



**T.C.  
İSTANBUL ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ**



**Yüksek Lisans Tezi**

**BÜYÜK ÖLÇEKLİ PRENATAL TEST GENOMİK VERİLERİNDEN  
TÜRK POPÜLASYONUNUN GENETİK PROFİLİNİN ORTAYA  
KONULMASI**

**Selim Can KURALAY**

**Biyoloji Anabilim Dalı**

**Zooloji Programı**

**DANIŞMAN  
Dr. Öğr. Üyesi Vahap ELDEM**

**II. DANIŞMAN  
Doç. Dr. Gökmen ZARARSIZ**

**Haziran, 2023**

**İSTANBUL**

Bu çalışma, 6.06.2023 tarihinde ařağıdaki jüri tarafından Biyoloji Anabilim Dalı, Zooloji Programında Yüksek Lisans tezi olarak kabul edilmiştir.

### **Tez Jürisi**

Dr. Öğr. Üyesi Vahap ELDEM(Danışman)  
İstanbul Üniversitesi  
Fen Fakültesi

Prof. Dr. Uğur AKSU  
İstanbul Üniversitesi  
Fen Fakültesi

Dr. Öğr. Üyesi Yavuz TURAN  
Marmara Üniversitesi  
Fen Fakültesi

## **İntihal Programı Beyanı**

20.04.2016 tarihli Resmi Gazete’de yayımlanan Lisansüstü Eğitim ve Öğretim Yönetmeliğinin 9/2 ve 22/2 maddeleri gereğince; Bu Lisansüstü teze, İstanbul Üniversitesi’nin aboneli olduğu intihal yazılım programı kullanılarak Fen Bilimleri Enstitüsü’nün belirlemiş olduğu ölçütlere uygun rapor alınmıştır.

## **Proje Destekleri**

Bu tez, İstanbul Üniversitesi Bilimsel Araştırma Projeleri Yürütücü Sekreterliğinin 39205 numaralı projesi ile desteklenmiştir.



## ÖNSÖZ

Girişimsel olmayan prenatal test (NIPT) amacıyla fetal anöploidilerin tespit edilmesi amacıyla üretilen genomik verinin popülasyon genetiği çalışmalarında kullanılması mümkündür. 13,044 bireye ait NIPT verileri kullanılarak yürütülen bu çalışmada, Türkiye popülasyonunun varyant profili incelenmiştir.

Yüksek lisans eğitimim süresince mentorluğuyla ve hiçbir zaman esirgemediği desteğiyle yanımda olan danışmanım Dr. Öğr. Üyesi Vahap ELDEM'e,

Her zaman yanımda olan İstanbul Üniversitesi Moleküler Biyoçeşitlilik Laboratuvarı üyelerine,

2210-A Yurt İçi Genel Yüksek Lisans Burs Programı kapsamında bursiyer olmam vesilesiyle maddi destek sağlayan TÜBİTAK'a,

Maddi ve manevi mümkün olan her türlü fedakarlığı gösteren çekirdek aileme sonsuz teşekkürlerimi sunarım.

Haziran 2023

Selim Can KURALAY

# İÇİNDEKİLER

Sayfa No

ÖNSÖZ .....	ii
İÇİNDEKİLER.....	iii
ŞEKİL LİSTESİ .....	v
TABLO LİSTESİ.....	vi
SİMGE VE KISALTMA LİSTESİ.....	vii
ÖZET .....	viii
SUMMARY .....	x
<b>1. GİRİŞ.....</b>	<b>1</b>
<b>2. GENEL KISIMLAR.....</b>	<b>3</b>
2.1. GİRİŞİMSEL OLMAYAN PRENATAL TEST (NIPT).....	3
2.2. NIPT VERİSİNİN POPÜLASYON GENETİĞİ ÇALIŞMALARINDA KULLANIMI .....	3
2.3. GENETİK VARYASYON VERİ TABANLARI .....	5
2.3.1. DBSNP.....	5
2.3.2. CLINVAR .....	5
2.3.3. GNOMAD.....	5
2.3.4. 1000 GENOMES – 1000 GENOM PROJESİ.....	6
2.4. VARYANTLARIN SINIFLANDIRILMASI .....	6
2.5. TÜRK POPÜLASYONUNA ÖZGÜ GENETİK VARYANT ÇALIŞMALARI.....	6
<b>3. MALZEME VE YÖNTEM.....</b>	<b>8</b>
3.1. ETİK DEĞERLENDİRMELER VE ETİK KURUL ONAYI.....	8
3.2. BİYOİNFORMATİK ANALİZLER.....	8
3.2.1. Genom Okumalarının Kalite Kontrolü .....	8
3.2.2. Genom Okumalarının Ön Filtreden Geçirilmesi .....	8
3.2.2. Genom Okumalarının İnsan Referans Genomuna Hizalanması.....	9
3.2.3. Hizalanan Genom Okumalarının Referans Genomu Kapsama Oranının Belirlenmesi .....	9
3.2.4. Varyantların Çağırılması .....	9
3.2.5. Varyantların Filtrelenmesi.....	10
3.2.6. Varyantların Karakterize Edilmesi .....	10
<b>4. BULGULAR.....</b>	<b>11</b>

4.1. OKUMA İSTATİSTİKLERİ .....	11
4.2. DNA OKUMALARININ İNSAN GENOMUNA HİZALANMASI .....	12
4.3. VARYANTLARIN ÇAĞIRILMASI VE FİLTRELENMESİ.....	13
4.4. VARYANTLARIN KARAKTERİZE EDİLMESİ .....	13
<b>5. TARTIŞMA VE SONUÇ .....</b>	<b>16</b>
<b>KAYNAKLAR.....</b>	<b>18</b>
<b>EKLER .....</b>	<b>21</b>
<b>ÖZGEÇMİŞ .....</b>	<b>24</b>



## ŞEKİL LİSTESİ

### Sayfa No

Şekil 4.1: Ham (A) ve filtrelenmiş (B) okumalara ait PHRED skoru dağılımı grafikleri .....12

Şekil 4.2: Tespit edilen varyantların genomik bölgelere göre dağılımı. ....13



## TABLO LİSTESİ

### Sayfa No

**Tablo 4.1:** 13044 bireye ait DNA okumalarına ilişkin istatistikler. .... 11

**Tablo 4.2:** Tespit edilen patojenik varyantlara ait bilgiler..... 14



## SİMGE VE KISALTMA LİSTESİ

<b>Kısaltmalar</b>	<b>Açıklama</b>
<b>Nipt</b>	: Non-invasive prenatal testing
<b>SnP</b>	: Single nucleotide polymorphism
<b>Ncbi</b>	: National Center for Biotechnology Information
<b>ts</b>	: Transisyon
<b>tv</b>	: Transversiyon
<b>PCR</b>	: Polymerase chain reaction
<b>Sam</b>	: Sequence alignment map
<b>Bam</b>	: Binary alignment map
<b>G.vcf</b>	: Genomic variant calling format
<b>Vcf</b>	: Variant calling format
<b>cffDNA</b>	: Cell free fetal DNA
<b>WHO</b>	: World Health Organisation

## ÖZET

### YÜKSEK LİSANS TEZİ

#### BÜYÜK ÖLÇEKLİ PRENATAL TEST GENOMİK VERİLERİNDEN TÜRK POPÜLASYONUNUN GENETİK PROFİLİNİN ORTAYA KONULMASI

**Selim Can KURALAY**

**İstanbul Üniversitesi**

**Fen Bilimleri Enstitüsü**

**Biyoloji Anabilim Dalı**

**Danışman : Dr. Öğr. Üyesi Vahap ELDEM**

**II. Danışman : Doç. Dr. Gökmen ZARARSIZ**

Bir popülasyonun genetik profilinin açığa çıkarılması o popülasyona ait genetik hastalıkların anlaşılabilmesi, tedavi edilebilmesi ve önleyici tıp stratejilerinin geliştirilebilmesi açısından önemli bir dönüm noktasıdır. Bu amaca yönelik yürütülebilecek büyük ölçekli genom sekanslama projelerinin gerektirdiği teknolojik altyapı ve maliyetler, özellikle Türkiye gibi gelişmekte olan ülkelerde bu projelerin hayata geçirilebilmeleri karşısında sınırlayıcı faktörler olarak karşımıza çıkmaktadırlar. Girişimsel olmayan prenatal test (NIPT) amacıyla üretilen düşük derinlikli tüm genom dizileme verilerinin farklı bir amaç için yeniden kullanımı ek bir maliyet gerektirmemektedir. Bahsi geçen dezavantajlar NIPT verileri için söz konusu olmadığından, popülasyon profilini yansıtabilecek yeterli sayıda örnek elde edilebildiği takdirde bu verilerin popülasyon genetiği çalışmalarında kullanılması alternatif bir çözüm olarak karşımıza çıkmaktadır. Türkiye popülasyonu özelinde günümüze kadar gerçekleştirilen en büyük ölçekli çalışma olan bu tez kapsamında, 13044 bireye ait NIPT kaynaklı tüm genom dizileme verisi biyoinformatik yöntemler kullanılarak analiz edilmiştir. Her bir bireye ait varyant profilleri elde edilmiş ve ortak bir veri tabanında depolanmıştır. Elde edilen 1,718,799 adet tek nükleotid polimorfizmlerinin (SNP'ler) insan genomu üzerinde hangi konumlarda gerçekleştiklerinin, potansiyel klinik etkilerinin ve Türk popülasyonunda görülme sıklıklarının tespit edilmesi hedeflenmiştir. Bu çerçevede Türk popülasyonunun genetik profilinin ortaya konulması ve Türk popülasyonu özelinde literatürde karşılaşılan bilgi eksikliğinin giderilmesine hizmet etmek amaçlanmıştır.

Haziran 2023, 37 sayfa.

**Anahtar kelimeler:** Genetik hastalıklar, non-invaziv prenatal test (NIPT), genom sekanslama, popülasyon genetiđi, tek nükleotid polimorfizmleri (SNP)



## **SUMMARY**

### **M.Sc. THESIS**

#### **GENETIC PROFILING OF TURKISH POPULATION FROM LARGE- SCALE PRENATAL TESTING DATA GENOMIC DATA**

**Selim Can KURALAY**

**İstanbul University**

**Institute of Graduate Studies in Sciences**

**Department of Biology**

**Supervisor : Assist. Prof. Dr. Vahap ELDEM**

**Co-Supervisor : Assoc. Prof. Dr. Gökmen ZARARSIZ**

Revealing the genetic profile of a population is a crucial milestone for understanding, treating and developing preventive medicine strategies for genetic diseases specific to that population. However, the technological infrastructure and costs required for large-scale genome sequencing projects that can achieve this goal can be limiting factors, especially in developing countries like Turkey. Reusing low-depth whole-genome sequencing data produced for non-invasive prenatal testing (NIPT) for a different purpose does not require additional costs. Thus, if a sufficient number of samples that can reflect the population profile can be obtained, using NIPT data in population genetics studies can be an alternative solution where the aforementioned limitations do not apply. In the scope of this thesis, which is the largest-scale study conducted to date on the Turkish population, the whole-genome sequencing data obtained from 13,044 individuals through NIPT were analyzed using bioinformatics methods. Variant profiles were obtained for each individual and stored in a common database. The objectives of this study are to ascertain the locations of 1,718,799 single nucleotide polymorphisms (SNPs) within the human genome, evaluate their possible clinical implications, and assess their prevalence among the Turkish population. The aim in this context is to uncover the genetic characteristics of the Turkish population and address the existing gaps in knowledge about the Turkish population found in literature.

June 2023, 37 pages.

**Keywords:** Genetic diseases, non-invasive prenatal testing (NIPT), genome sequencing, population genetics, single nucleotide polymorphisms (SNPs)



## 1. GİRİŞ

İnsana özgü genetik varyasyonları ortaya çıkarmak ve tanımlamak, kalıtsal hastalıkların altında yatan sebepleri anlamamıza önemli ölçüde yardımcı olmaktadır. Bu amaca yönelik dünya çapında gerçekleştirilen birçok çalışma ve veri tabanı olmasına karşın, bu çalışmalarda ve veri tabanlarında incelenen bireyler çoğunlukla Avrupa kökenlidir ve sunulan veri sınırlı genetik çeşitliliğe sahiptir. Bu eğilim, Avrupa kökenli olmayan popülasyonlara özgü genetik varyantlar ve bu varyantlarla ilişkili hastalık riskleri ile ilgili yanlış çıkarımlara yol açabilmektedir. Veri tabanlarında az sayıda bireyle temsil edilen popülasyonların sahip olduğu genetik çeşitliliği daha iyi anlamak ve gelecekte yapılacak varyant analizleri için bir referans noktası üretebilmek için, bu popülasyonları temsil eden bir veri setinin oluşturulması ihtiyacı doğmaktadır. Genom çapında ilişkilendirme çalışmaları (İng: Genome wide association studies, GWAS) dahilinde gerçekleştirilen varyant analizleri bu amaca hizmet edebilmektedir. Buna karşın, GWAS analizlerini gerçekleştirmek için standart prosedür olan tüm genom dizileme (İng: Whole genome sequencing, WGS) ve tüm ekzom dizileme (İng: Whole exome sequencing, WES) tekniklerinin gerektirdiği teknolojik ve maddi yük, popülasyon düzeyinde veri elde edebilmek için ulaşılmaması gereken birey sayısı da dikkate alındığında, gelişmekte olan ülkeler için oldukça yüksek boyutlara ulaşmaktadır ve araştırmacıları alternatif yöntemleri değerlendirmelerine sebep olmaktadır [1]. Maternal dolaşımda hücre dışı fetal DNA'nın keşfi (İng: cffDNA), maternal plazma kullanılarak fetal genetik bozuklukların tespit edilebilmesine olanak sağlamıştır [1,2]. Gebelikte müdahale gerektirmeyen bir genetik tanı testi olan NIPT (İng: Non-invasive prenatal testing), anne adaylarının kanındaki hücre dışı DNA'nın ultra düşük derinlikte (0.1-0.2x) dizilenmesi ve analiz edilmesi ile cenine ait kromozomal anöploidi bilgisinin elde edilmesi amacıyla kullanılmaktadır. Hamile kadınların dolaşım sisteminde mevcut olan cenine ait hücre dışı DNA (İng: Cell-free fetal DNA, cffDNA) dolaşımdaki bolluk düzeyinin yüksekliğiyle ve kararlı yapısıyla hamileliğe özgü testlere elverişli olmasıyla birlikte [3], NIPT verisinin popülasyon genetiği çalışmalarında yeniden kullanılması da yakın geçmişte tartışılmakta ve olumlu sonuçlar alınmaktadır [4-6]. DNA dizileme teknolojilerinde yaşanan gelişmeler ve dizileme maliyetlerinin azalması NIPT'in dünya çapında popülerliğini artırarak üretilen veri miktarının gün geçtikçe artmasına [7], dolaylı olarak bu verinin popülasyon genetiği çalışmalarında yeniden kullanımının da artış göstermesine ön ayak olmuştur.

Günümüze dek küresel ölçekte 20 milyonun üzerinde NIPT testinin gerçekleştirildiği ve her yıl yaklaşık 5 milyon anne adayının NIPT testi hizmeti aldığı tespit edilmiştir [8, 9].

Bu tez kapsamında Genoks Genetik Hastalıklar Değerlendirme Merkezi'nde 2020-2022 yılları arasında üretilen 13,044 bireye ait NIPT verisi incelenmiştir. NIPT kaynaklı genomik okumalar insan referans genomuna hizalanmıştır ve bu veri kullanılarak popülasyon ölçeğinde varyant seti elde edilmiştir.



## 2. GENEL KISIMLAR

### 2.1. GİRİŞİMSEL OLMAYAN PRENATAL TEST (NIPT)

Anne plazmasında bulunan cffDNA, gebeliğin erken dönemlerinde (5-7 hafta gibi) tespit edilebilir yoğunluğa ulaşmaktadır [4]. Fetal DNA'nın plazmadaki yoğunluğunun geçmiş dönemlerde %3-6 olduğu tespit edilmişse de [11], sonraki çalışmalar plazmada bulunan fetal DNA oranının %10-20 seviyelerine kadar çıkabildiğini göstermiştir [12]. NIPT, fetal anöploidi olasılığını belirlemek için anne adayının kan örneğinde bulunan hücre dışı fetal DNA'nın (cffDNA) ultra düşük derinlikte (0.1-0.2x) dizilenmesini takiben analiz edilmesini kapsayan bir testtir. Bu yöntem kullanılarak trizomi 21 (Down Sendromu), 13 (Patau Sendromu) ve 18 (Edwards Sendromu) gibi yaygın genetik hastalıkların doğum öncesi tespiti mümkün hale gelmektedir [13].

### 2.2. NIPT VERİSİNİN POPÜLASYON GENETİĞİ ÇALIŞMALARINDA KULLANIMI

NIPT verisinin popülasyon genetiği çalışmalarında kullanımı henüz yaygın bir yaklaşım olmamakla birlikte, gerçekleştirilmiş birkaç çalışma literatürde mevcuttur. Bunlardan en geniş kapsamlı olanı Liu ve diğ. [7] tarafından Çin popülasyonu üzerinde gerçekleştirilen “Genomic Analyses from Non-invasive Prenatal Testing Reveal Genetic Associations, Patterns of Viral Infections, and Chinese Population History” başlıklı çalışmadır. Bu çalışmada Çin popülasyonuna ait 141,431 anne adayına ait NIPT amacı ile üretilen düşük derinlikli DNA okumaları kullanılmıştır. Okumalar kalite kontrol ve filtreleme işlemlerinden geçirildikten sonra, insan genomuna hizalanmışlardır ve varyant analizine tabi tutulmuşlardır. Sonuç olarak bulunan 9,04 milyon SNP 1000 Genom Projesi'nde yer alan popülasyonlara ait varyantlarla karşılaştırılmıştır. Bu çalışmada tespit edilen transisyon/transversiyon (ts/tv) oranı 2.2 olmuştur ki bu oran, bu tez kapsamında tespit edilen varyantlara ait ts/tv oranıyla (2.15) oldukça yakındır. Benzer şekilde, 1000 Genom Projesinde [14] yer alan oranlarla da uyumludur.

Bir başka çalışma da Vietnam popülasyonu için Tran ve diğ. [6] tarafından gerçekleştirilen “Genetic profiling of Vietnamese population from large-scale genomic analysis of non-invasive prenatal testing data” başlıklı çalışmadır. Bu çalışmaya konu olan 2683 anne adayı bireye ait varyant profilleri, NIPT analizleri amacıyla üretilen düşük derinlikli DNA okumaları

kullanılarak analiz edilmiştir. Bu çalışmanın Liu ve diğ. tarafından gerçekleştirilen çalışmadan farkı, kullanılan DNA verisinin “pair-end” okuma stratejisi kullanılarak üretilmesidir. Liu ve diğ. [7] tarafından gerçekleştirilen çalışmaya benzer şekilde, genomik okumalar öncelikle kalite kontrol işleminden geçirilmiştir. Kalite kontrol işlemini takiben filtreleme işlemi gerçekleştirilmiş ve okumalar insan genomuna hizalanmıştır. Hizalama verileri kullanılarak gerçekleştirilen varyant çağırma analizi sonucunda 8,054,515 SNP tespit ve karakterize edilmiştir ve Vietnam popülasyonuna ait varyant veri tabanı oluşturulmuştur. Bu çalışmadaki varyant veri setine ait ts/tv oranı 2.0 olarak hesaplanmıştır. Bulunan varyantlar 1000 Genom Projesi veri setine dahil olan Vietnam popülasyonu verileriyle karşılaştırılmış ve bulunan varyantların %91,8’inin daha önce karakterize edildiği saptanmıştır. Daha önce karakterize edilen Vietnam popülasyon verileriyle örtüşmeyen kalan varyantlar ise yine 1000 Genom Projesi [14] veri setine dahil olan Doğu Asya (EAS) setiyle karşılaştırılmış ve varyantların küçük bir bölümünün de bu grupta karakterize edildiği belirlenmiştir (%0,8).

NIPT verisi kullanılarak gerçekleştirilen çalışmalar yalnızca tüm genom varyant analizleri ile sınırlı kalmamaktadır. Morshneva ve diğ. [15] tarafından gerçekleştirilen “Pilot Screening of Cell-Free mtDNA in NIPT: Quality Control, Variant Calling, and Haplogroup Determination” başlıklı çalışmada 645 bireye ait NIPT verilerinin içerdiği mitokondriyal genom verileri kullanılarak varyant analizi, soy araştırmaları, popülasyon çalışmaları gibi analizler yapılabileceği ve NIPT amaçlı üretilen verinin kullanım alanlarının genişletilebileceği tartışılmaktadır.

Ye ve diğ. [16] tarafından gerçekleştirilen “Identification of copy number variants by NGS-based NIPT at low sequencing depth” başlıklı bir başka çalışmada NIPT verisi kopya sayı varyasyonlarını (İng: Copy number variation, CNV) tespit etmek için kullanılmıştır. 873 anne adayına ait NIPT verisi kullanılarak uzunlukları 0.1-47.3 milyon baz arasında değişen 52 kopya sayı varyasyonu tespit edilmiştir.

## **2.3. GENETİK VARYASYON VERİ TABANLARI**

### **2.3.1. DBSNP**

Farklı türlere ait varyasyon bilgilerini içeren dbSNP [17], NCBI tarafından geliştirilmiş halka açık bir veri tabanıdır. SNP bilgileri dışında INDEL, mikrosatelit işaretleyiciler ve heterozigot dizilere ait bilgiler de içeren dbSNP'nin son versiyonu 1 milyonun üzerinde SNP bilgisi içermektedir ve bu miktar araştırmacılar ve dizileme laboratuvarları yeni bilgiler ekledikçe kontrol edilmekte ve güncellenmektedir.

### **2.3.2. CLINVAR**

ClinVar [18], genetik varyantlar ve bunların klinik önemleri hakkında bilgiler içeren halka açık bir veri tabanıdır. Ulusal Biyoteknoloji Bilgi Merkezi (NCBI) tarafından oluşturulan bu veri tabanında, tıp literatürü, klinik testler ve araştırmalarla ilgili olarak bildirilen varyantlar hakkındaki bilgiler depolanmaktadır. Varyantların her birinin hangi hastalık veya durumla ilişkili olduğu ve hastanın sağlığı üzerinde nasıl bir etkisi olabileceği gibi bilgiler, ClinVar veri tabanında yer almaktadır. Araştırmacılar için kritik öneme sahip bu veri tabanı, klinik uzmanlar ve genetik danışmanlar tarafından genetik test sonuçlarını yorumlamakta ve klinik karar verme süreçlerinde aktif olarak kullanılmaktadır. ClinVar, hızla büyüyen ve güncellenen veri tabanıyla, insan hastalıklarının genetik temeli hakkındaki bilgilerimizi artırmada önemli bir role sahiptir.

### **2.3.3. GNOMAD**

gnomAD [19] (Genome Aggregation Database) veri tabanı, farklı popülasyonlar arasındaki genetik çeşitliliği detaylı bir şekilde inceleyen kapsamlı bir insan genetik varyasyon kataloğudur. MIT ve Harvard Üniversitesi Broad Enstitüsü tarafından geliştirilmiş olan bu veri tabanı, 140.000'den fazla insanın dizileme verilerini içerir ve insan genomunda bulunan milyonlarca varyanta ait bilgiler içerir. gnomAD, farklı popülasyonlar arasındaki genetik varyantların dağılımını daha iyi anlamak, nadir ve yaygın genetik varyantları tespit etmek ve insan hastalıklarındaki genetik varyasyonların rolünü anlamak için araştırmacılara kaynak sağlamak amacıyla tasarlanmıştır. Veri tabanı ayrıca, her bir varyant için tahmini protein fonksiyonu etkisi, farklı popülasyonlardaki varyant sıklığı ve hastalıklar ile ilişkisi gibi ayrıntılı

açıklamalar da içerir. gnomAD, genetik, epidemiyolojik, evrimsel biyoloji gibi farklı alanlarda çalışan araştırmacılar için değerli bir araçtır.

#### **2.3.4. 1000 GENOMES – 1000 GENOM PROJESİ**

1000 Genom Projesi [14], insan genetik varyasyonlarının kapsamlı bir kataloğunu oluşturmayı hedefleyen uluslararası bir projedir. Bu proje, dünya çapındaki farklı popülasyonlardan binlerce bireyin genomlarının dizilenmesini içermektedir. Ortaya çıkan veri seti, 1000 Genom veri tabanı olarak bilinir ve insan genetik varyasyonunun en geniş ve çeşitli kataloglarından biridir. 1000 Genom veri tabanı, tüm insan genomu boyunca milyonlarca genetik varyantı içeren bir kaynaktır. Veri tabanı hem yaygın hem de nadir olan varyantları içerir ve farklı popülasyonlardaki genetik varyasyonların dağılımını bünyesinde bulundurur. Bu veri tabanı, araştırmacılar tarafından insan hastalıklarının genetik temelini araştırmak için kullanılan bir kaynak olmakla birlikte, insan popülasyonlarının evrimsel tarihine ilgi duyanlar için de önemli bir kaynaktır. 1000 Genom veri tabanı ücretsiz olarak kullanılabilir ve genetik, epidemiyoloji ve antropoloji gibi çeşitli alanlarda çalışan araştırmacılar için değerli bir araç olarak kendine yer bulmaktadır.

#### **2.4. VARYANTLARIN SINIFLANDIRILMASI**

Varyantların sınıflandırılması allel frekans bilgileri, işlevsel tahminler, türler arası korunma ve hastalık/fenotip ile ilişkili olma gibi birçok parametreyi içermektedir. Varyantlar, genellikle patojenik, muhtemel patojenik, belirsiz önemli, muhtemelen zararsız ve zararsız gibi beş kategoriye ayrılırlar. Sınıflandırma kriterleri, ilgili hastalık veya fenotipe bağlı olarak değişebilir.

#### **2.5. TÜRK POPÜLASYONUNA ÖZGÜ GENETİK VARYANT ÇALIŞMALARI**

Genetik varyasyon veri tabanları bölümünde değinilen veri tabanları yakın zamana kadar Türk popülasyonuna ait bir veri içermemekteydi. Kars ve diğ. [20] tarafından 2021 yılında yayınlanan “The genetic structure of the Turkish population reveals high levels of variation and admixture” başlıklı çalışma ile birlikte Türk variomu (TRV) ilk kez ortaya koyulmuştur. Akrabalık ilişkisi bulunmayan 3,362 birey ile tüm genom ve tüm ekzom okumaları kullanılarak yürütülen bu çalışma sonucunda Türk popülasyonunun tarihsel geçmişi de göz önünde bulundurularak Balkan, Kafkas, Orta Doğu ve Avrupa popülasyonlarıyla yakın genetik

benzerliklere sahip olduđu gözlenmiştir. Çalışma kapsamında tüm genom sekanslarından 45,981,721 ve tüm ekzom sekanslarından 1,123,248 varyant tespit edilmiştir. Bulunan varyantlardan nadir olanlarının (allel frekansı  $< 0.005$ ) tüm genom varyantları için %49'unun, tüm ekzom varyantları için ise %27'inin Türk popülasyonuna özgü olduđu tespit edilmiştir. Daha sonra varyantlar işlevsel durumlarına göre sınıflandırılmıştır. Bu çalışma gelecek genetik varyant çalışmaları için önemli bir kaynak olarak literatürde yer almaktadır.



### 3. MALZEME VE YÖNTEM

#### 3.1. ETİK DEĞERLENDİRMELER VE ETİK KURUL ONAYI

NIPT amacı ile üretilen genomik verileri kullanılan tüm bireyler bilgilendirilmiş gönüllü onam formlarını okumuş, onaylamış ve imzalamışlardır (**Ek 1**). Çalışma, İstanbul Üniversitesi İstanbul Tıp Fakültesi Etik Kurulu onayı ile gerçekleştirilmiştir (**Ek 2**). NIPT verilerinin bu tez kapsamında kullanılması konusunda T.C. Sağlık Bakanlığı'nca ruhsatlı Genoks Genetik Hastalıklar Değerlendirme Merkezi ile genetik veri paylaşım protokolü imzalanmıştır (**Ek 3**).

#### 3.2. BİYOİNFORMATİK ANALİZLER

Bu çalışmada kullanılan hesaplama kaynakları Ulusal Yüksek Başarımli Hesaplama Merkezi'nin (UHeM) desteğiyle sağlanmıştır. Kullanılan bilgisayarların donanım bilgileri aşağıdaki gibidir:

İşletim Sistemi: Rocky Linux release 8.5.

İşlemci: Intel XEON 8362 2.80GHz 64 çekirdek.

Hafıza Miktarı (RAM): 512GB.

Yerel Disk Boyutu: 10TB.

##### 3.2.1. Genom Okumalarının Kalite Kontrolü

Genom verilerine ait okuma sayısı, baz sayısı, %GC içeriği gibi genel istatistikler seqkit (versiyon 2.3.0) [21] yazılımı kullanılarak elde edilmiştir. Kalite kontrol işlemi ise FastQC (versiyon 0.11.9) [22] yazılımı kullanılarak gerçekleştirilmiştir. FastQC yazılımı ile elde edilen çıktılar, MultiQC (versiyon 1.14) [23] yazılımı kullanılarak birleştirilmiş ve tüm bireylere ait veriyi kapsayacak şekilde raporlanmıştır.

##### 3.2.2. Genom Okumalarının Ön Filtreden Geçirilmesi

Okumalar fastp (versiyon 0.23.2) [24] yazılımı kullanılarak kalite skoruna (PHRED) göre filtreleme işleminden geçirilmiştir. PHRED skoru 20'nin altında olan bazlar bir okumanın %40'ından fazlasını kapsadığı takdirde o okuma analiz dışı bırakılmıştır.

### 3.2.2. Genom Okumalarının İnsan Referans Genomuna Hizalanması

Okumaların referans genoma hizalanması işlemi için bwa (versiyon 0.7.17-r1188) [25] yazılımı kullanılmıştır. İnsan referans genom verisi olarak USCS [26] veri tabanından indirilen 2014/01/15 yayın tarihli hg38 versiyonlu insan genomu kullanılmıştır. Hizalanma işlemi tamamlandıktan sonra her bir örnek için *.sam* formatında hizalanma dosyaları elde edilmiştir. Samtools (versiyon 1.17) [27] yazılımı kullanılarak genomik koordinatlarına göre sıralanan ve *.bam* formatına dönüştürülen bu dosyalar, GATK (versiyon 4.2.6.1) [28] yazılımının MarkDuplicates fonksiyonu kullanılarak potansiyel PCR duplikasyonlarından arındırılmışlardır.

### 3.2.3. Hizalanan Genom Okumalarının Referans Genomu Kapsama Oranının Belirlenmesi

Genom okumalarının referans genomu kapsama oranının belirlenmesi için Lander/Waterman denklemi kullanılmıştır. Bu eşitliğe göre üretilen okuma sayısı ve okuma uzunluğunun çarpımının haploid genom uzunluğuna bölünmesiyle, okumaların genomu ne oranda kapsadığı teorik olarak belirlenebilmektedir. Eşitlik şu şekilde gösterilmektedir:

$$C = LN/G$$

Eşitlikteki “C” simgesi, okumaların genomu kapsama oranını ifade etmektedir. “L” simgesi okuma uzunluğunu, “N” simgesi okuma sayısını, “G” simgesi ise haploid genom boyutunu temsil etmektedir.

### 3.2.4. Varyantların Çağırılması

Varyantları çağırılması amacı ile GATK yazılımının germ hattı varyantların çağırılması amacı ile önerilen iş akışı takip edilmiştir (GATK Best Practices) [28]. Bu işlem için, PCR duplikasyonlarından arındırılmış hizalama dosyaları GATK yazılımının HaplotypeCaller fonksiyonu ile analize tabi tutulmuştur. Sonuç olarak, her bir örnek için bireysel varyantları içeren *.g.vcf* formatında dosyalar elde edilmiştir. Popülasyon düzeyinde varyantların eldesi için

GATK yazılımının GenomicsDBImport ve GenotypeGVCFs fonksiyonları kullanılmıştır. GenomicsDBImport fonksiyonu, her bir bireye ait varyant bilgilerini genel bir veri tabanına işlemektedir. Bu işlem tamamlandıktan sonra popülasyona ait tüm varyantları elde etmek için GenotypeGVCFs fonksiyonu kullanılmıştır. Sonuç olarak, filtrelenmeye hazır ve tüm popülasyona ait varyant bilgilerini içeren *.vcf* formatında varyant çağırma dosyası elde edilmiştir.

### 3.2.5. Varyantların Filtrelenmesi

Varyantların filtrelenmesi amacıyla GATK yazılımının VariantRecalibrator ve ApplyVQSR araçları kullanılmıştır. Varyant Kalite Skoru Rekalibrasyonu (VSQR) adı verilen iki adımlı bu işlemin ilk aşaması VariantRecalibrator aracı kullanılarak yapılan model oluşturma aşamasıdır. 1000 Genom Projesinin [14] de dahil olduğu veri tabanlarında mevcut olan bilinen insan SNP' verilerini girdi olarak kullanan VariantRecalibrator aracı sonuç olarak her bir varyanta kalite skoru atamaktadır. Girdi olarak verilen bilinen varyant verisini kullanarak kendini eğiten araç, sorgulanan her bir varyantın gerçek varyant veya veri işleme artefaktı olup olmadığını sorgulamaktadır. Filtreleme işleminin ikinci aşamasında ApplyVQSR aracı kullanılmıştır. VariantRecalibrator aracı kullanılarak oluşturulan varyant kalite puanlarına göre filtreleme işlemi yapan ApplyVQSR aracı kullanılırken doğruluk duyarlılığı seviyesi 0.99 olarak belirlenmiştir ve bu puanın altında kalan varyantlar yalancı pozitif olarak değerlendirilerek analiz dışı bırakılmışlardır.

### 3.2.6. Varyantların Karakterize Edilmesi

Filtreleme işleminden geçen varyantlar SnpEff (versiyon 5.1) [29] yazılımı kullanılarak karakterize edilmiştir. SnpEff yazılımı birçok kaynakta mevcut olan insan referans genom versiyonlarını kullanarak analiz gerçekleştirebilmektedir. Bu analizler için ENSEMBL [30] veri tabanından sağlanan GRCh38 versiyonlu insan referans genomu kullanılmıştır. Dünya popülasyonlarında bulunan diğer varyantlar ile ilgili bilgiler 1000 Genom Projesi verilerden elde edilmiştir. Varyantların klinik öneminin değerlendirilmesi ise ClinVar [18] veri tabanı kullanılarak gerçekleştirilmiştir. Karakterizasyon işlemi tamamlandıktan sonra ise varyantların bulunduğu bölgeler, hangi gen üzerinde bulunduğu, patojenik önemi gibi bilgiler elde edilmiştir.

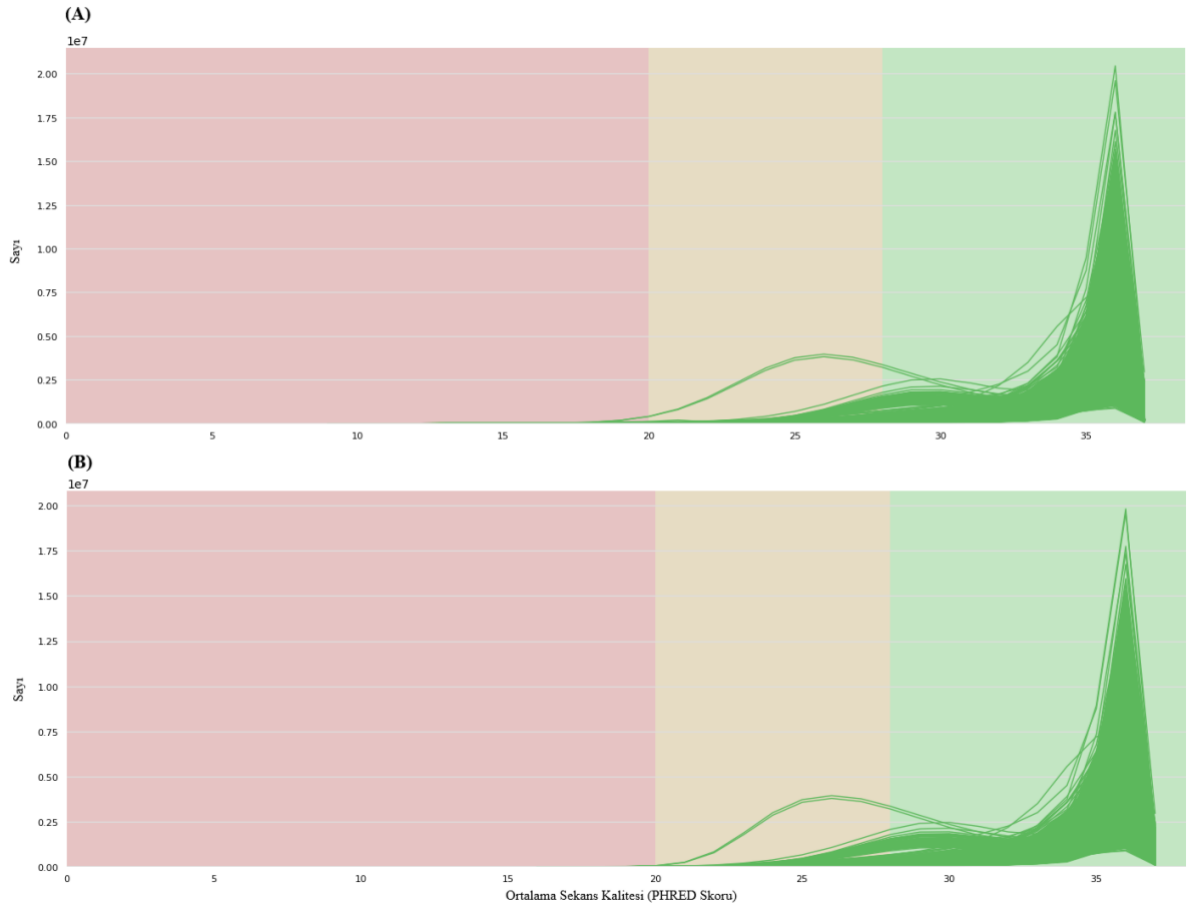
## 4. BULGULAR

### 4.1. OKUMA İSTATİSTİKLERİ

13,044 bireye ait ham okumalar içerdikleri okuma uzunluğu, okuma sayısı, baz sayısı ve baz kalitesi bakımlarından değerlendirilmiştir. Buna göre “single-end” okuma stratejisi ile üretilen DNA okumalarına ait okuma uzunluğu 35 baz, toplam okuma sayısı 173,73 milyar, toplam baz sayısı 6150,50 milyar, belirlenen kalite skorunun altında kalan okuma sayısı ise ( $Q < 20$ ) 0,70 milyar olarak tespit edilmiştir. Filtreleme işlemi gerçekleştirildikten sonra analize hazır temiz okuma sayısı 175,04 milyar, baz sayısı ise 6125,46 milyar olarak tespit edilmiştir. Filtreleme işlemi sonrası birey başına ortalama okuma sayısı 13,42 milyon, ortalama baz sayısı ise 469,60 milyon olarak tespit edilmiştir. Okuma istatistikleri (**Tablo 4.1**) ve filtreleme işlemine dair grafik (**Şekil 4.1**) aşağıdaki gibidir:

**Tablo 4.1:** 13044 bireye ait DNA okumalarına ilişkin istatistikler.

	HAM	FİLTRELENMİŞ ( $Q > 20$ )
Toplam Okuma Sayısı (Milyar)	175,73	175,04
Toplam Baz Sayısı (Milyar)	6150,60	6125,46
Ortalama Okuma Sayısı (Milyon)	13,47	13,42
Ortalama Baz Sayısı (Milyon)	471,53	469,60



**Şekil 4.1:** Ham (A) ve filtrelenmiş (B) okumalara ait PHRED skoru dağılımı grafikleri.

## 4.2. DNA OKUMALARININ İNSAN GENOMUNA HİZALANMASI

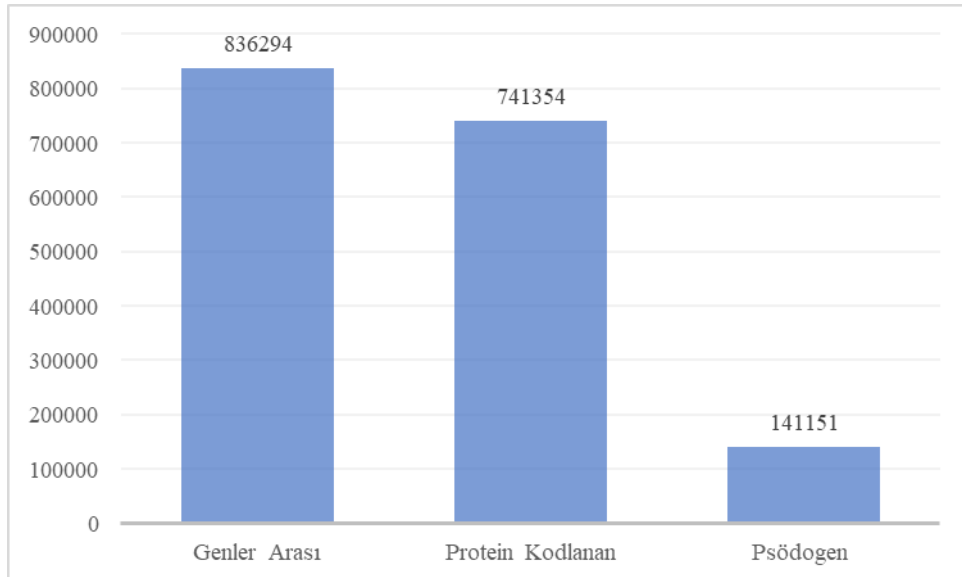
Filtreleme işlemi sonrası elde edilen temiz okumalar, insan referans genomuna hizalanmıştır. 175,04 milyar temiz okumanın 169,83’ü insan genomuna başarılı bir şekilde hizalanmıştır. Birey başına ortalama 13,42 milyon okumanın 13,01 milyonunun insan genomuna hizalandığı gözlenmiştir. Toplam hizalanma oranı %96,94 olarak tespit edilmiştir. Okumaların insan genomunu kapsama oranı ortalama 0,15x olarak tespit edilmiştir. Bu oran birey başına çok düşük olsa da örneklerimize ait okumalar kombine edildiği takdirde kapsama oranı popülasyon düzeyinde varyant analizleri gerçekleştirebilecek düzeyde artmaktadır. Kombine edilmiş kapsama oranı 1978x olarak tespit edilmiştir.

### 4.3. VARYANTLARIN ÇAĞIRILMASI VE FİLTRELENMESİ

Her bir örneğe ait varyant bilgileri ortak bir veri tabanında depolandıktan sonra, popülasyonu yansıtan bütün varyantların çağırılması işlemi gerçekleştirilmiştir. Bunun sonucunda 13,044 bireye ait 1,793,856 adet tek nükleotid polimorfizmi (SNP) verisi elde edilmiştir. Filtreleme işlemi sonrası bu varyantlardan 75,057 tanesi GATK yazılımının VariantRecalibrator aracı tarafından yalancı pozitif olarak değerlendirilmiş ve analiz dışı bırakılmıştır. Filtreleme işlemi öncesi tüm varyantları kapsayan transisyon/transversiyon (ts/tv) oranı 2.17 olarak gözükürken, filtreleme işlemi sonrası elde edilen varyantlar için bu değer 2.15 olarak güncellenmiştir. Yalancı pozitif olarak değerlendirilen varyantlar analiz dışı bırakıldıktan sonra kalan 1,718,799 SNP Türk popülasyonuna ait güvenilir varyantlar olarak değerlendirilip karakterizasyon aşamasına geçilmiştir.

### 4.4. VARYANTLARIN KARAKTERİZE EDİLMESİ

1,718,799 SNP’den oluşan veri setinde bulunan 836,294 SNP’nin genler arası bölgelerde olduğu tespit edilmiştir. Bunu takiben 741,354 SNP’nin protein kodlanan bölgelerde, 141,151 SNP’nin ise psödogen bölgelerinde olduğu tespit edilmiştir (**Şekil 4.2**). Protein kodlanan bölgelerde bulunan 741,354 SNP’nin toplam 4464 gen bölgesinde gerçekleştiği tespit edilmiştir. SNP tespiti yapılan intergenik bölge sayısı ise 2989 olarak hesaplanmıştır. Bu değerlere göre bölge başına 280 SNP içermesiyle intergenik bölgeler en yüksek varyasyonun gözlemlendiği bölgeler olmuşturlar. Protein kodlanan bölgelerde ise bölge başına varyasyon oranı 166 olarak hesaplanmıştır.



**Şekil 4.2:** Tespit edilen varyantların genomik bölgelere göre dağılımı.

ClinVar veri tabanı kullanılarak yapılan patojenite testinde 1822 varyanta ait bilgi elde edilmiştir. Bu varyantlardan 14 tanesi patojenik/muhtemel patojenik olarak sınıflandırılırken 1808 tanesi iyi huylu/muhtemel iyi huylu olarak sınıflandırılmışlardır. Patojenik varyantların tamamının protein kodlanan bölgelerde bulunduğu gözlenmiştir. Tespit edilen patojenik varyantlara ait bilgiler aşağıdaki gibidir (**Tablo 4.2**).

**Tablo 4.2:** Tespit edilen patojenik varyantlara ait bilgiler.

Kromozom	Pozisyon	Referans Allel	Alternatif Allel	Allel Frekansı	Klinik Statü
1	226735804	G	T	0.015	Muhtemel patojenik
1	226736237	A	C	0.00138	Muhtemel patojenik
1	976215	A	G	0.0007666	Patojenik
1	169549811	C	T	0.00011	Patojenik
1	17262194	T	A	0.00007	Patojenik/Muhtemel patojenik
1	21564110	C	T	0.00007	Patojenik/Muhtemel patojenik
1	43430098	C	T	0.00007	Patojenik
2	21002271	A	G	0.00007	Patojenik
2	44312653	T	C	0.00007	Patojenik/Muhtemel patojenik
2	218661516	G	A	0.00007	Muhtemel patojenik
2	240797674	T	C	0.00007	Muhtemel patojenik
4	39504503	G	A	0.00007	Muhtemel patojenik
15	44651712	G	C	0.00007	Patojenik/Muhtemel patojenik
16	3243407	T	C	0.00007	Patojenik/Muhtemel patojenik

Allel frekansı en yüksek olan ve birinci kromozomun 226735804 pozisyonunda Guanin bazının Timin bazına deęişimiyle karakterize edilen SNP, PERM1 geni üzerinde bulunmaktadır. ClinVar [18] veri tabanında klinik önemi belirtilmemekle birlikte, muhtemel patojenik varyant olarak sınıflandırılmışır.

En yüksek ikinci allel frekansına sahip ve birinci Kromozomun 226736237 pozisyonunda bulunan, Adenin bazının Sitozin bazına deęişimiyle karakterize edilen varyant ise ITPKB genini etkilemektedir. Bu varyantın klinik önemi bilinmemekle birlikte, Ferguson ve dię. [31] tarafından yürütölen ve kronik böbrek rahatsızlıęı bulunan hastalarda koroner arter kalsifikasyonunu inceleyen çalıřmaya göre bu varyantın kronik böbrek rahatsızlıęı ile iliřkili olabileceęi belirtilmektedir.

En yüksek dördüncü allel frekansına sahip ve birinci kromozomun 169549811 pozisyonunda bulunan, Sitozin bazının Timin bazına deęişimiyle karakterize edilen SNP, F5 transkripti üzerinde etki göstermektedir. ClinVar veri tabanında iskemik inme, Budd-Chiari Sendromu, yinelenen düşük riski ve Faktör V eksiklięi üzerinde risk faktörü olarak tanımlanan bu varyant aynı zamanda trombofili ve buna baęlı olarak řiddetli COVID-19 hastalıęına karşı duyarlılık etkileriyle patojenik olarak tanımlanmaktadır [32].

En yüksek allel frekansına sahip 4 varyant dıřında kalan 10 varyantın tamamının allel frekansı eřit olarak gözlenmektedir. Bu varyantlar ClinVar [18] veri tabanında yüksek kolesterol, sistinüri, GRACILE sendromu, herediter spastik parapleji, epileptik ensefalopadi ve ailevi Akdeniz ateři hastalıklarıyla iliřkilendirilmektedir.

## 5. TARTIŞMA VE SONUÇ

Bu çalışma kapsamında 13,044 anne adayi bireye ait NIPT amacı ile üretilmiş genomik okumalar, Türk popülasyonuna ait varyant profilini ortaya koymak için analiz edilmiştir. Dizileme derinliği tüm örnekler birleştirildiğinde 1978x olarak tespit edilmiştir. Popülasyonu temsil eden 1,718,799 adet varyant tanımlanmış ve karakterize edilmiştir. Tanımlanan varyantların büyük çoğunluğu iyi huylu olarak etiketlenmiştir ve patojenik etkileri bulunmamaktadır. Buna karşın, tespit edilen 14 adet varyantın patojenik etkilerinin olduğu tanımlanmıştır ve bu varyantların hangi genler üzerinde etki gösterip, hangi hastalıklar ile ilişkili olduğu incelenmiştir

Birinci kromozomun 169549811 pozisyonunda bulunan ve F5 geni üzerinde etki gösteren varyantın ilişkili olduğu hastalıklar özellikle dikkat çekmektedir. F5 geninin COVID-19 hastalığında kritik öneme sahip olan koagülasyon sürecinde rol aldığı tespit edilmiştir [33]. Aynı zamanda F5 geni, hastalığının ağır geçirilmesine sebep olan yan hastalıklarla (böbrek hastalıkları, karaciğer hastalıkları, diyabet, akciğer hastalıkları ve kardiyovasküler hastalıklar) ortak ilişkili bir gen olarak karşımıza çıkmaktadır [34]. Bu çalışmada bu varyanta ait allel frekansı 0.00011 olarak tespit edilmiştir. GnomAD [19] veri tabanına göre Doğu Asya popülasyonunda frekansı 0.00 olan bu varyantın frekansı Avrupa popülasyonunda 0.02700, Afrika popülasyonunda 0.00466, Amerikan popülasyonunda ise 0.00959 olarak karşımıza çıkmaktadır. Bu varyantın küresel frekansı ise 0.017619 olarak tespit edilmiştir. Bu çalışmaya göre Türk popülasyonunda gözlenen frekansı diğer dünya popülasyonlarına göre oldukça düşük olan bu varyant, Dünya Sağlık Örgütü (WHO) verilerine göre açıklanan COVID-19 ölüm oranlarının Türkiye’de Avrupa ülkelerine ve Amerika’ya göre düşüklüğünü açıklıyor olabilir. Avrupa ülkelerinin büyük çoğunluğunda ölüm oranı 100 binde 200 değerinin üzerindeyken, Türkiye’de bu değer 100 binde 124,25 olarak karşımıza çıkmaktadır.

Çalışma kapsamında tespit edilen Türk popülasyonuna ait varyant bilgileri, var olan veri tabanlarında Türk popülasyonu özelinde mevcut olan eksikliği gidermek konusunda destekleyici bir kaynak olarak ortaya koyulmuştur. NIPT verilerinin popülasyon genetiği çalışmalarında kullanımı her ne kadar popüler olmasa da bu çalışma ile birlikte bu yöntemin uygulanabilirliği literatürdeki diğer çalışmalarla birlikte desteklenmiştir. Okuma derinliğinin düşük olmasıyla birlikte NIPT kaynaklı genomik okumalar yalnızca popülasyon düzeyinde

anlamalı veri sunmaktadır [7], bu durum kişisel veri güvenliği konusundaki olası tartışmaları de minimum düzeye indirmektedir [6].

NIPT kaynaklı üretilen genomik veriler gün geçtikçe artarken, popülasyon genetiği çalışmalarında bu verilerin kullanılması dünya popülasyonlarına dair bilgilerimizi genişletmek amacı ile alternatif bir kaynak olarak kullanılabilir. Maliyet avantajı ve popülasyon hakkında bilgi sağlama gücüyle bu verilerin dünya devletleri tarafından sağlık politikalarını şekillendirmek amacıyla kullanımın da olası bir seçenek olarak düşünülmesi gündeme getirilebilir. Bu çalışma, NIPT amacı ile üretilen verilerin genetik varyantları tespit etmek konusunda güçlü bir kaynak olduğunu ortaya koymakta ve Türk popülasyonundan sağlanan büyük ölçekli bir veri setiyle elde edilen bulgularıyla da bunu desteklemektedir.

## KAYNAKLAR

- [1]. Gurdasani, D., Barroso, I., Zeggini, E., & Sandhu, M. S., 2019, Genomics of disease risk in globally diverse populations, *Nature Reviews Genetics*, 20 (9), 520-535.
- [2]. Chiu, R. W., Chan, K. A., Gao, Y., Lau, V. Y., Zheng, W., Leung, T. Y., ... & Lo, Y. D., 2008, Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallel genomic sequencing of DNA in maternal plasma, *Proceedings of the National Academy of Sciences*, 105 (51), 20458-20463.
- [3]. Fan, H. C., Blumenfeld, Y. J., Chitkara, U., Hudgins, L., & Quake, S. R., 2008, Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood, *Proceedings of the National Academy of Sciences*, 105 (42), 16266-16271.
- [4]. Lo, Y. D., Corbetta, N., Chamberlain, P. F., Rai, V., Sargent, I. L., Redman, C. W., & Wainscoat, J. S., 1997, Presence of fetal DNA in maternal plasma and serum, *The lancet*, 350 (9076), 485-487.
- [5]. Budis, J., Gazdarica, J., Radvanszky, J., Harsanyova, M., Gazdaricova, I., Strieskova, L., ... & Szemes, T., 2019, Non-invasive prenatal testing as a valuable source of population specific allelic frequencies, *Journal of biotechnology*, 299, 72-78.
- [6]. Tran, N. H., Vo, T. B., Nguyen, V. T., Tran, N. T., Trinh, T. H. N., Pham, H. A. T., ... & Phan, M. D., 2020, Genetic profiling of Vietnamese population from large-scale genomic analysis of non-invasive prenatal testing data, *Scientific reports*, 10 (1), 19142.
- [7]. Liu, S., Huang, S., Chen, F., Zhao, L., Yuan, Y., Francis, S. S., ... & Xu, X., 2018, Genomic analyses from non-invasive prenatal testing reveal genetic associations, patterns of viral infections, and Chinese population history, *Cell*, 175 (2), 347-359.
- [8]. Allyse, M., Minear, M. A., Berson, E., Sridhar, S., Rote, M., Hung, A., & Chandrasekharan, S., 2015, Non-invasive prenatal testing: a review of international implementation and challenges, *International journal of women's health*, 7, 113-126.
- [9]. Qiao, L., Zhang, Q., Liang, Y., Gao, A., Ding, Y., Zhao, N., ... & Wang, T., 2019, Sequencing of short cfDNA fragments in NIPT improves fetal fraction with higher maternal BMI and early gestational age, *American Journal of Translational Research*, 11 (7), 4450.
- [10]. Bianchi, D. W., & Chiu, R. W., 2018, Sequencing of circulating cell-free DNA during pregnancy, *New England Journal of Medicine*, 379 (5), 464-473.
- [11]. Lo, Y. D., Tein, M. S., Lau, T. K., Haines, C. J., Leung, T. N., Poon, P. M., ... & Hjelm, N. M., 1998, Quantitative analysis of fetal DNA in maternal plasma and serum: implications for noninvasive prenatal diagnosis, *The American Journal of Human Genetics*, 62 (4), 768-775.
- [12]. Wright, C. F., & Burton, H., 2009, The use of cell-free fetal nucleic acids in maternal blood for non-invasive prenatal diagnosis, *Human reproduction update*, 15 (1), 139-151.

- [13]. Agarwal, A., Sayres, L. C., Cho, M. K., Cook-Deegan, R., & Chandrasekharan, S., 2013, Commercial landscape of noninvasive prenatal testing in the United States, *Prenatal diagnosis*, 33 (6), 521-531.
- [14]. Siva, N., 2008, 1000 Genomes Project, *Nature biotechnology*, 26 (3), 256-257.
- [15]. Morshneva, A., Kozyulina, P., Vashukova, E., Tarasenko, O., Dvoynova, N., Chentsova, A., ... & Glotov, A., 2021, Pilot screening of cell-free mtDNA in NIPT: quality control, variant calling, and Haplogroup determination, *Genes*, 12 (5), 743.
- [16]. Ye, X., Lin, S., Song, X., Tan, M., Li, J., Wang, J., ... & Chen, M., 2021, Identification of copy number variants by NGS-based NIPT at low sequencing depth, *European Journal of Obstetrics & Gynecology and Reproductive Biology*, 256, 297-301.
- [17]. Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., & Sirotkin, K., 2001, dbSNP: the NCBI database of genetic variation, *Nucleic acids research*, 29 (1), 308-311.
- [18]. Landrum, M. J., Lee, J. M., Benson, M., Brown, G. R., Chao, C., Chitipiralla, S., ... & Maglott, D. R., 2018, ClinVar: improving access to variant interpretations and supporting evidence, *Nucleic acids research*, 46 (D1), D1062-D1067.
- [19]. Koch, L., 2020, Exploring human genomic diversity with gnomAD, *Nature Reviews Genetics*, 21 (8), 448-448.
- [20]. Kars, M. E., Başak, A. N., Onat, O. E., Bilguvar, K., Choi, J., Itan, Y., ... & Özçelik, T., 2021, The genetic structure of the Turkish population reveals high levels of variation and admixture, *Proceedings of the National Academy of Sciences*, 118 (36), e2026076118.
- [21]. Shen, W., Le, S., Li, Y., & Hu, F., 2016, SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation, *PloS one*, 11 (10), e0163962.
- [22]. Wingett, S. W., & Andrews, S., 2018, FastQ Screen: A tool for multi-genome mapping and quality control, *F1000Research*, 7, 1338.
- [23]. Ewels, P., Magnusson, M., Lundin, S., & Käller, M., 2016, MultiQC: summarize analysis results for multiple tools and samples in a single report, *Bioinformatics*, 32 (19), 3047-3048.
- [24]. Chen, S., Zhou, Y., Chen, Y., & Gu, J., 2018, fastp: an ultra-fast all-in-one FASTQ preprocessor, *Bioinformatics*, 34 (17), i884-i890.
- [25]. Li, H., 2013, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM., *arXiv preprint arXiv:1303.3997*.
- [26]. Karolchik, D., Baertsch, R., Diekhans, M., Furey, T. S., Hinrichs, A., Lu, Y. T., ... & Kent, W. J., 2003, The UCSC genome browser database, *Nucleic acids research*, 31 (1), 51-54.

- [27]. Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., ... & Li, H., 2021, Twelve years of SAMtools and BCFtools, *Gigascience*, 10 (2), 2047-217X.
- [28]. Van der Auwera, G. A., & O'Connor, B. D., 2020, *Genomics in the cloud: using Docker, GATK, and WDL in Terra*, O'Reilly Media., USA, ISBN: 1491975199.
- [29]. Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., ... & Ruden, D. M., 2012, A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3, *fly*, 6 (2), 80-92.
- [30]. Howe, K. L., Achuthan, P., Allen, J., Allen, J., Alvarez-Jarreta, J., Amode, M. R., ... & Flicek, P. (2021). Ensembl 2021, *Nucleic acids research*, 49(D1), D884-D891.
- [33]. Manderstedt, E., Halldén, C., Lind-Halldén, C., Elf, J., Svensson, P. J., Engström, G., ... & Zöller, B., 2022, Thrombomodulin (THBD) gene variants and thrombotic risk in a population-based cohort study, *Journal of Thrombosis and Haemostasis*, 20 (4), 929-935.
- [34]. Dolan, M. E., Hill, D. P., Mukherjee, G., McAndrews, M. S., Chesler, E. J., & Blake, J. A., 2020, Investigation of COVID-19 comorbidities reveals genes and pathways coincident with the SARS-CoV-2 viral disease, *Scientific reports*, 10 (1), 1-11.

## EKLER

## Ek 1. Hasta bilgilendirme ve onam formu örneği.



Genoks Genetik Hastalıklar Değerlendirme Merkezi  
WES/CES/PANEL Hasta Bilgilendirme ve Onam Formu

444 8 732  
www.genoks.com.tr

**Tüm Ekzom Dizileme (Whole Exome Sequencing-WES) / Klinik Ekzom Dizileme (Clinical Exome Sequencing-CES) / Panel Testlerinin Tanımı ve Amacı**

İnsan genomunda bulunan ortalama 20.000 genin yaklaşık 180.000 ekzonunun DNA dizisinin çıkarılması Tüm Ekzom Dizileme (Whole Exome Sequencing-WES) adını almaktadır. İnsanın genomunda sadece hastalık yaptığı kanıtlanmış olan yaklaşık 5.000 genin sekanslanması Klinik Ekzom Dizileme (Clinical Exome Sequencing-CES) olarak adlandırılır. Hastanın endikasyonu ile ilişkilendirilmiş belirli gen gruplarının sekanslanması ise 'Panel Dizileme' olarak adlandırılır. Panel testlerini oluşturan genlerin sayısı klinik duruma göre değişiklik göstermektedir. WES, CES ve panel testlerinde Yeni Nesil Sekanslama (Next Generation Sequencing) teknolojisi ile elde edilen DNA dizisi, referans insan genomu ile karşılaştırılır. Yapılan biyoinformatik analizler sonucunda şüphelenilen bir hastalığın nedeni olabilecek moleküler genetik değişimler değerlendirilir.

**Test Limitasyonları**

Test ile ilgili işlemler, ilgililenen genetik duruma özeldir ve bireyin tüm genetik yapısı hakkında bilgi vermez. Tüm Ekzom Dizileme testi, çeşitli teknik nedenlerden dolayı insan genomundaki tüm ekzonları analiz edemez. Bu test, testi yapılan kişilerin, diğer aile üyelerinin ve/veya doğacak çocuklarının tamamen sağlıklı olacağını garantiye alamaz. Test sonuçlarının hastalık oluşturuca değişikliği açıklayamadığı ve normal olarak yorumlandığı durumlarda hastada genetik veya genetik olmayan başka hastalıklar dışlanamamaktadır. Bu testin doğruluğu, daha önce yapılmış kemik iliği naklinden, yakın tarihli kan infüzyonundan veya kan örneğine harici DNA'yı ekleyebilecek diğer terapilerden etkilenebilir. Bu testte kullanılan DNA, germ hücreleri yerine kandan çıkarılır ve bu nedenle nadiren de olsa gözlemlenen kimerizm'e bağlı olarak sonuç değişimi olasılığı vardır. Bu test, ekzon ve ekzon-intron bağlantı noktalarındaki nokta mutasyonları ve küçük delesyon/duplikasyonları (<20bp) tespit edilebilir, gende incelenemeyen bölgelerdeki (regülatör, promotör, derin intronik vs.) değişimleri, büyük delesyon/duplikasyonları ve kompleks yeniden düzenlenmeleri, dengeli translokasyonları, tetraploidi ve diğer bölgelemleri, kopya sayısı değişimlerini, mozaisizimleri, mitokondriyal DNA mutasyonlarını, metilasyon bozuklukları gibi epigenetik değişimleri tespit edemez. Ekzon-intron bağlantılı noktalarının kapsamları ve okunma derinlikleri değişiklik gösterebilir. DNA tabanlı olarak yapılan bu testin biyoinformatik analizlerinin tamamına yakını yurtdışı kaynaklı veri tabanları esas alınarak yapılmaktadır ve bu hastalıkların çoğunda ülkemiz nüfusu için tanımlanmış mutasyon profilleri ve/veya polimorfik özellikleri çok az bulunmakta ya da bulunmamaktadır. Bu test raporlarında ACMG (American College of Medical Genetics and Genomics) kriterlerine göre sınıflandırılmış fenotip ile ilişkili olduğu düşünülen Patojenik, Olası Patojenik veya Klinik Önemi Kesin Bilinmeyen Varyantlar olarak sınıflandırılmaktadır. Yapılacak ileri ek çalışmalar ile hastalığa neden olabileceği ispatlanabilecek varyantlar 'Güçlü Klinik Önemi Kesin Bilinmeyen Varyantlar' sınıfında değerlendirilmektedir. Güçlü Klinik Önemi Kesin Bilinmeyen Varyantlar raporlanmakta iken, zayıf Klinik Önemi Kesin Bilinmeyen Varyantlar raporlanmamaktadır. Bu varyantlar test edilen hastalıklarla ilişkili açısından güncel verilerle değerlendirilmelidir. Rapor tarihinde patojenitesi anlaşılmasın ve raporda listelenmemiş olan bir varyasyon zaman içerisinde yapılan bilimsel çalışmalarla patojenik olarak sınıflandırılabilir. Nadir polimorfizmler yanlış pozitif/negatifte yol açabilir. Bu nedenle pozitif/negatif sonuçların klinik bulgular doğrultusunda hekim tarafından değerlendirilmesi ve gerekli görüldüğü takdirde diğer tanı yöntemleri ile doğrulanması gerekmektedir.

**Numune Alımı**

Bu testleri yapacak kişiden en az 2 ml EDTA'lı tüpe kan örneği alınması gereklidir. Bu kan örneği genellikle koldaki toplardamarlardan alınmaktadır. Alınan örneğin yeterli olması veya elde edilen DNA'nın istenilen kriterleri sağlamayıp analiz edilemeyecek düzeyde olması gibi durumlarda yeniden örnek istenebilir ve test tekrar edilebilir. Ayrıca ilk analizler sonrasında herhangi şüpheli bir bulgu varlığına veya taşıyıcılık analizi yapılması gerektiğinde anneden, babadan ve diğer aile üyelerinden ilave test yapılması önerilebilir.

**Sonuçlar**

Bu testin sonuçlarına süresi teknik veya analize ait faktörler nedeniyle değişiklik gösterebilir. Sonuçların genetik danışma eşliğinde verilmesi uygundur. Tüm genetik veriler kişiseldir, sonuçlar yalnızca imzası alınan kişi(ler) ile paylaşılacaktır. Genetik test sonuçlarının aile üyeleri için de sonuçları vardır. Analiz edilen genlerden herhangi birinde bir mutasyon / varyant taşıdığı tespit edilirse, bu mutasyon/varyantların klinik öneminin belirlenmesi için anne ve babadan alınan kandan elde edilen DNA ile sadece bu mutasyon/varyantların test edilmesi gerekebilir. Bu nedenle bu onam formunda anne ve babanın da onayları alınmaktadır. Bu sonuçlar aynı zamanda diğer aile bireyleri için de etiklere neden olabilir. Bu durum, klinisyen ile görüşülmelidir, sağlık hizmeti sunucuları arasında yönetim seçenekleri ile ilgili farklı görüşler olabilir, izin alınmadan bu sonuçlar diğer aile bireyleriyle paylaşılmalıdır. Genoks GHDM test sırasında veya sonuçların bilmesinden doğabilecek kaygı veya psikolojik sorunlardan sorumlu değildir. Genoks GHDM hastalıklarının daha iyi anlaşılmasına neden olabilecek biyolojik mekanizmaların araştırılması amacıyla kalan kan numunesini ve verileri kişisel bilgileri gizleyerek araştırma amaçlı kullanabilir.

**Rastantisal Sonuçlar**

WES/CES testinde, kanser veya kalp hastalıkları gibi ciddi durumlara yatkınlıkta ilişkilendirilmiş ve ACMG kılavuzunda listelenmiş olan bazı genlerde, endikasyon sebebinin bağımsız olarak varyant saptanması "Rastantisal Sonuçlar" olarak tanımlanmaktadır. Bu sonuçların raporlanması, kişiden alınan onam formundaki talebine göre yapılmaktadır. Bu sonuçların, diğer aile bireyleri için risk yaratabileceği unutulmamalıdır. Bu sonuçların kişide oluşturabileceği psikolojik durumlardan Genoks GHDM sorumlu değildir.

**Onam Bildirgesi**

Bu testin yapılabilmesi için hastadan veya yasal olarak yetkilendirilmiş veli/yakınından imzalanmış onam formu alınması şarttır. Bu form aynı zamanda hekim tarafından da imzalanarak Genoks GHDM'ye teslim edilmelidir.

- 1) Şahsına veya çocuğuna uygulanan genetik test için test bilgi formunu okudum/birine okuttum ve gerekli açıklama alınmıştır. Genetik analizlerin yazılı açıklamasını gördüm, okudum ve anladım. Test edilen hastalıklarla ilgili genetik temel, olasılık, korunma ve tedavileri ile ilgili bilgileri öğrendim. Test sürecinde karşılaşılabilecek riskleri ve testin sınırları ile ilgili bilgi aldım. Tüm sorularına cevap aldım ve sorunun çıkması için ne kadar süreye ihtiyacım olduğunu biliyorum. Test hakkında başka soru veya endişe olması durumunda iletişime geçebileceğim kişi hakkında bilgilendirildim.
- 2) Test sonuçlarının aşağıda imzası bulunan hekimimle paylaşılmasını kabul ediyorum. DNA tabanlı testlerin karmaşıklığı ve test sonuçlarının önemi nedeniyle, sonuçların hekimler aracılığıyla rapor edilmesini ve test sonuçları için hekimlerle iletişime geçmem gerektiğini kabul ediyorum.
- 3) Bu formu imzalatılarak, kan örneğimde testin yapılması için Genoks GHDM'ye onay veriyorum. Nedenini belirtmeksizin herhangi bir zamanda onayımı tamamen veya kısmen geri çekebileceğimi ve test sonuçlarını öğrenmeme hakkına sahip olduğum bilgisini aldım.
- 4) İsimli olan test materyalinin ve o zamana kadar elde edilen sonuçların istediğim zaman imha edilmesini talep edebileceğimi biliyorum.
- 5) Kullanılmayan test materyalinin laboratuvarında diğer genetik testler için kişisel bilgilerimin saklı kalması koşulu ile saklanmasına ve kullanılmasına onay veriyorum.
- 6) Test sonuçlarının bilimsel amaçlarla istatistiksel veri tabanında saklanması ve kullanılmasına için rıza göstermekteyim. Veri analizlerimin kişisel bilgilerimin gizlenerek Genoks GHDM tarafından herhangi bir rapor ya da yayında kullanılmasına izin veriyorum.
- 7) Kan örneğimin genetik testini, kendi laboratuvarında yapması ya da gerekli görüldüğü takdirde yurt içi veya yurt dışında bir laboratuvarında yaptırması için Genoks GHDM'ye onay veriyorum.
- 8) Endikasyonum dışında bulunan ve ACMG kılavuzunda listelenmiş genlerde saptanan Rastantisal Bulguların raporlandırılmasına onay veriyorum (Yalnız WES/CES yapılan hastalar için cevaplanması gerekmektedir).  Evet  Hayır

**Tarih:**

Hasta Ad/Soyad:

\*İmza:

Anne Ad/Soyad:

Kimlik No / İmza:

Baba Ad/Soyad:

Kimlik No / İmza:

\*18 yaşından küçükler için veli imzası gerekir.

Onam formunun geçerli olabilmesi ve numunenin işleme alınabilmesi için imza alanının eksiksiz doldurulması zorunludur.

**Ek 2. Genetik veri paylaşım protokolü.****GENOKS®**

Genetik Hastalıklar Değerlendirme Merkezi

27.09.2022

**GENETİK VERİ PAYLAŞIM PROTOKOLÜ**

“Büyük ölçekli prenatal test genomik verilerinden Türk popülasyonunun genetik profilinin ortaya konulması” [Genetic profiling of Turkish population from large-scale prenatal testing genomic data] başlıklı tez ve proje kapsamında T.C. Sağlık Bakanlığı'nca ruhsatlı Genoks Genetik Hastalıklar Tanı Merkezi - Genoks GHDM'de yürütülen İnsan Düşük ölçekli tüm genoma dayalı (ing. low-pass whole-genome sequencing) verilerinin Hasta Onam Bildirgesi'nin 11. Maddesinde belirtilen “Kalan test numunesi ve datalarının kişisel bilgilerim anonim hale getirilerek araştırma amaçlı kullanılması için Genoks GHDM'ye onay vermemin, test yapılmasını ve daha fazla tedavi alma hakkımı etkilemeyeceğini anladım ve kabul ediyorum.” husus ile uyumlu olmak kaydı ile sadece bu çalışma kapsamında kullanılmasını izin verilmektedir.

**Ek 3. Etik kurul onayı.**

Tarih ve Sayı: 16.12.2022-1475626



T.C.  
**İSTANBUL ÜNİVERSİTESİ**  
İstanbul Tıp Fakültesi Dekanlığı  
Klinik Araştırmalar Etik Kurulu



Sayı :E-29624016-050.99-1475626  
Konu :Dr. Öğr. Üyesi Vahap ELDEM hk.

**Sayın Dr. Öğr. Üyesi Vahap ELDEM**  
Fen Fakültesi Biyoloji Bölümü

İlgi : Fen Bilimleri Enstitüsünün 27/07/2022 gün ve 1082741 sayılı yazısı

Sorumlu araştırmacılığımı üstlendiğiniz ve Selim Can KURALAY' ın yürüteceği 2022/1867 dosya numaralı "Büyük Ölçekli Prenatal Test Genomik Verilerinden Türk Popülasyonunun Genetik Profilinin Ortaya Konulması" başlıklı çalışma, kurulumuzun 09/12/2022 tarih ve 22 sayılı toplantısında görüşülerek etik yönden uygun bulunmuş olup, tutanaklar ekte sunulmuştur.

Bilgilerinizi rica ederim.

Prof. Dr. Ali Yağız ÜRESİN  
Kurul Başkanı

**Bu belge, güvenli elektronik imza ile imzalanmıştır.**

Belge Doğrulama Kodu :BSCTS40RZC Pin Kodu :09482

Belge Takip Adresi : <https://www.turkiye.gov.tr/istanbul-universitesi-ebys>

İstanbul Tıp Fakültesi Dekanlığı Çapa/Fatih/İSTANBUL

Tel : 0 212 414 21 38/414 20 00-31561 Faks : 0 212 414 21 38 / 635 11 93

e-posta : if-dekanlik@istanbul.edu.tr Elektronik Ağ : <http://istanbul.tip.istanbul.edu.tr>

Keş Adresi: istanbuluniversitesi@hs01.kep.tr

Bilgi için : Cihan KILIÇ

Dahili : 31346



## ÖZGEÇMİŞ

Kişisel Bilgiler	
Adı Soyadı	Selim Can Kuralay
Doğum Yeri	
Doğum Tarihi	Tarih girmek için tıklayın veya dokununuz.
Uyruğu	<input checked="" type="checkbox"/> T.C. <input type="checkbox"/> Diğer:
E-Posta Adresi	
Web Adresi	

Eğitim Bilgileri	
Lisans	
Üniversite	İstanbul Üniversitesi
Fakülte	Fen Fakültesi
Bölümü	Biyoloji
Mezuniyet Yılı	01.09.2020

Yüksek Lisans	
Üniversite	İstanbul Üniversitesi
Enstitü Adı	Fen Bilimleri Enstitüsü
Anabilim Dalı	Biyoloji Anabilim Dalı
Programı	Zooloji Programı

Makale ve Bildiriler	
<p>Cay, S. B., Cinar, Y. U., Kuralay, S. C., Inal, B., Zararsız, G., Ciftci, A., ... &amp; Erol, O., 2022, Genome skimming approach reveals the gene arrangements in the chloroplast genomes of the highly endangered <i>Crocus L.</i> species: <i>Crocus istanbulensis</i> (B. Mathew) Rukšāns, <i>Plos one</i>, 17(6), e0269747.</p>	
<p>Eldem, V., Kuralay, S. C., Özdoğan, G., Özçelik, G. H., Aydın, D., Çakmak, G., ... &amp; Tekin, I. M. (2023). Comprehensive analysis of circulating viral DNA in maternal plasma at population-scale using low-pass whole-genome sequencing, <i>Genomics</i>, 110556.</p>	