



**T. C.
ULUDAĞ ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ
BİYOİSTATİSTİK ANABİLİM DALI**

MEME KANSERİNDE GEN EKSPRESYON VERİLERİNİN MİKRODİZİ ANALİZİ

Sema ATIŞ

(YÜKSEK LİSANS TEZİ)

Bursa-2014



T. C.
ULUDAĞ ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ
BİYOİSTATİSTİK ANABİLİM DALI

MEME KANSERİNDE GEN EKSPRESYON VERİLERİNİN MİKRODİZİ ANALİZİ

Sema ATIŞ

(YÜKSEK LİSANS TEZİ)

Danışman: Prof. Dr. İLKER ERCAN

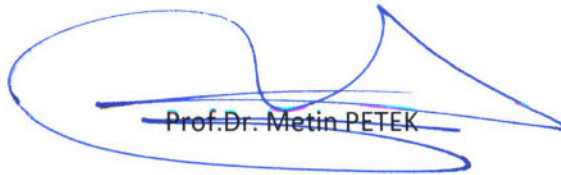
Bursa-2014

SAĞLIK BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜ'NE

Biyostatistik Anabilim Dalı Yüksek Lisans öğrencisi Sema Atış tarafından hazırlanan Meme Kanserinde Gen Ekspresyon Verilerinin Mikrodizi Analizi konulu Yüksek Lisans tezi 06/01/2015 günü, 09:30-11:00 saatleri arasında yapılan tez savunma sınavında jüri tarafından oybirliği/oyçokluğu ile kabul edilmiştir.

	<u>Adı-Soyadı</u>	<u>İmza</u>
Tez Danışmanı	Prof. Dr. İlker ERCAN	
Üye	Prof. Dr. Sezai Türkel	
Üye	Doç. Dr. Güven Özkaya	
Üye	Yrd. Doç. Dr. Bülent Ediz	
Üye	Yrd. Doç. Dr. Elif Uz	

Bu tez Enstitü Yönetim Kurulu'nun 07.01.2015 tarih ve 2015/01 sayılı toplantısında alınan 13 numaralı kararı ile kabul edilmiştir.++


Prof. Dr. Metin PETEK

Enstitü Müdürü

İÇİNDEKİLER

TÜRKÇE ÖZET	III
İNGİLİZCE ÖZET	IV
1. GİRİŞ	1
2. GENEL BİLGİLER	2
2.1. Mikroarray Yapımı ve Kullanımı	2
2.1.1. Spotlanmış Mikroarray	4
2.1.2. In-situ Sentezlenmiş Oligonükleotit Arrayler	6
2.1.2.1. Affymetrix Teknolojisi	7
2.1.3. Gen Ekspresyon Mikroarrayleri	8
2.1.3.1. Örneklem Hazırlanması ve Etiketleme	9
2.1.3.2. Hibridizasyon ve Yıkama	10
2.1.3.3. Tarama ve Görüntü Elde Etme	10
2.2. Mikroarray Verilerinin Ön İşlemesi	11
2.2.1. Görüntü Analizi	12
2.2.2. Arka Plan Düzeltmesi	12
2.2.2.1. RMA Arka Plan Düzeltmesi	14
2.2.2.2. MAS 5.0 Arka Plan Düzeltmesi	14
2.2.3. Kalite Değerlendirmesi	15
2.2.4. Normalizasyon	20
2.2.4.1. Scaling Normalizasyon Yöntemi	21
2.2.4.2. Lineer Olmayan Normalizasyon Yöntemleri	21
2.2.4.3. Çeyreklik Normalizasyonu	22
2.2.4.4. Cyclic Loess Normalizasyonu	22

2.2.4.5. Kontrast Normalizasyonu	22
2.3. Farklı Gen Ekspresyon Analizinde Kullanılan İstatistiksel Yöntemler.....	22
2.3.1. Fold Change Kriteri	24
2.3.2. Student t-Testi.....	25
2.3.3. Satterthwaite-Welch t-Testi	25
2.3.4. Bağımlı Örneklem t-Testi	26
2.3.5. Moderated t-Testi.....	27
2.3.6. Wilcoxon Rank Sum Testi.....	27
2.3.7. ANOVA F-Testi	28
2.3.8. Moderated F Testi	29
2.3.9. Çoklu Hipotez Testleri.....	30
2.3.9.1. Tip I ve Tip II Hata Oranları.....	30
2.3.9.2. Yanlış Bulgu Oranı	31
3. GEREÇ VE YÖNTEM.....	33
3.1. Analizde Kullanılacak Veri Seti	33
3.2. Kullanılacak İstatistiksel Yöntemler.....	34
4. BULGULAR.....	37
4.1. Kalite Kontrol Sonuçları	37
4.2. Dizayn ve Kontrast Matrisi.....	42
4.3. Farklı Gen Ekspresyon Analizi Sonuçları	43
5. TARTIŞMA VE SONUÇ	50
6. KAYNAKLAR	53
7. TEŞEKKÜR.....	55
8. ÖZGEÇMİŞ	56

ÖZET

Son yıllarda, moleküler biyoloji ve bilişim teknolojisindeki gelişmeler çeşitli canlı türlerinin genomlarının büyük bir kısmının incelenmesine olanak sağlamıştır. Moleküler biyoloji alanında üretilen veriler çok büyük miktarda olduklarından, güncel biyoinformatik çalışmaları genlerin ve proteinlerin yapısal ve işlevsel yönleri ile ilgilenmektedir. Bu çalışmaların çoğu İnsan Genom Projesi ile ilgilidir. Biyoinformatik alanında farklı bilgisayar bilimi teknolojileri ve istatistiksel yöntemler dikkate alınmaktadır. Bu multidisipliner yaklaşım istatistiksel yöntemler, biyoinformatik uygulamaları ve bilgisayar bilimi teknolojileri arasındaki ilişkinin anlaşılmasına olanak sağlamaktadır. Biyoinformatik araçları, mikroarray analizinin gerçekleştirilmesindeki ihtiyaçları karşılamaktadır. Farklı gen ekspresyon analizi ile biyolojik yolların iç yüzü anlaşılmaktadır ve ileriki hipotez geliştirme için bir platform sağlanmaktadır. Gen ekspresyon verilerini analiz etmedeki esas, ekspresyon modelleri fenotip veya deneysel koşula göre farklılık gösteren genleri saptama ihtiyacıdır. Basit bir mikroarray deneyi iki koşul arasındaki ekspresyon farklılıklarını tespit etmek için gerçekleştirilmektedir. Her bir koşul bir veya daha fazla RNA örnekleme tarafından temsil edilebilmektedir. Mikroarrayler geniş çeşitlilikteki deneysel ortamlarda farklı gen ekspresyonunun tespiti için kullanılmaktadır.

Anahtar Kelimeler: Mikrodizi, Gen Ekspresyonu, Farklı Gen Ekspresyon Analizi, Göğüs Kanseri

SUMMARY

ANALYSIS OF MICROARRAY GENE EXPRESSION DATA IN BREAST CANCER

Recent advancements in the fields of molecular biology and information technology led to further study of the most genomes of various species. Since the data generated in the field molecular biology is in large quantity, current bioinformatic researches deal with the structural and functional aspects of genes and proteins. This multidisciplinary approach comprehends the relation between statistical methods, bioinformatic applications and computer science. Bioinformatic tools provide the needs for carrying out microarray analysis. Inside story of the biological pathways understood through different gene expression analysis that provides a platform for developing further hypothesis. Principle in analysing the gene expression data is the need for determine the genes of which expression models differ by phenotype or the experimental condition. A simple microarray experiment is conducted to find the expression differences between two conditions. Every single condition is represented by a RNA sample or more. Microarrays are employed to detect the distinctive gene expression in various experimental conditions.

Keywords: Mikroarray, Gene Expression, Differentially Gene Expression Analysis, Breast Cancer

1.GİRİŞ

Biyoinformatik “biyolojik ve bilişimsel bilimler arasındaki ara birim” olarak tanımlanabilecek yeni gelişen disiplinler arası bir alandır. Biyoinformatik çalışmaları, bilişimsel teknolojiler yardımıyla büyük miktardaki biyolojik verilerin analizi ile ilgilenmektedir. Son yıllarda, moleküler biyoloji ve bilişim teknolojisindeki gelişmeler çeşitli canlı türlerinin genomlarının büyük bir kısmının incelenmesine olanak sağlamıştır. Moleküler biyoloji alanında üretilen veriler çok büyük miktarda olduklarından, güncel biyoinformatik çalışmaları genlerin ve proteinlerin yapısal ve işlevsel yönleri ile ilgilenmektedir. Bu çalışmaların çoğu İnsan Genom Projesi ile ilgilidir. Kısaca, biyolojiye uygulanmakta olan bilişim bilimi biyoinformatik olarak adlandırılmaktadır. Bu alan bilişim teknolojisi, bilgisayar bilimi, biyoloji ve istatistik olmak üzere birden çok akademik disiplini ilgilendirmektedir (1).

Biyoinformatik alanında farklı bilgisayar bilimi teknolojileri ve istatistiksel yöntemler dikkate alınmaktadır. Bu multidisipliner yaklaşım istatistiksel yöntemler, biyoinformatik uygulamaları ve bilgisayar bilimi teknolojileri arasındaki ilişkinin anlaşılmasına olanak sağlamaktadır (2).

Biyoinformatik alanında gerçekleştirilen en basit görevler biyolojik bilgi veri tabanlarının oluşturulması ve geliştirilmesidir. Bu veri tabanlarının çoğunluğunu nükleik asit dizilimleri ve onlardan türetilen protein dizilimleri oluşturmaktadır. Biyoinformatik alanındaki en önemli görevlerden biri de dizilim bilgilerinin analizidir. Biyoinformatik, kompleks biyolojik sorulara açıklık getirmek için, yeni bilişim teknolojilerinin gelişmesi aracılığıyla tüm canlı organizmaların genetik yapılarını incelemektedir (1).

Biyoinformatik, özellikle mikroarray verilerin analizinde, moleküler biyologlar için giderek daha önemli bir araç haline gelmiştir. Mikroarrayler, verileri yorumlanabilir hale getirmek için bir dizi analizi gerektiren büyük miktardaki bilgiyi üretebilmektedirler. Mikroarray sonuçları örneklem koşullarındaki farklılıkları ortaya çıkartmak için doğrudan yorumlanamamaktadır. Mikroarray deneylerini yorumlanabilir hale getirmek için, bir dizi algoritma ve yaklaşımın uygulanması gerekmektedir. Ortaya çıkan verilerin normalizasyonundan sonra, genellikle anlamlılık analizi, ilgili biyolojik bileşenlerin ve örneklemelerin kümelenmesi ve görselleştirme yapılmaktadır (3). Birçok mikroarray çalışması farklı fenotipler ile ilişkili genleri tespit etmek amacıyla tasarlanmaktadır. Bu çalışmada mikroarray verilerinin farklı gen ekspresyon analizine odaklanılmıştır (4).

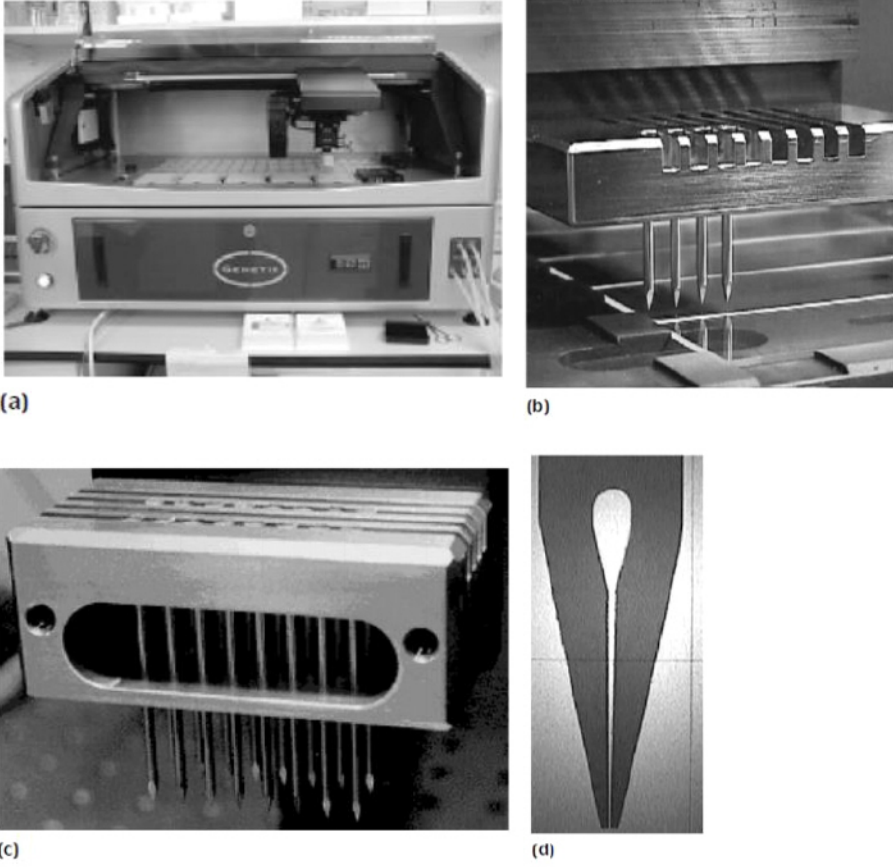
2. GENEL BİLGİLER

2.1. Mikroarray Yapımı ve Kullanımı

DNA mikroarrayleri, DNA moleküllerinin kimyasal olarak bağlanmış olduğu katı bir yüzeyden oluşmaktadır. Bu katı yüzey için genellikle mikroskop lamı kullanılmaktadır. Bir mikroarray'in amacı biyolojik bir örneklem içinde etiketlenmiş nükleik asitlerin varlığını ve çokluğunu tespit etmektir. Mikroarray deneylerinin çoğunda, etiketli nükleik asitler bir örneklem veya dokunun mRNA'sından elde edilmektedir ve böylece mikroarrayler gen ekspresyonunu ölçmektedir (5).

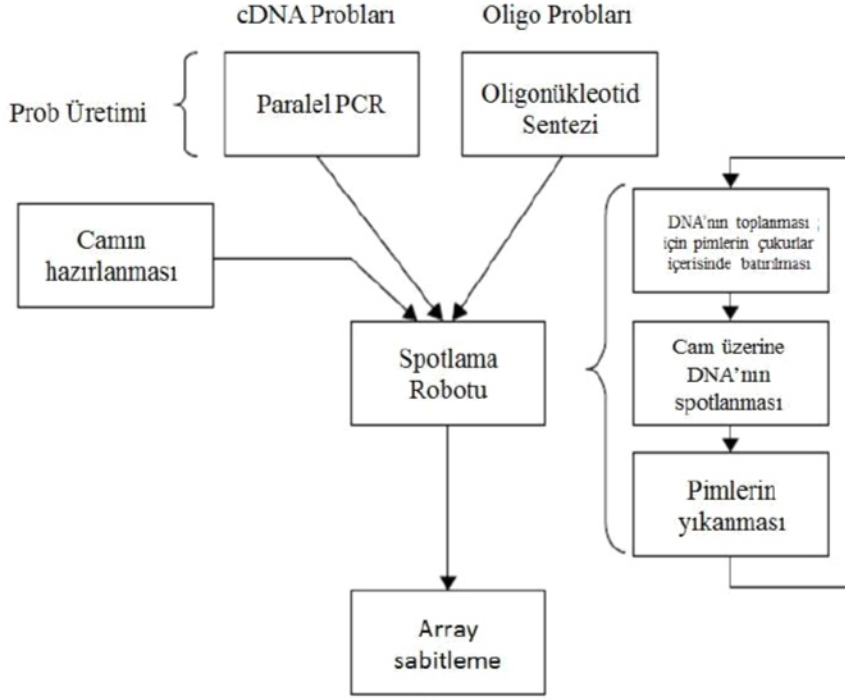
Bir mikroarray'in gücü, bir array'e bağlanmış binlerce farklı DNA molekülü olabilmesidir ve böylece binlerce genin ekspresyonunu eş zamanlı olarak ölçmek mümkün hale gelmektedir. Mikroarray yapımında kullanılan iki ana teknoloji vardır; robotik spotlama ve in-situ sentezi (5).

Piyasada farklı spotlama robotları mevcuttur; Şekil-2.1'de, Oxfordshire'de bulunan Fare Genetiği Birimindeki bir Genetix stoplama robotu gösterilmektedir. Pimler dikdörtgen bir sistem içindeki bir kasette tutulmaktadır. Bu dikdörtgen sistem, sıvının bırakılması için mikrotitre plakaları ve cam arrayleri arasında hareket ettirilebilen bir robot kol içerisinde yer almaktadır. Kasetteki pim sayısı değişebilmektedir. Ne kadar çok pim olursa robotun işlem hacmi o kadar fazla olacaktır, fakat aynı zamanda pim'den pime değişkenlik eğilimi de artacaktır. Modern kullanımdaki pimler, örnekleme dayalı bir rezervuara sahip olduklarından birçok özelliği ortaya çıkarabilmektedirler. Eski robotlar ise yalnızca bir özelliği ortaya çıkarabilen katı pimleri kullanmaktadır (5).



Şekil-2.1: Oxfordshire, Harwell’de bulunan Fare Genetiği Birimindeki bir Genetix stoplama robotu (5).

Spotlanmış array’lerin sentezinde yer alan çeşitli adımlar vardır. cDNA prob’ları yüksek derecede paralel polimeraz zincir tepkimesi (PCR) aracılığıyla yapılmaktadır; oligonükleotit problemleri ön sentezleme gerektirmektedir. Bir sonraki adım spotlama aşamasıdır. Robot kolu, problemleri içeren mikrotiter plakaların biri üzerinden pimleri içeren kaseti hareket ettirmektedir ve DNA’yı toplamak için pimleri çukurlar içerisine batırmaktadır. Daha sonrasında, kol ilk array üzerine taşınmakta ve kaset aşağı hareket etmektedir, böylece pimler cama değmekte ve DNA yüzey üzerine bırakılmaktadır. Eğer birden fazla array sentezleniyorsa kaset sonraki arraylere taşınmaktadır. Spotlanacak bir sonraki DNA alınmadan önce, herhangi bir bulaşma olmaması için pimler yıkanmaktadır. Array üretiminin son aşaması ise cam yüzeyin düzeltildiği sabitleme aşamasıdır (5).



Şekil-2.2: Spotlanmış array sentezi (5).

2.1.1. Spotlanmış Microarray

Bu yöntem ilk mikroarraylerin üretildiği teknolojidir. Array üç ana adım aracılığıyla (şekil-2.2) bir spotlama robotu (Şekil-2.1 a) kullanılarak yapılmaktadır (5).

1. Array üzerine yerleşecek DNA problemlerini yapmak
2. Spotlama robotu ile DNA'yı array'in cam yüzeyi üzerine spotlamak
3. Cam yüzeyin post-spotting işlemi

Spotlanmış array'lerin 3 ana türü vardır. DNA prob türüne göre veya prob kimyasının cama bağlanması ile elde edilmektedirler (Tablo-2.1).

Spotlanmış bir array üzerinde kullanılan DNA problemleri, polimeraz zincir tepkimesi üretimleri (PCR) veya oligonükleotitler olabilmektedir. İlk durumda, bir klon kütüphanesinden DNA'yı büyütmek için yüksek derecede paralel PCR kullanılmaktadır ve büyütülmüş DNA arıtılmaktadır. İkinci durumda ise DNA oligonükleotitleri array üzerinde kullanılmak için sentezlenmektedir (5).

Tablo-2.1: Spotlanmış microarraylerin üç türü

DNA probları	Yüzey Kimyası	
	Kovalent	Kovalent Olmayan
Oligonükleotit	✓	
cDNAs	✓	✓

DNA probları, oligonükleotitler veya cDNA'ler; yüzey kimyası ise kovalent veya kovalent olmayan olarak ikiye ayrılmaktadır. Oligonükleotit probları yalnızca kovalent olarak bağlanmış olabilir; cDNA probları ise kovalent veya kovalent olmayan şekillerde bağlanabilirler. Kovalent bağ, bir alifatik aminin (NH₂) DNA probunun 5' sonuna eklenmesi yoluyla oluşmaktadır. Dolayısıyla problemler 5' sonundan cama bağlanmaktadır. Kovalent olmayan bağ ise cam slayt (lam) üzerindeki amin grupları ve DNA prob omurgası üzerindeki fosfat grupları arasındaki elektrostatik çekim aracılığıyla oluşmaktadır; bu şekilde DNA probu omurgası aracılığıyla cam'a bağlanmaktadır (5).

Bağlanma kimyası, kovalent veya kovalent olmayan şekillerde olabilmektedir. Kovalent bağlanma ile, ilk alifatik amin (NH₂) grubu DNA probuna eklenmektedir. Cam üzerindeki kimyasal bağlayıcılar ile bu grup arasında bir kovalent bağ yapılarak prob cam'a bağlanmaktadır. Kovalent olmayan bağlanma ile, probun array'e bağlanması, DNA probunun fosfat omurgası ve camın yüzeyine bağlı NH₂ grupları arasındaki elektrostatik çekim ile gerçekleşmektedir. Etkileşim DNA omurgası boyunca çeşitli yerlerde gerçekleşmektedir. Böylece prob cama farklı noktalarda bağlanmaktadır. Çoğu oligonükleotit probu cDNA'lerden daha kısa olduğundan, bu etkileşimler oligonükleotit problemlerini cama tutturmak için yeterince kuvvetli değildirler. Bu nedenle, kovalent olmayan bağ çoğu zaman yalnızca cDNA mikroarray'leri için kullanılmaktadır (5).

DNA probları microtiter plakalar içinde oluşturulmaktadır. Spotlama robotu bir kaset içinde tutulan ve ızgara şeklinde sıralanmış bir dizi pim içermektedir (Şekil-2.1b ve Şekil-2.1c). Pimler sıvının microtiter plakalardan cam array'e aktarılmasında kullanılmaktadır. İlk spotlama robotları katı pimleri kullanmaktaydı; bu pimler yalnızca array üzerindeki bir spot için yeterli sıvıyı tutabilmekte ve böylece bir sonraki spotu basmadan önce pim kasetinin probu içeren plakaya geri döndürülmesi gerekmektedir (5).

Günümüzde array üretim robotlarındaki pimler, sıvıyı tutmakta kullanılan bir rezervuara sahiptir (Şekil-2.1d). Böylece her bir prob birden fazla array üzerinde spotlanabileceğinden, daha yüksek verimlilikte array üretimi yapılabilmektedir (5).

Baskı işlemi temel olarak beş adımda yapılmaktadır:

1. İlk DNA dizisini toplamak için pimler çukurlar içerisine batırılmaktadır.
2. Yapılacak array sayısına ve pimlerin tutabileceği sıvı miktarına bağlı olarak, bu DNA farklı arrayler üstüne spotlanmaktadır.
3. Pimler, bir sonraki örnekleme bir bulaşma olmasını önlemek için herhangi bir artık çökeltinin kalmaması amacıyla yıkanmaktadır.
4. Pimler bir sonraki çukur setine batırılmaktadır.
5. İkinci adıma dönmekte ve array tamamlanana kadar bu döngü tekrar edilmektedir.

Array üretiminin son aşamasında, array'in yüzeyi sabitlenebilir. Böylece daha fazla DNA yüzeye eklenemez. Cam yüzeyin üzerindeki yapının kimyasına bağlı olarak farklı sabitleme işlemleri vardır. İstenen sonuç her zaman aynıdır: hibridizasyon sırasında array'in camına örneklemeden DNA hedefinin yapışmaması istenir. Bu durumun ortaya çıkmaması için de yüzey değiştirilmelidir. Yüzeyin değişmesi aynı zamanda camın daha su emici hale gelmesini sağlamaktadır ve hibridizasyon sırasında hedef çözeltisinin karıştırılmasına yardımcı olmaktadır (5).

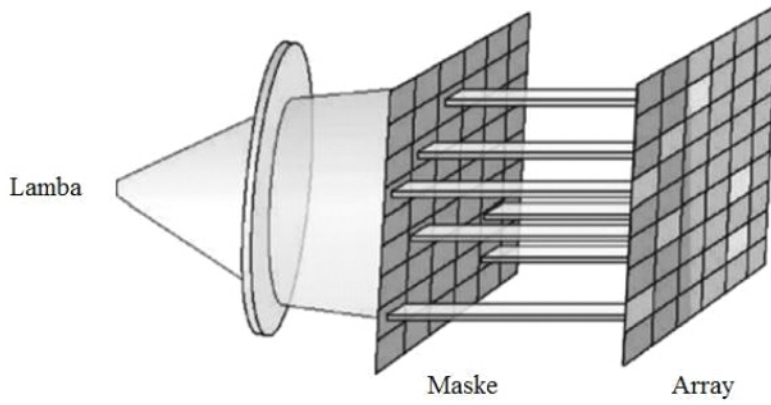
2.1.2 In-Situ Sentezlenmiş Oligonükleotit Arrayler

Bu arrayler temelde spotlanmış arraylerden farklıdır. Fark, önsentezleme oligonükleotitleri yerine arrayin yüzeyinde bazdan baza oligolar oluşturulmasıdır. İşlem, eklenecek son nükleotidin şekerinin 5 hidroksil grubu ve bir sonraki nükleotidin fosfat grubu arasındaki kovalent reaksiyon ile gerçekleşmektedir. Cam üzerinde oligonükleotide eklenecek her bir nükleotidin, 5' konumu üzerinde koruyucu bir grubu vardır. Koruyucu grup, her bir sentez sırasında birden fazla bazın eklenmesini önlemeye yaramaktadır. Daha sonrasında bu koruyucu grup bir sonraki sentezden önce asit veya ışık ile bir hidroksil gruba dönüşmektedir (5).

Koruyucu grubun uzaklaştırılması (deprotection) için farklı yöntemler in-situ sentezlenmiş arraylerin yapılabileceği üç ana teknolojiye yol açmıştır:

1. Affymetrix Teknolojisi
2. Maskless Photodeprotection Teknolojisi
3. Agilent ve Oxford Gen Teknolojisi

Affymetrix arrayleri ışık-aracılı deprotection aşaması ile birlikte in-situ sentez kullanılarak üretilmektedir. Her bir sentez sırasında, arrayin uygun bölümlerine tek bir baz eklenmektedir. Işığı arrayin uygun bölgelerine doğrultmak için bir maske kullanılmaktadır. Böylece baz doğru özelliklere eklenmiş olmaktadır. Her bir aşama farklı bir maske gerektirmektedir (5).



Şekil-2.3: Affymetrix Teknolojisi (5).

2.1.2.1. Affymetrix Teknolojisi

Affymetrix arrayler, terminal nükleotit üzerindeki koruyucu grubu, daha fazla bazın eklenebileceği bir hidroksil grubuna dönüştürmek için ışığı kullanmaktadırlar. Işık maske kullanılarak uygun özelliklere doğru yönlendirilmektedir. Maske ışığın, arrayin yalnızca belirli alanlarına geçmesini sağlamaktadır (Şekil-2.3). Bu teknik fotolitografi olarak da bilinmektedir ve ilk defa silikon çiplerin üretilmesi için kullanılmıştır. Ancak maske takımı tasarlandığında ve yapıldığında, çok sayıda özdeş arrayin üretilebileceği anlaşılmıştır. Affymetix teknolojisi çok sayıda “standart” array yapılması için oldukça uygundur (5).

2.1.3. Gen Ekspresyon Mikroarrayleri

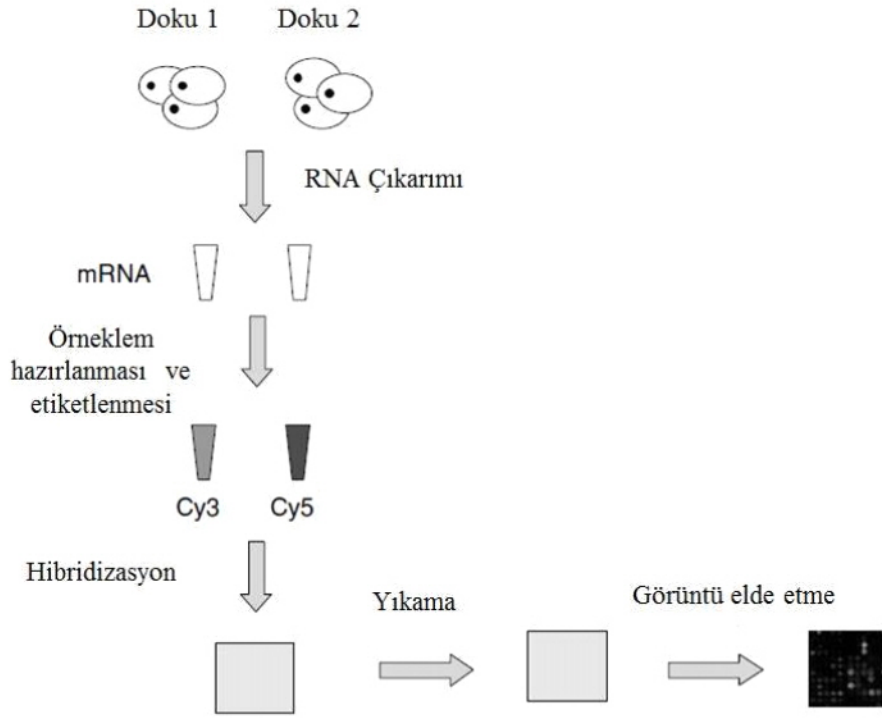
Gen ekspresyonu, her bir genin kalıp DNA'sından protein sentezlenmesi sürecidir. Bu sürecin ilk aşaması transkripsiyondur. Transkripsiyon, DNA'nın bir sarmalının RNA olarak kopyalanması sürecidir. Ökaryotlarda bu süreç RNA uçbirleşmesi olarak izlenmektedir (6).

Uçbirleşme, intronların birincil kopyadan ayrılması ve ergin mRNA'nın yapılmasıdır. RNA'nın transkripsiyonu ve uçbirleştirmesi nükleus içerisinde gerçekleşmektedir. Gen ekspresyonunun bir sonraki aşaması mRNA'nın proteine dönüşmesidir. Bu aşama sitoplazma içerisinde gerçekleşmektedir. Gen ekspresyonu sürecinde, RNA yalnızca temel alt maddeyi (mRNA) değil aynı zamanda protein sentez araçlarının (tRNA, rRNA) bileşenlerini de sağlamaktadır (6).

Bazı protein kodlayan genler hemen hemen sabit düzeyde kopyalanmaktadırlar; bu genler *housekeeping genler* olarak adlandırılmaktadır ve temel reaksiyonlar için her zaman gereklidirler. Bir geni maskeleyecek sinyal, hücrenin dışından da gelebilmektedir; örneğin, bir besinden veya bir hormondan. DNA içindeki özel düzenleyici dizilimler, bir genin sinyallere yanıt verip vermeyeceğini belirlemektedirler ve dolayısıyla protein kodlayan genin transkripsiyonunu etkilemektedirler. Hangi genlerin hangi koşullar altında ifade edildiğini anlamak, hücredeki biyolojik süreçler hakkında çok önemli bilgiler vermektedir. Mikroarray teknolojisinin gücü binlerce genin ekspresyonunu eşzamanlı olarak ölçme yeteneğinde yatmaktadır (6).

Microarray kullanılarak bir örnekteki gen ekspresyonunu ölçmek için dört laboratuvar adımının gerçekleştirilmesi gereklidir (şekil-2.4);

1. Örnek hazırlanması ve etiketleme
2. Hibridizasyon
3. Yıkama
4. Tarama ve Görüntü elde etmek



Şekil-2.4: Bir mikroarray'in kullanılmasındaki adımlar (5).

Bir mikroarrayin kullanılmasında ilk aşama ilgili dokudan ya da dokulardan RNA'nın çıkartılmasıdır. Çoğu mikroarray teknolojisinde, iki örneklem hazırlanır ve bu örneklem genellikle Cy3 (yeşil) ve Cy5 (kırmızı) olmak üzere iki farklı boya ile etiketlenir. Daha sonrasında bu örneklem array'e eş zamanlı olarak melezleştirilir (hibridizasyon). Örneklem 45°C ve 65°C arasında bir ısıda 12 ila 24 saat boyunca inkübe edilir ve sonrasında array yıkanır (5).

2.1.3.1. Örneklem Hazırlanması ve Etiketleme

Microarray deneylerinde örneklemin hazırlanabileceği ve etiklenebileceği çeşitli yöntemler vardır. Bütün yöntemlerde ilk aşama ilgili dokudan RNA'nın çıkartılmasıdır. Bu işlem yinelenmesi zor bir işlemdir ve RNA çıkarımını gerçekleştiren uzmanlar arasında çok büyük değişkenlik vardır. Eskiden DNA radyoaktif olarak etiketlenmekteyken, günümüzde çoğu laboratuvar Cy3 ve Cy5 boya ile floresan etiketleme yöntemini kullanmaktadır. En yaygın deneylerde, her biri boya ile etiketlenmiş iki örneklem arraylere melezleştirilmektedir; bu her iki örneklemin de eş zamanlı ölçülmesini sağlamaktadır (5).

Standart mikroarray deneylerinde, iki farklı biyolojik örnekteki mRNA ekspresyonları karşılaştırılmaktadır. Affymetrix GeneChip ve CodeLink Bioarray sistemleri yalnızca bir örneğin her bir çiple melezleştiği tek-renkli saptama planı kullanılmaktadır. Klasik spotlanmış mikroarrayler ve Agilent sistemi aynı arrayin, her biri farklı floresan boyası (genellikle Cy3 ve Cy5) ile etiketlenmiş iki örnek ile melezleştirildiği iki-renkli bir plan kullanılmaktadır. Floresan sinyallerinin oranı, aynı mikroarray üzerindeki iki farklı biyolojik örneğin transkript miktarını temsil etmektedir. Genellikle, ilgili örnek mRNA'sı bir floresan boyası ile etiketlenmekteyken bir referans mRNA'sı da başka bir boya ile etiketlenmektedir. Bu referanslar mikroarray cDNA'lerinin referans sinyallerini elde etmek için normal hücreler veya dokulardan üretilmektedir (7).

2.1.3.2. Hibridizasyon ve Yıkama

Manüel ve robotik olmak üzere iki ana hibridizasyon yöntemi vardır. Manüel hibridizasyon yönteminde, array bir hibridizasyon haznesine yerleştirilmektedir. Laborant, hedefi içeren hibridizasyon çözeltisini array üzerine enjekte etmektedir. Hazne, arrayi doğru ısıda tutacak bir kuvöz içerisine yerleştirilmektedir. Alternatif olarak hibridizasyon işlemi bir robot aracılığıyla mekanik olarak yapılabilmektedir. Robotik hibridizasyon, ısının kontrol edilmesi açısından manüel hibridizasyon yöntemine karşı avantajlıdır. Hibridizasyon aşamasından sonra, slaytlar yıkanmaktadır. Bunun yapılmasının sebebi, array üzerinde etiketlenmiş hedefin spesifik özelliklere bağlı hedef olduğunu kesinleştirmektir. Böylece bu hedef ölçülmek istenen DNA'yı temsil etmektedir (5).

2.1.3.3. Tarama ve Görüntü Elde Etme

Melezleştirilmiş array yüzeyinin görüntüsünü üretmek laboratuvar sürecinin son aşamasıdır. Array üzerindeki heterodupleks, belirli bir dalga boyundaki ışık ile uyarılan boyayı içermektedir. Bu aşamada slayt, yüzeyi okuyabilen bir tarayıcı içerisine yerleştirilmektedir. Tarayıcı, array üzerine odaklanan bir veya daha fazla lazeri içermektedir. Çoğu tarayıcı iki-renkli arrayler için iki lazer kullanılmaktadır. Dijital görüntü üzerindeki her bir piksel, array üzerindeki noktaya lazerin odaklanması ile uyarılan floresan yoğunluklarını temsil etmektedir. Bu nokta üzerindeki boya lazer tarafından uyarılmaktadır ve floresan ışığı yaymaktadır. Tüm array'i taramak amacıyla, lazer array üzerindeki her bir noktaya odaklanmalıdır (5).

Taranmış bir görüntüdeki her bir piksel değeri, slayt üzerindeki belirli bir konumdaki hibridizasyon seviyesini temsil etmektedir. Belirli bir taranmış DNA dizilimi için hibridizasyonun toplam miktarı spottaki toplam floresan ile orantılıdır. Bu yüzden spot yoğunluğunun doğal ölçümü spot maskesi içerisindeki piksel yoğunluğunun toplamına eşittir (8). Piksel ile ifade edilen fiziksel alanın büyüklüğünün lazerin spot boyutu ile aynı olması olağan bir durumdur. Bu durum pikselde ölçülen ışığın yoğunluğunun arrayin üzerindeki noktadan geldiğini göstermektedir. Aynı zamanda lazer spotundan küçük piksel boyutunun elde edilmesi de mümkündür. Bu durum her bir piksel üzerindeki ışığın arrayin komşu alanlarından geldiğini göstermektedir. Böylece görüntü bulanmaktadır ve özellik içindeki küçük düzensizlikler gizlenebilmektedir (5).

İki-renkli arrayler ile, tarayıcı çıktısı genellikle iki monokrom görüntüsü olarak elde edilmektedir. İki monokrom görüntüsü, mikroarraylerin kırmızı-yeşil yanlı renkli görüntülerini oluşturmak için birleştirilmektedir (5).

2.2. Mikroarray Verilerinin Ön İşlemesi

Mikroarray teknolojisi, nükleik asidin hibridizasyon özelliklerinden yararlanmaktadır ve hedef örnekleme bulunan ilgili nükleik asit transkripsiyonlarının miktarını ölçmek için, katı bir yüzeye bağlanmış prob olarak adlandırılan tamamlayıcı molekülleri kullanmaktadır. Hedefteki moleküller etiketlenmektedir ve yoğunluk olarak adlandırılan her bir probta melezleşmiş hedef miktarını ölçmek için özel bir tarayıcı kullanılmaktadır (6).

Çoğu üretici, arraylerde spesifik olmayan bağlanım ve optik gürültünün etkilerine rastlamaktadır. Uygulamada, çeşitli değişkenlik kaynaklarının ölçülmesi gerekmektedir. Söz konusu değişkenlik ve gürültü sebeplerinden dolayı ham veriler ayarlanmalıdır (5).

İki-renk olarak spotlanmış platformlar her bir microarray (kırmızı ve yeşil kanallar) için prob-seviyesinde iki adet veri seti üretmektedir ve cam slaytta prob içermeyen alanlardan lokal arka plan gürültü seviyeleri ölçülmektedir (5). Spotlanmış arrayler için veri ön işleminin amacı, kırmızı ve yeşil kanallar için her bir spot veya genin ekspresyon değerinin spotun kendi bölgesinde ölçülmesidir (7). Ön işleme temel olarak 6 bölüme ayrılabilir: görüntü analizi, veri alımı, arka plan ayarlaması, normalizasyon, özetleme ve kalite değerlendirmesi (7).

Görüntü analizi taranan görüntülerdeki piksel yoğunluklarının prob-seviyesinde verilere dönüştürülmesine olanak sağlamaktadır. Veriler farklı formatlar halinde olabildiklerinden genellikle veri alımı yöntemlerine gerek duyulmaktadır. Spesifik olmayan hibridizasyonlar ve optik tarama sistemindeki gürültü nedeniyle ölçülmüş prob yoğunlukları ortaya çıktığından dolayı arka plan ayarlaması gerekmektedir. Doğru spesifik hibridizasyon ölçümleri elde etmek için gözlenen yoğunluklar ayarlanmalıdır. Birçok değişkenlik kaynağı olduğundan uygun bir normalizasyon yapılmadan farklı array hibridizasyonlarından elde edilen ölçümleri karşılaştırmak mümkün değildir. Bu değişkenlikler ters transkripsiyon, etiketleme, hibridizasyon reaksiyonları, array ile ilgili fiziksel problemler, dizi etkileri (batch effects) ve laboratuvar koşulları gibi sebeplerden kaynaklanmaktadır. Bazı platformlarda, transkripsiyonlar birden çok prob tarafından temsil ediliyorsa özetleme aşaması gerekli olmaktadır. Her bir gen için, ayarlanmış ve normalleştirilmiş yoğunluklar RNA transkripsiyon miktarına orantılı bir niceliğe özetlenmelidir. Kalite kontrol, kabul edilebilir rastgele dalgalanma seviyesini aşan aykırı ölçümleri tespit etmek için oldukça önemli bir süreçtir (5).

2.2.1. Görüntü Analizi

Etiketlenmiş kontrol ve sorgu DNA'ları ile hibridizasyon işlemi gerçekleştirildikten sonra array taranmaktadır. İki floresan boyasına ilişkin iki görüntü elde edilmektedir. Bu görüntüler, problemlere ilişkin spotları tespit etmek ve miktarlarını belirlemek için analiz edilmektedir (9).

Taranan görüntülerin işlenmesi genellikle üç görevi içermektedir.(1) Izgaralama olarak adlandırılan spotların her birine koordinat atanması işlemi, (2) piksellerin ön plan veya arka plan olarak sınıflandırılmasını sağlayan segmentasyon işlemi; ve (3) array üzerindeki her bir spot için yoğunluk çıkarımı. Bu yoğunluklar, kırmızı ve yeşil ön plan floresan yoğunlukları, arka plan yoğunlukları ve bazı durumlarda kalite ölçümleri olarak ifade edilebilirler (9).

2.2.2. Arka Plan Düzeltmesi

Gen ekspresyonu ölçümü genellikle floresan yoğunluğu ölçümü olarak elde edilmektedir (6). Arka plan Floresanı; spesifik olmayan bir bağlanma, yıkama aşamasından sonra arta kalan çökeltiler, tarayıcıdan kaynaklanan optik gürültüler gibi birçok kaynaktan meydana gelebilmektedir (4).

Ölçüm değerleri, kalibrasyon gibi cihaz sistemi içerisinde çeşitli ayarlamalar geçirmiş olabilir. Bu nedenle, gen ekspresyon verileri sunulurken değerlerin cihaz sistemi tarafından nasıl üretildiği açıklanmalıdır. Bu ekspresyon ölçümleri her zaman “arka plan gürültüsü” olarak adlandırılan bir bileşen içermektedir (6). Yerel arka plan gürültü seviyeleri prob içermeyen cam slayttaki alanlardan ölçülmektedir. Arka plan düzeltilmesi, belirsiz arka plan gürültüsünü ve her bir çip üzerindeki genel sinyal seviyesinin yerel değişimlerini gidermeye çalışmaktadır (10). Arka plan etkisini ortadan kaldırmak için en yaygın olarak kullanılan yöntem spotlar etrafında ölçülen floresan yoğunluğunun çıkartılmasıdır (9).

Mikroarray gen ekspresyon veri setleri, $g = 1, \dots, G$ tüm genler ve $n = 1, \dots, N$ tüm örneklemeler olmak üzere gen ekspresyon ölçümleri ω_{gn} değerlerinden oluşmaktadır. Genler satırlara örneklemeler ise sütunlara karşılık gelmek üzere ω_{gn} değerleri $G \times N$ ’lik bir matris içerisinde yer almaktadır (6). Okunan ω_{gn} gen ekspresyon değeri, istatistiksel olarak gerçek gen ekspresyon değeri x_{gn} ve arka plan gürültüsü B_{gn} bileşenlerinin toplamı olarak tanımlanabilmektedir (6);

$$\omega_{gn} = x_{gn} + B_{gn} \quad (2.1)$$

Arka plan gürültüsünün yapısı ve düzeltilmesi, kullanılan mikroarray teknolojisine bağlıdır. Spotlanmış array verileri, düzeltilmemiş ekspresyon yoğunlukları ω_{gn} ’ler ile birlikte arka plan gürültüsü B_{gn} ’nin bir tahminini sağlamaktadır. Arka plan tahmini \hat{B}_{gn} olarak ifade edilirse arka plan düzeltilmesi yapılmış ekspresyon değerleri $\omega_{gn}^{(c)}$ aşağıdaki gibi elde edilmektedir (6);

$$\omega_{gn}^{(c)} = \omega_{gn} - \hat{B}_{gn} = (x_{gn} + B_{gn}) - \hat{B}_{gn} \quad (2.2)$$

Mikroarray analizlerinde arka plan düzeltilmesi için kullanılan en yaygın yöntemler; “Robust Multi Array Average (RMA) Arka Plan Düzeltilmesi” yöntemi ve “MAS 5.0 Arka Plan Çıkarması” yöntemleridir (10).

2.2.2.1. RMA Arka Plan Düzeltmesi

RMA arka plan düzeltmesi, yalnızca Perfect Match (PM) yoğunluklarını kullanan bir yöntemdir. PM değerleri, prob yoğunluklarının dağılımı için küresel bir model kullanılarak, düzeltilmektedir (4).

Model, prob yoğunluklarının deneysel dağılımını esas almaktadır. Gözlenen PM problemleri, μ ortalamalı ve σ^2 varyanslı bir Gauss gürültü bileşeni olarak modellenmektedir (4).

Negatif ekspresyon değerlerini önlemek için, normal dağılım sıfırda kesilmektedir. Gözlenen yoğunluğun Y olduğu düşünülürse, düzeltme aşağıdaki gibi olacaktır;

$$E(S|Y = y) = a + b \frac{\phi\left(\frac{a}{b}\right) - \phi\left(\frac{y-a}{b}\right)}{\phi\left(\frac{a}{b}\right) + \phi\left(\frac{y-a}{b}\right) - 1} \quad (2.3)$$

S , α ortalamalı üstel bir sinyal bileşeni iken $\alpha = s - \mu - \sigma^2\alpha$ ve $b = \sigma$ 'dir. ϕ ve Φ sırasıyla standart normal yoğunluk ve dağılım fonksiyonlarıdır (4).

2.2.2.2. MAS 5.0 Arka Plan Düzeltmesi

MAS 5.0 Arka Plan Düzeltmesi yönteminde çip, k adet dikdörtgen biçiminde bölgesi olan bir kafese bölünmektedir. Her bir bölgede, prob yoğunluklarının en az %2'si bu kafes için bir arka plan değeri hesaplamak amacıyla kullanılmaktadır. Daha sonra, her bir prob yoğunluğu, arka plan değerlerinin ağırlıklandırılmış bir ortalamasına dayanarak düzeltilmektedir. Ağırlıklar, problemler arasındaki mesafeye ve kafesin ağırlık merkezine bağlıdır (4). $d_k(x, y)$, (x, y) konumundan k bölgesinin ağırlık merkezine kadar olan öklit uzaklığı iken ağırlıklar aşağıdaki gibi hesaplanmaktadır (4);

$$\omega_k(x, y) = \frac{1}{d_k^2(x, y) + s_0} \quad (2.4)$$

Düzeltilme katsayısı s_0 için geçerli değer 100'dür. MAS 5.0 Arka Plan Düzeltmesi yönteminde, hem Perfect Match (PM) hem de Mismatch (MM) problemleri düzeltilmektedir (4).

2.2.3. Kalite Değerlendirmesi

Arraylerin normalleştirilmesinden önce verilerin kalitesini değerlendirmek gerekmektedir. Genel hibridizasyon değerlendirme, array hibridizasyon performansının genel değerlendirilmesini göstermektedir. Kalite kontrol değerlendirme deneysel verilerin kalitesinin kabul edilir olup olmadığını ve herhangi bir hibridizasyonun tekrarlanmasının gerekip gerekmediğini belirlemek için yapılmalıdır (4).

Kalite kontrol değerlendirme işleminde hibridizasyon veya diğer deneysel yapılar ile ilgili potansiyel problemleri saptamak için çeşitli açıklayıcı veri grafikleri çizilmektedir. Kalite kontrol grafikleri temelde diyagnostik ve spot istatistikleri olmak üzere ikiye ayrılmaktadır (4).

- i. Diyagnostik Grafikler
- ii. Spot İstatistikleri Grafikleri

i. Diyagnostik Grafikler

Diyagnostik grafikler, yoğunluk yanlılıklarını değerlendirmeye yarayan MA-grafikleri ve her bir kanal için sinyal-gürültü oranlarını inceleyen histogramlar gibi çeşitli grafikleri içermektedir. Diyagnostik Grafikler genellikle iki kanal arasındaki lineer olmayan eğilimleri gözlemlemek için kullanılmaktadır (4).

i.a. MA Grafikleri

M ve A iki renkli arraylerin analizinde çok yaygın olarak kullanılan değişkenlerdir. A aşağıdaki gibi tanımlanmaktadır;

$$A = \log_2 \sqrt{Cy5 \cdot Cy3} = \frac{1}{2} [\log_2(Cy5) + \log_2(Cy3)] \quad (2.5)$$

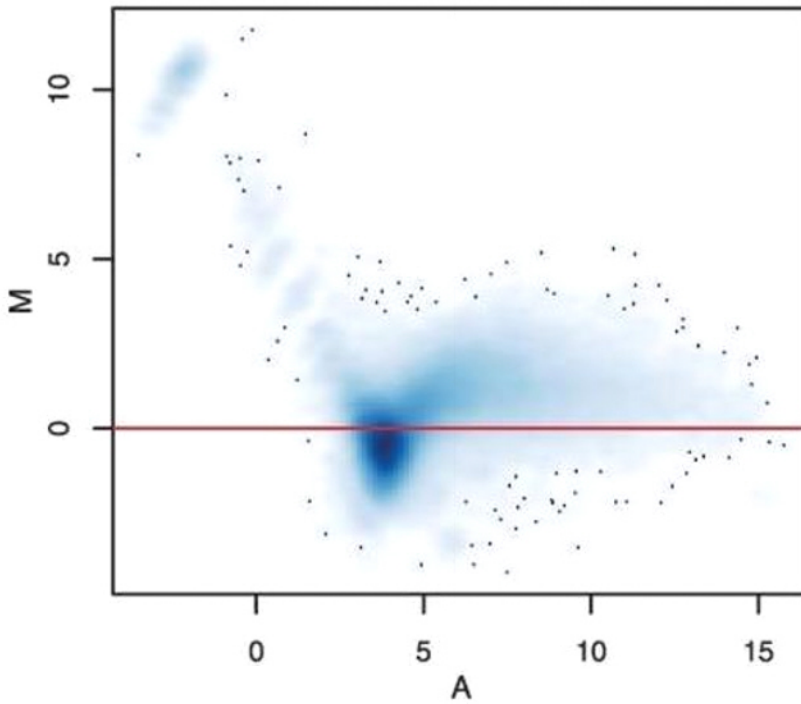
$Cy5$ ve $Cy3$, sırasıyla belirli bir spotun yeşil ve kırmızı boya yoğunluklarını göstermektedir. A değişkeni bir spotun logaritmik dönüşüm uygulanmış toplam yoğunluğunun ölçümüdür. Böylece, birleştirilmiş kırmızı ve yeşil yoğunluklar belirli bir spot için yüksek ise A değeri de yüksek olacaktır (4).

M deęiřkeni ise ařaęıdaki gibi tanımlanmaktadır;

$$M = \log_2 \frac{Cy5}{Cy3} = \log_2(Cy5) - \log_2(Cy3) \quad (2.6)$$

M deęiřkeni, kırmızı boya yoęunluklarının yeřil boya yoęunluklarına oranının logaritmik donüşüm uygulanmış halidir. M deęiřkeni kırmızı ve yeřil boyalardan hangisinin belirli bir spottaki array'e daha fazla baęlandığını göstermektedir (4).

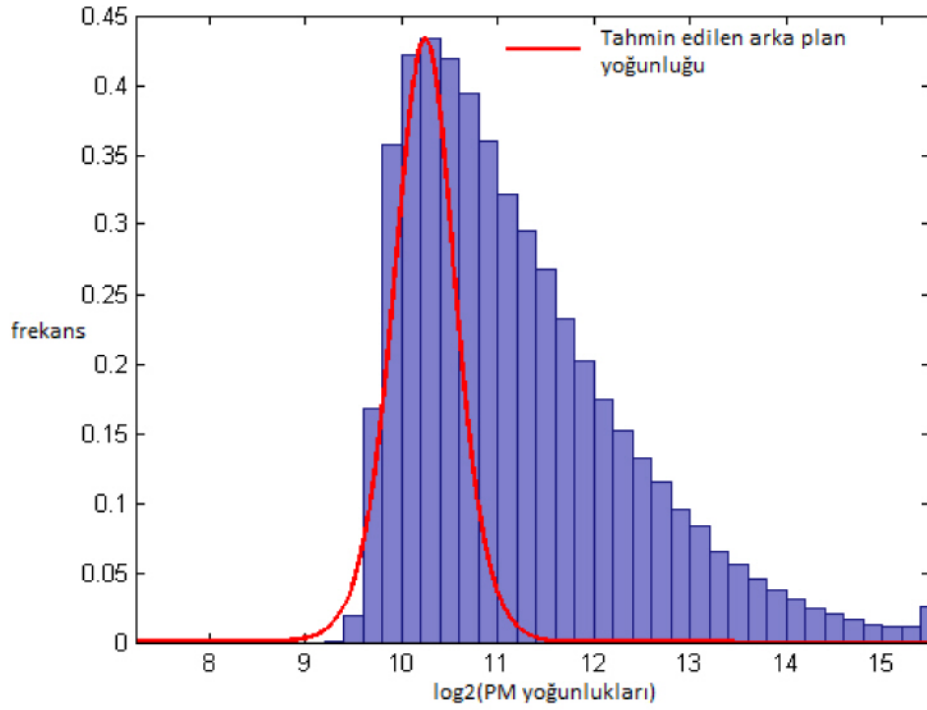
MA grafikleri yoęunluk yanlılıđını arařtırmaya olanak saęlayan grafiklerdir. Grafik üzerinde x-ekseninin altında veya üstünde orantısız miktarda spot bulunması array içerisinde bir problem olduđunun belirtisidir. MA grafikleri array içerisinde normalizasyonun gerekip gerekmediđinin bir göstergesidir (4).



řekil-2.5: MA Grafiđi. M deęerlerinin sıfır ortalama üzerinde merkezlenmemiř olması log-oranı ve yoęunluklar arasında bir baęımlılık olduđunu göstermektedir (11).

i.b. Histogramlar

Mikroarray tasarımlarda, farklı gen ekspresyonunu değerlendirmek için yapılan testlerin p-değerlerinin histogramlarını elde etmek oldukça önemlidir. Histogramlar yorumlaması kolay ve önemli bilgiler içeren grafiklerdir. Bir histogram, verilerde sinyal olup olmadığının, genlerin farklı ifade edilip edilmediklerinin göstergesidir. Ayrıca, histogramlar gerçekte ne kadar genin farklı ifade edildiği hakkında kestirim yapmaya da olanak sağlamaktadır (12).



Şekil-2.6: Histogram Grafiği.

ii. Spot İstatistikleri Grafikleri

Spot İstatistikleri, spot ve hibridizasyon yapılarını öngörmeye yaramaktadır. Uzaysal Grafikler, Kutu grafikleri, Serpme Diyagramları ve volkan grafikleri, spot istatistikleri ile elde edilebilecek grafiklerin başlıcalarıdır (4).

ii.a. Spot İstatistiklerinin Uzaysal Grafikleri

Uzaysal grafikler, düzgün olmayan spot ve hibridizasyon yapılarını ortaya çıkarmak için kullanılmaktadır. Her bir array üzerindeki yoğunlukların uzaysal dağılımlarını gözlemlemek ve bozuklukları tespit etmek için uzaysal grafikler kullanılmaktadır. Arrayler üzerinde herhangi bir boyutsal bozukluk olup olmadığına göre gerekli arka plan düzeltmesine karar verilmesinde uzaysal grafikler önemli rol oynamaktadır (4).

ii.b. Spot İstatistiklerinin Kutu Grafikleri

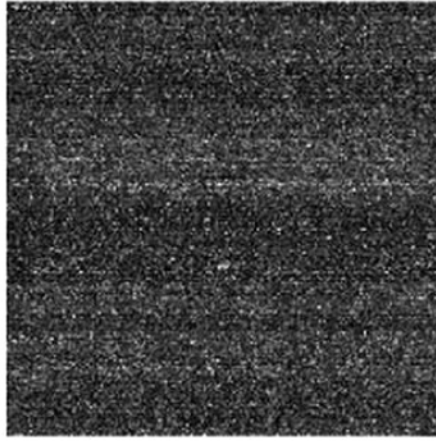
Kutu grafikleri, spot ve hibridizasyon yapılarını gözlemlemek için kullanılan grafiklerden biridir. Aynı zamanda farklı arrayler arasındaki ölçek farklılıklarını anlamak için kutu grafikleri çizilebilmektedir. Arrayler arası normalizasyonun gerekli olup olmadığını anlamak için kutu grafiklerini değerlendirmek gerekmektedir. Arraylerin homojenliği kutu grafiklerinden oldukça açık bir şekilde gözlemlenebilmektedir (4).

ii.c. Spot İstatistiklerinin Serpme Diyagramı Grafikleri

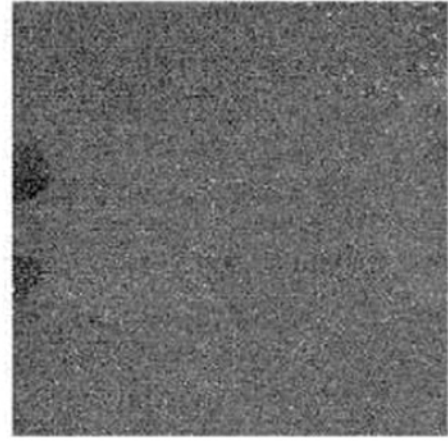
İki örnekleme ait ekspresyon değerlerini karşılaştırmak için kullanılan serpme diyagramları mikroarray verilerin görselleştirilmesinde en yaygın olarak kullanılan grafiklerdir (3). Mikroarray verilerin analizinin ilk aşamasında, genel yapıların ve değişkenliğin incelenmesi için iki yoğunluk kanalı arasında serpme diyagramı çizilmektedir. Serpme diyagramları aynı zamanda anormal ekspresyon yapılarına sahip, merkezden uzak slaytları incelemek için yaygın olarak kullanılmaktadır (13).

ii.d. Spot İstatistiklerinin Volkan Grafikleri

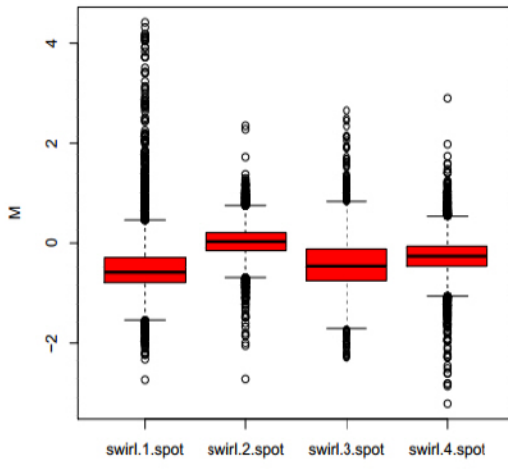
Volkan Grafikleri, fold change ve t-testi kriterlerini özetleyen grafiklerdir. Volkan Grafikleri, Log₂ fold change'e karşı gene özgü t-testinden negatif log₁₀ dönüşümü yapılmış p değerlerinin serpme diyagramı olarak çizilmektedir. Mikroarray üzerinde ölçülmüş her bir gen için, grafik üzerinde o genin t istatistiğini ve fold change'ini temsil eden bir nokta bulunmaktadır (12).



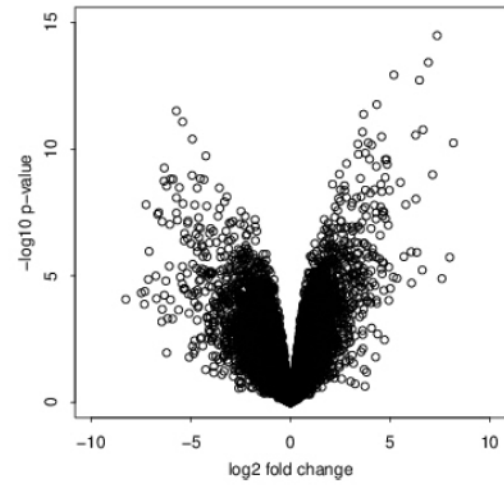
(a)



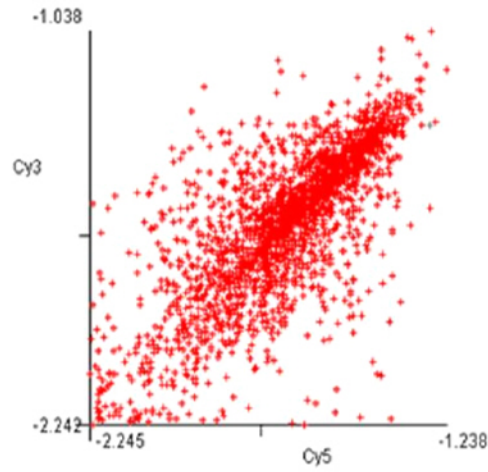
(b)



(c)



(d)

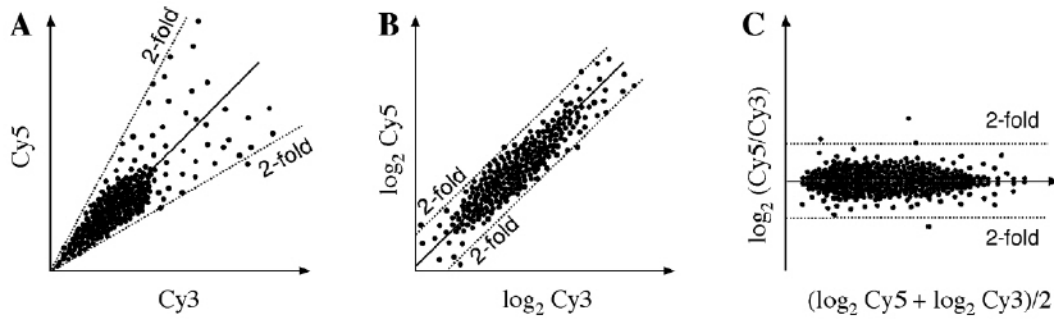


(e)

Şekil-2.7: (a) uzaysal grafik. (b) hatalı uzaysal grafik. (c) kutu grafiği. (d) volkan grafiği. (e) serpmme diyagramı.

2.2.4. Normalizasyon

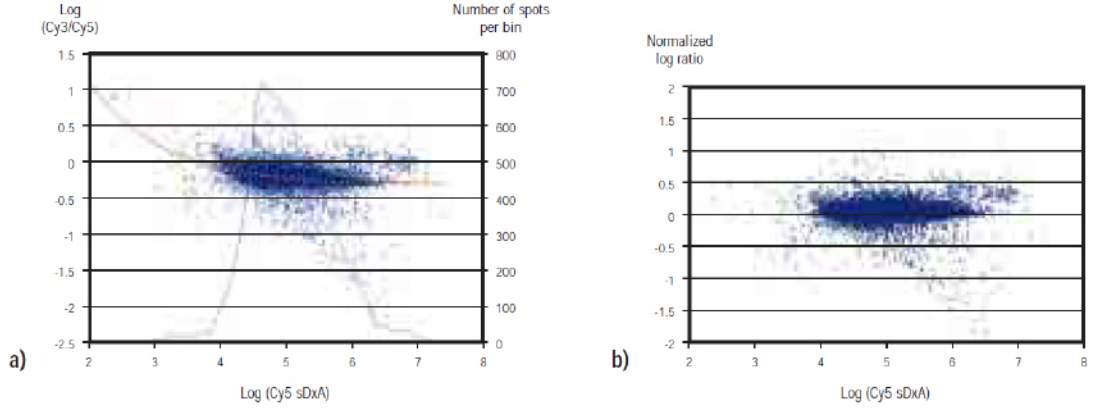
Normalizasyon aşamasının amacı, örnekler arasındaki biyolojik farklılıkların aksine teknik varyasyona göre verilerin ayarlanmasıdır. Arraylerin hibridizasyon süreçleri arasında her zaman az da olsa bir tutarsızlık olacaktır. Varyasyonlar çeşitli arraylerin genel floresan yoğunluk seviyeleri arasında ölçme farklılıklarına neden olabilmektedir. Normalizasyon işlemi farklı arraylerden elde edilen ölçüm değerlerini karşılaştırılabilir hale getirmek için gereklidir (4). Normalizasyon yöntemleri kullanılan DNA mikroarray teknolojilerine göre farklılık göstermektedir. Genel olarak, ileriki analizler için logaritmik dönüşüm uygulanmış veriler kullanılmaktadır. Normalizasyon fonksiyonlarının çoğu verileri bu formata getirmektedir (14).



Şekil-2.8: Veri normalizasyon sürecini gösteren gen ekspresyon analizinin serpmeye diyagramı (4).

En yaygın olarak kullanılan normalizasyon yöntemleri aşağıdaki gibidir (4);

1. Scaling Normalizasyon Yöntemi
2. Lineer Olmayan Normalizasyon Yöntemleri
3. Çeyreklik Normalizasyonu
4. Cyclic Loess Normalizasyonu
5. Kontrast Normalizasyonu



Şekil-2.9: Mikroarray verilere lineer olmayan normalizasyon uygulanması. (a) Bir mikroarray deneyinden elde edilmiş Cy5 sinyal değerlerine karşı $\log(\text{Cy3}/\text{Cy5})$ değerlerinin dağılımı. (b) Aynı verilerin lineer olmayan normalizasyon algoritması kullanıldıktan sonraki halini göstermektedir (14)

2.2.4.1. Scaling Normalizasyon Yöntemi

Scaling Normalizasyon Yöntemi, referans bir array seçerek diğer tüm arrayleri aynı ortalama yoğunluğa sahip olacak şekilde ölçeklemektedir. Yapılan işlem referans bir array seçip, sonrasında sabit terim olmadan her bir array ve seçilen array arasında bir lineer regresyon oluşturmaya denktir. Daha sonrasında regresyon çizgisi normalizasyon ilişkisi olarak kullanılmaktadır (4).

2.2.4.2. Lineer Olmayan Normalizasyon Yöntemleri

Lineer Olmayan Normalizasyon Yöntemleri, arrayler arasındaki lineer olmayan düzeltmeleri gerçekleştirmektedir. Bu yöntemler genellikle scaling yöntemi gibi lineer düzeltmelerden daha iyi performans göstermektedir (4).

Literatürde birçok lineer olmayan normalizasyon yöntemi vardır. Workman ve arkadaşları (14), array sinyal dağılım analizi ve kübik eğri cetvelleri kullanarak lineer olmayan bir normalizasyon yöntemi önermişlerdir. Chen ve arkadaşları (14), lokasyon yanlılıkları için düzeltmenin sağlanabileceği alt küme normalizasyonunu sunmuşlardır. Edwards (14), tek kanal cDNA mikroarraylerinde uzaysal heterojenliği düzeltebilecek lineer olmayan LOWESS normalizasyon yöntemini önermiştir.

2.2.4.3. Çeyreklik Normalizasyonu

Çeyreklik normalizasyonunun amacı her bir arraye aynı deneysel yoğunluk dağılımını empoze etmektir. Eğer iki veri vektörü aynı dağılıma sahipse, bir Q-Q grafiği 1 eğim ve 0 kesen ile düz bir diyagonal çizgiye sahip olmaktadır (4).

İki veri vektörünün birbirlerine karşı çeyrekliklerini çizmek ve her bir noktayı 45 derecelik diyagonal çizgi üzerine tasarlamak, her iki veri vektörünün aynı dağılıma sahip olmasını sağlayan bir dönüşüme neden olmaktadır (4).

2.2.4.4. Cyclic Loess Normalizasyonu

Cyclic loess yöntemi, Cy5 ve Cy3 kanal yoğunluklarının MA grafikleri kullanılarak normalleştirildiği küresel loess yönteminin bir genellemesidir. Tek kanal array verileri ile ilgilenirken, array çiftleri birbirlerine göre normalleştirilmektedir. Cyclic Loess yöntemi, ikili şekilde çalışarak bir array seti için yoğunlukları normalleştirilmektedir (4).

2.2.4.5. Kontrast Normalizasyonu

Kontrast Normalizasyonunda, veriler bir kontrast setine dönüştürülmekte ve lineer olmayan bir MA-grafiği normalizasyonu gerçekleştirilmektedir. Daha sonrasında ters dönüşüm uygulanmaktadır (4).

2.3. Farklı Gen Ekspresyon Analizinde Kullanılan İstatistiksel Yöntemler

Gen ekspresyon verilerini analiz etmedeki esas, ekspresyon modelleri fenotip veya deneysel koşula göre farklılık gösteren genleri saptama ihtiyacıdır. Gen ekspresyonu iyi koordine edilmiş bir sistemdir ve bu yüzden farklı genlerdeki ölçümler genel olarak bağımsız değildir. Belirli etkileşim ve transkripsiyonel kontroller dikkate alındığında, örnekler arasındaki anlamlı karşılaştırmalar belirli gen setlerinin ortak dağılımı göz önünde tutularak yapılabilmektedir (4).

Genleri seçmek için basit bir yaklaşım “fold change” kriterini kullanmaktır. Bu yalnızca hiç tekrarlanmanın olmadığı ya da çok az olduğu durumlarda olasıdır. Ancak, yalnızca fold change temeline dayanan bir analiz, genden gene değişiklik gösterebilen biyolojik ve deneysel varyasyonların varlığında ekspresyon farklılıklarının anlamlılığının değerlendirilmesine olanak sağlamamaktadır. Bu durum, farklı ekspresyonların değerlendirilmesi için istatistiksel testlerin kullanılmasının ana sebebidir. Parametrik testler genellikle, altta yatan model varsayımları karşılanıyorsa, daha yüksek bir güce sahiptirler (9).

Mikroarray verilerin istatistiksel analizini yaparken, önemli bir soru da verilerin hangi ölçek üzerinde analiz edileceğini belirlemektir. Genellikle, her bir gen için tekrarlanmış ölçümlerin dağılımını yaklaşık olarak simetrik ve normale yakın hale getirebilmek için logaritmik ölçek kullanılmaktadır. Mikroarray ölçümleri için bir hata modelinden türetilen varyans dengeleyici dönüşüm, ölçülen yoğunlukların varyansını bunların beklenen değerinden bağımsız hale getirmek için kullanılabilir. Bu varyans homojenliğine dayanan gen odaklı istatistiksel testler için avantajlı olabilir. Çünkü bu yoğunluk düzeyindeki farklılıklardan dolayı ortaya çıkan deneysel koşullar arasındaki varyanstaki farklılıkları azaltacaktır. Ancak tabii ki de koşullar arasındaki varyanstaki farklılıkların gene-özel biyolojik sebepleri de olabileceği unutulmamalıdır (4).

Bir veya iki grup için t-test karşılaştırmaları, çoklu gruplar için varyans analizi ve daha genel trend testleri farklı gen ekspresyonunu değerlendirmek için sıklıkla kullanılan modellerdir. Parametrik bir yöntem olarak, lineer modelleme yukarıda tartışılan uyarılara bağlıdır. Fakat model parametrelerinin uygun yorumlanabilirliği sıklıkla bunu mikroarray analizi için seçilen yöntem yapmaktadır. Genlerin ortak düzenlenmesine ilişkin bilgi eksikliği nedeniyle, lineer modeller genellikle her bir gen için ayrı ayrı hesaplanır. İlgili genlerin listeleri tespit edildiğinde, araştırmacılar bunların ortak hareketlerinin daha ileri modellemesi için koordineli düzenleme çalışmalarına başlayabilirler (4).

Her bir gen için bir istatistiksel test yapma yaklaşımı yaygındır. Ancak bu yaklaşımın bir takım zorlukları vardır. En önemlisi çok fazla sayıda hipotez testinin uygulanıyor olmasıdır. Bu durum potansiyel olarak çok sayıda yanlış pozitif sonuca yol açacaktır. Çoklu test prosedürleri, bir hipotez testleri grubunun sonuçlarının bütün anlamlılığını değerlendirmeye olanak sağlamaktadır. Bunlar deneysel hata oranı veya yanlış bulgu oranı (false discovery rate) gibi tip I (yanlış pozitif) hata oranlarını kontrol ederek tanımlayıcılığa odaklanmaktadır. Bu kontroller çoklu karşılaştırmalar için düzeltme yapma amacıyla çoklu hipotez testlerinde kullanılan istatistiksel yöntemlerdir. Yine de çoklu hipotez testi yapmak bir sorun olmaya devam etmektedir. Çünkü tanımlayıcılıktaki bir artış, p değeri düzeltme yöntemleri tarafından sağlandığı gibi, bir duyarlılık kaybı ile de ilgilidir. Dolayısıyla doğru pozitifliği tespit etme şansı azalmaktadır (4).

Ayrıca, ekspresyondaki en etkili deęişimlere sahip olan genler mutlaka ilgili biyolojik süreçlerde büyük rol oynayan genler olacak deęildir. Bu sorunun üzerine yalnızca önceki biyolojik bilgileri, mikroarray verilerin analizine dahil ederek gidilebilir. Bu çözüm, analizi belirli bir gen seti üzerine odaklamaya yol açabilmektedir. Eđer böyle bir biyolojik ön eleme uygulanamazsa, test edilecek hipotezlerin sayısı spesifik olmayan filtreleme işlemleri ile makul derecede azaltılabilir. Örneğin; örnekler arasında sürekli olarak düşük yoğunluk deęerlerine sahip olan genler veya düşük varyansa sahip olan genler çıkartılabilir (4).

Çoęu mikroarray deneyi her bir durum için yalnızca birkaç tekrarlama içermektedir, bu gene-özgü varyansları tahmin etmeyi zorlaştırmaktadır. Tüm genlerin verilerinden sağlanan varyans bilgilerinden yararlanmak için farklı yöntemler geliştirilmiştir (4).

2.3.1. Fold Change Kriteri

Basit bir mikroarray deneyi iki koşul arasındaki ekspresyon farklılıklarını tespit etmek için gerçekleştirilmektedir. Her bir koşul bir veya daha fazla RNA örnekleme tarafından temsil edilebilmektedir. Genlerin ekspresyon farklılıklarını test etmek için kullanılan yöntemlerden en temeli “Fold Change” kriteridir (15).

Yöntem, iki koşulun ekspresyon düzeyleri arasındaki log oranını hesaplamaktadır. Daha sonrasında tüm genlerin farklı ekspresyon için belirlenmiş bir eşik deęerinden büyük olup olmadığı deęerlendirmektedir. Fold change yöntemi istatistiksel deęişkenliği hesaba katmadığı için güvenilir bir yöntem olarak nitelendirilmemektedir. Veriler uygun bir şekilde normalleştirilmediğinde fold change kriteri yanlılıęa eğilimli olmaktadır (15).

x_i ve y_i sırasıyla i geni için kontrol ve tedavi gruplarındaki ekspresyon seviyeleri ise, i geni için fold change aşağıdaki gibidir (15);

$$FC = \frac{\bar{x}_i}{\bar{y}_i} \quad (2.7)$$

2.3.2. Student t-Test

Student t-testi, eş varyans olması durumunda, farklı ifade edilen genlerin tespit edilmesinde kullanılan en temel istatistiksel yöntemlerden biridir. Genler birbirinden bağımsız ve gözlemler normal dağılıyor ise Student t-dağılımı kullanılmalıdır (15).

x_1 ve x_2 iki koşul altında birbirinden bağımsız iki gen ekspresyon verisi olsun; örneğin normal ve hastalıklı, belirli bir gen için iki örneklem t istatistiği eşitlik (2.8)'deki gibidir;

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad (2.8)$$

Burada s , pooled standart sapma, n_1 ve n_2 ise her bir koşuldaki örneklem ölçümüdür. Pooled varyans, s^2 eşitlik (2.9)'daki gibi hesaplanmaktadır;

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \quad (2.9)$$

Burada s_1^2 ve s_2^2 her bir koşul altındaki tekrarlanmış gözlemlerin varyanslarıdır.

Hesaplanan test istatistiği, $n_1 + n_2 - 2$ serbestlik derecesi ile birlikte kritik değer ile karşılaştırılmaktadır. t-testi örneklemelerin varyanslarından faydalandığından, dolayısıyla fold change yaklaşımının eksikliklerine de dikkat çekmektedir (15).

2.3.3. Satterthwaite-Welch t-Testi

Varyansların homojenliği mikroarray deneylerinde nadir rastlanan bir durumdur. Mikroarray deneylerinde heterojen denek örneklemeleri veya hücre örneklemeleri nedeniyle heterojen varyanslılık ortaya çıkabilmektedir. Transkripsiyon faktörünün ekspresyon değişiminin korelasyon yapısını değiştirmesi de heterojen varyansa neden olabilmektedir. Dolayısıyla, farklı varyanslara sahip bağımsız örneklemeler için Welch (Satterthwaite's) t-test yöntemi çok daha uygun olacaktır (16).

x_1 ve x_2 iki koşul altında birbirinden bağımsız iki gen ekspresyon verisi iken Welch t -istatistiği belirli bir gen için aşağıdaki gibi hesaplanmaktadır;

$$t_w = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad (2.10)$$

Burada \bar{x}_1 ve \bar{x}_2 belirli bir genin 1 ve 2 koşulları altındaki ekspresyon değerlerinin örneklem ortalamalarını, n_1 ve n_2 örneklem ölçümlerini, s_1^2 ve s_2^2 ise iki örneklemin varyanslarını ifade etmektedir (16).

Welch yöntemi farklı örneklemlerin varyansları altında serbestlik derecesi için özel bir düzeltme ye sahiptir;

$$\mathcal{V} = \frac{(\omega_1 + \omega_2)^2}{\omega_1^2/(n_1 - 1) + \omega_2^2/(n_2 - 1)} \quad (2.11)$$

ω_i niceliği, i örneklemini için tahmin edilen karesi alınmış standart hatadır;

$$\omega_i = \frac{s_i^2}{n_i} \quad (2.12)$$

2.3.4. Bağımlı Örneklem t -Testi

Bağımlı tasarım, genellikle deneysel birimlerin doğal bir eşleşmesini içeren gen ekspresyon karşılaştırmalarının yapıldığı iki-kanallı deneylerde kullanılmaktadır. Student t -testi ve Welch t -testi bağımsız örneklemler için kullanılmaktadır. Örneklem bağımlı olduğunda bağımlı örneklem t -testi kullanılmalıdır (17).

\bar{x}_j ve s_j , sırasıyla ortalama fark ve farkların standart sapması iken, $n-1$ serbestlik derecesi ile, bağımlı örneklem test istatistiği eşitlik (2.13)'de verilmiştir;

$$t = \frac{\bar{x}_d}{s_d/\sqrt{n}} \quad (2.13)$$

2.3.5. Moderated t -Testi

Günümüzde moderated t -testi mikroarray çalışmalarında kullanılan en yaygın yöntemlerden biridir. Değişiklik, t -istatistiği hesaplanırken standart sapmaya, t -değerinin değişkenliğini azaltacak küçük bir değer eklenmesidir. Moderated t -testinin asıl amacı yanlış pozitifleri önlemek amacıyla küçük bir standart sapma ile genlerin istatistiksel anlamlılığını azaltmaktır (15).

s_0 , T'_i değişkenliğini azaltmak için seçilmiş bir sabit iken moderated t -istatistiği aşağıdaki gibi tanımlanmaktadır;

$$T'_i = \frac{\bar{x}_i - \bar{y}_i}{s_i + s_0} \quad (2.14)$$

İki gen yaklaşık olarak eşit istatistiksel anlamlılığa sahip olabilir, ancak ileriki çalışmaya aday olmak için eşit ilgiye sahip olmayabilir. Eğer eşit anlamlılığın çeşitli ekspresyon profilleri varsa, genel anlamda daha fazla dinamik açıklık gösteren bu profillerle ilgilenilir. t istatistiği istatistiksel anlamda çok güçlüdür. Fakat amaç kısa bir sıralanmış liste oluşturmak olabilir. Bu noktada Moderated t -istatistiği sıralama performansını arttırmaktadır (15).

2.3.6. Wilcoxon Rank Sum Testi

Wilcoxon rank sum testi, x_i ekspresyon değerleri normal dağılmadığında, iki grup için gen ekspresyonların karşılaştırılmasında kullanılan parametrik olmayan bir yöntemdir. Örneklemelerin karşılaştırılmasında elde edilen değerlerin yerine gözlemlerin sıralamaları kullanılır. Bu işlem, Wilcoxon rank sum testini aykırı değerlere çok duyarlı olmaması bakımından, t -istatistiğinden daha dayanıklı yapmaktadır (16).

$Rank(x_{g1k})$ x_{g1k} 'in tüm x_{g1k} ve x_{g2k} değerleri arasındaki sıralaması iken belirli bir g geni için Wilcoxon rank sum istatistiği aşağıdaki gibi hesaplanmaktadır;

$$z_g = \sum_k Rank(x_{g1k}) \quad (2.15)$$

Sıralama en küçük değer için 1'den başlamaktadır ve en yüksek değer $n_{g1} + n_{g2}$ olmaktadır. Eğer tüm x_{g1k} değerleri x_{g2k} değerlerinden küçükse, $z_g, \sum_{k=1}^{n_{g1}} k$ olacaktır. Eğer tüm x_{g1k} değerleri x_{g2k} değerlerinden büyükse, $z_g, \sum_{k=n_{g2}+1}^{n_{g1}+n_{g2}} k$ olacaktır. Diğer tüm olasılıklar ise bu iki değer arasında olmaktadır (16).

2.3.7. ANOVA F-Testi

ANOVA, mikroarray veri analizlerinde gen ekspresyonunu etkileyebilecek faktörlerin etkilerinin anlamlılığını araştırmak için kullanılmaktadır. Mikroarray veriler için ANOVA modeli iki aşamada belirlenebilmektedir (18). İlk aşama normalizasyon modelidir;

$$y_{ijgr} = \mu + A_i + D_j + AD_{ij} + r_{ijgr} \quad (2.16)$$

y_{ijgr} , bir sinyal yoğunluğunun logaritmasıdır. (i) indisi array'i, (j) indisi boyayı, (g) geni ve (r) ise ölçümü ifade etmektedir. μ , genel ortalama ekspresyon seviyesidir. A ölçülen yoğunlukta array'in etkisi; D boyanın etkisi; AD ise boya ve array arasındaki etkileşimin etkilerini göstermektedir. İlk aşama ölçülen yoğunluklardan r_{ijgr} terimini oluşturmaktadır (18). İkinci aşamada normalizasyon yönteminin artıkları açısından gene-özgü etkiler modellenmektedir (18). Gene-özgü model (1.12)'de ifade edilmiştir;

$$r_{ijgr} = G + VG_{ij} + DG_j + AG_i + \epsilon_{ijr} \quad (2.17)$$

G belirli bir gene ilişkin ortalama yoğunluk; AG_i arrayin bu gen üzerindeki etkisi; DG_j boyanın bu gen üzerindeki etkisi ve ϵ_{ijr} artık değeridir. VG (variety-by-gene) terimi analizdeki esas ilgilenilen terimdir ve belirli bir genin örneklem arasındaki ekspresyon seviyelerindeki değişkenlikleri yansıtmaktadır. VG_{ij} örneklemle ilişkin etkiler için bir "catch-all" terimi olarak ifade edilmektedir. En basit durumda, VG_{ij} boyası ile i arrayi üzerindeki örneklem tarafından temsil edilen koşulun bir göstergesidir. Daha karmaşık deneylerde, biyolojik örneklem seviyesindeki tasarım yapısı VG terimlerinde yansıtılmaktadır (18).

Bir array içerisinde kopyalanmış spotlar olduğunda, model etiketleme ve spot etkileri için ek terimleri içermelidir. Gene-özgü model tek-renkli veriler için DG ve AG terimleri kaldırılarak değiştirilebilmektedir (18).

Hipotez testi iki modelin karşılaştırılmasını içermektedir. Sıfır hipotezi farklı ekspresyonun olmadığını (VG değerlerinin sıfıra eşit olduğunu), alternatif hipotez ise koşullar arasında farklı ekspresyon olduğunu (VG değerlerinin sıfıra eşit olmadığını) ileri sürmektedir (18). F istatistikleri artık kareler toplamları ile aşağıdaki gibi hesaplanmaktadır;

$$F1 = \frac{(AKT_0 - AKT_1)/(v_0 - v_1)}{AKT_1/v_1} \quad (2.18)$$

$$F2 = \frac{(AKT_0 - AKT_1)/(v_0 - v_1)}{\sigma_{pool}^2} \quad (2.19)$$

$$F1 = \frac{(AKT_0 - AKT_1)/(v_0 - v_1)}{(AKT_1/v_1 + \sigma_{pool}^2)/2} \quad (2.20)$$

AKT_0 , v_0 ve AKT_1 , v_1 sırasıyla sıfır ve alternatif modeller için artık kareler toplamları ve serbestlik dereceleridir. σ_{pool}^2 tüm genler arasındaki pooled hata varyansıdır (18).

2.3.8. Moderated F Testi

Moderated t istatistikleri, doğal olarak karşılaştırmaların herhangi bir seti hakkındaki test hipotezleri için kullanılabilir F istatistiklerine yol açmaktadır. Moderated t istatistiklerinin uygun karesel formları F dağılımlarını izlemektedir (19).

Kontrast tahmincilerinin ortalamasının β_g olduğu durumda, belirli bir gen için tüm karşılaştırmaların test edilmek istendiği varsayılınsın. U_g diyagonal matris, C kontrast matrisi, V_g kesin artı matrisi iken $\hat{\beta}_g$ 'nin korelasyon matrisi $R_g = U_g^{-1}C^T V_g C U_g^{-1}$ olmaktadır. r , C'nin sütun sırası iken $Q_g^T R_g Q_g = I_r$ ve $q_g = Q_g^T t_g$ olmaktadır. F_g , aşağıdaki eşitlikle bulunmaktadır;

$$F_g = q_g^T q_g / r = t_g^T Q_g Q_g^T t_g / r \sim F_{r, d_0 + d_g} \quad (2.21)$$

Eğer Q_g 'nin sütunları R_g 'nin aralık uzayını kapsayan özvektörler olacak şekilde seçilirse diyagonal matris $Q_g^T R_g Q_g = \Lambda_g$ olmaktadır. F_g ise aşağıdaki gibi bulunmaktadır.

$$F_g = q_g^T \Lambda_g^{-1} q_g / r \sim F_{r, d_0 + d_g} \quad (2.22)$$

Moderated F istatistiđi, belirli bir gen için karřılařtırmalardan herhangi birinin sıfır olup olmadığını test etmektedir. Testin sonucunda genin herhangi bir karřılařtırmada farklı olarak ifade edilip edilmediđi bulunmaktadır. Çok fazla karřılařtırma ieren kompleks deneylerde, genlerin öncelikle moderated F istatistiklerine dayalı seilmesi tercih edilmektedir (19).

2.3.9. Çoklu Hipotez Testleri

m tane sıfır hipotezinin eř zamanlı olarak test edildiđi çoklu hipotez testlerinde, yanlış bulgular kaçınılmaz olduđundan farklı ekspresyonların deđerlendirilmesi bir sorun haline gelmektedir. Dolayısıyla çoklu test problemlerinde yanlış bulguların hata oranını kontrol etmek oldukça önemlidir (17). Çoklu test problemlerine yönelik standart yaklaşım iki durumdan meydana gelmektedir;

1. Her bir j geni için bir T_j test istatistiđini hesaplamak
2. Öncesinde uygun bir şekilde belirlenmiř bir Tip I hata oranını kontrol ederek hangi hipotezlerin reddedileceđini belirlemek için çoklu test prosedürlerini uygulamak (17).

2.3.9.1. Tip I ve Tip II Hata Oranları

Mikroarray deneylerinde herhangi bir test durumunda, iki tip hata yapılabilir: genin farklı ifade edilmediđini öne süren dođru sıfır hipotezinin reddedilmesiyle yapılan *Tip I hata* (Yanlış Pozitif) ve test prosedürü yanlış sıfır hipotezini reddetmede başarısız olduđunda yapılan *Tip II hata* (Yanlış Negatif). *Tip I hata* ve *Tip II hata* durumları Tablo-2.2’de özetlenmektedir (17).

Tablo-2.2 – Çoklu Hipotez testlerinde Tip I ve Tip II hatalar

	Kabul	Red	
Dođru Sıfır Hipotezi	$h_0 - V_n$	V_n	h_0
Yanlış Sıfır Hipotezi	U_n	$h_1 - U_n$	h_1
	$m - R_n$	R_n	m

V_n , Tip I hataların sayısını; U_n , Tip II hataların sayısını göstermektedir. Doğru ve yanlış hipotezlerin sayısı h_0 ve h_1 bilinmeyen parametrelerdir. Belirli bir test uygulanarak reddedilen hipotezlerin sayısı R_n gözlenebilir bir rastgele değişkendir. U_n , $h_1 - U_n$ ve V_n ise gözlenemeyen rastgele değişkenlerdir (17).

Çoklu testlerde yaygın olarak kullanılan Tip I hata oranları aşağıdaki gibidir;

- Karşılaştırma başına hata oranı (Per comparison error rate, PCER): Hipotez sayısına karşı Tip I hataların sayısının beklenen değeri,

$$PCER = \frac{E(V_n)}{m} \quad (2.23)$$

- Deneysel hata oranı (Family-wise error rate, FWER): En az bir Tip I hata olma olasılığı,

$$FWER = Pr(V_n = 1) \quad (2.24)$$

- Yanlış bulgu oranı (False discovery rate, FDR): Reddedilen hipotezler arasında Tip I hataların beklenen oranı,

$$FDR = E\left(\frac{V_n}{R_n}; R_n > 0\right) = E\left(\frac{V_n}{R_n} \middle| R_n > 0\right) \times Pr(R_n > 0) \quad (2.25)$$

2.3.9.2. Yanlış Bulgu Oranı (False Discovery Rate)

Literatürde hata oranı ile ilgili tartışmaların büyük çoğunluğu deneysel ve karşılaştırma başına hata oranları ile ilgilidir. Son yıllarda, Benjamini ve Hochberg (20) tarafından önerilen yanlış bulgu oranı (YBO) çok yaygın hale gelmiştir. YBO, yazarlar tarafından hatalı retlerin toplam retlere oranının beklenen değeri olarak tanımlanmıştır. Benjamini ve Hochberg (20), YBO'nun kontrol edilmesinin deneysel veya karşılaştırma başına hatanın kontrol edilmesinden daha makul olduğunu belirtmişlerdir.

$\mu_1, \mu_2, \dots, \mu_J$, J tane ortalama ise, istenilen $m=J(J-1)/2$ ikili hipotezler ailesinin testidir. U_n , $m - R_n$ adet ret kümesinden doğru biçimde reddedilen hipotezlerin sayısıdır. $h_0 - V_n$ yanlış biçimde reddedilen çift sayısıdır (20).

Benjamini ve Hochberg (20), yanlış reddedilen sıfır hipotezlerin Q raslantı değişkeni ile ifade edilebileceğini belirtmişlerdir.

$$Q = \frac{h_0 - V_n}{(h_0 - V_n) + V_n} \quad (2.26)$$

$m - R_n = 0$ iken Q sıfır olarak tanımlanmaktadır. Bunun anlamı reddedilme olmadığında hata oranı sıfırdır. Benjamini ve Hochberg (20) tarafından YBO, Q'nun ortalaması olarak tanımlanmıştır.

Buna bağlı olarak;

$$\text{Karşılaştırma Başına Hata Oranı} = E\left(\frac{h_0 - V_n}{m}\right) \quad (2.27)$$

$$P(h_0 - V_n \geq 1) = \text{Deneysel Hata Oranı} \quad (2.28)$$

$$E(Q) = E\left(\frac{h_0 - V_n}{(h_0 - V_n) + U_n}\right) = E\left(\frac{h_0 - V_n}{m - R_n}\right) = E\left(\frac{\text{Yanlış Ret Sayısı}}{\text{Toplam Ret Sayısı}}\right) \quad (2.29)$$

3. GEREÇ VE YÖNTEM

3.1. Analizde kullanılacak Veri Seti

Veriler, Affymetrix HGU95av2 arrayleri kullanılarak MCF7 göğüs kanseri hücreleri üzerinde yapılan 2×2 'lik bir faktöryel deneyin sonuçlarını içermektedir (Wolfgang Huber, Robert Gentleman, 2006) (21). İlgili veriler, Bioconductor.org kaynağından alınmıştır (21). Deneydeki faktörler östrojen (var ya da yok) ve maruziyet süresidir (10 ya da 48 saat). Veriler, östrojen reseptör pozitif (ER+) göğüs kanseri hücrelerindeki gen ekspresyon ölçümlerini içermektedir. Bütün veri seti 12,625 prob setini içermektedir.

Yapılan deneyde tüm sekiz örneklemin serum starvasyonundan sonra, dört örneklem östrojene maruz bırakılmıştır. Daha sonrasında bu örneklemelerin ikisinden 10 saat sonra, diğer ikisinden ise 48 saat sonra mRNA transkripsiyon miktarları ölçülmüştür. Kontrol olarak adlandırılan diğer dört örneklem ise östrojene maruz bırakılmamıştır. Kontrol örneklemelerinin ikisinden 10 saat sonra, diğer ikisinden ise 48 saat sonra mRNA transkripsiyon miktarları ölçülmüştür. Tablo-3.1 ilgili örneklemeleri ve deney tasarımı göstermektedir.

Tablo-3.1: Çalışmanın Deney Tasarımı

Zaman	Östrojen	
	Yok	Var
10 Saat	low10-1.cel	high10-1.cel
	low10-2.cel	high10-2.cel
48 Saat	low48-1.cel	high48-1.cel
	low48-2.cel	high48-2.cel

Dört deneysel koşul altında mRNA örneklemeleri elde edilmiştir ve mRNA'lar ters transkriptaz enzimi ile cDNA'ya çevrilmiştir. Sonrasında örneklemeler etiketlenmiş ve cRNA'ya parçalanmıştır. Son olarak etiketlenen örneklemeler Affymetrix Gen çip arraylerine melezleştirilmiştir. İlgili genler prob setleri kullanılarak arrayler üzerinde temsil edilmektedir.

Arraylerin görüntüsünü elde etmek için bir tarayıcı kullanılmıştır. Her bir spotun yoğunluğu o spotta ne kadar bağlanma olduğunu göstermektedir. Bir CEL dosyası, her bir spot için işlenmiş yoğunluk değerlerini içermektedir ve taranmış görüntünün nicel bir özeti olarak ifade edilebilir. CEL dosyaları ham verileri ifade etmektedir.

3.2. Kullanılacak İstatistiksel Yöntemler

İstatistiksel analiz, ham veriler olarak tanımlanan CEL dosyaları kullanılarak RStudio yazılım programı aracılığıyla başlamıştır. Analiz başlamadan önce ilk adım olarak ham verilerin kalite değerlendirmesi yapılmıştır. Kalite kontrol, ölçek faktörlerinin ve her bir array için ortalama arka planın değerlendirilmesini, GAPDH ve beta-actin 3'/5' oranları ile array üzerindeki gen sayılarının gözlenmesini sağlayan *simpleaffy* paketi içerisindeki *qc* fonksiyonları ile başlamıştır. Arraylerin homojenliğini değerlendirmek ve arrayler arasındaki yoğunluk dağılımını karşılaştırmak için ham verilerin logaritma2 tabanındaki yoğunlukları kullanılarak her bir arraye ilişkin kutu grafikleri çizilmiştir. Her bir array üzerindeki yoğunlukların boyutsal dağılımlarını gözlemlemek ve boyutsal bozuklukları tespit etmek için CEL dosyalarının görüntüleri elde edilmiştir. Arraylerin karşılaştırılması için tekrarlama arrayleri göz ardı edilerek dört array için mümkün tüm karşılaştırmaların MA grafikleri çizilmiştir. MA grafiklerinin çizilmesinde logaritmik oranların (M) ve logaritmik yoğunlukların hesaplanması için Perfect Match (PM) problemleri kullanılmıştır. MA grafikleri ortalama logaritmik yoğunluk değerleri A'lara karşı logaritmik oranlar M'lerin çizilmesi ile oluşturulmuştur. Normalizasyon öncesi kalite kontrol aşaması her bir arrayin yoğunluk dağılımlarının değerlendirilmesi için çizilen histogramlar ile tamamlanmıştır.

Ham verilerin kalite kontrolünden sonra, RMA (Robust Multiple-Array Average) yöntemi kullanılarak verilere arka plan düzeltmesi ve normalizasyon uygulanmıştır. RMA yöntemi ile prob-seviyesindeki sinyal arka plan sinyalinden ayrılmıştır. RMA yöntemi ile çeyreklik normalizasyonu uygulanmıştır ve tüm arraylerin aynı çeyrekliğe sahip olması sağlanmıştır. RMA yöntemi kullanılarak normalleştirilen ve arka plan düzeltmesi yapılan yoğunluklar ile analizde kullanılacak ekspresyon seti oluşturulmuştur. Arka plan düzeltmesi ve normalizasyonun uygulanmasından sonra kalite kontrolün yeniden değerlendirilmesi için her bir arraye ilişkin kutu grafikleri çizilmiştir.

Deneyde dört çift tekrarlanmış array bulunmaktadır, böylece lineer modelde dört parametre bulunacaktır. Sekiz array ve 4 deneysel koşul için deneysel tasarımı açıklayacak dizayn matrisi oluşturulmuştur. Dizayn matrisinin ve lineer modelin oluşturulmasından sonra dört tedavi kombinasyonu arasındaki ilgili karşılaştırmaları içeren kontrast matrisi oluşturulmuştur. Kontrastlar lineer modelin parametrelerinin lineer kombinasyonları olarak oluşturulmuştur. Her bir gen için kontrast matrisindeki karşılaştırmaların anlamlılığını, moderated t-istatistiği kullanılarak hesaplanmıştır. Karşılaştırmalar için moderated t-istatistikleri eBayes fonksiyonu ile elde edilmektedir. Farklı ekspresyonlara sahip genler log-odds'larına göre sıralanmıştır. Her bir karşılaştırma için moderated t istatistiğinin yanında kontrast katsayıları, ortalama ekspresyon değerleri hesaplanmıştır. Kontrast katsayıları (M değerleri) iki deneysel koşul arasındaki log₂ tabanındaki fold change hesaplanarak elde edilmiştir. Ortalama ekspresyon değerleri (A değerleri) tüm arrayler arasındaki bir gen için ortalama logaritma iki tabanındaki ekspresyon seviyesini göstermektedir.

Çalışmanın amacı ER+ göğüs kanseri hücrelerindeki östrojen faktörüne yanıt veren genleri tespit etmektir. Östrojen faktörünün yanında zaman faktörü de değerlendirmeye alınmıştır ve östrojene erken yanıt veren ve geç yanıt veren genler tespit edilmiştir. Erken yanıt veren genlerin doğrudan-hedef genler kabul edilmektedir. Geç yanıt veren genler büyük bir olasılıkla moleküler yolaktaki akış aşağı hedeflerdir. Genler için herhangi bir karşılaştırmanın sıfırdan farklı olup olmadığını test etmek amacıyla moderated F-istatistiği kullanılmıştır. Çoklu sınama için p değerleri düzeltilmiştir. Yanlış bulgu oranını kontrol etmek için Benjamini ve Hochberg yöntemi kullanılarak p değerleri düzeltilmiştir.

İstatistiksel analiz, RStudio programı ile yapılmıştır. Analizin yapılmasında kullanılan veri setinin alınması için "estrogen" paketi kullanılmıştır. Mikroarray verilerin gen ekspresyon analizi ve lineer modellerin oluşturulması için "limma" paketi kullanılmıştır. Analizin uygulanmasında limma paketi ile tüm genlerin varyanslarına dayalı hesaplanan küresel varyans tahmin edicisini kullanan, Denemeli Bayes (Empirical Bayes) yaklaşımı uygulanmıştır.

Limma paketi, özellikle tasarlanmış deneylerin analizi için ve farklı ekspresyonun değerlendirilmesinde lineer modelleri kullanmaktadır. Grafiklerin çizilmesi ve kalite kontrolün değerlendirilmesi için “affy” ve “marray” paketleri kullanılmıştır. RMA normalizasyonu “affy” paketi ile yapılmıştır. qc fonksiyonu ile kalite kontrol değerlendirmesinin yapılmasında “Simpleaffy” paketi kullanılmıştır.

4.BULGULAR

4.1. Kalite Kontrol Sonuçları

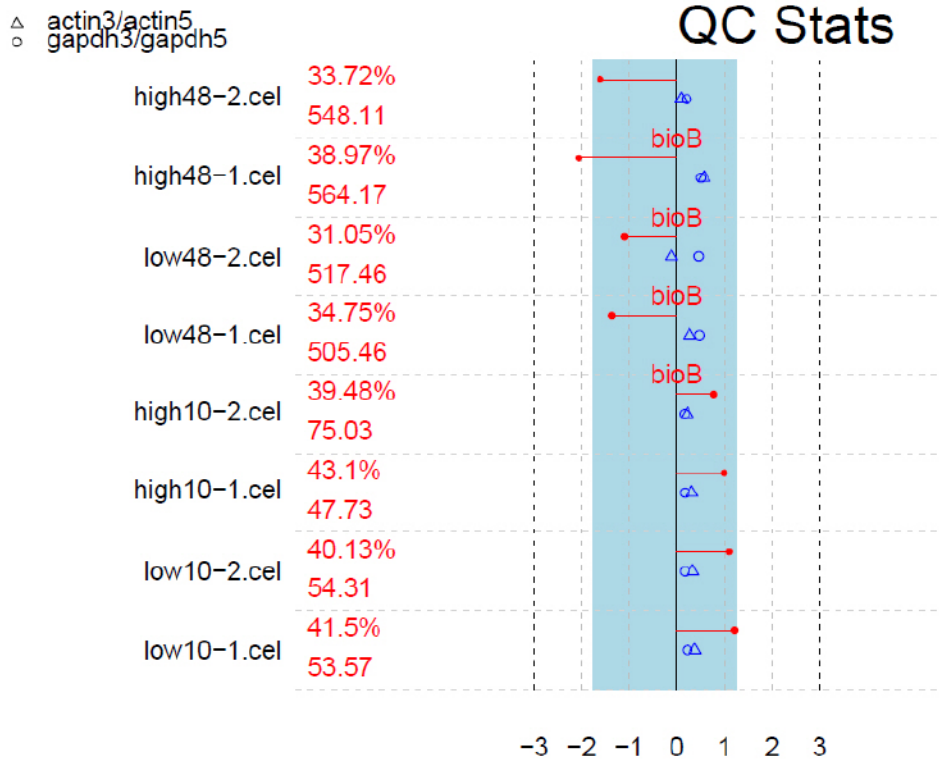
Şekil-4.1, ölçek faktörlerinin ve her bir array için ortalama arka planın değerlendirilmesini, GAPDH ve beta-actin 3'/5' oranları ile array üzerindeki gen sayılarının gözlenmesini sağlayan *simpleaffy* paketi içerisindeki *qc* fonksiyonları ile elde edilen QC grafiğini göstermektedir. Her bir array şekil üzerinde ayrı bir çizgi ile temsil edilmektedir. Dik merkez çizgisi 0 fold change'e karşılık gelmektedir. GAPDH oranları yuvarlaklar, beta-actin 3'/5' oranları ise üçgenler ile gösterilmiştir. Affymetrix, beta-actin'in 3'ün altında, GAPDH'nin ise 1'in etrafında olması gerektiğini bildirmektedir. beta-actin ve GAPDH oranları söz konusu eşik değerleri içerisinde gözlenmiştir. Ölçek faktörleri merkezden çizilen noktalı çizgiler ile gösterilmiştir. Mavi bölge içerisinde bulunan çizgiler ölçek faktörlerinin uyumlu olduğunu göstermektedir. Elde edilen sonuçlara göre ikinci array için noktalı çizgi mavi bölgenin dışında gözlenmiştir. Grafiğin solunda bulunan yüzdeler ortalama arka planı, sayılar ise ilgili array üzerindeki gen sayısını ifade etmektedir. Her bir array üzerindeki gen sayıları ve ortalama arka plan yüzdeleri birbirine yakın olmalıdır. Önemli derecede farklı olan değerler grafik üzerinde kırmızı ile gösterilmektedir. Gözlenen değerler arrayler üzerindeki ortalama arka planın ve gen sayılarının önemli derecede farklı olduğunu göstermektedir.

Şekil-4.2 arraylerin homojenliğini değerlendirmek ve arrayler arasındaki yoğunluk dağılımını karşılaştırmak için ham verilerin logaritma2 tabanındaki yoğunlukları kullanılarak her bir arraye ilişkin çizilen kutu grafiklerini göstermektedir. Her bir kutu bir arrayi temsil etmektedir. 8 array için çizilen kutu grafiklerinin genişliklerinin ve ortalamalarının birbirinden farklı olduğu gözlenmiştir. Bu durumda arraylerin homojen olmadığı söylenebilir.

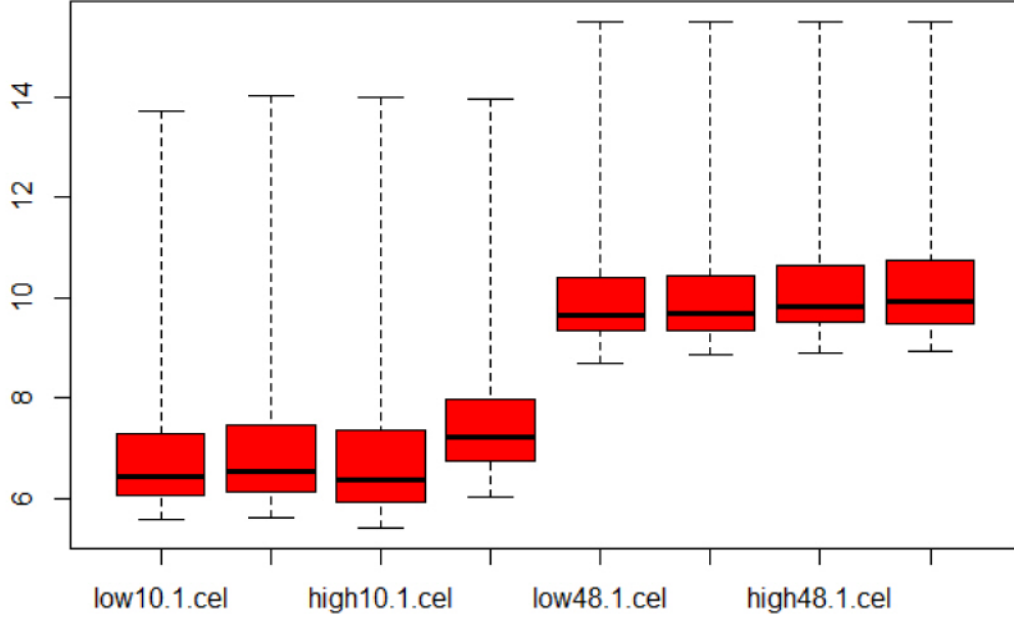
Şekil-4.3, arraylerin karşılaştırılması için tekrarlama arrayleri göz ardı edilerek dört array için mümkün tüm karşılaştırmaların MA grafiklerini göstermektedir. 8 array için mümkün 28 karşılaştırma söz konusudur. Mümkün tüm karşılaştırmalar yerine tekrarlama arrayleri göz ardı edilerek yalnızca low10-1.cel, high10-1.cel, low48-1.cel ve high48-1.cel arrayleri için MA grafikleri çizilmiştir. log oranları (M) ve log yoğunluklarının (A)

hesaplanması için PM (Perfect Match) problemleri kullanılmıştır. MA grafikleri anlamlı bir zaman etkisinin olduğunu göstermektedir. 10 saat ve 48 saat arasındaki karşılaştırma için çizilen MA grafiği $M=0$ 'da merkezlenmemektedir.

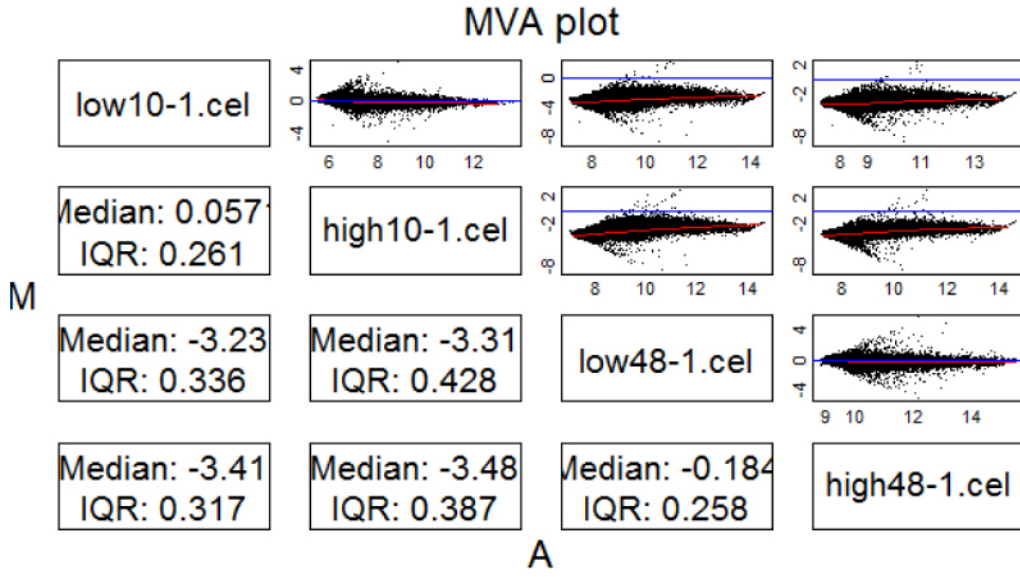
Şekil-4.4, Her bir array üzerindeki yoğunlukların boyutsal dağılımlarını gözlemlemek ve boyutsal bozuklukları tespit etmek için CEL dosyalarının görüntüleri göstermektedir. Arrayler üzerinde herhangi bir boyutsal bozukluk olmadığı gözlenmiştir. Şekil-4.5, her bir array'e ilişkin histogram grafiklerini göstermektedir. Histogram grafiklerinden tüm arrayler için normal dağılımın gözlenmediği sonucuna varılmıştır.



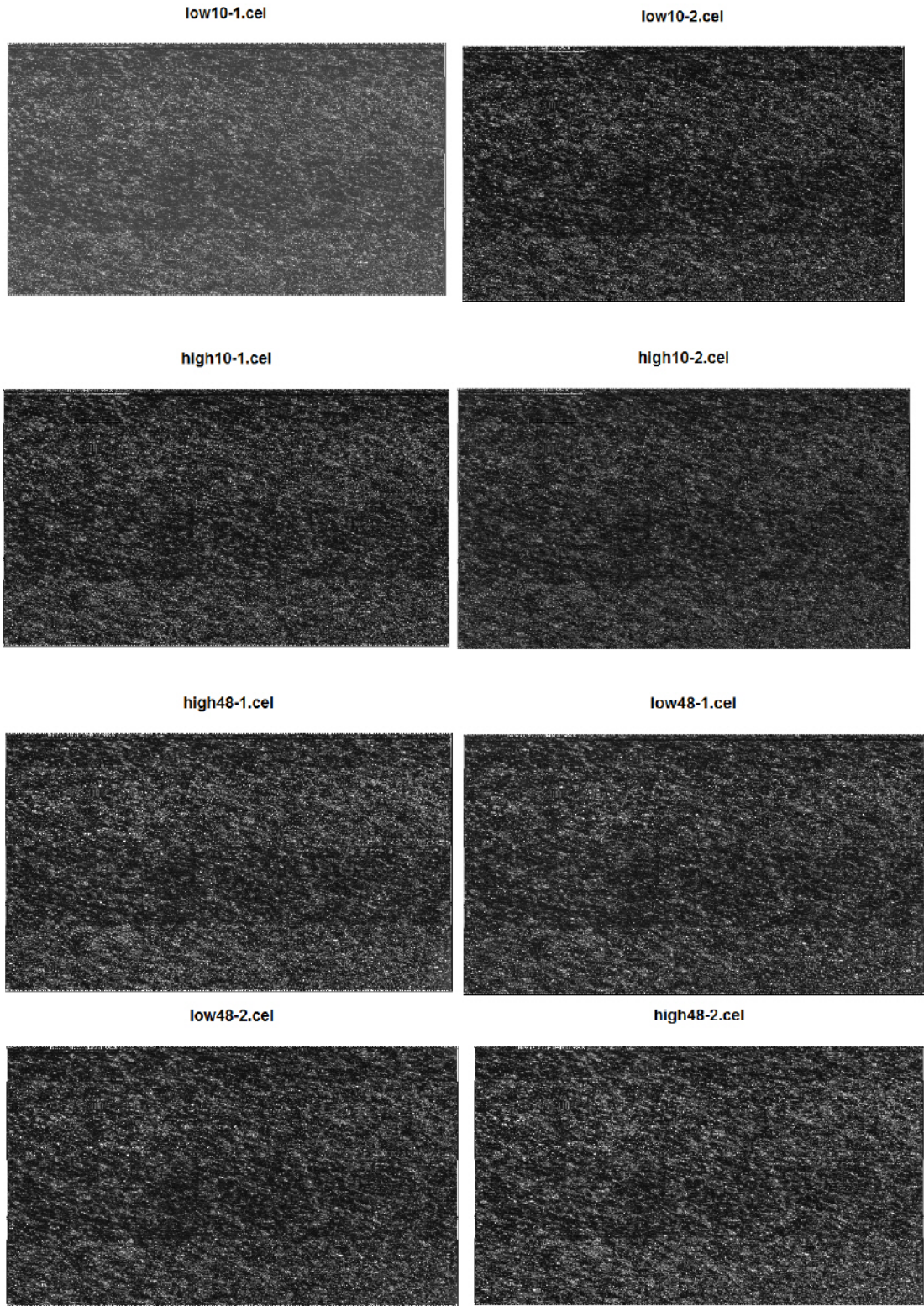
Şekil-4.1: QC grafiği



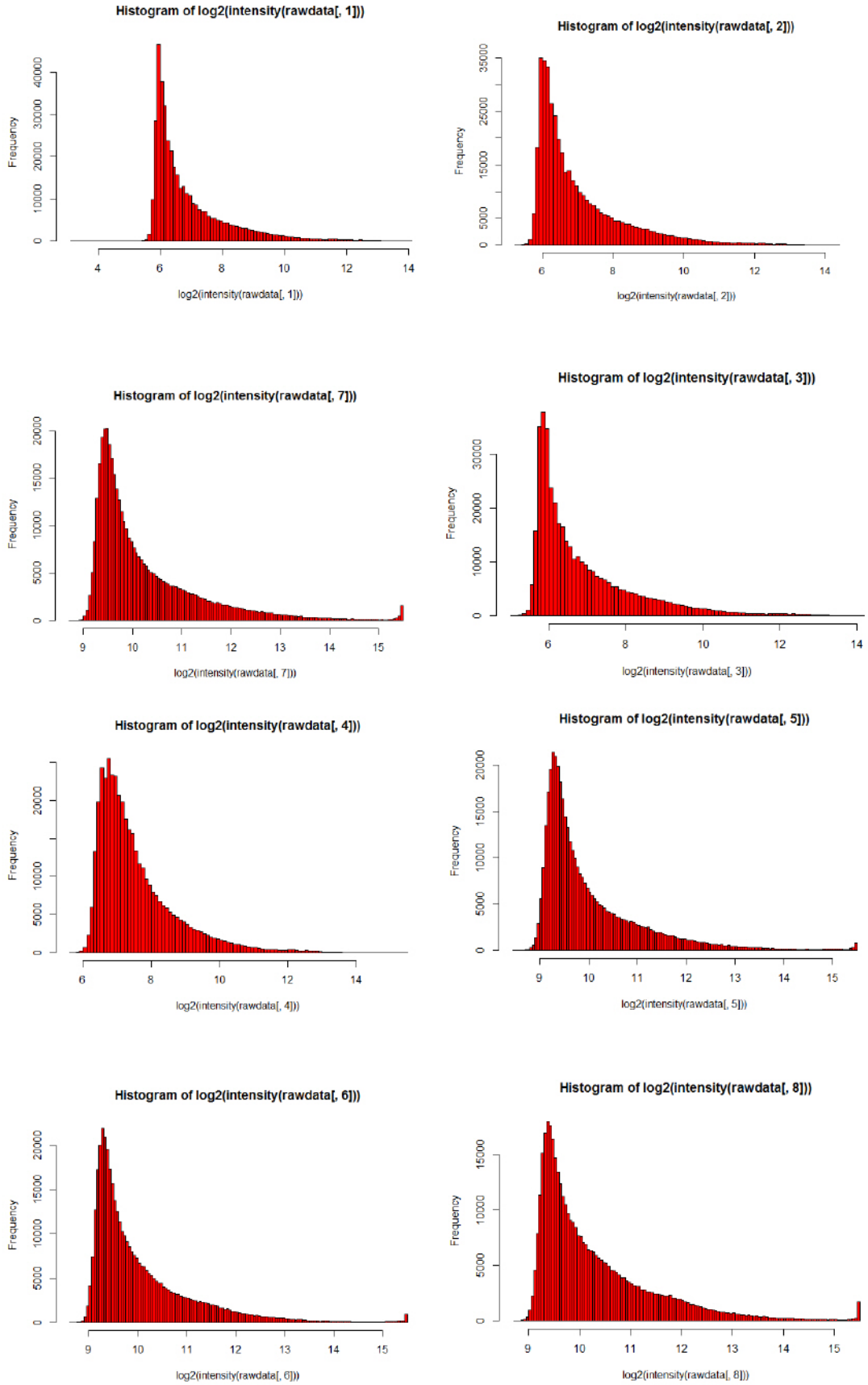
Şekil-4.2: Her bir array için kutu grafikleri



Şekil-4.3: MA Grafikleri

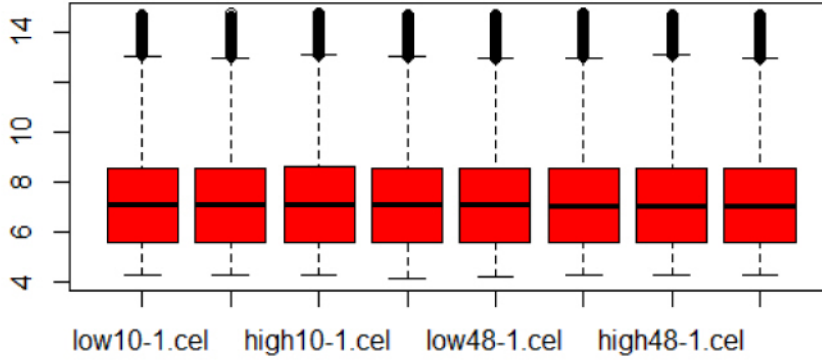


Şekil-4.4: Arraylerin boyutsal dağılımları



Şekil-4.5: Her bir array için ham verilerin histogramları

Şekil-4.6, RMA normalizasyonu sonrasında elde edilen ekspresyon seti için her bir arraye ait kutu grafiklerini göstermektedir. Normalizasyon sonrası arraylerin homojenliğinin sağlandığı gözlenmiştir.



Şekil-4.6: Normalizasyon sonrası arraylere ilişkin kutu grafikleri

4.2. Dizayn ve Kontrast Matrisi

Sekiz array ve dört deneysel koşul için Dizayn matrisi aşağıdaki gibi elde edilmiştir. Matrisin sütunları deneysel koşulları, satırları ise arrayleri temsil etmektedir.

$$X = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Kontrast matrisi tüm mümkün karşılaştırmalar için aşağıdaki gibi elde edilmiştir. Sütunlar C1, C2 ve C3 olarak mümkün üç karşılaştırmayı temsil etmektedir. Satırlar ise 4 deneysel koşulu göstermektedir.

$$C = \begin{pmatrix} -1 & 0 & -1 \\ 0 & -1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

C1: 10 saatte Östrojen etkisi

C2: 48 saatte östrojen etkisi

C3: Östrojen olmadan zamanın etkisi

4.3. Farklı Gen Ekspresyon Analizi Sonuçları

Tablo-4.1, 10 saatte östrojen varlığı ve yokluğu arasında farklı ifade edilen ilk 50 genin listesini göstermektedir. Tablo her bir gene ait Affymetrix gen ID'lerini, logaritmik "fold change" değerlerini, ortalama ekspresyon değerlerini, moderated t-istatistiği değerlerini, p-değerlerini, ayarlanmış p-değerlerini ve log-odd'ları göstermektedir. Logaritmik "fold change" değerleri (M değerleri) genelleştirilmiş doğal bir logaritmik ölçek üzerindeki ekspresyon farklılıklarını içermektedir. Ortalama ekspresyon değerleri mutlak yoğunlukları ifade etmemektedir. İlk karşılaştırma için, 10 saatte östrojen varlığı ve yokluğu arasında farklı ifade edilen 326 gen bulunmuştur. Karşılaştırma 115 gen için negatif anlamlı ve 211 gen için pozitif anlamlıdır. Karşılaştırmanın anlamlı olmadığı gen sayısı 12299 olarak elde edilmiştir. Tablo-4.2, ilk karşılaştırma için farklı ifade edilen ilk 50 genin Affy ID'leri ile birlikte Entrez ID'lerini, gen isimleri ve gen sembollerini göstermektedir.

Tablo-4.3, 48 saatte östrojen varlığı ve yokluğu arasında farklı ifade edilen ilk 50 genin listesini göstermektedir. 48 saatte östrojenin etkisini temsil eden ikinci karşılaştırma için 260 negatif anlamlı, 378 pozitif anlamlı ve 11987 anlamlı olmayan karşılaştırma elde edilmiştir. Toplamda 48 saatte östrojen varlığı ve yokluğu arasında farklı ifade edilen 638 gene ulaşılmıştır. Tablo 4.4, ikinci karşılaştırma için farklı ifade edilen ilk 50 genin Affy ID'leri ile birlikte Entrez ID'lerini, gen isimleri ve gen sembollerini göstermektedir.

Tablo-4.5, östrojen faktörü olmadan 48 saat ve 10 saat arasında farklı ifade edilen ilk 50 genin listesini göstermektedir. Östrojen yokluğunda zamanın etkisini gösteren 3. karşılaştırma için 389 negatif anlamlı, 201 pozitif anlamlı ve 12035 tane anlamlı olmayan karşılaştırma elde edilmiştir. 48 saat ve 10 saat arasında toplamda 590 gen farklı ifade edilmektedir.

Tablo-4.1: 10 saatte östrojen varlığı ve yokluğu arasında farklı ifade edilen ilk 50 genin listesi

	Affy ID	Entrez ID	log FC	Ort_Eks	t-istatistiği	p-değeri	adj. p-değeri	Log-Odd
1	39642_at	54898	2,93942771	7,876515498	23,7171523	4,74E-09	3,13E-05	9,966809693
2	910_at	7083	3,113732544	9,660237759	23,59225083	4,96E-09	3,13E-05	9,942522403
3	31798_at	7031	2,800194536	12,11577812	16,38508504	1,03E-07	3,51E-04	7,977290287
4	41400_at	7083	2,381039941	10,04155327	16,22463361	1,11E-07	3,51E-04	7,916920592
5	40117_at	4175	2,555281868	9,676557206	15,68069724	1,47E-07	3,58E-04	7,705092523
6	1854_at	4605	2,507615838	8,53209895	15,15847605	1,95E-07	3,58E-04	7,490766357
7	39755_at	7494	1,679330929	12,13183867	15,06365206	2,05E-07	3,58E-04	7,450642601
8	1824_s_at	5111	1,914637429	9,238869693	14,87915209	2,27E-07	3,58E-04	7,37147496
9	1126_s_at	960	1,782825046	6,879918348	13,83039753	4,12E-07	5,78E-04	6,892307434
10	1536_at	990	2,66225839	5,937221965	13,26247284	5,80E-07	7,32E-04	6,610485643
11	981_at	4173	1,818408246	7,781305836	13,08578423	6,46E-07	7,42E-04	6,519352125
12	33252_at	4172	1,740212749	8,000356058	12,58547682	8,86E-07	9,20E-04	6,251908441
13	1505_at	7298	2,395860925	8,764755727	12,48145498	9,48E-07	9,20E-04	6,19449836
14	34363_at	6414	-1,747957273	5,553958839	-12,19735114	1,14E-06	1,03E-03	6,034407916
15	1884_s_at	5111	2,799396394	9,034795954	12,05054444	1,26E-06	1,06E-03	5,949748627
16	36134_at	10439	2,491535608	8,275733726	11,78584879	1,50E-06	1,19E-03	5,793668332
17	37485_at	11001	1,607306026	6,672714099	11,38503852	1,99E-06	1,48E-03	5,548589215
18	239_at	1509	1,569567734	11,24917868	10,4039124	4,07E-06	2,66E-03	4,900708933
19	38116_at	9768	2,318493245	9,513109474	10,39724624	4,09E-06	2,66E-03	4,896058448
20	40533_at	332	1,256465834	8,465999684	10,3598426	4,21E-06	2,66E-03	4,869899446
21	35249_at	9134	2,223901904	6,075580411	10,0910298	5,18E-06	3,12E-03	4,67859069
22	673_at	4522	1,432474084	9,828893528	9,89789874	6,04E-06	3,27E-03	4,537469521
23	37043_at	3399	-1,320388746	11,52943354	-9,896422034	6,04E-06	3,27E-03	4,53637839
24	1775_at	23649	1,793191228	8,677821649	9,840457302	6,32E-06	3,27E-03	4,49488933
25	36833_at	2717	1,518012363	8,709900072	9,811747757	6,47E-06	3,27E-03	4,473501683
26	36634_at	7832	-1,582995155	8,671715755	-9,694611336	7,11E-06	3,45E-03	4,385500727
27	41583_at	2237	2,164070654	9,771353738	9,594135406	7,71E-06	3,55E-03	4,309060052
28	947_at	4176	2,302831419	10,97252402	9,570022221	7,86E-06	3,55E-03	4,290582143
29	2042_s_at	4602	1,998635672	8,178149682	9,320854741	9,66E-06	4,21E-03	4,096576803
30	37686_s_at	7374	1,858270785	8,799777347	9,256049612	1,02E-05	4,26E-03	4,04518642
31	1823_g_at	NA	1,433357217	7,941790261	9,163845093	1,10E-05	4,26E-03	3,971391823
32	37899_at	7298	1,728186284	6,046277064	9,144056502	1,12E-05	4,26E-03	3,955449835
33	38827_at	10551	2,93219975	6,829772698	9,110154492	1,15E-05	4,26E-03	3,928051448
34	32027_at	5174	2,022415207	5,658372982	9,03318102	1,23E-05	4,26E-03	3,865436466
35	35312_at	4171	2,267497578	10,12458863	9,026360642	1,24E-05	4,26E-03	3,859860884
36	32441_at	1503	1,080083437	5,799144609	8,987117509	1,28E-05	4,26E-03	3,827692666
37	41439_at	4430	-1,224572328	8,996632053	-8,978553825	1,29E-05	4,26E-03	3,820653032
38	36922_at	6241	1,792699498	7,373168178	8,97232151	1,30E-05	4,26E-03	3,815525378
39	36617_at	3397	-1,253134208	9,099454243	-8,958620993	1,32E-05	4,26E-03	3,80423995
40	39756_g_at	7494	1,686882707	10,22772343	8,920764982	1,36E-05	4,29E-03	3,772961819
41	32239_at	4147	-1,355703214	6,257010396	-8,791860854	1,52E-05	4,68E-03	3,665397497
42	35437_at	4886	1,001041656	7,354160662	8,611189488	1,78E-05	5,26E-03	3,51183387
43	653_at	5985	1,014629959	7,315216747	8,605786569	1,79E-05	5,26E-03	3,50719045
44	40215_at	7357	1,334946483	11,00037137	8,491259622	1,99E-05	5,29E-03	3,408053577
45	32043_at	8614	1,315642927	8,038722165	8,482199076	2,00E-05	5,29E-03	3,400152397
46	160043_at	4603	1,940053699	5,841700413	8,465367843	2,03E-05	5,29E-03	3,385452048
47	1178_at	NA	1,18762604	5,403992849	8,465361994	2,03E-05	5,29E-03	3,385446934
48	35995_at	11130	1,828735319	8,872008312	8,417625849	2,12E-05	5,29E-03	3,34359245
49	40425_at	1942	-1,289385811	9,691118189	-8,400851831	2,16E-05	5,29E-03	3,328828144
50	34678_at	26509	1,421563714	9,997873993	8,384135928	2,19E-05	5,29E-03	3,314085356

Tablo-4.2: İlk karşılaştırma için farklı ifade edilen ilk 50 genin İsimleri

	Gen ID	Gen Sembolü	Entrez ID	Gen İsmi
1	910_at	TK1	7083	thymidine kinase 1, soluble
2	31798_at	TFF1	7031	trefoil factor 1
3	1854_at	MYBL2	4605	v-myb myeloblastosis viral oncogene homo
4	38116_at	KIAA0101	9768	KIAA0101
5	38065_at	HMGB2	3148	high-mobility group box 2
6	39755_at	XBP1	7494	X-box binding protein 1
7	1592_at	TOP2A	7153	topoisomerase (DNA) II alpha 170kDa
8	41400_at	TK1	7083	thymidine kinase 1, soluble
9	33730_at	GPRC5A	9052	G protein-coupled receptor, family C, gr
10	1651_at	UBE2C	11065	ubiquitin-conjugating enzyme E2C
11	38414_at	CDC20	991	cell division cycle 20 homolog (S. cerev
12	1943_at	CCNA2	890	cyclin A2
13	40117_at	MCM6	4175	minichromosome maintenance complex compo
14	40533_at	BIRC5	332	baculoviral IAP repeat-containing 5 (sur
15	39642_at	ELOVL2	54898	elongation of very long chain fatty acid
16	34851_at	AURKA	6790	aurora kinase A
17	1824_s_at	PCNA	5111	proliferating cell nuclear antigen
18	35995_at	ZWINT	11130	ZW10 interactor
19	893_at	UBE2S	27338	ubiquitin-conjugating enzyme E2S
20	40079_at	NA	NA	NA
21	40619_at	UBE2S	27338	ubiquitin-conjugating enzyme E2S
22	37043_at	ID3	3399	inhibitor of DNA binding 3, dominant neg
23	32263_at	CCNB2	9133	cyclin B2
24	1803_at	CDC2	983	cell division cycle 2, G1 to S and G2 to
25	41439_at	MYO1B	4430	myosin IB
26	40697_at	CCNA2	890	cyclin A2
27	2031_s_at	CDKN1A	1026	cyclin-dependent kinase inhibitor 1A (p2
28	36134_at	OLFM1	10439	olfactomedin 1
29	757_at	ANXA2	302	annexin A2
30	32272_at	TUBA1B	10376	tubulin, alpha 1b
31	39109_at	TPX2	22974	TPX2, microtubule-associated, homolog (X
32	1884_s_at	PCNA	5111	proliferating cell nuclear antigen
33	40898_at	SQSTM1	8878	sequestosome 1
34	349_g_at	KIFC1	3833	kinesin family member C1
35	40412_at	PTTG1	9232	pituitary tumor-transforming 1
36	846_s_at	BAK1	578	BCL2-antagonist/killer 1
37	1945_at	CCNB1	891	cyclin B1
38	2049_s_at	JUNB	3726	jun B proto-oncogene
39	36833_at	GLA	2717	galactosidase, alpha
40	38131_at	PTGES	9536	prostaglandin E synthase
41	1544_at	BLM	641	Bloom syndrome
42	38254_at	TBC1D9	23158	TBC1 domain family, member 9 (with GRAM
43	673_at	MTHFD1	4522	methylenetetrahydrofolate dehydrogenase
44	33308_at	GUSB	2990	glucuronidase, beta
45	2042_s_at	MYB	4602	v-myb myeloblastosis viral oncogene homo
46	40195_at	H2AFX	3014	H2A histone family, member X
47	1087_at	EPOR	2057	erythropoietin receptor
48	34238_at	IGSF1	3547	immunoglobulin superfamily, member 1
49	981_at	MCM4	4173	minichromosome maintenance complex compo
50	36374_at	THSD4	79875	thrombospondin, type I, domain containin

Tablo-4.3: 48 saatte östrojen varlığı ve yokluğu arasında farklı ifade edilen ilk 50 genin listesi

	Gen ID	log FC	Ort_Eks	t-istatistiği	p-değeri	adj.P.Val	Log-odd
1	910_at	3,85506143	9,66023776	29,2091806	8,266125E-10	1,043598E-05	11,6061927
2	31798_at	3,59733384	12,1157781	21,0494736	1,284430E-08	7,631722E-05	9,89055749
3	1854_at	3,34089643	8,53209895	20,1956367	1,813478E-08	7,631722E-05	9,64139864
4	38116_at	3,75889126	9,51310947	16,8566883	8,116230E-08	2,511100E-04	8,48019712
5	38065_at	2,99364143	9,0971826	16,2091356	1,121213E-07	2,511100E-04	8,21417464
6	39755_at	1,76524944	12,1318387	15,8343438	1,359405E-07	2,511100E-04	8,0531341
7	1592_at	2,2964836	8,31132952	15,7884078	1,392293E-07	2,511100E-04	8,03302509
8	41400_at	2,24351036	10,0415533	15,2874939	1,814762E-07	2,752126E-04	7,80829476
9	33730_at	-2,0413903	8,5734698	-15,142985	1,961911E-07	2,752126E-04	7,74155644
10	1651_at	2,96828252	10,5042757	14,7809652	2,392480E-07	3,020507E-04	7,57047037
11	38414_at	2,01702192	9,46116996	14,5906245	2,660400E-07	3,053413E-04	7,47822215
12	1943_at	2,19006224	7,59657903	14,0031036	3,722630E-07	3,689505E-04	7,18304085
13	40117_at	2,27622789	9,67655721	13,9682595	3,799095E-07	3,689505E-04	7,16502273
14	40533_at	1,6404117	8,46599968	13,5255623	4,939428E-07	4,454305E-04	6,93089142
15	39642_at	1,6144453	7,8765155	13,0263605	6,705257E-07	5,180349E-04	6,65481132
16	34851_at	1,95726413	9,96111414	12,8459337	7,508260E-07	5,180349E-04	6,55173082
17	1824_s_at	1,64136588	9,23886969	12,7554868	7,950641E-07	5,180349E-04	6,49937685
18	35995_at	2,75557929	8,87200831	12,6838669	8,321542E-07	5,180349E-04	6,45759359
19	893_at	1,54063186	10,9540793	12,6640207	8,427698E-07	5,180349E-04	6,44596366
20	40079_at	-2,4141533	8,23054727	-12,628114	8,623610E-07	5,180349E-04	6,42486506
21	40619_at	1,62573693	9,96015846	12,5752987	8,921047E-07	5,180349E-04	6,39369693
22	37043_at	-1,6735079	11,5294335	-12,543079	9,108090E-07	5,180349E-04	6,37460413
23	32263_at	2,4231151	7,2659787	12,4880923	9,437467E-07	5,180349E-04	6,3418803
24	1803_at	1,59202672	8,72651986	12,1917886	1,145687E-06	6,026792E-04	6,16247464
25	41439_at	-1,6116453	8,99663205	-11,81657	1,473653E-06	7,336152E-04	5,9276163
26	40697_at	1,84694799	8,33621741	11,7800226	1,510811E-06	7,336152E-04	5,90426814
27	2031_s_at	-1,7234747	10,7529057	-11,560564	1,757079E-06	7,980454E-04	5,76225166
28	36134_at	2,44169229	8,27573373	11,5500722	1,769922E-06	7,980454E-04	5,75538358
29	757_at	-1,7076407	8,74978636	-11,448914	1,899255E-06	8,268308E-04	5,68878764
30	32272_at	1,54669681	11,6301314	11,3132185	2,089540E-06	8,515524E-04	5,59837595
31	39109_at	2,46502137	9,31627775	11,3122711	2,090940E-06	8,515524E-04	5,59774036
32	1884_s_at	2,58669798	9,03479595	11,1349429	2,372500E-06	9,360254E-04	5,47768156
33	40898_at	-1,2810346	11,1798894	-11,038272	2,543562E-06	9,690933E-04	5,41130892
34	349_g_at	1,68072357	7,5304116	10,9918723	2,630499E-06	9,690933E-04	5,3792173
35	40412_at	2,34536704	9,38985297	10,9414778	2,728677E-06	9,690933E-04	5,34418883
36	846_s_at	-2,4235919	8,88317379	-10,924157	2,763355E-06	9,690933E-04	5,33210741
37	1945_at	2,74540484	7,98669717	10,737749	3,169217E-06	1,057770E-03	5,20070875
38	2049_s_at	-1,58029	9,38579911	-10,73156	3,183783E-06	1,057770E-03	5,19630266
39	36833_at	1,64507305	8,70990007	10,6330108	3,426029E-06	1,093976E-03	5,12575773
40	38131_at	1,23706983	8,69934805	10,5939963	3,527528E-06	1,093976E-03	5,09762944
41	1544_at	2,05498674	6,69477238	10,5668443	3,600135E-06	1,093976E-03	5,077986
42	38254_at	-1,3274076	11,8682758	-10,533041	3,692858E-06	1,093976E-03	5,05345257
43	673_at	1,5203609	9,82889353	10,5051661	3,771313E-06	1,093976E-03	5,03315688
44	33308_at	-1,6315008	12,0120862	-10,449202	3,934475E-06	1,093976E-03	4,99223002
45	2042_s_at	2,22671777	8,17814968	10,3845404	4,132823E-06	1,093976E-03	4,94464255
46	40195_at	2,00172086	10,4144193	10,3805478	4,145430E-06	1,093976E-03	4,9416936
47	1087_at	-1,5615035	7,93242857	-10,378425	4,152151E-06	1,093976E-03	4,94012487
48	34238_at	1,26188059	7,58114621	10,3761777	4,159276E-06	1,093976E-03	4,93846444
49	981_at	1,43382139	7,78130584	10,3181876	4,347965E-06	1,120266E-03	4,89547345
50	36374_at	1,37473264	7,42194324	10,2627986	4,537094E-06	1,145616E-03	4,85416506

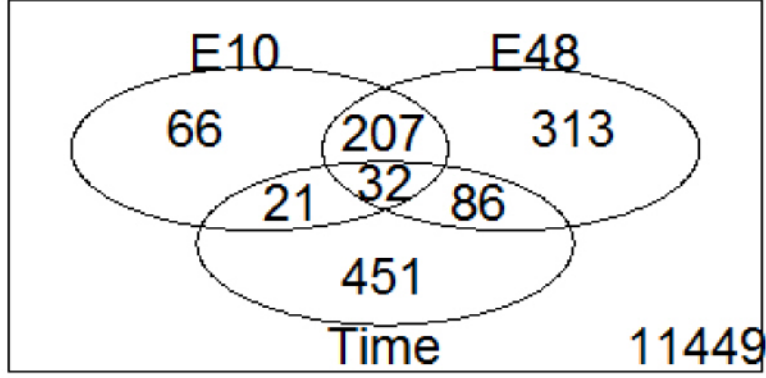
Tablo-4.4: İkinci karşılaştırma için farklı ifade edilen ilk 50 genin İsimleri

	Gen ID	Gen Sembolü	Gen İsmi	Entrez ID
1	910_at	TK1	thymidine kinase 1, soluble	7083
2	31798_at	TFF1	trefoil factor 1	7031
3	1854_at	MYBL2	v-myb myeloblastosis viral oncogene homo	4605
4	38116_at	KIAA0101	KIAA0101	9768
5	38065_at	HMGB2	high-mobility group box 2	3148
6	39755_at	XBP1	X-box binding protein 1	7494
7	1592_at	TOP2A	topoisomerase (DNA) II alpha 170kDa	7153
8	41400_at	TK1	thymidine kinase 1, soluble	7083
9	33730_at	GPRC5A	G protein-coupled receptor, family C, gr	9052
10	1651_at	UBE2C	ubiquitin-conjugating enzyme E2C	11065
11	38414_at	CDC20	cell division cycle 20 homolog (S. cerev	991
12	1943_at	CCNA2	cyclin A2	890
13	40117_at	MCM6	minichromosome maintenance complex compo	4175
14	40533_at	BIRC5	baculoviral IAP repeat-containing 5 (sur	332
15	39642_at	ELOVL2	elongation of very long chain fatty acid	54898
16	34851_at	AURKA	aurora kinase A	6790
17	1824_s_at	PCNA	proliferating cell nuclear antigen	5111
18	35995_at	ZWINT	ZW10 interactor	11130
19	893_at	UBE2S	ubiquitin-conjugating enzyme E2S	27338
20	40079_at	NA	NA	NA
21	40619_at	UBE2S	ubiquitin-conjugating enzyme E2S	27338
22	37043_at	ID3	inhibitor of DNA binding 3, dominant neg	3399
23	32263_at	CCNB2	cyclin B2	9133
24	1803_at	CDC2	cell division cycle 2, G1 to S and G2 to	983
25	41439_at	MYO1B	myosin IB	4430
26	40697_at	CCNA2	cyclin A2	890
27	2031_s_at	CDKN1A	cyclin-dependent kinase inhibitor 1A (p2	1026
28	36134_at	OLFM1	olfactomedin 1	10439
29	757_at	ANXA2	annexin A2	302
30	32272_at	TUBA1B	tubulin, alpha 1b	10376
31	39109_at	TPX2	TPX2, microtubule-associated, homolog (X	22974
32	1884_s_at	PCNA	proliferating cell nuclear antigen	5111
33	40898_at	SQSTM1	sequestosome 1	8878
34	349_g_at	KIFC1	kinesin family member C1	3833
35	40412_at	PTTG1	pituitary tumor-transforming 1	9232
36	846_s_at	BAK1	BCL2-antagonist/killer 1	578
37	1945_at	CCNB1	cyclin B1	891
38	2049_s_at	JUNB	jun B proto-oncogene	3726
39	36833_at	GLA	galactosidase, alpha	2717
40	38131_at	PTGES	prostaglandin E synthase	9536
41	1544_at	BLM	Bloom syndrome	641
42	38254_at	TBC1D9	TBC1 domain family, member 9 (with GRAM	23158
43	673_at	MTHFD1	methylenetetrahydrofolate dehydrogenase	4522
44	33308_at	GUSB	glucuronidase, beta	2990
45	2042_s_at	MYB	v-myb myeloblastosis viral oncogene homo	4602
46	40195_at	H2AFX	H2A histone family, member X	3014
47	1087_at	EPOR	erythropoietin receptor	2057
48	34238_at	IGSF1	immunoglobulin superfamily, member 1	3547
49	981_at	MCM4	minichromosome maintenance complex compo	4173
50	36374_at	THSD4	thrombospondin, type I, domain containin	79875

Tablo-4.5: Östrojen faktörü olmadan 48 saat ve 10 saat arasında farklı ifade edilen ilk 50 genin listesi

	Affy ID	log FC	Ort_Eks	t-istatistiği	p-değeri	Adj. p-değeri	Log Odd
1	AFFX-CreX-5_at	-6,826151212	9,921025986	-32,8101093	3,108614E-10	2,703546E-06	10,6551118
2	AFFX-CreX-3_at	-6,526323934	10,4063428	-31,58458654	4,282844E-10	2,703546E-06	10,5479856
3	AFFX-BioDn-5_at	-3,734781347	8,348423618	-19,15514612	2,815350E-08	1,184793E-04	8,57531879
4	AFFX-BioB-M_at	-3,408360438	8,270104515	-18,28381354	4,143024E-08	1,307642E-04	8,33645777
5	AFFX-BioDn-3_at	-2,485796741	11,40108019	-15,5252204	1,598695E-07	4,036706E-04	7,42747525
6	39581_at	-2,673757028	6,62332121	-13,92948732	3,886242E-07	6,755348E-04	6,76973048
7	AFFX-BioC-3_at	-2,986652659	8,159920488	-13,92124783	3,905047E-07	6,755348E-04	6,76603337
8	37014_at	-1,516742013	7,820017131	-13,76536484	4,280617E-07	6,755348E-04	6,69545904
9	2004_at	-2,064095639	6,826915667	-12,39872824	1,000140E-06	1,402974E-03	6,02191194
10	AFFX-BioC-5_at	-2,053876343	8,739444433	-11,72355149	1,570292E-06	1,982494E-03	5,64898244
11	34363_at	-1,654036504	5,553958839	-11,54196635	1,779917E-06	1,997433E-03	5,54366734
12	38065_at	-2,111220862	9,097182596	-11,43125054	1,922899E-06	1,997433E-03	5,47836272
13	40071_at	-1,732101181	8,096773088	-11,30276674	2,105052E-06	1,997433E-03	5,40151988
14	33730_at	1,511442816	8,573469796	11,21184666	2,245526E-06	1,997433E-03	5,34644537
15	32597_at	-1,516480513	8,063733282	-11,13453889	2,373188E-06	1,997433E-03	5,29915522
16	1034_at	1,312306585	10,05448203	10,90338481	2,805591E-06	2,213787E-03	5,15517764
17	41386_i_at	1,244845164	9,629715209	10,65598281	3,367785E-06	2,501075E-03	4,99668354
18	910_at	-1,333710375	9,660237759	-10,10530907	5,126632E-06	3,595762E-03	4,62674807
19	AFFX-BioB-3_at	-2,610904824	7,815168049	-9,521700472	8,181516E-06	5,436402E-03	4,20703579
20	35059_at	1,077954019	7,075002482	9,340293759	9,507995E-06	6,001922E-03	4,07039109
21	33899_at	-1,584573505	7,698742005	-9,131207133	1,134059E-05	6,540820E-03	3,90908004
22	38116_at	-2,034858149	9,513109474	-9,125289141	1,139786E-05	6,540820E-03	3,90445364
23	1651_at	-1,784304444	10,50427568	-8,885185892	1,401516E-05	7,399336E-03	3,71387186
24	40079_at	1,68999851	8,230547268	8,840156762	1,457665E-05	7,399336E-03	3,67749482
25	1197_at	-1,288722595	7,985672566	-8,834248479	1,465215E-05	7,399336E-03	3,67270668
26	38814_at	-1,116091615	8,384875609	-8,615012698	1,778452E-05	8,512578E-03	3,49252417
27	40759_at	-1,038523025	7,250358939	-8,588865981	1,820512E-05	8,512578E-03	3,47070455
28	39397_at	-1,218686945	8,592946213	-8,509181529	1,955631E-05	8,817799E-03	3,40376585
29	35995_at	-1,813721585	8,872008312	-8,348517982	2,263098E-05	9,515541E-03	3,26675233
30	AFFX-BioB-5_at	-2,545266405	8,253587852	-8,333507343	2,294450E-05	9,515541E-03	3,25380962
31	35224_at	0,98132609	6,672440117	8,313731376	2,336489E-05	9,515541E-03	3,23672075
32	37026_at	1,138842837	8,62024349	8,260994393	2,452819E-05	9,562747E-03	3,19094116
33	36617_at	-1,152119106	9,099454243	-8,236466893	2,509092E-05	9,562747E-03	3,16954573
34	34717_s_at	-1,007749747	6,022097393	-8,208369559	2,575314E-05	9,562747E-03	3,14495495
35	34472_at	-1,128904935	6,738772414	-8,093626014	2,866528E-05	1,007709E-02	3,04362197
36	40690_at	-1,173370973	9,687965487	-8,064259288	2,946782E-05	1,007709E-02	3,01745062
37	39560_at	-0,938545262	8,315207215	-8,053620373	2,976467E-05	1,007709E-02	3,00794532
38	1945_at	-2,049194369	7,986697166	-8,014750499	3,087760E-05	1,007709E-02	2,97310824
39	2010_at	-0,921013373	9,133447807	-8,006176883	3,112922E-05	1,007709E-02	2,96540105
40	32239_at	-1,214045669	6,257010396	-7,873198564	3,533636E-05	1,115304E-02	2,84478311
41	36631_at	-1,589391326	7,777589716	-7,807216721	3,765391E-05	1,159465E-02	2,78417473
42	31867_at	-1,535281454	6,025935955	-7,750657809	3,977440E-05	1,195600E-02	2,73181585
43	881_at	1,03716904	6,204206132	7,625638667	4,494489E-05	1,278680E-02	2,61473455
44	38368_at	-1,762218504	9,484085201	-7,607137759	4,577116E-05	1,278680E-02	2,59724934
45	34736_at	-1,791017219	7,756538393	-7,606574537	4,579657E-05	1,278680E-02	2,5967164
46	37409_at	-0,973308079	5,394338029	-7,589173501	4,658952E-05	1,278680E-02	2,58023181
47	37895_at	-1,315537248	9,304807001	-7,499522495	5,092224E-05	1,367858E-02	2,49471976
48	39092_at	-1,54764699	9,189420576	-7,465519913	5,268021E-05	1,379620E-02	2,46203003
49	960_g_at	-1,019037803	10,83362323	-7,449236776	5,354567E-05	1,379620E-02	2,4463252
50	32448_at	0,895149155	8,357231373	7,384704063	5,713304E-05	1,422473E-02	2,38376182

Son olarak, tüm karşılaştırmalar için farklı ifade edilen genlerin Ven Şeması elde edilmiştir. Şekil-4.7 karşılaştırmaların Ven şemasını göstermektedir.



Şekil-4.7: Östrojen10, östrojen48 ve zaman faktörleri için tüm karşılaştırmalara ait Ven Şeması

5. TARTIŞMA VE SONUÇ

Mikroarrayler geniş çeşitlilikteki deneysel ortamlarda farklı gen ekspresyonunun tespiti için kullanılmaktadır. Biyoinformatik araçları, mikroarray analizinin gerçekleştirilmesindeki ihtiyaçları karşılamaktadır. Farklı gen ekspresyon analizi ile biyolojik yolların iç yüzü anlaşılabilir, ileriki hipotez geliştirme için bir platform sağlanmaktadır (4).

Çok faktörlü mikroarray deneyleri genellikle genetik olarak birbiriyle aynı hücre dizileri gibi model organizmalara kombinasyon halindeki tedavilerin uygulanmasını içermektedir. Bu deneylerin başladığı eşit referans noktası teorik olarak, doğal bir şekilde meydana gelen bireyler arası değişkenliği sınırlamaktadır, böylece farklı gen ekspresyonunu araştırma altındaki tedavilere veya deneysel koşullara bağlama imkanı ortaya çıkmaktadır. Sıklıkla, bu deneyler çeşitli tedavi kombinasyonları ile genetik ağların karışıklığını araştırmak için tasarlanmaktadır. Böylece genetik ağ yeniden yapılandırılmasının ilk adımları sağlanmaktadır (4).

Faktöriyel tasarımlarda, tedavilerin etkileri ve birbirleriyle etkileşimleri uygun olarak doğrusal bir modelde ölçülebilmektedir. Transkripsiyon ve çeviri mekanizmaları için ilgili farklılıklar dikkatli bir hesaplama ile belirlendiği takdirde, araştırmacılar sonuçlarına çok anlamlı biyolojik yorumlamalar yükleyebilmektedirler. Çok faktörlü lineer modeller mikroarray çalışmalarda çeşitli amaçlar için kullanılmaktadır. Kombinasyon olarak uygulanan tedaviler nedeniyle farklı ifade edilmiş genleri belirlemenin dışında, cDNA mikroarraylerinin veri önışlemesi içinde lineer modeller oldukça faydalıdır (9).

Gordon K. Smyth'in (4) RStudio paket programı için tasarladığı "Limma" paketi, mikroarray verilerin gen ekspresyon veri analizinde en çok kullanılan araçlardan biri haline gelmiştir. Farklı genlerin çıkarımı için moderated t-istatistiği ve fold change kriteri literatürde en çok kullanılan yöntemler arasında bulunmaktadır. İki renkli mikroarray tasarımları için ANOVA modelinin kullanılması ilk defa Kerr ve arkadaşları tarafından önerilmiştir ve literatürde önemli bir yer almıştır (4).

Wolfgang Huber'in (21) 74 böbrek hücreli kanser hasta biyopsi örneklerinin gen ifade verilerini analiz ettiği çalışması literatürde gen ekspresyon analizi için gösterilebilecek örneklerden biridir. Yapılan çalışmada Böbrek hücreli kanser örneklerinin üç farklı histolojik türünde farklılaşan gen ekspresyonları elde edilmiştir. Wolfgang Huber araştırmasında Çoklu karşılaştırmada Yanlış bulgu oranını önlemek amacıyla Benjamini and Hochberg (1995) yöntemini kullanmıştır (21). Benjamini and Hochberg yöntemi gen ekspresyon analizlerinde yanlış bulgu oranını kontrol etmek için kullanılan en yaygın ve en etkili yöntemlerden biridir (4).

Göğüs kanseri teşhisi konulmuş kadınlardan yaklaşık %15'i, ulusal klinik klavuzlara göre tedavi edilmelerine rağmen, hastalıklarından dolayı 5 yıl içerisinde ölmektedir. Çeşitli gen işaretlerinden elde edilen bilgileri birleştirerek, hastalığın genel tanımlaması ortaya çıkartılabilir. Bu bağlamda, Mikroarray binlerce genin eşzamanlı olarak incelenmesini mümkün hale getirmiştir. Margit Riss ve arkadaşlarının (22) yaptıkları gen ekspresyon profil analizi çalışmasında göğüs kanserinin T1 ve T2 kategorilerini tanımlayan moleküler işaretleri araştırılmıştır. Margit Riss ve arkadaşları (22) çalışmalarında, bu tezin uygulanmasında kullanılan ER+ göğüs kanseri hücrelerine de yer vermiştir. Çoğalma ve çoğalmaya bağlı genlerin seviyesi, ER pozitif kanser hücrelerindeki en güçlü prognostik faktörlerdir. Çoğalmaya bağlı genler, ER negatif kanser hücrelerinde yüksek derecede ifade edilmektedir (22).

Bu tez çalışmasında, Wolfgang Huber ve Robert Gentleman'ın (21) Affymetrix gen çipleri ile yaptıkları deneyde elde edilen gen ekspresyon verileri kullanılmıştır. Çalışmanın amacı ER+ göğüs kanseri hücrelerindeki östrojen faktörüne yanıt veren genleri tespit etmektir. Östrojen faktörünün yanında zaman faktörü de değerlendirmeye alınarak östrojene erken yanıt veren ve geç yanıt veren genler tespit edilmiştir. Erken yanıt veren genler doğrudan-hedef genler kabul edilmektedir. Geç yanıt veren genler ise büyük bir olasılıkla moleküler yolaktaki akış aşağı hedeflerdir.

Östrojene bağlanma üzerine, östrojen reseptörü (ER), ekspresyonlarını uyarıcı ya da bastırıcı belirli genler için, bir transkripsiyon faktörü olarak görev yapmaktadır. Yapılan çalışmada, hücrelerdeki östrojenin birincil ve ikincil hedeflerini belirlemek ile ilgilenilmiştir. 10 saatte etkilenen genler, östrojenin doğrudan hedeflerini oluşturmaktadır. Doğrudan hedefler, östrojen-bağlı ER tarafından doğrudan uyarılan veya engellenen genlerdir.

Sonu olarak, 10 saatte strojen varlıđı ve yokluđu arasında farklılařan toplam 326 gen bulunmuřtur. 48 saat hedefleri, strojenden etkilenmiř yolaklardaki akıř ařađı genler olabilir. Toplamda 638 gen 48 saat ierisinde strojenin varlıđı ve yokluđu arasında farklı ifade edilmiřtir. strojen etkisi olmadıđında yalnızca zaman faktr iin farklı ifade edilen toplam 590 gene ulařılmıřtır. Zaman ve strojen faktrlerinin gen ekspresyonunu etkilediđi sylenebilir. Elde edilen genlere gre strojen faktrne ge yanıt veren, byk ihtimalle molekler yolaktaki akıř ařađı hedefler olan genler iin toplamda farklılařan gen sayısı daha fazladır.

KAYNAKLAR

1. CHIKHALE NJ, GOMASE VS. Bioinformatics theory and practice, 1st edition, Himalaya Publishing House PVT.LTD, Mumbai, page 1-4, 2007.
2. ARHIPOVA I. The role of statistical methods in computer science and bioinformatics. 7th International Conference on Teaching, Brazil, page 1-2, 2006.
3. LOEWE RP, NELSON PJ. Microarray bioinformatics. *Methods in Molecular Biology*, 671: 295-320, 2010.
4. GENTLEMAN R, CAREY VJ. Bioinformatics and computational biology solutions using R and bioconductor, 1st edition, Springer Science+Business Media, Inc., China, page 14-24, 2005.
5. STEKEL D. Microarray bioinformatics, 1st edition, United States of America by Cambridge University Press, New York, page 4-15, 2003.
6. LEE T. Analysis of Microarray Gene Expression Data, 1st edition, Kluwer Academic Publishers, New York, page 32-40, 2004.
7. WOLF K. Gene expression profiling by microarrays, 1st edition, Cambridge University Press, Cambridge, page 76-80, 2006.
8. DANIEL P. Practical Approach to Microarray Data Analysis, 1st edition, Kluwer Academic Publishers, Dordrecht, page 102-150, 2002.
9. SIMON M. Methods of microarray data analysis II, 1st edition, Kluwer Academic Publishers, Dordrecht, page 46-56, 2002.
10. FREUDENBERG JM. Comparison of background correction and normalization procedures for high-density oligonucleotide microarrays, Interdisciplinary Centre for Bioinformatics, Leipzig, page 19-25, 2005.
11. KAUFFMANN A, GENTLEMAN R, HUBER W. Arrayqualitymetrics – a bioconductor package for quality assessment of microarray data. *Bioinformatics*, 25: 415-416, 2009.
12. GOHLMANN H, TALLOEN W. Gene expression studies affymetrix microarray, 1st edition, Taylor&Francis Group LLC, New York, page 185-186, 2009.
13. PARK T, GON YI S, LEE S, LEE K. Diagnostic plots for detecting outlying slides in a cDNA microarray experiment. *BioTechniques*, 38: 463-471, 2005.
14. TUIMALA J. DNA microarray data analysis using Bioconductor. CSC - IT center for science, Espoo, page 63-82, 2008.
15. SREEKUMAR J, JOSE KK. Statistical tests for identification of differentially expressed genes in cDNA microarray experiments. *Indian Journal of Biotechnology*, 7: 423-436, 2008
16. WIT E, MCCLURE J. Statistics for microarray: design, analysis, and inference, 1st edition, John Wiley & Sons Ltd, Chichester, page 150-157, 2004.
17. JAE K. Statistical bioinformatics: for biomedical and life science researchers, 1. edition, John Wiley&Sons, New Jersey, page 78-90, 2010.
18. CUI X, CHURCHILL GA. Statistical tests for differential expression in cDNA microarray experiments. *Genome Biology*, 4:210, 2003.
19. SMYTH GK. Linear models and empirical bayes methods for assessing differentially expression in microarray experiments. *Statistical application in genetics and molecular biology*, Vol 3, 2004.

20. OZKAYA G, ERCAN I. Examining multiple comparison procedures according to error rate, power typei and false discovery rate. *Journal of modern applied statistical methods*, 11(2):348-60, 2012.
21. Bioconductor open source software for bioinformatics, hosting provided by Fred Hutchinson Cancer Research Center, 2003-2014.
22. RIIS MLH, ZHAO X, KAHEV F. Gene expression profile analysis of T1 and T2 breast cancer reveals different activation pathways. *ISRN Oncology*, 2013: 1-12, 2013

TEŐEKKÜR

Yüksek lisans eğitimim boyunca ve tezimi gerçekleřtirmem sırasında benden desteęini hiç esirgemeyen danışmanım Prof. Dr. İlker ERCAN'a güveni ve sabrı için çok teşekkür ederim.

Masaryk Üniversitesindeki stajım boyunca, değerli katkılarını benimle paylaşan Erasmus danışmanım Dr. Eva Budinska'ya çok teşekkür ederim. Yüksek lisans eğitimim boyunca benimle bilgilerini paylaşan Anabilim Dalındaki öğretim üyesi ve elemanlarına da teşekkürlerimi sunarım.

Yüksek lisans eğitimim boyunca ve tez çalışmam sırasında göstermiş oldukları desteklerinden dolayı aileme çok teşekkür ederim.

8. ÖZGEÇMİŞ

1989 yılında Bursa'da doğdum. Lise öğrenimimi 2007 yılında İnegöl Nuh Mehmet Küçük Çalık Lisesi'nde tamamladım. Ege Üniversitesi İstatistik bölümünden 2013 yılında mezun oldum. 2013 güz döneminde Uludağ Üniversitesi Sağlık Bilimleri Enstitüsü Biyoistatistik Anabilim dalında yüksek lisans eğitimime başladım. Haziran-Eylül 2013 tarihleri arasında Masaryk Üniversitesi Biyoistatistik Anabilim Dalında biyoinformatik araştırma grubunda staj yaptım.