

**T.C.
AKDENİZ ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ
Biyostatistik ve Tıbbi Bilişim Anabilim Dalı**

**GENETİK VERİ İÇEREN KLİNİK VERİ
TABANLARINA YÖNELİK ANALİZ YÖNTEMLERİ**

Özgür TOSUN

Doktora Tezi

Antalya, 2012

**T.C.
AKDENİZ ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ
Biyostatistik ve Tıbbi Bilişim Anabilim Dalı**

**GENETİK VERİ İÇEREN KLİNİK VERİ
TABANLARINA YÖNELİK ANALİZ YÖNTEMLERİ**

Özgür TOSUN

Doktora Tezi

**Tez Danışmanı
Prof. Dr. Osman SAKA**

“Kaynakça Gösterilerek Tezimden Yararlanılabilir”

Antalya, 2012

Sağlık Bilimleri Enstitüsü Müdürlüğü'ne;

Bu çalışma, jürimiz tarafından Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı, Tıp Bilişimi Programında Doktora tezi olarak kabul edilmiştir./..../.....

Tez danışmanı: Prof. Dr. Osman SAKA
Akdeniz Üniversitesi Tıp Fakültesi
Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı

Üye: Prof. Dr. Sadi ÖZDEM
Akdeniz Üniversitesi Tıp Fakültesi
Tıbbi Farmakoloji Anabilim Dalı

Üye. Prof. Dr. Ergun KARAAĞAOĞLU
Hacettepe Üniversitesi Tıp Fakültesi
Biyoistatistik Anabilim Dalı

Üye: Yrd. Doç. Dr. Kemal Hakan GÜLKESEN
Akdeniz Üniversitesi Tıp Fakültesi
Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı

Üye: Yrd. Doç. Dr. Uğur BİLGE
Akdeniz Üniversitesi Tıp Fakültesi
Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı

ONAY:

Bu tez, Enstitüsü Yönetim Kurulunca belirlenen yukarıdaki jüri üyeleri tarafından uygun görülmüş ve Enstitü Yönetim Kurulu'nun / / 2012 tarih ve /..... sayılı kararıyla kabul edilmiştir.

Prof. Dr. İsmail ÜSTÜNEL
Enstitü Müdürü

Saęlık Bilimleri Enstitüsü Kurulu ve Akdeniz Üniversitesi Senato Kararı

Saęlık Bilimleri Enstitüsü'nün 22/06/2000 tarih ve 02/09 sayılı enstitü kurul kararı ve 23/05/2003 tarih ve 04/44 sayılı senato kararı gereęince "Saęlık Bilimleri Enstitülerinde lisansüstü eğitim gören doktora öğrencilerinin tez savunma sınavına girebilmeleri için doktora bilim alanında SCI tarafından taranan dergilerde en az bir yurtdışı yayın yapması gerektięi" ilkesi gereęince yapılan yayınların listesi ařaęıdadır (orjinalleri ekte sunulmuřtur).

1. Oęus C, Ket S, Bilgen T, Keser I, Cilli A, Gocmen AY, Tosun O, Gumuslu S. Insertion/deletion polymorphism and serum activity of the angiotensin-converting enzyme in Turkish patients with obstructive sleep apnea syndrome. Biochem Genet 2010, 48(5-6):516-523.

ÖZET

Genomik ve bilgisayar teknolojilerindeki gelişmelerin ardından genetik arařtırmalarda biyoinformatik disiplininin rolü daha da artmıřtır. Karmařık arařtırma sorularının çözüümü için yeni veri analiz tekniklerine duyulan ihtiyaç, biyoinformatik profesyonellerinin karřılařtıđı en önemli sorundur.

Hasta ve kontrol grupları arasındaki allel frekansı farklılıklarını arařtırmada en çok kullanılan arařtırma tasarımı Genom Çađında İlgi Çalıřmalarıdır (GWAS). İlk GWAS arařtırması 2005 senesinde gerçekteřirilmifitir ve o zamandan beri gittikçe artan sayıda GWAS arařtırması yayınlanmıřtır. GWAS, tüm kısıtlılıklarına rađmen vaka – kontrol temelli genetik arařtırmalar için en önemli yöntemlerden birisidir.

Genomik verilerin analizi için yakın zamanda önerilen en yeni yöntemlerden birisi de Çok Deđişkenli Uzaklık Matrisi Regresyonu'dur (MDMR). MDMR, özellikle bađımsız deđişken sayısının olgu sayısına kıyasla çok fazla olduđu durumlar için oldukça güçlü bir analiz yöntemidir. Bireyler arasındaki uzaklıklar hesaplanarak bir bađımlı deđişkenler matrisi oluşturulur ve genetik bilgi içeren deđişkenler de modele bađımsız deđişken olarak eklenebilir. Böylece bađımsız deđişkendeki varyasyon miktarı açıklanmaya çalıřılır. Yakın zamana kadar, yöntemin genetik alanındaki kullanımını gen ekspresyonu verisi ile sınırlı kalmıřtır.

Bu çalıřmada GWAS yaklaşımı ve MDMR yöntemi birlikte bipolar bozukluk veri setine uygulanmıřtır. İlk basamakta GWAS yapılarak bipolar hastası bireylerle kontrol olgularını birbirinden ayıran istatistiksel olarak anlamlı SNPler bulunmaya çalıřılmıřtır. İkinci basamakta ise bipolar bozukluk hastalarına MDMR uygulanmıř ve hastalığın seyrine etki etmesi olası fenotip X genotip etkileřimleri incelenmifitir. En son basamakta ise bu iki uygulamanın sonuçları kıyaslanmıř ve kesiřim SNPleri arařtırılmıřtır.

Bazı kısıtlılıkları olmasına rađmen MDMR yöntemi, genetik faktörlerin kompleks hastalıklara etkisini incelemek için güçlü bir analiz aracıdır. Yöntemin esnekliđi hem gen ekspresyonu verisi ile hem de SNP verisi ile genomik arařtırmaların yürütülmesine olanak tanımaktadır. Bilindiđi kadarıyla bu çalıřma MDMR yönteminin fenotip verisi ile oluşturulan uzaklık matrisleri ile ilk uygulamasıdır. Ayrıca yöntem bipolar bozukluk verisi için ilk kez kullanılmaktadır.

GWAS ve MDMR analizleri sonucunda toplam 12 SNP ortak olarak bulunmuřtur. Bu SNPler hem hastalığın ortaya çıkması hem de seyri açısından önemlidirler. Ayrıca analizler sonucunda istatistiksel anlamlılıđı gösterilen bazı SNP ve genlerin önceki arařtırmalarda da bipolar bozukluk ile iliřkili olduđu gösterilmifitir.

Sonuç olarak, MDMR etkin bir analiz aracıdır ve kısıtlılıkları üzerinde çalıřılarak genomik arařtırmalarda verimli řekilde kullanılabileceđi düşünölmektedir.

Anahtar kelimeler: GWAS, SNP, bipolar bozukluk, fenotip, genotip, uzaklık matrisi, çok deđişkenli uzaklık matrisi regresyonu

ABSTRACT

Following the advances in genomics and computer technologies, the role of bioinformatics in genetic studies dramatically increased. The need for new data analysis techniques to handle complex problems is one of the most challenging tasks that bioinformatics professionals have to deal with.

Most frequent study design to compare allele frequency distances between patients and controls is called GWAS (Genome Wide Association Study). Since the first GWAS in 2005, increasing numbers of similar research were published. GWAS has its own strengths and weaknesses and still is a major design for most case – control genetic research.

One of the recent data analysis methods proposed for the analysis of genomic data is Multivariate Distance Matrix Regression (MDMR). MDMR is a statistically powerful tool, especially for the problems where the number of independent variables exceeds the total number of cases. A dependent variable matrix is calculated based on statistical distances between individuals and genetic information is added to the model as independent variables to explain the variation in the distance matrix. Until recently, its use in genetic research was limited with gene expression studies.

This study utilizes traditional GWAS approach and MDMR method on a bipolar disorder data set. As the first step, we applied the GWAS technique to determine the statistically significant SNPs those may distinguish bipolar patients and control individuals. In the second step, we applied MDMR model to understand the phenotype X genotype interactions those may affect the disease course in bipolar disorder. As the last step, we compared the outcomes collected in previous steps and determined the overlapping SNPs.

Despite several limitations, MDMR is a useful analysis tool, especially for investigating the effects of genetic factors on complex diseases. Flexibility of the method is an advantage as it can be used in both gene expression and SNP based genomic research. To our knowledge, this study is the first MDMR application on phenotype based distance matrices. Moreover, this is the first study to apply the method on bipolar disorder data.

A total of 12 SNPs were common in both GWAS and MDMR runs. These SNPs are crucial in terms of both occurrence and course of bipolar disorder. Several SNPs and genes those found to be statistically significant in our analysis were previously reported to be associated with bipolar disorder.

As a conclusion, MDMR is a useful tool and it should be tweaked in terms of its limitations so that it can be used more effectively in genomic research.

Key Words: GWAS, SNP, bipolar disorder, phenotype, genotype, distance matrix, multivariate distance matrix regression

TEŐEKKÜR

Danıőmanım sayın Prof. Dr. Osman SAKA'ya, yardımlarından ötürü Prof. Dr. Nicholas SCHORK'a ve Dr. Ondrej LIBIGER'e çok teőekkür ederim.

Bu çalıőmanın var olabilmesi için en önemli desteęi veren babam Muhittin TOSUN'a sonsuz teőekkür ederim.

Son olarak, verilen emeęin çok büyük bir kısmına sahip olan çalıőma arkadaőım Mehmet Kemal SAMUR'a özel olarak teőekkürlerimi sunarım.

İÇİNDEKİLER

	Sayfa
ÖZET	v
ABSTRACT	vi
TEŞEKKÜR	vii
İÇİNDEKİLER DİZİNİ	viii
SİMGELER ve KISALTMALAR DİZİNİ	x
ŞEKİLLER DİZİNİ	xi
ÇİZELGELER DİZİNİ	xii
GİRİŞ	1
GENEL BİLGİLER	4
2.1. Genetik ve Biyoinformatik Tarihçesindeki Temel Adımlar	4
2.2. Genom Çapında İlişki Araştırmaları	6
2.2.1. GWAS Nedir?	6
2.2.2. GWAS Araştırmalarının Kısıtlılıkları	9
2.3. Çok Değişkenli Uzaklık Matrisi Regresyonu	12
2.4. Bipolar Davranış Bozukluğu ve Genetiğine İlişkin Araştırmalar	14
GEREÇLER VE YÖNTEMLER	21
3.1. Veri Seti	21
3.1.1. Bipolar Bozukluk Hastaları	21
3.1.2. Kontroller	22
3.1.3. Klinik Değerlendirme	22
3.1.4. Genotipleme ve Kalite Kontrolü	22
3.2. İstatistiksel Yöntemler	23
3.2.1. Genom Çapında İlişki Araştırması	23
3.2.2. GWAS Sonuçlarının SNP'den Gen Düzeyine İlerletilmesi	24
3.2.3. Çok Değişkenli Uzaklık Matrisi Regresyonu	25
3.2.4. GWAS ve MDMR Sonuçlarının Birlikte Değerlendirilmesi	32
BULGULAR	33
4.1. GWAS Bulguları	33
4.2. MDMR Bulguları	45
4.3. GWAS ve MDMR Bulgularının Birleştirilmesi	50
TARTIŞMA	52
SONUÇLAR	59
KAYNAKLAR	60

ÖZGEÇMİŞ

74

EKLER

75

EK-1: Insertion/deletion polymorphism and serum activity of the angiotensin-converting enzyme in Turkish patients with obstructive sleep apnea syndrome.

SİMGELER VE KISALTMALAR DİZİNİ

DNA	: Deoksiribonükleik asit
GWAS	: Genome Wide Association Study
SNP	: Single Nucleotid Polymorphism
MDMR	: Multivariate Distance Matrix Regression
LD	: Linkage disequilibrium
MANOVA	: Multivariate Analysis of Variance
MANCOVA	: Multivariate Analysis of Covariance
DSM-IV	: Diagnostic and Statistical Manual of Mental Disorders-IV
ABD	: Amerika Birleşik Devletleri
NIH	: National Institute of Health
NCBI	: The National Center for Biotechnology
dbGaP	: The Database of Genotypes and Phenotypes
GAIN	: Genetic Association Information Network
BPI	: Bipolar I
SABP	: Schizo-Affective, Bipolar Type
NIMH	: The National Institute of Mental Health
DIGS	: Diagnostic Interview for Genetic Studies
PVE	: Proportion of Variation Explained
FDR	: False Discovery Rate

ŞEKİLLER DİZİNİ

Şekil	Sayfa
4.1. Tüm SNPLer için GWAS ile Elde Edilen p Değerlerine ait Q-Q Grafiği	33
4.2. Tüm SNPLer için Kromozomlara Göre GWAS Sonuçlarına ait Anlamlılık Grafiği	39
4.3. Tüm Genler için GWAS ile Elde Edilen p Değerlerine ait Q-Q Grafiği	43
4.4. Tüm Genler için Kromozomlara Göre GWAS Sonuçlarına ait Anlamlılık Grafiği	44
4.5. Anlamlı Bulunan Genlerin Kromozomdaki Pozisyonlarına Göre Gösterimi	45
4.6. Hastalara ait Uzaklık Matrisinin Grafiksel Gösterimi	46
4.7. MDMR Analizi Sonucunda Tüm SNPLer'e ait F İstatistiği Değerlerinin Histogramı	47

ÇİZELGELER DİZİNİ

Tablo		Sayfa
2.1.	Bazı Hastalık ve Fenotiplere İlişkin Tahmin Edilen Kalıtsallık Yüzdeleri	9
4.1.	GWAS Analizi Sonucunda $p < 0,0001$ Düzeyinde Anlamlı Olduğu Belirlenen Toplam 91 SNP'e ait Özellikler	35
4.2.	GWAS ile Hesaplanan SNP p Değerlerinin Gen Düzeyinde Değerlendirilmesi	40
4.3.	MDMR Sonuçlarına Göre İstatistiksel Olarak Anlamlı Bulunan İlk 100 SNP	48
4.4.	GWAS ve MDMR Analizlerinde En Üst 2.500 SNP Arasında Ortak SNPler	51

GİRİŞ

Bilgi teknolojilerindeki hızlı gelişimin etkisi, her alanda olduğu gibi biyolojik bilimler ve tıp alanlarında da önemli değişimler ve yenilikler getirmiştir. Bu teknolojilerin sağladığı olanaklarla birlikte gittikçe artan miktarlarda veri elde edilmektedir. Elde edilen bu verilerin saklanması, işlenmesi, çözümlenmesi, sunulması ve iletilmesi, yine bilişim teknolojilerinin yardımı ile daha hızlı ve kolay bir şekilde yapılabilmeye başlanmıştır.

Bilişim alanındaki gelişmelere paralel olarak istatistiksel yöntemler de zaman içerisinde çağın gereksinimlerine uygun biçimde gelişmektedir. Bilinen yöntemlerin uygulanabilirliği yüksek kapasiteli bilgisayarlar ve işlevsel yazılımlar sayesinde kolaylaşırken yeni çözümlene yöntemleri de geliştirilmekte, bu yeni yöntemler ışığında gerçekleştirilen yeni bilimsel çalışmalar sonucunda da sağlık alanındaki bilgi birikimi hızla artmaktadır. Bu çözümlenmeler sonucunda tıpta kanıta dayalı uygulamaların oranları hızla yükselmektedir. Tıp alanında yapılan bilimsel çalışmaların ve yayınlanan makalelerin sayısı son senelerde oldukça büyük artış göstermiştir.

Bilgi teknolojilerinin tıp alanındaki önemli etkilerinden birisi de genetik çalışmalar üzerinde olmuştur. Kişilere ait genetik verilerin toplanabilmesi ve bu veriler ile birçok hastalık arasındaki olası ilişkilerin araştırılmaya başlanması ile birlikte hem mevcut veri tabanlarının boyutlarında büyük artışlar gerçekleşmiş hem de bu büyük veri tabanlarındaki verilerin çözümlenmesine yönelik yöntemler geliştirilmesi gereksinimi artmıştır. Kişisel genetik bilgileri, demografik bilgileri, hastalık, tanı ve tedavi bilgilerini ve bireye ait fenotipik özellikleri içeren veri tabanlarının çözümlenebilmesi için yeni istatistiksel yaklaşımlar ortaya çıkmaktadır.

DNA mikro dizi analizleri ve proteomik platformlar gibi ileri teknoloji kullanılarak gerçekleştirilen yöntemler ile araştırmacılar oldukça karmaşık ve büyük veri toplulukları üzerinde genetik faktörlerin etkilerini incelemek durumunda kalmaktadırlar. Bu gereksinim, birden çok disiplinin işbirliğini artırmış ve biyoinformatik alanının artan bir ivme ile önem kazanmasına neden olmuştur. On binlerce gen ve proteinin ekspresyonuna ilişkin verilerin eş zamanlı olarak çözümlenmesi ve bu verilerden anlamlı bilgilerin elde edilmesi bu yeni teknolojiler ve teknikler yoluyla gerçekleşmektedir [1, 2]. Bununla birlikte bu teknolojiler sayesinde üretilen büyük miktarlardaki verinin istatistiksel ve biyolojik olarak önemliliğini tespit etmek de oldukça zordur [3]. Bu sebeple, çok sayıda akıllı ve

kullanışlı veri çözümlene stratejileri geliştirilerek çok boyutlu genomik verilerin analizine yönelik ilerlemeler kaydedilmektedir [4].

Özellikle genetik verilerin analizine yönelik geliştirilmiş olan çözümlene yöntemleri, birçok hastalığın incelenmesinde kullanılmaktadır. Bu yöntemler temel olarak bireylerin genetik verilerini de içeren geniş veri tabanlarında uygulanan istatistiksel yaklaşımlardır. Veri madenciliği uygulamalarının da katkısı sonucunda bu tip büyük hastalık veri tabanlarından yeni bilgiler üretilmekte ve bilimsel çalışmalar sonucunda hastalıkların tanı ve tedavi süreçlerinde önemli ilerlemeler kaydedilmektedir. Bu yolla çok sayıda hastalığın genler vasıtasıyla nesilden nesle ne şekilde taşındığı anlaşılmıştır. Gelecek nesillerin ebeveynlerinin taşıdıkları hastalıkların etkisinde kalacağı düşünülerek çeşitli önlemlerin alınması yoluna gidilmektedir. Genetik alanındaki gelişmeler hastalıkların etiyolojisinin anlaşılması, tanı ve tedavi süreçlerinin şekillendirilmesi bakımından yeni imkânlar yaratmaktadır.

Yukarıda bahsedilen klinik problemlerden birisi de bipolar davranış bozukluğu (bipolar disorder) olarak bilinen psikiyatrik hastalıktır. Bipolar bozukluk, kişinin yaşam kalitesini etkileyen ve sıklıkla da hayati tehlike arz edebilen bir hastalıktır. Toplam dünya nüfusunun yaklaşık %1'ini etkilemektedir. Hastaların yaşam süreleri boyunca %17 oranında intihar etme sıklığı görülmektedir [5].

Her ne kadar bipolar bozukluğun nedenleri tam olarak belirlenememiş olsa da, genetik faktörlerin önemli bir rolü olduğu düşünülmektedir. Bipolar bozukluk tanısı almış kişilerin %80-90'ında ailede bipolar veya majör depresyon öyküsü vardır [5]. Bununla birlikte, aile öyküsü olanların hastalığa ilişkin genetik eğilime mutlaka sahip olacağı söylenemeyeceği gibi, genetik yatkınlığı taşıyıcılar bile hastalık semptomlarını kesin olarak geliştirecekleri de iddia edilemez. Mevcut yöntemler ile hastalığın tanısı standartlara dayalı şekilde konulabilmekte ve doğru tedavi ile hastaların günlük yaşamlarındaki stabilizasyonu sağlanabilmektedir.

Genotipleme teknolojilerindeki mevcut gelişmeler ve kamu erişimine açık veri tabanlarının oluşturulması ile hastalıkların genetik nedenlerini araştıran geniş ölçekli “Genom Çapında İlişki Çalışmaları” (GWAS) yaygınlaşmaya başlamıştır. Son 10 yılda bipolar bozukluk ile ilgili GWAS araştırmalarının sayısı önemli ölçüde artmıştır. Çalışmalar sonucunda 200'ün üzerinde gen veya SNP'in (Tek Nükleotid Polimorfizm - Single Nucleotid Polymorphism) hastalık riskini artırdığı gösterilmiştir. Bu genlerden bazıları (DAOA, Alg9, DGKH, ANK3, CACNA1C, DOK5, RGS4 vb.) en az iki çalışmadan replike edilebilmiştir. Ayrıca, tüm otozomal kromozomların bipolar bozukluk ile ilişkili olduğu ve X kromozomunun da bazı ilişkili genleri (MAOA, NLGN3 ve HTR2C) taşımakta olduğu gösterilmiştir [5-12]. Özetle, bu çalışmalar bipolar bozukluk genetiğine ilişkin daha çok bilgiye erişmemizi sağlamış da olsa hastalıktan tek başına sorumlu olan tek bir gen / SNP tespit edilememiştir.

Bireylerin genetik yapıları çok çeşitlilik göstermektedir. Bu bakımdan sadece hastaları sağlıklı bireylerden ayırabilmek değil, hasta bireyler arasındaki ortak fenotipik özelliklerin belirlenebilmesi de biyoinformatiğin bir diğer amacıdır. Aynı hastalığa sahip bireylerin birbirlerinden farklı seyreden özelliklerinin tespit edilebilmesi ve bu özelliklere neden olabilecek genetik faktörlerin bulunabilmesi kişiselleştirilmiş tedavi yöntemleri bakımından oldukça önemlidir. Bireyler arasındaki genetik çeşitliliğe katkıda bulunan veya bu çeşitlilik ile ilişkili olan faktörlerin belirlenmesinde kullanılan istatistiksel yöntemler gün geçtikçe çoğalmaktadır. Bu yöntemlerin gelişmesi yeni moleküler analizler ve teknolojiler ile birlikte daha da hızlanmıştır [13].

Bu yöntemlerden biridi de Çok Değişkenli Uzaklık Matrisi Regresyonu'dur (Multivariate Distance Matrix Regression - MDMR). Bu yöntem, esas olarak bireylerden toplanmış olan gen ekspresyonu veya SNP verilerinin çalışmadaki her birey çifti için benzerlik / ayrılık özelliklerini yansıtmak üzere benzerlik veya uzaklık matrislerinin yapılandırılması yaklaşımına dayanır [14]. Bu yaklaşım gen ekspresyonu ve SNP analizi çalışmasında kullanılmaya başlanan bir yaklaşımdır [15-18]. Yöntem, bir uzaklık matrisi ile gen ekspresyonu veya SNP verileri toplanmış olan bireylerden alınan tahmin ediciler arasındaki varyasyonların ilişkilerini test eder. Benzerlik veya uzaklık matrisleri içerisindeki örüntülere ilişkin hipotezleri test ederek çok büyük veri gruplarındaki tanımlaması zor ilişkilerin ortaya çıkarılmasında kullanılır. Oldukça esnek bir yöntem olup mevcut tek değişkenli yaklaşımların ötesinde çözümlene olanakları sunar [14].

Bu özellik göz önüne alınacak olursa, MDMR ile benzer davranış gösteren gen veya SNP kümelerinin tanımlanması ve böylece bireylerin benzerlik veya farklılıklarının incelenmesi mümkün olacaktır [14]. Bu yöntemin bilinen klasik kümeleme yöntemlerinden en büyük farkı, klasik yöntemlerin sıklıkla tek boyutlu olması, buna karşın MDMR'da hem genlerin hem de fenotiplerin eş zamanlı değerlendirilmesinin yapılabilmesidir. Yani; bir tek gen / SNP birden çok biyolojik özelliği etkileyebilir ve bir biyolojik özellik de aynı anda birden çok gen / SNP tarafından etkilenebilir.

Araştırmamızda amaç; bahsedilen çift yönlü ilişkilerin ortaya konması için MDMR yönteminin kullanılarak bipolar bozukluk hastalarının fenotipik özellikler bakımından değerlendirilmesi ve genetik farklılıklarının neden olduğu etkilerin incelenmesidir. Ayrıca standart GWAS yaklaşımı ile bipolar bozukluk hastalarının normal kontrol olgularından SNP ve gen seviyesinde gösterdikleri farklılıklar incelenecektir. Böylece MDMR ve GWAS bulguları birleştirilerek hastalığın ortaya çıkışı ve seyri için önemli bulunan ortak SNPler'in varlığı araştırılacaktır.

GENEL BİLGİLER

2.1. Genetik ve Biyoinformatik Tarihçesindeki Temel Adımlar

Gregor Mendel 1866 senesinde kalıtımla ilgili teorilerini bahçe bezelyesi ile gerçekleştirdiği denemelere dayanarak yayınladı. Her ne kadar bu yayın 1900ler'in başına kadar yeterince dikkate alınmamış da olsa, Mendel'in çalışması aslında modern genetik biliminin temellerini oluşturmaktadır. Mendel, beyaz çiçekli ve mor çiçekli bitkileri çaprazladığında karışım renkte çiçeklere sahip bitkiler yerine yine mor çiçekli bitkiler elde ettiğini gözlemledi. Birbirini takip eden birkaç çaprazlama sonucunda baskın ve çekinik karakterli kalıtıma dayanan fikirlerini şekillendirmiş oldu. Baskın karakterler (bu örnekte mor çiçekli olmak) çekinik karakterin (beyaz renkte çiçek) gizli kalmasını sağlamaktaydı [19].

Daha sonra, 1900ler'in başlarında Sir Archibald Garrod "metabolizmanın doğuştan gelen hataları" (örneğin albinizm) şeklinde adlandırdığı çeşitli hastalıkları tanımlarken Mendel'in denemelerindeki benzer bir kalıtım örüntüsü ileri sürdü. Böylece hastalıkların kalıtımı açısından önemli bir temel inşa etmiş oldu [20]. Garrod, otozomal çekinik kalıtımın doğasını bu koşullara dayanarak açıkladı (hastalığın ortaya çıkması için normal olmayan genetik varyant ya da "allel" çiftinin gerekli olduğu). Mendel prensiplerine dayanarak, hastalık alleli bakımından heterozigot olan iki bireyin çiftleşmesi ile doğacak çocuğun dörtte bir ihtimalle her iki çekinik hastalık allelini taşıyacağını ve hasta olacağını ileri sürdü [19].

Wilhelm Johannsen 1909 senesinde "gen" terimini ilk kez kullandı ve 1915'de de Thomas Morgan Hunt kromozomlar üzerinde yer alan genlerin kalıtımın yapı taşları olduğunu ileri sürdü [21]. Takip eden senelerde otozomal çekinik karakterde kalıtımı gerçekleşen çeşitli hastalıkların varlığına ilişkin örnekler sunuldu. Bunların arasında fenilketonüri ve kistik fibrozis de vardı. Aynı zamanda otozomal baskın hastalıklar da (hastalığın ortaya çıkması için sadece bir tane normal olmayan allelin varlığının yeterli olduğu) tanımlanmaktaydı. O dönemde ortaya çıkarılan otozomal baskın karakterli bu hastalıkların belli başlıları Huntington hastalığı ve Marfan sendromuydu. Bu şekilde bir kalıtım örüntüsü gösteren hastalıklar "Mendeliyen hastalıklar" olarak adlandırıldılar ve bu hastalıkların analizi ve anlaşılması görece daha kolaydı.

Bununla birlikte, benzer hastalıkların kalıtım modellerine dayanılarak oluşan genel kanı da gelecekte yapılacak araştırmaların esas çıkış noktasını oluşturdu:

genler belirleyicidir ve eğer bir birey bir hastalığa ilişkin allele sahipse, o kişi istenmeyen bu özelliği veya hastalığı sonraki nesillere taşımış olur. Zaman ilerledikçe bu yargının tam olarak doğru olmadığı görüldü. Çünkü yakın zamanda yürütülen genomik temelli araştırmalarda görülmektedir ki spesifik bir genetik varyasyona sahip olmak, hastalığın ortaya çıkmasını veya taşınmasını kesin olarak sağlamayabilir. Çünkü çevresel ve / veya diğer genetik faktörler de hastalıkların ortaya çıkmasında önemli birer belirleyicidirler [19, 22].

Hans Winkler 1920 senesinde gen ve kromozom kelimelerini birleştirerek “genom” kavramını ortaya attı. Genom, bir organizmadaki tüm genleri temsil eden bir kavramdı [23]. DNA'nın kromozom ve genlerin en temel bileşeni olduğu bulgusunu ortaya çıkaran James Watson ve Francis Crick isimindeki iki bilim adamı, Rosalind Franklin'in de katkılarıyla 1953 yılında DNA'nın çift sarmallı yapısını keşfettiler. Marshall Nirenberg ise 1966 senesinde “genetik kodu kırarak” DNA'nın protein yapımındaki rollerini tanımladı [19, 24].

Tüm bu önemli bulguların ışığında, genomik profillerin tıp ve diğer alanlarda ne şekilde kullanılabileceğine odaklanan yeni çalışmalar yapılmaya başlandı. Örneğin 1978 yılında bir biyoteknoloji firması olan Genentech, genetik mühendisliği yöntemleri ile diyabet tedavisinde kullanılmak üzere, bakteriden insan insülini üretmeyi başardı (rekombinant teknoloji kullanılarak üretilen ilk ilaç) [19, 25]. Ayrıca 1983 senesinde bilim adamları Huntington hastalığından sorumlu olan geni buldular. Böylece hastalıklar için genetik test rutin olarak kullanılmaya başlandı. Sir Alec John Jeffreys 1984 senesinde DNA profillemeye ya da “parmak izi” yöntemini geliştirerek genetik bilginin babalık testlerinde ve adli vakalarda kullanımına olanak tanıdı. Bu yöntemin adli olarak kullanıldığı ve DNA kanıtlarının incelenmesi sonucunda zanlının beraat ettiği ilk dava 1986 yılında Richard Buckland'ın suçsuzluğunun kanıtlandığı dava olarak tarihe geçti. Daha sonra dünyanın en büyük gen çipi üretici firmalarından biri olacak Affymetrix'i kuracak olan Stephen Fodor 1989 senesinde ilk DNA “mikro dizisini” ve tarayıcısını üretti. Bu buluş ileriki yıllarda yüz binlerce genetik varyantın eş zamanlı olarak test edilebilmesine olanak tanıdı ve GWAS döneminin de fitilini ateşlemiş oldu [19, 26].

Yakın zamanda ise iki temel araştırmanın öncülüğünde “post genomik” çağa giriş yapılmış oldu: Human Genome Project (HGP) ve International HapMap Project. HGP'nin amacı DNA'yı oluşturan tüm kimyasal baz çiftlerini tanımlayarak referans bir insan genomunu ortaya çıkarabilmektir. Ayrıca yaklaşık 25.000 adet genin lokasyonlarını belirleyerek insan genomunu keşfetmek hedeflenmişti. 2000 yılında öncül bir çalışma yayınlandı [27, 28] ve en son versiyon da 2003 yılında sunuldu. International HapMap Project'in amacı ise bireyler ve popülasyonlar genelinde genom çapında bir veri tabanı yayınlamak ortak insan sekans örüntülerini kataloglayabilmektir [29, 30]. HapMap projesinin esas hedeflerinden birisi de “ortak hastalık – ortak genetik varyasyon” hipotezine [31-34] uygun şekilde hastalıklara

neden olan genetik varyantların tanımlanabilmesini kolaylaştırmaktı. Bu hipotez daha sonraki arařtırmaların sonuçları ile birlikte revize edilmek durumunda kalmıřtır [33, 35-37]. Bu hipoteze gre kompleks hastalıkların genetik nedenleri, toplumun yaklařık %5’inde grlmekte olan ortak allelik varyantlarla iliřkilendirilmelidir. Bu varyantlar “risk faktrleri” ya da “hastalığa karřı hassaslık varyantları” olarak adlandırılır. Bu anlamda Mendel prensiplerinin deterministik yaklařımından farklıdır. Zaman ierisinde grlen teknolojik geliřimler ve maliyet azalmaları sonucunda, DNA mikro dizi teknikleri yz binlerce genetik varyantın aynı anda test edilebilmesine olanak tanımıřtır ve bu iki nemli projenin atıđı yol sayesinde ilk GWAS alıřması yapılabilmıřtir [38]. Bu ilk GWAS alıřmasının ardından gittike artan sayıda GWAS arařtırması yapılmıřtır. Her ne kadar GWAS arařtırmalarına zel nemli kısıtlılıklar olduđu bilinse de, tm bu arařtırmalar genomik ve kiřiselleřtirilmiř tıp alanının altyapısını hazırlamaktadırlar [19].

2.2. Genom apında İliřki Arařtırmaları

2.2.1. GWAS Nedir?

İnsan genom sekansı yaklařık olarak 3 milyar nkleotid bazından (ift sarmal yapı dřnlecek olursa 6 milyar baz) meydana gelmektedir. Her ne kadar bu dizinin %99’undan fazlası bireyden bireye farklılık gstermese de, farklılıkların ortaya ıktıđı grece az sayıdaki blge nemlidir. nk bu farklılıklar evresel / davranıřsal faktrlerle birlikte fenotipik deđiřkenliğe neden olmaktadır. DNA dizisindeki varyasyonlar farklı formlarda olabilir. Bu varyasyonlardan en sık grleni, bireylerin DNA dizisinin herhangi bir noktasında gsterdikleri tek baz temelli farklılıklardır. Bu farklılıklara genel olarak SNP adı verilir. SNPler ortalama 300 bazda bir blgeye denk dřecek Őekilde grlrler [39].

İnsan genomunda 10 milyondan fazla SNP olduđu dřnlmektedir. GWAS alıřmaları temel olarak yksek kapasiteli genotipleme (highthroughput genotyping) teknolojilerini kullanarak genomdaki SNP varyasyonlarını arařtırır ve bir hastalığa (veya belirli bir fenotipe) sahip olan bireylerle normal bireyler arasındaki allel frekansı farklılıklarını bulmayı amalar. GWAS arařtırmaları zellikle son 10 yılda teknolojideki geliřmeler ve maliyetlerdeki azalmalar sonucunda hastalıkların genetik nedenlerini inceleyen arařtırmaların en nemli yntemlerinden birisi haline gelmiřtir [40]. Hastalıklara kalıtsal yatkınlık ve yakalanmanın nedenlerini inceleyen bu alıřmalarda ok sayıda SNP’in genotip ve allel frekansları bakımından hasta ve kontrol grupları arasında istatistiksel bir farklılık gsterip gstermediđi eř zamanlı olarak arařtırılır [41-44].

On milyonun zerindeki blgenin genotiplenmesinin ok pahalı olacađı ve fazla zaman alacađı dřnlebilir. Fakat genom, bađlantı eřitsizliđi (Linkage Disequilibrium – LD) adı verilen yapısal bir zellik gstermektedir. Bir fenotipin genetik kaynađını arařtırabilmek iin yaklařık ka tane SNP incelenmelidir sorusuna

bağlantı eşitsizliği açısından bakmak gerekir. Buna göre belirli bir kromozom üzerinde DNA sekansının büyük bölümleri yüksek korelasyon gösterir. Bu yapısal özellik GWAS çalışmalarının maliyetini düşüren bir kısa yol sunar ve böylece korelasyon gösteren bölge içindeki temsilci SNPLer (“etiket” SNP) genotiplenerek aynı bölgeye ait diğer SNPLer’in ölçülmeyen bazıları tahmin edilebilir. Böylesi korelasyon gösteren bölgelere “haplotip” adı verilir.

Bilinen “etiket SNPLer’in” oranını araştıran bir çalışmada Afrika popülasyonunda yaklaşık 10.000 – 30.000 arası baz çiftinin Avrupa ve Asya popülasyonlarında ise yaklaşık 30.000 - 50.000 baz çiftinin ortak etiket SNP taşıdığı gösterilmiştir [45]. Bu bulgu HapMap projesinin de üzerinde çalıştığı bir araştırma ile daha da şekillenmiştir ve proje sonucunda yaklaşık 4 milyon SNP onaylanarak kataloglanmıştır [46]. Genel anlamda dünya popülasyonlarında var olduğu tahmin edilen 10 milyonun üzerinde SNP’den ortak olduğu düşünülen 2,8 milyonu da bu kataloglanan SNPLer arasındadır.

Bu sayede dünya genelinde biyoteknoloji firmaları arasındaki rekabet de artmış ve yüksek kapasiteli genotipleme araçları üretmek amacıyla teknolojik gelişmeler birbirini izlemiştir. Şu an itibarıyla Avrupa popülasyonunda yaklaşık 500.000 ve Afrika popülasyonunda da yaklaşık 1 milyon adet SNP’in etiket görevi üstlendiği ve böylece toplam SNPLer’in yaklaşık %80’inin tahmin edilebilmesini sağlayacakları kabul edilmektedir [47].

Bu şekilde tüm haplotipler genelinde yapılacak bir genotipleme ile 500.000 ile 1 milyon arasında SNP incelenerek tüm genomu ait yeterli bilgi toplanmış olur. Tüm genomdan toplanan ve sayısı genellikle 500.000’in üzerinde olan SNP verisi ile hastalığa neden olan varyasyonların tespiti amaçlanır.

Son yıllarda genotipleme teknolojilerinin maliyeti neredeyse 2.000’de birine kadar gerilerken sonuçların hassasiyeti yeni üretilen çipler sayesinde binde bir hatanın altına kadar gelişim göstermiştir. Tüm bu gelişmeler GWAS araştırmalarının yaygınlaşmasını sağlamıştır [48].

Şu ana kadar yapılan çok sayıda GWAS araştırması ile hastalıklarla ilişkili birçok polimorfizmin varlığı ampirik olarak kanıtlanmıştır [39, 42, 49-51]. 2005 senesinden beri sayısı gittikçe artan GWAS araştırmasının hemen hepsinde bu yöntem kullanılmıştır. Bu çalışmalara ilişkin bilgilere “<http://www.genome.gov/gwastudies>” web adresinden de erişilebilir [52-54].

İlk örneğinin yayınlandığı 2005 yılından beri [38] GWAS çalışmaları ile 40’ın üzerinde önemli hastalığa ilişkin genetik yatkınlığa neden olan çok sayıda spesifik DNA sekansları belirlenebilmiştir [39]. Bu bulguların birçoğu, önceden düşünülmeyen yeni aday genlere odaklanılmasına neden olmuş ve böylelikle yeni patofizyolojik hipotezler ortaya atılabilmektedir. GWAS yöntemi birçok açıdan

uygulama avantajları taşımaktadır. Artık milyonlarca insan DNA sekansları kataloglanabilmektedir ve aynı anda milyonlarca varyasyonun çabuk ve doğru bir biçimde belirlenebilmesi mümkün hale gelmiştir.

Nöropsikiyatrik bozukluklar, bilinen genetik kökenleri nedeniyle üzerinde en çok çalışılan konuların başında gelmektedir. Bugüne kadar başta otizm, bipolar bozukluk, şizofreni, majör depresif bozukluk ve dikkat eksikliği hiperaktivite bozukluğu olmak üzere birçok mental hastalık için çok sayıda GWAS araştırması yürütülmüştür [48].

En temel şekliyle bir GWAS araştırması; kalıtsal olduğu düşünülen fenotipi taşıdığı doğru bir şekilde tanılanmış olgularla (hastalar), bu hastaların karşılaştırılacağı yine net bir şekilde tanılanmış kontrol olgularını içerir. SNPLer'in frekansları farklı toplumlarda farklılıklar gösterebildiğinden, araştırmaya dâhil edilen olguların hepsinin aynı kökenden gelmesine (Avrupa kökenli, Asya kökenli gibi) dikkat edilmelidir. Varyasyonların nedenselliği homojen popülasyonlarda daha net görülebilmektedir. Her bir olgunun doku veya kan örneği çipler kullanılarak genotiplenir. Bu aşamada, toplanan verinin dikkatli bir kalite kontrol sürecinden geçmesi gerekir. Veriler yanlış pozitifliği veya yanlış negatifliği artırabilecek hatalardan, yanlış ölçmelere bağlı beklenmeyen uç değerlerden temizlenmelidir. Bireyler arasındaki akrabalık da bu yöntemlerde tespit edilebilir ve bu durumda bazı olgular araştırma dışı tutulabilir. Bu aşamada ortaya çıkabilecek bazı sorunlar da istatistiksel düzeltme yöntemleri ile ortadan kaldırılabilir [55]. Birçok çalışmada her bir SNP, fenotipe göre test edilir ve HapMap referans dizisi kullanılarak veri imputasyonu da yapılabilir [48, 50, 56, 57].

Sistemik taraf tutmanın önüne geçebilmek için hasta ve kontrol olgularının iyi tanımlanması ve dikkatle aynı popülasyon içinden seçilmesi gerekir. Nadir rastlanan veya klinik seyri ağır olan hastalıklar için vakalarla kontrollerin benzer yapıda olması durumu zor sağlanabilmektedir. Araştırılan fenotipe bağlı olarak, karşılaştırılacak grupların yaş, cinsiyet gibi demografik özellikler açısından benzer olması gerekebilir. Örneğin Alzheimer ile ilgili bir araştırma tasarımında kontrollerle hastaların yaş olarak birbirine yakın olması önemlidir. Eğer gen X çevre etkileşimi ile ilgili bilinen etkiler varsa, bu etkiler göz önünde bulundurulmalıdır (örneğin madde bağımlılığını araştıran bir çalışmada kontrol bireyleri daha önce maddeyi kullanmış fakat bağımlı olmamış bireylerden seçilebilir). GWAS araştırmasının istatistiksel gücü, araştırılan hastalığın yaygınlığına bağlı olarak da değişim gösterebilir. Eğer sık rastlanmayan bir hastalık inceleniyorsa (prevalansı %5'den az), kontrol olgularını klinik değerlendirmeye tabi tutmadan seçmek istatistiksel güçte az bir kayba neden olabilir. Fakat daha sık rastlanan bir hastalık için, kontrol olguları arasında hiç hasta birey olmamasına dikkat etmek gerekir [48, 57].

2.2.2. GWAS Arařtırmalarının Kısıtlılıkları

GWAS, genomdan elde edilen bilgiler ile yürütölen ve sık görölen, kompleks fenotiplerin ortaya çıkmasında etkili olan SNPler'in belirlenmesinde kuvvetli bir arařtırma çeřididir. Fakat GWAS ile tespit edilebilen hastalıęa yatkınlık riski ile (genellikle %1-10 arası), hastalıkların kalıtsal kökenlerine iliřkin tahminler arasında gözle göröölür bir fark vardır. (Tablo 2.1) [58, 59]. Bazı hastalıklara yatkınlıęın kalıtsallıęına iliřkin tahminler %50'nin üzerinde olabilmektedir. Ayrıca kronik, yařlanma ile ilgili hastalıklara kıyasla nöropsikiyatrik bozuklukların genetik temellerini arařtıran GWAS çalıřmalarında bu durum daha da belirgin olabilmektedir [60]. Bu eksiklięin belki de en öncelikli sebebi, az rastlanan varyantların istatistiksel olarak incelenebilmesi için çok büyük örneklemlere ihtiyaç duyuluyor olmasıdır. Etkisi küçük fakat önemli olan SNPler'i tespit edebilmek için çok fazla sayıda olguya ihtiyaç duyabilmektedir. Ek olarak, řu an itibarıyla genotiplemede kullanılan çiplerin büyük kısmı nadir varyasyonları (frekans %1'den daha az olan) tespit edemeyebilmektedirler. Son olarak, gen X gen etkileřimini belirleyebilmede mevcut yaklařımların düşük istatistiksel gücü, gen X çevre etkileřimini yeterli olarak inceleyemiyor olmamız, fenotipik heterojenlik ya da yeterince hassas ve belirgin olmayan tanı kriterleri ve epigenomik deęiřimler gibi faktörler de GWAS sonuçlarının tatmin edicilięini azaltabilmektedir [59].

Tablo 2.1. Bazı Hastalık ve Fenotiplere İliřkin Tahmin Edilen Kalıtsallık Yüzdeleri [19]

Hastalık / Fenotip	Tahmin Edilen Kalıtsallık Yüzdesi	Arařtırmalar
Obezite	77	[61, 62]
Koroner Kalp Hastalıęı	57	[63]
Tip II Diyabet	64	[64]
Kolorektal Kanserler	35	[65]
Prostat Kanseri	42	[65]
Meme Kanseri	27	[65]
Alzheimer	74	[66]
Uzun Ömürlölük	25	[67, 68]
řizofreni	81	[69]
Bipolar Bozukluk	85	[70, 71]
Ünipolar Depresyon	37	[72, 73]
Anksiyete Bozuklukları	32	[74, 75]

Bu faktörlerin çoęu nöropsikiyatrik bozukluklarda dięer bazı hastalıklara kıyasla daha önemli rol oynamaktadırlar. Ayrıca yine nöropsikiyatrik bozukluklar için "ortak hastalık – ortak genetik varyasyon" hipotezi yetersiz kalabilmektedir. Dolayısıyla böylesi hastalıkların genetik temellerini arařtırırken bu yaklařımın yerine gen X çevre etkileřimine daha fazla önem verebilecek arařtırma yöntemlerine gerek duyulduęu düşünölmektedir [37]. Tüm bu sıralanan sebepler yüzünden mental bozuklukları inceleyen GWAS arařtırmalarının klinikte etkin řekilde kullanılabilir

genetik bilgiyi sunduğunu tam olarak söyleyemeyiz. Bu sebeple, özellikle nöropsikiyatrik bozukluklar için olmak üzere, neredeyse tüm kompleks ve Non-Mendeliyen hastalıklar için genomik bilginin daha etkin olarak kullanılabilceği yeni biyoinformatik yaklaşımlarına ihtiyaç duyulduğı söylenebilir [19].

Yeni jenerasyon sekanslama araçlarının daha da gelişip maliyetlerinin de azalması ile yakın gelecekte DNA'nın tamamıyla incelenmesi araştırmalar açısından belki de bu sorunları ortadan kaldıracaktır. Bu şekilde, sadece “etiket SNPlar” veya “aday genler” değil, tüm nadir varyasyonlar, insersiyonlar, delesyonlar, kopya sayısı varyasyonları da tespit edilebilecektir [19]. GWAS araştırmaları, tüm genom araştırmalarına evrilecek ve maliyetler açısından daha ulaşılabilir olacaktır. Bu da özellikle yaşlanma ve nöropsikiyatrik bozukluklar açısından yepyeni bulgulara ulaşılmasını sağlayabilir. Örneğin Halaschek- Wiener ve arkadaşlarının yürüttüğü bir araştırmada 85 yaş ve üstü sağlıklı bireylerde yaşlanmayla ilgili 24 adet aday gen tamamıyla sekanslanmış ve tespit edilen istatistiksel olarak anlamlı varyasyonların %41'inin mevcut genetik veri tabanlarında kayıtlı olmadığı görülmüştür [76]. Bu anlamda, tekniklerin gelişmesi ve maliyetlerin düşmesi, bilgi birikimini artıracak ve dolaylı olarak bulguların daha kapsamlı ve hassas olmasına yol açabilecektir. Yani GWAS başta olmak üzere, bu güne kadar kullanılan biyoinformatik yöntemleri, süreç içerisinde hep evrilmek durumundadır ve bu anlamda genetik araştırmalarda yöntem her zaman için dinamiktir.

Yeni teknikler ve daha fazla sekanslama bilgisi içeren veriler ile şizofreni, bipolar bozukluk ve anoreksiya nervosa gibi hastalıkların genetiğine ilişkin bilgi birikiminin klinik kullanımda gittikçe daha fazla etkinleşeceği öngörülebilir [19]. Ayrıca veri setlerinin birleştirilmesine dayanan meta analizi gibi teknikler de en az teknolojik gelişmeler kadar önemlidir. Bu bakımdan elde edilen verilerin bilimsel kullanıma açık veri tabanlarında tutulması tercih edilmektedir. Böylelikle araştırmaların gücünü ve örneklem büyüklüğünü artırmak mümkün olabilecektir [77, 78].

Ayrıca gelecekteki GWAS araştırmalarından daha kullanışlı bulgular elde edebilmek için önerilebilecek bir diğer etkin strateji de klasik vaka – kontrol araştırmalarının yanı sıra, fenotip bilgilerinin de doğru şekilde toplanmasıdır. Çevresel faktörler de bir diğer önemli değişken grubudur. Gen X gen etkileşimlerini ortaya koyabilmek için a priori hipotez testleri kullanılabilir. Gen X çevre etkileşimini daha net ortaya koyabilmek için olguların ailelerinden de etkin veri toplamak gerekmektedir. Tüm bu önerilebilecek stratejiler, maliyet azalmaları, teknik alanda kaydedilen gelişmeler ve araştırma yöntemlerinin optimize edilmesi ile birlikte, kompleks fenotiplerin genetiğine ilişkin daha tatmin edici sonuçlara ulaşılacağını öngörebiliriz [19].

Bu yöntemin kullanıldığı çalışmalar ile hastalıkların ve bazı kompleks fenotipik özelliklerin kalıtsal nedenlerine ilişkin çok önemli bulgulara erişilmiş de olsa, GWAS araştırmalarının bazı önemli uygulama kısıtlılıkları da vardır [79]. Bunların en başında yöntemin SNPlar'ın / genom bölgelerinin tek başlarına gösterdiği etkiyi bulmaya odaklanması gelmektedir. Oysa kompleks fenotipik özellikler birçok SNP'in / genin etkileşimi sonucunda ortaya çıkmaktadır [43, 54, 80]. Bu etkileşime diğer genetik ve çevresel faktörler de katkıda bulunur [81].

Öte yandan test edilen SNP sayısı fazla olduğundan çoklu testler için yapılan istatistiksel düzeltmeler sonucunda az sayıda SNP anlamlılık düzeyini sağlayabilmektedir [43, 82]. Bir SNP'in genom çapında anlamlı bulunabilmesi için genellikle çok düşük p değerlerinde ($p < 10^{-6}$ ya da $p < 10^{-8}$ gibi) önemliliği ispatlanmak durumunda kalınmaktadır [83]. Bu sorunu giderecek optimal bir p değerinin tespiti de zordur [84]. Anlamlılık düzeyinin düşük tutulması, önemli bulunan SNP sayısını kısıtlayıp biyolojik ağlarda rol oynayan belirli polimorfizmlerin dikkate alınmamasına neden olur [85]. Yüksek tutulacak p değeri ise çok sayıda SNP'in önemli bulunmasına yol açacaktır ve bu durumda da yanlış pozitiflik artar [84]. Özetle, GWAS yaklaşımında etkisi anlamlı bulunmayan fakat biyolojik ağlarda etkileşimler sayesinde önemli rol oynayan SNPlar göz ardı edilmiş olur [41, 80, 81]. İstatistiksel olarak "gürültü" kabul edilen bu SNPlar'e ilişkin bilgilerin de değerlendirilebileceği yeni analiz yaklaşımlarına ihtiyaç olduğu savunulmaktadır [43, 49, 81, 86].

Her ne kadar GWAS yoluyla etkisi tespit edilen aday genlerden birçoğu hastalıkların patofizyolojisini anlayabilmede önemli protein yollarının keşfine olanak tanımış da olsa [50, 51, 80, 87-93], SNPlar'ın GWAS ile bulunan bireysel etkileri görece düşüktür (odds oranı $\approx 1,2 - 1,4$) [43, 54, 80, 83, 94]. Yani istatistiksel anlamlılık fenotipik varyasyonların küçük bir kısmını açıklayabilmektedir [54]. Bununla birlikte, bir GWAS sonucunda çok sayıda SNP hastalıkla ilişkili bulunabilir. Fakat bu polimorfizmlerden hangilerinin etkisinin nedensel olduğu, hangilerinin ise veriden kaynaklanan yanıltıcı bulgu olduğunu bilmek zordur [54]. Öte yandan nedensel etkileri olan SNPlar ile kuvvetli LD gösteren SNPlar de GWAS sonucunda önemli bulunabilir. Bu SNPlar'i de tespit etmek zordur.

GWAS çalışmaları ile önemliliği tespit edilen SNPlar'ın başka benzer çalışmalarda replike edilmesi de genellikle mümkün olmamaktadır. Bunun nedeni; SNPlar'ın bireysel etkilerinin düşük olması, kompleks fenotipik özelliklerin genetiğindeki heterojenlik ve GWASlar'da genom çapında anlamlılık için kabul edilen p değeri sınırının çok düşük olmasıdır [80, 84, 95, 96]. Bulguların replike edilememesi, o bulguya ilişkin kanıtın gücünü azaltır.

GWAS çalışmalarının bu anlatılan sebepler yüzünden gelecekte üç temel basamağa dayalı olarak evrimleşeceği öngörülebilir: SNP, gen ve yolak temelli ilişki analizleri [80].

2.3. Çok Değişkenli Uzaklık Matrisi Regresyonu

MMDR yöntemi, ilk olarak McArdle ve Anderson tarafından ekoloji alanındaki çalışmalardaki karmaşık veri yapısına uygulanmakta olan çok değişkenli analiz yöntemlerini (MANOVA, MANCOVA, kısmi ya da çoklu regresyon gibi) daha da geliştirebilmek amacıyla temelleri atılan bir yöntemdir [97]. Daha sonra Zapala ve Schork, bu yöntemi genetik alanına adapte ederek kullanımını test etmişlerdir [14]. Yöntem şu ana kadar özellikle gen ekspresyonu temelli araştırmalar olmak üzere birçok genomik araştırmada kullanılmıştır [15-18, 98]. Ayrıca bir görselleştirme aracı olarak da uygulanabilmektedir [99, 100].

Yöntem, bireyler arasında değişkenlere dayalı benzerlikleri veya uzaklıkları belirleyerek bir benzerlik / uzaklık matrisi oluşturma temeline dayanır. Oluşturulan bu matris, regresyon analizinde bağımlı değişken matrisi olarak kullanılır. Temel olarak, oluşturulmuş olan benzerlik / uzaklık matrisindeki varyasyon ile bağımsız değişkenler arasındaki ilişki incelenir. Bu yöntemde bağımlı değişken matrisini oluşturmada kullanılacak olan değişkenler ve bağımsız değişkenlerin ne olacağı seçimi, yöntemin uygulanabileceği alanların çeşitlilik göstermesini sağlar. Örneğin yöntem, ölçülen klinik değişkenleri bakımından bireylerin arasındaki matematiksel uzaklıklar bir matrise dönüştürülerek bu matristeki varyasyonun gen ekspresyonu değişiklikleri ile açıklanması için kullanılabilir. Yöntem, ağaç temelli gösterimler ve ısı haritaları gibi görselleştirme yaklaşımlarına uygundur. Matris içindeki olası örüntülerin test edildiği hipotezler ile kümeleme analizlerinin kısıtlılıklarından kaçınılmakta ve bağımsız değişken sayısının olgu sayısından çok daha fazla olduğu genetik çalışmalar için istatistiksel gücü fazla olan bir çözümlenme yöntemi sunmaktadır [14, 97, 98].

Yakın zamanda gerçekleştirilen genomik araştırmalarının birçoğu, kompleks hastalıkların ve fenotiplerin genomun tek bir bölgesindeki spesifik bir mutasyona bağlı olmadığını, böylesi fenotiplerin ortaya çıkmasında etkileşimlerin önemli olabileceğini göstermektedir. Böylesi kompleks fenotiplerde, tek başına etkisi istatistiksel olarak düşük olan, popülasyon bazında nadir görülen fakat bireyin taşıyabileceği başka benzer değişimlerle birlikte toplamsal etkisi önemli olabilecek çok sayıdaki genetik varyasyonun birlikte görülmesinin önemli olabileceği yönündeki hipotezler ağırlık kazanmaktadır [19, 37]. Organizmaların taşımakta oldukları genetik materyalin yanı sıra, çevre ile etkileşimin de belirleyici olduğu görülmektedir. Bu açıdan epigenomik alanı özellikle kompleks fenotiplerin nedenselliğini anlamada gün geçtikçe değer kazanmaktadır [59, 101].

Oluşturulmuş benzerlik / uzaklık matrisini açıklamak için uygulanabilecek bir modelde kullanılacak bağımsız değişkenler SNPLer'e ait genotip verileri de olabilir. Hastalardan toplanacak olan klinik anlamda değerli fenotip verisi ile uzaklık matrisleri oluşturulabilir. Böylece hastalar arasında görülen fenotip varyasyonlarına bağlı benzerlik veya farklılıkların SNP temelli kökenleri araştırılacak ve gen ve kromozom bazında genetik nedenler incelenebilecektir.

Uzaklık matrisleri birçok istatistiksel analiz yönteminde kullanılmaktadır. Bununla birlikte, uzaklık matrisinin oluşturulması için birçok farklı yöntem uygulanabilmektedir. Bu yöntemler, olgular arasındaki benzerlik veya farklılıklara odaklanabilir. Benzerlikler için en sık kullanılan uzaklık matrisi oluşturma yaklaşımı korelasyon matrisleridir. Farklılıklar için ise sıklıkla Öklit uzaklık matrisi kullanılmaktadır.

Uzaklık matrisi oluşturma yöntemlerinden başlıcaları şöyle sıralanabilir:

- Bray-Curtis farklılıklar matrisi
- Bray-Curtis karekök yöntemi
- Öklit uzaklığı
- Orloci Chord uzaklığı
- Ki kare uzaklığı
- Hellinger uzaklığı
- Gower farklılık matrisi
- Canberra uzaklığı
- Binomial tabanlı uzaklık
- Kulczynski farklılıklar matrisi
- McArdle-Gower farklılıklar matrisi
- Manhattan uzaklığı
- Jaccard farklılıklar matrisi

- Minkowski uzaklığı
- Chebyshev uzaklığı
- Cavalli-Sforza ve Edwards uzaklığı
- Wright uzaklık ölçütleri
- Latter uzaklığı
- Reynolds uzaklığı
- Nei'nin standart genetik uzaklığı
- Nei minimum uzaklık ölçütü

Önceki çalışmalarda MDMR çözümlenmeleri için farklı uzaklık ölçütlerinin farklı varyasyonlar ortaya çıkardığı görülmüştür. Fakat uzaklık ölçütü seçiminin bağımsız değişkenlerin matristeki varyasyonu açıklamada istatistiksel olarak önemli bir farka neden olmadığını gösteren araştırmalar yayınlanmıştır. Ayrıca yine önceki çalışmalarda istatistiksel güç açısından da uzaklık matrisi ölçütlerinin önemli farklılıklara neden olmadığı gösterilmiştir [14, 102, 103]. Yine de matrisin oluşturulmasında verilerin özellikleri ve alternatif yöntemlerin performansları dikkate alınmalıdır [104].

2.4. Bipolar Davranış Bozukluğu ve Genetiğine İlişkin Araştırmalar

Bipolar bozukluk veya iki uçlu duygu-durum bozukluğu, eskiden manik depresyon, manik atak veya manik depresif bozukluk olarak bilinen hastalıktır. Kişinin depresif eğilimlerin yoğun yaşandığı dönemlerle, taşkınlık, coşkunluk olarak tanımlanabilecek mani dönemleri yaşadığı duygu-durum bozuklukları sınıfını tanımlayan tanısız kategoridir. Bipolar bozukluk, kişinin yaşam kalitesini etkileyen ve sıklıkla da hayati tehlike arz edebilen bir hastalıktır. Bozukluk, önemli morbidite ve mortaliteye neden olup (intihar girişimleri başta olmak üzere) toplumsal ve mesleki işlevleri etkileyerek boşanma, işsizlik, yaşam kalitesinde azalma gibi olumsuz sonuçlara neden olabilmektedir [105]. DSM-IV (Diagnostic and Statistical Manual of Mental Disorders – IV) adlı tanı ve istatistik kriteri ile teşhis konur [105, 106].

Bipolar davranış bozukluğu, yüksek derecede kalıtsal temeli olan psikiyatrik bir hastalıktır. 20. yüzyılın başlarından itibaren yürütülen araştırmalar ve ikizlerle gerçekleştirilen bilimsel çalışmalar göstermiştir ki başta bipolar bozukluk

olmak üzere birçok mental rahatsızlık aile düzeyinde belirgindir ve kalıtsal bir temele sahiptirler [107, 108]. Bipolar bozukluk tanısı almış kişilerin %80-90'ında ailede bipolar veya majör depresyon öyküsü vardır [109]. Bipolar bozukluğun genel popülasyonda görülme sıklığı çeşitli çalışmalarda %1 ile %2 civarında rapor edilmektedir [5, 109]. Ortalama başlama yaşı 21'dir. Erkek ve kadınlarda görülme sıklığı benzerdir. ABD'de 18 yaş ve üstündeki yaklaşık 5,7 milyon birey (popülasyonun %2,6'sı) hastalıktan etkilenmektedir [106, 110, 111]. Ayrıca birinci dereceden bipolar bozukluk hastası akrabası olan bipolar tanısı almış bireylerde morbidite riskinin %10-20 arası olduğu gösterilmiştir [112].

En kayda değerleri bipolar I ve bipolar II bozukluk olmak üzere, birkaç bipolar bozukluk tipi tanımlanmıştır. Genellikle rahatsızlık depresyon, mani (bipolar I) veya hipomani (bipolar II) veya karma depresyon ve mani ataklarıyla karakterizedir. Atakların sıklığı hastadan hastaya değişmekle birlikte, bazı hastalar sık sık ataklar geçirebilmektedirler (hızlı döngüsellik) [113].

Bipolar bozukluk tanısı almış bireylerde mani ataklarından çok depresyon gözlenmektedir. Bu nedenle unipolar depresyon ile karıştırılarak yanlış tanılanabilmektedir. Bu hatalı tanı, semptomları kontrol altına almada yetersizliğe neden olmakla birlikte, aynı zamanda hastalığı kötüleştirebilen antidepresan tedavisi kullanımına da yol açabilmektedir [113].

Hastalığın DSM-IV kriterlerine göre ana bileşenleri manik depresif ataklar, manik ataklar hipomanik ataklar ve karma ataklardır.

Majör depresif ataklar 2 hafta boyunca depresif duygu-durum, ilgisizlik veya hiçbir şeyden zevk almama şeklinde tanımlanmaktadır. Tabloya iştah ve beden ağırlığında değişiklik, uyku bozuklukları, psikomotor ajitasyon veya psikomotor gelişme geriliği, bitkinlik veya enerji kaybı, suçluluk, konsantrasyonda azalma, yinelenen ölüm veya intihar saplantısı gibi semptomlar eşlik etmektedir. Majör depresif atak 2 aydan uzun sürebilir. Bu durumdaki hastalar, belirgin işlevsel bozukluklar, kendini değersiz hissetme saplantısı, intihar düşüncesi, psikotik semptomlar veya psikomotor gerileme gösterebilirler. Yine de bipolar olgularda görülen bu durum, unipolar depresyonda benzerleri görülen hastalık seyrinden iyi ayırt edilmelidir [113].

Manik ataklar 1 haftadan uzun süren (veya hastaneye yatış gerektirecek seviyeden daha kısa süreli) belirgin, normal dışı, sürekli keyifli, taşkın veya huzursuz mizaçlı, üstünlük duygusu, uykusuzluk, kısık konuşma, karmaşık düşünceler, ilgisizlik, ajitasyon, acı sonuçlara yol açabilen hedeflere yönelik davranışlarda artış (çılginca para harcamak, uygunsuz cinsel yaşam gibi) şeklinde tanımlanmaktadır. Hipomanik ataklar ağır derecede olmayıp sosyal veya mesleki işlevsellikte bozulma ve psikotik özellikler göstermemekte ve hastaneye yatışı

gerektirmemektedirler. Karma ataklar en az 1 hafta süren ve hemen hemen her gün hem manik hem de depresif semptomların varlığıyla tanımlanmaktadır [113].

Bipolar I bozukluk, her zaman olmasa bile, genellikle majör depresif semptomların eşlik ettiği bir veya birden fazla sayıda manik veya karma atakla karakterizedir. Hastalar psikotik semptomlar da gösterebilirler. Bipolar II bozukluk en azından bir hipomanik atakla (manik ve karma ataklar dışında) karakterizedir. Bipolar bozukluk ile majör depresif bozukluk arasındaki ana farklılık, mani veya hipomaninin varlığı veya yokluğudur. İki bozukluğun tedavi yaklaşımları belirgin derecede farklı olduğundan, bu ayrım önemlidir [113].

Eşlik eden psikiyatrik rahatsızlıklar bipolar bozukluğu daha karmaşık hale getirebilir ve klinik yaklaşımı zorlaştırabilir. Bu rahatsızlıklar arasında kişilik bozuklukları, alkol ve madde bağımlılığı veya kötüye kullanımı anksiyete bozuklukları da yer alabilir. Bir araştırmada, bipolar bozukluğu olan 60 hastanın 23'ü (%38) en büyük olasılıkla özseverlik, sınırda, anti sosyal kişilik, obsesif-kompulsif bozukluk veya kaçınma davranışı olmak üzere en azından bir kişilik bozukluğu için tanısal kriterleri yerine getirmişlerdir [114]. Bu bozukluklar veya kişilik özelliklerinin varlığı, bipolar bozukluğu olan hastaların başlangıç durumlarına dönüp dönmediklerini belirlemekte hekimleri zorlayabilmektedir. Semptomları tedavi edilemeyecek kadar karmaşık veya şiddetli olabilir. Bipolar ve kişilik bozuklukları olan kişilerde intihar riski, yalnız bipolar bozukluğu olanlara göre daha yüksektir [113].

Bipolar bozukluğu olan erişkin ve çocuklarda birçok benzer semptom saptanmasına rağmen, bu yaş gruplarına özel semptomlarda birkaç ana farklılık saptanabilir. Örneğin çocuklarda huzursuzluk, üstünlük kompleksi veya coşku daha sık görülmektedir [115].

Bipolar bozukluğu olan kişilerde alkol ve madde kötüye kullanımı veya bağımlılığı riski artmış olup, bipolar I bozukluğu olanlarda riskin %61 kadar daha yüksek olduğu tahmin edilmektedir [116]. Hastalar alkol veya madde kullanımının tetiklediği depresyonlar veya maniye benzer klinik rahatsızlıklar gösterebilirse de, daha yakından incelemede bu kişilerin gerçekte bipolar bozukluğu olmadığı anlaşılabilir [117]. Gerçekten bipolar bozukluğu olan hastalarda yalnızca bir duygu-durum bozukluğu sırasında veya ataklar sırasında alkol veya madde kötüye kullanımı saptanabilir. Ayrıca bir araştırmada da hastaneye yatan ve çoğu bipolar I bozukluğu olan hastaların %24'üne anksiyete bozuklukları eşlik etmiştir [118].

Yapılan bir diğer araştırmada ise ilk kez hastaneye yattıktan sonra 4 yıl hastanede kalmış manik veya karma epizotlu 166 erişkin hastanın %98'inde 2 yılda ve bunların da yarısında 54. haftada sendromal iyileşme sağlanmıştır. Hastaların %72'sinde ise semptomatik (değerlendirme ölçekleriyle ölçülen semptomlar belirgin derecede hafiflemiştir) düzelme görülmüştür (mesleki ve genel durumda hastalık

öncesi iyileşme haline geri dönüş). Sendromal düzelme sağlandıktan sonra hastaların %40'ı yeni bir mani veya depresyon atağı geçirmiş veya herhangi bir düzelme olmaksızın hastalığın fazları değişip durmuştur [119].

Judd ve arkadaşlarının yapmış olduğu bir araştırmadaysa bipolar I bozukluğun kronik bir seyir izleyebildiği, ortalama 13 yıllık izlemde hastaların %47'sinin semptomatik olduğu gösterilmiştir. Manik-hipomanik semptomlar (%9) veya döngüsel ya da karma semptomlara (%6) göre depresif semptomlar (gözlenen haftaların %32'sinde) daha sık görülmüştür [120]. İleriye dönük ortalama 13 yıllık izlem süresince takip süresinin %54'üne varan zaman diliminde bipolar II bozukluğu olanlarda benzer kalıplar gözlenmiştir. Bu çalışmada da hipomanik (%1), döngüsel veya karma (%2) semptomlardan çok (gözlem yapılan haftaların %50'sinde) depresif semptomlar görülmüştür [121].

Çocuklarda bipolar bozukluğun doğal seyri biraz farklıdır. Erişkinlerde 3., 6. aylarda, 1. ve 2. yıllarda sırasıyla %77, 86, 92 ve 98 gibi yüksek oranlarda sendromal düzelme görülürken [119], ergenlik öncesi veya erken erişkin dönemde mani geçirenlerde bu oranlar çok daha düşük düzeylerde bulunmuştur [122]. Toplam 86 hastayı inceleyen bir çalışmada (başlangıçta yaş ortalaması 7 yıl), sendromal düzelme oranları 6. aylar ve 2. yıllarda sırasıyla ancak %14 ve %65 düzeylerinde bulunmuştur [122].

Bipolar bozukluk ile ilgili onlarca yıl süren çalışma ve gözlemler sonucunda elde edilen tüm bu bilgiler, genetik alanındaki ilerlemelerle birlikte ortaya çıkan yeni araştırma yaklaşımlarının temelini oluşturmuştur. Bu süreçte hastalığın tanı kriterleri birçok kez yenilenmiş ve toplumların etnik kökenlerine ilişkin özellikleri sürekli bir değişim göstermiştir. Biyoinformatik yöntemlerinin yaygın olarak kullanıldığı genomik çalışmalar öncesinde genetik epidemiyoloji yöntemlerini temel alan yaklaşımlar yoluyla kalıtsal temelleri olabilecek bazı fenotiplerin varlığı (hastalıklar veya özellikler) tespit edilmeye çalışılmaktaydı. Özetleyecek olursak, ikiz çalışmaları, aile ağacı çalışmaları ve popülasyon temelli araştırmalar ile herhangi bir hastalığın kalıtsal kökenleri olup olmadığı tespit edilebiliyordu. En çok aktarılan fenotipler inceleniyor ve genetik ve çevresel faktörler arasındaki etkileşimlerle ilgili hipotezler üretilabiliyordu. Bugün majör psikiyatrik bozukluklar için hala kullanılmakta olan tanı kriterleri de temel olarak benzer ikiz ve aile araştırmalarına dayanmaktadır [48].

Genetik araştırmaların yoğunlaşmaya başladığı ilk yıllarda bipolar bozukluk ile ilgili yapılan araştırmalarda da diğer tüm hastalıklarda olduğu gibi çeşitli kısıtlılıklar göze çarpmaktadır [123]. Bu kısıtlılıkların başında yetersiz örneklem büyüklükleri, standart araştırma yöntemlerinin geliştirilememiş olması ve genetik haritalama ile ilişkin bilgilerin henüz tam şekillenmemiş olması gelmekteydi [109].

Bununla birlikte, teknoloji alanında kaydedilen gelişmeler, maliyetlerin hızla düşmesi, çok merkezli çalışmaların sayısındaki artışlar ve uygulanmaya başlanan yeni analiz yöntemlerinin ortaya çıkışı ile bu sorunların üstesinden gelinebileceği ortaya çıktı. Bu gelişmelerle birlikte özellikle 2005 yılından itibaren kompleks hastalıkların genetik kökenlerine ilişkin daha standart araştırmalar yapılmaya başlandı. O tarihlerde özellikle “ortak hastalık – ortak genetik varyasyon” yaklaşımından hareketle birçok araştırmacı kompleks hastalıklara veya fenotiplere sahip benzer bireylerde genetik olarak belirgin benzerlikler gösteren varyasyonların görülmesi gerektiğini varsaymaktaydılar [32-34, 124]. Böylece en sonunda büyük ölçekli araştırmalar yapılarak kompleks fenotiplerin genetik nedenini açıklayan genlerin bulunacağı ya da en azından sistematik moleküler araştırmalar için belirgin başlangıç noktalarının tanımlanabileceği öngörülmüyordu. Bipolar bozukluk genetiğine odaklanan araştırmaların da ortak amacı buydu [123].

Bununla birlikte, o tarihten sonra sayısı giderek artan araştırmalar sonucunda bu beklentilerin tam anlamı ile karşılanmadığı ortaya çıktı. Çok geniş ölçekli araştırmalar veya birçok araştırmaya ait bulguları birleştirmeyi amaçlayan meta analizi uygulamaları sonucunda, bulguların öngörüldüğü gibi birden çok çalışmada kuvvetli bir biçimde replike edilemediği görüldü. Her ne kadar bazı genom bölgelerinin, genlerin veya SNPler’in birden çok araştırmada kompleks fenotiplerle ilişkilendirilebilmesi mümkün olsa da, umulduğu gibi net ve tekrarlanabilir bulgular ortaya çıkmamıştır [33, 48, 109].

Böylece birçok farklı çalışma sonucunda farklı aday gen ve SNPler rapor edilmeye başlandı. 21. yüzyılın başlarına kadar genel hâkim görüş olan “ortak hastalık – ortak genetik varyasyon” fikri de revize edilmek durumunda kaldı. Artık kompleks hastalıkların genetiğine ilişkin daha baskın yaklaşım “ortak hastalık – çok sayıda, nadir görülen varyasyon” hipotezi ile ifade edilmektedir [33, 35, 36]. Yani kompleks fenotiplerin bireysel etkileri küçük fakat toplamsal olarak istatistiksel anlamlılık gösterebilecek çok sayıda genetik varyasyona bağlı olarak ortaya çıktığı fikri ağırlık kazandı [125]. Özellikle de psikiyatrik bozukluklara ilişkin elde ettiğimiz bilgiler göstermektedir ki her biri görece az risk etkisi olan çok sayıda genetik varyasyonun çevresel faktörlerle etkileşimi, kalıtsallığın en temel belirleyicisi olmaktadır [48].

Bununla birlikte, bipolar bozukluğun genetik heterojenliği düşünülecek olursa yürütülen ilişki araştırmalarının mevcut bulguları ve ileride belki de daha büyük örneklerle gerçekleştirilecek araştırmaların sonuçları hala büyük önem taşımaktadır. Tüm bu bulguları toparlamak ve özellikle de hastaların ailelerinden daha titiz bir şekilde fenotip verisi elde ederek bu veriler ile genetik bilgileri ilişkilendirmeye çalışmak bipolar bozukluk araştırmalarının temel amaçlarından biri olmaya devam etmektedir [109, 126]. Ayrıca bugün yürütülmekte olan moleküler temelli genetik araştırmalarda da karşılaştırma gruplarını uygun şekilde

belirleyebilmemiz için hala kaliteli epidemiyolojik verilere ihtiyaç duyulmaktadır [48].

Her ne kadar yukarıda bahsedilen zorluklar nedeniyle, bipolar bozukluğa neden olan kalıtsal faktörler yapılan her araştırmada benzer istatistiksel anlamlılıkla replike edilememiş de olsa, konu ile ilgili birçok GWAS araştırmasında yatkinlığa neden olan aday gen ve SNPlar'ın varlığı deneysel olarak kanıtlanmıştır [6, 9, 127-130].

Şu ana kadar bipolar bozukluk hastaları ile yürütülen GWAS araştırmalarında bazı aday genlerin varlığından söz edilebilmektedir [5-10, 50]. Bu genlerden başlıcaları DGKH, CACNA1C ve ANK3 genleri olup çeşitli araştırmalarda bu genlerin genom çapında anlamlı oldukları tekrarlı biçimde gösterilebilmiştir [5-12]. DGKH geni kromozom 13q14'ün bipolar ilişki bölgesinde yer alır ve diaçil gliserol kinaz eta proteininin üretimini kodlamakla görevlidir. Bu protein lityuma duyarlı fosfatidil inositol yolağında anahtar görev üstlenmektedir. CACNA1C geni voltaja bağımlı kalsiyum kanalının bir alfa-1 alt ünitesini kodlamakla sorumludur. ANK3 geni ise ankrin-G'yi kodlar. Bu büyük protein, nöral-spesifik izoformları aksonal başlangıç segmentinde ve Ranvier nodlarında yer alır ve iyon kanalları ile hücre adezyon moleküllerinin devamlılığını sağlar [109]

Bipolar bozukluk ile ilgili şu ana kadar yürütülmüş olan GWAS araştırmalarının ortaya çıkardığı bazı önemli bulgular şöyle sıralanabilir [109]:

- Bipolar bozukluk, poligenik bir hastalıktır. Yani her bir lokusun hastalık riskine katkısı görece düşüktür. Hasta bireylerde risk allelleri kontrollere kıyasla anlamlı miktarda fazladır. Dolayısıyla hastalık riski, taşınmakta olan risk allellerinin miktarına bağlı olarak artış gösterebilmektedir.
- GWAS ile bulunacak en açıklayıcı bulguların önceden üzerinde çokça çalışılmış olan genlerden birisi olması şart değildir. Çünkü bu şüpheli genler genellikle ya sinirsel iletim ile ilgili hipotetik nedenselliğe ya da ilişki çalışmalarının bulgularına göre belirlenmişlerdir ve yeni araştırma bulguları bakımından kesin bağlayıcılıkları yoktur.
- Genellikle GWAS araştırmalarında uygulanan bir yöntem olan “belirli sayıdaki en anlamlı bulguyu değerlendirme” stratejisi birçok kompleks hastalıkta olduğu gibi bipolar bozuklukta da yeterli olmayabilir. Yani bir GWAS araştırmasında kabul edilen istatistiksel anlamlılık sınırı risk genlerinin ve polimorfizmlerin belirlenmesinde bazen negatif etki gösterebilmektedir. Yapılan meta analizi araştırmaları göstermektedir ki, tek başlarına değerlendirilen GWAS

arařtırmalarında yüksek düzeyde anlamlı olmayan lokuslar, birden çok alıřma ortak deęerlendirildięinde nemli bir bulguya dnüşebilmektedir. Dolayısıyla benzer alıřmaların tümü gözetilerek düşünülecek istatistiksel anlamlılık düzeyi belki de kompleks genetik zelliklerin anlaşılması için daha deęerli olabilir [131].

- Bipolar bozukluk gibi kompleks hastalıklarda allelik heterojenite nemli bir faktr olabilir. Allelik heterojenite; herhangi bir fenotipe bir gen üzerindeki farklı allellerin neden olabileceęidir. Bu kavram, kistik fibrozis [132] ve BRCA1/2 ile iliřkili meme kanseri [133] gibi monogenik hastalıklarda net řekilde gözlenebilmektedir. zellikle ANK3 geni ve bipolar bozukluk iliřkisi düşünöldüğünde, eřitli sayıda allelin ve haplotipin baęımsız risk faktrleri olabildięi gözlenmiřtir [5, 12].
- Dięer tüm kompleks fenotipte de göröldüğü gibi, bipolar bozukluk ile ilgili GWASlar sonucunda belirlenebilen varyasyonların da genetik deęiřkenlik aısından sadece küçük yüzdelerle etki gösterdikleri netlik kazanmıřtır. Bu kavrama “eksik genetik aktarılabirlik” adı verilmektedir [134]. Ayrıca epigenetik kavramının da bu noktada önem kazandıęı görölmektedir [101].

Postmortem beyin örneklerinden alınan dokularla gerekleřtirilen mikrodizi analizleri temelli arařtırmalar da bulunmaktadır. Her ne kadar bazı SNP ve genler birden çok alıřmada replike edilebilse de, bipolar davranıř bozukluęuna iliřkin genetik risk faktrleri hala tam olarak ortaya konulamadıęı söylenebilir [106, 135, 136]. Bunun nedeni bu nöropsikiyatrik hastalıęın poligenik yapısı, fenotipik kompleksitesi ve heterojenlięi olabilir [106]. Fakat GWAS ile gen ekspresyonu verilerinin entegre edilebileceęi ve yeni nesil sekanslama teknolojilerinin de kullanılacağı alıřmalarla bu konudaki bilgi birikiminin artacağı öngörülmektedir [106].

GEREÇLER VE YÖNTEMLER

3.1. Veri Seti

Araştırmada kullanılan veri seti, doktora tez çalışması dâhilinde işbirliği yürütülen akademik kurumların desteği ve ortak çalışma girişimleri sonucunda mevcut araştırmayı desteklemek amacıyla sağlanmıştır. Veri seti ABD’de oluşturulmuş ve çeşitli bilimsel araştırma projeleri için belirli akademik izinler dâhilinde araştırmacıların kullanımına sunulmaktadır. Şu an itibarı ile bahsedilen akademik izin süreçleri de yerine getirilerek veri üzerinde yeni yöntem uygulamalarına destek verilmektedir.

ABD’de sağlık alanındaki bilimsel araştırmaları destekleyen ve organize eden en büyük kurum olan National Institute of Health (NIH) bünyesinde kurulmuş olan The National Center for Biotechnology (NCBI) tarafından çevrimiçi sunulan büyük veri tabanlarından birisi olan dbGAP (The database of Genotypes and Phenotypes) bünyesinde çalışmamıza sunulan veriye erişim mümkündür [78].

Veri setinin çalışmada kullanılan rafine edilmiş haline ilişkin hasta ve kontrol seçim yöntemleri, klinik değerlendirme, genotipleme ve kalite kontrolü gibi basamakları aşağıda özetlenmiştir. Bu yöntemler Smith ve arkadaşlarının çalışmasında detaylandırılmıştır [5].

3.1.1. Bipolar Bozukluk Hastaları

Bu hastalar ABD’de uzun süredir genetik çalışmalara destek veren National Human Genome Research Institute bünyesinde kurulmuş olan The Genetic Association Information Network (GAIN) çalışma grubunun kollarından birisi olan Bipolar Konsorsiyumu’nun 1991 senesinden beri Amerika’da yürüttüğü çalışma sonucunda belirlenmiş hastalardır [78]. Hasta bireylerin toplanması toplam 5 basamakta gerçekleşmiştir. İlk dört basamakta bağlantı (linkage) çalışmaları için toplanan aileler vardır. Beşinci basamakta ise geniş ölçekli ilişki çalışmaları için toplanmış olan ve önceki toplanan bireyler ile akrabalıkları olmayan hastalar vardır.

Bu basamaklardan 1-4 arasındakiler bipolar aileler veri setini oluşturmaktadır ve toplam 646 soyağacından (pedigree) 2.936 adet birey toplanmıştır. Bu bireylerin ortak özelliği, kendilerinin Bipolar I (BPI) tanısı almış olmaları ve ek olarak da soyağaçlarında en az bir adet birinci dereceden BPI ya da bipolar tip şizoaffektif bozukluk (SABP) tanısı almış yakınlarının bulunmasıdır. Bizim çalışmamız için bu

olgular arasından belirli özelliklere sahip bir alt grup belirlenmiştir. Bu soyağaçlarından oluşan grup içerisinde Diagnostic DSM-IV kriterlerine uygun ve birbiri ile ilişkili olmayan BPI tanısı almış bireyler seçilmiştir. Bu seçim sonucunda 1. ve 2. Basamağa ait toplam 175 adet akraba olmayan bireye erişilmiştir. 3. ve 4. Basamaktan ise 396 birey seçilmiştir. Araştırmanın son basamağından ise toplam 430 vaka çalışmamıza dâhil edilmiştir. Bu sayede toplamda 1.001 adet hastaya ulaşılmıştır. Bu hastaların ortak özelliği birbirleri ile akrabalık ilişkilerinin olmaması ve Avrupa kökenli Amerikalılar olmalarıdır. Kişilerin Avrupa kökenli Amerikalı olup olmadıklarına bakılırken kendi ifadelerine ve hem anne hem de baba tarafındaki dört büyükanne ve büyükbabalarının hepsinin birden Avrupa kökenli Amerikalı olmalarına bakılmıştır. Toplam 1.001 vakanın 951'i BPI tanısı almışlarken geriye kalan 50'si ise SABP tanısı almışlardır.

3.1.2. Kontroller

Kontrol bireyleri Dr. Pablo Gejman öncülüğünde ve Knowledge Networks Inc. kurumunun The National Institute of Mental Health (NIMH) tarafından desteklenen çalışması sonucunda toplanan bireyler arasından seçilmişlerdir. Bu çalışmada Amerika Birleşik Devletleri genelinde toplam 4.586 adet gönüllü birey bir tıbbi anket amacıyla kan örnekleri vermeyi kabul etmişlerdir. Bu bireyler arasından psikiyatrik anket bölümünü tamamıyla dolduran ve psikoz veya bipolar bozukluk öyküsü yaşamadıklarını rapor eden bireyler çalışmamıza dâhil edilmişlerdir. Toplamda 1.033 adet Avrupa kökenli Amerikalı birey araştırmamızda kontrol grubunu oluşturmuştur.

3.1.3. Klinik Değerlendirme

Tüm hasta bireyler ile Diagnostic Interview for Genetic Studies (DIGS Version 4.0) temelli görüşmeler düzenlenmiştir [137]. DIGS, posttravmatik stres bozukluğu ve yetişkinlerde dikkat eksikliği bozukluğu gibi konularda ek veriler toplamak ve bipolar bozukluğa ilişkin yeni fenotipik özellikler ekleyebilmek amacıyla revize edilmiştir. Ayrıca kullanacak olan klinisyenin tanı koyma sürecinde çok sayıda fenotipik göstergenin bulunması durumunda yargısını kolaylaştırabilmek amacıyla da değişiklikler yapılmıştır. Bu ek bilgilerin arasında başlangıç yaşı, depresyon epizodu sayısı, hipomani ve mani, madde bağımlılığı ve psikozla ilgili geçici duygudurum bozuklukları ve aile öyküsü özeti gibi bilgiler bulunmaktadır. Tüm göstergeler uzman bir klinisyen tarafından (genellikle psikiyatrist) değerlendirilmiştir. Bu süreçte eldeki tüm bilgilere, hastaların sağlık kayıtlarına, görüşme yapan kişinin gözlemlerine ve DIGS cevaplarına dikkat edilmiştir.

3.1.4. Genotipleme ve Kalite Kontrolü

Alınan örneklerle gerçekleştirilen genotipleme The Broad Institute Center for Genotyping and Analysis bünyesinde gerçekleştirilmiştir. Genotipleme Affymetrix Genome-Wide Human SNP Array 6.0 kullanılarak yapılmıştır. Kalite kontrol

sonucunda %86'nın altında genotiplendiği görülen örnekler tekrar genotipleme sürecine sokulmuştur. Belli başlı bazı kalite metriklerine uygunluk göstermeyen örnekler çalışma dışında tutulmuştur. Bu metriklerden bazıları;

- düşük çağrı oranı (call rate = örnek başına çağrılan SNP sayısının veri setindeki toplam SNP sayısı içindeki yüzdesi) (%98,5'in altındaysa)
- aşırı yüksek ya da düşük heterozigotluk (0,344 ve 0,363 arası)
- rapor edilen cinsiyet ile genetik analizle belirlenen cinsiyet arası uyumsuzluk.

Ayrıca araştırmaya dâhil edilebilmeleri için SNPlar'ın de bazı kalite kontrol ölçütlerini sağlamaları istendi. Bu ölçütler;

- minör allel frekansının 0,01'den fazla olması
- çağrı oranının %95'in üstünde olması
- kontrol örneklerinde SNP'in Hardy Weinberg eşitliğini ($p < 1 \times 10^{-6}$) düzeyinin altında kalmayacak şekilde sağlaması
- en fazla 3 Mendelyen hatanın görülmesi
- tekrarlanan örnek ölçümlerinde birden çok uyumsuzluk görülmemesi.

Bütün bu genotipleme ve kalite kontrol aşamalarının ardından toplamda 724.067 adet SNP araştırmaya katılmıştır.

3.2. İstatistiksel Yöntemler

3.2.1. Genom Çapında İlişki Araştırması

Araştırmanın bu bölümünde bipolar bozukluk hastası olan 1.001 birey ile kontrol grubundaki 1.033 birey arasında SNP düzeyindeki olası farklılıklar incelenmiştir. Tüm kalite kontrol süreçleri sonucunda analize giren toplam 724.067 adet SNP incelenmiştir.

Analizler PLINK (Version 1.05) kullanılarak yürütülmüştür. Her SNP için aşağıdaki özellikler ve hesaplamalar elde edilmiştir:

- Kromozom
- SNP numarası (rs number)

- Ait olduđu genin ismi (biliniyorsa)
- İlk allelin harf kodu (tüm örneklem frekanslarına göre belirlenen minör allel)
- Bu allelin hastalarda görülme sıklığı
- Bu allelin kontrollerde görülme sıklığı
- Diğer allelin harf kodu
- Teste ait ki-kare istatistiđi deđeri (serbestlik derecesi = 1)
- Testle hesaplanan asimptotik anlamlılık deđeri (düzeltilmemiş p deđeri)
- Genomik kontrole göre düzeltilmiş anlamlılık deđeri. Bu deđer ortanca ki-kare istatistiđi deđeri kullanılarak hesaplanmaktadır.
- Testle hesaplanan odds oranı

Tüm hesaplamalar sonucunda düzeltilmiş p deđerleri dikkate alınarak hasta ve kontrol bireyleri arasında allel frekansları bakımından istatistiksel olarak anlamlı olduđu görülen SNPLer belirlenmiştir. İstatistiksel önemlilik için alfa deđeri 0,0001 olarak belirlenmiştir.

3.2.2. GWAS Sonuçlarının SNP'den Gen Düzeyine İlerletilmesi

GWAS analizi sonucunda elde edilen tüm SNPLer'e ait anlamlılık düzeyleri, çevrimiçin bir uygulama kullanılarak gen düzeyine ilerletilmiştir. Liu ve arkadaşlarının geliştirerek çevrimiçi hizmete soktukları VEGAS isimli uygulama aracı, GWAS ile elde edilmiş olan SNP p deđerlerini alarak bu SNPLer'i genom üzerindeki konumlarına göre referans gen dizisini temel alarak ait oldukları genlere haritalamaktadır. Böylece veri setindeki SNPLer'den haritalama suretiyle gen bilgisine ulaşılmaktadır. Her gene ait tüm SNPLer deđerlendirilerek SNPLer'e ait hesaplanmış olan p deđerleri matematiksel modelleme ve simülasyonlar yoluyla gen düzeyinde tekrar hesaplanmaktadır [138].

Araştırmamızın GWAS basamađında kullanılan 724,067 adet SNP'e ait p deđerleri hesaplandıktan sonra VEGAS çevrimiçi aracı kullanılarak gen düzeyinde anlamlılıklar incelenmiştir. Gen düzeyindeki p deđerleri SNPLer'den hesaplanan p deđerlerinin güvenli bir kestirimi olduđu için ve her gende birden çok SNP p deđerleri

dikkate alındığı için gen düzeyindeki p değerlerinin SNP seviyesine göre daha yüksek olması beklenmektedir [138]. Bu nedenle çalışmamızda gen düzeyinde dikkate alınacak alfa değerinin 0,005 olması öngörülmüştür. Bu aşamada her gene ait aşağıdaki özellikler sunulmuştur:

- Genin üzerinde yer aldığı kromozom
- Genin ismi
- Gen üzerinde veri setinde kullanılmış olan SNP sayısı
- Genin başlangıç pozisyonu
- Genin sonlanış pozisyonu
- Gene ait düzeltilmiş p değeri

Hem SNP düzeyinde hem de gen düzeyinde elde edilen bulguların grafiksel gösterimleri için Q-Q grafikleri ve Manhattan grafikleri çizilmiştir. Bu grafiksel gösterimler için R yazılımının (Version: x64 2.14.0) “graphics” paketi (Version 2.16.0) kullanılmıştır [139, 140]. Stephen Turner’ın paket için özelleştirdiği Manhattan ve Q-Q grafikleri çizim scriptlerinden faydalanılmıştır [141].

3.2.3. Çok Değişkenli Uzaklık Matrisi Regresyonu

Y matrisinin N x P boyutunda ve gen ekspresyonu değerlerinden oluşan bir matris olduğunu düşünelim. N kişi sayısını P ise gen sayısını ifade etsin. X matrisi ise N x M boyutunda olsun ve M de genlerin ekspresyonu ile ilişkili olduğu düşünülen tahmin edici değişkenlerin sayısı olsun. M’in ilk sütunu tüm elemanları 1 olan bir vektör olsun. Bu tahmin edici değişkenler, yaş, klinik tanı, grup, hücre tipi, genotip bilgisi gibi birçok farklı değişken olabilir. Standart çoklu regresyon modeli bu durumda şöyle yazılacaktır:

$$Y = X\beta + \varepsilon$$

Bu denklemde β regresyon katsayılarından oluşan M x P boyutunda bir matristir ve ε da hata terimidir. β için üretilen en düşük kareler çözümü:

$$\hat{\beta} = (X'X)^{-1} X'Y$$

olacaktır. Modelin artık hata matrisi ise:

$$R = Y - \hat{Y} = Y - X\hat{\beta} = (I - H) Y$$

olur. Burada;

$$H = (X'X)^{-1} X'$$

olduğunu ve geleneksel “şapka” matrisi olduğunu hatırlatalım. Modelde verisi olan kişi sayısı, modele sokulan değişken sayısından çok daha az olduğu durumlarda ($N < P$) bu modeli çözmek karmaşıklaşır. Genomik araştırmalarda toplanan verilerde durum genellikle böyledir. Bu karmaşıklık için önerilebilecek alternatif yaklaşım ise M sayıda tahmin edici değişkenin, çalışmaya katılan olguların P sayıdaki gen ekspresyonu değerlerine göre birbirleri ile olan benzerlik veya uzaklıkları ile olabilecek olası ilişkilerine bakmaktır.

Bu durumda D matrisinin $N \times N$ boyutunda bir matris olduğunu ve d_{ij} teriminin de i ve j sayılı bireylerin P gen ekspresyonu değerleri açısından birbirleri ile olan uzaklıklarını (ya da farklılıkların) gösterdiğini düşünelim. Örneğin d_{ij} terimi iki birey arasındaki Öklit uzaklığının kendisi veya korelasyon katsayısının bir fonksiyonu olarak hesaplanabilir.

$$A = (a_{ij}) = \left(-\frac{1}{2} d_{ij}^2\right)$$

olsun. Bu durumda A 'yı kullanarak Gower matrisini şu şekilde hesaplayabiliriz:

$$G = \left(I - \frac{1}{n} 11'\right) A \left(I - \frac{1}{n} 11'\right)$$

Burada 1 N boyutlu ve her elemanı 1 olan bir sütun vektörüdür. I ise $N \times N$ birim matristir. M sayıdaki tahmin değişkeni ile N bireyin P değişkene göre farklılıklarının arasındaki ilişkiyi incelerken kullanılacak uygun F istatistiği aşağıdaki denklemle hesaplanabilir:

$$F = \frac{\text{tr}(HGH)/(M-1)}{\text{tr}[(I-H)G(I-H)]/(N-M)}$$

Yukarıda da vurgulandığı gibi H şapka matrisi, G Gower matrisi ve I da birim matristir. M bir skalerdir ve tahmin edici değişken sayısını göstermektedir. N de birey sayısıdır. Eğer $P = 1$ ise ve uzaklık matrisi standart Öklit uzaklığına göre hesaplanmışsa, bu durumda yukarıdaki denklemde hesaplanan F , standart F istatistiğidir ve varyans analizi bağlamındaki F istatistiği ile aynı özellikleri gösterir. Farklı uzaklık hesaplama yöntemleri kullanılarak oluşturulan uzaklık matrisleri için yapılacak çözümlerde F istatistiğinin özellikleri de değişim gösterecektir.

F istatistiğinin dağılımının kullanılan uzaklık matrisi yöntemine göre değişim göstermesi ve dolayısıyla karmaşık olması sebebiyle, istatistiksel anlamlılığı belirleyebilmek için permütasyon testleri uygulanabilir [142, 143]. Permütasyon testleri ham veriyi permüte ederek de gerçekleştirilebilir, G matrisinin satır ve sütunlarını eşzamanlı permüte ederek de gerçekleştirilebilir [97]. Ek olarak, eğer permütasyon testleri uygulanacaksa F istatistiğinin hesaplandığı eşitliğin payındaki (M – 1) ve paydasındaki (N – M) serbestlik derecesi terimlerini kullanmaya gerek yoktur. Ayrıca tıpkı standart çoklu regresyon analizlerinde olduğu gibi basamaklı (step-wise) veya değişken seçme prosedürleri de bu yöntemde uygulanabilir [144]. P değeri hesabının dışında hesaplanabilecek bir diğer rakamsal gösterge de belirli tahmin edici değişkenlerin uzaklık matrisindeki varyasyonun ne kadarını açıklayabildiğidir. Bu ölçüte kısaca varyasyon açıklama oranı (proportion of variation explained – PVE) diyecek olursak;

$$PVE = \text{tr}(HGH) / \text{tr}(G)$$

şeklinde bulunabilir.

Zapala ve Schork yöntemi teorik olarak geliştirip ortaya koydukları çalışmalarında hesaplamalara ilişkin yukarıda sunulan ayrıntıları detaylı şekilde açıklamışlardır. Ayrıca yaptıkları gerçek verili ve simüle uygulamalarla yöntemin istatistiksel gücünün de yüksek olduğunu göstermişlerdir [14].

Çalışmamızda MDMR yöntemi kullanılarak bipolar bozukluk hastalarının fenotipik özellikleri temel alınmış ve bireylerin bu özellikler bakımından birbirleri ile uzaklıkları hesaplanarak bir uzaklık matrisi oluşturulmuştur. Uzman doktorların görüşüne dayanılarak, toplanan fenotip verisi içinden çalışmaya dâhil edilmesi konusunda karar alınan toplam 44 adet farklı fenotipik özellik değişken belirlenmiştir. Bu 44 değişkenden tüm hastaların %5'inden fazlası için eksik veri içeren değişkenler, uzaklık matrisi hesabına dâhil edilmemişlerdir. Bu kurala uyan ve matris oluşumuna katılan fenotipik değişken sayısı 29'dur. Bu 29 adet fenotipik özellik aşağıda sıralanmıştır.

- **Migren** (DIGS – Tıbbi Öykü – Soru 1: Bir doktora tarafından migren sebepli baş ağrısı tanısı aldınız mı?)
- **Ağır manik konuşkanlık** (DIGS – Mani / Hipomani – Soru G7: En ağır manik epizodunuzda normalden daha fazla mı konuşuyordunuz ya da konuşmaya devam etmeniz yönünde bir baskı hissettiniz mi?)

- **Ađır manik dūőüne uęuőması** (DIGS – Mani / Hipomani – Soru G8: En ađır manik epizodunuzda dūőünceleriniz uęuőuyor muydu ya da insanların sizi takip etmekte zorlandığı kadar hızlı konuőuyor muydunuz?)
- **Ađır manik grandiozite** (DIGS – Mani / Hipomani – Soru G9: En ađır manik epizodunuzda kendinizin ęok önemli bir insan olduđunu ya da özel güçler, yetenekler ve planlarınız olduđunu dūőündünüz mü?)
- **Ađır manik az uyuma** (DIGS – Mani / Hipomani – Soru G10: En ađır manik epizodunuzda uykuya normalden daha az ihtiyaę hissettiniz mi?)
- **Ađır manik dikkat dađılması** (DIGS – Mani / Hipomani – Soru G11: En ađır manik epizodunuzda dikkatiniz bir konudan baőka bir konuya atlayıp durduđu için konsantrasyonunuzda normale göre azalma oldu mu?)
- **Ađır manik sıkıntıya neden olma** (DIGS – Mani / Hipomani – Soru G12: En ađır manik epizodunuzda sizin için sıkıntıya neden olabilecek bir Őey yaptığımız oldu mu; örneđin gereksiz satın almalar, ticari yatırımlar yapma, cinsellikle ilgili dūőüncesiz hareketler yapma, dikkatsiz araba sürme gibi?)
- **Ađır Mani Semptomları Sayısı** (DIGS – Mani / Hipomani – Soru G1-72)
- **Duysal halüsinasyonlar** (DIGS - Psikoz – Soru K1a: Hię sesler duyduđunuz oldu mu? Örneđin bazı insanlar, aslında geręekte olmayan kiőilerin kendilerine fısıldadıklarını veya kendileriyle konuőtuklarını tecrübe ettiklerini söylerler)
- **Görsel halüsinasyonlar** (DIGS - Psikoz – Soru K1b: Hię baőka insanlarca görülmeyen Őeyleri gördüğünüz oldu mu?)
- **Delüzyonlar** (DIGS - Psikoz – Soru K1c: Hię baőkalarının sizinle herhangi bir Őey paylaőmadığı yönünde inanę ve fikirlere kapıldığımız fakat daha sonra bunun dođru olmadığını fark ettiđiniz oldu mu? Örneđin insanların size karőı olduklarına, size zarar vermeye ęalıőtıklarına, hakkınızda konuőtuklarına ya da size özel mesajlar

verildiğini düşündüklerine (televizyon veya radyo gibi yollarla) inandınız mı?)

- **İntihar girişimi** (DIGS – İntihar Davranışı – Soru O1: Hiç kendinizi öldürmeyi denediniz mi?)
- **Obsesiflik** (DIGS – Anksiyete Bozuklukları – Soru P1: Hiç aslında anlamı olmayan ve düşünmemeye çalıştığınız halde sürekli aklınızı kurcalayan düşüncelerle boğuştuğunuz oldu mu?)
- **Kompulsiflik** (DIGS – Anksiyete Bozuklukları – Soru P2: Hiç daha az anksiyete hissedebilmek amacıyla aynı hareketi direnç göstermeden defalarca tekrarladığınız oldu mu? Örneğin el yıkama, bir şeyleri sayma ya da sürekli kontrol etme gibi)
- **Panik Bozukluk** (DIGS – Anksiyete Bozuklukları – Soru P11: Hiç aslında herhangi bir tehdit olmayan durumlarda aniden korkarak panik atak veya anksiyete atağı geçirdiğin oldu mu?)
- **Agorafobi** (DIGS – Anksiyete Bozuklukları – Soru P28a: Hiç dışarıya yalnız çıkmaktan, kalabalık yerlerde veya mağazalarda yalnız bulunmaktan ya da kaçamayacağını veya yardım bulamayacağını hissettiğin yerlerde tek başına bulunmaktan ötürü aşırı korkmuş hissettin mi?)
- **Sosyal fobi** (DIGS – Anksiyete Bozuklukları – Soru P28b: Hiç insanlar önünde konuşmak, yemek yemek veya yazmak gibi eylemleri gerçekleştirme fikrinden aşırı korkmuş hissettin mi?)
- **Basit fobi** (DIGS – Anksiyete Bozuklukları – Soru P28c: Hiç çeşitli hayvanlardan, yüksekten veya kapalı kalmaktan aşırı korkmuş hissettin mi?)
- **Anoreksiya Bulimia** (DIGS – Yeme Bozuklukları – Soru Q14-18)
- **Alkolizm** (DIGS – Alkol İstismarı ve Bağımlılığı – Soru I1-23)
- **Madde istismarı** (DIGS – Tütün, Uyuşturucu Madde İstismarı ve Bağımlılığı – Soru J1-53)

- **Ađır Depresyon Semptomları Sayısı** (DIGS – Majör Depresyon – Soru F1-78)
- **Hastalık Bařlangıç Yařı** (DIGS - Psikiyatrik Rahatsızlıđa Genel Bakıř – Soru E2b: Duygusal problemleriniz iin bir profesyonele danıřtıđınızda ka yařındaydınız?)
- **Tedavi Uygulandıđı Sırada alıřma Durumu** (DIGS – Psikiyatrik Rahatsızlıđa Genel Bakıř – Soru E2c: Rahatsızlıkla ilgili destek aldıđınız sırada alıřıyor muydunuz?)
- **Psikiyatrik Hastaneye Yatma Gemiři** (DIGS – Psikiyatrik Rahatsızlıđa Genel Bakıř – Soru E4: Ruh haliniz, duygularınız ya da davranıř řekliniz yüzünden bir hastaneye yatırıldıđınız oldu mu?)
- **Görüşme Anındaki Yař** (DIGS – Demografik Bilgiler – Soru 2: řu an ka yařındasınız?)
- **Medeni durum** (DIGS – Demografik Bilgiler – Soru 7: řu anki medeni durumunuz nedir?)
- **Meslek** (DIGS – Demografik Bilgiler – Soru 10a: řu anki iřiniz nedir?)
- **Okul Yılı** (DIGS – Demografik Bilgiler – Soru 11: Toplam ka yıllık okul eđitimi tamamladınız?)

Bu deđiřkenlerin dört tanesi (görüşme anındaki yař, medeni durum, meslek ve okul yılı) hastalıđa yönelik birer fenotipik özellik olmadıklarından, fakat arařtırmanın hedeflerinden birisi olan fenotip x genotip etkileřimini etkileyebilecekleri düşünöldüđünden kovaryant olarak atanmıř ve regresyon denkleminin bađımsız deđiřkenleri olarak deđerlendirilmiřlerdir.

Kovaryant olarak atanan dört deđiřken ıkarılınca geriye kalan toplam 25 adet deđiřken, hastalara ait uzaklık matrisini oluřturmak iin kullanılmıřtır.

Uzaklık matrisine sokulan 25 deđiřken iin, uzaklıkların hesaplanma ařamasından önce son bir basamak filtre daha uygulanmıřtır. Bu basamakta, uzaklıkların hesaplanmasında eksik verisi olan hastaların hesaplamalara dâhil edilmemesi iin tüm deđiřkenler incelenmiř ve verisi bulunmayan hastalar alıřma

dışı bırakılmıştır. Bu basamakta elenen hastalar çıkarıldıktan sonra uzaklık matrisi 25 değişken üzerinden toplam 873 hasta ile oluşturulmuştur.

Uzaklık matrisinin oluşturulmasında R yazılımı (Version: x64 2.11.1) kullanılmıştır. Uzaklık matrisi hesaplamak için ise FD paketi (Version 1.01-11) içerisindeki “gowdis” fonksiyonu kullanılmıştır [145, 146]. Bu fonksiyon farklı tiplerde değişkenleri içeren verilerde (sürekli, sıralı, ikili) bireyler arasındaki ikili farklılıkları ölçmek için Gower’ın benzerlik hesaplama yöntemini temel alır [147]. Fonksiyon, Podani’nin çalışmasında özelleştirdiği Gower benzerlik hesaplamasını gerçekleştirir ve hesaplanan benzerlikleri uzaklığa dönüştürür [148]. Ayrıca Legrende ve Legrende’nin yöntemine göre değişken ağırlıklarını hesaplar ve uzaklık matrisini oluşturur [149].

Oluşturulan uzaklık matrisinin bağımlı değişken olduğu ve bu matris içindeki varyasyonun GWAS analizinde de kullanılan toplam 724.067 adet SNP ile incelendiği MDMR modeli çalıştırılmıştır. Belirlenmiş olan kovaryantlar da regresyon modeline eklenmiştir. Modelin çalıştırılmasında yine R yazılımı (Version: x64 2.11.1) kullanılmıştır. Araştırmada işbirliği yürütülen The Scripps Research Institute araştırma enstitüsüne bağlı Translational Science Institute bünyesinde Prof. Dr. Nicholas Schork önderliğindeki araştırma laboratuvarında Dr. Rany M. Salem tarafından geliştirilmiş olan MDMR R scripti kullanılarak tüm analizler iki basamakta yapılmıştır. Bu script, talep edildiği takdirde araştırmacılarla bilimsel kullanım amacıyla paylaşılmaktadır [102].

Birinci basamakta tüm 724.097 SNP için F istatistikleri hesaplanmıştır. Permütasyon testleri gerçekleştirilmeden yapılan bu analizde amaç en yüksek F istatistiğine sahip SNPler’i tespit etmek ve F istatistiği değerine göre SNPler’i sıralayarak en tepedeki 2.500 SNP’i seçebilmektir. F istatistiği küçüldükçe p değerinin 1’e yaklaşacağı bilindiğinden, istatistiksel olarak anlamlı olma olasılığı daha fazla olan yüksek F istatistiği hesaplanmış SNPler’e odaklanmak ve böylece anlamlı olmayacağı kesin olan çok sayıda SNP’i büyük bilgisayar kaynağı ve zaman tüketimi gerektiren permütasyon testlerine sokmamak esas alınmıştır.

İkinci basamakta ise ilk basamakta F istatistik değerlerine göre belirlenmiş olan 2.500 adet SNP ile permütasyon testleri gerçekleştirilmiştir. Böylece her SNP için p değerlerini hesaplamak amaçlanmıştır.

Sadece 1 adet SNP için tek bilgisayar işlemcisi kullanılarak 1.000 permütasyonla çalıştırılan analiz yaklaşık 80 dakika sürmektedir. Buradan hareketle, sıralı listeden bazı SNPler belirlenerek bu SNPler’e 1.000 permütasyonlu analiz uygulanmıştır. Burada amaç tüm analiz için seçilmesi gereken SNP sayısını belirleyebilmektir. Böylece 500üncü, 1.000inci, 2.500üncü ve 5.000inci SNP için

olmak üzere toplam 4 adet MDMR analizi uygulanmıştır. Sonuç olarak 500üncü SNP ve 1.000inci SNP anlamlı bulunurken 2.500üncü SNP ve 5.000inci SNP'in anlamlı olmadıkları gözlenmiştir. Böylece asıl analiz için sıralı listedeki ilk 2.500 SNP seçilmiştir ve her SNP için 1.000 permütasyon testi yapılmıştır. Bu basamakta alfa değeri 0,005 olarak belirlenmiştir.

3.2.4. GWAS ve MDMR Sonuçlarının Birlikte Değerlendirilmesi

Araştırmada temel olarak kullanılan iki yöntem yukarıda açıklanan GWAS analizi ve MDMR modelidir. GWAS sonucunda bipolar bozukluk hastalarının normal kontrol bireylerine göre toplam 724.067 SNP bakımından allel frekansı farklılıkları araştırılmıştır. MDMR uygulamasında ise bipolar bozukluk hastaları 25 adet fenotipik özellik bakımından uzaklık matrisi oluşturularak incelenmiş, 724.067 SNP'in bu uzaklık matrisindeki varyasyona etkileri araştırılmıştır.

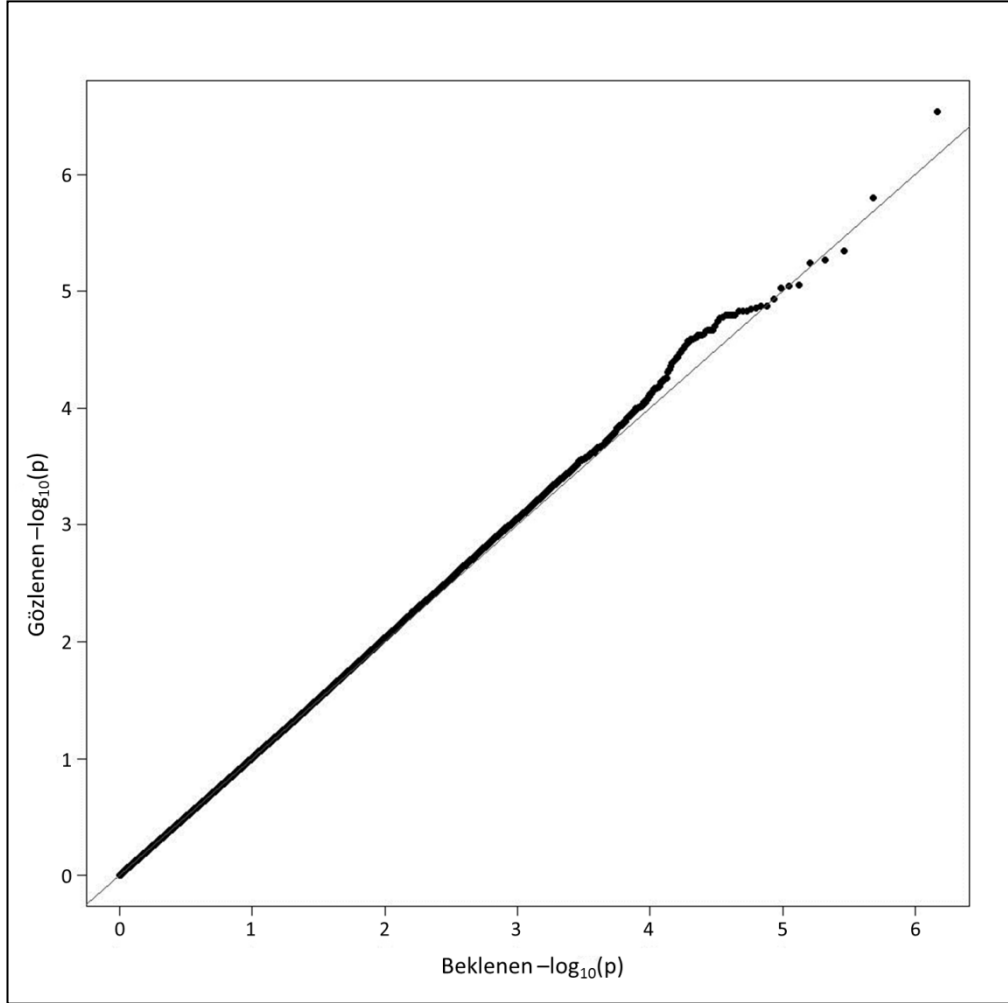
Her iki yöntem sonucunda elde edilen istatistiksel olarak anlamlı SNPLer yöntemlerin amacına uygun şekilde ayrı ayrı değerlendirilmiştir. Bunun yanında her iki araştırmada da anlamlı bulunması olası SNPLer'in varlığına bakılmıştır. İki temel araştırma sorusuna da istatistiksel olarak anlamlı bulunarak cevap veren SNPLer bir üçüncü değerlendirmeye tabi tutulmuştur.

BULGULAR

4.1. GWAS Bulguları

Araştırmanın ilk basamağında bipolar bozukluğu tanısı almış 1.001 kişi ile kontrol olguları (1.033 kişi) arasında genom çapında ilişki çalışması analizi gerçekleştirilmiştir. Bu analizde incelenen toplam 724.067 adet SNP için anlamlılık düzeyi 0,0001 olarak kabul edilmiştir.

Hesaplanan p değerlerine ilişkin Q-Q grafiği Şekil 4.4. ile gösterilmiştir. Beklenen ve gözlenen p değerleri arasındaki dağılımın uyumu bu grafikte incelenmiştir.



Şekil 4.1. Tüm SNPlar için GWAS ile Elde Edilen p Değerlerine ait Q-Q Grafiği

Tablo 4.1 ise yapılan GWAS analizi sonucunda anlamlı bulunan tüm SNPlere ait özellikleri göstermektedir. Bu tablo düzeltilmiş p değerlerine göre sıralanmıştır. SNPlerin rs sayıları, hangi kromozom üzerinde oldukları, eğer biliniyorsa hangi gene ait oldukları, allel bazları, birinci allelin hasta ve kontrollerdeki görülme sıklığı, analiz sonucunda elde edilen ki kare istatistiği değeri, düzeltilmemiş ve düzeltilmiş p değerleri ve odds oranları tabloda görülebilir.

Örneğin ilk sırada gösterilen rs1825828 numaralı SNP 3. kromozom üzerindedir ve hangi gen üzerinde konumlandığı belirlenmemiştir. Birinci allel sitozin ikinci allel ise timindir. Sitozin alleli veri grubunda hastalarda %22 sıklıkta görülmüştür. Kontrollerde görülme sıklığı ise %29,27 olarak bulunmuştur. Bu durumda allelin odds oranınının 0,68 olduğu gözlenmektedir. Ki kare değeri 27,05 olarak hesaplanmıştır. Düzeltilmiş p değerine göre GWAS analizinin sonunda hasta ve kontrol grupları arasında görülme sıklığı bakımından istatistiksel anlamlılığı en düşük bulunan SNP rs1825828 olmuştur ($p=0,000000288$).

Odds oranlarına bakılarak Tablo 4.1 bulguları incelendiğinde, en yüksek odds oranına sahip SNP'in rs1956817 olduğu görülmektedir (odds oranı=3,79). Bu SNP için ilk allel olan guanin hastalarda %1,82 sıklıkta görülürken kontrollerde görülme sıklığı %0,49 bulunmuştur.

Yine odds oranlarına bakılarak Tablo 4.1'de görülen en düşük odds oranlı iki SNP'in rs2129554 ve rs8013833 oldukları görülebilir. Her iki SNP için de birinci allelin hastalarda görülme sıklığı kontrollerde görülme sıklığına göre daha düşüktür. rs2129554 numaralı SNP hastalarda %0,71 sıklıkla görülürken kontrol bireylerinde görülme sıklığı %2,03 bulunmuştur. rs8013833 numaralı SNP için ise hastalarda görülme sıklığı %0,65 ve kontrollerde görülme sıklığı ise %2,13 olarak bulunmuştur.

Anlamlı bulunan tüm 91 SNP'den 40'ı için odds oranı 1'den düşüktür (minimum=0,3 ve maksimum=0,78). Geriye kalan 51 SNP için ise odds oranı 1'den büyük bulunmuştur (minimum=1,3 ve maksimum=3,79).

Tablo 4.1 ile gösterilen kromozom bilgisi dikkate alındığında anlamlı bulunmuş olan 91 adet SNP'in 20'sinin (%21,98) 2 numaralı kromozom üzerinde olduğu görülmektedir. Ayrıca SNPlerin bilinen genlere göre dağılımına bakıldığında da 10 adet SNP'in NAP5 geni üzerinde olduğu görülmektedir (%10,99).

Şekil 4.2 ise analize katılmış olan tüm SNPler için hesaplanan p değerlerinin logaritmik dönüşümle gösterildiği grafikdir. Analiz için kabul edilen p değerine (0,0001) ait referans çizgisi işaretlenmiştir. Tüm SNPler genom üzerindeki fiziksel konumlarına göre konumlandırılmıştır. Böylece her kromozom için SNPler işaretlenmiş ve referans çizgisinin üzerinde kalan anlamlı SNPler gösterilmiştir.

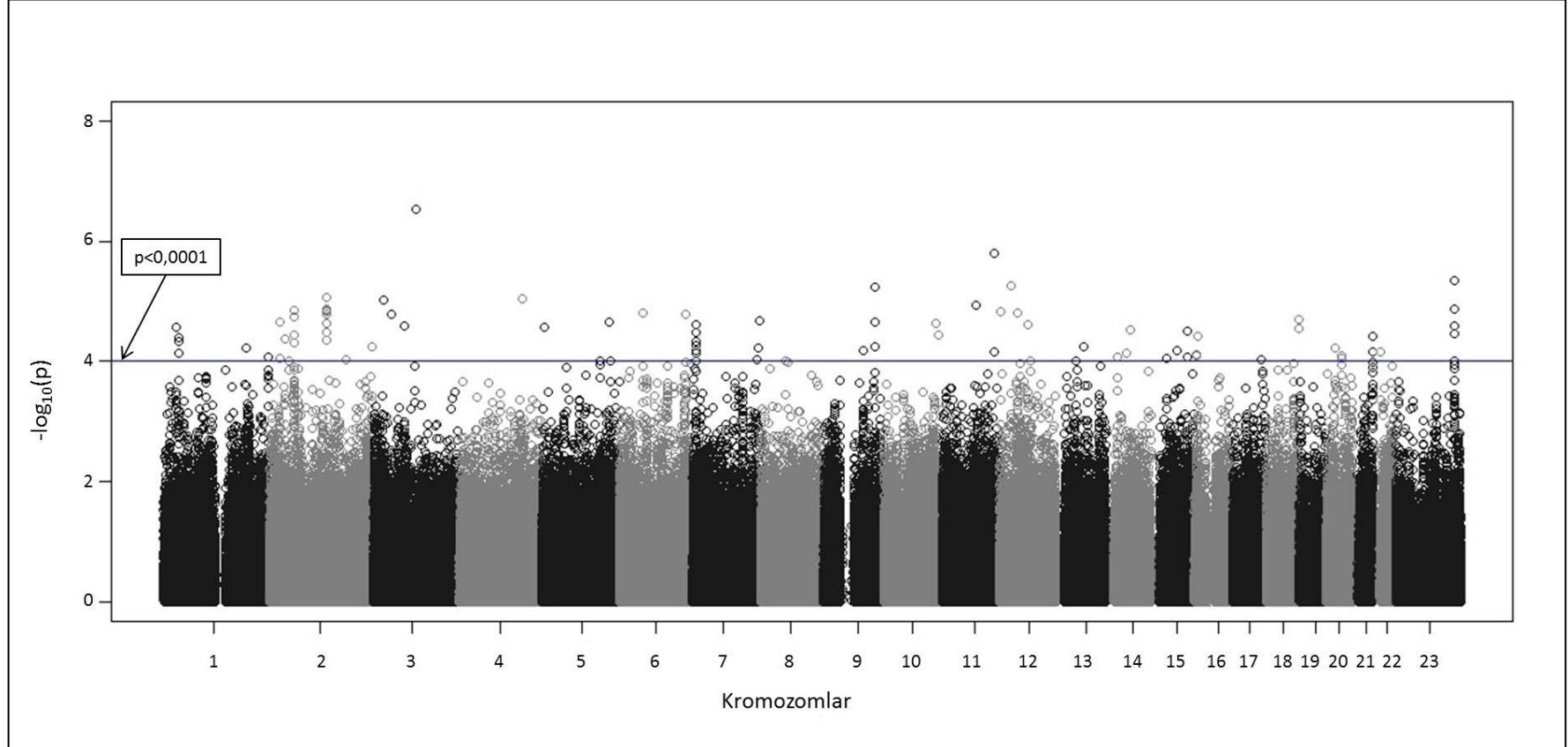
Tablo 4.1. GWAS Analizi Sonucunda $p < 0,0001$ Düzeyinde Anlamlı Olduğu Belirlenen Toplam 91 SNP'e ait Özellikler

Kromozom	SNP	Gen	Allel 1	Hastalarda Allel 1 Frekansı (%)	Kontrollerde Allel 1 Frekansı (%)	Allel 2	Ki Kare	p	Düzeltilmiş p	Odds Oranı
3	rs1825828	-	C	22	29,24	T	27,05	0,000000198	0,000000288	0,68
11	rs4936819	-	T	3,873	7,415	G	23,66	0,00000115	0,000001597	0,50
23	rs5907577	-	T	34,62	26,83	C	21,6	0,00000336	0,000004549	1,44
12	rs4964010	ITPR2	G	31,03	38,15	T	21,27	0,00000398	0,000005364	0,73
9	rs11789399	DBC1	G	51,35	44,14	A	21,15	0,00000424	0,000005704	1,34
2	rs10193871	NAP5	G	10,43	15,16	A	20,31	0,0000066	0,00000877	0,65
4	rs7690204	GYP A	A	45,11	52,19	T	20,27	0,00000673	0,00000894	0,75
3	rs620918	UBE2E2	G	47,37	40,33	A	20,17	0,00000709	0,000009399	1,33
11	rs7931512	-	C	21,23	27,2	T	19,74	0,00000886	0,00001168	0,72
2	rs13430905	NAP5	A	10,64	15,3	G	19,48	0,0000102	0,00001334	0,66
23	rs613278	-	G	52,6	44,59	A	19,48	0,0000102	0,00001336	1,38
2	rs1016771	VRK2	G	45,04	38,22	A	19,41	0,0000105	0,00001384	1,33
2	rs13431418	NAP5	T	10,24	14,81	C	19,35	0,0000109	0,00001431	0,66
12	rs4765864	CACNA2D4	G	24,93	19,22	A	19,3	0,0000112	0,00001467	1,40
2	rs10496702	NAP5	A	10,29	14,86	G	19,27	0,0000114	0,00001489	0,66
2	rs7600871	NAP5	A	10,29	14,86	G	19,27	0,0000114	0,00001489	0,66
6	rs9475671	-	T	47,47	40,62	C	19,15	0,0000121	0,00001584	1,32
12	rs11532502	-	A	16,35	21,75	G	19,14	0,0000121	0,00001588	0,70
6	rs12523857	TIAM2	A	48,25	41,43	T	19,12	0,0000123	0,00001607	1,32
6	rs12528887	TIAM2	T	48,25	41,43	C	19,12	0,0000123	0,00001607	1,32
2	rs13399558	NAP5	A	12,26	17,12	T	19,06	0,0000127	0,00001655	0,68
3	rs14380	VIPR1	A	19,59	14,41	T	19,03	0,0000129	0,00001679	1,45
2	rs2043890	VRK2	C	30,06	36,55	T	18,89	0,0000138	0,00001801	0,75
18	rs537962	KCNG2	C	37,86	31,41	G	18,69	0,0000154	0,00002001	1,33

Kromozom	SNP	Gen	Allel 1	Hastalarda Allel 1 Frekansı (%)	Kontrollerde Allel 1 Frekansı (%)	Allel 2	Ki Kare	p	Düzeltilmiş p	Odds Oranı
7	rs10949808	C7orf13	T	42,06	35,48	G	18,55	0,0000166	0,00002152	1,32
9	rs11789407	ADAM19	G	50,3	43,56	T	18,53	0,0000167	0,00002167	1,31
5	rs11740562	DBC1	C	7,792	4,55	T	18,53	0,0000167	0,00002168	1,77
2	rs41523444	CENPO	G	3,719	1,556	C	18,53	0,0000167	0,0000217	2,44
10	rs7068008	FGFR2	A	42,08	48,83	G	18,4	0,0000179	0,00002318	0,76
2	rs17800749	NAP5	T	10,64	15,15	C	18,37	0,0000181	0,00002349	0,67
2	rs12469282	NAP5	T	12,37	17,17	A	18,37	0,0000182	0,0000236	0,68
12	rs10784460	MSRB3	T	47,94	41,17	C	18,34	0,0000185	0,00002395	1,32
7	rs4724977	C1GALT1	T	25,47	19,86	C	18,27	0,0000192	0,00002478	1,38
23	rs677033	-	G	52,53	44,78	A	18,24	0,0000194	0,00002512	1,37
23	rs4825220	-	T	52,26	44,52	C	18,2	0,0000199	0,00002565	1,36
3	rs6775777	RYBP	A	15,21	10,71	G	18,19	0,00002	0,00002577	1,50
5	rs6891243	LOC442132	T	49,1	42,45	A	18,13	0,0000207	0,00002667	1,31
1	rs11580589	PTPRU	C	24,69	30,77	T	18,12	0,0000208	0,0000268	0,74
18	rs4799092	KCNG2	G	37,86	31,52	C	18	0,0000221	0,00002844	1,32
14	rs1092015	-	T	3,38	1,361	C	17,94	0,0000228	0,00002936	2,54
15	rs6603020	LOC648809	G	30,31	24,39	A	17,86	0,0000238	0,00003064	1,35
2	rs13412008	NAP5	T	12,45	17,15	G	17,77	0,000025	0,00003209	0,69
7	rs7800159	FLJ20323	A	24,38	18,94	C	17,7	0,0000259	0,00003327	1,38
23	rs5909027	-	T	52,28	44,64	C	17,64	0,0000266	0,00003415	1,36
10	rs17153882	FANK1	G	1,236	3,202	C	17,52	0,0000285	0,00003648	0,38
2	rs848291	FANCL	T	44,01	37,56	C	17,5	0,0000288	0,00003682	1,31
16	rs4786850	LOC440337	C	32,12	26,17	A	17,44	0,0000297	0,00003795	1,34
21	rs11701130	SLC19A1	A	20,23	25,75	C	17,4	0,0000302	0,00003863	0,73

Kromozom	SNP	Gen	Allel 1	Hastalarda Allel 1 Frekansı (%)	Kontrollerde Allel 1 Frekansı (%)	Allel 2	Ki Kare	p	Düzeltilmiş p	Odds Oranı
1	rs577483	CLSPN	A	10,56	14,92	G	17,33	0,0000314	0,00004007	0,67
2	rs1158219	EIF2AK2	T	35,41	41,76	A	17,28	0,0000323	0,0000412	0,76
2	rs6749561	NAP5	C	16,48	21,59	T	17,16	0,0000344	0,0000438	0,72
7	rs10486158	FLJ20323	C	24,27	18,94	T	17,05	0,0000363	0,00004622	1,37
1	rs535638	CLSPN	C	10,59	14,91	T	17,01	0,0000372	0,0000473	0,68
2	rs2717055	FANCL	G	34,17	40,41	A	16,92	0,000039	0,00004951	0,77
7	rs13233490	-	G	2,198	0,6776	C	16,72	0,0000434	0,00005492	3,29
2	rs12692245	ASB1	C	48,02	41,62	T	16,69	0,000044	0,0000557	1,30
13	rs2129554	-	A	0,7136	2,303	G	16,67	0,0000446	0,0000564	0,30
9	rs10984107	DBC1	C	42,36	36,11	T	16,66	0,0000447	0,00005651	1,30
1	rs10922013	KCNT2	A	19,84	25,33	C	16,6	0,0000462	0,00005846	0,73
7	rs4726211	ACTR3B	G	44,21	50,58	A	16,57	0,0000468	0,0000592	0,77
20	rs6046396	RIN2	G	31,52	25,75	A	16,55	0,0000475	0,00006002	1,33
9	rs7858079	DIRAS2	A	30,92	36,93	G	16,38	0,0000517	0,0000652	0,76
7	rs2881814	-	T	24,37	19,14	C	16,36	0,0000525	0,00006616	1,36
9	rs10821402	DIRAS2	A	31,02	37,03	C	16,34	0,0000528	0,00006654	0,76
7	rs6955140	RORA	A	24,02	18,8	G	16,32	0,0000534	0,00006725	1,37
15	rs3934516	-	G	4,45	7,454	C	16,32	0,0000535	0,00006739	0,58
11	rs11219172	-	T	3,846	6,68	G	16,31	0,0000539	0,00006786	0,56
22	rs9604779	BCL2L13	G	10,85	7,226	A	16,26	0,0000551	0,00006936	1,56
21	rs2839020	PCBP3	A	24,51	30,23	G	16,23	0,0000562	0,00007067	0,75
1	rs6688464	TFAP2E	G	10,94	15,2	A	16,2	0,000057	0,00007163	0,69
7	rs2159710	COL28A1	A	24,14	18,95	C	16,14	0,0000589	0,00007396	1,36
14	rs8013833	LOC283551	A	0,6494	2,13	G	16,13	0,0000592	0,00007437	0,30

Kromozom	SNP	Gen	Allel 1	Hastalarda Allel 1 Frekansı (%)	Kontrollerde Allel 1 Frekansı (%)	Allel 2	Ki Kare	p	Düzeltilmiş p	Odds Oranı
16	rs7188257	A2BP1	C	32,92	27,15	T	16,08	0,0000606	0,00007608	1,32
20	rs6023059	-	C	44,51	50,77	T	16,02	0,0000628	0,00007874	0,78
16	rs13336322	A2BP1	C	15,25	11,01	G	15,99	0,0000635	0,00007966	1,45
1	rs2275313	SMYD3	G	7,9	11,62	C	15,9	0,0000669	0,00008377	0,65
14	rs1956817	-	G	1,82	0,4869	A	15,88	0,0000675	0,00008456	3,79
15	rs7178655	-	A	35,01	29,19	C	15,86	0,0000683	0,00008552	1,31
20	rs6022716	FLJ35695	G	44,55	50,78	A	15,78	0,0000712	0,00008905	0,78
2	rs35213472	CENPO	G	3,497	1,549	A	15,78	0,0000713	0,00008911	2,30
15	rs2643217	C20orf77	T	24,25	19,12	C	15,78	0,0000713	0,00008917	1,35
2	rs2969363	MTX2	C	39,75	33,71	T	15,72	0,0000735	0,00009189	1,30
7	rs2969637	-	G	38,46	44,58	T	15,66	0,0000756	0,00009441	0,78
17	rs2193055	ATP6V0E2	C	21,33	16,47	A	15,66	0,000076	0,00009484	1,38
5	rs952037	C2orf34	A	32,22	26,57	G	15,62	0,0000773	0,00009647	1,31
2	rs17498753	ADRA1B	C	12,97	17,48	T	15,62	0,0000776	0,00009685	0,70
13	rs7336303	-	A	3,378	1,45	C	15,6	0,0000784	0,00009781	2,38
7	rs17158578	-	G	2,585	0,9406	C	15,59	0,0000788	0,00009823	2,80
23	rs1244670	-	C	47,8	40,68	T	15,57	0,0000793	0,00009892	1,34
5	rs2673925	TRPC7	T	32,67	27,01	C	15,56	0,0000797	0,00009943	1,31
12	rs328764	TBC1D15	A	20,08	15,36	G	15,56	0,0000798	0,0000995	1,39



Şekil 4.2. Tüm SNPler için Kromozomlara Göre GWAS Sonuçlarına ait Anlamlılık Grafiği

Hasta ve kontrol grupları SNP seviyesinde karşılaştırıldıktan sonra, tüm SNPlar'ın fiziksel olarak genom üzerinde buldukları pozisyonlar esas alındı ve böylece sonuçların gen düzeyinde de anlamlılıkları incelendi. Aşağıda sunulan Tablo 4.2'de gen düzeyinde yapılan analiz sonucunda alfa=0,005 düzeyinde anlamlı bulunan genlere (101 gen) ait bulgular gösterilmiştir.

Tablo 4.2. GWAS ile Hesaplanan SNP p Değerlerinin Gen Düzeyinde Değerlendirilmesi

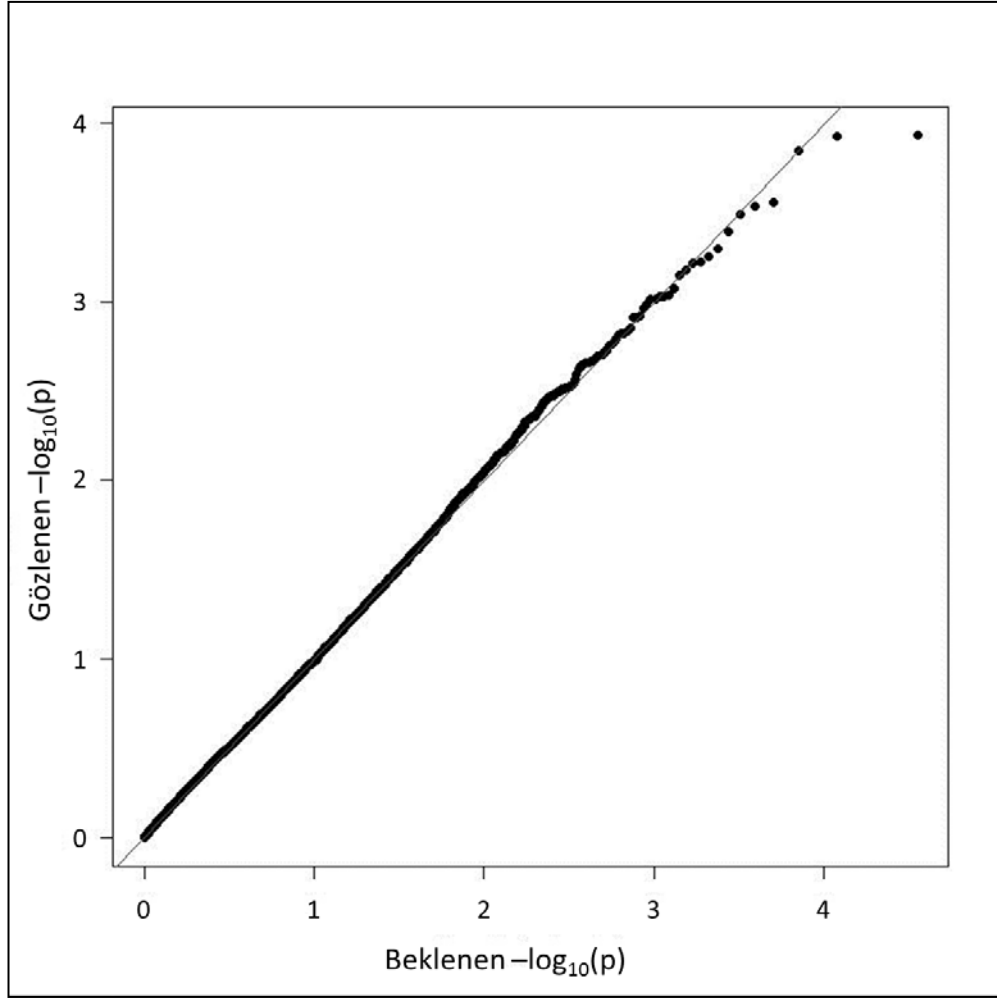
Kromozom	Gen	SNP Sayısı	Başlangıç	Bitiş	p
20	STK4	46	43028533	43142007	0,000117
20	KCNS1	23	43154363	43163167	0,000118
12	OR10AD1	28	46882388	46883342	0,000143
20	WFDC5	23	43171506	43177217	0,000275
17	RPL38	21	69711389	69717614	0,000291
12	LOC387856	28	46863665	46865976	0,0004
12	PIP4K2C	19	56271323	56283298	0,0005
12	SLC26A10	19	56299959	56306201	0,000557
12	GEFT	17	56290229	56297293	0,000595
1	EIF2C1	10	36121396	36162486	0,000611
12	DTX3	16	56284870	56289850	0,000663
12	B4GALNT1	17	56305817	56313252	0,000833
18	CDH19	30	62322300	62422196	0,000916
17	C17orf55	12	76891218	76897643	0,000926
1	OR2C3	20	245760056	245763764	0,000934
2	COPS8	22	237658822	237672228	0,001028
17	TMEM105	15	76899668	76919069	0,001083
16	CREBBP	29	3715056	3870122	0,001203
1	EIF2C3	11	36169358	36294650	0,001221
6	SYNE1	235	152484514	153000227	0,001221
12	ASB8	31	46827838	46837644	0,00139
1	EIF2C4	6	36046414	36093775	0,001444
7	SLC26A4	26	107088315	107145488	0,001499
12	KIF5A	19	56230113	56264821	0,00151
7	WNT2	39	116704517	116750579	0,00153
2	QPCT	40	37425256	37453969	0,00161
7	ASZ1	25	116790511	116854813	0,00165
20	WFDC12	22	43185480	43186520	0,001719
17	SLC38A10	17	76833393	76883691	0,001751
19	NCLN	22	3136874	3160573	0,001876
15	ZNF592	19	83092821	83150667	0,001882
2	PRKD3	45	37331149	37397726	0,001957
12	ZNF641	20	47022178	47030268	0,001979
19	BRUNOL5	24	3175700	3248073	0,00199
2	MBOAT2	33	8914151	9061327	0,00201

20	CABLES2	29	60397080	60415734	0,002082
1	OLFML2B	84	160219605	160260268	0,002138
7	CFTR	45	116907252	117095954	0,00216
12	H1FNT	20	47009029	47010329	0,00217
22	TCF20	21	40885962	40941389	0,00217
12	TBC1D15	45	70519805	70604362	0,00218
1	NEK7	43	196468304	196555392	0,0022
1	C1orf216	8	35952063	35957377	0,002267
1	CLSPN	8	35974404	36008138	0,002275
20	KIAA0406	25	36044836	36095247	0,00229
12	RAB21	16	70434924	70467417	0,00236
15	CDAN1	19	40803051	40816709	0,002529
13	CCDC122	39	43308488	43351826	0,002707
17	C17orf89	14	76827705	76829693	0,00285
13	C13orf31	30	43351419	43366068	0,00289
11	C11orf10	17	61313177	61316661	0,002941
11	C11orf9	18	61276696	61312565	0,00296
11	POLD3	37	73981276	74031413	0,003
17	KRTAP9-4	12	36659464	36660431	0,003
11	FEN1	17	61316725	61321286	0,00304
20	C20orf151	28	60418687	60435984	0,00307
12	TMEM19	16	70366144	70384106	0,00308
17	KRTAP9-8	14	36647795	36648782	0,00308
12	BHLHB3	40	26164225	26169113	0,00312
20	C20orf77	27	36095361	36154180	0,00314
1	OR2W5	31	245721052	245722015	0,00316
7	CTTNBP2	44	117137941	117300797	0,003171
1	JMJD4	23	225985557	225989735	0,00324
17	C17orf56	15	76816673	76827450	0,00327
22	PHF21B	114	43655705	43784245	0,00331
1	SEMA4A	21	154390011	154414159	0,00333
7	MAD1L1	109	1821953	2239109	0,00334
12	MSRB3	38	63958754	64146954	0,00334
1	PDZK1IP1	26	47421847	47428358	0,00337
20	RPS21	19	60395515	60396971	0,00341
1	C1orf142	27	225989319	226035550	0,003457
6	HIST1H1A	15	26125238	26126019	0,00352
10	ABLIM1	130	116180858	116434404	0,00352
17	KRTAP9-9	19	36642277	36666142	0,00359
21	ZNF295	53	42280008	42303565	0,003638
15	CTDSPL2	8	42506870	42606721	0,00365
15	EIF3J	6	42616557	42642293	0,00374
6	RPP40	49	4940278	4949270	0,00386

3	THOC7	13	63794585	63824637	0,00391
19	DIRAS1	11	2665564	2672369	0,00396
7	TNRC18	37	5312948	5429703	0,00404
1	TMEM51	81	15352815	15419459	0,00412
17	CCT6B	16	30279050	30312619	0,00418
22	SEPT3	21	40702876	40724171	0,00422
6	GLULD1	14	64047518	64087841	0,00435
6	HIST1H3A	16	26128696	26129165	0,00437
22	WBP2NL	20	40724737	40754423	0,00438
11	FADS1	19	61323676	61340886	0,0044
5	TRPC7	65	135577021	135720972	0,00441
6	PHF3	13	64414389	64482364	0,00441
19	VMAC	13	5855851	5861263	0,00445
8	ARC	18	143689411	143692835	0,00451
22	MKL1	31	39136237	39362636	0,004519
6	HIST1H4A	16	26129885	26130257	0,00452
2	SULT6B1	22	37248466	37269194	0,0046
14	PPM1A	36	59782222	59835559	0,00461
8	GPR20	18	142435768	142446547	0,00467
20	LAMA5	24	60317515	60375763	0,00467
11	NLRX1	14	118544649	118559936	0,00494
1	CYP4A22	20	47375693	47387113	0,00497
17	KRTAP17-1	27	36724694	36725473	0,00498

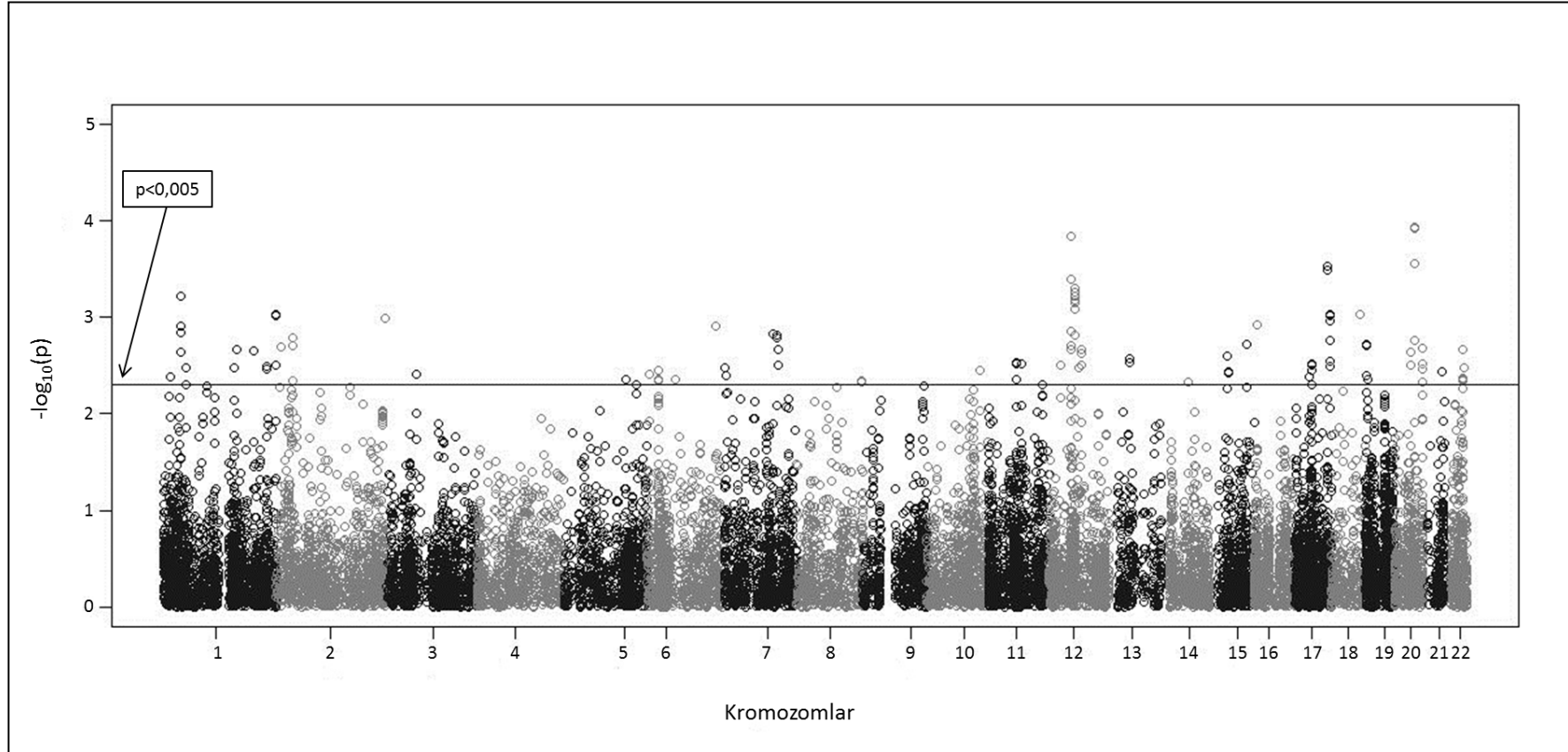
Tablo 4.2’de görülen anlamlı 101 gen için kromozom temelli bakıldığında, 12. kromozomun toplam 16 adet anlamlı gen ile en sık anlamlı gen bulunan kromozom olduğu görülmektedir (%15,8). Ayrıca 1 numaralı kromozom (1 gen %14,9) ve 17 numaralı kromozom da (11 gen ve %10,9) anlamlı bulunan gen sayısı bakımından ikinci ve üçüncü sıradadırlar.

Bu analiz sonucunda elde edilen p değerlerinin yanlılık ve dağılımlarını incelemek amacıyla Şekil 4.3’de sunulan Q-Q grafiği çizilmiştir.

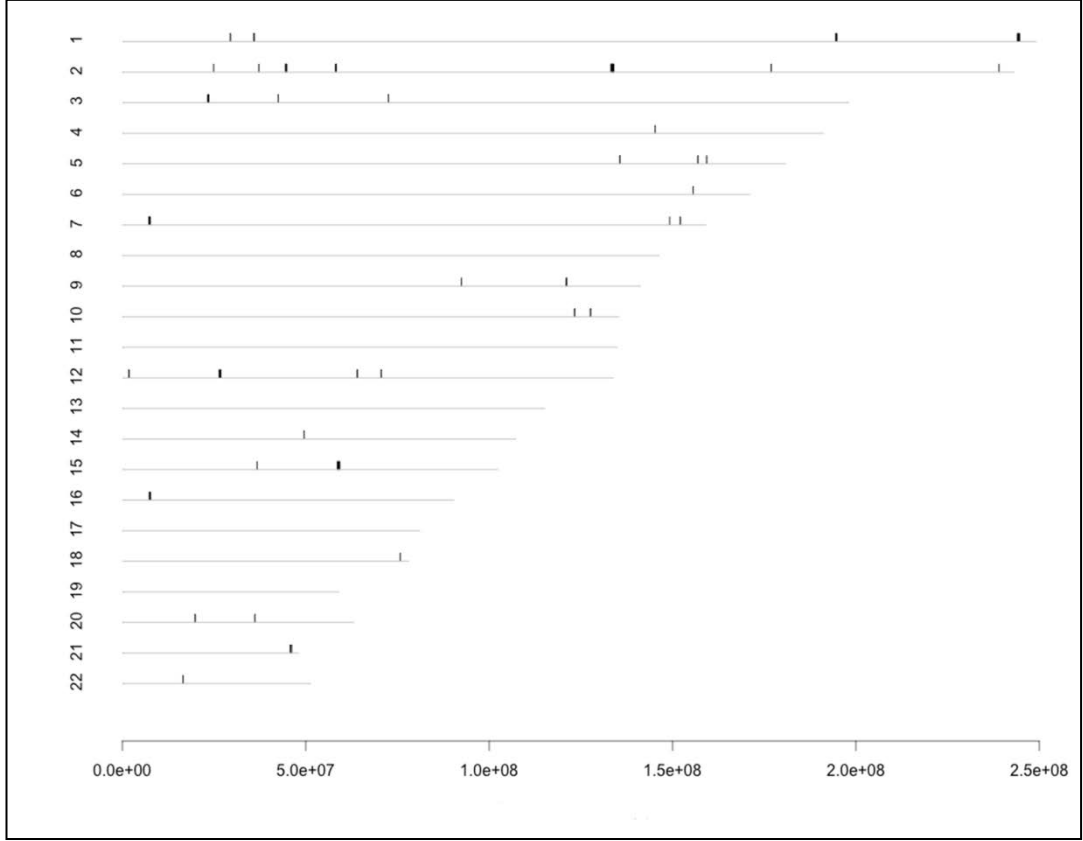


Şekil 4.3. Tüm Genler için GWAS ile Elde Edilen p Değerlerine ait Q-Q Grafiği

Son olarak, tüm 724.067 SNP'in toplam 17.724 adet gene haritalandığı ve tüm genler için p değerinin hesaplandığı analiz sonucunda elde edilen anlamlılık düzeyleri Şekil 4.4 ile gösterilmiştir. Burada mevcut analiz için belirlenen 0,005 anlamlılık düzeyi referans çizgisi ile işaretlenmiş olup bu referansın üzerinde kalan noktalar ile anlamlı bulunan genlerin tümü işaretlenmiştir. Genom üzerindeki fiziksel konumlarına göre yerleştirilen genler için kromozom bilgileri de grafikte görülebilmektedir.



Şekil 4.4. Tüm Genler için Kromozomlara Göre GWAS Sonuçlarına ait Anlamlılık Grafiği



Şekil 4.5. Anlamli Bulunan Genlerin Kromozomdaki Pozisyonlarına Göre Gösterimi

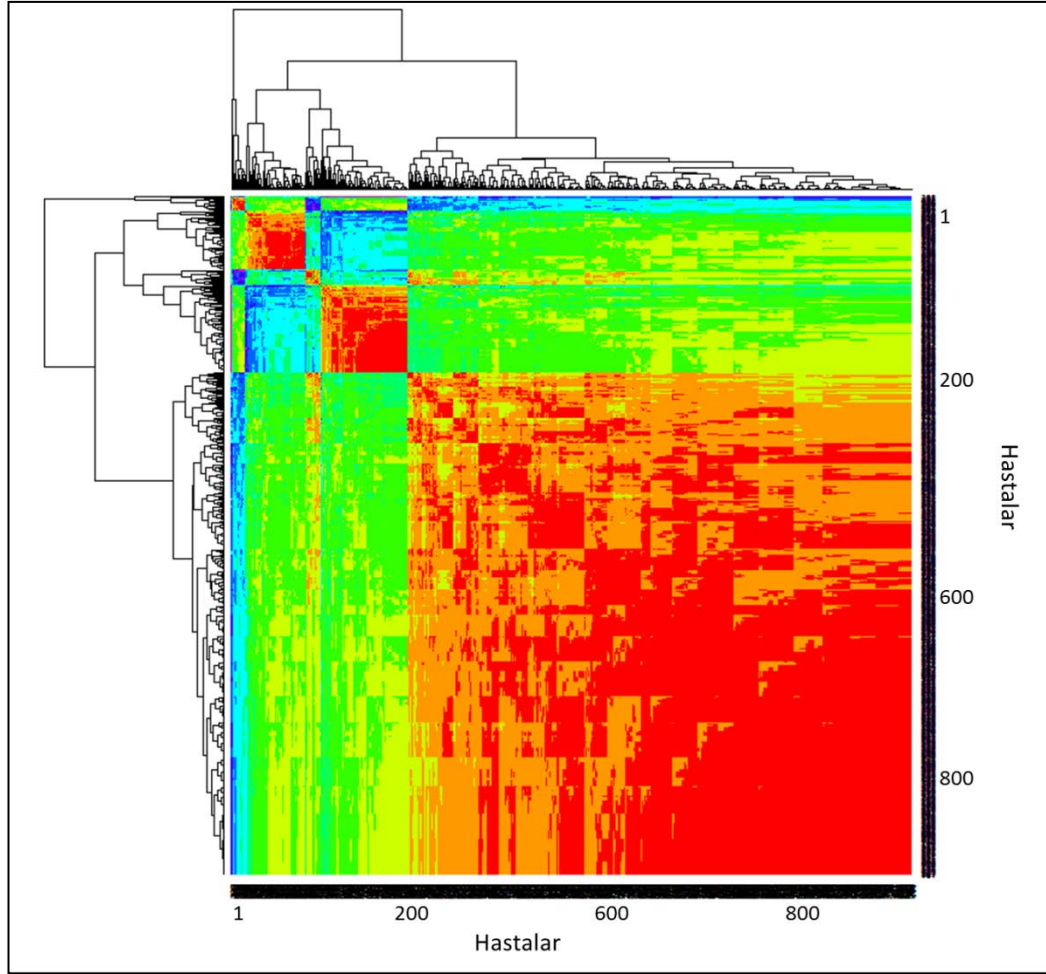
Şekil 4.5’de görüldüğü üzere, 8, 11, 13, 17 ve 19 numaralı kromozomlarda gen düzeyinde anlamlılık tespit edilememiştir. Daha koyu işaretlemelerin görüldüğü bazı bölgelerde ise fiziksel konum olarak birbirine yakın genlerin istatistiksel anlamlı olduğu görülmektedir. Örneğin 1, 2, 3, 7, 12 ve 15 numaralı kromozomlar üzerinde böylesi anlamlı ve birbirine yakın gen lokasyonları olduğu gözlenmektedir.

4.2. MDMR Bulguları

Araştırmanın ikinci ve temel basamağı olan MDMR uygulaması ise bipolar bozukluk tanısı almış hastaların fenotipik özellikleri bakımından birbirleri ile ikili farklılıklarına dayanan bir uzaklık matrisi temelli regresyon modeli ile incelenmiştir.

Uzaklık matrisi hesaplaması için hastalık seyri ile ilişkili olduğu düşünülen toplam 25 adet fenotip kullanılmıştır. Kayıp verilerle ilgili filtreleme ve çalışmaya dâhil edilme ölçütlerinin uygulanması sonucunda toplam 873 adet hastaya ait veriler kullanılmıştır.

Şekil 4.6 ile hesaplanmış olan 873 x 873 boyutundaki uzaklık matrisinin ısı grafiği biçimindeki gösterimi sunulmuştur. Kırmızı tam benzerlik mavi ise tam farklılık ifade eden renk kodudur.



Şekil 4.6. Hastalara ait Uzaklık Matrisinin Grafiksəl Gösterimi

Şekil 4.6’da grafiksəl gösterimi sunulmuş olan uzaklık matrisinin bağımlı deęişken olduęu ve veri setindeki tüm SNPler’in de bağımsız deęişkenler olarak uzaklık matrisindeki varyasyonu açıklamak üzere düşünöldüęü MDMR analizinin ilk temel basamağında permütasyon testleri kullanılmadan tüm 724.067 SNP için F istatistięi hesaplanmıştır.

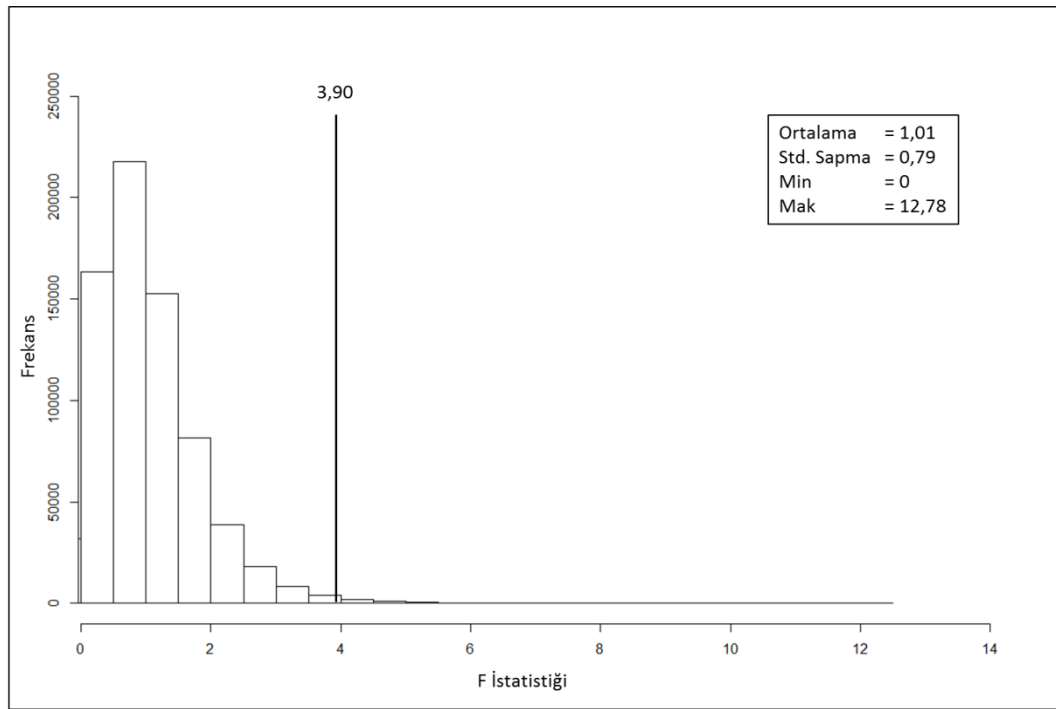
Hiç permütasyon testi uygulamadan tek bir SNP için tek bilgisayar işlemcisi ayrılarak yürütölen analiz süresi yaklaşık 1 saniye sürmektedir. Tüm SNPler’in F istatistięinin hesaplanma süresi, tek işlemci kullanarak yaklaşık 200 saattir.

SNPler’in F istatistięi deęerleri hesaplandıktan sonra permütasyon testlerine geçilerek istatistiksel olarak anlamlı SNPler’i tespit etmek için filtreleme uygulanmıştır. F istatistięi düşük olan çok sayıda SNP’in istatistiksel olarak anlamlı olamayacağı, yani p deęeri belirlenmiş alfa deęeri olan 0,005’in altında olacağı için, düşük F istatistięi hesaplanan SNPler’e permütasyon testi yapılmamıştır. Böylece çok fazla zaman ve bilgisayar kaynağı gerektiren ve istatistiksel olarak önemli olmayacağı kestirilen SNPler dışarıda bırakılmıştır.

Bu aşamada F istatistięine göre sıralanan SNPler’den en yüksek F istatistięine sahip belirli sayıda SNP permütasyon testlerine dâhil edilmiştir. Bu sayıyı

belirleyebilmek amacıyla, sıralı listeden bazı SNPlar seçilmiş ve bu SNPlar'e 1.000 permütasyonlu MDMR analizi uygulanmıştır. Seçilen SNPlar, F istatistiği en yüksek olan 500üncü, 1.000inci, 2.500üncü ve 5.000inci SNPlar'dir. Bu analizlerin sonunda elde edilen p değerlerine bakıldığında 500üncü SNP ve 1.000inci SNP anlamlı bulunurken 2.500üncü SNP ve 5.000inci SNP'in anlamlı olmadıkları gözlenmiştir. Böylece asıl analiz için sıralı listedeki ilk 2.500 SNP'in seçilmesine karar verilmiştir. Seçilen bu 2.500 SNP 1.000'er permütasyon testi ile analiz edilmiştir. Bu basamakta alfa değeri 0,005 olarak belirlenmiştir.

Şekil 4.7'de 724.067 SNP'e ait F istatistiği dağılımı grafiği çizilmiştir. Hesaplanan F istatistiklerinin ortalaması 1,01 ve standart sapması da 0,79 olarak bulunmuştur. En düşük F istatistiği 0 ve en yüksek F istatistiği de 12,78'dir.



Şekil 4.7. MDMR Analizi Sonucunda Tüm SNPlar'e ait F İstatistiği Değerlerinin Histogramı

Yukarıda açıklandığı gibi, en yüksek F istatistiğine sahip 2.500 SNP seçilmiştir. 2.500üncü SNP'in F istatistiği 3,90'dır (Şekil 4.7). Dolayısıyla F istatistiği 3,90 ile 12,78 arasında hesaplanan toplam 2.500 adet SNP permütasyon testlerine tabi tutulmuştur.

Bir adet SNP için 1.000 permütasyonlu tek bilgisayar işlemcisi ayrılan MDMR analizi için geçen süre yaklaşık 80 dakikadır. Toplam 2.500 SNP için analizin tamamlanması yaklaşık 3.300 saat (~140 gün) sürmektedir. Bu sebeple analizler 10 adet paralel işlemci ile yürütülmüş ve kesintisiz yaklaşık 2 hafta zaman almıştır.

Belirlenmiş olan 0,005 düzeyindeki istatistik anlamlılığı 2.500 SNP'den 1.728 tanesi sağlamıştır. Tablo 4.3'de bu SNPlar'den ilk 100 tanesi gösterilmiştir.

Tablo 4.3. MDMR Sonuçlarına Göre İstatistiksel Olarak Anlamlı Bulunan İlk 100 SNP

SNP	N	F İstatistiği	p	Açıklanan Varyasyon Yüzdesi
rs9352737	865	12,775	0	0,0145
rs12035324	867	12,201	0	0,0139
rs12332357	862	11,259	0	0,0129
rs1351959	873	10,113	0	0,0114
rs9375754	873	9,492	0	0,0213
rs4664525	871	9,051	0	0,0204
rs9946845	861	9,543	0	0,0109
rs9361538	864	9,663	0	0,0110
rs4706789	873	9,676	0	0,0109
rs7278456	867	9,008	0	0,0204
rs871436	872	8,961	0	0,0201
rs9964530	873	9,173	0	0,0206
rs1440752	873	9,056	0	0,0203
rs16844893	870	8,998	0	0,0102
rs17211228	873	8,790	0	0,0100
rs9577730	853	8,519	0	0,0099
rs10104788	873	8,179	0	0,0184
rs10506556	873	8,394	0	0,0189
rs2469354	871	8,061	0	0,0182
rs1994069	873	8,447	0	0,0096
rs6828388	872	8,435	0	0,0096
rs6027005	873	8,195	0	0,0184
rs2028945	865	8,147	0	0,0185
rs3783124	825	8,645	0	0,0104
rs6791443	870	8,278	0	0,0187
rs6462731	871	7,872	0	0,0178
rs11074925	871	8,381	0	0,0095
rs16932028	872	7,910	0	0,0178
rs10914351	865	8,801	0	0,0101
rs12370025	870	8,257	0	0,0186
rs7485934	873	7,793	0	0,0175
rs1525299	867	8,049	0	0,0182
rs17251516	873	8,041	0	0,0181
rs678262	873	7,830	0	0,0176
rs7159893	873	7,949	0	0,0179
rs11116610	872	7,846	0	0,0177
rs1782938	870	7,652	0	0,0173
rs1735318	873	7,483	0	0,0168
rs16982391	872	7,667	0	0,0173

rs6562214	863	7,808	0	0,0178
rs16855571	869	7,729	0	0,0088
rs6739985	869	7,520	0	0,0170
rs17345849	873	7,395	0	0,0167
rs2888810	873	7,284	0	0,0164
rs7242702	873	7,746	0	0,0174
rs628299	869	7,342	0	0,0166
rs9955571	871	7,735	0	0,0174
rs4752594	873	7,453	0	0,0085
rs10788208	873	7,453	0	0,0085
rs10861139	873	7,265	0	0,0164
rs8058644	857	7,039	0	0,0162
rs7091811	872	7,422	0	0,0084
rs10749429	872	7,433	0	0,0084
rs10861137	873	7,220	0	0,0163
rs7298655	873	7,230	0	0,0163
rs1659132	857	7,314	0	0,0168
rs3850016	873	7,239	0	0,0163
rs7295499	873	7,239	0	0,0163
rs7314790	873	7,239	0	0,0163
rs7964654	873	7,239	0	0,0163
rs7465143	866	7,293	0	0,0166
rs6715405	858	7,467	0	0,0171
rs9289348	867	6,977	0	0,0158
rs7008053	873	7,551	0	0,0170
rs16869462	873	7,186	0	0,0162
rs17059290	873	7,061	0	0,0159
rs12329322	872	7,275	0	0,0164
rs12570407	873	8,091	0	0,0092
rs10958627	873	7,395	0	0,0084
rs6558795	865	7,235	0	0,0164
rs9576468	842	7,073	0	0,0165
rs7301439	864	7,065	0	0,0161
rs10197160	871	7,460	0	0,0085
rs17113474	873	7,593	0	0,0086
rs9491951	854	7,046	0	0,0082
rs9976065	861	7,314	0	0,0084
rs472730	865	7,260	0	0,0165
rs9372972	873	6,981	0	0,0157
rs995322	872	7,141	0	0,0161
rs9352744	872	7,776	0	0,0088
rs6790032	859	6,850	0	0,0157
rs10829631	872	7,133	0	0,0161

rs12680875	873	7,017	0	0,0158
rs17099529	858	7,594	0	0,0088
rs4943561	863	6,866	0	0,0157
rs727389	867	6,732	0	0,0153
rs528765	865	6,965	0	0,0158
rs9682236	873	6,727	0	0,0152
rs17127810	873	7,029	0	0,0158
rs9321245	870	6,743	0	0,0153
rs13358776	849	7,008	0	0,0162
rs3850017	872	6,745	0	0,0152
rs11250936	847	6,633	0	0,0154
rs17071652	805	6,900	0	0,0169
rs11263754	872	6,521	0	0,0147
rs17781327	873	6,791	0	0,0153
rs16992856	870	6,567	0	0,0149
rs1516555	873	6,772	0	0,0153
rs2162075	872	7,197	0	0,0162
rs12424244	871	6,804	0	0,0154

Tablo 4.3’de görüldüğü gibi bazı SNPler için analize giren kişi sayısı 873’de daha azdır. Bunun nedeni, verinin genotip kısmındaki eksikliklerdir. Bireylerden alınan doku veya kan örneklerinin sekanslama sırasında çeşitli faktörlere bağlı olarak çalışılan tüm genom bölgeleri için genotiplenememesinden kaynaklanmaktadır.

Tablo 4.3’de ayrıca “Açıklanan Varyasyon Yüzdesi” sütunu sunulmuştur. Bu sütundaki değerler, her bir SNP’in uzaklık matrisinde araştırılan varyasyonun ne kadarlık bir kısmını açıklayabildiğini göstermektedir. Örneğin F istatistiği en yüksek bulunan rs9352737 numaralı SNP, uzaklık matrisinde gözlenen varyasyonun %0,0145 kadarını açıklamaktadır.

Analiz sonucunda 0,005 düzeyinde anlamlı bulunan 1.728 SNP için açıklanan varyasyon yüzdesi kısmı toplandığında %19,55 bulunmaktadır. Yani MDMR analizi ile hastaların fenotipik özelliklerini temel alan bir uzaklık matrisi oluşturulup 0,005 düzeyinde istatistiksel olarak anlamlı olan tüm SNPler incelendiğinde, bu 1.728 SNP’in toplam varyasyonun toplamsal olarak %19,55 kadarını açıkladığı gözlenmektedir.

4.3. GWAS ve MDMR Bulgularının Birleştirilmesi

GWAS ve MDMR basamaklarından oluşan tüm çözümlerinin sonucunda elde edilen 2 küme SNP vardır: bu kümelerden birincisi GWAS ile tespit edilen SNPler, ikincisi ise MDMR ile tespit edilen SNPler’dir.

Her iki analiz sonucunda elde edilen sıralı listelerdeki ilk 2.500 SNP içinde herhangi bir kesişim olup olmadığı incelendiğinde ise, her iki analizde ortak olarak en tepedeki 2.500 SNP listesine giren 12 adet SNP olduğu görülmüştür. Bu SNPler Tablo 4.4 ile gösterilmiştir.

Tablo 4.4. GWAS ve MDMR Analizlerinde En Üst 2.500 SNP Arasında Ortak SNPler

SNP	GWAS Sıralaması	GWAS p Deęeri	MDMR Sıralaması	MDMR p Deęeri
rs620918	8	0,000009399	397	0
rs8044639	161	0,0002019	2.489	0,029
rs13338967	223	0,0002667	2.485	0,024
rs41343045	464	0,0005569	2.245	0,009
rs17087294	958	0,001148	2.175	0,009
rs9392129	1.523	0,001876	2.085	0,008
rs11243875	1.588	0,001952	1.002	0,002
rs7863484	1.873	0,002294	1.621	0,004
rs11814078	1.922	0,002369	2.293	0,010
rs9649850	1.967	0,002424	1.414	0,003
rs11838644	2.236	0,002768	2.474	0,020
rs17295058	2.344	0,002924	2.384	0,014

TARTIŞMA

Genetik ilişki arařtırmaları, kompleks hastalıkların ve fenotiplerin ortaya çıkmasına etki eden kalıtsal faktörlerin neler olduğunu anlayabilmek için özellikle 20. yüzyılın ikinci yarısından itibaren çok sayıda bilimsel çalışmada uygulanmıştır.

DNA'nın yapısal olarak keşfi 1953 senesinde James D. Watson ve Francis Crick tarafından gerçekleştirilmiştir. Bu çalışma ile birlikte bilim dünyasında yepyeni bir devir açılmış, biyolojik bilimler ve tıp alanında birbirini takip eden birçok yeni buluş görülmüştür. Canlıların genetiğine ilişkin bilinenler arttıkça, yüzyıllardır nesilden nesle iletiildiği gözlenen özelliklerin hangi mekanizmalarla aktarıldığına ilişkin soruların da cevabını arayan arařtırmalar tasarlanarak uygulanmaya başlanmıştır.

Öte yandan, teknoloji alanında kaydedilen yüksek ivmeli gelişmeler, tüm bilim alanlarında olduğu gibi, moleküler biyoloji ve genetik alanında da kendini göstermeye başlamıştır. Önceleri laboratuvar ortamında emek-yoğun, yüksek konsantrasyon isteyen, zaman alıcı ve görece düşük güvenilirlikli sonuçlar veren yöntemler ile yürütülen arařtırmalar yapılmaktaydı. Fakat yüksek mikrodizi teknolojileri, DNA sekanslama teknikleri ve mikroçipler kullanılmaya başlandıkça benzer arařtırmaların sayısı ve kalitesi de artmaya başladı.

Yeni milenyum ile birlikte Human Genome Project ve International HapMap Project gibi iki büyük arařtırmanın şekillenmesi, genetik çalışmaların bilimsel arařtırmaların en önemli odaklarından biri haline dönüşmesini sağlamıştır. Genetik arařtırmalar sadece klinik tıbbi ilgilendiren alanlarda değil, biyoloji, gıda mühendisliği, viroloji, ziraat ve daha birçok dala da damga vurmaya başlamıştır.

Tüm bu gelişmelerin neticesinde ortaya çıkan en büyük sonuçlardan birisi devasa boyutlarda verinin üretilmesi olmuştur. Verilerin işlenmesi, kalite kontrolü, depolanması ve analiz edilmesi başlı başına yeni bir disiplinin evrilmesine yol açmıştır.

Temelleri matematik, istatistik, biyolojik bilimler ve enformasyon teknolojileri üzerine kurulu olan biyoinformatik alanı, varoluş sebebi gereği çok disiplinli bir alan olarak ortaya çıkıp gün geçtikçe daha da fazla önem kazanmaya başlamıştır. Fakat böyle arařtırmalardan elde edilecek verileri işleyerek bilgiye

dönüştürecek şekilde analiz edebilmek için yeni ve etkin stratejilere ihtiyaç duyulmaktadır.

GWAS arařtırmaları, genetik verilerin etkin kullanıldıđı temel yöntemlerin bařında gelmektedir. İlki 2005 yılında yayınlanan bu arařtırmalar, birçok hastalık ve fenotipin kalıtsallıđına iliřkin önemli bulgulara eriřmemize olanak tanımıřtır [53]. Bununla birlikte, zaman ierisinde GWAS alıřmalarının kısıtlılıkları daha da netleřmiř ve yeni alternatif yaklařımlara gereksinim duyulduđu görülmüřtür [79].

Geleneksel istatistiksel yaklařımlar ve yaygın kullanılan ok deđiřkenli yöntemlerin genomik verilere uygulanması ise bazı aılardan sorgulanmaktadır. Örneđin tek deđiřkenli geleneksel istatistiksel özümleme yöntemlerinin, birçok faktörden etkilenen ve bu faktörlerin toplamsal etkisi sonucunda ortaya ıktıđı düşünölen problemlere yaklařımı yetersiz kalabilmektedir. Yapılan alıřmalar göstermiřtir ki Mendeliyen özellik gösteren sınırlı sayıdaki hastalık ve fenotipler diřında birçok kalıtsal özellik kompleks yapıdadır. Genetik bilginin evre faktörleri ve davranıřsal özelliklerle etkileřimi belirleyici olabilmektedir [59, 101]. Bu aıdan tek deđiřkenli yöntemlerin özömlenmeleri birçok genetik problemi aısından tatmin edici olmamaktadır.

Standart ok deđiřkenli istatistiksel analizlerde ise genetik arařtırmalara özgü bazı zorluklar görölebilmektedir. Bu zorluklardan biri, ok fazla sayıda deđiřkenle yürütölen fakat örnekleme büyüklüđu görece az olan arařtırma tasarımlarıdır. Günümüzde birçok genetik arařtırmada 1.000 – 2.000 civarında bireye ulařılabilmektedir. Maliyetlerin ok fazla sayıda olgu ile alıřılabilmesini kısıtlaması, büyük örneklemlere eriřmenin dođal zorlukları, arařtırılan bazı hastalık ve fenotiplerin toplumda nadir görölmesi gibi sebeplerle kiři sayısı genellikle belirli düzeylerin üstüne ıkamamaktadır. Bununla birlikte hemen hemen her genetik arařtırmada ok fazla sayıda faktörün etkisi arařtırılmak istenmektedir. Örneđin gen ekspresyonu temelli arařtırmalarda yaklařık 20.000 civarı gene ait veri toplanmaktadır. Yine standart bir GWAS iin bir milyon civarında SNP'in etkisi incelenmektedir. Örnekleme sayısının kısıtlı, incelenen bađımsız deđiřken sayısınına örnekleme sayısına göre ok fazla olması, geleneksel ok deđiřkenli analiz yöntemlerinin uygulanma zorluklarının bařlıcalarındandır [14]. Böylesi durumlarda alıřmanın istatistiksel gücünün düřük olması istenmeyen bir durumdur.

Bir diđer önemli konu da istatistiksel hatanın oklu testler neticesinde büyümesidir. ok sayıda deđiřkenin incelendiđi arařtırmalarda hesaplanan istatistiksel hataların oklu testlere göre düzeltilmesi, birçok arařtırmanın odaklandıđı bir konudur [150]. Örneđin yanlış bulunuřluk oranı (false discovery rate – FDR) yaklařımı ile oklu karşılařtırmalarda düzeltmeler yapılabilmektedir. FDR, hata sonucunda reddedilen yokluk hipotezlerinin (tip I hata) toplam reddedilen hipotezler iindeki beklenen oranını esas alarak düzeltme uygular. Standart

istatistiksel çözümlerden sık kullanılan hata düzeltme yaklaşımlarından birisi olan Bonferroni düzeltmesine kıyasla daha esnektir. Bu nedenle genetik arařtırmalarda kullanılma oranı da fazladır. Çok sayıda FDR düzeltmesi yaklařımı öne sürölmüřtür. Genetik arařtırmalarda en sık kullanılan yöntemlerin bařında Benjamini-Hochberg yöntemi gelmektedir [151].

Çok deęişkenli istatistiksel yöntemler, yoğun hesaplama gerektiren matematiksel uygulamalardır. Her ne kadar bilgisayar donanım ve yazılım teknolojilerinde büyük ilerlemeler gerçekteşmiş de olsa, çok boyutlu verilerin analizinde bu yöntemlerin kullanılması büyük miktarlarda bilgisayar kaynağına ve uzun zamanlara gereksinim duymaktadır. Bu sorundan kaçınabilmek için veri indirgeme yaklaşımları önerilebilir. Fakat birçok kompleks fenotipin ortaya çıkışında popölasyonda nadir görölen fakat etkilenen bireyin genomunda çok sayıda bulunması muhtemel varyantların toplamsal etkisi olduęu düşünöldüęünde, indirgeme yaklaşımlarının da yetersiz olacağı görölebilir [33, 35, 36]. Bu nedenle genomik arařtırmalar için, elde edilmiş verileri her detayı ile deęerlendirebilecek çözümlerlerin kullanılması gerekmektedir.

Bütün bu sıralanan sebeplerin etkisiyle, genomik arařtırmalarla elde edilen verilerin analizi için yeni yaklaşımlara gereksinim duyulmaktadır. Bu yaklaşımların hem istatistiksel olarak güçlü testler olması hem de kaynak kullanımını bakımından etkin olması gerekmektedir.

MDMR yöntemi de benzer problemler için şekillendirilmiş çok deęişkenli bir çözümlerme aracıdır. Yöntemin bireyler arasında ölçölen benzerlik / uzaklık temelli bir bağımlı deęişken matrisini temel alarak çok sayıda genetik deęişkeni kullanıp bu matris içindeki varyasyonu açıklaması hedeflenmektedir. İlk teorik temelleri McArdle ve arkadaşları tarafından atılan [97] ve ardından Schork ve arkadaşları tarafından genomik arařtırmalarda da kullanılabilen üzere şekillendirilen [14] bu yöntem řu ana kadar çeşitli çalışmalarda uygulanmıştır [15-18, 98]. Yöntemin ağırlıklı olarak uygulandığı genetik arařtırmaların ortak özelliğı, gen ekspresyonu temelli arařtırmalar olmalarıdır. Bireyler arasındaki benzerlik veya farklılıklar, genlerin ekspresyon deęerlerine göre hesaplanmış ve bağımlı deęişken matrisi böyle oluşturulmuştur [14]. Matrisin grafiksel gösterimi, bireyler arasındaki küme ve örüntüleri de şekilsel olarak ortaya koyabilmektedir (Şekil 4.6).

Fakat yöntemin esnekliğı ve istatistiksel güç açısından diđer birçok çok deęişkenli çözümlerme yöntemine göre büyük sayılarda deęişken içeren arařtırmalara uygun olduęunun gösterilmiş olması [14], yöntemin SNP temelli arařtırmalar için de uygun olabileceğini düşöndürmektedir. Yukarıda da vurgulandığı gibi standart bir SNP temelli arařtırmada, SNP sayısı bir milyona kadar ulaşabilmektedir.

Ayrıca birçok arařtırmacı sadece belirli bir fenotipin / hastalığın varlığı veya yokluęunun nedenselliğini deęil, hastalıkların seyrini belirleyen kalıtsal faktörleri

incelemektedir. Geleneksel vaka – kontrol çalışmalarının ötesinde, bireylerden alınan fenotipik bilgilerin önemli olduğu benzer arařtırmalarda toplanan bu bilgilerin analizi için MDMR yönteminin uygun olduđu düşünülebilir.

Bu arařtırmada MDMR yöntemi ilk kez fenotip X genotip etkileşimini inceleyebilmek amacıyla kullanılmıştır. Bipolar bozukluk hastalarından toplanmış olan ve her biri klinik olarak hastalık seyriyle yakın ilişkili kabul edilen çok sayıda farklı fenotip temel alınmış ve bireylerin birbirleri ile gösterdikleri farklılıklar incelenmiştir. Bu amaçla fenotip temelli bir uzaklık matrisi oluşturulmuş ve GWAS arařtırması için toplanmış olan 725.000 civarı SNP bilgisi bu matristeki varyasyonu açıklamada kullanılmıştır.

Hepsi aynı tanıyı almış bireylerin bütünüyle benzer hastalık seyri göstermeyebileceği bilinmektedir. Birçok kompleks hastalık için farklı hastalık evreleri ve karakteristikleri vardır. Nöropsikiyatrik bozukluklar da bu anlamda oldukça heterojen seyre sahiptirler. Bir hastalığın patofizyolojisi biliniyorsa (örneğin bir enzim eksikliği), bu durumda aday genleri belirlemek ve DNA sekansının hangi bölgesinin hastalığa neden olduğunu tespit etmek görece kolaydır. Psikiyatrik bozukluklar içinse patofizyoloji genellikle bilinmemektedir [123]. Ayrıca kompleks hastalıkların nadir görülen ama bireyde fazla sayıda bulunabilen genetik varyantların toplamsal etkisinden kaynaklanabileceği de düşünüldüğünde, hasta bireylerin taşıdıkları genetik varyasyon miktarına göre hastalık seyirlerinin deęişiklik gösterebileceği varsayılabilir.

Buradan hareketle arařtırmamız hasta bireylerin fenotipik verilerini genomlarından elde edilmiş SNP verileri ile açıklayabilme amacına odaklanmıştır. MDMR yönteminin basamaklandırılması ve permütasyon testleri için belirli bir strateji izlenmesi sonucunda toplam 1.728 adet SNP'in 0,005 seviyesinde istatistiksel olarak anlamlı olduđu görülmüştür. Tablo 4.3'de F istatistiğine göre sıralanmış şekliyle en anlamlı bulunan 100 SNP gösterilmiştir. Ayrıca MDMR analizi sonucunda anlamlı bulunan bu 1.728 SNP'in, fenotip bilgileri ile oluşturulmuş uzaklık matrisindeki varyasyonun toplam %19,55'ini açıkladığı bulunmuştur. Bu bulgu, genetik varyasyonlarının etkisinin toplamsallığı bakımından önemlidir. İncelenen SNPler'den yaklaşık %0,24'ü istatistiksel olarak önemli bulunurken bu SNPler toplamda fenotipik varyasyonun yaklaşık %20'sini açıklayabilmektedir.

Şekil 4.6'da hastaların fenotipik bilgileri ile oluşturulan uzaklık matrisinin ısı grafięi ile gösterimi sunulmuştur. Bu grafikten hastaların birbirleri ile oluşturdukları doğal kümeler ve örüntüler de görülmektedir. Bu gösterim, MDMR analizinin görsellik bakımından sağladığı önemli bir avantajdır.

Çalışmamızın bir dięer amacı da hem hasta – kontrol arasında anlamlı etki gösteren hem de sadece hasta grubu içerisindeki fenotipik varyasyonu açıklayan SNPler'in var olup olmadığını inceleyebilmektir. Tablo 4.4'de bu koşulu yerine

getiren 12 adet SNP'e ilişkin bilgiler sunulmuştur. Bu 12 SNP'in hem hastalığın ortaya çıkmasında hem de seyrinde önemli olduğu görülmektedir. Çalışmamız bu özelliğiyle de bipolar bozukluk genetiği alanında yapılmış tek araştırmadır.

Araştırmamızla ilgili en önemli kısıtlılık, MDMR yönteminin çok yoğun kaynak kullanımı gerektirmesidir. Toplam 1.000 adet permütasyon testine tabi tutulan tek bir SNP için harcanan sürenin 80 dakikayı bulması, bütün genomdan alınmış olan yaklaşık 725.000 SNP'in hepsi için permütasyon testlerinin yapılmasını olanaksız kılmaktadır. Yine de F istatistiği temel alınarak uygulamış olduğumuz filtre basamağı ile göz ardı edilebilir etkisi olan SNPleri eleme yönündeki stratejimizin verimli olduğu savunulabilir. Anlamlı bulunan SNP oranının açıkladığı yüksek miktardaki fenotipik varyasyon da bu görüşümüzü desteklemektedir.

Bununla birlikte, genom çapındaki anlamlılık için 1.000'den fazla permütasyon testine ihtiyaç duyulabilir. Genom çapında istatistiksel olarak anlamlı kabul edilebilecek p değerinin 10^{-6} seviyesinde hesaplanması planlanacak olursa permütasyon sayısını bin katına çıkarmak gerekecektir. Bu durumda benzer bir çalışmanın tamamlanması için 1.000 paralel işlemci kullanıldığı varsayılırsa kabaca 140 güne ihtiyaç duyulacaktır. Paralel işlemci kullanımı ve analize ilişkin bir planlama stratejisi şarttır. Bu açıdan yöntemin diğer benzer çok değişkenli yaklaşımlara kıyasla bir avantaj sağlamadığı söylenebilir.

Tüm genomdan SNP düzeyinde toplanan veriler genellikle GWAS araştırmalarının temelini oluşturmaktadır. Gen ekspresyonu ölçülerek yapılan araştırmalarda ise etkiler gen seviyesinde özetlenmektedir. Araştırmamızda uyguladığımız GWAS yaklaşımında ise hem SNP hem de gen düzeyinde istatistiksel etkiler incelenmiştir. Tablo 4.1'de anlamlı bulunan SNP'ler görülmektedir. Ek olarak, tüm SNPlere ait GWAS sonuçları ile yapılan yeni hesaplamalar ile gen düzeyinde istatistiksel anlamlılık incelenmiştir. Bu ek analize ilişkin bulgular ise Tablo 4.2 ile sunulmuştur. Her iki düzeyde etkinin birlikte sunulduğu araştırmaların daha açıklayıcı olacağı ve gelecek araştırmalarda da buna benzer bir stratejinin benimseneceği düşünülmektedir. Bu bakımdan araştırmamız büyük bir GWAS verisi ile hem SNP hem de gen bazında analiz gerçekleştirilen ilk uygulamalardandır.

Şekil 4.1 ve Şekil 4.3'de GWAS araştırmasının hem SNP temelli hem de gen temelli basamakları için Q-Q grafikleri sunulmuştur. Bu grafikler her iki basamakta da elde edilen p değerlerinin yanlılıktan uzak olduğunu göstermektedir. Özellikle Şekil 4.3'de gösterilen gen temelli basamağa ait sonuçlar bu bakımdan daha da önemlidir. Bunun nedeni, gen temelli genetik araştırmaların esas olarak genlerin ekspresyon düzeyleri ölçülerek gerçekleştirilmesi fakat bizim araştırmamızda bunu SNPler bazında hesaplanan p değerlerinden yola çıkılarak yapılmasıdır. Bu basamakta kullandığımız istatistiksel yöntemin, gerçek veriler ile böylesine yanlılıktan uzak bulgular sağlaması dikkat çekicidir.

Biyoinformatik alanının çok disiplinli doğası gereği, ele alınan araştırma sorusunun ait olduğu birincil disiplinle işbirliği önemlidir. Bu anlamda bulgularımızın, araştırmamızda kalıtsal faktörlerin etkisinin incelendiği önemli bir nöropsikiyatrik hastalık olan bipolar bozukluk tanısının klinik düzeyde yetkin uzmanlarınca değerlendirilmesi önemlidir. Biyoinformatik yaklaşımlar benzer problemler için güvenilir ve etkin çözüm stratejileri üretmek ve uygulamak amacıyla şekillendirilmektedir. Bu stratejilerin bilimsel işbirliği dâhilinde ilgili disiplinlerle beraber uygulanması esastır. Bu bakımdan araştırmamızın bulgularının klinik önemliliği ve yorumlanması temel olarak uzman psikiyatristlerce değerlendirilmelidir.

Bununla birlikte, başta bipolar bozukluk olmak üzere birçok nöropsikiyatrik rahatsızlıkla ilgili şu ana kadar yayınlanmış genetik araştırma bulunmaktadır. Bu araştırmalar sonucunda bu kompleks hastalığın kalıtımını doğrudan açıklayan SNP veya genlerin olmadığı görülmektedir. Fakat bazı araştırmalarda istatistiksel olarak önemli olduğu gösterilen bazı genom bölgelerinin tekrarlı olarak rapor edildiği görülebilir. Bu genlerden başlıcaları DGKH, CACNA1C ve ANK3 genleri olup çeşitli araştırmalarda bu genlerin genom çapında anlamlı oldukları tekrarlı biçimde gösterilebilmiştir [5-12].

Bizim araştırmamızda ulaştığımız bulgular da eski benzer çalışmalara bakılarak değerlendirilebilir. Bu değerlendirme her ne kadar alanın klinik uzmanları tarafından yorumlanmak durumunda da olsa, bulguların karşılaştırmalı incelenmesi araştırmamız bakımından değerlidir.

GWAS basamağında anlamlı bulunan rs11789399, rs10496702, rs10949808, rs11740562 ve rs6046396 numaralı SNPLer daha önce gerçekleştirilen bir meta analiz çalışmasında da bipolar bozukluk ile ilişkilendirilmiştir [152]. Smith ve arkadaşlarının bipolar bozukluğu farklı ırklarda inceledikleri bir diğer araştırmada yine rs10193871 ve rs5907577 numaralı SNPLer raporlanmışlardır [5]. Graae ve arkadaşlarının bipolar genetiğine ilişkin araştırmalarında da rs6023059 SNP'i önemli bulunmuştur [153].

Çok sayıda SNP'inin GWAS sonucunda istatistiksel olarak anlamlı olduğunu göstermiş olduğumuz NAP5 geninin birden çok bipolar çalışmasında hastalıkla ilişkilendirilmiş olduğu görülmektedir [5, 152]. Ayrıca bu gen bir başka psikiyatrik rahatsızlık olan dikkat eksikliği hiperaktivite bozukluğu ile de ilişkilendirilmiştir [154]. Crisafulli ve arkadaşlarının yakın zamanda yaptıkları araştırmada, bizim çalışmamızı destekler şekilde CREBBP geninin bipolar bozukluk ve depresif bozukluk gibi nöropsikiyatrik hastalıklarla ilişkili olduğu gösterilmiştir [155].

GWAS bulgularımıza göre SNP bazında anlamlılığa bakıldığında 8, 17 ve 18 numaralı kromozomlar üzerinde yer alan SNPLer'de istatistiksel anlamlılık görülmediği gözlenmektedir (Şekil 4.2). Sonuçlar gen seviyesinde

değerlendirildiğinde ise 4, 8, 9, 10, 14, 16 ve 21 numaralı kromozomlarda istatistiksel anlamlılık gösteren gen sayısının hiç olmadığı ya da çok az olduğu gözlemlenmiştir (Şekil 4.4).

Araştırmamızın klinik anlamda en önemli basamaklarından biri GWAS uygulaması ile MDMR uygulamasının bulgularını birlikte değerlendirmektir. Tablo 4.4 ile sunulmuş olan 12 adet SNP, hem hasta – kontrol farklılıklarını inceleyen GWAS araştırmasında hem de sadece hasta olgular arasındaki varyasyona odaklanan MDMR analizinde istatistiksel anlamlılığı en yüksek olan ilk 2.500 SNP arasında ortak bulunmaktadır. Bu SNPler'den en anlamlısı olan rs620918, 3'üncü kromozom üzerinde yer almaktadır. Bu SNP, GWAS sonucunda p değeri en düşük bulunan 8. SNP'dir. MDMR analizinde de F istatistiği bakımından en yüksek 397. SNP olduğu görülmüştür. rs620918 numaralı SNP daha önce Smith ve arkadaşlarının yapmış olduğu GWAS araştırmasında da bipolar bozukluk ile ilişkilendirilmiştir [5]. Tablo 4.4'de sunulmuş olan diğer 11 SNP için literatürde nöropsikiyatrik bozukluklarla ilişki rapor eden herhangi bir çalışmaya rastlanmamış olmakla birlikte, bizim araştırmamızda bu SNPler'in hem hastalığın ortaya çıkışıyla hem de seyriyle ilgili istatistiksel önemlilik göstermesinin değerli bir bulgu olabileceği düşünülebilir.

Araştırmamız, genomik çalışmalarda uygulanan yöntemlerin gerçek veriler ile test edilmesi açısından önemlidir. Özellikle MDMR uygulamasının fenotip X genotip etkileşimi açısından yeni bir yaklaşım sunduğu gerçektir. Hastalıkların veya insan topluluklarının fenotipik açıdan benzerlikleri veya farklılıklarının SNP seviyesinde incelenmesi önemlidir. Özellikle gelecek yıllarda daha da fazla önem kazanacak olan kişiselleştirilmiş tıp uygulamaları bakımından bireylerin taşıdıkları fenotipik özellikler daha değerlendirilecektir. Genetik araştırmalarında bireye ve aileye ait fenotip verilerinin daha fazla toplanıp değerlendirileceği öngörülebilir. MDMR yöntemi, bu veriler için istatistiksel olarak güçlü ve esnek bir çözüm sunmaktadır. Yöntemin kaynak harcaması bakımından optimize edilmesi ve böylelikle daha kısa sürede bulgu vermesi yönünde çalışmalara ihtiyaç vardır. Yazılım açısından da görece daha ağır çalışan R scripti gözden geçirilebilir veya C++ gibi temel programlama dilleri ile yeni bir kod geliştirilebilir. Bizim çalışmamızda kullanılan R scripti SNP verilerindeki eksikliklere izin vermektedir. Bu açıdan mevcut alternatiflerin en uygunudur.

Özetle, biyoinformatik alanının giderek geliştiği ve artan ihtiyaca göre etkin çözümleme tekniklerine gün geçtikçe daha fazla ihtiyaç duyduğu düşünülürse, çalışmamızın MDMR uygulaması bakımından değerli ve gelecekteki araştırmalara ışık tutucu olduğu görülebilir. Kanıta dayalı ve kişiselleştirilmiş tıp alanlarında genomik bilginin daha fazla kullanımına olanak tanıyacağı düşünülen bu yöntemin gerçek verilerle uygulanması değerlidir. Yöntemin kısıtlılıkları giderilir ve gelecekteki optimizasyonu düşünülürse, birçok klinik problemde verimli şekilde kullanılabilmesi gözlenmiştir.

SONUÇLAR

1. GWAS arařtırmaları kısıtlılıklarına raęmen en temel genomik arařtırma yntemidir. Gelecekteki GWAS alıřmalarında teknolojidaki geliřmelerle birlikte genomdan toplanacak olan daha detaylı verinin analizi ile nadir grlen varyantların da etkisi incelebilecektir.
2. Birok kompleks hastalıkta ok sayıda genetik varyasyonun toplamsal etkisi belirleyici olabilmektedir.
3. Genetik arařtırmalar sadece vaka – kontrol tasarımında deęil, fenotipik zelliklerin de arařtırılmasına ynelik dřnlmelidir. Epigenomik yaklařım ile bireylerde ortaya ıkan fenotipik zelliklerin gen X evre etkileřimi ile řekillendięi dřnlebilir. Bu yzden kompleks hastalıkların anlařılması amacıyla fenotipik zelliklere de odaklanmak gerekir.
4. MDMR yntemi bireylerin eřitli fenotipik zellikler bakımından birbirleri arasındaki varyasyonlarının anlařılabileceęi, kolay grselleřtirilebilme avantajı sunan, esnek ve istatistiksel olarak gl bir yntemdir. alıřmamızda tartıřılan kısıtlılıkları da dikkate alınarak daha da geliřtirilmelidir.
5. Epigenomik, kanıta dayalı tıp ve kiřiselleřtirilmiř tedavi gibi alanlar aısından MDMR yntemi ile elde edilebilecek bulgular deęerlidir.
6. Bipolar bozukluk verisi ile gerekleřtirilen uygulamalar sonucunda klinik olarak deęerli olabilecek bulgulara eriřilmiřtir. Hem SNP dzeyinde hem de gen dzeyinde ulařılan bu bulguların deęerlendirilmesi, hastalık genetięine iliřkin yeni bilgiler sunabilir.
7. zellikle GWAS ve MDMR bulgularının birleřtirilmesi ve elde edilen sonuların literatrle birlikte deęerlendirilmesi, bulguların deęerini artırmaktadır. Biyoinformatik zmlmeleri aısından birden ok yntemin eř zamanlı uygulanması ve bulgularının birleřtirilmesi nemlidir.
8. Arařtırmamız, hem grece yeni bir analiz ynteminin fenotip X genotip etkileřimi problemi zerinde uygulanması bakımından hem de yerleřmiř genetik arařtırma teknikleri ile birlikte deęerlendirmeye olanak tanınması aısından ilk olma zellięindedir. Bulguların klinik nemi psikiyatri alanının kalıtımla ilgilenen uzman hekimlerince deęerlendirilmelidir.

KAYNAKLAR

1. Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS, Mittmann M, Wang C, Kobayashi M, Horton H, Brown EL. Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat Biotechnol* 1996, 14(13):1675-1680.
2. Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* 1999, 17(10):994-999.
3. Bassett DE, Jr., Eisen MB, Boguski MS. Gene expression informatics--it's all in your mine. *Nat Genet* 1999, 21(1 Suppl):51-55.
4. Allison DB, Cui X, Page GP, Sabripour M. Microarray data analysis: from disarray to consolidation and consensus. *Nat Rev Genet* 2006, 7(1):55-65.
5. Smith EN, Bloss CS, Badner JA, Barrett T, Belmonte PL, Berrettini W, Byerley W, Coryell W, Craig D, Edenberg HJ, Eskin E, Foroud T, Gershon E, Greenwood TA, Hipolito M, Koller DL, Lawson WB, Liu C, Lohoff F, McInnis MG *et al.* Genome-wide association study of bipolar disorder in European American and African American individuals. *Mol Psychiatry* 2009, 14(8):755-763.
6. Sklar P, Smoller JW, Fan J, Ferreira MA, Perlis RH, Chambert K, Nimgaonkar VL, McQueen MB, Faraone SV, Kirby A, de Bakker PI, Ogdie MN, Thase ME, Sachs GS, Todd-Brown K, Gabriel SB, Sougnez C, Gates C, Blumenstiel B, Defelice M *et al.* Whole-genome association study of bipolar disorder. *Mol Psychiatry* 2008, 13(6):558-569.
7. Baum AE, Akula N, Cabanero M, Cardona I, Corona W, Klemens B, Schulze TG, Cichon S, Rietschel M, Nothen MM, Georgi A, Schumacher J, Schwarz M, Abou Jamra R, Hofels S, Propping P, Satagopan J, Detera-Wadleigh SD, Hardy J, McMahon FJ. A genome-wide association study implicates diacylglycerol kinase eta (DGKH) and several other genes in the etiology of bipolar disorder. *Mol Psychiatry* 2008, 13(2):197-207.
8. Ferreira MA, O'Donovan MC, Meng YA, Jones IR, Ruderfer DM, Jones L, Fan J, Kirov G, Perlis RH, Green EK, Smoller JW, Grozeva D, Stone J, Nikolov I, Chambert K, Hamshere ML, Nimgaonkar VL, Moskvina V, Thase ME, Caesar S *et al.* Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. *Nat Genet* 2008, 40(9):1056-1058.

9. Scott LJ, Muglia P, Kong XQ, Guan W, Flickinger M, Upmanyu R, Tozzi F, Li JZ, Burmeister M, Absher D, Thompson RC, Francks C, Meng F, Antoniadou A, Southwick AM, Schatzberg AF, Bunney WE, Barchas JD, Jones EG, Day R *et al.* Genome-wide association and meta-analysis of bipolar disorder in individuals of European ancestry. *Proc Natl Acad Sci U S A* 2009, 106(18):7501-7506.
10. Smith EN, Koller DL, Panganiban C, Szlinger S, Zhang P, Badner JA, Barrett TB, Berrettini WH, Bloss CS, Byerley W, Coryell W, Edenberg HJ, Foroud T, Gershon ES, Greenwood TA, Guo Y, Hipolito M, Keating BJ, Lawson WB, Liu C *et al.* Genome-wide association of bipolar disorder suggests an enrichment of replicable associations in regions near genes. *PLoS Genet* 2011, 7(6):e1002134.
11. Ollila HM, Soronen P, Silander K, Palo OM, Kieseppa T, Kaunisto MA, Lonnqvist J, Peltonen L, Partonen T, Paunio T. Findings from bipolar disorder genome-wide association studies replicate in a Finnish bipolar family-cohort. *Mol Psychiatry* 2009, 14(4):351-353.
12. Schulze TG, Detera-Wadleigh SD, Akula N, Gupta A, Kassem L, Steele J, Pearl J, Strohmaier J, Breuer R, Schwarz M, Propping P, Nothen MM, Cichon S, Schumacher J, Rietschel M, McMahon FJ. Two variants in Ankyrin 3 (ANK3) are independent genetic risk factors for bipolar disorder. *Mol Psychiatry* 2009, 14(5):487-491.
13. Nievergelt CM, Libiger O, Schork NJ. Generalized analysis of molecular variance. *PLoS Genet* 2007, 3(4):e51.
14. Zapala MA, Schork NJ. Multivariate regression analysis of distance matrices for testing associations between gene expression patterns and related variables. *Proc Natl Acad Sci U S A* 2006, 103(51):19430-19435.
15. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* 1998, 95(25):14863-14868.
16. Spellman PT, Sherlock G, Zhang MQ, Iyer VR, Anders K, Eisen MB, Brown PO, Botstein D, Futcher B. Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol Biol Cell* 1998, 9(12):3273-3297.
17. Alon U, Barkai N, Notterman DA, Gish K, Ybarra S, Mack D, Levine AJ. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc Natl Acad Sci U S A* 1999, 96(12):6745-6750.

18. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 1999, 286(5439):531-537.
19. Bloss CS, Jeste DV, Schork NJ. Genomics for disease treatment and prevention. *Psychiatr Clin North Am* 2011, 34(1):147-166.
20. Nebert DW, Zhang G, Vesell ES. From human genetics and genomics to pharmacogenetics and pharmacogenomics: past lessons, future directions. *Drug Metab Rev* 2008, 40(2):187-224.
21. Outsmart Your Genes [<http://www.outsmartyourgenes.com/history.html>]
22. Lander ES, Schork NJ. Genetic dissection of complex traits. *Science* 1994, 265(5181):2037-2048.
23. Lederberg J, McCray AT. Ome Sweet ‘Omics -- A Genealogical Treasury of Words. *The Scientist* 2001, 15(7).
24. Leder P. Retrospective. Marshall Warren Nirenberg (1927-2010). *Science* 2010, 327(5968):972.
25. First Successful Laboratory Production of Human Insulin Announced [<http://www.gene.com/gene/news/press-releases/display.do?method=detail&id=4160>]
26. Lander ES, Weinberg RA. Genomics: journey to the center of biology. *Science* 2000, 287(5459):1777-1782.
27. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M *et al*. The sequence of the human genome. *Science* 2001, 291(5507):1304-1351.
28. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K *et al*. Initial sequencing and analysis of the human genome. *Nature* 2001, 409(6822):860-921.
29. The International HapMap Project. *Nature* 2003, 426(6968):789-796.
30. A haplotype map of the human genome. *Nature* 2005, 437(7063):1299-1320.
31. Collins FS, Guyer MS, Charkravarti A. Variations on a theme: cataloging human DNA sequence variation. *Science* 1997, 278(5343):1580-1581.

32. Pritchard JK, Cox NJ. The allelic architecture of human disease genes: common disease-common variant...or not? *Hum Mol Genet* 2002, 11(20):2417-2423.
33. Gershon ES, Alliey-Rodriguez N, Liu C. After GWAS: searching for genetic risk for schizophrenia and bipolar disorder. *Am J Psychiatry* 2011, 168(3):253-256.
34. Lander ES. The new genomics: global views of biology. *Science* 1996, 274(5287):536-539.
35. Singleton A, Hardy J. A generalizable hypothesis for the genetic architecture of disease: pleomorphic risk loci. *Hum Mol Genet* 2011, 20(R2):R158-162.
36. Schork NJ, Murray SS, Frazer KA, Topol EJ. Common vs. rare allele hypotheses for complex diseases. *Curr Opin Genet Dev* 2009, 19(3):212-219.
37. Uher R. The role of genetic variation in the causation of mental illness: an evolution-informed framework. *Mol Psychiatry* 2009, 14(12):1072-1082.
38. Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, Henning AK, SanGiovanni JP, Mane SM, Mayne ST, Bracken MB, Ferris FL, Ott J, Barnstable C, Hoh J. Complement factor H polymorphism in age-related macular degeneration. *Science* 2005, 308(5720):385-389.
39. Manolio TA, Brooks LD, Collins FS. A HapMap harvest of insights into the genetics of common disease. *J Clin Invest* 2008, 118(5):1590-1605.
40. Juran BD, Lazaridis KN. Genomics in the post-GWAS era. *Semin Liver Dis* 2011, 31(2):215-222.
41. Baranzini SE, Galwey NW, Wang J, Khankhanian P, Lindberg R, Pelletier D, Wu W, Uitdehaag BM, Kappos L, Polman CH, Matthews PM, Hauser SL, Gibson RA, Oksenberg JR, Barnes MR. Pathway and network-based analysis of genome-wide association studies in multiple sclerosis. *Hum Mol Genet* 2009, 18(11):2078-2090.
42. Altshuler D, Daly MJ, Lander ES. Genetic mapping in human disease. *Science* 2008, 322(5903):881-888.
43. Jia P, Zheng S, Long J, Zheng W, Zhao Z. dmGWAS: dense module searching for genome-wide association studies in protein-protein interaction networks. *Bioinformatics* 2011, 27(1):95-102.
44. Holmans P, Green EK, Pahwa JS, Ferreira MA, Purcell SM, Sklar P, Owen MJ, O'Donovan MC, Craddock N. Gene ontology analysis of GWA study

data sets provides insights into the biology of bipolar disorder. *Am J Hum Genet* 2009, 85(1):13-24.

45. Pe'er I, Chretien YR, de Bakker PI, Barrett JC, Daly MJ, Altshuler DM. Biases and reconciliation in estimates of linkage disequilibrium in the human genome. *Am J Hum Genet* 2006, 78(4):588-603.
46. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, Yu F, Yang H, Zeng C, Gao Y, Hu H, Hu W, Li C *et al.* A second generation human haplotype map of over 3.1 million SNPs. *Nature* 2007, 449(7164):851-861.
47. Kruglyak L. The road to genome-wide association studies. *Nat Rev Genet* 2008, 9(4):314-318.
48. Cichon S, Craddock N, Daly M, Faraone SV, Gejman PV, Kelsoe J, Lehner T, Levinson DF, Moran A, Sklar P, Sullivan PF. Genomewide association studies: history, rationale, and prospects for psychiatric disorders. *Am J Psychiatry* 2009, 166(5):540-556.
49. Eleftherohorinou H, Wright V, Hoggart C, Hartikainen AL, Jarvelin MR, Balding D, Coin L, Levin M. Pathway analysis of GWAS provides new insights into genetic susceptibility to 3 inflammatory diseases. *PLoS One* 2009, 4(11):e8068.
50. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007, 447(7145):661-678.
51. Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H, Timpson NJ, Perry JR, Rayner NW, Freathy RM, Barrett JC, Shields B, Morris AP, Ellard S, Groves CJ, Harries LW, Marchini JL, Owen KR, Knight B, Cardon LR *et al.* Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* 2007, 316(5829):1336-1341.
52. Manolio TA, Collins FS. The HapMap and genome-wide association studies in diagnosis and therapy. *Annu Rev Med* 2009, 60:443-456.
53. A Catalog of Published Genome-Wide Association Studies [<http://www.genome.gov/gwastudies/>]
54. Xiong Q, Ancona N, Hauser ER, Mukherjee S, Furey TS. Integrating genetic and gene expression evidence into genome-wide association analysis of gene sets. *Genome Res* 2011.

55. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006, 38(8):904-909.
56. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* 2007, 39(7):906-913.
57. McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, Hirschhorn JN. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* 2008, 9(5):356-369.
58. Vineis P, Pearce N. Missing heritability in genome-wide association study research. *Nat Rev Genet* 2010, 11(8):589.
59. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M *et al.* Finding the missing heritability of complex diseases. *Nature* 2009, 461(7265):747-753.
60. Bloss CS, Schiabor KM, Schork NJ. Human behavioral informatics in genetic studies of neuropsychiatric disease: multivariate profile-based analysis. *Brain Res Bull* 2010, 83(3-4):177-188.
61. Stunkard AJ, Foch TT, Hrubec Z. A twin study of human obesity. *JAMA* 1986, 256(1):51-54.
62. Stunkard AJ, Sorensen TI, Hanis C, Teasdale TW, Chakraborty R, Schull WJ, Schulsinger F. An adoption study of human obesity. *N Engl J Med* 1986, 314(4):193-198.
63. Zdravkovic S, Wienke A, Pedersen NL, Marenberg ME, Yashin AI, De Faire U. Heritability of death from coronary heart disease: a 36-year follow-up of 20 966 Swedish twins. *J Intern Med* 2002, 252(3):247-254.
64. King RA, Rotter JI, Motulsky AG. *The Genetic Basis of Common Diseases*: Oxford University Press; 2002.
65. Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, Pukkala E, Skytthe A, Hemminki K. Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med* 2000, 343(2):78-85.
66. Gatz M, Pedersen NL, Berg S, Johansson B, Johansson K, Mortimer JA, Posner SF, Viitanen M, Winblad B, Ahlbom A. Heritability for Alzheimer's disease: the study of dementia in Swedish twins. *J Gerontol A Biol Sci Med Sci* 1997, 52(2):M117-125.

67. Cournil A, Kirkwood TB. If you would live long, choose your parents well. *Trends Genet* 2001, 17(5):233-235.
68. Gudmundsson H, Gudbjartsson DF, Frigge M, Gulcher JR, Stefansson K. Inheritance of human longevity in Iceland. *Eur J Hum Genet* 2000, 8(10):743-749.
69. Sullivan PF, Kendler KS, Neale MC. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch Gen Psychiatry* 2003, 60(12):1187-1192.
70. McGuffin P, Rijsdijk F, Andrew M, Sham P, Katz R, Cardno A. The heritability of bipolar affective disorder and the genetic relationship to unipolar depression. *Arch Gen Psychiatry* 2003, 60(5):497-502.
71. Kieseppa T, Partonen T, Haukka J, Kaprio J, Lonnqvist J. High concordance of bipolar I disorder in a nationwide sample of twins. *Am J Psychiatry* 2004, 161(10):1814-1821.
72. Sullivan PF, Neale MC, Kendler KS. Genetic epidemiology of major depression: review and meta-analysis. *Am J Psychiatry* 2000, 157(10):1552-1562.
73. McGuffin P, Katz R, Watkins S, Rutherford J. A hospital-based twin register of the heritability of DSM-IV unipolar depression. *Arch Gen Psychiatry* 1996, 53(2):129-136.
74. Hetttema JM, Prescott CA, Kendler KS. A population-based twin study of generalized anxiety disorder in men and women. *J Nerv Ment Dis* 2001, 189(7):413-420.
75. Hetttema JM, Neale MC, Kendler KS. A review and meta-analysis of the genetic epidemiology of anxiety disorders. *Am J Psychiatry* 2001, 158(10):1568-1578.
76. Halaschek-Wiener J, Amirabbasi-Beik M, Monfared N, Pieczyk M, Sailer C, Kollar A, Thomas R, Agalaridis G, Yamada S, Oliveira L, Collins JA, Meneilly G, Marra MA, Madden KM, Le ND, Connors JM, Brooks-Wilson AR. Genetic variation in healthy oldest-old. *PLoS One* 2009, 4(8):e6641.
77. Park JH, Wacholder S, Gail MH, Peters U, Jacobs KB, Chanock SJ, Chatterjee N. Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nat Genet* 2010, 42(7):570-575.
78. [<http://www.ncbi.nlm.nih.gov/gap>]

79. Wang K, Li M, Bucan M. Pathway-based approaches for analysis of genomewide association studies. *Am J Hum Genet* 2007, 81(6):1278-1283.
80. Peng G, Luo L, Siu H, Zhu Y, Hu P, Hong S, Zhao J, Zhou X, Reville JD, Jin L, Amos CI, Xiong M. Gene and pathway-based second-wave analysis of genome-wide association studies. *Eur J Hum Genet* 2010, 18(1):111-117.
81. Chen L, Zhang L, Zhao Y, Xu L, Shang Y, Wang Q, Li W, Wang H, Li X. Prioritizing risk pathways: a novel association approach to searching for disease pathways fusing SNPs and pathways. *Bioinformatics* 2009, 25(2):237-242.
82. Wang WY, Barratt BJ, Clayton DG, Todd JA. Genome-wide association studies: theoretical and practical concerns. *Nat Rev Genet* 2005, 6(2):109-118.
83. Hsu YH, Zillikens MC, Wilson SG, Farber CR, Demissie S, Soranzo N, Bianchi EN, Grundberg E, Liang L, Richards JB, Estrada K, Zhou Y, van Nas A, Moffatt MF, Zhai G, Hofman A, van Meurs JB, Pols HA, Price RI, Nilsson O *et al.* An integration of genome-wide association study and gene expression profiling to prioritize the discovery of novel susceptibility Loci for osteoporosis-related traits. *PLoS Genet* 2010, 6(6):e1000977.
84. Akula N, Baranova A, Seto D, Solka J, Nalls MA, Singleton A, Ferrucci L, Tanaka T, Bandinelli S, Cho YS, Kim YJ, Lee JY, Han BG, McMahon FJ. A network-based approach to prioritize results from genome-wide association studies. *PLoS One* 2011, 6(9):e24220.
85. Elbers CC, van Eijk KR, Franke L, Mulder F, van der Schouw YT, Wijmenga C, Onland-Moret NC. Using genome-wide pathway analysis to unravel the etiology of complex diseases. *Genet Epidemiol* 2009, 33(5):419-431.
86. Zhang M, Liang L, Morar N, Dixon AL, Lathrop GM, Ding J, Moffatt MF, Cookson WO, Kraft P, Qureshi AA, Han J. Integrating pathway analysis and genetics of gene expression for genome-wide association study of basal cell carcinoma. *Hum Genet* 2011.
87. Rioux JD, Xavier RJ, Taylor KD, Silverberg MS, Goyette P, Huett A, Green T, Kuballa P, Barmada MM, Datta LW, Shugart YY, Griffiths AM, Targan SR, Ippoliti AF, Bernard EJ, Mei L, Nicolae DL, Regueiro M, Schumm LP, Steinhardt AH *et al.* Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. *Nat Genet* 2007, 39(5):596-604.
88. Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, Boutin P, Vincent D, Belisle A, Hadjadj S, Balkau B, Heude B, Charpentier G, Hudson TJ, Montpetit A, Pshezhetsky AV, Prentki M, Posner BI, Balding DJ, Meyre D *et*

al. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 2007, 445(7130):881-885.

89. Zanke BW, Greenwood CM, Rangrej J, Kustra R, Tenesa A, Farrington SM, Prendergast J, Olschwang S, Chiang T, Crowdy E, Ferretti V, Laflamme P, Sundararajan S, Roumy S, Olivier JF, Robidoux F, Sladek R, Montpetit A, Campbell P, Bezieau S *et al.* Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat Genet* 2007, 39(8):989-994.
90. Gudmundsson J, Sulem P, Steinthorsdottir V, Bergthorsson JT, Thorleifsson G, Manolescu A, Rafnar T, Gudbjartsson D, Agnarsson BA, Baker A, Sigurdsson A, Benediktsdottir KR, Jakobsdottir M, Blondal T, Stacey SN, Helgason A, Gunnarsdottir S, Olafsdottir A, Kristinsson KT, Birgisdottir B *et al.* Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. *Nat Genet* 2007, 39(8):977-983.
91. Moffatt MF, Kabesch M, Liang L, Dixon AL, Strachan D, Heath S, Depner M, von Berg A, Bufe A, Rietschel E, Heinzmann A, Simma B, Frischer T, Willis-Owen SA, Wong KC, Illig T, Vogelberg C, Weiland SK, von Mutius E, Abecasis GR *et al.* Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature* 2007, 448(7152):470-473.
92. Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, Erdos MR, Stringham HM, Chines PS, Jackson AU, Prokunina-Olsson L, Ding CJ, Swift AJ, Narisu N, Hu T, Pruim R, Xiao R, Li XY, Conneely KN, Riebow NL *et al.* A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 2007, 316(5829):1341-1345.
93. Plenge RM, Seielstad M, Padyukov L, Lee AT, Remmers EF, Ding B, Liew A, Khalili H, Chandrasekaran A, Davies LR, Li W, Tan AK, Bonnard C, Ong RT, Thalamuthu A, Pettersson S, Liu C, Tian C, Chen WV, Carulli JP *et al.* TRAF1-C5 as a risk locus for rheumatoid arthritis--a genomewide study. *N Engl J Med* 2007, 357(12):1199-1209.
94. Lesnick TG, Papapetropoulos S, Mash DC, Ffrench-Mullen J, Shehadeh L, de Andrade M, Henley JR, Rocca WA, Ahlskog JE, Maraganore DM. A genomic pathway approach to a complex disease: axon guidance and Parkinson disease. *PLoS Genet* 2007, 3(6):e98.
95. Moonesinghe R, Khoury MJ, Liu T, Ioannidis JP. Required sample size and nonreplicability thresholds for heterogeneous genetic associations. *Proc Natl Acad Sci U S A* 2008, 105(2):617-622.
96. Kraft P, Zeggini E, Ioannidis JP. Replication in genome-wide association studies. *Stat Sci* 2009, 24(4):561-573.

97. McArdle BH, Anderson MJ. Fitting Multivariate Models to Community Data: A Comment on Distance-Based Redundancy Analysis. *Ecology* 2001, 82(1):290-297.
98. Chung D, Zhang Q, Kraja AT, Borecki IB, Province MA. Distance-based phenotypic association analysis of DNA sequence data. *BMC Proc* 2011, 5 Suppl 9:S54.
99. Slonim DK. From patterns to pathways: gene expression data analysis comes of age. *Nat Genet* 2002, 32 Suppl:502-508.
100. D'Haeseleer P. How does gene expression clustering work? *Nat Biotechnol* 2005, 23(12):1499-1501.
101. Feinberg AP. Genome-scale approaches to the epigenetics of common human disease. *Virchows Arch* 2010, 456(1):13-21.
102. Salem RM, O'Connor DT, Schork NJ. Curve-based multivariate distance matrix regression analysis: application to genetic association analyses involving repeated measures. *Physiol Genomics* 2010, 42(2):236-247.
103. Wessel J, Schork NJ. Generalized genomic distance-based regression methodology for multilocus association analysis. *Am J Hum Genet* 2006, 79(5):792-806.
104. Libiger O, Nievergelt CM, Schork NJ. Comparison of genetic distance measures using human SNP genotype data. *Hum Biol* 2009, 81(4):389-406.
105. Baldessarini RJ, Vieta E, Calabrese JR, Tohen M, Bowden CL. Bipolar depression: overview and commentary. *Harv Rev Psychiatry* 2010, 18(3):143-157.
106. Ginsberg SD, Hemby SE, Smiley JF. Expression profiling in neuropsychiatric disorders: Emphasis on glutamate receptors in bipolar disorder. *Pharmacol Biochem Behav* 2011.
107. Schulze TG, Fangerau H, Propping P. From degeneration to genetic susceptibility, from eugenics to genethics, from Bezugsziffer to LOD score: the history of psychiatric genetics. *Int Rev Psychiatry* 2004, 16(4):246-259.
108. Barnett JH, Smoller JW. The genetics of bipolar disorder. *Neuroscience* 2009, 164(1):331-343.
109. Schulze TG. Genetic research into bipolar disorder: the need for a research framework that integrates sophisticated molecular biology and clinically informed phenotype characterization. *Psychiatr Clin North Am* 2010, 33(1):67-82.

110. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry* 2005, 62(6):593-602.
111. Kessler RC, Chiu WT, Demler O, Merikangas KR, Walters EE. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry* 2005, 62(6):617-627.
112. Shih RA, Belmonte PL, Zandi PP. A review of the evidence from family, twin and adoption studies for a genetic contribution to adult psychiatric disorders. *Int Rev Psychiatry* 2004, 16(4):260-283.
113. Citrome L, Goldberg JF. The many faces of bipolar disorder. How to tell them apart. *Postgrad Med* 2005, 117(2):15-16, 19-23.
114. Brieger P, Ehrt U, Marneros A. Frequency of comorbid personality disorders in bipolar and unipolar affective disorders. *Compr Psychiatry* 2003, 44(1):28-34.
115. Haugaard JJ. Recognizing and treating uncommon behavioral and emotional disorders in children and adolescents who have been severely maltreated: bipolar disorders. *Child Maltreat* 2004, 9(2):131-138.
116. Regier DA, Farmer ME, Rae DS, Locke BZ, Keith SJ, Judd LL, Goodwin FK. Comorbidity of mental disorders with alcohol and other drug abuse. Results from the Epidemiologic Catchment Area (ECA) Study. *JAMA* 1990, 264(19):2511-2518.
117. Raimo EB, Schuckit MA. Alcohol dependence and mood disorders. *Addict Behav* 1998, 23(6):933-946.
118. Henry C, Van den Bulke D, Bellivier F, Etain B, Rouillon F, Leboyer M. Anxiety disorders in 318 bipolar patients: prevalence and impact on illness severity and response to mood stabilizer. *J Clin Psychiatry* 2003, 64(3):331-335.
119. Tohen M, Zarate CA, Jr., Hennen J, Khalsa HM, Strakowski SM, Gebre-Medhin P, Salvatore P, Baldessarini RJ. The McLean-Harvard First-Episode Mania Study: prediction of recovery and first recurrence. *Am J Psychiatry* 2003, 160(12):2099-2107.
120. Judd LL, Akiskal HS, Schettler PJ, Endicott J, Maser J, Solomon DA, Leon AC, Rice JA, Keller MB. The long-term natural history of the weekly symptomatic status of bipolar I disorder. *Arch Gen Psychiatry* 2002, 59(6):530-537.

121. Judd LL, Akiskal HS, Schettler PJ, Coryell W, Endicott J, Maser JD, Solomon DA, Leon AC, Keller MB. A prospective investigation of the natural history of the long-term weekly symptomatic status of bipolar II disorder. *Arch Gen Psychiatry* 2003, 60(3):261-269.
122. Geller B, Tillman R, Craney JL, Bolhofner K. Four-year prospective outcome and natural history of mania in children with a prepubertal and early adolescent bipolar disorder phenotype. *Arch Gen Psychiatry* 2004, 61(5):459-467.
123. Schulze TG, McMahon FJ. Genetic linkage and association studies in bipolar affective disorder: a time for optimism. *Am J Med Genet C Semin Med Genet* 2003, 123C(1):36-47.
124. Reich DE, Lander ES. On the allelic spectrum of human disease. *Trends Genet* 2001, 17(9):502-510.
125. Bhatia G, Bansal V, Harismendy O, Schork NJ, Topol EJ, Frazer K, Bafna V. A covering method for detecting genetic associations between rare variants and common phenotypes. *PLoS Comput Biol* 2010, 6(10):e1000954.
126. Clerget-Darpoux F, Elston RC. Are linkage analysis and the collection of family data dead? Prospects for family studies in the age of genome-wide association. *Hum Hered* 2007, 64(2):91-96.
127. Belmonte Mahon P, Pirooznia M, Goes FS, Seifuddin F, Steele J, Lee PH, Huang J, Hamshere ML, Depaulo JR, Jr., Kelsoe JR, Rietschel M, Nothen M, Cichon S, Gurling H, Purcell S, Smoller JW, Craddock N, Schulze TG, McMahon FJ, Potash JB *et al.* Genome-wide association analysis of age at onset and psychotic symptoms in bipolar disorder. *Am J Med Genet B Neuropsychiatr Genet* 2011, 156B(3):370-378.
128. Ruberto G, Vassos E, Lewis CM, Tatarelli R, Girardi P, Collier D, Frangou S. The cognitive impact of the ANK3 risk variant for bipolar disorder: initial evidence of selectivity to signal detection during sustained attention. *PLoS One* 2011, 6(1):e16671.
129. Takata A, Kim SH, Ozaki N, Iwata N, Kunugi H, Inada T, Ujike H, Nakamura K, Mori N, Ahn YM, Joo EJ, Song JY, Kanba S, Yoshikawa T, Kim YS, Kato T. Association of ANK3 with bipolar disorder confirmed in East Asia. *Am J Med Genet B Neuropsychiatr Genet* 2011.
130. McMahon FJ, Akula N, Schulze TG, Muglia P, Tozzi F, Detera-Wadleigh SD, Steele CJ, Breuer R, Strohmaier J, Wendland JR, Mattheisen M, Muhleisen TW, Maier W, Nothen MM, Cichon S, Farmer A, Vincent JB, Holsboer F, Preisig M, Rietschel M. Meta-analysis of genome-wide association data identifies a risk locus for major mood disorders on 3p21.1. *Nat Genet* 2010, 42(2):128-131.

131. Baum AE, Hamshere M, Green E, Cichon S, Rietschel M, Noethen MM, Craddock N, McMahon FJ. Meta-analysis of two genome-wide association studies of bipolar disorder reveals important points of agreement. *Mol Psychiatry* 2008, 13(5):466-467.
132. O'Sullivan BP, Freedman SD. Cystic fibrosis. *Lancet* 2009, 373(9678):1891-1904.
133. Narod SA, Foulkes WD. BRCA1 and BRCA2: 1994 and beyond. *Nat Rev Cancer* 2004, 4(9):665-676.
134. Maher B. Personal genomes: The case of the missing heritability. *Nature* 2008, 456(7218):18-21.
135. Choi KH, Higgs BW, Wendland JR, Song J, McMahon FJ, Webster MJ. Gene expression and genetic variation data implicate PCLO in bipolar disorder. *Biol Psychiatry* 2011, 69(4):353-359.
136. Askland K, Read C, Moore J. Pathways-based analyses of whole-genome association study data in bipolar disorder reveal genes mediating ion channel activity and synaptic neurotransmission. *Hum Genet* 2009, 125(1):63-79.
137. Nurnberger JI, Jr., Blehar MC, Kaufmann CA, York-Cooler C, Simpson SG, Harkavy-Friedman J, Severe JB, Malaspina D, Reich T. Diagnostic interview for genetic studies. Rationale, unique features, and training. NIMH Genetics Initiative. *Arch Gen Psychiatry* 1994, 51(11):849-859; discussion 863-844.
138. Liu JZ, McRae AF, Nyholt DR, Medland SE, Wray NR, Brown KM, Hayward NK, Montgomery GW, Visscher PM, Martin NG, Macgregor S. A versatile gene-based test for genome-wide association studies. *Am J Hum Genet* 2010, 87(1):139-145.
139. Becker RA, Chambers JM, Wilks AR. The new S language: a programming environment for data analysis and graphics: Wadsworth and Brooks/Cole Advanced Books & Software; 1988.
140. Murrell P. R Graphics: Chapman & Hall/CRC; 2006.
141. Getting Genetics Done
[<http://gettinggeneticsdone.blogspot.com/2011/04/annotated-manhattan-plots-and-q-q-plots.html>]
142. Edgington ES. Randomization Tests: M. Dekker; 1995.
143. Manly BFJ. Randomization, Bootstrap and Monte Carlo Methods in Biology: Chapman & Hall; 1997.

144. Neter J, Wasserman W, Kutner MH. Applied linear statistical models: regression, analysis of variance, and experimental designs: R.D. Irwin; 1985.
145. Laliberté E, Legendre P. A distance-based framework for measuring functional diversity from multiple traits. *Ecology* 2010, 91(1):299-305.
146. FD: measuring functional diversity from multiple traits, and other tools for functional ecology [<http://cran.r-project.org/web/packages/FD/>]
147. Gower JC. A General Coefficient of Similarity and Some of Its Properties. *Biometrics* 1971, 27(4):857--871.
148. Podani J. Extending Gower's General Coefficient of Similarity to Ordinal Characters. *Taxon* 1999, 48(2):331-340.
149. Legendre P, Legendre L. *Numerical Ecology*: Elsevier; 1998.
150. Rice TK, Schork NJ, Rao DC. Methods for handling multiple testing. *Adv Genet* 2008, 60:293-308.
151. Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I. Controlling the false discovery rate in behavior genetics research. *Behav Brain Res* 2001, 125(1-2):279-284.
152. Wang KS, Liu XF, Aragam N. A genome-wide meta-analysis identifies novel loci associated with schizophrenia and bipolar disorder. *Schizophr Res* 2010, 124(1-3):192-199.
153. Graae L, Karlsson R, Paddock S. Significant association of estrogen receptor binding site variation with bipolar disorder in females. *PLoS One* 2012, 7(2):e32304.
154. Lasky-Su J, Neale BM, Franke B, Anney RJ, Zhou K, Maller JB, Vasquez AA, Chen W, Asherson P, Buitelaar J, Banaschewski T, Ebstein R, Gill M, Miranda A, Mulas F, Oades RD, Roeyers H, Rothenberger A, Sergeant J, Sonuga-Barke E *et al*. Genome-wide association scan of quantitative traits for attention deficit hyperactivity disorder identifies novel associations and confirms candidate gene associations. *Am J Med Genet B Neuropsychiatr Genet* 2008, 147B(8):1345-1354.
155. Crisafulli C, Shim DS, Andrisano C, Pae CU, Chiesa A, Han C, Patkar AA, Lee SJ, Serretti A, De Ronchi D. Case-control association study of 14 variants of CREB1, CREBBP and CREM on diagnosis and treatment outcome in major depressive disorder and bipolar disorder. *Psychiatry Res* 2012.

ÖZGEÇMİŞ

Özgür Tosun, 1977 yılında İstanbul'da doğdu. İlk ve orta öğrenimini Elazığ'da, lise öğrenimini Ankara'da tamamlayarak 1997 yılında Hacettepe Üniversitesi İktisadi ve İdari Bilimler Fakültesi İktisat (İng) Bölümü'nde üniversite eğitimine başladı. Lisans eğitimini 2001 yılında tamamladı ve 2002 yılında Akdeniz Üniversitesi Sağlık Bilimleri Enstitüsü Biyoistatistik ve Tıp Bilişimi Anabilim Dalı'nda Yüksek Lisans eğitimine başladı. 2005 yılında yüksek lisans tezini verdikten sonra 2006 yılında aynı anabilim dalında doktora eğitimine başladı. Halen Akdeniz Üniversitesi Sağlık Bilimleri Enstitüsü Biyoistatistik ve Tıp Bilişimi Anabilim Dalı'nda Araştırma Görevlisi olarak çalışmaktadır. Yabancı dili İngilizce'dir.

EKLER

Insertion/Deletion Polymorphism and Serum Activity of the Angiotensin-Converting Enzyme in Turkish Patients with Obstructive Sleep Apnea Syndrome

Candan Ogus · Serap Ket · Turker Bilgen ·
Ibrahim Keser · Aykut Cilli · Ayse Yesim Gocmen ·
Ozgur Tosun · Saadet Gumuslu

Received: 20 December 2008 / Accepted: 2 February 2010
© Springer Science+Business Media, LLC 2010

Abstract This study determined the allelic frequency and genotypic distribution of an angiotensin-converting enzyme (ACE) polymorphism and serum ACE activity in Turkish patients with obstructive sleep apnea syndrome (OSAS). A colorimetric assay measured serum ACE activity in 73 of 97 subjects. Frequencies for II, ID, and DD genotypes were 19.6, 53.6, and 26.8% in the OSAS group and 15, 38, and 47% in the control group, respectively ($P = 0.02$). The I allele frequency was higher in the OSAS group than in the healthy control group ($P = 0.02$). Carrying the I allele (II or ID genotypes) increased OSAS risk 2.41 times in the Turkish population. Mean ACE activity was significantly lower in patients with the II genotype than in the DD genotype ($P = 0.011$), and ACE activity was significantly lower in patients with severe OSAS than in those with mild OSAS ($P = 0.006$). Our results suggest that II and ID genotypes of the *ACE* gene increase the risk of developing OSAS in the Turkish population.

Keywords ACE I/D polymorphism · Angiotensin-converting enzyme (ACE) serum activity · Obstructive sleep apnea syndrome (OSAS) · Turkish population

Introduction

The modern obesity pandemic is likely to increase the prevalence of obstructive sleep apnea syndrome (OSAS), the most common form of sleep-disordered breathing (Taheri 2004; Taheri and Mignot 2002). OSAS is associated with snoring, apnea, daytime sleepiness, and significant mortality due to accidents and cardiovascular events (Ursavas et al. 2007). Therefore, understanding the pathophysiologic basis of OSAS is essential for the development of prevention, screening, and therapeutic strategies.

First-degree relatives of patients with OSAS have been shown to be at high risk for development of this disorder. Familial aggregation studies indicate that most of the OSAS risk factors were obesity, ventilatory control abnormalities, and craniofacial dysmorphism (Taheri and Mignot 2002; Kaparianos et al. 2006). The study of Palmer et al. suggests the involvement of multiple genetic factors associated with development of OSAS (Palmer et al. 2004). Although several genes may increase the risk of OSAS, the molecular basis of OSAS development has not been clearly elucidated (Taheri 2004; Riha et al. 2005; Tafti et al. 2007; Bayazit et al. 2006a, b, 2007; Hanaoka et al. 2008; Pierola et al. 2007; Barcelo et al. 2002).

Circulating angiotensin-converting enzyme (ACE) activity shows extensive interindividual variability, and ACE insertion (I)/deletion (D) polymorphism accounts for 47% of the total variance of serum ACE levels (Rigat et al. 1990). Plasma and tissue levels of ACE activity are higher in patients with the DD genotype than in those with the II genotype, and patients with the ID genotype have intermediate ACE levels (Seckin et al. 2006; Ozen et al. 1997). Experimental and anthropological studies indicate that I polymorphism in the ACE gene, which produces reduced serum and tissue ACE activity, is more frequent in individuals with greater endurance and better adaptation to high altitude (Palmer and Redline 2003). Of the few studies investigating the relationship between ACE I/D polymorphism and OSAS, Barcelo et al. (2001) did not find any difference in frequency distribution of the DD, II, and ID genotypes between OSAS patients and healthy subjects. Rubinsztajn et al. (2004) also reported no association between ACE polymorphisms and OSAS. Xiao et al. (1999), however, described the I allele as a risk factor for OSAS in a Chinese population. In addition, a high frequency of the I allele and the II genotype has been closely associated with hypertensive patients who show more severe forms of OSAS (Zhang et al. 2000). In our study, we determined the allelic frequency, genotypic distribution, and serum levels of ACE activity in Turkish patients who presented with OSAS.

Materials and Methods

The study included 97 unrelated Turkish patients (nine women, 88 men) with OSAS, who were diagnosed with polysomnography between 2001 and 2004. The

study was conducted at the Sleep Unit of the Department of Chest Diseases of the Akdeniz University Medical Faculty. Polysomnography was performed with 16-channel EMBLA SX Proxy 3.0 (Medcare, Iceland) with continuous sleep-technician monitoring, consisting of four channels of EEG, two channels of EOG, submental EMG, oronasal air flow, thoracic and abdominal movements, pulse oximeter saturation, tibial EMG, body position detector, electrocardiogram, and tracheal sound. Records were scored at intervals of 30 s. Apnea was defined as complete cessation of airflow lasting ≥ 10 s. Hypopnea was defined as 70% or more reduction in respiratory airflow lasting ≥ 10 s and accompanied by a decrease of $\geq 4\%$ in oxygen saturation. An apnea–hypopnea index (AHI; average number of episodes of apnea and hypopnea per hour of sleep) was used to determine the degree of OSAS. An index of 5–15 was considered mild, 15–30 moderate, and more than 30 severe OSAS. Sleep data were staged according to the system described by Rechtschaffen and Kales (1968).

All the subjects underwent a physical examination, including body mass index (kg/m^2), sex, AHI, age, neck circumference, and sleep parameters. We evaluated the healthy control groups in the study of Berdeli and Cam (2009), and 79 age-matched healthy Turkish volunteers (mean age 60.1 ± 10 years) without OSAS from the study group of Tuncer et al. (2006) were adopted as a control group for our study. Patients with sarcoidosis, chronic obstructive pulmonary disease, diabetes mellitus, liver cirrhosis, thyroid dysfunction, or renal failure were excluded. Patients who used ACE inhibitors, AT receptor blockers, and continuous positive airway pressure were also excluded from the study. All participants signed an informed consent form, and this study was approved by the local ethics committee of the Medical Faculty of Akdeniz University.

Genotyping for ACE I/D Polymorphism

Genomic DNA was extracted from 10 ml peripheral blood samples with K_3 -EDTA by a salting-out method (Miller et al. 1988). The D and I alleles were identified by polymerase chain reaction (PCR) performed in a final volume of 50 μl containing 10 pmol of each primer (Forward 5'-CTGGAGACCACTCCCATCCTTTCT-3' and Reverse 5'-GATGTGGCCATCACATTCGTCAGAT-3'), 20 mM dNTP, 1.5 mM MgCl_2 , 0.5 μg DNA, 5 μl $10 \times$ PCR buffer, and 1 U *Taq* DNA polymerase. The thermal cycling procedure consisted of initial denaturation at 95°C for 5 min, denaturation at 94°C for 1 min, annealing at 63°C for 1 min, and extension at 72°C for 2 min, repeated for 35 cycles (Jeng et al. 1998). The DNA products were visualized in 2% agarose gel stained with ethidium bromide. A fragment of 190 bp represented the D allele, and a fragment of 490 bp represented the presence of the I allele. To prevent mistyping of ID genotypes as DD genotypes, because of the selective amplification of the short fragment, each sample that had the DD genotype was reamplified with insertion-specific primers (Forward 5'-TGGGACCACAGCGCCCGCCACTAC-3' and Reverse 5'-TCGCCAGCCCTCCCATGCCCA TAA-3'), which recognizes the inserted DNA sequence in 25 ml of the reaction mixture, with 1 min at 94°C, followed by 30 cycles of 30 s at 94°C, 45 s at 67°C,

and 2 min at 72°C. The 335-bp product was observed in the presence of the I allele (Shanmugam et al. 1993).

Serum ACE Activity

Blood samples were collected, and serum was prepared, and stored at -80°C until analysis. Serum ACE activity was determined with an ACE colorimetric assay kit (KK-ACE, Böhlmann Laboratories AG, Switzerland). One unit of ACE activity was defined as the amount of enzyme required to release 1 μmol of hippuric acid/min/liter of serum at 37°C .

Statistical Analysis

The distribution of I/D polymorphisms of the *ACE* gene in OSAS patients and in the control groups was compared by a chi-square test. The association between the distribution of I and D alleles and OSAS severity subgroups, as well as case and control groups, was assessed by a chi-square test. The ACE plasma activity for three OSAS severity subgroups and three genotypic subgroups was assessed by Kruskal–Wallis variance analysis. A Mann–Whitney *U*-test was used to determine the multiple comparisons in these subgroups. Pearson correlation analysis was used to determine the possible relationship between the study variables and ACE plasma activity. All the statistical analyses were carried out with MedCalc software (version 10.2.0.0). *P*-values lower than 0.05 were considered statistically significant.

Results

Angiotensin-converting enzyme (ACE) genotype distribution was consistent with Hardy–Weinberg equilibrium in both the patient and control groups (Table 1). The I allele was observed more frequently in OSAS patients than in the control group [$P = 0.02$; OR = 1.68 (1.08–2.57)]. Carrying the I allele (genotype II or ID) increases the OSAS risk 2.41 times in the Turkish population [$P = 0.006$; OR = 2.41 (91.28–4.52)].

Table 1 Frequency of *ACE* I/D genotypes and alleles in OSAS patients and control subjects without OSAS

Group (<i>n</i>)	Genotype %*			Allele*	
	II	ID	DD	I	D
OSAS patients (97)	19.6	53.6**	26.8**	0.46	0.54
Control subjects (79)	15	38	47	0.34	0.66

* Difference between groups significant at $P < 0.05$. Allele frequency, OR = 1.68 (1.08–2.57)

** Significant at $P < 0.01$. Difference between ID or DD genotype, OR = 2.47 (1.26–4.84). Difference between carrying I allele or not, OR = 2.41 (1.28–4.52)

There were no significant differences in the mean value of neck circumference, body mass index, age, sex, mean duration apnea, and minimum SpO₂ between the I and D genotypes of OSAS patients ($P > 0.05$). Also, we found no statistically significant differences related to degree of severity of OSAS and ACE gene polymorphism (Table 2, $P = 0.831$).

Although ACE genotype distribution was determined in all of the patients, serum ACE activity was determined in only 73 patients. We found a statistically significant difference among the ACE genotype subgroups in terms of the mean ACE activity ($P = 0.043$), and the mean ACE activity was lower in the II genotype than in the DD genotype (Table 3, $P = 0.011$). When we compared ACE activity with the severity of OSAS ($P = 0.019$), the ACE activity was significantly lower in the severe OSAS group than in the mild OSAS group (Table 4, $P = 0.006$).

Our investigation of possible relationships between ACE activity and demographic and/or polysomnography variables found significant correlations for both minimum SpO₂ and AHI (Table 5). The level of minimum SpO₂ was high and AHI tended to be low for patients who had higher ACE activity (minimum SpO₂: $r = 0.30$, $P = 0.01$; AHI: $r = -0.31$, $P = 0.007$).

Discussion

Our study showed that the genotypic distribution of ACE gene polymorphism was significantly different in the OSAS patients, and the ACE II genotype increased

Table 2 Relationship between ACE gene polymorphism and OSAS severity

OSAS (n)	Genotype n (%)		
	II	ID	DD
Mild (44)	10 (22.7)	21 (47.7)	13 (29.5)
Moderate (22)	3 (13.6)	13 (59.1)	6 (27.3)
Severe (31)	6 (19.4)	18 (58.1)	7 (22.6)

Table 3 ACE activity level in three ACE genotypes of OSAS patients

Genotype	ACE activity, mean IU/l ± SD
II	33.53 ± 9.31*
ID	38.25 ± 10.80
DD	41.81 ± 9.62

* Difference between II and DD genotypes significant at $P < 0.05$

Table 4 Relationship between mean ACE activity and OSAS severity

OSAS (n)	ACE activity, mean IU/l ± SD
Mild (36)	41.41 ± 10.16*
Moderate (18)	36.96 ± 10.44
Severe (19)	33.31 ± 9.34

* Difference between mild and severe OSAS patients significant at $P < 0.05$

Table 5 Correlation analysis of ACE plasma activity and patient variables

Variable	OSAS study population ^a Mean \pm SD	ACE plasma activity, r (P)
Age in years	51.27 \pm 9.97	0.80 (0.50)
Body mass index in kg/m ²	30.58 \pm 5.79	-0.03 (0.80)
Neck circumference in cm	41.82 \pm 3.31	-0.17 (0.16)
Apnea-hypopnea index	24.24 \pm 18.34	-0.31 (0.01)*
Minimum SpO ₂ percentage	76.52 \pm 11.40	0.30 (0.01)*
Epworth score	11.34 \pm 4.7	-0.11 (0.35)
Mean duration apnea in seconds	21.55 \pm 5.89	0.02 (0.88)

^a Of the 97 unrelated Turkish patients, 9 (9.3%) were women, 88 (90.7%) were men

* Statistically significant correlation at 0.05 level

OSAS risk 2.4 times in the Turkish population (Table 1). Even though the II genotype was shown to be a high risk factor for OSAS in a Chinese population, no association was found between *ACE* polymorphism and OSAS in Spanish and Polish populations (Barcelo et al. 2001; Rubinsztajn et al. 2004; Xiao et al. 1999).

Barley et al. (1994) studied *ACE* gene polymorphism in different populations, including white Europeans, black Nigerians, Samoan Polynesians, and Yanomami Indians. They found that I allele frequency was higher in the latter two populations, and they concluded that *ACE* gene polymorphism was associated with ethnic origin. In the control group adopted for our study, the frequency of the I allele was found to be 0.34 (Tuncer et al. 2006).

A correlation between homozygote gene deletion (DD genotype) and high ACE activity has been reported in many studies (Seckin et al. 2006; Ozen et al. 1997). Accordingly, we found the highest ACE activity in the DD genotype and the lowest in the II genotype (Table 3). Barcelo et al. (2001) compared ACE activity in patients with OSAS and control subjects and showed that ACE activity is increased in patients with OSAS regardless of the presence or absence of arterial hypertension. We did not evaluate ACE activity in control subjects, which might be a limitation of our study. When we compared ACE activity with the severity of OSAS, the ACE activity was significantly lower in the severe OSAS group than in the mild OSAS group, although genotype differences do not exist between the two groups (Table 4). We found that the level of minimum SpO₂ was significantly higher and AHI was significantly lower in patients with high ACE activity (Table 5).

Several possible mechanisms may contribute to our results. First, different *ACE* genotype distributions can result in different ACE activity levels. We found no significant correlations, however, between OSAS severity and *ACE* gene polymorphism. Second, morbid obesity can negatively influence ACE activity. Previous results have shown the involvement of ACE in adipocyte growth, function, and inhibition of adipocyte differentiation by ACE-processed angiotensin II, but Bell et al. (2007) reported no correlation between ACE gene variations and development of severe obesity. To our knowledge, no study has defined ACE serum activity in

obese individuals. Third, decreased serum activity of ACE depends on intermittent hypoxemia related to high frequency apnea/hypopnea episodes in severe OSAS. In our study, we found that SpO₂ was higher and AHI was lower in patients with high ACE activity (Table 5). The ACE enzyme activity can change in pulmonary diseases due to vascular endothelial damage. Mean serum ACE activity has been reported for a wide variety of chronic airway diseases such as asthma, chronic bronchitis, emphysema, and cystic fibrosis. Rohatgi (1982) reported lower ACE activity in OSAS patients than in healthy controls. Ashutosh and Keighley (1976) and Kanazawa et al. (2000), however, reported higher ACE activity in chronic hypoxia. Further studies are needed to clarify these discrepancies.

In conclusion, our data demonstrate that the frequency of the II genotype of the ACE gene is significantly higher in OSAS patients than in healthy subjects, and the II genotype increases the risk of development of OSAS by 2.4 fold in this Turkish population.

Acknowledgments This study was supported by Research Foundation Management of Akdeniz University (2004.04.103.10).

References

- Ashutosh K, Keighley JF (1976) Diagnostic value of serum angiotensin converting enzyme activity in lung diseases. *Thorax* 31:552–557
- Barcelo A, Elorza MA, Barbe F, Santos C, Mayoralas LR, Agusti AG (2001) Angiotensin converting enzyme in patients with sleep apnoea syndrome: plasma activity and gene polymorphisms. *Eur Respir J* 17:728–732
- Barcelo A, Llompert E, Barbe F, Morla M, Vila M, Agustí AG (2002) Plasminogen activator inhibitor-I (PAI-I) polymorphisms in patients with obstructive sleep apnoea. *Respir Med* 96:193–196
- Barley J, Blackwood A, Carter ND, Crews DE, Cruickshank JK, Jeffery S, Ogunlesi AO, Sagnella GA (1994) Angiotensin converting enzyme insertion/deletion polymorphism: association with ethnic origin. *J Hypertens* 12:955–957
- Bayazit YA, Erdal ME, Yilmaz M, Ciftci TU, Soylemez F, Gokdoğan T, Kokturk O, Kemaloglu YK, Koybasioglu A (2006a) Insulin receptor substrate gene polymorphism is associated with obstructive sleep apnea syndrome in men. *Laryngoscope* 116:1962–1965
- Bayazit YA, Yilmaz M, Ciftci T, Erdal E, Kokturk O, Gokdogan T, Kemaloglu YK, Inal E (2006b) Association of the -1438G/A polymorphism of the 5-HT2A receptor gene with obstructive sleep apnea syndrome. *ORL J Otorhinolaryngol Relat Spec* 68:123–128
- Bayazit YA, Yilmaz M, Kokturk O, Erdal ME, Ciftci T, Gokdogan T, Kemaloglu Y, Ileri F (2007) Association of GABA(B)R1 receptor gene polymorphism with obstructive sleep apnea syndrome. *ORL J Otorhinolaryngol Relat Spec* 69:190–197
- Bell CG, Meyre D, Petretto E, Levy-Marchal C, Hercberg S, Charles MA, Boyle C, Weill J, Tauber M, Mein CA, Aitman TJ, Froguel P, Walley AJ (2007) No contribution of angiotensin-converting enzyme (ACE) gene variants to severe obesity: a model for comprehensive case/control and quantitative cladistic analysis of ACE in human diseases. *Eur J Hum Genet* 15:320–327
- Berdeli A, Cam FS (2009) Prevalence of the angiotensin I converting enzyme gene insertion/deletion polymorphism in a healthy population. *Biochem Genet* 47:412–420
- Hanaoka M, Yu X, Urushihata K, Ota M, Fujimoto K, Kubo K (2008) Leptin and leptin receptor gene polymorphisms in obstructive sleep apnea syndrome. *Chest* 133:79–85
- Jeng JR, Harn HJ, Yueh KC, Jeng CY, Shieh SM (1998) Plasminogen activator inhibitor-1 and angiotensin I converting enzyme gene polymorphism in patients with hypertension. *AJH* 11:235–239
- Kanazawa H, Okamoto T, Hirata K, Yoskikwa J (2000) Deletion polymorphisms in the angiotensin converting enzyme gene are associated with pulmonary hypertension evoked by exercise challenge in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 162:1235–1238

- Kaparianos A, Sampsonas F, Karkoulis K, Spiropoulos K (2006) Obstructive sleep apnoea syndrome and genes. *Neth J Med* 64:280–289
- Miller SA, Dykes DD, Polesky MF (1988) A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic acids Res* 16:1215
- Ozen S, Alikasifoglu M, Tuncbilek E, Bakaloglu A, Besbas N, Aran B, Saatci U (1997) Polymorphisms in angiotensin converting enzyme gene and reflux nephropathy: a genetic predisposition to scar formation. *Nephrol Dial Transplant* 12:2031–2033
- Palmer LJ, Redline S (2003) Genomic approaches to understanding obstructive sleep apnea. *Respir Physiol Neurobiol* 135:187–205
- Palmer LJ, Buxbaum SG, Larkin EK, Patel SR, Elston RC, Tishler PV, Redline S (2004) Whole genome scan for obstructive sleep apnea and obesity in African-American families. *Am J Respir Crit Care Med* 169:1314–1321
- Pierola J, Barcelo A, de la Pena M, Barbé F, Soriano JB, Sánchez Armengol A, Martínez C, Agustí A (2007) Beta3-Adrenergic receptor Trp64Arg polymorphism and increased body mass index in sleep apnoea. *Eur Respir J* 30:743–747
- Rechtschaffen A, Kales A (eds) (1968) A Manual of standardized terminology: techniques and scoring system for sleep stages of human subjects. UCLA Brain Information Service/Brain Research Institute, Los Angeles, CA
- Rigat B, Hubert C, Alhenc-Gelas F, Cambien F, Corvol P, Soubrier F (1990) An insertion/deletion polymorphism in the angiotensin I-converting enzyme gene accounting for half the variance of serum enzyme levels. *J Clin Invest* 86:1343–1346
- Riha RL, Brander P, Vennelle M, McArdle N, Kerr SM, Anderson NH, Douglas NJ (2005) Tumour necrosis factor- α (–308) gene polymorphism in obstructive sleep apnoea–hypopnoea syndrome. *Eur Respir J* 26:673–678
- Rohatgi PK (1982) Serum angiotensin converting enzyme in pulmonary disease. *Lung* 160:287–301
- Rubinsztajn R, Kumor M, Byskiniewicz K, Chazan R (2004) Angiotensin-converting enzyme gene polymorphism in patients with obstructive sleep apnea. *Pol Arch Med Wewn* 112:817–822
- Seckin D, Ilhan N, Ilhan N, Ozbay Y (2006) The relationship between ACE insertion/deletion polymorphism and coronary artery disease with or without myocardial infarction. *Clin Biochem* 39:50–54
- Shanmugam V, Sell KW, Saha BK (1993) Mistyping ACE heterozygotes. *PCR Methods Appl* 3:120–121
- Tafti M, Dauvilliers Y, Overeem S (2007) Narcolepsy and familial advanced sleep-phase syndrome: molecular genetics of sleep disorders. *Curr Opin Genet Dev* 17:222–227
- Taheri S (2004) The genetics of sleep disorders. *Minerva Med* 95:203–212
- Taheri S, Mignot E (2002) The genetics of sleep disorders. *Lancet Neurol* 1:242–250
- Tuncer N, Tuglular S, Kilic G, Sazci A, Us O, Kara I (2006) Evaluation of the angiotensin-converting enzyme insertion/deletion polymorphism and the risk of ischaemic stroke. *J Clin Neurosci* 13: 224–227
- Ursavas A, Karadag M, Ozarda Ilcol Y, Burgazlioglu B, Ercan I, Gozü RO (2007) Relationship between serum Substance P levels and daytime sleepiness in obstructive sleep apnea syndrome. *Chest* 131:1400–1405
- Xiao Y, Huang X, Qiu C, Zhu X, Liu Y (1999) Angiotensin I-converting enzyme gene polymorphism in Chinese patients with obstructive sleep apnea syndrome. *Chin Med J* 112:701–704
- Zhang J, Zhao B, Gesongluobu SunY, Wu Y, Pei W, Ye J, Hui R, Liu L (2000) Angiotensin-converting enzyme gene insertion/deletion (I/D) polymorphism in hypertensive patients with different degrees of obstructive sleep apnea. *Hypertens Res* 23:407–411