

T.C.  
HACETTEPE ÜNİVERSİTESİ  
SAĞLIK BİLİMLERİ ENSTİTÜSÜ

KOMPLEKS HASTALIKLARLA İLİŞKİLİ GENETİK VARYASYONUN  
ZAMAN-SERİSİ GENOM DİZİLEME VERİLERİNDE POPÜLASYON  
GENETİĞİ YAKLAŞIMLARI İLE İNCELENMESİ

İdil TAÇ

Biyoinformatik Programı  
YÜKSEK LİSANS TEZİ

ANKARA  
2023



T.C.  
HACETTEPE ÜNİVERSİTESİ  
SAĞLIK BİLİMLERİ ENSTİTÜSÜ

KOMPLEKS HASTALIKLARLA İLİŞKİLİ GENETİK VARYASYONUN  
ZAMAN-SERİSİ GENOM DİZİLEME VERİLERİNDE POPÜLASYON  
GENETİĞİ YAKLAŞIMLARI İLE İNCELENMESİ

İdil TAÇ

Biyoinformatik Programı

YÜKSEK LİSANS TEZİ

TEZ DANIŞMANI

Dr. Öğr. Üyesi İdil YET

İKİNCİ DANIŞMAN

Dr. Öğr. Üyesi Gülşah Merve KILINÇ

ANKARA

2023

**HACETTEPE ÜNİVERSİTESİ**  
**SAĞLIK BİLİMLERİ ENSTİTÜSÜ**  
**KOMPLEKS HASTALIKLARLA İLİŞKİLİ GENETİK VARYASYONUN ZAMAN-SERİSİ**  
**GENOM DİZİLEME VERİLERİNDE POPÜLASYON GENETİĞİ YAKLAŞIMLARI İLE**  
**İNCELENMESİ**

**Öğrenci: İdil TAÇ**

**Danışman: Dr. Öğr. Üyesi İdil YET**

**İkinci Danışman: Dr. Öğr. Üyesi Gülşah Merve KILINÇ**

Bu tez çalışması 01/06/2023 tarihinde jürimiz tarafından “Biyoinformatik Programı” nda yüksek lisans tezi olarak kabul edilmiştir.

**Jüri Başkanı:** **Prof.Dr. Mehmet SOMEL**  
(*Ortadoğu Teknik Üniversitesi*)

**Tez Danışmanı:** **Dr. Öğr. Üyesi İdil YET**  
(*Hacettepe Üniversitesi*)

**Üye:** **Dr. Öğr. Üyesi Arda ÇETİNKAYA**  
(*Hacettepe Üniversitesi*)

Bu tez Hacettepe Üniversitesi Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğinin ilgili maddeleri uyarınca yukarıdaki jüri tarafından uygun bulunmuştur.

Prof. Dr. Müge YEMİŞÇİ ÖZKAN

**Enstitü Müdürü**

## YAYIMLAMA VE FİKRİ MÜLKİYET HAKLARI BEYANI

Enstitü tarafından onaylanan lisansüstü tezimin/raporumun tamamını veya herhangi bir kısmını, basılı (kağıt) ve elektronik formatta arşivleme ve aşağıda verilen koşullarla kullanıma açma iznini Hacettepe Üniversitesine verdiğimi bildiririm. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet haklarım bende kalacak, tezimin tamamının ya da bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanım hakları bana ait olacaktır.

Tezin kendi orijinal çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanılması zorunlu metinlerin yazılı izin alınarak kullandığımı ve istenildiğinde suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

Yükseköğretim Kurulu tarafından yayınlanan “Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge” kapsamında tezimin aşağıda belirtilen koşullar haricince YÖK Ulusal Tez Merkezi / H.Ü. Kütüphaneleri Açık Erişim Sisteminde erişime açılır.

- Enstitü / Fakülte yönetim kurulu kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren 2 yıl ertelenmiştir. <sup>(1)</sup>
- Enstitü / Faksülte yönetim kurulunun gerekçeli kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren ... ay ertelenmiştir. <sup>(2)</sup>
- Tezimle ilgili gizlilik kararı verilmiştir.

...../...../.....

İdil TAÇ

1“Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge”

(1) Madde 6. 1. Lisansüstü tezle ilgili patent başvurusu yapılması veya patent alma sürecinin devam etmesi durumunda, tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulu iki yıl süre ile tezin erişime açılmasının ertelenmesine karar verebilir.

(2) Madde 6. 2. Yeni teknik, materyal ve metotların kullanıldığı, henüz makaleye dönüşmemiş veya patent gibi yöntemlerle korunmamış ve internetten paylaşılması durumunda 3. şahıslara veya kurumlara haksız kazanç imkanı oluşturabilecek bilgi ve bulguları içeren tezler hakkında tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulunun gerekçeli kararı ile altı ayı aşmamak üzere tezin erişime açılması engellenebilir.

(3) Madde 7. 1. Ulusal çıkarları veya güvenliği ilgilendiren, emniyet, istihbarat, savunma ve güvenlik, sağlık vb. konulara ilişkin lisansüstü tezlerle ilgili gizlilik kararı, tezin yapıldığı kurum tarafından verilir \*. Kurum ve kuruluşlarla yapılan işbirliği protokolü çerçevesinde hazırlanan lisansüstü tezlere ilişkin gizlilik kararı ise, ilgili kurum ve kuruluşun önerisi ile enstitü veya fakültenin uygun görüşü üzerine üniversite yönetim kurulu tarafından verilir. Gizlilik kararı verilen tezler Yükseköğretim Kuruluna bildirilir. Madde 7.2. Gizlilik kararı verilen tezler gizlilik süresince enstitü veya fakülte tarafından gizlilik kuralları çerçevesinde muhafaza edilir, gizlilik kararının kaldırılması halinde Tez Otomasyon Sistemine yüklenir

\* Tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulu tarafından karar verilir.

## ETİK BEYAN

Bu alıřmadaki bütn bilgi ve belgeleri akademik kurallar erevesinde elde ettiđimi, grsel, iřitsel ve yazılı tm bilgi ve sonuları bilimsel ahlak kurallarına uygun olarak sunduđumu, kullandıđım verilerde herhangi bir tahrifat yapmadıđımı, yararlandıđım kaynaklara bilimsel normlara uygun olarak atıfta bulunduđumu, tezimin kaynak gsterilen durumlar dıřında zgn olduđunu, Tez Danıřmanı Dr. đr. yesi İdil YET ve ikinci Danıřman Dr. đr. yesi Glřah Merve KILIN'ın danıřmanlıđında tarafımdan retildiđini ve Hacettepe niversitesi Sađlık Bilimleri Enstits Tez Yazım Ynergesine gre yazıldıđını beyan ederim.

**İdil TA**

## TEŞEKKÜR

Başta tez çalışmasının başından sonuna tüm aşamalarında kritik değerlendirmeleri ve yönlendirmeleriyle desteklerini sunan, ayrıca yüksek lisans sürem boyunca hem akademik kariyerim hem özel hayatımda desteklerini hiç esirgemeyen çok değerli tez danışmanlarım Dr. Öğr. Üyesi Gülşah Merve KILINÇ ve Dr. Öğr. Üyesi İdil YET'e, tez kapsamında yayınlanmış genomik verilerin işlenmesi süresince kolaylaştırıcı katkıları ve verdikleri tüm destek sebebiyle tüm CompEvo ekibine, ders ve tez süresince tüm yardımlarından dolayı Hacettepe Biyoinformatik bölümü öğrencilerine ve Menekşe Tarla'ya minnetle teşekkür ederim.

Eğitim hayatım boyunca aldığım her kararda yanımda olup beni destekleyen, hayatımın hiçbir anında beni asla sevgisiz bırakmayan canım ailem, annem Ayfer, babam Ata, kardeşim İmge'ye en içten teşekkürlerimi sunarım.

Yüksek lisans tez çalışmalarım sırasında sağladığı burs nedeniyle TÜBİTAK BİDEB 2210-A ve TÜBİTAK ARDEB 3501 (Yürütücü: Dr. Öğr. Üyesi Gülşah Merve KILINÇ, 121Z025) projeleri ile tez çalışmalarında deneysel işlemlerin gerçekleştirilmesi için kaynak sunması nedeniyle Hacettepe Üniversitesi BAP Koordinasyon Birimi tarafından Yürütücü: Dr. Öğr. Üyesi İdil YET, TYL-2022-19809 numaralı lisansüstü tez destek projesi kapsamındaki destekleri nedeniyle destekleyen kurumlara teşekkür ederim.

## ÖZET

**Taç, İ., Kompleks Hastalıklarla İlişkili Genetik Varyasyonun Zaman-Serisi Genom Dizileme Verilerinde Popülasyon Genetiği Yaklaşımları ile İncelenmesi, Hacettepe Üniversitesi Sağlık Bilimleri Enstitüsü Biyoinformatik Programı Yüksek Lisans Tezi, Ankara, 2023.** Genetik hastalıklar evrimsel sürecin bir ürünü olarak ortaya çıkar. Anatomik olarak modern insanın son 200.000 yıllık demografik tarihi, bugün yaygın görülen hastalıklarla ilişkili genetik yapıyı şekillendirmiştir. Bu süreçte genomu etkileyen evrimsel güçlerin dinamiklerini zaman-serisi genomik veriler ışığında anlamak mümkündür. Antik DNA zaman-serisi genomik bilgi sağlama potansiyeline sahiptir. Bu şekilde üretilen genomik veriyle günümüzde yaygın görülen hastalıklarla ilgili genetik yapıyı etkileyen evrimsel güçlerin çalışılması mümkündür. Bu tez kapsamında, Obezite ve Tip 2 diyabet olmak üzere tüm dünyada yüksek prevalansa sahip iki kompleks hastalıkla ilişkili genetik yapının son 10.000 yılı kapsayan değişim dinamikleri araştırılmıştır. Anadolu'dan yayınlanmış 319 antik ve 16 modern insan genom dizileme verisi Obezite ile ilişkili 18, Tip 2 diyabet ile ilişkili 415 ve 1000 nötral varyant için analiz edilmiştir. Tüm genetik varyantlar için günümüz, Neolitik (G.Ö. 10326-6700) ve Neolitik sonrası (G.Ö. 7720-50) olmak üzere üç farklı dönem için maksimum olabilirlik temelli bir yaklaşımla alel frekans yörüngeleri belirlenmiştir.  $F_{ST}$  ile hastalıklarla ilişkili varyantlar için antik ve modern Anadolu popülasyonları arasındaki genetik mesafe hesaplanmış ve elde edilen sonuçlar evrimsel olarak nötral bölgelerde yer alan varyantlarla karşılaştırılmıştır. Karşılaştırılan ikili dönemler arasında alel frekanslarının ve  $F_{ST}$  dağılımlarının nötral varyantlardan önemli ölçüde farklılık gösterdiği dönemler; Neolitik ile Günümüz, Neolitik ile Neolitik Sonrası ( $p < 0.05$ , ANOVA)'dır. Pozitif seçilimi test etmek için popülasyon dal istatistiği (PDI) hesaplanmıştır. PDI sonucunda seçim sinyali yakalanan varyantlardan Neolitik'ten günümüze alel frekansı artan Obezite ile ilişkili bir, Tip 2 diyabet ile ilişkili dokuz varyant; Neolitik ve Neolitik sonrası dönemde alel frekansı artan Tip 2 diyabet ile ilişkili bir varyant bulunmuştur. Bu varyantlardan rs35011184 ile ilişkili gen *TCF7L2*, seçilime dair izler taşıyan Tip 2 diyabetle ilişkili genlerden biridir. Anadolu antik ve günümüz popülasyonları ileri zamanlı genetik benzetim aracı kullanılarak nötral senaryo altında modellenmiştir. Üretilen sentetik genom verisi ile hesaplanan alel frekans dağılımları gerçek veriden elde edilen frekanslar ile karşılaştırılarak nötral model ile farklılık test edilmiştir. Alel frekanslarının Neolitik ile günümüz arasındaki artışının nötral senaryo altında gözlenen artıştan daha yüksek olduğu bulunmuştur. Son 10.000 yılda Obezite ve Tip 2 diyabetle ilişkili alellerin bir kısmının frekansının arttığı, bu değişimin nötral bölgelere göre önemli ölçüde farklılık göstermesi ile bu değişimin poligenik seçilimin izlerini taşıyabileceği gösterilmiştir.

**Anahtar Kelimeler:** Antik DNA, Obezite, Popülasyon Genetiği, Diyabet

Bu tez TÜBİTAK BİDEB 2210-A, TÜBİTAK ARDEB 3501 (121Z025) ve Hacettepe Üniversitesi BAP Koordinasyon Birimi (TYL-2022-19809) tarafından desteklenmiştir.



**ABSTRACT**

**Taç, İ., Investigation of Genetic Variation Associated with Complex Diseases in Time-Series Genome Sequencing Data with Population Genetics Approaches, Hacettepe University Graduate School of Health Sciences Department of Bioinformatics Master's Thesis, Ankara, 2023.** Genetic diseases are the products of evolutionary process. The demographic history of anatomically modern humans over the last 200.000 years has shaped the genetic structure associated with common diseases. It is possible to understand the dynamics of evolutionary forces affecting the genome in light of time-series genomic data. Ancient DNA provides time-series genomic information. With the genomic data generated in this way, it is possible to study the evolutionary forces affecting the genetic structure of common diseases. In this thesis, we investigated the temporal dynamics of genetic structure behind Obesity and Type 2 diabetes, two complex diseases with high prevalence worldwide over the last 10.000 years. A total of 319 ancient and 16 modern human genome sequencing data from Anatolia were analyzed for 18 Obesity-related, 415 Type 2 diabetes-related and 1000 neutral variants. Allele frequency trajectories were determined for all genetic variants using a maximum likelihood-based approach for three different periods: Neolithic (10326-6700 BC), post-Neolithic (7720-50 BC), and present-day. The genetic distance between ancient and modern Anatolian populations for variants associated with disease was calculated using  $F_{ST}$  and the results were compared with variants located in evolutionarily neutral regions. The periods in which allele frequencies and  $F_{ST}$  distributions differed significantly from neutral variants among the compared binary periods were Neolithic and Modern, Neolithic and Post-Neolithic ( $p < 0.05$ , ANOVA). To test for positive selection, the population branch statistic (PBS) was calculated. As a result of the PBS, among the variants with selection signals, one variant was associated with Obesity and nine variants were associated with Type 2 diabetes with increasing allele frequencies from Neolithic to the present, and one variant was associated with Type 2 diabetes with increasing allele frequencies in Neolithic and post-Neolithic periods were discovered. Among these variants, *TCF7L2*, gene associated with rs35011184, is one of the genes associated with Type 2 diabetes with traces of selection. Anatolian ancient and modern populations were modelled under a neutral scenario using a forward-time genetic simulation tool. The allele frequency distributions calculated with the synthetic genome data were compared with the frequencies obtained from real data and the difference with the neutral model was tested. It was found that the increase in allele frequencies between the Neolithic and present-day was higher than the increase observed under the neutral scenario. It has been shown that the frequency of some of the alleles associated with Obesity and Type 2 diabetes has increased over the last 10,000 years, and that this change may imply for polygenic selection.

**Keywords:** Ancient DNA, Population Genetics, Obesity, Diabetes Mellitus

This thesis has been supported by TÜBİTAK BİDEB 2210-A, TÜBİTAK ARDEB 3501 (121Z025) and Hacettepe University Research Projects Unit (TYL-2022-19809).

**İÇİNDEKİLER**

ONAY SAYFASI	iii
YAYIMLAMA VE FİKRİ MÜLKİYET HAKLARI BEYANI	iv
ETİK BEYANI	v
TEŞEKKÜR	vi
ÖZET	vii
ABSTRACT	viii
İÇİNDEKİLER	ix
SİMGELER VE KISALTMALAR	xii
ŞEKİLLER	xiv
TABLolar	xix
<b>1 . GİRİŞ</b>	<b>1</b>
<b>2 . GENEL BİLGİLER</b>	<b>3</b>
2.1 . İnsan Genomu ve Genetik Varyasyon	3
2.2 . Kompleks Hastalıklar	4
2.3 . Obezite ve Tip 2 Diyabet	6
2.4 . Hastalıkların Genetik Arka Planını Etkileyen Evrimsel Süreçler	9
2.5 . Antik DNA ve Paleogenetik	12
2.6 . Kompleks Fenotipleri Antik DNA ile Araştırmak	13
2.7 . Anadolu ve Paleogenetik Araştırmalar	14
<b>3 . GEREKÇE VE YÖNTEM</b>	<b>16</b>

3.1 . İlgilenilen Genomik Bölgelerin Belirlenmesi	16
3.1.1 . Kompleks Fenotiplerle İlişkili Varyantların Belirlenmesi	16
3.1.2 . Nötral Varyantların Belirlenmesi	17
3.2 . Zaman Serisi Genomik Veri Setleri	20
3.2.1 . Antik Genomik Veri Seti için Örnek Seçimi	20
3.2.2 . Antik Genom Verilerinin Hizalanması	22
3.2.3 . Modern Genom Veri Setlerinin İşlenmesi	22
3.3 . Veri Setlerinin Hazırlanması	22
3.3.1 . Alel Frekans Hesaplamaları için Veri Seti Hazırlanması	22
3.3.2 . Biyoinformatik Yazılımların Karşılaştırması için Veri Seti	23
3.4 . Alel Frekans Hesaplamalarının Yapılması	25
3.5 . $F_{ST}$ Hesaplanması	25
3.6 . Popülasyon Dal İstatistiği Skorlarının Hesaplanması	28
3.7 . İleri Zamanlı Genetik Benzetim	29
<b>4 . BULGULAR</b>	<b>32</b>
4.1 . Fenotip ve Nötral Bölgelerle İlişkili Hedeflenen TNP Sayıları ve Genotiplenme Oranları	32
4.2 . Alel Frekans Yörüngelerinin Belirlenmesi	33
4.3 . Seçilim Analizleri Sonuçları	45
4.4 . İleri Zamanlı Genetik Benzetim ile Anadolu Popülasyonlarında Nötral Senaryo Altında Alel Frekans Değişimlerinin Modellenmesi	55
4.5 . Farklı Biyoinformatik Yaklaşımlarla Hazırlanan Veri Setlerinin Test Edilmesi	62
<b>5 . TARTIŞMA</b>	<b>63</b>
<b>6 . SONUÇ VE ÖNERİLER</b>	<b>68</b>
<b>7 . KAYNAKLAR</b>	<b>70</b>

## **8 . EKLER**

**EK-1:** Tez çalışmasının etik kurul izni

**EK-2:** Tez çalışması orijinallik raporu

**EK-3:** Nötral senaryo altında Anadolu modellemesi için SLiM programında yürütülen komutlar.

## **9 . ÖZGEÇMİŞ**

## SİMGELER VE KISALTMALAR

<b>aDNA</b>	Antik DNA
<b>ANGSD</b>	Yeni Nesil Dizileme Verilerinin Analizi ( <i>Analysis of Next Generation Sequencing Data</i> )
<b>ANOVA</b>	Varyans Analizi ( <i>Analysis of Variance</i> )
<b>BAM</b>	İkili hizalama /Harita Dosyası ( <i>Binary Alignment Map</i> )
<b>bç</b>	Baz Çifti ( <i>Base Pair</i> )
<b>BCE</b>	Milattan Önce ( <i>Before Common Era</i> )
<b>BED</b>	( <i>Browser Extensible Data</i> )
<b>csv</b>	Virgülle Ayrılmış Değer ( <i>Comma Separated Value</i> )
<b>DNA</b>	DeoksiriboNükleik Asit
<b>F<sub>ST</sub></b>	Fiksasyon İndeksi ( <i>Fixation Index</i> )
<b>G.Ö.</b>	Günümüzden Önce
<b>GA</b>	Güven Aralığı
<b>GATK</b>	Genom Analiz Araç Kiti ( <i>Genome Analysis Toolkit</i> )
<b>GWAS</b>	Genom Boyu İlişkilendirme Çalışmaları ( <i>Genome Wide Association Study</i> )
<b>HGP</b>	İnsan Genom Projesi ( <i>Human Genome Project</i> )
<b>IBD</b>	Mesafeye Göre İzolasyon ( <i>Isolation by Distance</i> )
<b>İndel</b>	İnsersiyon/ Delesyon
<b>Kb</b>	Kilobaz
<b>lncRNA</b>	Uzun kodlanmayan RNA ( <i>Long non-coding RNA</i> )
<b>MAF</b>	Minör Alel Frekansı
<b>Mb</b>	Mega Baz
<b>miRNA</b>	Mikro RNA
<b>NCBI</b>	Ulusal Biyoteknoloji Bilgi Merkezi ( <i>National Center for Biotechnology Information</i> )
<b>NGS</b>	Yeni Nesil Dizileme ( <i>Next Generation Sequencing</i> )
<b>NRE</b>	Neutral Region Explorer
<b>OECD</b>	İktisadi İşbirliği ve Gelişme Teşkilatı ( <i>Organisation for Economic Co-operation and Development</i> )
<b>OMIM</b>	Online Mendelian Inheritance in Man
<b>PBS</b>	Popülasyon Dal İstatistiği ( <i>Population Branch Statistic</i> )
<b>PCR</b>	Polimeraz Zincir Reaksiyonu ( <i>Polymerase Chain Reaction</i> )
<b>RAF</b>	Risk Alel Frekansı
<b>snRNA</b>	Küçük Nükleer RNA ( <i>Small Nuclear RNA</i> )
<b>TSV</b>	Sekme ayrılmış değerler ( <i>Tab Separated Values</i> )
<b>TNP</b>	Tek Nükleotid Polimorfizmi ( <i>Single Nucleotide Polymorphism</i> )

<b>TÜBİTAK</b>	Türkiye Bilimsel ve Teknolojik Araştırma Kurumu
<b>TÜİK</b>	Türkiye İstatistik Kurumu
<b>VCF</b>	Varyant Çağrı Formatı ( <i>Variant Call Format</i> )
<b>WGS</b>	Tüm Genom Dizileme ( <i>Whole Genome Sequencing</i> )

## ŞEKİLLER

Şekil	Sayfa
2.1. Hastalıklar için genetik modeller.	5
2.2. Bolluk-kıtlık/egzersiz-dinlenme döngülerinde vücutta meydana gelen reaksiyon.	8
3.1. Veri setine dahil edilen yayınlanmış Anadolu antik genomik verilerinin veri üretilme yöntemi ve okuma derinliklerine (X) göre dağılımının kutu grafiği ile gösterilmesi.	20
3.2. Veri setine dahil edilen Anadolu yayınlanmış antik ve günümüz genom örneklerinin, A panelinde lokasyonu ve B panelinde karbon 14 tarihlendirmeleri ve ortalama okuma derinliğine göre dağılımı gösterilmiştir. B panelinde y eksenini log (10) tabanına göre ölçeklendirilmiştir.	21
3.3. Farklı varyant çağırma yaklaşımlarının test edilmesi için uygulanan iş akışı.	24
3.4. Karşılaştırmalı analizde kullanılan bireylerin A) lokasyonu, B) yaş ve ortalama okuma derinliklerinin dağılımı.	24
3.5. $F_{ST}$ tahminlerinden popülasyon dal istatistiği skorlarının hesaplanması.	28
3.6. İleri zamanlı genetik benzetim analizinde kullanılan demografik modelin şemasal gösterimi.	30
4.1. Seçilen 18 Obezite, 415 Tip 2 diyabet (T2D) ve 1000 nötral TNP'nin kromozomlar üzerinde dağılımının gösterimi.	32
4.2. Obezite (n=18, mavi) ve Tip 2 diyabet (n= 404, kırmızı) ile ilişkili varyantın Anadolu Neolitik, Anadolu Neolitik sonrası ve günümüz Anadolu'daki risk alel frekanslarının dağılımı. Arka planda gri renk ile gösterilen çizgiler 1000 adet nötral varyantın minör alel frekansının dağılımını göstermektedir.	34
4.3. Anadolu Neolitikten günümüze frekansı düzenli olarak artış ve azalış gösteren Obezite ve Tip 2 diyabet ile ilişkili varyantlar. A panelinde risk alel frekans değişimlerinin çizgi grafiği ile gösterilmiştir. B panelinde GWAS Catalog'da anlamlı ilişkide raporlanan (p-değeri $<5 \times 10^{-8}$ ) fenotip bilgilerinin artan ve azalana göre sayıca dağılımının listelenmesi.	35
4.4. Risk alel frekansı Neolitikten günümüze düzenli artış ve azalış gösteren Obezite ve Tip 2 diyabetle ilişkili TNP'lerin ilişkili olduğu genlerin gen ontolojisi analizi. (A) Hücresel Bileşenler (B) Sinyal Yolakları (C) Biyolojik Süreçler (D) Moleküler Fonksiyon.	37

- 4.5.** Anadolu antik ve günümüz dönemlerinden seçilen bireylerin taşıdığı Obezite ve Tip 2 diyabet ile ilişkili toplam risk alel sayılarının kutu grafiği ile gösterimi. Y eksenini  $\log(10)$  tabanına göre ölçeklendirilmiştir. **38**
- 4.6.** Neolitikten Neolitik sonrası döneme risk alel frekans artışı (sol panel) ve azalışı (sağ panel) gözlenen Obezite (mavi) ve Tip 2 diyabet (kırmızı) varyantlarının nötral varyantların alel frekansları (sarı) ile karşılaştırılması. Obezite ve Tip 2 diyabet ile nötral alel frekanslarının ortalamaları Mann-Whitney U testi ile karşılaştırılmış ve Bonferroni düzeltmesi sonucu p-değeri anlamlılıklarına göre anlamlı sonuçlar (p-değeri  $< 0,05$ ) kutu grafikler üzerinde gösterilmiştir. Obezite, Tip 2 diyabet ve nötral alel frekanslarının dönemler arası değişimi ise Wilcoxon testi ile karşılaştırılmış Bonferroni düzeltmesi sonucu p-değeri anlamlılıklarına göre anlamlı sonuçlar (p-değeri  $< 0,05$ ) kutu grafikler üzerinde gösterilmiştir. **40**
- 4.7.** Neolitik sonrası dönemden günümüze risk alel frekans artışı (sol panel) ve azalışı (sağ panel) gözlenen Obezite (mavi) ve Tip 2 diyabet (kırmızı) varyantlarının nötral varyantların alel frekansları (sarı) ile karşılaştırılması. Obezite ve Tip 2 diyabet ile nötral alel frekanslarının ortalamaları Mann-Whitney U testi ile karşılaştırılmış ve Bonferroni düzeltmesi sonucu p-değeri anlamlılıklarına göre anlamlı sonuçlar (p-değeri  $< 0,05$ ) kutu grafikler üzerinde gösterilmiştir. Obezite, Tip 2 diyabet ve nötral alel frekanslarının dönemler arası değişimi ise Wilcoxon testi ile karşılaştırılmış Bonferroni düzeltmesi sonucu p-değeri anlamlılıklarına göre anlamlı sonuçlar (p-değeri  $< 0,05$ ) kutu grafikler üzerinde gösterilmiştir. **41**
- 4.8.** Neolitikten günümüze risk alel frekans artışı (sol panel) ve azalışı (sağ panel) gözlenen Obezite (mavi) ve Tip 2 diyabet (kırmızı) varyantlarının nötral varyantların alel frekansları (sarı) ile karşılaştırılması. Obezite ve Tip 2 diyabet ile nötral alel frekanslarının ortalamaları Mann-Whitney U testi ile karşılaştırılmış ve Bonferroni düzeltmesi sonucu p-değeri anlamlılıklarına göre anlamlı sonuçlar (p-değeri  $< 0,05$ ) kutu grafikler üzerinde gösterilmiştir. Obezite, Tip 2 diyabet ve nötral alel frekanslarının dönemler arası değişimi ise Wilcoxon testi ile karşılaştırılmış Bonferroni düzeltmesi sonucu p-değeri anlamlılıklarına göre anlamlı sonuçlar (p-değeri  $< 0,05$ ) kutu grafikler üzerinde gösterilmiştir. **42**
- 4.9.** Obezite ve Tip 2 diyabet ile ilişkili varyantların risk alelleri için dönem çiftleri arasında alel frekansı değişiklikleri. A panelinde Neolitik-Neolitik sonrası dönem çiftleri, B panelinde Neolitik sonrası günümüz dönem çiftleri, C panelinde Neolitik ve günümüz dönem



- çiftleri incelenmektedir. Sol panellerde alel frekansının artış gösterdiği, sağ panellerde alel frekansının azalış gösterdiği Obezite (mavi), Tip 2 diyabet (kırmızı) ve nötral (gri) ile ilişkili varyantların alel frekanslarının dağılımları çizgi grafiği ile gösterilmektedir. **44**
- 4.10.** Fenotip ile ilişkili ve nötral TNP'lerin  $F_{ST}$  dağılımları. Kırmızı kesikli dikey çizgi, nötral TNP'lerin ortalama  $F_{ST}$  'sini göstermektedir. ANOVA testini takiben uygulanan Tukey testine göre nötral TNP'lerden farklılık gösteren ( $p$  değeri $<0.05$ ) fenotipler turuncu, farklılık göstermeyen ( $p$  değeri $> 0.05$ ) fenotipler mavi ile gösterilmiştir. Kutu grafiklerinin sağ tarafında  $p$  değerlerini göstermekte iken sol tarafında TNP sayılarını temsil etmektedir. **46**
- 4.11.** Anadolu Neolitik, Anadolu Neolitik sonrası ve Yoruba popülasyonları arasındaki popülasyon dal istatistiği analizi için Manhattan grafiği. Hedef popülasyon Anadolu Neolitik sonrası ve dış grup Yoruba seçilmiştir. **51**
- 4.12.** Anadolu Günümüz, Anadolu Neolitik ve Yoruba popülasyonları arasındaki popülasyon dal istatistiği analizi için Manhattan grafiği. Hedef popülasyon Anadolu Günümüz ve dış grup Yoruba seçilmiştir. **52**
- 4.13.** Popülasyon dal istatistiği ve  $F_{ST}$  analizi ile seçim izi yakalanan ve alel frekans artışı gözlenen *TCF7L2* genindeki varyantların risk alel frekanslarının değişimi. **53**
- 4.14.** Popülasyon dal istatistiği analizleri ile seçim sinyali yakalanan varyantların risk alel frekans dağılımları. A panelinde Anadolu Neolitik sonrası dalında seçim sinyali yakalanan ve risk alel frekansı değişimi gözlenen üç Tip 2 diyabet (kırmızı) varyantının frekans dağılımları çizgi grafik ile gösterilmektedir. B panelinde Anadolu günümüz dalında dalında seçim sinyali yakalanan ve risk alel frekansı değişimi gözlenen 9 Tip 2 diyabet (kırmızı) ve bir Obezite (mavi) varyantının frekans dağılımları çizgi grafik ile gösterilmektedir. **54**
- 4.15.** İleri zamanlı genetik benzetim ile Anadolu popülasyonlarında nötral senaryo altında hesaplanan alel frekanslarının dağılımı. A panelinde her dönem için ortalama alel frekansları kırmızı elmas şekli ile gösterilmiştir. **56**
- 4.16.** Neolitikten Neolitik sonrası döneme risk alel frekans artışı (sol panel) ve azalışı (sağ panel) gözlenen Obezite (mavi) ve Tip 2 diyabet (kırmızı) varyantlarının ileri zamanlı genetik benzetim ile nötral senaryo altında hesaplanan alel frekansları (açık sarı) ile karşılaştırılması. Obezite ve Tip 2 diyabet ile benzetimle hesaplanan alel frekanslarının ortalamaları Mann-Whitney U testi ile karşılaştırılmış ve Bonferroni düzeltmesi sonucu  $p$ -değeri anlamlılıklarına göre anlamlı sonuçlar ( $p$ -değeri  $< 0,05$ ) kutu

grafikler üzerinde gösterilmiştir. Obezite, Tip 2 diyabet ve benzetimle hesaplanan alel frekanslarının dönemler arası değişimi ise Wilcoxon testi ile karşılaştırılmış Bonferroni düzeltmesi sonucu p-değeri anlamlılıklarına göre anlamlı sonuçlar (p-değeri <0,05) kutu grafikler üzerinde gösterilmiştir.

57

**4.17.** Neolitikten günümüze risk alel frekans artışı (sol panel) ve azalışı (sağ panel) gözlenen Obezite (mavi) ve Tip 2 diyabet (kırmızı) varyantlarının ileri zamanlı genetik benzetim ile nötral senaryo artında hesaplanan alel frekansları (açık sarı) ile karşılaştırılması. Obezite ve Tip 2 diyabet ile benzetimle hesaplanan alel frekanslarının ortalamaları Mann-Whitney U testi ile karşılaştırılmış ve Bonferroni düzeltmesi sonucu p-değeri anlamlılıklarına göre anlamlı sonuçlar (p-değeri < 0,05) kutu grafikler üzerinde gösterilmiştir. Obezite, Tip 2 diyabet ve benzetimle hesaplanan alel frekanslarının dönemler arası değişimi ise Wilcoxon testi ile karşılaştırılmış Bonferroni düzeltmesi sonucu p-değeri anlamlılıklarına göre anlamlı sonuçlar (p-değeri <0,05) kutu grafikler üzerinde gösterilmiştir.

58

**4.18.** Neolitik sonrası dönemden günümüze risk alel frekans artışı (sol panel) ve azalışı (sağ panel) gözlenen Obezite (mavi) ve Tip 2 diyabet (kırmızı) varyantlarının ileri zamanlı genetik benzetim ile nötral senaryo artında hesaplanan alel frekansları (açık sarı) ile karşılaştırılması. Obezite ve Tip 2 diyabet ile benzetimle hesaplanan alel frekanslarının ortalamaları Mann-Whitney U testi ile karşılaştırılmış ve Bonferroni düzeltmesi sonucu p-değeri anlamlılıklarına göre anlamlı sonuçlar (p-değeri < 0,05) kutu grafikler üzerinde gösterilmiştir. Obezite, Tip 2 diyabet ve benzetimle hesaplanan alel frekanslarının dönemler arası değişimi ise Wilcoxon testi ile karşılaştırılmış Bonferroni düzeltmesi sonucu p-değeri anlamlılıklarına göre anlamlı sonuçlar (p-değeri <0,05) kutu grafikler üzerinde gösterilmiştir.

59

**4.19.** Anadolu antik ve modern dönemlerinde minor alel frekans değerleri hesaplanan nötral (turuncu) varyantlarının ileri zamanlı genetik benzetim ile nötral senaryo artında hesaplanan alel frekansları (açık sarı) ile karşılaştırılması. A panelinde Neolitik-Neolitik sonrası dönem çiftleri, B panelinde Neolitik sonrası günümüz dönem çiftleri, C panelinde Neolitik ve günümüz dönem çiftleri incelenmektedir. Nötral ile benzetimle hesaplanan alel frekanslarının ortalamaları Mann-Whitney U testi ile karşılaştırılmış ve Bonferroni düzeltmesi sonucu p-değeri anlamlılıklarına göre anlamlı sonuçlar (p-değeri < 0,05) kutu grafikler üzerinde gösterilmiştir. Nötral ve benzetimle hesaplanan alel frekanslarının dönemler arası değişimi ise Wilcoxon testi ile karşılaştırılmış Bonferroni düzeltmesi sonucu p-değeri

anlamlılıklarına göre anlamlı sonuçlar ( $p$ -değeri  $<0,05$ ) kutu grafikler üzerinde gösterilmiştir.

**61**

**4.20.** Farklı varyant çağırma yaklaşımları ile hesaplanan alel frekanslarının dağılımları. ANOVA sonuçlarını Tukey testi sonuçları takip etmektedir ve anlamlılık seviyeleri kutu grafiklerinde sunulmuştur.

**62**

**TABLULAR**

<b>Tablo</b>	<b>Sayfa</b>
<b>3.1.</b> GWAS Catalog'dan elde edilen TNP'lerin filtreleme basamakları.	<b>19</b>
<b>4.1.</b> Veri setine dahil edilen Anadolu yayınlanmış antik genom örneklerinin veri üretilme yöntemlerine göre örneklem büyüklükleri, ortalama ve medyan okuma derinlikleri.	<b>33</b>
<b>4.2.</b> Risk alel frekanslarında değişim gözlenen varyantların yolak analizleri.	<b>36</b>
<b>4.3.</b> İkili dönemler arası karşılaştırmalarda ilgili fenotiple ilişkili alellerin frekansının artışı ve azalışına göre TNP sayıları listelenmiştir.	<b>39</b>
<b>4.4.</b> Risk alel frekansı artış gösteren Obezite ve Tip 2 diyabet ile ilişkili varyantların $F_{ST}$ değerlerinin nötral varyantların $F_{ST}$ değerleri ile karşılaştırmasının listesi.	<b>48</b>
<b>4.5.</b> Popülasyon dal istatistiği karşılaştırmalarında belirlenen eşik değerinin üzerinde olan ve risk alel frekans değişimi hesaplanabilen TNP'lerin listesi.	<b>50</b>

## 1. GİRİŞ

Genetik, çevresel ve yaşam tarzı faktörlerinin etkisiyle ortaya çıkan ve tüm dünyada insan popülasyonlarında yaygın görülen hastalıklar kompleks hastalıklar olarak tanımlanmaktadır. Bu hastalıklarla ilişkili genetik yapının anlaşılması için yapılan araştırmalar, aynı anda birçok örnekten yüksek verimli genomik verilerin kısa sürede, daha az maliyetle üretilmesini ve analizini sağlayan yeni nesil dizileme teknolojileri ile ivme kazanmıştır (1). Bu gelişmelere paralel olarak genom boyu ilişkilendirme çalışmaları ile (İng. Genome wide association study, GWAS) bu hastalıklarla ilişkili çok sayıda varyant keşfedilmiştir (2).

Günümüze kadar yapılan genom boyu ilişkilendirme çalışmaları Avrupa popülasyonları odaklı olduğu için (3, 4), Avrupa dışındaki bölgelerde kompleks hastalıklarla ilişkili varyantların etki büyüklüğü (İng. effect size), risk alelleri ve risk alel frekanslarının dağılımı hakkında daha az sayıda literatüre ulaşılabilmektedir.

GWAS Catalog veri tabanında yayınlanan çalışmalara dahil edilen popülasyonların dağılımının eşit olmadığı ve %83,19 oranı ile Avrupa popülasyonlarının baskın olduğu görülmektedir (5). Kompleks hastalıklarda, minor alel frekansı (MAF) %5'ten yüksek ve hastalık üzerindeki tekil etki büyüklüğü düşük olan varyantların poligenik kalıtımla kümülatif olarak hastalık riskini arttırdığı bir süreç görülmektedir (6, 7).

Hem hastalıkların toplumdaki yaygınlığı hem de hastalıklarla ilişkili varyantların frekansının ortalama veya yüksek olması düşünüldüğünde hangi evrimsel güçlerin bu hastalıklarla ilişkili genetik varyasyonu etkilediği ve bu hastalıkların neden günümüzde yaygın olabileceği evrimsel biyoloji alanında temel sorulardan biridir. Bu doğrultuda bugüne dek modern insan genomları ile kapsamlı araştırmalar yürütülmüştür.

Antik DNA temporal genomik bilgi üreterek genomda etki eden evrimsel güçlerin dinamiklerini belirlememize olanak sağlar. Paleogenetik yaklaşımlarla

üretile temporal (zaman-serisi) veri setleri ile popülasyonların zaman içinde deęişen alel frekans yörüngeleri incelenebilir, böylelikle genomun yalnızca statik anlık görüntüsü ile erişilemeyen bilgilere ulaşılmaktadır (8, 9). Bu tez çalışmasında çevresel ve kültürel faktörlerden etkilendięi bilenen ve günümüzde Türkiye’de yaygın bir halk saęlığı problemi olan Obezite ve Tip 2 diyabet seçilmiştir.

Bu tez kapsamında ilk defa bu iki hastalıkla ilişkili genetik yapının Anadolu’da son 10.000 yıllık süreçteki deęişim dinamikleri, paleogenomik, popülasyon genetięi ve ileri zamanlı genetik benzetim yaklaşımları ile araştırılmıştır. Yayınlanmış antik genomik veriler ile Anadolu için günümüze kadar oluşturulmuş en kapsamlı veri seti oluşturularak, düşük okuma derinlikli antik genomik verilerden maksimum verimi almamızı saęlayan alel frekansı hesaplama ve seçim analizleri yöntemleri kullanılarak Obezite ve Tip 2 diyabetin neden yaygın olduęuna dair uzun süre tartışılan hipotezlerin doğrudan test edilmesi hedeflenmiştir.

## 2. GENEL BİLGİLER

### 2.1. İnsan Genomu ve Genetik Varyasyon

DNA (deoksiribonükleik asit), adenin, sitozin, guanin ve timin bazların oluşan çift iplikli bir moleküldür. Çift iplik boyunca yaklaşık 3,2 milyar eşleşen ikili baz çifti bulunur. Bu baz çiftlerinin birkaç yüz baz çiftinden 2,4 milyon baz çiftine kadar çeşitli kombinasyonları genleri oluşturur. İnsan genomu, hücre çekirdeğinde 23 çift,  $2n=46$ , kromozomdan oluşmaktadır.

Genler DNA molekülünün fonksiyonel birimleridir. En güncel GRCh38.p14 genom derlemesi itibariyle proteinleri kodlayan 67.007 gen bulunmaktadır. Genler kodladıkları proteinler aracılığı ile hücre yapıları, metabolizma gibi süreçlere dahil olurlar. DNA üzerindeki değişiklikler metabolizma ve fenotip gibi süreçlerde değişikliklere neden olabilmektedir bu yüzden insan DNA'sını anlamak ve olası değişiklikleri tanımlayabilmek oldukça önemli hale gelmektedir.

İnsan genomunda kromozomların neredeyse tüm genetik içeriğinin dizileri 1990-2003 yılları arasında yürütülen ve uluslararası bir iş birliği olan İnsan Genom Projesi (İng. Human Genome Project, HGP) ile belirlenmiş ve kamuya açık hale getirilmiştir (10). Bu proje ile, insan genomunun dizisinin belirlenmesi, kromozomlarının başlıca bölümleri için genlerin lokasyonlarının haritalanması ve elde edilen genetik bilginin analiz edilmesi için araştırma araçları geliştirilmesi amaçlanmıştır (11). İnsan Genom Projesinin tamamlanmasından günümüze, insan genetik varyasyonunu anlamayı amaçlayan birçok başka proje başlatılmıştır: HapMap Projesi, 1000 Genom Projesi gibi (12-14). Tüm bu projeler, insan genomuna dair önemli bilgi sağlamıştır.

Genomun anlaşılmasıyla birlikte hastalıklarla ilişkili genetik yapının anlaşılmasına yönelik çalışmalar artmıştır. Alel, genomda belirli pozisyondaki genetik dizinin bir veya daha fazla versiyonunu anlatır ve popülasyondaki en yaygın ikinci alelin frekansı minör alel frekansı olarak tanımlanmaktadır. Geniş kapsamlı olarak

varyant DNA sekansındaki deęişiklikleri ifade etmektedir. Şayet DNA dizisindeki bu tek nükleotid varyasyonu popülasyonun en az %1'inde görülürse bu durum da tek nükleotid polimorfizmi (TNP) olarak adlandırılmaktadır.

Alel frekansı, bir alelin ilgili popülasyondaki sıklığını gösteren bir niceliktir. Kompleks hastalıkların toplumdaki yaygınlığı ve ilişkili varyantların frekansının ortalama veya yüksek olduğu düşünöldüğünde hangi evrimsel güçlerin bu hastalıklarla ilişkili alel frekansını etkilediğı ve bu hastalıkların neden günümüzde yaygın olabileceğı evrimsel biyoloji alanındaki temel sorulardan biridir.

## 2.2. Kompleks Hastalıklar

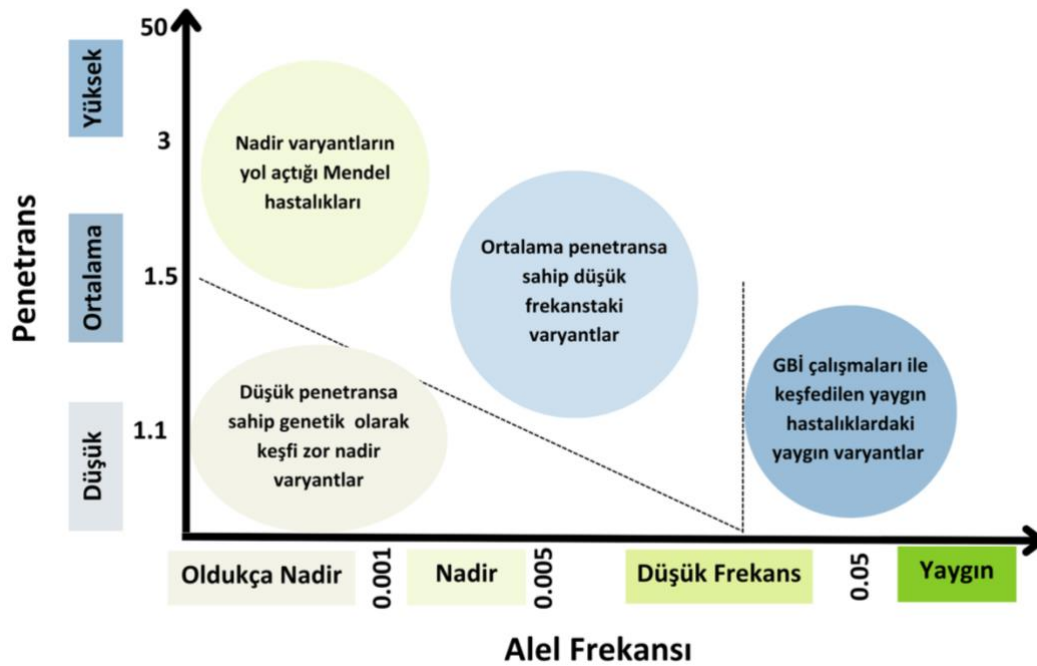
Kompleks hastalıklar; genetik ve çevresel faktörlerin etkisi altında ortaya çıkmaktadır. Bu hastalıklarla ilişkili genetik varyantların çoğunlukla minör alel frekansı %5'ten yüksek ve fenotip üzerindeki tekil etki büyüklüğü düşük olduğu bilinmektedir. (6, 15, 16). Günümüzde insanları büyük ölçüde etkileyen metabolik, psikiyatrik ve otoimmün hastalıklar gibi hastalıkların önemli bir çoğunluğu ve boy uzunluğu, cilt pigmentasyonu gibi fiziksel özellikler kompleks fenotiplerdir (17-21).

Kompleks hastalıklarla ilişkili genetik yapıyı anlamaya yönelik araştırmalar, yeni nesil dizileme (İng. Next Generation Sequencing, NGS) teknolojilerindeki gelişmelere paralel olarak ivme kazanmıştır. Yeni nesil dizileme, tüm genomdaki nükleotit dizisini yüksek verim ve yük hız ile belirlemeyi sağlamaktadır ve Sanger dizilemeden sonra ortaya çıkan tüm dizileme teknolojilerini kapsamaktadır (22). Bu teknolojiler kullanılarak yürütölen genom boyu ilişkilendirme çalışmaları ile kompleks hastalıklarla ilişkili olduğu bilinen genetik varyantlar keşfedilmiştir (1, 7, 23). Genom boyu ilişkilendirme çalışmaları, bireysel fenotipleri tahmin etmek, hastalıkla ilişkili varyantları belirlemek ve kompleks hastalıkların genetik arka planını anlamak için kullanılmaktadır. GBİ çalışmaları ile keşfedilen varyantlar, 'GWAS Catalog' gibi çeşitli kullanıcıların kolay arama yapabildiğı, görselleştirilebilen ve diđer kaynaklarla entegre edilebilen veri tabanlarında saklanarak kullanıcılara sunulmaktadır (24). 'GWAS Catalog', Avrupalılar, Güney ve Doęu Asyalılar, Afrikalılar, Hispanik-Latin Amerikalılar



dahil olmak üzere çeşitli popülasyonlar için GBİÇ ile keşfedilen genetik ilişkileri içermektedir (2). GWAS Catalog'da 6367 araştırma ve 516944 genetik ilişki bulunmaktadır (<https://www.ebi.ac.uk/GWA/>, erişim tarihi Mayıs 2023).

Kompleks hastalıkların genetik arka planını açıklamaya dair çeşitli hipotezler bulunmaktadır (20, 25-27). Bunlar arasında en yaygın hipotez aynı zamanda genom boyu ilişkilendirme çalışmalarının da altında yatan mantık olan yaygın hastalıkların popülasyonun %1-5'inden fazla bulunan düşük penetransa sahip (ilgili varyanta sahip taşıyıcının fenotipte gösterme olasılığı) alellik varyantlarla açıklanabileceğini söyleyen "yaygın hastalık-yaygın varyant" hipotezidir (27-30). Bu görüşün rakibi olarak "yaygın hastalık-nadir varyant" hipotezi mevcuttur. Yaygın hastalık-nadir varyant hipotezi, popülasyonda yaygın olan (%1-5'ten yüksek prevalans) hastalıklarla ilişkili süreçte çok sayıda yüksek penetransa sahip nadir varyantın genetik duyarlılığa katkıda bulunduğunu savunmaktadır (31, 32).



**Şekil 2.1.** Hastalıklar için genetik modeller.

Şekil 2.1.'de hastalıklarla ilgili "yaygın hastalık-yaygın varyant" ve "yaygın hastalık-nadir varyant" genetik modellerini şematize hali görülmektedir. Sağ alt bölgede genom boyu ilişkilendirme çalışmaları, yaygın, düşük penetranslı, risk

varyantlarını tanımlanmıştır. Sol bölgede, tüm genom dizilimi ile daha kolay tespit edilebilecek nadir risk varyantlarını gösterilmektedir.

Daha kapsamlı bir model olarak Omnigenik model genleri fenotipe nedensellik açısından yakınlıklarına göre gruplamaktadır. Bu modele göre gen ürününün fenotip üzerinde doğrudan etkisi olan çekirdek genler üzerinde yer alan çekirdek varyantlar kalıtımın az bir miktarını açıklarken; çekirdek genlerle bir etkileşim ağı içinde olan periferik varyantlar kalıtsallığı daha çok açıklamaktadır (15, 33). Kompleks hastalıklarla ilişkili varyantların frekansı ve etki büyüklüğü popülasyonlar arasında farklılık göstermektedir (34). Bugüne kadar yapılan GBİ çalışmaları çoğunlukla Avrupa popülasyonları ile sınırlı kalmıştır (3, 4, 35), GWAS Catalog'da yayınlanan çalışmalara dahil edilen popülasyonların dağılımının eşit olmadığı ve %83,19 oranı ile Avrupa popülasyonlarının baskın olduğu görülmektedir (5). Dolayısıyla Avrupa dışındaki bölgelerde kompleks hastalıklarla ilişkili varyantların etki büyüklüğü (İng. Effect size), risk alelleri ve risk alel frekanslarının dağılımı hakkında bilgimiz sınırlıdır.

### **2.3. Obezite ve Tip 2 Diyabet**

Bu tez kapsamında günümüz Türkiye popülasyonunda Obezite ve Tip 2 diyabet olmak üzere iki farklı kompleks hastalıkla ilişkili genetik yapının temporal değişimi araştırılacaktır. Obezite, vücut ağırlığının ve adipoz dokunun birçok olumsuz sağlık sorununu tetikleyecek ölçüde atması ve yetişkinlerde vücut kitle endeksinin 30 ve daha yüksek olması olarak tanımlanır (36-38). Tip 2 diyabet kronik hiperglisemi ile karakterize olan pankreasın yeterince insülin hormonu sentezleyemediği veya sentezlendiği durumlarda hedef hücrelerin insüline yeterli ölçüde tepki vermediği durumlarda ortaya çıkan metabolik bir hastalıktır (39).

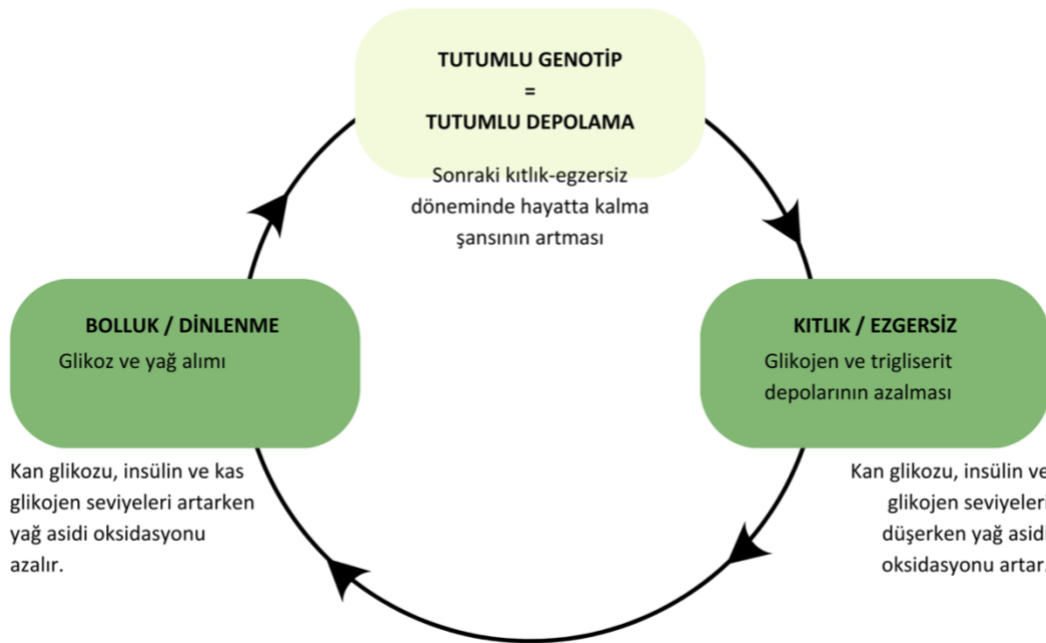
OECD (İktisadi İşbirliği ve Gelişme Teşkilatı) verilerine göre Türkiye'de 2017 yılında 15 yaş üstü aşırı kilolu ve obez nüfus yüzdesi %28,8'dir (40), TÜİK (Türkiye İstatistik Kurumu) 2019 verilerine göre kadınların %30,4'ünün obez öncesi ve %24,8'inin obez olduğu erkeklerde ise bu oranların %39,7 ve %17,3 olarak değiştiği bildirilmiştir (41). Türkiye'de Tip 2 diyabet prevalansı 2013 yılında %16,5 olarak

bulunmuştur (42) . OECD 2018 verilerine göre; ölçülen 15 yaş ve üstü ölçülen obez nüfus yüzdesi Şili %74,2, Amerika Birleşik Devletleri %71, Yeni Zelanda % 66,8, Türkiye % 64,4, Birleşik Krallık %64,3, Kanada % 59,1, Fransa % 49 olarak açıklanmıştır (41). Kalıtım oranları Obezite için %40-%70 ve Tip 2 diyabet için sırası ile ve %20-%80 aralığındadır (43). Tip 2 diyabet ve komorbiditelerin neden olduğu yüksek ölüm oranları ile ilişkilendirilen Obezite (44) genetik arka plan, yaşam tarzı ve çevrenin ortak etkisi ile ortaya çıkmaktadır. Mayıs 2023 itibari ile GBİÇ GWAS Catalog'da Obezite (EFO\_0001073) ile ilişkili 45 çalışma ve 294 ilişki, Tip 2 diyabet (MONDO\_0005148) ile ilişkili 224 çalışma ve 5348 ilişki girilmiştir (<https://www.ebi.ac.uk/gwas/home>, erişim tarihi 17 Mayıs 2023). Metabolizma ile ilişkili bu iki kompleks hastalığın günümüzde hala yaygın görülmekte olmasını açıklamak için farklı evrimsel hipotezler öne sürülmüştür.

**Tutumlu gen hipotezi.** Obezite ve Tip 2 diyabetin günümüzde prevalansının yüksek olmasını açıklayan hipotezlerden biri tutumlu gen hipotezidir (İng. Thrifty Gene Hypothesis) (45). İnsan toplulukları avcı-toplayıcı olarak yaşadıkları dönemde çok sayıda birbirini takip eden kıtlık ve bolluk döngülerine maruz kalmışlardır. Bu iki uç çevresel faktörle başa çıkabilen bireylerin hayatta kalma şansı daha yüksektir. Bu bireylerin üzerindeki seçim baskısı enerji kaybını en aza indirip tasarruf yaparak bolluk dönemlerinde enerjiyi yağ olarak depolamalarına ve sonraki kıtlık döneminde yağ şeklinde depolanan enerjinin hayatta kalmayı kolaylaştırmasına dayanır. Neel tarafından Tutumlu Genotip (45) olarak açıklanan bu hipotezin önermesi, doğal seçim ile besini depolama ve kullanma konusunda verimli olacak şekilde evrimleştiğimizdir. Chakravarthy ve Booth (2004), fiziksel aktivite- dinlenme döngüsünün kıtlık-bolluk döngüsü kadar önemli olduğunu vurgulamıştır (46). Avcı-toplayıcı toplumlarda yiyecek aramanın fiziksel aktivite ile bağlantılı olması bu konuda yetenekli bireylerin hayatta kalarak genlerinin aktarılmasına neden olur. Vücudun bolluk ve egzersiz sonrası dinlenerek toparlanma aşamasında verdiği tepki benzerdir. Şekil 2.2.'de şematize edilen bolluk-kıtlık/egzersiz-dinlenme döngüsüne göre dinlenerek toparlanma süresince yağ asidi oksidasyonunu azaltırken kan glikozunu, insülini ve kas glikojen seviyelerini artırmaktır. Kıtlık ve egzersiz sırasında

ise kan glukozu, insülin ve glikojen seviyeleri düşerken yağ asidi oksidasyonu artar (47).

Tarih boyunca çoğunlukla insan için avantaj olan bu adaptasyon sürekli besin bolluğunun, yüksek kalorili diyetlerin ve düşük aktivite seviyesinin olduğu modern toplumlarda avantaj olmaktan çıkıp Obezite ve Tip 2 diyabet gibi insülin direnci ile ilişkili metabolik hastalıkların görülmesine neden olduğu düşünülmektedir (46, 48).



**Şekil 2.2.** Bolluk-kıtlık/egzersiz-dinlenme döngülerinde vücutta meydana gelen reaksiyon.

**Sürüklenen gen hipotezi.** Tutumlu gen hipotezine yönelik bir eleştiri, şiddetli bolluk- kıtlık döngülerinin varsayıldığı sıklıkta yaşanmamış olabileceğidir. Speakman'ın sürüklenen gen hipotezine (İng. Drifty Gene Hypothesis) (49) göre, erken insan popülasyonlarının predatörlerle yaşadıkları süreç boyunca av olmamak için kilo almama açısından seçim baskısı altında olduğunu öne sürmüştür. 2 milyon yıl önce insan atalarının ateş ve silah kullanma becerileri geliştikçe seçim baskısının kaybolmasına neden olmuş ve bu da metabolizmayı düzenleyen ve kilo alması şımanlık üst sınırını kontrol eden regülatör sistemlerin kodlandığı genlerde genetik sürüklenme ile günümüzde obezite prevalansının yüksek seyretmesine neden olmuştur (50). Günümüz popülasyonlarının genomik verilerini kullanarak kompleks

hastalıklar üzerindeki seçilimi incelemedeki son çalışmalar umut verici görünmektedir. Uzun yıllar üzerinde tartışılan bu iki hipotez de günümüze kadar alel frekansı ve haplotip bazlı seçim testleri ile günümüz popülasyonları üzerinde incelenmiştir.

#### **2.4. Hastalıkların Genetik Arka Planını Etkileyen Evrimsel Süreçler**

Hastalıkların oluşumu evrimsel süreçler perspektifinden incelendiğinde, değişen çevre koşulları nedeniyle oluşan seçim baskısı günümüz popülasyonlarının genetik çeşitliğini ve birçok hastalıkla ilişkili varyantları şekillendirmiştir (20, 51). İnsan evrimsel tarihinin son 100.000 yılı incelendiğinde yaklaşık 100.000 yıl önce Afrika'dan çıkarak Avrasya ve zamanla dünyanın geri kalanına yayıldıkları bu süreçte de introgresyon, darboğaz ve popülasyon genişlemesi gibi demografik olayların popülasyonun genetik yapısını şekillendirdiği bir süreç görülmektedir (20). Hızla değişen çevre koşullarına adaptasyonda yaşanan zorluklar potansiyel evrimsel uyumsuzluk (İng. Mismatch) oluşturmuştur (52, 53).

Son 10.000 yılda Neolitikleşme ile insanların yeni ortamlara maruz kalması, tarım ve kentleşme gibi yaşam tarzı değişikliklerine, yeni beslenme alışkanlıkları edinmelerine ve yeni kültürel normlarla tanışmaları adaptasyon için bir fırsat oluşturmuştur (54-57). Antik DNA çalışmalarının ilerlemesi hem insan atalarının genetik yapısı hakkında hem de o dönemki çevresel koşullar hakkında bilgi edinilmesini sağlamıştır.

Bu çalışmalar sayesinde hızlı değişen çevresel faktörlere kıyasla genetik değişimin bu hızı yakalayamaması ve yeni hastalık modellerinin ortaya çıkışını görebilmekteyiz. Lokal adaptasyonun popülasyonlar arası fenotipik farkların evrilmesi açısından önemli katkıda bulunduğu, popülasyonların diyetlere, patojenlere ve ortam sıcaklıklarına yerel adaptasyon gösterdikleri bilinmektedir (20, 58). Beslenme ile ilgili günümüz çeşitli insan popülasyonlarında lokal adaptasyondan etkilenen farklı genler olduğu bilinmektedir. Lokal adaptasyon genetik değişimle bir popülasyondaki bireylerin yerel çevrelerinde aynı türün diğer popülasyonlarından ortalama daha

yüksek uyum gücüne (İng. Fitness) sahip olmaları ile görülmektedir. Zamanla popülasyonlar arasında genetik ve fenotipik farklılaşmaya yol açan doğal seçilim ile yönlendirilir (58, 59). Örneğin Avrasya ve Afrika popülasyonlarında laktaz direnci ile ilişkili *LCT* (60, 61) ve çinko seviyeleri ile ilişkili *SLC30A9* ve *SLC39A8* (61, 62); Avrupalılarda ergotiyonin eksikliği ile ilişkili *IBD5* (63, 64) ve düşük demir seviyeleri ile ilişkili *HFE* (65, 66), Asyalılarda alkol tüketimi ile ilişkili *ADH1B* (67, 68) geni gibi.

Belirli bir popülasyondan bağımsız beslenme alışkanlıkları ile yapılan bir diğer çalışmada yüksek nişasta tüketimine sahip popülasyonların, düşük olanlara kıyasla ortalama daha fazla *AMY1* (tükürük amilaz geni) kopyasına sahip olduğu bulunmuştur. Neolitikleşme ile tarımla ilgilenen topluluklarda ve kurak ortamda yaşayan avcı-toplayıcı topluluklarda nişasta tüketiminin ayırıcı özelliklerden birisidir. Yüksek *AMY1* kopya sayılarının nişastalı gıdaların sindirimine yardımcı olarak bağırsak için koruyucu tampon görevi görmektedir. Perry ve arkadaşlarının kopya sayısı değişikliklerinin pozitif seçilim örneği olarak bildirdiği bu çalışma insan genomunda bulunan ilk örneklerden birisidir (69).

Küçük ölçekli mikro evrimsel süreçlerde popülasyondaki alel frekansı değişimini etkileyen üç temel evrimsel süreçten bahsedebiliriz: gen akışı, genetik sürüklenme ve doğal seçilim (51).

Gen akışı, bir popülasyondan diğerine alellerin göç yolu ile aktarılmasıdır. 1943 yılında Wright tarafından ortaya atılan Mesafeye Göre İzolasyon (İng. Isolation By Distance, IBD) teorisi yerel genetik farklılıkların yerel uzamsal dağılımı altında birikimine dayanmaktadır (70). Teori sürüklenmenin popülasyonların daha uzak mesafelerde daha fazla farklılaşmasına neden olacağını ve bu popülasyon büyüklüğü azaldıkça daha belirgin olması gerektiğini öne sürmüştür. Coğrafi izolasyona neden olan sıra halinde dağlar, çöller ve büyük su kütleleri popülasyonlar arasındaki gen akışını kısıtlayabilir. Gen akışı yalnızca göç eden bireyler çoğaldığında ve göç edilen popülasyonun gen havuzuna katkıda bulunduğu gerçekleşir (71). Bireylerin popülasyonlar arasındaki sınırlı göçü, bu popülasyonlar arasında genetik

farklılaşmaya yol açan genetik çeşitlilik üzerinde bir etkiye sahip olabilir ve izolasyonda adaptif evrime yol açabilmektedir (51, 70, 72).

Genetik sürüklenme, bir popülasyondaki alel frekanslarının rastlantısal olarak değişmesidir. Bir popülasyondaki genetik karakteristiğin yok olması ya da hayatta kalmasına ve alellerin diğer alellerden bağımsız olarak o popülasyonda yaygın görülmesine neden olur. Özellikle izole olmuş popülasyonlar arası farklılaşmanın başlıca bir mekanizmasıdır (73).

Doğal seçimde bir canlının hayatta kalma ve üreme şansını artıran genomik varyasyonlara sahip bireylerin katısal özelliklerinin gelecek nesillere daha fazla aktarılması sonucunda uyumlu bireylerin popülasyonda çoğalması ile genetik havuzun değişmesidir (51, 74). Doğal seçim; hayatta kalma, doğurganlık, çiftleşme başarısı, uyum gücü veya yaşam döngüsündeki herhangi bir farklılıktan kaynaklanabilmektedir (75).

Popülasyondaki alel frekansı genetik sürüklenme ve seçim olmak üzere iki temel evrimsel güç tarafından şekillendirilmektedir. Kimura tarafından öne sürülen moleküler evrimin nötral teorisi, popülasyondaki genetik varyasyonun çoğunluğunun seçim değil mutasyon ve genetik sürüklenme sonucu gerçekleştiğinden bahsetmektedir (76) (77). Seçilime dayalı modellerde günümüzde veya geçmiş dönemlerde avantaja sahip alellerin pozitif seçim ile frekanslarının yükselmesine, dezavantajlı alellerin ise negatif (arındırıcı) seçim ile popülasyon gen havuzundan uzaklaştırılarak frekanslarının azalmasına dayanır (51, 78).

Antik DNA çalışmaları bu noktada modern genomlarla cevaplayamadığımız birçok soruya yanıt olabilmektedir. Evrime dair zaman serisi veri sağlayarak, zaman içindeki alel frekansı değişikliklerini takip etmeye ve gen-çevre etkileşimine bakabilmemize aracı olmaktadır. Bu sayede seçilimin ne zaman ve nerede gerçekleştiği hakkında daha geniş açıdan bakmaya olanak sağlamaktadır. Ancak literatür taramalarında gözlemlediğimiz özellikle kompleks hastalıklar için antik DNA verilerinden son yıllara kadar yeterince yararlanılmaması olmuştur. Bu durum antik

DNA çalışmalarının limitasyonları ile de açıklanabilmektedir: antik DNA'nın oldukça hasar görmüş fragmente yapısı, üretilen çoğu genomik verinin düşük derinlikte olması, elde edilen örneklem sayısının istatistiksel yorumlama yapabilmemiz için yeterli olmaması. Bu durum uzun yıllar boyunca seçim analizlerinin modern popülasyonlar ile yapılması, kompleks hastalıklardansa tek gen hastalıklarına veya az sayıda fenotipe odaklanması gibi sonuçlar doğurmuştur (63, 79-83). Seçim analizlerinde antik DNA zaman-serisi verilerinin doğal seçim sinyallerini ortaya çıkarma gücünü vurgulayan çalışmalar mevcuttur (80, 84, 85).

## **2.5. Antik DNA ve Paleogenetik**

Antik DNA çalışmaları yaklaşık 37 yıl önce Higuchi ve ark.'ları (1984) tarafından 19. yüzyılda soyu tükenmiş bir Güney Afrikalı quagga ve 36 yıl önce Pääbo ve ark.'ları (1985) tarafından Mısır mumyasının DNA materyalinin elde edilip dizilenmesi ile başlayan (86-89) günümüzde yaşamayan insan ve diğer organizmalardan elde edilen DNA'larla çalışılan bilim alanıdır. Antik DNA kullanılarak tarih öncesi dönem popülasyon genetik yapısındaki dinamiklerin araştırıldığı araştırma alanı olan paleogenetik, arkeoloji, antropoloji, biyolojik bilimleri birleştiren multidisipliner bir bilim alanıdır.

Antik DNA çalışmaları sayesinde insan ortak akrabaları Neandertal ve Denisovalılar hakkında araştırmalarla insan soy ağacında yeni dallar keşfedilmiş (90-92), modern insanın Afrika'dan çıkış sonrası göç yolları, Neandertal ve Denisovalılarla karşılaşmaları ve genetik karışımları ortaya çıkmıştır (93, 94). Arkeobotanik ve arkeozoolojik kalıntıların incelenmesi ile yabancı türlerin evcilleştirilmesi, insanların göç yolları, avcı-toplayıcı hayat tarzından tarım ve çobanlığa geçiş hakkında bilgi sahibi olmamızı sağlamıştır (95, 96). Neolitikleşme ile çobanlık hayat tarzı ile tanışan insan popülasyonları hayvan patojenleri ile tanışmış ve antik DNA ile yapılan bulaşıcı hastalık çalışmaları ile de bulaşıcı hastalıkların ortaya çıkışı, koruyucu genler ve hastalıkların yayılım süreci hakkında anlayışımızı değiştirmiştir (97-101).



Antik DNA çalışmalarındaki bu ilerlemeler; yıllar içerisinde yeni nesil dizileme teknolojisindeki gelişmeler (102), antik DNA'nın daha saf ve yüksek derinlikli üretilmesine dair optimize edilmiş laboratuvar süreçleri ve elde edilen verinin kalitesinin artırılmasına dair biyoinformatik yaklaşımlar (impütasyon, ölüm sonrası DNA hasarının (İng. post-mortem DNA damage) tespiti ve giderilmesi, haplodizasyon gibi) sayesinde olmuştur. Yeni nesil dizileme ayrıca antik DNA analizinde sıklıkla karşılaşılan önemli limitasyonlar olan, kontaminasyon ve ölüm sonrası DNA hasarı gibi sorunların üstesinden gelmek için istatistiksel çözümler sunmuştur (103-106).

Paleogenomik yaklaşımlar ile antik dönemlerdeki popülasyonların demografik geçmişleri, göç hareketleri ve patojenlerle ilişkileri hakkında bilgi edinmek mümkün hale gelmiştir (55, 63, 107, 108). Paleogenetik yaklaşımlarla oluşturulan zaman-serisi genomik veri setleri ile popülasyonlardaki alel frekans dinamikleri, haploblok ve bağlantı dengesizliği bloklarının belirlenmesi ve zaman içindeki değişimi incelenerek, genomun yalnızca günümüzdeki hali ile erişilemeyen bilgilere ulaşılabilir (8, 9).

## **2.6. Kompleks Fenotipleri Antik DNA ile Araştırmak**

Antik DNA, evrime dair zamansal genetik bilgi sağlayarak ve zaman içindeki alel frekansı değişikliklerini takip ederek evrimsel güçlerin temporal dinamiklerinin incelenmesini mümkün kılmaktadır (85). Günümüze kadar antik DNA çalışmaları hem dünya genelinde hem Anadolu antik dönemleri üzerine yapılan çalışmalarda demografik süreçlere odaklanmış, insan popülasyonlarının göçleri, birbirleri ile etkileşimi ve akrabalık ilişkileri gibi birçok soruyu yanıtlamaya yardımcı olmuştur (55, 63, 107-109). Son yıllarda demografik odaklı çalışmalara ek olarak hastalıkların ve fenotipik özelliklerin arka planında evrimsel mekanizmaları anlamaya yönelik araştırmalarda artış gözlenmektedir.

Antik DNA seçilim analizlerinde oldukça bilgi verici olsa da yakın zamana kadar, bu çalışmaları yüksek istatistiksel hassasiyetle yürütmek için gereken büyük örneklem boyutları ve yüksek derinlikli genomik veri mevcut değildi. Bu nedenle, antik DNA verileri ile doğal seçilimi incelemeye yönelik ilk çabalar sınırlı kalarak

seçilim analizlerinde modern popülasyonlar kullanılmış (63, 79) ve genellikle aday lokuslara veya sınırlı sayıda fenotipe odaklanmıştır (80-82, 84).

Son yıllarda hem dizileme teknolojilerindeki gelişmeler hem de antik DNA çalışmalarının optimize edilmesi üzerine yapılan çalışmalar daha otantik ve yüksek derinlikli genomik verilere ulaşmamızı sağlamıştır (110-112). Bu sayede kompleks hastalık ve özellikleri çalışabilmek için gerekli olan hem daha yüksek derinlikli genom boyu dizileme yapma ve daha yüksek istatistiksel hassasiyetle yürütmek için gerekli örneklem boyutlarına ulaşma şansı artmıştır (80, 84).

## **2.7. Anadolu ve Paleogenetik Araştırmalar**

Bu tez kapsamında 2 kompleks özellik olan Obezite ve Tip 2 diyabet ile ilişkili varyantların Anadolu'da yaşamış antik insan topluluklarından elde edilecek genomik veri kullanılarak incelenecektir. Anadolu'nun önemi şudur: Günümüzden 10.000 yıl önce avcılık, balıkçılık ve toplayıcılığa dayalı bir yaşam tarzından çiftçilik ve çobanlığa dayalı başka bir yaşam tarzına geçişin başladığı Neolitikleşme süreçleri önemli bir dönüm noktasıdır (113). Avcı-toplayıcı yaşam biçimin görüldüğü Paleolitik dönemden kademeli olarak insanların daha büyük gruplar halinde yaşamaya başladığı tarım-hayvancılık yaptığı Neolitik dönemde yaşam tarzındaki değişiklikler ile insan popülasyonları beslenmelerinde, kültürel ve sosyal ilişkilerinde değişimler yaşamış, hayvancılık ve yerleşik yaşamla beraber yeni patojenlerle karşılaşmışlardır (20, 54, 55). Yetersiz beslenme ve bulaşıcı hastalıkların artması ile fizyolojik stres olduğu Neolitik popülasyonların iskelet analizleri ile görülmüştür (56, 57, 114, 115).

Anadolu Neolitikleşme sürecinin gerçekleştiği ana bölgenin içinde yer alması ve değişim sürecinin kademeli yaşanması sayesinde tarih öncesi kalıntıların ait oldukları kültürlerin net bilinmesi bu sürecin çalışması için eşsiz bir bölge olmasını sağlamaktadır (116). Günümüze kadar Anadolu antik popülasyonları ile yapılan çalışmalar popülasyon hareketleri, akrabalık ilişkileri, göç yolları üzerine odaklanmış (55, 117-120) olup elde edilen genomik verinin düşük okuma derinliğine sahip olması ve örneklem büyüklüğünün düşük olması geniş kapsamlı seçim taramalarının önüne

geçmiştir. Ancak geçtiğimiz 2 yıl süresince yayınlanan çalışmalar sayesinde Neolitik, Neolitik sonrası ve tarihi dönemler için daha fazla genomik veriye ve daha yüksek derinlikli genomlara ulaşabilmekteyiz (117, 119, 121). Bu tez çalışmasında Ocak 2023'e kadar yayınlanmış Anadolu antik genomlarını dahil ederek Obezite ve Tip 2 diyabet hastalıkları üzerine alandaki bu açığı kapatmayı hedeflenmektedir.

### 3. GEREKÇE VE YÖNTEM

Bu tez kapsamında, Obezite ve Tip 2 diyabet ile ilişkili genetik yapının son Neolitikleşme süreciyle beraber nasıl değiştiği, paleogenomik yaklaşımlarla farklı popülasyon genetiği yaklaşımları ile incelenmiştir. Bunun için i) Literatürden fenotiple ilişkili ve nötral bölgelerden varyantlar seçilmiş, ii) Son 10.000 yıllık Anadolu antik genom verileri ve Günümüz Anadolu genom verileri ile veri seti oluşturulmuştur, iii) Seçilen varyantlar çeşitli popülasyon genetiği yöntemleri ile oluşturulan veri setleri kullanılarak incelenmiştir.

#### 3.1. İlgilenilen Genomik Bölgelerin Belirlenmesi

##### 3.1.1. Kompleks Fenotiplerle İlişkili Varyantların Belirlenmesi

Bu tez çalışmasında, Bölüm 2.3.'te bahsedilen günümüzde Türkiye'de yaygın görülen, genetik arkaplan, yaşam tarzı ve çevrenin ortak etkisi ile ortaya çıkan Obezite ve Tip 2 diyabetle ilişkili TNP'lere odaklanılmıştır. Bu iki fenotiple ilişkili TNP'ler GWAS Catalog'dan (<https://www.ebi.ac.uk/gwas/>, erişilme tarihi 2021-09-7) (2) indirilmiştir. TSV formatında indirilen TNP'ler Tablo 3.1.'de belirtilen sırayla filtrelenmiştir. İlk olarak risk alel bilgisi mevcut olmayan ilişkiler filtrelenmiş daha sonra risk aleli indel olan ilişkiler, birden fazla farklı risk alel bilgisi mevcut olan ilişkiler, Bonferroni düzeltmesi sonrasında ortak varyantları belirlemek için en katı eşik değeri olan p-değeri  $5 \times 10^{-8}$ 'den büyük varyantlar filtrelenmiştir. Bu filtreleme sonucunda elde edilen ilişkiler içinden duplike olan TNP'ler manuel filtreleme ile çıkarılmış elde edilen TNP'ler içinde bialellik olmayanlar Ensembl Biomart (<https://www.ensembl.org/biomart/martview>) (Ensembl release 106) ile filtrelenmiş ve kalan TNP'lerden Nucleotide BLAST (<https://blast.ncbi.nlm.nih.gov>) sonucu genomda tek bir noktaya hizalanan TNP'ler seçilmiştir. En son filtre olarak Plink (v.1.9) (122) yazılımı ile 1000 Genom Projesi Faz 3 GRCh38 (123) ([http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data\\_collections/1000\\_genomes\\_project/release/20181203\\_biallelic\\_SNV/](http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data_collections/1000_genomes_project/release/20181203_biallelic_SNV/)) veri setinde gözlenen ve “-indep-pairwise 200 25 0.4” parametresi (63) ile bağlantı dengesizliğinde bulunmayan TNP'ler seçilmiştir.

Toplamda 432 fenotip ile ilişkili TNP hedeflenmiştir hem Obezite hem Tip 2 diyabet ile ilişkili bir TNP mevcuttur: rs476828 (Tablo 3.1.). Genom boyu ilişkilendirme çalışmaları, genom boyunca tek nükleotid polimorfizmlerini karşılaştırır ve hangi TNP'lerin fenotip ile kalıtılmasının muhtemel olduğunu saptarken fenotipe sahip olan ve olmayan birçok bireyi karşılaştırmak için bunları belirteç olarak kullanmaktadır. Genom boyu ilişkilendirme çalışmalarında, çoklu testleri hesaba katmak için, ortak bir genetik varyant ile ilgilenilen bir hastalık arasındaki ilişkiyi tanımlamak için  $5 \times 10^{-8}$  lik sabit bir p-değeri eşiği yaygın olarak kullanılmaktadır (124). İlk olarak kompleks hastalıkların genetiği incelenirken aday gen çalışmaları tarafından bildirilen birçok yanlış pozitif keşif nedeniyle bu katı p-değeri eşiği Risch ve Merikangas (1996) tarafından önerilmiştir (125). Genom çapında anlamlı ilişki terimi, bu eşiğe ulaşan varyantlar için kullanılmaktadır.

### 3.1.2. Nötral Varyantların Belirlenmesi

Genomdaki nötral bölgeleri bulmak için Neutral Region Explorer (NRE) (<http://nre.cb.bscb.cornell.edu>) programı kullanılmıştır. Temel parametreler ile minimum bölge boyutu 200 baz çifti (bç), rekombinasyon hızı 0.9 centiMorgan/Megabaz (cM/MB) olarak ayarlanarak HapMap genetik haritası verisi kullanılmış, minimum arka plan seçilim katsayısını 0,95 olarak, HapMap popülasyonlarından CEU (CEPH koleksiyonundan Kuzey ve Batı Avrupa kökenli Utah bireyleri) ve Mask parametresi katı (İng. Strick) olarak ayarlanmıştır. Elde edilen nötral bölgeler 1000 Genom Projesi Faz 3 veri seti ile BEDTools (126) ile kesiştirilerek nötral varyantların pozisyonları ve kimliklerine dair bilgi edinilmiştir. Nötral varyantlar şu şekilde filtrelenmiştir: Biallelik olmayan TNP'ler Ensembl Biomart (<https://www.ensembl.org/biomart/martview>) (Ensembl release 106) ile filtrelenmiş ve kalan TNP'lerden Nucleotide BLAST (<https://blast.ncbi.nlm.nih.gov>) sonucu genomda tek bir noktaya hizalanan TNP'ler seçilmiştir. Toplamda 1000 nötral TNP için Ensembl Biomart ile (<https://www.ensembl.org/biomart/martview>) (Ensembl release 106) kromozom, genomik pozisyon (GRCh38), varyant kimlik bilgisi ve global

minör alel bilgisi çekilerek BED dosya formatında çalışmanın sonraki basamakları için saklanmıştır.

**Tablo 3.1.** GWAS Catalog'dan elde edilen TNP'lerin filtreleme basamakları.

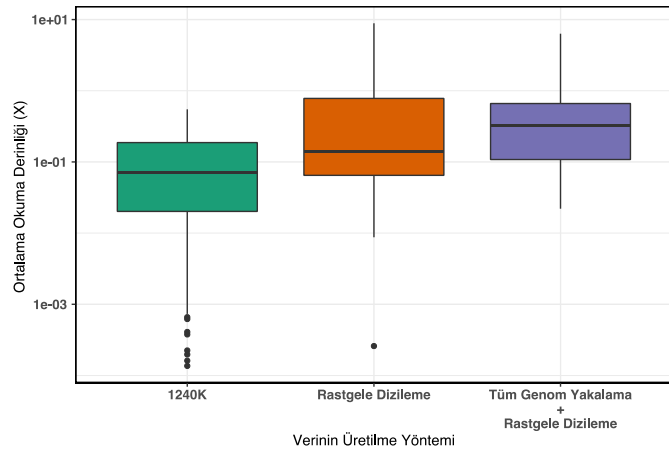
Fenotip	GWAS Catalog Numarası	GWAS Catalog'daki ilişki sayısı	Risk alel bilgisi mevcut olmayan ilişkilerin filtrelenmesi	Risk aleli indel* olan ilişkilerin filtrelenmesi	Birden fazla risk aleli olan ilişkilerin filtrelenmesi	$p < 5 \times 10^{-8}$ olan ilişkiler	Manuel filtreleme**	Bialellik TNP'ler	BLAST ve bağlantı dengesizliği filtresi
Obezite	EFO_0001073	289	209	209	101	101	77	41	18
Tip 2 diyabet	MONDO_0005148	3796	3274	3109	2471	1670	1329	818	415

\*1 baz uzunluğundan fazla silinme ve eklenme mutasyonlardır, \*\*Manuel filtreleme basamağı TNP'lerin son kez incelenmesi ile veride kalan duplikelerin filtrelenmesidir.

## 3.2. Zaman Serisi Genomik Veri Setleri

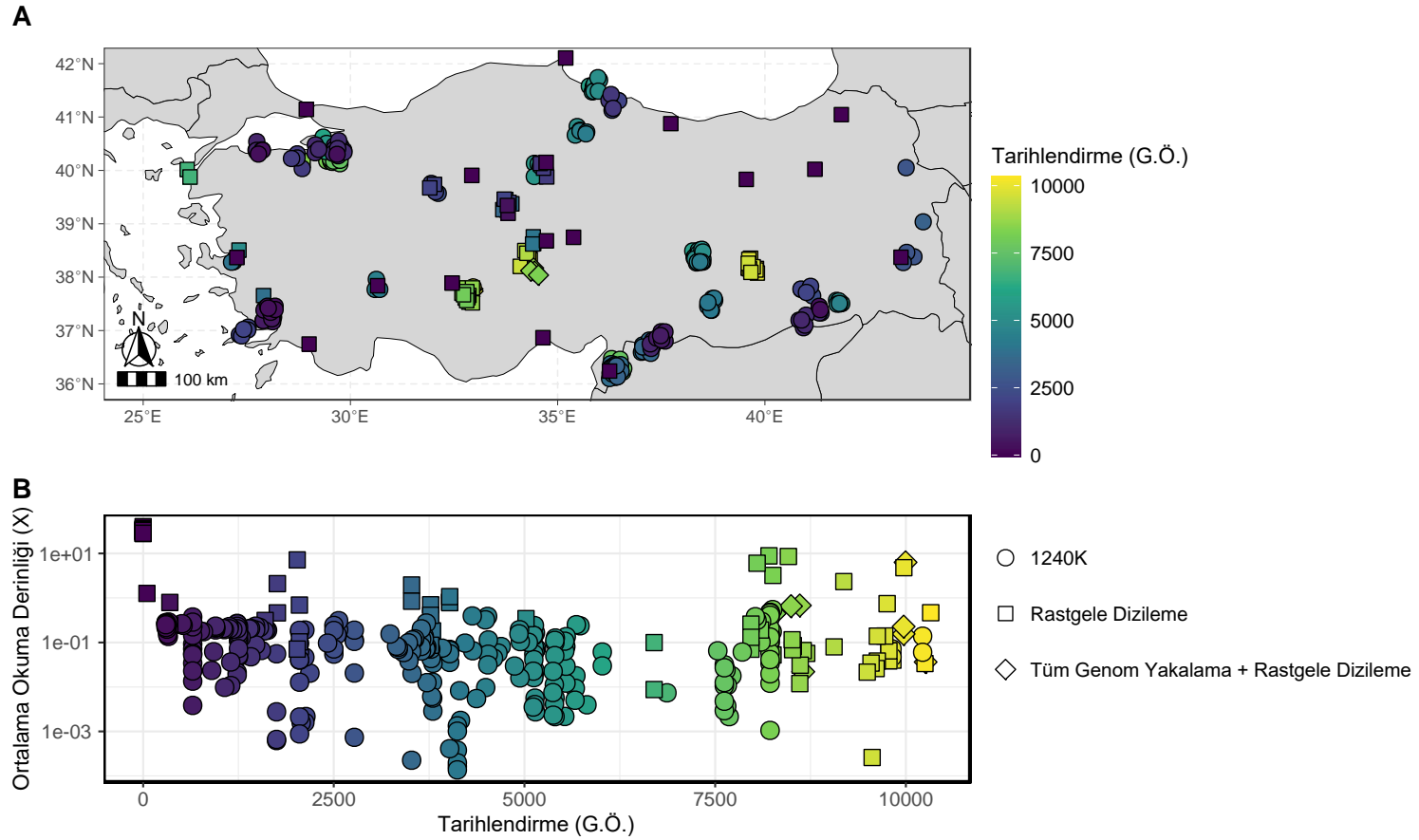
### 3.2.1. Antik Genomik Veri Seti için Örnek Seçimi

Veri setimize Anadolu'dan yayınlanmış tüm antik insan genom dizileme verileri (n= 319) (Ocak 2023 itibariyle) dahil edilmiştir (55, 63, 117, 119-121, 127-133). Toplam 319 örneğin karbon 14 tarihlendirme yaşları günümüzden önce (G.Ö.) 10326-50 arasında değişmektedir. Ayrıca Anadolu'dan 16 günümüz bireyi (134) ve popülasyon dal istatistiği analizinde dış grup olarak kullanmak için HGDP (Human Genome Diversity Project) veri setinde yer alan günümüz Nijerya popülasyonunu temsilen Yoruba (n=18) popülasyonu dahil edilmiştir. Anadolu popülasyonlarını arkeolojik dönemlendirmeye göre üç gruba ayrılmıştır: Neolitik (n= 81) (G.Ö. 10326-6700) (119, 121, 127, 130, 133), Neolitik sonrası (n= 238) (G.Ö. 7720- 50) (119, 121, 127, 130, 133), günümüz (n= 16) (134). Anadolu antik genomik veri setinde Neolitik dönem için 36 tane 1240K, 37 tane rastgele (İng. Shotgun) dizileme ve 8 tane tüm genom yakalama (İng. Whole Genome Capture) ile birleştirilmiş rastgele dizileme yöntemi ile; Neolitik sonrası dönemde 218 tane 1240K ve 20 tane rastgele dizileme yöntemleri ile üretilmiş genomik veriler kullanılmıştır (Şekil 3.1.).



**Şekil 3.1.** Veri setine dahil edilen yayınlanmış Anadolu antik genomik verilerinin veri üretim yöntemi ve okuma derinliklerine (X) göre dağılımının kutu grafiği ile gösterilmesi.





**Şekil 3.2.** Veri setine dahil edilen Anadolu yayınlanmış antik ve günümüz genom örneklerinin, A panelinde lokasyonu ve B panelinde karbon 14 tarihlendirmeleri ve ortalama okuma derinliğine göre dağılımı gösterilmiştir. B panelinde y eksenini log (10) tabanına göre ölçeklendirilmiştir.

### 3.2.2. Antik Genom Verilerinin Hizalanması

Bu çalışmada yayınlanmış genomik veriler Kıvılcım Başak Vural tarafından Orta Doğu Teknik Üniversitesi CompEvo Laboratuvarı sunucusuna fastq dosya formatında indirilerek referans genoma hizalanmıştır. Öncelikle AdapterRemoval (135) programı kullanılarak adaptörler çıkarılmış ve iki uçlu okumalar birleştirilmiştir. Birleştirilen okumalar Burrows-Wheeler Aligner (BWA) (136) aln ve samse modülleri kullanılarak insan referans genomuna NCBI GRCh38 (assembly GRCh38.p13, Annotation Release 109) '-l 16500 -o 0.02 -q30' parametreleri kullanılarak hizalanmıştır. Polimeraz zincir reaksiyonu duplikeleri Picard (<https://github.com/broadinstitute/picard>) kullanılarak filtrelenmiştir. 35 nükleotitten daha kısa olan ve toplamda %10'dan daha yüksek oranda yanlış eşleşme (İng. mismatch) içeren okumalar Kılınç vd. (55) 2016 makalesinde yayınlanan bir shell kodu ile filtrelenmiştir. Hizalama dosyaları 'BAM' dosya formatında saklanmıştır.

### 3.2.3. Modern Genom Veri Setlerinin İşlenmesi

İnsan Genomu Çeşitliliği Projesi'nden (İng. Human Genome Diversity Project, HGDP) günümüz Nijerya bireylerinin (137) (Yoruba, n=18) GRCh38 referans genomuna hizalı CRAM dosyaları <https://www.internationalgenome.org/data-portal/population/YorubaHGDP> adresinden Kıvılcım Başak Vural tarafından indirilmiş ve CRAM dosyaları BAM dosya formatına dönüştürülerek saklanmıştır. Günümüz Anadolu bireylerinin (n=16) yayınlanmış genomik veriler fastq dosya formatında indirilmiştir (134).

## 3.3. Veri Setlerinin Hazırlanması

### 3.3.1. Alel Frekans Hesaplamaları için Veri Seti Hazırlanması

Maksimum olabilirlik bazlı frekans hesaplama yöntemi için diploid varyant çağırma işlemi gerçekleştirilmiştir. Bölüm 3.2.2.'de açıklandığı gibi aynı iş hattından geçirilerek insan referans genomu GRCh38'e hizalanmış her BAM dosyasından SAMtools (v.1.9), mpileup (138) kullanılarak VCF dosyaları oluşturulmuştur.

samtools mpileup -R -B -q30 -Q30 -f ref.fa -g bamfile | bcftools call -m > outfile

VCF (Varyant Çağrı Formatı) dosyaları bir başlık bölümüne ve bir ana gövde bölümüne sahip olan her satırın bir varyant ve dosyadaki bireyler tarafından taşınan genotipler hakkında bilgilerden oluşmaktadır (139). VCF dosyalarından gerekli bilgiler GATK (140) “-VariantsToTable” fonksiyonu ile üretilerek matematiksel denklemde (Formül 3.1) kullanılmak üzere saklanmıştır

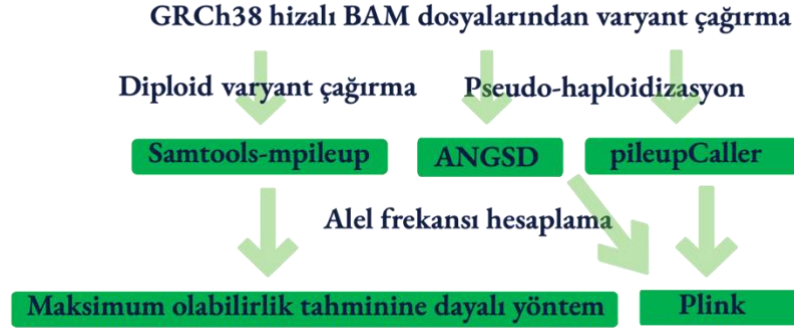
### 3.3.2. Biyoinformatik Yazılımların Karşılaştırması için Veri Seti

Yüksek okuma derinliğine sahip olmaları nedeniyle modern genomlarda diploid varyant çağırma işlemini güvenilir bir şekilde gerçekleştirilirken, antik genomlar genellikle sözde-haploidizasyon prosedüründen geçmektedir. Tezin bu kısmında farklı yaklaşımlar kullanan programlar ve algoritmalar ile antik genomlardan varyantlar çağırarak yaklaşımlar arasında farklılıklar olup olmadığını incelemek amaçlanmıştır.

Bunun için iki farklı program kullanılarak sözde-haploidizasyon ile varyant çağırma ve diploid varyant çağırma uygulanmıştır. ANGSD ile varyant çağırma işleminde ANGSD (v.0.934) (141) aracı “-doHaploCall 1” seçeneği kullanılarak tüm BAM dosyalarından her TNP başına bir alel rastgele seçilerek plink (v.1.90) (122) yazılımında girdi olarak kullanılacak tped/tfam dosya formatında sözde-haploid genotipler oluşturulmuştur. Açık kaynak kodlu C/C++ programı olan ANGSD, Linux sistemlerinde kullanılabilen BAM, VCF ve impute edilmiş beagle genotip dosyaları gibi farklı dosya formatlarını girdi olarak kabul eden yeni nesil dizileme verilerini analiz etmeye yönelik oluşturulmuş bir yazılımdır (141).

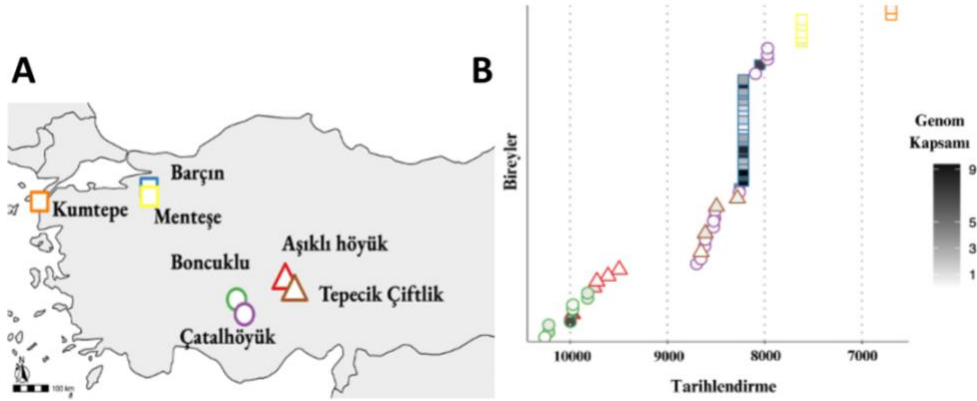
pileupCaller ile varyant çağırma işlemi iki adımda gerçekleştirilmiştir: Pileup dosyaları SAMtools (v.1.9) mpileup (138) yazılımı ile oluşturulmuş ve PileupCaller (v.1.4.0) aracı (<https://github.com/stschiff/sequenceTools>) ve “-randomHaploid” seçeneği kullanılarak tüm BAM dosyalarından her TNP başına bir alel rastgele seçilerek sözde-haploid genotipler oluşturulmuştur (base quality>30 and MAPQ>30).

Allel frekansı hesaplanırken Bölüm 3.1.1.'de seçilen Tip 2 diyabet ile ilişkili varyantlar kullanılmış ve her TNP için iki farklı yöntem kullanılarak hesaplanmıştır.



**Şekil 3.3.** Farklı varyant çağırma yaklaşımlarının test edilmesi için uygulanan iş akışı.

Yukarıdaki iş akışı grafiğinde iki farklı yaklaşım ile varyant çağırma işlemi ve sonrasında uygulanan allel frekansı hesaplama yöntemi gösterilmiştir. Varyantları çağırma ve allel frekanslarını hesaplamak için 58 Neolitik Anadolu bireyine ait antik genom dizileme verileri (BAM dosyaları) kullanılmıştır: Kumtepe'den 2 birey (132), Menteşe'den 5 birey (63), Barçın'dan 21 birey (63, 129), Tepecik Çiftlik'ten 4 birey (128), Boncuklu'dan 8 (55), Aşıklı Höyük'ten 5 birey (120), Çatalhöyük'ten 13 birey (120).



**Şekil 3.4.** Karşılaştırmalı analizde kullanılan bireylerin A) lokasyonu, B) yaş ve ortalama okuma derinliklerinin dağılımı.

### 3.4. Alel Frekans Hesaplamalarının Yapılması

Alel frekanslarını hesaplamak için seçilen fenotiplerle (Tablo 3.1) ile ilişkili TNP'lerin risk alel bilgileri GWAS Catalog'dan (2) elde edilmiş ve oluşturulan veri setlerinde risk alel frekansı hesaplanmıştır. Çalışılması hedeflenen antik bireyler düşük okuma derinliğine (İng. coverage) sahip olduğu (ortalama= 0,29) için sekans verilerinden maksimum verimi almak hedeflenmiştir. Bu amaçla binomial olabirlik dağılımını kullanarak her popülasyon için hesaplanan log-olabirlik dağılımını maksimize eden değerleri bulan Mathieson vd., 2015'ten (63) uyarlanan eşitlik (3.1.) kullanılmıştır. Bu formül şu şekilde çalışmaktadır: Belirli bir lokusta, her bir popülasyon için sekans seviyesi verilerine sahip N örneğimiz olduğunu varsayalım. Sekans seviyesi verilerine sahip her bir örnek i için,  $T_i$  toplam sekans sayısı,  $R_i$  risk alele sahip sekans sayısını temsil eder. Her bir popülasyon için risk alel frekansı  $p$ ,

$$L(p) = \prod_{i=1}^N \{ p^2 B(R_i, T_i, 1 - \epsilon) + 2p(1 - p) B(R_i, T_i, 0.5) + (1 - p)^2 B(R_i, T_i, \epsilon) \} \quad (3.1.)$$

eşitliği ile hesaplanmıştır.

$L(p)$  değerini maksimize eden  $p$  değeri R yazılımı ile bulunmuş ve risk alel frekansı olarak kabul edilmiştir ([https://github.com/idiltac/YLisans\\_Tez\\_IdilTac/blob/main/AlelFrekansHesaplama.R](https://github.com/idiltac/YLisans_Tez_IdilTac/blob/main/AlelFrekansHesaplama.R)).

$$B(k, n, p) = \binom{n}{k} p^k (1 - p)^{n-k} \quad (3.2.)$$

Binomial olabirlik dağılımını (3.2.),  $\epsilon$  hata olasılığını 0,01) ve  $L(p)$  log-olabirlik değerini temsil etmektedir.

### 3.5. $F_{ST}$ Hesaplanması

$F_{ST}$ , sabitlenme indeksi, popülasyon genetiği analizlerinde sıklıkla kullanılan iki popülasyon arasındaki genetik mesafe hakkında bilgi veren bir istatistiktir (142). Popülasyon genetiği analizlerinde ilgilenilen popülasyonlar arasındaki farklılaşma miktarını ölçebilmek oldukça önemlidir.  $F_{ST}$  değerleri 0 ve 1 arasında ve 1'e yaklaşan

değer iki popülasyon arasındaki farklılaşma/uzaklaşma seviyesinin yüksek olduğunu gösterir. Ayrıca gen ve lokus bazlı seçilim analizlerinde,  $F_{ST}$  dağılımının üst kuyruğundaki değerler varsayımsal olarak seçilim altında kabul edilmektedir (143, 144). Alel frekanslarının değişimine dayanan popülasyon farklılaşma ve seçilim analizlerinde kullanılan  $F_{ST}$  tahmini için farklı tahminciler vardır: Wright'ın  $F_{ST}$  (145), Weir-Cockerham'ın  $F_{ST}$  (146), Nei'nin  $F_{ST}$  (147), Hudson'ın  $F_{ST}$  (148).

Sıklıkla kullanılanlardan biri olan Nei'nin  $F_{ST}$ 'si (147) aşağıdaki gibi bir formül ile hesaplanmaktadır:

$$F_{ST} = \frac{H_T - H_S}{H_T} \quad (3.3.)$$

Formülde  $H_T$  toplam popülasyonun heterozigotluğunu temsil etmekte ve  $H_S$  alt popülasyonlar arasında ortalaması alınmış heterozigotluğu etmektedir.  $F_{ST}$  istatistiği; son 10.000 yıldaki antik Anadolu popülasyonları arasındaki genetik farklılaşmayı değerlendirmek için hesaplanmıştır. Popülasyonların okuma derinlikleri ve örneklem büyüklüklerinin farklı olması nedeniyle ikili  $F_{ST}$  genotip olabilirliklerine dayalı bir yaklaşım kullanan ANGSD programı ile hesaplanmıştır.

Özellikle düşük ve orta derinlikteki veriler için kullanışlı olması antik DNA verilerinde  $F_{ST}$  analizini yapmak için seçme nedenimizdir. Ancak yine de düşük okuma derinliğine sahip genomik veri ile çalışırken yanlış pozitif sonuçları engellemek için genomdaki strick mask (<https://www.internationalgenome.org/announcements/genome-accessibility-masks/>) filtresinden geçmiş erişilebilirlik maskeli (İng. accessibility-masked) bölgelerini kullanarak  $F_{ST}$  hesaplanmıştır (149).

ANGSD (141) ile her TNP için  $F_{ST}$  istatistiği tahmini yapabilmek için popülasyon çiftleri arasındaki sonlu olasılık tahmini (İng. posterior expectation) genetik varyansı (A) ve beklenen toplam varyansı (B) hesaplanmıştır (150). A ve B değerleri Reynolds ve ark., (142)'nin Alfa ve Beta tahmin değerleri ile ilişkilidir (144). *Homo sapiens* atasal referans dosyası Ensembl'dan indirilmiştir (<http://ftp.ensembl.org/pub/release->

107/fasta/ancestral\_alleles/homo\_sapiens\_ancestor\_GRCh38.tar.gz, Ensembl Release 107, erişim tarihi 2022-08-20). Her kromozom için ayrı ayrı olan atasal referans fasta dosyaları kromozom sıralamaları göz önünde bulundurularak birleştirilmiş ve tüm kromozomlar için tek bir atasal referans fasta dosyası olarak kullanılmıştır.

Her kromozom için ayrı ayrı olan atasal referans fasta dosyaları kromozom sıralamaları göz önünde bulundurularak birleştirilmiş ve tüm kromozomlar için tek bir atasal referans fasta dosyası olarak kullanılmıştır.

Her TNP başına  $F_{ST}$  hesaplayabilmek için, tüm TNP'ler için bilgi içeren alel frekans spektrumunu (İng. site-frequency spectrum, SFS) değerleri kullanılmaktadır. İlk olarak, ANGSD'da antik popülasyonlar için minimum 20 baz kalitesi ve minimum 20 eşleme kalitesi, günümüz popülasyonları için minimum 30 baz kalitesi ve minimum 30 eşleme kalitesi kullanarak "site allele frequency (SAF)" uzantılı dosyalar elde edilmiştir.

#### **Antik popülasyonlar için:**

ANGSD -b bamlist -ref ref.fa -anc anc.fa -out outfile/ -dosaf 1 -gl 2 -doCounts 1 -uniqueOnly 1 -remove\_bads 1 -only\_proper\_pairs 1 -minMapQ 30 -minQ 30 -minInd 5 -setMinDepth 7

#### **Günümüz popülasyonları için:**

ANGSD -b bamlist -ref ref.fa -anc anc.fa -out outfile/ -dosaf 1 -gl 2 -doCounts 1 -uniqueOnly 1 -remove\_bads 1 -only\_proper\_pairs 1 -minMapQ 20 -minQ 20 -minInd 5 -setMinDepth 7

Üretilen SAF dosyaları kullanarak, ANGSD "realSFS" ile popülasyonlar arası örneklem büyüklüğüne farklılıklarına daha az duyarlı olan Hudson-Bhatia  $F_{ST}$  tahmincisini kullanılarak, her TNP için çift yönlü  $F_{ST}$  hesaplanmıştır (150).

```
realSFS pop1.saf.idx pop2.saf.idx > pop1.pop2.ml
```

```
realSFS fst index pop1.saf.idx pop2.saf.idx -sfs pop1.pop2.ml -fstout
pop1.pop2
```

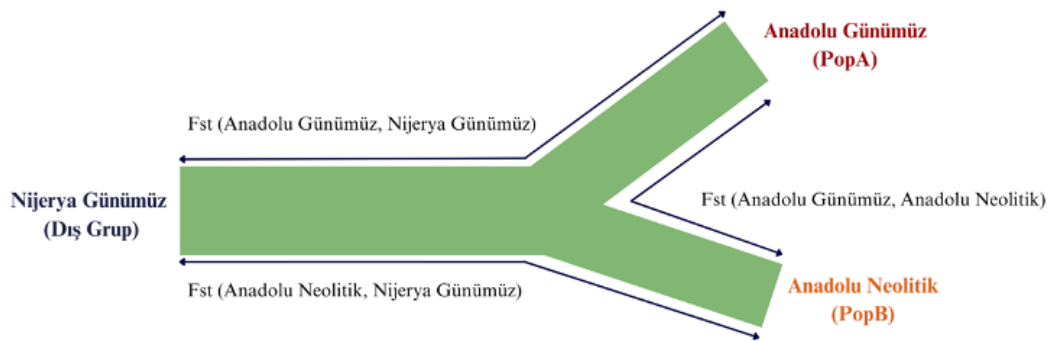
```
realSFS fst print pop1.pop2.fst.idx > pop1.pop2.txt
```

$F_{ST}$  tahmini, elde edilen A ve B değerleri R programlama dili yazılımı kullanarak A/B olarak orantılanarak hesaplanmıştır.

### 3.6. Popülasyon Dal İstatistiği Skorlarının Hesaplanması

Anadolu'da son 10.000 yılda pozitif seçilim altındaki bölgeleri tespit etmek için popülasyon dal istatistiği (PDI) (İng. Population branch statistics, PBS) kullanılmıştır (151). PDI analizinde üç popülasyondan oluşan bir set kullanır (POP1, POP2, POP3) ve bunların ikisi zamansal seçilimi araştırmak istediğimiz popülasyonlardan ve biri dış grup olarak atadığımız popülasyondan oluşur (POP1, POP2), POP3).

Bu çalışmada POP1'in günümüz Anadolu bireyleri (134), POP2'nin antik Anadolu bireyleri (Anadolu Neolitik ve Anadolu Neolitik Sonrası popülasyonları) ve POP3'ün günümüz Nijerya bireyleri (HGDP, Yoruba) (152) olduğu durumla ilgilenilmiştir.



**Şekil 3.5.**  $F_{ST}$  tahminlerinden popülasyon dal istatistiği skorlarının hesaplanması.



Neolitik dönemden günümüze seçilen popülasyonlarda her lokus için popülasyon dal istatistiği skorları aşağıdaki formül kullanılarak hesaplanmıştır:

$$PD\dot{I}_{POP1} = \frac{T_{POP1,POP2} + T_{POP1,POP3} - T_{POP2,POP3}}{2} \quad (3.4.)$$

Şekil 3.5.'te şematize edilen gösterimi bulunan T değerleri ise aşağıdaki formül kullanılarak Bölüm 3.4'te hesaplanan  $F_{ST}$  değerleri kullanılarak hesaplanmıştır:

$$T = -\log(1 - F_{ST}) \quad (3.5.)$$

Genel popülasyon dal istatistiği skorlarının dağılımının %99,5 yüzdilik diliminde bulunan varyantlara odaklanılmıştır.

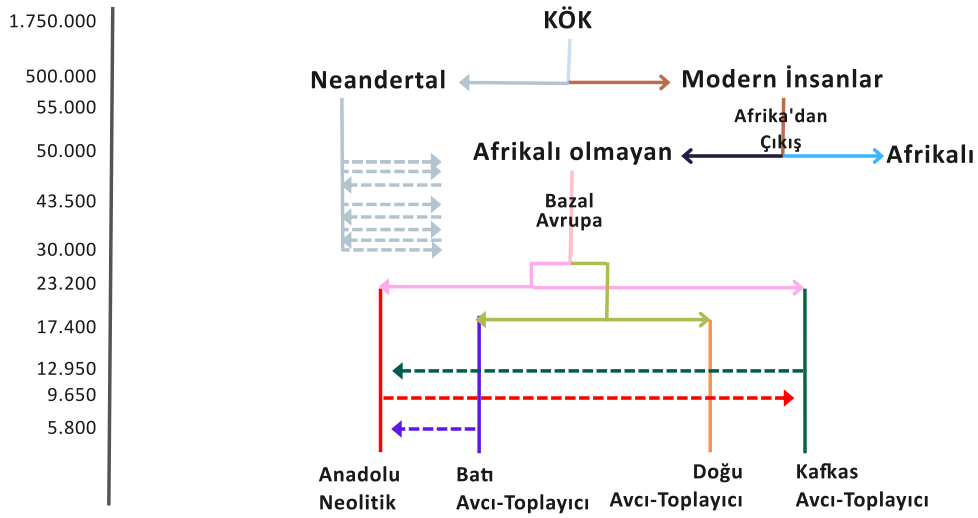
Her popülasyon çifti ve seçilen her lokus için popülasyon dal istatistiği skorları ANGSD (v.0.934) (141) ile htlib (versiyon 1.13; <http://www.htslib.org/>) kullanılarak hesaplanmıştır.  $F_{ST}$  ilk olarak polimorfizm verilerinden popülasyon farklılaşmasının bir ölçüsü olarak ANGSD'dan realSFS  $F_{ST}$  indeksi ile hesaplanmıştır.

SFS hesaplamak için gerekli olan olabilirlik dosyaları Bölüm 3.5.'te hesaplanmıştır. Ardından, realSFS fonksiyonu ile her popülasyon çifti için 2D-SFS hesaplanmıştır. Karşılaştırma için seçilen 2 popülasyon için hazırlanan ve bir dış grup popülasyonu için hazırlanan olmak üzere üç dosya, her popülasyon için popülasyon dal istatistiği çıktısı almak üzere RealSFS  $F_{ST}$  programının girdisi olarak kullanılmış ve 1 bazlık adımlarla 1 bazlık genom pencereleri “-win 1 -step 1” başına indeksi hesaplamak için "stats2" seçeneği kullanılmıştır.

### 3.7. İleri Zamanlı Genetik Benzetim

Obezite ve Tip 2 diyabetin günümüzdeki yaygınlığını nötral model altında incelemek ve bunu insan genomunun gerçekçi bir modelini oluşturarak yapabilmek için literatür taraması ile oluşturulan demografik modelden çıkarılan parametreler ileri zamanlı genetik benzetim aracı SLiM'e (153) uygulanmıştır.

Tüm simülasyonlar Wright-Fisher modeli altında SLiM (v.3.7) (153) ile yürütülmüştür. Her simülasyon, değişen rekombinasyon ve mutasyon oranları ile karakterize edilen 10 Mb uzunluğunda DNA dizilerinden oluşmuştur. Her simülasyon için,  $1 \times 10^{-8}$  rekombinasyon oranı ve  $1 \times 10^{-7}$  mutasyon oranı ile insan genomunda rastgele bir 10 Mb genomik pencerenin fiziksel koordinatları örneklenerek 10 Mb'lık DNA dizileri simüle edilmiştir (154, 155). Demografik model için parametreler literatür taramalarından elde edilmiştir (Şekil 3.6.). Parametreler sırasıyla aşağıdaki gibidir (Bkz. EK-3).



**Şekil 3.6.** İleri zamanlı genetik benzetim analizinde kullanılan demografik modelin şemasal gösterimi.

Simülasyonlar, önceki çalışmalarda kullanılan demografik modelleri, karışım ve ayrılma zamanlarının bir kombinasyonu kullanılarak yürütülmüştür. İlk olarak mutasyonlu simüle edilmiş genomların bir denge durumuna ulaşmasını sağlamak için 70.000 nesillik bir başlangıç (İng. burn-in) periyodu ile başlayarak, modern insan atalarının ve Neandertallerin basitleştirilmiş bir demografisini simüle edilmiştir (94). Başlangıç periyodunun uzunluğu simülasyonun doğasına göre seçilmektedir (156).

Neandertallerin ve modern insanların ayrılması 500.000 yıl öncesine ayarlanmıştır (157, 158), Neandertallerin ve modern insan atalarının popülasyon büyüklükleri (İng. Population size, N) sırasıyla 1.000 ve 10.000 olarak ayarlanmıştır

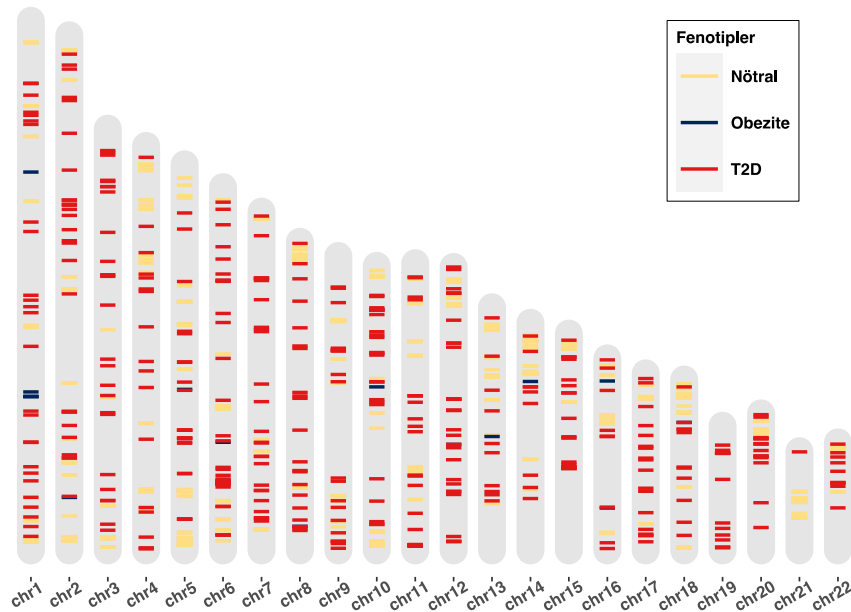
(94). Bařlangıç periyodu ařamasının sonunda, Afrikalı ve Afrikalı olmayan popülasyonların 55.000 yıl önce ayrılmasını simüle edilmiştir. Ayrılmadan sonra, Afrikalı olmayan popülasyon Afrika'dan çıkış (İng. Out of Africa) 5000 yıl süresince  $N=1861$  ile bir darboğaza (İng. Bottleneck) uğramıştır (93, 94). Günümüzden 50.000-30.000 yıl önceki aralıkta Neandertallerden Afrikalı olmayan popülasyona %3 oranında karışım simüle edilmiştir. Afrika dışı yayılımı simüle etmek için ilk olarak 43.500 yıl önce Afrikalı olmayan popülasyondan  $N=5000$  ile Bazal Avrupa ayrılması simüle edilmiştir (159). Neandertal-Afrikalı olmayan popülasyon karışımının bitiři ile Neandertal popülasyonu birey sayısı sıfıra getirilerek silinmiştir (93, 94). Bazal Avrupa popülasyonundan 23.200 yıl önce popülasyon büyüklükleri 10.000 olmak üzere Kafkas Avcı Toplayıcı (CHG) ve Anadolu Neolitik popülasyonunun atalarının ayrılması simüle edilmiştir (159). Bundan 5800 yıl sonra ise Bazal Avrupa popülasyonundan Batı Avcı Toplayıcı (WHG) ve Doęu Avcı Toplayıcı (EHG) popülasyonlarının ayrılması simüle edilmiştir (159).

Anadolu Neolitik popülasyonuna 12950 yıl önce Kafkas Avcı Toplayıcı (160), 5800 yıl önce Batı Avcı Toplayıcı karışımı (159) simüle edilmiştir. Kafkas Avcı Toplayıcı popülasyonuna 9650 yıl önce Anadolu Neolitik karışımı (160) simüle edilmiştir.

## 4. BULGULAR

### 4.1. Fenotip ve Nötral Bölgelerle İlişkili Hedeflenen TNP Sayıları ve Genotiplenme Oranları

Fenotiplerle ilişkili varyantlar Bölüm 3.1.'de belirtilen şekilde NHGRI-EBI GWAS Catalog'dan toplanmış ve Tablo 3.1'de belirtilen filtrelemelerden geçtikten sonra kalan TNP'lerin kromozomlar üzerinde dağılımı Şekil 4.1.'de gösterilmiştir. Toplam 1432 TNP'nin 1420'si Neolitik dönem bireylerinden (Obezite 18, Tip 2 diyabet 402, nötral 1000), 1422'si Neolitik sonrası dönem bireylerinden (Obezite 18, Tip 2 diyabet 404, nötral 1000), 1423'ü ise günümüz bireylerinden başarıyla çağrılmıştır (Obezite 18, Tip 2 diyabet 405, nötral 1000). Her bir TNP için en az 5 bireyde dizilenmiş olma şartı eklendiğinde Neolitik dönemde 1417 (Obezite 17, Tip 2 diyabet 400, nötral 1000), Neolitik sonrası dönemde 1418 (Obezite 17, Tip 2 diyabet 402, nötral 999) ve günümüzde 1421 TNP (Obezite 18, Tip 2 diyabet 403, nötral 1000) kalmıştır (Şekil 4.1.).



**Şekil 4.1.** Seçilen 18 Obezite, 415 Tip 2 diyabet (T2D) ve 1000 nötral TNP'nin kromozomlar üzerinde dağılımının gösterimi.

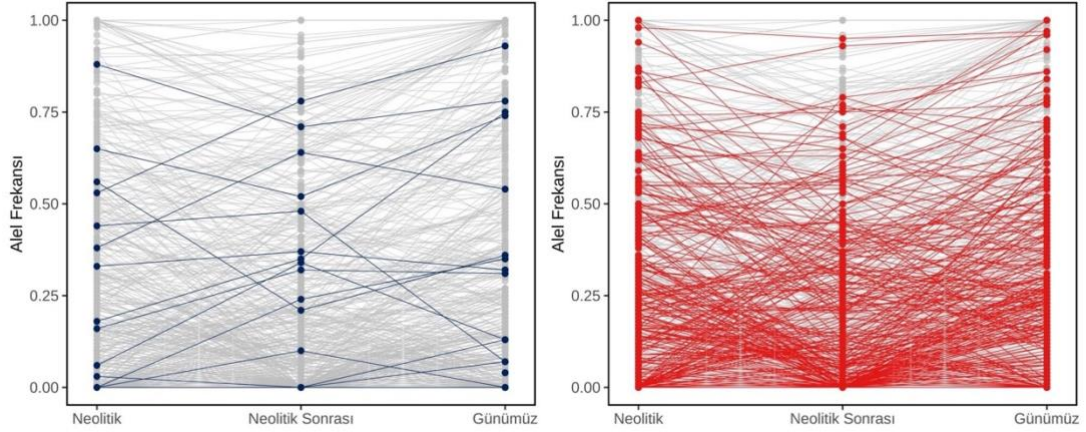
## 4.2. Alel Frekans Yörüngelerinin Belirlenmesi

Anadolu'da son 10.000 yılı tarihsel olarak üç döneme ayrılarak varyantların risk alel frekansları Bölüm 3.4.'te açıklandığı gibi hesaplanmış ve değişimleri incelenmiştir.

Anadolu Neolitik (n= 81), Anadolu Neolitik sonrası (n= 238) ve günümüz Anadolu (n=16) popülasyonlarında Obezite, Tip 2 diyabet ve nötral varyantların alel frekansı hesaplanmıştır. Antik popülasyonlarda ortalama okuma derinlikleri düşük olduğu için Bölüm 3.3.'te açıklanan maksimum olabilirlik temelli alel frekans hesaplama yöntemi tercih edilmiştir (Tablo 4.1.).

**Tablo 4.1.** Veri setine dahil edilen Anadolu yayınlanmış antik genom örneklerinin veri üretilme yöntemlerine göre örneklem büyüklükleri, ortalama ve medyan okuma derinlikleri.

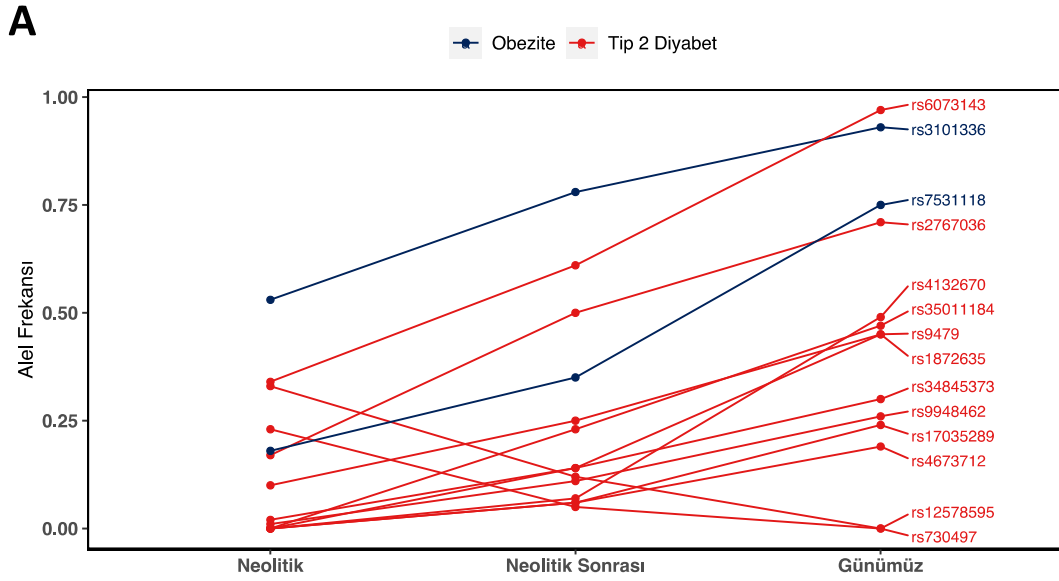
Veri Üretilme Yöntemleri	Neolitik			Neolitik Sonrası		
	Örneklem Büyüklüğü	Ortalama Okuma Derinliği	Medyan Okuma Derinliği	Örneklem Büyüklüğü	Ortalama Okuma Derinliği	Medyan Okuma Derinliği
1240K	36	0,17	0,1	218	0,1	0,07
Rastgele Dizileme	37	1,31	0,07	20	1,02	0,7
Tüm Genom Yakalama + Rastgele Dizileme	8	1,06	0,34	-	-	-



**Şekil 4.2.** Obezite (n=18, mavi) ve Tip 2 diyabet (n= 404, kırmızı) ile ilişkili varyantın Anadolu Neolitik, Anadolu Neolitik sonrası ve günümüz Anadolu'daki risk alel frekanslarının dağılımı. Arka planda gri renk ile gösterilen çizgiler 1000 adet nötral varyantın minör alel frekansının dağılımını göstermektedir.

Yukarıdaki çizgi grafiğinde Obezite ve Tip 2 diyabet TNP'leri için risk alel frekansının, nötral bölgelerden seçilmiş TNP'ler için minör alel frekansının dağılımı görülmektedir. Anadolu Neolitikten günümüze frekansı düzenli olarak artış gösteren 2 Obezite, 10 Tip 2 diyabet; Neolitikten Neolitik sonrası döneme artış sonrasında günümüze azalış gösteren 2 Obezite, 17 Tip 2 diyabet; Neolitikten Neolitik sonrası döneme azalış sonrasında günümüze artış gösteren 48 Tip 2 diyabet; Neolitikten günümüze düzenli azalış gösteren 2 Tip 2 diyabet ve frekansı Neolitikten günümüze değişim göstermeyen 2 Obezite ve 106 Tip 2 diyabet varyantı bulunmuştur.

Risk alel frekansı Anadolu Neolitikten günümüze düzenli artış gösteren Obezite ile ilişkili 2 ve Tip 2 diyabet ile ilişkili 10 varyantın ve düzenli azalış gösteren Tip 2 diyabetle ilişkili 2 varyantın frekans dağılımları ve diğer fenotiplerle ilişkileri Şekil 4.3.'te gösterilmektedir. Alel frekansı artış gösteren varyantların ilişkili oldukları diğer fenotipler arasında beslenme ve metabolizma ile ilişkili olduğu bilinen fenotipler bulunmaktadır: Vücut kitle indeksi, alkol tüketimi, bel çevresi, boy uzunluğu, kalça çevresi, tat beğeni ölçümü. Alel frekans artışı gözlenen



**B**

Fenotipler	Neolitikten Günümüze Risk Alel Frekans Değişimi	
	Artış	Azalış
Tip 2 diyabet	9	2
Vücut kitle indeksi	3	
Obezite	2	
Zeka	2	
Kolorektal kanser	1	
Alkol tüketimi	1	
Bel çevresi	1	
Boy Uzunluğu	1	
Eritrosit sayısı	1	
Kalça çevresi	1	
Lökosit sayısı	1	
Menarş	1	
Nörolojik bozukluklar	1	
Retinopati x Tip 2 diyabet etkileşimi	1	
Tat beğeni ölçümü	1	
Glikoz seviyeleri		1
Hemoglobin A1c seviyeleri		1
İlaç kullanımı		1
<b>Toplam</b>	<b>27</b>	<b>5</b>

**Şekil 4.3.** Anadolu Neolitikten günümüze frekansı düzenli olarak artış ve azalış gösteren Obezite ve Tip 2 diyabet ile ilişkili varyantlar. A panelinde risk alel frekans değişimlerinin çizgi grafiği ile gösterilmiştir. B panelinde GWAS Catalog'da anlamlı ilişkide raporlanan ( $p$ -değeri  $<5 \times 10^{-8}$ ) fenotip bilgilerinin artan ve azalana göre sayıca dağılımının listelenmesi.

Risk alel frekansı Anadolu Neolitikten günümüze düzenli artış gösteren Obezite ile ilişkili 2 ve Tip 2 diyabet ile ilişkili 10 varyantın ve düzenli azalış gösteren Tip 2 diyabetle ilişkili 2 varyantın ilişkili olduğu genlerin rol oynadıkları sinyal yolları,

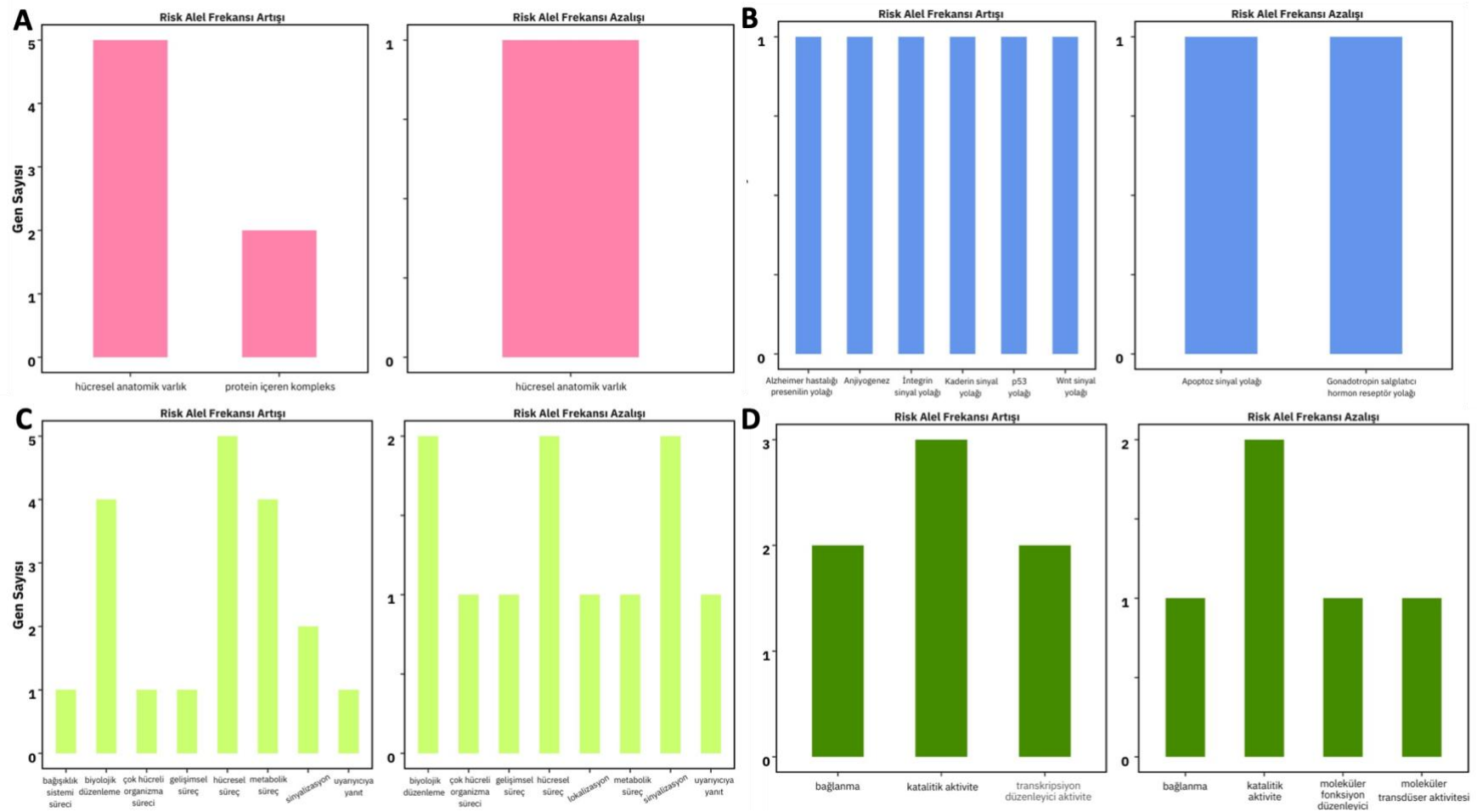
moleküler fonksiyon, hüresel bileşen ve biyolojik süreçlerin analizi için PANTHER (161) kullanılmıştır. Şekil 4.4.'te fonksiyonel sınıflandırma sonuçları gösterilmektedir.

**Tablo 4.2.** Risk alel frekanslarında değişim gözlenen varyantların yolak analizleri.

Risk Alel Frekansının Değişimi	Varyant	Fenotip	Gen	Panther Protein Sınıfı	Metabolik Yolak
Artanlar	rs7531118	Obezite	-	-	-
	rs9948462	Tip 2 Diyabet	<i>LAMA1</i>	hücre yapışma molekülü(PC00069)	İntegrin sinyal yolağı
	rs2767036	Tip 2 Diyabet	<i>PDHX</i>	asetiltransferaz(PC00038)	-
	rs6073143	Tip 2 Diyabet	<i>IFT52</i>	-	-
	rs1872635	Tip 2 Diyabet	<i>SMUG1</i>	DNA glikosilaz(PC00010)	-
	rs4673712	Tip 2 Diyabet	<i>IKZF2</i>	C2H2 çinko parmak transkripsiyon faktörü (PC00248)	-
	rs9479	Tip 2 Diyabet	<i>PML</i>	ubikitin-protein ligaz(PC00234)	p53 yolağı
	rs17035289	Tip 2 Diyabet	<i>TET2</i>	-	-
	rs3101336	Obezite	<i>NEGR1</i>	immünoglobulin(PC00123)	-
	rs34845373	Tip 2 Diyabet	<i>DTNB</i>	ubikitin-protein ligaz(PC00234)	-
rs4132670, rs35011184	Tip 2 Diyabet	<i>TCF7L2</i>	DNA-bağlayıcı transkripsiyon faktörü(PC00218)	Alzheimer hastalığı-presenilin yolağı, Wnt sinyal yolağı, Kadherin sinyal yolağı, Anjiyogenez	
Azalanlar	rs12578595	Tip 2 Diyabet	<i>PTH1H</i>	peptit hormonu (PC00179)	-
	rs730497	Tip 2 Diyabet	<i>GCK</i>	reseptör olmayan serin/treonin protein kinaz(PC00167)	Gonadotropin salgılatıcı hormon reseptör yolağı, Apoptoz sinyal yolağı

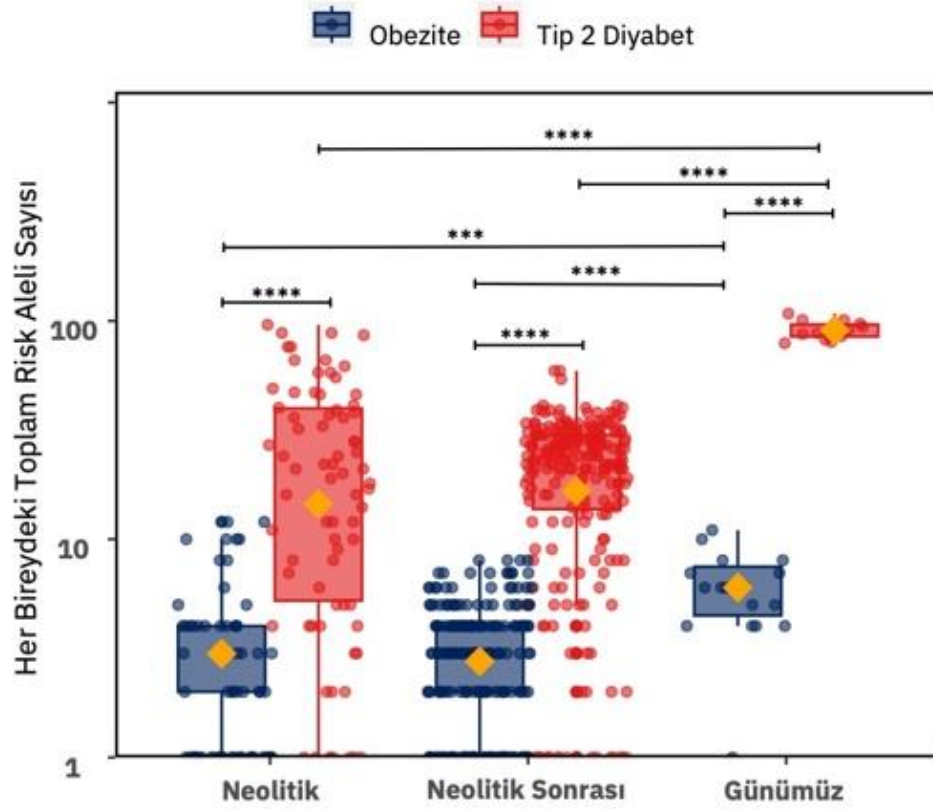
Tablo 4.2.'de risk alel değişimlerine göre ilişkili genlerin yolak analizleri sonuçları listelenmektedir. Risk alel frekansının Anadolu'da Neolitikten günümüze artış gösterdiği varyantlarda *LAMA1* geni integrin sinyal yolağı ile, *PML* geni p53 yolağı ile ve *TCF7L2* geni Wnt sinyal yolağı, anjiyogenez, Alzheimer hastalığı-presenilin yolağı ve kaderin sinyal yolağıyla ilişkilidir. Risk alel frekansının Anadolu'da Neolitikten günümüze azalma gösterdiği varyantlarda *GCK* geni apoptoz sinyal yolağı ve Gonadotropin salgılatıcı hormon reseptör yolağında rol almaktadır.





**Şekil 4.4.** Risk alel frekansı Neolitikten günümüze düzenli artış ve azalış gösteren Obezite ve Tip 2 diyabetle ilişkili TNP'lerin ilişkili olduğu genlerin gen ontolojisi analizi. (A) Hücrel Bileşenler (B) Sinyal Yolakları (C) Biyolojik Süreçler (D) Moleküler Fonksiyon.

Antik ve günümüz Anadolu bireyleri üzerindeki risk alel yükünün dağılımı incelenmek istenmiştir. Bu amaçla Bölüm 3.4.'te alel frekans hesaplaması için kullanılan risk alel sahip sekans sayılarından yararlanılmıştır. Her lokus için T toplam sekans sayısı, R risk alele sahip sekans sayısını ifade ederken risk alel sayısı ( $R/T > 0.5$ ) olan varyantlarda risk alel sayısı 1 olarak kabul edilerek hesaplanmıştır.



**Şekil 4.5.** Anadolu antik ve günümüz dönemlerinden seçilen bireylerin taşıdığı Obezite ve Tip 2 diyabet ile ilişkili toplam risk alel sayılarının kutu grafiği ile gösterimi. Y eksenini log<sub>10</sub> tabanına göre ölçeklendirilmiştir.

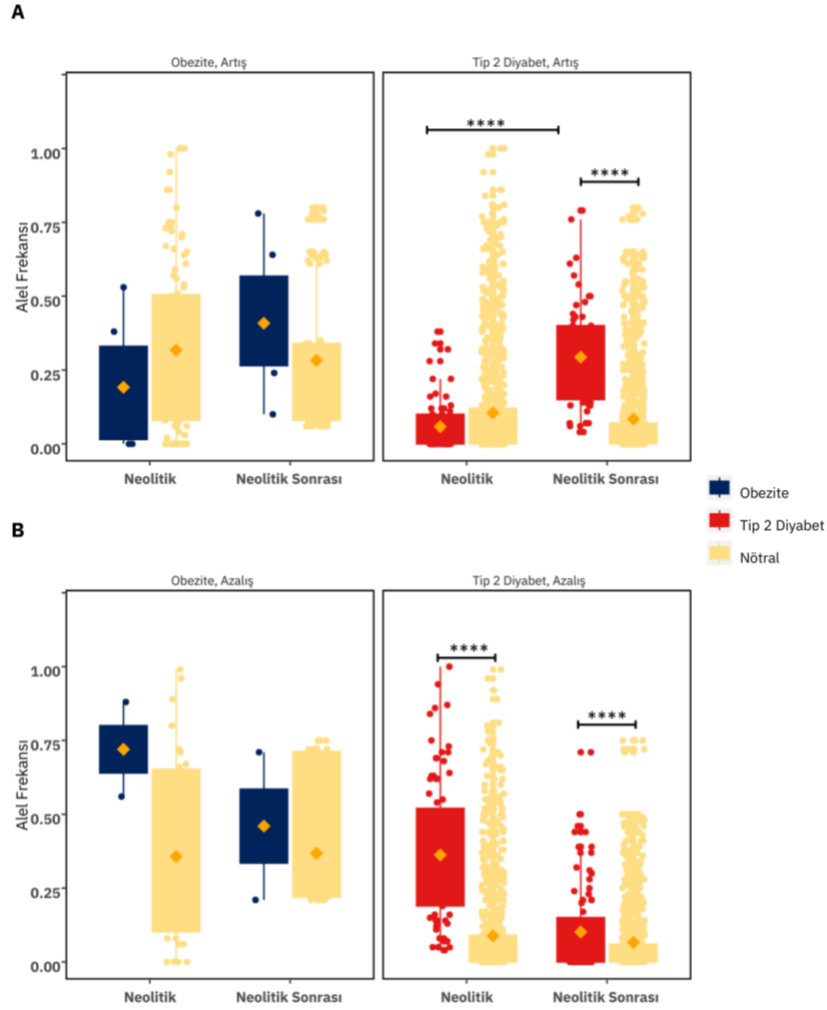
İkili dönemler arası alel frekansı karşılaştırılırken nötral varyantların minör alel frekansları ile Obezite ve Tip 2 diyabet ile ilişkili varyantların risk alel frekansları karşılaştırılmıştır. Obezite, Tip 2 diyabet ile ilişkili varyantlardan dönemler arasında risk alel frekansı artış ve azalış gösterenler karşılaştırmaya dahil edilmiştir. Dönemler arası risk alel frekansı artış ve azalış gösteren Obezite ve Tip 2 diyabet varyantlarının sayıları Tablo 4.3.'te listelenmiştir. Karşılaştırma için alel frekans değerleri sıfır ile bir arasında 0,05'erli gruplara bölünerek atanmış ve frekans artışı gösterdikleri dönemde aynı gruba atanan nötral varyantların ortalama alel frekansı ile Obezite ve Tip 2 diyabet varyantlarının ortalama risk alel frekansları karşılaştırılmıştır.

**Tablo 4.3.** İkili dönemler arası karşılaştırmalarda ilgili fenotiple ilişkili alellerin frekansının artışı ve azalışına göre TNP sayıları listelenmiştir.

Fenotipler	Neolitik - Neolitik Sonrası		Neolitik Sonrası - Günümüz		Neolitik - Günümüz	
	Frekans Artışı	Frekans Azalışı	Frekans Artışı	Frekans Azalışı	Frekans Artışı	Frekans Azalışı
Obezite	6	2	5	3	4	2
Tip 2 Diyabet	59	83	121	35	107	56

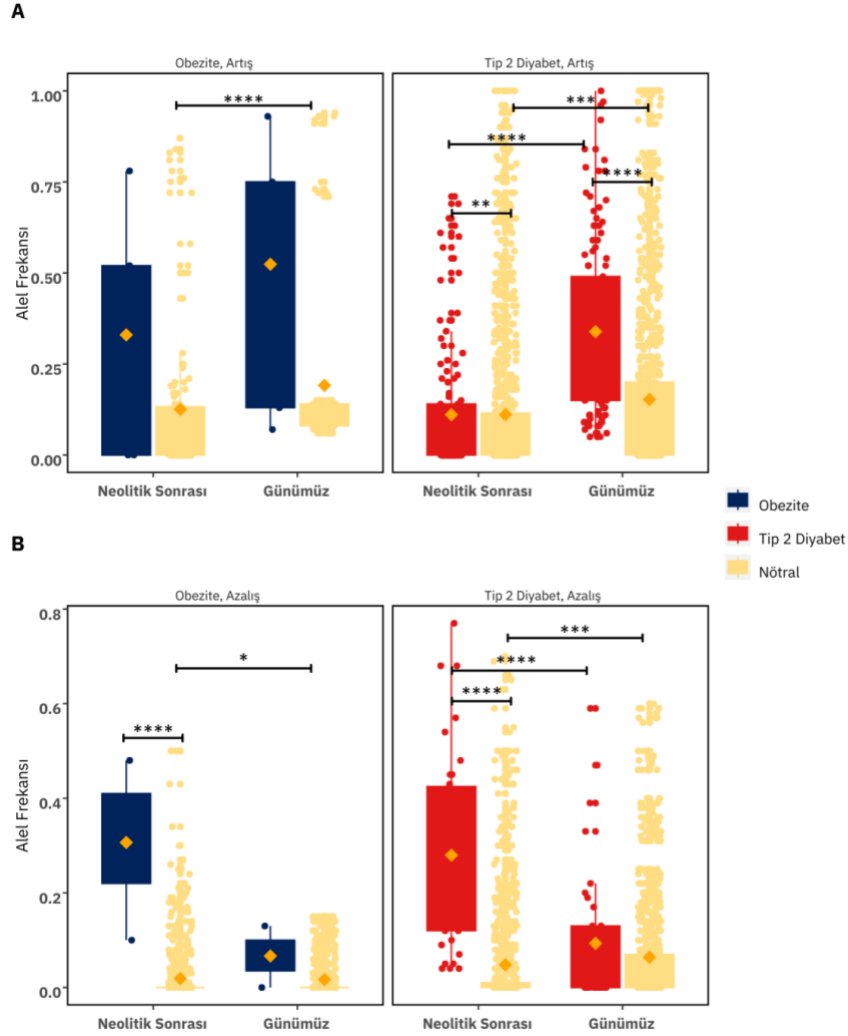
Alel frekansı değişikliklerinin anlamlılığını test etmek için non-parametrik Wilcoxon testi kullanılmış ve "rstatix" paket fonksiyonları (v0.7.2) kullanılarak tüm özellikler için p değerlerine "Bonferroni" çoklu test düzeltme yöntemi uygulanmıştır. Sonrasında fenotiple ilişkili alellerin aynı dönemdeki nötral alellerle frekanslarını karşılaştırmak için non-parametrik Mann-Whitney U testi kullanılmış ve "rstatix" paket fonksiyonları kullanılarak tüm özellikler için p değerlerine "Bonferroni" çoklu test düzeltme yöntemi uygulanmıştır.

Neolitikten Neolitik sonrası döneme Tip 2 diyabetle ilişkili artış gösteren risk alellerinin ortalama frekans değişimleri frekans artışı gösterdikleri dönemde aynı frekanstaki nötral varyantların ortalama alel frekans değişimi ile kıyaslandığında anlamlı farklılık göstermektedir (Şekil 4.6.).

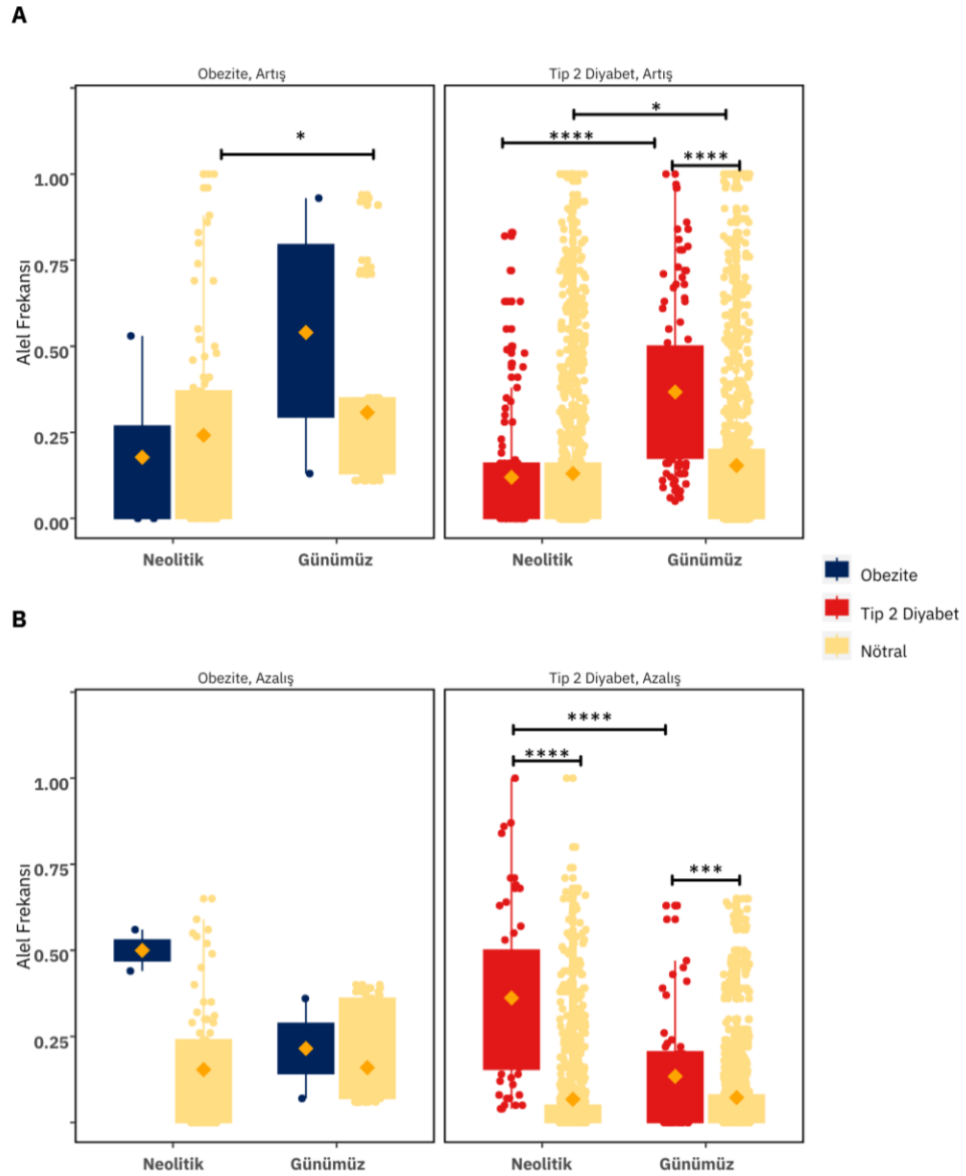


**Şekil 4.6.** Neolitikten Neolitik sonrası döneme risk alel frekans artışı (sol panel) ve azalışı (sağ panel) gözlenen Obezite (mavi) ve Tip 2 diyabet (kırmızı) varyantlarının nötral varyantların alel frekansları (sarı) ile karşılaştırılması. Obezite ve Tip 2 diyabet ile nötral alel frekanslarının ortalamaları Mann-Whitney U testi ile karşılaştırılmış ve Bonferroni düzeltmesi sonucu p-değeri anlamlılıklarına göre anlamlı sonuçlar ( $p$ -değeri  $< 0,05$ ) kutu grafikler üzerinde gösterilmiştir. Obezite, Tip 2 diyabet ve nötral alel frekanslarının dönemler arası değişimi ise Wilcoxon testi ile karşılaştırılmış Bonferroni düzeltmesi sonucu p-değeri anlamlılıklarına göre anlamlı sonuçlar ( $p$ -değeri  $< 0,05$ ) kutu grafikler üzerinde gösterilmiştir.

Neolitikten günümüze Tip 2 diyabetle ilişkili azalış gösteren risk alellerinin ortalama frekans değişimleri frekansın azalma gösterdiği dönemde aynı frekanstaki nötral varyantların ortalama alel frekans değişimi ile kıyaslandığında anlamlı farklılık göstermektedir (Şekil 4.8.).

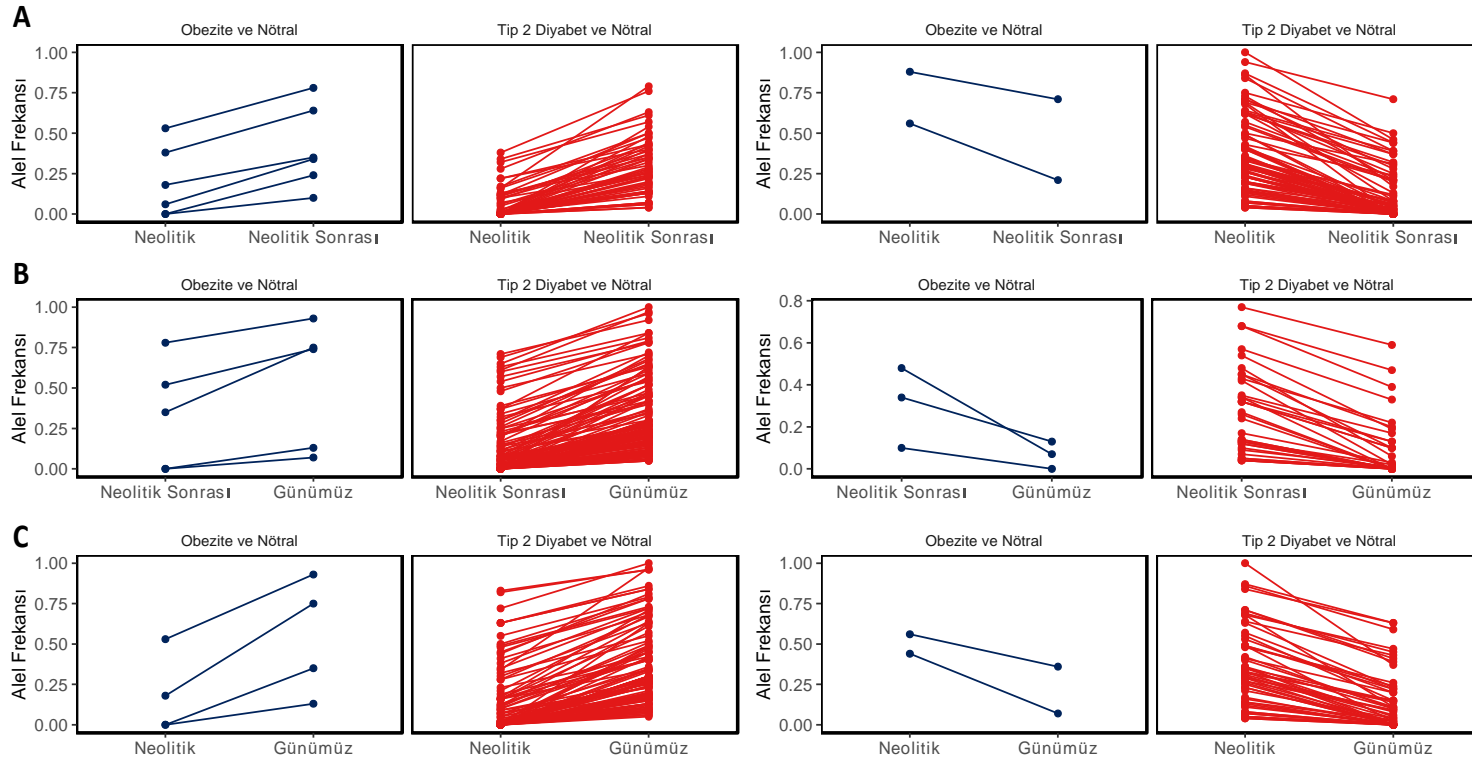


**Şekil 4.7.** Neolitik sonrası dönemden günümüze risk alel frekans artışı (sol panel) ve azalışı (sağ panel) gözlenen Obezite (mavi) ve Tip 2 diyabet (kırmızı) varyantlarının nötral varyantların alel frekansları (sarı) ile karşılaştırılması. Obezite ve Tip 2 diyabet ile nötral alel frekanslarının ortalamaları Mann-Whitney U testi ile karşılaştırılmış ve Bonferroni düzeltmesi sonucu p-değeri anlamlılıklarına göre anlamlı sonuçlar (p-değeri < 0,05) kutu grafikler üzerinde gösterilmiştir. Obezite, Tip 2 diyabet ve nötral alel frekanslarının dönemler arası değişimi ise Wilcoxon testi ile karşılaştırılmış Bonferroni düzeltmesi sonucu p-değeri anlamlılıklarına göre anlamlı sonuçlar (p-değeri <0,05) kutu grafikler üzerinde gösterilmiştir.



**Şekil 4.8.** Neolitikten günümüze risk allel frekans artışı (sol panel) ve azalışı (sağ panel) gözlenen Obezite (mavi) ve Tip 2 diyabet (kırmızı) varyantlarının nötral varyantların allel frekansları (sarı) ile karşılaştırılması. Obezite ve Tip 2 diyabet ile nötral allel frekanslarının ortalamaları Mann-Whitney U testi ile karşılaştırılmış ve Bonferroni düzeltmesi sonucu p-değeri anlamlılıklarına göre anlamlı sonuçlar ( $p$ -değeri  $< 0,05$ ) kutu grafikler üzerinde gösterilmiştir. Obezite, Tip 2 diyabet ve nötral allel frekanslarının dönemler arası değişimi ise Wilcoxon testi ile karşılaştırılmış Bonferroni düzeltmesi sonucu p-değeri anlamlılıklarına göre anlamlı sonuçlar ( $p$ -değeri  $< 0,05$ ) kutu grafikler üzerinde gösterilmiştir.

Anadolu'da Neolitikten günümüze alel frekans deęişikliklerinin yönlü bir deęişimi işaret edip etmedięini anlamak için frekansa dayalı seçim testleri ile her popülasyon çifti test edilmiştir. Bu tez çalışmasında fenotipi ortaya çıkarma olasılıęını artıran alellerin yaygınlıklarının günümüzde daha yüksek olduęu varsayılarak bu artışın pozitif seçim kaynaklı olup olmadığını test edip karşılaştırılan popülasyon çiftleri içinde varyantların zaman içindeki yönlü deęişimi de anlatılmak istendięi için  $F_{ST}$  istatistięine ek olarak popülasyon dal istatistięi de uygulanmıştır.

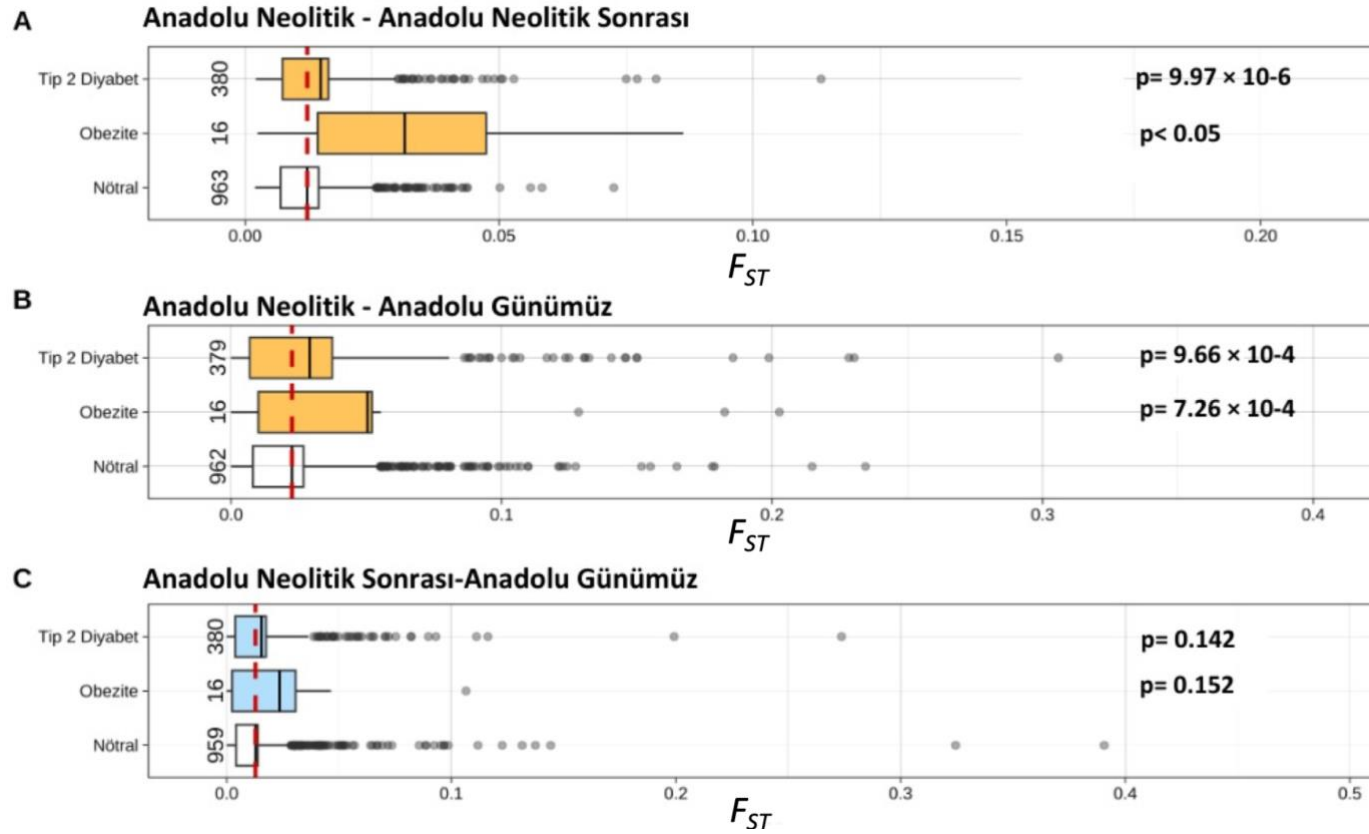


**Şekil 4.9.** Obezite ve Tip 2 diyabet ile ilişkili varyantların risk alelleri için dönem çiftleri arasında alel frekansы değişiklikleri. A panelinde Neolitik-Neolitik sonrası dönem çiftleri, B panelinde Neolitik sonrası günümüz dönem çiftleri, C panelinde Neolitik ve günümüz dönem çiftleri incelenmektedir. Sol panellerde alel frekansının artış gösterdiği, sağ panellerde alel frekansının azalış gösterdiği Obezite (mavi), Tip 2 diyabet (kırmızı) ve nötral (gri) ile ilişkili varyantların alel frekanslarının dağılımları çizgi grafiği ile gösterilmektedir.



### 4.3. Seçilim Analizleri Sonuçları

$F_{ST}$  istatistiğini analiz edilirken amaçlanan Tip 2 diyabet ve Obezite TNP'lerinin nötral bölgelerdeki TNP'lerden hangi dönem çiftleri arasında önemli ölçüde farklı olduğunu bulmak olmuştur. Popülasyonların farklılaşma ölçüsü olarak sıkça kullanılan  $F_{ST}$  değerleri Bölüm 3.4'te açıklandığı gibi ANGSD kullanılarak her bir TNP için her popülasyon çifti arasında ölçülmüştür. Daha sonra, alel frekansında önemli değişiklikler olup olmadığını test etmek için her fenotiple ilişkili TNP'lerin  $F_{ST}$  değerlerinin dağılımını nötr bölgelerdeki TNP'lerle R programlama dili (versiyon 4.3), <https://www.r-project.org/>) ile ANOVA testini takip eden post-hoc Tukey's testi kullanılarak karşılaştırılmıştır. Obezite ve Tip 2 diyabet varyantlarının popülasyonlar arasında nötral bölgeden farklılık göstermesinin evrimsel arka planı demografik yollarla popülasyonlar arası karışımın bir sonucu veya genetik sürüklenme olarak açıklanabilmektedir.



**Şekil 4.10.** Fenotip ile ilişkili ve nötral TNP'lerin  $F_{ST}$  dağılımları. Kırmızı kesikli dikey çizgi, nötral TNP'lerin ortalama  $F_{ST}$  'sini göstermektedir. ANOVA testini takiben uygulanan Tukey testine göre nötral TNP'lerden farklılık gösteren ( $p$  değeri $<0.05$ ) fenotipler turuncu, farklılık göstermeyen ( $p$  değeri $> 0.05$ ) fenotipler mavi ile gösterilmiştir. Kutu grafiklerinin sağ tarafında  $p$  değerlerini göstermekte iken sol tarafında TNP sayılarını temsil etmektedir.

Şekil 4.10.'da fenotiplerle ilişkili TNP'lerin ve nötr bölgelerdeki TNP'lerin  $F_{ST}$  dağılımları aynı şekilde sırasıyla mavi ve şeffaf renkli kutu grafikleriyle gösterilmiştir. ANOVA testini takiben uygulanan Tukey testine göre nötral TNP'lerden önemli ölçüde farklılık gösteren (p-değeri<0,05) fenotipler turuncu olarak gösterilmiştir. Karşılaştırdığımız dönemler arasında  $F_{ST}$  değerlerinin dağılımları nötral bölgelerden anlamlı ölçüde (p-değeri< 0,05) farklılık gösterenler: Anadolu Neolitik ve Anadolu Post Neolitik arasında nötral bölge için ortalama  $F_{ST}$  değeri 0,01223509 iken Tip 2 diyabet için 0,01488753 (p-değeri =  $9.97 \times 10^{-6}$ ) ve Obezite için 0,03142205 (p-değeri <0,05), Anadolu Neolitik ve Anadolu Günümüz arasında nötral bölge için ortalama  $F_{ST}$  değeri 0,0224476 iken, Tip 2 diyabet için 0,02903933 (p-değeri = 0,000966) ve Obezite için 0,05041716 (p-değeri = 0,000726). Nötral bölgelerden daha yüksek  $F_{ST}$  değerleri, alel frekansındaki değişikliklerin adaptif olduğunu fikri ile uyumludur (162, 163). Sonuçlarımız bu fenotipler için karşılaştırma yapılan popülasyon çiftlerini farklılaştıran poligenik adaptasyon sinyali olabileceğini işaret etmektedir.

Çalışmanın sonraki kısımlarında dönemler arasında fenotiple ilişkili TNP'lerin  $F_{ST}$  değerleri nötral bölgelere kıyasla farklılık gösterdiği Anadolu Neolitik-Anadolu Post Neolitik ve Anadolu Neolitik-Anadolu Günümüz popülasyon karşılaştırmalarına odaklanılmıştır.

Poligenik adaptasyon sinyali görülen fenotiplerde risk alel frekansının arttığı gözlemlenen varyantların her birinin  $F_{ST}$  değerlerini frekansın artış gösterdiği dönem için aynı alel frekansına sahip nötral varyantların  $F_{ST}$  değerleri ile karşılaştırdık. Her dönem çifti arasında Obezite ve Tip 2 diyabet ile ilişkili varyantlarla nötral varyantlar arasındaki  $F_{ST}$  değişikliklerinin önemini test etmek için Student's t-testi kullanılmıştır. İstatistiksel olarak anlamlı farklılık (p-değeri <0,05) gösteren varyantlar içinden alel frekans artış gözlenenler Tablo 4.4.'te listelenmiştir.

**Tablo 4.4.** Risk alel frekansı artış gösteren Obezite ve Tip 2 diyabet ile ilişkili varyantların  $F_{ST}$  değerlerinin nötral varyantların  $F_{ST}$  değerleri ile karşılaştırmasının listesi.

TNP	Fenotip	Lokasyon	Risk Aleli	Nötral TNP Sayısı	p değeri (t test)	Karşılaştırılan Popülasyonlar	Gen İsmi	Gen Türü	Neolitik RAF	Neolitik Sonrası RAF	Günümüz RAF
rs2112347	Obezite	5:75719417	T	5	0.0000453	Anadolu Neolitik - Anadolu Günümüz	SLC25A5P9 - POC5	psödogen - protein kodlayan	0	0,24	0,35
rs4132670	Tip 2 Diyabet	10:113008012	A	9	0.01264257	Anadolu Neolitik - Anadolu Neolitik Sonrası	TCF7L2	protein kodlayan	0	0,07	0,49
rs10993072	Tip 2 Diyabet	9:94152720	C	6	0.0000276	Anadolu Neolitik - Anadolu Günümüz	MIRLET7A1HG - PTPDC1	lncRNA - protein kodlayan	0	0	0,15
rs11181613	Tip 2 Diyabet	12:42652647	C	11	0.00052164	Anadolu Neolitik - Anadolu Günümüz	LINC02451	lncRNA	0	0	0,08
rs12681990	Tip 2 Diyabet	8:37001668	C	10	0.00658774	Anadolu Neolitik - Anadolu Günümüz	SMARCE1P4 - KCNU1	psödogen - protein kodlayan	0	0,48	0,06
rs17035289	Tip 2 Diyabet	4:105127134	C	7	0.01275505	Anadolu Neolitik - Anadolu Günümüz	TET2 - RNU6 - 351P	protein kodlayan - snRNA	0	0,06	0,24
rs17684074	Tip 2 Diyabet	18:57008153	G	6	3.72E-08	Anadolu Neolitik - Anadolu Günümüz	WDR7	protein kodlayan	0,02	0,06	0,15
rs1922879	Tip 2 Diyabet	7:147961447	G	5	0.00117829	Anadolu Neolitik - Anadolu Günümüz	CNTNAP2	protein kodlayan	0	0,01	0,35
rs2240716	Tip 2 Diyabet	22:19982173	C	4	0.02776913	Anadolu Neolitik - Anadolu Günümüz	ARVCF	protein kodlayan	0	0,23	0,18
rs2820444	Tip 2 Diyabet	1:219568478	G	4	1.03E-09	Anadolu Neolitik - Anadolu Günümüz	LYPLAL1 - AS1 - ZC3H11B	- - protein kodlayan	0	0	0,3
rs34845373	Tip 2 Diyabet	2:25412902	G	4	0.00027164	Anadolu Neolitik - Anadolu Günümüz	DTNB	protein kodlayan	0	0,14	0,3
rs4673712	Tip 2 Diyabet	2:212964997	C	5	0.0029816	Anadolu Neolitik - Anadolu Günümüz	MIR4776 - 1 - IKZF2	miRNA - protein kodlayan	0	0,06	0,19
				7	0.00746172	Anadolu Neolitik - Anadolu Neolitik Sonrası					
rs510062	Tip 2 Diyabet	8:96126510	G	4	0.00000284	Anadolu Neolitik - Anadolu Günümüz	GAPDHP30 - GDF6	psödogen - protein kodlayan	0	0,28	0,3
				4	0.00000013	Anadolu Neolitik - Anadolu Neolitik Sonrası					
rs55812705	Tip 2 Diyabet	6:111417590	C	10	0.02868825	Anadolu Neolitik - Anadolu Günümüz	MFSD4B - REV3L	protein kodlayan - protein kodlayan	0	0	0,06
rs59489841	Tip 2 Diyabet	3:173992905	C	4	0.000000132	Anadolu Neolitik - Anadolu Günümüz	NLGN1	protein kodlayan	0	0	0,3
rs60573766	Tip 2 Diyabet	1:20361859	C	11	0.00954321	Anadolu Neolitik - Anadolu Günümüz	LINC01141	lncRNA	0	0	0,08
rs7029718	Tip 2 Diyabet	9:23358497	G	4	0.00000131	Anadolu Neolitik - Anadolu Günümüz	-	-	0	0,16	0,3
				6	0.000000847	Anadolu Neolitik - Anadolu Neolitik Sonrası					
rs73184014	Tip 2 Diyabet	7:104875827	G	5	0.0254681	Anadolu Neolitik - Anadolu Neolitik Sonrası	LHFPL3	protein kodlayan	0	0,19	0,16
rs7330796	Tip 2 Diyabet	13:75324027	T	17	0.03033099	Anadolu Neolitik - Anadolu Günümüz	TBC1D4	protein kodlayan	0,72	0,65	1
rs7645613	Tip 2 Diyabet	3:115344825	C	5	0.02353322	Anadolu Neolitik - Anadolu Günümüz	EIF4E2P2	psödogen	0	0	0,19
rs7660000	Tip 2 Diyabet	4:88830707	C	4	0.03638198	Anadolu Neolitik - Anadolu Günümüz	FAM13A	protein kodlayan	0,01	0,13	0,17
				6	0.01579997	Anadolu Neolitik - Anadolu Neolitik Sonrası					
rs8071043	Tip 2 Diyabet	17:4085157	C	5	0.03561473	Anadolu Neolitik - Anadolu Günümüz	ZZEF1	protein kodlayan	0	0,54	0,19
rs9385400	Tip 2 Diyabet	6:126443044	T	7	0.0000906	Anadolu Neolitik - Anadolu Neolitik Sonrası	MIR588 - CENPW	miRNA - protein kodlayan	0,12	0,5	0,61

TNP'lerle ilişkili genomik lokasyonu, risk aleli, risk alel frekansı, varyanta en yakın genlerin isimleri, genlerin türleri listelenmiştir. Risk alel frekansı, RAF.

Risk alel frekansı Anadolu Neolitik ile Anadolu Neolitik Sonrası dönemleri arasında artış gösteren ve  $F_{ST}$  karşılaştırma yaklaşımını kullanarak nötral lokuslardan farklı olan 8 TNP bulunmuştur. Bu 8 varyant Tip 2 diyabet ile ilişkilidir ve varyantların genomik lokasyon olarak en yakın oldukları 4 gen Tablo 4.4'te listelenmiştir. Bu genlerden yedisi protein kodlayan, ikisi miRNA, kalan üç tanesinin ikisi psödogen ve snRNA'dır. rs7029718 varyantı ile ilişkili gen bulunmamaktadır.

Risk alel frekansı Anadolu Neolitik ile Anadolu Günümüz dönemleri arasında artış gösteren ve  $F_{ST}$  karşılaştırma yaklaşımını kullanarak nötral lokuslardan farklı olan 20 TNP bulunmuştur. Bu 20 varyanttan 19'u Tip 2 diyabet ile biri Obezite ile ilişkilidir. Varyantların genomik lokasyon olarak en yakın oldukları 24 gen Tablo 4.4'te listelenmiştir. Bu genlerden on altısı protein kodlayan, dördü psödogen, üçü lncRNA ve biri miRNA'dır. rs7029718 varyantı ile ilişkili gen bulunmamaktadır.

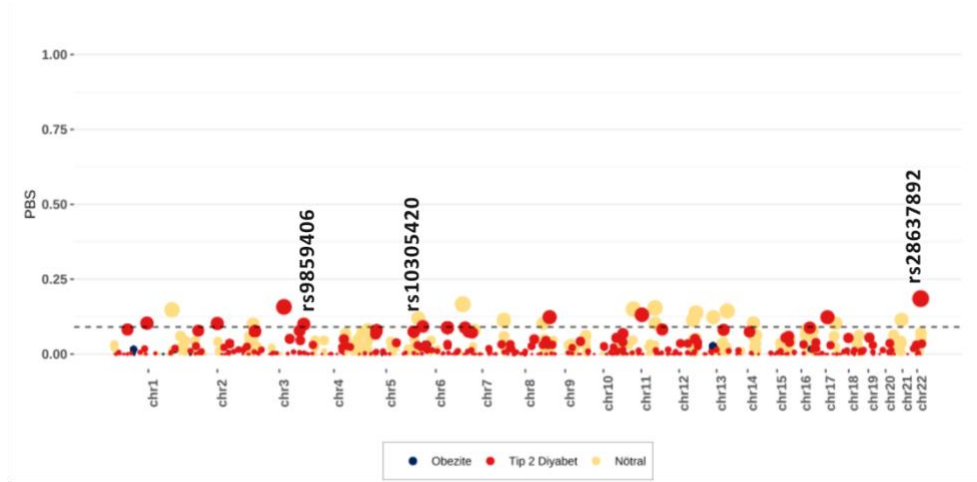
$F_{ST}$  tahminine dayalı analize ek olarak, aynı dönemler arasında popülasyon dal istatistiğine (PDİ) dayalı analiz ile risk alel frekans artışının gözlemlendiği popülasyon dallarındaki seçilim sinyalleri araştırılmıştır. İki karşılaştırma için de Anadolu popülasyonlarına dış grup olarak 1000 Genom Projesi Faz 3'ten (123) Yoruba popülasyonu seçilmiştir. PDİ skorları, ANGSD (v0.934) (141) ile hesaplanan  $F_{ST}$  tahminleri kullanılarak Gereç ve Yöntemler 3.5. Bölümünde açıklandığı gibi hesaplanmıştır.

Tez kapsamında seçilen Obezite ve Tip 2 diyabet ile ilişkili TNP'lerden popülasyon dal istatistiği dağılımının %99,5 yüzdelik diliminin üzerinde PDİ skorlarına sahip TNP'ler, kromozom numarası, pozisyon, risk aleli, risk alel frekansı, varyanta en yakın genlerin isimleri, genlerin türleri ve mutasyon sonuçları Tablo 4.5.'te gösterilmiştir.

**Tablo 4.5.** Popülasyon dal istatistiği karşılaştırmalarında belirlenen eşik değerin üzerinde olan ve risk alel frekans değişimi hesaplanabilen TNP'lerin listesi.

TNP	Fenotip	Lokasyon	Risk Aleli	Popülasyon Dal İstatistiği Karşılaştırılan Popülasyonlar	Popülasyon Dal İstatistiği Skoru	Frekans Değişimi	Neolitik RAF	Neolitik Sonrası RAF	Gen İsmi	Gen Türü	Mutasyonun En Ciddi Sonuçları
rs28637892	Tip 2 Diyabet	22:45917738	T	Anadolu Neolitik Sonrası - Anadolu Neolitik	0,1851349	Azalma	0,05	0	ATXN10 - WNT7B	protein kodlayan - protein kodlayan	intergenik varyant
rs9859406	Tip 2 Diyabet	3:185816694	G	Anadolu Neolitik Sonrası - Anadolu Neolitik	0,09984142	Azalma	0,24	0	IGF2BP2	protein kodlayan	intron varyant
rs10305420	Tip 2 Diyabet	6:39048860	C	Anadolu Neolitik Sonrası - Anadolu Neolitik	0,0921697	Artma	0	0,4	GLP1R	protein kodlayan	missense varyantı
TNP	Fenotip	Lokasyon	Risk Aleli	Popülasyon Dal İstatistiği Karşılaştırılan Popülasyonlar	Popülasyon Dal İstatistiği Skoru	Frekans Değişimi	Neolitik RAF	Günümüz RAF	Gen İsmi	Gen Türü	Mutasyonun En Ciddi Sonuçları
rs7531118	Obezite	1:72371556	C	Anadolu Günümüz - Anadolu Neolitik	0,20100855	Artma	0,18	0,75	RNU6 - 1246P - RPL31P12	snRNA - psödogen	intron varyant
rs35011184	Tip 2 Diyabet	10:112989975	G	Anadolu Günümüz - Anadolu Neolitik	0,17090062	Artma	0	0,47	TCF7L2	protein kodlayan	intron varyant
rs1815591	Tip 2 Diyabet	20:62645662	T	Anadolu Günümüz - Anadolu Neolitik	0,16220126	Artma	0,48	0,72	SLCO4A1	protein kodlayan	intron varyant
rs2820444	Tip 2 Diyabet	1:219568478	G	Anadolu Günümüz - Anadolu Neolitik	0,15449761	Artma	0	0,3	LYPLAL1 - AS1 - ZC3H11B	lncRNA - protein kodlayan	intergenik varyant
rs1922879	Tip 2 Diyabet	7:147961447	G	Anadolu Günümüz - Anadolu Neolitik	0,14605817	Artma	0	0,35	CNTNAP2	protein kodlayan	intron varyant
rs6712905	Tip 2 Diyabet	2:196087286	C	Anadolu Günümüz - Anadolu Neolitik	0,14535714	Azalma	0,49	0,02	DNAH7 - STK17B	protein kodlayan - protein kodlayan	intergenik varyant
rs7107784	Tip 2 Diyabet	11:2193859	G	Anadolu Günümüz - Anadolu Neolitik	0,14191428	Artma	0,63	0,84	ASCL2 - MIR4686	protein kodlayan - miRNA	regülatör bölge varyantı
rs917195	Tip 2 Diyabet	7:30688836	C	Anadolu Günümüz - Anadolu Neolitik	0,13738093	Artma	0,05	0,23	CRHR2	protein kodlayan	kodlamayan transkript ekzon varyantı
rs2257883	Tip 2 Diyabet	12:65822382	G	Anadolu Günümüz - Anadolu Neolitik	0,12029167	Artma	0	0,2	RPSAP52	psödogen	intron varyant
rs13005841	Tip 2 Diyabet	2:211437848	T	Anadolu Günümüz - Anadolu Neolitik	0,11058333	Artma	0	0,33	ERBB4	protein kodlayan	intron varyant

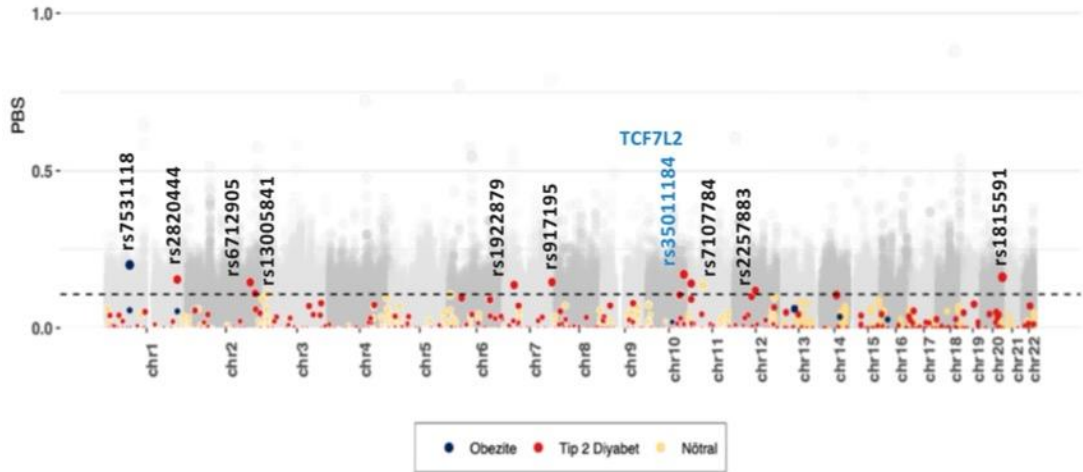
TNP'lerle ilişkili genomik lokasyonu, risk aleli, risk alel frekansı, varyanta en yakın genlerin isimleri, genlerin türleri ve mutasyona dair bilgiler listelenmiştir. Risk alel frekansı, RAF.



**Şekil 4.11.** Anadolu Neolitik, Anadolu Neolitik sonrası ve Yoruba popülasyonları arasındaki popülasyon dal istatistiği analizi için Manhattan grafiği. Hedef popülasyon Anadolu Neolitik sonrası ve dış grup Yoruba seçilmiştir.

Şekil 4.11.'de gösterilen PDİ skoru dağılımı, Anadolu Neolitik Sonrası dalında tez kapsamında incelemek için hedeflenen TNP'lerden Tip 2 diyabet ile ilişkili 9 TNP için sinyali göstermektedir. Bunlar arasında alel frekans değişimi istatistiksel olarak güven aralıklarına uygun bir şekilde hesaplayabilen 3 TNP'den ikisinin rs28637892, rs9859406 risk alel frekansında azalma ve rs10305420 varyantının alel frekansında artış gözlemlenmiştir (Tablo 4.5.). Sinyal yakalanan varyantların genomik lokasyon olarak en yakın oldukları 4 gen Tablo 4.5.'te listelenmiştir. Bu genlerin hepsi protein kodlayan genlerdir.

En güçlü seçim sinyali (PDİ skoru= 0,1851349), GWAS Catalog'da Tip 2 diyabet ile ilişkilendirilmiş, risk alel frekansında azalma gözlemlediğimiz bir intergenik varyantdır (rs28637892). En yüksek ikinci seçim sinyali (PDİ skoru= 0,09984142) *IGF2BP2* geninde bulunan Tip 2 diyabet ile ilişkili risk alel frekansında azalma gözlemlediğimiz bir intron varyantı olan rs9859406'dır. Yakaladığımız son sinyal (PDİ skoru= 0,0921697) Tip 2 diyabet ile ilişkili Neolitik dönemden Neolitik Sonrası döneme risk alel frekansında artış gözlemlenen bir missense varyant olan rs10305420'dır. GWAS Catalog'da Tip 2 diyabet ve sigaraya başlama özelliği ile de ilişkilendirilmiştir.



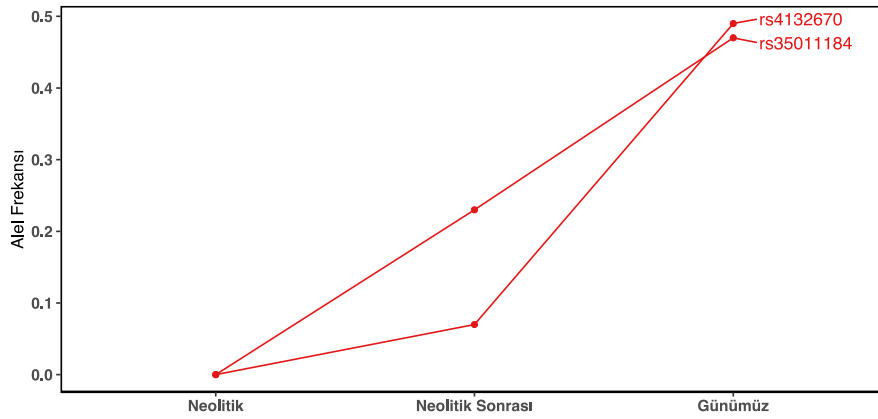
**Şekil 4.12.** Anadolu Günümüz, Anadolu Neolitik ve Yoruba popülasyonları arasındaki popülasyon dal istatistiği analizi için Manhattan grafiği. Hedef popülasyon Anadolu Günümüz ve dış grup Yoruba seçilmiştir.

Şekil 4.12.'de gösterilen PDI skoru dağılımı, Anadolu Günümüz dalında tez kapsamında incelemek için hedeflenen TNP'lerden Obezite için 1, Tip 2 diyabet ile ilişkili 9 TNP için sinyali göstermektedir. Bunlar arasında alel frekans değişimi istatistiksel olarak güven aralıklarına uygun bir şekilde 10 TNP'den birinin rs6712905 risk alel frekansında azalma ve rs13005841, rs1815591, rs1922879, rs2257883, rs2820444, rs3501184, rs7107784, rs7531118, rs917195 varyantlarının alel frekansında artış gözlemlenmiştir (Tablo 4.5.). Sinyal yakalanan varyantların genomik lokasyon olarak en yakın oldukları 14 gen Tablo 4.5.'de listelenmiştir. Bu genlerden dokuzu protein kodlayan, ikisi psödogen, kalan üç tanesi lncRNA, snRNA ve miRNA'dır.

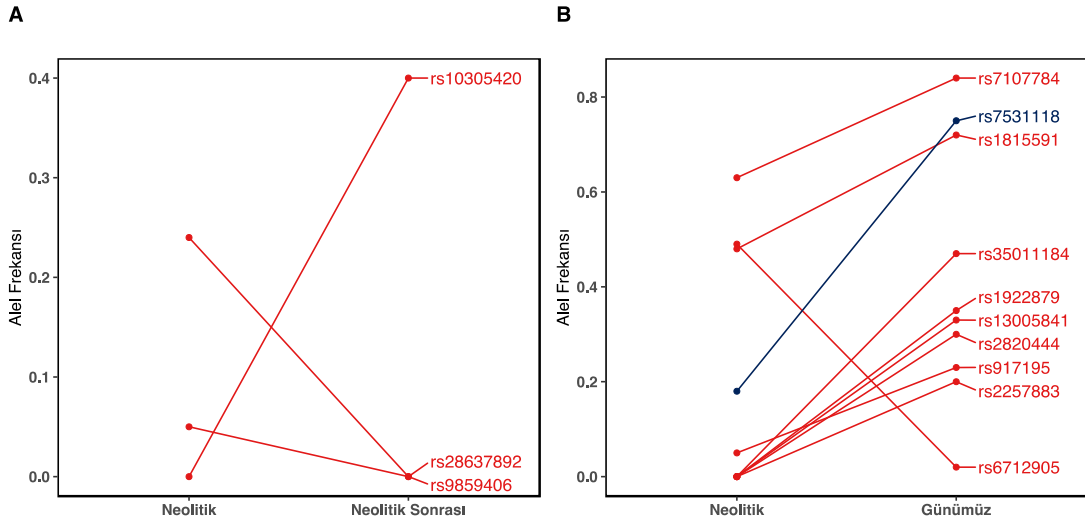
En güçlü seçilim sinyali (PDI skoru= 0,20100855), GWAS Catalog'da obezite ile ilişkilendirilmiş bir intron varyantıdır (rs7531118). rs7531118 aynı zamanda kalça-bel çevresi ölçüsü, vücut kitle indeksi ve bazı nöronal bozukluklarla ilişkilendirilmiştir (dikkat eksikliği ve hiper aktivite bozukluğu, major depresif bozukluk gibi).



En yüksek ikinci seçilim sinyali (PDİ skoru= 0,17090062) *TCF7L2* geninde bulunan bir intron varyantı olan rs35011184'tür. Şekil 4.13.'te risk alel frekans değişimi gösterilen varyant ile ilişkili GWAS Catalog'da bulunan Tip 2 diyabet ile ilişkili tek çalışma bulunmaktadır (164). GWAS Catalog'a göre *TCF7L2* geni hemogloblin A1c seviyeleri, Tip 2 diyabet, açlık kan şekeri, açlık insülin seviyeleri ve vücut kitle indeksi ile ilişkilidir. *TCF7L2* geni fonksiyonel domainleri yüksek oranda korunmuş sekanslardan oluşan Wnt sinyal yolağının parçasını oluşturan transkripsiyon faktörüdür ve Tip 2 diyabet çalışmalarında sıklıkla çalışılan bir genidir.



**Şekil 4.13.** Popülasyon dal istatistiği ve  $F_{ST}$  analizi ile seçilim izi yakalanan ve alel frekans artışı gözlenen *TCF7L2* genindeki varyantların risk alel frekanslarının değişimi.



**Şekil 4.14.** Popülasyon dal istatistiği analizleri ile seçilim sinyali yakalanan varyantların risk alel frekans dağılımları. A panelinde Anadolu Neolitik sonrası dalında seçilim sinyali yakalanan ve risk alel frekansı değişimi gözlenen üç Tip 2 diyabet (kırmızı) varyantının frekans dağılımları çizgi grafik ile gösterilmektedir. B panelinde Anadolu günümüz dalında dalında seçilim sinyali yakalanan ve risk alel frekansı değişimi gözlenen 9 Tip 2 diyabet (kırmızı) ve bir Obezite (mavi) varyantının frekans dağılımları çizgi grafik ile gösterilmektedir.

Günümüze kadar Obezite ve Tip 2 diyabet için yapılmış çalışmalar modern insan popülasyonları temel alınarak yapılan çalışmalardır. Tip 2 diyabet ile ilişkili alellerle ilişkili bazı seçilim çalışmalarının sonuçları tutumlu gen hipoteziyle çelişmiştir (162, 165). Örneğin pozitif seçilim altında olan alellerin risk alellerine kıyasla koruyucu aleller olduğuna dair veri sunulmuştur (78, 166, 167). Bu çalışmalardan birinde vücut kitle indeksi ile ilişkili 165 varyant analiz edilmiş ve pozitif seçilim sinyali yakalanan dokuz varyanttan dördü yağ depolanmasını teşvik eden risk aleli için seçilimi, kalan beş varyantta zayıflıkla ilişkili aleller üzerinde seçilime dair veri sunmuştur (166).

İnsan Genomu Çeşitliliği Projesi (HGDP) popülasyonları kullanılarak yapılan bir çalışmada (168) popülasyonlar arasında Tip 2 diyabetle ilişkili lokuslarda yüksek derecede farklılaşma olduğu ve bu farklılaşmanın en çok Doğu Asya ve Sahra altı Afrika popülasyonlarında belirgin olduğuna dair veri sunmuştur. Aynı çalışmada Doğu Asyalılarda haplotip bazlı analizlerde genişletilmiş haplotip homozigotluğu (İng. Extended haplotype homozygosity) ve diğer popülasyonlara kıyasla yüksek

farklılaşma gösteren Tip 2 diyabetle ilişkili *HHEX*, *TCF7L2* geni gibi doğal seçilime maruz kalmış lokuslar tanımlanmıştır (168).

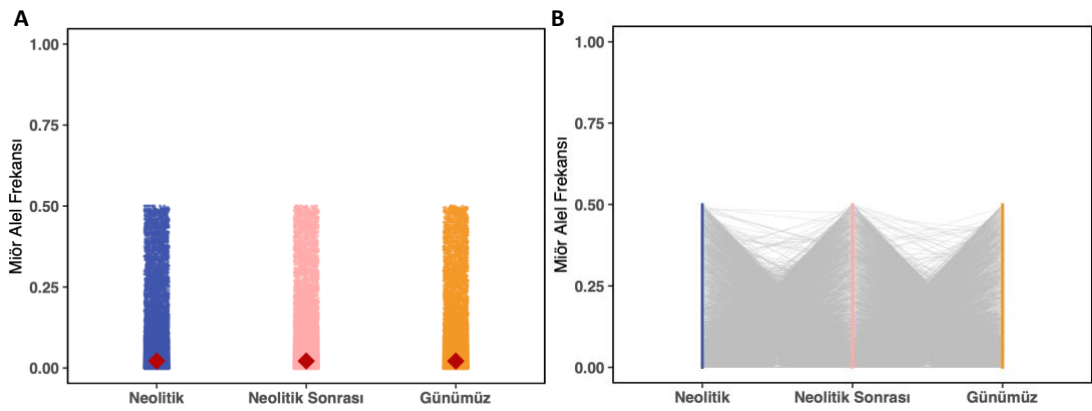
Afrika genomlarında seçilimin enerji metabolizması, yağ dokusu biyolojisi, insülin etkisi ve tuz tutma dahil olmak üzere Tip 2 diyabet ve hipertansiyonun altında yatan mekanizmalar üzerindeki etkisine dair yapılan bir çalışmada *PPARA* ve *TCF7L2* genlerindeki varyantlar için bulunan seçilim etkileri ayrıca *MPHOSPH9*, *TMEM127*, *ZRANB3* ve *MC3R* dahil olmak üzere çeşitli genler için insülin direnci ve salgılanması dahil olmak üzere insülin etkisi üzerinde seçilim etkilerine dair kanıtlar bulunmuştur (169).

Segurel ve arkadaşları (167) tarafından Tutumlu gen hipotezinin aksine, Tip 2 diyabet ile ilişkili risk alelleri yerine, bazı koruyucu alellerin pozitif seçilim baskısı altında olduğu tespit edilmiştir. Orta Asya'dan bir çoban ve bir çiftçi popülasyonunu karşılaştırarak Tip 2 diyabet ile ilişkili 10 varyanttaki seçilim paternlerini inceledikleri çalışmada iki popülasyon için de Neolitik dönemden itibaren koruyucu aleller üzerinde pozitif seçilim izlerini tespit etmişlerdir. Tanımladıkları koruyucu aleller 5500 ila 12.000 yıl önce, ilk tarım devrimi zamanında ortaya çıkmış olduğu için bu geçiş süresince seçilim baskısına uğramış olabilecekleri düşünülmektedir. Çalışmada Orta Asya'da iki popülasyonda Tip 2 diyabet ile ilişkili *LEPR* (rs1137100), *HHEX* (rs1111875) ve *PON1* (rs3917498) genlerindeki varyantlar için bu seçilim etkileri bulunmuştur (167).

#### **4.4. İleri Zamanlı Genetik Benzetim ile Anadolu Popülasyonlarında Nötral Senaryo Altında Alel Frekans Değişimlerinin Modellenmesi**

Fenotiple ilgili seçilen varyantların alel frekanslarındaki değişikliklerin hangi evrimsel senaryo altında gerçekleştiği test edilmek istenmiştir. Bu amaçla yapılan ileri zamanlı genetik benzetim analizleri SLiM (v.3.7) (153) kullanarak simüle edilmiş veriler üzerinde alel frekansı hesaplanmıştır. Kullanılan parametre ve değerler Gerekçe ve Metot Bölüm 3.5'te açıklanmıştır.

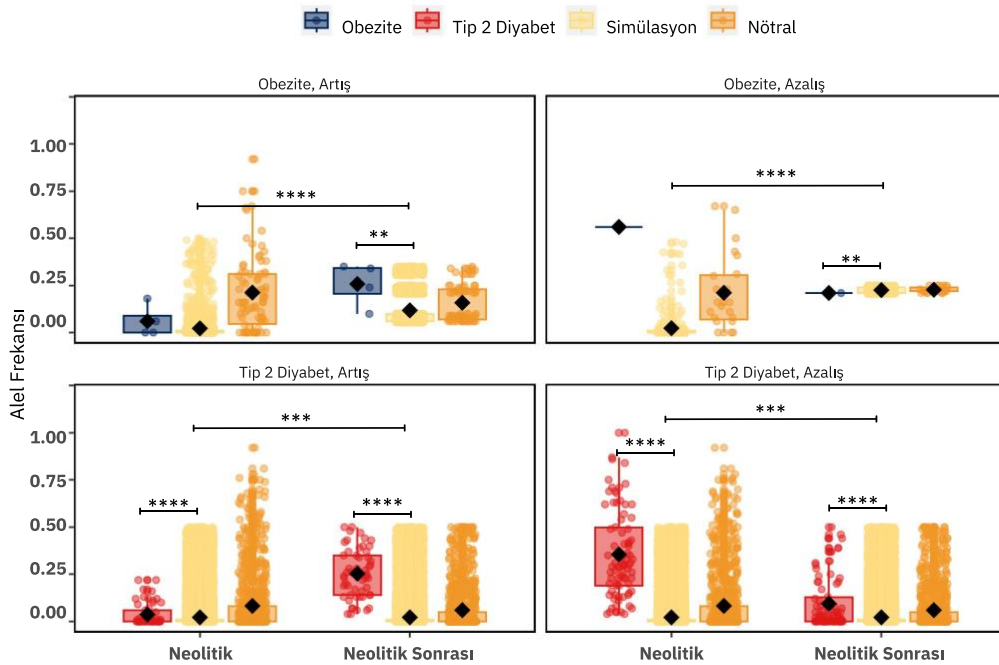
Neolitik dönemi temsilen günümüzden 10.000-6000 yıl öncesinden 100 birey, Neolitik sonrası dönemi temsil eden günümüzden 5200-3200 yıl öncesinden 100 birey ve günümüz Anadolu dönemini temsil eden 100 birey için 10 Mb'lık genom dizisi varyant çağırma dosyası (VCF) formatında üretilmiştir. Anadolu'daki Neolitik, Neolitik sonrası ve günümüzdeki dönemler arasındaki alel frekansları arasındaki değişimi gözlemlemek amaçlanarak simüle edilmiş popülasyonların VCF dosyalarından Plink (v.1.9) (122) ile toplam 55.921 TNP için minör alel frekansı hesaplanmıştır (Şekil 4.15.).



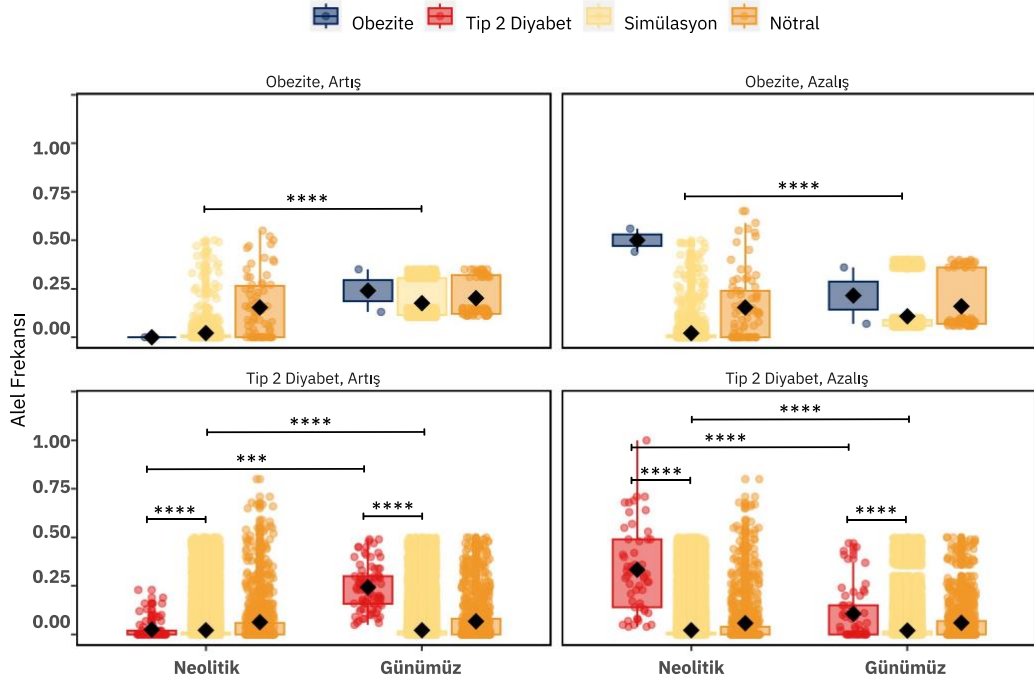
**Şekil 4.15.** İleri zamanlı genetik benzetim ile Anadolu popülasyonlarında nötral senaryo altında hesaplanan alel frekanslarının dağılımı. A panelinde her dönem için ortalama alel frekansları kırmızı elmas şekli ile gösterilmiştir.

Alel frekans hesaplamalarından sonra, bu frekanslar R'da (v4.2.2, <https://www.r-project.org/>) artan ve azalan bütün aleller için her dönem çifti arasında karşılaştırılmıştır. Alel frekansı değişikliklerinin anlamlılığını test etmek için non-parametrik Wilcoxon testi kullanılmış ve "rstatix" paket fonksiyonları (v0.7.2) kullanılarak tüm özellikler için p değerlerine "Bonferroni" çoklu test düzeltme yöntemi uygulanmıştır. Sonrasında gerçek veri ilişkili alellerin aynı dönemdeki nötral alellerle frekanslarını karşılaştırmak için non-parametrik Mann-Whitney U testi kullanılmış ve "rstatix" paket fonksiyonları kullanılarak tüm özellikler için p değerlerine "Bonferroni" çoklu test düzeltme yöntemi uygulanmıştır. İkili dönemler arası alel frekansı karşılaştırılırken nötral senaryo altında hesaplanan alel frekansları ile gerçek veriden maksimum olabilirlik temelli yaklaşım kullanılarak hesaplanan

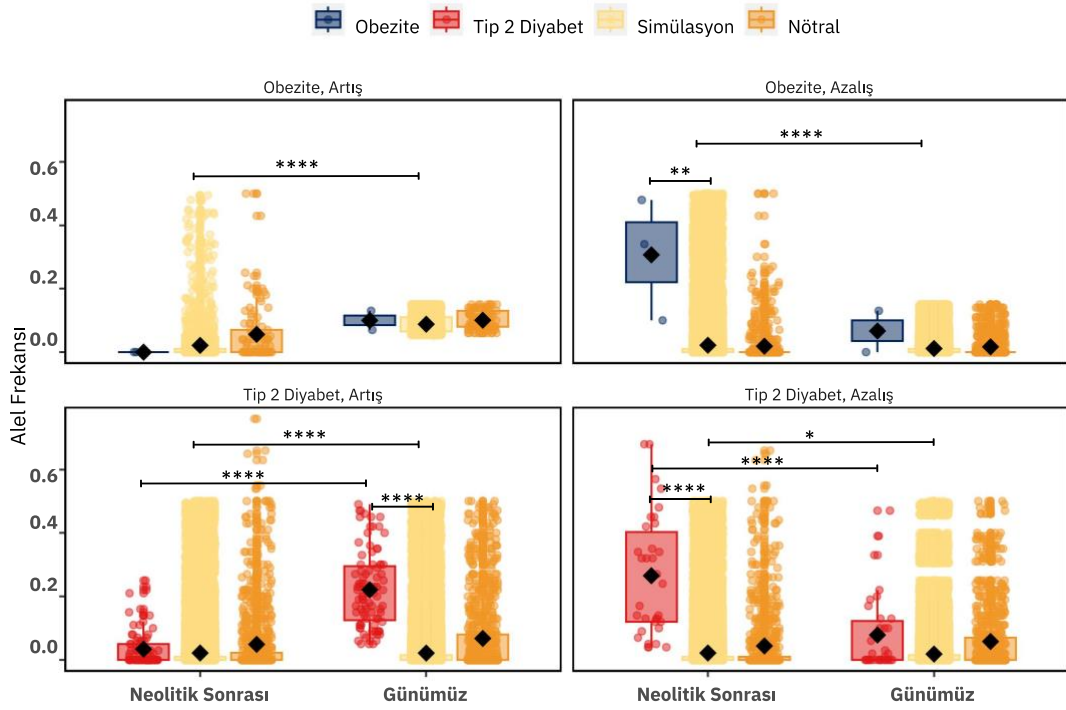
Obezite, Tip 2 diyabet ve nötral varyantların alel frekansları karşılaştırılmıştır. Gerçek veriden hesaplanan Obezite, Tip 2 diyabet ile ilişkili varyantların dönemler arasında risk alel frekansı artış ve azalma gösterenler karşılaştırmaya dahil edilirken, nötral varyantların tamamı karşılaştırmaya dahil edilmiştir. Karşılaştırma için alel frekans değerleri sıfır ile bir arasında 0,05'erli bölünen gruplara atanmış ve frekans değişimi gözlenen dönemlerde aynı gruba atanan alel frekans değerleri seçilmiştir. Neolitik, Neolitik sonrası ve günümüz popülasyonları için nötral senaryo altında simüle edilen varyantların alel frekansları, yayınlanmış genom verisinden hesaplanan alel frekansları ile karşılaştırılarak frekans değişimi gözlemlenen varyantların nötral süreçten farklı davranıp davranmadığı incelenmiştir.



**Şekil 4.16.** Neolitikten Neolitik sonrası döneme risk alel frekans artışı (sol panel) ve azalışı (sağ panel) gözlenen Obezite (mavi) ve Tip 2 diyabet (kırmızı) varyantlarının ileri zamanlı genetik benzetim ile nötral senaryo altında hesaplanan alel frekansları (açık sarı) ile karşılaştırılması. Obezite ve Tip 2 diyabet ile benzetimle hesaplanan alel frekanslarının ortalamaları Mann-Whitney U testi ile karşılaştırılmış ve Bonferroni düzeltmesi sonucu p-değeri anlamlılıklarına göre anlamlı sonuçlar (p-değeri < 0,05) kutu grafikler üzerinde gösterilmiştir. Obezite, Tip 2 diyabet ve benzetimle hesaplanan alel frekanslarının dönemler arası değişimi ise Wilcoxon testi ile karşılaştırılmış Bonferroni düzeltmesi sonucu p-değeri anlamlılıklarına göre anlamlı sonuçlar (p-değeri <0,05) kutu grafikler üzerinde gösterilmiştir.



**Şekil 4.17.** Neolitikten günümüze risk alel frekans artışı (sol panel) ve azalışı (sağ panel) gözlenen Obezite (mavi) ve Tip 2 diyabet (kırmızı) varyantlarının ileri zamanlı genetik benzetim ile nötral senaryo altında hesaplanan alel frekansları (açık sarı) ile karşılaştırılması. Obezite ve Tip 2 diyabet ile benzetimle hesaplanan alel frekanslarının ortalamaları Mann-Whitney U testi ile karşılaştırılmış ve Bonferroni düzeltmesi sonucu p-değeri anlamlılıklarına göre anlamlı sonuçlar (p-değeri < 0,05) kutu grafikler üzerinde gösterilmiştir. Obezite, Tip 2 diyabet ve benzetimle hesaplanan alel frekanslarının dönemler arası değişimi ise Wilcoxon testi ile karşılaştırılmış Bonferroni düzeltmesi sonucu p-değeri anlamlılıklarına göre anlamlı sonuçlar (p-değeri <0,05) kutu grafikler üzerinde gösterilmiştir.



**Şekil 4.18.** Neolitik sonrası dönemden günümüze risk alel frekans artışı (sol panel) ve azalışı (sağ panel) gözlenen Obezite (mavi) ve Tip 2 diyabet (kırmızı) varyantlarının ileri zamanlı genetik benzetim ile nötral senaryo altında hesaplanan alel frekansları (açık sarı) ile karşılaştırılması. Obezite ve Tip 2 diyabet ile benzetimle hesaplanan alel frekanslarının ortalamaları Mann-Whitney U testi ile karşılaştırılmış ve Bonferroni düzeltmesi sonucu p-değeri anlamlılıklarına göre anlamlı sonuçlar (p-değeri < 0,05) kutu grafikler üzerinde gösterilmiştir. Obezite, Tip 2 diyabet ve benzetimle hesaplanan alel frekanslarının dönemler arası değişimi ise Wilcoxon testi ile karşılaştırılmış Bonferroni düzeltmesi sonucu p-değeri anlamlılıklarına göre anlamlı sonuçlar (p-değeri <0,05) kutu grafikler üzerinde gösterilmiştir.

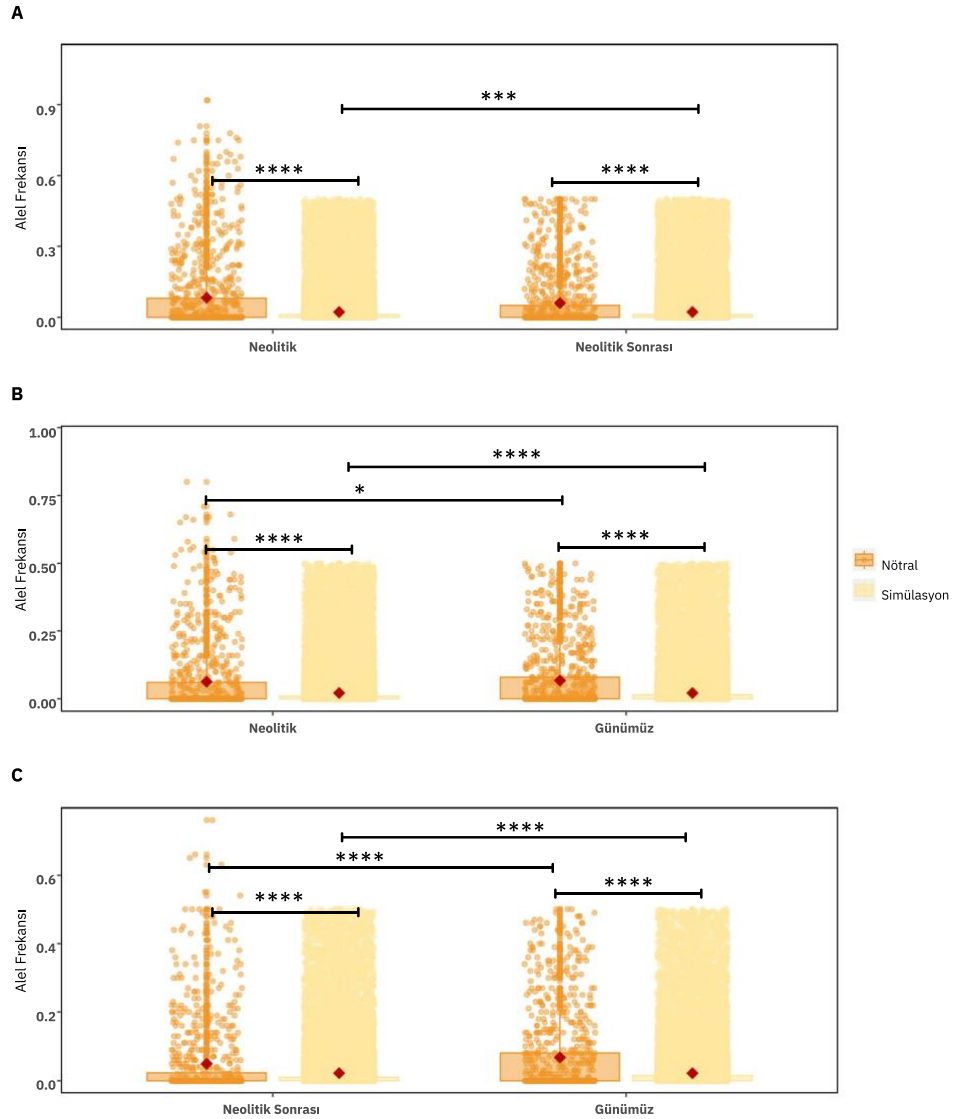
Şekil 4.16.'da Neolitikten Neolitik sonrası döneme Obeziteyle ilişkili frekansı değişim gösteren risk alellerinin ve değişim gösterdiği dönemde aynı frekanstaki nötral senaryo altında hesaplanan alel frekans değişimi ile kıyaslandığında anlamlı farklılık göstermektedir. Ancak karşılaştırmada Obezite ile ilişkili varyant sayısının az olması bu sonucun güvenilirliğini düşürmektedir. Neolitikten Neolitik sonrası döneme Tip 2 diyabetle ilişkili frekansı değişim gösteren risk alellerinin ve değişim gösterdiği dönemde aynı frekanstaki nötral senaryo altında hesaplanan alel frekans değişimi ile kıyaslandığında anlamlı farklılık göstermektedir.

Şekil 4.17.'de Neolitikten günümüze Obeziteyle ilişkili frekans değişim gösteren risk alellerinin ve değişim gösterdiği dönemde aynı frekanstaki nötral senaryo altında hesaplanan alel frekans değişimi ile kıyaslandığında anlamlı farklılık göstermektedir. Ancak karşılaştırmada Obezite ile ilişkili varyant sayısının az olması bu sonucun güvenilirliğini düşürmektedir. Neolitikten günümüze Tip 2 diyabetle ilişkili frekans değişim gösteren risk alellerinin ve değişim gösterdiği dönemde aynı frekanstaki nötral senaryo altında hesaplanan alel frekans değişimi ile kıyaslandığında benzer bir değişim gözlenmektedir

Şekil 4.18.'de Neolitik sonrası dönemden günümüze Obeziteyle ilişkili frekans değişim gösteren risk alellerinin ve değişim gösterdiği dönemde aynı frekanstaki nötral senaryo altında hesaplanan alel frekans değişimi ile kıyaslandığında anlamlı farklılık göstermektedir. Ancak karşılaştırmada Obezite ile ilişkili varyant sayısının az olması bu sonucun güvenilirliğini düşürmektedir. Neolitik sonrası dönemden günümüze Tip 2 diyabetle ilişkili frekans değişim gösteren risk alellerinin ve değişim gösterdiği dönemde aynı frekanstaki nötral senaryo altında hesaplanan alel frekans değişimi ile kıyaslandığında benzer bir değişim gözlenmektedir.

Şekil 4.19.'da nötral varyantlar ile benzetimle hesaplanan alel frekanslarının dönemler arası değişimi incelenmiştir. Yalnızca Neolitik-Neolitik sonrası dönem arasındaki nötral varyantların minör alel frekansındaki değişim ile benzetimle hesaplanan alel frekansları arasında farklılık gözlenmiştir. Diğer dönem çiftleri arasında istatistiksel olarak bir farklılık işaret etmemektedir.

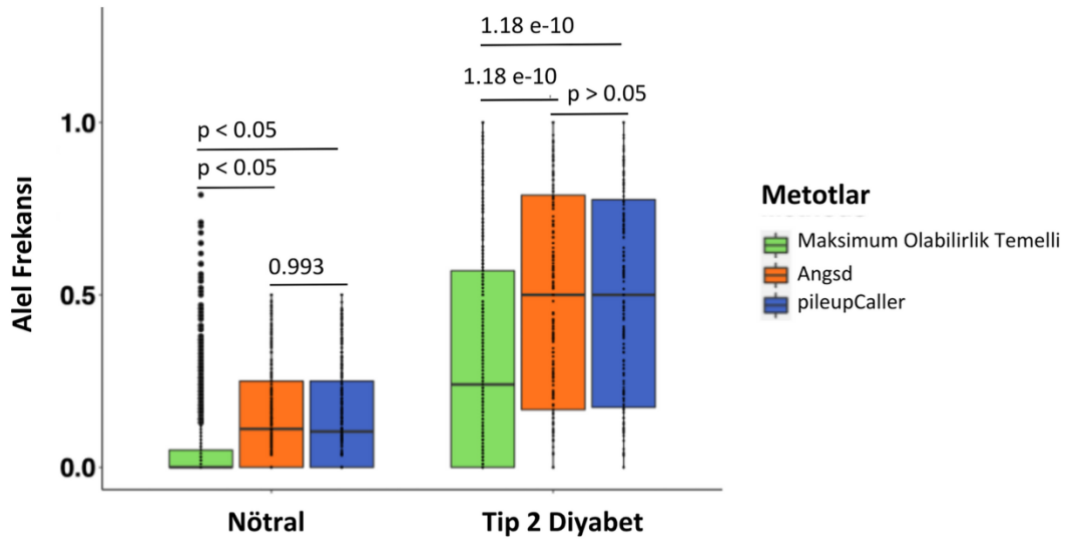




**Şekil 4.19.** Anadolu antik ve modern dönemlerinde minor alel frekans değerleri hesaplanan nötral (turuncu) varyantlarının ileri zamanlı genetik benzetim ile nötral senaryo altında hesaplanan alel frekansları (açık sarı) ile karşılaştırılması. A panelinde Neolitik-Neolitik sonrası dönem çiftleri, B panelinde Neolitik sonrası günümüz dönem çiftleri, C panelinde Neolitik ve günümüz dönem çiftleri incelenmektedir. Nötral ile benzetimle hesaplanan alel frekanslarının ortalamaları Mann-Whitney U testi ile karşılaştırılmış ve Bonferroni düzeltmesi sonucu p-değeri anlamlılıklarına göre anlamlı sonuçlar ( $p$ -değeri  $< 0,05$ ) kutu grafikler üzerinde gösterilmiştir. Nötral ve benzetimle hesaplanan alel frekanslarının dönemler arası değişimi ise Wilcoxon testi ile karşılaştırılmış Bonferroni düzeltmesi sonucu p-değeri anlamlılıklarına göre anlamlı sonuçlar ( $p$ -değeri  $< 0,05$ ) kutu grafikler üzerinde gösterilmiştir.

#### 4.5. Farklı Biyoinformatik Yaklaşımlarla Hazırlanan Veri Setlerinin Test Edilmesi

Sözde-haploidizasyon yöntemleri nötral fenotip alellerinin frekans ortalamalarına çok yakındır (ANOVA, Tukey post hoc,  $p < 0.05$ ). Haploidizasyon temelli yöntemlerle karşılaştırıldığında, maksimum olabilirlik tahmini temelli yöntemler haploidizasyon temelli yöntemlerden önemli ölçüde farklı alel frekansları ortalamaları elde etmiştir (ANOVA, Tukey post hoc,  $p < 0.05$ ) (Şekil 4.20.). Mevcut bulgulara göre, düşük okuma derinliğine sahip antik DNA verilerinde ANGSD ve pileupCaller ile varyant çağırma, alel frekansı hesaplama gibi popülasyon genetiği analizlerinde farklılıklara yol açmamaktadır. İstatistiksel olarak farklı olan sözde-haploidizasyon ve maksimum olabilirlik tahmini tabanlı yöntemlerle karşılaştırıldığında hangi yöntemin daha güvenilir olduğunu belirlemek için daha fazla çalışmaya ihtiyaç vardır.



**Şekil 4.20.** Farklı varyant çağırma yaklaşımları ile hesaplanan alel frekanslarının dağılımları. ANOVA sonuçlarını Tukey testi sonuçları takip etmektedir ve anlamlılık seviyeleri kutu grafiklerinde sunulmuştur.

## 5. TARTIŞMA

Kompleks hastalıklar, birden fazla gen ve çevresel faktörlerin etkileşimi sonucu ortaya çıkan hastalıklardır. Evrimsel biyolojinin sorularından birisi bu hastalıkların toplumdaki yaygınlığı ve genom boyu ilişkilendirme çalışmaları ile tespit edilen ilişkili varyantların frekansının ortalama veya yüksek olması düşünüldüğünde hangi evrimsel mekanizmaların genetik varyasyonu ortalama frekansta tuttuğunun cevaplanmasıdır.

Neolitikleşme ile insan toplulukları geçim stratejilerinde, beslenme alışkanlıklarında, kültürel ve sosyal ilişkilerinde köklü değişimler yaşamış, hayvancılıkla birlikte yeni patojenlerle karşılaşmış ve yavaş yavaş yerleşik hayata geçmiştir. Coğrafi olarak Doğu ile Batı arasında bir köprü görevi gören Anadolu'daki antik popülasyonlar, kültürel süreçteki bu köklü değişimlere yakından tanıklık etmiştir.

Bu tez kapsamında, Türkiye'de yüksek prevalansa sahip ve beslenme ile yakından ilişkili iki kompleks metabolik hastalık olan Obezite ve Tip 2 diyabete odaklanılmıştır. Neolitikleşme süreci ile bu hastalıkların evrimsel tarihini anlamak için antik ve günümüz genomları paleogenomik, popülasyon genetiği ve ileri zamanlı genetik simülasyon yaklaşımları ile araştırılmıştır.

Alel frekansı ve seçilim analizlerinde kullanılacak hesaplama yöntemi seçilirken örneklerin düşük ortalama okuma derinliğine sahip olduğu göz önünde bulundurulmuştur. Alel frekans hesaplamasında her popülasyondaki frekansın olasılığını maksimum olabilirlik tahmini ile hesaplayan Bölüm 3.4.'de açıklanan yöntem kullanılmıştır. Diploid varyant çağırma ile lokus hakkında elde edilen tüm bilgiyi kullanarak sekans verilerinden maksimum verimi almayı amaçlayan maksimum olabilirlik temelli bu yöntem, güvenilir diploid varyant çağırma için yeterli okuma derinliğine sahip olunmayan örnekler için sıklıkla kullanılmaktadır (63, 83, 170, 171).

Seçilim analizi için temel olarak kullandığımız  $F_{ST}$  istatistiğini hesaplarken özellikle düşük ve orta okuma derinliğine sahip genomik veriler için kullanışlı olan ve

genotip olabilirlikleri üzerinden hesaplama yapan ANGSD (141) yazılımı kullanılmıştır. Düşük okuma derinliğine sahip verilerde genotip bilgisine dair belirsizlik, haritalama ve dizileme hataları ve haploidizasyon gibi nedenlerden kaynaklanabilmektedir. Antik DNA çalışmalarında sıkça uygulanan haploidizasyon yöntemi heterozigot pozisyonlar yerine diploid kromozomlardan rastgele tek bir alelin seçilmesi ve sözde-haploid sekans oluşturulmasıdır. Düşük ve orta okuma derinlikli veriler için downstream analizlerde genotiplere kıyasla ham veri veya genotip olabilirliklerini kullanmak genotip bilgisine dair istatistiksel belirsizliği hesaba katmayı kolaylaştırmaktadır (141, 150, 172). Ayrıca ANGSD'ın kullandığı  $F_{ST}$  tahminicisinin, Hudson-Bhatia, popülasyonlar arası örneklem büyüklüğüne farklılıklarına duyarlılığı az olması nedeni ile de modern Anadolu (n=16) ve antik dönem popülasyonları (Neolitik n= 81, Neolitik sonrası= 238) arasında  $F_{ST}$  hesaplarırken ANGSD yazılımı kullanmak avantaj sağlamıştır.

$F_{ST}$  istatistiğini kullanarak hastalıkla ilişkili TNP'ler için antik ve günümüz Anadolu popülasyonları arasındaki genetik mesafe hesaplanmış ve sonuçları evrimsel olarak nötral bölgelerle karşılaştırılarak olası seçim sinyalleri yakalanmaya çalışılmıştır. Aynı karşılaştırma hastalıkla ilişkili varyantlar ve nötral varyantların alel frekansları ile de yapılmıştır. İlk olarak hastalıklar arasında nötral bölgelerle anlamlı bir farklılık olup olmadığı  $F_{ST}$  dağılımları ANOVA ve ardından post-hoc Tukey's testi ile karşılaştırılarak incelenmiştir. Bunun sonucunda hastalık ile ilişkili TNP'ler ve nötral TNP'ler arasında bu  $F_{ST}$  karşılaştırma yaklaşımını kullanılarak bu hastalıkların nötral bölgelerden daha farklı olduğu iki popülasyon çifti bulunmuştur: Anadolu Neolitik ve Anadolu Post Neolitik arasında nötral bölge için ortalama  $F_{ST}$  değeri 0,01223509 iken Tip 2 diyabet için 0,01488753 (p-değeri =  $9.97 \times 10^{-6}$ ) ve Obezite için 0,03142205 (p-değeri < 0,05), Anadolu Neolitik ve Anadolu Günümüz arasında nötral bölge için ortalama  $F_{ST}$  değeri 0,0224476 iken, Tip 2 diyabet için 0,02903933 (p-değeri = 0,000966) ve Obezite için 0,05041716 (p-değeri = 0,000726). Popülasyonda zaman içinde alel frekansının değişmesine gen akışı, genetik sürüklenme ve doğal seçim gibi evrimsel süreçler katkı sağlamaktadır. Bu yüzden zaman içinde artan alel frekansının başka bir popülasyonla karışım, gen akışı, gibi nötral süreçlerle ilişkili olması da beklenebilmektedir. Genomdaki nötral bölgelerle kıyaslama yapmak ve hastalıkla

ilişkili alellerin frekansının nötral varyantlarla kıyaslandığında anlamlı farklılık gösterdiğini gözlemlemek, gen akışı hakkında yorum yapabilmeyi sağlamaktadır. Nötral bölgelerden daha yüksek  $F_{ST}$  değerleri, alel frekansındaki değişikliklerin adaptif dolayısıyla rastgele olmayan süreçler sonucunda farklılaşma olduğunu ve poligenik seçilimi işaret eden bir süreç olduğunu gösterebilmektedir (162, 163).

$F_{ST}$  dağılımı ve alel frekans karşılaştırmaları ile popülasyonlar arasında farklılaşma belirtileri gördükten sonra, ilgilenilen özelliklerin mevcut yaygınlığını açıklayan pozitif seçim hipotezini test etmek için Anadolu popülasyonları ve bir dış popülasyon kullanılarak popülasyon dal istatistiği gerçekleştirilmiştir.

Anadolu Neolitik-Anadolu Neolitik Sonrası popülasyonlar ve Yoruba dış popülasyonu kullanılarak yapılan PDİ analizinde, Anadolu Neolitik Sonrası dalında tez kapsamında incelemek için hedeflenen TNP'lerden Tip 2 diyabet ile ilişkili 9 TNP için sinyal yakalanmıştır. Bunlar arasında alel frekans değişimi gözlenen 3 TNP'den ikisinin rs28637892, rs9859406 risk alel frekansında azalma ve rs10305420 varyantının alel frekansında artış gözlemlenmiştir (Tablo 4.5.).

Anadolu Neolitik-Anadolu Günümüz popülasyonları ve Yoruba dış popülasyonu kullanılarak yapılan PDİ analizinde, Anadolu Günümüz dalında tez kapsamında incelemek için hedeflenen TNP'lerden Tip 2 diyabet ile ilişkili 9 TNP için sinyal yakalanmıştır. Bunlar arasında alel frekans değişimi gözlenen bir şekilde 10 TNP'den birinin (rs6712905) risk alel frekansında azalma ve rs13005841, rs1815591, rs1922879, rs2257883, rs2820444, rs35011184, rs7107784, rs7531118, rs917195 varyantlarının alel frekansında artış gözlemlenmiştir (Tablo 4.5.).

Tip 2 diyabet ile farklı popülasyonlarda güçlü etkiye sahip farklı genetik lokuslar tanımlanmıştır (173): Avrupa popülasyonlarında *TCF7L2* (174), Batı Asya popülasyonlarında *KCNQ1* (175, 176), Meksika popülasyonlarında *SCL16A11* (177) ve Grönland popülasyonlarında *TBC1D4* (178).

Popülasyon dal istatistiği analizinde rs35011184 (Tablo 4.5.) ve  $F_{ST}$  karşılaştırmalarında rs4132670 (Tablo 4.4.) seçim adayı olan ve risk alel frekansında

artış gözlenen iki varyantla ilişkili gen olan *TCF7L2*, Tip 2 diyabet ile ilişkilidir ve Asya popülasyonları yapılan çalışmalarda koruyucu haplotip üzerinde yakın zamanlı seçilim görüldüğüne işaret edilmektedir (165, 168, 179). Seçilim adayı olan diğer varyantlar için literatürde bir çalışma bulunamamıştır.

Literatürdeki önceki çalışmalar, Tip 2 diyabet ve Obezite gibi bazı hastalıklar üzerindeki seçilimin, bu hastalıklarla ilişkili genlerin bolluk-kıtlık/egzersiz-dinlenme döngülerinde için pozitif seçilim altında evrimleştiğini varsayan "Tutumlu Genotip Hipotezi" ile açıklanabileceğini öne sürmektedir (45, 46, 162). Bu bireylerin üzerindeki seçilim baskısı enerji kaybını en aza indirip tasarruf yaparak bolluk dönemlerinde enerjiyi yağ olarak depolamalarına ve sonraki kıtlık-dinlenme döneminde yağ şeklinde depolanan enerjinin hayatta kalmayı kolaylaştırmasına dayanmaktadır. Bu hipotezin önermesi, doğal seçilim ile besini depolama ve kullanma konusunda verimli olacak şekilde evrimleştiğimizdir. Bu hipotez, tez kapsamında hem alel frekans değişimlerinde hem de seçilim analizlerinde gözlemlediğimiz frekans değişiklikleri için bir açıklama olabilir.

Bu çalışmada, alel frekansı ve seçilim analizlerinde kullanılacak hesaplama yöntemi seçilirken örneklerin düşük ortalama okuma derinliğine sahip olduğu göz önünde bulundurularak ham dizileme verisinden maksimum verimi almak hedeflenmiştir. Ancak son yayınlar seçilim analizinde  $F_{ST}$  tahmini için kullanılan genotip olabilirliklerine dayanan yöntemin de (141) düşük okuma derinliklerinde yeterli güvenilirlikte olmadığına dair sonuçlar sunmuştur (99, 180).

Özet olarak, bu çalışma,  $F_{ST}$  analizlerinde Tip 2 diyabet ve Obezite için poligenik seçilimin izi yakalanmış, popülasyon dal istatistiği analizi ve alel frekansında da bu hastalıkların fenotipte görülme riskini artıran bazı alellerde pozitif seçilim sinyalleri yakalanmıştır. Anadolu'da Neolitikleşmenin oluşturduğu seçilim baskısı incelenen başka bir tez çalışmasında Neolitik ve günümüz Anadolu popülasyonları incelenmiş ve lipid metabolizması genlerini içeren varyantlarda seçilim sinyalleri yakalanmıştır (181). Gelecekteki çalışmalarda; lipid metabolizması, leptin seviyesi, vücut kitle indeksi gibi fenotipler de dahil edilerek beslenme ve metabolizma ile ilişkili

daha geniş kapsamlı bir çalışma yapılmalıdır. Çalışma kapsamında GWAS Catalog'dan seçilen varyantların genotiplenme oranlarının düşük olduğu göz önüne alındığında varyantların ilişkili olduğu genlerde seçilim taraması yapmak antik DNA ile çalışmanın neden olduğu bu limitasyonu aşmaya yardımcı olabilir.

## 6. SONUÇ VE ÖNERİLER

### SONUÇLAR

1. Tez kapsamında yayınlanmış antik genomları ile Anadolu popülasyonları için oluşturulmuş en kapsamlı genomik veri seti derlenmiştir ([https://github.com/idiltac/YLisans\\_Tez\\_IdilTac/blob/main/tez\\_bireyler.csv](https://github.com/idiltac/YLisans_Tez_IdilTac/blob/main/tez_bireyler.csv)).
2. Genotip olabirlikleri, diploid varyant çağırma ve sözde-haploidizasyon varyant çağırma yaklaşımları ile üç farklı genomik veri seti oluşturularak bu yaklaşımların etkinliği alel frekans hesaplama sonuçları üzerinden test edilmiştir.
3. Genom boyu ilişkilendirme çalışmaları ile Avrupa popülasyonlarında keşfedilen varyantların Avrupa-dışı popülasyonlarda ne ölçüde bilgi verdiğini Anadolu popülasyonunu model olarak kullanılarak test edilmiştir.
4. Obezite ve Tip 2 diyabet ile ilişkili varyantların Anadolu'da son 10,000 yıldaki alel frekans yörüngeleri belirlenmiştir.
5. İleri zamanlı genetik benzetim yöntemleri ile nötral senaryo altında Anadolu popülasyonları modellenmiştir. İleri dönem çalışmalarında farklı evrimsel senaryoları denemek için popülasyon demografik modeli çıkarılmıştır.
6. Anadolu'da ilk defa antik popülasyonlar kullanılarak seçilim analizleri uygulanmıştır.

### GELECEK ÇALIŞMALAR

Daha ileri analizlerde, modern ve antik örnekler arasındaki karşılaştırmaların gücünü artırmak için i) Anadolu Neolitik ve Tunç Çağı'ndan daha fazla genom dizileme verileri oluşturmak ve daha fazla örnekleme hedeflenen hastalık varyantları için daha yüksek derinlikli genomik veri elde edilmesi, ii) Yayınlanmış antik genomik verilerin kalitesini ve elde edilen genotip bilgisini arttırmak için imputasyon uygulaması yapılması ve bu sayede haplotip bazlı seçilim analizlerini de yapabilme imkanı, iii) Alel



frekans hesaplamasında uygulanan dizileme verilerinden maksimum olabilirlik temelli algoritma (Bölüm 3.4.) ile genom boyu seçim analizi hesaplamak (63) iv) demografik modeli tez kapsamında oluşturulan Anadolu popülasyonları için nötral senaryoya ek olarak farklı seçim modelleri ile farklı evrimsel senaryoların test edilmesi planlanmaktadır.

## 7. KAYNAKLAR

1. Day-Williams AG, Zeggini E. The effect of next-generation sequencing technology on complex trait research. *Eur J Clin Invest.* 2011;41(5):561-7.
2. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* 2019;47(D1):D1005-D12.
3. Popejoy AB, Fullerton SM. Genomics is failing on diversity. *Nature.* 2016;538(7624):161-4.
4. Wojcik GL, Graff M, Nishimura KK, Tao R, Haessler J, Gignoux CR, et al. Genetic analyses of diverse populations improves discovery for complex traits. *Nature.* 2019;570(7762):514-8.
5. Mills MC, Rahal C. A scientometric review of genome-wide association studies. *Communications Biology.* 2019;2(1):9.
6. Giudicessi JR, Ackerman MJ. Genetic testing in heritable cardiac arrhythmia syndromes: differentiating pathogenic mutations from background genetic noise. *Curr Opin Cardiol.* 2013;28(1):63-71.
7. Goddard ME, Kemper KE, MacLeod IM, Chamberlain AJ, Hayes BJ. Genetics of complex traits: prediction of phenotype, identification of causal polymorphisms and genetic architecture. *Proc Biol Sci.* 2016;283(1835).
8. Steinrucken M, Bhaskar A, Song YS. A Novel Spectral Method for Inferring General Diploid Selection from Time Series Genetic Data. *Ann Appl Stat.* 2014;8(4):2203-22.
9. Terhorst J, Schlotterer C, Song YS. Multi-locus analysis of genomic time series data from experimental evolution. *PLoS Genet.* 2015;11(4):e1005069.
10. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. *Nature.* 2001;409(6822):860-921.
11. Hood L, Rowen L. The Human Genome Project: big science transforms biology and medicine. *Genome Medicine.* 2013;5(9):79.
12. Altshuler D, Donnelly P, The International HapMap C. A haplotype map of the human genome. *Nature.* 2005;437(7063):1299-320.
13. Altshuler DM, Gibbs RA, Peltonen L, Altshuler DM, Gibbs RA, Peltonen L, et al. Integrating common and rare genetic variation in diverse human populations. *Nature.* 2010;467(7311):52-8.

14. McVean GA, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, Chakravarti A, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012;491(7422):56-65.
15. Boyle EA, Li YI, Pritchard JK. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell*. 2017;169(7):1177-86.
16. Xing C, Garcia CK. Epigenetic inheritance of telomere length obscures identification of causative PARN locus. *J Med Genet*. 2016;53(5):356-8.
17. Loh PR, Bhatia G, Gusev A, Finucane HK, Bulik-Sullivan BK, Pollack SJ, et al. Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nat Genet*. 2015;47(12):1385-92.
18. Shi H, Kichaev G, Pasaniuc B. Contrasting the Genetic Architecture of 30 Complex Traits from Summary Association Data. *Am J Hum Genet*. 2016;99(1):139-53.
19. Barroso I, McCarthy MI. The Genetic Basis of Metabolic Disease. *Cell*. 2019;177(1):146-61.
20. Benton ML, Abraham A, LaBella AL, Abbot P, Rokas A, Capra JA. The influence of evolutionary history on human health and disease. *Nat Rev Genet*. 2021;22(5):269-83.
21. Sella G, Barton NH. Thinking About the Evolution of Complex Traits in the Era of Genome-Wide Association Studies. *Annu Rev Genomics Hum Genet*. 2019;20:461-93.
22. Behjati S, Tarpey PS. What is next generation sequencing? *Arch Dis Child Educ Pract Ed*. 2013;98(6):236-8.
23. Schwarze K, Buchanan J, Fermont JM, Dreau H, Tilley MW, Taylor JM, et al. The complete costs of genome sequencing: a microcosting study in cancer and rare diseases from a single center in the United Kingdom. *Genet Med*. 2020;22(1):85-94.
24. Sollis E, Mosaku A, Abid A, Buniello A, Cerezo M, Gil L, et al. The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Res*. 2023;51(D1):D977-D85.
25. Cavalli-Sforza LL, Feldman MW. Cultural transmission and evolution: a quantitative approach. *Monogr Popul Biol*. 1981;16:1-388.
26. Behnke JM, Stearns SC, Koella JC: Evolution in Health and Disease. *Parasites & Vectors*. 2009;2(1):4.
27. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. *Nature*. 2009;461(7265):747-53.

28. Reich DE, Lander ES. On the allelic spectrum of human disease. *Trends Genet.* 2001;17(9):502-10.
29. Collins FS, Guyer MS, Charkravarti A. Variations on a theme: cataloging human DNA sequence variation. *Science.* 1997;278(5343):1580-1.
30. Pritchard JK. Are rare variants responsible for susceptibility to complex diseases? *Am J Hum Genet.* 2001;69(1):124-37.
31. Schork NJ, Murray SS, Frazer KA, Topol EJ. Common vs. rare allele hypotheses for complex diseases. *Curr Opin Genet Dev.* 2009;19(3):212-9.
32. El-Fishawy P. Common Disease-Rare Variant Hypothesis. In: Volkmar FR, editor. *Encyclopedia of Autism Spectrum Disorders.* New York, NY: Springer New York; 2013. p. 720-2.
33. Mathieson I. The omnigenic model and polygenic prediction of complex traits. *Am J Hum Genet.* 2021;108(9):1558-63.
34. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am J Hum Genet.* 2017;101(1):5-22.
35. Wojcik GL, Graff M, Nishimura KK, Tao R, Haessler J, Gignoux CR, et al. The PAGE Study: How Genetic Diversity Improves Our Understanding of the Architecture of Complex Traits. *bioRxiv.* 2018:188094.
36. Spiegelman BM, Flier JS. Obesity and the regulation of energy balance. *Cell.* 2001;104(4):531-43.
37. Danaei G, Ding EL, Mozaffarian D, Taylor B, Rehm J, Murray CJ, et al. The preventable causes of death in the United States: comparative risk assessment of dietary, lifestyle, and metabolic risk factors. *PLoS Med.* 2009;6(4):e1000058.
38. Genné-Bacon EA. Thinking evolutionarily about obesity. *Yale J Biol Med.* 2014;87(2):99-112.
39. Stumvoll M, Goldstein BJ, van Haeften TW. Type 2 diabetes: pathogenesis and treatment. *Lancet.* 2008;371(9631):2153-6.
40. OECD. Overweight or obese population (indicator) 2023 [Available from: <https://data.oecd.org/healthrisk/overweight-or-obese-population.htm>].
41. Ünlü TND, D. TÜRKİYE'DE UYGULANAN OBEZİTE POLİTİKALARI; ALMANYA UYGULAMA KARŞILAŞTIRILMASI. *Selçuk Sağlık Dergisi.* 2022;3(1):62-78.

42. Satman I, Omer B, Tutuncu Y, Kalaca S, Gedik S, Dinccag N, et al. Twelve-year trends in the prevalence and risk factors of diabetes and prediabetes in Turkish adults. *Eur J Epidemiol.* 2013;28(2):169-80.
43. Herrera BM, Lindgren CM. The genetics of obesity. *Curr Diab Rep.* 2010;10(6):498-505.
44. Abdelaal M, le Roux CW, Docherty NG. Morbidity and mortality associated with obesity. *Ann Transl Med.* 2017;5(7):161.
45. Neel JV. Diabetes mellitus: a "thrifty" genotype rendered detrimental by "progress"? *Am J Hum Genet.* 1962;14:353-62.
46. Chakravarthy MV, Booth FW. Eating, exercise, and "thrifty" genotypes: connecting the dots toward an evolutionary understanding of modern chronic diseases. *J Appl Physiol (1985).* 2004;96(1):3-10.
47. Molé PA, Oscai LB, Holloszy JO. Adaptation of muscle to exercise. Increase in levels of palmitoyl Coa synthetase, carnitine palmitoyltransferase, and palmitoyl Coa dehydrogenase, and in the capacity to oxidize fatty acids. *J Clin Invest.* 1971;50(11):2323-30.
48. Eaton SB, Konner M, Shostak M. Stone agers in the fast lane: chronic degenerative diseases in evolutionary perspective. *Am J Med.* 1988;84(4):739-49.
49. Speakman JR. Thrifty genes for obesity, an attractive but flawed idea, and an alternative perspective: the 'drifty gene' hypothesis. *Int J Obes (Lond).* 2008;32(11):1611-7.
50. Prentice AM, Hennig BJ, Fulford AJ. Evolutionary origins of the obesity epidemic: natural selection of thrifty genes or genetic drift following predation release? *Int J Obes (Lond).* 2008;32(11):1607-10.
51. Saeb AT, Al-Naqeb D. The Impact of Evolutionary Driving Forces on Human Complex Diseases: A Population Genetics Approach. *Scientifica (Cairo).* 2016;2016:2079704.
52. Esteller-Cucala P, Maceda I, Børghlum AD, Demontis D, Faraone SV, Cormand B, et al. Genomic analysis of the natural history of attention-deficit/hyperactivity disorder using Neanderthal and ancient Homo sapiens samples. *Scientific Reports.* 2020;10(1):8622.
53. Crawford C, Salmon C. Psychopathology or adaptation? Genetic and evolutionary perspectives on individual differences and psychopathology. *Neuro Endocrinol Lett.* 2002;23 Suppl 4:39-45.

54. Gluckman PD, Low FM, Buklijas T, Hanson MA, Beedle AS. How evolutionary principles improve the understanding of human health and disease. *Evol Appl.* 2011;4(2):249-63.
55. Kilinc GM, Omrak A, Ozer F, Gunther T, Buyukkarakaya AM, Bicakci E, et al. The Demographic Development of the First Farmers in Anatolia. *Curr Biol.* 2016;26(19):2659-66.
56. Larsen CS. The agricultural revolution as environmental catastrophe: Implications for health and lifestyle in the Holocene. *Quaternary International.* 2006;150(1):12-20.
57. Latham K. Human Health and the Neolithic Revolution: an Overview of Impacts of the Agricultural Transition on Oral Health, Epidemiology, and the Human Body. 2013.
58. Rees JS, Castellano S, Andrés AM. The Genomics of Human Local Adaptation. *Trends in Genetics.* 2020;36(6):415-28.
59. Savolainen O, Lascoux M, Merila J. Ecological genomics of local adaptation. *Nat Rev Genet.* 2013;14(11):807-20.
60. Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, et al. Genome-wide detection and characterization of positive selection in human populations. *Nature.* 2007;449(7164):913-8.
61. Tishkoff SA, Reed FA, Ranciaro A, Voight BF, Babbitt CC, Silverman JS, et al. Convergent adaptation of human lactase persistence in Africa and Europe. *Nature Genetics.* 2007;39(1):31-40.
62. Zhang C, Li J, Tian L, Lu D, Yuan K, Yuan Y, et al. Differential Natural Selection of Human Zinc Transporter Genes between African and Non-African Populations. *Scientific Reports.* 2015;5(1):9658.
63. Mathieson I, Lazaridis I, Rohland N, Mallick S, Patterson N, Roodenberg SA, et al. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature.* 2015;528(7583):499-503.
64. Wang MH, Okazaki T, Kugathasan S, Cho JH, Isaacs KL, Lewis JD, et al. Contribution of higher risk genes and European admixture to Crohn's disease in African Americans. *Inflamm Bowel Dis.* 2012;18(12):2277-87.
65. Heath KM, Axton JH, McCullough JM, Harris N. The evolutionary adaptation of the C282Y mutation to culture and climate during the European Neolithic. *Am J Phys Anthropol.* 2016;160(1):86-101.

66. Toomajian C, Ajioka RS, Jorde LB, Kushner JP, Kreitman M. A method for detecting recent selection in the human genome from allele age estimates. *Genetics*. 2003;165(1):287-97.
67. Han Y, Gu S, Oota H, Osier MV, Pakstis AJ, Speed WC, et al. Evidence of positive selection on a class I ADH locus. *Am J Hum Genet*. 2007;80(3):441-56.
68. Polimanti R, Yang BZ, Zhao H, Gelernter J. Evidence of Polygenic Adaptation in the Systems Genetics of Anthropometric Traits. *PLoS One*. 2016;11(8):e0160654.
69. Perry GH, Dominy NJ, Claw KG, Lee AS, Fiegler H, Redon R, et al. Diet and the evolution of human amylase gene copy number variation. *Nature Genetics*. 2007;39(10):1256-60.
70. Slatkin M. Isolation by Distance in Equilibrium and Non-Equilibrium Populations. *Evolution*. 1993;47(1):264-79.
71. Rockman MV. Human evolutionary genetics: origins, peoples, and disease (2004). *Human Genetics*. 2004;115(2):167-8.
72. Sexton JP, Hangartner SB, Hoffmann AA. GENETIC ISOLATION BY ENVIRONMENT OR DISTANCE: WHICH PATTERN OF GENE FLOW IS MOST COMMON? *Evolution*. 2014;68(1):1-15.
73. Kimura M, Ohta T. The Average Number of Generations until Fixation of a Mutant Gene in a Finite Population. *Genetics*. 1969;61(3):763-71.
74. Jones AG, Bürger R, Arnold SJ. Epistasis and natural selection shape the mutational architecture of complex traits. *Nat Commun*. 2014;5:3709.
75. Vitti JJ, Grossman SR, Sabeti PC. Detecting Natural Selection in Genomic Data. *Annual Review of Genetics*. 2013;47(1):97-120.
76. Kimura M. Evolutionary rate at the molecular level. *Nature*. 1968;217(5129):624-6.
77. Kimura M. *The Neutral Theory of Molecular Evolution*. Cambridge: Cambridge University Press; 1983.
78. Ayub Q, Moutsianas L, Chen Y, Panoutsopoulou K, Colonna V, Pagani L, et al. Revisiting the thrifty gene hypothesis via 65 loci associated with susceptibility to type 2 diabetes. *Am J Hum Genet*. 2014;94(2):176-85.
79. Marnetto D, Pankratov V, Mondal M, Montinaro F, Parna K, Vallini L, et al. Ancestral genomic contributions to complex traits in contemporary Europeans. *Curr Biol*. 2022;32(6):1412-9 e3.

80. Le MK, Smith OS, Akbari A, Harpak A, Reich D, Narasimhan VM. 1,000 ancient genomes uncover 10,000 years of natural selection in Europe. *bioRxiv*. 2022.
81. Mathieson S, Mathieson I. FADS1 and the Timing of Human Adaptation to Agriculture. *Mol Biol Evol*. 2018;35(12):2957-70.
82. Dominguez-Andres J, Kuijpers Y, Bakker OB, Jaeger M, Xu CJ, Van der Meer JW, et al. Evolution of cytokine production capacity in ancient and modern European populations. *Elife*. 2021;10.
83. Cheng JY, Stern AJ, Racimo F, Nielsen R. Detecting Selection in Multiple Populations by Modeling Ancestral Admixture Components. *Molecular Biology and Evolution*. 2021;39(1).
84. He Z, Dai X, Lyu W, Beaumont M, Yu F. Estimating Temporally Variable Selection Intensity from Ancient DNA Data. *Mol Biol Evol*. 2023;40(3).
85. Dehasque M, Avila-Arcos MC, Diez-Del-Molino D, Fumagalli M, Guschanski K, Lorenzen ED, et al. Inference of natural selection from ancient DNA. *Evol Lett*. 2020;4(2):94-108.
86. Higuchi R, Bowman B, Freiberger M, Ryder OA, Wilson AC. DNA sequences from the quagga, an extinct member of the horse family. *Nature*. 1984;312(5991):282-4.
87. Pääbo S. Molecular cloning of Ancient Egyptian mummy DNA. *Nature*. 1985;314(6012):644-5.
88. Hummel S. Ancient DNA. In: Henke W, Tattersall I, editors. *Handbook of Paleoanthropology: Vol I:Principles, Methods and Approaches Vol II:Primate Evolution and Human Origins Vol III:Phylogeny of Hominids*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2013. p. 1-24.
89. Rizzi E, Lari M, Gigli E, De Bellis G, Caramelli D. Ancient DNA studies: new perspectives on old samples. *Genetics Selection Evolution*. 2012;44(1):21.
90. Krause J, Fu Q, Good JM, Viola B, Shunkov MV, Derevianko AP, et al. The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature*. 2010;464(7290):894-7.
91. Reich D, Green RE, Kircher M, Krause J, Patterson N, Durand EY, et al. Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature*. 2010;468(7327):1053-60.
92. Meyer M, Kircher M, Gansauge MT, Li H, Racimo F, Mallick S, et al. A high-coverage genome sequence from an archaic Denisovan individual. *Science*. 2012;338(6104):222-6.



93. Gravel S, Henn BM, Gutenkunst RN, Indap AR, Marth GT, Clark AG, et al. Demographic history and rare allele sharing among human populations. *Proceedings of the National Academy of Sciences*. 2011;108(29):11983-8.
94. Petr M, Pääbo S, Kelso J, Vernot B. Limits of long-term selection against Neandertal introgression. *Proc Natl Acad Sci U S A*. 2019;116(5):1639-44.
95. Frantz LAF, Bradley DG, Larson G, Orlando L. Animal domestication in the era of ancient genomics. *Nat Rev Genet*. 2020;21(8):449-60.
96. Nilson SM, Gandolfi B, Grahn RA, Kurushima JD, Lipinski MJ, Randi E, et al. Genetics of randomly bred cats support the cradle of cat domestication being in the Near East. *Heredity*. 2022;129(6):346-55.
97. Spyrou MA, Bos KI, Herbig A, Krause J. Ancient pathogen genomics as an emerging tool for infectious disease research. *Nat Rev Genet*. 2019;20(6):323-40.
98. Key FM, Posth C, Esquivel-Gomez LR, Hübler R, Spyrou MA, Neumann GU, et al. Emergence of human-adapted *Salmonella enterica* is linked to the Neolithization process. *Nature Ecology & Evolution*. 2020;4(3):324-33.
99. Klunk J, Vilgalys TP, Demeure CE, Cheng X, Shiratori M, Madej J, et al. Evolution of immune genes is associated with the Black Death. *Nature*. 2022;611(7935):312-9.
100. Warinner C, Rodrigues JF, Vyas R, Trachsel C, Shved N, Grossmann J, et al. Pathogens and host immunity in the ancient human oral cavity. *Nat Genet*. 2014;46(4):336-44.
101. Warinner C, Herbig A, Mann A, Fellows Yates JA, Weiss CL, Burbano HA, et al. A Robust Framework for Microbial Archaeology. *Annu Rev Genomics Hum Genet*. 2017;18:321-56.
102. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet*. 2016;17(6):333-51.
103. Krause J, Briggs AW, Kircher M, Maricic T, Zwyns N, Derevianko A, et al. A complete mtDNA genome of an early modern human from Kostenki, Russia. *Curr Biol*. 2010;20(3):231-6.
104. Skoglund P, Northoff BH, Shunkov MV, Derevianko AP, Paabo S, Krause J, et al. Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proc Natl Acad Sci U S A*. 2014;111(6):2229-34.
105. Renaud G, Schubert M, Sawyer S, Orlando L. Authentication and Assessment of Contamination in Ancient DNA. *Methods Mol Biol*. 2019;1963:163-94.

106. Orlando L, Allaby R, Skoglund P, Der Sarkissian C, Stockhammer PW, Ávila-Arcos MC, et al. Ancient DNA analysis. *Nature Reviews Methods Primers*. 2021;1(1):14.
107. Skoglund P, Sjödin P, Skoglund T, Lascoux M, Jakobsson M. Investigating population history using temporal genetic differentiation. *Mol Biol Evol*. 2014;31(9):2516-27.
108. Lazaridis I, Patterson N, Mittnik A, Renaud G, Mallick S, Kirsanow K, et al. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature*. 2014;513(7518):409-13.
109. Günther T, Malmström H, Svensson EM, Omrak A, Sánchez-Quinto F, Kilinç GM, et al. Population genomics of Mesolithic Scandinavia: Investigating early postglacial migration routes and high-latitude adaptation. *PLoS Biol*. 2018;16(1):e2003703.
110. Maixner F, Mitterer C, Jäger HY, Sarhan MS, Valverde G, Lücken S, et al. Linear polyacrylamide is highly efficient in precipitating and purifying environmental and ancient DNA. *Methods in Ecology and Evolution*. 2022;13(3):653-67.
111. Rohland N, Mallick S, Mah M, Maier R, Patterson N, Reich D. Three assays for in-solution enrichment of ancient human DNA at more than a million SNPs. *Genome Res*. 2022;32(11-12):2068-78.
112. Au - Parker CE, Au - Bos KI, Au - Haak W, Au - Krause J. Optimized Bone Sampling Protocols for the Retrieval of Ancient DNA from Archaeological Remains. *JoVE*. 2021(177):e63250.
113. Diamond J. Evolution, consequences and future of plant and animal domestication. *Nature*. 2002;418(6898):700-7.
114. Armelagos GJ. Human evolution and the evolution of disease. *Ethn Dis*. 1991;1(1):21-5.
115. Ulijaszek SJ. Human dietary change. *Philos Trans R Soc Lond B Biol Sci*. 1991;334(1270):271-8; discussion 8-9.
116. Özdoğan M. Archaeological Evidence on the Westward Expansion of Farming Communities from Eastern Anatolia to the Aegean and the Balkans. *Current Anthropology*. 2011;52(S4):S415-S30.
117. Altınışık NE, Kazancı DD, Aydoğan A, Gemici HC, Erdal ÖD, Sarıaltun S, et al. A genomic snapshot of demographic and cultural dynamism in Upper Mesopotamia during the Neolithic Transition. *Science Advances*. 2022;8(44):eabo3609.

118. Kilinc GM, Koptekin D, Atakuman C, Sumer AP, Donertas HM, Yaka R, et al. Archaeogenomic analysis of the first steps of Neolithization in Anatolia and the Aegean. *Proc Biol Sci.* 2017;284(1867).
119. Koptekin D, Yüncü E, Rodríguez-Varela R, Altınışık NE, Psonis N, Kashuba N, et al. Spatial and temporal heterogeneity in human mobility patterns in Holocene Southwest Asia and the East Mediterranean. *Current Biology.* 2023;33(1):41-57.e15.
120. Yaka R, Mapelli I, Kaptan D, Dogu A, Chylenski M, Erdal OD, et al. Variable kinship patterns in Neolithic Anatolia revealed by ancient genomes. *Curr Biol.* 2021;31(11):2455-68 e18.
121. Lazaridis I, Alpaslan-Roodenberg S, Acar A, Açikkol A, Agelarakis A, Aghikyan L, et al. The genetic history of the Southern Arc: A bridge between West Asia and Europe. *Science.* 2022;377(6609):eabm4247.
122. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81(3):559-75.
123. Lowy-Gallego E, Fairley S, Zheng-Bradley X, Ruffier M, Clarke L, Flicek P. Variant calling on the GRCh38 assembly with the data from phase three of the 1000 Genomes Project. *Wellcome Open Res.* 2019;4:50.
124. Chen Z, Boehnke M, Wen X, Mukherjee B. Revisiting the genome-wide significance threshold for common variant GWAS. *G3 (Bethesda).* 2021;11(2).
125. Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science.* 1996;273(5281):1516-7.
126. Quinlan AR. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr Protoc Bioinformatics.* 2014;47:11 2 1-34.
127. de Barros Damgaard P, Martiniano R, Kamm J, Moreno-Mayar JV, Kroonen G, Peyrot M, et al. The first horse herders and the impact of early Bronze Age steppe expansions into Asia. *Science.* 2018;360(6396).
128. Feldman M, Fernandez-Dominguez E, Reynolds L, Baird D, Pearson J, Hershkovitz I, et al. Late Pleistocene human genome suggests a local origin for the first farmers of central Anatolia. *Nat Commun.* 2019;10(1):1218.
129. Hofmanova Z, Kreutzer S, Hellenthal G, Sell C, Diekmann Y, Diez-Del-Molino D, et al. Early farmers from across Europe directly descended from Neolithic Aegeans. *Proc Natl Acad Sci U S A.* 2016;113(25):6886-91.
130. Lazaridis I, Mittnik A, Patterson N, Mallick S, Rohland N, Pfrenkle S, et al. Genetic origins of the Minoans and Mycenaeans. *Nature.* 2017;548(7666):214-8.

131. Marchi N, Winkelbach L, Schulz I, Brami M, Hofmanova Z, Blocher J, et al. The genomic origins of the world's first farmers. *Cell*. 2022;185(11):1842-59 e18.
132. Omrak A, Gunther T, Valdiosera C, Svensson EM, Malmstrom H, Kieseewetter H, et al. Genomic Evidence Establishes Anatolia as the Source of the European Neolithic Gene Pool. *Curr Biol*. 2016;26(2):270-5.
133. Skourtanioti E, Erdal YS, Frangipane M, Balossi Restelli F, Yener KA, Pinnock F, et al. Genomic History of Neolithic to Bronze Age Anatolia, Northern Levant, and Southern Caucasus. *Cell*. 2020;181(5):1158-75.e28.
134. Alkan C, Kavak P, Somel M, Gokcumen O, Ugurlu S, Saygi C, et al. Whole genome sequencing of Turkish genomes reveals functional private alleles and impact of genetic interactions with Europe, Asia and Africa. *BMC Genomics*. 2014;15(1):963.
135. Lindgreen S. AdapterRemoval: easy cleaning of next-generation sequencing reads. *BMC Res Notes*. 2012;5:337.
136. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754-60.
137. Bergström A, McCarthy SA, Hui R, Almarri MA, Ayub Q, Danecek P, et al. Insights into human genetic variation and population history from 929 diverse genomes. *Science*. 2020;367(6484):eaay5012.
138. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078-9.
139. Zhang H. Overview of Sequence Data Formats. In: Mathé E, Davis S, editors. *Statistical Genomics: Methods and Protocols*. New York, NY: Springer New York; 2016. p. 3-17.
140. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297-303.
141. Korneliussen TS, Albrechtsen A, Nielsen R. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics*. 2014;15:356.
142. Reynolds J, Weir BS, Cockerham CC. Estimation of the coancestry coefficient: basis for a short-term genetic distance. *Genetics*. 1983;105(3):767-79.
143. Berg JJ, Coop G. A population genetic signal of polygenic adaptation. *PLoS Genet*. 2014;10(8):e1004412.
144. Bhatia G, Patterson N, Sankararaman S, Price AL. Estimating and interpreting FST: the impact of rare variants. *Genome Res*. 2013;23(9):1514-21.

145. Wright S. THE INTERPRETATION OF POPULATION STRUCTURE BY F-STATISTICS WITH SPECIAL REGARD TO SYSTEMS OF MATING. *Evolution*. 1965;19(3):395-420.
146. Weir BS, Cockerham CC. Estimating F-Statistics for the Analysis of Population Structure. *Evolution*. 1984;38(6):1358-70.
147. Nei M. Definition and Estimation of Fixation Indices. *Evolution*. 1986;40(3):643-5.
148. Hudson RR, Kreitman M, Aguade M. A test of neutral molecular evolution based on nucleotide data. *Genetics*. 1987;116(1):153-9.
149. Gopalakrishnan S, Ebenesersdottir SS, Lundstrom IKC, Turner-Walker G, Moore KHS, Luisi P, et al. The population genomic legacy of the second plague pandemic. *Curr Biol*. 2022;32(21):4743-51 e6.
150. Nielsen R, Korneliussen T, Albrechtsen A, Li Y, Wang J. SNP calling, genotype calling, and sample allele frequency estimation from New-Generation Sequencing data. *PLoS One*. 2012;7(7):e37558.
151. Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZX, Pool JE, et al. Sequencing of 50 human exomes reveals adaptation to high altitude. *Science*. 2010;329(5987):75-8.
152. Bergström A, McCarthy SA, Hui R, Almarri MA, Ayub Q, Danecek P, et al. Insights into human genetic variation and population history from 929 diverse genomes. *Science*. 2020;367(6484).
153. Haller BC, Galloway J, Kelleher J, Messer PW, Ralph PL. Tree-sequence recording in SLiM opens new horizons for forward-time simulation of whole genomes. *Mol Ecol Resour*. 2019;19(2):552-66.
154. Lohmueller KE, Albrechtsen A, Li Y, Kim SY, Korneliussen T, Vinckenbosch N, et al. Natural selection affects multiple aspects of genetic variation at putatively neutral sites across the human genome. *PLoS Genet*. 2011;7(10):e1002326.
155. Cuadros-Espinoza S, Laval G, Quintana-Murci L, Patin E. The genomic signatures of natural selection in admixed human populations. *The American Journal of Human Genetics*. 2022;109(4):710-26.
156. Peng B, Amos CI, Kimmel M. Forward-time simulations of human populations with complex diseases. *PLoS Genet*. 2007;3(3):e47.
157. Vernot B, Akey JM. Resurrecting surviving Neandertal lineages from modern human genomes. *Science*. 2014;343(6174):1017-21.
158. Harris K, Nielsen R. The Genetic Cost of Neanderthal Introgression. *Genetics*. 2016;203(2):881-91.

159. Allentoft ME, Sikora M, Refoyo-Martínez A, Irving-Pease EK, Fischer A, Barrie W, et al. Population Genomics of Stone Age Eurasia. *bioRxiv*. 2022:2022.05.04.490594.
160. Chintalapati M, Patterson N, Moorjani P. The spatiotemporal patterns of major human admixture events during the European Holocene. *eLife*. 2022;11:e77625.
161. Mi H, Muruganujan A, Casagrande JT, Thomas PD. Large-scale gene function analysis with the PANTHER classification system. *Nature Protocols*. 2013;8(8):1551-66.
162. Myles S, Hradetzky E, Engelken J, Lao O, Nürnberg P, Trent RJ, et al. Identification of a candidate genetic variant for the high prevalence of type II diabetes in Polynesians. *Eur J Hum Genet*. 2007;15(5):584-9.
163. Stephan W. Signatures of positive selection: from selective sweeps at individual loci to subtle allele frequency changes in polygenic adaptation. *Mol Ecol*. 2016;25(1):79-88.
164. Vujkovic M, Keaton JM, Lynch JA, Miller DR, Zhou J, Tcheandjieu C, et al. Discovery of 318 new risk loci for type 2 diabetes and related vascular outcomes among 1.4 million participants in a multi-ancestry meta-analysis. *Nat Genet*. 2020;52(7):680-91.
165. Helgason A, Palsson S, Thorleifsson G, Grant SF, Emilsson V, Gunnarsdottir S, et al. Refining the impact of TCF7L2 gene variants on type 2 diabetes and adaptive evolution. *Nat Genet*. 2007;39(2):218-25.
166. Wang G, Speakman JR. Analysis of Positive Selection at Single Nucleotide Polymorphisms Associated with Body Mass Index Does Not Support the "Thrifty Gene" Hypothesis. *Cell Metab*. 2016;24(4):531-41.
167. Segurel L, Austerlitz F, Toupance B, Gautier M, Kelley JL, Pasquet P, et al. Positive selection of protective variants for type 2 diabetes from the Neolithic onward: a case study in Central Asia. *Eur J Hum Genet*. 2013;21(10):1146-51.
168. Klimentidis YC, Abrams M, Wang J, Fernandez JR, Allison DB. Natural selection at genomic regions associated with obesity and type-2 diabetes: East Asians and sub-Saharan Africans exhibit high levels of differentiation at type-2 diabetes regions. *Hum Genet*. 2011;129(4):407-18.
169. Meeks KAC, Bentley AR, Adeyemo AA, Rotimi CN. Evolutionary forces in diabetes and hypertension pathogenesis in Africans. *Hum Mol Genet*. 2021;30(R1):R110-r8.

170. Childebayeva A, Rohrlach AB, Barquera R, Rivollat M, Aron F, Szolek A, et al. Population Genetics and Signatures of Selection in Early Neolithic European Farmers. *Mol Biol Evol.* 2022;39(6).
171. Mathieson I. Estimating time-varying selection coefficients from time series data of allele frequencies. *bioRxiv.* 2020:2020.11.17.387761.
172. Fumagalli M, Vieira FG, Korneliussen TS, Linderoth T, Huerta-Sánchez E, Albrechtsen A, et al. Quantifying population genetic differentiation from next-generation sequencing data. *Genetics.* 2013;195(3):979-92.
173. Tam V, Patel N, Turcotte M, Bosse Y, Pare G, Meyre D. Benefits and limitations of genome-wide association studies. *Nat Rev Genet.* 2019;20(8):467-84.
174. Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature.* 2007;445(7130):881-5.
175. Unoki H, Takahashi A, Kawaguchi T, Hara K, Horikoshi M, Andersen G, et al. SNPs in *KCNQ1* are associated with susceptibility to type 2 diabetes in East Asian and European populations. *Nat Genet.* 2008;40(9):1098-102.
176. Yasuda K, Miyake K, Horikawa Y, Hara K, Osawa H, Furuta H, et al. Variants in *KCNQ1* are associated with susceptibility to type 2 diabetes mellitus. *Nat Genet.* 2008;40(9):1092-7.
177. Consortium STD, Williams AL, Jacobs SB, Moreno-Macias H, Huerta-Chagoya A, Churchhouse C, et al. Sequence variants in *SLC16A11* are a common risk factor for type 2 diabetes in Mexico. *Nature.* 2014;506(7486):97-101.
178. Moltke I, Grarup N, Jorgensen ME, Bjerregaard P, Treebak JT, Fumagalli M, et al. A common Greenlandic *TBC1D4* variant confers muscle insulin resistance and type 2 diabetes. *Nature.* 2014;512(7513):190-3.
179. Hancock AM, Witonsky DB, Gordon AS, Eshel G, Pritchard JK, Coop G, et al. Adaptations to climate in candidate genes for common metabolic disorders. *PLoS Genet.* 2008;4(2):e32.
180. Barton AR, Santander CG, Skoglund P, Moltke I, Reich D, Mathieson I. Insufficient evidence for natural selection associated with the Black Death. *bioRxiv.* 2023.
181. Fer E. Testing natural selection on polygenic trait-associated alleles in anatolia using neolithic and present-day human genomes [Postgraduate]: Middle East Technical University; 2019.

## 8. EKLER

## EK-1: Tez çalışmasının etik kurul izni



T.C.  
HACETTEPE ÜNİVERSİTESİ  
Girişimsel Olmayan Klinik Araştırmalar Etik Kurulu

Sayı : 16969557 - 1938

Konu :

05.10.2021

Dr. Öğr. Üyesi İdil YET  
Sağlık Bilimleri Enstitüsü  
Biyoinformatik Anabilim Dalı  
Öğretim Üyesi

Sayın Dr. Öğr. Üyesi YET,

Kurulumuza değerlendirilmek üzere sunduğunuz GO 21/1060 kayıt numaralı ve "*Kompleks Hastalıklarla İlişkili Genetik Varyasyonun Zaman Serisi Genom Dizileme Verilerinde Popülasyon Genetiği Yaklaşımları ile İncelenmesi*" başlıklı proje Kurulumuzun 05.10.2021 tarihli toplantısında değerlendirilmiş olup, çalışmanın erişime açık veri tabanlarından veri toplanması yolu ile yapılacağı görülmüştür. Gönüllü insanlar üzerinde gerçekleştirilecek nitelikte olmayan bu tip çalışmalar Etik Kurulların kapsamı dışında kalmaktadır.

Bu yazı ilgili protokolün bilimsel ve etik açıdan incelendiğini belirtmek için Etik Kurul kararı yerine geçmek üzere hazırlanmıştır.

Prof. Dr. G. Burça AYDIN  
Başkan

EK :  
Toplantı Katılım Tutanağı.



**EK-2: Tez çalışmasının orijinallik raporu**



## Dijital Makbuz

Bu makbuz ödevinizin Turnitin'e ulaştığını bildirmektedir. Gönderiminize dair bilgiler şöyledir:

Gönderinizin ilk sayfası aşağıda gönderilmektedir.

Gönderen:	İdil Taç
Ödev başlığı:	KOMPLEKS HASTALIKLARLA İLİŞKİLİ GENETİK VARYASYONUN ...
Gönderi Başlığı:	KOMPLEKS HASTALIKLARLA İLİŞKİLİ GENETİK VARYASYONUN ...
Dosya adı:	I_diiTac__Yu_ksekLisansTez_Du_zeltmeler_turnitine_yu_kleme...
Dosya boyutu:	7.83M
Sayfa sayısı:	68
Kelime sayısı:	13,906
Karakter sayısı:	91,278
Gönderim Tarihi:	19-Haz-2023 04:50ÖS (UTC+0300)
Gönderim Numarası:	2119061635



Copyright 2023 Turnitin. Tüm hakları saklıdır.

## KOMPLEKS HASTALIKLARLA İLİŞKİLİ GENETİK VARYASYONUN ZAMAN-SERİSİ GENOM DİZİLEME VERİLERİNDE POPÜLASYON GENETİĞİ YAKLAŞIMLARI İLE İNCELENMESİ

### ORJİNALLİK RAPORU

% <b>4</b>	% <b>4</b>	% <b>1</b>	% <b>1</b>
BENZERLİK ENDEKSİ	İNTERNET KAYNAKLARI	YAYINLAR	ÖĞRENCİ ÖDEVLERİ

### BİRİNCİL KAYNAKLAR

<b>1</b>	<b>mediplatform.com</b> İnternet Kaynağı	% <b>1</b>
<b>2</b>	<b>arkadasimdiyabet.com</b> İnternet Kaynağı	% <b>1</b>
<b>3</b>	<b>"SÖZLÜ SUNUM ÖZETLERİ / ORAL PRESENTATION ABSTRACTS", Turkish Journal of Biochemistry, 2017</b> Yayın	<% <b>1</b>
<b>4</b>	<b>www.biorxiv.org</b> İnternet Kaynağı	<% <b>1</b>
<b>5</b>	<b>lab.hacettepe.edu.tr</b> İnternet Kaynağı	<% <b>1</b>
<b>6</b>	<b>github.com</b> İnternet Kaynağı	<% <b>1</b>
<b>7</b>	<b>www.popgen.dk</b> İnternet Kaynağı	<% <b>1</b>
<b>8</b>	<b>www.medrxiv.org</b> İnternet Kaynağı	

**EK-3:** Nötral senaryo altında Anadolu modellemesi için SLiM programında yürütülen komutlar.

```

initialize() {
    // 10 Mb segmenti simüle edin
    defineConstant("L", 10000000);

    //mutasyon oranını ayarlayın
    defineConstant("mu", 1e-8);

    //Nükleotid bazlı simülasyon
    initializeSLiMOptions(nucleotideBased=T);

    //yukarıda tanımlanan "L" uzunluğundaki ACGT'lerin rastgele dizisini oluşturun
    initializeAncestralNucleotides(randomNucleotides(L));

    //Nükleotid bazlı nötral mutasyonların tanıtılması -> baskınlık katsayısı (İng.
    dominance coeff) 0,5'tir, DFE sabit uygunluk (İng. fitness effect) etkisine
    ayarlanmıştır, seçilim katsayısı sıfır -> nötral

    initializeMutationTypeNuc("m1", 0.5, "f", 0.0);

    m1.convertToSubstitution = F;

    initializeGenomicElementType("g1", m1, 1.0, mmJukesCantor(mu));

    initializeGenomicElement(g1, 0, L-1);

    //rekombinasyon oranını ayarlayın
    initializeRecombinationRate(1e-8);
}

// nesil süresi = 25 yıl

// p1= modern insanlar ve Neandertaller öncesinde, 70.000 nesillik bir başlangıç
(burn-in) dönemi

1 early() {
    sim.addSubpop("p1", 10000);
}

```

```

}

// p2=Neandertal, p3 şu anda tüm Modern İnsanların atasını temsil etmektedir,
// İnsan ve Neandertal ayrılma zamanı 500.000 yıl önce (20000 nesil), Neandertallerin
// ve modern insanların Ne'si 1.000 ve 10.000 sabit değerlerine ayarlanmıştır

50000 early() {
    sim.addSubpopSplit("p2",1000,p1); // p2= Neandertal, p1= Root
    sim.addSubpopSplit("p3",10000,p1); // p3= Modern Humans, p1= Root
}

// Afrikalı ve Afrikalı olmayan popülasyonların 60 kya'da ayrılması, p4 günümüz
// Afrika'nın atasını temsil ediyor, p5= Afrikalı olmayan, yaklaşık 55 kya 2400 gen., p4=
// Afrika, p3= Modern İnsanlar , p5= Afrikalı olmayan

67600 early() {
    sim.addSubpopSplit("p4",10000,p3);
    sim.addSubpopSplit("p5",10000,p3);
}

// Ayrılmanın ardından, Afrikalı olmayan popülasyon N = 1861, 5 kya nesil - 200 nesil
// ile bir darboğaz yaşadı.

67800 early() {
    p5.setSubpopulationSize(1861);
}

// Afrikalı olmayan Neandertal göçleri 50 kya= 2000 nesil önce başlar, p5=Afrikalı
// olmayan, p2=Neandertal

68000 early() {
    p5.setMigrationRates(p2,0.03);
}

// p6= Bazal Avrupalı, Bazal Avrupa bölünmesinin zamanı 1740 nesil, Anadolu
// Çiftçileri ve CHG ayrışmasının zamanı 928 nesil, WHG ve EHG ayrılmasının zamanı
// 696 nesil önce, p6=Bazal Avrupalı, p5=Afrikalı olmayan

```

```

68260 early() {
    sim.addSubpopSplit("p6",5000,p5);
}

68800 early(){
// Afrika dışına Neandertal göçleri 30 kya= 1200 nesil önce sona erdi, Neandertal
popülasyonunu çıkarın

    p5.setMigrationRates(p2,0);

    p2.setSubpopulationSize(0);
}

//p7= CHG 10.000 kişi. Anadolu Neolitik ve CHG ayrılma zamanı 928 gen., p7=CHG,
p8=Anadolu Çiftçisinin Atasının Atası

69072 early() {
    sim.addSubpopSplit("p7",10000,p6);
    sim.addSubpopSplit("p8",10000,p6);
}

// p9 WHG popülasyon büyüklüğü 10.000, WHG ve EHG'nin 696 nesil önce ayrıldığı
zamanı, p10= EHG pop. büyüklüğünü 10.000 temsil eder.

69304 early() {
    sim.addSubpopSplit("p9",10000,p6);
    sim.addSubpopSplit("p10",10000,p6);
}

//Anadolu Neolitike 518 nesil önce (MÖ 10899) karışım eklendi, s7= ÇHG, s8=
Anadolu Çiftçisinin Atası, s9 = WHG

//Anadolu Neolitike

69482 early() {
p8.setMigrationRates(c(p9, p7), c(0.274, 0.387));
}

```

```
//Anadolu Kalkolitik 340 kuşak önce karışım eklendi(8500 BP), p7= CHG

//Anadolu Kalkolitik

69660 early() {

p8.setMigrationRates(p7, 0.329);

}

//Anadolu Neolitik 10000- 6000 BP = 400-240 gen = 69600-69760 için vcf dosyası
oluşturun

69600:69760 late() {

    p8.outputVCFsample(100, replace=T,
filePath="p8.69600_69760.Scenario1a.vcf", simplifyNucleotides=T); }

//Anadolu Neolitik Sonrası için vcf dosyası oluşturun 5200-3200 BP= 208 gen-128
gen = 69792 - 69892 gen

69792:69892 late() {

    p8.outputVCFsample(100, replace=T,
filePath="p8.69792_69892.Scenario1a.vcf", simplifyNucleotides=T); }

//Anadolu günümüz 0 BP 70000 nesli için vcf dosyası oluşturun

70000 late() {

    p8.outputVCFsample(100, replace=T, filePath="p8.70000.Scenario1a.vcf",
simplifyNucleotides=T); }
```

## 9. ÖZGEÇMİŞ