



Hedef Gen Çıkarımı için Makine Öğrenmesi Tabanlı Mikrodizi Verilerinin Analizi

Machine Learning Based Microarray Data Analysis For Target Gene Inference

Zerrin YILDIZ ÇAVDAR
İstanbul Üniversitesi Cerrahpaşa
Bilgisayar Mühendisliği Bölümü
İstanbul, Türkiye
zerrintt@gmail.com
ORCID: 0000-0003-4299-5344

Ahmet SERTBAŞ
İstanbul Üniversitesi Cerrahpaşa
Bilgisayar Mühendisliği Bölümü
İstanbul, Türkiye
asertbas@iuc.edu.tr
ORCID: 0000-0001-8166-1211

Tolga ENSARI
Arkansas Tech University
Department of Engineering and
Computing Sciences
Russellville, USA
tensari@atu.edu
ORCID: 0000-0003-0896-3058

Öz

Mikrodizi teknolojisi gen ifadesindeki farklılıkların tespit edilmesinde kullanılır. Bu teknoloji ilaç geliştirme süreçlerinden tedavi süreçlerinin iyileştirilmesine birçok alanda katkı sağlamaktadır. Bu çalışmada, kronik hipoksi tedavisinin fare beyni üzerindeki etkisi ve oksidatif strese maruz kalan fare nöronlarının gen üzerindeki etkisi ile ilgili veri kümeleri üzerinde mikrodizi analizi yapılmıştır. Çalışmada açık erişim sağlanabilen iki farklı mikrodizi veri kümesi kullanılmıştır ve makine öğrenmesi yöntemleri ile çalışılmıştır. İlk adım olarak veri kümelerinin indirilmesi, ardından ön işleme tabi tutularak normalizasyon uygulanması sağlanmıştır. Bu aşamalar ile veri kümeleri gen çıkarımı için çalışmaya uygun hale getirilmiştir. Hazırlanan bu gen ekspresyon ifadeleri üzerinden istatistiksel ve makine öğrenmesi yöntemleri kullanılarak analizler gerçekleştirilmiş ve hedef gen çıkarımı sağlanmıştır.

Anahtar sözcükler: Mikrodizi, makine öğrenmesi, hedef gen çıkarımı, gen ekspresyon analizi

Abstract

Microarray technology is used to detect differences in gene expression. This technology contributes to many fields, from drug development processes to the improvement of treatment procedures. In this study, microarray analysis was performed on datasets related to the effect of chronic hypoxia treatment on the mouse brain and the effect of oxidative stress on the genes of mouse neurons. Two

different publicly accessible microarray datasets were used in the study, and machine learning methods were employed. The first step involved downloading the datasets, followed by preprocessing and normalization. These steps made the datasets suitable for gene extraction. Statistical and machine learning methods were applied to the prepared gene expression profiles, leading to the inference of target genes.

Keywords: Microarray, machine learning, target gene inference, analyzing gene expression

1. Giriş

Mikrodizin, biyoçip, gençip olarak da adlandırılan mikrodizi teknolojisi, gen ekspresyonunun incelenmesinde yaygın olarak kullanılan bir teknolojidir. Mikrodizi veri analizi, kanser araştırmaları, tedavi süreçlerinin iyileştirilmesi veya ilaç geliştirme gibi birçok alanda önemli bir araç olmuştur. Bu yaygın çalışmada mikrodizi teknolojisi yardımıyla, hastalık tedavi süreçlerinin etkili olduğu gen ifadesi farklılıkları incelenmiştir.

Mikrodiziler genomun (bir canlıya ait tüm DNA bilgisi) bir çip üzerinde görüntülenmesini sağlayarak gen ifadesindeki farklılıkları ortaya çıkarmayı sağlayan bir teknolojidir [1]. Değişkenlerin sayısından daha az deney replikasyonu yapılması veya gen düzenlemesi hakkında sınırlı bilgi sahibi olunması gibi faktörler mikrodizi analizini karmaşık hale getirmektedir [2]. Ancak kullanılan yazılımların hızla gelişmesi ve yeni teknikler ile bu olumsuzluklar her geçen gün daha da aşılmakta ve başarılı sonuçlar elde edilmektedir. Illumina BeadStudio [3], rastgele düzenlenmiş boncuklara dayanan mikrodizi teknolojisi, prob replikasyonu ve düşük hibritleşme gibi avantajları ile mikrodizi analizlerinde tercih edilen bir

platformudur. Gunderson ve arkadaşları [4] çalışmalarında bu platformun, geliştirdiği rastgele düzenlenmiş boncuklara dayanan mikrodizi teknolojisi; prob replikasyonu, düşük hibritleşme maliyeti gibi özellikleri sayesinde her geçen gün daha popüler olduğunu ifade etmişlerdir. Wang ve arkadaşları [5] ise Illumina BeadStudio'nun rastgele düzenlenmiş boncuk tabanlı mikrodizi teknolojisi ile gen ekspresyon analizi ve genotipleme çalışmalarında yaygın kullanıldığını belirtmişlerdir. Platform, her bir boncuğun yüzeyinde rastgele ancak yüksek yoğunlukta ve çoğaltılmış prob dizilerinin bulunmasıyla, sinyal-gürültü oranını artırarak yüksek hassasiyetli sonuçlar elde edilmesini sağlar [6]. Böylece geleneksel sabit prob dizilerine göre daha hassas veri elde edilmesine olanak sağlar. Illumina BeadStudio, prob replikasyonu ve düşük hibritleşme maliyetleri gibi avantajlara sahip olmasının yanında, genetik varyasyonların ve gen ekspresyon profillerinin güvenilir ve tekrarlanabilir analizini mümkün kılmaktadır [7]. Illumina BeadStudio'da kullanılan rastgele düzenlenmiş boncuk yapısı, yüzeye yüksek yoğunlukta yerleştirilmiş mikroskobik boncuklar üzerine kuruludur. Her boncuk, belirli bir prob dizisini taşır. Bu dizilerin başka kopyaları da yüzeyde bulunur. Bu durum sinyal-gürültü oranını artırarak daha doğru sonuçlar elde edilmesine yardımcı olur [8]. Bu yaklaşım, geleneksel mikrodizi platformlarına göre daha düşük maliyetli ve daha güvenilir bir analiz seçeneği sağlar. Bu çalışmada Illumina BeadStudio ile hazırlanmış mikrodizi verileri kullanılmıştır.

NCBI (National Center of Biotechnology Information – Ulusal Biyoteknoloji Bilgi Merkezi) geniş bir veri tabanı koleksiyonuna sahip bir platformdur. <https://ncbi.nlm.nih.gov> adresinden bu platforma erişim sağlanabilmektedir. NCBI altındaki GEO (Gene Expression Omnibus – Gen İfade Dizini) açık erişimli bir genomik veri tabanıdır. Bu çalışmada GEO altından erişilebilen GSE8262 ve GSE10470 nolu veri kümeleri kullanılmıştır [9]. Bu veri kümeleri fareler üzerinde yapılan çalışmalardan oluşmaktadır. GSE8262 veri kümesi kronik hipoksi tedavisinin fare beyni üzerindeki etkisini incelemek için hazırlanmıştır. GSE10470 veri kümesi ise oksidatif strese maruz kalan fare nöronlarının gen üzerindeki etkisini incelemek için hazırlanmıştır.

Mikrodizi verilerinin hazırlanması için birçok farklı üreticinin hazırladığı yazılımlar kullanılmaktadır. Yayın çalışmasına konu olan GSE8262 ve GSE10470 veri kümeleri Illumina firmasının BeadStudio yazılımı ile hazırlanmıştır. Veri kümelerinin incelenmesinde ise R [10] programlama dilinin Bioconductor aracı kullanılmıştır [11]. Veri kümeleri GEO'dan elde edildikten sonra ön işleme tabi tutularak normalizasyon uygulanması ile gen tahmini için çalışmaya uygun hale getirilmesi sağlanmıştır.

2. Kaynak Taraması

2.1 Veri Kümesi

Çalışmada Illumina BeadStudio veri kümeleri kullanılmıştır. Üzerinde çalışma yapılan ilk veri kümesi; GSE8262 erişim numarası ile GEO platformundan paylaşımına açılan, kronik hipoksi tedavisinden sonra gelişen fare beynindeki gen ekspresyon profilini içermektedir. Veri kümesi; geliştirmekte

olan fare beyninin kortikal ve hipokampal bölgelerinde CCH (Chronic Consistent Hypoxia – Kronik Tutarlı Hipoksi) ve CIH (Chronic Intermittent Hypoxia – Kronik Aralıklı Hipoksi)'nin etkisini incelemek için hazırlanmıştır. Veri kümelerinde 18 örnek bulunmaktadır: Kortikal tipi için 3, hipokampus tipi için 3 olmak üzere 6 kontrol grubu, CCH tedavisi uygulanan 6, CIH tedavisi uygulanan 6 olmak üzere 12 deney grubu bulunmaktadır. Veri kümesine ve detay bilgiye ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE8262 [12]'den erişim sağlanabilmektedir.

İkinci veri kümesi GSE10470 erişim numarası ile GEO platformundan paylaşımına açılan, oksidatif strese maruz kalan WT ve ATF4 KO embriyonik nöronlarının gen ekspresyon profilini içermektedir. Veri kümesinde 24 örnek bulunmaktadır: WT tipi için 8, ATF4 tipi için 8 olmak üzere 16 kontrol grubu (tedavi edilmemiş nöronlar), WT tipi için 4, ATF4 tipi için 4 olmak üzere 8 deney grubu (glutamat homolog homostein ile oksidatif strese maruz bırakılan nöronlar) bulunmaktadır. Veri kümesine ve detay bilgiye ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE19470 [13]'den erişim sağlanabilmektedir.

2.2 Kaynak İnceleme

Seçilen veri kümelerini kullanan çalışmalar incelemiş ve aşağıdaki çalışmalara dahil edildiği görülmüştür. Uygulanan yöntemler ile bu verileri farklı açılardan değerlendirerek tedavi ve koruma süreçlerine katkı sağlayacak çıktılar elde edilmeye çalışılmıştır.

Zhou ve arkadaşlarının yaptığı çalışmada [14], kronik hipoksinin (oksijen yetmezliği) fare beynindeki gen ifadelerini nasıl değiştirdiği araştırılmıştır. Özellikle sarcospan (sspn) geninin hipoksi altındaki glial hücre ölümündeki rolü incelenmiştir. Çalışma kapsamında; denek farelerinin seçimi, hipoksi tedavisi ve incelenecek dokuların hazırlanması, RNA (RiboNükleikAsid)'ların çıkarılarak mikrodizi analizlerinin yapılması adımları izlenmiştir. CIH etkisi ile 80, CCH etkisi 137 genin etkilendiğine dikkat çekilmiştir. Hipoksi ile ilişkili nörolojik bozuklukların tedavisine olumlu katkılar sağlayabilecek çıktılar olan bu çalışma, kronik hipoksi koşullarının kortikal ve hipokampus bölgelerinde gen ifadelerinde önemli değişikliklere sebep olduğunu göstermiştir. Bu deneysel çalışma sırasında hazırlanan veri, GSE8262 koduyla GEO [9] platformundan erişime açılmıştır.

GSE8262 veri kümesi ile GSE2385 ve GSE1357 veri kümelerinin kullanıldığı, Wang ve arkadaşlarının yaptığı çalışmada [15], aralıklı hipoksi ile tetiklenen nöroenflamasyon ve bilişsel gerilemede SENP1 (SUMP spesifik proteaz 1) proteininin etkisini incelemiştir. Mikroglia da meydana gelen değişiklikler hem laboratuvar ortamında (in vitro) hem de canlı organizma içinde (in vivo) incelenmiştir. Yaptıkları çalışmada SENP1 proteininin, NEMO proteininin SUMOlaşmasını (SUMO: 'Small Ubiquitin like MOdifiers' proteininin hücrelerdeki diğer proteinlere bağlanarak işlevlerini değiştirmesi) modüle ederek aralıklı hipoksiye bağlı nöroenflamasyonu düzenlediğini ve NF-kB yolunu aktive ettiğini ifade etmişlerdir. Bu bulguların mikroglia da SENP1'in modüle edilmesinin aralıklı hipoksinin neden olduğu bilişsel

bozuklukların tedavisinde potansiyel bir hedef olabileceğini gösterdiğini belirtmişlerdir.

Wang ve arkadaşlarının yaptığı çalışmada [16], SENP1'in OSAS (Obstructive Sleep Apnea Syndrome – Obstrüktif Uyku apnesi Sendromu) ile ilişkili CIH'in neden olduğu nöroiltihaplanma ve nöronal hasara etkisini araştırmışlardır. Çalışma için SENP1 aşırı eksprese edilmiş mikroglia ve SENP1 eksik fareler kullanılarak, CIH mikroglia ve fare modelleri oluşturulmuştur. Çalışmada GSE8262, GSE2385 ve GSE1357 veri kümeleri kullanılmıştır. Yapılan çalışma sonucunda CIH'in SENP1 ve TOM1 seviyelerini azalttığını, TOM1 proteininin SUMOlaşmasını artırdığını ve mikroglial göç, nöroiltihaplanma, A β 42 (nöronal beta-amiloid42) birikimi ve apoptozu teşvik ettiğini belirtmişlerdir. SENP1'in TOM1 üzerindeki etkisinin, CIH'in neden olduğu bilişsel işlev bozukluklarının tedavisinde yeni yaklaşımlar sunabileceğini ifade etmişlerdir.

Lange ve arkadaşlarının yaptığı çalışmada [17], ATF4 (Activating Transcription Factor 4 – Aktive Edici Transkripsiyon Faktörü 4) proteininin oksidatif stres altında nöronlarda nasıl bir transkripsiyon faktörü olduğu incelenmiştir. Birçok nörodejeneratif hastalığın temelinde yatan, vücutta serbest radikallerin artması sonucu meydana gelen oksidatif strese tepki olarak glutasyon gibi antioksidanların tükenmesiyle ATF4 oluştuğuna dikkat çekilen çalışmada, ATF4'ün nöronlarda hücre ölümünü tetikleyen genlerin ifadesini artırdığı öne sürülmüştür. Çalışma kapsamında denek farelerinin seçimi, fare beyinlerinden elde edilen nöron kültürlerinin (oksidatif stres koşullarını sağlamak için) glutasyon tükenmesi gibi çeşitli stres faktörlerine maruz bırakılması hem canlı organizmada hem de laboratuvar ortamında incelemelerin yapılması, RNA'ların çıkarılarak mikrodizi analizlerinin yapılması süreçleri uygulanmıştır. Çalışma çıktısı olarak, ATF4'ün sinir sistemi hücrelerinde oksidatif stres ile ilişkili hücre ölümü süreçlerini başlattığını ve oksidatif stresin rol oynadığı hastalıkların tedavisinde hedeflenebileceğini belirtmişlerdir. Bu deneysel çalışma sırasında hazırlanan veri, GSE10470 koduyla GEO platformundan erişime açılmıştır.

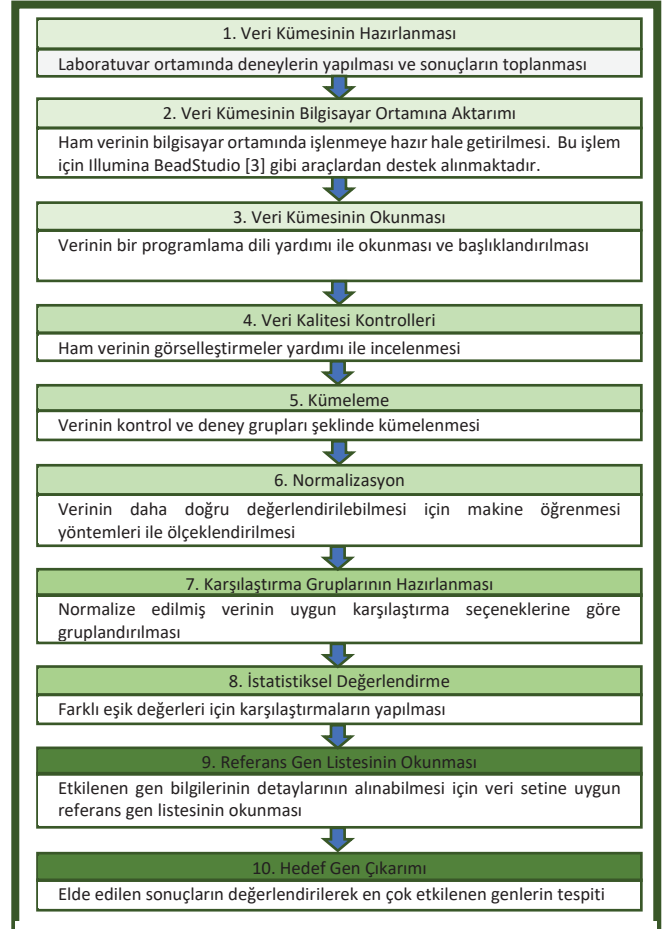
Woo ve arkadaşlarının yaptığı çalışmada [18], merkezi sinir sisteminde meydana gelen iltihaplanmada GRM8 (Glutamate Receptor, Metabotropic 8 – Metabotropik Glutamat Reseptörü 8)'in nöroprotektif rolü incelenmiştir. Bu çalışmada GSE10470 veri kümesi dışında birçok veri kümesi kullanılmıştır. MS, Alzheimer, Parkinson gibi nörodejeneratif hastalıklarda glutamat seviyesinin artması, nöronların aşırı uyarılmasına ve ölümüne neden olmaktadır. Çalışma, GRM8'in nörodejeneratif hastalıklarda glutamat seviyesini düzenleyerek nöronları koruduğunu öne sürmektedir. Çalışma çıktısı olarak elde ettikleri bulgular ile GRM8'in nörodejeneratif hastalıkların tedavisinde hedef olarak değerlendirilebileceğini belirtmişlerdir.

Bloomingdale ve arkadaşlarının yaptığı çalışmada [19], bortezomib isimli kemoterapi ilacının neden olduğu periferik nöropatiyi (el ve ayaklarda karıncalanma, uyuşma, ağrı gibi belirtiler) iyileştirmek için yeni bir strateji geliştirilmesi hedeflenmiştir. Kemoterapi kaynaklı nöropatik ağrının tedavi

edilmesi için ilaç geliştirme amaçlı çalışmada; GSE10470 veri kümesi ile bortezomib ile tetiklenen nöropatik ağrı yaşayan farelerde, sinir iletim hızları ve ağrı davranışları ölçülmüştür. Ayrıca laboratuvar ortamında, sinir hücreleri üzerinde bortezomib ve dexanabinol tedavisi sonrası çıktılar değerlendirilmiştir. Çalışmanın çıktısı olarak, dexanabinol'ün kemoterapiye bağlı nöropatik tedavide iyileştirici etkisi olabileceği belirtilmiştir.

3. Araçlar ve Yöntem

Mikrodizi verilerinin makine öğrenmesi yöntemleri ile analiz edilmesinde bu çalışmada da kullanılan Şekil-1'deki adımlar izlenebilmektedir. Şekil-1'deki 1. ve 2.adımlar [14] ve [17] yayınlarında detaylandırıldığı şekilde hazırlanmıştır ve çalışmada bu veri kümeleri kullanılmıştır. Diğer adımlar bu bölüm altındaki alt başlıklarda detaylandırıldığı şekilde uygulanmıştır. Veri kümesinin okunması süreci 3.1 başlığı altında açıklandığı şekilde GEO platformundan verinin elde edilmesi ve R programlama dili ile çalışmaya hazır hale getirilmesi şeklinde uygulanmıştır. Veri kalitesi kontrolleri için görselleştirme desteği de kullanılarak kontroller sağlanmıştır, çalışma detaylarına 3.2 başlığı altında değinilmiştir. Kümeleme çalışması 3.3 başlığı altında detaylandırıldığı gibi pearson korelasyona bağlı kümeleme kullanılarak yapılmıştır. Kümeleme işleminin ardından uygulanan niceliksel normalizasyon süreci 3.4 başlığı altında açıklanmıştır. Normalizasyon işlemi tamamlandıktan sonra 3.5 başlığında



Şekil-1: Mikrodizi verilerinin makine öğrenmesi yöntemleri ile analiz aşamaları

değindildiği gibi karşılaştırma grupları hazırlanmıştır. Hazırlanan karşılaştırma grupları üzerinde istatistiksel değerlendirmeler yapılmıştır. Yapılan değerlendirmeler 3.6 başlığı altında görselleştirme desteği ile detaylandırılmıştır. Değerlendirmelerin tamamlanması sonrasında sonuç verilerinin anlamlandırılabilmesi amacıyla veri kümelerine uygun referans gen listeleri okunmuştur (3.7 başlığı altında detay bilgiler bulunmaktadır). Son adım olarak hedef gen listeleri çıkarılarak 3.8 başlığı altında en çok etkilenen 20 gen listelenmiştir.

3.1 Veri Kümesinin Okunması

Veri kümesinin analizler için hazırlanması için kullanılacak yöntem belirlenirken veri kümesi dosyasının hangi platformda hazırlandığı, biçimi ve veri kümesinin açık erişim platformunda olup olmaması gibi bilgiler dikkate alınır. Çalışmaya konu [12] ve [13] veri kümeleri açık erişim platformunda olup, NCBI - GEO platformundan elde edilmiştir. Veri kümeleri üzerinde analizlerin yapılabilmesi için gerekli olan başlıklandırma işlemi R programlama dilinin Bioconductor aracı kullanılarak gerçekleştirilerek veri çerçevesi (dataframe) oluşturulmuştur.

Bu veri çerçevesinde;

- Sıra başlığı her bir deneyin sıra numarasını göstermektedir.
- Grup başlığı deneylerin gerçekleştirildiği grup bilgisidir. Veri kümeleri gruplar halinde alınarak işlenmektedir. Buradaki bilgi ilgili deneyin hangi grupta olduğu bilgisini gösterir.
- GenTipi başlığı deneylere konu gen tiplerini göstermektedir.
- Replika başlığı aynı koşullar altında tekrarlanan deneylerin numarasıdır. Birden çok replika kullanılarak verinin daha doğru değerlendirilmesi hedeflenmektedir.
- İlaç başlığı deney gruplarına uygulanan işlemi temsil eder. Bu işlem bir ilacın veya tedavinin uygulanması şeklinde olabilmektedir. Buradaki “ctrl” kontrol gruplarını ifade eder.
- Dizi kodu başlığı her bir deneyin ayırt edici kimlik bilgisidir. Her bir deney için benzersiz/tanımlayıcı bilgi niteliğindedir.

GSE8262 veri kümesi gelişen fare beyninin kortikal ve hipokampal bölgesinde CCH (kronik tutarlı hipoksi) ve CIH (kronik aralıklı hipoksi) tedavisi uygulanan deneye ait veri kümesi içermektedir. Veri kümesi 18 örnek içerip 145 kolon ve 46120 satırdan oluşmaktadır.

Bu veri kümesi için hazırlanan veri çerçevesi ise Çizelge-1’deki gibidir. Bu çerçeve her bir deney için veri çerçevesi bilgisini gösterir. Bu veri kümesindeki gen tipleri “Cortex” fare beynin kortikal bölgesine ait örnekleri, “HCampus” ise fare beyninin hipokampal bölgesine ait örnekleri ifade etmektedir. İlaç kısmındaki “ctrl” kontrol gruplarını, “cih” kronik aralıklı

hipoksi tedavisi uygulanan deneyleri, “cch” ise kronik tutarlı hipoksi tedavisi uygulanan örnekleri ifade etmektedir.

Çizelge-1: Veri Çerçevesi İçeriği – GSE8262

Sıra	Grup	GenTipi	Replika	İlaç	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
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

GSE10470 veri kümesi ise ATF4 ve WT genotiplerine HCA uygulanan deneye ait veri kümesini içermektedir. Veri kümesi 24 örnek içerip 193 kolon ve 24048 satırdan oluşmaktadır.

Bu veri kümesi için hazırlanan veri çerçevesi ise Çizelge-2’deki gibidir. Bu veri kümesindeki gen tipleri “wt (wild type)” normal farelere ait örnekleri, “ko (knockout)” ise ATF4 eksikliği olan farelere ait örnekleri ifade etmektedir. İlaç kısmındaki “ctrl” kontrol gruplarını, “hca” ise glutamat homolog homostein ile oksidatif strese maruz bırakılan örnekleri ifade etmektedir.

Çizelge-2: Veri Çerçevesi İçeriği – GSE10470

Sıra	Grup	GenTipi	Replika	İlaç	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
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

3.2 Veri Kalitesi Kontrolleri

Okunan veri kümesinin hazırlanan veri çerçevesi ile eşleştirilmesi sonrası veri gruplaması Çizelge-3 ve Çizelge-4'teki gibi olmuştur.

Çizelge-3: Veri Gruplaması – GSE8262

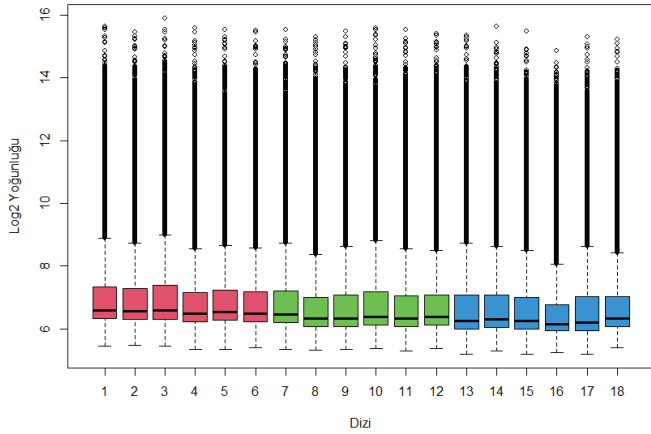
"Cortex.ctrl.1"	"Cortex.ctrl.7"	"Cortex.ctrl.13"
"Cortex.cih.2"	"Cortex.cih.8"	"Cortex.cih.14"
"Cortex.cch.3"	"Cortex.cch.9"	"Cortex.cch.15"
"HCampus.ctrl.4"	"HCampus.ctrl.10"	"HCampus.ctrl.16"
"HCampus.cih.5"	"HCampus.cih.11"	"HCampus.cih.17"
"HCampus.cch.6"	"HCampus.cch.12"	"HCampus.cch.18"

Çizelge-4: Veri Gruplaması – GSE10470

"wt.ctrl.1"	"ko.ctrl.7"	"ko.ctrl.13"	"wt.hca.19"
"wt.ctrl.2"	"ko.ctrl.8"	"wt.hca.14"	"ko.ctrl.20"
"wt.ctrl.3"	"ko.ctrl.9"	"ko.hca.15"	"wt.ctrl.21"
"wt.ctrl.4"	"wt.hca.10"	"wt.ctrl.16"	"ko.hca.22"
"ko.ctrl.5"	"ko.hca.11"	"wt.ctrl.17"	"wt.hca.23"
"ko.ctrl.6"	"wt.ctrl.12"	"ko.hca.18"	"ko.ctrl.24"

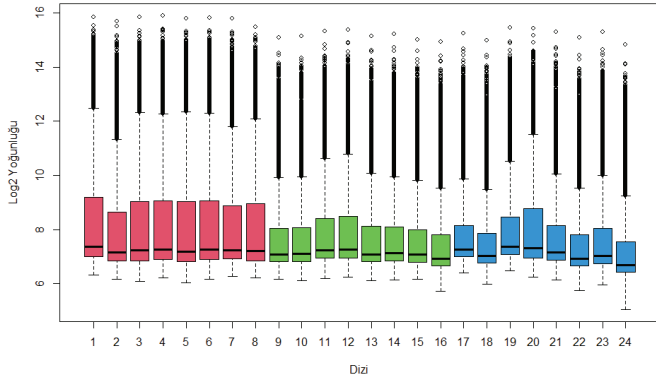
Sinyal dağılımı kontrol edildiğinde Şekil-2 ve Şekil-3'teki sonuçlara ulaşılmıştır.

Normalize Edilmemiş Verideki Sinyal Dağılımı



Şekil-2: Normalize Edilmemiş Verideki Sinyal Dağılımı – GSE8262

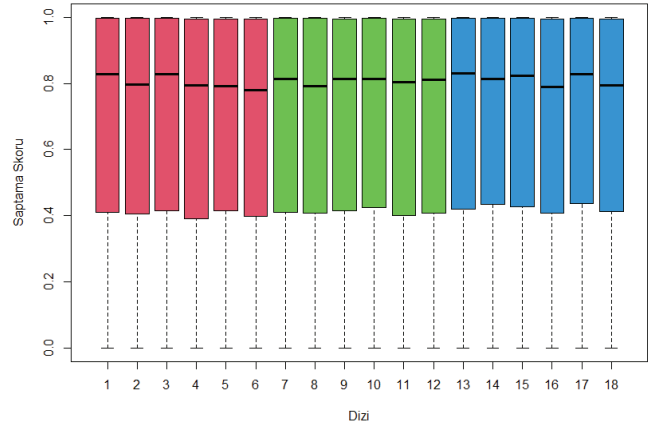
Normalize Edilmemiş Verideki Sinyal Dağılımı



Şekil-3: Normalize Edilmemiş Verideki Sinyal Dağılımı – GSE10470

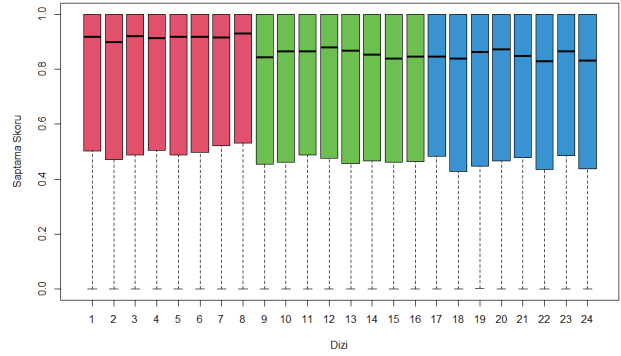
Veri kümelerinin saptama skorları ise Şekil-4 ve Şekil-5'teki gibidir.

Saptama Skorları



Şekil-4: Saptama Skorları – GSE8262

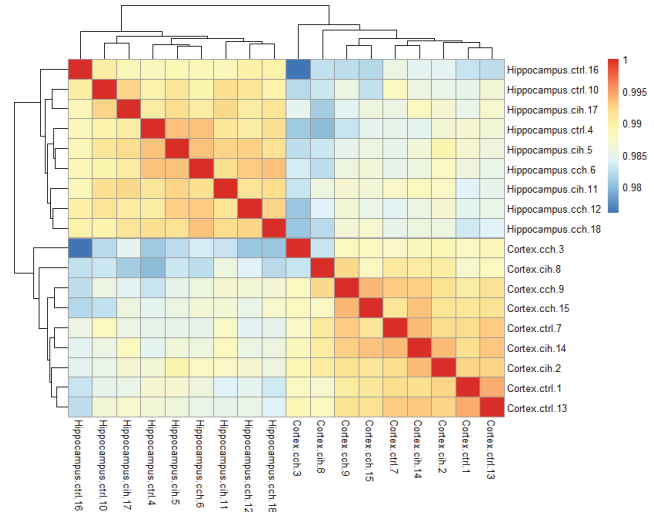
Saptama Skorları



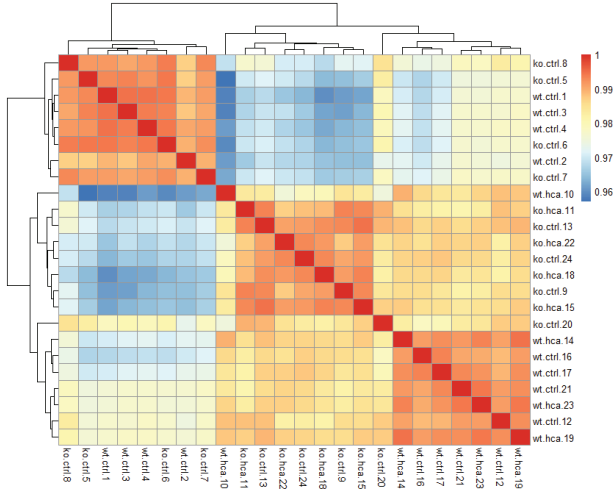
Şekil-5: Saptama Skorları – GSE10470

3.3 Kümeleme

Makine öğrenmesinde benzer veri gruplarını bir araya getirme işlemi olan kümeleme(clustering); verinin keşfi, anomali tespiti, özellik seçimi, segmentasyon gibi birçok süreçte kullanılmaktadır. Hiyerarşik kümeleme, yoğunluk tabanlı kümeleme (DBSCAN), K-Means kümeleme, spektral kümeleme, kavramsal kümeleme gibi birçok kümeleme tekniği mevcuttur. Bu çalışmada veri kümeleri üzerinde hiyerarşik kümeleme grubunda olan "pearson korelasyona bağlı kümeleme" kullanılmıştır.



Şekil-6: Pearson Korelasyona Bağlı Kümeleme – GSE8262



Şekil-7: Pearson Korelasyona Bağlı Kümeleme – GSE10470

3.4 Normalizasyon

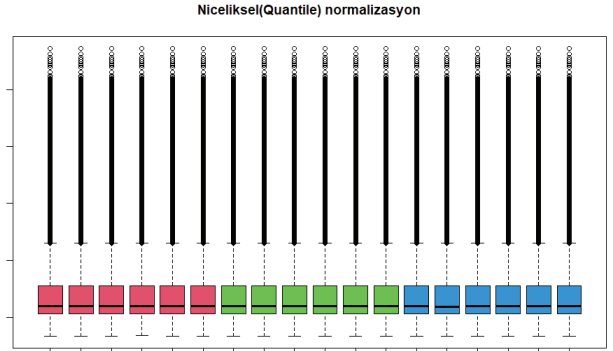
Normalizasyon veri kümesindeki özelliklerin belli bir aralığa veya ölçeğe dönüştürülmesi işlemidir. Veri kümesindeki büyük değer farklılıkları olan özelliklerde değerler 0-1 arasına ölçeklendirilerek daha doğru ve kolay değerlendirilebilir. Böylece farklı ölçeklerde olan verilerin karşılaştırılması mümkün olur. Min-Max normalizasyonu, Z-Score normalizasyonu, maksimum mutlak değer normalizasyonu (maks abs), niceliksel (quantile) normalizasyon gibi birçok farklı normalizasyon yöntemi mevcuttur.

Çalışmada kullanılan Bioconductor paketi altındaki Limma (linear models for microarray data) kütüphanesi gen ekspresyon analizlerinde tercih edilen bir araçtır. Bu kütüphane veri dağılımını standart hale getirme için çeşitli normalizasyon yöntemleri sağlamaktadır. Limma kütüphanesi niceliksel, ölçek, medyan, loess normalizasyon yöntemlerini sunmaktadır. Niceliksel normalizasyon örnekler arasındaki dağılımları eşitleyerek verileri standart hale getirirken, ölçek normalizasyon verilerin ortalama ve varyansını standartlaştırarak tutarlılığı artırır. Medyan normalizasyon her örneği medyan değerine göre ayarlayarak aşırı değer sorununu gidermeye yardımcı olurken, loess normalizasyon iki renkli mikrodizelerde yerel regresyon kullanarak dağılımı dengelemeye yardımcı olur.

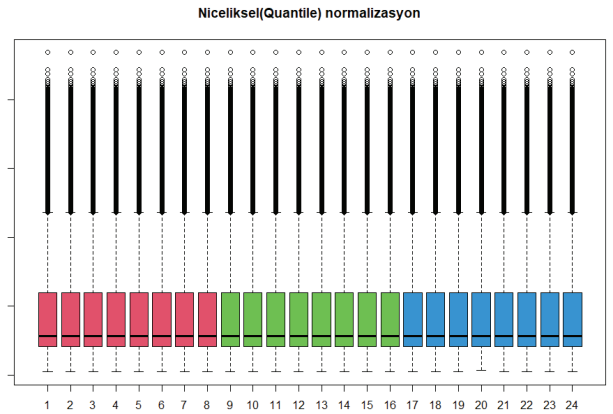
Niceliksel normalizasyon, yüksek boyutlu veri analizinde yaygın olarak kullanılan önemli bir normalizasyon tekniğidir [20]. Biyoformatikte gen ekspresyon verilerinin sıralanarak, tüm veri kümesinin aynı dağılımda olmasını sağlayan bir yöntem olarak sıkça kullanılmaktadır. Niceliksel normalizasyon, laboratuvar koşulları veya farklı örnek hazırlama teknikleri gibi sebeplerle mikrodizi deneylerinde oluşan teknik varyasyonları azaltarak verinin standart dağılımını sağlar. Bu yaklaşım, örnekler arasındaki teknik varyasyonları azaltarak değişimin daha doğru şekilde tespit edilmesini sağlar [21]. Ayrıca deneylerdeki gen ekspresyon dağılımını eşitleyerek, verilerin karşılaştırılabilirliğini artırır. Böylece farklı koşullar altında elde edilen verilerin daha anlamlı bir şekilde karşılaştırılması mümkün olur [22]. Niceliksel normalizasyon, biyolojik gürültüyü azaltarak daha tutarlı sonuçlar elde edilmesini sağlar. Biyolojik gürültünün

azaltılması özellikle düşük sinyalli genlerin analizinde homojen dağılım sağlayarak yanlış pozitif sonuçları azaltır [23]. Çalışmada biyolojik gürültü, teknik varyasyonu azaltması, veri karşılaştırma ve ölçeklenme sağlaması gibi avantajları sebebiyle niceliksel normalizasyon tercih edilmiştir.

Normalizasyon işlemi sonrasında sinyal dağılımları Şekil-8 ve Şekil-9'daki gibi değişiklik göstermiştir.

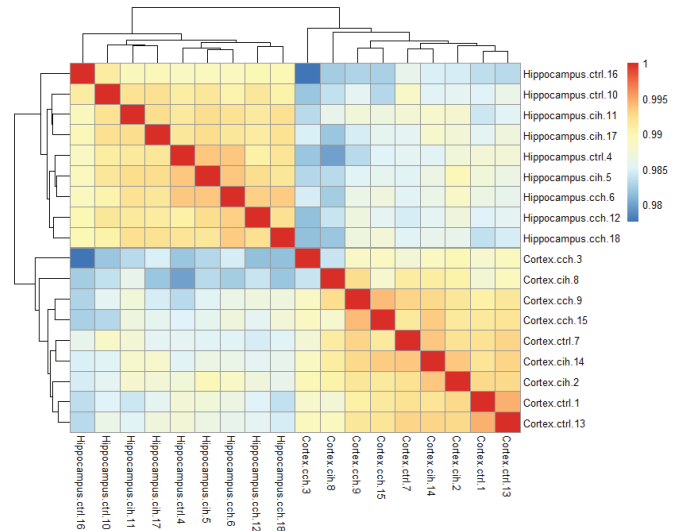


Şekil-8: Normalizasyon Sonrası Sinyal Dağılımı – GSE8262

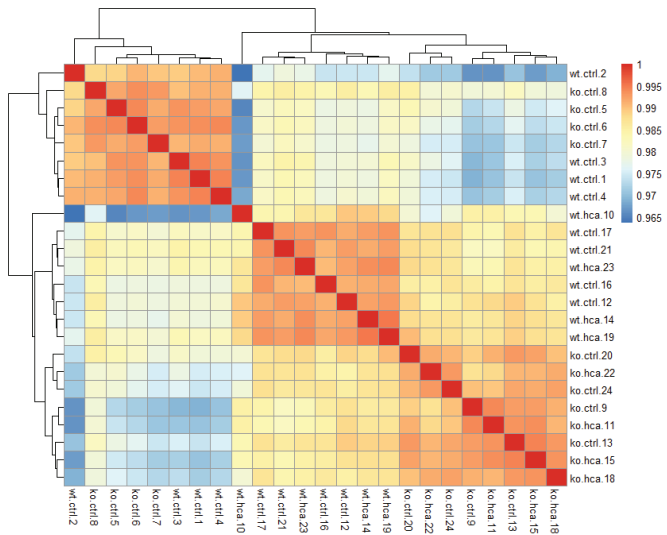


Şekil-9: Normalizasyon Sonrası Sinyal Dağılımı – GSE10470

Normalizasyon sonrası sıcaklık haritası değişimleri ise Şekil-10 ve Şekil-11'deki gibidir.



Şekil-10: Normalizasyon Sonrası Kümelmiş Veri Sıcaklık Haritası – GSE8262



Şekil-11: Normalizasyon Sonrası Kümelmiş Veri Sıcaklık Haritası – GSE10470

3.5 Karşılaştırma Gruplarının Hazırlanması

Normalizasyon işlemi sonrasında, gen ekspresyon ifadelerindeki değişikliklerin değerlendirilebilmesi için karşılaştırma grupları tanımlanarak bir matris hazırlanmıştır. Her karşılaştırma grubu iki veya daha fazla koşulun birbirinden çıkarılması ile elde edilmiş ve her gruba bir başlık tanımlanmıştır. Hazırlanan karşılaştırma matrisi farklı koşullar arasındaki farkların istatistiksel olarak test edilmesi için kullanılmıştır.

GSE8262 için karşılaştırma grupları aşağıdaki şekilde hazırlanmıştır.

Karşılaştırma Grupları Oluşturulması	
1.	Cortex.ctrl ile Hippocampus.ctrl karşılaştırma grubunun oluşturularak CortexwsHippocampus başlığı verilmesi
2.	Hippocampus.cih ile Hippocampus.ctrl karşılaştırma grubunun oluşturularak cihHippocampus başlığı verilmesi
3.	Cortex.cih ile Cortex.ctrl karşılaştırma grubunun oluşturularak cihCortex başlığı verilmesi
4.	(Hippocampus.cih- Hippocampus.ctrl) ile (Cortex.cih- Cortex.ctrl) karşılaştırma grubunun oluşturularak cihHippocampusnotcihCortex başlığı verilmesi
5.	(Cortex.cih- Cortex.ctrl) ile (Hippocampus.cih- Hippocampus.ctrl) karşılaştırma grubunun oluşturularak cihCortexnotcihHippocampus başlığı verilmesi
6.	Hippocampus.cch ile Hippocampus.ctrl karşılaştırma grubunun oluşturularak cchHippocampus başlığı verilmesi
7.	Cortex.cch ile Cortex.ctrl karşılaştırma grubunun oluşturularak cchCortex başlığı verilmesi
8.	(Hippocampus.cch- Hippocampus.ctrl) ile (Cortex.cch- Cortex.ctrl) karşılaştırma grubunun oluşturularak cchHippocampusnotcihCortex başlığı verilmesi
9.	(Cortex.cch- Cortex.ctrl) ile (Hippocampus.cch- Hippocampus.ctrl) karşılaştırma grubunun oluşturularak cchCortexnotcihHippocampus başlığı verilmesi

Şekil-12: Karşılaştırma Gruplarının Hazırlanması Süreci – GSE8262

Çizelge-5: Karşılaştırma Matrisi – GSE8262

	CortexwsHippocampus	cihHippocampus
Cortex.ctrl	1	0
Cortex.cih	0	0
Cortex.cch	0	0
Hippocampus.ctrl	-1	-1
Hippocampus.cih	0	1
Hippocampus.cch	0	0
	cihCortex	cihHippocampusnotcihCortex
Cortex.ctrl	-1	1
Cortex.cih	1	-1
Cortex.cch	0	0
Hippocampus.ctrl	0	-1
Hippocampus.cih	0	1
Hippocampus.cch	0	0

	cihCortexnotcihHippocampus	cchHippocampus
Cortex.ctrl	-1	0
Cortex.cih	1	0
Cortex.cch	0	0
Hippocampus.ctrl	1	-1
Hippocampus.cih	-1	0
Hippocampus.cch	0	1
	cchCortex	cchHippocampusnotcihCortex
Cortex.ctrl	-1	1
Cortex.cih	0	0
Cortex.cch	1	-1
Hippocampus.ctrl	0	-1
Hippocampus.cih	0	0
Hippocampus.cch	0	1
	cchCortexnotcihHippocampus	
Cortex.ctrl	-1	
Cortex.cih	0	
Cortex.cch	1	
Hippocampus.ctrl	1	
Hippocampus.cih	0	
Hippocampus.cch	-1	

GSE10470 için karşılaştırma grupları aşağıdaki şekilde hazırlanmıştır.

Karşılaştırma Grupları Oluşturulması	
1.	ko.ctrl ile wt.ctrl karşılaştırma grubunun oluşturularak KOwsWT başlığı verilmesi
2.	wt.hca ile wt.ctrl karşılaştırma grubunun oluşturularak HCAwt başlığı verilmesi
3.	ko.hca ile ko.ctrl karşılaştırma grubunun oluşturularak HCAko başlığı verilmesi
4.	(wt.hca- wt.ctrl) ile (ko.hca- ko.ctrl) karşılaştırma grubunun oluşturularak HCAwtnotHCAko başlığı verilmesi

Şekil-13: Karşılaştırma Gruplarının Hazırlanması Süreci – GSE10470

Çizelge-6: Karşılaştırma Matrisi – GSE10470

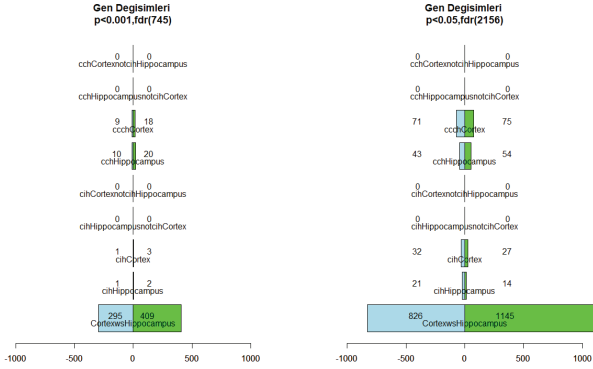
	KOwsWT	HCAwt	HCAko	HCAwtnotHCAko
wt.ctrl	-1	-1	0	-1
ko.ctrl	1	0	-1	1
wt.hca	0	1	0	1
ko.hca	0	0	1	-1

3.6 İstatistiksel Değerlendirme

Doğrusal modelleme yapılarak karşılaştırma matrisleri oluşturulduktan sonra istatistiksel değerlendirmeler yapılmıştır. Çalışmamızda mikrodizi verilerinin değerlendirilmesinde sıklıkla tercih edilen eBayes (empirical bayes – amprik bayes) yöntemi kullanılmıştır [24].

Çalışmada kullanılan FDR (False Discovery Rate – Yanlış Keşif Oranı) yöntemi, mikrodizi analizlerinde sıklıkla kullanılan istatistiksel yöntemlerdendir. Temel amacı, yapılan çoklu testlerde yanlış pozitif sonuçların oranını kontrol etmektir [25]. Bir test sonucu istatistiksel olarak anlamlı olduğunda bu sonuçtaki hata olasılığını azaltmak için kullanılır. FDR istatistikğinde sıklıkla kullanılan p-değeri (p-value) ise elde edilen verinin gözlemlenen sonuçtan daha uç bir sonuç verme olasılığını ifade eder. P-değerinin düşük olması sonucun anlamlı olduğunu gösterir. Kullanılan verinin içeriğine bağlı olarak farklı eşik değerleri anlamlı kabul edilebilmekle birlikte genel olarak p-değeri 0,05 altında olan sonuçlar anlamlı/doğru kabul edilir [26].

p-değeri<0,05 için çok fazla gen elde edilmesi durumunda sonuçları filtrelemek/azaltmak için p-değerinin düşürülerek gen sayılarının alınması yoluna gidilebilir. Çalışmanın değerlendirme kısmında farklı eşik değerleri için (0,05 ve 0,0001) diferansiyel olarak ifade edilen genlerin sayıları çıkarılmıştır. FDR analizlerinde p<0,001 ve p<0,05 eşik değerleri için etkilenen gen sayıları Şekil-14, Şekil-15, Şekil-16, Şekil-17, Şekil-18, Şekil-19'da gösterilmiştir. Yeşil çubuklar yukarı regüle (up-regulated) gen sayısını, mavi çubuklar ise aşağı regüle (down-regulated) gen sayısını göstermektedir.



Şekil-14: Farklı eşik değerleri için gruplar arası etkilenen gen sayıları – GSE8262

GSE8262 için p<0,05 genler alındığında toplam 2156 adet genin etkilendiği görülmüştür (cchCortex: 71 gen aşağı - 75 gen yukarı regüle, cchHippocampus: 43 gen aşağı - 54 gen yukarı regüle, cihCortex: 32 gen aşağı - 27 gen yukarı regüle, cihHippocampus: 21 gen aşağı - 14 gen yukarı regüle olmuştur).

	CortexwHippocampus	cihHippocampus	cihCortex
Down	826	21	32
NotSig	44149	46085	46061
Up	1145	14	27

	cihHippocampusnotcihCortex	cihCortexnotcihHippocampus	cchHippocampus
Down	0	0	43
NotSig	46120	46120	46023
Up	0	0	54

	ccchCortex	cchHippocampusnotcihCortex	cchCortexnotcihHippocampus
Down	71	0	0
NotSig	45974	46120	46120
Up	75	0	0

Şekil-15: p<0,05 için gruplar arası etkilenen gen sayıları – GSE8262

Hassasiyeti artırmak için p<0,001 genler alındığında ise toplam 745 genin etkilendiği görülmüştür (cchCortex: 9 gen aşağı – 18 gen yukarı regüle, cchHippocampus: 10 gen aşağı - 20 gen yukarı regüle, cihCortex: 1 gen aşağı - 3 gen yukarı regüle, cihHippocampus: 1 gen aşağı - 2 gen yukarı regüle olmuştur).

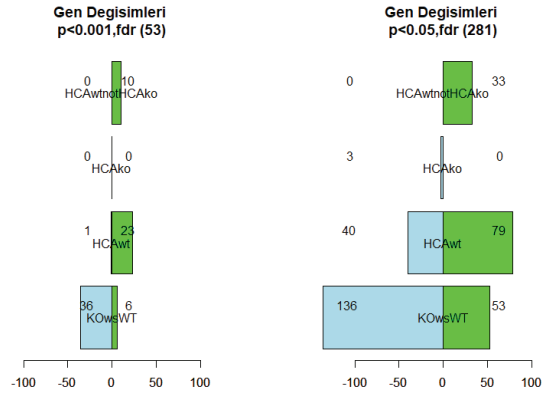
	CortexwHippocampus	cihHippocampus	cihCortex
Down	295	1	1
NotSig	45416	46117	46116
Up	409	2	3

	cihHippocampusnotcihCortex	cihCortexnotcihHippocampus	cchHippocampus
Down	0	0	10
NotSig	46120	46120	46090
Up	0	0	20

	ccchCortex	cchHippocampusnotcihCortex	cchCortexnotcihHippocampus
Down	9	0	0
NotSig	46093	46120	46120
Up	18	0	0

Şekil-16: p<0,001 için gruplar arası etkilenen gen sayıları – GSE8262

GSE10470 için yapılan istatistiksel çalışma sonuçları aşağıdaki şekilde olmuştur.



Şekil-17: Farklı eşik değerleri için gruplar arası etkilenen gen sayıları – GSE10470

GSE10470 için p<0,05 genler alındığında toplam 281 adet genin etkilendiği görülmüştür (HCAwtnotHCAko: 33 gen yukarı regüle, HCAko: 3 gen aşağı regüle, HCAwt: 40 gen aşağı - 79 gen yukarı regüle olmuştur).

	KOWsWT	HCAwt	HCAko	HCAwtnotHCAko
Down	136	40	3	0
NotSig	23859	23929	24045	24015
Up	53	79	0	33

Şekil-18: p<0,05 için gruplar arası etkilenen gen sayıları – GSE10470

Hassasiyeti artırmak için p<0,001 genler alındığında ise toplam 53 genin etkilendiği görülmüştür (HCAwtnotHCAko: 10 gen yukarı regüle, HCAwt: 1 gen aşağı - 23 gen yukarı regüle, KOWsWT: 36 gen aşağı - 6 gen yukarı regüle olmuştur).

	KOWsWT	HCAwt	HCAko	HCAwtnotHCAko
Down	36	1	0	0
NotSig	24006	24024	24048	24038
Up	6	23	0	10

Şekil-19: p<0,001 için gruplar arası etkilenen gen sayıları – GSE10470

3.7 Referans Gen Listesi Okunması

İstatistiksel değerlendirmelerle elde edilen sonuçların gen bilgilerinin açıklanmaya ihtiyacı vardır. Yapılan çalışma ile modeldeki her bir karşılaştırma grubu için p-değeri ve gen kod listeleri oluşturmuştur. Çalışma sonucunda ulaşılan gen kodlarının hangi genlere ait olduğu ve bu genlerin özellikleri veri tabanlarından manuel olarak araştırılabilir. Ancak gen bilgilerinin veri tabanlarından manuel olarak araştırılması zaman alan bir süreçtir. Bu süreci kolaylaştırmak amacıyla programlama yöntemleri kullanılarak gen açıklamalarının elde edilmesi mümkündür. Bu bağlamda elde edilen gen kodları uygun referans gen veri tabanları ile karşılaştırılarak analiz edilmiştir. Ulaşılabilecek referans gen listesi bilgileri Illumina BeadStudio ham veri dosyasından bulunabilmektedir. GSE8262 için Mouse-6_V1, GSE10470 için ise Mouse_Ref-8-V1 gen referans listeleri kullanılmıştır. Bu listelerde gen sembolü, gen açıklaması, gen numarası, prob bilgisi gibi bilgiler bulunmaktadır. Son aşama olarak elde edilen gen kodları ile referans gen listesi birleştirilerek gen açıklamaları elde edilmiştir.


```

1 Illumina Inc. BeadStudio version 1.5.1.3
2 Normalization = none
3 Array Content = Mouse-6_V1.xml
4 Error Model = none
5 Date/Time = 12/15/2005 3:15 PM
6 Local Settings = en-US
7
8 TargetID MIN_Signal-1412023004_A AVG_Signal-1412023004_A MAX_Signal-1412023004_A
9 10181072_239_rc-S 78.4 78.4 78.4 1 NaN 5.1 42 0.3539604 9
10 10181072_290-S 94 94 94 1 NaN 4 48 0.79022277 86.1 86.1 8
11 10181072_290_rc-S 99.1 99.1 99.1 1 NaN 6.7 37 0.84777228 9
12 10181072_311-S 71.9 71.9 71.9 1 NaN 2.8 48 0.1615099 73.3

```

Şekil-20: Referans Gen Listesi Bilgisinin Tespiti – GSE8262

```

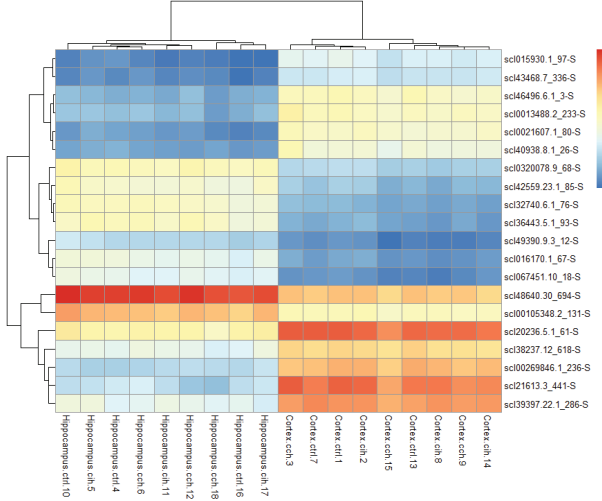
1 Illumina Inc. BeadStudio version 1.5.1.3
2 Normalization = none
3 Array Content = Mouse_Ref-8_V1.xml
4 Error Model = none
5 Date/Time = 3/3/2006 12:51 PM
6 Local Settings = en-US
7
8 TargetID MIN_Signal-1412066061_A AVG_Signal-1412066061_A MAX_Signal-1412066061_A
9 182_rRNA_X00696_529-S 6771.3 6771.3 6771.3 1 NaN 180.7 43 1 8179.4
10 GI_11079652-S 128.7 128.7 128.7 1 NaN 7.3 35 0.50542823 141.6 1
11 GI_21070949-S 17559.8 17559.8 17559.8 1 NaN 501.7 32 1 14079.4 14079.9
12 GI_21746160-S 14426.5 14426.5 14426.5 1 NaN 587.6 35 1 14499.9 14499.9

```

Şekil-21: Referans Gen Listesi Bilgisinin Tespiti – GSE10470

3.8 Hedef Gen Çıkarımı

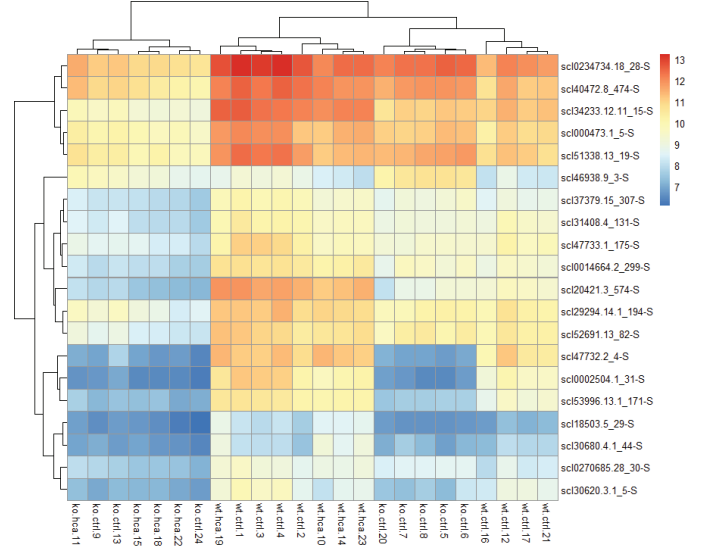
Referans gen listesinin okunması ile ID'lere karşılık gen bilgilerinin alınabilmesi mümkün olmaktadır. Elde edilen diferansiyel ekspresyon sonuçları istatistiksel öneme göre sıralanmış ve bir ifade matrisi hazırlanarak en çok etkilenen 20 gen bilgisi seçilmiştir.



Şekil-22: Belirlenen Kriterlere Göre En Çok Etkilenen 20 Genin Matris Halinde Gösterimi – GSE8262

Transkript ID	Symbol	Definition
sci015930.1_97-S	Indo	Mus musculus indoleamine-pyrrole 2
sci43468.7_336-S	Ist1	Mus musculus ISL1 transcription factor
sci46496.6.1_3-S	Tnnc1	Mus musculus troponin C
sci0013488.2_233-S	Drd1a	Mus musculus dopamine receptor D1A (Drd1a)
sci0021607.1_80-S	Tcrb-V8.2	
sci40938.8.1_26-S	Ppp1r1b	Mus musculus protein phosphatase 1
sci0320078.9_68-S	Olfml2b	Mus musculus olfactomedin-like 2B (Olfml2b)
sci42559.23.1_85-S	Slc26a4	Mus musculus solute carrier family 26
sci32740.6.1_76-S	Klk8	Mus musculus kallikrein related-peptidase 8 (Klk8)
sci36443.5.1_93-S	Spink8	Mus musculus serine peptidase inhibitor
sci49390.9.3_12-S	Pkp2	Mus musculus plakophilin 2 (Pkp2)
sci016170.1_67-S	Il16	Mus musculus interleukin 16 (Il16)
sci067451.10_18-S	Pkp2	Mus musculus plakophilin 2 (Pkp2), mRNA.
sci48640.30_694-S	Dgkg	
sci00105348.2_131-S	Golph2	Mus musculus golgi phosphoprotein 2 (Golp2)
sci20236.5.1_61-S	6330527006Rik	
sci38237.12_618-S	A130090K04Rik	
sci00269846.1_236-S	LOC665506	PREDICTED: Mus musculus similar to T-cell receptor beta 2 chain C region (LOC665506)
sci21613.3_441-S	Gpr88	Mus musculus G-protein coupled receptor 88 (Gpr88)
sci39397.22.1_286-S	Rgs9	Mus musculus regulator of G-protein signaling 9 (Rgs9)

Şekil-23: En Çok Etkilenen Genler Sembol ve Tanımları – GSE8262. Mouse-6.V1 referans gen listesindeki bilgilerden oluşturulmuştur.



Şekil-24: Belirlenen Kriterlere Göre En Çok Etkilenen 20 Genin Matris Halinde Gösterimi – GSE10470

Transkript ID	Symbol	Definition
sci0234734.18_28-S	Aars	Mus musculus alanyl-tRNA synthetase (Aars) mRNA.
sci40472.8_474-S	Slc1a4	Mus musculus solute carrier family 1 (glutamate/neutral amino acid transporter) member 4 (Slc1a4) mRNA.
sci34233.12.11_15-S	Slc7a5	Mus musculus solute carrier family 7 (cationic amino acid transporter y+ system) member 5 (Slc7a5) mRNA.
sci000473.1_5-S	Slc3a2	Mus musculus solute carrier family 3 (activators of dibasic and neutral amino acid transport) member 2 (Slc3a2) mRNA.
sci51338.13_19-S	Nars	Mus musculus asparaginyl-tRNA synthetase (Nars) mRNA.
sci46938.9_3-S	Slc25a17	Mus musculus solute carrier family 25 (mitochondrial carrier; peroxisomal membrane protein) member 17 (Slc25a17) mRNA.
sci37379.15_307-S	Shmt2	Mus musculus serine hydroxymethyltransferase 2 (mitochondrial) (Shmt2) mRNA.
sci31408.4_131-S	Atf5	Mus musculus activating transcription factor 5 (Atf5) mRNA.
sci47732.2_4-S	Mgat3	Mus musculus mannoside acetylglucosaminyltransferase 3 (Mgat3) mRNA.
sci0014664.2_299-S	Slc6a9	Mus musculus solute carrier family 6 (neurotransmitter transporter glycine) member 9 (Slc6a9) mRNA.
sci20421.3_574-S	1810008K03Rik	Mus musculus RIKEN cDNA 1810008K03 gene (1810008K03Rik) mRNA.
sci29294.14.1_194-S	Asns	Mus musculus asparagine synthetase (Asns) mRNA.
sci52691.13_82-S	Psat1	Mus musculus phosphoserine aminotransferase 1 (Psat1) mRNA.
sci47732.2_4-S	Atf4	Mus musculus activating transcription factor 4 (Atf4) mRNA.
sci0002504.1_31-S	Atf4	Mus musculus activating transcription factor 4 (Atf4) mRNA.
sci53996.13.1_171-S	Slc7a3	Mus musculus solute carrier family 7 (cationic amino acid transporter y+ system) member 3 (Slc7a3) mRNA.
sci18503.5_29-S	Trib3	
sci30680.4.1_44-S	Nupr1	Mus musculus nuclear protein 1 (Nupr1) mRNA.
sci0270685.28_30-S	Fthfscd1	Mus musculus formyltetrahydrofolate synthetase domain containing 1 (Fthfscd1) mRNA.
sci30620.3.1_5-S	Cox6a2	Mus musculus cytochrome c oxidase subunit VI a polypeptide 2 (Cox6a2) mRNA.

Şekil-25: En Çok Etkilenen Genler Sembol ve Tanımları – GSE10470. Mouse_Ref-8_V1 referans gen listesindeki bilgilerden oluşturulmuştur.

4. Sonuç ve Tartışma

Bu çalışmada mikrodizi veri kümelerinde gen ekspresyon analizi çalışması yapılmıştır. Illumina BeadStudio ile hazırlanmış olan, kronik hipoksi tedavisi ve oksidatif stresin fareler üzerindeki etkisini inceleyen iki farklı veri kümesi kullanılmıştır. "Mikrodizi verilerinin makine öğrenmesi yöntemleri ile analizi" işlem hattı her iki veri kümesinde de başarıyla uygulanarak sonuçlar elde edilmiştir. Bu işlem hattı ile verinin okunmasından gen listelerinin elde edilmesine kadar tüm süreç adım adım incelenmiştir. Doğru yöntem ve değerlendirme metodolojileri belirlenerek uygulanmıştır. Görselleştirmeler kullanılarak sürecin anlaşılabilirliği artırılmıştır. Çalışmamız mikrodizi analizinde kullanılabilecek yöntemler ve izlenmesi gereken adımları göstermesi açısından benzer çalışmalara katkı sağlayabilir.

Mikrodizi teknolojisi ile gen ekspresyon analizi yaygın bir yöntem olup, kanser gibi hastalıkların tedavi süreçlerinin iyileştirilmesi, hastalığın ilerlemesinin yavaşlatılması/durdurulması, hastalığın önlenmesi, etkin ilaç geliştirilmesi veya ilaçlara bağlı yan etkilerin azaltılması/giderilmesi gibi birçok süreçte fayda

sağlayabilmektedir. Schena ve arkadaşlarının [27], Lockhart ve Winzelerin [28], Alizadeh ve arkadaşlarının [29], Irizarry ve arkadaşlarının [30], Golub ve arkadaşlarının [31], Van 't Veer ve arkadaşlarının [32] yaptığı çalışmalar gibi birçok çalışmada mikrodizi gen ekspresyon analizinin bu yöndeki etkilerine değinilmiştir. Çalışmamızda elde ettiğimiz sonuçların da tanı, tedavi ve ilaç geliştirme süreçlerine olumlu katkı sağlayabileceği düşünülmektedir.

Kaynakça

- [1] Yıldız Çavdar Z., Ensari T., Türker Şener L., Sertbaş A. Makine Öğrenmesi Tabanlı Mikrodizi Tekniği ile MikroRNA Hedef Tahmini: Araştırma Çalışması, Avrupa Bilim ve Teknoloji Dergisi, 2022, pp. 39-45.
- [2] Cordero F., Botta M., Calogero R.A. Microarray data analysis and mining approaches, Briefings in Functional Genomics and Proteomics, 2008, pp. 265-281.
- [3] Illumina BeadStudio, <https://www.illumina.com> (Erişim Tarihi: 22.05.2024).
- [4] Gunderson K.L., Kruglyak S., Graige M.S., Garcia F., Kermani B.G., Zhao C., Che D., Dickinson T., Wickham E., Bierle J., Doucet D., Milewski M., Yang R., Siegmund C., Haas J., Zhou L., Oliphant A., Fan J.B., Barnard S., Chee M.S. Decoding Randomly Ordered DNA Arrays, Genome Research, 2004, pp. 870-877.
- [5] Wang Y., Mooradian A., Cheng Y. High-throughput SNP genotyping using the BeadArray technology, microarray data analysis and mining approaches, Genomics, 2008, pp.99-110.
- [6] Bibikova M., Fan J.B. Genome-wide DNA methylation profiling using Infinium assay, Epigenomics, 2006, pp.177-200.
- [7] Steemers F.J., Chang W., Lee G. Whole-genome genotyping with the single-base extension assay, Nature Methods, pp.851-856.
- [8] Gunderson K.L., Steemers F.J., Lee G. A genome-wide scalable SNP genotyping assay using microarray technology, Genomics, 2004, pp.549-554.
- [9] National Center of Biotechnology Information, <https://ncb.nlm.nih.gov> (Erişim Tarihi: 22.05.2024).
- [10] R Programlama Dili, <https://www.r-project.org/> (Erişim Tarihi: 22.05.2024).
- [11] Bioconductor, <https://www.bioconductor.org/> (Erişim Tarihi: 22.05.2024).
- [12] GEO GSE8262 Veri Kümesi, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE8262> (Erişim Tarihi: 22.05.2024).
- [13] GEO GSE10470 Veri Kümesi, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE10470> (Erişim Tarihi: 22.05.2024).
- [14] Zhou D., Wang J., Zapala M.A., Xue J., Schork N.J., Haddad G.G. Gene expression in mouse brain following chronic hypoxia: role of sarcospan in glial cell death, Physiol Genomics, 2008, pp. 370-379.
- [15] Wang H., Yang T., Sun J., Zhang S., Liu S. SENP1 modulates microglia-mediated neuroinflammation toward intermittent hypoxia-induced cognitive decline through the de-SUMOylation of NEMO, Journal of Cellular and Molecular Medicine, 2021, pp. 6463-7099.
- [16] Wang H., Wang X., Shen Y., Wang Y., Yang T., Sun J., Liu S. SENP1 modulates chronic intermittent hypoxia-induced inflammation of microglia and neuronal injury by inhibiting TOM1 pathway, International Immunopharmacology, 2023, pp. 1-13.
- [17] Lange P.S., Chavez J.C., Pinto J.T., Coppola G., Sun C.W., Townes T.M., Geschwind D.H., Ratan R.R. ATF4 is an oxidative stress-inducible, prodeath transcription factor in neurons in vitro and in vivo, The Journal of Experimental Medicine, 2008, pp. 1227-1242.
- [18] Woo M.S., Ufer F., Rothhammer N., Liberto G.D., 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., Friese M.A. Neuronal metabotropic glutamate receptor 8 protects against neurodegeneration in CNS inflammation, The Journal of Experimental Medicine, 2021, pp. 1-30.
- [19] 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., Mager D.E. Systems Pharmacology Modeling Identifies a Novel Treatment Strategy for Bortezomib-Induced Neuropathic Pain, Frontiers in Pharmacology, 2022, pp. 1-16.
- [20] Zhao Y., Wong L., Goh W.W.B. How to do quantile normalization correctly for gene expression data analyses. Scientific Reports, 2020, pp. 1-11.
- [21] Bolstad B.M., Irizarry R.A., Astrand M., Speed T.P. A comparison of normalization methods for high-density oligonucleotide array data based on variance and bias. Bioinformatics, 2003, pp. 185-193.
- [22] Irizarry R.A., Hobbs B., Collin F., Beazer-Barclay Y.D., Antonellis K.J., Scherf U., Speed T.P. Exploration, normalization, and summaries of high-density oligonucleotide array probe level data. Biostatistics, 2003, pp. 249-264.
- [23] Smyth G.K., Speed T., W.W.B. Normalization of cDNA microarray data. Methods, 2003, pp. 265-273.
- [24] Phipson B., Lee S., Majewski I.J., Alexander W.S., Smyth G.K. Robust Hyperparameter Estimation Protects Against Hypervariable Genes and Improves Power to Detect Differential Expression. The Annals of Applied Statistics, 2016, pp. 946-963.
- [25] Aubert J., Bar-Hen A., Daudin J.J., Robin S. Determination of the differentially expressed genes in microarray experiments using local FDR. BMC Bioinformatics, 2004.
- [26] Pawitan Y., Michiels S., Koscielny S., Gusnanto A., Ploner A. False discovery rate, sensitivity and sample size for microarray studies. Bioinformatics, 2005, pp. 3017-3024.
- [27] Schena M., Shalon D., Davis R.W., Brown P.O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science, 1995, pp. 467-470.
- [28] Lockhart D.J., Winzeler E.A. Genomics, gene expression and DNA arrays. Nature, 2000, pp. 827-836.
- [29] Alizaden A.A., Eisen M.B., Davis R.E., Ma C., Lossos I.S., Rosenwald A., Staudt L.M. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature, 2000, pp. 503-511.
- [30] Irizarry R.A., Hobbs B., Collin F., Beazer Barclay Y.D., Antonellis K.J., Scherf U., Speed T.P. Exploration, normalization, and summaries of high-density oligonucleotide array probe level data. Biostatistics, 2003, pp. 249-264.
- [31] Golub T.R., Slonim D.K., Tamayo P., Huard C., Gaasenbeek M., Mesirov J.P., Lander E.S. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science, 1999, pp. 531-537.
- [32] Van't Veer L.J., Dai H., Van de Vijver M.J., He Y.D., Hart A.A., Mao M., Friend S.H. Gene expression profiling predicts clinical outcome of breast cancer. Nature, 2002, pp. 530-536.