

T.C.  
HACETTEPE ÜNİVERSİTESİ  
SAĞLIK BİLİMLERİ ENSTİTÜSÜ

GENOTİP İMPÜTASYONUNUN  
ANTİK ANADOLU İNSAN GENOMLARINDA  
DEĞERLENDİRİLMESİ

Hande ÇUBUKCU

Biyoinformatik Programı  
YÜKSEK LİSANS TEZİ

ANKARA

2024



T.C.  
HACETTEPE ÜNİVERSİTESİ  
SAĞLIK BİLİMLERİ ENSTİTÜSÜ

GENOTİP İMPÜTASYONUNUN  
ANTİK ANADOLU İNSAN GENOMLARINDA  
DEĞERLENDİRİLMESİ

Hande ÇUBUKCU

Biyoinformatik Programı  
YÜKSEK LİSANS TEZİ

TEZ DANIŞMANI  
Dr. Öğr. Üyesi Gülşah Merve KILINÇ

ANKARA

2024

# Genotip İmpütasyonunun Antik Anadolu İnsan Genomlarında Değerlendirilmesi

**Öğrenci: Hande ÇUBUKCU**

**Danışman: Dr. Öğretim Üyesi Gülşah Merve KILINÇ**

Bu tez çalışması 16/05/2024 tarihinde jürimiz tarafından  
“Biyoinformatik Programı”nda yüksek lisans tezi olarak kabul  
edilmiştir.

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

**Tez Danışmanı: Dr. Öğretim Üyesi Gülşah Merve KILINÇ**  
( *Hacettepe Üniversitesi* )

**Üye: Dr. Öğretim Üyesi İdil YET**  
( *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 tezim aşağıda belirtilen koşullar haricince YÖK Ulusal Tez Merkezi / H.Ü. Kütüphaneleri Açık Erişim Sisteminde erişime açılır.

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

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

Hande ÇUBUKCU

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

(1) Madde 6. 1. Lisansüstü teze 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 metodların kullanıldığı, henüz makaleye dönüşmemiş veya patent gibi yöntemlerle korunmamış ve internette 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 çalışmadaki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi, görsel, işitsel ve yazılı tüm bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu, kullandığım verilerde herhangi bir tahrifat yapmadığımı, yararlandığım kaynaklara bilimsel normlara uygun olarak atıfta bulunduğumu, tezimin kaynak gösterilen durumlar dışında özgün olduğunu, Dr. Öğretim Üyesi Gülşah Merve Kılınç danışmanlığında tarafımdan üretildiğini ve Hacettepe Üniversitesi Sağlık Bilimleri Enstitüsü Tez Yazım Yönergesine göre yazıldığını beyan ederim.

Hande ÇUBUKCU

## TEŞEKKÜR

Yüksek lisans eğitimim süresince bana her konuda yardımcı olan, yol gösteren ve araştırma imkânı sağlayan değerli danışmanım Dr. Öğr. Üyesi Gülşah Merve Kılınç'a, süreç boyunca yardımları ve desteği için değerli hocam Dr. Öğr. Üyesi İdil Yet'e, araştırmalarım hakkında soruları, önerileri ve desteği için değerli hocam Prof. Dr. Mehmet Somel'e, çalışmalarına sundukları katkılar için CompEvo ekibine, tüm yardımları için Menekşe Tarla'ya, biyoinformatik bölümüne ve öğrencilerine,

Bana her zaman destek olan sevgili anneme,

Yüksek lisans eğitimim sırasında sağladıkları burs ve araştırma deneyimi için TÜBİTAK BİDEB 2210-A ve TÜBİTAK ARDEB-3501 (Yürütücü: Dr. Öğretim Üyesi Gülşah Merve Kılınç, Proje Nu:121Z025) programlarına teşekkür ederim.

## ÖZET

**Cubukcu, H., Genotip İmpütasyonunun Antik Anadolu İnsan Genomlarında Değerlendirilmesi, Hacettepe Üniversitesi Sağlık Bilimleri Enstitüsü Biyoinformatik Programı Yüksek Lisans Tezi, Ankara, 2024.** Genotip impütasyonu düşük derinlikli antik DNA verilerinde yüksek miktardaki eksik bilgiye bir çözüm sunmaktadır. Bu tezde, fazlama ve impütasyon metodu GLIMPSE kullanarak, otozomal biallelik TNP'lerde impütasyonun doğruluğu ve kayıp miktarları, Anadolu ve Batı Avrasya'dan on yüksek derinlikli ve yüzün üzerinde düşük derinlikli örnek ile dört farklı genom derinliğinde farklı filtreler kullanılarak değerlendirilmiştir. İmpütasyon verisini sıklıkla kullanılan popülasyon genetiği analizlerinde kullanmanın etkisi araştırılmıştır. Yüksek kaliteli veri ile genotip uyumluluğu olarak ölçülen impütasyon doğruluğu, 0.1x derinliğinde yaygın varyantlarda %80 ve %99 genotip olasılığı filtreleri ile sırasıyla %70'e ve %90'a ulaşırken tüm varyantların %5'ten ve %15'ten azı kaybedilmektedir. Yaygın kullanılan Human Origins veriseti pozisyonlarında 0.1x verisi için yaygın varyantlarda kayıp oranları aynı filtreler ile %20 ve %50'ye ulaşmaktadır. Temel bileşenler analizi ve f4 istatistiklerine göre sözde-haploid veride daha fazla referans yanlılığı görülmektedir, bu da impütasyonun referans yanlılığını azalttığına işaret etmektedir. Diğer yandan, impütasyon verisinde genom derinliği düştükçe referans panel popülasyonlarına eğilim görülmektedir. Alel frekans hesabı karşılaştırmalarına göre, Anadolu Neolitik dönem örneklerinde impütasyon veya hizalama verisinden hesaplanan sıklık değerleri arasında yaygın varyantlarda yüzde 15'e kadar fark görülebilmektedir. Elde edilen sonuçlara göre, ilgi alanı bölgeden düşük derinlikli antik genom örneklerinde impütasyon kullanarak yaygın varyantlar için güvenilir sonuçlar almak mümkündür.

**Anahtar Kelimeler:** antik DNA, genotip impütasyonu, lc-WGS, referans yanlılığı

Bu tez TÜBİTAK BİDEB 2210-A ve TÜBİTAK ARDEB 3501 (121Z025) programları tarafından desteklenmiştir.



## ABSTRACT

**Cubukcu, H., Evaluation of Genotype Imputation on Ancient Anatolian Human Genomes, Hacettepe University Graduate School of Health Sciences Department of Bioinformatics Master's Thesis, Ankara 2024.** Genotype imputation presents a solution to the high amount of missing information in low coverage ancient DNA samples. In this thesis, by using the phasing and imputation tool GLIMPSE, the accuracy of genotype imputation and the amount of missing variants on autosomal biallelic SNPs are evaluated using different genotype posterior probability filters at four different coverages using ten high coverage and over a hundred low coverage genomes from Anatolia and West Eurasia. The effect of using imputed data in frequently used population genetics analyses is investigated. The accuracy of imputation measured as genotype concordance with high coverage data reaches 70% to 90% in common variants at 0.1x genome coverage while losing less than 5% to 15% of all variants using filters between 80% and 99%. Missing amount in common variants using the same data reaches %20 and %50 in commonly used Human Origins dataset positions. Pseudohaploid data have more affinity towards the reference genome compared to the imputed data, pointing to mitigation of reference bias by imputation. On the other hand, imputed data show tendency towards the reference panel populations as the coverage decreases. Allele frequency calculations using imputation and alignment data show up to 15% difference in low coverage Anatolian Neolithic samples. Results show that it is possible to get reliable results using imputed data for common variants for low coverage ancient genome samples from the region of interest.

**Keywords:** ancient DNA, genotype imputation, lc-WGS, reference bias

This thesis has been supported by TÜBİTAK BİDEB 2210-A and TÜBİTAK ARDEB 3501 (121Z025) programs.

## 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	xiii
TABLolar	xv
<b>1. GİRİŞ</b>	<b>1</b>
<b>2. GENEL BİLGİLER</b>	<b>3</b>
2.1 Antik DNA Çalışmaları	3
2.2 Antik DNA Verilerinde Karşılaşılan Bazı Sorunlar	4
2.3 Genotip İmpütasyonu	5
2.3.1 Referans paneller	6
2.3.2 Antik DNA Verisi ile Genotip İmpütasyonu Çalışmaları	7
2.3.3 Saklı Markov Modelleri ve GLIMPSE	7
2.3.4 Genotip İmpütasyonu Doğruluğu Ölçüleri	8

2.4 Anadolu'dan Elde Edilen Antik İnsan Genetik Verisi	9
2.5 Bu Çalışmanın Kapsamı	10
2.5.1 Çalışmada Kullanılan Veri Kaynakları	10
<b>3. GEREÇ VE YÖNTEM</b>	<b>12</b>
3.1 Veriler	12
3.2 Ön-işleme	14
3.3 Veri Hazırlama	15
3.4 Genotip İmpütasyonu	16
3.5 Genotipleme	16
3.6 Karşılaştırma Veriseti	17
3.7 Temel Bileşenler Analizi	17
3.8 F-İstatistikleri	18
3.9 Alel Sıklığı Hesabı Karşılaştırması	18
<b>4. BULGULAR</b>	<b>21</b>
4.1 Doğruluğun değerlendirilmesi	21
4.2 Kayıp TNP sayıları	25
4.3 Genotip İmpütasyonunun TBA Sonucuna Etkisi	25
4.3.1 Modern örnekler için TBA analizi	27
4.4 F-İstatistikleri ve Referans Yanlılığı	28
4.5 Düşük Derinlikte Antik Genomların İmpütasyonu	30
4.6 Alel Sıklık Değerleri Karşılaştırması	33

<b>5. TARTIŞMA</b>	36
<b>6. SONUÇ VE ÖNERİLER</b>	40
<b>7. KAYNAKLAR</b>	42
<b>8. EKLER</b>	
<b>EK-1:</b> Tez Çalışması ile İlgili Etik Kurul İzni	
<b>EK-2:</b> Tez Çalışması Orijinallik Raporu	
<b>EK-3:</b> AKT16 ve BAR25 Örnekleri için Doğruluk Grafikleri	
<b>EK-4:</b> Ek TBA Görselleri	
<b>EK-5:</b> Yüksek Derinlikli İmpütasyon Örnekleri için Kayıp Tablosu – 1	
<b>EK-6:</b> Yüksek Derinlikli İmpütasyon Örnekleri için Kayıp Tablosu – 2	
<b>EK-7:</b> Düşük Derinlikli Örnekler için Bilgi ve Kayıp Tablosu	
<b>EK-8:</b> Yeni Üretilen Veri için Bilgi ve Kayıp Tablosu	
<b>EK-9:</b> Alel Sıklık Hesapları Tablosu – Yaygın Varyantlar	
<b>EK-10:</b> Alel Sıklık Hesapları Tablosu – Az Yaygın Varyantlar	
<b>9. ÖZGEÇMİŞ</b>	62

## SİMGELER VE KISALTMALAR

<b>BAM</b>	<i>Binary Alignment Map</i>
<b>BCF</b>	İkili Çağrı Formatı, <i>Binary Call Format</i>
<b>DNA</b>	Deoksiribonükleik Asit
<b>GL</b>	Genotip Olabilirliği, <i>Genotype Likelihood</i>
<b>GLIMPSE</b>	<i>Genotype Likelihoods Imputation and Phasing Method</i>
<b>GP</b>	Ardıl Genotip Olasılığı, <i>Posterior Genotype Probability</i>
<b>GWAS</b>	Genom Boyu İlişkilendirme Çalışması, <i>Genome Wide Association Study</i>
<b>HapMap</b>	Uluslararası Haplotip Haritası Projesi, <i>International Haplotype Map Project</i>
<b>HGDP</b>	İnsan Genom Çeşitliliği Projesi, <i>Human Genome Diversity Project</i>
<b>HMM</b>	Saklı Markov Modeli, <i>Hidden Markov Model</i>
<b>HRC</b>	Haplotip Referans Konsorsiyumu, <i>Haplotype Reference Consortium</i>
<b>KGP</b>	Bin Genom Projesi, <i>Thousand Genomes Project</i>
<b>LC-WGS</b>	Düşük Derinlikli Tüm Genom Dizileme, <i>Low Coverage Whole Genome Sequencing</i>
<b>MAF</b>	Minör Alel Sıklığı, <i>Minor Allele Frequency</i>
<b>NRD</b>	Referans Dışı Uyumsuzluk, <i>Non Reference Discordance</i>
<b>PBWT</b>	<i>Positional Burrows-Wheeler Transform</i>
<b>PCR</b>	Polimeraz Zincir Tepkimesi, <i>Polymerase Chain Reaction</i>
<b>SG</b>	Rastgele (Dizileme), <i>Shotgun (Sequencing)</i>
<b>TNP</b>	Tek Nükleotit Poliformizmi, <i>SNP, Single Nucleotide Polymorphism</i>
<b>TBA</b>	Temel Bileşenler Analizi, <i>Principal Components Analysis</i>
<b>TOPMed</b>	<i>Trans Omics for Precision Medicine</i>
<b>TÜBİTAK</b>	Türkiye Bilimsel ve Teknik Araştırma Kurumu
<b>VCF</b>	Varyant Çağrı Formatı, <i>Variant Call Format</i>

## ŞEKİLLER

Şekil	Sayfa
3.1. Yüksek derinlikli örneklerin A) haritada konumları B) yaş ve hasar onarım bilgileri.	14
4.1. İki sürümde A) $r^2$ , B) heterozigot hata ve C) NRD ile ölçülen genotip impütasyonu doğruluk değerleri.	21
4.2. İki panelde A) $r^2$ , B) heterozigot hata ve C) NRD ile ölçülen genotip impütasyonu doğruluk değerleri.	22
4.3. GP filtresine göre A) genotip uyumluluk, B) heterozigot hata oranı ve C) NRD değerleri.	23
4.4. On örnek için A) KGP B) KGP+HGDP panelinde GP filtresine göre 22. kromozomda kayıp TNP sayıları ve oranları.	24
4.5. On örnek için impütasyon yapılmış verilerin Human Origins verisetindeki yaygın varyantlarda GP filtresine göre kayıp sayıları.	25
4.6. Yüksek derinlikli örnekler için A) GP80 filtreli impütasyon verisi ve B) GP99 filtreli impütasyon verisi ile çizilen TBA grafikleri.	27
4.7. Modern TGP örnekleri için TBA grafiği.	28
4.8. f-istatistikleri. A) f(DışGrup, Href, İmpütasyon, Diploid/Sözde-haploid) ve B) f(DışGrup, ModernPop, İmpütasyon, Diploid/Sözde-haploid) topolojisi ile yapılan hesaplamalar.	29
4.9. Düşük derinlikli yayınlanmış örneklerin TBA grafiği. Referans genom mavi elmas ile gösterilmiş ve aynı örnek için iki veri tipi arasına çizgi eklenmiştir.	31
4.10. Düşük derinlikli örnekler için A) derinliğe ve B) veri dizileme yöntemine göre Human Origins verisetinde kayıp oranlarının impütasyon ve sözde-haploid verilerde karşılaştırması.	32
4.11. Proje örneklerinde veriseti pozisyonlarında kayıp oranlarının sözdehaploid veri ile A) örneğe göre ve B) derinlik grubuna göre karşılaştırması.	33
4.12. Yaygın (A, B) ve az yaygın (C, D) varyantlar için GP80 ve GP99 filtreli impütasyon veya hizalama verisinden hesaplanan alel sıklık	

hesabı farkları.	34
<b>4.13.</b> Yaygın (A) ve az yaygın (B) varyantlar için GP80 veya GP99 filtreli veriyle ve hizalama verisiyle hesaplanan değerlerin fark dağılımı.	35

**TABLolar**

<b>Tablo</b>		<b>Sayfa</b>
<b>2.1.</b>	Bazı genotip impütasyonu araçları	6
<b>2.2.</b>	Türkiye'den elde edilen iyi kalitede veriler.	9
<b>3.1.</b>	Karşılaştırma için kullanılan yüksek derinlikli örnekler.	13
<b>3.2.</b>	Yayınlanmış düşük derinlikli örneklerin ülkeye göre dağılımı ve genom boyu derinlik aralıkları.	13
<b>3.3.</b>	Yeni dizileme verisi üretilen örneklerin genom derinlik aralıkları, konuma göre sayıları ve dönemleri.	13
<b>3.4.</b>	Alel sıklık hesabı için kullanılan örnekler.	19



## 1. GİRİŞ

Genetik arařtırmaların kapsamını binlerce yıl öncesinde yařamıř canlıları içerecek řekilde geniřleten antik DNA arařtırmaları ile son yıllarda giderek artan bir hızla onlarca ülkeden yüzlerce antik DNA dizileme verisi üretilmektedir. Antik DNA içeren kaynak malzemeler genellikle çevresel kořullar nedeniyle az korunmuřtur ve az miktarda DNA içermektedir. Elde edilen az miktarda veriyi verimli biçimde kullanabilmek için farklı yöntemler geliřtirilmektedir. Yaygın kullanılan bir yöntem sözde-haploidizasyon iřlemidir. Bu yöntem eksik aleller nedeniyle çağrılmayan diploid genotipler yerine mevcut olan tek alelin veya ikisinden birinin diploid olarak kullanılmasını temel alır, ancak sözde-haploidizasyon yönteminin referans yanlılıđına neden olduđu bilinmektedir.

Son yıllarda büyük çaplı genomik dizileme projeleri ile yüksek kalitede insan ekzom ve genom verisi üretilmesi arařtırmalara kaynak olacak büyük miktarda veri elde edilmesini sađlamıřtır. Bu veriler birçok popülasyondaki genetik çeřitliliđin belgelenmesine ve referans panel olarak kullanılmasına imkan vermektedir. Bir referans panel kullanarak dizileme verilerindeki eksik genotiplerin tahmin edilmesi genotip impütasyonu olarak adlandırılır. Genotip impütasyonu seyrek genotip verisi içeren TNP *array* veya düşük derinlikli tüm genom dizileme verileri üzerinde uygulanabilir.

2007 yılında ilk kez GWAS çalıřmalarında kullanılan genotip impütasyonu, son yıllarda üretilen veri miktarı hızla artsa da iyi kalite veri elde etme sorunu yařanan antik DNA arařtırmalarında verideki eksik genotipleri *doldurmak için* güçlü bir aday olmuřtur. Son yıllarda geliřtirilen hızlı ve dođru bir řekilde genotip ataması yapan araçlar ile antik DNA verilerine impütasyon uygulamak daha kolay hâle gelmiř ve özellikle referans panellerde iyi temsil edilen örnekler için iyi sonuçlar alınmaya başlanmıřtır.

Bu çalıřmada, genotip impütasyonunun Türkiye'den antik insan örneklerinde kullanılabilirliđi hakkında bir inceleme yapmak hedeflenmiřtir. İklim kořulları nedeniyle az miktarda korunmuř DNA elde edilebilen bölgelerden olan Türkiye'de elde edilen antik genom örneklerinden bu yöntemle daha fazla bilgi elde etmek

mümkün görünmektedir. Bu nedenle yöntemin performansını farklı parametreler ile değerlendirmek ve Türkiye ve çevresinden yayınlanmış ve bazı yayınlanmamış verilere genotip impütasyonu uygulamak amaçlanmıştır.

## 2. GENEL BİLGİLER

### 2.1 Antik DNA Çalışmaları

Canlı dokulardan izole edilerek elde edilen DNA'nın aksine antik DNA molekülleri genellikle kazılarda çıkarılan dokular veya malzemelerden elde edilir. Bu durumda sıklıkla toprağın asitliği, nem gibi çevresel koşulların etkisiyle DNA'nın korunabilen eser miktarda bir kısmı elde edilmektedir. Deoksiriboz ve baz arasındaki  $\beta$ -n-glikozidik bağın kopması sonucu bölgelerin tek parça hâlinde kopyalanmasını engelleyen kesikler, oksidasyon sonucu açılan halkalar veya pürin dimerizasyonu gibi engel bölgeleri (İng. blocking lesion) olduğu için antik DNA moleküllerinin PZR çıktıları kısa uzunluktadır. Bunu yanısıra DNA parçalarının uçlarında bazlardaki amin grubunun (NH<sub>2</sub>-) kaybedilmesi sonucunda sitozin timin bazına dönüşür, bu C-T deaminasyonu antik DNA dizileme ile üretilen verilerin tanımlayıcı özelliklerindedir (1).

Çalışmaların hız kazandığı, farklı ülkelerden ilk defa antik DNA verilerinin yayınlandığı dönemde arkeolojik buluntuların çerçevesinde antik örnekleri günümüz popülasyonları ve antik örneklerin popülasyonlarıyla karşılaştıran demografi temalı araştırma çalışmaları yaygındı. Neandertal ve Denisova genomlarının keşfi ve modern popülasyonlara olan genetik benzerlikleri, ilk tarım topluluklarının göç hareketleri ve ilişkileri, bozkır halklarının Avrasya'ya nasıl bir etki ettiği gibi konular ile ilgili araştırmalar yayınlandı. Yayınlanan veriler arttıkça yapılabilecek çalışmaların kapsamı da genişlemiş, son yıllarda araştırmalar eski çağlarda insanların hangi sağlık sorunlarıyla mücadele ettiği, bağırsak mikrobiyomlarının ve patojenlerin çeşitliliği, belirli fenotiplerin seçilime uğrayıp uğramadığı, toplumlarda ne çeşit akrabalık ilişkileri olduğu gibi sorular üzerine yoğunlaşmıştır (2,3).

Yayınlanan örnek sayısının artmasında daha fazla sayıda araştırmacının antik DNA araştırmalarına dahil olması ve araştırmalar için kaynak bulması ile beraber korunmuş DNA içeren daha çok sayıda örneğin elde edilmesi, malzemelerden DNA eldesindeki ıslak laboratuvar yöntemlerinin geliştirilmesi ile daha fazla ve daha

kullanılabilir DNA elde edilmesi, dizileme teknolojilerinin gelişmesi ile elde edilen DNA'nın daha verimli bir şekilde, daha hızlı ve daha ucuz dizilenmesi gibi etkenler ayrı ayrı önem taşır.

Temel Bileşenler Analizi, popülasyon genetiği çalışmalarında sıklıkla başvurulan bir boyut düşürme yöntemidir. Antik DNA verilerine sıklıkla TBA uygulanarak örneklerin genetik yapısı hakkında analizler için öncül bir sonuç elde edilir. Düşük derinliğe sahip örneklerin konumları yeterli sayıda TNP içermemeleri nedeniyle sapma gösterebilir. TBA'nın yanı sıra çok boyutlu ölçekleme (MDS, *multi-dimensional scaling*) kullanarak da çok boyutlu genetik veriyi basitleştirmek ve örnekleri gruplamak mümkündür. Çok az veri içeren (otuz bin okumaya kadar) örnekler için TBA yerine MDS ile sonuç alınabilir (4). Sık kullanılan bir başka yöntem olan f-istatistikleri alel paylaşım miktarlarına göre iki, üç veya dört grubun birbirlerine olan genetik benzerliğini veya yakınlığını ölçmek için kullanılır (5).

## 2.2 Antik DNA Verilerinde Karşılaşılan Bazı Sorunlar

İyi korunmuş antik DNA örnekleri genellikle soğuk havaya ve az nemli iklime sahip olan bölgelerde bulunur. Ekvatorial kuşağa doğru yaklaşıldıkça iklim koşulları nedeniyle toprak içinde kemiklerin korunmuşlukları ile beraber içerdikleri DNA da azalmaktadır. Kötü korunmuş malzemelerden elde edilen antik DNA verisi genellikle düşük genom boyu derinliğine sahiptir.

Elde edilen az miktarda veriyi verimli kullanabilmek için diploid ve eksiksiz veriye uygun olarak tasarlanmış yöntemleri uygulayabilmek amacıyla sözde-haploidizasyon (İng. pseudohaploidization) işlemine başvurulur. Bu işlemde temelde her pozisyon için rastgele bir nükleotit seçilerek sahte haploid bir veri elde edilir. Ancak özellikle birçok farklı hata kaynağı içerebilen antik DNA verileri için bu işlem analizlerde yanlışlıklara yol açabilir (6).

Dizileme sonrası DNA dizi parçalarının hizalaması genellikle lineer haploid bir referans genom kullanılarak yapılır. Bu sırada haploid veride kullanılan referans alel dışındaki alellerin (alternatif) varlığı hizalanan okumaların hizalama eşleşme skorlarını (İng. mapping score) düşürür ve bazı durumlarda bu okumaların

kullanılmamasına yol açar. Özellikle düşük derinlikli dizileme verileri için bu durum sıklıkla alternatif alellerin eşleşme hatası sebebiyle veriden atılmasına yol açan referans yanlılığına (İng. reference bias) neden olur. (7,8) Büyük oranda düşük derinlikli olmasının yanında bir de kısa okuma uzunluklarına sahip olan antik DNA okumaları bu sorundan daha fazla etkilenmektedir (8). Çoğunlukla sözde-haploid hâle getirilerek kullanılan antik DNA verisi için bu yanlılık sebebiyle alternatif alelleri içeren genotipler eksik miktarda tespit edilebilmektedir. Yakın zamanda referans genoma hizalama yerine pangenom hizalama yapma ve impütasyon uygulaması ile referans yanlılığı sorunundan kurtulmanın mümkün olduğu görülmüştür (9).

Popülasyon genetiği araştırmalarının sonuçlarını etkileyen bir diğer yanlılık, önceden seçilmiş bir TNP seti ile yapılan analizlerde görülebilen TNP tespit yanlılığıdır (İng. SNP ascertainment bias). TNP'lerin seçimi sırasında bir adet popülasyonun kullanılması ile bu popülasyonun genetik çeşitliliğininin daha fazla temsil edilmesi sonucu ortaya çıkar. Bu yanlılığa ayrıca daha az görülen alellerin belirlenmesini zorlaştıracığı için küçük örneklem büyüklüğü de neden olabilir (10). İmpütasyon kullanmanın TNP tespit yanlılığını azaltabileceği gösterilmiştir (11). TNP tespit yanlılığına sahip olmayan bir veri seti kullanmak araştırma sonuçlarının güvenilirliğini artıracaktır.

### 2.3 Genotip İmpütasyonu

İmpütasyon metodu, istatistikte bir verideki eksik bilgilerin tahmin edilmesini ve atanmasını ifade eder. Genotip impütasyonu, genom boyu ilişkilendirme, fenotip tahmini, popülasyon genetiği gibi farklı genetik araştırma alanlarında sıklıkla kullanılan genotip verilerindeki eksikleri tamamlama işlemidir (12). İlk kez bir Tip 2 Diyabet GWAS araştırmasında HapMap verisi referans olarak kullanılarak iki milyondan fazla otozomal pozisyonda genotip impütasyonu uygulanmıştır (13). Düşük derinlikli tüm genom dizilemenin (lc-WGS) yaygınlaşmasıyla bu tip veriler için de genotip impütasyonu kullanılmaya başlanmıştır. Genotip impütasyonu işleminde eksik genotipler içeren haplotipler, eksik veri içermeyen çok sayıda haplotip içeren bir referans panel kullanılarak doldurulur. Çok yüksek hesaplama işlem yükü içeren impütasyon işleminden önce haplotiplere ön-fazlama (İng.

prephasing) işlemi yapmak impütasyon ve fazlama süreçlerini daha hızlı ve kolay bir hâle getirmektedir (14). Genotip impütasyonu için kullanılan bazı araçlar Tablo 2.1’de gösterilmiştir. Araçların hepsi, kromozomlardaki uzun bölgeleri daha kısa haplotip bölgelerin mozaikleri olarak ele alan 2003 tarihli Li ve Stephens modelini temel almaktadır (14,15).

**Tablo 2.1.** Bazı genotip impütasyonu araçları

Yazılım	Yayın Tarihi	Kaynak
IMPUTE2	2009	(16)
minimac3	2016	(17)
BEAGLE5	2018	(18)
QUILT	2021	(19)
GLIMPSE2	2023	(20)

### 2.3.1 Referans paneller

Genotip impütasyonu işlemi için kullanılacak referans panel oldukça önem taşır. Referans panel impütasyon yapılacak hedef (İng. target) örnekler için kaynak haplotip havuzunu oluşturur ve kaynak haplotipler örneğin haplotiplerine ne kadar benzer ise, bir başka deyişle örnek panelde ne kadar iyi temsil edilirse impütasyonun başarısı o kadar artacaktır. Bu nedenle kullanılan referans panele göre elde edilecek sonuç değişebilir. HRC (21), TOPmed (22), KGP, (23) HGDP (24) gibi büyük dizileme verisi kaynakları çok sayıda varyant için çok sayıda haplotip içermeleri nedeniyle referans panel olarak kullanılabilir. Farklı popülasyonlar için genotip impütasyonunun performansı değerlendirilmekte ve geniş kapsamlı projelerde yeterince temsil edilmeyen popülasyonlar için mevcut referans paneller ayrı ayrı veya birlikte kullanılarak en iyi sonuç aranmaktadır (25–28). Çalışılan popülasyona ait biyobanka verilerinin referans panel olarak kullanılmasının daha az sayıda örneğe rağmen impütasyon sonucunu belirgin biçimde iyileştireceği öngörülmüş ve gözlenmiştir (29,30). Referans panellerin geniş olması impütasyon üzerinde genel olarak olumlu etki yaparken paneldeki popülasyon çeşitliliğinin artmasının her zaman doğrudan daha iyi bir sonuca yol açmayacağına dair bir bulgu da mevcuttur (31).

### 2.3.2 Antik DNA Verisi ile Genotip İmpütasyonu Çalışmaları

Antik DNA verileri için araçların ve panellerin gelişimiyle genotip impütasyonu son yıllarda daha kullanılabilir hâle gelmiştir. Avrupa'dan bir neolitik dönem örneği kullanarak çok düşük derinlikli antik genomlar için bir impütasyon iş-akışı sunan bir çalışma (32), BEAGLE aracı ile, referans panel seçiminin yine bir adet Avrupa'dan antik örnekte etkisini gösteren bir çalışma (33), GLIMPSE ile yapılmış, dünyanın farklı bölgelerinden 42 iyi kalite antik genom kullanarak değerlendirmeler yapan kapsamlı bir çalışma (34), ve yine GLIMPSE ile antik örnekler için ölüm sonrası hasarın (İng. PMD, post mortem damage) impütasyon sonuçlarına etkisini inceleyen bir çalışma (35) ile antik DNA verilerinde impütasyon uygulaması hakkında bilgi elde edilmiş ve uygulamanın sınırları araştırılmıştır. Teknik araştırmalarla birlikte analizlerde impütasyon uygulanmış antik DNA verisi kullanılan araştırmalar da artmaktadır (36–41).

### 2.3.3 Saklı Markov Modelleri ve GLIMPSE

Saklı Markov Modelleri (HMM, Hidden Markov Models), gözlenebilir durumların gerçekleşme olasılıklarının doğrudan gözlenemeyen faktörlere bağlı olduğu istatistiksel modellerdir. Bu modeller biri Markov zinciri içeren, görünmeyen bir saklı süreç diğeri ise gözlemlenebilir simgeleri içeren süreç olmak üzere iki stokastik süreçten oluşur (42). Saklı Markov Modelleri protein veya DNA dizilerindeki örüntüleri saptamak ve tahmin etmekte oldukça başarılı bulunmuştur ve birçok hesaplamalı biyoloji analiz aracında kullanılmaktadır (42). DNA dizilerinde genotip impütasyonu yapan birçok araç Saklı Markov Modelleri kullanmaktadır (14).

GLIMPSE, (İng. Genotype Likelihoods Imputation and Phasing Method) ilk olarak 2021 yılında yayınlanmış, düşük derinlikli tüm genom dizileme (İng. lc-WGS) verisinde hızlı bir şekilde genotip impütasyonu yapabilen bir yazılımdır (43). Aynı zamanda GLIMPSE yazılımının ikinci sürümünün önceki yazılımlara göre de daha yüksek doğrulukla sonuç verdiği görülmüştür (20).

GLIMPSE impütasyon modeli Li ve Stephens Saklı Markov Modeli'ni esas alır. PBWT veri yapısını kullanarak lineer zamanda haploid impütasyonunu temel alarak

çalışır. Bu özellikleri nedeniyle kuadratik zamanda diploid impütasyon yapan araçlardan daha hızlı bir şekilde atamaları yapabilmektedir. Programın girdileri bir referans panel ve bir genotip olabilirliği matrisidir. Her bir haplotip için iki sözde-haploid olabilirlik vektörüne Li ve Stephens HMM modeli temelli impütasyon yaptıktan sonra fazlama (İng. phasing) yapar ve diplotipi oluşturur (43). Girdi olarak GL bilgisi içeren bir VCF veya BCF dosyası vermek veya ikinci sürümü itibariyle doğrudan BAM dosyası vermek mümkündür.

Yazılım, VCF veya BCF dosyasında, GT:DS:GP formatında her bir pozisyon için çıktı olarak iki nokta ile ayrılmış üç değer verir. İlki tahmin edilmiş fazlanmış (İng. phased) genotiptir. Biallelik bir TNP pozisyonu için, 0 referans, 1 alternatif aleli ifade eder; böylece genotip 0|0, 0|1, 1|0 veya 1|1 değerlerine sahip olabilir. İkinci değer DS, alel doz (İng. allele dosage) değeridir, 0 ile 2 arasında, genotipteki alternatif alel miktarını ifade eden bir değerdir. DS 0 değeri pozisyon için hiç alternatif alel bulunmadığını, 2 ise 2 alternatif aleli, yani alternatif homozigot genotipi ifade eder. Sonucu değer GP ise genotiplerin her biri için ardıl genotip olasılık (İng. genotype posterior probability) değeridir ve üç genotip için sırayla yüzdelik olasılık değerlerini virgülle ayrılmış olarak içerir. Örnek olarak, bir TNP için bir örneğe ait çıktı değeri '0|0:0.3:0.85,0.15,0' dır. Bu durumda bu varyant için %85 olasılıkla ve 0.3 alternatif genotip doz değeri ile 0|0 yani referans homozigot genotipi çıkarımı yapılmıştır.

### 2.3.4 Genotip İmpütasyonu Doğruluğu Ölçüleri

İmpütasyon işleminin doğruluğunu ölçmek için sıklıkla kullanılan yöntem impütasyon verisinin asıl veri (validasyon) ile karşılaştırılmasını esas alır. Bunun için iyi kalitede, yüksek genom okuma derinliğine sahip bir örneğin genom dizileme verisi düşük okuma derinliğine düşürüldükten sonra impütasyon yapılır ve sonrasında iki verideki alel dozlarını kullanarak genotiplerin birbirleri ile ne kadar 'uyumlu' oldukları ölçülür. Bunun için sıklıkla lineer regresyon uygulanarak elde edilen Pearson korelasyon katsayısı  $r^2$ , genotip uyumluluk (İng. genotype concordance) değeri olarak kullanılır. Bu ölçüm karşılaştırılacak bir iyi kalitede bir veri olmadığı durumlarda, yani düşük okuma derinliğine sahip dizileme verilerinde kullanılamaz.



İmpütasyon işleminde herhangi bir pozisyonda referans alellerin atanma olasılıkları daha yüksek olduğu için referans homozigot genotipin doğru tahminlerini hesaba katmayarak heterozigot ve homozigot alternatif genotiplerdeki hata miktarını öne çıkaran referans-dışı uyumsuzluk (İng. NRD, non-reference discordance) değeri ve yalnızca heterozigot genotipleri değerlendiren heterozigot hata oranı (İng. heterozygous mismatch) da atama işleminin doğruluğunun bir ölçüsü olarak kullanılmaktadır.

İmpütasyon çıktısında elde edilen genotiplerin güvenilirliğini yorumlamada GP gibi olasılık çıktıları bir araç olarak kullanılır. Analizlerde kullanılmak üzere genellikle %80 veya %90 gibi bir GP alt sınırı belirleyerek, sınırın altında kalan genotip atamaları sonraki analizlere dahil edilmeyebilir. Bunun dışında impütasyon araçlarında her pozisyon için belirtilen INFO skoru da varyantlara göre impütasyon işleminin güvenilirliğine dair bir olasılık olarak değerlendirilebilir.

#### 2.4 Anadolu'dan Elde Edilen Antik İnsan Genetik Verisi

Türkiye arkeolojik buluntular açısından oldukça zengin olmasına rağmen iklim koşulları sebebiyle Türkiye'de elde edilen DNA içeren malzemeler çoğunlukla iyi korunmuş değildir ve az miktarda DNA içermektedir. Şimdiye dek elde edilen antik insan genom verileri için tüm genom derinliği birkaç örnek hariç 4X'in altındadır. En yüksek ortalama genom boyu derinliğine sahip örnekler Tablo 2.2'de gösterilmiştir.

**Tablo 2.2.** Türkiye'den elde edilen en iyi kalitede veriler.

Örnek Adı	Genom Boyu Derinlik	Kaynak	Yer
BAR25	8.86	(44)	Barcın Höyük, Bursa
AKT16	8.58	(44)	Aktopraklık Höyük, Bursa
Gor001	7.19	(45)	Gordion, Ankara
Bon002	6.32	(46)	Boncuklu Höyük, Konya
Bar8	6.02	(47)	Barcın Höyük, Bursa
Ash128	4.79	(48)	Aşıklı Höyük, Aksaray

## 2.5 Bu Çalışmanın Kapsamı

Bu çalışmada, Anadolu'dan ve Batı Avrasya'dan antik genomlarda impütasyon kullanımının sınırları hakkında bir değerlendirme yapmak ve belirlenen koşullarda Anadolu ve çevresinden uygun örneklerle uygulamak hedeflenmiştir. Genotip uyumluluğu, heterozigot hata oranı ve referans dışı uyumsuzluk (NRD) değerlerini kullanarak atamaların doğruluğu hakkında yorum yapmak, üretilen veriyi yaygın antik DNA analizlerinde kullanarak değerlendirme yapmak amaçlanmıştır. Bunun için ikisi Türkiye'den toplam on yayınlanmış yüksek derinlikli genom impütasyon doğruluğunu değerlendirmek üzere, büyük kısmı Türkiye'den yüz on yayınlanmış genom ve araştırma projesi kapsamında yeni üretilmiş otuz dört genom verisi de impütasyon uygulamak ve analizlerde kullanmak üzere derlenmiştir.

### 2.5.1 Çalışmada Kullanılan Veri Kaynakları

1000 Genom Projesi, KGP, 2010 yılında başlatılmış uluslararası bir çabanın ürünü, dünyadaki farklı kıtalardan farklı popülasyonların genetik çeşitliliğini belgelemek amacıyla DNA dizilemesi yapmış ve açık erişime sunmuş bir projedir. Başlangıçta verilerin derinleme derinliği daha düşükken Faz 3 itibariyle, mevcut örneklerin bazı akrabalarının da verisini içeren (602 üçlü akraba örneği), 26 popülasyondan 3202 örnek için 30X derinliğinde tüm veriler tekrar dizilenmiştir ve en güncel insan referans genom sürümü GRCh38'e hizalanmıştır (23). Referans panel olarak kullanım için mevcut en kapsamlı veriyi sunan projede yaklaşık 60 milyon biallelik TNP pozisyonu mevcuttur.

HGDP (İng. Human Genome Diversity Project), KGP'ye kıyasla dünyadaki genetik çeşitliliği daha iyi yansıtmak amacına sahip bir başka büyük çaplı genom dizileme projesidir (24). HGDP+KGP ise, KGP ve HGDP projelerinde üretilen verilerin birlikte işlenmesi sonucu elde edilen daha büyük çapta bir veridir (49). Bu veride iki veri setinin de örneklerinin mevcut olmasının yanında yeniden birlikte işlenmeleri sebebiyle çok daha fazla sayıda nadir varyant keşfedilmiştir.

Human Origins, popülasyon genetiği araştırmalarına uygun olarak geliştirilmiş bir TNP *array* setidir. İlk olarak 2012 yılında yayınlanan verisetinin ikinci versiyonu

2014 yılında paylaşılmıştır (5,50). Human Origins veriseti 13 farklı popülasyondan bireylerin heterozigot olduğu pozisyonları temel alarak seçilmiş yaklaşık 600 bin TNP içeren 14 farklı panel içermektedir. Sıklıkla insan antik DNA'sı çalışmalarında temel bileşenler analizi yürütmek için tercih edilmektedir. Bu verisetindeki TNP'lerin seçilmesinde kullanılan adımlar detaylı bir biçimde belgelenmiştir.

gnomAD (İng. Genome Aggregate Database) 2014 yılında hayata geçmiş, büyük çaplı ekzom ve genom dizileme projelerine ait verileri bilim dünyasına açan bir projedir (<https://gnomad.broadinstitute.org/about>). Farklı dizileme projelerinden elde edilen farklı popülasyonlara ait alel frekansı ve genotip verilerini içerir ve insan genetik çeşitliliği için bir veritabanı görevi görür. Milyonlarca pozisyon için veri içermektedir.

### 3. Gereç ve Yöntem

Bu tez kapsamında antik DNA verileri için genotip impütasyonunun doğruluğunu ve impütasyon çıktısının sıklıkla kullanılan analizlerdeki performansını değerlendirmek amacıyla antik ve modern dizileme verileri derlenmiştir. GLIMPSE2 aracı ile, KGP 3.Faz 30X verisi referans panel olarak kullanılarak otozomal biallelik TNP pozisyonlarında örnekler genotip impütasyonu yapılmıştır. Referans panel karşılaştırması amacıyla HGDP+KGP verisi 22. kromozomda referans panel olarak kullanılmıştır. GP ve MAF değerlerine göre genotip uyumluluk değerleri incelenmiştir. İmpütasyon çıktısı verinin antik DNA çalışmalarında en çok kullanılan metotlardan TBA ve f-istatistiklerinde nasıl davrandığı araştırılmıştır. İmpütasyon çıktısı ile hesaplanan alel sıklık değerleri BAM dosyasından hesaplanan değerler ile karşılaştırılmıştır. Kullanılan kodlar ve komutlar ile ilgili verilere <https://github.com/hande-j> adresinden ulaşılabilir.

#### 3.1 Veriler

Çalışma kapsamında yüz on sekiz adet yayınlanmış, otuz dört yeni üretilmiş antik insan genom dizileme verisi derlenmiştir:

Genom derinlikleri 0.1x ile 6.3x arasında olan 108 antik genomun dizileme verisi elde edilmiştir. Veriler üretim biçimine göre 74 tanesi yakalama (1240K), 30 tanesi rastgele dizileme ve 6 tanesi yakalama sonrası rastgele dizileme verisi olmak üzere üçe ayrılmaktadır. Bunun dışında genom derinliği 8.58 ve 53.46x arasında olan 10 yüksek derinlikli genoma ait dizileme verisi impütasyon sonuçlarını değerlendirmek amacıyla kullanılmıştır. Bu on örneğin hepsi için tüm genom rastgele dizileme ile veri üretilmiştir. Tablo 3.1’de yüksek kaliteli örneklerin bilgileri, Tablo 3.2’de düşük kaliteli yayınlanmış örneklerin ülkeye göre dağılımı gösterilmiştir. Tablo 3.2’deki örneklerin detaylı bilgileri EK-7’de görülebilir.

TÜBİTAK ARDEB 3501 programı ile desteklenen 121Z05 numaralı araştırma projesi kapsamında tüm genom rastgele dizileme ve seçili TNP’ler için yakalama (İng. capture) dizilemesi yapılarak Anadolu’dan neolitik ve geç kalkerolitik dönemlerine ait üç farklı konumdan 34 örnek için insan genom dizileme verisi

üretimiştir. Tablo 3.3'te örneklerin yere göre bilgisi gösterilmiştir. Örneklerin detaylarını içeren tablo EK-8'de görülebilir.

**Tablo 3.1.** Karşılaştırma için kullanılan yüksek derinlikli örnekler.

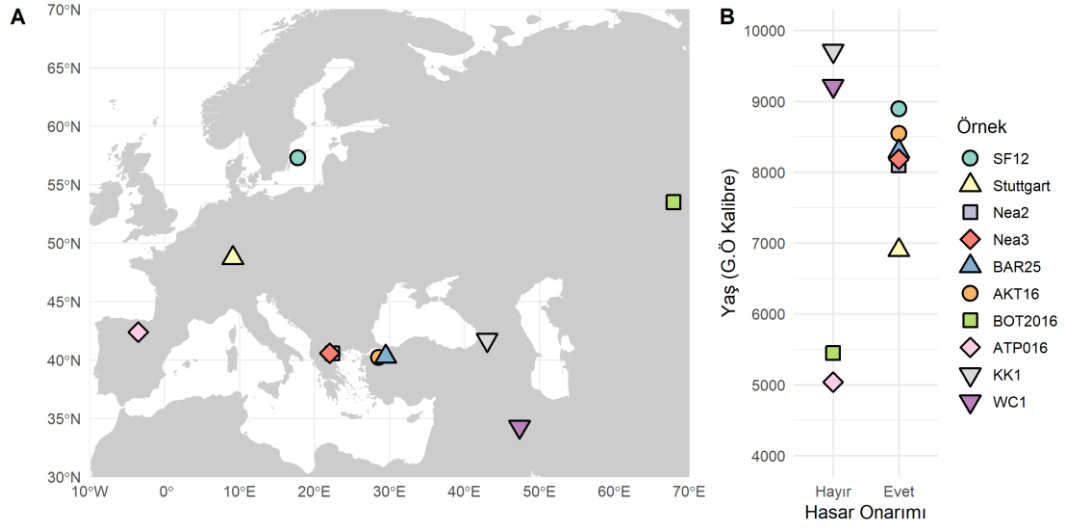
Örnek	Kalite Filtreli Genom Derinliği	Konum	Kaynak	Hasar Onarımı
SF12	53.46	Stora Karlsö, İsveç	(51)	Var, UDG
Stuttgart	13.65	Stuttgart, Almanya	(50)	Var, UDG
Nea2	10.82	Nea Nikomedeia, Yunanistan	(44)	Var, USER
Nea3	9.76	Nea Nikomedeia, Yunanistan	(44)	Var, USER
AKT16	8.58	Aktopraklık, Türkiye	(44)	Var, USER
BAR25	8.86	Barcın, Türkiye	(44)	Var, USER
ATP016	12.36	El Portalon, Atapuerca, İspanya	(52)	Yok
BOT2016	12.36	Botay, Kazakistan	(53)	Yok
KK1	9.41	Kotias Klde, Gürcistan	(37)	Yok
WC1	8.73	Wezmeh Mağarası, İran	(54)	Yok

**Tablo 3.2.** Yayınlanmış düşük derinlikli örneklerin ülkeye göre dağılımı ve genomboyu derinlik aralıkları.

Ülke	En Yüksek Derinlik	En Düşük Derinlik	Örnek Sayısı
Türkiye	6.31	0.11	72
İran	1.69	0.15	24
Yunanistan	4.36	0.14	12

**Tablo 3.3.** Yeni dizileme verisi üretilen örneklerin genom derinlik aralıkları, konuma göre sayıları ve dönemleri.

Örnek Sayısı	Derinlik Aralığı (x)	Yer	Dönem
14	0.02 - 0.12	İkiztepe	Geç Kalkolitik
15	0.10 - 5.28	Çatalhöyük	Neolitik
5	0.03 - 1.20	Çayönü	Neolitik



**Şekil 3.1.** Yüksek derinlikli örneklerin A) haritada konumları B) yaş ve hasar

onarım bilgileri.

### 3.2 Ön-işleme

Bu tez çalışması kapsamında derlenen yayınlanmamış ve yayınlanmış tüm dizileme verilerinin ön işleme Kıvılcım Başak Vural tarafından Orta Doğu Teknik Üniversitesi Biyolojik Bilimler bölümünde CompEvo araştırma grubuna ait sunucuda yapılmıştır. Bu işlemler aşağıdaki gibidir.

Her bir örnek için dizileme çıktısı (fasta) bwa (55) aracı alın komutu tek uç modunda antik DNA verisi için uygun olan '-n 0.01 -o 2 -l 16500' komutu kullanılarak hg38 insan referans genomuna hizalanmıştır. PCR kopya okumaları FilterUniqSAMCons.py (56) kodu kullanılarak atılmış, sonrasında hizalama (*mapping*) kalitesi 30'dan düşük, uzunluğu 35'ten az, referans genoma eşleşme (*mismatch*) oranı %10'dan yüksek olarak hizalanan okumalar atılmıştır. Son olarak her okumanın iki ucundan 10 baz bamUtil paketinin trimBam komutu ile çıkarılmıştır (57).

### 3.3 Veri Hazırlama

Genom boyu derinlikleri Samtools (58) view ve BEDtools genomeCoverageBed (59) komutları kullanılarak hizalama kalitesi 30'dan yüksek okumalar kullanılarak hesaplandı. Bunun için kullanılan komut aşağıdaki gibidir:

```
'samtools view -b -q 30 -F 4 ${bamfile} | ${genomeCov} -ibam - -g ${refgenome} >
coverages/${name}.cov.txt
```

```
coverage=$(grep genome coverages/${name}.cov.txt | awk '{NUM+=$2*$3;
DEN+=$3} END {print NUM/DEN}')
```

Okumaları seyreltilerek genom boyu derinlikleri azaltılacak on yüksek derinlikli genom için, hg38 referans genomunu (\$ref) ve 'ratio=\$(awk "BEGIN {print \$new/\$depth}")' komutu ile hesaplanan seyreltme oranını kullanarak 'samtools view -T \${ref} -s \${ratio}' komutu ile yeni genom okuma derinliği ortalaması (\$new) 0.1x, 0.25x, 0.5x ve 1x olacak şekilde yeni BAM ve indeks dosyaları üretildi.

Referans panel olarak 1000 Genom projesi (KGP) üçüncü faz (30X) verisi kullanıldı (23). Otozomal kromozomlar için veri BCF formatında [http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data\\_collections/1000G\\_2504\\_high\\_coverage/working/20201028\\_3202\\_phased/](http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data_collections/1000G_2504_high_coverage/working/20201028_3202_phased/) adresinden indirildi. bcftools (60) kullanılarak önce multi allelik pozisyonlar biallelik pozisyonlara dönüştürüldü ve sonrasında biallelik TNP pozisyonları seçildi.

Ek olarak, karşılaştırma amacıyla 1000 Genom verisi ve HGDP verisi ile birlikte işlenerek hazırlanan HGDP+KGP verisinin (49) 22. kromozom fazlanmış haplotip verisi [https://console.cloud.google.com/storage/browser/gcp-public-data--gnomad/resources/hgdp\\_1kg/phased\\_haplotypes](https://console.cloud.google.com/storage/browser/gcp-public-data--gnomad/resources/hgdp_1kg/phased_haplotypes) adresinden indirildi. Bu veri toplamda 4099 bireyin dizileme verisini içermektedir. Veri, indirildikten sonra KGP referans paneline benzer şekilde biallelik TNP pozisyonları seçildi. 22. kromozomda KGP panelinde 900 bin birleşik panelde 1.800 bin TNP bulunmaktadır.

### 3.4 Genotip impütasyonu

Genotip impütasyonu GLIMPSE (20) aracının ikinci sürümü kullanılarak yapıldı. İlk olarak dört farklı genom derinliğine seyreltilen on örneğe otozomlarda genotip impütasyonu yapıldı. Sırasıyla GLIMPSE\_split\_reference ile referans paneli uygun formata çevirme, GLIMPSE\_chunk ile kromozomları kısa parçalara bölme, GLIMPSE\_phase ile impütasyon ve fazlama, GLIMPSE\_ligate ile kısa parçaları kromozomlara birleştirme adımlarını içeren impütasyon iş akışı yürütüldü ve GLIMPSE\_concordance aracı ile doğruluk oranları incelendi. İş akışı sırasında varsayılan parametreler kullanılmıştır. On örnekte genotip uyumluluğu değerlendirmesinde panel karşılaştırması amacıyla ikinci sürüm kullanılarak KGP'nin yanında 22. kromozom için HGDP+KGP verisi, sürüm karşılaştırması amacıyla KGP paneli ile GLIMPSE aracının ilk sürümü (43) kullanıldı. Düşük derinlikli örneklere KGP paneli kullanılarak ikinci sürüm ile impütasyon yapıldı.

*Concordance* modülü ile doğruluk karşılaştırmaları yapılırken alel frekans verisi gnomAD (3.0) verisinin referans panelle kesişim noktaları alınarak hazırlandı. Alel frekans için mevcut farklı gruplamalar içinden AF\_NFE (*Allele Frequency\_NonFinnishEuropean*) değeri kullanıldı. İmpütasyon çıktısı, diploid genotiplenmiş ve filtrelenmiş yüksek kalite genom verisi ile karşılaştırıldı. Çıktı dosyalarında elde edilen  $r^2$  Pearson korelasyon katsayısı genotip uyumluluğu için anahtar sonuç olarak değerlendirildi. Bunun dışında NRD, ve heterozigot hata oranı değerleri kullanıldı.

### 3.5 Genotipleme

Yüksek derinlikli örneklerin (n=10) diploid genotipleri en güvenilir genotipleri elde etmek amacıyla Sousa de Mota ve arkadaşlarının 2023 yılı çalışmasında uygulanan yöntemle benzer olarak beş adımda filtrelendi. Önce 'bcftools mpileup -q 30 -Q 20 -C 50' komutu çalıştırılarak otozomal kromozomlarda KGP panelinde bulunan biallelik TNP varyantları için genotipler elde edildi. Sonrasında VCF dosyasında QUAL değeri 30'dan büyük olan, okuma derinliği 8'den büyük ve ortalama genomboyu derinliğin iki katından küçük olan, UCSC RepeatMask aracına göre tekrarlı bölgelerde bulunmayan, 1000 genom *strict mask*



[https://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data\\_collections/1000\\_genomes\\_project/working/20160622\\_genome\\_mask\\_GRCh38/StrictMask/20160622.allChr.mask.bed](https://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data_collections/1000_genomes_project/working/20160622_genome_mask_GRCh38/StrictMask/20160622.allChr.mask.bed)

içinde bulunan pozisyonlar çekildi.

108 düşük derinlik, 10 yüksek derinlikli genom için sözde-haploid genotipleme Human Origins pozisyonları için EIGENSTRAT veriseti ve .bed uzantılı pozisyon dosyası ile hg38 insan referans genomu kullanılarak önce ‘samtools mpileup -R -B -q30 -Q30’ komutu ve sonra sequenceTools (<https://github.com/stschiff/sequenceTools>) yazılımının ‘pileupCaller -randomHaploid’ komutu kullanılarak yapıldı.

### 3.6 Karşılaştırma Veriseti

Popülasyon genetiği analizleri için Human Origins TNP Array verisetinin ikinci sürümünü kullanarak karşılaştırmalı bir veriseti hazırlandı. Bunun için toplamda 592,146 TNP’nin hg19 (GRCh37) genomik konumları UCSC LiftOver (<https://genome.ucsc.edu/goldenPath/help/hgTracksHelp.html#Liftover>) aracı ile ‘hg19ToHg38.over.chain.gz’ dosyası kullanılarak GRCh38/hg38 koordinatlarına çevrildi ve 213 popülasyondan 2583 birey için 592,006 TNP konumu elde edildi. PLINK ile çevrilen pozisyonlardan PLINK ve EIGENSTRAT formatı verisetleri ve .bed uzantılı pozisyon dosyaları oluşturuldu.

PLINK (61) ve EIGENSOFT yazılım paketinden (<https://github.com/DReichLab/EIG>) (62) convertf ve mergeit araçları kullanılarak antik örneklerin i) genotip impütasyonu çıktısı ve ii) sequenceTools pileupCaller ile çağrılmış sözde-haploid genotipleri verisetine eklendi. Referans yanlılığını değerlendirmek için KGP referans panelinin homozigot referans genotipleri, hg38 sürümü insan referans genomunu temsilen verisetine PLINK ile eklendi.

### 3.7 Temel Bileşenler Analizi

Karşılaştırma verisetini kullanarak EIGENSOFT yazılım paketinden smartpca aracı ile temel bileşenler analizi yürütüldü. Human Origins verisetindeki 213 modern popülasyon için on temel bileşen hesaplaması ve antik örneklerin yanısıra insan referans genomunu yansıtmaya işlemi ‘outliermode:2, numoutevec:10,

lsqproject:YES' parametreleri kullanılarak yapıldı. Daha sonra .avec uzantılı çıktıdan birinci ve ikinci bileşen değerleri RStudio (<https://posit.co/products/open-source/rstudio/>) üzerinden ggplot2 ve dplyr paketleri ile görselleştirildi.

### 3.8 F-İstatistikleri

AdmixTools yazılım paketinin qpDstat aracı (<https://github.com/DReichLab/AdmixTools/>) kullanılarak 'f4mode:YES' ayarı ile iki farklı topoloji kullanılarak Human Origins veri seti pozisyonları kullanılarak f4 istatistikleri hesaplandı. Modern Mbuti Pigme popülasyonu istatistikleri hesaplamada dış grup (İng. outgroup) olarak kullanıldı. İlk topoloji f(Mbuti, Modern Popülasyon, İmpütasyon, Genotipleme), ikincisi f(Mbuti, Referans Genom, İmpütasyon, Genotipleme) şeklindedir. İlk topoloji için Human Origins verisetinde bulunan popülasyonlardan KGP referans panelinde bulunan popülasyonları temsil etmesi amacıyla Bengal (Bengali), Pencap (Punjabi), Japon (Japanese), Han (Han), Fin (Finnish), İngiliz (English), İspanyol (Spanish), İtalyan (Italian\_North) popülasyonları kullanıldı. Girdi olarak kullanılan kombinasyon dosyalarının hazırlanması, çıktıların düzenlenmesi ve görselleştirilmesi RStudio ile ggplot2 ve dplyr, paketleri ile yapıldı.

### 3.9 Alel Sıklığı Hesabı Karşılaştırması

Alel frekansları hesaplamak için kullanılmak üzere referans panel olarak kullanılan KGP verisinden rastgele yaygın TNP'ler seçilmiştir. Bu işlem için önce KGP verisi tek bir VCF dosyasında birleştirilmiştir. Sonrasında önce bu dosya minör alel değeri 0.05'ten büyük olan TNP'leri seçmek için 'bcftools view -q:0.05:minor' komutu ile filtrelenmiştir. Aynı şekilde alel frekansı panelde 0.01 ile 0.05 arasında olan TNP'ler çekilmiştir. Sonrasında bu TNP'lerden Human Origins verisetinde bulunanlar 'bcftools view -r' komutu ile çekilmiştir. Sonunda, sıklığı yüzde bir ile beş arasında 42803, beşten büyük 390767 adet TNP listelenmiştir. R base paketinden rastgele seçim yapmak için sample() fonksiyonu kullanılmıştır.

EK-5 ve EK-6'da yer alan örneklerden Türkiye'den, neolitik döneme tarihlenen ve veri üretim tipi fark etmeksizin genom boyu ortalama derinliği 0.1x'in

üzerinde olan 39 örnek, ve yeni 2 örnek (mus005, cay007) (45,63) belirlenerek toplam 41 örnek karşılaştırma için kullanılmak üzere GP80 ve GP99 filtrelili impütasyon verileri tek birer VCF dosyasında birleştirilmiştir. Alel sıklık hesabı için kullanılan örnekler Tablo 3.4’de gösterilmiştir. SG+CP rastgele ve yakalama ile dizilemeyi ifade etmektedir.

Rastgele seçili 100 yaygın, 100 az yaygın TNP için alel sıklık değerleri 41 Anadolu neolitik döneme ait düşük derinlikli örnek için, ANGSD (64) aracı ile BAM dosyalarından ve PLINK ile impütasyon çıktısı VCF dosyalarından hesaplanmıştır. ANGSD kullanılırken ‘-doMAF 1 -doMAF 2’ komutları, PLINK ile ‘—freq’ komutu ile frekans değerleri hesaplanmıştır. İmpütasyon verileri için GP80 ve GP99 filtrelili olarak iki farklı veri kullanılmıştır. Hesap sonuçları EK-9 ve EK-10’da tablo olarak sunulmuştur.

**Tablo 3.4.** Alel sıklık hesabı için kullanılan örnekler.

Örnek	Derinlik	Derinlik Grubu	Veri Tipi
Ash128	4.79937	>1X	SG
Ash129	0.744252	0.5X-1X	SG
Bar31	3.22912	>1X	SG
Bar8	6.01755	>1X	SG
Bon001	0.1555	0.1X-0.25X	SG+CP
Bon002	6.3169	>1X	SG+CP
Bon004	0.227853	0.1X-0.25X	SG+CP
cay007	0.464762	0.25X-0.5X	SG
cay013	0.138491	0.1X-1X	SG+CP
cay034	1.20372	>1X	SG+CP
cch146	0.108099	0.1X-1X	SG+CP
cch151	0.172258	0.1X-1X	SG+CP
cch152	0.126706	0.1X-1X	SG+CP
cch163	0.1706	0.1X-1X	SG+CP
cch170	0.348318	0.1X-1X	SG+CP
cch226	0.112031	0.1X-1X	SG+CP
cch242	0.156732	0.1X-1X	SG+CP
cch251	1.07179	>1X	SG+CP
cch255	0.115971	0.1X-1X	SG+CP
cch290	0.116671	0.1X-1X	SG+CP
cch294	5.28131	>1X	SG+CP
CCH294	0.257745	0.25X-0.5X	SG
cch312	0.379924	0.1X-1X	SG+CP
cch318	0.134756	0.1X-1X	SG+CP
cch328	0.508437	0.1X-1X	SG+CP
cch510	0.481299	0.1X-1X	SG+CP

**Tablo 3.4. (Devam)** Alel sıklık hesabı için kullanılan örnekler.

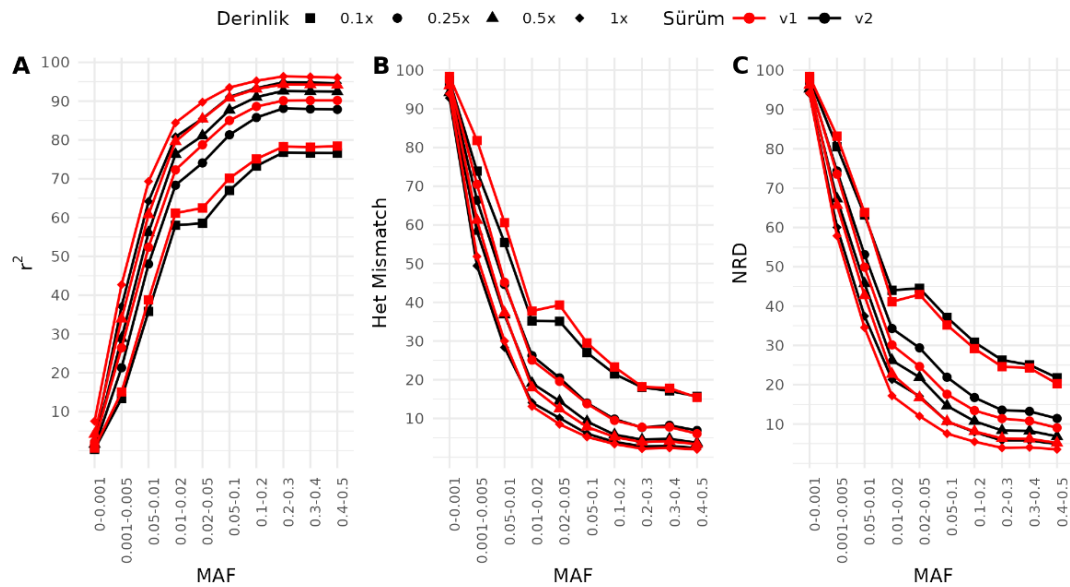
cth739	0.189873	0.1X-0.25X	SG
cth747	0.115017	0.1X-0.25X	SG
I0707	0.469355	0.25X-0.5X	1240K
I0708	0.352023	0.25X-0.5X	1240K
I0709	0.477074	0.25X-0.5X	1240K
I0745	0.448817	0.25X-0.5X	1240K
I0746	0.459065	0.25X-0.5X	1240K
I1579	0.313284	0.25X-0.5X	1240K
I1580	0.205574	0.1X-0.25X	1240K
I1583	0.408621	0.25X-0.5X	1240K
I1585	0.379656	0.25X-0.5X	1240K
mus005	2.33328	>1X	SG
Tep002	0.678236	0.5X-1X	SG+CP
Tep003	0.655736	0.5X-1X	SG+CP
Tep004	0.447504	0.25X-0.5X	SG+CP

## 4. BULGULAR

### 4.1 Doğruluğun değerlendirilmesi

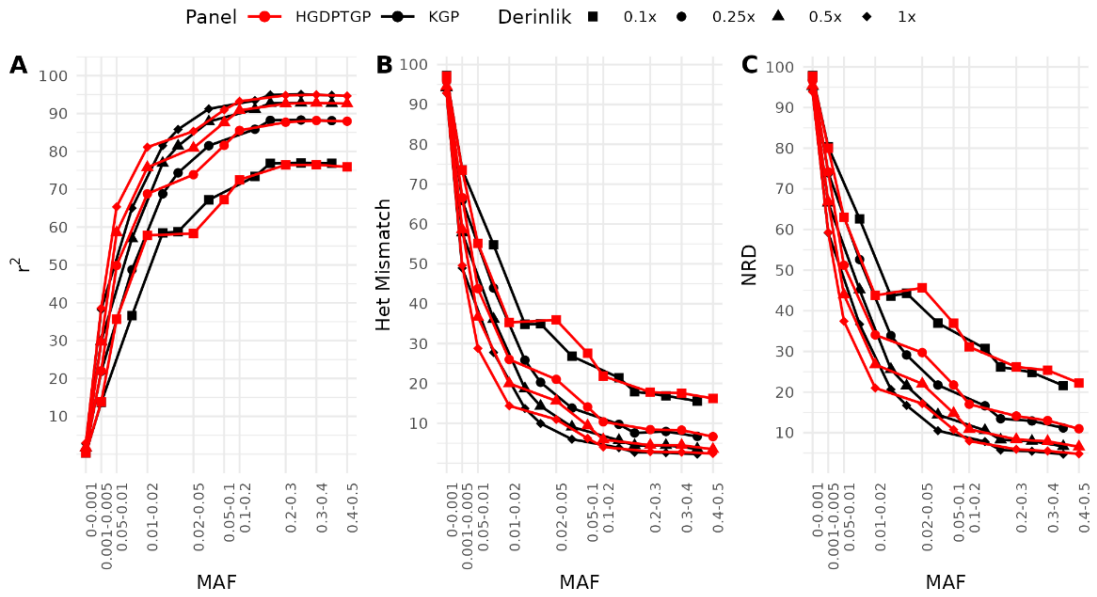
Genotip impütasyonunun doğruluğunu değerlendirmede GLIMPSE yazılımının GLIMPSE\_concordance aracı kullanıldı. Bu araç, impütasyon çıktısını bir doğrulama verisi ile karşılaştırarak doğru ve yanlış atanan homozigot ve heterozigot genotip sayılarını, Pearson  $r^2$  korelasyon katsayısı ile genotip uyumluluğu bilgisini bildirmektedir.

On örneğin genotip uyumluluk değerlerini toplu olarak, minör alel frekans değerlerini x ekseninde gösteren grafikler çizilip incelendiğinde minör alel frekansı %5'ten büyük olan (yaygın) varyantlarda genotip uyumluluk değerlerinin en az; 0.1x için %65, 0.25x için %80, 0.5 için %85, ve 1x için %90 oranlarında olduğu gözlenmiştir. Yaygın varyantlarda heterozigot hata oranı aynı derinlikler için sırasıyla %25, %15, %10 ve %5 değerlerinin altındadır. Son olarak referans dışı uyumsuzluk değerleri yaygın varyantlarda aynı derinlikler için sırayla %35, %20, %15 ve %10'un altındadır (Şekil 4.1, Şekil 4.2).

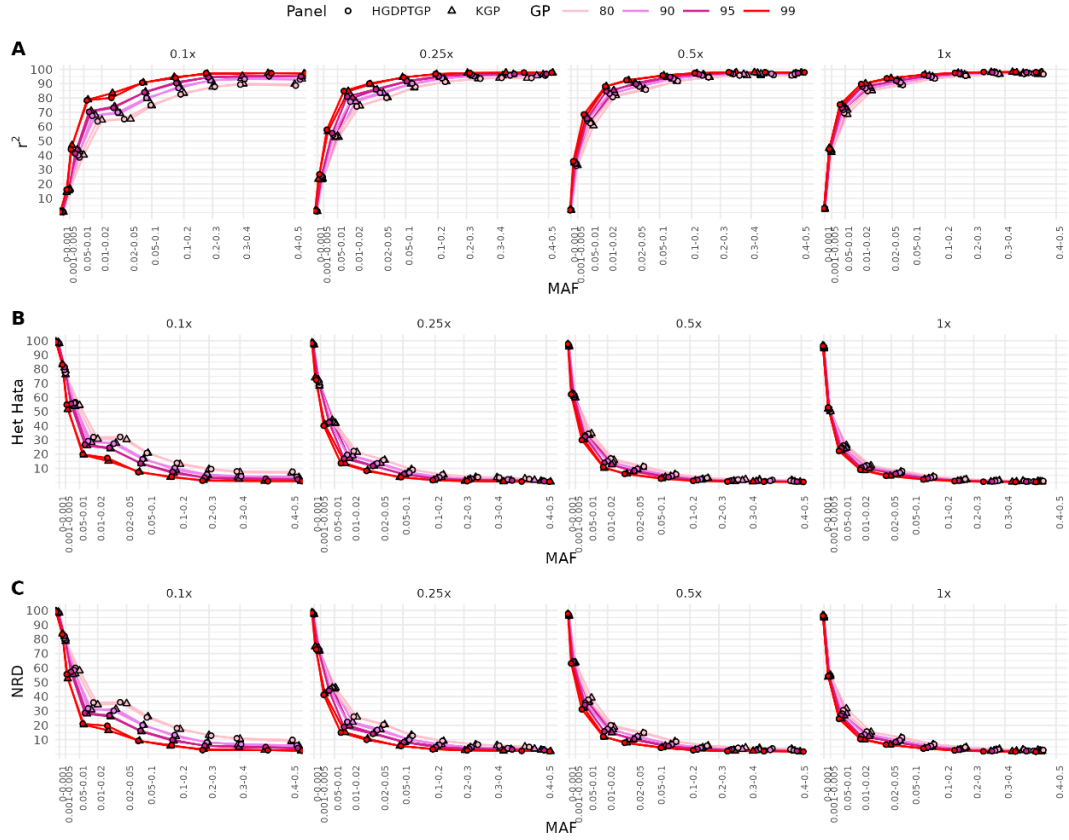


**Şekil 4.1.** İki sürümde A)  $r^2$ , B) heterozigot hata ve C) NRD ile ölçülen genotip impütasyonu doğruluk değerleri.

Ardıl genotip olasılığı filtresi uygulanarak en az %80, %90, %95 ve %99 olasılıkla atanan genotipler uyumluluk karşılaştırmasında kullanıldığında ise yaygın varyantlarda 0.1x için en az %75, 0.25x için %85, 0.5x için %90, 1x için ise %90 olduğu görülmüştür. Ardıl genotip olasılığı alt sınırı yüzde doksan dokuz olarak belirlendiğinde yüzde seksen sınıra göre uyumluluk değerlerinin yaygın varyantlarda %5 ile %15 arası bir oranda yükseldiği görülmüştür. Heterozigot hata oranları ve NRD de benzer biçimde filtrelenmemiş veriye göre azalmıştır (Şekil 4.3).



Şekil 4.2. İki panelde A)  $r^2$ , B) heterozigot hata ve C) NRD ile ölçülen genotip impütasyonu doğruluk değerleri.



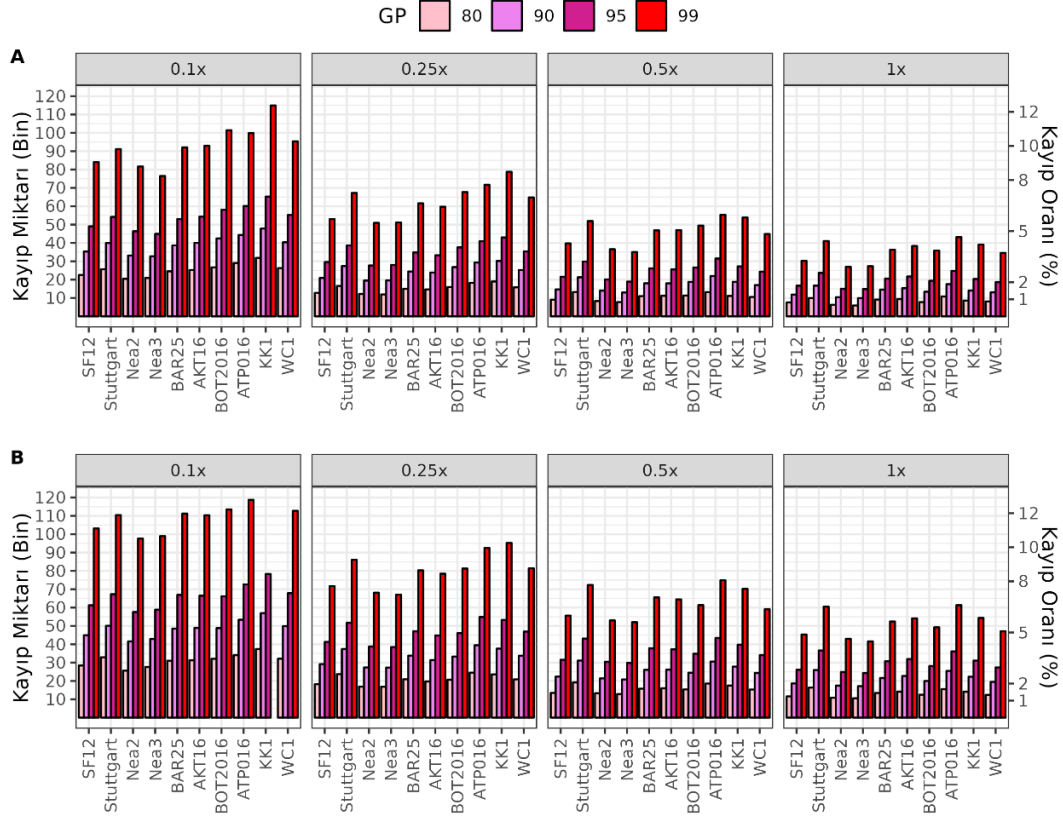
**Şekil 4.3.** GP filtresine göre A) genotip uyumluluk, B) heterozigot hata oranı ve C) NRD değerleri.

Şimdiye dek Türkiye’den elde edilmiş antik genomlardan en yüksek derinliğe sahip olan AKT16 ve BAR25 genomlarının genotip uyumluluk grafikleri incelenmiştir. İki genomun da Anadolu ve Avrupa Neolitik genomlarına benzer (Nea2, Nea3, Stuttgart) doğruluk örüntülerine sahip oldukları görülmüştür. Ardıl genotip filtresi kullanılmayan veri için; AKT16 genomunun  $r^2$  değeri %70-%90, BAR25’in ise %62-%92 aralığındadır. Heterozigot hata oranı AKT16 için %5-%25, Bar25 için %5-%30 aralığında ve NRD değerli ikisi için de %10-%40 aralığındadır (Bkz. EK-3).

## 4.2 Kayıp TNP sayıları

Dört farklı başlangıç genom derinliğinde tüm genomdaki TNP pozisyonlarda impütasyon yapılan toplamda kırk örnek dört farklı genotip olabilirlik değeri ile

filtrelendikten sonra kalan varyant sayıları görselleştirilmiş ve karşılaştırılmıştır. 22. kromozomdaki 900 bin pozisyon içinde, 0.1x için GP99 filtresi 75-115 bin, 1x'te ise 30-40 bin kayba yol açmaktadır. GP80 filtresi ise 0.1x için 20-25 bin, 1x için 10 binin altında kayba neden olmaktadır. %99 olasılıkla filtreleme %80'e göre yaklaşık 20-60 bin arası miktarda daha fazla kayba neden olmaktadır. Genom derinliği 1x'ten 0.1x'e doğru azaldıkça kayıp TNP sayısı 15-40 bin artmaktadır (Şekil 4.4).

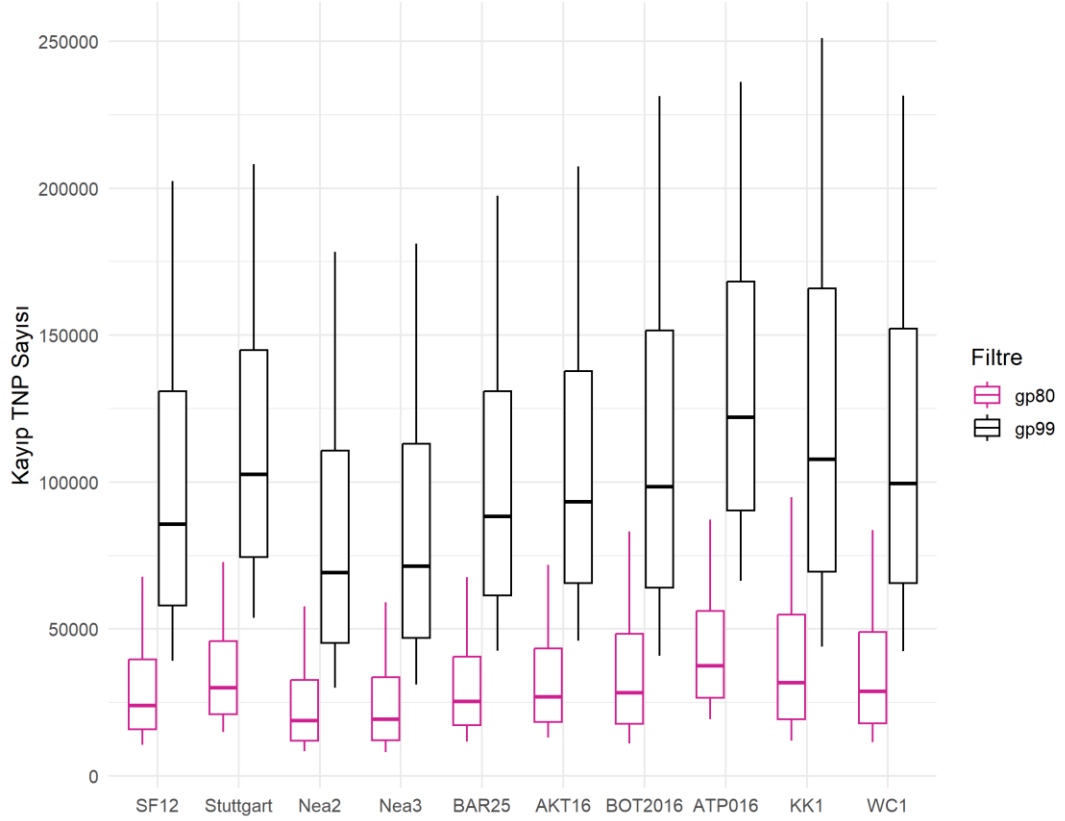


**Şekil 4.4.** On örnek için A) KGP B) KGP+HGDP panelinde GP filtresine göre 22. kromozomda kayıp TNP sayıları ve oranları.

Temel Bileşenler Analizi için kullanılan Human Origins veriseti pozisyonlarında hesaplanan kayıp sayılarına göre, verisetindeki yaklaşık 600 bin pozisyon içinde on örneğin 0.1x derinlikte kayıp sayıları %99 filtrelili veride 200 bin ile 250 bin arasında, %80 filtrelili veride 50 ile 80 bin arasındadır. 1x derinlikteki veriler için değerler çok daha düşük, %99 filtresi için 30-70 bin, %80 filtresi için 15-25 bin arasındadır. (Şekil 4.5) Bazı örnekler, aynı derinlik ve filtre değerlerinde diğerlerine göre daha fazla kayıp vermiştir. (Şekil 4.4 ve 4.5) Hasar onarımına göre



gruplama yapıldığında hasar onarımı (UDG veya USER enzimi uygulaması) yapılmamış örnekler, KK1, WC1, ATP016 ve BOT2016 için kayıp miktarı yaklaşık 0.1x derinliğinde GP99 filtrelili veride 25-50 bin daha fazladır.



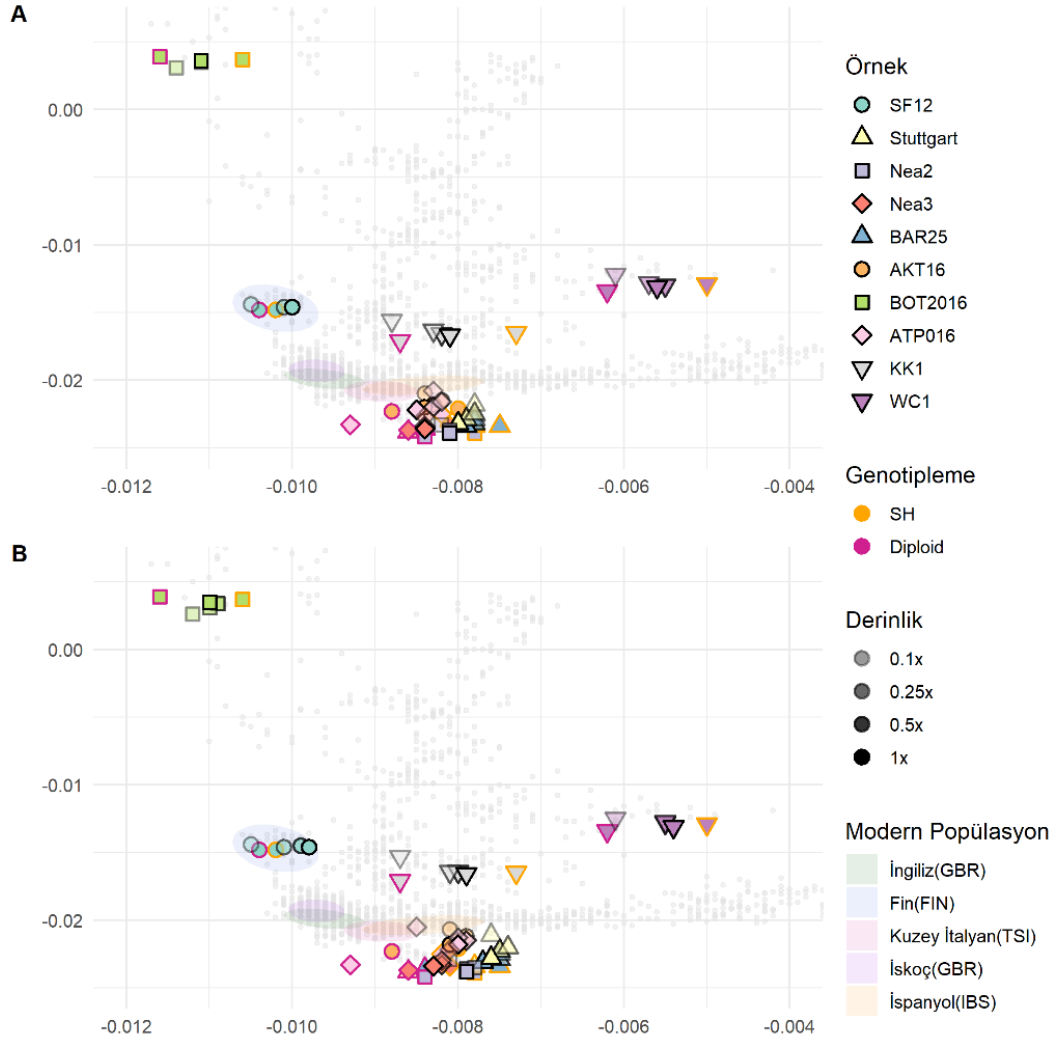
**Şekil 4.5.** On örnek için impütasyon yapılmış verilerin Human Origins verisetindeki yaygın varyantlarda GP filtresine göre kayıp sayıları.

### 4.3 Genotip İmpütasyonunun TBA Sonucuna Etkisi

İmpütasyon çıktısı verinin Human Origins pozisyonlarındaki genotipleri verisine eklenerek smartpca ile TBA yürütüldü. Human Origins verisetindeki modern popülasyonlarla hesaplanan birinci ve ikinci temel bileşenleri eksen alan temel bileşenler alanına (İng. PC space) yansıtılan impütasyon yapılmış ve yapılmamış genomlar incelendi. Verisine insan referans genomunu temsilen eklenen homozigot referans aleller de yansıtılarak farklı yöntemlerle çağrılan genomların konumlanmaları incelendi. Şekil 4.6'da gösterilen sonuçta her iki panelde kırmızı kareler örneğin diploid genotiplenmiş verisini, siyah kare ise sözde-

haploid genotiplenmiş verisini göstermektedir. Modern popülasyon isimlerinin yanında parantez içinde referans panelde mevcut en yakın popülasyon kodu belirtilmiştir.

TBA görselinde, ilk olarak impütasyon verilerinin sözde-haploid veriye göre referans genomdan daha uzakta konumlandıkları gözlemlendi. İkinci olarak impütasyon verilerinin başlangıç genom derinlikleri düştükçe ( $1x$ 'ten  $0.1x$ 'e) modern popülasyonlara daha yakın ve aynı zamanda referans genomdan daha uzak konumlandıkları gözlemlendi. Özellikle  $0.1x$  derinliğinden impütasyon yapılan genomlardan Anadolu ve Avrupa Neolitik kültürlerine ait olanlarda TBA alanında kendilerine en yakın modern popülasyonlara (Güney Avrupalı popülasyonlar) doğru kaydıkları gözlemlendi. Genotip uyumluluğu incelenmesinde impütasyon verisinin karşılaştırıldığı asıl veri olarak kullanılan filtrelenmiş diploid örnekler, TBA alanına yansıtıldığında hem sözde-haploid hem de impütasyon verisine göre birincil bileşende eksi yönde konumlanmışlardır. TBA görsellerinin daha detaylı ve tam hâlleri EK-4'te gösterilmiştir.

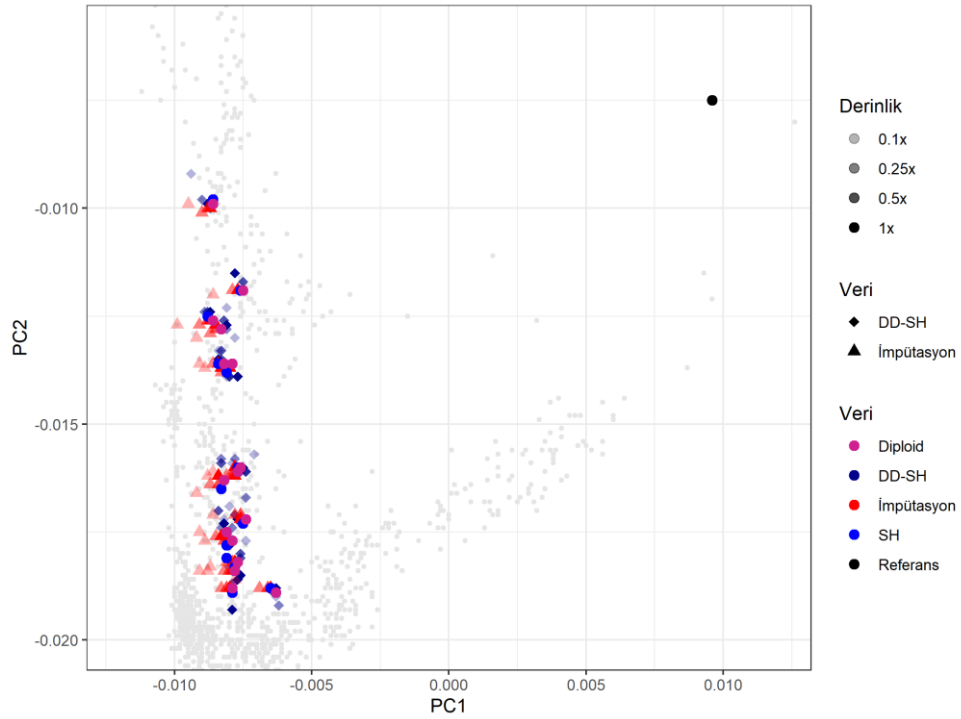


**Şekil 4.6.** Yüksek derinlikli örnekler için A) GP80 filtreli impütasyon verisi ve B) GP99 filtreli impütasyon verisi ile çizilen TBA grafikleri.

### 4.3.1 Modern Örnekler için TBA analizi

Görülen farkın modern veride incelenebilmesi amacıyla Türkiye’den örneklenmiş on dört modern genomun (65) okumaları düşük derinlik (DD) seviyelerine, 0.1x, 0.25x, 0.5x ve 1x genom derinliklerine, seyreltilmiş ve örneklerle impütasyon yapılmıştır. Human Origins verisetine eklenerek öncekilerle aynı biçimde TBA analizi yürütülmüştür. Şekil 4.7’de sunulan TBA görseline göre, modern örnekler için diploid ve sözde-haploid (SH) verilerin konumlarının farklı olmadığı, genotipleme yöntemlerinin bir fark yaratmadığı görülmüştür. İmpütasyon verilerinin ise derinlikleri düştükçe birinci bileşen doğrultusunda eksi yönde

kaydıkları görülmüştür. Antik impütasyon örneklerinde görülen kayma modern veride de görülmüş olmakla birlikte sözde-haploid verinin referans genoma daha yakın olması ve diploid verinin sözde-haploid veri ile yakın konumlanmaması modern örneklerde söz konusu değildir. Bu durum TBA sonuçlarında antik veride yapılan sözde-haploid genotiplemelerin güvenilirliğine şüphe düşürmekte ve antik örnekler için üretilen sözde-haploid veride referans yanlılığına işaret etmektedir.



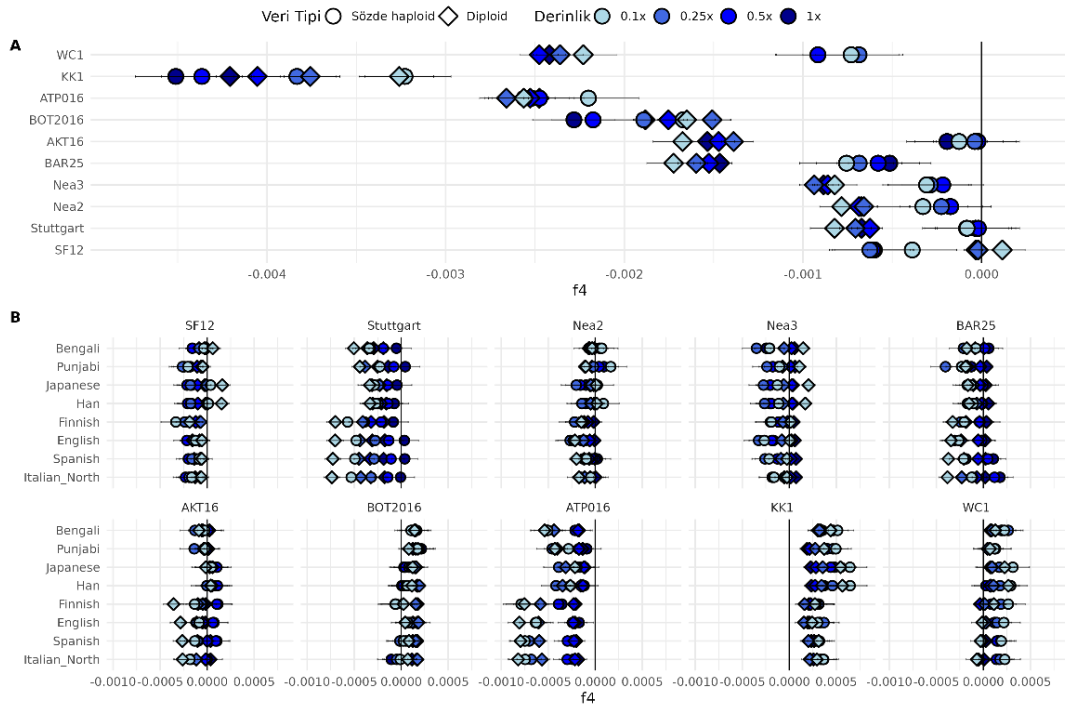
Şekil 4.7. Modern TGP örnekleri için TBA grafiği.

#### 4.4 F-İstatistikleri ve Referans Yanlılığı

TBA sonucuna göre impütasyon ve sözde-haploid verinin referans genoma uzaklıklarının farklı gözlemlenmesi sonrasında bu sonucu test etmek amacıyla popülasyonların ve örneklerin birbirlerine benzerliklerini ölçmek için kullanılan istatistiksel karşılaştırma yöntemi f-istatistikleri hesaplanmıştır.

Şekil 4.8'in A panelinde görülen, f(Dış Grup, Referans Genom, İmpütasyon, Genotipleme) topolojisiyle hesaplanan f4 istatistiklerine göre, on örneğin tamamında referans genom verisi impütasyon yerine sözde-haploid veriye daha yakın

görülmüştür. İmpütasyon verilerinin referans paneldeki popülasyonlara daha çok benzeyip benzemediğini test etmek amacıyla  $f_4$  (Dış Grup, ModernPopülasyon, İmpütasyon, Genotipleme) topolojisi ile  $f_4$  istatistiği hesaplanmıştır. B panelinde gösterilen sonuçlarda, örneklerin hepsinde farklı popülasyonlara yakınlık olduğu görülmüştür. Genel bir örüntü olarak modern popülasyonların impütasyon verisine daha yakın oldukları söylenebilir.  $f$ -istatistikleri sonucunda gözlenen örüntü TBA sonucu ile uyumludur. Sözde-haploid veriler referans genoma impütasyon verisinden daha yakın görünmektedir. Tüm hesaplamalar Human Origins verisetinde, MAF5 ve GP80 filtreli impütasyon verisi ile yapılmıştır.



**Şekil 4.8.**  $f_4$ -istatistikleri. A)  $f_4$  (DışGrup, Href, İmpütasyon, Diploid/Sözde-haploid) ve B)  $f_4$  (DışGrup, ModernPop, İmpütasyon, Diploid/Sözde-haploid) topolojisi ile yapılan hesaplamalar.

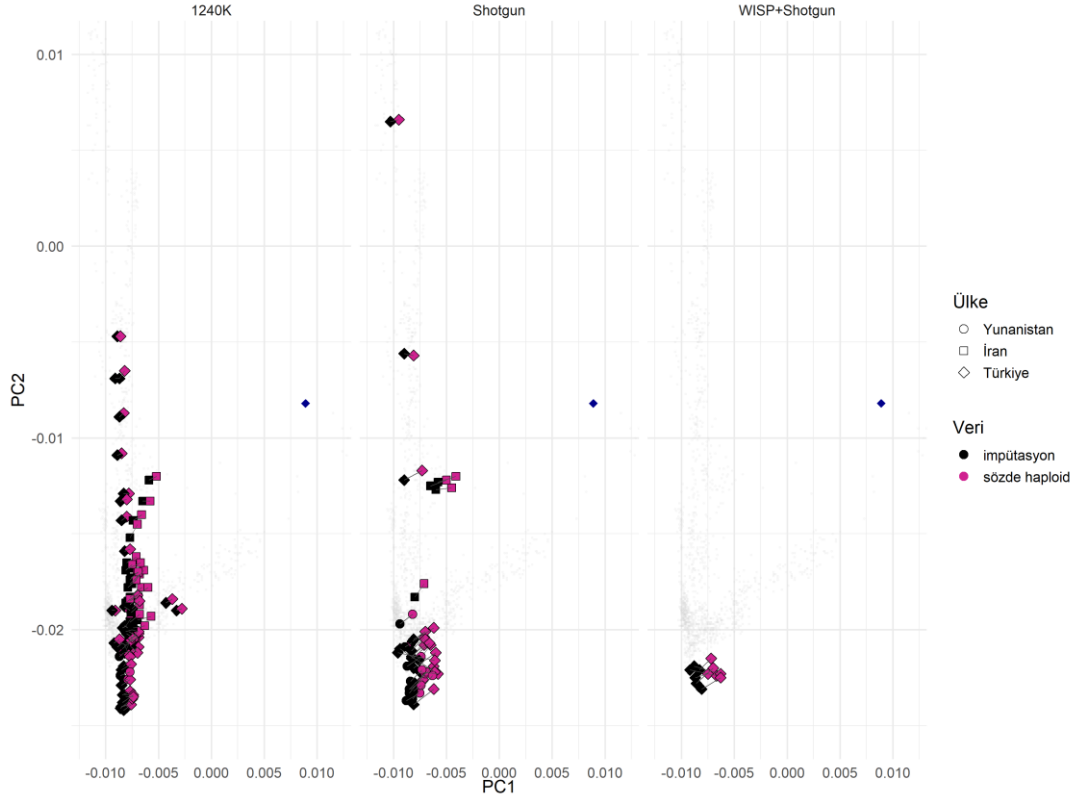
AKT16 ve BAR25 için sonuçlarda, referans sözde-haploid veride impütasyon verisinden biraz daha fazla olsa da referans genoma belirgin bir yakınlık görülmemiştir. İlk topoloji için AKT16 ve BAR25 için diploid veri için sonuçlar benzer iken sözde-haploid ile yapılan karşılaştırmada AKT16'nın  $f_4$ -skoru BAR25 örneğine göre daha düşüktür. AKT16 sözde-haploid verisi Stuttgart örneği ile yakın

bir f4 değerine sahiptir. İkinci topolojide de AKT16'nın skorları daha az, BAR25 skorları daha fazla sapma göstermektedir.

#### 4.5 Düşük Derinlikte Antik Genomların İmpütasyonu

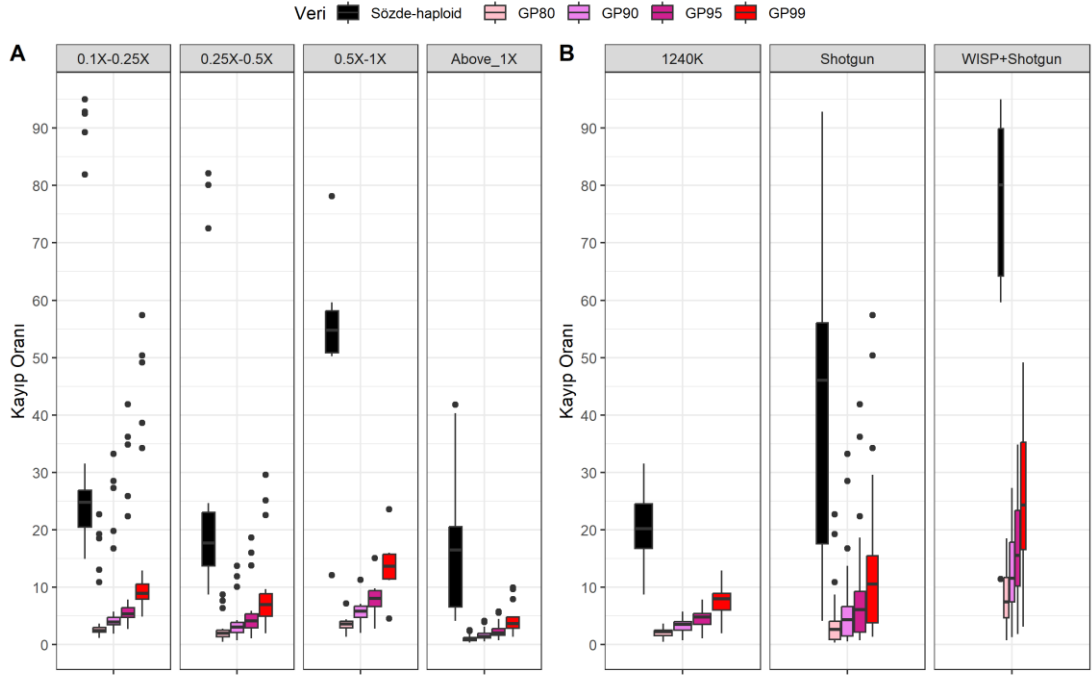
Çalışma kapsamında yayınlanmış 108 yayınlanmış, 34 yayınlanmamış düşük derinlikli antik genoma impütasyon uygulanmıştır. Bu sayının yaklaşık dörtte üçünü Türkiye'den elde edilen genomlar oluşturmaktadır. Genom derinlikleri 0.1x-8x arası olan bu genomların birçoğu, diploid veya sözde-haploid genotipleme yöntemleri ile oldukça az sayıda TNP için genotiplenebilmektedir (Bkz. EK-7). İmpütasyon yapılarak bu genomlardan daha fazla bilgi elde etmek ve analizlere dahil edebilmek amacıyla genotip ataması uygulanmış ve sonuçları öncelikle TBA ile incelenmiştir. Sözde-haploid genotiplenmiş örnekler karşılaştırma yapmak amacıyla impütasyon verisi ile beraber verisetine dahil edilmiş ve Human Origins veriseti modern popülasyonları ile hesaplanan temel bileşenler alanına yansıtılmıştır. İmpütasyon verileri verisetindeki yaygın varyantları içermektedir (MAF > 0.5) ve GP99 filtrelidir.

Şekil 4.9'da sunulan TBA görseli incelendiğinde impütasyon verilerinin sözde-haploid verilerden her bir örnek için farklı bir yönde konumlandığı, referans genoma her birinin haploide göre daha uzak olduğu görülmektedir. Bu örüntü genom derinlikleri düşürülüp impütasyon yapılan örneklerdeki ile tutarlıdır, impütasyon verisi sözde-haploid veriye göre referans genomdan daha uzakta konumlanmaktadır. Düşük derinlikli genom örnekleme veri üretimi yönünden tek tip olmadığı için yakalama verisi ile üretilen ve rastgele dizileme ile üretilen verileri de karşılaştırmak mümkün olmuştur. TNP yakalama yöntemi ile üretilen örnekler Human Origins verisetinde daha az eksik TNP'ye sahip olmalarının yanında TBA grafiğinde sözde haploid veri ile daha yakın konumlanmışlardır.



**Şekil 4.9.** Düşük derinlikli yayınlanmış örneklerin TBA grafiği. Referans genom mavi elmas ile gösterilmiş ve aynı örnek için iki veri tipi arasında çizgi eklenmiştir.

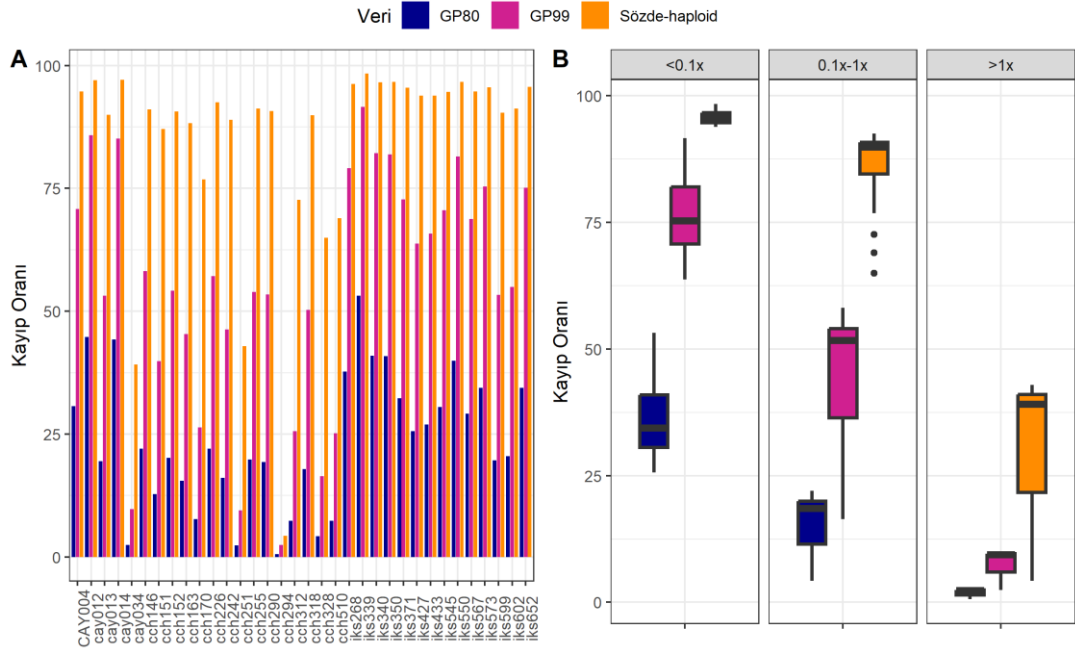
Şekil 4.10'da örneklerin veri tipine ve derinlik grubuna göre kayıp oranları gösterilmiştir. A panelinde derinlik ile kayıp oranı arasında doğru orantı görülmezken bunu açıklayan sonuç B panelinde görülmektedir. TNP yakalama (1240K) örnekleri bu veriseti pozisyonlarında sözde haploid veride de çok daha az eksik TNP'ye sahiptirler, yakalama örnekleri %10-30 aralığında, rastgele dizileme (SG) %5-%95 aralığında kayba sahiptir. İmpütasyon kayıp oranlarını azaltsa da veri tipinden kaynaklı fark belirgindir. Her bir örnek için detaylı kayıp oranları (SH, GP80, GP90, GP99) EK-7'de görülebilir.



**Şekil 4.10.** Düşük derinlikli örnekler için A) derinliğe ve B) veri dizileme yöntemine göre Human Origins verisetinde kayıp oranlarının impütasyon ve sözde haploid verilerde karşılaştırması.

Araştırma projesi kapsamında genomları dizilenmiş 34 Neolitik ve Geç Kalkolitik çağlara ait antik genomlarda da yayınlanmış rastgele dizileme genomlara benzer bir sonuç elde edilmiştir. Bu örnekler arasında on beş tanesi 0.1x genom derinliğinin altında olduğu için bu grupta 0.1x altı örnekler ile bir karşılaştırma yapılmıştır. Her bir örnek için sonuçlar Şekil 4.11A'da görülmektedir. Derinlik grubuna göre inceleme yapıldığında, GP80 filtresi uygulanmış impütasyon verisi için yaygın varyantlarda, 0.1x altı genomlarda HO verisetinde kayıp oranı %25-%55 aralığında iken GP99 filtreli veri %75 medyan değere sahiptir (Şekil 4.11B). 0.1x-1x aralığındaki örneklerde ise kayıp oranı GP80 için %25'in altında ve GP99 için medyan %50 değere sahiptir. 1x üzerindeki birkaç genom içinse kayıp oranları çok daha düşüktür (<%10). Bu oranlar aynı pozisyonlarda sözde-haploid genotiplenmiş veride örneklerin çoğunda kayıp oranı olan %80'den yüksektir. 0.1x altı genomlara impütasyon yaparak yaygın varyantlarda GP80 filtresi ile yaklaşık %25-%50 kayıp, GP99 ile %65-%85 kayıp elde etmek mümkündür. Tüm örnekler için kayıp oranları tablosu EK-8'de görülebilir.



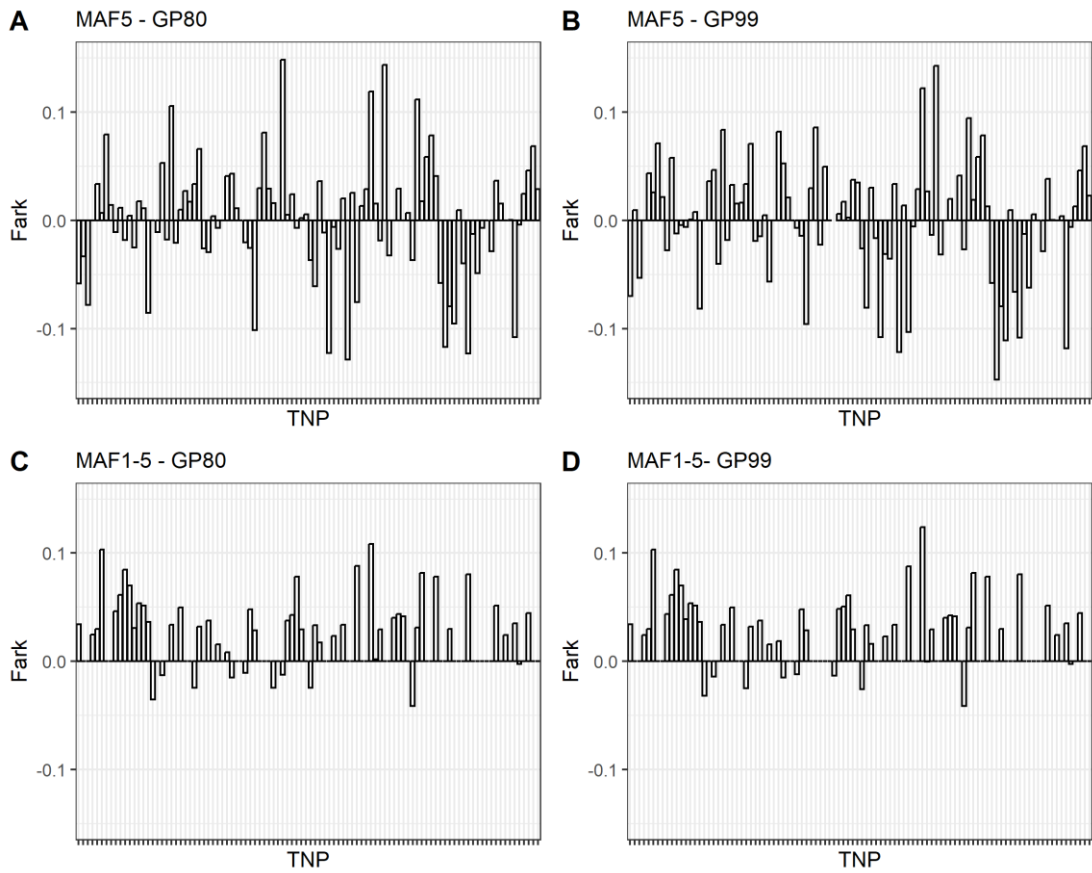


**Şekil 4.11.** Proje örneklerinde veriseti pozisyonlarında kayıp oranlarının söздеhaploid veri ile A) örneğe göre ve B) derinlik grubuna göre karşılaştırması.

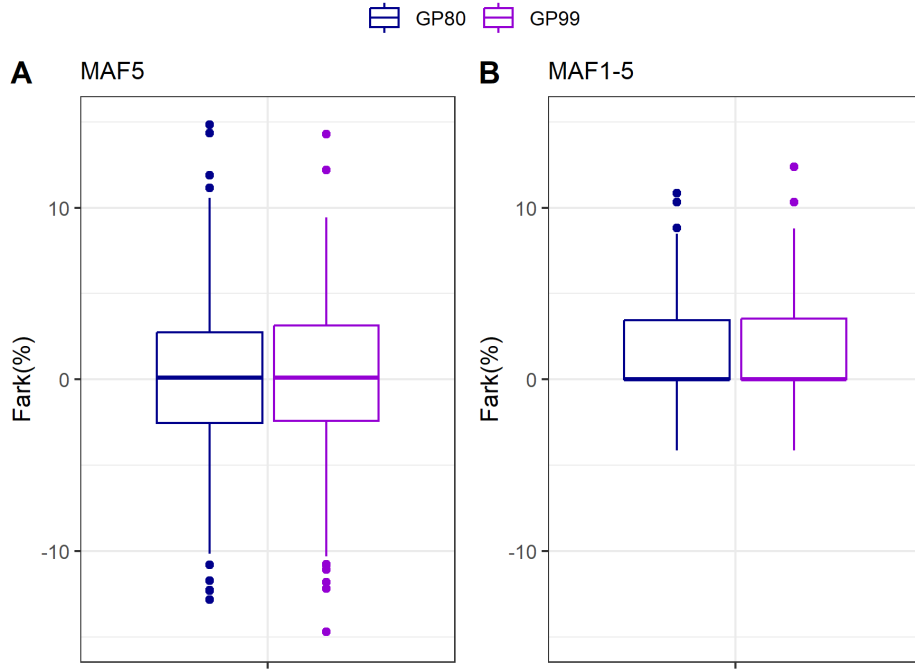
#### 4.6 Alel Sıklık Değerleri Karşılaştırması

41 Neolitik dönem Anadolu örneğinde alel sıklıklarının impütasyon çıktısı veya BAM dosyasından hesaplanması arasında fark olup olmadığı, iki değer arasındaki farklar karşılaştırılarak değerlendirilmiştir. Şekil 4.12 A ve B panellerinde görülen sonuçlara göre yaygın varyantlar için %15'e kadar farklı sıklık değerleri görülmüştür. C ve D panellerinde görülen az yaygın (MAF1-5) varyantlar için ise %10'a kadar fark görülmüştür. Çoğu TNP pozisyonu için elde edilen frekans değerleri yaygın varyantlar için %10'dan ve az yaygın olanlar için %5'ten az farklıdır. Şekil 4.13'te farkların dağılımı kutu-çizgi grafiği ile gösterilmiştir. MAF değerleri %1 ile %5 arasında olan TNP'ler için çoğunda TNP'nin sıklığı impütasyon çıktısı veri için 0 olarak rapor edilmiştir. Bazı TNPlerde impütasyon çıktısından PLINK ile hesaplanan sonuç, bazılarında ise ANGSD ile BAM dosyalarından hesaplanan sonuç daha yüksektir. Bu sonuçları elde ederken kullanılan genotip sayısı veya örneklem büyüklüğü farklıdır, birçok pozisyon için impütasyon verisinde çok daha fazladır. Bazı TNP'lerde ise aynı frekans değerleri bulunmuştur. Örneğin 8.

Kromozom – 142655246 pozisyonundaki G referans, A alternatif aleline sahip TNP için; genotiplenmiş veriden elde edilen frekans değeri 0.436010, örnek sayısı ise 11’dir. Aynı TNP için GP80 filtreli impütasyon verisi ile taranmış frekans değeri 0.4359, genotip sayısı ise 39’dur. Bu durumda MAF değeri %5’ten büyük olan bu TNP için GP80 filtreli, en az 0.1x başlangıç genom derinlikli verinin impütasyon çıktısı daha büyük örneklem ile güvenli bir şekilde bilgi vermiştir. Tüm alel sıklık hesaplama sonuçları EK-9 ve EK-10’da görülebilir.



**Şekil 4.12.** Yaygın (A, B) ve az yaygın (C, D) varyantlar için GP80 ve GP99 filtreli impütasyon veya hizalama verisinden hesaplanan alel sıklık hesabı farkları.



**Şekil 4.13.** Yaygın (A) ve az yaygın (B) varyantlar için GP80 veya GP99 filtrelili veriyile ve hizalama verisiyle hesaplanan değerlerin fark dağılımı.

## 5. TARTIŞMA

Bu çalışmada antik genom arařtırmalarında kullanılabilen verileri artırmak amacıyla düşük genom kapsamına ve okuma derinliđine sahip olan, parçalı, ve az miktarda veriye sahip olan antik genom dizileme verilerine impütasyon uygulanmıřtır.

Dođruluk deđerlendirmesi yapmak amacıyla filtreli genom boyu okuma derinliđi 8x'in üzerinde ikisi Türkiye'den on adet genom kullanılmıřtır. Her bir örnek için dört farklı genomboyu okuma derinliđine sahip olacak řekilde okumaları seyreltilerek impütasyon uygulanmıřtır. Asıl verilere diploid genotipleme yapılmıř, belirli filtreler kullanılarak karřılařtırma verisi elde edilmiřtir. Tüm örnekler için MAF deđerleri %5 ve üzeri varyantlarda %80 üzerinde genotip olasılıđı filtreli sonuřlarda daha yüksek genotip uyumluluđu oranları görölmüřtür. Buna göre impütasyon sonrası ardıl genotip olasılıklarına göre filtreleme yaparak yaygın varyantlarda %90 ve üzeri dođrulukla genotipleri tahmin etmek mümkündür.

Yüksek dođruluk oranlarında genotip atamaları yapmak istenen sonuř olmakla birlikte bu sonucun ne kadar genotip için geđerli olduđu da önemlidir. Bir bařka deyiřle ardıl genotip olasılıklarına göre filtrelenen verilerde ne kadar genotipin elde edildiđi de dikkate alınmalıdır. Bu çalışmada kullanılan Human Origins veri seti pozisyonları için yapılan karřılařtırmada 0.25, 0.5x ve 1x genom derinliđinden impütasyon yapılan ve GP80 filtreli verilerde 600 bin küsur pozisyon içinden en fazla 50 bin kadarının kaybolduđu, GP99 için de en fazla 120-160 bin kadarının kaybolduđu görölmüřtür. 0.1 x derinlikli genomlar için bu sayılar, GP80 filtresi için 60-80 ve GP99 filtresi için 125-250 bin aralıklarında olup belirgin bir biçimde daha fazladır. Bu sonuřlara göre kullanılacak örneđe ve bařlangıç derinliđine bađlı olarak kaybedilen varyant sayısının deđiřtiđi ve yüksek güvenilirlikle filtrelenen verilerde daha fazla veri kaybedildiđi görölmüřtür. Filtrelere bađlı kayıplara rađmen impütasyon yapılan örneklerde sözde haploid veriye kıyasla belirgin miktarda fazla sayıda varyant elde edilmektedir.

Düşük derinlikli örnekler için, 1240K yakalama yöntemi ile üretilmiř örneklerin rastgele dizileme ile üretilmiř örneklere kıyasla Human Origins veriseti

pozisyonlarında çok daha az kayıp göstermeleri bu örneklerin dizilendiği pozisyonların Human Origins veriseti pozisyonlarının birçoğunu içermesinden kaynaklanmaktadır. Ayrıca bu örneklerin birçoğu 1240K yakalama pozisyonlarında yüksek derinlikle dizilenmiştir. Bu sonuç veri tipinin ve analizde kullanılan verisetinin kayıp miktarlarını belirgin biçimde değiştirebildiğini göstermektedir.

Antik DNA araştırmalarında yaygın kullanılan bir boyut düşürme tekniği olan TBA ile çok sayıda popülasyonu bir arada incelemek mümkündür. Az miktarı ve parçalı yapısından dolayı antik verilerde diploid genotip çağırma oldukça az miktarda güvenilir veri vermekte, bunun yerine sözde-haploid veriler analizlere dahil edilmektedir. İmpütasyon verisinde ise genotipler diploid olarak tahmin edilmektedir. Bunu yanı sıra sözde-haploid veride referans genom kaynak olarak kullanılarak genotipler çağırılmakta iken impütasyon işleminde referans olarak çok sayıda genomu içeren paneller kullanılmaktadır. Bu iki veri tipini temel bileşenler analizine birlikte dahil ederek eğilimleri karşılaştırmalı olarak incelenmiştir. TBA bir ve ikincil bileşenleriyle oluşturulan grafikte birkaç farklı sonuç görülmüştür: Öncelikle, impütasyon verisi sözde-haploid veriye göre referans genoma daha uzak konumlanmıştır. İkinci olarak, impütasyon verisinin başlangıç genom derinliği düştükçe referans paneldeki popülasyonlara yaklaşmışlardır.

TBA sonucuna göre sözde-haploid verinin referans yanlılığı içerdiği düşünülmüş ve ikinci bir test olarak f-istatistikleri hesaplanarak devam edilmiştir. f4 istatistikleri ile iki farklı grubun bir üçüncüye yakınlığı paylaşılan alel sayıları üzerinden sayısal bir ölçü hâline getirilmektedir. On örnek iki farklı topolojide karşılaştırılmıştır. Birinde referans genomun her örnek için impütasyon çıktısı, diploid ve sözde-haploid verilere yakınlığı, diğerinde ise kullanılan referans panelde bulunan modern popülasyonlar veya onlara çok yakın popülasyonların hangi veri tipine daha yakın oldukları test edilmiştir. Referans genomun impütasyon verisinden çok sözde-haploid veriye benzerlik gösterdiği görülmüştür. TBA sonucu ile uyuşan bu sonuç örneklerin sözde-haploid genotip verisinin impütasyon verisine göre referans genoma daha çok benzediği, referans yanlılığına sahip olduğu şeklinde yorumlanabilir. İkinci karşılaştırmada modern popülasyonların örneklere göre farklı sonuçlar verdiği görülmüştür. Referans paneldeki modern popülasyonlar ile yeterince

temsil edilmeyen bir bölgede yer alan ve hasar onarımı görmemiş KK1, WC1, BOT2016 örnekleri için modern popülasyonların sözde-haploid veriden çok impütasyon verisine benzerlik gösterdiği söylenebilir.

Antik verilerde görülen sapmanın modern veride test edilmesi amacıyla on dört modern genoma aynı yöntemle impütasyon yapılmıştır. Modern örnekler için diploid, sözde-haploid ve impütasyon verisi TBA görselinde incelendiğinde impütasyon verisinin başlangıç genom derinliği düştükçe bir yönde kaydıkları görülmüştür. Sözde-haploid ve diploid veriler farklı konumlanmamakta ve impütasyon verisi ile sözde-haploid veri arasında referans genoma doğru veya zıttına bir eğilim görülmemiştir. Modern genomlarda böyle bir sonuç görülmemesi üzerine referansa yakınlığın antik verilerde görüldüğü düşünülmüştür.

İmpütasyon verisinin doğruluk ve yaygın analizlerde kullanılabilirliği değerlendirildikten sonra Anadolu ve çevre bölgelerden, genom derinliği alt sınırı 0.1x olarak belirlenmiş bir genom örnekleme impütasyon yapmak üzere seçilmiştir. Önceki sonuçlara göre impütasyonun 0.1x derinliğindeki veride yüksek oranda doğru sonuç vermesi için GP99 filtresi ve MAF değeri %5'ten büyük varyantlar seçilerek Human Origins pozisyonlarında bir veri seti hazırlanmış ve TBA yürütülmüştür. Önceki antik veri TBA sonucu ile benzer olarak sözde-haploid ve impütasyon verisinde bir ve ikincil bileşenlere göre konumlanmada fark görülmüştür. Sözde-haploid veri referans genom yönüne doğru konumlanmış görünürken impütasyon verisi referans panele daha yakın konumlanmıştır.

Farklı güvenilirlik eşikleri olarak yorumlanabilecek ardıl genotip olasılığı filtrelerini kullanarak impütasyon yapılmış düşük derinlikli veriden farklı TNP'ler için alel sıklık değerleri elde etmek de mümkündür. Çağrılmış genotipler ile hesaplanan veride düşük derinlikli örnekler için daha fazla eksik genotip vardır. Güvenilir bir alel sıklık hesabı yapmak için çok sayıda örnekten aynı genomik pozisyonlarda veri elde edilmesi gerekmektedir. Dolayısıyla örneklerin düşük derinlikte olduğu gruplar için alel sıklığı taraması yapmak oldukça zordur. Bu noktada impütasyon verisinin bir olumlu etkisi örneklem boyutunu büyütmesidir. İmpütasyon verisi kullanılarak, yaygın varyantlarda ve GP filtreli veri kullanarak alel sıklığı bilgisini elde etmek mümkün görünmektedir. Ancak genotiplenmiş veri ile

görülen %15'e varan farkın nedenleri ve impütasyon verisinin bu işlem için güvenilirliği üzerine daha fazla inceleme yapmak gerekmektedir.

Bu çalışmada Anadolu ve çevresi bölgelerden antik genom örneklerinde impütasyon kullanmanın olanakları araştırılmıştır, ancak impütasyon doğruluğunu değerlendirmek için bu bölgeden yalnızca birkaç adet iyi kalitede genom verisi mevcuttur. Farklı tarih dönemlerinden elde edilecek yüksek kalitede genomlar impütasyon doğruluğu bilgisini daha güvenilir hale getirecektir. Mevcut iki genom Erken Neolitik döneme tarihlenmiştir ancak Neolitik dönem sonrasında Anadolu ve yakın çevresinde görülen hareketliliğe bağlı olarak genetik çeşitliliğin bir miktar artış gösterdiği düşünülmektedir. Bölgedeki insan hareketliliğinin ve genetik çeşitliliğin daha da arttığı Tunç ve Demir Çağlarına tarihlenen örnekler için, impütasyonun bu panel ile Neolitik dönem örneklerindeki kadar iyi bir sonuç verip vermeyeceği bu dönemlerden iyi kalitede bir genom mevcut olmadığı için incelenememiştir. Yine de neolitik dönem örneklerini referans olarak diğer dönemlerden örnekler için tahmin ve yorum yapılabilir.

Bu çalışma açık erişimli en geniş kapsamlı ve yüksek kaliteli genomik veri kaynaklarından 1000 Genom projesi verisi referans panel olarak kullanılarak yapılmıştır. Elde edilen sonuçlara göre bu panel ile hedef örneklerde yaygın varyantlar için iyi sonuçlar alınmakla birlikte düşük derinlikli örnekler için yaygın olmayan varyantlar için iyi sonuç elde etmek oldukça zordur. Bunun için daha fazla örnek içeren panellere ihtiyaç vardır. Büyük genom verisi işleme projelerinin modern genomik araştırmalarına sunacağı katkı kadar antik genom çalışmalarına da fayda sağlayacağı açıktır.

## 6. SONUÇ VE ÖNERİLER

Bu tez çalışmasında Türkiye ve Batı Avrasya'dan düşük genom derinlikli insan antik genom verileri için genotip impütasyonu uygulamanın sınırları araştırılmıştır. Çalışma şu şekilde özetlenebilir:

1- İmpütasyon doğruluğu; genotip uyumluluk, heterozigot hatası ve referans-dışı uyumsuzluk ölçütleri ile ardıl genotip olasılığı ve minör alel sıklığı parametrelerine göre 10 adet yüksek genom derinlikli (8x-63x) örnek dört farklı başlangıç genom derinliğinde GLIMPSE2 ile KGP 3. Faz 30X verisi referans panel olarak kullanılarak impütasyon yapılarak değerlendirilmiş, dört farklı ardıl genotip olasılığı filtresine göre kaybedilen varyant sayıları incelenmiştir.

2- Genom boyu derinlikleri 0.1x ile 8x arasında olan 108 yayınlanmış ve proje kapsamında üretilen, çoğunluğu 1x altında olan 34 örneğe GLIMPSE2 yazılımı ile tüm otozomlarda KGP Paneli 3. Faz 30X referans paneli kullanılarak impütasyon uygulanmıştır.

3- İmpütasyon verisinin temel bileşenler analizinde nasıl sonuç verdiği genom derinliğine ve ardıl genotip olasılığı filtrelerine göre sözde-haploid veri ile karşılaştırılarak incelenmiştir.

4- İmpütasyon verisinin f4 istatistiklerine göre referans genoma ve modern popülasyonlara yakınlığı sözde-haploid veri ile karşılaştırılarak incelenmiştir.

5 – Türkiye'den modern on dört genoma genom boyu derinlikleri düşürülüp impütasyon uygulanarak TBA analizinde nasıl sonuç verdikleri incelenmiştir.

6- Rastgele seçilen MAF değeri KGP panelindeki popülasyonlarda %5'ten büyük veya %1 ile %5 arasında olan biallelik TNPlar için impütasyon çıktısından ve hizalama verisinden elde edilen minör alel sıklık değerleri karşılaştırılmıştır.

Elde edilen sonuçlara göre mevcut referans panelleri kullanarak Anadolu ve Batı Avrasya'dan düşük derinlikli antik insan genomlarında yaygın varyantlar için yüksek oranda doğrulukla genotip ataması yapmak ve bu veriyi yaygın kullanılan



popülasyon genetiği analizlerinde kullanmak mümkündür. İmpütasyon verisi analizlerde sözde-haploid veri yerine tercih edilebilir.

Genom derinliği 0.25x ve üstü olan veri, GP değeri en az 80 olacak şekilde filtrelenerek analizlerde MAF değeri %1 üzeri olan varyantlar için güvenilir sonuçlar elde edilebilir. 0.1x derinliğine sahip veri için GP99 filtreli veri ve MAF değeri %5'in üzerindeki varyantlar için sonuç daha güvenilir görünmekle birlikte, TBA sonucunda referans panele doğru bir sapma göstermektedir. 0.25x altı derinliklerde kullanılacak örneklere göre farklı filtrelemeler ile değerlendirme yapmak gerekebilir.

İleride düşük genom boyu derinliğe sahip antik insan genom örneklerine, ve genel olarak tüm impütasyon yapılması planlanan çalışmalarda;

- örneklerin mümkün olduğunca iyi temsil edildiği geniş kapsamlı, öncelikle olabildiğince çok sayıda örnek, ve farklı popülasyonlardan veri içeren referans paneller kullanılması, bunun için güncel verilerin takip edilmesi,
- yapılacak analizlere göre genotip uyumluluk değerlerini her örnek için inceleyerek, uygun alt sınır MAF değerleri ve GP filtreleri belirlenmesi, gerekliyse birkaç farklı filtrenin karşılaştırılması,
- seçilen filtrelere göre kaybedilen genotip sayılarının göz önünde bulundurulması faydalı olacaktır.

## 7. KAYNAKLAR

1. Brown T, Brown K. *Biomolecular Archaeology An Introduction*. Wiley-Blackwell; 2011.
2. Liu Y, Mao X, Krause J, Fu Q. Insights into human history from the first decade of ancient human genomics. *Science*. 2021 Sep 24;373(6562):1479–84.
3. Orlando L, Allaby R, Skoglund P, Der Sarkissian C, Stockhammer PW, Ávila-Arcos MC, et al. Ancient DNA analysis. *Nat Rev Methods Primers*. 2021 Feb 11;1(1):1–26.
4. Malaspinas AS, Tange O, Moreno-Mayar JV, Rasmussen M, DeGiorgio M, Wang Y, et al. bammds: a tool for assessing the ancestry of low-depth whole-genome data using multidimensional scaling (MDS). *Bioinformatics*. 2014 Oct 15;30(20):2962–4.
5. Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, et al. Ancient Admixture in Human History. *Genetics*. 2012 Nov 1;192(3):1065–93.
6. Barlow A, Hartmann S, Gonzalez J, Hofreiter M, Paijmans JLA. Consensify: A Method for Generating Pseudohaploid Genome Sequences from Palaeogenomic Datasets with Reduced Error Rates. *Genes (Basel)*. 2020 Jan 2;11(1):50.
7. Ros-Freixedes R, Battagin M, Johnsson M, Gorjanc G, Mileham AJ, Rounsley SD, et al. Impact of index hopping and bias towards the reference allele on accuracy of genotype calls from low-coverage sequencing. *Genetics Selection Evolution*. 2018 Dec 13;50(1):64.
8. Günther T, Nettelblad C. The presence and impact of reference bias on population genomic studies of prehistoric human populations. *PLOS Genetics*. 2019 Tem;15(7):e1008302.
9. Vaddadi NSK, Mun T, Langmead B. Minimizing Reference Bias with an Impute-First Approach [Internet]. *bioRxiv*; 2023 [cited 2024 Apr 2]. p. 2023.11.30.568362. Available from: <https://www.biorxiv.org/content/10.1101/2023.11.30.568362v1>
10. Lachance J, Tishkoff SA. SNP ascertainment bias in population genetic analyses: Why it is important, and how to correct it. *Bioessays*. 2013 Sep;35(9):780–6.
11. Geibel J, Reimer C, Pook T, Weigend S, Weigend A, Simianer H. How imputation can mitigate SNP ascertainment Bias. *BMC Genomics*. 2021 May 12;22(1):340.
12. Li Y, Willer C, Sanna S, Abecasis G. Genotype imputation. *Annu Rev Genomics Hum Genet*. 2009;10:387–406.
13. Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, et al. A Genome-Wide Association Study of Type 2 Diabetes in Finns Detects Multiple Susceptibility Variants. *Science*. 2007 Jun 1;316(5829):1341–5.
14. Das S, Abecasis GR, Browning BL. Genotype Imputation from Large Reference Panels. *Annu Rev Genomics Hum Genet*. 2018 Aug 31;19:73–96.
15. Li N, Stephens M. Modeling Linkage Disequilibrium and Identifying Recombination Hotspots Using Single-Nucleotide Polymorphism Data. *Genetics*. 2003 Dec 1;165(4):2213–33.
16. Howie BN, Donnelly P, Marchini J. A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies. *PLOS Genetics*. 2009 Haz;5(6):e1000529.

17. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. *Nat Genet.* 2016 Oct;48(10):1284–7.
18. Browning BL, Zhou Y, Browning SR. A One-Penny Imputed Genome from Next-Generation Reference Panels. *Am J Hum Genet.* 2018 Sep 6;103(3):338–48.
19. Davies RW, Kucka M, Su D, Shi S, Flanagan M, Cunniff CM, et al. Rapid genotype imputation from sequence with reference panels. *Nat Genet.* 2021 Jul;53(7):1104–11.
20. Rubinacci S, Hofmeister RJ, Sousa da Mota B, Delaneau O. Imputation of low-coverage sequencing data from 150,119 UK Biobank genomes. *Nat Genet.* 2023 Jul;55(7):1088–90.
21. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet.* 2016 Oct;48(10):1279–83.
22. Taliun D, Harris DN, Kessler MD, Carlson J, Szpiech ZA, Torres R, et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature.* 2021 Feb;590(7845):290–9.
23. Byrska-Bishop M, Evani US, Zhao X, Basile AO, Abel HJ, Regier AA, et al. High-coverage whole-genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios. *Cell.* 2022 Sep 1;185(18):3426–3440.e19.
24. 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 Mar 20;367(6484):eaay5012.
25. Schurz H, Müller SJ, van Helden PD, Tromp G, Hoal EG, Kinnear CJ, et al. Evaluating the Accuracy of Imputation Methods in a Five-Way Admixed Population. *Front Genet.* 2019 Feb 5;10:34.
26. Hwang MY, Choi NH, Won HH, Kim BJ, Kim YJ. Analyzing the Korean reference genome with meta-imputation increased the imputation accuracy and spectrum of rare variants in the Korean population. *Front Genet.* 2022;13:1008646.
27. Shi M, Tanikawa C, Munter HM, Akiyama M, Koyama S, Tomizuka K, et al. Genotype imputation accuracy and the quality metrics of the minor ancestry in multi-ancestry reference panels. *Briefings in Bioinformatics.* 2024 Jan 1;25(1):bbad509.
28. Cahoon JL, Rui X, Tang E, Simons C, Langie J, Chen M, et al. Imputation accuracy across global human populations. *The American Journal of Human Genetics.* 2024 May 2;111(5):979–89.
29. Chattopadhyay A, Lu TP. Overcoming the challenges of imputation of rare variants in a Taiwanese cohort. *Transl Cancer Res.* 2020 Jul;9(7):4065–9.
30. Razali RM, Rodriguez-Flores J, Ghorbani M, Naeem H, Aamer W, Aliyev E, et al. Thousands of Qatari genomes inform human migration history and improve imputation of Arab haplotypes. *Nat Commun.* 2021 Oct 12;12(1):5929.
31. Dekeyser T, Génin E, Herzig AF. Opening the Black Box of Imputation Software to Study the Impact of Reference Panel Composition on Performance. *Genes.* 2023 Feb;14(2):410.
32. Hui R, D’Atanasio E, Cassidy LM, Scheib CL, Kivisild T. Evaluating genotype imputation pipeline for ultra-low coverage ancient genomes. *Sci Rep.* 2020 Oct 29;10(1):18542.

33. Ausmees K, Sanchez-Quinto F, Jakobsson M, Nettelblad C. An empirical evaluation of genotype imputation of ancient DNA. *G3 Genes|Genomes|Genetics*. 2022 Jun 1;12(6):jkac089.
34. Sousa da Mota B, Rubinacci S, Cruz Dávalos DI, G. Amorim CE, Sikora M, Johannsen NN, et al. Imputation of ancient human genomes. *Nat Commun*. 2023 Jun 20;14(1):3660.
35. Garrido Marques A, Rubinacci S, Malaspinas AS, Delaneau O, Sousa da Mota B. Assessing the impact of post-mortem damage and contamination on imputation performance in ancient DNA. *Sci Rep*. 2024 Mar 14;14(1):6227.
36. Gamba C, Jones ER, Teasdale MD, McLaughlin RL, Gonzalez-Fortes G, Mattiangeli V, et al. Genome flux and stasis in a five millennium transect of European prehistory. *Nat Commun*. 2014 Oct 21;5(1):5257.
37. Jones ER, Gonzalez-Fortes G, Connell S, Siska V, Eriksson A, Martiniano R, et al. Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat Commun*. 2015 Nov 16;6(1):8912.
38. Martiniano R, Cassidy LM, Ó'Maoldúin R, McLaughlin R, Silva NM, Manco L, et al. The population genomics of archaeological transition in west Iberia: Investigation of ancient substructure using imputation and haplotype-based methods. *PLOS Genetics*. 2017 Tem;13(7):e1006852.
39. Cassidy LM, Maoldúin RÓ, Kador T, Lynch A, Jones C, Woodman PC, et al. A dynastic elite in monumental Neolithic society. *Nature*. 2020 Jun;582(7812):384–8.
40. Gelabert P, Blazyte A, Chang Y, Fernandes DM, Jeon S, Hong JG, et al. Northeastern Asian and Jomon-related genetic structure in the Three Kingdoms period of Gimhae, Korea. *Curr Biol*. 2022 Aug 8;32(15):3232-3244.e6.
41. Morez A, Britton K, Noble G, Günther T, Götherström A, Rodríguez-Varela R, et al. Imputed genomes and haplotype-based analyses of the Picts of early medieval Scotland reveal fine-scale relatedness between Iron Age, early medieval and the modern people of the UK. *PLoS Genet*. 2023 Apr;19(4):e1010360.
42. Yoon BJ. Hidden Markov Models and their Applications in Biological Sequence Analysis. *Curr Genomics*. 2009 Sep;10(6):402–15.
43. Rubinacci S, Ribeiro DM, Hofmeister RJ, Delaneau O. Efficient phasing and imputation of low-coverage sequencing data using large reference panels. *Nat Genet*. 2021 Jan;53(1):120–6.
44. Marchi N, Winkelbach L, Schulz I, Brami M, Hofmanová Z, Blöcher J, et al. The genomic origins of the world's first farmers. *Cell*. 2022 May 26;185(11):1842-1859.e18.
45. 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 Jan 9;33(1):41-57.e15.
46. Kılınc GM, Omrak A, Özer F, Günther T, Büyükkarakaya AM, Bıçakçı E, et al. The Demographic Development of the First Farmers in Anatolia. *Current Biology*. 2016 Oct;26(19):2659–66.
47. Hofmanová Z, Kreutzer S, Hellenthal G, Sell C, Diekmann Y, Díez-del-Molino D, et al. Early farmers from across Europe directly descended from Neolithic Aegeans. *Proceedings of the National Academy of Sciences*. 2016 Jun 21;113(25):6886–91.

48. Yaka R, Mapelli I, Kaptan D, Doğu A, Chyleński M, Erdal ÖD, et al. Variable kinship patterns in Neolithic Anatolia revealed by ancient genomes. *Current Biology*. 2021 Jun;31(11):2455-2468.e18.
49. Koenig Z, Yohannes MT, Nkambule LL, Goodrich JK, Kim HA, Zhao X, et al. A harmonized public resource of deeply sequenced diverse human genomes [Internet]. *bioRxiv*; 2023 [cited 2024 Jan 14]. p. 2023.01.23.525248. Available from: <https://www.biorxiv.org/content/10.1101/2023.01.23.525248v3>
50. 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 Sep;513(7518):409–13.
51. Günther T, Malmström H, Svensson EM, Omrak A, Sánchez-Quinto F, Kılınç GM, et al. Population genomics of Mesolithic Scandinavia: Investigating early postglacial migration routes and high-latitude adaptation. *PLOS Biology*. 2018 Oct;16(10):e2003703.
52. Valdiosera C, Günther T, Vera-Rodríguez JC, Ureña I, Iriarte E, Rodríguez-Varela R, et al. Four millennia of Iberian biomolecular prehistory illustrate the impact of prehistoric migrations at the far end of Eurasia. *Proceedings of the National Academy of Sciences*. 2018 Mar 27;115(13):3428–33.
53. 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 Jun 29;360(6396):eaar7711.
54. Broushaki F, Thomas MG, Link V, López S, van Dorp L, Kirsanow K, et al. Early Neolithic genomes from the eastern Fertile Crescent. *Science*. 2016 Jul 29;353(6298):499–503.
55. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009 Jul 15;25(14):1754–60.
56. Kircher M, Sawyer S, Meyer M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Research*. 2012 Jan 1;40(1):e3.
57. Jun G, Wing MK, Abecasis GR, Kang HM. An efficient and scalable analysis framework for variant extraction and refinement from population-scale DNA sequence data. *Genome Res*. 2015 Jun;25(6):918–25.
58. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009 Aug 15;25(16):2078–9.
59. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010 Mar 15;26(6):841–2.
60. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. *Gigascience*. 2021 Feb 16;10(2):giab008.
61. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am J Hum Genet*. 2007 Sep;81(3):559–75.
62. Patterson N, Price AL, Reich D. Population Structure and Eigenanalysis. *PLOS Genetics*. 2006 Apr;2(12):e190.

63. 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 Nov 4;8(44):eabo3609.
64. Korneliussen TS, Albrechtsen A, Nielsen R. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics*. 2014 Nov 25;15(1):356.
65. 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 Nov 7;15(1):963.
66. 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 May 28;181(5):1158-1175.e28.
67. Lazaridis I, Mitnik A, Patterson N, Mallick S, Rohland N, Pfrengle S, et al. Genetic origins of the Minoans and Mycenaeans. *Nature*. 2017 Aug;548(7666):214–8.
68. 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 Dec;528(7583):499–503.
69. 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 Aug 26;377(6609):eabm4247.
70. Narasimhan VM, Patterson N, Moorjani P, Rohland N, Bernardos R, Mallick S, et al. The formation of human populations in South and Central Asia. *Science*. 2019 Sep 6;365(6457):eaat7487.
71. Clemente F, Unterländer M, Dolgova O, Amorim CEG, Coroado-Santos F, Neuenschwander S, et al. The genomic history of the Aegean palatial civilizations. *Cell*. 2021 May 13;184(10):2565-2586.e21.

## 8. EKLER

**EK-1** Etik Kurul İzni

**EK-2** Orijinallik Raporu

**EK-3** AKT16 ve BAR25 Örnekleri için Doğruluk Grafikleri

**EK-4** Ek TBA Görselleri

**EK-5** Kayıp Sayıları Tablosu 1

**EK-6** Kayıp Sayıları Tablosu 2

**EK-7** Düşük Derinlikli Örneklerin Bilgi ve Kayıp Sayıları Tablosu

**EK-8** Yeni Üretilen Verinin Bilgi ve Kayıp Sayıları Tablosu

**EK-9** Alel Sıklık Hesaplamaları Tablosu 1

**EK-10** Alel Sıklık Hesaplamaları Tablosu 2

## EK-1 Etik Kurul İzni



T.C.  
HACETTEPE ÜNİVERSİTESİ REKTÖRLÜĞÜ  
Sağlık Bilimleri Araştırma Etik Kurulu



Sayı : E-16969557-050.01.04-00003125090  
Konu : Kurul Kararı (SBA 23/213)

10/10/2023

## Sayın Dr. Öğr. Üyesi Gülşah Merve KILINÇ

Kurulumuza değerlendirilmek üzere sunduğunuz SBA 23/213 kayıt numaralı *“Genotip İmpütasyonunun Antik Anadolu İnsan Genomlarında Değerlendirilmesi”* başlıklı proje önerisi Kurulumuzun 03.10.2023 tarihli toplantısında değerlendirilmiştir. İnsan katılımcı, biyolojik materyal veya sağlık verisi kullanılmayan bu tip çalışmalar Etik Kurulların kapsamı dışında kalmaktadır.

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

Bilgilerinizi rica ederim.

Prof. Dr. Nüket Paksoy ERBAYDAR  
Kurul Başkanı

**EK** :  
1. Toplantı Katılım Tutanağı

Bu belge güvenli elektronik imza ile imzalanmıştır.

Belge Doğrulama Kodu: 70069C87-B189-4AAD-86A3-166D6A85D5E1

Belge Doğrulama Adresi: <https://www.turkiye.gov.tr/hu-ebys>

Adres: Hacettepe Üniversitesi Etik Kurulları ve Komisyonları 06100 Sıhhiye/Ankara

Bilgi için: Burak CİHAN

E-posta: [sbaek@hacettepe.edu.tr](mailto:sbaek@hacettepe.edu.tr) Elektronik Ağ: [www.hacettepe.edu.tr](http://www.hacettepe.edu.tr)

Bilgisayar İşletmeni

Telefon: 0 (312) 305 1082 Faks: 0 (312) 310 0580

Telefon: 03123051082

Kep:





## EK-2 Orijinallik Raporu

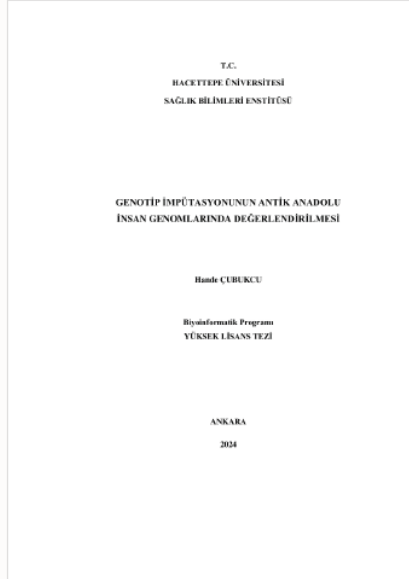


### Dijital Makbuz

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

Gönderinizin ilk sayfası aşağıda gönderilmektedir.

Gönderen: Hande Cubukcu  
Ödev başlığı: Tez Metni Final  
Gönderi Başlığı: GENOTİP İMPÜTASYONUNUN ANTİK ANADOLU İNSAN GENO...  
Dosya adı: HandeCubukcu\_Tez2024\_t2.docx  
Dosya boyutu: 5.32M  
Sayfa sayısı: 44  
Kelime sayısı: 7,834  
Karakter sayısı: 54,256  
Gönderim Tarihi: 07-Haz-2024 02:11ÖS (UTC+0300)  
Gönderim Numarası: 2397554739



## GENOTİP İMPÜTASYONUNUN ANTİK ANADOLU İNSAN GENOMLARINDA DEĞERLENDİRİLMESİ

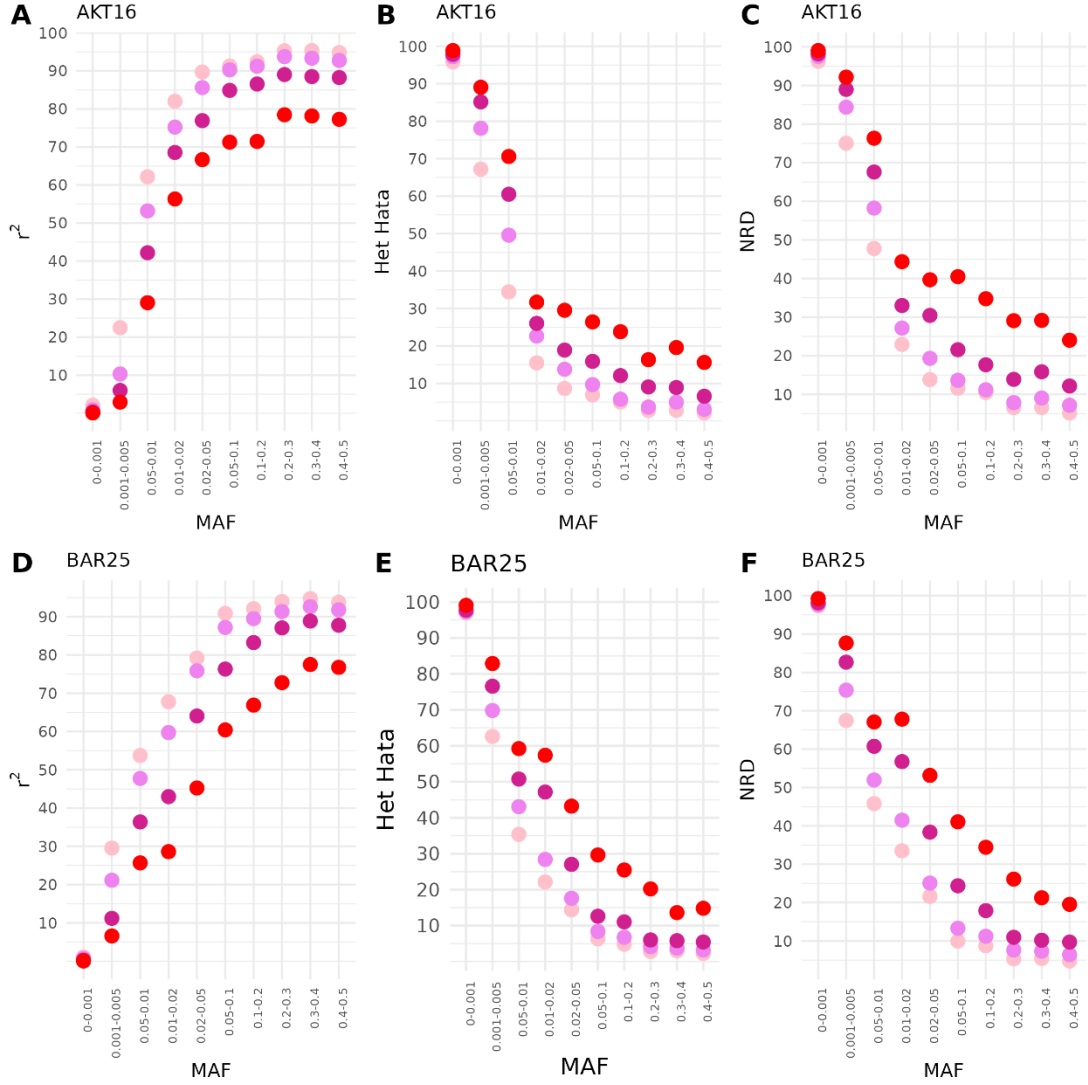
### ORJİNALLİK RAPORU

% <b>1</b>	% <b>1</b>	% <b>1</b>	% <b>0</b>
BENZERLİK ENDEKSİ	İNTERNET KAYNAKLARI	YAYINLAR	ÖĞRENCİ ÖDEVLERİ

### BİRİNCİL KAYNAKLAR

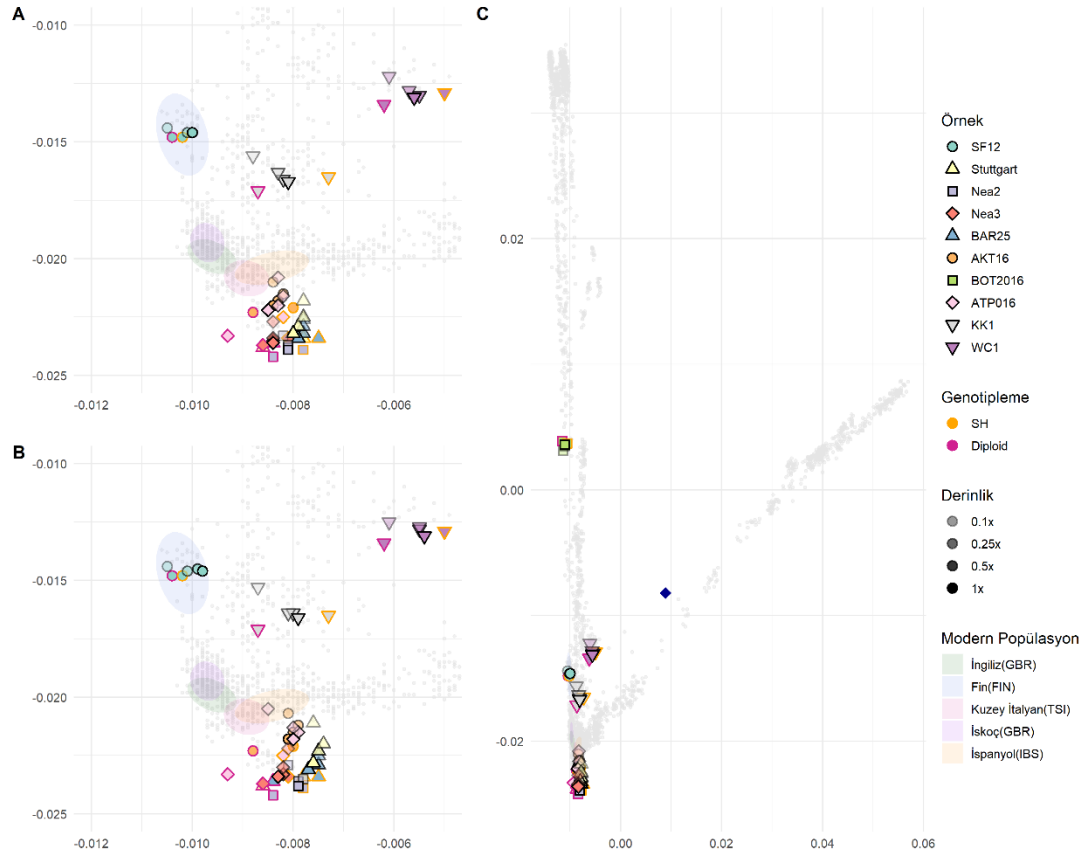
<b>1</b>	<a href="https://journals.plos.org">journals.plos.org</a> İnternet Kaynağı	<% <b>1</b>
<b>2</b>	<a href="https://mouseion.jax.org">mouseion.jax.org</a> İnternet Kaynağı	<% <b>1</b>
<b>3</b>	Kongyang Zhu, Haifeng He, Le Tao, Hao Ma, Xiaomin Yang, Rui Wang, Jianxin Guo, Chuan-Chao Wang. "Protocol for a comprehensive pipeline to study ancient human genomes", STAR Protocols, 2024 Yayın	<% <b>1</b>
<b>4</b>	<a href="http://www.mdpi.com">www.mdpi.com</a> İnternet Kaynağı	<% <b>1</b>
<b>5</b>	<a href="https://mendeliandisorder.blogspot.com">mendeliandisorder.blogspot.com</a> İnternet Kaynağı	<% <b>1</b>
<b>6</b>	<a href="http://www.nature.com">www.nature.com</a> İnternet Kaynağı	<% <b>1</b>
<b>7</b>	<a href="https://openaccess.hacettepe.edu.tr">openaccess.hacettepe.edu.tr</a> İnternet Kaynağı	<% <b>1</b>

### EK-3 AKT16 ve BAR25 Örnekleri için Doğruluk Grafikleri



**Şekil.** Genotip uyumluluk, heterozigot hata ve NRD değerleri gösterilmiştir. Pembe, koyu pembe, mor ve kırmızı sırayla 1x, 0.5x, 0.25x, 0.1x derinliklerini ifade etmektedir. KGP paneli ile impütasyon yapılmış örneklere *concordance* hesaplaması NFE tag kullanılarak KGP ve HGDP+KGP ortak pozisyonlarında yapılmıştır.

### EK-4 Ek TBA Görselleri



**Şekil.** TBA grafiklerinin A) GP80, B) GP99’lu veri için daha fazla yakınlaştırılmış hâlleri ve C) GP80’li veri için tam TBA görseli. Referans genom C) panelinde lacivert elmas ile işaretlenmiştir.

### EK-5 Kayıp Sayıları Tablosu 1

**Tablo:** Yüksek kaliteli örneklerin her bir derinlikte GP80 filtreli impütasyon verisinde Human Origins verisetindeki yaygın varyantlarda (MAF>0.5) varyant istatistikleri.

Örnek	n RefHom	nNon RefHom	nHets	n Kayıp	nTrans itions	nTrans versions	nSingle tons	Toplam	Kayıp Oranı
WC1-0.1x-gp80	189040	65968	91988	83654	128480	29476	160	430650	0.194250551
WC1-0.25x-gp80	202972	76361	113918	37399	154805	35474	77	430650	0.086843144
WC1-0.5x-gp80	208440	80543	121635	20032	164556	37622	73	430650	0.046515732
WC1-1x-gp80	211128	82692	125268	11562	169428	38533	140	430650	0.026847788
AKT16-0.1x-gp80	195976	65956	96917	71801	132525	30351	141	430650	0.166727041
BAR25-0.1x-gp80	195763	67194	100108	67585	135872	31431	143	430650	0.156937188
Nea2-0.1x-gp80	200032	70065	102896	57657	140480	32481	109	430650	0.133883664
Nea3-0.1x-gp80	199050	69586	102989	59025	140344	32233	115	430650	0.137060258
AKT16-0.5x-gp80	212026	77436	121123	20065	161607	36954	61	430650	0.04659236
BAR25-0.5x-gp80	210789	77165	123623	19073	163294	37495	53	430650	0.044288866
Nea2-0.5x-gp80	212826	79062	125485	13277	166372	38178	34	430650	0.03083014
Nea3-0.5x-gp80	212759	78879	125525	13487	166225	38183	41	430650	0.031317775
AKT16-0.25x-gp80	207582	74253	114973	33842	153888	35339	75	430650	0.078583537
BAR25-0.25x-gp80	206713	74461	117971	31505	156395	36038	65	430650	0.073156856
Nea2-0.25x-gp80	209485	76614	120136	24415	159958	36793	48	430650	0.05669337
Nea3-0.25x-gp80	209027	76483	120034	25106	159751	36769	51	430650	0.058297922
AKT16-1x-gp80	214189	79240	124095	13126	165611	37727	107	430650	0.030479508
BAR25-1x-gp80	213385	78695	126836	11734	167160	38374	133	430650	0.027247184
Nea2-1x-gp80	214389	80106	127758	8397	169089	38777	60	430650	0.019498433
Nea3-1x-gp80	214475	80008	128006	8161	169207	38813	89	430650	0.018950424
atp016-0.1x-gp80	188503	61983	92907	87257	125851	29039	225	430650	0.202616974
BOT2016-0.1x-gp80	193588	67621	86248	83193	125128	28743	127	430650	0.193180077
KK1-0.1x-gp80	191673	63663	80532	94782	117211	26985	110	430650	0.220090561
LBK-0.1x-gp80	193464	66283	98073	72830	133605	30751	182	430650	0.169116452
sf12-0.1x-gp80	202442	72379	88105	67724	130576	29908	87	430650	0.157259956
atp016-0.5x-gp80	206083	73995	121570	29002	159054	36511	110	430650	0.067344711
BOT2016-0.5x-gp80	211987	82406	116306	19951	161740	36973	51	430650	0.046327644
KK1-0.5x-gp80	213818	81564	113565	21703	158751	36380	41	430650	0.050395913
LBK-0.5x-gp80	208589	76909	122074	23078	161953	37032	82	430650	0.053588761
sf12-0.5x-gp80	217320	83499	112241	17590	159211	36529	47	430650	0.040845234
atp016-0.25x-gp80	201000	70264	113527	45859	149420	34371	153	430650	0.106487867
BOT2016-0.25x-gp80	207009	78376	108474	36791	152017	34834	61	430650	0.085431325
KK1-0.25x-gp80	208058	76393	104514	41685	147165	33745	64	430650	0.096795542
LBK-0.25x-gp80	204172	73904	115638	36936	154273	35269	87	430650	0.085768025
sf12-0.25x-gp80	213331	80771	106341	30207	152225	34888	51	430650	0.070142807
atp016-1x-gp80	209054	76313	125967	19316	164532	37751	122	430650	0.044853129
BOT2016-1x-gp80	214681	84700	120203	11066	166741	38163	186	430650	0.025696041
KK1-1x-gp80	216856	84321	117549	11924	164270	37604	145	430650	0.027688378
LBK-1x-gp80	211243	78852	125666	14889	166489	38029	135	430650	0.034573319
sf12-1x-gp80	219584	85317	115137	10612	163035	37420	95	430650	0.024641821

## EK-6 Kayıp Sayıları Tablosu 2

**Tablo:** Yüksek kaliteli örneklerin her bir derinlikte GP99 filtreli impütasyon verisinde Human Origins verisetindeki yaygın varyantlarda (MAF>0.5) varyant istatistikleri.

Örnek	nRef Hom	nNon RefHom	nHets	n Kayıp	nTrans itions	nTrans versions	nSingle tons	Toplam	Kayıp Oranı
WC1-0.1x-gp99	116585	34872	47765	231428	67195	15442	0	430650	0.537392314
WC1-0.25x-gp99	164508	57106	83281	125755	114125	26262	0	430650	0.292012075
WC1-0.5x-gp99	186226	68704	102370	73350	139171	31903	0	430650	0.170323929
WC1-1x-gp99	198152	75571	114542	42385	154714	35399	0	430650	0.098420992
AKT16-0.1x-gp99	129667	37863	55799	207321	76084	17578	0	430650	0.481414141
BAR25-0.1x-gp99	134115	40238	58840	197457	80243	18836	0	430650	0.45850923
Nea2-0.1x-gp99	143832	44710	63798	178310	87949	20559	0	430650	0.414048531
Nea3-0.1x-gp99	141764	43922	63895	181069	87444	20375	0	430650	0.420455126
AKT16-0.5x-gp99	190017	65760	102745	72128	136967	31540	0	430650	0.167486358
BAR25-0.5x-gp99	190194	66289	106405	67762	140221	32473	0	430650	0.157348195
Nea2-0.5x-gp99	197547	70862	111843	50398	148385	34321	0	430650	0.117027749
Nea3-0.5x-gp99	196477	70502	111434	52237	147795	34145	0	430650	0.121298038
AKT16-0.25x-gp99	172286	56763	87012	114589	116804	26972	0	430650	0.266083827
BAR25-0.25x-gp99	173195	57969	90724	108762	120577	28116	0	430650	0.252553117
Nea2-0.25x-gp99	182627	62940	96966	88117	129736	30171	0	430650	0.204613956
Nea3-0.25x-gp99	181578	62413	96230	90429	128868	29777	0	430650	0.209982584
AKT16-1x-gp99	200063	71225	113338	46024	150134	34431	0	430650	0.106871009
BAR25-1x-gp99	199861	71495	116612	42682	152811	35297	0	430650	0.099110647
Nea2-1x-gp99	205001	75143	120457	30049	159010	36592	0	430650	0.06977592
Nea3-1x-gp99	204478	74870	120218	31084	158552	36540	0	430650	0.072179264
atp016-0.1x-gp99	115346	31992	47168	236144	64140	15020	0	430650	0.548343202
BOT2016-0.1x-gp99	119350	35507	44490	231303	64814	15183	0	430650	0.537102055
KK1-0.1x-gp99	110378	29994	39132	251146	55928	13198	0	430650	0.583178916
LBK-0.1x-gp99	128376	38395	55648	208231	76309	17734	0	430650	0.483527226
sf12-0.1x-gp99	136186	42295	49738	202431	74751	17282	0	430650	0.470059213
atp016-0.5x-gp99	176388	58789	97082	98391	126728	29143	0	430650	0.228470916
BOT2016-0.5x-gp99	190385	70779	97767	71719	137104	31443	0	430650	0.166536631
KK1-0.5x-gp99	190356	68218	94010	78066	131879	30350	0	430650	0.181274817
LBK-0.5x-gp99	183609	64042	101587	81412	134689	30940	0	430650	0.189044468
sf12-0.5x-gp99	197660	72754	95879	64357	137083	31550	0	430650	0.149441542
atp016-0.25x-gp99	156852	49492	78652	145654	104009	24135	0	430650	0.338218971
BOT2016-0.25x-gp99	168636	58786	78184	125044	111193	25777	0	430650	0.290361082
KK1-0.25x-gp99	164971	54808	73407	137464	104169	24047	0	430650	0.319201207
LBK-0.25x-gp99	166194	55473	85210	123773	114286	26397	0	430650	0.287409729
sf12-0.25x-gp99	180230	63038	80315	107067	116522	26831	0	430650	0.248617207
atp016-1x-gp99	188625	65536	110154	66335	142809	32882	0	430650	0.154034599
BOT2016-1x-gp99	202268	77755	109679	40948	152439	34996	0	430650	0.095084175
KK1-1x-gp99	203791	76381	106516	43962	148678	34221	0	430650	0.102082898
LBK-1x-gp99	194304	69655	112882	53809	148474	34063	0	430650	0.124948334
sf12-1x-gp99	207336	78265	105913	39136	149710	34468	0	430650	0.090876582

### EK-7 Düşük Derinlikli Örneklerin Bilgi ve Kayıp Sayıları Tablosu

**Tablo:** Düşük genomboyu derinliğe sahip yayınlanmış örnekler (N=108) için örnek bilgileri ile Human Origins veriseti pozisyonlarından sözde haploid (SH) veride toplam 591957 pozisyon için, filtreli impütasyon verilerinde (GP80, GP90, GP99) toplam 449840 pozisyon için (MAF > 0.5) eksik TNP oranları (E.O).

Örnek	Ülke	Veri Tipi	Genomboyu Derinlik	Derinlik Grubu	Kaynak	GP80 E.O	GP90 E.O	GP99 E.O	SH E.O
AH1	İran	SG	1.03817	Above_1X	(54)	0.0237	0.0392	0.0968	0.4029
AH2	İran	SG	0.572561	0.5X-1X	(54)	0.0406	0.0666	0.1579	0.5814
AH4	İran	SG	0.774265	0.5X-1X	(54)	0.0313	0.0517	0.1253	0.5026
ALA013	Türkiye	1240K	0.272733	0.25X-0.5X	(66)	0.0120	0.0199	0.0505	0.1753
ALA026	Türkiye	1240K	0.20314	0.1X-0.25X	(66)	0.0151	0.0248	0.0628	0.1848
ART009	Türkiye	1240K	0.385299	0.25X-0.5X	(66)	0.0072	0.0118	0.0311	0.1072
ART015	Türkiye	1240K	0.240067	0.1X-0.25X	(66)	0.0112	0.0189	0.0489	0.1617
ART042	Türkiye	1240K	0.236839	0.1X-0.25X	(66)	0.0134	0.0222	0.0561	0.1491
Ash128	Türkiye	SG	4.79937	Above_1X	(48)	0.0080	0.0131	0.0313	0.0415
Ash129	Türkiye	SG	0.744252	0.5X-1X	(48)	0.0439	0.0710	0.1602	0.5940
Bar31	Türkiye	SG	3.22912	Above_1X	(47)	0.0117	0.0193	0.0475	0.1570
Bar8	Türkiye	SG	6.01755	Above_1X	(47)	0.0041	0.0068	0.0171	0.0458
Bon001	Türkiye	WISP+SG	0.1555	0.1X-0.25X	(46)	0.1853	0.2731	0.4914	0.9503
Bon002	Türkiye	WISP+SG	6.3169	Above_1X	(46)	0.0076	0.0126	0.0314	0.1143
Bon004	Türkiye	WISP+SG	0.227853	0.1X-0.25X	(46)	0.1302	0.1979	0.3864	0.9250
CCH294	Türkiye	SG	0.257745	0.25X-0.5X	(48)	0.0871	0.1371	0.2959	0.8012
cth739	Türkiye	SG	0.189873	0.1X-0.25X	(48)	0.1924	0.2849	0.5040	0.8931
cth747	Türkiye	SG	0.115017	0.1X-0.25X	(48)	0.2268	0.3321	0.5742	0.9286
F38	İran	SG	1.69658	Above_1X	(54)	0.0097	0.0163	0.0406	0.1955
I0071	Yunanistan	1240K	0.459461	0.25X-0.5X	(67)	0.0056	0.0091	0.0229	0.1049
I0707	Türkiye	1240K	0.469355	0.25X-0.5X	(68)	0.0045	0.0075	0.0196	0.0871
I0708	Türkiye	1240K	0.352023	0.25X-0.5X	(68)	0.0073	0.0122	0.0313	0.1059
I0709	Türkiye	1240K	0.477074	0.25X-0.5X	(68)	0.0049	0.0082	0.0211	0.0940
I0745	Türkiye	1240K	0.448817	0.25X-0.5X	(68)	0.0052	0.0085	0.0221	0.0921
I0746	Türkiye	1240K	0.459065	0.25X-0.5X	(68)	0.0047	0.0078	0.0201	0.0878
I10430	Türkiye	1240K	0.204044	0.1X-0.25X	(69)	0.0223	0.0352	0.0825	0.3157
I13518	Yunanistan	1240K	0.230754	0.1X-0.25X	(69)	0.0236	0.0368	0.0851	0.1780
I14820	Türkiye	1240K	0.249244	0.1X-0.25X	(69)	0.0217	0.0343	0.0796	0.2539
I14822	Türkiye	1240K	0.211143	0.1X-0.25X	(69)	0.0243	0.0382	0.0881	0.2648
I14823	Türkiye	1240K	0.248867	0.1X-0.25X	(69)	0.0215	0.0337	0.0771	0.2480
I14824	Türkiye	1240K	0.203218	0.1X-0.25X	(69)	0.0274	0.0428	0.0973	0.2714
I14831	Türkiye	1240K	0.215714	0.1X-0.25X	(69)	0.0227	0.0361	0.0835	0.2599
I14832	Türkiye	1240K	0.232116	0.1X-0.25X	(69)	0.0224	0.0352	0.0806	0.2531
I14839	Türkiye	1240K	0.207197	0.1X-0.25X	(69)	0.0242	0.0383	0.0865	0.2410
I1579	Türkiye	1240K	0.313284	0.25X-0.5X	(69)	0.0142	0.0216	0.0474	0.1140
I1580	Türkiye	1240K	0.205574	0.1X-0.25X	(68)	0.0148	0.0247	0.0639	0.2015
I1583	Türkiye	1240K	0.408621	0.25X-0.5X	(68)	0.0068	0.0113	0.0294	0.1414
I1584	Türkiye	1240K	0.251631	0.25X-0.5X	(69)	0.0191	0.0298	0.0696	0.1942
I1585	Türkiye	1240K	0.379656	0.25X-0.5X	(68)	0.0130	0.0199	0.0432	0.1046
I1947	İran	1240K	0.153435	0.1X-0.25X	(70)	0.0167	0.0274	0.0705	0.1920
I1954	İran	1240K	0.185066	0.1X-0.25X	(70)	0.0143	0.0242	0.0602	0.1822
I20000	Türkiye	1240K	0.233312	0.1X-0.25X	(69)	0.0313	0.0497	0.1118	0.2451
I20145	Türkiye	1240K	0.227434	0.1X-0.25X	(69)	0.0289	0.0455	0.1027	0.2662
I20146	Türkiye	1240K	0.373684	0.25X-0.5X	(69)	0.0204	0.0314	0.0716	0.1973
I20187	Türkiye	1240K	0.224807	0.1X-0.25X	(69)	0.0279	0.0440	0.0992	0.2463
I20229	Türkiye	1240K	0.240746	0.1X-0.25X	(69)	0.0265	0.0414	0.0931	0.2380
I20259	Türkiye	1240K	0.269086	0.25X-0.5X	(69)	0.0261	0.0408	0.0914	0.2350
I20260	Türkiye	1240K	0.262683	0.25X-0.5X	(69)	0.0257	0.0410	0.0919	0.2382
I20264	Türkiye	1240K	0.231489	0.1X-0.25X	(69)	0.0282	0.0444	0.1014	0.2592

I20266	Türkiye	1240K	0.203836	0.1X-0.25X	(69)	0.0327	0.0515	0.1161	0.2835
I20323	Türkiye	1240K	0.278924	0.25X-0.5X	(69)	0.0252	0.0396	0.0894	0.2403
I20324	Türkiye	1240K	0.272082	0.25X-0.5X	(69)	0.0266	0.0417	0.0942	0.2335
I20325	Türkiye	1240K	0.217641	0.1X-0.25X	(69)	0.0306	0.0481	0.1085	0.2728
I20326	Türkiye	1240K	0.267031	0.25X-0.5X	(69)	0.0251	0.0395	0.0884	0.2463
I20328	Türkiye	1240K	0.240601	0.1X-0.25X	(69)	0.0288	0.0459	0.1029	0.2580
I20329	Türkiye	1240K	0.262083	0.25X-0.5X	(69)	0.0274	0.0429	0.0965	0.2363
I20571	Türkiye	1240K	0.288793	0.25X-0.5X	(69)	0.0242	0.0378	0.0862	0.2285
I20572	Türkiye	1240K	0.242772	0.1X-0.25X	(69)	0.0254	0.0402	0.0924	0.2361
I20573	Türkiye	1240K	0.264027	0.25X-0.5X	(69)	0.0249	0.0390	0.0887	0.2400
I20574	Türkiye	1240K	0.286566	0.25X-0.5X	(69)	0.0237	0.0374	0.0837	0.2279
I2495	Türkiye	1240K	0.394119	0.25X-0.5X	(69)	0.0152	0.0229	0.0503	0.1476
I2683	Türkiye	1240K	0.322247	0.25X-0.5X	(69)	0.0192	0.0299	0.0679	0.1953
I3911	İran	1240K	0.291792	0.25X-0.5X	(69)	0.0245	0.0383	0.0853	0.1785
I3912	İran	1240K	0.346935	0.25X-0.5X	(69)	0.0185	0.0292	0.0666	0.1687
I3913	İran	1240K	0.342205	0.25X-0.5X	(69)	0.0192	0.0295	0.0669	0.1698
I3914	İran	1240K	0.272479	0.25X-0.5X	(69)	0.0232	0.0369	0.0832	0.1750
I3915	İran	1240K	0.386296	0.25X-0.5X	(69)	0.0179	0.0279	0.0635	0.1630
I3920	Yunanistan	1240K	0.140594	0.1X-0.25X	(68)	0.0178	0.0295	0.0759	0.1810
I4029	Türkiye	1240K	0.308767	0.25X-0.5X	(69)	0.0155	0.0240	0.0547	0.1251
I4097	İran	1240K	0.473988	0.25X-0.5X	(69)	0.0168	0.0261	0.0569	0.1449
I4100	İran	1240K	0.560061	0.5X-1X	(69)	0.0133	0.0204	0.0450	0.1211
I4232	İran	1240K	0.295279	0.25X-0.5X	(69)	0.0222	0.0351	0.0800	0.1938
I4233	İran	1240K	0.343369	0.25X-0.5X	(69)	0.0195	0.0306	0.0694	0.1745
I4236	İran	1240K	0.221315	0.1X-0.25X	(69)	0.0368	0.0574	0.1289	0.2874
I4238	İran	1240K	0.219734	0.1X-0.25X	(69)	0.0265	0.0419	0.0963	0.2183
I4239	İran	1240K	0.288928	0.25X-0.5X	(69)	0.0223	0.0347	0.0799	0.2014
I4269	İran	1240K	0.356482	0.25X-0.5X	(69)	0.0170	0.0261	0.0586	0.1568
I4272	İran	1240K	0.260934	0.25X-0.5X	(69)	0.0247	0.0388	0.0887	0.2023
I4273	İran	1240K	0.208623	0.1X-0.25X	(69)	0.0347	0.0545	0.1212	0.2463
I4280	İran	1240K	0.236555	0.1X-0.25X	(69)	0.0252	0.0390	0.0890	0.2137
I4349	İran	1240K	0.262163	0.25X-0.5X	(70)	0.0194	0.0320	0.0794	0.2290
I4532	Türkiye	1240K	0.227203	0.1X-0.25X	(69)	0.0237	0.0373	0.0853	0.2477
I4534	Türkiye	1240K	0.206343	0.1X-0.25X	(69)	0.0246	0.0391	0.0898	0.2543
I6429	İran	1240K	0.211933	0.1X-0.25X	(69)	0.0221	0.0342	0.0784	0.2033
I8369	Türkiye	1240K	0.233347	0.1X-0.25X	(69)	0.0218	0.0343	0.0791	0.1868
I8372	Türkiye	1240K	0.217982	0.1X-0.25X	(69)	0.0228	0.0358	0.0812	0.2060
Klei10	Yunanistan	SG	1.83224	Above_1X	(47)	0.0087	0.0142	0.0363	0.1838
Kou01	Yunanistan	SG	2.2002	Above_1X	(71)	0.0089	0.0146	0.0371	0.2003
Kou03	Yunanistan	SG	2.45156	Above_1X	(71)	0.0082	0.0134	0.0334	0.1397
Log02	Yunanistan	SG	3.89828	Above_1X	(71)	0.0034	0.0055	0.0142	0.0557
Log04	Yunanistan	SG	4.362	Above_1X	(71)	0.0032	0.0052	0.0132	0.0426
MA2195	Türkiye	SG	0.791846	0.5X-1X	(53)	0.0287	0.0469	0.1122	0.5358
MA2196	Türkiye	SG	0.244202	0.1X-0.25X	(53)	0.1085	0.1678	0.3422	0.8191
MA2198	Türkiye	SG	0.694206	0.5X-1X	(53)	0.0423	0.0686	0.1574	0.5644
MA2200	Türkiye	SG	1.96354	Above_1X	(53)	0.0122	0.0201	0.0496	0.2210
MA2203	Türkiye	SG	0.845682	0.5X-1X	(53)	0.0397	0.0644	0.1479	0.5087
MA2205	Türkiye	SG	0.704461	0.5X-1X	(53)	0.0318	0.0519	0.1253	0.5477
MA2206	Türkiye	SG	0.408031	0.25X-0.5X	(53)	0.0632	0.1007	0.2252	0.7253
MA2210	Türkiye	SG	0.809377	0.5X-1X	(53)	0.0289	0.0471	0.1141	0.5082
MA2212	Türkiye	SG	0.780556	0.5X-1X	(53)	0.0358	0.0581	0.1365	0.5480
MA2213	Türkiye	SG	1.07314	Above_1X	(53)	0.0250	0.0411	0.0994	0.4183
Mik15	Yunanistan	SG	2.76871	Above_1X	(71)	0.0104	0.0170	0.0416	0.1728
Pal7	Yunanistan	SG	1.16989	Above_1X	(47)	0.0192	0.0313	0.0790	0.3625
Pta08	Yunanistan	SG	3.56335	Above_1X	(71)	0.0047	0.0078	0.0197	0.0685
Rev5	Yunanistan	SG	0.972988	0.5X-1X	(47)	0.0278	0.0455	0.1118	0.5367
Tep002	Türkiye	WISP+SG	0.678236	0.5X-1X	(46)	0.0719	0.1125	0.2356	0.7810
Tep003	Türkiye	WISP+SG	0.655736	0.5X-1X	(46)	0.0378	0.0619	0.1424	0.5962
Tep004	Türkiye	WISP+SG	0.447504	0.25X-0.5X	(46)	0.0765	0.1192	0.2512	0.8210



### EK-8 Yeni Üretilen Verinin Bilgi ve Kayıp Sayıları Tablosu

**Tablo:** Yeni üretilen genom örnekleri için impütasyon çıktısı ve sözde-haploid veri için Human Origins verisindeki eksik oranı değerleri ile örneklerin tanımlayıcı bilgileri. Oranlar impütasyon verilerinde yaygın varyantlar, 430650 pozisyon için, sözde haploid veride ise 591957 pozisyon içindir. (Geç Kalk.: Geç Kalkolitik.)

Örnek	GP80 Eksik Oranı	GP99 Eksik Oranı	Sözde Haploid Eksik Oranı	Derinlik	Derinlik Grubu	Dönem
cay004	0.306864	0.70815	0.94718	0.072725	<0.1x	Neolitik
cay012	0.447215	0.858342	0.969817	0.038932	<0.1x	Neolitik
cay013	0.194606	0.531922	0.899599	0.138491	0.1x-1x	Neolitik
cay014	0.442517	0.851397	0.971065	0.040106	<0.1x	Neolitik
cay034	0.024572	0.097638	0.391561	1.20372	>1x	Neolitik
cch146	0.220416	0.581605	0.910955	0.108099	0.1x-1x	Neolitik
cch151	0.127428	0.398251	0.870908	0.172258	0.1x-1x	Neolitik
cch152	0.201481	0.542101	0.90678	0.126706	0.1x-1x	Neolitik
cch163	0.155314	0.453844	0.883108	0.1706	0.1x-1x	Neolitik
cch170	0.076728	0.26313	0.768002	0.348318	0.1x-1x	Neolitik
cch226	0.219868	0.571757	0.925079	0.112031	0.1x-1x	Neolitik
cch242	0.160469	0.463209	0.889642	0.156732	0.1x-1x	Neolitik
cch251	0.0231	0.094457	0.429077	1.07179	>1x	Neolitik
cch255	0.198342	0.539294	0.912688	0.115971	0.1x-1x	Neolitik
cch290	0.192746	0.53373	0.907468	0.116671	0.1x-1x	Neolitik
cch294	0.006049	0.024498	0.042674	5.28131	>1x	Neolitik
cch312	0.073305	0.255813	0.726514	0.379924	0.1x-1x	Neolitik
cch318	0.17905	0.502355	0.898496	0.134756	0.1x-1x	Neolitik
cch328	0.04212	0.164015	0.64928	0.508437	0.1x-1x	Neolitik
cch510	0.073877	0.251295	0.68958	0.481299	0.1x-1x	Neolitik
iks268	0.377297	0.791304	0.962104	0.046204	<0.1x	Geç Kalk.
iks339	0.531613	0.916011	0.983671	0.024663	<0.1x	Geç Kalk.
iks340	0.409154	0.821897	0.96607	0.041069	<0.1x	Geç Kalk.
iks350	0.408861	0.818793	0.966954	0.042689	<0.1x	Geç Kalk.
iks371	0.322961	0.727463	0.954566	0.065357	<0.1x	Geç Kalk.
iks427	0.255934	0.637418	0.938903	0.085535	<0.1x	Geç Kalk.
iks433	0.269781	0.658029	0.938649	0.078947	<0.1x	Geç Kalk.
iks545	0.305051	0.705775	0.946006	0.064489	<0.1x	Geç Kalk.
iks550	0.399633	0.815254	0.966276	0.041154	<0.1x	Geç Kalk.
iks567	0.291169	0.687243	0.947074	0.072186	<0.1x	Geç Kalk.
iks573	0.344215	0.753698	0.955581	0.053521	<0.1x	Geç Kalk.
iks599	0.196484	0.533069	0.904258	0.114951	0.1x-1x	Geç Kalk.
iks602	0.205132	0.549544	0.912622	0.126154	0.1x-1x	Geç Kalk.
iks652	0.343859	0.75136	0.956289	0.056508	<0.1x	Geç Kalk.

### EK-9 Alel Sıklık Hesaplamaları Tablosu 1

**Tablo:** 41 düşük derinlikli Anadolu Neolitik örneğini kullanarak Human Origins verisetinden rastgele seçilen 100 yaygın TNP (MAF>0.5) için ANGSD ile BAM dosyalarından ve PLINK ile filtreli (GP80/GP99) impütasyon çıktısı VCF dosyasından hesaplanan minör alel frekansları. GP80/GP99Diff sütunları ANGSD knownEM – GP80/GP99MAF sütunu farkını göstermektedir.

chromo: position	major	minor	ANGSD knownEM	nInd	GP80 MAF	nInd GP80	GP80 Diff	GP99 MAF	nInd GP99	GP99 Diff
chr1:117624255	C	T	0.351905	16	0.4103	39	-0.058395	0.4219	32	-0.069995
chr1:165989494	A	G	0.453870	23	0.4872	39	-0.03333	0.4444	36	0.00947
chr1:212409534	A	G	0.049832	24	0.1282	39	-0.078368	0.1029	34	-0.053068
chr1:224448673	C	T	0.000014	21	0	41	0.000014	0	41	0.000014
chr1:240309961	C	T	0.094679	20	0.06098	41	0.033699	0.05128	39	0.043399
chr10:35673959	G	A	0.353894	22	0.3472	36	0.006694	0.3281	32	0.025794
chr10:57685679	G	T	0.386797	20	0.3077	39	0.079097	0.3158	38	0.070997
chr10:76348302	C	T	0.089011	20	0.075	40	0.014011	0.06757	37	0.021441
chr10:87697314	C	A	0.317086	19	0.3281	32	-0.011014	0.3448	29	-0.027714
chr10:96658173	G	A	0.269251	19	0.2576	33	0.011651	0.2115	26	0.057751
chr11:2696707	T	C	0.076154	25	0.09459	37	-0.018436	0.08824	34	-0.012086
chr11:61439838	G	T	0.174914	21	0.1707	41	0.004214	0.1795	39	-0.004586
chr12:101910172	G	A	0.047918	15	0.07317	41	-0.025252	0.05405	37	-0.006132
chr12:119727350	A	G	0.417579	23	0.4	40	0.017579	0.4167	36	0.000879
chr12:19866718	C	T	0.211243	15	0.2	30	0.011243	0.2037	27	0.007543
chr12:21180106	A	G	0.264636	18	0.35	40	-0.085364	0.3462	39	-0.081564
chr12:59024255	T	A	0.000000	13	0	41	0	0	37	0
chr13:107006753	T	C	0.489151	20	0.5	39	-0.010849	0.4531	32	0.036051
chr13:22085614	C	T	0.093289	21	0.04054	37	0.052749	0.04688	32	0.046409
chr13:99437410	C	A	0.121160	22	0.1389	36	-0.01774	0.1613	31	-0.04014
chr14:34224346	G	A	0.183673	19	0.07812	32	0.105553	0.1	25	0.083673
chr15:53900295	A	G	0.464749	19	0.4857	35	-0.020951	0.4828	29	-0.018051
chr15:67940161	A	C	0.472109	16	0.4625	40	0.009609	0.4394	33	0.032709
chr15:71247983	C	A	0.360320	23	0.3333	36	0.02702	0.3448	29	0.01552
chr16:19677710	T	C	0.273964	21	0.2568	37	0.017164	0.2576	33	0.016364
chr16:20239656	C	T	0.033497	18	0	41	0.033497	0	40	0.033497
chr16:331590	C	A	0.258318	25	0.1923	39	0.066018	0.1875	32	0.070818
chr16:5950036	G	A	0.405703	18	0.4318	22	-0.026097	0.425	20	-0.019297
chr16:65796358	T	C	0.073374	25	0.1026	39	-0.029226	0.08824	34	-0.014866
chr17:1792679	T	C	0.198104	22	0.1944	36	0.003704	0.1935	31	0.004604
chr17:38835723	G	A	0.443187	23	0.45	30	-0.006813	0.5	23	-0.056813
chr17:6823352	T	A	0.000001	19	0	41	0.000001	0	39	0.000001
chr17:8153288	G	T	0.172585	21	0.1316	38	0.040985	0.09091	33	0.081675
chr19:53689298	A	G	0.095479	21	0.05263	38	0.042849	0.04286	35	0.052619
chr2:118762771	T	G	0.073609	17	0.0625	40	0.011109	0.05263	38	0.020979
chr2:118994874	A	C	0.000000	19	0	40	0	0	37	0
chr2:124668185	C	T	0.236005	15	0.2564	39	-0.020395	0.2429	35	-0.006895
chr2:141116841	A	C	0.000001	15	0.02564	39	-0.025639	0.01429	35	-0.014289
chr2:142881000	T	G	0.035937	19	0.1375	40	-0.101563	0.1316	38	-0.095663
chr2:170870967	C	T	0.029809	21	0	38	0.029809	0	35	0.029809
chr2:179290178	T	C	0.291426	18	0.2105	38	0.080926	0.2059	34	0.085526
chr2:200469681	C	A	0.459762	21	0.4306	36	0.029162	0.4821	28	-0.022338
chr2:214389813	G	T	0.442410	23	0.4265	34	0.01591	0.3929	28	0.04951
chr2:220042375	G	A	0.000001	22	0	40	0.000001	0	38	0.000001
chr2:224013852	T	C	0.460771	14	0.3125	32	0.148271	0.3103	29	0.150471
chr2:2620749	G	A	0.373372	19	0.3684	38	0.004972	0.3676	34	0.005772
chr2:67792210	T	C	0.061452	20	0.0375	40	0.023952	0.04412	34	0.017332
chr2:68827978	G	A	0.042855	27	0.05	40	-0.007145	0.04054	37	0.002315

chr2:80582941	G	A	0.361069	22	0.359	39	0.002069	0.3235	34	0.037569
chr20:48686904	T	C	0.192938	19	0.1875	40	0.005438	0.1579	38	0.035038
chr20:8512003	G	A	0.116907	17	0.1538	26	-0.036893	0.1429	21	-0.025993
chr21:17872211	T	C	0.209313	19	0.2703	37	-0.060987	0.2903	31	-0.080987
chr21:24423272	A	G	0.075461	15	0.03947	38	0.035991	0.04545	33	0.030011
chr21:33788837	A	G	0.316904	18	0.3281	32	-0.011196	0.3333	27	-0.016396
chr21:36935865	C	T	0.351638	21	0.4744	39	-0.122762	0.4595	37	-0.107862
chr21:40177830	G	A	0.191034	22	0.197	33	-0.005966	0.2222	27	-0.031166
chr21:40652225	T	C	0.064452	22	0.09091	33	-0.026458	0.1	25	-0.035548
chr22:28908265	C	T	0.276766	20	0.2564	39	0.020366	0.2432	37	0.033566
chr22:42752253	T	C	0.253146	22	0.3816	38	-0.128454	0.375	28	-0.121854
chr22:44954783	A	G	0.207158	19	0.1818	33	0.025358	0.1935	31	0.013658
chr3:141940399	T	C	0.275678	19	0.3514	37	-0.075722	0.3788	33	-0.103122
chr3:14197243	T	G	0.420975	20	0.4079	38	0.013075	0.4265	34	-0.005525
chr3:162367159	C	T	0.169768	23	0.141	39	0.028768	0.141	39	0.028768
chr3:187036934	G	A	0.247071	15	0.1282	39	0.118871	0.125	32	0.122071
chr3:24546089	G	A	0.092364	16	0.07692	39	0.015444	0.06579	38	0.026574
chr3:29560650	A	G	0.078696	16	0.09756	41	-0.018864	0.09211	38	-0.013414
chr3:37527286	G	A	0.168530	24	0.025	40	0.14353	0.02564	39	0.14289
chr4:111983310	A	G	0.454865	22	0.4872	39	-0.032335	0.4865	37	-0.031635
chr4:181828911	C	A	0.000002	18	0	41	0.000002	0	39	0.000002
chr4:24997677	T	G	0.186411	16	0.1571	35	0.029311	0.1667	30	0.019711
chr4:74703727	G	A	0.000002	23	0	37	0.000002	0	34	0.000002
chr5:174224906	C	T	0.374556	26	0.3676	34	0.006956	0.3333	27	0.041256
chr5:29061609	G	A	0.000001	19	0.03659	41	-0.036589	0.02703	37	-0.027029
chr5:30790970	C	T	0.270024	24	0.1585	41	0.111524	0.1757	37	0.094324
chr5:41092707	T	C	0.493187	21	0.4756	41	0.017587	0.4744	39	0.018787
chr5:55143638	A	G	0.278002	20	0.2195	41	0.058502	0.2195	41	0.058502
chr6:103829385	C	T	0.078373	23	0	39	0.078373	0	37	0.078373
chr6:12578627	G	A	0.262983	20	0.2222	36	0.040783	0.25	28	0.012983
chr6:146849198	G	A	0.092181	17	0.15	40	-0.057819	0.15	40	-0.057819
chr6:150380974	T	C	0.220594	22	0.3378	37	-0.117206	0.3676	34	-0.147006
chr6:162590411	G	A	0.042634	17	0.122	41	-0.079366	0.122	41	-0.079366
chr6:87950289	G	A	0.076645	9	0.1719	32	-0.095255	0.1875	24	-0.110855
chr6:97384573	G	A	0.107101	24	0.09756	41	0.009541	0.09756	41	0.009541
chr7:112562249	C	T	0.293426	18	0.3333	39	-0.039874	0.3594	32	-0.065974
chr7:126484999	T	C	0.312890	21	0.4359	39	-0.12301	0.4211	38	-0.10821
chr7:15085115	G	A	0.000001	22	0.0125	40	-0.012499	0.01282	39	-0.012819
chr7:27177740	A	G	0.415284	23	0.4643	28	-0.049016	0.4773	22	-0.062016
chr7:27351334	G	A	0.411681	26	0.4189	37	-0.007219	0.4062	32	0.005481
chr7:45065547	A	C	0.000001	24	0	36	0.000001	0	32	0.000001
chr7:46083137	T	G	0.056717	21	0.08537	41	-0.028653	0.08537	41	-0.028653
chr7:54828368	C	T	0.123933	27	0.0875	40	0.036433	0.08571	35	0.038223
chr7:78419309	C	T	0.056084	22	0.04054	37	0.015544	0.05556	27	0.000524
chr8:127642338	A	C	0.000001	19	0	39	0.000001	0	31	0.000001
chr8:142655246	G	A	0.436010	11	0.4359	39	0.00011	0.4324	37	0.00361
chr8:143381516	T	C	0.362558	16	0.4706	34	-0.108042	0.4808	26	-0.118242
chr8:69052379	A	G	0.072819	21	0.07692	39	-0.004101	0.07895	38	-0.006131
chr8:75299478	T	C	0.498018	27	0.4737	38	0.024318	0.4853	34	0.012718
chr9:103959523	C	T	0.109939	18	0.0641	39	0.045839	0.0641	39	0.045839
chr9:109091406	G	A	0.068377	20	0	41	0.068377	0	41	0.068377
chr9:92021318	T	C	0.279005	20	0.25	40	0.029005	0.2564	39	0.022605

## EK-10 Alel Sıklık Hesaplamaları Tablosu 2

**Tablo:** 41 düşük derinlikli Anadolu Neolitik örneğini kullanarak Human Origins verisetinden rastgele seçilen 100 az yaygın TNP ( $0.1 < \text{MAF} < 0.5$ ) için ANGSD ile BAM dosyalarından ve PLINK ile filtreli (GP80/GP99) impütasyon çıktısı VCF dosyasından hesaplanan minör alel sıklıkları. GP80/GP99Diff sütunları ANGSD knownEM – GP80/GP99MAF sütunu farkını göstermektedir.

chromo:position	major	minor	ANGSD knownEM	nInd	GP80 MAF	nInd GP80	GP80 Diff	GP99 MAF	nInd GP99	GP99 Diff
chr1:100115466	G	A	0.034397	24	0	41	0.034397	0	41	3.4397
chr1:105158250	C	A	0.000002	16	0	41	0.000002	0	38	0.0002
chr1:166295632	A	C	0.000002	27	0	41	0.000002	0	41	0.0002
chr1:167181594	G	A	0.036710	22	0.0122	41	0.02451	0.0125	40	2.421
chr1:177735612	C	T	0.029902	20	0	40	0.029902	0	39	2.9902
chr1:182633903	G	A	0.103113	23	0	41	0.103113	0	41	10.3113
chr1:184329259	G	A	0.000002	19	0	41	0.000002	0	41	0.0002
chr1:200726929	C	A	0.000002	20	0	41	0.000002	0	41	0.0002
chr1:215677475	C	T	0.059809	21	0.01351	37	0.046299	0.01613	31	4.3679
chr1:216631013	C	T	0.061151	27	0	41	0.061151	0	41	6.1151
chr1:235109113	G	A	0.084661	23	0	41	0.084661	0	40	8.4661
chr1:245454043	C	T	0.070152	20	0	41	0.070152	0	40	7.0152
chr1:24586219	G	A	0.175383	21	0.1447	38	0.030683	0.1364	33	3.8983
chr1:36365674	C	T	0.053760	25	0	40	0.05376	0	40	5.376
chr1:43060010	C	T	0.051495	28	0	41	0.051495	0	41	5.1495
chr1:80693187	C	T	0.036146	18	0	41	0.036146	0	41	3.6146
chr1:80744440	G	A	0.030741	21	0.06579	38	-0.03505	0.0625	32	-3.1759
chr10:124616936	A	C	0.000001	17	0	41	0.000001	0	40	0.0001
chr10:128477089	G	A	0.000002	22	0.01282	39	-0.01282	0.01429	35	-1.4288
chr10:60281641	A	C	0.000001	20	0	41	0.000001	0	41	0.0001
chr10:61774108	G	A	0.033943	26	0	39	0.033943	0	37	3.3943
chr10:83375655	A	C	0.000000	17	0	41	0	0	41	0
chr10:89068259	G	A	0.049545	24	0	41	0.049545	0	40	4.9545
chr11:120167777	G	A	0.000003	26	0	41	0.000003	0	40	0.0003
chr11:122855645	C	T	0.000002	18	0	41	0.000002	0	40	0.0002
chr11:4937638	T	A	0.000002	20	0.02439	41	-0.02439	0.025	40	-2.4998
chr11:78542486	T	C	0.032086	18	0	41	0.032086	0	41	3.2086
chr12:123747306	G	A	0.000002	18	0	41	0.000002	0	40	0.0002
chr12:135461	G	A	0.037527	17	0	39	0.037527	0	38	3.7527
chr12:41215639	C	A	0.000003	19	0	41	0.000003	0	41	0.0003
chr12:5135691	G	A	0.028000	21	0.0122	41	0.0158	0.0122	41	1.58
chr12:7789438	G	A	0.000001	22	0	40	0.000001	0	40	0.0001
chr13:29562098	G	A	0.093795	18	0.08537	41	0.008425	0.075	40	1.8795
chr13:69731535	C	T	0.033599	19	0.04878	41	-0.01518	0.04878	41	-1.5181
chr13:86089795	T	A	0.000001	19	0	41	0.000001	0	39	0.0001
chr14:33440204	A	C	0.000001	19	0	40	0.000001	0	38	0.0001
chr14:88904335	A	G	0.028673	21	0.03947	38	-0.0108	0.04054	37	-1.1867
chr15:27426017	C	T	0.048038	23	0	41	0.048038	0	41	4.8038
chr15:38043761	G	A	0.028766	22	0	41	0.028766	0	37	2.8766
chr16:50640120	G	T	0.000002	19	0	40	0.000002	0	39	0.0002
chr16:80814841	T	A	0.000000	18	0	41	0	0	41	0
chr17:28500820	C	A	0.000000	16	0	41	0	0	38	0
chr17:34078814	T	A	0.000000	16	0.02439	41	-0.02439	0	39	0
chr17:43959928	G	A	0.000005	15	0	41	0.000005	0	41	0.0005
chr17:6125087	A	C	0.000001	21	0.0125	40	-0.0125	0.01316	38	-1.3159
chr18:55235854	G	A	0.064066	23	0.02632	38	0.037746	0.01562	32	4.8446
chr18:7441313	C	T	0.126349	17	0.08333	36	0.043019	0.07576	33	5.0589
chr19:1374642	G	A	0.132283	19	0.05405	37	0.078233	0.07143	28	6.0853

chr19:40384933	C	T	0.029544	20	0	39	0.029544	0	37	2.9544
chr2:134325983	A	C	0.000002	18	0	41	0.000002	0	38	0.0002
chr2:136393703	A	C	0.000000	18	0.02439	41	-0.02439	0.02564	39	-2.564
chr2:15692597	C	T	0.033313	18	0	41	0.033313	0	40	3.3313
chr2:204932787	A	G	0.029704	20	0.0122	41	0.017504	0.01351	37	1.6194
chr2:209259937	A	C	0.000002	21	0	41	0.000002	0	41	0.0002
chr2:209852772	T	A	0.000003	20	0	41	0.000003	0	41	0.0003
chr2:22347083	C	T	0.035649	18	0.0122	41	0.023449	0.01282	39	2.2829
chr2:234931295	A	C	0.000001	18	0	41	0.000001	0	40	0.0001
chr2:27571227	G	A	0.033966	19	0	41	0.033966	0	41	3.3966
chr2:41307713	A	C	0.000001	22	0	41	0.000001	0	40	0.0001
chr2:42373546	T	A	0.000003	23	0	41	0.000003	0	41	0.0003
chr2:51283629	G	A	0.100375	21	0.0122	41	0.088175	0.0125	40	8.7875
chr2:57385797	A	C	0.000002	23	0	41	0.000002	0	41	0.0002
chr20:37189672	G	A	0.000002	19	0	41	0.000002	0	41	0.0002
chr20:774304	C	T	0.123930	21	0.01562	32	0.10831	0	24	12.393
chr22:35648305	G	T	0.026793	25	0.025	40	0.001793	0.02703	37	-0.0237
chr3:114491741	C	T	0.029415	22	0	40	0.029415	0	38	2.9415
chr3:116154400	A	C	0.000002	27	0	41	0.000002	0	38	0.0002
chr3:169353462	A	C	0.000001	23	0	41	0.000001	0	41	0.0001
chr3:17299426	G	A	0.040256	16	0	41	0.040256	0	41	4.0256
chr4:120342770	C	T	0.104814	14	0.06098	41	0.043834	0.0625	40	4.2314
chr4:1318467	G	A	0.041656	17	0	41	0.041656	0	40	4.1656
chr4:149121568	T	A	0.000003	11	0	41	0.000003	0	41	0.0003
chr4:155302847	T	G	0.031726	19	0.07317	41	-0.04144	0.07317	41	-4.1444
chr4:166258049	C	T	0.031050	19	0	41	0.03105	0	41	3.105
chr4:167803529	G	A	0.081816	24	0	41	0.081816	0	41	8.1816
chr4:21641817	C	T	0.000002	17	0	41	0.000002	0	37	0.0002
chr4:48809840	A	C	0.000002	19	0	41	0.000002	0	39	0.0002
chr4:95219000	C	T	0.140535	21	0.0625	40	0.078035	0.0625	40	7.8035
chr5:40887924	G	A	0.000002	16	0	41	0.000002	0	41	0.0002
chr5:51587581	A	C	0.000002	20	0	41	0.000002	0	40	0.0002
chr6:10475514	G	A	0.029949	20	0	41	0.029949	0	41	2.9949
chr6:124929024	A	C	0.000001	19	0	40	0.000001	0	38	0.0001
chr6:138155299	A	C	0.000001	16	0	41	0.000001	0	41	0.0001
chr6:152820939	A	C	0.000001	15	0	41	0.000001	0	40	0.0001
chr6:24735857	C	T	0.080424	21	0	41	0.080424	0	40	8.0424
chr6:35565388	A	C	0.000005	26	0	41	0.000005	0	41	0.0005
chr6:75060325	A	C	0.000000	17	0	41	0	0	40	0
chr6:79105453	G	A	0.000001	15	0	40	0.000001	0	40	0.0001
chr6:98383640	A	C	0.000001	19	0	40	0.000001	0	40	0.0001
chr7:102048380	T	A	0.000000	17	0	41	0	0	41	0
chr7:114988210	G	A	0.051592	23	0	40	0.051592	0	40	5.1592
chr7:21949846	T	A	0.000000	17	0	41	0	0	41	0
chr7:37667307	G	A	0.024462	23	0	41	0.024462	0	40	2.4462
chr8:120266711	A	C	0.000001	21	0	41	0.000001	0	41	0.0001
chr8:28532099	C	T	0.034973	17	0	41	0.034973	0	40	3.4973
chr8:42728028	A	G	0.095006	21	0.09756	41	-0.00255	0.09756	41	-0.2554
chr8:61329249	A	C	0.000002	23	0	41	0.000002	0	41	0.0002
chr8:95216427	C	T	0.044569	18	0	41	0.044569	0	39	4.4569
chr9:73438062	T	A	0.000001	22	0	41	0.000001	0	41	0.0001
chr9:942952	A	C	0.000001	19	0	41	0.000001	0	41	0.0001

## 9. ÖZGEÇMİŞ