



T.C.  
İSTANBUL ÜNİVERSİTESİ-CERRAHPAŞA  
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ



**DOKTORA TEZİ**

**MAKİNE ÖĞRENMESİ YÖNTEMLERİ ile MİKRODİZİ HEDEF GEN  
ÇIKARIMI**

**Zerrin YILDIZ ÇAVDAR**

**DANIŞMAN**

**Prof. Dr. Ahmet SERTBAŞ**

(danışman imzalı olacaktır)

**II. DANIŞMAN**

**Doç. Dr. Leyla TÜRKER ŞENER**

**Bilgisayar Mühendisliği Anabilim Dalı**

**Bilgisayar Mühendisliği, Doktora Programı**

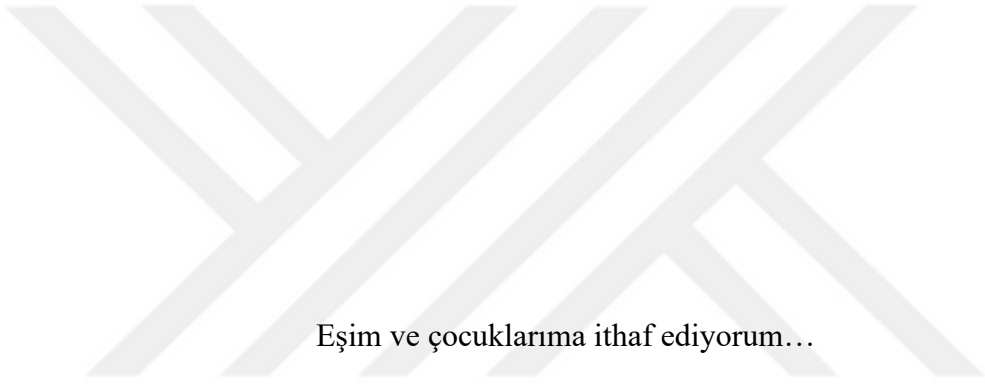
**Ocak, 2025**

## TEZ KABUL VE ONAYI

Zerrin YILDIZ ÇAVDAR tarafından, Prof. Dr. Ahmet SERTBAŞ danışmanlığında hazırlanan "MAKİNE ÖĞRENMESİ YÖNTEMLERİ ile MİKRODİZİ HEDEF GEN ÇIKARIMI" başlıklı bu çalışma, jürimiz tarafından 15/01/2025 tarihinde yapılan sınav sonucunda oy birliği ile başarılı bulunarak **Doktora Tezi** olarak kabul edilmiştir.

### Tez Jürisi

	İmza	Sonuç
DANIŞMAN	Prof. Dr. Ahmet SERTBAŞ	<input checked="" type="checkbox"/>
	İstanbul Üniversitesi Cerrahpaşa Bilgisayar Mühendisliği Anabilim Dalı	Kabul <input type="checkbox"/> Ret
ÜYE	Prof. Dr. Muhammed Ali AYDIN	<input checked="" type="checkbox"/>
	İstanbul Üniversitesi Cerrahpaşa Bilgisayar Mühendisliği Anabilim Dalı	Kabul <input type="checkbox"/> Ret
ÜYE	Prof. Dr. Nizamettin AYDIN	<input checked="" type="checkbox"/>
	İstanbul Teknik Üniversitesi Bilgisayar Mühendisliği Anabilim Dalı	Kabul <input type="checkbox"/> Ret
ÜYE	Dr. Öğr. Üyesi Muhammed Erdem İSENKUL	<input checked="" type="checkbox"/>
	İstanbul Üniversitesi Cerrahpaşa Bilgisayar Mühendisliği Anabilim Dalı	Kabul <input type="checkbox"/> Ret
ÜYE	Dr. Öğr. Üyesi Tolga ENSARİ	<input checked="" type="checkbox"/>
	Arkansas Tech Üniversitesi Engineering and Computing Sciences Anabilim Dalı	Kabul <input type="checkbox"/> Ret



Eşim ve çocuklarıma ithaf ediyorum...

## **BÜTÇE DESTEKLERİ**

### **MAKİNE ÖĞRENMESİ YÖNTEMLERİ ile MİKRODİZİ HEDEF GEN ÇIKARIMI**

Bu tez çalışması için herhangi bir kurumdan bütçe desteği alınmamıştır.



## TEŐEKKÜR

Doktora alıŐmam boyunca her zaman bana destek olan eŐim Mehmet AVDAR'a ve daha az zaman ayırmak zorunda kaldıĐım, enerji kaynaklarım kızım Azra AVDAR ve oĐlum Burak AVDAR'a, bu zorlu yolculukta yolumu aydınlatan saygıdeĐer hocam Prof.Dr. Ahmet SERTBAŐ, uzak mesafelerden desteĐini hi esirgemeyen Dr.ŐĐr.Üyesi Tolga ENSARİ, enerjisi ve yaratıcı fikirleri ile ilham kaynaĐı olan Do.Dr.Leyla TÜRKER ŐENER hocalarıma ve bu süreçte tanıştıĐım yol arkadaŐım, moral ve motivasyon kaynaĐım olan deĐerli arkadaŐım Müberra Nur AKAMAN'a teŐekkür ederim.

Ocak 2025

Zerrin YILDIZ AVDAR

# İÇİNDEKİLER

Sayfa No

TEZ KABUL VE ONAYI.....	ii
BÜTÇE DESTEKLERİ .....	iv
TEŞEKKÜR.....	v
İÇİNDEKİLER.....	vi
ŞEKİL LİSTESİ .....	viii
TABLO LİSTESİ.....	x
SİMGE VE KISALTMA LİSTESİ.....	xii
ÖZET .....	xiv
ABSTRACT .....	xvi
<b>1. GİRİŞ.....</b>	<b>1</b>
<b>2. KAVRAMSAL ÇERÇEVE .....</b>	<b>2</b>
2.1. MİKORNA.....	2
2.2. MİKRODİZİ .....	3
2.2.1. Gen Ekspresyon Analizi.....	4
2.2.2. Mikrodizi Hastalık Veritabanları .....	5
2.2.3. Mikrodizi Veriseti Hazırlama Süreci .....	6
2.2.4. Mikrodizi Verisetleri.....	8
2.3. MAKİNE ÖĞRENMESİ VE BİYOENFORMATİK.....	11
2.3.1. Makine Öğrenmesi Çeşitleri.....	12
2.3.2. Makine Öğrenmesi Yöntemleri.....	13
<b>3. YÖNTEM .....</b>	<b>16</b>
3.1. ÖNERİLEN HEDEF TAHMİN YÖNTEMİ .....	16
3.1.1. Veriseti Ön Hazırlık .....	16
3.1.2. Veri Kalitesi Kontrolleri.....	17
3.1.3. Kümeleme .....	18
3.1.4. Normalizasyon .....	19
3.1.5. Gruplama.....	20
3.1.6. Değerlendirme.....	21

3.2. ÖZGÜN VERİSETİ.....	23
3.3. VERİSETİNİN AYRI DENEYLER ŞEKLİNDE AYRIŞTIRILMASI.....	23
<b>4. BULGULAR .....</b>	<b>28</b>
4.1. ÖZGÜN VERİSETİ ANALİZ VE DEĞERLENDİRME.....	28
4.1.1. Deney Grubu 1 – Nikotin .....	28
4.1.2. Deney Grubu 2 – M100.....	33
4.1.3. Deney Grubu 3 – M1000.....	39
4.1.4. Deney Grubu 4 – Işın .....	45
4.1.5. Deney Grubu 5 – Nikotin ve M100.....	49
4.1.6. Deney Grubu 6 – Nikotin ve M1000.....	54
4.1.7. Deney Grubu 7 – Nikotin ve Işın .....	60
4.1.8. Deney Grubu 8 – Nikotin, Işın ve M100.....	65
4.1.9. Deney Grubu 9 – Nikotin, Işın ve M1000.....	70
4.2. GSE10470 VERİSETİ ANALİZ VE DEĞERLENDİRME.....	76
4.3. GSE8262 VERİSETİ ANALİZ VE DEĞERLENDİRME.....	83
<b>5. TARTIŞMA.....</b>	<b>90</b>
<b>6. SONUÇ VE ÖNERİLER .....</b>	<b>93</b>
<b>KAYNAKLAR.....</b>	<b>94</b>
<b>İNTİHAL RAPORU İLK SAYFASI .....</b>	<b>97</b>

## ŞEKİL LİSTESİ

Sayfa No

Şekil 2.1: DNA – RNA – Protein – Canlı süreci.....	6
Şekil 2.2: MikroRNA.....	7
Şekil 2.3: Mikrodizi çip yapısı.....	7
Şekil 2.4: Özgün veriseti hazırlık süreci: floresan boyama ve tarama süreci .....	8
Şekil 2.5: Makine öğrenmesi ve biyoenformatik .....	12
Şekil 2.6: Makine Öğrenmesi Yöntemleri .....	13
Şekil 3.1: SMATP Yöntemi .....	16
Şekil 4.1: Normalizasyon Öncesi Sinyal Dağılımı - Nikotin.....	30
Şekil 4.2: Saptama Skorları - Nikotin .....	30
Şekil 4.3: Normalizasyon Öncesi Kümeleme Isı Haritası- Nikotin .....	31
Şekil 4.4: Normalizasyon Sonrası Sinyal Dağılımı- Nikotin.....	31
Şekil 4.5: Normalizasyon Sonrası Kümelenmiş Veri Isı Haritası - Nikotin .....	32
Şekil 4.6: Gruplar Arası Karşılaştırmalar - Nikotin.....	33
Şekil 4.7: Normalizasyon Öncesi Sinyal Dağılımı – M100.....	35
Şekil 4.8: Saptama Skorları – M100 .....	35
Şekil 4.9: Normalizasyon Öncesi Kümeleme Isı Haritası - M100.....	36
Şekil 4.10: Normalizasyon Sonrası Sinyal Dağılımı – M100.....	36
Şekil 4.11: Normalizasyon Sonrası Kümelenmiş Veri Isı Haritası – M100 .....	37
Şekil 4.12: Gruplar Arası Karşılaştırmalar – M100 .....	38
Şekil 4.13: En Çok Etkilenen 10 Gen – M100.....	39
Şekil 4.14: Normalizasyon Öncesi Sinyal Dağılımı – M1000.....	40
Şekil 4.15: Saptama Skorları – M1000 .....	41
Şekil 4.16: Normalizasyon Öncesi Kümeleme Isı Haritası - M1000.....	41
Şekil 4.17: Normalizasyon Sonrası Sinyal Dağılımı – M1000.....	42
Şekil 4.18: Normalizasyon Sonrası Kümelenmiş Veri Isı Haritası – M1000 .....	42
Şekil 4.19: Gruplar Arası Karşılaştırmalar – M1000 .....	44
Şekil 4.20: En Çok Etkilenen 10 Gen – M1000.....	44
Şekil 4.21: Normalizasyon Öncesi Sinyal Dağılımı – Işın .....	46
Şekil 4.22: Saptama Skorları – Işın.....	46
Şekil 4.23: Normalizasyon Öncesi Kümeleme Isı Haritası - Işın .....	47
Şekil 4.24: Normalizasyon Sonrası Sinyal Dağılımı – Işın .....	47
Şekil 4.25: Normalizasyon Sonrası Kümelenmiş Veri Isı Haritası – Işın.....	48
Şekil 4.26: Gruplar Arası Karşılaştırmalar – Işın.....	49
Şekil 4.27: Normalizasyon Öncesi Sinyal Dağılımı – Nikotin ve M100.....	51
Şekil 4.28: Saptama Skorları – Nikotin ve M100 .....	51
Şekil 4.29: Normalizasyon Öncesi Kümeleme Isı Haritası – Nikotin ve M100 .....	52
Şekil 4.30: Normalizasyon Sonrası Sinyal Dağılımı – Nikotin ve M100.....	52
Şekil 4.31: Normalizasyon Sonrası Kümelenmiş Veri Isı Haritası – Nikotin ve M100 .....	53
Şekil 4.32: Gruplar Arası Karşılaştırmalar – Nikotin ve M100 .....	54
Şekil 4.33: Normalizasyon Öncesi Sinyal Dağılımı – Nikotin ve M1000.....	56
Şekil 4.34: Saptama Skorları – Nikotin ve M1000 .....	56
Şekil 4.35: Normalizasyon Öncesi Kümeleme Isı Haritası – Nikotin ve M1000 .....	57
Şekil 4.36: Normalizasyon Sonrası Sinyal Dağılımı – Nikotin ve M1000.....	57
Şekil 4.37: Normalizasyon Sonrası Kümelenmiş Veri Isı Haritası – Nikotin ve M1000 .....	58

Şekil 4.38: Gruplar Arası Karşılaştırmalar – Nikotin ve M1000.....	59
Şekil 4.39: En Çok Etkilenen 10 Gen – Nikotin ve M1000.....	60
Şekil 4.40: Normalizasyon Öncesi Sinyal Dağılımı – Nikotin ve Işın .....	61
Şekil 4.41: Saptama Skorları – Nikotin ve Işın.....	62
Şekil 4.42: Normalizasyon Öncesi Kümeleme Isı Haritası – Nikotin ve Işın.....	62
Şekil 4.43: Normalizasyon Sonrası Sinyal Dağılımı – Nikotin ve Işın .....	63
Şekil 4.44: Normalizasyon Sonrası Kümelenmiş Veri Isı Haritası – Nikotin ve Işın.....	63
Şekil 4.45: Gruplar Arası Karşılaştırmalar – Nikotin ve Işın.....	64
Şekil 4.46: En Çok Etkilenen 10 Gen – Nikotin ve Işın .....	65
Şekil 4.47: Normalizasyon Öncesi Sinyal Dağılımı – Nikotin, Işın ve M100.....	66
Şekil 4.48: Saptama Skorları – Nikotin, Işın ve M100 .....	67
Şekil 4.49: Normalizasyon Öncesi Kümeleme Isı Haritası – Nikotin, Işın ve M100 .....	67
Şekil 4.50: Normalizasyon Sonrası Sinyal Dağılımı – Nikotin, Işın ve M100.....	68
Şekil 4.51: Normalizasyon Sonrası Kümelenmiş Veri Isı Haritası – Nikotin, Işın ve M100 ..	68
Şekil 4.52: Gruplar Arası Karşılaştırmalar – Nikotin, Işın ve M100 .....	70
Şekil 4.53: En Çok Etkilenen 10 Gen – Nikotin, Işın ve M100.....	70
Şekil 4.54: Normalizasyon Öncesi Sinyal Dağılımı – Nikotin, Işın ve M1000.....	72
Şekil 4.55: Saptama Skorları – Nikotin, Işın ve M1000 .....	72
Şekil 4.56: Normalizasyon Öncesi Kümeleme Isı Haritası – Nikotin, Işın ve M1000 .....	73
Şekil 4.57: Normalizasyon Sonrası Sinyal Dağılımı – Nikotin, Işın ve M1000.....	73
Şekil 4.58: Normalizasyon Sonrası Kümelenmiş Veri Isı Haritası – Nikotin, Işın ve M1000	74
Şekil 4.59: Gruplar Arası Karşılaştırmalar – Nikotin, Işın ve M1000 .....	75
Şekil 4.60: En Çok Etkilenen 10 Gen – Nikotin, Işın ve M1000.....	76
Şekil 4.61: Normalizasyon Öncesi Sinyal Dağılımı – GSE10470.....	78
Şekil 4.62: Saptama Skorları – GSE10470 .....	79
Şekil 4.63: Normalizasyon Öncesi Kümeleme Isı Haritası – GSE10470 .....	79
Şekil 4.64: Normalizasyon Sonrası Sinyal Dağılımı – GSE10470.....	80
Şekil 4.65: Normalizasyon Sonrası Kümelenmiş Veri Isı Haritası – GSE10470 .....	81
Şekil 4.66: Gruplar Arası Karşılaştırmalar – GSE10470 .....	82
Şekil 4.67: En Çok Etkilenen 10 Gen – GSE10470.....	83
Şekil 4.68: Normalizasyon Öncesi Sinyal Dağılımı – GSE8262.....	85
Şekil 4.69: Saptama Skorları – GSE8262 .....	86
Şekil 4.70: Normalizasyon Öncesi Kümeleme Isı Haritası – GSE8262 .....	86
Şekil 4.71: Normalizasyon Sonrası Sinyal Dağılımı – GSE8262.....	87
Şekil 4.72: Normalizasyon Sonrası Kümelenmiş Veri Isı Haritası – GSE8262 .....	87
Şekil 4.73: Gruplar Arası Karşılaştırmalar – GSE8262 .....	89
Şekil 4.74: En Çok Etkilenen 10 Gen – GSE8262.....	89

## TABLO LİSTESİ

Sayfa No

<b>Tablo 3.1:</b> Veri Çerçevesi .....	17
<b>Tablo 3.2:</b> Özgün veriseti ham veri karşılaştırma grupları .....	23
<b>Tablo 3.3:</b> Özgün veriseti deney grupları örnek sayıları .....	24
<b>Tablo 3.4:</b> Özgün veriseti yeniden başlıklandırma çalışması .....	25
<b>Tablo 3.5:</b> Özgün veriseti deney grupları .....	27
<b>Tablo 4.1:</b> Veri Çerçevesi - Nikotin .....	28
<b>Tablo 4.2:</b> Örneklerin Sıralanması - Nikotin .....	29
<b>Tablo 4.3:</b> Tasarım Matrisi - Nikotin.....	29
<b>Tablo 4.4:</b> Kontrast Matrisi – Nikotin .....	32
<b>Tablo 4.5:</b> Veri Çerçevesi – M100 .....	34
<b>Tablo 4.6:</b> Örneklerin Sıralanması – M100 .....	34
<b>Tablo 4.7:</b> Tasarım Matrisi – M100.....	34
<b>Tablo 4.8:</b> Kontrast Matrisi – M100 .....	37
<b>Tablo 4.9:</b> Veri Çerçevesi – M1000 .....	39
<b>Tablo 4.10:</b> Örneklerin Sıralanması – M1000 .....	40
<b>Tablo 4.11:</b> Tasarım Matrisi – M1000.....	40
<b>Tablo 4.12:</b> Kontrast Matrisi – M1000 .....	43
<b>Tablo 4.13:</b> Veri Çerçevesi – Işın .....	45
<b>Tablo 4.14:</b> Örneklerin Sıralanması – Işın.....	45
<b>Tablo 4.15:</b> Tasarım Matrisi – Işın .....	45
<b>Tablo 4.16:</b> Kontrast Matrisi – Işın .....	48
<b>Tablo 4.17:</b> Veri Çerçevesi – Nikotin ve M100 .....	50
<b>Tablo 4.18:</b> Örneklerin Sıralanması – Nikotin ve M100 .....	50
<b>Tablo 4.19:</b> Tasarım Matrisi – Nikotin ve M100.....	50
<b>Tablo 4.20:</b> Kontrast Matrisi – Nikotin ve M100 .....	53
<b>Tablo 4.21:</b> Veri Çerçevesi – Nikotin ve M1000 .....	55
<b>Tablo 4.22:</b> Örneklerin Sıralanması – Nikotin ve M1000 .....	55
<b>Tablo 4.23:</b> Tasarım Matrisi – Nikotin ve M1000.....	55
<b>Tablo 4.24:</b> Kontrast Matrisi – Nikotin ve M1000 .....	58
<b>Tablo 4.25:</b> Veri Çerçevesi – Nikotin ve Işın .....	60
<b>Tablo 4.26:</b> Örneklerin Sıralanması – Nikotin ve Işın.....	61
<b>Tablo 4.27:</b> Tasarım Matrisi – Nikotin ve Işın .....	61
<b>Tablo 4.28:</b> Kontrast Matrisi – Nikotin ve Işın .....	64
<b>Tablo 4.29:</b> Veri Çerçevesi – Nikotin, Işın ve M100 .....	65
<b>Tablo 4.30:</b> Örneklerin Sıralanması – Nikotin, Işın ve M100 .....	66
<b>Tablo 4.31:</b> Tasarım Matrisi – Nikotin, Işın ve M100.....	66
<b>Tablo 4.32:</b> Kontrast Matrisi – Nikotin, Işın ve M100 .....	69
<b>Tablo 4.33:</b> Veri Çerçevesi – Nikotin, Işın ve M1000 .....	71
<b>Tablo 4.34:</b> Örneklerin Sıralanması – Nikotin, Işın ve M1000 .....	71
<b>Tablo 4.35:</b> Tasarım Matrisi – Nikotin, Işın ve M1000.....	72
<b>Tablo 4.36:</b> Kontrast Matrisi – Nikotin, Işın ve M1000 .....	74
<b>Tablo 4.37:</b> Veri Çerçevesi – GSE10470 .....	76
<b>Tablo 4.38:</b> Örneklerin Sıralanması – GSE10470 .....	77
<b>Tablo 4.39:</b> Tasarım Matrisi – GSE10470.....	77

<b>Tablo 4.40:</b> Kontrast Matrisi – GSE10470 .....	81
<b>Tablo 4.41:</b> Veri Çerçevesi – GSE8262 .....	83
<b>Tablo 4.42:</b> Örneklerin Sıralanması – GSE8262 .....	84
<b>Tablo 4.43:</b> Tasarım Matrisi – GSE8262.....	84
<b>Tablo 4.44:</b> Kontrast Matrisi – GSE8262.....	88



## SİMGE VE KISALTIMA LİSTESİ

<b>Kısaltmalar</b>	<b>Açıklama</b>
<b>AI</b>	: yapay zeka (Artificial Intelligence)
<b>ANN</b>	: yapay sinir ağları (Artificial Neural Networks)
<b>ATF4</b>	: Activating Transcription Factor 4
<b>cch</b>	: kronik sürekli hipoksi (chronic consistent hypoxia)
<b>cih</b>	: kronik aralıklı hipoksi (chronic intermittent hypoxia)
<b>CNN</b>	: evrimsel sinir ağı (Convolutional Neural Networks)
<b>ctrl</b>	: kontrol
<b>cx</b>	: cortex
<b>dbDEMC</b>	: insan kanserinde diferansiyel olarak ifade edilen mikrornalar (differentially expressed MiRNAs in human cancer)
<b>eBayes</b>	: ampirik bayes (empirical bayes)
<b>FDR</b>	: yanlış keşfi oranı (False Discovery Rate)
<b>GRM8</b>	: Metabotropik Glutamat Reseptörü 8
<b>GEO</b>	: gen ifadesi omnibusu (Gen Expression Omnibus)
<b>hc</b>	: hipocampus
<b>HCA</b>	: homosisteat (oksidatif stres)
<b>HMDD</b>	: insan mikrorna hastalık veritabanı (Human MicroRNA Disease Database)
<b>ICGC</b>	: uluslararası kanser genom konsorsiyumu (International Cancer Genome Consortium)
<b>KNN</b>	: K-en yakın komşu (K-Nearest Neighbour)
<b>ko</b>	: knock-out, ATF4 geni olmayan fare
<b>limma</b>	: mikrodizi verileri için lineer modeller (linear models for microarray data)
<b>M100</b>	: Metformin (şeker hastalarına verilen bir ilaç türü). M100: 100 milimolar metformin
<b>M1000</b>	: Metformin (şeker hastalarına verilen bir ilaç türü). M1000: 1000 milimolar metformin
<b>MCF7</b>	: meme kanserine ait MCF7 isimli hücre soyu
<b>MDA</b>	: meme kanserine ait MDA-N1 isimli hücre soyu
<b>mikroRNA</b>	: mikro ribonükleik asit (Micro Ribonucleic Acids)
<b>ML</b>	: Makine öğrenmesi (machine learning)
<b>mRNA</b>	: messengerRNA

<b>NCBI</b>	: ulusal biyoteknoloji bilgi merkezi (National Center for Biotechnology Information)
<b>nt</b>	: nükleotit
<b>ÖzgünVeri</b>	: “Nikotin ve Radrasyon İndüklemeli Meme Kanseri Kök Hücreleri Üzerine Metforminin Antikanser Etkisi” proje veriseti
<b>qPCR</b>	: quantitative Polymerase Chain Reaction
<b>PCA</b>	: temel bileşenler analizi (Principal Component Analysis)
<b>RNN</b>	: evrişimsel sinir ağı (Recurrent Neural Networks)
<b>SEN1</b>	: sumo spesifik proteaz 1
<b>SMATP</b>	: akıllı mikrodizi analiz ve hedef çıkarımı (Smart Microarray Analyze and Target Prediction)
<b>SVM</b>	: destek vektör makineleri (Support Vector Machines)
<b>TCGA</b>	: kanser genom atlası (The Cancer Genome Atlas)
<b>t-SNE</b>	: t-dağılımlı stokastik komşu yerleştirme (t-Distributed Stochastic Neighbor Embedding)
<b>UCSC</b>	: California Santa Cruz Üniversitesi Genom Tarayıcısı
<b>wt</b>	: wild type, normal fare

## ÖZET

### [DOKTORA TEZİ]

#### [MAKİNE ÖĞRENMESİ YÖNTEMLERİ ile MİKRODİZİ HEDEF GEN ÇIKARIMI]

[Zerrin YILDIZ ÇAVDAR]

**İstanbul Üniversitesi-Cerrahpaşa**

**Lisansüstü Eğitim Enstitüsü**

**Bilgisayar Mühendisliği Anabilim Dalı**

**Bilgisayar Mühendisliği, Doktora Programı**

[Danışman : Prof. Dr. Ahmet SERTBAŞ]

[II. Danışman : Doç. Dr. Leyla TÜRKER ŞENER]

[Mikrodizi, mikrona, dna, rna verileri gibi büyük tıbbi verilerin işlenmesi ve değerlendirilmesi zorlu bir süreçtir. Biyoenformatik, bu tür tıbbi bilgilerin makine öğrenmesi gibi bilgisayar teknolojileri kullanılarak işlenmesini sağlayan çok disiplinli bir araştırma alanıdır. Biyoloji, matematik, bilgisayar bilimlerinin bir arada çalıştığı biyoenformatik alanındaki çalışmalar; hastalık teşhisi, önleyici tedavi oluşturma, ilaç geliştirme ve kişiselleştirilmiş tedavi süreçleri gibi birçok konuda destek sağlamaktadır. Farklı organizmalarda görülebilen ve kanser ile ilişkili olduğu düşünülen mikroRNA(miRNA, miR olarak da isimlendirilir)ların genler üzerindeki etkisini araştırmada biyoenformatik yöntemler tercih edilmektedir. Mikrodizi teknolojisi mikroRNA, DNA, RNA gibi verilerin değerlendirilmesinde kullanılan tekniklerden biridir. Mikrodizi yöntemi ile gen ifadesinin analizi, hedef gen çıkarımı, mutasyon analizi ve biyomarker keşfi gibi birçok araştırma mümkün olabilmektedir. Mikrodizi verilerinin hazırlanmasında farklı yöntemler kullanılabilir ve bu yöntemlere bağlı olarak verinin incelenme şekli değişiklik göstermektedir. Bu çalışmada mikrodizi tekniği ile hazırlanmış biyolojik verilerin analizi ve hedef gen çıkarım çalışması yapılmıştır. Bu amaçla makine öğrenmesi yöntemlerini kullanan bir metot geliştirilmiş ve farklı verisetlerinde uygulanmıştır.

Çalışmada açık erişimli iki verisetinin yanında özgün bir veriseti de kullanılmıştır. 3 farklı veriseti grubuna ait 11 deney için; verilerin bilgisayar ortamında işlenmeye hazır hale getirilmesinden hedef genlerin tespitine kadar devam eden süreç detaylı olarak incelemektedir. Çalışma ham verinin bilgisayar ortamına aktarılması, veri kalitesi kontrolleri, kümeleme, normalizasyon, karşılaştırma grupları hazırlanması ve hedef gen çıkarım süreçlerinden oluşmaktadır. |

Ocak 2025 , |118| sayfa.

**Anahtar kelimeler:** |Mikrodizi, gen ekspresyon analizi, hedef gen çıkarımı, mikroRNA, miRNA, makine öğrenmesi |



## **ABSTRACT**

**[Ph.D. THESIS]**

**[MICROARRAY TARGET GEN EXTRACTION with MACHINE LEARNING  
METHODS ]**

**[Zerrin YILDIZ ÇAVDAR]**

**İstanbul University-Cerrahpaşa**

**Institute of Graduate Studies**

**Department of Computer Engineering**

**Computer Engineering, Phd Program**

**[Supervisor : Prof. Dr. Ahmet SERTBAŞ**

**Co-Supervisor: Assist. Prof. Dr. Leyla TÜRKER ŞENER ]**

[Processing and analysing large medical data such as microarray, microRNA, dna, rna data is a challenging process. Bioinformatics is a multidisciplinary field of research that enables the processing of such medical data using computational technologies such as machine learning. Studies in the field of bioinformatics, where biology, mathematics and computer science work together, provide support in many areas such as disease diagnosis, preventive treatment, drug development and personalised treatment procedures. Bioinformatics methods are preferred to study the effect of microRNAs (also called miRNA, miR), which are found in various organisms and are thought to be associated with cancer, on genes. Microarray technology is one of the techniques used in the analysis of data such as microRNA, DNA, RNA. Many types of research such as the analysis of gene expression, the extraction of target genes, mutation analysis and the discovery of biomarkers are possible using the microarray method. Different methods can be used to prepare microarray data, and the way in which the data is analysed varies depending on these methods. In this study, biological data prepared using the microarray

technique are analysed and a study is conducted to extract target genes. For this purpose, a method using machine learning techniques is developed and applied to different datasets. In addition to two open access datasets, a unique dataset is also used in the study. For 11 experiments belonging to 3 different dataset groups, the process from preparing the data for processing in a computer environment to recognising the target genes is studied in detail. The study includes computerization of raw data, data quality checking, clustering, normalisation, creation of comparison groups and prediction of target genes. |

January 2025, [118] pages.

**Keywords:** [Microarray, gene expression analysis, target gene inference, microRNA, miRNA, machine learning]

## 1. GİRİŞ

Beslenme şekli, sigara kullanımı veya kullanılan ilaçlar gibi birçok etken hastalıkların seyrini değiştirebilmektedir. Sağlıklı bir organizmada belli bir maddenin olumlu/olumsuz etkilerini veya hasta bir organizmada uygulanan tedavinin etkilerini araştırmak zorlu bir süreçtir. Bu süreçte bilgisayar teknolojilerinden destek alarak ilerlemek tıbbi süreçleri değerlendirmeyi hızlandırmanın yanında daha doğru sonuçlar alınabilmesini de sağlamaktadır. Bilgisayar, matematik, istatistik bilimleri desteği ile tıbbi verilerin incelenmesi süreci olan biyoenformatik, tedavilerin genler üzerindeki etkisini incelemede de sıklıkla kullanılmaktadır. Canlı bir organizmada farklı tedavilerin etkilerini incelemek için kullanılacak yöntemlerden biri olan “gen ekspresyon analizi”, gen değişimlerini ortaya çıkararak biyolojik süreçleri anlamayı kolaylaştırır.

Gen değişimlerinin sebepleri araştırılırken sıklıkla karşılaşılan mikroRNA’lar, önemli gen düzenleyicileri olup farklı organizmalarda görülebilmektedir. mikroRNA’lar genleri baskılayarak hücrenin gelişim, çoğalma, ölüm gibi süreçlerini değiştirebilmektedir. Kodlanamayan RNA hücreleri olan mikroRNA’ların incelenmesinde; DNA, RNA’ların da incelenmesinde sıklıkla kullanılan mikrodizi yöntemi tercih edilebilmektedir. Mikrodiziler mikroRNA, DNA, RNA gibi büyük tıbbi verilerin bir çip üzerinde gösterilerek analiz edilebilmesine izin veren bir teknolojidir. Mikrodizi teknolojisi kullanılarak gen ifadesinde ortaya çıkan farklılıklar tespit edilebilmektedir. Biyoenformatik yöntemlerle daha doğru ve hızlı şekilde değerlendirilebilen mikrodizi verisi, tedavi süreçlerine bağlı olarak etkilenen genlerin tespit edilebilmesini ve buna bağlı olarak farklı tedavi/ilaçların geliştirilebilmesini sağlamada yardımcı olmaktadır.

Bu çalışmada meme kanseri hücre soyuna ait mikroRNA, kronik hipoksi tedavisi uygulanan fare beyni hücresi ve oksidatif strese maruz kalan fare nöronlarına ait mikrodizi verileri üzerinde gen ekspresyon analizi yapılmıştır. Çalışma sonucunda elde edilen çıktılardan hem gen ekspresyon analizi sürecinin anlaşılması hem de incelenen hastalık süreçlerine ait tedavi/ilaç geliştirme süreçlerine katkı sağlayacağı düşünülmektedir.

## 2. KAVRAMSAL ÇERÇEVE

Bu bölümde çalışma ilgili biyoenformatik, mikrorna, mikrodizi, gen ekspresyon analizi gibi kavramlara, verisetlerinin incelenmesine, literatür tarama bilgilerine ve hedef gen çıkarımı sürecinde kullanılan makine öğrenme yöntemlerine değinilmiştir.

### 2.1. MİKORNA

Biyoenformatik, biyolojik bilgilerin bilgisayar ile incelenmesi ve işlenmesidir. Bioinformatics kelimesi dilimize biyoenformatik, biyoformatik, biyobilişim, biyoteknoloji gibi birçok şekilde çevrilmekle birlikte, en yaygın kullanımı “biyoenformatik” şeklindedir. Biyoenformatik; biyoloji, moleküler biyoloji ve genetik, biyokimya, matematik, istatistik ve bilgisayar bilimleri gibi birçok alanın, DNA ve protein veritabanları üzerine odaklanması sonucu ortaya çıkan çok disiplinli bir araştırma alanıdır.

MikroRNA ilk olarak 1993'te keşfedilmiş bir RNA molekülüdür. RNA'lar protein kodlayan RNA'lar ve kodlamayan RNA'lar (ncRNA'lar) olarak ikiye ayrılabilir. mikroRNA'lar protein kodlamayan RNA sınıfındadır. Yaklaşık 21-23 nükleotid uzunluğunda olup, yine bir RNA türü olan mRNA ile etkileşime girerek protein oluşumunu engeller. İlk keşfedilen küçük RNA, miRNA'dır (Narry, 2005). Birçok farklı organizmada görülen mikroRNA'lar hücre üzerinde birçok olayda etkili olabilmektedir. MikroRNA(miRNA)'ların hedef genler üzerindeki etkisi ile hücrenin çoğalması, ölümü, değişmesi gibi süreçler etkilenebilmektedir. MikroRNA'lar önemli gen düzenleyicileridir. Yapılan çalışmalar mikroRNA'ların hastalıkların hem teşhis hem de tedavisinde etkili olduklarını göstermiştir.

MikroRNA'lar bir ya da birden fazla hedef geni baskılayarak hücrenin gelişim, farklılaşma, çoğalma, ölümü gibi farklı olaylarda rol oynayabilir. Yapılan çalışmalar miRNA'ların yeni bir onkogen veya tümör baskılayıcı gen sınıfı oluşturabileceğini göstermiştir. Normal ve patolojik dokular arasında farklı seviyede ifade edilen miRNA'lar tespit edilerek, insan kanserlerinde tanı ve tedavide etkili olabilecek yeni miRNA'lar belirlenebilir (Karagün ve diğ., 2014).

## 2.2. MİKRODİZİ

Moleküler biyoloji ile bilgisayar bilimlerinin bir arada kullanılması ile ortaya çıkan mikrodizi teknolojisi (bazı kaynaklarda gençip, biyoçip, mikrodizin olarak da adlandırılır), tüm genomun bir çip üzerinde görüntülenmesi ve gen ifadesindeki farklılıkların tespit edilmesinde kullanılan bir teknolojidir. DNA ve RNA çalışmalarında kullanılan mikrodizi teknolojisi, ekspresyon analizi denilen bir teknik kullanılarak proteine dönüştürülen ve dönüştürülemeyen RNA'ları tespit etmeyi mümkün kılmaktadır. Bu teknolojiye mikroskobik DNA bilgisi çiplere yerleştirilir ve floresan ışımaya benzer bir yöntem ile hangi genlerin etkilendiğinin tespiti yapılmaya çalışılır. Cam, plastik veya silikon gibi maddelerden hazırlanmış çiplere tutturulan DNA parçaları prob olarak adlandırılır ve yaklaşık 20-100 nükleotid uzunluğundadır. Gen ekspresyon analizi için tercih edilen yöntemlerden biri olan mikrodizi analizi, yüksek doğruluk ve hızda değerlendirme sağlayan bir tekniktir. Mikrodizi ile gen ekspresyon analizi, genetik hastalıklar, kanser araştırmaları, önleyici tedavi ve ilaç geliştirme süreçleri gibi birçok alanda yaygın olarak kullanılmaktadır (Sчена ve diğ., 1995). Mikrodizi teknolojisi, cam veya silikon bir çip üzerine gen veya DNA parçacağının birden çok replika içeren diziler halinde yerleştirilmesi ile hazırlanır. Bu diziler hedef genetik materyale özel şekilde tasarlanır. Bu yöntem ile gen ekspresyon analizi için aşağıdaki adımlar izlenir:

- 1.Laboratuvar ortamında örneklerin hazır hale getirilmesi. İncelenecek örnekten RNA izolasyonu sağlanarak cDNA'ya dönüştürülür.
- 2.Etiketleme: cDNA floresan boya ile etiketlenir.
- 3.Hibritizasyon: Etiketlenmiş cDNA, çip üzerindeki dizilere bağlanır. Çip üzerindeki her prob, spesifik bir geni hedef alır ve o genin ekspresyon seviyesini belirler.
- 4.Görüntüleme: Hibritizasyona uğramış örnekler okunarak görüntülenir.
- 5.Analiz: Her bir gen için gen ekspresyon seviyesi belirlenir ve biyoenformatik yöntemlerle analiz edilir.

Mikrodizi süreci için çipleri hazırlayan Illumina (<https://www.illumina.com>, Erişim Tarihi: 22.11.2024), Affymetrix (<https://www.thermofisher.com/tr/en/home/life-science/microarray-analysis.html>, Erişim Tarihi: 22.11.2024.), Agilent (<https://www.agilent.com/en/product/software-informatics/genomics-software-informatics/gene-expression/genespring-gx>, Erişim Tarihi: 22.11.2024) gibi birçok firma bulunmaktadır. Bu tez çalışmasında kullanılan veri setinin oluşturulması sırasında Illumina tarafından hazırlanmış çipler kullanılmıştır.

Mikrodiziler, genellikle gen anlatım farklarını (sağlıklı ve hastalıklı doku, kanserli hücre ve ilaç uygulanmış kanserli hücre gibi başlangıç durumu ve deney grubu arasındaki farkları) tespit edebilmek amacı ile tüm genlerin eşzamanlı anlatımlarını belirlemek için kullanılmaktadır (Özkan ve Erol, 2017). Tez çalışmasında kullanılan özgün verisetinde MCF-7 ve MDA-N1 kanser hücreleri ile bu kanser hücrelerine nikotin, ışın, metformin gibi maddelerin uygulandığı hücreler karşılaştırılmıştır. İkinci verisetinde fare beyin hücresine kronik hipoksi tedavisi uygulanan mikrodizi örnekleri karşılaştırılmıştır. Son verisetinde ise oksidatif strese maruz kalan fare nöronlarını içeren mikrodizi verisi incelenmiştir.

### 2.2.1. Gen Ekspresyon Analizi

Biyolojik süreçleri anlamak için kritik bir önemi olan gen ekspresyon analizi, belirli genlerin transkripsiyon sürecinde ne kadar aktif olduğunu incelemektedir. Bu analiz, özellikle hastalık mekanizmaları, biyomarker keşfi ve kişiselleştirilmiş tıp alanında geniş bir uygulama alanına sahiptir. Biyolojik süreçleri anlamak için kullanılan yöntemlerden biri olan gen ekspresyon analizi, bir hücrede veya dokuda genlerin ne kadar baskın olduğunu tespit etmek için kullanılır. Biyomedikal araştırmalarda yaygın olarak kullanılan bu teknoloji, ilaç geliştirme, biyomarker keşfi, önleyici tedavi süreçleri, kanser gibi hastalıkların mekanizmalarının anlaşılması gibi birçok alanda tercih edilmektedir (Golub ve diğ., 1999). Gen ekspresyon analizinin başlıca kullanım alanları aşağıdaki şekildedir:

1. Hastalık Araştırmaları: Kanser veya nörodejeneratif (Parkinson, Alzheimer, MS gibi) hastalıkların incelenmesinde bu yöntem yaygın olarak kullanılmaktadır (Hanahan ve diğ., 2011). Normal hücreler ile hastalıklı hücrelerin karşılaştırılması ile değişen gen ekspresyon profilleri incelenir. Böylece hem hastalığın mekanizmasının anlaşılması hem de tedavi geliştirme süreçlerine katkı sağlanabilmektedir.
2. Kişiselleştirilmiş Tıp: Farklı bireylerde gen ekspresyon farklılıkları, kişiselleştirilmiş ilaç/tedavi geliştirilmesine olanak sağlar. Bireylerin genetik yapısına bağlı olarak ilaçlara verdiği tepki değişebilmektedir (Collins ve diğ., 2015). Gen ekspresyon analizi ile ilaç yan etkilerinin azaltılması veya daha etkin tedavi sağlamak mümkün olabilmektedir.
3. Biyomarker Keşfi: Gen ekspresyon analizi hastalıklara özel biyomarker keşfi için kullanılabilir. Bazı kanser hücrelerinde belirgin şekilde ifade edilen genler veya kan örneklerinde tespit edilen gen ekspresyonları dikkate alınarak, tanı ve tedavi süreçlerine katkı sağlayabilecek biyomarkerler tespit edilebilir.

Gen ekspresyon analizini gerçekleştirmek için farklı yöntemler kullanılabilir. En yaygın yöntemler aşağıdaki gibidir:

1. Mikrodizi Analizi: Mikrodiziler bir canlıya ait tüm DNA bilgisinin (genom) bir çip üzerinde görüntülenmesini sağlayarak gen ifadesindeki farklılıkları ortaya çıkarmayı sağlayan bir teknolojidir (Çavdar ve diğ., 2022). Mikrodizi verileri yaygın olarak biyoenformatik teknikleri ile analiz edilerek biyolojik süreçlerdeki farklılıklar değerlendirilir.
2. RNA Sekanslama: Hücre veya dokudaki transkriptomu (belli bir durumdaki tüm RNA moleküllerinin toplamı) inceleyerek, hangi genlerin ne derece aktifleştüğünü ölçer (Wang ve diğ., 2009).
3. qPCR: İncelenen örnekteki RNA'nın veya genin miktarını hassas ve kesin şekilde ölçmeyi sağlar. "Gerçek Zamanlı PCR" olarak da adlandırılan yöntem, gen ekspresyon analizinde yaygın olarak kullanılmaktadır (Bustin ve diğ., 2000).

### 2.2.2. Mikrodizi Hastalık Veritabanları

Mikrodizi verilerinin paylaşıldığı çeşitli veritabanları bulunmaktadır. Bu veritabanlarında MIAME (Minimum Information About a Microarray Experiment) kriterlerine uygun veri setleri paylaşılmaktadır (Öztemur ve diğ., 2014., Brazma ve diğ., 2001).

GEO veritabanı başta olmak üzere mikrodizi veriseti paylaşan birçok veritabanı bulunmaktadır (Çavdar ve diğ., 2022). Açık erişim olan veritabanlarından birkaç tanesi aşağıdaki gibidir:

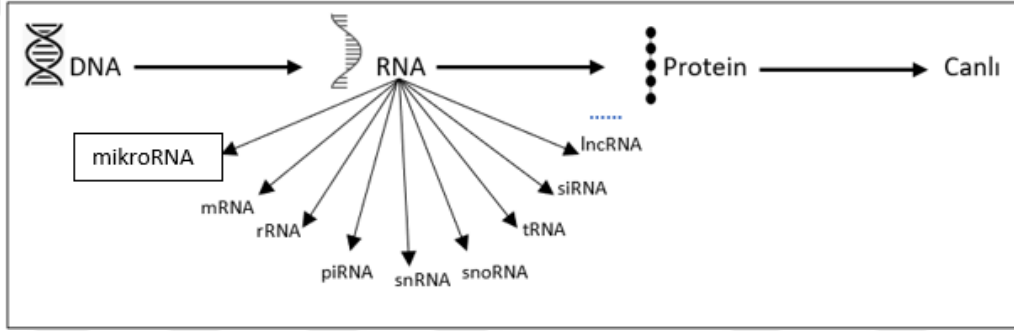
- NCBI altındaki GEO (<https://www.ncbi.nlm.nih.gov/geo/>, Erişim Tarihi: 22.11.2024).
- ArrayExpress (<https://www.ebi.ac.uk/arrayexpress/>, Erişim Tarihi: 22.11.2024).
- DataMed (<https://datamed.org/>, Erişim Tarihi: 22.11.2024).

MikroRNA verilerinin de incelendiği tez çalışmasında biyoenformatiğin "veri madenciliği ve bilgi çıkarımı" alanında çalışılarak hedef çıkarımı yapılmıştır.

MikroRNA mikrodizi verileri için özgün veriseti kullanılırken diğer veri setleri NCBI altından GEO'dan erişilerek kullanılmıştır. Çalışma; verisetlerinin incelenmesi, benzer çalışmaları ve yöntemleri inceleyen literatür taraması, veri setlerinin makine öğrenmesi yöntemleri ile değerlendirilmesi, hedef gen çıkarımlarının yapılması bölümlerinden oluşmaktadır.

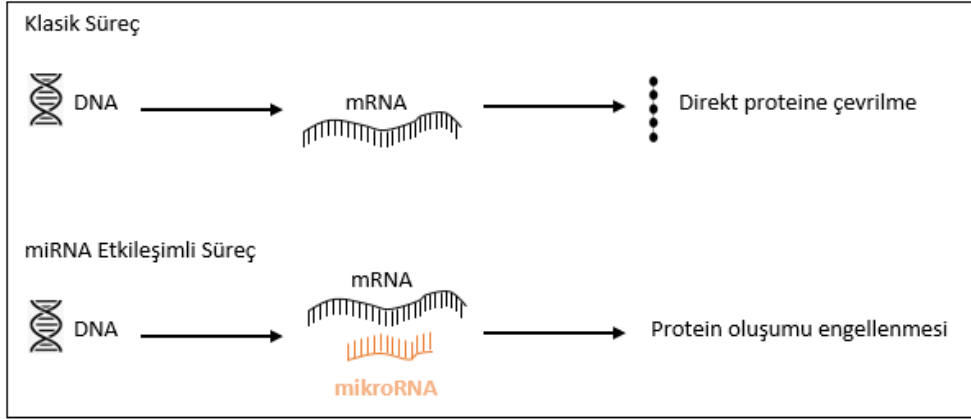
### 2.2.3. Mikrodizi Veriseti Hazırlama Süreci

DNA; tüm canlılarda bulunan ve kalıtsallığı sağlayan moleküldür, genetik bilgiyi taşır. DNA nükleotid denilen 4 değişik yapı taşından oluşur: A (Adenin), T (Timin), G (Guanin) ve C (Sitozin). Bir insan genomunda yaklaşık 3 milyar nükleotid vardır ve bu verinin elle yazılması/işlenmesi mümkün olmadığından biyoenformatiğe ihtiyaç duyulmaktadır. RNA; DNA'ya benzer şekilde nükleotidlerden oluşan bir yapıdır. DNA'daki genetik bilginin proteine dönüştürülmesinde rol oynar. Birden çok RNA çeşidi bulunmaktadır. DNA genetik bilgiyi taşırken, RNA ise DNA tarafından taşınan bilginin kullanılmasını sağlar. DNA'daki verinin RNA aracılığı ile taşınması, ardından proteine ve canlıya dönüşmesi sürecinde, yolunda gitmeyen durumlar olup RNA'nın proteine dönüşemeyebilmektedir. Bu duruma bir RNA çeşidi olan, mikroRNA adı verilen moleküllerin sebep olduğu anlaşılmıştır.



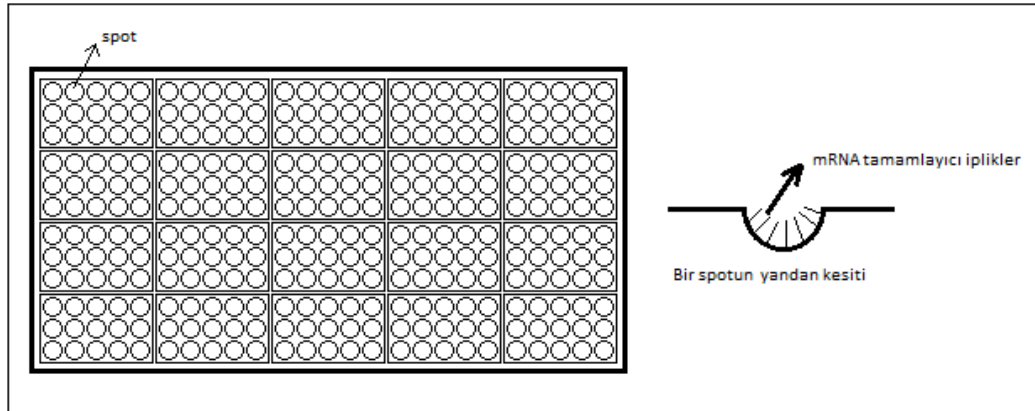
Şekil 2.1: DNA – RNA – Protein – Canlı süreci

MikroRNA'lar kanserde biyobelirteçler olarak potansiyeli en çok incelenen moleküller arasındadır (Sherafatian, 2018). MikroRNA'lar farklılaşma, metabolizma, gelişme ve hücrel sinyalizasyon gibi birçok biyolojik süreçte önemli düzenleyici roller oynarlar. Bu nedenle, gen hedeflerinin tanımlanması mikroRNA'ların fonksiyonel karakterizasyonu için önemlidir (Sticht ve diğ., 2018). MikroRNA'ların işlev bozuklukları hedeflerini bozabilir ve bu nedenle kanser, nörolojik bozukluklar, diyabet ve kardiyovasküler hastalıklar gibi birçok hastalığa yol açabilir. Son zamanlarda mikroRNA-hastalık ilişkisi tahmini için hesaplama yöntemleri önerilmiştir. Bu nedenle mikroRNA hedeflerinin doğru tahmin edilmesi, karmaşık hastalık mekanizmasının anlaşılmasına katkıda bulunabilir (Jiang ve diğ., 2018).



Şekil 2.2: MikroRNA

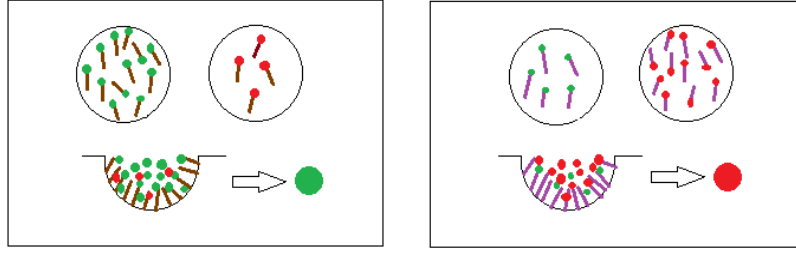
Bu tez çalışmasında kullanılan özgün veriseti mikroRNA hücre soylarından oluşan mikrodizi verisetidir. Mikrodizi verisi hazırlanırken çipler kullanılmaktadır. Bir çipte 6000 kadar spot bulunabilir. Her spot farklı bir geni temsil eder. Tez çalışmasında kullanılan verisetinin hazırlanması sırasında 47323 adet proba çalışmaya başlanmıştır. Bu problemler ekspresyon seviyelerine göre filtrelenmiştir. Normalize datalardan %20'nin altında ışığa alınan problemler filtreye takılmış ve çalışmaya 47288 proba devam edilmiştir.



Şekil 2.3: Mikrodizi çip yapısı

Teze konu olan özgün veriseti 9 farklı deneyden üretilen mikrodizi verilerinden oluşmaktadır. Örneğin deneylerden MCF-7 kanser hücre soyu içeren kontrol örneği ve MCF-7 hücresine Nikotin uygulanmış iki örneğin, spotlarda bulunan A ve B genleri için değerlendirilmesi için temsili bir gösterim yapılmak istenirse aşağıdaki gibi bir süreç gelişecektir. A geninin kahverengi, B geninin ise mor ile gösterildiği, kontrol örneğinin (MCF-7 kanser hücresi) yeşil, test örneğinin (MCF-7 kanser hücresine Nikotin uygulanmış hücre) kırmızı ile işaretlendiği kabul edilirse; Kontrol örneklerinin bulunduğu hücrede A geni daha fazla olduğu varsayılır, spot tarandığında yeşil renk daha baskın olduğu için yeşil bu spot bilgisayar ortamına yeşil olarak aktarılmıştır. Nikotin uygulanmış örneklerin bulunduğu

hücrede B geni daha fazla olduğu varsayılır, spot tarandığında kırmızı renk daha baskın olduğu için bu spot bilgisayar ortamına kırmızı olarak aktarılır.



**Şekil 2.4:** Özgün veriseti hazırlık süreci: floresan boyama ve tarama süreci

Burada nikotin uygulanan renkler daha baskın çıkıyorsa, gen yukarı düzenleniyor (up regulation) yani protein kodlanıyor, kontrol grubu renkleri daha baskın çıkıyorsa gen aşağı düzenleniyor (down regulation) yani protein kodlanmıyor, renk değişimi olmuyorsa (renk sayıları neredeyse aynı ise) gen değişmiyor şeklinde yorumlanabilir.

Tarama sonucunda bilgisayara aktarılan sonuçlarda kullanılan renkler aşağıdaki gibidir:

**Yeşil:** MCF-7 veya MDA-N1 kanser hücre soyuna ait miRNA'lardan etkilenen gen.

**Kırmızı:** MCF-7, MDA-N1 hücrelerine nikotin, Işın, Metformin vs maddelerinin uygulandığı örneklerden hibridize olan gen.

**Sarı:** Eşit hibridize olan gen.

**Siyah:** Hem kontrol hem örnekten etkilenmeyen gen.

#### 2.2.4. Mikrodizi Verisetleri

Bu bölümde tez çalışmasında kullanılan özgün veriseti dışında kullanılan mikrodizi verisetleri ile ilgili bilgi paylaşılmaktadır.

##### 2.2.4.1. GSE10470 Veriseti

Fare nöronlarından hazırlanan mikrodizi veriseti, Illumina BeadStudio türünde olup GSE10470 kodu ile GEO ortamından erişilmiştir. Veriseti HCA'nın gen ekspresyonunda sebep olduğu değişimi incelemektedir. Bu yapılırken ATF4 geninin süreci ne yönde etkilediği tespit edilmeye çalışılmıştır. Verisetinde ATF4 geni olmayan ve sağlıklı iki farklı tür fareden alınan örneklere HCA uygulanmış ve kontrol grupları ile HCA uygulanan grupların karşılaştırılması sağlanmıştır. GSE10470 verisetinde 24 örnek bulunmaktadır:

- Wild-type HCA: 6 adet örnek içerir. HCA uygulanmış normal farelerden alınan nöronlardan oluşur.
- Knock-out ATF4: 6 adet örnek içerir. HCA uygulanmış, ATF4 geni olmayan farelerden alınan nöronlardan oluşur.

- Wild-type Kontrol: 6 adet örnek içerir. HCA uygulanmamış, normal fare embriyonik nöronlarını içeren kontrol grubudur.
- Knock-out Kontrol: 6 adet örnek içerir. ATF4 geni olmayan ve HCA uygulanmamış fare embriyonik nöronlarını içeren kontrol grubudur.

#### **2.2.4.2. GSE10470 Verisetini Kullanan Çalışmalar**

Bu verisetini kullanan akademik yayınlar incelendiğinde literatürde 4 çalışma ile karşılaşılmıştır. Bu çalışmalardan ilki Lange ve diğ. (2008) yaptığı çalışmadır. Çalışma ilgili verisetinin laboratuvar ortamında hazırlanmasını da içeren deneylerinin sonucunu kapsamaktadır. ATF4'ün oksidatif stres kaynaklı rahatsızlıklarda etkisini ölçmek için yapılan çalışmada, denek fareleri üzerinde glutasyon tükenmesi gibi stress faktörlerini ölçmek için alınan örnekler analiz edilmiştir. Çalışma ATF4'ün oksidatif strese bağlı hastalıkların tedavisinde etkili olabileceğini belirtmektedir. Bu deneysel çalışma sürecinde hazırlanan veri GSE10470 kodu ile GEO üzerinden paylaşılmıştır.

Woo ve diğ. (2021) yaptığı çalışmada ise GRM8 proteininin merkezi sinir sistemindeki nöroprotektif etkisi araştırılmıştır. MS gibi nörodejeneratif hastalıklarda glutamat seviyelerinin artışının nöron ölümüne sebep olduğu vurgulanan çalışmada, GRM8'in glutamat seviyesini kontrol altına alarak nöronları koruyabilme potansiyelinin olduğu belirtilmiştir.

Diğer bir çalışma Bloomingdale ve diğ. (2022) tarafından yapılmıştır. Çalışmada GSE10470 veriseti de kullanılmakta ve bortozomib ilacının neden olduğu nöropatinin tedavi süreci için seçenekler araştırılmaktadır. Dexanabinol tedavisinin uygulandığı çalışmada, dexanabinolün kemoterapiye bağlı nöropatik ağrının tedavisinde iyileştirici bir potansiyeli olduğu ifade edilmektedir.

Bu tez çalışmasında geliştirilen SMATP yöntemi ile yapılan yayın çalışmasında (Çavdar ve diğ., 2024), hazırlanan makine öğrenmesi tabanlı yöntemin doğruluğunu vurgulamak için GSE8262 veriseti ile birlikte bu veriseti de kullanılmış ve en çok etkilenen genler çıkarılmaya çalışılmıştır.

#### **2.2.4.3. GSE8262 Veriseti**

Bu verisetinde fare beyninin kortikal ve hipokampal bölgelerinden alınan örneklerle hazırlanan mikrodizi verisi üzerinde hipoksi etkisi incelenmiştir. Illumina BeadStudio türünde olan açık erişimli verisetine GEO platformundan GSE8262 kodu ile erişim sağlanmıştır. Deneysel uygulanan tedavi modelleri:

- CIH: Aralıklı düşük oksijen seviyelerine maruz bırakılma işlemi.
- CCH: Sürekli düşük oksijen seviyelerine maruz bırakılma işlemi.

GSE8261 verisetinde toplam 18 örnek bulunmaktadır:

- CIH grubu: 6 adet örnek içerir. Kronik aralıklı hipoksi tedavisi uygulanmış farelerden alınan kortikal ve hipokampal örneklerden oluşur.
- CCH grubu: 6 adet örnek içerir. Kronik sürekli hipoksi tedavisi uygulanmış farelerden alınan kortikal ve hipokampal örneklerden oluşur.
- Kontrol grubu: 6 adet örnek içerir. Tedavi uygulanmamış farelerden alınan 3 adet kortikal ve 3 adet hipokampal örnekten oluşur.

#### **2.2.4.4. GSE8262 Verisetini Kullanan Çalışmalar**

GSE8262 veriseti 4 farklı çalışmada kullanılmıştır. Bu çalışmalardan Zhou ve diğ. (2008) yaptığı çalışma verisetinin laboratuvar ortamında hazırlandığı ve ardından değerlendirildiği çalışmadır. Yapılan bu deneysel çalışmada üretilen veri GEO üzerinden GSE8262 kodu ile erişime açılmıştır. Çalışma deney farelerinin seçiminden başlayarak, beynin kortikal ve hipokampal bölgelerinden örnekler alınması, örneklere aralıklı ve sürekli hipoksinin uygulanması, deney sonuçlarının uygun araçlar ve Illumina BeadStudio yardımı ile bilgisayar ortamına aktarılması ve gen ekspresyon değişiminin incelenmesi süreçlerini kapsamaktadır.

Bu verisetinin dahil edildiği ikinci çalışma Wang ve diğ. (2021) yaptığı SENP1 etkisi ile ilgili çalışmadır. Çalışmada hipoksinin SENP1 proteini üzerindeki etkisi incelenmiştir. Mikroglia'da gözlemlenen değişikliklerinin değerlendirildiği çalışmada, SENP1 proteinin hipoksiye bağlı sorunların tedavisinde etkili olabileceği ispatlanmaya çalışılmıştır.

GSE8262 verisetini kullanan diğer bir çalışma Wang ve diğ. (2023) tarafından yapılmıştır. Çalışmada SENP1 proteininin uyku apnesi ile ilişki hipoksinin sebep olduğu hasar üzerindeki etkisi incelenmiştir. Çalışmada SENP1'in az veya çok kullanıldığı durumlar karşılaştırılmıştır. Çalışma SENP1 proteininin tedaviyi olumlu etkileyebileceği teorisini sunmaktadır.

Bu veriseti ile ilgili yayınlanmış olan son akademik yayın tez çalışması kapsamında uygulanan makine öğrenmesi tabanlı hedef gen çıkarımı çalışmasını içeren çalışmadır (Çavdar ve diğ., 2024). Yayında bu verisetinin değerlendirilmesi sürecince makine öğrenmesi yöntemlerinin nasıl kullanılabileceği ele alınmakta ve sonuç gen listesi paylaşılmaktadır.

### 2.3. MAKİNE ÖĞRENMESİ VE BİYOENFORMATİK

Gen ekspresyon analizi, hastalık sebeplerinin anlaşılması, koruyucu-tedavi edici yöntemlerin geliştirilmesi, ilaç geliştirme süreçleri gibi birçok alanda önemli çıktılar sağlamaktadır. Gen ekspresyon verisi gibi büyük ve karmaşık verilerin analizi için makine öğrenmesi yöntemlerinin kullanılması doğru ve hızlı sonuç alabilmeyi sağlaması açısından önemli bir yere sahiptir. Bu bölümde gen ekspresyon analizi sürecinde kullanılan teknik ve yöntemlerin temel kavramlarına ve kullanım örneklerine yer verilmiştir.

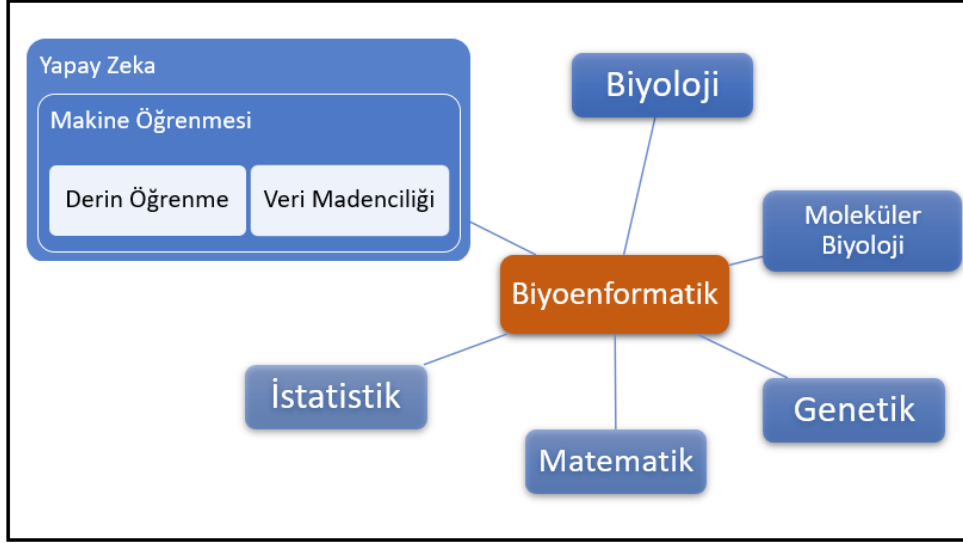
Yapay Zeka (AI) bilgisayarın insan zekasını taklit ederek öğrenmesi, karar vermesi, problem çözmesi gibi işleri yapabilmesini hedefleyen bir disiplindir. Gen ekspresyon analizi sürecinde yapay zeka, büyük ve karmaşık verilerin değerlendirilmesi, kanser gibi hastalıkların tespiti, bağışıklık sisteminin anlaşılması gibi birçok alanda kullanılmaktadır. Yapay zekanın evrimsel sinir ağı (CNN), tekrarlı sinir ağları (RNN) gibi derin öğrenme yöntemleri bu tür büyük ve karmaşık verilerin değerlendirilmesinde yüksek doğruluk oranının yanında hız da sağlamaktadır (LeCun ve diğ., 2015).

Yapay zekanın bir alt dalı olan makine öğrenmesi (ML) kendisine sağlanan veriden öğrenerek tahminde ve çıkarımda bulunma yeteğine sahiptir. Makine öğrenmesi kendisine verilen verilerden ortaya çıkardığı örüntü ve desenleri kullanarak yeni veriler üzerinde tahminde bulunmayı hedefler. Büyük veriler üzerinde verimli bir şekilde çalışmayı mümkün kılan bu disiplin gen ekspresyon analizi sürecinde de tercih edilmektedir.

Makine öğrenmesi teknikleri ile çalışan veri madenciliği, büyük verilerin değerlendirilmesinde, örüntü ve desenlerin tespitinde, veriler arasındaki ilişkinin ortaya çıkarılmasında etkili bir yöntemdir. Bu yöntem verideki örüntüleri bularak çıkarım ve tahmin yapmak için çeşitli algoritmalar kullanır. Veri madenciliği istatistik, matematik, makine öğrenmesi, yapay zeka teknikleri gibi disiplinlerden destek alarak verinin analizini ve etkili olarak yorumlanabilmesini sağlamaktadır (Han ve diğ., 2012).

Makine öğrenmesinin başka bir alt dalı olan derin öğrenme, yapay sinir ağlarını farklı yöntemlerle kullanarak karmaşık örüntüleri öğrenmeyi amaçlar. Özellikle büyük ve karmaşık verisetlerinde etkili bir yöntem olup biyoenformatik çalışmalar dışında doğal dil işleme, görüntü işleme gibi birçok alanda başarılı sonuçlar vermektedir (LeCun ve diğ., 2015).

Bu tez çalışmasında makine öğrenmesinin veri madenciliği yöntemleri kullanılarak hedef gen çıkarımı gerçekleştirilmiştir.



Şekil 2.5: Makine öğrenmesi ve biyoenformatik

### 2.3.1. Makine Öğrenmesi Çeşitleri

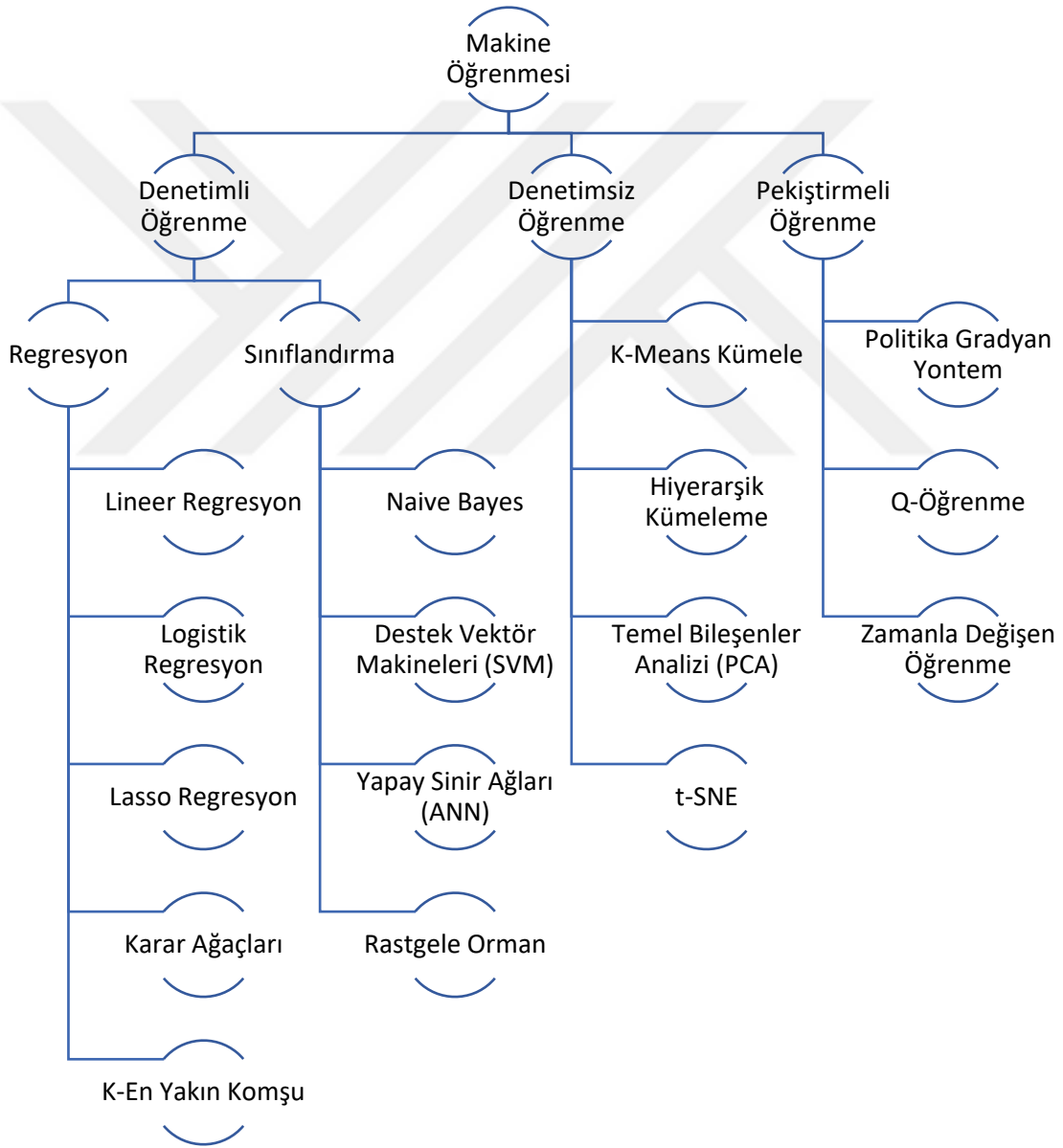
Gen ekspresyon analizi, hedef gen çıkarımı gibi süreçlerde yaygın olarak kullanılan makine öğrenmesi; denetimli, denetimsiz ve pekiştirmeli olarak 3 kategoriye ayrılabilir.

- Denetimli Öğrenme (Supervised Learning): Denetimli öğrenme, girdi ve çıktılarının (eğitim verisi) verilerek sistemin öğrenmesinin sağlandığı ve buradan sağlanan öğrenme ile çıktıları olmayan yeni veriler üzerinde sonuçları tahmin etmeyi sağlayan yaklaşımdır. Eğitim sürecinde verilen her giriş verisinin bir çıktısı (etiket olarak da isimlendirilir) vardır ve model buradan öğrenir. Test verisi ile bu öğrenmenin başarısı test edilir ve süreç ihtiyaca göre tekrar uygulanır. Öğrenme sağlandıktan sonra bu yapı yeni gelen veriler üzerinde çıktıları tahmin edebilmektedir. Karar ağaçları, destek vektör makineleri, naive bayes, logistik regresyon yaygın olarak kullanılan denetimli makine öğrenmesi yöntemleridir (Lavine ve diğ., 2004).
- Denetimsiz Öğrenme (UnSupervised Learning): Öğrenme sürecinde çıktıların (etiketlerin) bulunmadığı öğrenme türüdür. Veriler arasındaki benzerlikler keşfedilerek öğrenme sağlanır. Verideki örüntü ve desenler bulunarak öğrenme süreci sağlanır. Bu yöntem gen ekspresyon verilerinin keşfinde tercih edilen yöntemler arasındadır. K-means, hiyerarşik gibi kümeleme algoritmaları bu yöntemde yaygın olarak kullanılmaktadır (Lain, 2010).
- Pekiştirmeli Öğrenme (Reinforcement Learning): Diğer öğrenme türlerine göre daha az bilinen/kullanılan pekiştirmeli öğrenme, sistemin çevresi ile etkileşime girip, aldığı geribildirimlerden öğrenmesi yöntemi ile çalışır. Deneme-yanılma ile en iyi sonuca ulaşmayı hedefleyen bu yöntem özellikle robotik ve oyun geliştirme süreçlerinde tercih

edilmektedir. Gen ekspresyon analizi sürecinde pek sık kullanılsa da bu alanda faydalı alternatifler sunabilecek bir yöntem olarak değerlendirilebilir (Sutton ve Barto, 2018).

### 2.3.2. Makine Öğrenmesi Yöntemleri

Makine öğrenmesi sürecinde kullanılacak çeşitli yöntemler/algoritmalar bulunmaktadır. Üzerinde çalışılmak istenen verinin içeriğine ve beklenen çıktılara göre farklı yöntemler daha etkili sonuçlar verebilmektedir. En bilinen makine öğrenmesi yöntemleri Şekil 2.6'da gösterilmiştir.



Şekil 2.6: Makine Öğrenmesi Yöntemleri

Gen ekspresyon analizi süreçlerinde sıklıkla tercih edilen makine öğrenmesi yöntemleri aşağıdaki gibidir:

- Karar Ağaçları (Decision Trees): Veriyi bir ağaç yapısında sınıflandıran bir yöntemdir. Hiyerarşik yapıda bir düğüm (node) bir özelliği gösterir, dallar özelliğin farklı değerlerini gösterir, yaprak düğümler ise tahmin-sonuçları gösterir. Gen ekspresyonu gibi biyolojik verilerde yaygın olarak kullanılan bir yöntemdir (Chen ve diğ., 2014-A)
- Destek Vektör Makineleri (SVM): Verilerli belirli sınıflara ayırmak için kullanılan bir algoritmadır. Sınıflandırma ve regresyon problemleri için etkili bir yöntemdir. Doğrusal olmayan verileri işleyebilmesi ve büyük verilerdeki performansı, kanser gibi hastalıkların sınıflandırılması gibi genetik çalışmalarda tercih edilmesini sağlamaktadır (Huang ve diğ., 2018).
- K-En Yakın Komşu (KNN): Verileri sınıflandırmak için mesafayı kullanarak, en yakın komşuları bulur. Küçük verisetlerinde etkili bir yöntemdir. Gen ekspresyon deseninin bulunmasında kullanılan yöntemlerdendir.
- Naive Bayes: Bayes teoremini kullanan, istatistiksel bir yöntemdir. Bu sınıflandırma yöntemi, metin sınıflandırmada olduğu gibi genetik veri analizinde de etkili bir yöntemdir (Chandra ve Gupta, 2011).
- Yapay Sinir Ağları (ANN): İnsan beyin hücrelerindeki sinir ağlarının çalışmasını taklit eden bu yöntem, özellikle derin öğrenme tekniklerinde kullanılmaktadır. Bu matematiksel model, büyük ve karmaşık verisetlerinin değerlendirilmesinde tercih edilmektedir. ANN, gen ekspresyon analizi ve hedef tahmini süreçlerinde tercih edilen yöntemlerdendir (Chen ve diğ., 2014-B).

Bu tez çalışmasında kullanılan normalizasyon, kümeleme ve modelleme gibi teknikler makine öğrenmesi süreçlerinin temel bileşenlerini oluşturmaktadır. Makine öğrenmesi süreçlerinde veri ön işleme (normalizasyon, ölçeklendirme gibi) model performansını doğrudan etkileyen önemli bir adımdır. Gen ekspresyon verilerindeki teknik varyasyonları en aza indirmek ve verisetlerini karşılaştırılabilir hale getirmek için uygulanan ön işlemlerden biri de normalizasyondur. Bu çalışmada geliştirilen SMATP yönteminde niceliksel normalizasyon kullanılmıştır. Makine öğrenmesinde kümeleme, verisetindeki benzerliklerin ortaya çıkarılması için kullanılmaktadır. Pearson korelasyon katsayısı, iki genin ekspresyon profilleri arasındaki doğrusal ilişkiyi ölçer. Bu istatistiksel metrik, kümeleme algoritmalarında benzerlik ölçütü olarak kullanılır. Özellikle hiyerarşik kümeleme veya k-means gibi denetimsiz öğrenme

yöntemlerinde, genler veya örnekler arasındaki benzerlikler pearson korelasyonuna dayalı olarak hesaplanabilmektedir. Geliştirilen SMATP yönteminde pearson korelasyona dayalı kümeleme işlemi yapılmıştır. Makine öğrenmesi özellik seçimi süreçlerinde eBayes istatistiği tercih edilen yöntemlerdendir. Bu yaklaşım, destek vektör makineleri, rastgele orman ve derin öğrenme gibi yöntemlerde, veri gürültüsünü azaltarak modelin genelleme kapasitesini artırır. eBayes yöntemi, gen ekspresyon analizinde diferansiyel gen ekspresyonunu belirlemek için kullanılan bir istatistiksel yaklaşımdır. Tez çalışmasında geliştirilen SMATP yönteminde eBayes istatistiği kullanılmıştır.



### 3. YÖNTEM

Bu bölümde mikrodizi verileri üzerinde hedef gen çıkarımı için geliştirilen yöntem ve özgün verisetinin üzerinde çalışmaya uygun hale getirilmesi hakkında detaylı bilgi verilmektedir. İlk bölümde mikrodizi verileri ile hedef gen çıkarımı için geliştirilen yöntem detaylandırılarak açıklanmaktadır. İkinci bölümde ise özgün verisetinin kullanıma hazır hale getirilmesi için izlenen adımlar detaylandırılmaktadır.

#### 3.1. ÖNERİLEN HEDEF TAHMİN YÖNTEMİ

Mikrodizi verilerinin makine öğrenmesi yöntemlerinden olan veri madenciliği yöntemleri ile incelenerek hedef gen çıkarımı yapılan bu tez çalışmasında, verilerin değerlendirilmesi için SMATP adı verilen bir yöntem geliştirilmiştir. Bu yöntem R programlama dili (<https://www.r-project.org/>, Erişim Tarihi: 22.05.2024) ortamında Bioconductor (<https://www.bioconductor.org/>, Erişim Tarihi: 22.05.2024) paketi desteği alınarak hazırlanmıştır. SMATP yöntemi adımları Şekil 3.1’te gösterildiği gibidir. Illumina BeadStudio/GenomeStudio ile hazırlanan mikrodizi verilerini değerlendirmek için verisetinde ön hazırlık yapılarak üzerinde çalışmaya uygun hale getirilmesi, görselleştirmeler desteği ile veri kalitesi kontrollerinin yapılarak verinin üzerinde çalışmaya uygun olup olmadığının değerlendirilmesi, kontrol ve deney gruplarının kümelenmesi, verinin normalize edilmesi, karşılaştırma gruplarının hazırlanması ve hedef gen çıkarımı süreçleri gerçekleştirilmektedir.



Şekil 3.1: SMATP Yöntemi

##### 3.1.1. Veriseti Ön Hazırlık

Gen ekspresyon analizi için geliştirilen SMATP metodunun ilk aşaması ham verisetinin kullanılma hazır hale getirilmesi sürecini içermektedir. Bu aşamada verisetinin kodlama

ortamına aktarımı, verinin tanımlama süreci için veri çerçevesi oluşturulması süreçlerini içermektedir. Illumina BeadStudio formatındaki mikrodizi verilerinin analizi için bu aşamada düz metin formatındaki ham veri alındıktan sonra bir veri çerçevesi hazırlanmaktadır. Hazırlanan veri çerçevesi verinin başlıklandırılmasını sağlamaktadır. Bu veri çerçevesinde bulunan başlıklar Tablo 3.1'deki gösterilmektedir.

**Tablo 3.1:** Veri Çerçevesi

Veri Çerçevesi Başlığı	Açıklama
<b>Sıra (Array)</b>	Verisetindeki örneğin sıra numarası bilgisidir.
<b>Grup (Batch)</b>	Verinin hangi grupta işleme alındığı bilgisidir.
<b>GenTipi (Genotype)</b>	Deneyde kullanılan gen tipi bilgisidir.
<b>Replika (Replicate)</b>	Verisetindeki örneğin hangi replikada bulunduğu bilgisidir.
<b>Tedavi (Drug)</b>	Uygulanan tedavi bilgisidir.
<b>DiziKodu (ArrayCode)</b>	Ham veri örnek kodu bilgisidir.

### 3.1.2. Veri Kalitesi Kontrolleri

Geliştirilen yöntemin ikinci aşaması bilgisayar ortamına aktarılan ve başlıklandırılan veri üzerinde kalite kontrollerinin yapılması aşamasıdır. Bu aşamada aşağıdaki adımlar görselleştirmeler desteği ile uygulanmakta ve verinin analize uygunluğu kontrol edilmektedir. Veri kalitesi kontrol aşamaları aşağıdaki gibidir:

- Mevcut verideki örneklerin sıralanması: doğru kontrollerin yapılabilmesi için tüm replikaları içeren deneyin sıralanması süreci.
- Tasarım matrisinin hazırlanması: kontrol/deney gruplarının bir matris halinde tasarlanması süreci. Doğrusal modelleme ile bir tasarım matrisi hazırlanmaktadır. Doğrusal modelleme, gen ekspresyon seviyelerini belirli deneysel gruplar arasında karşılaştırmayı sağlar. Bunun için bir tasarım matrisi oluşturulur. Tasarım matrisi, deney gruplarının ve koşullarının modelde nasıl gösterileceğini ifade eder.

$$T = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ \vdots & \vdots & \vdots \end{bmatrix} \quad (3.1)$$

Burada her satır bir örneği gösterir. Her sütun bir deneysel grup/koşulu gösterir. 1 değeri örneğin ilgili gruba ait olduğunu belirtir. Örneğin “kontrol”, “tedavi1”, “tedavi2” şeklinde bir deneyde ilk sütun “kontrol”, ikinci sütun “tedavi1”, üçüncü sütun “tedavi2” anlamına gelir.

- Normalizasyon öncesi sinyal dağılım durumu kontrolü: sinyal dağılımının kutu grafik (boxplot) ile görselleştirilmesi ve sinyal dağılım durumunun değerlendirilmesi süreci.
- Saptama skorlarının kontrolü: beadstudio türündeki mikrodizi verisindeki genlerin 0-1 arasında ifade edilme olasılığının (0: ifade edilmiyor, 1: ifade ediliyor) kontrolü. Saptama skorlarının kutu grafik (boxplot) ile görselleştirilerek değerlendirilmesi süreci.

### 3.1.3. Kümeleme

Mikrodizi analizinde kümeleme(clustering) işlemi, hedef gen tespit sürecinin önemli adımlarından biridir. Gen ekspresyon verisindeki örneklerin ne kadar benzediğini görmek için etkili bir yöntemdir. Çalışmada kullanılan verisetlerinde olduğu gibi, büyük veri içeren ve birçok çok replikası olan verilerin değerlendirilmesini kolaylaştırmaktadır. Mikrodizi verilerini incelemek için kullanılacak birçok korelasyon yöntemi bulunmaktadır (Bishara ve Hittner, 2012).

- a) Pearson Korelasyon: Değişkenler arasındaki doğrusal ilişkiyi ölçer (Bolstad ve diğ., 2003). Pearson korelasyonunun 3 değer aralığı bulunur:
  - Pozitif yönlü doğrusal ilişki (+1'e yakın)
  - Negatif yönlü doğrusal ilişki (-1'e yakın)
  - Doğrusal ilişki yok (0'a yakın)
- b) Spearman Korelasyon: Değişkenler arasındaki sıralı ilişkiyi ölçer. Veriler normal dağılıyorken Pearson korelasyon tercih edilirken, normal dağılım olmadığında spearman tercih edilebilir.
- c) Kendall Tau Korelasyon: Spearman'den daha hassas sonuç verebilen bu yöntem, veriler arasındaki sıralı ilişkiyi ölçer. Küçük veri setleri için daha uygundur.
- d) Hiyerarşik: Veriler arasındaki ilişkiyi ölçmek için öklit mesafesi, manhattan mesafesi, kosinüs benzerliği gibi farklı mesafe metriklerini kullanarak değerlendirme sağlar.
- e) K-Means: Verilerin birbirine yakınlığını ölçer.

Çalışmada geliştirilen SMATP yönteminin kümeleme aşamasında "Pearson Korelasyona Bağlı Kümeleme" kullanılmıştır. Pearson korelasyon matematiksel olarak iki değişkenin (x ve y gibi) ortak varyansının, bu değişkenlerin ayrı ayrı standart sapmalarının

çarpımına oranı olarak ifade edilir. Pearson korelasyon katsayısı (r) aşağıdaki şekilde hesaplanmaktadır.

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} \quad (3.2)$$

Burada;

- x ve y karşılaştırılan iki değişken,
- n veri noktalarının sayısı,
- $\sum xy$  eşleştirilmiş puanların çarpımının toplamı,
- $\sum x^2$ ,  $\sum xy^2$  her bir değişken için kareler toplamıdır.

Korelasyon matrisi hazırlanırken yöntemin önceki adımlarında hazırlanan veri çerçevesi ve veri matrisi kullanılmıştır. Hazırlanan matrisin tek boyutlu, düz bir liste haline getirilmesi için, matris bir vektöre dönüştürülmüştür. Önceki adımlar ile kontrolleri sağlanarak analize uygun olduğuna karar verilen ve kümelenen veriseti bu aşamada sıcaklık haritası kullanılarak görselleştirilmiştir. Birden çok replika içeren bu şekildeki büyük gen ekspresyon verilerinde benzer veri gruplarını bir araya getirme işlemi olan kümeleme verinin keşfi, anomali tespiti gibi işlemler için önemli bir adımdır. Yapılan kümeleme işlemi verinin durumunu görmek ve sonraki adımlarda gerçekleştirilen gruplama sürecine destek olmak açısından önemli bir adım olmaktadır. Bu aşamada kullanılan sıcaklık haritası normalizasyon öncesi durumu gösterirken, normalizasyon işlemi sonrasında tekrar sıcaklık haritası alınarak karşılaştırma yapılmaktadır.

### 3.1.4. Normalizasyon

Mikrodizi veri analizinde doğru sonuçlar elde etmek için verilerin uygun şekilde normalize edilmesi önemli bir adımdır. Normalizasyon ile teknik ve biyolojik varyasyonu azaltarak daha güvenilir sonuçlar elde edilebilir. Normalizasyon veri analiz sürecinde değerlerin belli bir aralığa getirilmesini sağlar. Farklı değerlerdeki veriyi 0-1 arasına ölçeklendirmek değerlendirme sürecini kolaylaştırarak sağlıklı karşılaştırmalar yapmayı mümkün kılar. Illumina BeadStudio ile hazırlanmış olan mikrodizi verilerinin analiz sürecinde kullanılacak farklı normalizasyon yöntemleri bulunmaktadır. Mikrodizi dizi veri analizinde, R programlama dilinde kullanılacak yaygın normalizasyon yöntemleri aşağıdaki gibidir:

- a) Niceliksel (Quantile) Normalizasyon: Tüm örneklerdeki yoğunluk dağılımlarını eşitlemeyi hedefleyen normalizasyon türüdür. Farklı örneklerdeki sistematik varyasyonları azaltarak verinin karşılaştırılabilir hale getirilmesini sağlar. (Zhao ve diğ., 2020)

b) Ölçek (Scale) Normalizasyon: Örnekleri belli bir ölçeğe göre normalize ederek sinyal yoğunluğunu aynı ölçüye indirmek için kullanılır.

c) Döngüsel(Cyclic Loess) Normalizasyon: Yüksek doğruluk gerektiren verilerde pairwise karşılaştırmalar yaparak normalizasyon sağlar. Daha hassas bir normalizasyon sağlarken işlem sürecinin uzaması dezavantajıdır.

d) Robust Spline Normalizasyon: Normalizasyon, arka plan düzeltme ve varyans stabilizasyonunu bir arada uygulayan normalizasyon şeklidir. Parazit sinyalleri ve teknik varyasyonu azaltmak için etkili bir yöntemdir.

SMATP metodumuzda biyolojik gürültü, teknik varyasyonu azaltması, veri karşılaştırma ve ölçekleme sağlaması gibi avantajları sebebiyle “niceliksel normalizasyon” tercih edilmiştir. Niceliksel normalizasyonun çalışması aşağıdaki adımlarla gerçekleştirilmektedir:

- Veri matrisi: Gen ekspresyon verisindeki genler satırları, örnekler sütunları oluşturacak şekilde bir veri matrisi hazırlanır.  $X \in \mathbb{R}^{a \times b}$  (a gen, b örnek)

- Sıralama: Her sütun  $X_f$  sıralı bir vektör  $X_f^{sırala}$  haline getirilir.

- Ortalama hesaplanması: Tüm sütunların sıralanmış değerlerinin ortalaması alınır.

$$Ort_i = \frac{1}{n} \sum_{f=1}^n X_{s,f}^{sırala} \quad (3.3)$$

- Değerlerin değiştirilmesi: Sıralı değerler, ortalama değer ile değiştirilir.

Yapılan normalizasyon işleminden sonra önceki adımlarda hazırlanan sinyal dağılımı ve sıcaklık haritası görselleştirmeleri tekrar hazırlanarak, normalizasyon öncesi ve normalizasyon sonrası durumun karşılaştırılması sağlanmıştır.

### 3.1.5. Gruplama

Normalize edilmiş verinin değerlendirilmesi için bir sonraki adım olarak karşılaştırma grupları hazırlanmıştır. Doğrusal modelleme ile bir karşılaştırma matrisi (contrast matrix) oluşturularak hangi grupların karşılaştırılacağı belirlenmiştir ve bu gruplara ayrı ayrı başlıklar verilmiştir. Bu matris, karşılaştırma grubundaki iki koşul arasındaki gen ekspresyon farklılıklarının tespiti için yardımcı olmaktadır. Yapılan karşılaştırmada p-değerleri kontrol edilerek hangi genlerin etkilendiği ve etkilenme yönü (aşağı regüle/yukarı regüle) tespit edilebilmektedir. Gruplar veya koşullar arasındaki karşılaştırmaları tanımlamak için kontrast matris hazırlamak etkili bir yöntemdir. Bu matris veri kalitesi kontrolleri adımı hazırlanan tasarım matrisindeki parametrelerin belirli kombinasyonlarını test etmek için kullanılır.

Karşılaştırma matrisi aşağıdaki şekilde ifade edilir:

$$K = \begin{bmatrix} k_{11} & k_{12} & \cdots & k_{1m} \\ k_{21} & k_{22} & \cdots & k_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ k_{n1} & k_{n2} & \cdots & k_{nm} \end{bmatrix} \quad (3.4)$$

n: test edilen karşılaştırmaların sayısı

m: tasarım matrisindeki sütunların sayısı

Bu matriste her satır bir hipotezi temsil eder. Örneğin;

\*  $k_{11} = 1, k_{12} = -1 \rightarrow k_{11}$  grubu ile  $k_{12}$  grubu arasındaki farkı test eder.

\*  $k_{11} = 1, k_{12} = 0, k_{13} = -1 \rightarrow k_{11}$  grubu ile  $k_{13}$  arasındaki farkı test eder.

Örnek: 3 gruptan (sırasıyla kontrol, tedavi1, tedavi2) oluşan bir karşılaştırma matrisi aşağıdaki gibi tanımlanabilir.

$$K = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \quad (3.5)$$

Bu matriste;

1.satır  $\rightarrow$  Kontrol grubu ile tedavi1 arasındaki fark

2.satır  $\rightarrow$  Tedavi1 ile tedavi2 arasındaki farkın test edileceğini gösterir.

Yapılacak karşılaştırmalar isimlendirilerek hipotezlerin anlaşılır olması sağlanır.

Yukarıdaki örnek için aşağıdaki benzer başlıklandırma yapılabilir:

\* KontrolvsTedavi1: kontrol grubu ile tedavi1 grubu arasındaki fark

\* Tedavi1vsTedavi2: tedavi1 grubu ile tedavi2 grubu arasındaki fark

Karşılaştırma matrisine bağlı olarak her bir hipotez için test istatistikleri (t-değeri, p-değeri) hesaplanır. Bu değerler, genlerin belirli gruplar arasında farklı ekspresyon gösterip göstermediğini belirlemek için kullanılır.

### 3.1.6. Değerlendirme

Normalizasyon ve karşılaştırma matrisinin hazırlanması işlemleri sonrasında, hazırlanan karşılaştırma matrisi farklı koşullar arasındaki farkların istatistiksel olarak test edilmesi için kullanılmıştır. Mikrodizi analizinde istatistiksel değerlendirme sürecinde kullanılacak farklı metotlar mevcuttur. Kullanılan verinin türüne ve içeriğine bağlı olarak farklı yöntemler daha etkili olabilmektedir. Bu süreçte kullanılacak bazı yöntemler;

- eBayes (Empirical Bayes): Diferansiyel gen ekspresyon analizinde kullanılan bir tekniktir. Doğrusal modeller için hazırlanmış olan bu yöntem, gen ekspresyon varyasyonlarını modellemek için Bayes teorisini kullanır. FDR (False Discovery Rate) kontrolü ile birlikte etkili bir çözüm sunan bu yöntem yüksek boyutlu verilerde hızlı

sonuç verebilmektedir (Aubert ve diğ., 2004). FDR'de amaç yapılan çoklu testlerde yanlış pozitif sonuçların oranını kontrol etmektir (Pawitan ve diğ., 2005). FDR istatistiğinde p-value, elde edilen verinin gözlemlenen sonuçtan daha uç bir sonuç verme olasılığını gösterir. P-value düşük olan kayıtlar etkilenen genleri göstermektedir. Verisine bağlı olarak farklı eşik değerleri kullanılabilir.

- b. T-Test: Genellikle iki gruplu (deney-kontrol) verisetlerinde tercih edilen basit ve yaygın bir yöntemdir.
- c. ANOVA: Üç veya daha çok grubun bulunduğu verilerde tercih edilen bir yöntemdir.
- d. DESeq / DESeq2: Illumina mikrodizi verilerinde tercih edilen yöntemlerdendir.
- j) EdgeR: Mikrodizi ve RNA-seq verilerinde tercih edilen başka bir yöntemdir.

SMATP yönteminde veri içeriğine uygun olacak şekilde eBayes ile istatistiksel değerlendirmeler yapılmıştır. eBayes yöntemi geleneksel bayes yönteminden uyarlanarak hazırlanmaktadır. Geleneksel bayes yönteminde, bir ön bilgi (prior) tanımlanır ve bu bilgi, verilerden elde edilen bilgi ile (likelihood) birleştirilerek bir son bilgi (posterior) elde edilir. eBayes yönteminde ise ön bilgi veri kümesinden tahmin edilmektedir. eBayes, her gen için bir lineer model kurar:

$$y_{ab} = \mu_a + \epsilon_{ab}, \quad \epsilon_{ab} \sim N(0, \sigma_a^2) \quad (3.6)$$

Bu modelde;

- \*  $y_{ab}$ : a.gen ve b.örnek için gen ekspresyon değeri,
- \*  $\mu_a$ : a.genin beklenen ekspresyon değeri,
- \*  $\epsilon_{ab}$ : hata terimi, genellikle normal dağılım  $N(0, \sigma_a^2)$  varsayılır,
- \*  $\sigma_a^2$ : a.genin varyansıdır.

eBayes yöntemi, genler arasındaki varyans tahminlerini birleştirerek daha kararlı bir varyans tahmini oluşturur. eBayes ile yapılan değerlendirme sürecinde FDR istatistiği kullanılarak p-değeri üzerinden kontroller sağlanmıştır. Hazırlanan yöntem farklı eşik değerleri için gen değişimlerini tespit etmeyi sağlayabilmektedir. Elde edilen sonuçlar grafik ve sıcaklık haritası ile görselleştirilmiştir.

### 3.2. ÖZGÜN VERİSETİ

Özgün veriseti olarak Doç.Dr. Leyla TÜRKER ŞENER hocanın “Nikotin ve Radrasyon İndüklemeli Meme Kanseri Kök Hücreleri Üzerine Metforminin Antikanser Etkisi” isimli proje kapsamında hazırlanmış olduğu microRNA mikrodizi verileri kullanılmıştır.

Mikrodizi araştırmalarında elde edilen veri kümesi çeşitli aşamalardan geçer. Ham veri elde edildiğinde veri analizleri için henüz kullanılabilir durumda değildir. Ham mikrodizi veri üreten çeşitli platformlar vardır: Affymetrix, Agilent, Illumina gibi. Bu tez çalışmasındaki veriler Illumina’da hazırlanmıştır.

Çalışmada 9 farklı deney aynı zamanda, aynı koşullar altında uygulanmıştır. Illumina ile oluşturulan ham veri dosyası tüm deneyleri bir arada sunmaktadır. Veriseti bu hali ile analize uygun olmayıp önce 9 deneyin ayrıştırılması ve sonrasında her bir deneyin ayrı ayrı değerlendirilmesi yoluna gidilmiştir. Çalışmada MCF7 ve MDA-N1 kanser hücre soyu mirna hücrelerine Nikotin, Metformin ve Işın’ın farklı bileşimlerde ve dozda uygulanması ile meydana gelen değişimin izlenmesi hedeflenmiştir. Tedavi (nikotin, metformin, ışın uygulanması) hem MCF7 hem de MDA-N1 hücrelerine uygulanmış ve birden çok replika kullanılmıştır.

### 3.3. VERİSETİNİN AYRI DENEYLER ŞEKLİNDE AYRIŞTIRILMASI

Özgün veriseti 9 deneyin aynı koşullar altında, eş zamanlı gerçekleştirildiği bir çalışmadır. Ham veride tüm deney verileri tek bir kaynaktır. Deneyleri değerlendirebilmek için ilk işlem olarak bu deney verilerinin ayrıştırılarak ayrı ayrı incelemeye hazır hale getirilmesi sağlanmıştır. Ham veriler Illumina GenomeStudio(BeadStudio yeni versiyon ismi) ile hazırlanmıştır. 9 farklı tedavi yöntemi için 18 farklı karşılaştırma gerçekleştirilmiştir. Deney sürecinde hazırlanan hedef ve kontrol grupları Tablo 3.2’de gösterildiği gibidir.

**Tablo 3.2:** Özgün veriseti ham veri karşılaştırma grupları

No	Hedef Grup	Kontrol Grup	Açıklama
1	MDA-K	MDA-N	MDA-N1 kontrol grubu ile MDA-N1 hüresine Nikotin uygulanmış karşılaştırma grubu
2	MDA-K	MDA-M1000	MDA-N1 kontrol grubu ile MDA-N1 hüresine M1000 uygulanmış karşılaştırma grubu
3	MDA-K	MDA-M100	MDA-N1 kontrol grubu ile MDA-N1 hüresine M100 uygulanmış karşılaştırma grubu
4	MDA-K	MDA-I	MDA-N1 kontrol grubu ile MDA-N1 hüresine Işın uygulanmış karşılaştırma grubu
5	MDA-K	MDA-N-M1000	MDA-N1 kontrol grubu ile MDA-N1 hüresine Nikotin+M1000 uygulanmış karşılaştırma grubu

**Tablo 3.2 (devam)**

6	MDA-K	MDA-N-M100	MDA-N1 kontrol grubu ile MDA-N1 hücresine Nikotin+M100 uygulanmış karşılaştırma grubu
7	MDA-K	MDA-N-I	MDA-N1 kontrol grubu ile MDA-N1 hücresine Nikotin+Işın uygulanmış karşılaştırma grubu
8	MDA-K	MDA-N-I-M1000	MDA-N1 kontrol grubu ile MDA-N1 hücresine Nikotin+Işın+M1000 uygulanmış karşılaştırma grubu
9	MDA-K	MDA-N-I-M100	MDA-N1 kontrol grubu ile MDA-N1 hücresine Nikotin+Işın+M100 uygulanmış karşılaştırma grubu
10	MCF7-K	MCF7-N	MCF7 kontrol grubu ile MCF7 hücresine Nikotin uygulanmış karşılaştırma grubu
11	MCF7-K	MCF7-M1000	MCF7 kontrol grubu ile MCF7 hücresine M1000 uygulanmış karşılaştırma grubu
12	MCF7-K	MCF7-M100	MCF7 kontrol grubu ile MCF7 hücresine M100 uygulanmış karşılaştırma grubu
13	MCF7-K	MCF7-I	MCF7 kontrol grubu ile MCF7 hücresine Işın uygulanmış karşılaştırma grubu
14	MCF7-K	MCF7-N-M1000	MCF7 kontrol grubu ile MCF7 hücresine Nikotin+M1000 uygulanmış karşılaştırma grubu
15	MCF7-K	MCF7-N-M100	MCF7 kontrol grubu ile MCF7 hücresine Nikotin+M100 uygulanmış karşılaştırma grubu
16	MCF7-K	MCF7-N-I	MCF7 kontrol grubu ile MCF7 hücresine Nikotin+Işın uygulanmış karşılaştırma grubu
17	MCF7-K	MCF7-N-I-M1000	MCF7 kontrol grubu ile MCF7 hücresine Nikotin+Işın+M1000 uygulanmış karşılaştırma grubu
18	MCF7-K	MCF7-N-I-M100	MCF7 kontrol grubu ile MCF7 hücresine Nikotin+Işın+M100 uygulanmış karşılaştırma grubu

Toplam 34 örnekten oluşan deneyler ve örnek sayıları Tablo 3.3'de gösterilmektedir.

**Tablo 3.3: Özgün veriseti deney grupları örnek sayıları**

Veri Seti İçeriği	MDA Örnek Sayı	MCF7 Örnek Sayı	Açıklama
Kontrol	2	2	Hem MDA hem MCF7 için 2 şer tane kontrol deney grubu bulunmaktadır.
Nikotin	2	1	MDA hücre soyuna Nikotin uygulanan 2; MCF7 hücre soyuna Nikotin uygulanan 1 deney grubu bulunmaktadır.
Işın	2	1	MDA hücre soyuna Işın uygulanan 2; MCF7 hücre soyuna Işın uygulanan 1 deney grubu bulunmaktadır.
M100	2	2	Hem MDA hem MCF7 hücre soyuna M100 uygulanan 2şer deney grubu bulunmaktadır.
M1000	2	2	Hem MDA hem MCF7 hücre soyuna M1000 uygulanan 2şer deney grubu bulunmaktadır.
Nikotin+M100	2	1	MDA hücre soyuna Nikotin+M100 uygulanan 2; MCF7 hücre soyuna Nikotin+M100 uygulanan 1 deney grubu bulunmaktadır.
Nikotin+M1000	1	1	Hem MDA hem MCF7 hücre soyuna Nikotin+M1000 uygulanan 1er deney grubu bulunmaktadır.

**Tablo 3.3 (devam)**

Nikotin+Işın	1	2	MDA hücre soyuna Nikotin+Işın uygulanan 1; MCF7 hücre soyuna Nikotin+Işın uygulanan 2 deney grubu bulunmaktadır.
Nikotin+Işın+M100	2	2	Hem MDA hem MCF7 hücre soyuna Nikotin+Işın+M100 uygulanan 2şer deney grubu bulunmaktadır.
Nikotin+Işın+M1000	1	3	MDA hücre soyuna Nikotin+Işın+M1000 uygulanan 1; MCF7 hücre soyuna Nikotin+Işın+M1000 uygulanan 3 deney grubu bulunmaktadır.

Bu veriler her bir deney grubunun örnekleri ayrı bir veriseti olacak şekilde ayrıştırılarak, işlem anlaşılabilirliği sağlamak için örnekler yeniden isimlendirilmiştir. Yeniden isimlendirilen veriseti başlıkları Tablo 3.4’de gösterilmektedir.

**Tablo 3.4: Özgün veriseti yeniden başlıklandırma çalışması**

Orijinal Başlık	Deney Adı	Yeni Başlık	Açıklama
101178090075_A	MDA-K	MDA_K_1	MDA kontrol verisi 1, Orijinal veride kolon başlıklarında geçen "101178090075_A" yerine "MDA_K_1" yazılmıştır
101178090075_B	MDA-N1	MDA_N_1	MDA hüresine Nikotin uygulanan veri grubu 1, Orijinal veride kolon başlıklarında geçen "101178090075_B" yerine "MDA_N_1" yazılmıştır
101178090075_C	MDA-M1000	MDA_M1000_1	MDA hüresine M1000 (1000 milimolar metformin) uygulanan veri grubu 1
101178090075_D	MDA-M100	MDA_M100_1	MDA hüresine M100 (100 milimolar metformin) uygulanan veri grubu 1
101178090075_E	MDA-I	MDA_I_1	MDA hüresine Işın uygulanan veri grubu 1
101178090075_F	MDA-N-M100	MDA_N_M100_1	MDA hüresine Nikotin ve M100 uygulanan veri grubu 1
101178090075_I	MDA-N-I-M100	MDA_N_I_M100_1	MDA hüresine Nikotin, Işın ve M100 uygulanan veri grubu 1
101178090075_J	MCF7-K	MCF7_K_1	MCF7 kontrol verisi 1
101178090075_K	MCF7-M100	MCF7_M100_1	MCF7 hüresine M100 uygulanan veri grubu 1
101178090075_L	MCF7-M1000	MCF7_M1000_1	MCF7 hüresine M1000 uygulanan veri grubu 1
101178090077_A	MCF7-N-I-M100	MCF7_N_I_M100_1	MCF7 hüresine Nikotin, Işın ve M100 uygulanan veri grubu 1
101178090077_B	MCF7-N-I-M1000	MCF7_N_I_M1000_1	MCF7 hüresine Nikotin, Işın ve M1000 uygulanan veri grubu 1

**Tablo 3.4 (devam)**

101178090077_C	MCF7-N-I	MCF7_N_I_1	MCF7 hücreesine Nikotin ve Işın uygulanan veri grubu 1
101178090077_D	MCF7-N1	MCF7_N_1	MCF7 hücreesine Nikotin uygulanan veri grubu 1
101178090077_E	MCF7-N-M1000	MCF7_N_M1000_1	MCF7 hücreesine Nikotin ve M1000 uygulanan veri grubu 1
101178090077_F	MCF7-N-M100	MCF7_N_M100_1	MCF7 hücreesine Nikotin ve M100 uygulanan veri grubu 1
101178090077_G	MCF7-M100	MCF7_M100_2	MCF7 hücreesine M100 uygulanan veri grubu 2
101178090077_H	MCF7-M1000	MCF7_M1000_2	MCF7 hücreesine M1000 uygulanan veri grubu 2
101178090077_I	MCF7-N-I-M100	MCF7_N_I_M100_2	MCF7 hücreesine Nikotin, Işın ve M100 uygulanan veri grubu 2
101178090077_J	MCF7-N-I-M1000	MCF7_N_I_M1000_2	MCF7 hücreesine Nikotin, Işın ve M1000 uygulanan veri grubu 2
101178090077_K	MCF7-N-I-M1000	MCF7_N_I_M1000_3	MCF7 hücreesine Nikotin, Işın ve M1000 uygulanan veri grubu 3
101178090077_L	MCF7-N-I	MCF7_N_I_2	MCF7 hücreesine Nikotin ve Işın uygulanan veri grubu 2
100896670083_A	MDA-K	MDA_K_2	MDA kontrol verisi 2
100896670083_B	MDA-N1	MDA_N_2	MDA hücreesine Nikotin uygulanan veri grubu 2
100896670083_C	MDA-M1000	MDA_M1000_2	MDA hücreesine M1000 uygulanan veri grubu 2
100896670083_D	MDA-M100	MDA_M100_2	MDA hücreesine M100 uygulanan veri grubu 2
100896670083_E	MDA-I	MDA_I_2	MDA hücreesine ışın uygulanan veri grubu 2
100896670083_F	MDA-N-M1000	MDA_N_M1000_1	MDA hücreesine Nikotin ve M1000 uygulanan veri grubu 1
100896670083_G	MDA-N-M100	MDA_N_M100_2	MDA hücreesine Nikotin ve M100 uygulanan veri grubu 2
100896670083_H	MDA-N-I	MDA_N_I_1	MDA hücreesine Nikotin ve Işın uygulanan veri grubu 1
100896670083_I	MDA-N-I-M1000	MDA_N_I_M1000_1	MDA hücreesine Nikotin, Işın ve M1000 uygulanan veri grubu 1
100896670083_J	MDA-N-I-M100	MDA_N_I_M100_2	MDA hücreesine Nikotin, Işın ve M100 uygulanan veri grubu 2
100896670083_K	MCF7-I	MCF7_I_1	MCF7 hücreesine Işın uygulanan veri grubu 1
100896670083_L	MCF7-K	MCF7_K_2	MCF7 kontrol verisi 2

Verilerin incelenerek, başlıklandırılması, her bir tedavi türü (uygulanan madde) özelinde gruplanarak ayrı ham verisetleri hazırlanması sonrasında oluşan deney grupları Tablo 3.5'deki gibi olmuştur.

**Tablo 3.5:** Özgün veriseti deney grupları

Uygulanan Tedavi	MCF7	MDA	Veri Setleri
Nikotin	√	√	Deney Grubu1
M100	√	√	Deney Grubu2
M1000	√	√	Deney Grubu3
Işın	√	√	Deney Grubu4
Nikotin + M100	√	√	Deney Grubu5
Nikotin + M1000	√	√	Deney Grubu6
Nikotin + Işın	√	√	Deney Grubu7
Nikotin + Işın + M100	√	√	Deney Grubu8
Nikotin + Işın + M1000	√	√	Deney Grubu9

## 4. BULGULAR

Bu bölümde biri özgün 3 farklı verisetine ait 11 farklı deneyin analiz ve hedef gen çıkarımı bulguları detaylandırılmıştır. Illumina BeadStudio yazılımı ile hazırlanarak bilgisayar ortamına aktarılmış ham veriler üzerinde, R programlama dili ve Bioconductor paketi altındaki kütüphanelerin desteği ile geliştirilen SMATP yöntemi çıktıları paylaşılmaktadır. 4.1. bölümünde özgün veriseti, 4.2. bölümünde GSE10470 veriseti, 4.3. bölümünde ise GSE8262 veriseti ile yapılan çalışmalara ait bulgular yer almaktadır.

### 4.1. ÖZGÜN VERİSETİ ANALİZ VE DEĞERLENDİRME

Bu bölümde Doç.Dr. Leyla TÜRKER ŞENER hocanın “Nikotin ve Radrasyon İndüklemeli Meme Kanseri Kök Hücreleri Üzerine Metforminin Antikanser Etkisi” isimli proje kapsamında hazırlanmış olduğu microRNA hücre soylarından hazırlanmış mikrodizi verileri değerlendirilmiştir. Çalışmada meme kanseri hücre soylarına (MCF7 ve MDA-N1) 9 farklı tedavi eş zamanlı ve aynı koşullar altında uygulanmıştır.

#### 4.1.1. Deney Grubu 1 – Nikotin

MCF7 ve MDA-N1 meme kanseri hücre soylarına Nikotin uygulanması ile hazırlanmış olan deneye ait veriseti 57 kolon ve 34694 satır içermektedir. Veriseti 3 Nikotin uygulanmış örnek, 4 kontrol örneği olmak üzere toplam 7 örnekten oluşmaktadır. Nikotin deneyinde SMATP yöntemi aşağıdaki gibi uygulanmıştır.

İlk adım olarak veriseti bilgisayar ortamına aktarılmıştır. Ardından verinin değerlendirilebilmesi için veri çerçevesi hazırlanmıştır. Hazırlanan veri çerçevesi Tablo 4.1’da gösterildiği gibidir.

**Tablo 4.1:** Veri Çerçevesi - Nikotin

Sıra	Grup	GenTipi	Replika	Tedavi	DiziKodu
1	1	MDA	1	Kontrol	MDA_K_1
2	1	MDA	1	Nikotin	MDA_N_1
3	1	MCF7	1	Kontrol	MCF7_K_1
4	2	MCF7	1	Nikotin	MCF7_N_1
5	3	MDA	2	Kontrol	MDA_K_2
6	3	MDA	2	Nikotin	MDA_N_2
7	3	MCF7	2	Kontrol	MCF7_K_2

Veri çerçevesi yardımıyla Nikotin deneyindeki örneklerin sıralaması Tablo 4.2’de gösterildiği gibi olmuştur.

**Tablo 4.2:** Örneklerin Sıralanması - Nikotin

Sıra No	Örnek Adı
1	MDA.Kontrol.1
2	MDA.Nikotin.2
3	MCF7.Kontrol.3
4	MCF7.Nikotin.4
5	MDA.Kontrol.5
6	MDA.Nikotin.6
7	MCF7.Kontrol.7

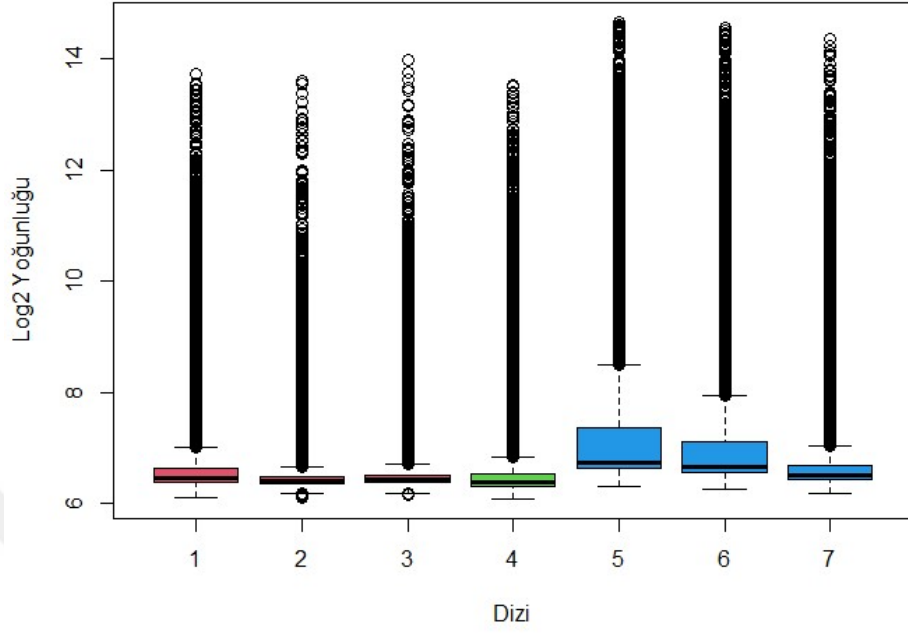
Sıralanmış verilerin log2 dönüşümü yapıldıktan sonra oluşturulan tasarım matrisi Tablo 4.3’de gösterilmektedir.

**Tablo 4.3:** Tasarım Matrisi - Nikotin

Örnek	Min	1st Qu.	Median	Mean	3rd Qu.	Max
MDA.Kontrol.1	6.102	6.385	6.451	6.733	6.636	13.722
MDA.Nikotin.2	6.102	6.361	6.409	6.514	6.482	13.615
MCF7.Kontrol.3	6.150	6.384	6.435	6.554	6.515	13.980
MCF7.Nikotin.4	6.083	6.320	6.380	6.615	6.525	13.526
MDA.Kontrol.5	6.306	6.634	6.736	7.311	7.376	14.663
MDA.Nikotin.6	6.259	6.562	6.664	7.119	7.116	14.562
MCF7.Kontrol.7	6.192	6.445	6.513	6.767	6.682	14.356

İşlenmemiş/ham verideki sinyal dağılımı kontrol edilmiştir. Normalizasyon işlemi öncesindeki sinyal dağılımı Şekil 4.1’deki kutu grafik ile gösterildiği gibidir.

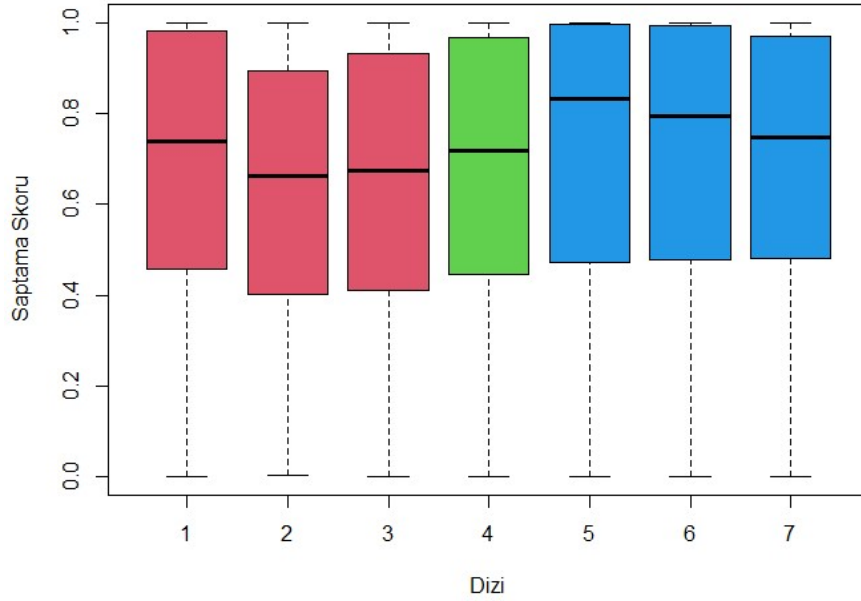
Normalizasyon Öncesi Sinyal Dağılımı



Şekil 4.1: Normalizasyon Öncesi Sinyal Dağılımı - Nikotin

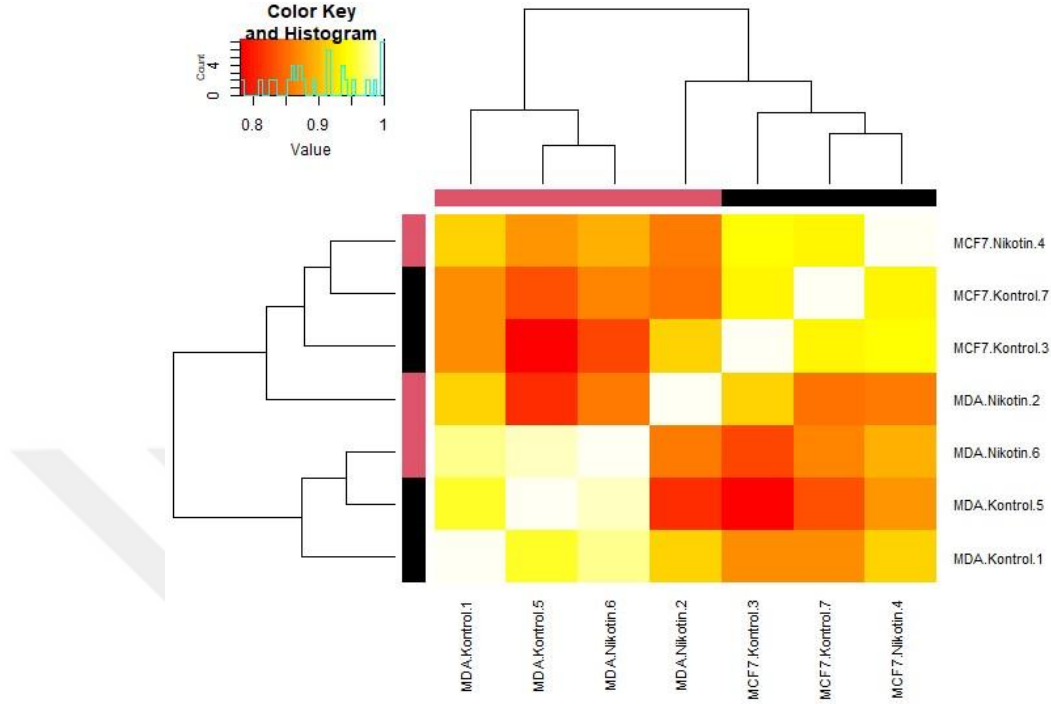
Sinyal dağılımının kontrolünün ardından saptama skoru kontrol işlemi gerçekleştirilerek ham verideki genlerin 0-1 arasında ifade edilme olasılığı Şekil 4.2’de gösterilmiştir. (0: ifade edilmiyor, 1: ifade ediliyor)

Saptama Skorları



Şekil 4.2: Saptama Skorları - Nikotin

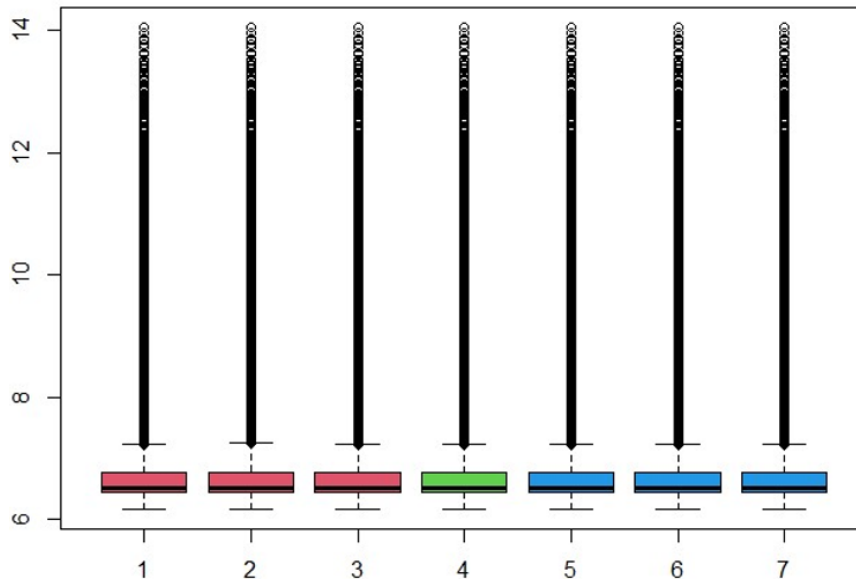
Veri kalitesi kontrolleri tamamlanan veriseti üzerinde pearson korelasyona bağlı kümeleme işlemi yapılmıştır. Kümelenmiş veri Şekil 4.3'deki ısı haritasındaki gibidir.



Şekil 4.3: Normalizasyon Öncesi Kümeleme Isı Haritası- Nikotin

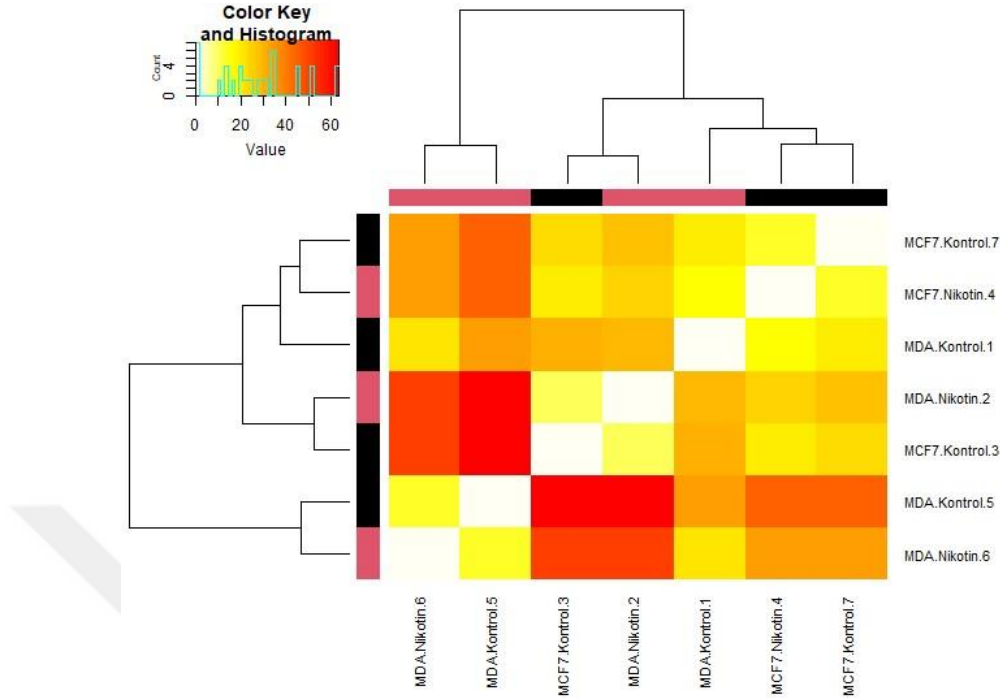
Niceliksel normalizasyon işlemi ile veri 0-1 arasına ölçeklendirilmiştir. Normalizasyon işlemi sonrasında sinyal dağılımı ve ısı haritası tekrar hazırlanarak değişim incelenmiştir. Niceliksel normalizasyon sonrası sinyal dağılımı Şekil 4.4'deki gibidir.

#### Niceliksel (Quantile) Normalizasyon



Şekil 4.4: Normalizasyon Sonrası Sinyal Dağılımı- Nikotin

Normalizasyon sonrası ısı haritası Şekil 4.5’deki gibidir.



**Şekil 4.5:** Normalizasyon Sonrası Kümelenmiş Veri Isı Haritası - Nikotin

Normalizasyon işleminin ardından karşılaştırmaların yapılabilmesi için gruplar hazırlanmıştır:

- MCF7wsMDA = MCF7.Kontrol – MDA.Kontrol
- NikotinMDA = MDA.Nikotin – MDA.Kontrol
- Nikotin MCF7 = MCF7.Nikotin – MCF7.Kontrol
- NikotinMDAnotNikotinMCF7 = (MDA.Nikotin-MDA.Kontrol)-(MCF.Nikotin-MCF7.Kontrol)

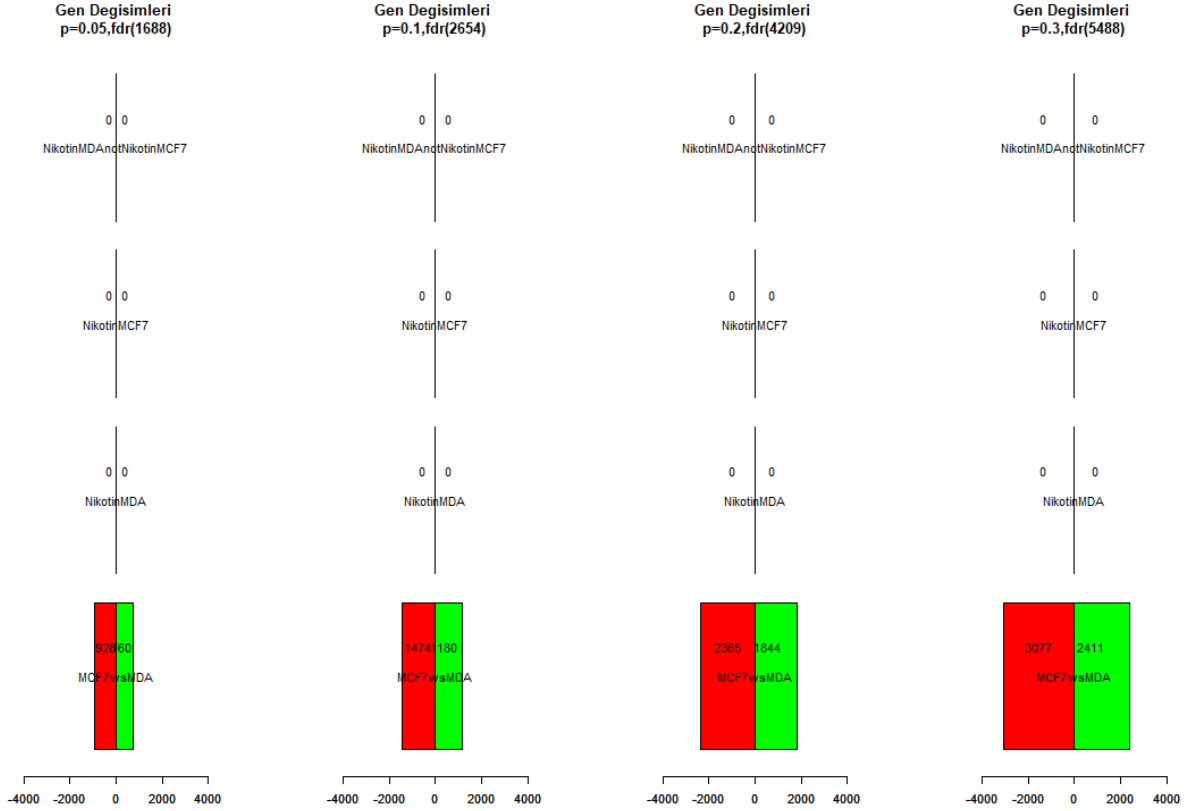
Karşılaştırma grupları için hazırlanan kontrast matrisi Tablo 4.4’da gösterilmektedir.

**Tablo 4.4:** Kontrast Matrisi – Nikotin

Seviye	MCF7wsMDA	NikotinMDA	NikotinMCF7	NikotinMDA notNikotinMCF7
MDA.Kontrol	-1	-1	0	-1
MDA.Nikotin	0	1	0	1
MCF7.Kontrol	1	0	-1	1
MCF7.Nikotin	0	0	1	-1

Karşılaştırma gruplarının hazırlanması sonrası SMATP yönteminin son aşaması uygulanarak veri değerlendirilmiştir. Değerlendirme sürecinde eBayes kullanılarak istatistiksel önem değerlendirilmiş, FDR istatistiği ile gözlem yapılmıştır. FDR’de sıklıkla kullanılan p-değeri ile kontroller yapılmıştır. P-değerinin 0’a yakın olması sonucun anlamlı olduğunu

göstermektedir. 4 farklı eşik değeri için kontroller yapıldığında gen değişimleri Şekil 4.6'daki gibi olmuştur. Farklı istatistiksel eşiklerde diferansiyel olarak ekprese edilen genlerin sayısının gösterildiği şekilde yeşil çubuklar yukarı düzenlenmiş genleri, kırmızı çubuklar ise aşağı düzenlenmiş genleri göstermektedir.



**Şekil 4.6:** Gruplar Arası Karşılaştırmalar - Nikotin

Sonuçlar incelendiğinde kontrol grubunda (hiçbir tedavi uygulanmamış kanser hücre soyu kontrol örnekleri) halihazırda aktive olan genler bulunurken nikotin uygulanan örneklerde belirgin bir gen değişiminin olmadığı gözlemlenmiştir.

#### 4.1.2. Deney Grubu 2 – M100

MCF7 ve MDA-N1 meme kanseri hücre soylarına M100 (100 milimolar metformin) uygulanması ile hazırlanmış olan deneye ait veriseti 65 kolon ve 34694 satır içermektedir. Veriseti 4 M100 uygulanmış örnek, 4 kontrol örneği olmak üzere toplam 8 örnekten oluşmaktadır. M100 deneyinde SMATP yöntemi aşağıdaki gibi uygulanmıştır.

İlk adım olarak veriseti bilgisayar ortamına aktarılmıştır. Ardından verinin değerlendirilebilmesi için veri çerçevesi hazırlanmıştır. Hazırlanan veri çerçevesi Tablo 4.5'de gösterildiği gibidir.

**Tablo 4.5:** Veri Çerçevesi – M100

Sıra	Grup	GenTipi	Replika	Tedavi	DiziKodu
1	1	MDA	1	Kontrol	MDA_K_1
2	1	MDA	1	M100	MDA_M100_1
3	1	MCF7	1	Kontrol	MCF7_K_1
4	1	MCF7	1	M100	MCF7_M100_1
5	2	MCF7	2	M100	MCF7_M100_2
6	3	MDA	2	Kontrol	MDA_K_2
7	3	MDA	2	M100	MDA_M100_2
8	3	MCF7	2	Kontrol	MCF7_K2

Veri çerçevesi yardımıyla M100 deneyindeki örneklerin sıralaması Tablo 4.6’da gösterildiği gibi olmuştur.

**Tablo 4.6:** Örneklerin Sıralanması – M100

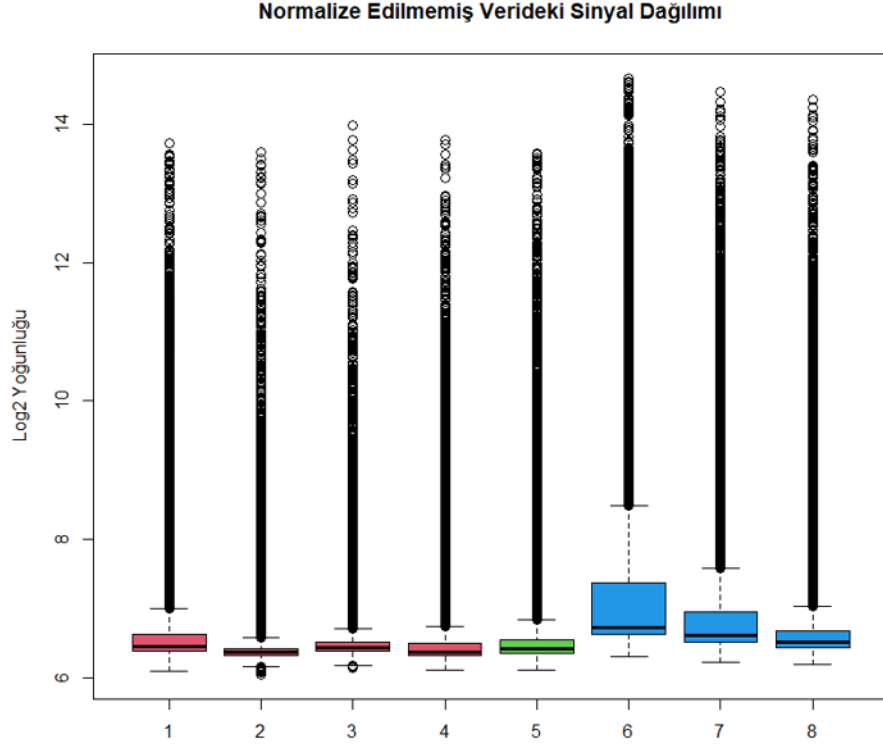
Sıra No	Örnek Adı
1	MDA.Kontrol.1
2	MDA.M100.2
3	MCF.Kontrol.3
4	MCF7.M100.4
5	MCF7.M100.5
6	MDA.Kontrol.6
7	MDA.M100.7
8	MCF7.Kontrol.8

Sıralanmış verilerin log<sub>2</sub> dönüşümü yapıldıktan sonra oluşturulan tasarım matrisi Tablo 4.7’de gösterilmektedir.

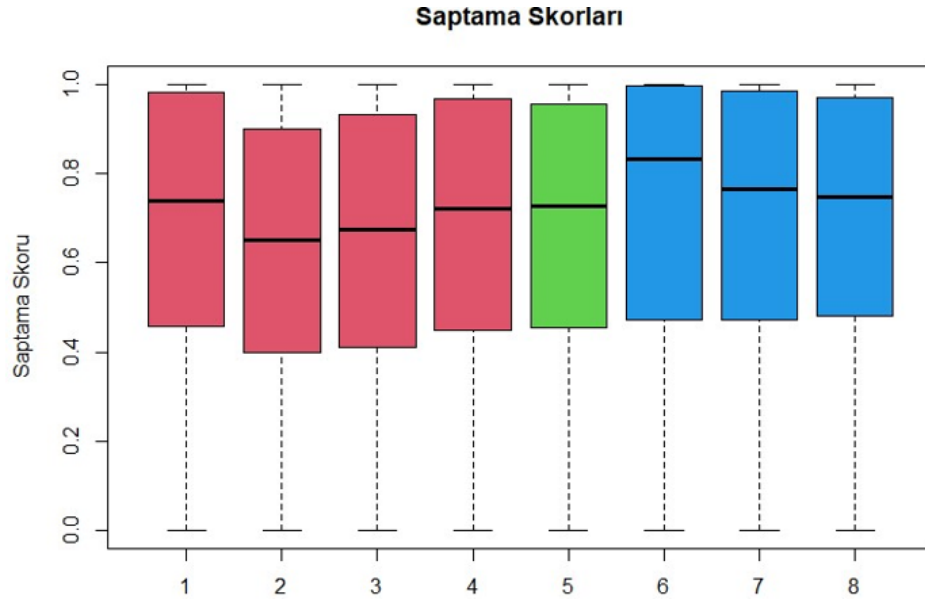
**Tablo 4.7:** Tasarım Matrisi – M100

Örnek	Min	1st Qu.	Median	Mean	3rd Qu.	Max
MDA.Kontrol.1	6.102	6.385	6.451	6.733	6.636	13.722
MDA.M100.2	6.049	6.324	6.369	6.468	6.431	13.597
MCF.Kontrol.3	6.150	6.384	6.435	6.554	6.515	13.980
MCF7.M100.4	6.119	6.326	6.380	6.581	6.497	13.776
MCF7.M100.5	6.111	6.365	6.425	6.639	6.558	13.573
MDA.Kontrol.6	6.306	6.634	6.736	7.311	7.376	14.663
MDA.M100.7	6.233	6.527	6.622	6.990	6.952	14.468
MCF7.Kontrol.8	6.192	6.445	6.513	6.767	6.682	14.356

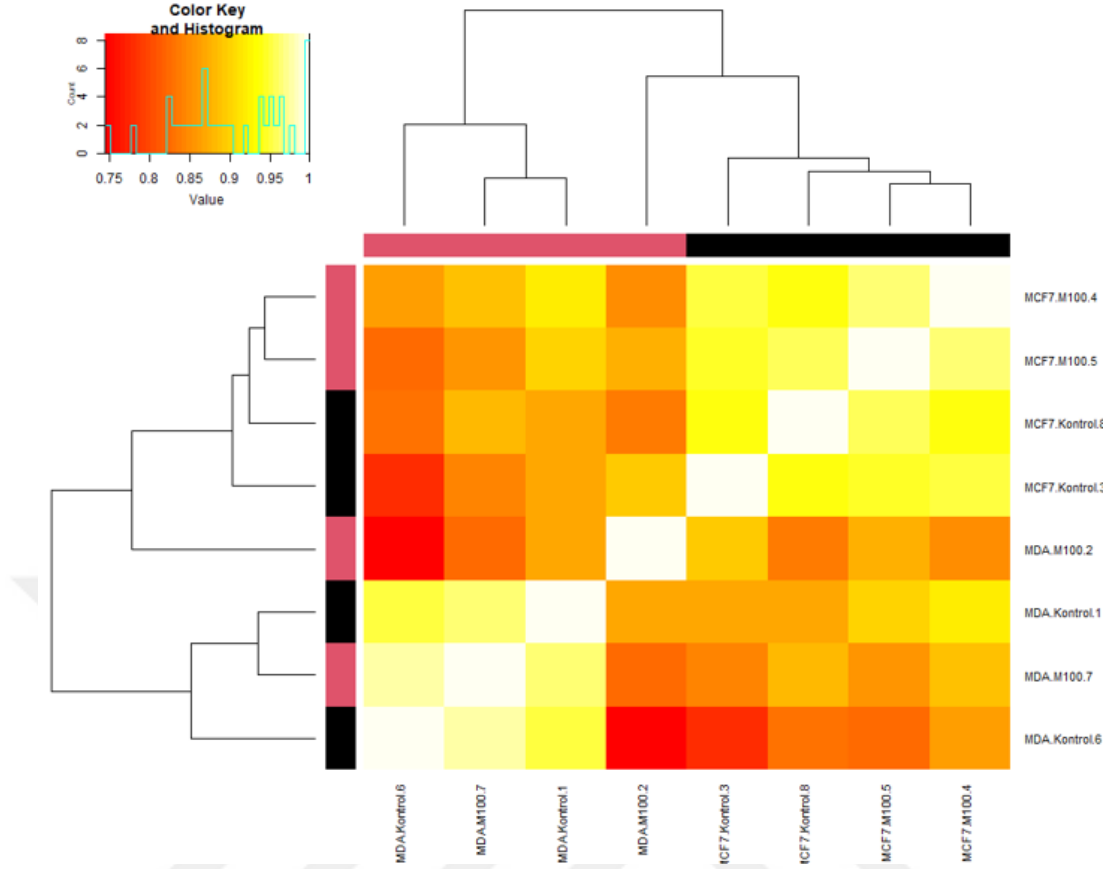
İşlenmemiş/ham verideki sinyal dağılımı kontrol edilmiştir. Normalizasyon işlemi öncesindeki sinyal dağılımı Şekil 4.7’deki kutu grafik ile gösterildiği gibidir.



Sinyal dağılımının kontrolünün ardından saptama skoru kontrol işlemi gerçekleştirilerek ham verideki genlerin 0-1 arasında ifade edilme olasılığı Şekil 4.8’de gösterilmiştir. (0: ifade edilmiyor, 1: ifade ediliyor)



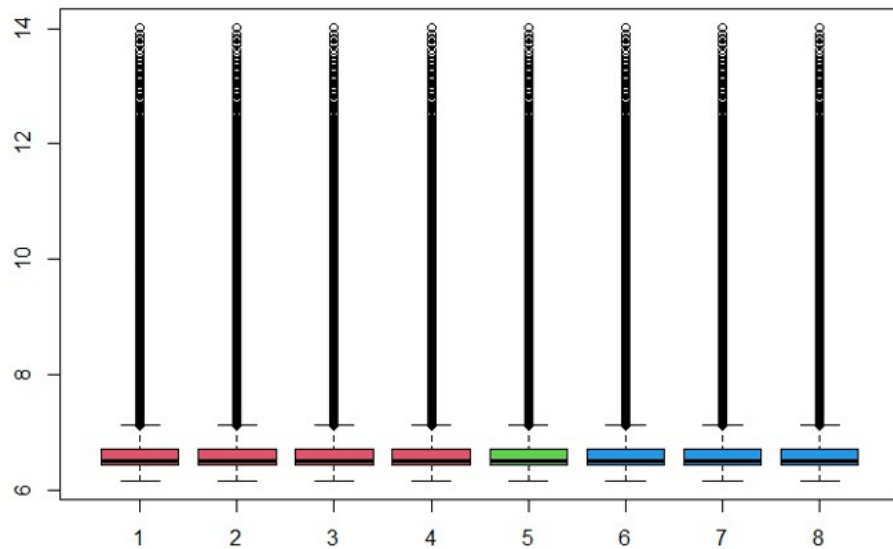
Veri kalitesi kontrolleri tamamlanan veriseti üzerinde pearson korelasyona bağlı kümeleme işlemi yapılmıştır. Kümelenmiş veri Şekil 4.9’daki ısı haritasındaki gibidir.



**Şekil 4.9:** Normalizasyon Öncesi Kümeleme Isı Haritası - M100

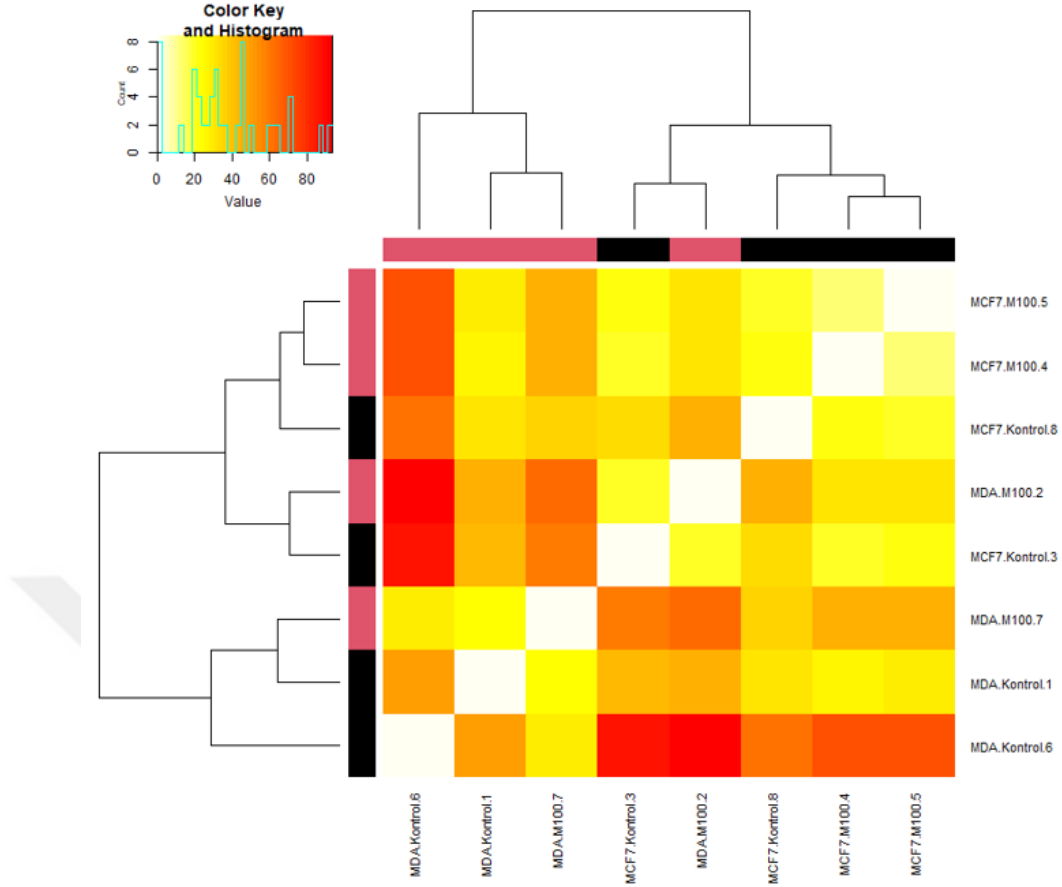
Niceliksel normalizasyon işlemi ile veri 0-1 arasına ölçeklendirilmiştir. Normalizasyon işlemi sonrasında sinyal dağılımı ve ısı haritası tekrar hazırlanarak değişim incelenmiştir. Niceliksel normalizasyon sonrası sinyal dağılımı Şekil 4.10'daki gibidir.

#### Niceliksel (Quantile) Normalizasyon



**Şekil 4.10:** Normalizasyon Sonrası Sinyal Dağılımı – M100

Normalizasyon sonrası ısı haritası Şekil 4.11'deki gibidir.



**Şekil 4.11:** Normalizasyon Sonrası Kümelmiş Veri Isı Haritası – M100

Normalizasyon işleminin ardından karşılaştırmaların yapılabilmesi için gruplar hazırlanmıştır:

- MCF7wsMDA = MCF7.Kontrol – MDA.Kontrol
- M100MDA = MDA.M100 – MDA.Kontrol
- M100MCF7 = MCF7.M100 – MCF7.Kontrol
- M100MDAnotM100MCF7 = (MDA.M100-MDA.Kontrol)-(MCF7.M100-MCF7.Kontrol)

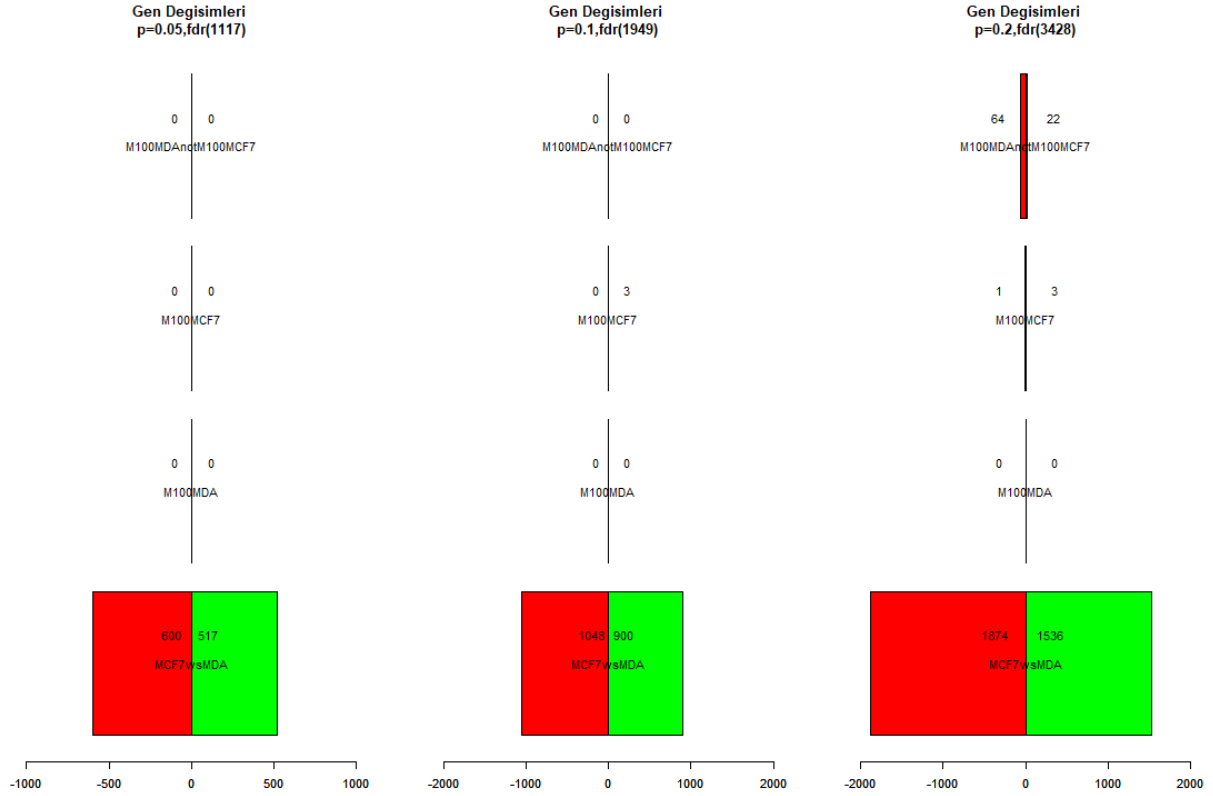
Karşılaştırma grupları için hazırlanan kontrast matrisi Tablo 4.8’de gösterilmektedir.

**Tablo 4.8:** Kontrast Matrisi – M100

Seviye	MCF7wsMDA	M100MDA	M100MCF7	M100MDA notM100MCF7
MDA.Kontrol	-1	-1	0	-1
MDA.M100	0	1	0	1
MCF7.Kontrol	1	0	-1	1
MCF7.M100	0	0	1	-1

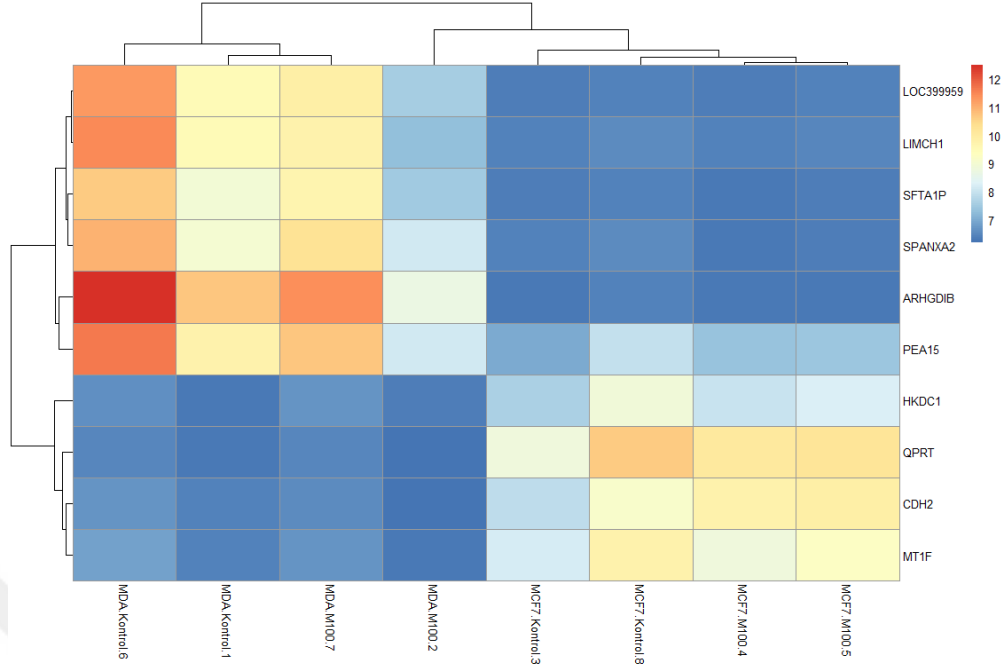
Karşılaştırma gruplarının hazırlanması sonrası SMATP yönteminin son aşaması uygulanarak veri değerlendirilmiştir. Değerlendirme sürecinde eBayes kullanılarak istatistiksel

önem değerlendirilmiş, FDR istatistiği ile gözlem yapılmıştır. FDR’de sıklıkla kullanılan p-değeri ile kontroller yapılmıştır. P-değerinin 0’a yakın olması sonucun anlamlı olduğunu göstermektedir. 3 farklı eşik değeri için kontroller yapıldığında gen değişimleri Şekil 4.12’deki gibi olmuştur. Farklı istatistiksel eşiklerde diferansiyel olarak eksprese edilen genlerin sayısının gösterildiği şekilde yeşil çubuklar yukarı düzenlenmiş genleri, kırmızı çubuklar ise aşağı düzenlenmiş genleri göstermektedir.



Şekil 4.12: Gruplar Arası Karşılaştırmalar – M100

Sonuçlar incelendiğinde M100 veriseti için en çok etkilenen 10 gen aşağıdaki gibi olmuştur.



**Şekil 4.13:** En Çok Etkilenen 10 Gen – M100

#### 4.1.3. Deney Grubu 3 – M1000

MCF7 ve MDA-N1 meme kanseri hücre soylarına M1000 (1000 milimolar metformin) uygulanması ile hazırlanmış olan deneye ait veriseti 65 kolon ve 34694 satır içermektedir. Veriseti 4 M1000 uygulanmış örnek, 4 kontrol örneği olmak üzere toplam 8 örnekten oluşmaktadır. M1000 deneyinde SMATP yöntemi aşağıdaki gibi uygulanmıştır.

İlk adım olarak veriseti bilgisayar ortamına aktarılmıştır. Ardından verinin değerlendirilebilmesi için veri çerçevesi hazırlanmıştır. Hazırlanan veri çerçevesi Tablo 4.9’da gösterildiği gibidir.

**Tablo 4.9:** Veri Çerçevesi – M1000

Sıra	Grup	GenTipi	Replika	Tedavi	DiziKodu
1	1	MDA	1	Kontrol	MDA_K_1
2	1	MDA	1	M1000	MDA_M1000_1
3	1	MCF7	1	Kontrol	MCF7_K_1
4	1	MCF7	1	M1000	MCF7_M1000_1
5	2	MCF7	2	M1000	MCF7_M1000_2
6	3	MDA	2	Kontrol	MDA_K_2
7	3	MDA	2	M1000	MDA_M1000_2
8	3	MCF7	2	Kontrol	MCF7_K2

Veri çerçevesi yardımıyla M1000 deneyindeki örneklerin sıralaması Tablo 4.10’da gösterildiği gibi olmuştur.

**Tablo 4.10:** Örneklerin Sıralanması – M1000

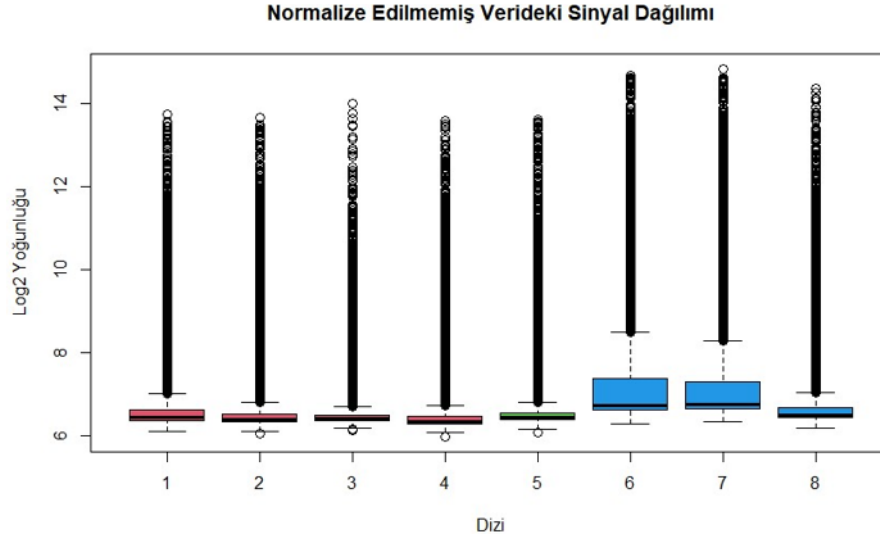
Sıra No	Örnek Adı
1	MDA.Kontrol.1
2	MDA.M1000.2
3	MCF.Kontrol.3
4	MCF7.M1000.4
5	MCF7.M1000.5
6	MDA.Kontrol.6
7	MDA.M1000.7
8	MCF7.Kontrol.8

Sıralanmış verilerin log2 dönüşümü yapıldıktan sonra oluşturulan tasarım matrisi Tablo 4.11’de gösterilmektedir.

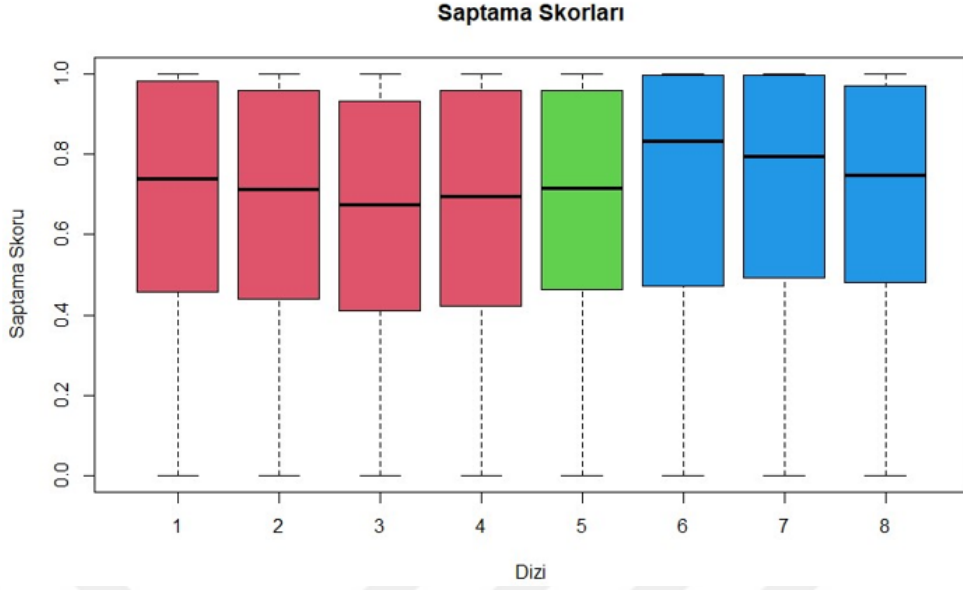
**Tablo 4.11:** Tasarım Matrisi – M1000

Örnek	Min	1st Qu.	Median	Mean	3rd Qu.	Max
MDA.Kontrol.1	6.102	6.385	6.451	6.733	6.636	13.722
MDA.M1000.2	6.075	6.350	6.411	6.617	6.533	13.654
MCF.Kontrol.3	6.150	6.384	6.435	6.554	6.515	13.980
MCF7.M1000.4	5.986	6.300	6.354	6.559	6.474	13.579
MCF7.M1000.5	6.094	6.394	6.450	6.648	6.564	13.598
MDA.Kontrol.6	6.306	6.634	6.736	7.311	7.376	14.663
MDA.M1000.7	6.350	6.661	6.764	7.290	7.312	14.825
MCF7.Kontrol.8	6.192	6.445	6.513	6.767	6.682	14.356

İşlenmemiş/ham verideki sinyal dağılımı kontrol edilmiştir. Normalizasyon işlemi öncesindeki sinyal dağılımı Şekil 4.14’deki kutu grafik ile gösterildiği gibidir.

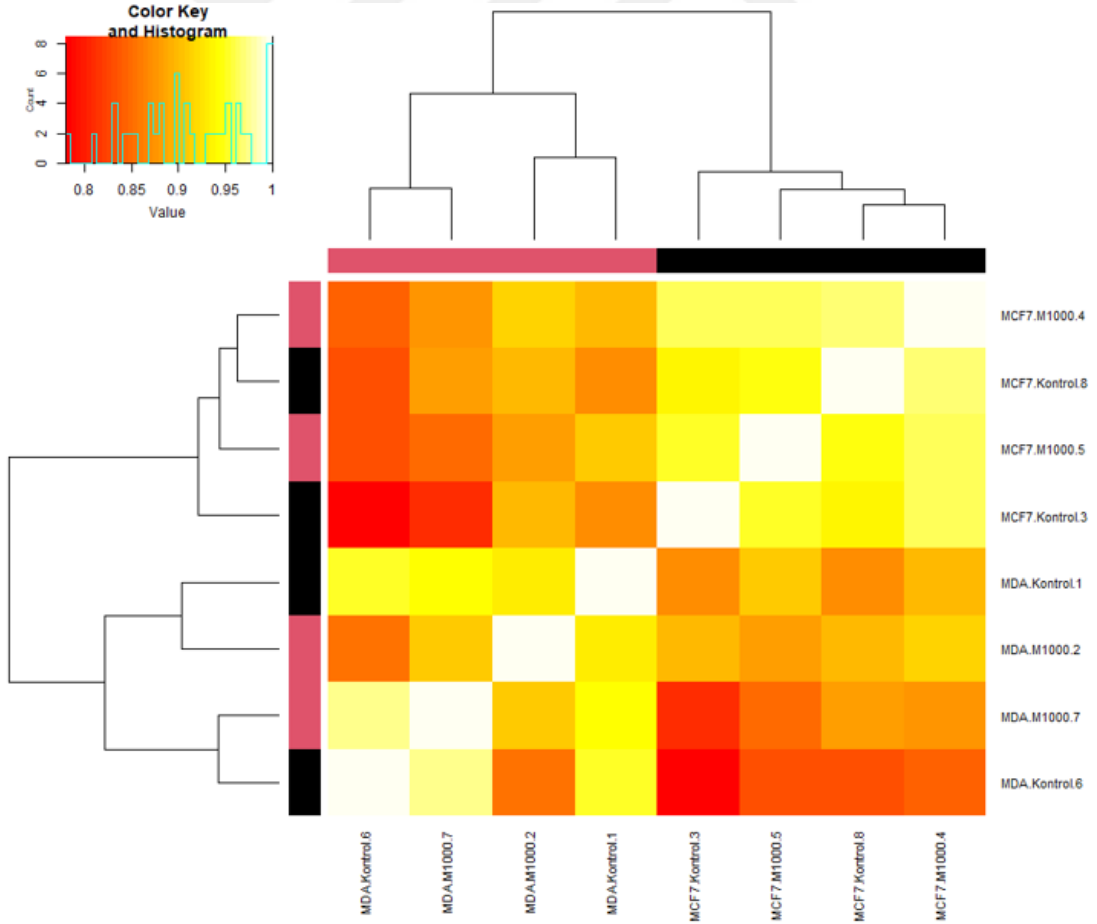
**Şekil 4.14:** Normalizasyon Öncesi Sinyal Dağılımı – M1000

Sinyal dağılımının kontrolünün ardından saptama skoru kontrol işlemi gerçekleştirilerek ham verideki genlerin 0-1 arasında ifade edilme olasılığı Şekil 4.15’de gösterilmiştir. (0: ifade edilmiyor, 1: ifade ediliyor)



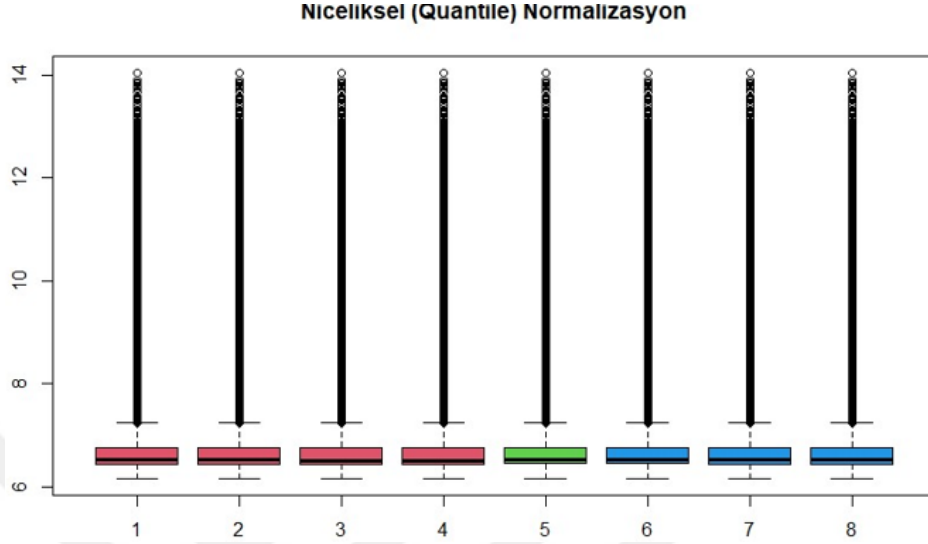
**Şekil 4.15:** Saptama Skorları – M1000

Veri kalitesi kontrolleri tamamlanan veriseti üzerinde pearson korelasyona bağlı kümeleme işlemi yapılmıştır. Kümelenmiş veri Şekil 4.16'daki ısı haritasındaki gibidir.



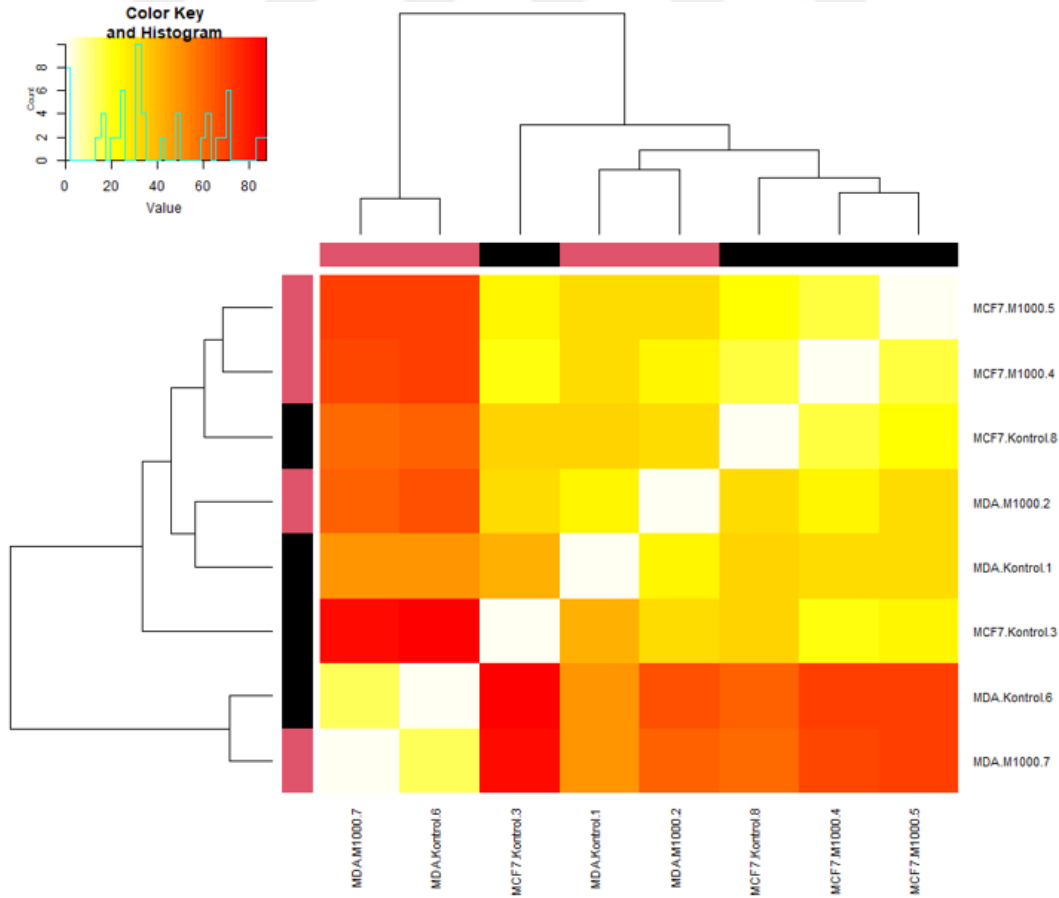
**Şekil 4.16:** Normalizasyon Öncesi Kümeleme Isı Haritası - M1000

Niceliksel normalizasyon işlemi ile veri 0-1 arasına ölçeklendirilmiştir. Normalizasyon işlemi sonrasında sinyal dağılımı ve ısı haritası tekrar hazırlanarak değişim incelenmiştir. Niceliksel normalizasyon sonrası sinyal dağılımı Şekil 4.17'deki gibidir.



**Şekil 4.17:** Normalizasyon Sonrası Sinyal Dağılımı – M1000

Normalizasyon sonrası ısı haritası Şekil 4.18'deki gibidir.



**Şekil 4.18:** Normalizasyon Sonrası Kümelmiş Veri Isı Haritası – M1000

Normalizasyon işleminin ardından karşılaştırmaların yapılabilmesi için gruplar hazırlanmıştır:

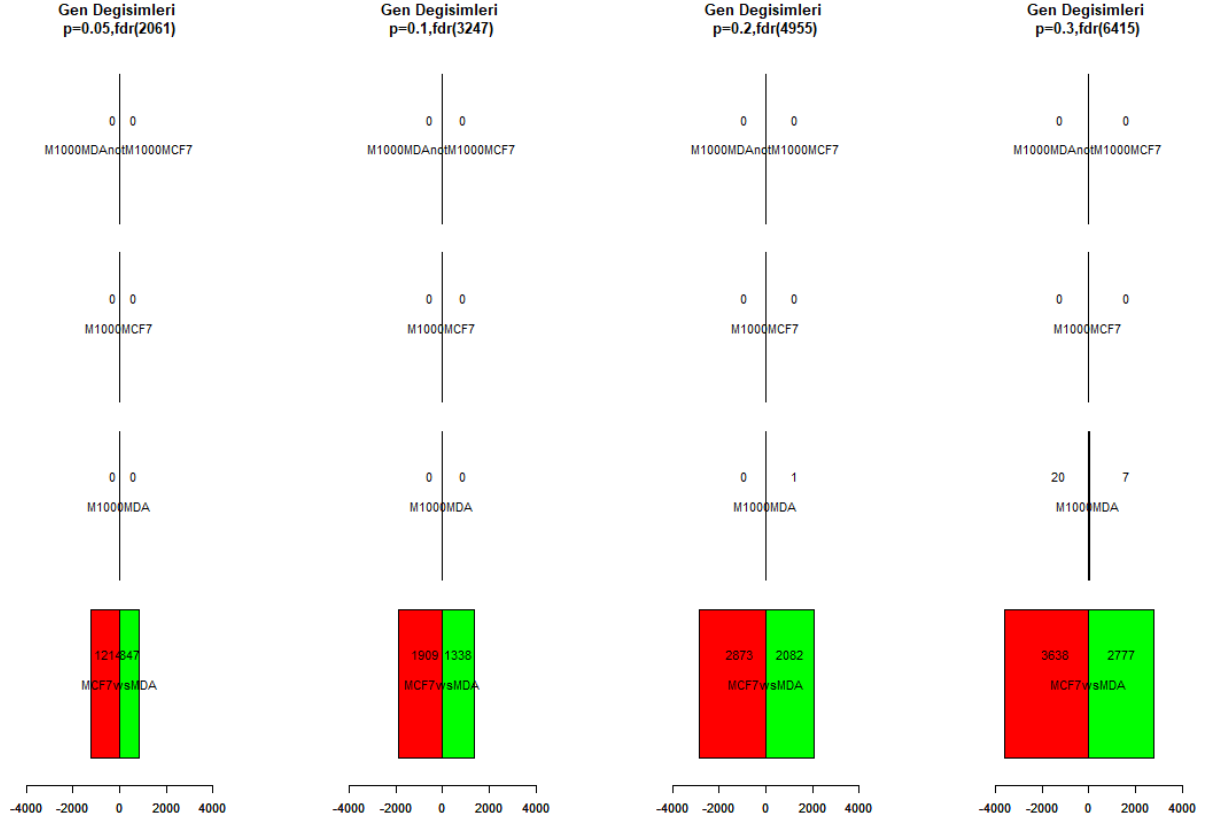
- $MCF7wsMDA = MCF7.Kontrol - MDA.Kontrol$
- $M1000MDA = MDA.M1000 - MDA.Kontrol$
- $M1000MCF7 = MCF7.M1000 - MCF7.Kontrol$
- $M1000MDAnotM1000MCF7 = (MDA.M1000-MDA.Kontrol)-(MCF7.M1000-MCF7.Kontrol)$

Karşılaştırma grupları için hazırlanan kontrast matrisi Tablo 4.12’de gösterilmektedir.

**Tablo 4.12:** Kontrast Matrisi – M1000

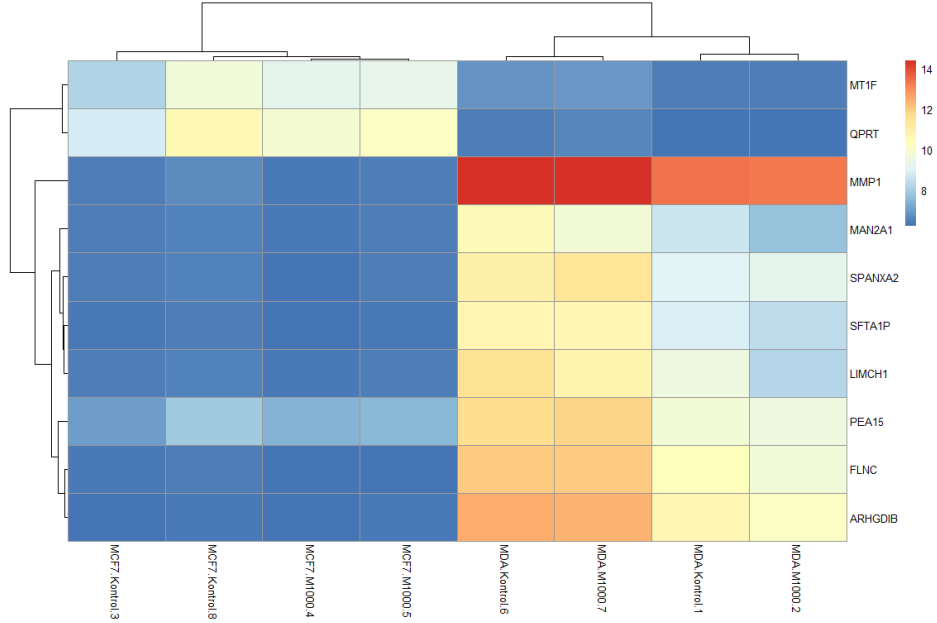
Seviye	MCF7wsMDA	M1000MDA	M1000MCF7	M1000MDA notM1000MCF7
MDA.Kontrol	-1	-1	0	-1
MDA.M1000	0	1	0	1
MCF7.Kontrol	1	0	-1	1
MCF7.M1000	0	0	1	-1

Karşılaştırma gruplarının hazırlanması sonrası SMATP yönteminin son aşaması uygulanarak veri değerlendirilmiştir. Değerlendirme sürecinde eBayes kullanılarak istatistiksel önem değerlendirilmiş, FDR istatistiği ile gözlem yapılmıştır. FDR’de sıklıkla kullanılan p-değeri ile kontroller yapılmıştır. P-değerinin 0’a yakın olması sonucun anlamlı olduğunu göstermektedir. 4 farklı eşik değeri için kontroller yapıldığında gen değişimleri Şekil 4.19’daki gibi olmuştur. Farklı istatistiksel eşiklerde diferansiyel olarak ekprese edilen genlerin sayısının gösterildiği şekilde yeşil çubuklar yukarı düzenlenmiş genleri, kırmızı çubuklar ise aşağı düzenlenmiş genleri göstermektedir.



Şekil 4.19: Gruplar Arası Karşılaştırmalar – M1000

Sonuçlar incelendiğinde M1000 veriseti için en çok etkilenen 10 gen aşağıdaki gibi olmuştur.



Şekil 4.20: En Çok Etkilenen 10 Gen – M1000

#### 4.1.4. Deney Grubu 4 – Işın

MCF7 ve MDA-N1 meme kanseri hücre soylarına Işın uygulanması ile hazırlanmış olan deneye ait veriseti 57 kolon ve 34694 satır içermektedir. Veriseti 3 Işın uygulanmış örnek, 4 kontrol örneği olmak üzere toplam 7 örnekten oluşmaktadır. Işın deneyinde SMATP yöntemi aşağıdaki gibi uygulanmıştır.

İlk adım olarak veriseti bilgisayar ortamına aktarılmıştır. Ardından verinin değerlendirilebilmesi için veri çerçevesi hazırlanmıştır. Hazırlanan veri çerçevesi Tablo 4.13’de gösterildiği gibidir.

**Tablo 4.13:** Veri Çerçevesi – Işın

Sıra	Grup	GenTipi	Replika	Tedavi	DiziKodu
1	1	MDA	1	Kontrol	MDA_K_1
2	1	MDA	1	Işın	MDA_I_1
3	1	MCF7	1	Kontrol	MCF7_K_1
4	3	MDA	2	Kontrol	MDA_K_2
5	3	MDA	2	Işın	MDA_I_2
6	3	MCF7	1	Işın	MCF7_I_1
7	3	MCF7	2	Kontrol	MCF7_K_2

Veri çerçevesi yardımıyla Işın deneyindeki örneklerin sıralaması Tablo 4.14’de gösterildiği gibi olmuştur.

**Tablo 4.14:** Örneklerin Sıralanması – Işın

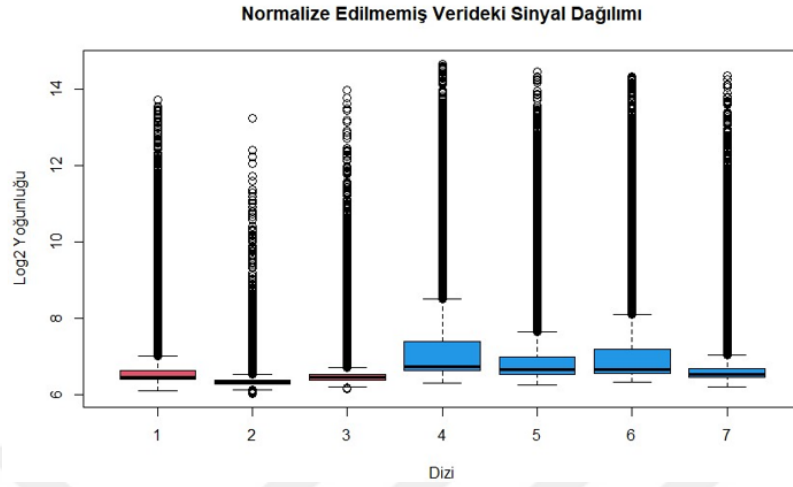
Sıra No	Örnek Adı
1	MDA.Kontrol.1
2	MDA.Isin.2
3	MCF7.Kontrol.3
4	MDA.Kontrol.4
5	MDA.Isin.5
6	MCF7.Isin.6
7	MCF7.Kontrol.7

Sıralanmış verilerin log<sub>2</sub> dönüşümü yapıldıktan sonra oluşturulan tasarım matrisi Tablo 4.15’de gösterilmektedir.

**Tablo 4.15:** Tasarım Matrisi – Işın

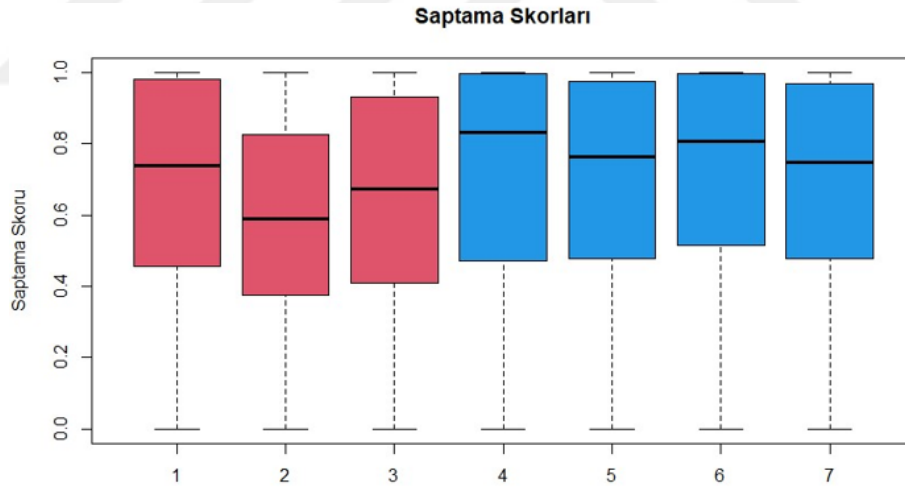
Örnek	Min	1st Qu.	Median	Mean	3rd Qu.	Max
MDA.Kontrol.1	6.102	6.385	6.451	6.733	6.636	13.722
MDA.Isin.2	6.018	6.269	6.313	6.354	6.366	13.234
MCF7.Kontrol.3	6.150	6.384	6.435	6.554	6.515	13.980
MDA.Kontrol.4	6.306	6.634	6.736	7.311	7.376	14.663
MDA.Isin.5	6.252	6.533	6.638	6.993	6.970	14.454
MCF7.Isin.6	6.309	6.561	6.653	7.142	7.176	14.328
MCF7.Kontrol.7	6.192	6.445	6.513	6.767	6.682	14.356

İşlenmemiş/ham verideki sinyal dağılımı kontrol edilmiştir. Normalizasyon işlemi öncesindeki sinyal dağılımı Şekil 4.21’deki kutu grafik ile gösterildiği gibidir.



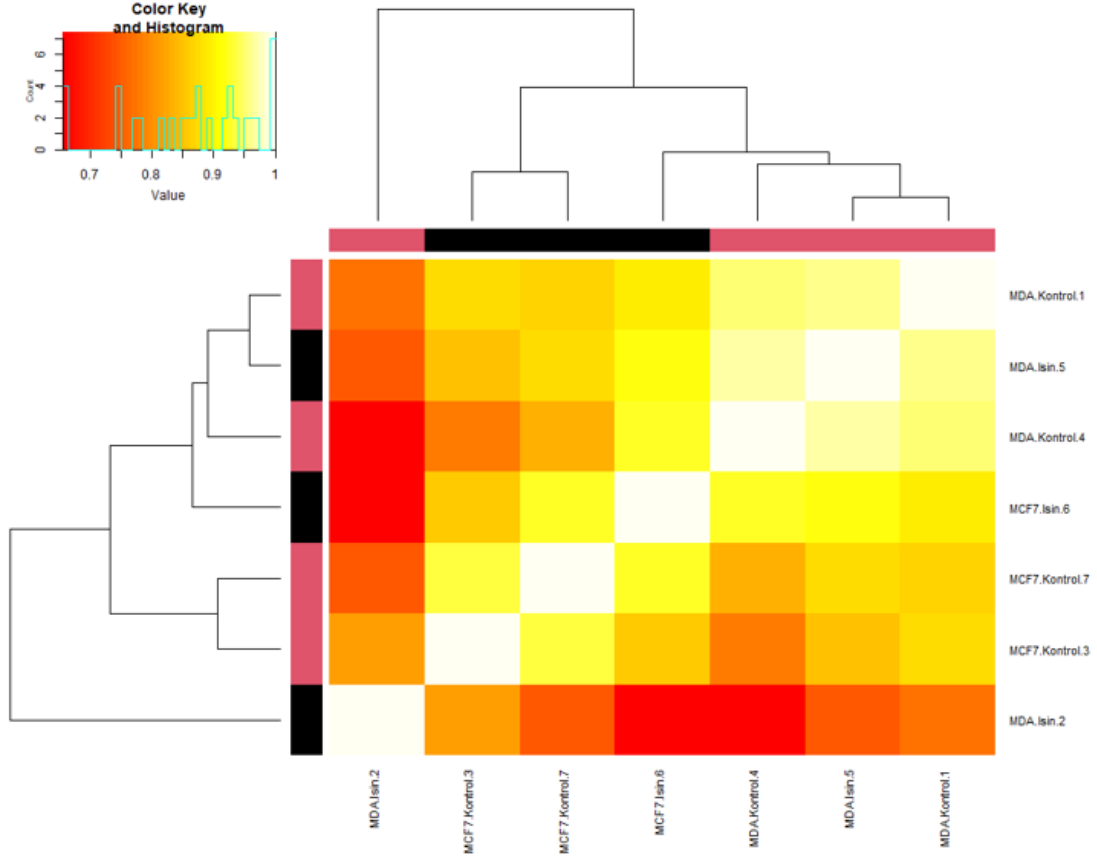
**Şekil 4.21:** Normalizasyon Öncesi Sinyal Dağılımı – Işın

Sinyal dağılımının kontrolünün ardından saptama skoru kontrol işlemi gerçekleştirilerek ham verideki genlerin 0-1 arasında ifade edilme olasılığı Şekil 4.22’da gösterilmiştir. (0: ifade edilmiyor, 1: ifade ediliyor)



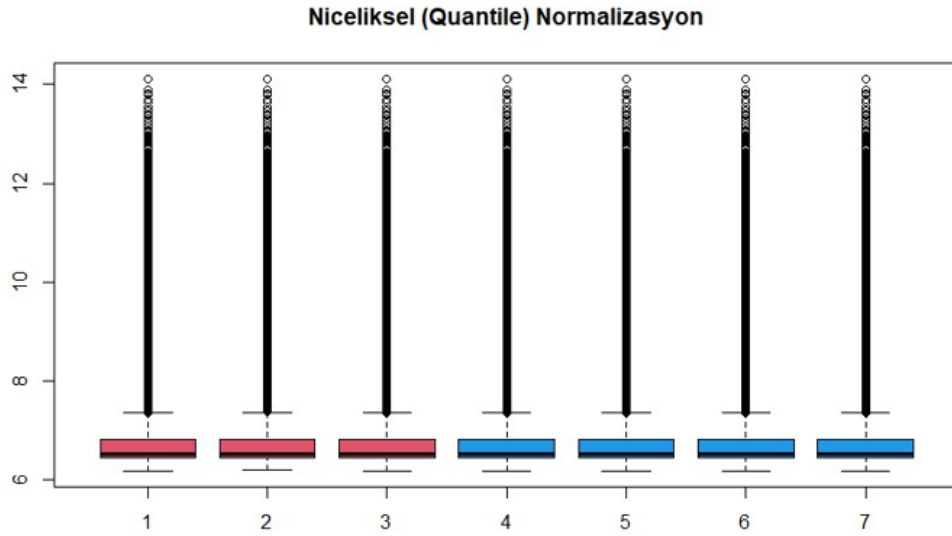
**Şekil 4.22:** Saptama Skorları – Işın

Veri kalitesi kontrolleri tamamlanan veriseti üzerinde pearson korelasyona bağlı kümeleme işlemi yapılmıştır. Kümelenmiş veri Şekil 4.23’deki ısı haritasındaki gibidir.



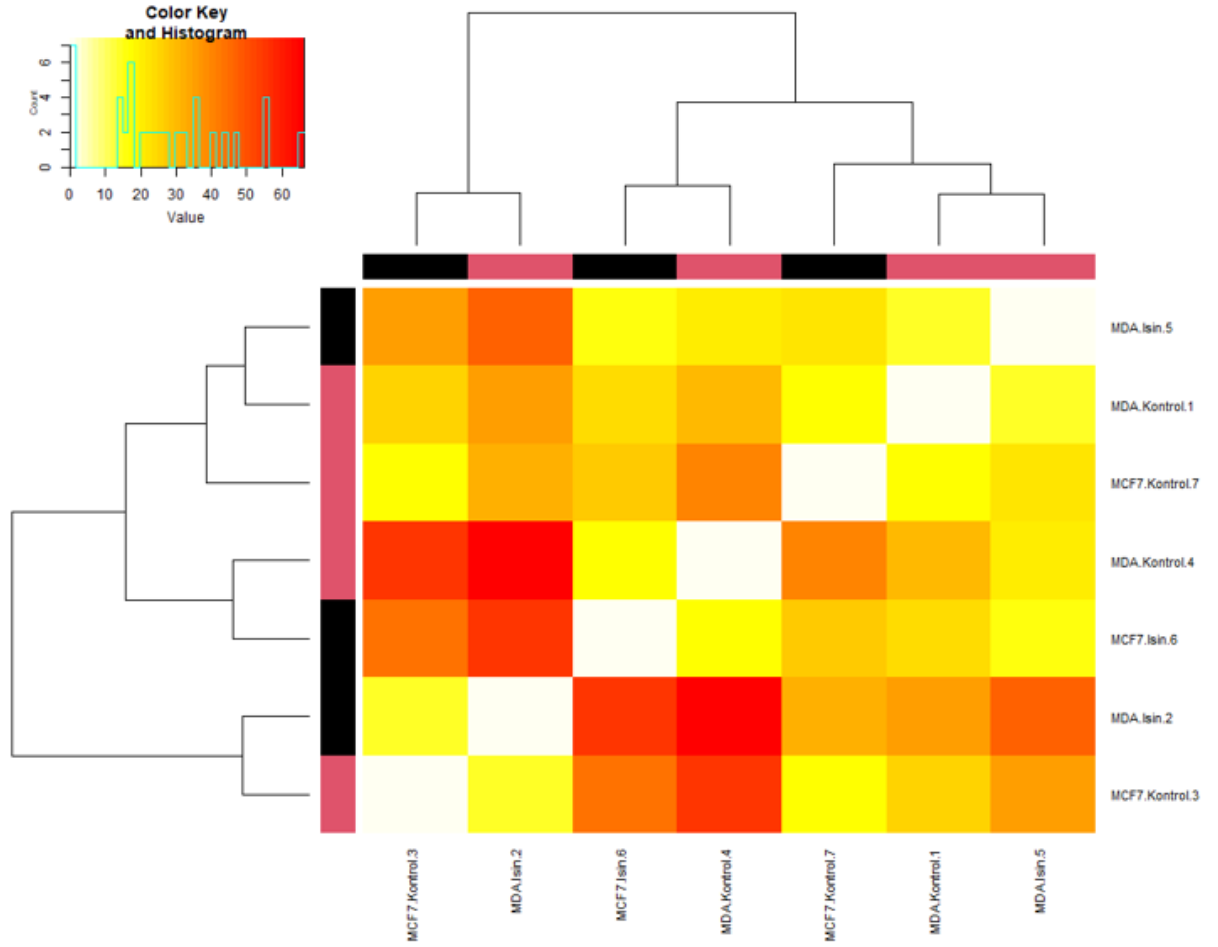
**Şekil 4.23:** Normalizasyon Öncesi Kümeleme Isı Haritası - Işın

Niceliksel normalizasyon işlemi ile veri 0-1 arasına ölçeklendirilmiştir. Normalizasyon işlemi sonrasında sinyal dağılımı ve ısı haritası tekrar hazırlanarak değişim incelenmiştir. Niceliksel normalizasyon sonrası sinyal dağılımı Şekil 4.24'deki gibidir.



**Şekil 4.24:** Normalizasyon Sonrası Sinyal Dağılımı – Işın

Normalizasyon sonrası ısı haritası Şekil 4.25'deki gibidir.



**Şekil 4.25:** Normalizasyon Sonrası Kümelenmiş Veri Isı Haritası – Işın

Normalizasyon işleminin ardından karşılaştırmaların yapılabilmesi için gruplar hazırlanmıştır:

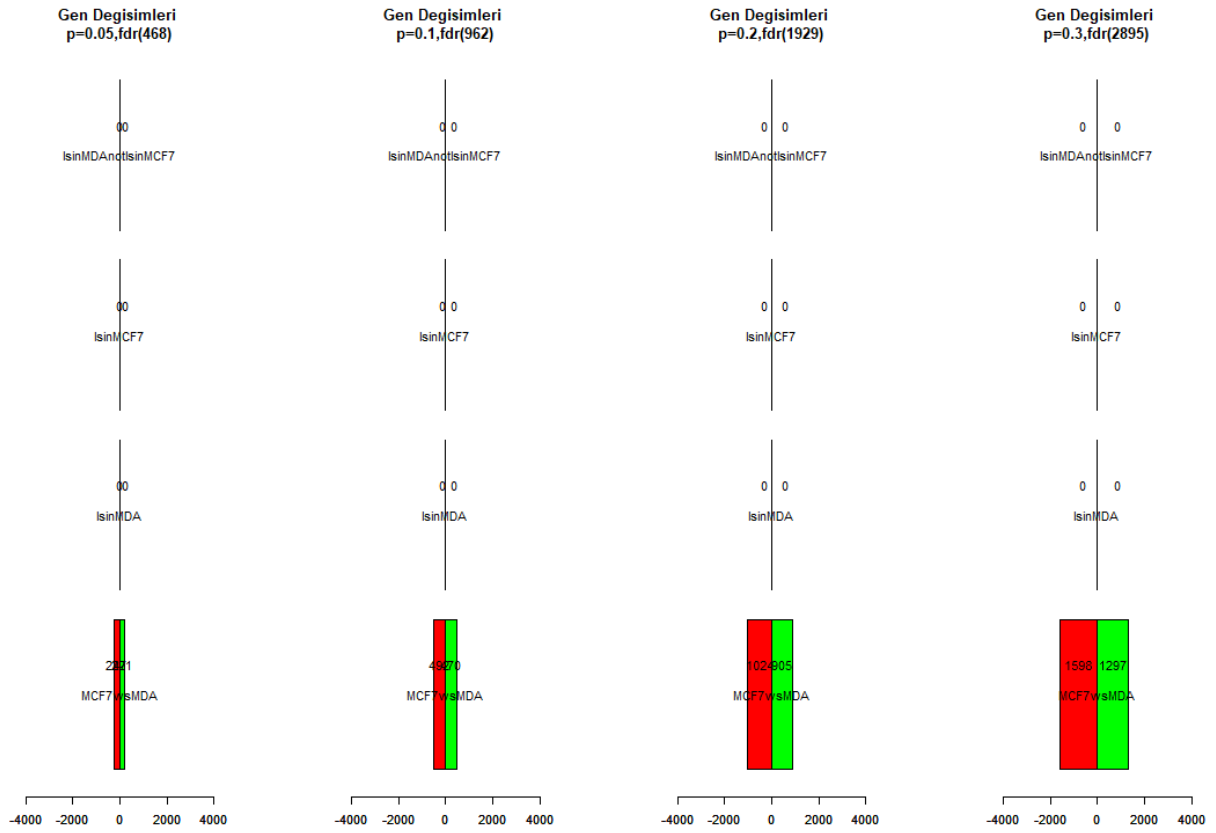
- $MCF7wsMDA = MCF7.Kontrol - MDA.Kontrol$
- $IsinMDA = MDA.Isin - MDA.Kontrol$
- $IsinMCF7 = MCF7.Isin - MCF7.Kontrol$
- $IsinMDAnotIsinMCF7 = (MDA.Isin - MDA.Kontrol) - (MCF7.Isin - MCF7.Kontrol)$

Karşılaştırma grupları için hazırlanan kontrast matrisi Tablo 4.16’da gösterilmektedir.

**Tablo 4.16:** Kontrast Matrisi – Işın

Seviye	MCF7wsMDA	IsinMDA	IsinMCF7	IsinMDA notIsinMCF7
MDA.Kontrol	-1	-1	0	-1
MDA.Isin	0	1	0	1
MCF7.Kontrol	1	0	-1	1
MCF7.Isin	0	0	1	-1

Karşılaştırma gruplarının hazırlanması sonrası SMATP yönteminin son aşaması uygulanarak veri değerlendirilmiştir. Değerlendirme sürecinde eBayes kullanılarak istatistiksel önem değerlendirilmiş, FDR istatistiği ile gözlem yapılmıştır. FDR’de sıklıkla kullanılan p-değeri ile kontroller yapılmıştır. P-değerinin 0’a yakın olması sonucun anlamlı olduğunu göstermektedir. 4 farklı eşik değeri için kontroller yapıldığında gen değişimleri Şekil 4.26’daki gibi olmuştur. Farklı istatistiksel eşiklerde diferansiyel olarak eksprese edilen genlerin sayısının gösterildiği şekilde yeşil çubuklar yukarı düzenlenmiş genleri, kırmızı çubuklar ise aşağı düzenlenmiş genleri göstermektedir.



Şekil 4.26: Gruplar Arası Karşılaştırmalar – Işın

Sonuçlar incelendiğinde kontrol grubunda (hiçbir tedavi uygulanmamış kanser hücre soyu kontrol örnekleri) halihazırda aktive olan genler bulunurken Işın uygulanan örneklerde belirgin bir gen değişimi olmadığı gözlemlenmiştir.

#### 4.1.5. Deney Grubu 5 – Nikotin ve M100

MCF7 ve MDA-N1 meme kanseri hücre soylarına Nikotin ve M100’ün birlikte uygulanması ile hazırlanmış olan deneye ait veriseti 57 kolon ve 34694 satır içermektedir. Veriseti 3 Nikotin ve M100’ün bir arada uygulandığı örnek, 4 kontrol örneği olmak üzere

toplam 7 örnekten oluşmaktadır. Nikotin ve M100 deneyinde SMATP yöntemi aşağıdaki gibi uygulanmıştır.

İlk adım olarak veriseti bilgisayar ortamına aktarılmıştır. Ardından verinin değerlendirilebilmesi için veri çerçevesi hazırlanmıştır. Hazırlanan veri çerçevesi Tablo 4.17’de gösterildiği gibidir.

**Tablo 4.17:** Veri Çerçevesi – Nikotin ve M100

Sıra	Grup	GenTipi	Replika	Tedavi	DiziKodu
1	1	MDA	1	Kontrol	MDA_K_1
2	1	MDA	1	N_M100	MDA_N_M100_1
3	1	MCF7	1	Kontrol	MCF7_K_1
4	2	MCF7	2	N_M100	MCF7_N_M100_1
5	3	MDA	2	Kontrol	MDA_K_2
6	3	MDA	1	N_M100	MDA_N_M100_2
7	3	MCF7	2	Kontrol	MCF7_K_2

Veri çerçevesi yardımıyla Nikotin ve M100 deneyindeki örneklerin sıralaması Tablo 4.18’de gösterildiği gibi olmuştur.

**Tablo 4.18:** Örneklerin Sıralanması – Nikotin ve M100

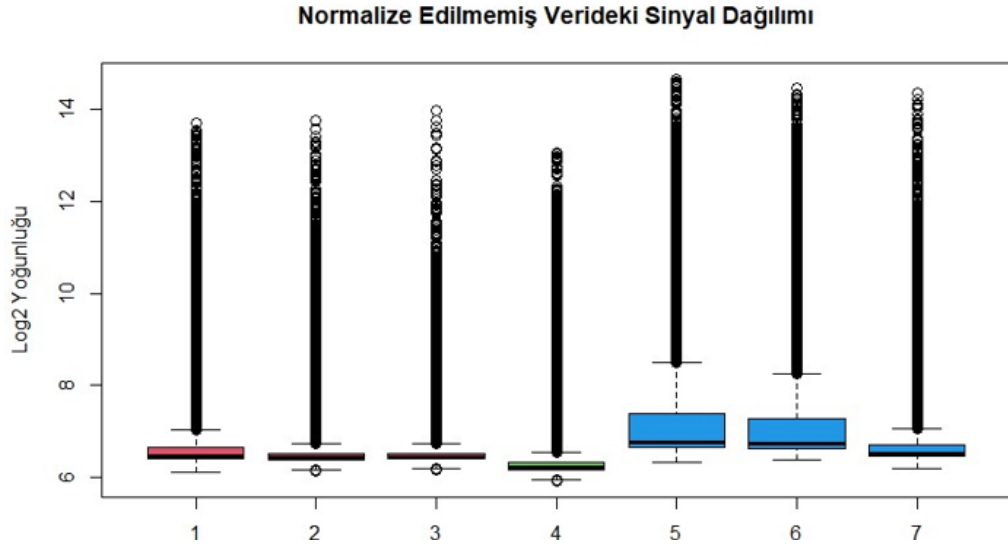
Sıra No	Örnek Adı
1	MDA.Kontrol.1
2	MDA.N_M100.2
3	MCF7.Kontrol.3
4	MCF7.N_M100.4
5	MDA.Kontrol.5
6	MDA.N_M100.6
7	MCF7.Kontrol.7

Sıralanmış verilerin log2 dönüşümü yapıldıktan sonra oluşturulan tasarım matrisi Tablo 4.19’da gösterilmektedir.

**Tablo 4.19:** Tasarım Matrisi – Nikotin ve M100

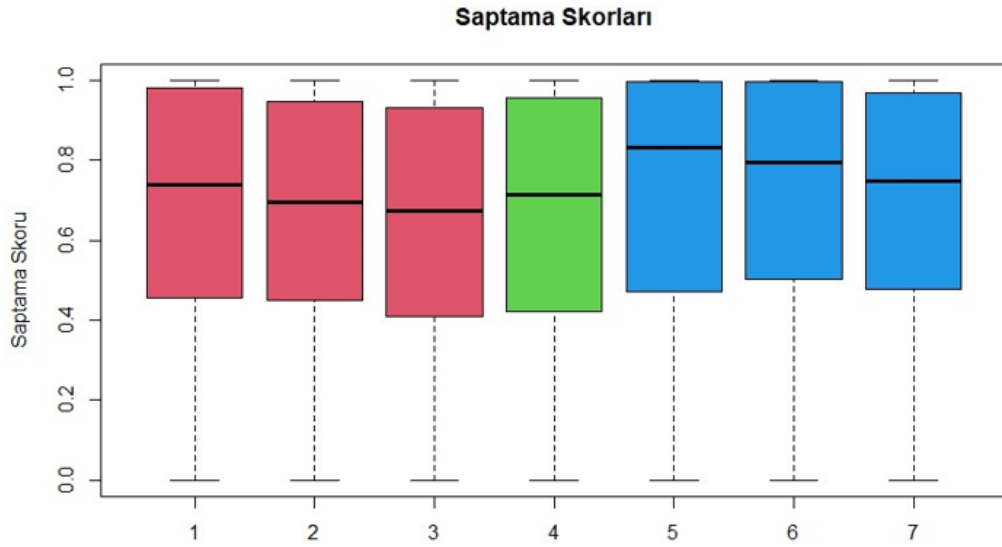
Örnek	Min	1st Qu.	Median	Mean	3rd Qu.	Max
MDA.Kontrol.1	6.102	6.385	6.451	6.733	6.636	13.722
MDA.N_M100.2	6.117	6.372	6.421	6.596	6.511	13.773
MCF7.Kontrol.3	6.150	6.384	6.435	6.554	6.515	13.980
MCF7.N_M100.4	5.914	6.156	6.208	6.407	6.311	13.062
MDA.Kontrol.5	6.306	6.634	6.736	7.311	7.376	14.663
MDA.N_M100.6	6.358	6.609	6.710	7.226	7.262	14.462
MCF7.Kontrol.7	6.192	6.445	6.513	6.767	6.682	14.356

İşlenmemiş/ham verideki sinyal dağılımı kontrol edilmiştir. Normalizasyon işlemi öncesindeki sinyal dağılımı Şekil 4.27’deki kutu grafik ile gösterildiği gibidir.



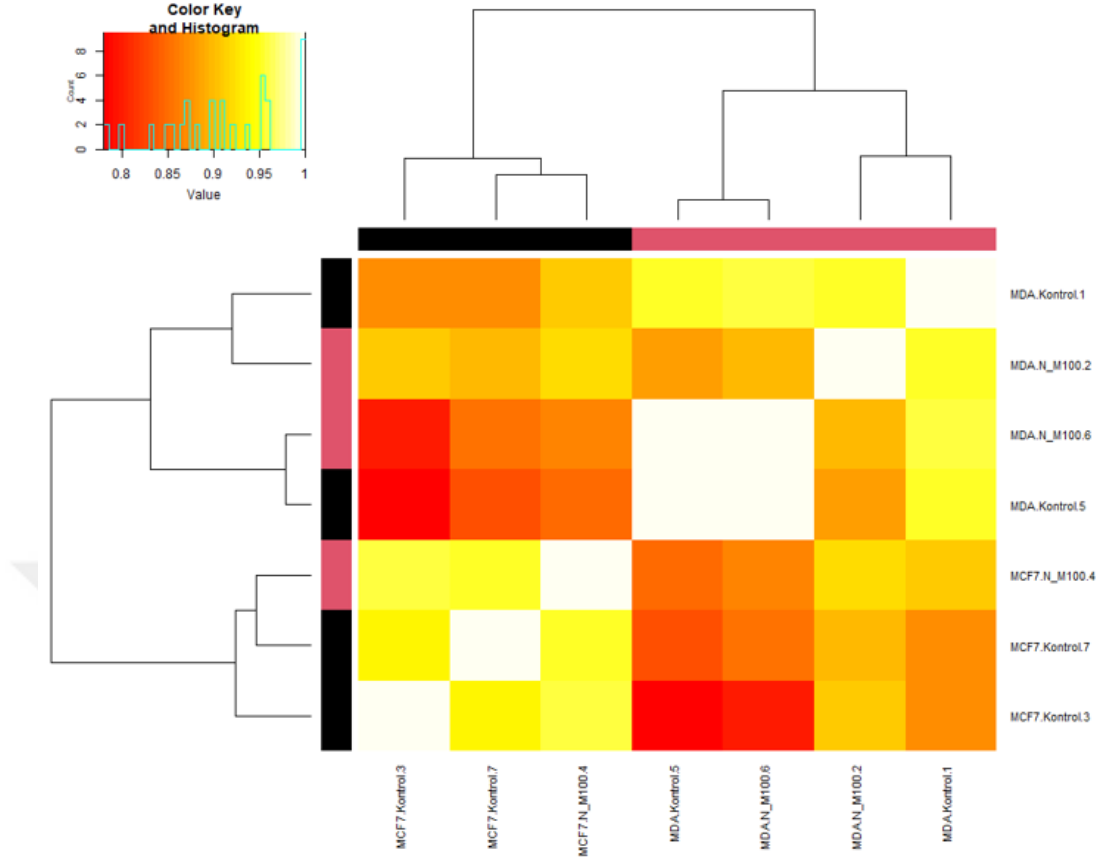
**Şekil 4.27:** Normalizasyon Öncesi Sinyal Dağılımı – Nikotin ve M100

Sinyal dağılımının kontrolünün ardından saptama skoru kontrol işlemi gerçekleştirilerek ham verideki genlerin 0-1 arasında ifade edilme olasılığı Şekil 4.28’de gösterilmiştir. (0: ifade edilmiyor, 1: ifade ediliyor)



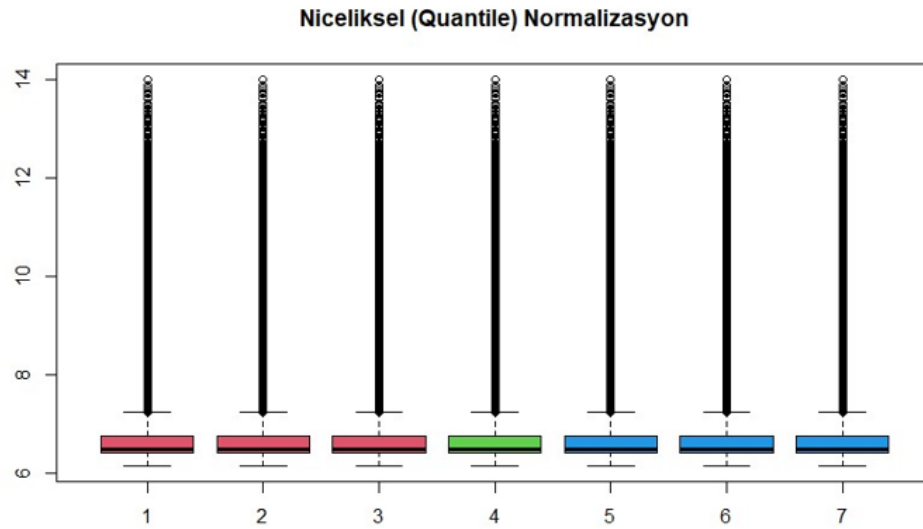
**Şekil 4.28:** Saptama Skorları – Nikotin ve M100

Veri kalitesi kontrolleri tamamlanan veriseti üzerinde pearson korelasyona bağlı kümeleme işlemi yapılmıştır. Kümelenmiş veri Şekil 4.29’daki ısı haritasındaki gibidir.



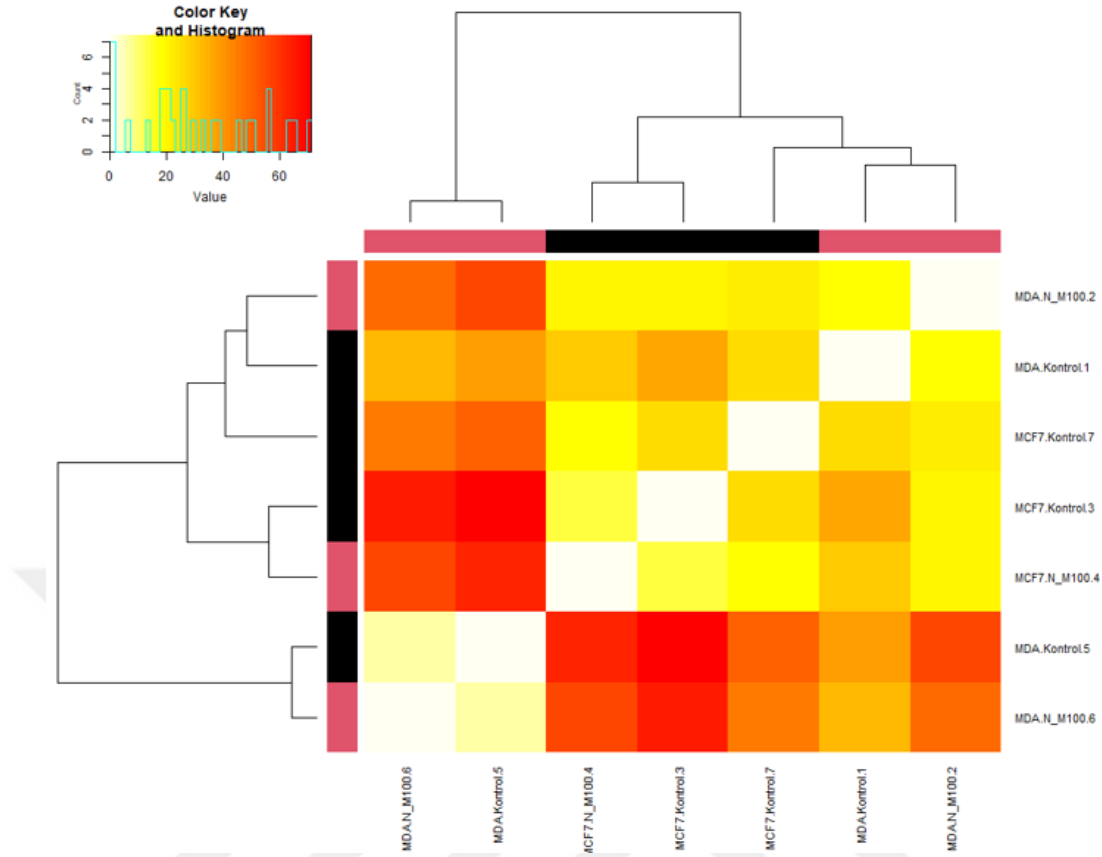
Şekil 4.29: Normalizasyon Öncesi Kümeleme ısı Haritası – Nikotin ve M100

Niceliksel normalizasyon işlemi ile veri 0-1 arasına ölçeklendirilmiştir. Normalizasyon işlemi sonrasında sinyal dağılımı ve ısı haritası tekrar hazırlanarak değişim incelenmiştir. Niceliksel normalizasyon sonrası sinyal dağılımı Şekil 4.30'deki gibidir.



Şekil 4.30: Normalizasyon Sonrası Sinyal Dağılımı – Nikotin ve M100

Normalizasyon sonrası ısı haritası Şekil 4.31'deki gibidir.



**Şekil 4.31:** Normalizasyon Sonrası Kümelmiş Veri Isı Haritası – Nikotin ve M100

Normalizasyon işleminin ardından karşılaştırmaların yapılabilmesi için gruplar hazırlanmıştır:

- MCF7wsMDA = MCF7.Kontrol – MDA.Kontrol
- N\_M100MDA = MDA.N\_M100 – MDA.Kontrol
- N\_M100MCF7 = MCF7.N\_M100 – MCF7.Kontrol
- N\_M100MDAnotN\_M100MCF7 = (MDA.N\_M100 - MDA.Kontrol) - (MCF7.N\_M100 - MCF7.Kontrol)

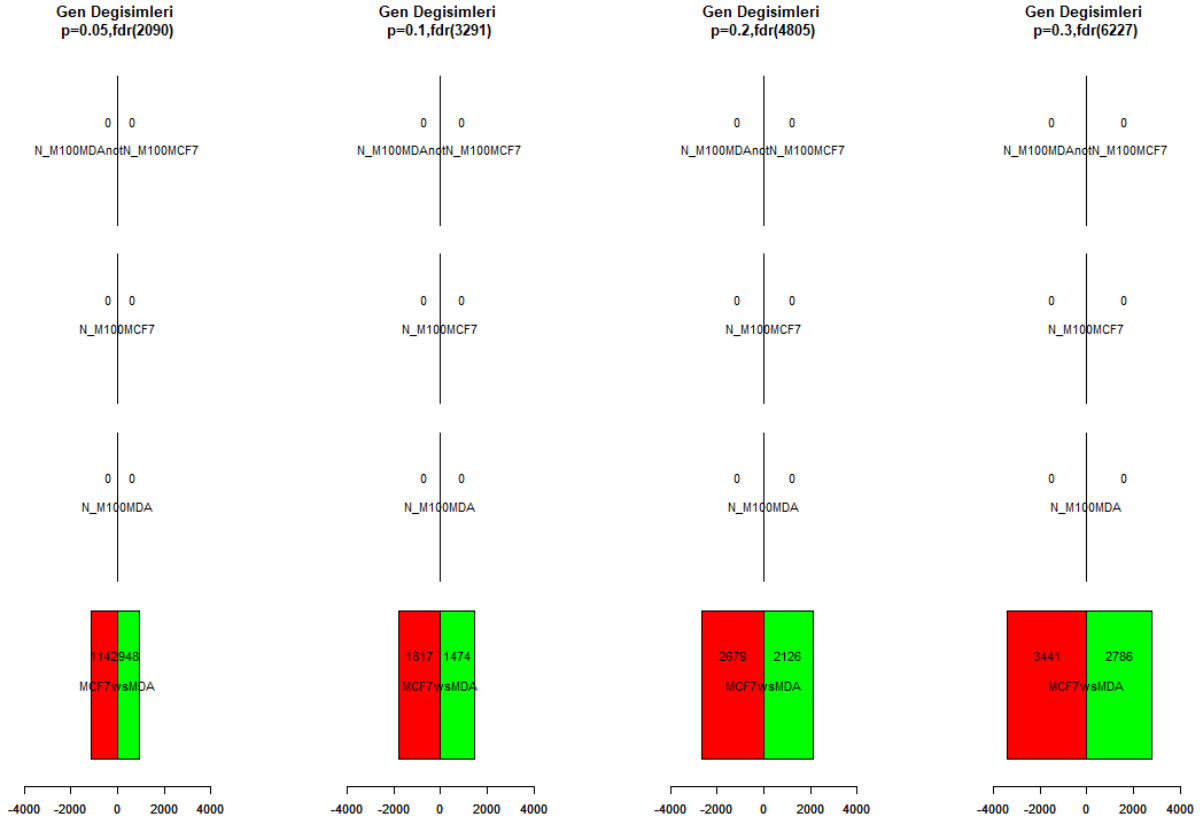
Karşılaştırma grupları için hazırlanan kontrast matrisi Tablo 4.20’de gösterilmektedir.

**Tablo 4.20:** Kontrast Matrisi – Nikotin ve M100

Seviye	MCF7wsMDA	N_M100MDA	N_M100MCF7	N_M100MDA notN_M100MCF7
MDA.Kontrol	-1	-1	0	-1
MDA.N_M100	0	1	0	1
MCF7.Kontrol	1	0	-1	1
MCF7.N_M100	0	0	1	-1

Karşılaştırma gruplarının hazırlanması sonrası SMATP yönteminin son aşaması uygulanarak veri değerlendirilmiştir. Değerlendirme sürecinde eBayes kullanılarak istatistiksel

önem değerlendirilmiş, FDR istatistiği ile gözlem yapılmıştır. FDR’de sıklıkla kullanılan p-değeri ile kontroller yapılmıştır. P-değerinin 0’a yakın olması sonucun anlamlı olduğunu göstermektedir. 4 farklı eşik değeri için kontroller yapıldığında gen değişimleri Şekil 4.32’deki gibi olmuştur. Farklı istatistiksel eşiklerde diferansiyel olarak eksprese edilen genlerin sayısının gösterildiği şekilde yeşil çubuklar yukarı düzenlenmiş genleri, kırmızı çubuklar ise aşağı düzenlenmiş genleri göstermektedir.



**Şekil 4.32:** Gruplar Arası Karşılaştırmalar – Nikotin ve M100

Sonuçlar incelendiğinde kontrol grubunda (hiçbir tedavi uygulanmamış kanser hücre soyu kontrol örnekleri) halihazırda aktive olan genler bulunurken Nikotin ve M100 birarada uygulanan örneklerde belirgin bir gen değişimi olmadığı gözlemlenmiştir.

#### 4.1.6. Deney Grubu 6 – Nikotin ve M1000

MCF7 ve MDA-N1 meme kanseri hücre soylarına Nikotin ve M1000’in bir arada uygulanması ile hazırlanmış olan deneye ait veriseti 49 kolon ve 34694 satır içermektedir. Veriseti 2 Nikotin ve M1000 bir arada uygulanmış örnek, 4 kontrol örneği olmak üzere toplam 6 örnekten oluşmaktadır. Nikotin ve M1000 deneyinde SMATP yöntemi aşağıdaki gibi uygulanmıştır.

İlk adım olarak veriseti bilgisayar ortamına aktarılmıştır. Ardından verinin değerlendirilebilmesi için veri çerçevesi hazırlanmıştır. Hazırlanan veri çerçevesi Tablo 4.21’de gösterildiği gibidir.

**Tablo 4.21:** Veri Çerçevesi – Nikotin ve M1000

Sıra	Grup	GenTipi	Replika	Tedavi	DiziKodu
1	1	MDA	1	Kontrol	MDA_K_1
2	1	MCF7	1	Kontrol	MCF7_K_1
3	2	MCF7	1	N_M1000	MCF7_N_M1000_1
4	3	MDA	2	Kontrol	MDA_K_2
5	3	MDA	1	N_M1000	MDA_N_M1000_1
6	3	MCF7	2	Kontrol	MCF7_K_2

Veri çerçevesi yardımıyla Nikotin ve M1000 deneyindeki örneklerin sıralaması Tablo 4.22’de gösterildiği gibi olmuştur.

**Tablo 4.22:** Örneklerin Sıralanması – Nikotin ve M1000

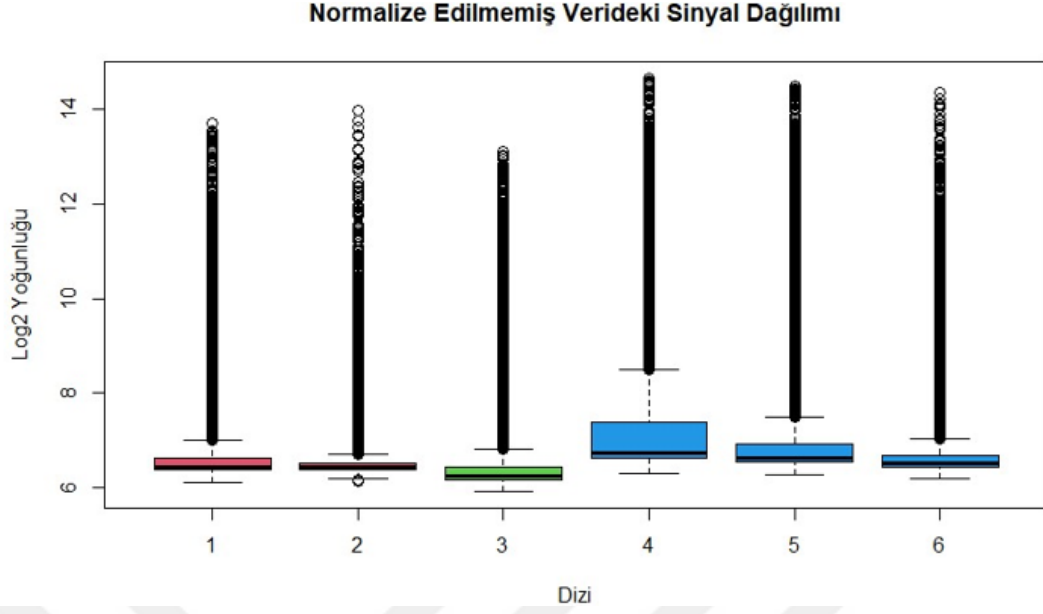
Sıra No	Örnek Adı
1	MDA.Kontrol.1
2	MCF7.Kontrol.2
3	MCF7.N_M1000.3
4	MDA.Kontrol.4
5	MDA.N_M1000.5
6	MCF7.Kontrol.6

Sıralanmış verilerin log<sub>2</sub> dönüşümü yapıldıktan sonra oluşturulan tasarım matrisi Tablo 4.23’de gösterilmektedir.

**Tablo 4.23:** Tasarım Matrisi – Nikotin ve M1000

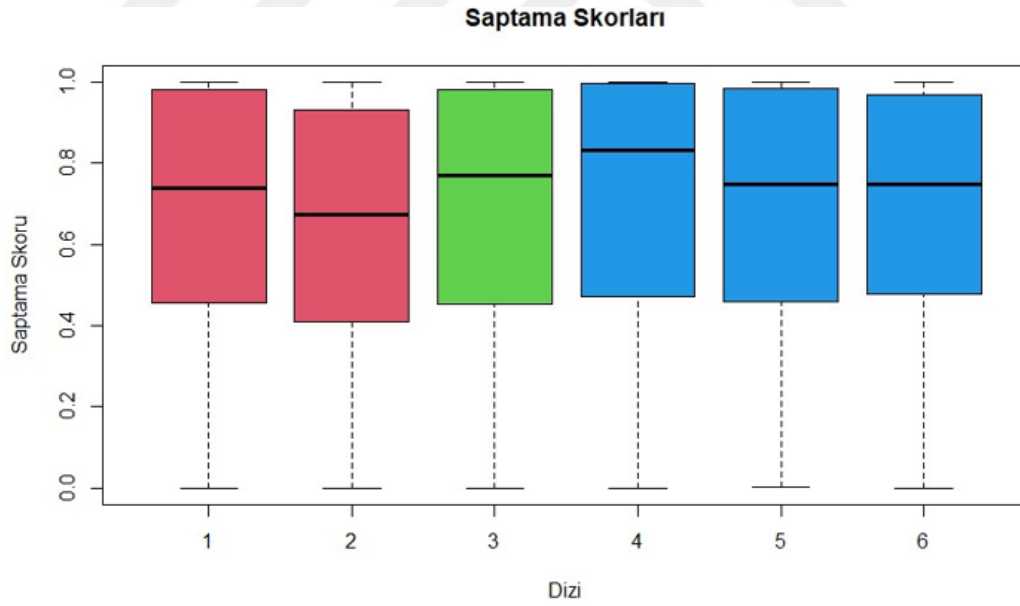
Örnek	Min	1st Qu.	Median	Mean	3rd Qu.	Max
MDA.Kontrol.1	6.102	6.385	6.451	6.733	6.636	13.722
MCF.Kontrol.2	6.150	6.384	6.435	6.554	6.515	13.980
MCF7.N_M1000.3	5.931	6.180	6.240	6.555	6.432	13.125
MDA.Kontrol.4	6.306	6.634	6.736	7.311	7.376	14.663
MDA.N_M1000.5	6.276	6.556	6.637	7.036	6.930	14.492
MCF7.Kontrol.6	6.192	6.445	6.513	6.767	6.682	14.356

İşlenmemiş/ham verideki sinyal dağılımı kontrol edilmiştir. Normalizasyon işlemi öncesindeki sinyal dağılımı Şekil 4.33’deki kutu grafik ile gösterildiği gibidir.



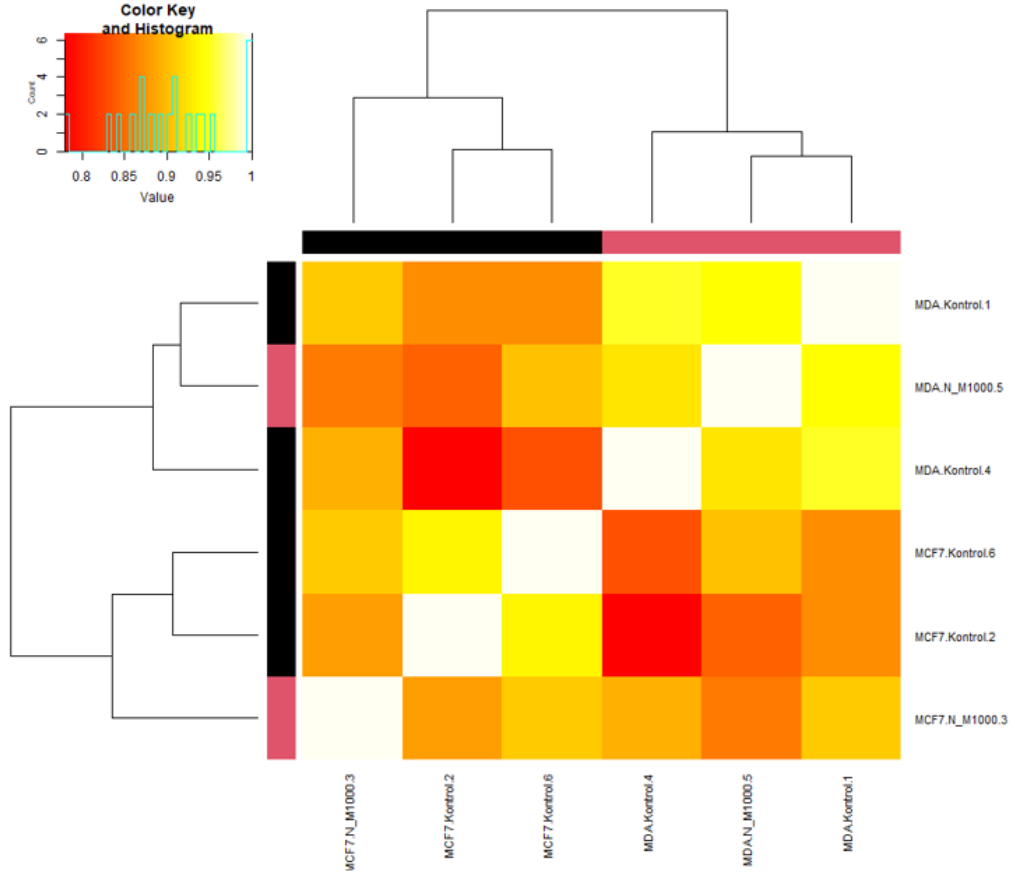
**Şekil 4.33:** Normalizasyon Öncesi Sinyal Dağılımı – Nikotin ve M1000

Sinyal dağılımının kontrolünün ardından saptama skoru kontrol işlemi gerçekleştirilerek ham verideki genlerin 0-1 arasında ifade edilme olasılığı Şekil 4.34’de gösterilmiştir. (0: ifade edilmiyor, 1: ifade ediliyor)



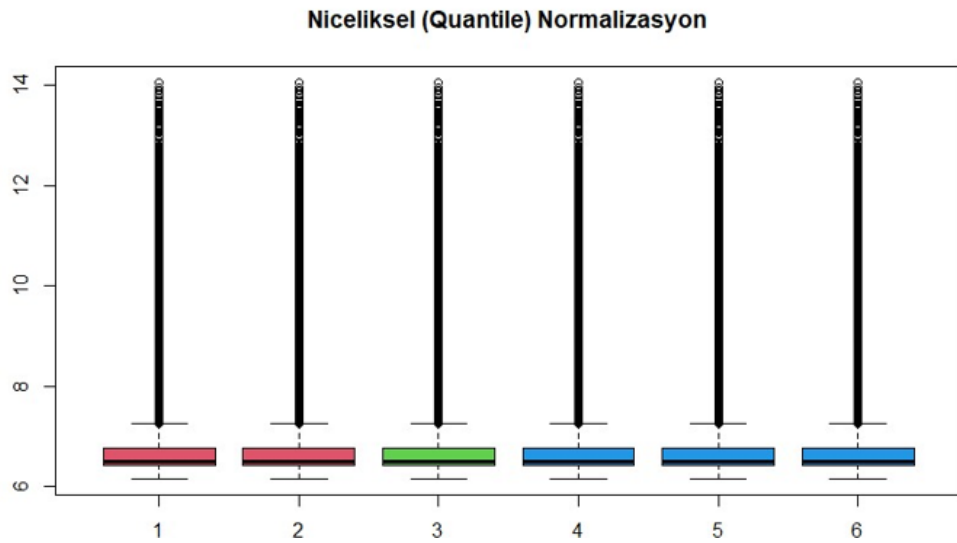
**Şekil 4.34:** Saptama Skorları – Nikotin ve M1000

Veri kalitesi kontrolleri tamamlanan veriseti üzerinde pearson korelasyona bağlı kümeleme işlemi yapılmıştır. Kümelenmiş veri Şekil 4.35’deki ısı haritasındaki gibidir.



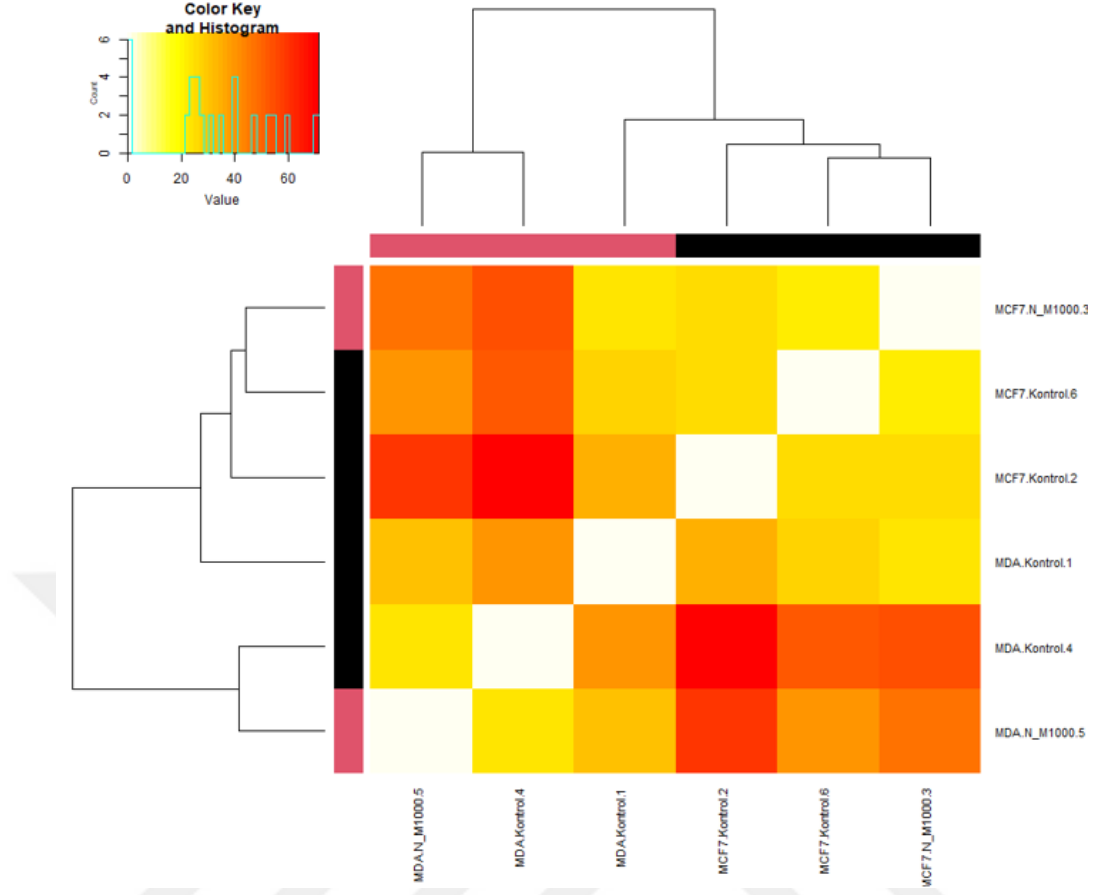
**Şekil 4.35:** Normalizasyon Öncesi Kümeleme Isı Haritası – Nikotin ve M1000

Niceliksel normalizasyon işlemi ile veri 0-1 arasına ölçeklendirilmiştir. Normalizasyon işlemi sonrasında sinyal dağılımı ve ısı haritası tekrar hazırlanarak değişim incelenmiştir. Niceliksel normalizasyon sonrası sinyal dağılımı Şekil 4.36'daki gibidir.



**Şekil 4.36:** Normalizasyon Sonrası Sinyal Dağılımı – Nikotin ve M1000

Normalizasyon sonrası ısı haritası Şekil 4.37'deki gibidir.



**Şekil 4.37:** Normalizasyon Sonrası Kümelenmiş Veri Isı Haritası – Nikotin ve M1000

Normalizasyon işleminin ardından karşılaştırmaların yapılabilmesi için gruplar hazırlanmıştır:

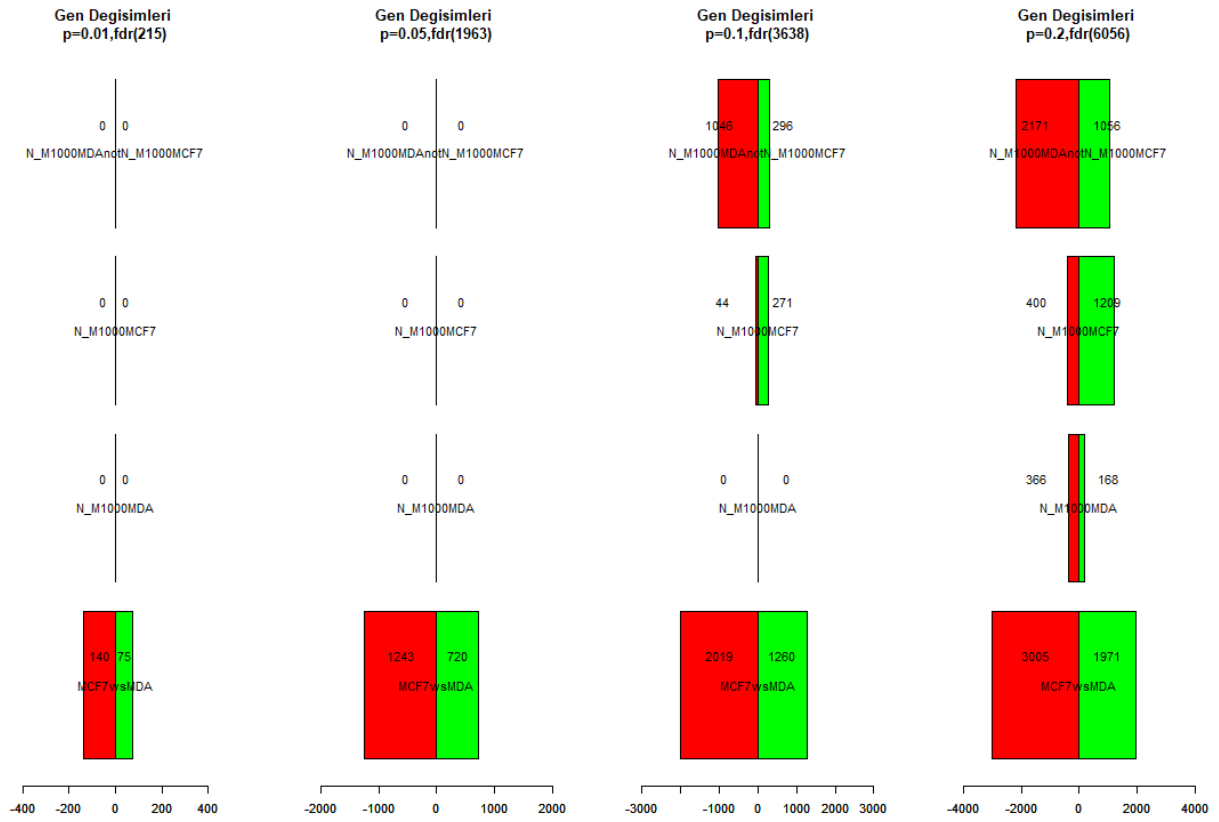
- MCF7wsMDA = MCF7.Kontrol – MDA.Kontrol
- N\_M1000MDA = MDA.N\_M1000 – MDA.Kontrol
- N\_M1000MCF7 = MCF7.N\_M1000 – MCF7.Kontrol
- N\_M1000MDAnotN\_M1000MCF7 = (MDA.N\_M1000 - MDA.Kontrol) - (MCF7.N\_M1000 - MCF7.Kontrol)

Karşılaştırma grupları için hazırlanan kontrast matrisi Tablo 4.24’de gösterilmektedir.

**Tablo 4.24:** Kontrast Matrisi – Nikotin ve M1000

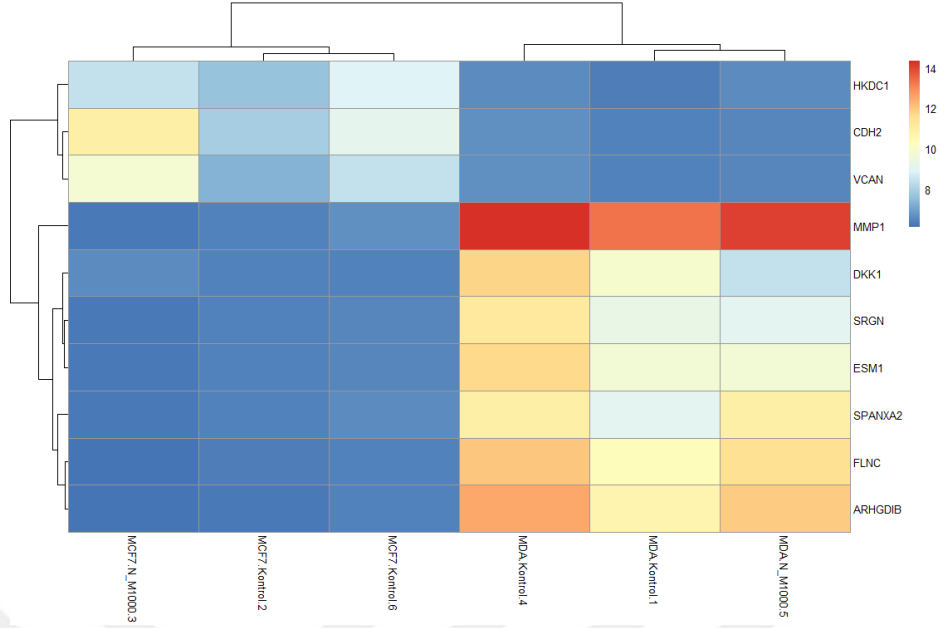
Seviye	MCF7wsMDA	N_M1000MDA	N_M1000MCF7	N_M1000MDA notN_M1000MCF7
MDA.Kontrol	-1	-1	0	-1
MCF7.Kontrol	1	0	-1	1
MCF7.N_M1000	0	0	1	-1
MDA.N_M1000	0	1	0	1

Karşılaştırma gruplarının hazırlanması sonrası SMATP yönteminin son aşaması uygulanarak veri değerlendirilmiştir. Değerlendirme sürecinde eBayes kullanılarak istatistiksel önem değerlendirilmiş, FDR istatistiği ile gözlem yapılmıştır. FDR’de sıklıkla kullanılan p-değeri ile kontroller yapılmıştır. P-değerinin 0’a yakın olması sonucun anlamlı olduğunu göstermektedir. 4 farklı eşik değeri için kontroller yapıldığında gen değişimleri Şekil 4.38’deki gibi olmuştur. Farklı istatistiksel eşiklerde diferansiyel olarak eksprese edilen genlerin sayısının gösterildiği şekilde yeşil çubuklar yukarı düzenlenmiş genleri, kırmızı çubuklar ise aşağı düzenlenmiş genleri göstermektedir.



Şekil 4.38: Gruplar Arası Karşılaştırmalar – Nikotin ve M1000

Sonuçlar incelendiğinde Nikotin ve M1000 veriseti için en çok etkilenen 10 gen aşağıdaki gibi olmuştur.



**Şekil 4.39:** En Çok Etkilenen 10 Gen – Nikotin ve M1000

#### 4.1.7. Deney Grubu 7 – Nikotin ve Işın

MCF7 ve MDA-N1 meme kanseri hücre soylarına Nikotin ve Işın bir arada uygulanması ile hazırlanmış olan deneye ait veriseti 57 kolon ve 34694 satır içermektedir. Veriseti 3 Nikotin ve Işın bir arada uygulanmış örnek, 4 kontrol örneği olmak üzere toplam 7 örnekten oluşmaktadır. Nikotin ve Işın deneyinde SMATP yöntemi aşağıdaki gibi uygulanmıştır.

İlk adım olarak veriseti bilgisayar ortamına aktarılmıştır. Ardından verinin değerlendirilebilmesi için veri çerçevesi hazırlanmıştır. Hazırlanan veri çerçevesi Tablo 4.25’de gösterildiği gibidir.

**Tablo 4.25:** Veri Çerçevesi – Nikotin ve Işın

Sıra	Grup	GenTipi	Replika	Tedavi	DiziKodu
1	1	MDA	1	Kontrol	MDA_K_1
2	1	MCF7	1	Kontrol	MCF7_K_1
3	2	MCF7	1	N_I	MCF7_N_I_1
4	2	MCF7	2	N_I	MCF7_N_I_2
5	3	MDA	2	Kontrol	MDA_K_2
6	3	MDA	1	N_I	MDA_N_I_1
7	3	MCF7	2	Kontrol	MCF7_K_2

Veri çerçevesi yardımıyla Nikotin ve Işın deneyindeki örneklerin sıralaması Tablo 4.26’da gösterildiği gibi olmuştur.

**Tablo 4.26:** Örneklerin Sıralanması – Nikotin ve Işın

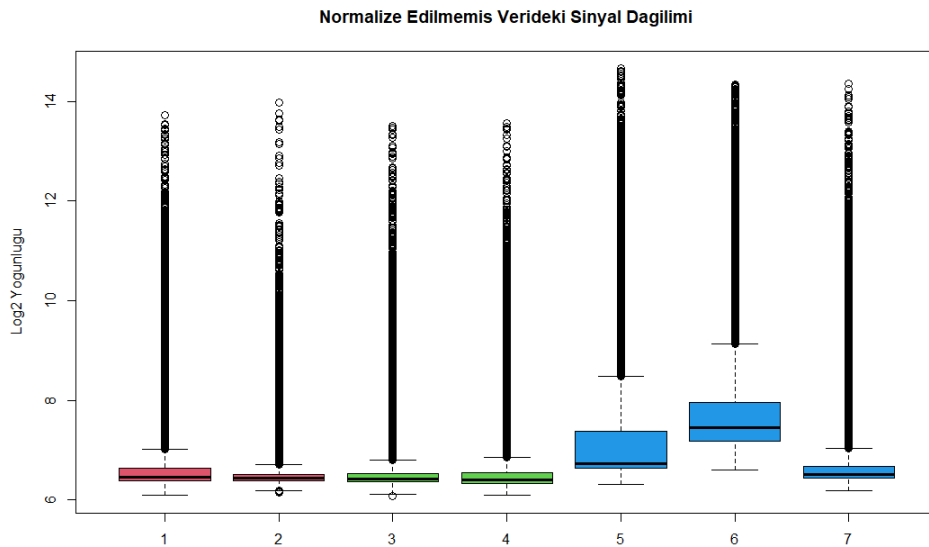
Sıra No	Örnek Adı
1	MDA.Kontrol.1
2	MCF7.Kontrol.2
3	MCF7.N_I.3
4	MCF7.N_I.4
5	MDA.Kontrol.5
6	MDA.N_I.6
7	MCF7.Kontrol.7

Sıralanmış verilerin log2 dönüşümü yapıldıktan sonra oluşturulan tasarım matrisi Tablo 4.27’de gösterilmektedir.

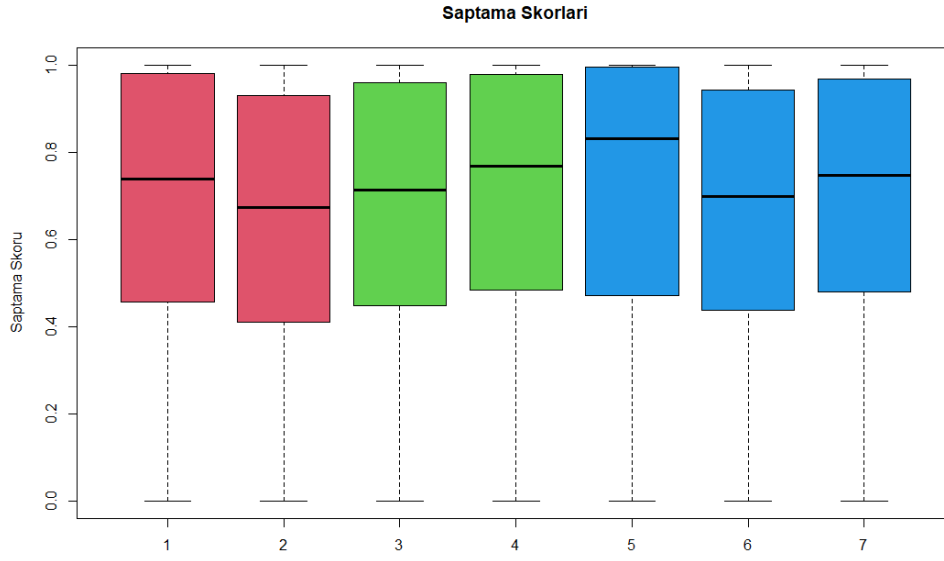
**Tablo 4.27:** Tasarım Matrisi – Nikotin ve Işın

Örnek	Min	1st Qu.	Median	Mean	3rd Qu.	Max
MDA.Kontrol.1	6.102	6.385	6.451	6.733	6.636	13.722
MCF7.Kontrol.2	6.150	6.384	6.435	6.554	6.515	13.980
MCF7.N_I.3	6.085	6.361	6.420	6.610	6.538	13.518
MCF7.N_I.4	6.104	6.338	6.396	6.641	6.544	13.571
MDA.Kontrol.5	6.306	6.634	6.736	7.311	7.376	14.663
MDA.N_I.6	6.606	7.187	7.445	7.835	7.964	14.349
MCF7.Kontrol.7	6.192	6.445	6.513	6.767	6.682	14.356

İşlenmemiş/ham verideki sinyal dağılımı kontrol edilmiştir. Normalizasyon işlemi öncesindeki sinyal dağılımı Şekil 4.40’daki kutu grafik ile gösterildiği gibidir.

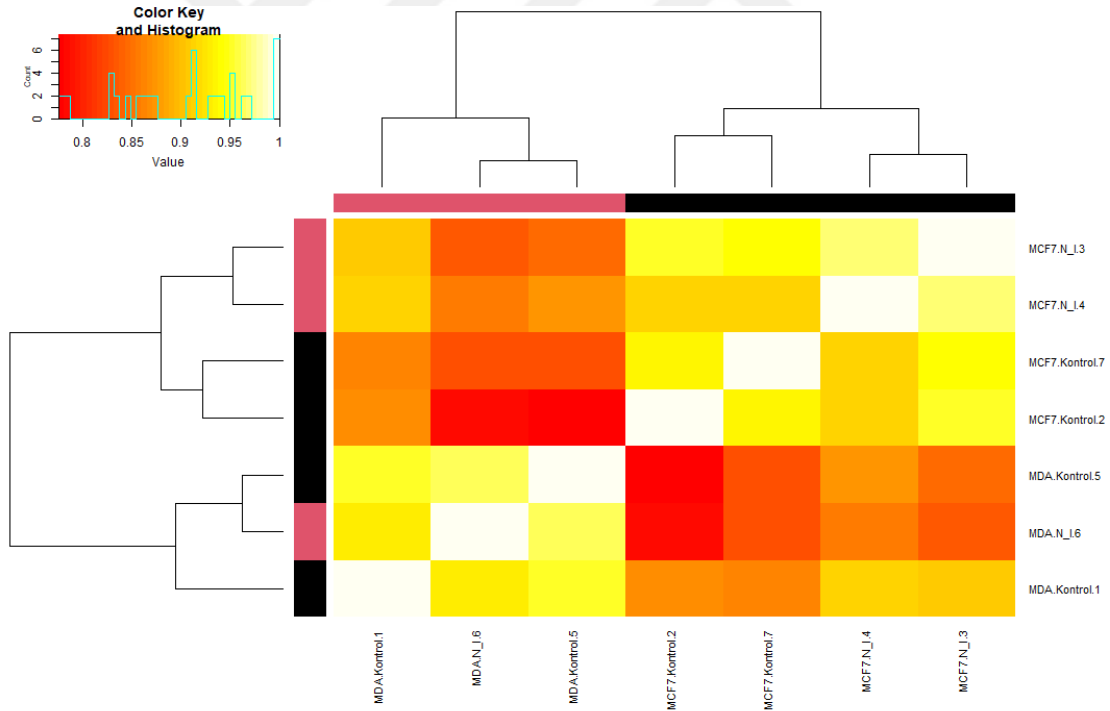
**Şekil 4.40:** Normalizasyon Öncesi Sinyal Dağılımı – Nikotin ve Işın

Sinyal dağılımının kontrolünün ardından saptama skoru kontrol işlemi gerçekleştirilerek ham verideki genlerin 0-1 arasında ifade edilme olasılığı Şekil 4.41’de gösterilmiştir. (0: ifade edilmiyor, 1: ifade ediliyor)



Şekil 4.41: Saptama Skorlari – Nikotin ve Işın

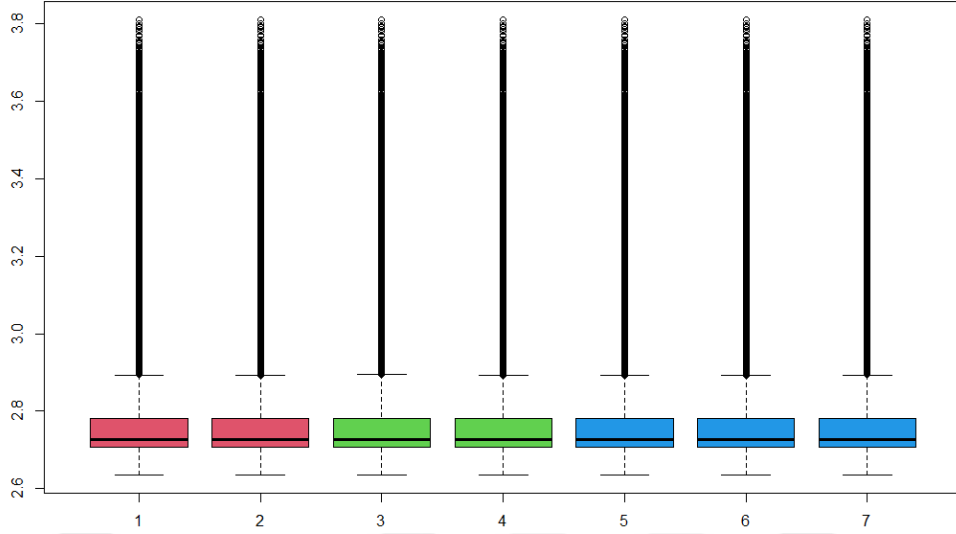
Veri kalitesi kontrolleri tamamlanan veriseti üzerinde pearson korelasyona bağlı kümeleme işlemi yapılmıştır. Kümeleşmiş veri Şekil 4.42'deki ısı haritasındaki gibidir.



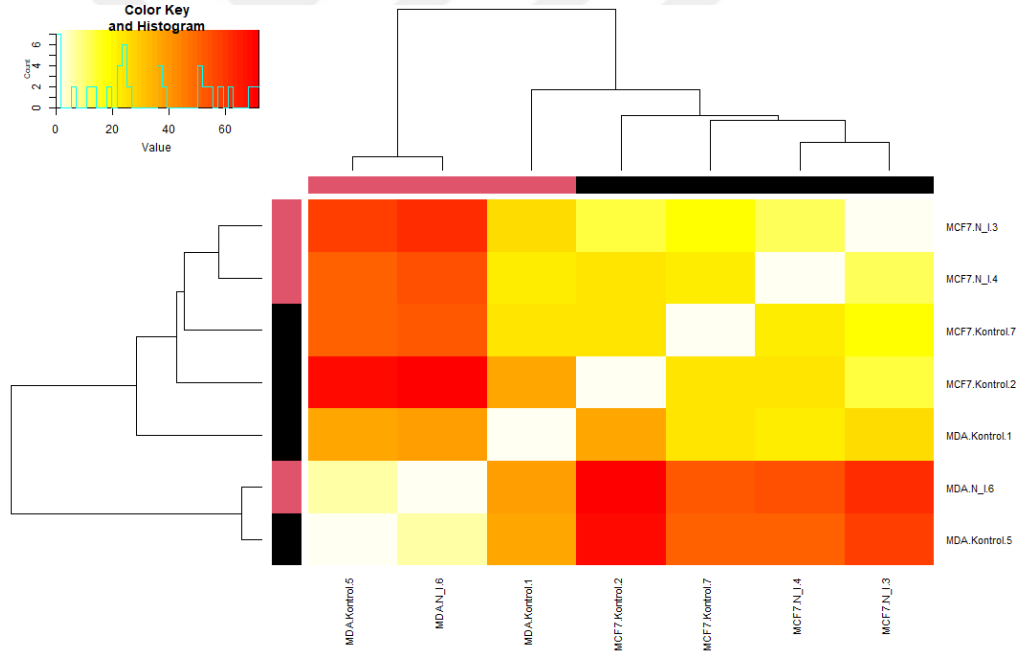
Şekil 4.42: Normalizasyon Öncesi Kümeleme Isı Haritası – Nikotin ve Işın

Niceliksel normalizasyon işlemi ile veri 0-1 arasına ölçeklendirilmiştir. Normalizasyon işlemi sonrasında sinyal dağılımı ve ısı haritası tekrar hazırlanarak değişim incelenmiştir. Niceliksel normalizasyon sonrası sinyal dağılımı Şekil 4.43'deki gibidir.

## Niceliksel (Quantile) Normalizasyon



**Şekil 4.43:** Normalizasyon Sonrası Sinyal Dağılımı – Nikotin ve Işın  
Normalizasyon sonrası ısı haritası Şekil 4.44'deki gibidir.



**Şekil 4.44:** Normalizasyon Sonrası Kümelmiş Veri Isı Haritası – Nikotin ve Işın  
Normalizasyon işleminin ardından karşılaştırmaların yapılabilmesi için gruplar hazırlanmıştır:

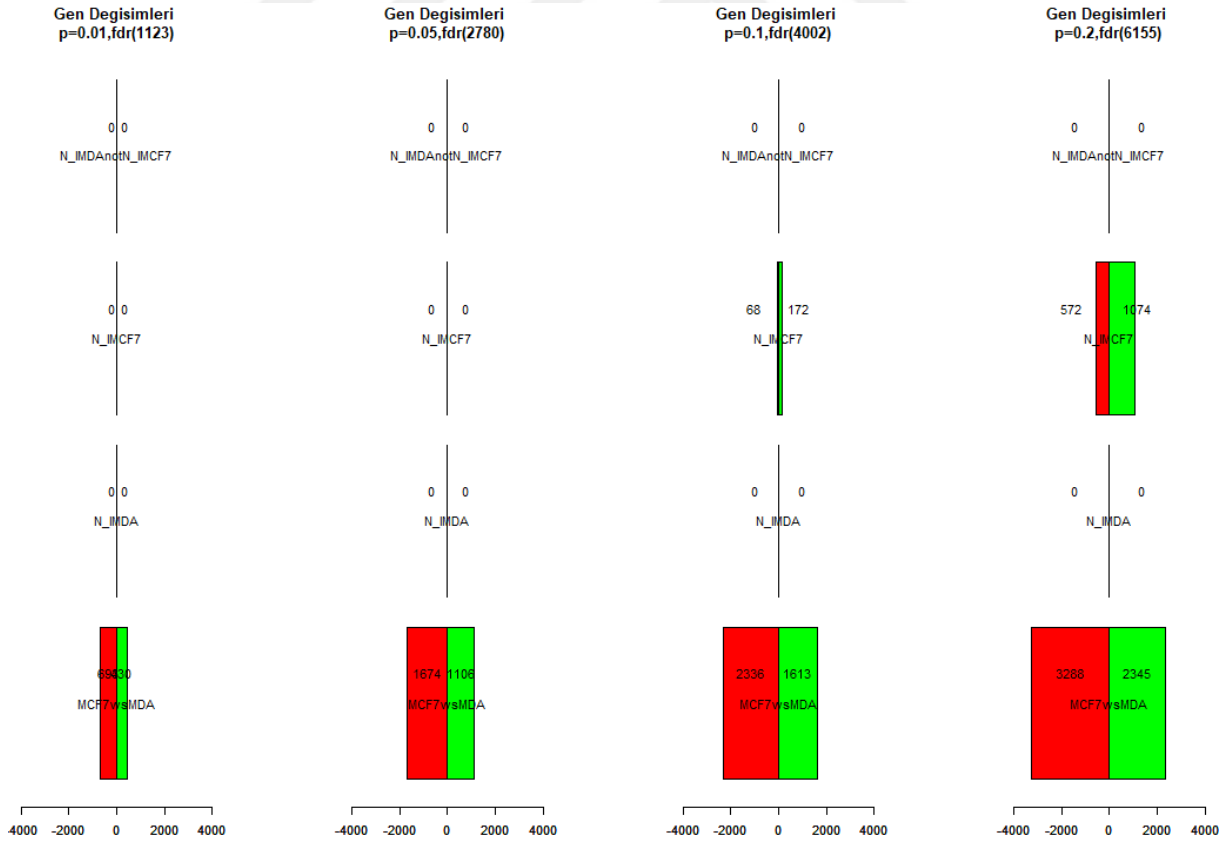
- $MCF7wsMDA = MCF7.Kontrol - MDA.Kontrol$
- $N\_IMDA = MDA.N\_I - MDA.Kontrol$
- $N\_IMCF7 = MCF7.N\_I - MCF7.Kontrol$
- $N\_IMDAnotN\_IMCF7 = (MDA.N\_I - MDA.Kontrol) - (MCF7.N\_I - MCF7.Kontrol)$

Karşılaştırma grupları için hazırlanan kontrast matrisi Tablo 4.28’de gösterilmektedir.

**Tablo 4.28:** Kontrast Matrisi – Nikotin ve Işın

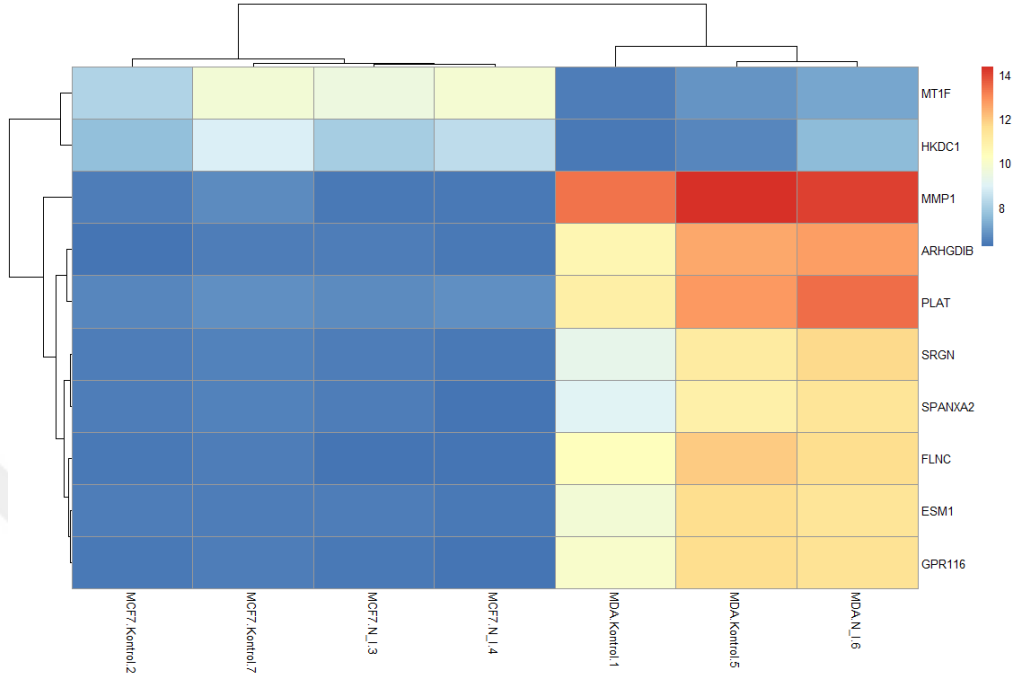
Seviye	MCF7wsMDA	N_IMDA	N_IMCF7	N_IMDA notN_IMCF7
MDA.Kontrol	-1	-1	0	-1
MCF7.Kontrol	1	0	-1	1
MCF7.N_I	0	0	1	-1
MDA.N_I	0	1	0	1

Karşılaştırma gruplarının hazırlanması sonrası SMATP yönteminin son aşaması uygulanarak veri değerlendirilmiştir. Değerlendirme sürecinde eBayes kullanılarak istatistiksel önem değerlendirilmiş, FDR istatistiği ile gözlem yapılmıştır. FDR’de sıklıkla kullanılan p-değeri ile kontroller yapılmıştır. P-değerinin 0’a yakın olması sonucun anlamlı olduğunu göstermektedir. 4 farklı eşik değeri için kontroller yapıldığında gen değişimleri Şekil 4.45’deki gibi olmuştur. Farklı istatistiksel eşiklerde diferansiyel olarak ekprese edilen genlerin sayısının gösterildiği şekilde yeşil çubuklar yukarı düzenlenmiş genleri, kırmızı çubuklar ise aşağı düzenlenmiş genleri göstermektedir.



**Şekil 4.45:** Gruplar Arası Karşılaştırmalar – Nikotin ve Işın

Sonuçlar incelendiğinde Nikotin ve Işın veriseti için en çok etkilenen 10 gen aşağıdaki gibi olmuştur.



Şekil 4.46: En Çok Etkilenen 10 Gen – Nikotin ve Işın

#### 4.1.8. DeneY Grubu 8 – Nikotin, Işın ve M100

MCF7 ve MDA-N1 meme kanseri hücre soylarına Nikotin, Işın ve M100'ün bir arada uygulanması ile hazırlanmış olan deneye ait veriseti 65 kolon ve 34694 satır içermektedir. Veriseti 4 Nikotin, Işın ve M100 bir arada uygulanmış örnek, 4 kontrol örneği olmak üzere toplam 8 örnekten oluşmaktadır. Nikotin, Işın ve M100 deneyinde SMATP yöntemi aşağıdaki gibi uygulanmıştır.

İlk adım olarak veriseti bilgisayar ortamına aktarılmıştır. Ardından verinin değerlendirilebilmesi için veri çerçevesi hazırlanmıştır. Hazırlanan veri çerçevesi Tablo 4.29'da gösterildiği gibidir.

Tablo 4.29: Veri Çerçevesi – Nikotin, Işın ve M100

Sıra	Grup	GenTipi	Replika	Tedavi	DiziKodu
1	1	MDA	1	Kontrol	MDA_K_1
2	1	MDA	1	N_I_M100	MDA_N_I_M100_1
3	1	MCF7	1	Kontrol	MCF7_K_1
4	2	MCF7	1	N_I_M100	MCF7_N_I_M100_1
5	2	MCF7	2	N_I_M100	MCF7_N_I_M100_2
6	3	MDA	2	Kontrol	MDA_K_2
7	3	MDA	2	N_I_M100	MDA_N_I_M100_2
8	3	MCF7	2	Kontrol	MCF7_K_2

Veri çerçevesi yardımıyla Nikotin, Işın ve M100 deneyindeki örneklerin sıralaması Tablo 4.30'da gösterildiği gibi olmuştur.

**Tablo 4.30:** Örneklerin Sıralanması – Nikotin, Işın ve M100

Sıra No	Örnek Adı
1	MDA.Kontrol.1
2	MDA. N_I_M100.2
3	MCF.Kontrol.3
4	MCF7. N_I_M100.4
5	MCF7. N_I_M100.5
6	MDA.Kontrol.6
7	MDA. N_I_M100.7
8	MCF7.Kontrol.8

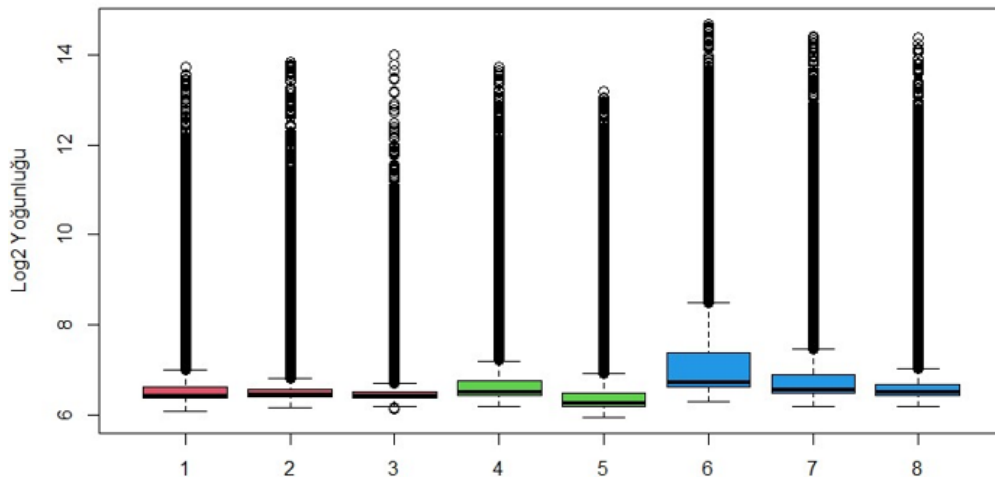
Sıralanmış verilerin log2 dönüşümü yapıldıktan sonra oluşturulan tasarım matrisi Tablo 4.31'de gösterilmektedir.

**Tablo 4.31:** Tasarım Matrisi – Nikotin, Işın ve M100

Örnek	Min	1st Qu.	Median	Mean	3rd Qu.	Max
MDA.Kontrol.1	6.102	6.385	6.451	6.733	6.636	13.722
MDA. N_I_M100.2	6.170	6.401	6.457	6.657	6.568	13.832
MCF.Kontrol.3	6.150	6.384	6.435	6.554	6.515	13.980
MCF7. N_I_M100.4	6.186	6.453	6.526	6.837	6.756	13.718
MCF7. N_I_M100.5	5.961	6.200	6.265	6.584	6.491	13.173
MDA.Kontrol.6	6.306	6.634	6.736	7.311	7.376	14.663
MDA. N_I_M100.7	6.188	6.505	6.586	6.951	6.890	14.393
MCF7.Kontrol.8	6.192	6.445	6.513	6.767	6.682	14.356

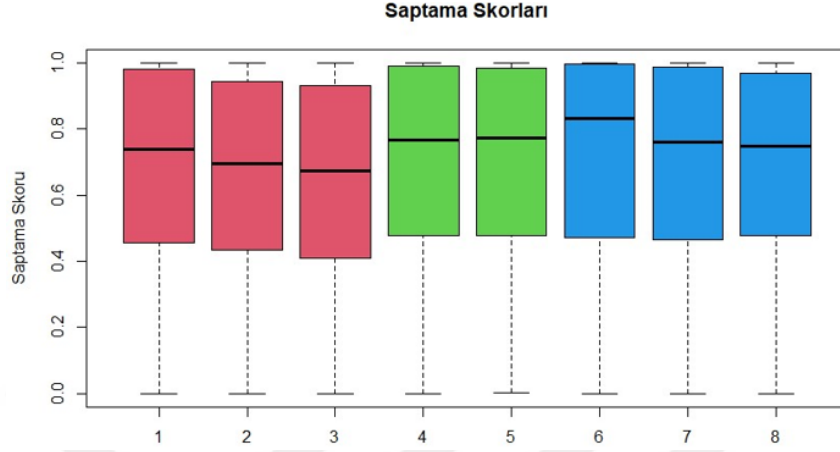
İşlenmemiş/ham verideki sinyal dağılımı kontrol edilmiştir. Normalizasyon işlemi öncesindeki sinyal dağılımı Şekil 4.47'deki kutu grafik ile gösterildiği gibidir.

**Normalize Edilmemiş Verideki Sinyal Dağılımı**



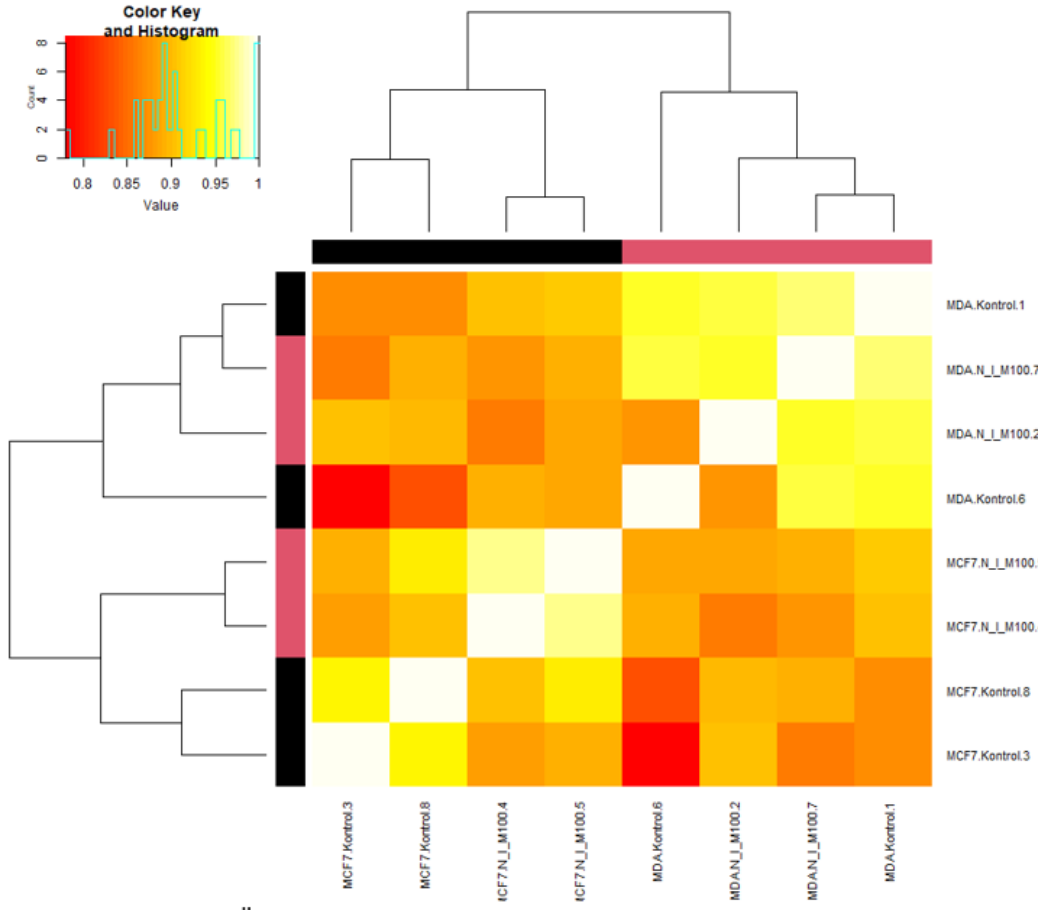
**Şekil 4.47:** Normalizasyon Öncesi Sinyal Dağılımı – Nikotin, Işın ve M100

Sinyal dağılımının kontrolünün ardından saptama skoru kontrol işlemi gerçekleştirilerek ham verideki genlerin 0-1 arasında ifade edilme olasılığı Şekil 4.48’de gösterilmiştir. (0: ifade edilmiyor, 1: ifade ediliyor)



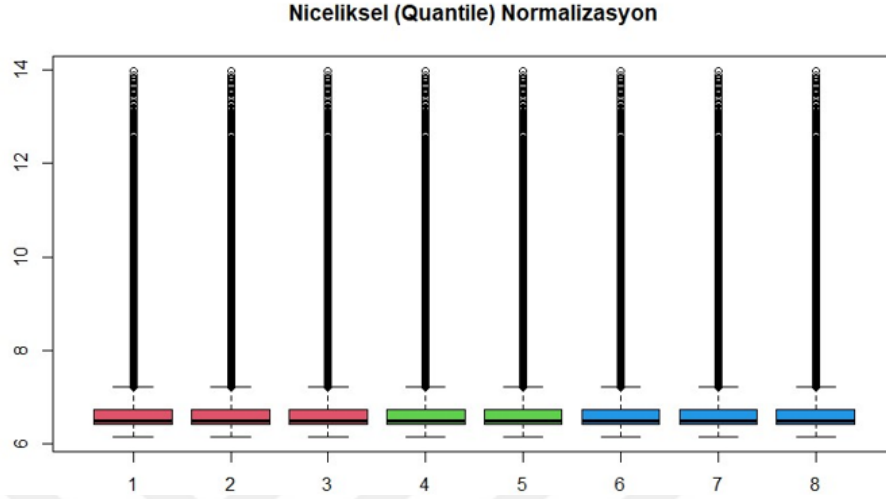
Şekil 4.48: Saptama Skorları – Nikotin, Işın ve M100

Veri kalitesi kontrolleri tamamlanan veriseti üzerinde pearson korelasyona bağlı kümeleme işlemi yapılmıştır. Kümelenmiş veri Şekil 4.49’daki ısı haritasındaki gibidir.



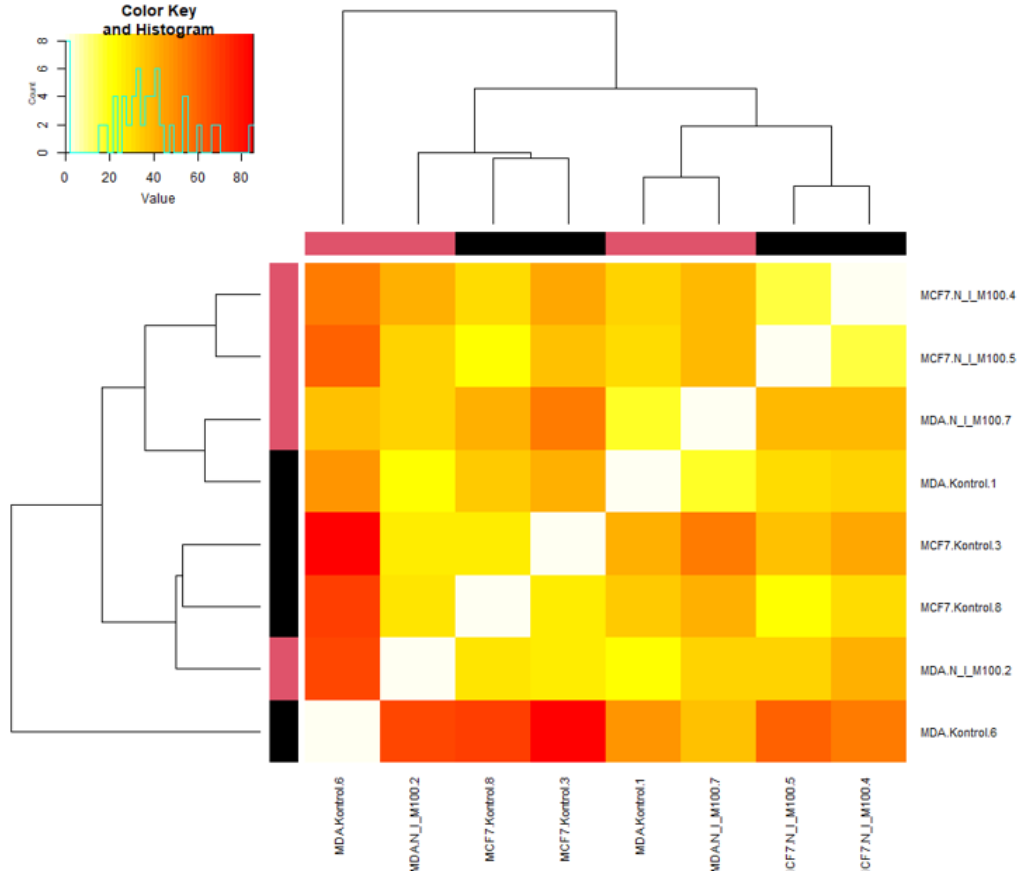
Şekil 4.49: Normalizasyon Öncesi Kümeleme Isı Haritası – Nikotin, Işın ve M100

Niceliksel normalizasyon işlemi ile veri 0-1 arasına ölçeklendirilmiştir. Normalizasyon işlemi sonrasında sinyal dağılımı ve ısı haritası tekrar hazırlanarak değişim incelenmiştir. Niceliksel normalizasyon sonrası sinyal dağılımı Şekil 4.50'deki gibidir.



**Şekil 4.50:** Normalizasyon Sonrası Sinyal Dağılımı – Nikotin, Işın ve M100

Normalizasyon sonrası ısı haritası Şekil 4.51'deki gibidir.



**Şekil 4.51:** Normalizasyon Sonrası Kümelmiş Veri Isı Haritası – Nikotin, Işın ve M100

Normalizasyon işleminin ardından karşılaştırmaların yapılabilmesi için gruplar hazırlanmıştır:

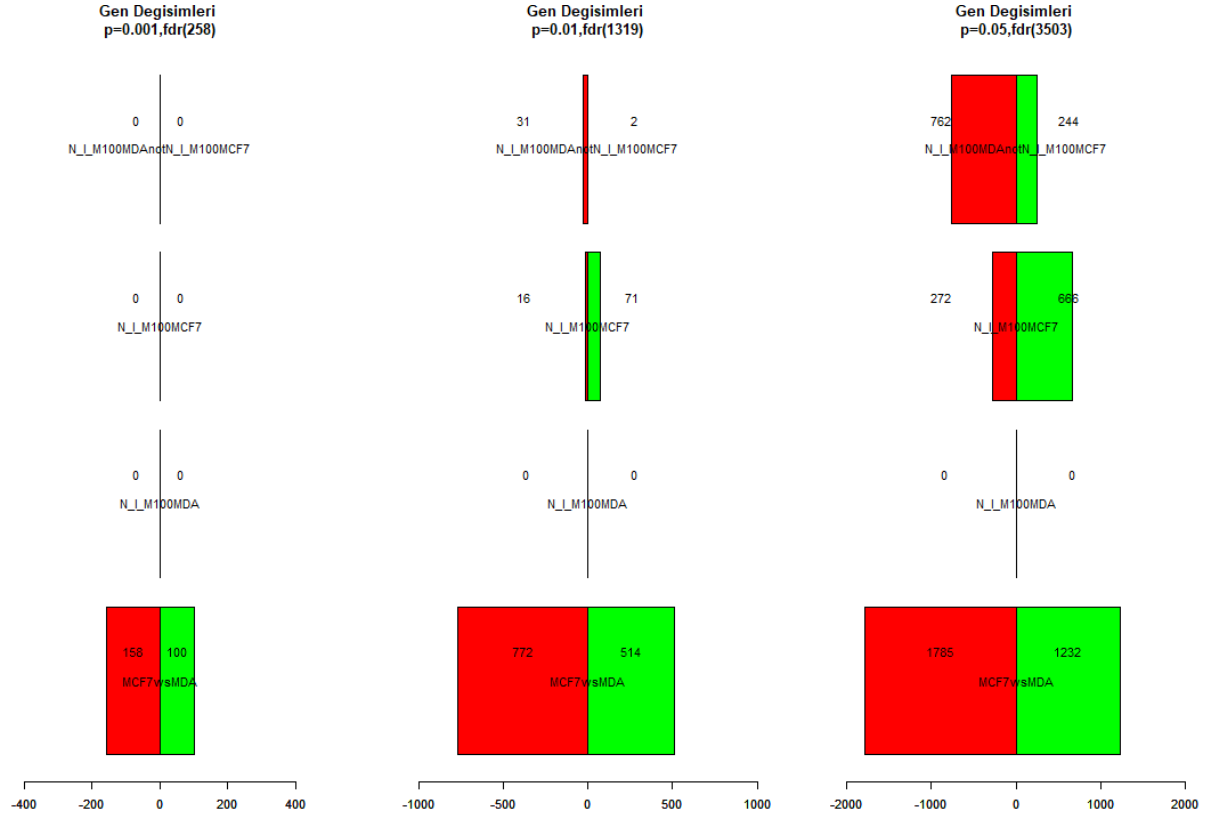
- $MCF7wsMDA = MCF7.Kontrol - MDA.Kontrol$
- $N\_I\_M100MDA = MDA.N\_I\_M100 - MDA.Kontrol$
- $N\_I\_M100MCF7 = MCF7.N\_I\_M100 - MCF7.Kontrol$
- $N\_I\_M100MDAnotN\_I\_M100MCF7 = (MDA.N\_I\_M100 - MDA.Kontrol) - (MCF7.N\_I\_M100 - MCF7.Kontrol)$

Karşılaştırma grupları için hazırlanan kontrast matrisi Tablo 4.32’de gösterilmektedir.

**Tablo 4.32:** Kontrast Matrisi – Nikotin, Işın ve M100

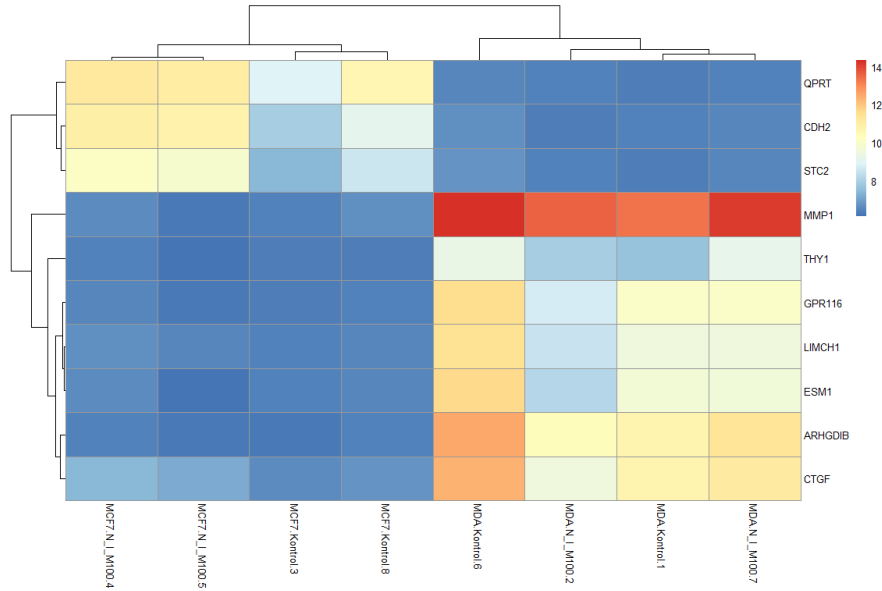
Seviye	MCF7wsMDA	N_I_M100MDA	N_I_M100MCF7	N_I_M100MDA Not N_I_M100MCF7
MDA.Kontrol	-1	-1	0	-1
MDA.N_I_M100	0	1	0	1
MCF7.Kontrol	1	0	-1	1
MCF7.N_I_M100	0	0	1	-1

Karşılaştırma gruplarının hazırlanması sonrası SMATP yönteminin son aşaması uygulanarak veri değerlendirilmiştir. Değerlendirme sürecinde eBayes kullanılarak istatistiksel önem değerlendirilmiş, FDR istatistiği ile gözlem yapılmıştır. FDR’de sıklıkla kullanılan p-değeri ile kontroller yapılmıştır. P-değerinin 0’a yakın olması sonucun anlamlı olduğunu göstermektedir. 3 farklı eşik değeri için kontroller yapıldığında gen değişimleri Şekil 4.52’deki gibi olmuştur. Farklı istatistiksel eşiklerde diferansiyel olarak ekprese edilen genlerin sayısının gösterildiği şekilde yeşil çubuklar yukarı düzenlenmiş genleri, kırmızı çubuklar ise aşağı düzenlenmiş genleri göstermektedir.



Şekil 4.52: Gruplar Arası Karşılaştırmalar – Nikotin, Işın ve M100

Sonuçlar incelendiğinde Nikotin, Işın ve M100 veriseti için en çok etkilenen 10 gen aşağıdaki gibi olmuştur.



Şekil 4.53: En Çok Etkilenen 10 Gen – Nikotin, Işın ve M100

#### 4.1.9. Deney Grubu 9 – Nikotin, Işın ve M1000

MCF7 ve MDA-N1 meme kanseri hücre soylarına Nikotin, Işın ve M1000'in bir arada uygulanması ile hazırlanmış olan deneye ait veriseti 65 kolon ve 34694 satır içermektedir.

Veriseti 4 Nikotin, Işın ve M1000 bir arada uygulanmış örnek, 4 kontrol örneği olmak üzere toplam 8 örnekten oluşmaktadır. Nikotin, Işın ve M1000 deneyinde SMATP yöntemi aşağıdaki gibi uygulanmıştır.

İlk adım olarak veriseti bilgisayar ortamına aktarılmıştır. Ardından verinin değerlendirilebilmesi için veri çerçevesi hazırlanmıştır. Hazırlanan veri çerçevesi Tablo 4.33'de gösterildiği gibidir.

**Tablo 4.33:** Veri Çerçevesi – Nikotin, Işın ve M1000

Sıra	Grup	GenTipi	Replika	Tedavi	DiziKodu
1	1	MDA	1	Kontrol	MDA_K_1
2	1	MCF7	1	Kontrol	MCF7_K_1
3	2	MCF7	1	N_I_M1000	MCF7_N_I_M1000_1
4	2	MCF7	2	N_I_M1000	MCF7_N_I_M1000_2
5	2	MCF7	3	N_I_M1000	MCF7_N_I_M1000_3
6	3	MDA	2	Kontrol	MDA_K_2
7	3	MDA	1	N_I_M1000	MDA_N_I_M1000_1
8	3	MCF7	2	Kontrol	MCF7_K2

Veri çerçevesi yardımıyla Nikotin, Işın ve M1000 deneyindeki örneklerin sıralaması Tablo 4.34'de gösterildiği gibi olmuştur.

**Tablo 4.34:** Örneklerin Sıralanması – Nikotin, Işın ve M1000

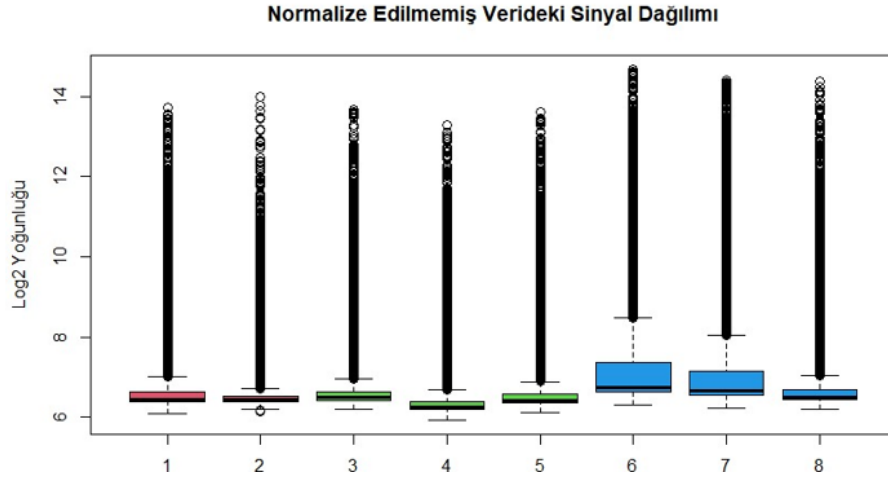
Sıra No	Örnek Adı
1	MDA.Kontrol.1
2	MCF7.Kontrol.2
3	MCF7.N_I_M1000.3
4	MCF7.N_I_M1000.4
5	MCF7.N_I_M1000.5
6	MDA.Kontrol.6
7	MDA.N_I_M1000.7
8	MCF7.Kontrol.8

Sıralanmış verilerin log<sub>2</sub> dönüşümü yapıldıktan sonra oluşturulan tasarım matrisi Tablo 4.35'de gösterilmektedir.

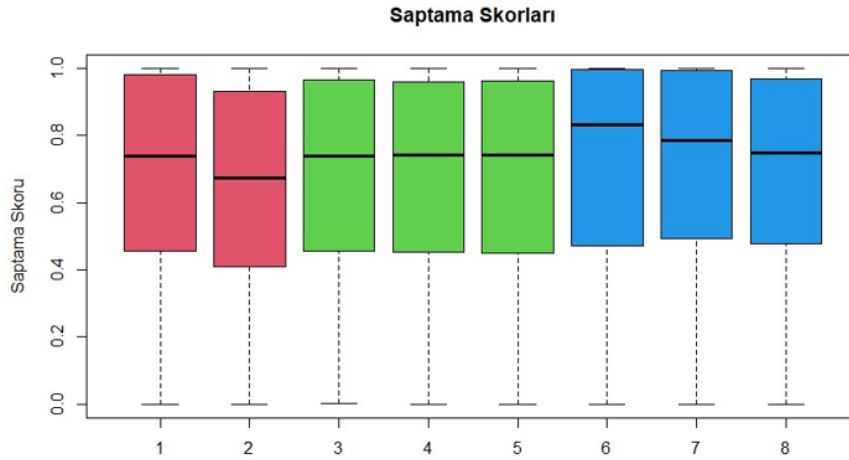
**Tablo 4.35:** Tasarım Matrisi – Nikotin, Işın ve M1000

Örnek	Min	1st Qu.	Median	Mean	3rd Qu.	Max
MDA.Kontrol.1	6.102	6.385	6.451	6.733	6.636	13.722
MCF7.Kontrol.2	6.150	6.384	6.435	6.554	6.515	13.980
MCF7.N_I_M1000.3	6.200	6.425	6.487	6.708	6.637	13.651
MCF7.N_I_M1000.4	5.938	6.200	6.257	6.483	6.397	13.289
MCF7.N_I_M1000.5	6.111	6.356	6.420	6.640	6.571	13.604
MDA.Kontrol.6	6.306	6.634	6.736	7.311	7.376	14.663
MDA.N_I_M1000.7	6.231	6.548	6.652	7.125	7.150	14.382
MCF7.Kontrol.8	6.192	6.445	6.513	6.767	6.682	14.356

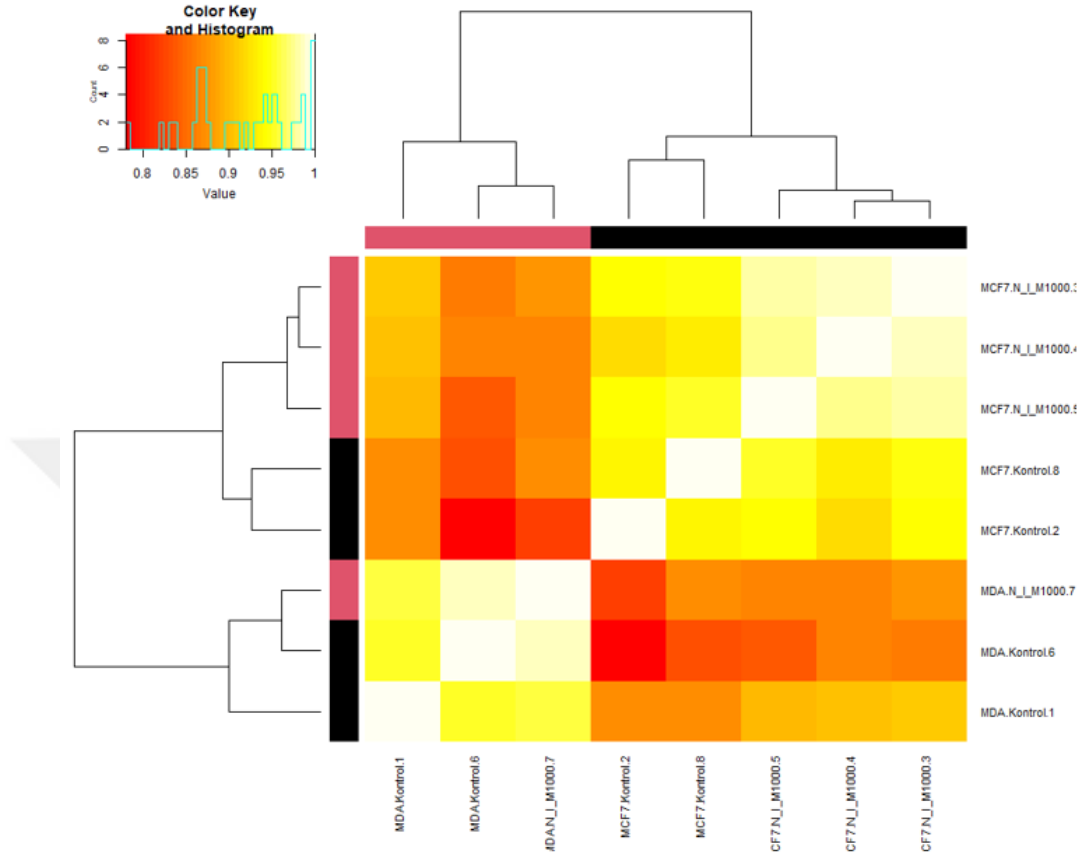
İşlenmemiş/ham verideki sinyal dağılımı kontrol edilmiştir. Normalizasyon işlemi öncesindeki sinyal dağılımı Şekil 4.54'deki kutu grafik ile gösterildiği gibidir.

**Şekil 4.54:** Normalizasyon Öncesi Sinyal Dağılımı – Nikotin, Işın ve M1000

Sinyal dağılımının kontrolünün ardından saptama skoru kontrol işlemi gerçekleştirilerek ham verideki genlerin 0-1 arasında ifade edilme olasılığı Şekil 4.55'de gösterilmiştir. (0: ifade edilmiyor, 1: ifade ediliyor)

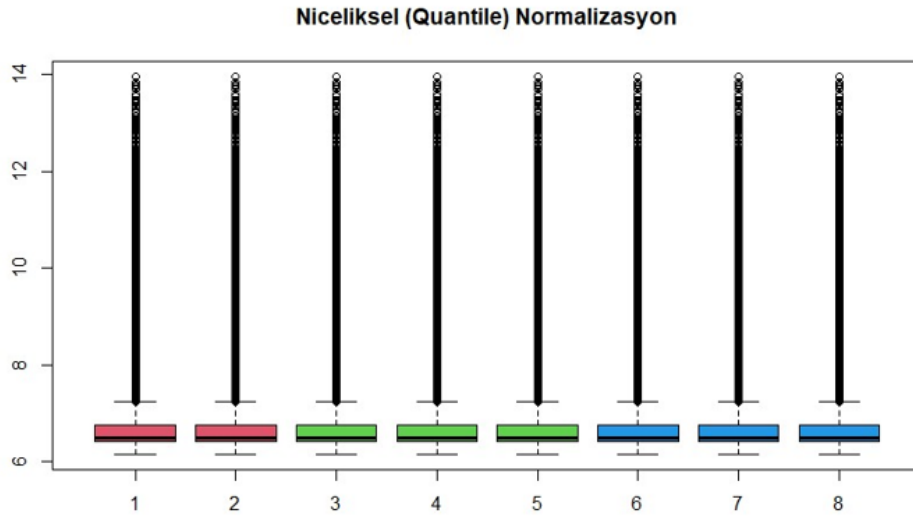
**Şekil 4.55:** Saptama Skorları – Nikotin, Işın ve M1000

Veri kalitesi kontrolleri tamamlanan veriseti üzerinde pearson korelasyona bağlı kümeleme işlemi yapılmıştır. Kümeleşmiş veri Şekil 4.56'daki ısı haritasındaki gibidir.



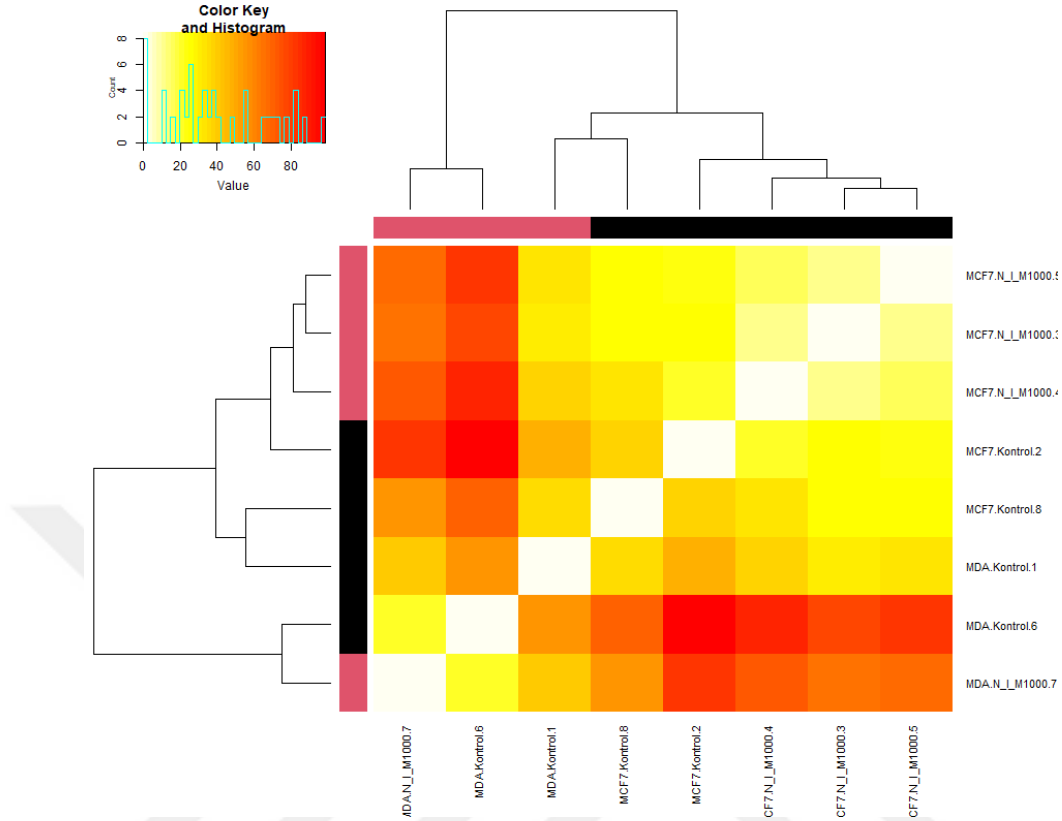
**Şekil 4.56:** Normalizasyon Öncesi Kümeleme Isı Haritası – Nikotin, Işın ve M1000

Niceliksel normalizasyon işlemi ile veri 0-1 arasına ölçeklendirilmiştir. Normalizasyon işlemi sonrasında sinyal dağılımı ve ısı haritası tekrar hazırlanarak değişim incelenmiştir. Niceliksel normalizasyon sonrası sinyal dağılımı Şekil 4.57'deki gibidir.



**Şekil 4.57:** Normalizasyon Sonrası Sinyal Dağılımı – Nikotin, Işın ve M1000

Normalizasyon sonrası ısı haritası Şekil 4.58'deki gibidir.



**Şekil 4.58:** Normalizasyon Sonrası Kümelmiş Veri Isı Haritası – Nikotin, Işın ve M1000

Normalizasyon işleminin ardından karşılaştırmaların yapılabilmesi için gruplar hazırlanmıştır:

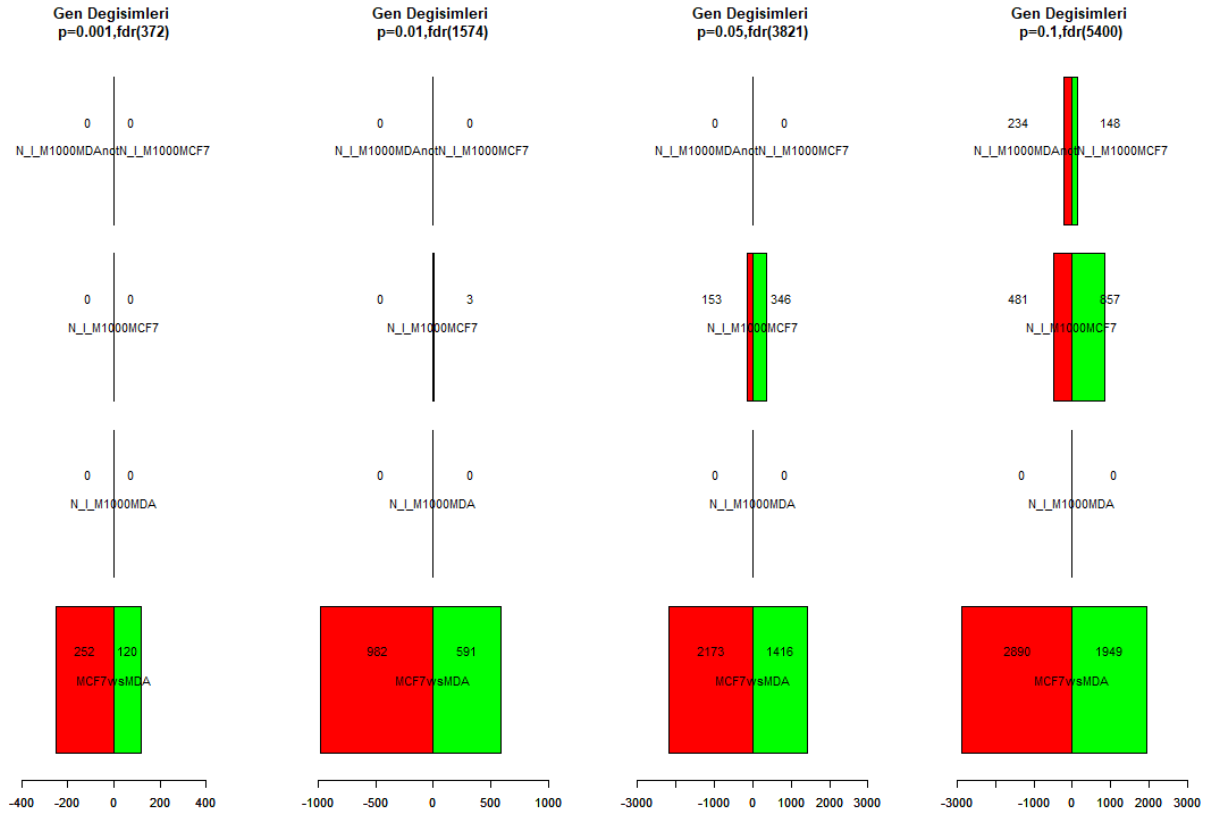
- MCF7wsMDA = MCF7.Kontrol – MDA.Kontrol
- N\_I\_M1000MDA = MDA.N\_I\_M1000 – MDA.Kontrol
- N\_I\_M1000MCF7 = MCF7.N\_I\_M1000 – MCF7.Kontrol
- $N_I_M1000MDA - N_I_M1000MCF7 = (MDA.N_I_M1000 - MDA.Kontrol) - (MCF7.N_I_M1000 - MCF7.Kontrol)$

Karşılaştırma grupları için hazırlanan kontrast matrisi Tablo 4.36'da gösterilmektedir.

**Tablo 4.36:** Kontrast Matrisi – Nikotin, Işın ve M1000

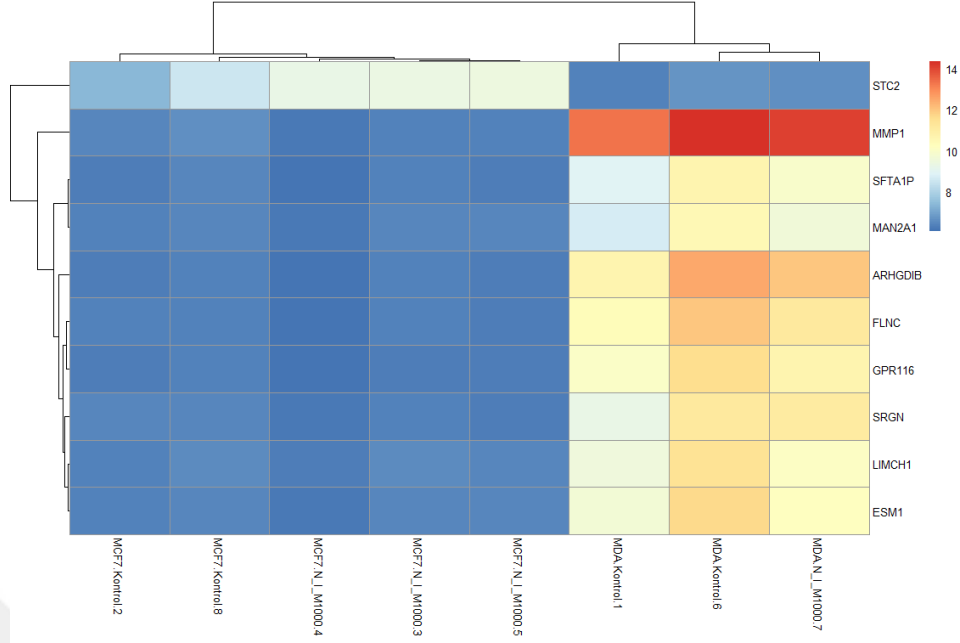
Seviye	MCF7 wsMDA	N_I_M1000MDA	N_I_M1000MCF7	N_I_M1000MDA not N_I_M1000MCF7
MDA.Kontrol	-1	-1	0	-1
MCF7.Kontrol	1	0	-1	1
MCF7.N_I_M1000	0	0	1	-1
MDA.N_I_M1000	0	1	0	1

Karşılaştırma gruplarının hazırlanması sonrası SMATP yönteminin son aşaması uygulanarak veri değerlendirilmiştir. Değerlendirme sürecinde eBayes kullanılarak istatistiksel önem değerlendirilmiş, FDR istatistiği ile gözlem yapılmıştır. FDR’de sıklıkla kullanılan p-değeri ile kontroller yapılmıştır. P-değerinin 0’a yakın olması sonucun anlamlı olduğunu göstermektedir. 4 farklı eşik değeri için kontroller yapıldığında gen değişimleri Şekil 4.59’daki gibi olmuştur. Farklı istatistiksel eşiklerde diferansiyel olarak eksprese edilen genlerin sayısının gösterildiği şekilde yeşil çubuklar yukarı düzenlenmiş genleri, kırmızı çubuklar ise aşağı düzenlenmiş genleri göstermektedir.



Şekil 4.59: Gruplar Arası Karşılaştırmalar – Nikotin, Işın ve M1000

Sonuçlar incelendiğinde Nikotin, Işın ve M1000 veriseti için en çok etkilenen 10 gen aşağıdaki gibi olmuştur.



Şekil 4.60: En Çok Etkilenen 10 Gen – Nikotin, Işın ve M1000

#### 4.2. GSE10470 VERİSETİ ANALİZ VE DEĞERLENDİRME

Fare nöronları üzerinde oksidatif stresin etkisini inceleyen çalışmaya ait verisetinde normal fareler ve ATF4 geni olmayan farelerden alınan örneklere HCA uygulanması ile hazırlanmış olan GSE10470 deneyine ait veriseti 193 kolon ve 24048 satır içermektedir. Veriseti 8 HCA uygulanmış örnek, 16 kontrol örneği olmak üzere toplam 24 örnekten oluşmaktadır. Oksidatif stres deneyinde SMATP yöntemi aşağıdaki gibi uygulanmıştır.

İlk adım olarak veriseti bilgisayar ortamına aktarılmıştır. Ardından verinin değerlendirilebilmesi için veri çerçevesi hazırlanmıştır. Hazırlanan veri çerçevesi Tablo 4.37’de gösterildiği gibidir.

Tablo 4.37: Veri Çerçevesi – GSE10470

Sıra	Grup	GenTipi	Replika	Tedavi	DiziKodu
1	1	wt	1	ctrl	1412066061_A
2	1	wt	2	ctrl	1412066061_B
3	1	wt	3	ctrl	1412066061_C
4	1	wt	4	ctrl	1412066061_D
5	1	ko	1	ctrl	1412066061_E
6	1	ko	2	ctrl	1412066061_F
7	1	ko	3	ctrl	1412066061_G
8	1	ko	4	ctrl	1412066061_H
9	2	ko	5	ctrl	1541554116_A
10	2	wt	1	hca	1541554116_B
11	2	ko	1	hca	1541554116_C
12	2	wt	5	ctrl	1541554116_D

**Tablo 4.37 (devam)**

13	2	ko	6	ctrl	1541554116_E
14	2	wt	2	hca	1541554116_F
15	2	ko	2	hca	1541554116_G
16	2	wt	6	ctrl	1541554116_H
17	3	wt	7	ctrl	1541554117_A
18	3	ko	3	hca	1541554117_B
19	3	wt	3	hca	1541554117_C
20	3	ko	7	ctrl	1541554117_D
21	3	wt	8	ctrl	1541554117_E
22	3	ko	4	hca	1541554117_F
23	3	wt	4	hca	1541554117_G
24	3	ko	8	ctrl	1541554117_H

Veri çerçevesi yardımıyla GSE10470 deneyindeki örneklerin sıralaması Tablo 4.38’de gösterildiği gibi olmuştur.

**Tablo 4.38: Örneklerin Sıralanması – GSE10470**

Sıra No	Örnek Adı	Sıra No	Örnek Adı	Sıra No	Örnek Adı
1	wt.ctrl.1	11	ko.hca.11	21	wt.ctrl.21
2	wt.ctrl.2	12	wt.ctrl.12	22	ko.hca.22
3	wt.ctrl.3	13	ko.ctrl.13	23	wt.hca.23
4	wt.ctrl.4	14	wt.hca.14	24	ko.ctrl.24
5	ko.ctrl.5	15	ko.hca.15		
6	ko.ctrl.6	16	wt.ctrl.16		
7	ko.ctrl.7	17	wt.ctrl.17		
8	ko.ctrl.8	18	ko.hca.18		
9	ko.ctrl.9	19	wt.hca.19		
10	wt.hca.10	20	ko.ctrl.20		

Sıralanmış verilerin log<sub>2</sub> dönüşümü yapıldıktan sonra oluşturulan tasarım matrisi Tablo 4.39’da gösterilmektedir.

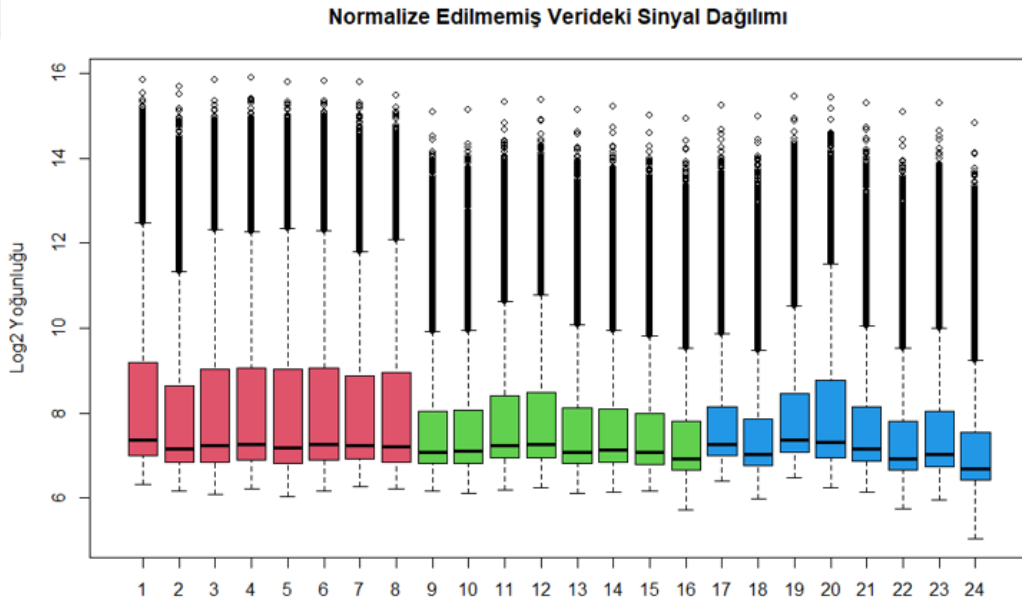
**Tablo 4.39: Tasarım Matrisi – GSE10470**

Örnek	Min	1st Qu.	Median	Mean	3rd Qu.	Max
wt.ctrl.1	6.317	7.008	7.365	8.279	9.196	15.828
wt.ctrl.2	6.162	6.833	7.167	7.971	8.633	15.681
wt.ctrl.3	6.081	6.844	7.225	8.119	9.033	15.847
wt.ctrl.4	6.217	6.898	7.265	8.158	9.046	15.895
ko.ctrl.5	6.042	6.806	7.187	8.094	9.023	15.795
ko.ctrl.6	6.154	6.896	7.266	8.159	9.049	15.816
ko.ctrl.7	6.276	6.908	7.237	8.084	8.863	15.778
ko.ctrl.8	6.208	6.848	7.204	8.087	8.946	15.478
ko.ctrl.9	6.164	6.809	7.075	7.666	8.050	15.070
wt.hca.10	6.119	6.823	7.100	7.657	8.067	15.129
ko.hca.11	6.184	6.941	7.234	7.901	8.411	15.330
wt.ctrl.12	6.242	6.948	7.259	7.920	8.486	15.373
ko.ctrl.13	6.115	6.812	7.089	7.697	8.117	15.133

Tablo 4.39 (devam)

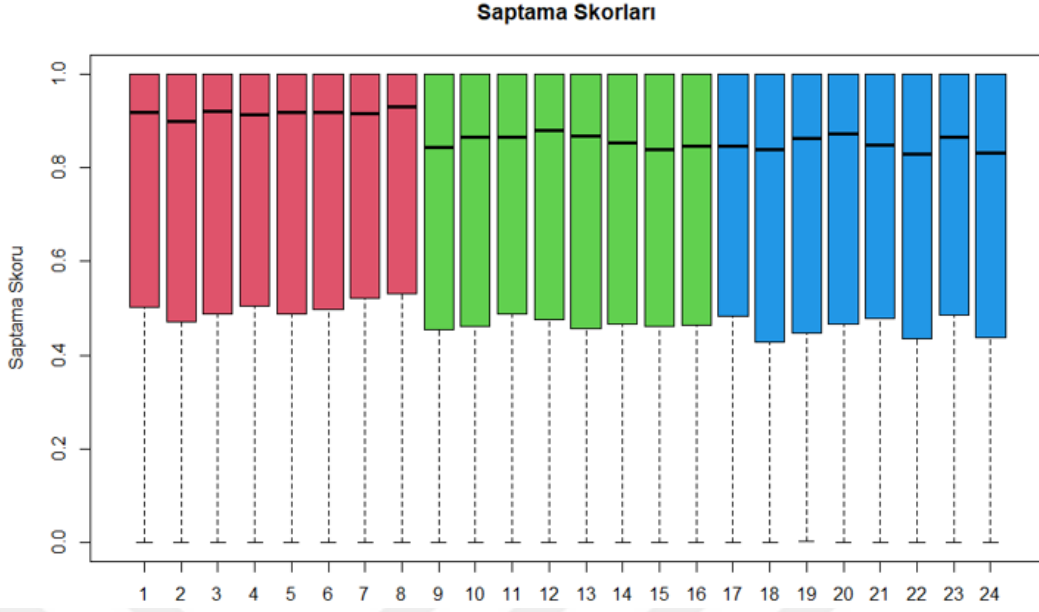
wt.hca.14	6.129	6.837	7.117	7.686	8.084	15.211
ko.hca.15	6.168	6.800	7.064	7.635	8.002	15.003
wt.ctrl.16	5.711	6.651	6.924	7.459	7.799	14.924
wt.ctrl.17	6.396	6.993	7.271	7.803	8.142	15.237
ko.hca.18	5.982	6.770	7.029	7.564	7.853	14.982
wt.hca.19	6.482	7.075	7.361	7.983	8.450	15.449
ko.ctrl.20	6.254	6.939	7.307	8.051	8.773	15.425
wt.ctrl.21	6.127	6.869	7.146	7.725	8.137	15.299
ko.hca.22	5.741	6.654	6.918	7.467	7.806	15.084
wt.hca.23	5.963	6.731	7.013	7.613	8.040	15.300
ko.ctrl.24	5.044	6.416	6.681	7.227	7.541	14.812

İşlenmemiş/ham verideki sinyal dağılımı kontrol edilmiştir. Normalizasyon işlemi öncesindeki sinyal dağılımı Şekil 4.61'deki kutu grafik ile gösterildiği gibidir.



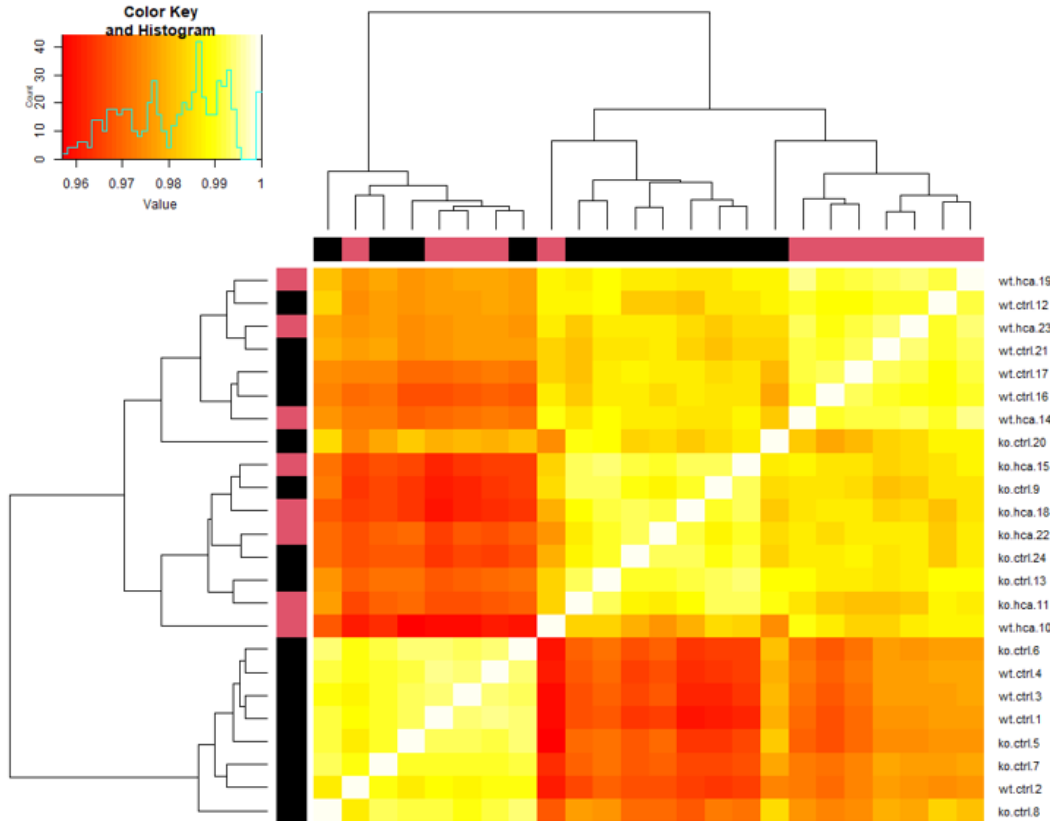
Şekil 4.61: Normalizasyon Öncesi Sinyal Dağılımı – GSE10470

Sinyal dağılımının kontrolünün ardından saptama skoru kontrol işlemi gerçekleştirilerek ham verideki genlerin 0-1 arasında ifade edilme olasılığı Şekil 4.62'de gösterilmiştir. (0: ifade edilmiyor, 1: ifade ediliyor)



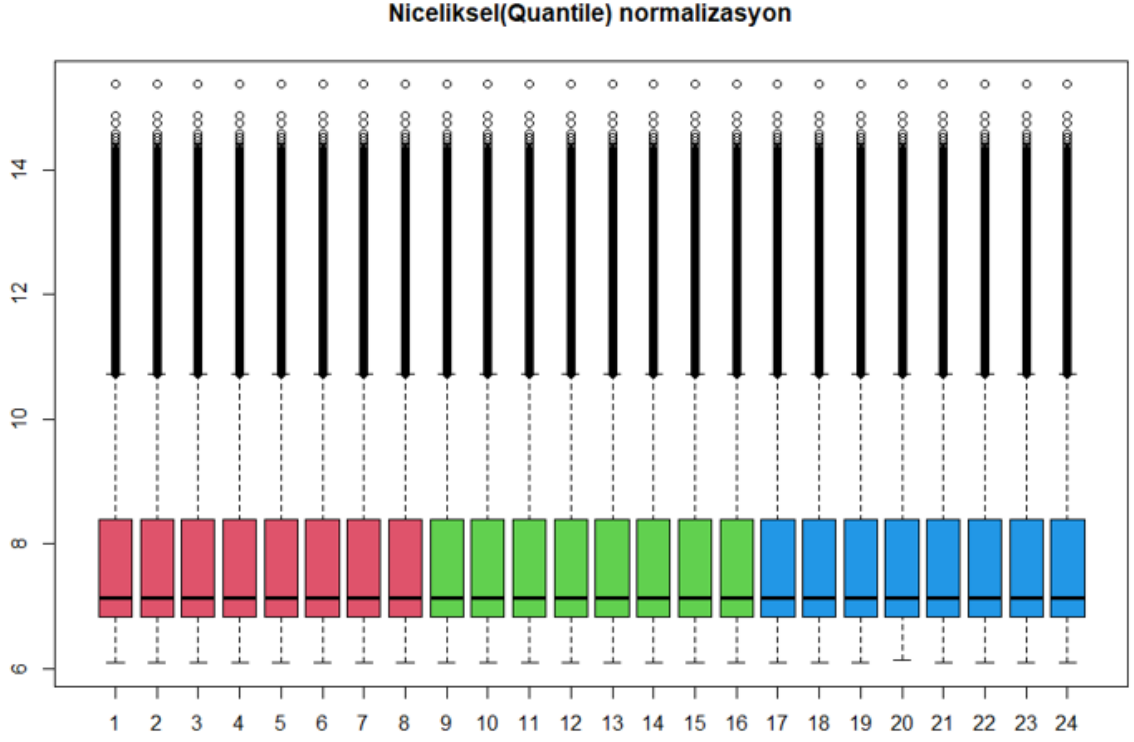
**Şekil 4.62:** Saptama Skorları – GSE10470

Veri kalitesi kontrolleri tamamlanan veriseti üzerinde pearson korelasyona bağlı kümeleme işlemi yapılmıştır. Kümelenmiş veri Şekil 4.63’deki ısı haritasındaki gibidir.



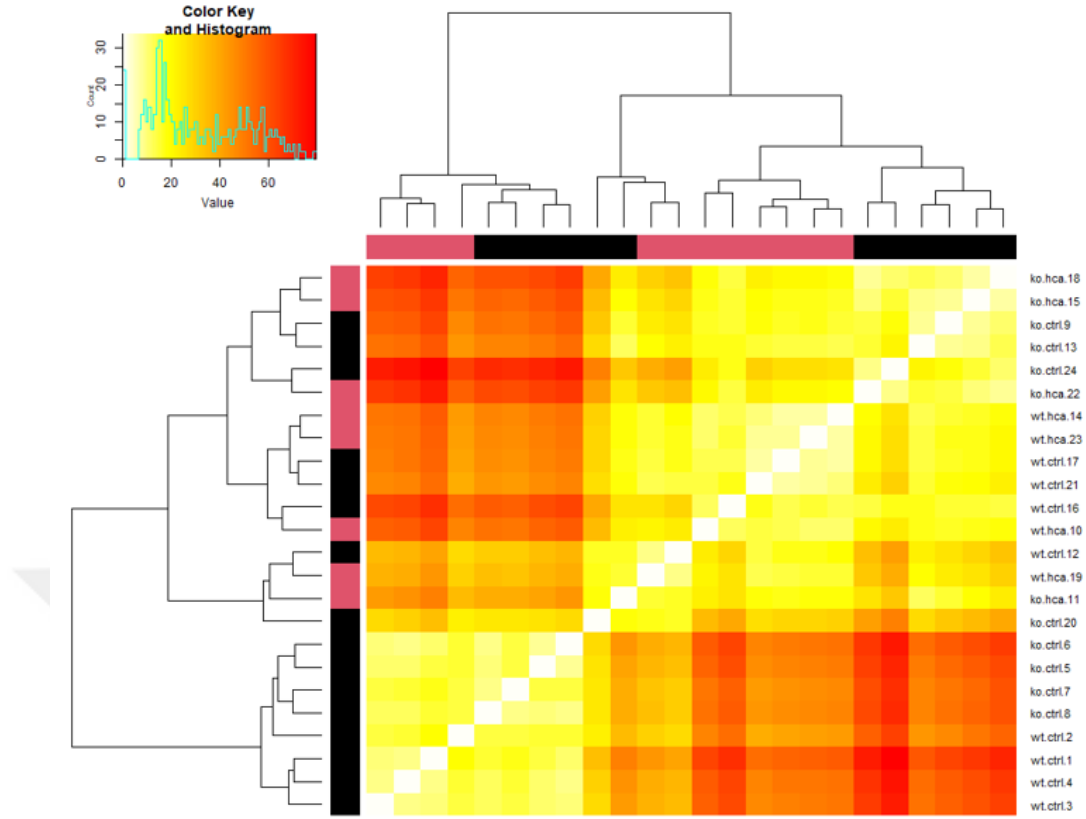
**Şekil 4.63:** Normalizasyon Öncesi Kümeleme Isı Haritası – GSE10470

Niceliksel normalizasyon işlemi ile veri 0-1 arasına ölçeklendirilmiştir. Normalizasyon işlemi sonrasında sinyal dağılımı ve ısı haritası tekrar hazırlanarak değişim incelenmiştir. Niceliksel normalizasyon sonrası sinyal dağılımı Şekil 4.64'deki gibidir.



**Şekil 4.64:** Normalizasyon Sonrası Sinyal Dağılımı – GSE10470

Normalizasyon sonrası ısı haritası Şekil 4.65'deki gibidir.



**Şekil 4.65:** Normalizasyon Sonrası Kümelenmiş Veri Isı Haritası – GSE10470

Normalizasyon işleminin ardından karşılaştırmaların yapılabilmesi için gruplar hazırlanmıştır:

- $KOwsWT = ko.ctrl - wt.ctrl$
- $HCAwt = wt.hca - wt.ctrl$
- $HCAko = ko.hca - ko.ctrl$
- $HCAwtnotHCAko = (wt.hca - wt.ctrl) - (ko.hca - ko.ctrl)$

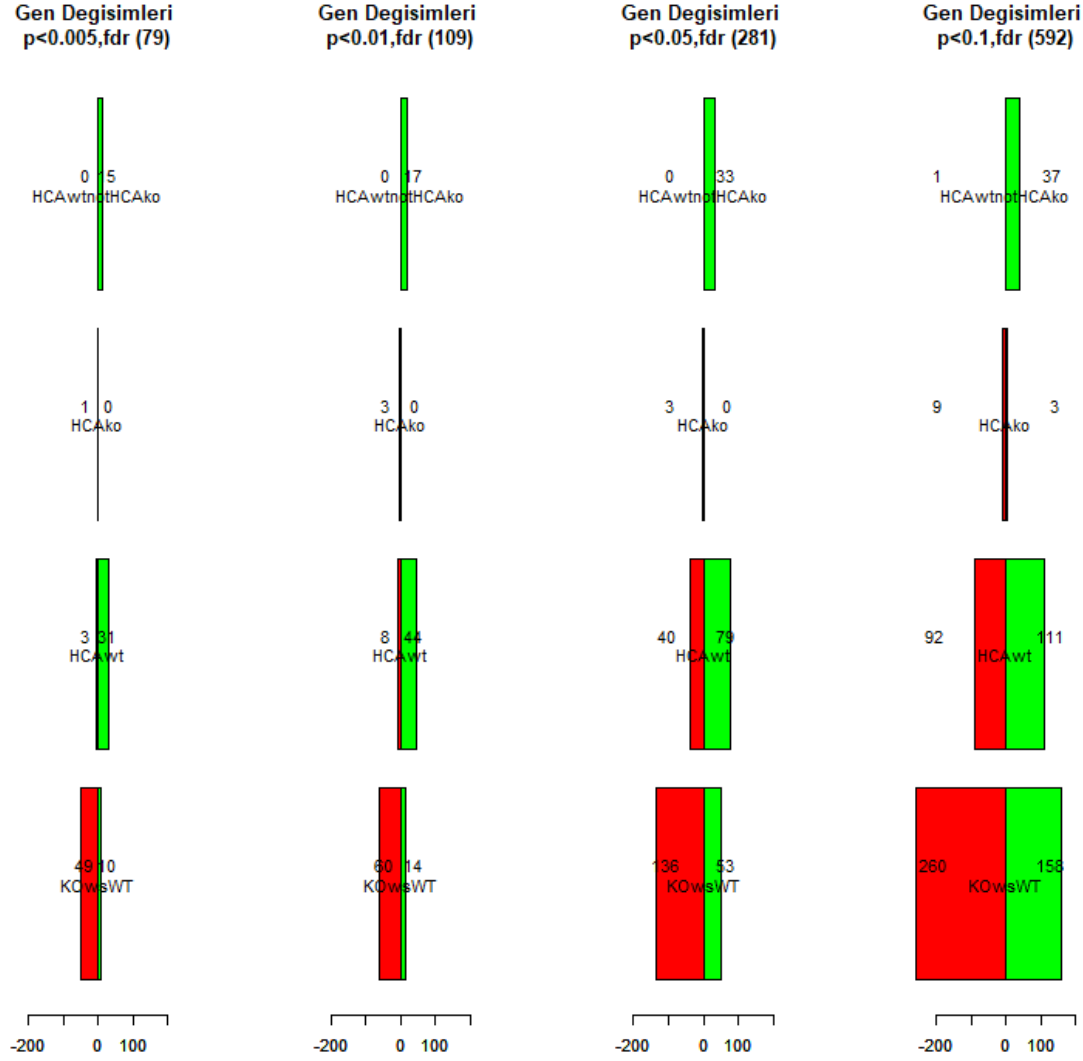
Karşılaştırma grupları için hazırlanan kontrast matrisi Tablo 4.40'da gösterilmektedir.

**Tablo 4.40:** Kontrast Matrisi – GSE10470

Seviye	KOwsWT	HCAwt	HCAko	HCAwtnotHCAko
wt.ctrl	-1	-1	0	-1
ko.ctrl	1	0	-1	1
wt.hca	0	1	0	1
ko.hca	0	0	1	-1

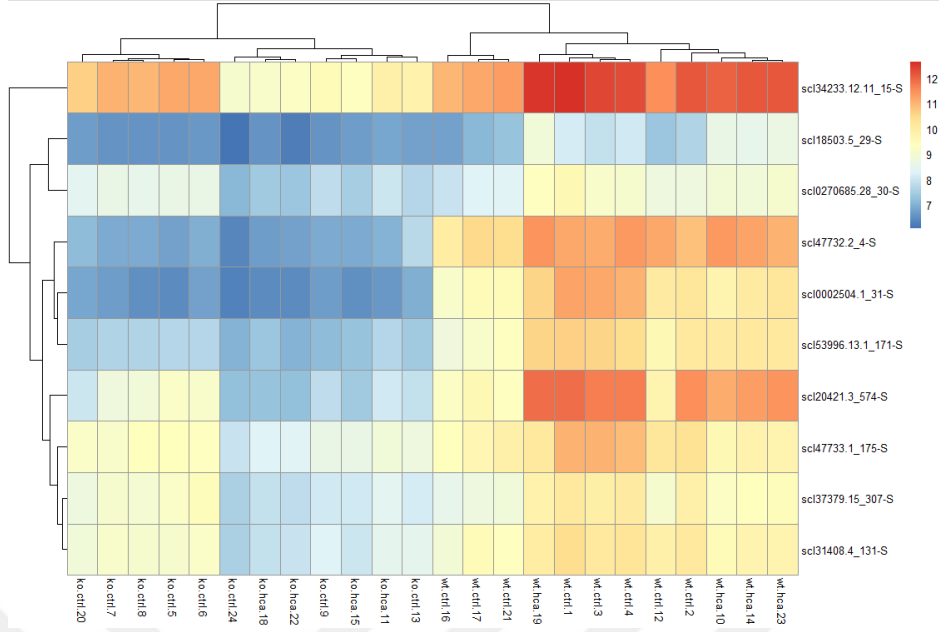
Karşılaştırma gruplarının hazırlanması sonrası SMATP yönteminin son aşaması uygulanarak veri değerlendirilmiştir. Değerlendirme sürecinde eBayes kullanılarak istatistiksel önem değerlendirilmiş, FDR istatistiği ile gözlem yapılmıştır. FDR'de sıklıkla kullanılan p-değeri ile kontroller yapılmıştır. P-değerinin 0'a yakın olması sonucun anlamlı olduğunu

göstermektedir. 4 farklı eşik değeri için kontroller yapıldığında gen değişimleri Şekil 4.66'daki gibi olmuştur. Farklı istatistiksel eşiklerde diferansiyel olarak ekprese edilen genlerin sayısının gösterildiği şekilde yeşil çubuklar yukarı düzenlenmiş genleri, kırmızı çubuklar ise aşağı düzenlenmiş genleri göstermektedir.



**Şekil 4.66:** Gruplar Arası Karşılaştırmalar – GSE10470

Sonuçlar incelendiğinde GSE10470 veriseti için en çok etkilenen 10 gen aşağıdaki gibi olmuştur.



**Şekil 4.67:** En Çok Etkilenen 10 Gen – GSE10470

### 4.3. GSE8262 VERİSETİ ANALİZ VE DEĞERLENDİRME

Fare beyninin kortikal ve hipokampal bölgesinde CCH (kronik tutarlı hipoksi) ve CIH (kronik aralıklı hipoksi) tedavisi uygulanan GSE8262 deneyine ait veriseti 145 kolon ve 46120 satır içermektedir. Veriseti 6 CCH uygulanmış örnek, 6 CIH uygulanmış örnek ve 6 kontrol örneği olmak üzere toplam 18 örnekten oluşmaktadır. Kronik hipoksi deneyinde SMATP yöntemi aşağıdaki gibi uygulanmıştır.

İlk adım olarak veriseti bilgisayar ortamına aktarılmıştır. Ardından verinin değerlendirilebilmesi için veri çerçevesi hazırlanmıştır. Hazırlanan veri çerçevesi Tablo 4.41’de gösterildiği gibidir.

**Tablo 4.41:** Veri Çerçevesi – GSE8262

Sıra	Grup	GenTipi	Replika	Tedavi	DiziKodu
1	1	Cortex	1	ctrl	1412023004_A
2	1	Cortex	1	cih	1412023004_B
3	1	Cortex	1	cch	1412023004_C
4	1	HCampus	1	ctrl	1412023004_D
5	1	HCampus	1	cih	1412023004_E
6	1	HCampus	1	cch	1412023004_F
7	2	Cortex	2	ctrl	1412023005_A
8	2	Cortex	2	cih	1412023005_B
9	2	Cortex	2	cch	1412023005_C
10	2	HCampus	2	ctrl	1412023005_D
11	2	HCampus	2	cih	1412023005_E
12	2	HCampus	2	cch	1412023005_F

**Tablo 4.41 (devam)**

13	3	Cortex	3	ctrl	1412023006_A
14	3	Cortex	3	cih	1412023006_B
15	3	Cortex	3	cch	1412023006_C
16	3	HCampus	3	ctrl	1412023006_D
17	3	HCampus	3	cih	1412023006_E
18	3	HCampus	3	cch	1412023006_F

Veri çerçevesi yardımıyla GSE8262 deneyindeki örneklerin sıralaması Tablo 4.42’de gösterildiği gibi olmuştur.

**Tablo 4.42: Örneklerin Sıralanması – GSE8262**

Sıra No	Örnek Adı
1	cx.ctrl.1
2	cx.cih.2
3	cx.cch.3
4	hc.ctrl.4
5	hc.cih.5
6	hc.cch.6
7	cx.ctrl.7
8	cx.cih.8
9	cx.cch.9
10	hc.ctrl.10
11	hc.cih.11
12	hc.cch.12
13	cx.ctrl.13
14	cx.cih.14
15	cx.cch.15
16	hc.ctrl.16
17	hc.cih.17
18	hc.cch.18

Sıralanmış verilerin log<sub>2</sub> dönüşümü yapıldıktan sonra oluşturulan tasarım matrisi Tablo 4.43’de gösterilmektedir.

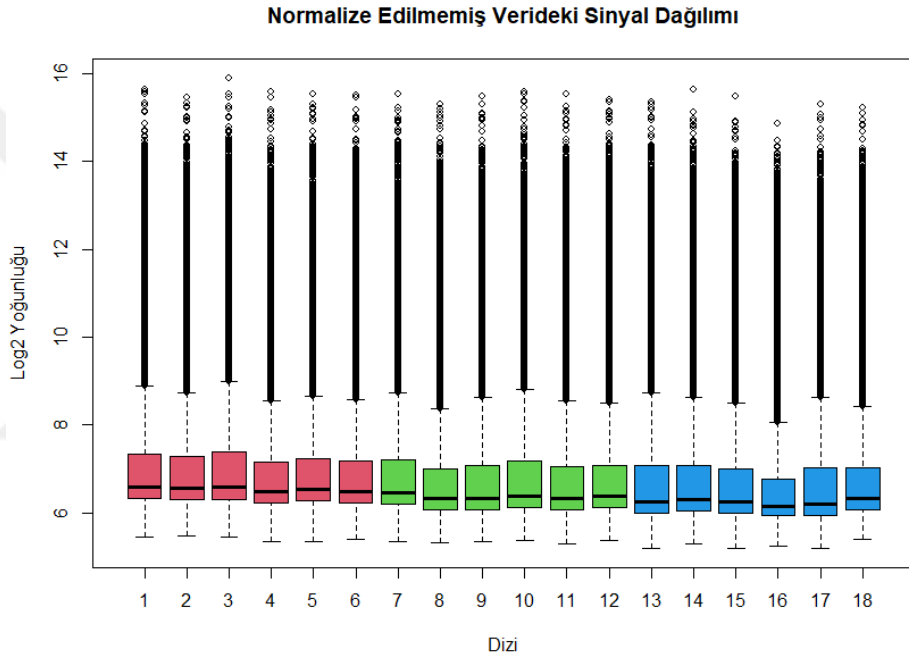
**Tablo 4.43: Tasarım Matrisi – GSE8262**

Örnek	Min	1st Qu.	Median	Mean	3rd Qu.	Max
cx.ctrl.1	5.440	6.322	6.591	7.141	7.345	15.634
cx.cih.2	5.472	6.307	6.559	7.106	7.276	15.445
cx.cch.3	5.443	6.309	6.580	7.155	7.379	15.896
hc.ctrl.4	5.354	6.235	6.492	6.999	7.159	15.569
hc.cih.5	5.347	6.285	6.536	7.069	7.229	15.534
hc.cch.6	5.385	6.236	6.492	7.013	7.173	15.501
cx.ctrl.7	5.336	6.194	6.458	7.020	7.212	15.531
cx.cih.8	5.329	6.081	6.326	6.856	6.992	15.290
cx.cch.9	5.350	6.064	6.331	6.883	7.085	15.481
hc.ctrl.10	5.365	6.117	6.370	6.982	7.190	15.576
hc.cih.11	5.293	6.057	6.326	6.866	7.058	15.532
hc.cch.12	5.382	6.133	6.382	6.928	7.077	15.396

Tablo 4.43 (devam)

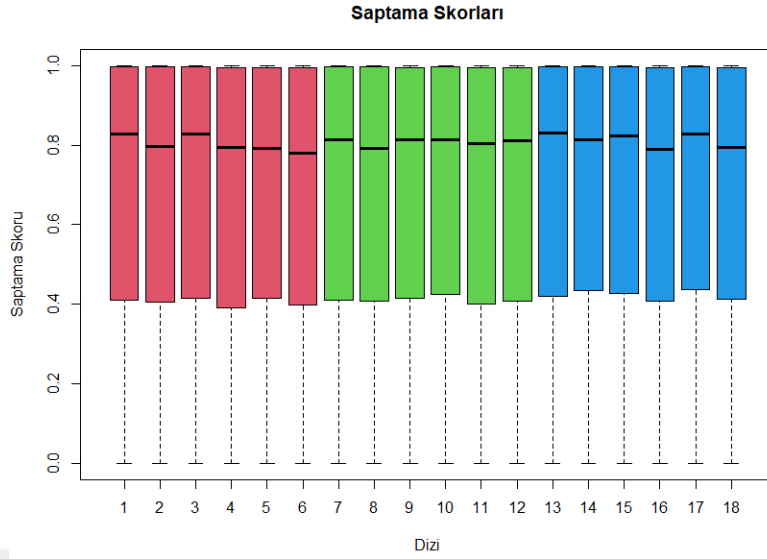
cx.ctrl.13	5.202	5.986	6.248	6.853	7.080	15.346
cx.cih.14	5.304	6.053	6.311	6.883	7.086	15.637
cx.cch.15	5.202	5.991	6.252	6.791	6.995	15.473
hc.ctrl.16	5.244	5.928	6.146	6.680	6.779	14.860
hc.cih.17	5.190	5.931	6.190	6.784	7.012	15.289
hc.cch.18	5.396	6.081	6.334	6.856	7.015	15.228

İşlenmemiş/ham verideki sinyal dağılımı kontrol edilmiştir. Normalizasyon işlemi öncesindeki sinyal dağılımı Şekil 4.68'deki kutu grafik ile gösterildiği gibidir.



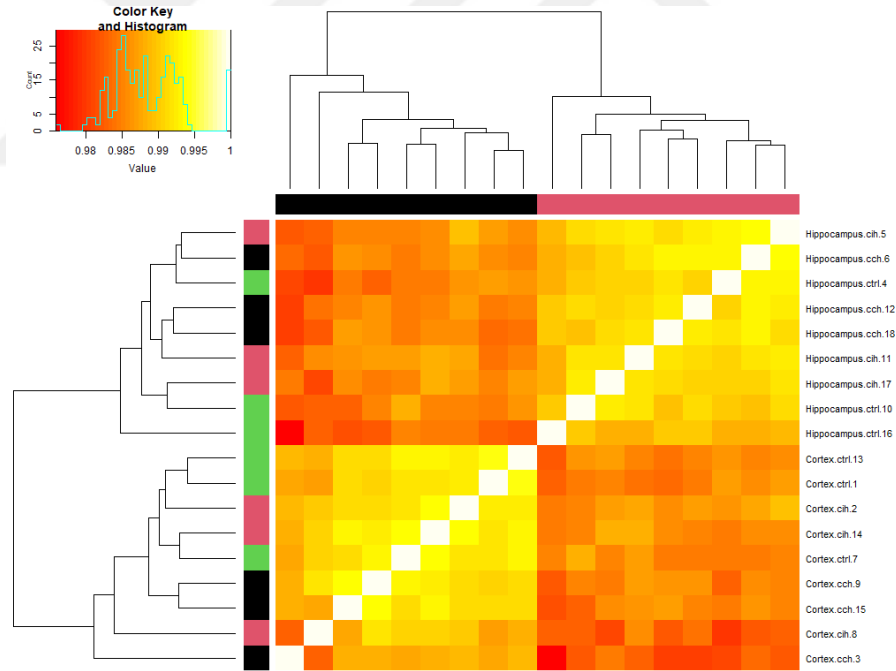
Şekil 4.68: Normalizasyon Öncesi Sinyal Dağılımı – GSE8262

Sinyal dağılımının kontrolünün ardından saptama skoru kontrol işlemi gerçekleştirilerek ham verideki genlerin 0-1 arasında ifade edilme olasılığı Şekil 4.69'da gösterilmiştir. (0: ifade edilmiyor, 1: ifade ediliyor)



**Şekil 4.69:** Saptama Skorları – GSE8262

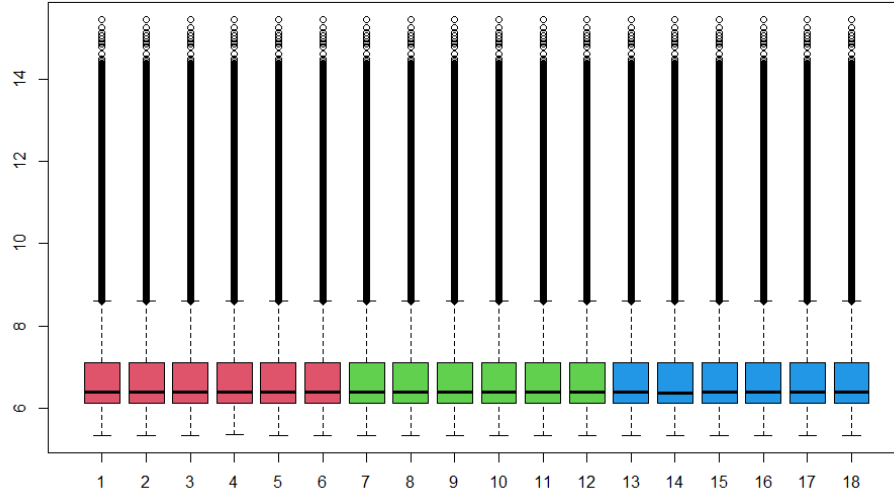
Veri kalitesi kontrolleri tamamlanan veriseti üzerinde pearson korelasyona bağlı kümeleme işlemi yapılmıştır. Kümelenmiş veri Şekil 4.70’deki ısı haritasındaki gibidir.



**Şekil 4.70:** Normalizasyon Öncesi Kümeleme Isı Haritası – GSE8262

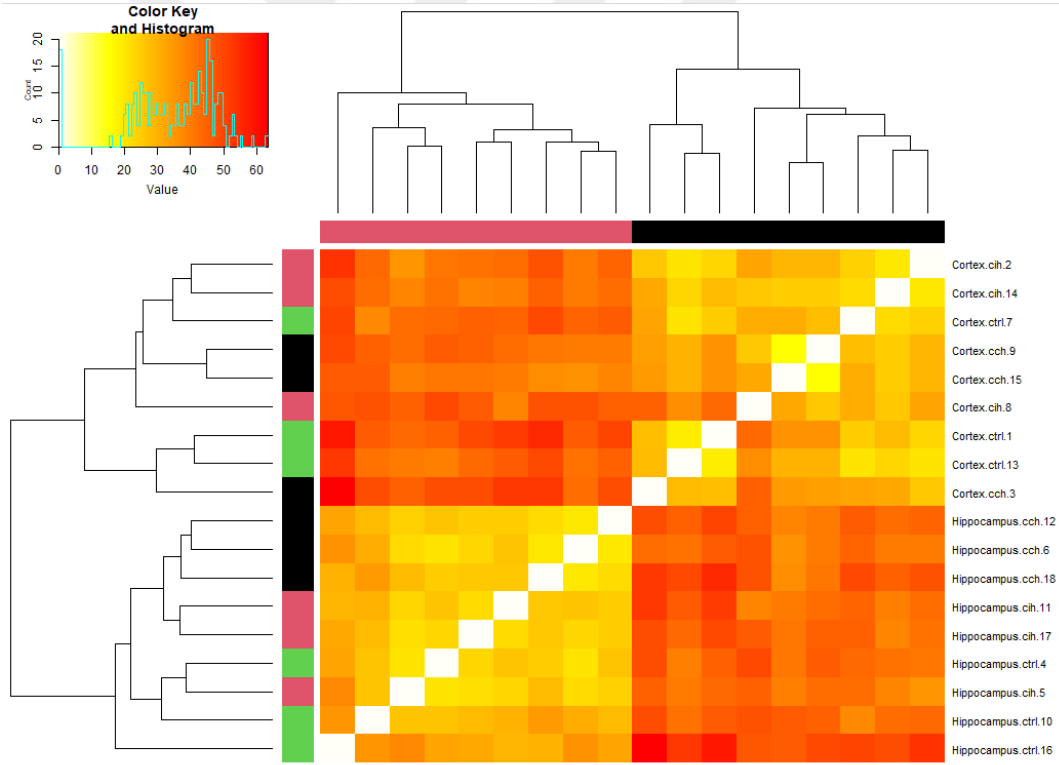
Niceliksel normalizasyon işlemi ile veri 0-1 arasına ölçeklendirilmiştir. Normalizasyon işlemi sonrasında sinyal dağılımı ve ısı haritası tekrar hazırlanarak değişim incelenmiştir. Niceliksel normalizasyon sonrası sinyal dağılımı Şekil 4.71’deki gibidir.

## Niceliksel(Quantile) normalizasyon



Şekil 4.71: Normalizasyon Sonrası Sinyal Dağılımı – GSE8262

Normalizasyon sonrası ısı haritası Şekil 4.72'deki gibidir.



Şekil 4.72: Normalizasyon Sonrası Kümelennmiş Veri Isı Haritası – GSE8262

Normalizasyon işleminin ardından karşılaştırmaların yapılabilmesi için gruplar hazırlanmıştır:

- CXwsHC= cx.ctrl- hc.ctrl
- cihHC= hc.cih- hc.ctrl

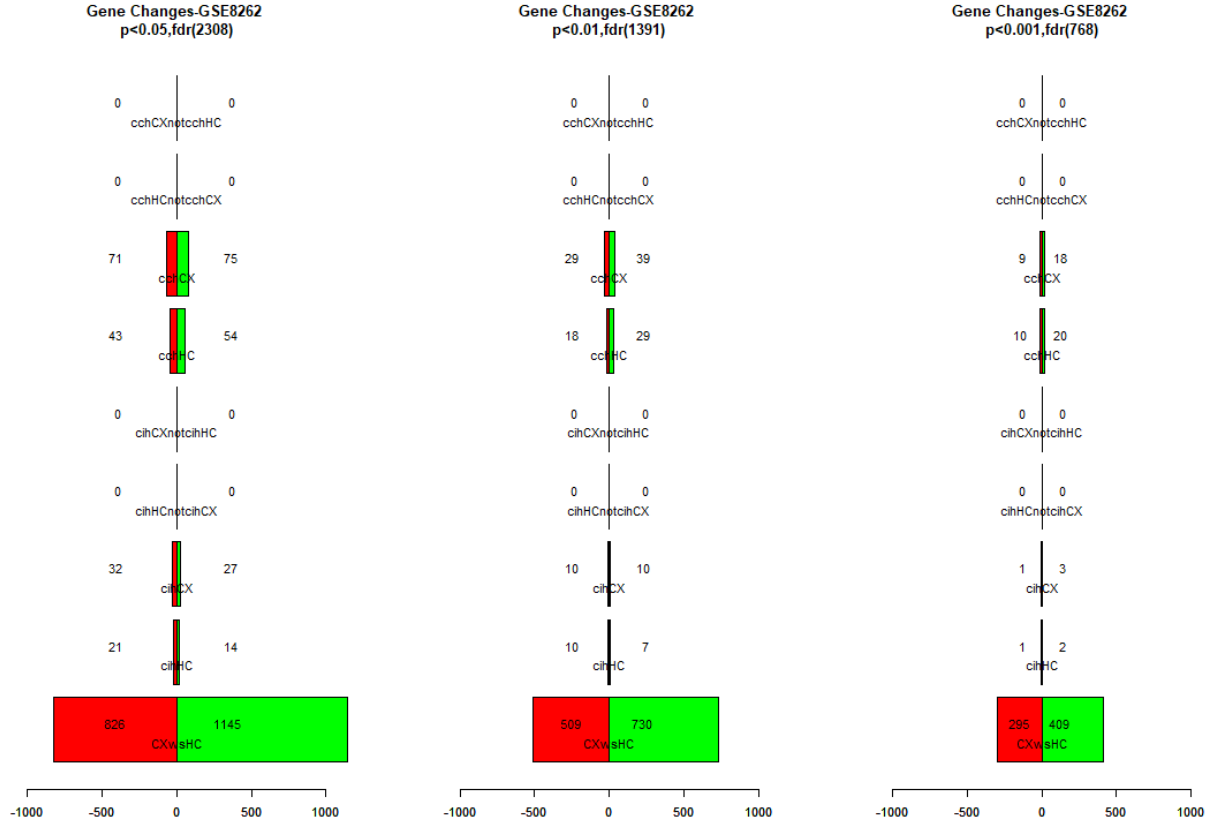
- $cihCX = cx.cih - cx.ctrl$
- $cihHCnotcihCX = (hc.cih - hc.ctrl) - (cx.cih - cx.ctrl)$
- $cihCXnotcihHC = (cx.cih - cx.ctrl) - (hc.cih - hc.ctrl)$
- $cchHC = hc.cch - hc.ctrl$
- $cchCX = cx.cch - cx.ctrl$
- $cchHCnotcchCX = (hc.cch - hc.ctrl) - (cx.cch - cx.ctrl)$
- $cchCXnotcchHC = (cx.cch - cx.ctrl) - (hc.cch - hc.ctrl)$

Karşılaştırma grupları için hazırlanan kontrast matrisi Tablo 4.44’de gösterilmektedir.

**Tablo 4.44:** Kontrast Matrisi – GSE8262

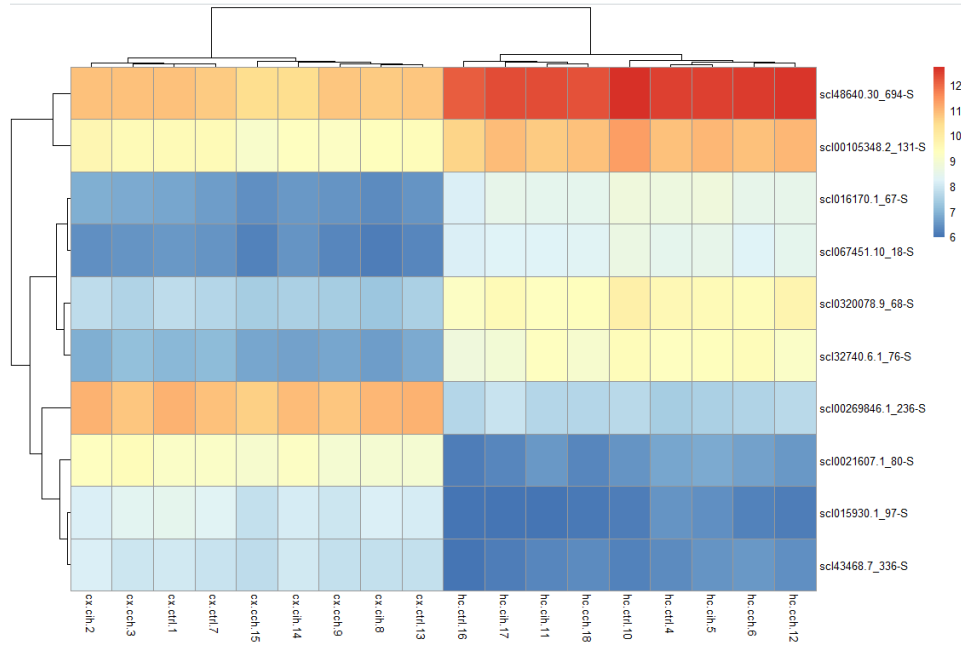
	CX Ws HC	cihHC	cihCX	cihHC not cihCX	cihCX not cihHC	cchHC	cchCX	cchHC not cchCX	cchCX not cchHC
cx.ctrl	1	0	-1	1	-1	0	-1	1	-1
cx.cih	0	0	1	-1	1	0	0	0	0
cx.cch	0	0	0	0	0	0	1	-1	1
hc.ctrl	-1	-1	0	-1	1	-1	0	-1	1
hc.cih	0	1	0	1	-1	0	0	0	0
hc.cch	0	0	0	0	0	1	0	1	-1

Karşılaştırma gruplarının hazırlanması sonrası SMATP yönteminin son aşaması uygulanarak veri değerlendirilmiştir. Değerlendirme sürecinde eBays kullanılarak istatistiksel önem değerlendirilmiş, FDR istatistiği ile gözlem yapılmıştır. FDR’de sıklıkla kullanılan p-değeri ile kontroller yapılmıştır. P-değerinin 0’a yakın olması sonucun anlamlı olduğunu göstermektedir. 3 farklı eşik değeri için kontroller yapıldığında gen değişimleri Şekil 4.73’deki gibi olmuştur. Farklı istatistiksel eşiklerde diferansiyel olarak ekprese edilen genlerin sayısının gösterildiği şekilde yeşil çubuklar yukarı düzenlenmiş genleri, kırmızı çubuklar ise aşağı düzenlenmiş genleri göstermektedir.



Şekil 4.73: Gruplar Arası Karşılaştırmalar – GSE8262

Sonuçlar incelendiğinde GSE8262 veriseti için en çok etkilenen 10 gen aşağıdaki gibi olmuştur.



Şekil 4.74: En Çok Etkilenen 10 Gen – GSE8262

## 5. TARTIŞMA

Bu tez çalışmasında kanser hücre soylarına farklı tedavi/ilaç uygulanarak gen değişimlerinin incelendiği özgün veriseti ile birlikte GSE10470 ve GSE8262 verisetleri kullanılmıştır. 9 farklı tedavi/ilaç uygulaması içeren özgün veriseti 25915 nolu proje kapsamında İstanbul Üniversitesi Bilimsel Araştırma Projeleri Koordinasyon Birimi desteği ile hazırlanmış projenin çıktısıdır. Bu veriseti ile daha önce yapılmış yayın/tez çalışması bulunmayıp, veriler bu çalışmada ilk kez değerlendirilmiştir. GSE10470 ve GSE8262 verisetleri ise GEO platformundan kullanıma sunulmuş, açık erişimli verisetleridir. Bu verisetlerini kullanan diğer çalışmalar incelendiğinde, laboratuvar çalışması ağırlıklı yayınlar olduğu görülmüş, 2.10 ve 2.11 bölümlerinde detaylarına yer verilmiştir. Tez çalışmasının yapıldığı tarih itibarıyla bu verisetlerini değerlendirmek için bilgisayar teknikleri ile yapılan ve gen bilgilerinin paylaşıldığı başka bir çalışma bulunmamakta, tez çalışması bu yönüyle özgün bir çalışma niteliği taşımaktadır.

Özgün veriseti altındaki 9 deney çıktısı ayrı ayrı değerlendirilmiştir. Çıktılar incelendiğinde “Nikotin”, “Işın” ve “Nikotin ve M100 (Nikotin ve M100’ün bir arada uygulandığı)” deneylerinde belirgin şekilde değişim gösteren gene rastlanmamıştır. Sadece Işın ve sadece Nikotin uygulanan deneylerde değişim gösteren gen gözlemlenemezken Nikotin ve Işın’ın bir arada uygulandığı durumda MMP1 ve ARHGDIB genlerinin aktifleştiği görülmüştür. Metformin uygulanan deneylerde, metformin dozu artırıldıkça etkilenen gen sayısının arttığı gözlemlenmiştir. M100 (100 milimolar metformin) tek başına uygulandığında ARHGDIB geninde büyük değişim görülürken M1000 (1000 milimolar metformin) dozunda uygulandığında MMP1, ARHGDIB ve FLNC genlerinin değişiminin güçlü olduğu görülmüştür. Nikotin, Işın ve M1000 maddelerinin bir arada uygulandığı deneyde ise MMP1, ARHGDIB en çok değişim gösteren genler olmuştur. Bu deney çıktıları Metformin ilacının meme kanseri sürecinde etkili olduğunu ve uygulanan doza bağlı olarak etkinin artabildiğini göstermektedir. Ayrıca Nikotin ve Işın ayrı ayrı tek başlarına uygulandıklarında büyük bir değişime sebep olmuyor gibi görünürken Nikotin ve Işın birarada uygulandığında gen değişimlerine sebep olabildikleri görülmektedir.

GSE10470 deneyinde oksidatif stresin fare nöronları üzerindeki etkisi incelenerek, bu süreçte hangi genlerin en çok etkilendiği bilgisi paylaşılmaya çalışılmıştır. Lange ve diğ. (2008) yaptığı çalışma laboratuvar ortamında denek fareler üzerinde yapılan çalışma ve bu çalışma sonucunda GSE10470 verisetinin oluşturulduğu deneysel çalışmadır. Çalışma tıbbi bir çalışma

olup kaynak verisetini hazırlamakta, bilgisayar teknolojileri ile verinin değerlendirilmesi sürecine değinilmemektedir. Woo ve diğ. (2021) yaptığı çalışmada, GRM8 proteininin merkezi sinir sistemindeki nöroprotektif etkisi araştırılmaktadır. Bu çalışmada GSE10470 dışında farklı verisetleri de kullanılmakta olup tez çalışmasındaki gibi bir teknolojik yaklaşım içermemektedir. Bu verisetini kullanan son çalışma Bloomingdale ve diğ. (2022) tarafından yapılmıştır. Bu çalışmada sadece GSE10470 değil başka verisetleri de kullanılarak bortozomib ilacının sebep olduğu nöropatinin tedavi süreci için seçenekler araştırılmıştır. Bu verini kullanan tüm çalışmalar laboratuvar çalışması ağırlıklı, tıbbi deneylerdir ve gen ekspresyon analiz sürecine odaklanılmamıştır. Tez çalışmasında GSE10470 veriseti gibi mikrodizi verilerinin makine öğrenmesi teknikleri ile incelenmesi ve hedef gen çıkarımı yapılmıştır. Çalışmada hem analiz süreci için geliştirilen yöntem hem de gen çıkarımının yapılarak gen bilgilerinin paylaşılması açısından özgün ve bu alanda yol gösterici nitelikte bir çalışma gerçekleştirilmiştir. Paylaşılan gen bilgileri alan uzmanları tarafından değerlendirilerek yeni araştırma-geliştirme süreçlerinin gerçekleştirilmesinde katkı sağlayabilecek niteliktedir.

GSE8262 verisetinde ise kronik hipoksinin fare beyni üzerindeki etkisi incelenerek, en çok etkilendiği düşünülen gen bilgileri paylaşılmaya çalışılmıştır. GSE8262 verisetinin deneysel olarak hazırlanması sürecini detaylandıran, Zhou ve diğ. (2008) yaptığı çalışma, verisetinin hazırlandığı ve sonucunda açık erişim ortamı üzerinden paylaşıldığı çalışmadır. Hipoksinin deneklere uygulanması ve sonuçların GEO üzerinden paylaşılması aşamalarını içermektedir. Wang ve diğ. (2021) yaptığı çalışma, GSE8262 dışında farklı verisetlerini de kullanan, SENP1 proteininin mikroglia üzerinde etkisini inceleyen bir çalışmadır. Bu verisetini kullanan diğer çalışma ise Wang ve diğ. (2023) tarafından yapılmış ve yine GSE8262 dışında farklı verisetleri kullanılarak SENP1 proteininin etkisi üzerine yapılmıştır. Bu çalışmalar deneysel çalışmalar olup analiz için bir yöntem veya gen bilgileri paylaşımı yapılmamıştır. Tez çalışmasında hem verinin analiz süreci için geliştirilen makine öğrenmesi teknikleri tabanlı bir yöntemi detaylandırılmış hem de gen çıkarım süreci sonucunda etkilenen gen bilgileri paylaşılmıştır. Geliştirilen SMATP analiz yöntemi benzer verisetlerini incelemek için araştırmacılara yol gösterebilecek nitelikte olması açısından diğer çalışmalardan ayrılmaktadır. Ayrıca çalışmada elde edilen genler, biyomedikal çalışmalar için kaynak oluşturmakta ve bu alanda destek sağlayabilecek nitelikte listelenmektedir. Bu yönleriyle çalışma deneysel diğer çalışmalardan ayrılmakta ve farklı bir bakış açısı sunmaktadır.

Bu deneylerin ortak özelliği Illumina BeadStudio/GenomeStudio formatında mikrodizi verileri olması olup, tez çalışmasında bu tür verilerin hem analiz edilmesi hem de hedef gen

ıkarımı srecinde kullanılabilir SMATP adı verilen yntem geliřtirilmiřtir. Birbirinden farklı ierikteki verisetlerinde bařarılı olan mikrodizi analiz ve gen ıkarım yntemi, bu trde verilerin incelenmesi ve gen ekspresyon farklılıklarının ortaya ıkarılmasında kullanılabilir.

Bu tez alıřma sunulmaya alıřan ıktıların, kiřileřtirilmiř tıp, ila/tedavi geliřtirme sreleri gibi alanlarda olumlu katkı saėlayabileceėi dřnlmektedir. Ayrıca SMATP ynteminin gen ekspresyon verilerinin analiz ve deėerlendirme srecinde yeni alıřmalara yardımcı olabileceėi dřnlmektedir.



## 6. SONUÇ VE ÖNERİLER

Bu çalışmada özgün verisetinin yanısıra açık erişimli verisetleri kullanılarak mikrodizi verileri üzerinde gen ekspresyon analizi ve hedef gen çıkarımı yapılmıştır. Mikrodizi türündeki gen ekspresyon verilerinin değerlendirilmesi için makine öğrenmesi tekniklerini kullanan bir yöntem geliştirilmiştir. Mikrodizi verileri gibi büyük tıbbi verilerin makine öğrenmesi teknikleri yardımıyla incelenmesi hızlı ve doğru sonuçlar elde edebilmek açısından katkı sağlamaktadır. Mikrodizi türündeki farklı verisetlerinde başarıyla uygulanan SMATP yönteminin oluşturulma süreci ayrı başlıklar altında detaylıca anlatılmıştır. Bu çalışma mikrodizi verilerinin değerlendirilmesi sürecinde izlenmesi gereken adımları sistematik bir şekilde bir araya getirmesi açısından benzer çalışmalara katkı sağlayabilecek niteliktedir.

Özgün verisetinin ilk kez değerlendirildiği bu çalışma tüm verisetleri için analiz sürecini detaylandırması açısından bu yöndeki çalışmalara örnek olabilecek niteliktedir. Ayrıca etkilenen gen bilgileri alan uzmanları tarafından değerlendirilerek hastalık seyrinin anlaşılması, ilaç/tedavi geliştirilmesi süreçlerinde katkı sağlayabilir.

## KAYNAKLAR

- Affymetrix, <https://www.thermofisher.com/tr/en/home/life-science/microarray-analysis.html>, Erişim Tarihi: 22.11.2024.
- Agilent, <https://www.agilent.com/en/product/software-informatics/genomics-software-informatics/gene-expression/genespring-gx>, Erişim Tarihi: 22.11.2024.
- ArrayExpress, <https://www.ebi.ac.uk/arrayexpress/>, Erişim Tarihi: 22.11.2024.
- Aubert, J., Bar-Hen, A., Daudin, J. J., and Robin, S., 2004. Determination of the differentially expressed genes in microarray experiments using local FDR. *BMC bioinformatics*, 5, 1-9.
- Bioconductor, <https://www.bioconductor.org/>, Erişim Tarihi: 22.05.2024
- Bishara, A. J., and Hittner, J. B., 2012. Testing the significance of a correlation with nonnormal data: comparison of Pearson, Spearman, transformation, and resampling approaches. *Psychological methods*, 17(3), 399. DOI: 10.1037/a0028087
- Bloomingdale, P., Meregalli, C., Pollard, K., Canta, A., Chiorazzi, A., Fumagalli, G., Monza, L., Pozzi, E., Alberti, P., Ballarini, E., Oggioni, N., Carlson, L., Liu, W., Ghandili, M., Ignatowski, T.A., Lee, K.P., Moore, M.J., Cavaletti, G. and Mager, D. E., 2022. Systems pharmacology modeling identifies a novel treatment strategy for bortezomib-induced neuropathic pain. *Frontiers in pharmacology*, 12, 817236.
- Bolstad, B. M., Irizarry, R. A., Åstrand, M., and Speed, T. P., 2003. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*, 19(2), 185-193. [DOI:10.1093/bioinformatics/19.2.185]
- Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., Aach, J., Ansorge, W., Ball, C.A., Causton, H.C., Gaasterland, T., Glenisson, P., Holstege, F.C.P., Kim, I.F., Markowitz, V., Matese, J.C., Parkinson, H., Robinson, A., Sarkans, U., Schulze-Kremer, S., Stewart, J., Taylor, R., Vilo, J and Vingron, M., 2001. Minimum information about a microarray experiment (MIAME)—toward standards for microarray data. *Nature genetics*, 29(4), 365-371.
- Bustin, S.A., 2000. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *Journal of molecular endocrinology*, 169-193.
- Chandra, B., and Gupta, M. (2011). Robust approach for estimating probabilities in Naïve-Bayes Classifier for gene expression data. *Expert systems with applications*, 38(3), 1293-1298.
- Chen, K. H., Wang, K. J., Tsai, M. L., Wang, K. M., Adrian, A. M., Cheng, W. C., Yang, T.S., Teng, N.C., Tan, K.P., and Chang, K. S., 2014-A. Gene selection for cancer identification: a decision tree model empowered by particle swarm optimization algorithm. *BMC bioinformatics*, 15, 1-10.
- Chen, Y. C., Ke, W. C., and Chiu, H. W., 2014-B. Risk classification of cancer survival using ANN with gene expression data from multiple laboratories. *Computers in biology and medicine*, 48, 1-7.

- Collins, F.S., and Varmus, H., 2015. A New Initiative on Precision Medicine. *The new England journal of medicine*, 793-795.
- Çavdar, Z.Y., Ensari, T., Şener, L.T., and Sertbaş, A., 2022. Makine Öğrenmesi Tabanlı Mikrodizi Tekniği ile MikroRNA Hedef Tanmini: Araştırma Çalışması. *Avrupa bilim ve teknoloji dergisi*, 39-45.
- Çavdar, Z. Y., Sertbaş, A., and Ensari, T., 2024. Hedef Gen Çıkarımı İçin Makine Öğrenmesi Tabanlı Mikrodizi Verilerinin Analizi. *Bilgisayar bilimleri ve mühendisliği dergisi*, 17(2), 174-183.
- DataMed, <https://datamed.org/>, Erişim Tarihi: 22.11.2024.
- GEO, <https://www.ncbi.nlm.nih.gov/geo/>, Erişim Tarihi: 22.11.2024.
- Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., and Lander, E.S., 1999. Fmolecular classification of cancer: class discovery and class prediction by gene expression monitoring, foundations of Constructive Matchematics. *Science*, 531-537.
- Han, J., Kamber, M., and Pei, J., 2012. Data Mining: Concepts and. Techniques, *Waltham: morgan kaufmann publishers*.
- Hanahan, D., and Weinberg, R.A., 2011. Hallmarks of Cancer: The Next Generation. *Cell*, 646-674.
- Huang, S., Cai, N., Pacheco, P. P., Narrandes, S., Wang, Y., and Xu, W., 2018. Applications of support vector machine (SVM) learning in cancer genomics. *Cancer genomics & proteomics*, 15(1), 41-51.
- Illumina BeadStuio, <https://www.illumina.com>, Erişim Tarihi: 22.011.2024.
- Jain, A. K., 2010. Data clustering: 50 years beyond K-means. *Pattern recognition letters*, 31(8), 651-666.
- Jiang, H., Wang, J., Li, M., Lan, W., Wu, F. X., and Pan, Y., 2018. miRTRS: a recommendation algorithm for predicting miRNA targets. *IEEE/ACM transactions on computational biology and bioinformatics*, 17(3), 1032-1041.
- Karagün, B.Ş., Antmen, B., Şaşmaz, İ., and Kılınç, Y., 2014. Mikro RNA ve Kanser. *Türk klinik biyokimya dergisi*, 12(1), 45-56.
- Lange, P. S., Chavez, J. C., Pinto, J. T., Coppola, G., Sun, C. W., Townes, T. M., Geschwind, D.H., and Ratan, R. R., 2008. ATF4 is an oxidative stress-inducible, prodeath transcription factor in neurons in vitro and in vivo. *The journal of experimental medicine*, 205(5), 1227-1242.
- Lavine, B. K., Davidson, C. E., and Rayens, W. S., 2004. Machine learning based pattern recognition applied to microarray data. *Combinatorial chemistry & high throughput screening*, 7(2), 115-131.
- LeCun, Y., Bengio, Y., and Hinton, G., 2015. Deep learning, *Nature*, 521(7553), 436-444.
- Narry, K.V., 2005. Small RNAs: Classification, Biogenesis, and Function. *Mol cells*, 19(1), 1-15.
- Özkan, Y., and Erol, Ç.S., 2017. *Biyoenformatik dna mikrodizi veri madenciliği*, 2.baskı, Papatya Bilim, ISBN-13: 978-6054220892.

- Öztemur, Y., Aydos, A., and GÜR-DEDEOĞLU, B., 2014. Meme kanseri mikrodizin verilerinin biyoinformatik yöntemler ile bir araya getirilmesi-Meta-analiz yaklaşımları. *Türk hijyen ve deneysel biyoloji dergisi*, 72(2), 155-162.
- Pawitan, Y., Michiels, S., Koscielny, S., Gusnanto, A., and Ploner, A., 2005. False discovery rate, sensitivity and sample size for microarray studies. *Bioinformatics*, 21(13), 3017-3024.
- R Programlama Dili, <https://www.r-project.org/>, Erişim Tarihi: 22.05.2024.
- Schena, M., Shalon, D., Davis, R.W., and Brown, P.O., 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, 467-470.
- Sherafatian, M., 2018. Tree-based machine learning algorithms identified minimal set of miRNA biomarkers for breast cancer diagnosis and molecular subtyping. *Gene*, 677, 111-118.
- Sticht, C., De La Torre, C., Parveen, A., and Gretz, N., 2018. miRWalk: an online resource for prediction of microRNA binding sites. *Plos one*, 13(10), e0206239.
- Sutton, R. S., and Barto, A. G., 2018. Reinforcement learning: An introduction. *MIT press*.
- Wang, H., Yang, T., Sun, J., Zhang, S., and Liu, S., 2021. SENP1 modulates microglia-mediated neuroinflammation toward intermittent hypoxia-induced cognitive decline through the de-SUMOylation of NEMO. *Journal of cellular and molecular medicine*, 25(14), 6841-6854.
- Wang, H., Wang, X., Shen, Y., Wang, Y., Yang, T., Sun, J., and Liu, S., 2023. SENP1 modulates chronic intermittent hypoxia-induced inflammation of microglia and neuronal injury by inhibiting TOM1 pathway. *International immunopharmacology*, 119, 110230.
- Wang, Z., Gerstein, M., and Snyder, M., 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews genetics*, 57-63.
- Woo, M. S., Ufer, F., Rothhammer, N., Di Liberto, G., Binkle, L., Haferkamp, U., Sonner, J.K., Engler, J.B., Hornig, S., Bauer, S., Wagner, I., Egervari, K., Raber, J., Duvoisin, R.M., Pless, O., Merkler, D. and Friese, M. A., 2021. Neuronal metabotropic glutamate receptor 8 protects against neurodegeneration in CNS inflammation. *Journal of experimental medicine*, 218(5), e20201290.
- Zhao, Y., Wong, L., and Goh, W. W. B., 2020. How to do quantile normalization correctly for gene expression data analyses. *Scientific reports*, 10(1), 15534.
- Zhou, D., Wang, J., Zapala, M. A., Xue, J., Schork, N. J., and Haddad, G. G., 2008. Gene expression in mouse brain following chronic hypoxia: role of sarcospan in glial cell death. *Physiological genomics*, 32(3), 370-379.

## İNTİHAL RAPORU İLK SAYFASI

Zerrin YILDIZ CAVDAR

ORJİNALLİK RAPORU

%6

BENZERLİK ENDEKSİ

%6

İNTERNET KAYNAKLARI

%3

YAYINLAR

%2

ÖĞRENCİ ÖDEVLERİ

BİRİNCİL KAYNAKLAR

1	Submitted to The Scientific & Technological Research Council of Turkey (TUBITAK) Öğrenci Ödevi	%2
2	acikbilim.yok.gov.tr İnternet Kaynağı	%1
3	www.semel.ucla.edu İnternet Kaynağı	%1
4	www.openaccess.hacettepe.edu.tr:8080 İnternet Kaynağı	<%1
5	www.slideshare.net İnternet Kaynağı	<%1
6	dergipark.org.tr İnternet Kaynağı	<%1
7	Çiftçi, Bekir. "Beden Eğitime ve Spor Yüksekokulu Öğrencilerinin Yeme Tutumlarının İncelenmesi (Balıkesir İl Örneği)", Balıkesir University (Turkey), 2024 Yayın	<%1
8	app.trdizin.gov.tr İnternet Kaynağı	<%1