kinoshita, “Dolichol-phosphate mannose synthase: Structure, function and regulation,” *Biochimica et Biophysica Acta (BBA) - General Subjects*, vol. 1780, no. 6, pp. 861–868, Jun. 2008. DOI: 10.1016/j.bbagen.2008.03.005.
- [176] E. S. Arnér, A. Holmgren, “The thioredoxin system in cancer,” *Seminars in Cancer Biology*, vol. 16, no. 6, pp. 420–426, Dec. 2006. DOI: 10.1016/j.semcancer.2006.10.009.
- [177] T. C. Karlenius, K. F. Tonissen, “Thioredoxin and cancer: A role for thioredoxin in all states of tumor oxygenation,” *Cancers*, vol. 2, no. 2, pp. 209–232, Mar. 2010. DOI: 10.3390/cancers2020209.
- [178] C. M. Mansbach, S. A. Siddiqi, “The biogenesis of chylomicrons,” *Annual Review of Physiology*, vol. 72, no. 1, pp. 315–333, Mar. 2010. DOI: 10.1146/annurev-physiol-021909-135801.
- [179] S. R. Nagarajan, L. M. Butler, A. J. Hoy, “The diversity and breadth of cancer cell fatty acid metabolism,” *Cancer & Metabolism*, vol. 9, no. 1, Jan. 2021. DOI: 10.1186/s40170-020-00237-2.

- [180] M. Liu *et al.*, “Inhibiting both proline biosynthesis and lipogenesis synergistically suppresses tumor growth,” *Journal of Experimental Medicine*, vol. 217, no. 3, Jan. 2020. DOI: 10.1084/jem.20191226.
- [181] E. J. Kay *et al.*, “Cancer-associated fibroblasts require proline synthesis by *pycr1* for the deposition of pro-tumorigenic extracellular matrix,” *Nature Metabolism*, vol. 4, no. 6, pp. 693–710, Jun. 2022. DOI: 10.1038/s42255-022-00582-0.
- [182] A. N. Bogner, K. M. Stiers, J. J. Tanner, “Structure, biochemistry, and gene expression patterns of the proline biosynthetic enzyme pyrroline-5-carboxylate reductase (*pycr*), an emerging cancer therapy target,” *Amino Acids*, vol. 53, no. 12, pp. 1817–1834, May 2021. DOI: 10.1007/s00726-021-02999-5.
- [183] O. N. Ivanova *et al.*, “Transcriptome analysis of redox systems and polyamine metabolic pathway in hepatoma and non-tumor hepatocyte-like cells,” *Biomolecules*, vol. 13, no. 4, p. 714, Apr. 2023. DOI: 10.3390/biom13040714.
- [184] D. J. Becker, J. B. Lowe, “Fucose: Biosynthesis and biological function in mammals,” *Glycobiology*, vol. 13, no. 7, 41R–53R, Mar. 2003. DOI: 10.1093/glycob/cwg054.
- [185] S. S. Pinho, C. A. Reis, “Glycosylation in cancer: Mechanisms and clinical implications,” *Nature Reviews Cancer*, vol. 15, no. 9, pp. 540–555, Aug. 2015. DOI: 10.1038/nrc3982.
- [186] M. W. Leal Quirino *et al.*, “FUCA2 and TSTA3 expression in gastric cancer: Candidate biomarkers of malignant transformation,” *Folia Histochemica et Cytobiologica*, vol. 60, no. 4, pp. 335–343, Dec. 2022. DOI: 10.5603/fhc.a2022.0031.
- [187] J. Niittymäki, P. Mattila, R. Renkonen, “Differential gene expression of GDP-l-fucose-synthesizing enzymes, GDP-fucose transporter and fucosyltransferase vii,” *APMIS*, vol. 114, no. 7–8, pp. 539–548, Aug. 2006. DOI: 10.1111/j.1600-0463.2006.apm_461.x.
- [188] D. M. Gilkes, S. Bajpai, P. Chaturvedi, D. Wirtz, G. L. Semenza, “Hypoxia-inducible factor 1 (HIF-1) promotes extracellular matrix remodeling under hypoxic conditions by inducing *p4ha1*, *p4ha2*, and *plod2* expression in fibroblasts,” *Journal of Biological Chemistry*, vol. 288, no. 15, pp. 10 819–10 829, Apr. 2013. DOI: 10.1074/jbc.m112.442939.
- [189] C. D. Madsen *et al.*, “Hypoxia and loss of *phd 2* inactivate stromal fibroblasts to decrease tumour stiffness and metastasis,” *EMBO reports*, vol. 16, no. 10, pp. 1394–1408, Aug. 2015. DOI: 10.15252/embr.201540107.
- [190] Y.-L. Chen, S. Eriksson, Z.-F. Chang, “Regulation and functional contribution of thymidine kinase 1 in repair of DNA damage,” *Journal of Biological Chemistry*, vol. 285, no. 35, pp. 27 327–27 335, Aug. 2010. DOI: 10.1074/jbc.m110.137042.

- [191] L. Slot Christiansen, B. Munch-Petersen, W. Knecht, “Non-viral deoxyribonucleoside kinases – diversity and practical use,” *Journal of Genetics and Genomics*, vol. 42, no. 5, pp. 235–248, May 2015. DOI: 10 . 1016/j.jgg.2015.01.003.
- [192] E. E. Bitter, M. H. Townsend, R. Erickson, C. Allen, K. L. O’Neill, “Thymidine kinase 1 through the ages: A comprehensive review,” *Cell & Bioscience*, vol. 10, no. 1, Nov. 2020. DOI: 10 . 1186/s13578-020-00493-1.
- [193] R. Mori, J. Ukai, Y. Tokumaru, Y. Niwa, M. Futamura, “The mechanism underlying resistance to 5-fluorouracil and its reversal by the inhibition of thymidine phosphorylase in breast cancer cells,” *Oncology Letters*, vol. 24, no. 3, Jul. 2022. DOI: 10 . 3892/ol.2022.13431.
- [194] L. Peng *et al.*, “Large-scale RNA-Seq transcriptome analysis of 4043 cancers and 548 normal tissue controls across 12 tcga cancer types,” *Scientific Reports*, vol. 5, no. 1, Aug. 2015. DOI: 10 . 1038/srep13413.
- [195] M. M. Alegre, R. A. Robison, K. L. O’Neill, “Thymidine kinase 1: A universal marker for cancer,” *Cancer and Clinical Oncology*, vol. 2, no. 1, May 2013. DOI: 10 . 5539/cco.v2n1p159.
- [196] H. Sun *et al.*, “P53 transcriptionally regulates sqle to repress cholesterol synthesis and tumor growth,” *EMBO reports*, vol. 22, no. 10, Aug. 2021. DOI: 10 . 15252/embr.202152537.
- [197] S. Feltrin *et al.*, “Sterol synthesis pathway inhibition as a target for cancer treatment,” *Cancer Letters*, vol. 493, pp. 19–30, Nov. 2020. DOI: 10 . 1016/j.canlet.2020.07.010.
- [198] C. Rao, “Chemopreventive effect of squalene on colon cancer,” *Carcinogenesis*, vol. 19, no. 2, pp. 287–290, Feb. 1998. DOI: 10 . 1093/carcin/19.2.287.
- [199] D. N. Brown *et al.*, “Squalene epoxidase is a bona fide oncogene by amplification with clinical relevance in breast cancer,” *Scientific Reports*, vol. 6, no. 1, Jan. 2016. DOI: 10 . 1038/srep19435.
- [200] G. Cirmena *et al.*, “Squalene epoxidase as a promising metabolic target in cancer treatment,” *Cancer Letters*, vol. 425, pp. 13–20, Jul. 2018. DOI: 10.1016/j.canlet.2018.03.034.

TEZDEN ÜRETİLMİŞ YAYINLAR

Konferans Bildirisi

1. M. Celik, A. Yilmaz, “Investigation of biochemical effects by combining breast cancer next-generation sequencing data,” in *int Conf. 2nd International Conference on Engineering and Applied Natural Sciences*, Oct. 2022, pp. 294–294. [Online]. Available: <https://www.iceans.org/2-nd-iceans-2022>.

