

ANKARA ÜNİVERSİTESİ
BİYOTEKNOLOJİ ENSTİTÜSÜ

TEMEL BİYOTEKNOLOJİ
DOKTORA TEZİ

MEME KANSERİNDE KODLANMAYAN RNA'LARIN DÜZENLEYİCİ
ROLLERİNİN BİYOİNFORMATİK OLARAK ARAŞTIRILMASI

Begüm Karaoğlu

Danışman Öğretim Üyesi
Prof. Dr. Bala Gür Dedeoğlu

Ocak

2025

ETİK BEYAN

Bu tez çalışmasının; akademik kural ve etik ilkelere bağılı kalınarak hazırlandığını, çalışmada yararlanılan ve bu çalışma ürünü olmayan bütün bilgiler için kaynak yayınlara atıfta bulunulmuş olduğunu beyan ederim.

Begüm Karaoğlu

İmzası

ONAY

Prof. Dr. Bala Gr Dedeoėlu danıřmanlıėında Begm Karaoėlu tarafından hazırlanan bu alıřma 30/01/2025 tarihinde ařaėıdaki jri tarafından Temel Biyoteknoloji Anabilim Dalı'nda doktora tezi olarak kabul edilmiřtir.

Bařkan: Do. Dr. zlen Konu Karakayalı

İmza:

ye: Prof. Dr. Bala Gr Dedeoėlu

İmza:

ye: Prof. Dr. Pelin Mutlu

İmza:

ye: Prof. Dr. Hasan Oėul

İmza:

ye: Prof. Dr. Emre Keskin

İmza:

Yukarıdaki sonucu onaylarım.

Prof. Dr. Demet Cansaran Duman

Enstit Mdr

ÖZET

Doktora Tezi

Meme Kanserinde Kodlanmayan RNA'ların Düzenleyici Rollerinin Biyoinformatik Olarak Araştırılması

Begüm Karaoğlu

Ankara Üniversitesi Biyoteknoloji Enstitüsü

Prof. Dr. Bala Gür Dedeoğlu

Meme kanseri, dünya çapında kadınlar arasında en yaygın görülen kanser türüdür. Bu durum, hastalığın tanı ve tedavisinde kullanılabilecek yeni ve etkin biyobelirteçlerin keşfinin önemini vurgulamaktadır. Gen ifadesinin kontrolünde görev alan moleküllerin (transkripsiyon faktörleri, miRNA'lar ve lncRNA'lar gibi) birbirleriyle olan etkileşimlerinin belirlenmesinin, meme kanser oluşumunda rol oynayan karmaşık süreçlerin sistem biyolojisi yaklaşımıyla çözülmesine katkı sağlayacağı düşünülmektedir.

Bu amaçla tez çalışması kapsamında, invaziv duktal meme kanseri (IDC) hastalarından elde edilen 179 tümör ve 51 normal doku içeren 5 Yeni Nesil Dizileme (NGS) çalışması ile 490 tümör ve 122 normal doku içeren 8 mikrodizin çalışmasına ait ifade verileri GEO veri tabanından elde edilmiştir. Diferansiyel ifade analizleri sonucunda mRNA, miRNA, lncRNA ve TF düzenleyici moleküllerine ait ifade verileri, tez kapsamında geliştirilen devre analiz programı *miRCuit*'a yüklenerek incelenmiştir.

Analizler sonucunda elde edilen devrelerden lncRNA CASC15, miR-130b-3p ve transkripsiyon faktörü KLF5'in meme kanseri oluşumunda önemli düzenleyici elemanlar olduğunu belirlenmiştir. Geliştirilen program ve bu çalışma sonucunda ortaya koyulan devrelerin, düzenleyici moleküller arasındaki ifade değişikliklerinin etkilerini anlamaya yardımcı olacağı ve tümör gelişimi sürecine dair kapsamlı bir bakış açısı sunacağı düşünülmektedir. Ayrıca, geliştirilen bu program, farklı çalışma alanlarından araştırmacıların düzenleyici devreler oluşturmaya olanak tanıyacaktır.

2025, 165 sayfa

Anahtar kelimeler: meme kanseri; gen regülasyonu; transkripsiyon faktörü; miRNA; lncRNA.

ABSTRACT

PhD Thesis

The Bioinformatics Investigation of the Regulatory Roles of Non-Coding RNAs in Breast Cancer

Begum Karaoglu

Ankara University Biotechnology Institute

Prof. Bala Gur Dedeoglu

Breast cancer is the most common type of cancer among women worldwide. This highlights the importance of discovering new and effective biomarkers that can be used for the diagnosis and treatment of the disease. It is believed that determining the interactions between molecules involved in gene expression regulation (such as transcription factors, miRNAs, and lncRNAs) will contribute to solving the complex processes involved in breast cancer development through a systems biology approach.

For this purpose, expression data from 5 Next Generation Sequencing (NGS) studies containing 179 tumors and 51 normal tissues from invasive ductal breast cancer (IDC) patients, as well as 8 microarray studies with 490 tumors and 122 normal tissues, were obtained from the GEO database. Following differential expression analyses, the expression data of mRNA, miRNA, lncRNA, and TF regulatory molecules were loaded into the circuit analysis program, miRCuit, developed within the scope of this thesis and analyzed.

The analysis results identified lncRNA CASC15, miR-130b-3p, and the transcription factor KLF5 as significant regulatory elements in the development of breast cancer. It is believed that the circuits developed in this study will help understand the effects of expression changes between regulatory molecules and provide a comprehensive perspective on the tumor development process. Moreover, the developed program will allow researchers from different fields to create regulatory circuits.

2025, 165 pages

Keywords: breast cancer; gene regulation; transcription factor; miRNA; lncRNA.

TEŞEKKÜR

Tezimin her aşamasında bilgi, deneyim ve önerileriyle bana rehberlik eden, her zaman titizlikle ve büyük bir özenle gerçekleştirdiğimiz çalışmalara akademik bilgi birikimiyle değer katan, karşılaştığım zorlukları aşmamda gösterdiği anlayış ve teşvik edici tutumuyla bana olan güvenini ve desteğini her an hissettiren danışman hocam Prof. Dr. Bala Gür Dedeoğlu'na sonsuz teşekkürlerimi sunarım.

TÜBİTAK “2244 Sanayi Doktora Programı” kapsamında 118C150 proje numarası ile desteklenen tez projemin sürdürülmesine katkı sağlayan Bilim İnsanı Destek Programları Başkanlığı (BİDEB) yönetimine ve programın diğer paydaşı İntergen Genetik Hastalıklar Değerlendirme Merkezi yöneticileri; değerli hocalarım Prof. Dr. Serdar Ceylaner ve Doç. Dr. Gülay Ceylaner'e teşekkürlerimi borç bilirim.

Lisans öğrenciliği yıllarımda staj günlerimden başlayarak doktoramı tamamladığım bu günlere kadar akademik yaşantımda ve özel hayatımda her zaman yanımda olan ve desteğini hiçbir zaman esirgemeyen değerli hocam Prof. Dr. Hilal Özdağ Sevgili'ye,

Çalışma ortamında sağladıkları kolaylıklar ve yardımlar ile değerli arkadaşlıklarını esirgemeyen ekip arkadaşlarım Senem Noyan'a, Kübra Nur Kaplan İlhan'a, Müge Öcal Demirtaş'a, Burcu Dolapçı Soydemir'e ve Ayşegül Yücel Polat'a,

Her zaman yanımda olan ve dostluklarıyla bana güç veren değerli arkadaşlarım Memet Gözübüyük'e, Beril Erdem'e ve B. Seçil Durel Avcıoğlu'na,

Bu anlamlı ve güzel süreçte bana sonsuz sabır ve anlayış gösteren, sevgileri, destekleri ve güvenleri ile her zaman yanımda olan, bana güç veren aileme; her zaman desteğini hissettiğim anneme, rol modelim olan babama, neşesiyle moral kaynağım olan kardeşime ve hayattaki en büyük şansım olan nişanlıma,

Bu yolculukta bana destek oldukları için sonsuz teşekkürlerimi sunarım.

İÇİNDEKİLER

ETİK BEYAN	i
ÖZET	iii
ABSTRACT	iv
TEŞEKKÜR.....	v
İÇİNDEKİLER.....	vi
ŞEKİLLER DİZİNİ.....	ix
ÇİZELGELER DİZİNİ.....	xii
SİMGELER DİZİNİ	xiii
1. GİRİŞ	1
2. KURAMSAL TEMELLER.....	2
2.1. MEME KANSERİ	2
2.1.1. HİSTOLOJİK SINIFLANDIRMA	3
2.1.2. PATOLOJİK SINIFLANDIRMA	6
2.1.3. MOLEKÜLER SINIFLANDIRMA	7
2.2. GEN İFADESİNİN REGÜLASYONU	10
2.2.1. TRANSKRİPSİYON FAKTÖRLERİ	10
2.2.2. MİKRO RNA’LAR	11
2.2.3. UZUN KODLAMAYAN RNA’LAR	13
2.3. BİYOLOJİK DEVRELER VE KANSERLEŞME SÜRECİNİN AYDINLATILMASINDA BİYOLOJİK DEVRELERİN ROLÜ	17
3. GEREKÇE VE AMAC.....	21

4. MATERYAL VE YÖNTEM..... 22

4.1. MATERYAL 22

4.2. YÖNTEM 24

4.2.1. VERİ SETLERİNİN OLUŞTURULMASI 25

4.2.2. VERİ SETLERİNİN ANALİZİ 26

4.2.2.1. Mikrodizin Verilerinin Analizi 26

4.2.2.2. Yeni Nesil Dizileme Verilerinin Analizi 30

4.2.3. DEVRE ANALİZ PROGRAMININ GELİŞTİRİLMESİ 34

4.2.3.1. Veri Tabanlarının Belirlenmesi 34

4.2.3.2. Veri Tabanlarından Elde Edilen Listelerin Programa Entegre Edilecek Şekilde Düzenlenmesi, İsimlendirme Sorunlarının Giderilmesi ve Kütüphanelerin Hazırlanması . 37

4.2.3.3. Kütüphanelerin Devre Analiz Programına Entegrasyonu ve Analiz Edilecek Dosyaların Programa Aktarılması Yoluyla Teorik ve Gerçek İkili İlişkilerin Belirlenmesi 38

4.2.3.4. Düzenleyici Devrelerin Oluşturulması 39

4.2.3.5. GSEA Zenginleştirme Analizi..... 42

4.2.4. DEVRE ANALİZ PROGRAMININ ARAYÜZ HALİNE GETİRİLMESİ..... 44

4.2.5. TCGA VERİ ANALİZİ..... 47

5. ARAŞTIRMA BULGULARI..... 48

5.1. MİKRODİZİN VERİ SETLERİNİN DİFERANSİYEL GEN İFADE ANALİZLERİ VE İSTATİSTİKSEL ANLAMLILIK DEĞERLENDİRMESİ 48

5.2. YENİ NESİL DİZİLEME VERİ SETLERİNİN DİFERANSİYEL GEN İFADE ANALİZLERİ VE İSTATİSTİKSEL ANLAMLILIK DEĞERLENDİRMESİ 49

5.3. MİKRODİZİN VE YENİ NESİL DİZİLEME VERİLERİ KULLANILARAK DEVRELERİN OLUŞTURULMASI	50
5.4. MİKRODİZİN VE YENİ NESİL DİZİLEME VERİLERİ KULLANILARAK OLUŞTURULAN DEVRELERİN GSEA ANALİZİ	53
5.5. MİKRODİZİN VE YENİ NESİL DİZİLEME VERİLERİNDEN ELDE EDİLEN SONUÇLARIN KARŞILAŞTIRILMASI.....	58
5.6. TCGA VERİ ANALİZİ	60
<u>6. TARTIŞMA.....</u>	<u>63</u>
<u>KAYNAKLAR.....</u>	<u>72</u>
<u>EKLER.....</u>	<u>92</u>
EK 1: VERİ SETLERİNE AİT TÜM NORMALİZASYON ÖNCESİ VE SONRASI BAR GRAFİKLERİ	92
EK 2: NGS VERİLERİNDEN ELDE EDİLEN MİRNA BAĞIMLI LNCRNA DÜZENLEYİCİ DEVRELER İLE TF BAĞIMLI LNCRNA DÜZENLEYİCİ DEVRELER.....	95
<u>ÖZGEÇMİŞ.....</u>	<u>129</u>
<u>TEZDEN ÇIKAN YAYINLAR.....</u>	<u>130</u>
<u>TEZDEN ÇIKAN BİLDİRİLER</u>	<u>142</u>

ŞEKİLLER DİZİNİ

Şekil 2.1. Her iki cinsiyet için meme kanserinin görülme sıklığı ve ölüm oranları.	2
Şekil 2.2. Meme dokusunun bölümlerini gösteren şematik çizim.....	3
Şekil 2.3. Meme kanserinin histolojik sınıflandırılması.....	5
Şekil 2.4. Meme kanserinin patolojik sınıflandırılması.....	6
Şekil 2.5. Meme kanserinin moleküler sınıflandırılması.....	9
Şekil 2.6. Transkripsiyon ön başlama kompleksinin oluşması.....	11
Şekil 2.7. miRNA biyogenezi.....	12
Şekil 2.8. lncRNA biyogenezi.	14
Şekil 2.9. lncRNA'ların hücrede katıldığı epigenetik, transkripsiyonel ve post-transkripsiyonel süreçler	16
Şekil 2.10. miRNA ve lncRNA'ların gen ifadesinin düzenlenmesindeki rollerinin şematize edilmiş örnek gösterimi.	17
Şekil 2.11. Farklı biyolojik devrelerin şematik gösterimi.	18
Şekil 2.12. mRNA-miRNA-lncRNA-TF düzenleyici devrelerinin ve devre elemanlarının birbirleri ile etkileşimlerinin şematik gösterimi.	19
Şekil 4.1. Verilerin elde edilmesinden regülatif devrelerin oluşturulmasına kadar izlenen basamakları gösteren iş akış şeması.....	24
Şekil 4.2. GSE73540 veri setine ait verilerin RMA normalizasyon öncesi ve sonrası bar grafikleri.	27
Şekil 4.3. GSE59247 veri setine ait verilerin quantile normalizasyon öncesi ve sonrası bar grafikleri.	27
Şekil 4.4. GSE73540 veri setinin diferansiyel gen ifade analizi.	28

Şekil 4.5. GSE59247 veri setinin diferansiyel gen ifade analizi.	29
Şekil 4.6. GSE71651 kodlu çalışmanın SRR2148235 kodlu hastasının ileri ve geri yönlü okumalarının kalite kontrol istatistikleri ve grafikleri.	31
Şekil 4.7. GSE71651 kodlu çalışmaya ait örnek feature count matrix görüntüsü.....	32
Şekil 4.8. GSE29173 kodlu çalışmaya ait örnek feature count matrix görüntüsü.....	32
Şekil 4.9. GSE71651 kodlu çalışmanın diferansiyel ifade analizi ve anotasyonu.	34
Şekil 4.10. lncRNA-miRNA, miRNA-mRNA ve TF-mRNA kütüphanelerinin örnek gösterimi.	38
Şekil 4.11. Kütüphanelerin devre analiz programına entegrasyonu ile teorik ve gerçek ikili ilişkilerin belirlenmesine yönelik şematik gösterim.	39
Şekil 4.12. Geliştirilen devre analiz programının kullanıcı arayüzü.	45
Şekil 4.13. Oluşturulan devrelerin txt formatındaki yapısı ve bu devrelerin csv formatında tablollaştırılmış versiyonlarına ait örnek ekran görüntüsü.	46
Şekil 5.1. Mikrodizin analizi sonucunda elde edilen farklı devre türlerine ait örnek gösterim.....	51
Şekil 5.2.NGS analizi sonucunda elde edilen farklı devre türlerine ait örnek gösterim 1...	52
Şekil 5.3. NGS analizi sonucunda elde edilen farklı devre türlerine ait örnek gösterim 2..	52
Şekil 5.4. NGS analizi sonucunda elde edilen farklı devre türlerine ait örnek gösterim 3..	53
Şekil 5.5. Mikrodizin verileri kullanarak oluşturulan devrelerin GSEA yöntemi ile analizi sonucunda zenginleşen yollar.	54
Şekil 5.6. Mikrodizin verileri kullanılarak oluşturulan devrelerin GSEA yöntemi ile analizi sonucunda zenginleşen yolların nokta grafiği.	55
Şekil 5.7. Mikrodizin verileri kullanılarak oluşturulan devrelerin GSEA yöntemi ile analizi sonucunda zenginleşen yollara ilişkin ağ yapısı.	56

Şekil 5.8. NGS verileri kullanılarak oluşturulan devrelerin GSEA yöntemi ile analizi sonucunda zenginleşen yollar.	57
Şekil 5.9. NGS verileri kullanılarak oluşturulan devrelerin GSEA yöntemi ile analizi sonucunda zenginleşen yolların nokta grafiği.	57
Şekil 5.10. NGS verileri kullanılarak oluşturulan devrelerin GSEA yöntemi ile analizi sonucunda zenginleşen yollara ilişkin ağ yapısı.	58
Şekil 5.11. Mikrodizin ve NGS veri setlerinin DEG analizleri sonucunda anlamlı ifade değişikliği gösteren mRNA, miRNA, lncRNA ve TF listelerinin karşılaştırılması.	59
Şekil 5.12. Gen Seti Zenginleştirme Analizi (GSEA) sonucunda belirlenen yolların mikrodizin ve NGS devreleri arasında karşılaştırılması.	60
Şekil 5.13. <i>miRCuit</i> tarafından oluşturulan CASC15 bağımlı düzenleyici devreler.	61
Şekil 5.14. TCGA'den elde edilen 1091 meme tümörü ve 105 normal doku örneğinde CASC15, KLF5, RASD1, CD55, DSC3, SOCS2, SORBS2 ve miR-130b-3p'nin ifadelerinin değerlendirilmesi.	62

ÇİZELGELER DİZİNİ

Çizelge 4.1. Veri setlerinin özellikleri.....	23
Çizelge 4.2. Devre analiz programında kullanılan veri tabanları ve moleküller arası etkileşim bilgileri.	36
Çizelge 5.1. Mikrodizin verilerinin analizi sonucunda elde edilen diferansiyel olarak anlamlı ifade gösteren moleküller.	49
Çizelge 5.2. Yeni Nesil Dizileme verilerinin analizi sonucunda elde edilen diferansiyel olarak anlamlı ifade gösteren moleküller.	50



SİMGELER DİZİNİ

BL-1	Basal-like 1 (Bazal hücre benzeri 1)
BL-2	Basal-like 2 (Bazal hücre benzeri 2)
CL	Claudin-low (Klaudin-düşük)
DCIS	Duktal karsinoma <i>in situ</i>
EGFR	Epidermal Growth Factor Receptor (Epidermal Büyüme Faktörü Reseptörü)
ER	Östrojen Reseptörü
GEO	Gene Expression Omnibus
GLOBOCAN	The Global Cancer Observatory (Küresel Kanser Gözlemevi)
GSEA	Gene Set Enrichment Analysis (Gen Seti Zenginleştirme Analizi)
HER	Human Epidermal Growth Receptor (İnsan epidermal büyüme reseptörü)
IARC	International Agency for Research on Cancer (Uluslararası Kanser Araştırmaları Ajansı)
IDC	Invasive ductal carcinoma (İnvaziv duktal karsinom)
IHC	Immunohistokimyasal
IM	İmmünomodülatör
LAR	Luminal Androjen Reseptörü
LCIS	Lobüler karsinoma <i>in situ</i>
lncRNA	Long non-coding RNA (Uzun kodlamayan RNA)
M	Mesenchymal
miRNA	Mikro RNA
MSigDB	The Molecular Signatures Database (Moleküler İmza Veri Tabanı)
MSL	Mesenchymal stemcell like (Mezenkimal kök hücre benzeri)
ncRNA	Non-coding RNA (Kodlanmayan RNA)
NCBI	National Library of Medicine

NGS	Next Generation Sequencing (Yeni Nesil Dizileme)
PIC	Preinitiation Complex (Ön başlatma kompleksi)
PR	Progesteron Reseptörü
RMA	Robust Multi-array Average (RMA)
SRA	Sequence Read Archive
TF	Transkripsiyon Faktörü
TNBC	Triple Negative Breast Cancer (Üçlü Negatif Meme Kanseri)
TNM	Tümör Nod Metastaz
WHO	World Health Organization (Dünya Sağlık Örgütü)



1. GİRİŞ

Meme kanseri kadınlarda en sık görülen kanser türüdür ve günümüzde kadınlar arasında kansere bağlı ölümlerde ilk sırada yer almaktadır (1). Oldukça karmaşık ve heterojen bir genetik hastalık olan meme kanserinin oluşumu ve gelişimi altında yatan mekanizmanın doğru anlaşılması, hastalığın prognozunu tahmini ve doğru tedavi yaklaşımlarının geliştirilmesi için son derece önemlidir (2,3). Günümüzde dizileme teknolojilerinin gelişmesi ve biyoinformatik alanındaki ilerlemeler ile meme kanserinin heterojen yapısını anlamaya katkı sağlayan pek çok çalışma yapılmış fakat hastalığın moleküler mekanizması henüz tam olarak aydınlatılmamıştır (2,4).

Hücrede biyolojik süreçlerin doğru ilerleyebilmesi için gen ifadesinin düzenlenmesi pek çok basamakta sıkı bir şekilde kontrol edilmektedir (5). TF'ler transkripsiyonel düzeyde gen ifadesinin düzenlenmesinden sorumlu temel moleküller iken, post-transkripsiyonel düzeyde düzenleme daha çok miRNA'lar tarafından gerçekleşmektedir (6,7). Son yıllarda artan çalışmalar, lncRNA'ların hem transkripsiyonel hem de post-transkripsiyonel seviyelerde düzenlemede görev alan önemli düzenleyici RNA'lar olduklarını ortaya koymaktadır (8–10). Sistem biyolojisi yaklaşımı ile moleküllerin birbirleri ile ilişkisinin ortaya koyulması, meme kanseri gibi karmaşık süreçlerin aydınlatılmasında önem kazanmaktadır (11–13).

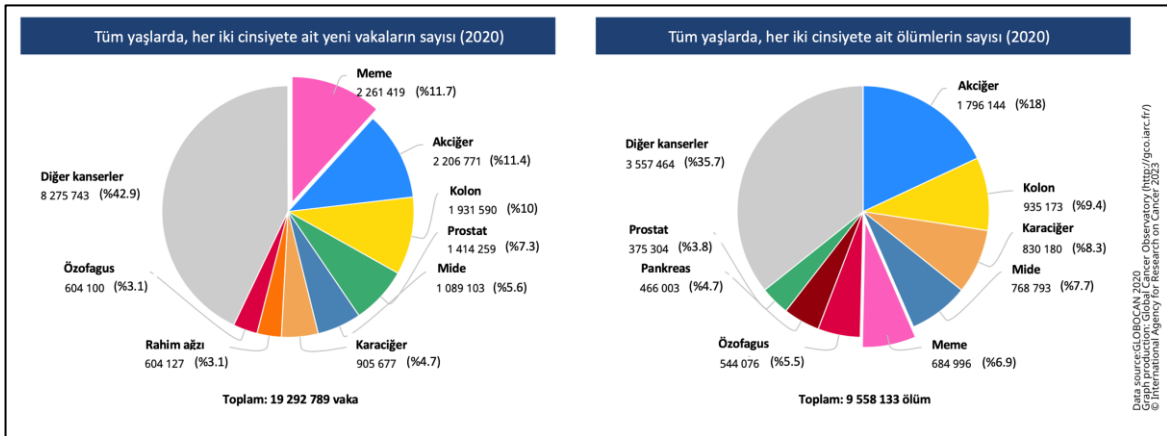
Meme kanseri ve normal dokuların lncRNA, miRNA ve mRNA ifade profil verileri kullanılarak meme kanseri oluşumunda ve gelişiminde rol oynayan düzenleyici lncRNA-TF-mRNA-miRNA devrelerini kurgulayacak bir program geliştirilmesi, moleküller arasında meydana gelecek bir ifade değişiminin sürece olan etkilerini ortaya koyacaktır. Bu yaklaşım ile kanserleşme sürecinin bütüncül bir yaklaşımla ele alınmasının sağlanması yeni biyobelirteçlerin önerilmesine ve etkin tedavi yaklaşımlarının geliştirilmesine olanak sağlayacaktır (4,14–16).

2. KURAMSAL TEMELLER

Meme kanseri pek çok faktör tarafından kontrol edilen kompleks, heterojen bir genetik hastalıktır ve dünya genelinde kadınlarda en çok görülen kanser türüdür (2,3). Farklı biyolojik özellikler, tedavi yanıtları ve prognostik modellere sahip birçok alt tipe ayrılan meme kanseri, günümüzde erken teşhis olanakları ve tedavideki önemli ilerlemelere rağmen halen kansere bağlı ölümlerde ilk sıralarda yer almaktadır (2,3). Söz konusu durum, hastalığın tanı ve tedavisinde kullanılabilecek yeni ve etkin biyomoleküllerin keşfinin önemini ortaya koymaktadır. Bunun için, gen ifadesinin kontrolünde görev alan mekanizmaların anlaşılması ve süreçte rol alan moleküllerin birbirleri ile etkileşimlerinin belirlenmesi, yeni biyomoleküllerin keşfine katkı sağlamaktadır (14–17).

2.1. MEME KANSERİ

Meme kanseri dünya çapında kadınları en çok etkileyen kanser türüdür ve kansere bağlı ölümlerde ilk sırada yer almaktadır (18). Erkeklerde görülen meme kanseri oranı ise, teşhisi gerçekleşen meme kanserlerinin %1'inden azını oluşturmaktadır (19). Uluslararası Kanser Araştırmaları Ajansı'nın (International Agency for Research on Cancer - IARC) sunmuş olduğu verilere göre her yıl dünya çapında toplam 2,26 milyon yeni vaka görülmekte, bu vakaların yaklaşık 680.000'i ise ölüm ile sonuçlanmaktadır (20). Son verilere ilişkin kadın ve erkeklerde meme kanseri görülme sıklığı ve ölüm oranları Şekil 2.1.'de verilmiştir (1).



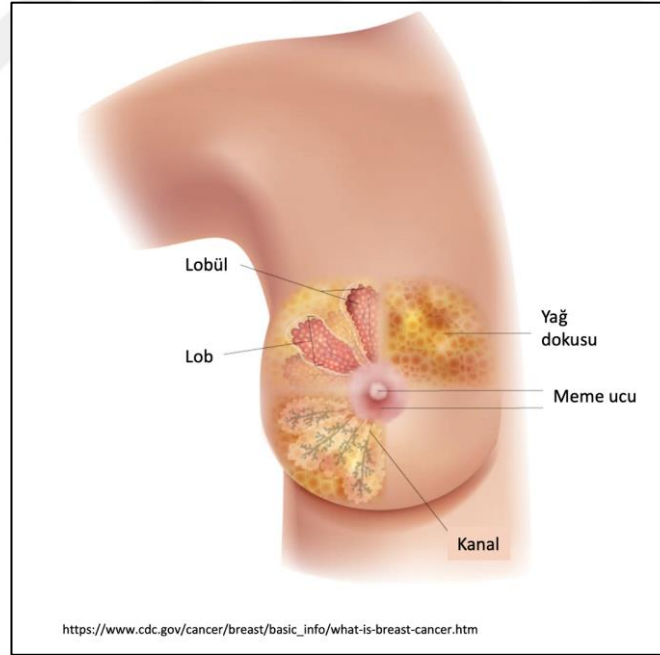
Şekil 2.1. Her iki cinsiyet için meme kanserinin görülme sıklığı ve ölüm oranları (1).

Dünya Sağlık Örgütü'nün (WHO) 2023 yılında elde ettiği son veriler, her yıl toplamda 2,3 milyondan fazla meme kanseri vakasının görüldüğünü ortaya koymaktadır. Bu oran yıllık kanser vakalarının %12,5'ini oluşturmaktadır (18,21).

Meme kanseri farklı morfolojik, biyolojik ve klinik özelliklere sahip heterojen bir hastalıktır. Bu heterojenlik tümörlerin farklı biyolojik davranışlar sergilemesine ve buna bağlı olarak farklı tedavi yanıtları geliştirmesine neden olmaktadır. Yüksek verimli moleküler tekniklerin ve biyoinformatik araçların gelişmesi meme kanserinin daha iyi anlaşılmasına olanak sağlamıştır (22). Bu kapsamda meme kanseri belirli özellikler göz önünde bulundurularak histolojik, patolojik ve moleküler alt sınıflara ayrılmaktadır (22,23).

2.1.1. HİSTOLOJİK SINIFLANDIRMA

Meme dokusu üç ana bölümden oluşmaktadır. Bunlar süt üreten bezler olan lobüller (lobul), sütü meme ucunda taşıyan kanallar (duct) ve tüm dokuyu çevreleyen bağ dokudur (Şekil 2.2.) (24).



Şekil 2.2. Meme dokusunun bölümlerini gösteren şematik çizim (24).

Meme kanseri histolojik olarak temelde üç sınıfa ayrılmaktadır. Bu sınıflar, tümör hücrelerinin proliferasyon yeteneklerinin meme lobülleri veya kanalları ile sınırlı olup olmadığı veya tümörü çevreleyen stromayı istila edip etmediği durumlara göre *in situ*,

mikroinvaziv veya invaziv olarak adlandırılmakta; tümörün loblarda veya kanallarda gelişmesine göre ise, lobüler veya duktal olarak daha alt sınıflara gruplandırılmaktadır. Bu gruplar arasındaki ayrımı yapmak için tümörün büyüme paternleri ve sitolojik özellikleri kullanılmaktadır (24–26).

In situ (non-invaziv) karsinomda tümör hücrelerinin proliferasyonu loblar veya kanallar ile sınırlıdır ve stromal invazyon bulunmamaktadır. *In situ* meme kanseri, tümörün geliştiği yere göre lobüler karsinoma *in situ* (LCIS) ve duktal karsinoma *in situ* (DCIS) olmak üzere iki gruba ayrılmaktadır. Ayrıca DCIS, farklı histolojik yapıları içeren alt tiplere sahiptir. Mikroinvaziv karsinom, tümör çapının 1 mm'den az invazyon gösterdiği sınıftır ve bileşenleri kanallar ve lobüller olabilmektedir (26). Invaziv karsinom ise, mikroinvazyonu aşan invaziv bileşenlere sahip ve bu invaziv bileşenlere göre histolojik tipin belirlendiği karsinomdur. Invaziv duktal karsinom (IDC), meme kanseri alt tipleri içerisinde en yaygın görülen alt tiptir (27,28) ve bu invaziv bileşenlerinin histolojik yapısına göre farklı alt tiplere ayrılabilir (26,27). Ayrıca IDC, özel tip olmayan invaziv karsinom (Invasive Carcinoma of No Special Type, NST) olarak da bilinmektedir (27,28). Diğer invaziv karsinomlar özel tip olarak adlandırılmaktadır ve invaziv duktal karsinomdan farklı sitopatolojik ve histopatolojik özelliklere sahiptir. Invaziv lobular karsinom, tubular karsinom, mucinous karsinom, medullary karsinom ve invaziv micropapillary karsinom başta olmak üzere pek çok özel invaziv karsinom tipi bulunmaktadır ve Şekil 2.3.'te detaylı olarak verilmiştir (25,26).

HİSTOLOJİK SINIFLANDIRMA	
<i>In situ Karsinom</i>	
<i>Duktal karsinom in situ (DCIS)</i>	Papillary Cribriform-papillary Cribriform Solid Comedo
<i>Lobular karsinom in situ (LCIS)</i>	
<i>Mikroinvaziv Karsinom</i>	
<i>İnvaziv Karsinom</i>	
<i>İnvaziv duktal karsinom (IDC)</i>	Tubule forming Solid Scirrhus Diğer tipler
<i>Özel tipler</i>	İnvaziv lobular karsinom Tubular karsinom İnvaziv cribriform karsinom Mucinous karsinom Medullary karsinom Apocrine karsinom Metaplastik karsinom İnvaziv micropapillary karsinom Secretory karsinom Adenoid cystic karsinom Diğer

Şekil 2.3. Meme kanserinin histolojik sınıflandırılması.

Tüm alt sınıflarından bağımsız olarak invaziv meme kanserleri histolojik olarak derecelendirilmektedir (grade). En yaygın kullanılan derecelendirme sistemi Bloom ve Richardson (29) tarafından geliştirilen, Elston ve Ellis (30) tarafından değiştirilen sistemdir. Nottingham derecelendirme sistemi adı verilen bu sistemde tümör tübül oluşturma derecesi, nükleer pleomorfizm ve mitotik indeks parametrelerine göre değerlendirilerek puanlanmakta ve daha sonra parametreler birleştirilerek tümörün derecesine karar verilmektedir. Buna göre 1. derece tümörler en farklılaşmış, 3. derece tümörler ise en az farklılaşmıştır. Başka bir deyişle, tümör derecesi daha küçük olan hastalarda daha iyi bir klinik sonuç elde edilirken tümör derecesi ileri olan hastalarda erken nüks ve metastaz görülmektedir. Bu nedenle tümör derecesi önemli bir prognostik belirteçtir (27,31–33).

Meme kanseri aynı zamanda Amerikan Ortak Kanser Komitesi/Uluslararası Kanser Kontrol Birliği (The American Joint Committee on Cancer/Union for International Cancer Control International Cancer Control) tarafından yayınlanan Tümör Nod Metastaz (TNM) Sistemi'ne göre evrelendirilmektedir (28,34). Bu sistemde tümör boyutu, bölgesel lenf nodu durumu ve uzak metastatik bölgelere yayılım gibi hem klinik hem de patolojik parametreler kullanılmaktadır (27).

2.1.2. PATOLOJİK SINIFLANDIRMA

Meme kanserinin patolojik sınıflandırmasında geleneksel morfolojik sınıflandırma temel öneme sahip olmasına rağmen, günümüzde yüksek işlem kapasiteli teknikler ve biyoinformatiğin gelişmesi ile araştırmacıların ilgisi, meme kanserini farklı gruplara ayırmak ve etkin tedavi yolunu belirlemek için tek gen ve çoklu gen testleri kullanımına odaklanmaktadır (Şekil 2.4.) (22).

PATOLOJİK SINIFLANDIRMA	
Morfolojik Sınıflandırma	Tümör Farklılaşması
	Tümörün tipi
	Tümörün derecesi
	Hastalığın Derecesi
Moleküler Sınıflandırma	Tümörün evresi
	Tek Gen ile Yapılan Sınıflandırmalar
	Tedavi kararının alınmasında etkili olan : Östrojen reseptörü (ER) ve HER2+
	Prognostik süreçte etkili olan: Progesteron reseptörü (PR) ve ki67
	Ailesel yatkınlık genleri: BRCA1, BRCA2, PALB2
	Çoklu Gen ile Yapılan Sınıflandırmalar
	Başta luminal sınıf genleri ve ilişkili diğer genler
Global Gen İfadesi ve Genomik Sınıflandırma	Luminal, bazal ve HER2 ile zenginleşmiş alt tipler
	Mutasyon imzaları
	Transkriptomik ve genomik özelliklerin kombinasyonuna dayalı sınıflandırmalar

Şekil 2.4. Meme kanserinin patolojik sınıflandırılması.

Patolojik sınıflandırmada tümörün farklılaşmasının doğru belirlenmesi kritik öneme sahiptir. Bu farklılaşma, tümörün tipi (sitolojik özellikler, tümör büyüme paterni ve yapısı, salgı özellikleri ve stromal özellikleri) ve histolojik derecesi belirlenerek morfolojik olarak ölçülmektedir. Hastalığın evresinin belirlenmesi ise en yaygın kullanılan ve primer tümörün

büyüklüğüne, nodal tutulumuna ve metastaz durumuna bağlı olarak değerlendirilen TNM sistemine göre gerçekleştirilmektedir. Meme kanserinin evresi en önemli prognostik belirteçlerden biri olarak kabul edilmektedir (22,28,34).

Benzer morfoloji ve evrelerdeki tümörlerin klinik davranışları ve tedavi yanıtlarında görülen farklılıklar, mevcut klinik ve morfolojik parametrelerin yetersiz olduğunu ortaya koymaktadır (35). Bu durum primer tümörün moleküler özelliklerinin anlaşılmasının, hastalığın prognozunu ve tedavi sürecinin tahmin edilmesinde ne kadar önemli olduğunu göstermektedir. Kullanılabilecek moleküler belirteçler, genler ve gen ürünlerini içeren (RNA ve protein) moleküllerin tekli veya bir arada karşılaştırılması sonucunda ortaya çıkan profillemeye dayanmaktadır (36–38). Bu kapsamda meme kanseri hastalarının tanı çalışmalarının önemli bir parçasını, önemli prognostik belirteçler olan östrojen reseptörleri (ER) ve insan epidermal büyüme faktörü reseptörlerinin (HER2) ifadesi oluşturmaktadır ve tedavi sürecinde belirleyici oldukları bilinmektedir (39–42). Benzer şekilde progesteron reseptörleri (PR) ve hücrenin proliferasyonunu düzenleyen Ki-67 genlerinin de meme kanserinde önemli biyobelirteçler oldukları kanıtlanmıştır. Bu biyobelirteçlerin ifadeleri birbirini etkilemekte ve prognostik değerlendirme sırasında kombinasyonlar halinde yorum yapılmaktadır (43–47). Bunlara ek olarak ailesel yatkınlık genleri olan *BRCA1*, *BRCA2* ve *PALB2* genleri de moleküler sınıflandırma için kullanılmaktadır(22).

Tek gen ile sınıflandırmaların yanı sıra, belirli bir tümör alt grubuna ait seçilmiş bir gen setinin ifadesine dayanan prognostik çoklu gen imzaları tanımlanmıştır. Bu imzalar süreci tahmin edebilmeleri ve tek gen üzerinden yapılan değerlendirmeler yerine gen kombinasyonları kullanılarak sonuçlar ortaya koymaları bakımından önemlidir (48,49). Karmaşık tümör biyolojisinin altında yatan genetik bozuklukları yansıtan durumu tahmin etmek için belirtilen tüm tek gen yaklaşımları ve çoklu gen imzaları tek başına yeterli olmamakta, özellikle riskli hasta gruplarında diğer bütün parametreler birlikte değerlendirildiğinde tamamlayıcı bilgiler sağlamaktadır (22).

2.1.3. MOLEKÜLER SINIFLANDIRMA

Meme kanserinin biyolojik ve fenotipik heterojen bir yapıya sahip olması, hastaların farklı klinik tablolar sergilemesine ve farklı tedavi yanıtları vermesine neden olmaktadır. Bu nedenle günümüzde tümörün yalnızca morfolojik özelliklerine göre sınıflandırılması veya

klirik patolojik parametrelere (tümör boyutu, metastazı vb) göre değeriendirilmesi tümör patolojisinin anlaşılmasında, hastalığın prognozunun belirlenmesinde ve tedavi sürecinin doğru planlanmasında oldukça yetersiz kalmaktadır (23,50). Yapılan pek çok çalışma, meme kanserinin moleküler paternlerini bulmak üzerine yoğunlaşmıştır (8,51–54). Perou, Sorlie ve arkadaşlarının öncülüğünü üstlendiği çalışmalar doğrultusunda meme kanseri, gen ifade profillerinde gözlenen benzerliklere göre *luminal A*, *luminal B*, *HER2+* ve *üçlü negatif* olmak üzere dört farklı moleküler alt sınıfa ayrılmıştır (23,55).

Meme kanseri alt sınıflarının ayrımı östrojen reseptörleri (ER), progesteron reseptörleri (PR), insan epidermal büyüme faktörü reseptörleri (HER2) ve hücre proliferasyonunu düzenleyen Ki-67 genlerinin varlığına göre yapılmaktadır (27). Bu dört adet (ER/PR/HER2/Ki-67) biyobelirteçten oluşan immunohistokimyasal panelin (IHC), meme kanserinin moleküler alt tiplerinin belirlenmesinde etkili olduğu belirlenmiştir (23,27).

Meme kanserinin moleküler alt sınıflarından biri olan ve en yaygın olarak görülen luminal A, östrojen ve progesteron reseptörlerinin varlığı (ER+, PR+), epidermal büyüme faktörü reseptörlerinin (HER2) ise yokluğu ile karakterizedir. Hücre proliferasyon belirteci Ki-67 ise, luminal A alt tipi meme kanserlerinde düşük ekspresyon seviyesine sahiptir (Ki-67<%20). Luminal A'ya göre daha kötü prognoz sergileyen luminal B ise, luminal A'ya benzeyen ER+, PR+ ve HER2+/- ifadelerinin yanı sıra yüksek Ki-67 (Ki-67>%20) ifadesine sahiptir. Diğer bir alt sınıf olan HER2+ meme kanseri, yeni teşhis edilen hastaların yaklaşık %15-20'sinde görülmektedir. HER2 reseptörlerinin aşırı ifadesi ile karakterize olmakla birlikte yüksek Ki-67 ifadesine sahiptir. Bu alt tipte ER ve PR reseptörlerinin ifadesi bulunmamaktadır. Tüm meme kanseri vakalarının yaklaşık %10-20'sini oluşturan üçlü meme kanseri (TNBC), hormon reseptörlerinin hiçbirinin ifade olmamasına karşın en yüksek Ki-67 ifadesi (Ki-67>30) ile en agresif meme kanseri alt sınıfı olarak bilinmektedir. Bu agresif kanser türü içerisindeki heterojenliğin daha iyi anlaşılması için pek çok gen ifade analizi çalışması yapılmış ve üçlü negatif meme kanseri 7 farklı alt tipe ayrılmıştır: İmmün yanıtta farklılıklar gösteren bazal gen ifadelerine sahip ve CK 5/6/14 genleri ile epidermal büyüme faktörü reseptörlerini (EGFR) ifade eden ve oldukça agresif olan bazal hücre benzeri 1 (BL-1) ve bazal hücre benzeri 2 (BL-2), immünolojik sinyalleşme yollarında önemli değişimler gösteren immünomodülatör tip (IM), mezenkimal ve mezenkimal kök hücre tipleri (M ve MSL), klaudin proteinlerini düşük ifade eden klaudin-düşük tip (CL) ve

androjen metabolizmasında rol alan genlerin farklı ifadesini gösteren luminal androjen reseptör tipi (LAR) (23,27,56–58). Meme kanserinin moleküler alt tipleri ve bu alt tip ayrımını gösteren gen ifade profilleri Şekil 2.5.’te verilmiştir.

MOLEKÜLER SINIFLANDIRMA	
Alt Tipler	İfade edilen reseptörler / genler
Luminal A	ER+ PR+ HER2- Ki-67 düşük seviye
Luminal B	ER+ PR+ HER2+/- Ki 67 yüksek seviye
HER2+	ER- PR- HER2+ Ki-67 yüksek seviye
Üçlü Negatif (TNBC)	
<i>Bazal hücre benzeri-1 (BL-1)</i>	ER-
<i>Bazal hücre benzeri-2 (BL-2)</i>	PR-
<i>İmmünomodülatör (IM)</i>	HER2-
<i>Mezenkimal (M)</i>	Ki-67 yüksek seviye
<i>Mezenkimal kök hücre benzeri (MSL)</i>	
<i>Klaudin-düşük (CL)</i>	
<i>Luminal androjen reseptör (LAR)</i>	

Şekil 2.5. Meme kanserinin moleküler sınıflandırılması.

Bunlara ek olarak tümör biyolojisinin daha iyi anlaşılmasına duyulan ihtiyaç Oncotype DX, Prosigna PAM50 ve Mammaprint gibi daha detaylı moleküler analizlerin gelişmesine yol açmıştır ve detaylı moleküler analizlere olanak sağlamaktadır (59–61).

Günümüzde meme kanseri pek çok alt sınıfa ayrılmıştır ve bu alt sınıflara ait biyolojik özellikler iyi tanımlanmıştır. Ancak eldeki veriler hastalığın tanı ve tedavisinde doğru yaklaşımları belirlemede hala yetersiz kalmaktadır. Bu kapsamda oldukça heterojen olan ve pek çok genetik faktör tarafından kontrol edilen meme kanserini anlamada gen ifadesini kontrol eden mekanizmaların iyi anlaşılması büyük önem taşımaktadır.

2.2. GEN İFADESİNİN REGÜLASYONU

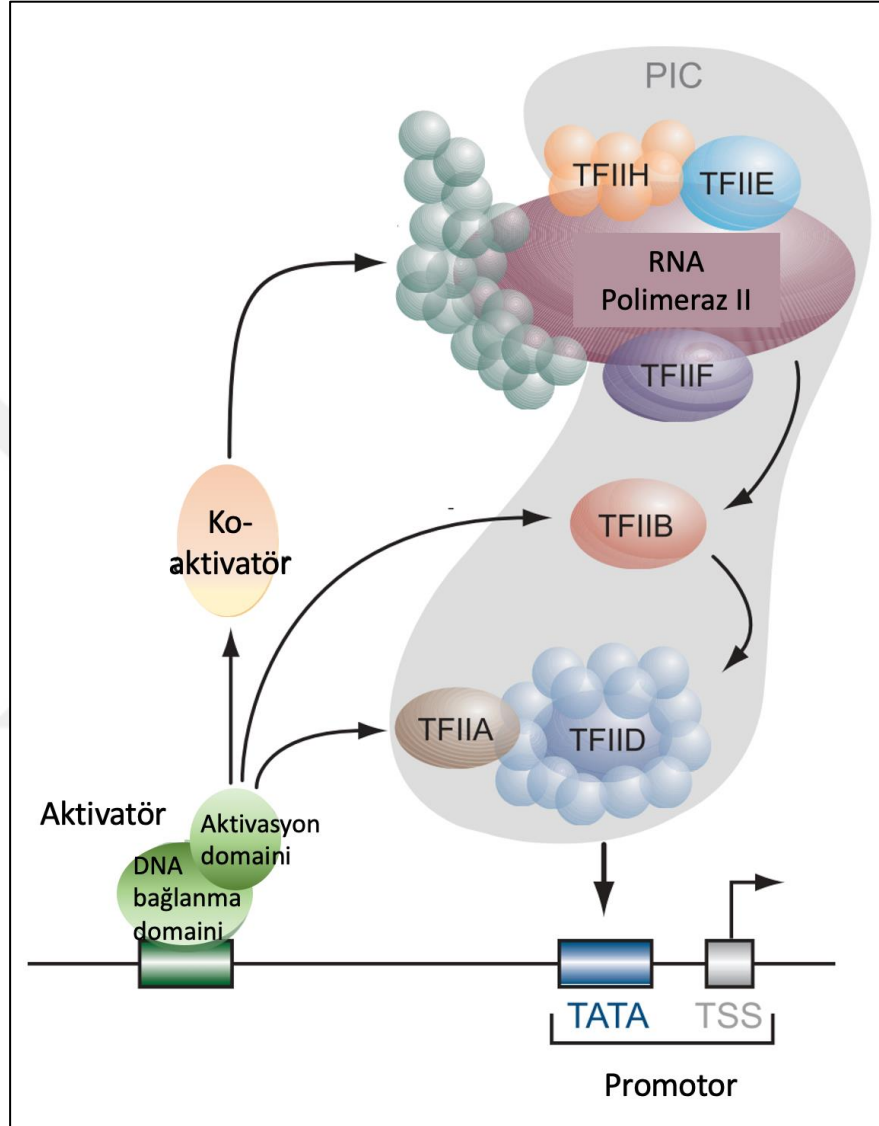
Gen ifadesi tüm organizmalar için gerekli olan karmaşık bir süreçtir. Bir gende kodlanan bilginin işlevsel bir ürüne dönüşmesinde önemli rol oynayan gen ifadesi, pek çok faktör tarafından dinamik bir şekilde düzenlenmektedir. Bu düzenleyici süreçler temelde transkripsiyonel ve post-transkripsiyonel (transkripsiyon sonrası) olmak üzere iki basamakta gerçekleşmektedir (5). Transkripsiyonel düzeyde düzenlemenin başlıca elemanları transkripsiyon faktörleri iken, post-transkripsiyonel düzeyde düzenleme temelde mikro RNA'lar tarafından kontrol edilmektedir (6,7). Genel olarak, transkripsiyon faktörleri gen ifade seviyelerini arttırırken miRNA'lar bu seviyeleri azaltmaktadır (62–65). Yüksek işlem kapasiteli tekniklerin gelişmesi ile keşfedilen ve hücredeki biyolojik görevleri ortaya konan uzun kodlamayan RNA'lar (lncRNA) ise, hem transkripsiyonel hem de post-transkripsiyonel düzenleyici süreçleri önemli ölçüde etkileyerek gen ifadesi düzenlenmesinin neredeyse her basamağında pozitif veya negatif etkiler gösterebilmektedir (8–10).

2.2.1. TRANSKRİPSİYON FAKTÖRLERİ

Transkripsiyon faktörleri (TF) belirli DNA dizilerine bağlanabilen ve gen ifadesini transkripsiyonel seviyede düzenleyen proteinlerdir (7). Transkripsiyon RNA polimeraz II enzimi tarafından gerçekleştirilmektedir. Transkripsiyon süreci ise TF'ler, aktivatörler, koaktivatörler ve baskılayıcılar tarafından düzenlenmektedir (66). Transkripsiyonun başlaması için gereken TF'ler, genel transkripsiyon faktörleri olarak adlandırılmaktadır (TFIIA, TFIIB, TFIID, TFIIIE, TFIIF ve TFIIH). Transkripsiyonun başlaması bu TF'lerin hedef geninin önünde yer alan ve transkripsiyonun bazal seviyede ve doğru şekilde gerçekleşmesi için gerekli olan promotor bölgelerindeki özgül DNA dizilerine bağlanması ile gerçekleşmektedir (67–69). Promotor bölgesinde düzenli bir şekilde bir araya gelen TF'ler transkripsiyon ön başlatma kompleksi (PIC) olarak adlandırılan bir kompleks oluşturmakta ve RNA polimeraz II'yi bu bölgeye çekerek transkripsiyonun başlamasını sağlamaktadır (70).

Transkripsiyonu düzenleyici elemanlardan olan aktivatörler, transkripsiyon başlama seviyesini arttırırken baskılayıcılar olarak adlandırılan diğer bir grup düzenleyici ise transkripsiyon seviyesini azaltmaktadır. Aktivatörler ve baskılayıcılar, çevresel uyarılara

karşı veya gelişimin farklı evrelerinde hedef genlerinin ifadelerini arttırmakta ya da azaltmaktadır. Koaktivatörler ise aktivatörler ile protein-protein kompleksi arasındaki etkileşimi sağlayarak gen ifadesinin düzenlenmesine katkıda bulunmaktadır (Şekil 2.6.) (66,71).

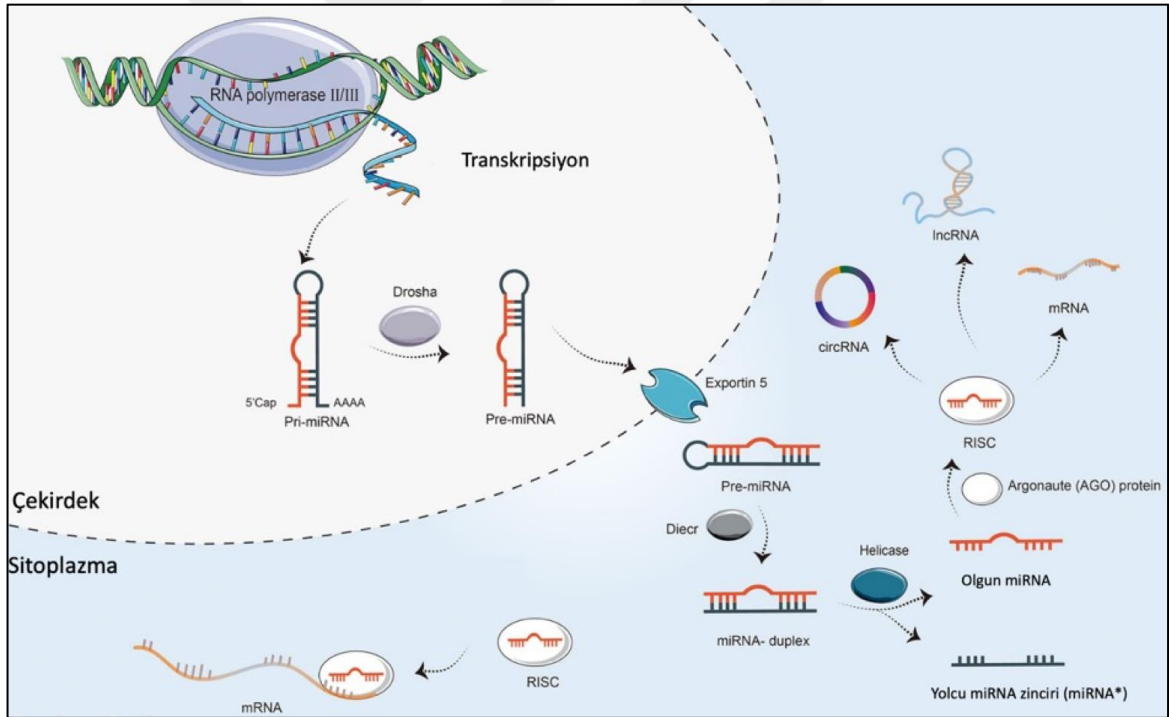


Şekil 2.6. Transkripsiyon ön başlama kompleksinin oluşması (Maston, 2006'dan uyarlanmıştır) (66).

2.2.2. MİKRO RNA'LAR

MikroRNA'lar (miRNA) 20-25 nükleotid uzunluğunda, türler arasında evrimsel olarak korunmuş, gen ifadesini post-transkripsiyonel seviyede kontrol eden küçük, protein kodlamayan RNA molekülleridir (72,73). RNA polimeraz II enzimi tarafından genomdaki

kendi miRNA genlerinden veya protein kodlayan genlerin intron bölgelerinden transkribe olmaktadır. Pri-miRNA olarak isimlendirilen bu birincil transkriptler saç tokası yapısındadır ve mRNA'ya benzer şekilde 5' şapka ve 3' poliA kuyruk içermektedir. Pri-miRNA'lar Drosha endonükleaz enzimi tarafından çekirdekte kesilerek yaklaşık 70 nükleotid uzunluğundaki pre-miRNA'ya dönüştürülmekte ve nükleer taşıma reseptörü exportin-5 tarafından sitoplazmaya taşınmaktadır. Dicer enzimi ile pre-miRNA'ların sitoplazmada kesilmesi sonucu saç tokası yapısı ortadan kaldırılarak yaklaşık 20 nükleotid uzunluğundaki miRNA/miRNA* dubleksleri oluşturulmaktadır. Bu dubleksten 5' daha stabil olan miRNA olgun miRNA, diğeri ise yolcu miRNA olarak adlandırılmaktadır ve yolcu ipliğin çoğunlukla yıkıldığı bilinmektedir. Olgun miRNA, RNA kaynaklı susturma kompleksine (RISC) dahil olmaktadır. miRNA'lar etkilerini hedeflerinin genellikle 3' UTR bölgelerine bağlanarak translasyonun baskılanması veya mRNA'nın degradasyonu şeklinde göstermektedir (Şekil 2.7.) (73,74).

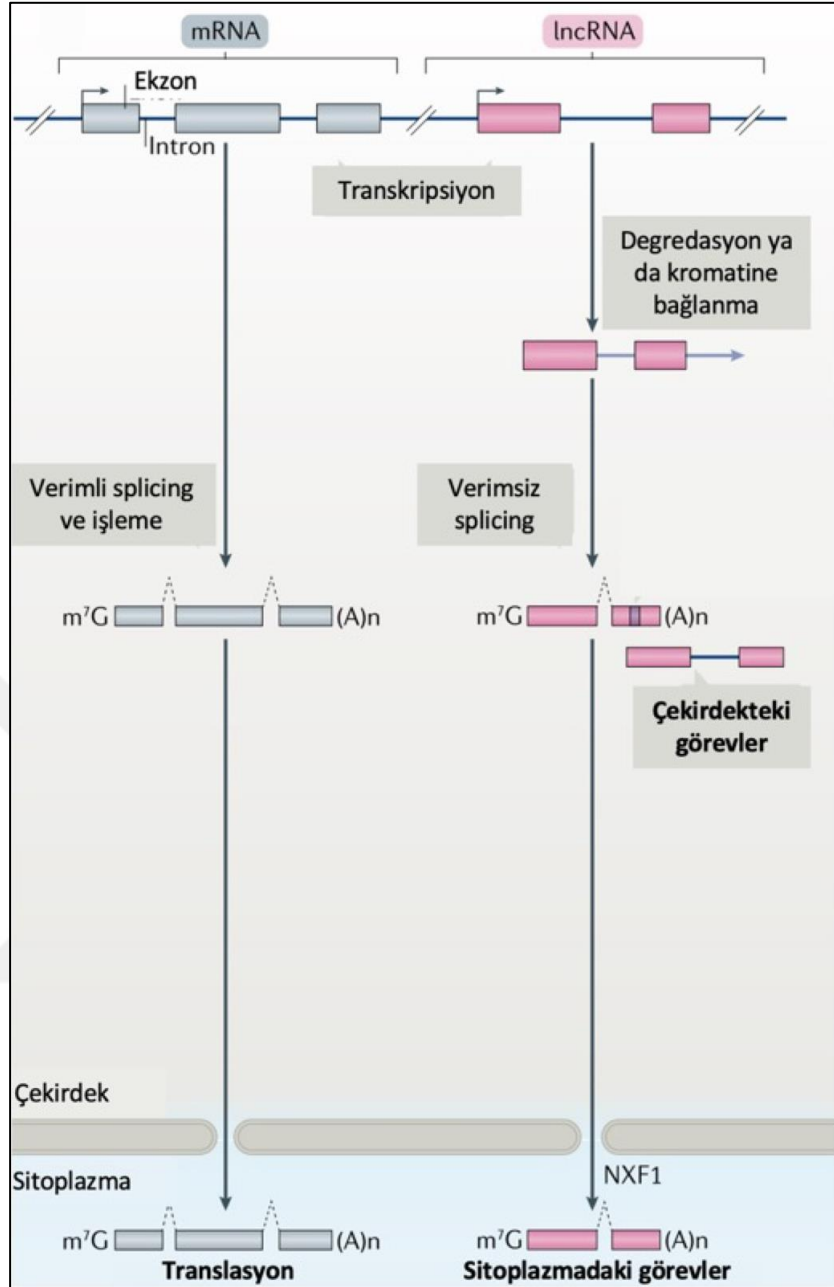


Şekil 2.7. miRNA biyogenezi (Jia, 2022'den uyarlanmıştır) (75).

miRNA'ların protein kodlayan genlerin yarısından fazlasının ifadesini kontrol ettiği ve pek çok hücre fonksiyonunun düzenlenmesine katkıda bulunduğu bilinmektedir (76). Ayrıca bir tümör baskılayıcı gen ya da onkogen olarak görev yaptığına ilişkin çalışmalar da bulunmaktadır (77,78).

2.2.3. UZUN KODLAMAYAN RNA'LAR

Teknolojinin ilerlemesi ve yüksek işlem kapasiteli dizileme teknolojilerinin gelişmesi ile hücrede önemli epigenetik düzenleyicilerden olan kodlamayan RNA'lar (ncRNA) tanımlanmıştır (13). Uzun kodlamayan RNA'lar (lncRNA), ncRNA'lar içerisinde yer alan önemli bir gruptur ve miRNA'lar gibi belirli hücrel süreçlerin ve gen ifadesinin düzenlenmesine katkısı olduğu bilinmektedir (75,79,80). LncRNA'lar 200 nükleotidden uzun, proteine çevrilmeyen RNA'lardır. mRNA'ya benzer şekilde RNA polimeraz II enzimi tarafından genomun intronik, intergenik, sense veya antisense zincirinden transkribe olmaktadır (9,79). Transkribe olduktan sonra daha az uç birleştirme sinyali (splicing) maruz kalmakta ve bu süreç mRNA'ya kıyasla daha az verimli şekilde gerçekleşmektedir (9). Transkriptlere mRNA'larda olduğu gibi 5'-7mG şapka ve 3'-poliA kuyruk eklenmektedir. LncRNA'lar çekirdekte sentezlendikten sonra bir kısmı çekirdekteki görevlerini yapmak için burada tutulurken bir kısmı RNA dışı aktarma faktörü (NXF1) tarafından sitoplazmaya taşınmaktadır (Şekil 2.8.) (9,81,82).



Şekil 2.8. lncRNA biyogenezi (Statello 2021’den uyarlanmıştır) (9).

Gen ifadesini hem transkripsiyonel hem de post-transkripsiyonel düzenlemeleri nedeniyle lncRNA’lar gen ifadesinin düzenlenmesinde anahtar role sahiptir. lncRNA’ların görev aldığı epigenetik, transkripsiyonel ve post-transkripsiyonel süreçler aşağıda özetlenmiştir (Şekil 2.9.) (83,84):

lncRNA’ların Görev Aldığı Epigenetik ve Transkripsiyonel Süreçler

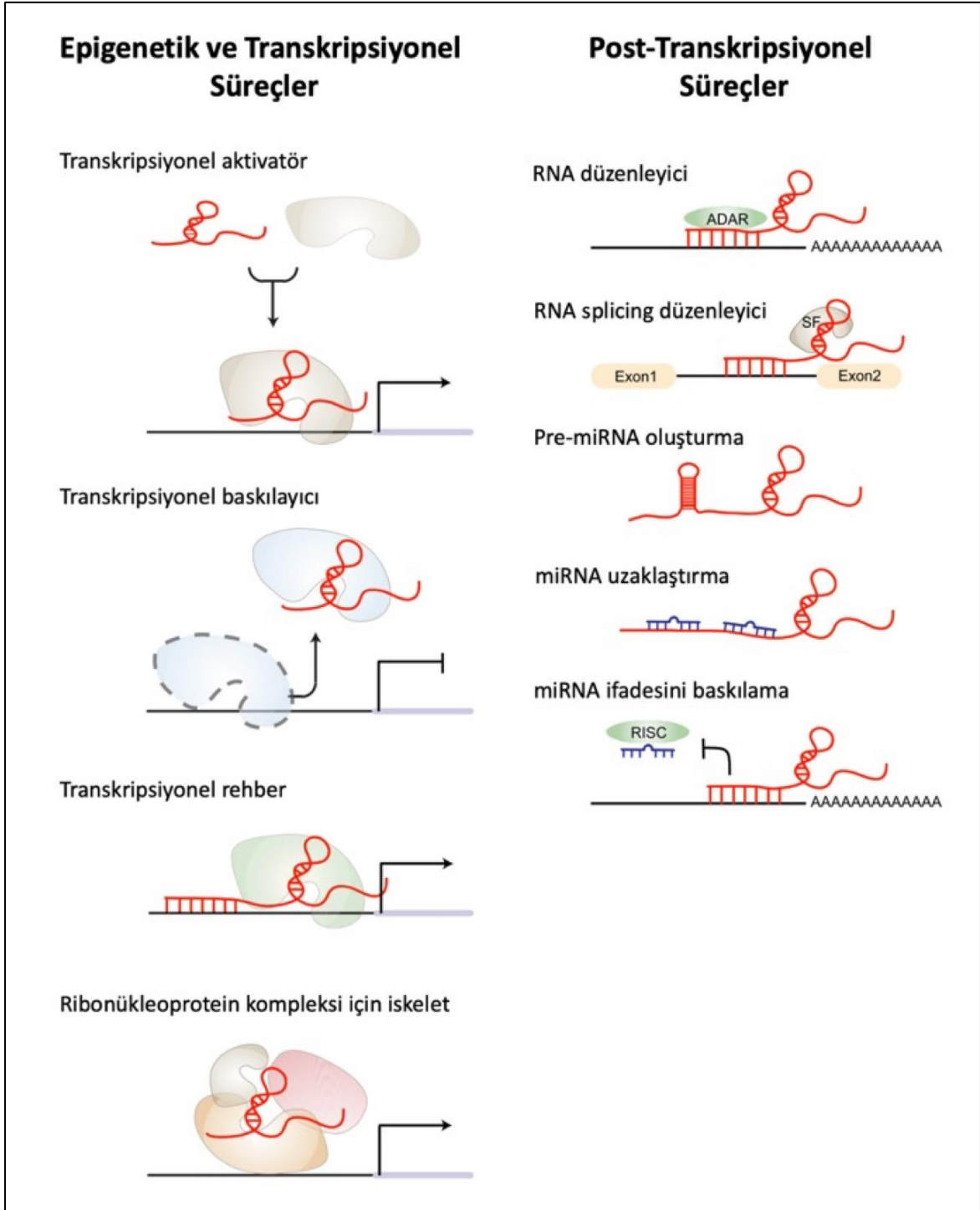
- Transkripsiyon faktörleri ile etkileşime girerek transkripsiyon faktörlerinin aktivasyonunun sağlanması,

- Transkripsiyon faktörlerine bağlanıp bu molekülleri genomdan uzaklaştırarak transkripsiyonel baskılanmanın sağlanması,
- Kromatin modifiye edici enzimleri hedef genlere yönlendirilerek rehber görevi görmesi,
- Genom üzerinde çoklu proteinleri bir araya getirerek ribonükleoprotein kompleksinin oluşmasını sağlayacak bir iskelet görevi görmesi.

lncRNA'ların Görev Aldığı Post-Transkripsiyonel Süreçler

- Eşleniği olan pre-mRNA'ya bağlanan lncRNA'nın RNA-RNA duplexi oluşturarak spesifik enzimlerin RNA'ya toplanmasını sağlaması ve RNA düzenlemesi (RNA editing) sürecine katkıda bulunması,
- Pre-mRNA'nın intron ve ekzonları arasına bağlanarak alternatif splicing mekanizmasının önlenmesi,
- Tam tersi şekilde, belirli splicing faktörlerini bu bölgeye taşıyarak alternatif splicing mekanizmasının aktifleştirilmesi,
- Yapısında bulundurduğu saç tokası yapısı ile pre-miRNA oluşumuna katkı sağlaması,
- Fonksiyonel miRNA'lar için tanıma bölgesi taşıyarak miRNA'ları hedefleri olan mRNA'dan uzaklaştırması ve gen ifadesinin artmasına katkıda bulunması,

mRNA'ların 3'-UTR bölgelerine bağlanmak için miRNA'lar ile rekabet ederek mRNA ifadesinde artışa katkıda bulunması (13).

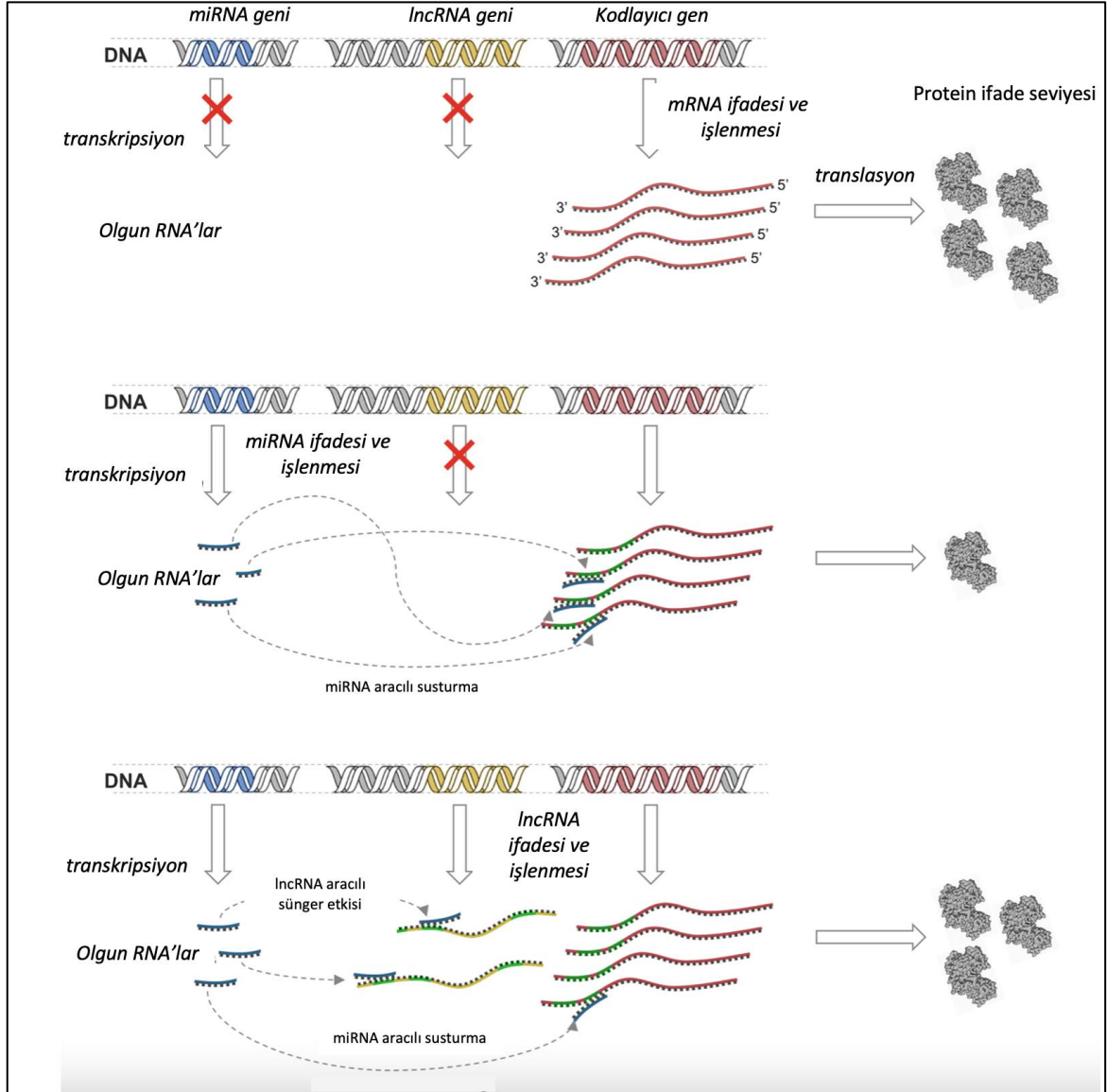


Şekil 2.9. lncRNA'ların hücrede katıldığı epigenetik, transkripsiyonel ve post-transkripsiyonel süreçler (Yang, 2014'ten uyarlanmıştır) (13).

lncRNA'lar gen ifadesinin düzenlenmesinin neredeyse her basamağında hem pozitif hem de negatif etki gösterebilen önemli biyolojik moleküllerdir. Bu nedenle, ifade farklılığı gösteren lncRNA'ların hücrede onkogen veya tümör baskılayıcı gibi davrandığı bilinmektedir (9,17,85).

2.3. BİYOLOJİK DEVRELER VE KANSERLEŞME SÜRECİNİN AYDINLATILMASINDA BİYOLOJİK DEVRELERİN ROLÜ

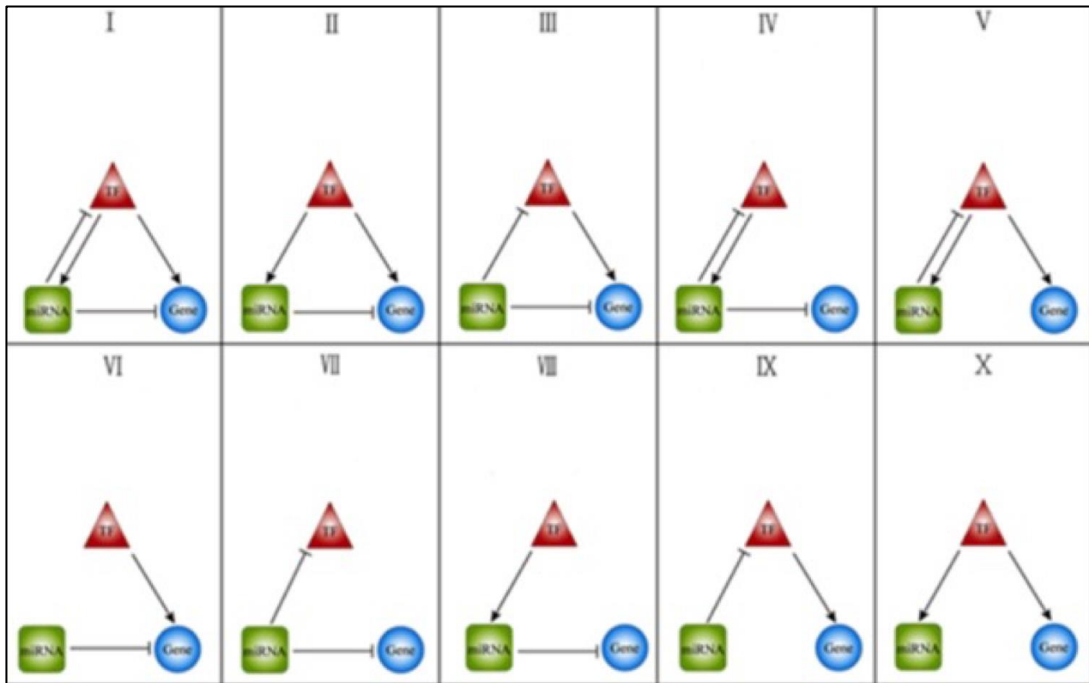
Dinamik biyolojik süreçlerde genlerin ifadesi pek çok faktör tarafından kontrol edilmektedir. Transkripsiyon faktörleri (TF), mikroRNA'lar (miRNA) ve uzun kodlanmayan RNA'lar (lncRNA) gen ifadesinin düzenlenmesinde ana düzenleyiciler olarak görev almaktadır. miRNA ve lncRNA aracılı gen ifadesinin düzenlenmesine ilişkin şematize edilmiş örnek Şekil 2.10.'da verilmiştir.



Şekil 2.10. miRNA ve lncRNA'ların gen ifadesinin düzenlenmesindeki rollerinin şematize edilmiş örnek gösterimi (Veneziano, 2019'dan uyarlanmıştır).

Bu düzenleyici moleküllerin ifadelerinde meydana gelen değişimler, diğer etkileşimde bulundukları moleküllerin ifadelerini değiştirebilmekte ve bu durum, meme kanseri de dahil olmak üzere çeşitli hastalıkların gelişimine neden olabilmektedir (11–13,86–89). Örneğin, transkripsiyon faktörleri meme kanseri hücrelerinin proliferasyonu, invazyonu ve migrasyonu gibi süreçlerinde kritik roller oynarken (90–92); miRNA'ların ifadesinin normal dokulara kıyasla tümör dokularında bozulduğu çeşitli kanserlerde ortaya konmuştur (93–95). Ayrıca miRNA'ların hedef genler üzerindeki baskılayıcı etkilerinin, meme kanseri hücrelerinde proliferasyon, invazyon ve metastaz gibi süreçlere katkıda bulunduğu bilinmektedir (96,97). Bunlara ek olarak, lncRNA'lar hücreSEL süreçlerde önemli düzenleyici roller oynamakta ve transkripsiyon faktörleri, miRNA'lar ve hedef genler ile etkileşime girerek onkojenik işlevler gösterebilmektedir (13,98–100).

Düzenleyici moleküller arasındaki bu ilişki biyolojik devreler (biyolojik düzenleyici devreler) olarak adlandırılmaktadır ve her molekülün birbiri ile arasındaki ilişkiyi (aktivasyon veya baskılama) ortaya koymaktadır. Biyolojik devreler üç veya daha fazla düzenleyici molekülden oluşabilmektedir. Biyolojik devreler, bu moleküllerin birbirlerini etkileme şekillerine göre açık devreler veya kapalı devreler olarak adlandırılmaktadır (Şekil 2.11) (101–106).



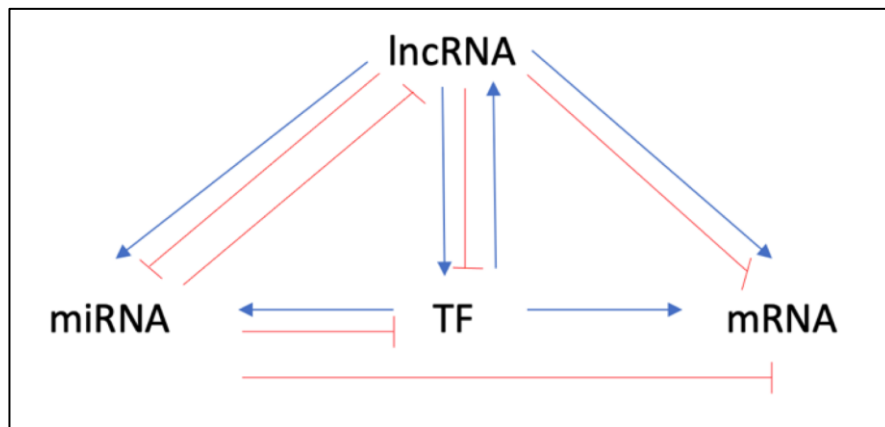
Şekil 2.11. Farklı biyolojik devrelerin şematik gösterimi (I, II, III kapalı devreleri; IV, V, VI, VII, VIII, IX, X açık devreleri göstermektedir).

Biyolojik devreler, sistem biyolojisi yaklaşımı ile bir genin ifadesinin hangi moleküller tarafından nasıl değiştiğini anlamak için oldukça önemlidir. Transkripsiyon faktörleri gen ifadesinin artması veya azalması şeklinde transkripsiyonel düzeyde görev yaparken, miRNA'lar ifadenin azalması yönünde post-transkripsiyonel seviyede görev almaktadır. TF'ler hem mRNA'ların hem de miRNA'ların transkripsiyonundan sorumlu moleküllerdir. Bunun yanı sıra, transkripsiyon sonucu oluşan miRNA'lar, TF'ler ile mRNA'ların ifadesinin azalmasına neden olmaktadır (107). Bu çerçeveden bakıldığında, bir TF – mRNA – miRNA'dan oluşan biyolojik devrede;

- TF, bir miRNA'nın ifadesini düzenliyor olabilir,
- miRNA, bir TF'i düzenliyor olabilir,
- TF ve miRNA molekülleri birlikte hedef genlerin ifadesini düzenliyor olabilir.

Hücrede pek çok düzenleyici göreve sahip olan lncRNA'ların TF-mRNA-miRNA düzenleyici devrelerindeki her bir molekül ile etkileşiminin olduğu bilinmektedir. Bu durumda, lncRNA'lar ile kurulacak yeni lncRNA-TF-mRNA-miRNA devrelerinde;

- lncRNA, bir TF'in hedef gene bağlanmasını arttırıyor veya bir TF'in hedef gene bağlanmasını baskılıyor olabilir,
- lncRNA, taşıdığı pre-miRNA yapısı sayesinde yeni miRNA oluşumuna katkıda bulunuyor veya bir miRNA'nın hedef gene bağlanmasını baskılıyor olabilir,
- lncRNA, mRNA'ya bağlanarak miRNA'ların bağlanmasını engelleme yolu ile genin ifadesini arttırıyor veya splicing faktörlerini hedef gene taşıyarak genin ifadesini arttırıyor olabilir (Şekil 2.12.).



Şekil 2.12. mRNA-miRNA-lncRNA-TF düzenleyici devrelerinin ve devre elemanlarının birbirleri ile etkileşimlerinin şematik gösterimi (Oklar pozitif regülasyonu, kesik uçlu oklar ise negatif regülasyonu temsil etmektedir).

Gen ifadesini d zenleyen bu molek ller arasında ortaya  ıkabilecek bir ifade deęiřimi, etkileřimdeki dięer molek llerin de ifadelerini deęiřtirerek kanserin de aralarında bulunduęu  eřitli hastalıkların oluřumuna neden olmaktadır (11–13,86–89). Bu nedenle, molek ler devrelerin ve aęların etkilerini anlamak hastalığın mekanizmasını a ıklamak ve hedefe y nelik yeni tedavi stratejilerinin geliřtirilmesini saęlamak amacıyla olduk a  nemlidir.



3. GEREKÇE VE AMAÇ

Gen ifadesi pek çok faktör tarafından düzenlenmektedir. Düzenleyici moleküllerin ifadesinde meydana gelen değişimin pek çok hastalık ile yakından ilişkilendirildiği bilinmektedir. Bu nedenle gen ifadesinin düzenlenmesinde rol oynayan moleküllerin ve bu moleküllerin birbirleri ile etkileşimlerinin ortaya koyulması pek çok açıdan önem taşımaktadır. Dinamik olan bu sürecin, kanser gibi heterojen ve çok faktörlü bir hastalığın oluşumunda ve gelişiminde rol alan moleküler mekanizmaların belirlenmesi sistem biyolojisi yaklaşımı ile bir bütün olarak araştırılması gerekmektedir. Bu çerçevede, lncRNA-TF-mRNA-miRNA biyolojik devrelerinin *in silico* olarak belirlenmesinin ve bu devreler kullanılarak biyolojik ağların oluşturulmasının, kanser oluşumu ve gelişiminde rol oynayan moleküler belirteçlerin ortaya konmasına katkı sağlayacağı düşünülmektedir.

Sistem biyolojisi yaklaşımı ile lncRNA ve miRNA'ların ifadelerinde meydana gelen değişimin TF'ler ve mRNA'lar üzerinden meme kanseri oluşumunda ve gelişiminde düzenleyici role sahip olduğu, doktora tez çalışmasının hipotezini oluşturmaktadır. Ayrıca tez çalışması kapsamında geliştirilecek olan programın, meme kanseri ve normal dokular arasında fark gösteren lncRNA-TF-mRNA-miRNA devreleri ile biyolojik ağlar kurgulayarak kanser tanı ve tedavisinde kullanılabilecek yeni biyobelirteçler önerebileceği, çalışmanın bir diğer hipotezidir.

Bu kapsamda, meme kanseri ve normal dokuların lncRNA, miRNA ve mRNA ifade profil verilerini kullanarak meme kanseri oluşumu ve gelişiminde rol alan düzenleyici lncRNA-TF-mRNA-miRNA devrelerini kurgulayacak bir program geliştirmek tez çalışmasının amacını oluşturmaktadır. Geliştirilen program, lncRNA ve miRNA'ların ifade değişimlerinin transkripsiyon faktörleri (TF'ler) ve mikro RNA'lar (miRNA'lar) üzerindeki düzenleyici rollerini ortaya çıkarmayı hedeflemektedir. Böylece meme kanserinde rol oynayan moleküllerin sistem biyolojisi yaklaşımıyla bütüncül şekilde ele alınması sağlanacaktır. Ayrıca programın yalnızca meme kanseri ile sınırlı kalmayıp farklı kanser gruplarında da benzer analizler yapabilecek şekilde tasarlanması ve böylece farklı disiplinlerden araştırmacıların kullanabileceği kapsamlı bir analiz aracı olarak sunulması, tez çalışmasının bir diğer amacını oluşturmaktadır.

4. MATERYAL VE YÖNTEM

4.1. MATERYAL

İnvaziv duktal meme kanseri (IDC), meme kanseri alt tipleri içerisinde en yaygın görülen alt tiptir (27,28). GEO veri tabanı kullanılarak yapılan kapsamlı araştırmada, yeterli sayıda hasta ve kontrol sayısına ulaşmak ve çalışmanın belirli bir alt tipe odaklanmasını sağlamak amacıyla, tez çalışmasında kullanılacak tümör ve normal/eşli normal dokuları invaziv duktal meme kanseri hastalarından seçilmiştir.

Tez çalışması kapsamında 179 invaziv duktal meme kanseri hastasından elde edilen tümör ve 51 normal/eşli normal dokularına ait 5 Yeni Nesil Dizileme (NGS) (108–112) ile 490 invaziv duktal meme kanseri hastasından elde edilen tümör ve 122 normal/eşli normal dokusuna ait 8 microarray (mikrodizin) (113–119) çalışması olmak üzere toplamda 13 adet veri seti kullanılmıştır (Çizelge 4.1.). Gene Expression Omnibus (GEO) veri tabanı kullanılarak yapılan kapsamlı araştırma sonucunda, IDC hastalarından taze dondurulmuş primer tümör ve normal/eşli normal dokulara ait mRNA/miRNA/lncRNA ifade verilerine ulaşılmıştır. GEO'dan elde edilen NGS ve mikrodizin verileri, analiz için geliştirilmiş özel paketler ve kütüphaneler kullanılarak Linux/Unix, R Studio ve Python ortamlarında analiz edilmiştir. Tümör ve normal dokular arasında anlamlı şekilde ifadesi değişen ve IDC oluşumunda rol oynayan regülatif devreleri bulmak amacıyla tasarlanacak olan devre analiz programının algoritması Python 3.12.0 yazılım dili kullanılarak oluşturulmuştur. Analiz edilen veriler ile literatür verilerinin karşılaştırılması amacıyla, program algoritmasına miRcode (120), DIANA TarBase (121), LncCeRBase (122), LncTarD (Zhao et al., 2023), miRDB (124), miRTarBase (124–126), TFLink (127), TRRUST (128), RNA Interactome Database (129), TargetScan (130), TransmiR (131) veri tabanlarından elde edilen ikili ilişki kütüphaneleri entegre edilmiştir. Literatür verilerinin program algoritması ile uyumlu hale getirilmesi ve entegrasyon süreci, Python 3.12.0 ortamında gerçekleştirilmiştir.

Çizelge 4.1. Veri setlerinin özellikleri.

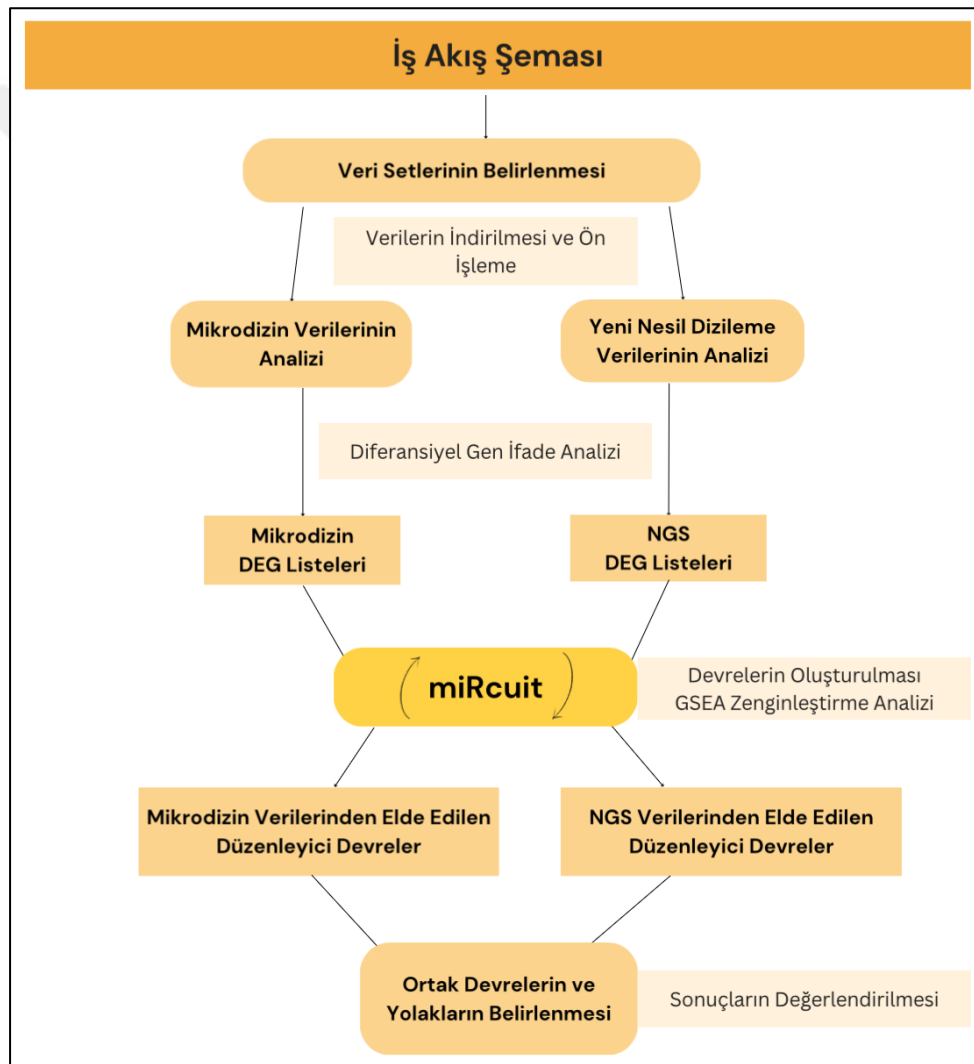
ÇALIŞMA NO	GEO ID	VERİ SETİ ÖZELLİKLERİ			TOPLAM BİREY SAYISI		DENEY TÜRÜ	PLATFORM BİLGİLERİ		YAYIN (PMID)
		Örnek Sayısı	Moleküller Altı Tip	Veri Seti	İnvaziv duktal karsinom	Normal meme dokusu		Platform Adı	Platform ID	
1	GSE40049	İnvaziv duktal karsinom (24) Eşli normal meme dokusu (14)	TNBC	miRNA	179 miRNA	51 miRNA	Yeni Nesil Dizileme (NGS)	AB SOLiD 4 System (Homo sapiens)	GPL13393	25888956
2	GSE39162	İnvaziv duktal karsinom (5) Eşli normal meme dokusu (5)	TNBC, luminal	miRNA	25 lncRNA	21 lncRNA		Illumina Genome Analyzer (Homo sapiens)	GPL9052	21199797
3	GSE29173*	İnvaziv meme karsinom (125) Normal meme dokusu (11)	TNBC, HER2+	miRNA	25 mRNA	21 miRNA		Illumina Genome Analyzer II (Homo sapiens)	GPL9115	21586611
4	GSE71651	İnvaziv duktal karsinom (15) Eşli normal ve sağlıklı normal meme dokusu (18)	TNBC, HER2+, luminal	Transkriptom	25 mRNA	21 miRNA		Illumina Genome Analyzer IIx (Homo sapiens)	GPL10999	28738804, 30764831, 30999914
5	GSE110114	İnvaziv duktal karsinom (10) Eşli normal meme dokusu (3)	Luminal	Transkriptom	25 TF	21 TF		Illumina HiSeq 2500 (Homo sapiens)	GPL9115	31395854, 31801944
6	GSE60689	İnvaziv duktal karsinom (2) Eşli normal meme dokusu (2)	TNBC	lncRNA			Mikrodizin	Agilent-045997 Arraystar human lncRNA microarray V3 (Probe Name Version)	GPL16791	30519573
7	GSE80266	İnvaziv duktal karsinom (5) Eşli normal meme dokusu (5)	TNBC, HER2+, luminal	lncRNA				Agilent-045997 Arraystar human lncRNA microarray V3 (Probe Name Version)	GPL16956	25416949
8	GSE113851	İnvaziv duktal karsinom (44) Normal meme dokusu (13)	ER/PR +	lncRNA mRNA	434 miRNA	119 miRNA		Arraystar Human lncRNA microarray V2.0 (Agilent-033010 Probe Name version)	GPL15314	27228351 34057025
9	GSE76250	İnvaziv duktal karsinom (165) Eşli normal meme dokusu (33)	TNBC	mRNA, lncRNA, miRNA, TF	470 lncRNA	79 lncRNA		Invitrogen NCode Human Non-coding RNA Microarray (NCRAH-02)	GPL16847	29765154
10	GSE45827	İnvaziv duktal karsinom (130) Normal meme dokusu (11)	TNBC, HER2+, luminal	mRNA, lncRNA, miRNA, TF	463 mRNA	72 mRNA		[HTA-2_0] Affymetrix Human Transcriptome Array 2.0 [transcript (gene) version]	GPL17586	26813360, 34615706 26921339
11	GSE73540	İnvaziv duktal karsinom (27) Normal meme dokusu (3)	ER+, HER2-	mRNA, lncRNA, miRNA, TF	419 TF	59 TF		[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	GPL570	27006338
12	GSE10797	İnvaziv duktal karsinom (28) Normal meme dokusu (5)	HER2+, luminal	mRNA, lncRNA, miRNA, TF				[HTA-2_0] Affymetrix Human Transcriptome Array 2.0 [transcript (gene) version]	GPL17586	27233495
13	GSE22544	İnvaziv duktal karsinom (13) Normal meme dokusu (4)	x	mRNA, lncRNA, miRNA, TF				[HG-U133A_2] Affymetrix Human Genome U133A 2.0 Array	GPL571	18373191
14	GSE59246**	İnvaziv duktal karsinom (56) Normal meme dokusu (3)	TNBC, HER2+, luminal	mRNA, lncRNA, TF				[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	GPL570	20799942
15	GSE59247**	İnvaziv duktal karsinom (15) Normal meme dokusu (4)	TNBC, HER2+, luminal	miRNA				Agilent-028004 SurePrint G3 Human GE 8x60K Microarray (Feature Number version)	GPL13607	27396337
16	GSE40525	İnvaziv duktal karsinom (56) Eşli normal meme dokusu (59)	TNBC, HER2+, luminal	miRNA				Agilent-031181 Unrestricted_Human_miRNA_V16.0_Microarray 030840 (miRBase release 14.0 miRNA ID version)	GPL15019	27396337
								Agilent-019118 Human miRNA Microarray 2.0 G4470B (miRNA ID version)	GPL8227	23125021

* Aynı çalışmanın hem array hem de NGS verisi olduğunu göstermektedir.

** Aynı çalışmanın hem mRNA hem de miRNA verisi olduğunu göstermektedir.

4.2. YÖNTEM

Tez çalışması kapsamında kullanılacak IDC ve normal/eşli normal dokular ile yapılan NGS ve mikrodizin çalışmalarının seçimine, seçilen veri setlerinin analizine, tümör ve normal dokular arasında anlamlı şekilde ifadesi değişen mRNA-miRNA-lncRNA-TF düzenleyici moleküllerinin belirlenmesine yönelik yöntem basamakları detaylı şekilde belirtilmiş ve iş akış şeması Şekil 4.1.'de verilmiştir. Ayrıca düzenleyici moleküllerin birbirleri ile olan regülatif ilişkilerini ortaya koyan devre analiz programının algoritması ve bu programın geliştirilme süreci kapsamlı şekilde ele alınmıştır.



Şekil 4.1. Verilerin elde edilmesinden regülatif devrelerin oluşturulmasına kadar izlenen basamakları gösteren iş akış şeması.

4.2.1. VERİ SETLERİNİN OLUŞTURULMASI

Geliştirilmekte olan devre analiz programı, araştırmacıların çalışma konusunu oluşturan spesifik iki durum arasında (tümör/normal, ilaç uygulaması var/yok vb.) anlamlı şekilde ifadesi değişen regülatif moleküller kullanarak düzenleyici biyolojik devreler tasarlamayı amaçlamaktadır. Söz konusu program, çalışılan konu üzerine elde edilmiş miRNA, mRNA, lncRNA ve TF ifade verilerini içeren veri dosyalarının sisteme yüklenmesi ile programın kullanıcılara açık ve/veya kapalı devreler kurgulaması üzerine tasarlanmaktadır.

Bu amaçla, Gene Expression Omnibus (GEO) veri tabanı üzerinden uygun arama kriterleri uygulanarak tez çalışmasının konusu olan invaziv duktal meme kanseri ve normal/eşli normal dokular ile yapılan yeni nesil dizileme (NGS) ve mikrodizin verilerine ulaşılmıştır. Çalışmanın kapsamına uygun veri setlerinin belirlenmesi için izlenen yol aşağıda belirtilmiştir:

1. GEO veri tabanında “*breast cancer non coding RNA*”, “*breast cancer transcriptome*” ve “*breast cancer RNA seq*” anahtar kelimeleri kullanılarak meme kanseri ile ilgili günümüze kadar yapılmış tüm NGS ve mikrodizin çalışmaları taranmıştır.
2. Taranan çalışmalar içerisinde invaziv duktal meme kanseri hastalarına ait taze dondurulmuş primer tümör dokuları ve bu hastaların eşli normal veya sağlıklı normal meme dokularını birlikte içeren çalışmalar seçilmiştir.
3. Tüm çalışmalar mRNA, miRNA ve lncRNA ifade verilerinden en az birini içermektedir.
4. Veri setlerinin oluşturulması sırasında hastaların moleküler alt tipleri arasındaki farklar gözlemlenmemiştir; ancak analizler sonucunda gerektiği takdirde daha spesifik araştırmalar yapabilmek adına kayıtları tutulmuştur.

Taramalar sonucunda 2 tanesi miRNA-mikrodizin (GSE59247 (118), GSE40525 (119), 6 tanesi transkriptom mikrodizin (GSE76250 (116), GSE45827 (115), GSE73540 (114), GSE10797 (113), GSE22544 (117), GSE59246 (118) olmak üzere toplamda 8 adet mikrodizin çalışması ile 3 tanesi yalnızca miRNA (GSE40049 (108), GSE39162 (109),

GSE29173 (110) verisini, 2 tanesi mRNA, miRNA ve lncRNA (GSE71651 (111), GSE110114 (112) verisini içeren toplamda 5 adet RNA-Seq çalışması veri setine dahil edilmiştir. Veri setlerinin özelliklerini, toplam birey sayısını, platform bilgilerini ve yayınlarını gösteren ayrıntılı tablo Çizelge 4.1.’de verilmiştir.

4.2.2. VERİ SETLERİNİN ANALİZİ

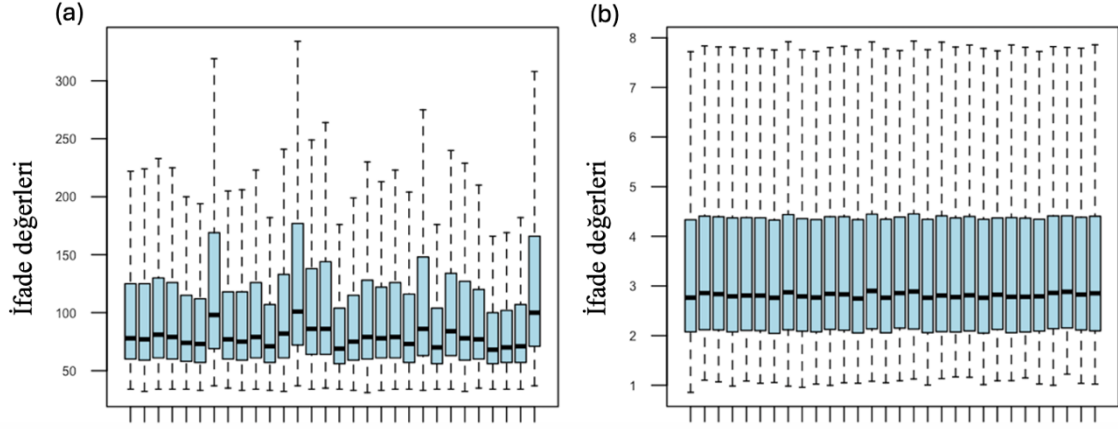
Tez çalışmasında kullanılan IDC ve normal dokulara ait ifade verileri, mikrodizin ve Yeni Nesil Dizileme (NGS) teknolojileri kullanılarak gerçekleştirilen çalışmalardan elde edilmiştir. İki farklı teknolojiye gelen verilerin analizi farklı yöntemlerle gerçekleştirilmiş olup analiz basamakları aşağıda başlıklar halinde sunulmuştur.

4.2.2.1. Mikrodizin Verilerinin Analizi

Tez çalışmasına dahil edilen IDC ve normal/eşli normal dokuların ifade verilerini içeren toplamda 8 adet mikrodizin çalışması bulunmaktadır. Bu çalışmalar Affymetrix ve Agilent mikrodizin çipleri ile yapılmış olup veri indirme ve ön işleme basamakları platforma göre farklılık göstermiştir.

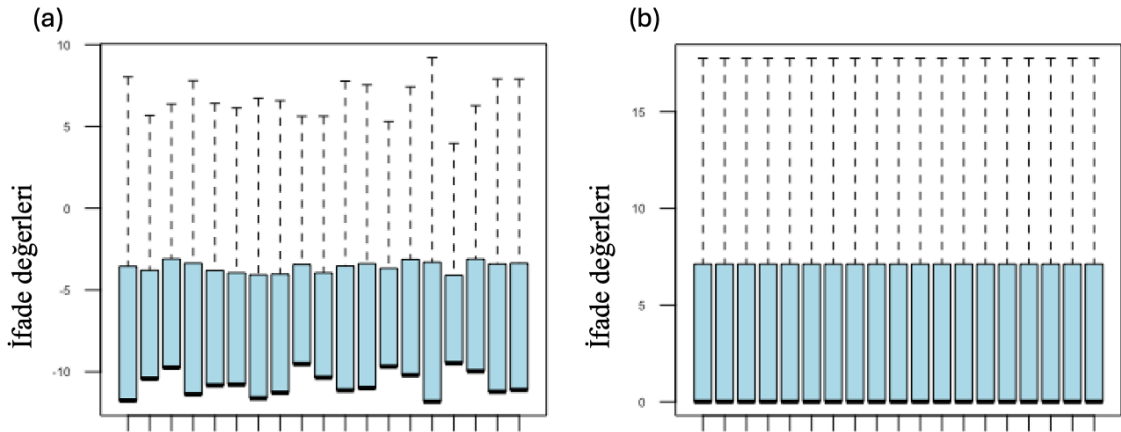
- *Verilerin İndirilmesi ve Ön İşleme*

Mikrodizin verilerinin indirilmesi R Studio’da *BiocManager* paket yöneticisi altında bulunan “*GEOquery*” paketi ile gerçekleştirilmiştir (132). Affymetrix platformuna ait veriler (GSE76250, GSE45827, GSE73540, GSE10797, GSE22544) .CEL dosya formatında indirilmiş ve “*affy*” (133) ve “*oligo*” (134) paketleri ile okunmuştur. Verilerin normalizasyonu için Robust Multi-array Average (RMA) normalizasyon yöntemi kullanılmıştır. Bu yöntem arka plan düzeltmesini, quantile normalizasyonu ve prob setin özetlenmesi basamaklarını içermektedir (135). Affymetrix mikrodizin veri setlerinden örnek olarak seçilen GSE73540 veri setine ait verilerin normalizasyon öncesi ve sonrası bar grafikleri Şekil 4.2.’de verilmiştir (Diğer veri setlerine ait tüm normalizasyon öncesi ve sonrası bar grafikleri EK 1’de sunulmuştur).



Şekil 4.2. GSE73540 veri setine ait verilerin RMA normalizasyon öncesi (a) ve sonrası (b) bar grafikleri.

Farklı teknoloji ile geliştirilmiş Agilent platformuna ait veriler (GSE59246, GSE59247, GSE40525) .txt dosya formatında indirilmiş ve herhangi bir ara işlem basamağına tabi tutulmamıştır. “limma” paketi (136) kullanılarak verilere ilk olarak “normexp” yöntemi ile arka plan düzeltmesi yapıp ardından “quantile” normalizasyon yöntemi uygulanmıştır. Agilent mikrodizin veri setlerinden örnek olarak seçilen GSE59247 veri setine ait verilerin normalizasyon öncesi ve sonrası bar grafikleri Şekil 4.3.’te verilmiştir (Diğer veri setlerine ait tüm normalizasyon öncesi ve sonrası bar grafikleri EK-1’de sunulmuştur).



Şekil 4.3. GSE59247 veri setine ait verilerin quantile normalizasyon öncesi (a) ve sonrası (b) bar grafikleri.

- *Diferansiyel Gen İfade Analizleri*

İnvaziv duktal meme kanseri dokusu ile normal doku arasındaki genlerin ifade değişimlerini belirlemek amacıyla *BiocManager* paket yöneticisi altında bulunan “limma” paketi

kullanılarak diferansiyel gen ifade analizleri gerçekleştirilmiştir. Normalize verilere *lmFit* fonksiyonu uygulanarak verideki her grup (IDC/normal) için genlerin ifade seviyeleri belirlenmiştir ve *makeContrasts* fonksiyonu ile karşılaştırma matrisleri oluşturulmuştur. *eBayes* fonksiyonu ile empirical Bayes düzeltmesi uygulanarak istatistiksel güç artırılmış ve yanlış pozitif değerler azaltılarak diferansiyel ifade istatistikleri ve p değerleri hesaplanmıştır. Sonuç olarak normalden dokulara kıyasla tümör dokularda ifadesi anlamlı olarak değişen genler elde edilmiştir.

Analiz sonucu elde edilen tablolarda, her bir gen için mikrodizinlerin prob ID'leri yer almaktadır. Prob ID'lerin anotasyon işlemi, her bir GSE kodlu çalışmanın kendi sayfasındaki anotasyon dosyası kullanılarak Python 3.12'de gerçekleştirilmiştir. Örnek olarak seçilen GSE73540 ve GSE59247 veri setlerinin diferansiyel gen ifade analizlerine ait ekran görüntüleri Şekil 4.4. ve (137–141).’te verilmiştir.

prob_ID	logFC	AveExpr	t	P.Value	adj.P.Val	B	GeneSymbol
TC09001920.hg.1	-1.88040087829967	2.36743615301551	-14.1854636410965	6.814852099030791e-15	2.29253159064308e-10	22.2659770356448	---
TC0X000943.hg.1	-1.62485586220697	2.59100866179674	-13.8991910590177	1.16517068976294e-14	2.29253159064308e-10	21.8311238144926	DMD
TC0X002026.hg.1	-2.82883565693606	2.9406032675566	-13.2346322755416	4.178716760137031e-14	5.48122727427174e-10	20.7831827473321	DMD
TC09000274.hg.1	-1.46289130783049	2.3375754953249	-13.0440359544352	6.0788246568963e-14	5.980195726838161e-10	20.4724075864013	CNTNAP3B
TC06003501.hg.1	-2.08827726623405	3.15350833852606	-12.7174587069354	1.16596495675896e-13	8.03084198317486e-10	19.9289892330533	---
TC09002584.hg.1	-2.10711302836489	2.50742886045636	-12.6931346850211	1.22449370788669e-13	8.03084198317486e-10	19.8879551073378	---
TC0X002025.hg.1	-2.6458282567683	2.84994718489327	-11.6940364588937	9.70009226664786e-13	5.227449260836481e-09	18.1332622908346	DMD
TC06003119.hg.1	-2.48712819084283	2.18769007960609	-11.6512147484116	1.06273269006358e-12	5.227449260836481e-09	18.0549596081915	SAMD5
TC03001145.hg.1	-1.39218832647839	4.95501858578091	-11.0370549029702	4.0329436184385296e-12	1.76333738143527e-08	16.9029188079346	OXTR
TC09001921.hg.1	-0.1695320731725	2.2221600117243	-10.9738172045325	4.6387103737742e-12	1.82537891918388e-08	16.7811749563781	---
TC09002504.hg.1	-1.22580298849235	2.96732092281875	-9.80365041261351	6.7702708454611e-11	2.42197207308854e-07	14.4199184367075	---
TC09001960.hg.1	-1.0372957967776	3.1342079132597	-9.63448327087483	1.01217357867664e-10	3.31917020787538e-07	14.0612249218783	---
TC09000253.hg.1	-0.643884979474552	2.75999207519257	-9.59002268092411	1.12570923076978e-10	3.40752184154011e-07	13.9662169224287	CNTNAP3B
TC09001937.hg.1	-0.747015129561574	3.6812290951978	-9.01011570899191	4.6129197834076006e-10	1.2965929028348e-06	12.6988791976563	---
TC15000953.hg.1	-2.08735800741599	5.34661159233287	-8.897250416578	6.101377679931171e-10	1.60063542055314e-06	12.4461466998739	SYNM
TC08001407.hg.1	-0.776289534543997	2.39592247862376	-8.85635357423278	6.75484787978615e-10	1.66131261823416e-06	12.3540814780542	C8orf88
TC21000892.hg.1	-2.80659581660429	3.07360148794887	-8.79291233660346	7.91316713619994e-10	1.83171199986238e-06	12.210753052613	ADAMTS5
TC09001097.hg.1	-0.835205847677174	3.40245373635184	-8.69287822286789	1.01670895611536e-09	2.21547421623684e-06	11.983489676591	CNTNAP3P2
TC21000040.hg.1	-0.766367696932727	2.4638765066131	-8.67266044760128	1.06970623639806e-09	2.21547421623684e-06	11.9373702953021	LINC00478
TC0X002024.hg.1	-1.41393201777378	3.51547640299151	-8.64947686216157	1.13396112275459e-09	2.2311252070758e-06	11.8844080594492	DMD
TC20001641.hg.1	-2.52801214745005	2.89164763411985	-8.51346101061002	1.59897387539689e-09	2.99624861765443e-06	11.5720207513688	---
TC0X000868.hg.1	-0.71959374241586	3.02283756439454	-8.45080398858651	1.8747494249291303e-09	3.35333021001755e-06	11.427163533677	GPM6B
TC09001092.hg.1	-0.482783767979176	3.11962478068792	-8.12520408508565	4.32089276702341e-09	7.39267179457122e-06	10.6648204505965	CNTNAP3
TC09002408.hg.1	-1.93677051627364	2.7131562883226	-8.01736179066012	5.7144678754681815e-09	9.36958439031452e-06	10.4088192823067	---
TC07001192.hg.1	-1.96352225411801	3.67337266996316	-7.9702366694136	6.45990704654724e-09	1.01681520875472e-05	10.296411033853	FAM126A
TC05000601.hg.1	-0.721992849019505	3.30256926545025	-7.94773164223266	6.850135672932941e-09	1.0367680340984e-05	10.2426141573339	GRAMD3
TC07001221.hg.1	-0.843008270625681	3.88929951888773	-7.87981752741145	8.17967352862513e-09	1.1921419741664e-05	10.0798199490715	HOXA5
TC0X000544.hg.1	-1.51888609180553	2.37006935388442	-7.78913745258626	1.03749048872314e-08	1.45808172220516e-05	9.8614081117033	---
TC07001811.hg.1	-0.654957475896868	2.57464365932878	-7.76861728978706	1.09498807587627e-08	1.48129489195287e-05	9.81181843341096	AASS
TC17001137.hg.1	-0.650673796003713	2.77825242048199	-7.75689000961361	1.12929396352281e-08	1.48129489195287e-05	9.78345073003894	TMEM220
TC15001679.hg.1	-1.60053656311362	2.26269579197708	-7.73693746538842	1.19019527101259e-08	1.51081851966505e-05	9.73514118786865	---
TC04002187.hg.1	-1.99758461954775	4.51132893520586	-7.68303129108172	1.37199687414974e-08	1.68717028108332e-05	9.60433786279565	FGF2
TC16000476.hg.1	-1.35028162900197	5.82049767313841	-7.6628839549087	1.44699079778819e-08	1.72547075405342e-05	9.55534419286573	MT1X
TC11001552.hg.1	-0.922551956897461	3.03171020517236	-7.52253730697347	2.09921472401387e-08	2.42959407660793e-05	9.21246722489961	PAMR1
TC13000763.hg.1	-1.17358860184738	4.1873930892117	-7.49505154830547	2.25851245594898e-08	2.53927782318435e-05	9.14499578064758	SPRY2
TC09000323.hg.1	-1.2546837893871	3.10651694428895	-7.4253406267824	2.71995242522002e-08	2.88674289019776e-05	8.97340445154565	MAMDC2
TC19000202.hg.1	-1.78210398804626	5.2810765917626	-7.4217950081336	2.7458348990537e-08	2.88674289019776e-05	8.96465924157565	CNN1
TC21000346.hg.1	-1.56161108735154	3.73987031775302	-7.41613993801012	2.78763512560074e-08	2.88674289019776e-05	8.95070756018701	ADAMTS5
TC17000693.hg.1	-0.766003706762508	3.8523938282584	-7.35210598495925	3.30861347648928e-08	3.33839099777769e-05	8.79242615835999	HLF
TC15002015.hg.1	-0.988300757417988	4.33808087383802	-7.29578410382804	3.84825327731753e-08	3.78581536789305e-05	8.65275181800466	---
TC03002373.hg.1	-0.887932171542819	2.59726658318936	-7.27145266922998	4.10827346851154e-08	3.93823436846586e-05	8.59228084647272	ADAMTS9-AS2

Şekil 4.4. GSE73540 veri setinin diferansiyel gen ifade analizi.

ID	logFC	AveExpr	t	P.Value	adj.P.Val	B
hsa-miR-517c	-2.28529887631579	0.481117174058297	-926.01408290743	4.5417449956294695e-4	6.21310715402112e-39	82.9925637429461
hsa-miR-1254	-7.12590690263158	1.50019254801952	-757.89046815783	1.45452892518234e-40	9.94897784824721e-38	80.8855588928696
hsa-miR-519b-3p	-1.99207761710526	0.419386382645555	-92.1640721836689	9.736481314408621e-25	4.43983547937033e-22	46.5903473969836
hsa-miR-520c-3p	-2.0790290881579	0.437691955498741	-81.185690844166	8.6973367650781e-24	2.9744891736567102e-2	44.2925580832933
hsa-miR-383	-3.75829353307018	0.820179067410098	-57.088823087951	3.76659944202666e-21	1.03054160733849e-18	37.9040169467582
hsa-miR-488	-3.63847119473684	0.76599556883782	-32.615268691972	5.54826448787585e-17	1.26500430323569e-14	27.7779329531609
hsa-miR-517b	-6.38215510824561	1.49009802779791	-22.169056325102	3.7982612300633e-14	7.42288766103799e-12	20.8860411963372
hsa-miR-520e	-3.40485369473684	0.716812925304835	-19.601874767283	2.95631362392922e-13	5.0552962969189694e-1	18.7209600380686
hsa-miR-190	-2.55490888289474	0.537877175443339	-15.138493679084	2.0415821813390297e-1	3.1032049156353304e-0	14.256300038532
hsa-miR-517a	-6.45038961701754	1.59506965245165	-13.870936781949	8.309487931305139e-11	1.03339813545686e-08	12.7781301847688
hsa-miR-501-3p	2.79557402482456	6.63552151311647	10.7723428305726	4.2744408301034805e-0	4.87286254631797e-07	8.63551362566506
hsa-miR-335*	-6.80674843324561	2.01994673306107	-10.390885448214	7.368462846470431e-09	7.75389013382426e-07	8.06407987797756
hsa-miR-520h	-1.57303739605263	0.331167388739738	-9.4926448391757	2.82148806388848e-08	2.65272292318088e-06	6.65662553679094
hsa-miR-1280	2.479515631	11.0891349429918	9.47294692441901	2.9086874157685e-08	2.65272292318088e-06	6.62474512969773
hsa-miR-720	2.82790810385965	17.0457166505126	8.93382346565479	6.80852663487272e-08	5.82129027281617e-06	5.73448358761304
hsa-miR-4286	3.38811845986842	14.0384352811774	8.691810032032	1.00862307198111e-07	8.11644919100097e-06	5.32347180739059
hsa-miR-501-3p	-5.98758579614035	1.84068365577575	-8.0770862843824	2.82757616168884e-07	2.14895788288351e-05	4.426633119164117
hsa-miR-4291	1.7321487291579	8.57547576785331	7.97912480457951	3.34722287570703e-07	2.41000047050906e-05	4.07057808064309
hsa-miR-411*	-1.0584864	0.222840863254973	-7.8763328782220	4.00082496453153e-07	2.73656427573957e-05	3.88450808658075
hsa-miR-33b*	2.87647810263158	5.75758834164833	7.51976114902949	7.50825573142358e-07	4.8910923504165e-05	3.22836580074499
hsa-miR-766	2.06637240438597	7.01749742724389	7.44986113450376	8.51122620826019e-07	5.29243520586361e-05	3.09778218440293
hsa-miR-210	3.49132117465965	9.75247326693364	7.29911197701844	1.11788320961129e-06	6.63693116376549e-05	2.81396540613798
hsa-miR-3200-5p	-3.8297840495614	1.18268359649597	-7.2767260802575	1.16437388837991e-06	6.63693116376549e-05	2.77156322141634
hsa-miR-335	-3.23774897302632	8.77951178915525	-6.7684148013581	2.99124885541242e-06	0.0001636811373681	1.79092902546971
hsa-miR-525-5p	-1.65407535561404	0.473216395664945	-6.5574042468233	4.47134124342837e-06	0.0002352613392696	1.37387843502822
hsa-miR-3925-5p	-4.61263070973684	1.32129206214694	-6.5331112465420	4.68496326516671e-06	0.0002373714721017	1.32549339868928
hsa-miR-1281	2.25100549263158	6.85768858375359	6.45741524754322	5.42106729805586e-06	0.0002648578594193	1.17423966712346
hsa-miR-371-3p	-1.51075909342105	0.31805616713309	-6.3912272586184	6.16281316473906e-06	0.0002907147727366	1.0413820966749
hsa-miR-3188	1.68846234087719	8.38563039815802	6.32514216453202	7.00879358535808e-06	0.0003101688648453	0.908174016001718
hsa-miR-21*	3.65051622456491	10.2970547681746	6.32368990213963	7.02868041681773e-06	0.0003101688648453	0.905240462058787
hsa-miR-1295	-4.01190571473684	1.29788693056799	-6.1873584885268	9.180201957033229e-06	0.0003925313366317	0.628670753845201
hsa-miR-4313	2.78055709350877	7.77664740286716	5.87501876048546	1.70959366382765e-05	0.0006878600388577	-0.013583776899432
hsa-miR-4323	2.80208385631579	5.14771345771481	5.859072948880055	1.7653217960321e-05	0.0006899886334205	-0.046686135492480
hsa-miR-329	-3.92678614552631	1.27659941283946	-5.8399624994023	1.83458763463011e-05	0.0006971433011594	-0.086397207569513
hsa-miR-337-3p	-5.79156977964912	2.0692593740583	-5.6225927197801	2.85156750873835e-05	0.0010543092843119	-0.541027189196974
hsa-miR-425*	2.44962398754386	6.3594335507896	5.5250224701768	3.48256813314527e-05	0.0012215777451648	-0.746798210022108
hsa-miR-1825	2.20549890122807	6.68458875632976	5.39238547980197	4.57852274479727e-05	0.0015276631987518	-1.02813102595178
hsa-miR-509-3-5p	-3.1772817722807	1.09244012724389	-5.3760459944589	4.73616331587756e-05	0.0015426360514572	-1.06291229788787
hsa-miR-129-5p	-1.65829302964912	0.530609479321455	-5.3148359263279	5.37800100524862e-05	0.0017109547384139	-1.19344236329677

Şekil 4.5. GSE59247 veri setinin diferansiyel gen ifade analizi.

Elde edilen listelerde pozitif logFC (logFoldChange) değerleri, normal dokuya göre tümör dokusunda ifadesi artan genleri; negatif değerler ise ifadesi azalan genleri göstermektedir.

- *Devre Analiz Programı için Verilerin Hazırlanması*

Tez çalışmasında kullanılan mikrodizin veri setleri, miRNA ve transkriptom mikrodizinleri olmak üzere iki çeşittir. miRNA mikrodizin çalışmaları yalnızca miRNA ifade verisi içerirken transkriptom çalışmaları mRNA, miRNA, lncRNA ve TF verilerini birlikte içermektedir. Mikrodizin çalışmalarının analizi sonucunda her bir düzenleyici molekül türüne ait ayrı bir dosya oluşturabilmek amacıyla, devre analiz programında kullanılmak üzere oluşturulmuş kütüphanelerden elde edilen mRNA, miRNA, lncRNA ve TF listeleri, oluşturulan gen listeleri ile karşılaştırılmış ve tüm veri setleri için düzenleyici moleüllere ait ayrı dosyalar oluşturulmuştur. Sonuç olarak her bir düzenleyici molekül türüne ait listeler birleştirilmiş ve tüm mikrodizin çalışmalarından elde edilen tek bir mRNA, miRNA, lncRNA ve TF listesi elde edilmiştir. Son aşamada miRNA'lar için isimlendirme problemini ortadan kaldırmak adına miRNA'ların isminden "hsa" eki kaldırılmış ve devre analiz programında uyumsuzluk yaşanmaması amacıyla tüm isimler "miR" eki ile başlayan tek bir isimlendirme sistemine uygun hale getirilmiştir.

4.2.2.2. Yeni Nesil Dizileme Verilerinin Analizi

Tez çalışmasına dahil edilen IDC ve normal/eşli normal dokuları içeren toplamda 5 adet RNA-Seq çalışmasının bir tanesi AB SOLID Sistem platformuna ait olup diğer çalışmalar Illumina platformlarına aittir. RNA-Seq çalışmaları içerisinde yalnızca miRNA verisini içeren çalışmaların (GSE29173, GSE39162, GSE40049) analizleri, transkriptom çalışmalarının (GSE71651, GSE110114) analizlerine göre farklılık göstermiştir.

- *Verilerin İndirilmesi ve Ön İşleme*

Transkriptom verilerinin indirilmesinden diferansiyel ifade analizine kadar olan tüm işlem basamakları *Ubuntu* işletim sisteminde, komut satırı üzerinden çalıştırılan biyoinformatik araçlar kullanılarak gerçekleştirilmiştir. Veri setlerinde bulunan hastaların SRR kodlu ham transkriptom verileri, National Library of Medicine (NCBI) web sitesindeki Sequence Read Archive (SRA) ara yüzünde bulunmaktadır. Her bir hastaya ait ham veriler, komut satırı üzerinden sra-tools paketinde yer alan *fasterq-dump* (versiyon 2.11.3) aracı ile SRA veri tabanına bağlanılarak indirilmiştir (142). Çift yönlü okumaya sahip transkriptom verileri *fastq*'ya çevrilmiş ve dizileme verilerinin kalite kontrolleri *FastQC* aracı (versiyon 0.11.9) ile gerçekleştirilmiştir (143). Örnek olarak seçilen GSE71651 kodlu çalışmanın SRR2148235 kodlu hastasının fastqc sonuçları Şekil 4.6'da gösterilmiştir.



Şekil 4.6. GSE71651 kodlu çalışmanın SRR2148235 kodlu hastasının ileri (a) ve geri (b) yönlü okumalarının kalite kontrol istatistikleri ve grafikleri.

Oluşturulan *fastq* dosyaları, RNA verilerinin hizalanmasında splice bölgelerine karşı hasasiyeti güçlü, hızlı işlem yapabilen ve düşük bellek tüketimine sahip *hisat2* aracı (versiyon 2.2.1) ile NCBI'dan indirilen en güncel insan referans genomu olan GRCh38.p14'e hizalanmıştır (144). Hizalama işlemi sonrasında elde edilen SAM dosyaları, *samtools* aracı (versiyon 1.13) kullanılarak BAM dosyalarına çevrilmiş, sıralanmış ve indekslenmiştir (145). Elde edilen BAM dosyalarının okuma sayıları, subread paketindeki *featureCounts* aracı (versiyon 2.0.3) ile GENCODE'tan indirilen “*Comprehensive gene annotation ALL gtf*” (versiyon 46) anotasyon dosyası kullanılarak elde edilmiştir (146). Her birey için elde edilen feature count matrix dosyası, gen ID'leri ve bu gen ID'lere karşılık gelen okuma sayılarını içeren .csv uzantılı dosyalardır. Veri setlerindeki tüm örnekler ait okuma sayıları, gen ID'ler referans alınarak birleştirilmiş ve her veri seti için diferansiyel ifade analizi gerçekleştirmek üzere hazırlanmış bir count matrix dosyası oluşturulmuştur. Örnek olarak seçilen GSE71651 kodlu çalışmaya ait feature count matrix dosyasının ekran görüntüsü Şekil 4.7.'de verilmiştir.

Genid	SRR2148235	SRR2148236	SRR2148237	SRR2148238	SRR2148239	SRR2148240	SRR2148241	SRR2148242	SRR2148243	SRR2148244	SRR2148245	SRR2148246	SRR2148247	SRR2148248	SRR2148249	SRR2148250	SRR2148251	SRR2148252	SRR2148253	SRR2148254	SRR2148255
ENSG000002272972	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ENSG00000227232	127	73	62	58	44	106	58	94	128	107	106	96	220	218	118	94	109	120	86	215	62
ENSG00000270267	30	21	3	8	2	14	7	22	13	17	26	28	15	5	8	15	4	16	12	49	14
ENSG00000243485	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ENSG00000274890	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ENSG00000227613	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ENSG00000268020	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ENSG00000240361	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ENSG00000186092	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ENSG00000238009	3	0	0	0	0	2	0	2	2	0	0	3	0	0	0	0	0	5	2	5	2
ENSG00000239545	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ENSG00000232750	4	4	9	0	6	4	5	4	7	6	2	4	10	3	2	8	0	1	0	16	4
ENSG00000268963	133	16	20	16	28	3	1	0	38	13	7	37	14	2	18	2	6	0	21	6	0
ENSG00000268961	95	7	5	7	10	2	2	3	29	4	31	3	3	0	2	8	2	2	1	8	0
ENSG00000239596	16	1	1	7	5	8	3	4	3	5	4	3	2	1	1	5	0	0	0	8	2
ENSG00000241890	31	33	16	11	2	6	10	7	22	13	6	8	14	6	19	35	7	6	5	4	12
ENSG00000222632	0	1	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0
ENSG00000241599	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ENSG00000277928	3	3	0	0	0	2	0	0	0	1	6	0	0	0	1	6	0	5	0	2	2
ENSG00000279457	556	594	231	269	263	201	170	249	496	379	812	253	357	336	208	468	236	331	445	649	187
ENSG00000273674	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ENSG00000275135	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ENSG00000229483	29	13	1	19	22	59	15	11	12	29	21	30	3	15	29	46	13	15	6	10	30
ENSG00000241670	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ENSG00000238679	16	4	4	17	15	6	2	24	6	7	9	2	11	4	9	4	5	13	9	8	14
ENSG00000238143	0	0	0	0	0	0	0	1	0	1	4	0	0	0	1	6	2	0	1	1	0
ENSG00000238601	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	2	0	0	0
ENSG00000237094	32	14	12	11	10	16	10	18	13	17	23	10	23	9	8	52	6	26	0	9	14
ENSG00000268732	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ENSG00000278566	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ENSG00000224813	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ENSG00000233633	0	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0	1	1	0	0	0
ENSG00000250575	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
ENSG00000271571	1	0	2	0	4	0	0	0	0	0	2	0	3	2	0	0	0	0	0	0	0
ENSG00000231709	7	0	4	5	0	2	4	1	2	2	10	3	1	2	2	6	4	16	2	1	0
ENSG00000235146	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
ENSG00000239664	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0
ENSG00000230021	5	2	0	1	1	6	0	6	0	3	0	0	6	2	5	2	0	3	0	1	0
ENSG00000222698	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ENSG00000225972	0	2	1	1	0	1	0	2	2	0	0	6	0	0	0	4	0	0	0	0	0
ENSG00000225626	673	615	419	432	184	402	327	684	775	571	285	1245	223	377	475	2333	184	253	207	1153	1533
ENSG00000271171	48	38	27	20	64	31	15	17	12	2	18	10	26	26	8	7	28	32	8	13	39
ENSG00000237973	184	361	434	165	87	335	339	347	415	230	152	795	182	259	354	155	146	346	170	374	485
ENSG00000273793	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ENSG00000229344	12	0	0	0	0	0	0	2	1	5	4	0	5	0	0	2	0	54	2	4	0
ENSG00000240409	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0
ENSG00000248527	1296	3880	1046	950	342	804	904	1089	1607	1007	609	3215	800	532	618	7944	599	407	468	1275	1081

Şekil 4.7. GSE71651 kodlu çalışmaya ait örnek feature count matrix görüntüsü.

Yalnızca miRNA sekans verisini içeren çalışmaların analiz yöntemi, transkriptom verisi için izlenen basamaklardan farklı gerçekleştirilmiştir. Seçim yöntemi “size fractionation” olan ve tek yönlü okunan miRNA çalışmalarında, her bireye ait feature count matrix dosyası GEO veri tabanından indirilmiştir. Veri集中的 tüm bireylere ait okuma sayıları, miRNA isimleri referans alınarak Python 3.12’de birleştirilmiş ve diferansiyel gen ifade analizine alınmak üzere hazır hale getirilmiştir. Örnek olarak seçilen GSE29173 kodlu çalışmaya ait örnek feature count matrix dosyasının ekran görüntüsü Şekil 4.8.’de verilmiştir.

Sequence	Annotation	Count_100_annot	Count_101_annot	Count_102_annot	Count_103_annot	Count_104_annot	Count_105_annot	Count_106_annot	Count_107_annot	Count_108_annot	Count_109_annot
AAAAAAAAAGCCCAACCT	hsa-miR-3613*	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
AAAAAACCGTTACCATTAAGT	hsa-miR-451-DICER1	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
AAAAAGTAGAAGCACTACT	hsa-miR-142-5p	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
AAAAAGCTGGGTTGAGAGGGCGA	hsa-miR-320-RNASEN	642.0	3511.0	579.0	1339.0	1613.0	2296.0	38.0	22.0	34.0	106.0
AAAAAGGCATAAAACCAAGACAA	hsa-miR-3910	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
AAAAAGGCTTACAGTGCAGGTA	hsa-miR-106a	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
AAAAAGTGTCTGTTCTGTCAGGTAA	hsa-miR-93	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0
AAAAAGTGTCTGTTCTGTCAGGTAG	hsa-miR-93	2.0	0.0	0.0	0.0	2.0	0.0	0.0	0.0	4.0	0.0
AAAAAGTGTCTTAAAGTGCAGTA	hsa-miR-106a	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
AAAAAGTGTCTTACAGTGCAGGT	hsa-miR-106a	18.0	10.0	2.0	8.0	7.0	3.0	1.0	0.0	1.0	1.0
AAAAAGTGTCTTACAGTGCAGGTA	hsa-miR-106a	22.0	9.0	3.0	4.0	5.0	0.0	0.0	2.0	2.0	3.0
AAAAAGTGTCTTACAGTGCAGGTAG	hsa-miR-106a	16.0	12.0	4.0	7.0	12.0	3.0	0.0	3.0	3.0	5.0
AAAAAGTGTCTTATAGTGCAGGTAG	hsa-miR-20a	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
AAAAAGTGTCTTACAGTGCAGGT	hsa-miR-106a	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
AAACAACAAAGCTGGCACTTCTT	hsa-miR-495	12.0	30.0	4.0	7.0	42.0	1.0	0.0	0.0	0.0	0.0
AAACAACAACTGGTGCACTTCTTT	hsa-miR-495	11.0	29.0	4.0	2.0	25.0	1.0	0.0	0.0	0.0	3.0
AAACATCTACACCTCTCAGCT	hsa-miR-30c-1	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
AAACCGTTTACCACTTACTGAGTTT	hsa-miR-451-DICER1	38290.0	1881.0	89.0	507.0	707.0	77.0	1.0	10.0	26.0	39.0
AAACCGTTTACCACTTACTGAGTTTA	hsa-miR-451-DICER1	789.0	27.0	3.0	16.0	32.0	1.0	1.0	0.0	5.0	3.0
AAACCGTTTACCACTTACTGAGTTTAA	hsa-miR-451	9.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
AAACCGTTTACCACTTACTGAGTTTAC	hsa-miR-451	4.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
AAACCGTTTACCACTTACTGAGTTTAC	hsa-miR-451	1801.0	97.0	7.0	26.0	57.0	3.0	0.0	1.0	0.0	0.0
AAACCGTTTACCACTTACTGAGTTTAT	hsa-miR-451	116.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
AAACCGTTTACCACTTACTGAGTTTCA	hsa-miR-451	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
AAACCGTTTACCACTTACTGAGTTTCC	hsa-miR-451	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
AAAGTGTCTGACAGTGCAGAT	hsa-miR-106b	16.0	13.0	1.0	0.0	17.0	1.0	3.0	10.0	16.0	2.0
AAAGTGTCTGACAGTGCAGATA	hsa-miR-106b	5.0	4.0	0.0	0.0	14.0	0.0	3.0	5.0	7.0	2.0
AAAGTGTCTGACAGTGCAGATAA	hsa-miR-106b	3.0	1.0	0.0	0.0	5.0	0.0	0.0	0.0	1.0	0.0
AAAGTGTCTGTTCTGTCAGGTAG	hsa-miR-93	42.0	22.0	0.0	2.0	28.0	3.0	2.0	10.0	17.0	7.0
AAAGTGTCTGTTCTGTCAGGTAGA	hsa-miR-93	2.0	4.0	1.0	1.0	2.0	1.0	0.0	0.0	0.0	0.0
AAAGTGTCTGTTCTGTCAGGTAT	hsa-miR-93	2.0	4.0	0.0	0.0	3.0	0.0	0.0	2.0	7.0	2.0
AAAGTGTCTTACAGTGCAGGT	hsa-miR-106a	11.0	5.0	2.0	1.0	5.0	3.0	4.0	3.0	1.0	0.0
AAAGTGTCTTACAGTGCAGGT	hsa-miR-17	11.0	5.0	2.0	1.0	5.0	3.0	4.0	3.0	1.0	0.0
AAAGTGTCTTACAGTGCAGGTA	hsa-miR-106a	14.0	3.0	0.0	1.0	5.0	2.0	4.0	2.0	5.0	1.0
AAAGTGTCTTACAGTGCAGGTA	hsa-miR-17	14.0	3.0	0.0	1.0	5.0	2.0	4.0	2.0	5.0	1.0
AAAGTGTCTTACAGTGCAGGTAG	hsa-miR-106a	11.0	17.0	11.0	2.0	23.0	2.0	6.0	3.0	11.0	2.0
AAAGTGTCTTACAGTGCAGGTAG	hsa-miR-17	11.0	17.0	11.0	2.0	23.0	2.0	6.0	3.0	11.0	2.0

dayanmaktadır. DESeq2 metodu, dağılım ve logaritmik kat değişimleri tahminleri ile farklı ifadeleri tespit edebilmeyi sağlamaktadır. Ayrıca dağılımı verinin kendisinden tahmin etmesi ve verinin gözlenen özelliklerine dayanarak değişim miktarını kontrol etmesi metodun güvenilirliğini arttırmaktadır (147). Tez çalışması kapsamında verilerin diferansiyel gen ifade analizleri, DESeq2 metodunun Python uyarlaması olan PyDESeq2 (versiyon 0.4.11) kullanılarak gerçekleştirilmiştir (148). Her bir veri seti için oluşturulan .csv formatındaki feature count matrix'ler Python 3.12'ye yüklenmiş ve okuma değerleri sıfır olan genler veri setinden çıkarılmıştır. Karşılaştırılacak olan gruplar belirlenmiş (IDC/Normal) ve metadata olarak tanımlanmıştır. Feature count matrix'ler ile metadatalar Python komutları ile birleştirilmiştir. Diferansiyel gen ifade analizleri, PyDESeq2 paketi kullanılarak gerçekleştirilmiş ve analiz sonucunda normal dokulara göre tümör dokularında ifadesi anlamlı şekilde değişen genler belirlenmiştir.

Elde edilen diferansiyel ifade tablolarında gen isimleri Ensembl ID'leri ile belirtilmiştir. Gen sembollerine ulaşmak için gerekli olan anotasyon işlemi Python 3.12'deki *sanbomics.tools* (versiyon 0.1.0) modülünün "*id_map*" fonksiyonu ile gerçekleştirilmiştir (149). Sonuç olarak Ensembl ID'leri gen isimleri ile eşleştirilmiş ve veri setleri standart bir formata getirilmiştir. Örnek olarak seçilen GSE71651 kodlu çalışmanın tamamlanmış analizini gösteren ekran görüntüsü Şekil 4.9.'da verilmiştir.

Geneid	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj	Symbol
ENSG00000227232	106.86686876993181	0.2227787823705861	0.2003654934023677	1.1118620206884064	0.26619748045972724	0.9994548905052463	WASH7P
ENSG00000278267	15.422473589554285	0.03276363725193899	0.303116478206973	0.10808926471350537	0.9139248837785413	0.9994548905052463	MIR6859-1
ENSG00000241860	19.46685734864815	-0.5111943986125638	0.4954820833210032	-1.031711167407402	0.30220744434863855	0.9994548905052463	AL627309.5
ENSG00000279457	380.3759292604917	-0.21645775733814615	0.23124710189605208	-0.9360452760849999	0.34924987687461617	0.9994548905052463	WASH9P
ENSG00000236679	10.425098571066615	-0.2708818175279463	0.3283318521842439	-0.8250244858239968	0.40935768942197925	0.9994548905052463	RPL23AP24
ENSG00000225630	797.95947271717377	-0.6815289552004172	0.38258349592234037	-1.7813861874971184	0.07484938313489412	0.9994548905052463	MTND2P28
ENSG00000237973	325.637183550341	0.0650085188917262	0.2799684838989927	0.23219941754294038	0.8163831301285839	0.9994548905052463	MTCO1P12
ENSG00000248527	1679.3192259090927	-0.9714975511982507	0.413928187568889	-2.347019556469724	0.018924261472132053	0.9994548905052463	MTATP6P1
ENSG00000229376	10.979057159259005	-0.16155472766177936	0.3194018916803516	-0.505803916225577	0.6129943090273959	0.9994548905052463	CICP3
ENSG00000237491	426.81121313825	-0.10436811693924362	0.1685741348706261	-0.6191229575008169	0.5358353605295318	0.9994548905052463	LINC01409
ENSG00000228794	77.4005729583901	-0.1562886578602434	0.2056032282200358	-0.7601465565590434	0.4471669863313297	0.9994548905052463	FAM87B
ENSG00000225890	673.465768669248	-0.15318710314099945	0.11879676301847056	-1.2894888652579015	0.19722818489221838	0.9994548905052463	LINC01128
ENSG00000225890	90.39262588075718	-0.38776301001466124	0.19256847348489404	-2.012696368867809	0.044146581627695886	0.9994548905052463	LINC00115
ENSG00000272438	11.141478387633851	-0.36359949724369217	0.6631279720653878	-0.5483096967109079	0.5834792703519462	0.9994548905052463	AL645608.6
ENSG00000230699	52.673776148274634	-0.1882448967938286	0.5679745003108082	-0.3314319510661426	0.7403182339274508	0.9994548905052463	AL645608.2
ENSG00000223764	245.41258407625622	0.9066635712727337	0.5112169174946202	1.7735398423747881	0.07613929595621088	0.9994548905052463	LINC02593
ENSG00000187634	411.27852756388	0.6371364570105501	0.5175048982123898	1.2311699062393457	0.2182593242387325	0.9994548905052463	SAMD11
ENSG00000188976	2482.965679832079	0.1441356345107119	0.21762865128958414	0.662300821406645	0.5077784505241801	0.9994548905052463	NOC2L
ENSG00000187961	390.56755327789614	-0.00323235736115405	0.21542495287504065	-0.01500456338978062	0.9880285397343452	0.9994548905052463	KLHL17
ENSG00000187583	260.17339659002255	-0.09237186003263151	0.2556039163592484	-0.3613867164022788	0.7178103767627333	0.9994548905052463	PLEKHN1
ENSG00000187642	63.81893261367405	-0.442859688581099	0.2796459439930816	-1.5836309389456624	0.1132777290290355	0.9994548905052463	PERM1
ENSG00000272512	131.00944851304342	-0.42984647099165607	0.23027460025489904	-1.8666690573595348	0.06194781735278559	0.9994548905052463	AL645608.7
ENSG00000188290	739.9502207473055	0.09044688792154244	0.2724114613835483	0.33202306342828786	0.7398718438484879	0.9994548905052463	HES4
ENSG00000187608	2333.3565113350464	0.8524961546729348	0.6038013649023506	0.4118817946209914	0.15798476905936776	0.9994548905052463	ISG15
ENSG00000188157	9221.70215665529	0.013452840947534949	0.2310493986444189	0.05822495547040416	0.9535694428479847	0.9994548905052463	AGRN
ENSG00000237330	102.61186141681644	0.40291513867368683	0.3476934332928913	1.1588229747620158	0.24652835061387723	0.9994548905052463	RNF223
ENSG00000131591	992.9512960611743	-0.0427148424432582	0.17161957343472822	-0.24889260349725706	0.8034438568559497	0.9994548905052463	C1orf159
ENSG00000223823	16.854810532767498	-0.23841200421772799	0.41861131170997845	-0.5695307258751388	0.5689960251837238	0.9994548905052463	LINC01342
ENSG00000207730	56.09941244837563	-0.10683288353244147	0.23619766760702762	-0.45230287248299167	0.6510508068822578	0.9994548905052463	MIR200B
ENSG00000207607	17.303058481472032	-0.2507722703855033	0.29500068859327133	-0.8500735085783228	0.39528421897462573	0.9994548905052463	MIR200A
ENSG00000198976	36.72149427584732	-0.288373924636681	0.2546520318232229	-1.1324234194089113	0.2574564634949995	0.9994548905052463	MIR429
ENSG00000272141	223.3026495191957	-0.09798886741244443	0.2198889363424928	-0.44562891176943864	0.6558653226522789	0.9994548905052463	AL390719.2
ENSG00000205231	16.219972717075026	-0.2808659831947273	0.34765282899274474	-0.80789212620096	0.4191526843935721	0.9994548905052463	TTL10-AS1
ENSG00000162571	41.15696208822165	-0.1803425338328894	0.34524734589598877	-0.5223574807356244	0.6014214545799876	0.9994548905052463	TTL10
ENSG00000186891	148.86493001758876	0.11046300203831237	0.3874496419438831	0.2851028626174636	0.775565336977963	0.9994548905052463	TNFRSF18
ENSG00000186827	286.58983581534426	0.12423804378535255	0.36074041612473523	0.343973512033486	0.7305474806066343	0.9994548905052463	TNFRSF4
ENSG00000078808	4428.178891382373	-0.01424077882151082	0.2243434604834544	-0.06347757492383495	0.9493862158429063	0.9994548905052463	SDF4
ENSG00000176022	796.8778612763706	0.2138380793989577	0.20745197336860594	1.0307335935665428	0.3026423389545133	0.9994548905052463	B3GALT6
ENSG00000184163	55.68305129563981	-0.07594687765029647	0.291707227405161	-0.2603530886974273	0.7945914252266418	0.9994548905052463	C1QTNF12
ENSG00000260179	32.12794291003825	-0.3204637572812076	0.3102419803997769	-1.0329477553884194	0.3016283486012007	0.9994548905052463	AL162741.1
ENSG00000160087	1293.72186738211	0.1305088446484648	0.15651222319395017	0.8338572028763406	0.4043614563138127	0.9994548905052463	UBE2J2
ENSG00000230415	48.83476702661487	0.06967914109129812	0.24499781837368081	0.2844071900469767	0.7760983495553586	0.9994548905052463	LINC01786

Şekil 4.9. GSE71651 kodlu çalışmanın diferansiyel ifade analizi ve anotasyonu.

Elde edilen listelerde pozitif log2FoldChange değerleri, normal dokuya göre tümör dokusunda ifadesi artan genleri; negatif değerler ise ifadesi azalan genleri göstermektedir.

- *Devre Analiz Programı için Verilerin Hazırlanması*

Yeni Nesil Dizileme çalışmalarından elde edilen sonuçlar, mikrodizin çalışmalarında olduğu gibi mRNA, miRNA, lncRNA ve TF verilerini birlikte içermektedir. Benzer şekilde, devre analiz programı kütüphanelerinden elde edilen listeler ile analiz sonucunda elde edilen gen listeleri kesiştirilerek düzenleyici moleküllere ait ayrı dosyalar oluşturulmuştur. Sonuçta her bir düzenleyici molekül listesi birleştirilerek NGS verilerine ait IDC ve normal dokular arasında ifadesi anlamlı şekilde değişen tek bir mRNA, miRNA, lncRNA ve TF listesi oluşturulmuştur. miRNA dosyalarındaki isimlendirme problemi, mikrodizin çalışmalarında olduğu gibi “hsa” ekinin kaldırılarak “miR” eki ile başlaması ile tek isimlendirme sistemine uygun hale getirilmesiyle ortadan kaldırılmıştır.

4.2.3. DEVRE ANALİZ PROGRAMININ GELİŞTİRİLMESİ

4.2.3.1. Veri Tabanlarının Belirlenmesi

Regülatif devreleri ortaya koyması için tasarlanan programda, bu devrelerin tahmini için gerekli en önemli basamağı moleküllerin birbirleri ile etkileşimi oluşturmaktadır. Başka bir

deyişle, moleküllerin birbirine karşı “aktivasyon” veya “baskılama” etkilerinin belirlenmesi sonucu bu devrelerin kurgulanması mümkün olmuştur (Şekil 2.12.).

Regülasyona yönelik örnek gösterim aşağıdaki şekildedir:

TF → mRNA *TF’in mRNA’nın ifadesini arttırması,*

miRNA ---| lncRNA *miRNA’nın lncRNA’nın ifadesini baskılaması.*

Bu etkinin belirlenebilmesi, programa verilen her bir ifade dosyasının (mRNA, miRNA, lncRNA ve TF dosyası) birbiri ile birleştirilmesi sonucu teoride olabilecek tüm ikili ilişkilerin tahmini esasına dayanmaktadır (TF – mRNA, TF – miRNA, TF – lncRNA, miRNA – lncRNA, miRNA – TF, miRNA – mRNA gibi). Söz konusu işlem basamağı hesaplamalı bir yaklaşım olup tamamen teoride olabilecek ikili ilişkileri göstermektedir. Regülasyonda rol oynayan gerçek ve güvenilir devrelerin ortaya çıkarılması için ise, teoride oluşturulan ikili ilişkilerin literatür verisi ile kesiştirilmesi gerekmektedir. Bunun için günümüzde kullanılan çeşitli veri tabanları mevcuttur ve çalışma kapsamında 11 farklı veri tabanı (*miRcode*, *DIANA TarBase*, *LncCeRBase*, *LncTarD*, *miRDB*, *miRTarBase*, *TFLink*, *TRRUST*, *RNA Interactome Database*, *TargetScan*, *TransmiR*) kullanılmıştır. Ayrıca miRNA’lar ve TF’lerin güncel molekül listelerinin kontrolü için sırasıyla *miRBase* (150,151) ve *Toronto The Human Transcription Factors* (152) veri tabanları incelenmiştir. Çalışma kapsamında kullanılan veri tabanları ve moleküller arası etkileşim bilgileri Çizelge 4.2.’de verilmiştir. Belirtilen veri tabanlarının sunduğu ikili ilişki listeleri devre analiz programının içerisine aktarılarak bir literatür kütüphanesi oluşturulmuştur. Programın analiz sırasında bulacağı teorik ikililer, literatürde yer alan bu ilişkiler ile otomatik olarak kesiştirilmekte ve devrelerin oluşturulması için ilk adım tamamlanmaktadır. Başka bir deyişle, devre analiz programı TF, mRNA, miRNA ve lncRNA’dan oluşan regülatif ilişki tahminlerini ortaya çıkarmakta ve daha sonra bu olasılıkları 11 farklı kaynaktan elde ettiği literatür verileri ile karşılaştırarak gerçek regülatif ilişkileri ortaya koymaktadır (Şekil 4.11.).

Çizelge 4.2. Devre analiz programında kullanılan veri tabanları ve moleküller arası etkileşim bilgileri.

VERİ TABANLARI	VERİ KAYNAĞI VE KAPSAMI	MOLEKÜLER ETKİLEŞİM	TOPLAM TRANSKRİPT	ETKİLEŞİM SAYISI	YÖNTEM
miRcode	GENCODE anotasyonuna dayalı olarak, 10.000'den fazla uzun kodlamayan RNA geni ve kodlayıcı transkript içeren insan mikroRNA hedef tahminlerini sağlamaktadır.	miRNA - lncRNA miRNA - mRNA	10.419 lncRNA 19.999 mRNA miRNA'lar bilinmiyor	179.905	Tahmin yaklaşımı
DIANA TarBase v7.0	Yarım milyondan fazla yüksek kaliteli, manuel olarak incelenmiş miRNA:gen etkileşimi sunmaktadır ve bu etkileşimler, deney sonuçları, metodolojiler ve koşullar gibi ayrıntılı meta-verilerle zenginleştirilmiştir.	miRNA - mRNA	Bilinmiyor	~ 65.000	Deneyisel veri
LncCeRBase	lncRNA-miRNA-mRNA etkileşim verilerini bir araya getiren kapsamlı bir kaynak olup, lncRNA'ların düzenleyici mekanizmalarına dair değerli bilgiler sunmaktadır.	lncRNA - miRNA - mRNA	130 lncRNA 245 mRNA 214 miRNA	432	Deneyisel veri
LncTarD 2.0	Deneyisel olarak desteklenmiş lncRNA-hedef düzenlemelerini, bunların etkilediği fonksiyonları ve insan hastalıklarında lncRNA aracılı düzenleyici mekanizmaları sunmaktadır.	lncRNA - miRNA lncRNA - mRNA	1.355 lncRNA 1.743 mRNA 506 miRNA	8.360	Deneyisel veri
miRDB	miRNA hedeflerinin tahmini ve işlevsel anotasyonları için geliştirilmiş bir çevrimiçi veritabanıdır ve gen düzenlemesi ile hastalık mekanizmalarındaki miRNA rollerini inceleyen araştırmacılar için bilgiler sunmaktadır.	miRNA - mRNA	2.656 miRNA 1.610.510 mRNA	1.642.327	Tahmin yaklaşımı
miRTarBase	Deneyisel olarak doğrulanmış mikroRNA-hedef etkileşimlerinin en büyük ve en güncel koleksiyonunu sunan kapsamlı bir veritabanıdır ve bu etkileşimler çeşitli deneyel yöntemlerle doğrulanmıştır.	miRNA - mRNA	4.296 miRNA 23.426 mRNA	~ 360.000	Deneyisel veri
TFLink	insanlar ve altı model organizma için transkripsiyon faktörü-hedef gen etkileşimleri, bağlanma bölgesi dizileri ve genomik konumlar hakkında bilgiler sağlamaktadır.	TF - mRNA	1.606 TF 20.139 mRNA	6.739.357	Deneyisel veri ve literatür verisi
TRRUST version 2	İnsanlar ve fareler için transkripsiyonel düzenleyici ağları sunan, manuel olarak derlenmiş bir veritabanıdır.	TF - mRNA	800 TF 2.067 mRNA	8.444	Deneyisel veri ve literatür verisi
RNAInter v4.0 RNA Interactome Database	Hücrel RNA'ların düzenleyici yapısını araştırmak için tasarlanmış, kapsamlı ve kolay erişilebilir bir RNA etkileşim platformudur.	RNA - RNA miRNA - TF lncRNA - TF	Bilinmiyor	4.452.093	Deneyisel veri ve hesaplamalı tahmin
TargetScan	Korunan bölgeleri arayarak miRNA'ların biyolojik hedeflerini tahmin eden bir veritabanıdır.	miRNA - mRNA	Bilinmiyor	718.234	Tahmin yaklaşımı
TransmiR v2.0	Transkripsiyon faktörü (TF)-mikroRNA (miRNA) düzenlemelerine yönelik bir veritabanıdır ve kullanıcıların TF'ler ile miRNA'lar arasındaki düzenleyici ilişkilerini ortaya koymaktadır.	TF - miRNA	~ 623 TF ~785 miRNA	~ 1.789.728	Deneyisel veri ve literatür verisi

4.2.3.2. Veri Tabanlarından Elde Edilen Listelerin Programa Entegre Edilecek Şekilde Düzenlenmesi, İsimlendirme Sorunlarının Giderilmesi ve Kütüphanelerin Hazırlanması

Programa entegre edilecek veri tabanları Çizelge 4.2.'de verildiği gibi farklı etkileşimleri içermektedir. Farklı veri tabanlarından elde edilen bu listeler, regülasyona dair pek çok veri içermekle birlikte farklı formatlarda oluşturulmuştur. Listelerin programa entegre edilebilmesi için her listenin aynı formatta düzenlenmesi, yalnızca regülasyona katılan moleküllerin isimlerini içeren sütunların kalacağı şekilde listelerin filtrelenmesi, farklı dosyalarda aynı bilgiyi içeren sütun isimlerinin aynı başlık ile adlandırılması ve son hallerinin kütüphane şeklinde hazırlanması gerekmektedir. Ayrıca aynı tip molekülün isimlendirmesi her dosyada aynı ve tek bir formatta olmalıdır. Örneğin miRNA'lar için isimlendirme bir dosyada miR kodu ile, başka bir dosyada hsa-miR kodu ile isimlendirildiğinde sorun çıkmaktadır ve program geliştirmeye başlamadan önce aşılması gereken önemli sorunlardan bir tanesini oluşturmaktadır. Tüm bu işlemler excel üzerinden tamamlandıktan sonra listeler Python 3.12'de okunmuş ve veri tabanları ile isimlendirilmiş kütüphane hazırlığı tamamlanmıştır. Oluşturulan kütüphanelerden lncRNA-miRNA, miRNA-mRNA ve TF-mRNA kütüphanelerinin örnek ekran görüntüsü Şekil 4.10.'da verilmiştir.

lncRNA	miRNA	miRNA	mRNA	TF	mRNA
ADNCR	miR-204	miR-196a-5p	CYP26B1	AHR	ABCG2
AF147447	miR-34c	miR-103a-3p	CYP26B1	AHR	BRCA1
AK015322	miR-19b-3p	miR-195-5p	CYP26B1	AHR	CA9
AK048794	miR-592	miR-107	CYP26B1	AHR	CCND1
APF	miR-188-3p	miR-16-5p	CYP26B1	AHR	CCNG2
ATB	miR-141-3P	miR-424-5p	CYP26B1	AHR	CYP1A1
ATB	miR-200	miR-15a-5p	CYP26B1	AHR	CYP1A2
ATB	miR-200c	miR-15b-5p	CYP26B1	AHR	CYP1B1
BC032469	miR-1207-5p	miR-181a-5p	USP28	AHR	CYP2B6
BC048612	miR-203	miR-505-5p	USP28	AHR	IL13
CASC2	miR-181a	miR-423-5p	USP28	AHR	IL1B
CASC2	miR-18a	miR-19b-3p	STARD3NL	AHR	IL6
CASC2c	miR-101	miR-486-5p	STARD3NL	AHR	MFSD2A
CCAT1	miR-155	miR-124-3p	STARD3NL	AHR	MT2A
CCAT1	miR-7	miR-1-3p	STARD3NL	AHR	PCNA
CCAT1	miR-218	miR-19a-3p	STARD3NL	AHR	RFC3
CCAT1	miR-490	miR-378b	CD74	AHR	UGT1A1
CCAT1	miR-181a	miR-191-5p	CD74	AHR	UGT1A6
CCR49	let-7	miR-335-5p	CD74	AHR	NTRK2
CHRF	miR-489	miR-95-3p	CD74	AHR	AGTR1
CRNDE	miR-136	miR-331-5p	RANBP9	AHR	CDK8
CTNNAP1	miR-141	miR-374a-5p	RANBP9	AHR	SETMAR
DANCR	miR-214	miR-19b-1-5p	RANBP9	AHR	EMC4
DANCR	miR-320a	miR-152-3p	RANBP9	AHR	HNRNPUL1
ECONEXIN	miR-411-5p	miR-148a-3p	RANBP9	AHR	GPR137C
EWSAT1	miR-326-5p	miR-148b-3p	RANBP9	AHR	CALHM1
EWSAT1	miR-330-5p	miR-19b-3p	RANBP9	AHR	RHOQ
FER1L4	miR-106a-5p	miR-410-3p	RANBP9	AHR	NUDCD2
FTH1P3	miR-224-5p	miR-19a-3p	RANBP9	AHR	KTN1
FTX	miR-374a	miR-101-3p	RANBP9	AHR	HSPA5
FTX	miR-545	miR-106a-5p	RANBP9	AHR	COL18A1

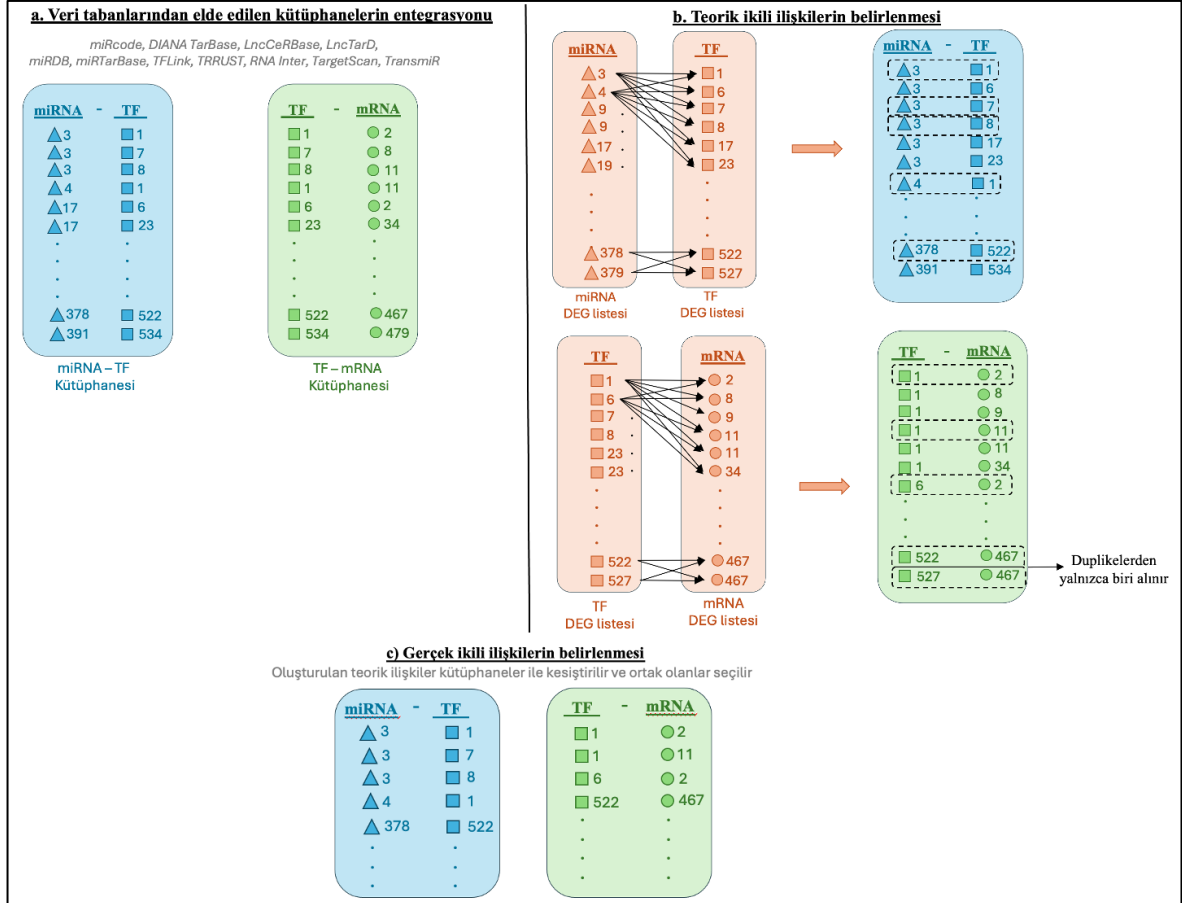
Şekil 4.10. lncRNA-miRNA, miRNA-mRNA ve TF-mRNA kütüphanelerinin örnek gösterimi.

4.2.3.3. Kütüphanelerin Devre Analiz Programına Entegrasyonu ve Analiz Edilecek Dosyaların Programa Aktarılması Yoluyla Teorik ve Gerçek İkili İlişkilerin Belirlenmesi

Tez kapsamında kullanılan ve Çizelge 4.2.'de listelenen veri tabanları kullanılarak oluşturulan kütüphaneler, Python 3.12 yazılım dili kullanılarak geliştirilen yazılım programına entegre edilmiştir. Bu kütüphaneler, teorik olarak önerilen tüm regülatif ikili ilişkilerin birbirleri ile olan ilişkilerini doğrulamak amacıyla kontrol dosyaları olma işlevini görmektedir.

Analize girecek olan GEO'dan elde edilmiş ve IDC/normal arasında farklılık gösteren mRNA, miRNA, lncRNA ve TF dosyaları Python 3.12 platformuna aktarılmış ve her bir molekül için belirli bir kat değişim eşik değeri (fold change threshold) 1.5 kat olarak belirlenmiştir. Mutlak log2FoldChange değeri 1.5 katın altında kalan moleküller analiz listesinden çıkarılmıştır. Log2FoldChange değerlerine göre filtreleme işleminin ardından listeler birbirleri ile birleştirilmiştir. Birleştirme işlemi, her bir dosyanın tüm satırlarının

diğer dosyanın tüm satırları ile ikili kombinasyonlar oluşturmasını sağlamaya yönelik olarak tasarlanmıştır. Başka bir deyişle, bu yöntemle ikili ilişki gösterebilecek tüm moleküler kombinasyonlar oluşturulmuş ve bu kombinasyonlar, ilgili kütüphanelerle aynı isim ve formatta, tüm teorik ilişkileri belirlemek amacıyla kullanılmıştır. Yapılan işlemlere yönelik şematik gösterim Şekil 4.11.'de verilmiştir.



Şekil 4.11. Kütüphanelerin devre analiz programına entegrasyonu ile teorik ve gerçek ikili ilişkilerin belirlenmesine yönelik şematik gösterim.

Sonuçta elde edilen tüm teorik ikili ilişkiler, programa entegre edilmiş olan aynı isim ve formattaki kütüphane dosyaları ile karşılaştırılmış ve ortak olan ikili ilişkiler belirlenerek devre analizine hazır dosyalar haline getirilmiştir.

4.2.3.4. Düzenleyici Devrelerin Oluşturulması

Devre analizi için hazırlanan dosyalar *miRNA-mRNA*, *miRNA-TF*, *TF-miRNA*, *TF-mRNA*, *TF-lncRNA*, *miRNA-lncRNA*, *lncRNA-mRNA*, *lncRNA-TF*, *lncRNA-miRNA* regülatif ilişkilerini içermektedir. Her bir devre, iki regülatif ilişki listesinin ortak molekül türü esas

alınarak listelerin birleştirilmesi ve regülasyon yönünün belirlenmesi esasına dayanmaktadır. Bu kapsamda her bir birleştirme işlemine regülatif listelerin dahil edilmesi ve farklı moleküllerin eklenmesi ile toplamda 8 farklı devre türü elde edilmiştir. Tüm analizler sonucunda ortaya çıkan devreler aşağıda verilmiştir:

- Devre 1: TF, miRNA, mRNA düzenleyici devreleri. Bir miRNA'nın aynı anda baskıladığı TF'leri ve mRNA'ları gösteren devrelerdir.

TF |----- miRNA -----| mRNA

- Devre 2: TF, miRNA, mRNA düzenleyici devreleri. Bir miRNA'nın aynı anda baskıladığı TF'ler ve mRNA'lar ile TF'lerin pozitif regüle ettiği miRNA'ları gösteren devrelerdir.

TF |----- miRNA -----| mRNA
----->

- Devre 3: TF, miRNA, mRNA düzenleyici devreleri. Bir miRNA'nın aynı anda baskıladığı TF'ler ve mRNA'lar ile TF'lerin regüle ettiği miRNA'lar ve mRNA'ları gösteren devrelerdir.

----->
TF |----- miRNA -----| mRNA
----->

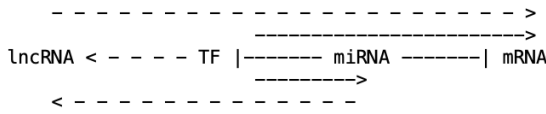
- Devre 4: lncRNA, TF, miRNA, mRNA düzenleyici devreleri. Üçlü devrelere ek olarak lncRNA'ları regüle eden TF'leri gösteren devrelerdir.

lncRNA < - - - TF |----- miRNA -----| mRNA
----->

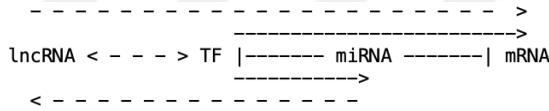
- Devre 5: lncRNA, TF, miRNA, mRNA düzenleyici devreleri. Üçlü devrelere ek olarak lncRNA'ları regüle eden TF'leri ve miRNA'ları gösteren devrelerdir.



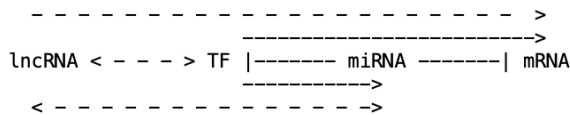
- Devre 6: lncRNA, TF, miRNA, mRNA düzenleyici devreleri. Üçlü devrelere ek olarak lncRNA'ları regüle eden TF'leri ve miRNA'ları gösteren; aynı zamanda lncRNA'ların da mRNA'ları regüle ettiği devrelerdir.



- Devre 7: lncRNA, TF, miRNA, mRNA düzenleyici devreleri. Üçlü devrelere ek olarak lncRNA'ları regüle eden TF'leri ve miRNA'ları gösteren; aynı zamanda lncRNA'ların da mRNA'ları ve TF'leri regüle ettiği devrelerdir.



- Devre 8: lncRNA, TF, miRNA, mRNA düzenleyici devreleri. Üçlü devrelere ek olarak lncRNA'ları regüle eden TF'leri ve miRNA'ları gösteren; aynı zamanda lncRNA'ların da mRNA'ları, TF'leri ve miRNA'ları regüle ettiği devrelerdir.



Devrelerin gösterimi lineer şekilde olup ok işaretlerinin anlamları aşağıdaki gibidir:

-----> : Pozitif regülasyon.

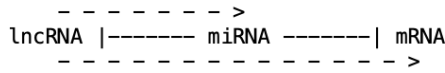
-----| : Negatif regülasyon.

- - - - -> : İlişkinin yönü belli; fakat regülasyon negatif veya pozitif olabilir.

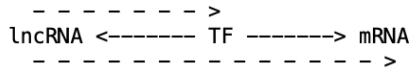
<- - - -> : Her iki molekül de birbirini regüle eder; fakat regülasyon negatif veya pozitif olabilir.

Devre analizi sonucunda elde edilen tüm devrelere ek olarak, *miRNA bağımlı lncRNA düzenleyici devreleri* ile *TF bağımlı lncRNA düzenleyici devreleri* olmak üzere iki farklı devre türü oluşturulmaktadır:

- *miRNA bağımlı lncRNA düzenleyici devreleri:* lncRNA, miRNA, mRNA moleküllerinin birbirleri üzerindeki regülatif etkilerini esas alarak oluşturulan devrelerdir.



- *TF bağımlı lncRNA düzenleyici devreleri:* lncRNA, TF, mRNA moleküllerinin birbirleri üzerindeki regülatif etkilerini esas alarak oluşturulan devrelerdir.



Analiz sonucunda elde edilen tüm devreler .txt formatında oluşturulmaktadır. Buna ek olarak, ileri analizler için her bir devre dosyasının .csv formatında tablolaştırılmış bir versiyonu program içerisine kaydedilmektedir.

4.2.3.5. GSEA Zenginleştirme Analizi

Gene Set Enrichment Analysis (GSEA), iki farklı biyolojik durumdan elde edilen ifade profilleri arasındaki anlamlı farklılıkları belirlemek için kullanılan bir hesaplama yöntemidir. GSEA, genler arasındaki bireysel değişiklikleri incelemek yerine, gen setlerindeki genel eğilimlere odaklanarak daha kapsamlı bir analiz sunmaktadır. Bu yöntem, gen setlerindeki koordineli değişiklikleri değerlendirerek ortaya çıkan küçük değişimlerin önemini hesaplamakta ve böylece biyolojik anlam taşıyan sonuçlara ulaşmayı sağlamaktadır (153,154).

The Molecular Signatures Database (MSigDB) ise, GSEA yazılımı ile birlikte kullanılan, anotasyonları yapılmış on binlerce gen setini içeren bir veri tabanıdır. MSigDB, özellikle zenginleştirme analiz yöntemleri için gen setleri sunmak amacıyla tasarlanmış ve bu nedenle GSEA yazılımına sorunsuz şekilde entegre edilmiştir. Ayrıca gen seti kaynakları ve türleri bakımından geniş bir kapsama sahiptir. Bu kaynaklar içerisinde *GO*, *KEGG* ve *TRANSFAC* gibi özel kaynaklardan üretilmiş setler bulunmaktadır ve toplamda en büyük gen seti sayısına sahiptir. Bu yönüyle MSigDB, gen setlerinin biyolojik anlamlarını ortaya koymak ve kapsamlı analizler yapmak için geniş bir kaynak sunmaktadır (155,156).

Tez çalışması kapsamında, 4. tür devre olarak tanımlanan lncRNA, TF, miRNA, mRNA düzenleyici devrelerine Gen Seti Zenginleştirme Analizi (GSEA) uygulanmıştır. Bu analiz, devre analiz programına entegre edilmiş Python kütüphanesi olan “*gseapy*” (versiyon 1.1.3) ile gerçekleştirilmiştir (157). Analizin gerçekleştirilmesi için devrelerin .csv formatındaki dosyalarından elde edilen mRNA bilgilerini içeren sütun kullanılmış ve *prerank* fonksiyonu ile analiz gerçekleştirilmiştir. Fonksiyonun çalışabilmesi için gerekli olan parametreler; genlerin zenginleşeceği gen seti *MSigDB_Hallmark_2020*, *permütasyon sayısı 1000*, *minimum gen sayısı 2* ve *maksimum gen sayısı 5000* olarak belirlenmiştir. Elde edilen sonuçlar içerisinde FDR q-val (False Discovery Rate) <0.25 kriterine uyan genler seçilmiştir (154). Bu genler normalize edilmiş zenginleşme skoru NES’e (Normalized Enrichment Score) göre büyükten küçüğe doğru sıralanmıştır. Son aşamada, genlerin zenginleştiği yollar, NES (Normalized Enrichment Score), NOM p-val (Nominal p-Value), FDR q-val (False Discovery Rate) ve zenginleşen genlerin bulunduğu tablolar .csv formatında kaydedilmiştir.

Zenginleştirme analizi sonuçlarının görselleştirilmesi amacıyla, Python’ın “*gseapy*” (versiyon 1.1.3) (158) ve “*matplotlib*” (versiyon 3.9.1) (159) kütüphaneleri kullanılarak GSEA (Gene Set Enrichment Analysis) grafiği ve nokta grafiği oluşturulmuştur. Ayrıca, gen zenginleştirme analizi sonuçlarının ağ yapısında daha ayrıntılı şekilde görselleştirilmesi için Python’ın “*gseapy*”, “*matplotlib*” ve “*networkx*” (versiyon 3.3) (160) kütüphanelerinden yararlanılmıştır. Tüm grafikler ve görseller .png formatında kaydedilmiştir.

Yapılan kapsamlı zenginleştirme ve yolak analizleri, diğer tüm analizler gibi hem mikrodizin hem de NGS veri setlerine uygulanmıştır. Her iki veri türü, 4. tür olarak adlandırılan

devrelerin analizine uygunluk göstermiştir. Bu uygunluk, her iki veri setinde de yeterli miktarda mRNA verisinin bulunması ile sağlanmış ve sonuçların birbirleri ile karşılaştırılarak kapsamlı bir şekilde analiz edilmesine olanak tanımıştır.

4.2.4. DEVRE ANALİZ PROGRAMININ ARAYÜZ HALİNE GETİRİLMESİ

Geliştirilen devre analiz programının kullanıcı etkileşimini kolaylaştırmak ve işlevselliğini arttırmak amacıyla görsel bir arayüz tasarlanmıştır. Bu arayüz, Python programlama dili kullanılarak “*PyQt5*” (versiyon 5.15.11) (161)kütüphanesi ile geliştirilmiştir. Söz konusu arayüz, kullanıcıların mRNA, miRNA, lncRNA ve TF verilerine ilişkin ifade dosyalarını, gen isimleri ve ifade değerlerini içeren uygun formatlarda sisteme yüklemeleri ile çalışmaktadır. Dosya yükleme işlemlerinin tamamlanmasının ardından kullanıcıların “*threshold seçme*” bölümünden her bir dosya için araştırılan iki biyolojik koşula uygun eşik değerini belirlemeleri istenmektedir. Dosya seçme ve eşik değeri belirleme işlemlerinin ardından sistem analize hazır hale gelmektedir (Şekil 4.12.).

miRcuit DEVRE ANALİZ PROGRAMI

miRcuit Devre Analiz Programı

miRcuit Devre Analiz Programı miRNA, lncRNA ve TF'lerin mRNA'lar üzerindeki düzenleyici etkilerini analiz eden bir programdır. miRcuit, analiz sonuçlarını *miRcode*, *DIANA TarBase*, *LncCeRBase*, *LncTarD*, *miRDB*, *miRTarBase*, *TFLink*, *TRRUST*, *RNAInter*, *TargetScan*, *TransmiR* kütüphaneleri ile karşılaştırarak moleküler etkileşimlerin olası yapısını ortaya koyar ve biyolojik süreçlerin daha iyi anlaşılmasına katkıda bulunur.

Programı çalıştırmadan önce analiz için gerekli dosyaları seçmeniz gerekmektedir. Dosyalar seçildikten sonra, belirlediğiniz threshold değerleri ve diğer parametreler kullanılarak analiz başlayacaktır.

Dosya Seçme

Araştırma sonucunda elde ettiğiniz mRNA, miRNA, lncRNA ve TF ifade dosyalarını *Gen sembolü* ve *log2FoldChange* değerlerini içerecek şekilde csv formatında sisteme yükleyiniz.

mRNA dosyası seçiniz. Örnek dosya formatı için [tıklayınız](#)

Dosya Seç

miRNA dosyası seçiniz. Örnek dosya formatı için [tıklayınız](#)

Dosya Seç

lncRNA dosyası seçiniz. Örnek dosya formatı için [tıklayınız](#)

Dosya Seç

TF dosyası seçiniz. Örnek dosya formatı için [tıklayınız](#)

Dosya Seç

Threshold Belirleme

Araştırma gruplarınız arasında olmasını istediğiniz kat değişimini tüm dosyalar için seçiniz.

mRNA Threshold: ☒ 1.0

miRNA Threshold: ☐ 1.5

lncRNA Threshold: ☐ 2.0

TF Threshold: ☐ 2.5

GSEA Zenginleştirme Analizi

Veri setinize GSEA veri tabanı üzerinden MSigDB zenginleştirme analizi yapmak için gerekli kutucuğu işaretleyiniz.

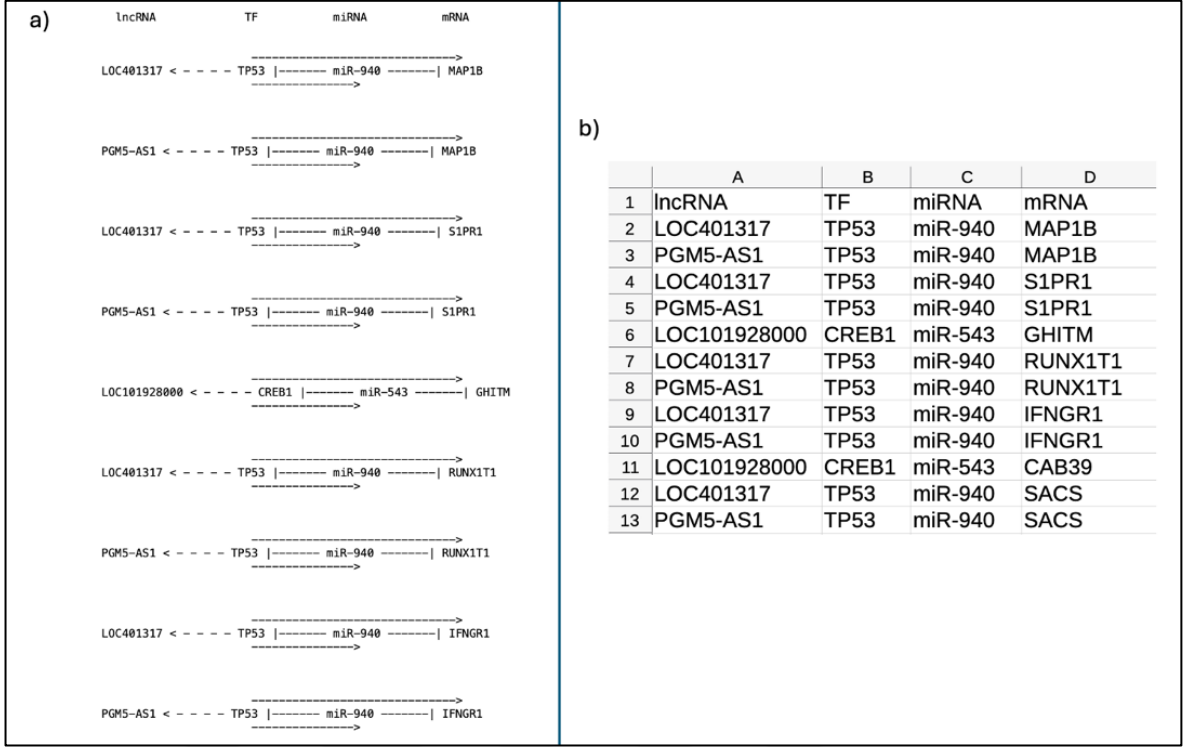
☐ GSEA zenginleştirme analizi yapılsın

☒ GSEA zenginleştirme analizi yapılsın

Analizi Başlat

Şekil 4.12. Geliştirilen devre analiz programının kullanıcı arayüzü.

Analiz sonucunda program, 8 farklı türde devre ile *miRNA* bağımlı *lncRNA* düzenleyici devreler ve *TF* bağımlı *lncRNA* düzenleyici devreler olmak üzere 2 tür üçlü devre oluşturmaktadır. Oluşturulan devreler .txt formatında sunulmakta olup her bir devrenin moleküllerini içeren ve ileri analizler için düzenlenmiş .csv formatındaki tablolar kullanıcıya sunulmaktadır (Şekil 4.13.).



Şekil 4.13. Oluşturulan devrelerin txt formatındaki yapısı (a) ve bu devrelerin csv formatında tablolatırılmış versiyonlarına (b) ait örnek ekran görüntüsü.

Analizin başlatılmasının ardından, otomatik olarak düzenlenen dosyalama sistemi, kullanıcıların gerekli verilere kolayca ulaşabilmesini sağlamak üzere tasarlanmıştır. Devreler, “*Circuits*” başlığı altında bulunmakta; farklı türdeki devreler “*Total_analysis_circuits*” başlığı altında yer alırken, özelleştirilmiş üçlü devreler ise kendi başlıkları altında yer almaktadır. Tablo formatındaki dosyalar ise, “*Molecular_interactions*” klasörü altında aynı alt klasör isimleri ile sunulmaktadır.

Programda, kullanıcılara ek analiz imkanı sunan bir “*GSEA analizi*” bölümü mevcuttur. Bu bölüm, varsayılan olarak “*GSEA analizi yapılmasın*” seçeneği ile işaretlenmiş olup kullanıcıların tercihiye göre devre analizlerine ek olarak GSEA analizi yapacak şekilde tasarlanmıştır (Şekil 4.12.). GSEA analiz sonuçları, “*GSEA_Results*” başlıklı ayrı bir klasörde sunulmaktadır. Bu sonuçlar, varsayılan olarak GSEA kütüphanesinin ürettiği .csv formatındaki “*prerank_report*” dosyasını ve “*prerank*” klasörü altındaki yolak grafiklerini içermektedir. Ek olarak, program koduna entegre edilmiş analizler ve grafikler sonucunda, $FDR < 0.25$ olan en anlamlı yolakların “*Top_Gene_Sets*” isimli .csv formatındaki tablosu ile “*GSEA_Plots*” klasörü altında yer alan ek GSEA grafiği, nokta grafiği ve ağ analizi sonuçları verilmektedir.

Programın ismi “*miRNA*” ve “*Circuits*” kavramlarının birleştirilmesi ile “***miRCuit***” olarak belirlenmiştir. Oluşturulan lokal program, “.py” dosyası formatında GitHub platformuna yüklenmiştir.

Programın GitHub bağlantısı: <https://github.com/miRcuit/miRcuit>.

4.2.5. TCGA VERİ ANALİZİ

miRCuit kullanılarak elde edilen devrelerin üyeleri arasındaki etkileşimleri doğrulamak amacıyla, The Cancer Genome Atlas (TCGA) meme kanseri veri analizi Xena Browser (162) kullanılarak gerçekleştirilmiştir.



5. ARAŞTIRMA BULGULARI

Tez çalışması kapsamında IDC ve normal/eşli normal dokular arasındaki gen ifade değişikliklerini incelemek amacıyla mikrodizin ve Yeni Nesil Dizileme (NGS) teknolojilerinden elde edilen veriler analiz edilmiştir. Elde edilen veriler doğrultusunda diferansiyel olarak ifade edilen gen listeleri (DEG listeleri) oluşturulmuştur. Bu analizlerde, gen ifadelerindeki değişikliğin anlamlılığı p değeri kullanılarak istatistiksel olarak değerlendirilmiştir.

Yeni Nesil Dizileme teknolojileri kullanılarak elde edilen verilerde yanlış keşif oranı (FDR) değerine göre filtreleme yapıldığında anlamlı gen sayısında büyük ölçüde azalma gözlenmiştir ve özellikle devre analiz programının verimli bir analiz gerçekleştirebilmesi için yeterli sayıda anlamlı gen elde edilememiştir. Bu nedenle analizlerde kullanılan p değeri, anlamlı genleri belirlemede daha geniş bir gen kümesi sunarak analizin kapsamını arttırmıştır.

Ayrıca her iki teknoloji (mikrodizin ve NGS) için veri analiz süreçlerinin tutarlı olmasını sağlamak adına, mikrodizin verilerinde de anlamlı genlerin belirlenmesinde p değeri temel alınmıştır. Bu yaklaşım, her iki veri setinin de aynı istatistiksel kriterlere göre değerlendirildiği bir analiz ortamı sağlamıştır. Böylelikle, IDC ve normal dokular arasındaki gen ifade farklılıkları tutarlı bir şekilde belirlenmiştir.

5.1. MİKRODİZİN VERİ SETLERİNİN DİFERANSİYEL GEN İFADE ANALİZLERİ VE İSTATİSTİKSEL ANLAMLILIK DEĞERLENDİRMESİ

Mikrodizin verilerinin diferansiyel ifade analizleri sonucunda, gen ifadelerindeki farklılıkların istatistiksel olarak anlamlılığını değerlendirmek amacıyla, gen listeleri içerisinde p değeri 0.05'ten küçük olanlar istatistiksel olarak anlamlı kabul edilmiştir. Analizler sonucunda yalnızca miRNA verisini içeren GSE40525 kodlu çalışmaya ait 296, GSE59247 kodlu çalışmaya ait ise 215 adet istatistiksel olarak anlamlı miRNA belirlenmiştir. Diğer transkriptom mikrodizin analizleri sonucunda elde edilen anlamlı moleküllerin sonuçları şu şekildedir: GSE76250 kodlu çalışmaya ait 9062 mRNA, 34 miRNA, 607 lncRNA, 783 TF; GSE45827 kodlu çalışmaya ait 8689 mRNA, 93 miRNA, 175 lncRNA, 765 TF; GSE73540 kodlu çalışmaya ait 2159 mRNA, 39 miRNA, 186

lncRNA, 203 TF; GSE10797 kodlu çalışmaya ait 1198 mRNA, 9 miRNA, 19 lncRNA, 118 TF; GSE22544 kodlu çalışmaya ait 4198 mRNA, 9 miRNA, 163 lncRNA, 396 TF ve GSE59246 kodlu çalışmaya ait 4212 mRNA, 51 lncRNA, 382 TF elde edilmiştir. Sonuçları gösteren özet tablo Çizelge 5.1.'de verilmiştir.

Çizelge 5.1. Mikrodizin verilerinin analizi sonucunda elde edilen diferansiyel olarak anlamlı ifade gösteren moleküller ($p < 0.05$).

Diferansiyel Olarak Anlamlı İfade Gösteren Moleküller ($p < 0.05$)					
GSE Numarası	mRNA	miRNA	lncRNA	TF	Çalışma Türü
GSE40525	x	296	x	x	miRNA-mikrodizin
GSE59247	x	215	x	x	miRNA-mikrodizin
GSE76250	9062	34	607	783	Transkriptom mikrodizin
GSE45827	8689	93	175	765	Transkriptom mikrodizin
GSE73540	2159	39	186	203	Transkriptom mikrodizin
GSE10797	1198	9	19	118	Transkriptom mikrodizin
GSE22544	4198	9	163	396	Transkriptom mikrodizin
GSE59246	4212	x	51	382	Transkriptom mikrodizin
Toplam (Mikrodizin)	13742	534	942	1213	

Sonuçta tüm veri setlerinden elde edilen düzenleyici molekül listeleri birleştirilmiş ve mikrodizin çalışmalarına göre toplamda 13742 mRNA, 534 miRNA, 942 lncRNA ve 1213 TF'in IDC ve normal dokular arasında ifadelerinin anlamlı şekilde değiştiği belirlenmiştir.

5.2. YENİ NESİL DİZİLEME VERİ SETLERİNİN DİFERANSİYEL GEN İFADE ANALİZLERİ VE İSTATİSTİKSEL ANLAMLILIK DEĞERLENDİRMESİ

Yeni Nesil Dizileme verilerinin diferansiyel ifade analizleri sonucunda, gen ifadelerindeki farklılıkların istatistiksel olarak anlamlılığını değerlendirmek amacıyla, gen listeleri içerisinde p değeri 0.05'ten küçük olanlar istatistiksel olarak anlamlı kabul edilmiştir. Analizler sonucunda yalnızca miRNA verisini içeren GSE29173 kodlu çalışmaya ait 156, GSE39162 kodlu çalışmaya ait 181 ve GSE40049 kodlu çalışmaya ait 232 adet istatistiksel olarak anlamlı miRNA belirlenmiştir. Diğer RNA-Seq analizleri sonucunda elde edilen anlamlı moleküllerin sonuçları şu şekildedir: GSE71651 kodlu çalışmaya ait 885 mRNA, 9 miRNA, 233 lncRNA, 59 TF ve GSE110114 kodlu çalışmaya ait 7783 mRNA, 21 miRNA, 1242 lncRNA ve 718 TF elde edilmiştir. Sonuçları gösteren özet tablo Çizelge 5.2.'de verilmiştir.

Çizelge 5.2. Yeni Nesil Dizileme verilerinin analizi sonucunda elde edilen diferansiyel olarak anlamlı ifade gösteren moleküller ($p < 0.05$).

Diferansiyel Olarak Anlamlı İfade Gösteren Moleküller ($p < 0.05$)

GSE Numarası	mRNA	miRNA	lncRNA	TF	Çalışma Türü
GSE29173	x	156	x	x	RNA-Seq (miRNA verisi)
GSE39162	x	181	x	x	RNA-Seq (miRNA verisi)
GSE40049	x	232	x	x	RNA-Seq (miRNA verisi)
GSE71651	885	9	233	59	RNA-Seq
GSE110114	7783	21	1242	718	RNA-Seq
Toplam (NGS)	8171	354	1389	740	

Sonuçta tüm veri setlerinden elde edilen düzenleyici molekül listeleri birleştirilmiş ve NGS çalışmalarına göre toplamda 8171 mRNA, 354 miRNA, 1389 lncRNA ve 740 TF'in IDC ve normal dokular arasında ifadelerinin anlamlı şekilde değiştiği belirlenmiştir.

5.3. MİKRODİZİN VE YENİ NESİL DİZİLEME VERİLERİ KULLANILARAK DEVRELERİN OLUŞTURULMASI

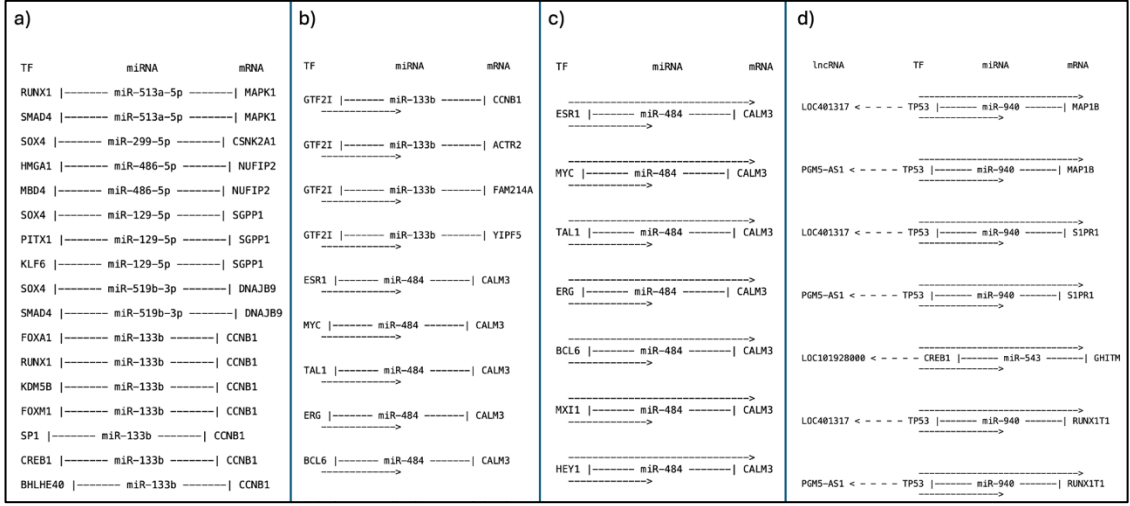
Mikrodizin ve NGS verilerinin analizleri sonucunda IDC ve normal dokular arasında anlamlı ifade değişimi gösteren moleküller belirlenmiş ve *miRCuit* devre analiz programına yüklenmeye hazır mRNA, miRNA, lncRNA ve TF listeleri elde edilmiştir. Listeler programa yüklenmiş, her iki teknolojiye elde edilen veriler için tüm dosyaların eşik değeri 1.5 kat olarak belirlenmiş ve analiz gerçekleştirilmiştir. Elde edilen sonuçlar aşağıda verilmiştir:

- *Mikrodizin verilerinden elde edilen devreler*

Mikrodizin verilerinden elde edilen analizler sonucunda 4 farklı türde devre oluşturulmuştur. Bu devreler, başlık “4.2.3.4. Düzenleyici Devrelerin Oluşturulması” altında tanımlanan türlerle uyumlu olarak isimlendirilmiştir. Her bir devre türünün toplamda içerdiği devre sayısı aşağıda belirtilmiştir.

- Devre 1: Toplamda 458.014 devre,
- Devre 2: Toplamda 129.459 devre,
- Devre 3: Toplamda 100.924 devre,
- Devre 4: Toplamda 6.267 adet devre bulunmuştur.

Analizler sonucunda 5, 6, 7, ve 8. türde devreler oluşturulamamıştır. Her bir devre türüne ait örnek gösterim Şekil 5.1.'de verilmiştir.



Şekil 5.1. Mikrodizin analizi sonucunda elde edilen farklı devre türlerine ait örnek gösterim (a) Devre 1, b) Devre 2, c) Devre 3, d) Devre 4).

Mikrodizin analizleri sonucunda, transkriptom array’lerden gelen lncRNA verilerinin sınırlılığı nedeniyle yeteri kadar molekül etkileşimine ulaşamamış ve bu nedenle *miRNA bağımlı lncRNA düzenleyici devreler* ile *TF bağımlı lncRNA düzenleyici devreler* oluşturulamamıştır.

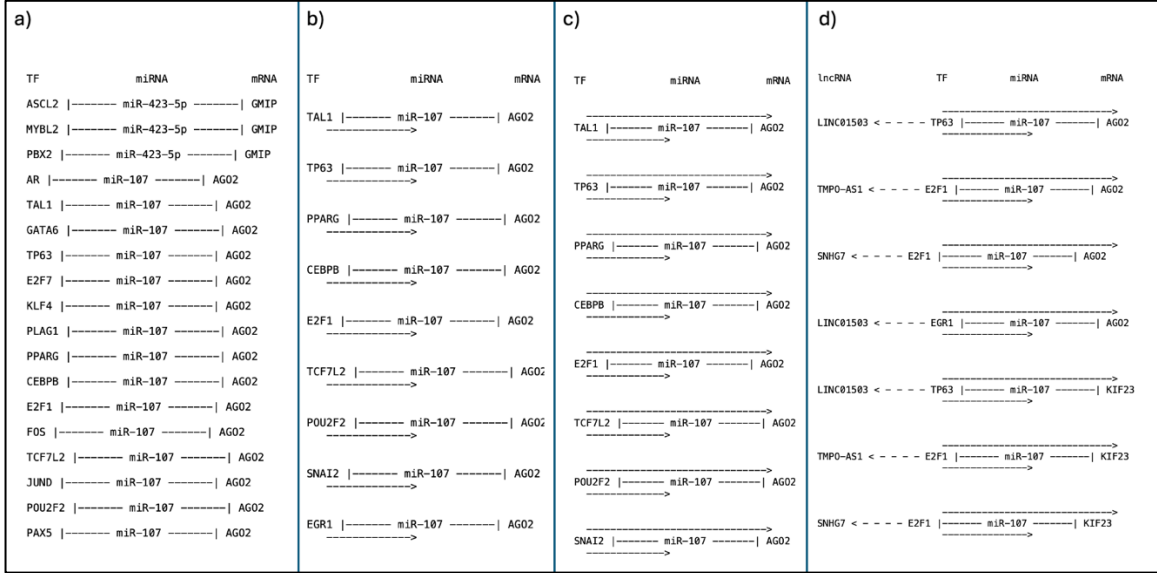
- *NGS verilerinden elde edilen devreler*

NGS verilerinden elde edilen analizler sonucunda 7 farklı türde devre oluşturulmuştur. Bu devreler, başlık “4.2.3.4. **Düzenleyici Devrelerin Oluşturulması**” altında tanımlanan türlerle uyumlu olarak isimlendirilmiştir. Her bir devre türünün toplamda içerdiği devre sayısı aşağıda belirtilmiştir.

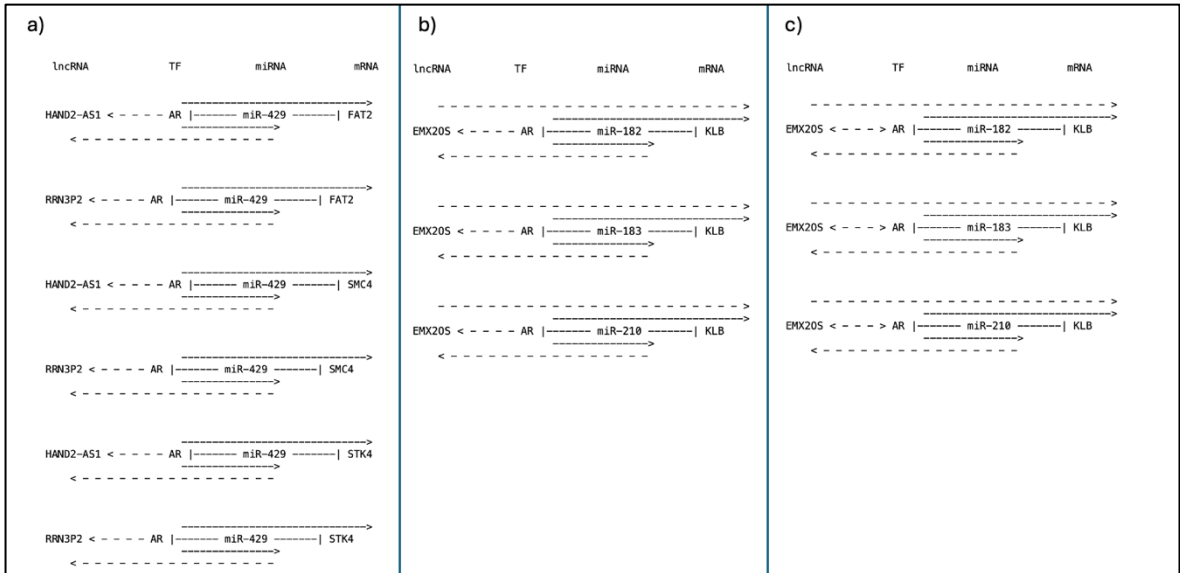
- Devre 1: Toplamda 202.635 devre,
- Devre 2: Toplamda 54.914 devre,
- Devre 3: Toplamda 39.583 devre,
- Devre 4: Toplamda 106.898 devre,
- Devre 5: Toplamda 4.443 devre,
- Devre 6: Toplamda 3 devre,

- Devre 7: Toplamda 3 adet devre bulunmuştur.

Analizler sonucunda 8. türde devre oluşturulamamıştır. Her bir devre türüne ait örnek gösterim Şekil 5.2. ve Şekil 5.3'te verilmiştir.

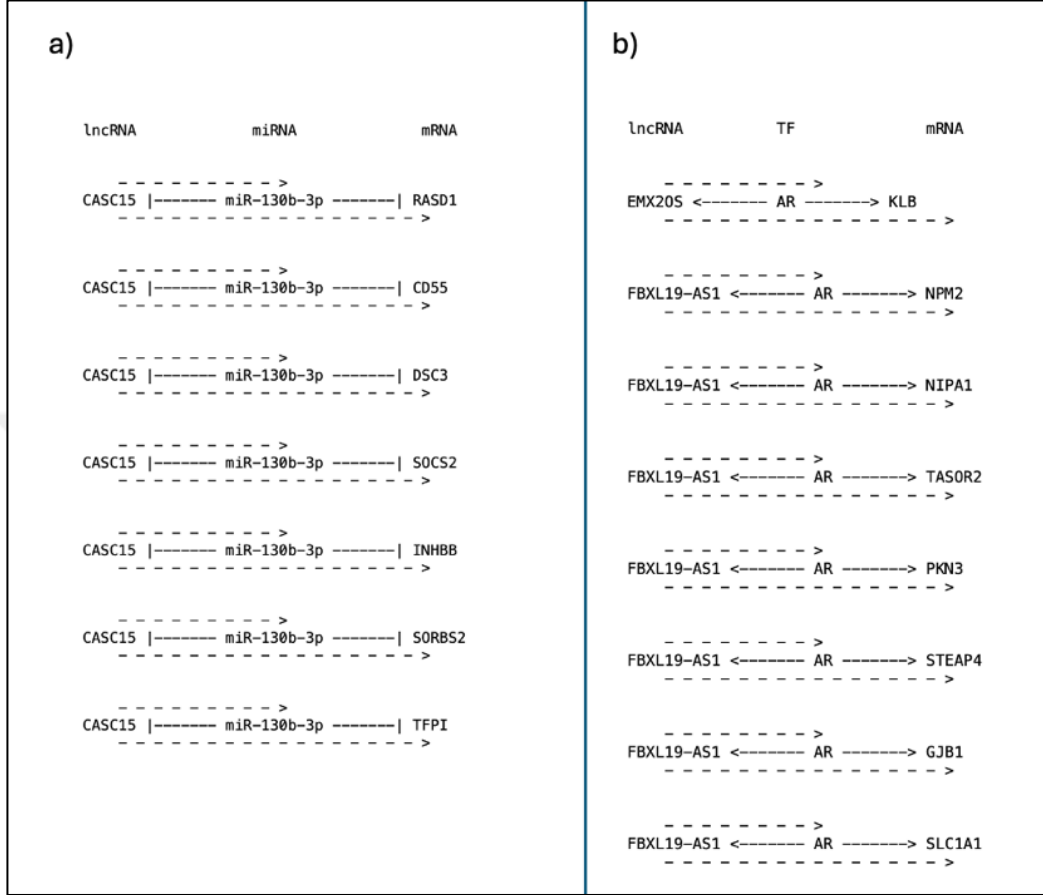


Şekil 5.2. NGS analizi sonucunda elde edilen farklı devre türlerine ait örnek gösterim 1 (a) Devre 1, b) Devre 2, c) Devre 3, d) Devre 4).



Şekil 5.3. NGS analizi sonucunda elde edilen farklı devre türlerine ait örnek gösterim 2 (a) Devre 5, b) Devre 6, c) Devre 7).

NGS analizleri sonucunda, *miRNA* bağımlı *lncRNA* düzenleyici devreler'e ait toplamda 7, *TF* bağımlı *lncRNA* düzenleyici devreler'e ait ise toplamda 169 adet devre oluşturulmuştur. Bu iki devre türüne ait örnek gösterim Şekil 5.4.'te verilmiştir.



Şekil 5.4. NGS analizi sonucunda elde edilen farklı devre türlerine ait örnek gösterim 3 (a) *miRNA* bağımlı *lncRNA* düzenleyici devreler, b) *TF* bağımlı *lncRNA* düzenleyici devreler).

Devre analizleri sonucunda hem mikrodizin hem de NGS verilerinden elde edilen 4. devre türüne ait mRNA listelerine GSEA uygulanmıştır; ancak miRNA, TF ve lncRNA listelerindeki molekül çeşitliliği GSEA için yetersiz olduğundan zenginleştirme işlemi yalnızca mRNA listeleri üzerinde gerçekleştirilmiştir.

5.4. MİKRODİZİN VE YENİ NESİL DİZİLEME VERİLERİ KULLANILARAK OLUŞTURULAN DEVRELERİN GSEA ANALİZİ

Mikrodizin ve NGS verileri kullanılarak oluşturulan 4. devre türü, her iki teknolojiye de gelen verilerin birlikte değerlendirilebilmesi ve karşılaştırılabilmesi için belirlenen en uygun devre türüdür ve GSEA yöntemi ile kapsamlı şekilde analiz edilmiştir. Her iki platformdan

elde edilen sonuçlar aşağıda sunulmuştur. Bu analiz, devredeki mRNA'ların biyolojik yollarla olan ilişkilerini ortaya koymak amacı ile gerçekleştirilmiştir.

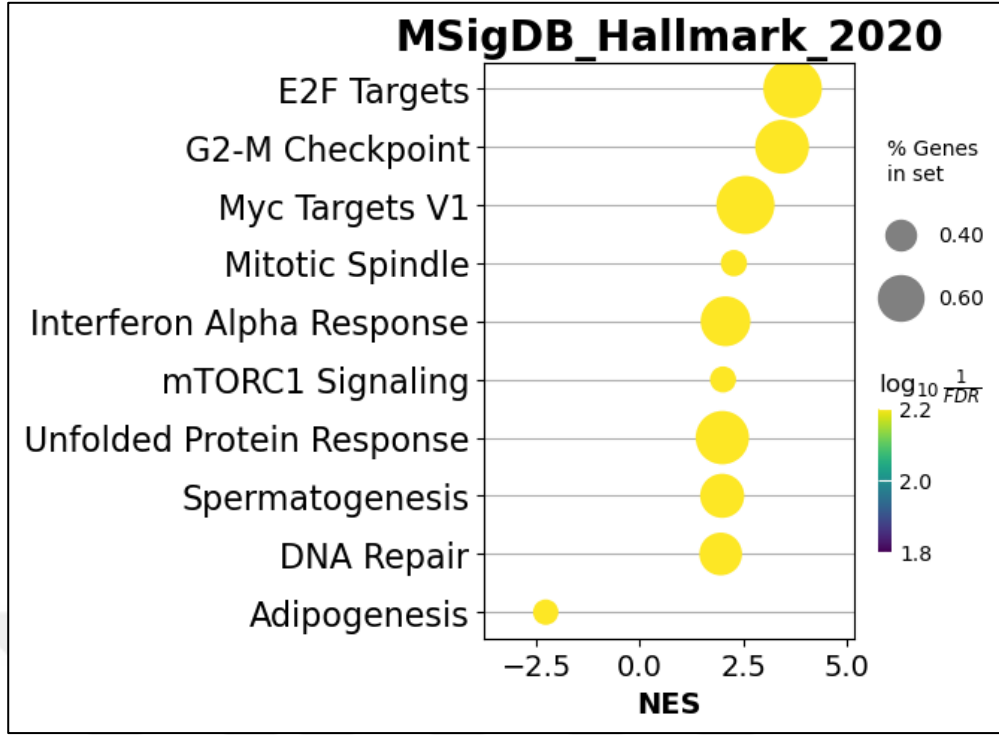
- Mikrodizin verileri kullanılarak oluşturulan devrelerin GSEA yöntemi ile analizi:

Mikrodizin verileri kullanılarak oluşturulan devrelerin GSEA yöntemi ile analizi sonucunda, FDR <0.25 olan en anlamlı ilk 5 yolak “E2F Targets, G2-M Checkpoint, Myc Targets, Mitotic Spindle” ve “Interferon Alpha Response” olarak belirlenmiş, listenin tamamı Şekil 5.5.’te verilmiştir.

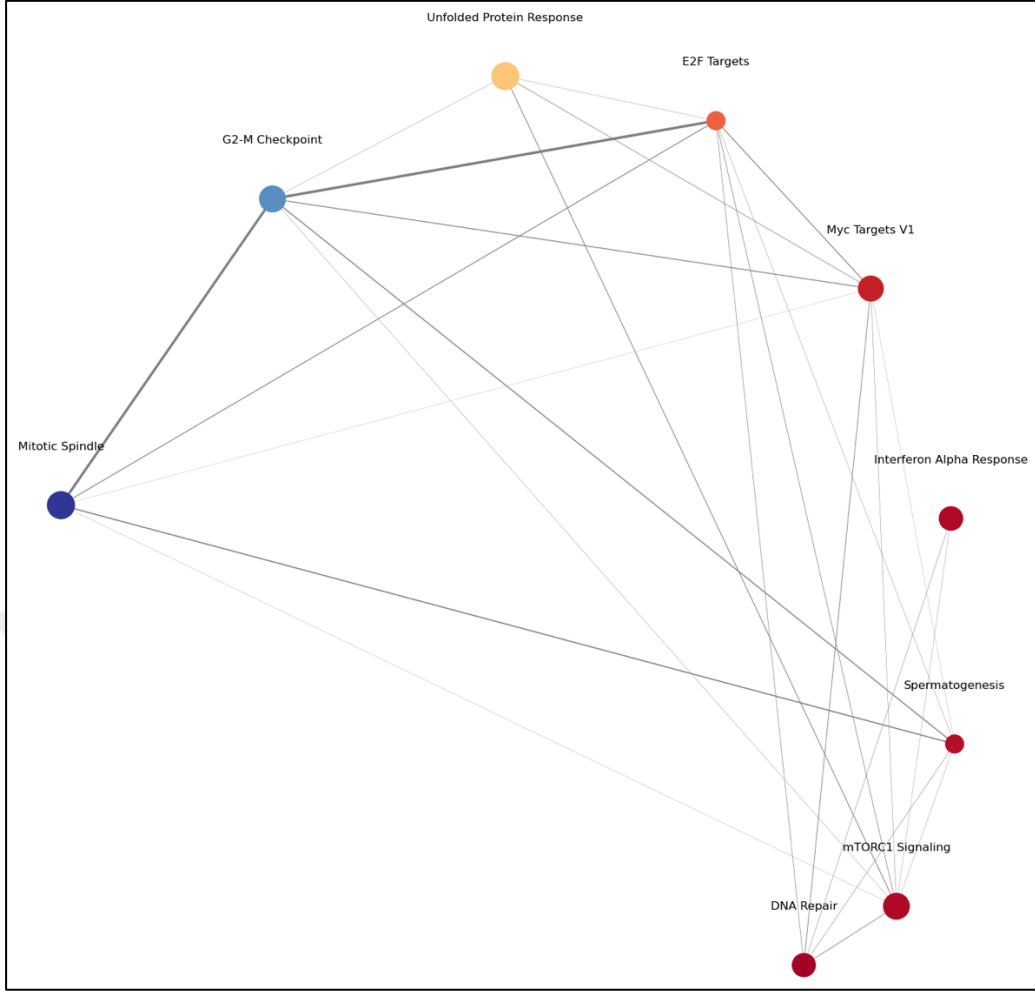
Term	NES	NOM p-val	FDR q-val	Lead_genes
E2F Targets	3.6916606483010463	0.0	0.0	RRM2;BUB1B;UBE2T;SPAG5;ATAD2;ANP32E;KPNA2;PTTG1;BIRC5;EZH2;NUDT21;KIF2
G2-M Checkpoint	3.4380015912818496	0.0	0.0	TTK;CENPF;NEK2;UBE2C;KPNA2;TPX2;PBK;SLC38A1;PTTG1;BIRC5;CCNA2;EZH2;KIF2
Myc Targets V1	2.553205207812253	0.0	0.0	YWHAE;KPNA2;CCNA2;MCM4;DXH15;HPRT1;PRPS2;ILF2;VDAC1;FAM120A;XPOT;ACP
Mitotic Spindle	2.2728693905836495	0.0	0.0	YWHAE;ANLN;TTK;CENPF;NEK2;TPX2;EZR;BIRC5;KIF2C;KIF23;KIF4A;ARFGEF1;PDLIN
Interferon Alpha Response	2.071510923122535	0.0030120481927710845	0.003239522170479854	C1S;IFI44L;RSAD2;CMPK2;IFI30;MX1;LAMP3;DDX60;OAS1;PNPT1;IFI27;PARP9;IFIH1;F
mTORC1 Signaling	2.0103965715199803	0.0	0.0038244358957053833	RRM2;PDP;ACTR2;DDIT4;M6PR;BUB1;SLC7A11;STIP1;RIT1;RAB1A;TFRC;MCM4;SOR
Unfolded Protein Response	1.9941178905410293	0.0	0.004049402713099818	DDIT4;KDEL3;YWHAZ;SLC7A5;PAIP1;DNAJC3;MTHFD2;VEGFA;PDIA6;LSM4;FKBP14;
Spermatogenesis	1.9903839241845297	0.003048780487804878	0.003711952487008166	TTK;NEK2;EZH2;KIF2C;BUB1;CLGN;MAP7;DBF4;GRM8;STAM2;STRBP;RFC4;NCAPH
DNA Repair	1.9539515403200194	0.0	0.004949269982677555	TMED2;ALYREF;PNP;NUDT21;CMPK2;HPRT1;RFC3;RFC5;ARL6P1;CANT1;RFC4;NME
Protein Secretion	1.679205900836155	0.01	0.03212526152392522	TPD52;TMED2;ANP32E;AP2B1;RAB14;ARFGEF1;M6PR;SNAP23;ARFGEF2;MAPK1
Myc Targets V2	1.6632061746105509	0.01937046004842615	0.033622313436049006	MCM4;SOR;TMEM97;NOLC1;CBX3;HSP1
Estrogen Response Late	1.4593314422632164	0.053497942386831275	0.10427211986232031	CDH1;FGFR3;SFN;CELSR2;TPD52L1;KIF20A;CA2;NR1P1;LLGL2;CDC6;PRLR;STIL;SER
Estrogen Response Early	1.43116211147613	0.03773584905660377	0.1097491966088848	RAB17;SFN;CELSR2;CELSR1;TTC39A;MUC1;TPD52L1;KRT8;REEP1;NR1P1;MLPH;RAS
Interferon Gamma Response	1.2471379288149291	0.14788732394366197	0.2452781071934747	C1S;IFI44L;PNP;RSAD2;CMPK2;IFI30;RNF213;MTHFD2;CASP3;MX1;OAS2;SPPL2A;GC
UV Response Up	1.2420174902122347	0.18181818181818182	0.2346853839058739	EPCAM;CA2;STIP1;RASGRP1;TFRC;PPAT;GGH;FKBP4;CASP3;HSPA13;AGO2;ATP6V10
Inflammatory Response	-1.397964211351397	0.08839050131926121	0.21251522260532293	NMUR1;ADM;PROK2;EREG;GPC3;SLC4A4;CXCL6;HBEGF;IL6;IL1R1;FZD5;PTGER2;NL
Apoptosis	-1.4052798009940388	0.08082191780821918	0.2207742375579115	CAV1;TGFB3;EREG;EMP1;IL6;BMP2;DCN;ANXA1;CLU;MMP2;CCND2;AVPR1A;BNIP3L
Apical Junction	-1.4730273833116785	0.06114130434782609	0.16728120244022812	CALB2;GNAI1;ADAMTS5;CLDN11;COL17A1;CD209;NEGR1;PECAM1;FBN1;MMP2;VCL;
Cholesterol Homeostasis	-1.5229103584231358	0.06485671191553545	0.13421254583179376	LPL;GNAI1;CLU;ANTXR2;AVPR1A;FABP5;ALDOC
Xenobiotic Metabolism	-1.5670458038762447	0.02560819462227913	0.11283858681749709	HSD17B2;CD36;IGF1;PDK4;NDRG2;AKR1C3;MAN1A1;XDH;IL1R1;ABCD2;HACL1;AOX1;
Bile Acid Metabolism	-1.6392172766620345	0.02857142857142857	0.07514282456596376	LIPE;SLCO1A2;BBOX1;ACSL1;ABCD2;HACL1;ALDH1A1;BMP6;ACSL5;ABCA9;OPTN
Myogenesis	-1.719817862274581	0.0013605442176870747	0.0477503339681096	CD36;IGF1;FHL1;MYH11;SORBS1;DMD;MYO1;HBEGF;CASQ2;ACSL1;FGF2;CLU;MYO
Fatty Acid Metabolism	-1.7414191042195857	0.00430416068866571	0.04618707898701077	CD36;ACOT2;GPD1;AOC3;CA6;MGLL;ACSL1;ME1;ALDH1A1;ACSL5;ACSL4;GABARAPL
Epithelial Mesenchymal Transition	-1.7763623482310302	0.0	0.039735099337748346	MFAP5;SRFP1;PCOLCE2;OXTR;TGFB3;MEST;MATN2;PTHLH;CXCL12;CXCL6;ABI3B
KRAS Signaling Up	-1.820922411134281	0.002628120893561104	0.031122985532785717	F13A1;PIGR;RELN;SNAP25;PTGS2;EREG;CFH;SLPI;TMEM100;AKAP12;EMP1;IGF2;HB
UV Response Dn	-1.8699720223842486	0.0013531799729364006	0.02529630787596282	CAV1;PLPP3;TGFB3;PRKAR2B;MGLL;DLCL1;EFEMP1;ADGRL2;VLDLR;ADD3;TFPI;PTP
Hypoxia	-1.971253842392812	0.0	0.014069294829889435	ADM;CAV1;PCK1;STBD1;AKAP12;GPC3;PLIN2;VLDLR;IL6;DCN;RRAGD;CDKN1C;PFKF
Adipogenesis	-2.273750862195346	0.0	0.0017053690702896285	LEP;CD36;LPL;ITIH5;LIPE;ENPP2;ANGPT1;SORBS1;PLIN2;MGLL;C3;LIFR;PTGER3;ME

Şekil 5.5. Mikrodizin verileri kullanarak oluşturulan devrelerin GSEA yöntemi ile analizi sonucunda zenginleşen yollar (FDR<0.25).

Analiz sonucunda zenginleşen yollar Şekil 5. 6.’da nokta grafiği ile görselleştirilmiş, bu yolların ağ yapısı ise Şekil 5.7.’de sunulmuştur.



Şekil 5.6. Mikrodizin verileri kullanılarak oluşturulan devrelerin GSEA yöntemi ile analizi sonucunda zenginleşen yolların nokta grafiği.



Şekil 5.7. Mikrodizin verileri kullanılarak oluşturulan devrelerin GSEA yöntemi ile analizi sonucunda zenginleşen yollara ilişkin ağ yapısı.

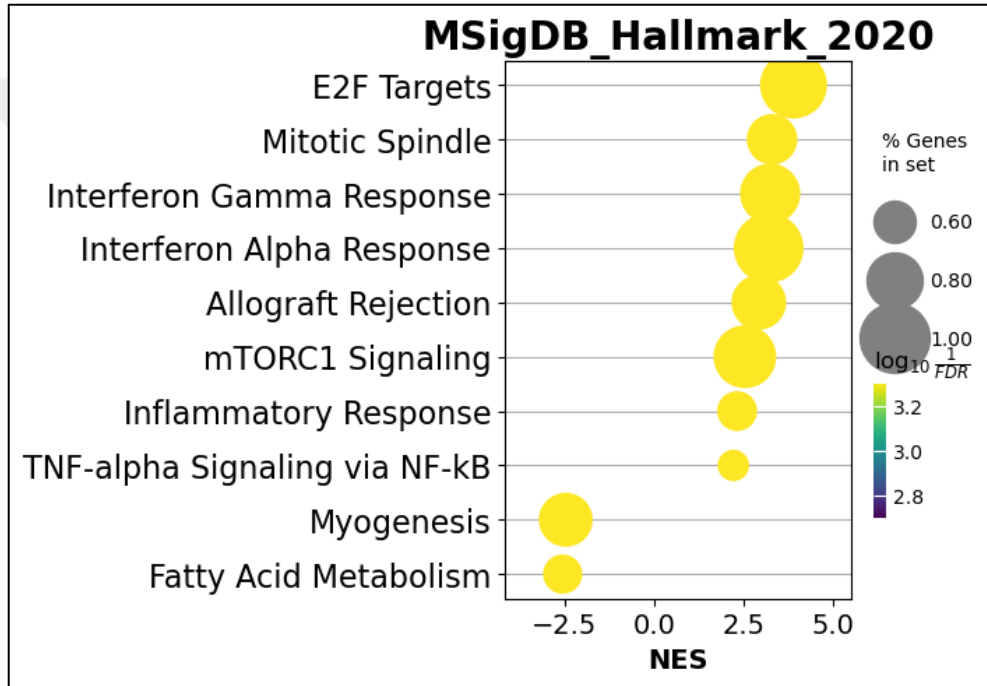
- *NGS verileri kullanılarak oluşturulan devrelerin GSEA yöntemi ile analizi:*

NGS verileri kullanılarak oluşturulan devrelerin GSEA yöntemi ile analizi sonucunda, FDR <0.25 olan en anlamlı ilk 5 yolak “*G2-M Checkpoint*, *E2F Targets*, *Mitotic Spindle*, *Interferon Gamma Response*” ve “*Interferon Alpha Response*” olarak belirlenmiş, listenin tamamı Şekil 5.8.’de verilmiştir.

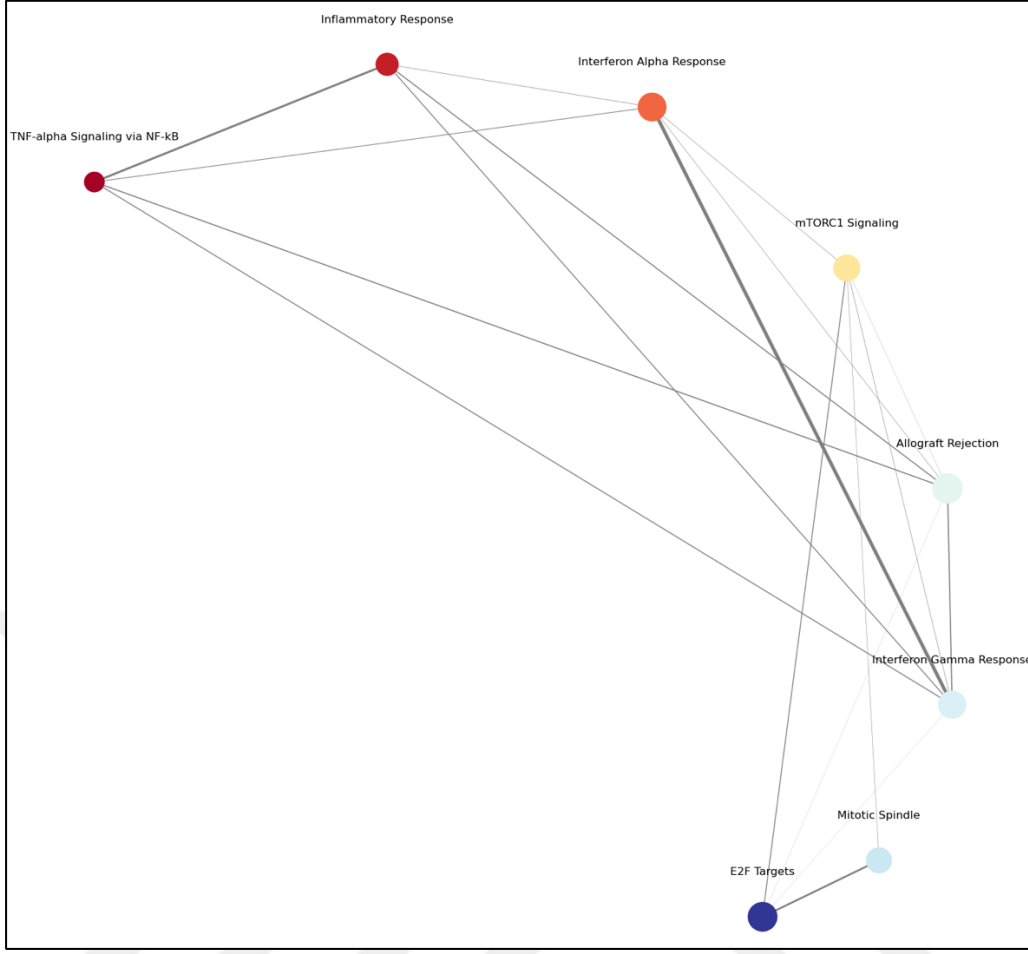
Term	NES	NOM p-val	FDR q-val	Lead_genes
G2-M Checkpoint	3.899564971250903	0.0	0.0	TOP2A;UBE2C;BUB1;RAD21;TPX2;KIF4A;EXO1;KIF15;AURKA;NUSAP1;MKI67;KNL1;NEK2;PRC1
E2F Targets	3.892667941128182	0.0	0.0	TOP2A;RRM2;RAD21;KIF4A;DLGAP5;AURKA;MKI67;BUB1B;ESPL1;MELK;KIF18B;RACGAP1;CIT
Mitotic Spindle	3.2931837422054975	0.0	0.0	TOP2A;BUB1;TPX2;KIF4A;KIF15;DLGAP5;AURKA;ANLN;NUSAP1;NEK2;PRC1;ESPL1;RACGAP1;C
Interferon Gamma Response	3.2402165395954046	0.0	0.0	TAP1;CD274;CCL5;ITGB7;OAS3;SLAMF7;OAS3;IL2RB;IFI44L;EPSTI1;NLRC5;SAMD9L;DDX60;V
Interferon Alpha Response	3.1972363075756762	0.0	0.0	TAP1;LAMP3;IFI44L;EPSTI1;SAMD9;SAMD9L;DDX60;OAS1;PARP14;MX1;IFI2;ISG20;PARP9;IFI
Allograft Rejection	2.9243164108392086	0.0	0.0	TAP1;TAP2;PTPRC;CD80;TRAT1;CCL5;CCR5;IL2RB;KLRD1;IL12B;CD3E;FYB1;MAP4K1L;CP2;IL
mTORC1 Signaling	2.534195497582047	0.0	0.0	RRM2;BUB1;AURKA;CXCR4;CORO1A;DAPP1;SLAC;CD2C5A;TFR3;ADIPOR2;TMEM97;STARD4;T
Inflammatory Response	2.319255579633638	0.0	0.0	CCL5;LAMP3;IL7R;SLAMF1;CD48;IL2RB;IL12B;CD69;LCP2;RGS1;CYBB;INHBA;GPR132;OLR1;O
TNF-alpha Signaling via NF-kB	2.2138237911787724	0.0	0.0004666083406240887	TAP1;CD80;CCL5;IL7R;IL12B;CD69;TNFAIP3;BIRC3;BCL2A1;IFIT2;INHBA;OLR1;SLC16A6;TNFA
Complement	2.0250601903377894	0.0	0.0054593175853018365	SERPINA1;CCL5;F5;HPCAL4;CR2;SPOCK2;PIK3CG;ITIH1;PLA2G7;HSPA1A;GNB4;TNFAIP3;LCP2
Spermatogenesis	1.9970846223750276	0.0	0.005821999522786925	LDHC;BUB1;AURKA;NEK2;NCAPH;CCNB2;KIF2C;SCG3;CDK1;CAMK4;TTK
Myc Targets V1	1.5347896665722058	0.03718199608610567	0.12554680664916884	CDC45;MAD2L1;KPN2;CCNA2;XPOT;MCM2
Xenobiotic Metabolism	-1.6487214223962547	0.02510460251046025	0.16905092133940955	AKR1C2;CYP26A1;CD36;ADH1C;F10;ALDH3A1;ACSM1;AKR1C3;PTGES;PDK4;MAOA;ARG2;IGFB
Adipogenesis	-2.2075362874300097	0.0019193857965451055	0.0009510600356647513	CIDEA;LEP;FABP4;ACADL;ADIPOQ;ANGPTL4;CD36;LIPE;LPL;GPX3;ITGA7;SNCG;TST;SORBS1;Y
Myogenesis	-2.4634088967815013	0.0	0.0	CFD;STC2;CD36;SLC6A8;SGCG;NCAM1;MYOM2;TNNT3;GPX3;ITGA7;TPM2;SORBS1;MYL3;SPB
Fatty Acid Metabolism	-2.549625125336403	0.0	0.0	CA4;CIDEA;ACADL;GPD1;G0S2;CD36;HMGS2;ADH1C;ALDH3A1

Şekil 5.8. NGS verileri kullanılarak oluşturulan devrelerin GSEA yöntemi ile analizi sonucunda zenginleşen yollar (FDR <0.25).

Analiz sonucunda zenginleşen yollar Şekil 5. 9.'da nokta grafiği ile görselleştirilmiş, bu yolların ağ yapısı ise Şekil 5.10.'da sunulmuştur.



Şekil 5.9. NGS verileri kullanılarak oluşturulan devrelerin GSEA yöntemi ile analizi sonucunda zenginleşen yolların nokta grafiği.



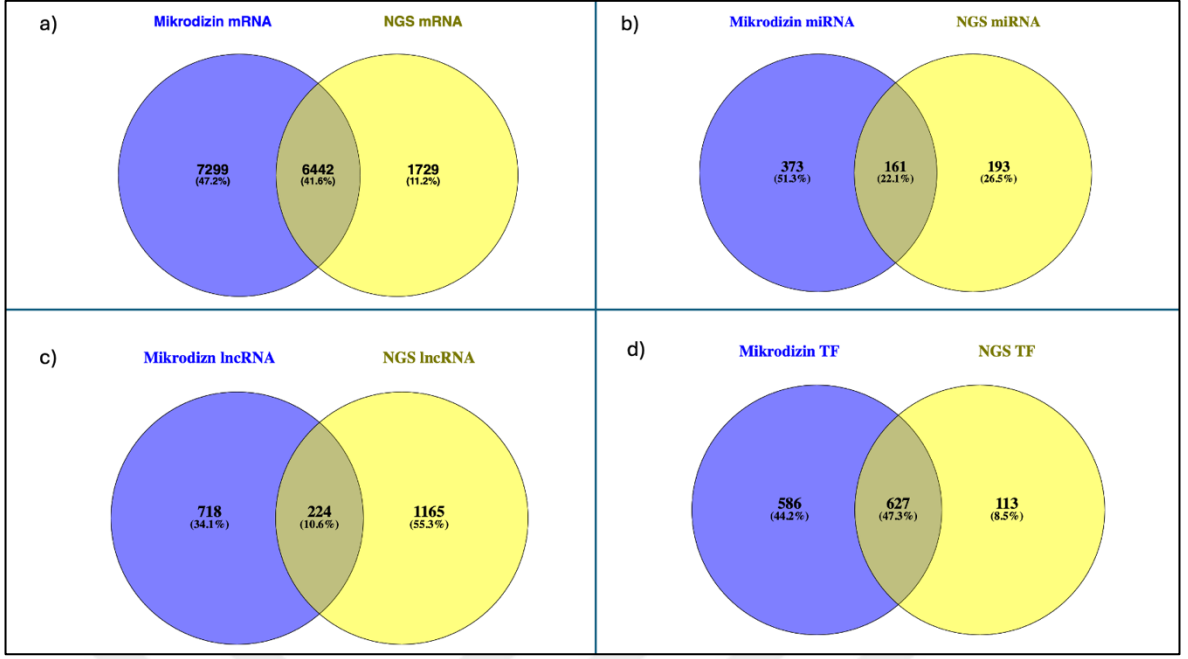
Şekil 5.10. NGS verileri kullanılarak oluşturulan devrelerin GSEA yöntemi ile analizi sonucunda zenginleşen yollara ilişkin ağ yapısı.

5.5. MİKRODİZİN VE YENİ NESİL DİZİLEME VERİLERİNDEN ELDE EDİLEN SONUÇLARIN KARŞILAŞTIRILMASI

Mikrodizin ve NGS platformlarından elde edilen tüm bulgular, kapsamlı bir şekilde değerlendirilmiş ve aralarındaki ilişki incelenmiştir. İki farklı platformdan (teknolojiden) elde edilen sonuçlar, bu analizlerin genetik regülasyon mekanizmaları üzerindeki etkilerini değerlendirmek amacıyla aşağıda sunulmuştur.

- DEG Listelerinin Karşılaştırılması

Mikrodizin ve NGS veri setlerinin DEG analizleri sonucunda anlamlı ifade değişikliği gösteren mRNA, miRNA, lncRNA ve TF listeleri karşılaştırılmıştır. Bu karşılaştırmanın sonuçları, Venny 2.1.0 veri tabanı kullanılarak oluşturulan Venn diyagramları ile görselleştirilmiştir (Şekil 5.11.).



Şekil 5.11. Mikrodivin ve NGS veri setlerinin DEG analizleri sonucunda anlamlı ifade değişikliği gösteren mRNA, miRNA, lncRNA ve TF listelerinin karşılaştırılması (a) mRNA verilerinin karşılaştırılması, b) miRNA verilerinin karşılaştırılması, c) lncRNA verilerinin karşılaştırılması, d) TF verilerinin karşılaştırılması.).

Sonuçta mikrodivin ve NGS verileri arasında mRNA'ların %41.6'sı (6442 mRNA), miRNA'ların %22.1'i (161 miRNA), lncRNA'ların %10.6'sı (224 lncRNA) ve TF'lerin %47.3'ü (627 TF) ortak olarak belirlenmiştir.

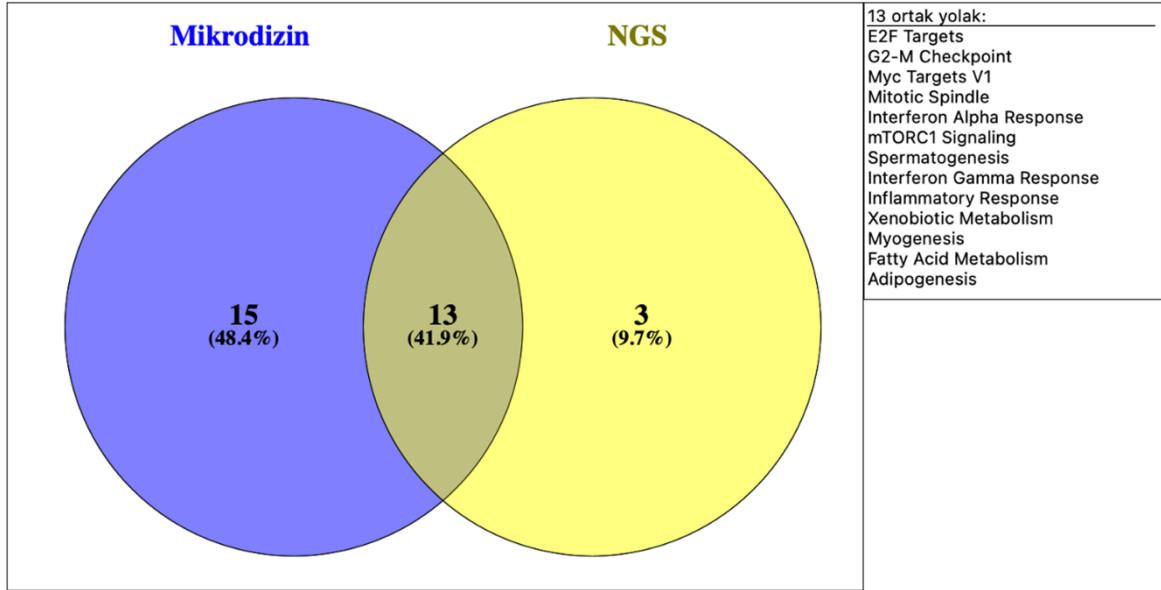
- Regülatif Devrelerin Karşılaştırılması

Mikrodivin ve NGS veri setlerinin diferansiyel ifade gösteren mRNA, miRNA, lncRNA ve TF listelerinden üretilen devreler karşılaştırılmıştır. Karşılaştırma işlemi, devrelerin .csv formatındaki tablo verileri kullanılarak Python 3.12.0 ortamında gerçekleştirilmiştir. Mikrodivin verileri kapsamında yalnızca dört farklı devre türü mevcut olduğundan, elde edilen sonuçlar bu dört devre türü ile sınırlı kalmıştır. Karşılaştırma sonucunda:

- Devre 1: Toplamda 23.475 adet regülatif devre,
- Devre 2: Toplamda 7.069 adet regülatif devre,
- Devre 3: Toplamda 5.195 adet regülatif devre ortak olarak bulunmuştur. Devre 4 ise, ortak bir regülatif devre içermemektedir.

- Genlerin Zenginleştigi Yolakların Karşılaştırılması

Devrelerin GSEA yöntemi ile gerçekleştirilen gen seti zenginleştirme analizi sonucunda, mikrodizin ve NGS devrelerinden elde edilen mRNA'ların zenginleştigi yollar değerlendirilmiş ve FDR <0.25 kriterini taşıyan en anlamlı yollar karşılaştırılmıştır. Sonuçlar, Venny 2.1.0 veri tabanı kullanılarak oluşturulan Venn diyagramları ile görselleştirilmiştir (Şekil 5.12.).

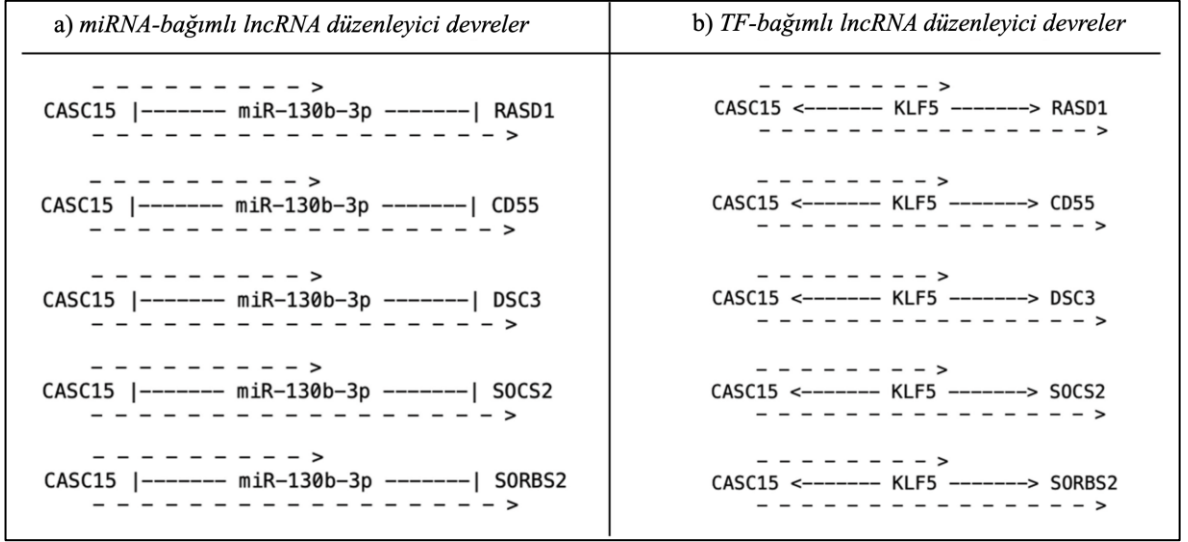


Şekil 5.12. Gen Seti Zenginleştirme Analizi (GSEA) sonucunda belirlenen yolların mikrodizin ve NGS devreleri arasında karşılaştırılması.

Sonuçlar, her iki platformdan elde edilen devreler arasında %41.9 oranında (13 yolak) ortak yolak bulunduğunu ortaya koymuştur.

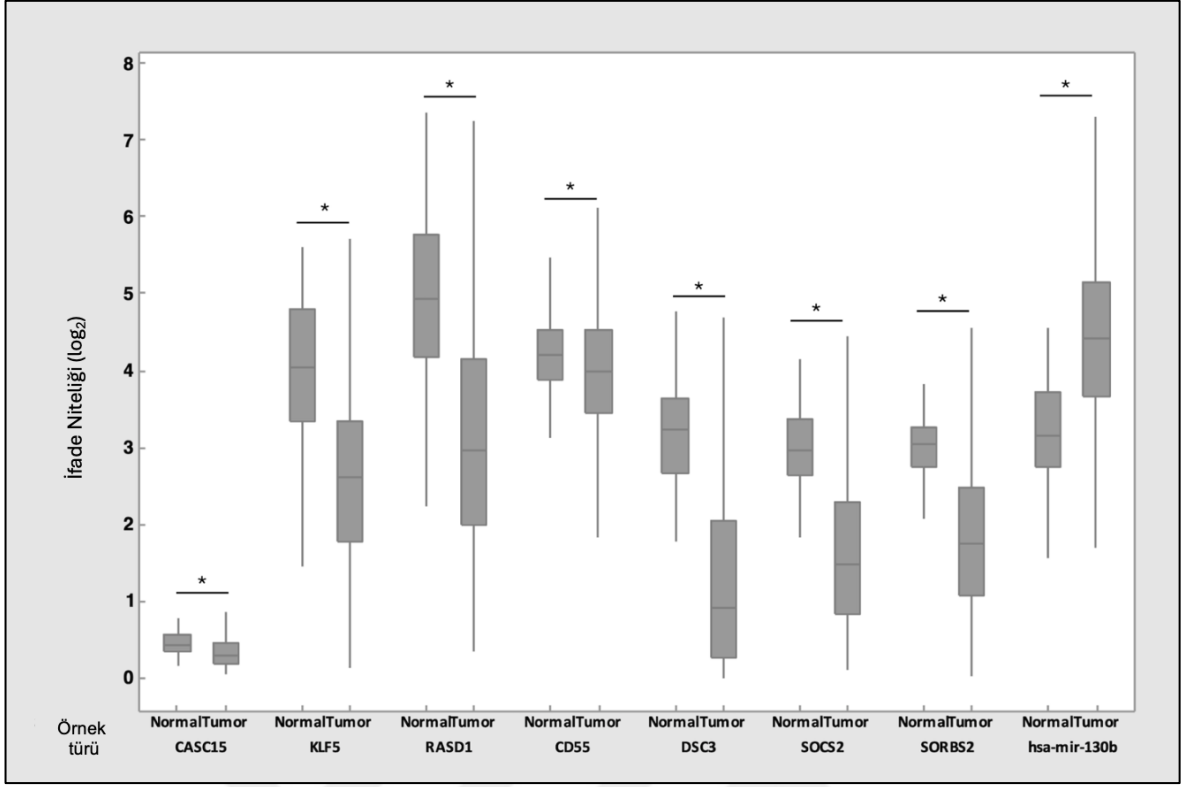
5.6. TCGA VERİ ANALİZİ

Veri analizi sonucunda yalnızca NGS veri setlerinden elde edilen özel düzenleyici devreler (miRNA-bağımlı lncRNA düzenleyici devreler ve TF-bağımlı lncRNA düzenleyici devreler) belirgin şekilde lncRNA CASC15'e odaklanmıştır. Diğer düzenleyici devre elemanları ise, transkripsiyon faktörü KLF5, miR-130b-3p ve hedef genler *RASD1*, *CD55*, *DSC3*, *SOCS2* ve *SORBS2* olarak belirlenmiştir (Şekil 5.13.).



Şekil 5.13. *miRCuit* tarafından oluşturulan CASC15 bağımlı düzenleyici devreler.

miRCuit kullanılarak elde edilen devrelerin üyeleri arasındaki ilişkilerin bağımsız veri setiyle doğrulanması amacıyla, Şekil 5.13'te sunulan CASC15, miR-130b-3p, KLF5 ve hedeflerinin ifade profilleri, TCGA meme kanseri veri seti (GDC BRCA) kullanılarak UCSC Xena aracılığıyla analiz edilmiştir. miR-130b-3p'nin, normal dokulara kıyasla meme tümörü dokularında anlamlı derecede fazla ifade edildiği, buna karşın diğer genlerin ve lncRNA CASC15'in tümör örneklerinde normal örneklere kıyasla anlamlı derecede az ifade edildiği bulunmuştur (Şekil 5.14.).



Şekil 5.14. TCGA'den elde edilen 1091 meme tümörü ve 105 normal doku örneğinde CASC15, KLF5, RASD1, CD55, DSC3, SOCS2, SORBS2 ve miR-130b-3p'nin ifadelerinin değerlendirilmesi (t-testi, $p < 0.01$).

6. TARTIŞMA

Gen ifadesi, çok sayıda düzenleyici molekülün karmaşık etkileşimleri tarafından kontrol edilen dinamik bir süreçtir. Düzenleyici moleküllerin ifadelerinde meydana gelen değişimlerin, meme kanseri başta olmak üzere pek çok hastalığın oluşumunda ve gelişiminde kritik bir role sahip olduğu bilinmektedir. Bu bağlamda, gen ifadesinin düzenlenmesine ilişkin süreçlerin sistem biyolojisi yaklaşımı ile bütüncül şekilde ele alınması ve *in silico* ortamda biyolojik devrelerin ve ağlarını kurgulanması, pek çok hastalığın tanısında ve yeni tedavi yaklaşımlarının geliştirilmesinde büyük önem taşımaktadır (5–7,10,17).

Tez çalışması kapsamında, biyolojik devrelerin ve ağların kurgulanması amacı ile **miRCuit** isimli yeni bir devre analiz programı geliştirilmiştir. Mikrodizin ve Yeni Nesil Dizileme (NGS) çalışmalarından elde edilen bulgular, yeni geliştirilen program kullanılarak kapsamlı bir şekilde analiz edilmiştir. Geliştirilen programın performansı ve özellikleri değerlendirilerek her iki teknolojiye elde edilen verilerin biyolojik çıktıları ve teknolojik farklılıkları detaylı bir şekilde aşağıda tartışılmıştır.

Veri Analizinde Teknik Platformların Etkisi ve “Batch Effect” Değerlendirmesi

Bu çalışmada, mikrodizin ve NGS verileri birleştirilmeksizin her iki platform kendi içinde ayrı olarak analiz edilmiş ve sonuçlar biyolojik karşılaştırmalar yapılabilecek şekilde değerlendirilmiştir. Mikrodizin ve NGS verilerinin ayrı analiz edilmesi, her iki platformun kendine özgü biyolojik ve teknolojik özelliklerinin daha net bir şekilde anlaşılmasını sağlamıştır. Bununla birlikte, platform içi meta-analiz yaklaşımlarının uygulanmasının, verilerdeki teknik varyasyonları ortadan kaldırarak istatistiksel güvenilirliği artırma ve biyolojik anlamı daha güçlü bir şekilde ortaya koyma potansiyeline sahip olduğu düşünülmektedir.

Verilerin birleştirilmesinde dikkate alınması gereken “Batch effect”, özellikle yüksek boyutlu gen ifade analizlerinde teknik değişkenliklerin biyolojik sinyalleri maskeleymesine neden olabilmektedir. Bu tür etkilerin giderilmesi amacıyla meta-analiz yaklaşımlarının kullanılmasının, platformlar arasında biyolojik sinyallerin daha doğru bir şekilde tanımlanmasını sağlayabileceği belirtilmektedir. Mikrodizin verilerinde sınırlı gen kapsamı

ve lncRNA'ların yeterince temsil edilmemesi gibi faktörlerin, meta-analiz modelleri aracılığıyla telafi edilebileceği öngörülmektedir. Benzer şekilde, NGS verilerinin geniş gen kapsamı sunmasının, birleştirilmiş bir analiz çerçevesinde daha fazla biyolojik çıktı sağlayabileceği değerlendirilmektedir.

Ayrıca, NGS ve mikrodizin teknolojilerinin biyolojik süreçleri farklı yönlerden yansıttığı göz önünde bulundurulmalıdır. NGS, RNA seviyelerindeki dinamik değişimleri yüksek çözünürlükte yakalayarak daha detaylı bir biyolojik görünüm sağlarken, mikrodizin belirli bir gen kümesi ile sınırlı olduğundan bazı moleküler özellikleri kapsam dışında bırakabilmektedir. Bu iki teknoloji arasındaki farklılıkların, batch effect'in minimize edilmesi yoluyla daha doğru bir şekilde değerlendirilmesi ve elde edilen sonuçların karşılaştırılabilirliğinin artırılması mümkün hale gelebilmektedir.

Bu tez kapsamında, her ne kadar meta-analiz uygulanmamış olsa da, platformların ayrı ayrı değerlendirilmesi, her bir teknolojinin biyolojik çıktılarının bağımsız olarak analiz edilmesine olanak tanımıştır. Bununla birlikte, bu çalışmada batch effect'in modellenmemesi nedeniyle, özellikle mikrodizin verilerindeki sınırlamalar dolayısıyla elde edilen bazı düzenleyici devrelerin kapsamının daralmış olabileceği düşünülmektedir.

Sonuç olarak, tez çalışması kapsamında *miRCuit* programının fonksiyonel olarak çalışıp çalışmadığının belirlenmesine odaklanıldığı için, batch effect'in doğurabileceği "istatistiksel yanlış pozitif / yanlış negatif" durumları göz ardı edilerek çalışma tamamlanmış ve sonuçlar buna göre değerlendirilmiştir.

miRCuit Programının Geliştirilmesi ve Biyolojik Ağ Analizine Katkısı

Biyolojik devreler ve ağların daha kapsamlı bir şekilde anlaşılması ve bunun sonucunda hücresel süreçlerin daha iyi modellenmesine duyulan ihtiyaç, pek çok biyoinformatik aracın geliştirilmesine olanak sağlamıştır. Bununla birlikte, düzenleyici devre analiz araçları, meme kanseri de dahil olmak üzere çeşitli hastalıkların altında yatan karmaşık moleküler etkileşimleri açıklama potansiyelleri nedeniyle giderek daha fazla dikkat çekmektedir. Günümüzde miRNA-hedef gen, lncRNA-miRNA, TF-hedef gen ve miRNA-TF etkileşimleri gibi ikili ilişkileri tanımlamak üzere tasarlanmış pek çok veri tabanı bulunmaktadır (137–141). Buna ek olarak, düzenleyici devreler üzerine yapılan araştırmalar,

çoğunlukla mevcut çevrimiçi araçlar veya programlar kullanılarak miRNA veya lncRNA merkezli üçlü ilişkilerin oluşturulmasına odaklanmaktadır (163–166). Ancak bu moleküllerin hücre içindeki düzenleyici rolleri göz önüne alındığında, bu tür devreler biyolojik süreçleri sistematik olarak açıklamada yetersiz kalabilmektedir (13,76–78).

Son yıllarda yapılan araştırmalar, mRNA, miRNA, lncRNA ve TF molekülleri arasındaki etkileşimlerin, meme kanseri de dahil olmak üzere çeşitli hastalıkların anlaşılmasında ve yeni tedavi yaklaşımlarının geliştirilmesindeki önemini vurgulamaktadır (94,97,167,168). Ancak günümüzde, bu etkileşimleri birlikte değerlendiren bütüncül bir yaklaşım ve bu etkileşimlerin sonucunda düzenleyici bir ağ kurgulayabilen aktif bir program bulunmamaktadır (15,100,169–175). Geliştirilen devre analiz programı *miRCuit* mRNA-miRNA-lncRNA-TF dörtlü devrelerine bütüncül bir yaklaşım getirerek literatürdeki bu boşluğu doldurmayı hedeflemektedir (Çizelge 6.1.).

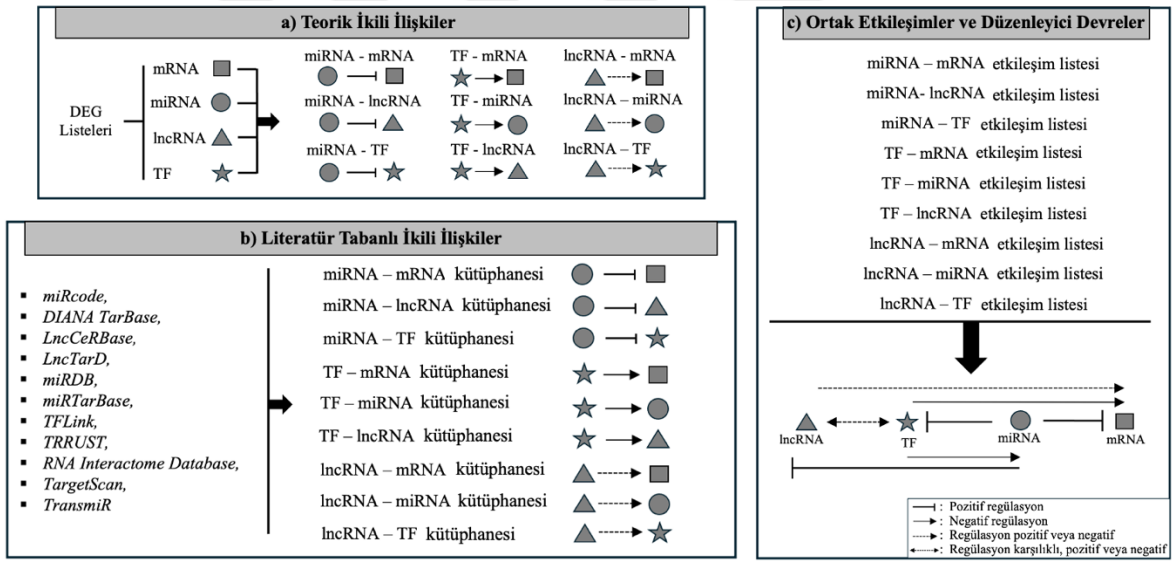
Çizelge 6.1. *miRCuit* ve farklı moleküler etkileşim programlarının karşılaştırılması.

Program Adı	Girdi dosyası	Entegre Edilen Veri Tabanı Sayısı	Etkileşim Tipi	Hedeflenen Hastalık	Çıktı Dosyası	Programlama Dili	Kullanıcı Arayüzü	Erişim
<i>GDCRNATools</i> (Li et al., 2018)	TCGA veya TARGET veri doyası	4 farklı veri tabanı	lncRNA-mRNA ilişkili endojen RNA'lar düzenleyici ağı	Kanser	DEG analizi, fonksiyonel zenginleştirme analizi	R/Bioconductor paketi	Hayır / R arayüzü	GitHub
<i>RegNetwork</i> (Liu et al., 2015)	TF, miRNA ve mRNA sembolü	25 farklı veri tabanı	TF-miRNA-hedef gen	Genel	Regülasyon ağları	Java, JavaScript ve Python	Evet	Erişim yok
<i>LncACTdb</i> (Wang et al., 2022)	mRNA, miRNA ve lncRNA ifade verisi	9 farklı veri tabanı	lncRNA-mRNA-miRNA	Farklı hastalıklar	Regülasyon ağları	MySQL, Java ve R	Hayır / R arayüzü	Web tabanlı
<i>miRGTf-net</i> (Nersisyan et al., 2021)	miRNA vernaRNA ifade dosyası	4 farklı veri tabanı	miRNA-gen-TF	Meme kanseri	Regülasyon ağları	Python	Hayır	Python arayüzü
<i>miRmapper</i> (da Silveira et al., 2018)	miRNA - mRNA etkileşim listesi ve toplam DEG listesi	3 farklı veri tabanı	TF-miRNA-hedef gen	Genel	Regülasyon ağları	R	Hayır	GitHub
<i>TFTF</i> (Wang, 2024)	TF sembolü	7 farklı veri tabanı	TF-hedef gen	Genel	TF-Hedef Regülasyon ağ analizi	R	Evet	GitHub
<i>miRCuit</i>	mRNA, miRNA, lncRNA, TF DEG listesi	11 farklı veri tabanı	mRNA-miRNA-lncRNA-TF dörtlü devre	Genel	İkili, üçlü, dörtlü etkileşimler, düzenleyici ağlar ve zenginleştirme analizi	Python	Evet	GitHub

Bununla birlikte *miRCuit*, web tabanlı bir program olmayıp kullanıcıların GitHub üzerinden erişim sağlayacakları bir program olarak sunulmuştur. Programın çalıştırılabilmesi için GitHub'ta detaylı olarak anlatıldığı gibi bazı programların/kütüphanelerin kullanıcı bilgisayarına indirilmesi ve kurulumlarının yapılması gerekmektedir. Bu durum, çalışmanın bir limitasyonu olarak değerlendirilebilir; fakat programın kurulum ve kullanımına ilişkin tüm detayların yer aldığı bir kullanım kılavuzu GitHub sayfasında yer almaktadır. Gerekli işlemler sonucunda *miRCuit*, kullanıcılara dosyalarını yükleyip kullanabilecekleri kolay bir ara yüz sunmaktadır.

Ayrıca *miRCuit*, farklı platformlardan gelen ifade verilerini analiz edebilecek şekilde tasarlanmıştır. Başka bir deyişle, girdi dosyası olarak farklı NGS veya mikrodizin platformlarından elde edilen hedef gen, miRNA, lncRNA ve TF ifade verilerini kullanmaktadır. Bu sayede, ifade verilerinin çeşitli hastalıklar için analiz edilebilmesi ve hastalıklar ile ilişkili devrelerin tasarlanması sağlanmaktadır.

miRCuit, devrelerin oluşturulması için ikili etkileşimlerin tamamını değerlendiren ve bu etkileşimlerin uygun şekilde birleştirilmesini sağlayan stratejilere sahip bir program olarak tasarlanmış olup kullanıcılara 4.2.3.4. *Düzenleyici Devrelerin Oluşturulması* başlığı altında sunulan, Devre 1’den Devre 8’e kadar farklı devre türlerini ara çıktılar olarak sunmaktadır. Ayrıca *miRCuit*, *miRNA-bağımlı lncRNA düzenleyici devreler* ve *TF-bağımlı lncRNA düzenleyici devreler* olmak üzere kullanıcılara iki özel düzenleyici devre sunmaktadır. Bu işlemler süresince, 11 farklı veri tabanından elde edilen veriler kullanılarak devrelerin oluşturulması sağlanmaktadır (Şekil 6.1.).



Şekil 6.1. *miRCuit* devre analiz programının ikili ve çoklu ilişkileri hesaplama ile düzenleyici devreleri kurgulamaya yönelik algoritmasının şematik gösterimi (a) Teorik ikili ilişkiler, b) Literatür tabanlı ikili ilişkiler, c) Ortak etkileşimler ve düzenleyici devreler.).

Devrelerin oluşturulması sırasında kullanılan veri tabanları, ikili molekül etkileşimlerini içeren (miRNA-mRNA, miRNA-lncRNA, miRNA-TF, TF-mRNA, TF-miRNA, lncRNA-mRNA, lncRNA-miRNA ve lncRNA-TF) veri tabanları olup elde edilen bu etkileşimler deneysel verilere, tahmin yaklaşımlarına, literatür verilerine ve hesaplamalı yaklaşımlara dayanmaktadır (Çizelge 4.2.). Tüm bu veri tabanlarının sisteme aktarılması ve biyolojik

devrelerin bu veri tabanlarının hepsinden gelen sonuçları temel olarak düzenleyici devreler ve ağlar kurgulaması, geliştirilen programın ve elde edilen devrelerin güvenilirliğini arttırmakta ve hedeflenen bütüncül analiz yaklaşımını desteklemektedir.

miRCuit devre analiz programı kullanılarak yapılan analiz sonucunda elde edilen devreler, kullanım ve analiz açısından kolaylık sağlayan doğrusal bir formatta kullanıcılara sunulmaktadır (Şekil 6.1c.). Lineer formatta verilen görsel dosyaların yanı sıra, kullanıcıların ileri düzey analizler için kullanabilecekleri tablo formatında dosyalar da her analiz için çıktı olarak verilmektedir. Ayrıca *miRCuit*, benzeri diğer araçlardan farklı olarak, kullanıcıların her bir ifade dosyası için eşik değerlerini seçmesine imkan sunmakta ve böylece araştırmacılara kendi örnek grupları arasında (örneğin tümör/normal, tedavi edilen/tedavi edilmeyen vb.) ne kadar ifade varyasyonunu dikkate almak istediklerini belirtebilecekleri bir bölüm sunmaktadır (Şekil 4.12.).

Programın bir diğer değerli uygulaması, analizden elde edilen gen listeleri ile Gen Zenginleştirme Analizi'nin (GSEA) gerçekleştirilebilmesidir. Bu şekilde, İnsan Moleküler İmza Veritabanı'na (The Molecular Signature Database, MSigDB) erişim sağlanmakta ve burada bulunan 9 ana koleksiyona ile alt koleksiyonlara ayrılmış toplamda 34.837 gen seti içerisinde devre elemanlarının zenginleştiği yolaklar belirlenmektedir (155,156). Analiz sonucunda zenginleşen genler ve yolaklar listelenerek bu yolaklar temelinde GSEA grafiği, nokta grafiği ve ağlar oluşturulmaktadır.

Bunlarla birlikte, programın veri dosyası olarak DEG listelerine ihtiyaç duyuyor olması, programın bir limitasyonu olarak değerlendirilebilir. Ayrıca, GSEA'nın kendi algoritmasının bir gereksinimi olarak Gen Zenginleştirme Analizi'ni gerçekleştirmek için yeterli sayıda gen içeren bir listeye duyulan gereksinim, programın bir diğer sınırlamasını oluşturmaktadır (154).

Böylelikle tez çalışması süresince tasarlanan programın literatürdeki 11 farklı veri tabanını tarayarak kanıtlanmış ikili ve üçlü ilişkileri kullanması ile bütünleşik devreler ortaya koyması, oluşturulan bu bütünleşik devrelerin karmaşık bir ağ yapısından ziyade okunması ve analiz edilmesi kolay lineer bir formatta sunulması, programın içerisinde kullanıcıya sunulan seçenekler ile analizin araştırmacının ihtiyaçlarına özgü kişiselleştirilmiş bir formatta veriliyor olması ve diğer araştırma alanlarından araştırmacılara farklı bir yaklaşım

getirme amacının bulunması ile *miRCuit*, biyolojik ağların ve gen düzenleme süreçlerinin daha derinlemesine anlaşılmasına katkı sağlayacak güçlü bir araç olma potansiyeline sahiptir.

miRCuit Kullanılarak Elde Edilen Sonuçların Değerlendirilmesi

Tez çalışması kapsamında, GEO'dan elde edilen 8 mikrodizin çalışmasına ait toplamda 490 invaziv duktal meme kanseri tümör ve 122 normal/eşli normal dokusu ile 5 Yeni Nesil Dizileme (NGS) çalışmasına ait toplamda 179 invaziv duktal meme kanseri tümör ve 51 normal/eşli normal dokusu ile DEG analizi gerçekleştirilmiş ve elde edilen sonuçlardan mRNA, miRNA, lncRNA ve TF veri listeleri oluşturulmuştur. Bu listeler hem mikrodizin hem de NGS çalışmaları için ayrı ayrı *miRCuit* programına yüklenerek analiz edilmiş, Gen Zenginleştirme Analizleri (GSEA) yapılmış ve grafikleri çizilmiştir. Elde edilen bulgular doğrultusunda, toplamda 8 farklı düzenleyici tip ile iki farklı özel tipte devre oluşturma kapasitesi olan *miRCuit* programından, mikrodizin verileri için 4, NGS verileri için 7 farklı tipte devre elde edilmiştir. Ayrıca, NGS verileri için miRNA bağımlı lncRNA düzenleyici devrelere ait toplam 7, TF bağımlı düzenleyici devrelere ait ise toplamda 169 devre oluşturulmuştur. GSEA analizi, mikrodizin ve NGS verilerinden elde edilen tüm düzenleyici elemanları içeren Devre 4'e uygulanmıştır. Sonuçta, her iki veri setinden gelen Devre 4'ün elemanları arasında ortak regülatif molekül bulunmamasına karşın, regüle ettikleri yolaklar kanser yolaklarında zenginleşmiş ve bu yolakların %41.9'u ortak bulunmuştur (Şekil 5.12.).

Bunun nedeninin, mikrodizin ve NGS yöntemlerinin farklı biyolojik özellikler ve ölçüm yöntemleri kullanıyor olmasından kaynaklandığı düşünülmektedir. NGS, mikrodizin teknolojisine göre daha geniş ve detaylı bir analiz sunarak RNA seviyelerindeki dinamik değişimleri yansıtmada daha hassas bir sonuç ortaya koymaktadır. Buna ek olarak, farklı analiz tekniklerinin kanser ile ilgili benzer yolaklar üzerinde zenginleşmesi, her iki veri setinden gelen moleküllerin de kanser biyolojisinde önemli olan yolakları vurguladığını göstermektedir (176,177).

Ayrıca mikrodizin verileri kullanılarak tüm devre türlerinin oluşturulamaması, mikrodizin teknolojisinin sınırlamalarından kaynaklanıyor olabilmektedir. Mikrodizin teknolojisinin sabit bir hedef kümesine odaklanması ve prob dizilerinin tanımlanmış lncRNA'ları yeterince kapsamıyor olması, bu sınırlamaların temel nedenleri arasında yer almaktadır. Başka bir

deyişle, NGS tabanlı çalışmalarda lncRNA'lar daha zengin bir şekilde ortaya koyulabilirken, mikrodizin tabanlı çalışmalarda bu molekül ile ilgili verilerin sınırlı olması, tüm devre türlerinin oluşturulamamasına neden olmuştur. İki farklı teknolojiye elde edilen sonuçların karşılaştırılması sürecinde, mikrodizin verilerindeki sınırlı lncRNA içeriğinin, analiz sonuçlarını etkileyen önemli bir faktör olduğu değerlendirilmiştir. Bu durum, mikrodizin tabanlı çalışmaların, özellikle lncRNA'ların kapsamlı bir şekilde analiz edilmesinde yeterli olmayabileceğini göstermektedir (176,177).

Meme Kanseri Gelişiminde lncRNA Tabanlı Düzenleyici Devrelerin Rolü ve Biyobelirteç Potansiyelleri

Meme kanseri oluşumu ve gelişiminde lncRNA'ların transkripsiyonel ve post-transkripsiyonel süreçler ile ilgili düzenleyici etkileri göz önünde bulundurulduğunda, bu rollerin devreler ve ağlar aracılığı ile ortaya koyulması önem kazanmaktadır. Bu amaçla tez çalışması kapsamında, miRCuit devre analiz programı, özel lncRNA tabanlı devreler oluşturacak şekilde tasarlanmış ve sonuçta NGS verilerinden miRNA bağımlı lncRNA düzenleyici devreler için toplamda 7, TF bağımlı lncRNA düzenleyici devreler için ise toplamda 169 devre oluşturulmuştur (EK 2).

NGS verilerinden elde edilen düzenleyici devrelerde, literatürle uyumlu olarak lncRNA CASC15, miR-130b-3p ve transkripsiyon faktörü KLF5, meme kanseri gelişiminde en önemli moleküller olarak öne çıkmıştır (Şekil 5.13.). LncRNA CASC15'in meme kanserinde yüksek oranda ifade edildiği, hücre proliferasyonunu, invazyonu ve tümör büyümesini tetiklediği ve onkogen olarak işlev gördüğü bilinmektedir (178–180). Transkripsiyon faktörü KLF5 ise, meme kanseri hücrelerinin göç ve invazyonunda kritik bir rol oynamaktadır. Zhou ve arkadaşlarının yaptığı çalışmada KLF5'in meme kanseri hücrelerinde baskılanmasının, hücre proliferasyonunu ve metastazı azalttığı gösterilmiştir (181). Diğer bir önemli devre elemanı olan miR-130b-3p, meme kanserinde hem onkojenik hem de tümör baskılayıcı etkiler gösterebilmektedir. miR-130b-3p'nin meme kanseri hastalarında yüksek oranda ifade edildiği, bununla birlikte PTEN genini hedef alarak hastalarda kemoterapi direnci geliştirdiği Miao ve arkadaşları tarafından yapılan çalışmalarda ortaya konmuştur (182). miR-130b-3p'nin meme kanserinde artan ifadesinin, hücre invazyonu ve göçünde önemli bir rol oynadığı gösterilmiştir. Bu nedenle, miR-130b-

3p'nin meme kanseri metastazı için potansiyel bir biyobelirteç olabileceği düşünülmektedir (183). Bununla birlikte Shui ve arkadaşlarının yaptığı çalışmada, invaziv meme kanseri MDA-MB-231 hücrelerinde, erken evre invaziv olmayan meme kanseri MCF-7 hücrelerine kıyasla miR-130b-3p'nin ifadesinin görece düşük olduğu gösterilmiş ve miR-130b-3p'nin hücre invazyonu ve göçünü *in vitro* ortamda anlamlı şekilde baskılayabileceğini ortaya koymuştur. Yapılan bu çalışmada, miR-130b-3p'nin meme kanseri hücrelerinin metastazını Notch ligandını (Delta-like 1) hedef alarak baskılamakta olduğu gösterilmiş ve MMP-9 ve VEGF gibi metastaz ile ilgili proteinlerin ifade seviyelerini azaltarak hücrelerin invazyonunu ve migrasyonunu inhibe ettiği kaydedilmiştir (183).

Meme kanseri oluşumu ve gelişimi mekanizmalarının derinlemesine aydınlatılmasında CASC15, KLF5 ve miR-130b-3p'nin regülatif ilişkilerinin incelenmesi önem kazanmaktadır. Bu bağlamda farklı düzenleyici devreler, sürecin aydınlatılmasına katkıda bulunsa da, devre analizi sonucu ortaya koyulan ve IDC/normal dokular arasında anlamlı sonuçlar veren devrelerin bütüncül yaklaşımına dair literatürde çalışmalar bulunmamaktadır. Örneğin, yapılan bir çalışmada CASC15'in meme kanserinde miR-153-3p'yi baskılayarak KLF5'in ifadesini arttırdığı ve bu mekanizmanın KLF5/CASC15/miR-153-3p arasında bir pozitif besleme döngüsü oluşturduğu ortaya konmuştur (178). Başka bir çalışma, miR-217'nin üçlü negatif meme kanseri hücrelerinde KLF5'in ifadesini baskıladığını ve bu durumun hücre proliferasyonu ve invazyonunun azalmasına neden olduğunu ortaya koymuştur. Bu bağlamda miR-217/KLF5 eksenini, terapötik bir hedef olarak değerlendirilmiştir (181). Yu ve arkadaşlarının yaptığı çalışma ise, CASC15'in, miR-130b-3p'yi baskılayarak metastaz ve proliferasyonun desteklendiğini belirtmektedir (178). Sonuçta CASC15, miR-130b-3p ve KLF5'in meme kanseri patogenezi ve progresyonunda önemli düzenleyici moleküller olduğu belirlenmiştir.

Buna ek olarak yapılan TCGA veri analizi, miR-130b-3p'nin hedefleri üzerindeki baskılayıcı etkisini ve KLF5 ile CASC15 arasındaki pozitif ilişkiyi doğrulamıştır. Bulgular, meme kanserinin patogenezinde lncRNA-miRNA-TF-mRNA etkileşimlerinin önemini vurgulayan düzenleyici devrelerle uyumlu bulunmuştur. Özellikle bu moleküllerin terapötik hedefler olarak yüksek bir potansiyele sahip olması ve meme kanseri tedavisinde yeni yaklaşımlar sunabilme potansiyeli, CASC15, KLF5 ve miR-130b-3p devrelerinin

düzenleyici rolü ve mekanizmalarının biyobelirteç olarak değerlendirilebilmesi açısından büyük önem taşımaktadır.

Tez çalışması kapsamında tasarlanmış olan *miRCuit* devre analiz programı kullanılarak oluşturulan devrelerin literatür ile uyumlu sonuçlar vermiş olması, geliştirilen bu yeni programın güvenilirliğini desteklemektedir. Bu bağlamda, *miRCuit* tarafından oluşturulan diğer devrelerin de biyobelirteç potansiyelleri açısından incelenmeye değer oldukları düşünülmektedir.

Sonuç olarak, *miRCuit*, ifade verilerinin analizine dayalı analitik bir yaklaşım benimseyerek çeşitli hastalıklar ve biyolojik koşullar için düzenleyici devreler ve ağlar tasarlamak amacıyla kullanılabilecek bir program özelliği taşımaktadır. Bu yaklaşımın, meme kanseri de dahil olmak üzere çeşitli hastalıklar için yeni biyobelirteçlerin önerilmesini ve gelecekte etkili tedavi stratejilerinin geliştirilmesini mümkün kılacağı düşünülmektedir.

KAYNAKLAR

1. GLOBOCAN. International Agency for Research on Cancer, Source [Internet]. 2020 [cited 2024 Oct 14]. Available from: <https://gco.iarc.fr/today/data/factsheets/cancers/>
2. Harbeck N, Gnant M. Breast cancer. *The Lancet*. 2017;389(10074):1134–50.
3. Polyak K. Heterogeneity in breast cancer. *Journal of Clinical Investigation*. 2011;121(10):3786–8.
4. Nasirpour MH, Sabery Anvar M, Alireza N, Mahdih S, Soheil S, Zarrin M. Gene Expression Profiles Reveal Potential Targets for Breast Cancer Diagnosis and Treatment [Internet]. 2022. Available from: <http://biorxiv.org/lookup/doi/10.1101/2022.09.03.504469>
5. Mitsis T, Efthimiadou A, Bacopoulou F, Vlachakis D, Chrousos GP, Eliopoulos E. Transcription factors and evolution: An integral part of gene expression (Review). *World Acad Sci J*. 2020;2(1):3–8.
6. Cannell IG, Kong YW, Bushell M. How do microRNAs regulate gene expression? *Biochem Soc Trans*. 2008;36(6):1224–31.
7. Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, et al. The Human Transcription Factors. *Cell* [Internet]. 2018;172(4):650–65. Available from: <https://doi.org/10.1016/j.cell.2018.01.029>
8. Shaath H, Elango R, Alajez NM. Molecular classification of breast cancer utilizing long non-coding rna (Lncrna) transcriptomes identifies novel diagnostic lncrna panel for triple-negative breast cancer. *Cancers (Basel)*. 2021;13(21).
9. Statello L, Guo CJ, Chen LL, Huarte M. Gene regulation by long non-coding RNAs and its biological functions. *Nat Rev Mol Cell Biol* [Internet]. 2021 Feb 22;22(2):96–118. Available from: <http://dx.doi.org/10.1038/s41580-020-00315-9>
10. Wu Z, Liu X, Liu L, Deng H, Zhang J, Xu Q, et al. Regulation of lncRNA expression. *Cell Mol Biol Lett*. 2014;19(4):561–75.

11. Dong H, Wang W, Mo S, Chen R, Zou K, Han J, et al. SP1-induced lncRNA AGAP2-AS1 expression promotes chemoresistance of breast cancer by epigenetic regulation of MyD88. *Journal of Experimental and Clinical Cancer Research*. 2018;37(1):1–15.
12. Wang Y, Zhou Y, Yang Z, Chen B, Huang W, Liu Y, et al. MiR-204/ZEB2 axis functions as key mediator for MALAT1-induced epithelial-mesenchymal transition in breast cancer. *Tumor Biology*. 2017;39(7):1–8.
13. Yang G, Lu X, Yuan L. LncRNA: A link between RNA and cancer. *Biochim Biophys Acta Gene Regul Mech* [Internet]. 2014;1839(11):1097–109. Available from: <http://dx.doi.org/10.1016/j.bbagr.2014.08.012>
14. Feng T, Shao F, Wu Q, Zhang X, Xu D, Qian K, et al. miR-124 downregulation leads to breast cancer progression via LncRNA-MALAT1 regulation and CDK4/E2F1 signal activation. *Oncotarget*. 2016;7(13):16205–16.
15. Wan X, Hao S, Hu C, Qu R. Identification of a novel lncRNA-miRNA-mRNA competing endogenous RNA network associated with prognosis of breast cancer. *J Biochem Mol Toxicol*. 2022 Aug 1;36(8).
16. Marrugo-Padilla A, Márquez-Lázaro J, Álviz-Amador A. Identification of prognostic biomarkers of invasive ductal carcinoma by an integrated bioinformatics approach. *F1000Res*. 2022 Sep 21;11:1075.
17. Lu C, Wei D, Zhang Y, Wang P, Zhang W. Long Non-Coding RNAs as Potential Diagnostic and Prognostic Biomarkers in Breast Cancer: Progress and Prospects. *Front Oncol*. 2021;11(August):1–17.
18. WHO. Launches new roadmap on breast cancer [Internet]. 2023 [cited 2023 Jul 10]. Available from: <https://www.who.int/news/item/03-02-2023-who-launches-new-roadmap-on-breast-cancer>
19. Houghton SC, Hankinson SE. Cancer Progress and Priorities: Breast Cancer. *Cancer Epidemiol Biomarkers Prev*. 2021 May 1;30(5):822–44.

20. IARC. International Agency for Research on Cancer [Internet]. 2020 [cited 2023 Jul 15]. Available from: <https://www.iarc.who.int/cancer-topics/>
21. Breast Cancer Facts and Statistics. BREASTCANCER.ORG [Internet]. 2023 [cited 2023 Jul 10]. Available from: <https://www.breastcancer.org/facts-statistics#section-how-common-is-breast-cancer>
22. Rakha EA, Tse GM, Quinn CM. An update on the pathological classification of breast cancer. *Histopathology*. 2023;82(1):5–16.
23. Nascimento RG do, Otoni KM. Histological and molecular classification of breast cancer: what do we know? *Mastology*. 2020;30:1–8.
24. Centers for Disease Control and Prevention What is Breast Cancer? CDC [Internet]. 2022 [cited 2023 Jul 17]. Available from: https://www.cdc.gov/cancer/breast/basic_info/what-is-breast-cancer.htm
25. Malhotra GK, Zhao X, Band H, Band V. Histological, molecular and functional subtypes of breast cancers. *Cancer Biol Ther*. 2010;10(10):955–60.
26. Tsuda H, Tsugawa K, Akiyama F, Horii R, Kurosumi M, Moriya T, et al. Histological classification of breast tumors in the General Rules for Clinical and Pathological Recording of Breast Cancer (18th edition). *Breast Cancer*. 2020;27(3):309–21.
27. Vuong D, Simpson PT, Green B, Cummings MC, Lakhani SR. Molecular classification of breast cancer. *Virchows Archiv*. 2014;465(1):1–14.
28. Lakhani SR, Ellis IO, Schnitt SJ, Tan PH, van de Vijver MJ. WHO Classification of Tumours of the Breast. WHO Classi. 2012.
29. Bloom HJ, Richardson WW. Histological grading and prognosis in breast cancer a study of 1409 cases of which 359 have been followed for 15 years. *Br J Cancer*. 1957;11(3):359–77.

30. Elston CW, Ellis O. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology*. 1991;19:403–10.
31. Blamey RW, Pinder SE, Ball GR, Ellis IO, Elston CW, Mitchell MJ, et al. Reading the prognosis of the individual with breast cancer. *Eur J Cancer*. 2007;43(10):1545–7.
32. Rakha EA, El-Sayed ME, Lee AHS, Elston CW, Grainge MJ, Hodi Z, et al. Prognostic significance of nottingham histologic grade in invasive breast carcinoma. *Journal of Clinical Oncology*. 2008;26(19):3153–8.
33. Ravdin PM, Siminoff LA, Davis GJ, Mercer MB, Hewlett J, Gerson N, et al. Computer program to assist in making decisions about adjuvant therapy for women with early breast cancer. *Journal of Clinical Oncology*. 2001;19(4):980–91.
34. Amin MB, Greene FL, Edge SB, Compton CC, Gershenwald JE, Brookland RK, et al. The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more “personalized” approach to cancer staging. *CA Cancer J Clin*. 2017;67(2):93–9.
35. Peto R, Boreham J, Clarke M, Davies C, Beral C. UK and USA breast cancer deaths down 25% in year 2000 at ages 20-69 years. *Lancet*. 2000;355(9217):1822.
36. Afzal S, Hassan M, Ullah S, Abbas H, Tawakkal F, Khan MA. Breast Cancer; Discovery of Novel Diagnostic Biomarkers, Drug Resistance, and Therapeutic Implications. *Front Mol Biosci*. 2022;9(February):1–10.
37. Breast cancer biomarkers. abcam [Internet]. 2023 [cited 2023 Jul 18]. Available from: <https://www.abcam.com/cancer/breast-cancer-biomarkers>
38. Hayes DF, Bast RC, Desch CE, Fritsche H, Kemeny NE, Jessup JM, et al. Tumor marker utility grading system: A framework to evaluate clinical utility of tumor markers. *J Natl Cancer Inst*. 1996;88(20):1456–66.

39. Ellis IO, Carder P, Hales S, Lee AH, Pinder SE, Rakha E, et al. Pathology reporting of breast disease in surgical excision specimens incorporating the dataset for histological reporting of breast cancer. The Royal College of Pathologists. 2016;G148(version 2):1–160.
40. Hammond MEH, Hayes DF, Dowsett M, Allred DC, Hagerty KL, Badve S, et al. American society of clinical oncology/college of American pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer (unabridged version). Arch Pathol Lab Med. 2010;134(7).
41. Rakha EA, Chmielik E, Schmitt FC, Tan PH, Quinn CM, Gallaghy G. Assessment of Predictive Biomarkers in Breast Cancer: Challenges and Updates. Pathobiology. 2022;89(5):263–77.
42. Rakha EA, Pinder SE, Bartlett JMS, Ibrahim M, Starczynski J, Carder PJ, et al. Updated UK recommendations for HER2 assessment in breast cancer. J Clin Pathol. 2015;68(2):93–9.
43. Rakha EA, Reis-Filho JS, Ellis IO. Combinatorial biomarker expression in breast cancer. Breast Cancer Res Treat. 2010;120(2):293–308.
44. Rakha EA, El-Sayed ME, Green AR, Paish EC, Powe DG, Gee J, et al. Biologic and clinical characteristics of breast cancer with single hormone receptor-positive phenotype. Journal of Clinical Oncology. 2007;25(30):4772–8.
45. Cuzick J, Dowsett M, Pineda S, Wale C, Salter J, Quinn E, et al. Prognostic value of a combined estrogen receptor, progesterone receptor, Ki-67, and human epidermal growth factor receptor 2 immunohistochemical score and comparison with the genomic health recurrence score in early breast cancer. Journal of Clinical Oncology. 2011;29(32):4273–8.
46. Regan MM, Viale G, Mastropasqua MG, Maiorano E, Golouh R, Carbone A, et al. Re-evaluating adjuvant breast cancer trials: Assessing hormone receptor status by immunohistochemical versus extraction assays. J Natl Cancer Inst. 2006;98(21):1571–81.

47. Rydén L, Jönsson PE, Chebil G, Dufmats M, Fernö M, Jirström K, et al. Two years of adjuvant tamoxifen in premenopausal patients with breast cancer: A randomised, controlled trial with long-term follow-up. *Eur J Cancer*. 2005;41(2):256–64.
48. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, et al. A Multigene Assay to Predict Recurrence of Tamoxifen-Treated, Node-Negative Breast Cancer. *New England Journal of Medicine* [Internet]. 2004 Dec 30;351(27):2817–26. Available from: <http://www.tandfonline.com/doi/abs/10.1080/14733400500093379>
49. Wang Y, Klijn JGM, Zhang Y, Sieuwerts AM, Look MP, Yang F, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet*. 2005;365(9460):671–9.
50. Masood S. Breast cancer subtypes: Morphologic and biologic characterization. *Women's Health*. 2016;12(1):103–19.
51. Burstein MD, Tsimelzon A, Poage GM, Covington KR, Contreras A, Fuqua SAW, et al. Comprehensive genomic analysis identifies novel subtypes and targets of triple-negative breast cancer. *Clinical Cancer Research*. 2015;21(7):1688–98.
52. Prat A, Perou CM. Deconstructing the molecular portraits of breast cancer. *Mol Oncol*. 2011;5(1):5–23.
53. Sørlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A*. 2003;100(14):8418–23.
54. Sørlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A*. 2001;98(19):10869–74.
55. Perou CM, Sørlie T, Eisen MB, Rijn M Van De, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. *Nature* [Internet]. 2000;533(May):747–52. Available from: www.stanford.edu/molecularportraits/

56. Goldhirsch A, Winer EP, Coates AS, Gelber RD, Piccart-Gebhart M, Thürlimann B, et al. Personalizing the treatment of women with early breast cancer: Highlights of the st gallen international expert consensus on the primary therapy of early breast Cancer 2013. *Annals of Oncology*. 2013;24(9):2206–23.
57. Mayrovitz HN. *Breast Cancer*. 2022. 31–42 p.
58. Zhang X. Molecular Classification of Breast Cancer Relevance and Challenges. *Arch Pathol Lab Med*. 2023;147(1):46–51.
59. Bernard PS, Parker JS, Mullins M, Cheung MCU, Leung S, Voduc D, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *Journal of Clinical Oncology*. 2009;27(8):1160–7.
60. Melisko M. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *Women's Oncology Review*. 2005;5(1):45–7.
61. Vijver V De. A Gene-Expression Signature As a Predictor of Survival in Breast Cancer. 2008;347(25):1999–2009.
62. Ghafouri-Fard S, Abak A, Fattahi F, Hussen BM, Bahroudi Z, Shoorei H, et al. The interaction between miRNAs/lncRNAs and nuclear factor- κ B (NF- κ B) in human disorders. Vol. 138, *Biomedicine and Pharmacotherapy*. Elsevier Masson s.r.l.; 2021.
63. Lozano-Velasco E, Garcia-Padilla C, Carmona-Garcia M, Gonzalez-Diaz A, Arequipa-Rendon A, Aranega AE, et al. MEF2C Directly Interacts with Pre-miRNAs and Distinct RNPs to Post-Transcriptionally Regulate miR-23a-miR-27a-miR-24-2 microRNA Cluster Member Expression. *Noncoding RNA*. 2024 Jun 1;10(3).
64. Pande A. Co-regulatory network of transcription factor and microrna: A key player of gene regulation. Vol. 5, *Biomedical and Biotechnology Research Journal*. Wolters Kluwer Medknow Publications; 2021. p. 374–9.
65. Rani V, Sengar RS. Biogenesis and mechanisms of microRNA-mediated gene regulation. Vol. 119, *Biotechnology and Bioengineering*. John Wiley and Sons Inc; 2022. p. 685–92.

66. Maston GA, Evans SK, Green MR. Transcriptional regulatory elements in the human genome. *Annu Rev Genomics Hum Genet.* 2006;7:29–59.
67. Dolores Delgado M, León J. Gene expression regulation and cancer. *Clinical and Translational Oncology.* 2006;8(11):780–7.
68. Latchman DS. Transcription factors: An overview. *International Journal of Biochemistry and Cell Biology.* 1997;29(12):1305–12.
69. Smale ST, Kadonaga JT. The RNA polymerase II core promoter. *Annu Rev Biochem.* 2003;72:449–79.
70. Hahn S. Structure and mechanism of the RNA polymerase II transcription machinery. *Nat Struct Mol Biol.* 2004;11(5):394–403.
71. Klug WS, Cummings MR, Spencer CA, Palladino MA, Killian D. *Concepts of Genetics.* 12th ed. 2019. 864 p.
72. Ambros V, Bartel B, Bartel DP, Burge CB, Carrington JC, Chen X, et al. A uniform system for microRNA annotation. 2003;277–9.
73. Bartel DP. MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. *Cell.* 2004;116(2):281–97.
74. Liu B, Li J, Cairns MJ. Identifying miRNAs, targets and functions. *Brief Bioinform.* 2014;15(1):1–19.
75. Jia Z, An J, Liu Z, Zhang F. Non-Coding RNAs in Colorectal Cancer: Their Functions and Mechanisms. *Front Oncol.* 2022;12(February):1–18.
76. Fabian MR, Sonenberg N, Filipowicz W. Regulation of mRNA translation and stability by microRNAs. *Annu Rev Biochem.* 2010;79:351–79.
77. Mendell JT, Olson EN. MicroRNAs in stress signaling and human disease. *Cell* [Internet]. 2012;148(6):1172–87. Available from: <http://dx.doi.org/10.1016/j.cell.2012.02.005>

78. Yong SL, Dutta A. MicroRNAs in cancer. *Annual Review of Pathology: Mechanisms of Disease*. 2009;4:199–227.
79. Schwerdtfeger M, Desiderio V, Kobold S, Regad T, Zappavigna S, Caraglia M. Long non-coding RNAs in cancer stem cells. *Transl Oncol* [Internet]. 2021;14(8):101134. Available from: <https://doi.org/10.1016/j.tranon.2021.101134>
80. Volders PJ, Helsens K, Wang X, Menten B, Martens L, Gevaert K, et al. LNCipedia: A database for annotated human lncRNA transcript sequences and structures. *Nucleic Acids Res*. 2013;41(D1):246–51.
81. Ma H, Hao Y, Dong X, Gong Q, Chen J, Zhang J, et al. Molecular mechanisms and function prediction of long noncoding RNA. *The Scientific World Journal*. 2012;2012(1).
82. Tani H, Imamachi N, Mizutani R, Imamura K, Kwon Y, Miyazaki S, et al. Genome-wide analysis of long noncoding RNA turnover. *Methods in Molecular Biology*. 2015;1262:305–20.
83. Hauptman N, Glavač D. Long non-coding RNA in cancer. *Int J Mol Sci*. 2013;14(3):4655–69.
84. Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. 2009 Mar 10;3(155–9).
85. Sun M, Gadad SS, Kim DS, Kraus WL. Discovery, Annotation, and Functional Analysis of Long Noncoding RNAs Controlling Cell-Cycle Gene Expression and Proliferation in Breast Cancer Cells. *Mol Cell* [Internet]. 2015;59(4):698–711. Available from: <http://dx.doi.org/10.1016/j.molcel.2015.06.023>
86. Poursheikhani A, Abbaszadegan MR, Kerachian MA. Mechanisms of long non-coding RNA function in colorectal cancer tumorigenesis. *Asia Pac J Clin Oncol*. 2021;17(1):7–23.

87. Rao MRS. *Advances in Experimental Medicine and Biology: Long Non Coding RNA Biology* [Internet]. 2017. 323 p. Available from: <http://www.springer.com/series/5584>
88. Yu Y, Nangia-Makker P, Farhana L, Majumdar APN. A novel mechanism of lncRNA and miRNA interaction: CCAT2 regulates miR-145 expression by suppressing its maturation process in colon cancer cells. *Mol Cancer*. 2017;16(1):1–11.
89. Zhu S, Chen CY, Hao Y. LncRNA KCNQ1OT1 acts as miR-216b-5p sponge to promote colorectal cancer progression via up-regulating ZNF146. *J Mol Histol* [Internet]. 2021;52(3):479–90. Available from: <https://doi.org/10.1007/s10735-020-09942-0>
90. Du L, Li J, Tian Y, Feng R. Pleckstrin-2-promoted PPM1B degradation plays an important role in transforming growth factor- β -induced breast cancer cell invasion and metastasis. *Cancer Sci*. 2023 Jun 1;114(6):2429–44.
91. Willis S, De P, Dey N, Long B, Young B, Sparano JA, et al. Enriched transcription factor signatures in triple negative breast cancer indicates possible targeted therapies with existing drugs. *Meta Gene*. 2015 Jun 1;4:129–41.
92. Wu Y, Li M, Lin J, Hu C. Hippo/TEAD4 signaling pathway as a potential target for the treatment of breast cancer (Review). Vol. 21, *Oncology Letters*. Spandidos Publications; 2021.
93. Denli AM, Tops BBJ, Plasterk RHA, Ketting RF, Hannon GJ. Processing of primary microRNAs by the Microprocessor complex. *Nature* [Internet]. 2004 Nov;432:231–5. Available from: www.nature.com/nature
94. He K, Li WX, Guan D, Gong M, Ye S, Fang Z, et al. Regulatory network reconstruction of five essential microRNAs for survival analysis in breast cancer by integrating miRNA and mRNA expression datasets. *Funct Integr Genomics*. 2019 Jul 1;19(4):645–58.
95. Meltzer Paul S. Small RNAs with big impacts. *Nature*. 2005 Jun 9;435:745–6.

96. Amir S, Simion C, Umeh-Garcia M, Krig S, Moss T, Carraway KL, et al. Regulation of the T-box transcription factor Tbx3 by the tumour suppressor microRNA-206 in breast cancer. *Br J Cancer*. 2016 May 10;114(10):1125–34.
97. Angelis ML De, Francescangeli F, Zeuner A, Baiocchi M. Colorectal Cancer Stem Cells : An Overview of Evolving Methods and Concepts. 2021;
98. Li A, Mallik S, Luo H, Jia P, Lee DF, Zhao Z. H19, a Long Non-coding RNA, Mediates Transcription Factors and Target Genes through Interference of MicroRNAs in Pan-Cancer. *Mol Ther Nucleic Acids*. 2020 Sep 4;21:180–91.
99. Liu Y, Chen X, Chen J, Luo Y, Chen Z, Lin D, et al. Gel-Free Single-Cell Culture Arrays on a Microfluidic Chip for Highly Efficient Expansion and Recovery of Colon Cancer Stem Cells. *ACS Biomater Sci Eng*. 2022;8(8):3623–32.
100. Zhang Z, Zhang J, Diao L, Han L. Small non-coding RNAs in human cancer: function, clinical utility, and characterization. *Oncogene* [Internet]. 2021 Mar 4;40(9):1570–7. Available from: <http://dx.doi.org/10.1038/s41388-020-01630-3>
101. Afshar AS, Xu J, Goutsias J. Integrative identification of deregulated MiRNA/TF-mediated gene regulatory loops and networks in prostate cancer. *PLoS One*. 2014;9(6).
102. Bai S, Wu YY, Yan Y, Shao S, Zhang J, Liu J, et al. Construct a circRNA/miRNA/mRNA regulatory network to explore potential pathogenesis and therapy options of clear cell renal cell carcinoma. *Sci Rep* [Internet]. 2020;10(1):1–15. Available from: <https://doi.org/10.1038/s41598-020-70484-2>
103. Chen H, Xu Z, Liu D. Small non-coding RNA and colorectal cancer. *J Cell Mol Med*. 2019;23(5):3050–7.
104. Dario Veneziano, Gioacchino P. Marceca, Sebastiano Di Bella, Giovanni Nigita, Rosario Distefano, Carlo M. Croce. Investigating miRNA–lncRNA Interactions: Computational Tools and Resources. *MIMB*,volum.

105. Öztemur Islakoğlu Y, Noyan S, Gür Dedeoğlu B. Hsa-miR-301a- and SOX10-dependent miRNA-TF-mRNA regulatory circuits in breast cancer. *Turkish Journal of Biology*. 2018;42(2):103–12.
106. Shen S, Li K, Liu Y, Liu X, Liu B, Ba Y, et al. Silencing lncRNA AGAP2-AS1 Upregulates miR-195-5p to Repress Migration and Invasion of EC Cells via the Decrease of FOSL1 Expression. *Mol Ther Nucleic Acids* [Internet]. 2020 Jun;20(127):331–44. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S2162253120300202>
107. Tjian R. Molecular machines that control genes. *Sci Am*. 1995;272(2):54–61.
108. Chang K, Kuo W. [GEO dataset]. 2015 [cited 2024 Oct 14]. Deregulated microRNAs in triple-negative breast cancer revealed by deep sequencing. Available from: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE40049>
109. Persson H, Kvist A, Rego N, Staaf J, Vallon-Christersson J, Luts L, et al. Identification of new microRNAs in paired normal and tumor breast tissue suggests a dual role for the ERBB2/Her2 gene. *Cancer Res*. 2011 Jan 1;71(1):78–86.
110. Farazi TA, Horling HM, ten Hoeve J, Mihailovic A, Halfwerk H, Morozov P, et al. [GEO dataset]. 2011 [cited 2024 Oct 14]. MicroRNA sequence and expression analysis in breast tumors by deep sequencing [miRNA sequence data]. Available from: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi>
111. Pang D XS. [GEO dataset]. 2017 [cited 2024 Oct 14]. Next Generation Sequencing Transcriptomes of breast cancer. Available from: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi>
112. Kang K. [GEO dataset]. 2018 [cited 2024 Oct 14]. A practical evaluation of alignment algorithms for RNA variant calling analysis. Available from: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi>
113. Casey T, Bond J, Tighe S, Hunter T, Lintault L, Patel O, et al. Molecular signatures suggest a major role for stromal cells in development of invasive breast cancer. *Breast Cancer Res Treat*. 2009 Mar;114(1):47–62.

114. Wen WX, Soo JSS, Kwan PY, Hong E, Khang TF, Mariapun S, et al. Germline APOBEC3B deletion is associated with breast cancer risk in an Asian multi-ethnic cohort and with immune cell presentation. *Breast Cancer Research*. 2016 May 27;18(1).
115. Gruosso T, Mieulet V, Cardon M, Bourachot B, Kieffer Y, Devun F, et al. Chronic oxidative stress promotes H2 AX protein degradation and enhances chemosensitivity in breast cancer patients . *EMBO Mol Med*. 2016 May;8(5):527–49.
116. Liu YR, Jiang YZ, Xu XE, Hu X, Yu K Da, Shao ZM. Comprehensive transcriptome profiling reveals multigene signatures in triple-negative breast cancer. *Clinical Cancer Research*. 2016 Apr 1;22(7):1653–62.
117. Hawthorn L, Luce J, Stein L, Rothschild J. Integration of transcript expression, copy number and LOH analysis of infiltrating ductal carcinoma of the breast [Internet]. 2010. Available from: <http://www.biomedcentral.com/1471-2407/10/460>
118. Lesurf R, Aure MR, Mørk HH, Vitelli V, Lundgren S, Børresen-Dale AL, et al. Molecular Features of Subtype-Specific Progression from Ductal Carcinoma In Situ to Invasive Breast Cancer. *Cell Rep*. 2016 Jul 26;16(4):1166–79.
119. Biagioni F, Bossel Ben-Moshe N, Fontemaggi G, Canu V, Mori F, Antoniani B, et al. miR-10b*, a master inhibitor of the cell cycle, is down-regulated in human breast tumours. *EMBO Mol Med*. 2012 Nov;4(11):1214–29.
120. Jeggari A, Marks DS, Larsson E. miRcode: A map of putative microrna target sites in the long non-coding transcriptome. *Bioinformatics*. 2012;28(15):2062–3.
121. Vlachos IS, Paraskevopoulou MD, Karagkouni D, Georgakilas G, Vergoulis T, Kanellos I, et al. DIANA-TarBase v7.0: Indexing more than half a million experimentally supported miRNA:mRNA interactions. *Nucleic Acids Res*. 2015;43(D1):D153–9.
122. Pian C, Zhang G, Tu T, Ma X, Li F. LncCeRBase: A database of experimentally validated human competing endogenous long non-coding RNAs. *Database*. 2018 Jan 1;2018(2018).

123. Zhao H, Yin X, Xu H, Liu K, Liu W, Wang L, et al. LncTarD 2.0: an updated comprehensive database for experimentally-supported functional lncRNA–target regulations in human diseases. *Nucleic Acids Res.* 2023;51(D1):D199–207.
124. Chen Y, Wang X. MiRDB: An online database for prediction of functional microRNA targets. *Nucleic Acids Res.* 2020;48(D1):D127–31.
125. Hsu S Da, Lin FM, Wu WY, Liang C, Huang WC, Chan WL, et al. MiRTarBase: A database curates experimentally validated microRNA-target interactions. *Nucleic Acids Res.* 2011;39(SUPPL. 1):163–9.
126. Huang HY, Lin YCD, Cui S, Huang Y, Tang Y, Xu J, et al. MiRTarBase update 2022: An informative resource for experimentally validated miRNA-target interactions. *Nucleic Acids Res.* 2022;50(D1):D222–30.
127. Liska O, Bohár B, Hidas A, Korcsmáros T, Papp B, Fazekas D, et al. TFLink: An integrated gateway to access transcription factor-target gene interactions for multiple species. *Database.* 2022;2022.
128. Han H, Cho JW, Lee S, Yun A, Kim H, Bae D, et al. TRRUST v2: An expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic Acids Res.* 2018 Jan 1;46(D1):D380–6.
129. Gong J, Shao D, Xu K, Lu Z, Lu ZJ, Yang YT, et al. RISE: A database of RNA interactome from sequencing experiments. *Nucleic Acids Res.* 2018;46(D1):D194–201.
130. Agarwal V, Bell GW, Nam JW, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. *Elife.* 2015 Jun 12;4(e05005):1–38.
131. Tong Z, Cui Q, Wang J, Zhou Y. TransmiR v2.0: An updated transcription factor-microRNA regulation database. *Nucleic Acids Res.* 2019;47(D1):D253–8.
132. Sean D, Meltzer PS. GEOquery: A bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics.* 2007 Jul 15;23(14):1846–7.

133. Gautier L, Cope L, Bolstad BM, Irizarry RA. Affy - Analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics*. 2004 Feb 12;20(3):307–15.
134. Carvalho BS, Irizarry RA. A framework for oligonucleotide microarray preprocessing. *Bioinformatics*. 2010 Aug 5;26(19):2363–7.
135. Kim C, Zheng H. RMA with quantile normalization mixes biological signals between different sample groups in microarray data analysis. 2014.
136. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015 Jan 6;43(7):e47.
137. Jung D, Kim B, Freishtat RJ, Giri M, Hoffman E, Seo J. MiRTarVis: An interactive visual analysis tool for microRNA-mRNA expression profile data. *BMC Proc*. 2015 Aug 13;9.
138. Nersisyan S, Galatenko A, Galatenko V, Shkurnikov M, Tonevitsky A. miRGTF-net: Integrative miRNA-gene-TF network analysis reveals key drivers of breast cancer recurrence. *PLoS One*. 2021 Apr 1;16(4 April).
139. da Silveira WA, Renaud L, Simpson J, Glen WB, Hazard ES, Chung D, et al. miRmapper: A tool for interpretation of miRNA–mRNA interaction networks. *Genes (Basel)*. 2018;9(9).
140. Wang J. TTF: An R-Based Integrative Tool for Decoding Human Transcription Factor–Target Interactions. *Biomolecules*. 2024 Jul 1;14(7).
141. Yang Y, Wang D, Miao YR, Wu X, Luo H, Cao W, et al. lncRNASNP v3: an updated database for functional variants in long non-coding RNAs. *Nucleic Acids Res*. 2023 Jan 6;51(1 D):D192–8.
142. Maurya A, Szymanski M, Karlowski WM. ARA: a flexible pipeline for automated exploration of NCBI SRA datasets. *Gigascience*. 2023;12.

143. Leggett RM, Ramirez-Gonzalez RH, Clavijo BJ, Waite D, Davey RP. Sequencing quality assessment tools to enable data-driven informatics for high throughput genomics. *Front Genet.* 2013;4(DEC):1–5.
144. Kim D, Langmead B, Salzberg SL. HISAT: A fast spliced aligner with low memory requirements. *Nat Methods.* 2015 Mar 31;12(4):357–60.
145. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009 Aug;25(16):2078–9.
146. Liao Y, Smyth GK, Shi W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics.* 2014 Apr 1;30(7):923–30.
147. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014 Dec 5;15(12).
148. Muzellec B, Teleńczuk M, Cabeli V, Andreux M. PyDESeq2: a python package for bulk RNA-seq differential expression analysis. *Bioinformatics.* 2023 Sep 1;39(9).
149. Sanbomics. Bioinformatic Services [Internet]. 2024 [cited 2024 Oct 14]. Available from: <https://www.sanbomics.com>
150. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.* 2006;34(Database issue).
151. Kozomara A, Birgaoanu M, Griffiths-Jones S. MiRBase: From microRNA sequences to function. *Nucleic Acids Res.* 2019 Jan 8;47(D1):D155–62.
152. Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, et al. The Human Transcription Factors. Vol. 172, Cell. Cell Press; 2018. p. 650–65.
153. Shen X, Kong S, Yang Q, Yin Q, Cong H, Wang X, et al. PCAT-1 promotes cell growth by sponging miR-129 via MAP3K7/NF- κ B pathway in multiple myeloma. *J Cell Mol Med.* 2020;24(6):3492–503.

154. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles [Internet]. 2005. Available from: www.pnas.org/cgi/doi/10.1073/pnas.0506580102
155. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics*. 2011 Jun;27(12):1739–40.
156. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Syst*. 2015 Dec 23;1(6):417–25.
157. Fang Z, Liu X, Peltz G. GSEAPy: a comprehensive package for performing gene set enrichment analysis in Python. *Bioinformatics*. 2023 Jan 1;39(1).
158. Feng J. Identification and validation of molecular subtypes and a 9-gene risk model for breast cancer. *Medicine (United States)*. 2023 Sep 22;102(38):E35204.
159. Shopbell PL, Britton MC, Ebert R, Barrett P, Hunter J, Miller JT, et al. matplotlib-A Portable Python Plotting Package. Vol. 347. 2005.
160. Hadaj P, Strzałka D, Nowak M, Łatka M, Dymora P. The use of PLANS and NetworkX in modeling power grid system failures. *Sci Rep*. 2022 Dec 1;12(1).
161. PyQt5 Reference Guide. PyQt Documentation v5.15.7 [Internet]. 2023 [cited 2024 Oct 15]. Available from: <https://www.riverbankcomputing.com/static/Docs/PyQt5/#>
162. Goldman M, Craft B, Hastie M, Repečka K, Kamath A, McDade F, et al. The UCSC Xena platform for public and private cancer genomics data visualization and interpretation [Internet]. 2018. Available from: <http://biorxiv.org/lookup/doi/10.1101/326470>
163. Ke X, Zhang J, Huang X, Li S, Leng M, Ye Z, et al. Construction and Analysis of the lncRNA-miRNA-mRNA Network Based on Competing Endogenous RNA in Atrial Fibrillation. *Front Cardiovasc Med*. 2022 Jan 24;9.

164. Liu W, Cai J, Tang M, Yang QJ. Bioinformatics-Based Identification of lncRNA-miRNA-mRNA Network in Dilated Cardiomyopathy and Drug Prediction. *J Nanomater*. 2021;2021.
165. Taheri M, Safarzadeh A, Hussen BM, Ghafouri-Fard S, Baniahmad A. lncRNA/miRNA/mRNA Network Introduces Novel Biomarkers in Prostate Cancer. *Cells*. 2022 Dec 1;11(23).
166. Xu T, Le TD, Liu L, Wang R, Sun B, Li J. Identifying cancer subtypes from miRNA-TF-mRNA regulatory networks and expression data. *PLoS One*. 2016 Apr 1;11(4).
167. Li WX, He K, Tang L, Dai SX, Li GH, Lv WW, et al. Comprehensive tissue-specific gene set enrichment analysis and transcription factor analysis of breast cancer by integrating 14 gene expression datasets. *Oncotarget* [Internet]. 2016 Dec 27;8:6775–86. Available from: www.impactjournals.com/oncotarget
168. Sideris N, Dama P, Bayraktar S, Stiff T, Castellano L. lncRNAs in breast cancer: a link to future approaches. *Cancer Gene Ther* [Internet]. 2022 Dec 4;29(12):1866–77. Available from: <https://www.nature.com/articles/s41417-022-00487-w>
169. Gao L, Li X, Nie X, Guo Q, Liu Q, Qi Y, et al. Construction of novel mRNA-miRNA-lncRNA regulatory networks associated with prognosis of ovarian cancer. *J Cancer*. 2020 Oct 17;11(23):7057–72.
170. Gao X, Zhao D, Han J, Zhang Z, Wang Z. Identification of microRNA-mRNA-TF regulatory networks in periodontitis by bioinformatics analysis. *BMC Oral Health*. 2022 Dec 1;22(1).
171. Guzzi PH, Di Martino MT, Tagliaferri P, Tassone P, Cannataro M. Analysis of miRNA, mRNA, and TF interactions through network-based methods. *EURASIP J Bioinform Syst Biol*. 2015 Dec 2;2015(1).
172. Hao W, Yang W, Yang Y, Cheng T, Wei T, Tang L, et al. Identification of lncRNA-miRNA-mRNA Networks in the Lenticular Nucleus Region of the Brain Contributes to Hepatolenticular Degeneration Pathogenesis and Therapy. *Mol Neurobiol*. 2024 Mar 1;61(3):1673–86.

173. Ye Y, Li SL, Wang SY. Construction and analysis of mRNA, miRNA, lncRNA, and TF regulatory networks reveal the key genes associated with prostate cancer. *PLoS One*. 2018 Aug 1;13(8).
174. Ye Z, Fang J, Wang Z, Wang L, Li B, Liu T, et al. Bioinformatics-based analysis of the lncRNA–miRNA–mRNA and TF regulatory networks reveals functional genes in esophageal squamous cell carcinoma. *Biosci Rep*. 2020 Aug 1;40(8).
175. Zhou RS, Zhang EX, Sun QF, Ye ZJ, Liu JW, Zhou DH, et al. Integrated analysis of lncRNA-miRNA-mRNA ceRNA network in squamous cell carcinoma of tongue. *BMC Cancer*. 2019 Aug 7;19(1).
176. Veneziano D, Di Bella S, Nigita G, Laganà A, Ferro A, Croce CM. Noncoding RNA: Current Deep Sequencing Data Analysis Approaches and Challenges. Vol. 37, *Human Mutation*. John Wiley and Sons Inc.; 2016. p. 1283–98.
177. Müller S, Raulefs S, Bruns P, Afonso-Grunz F, Plötner A, Thermann R, et al. Next-generation sequencing reveals novel differentially regulated mRNAs, lncRNAs, miRNAs, sdRNAs and a piRNA in pancreatic cancer. *Mol Cancer*. 2015 Apr 25;14(1).
178. Yu L, Xu Q, Yu W, Duan J, Dai G. LncRNA cancer susceptibility candidate 15 accelerates the breast cancer cells progression via miR-153–3p/KLF5 positive feedback loop. *Biochem Biophys Res Commun*. 2018 Dec 2;506(4):819–25.
179. Gu X, Chu Q, Zheng Q, Wang J, Zhu H. The dual functions of the long noncoding RNA CASC15 in malignancy. Vol. 135, *Biomedicine and Pharmacotherapy*. Elsevier Masson s.r.l.; 2021.
180. Shen Y, Xu Y, Li Y, Chen Z, Chen Z, Liao W, et al. Regulatory role of LncRNA CASC15 in malignant tumors. 2024; Available from: <https://doi.org/10.22541/au.171923176.66947256/v1>
181. Zhou W, Song F, Wu Q, Liu R, Wang L, Liu C, et al. MIR-217 inhibits triple-negative breast cancer cell growth, migration, and invasion through targeting KLF5. *PLoS One*. 2017 Apr 1;12(4).

182. Miao Y, Zheng W, Li N, Su Z, Zhao L, Zhou H, et al. MicroRNA-130b targets PTEN to mediate drug resistance and proliferation of breast cancer cells via the PI3K/Akt signaling pathway. *Sci Rep*. 2017 Feb 6;7.
183. Shui Y, Yu X, Duan R, Bao Q, Wu J, Yuan H, et al. miR-130b-3p inhibits cell invasion and migration by targeting the Notch ligand Delta-like 1 in breast carcinoma. *Gene*. 2017 Apr 20;609:80–7.

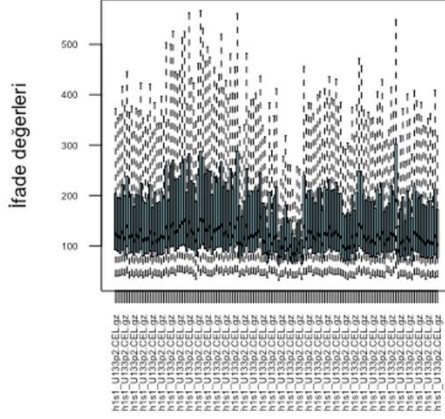


EKLER

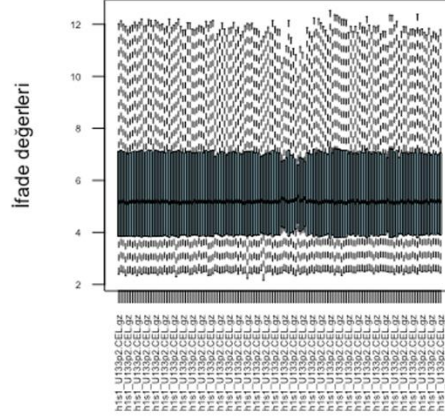
EK 1: VERİ SETLERİNE AİT TÜM NORMALİZASYON ÖNCESİ VE SONRASI BAR GRAFİKLERİ

Mikrodizin verilerinin normalizasyon öncesi (a) ve sonrası (b) grafikleri.

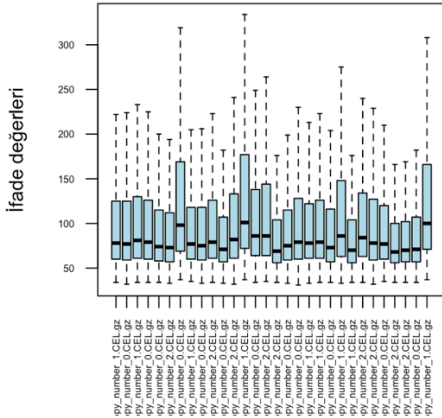
(a) GSE45827 ifade verisi grafiği



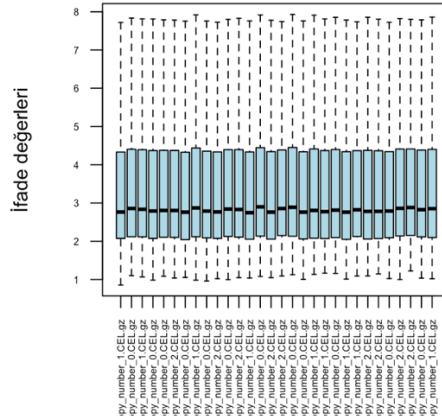
(b) GSE45827 normalize ifade verisi grafiği



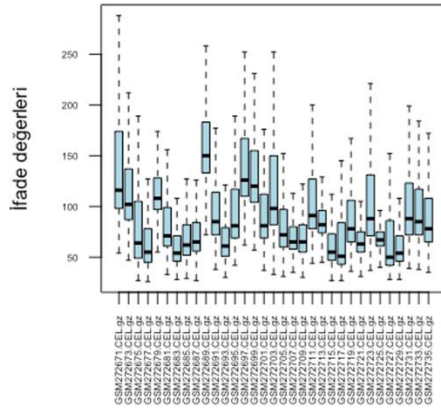
(a) GSE73540 ifade verisi grafiği



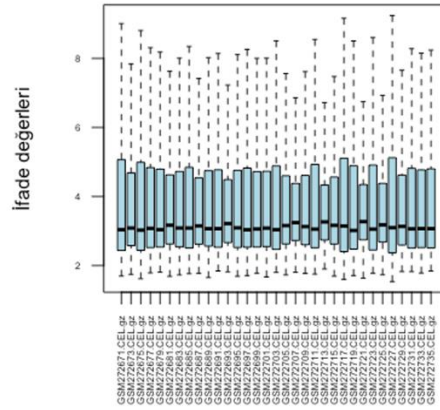
(b) GSE73540 normalize ifade verisi grafiği



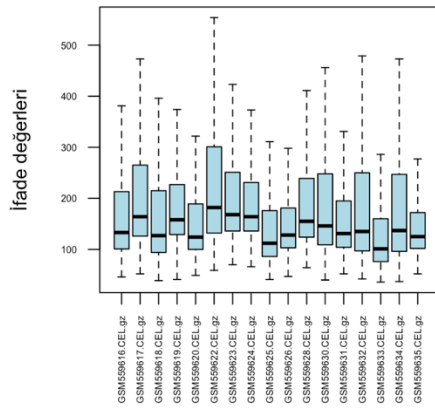
(a) **GSE10797 ifade verisi grafiği**



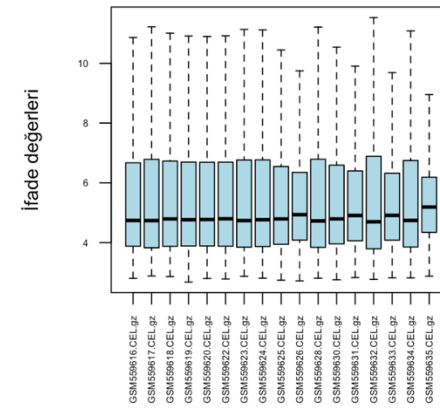
(b) **GSE10797 normalize ifade verisi grafiği**



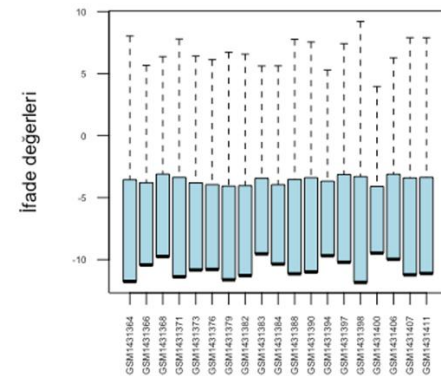
(a) **GSE22544 ifade verisi grafiği**



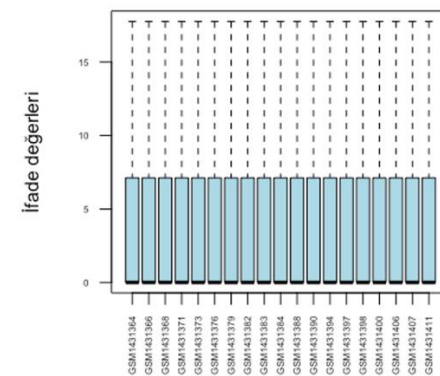
(b) **GSE22544 normalize ifade verisi grafiği**



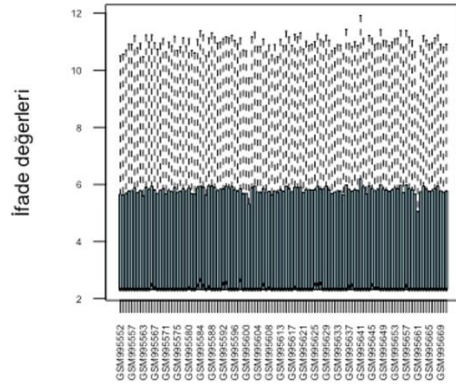
(a) **GSE59247 ifade verisi grafiği**



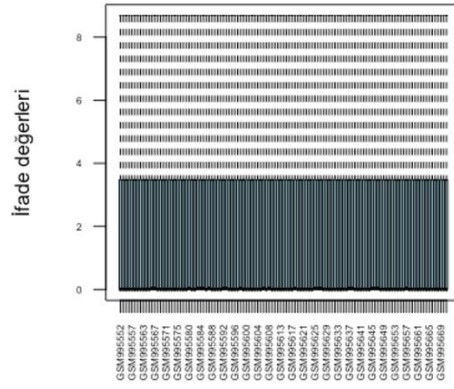
(b) **GSE59247 normalize ifade verisi grafiği**



(a) GSE40525 ifade verisi grafiği

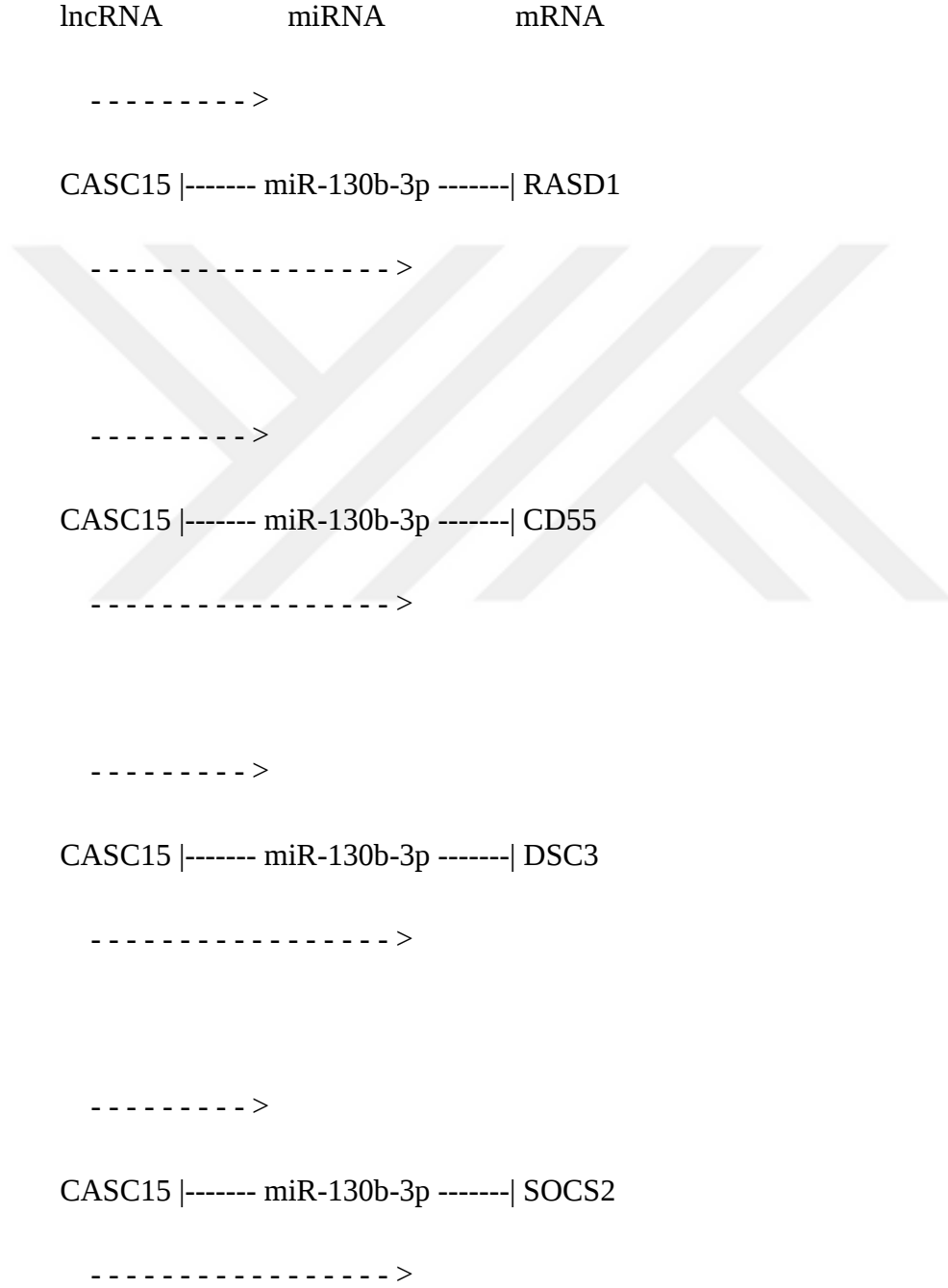


(b) GSE40525 normalize ifade verisi grafiği



EK 2: NGS VERİLERİNDEN ELDE EDİLEN MİRNA BAĞIMLI LNCRNA DÜZENLEYİCİ DEVRELER İLE TF BAĞIMLI LNCRNA DÜZENLEYİCİ DEVRELER

1. miRNA-bağımlı lncRNA Düzenleyici Devreler



----->

CASC15 |----- miR-130b-3p -----| INHBB

----->

----->

CASC15 |----- miR-130b-3p -----| SORBS2

----->

----->

CASC15 |----- miR-130b-3p -----| TFPI

----->

1. TF-bağımlı lncRNA Düzenleyici Devreler

lncRNA TF mRNA

----->

EMX2OS <----- AR -----> KLB

----->

----->

FBXL19-AS1 <----- AR -----> NPM2

----->

----->

FBXL19-AS1 <----- AR -----> NIPA1

----->

----->

FBXL19-AS1 <----- AR -----> TASOR2

----->

----->

FBXL19-AS1 <----- AR -----> PKN3

----->

----->

FBXL19-AS1 <----- AR -----> STEAP4

----->

----->

FBXL19-AS1 <----- AR -----> GJB1

----->

----->

FBXL19-AS1 <----- AR -----> SLC1A1

----->

----->

LEF1-AS1 <----- AR -----> PLLP

----->

----->

CASC15 <----- KLF5 -----> SLC7A11

----->

----->

CASC15 <----- KLF5 -----> RASD1

----->

----->

CASC15 <----- KLF5 -----> TSPAN7

----->

----->

CASC15 <----- KLF5 -----> TNFAIP3

----->

----->

CASC15 <----- KLF5 -----> RIPOR2

----->

----->

CASC15 <----- KLF5 -----> SERINC2

----->

----->

CASC15 <----- KLF5 -----> DSC3

----->

----->

CASC15 <----- KLF5 -----> PROM1

----->

----->

CASC15 <----- KLF5 -----> NPL

----->

----->

CASC15 <----- KLF5 -----> MX1

----->

----->

CASC15 <----- KLF5 -----> CD55

----->

----->

CASC15 <----- KLF5 -----> PLAC8

----->

----->

CASC15 <----- KLF5 -----> DOCK8

----->

----->

CASC15 <----- KLF5 -----> ID3

----->

----->

CASC15 <----- KLF5 -----> SORBS2

----->

----->

CASC15 <----- KLF5 -----> TMEM52

----->

----->

CASC15 <----- KLF5 -----> SOCS2

----->

----->

CASC15 <----- KLF5 -----> DHRS3

----->

----->

CASC15 <----- KLF5 -----> TBC1D10C

----->

----->

LINC00958 <----- KLF5 -----> SGCA

----->

----->

LINC00958 <----- KLF5 -----> PLIN5

----->

----->

LINC00958 <----- KLF5 -----> PPARGC1A

----->

----->

LINC00958 <----- KLF5 -----> TNFSF12

----->

----->

LINC00958 <----- KLF5 -----> SLC1A1

----->

----->

LINC00958 <----- KLF5 -----> RBL1

----->

----->

LINC00958 <----- KLF5 -----> PRPH

----->

----->

LINC00958 <----- KLF5 -----> SPSB4

----->

----->

LINC00958 <----- KLF5 -----> MFAP5

----->

----->

LINC00958 <----- KLF5 -----> NPR3

----->

----->

LINC00958 <----- KLF5 -----> SSTR3

----->

----->

LINC00958 <----- KLF5 -----> TMEM100

----->

----->

LINC00958 <----- KLF5 -----> TCN1

----->

----->

LINC00958 <----- KLF5 -----> PCK1

----->

----->

LINC00958 <----- KLF5 -----> ZDHHC11

----->

----->

LINC00958 <----- KLF5 -----> RNF128

----->

----->

LINC00958 <----- KLF5 -----> RGS1

----->

----->

LINC00958 <----- KLF5 -----> SHISAL1

----->

----->

LINC00958 <----- KLF5 -----> SERPINB9

----->

----->

FENDRR <----- E2F7 -----> OAS1

----->

----->

FENDRR <----- E2F7 -----> TMC6

----->

----->

FENDRR <----- E2F7 -----> SPRY2

----->

----->

FENDRR <----- E2F7 -----> PDCD1LG2

----->

----->

FENDRR <----- E2F7 -----> TUSC1

----->

----->

FENDRR <----- E2F7 -----> TSNAXIP1

----->

----->

FENDRR <----- E2F7 -----> TLE1

----->

----->

FENDRR <----- E2F7 -----> MDK

----->

----->

FENDRR <----- E2F7 -----> INAFM1

----->

----->

FENDRR <----- E2F7 -----> VEGFC

----->

----->

FENDRR <----- E2F7 -----> TAGLN

----->

----->

FENDRR <----- E2F7 -----> RNASE4

----->

----->

FENDRR <----- E2F7 -----> TRABD2A

----->

----->

FENDRR <----- E2F7 -----> IER2

----->

----->

FENDRR <----- E2F7 -----> RASSF2

----->

----->

FENDRR <----- E2F7 -----> GPC1

----->

----->

FENDRR <----- E2F7 -----> HEBP1

----->

----->

FENDRR <----- E2F7 -----> MTFR2

----->

----->

FENDRR <----- E2F7 -----> PTN

----->

----->

FENDRR <----- E2F7 -----> VNN2

----->

----->

FENDRR <----- E2F7 -----> FERMT1

----->

----->

FENDRR <----- E2F7 -----> PTX3

----->

----->

FENDRR <----- E2F7 -----> GALNT3

----->

----->

FENDRR <----- E2F7 -----> TMEM52B

----->

----->

FENDRR <----- E2F7 -----> RAB26

----->

----->

FENDRR <----- E2F7 -----> KRT15

----->

----->

FENDRR <----- E2F7 -----> WDFY1

----->

----->

FENDRR <----- E2F7 -----> LCAT

----->

----->

FENDRR <----- E2F7 -----> NUDCD1

----->

----->

FENDRR <----- E2F7 -----> SGTB

----->

----->

FENDRR <----- E2F7 -----> KIF2A

----->

----->

FENDRR <----- E2F7 -----> SNAP23

----->

----->

FENDRR <----- E2F7 -----> PER1

----->

----->

FENDRR <----- E2F7 -----> SLFN5

----->

----->

FENDRR <----- E2F7 -----> IGFBP5

----->

----->

FENDRR <----- E2F7 -----> SH3RF2

----->

----->

FENDRR <----- E2F7 -----> ZNFX1

----->

----->

FENDRR <----- E2F7 -----> FMN2

----->

----->

FENDRR <----- E2F7 -----> SLC25A21

----->

----->

FENDRR <----- E2F7 -----> RECQL

----->

----->

FENDRR <----- E2F7 -----> SMAP2

----->

----->

FENDRR <----- E2F7 -----> PYROXD2

----->

----->

FENDRR <----- E2F7 -----> NIPA1

----->

----->

FENDRR <----- E2F7 -----> RCAN1

----->

----->

FENDRR <----- E2F7 -----> LAP3

----->

----->

FENDRR <----- E2F7 -----> SLC22A4

----->

----->

FENDRR <----- E2F7 -----> SMTN

----->

----->

FENDRR <----- E2F7 -----> STC2

----->

----->

FENDRR <----- E2F7 -----> MCAM

----->

----->

FENDRR <----- E2F7 -----> TRIM59

----->

----->

FENDRR <----- E2F7 -----> KIF18A

----->

----->

FENDRR <----- E2F7 -----> TUBB2B

----->

----->

FENDRR <----- E2F7 -----> RAD54B

----->

----->

FENDRR <----- E2F7 -----> GULP1

----->

----->

FENDRR <----- E2F7 -----> IGF2BP3

----->

----->

FENDRR <----- E2F7 -----> GIT2

----->

----->

FENDRR <----- E2F7 -----> TTI1

----->

----->

FENDRR <----- E2F7 -----> NPL

----->

----->

FENDRR <----- E2F7 -----> MRC2

----->

----->

FENDRR <----- E2F7 -----> GPX4

----->

----->

FENDRR <----- E2F7 -----> SHF

----->

----->

FENDRR <----- E2F7 -----> PCOLCE

----->

----->

FENDRR <----- E2F7 -----> HMOX2

----->

----->

FENDRR <----- E2F7 -----> IFIH1

----->

----->

FENDRR <----- E2F7 -----> RTEL1

----->

----->

FENDRR <----- E2F7 -----> PNPLA3

----->

----->

FENDRR <----- E2F7 -----> STRIP2

----->

----->

FENDRR <----- E2F7 -----> MX1

----->

----->

FENDRR <----- E2F7 -----> PCBP4

----->

----->

FENDRR <----- E2F7 -----> SLC35G2

----->

----->

FENDRR <----- E2F7 -----> NBN

----->

----->

FENDRR <----- E2F7 -----> KNTC1

----->

----->

FENDRR <----- E2F7 -----> LY6K

----->

----->

FENDRR <----- E2F7 -----> UGCG

----->

----->

FENDRR <----- E2F7 -----> ISYNA1

----->

----->

FENDRR <----- E2F7 -----> XDH

----->

----->

FENDRR <----- E2F7 -----> LMNA

----->

----->

FENDRR <----- E2F7 -----> PLPP6

----->

----->

FENDRR <----- E2F7 -----> SPAG1

----->

----->

FENDRR <----- E2F7 -----> STK17B

----->

----->

FENDRR <----- E2F7 -----> LOXL4

----->

----->

FENDRR <----- E2F7 -----> MFAP4

----->

----->

FENDRR <----- E2F7 -----> HSPA1A

----->

----->

FENDRR <----- E2F7 -----> PLAC8

----->

----->

FENDRR <----- E2F7 -----> RMND1

----->

----->

FENDRR <----- E2F7 -----> SLC22A17

----->

----->

FENDRR <----- E2F7 -----> MEX3D

----->

----->

FENDRR <----- E2F7 -----> NAV3

----->

----->

FENDRR <----- E2F7 -----> LPXN

----->

----->

FENDRR <----- E2F7 -----> GPSM1

----->

----->

FENDRR <----- E2F7 -----> OAS3

----->

----->

FENDRR <----- E2F7 -----> UBE2T

----->

----->

FENDRR <----- E2F7 -----> OLR1

----->

----->

FENDRR <----- E2F7 -----> LZTS2

----->

----->

FENDRR <----- E2F7 -----> NEMP1

----->

----->

FENDRR <----- E2F7 -----> SLC20A1

----->

----->

FENDRR <----- E2F7 -----> KIF23

----->

----->

FENDRR <----- E2F7 -----> UBR7

----->

----->

FENDRR <----- E2F7 -----> ISG20

----->

----->

FENDRR <----- E2F7 -----> IL4I1

----->

----->

FENDRR <----- E2F7 -----> ITGAE

----->

----->

FENDRR <----- E2F7 -----> PDLIM4

----->

----->

FENDRR <----- E2F7 -----> SDC2

----->

----->

FENDRR <----- E2F7 -----> MRGPRF

----->

----->

FENDRR <----- E2F7 -----> HERC5

----->

----->

FENDRR <----- E2F7 -----> LMOD1

----->

----->

FENDRR <----- E2F7 -----> LTBP4

----->

----->

FENDRR <----- E2F7 -----> GALNT15

----->

----->

FENDRR <----- E2F7 -----> SPON2

----->

----->

FENDRR <----- E2F7 -----> PTMS

----->

----->

FENDRR <----- E2F7 -----> PHYHD1

----->

----->

FENDRR <----- E2F7 -----> TM4SF1

----->

----->

FENDRR <----- E2F7 -----> FHL2

----->

----->

FENDRR <----- E2F7 -----> FHL1

----->

----->

FENDRR <----- E2F7 -----> RNF19A

----->

----->

FENDRR <----- E2F7 -----> IFIT1

----->

----->

FENDRR <----- E2F7 -----> MET

----->

----->

FENDRR <----- E2F7 -----> PACSIN3

----->

----->

FENDRR <----- E2F7 -----> RBPMS

----->

----->

FENDRR <----- E2F7 -----> SQLE

----->

----->

FENDRR <----- E2F7 -----> FLVCR2

----->

----->

FENDRR <----- E2F7 -----> HSPB6

----->



ÖZGEÇMİŞ



TEZDEN ÇIKAN YAYINLAR

1) A Regulatory Circuits Analysis Tool, “miRCuit”, Helps Reveal Breast Cancer Pathways: Towards Systems Medicine in Oncology, OMICS: A Journal of Integrative Biology (Accepted at 31 Dec 2024.)





A Regulatory Circuits Analysis Tool, “miRCuit,” Helps Reveal Breast Cancer Pathways: Toward Systems Medicine in Oncology

Begum Karaoglu^{1,2} and Bala Gur Dedeoglu¹

Abstract

A systems medicine understanding of the regulatory molecular circuits that underpin breast cancer is essential for early cancer detection and precision/personalized medicine in clinical oncology. Transcription factors (TFs), microRNAs (miRNAs), and long non-coding RNAs (lncRNAs) control gene expression and cell biology, and by extension, serve as pillars of the regulatory circuits that determine human health and disease. We report here the development of a regulatory circuit analysis program, *miRCuit*, constructing 10 different types of regulatory elements involving messenger RNA, miRNA, lncRNA, and TFs. Using the *miRCuit*, we analyzed expression profiling data from 179 invasive ductal breast carcinoma and 51 normal tissue samples from the Gene Expression Omnibus database. We identified eight circuit types along with two special types of circuits, one of which highlighted the significant roles of lncRNA CASC15, miR-130b-3p, and TF *KLF5* in breast cancer development and progression. These findings advance our understanding of the regulatory molecules associated with breast cancer. Moreover, *miRCuit* offers a new avenue for users to construct circuits from regulatory molecules for potential applications to decipher disease pathogenesis.

Keywords: systems medicine, breast cancer, genomics regulation, integrative biology, data analysis program, network medicine

Introduction

Breast cancer is a complex, multifactorial disease with heterogeneous clinical phenotypes, and the most frequently diagnosed cancer type among women worldwide (GLOBOCAN, 2020; Polyak, 2011). Breast cancer encompasses numerous subtypes, each exhibiting distinct biological characteristics, treatment responses, and prognostic models (Denkert et al., 2018; Feng, 2023; Kim et al., 2020). Notably, invasive ductal carcinoma (IDC) is recognized as the most prevalent subtype. There is a growing need for new molecular targets that can help foster diagnostic and therapeutic innovation in breast cancer (Marrugo-Padilla et al., 2022; Xu et al., 2024; Zhao et al., 2022). In this context, understanding the mechanisms of gene regulation and identifying the molecular interactions involved in

these processes significantly contribute to the identification of new biomarkers (Lu et al., 2021).

For example, transcription factors (TFs), microRNAs (miRNAs), and long non-coding RNAs (lncRNAs) play critical roles in the regulation of gene expression and cell biology, and by extension, serve as pillars of the regulatory circuits that determine human health and disease (Cannell et al., 2008; Lambert et al., 2018; Mitsis et al., 2020; Shaath et al., 2021; Wu et al., 2014).

The regulation of gene expression occurs primarily at two levels: transcriptional and post-transcriptional (Mitsis et al., 2020). While TFs serve as the principal elements in regulating gene expression at the transcriptional level, miRNAs modulate regulatory processes at the post-transcriptional level (Cannell et al., 2008; Lambert et al., 2018). Generally, TFs enhance gene expression levels, whereas miRNAs

¹Biotechnology Institute, Ankara University, Ankara, Turkey.

²Intergen Genetics and Rare Diseases Diagnosis Center, Ankara, Turkey.

typically reduce them (Ghafari-Fard et al., 2021; Lozano-Velasco et al., 2024; Pande, 2021; Rani and Sengar, 2022). Additionally, lncRNAs significantly influence both transcriptional and post-transcriptional regulatory processes, exerting either positive or negative effects at nearly every stage of gene expression regulation (Shaath et al., 2021; Statello et al., 2021; Wu et al., 2014). Change in the expression of these regulators may alter the expression of other interacting molecules, leading to the development of various diseases, including breast cancer (Dong et al., 2018; Poursheikhani et al., 2021; Rao, 2017; Wang et al., 2017; Yang et al., 2014; Yu et al., 2017; Zhu et al., 2021).

TFs play a critical role in processes such as the proliferation, invasion, and migration of breast cancer cells. For instance, it has been demonstrated that TFs increase invasion in breast cancer cells through signaling pathways (Du et al., 2023; Willis et al., 2015; Wu et al., 2021). Moreover, miRNAs play significant roles in various cancers, with studies revealing that miRNA expression is disrupted in tumor tissues compared with normal tissues (Denli et al., 2004; He et al., 2019; Meltzer, 2005). The repressive effects of miRNAs on target genes contribute to processes such as proliferation, invasion, and metastasis in breast cancer cells (Amir et al., 2016; An et al., 2021). Furthermore, lncRNAs also play a significant regulatory role in cellular processes and can engage with TFs to exert oncogenic functions. Taken together, changes in the expression of these regulatory molecules can alter the expression of other interacting molecules, ultimately laying the foundation for the development of breast cancer. Consequently, understanding the effects of these molecular networks and circuits in breast cancer is crucial for elucidating the pathological mechanisms of the disease and developing targeted and precision/personalized medicine strategies.

Regulatory circuit analysis tools have increasingly gained attention for their potential to elucidate the complex molecular interactions underlying various diseases, including breast cancer. Existing programs, which we summarized in Table 1, are mostly designed for identifying binary relationships, including miRNA-target gene, lncRNA-miRNA, TF-target gene, and miRNA-TF interactions (Li et al., 2018; Liu et al., 2015; Nersisyan et al., 2021; da Silveira et al., 2018; Wang et al., 2022; Wang, 2024). However, these tools are often limited to either binary or triple relationships, leaving a gap in the comprehensive analysis of multi-layered regulatory networks. We report here the development of a regulatory circuit analysis program, *miRCuit*, constructing 10 different types of regulatory circuits involving messenger RNA (mRNA), miRNA, lncRNA, and TFs, and a robust mechanism for integrating quadruple circuits that helps address this knowledge gap in particular.

Materials and Methods

The present study used publicly available data and did not require informed consent and research ethics board approval. The study was conducted under the overall research ethics oversight of the authors' institutions.

We identified mRNAs, miRNAs, lncRNAs, and TFs that exhibit significant differential expression between IDC and normal or adjacent normal tissues. Additionally, we developed a circuit analysis tool called *miRCuit* that elucidates the regulatory interactions among these molecules. The steps both

for the differential expression analysis and program development are presented in Figure 1.

Construction and Analysis of Datasets

Selection of datasets

Next-generation sequencing (NGS) datasets containing invasive breast cancer (IDC) and normal/adjacent normal breast tissues were searched in the Gene Expression Omnibus (GEO) (GEO, 2024) database using the search terms “breast cancer non-coding RNA,” “breast cancer transcriptome,” and “breast cancer RNA-Seq.” Based on these keywords, five studies were identified. The datasets were selected from fresh frozen tissue samples, excluding formalin-fixed paraffin-embedded tissues and cell lines. The datasets were generated without considering differences among the molecular subtypes of the patients. Among five studies identified, three datasets, GSE40049 (Chang and Kuo, 2015), GSE39162 (Persson et al., 2011), and GSE29173 (Farazi et al., 2011), comprised only miRNA data, while the remaining two datasets, GSE71651 (Pang, 2017) and GSE110114 (Kang, 2018), included mRNA, miRNA, and lncRNA data. A total of 179 IDC and 51 normal/adjacent normal samples were included. Details of the datasets are presented in Supplementary Data S1.

RNA-Seq data analysis

Sequence read archive fastq files for the RNA-Seq datasets were downloaded and processed using the fasterq-dump tool (version 2.11.3) (<https://hpc.nih.gov/apps/sratoolkit.html>) (Maurya et al., 2022). Quality control of the data was performed using FastQC (version 0.11.9) (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) (Leggett et al., 2013). Given that the Phred score of the data was 30 or higher and that adapter contamination was absent, no trimming was applied. High-quality paired reads were mapped to the human reference genome (NCBI GRCh38.p14) using HISAT2 (version 2.2.1) (Kim et al., 2015). The Sequence Alignment Map (SAM) files obtained were converted to Binary Alignment Map (BAM) format, followed by sorting and indexing using SAMtools (version 1.13) (<http://www.htslib.org>) (Li et al., 2009). Read counts were obtained using featureCounts (2.0.3) (GENCODE comprehensive gene annotation, version 46) (Liao et al., 2014). Moreover, for RNA-Seq datasets containing only miRNA data, feature counts matrix files were retrieved from the GEO database. DESeq2 method was used to identify differentially expressed genes (DEGs) using the PyDESeq2 (version 0.4.11) (Muzellec et al., 2023) package in Python (version 3.12.0). Ensembl IDs were converted to gene symbols using the sanbomics tool (version 0.1.0) (Sanbomics, 2024) in Python (version 3.12.0).

Preparation of the data for circuit analysis program

The results obtained from the analysis involved mRNAs, miRNAs, lncRNAs, and TFs data collectively. To prepare the data for the circuit analysis program, the four regulatory molecules were separated into distinct lists. Subsequently, the lists of the four types of molecules from all studies were merged to create a single comprehensive list that included differentially expressed mRNAs, miRNAs, lncRNAs, and TFs. These lists

TABLE 1. COMPARISON OF *miRCUIT* AND DIFFERENT INTERACTION PROGRAMS

Program name	Input type	Integrated databases	Interaction type	Disease type	Output type	Programming language	User interface	Accession
<i>GDCRNA Tools</i> (Li et al., 2018)	Downloaded data from projects like TCGA or TARGET	4 different databases	lncRNA-mRNA related competing endogenous RNAs regulatory network	Cancer	DEG analysis, functional enrichment analysis	R/Bioconductor package	No/R interface	GitHub
<i>RegNetwork</i> (Liu et al., 2015)	TF, mRNA, and mRNA symbols	25 different databases	TF-miRNA-target gene	General	Regulatory networks	Java, JavaScript, and Python	Yes	Unavailable
<i>LncACTdb</i> (Wang et al., 2022)	mRNA, miRNA, and lncRNA expression data	9 different databases	lncRNA-mRNA-miRNA	Different diseases	Gene regulation networks	MySQL, Java, and R	No	Web-based platform
<i>miRGTF-net</i> (Nersisyan et al., 2021)	miRNA and mRNA expression profile	4 different databases	miRNA-gene-TF	Breast cancer	Regulatory networks	Python	No	Python interface
<i>miRmapper</i> (da Silva et al., 2018)	miRNA-mRNA interaction list and total DEG list	3 different databases	TF-miRNA-target gene	General	Regulatory networks	R	No	GitHub
<i>TFTF</i> (Wang, 2024)	TF symbol	7 different databases	TF-target gene	General	TF-target regulation network analysis	R	Yes	GitHub
<i>miRCuit</i>	mRNA, miRNA, lncRNA, TF DEG list	11 different databases	mRNA-miRNA-lncRNA-TF quadruple circuits	General	Binary, ternary and quadruple interactions, regulatory networks, and enrichment analysis	Python	Yes	GitHub

DEG, differentially expressed genes; lncRNAs, long non-coding RNAs; mRNA, messenger RNA; miRNA, microRNAs; TCGA, The Cancer Genome Atlas; TF, transcription factors.

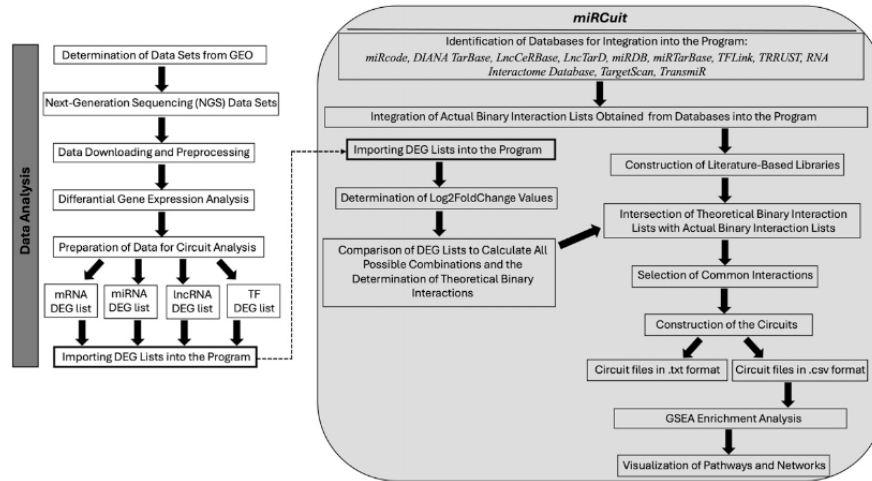


FIG. 1. Workflow of the study.

were named as mRNA DEG list, miRNA DEG list, lncRNA DEG list, and TF DEG list throughout the text. Additionally, miRNA names in different formats were standardized by removing the “hsa” prefix and replacing it with the “miR” prefix.

The Development of the Circuit Analysis Program

Determination of the databases and construction of the literature-based libraries

Eleven different databases demonstrating molecular interactions were identified to explore the regulatory circuits. These databases include miRcode (Jeggari et al., 2012), DIANA TarBase (Vlachos et al., 2015), LncCeRBase (Pian et al., 2018), LncTarD (Zhao et al., 2023), miRDB (Chen and Wang, 2020), miRTarBase (Hsu et al., 2011; Huang et al., 2022), TFLink (Liska et al., 2022), TRRUST (Han et al., 2018), RNA Interactome Database (Kang et al., 2022), TargetScan (Agarwal et al., 2015), and TransmiR (Tong et al., 2019), and the interactions between molecules are presented in Supplementary Data S2.

The literature-based libraries were constructed by importing binary molecular interactions from the databases into the circuit analysis program and lists with different formats were standardized to a uniform format using Python (version 3.12.0).

Integration of DEG list files into the program and identification of binary interactions

DEG lists of mRNAs, miRNAs, lncRNAs, and TFs were integrated into the circuit analysis program. Log2Fold-Change values were determined to be 1.5-fold between IDC and normal/adjacent normal tissues for each molecule type. Subsequently, all possible combinations that could reveal potential interactions between the molecules were determined. As a result, theoretical binary interactions were

identified, including TF-mRNA, TF-miRNA, TF-lncRNA, miRNA-mRNA, miRNA-TF, miRNA-lncRNA, lncRNA-mRNA, lncRNA-miRNA, and lncRNA-TF (Fig. 2A). The actual binary interaction lists obtained from the literature-based library (Fig. 2B) were intersected with the theoretical binary interaction lists to identify common interactions. Finally, all interaction lists were prepared for the circuit analysis (Fig. 2C).

Construction of regulatory circuits

The regulatory circuits were based on the principle of merging two interaction lists by identifying a common molecule and determining the direction of regulation. The proposed method is based on the following principles: (1) miRNAs repress their target genes, (2) TFs act to activate their target genes, and (3) lncRNAs can exert either positive or negative effects on their targets, as they are involved in both transcriptional and post-transcriptional regulation (Bartel, 2004; Hauptman and Glavač, 2013; Klug et al., 2019; Liu et al., 2014; Maston et al., 2006; Mercer et al., 2009).

Consequently, eight distinct types of regulatory circuits were constructed using the circuit analysis tool (Fig. 3A). In addition, two special lncRNA-containing circuit types were also generated: the miRNA-dependent lncRNA regulatory circuit and TF-dependent lncRNA regulatory circuit, both were created using the circuit analysis tool (Fig. 3B). All circuits were initially obtained in .txt format. However, for further analysis, the circuit files were converted to .csv format and saved as tables in .csv format within the circuit analysis program.

Integration of gene set enrichment analysis

Gene set enrichment analysis (GSEA) was integrated into the circuit analysis program. GSEA was performed using

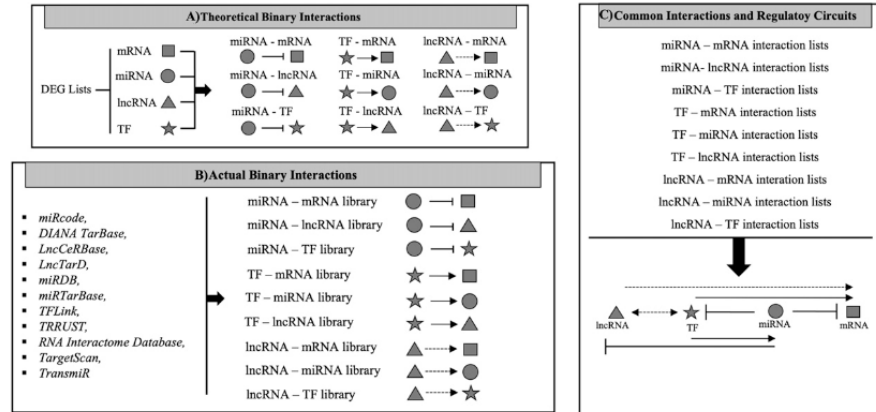


FIG. 2. Theoretical and actual binary molecular interactions and the formation of regulatory circuits. (----->) Positive regulation, (-----) negative regulation, (----->) the direction of the interaction is known, but the regulation can be either positive or negative, (<----->) Both molecules regulate each other, but the regulation can be either positive or negative.

the pre-rank module (Yoon et al., 2016) from the *gseapy* library (version 1.1.3) in Python (version 3.12.0). For the analysis, the column containing mRNA data extracted from the circuits .csv files was used. The parameters for the pre-rank module were set as follows: MSigDB_Hallmark_2020 for the gene set, 1000 for the number of permutations, 2 for the minimum gene number, 5000 for the maximum gene

number. The genes were selected based on the condition of complying with a false discovery rate (FDR q-val) <0.25 among the obtained results (Subramanian et al., 2005). The results were visualized using the *gseapy* (version 1.1.3) (Fang et al., 2023), *matplotlib* (version 3.9.1) (Shoppell et al., 2005), and *networkx* (version 3.3) (Hadaj et al., 2022) libraries in Python (version 3.12.0).

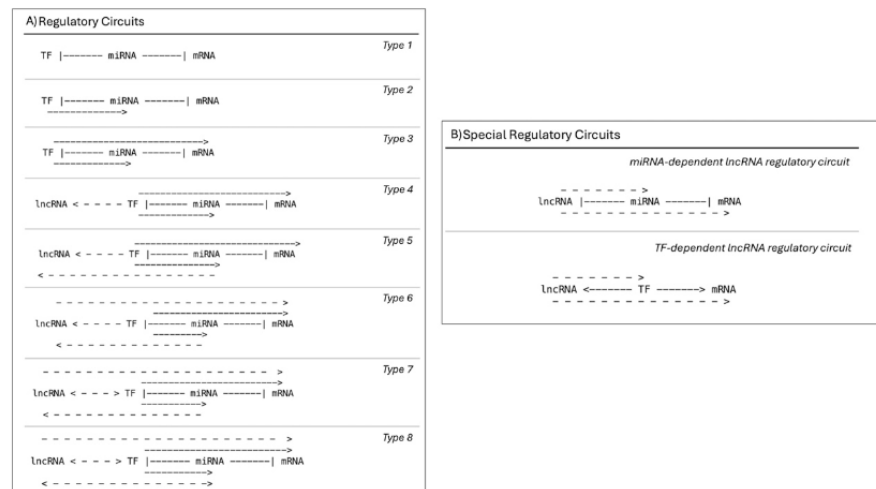


FIG. 3. Eight different types of regulatory circuits and special regulatory circuits (miRNA-dependent lncRNA regulatory circuit and TF-dependent lncRNA regulatory circuit). (----->) Positive regulation, (-----) negative regulation, (----->) the direction of the interaction is known, but the regulation can be either positive or negative, (<----->) both molecules regulate each other, but the regulation can be either positive or negative.

Development of a User Interface for the Circuit Analysis Program

A user interface for the circuit analysis program was developed using the *PyQt5* library (version 5.15.11) (*PyQt5 Reference Guide*, 2023) in Python (version 3.12.0), as shown in Supplementary Data S3. The user interface operates by allowing users to upload the DEG lists of mRNAs, miRNAs, lncRNAs, and TFs, including gene symbols and expression values, in .csv formats. Following the file uploads, threshold values are set for the uploaded data, corresponding to the two investigated conditions (e.g., IDC/normal tissue) and start the analysis. As a result of the analysis, the program constructs circuits in eight distinct types along with two special types and saves the resulting files in both .txt and .csv formats. In addition, the option to conduct GSEA is available to users. Utilizing this option, users can enrich their selected gene sets with the Molecular Signatures Database (MSigDB) (Liberzon et al., 2015). The results and plots obtained from the GSEA analysis are documented under a distinct section.

The name of circuit analysis program is derived from combination of “miRNA” and “Circuits,” resulting in “*miRCuit*.” The developed local program is uploaded to the GitHub platform. The GitHub link for the program is <https://github.com/miRCuit/miRCuit>, where installation documentation and a detailed manual are provided.

The Cancer Genome Atlas Data Analysis

To validate the interactions among the circuit members identified by *miRCuit*, The Cancer Genome Atlas (TCGA) breast cancer data analysis was performed by using Xena Browser (Goldman et al., 2018).

Results

Differential expression gene analysis of NGS datasets

We identified 156 statistically significant miRNAs in GSE29173, 181 in GSE39162, and 232 in GSE40049, which contains only miRNA data. The results of DEG lists obtained from other RNA-Seq analyses were as follows: 885 mRNAs, 9 miRNAs, 233 lncRNAs, and 59 TFs were obtained from GSE71651, while GSE110114 produced 7783 mRNAs, 21 miRNAs, 1242 lncRNAs, and 718 TFs. A summary table is presented in Supplementary Data S4. The lists of regulatory molecules obtained from all datasets were combined, revealing a total of unique 8171 mRNAs, 354 miRNAs, 1389 lncRNAs, and 740 TFs.

Designing regulatory circuits with “*miRCuit*”

DEG list files, including gene symbols and Log2Fold-Change values, were uploaded to the circuit analysis program *miRCuit*. A 1.Fivefold threshold option was selected for all files and circuit analysis was conducted resulting in the identification of seven distinct type regulatory circuits and two special-type circuits. These circuits were named according to the types defined in Figure 3. The total number of regulatory circuits for each type is as follows: *Type 1* comprises 202,635 circuits, *Type 2* comprises 54,914 circuits, *Type 3* comprises 39,583 circuits, *Type 4* comprises 106,894 circuits, *Type 5* comprises 4443 circuits, and *Types 6* and *7* each comprise 3 circuits, while no circuits were generated for *Type 8*.

GSEA of identified regulatory circuits

Among all regulatory circuit types, *Type 4* was selected for GSEA analysis, due to its higher number of circuits (106,894) with the inclusion of all regulatory elements. The GSEA analysis results using the mRNA list from circuits *Type 4* show that the genes were predominantly enriched in “G2-M Checkpoint,” “E2F Targets,” “Mitotic Spindle,” “Interferon Gamma Response,” and “Interferon Alpha Response” pathways (FDR < 0.25), according to MSigDB. All enriched pathways are listed in Supplementary Data S5.

The dot plot and network visualization of the enriched pathways resulting from the analysis are presented in Supplementary Data S6.

We aimed to decipher the roles of lncRNAs in breast cancer in this study. Hence, we designed *miRCuit* to generate special lncRNA-based circuits. As a result, a total of 7 circuits were created for *miRNA-dependent lncRNA regulatory circuits*, while a total of 169 circuits were created for TF-dependent lncRNA regulatory circuits (Supplementary Data S7). Among these specialized regulatory circuits, CASC15 emerged as a prominent lncRNA shared by two distinct circuit types. miR-130b-3p and *KLF5* were identified as its key regulatory partners (Fig. 4).

TCGA Data Analysis

For the verification of relations between members of circuits obtained from *miRCuit* with an independent data set, we analyzed the expression profiles of CASC15, miR130b-3p, *KLF5*, and their targets provided in Figure 4 by using TCGA breast cancer data set (GDC BRCA) via UCSC (University of California, Santa Cruz) Xena Browser. miR-130b-3p was found to be significantly upregulated in breast tumor tissues compared with normal tissues while the other genes and lncRNA CASC15 were significantly downregulated in tumor samples compared with normal samples (Supplementary Data S8).

Discussion

Breast cancer is a major planetary health burden. A systems medicine approach that integrates the role of regulatory molecules and networks in cancer pathogenesis is essential for precision/personalized medicine in the oncology clinic. In the present study, we report, first, the development of *miRCuit*, and second, our findings from expression profiling data from 179 invasive ductal breast carcinoma (IDC) and 51 normal tissue samples from the GEO database.

Currently, numerous web tools have been designed to identify binary relationships, such as miRNA-target gene, lncRNA-miRNA, TF-target gene, and miRNA-TF interactions (Jung et al., 2015; Nersisyan et al., 2021; da Silveira et al., 2018; Wang, 2024; Yang et al., 2023). Moreover, research on regulatory circuits involving triple relationships primarily focuses on constructing miRNA or lncRNA-centric circuits by using currently available online tools or programs (Ke et al., 2022; Liu et al., 2021; Taheri et al., 2022; Xu et al., 2016). Considering the regulatory roles of these molecules within cells, such circuits may prove inadequate in explaining biological processes, systemically (Fabian et al., 2010; Mendell and Olson, 2012; Yang et al.,

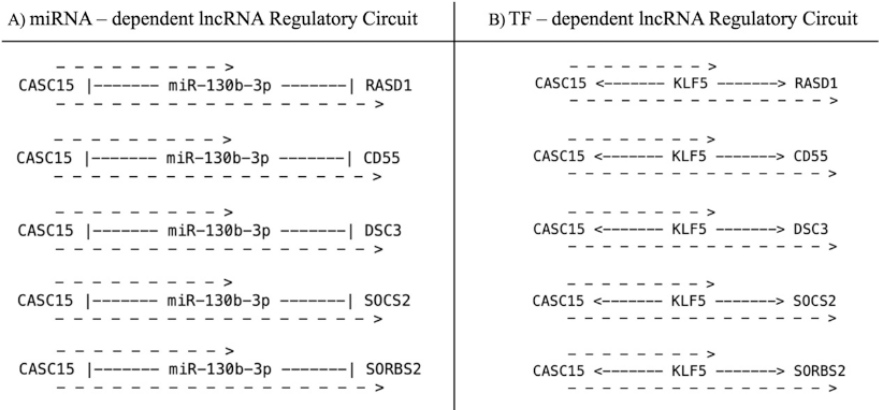


FIG. 4. miRCuit-generated CASC15-dependent special circuits.

2014; Yong and Dutta, 2009). Recent studies emphasize the importance of interactions among mRNA, miRNA, lncRNA, and TF molecules in understanding various diseases, including breast cancer and in developing new therapeutic approaches (An et al., 2021; He et al., 2019; Li et al., 2017; Sideris et al., 2022).

However, there is currently no active program that constructs a comprehensive regulatory interaction encompassing these interactions (Gao et al., 2020, 2022; Guzzi et al., 2015; Hao et al., 2024; Wan et al., 2022; Wang et al., 2021; Wang et al., 2022; Ye et al., 2018, 2020; Zhao et al., 2021; Zhou et al., 2019). Other than the current programs mentioned in Table 1 (Li et al., 2018; Liu et al., 2015; Nersisyan et al., 2021; da Silveira et al., 2018; Wang et al., 2022; Wang, 2024), our circuit analysis program, *miRCuit*, presents this gap by providing a comprehensive framework for analyzing these regulatory networks, specifically presenting a holistic approach to mRNA-miRNA-lncRNA-TF quadruple circuits. However, the program is not web-based and requires users to access it via GitHub, necessitating specific installations and downloads before reaching the interface. This could be perceived as a limitation of the study. Nonetheless, the comprehensive instructions available on GitHub greatly enhance the usability and accessibility of *miRCuit* for users. On the other hand, *miRCuit* is compatible with any expression data, as it utilizes DEGs, miRNAs, lncRNAs, and TFs as input. Therefore, expression data obtained from NGS or microarray platforms across various diseases can be analyzed, and disease-related circuits can be identified.

miRCuit has strategies that define all binary interactions to construct circuits, thereby presenting all relationships to users as intermediate outputs, thus providing them with different circuit types ranging from Type 1 to Type 8, which is presented in Figure 3A. Additionally, *miRCuit* offers users two special regulatory circuits: miRNA-dependent lncRNA regulatory circuits and TF-dependent lncRNA regulatory circuits (Fig. 3B). In this process, the data obtained from 11 distinct databases play an essential role in constructing the circuits.

The obtained circuits are presented in a linear format, providing convenience for both usage and analysis. In addition to visual files provided in linear format, there are also table format files among the outputs that users can utilize for advanced analysis. Additionally, *miRCuit* sets itself apart from other tools by allowing users to choose threshold values for each expression file providing a section for researchers to indicate how much expression variation they want to consider in their analyses between groups of samples such as disease/control or treatment/non-treatment (Supplementary Data S3).

Another potentially valuable application of the program is its ability to perform GSEA using gene lists obtained from the analysis, allowing users to list enriched genes and create graphs and networks based on the most enriched pathways.

On the other hand, the need of gene expression data is the main limitation for the program. Additionally, the requirement for a list containing a sufficient number of genes suitable for performing GSEA constitutes another limitation of *miRCuit* (Fang et al., 2023).

We conducted GSEA exclusively on Type 4 regulatory circuits because of the limitations associated with GSEA mentioned above. Furthermore, these circuits included the maximum number of interactions with four regulatory elements (mRNA, miRNA, lncRNA, and TF) simultaneously, which is crucial for capturing the complexity of regulatory interactions. When we analyze the pathways identified through GSEA, it becomes evident that they are significantly associated with cancer-related pathways like “G2-M Checkpoint,” “E2F Targets,” and mTORC1 Signaling” pathways, which underscores the reliability of the tool we developed.

In this study, we considered the expression of 179 IDC samples and 51 normal/adjacent normal tissues. Based on the results obtained, we focused on special-type circuits and emphasized the role of lncRNAs in cancer development, driven by the increasing interest in their regulatory functions. We identified 7 lncRNA-miRNA-mRNA circuits and 169 lncRNA-TF-mRNA circuits. In these circuits, concordant

with the literature, the lncRNA CASC15, miR-130b-3p, and TF *KLF5* emerged as the most prominent molecules in breast cancer development. The TCGA data analysis that we performed validated the inhibitory effect of miR-130b-3p on its targets together with the positive relation between *KLF5* and CASC15. The findings are consistent with the identified regulatory circuits, highlighting the importance of understanding lncRNA-miRNA-TF-mRNA interactions in the pathogenesis of breast cancer.

LncRNA CASC15 is known to function mostly as an oncogene in various cancers, including breast cancer, and shows high-expression levels along with TF *KLF5* in breast cancer tissues and cells (Gu et al., 2021; Shen et al., 2024; Shi et al., 2019; Yu et al., 2018). Moreover, it has been demonstrated that miR-130b-3p plays an important role in breast cancer cell invasion and migration. Therefore, it is considered that the miR-130b-3p may be a potential biomarker for breast cancer metastasis (Shui et al., 2017). Hence it could be valuable to reveal the biomarker potential of the circuit, CASC15, *KLF5*, and miR-130b-3p in breast cancer. Additionally, the regulatory circuits provided by *miRCuit* that are concordant with the literature supported the robustness and reliability of our program. In this context, the other circuits constructed by *miRCuit* are worth to investigate for their potential biomarker roles.

Conclusions

We identified eight circuit types along with two special types of circuits one of which highlighted the significant roles of lncRNA CASC15, miR-130b-3p, and TF *KLF5* in breast cancer development and progression. These findings advance our understanding of the regulatory molecules associated with breast cancer. Moreover, *miRCuit* offers a new avenue for users to construct circuits from regulatory molecules for potential applications to decipher disease pathogenesis.

Acknowledgments

The authors thank Efe Dalı for technical support in data acquisition and acknowledge Prof. Dr. Hasan Oğul for his valuable assistance in supporting the Python learning process.

Authors' Contributions

B.K.: investigation, methodology, data curation, formal analysis, visualization and writing—original draft. B.G.D.: conceptualization, methodology, data curation, formal analysis, supervision, writing—review & editing.

Author Disclosure Statement

The authors declare they have no conflicting financial interests.

Funding Information

The authors received no funding for this article.

Supplementary Material

Supplementary Data S1
Supplementary Data S2
Supplementary Data S3

Supplementary Data S4
Supplementary Data S5
Supplementary Data S6
Supplementary Data S7
Supplementary Data S8

References

- Agarwal V, Bell GW, Nam JW, et al. Predicting effective MicroRNA target sites in mammalian mRNAs. *Elife* 2015; 4(e05005):1–38; doi: 10.7554/eLife.05005
- Amir S, Simion C, Umeh-Garcia M, et al. Regulation of the T-Box transcription factor tbx3 by the tumour suppressor MicroRNA-206 in breast cancer. *Br J Cancer* 2016;114(10): 1125–1134; doi: 10.1038/bjc.2016.73
- An SQ, He K, Liu F, et al. Low MicroRNA150 expression is associated with activated carcinogenic pathways and a poor prognosis in patients with breast cancer. *Oncol Rep* 2021; 45(3):1235–1248; doi: 10.3892/or.2021.7945
- Bartel DP. MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell* 2004;116(2):281–297; doi: 10.1016/S0092-8674(04)00045-5
- Cannell IG, Kong YW, Bushell M. How do MicroRNAs regulate gene expression? *Biochem Soc Trans* 2008;36(Pt 6): 1224–1231; doi: 10.1042/BST0361224
- Chang K, Kuo W. Deregulated MicroRNAs in triple-negative breast cancer revealed by deep sequencing. 2015. Available from: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE40049> [Last accessed: October 14, 2024].
- Chen Y, Wang X. MiRDB: An online database for prediction of functional MicroRNA targets. *Nucleic Acids Res* 2020; 48(D1):D127–D131; doi: 10.1093/nar/gkz757
- da Silveira WA, Renaud L, Simpson J, et al. MiRMapper: A tool for interpretation of MiRNA–mRNA interaction networks. *Genes (Basel)* 2018;9(9); doi: 10.3390/genes9090458
- Denkert C, von Minckwitz G, Darb-Esfahani S, et al. Tumour-Infiltrating Lymphocytes and prognosis in different subtypes of breast cancer: A pooled analysis of 3771 patients treated with neoadjuvant therapy. *Lancet Oncol* 2018;19(1):40–50; doi: 10.1016/S1470-2045(17)30904-X
- Denli AM, Tops BBJ, Plasterk RHA, et al. Processing of primary MicroRNAs by the microprocessor complex. *Nature* 2004;432(7014):231–235.
- Dong H, Wang W, Mo S, et al. SP1-Induced LncRNA AGAP2-AS1 expression promotes chemoresistance of breast cancer by epigenetic regulation of MyD88. *J Exp Clin Cancer Res* 2018;37(1):202–215; doi: 10.1186/s13046-018-0875-3
- Du L, Li J, Tian Y, et al. Pleckstrin-2-Promoted PPM1B degradation plays an important role in transforming growth factor- β -induced breast cancer cell invasion and metastasis. *Cancer Sci* 2023;114(6):2429–2444; doi: 10.1111/cas.15791
- Fabian MR, Sonenberg N, Filipowicz W. Regulation of mRNA translation and stability by MicroRNAs. *Annu Rev Biochem* 2010;79:351–379; doi: 10.1146/annurev-biochem-060308-103103
- Fang Z, Liu X, Peltz G. GSEAPy: A comprehensive package for performing gene set enrichment analysis in python. *Bioinformatics* 2023;39(1); doi: 10.1093/bioinformatics/btac757
- Farazi TA, Horling HM, ten Hoeve J, et al. MicroRNA sequence and expression analysis in breast tumors by deep sequencing [MiRNA Sequence Data]. 2011. Available from: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi> [Last accessed: October 14, 2024].

- Feng J. Identification and validation of molecular subtypes and a 9. *Medicine* (Baltimore) 2023;102(38):E35204; doi: 10.1097/MD.00000000000035204
- Gao L, Li X, Nie X, et al. Construction of novel mRNA-MiRNA-LncRNA regulatory networks associated with prognosis of ovarian cancer. *J Cancer* 2020;11(23):7057–7072; doi: 10.7150/JCA.49557
- Gao X, Zhao D, Han J, et al. Identification of MicroRNA-MRNA-TF regulatory networks in periodontitis by bioinformatics analysis. *BMC Oral Health* 2022;22(1):118; doi: 10.1186/s12903-022-02150-0
- GEO. Gene Expression Omnibus. 2024. Available from: www.ncbi.nlm.nih.gov/geo/ [Last accessed: October 14, 2024].
- Ghafouri-Fard S, Abak A, Fattahi F, et al. The interaction between MiRNAs/LncRNAs and nuclear Factor-KB (NF-KB) in human disorders. *Biomed Pharmacother* 2021;138:111519; doi: 10.1016/j.biopha.2021.111519
- GLOBOCAN. International Agency for Research on Cancer, Source. 2020. Available from: <https://geo.iarc.fr/today/data/factsheets/cancers/> [Last accessed: October 14, 2024].
- Goldman M, Craft B, Hastie M, et al. The UCSC Xena platform for public and private cancer genomics data visualization and interpretation. *bioRxiv* 2018; doi: 10.1101/326470
- Gu X, Chu Q, Zheng Q, et al. The dual functions of the long noncoding RNA CASC15 in malignancy. *Biomed Pharmacother* 2021;135:111212–113322; doi: 10.1016/j.biopha.2020.111212
- Guzzi PH, Di Martino MT, Tagliaferri P, et al. Analysis of MiRNA, mRNA, and TF interactions through network-based methods. *EURASIP J Bioinform Syst Biol* 2015;2015(1):4; doi: 10.1186/s13637-015-0023-8
- Hadaj P, Strzalka D, Nowak M, et al. The use of plans and networks in modeling power grid system failures. *Sci Rep* 2022; 12(1):17445; doi: 10.1038/s41598-022-22268-z
- Han H, Cho JW, Lee S, et al. TRRUST v2: An expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic Acids Res* 2018;46(D1):D380–D386; doi: 10.1093/nar/gkx1013
- Hao W, Yang W, Yang Y, et al. Identification of LncRNA-MiRNA-MRNA networks in the lenticular nucleus region of the brain contributes to hepatolenticular degeneration pathogenesis and therapy. *Mol Neurobiol* 2024;61(3):1673–1686; doi: 10.1007/s12035-023-03631-1
- Hauptman N, Glavač D. Long non-coding RNA in cancer. *Int J Mol Sci* 2013;14(3):4655–4669; doi: 10.3390/ijms14034655
- He K, Li WX, Guan D, et al. Regulatory network reconstruction of five essential MicroRNAs for survival analysis in breast cancer by integrating MiRNA and mRNA expression datasets. *Funct Integr Genomics* 2019;19(4):645–658; doi: 10.1007/s10142-019-00670-7
- Hsu SD, Lin FM, Wu WY, et al. MiRTarBase: A database curates experimentally validated MicroRNA-Target interactions. *Nucleic Acids Res* 2011;39(Database issue):D163–D169; doi: 10.1093/nar/gkq1107
- Huang HY, Lin YCD, Cui S, et al. MiRTarBase update 2022: An informative resource for experimentally validated MiRNA-Target interactions. *Nucleic Acids Res* 2022; 50(D1):D222–D230; doi: 10.1093/nar/gkab1079
- Jeggari A, Marks DS, Larsson E. MiRcode: A map of putative microRNA target sites in the long non-coding transcriptome. *Bioinformatics* 2012;28(15):2062–2063; doi: 10.1093/bioinformatics/bts344
- Jung D, Kim B, Freishtat RJ, et al. MiRTarVis: An interactive visual analysis tool for MicroRNA-MRNA expression profile data. *BMC Proc* 2015;9([Suppl 6] Proceedings of the 5th Symposium on Biological Data):S2; doi: 10.1186/1753-6561-9-S6-S2
- Kang J, Tang Q, He J, et al. RNAInter v4.0: RNA interactome repository with redefined confidence scoring system and improved accessibility. *Nucleic Acids Res* 2022;50(D1): D326–D332; doi: 10.1093/nar/gkab997
- Kang K. A practical evaluation of alignment algorithms for RNA Variant calling analysis. 2018. Available from: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi> [Last accessed: October 14, 2024].
- Ke X, Zhang J, Huang X, et al. Construction and analysis of the LncRNA-MiRNA-MRNA network based on competing endogenous RNA in atrial fibrillation. *Front Cardiovasc Med* 2022;9:791156; doi: 10.3389/fcvm.2022.791156
- Kim D, Langmead B, Salzberg SL. HISAT: A fast spliced aligner with low memory requirements. *Nat Methods* 2015; 12(4):357–360; doi: 10.1038/nmeth.3317
- Kim J, Yu D, Kwon Y, et al. Genomic characteristics of triple-negative breast cancer nominate molecular subtypes that predict chemotherapy response. *Mol Cancer Res* 2020;18(2): 253–263; doi: 10.1158/1541-7786.MCR-19-0453
- Klug WS, Cummings MR, Spencer CA, et al. *Concepts of Genetics*. 12th ed. Pearson; 2019.
- Lambert SA, Jolma A, Campitelli LF, et al. The human transcription factors. *Cell* 2018;172(4):650–665; doi: 10.1016/j.cell.2018.01.029
- Leggett RM, Ramirez-Gonzalez RH, Clavijo BJ, et al. Sequencing quality assessment tools to enable data-driven informatics for high throughput genomics. *Front Genet* 2013;4:288–285; doi: 10.3389/fgenet.2013.00288
- Li H, Handsaker B, Wysoker A, et al; 1000 Genome Project Data Processing Subgroup. The sequence alignment/map format and SAMtools. *Bioinformatics* 2009;25(16):2078–2079; doi: 10.1093/bioinformatics/btp352
- Li R, Qu H, Wang S, et al. GDCRNATools: An R/Bioconductor package for integrative analysis of LncRNA, MiRNA and mRNA Data in GDC. *Bioinformatics* 2018;34(14):2515–2517; doi: 10.1093/bioinformatics/bty124
- Li W-X, He K, Tang L, et al. Comprehensive tissue-specific gene set enrichment analysis and transcription factor analysis of breast cancer by integrating 14 gene expression datasets. *Oncotarget* 2017;8(4):6775–6786; doi: 10.18632/oncotarget.14286
- Liao Y, Smyth GK, Shi W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 2014;30(7):923–930; doi: 10.1093/bioinformatics/btt656
- Liberzon A, Birger C, Thorvaldsdóttir H, et al. The molecular signatures database hallmark gene set collection. *Cell Syst* 2015;1(6):417–425; doi: 10.1016/j.cels.2015.12.004
- Liska O, Bohár B, Hidas A, et al. TFLink: An integrated gateway to access transcription factor-target gene interactions for multiple species. *Database* 2022;2022; doi: 10.1093/database/baac083
- Liu B, Li J, Cairns MJ. Identifying MiRNAs, targets and functions. *Brief Bioinform* 2014;15(1):1–19; doi: 10.1093/bib/bbs075
- Liu W, Cai J, Tang M, et al. Bioinformatics-Based identification of LncRNA-MiRNA-MRNA network in dilated

- cardiomyopathy and drug prediction. *J Nanomater* 2021; 2021:1–10; doi: 10.1155/2021/5566316
- Liu Z-P, Wu C, Miao H, et al. RegNetwork: An integrated database of transcriptional and post-transcriptional regulatory networks in human and mouse. *Database* 2015;2015:1–12; doi: 10.1093/database/bav095
- Lozano-Velasco E, Garcia-Padilla C, Carmona-Garcia M, et al. MEF2C directly interacts with pre-MiRNAs and Distinct RNP to post-transcriptionally regulate MiR-23a-MiR-27a-MiR-24-2 MicroRNA cluster member expression. *Noncoding RNA* 2024;10(3); doi: 10.3390/ncrna10030032
- Lu C, Wei D, Zhang Y, et al. Long non-coding RNAs as potential diagnostic and prognostic biomarkers in breast cancer: Progress and prospects. *Front Oncol* 2021;11(August): 710538–710517; doi: 10.3389/fonc.2021.710538
- Marrugo-Padilla A, Márquez-Lázaro J, Álviz-Amador A. Identification of prognostic biomarkers of invasive ductal carcinoma by an integrated bioinformatics approach. *F1000Res* 2022;11:1075; doi: 10.12688/f1000research.123714.1
- Maston GA, Evans SK, Green MR. Transcriptional regulatory elements in the human genome. *Annu Rev Genomics Hum Genet* 2006;7:29–59; doi: 10.1146/annurev.genom.7.080505.115623
- Maurya A, Szymanski M, Karlowski WM. ARA: A flexible pipeline for automated exploration of NCBI SRA datasets. *Gigascience* 2022;12; doi: 10.1093/gigascience/giad067
- Meltzer PS. Small RNAs with big impacts. *Nature* 2005; 435(7043):745–746; doi: 10.1038/435745a
- Mendell JT, Olson EN. MicroRNAs in stress signaling and human disease. *Cell* 2012;148(6):1172–1187; doi: 10.1016/j.cell.2012.02.005
- Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: Insights into functions. *Nat Rev Genet* 2009;10(3):155–159; doi: 10.1038/nrg2521
- Mitsis T, Efthimiadou A, Bacopoulou F, et al. Transcription factors and evolution: An integral part of gene expression (review). *World Acad Sci J* 2020;2(1):3–8; doi: 10.3892/wasj.2020.32
- Muzellec B, Teleniczuk M, Cabeli V, et al. PyDESeq2: A python package for bulk RNA-Seq differential expression analysis. *Bioinformatics* 2023;39(9); doi: 10.1093/bioinformatics/btad547
- Nersisyan S, Galatenko A, Galatenko V, et al. MiRGTF-Net: Integrative MiRNA-Gene-TF network analysis reveals key drivers of breast cancer recurrence. *PLoS One* 2021;16(4): e0249424; doi: 10.1371/journal.pone.0249424
- Pande A. Co-Regulatory network of transcription factor and microRNA: A key player of gene regulation. *Biomedical and Biotechnology Research Journal* 2021;5(4):374–379; doi: 10.4103/bbrj.bbrj_182_21
- Pang D. XS. Next Generation Sequencing Transcriptomes of Breast Cancer. 2017. Available from: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi> [Last accessed: October 14, 2024].
- Persson H, Kvist A, Rego N, et al. Identification of new MicroRNAs in paired normal and tumor breast tissue suggests a dual role for the ERBB2/Her2 gene. *Cancer Res* 2011;71(1): 78–86; doi: 10.1158/0008-5472.CAN-10-1869
- Pian C, Zhang G, Tu T, et al. LncCeRBase: A database of experimentally validated human competing endogenous long non-coding RNAs. *Database (Oxford)* 2018;2018(2018); doi: 10.1093/database/bay061
- Polyak K. Heterogeneity in breast cancer. *J Clin Invest* 2011; 121(10):3786–3788; doi: 10.1172/JCI60534
- Poursheikhani A, Abbaszadegan MR, Kerachian MA. Mechanisms of long non-coding RNA function in colorectal cancer tumorigenesis. *Asia Pac J Clin Oncol* 2021;17(1):7–23; doi: 10.1111/ajco.13452
- PyQt5 Reference Guide. PyQt Documentation v5.15.7. 2023. Available from: <https://www.riverbankcomputing.com/static/Docs/PyQt5/#> [Last accessed: October 15, 2024].
- Rani V, Sengar RS. Biogenesis and Mechanisms of MicroRNA-Mediated Gene Regulation. *Biotechnol Bioeng* 2022;119(3): 685–692; doi: 10.1002/bit.28029
- Rao MRS. Advances in Experimental Medicine and Biology: Long Non Coding RNA Biology. Springer; 2017.
- Sanbomics. Bioinformatic Services. 2024. Available from: <https://www.sanbomics.com> [Last accessed: October 14, 2024].
- Shaath H, Elango R, Alajezi NM. Molecular classification of breast cancer utilizing long non-coding Rna (Lncrna) transcriptomes identifies novel diagnostic lncrna panel for triple-negative breast cancer. *Cancers (Basel)* 2021;13(21); doi: 10.3390/cancers13215350
- Shen Y, Xu Y, Li Y, et al. Regulatory role of LncRNA CASC15 in malignant tumors. 2024; doi: 10.22541/au.171923176.66947256/v1
- Shi Y, Gao S, Zheng Y, et al. Lncrna Casc15 functions as an unfavorable predictor of ovarian cancer prognosis and inhibits tumor progression through regulation of Mir-221/Arid1a axis. *Onco Targets Ther* 2019;12:8725–8736; doi: 10.2147/OTT.S219900
- Shopbell PL, Britton MC, Ebert R, et al. Matplotlib—A Portable Python Plotting Package. 2005.
- Shui Y, Yu X, Duan R, et al. MiR-130b-3p inhibits cell invasion and migration by targeting the notch ligand Delta-like 1 in breast carcinoma. *Gene* 2017;609:80–87; doi: 10.1016/j.gene.2017.01.036
- Sideris N, Dama P, Bayraktar S, et al. LncRNAs in breast cancer: A link to future approaches. *Cancer Gene Ther* 2022; 29(12):1866–1877; doi: 10.1038/s41417-022-00487-w
- Statello L, Guo C-J, Chen L-L, et al. Gene regulation by long non-coding RNAs and its biological functions. *Nat Rev Mol Cell Biol* 2021;22(2):96–118; doi: 10.1038/s41580-020-00315-9
- Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;102(43):15545–15550.
- Taheri M, Safarzadeh A, Hussien BM, et al. LncRNA/MiRNA/MRNA network introduces novel biomarkers in prostate cancer. *Cells* 2022;11(23); doi: 10.3390/cells11233776
- Tong Z, Cui Q, Wang J, et al. TransmiR v2.0: An updated transcription factor-microRNA regulation database. *Nucleic Acids Res* 2019;47(D1):D253–D258; doi: 10.1093/nar/gky1023
- Vlachos IS, Paraskevopoulou MD, Karagkouni D, et al. DIANA-TarBase v7.0: Indexing more than half a million experimentally supported MiRNA:MRNA interactions. *Nucleic Acids Res* 2015;43(Database issue):D153–D159; doi: 10.1093/nar/gku1215
- Wan X, Hao S, Hu C, et al. Identification of a novel LncRNA-MiRNA-MRNA competing endogenous RNA network associated with prognosis of breast cancer. *J Biochem Mol Toxicol* 2022;36(8):e23089; doi: 10.1002/jbt.23089

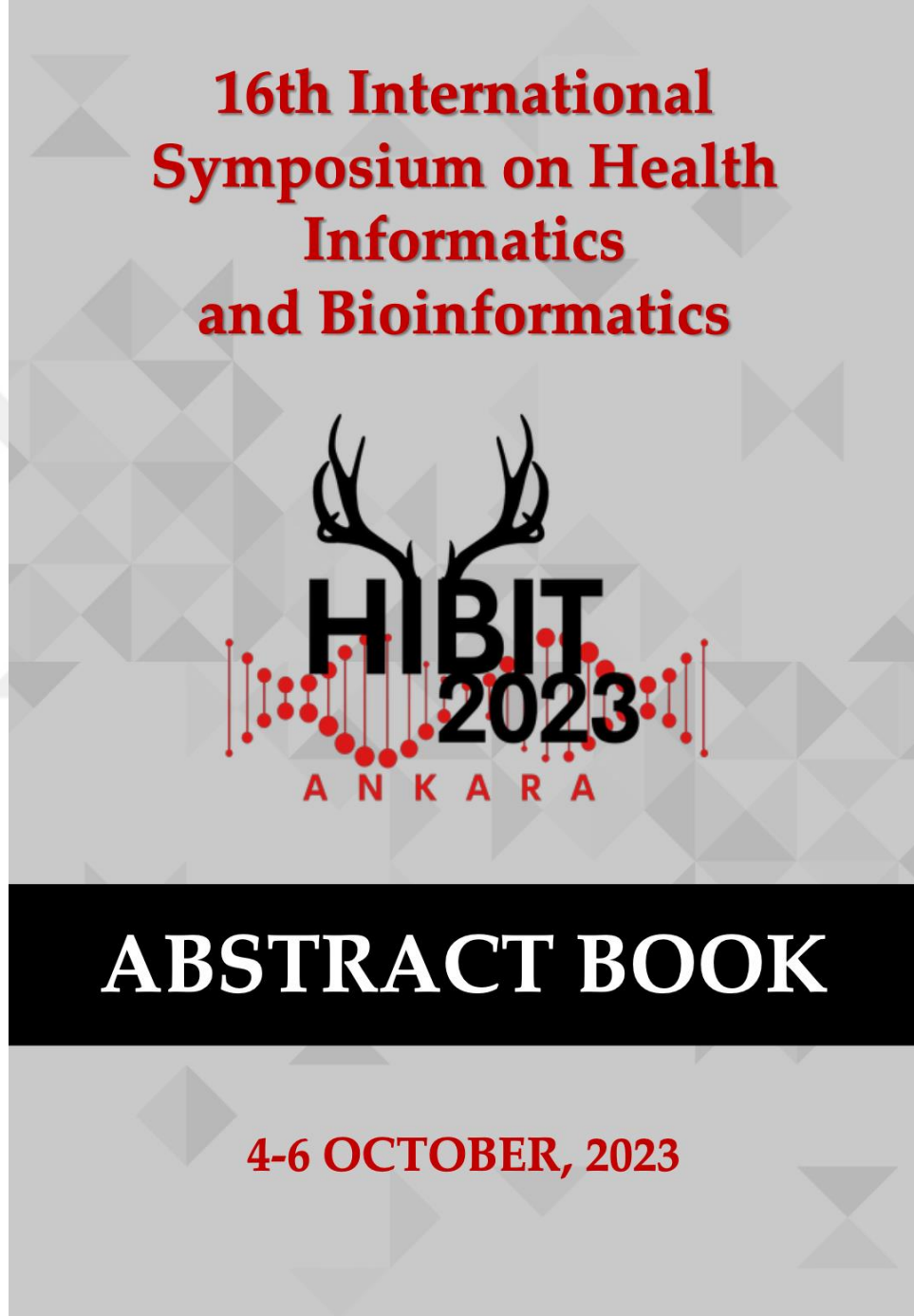
- Wang D, Wang L, Han J, et al. Bioinformatics-Based analysis of the LncRNA-MiRNA-MRNA network and TF regulatory network to explore the regulation mechanism in spinal cord ischemia/reperfusion injury. *Front Genet* 2021;12:650180; doi: 10.3389/fgene.2021.650180
- Wang J. TTF: An R-Based integrative tool for decoding human transcription factor–target interactions. *Biomolecules* 2024;14(7); doi: 10.3390/biom14070749
- Wang P, Guo Q, Qi Y, et al. LncACTdb 3.0: An updated database of experimentally supported CeRNA interactions and personalized networks contributing to precision medicine. *Nucleic Acids Res* 2022;50(D1):D183–D189; doi: 10.1093/nar/gkab1092
- Wang Y, Zhou Y, Yang Z, et al. MiR-204/ZEB2 Axis functions as key mediator for MALAT1-induced epithelial-mesenchymal transition in breast cancer. *Tumour Biol* 2017;39(7):101042831769099–101042831769098; doi: 10.1177/1010428317690998
- Willis S, De P, Dey N, et al. Enriched transcription factor signatures in triple negative breast cancer indicates possible targeted therapies with existing drugs. *Meta Gene* 2015;4: 129–141; doi: 10.1016/j.mgene.2015.04.002
- Wu Y, Li M, Lin J, et al. Hippo/TEAD4 signaling pathway as a potential target for the treatment of breast cancer (Review). *Oncol Lett* 2021;21(4):313; doi: 10.3892/ol.2021.12574
- Wu Z, Liu X, Liu L, et al. Regulation of LncRNA expression. *Cell Mol Biol Lett* 2014;19(4):561–575; doi: 10.2478/s11658-014-0212-6
- Xu G, Yu J, Lyu J, et al. Proteogenomic landscape of breast ductal carcinoma reveals tumor progression characteristics and therapeutic targets. *Adv Sci (Weinh)* 2024;11(46): e2401041; doi: 10.1002/adv.202401041
- Xu T, Le TD, Liu L, et al. Identifying cancer subtypes from MiRNA-TFmRNA Regulatory networks and expression data. *PLoS One* 2016;11(4):e0152792; doi: 10.1371/journal.pone.0152792
- Yang G, Lu X, Yuan L. LncRNA: A link between RNA and cancer. *Biochim Biophys Acta* 2014;1839(11):1097–1109; doi: 10.1016/j.bbagr.2014.08.012
- Yang Y, Wang D, Miao YR, et al. LncRNASNP v3: An updated database for functional variants in long non-coding RNAs. *Nucleic Acids Res* 2023;51(D1):D192–D198; doi: 10.1093/nar/gkac981
- Ye Y, Li SL, Wang SY. Construction and analysis of MRNA, MiRNA, LncRNA, and TF regulatory networks reveal the key genes associated with prostate cancer. *PLoS One* 2018; 13(8):e0198055; doi: 10.1371/journal.pone.0198055
- Ye Z, Fang J, Wang Z, et al. Bioinformatics-Based analysis of the LncRNA-MiRNA-MRNA and TF regulatory networks reveals functional genes in esophageal squamous cell carcinoma. *Biosci Rep* 2020;40(8); doi: 10.1042/BSR20201727
- Yong SL, Dutta A. MicroRNAs in cancer. *Annu Rev Pathol* 2009;4:199–227; doi: 10.1146/annurev.pathol.4.110807.092222
- Yoon S, Kim SY, Nam D. Improving gene-set enrichment analysis of RNA-Seq data with small replicates. *PLoS One* 2016; 11(11):e0165919; doi: 10.1371/journal.pone.0165919
- Yu L, Xu Q, Yu W, et al. LncRNA cancer susceptibility candidate 15 accelerates the breast cancer cells progression via MiR-153–3p/KLF5 positive feedback loop. *Biochem Biophys Res Commun* 2018;506(4):819–825; doi: 10.1016/j.bbrc.2018.10.131
- Yu Y, Nangia-Makker P, Farhana L, et al. A novel mechanism of LncRNA and MiRNA interaction: CCAT2 Regulates MiR-145 expression by suppressing its maturation process in colon cancer cells. *Mol Cancer* 2017;16(1):155–111; doi: 10.1186/s12943-017-0725-5
- Zhao CM, Li LL, Xu JW, et al. LINC00092 suppresses the malignant progression of breast invasive ductal carcinoma through modulating SFRP1 expression by sponging MiR-1827. *Cell Transplant* 2022;31:9636897221086967; doi: 10.1177/09636897221086967
- Zhao H, Yin X, Xu H, et al. LncTarD 2.0: An updated comprehensive database for experimentally-supported functional LncRNA–target regulations in human diseases. *Nucleic Acids Res* 2023;51(D1):D199–D207; doi: 10.1093/nar/gkac984
- Zhao X, Tang D, Chen X, et al. Functional LncRNA-MiRNA-MRNA networks in response to baicalein treatment in hepatocellular carcinoma. *Biomed Res Int* 2021;2021:8844261; doi: 10.1155/2021/8844261
- Zhou RS, Zhang EX, Sun QF, et al. Integrated analysis of LncRNA-MiRNA-MRNA CeRNA network in squamous cell carcinoma of tongue. *BMC Cancer* 2019;19(1):779; doi: 10.1186/s12885-019-5983-8
- Zhu S, Chen CY, Hao Y. LncRNA KCNQ1OT1 Acts as MiR-216b-5p sponge to promote colorectal cancer progression via up-regulating ZNF146. *J Mol Histol* 2021;52(3):479–490; doi: 10.1007/s10735-020-09942-0

Abbreviations Used

DEG	=	Differentially Expressed Gene
FDR	=	False Discovery Rate
GEO	=	Gene Expression Omnibus
GSEA	=	Gene Set Enrichment Analysis
IDC	=	Invasive Ductal Carcinoma
lncRNA	=	long non-coding RNA
miRNA	=	microRNA
mRNA	=	messenger RNA
MSigDB	=	Molecular Signatures Database
NGS	=	Next Generation Sequencing
TCCA	=	The Cancer Genome Atlas
TF	=	Transcription Factor

TEZDEN ÇIKAN BİLDİRİLER

1) Begüm Karaoğlu1 , Bala Gür Dedeoğlu1, “Therapeutic Targets for Metastatic Breast Cancer: Insights from miRNA and Gene Regulation Circuits”, 16th International Symposium on Health Informatics and Bioinformatics, 4-6 October 2023.



Therapeutic Targets for Metastatic Breast Cancer: Insights from miRNA and Gene Regulation Circuits

Begüm Karaoğlu¹, Bala Gür Dedeoğlu¹

¹Ankara University, Biotechnology Institute, Ankara, Turkey

Breast cancer is the most common cancer type in women and takes part in first in cancer-related deaths. With early diagnosis and treatment, breast cancer has a high chance of successful treatment. However, when breast cancer is not diagnosed or treated early, it can metastasize to various tissues and distant organs. This emphasizes importance of accurate and usable biomarkers in diagnosing breast cancer and predicting its prognosis. Understanding regulatory relationship between key molecules, such as transcription factors (TFs) and microRNAs (miRNAs), which are essential for gene expression at transcriptional and post-transcriptional levels, contributes to a better understanding of cancer process. In this study, we used a local circuit analysis program called miRcuit, which constructs TF-mRNA-miRNA regulatory circuits, to investigate the differences in regulatory circuits between metastatic cancer patients and healthy control individuals. For this purpose, we downloaded gene expression data of mRNA (GSE45498, GSE41970) and miRNA (GSE45498, GSE38167) from GEO for metastatic and healthy individuals. Significant changes in expression were used to create lists of individual mRNAs and miRNAs, which were then inputted into miRcuit program to obtain the biological circuits between metastatic patients and control groups. As a result, we identified 7 distinct open circuits that play a role in metastasis process and show significant expression changes during this process. These circuits were: *ESR1* |---- *miR-203* ----| *BIRC5*, *ESR1* |---- *miR-203* ----| *FGF2*, *ESR1* |---- *miR-203* ----| *IGF1*, *ESR1* |---- *miR-203* ----| *TYMS*, *FOS* |---- *miR-101* ----| *DLC1*, *FOS* |---- *miR-101* ----| *FGF2*, *FOS* |---- *miR-155* ----| *FGF2*. Understanding the regulatory networks of these genes and miRNAs may provide valuable insights into mechanisms of breast cancer metastasis and could potentially lead to development of novel diagnostic and therapeutic strategies for metastatic breast cancer.

Begüm Karaoğlu¹, Bala Gür Dedeoğlu¹

¹ Ankara University, Biotechnology Institute, Ankara, Turkey

Abstract

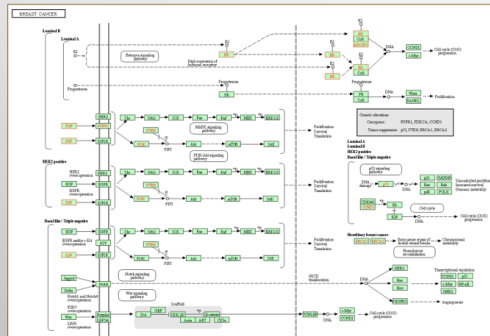
Breast cancer is the most common cancer type in women and takes part in first in cancer-related deaths. With early diagnosis and treatment, breast cancer has a high chance of successful treatment. However, when breast cancer is not diagnosed or treated early, it can metastasize to various tissues and distant organs. This emphasizes importance of accurate and usable biomarkers in diagnosing breast cancer and predicting its prognosis. Understanding regulatory relationship between key molecules, such as transcription factors (TFs) and microRNAs (miRNAs), which are essential for gene expression at transcriptional and post-transcriptional levels, contributes to a better understanding of cancer process. In this study, we used a local circuit analysis program called *miRcuit*, which constructs TF-miRNA-miRNA regulatory circuits, to investigate the differences in regulatory circuits between metastatic cancer patients and healthy control individuals. For this purpose, we downloaded gene expression data of mRNA and miRNA from GEO for metastatic and healthy individuals and inputted into miRcuit program to obtain the biological circuits between metastatic patients and control groups. As a result, we identified 7 distinct open circuits that play a role in the metastasis process and show significant expression changes during this process. To investigate the biological roles of these circuits, we performed pathway enrichment analysis.

Material and Method

In the study, mRNA (GSE45498, GSE41970) and miRNA (GSE45498, GSE38167) expression data were downloaded from GEO for metastatic and healthy individuals. TFs were extracted from the mRNA data, and molecules showing significant expression changes were subjected to analysis (a significance threshold of 1.5-fold change was selected). Biological circuits involving TF-miRNA-miRNA were constructed using miRcuit, a local circuit analysis program developed using Python 3.7.4. Genes obtained from this analysis were further subjected to pathway enrichment analysis using the Webgestalt web site.

Result and Discussion

As a result of the miRcuit analysis, 7 circuits were obtained. It was determined that these circuits focused on genes involved in metastatic processes such as cancer development and apoptosis inhibition. The roles of the obtained genes in biological processes were identified through pathway enrichment analysis in Webgestalt.



miRcuit: A Tool For Integrating Analysis of mRNA and miRNA Expression Data to Infer Regulatory Circuits

Home Flow Chart About Contact

[Click to see examples](#)

Data with:

- ☐ Raw Data
- ☐ Differentially Expression Values

Please do not select threshold values if you use differentially expression values.

Analysis method...

- ☐ without GSEA and miSEA
- ☐ with GSEA and miSEA
- ☐ with miSEA
- ☐ with GSEA

Select file...

mRNA File:

Threshold for logFC:

miRNA File:

Threshold for logFC:

Please do not panic, it takes time to process! 🕒

[Click to See Results](#)

TF	miRNA	mRNA
ESR1	miR-203	BIRC5
ESR1	miR-203	FGF2
ESR1	miR-203	IGF1
ESR1	miR-203	TYMS
FGF5	miR-101	DLG1
FGF5	miR-101	FGF2
FGF5	miR-155	FGF2

Conclusion

Understanding the regulatory networks of genes and miRNAs that differ between metastatic cancer patients and healthy control may provide valuable insights into mechanisms of breast cancer metastasis. This understanding could potentially lead to the development of novel diagnostic and therapeutic strategies for metastatic breast cancer.

- 1) B. Karaoglu, E. Dalli, V. Mazlum, H. Ogul, B. Gur Dedeoglu. “miRcuit: A tool for integrating analysis of mRNA and miRNA expression data to infer regulatory circuits” European Association for Cancer Research, EACR 2023: Innovative Cancer Science, 12-15 June 2023, Torino, Italy. *Molecular Oncology* 17 (Suppl. 1) (2023) pp.24.



EACR23-0215

miRcui: A tool for integrating analysis of mRNA and miRNA expression data to infer regulatory circuits

B. Karaoglu¹, E. Dallr², V. Mazlum³,
H. Ogu⁴, B. Gur Dedeoglu¹

¹Ankara University Biotechnology Institute,
System Biotechnology, Ankara, Türkiye

²Ege University, Department of Biology, Izmir, Türkiye

³Cankaya University,

Department of Computer Engineering, Ankara, Türkiye

⁴Østfold University College,

Department of Computer Science and Communication,
Halden, Norway

Introduction

Gene expression is controlled by many factors. While transcription factors (TFs) perform this control at the transcriptional level, microRNAs (miRNAs) are involved at the post-transcriptional level. Biological networks and circuits based on TFs, miRNAs and their targets are extremely important for a better understanding of cellular processes. For this purpose, we developed an algorithm and designed a web tool that creates biological circuits by using experiment-specific and literature-based data together.

Material and Methods

Circuit analysis tool was created using Python software language (v.3.7.4) to construct regulatory circuits from the high-throughput data containing expression values of genes and miRNAs. The datasets to be uploaded are mRNA (also including TFs) and miRNA files that show significant expression changes under two different conditions like tumor and normal tissues. Then, the absolute log(FC) values were calculated by performing the log₂ transformation for each molecule in the sets (mRNA list and miRNA list) by Pandas and Math libraries offered by Python. Each molecule is labelled "up" or "down" and the direction of the intermolecular relationship is determined.

Results and Discussions

The mRNA (TF) and miRNA lists showing significant expression changes were combined and all possible mRNA (TF)-miRNA pairs were obtained. Significant relationships were extracted by comparing all possible pairs with known pairs in databases (Data bases and direction of the relationship: TargetScan: miRNA-mRNA relationship, miRNA ----| mRNA or miRNA ----| TF. TransmiR: TF-miRNA relationship, TF ----> miRNA. hTFtarget: TF-mRNA relationship, TF ----> mRNA). All TF-mRNA, TF-miRNA, miRNA-mRNA and miRNA-TF pairs showing significant expression changes were obtained and intersected with each other. As a result, two different types of open circuit and one closed type circuit files were created.

Conclusion

The algorithm developed to create biological circuits extracting the TF – miRNA – mRNA relationships will soon be available as a web tool under the name "Integrative Regulatory Circuit Finder".

novel Rho-associated protein kinases (ROCK) inhibitors

M. Beljkaš¹, S. Oljačić¹, M. Petković², K. Nikolić¹

¹Faculty of Pharmacy- University of Belgrade,
Department of Pharmaceutical Chemistry, Belgrade, Serbia

²Faculty of Pharmacy- University of Belgrade,
Department of Organic Chemistry, Belgrade, Serbia

Introduction

Rho-associated protein kinases (ROCK1 and ROCK2) are involved in various cellular functions as the key regulators of numerous substrates. Alterations in ROCK activities are associated with many diseases, including cancer. ROCKs play a critical role in the initiation and growth of cancer through the activation and stabilization of some oncogenes and in cancer progression through influence on migration, invasion, and angiogenesis. Recent studies have shown an association between the overactivity of ROCKs and lower survival in patients with the most aggressive carcinomas, such as pancreatic ductal adenocarcinoma (PDAC).

Moreover, the activity of ROCK is related to tumor cells and the tumor environment, which contributes to tumor progression. Therefore, ROCKs are very good drug targets in anticancer research. In this study 3D-quantitative structure-activity relationship (3D-QSAR) method was applied to examine the structures of ROCK1 and ROCK2 pharmacophores and design dual ROCK1 and ROCK2 inhibitors as novel anticancer agents.

Material and Methods

The dataset of structurally diverse ROCK1 and ROCK2 inhibitors with IC₅₀ values was downloaded from the ChEMBL database. Conformations of all compounds were optimized by use of Hartree-Fock with 3-21G basis set method. The bioactive conformations of the ligands for the 3D-QSAR study were generated by a molecular docking study that was performed by GOLD software 2022.1.0. Pentacle software 1.07 was used to generate 3D-QSAR models for ROCK1 and ROCK2 isoforms. All compounds were divided into a training set and a test set to prepare both 3D-QSAR models.

Results and Discussions

The molecular determinants responsible for ROCK1 and ROCK2 inhibiting activities have been identified by 3D-QSAR modeling and comparative analysis of the ROCK1 and ROCK2 pharmacophores. Both created 3D-QSAR models met all the internal and external validation criteria. Validated 3D-QSAR models were used to design new ROCK1 and ROCK2 inhibitors and predict their activities. The most active designed compounds were further evaluated using *in silico* ADMET (Absorption, Distribution, Metabolism, Excretion, Toxicity) profiling. Based on these findings, novel dual ROCK1/ROCK2 inhibitors were selected for further synthesis.

Conclusion

According to the obtained results, both 3D-QSAR models can be considered as a reliable prediction of the ROCK1 and ROCK2 inhibiting activity of novel compounds and thus significantly facilitate the selection of the best candidates for further synthesis and biological evaluation.

- 2) Karaoglu B, Noyan S, Gurdal H, Gur Dedeoglu B. “miR-16-5p and E2F-dependent miRNA-TF-mRNA Regulatory Circuits in Breast Cancer.” European Association for Cancer Research 2021 Virtual Congress (EACR 2021).

