

ANKARA ÜNİVERSİTESİ  
BİYOTEKNOLOJİ ENSTİTÜSÜ

TEMEL BİYOTEKNOLOJİ  
DOKTORA TEZİ

GENOM EBADINDAKİ TÜRK POPÜLASYONU TNP VERİLERİNİN VERİ  
TABANININ HAZIRLANMASI VE SONUÇLARIN HAPMAP IŞIĞINDA  
DEĞERLENDİRİLMESİ

Pelin FİDANOĞLU

Danışman Öğretim Üyesi

Prof. Dr. Hilâl ÖZDAĞ

Ekim

2013

## **ETİK BEYAN**

Bu tez çalışmasının; akademik kural ve etik ilkelere baęlı kalınarak hazırlandığını, çalışmada yararlanılan ve bu çalışma ürünü olmayan bütün bilgiler için kaynak yayınlara atıfta bulunulmuş olduğunu beyan ederim.

Pelin FİDANOĞLU

İmzası

Prof. Dr. Hilâl ÖZDAĞ danışmanlığında, Pelin FİDANOĞLU tarafından hazırlanan bu çalışma .../.../.... tarihinde aşağıdaki jüri tarafından Biyoteknoloji Anabilim Dalı'nda doktora tezi olarak kabul edilmiştir.

Başkan : Prof. Dr. Nejat AKAR İmza:

Üye :Prof. Dr. Nurten AKARSU İmza:

Üye :Prof. Dr. Hilal ÖZDAĞ İmza:

Üye :Doç. Dr. Işık YULUĞ İmza:

Üye :Yrd. Doç Dr. Bala GÜR DEDEOĞLU İmza:

Yukarıdaki sonucu onaylarım.

Prof. Dr. Aykut ÖZKUL

Enstitü Müdürü

# ÖZET

Doktora Tezi

Genom Ebadındaki Türk Popülasyonu TNP Verilerinin Veri Tabanının Hazırlanması ve  
Sonuçların HapMap Işığında Değerlendirilmesi

Pelin FİDANOĞLU

Ankara Üniversitesi Biyoteknoloji Enstitüsü

Danışman: Prof. Dr. Hilâl ÖZDAĞ

Genom projelerinin en önemli çıktıları popülasyon ve bireyler arası varyasyonlar ile bu varyasyonların anlamı olmuştur. Uluslararası HapMap projesi ile dünya popülasyonlarını temsil etmek üzere seçilen 4 popülasyonun Tek Nükleotid Polimorfizm (TNP) dökümlerinin ortaya çıkarılması suretiyle insan genomunda yürütülecek olan genomik araştırmalar için temel bir haplotip haritasının çıkarılması hedeflenmiştir. Bugün genomboyu asosiyasyon çalışmalarında (GBAÇ) kullanılan standart TNP mikrodizineri HapMap projesi kapsamında analiz edilen TNP'leri içermektedir.

Bu tez çalışması ile 4 farklı araştırma projesi kapsamında önceden çalışılmış 227 sağlıklı ve 40 hasta örneğin, Affymetrix GeneChip 250K TNP tarama sonuçları MySQL veri tabanında toplanarak ilk kez Türk popülasyonuna özgü bir TNP veri tabanı oluşturulmuştur. TNP veri tabanından sorgular yardımı ile çekilen Türk popülasyon verileri ile 11 HapMap popülasyonunun internet sitesinden indirilen 2009-02 (faz II+III) verilerinin Minör Allel Frekansları (MAF) hesaplanarak, Türk ve HapMap popülasyonlarındaki karşılaştırmaları yapılmıştır. 267 bireyin karşılaştırılabilen 168.993 referans TNP dizisi için, Türk popülasyon verisi bütün olarak ve coğrafi bölgelere ayrılarak analiz edildiğinde HapMap popülasyonlarından farklı olduğu gözlenmiştir. Sonuç olarak bu tez çalışması Türk popülasyonunda yüksek rezolüsyonlu genomik analizlerin yapılması gereğini kantitatif sonuçlarla ortaya koymuştur.

2013, 254 sayfa

**Anahtar kelimeler:** Türk TNP Veri Tabanı, popülasyon genetiği, Haplotip, HapMap.

## **ABSTRACT**

Ph.D. Thesis

Turkish Population Database Design for Genom-wide Association Studies and Comparison  
with HapMap

Pelin FİDANOĞLU

Ankara University Biotechnology Institute

Supervisor: Prof. Dr. Hilâl ÖZDAĞ

One of the most important outcome of the human genome project is inter-individual and inter-population variations in the genome and their meanings. The International HapMap project aimed to reveal the haplotype map of the human genome from Single Nucleotide Polymorphism (SNP) status of the chosen four populations representing human population. Today standard SNP arrays used in genomwide association studies (GWAS) hold the SNP's analyzed in the HapMap project.

In this dissertation 250K array genotype data of a total of 227 healthy individuals and 40 patients was gathered for the first time in a database. The data is then compared to HapMap dataset. Turkish population data was selected with queries and 11 HapMap population data (2009-02 phase II+III) was downloaded from the website. Calculated Minor Allele Frequencies were compared between Turkish and HapMap populations. MAFs of Turkish population were analyzed as a whole and depending on geographic regions and significant differences were found compared to HapMap populations. In conclusion this study quantitatively showed that high-throughput and high resolution genome analysis should be carried out in Turkish Population.

2013, 254 pages

**Keywords:** Turkish SNP Database, population genetics, Haplotype, HapMap

## TEŞEKKÜR

Bu tezin oluşması ve sonuçlanmasında zamanını ve emeğini esirgemeyen sevgili hocam değerli danışmanım Ankara Üniversitesi Biyoteknoloji Enstitüsü Müdür Yardımcısı Prof. Dr. Hilâl Özdağ'a verdiği bilimsel desteği ve sağladığı veriler için,

Hacettepe Üniversitesi Tıp Fakültesi Çocuk Sağlığı ve Hastalıkları Anabilim Dalı Hematoloji Bilim Dalı, Gen Haritalama Laboratuvarı Öğretim Üyesi Prof. Dr. Ayşe Nurten Akarsu'ya bilimsel desteği ve sağladığı veriler için,

Çukurova Üniversitesi, Tıp Fakültesi, Çocuk Sağlığı ve Hastalıkları Anabilim Dalı Öğretim Üyesi Prof. Dr. Ali Kemal Topaloğlu'na, Ankara Üniversitesi, Tıp Fakültesi, Dahili Tıp Bilimleri Bölümü, Çocuk Sağlığı ve Hastalıkları Anabilim Dalı öğretim üyesi Prof. Dr. Mustafa Tekin'e sağladıkları verileri için,

Ankara Üniversitesi Biyoteknoloji Enstitüsü Müdürü iken Prof. Dr. Nejat Akar'a beni bu çalışmaya yönlendirdiği, teşvik ettiği ve çalışmam süresince desteklerini esirgemediği için,

Çalışmamın son zamanlarında beni her zaman destekleyen ve yüreklendiren Yrd. Doç. Dr. Farid Rajablı'ya (Turgut Özal Üniversitesi Mühendislik Fakültesi) ve Dr. Beyza Doğanay Erdoğan'a (Ankara Üniversitesi Tıp Fakültesi Biyoistatistik ABD) bilimsel destekleri için,

Başkent Üniversitesi Bilgi İşlem Daire Başkanı sevgili arkadaşım Ramazan Tekinarslan'a bilimsel desteğini hiçbir zaman esirgemediği için,

Ankara Üniversitesi Biyoteknoloji Enstitüsü doktora öğrencisi sevgili arkadaşım Mehmet Korkmaz'a bilimsel desteğini esirgemediği için,

Tübitak Ulakbim'e numerik hesaplamalarda yüksek başarımlar ve grid hesaplama merkezi (TRUBA) altyapısını kullandığı için,

Ankara Üniversitesi Genombilim grubundaki tüm arkadaşlarıma desteklerini esirgemedikleri için,

Sevgili aileme hayatım boyunca desteklerini ve anlayışlarını esirgemedikleri için sonsuz teşekkürlerimi sunarım.

## İÇİNDEKİLER

ETİK BEYAN .....	i
ONAY SAYFASI .....	ii
ÖZET .....	iii
ABSTRACT .....	iv
İÇİNDEKİLER.....	vi
ŞEKİLLER DİZİNİ.....	xix
ÇİZELGELER DİZİNİ.....	xx
SİMGELER DİZİNİ .....	xxii
1. GİRİŞ .....	1
2. KURAMSAL TEMELLER .....	3
2.1. İnsan Genom Projesi .....	3
2.2. İnsan Genomundaki Varyasyonlar.....	5
2.2.1. Tek Nükleotit Polimorfizmleri.....	6
2.2.2. Diğer Küçük Çaplı ve Orta Ölçekli Yapısal Varyasyonlar .....	9
2.3. Varyasyonların Kullanım Alanları.....	12
2.3.1. Bağlantı (Linkage) .....	13
2.3.2. Asosiyasyon .....	14
2.3.3. TNP Polimorfizmleri ve Farmakogenomik .....	21
2.3.4. Genetik Antropoloji.....	21
2.4. HapMap.....	23
2.5. HapMap Projesi ve Diğer Popülasyonlar.....	26
2.5.1. Çin İnsanı Genom Varyasyon Projesi.....	26
2.5.2. Hindistan Varyasyon Projesi (IGVDB) .....	27
2.5.3. Japon Popülasyon Yapısının TNP Profilleri ile Belirlenmesi.....	27

2.5.4. Estonya Genom Projesi .....	29
2.5.5. Pasifik Pan-Asya TNP Girişimi.....	30
2.5.6. Singapur Genom Varyasyon Projesi.....	31
2.5.7. HapMap TNP verilerinin Singapur Çinli Popülasyonuna Transfer Edilebilirliğinin Değerlendirilmesi .....	31
2.5.8. UK10K - Sağlık ve Hastalıkta Nadir Genetik Varyantlar .....	31
2.5.9. 1000 Genom Projesi.....	32
2.6. Araştırmalar için TNP veri tabanları.....	34
2.6.1. Tek Nükleotit Polimorfizm Veri Tabanı (Short Genetic Variations-dbSNP)35	
2.6.2. TNP Konsorsiyum - Uluslararası HapMap Projesi (The SNP Consortium–TSC-International HapMap Project).....	36
2.6.3. Japon Veri Tabanı (Japanese Single Nucleotide Polimorphisms Database-JSNP) 36	
2.6.4. Pan-Asya TNP Genotip Veri tabanı (the Pan-Asian SNP genotyping database-PanSNPdb) .....	37
2.6.5. İnsan Genom Varyasyonları Veri Tabanı (Human Genome Variation Database-HGVbase).....	38
2.6.6. Genotip ve Fenotip Veri Tabanı (The database of Genotypes and Phenotypes-dbGaP).....	38
2.6.7. Varyasyon Veri Tabanı (Database of Genomic Structural Variation-dbVar) .....	38
2.7. Araştırmalar için diğer veri tabanları.....	38
2.7.1. Kaliforniya Üniversitesi, Santa-Cruz Genom Web Tarayıcısı Veri Tabanı (UCSC Genome Browser Database).....	38
2.7.2. SNPper .....	39
2.7.3. TNP İlişkili Proje Dizilerin Sıralanması ve Değerlendirilmesi (Project Aligned Related Sequences and Evaluate SNP-PARSESNP) .....	40
2.7.4. Fonksiyonel Elemanların TNP Veri Tabanı Functional Element SNPs Database in human (FESD).....	40
2.7.5. Genetik Asosiyasyon Veri Tabanı (the Genetic Association Database-GAD) .....	41



2.8. Genom Boyunca İlişkilendirme Çalışmaları.....	41
2.9. GBAÇ Çalışmalarının Değerlendirilmesi .....	42
2.10. GBAÇ Çalışmalarında TNP Seçim Metotları.....	44
2.11. GBAÇ Çalışmalarında Kullanılan İstatistiksel Analizler .....	45
2.11.1. TNP Frekanslarının Hesaplanması.....	45
2.11.2. R Project ile İstatistiksel Analizler.....	45
2.11.3. Haploview ile Tag TNP analizi .....	46
2.11.4. GBAÇ Çalışmalarında Kullanılan Diğer İstatistiksel Analizler .....	47
2.11.5. Partial Component Analizi .....	47
2.11.6. K-Means (K-Ortalama).....	48
3. GEREKÇE VE AMAÇ .....	51
4. MATERYAL VE YÖNTEM.....	53
4.1. Materyal .....	53
4.1.1. HapMap Popülasyonları Yüksek Çözünürlüklü Verileri.....	54
4.1.2. Türk Popülasyonu Düşük Çözünürlüklü Verileri.....	55
4.1.3. HapMap ve Türk Popülasyonu Karşılaştırılabilir Düşük Çözünürlüklü Örnek ve Varyasyon Sayıları.....	56
4.1.4. Türk Popülasyonu Bölgelerinin kendi aralarında ve HapMap ile Karşılaştırılabilir Düşük Çözünürlüklü Örnek ve Varyasyon Sayıları.....	57
4.2. Yöntem .....	58
4.2.1. Veri Toplama .....	58
4.2.2. Veri Tabanında Saklanacak Verilerin İşlenmesi ve Veri Yığınının Seçilimi .....	58
4.2.3. Veri Tabanında Saklanacak Veri Kaynakları ve Sayıları .....	60
4.2.4. Veri Tabanı Mantıksal Tasarımı.....	61
4.2.5. Veri Tabanı Fiziksel Tasarımı.....	65
4.2.6. Veri tabanı Araçları ile Türk TNP Veri Tabanı Oluşturulması .....	70

<b>4.2.7. Yüksek Çözünürlüklü HapMap Popülasyon Verilerinin Minör Allel Frekanslarının Hesaplanması .....</b>	<b>71</b>
<b>4.2.8. HapMap ve Düşük Çözünürlüklü Türkiye Popülasyon Verilerinin Minör Allel Frekanslarının Hesaplanması .....</b>	<b>72</b>
<b>4.2.9. Türkiye'nin 7 Coğrafi Bölgesinin Minör Allel Frekanslarının Hesaplanması.....</b>	<b>72</b>
<b>4.2.10. HapMap Popülasyonlarının ve Türkiye'nin 7 Coğrafi Bölgesinin için Minör Allel Frekanslarının Hesaplanması .....</b>	<b>72</b>
<b>4.2.11. HapMap Popülasyonları Yüksek Çözünürlüklü Verileri ile Hiyerarşik Kümeleme Analizi .....</b>	<b>73</b>
<b>4.2.12. HapMap Popülasyonları Düşük Çözünürlüklü Verileri ile Hiyerarşik Kümeleme Analizi .....</b>	<b>74</b>
<b>4.2.13. 7 Coğrafik Bölge Verileri (188) ile Hiyerarşik Kümeleme Analizi (Düşük Çözünürlük).....</b>	<b>74</b>
<b>4.2.14. 7 Coğrafik Bölgenin Verileri (188) ve Memleketi Bilinmeyen (79) veri ile Hiyerarşik Kümeleme Analizi (Düşük Çözünürlük) .....</b>	<b>74</b>
<b>4.2.15. 11 HapMap Popülasyonu ve 7 Coğrafik Bölge Verisi (188) ile Hiyerarşik Kümeleme Analizi (Düşük Çözünürlük).....</b>	<b>75</b>
<b>4.2.16. 11 HapMap Popülasyonu ve 267 Türk Popülasyon Verisinin ile Hiyerarşik Kümeleme Analizi (Düşük Çözünürlük).....</b>	<b>75</b>
<b>5. ARAŞTIRMA BULGULARI.....</b>	<b>76</b>
<b>5.1. Yüksek Çözünürlüklü HapMap Popülasyon Verilerinin MAF Karşılaştırması .....</b>	<b>76</b>
<b>5.2. Düşük Çözünürlüklü HapMap ve Türkiye Popülasyon Verilerinin Minör Allel Frekans Karşılaştırmaları .....</b>	<b>78</b>
<b>5.3. Türkiye'nin 7 Coğrafi Bölgesinin Minör Allel Frekanslarının Hesaplanması.....</b>	<b>80</b>
<b>5.4. HapMap Popülasyonlarının ve Türkiye'nin 7 Coğrafi Bölgesinin MAF Karşılaştırmaları .....</b>	<b>80</b>
<b>5.5. HapMap Popülasyonları Yüksek Çözünürlüklü Verileri ile Hiyerarşik Kümeleme Analizi .....</b>	<b>81</b>
<b>5.6. HapMap Popülasyonları Düşük Çözünürlüklü Verileri ile Hiyerarşik Kümeleme Analizi .....</b>	<b>82</b>

5.7. 7 Coğrafik Bölge Verileri (188) ile Hiyerarşik Kümeleme Analizi (Düşük Çözünürlük).....	83
5.8. 23 Kromozom Boyunca Türkiye (Bölgesi Bilinmeyenler dahil) Hiyerarşik Kümeleme .....	84
5.9. 11 HapMap Popülasyonu ve 7 Coğrafik Bölge Verisi (188) ile Hiyerarşik Kümeleme Analizi (Düşük Çözünürlük).....	85
5.10. 11 HapMap Popülasyonu ve 267 Türk Popülasyon Verisinin ile Hiyerarşik Kümeleme Analizi (Düşük Çözünürlük).....	86
6. TARTIŞMA ve SONUÇ .....	88
6.1. Çalışmada Kullanılan Verinin Yapısı ve Veri Tabanının Oluşturulması .	88
6.2. Tag TNP Analizleri.....	89
6.3. Popülasyonlar içi ve Popülasyonlar arası MAF karşılaştırmaları.....	90
6.4. Kümeleme Analizleri .....	92
6.5. Sonuç.....	93
KAYNAKLAR.....	95
EKLER.....	100
EK1. Genomik diziler için Genom Tarayıcıları.....	100
EK2. Genom Varyasyon Veritabanları .....	102
EK3. Türkiye Bölge Kodları.....	103
EK4: Çipte her rs'e karşılık gelen en güncel allelleri (AllelA/B) NCBI web sitesinden çeker108	
EK5. Matlab ile HapMap Popülasyon Frekanslarının Hesaplanması .....	110
EK6. Matlab ile HapMap Popülasyonlarının Ortak rs'lerin Seçilimi (Yüksek Çözünürlük) .....	117
EK7. Türk Popülasyon Frekanslarının Hesaplanması .....	119
EK8. HapMap ve Türk Popülasyon rs'lerinin Ortak olanlarının Seçilimi.....	121
EK9. R 2.15.3 (2013-03-01) Programı ile Öklid Uzaklığına Göre Hiyerarşik Kümeleme Kodu .....	123

<b>EK10. Güneybatı ABD’de Afrika Kökenlilerin (ASW) Minör Allel Sayıları (Yüksek Çözünürlük) .....</b>	<b>124</b>
<b>EK11. Webuye’da Luhya Kabilesi, Kenya (LWK ) Minör Allel Sayıları (Yüksek Çözünürlük) .....</b>	<b>125</b>
<b>EK12. Kinyawa’da Yaşayan Maasai’lerin, Kenya (MKK) Minör Allel Sayıları (Yüksek Çözünürlük) .....</b>	<b>126</b>
<b>EK13. Ibida’da Yaşayan Yorubalı’ların, Nijerya (YRI) Minör Allel Sayıları (Yüksek Çözünürlük) .....</b>	<b>127</b>
<b>EK14. Tokyoda yaşayan Japon’ların, Japonya (JPT) Minör Allel Sayıları (Yüksek Çözünürlük) .....</b>	<b>128</b>
<b>EK15. Beijing’li Han Çinlilerinin, Çin (CHB) Minör Allel Sayıları (Yüksek Çözünürlük) .....</b>	<b>129</b>
<b>EK16. Denver Metropolünde Yaşayan Çinlilerin, Kolarado (CHD) Minör Allel Sayıları (Yüksek Çözünürlük).....</b>	<b>130</b>
<b>EK17. Houston’da Yaşayan Gujarati Hintlilerinin, Teksas (GIH) Minör Allel Sayıları (Yüksek Çözünürlük).....</b>	<b>131</b>
<b>EK18. Los Angeles’da yaşayan Meksika Kökenlilerin, Kaliforniya (MEX) Minör Allel Sayıları (Yüksek Çözünürlük).....</b>	<b>132</b>
<b>EK19. CEPH Koleksiyonundan Kuzey ve Batı Avrupa Kökenli olup Utah’a Yerleşenlerin (CEU) Minör Allel Sayıları (Yüksek Çözünürlük).....</b>	<b>133</b>
<b>EK20. İtalya Yaşayan Toskanalıların (TSI) Minör Allel Sayıları (Yüksek Çözünürlük) .....</b>	<b>134</b>
<b>EK21. HapMap Afrika Popülasyonlarının Minör Allel Sayıları (Düşük Çözünürlük)..</b>	<b>135</b>
<b>EK22. HapMap Asya Popülasyonlarının Minör Allel Sayıları (Düşük Çözünürlük).....</b>	<b>136</b>
<b>EK23. HapMap ABD+Avrupa Popülasyonlarının Minör Allel Sayıları (Düşük Çözünürlük) .....</b>	<b>137</b>
<b>EK24. HapMap 4A Popülasyonları Minör Allel Sayıları (Düşük Çözünürlük).....</b>	<b>138</b>
<b>EK25. HapMap Afrika ve Türk Popülasyonunun Ortak Minör Allel Sayıları (Düşük Çözünürlük) .....</b>	<b>139</b>
<b>EK26. HapMap Asya ve Türk Popülasyonunun Ortak Minör Allel Sayıları (Düşük Çözünürlük) .....</b>	<b>140</b>

<b>EK27. HapMap ABD+Avrupa ve Türk Popülasyonunun Ortak Minör Allel Sayıları (Düşük Çözünürlük).....</b>	<b>141</b>
<b>EK28. 1. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Yüksek Çözünürlük).....</b>	<b>142</b>
<b>EK29. 2. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Yüksek Çözünürlük).....</b>	<b>142</b>
<b>EK30. 3. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Yüksek Çözünürlük).....</b>	<b>143</b>
<b>EK31. 4. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Yüksek Çözünürlük).....</b>	<b>143</b>
<b>EK32. 5. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Yüksek Çözünürlük).....</b>	<b>144</b>
<b>EK33. 6. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Yüksek Çözünürlük).....</b>	<b>144</b>
<b>EK34. 7. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Yüksek Çözünürlük).....</b>	<b>145</b>
<b>EK35. 8. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Yüksek Çözünürlük).....</b>	<b>145</b>
<b>EK36. 9. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Yüksek Çözünürlük).....</b>	<b>146</b>
<b>EK37. 10. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Yüksek Çözünürlük).....</b>	<b>146</b>
<b>EK38. 11. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Yüksek Çözünürlük).....</b>	<b>147</b>
<b>EK39. 12. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Yüksek Çözünürlük).....</b>	<b>147</b>
<b>EK40. 13. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Yüksek Çözünürlük).....</b>	<b>148</b>
<b>EK41. 14. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Yüksek Çözünürlük).....</b>	<b>148</b>
<b>EK42. 15. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Yüksek Çözünürlük).....</b>	<b>149</b>
<b>EK43. 16. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Yüksek Çözünürlük).....</b>	<b>149</b>

<b>EK44. 17. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Yüksek Çözünürlük) .....</b>	<b>150</b>
<b>EK45. 18. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Yüksek Çözünürlük) .....</b>	<b>150</b>
<b>EK46. 19. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Yüksek Çözünürlük) .....</b>	<b>151</b>
<b>EK47. 20. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Yüksek Çözünürlük) .....</b>	<b>151</b>
<b>EK48. 21. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Yüksek Çözünürlük) .....</b>	<b>152</b>
<b>EK49. 22. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Yüksek Çözünürlük) .....</b>	<b>152</b>
<b>EK50. X. Kromozomu HapMap Popülasyonları Hiyerarşik Kümeleme (Yüksek Çözünürlük) .....</b>	<b>153</b>
<b>EK51. 1. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Düşük Çözünürlük) .....</b>	<b>153</b>
<b>EK52. 2. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Düşük Çözünürlük) .....</b>	<b>154</b>
<b>EK53. 3. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Düşük Çözünürlük) .....</b>	<b>154</b>
<b>EK54. 4. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Düşük Çözünürlük) .....</b>	<b>155</b>
<b>EK55. 5. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Düşük Çözünürlük) .....</b>	<b>155</b>
<b>EK56. 6. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Düşük Çözünürlük) .....</b>	<b>156</b>
<b>EK57. 7. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Düşük Çözünürlük) .....</b>	<b>156</b>
<b>EK58. 8. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Düşük Çözünürlük) .....</b>	<b>157</b>
<b>EK59. 9. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Düşük Çözünürlük) .....</b>	<b>157</b>
<b>EK60. 10. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Düşük Çözünürlük) .....</b>	<b>158</b>

<b>EK61. 11. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Düşük Çözünürlük) .....</b>	<b>158</b>
<b>EK62. 12. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Düşük Çözünürlük) .....</b>	<b>159</b>
<b>EK63. 13. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Düşük Çözünürlük) .....</b>	<b>159</b>
<b>EK64. 14. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Düşük Çözünürlük) .....</b>	<b>160</b>
<b>EK65. 15. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Düşük Çözünürlük) .....</b>	<b>160</b>
<b>EK66. 16. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Düşük Çözünürlük) .....</b>	<b>161</b>
<b>EK67. 17. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Düşük Çözünürlük) .....</b>	<b>161</b>
<b>EK68. 18. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Düşük Çözünürlük) .....</b>	<b>162</b>
<b>EK69. 19. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Düşük Çözünürlük) .....</b>	<b>162</b>
<b>EK70. 20. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Düşük Çözünürlük) .....</b>	<b>163</b>
<b>EK71. 21. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Düşük Çözünürlük) .....</b>	<b>163</b>
<b>EK72. 22. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Düşük Çözünürlük) .....</b>	<b>164</b>
<b>EK73. X. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Düşük Çözünürlük) .....</b>	<b>164</b>
<b>EK74. 1. Kromozom Türkiye (7 Bölge) Hiyerarşik Kümeleme .....</b>	<b>165</b>
<b>EK75. 2. Kromozom Türkiye (7 Bölge) Hiyerarşik Kümeleme .....</b>	<b>165</b>
<b>EK76. 3. Kromozom Türkiye (7 Bölge) Hiyerarşik Kümeleme .....</b>	<b>166</b>
<b>EK77. 4. Kromozom Türkiye (7 Bölge) Hiyerarşik Kümeleme .....</b>	<b>166</b>
<b>EK78. 5. Kromozom Türkiye (7 Bölge) Hiyerarşik Kümeleme .....</b>	<b>167</b>
<b>EK79. 6. Kromozom Türkiye (7 Bölge) Hiyerarşik Kümeleme .....</b>	<b>167</b>

<b>EK80. 7. Kromozom Türkiye (7 Bölge) Hiyerarşik Kümeleme .....</b>	<b>168</b>
<b>EK81. 8. Kromozom Türkiye (7 Bölge) Hiyerarşik Kümeleme .....</b>	<b>168</b>
<b>EK82. 9. Kromozom Türkiye (7 Bölge) Hiyerarşik Kümeleme .....</b>	<b>169</b>
<b>EK83. 10. Kromozom Türkiye (7 Bölge) Hiyerarşik Kümeleme .....</b>	<b>169</b>
<b>EK84. 11. Kromozom Türkiye (7 Bölge) Hiyerarşik Kümeleme .....</b>	<b>170</b>
<b>EK85. 12. Kromozom Türkiye (7 Bölge) Hiyerarşik Kümeleme .....</b>	<b>170</b>
<b>EK86. 13. Kromozom Türkiye (7 Bölge) Hiyerarşik Kümeleme .....</b>	<b>171</b>
<b>EK87. 14. Kromozom Türkiye (7 Bölge) Hiyerarşik Kümeleme .....</b>	<b>171</b>
<b>EK88. 15. Kromozom Türkiye (7 Bölge) Hiyerarşik Kümeleme .....</b>	<b>172</b>
<b>EK89. 16. Kromozom Türkiye (7 Bölge) Hiyerarşik Kümeleme .....</b>	<b>172</b>
<b>EK90. 17. Kromozom Türkiye (7 Bölge) Hiyerarşik Kümeleme .....</b>	<b>173</b>
<b>EK91. 18. Kromozom Türkiye (7 Bölge) Hiyerarşik Kümeleme .....</b>	<b>173</b>
<b>EK92. 19. Kromozom Türkiye (7 Bölge) Hiyerarşik Kümeleme .....</b>	<b>174</b>
<b>EK93. 20. Kromozom Türkiye (7 Bölge) Hiyerarşik Kümeleme .....</b>	<b>174</b>
<b>EK94. 21. Kromozom Türkiye (7 Bölge) Hiyerarşik Kümeleme .....</b>	<b>175</b>
<b>EK95. 22. Kromozom Türkiye (7 Bölge) Hiyerarşik Kümeleme .....</b>	<b>175</b>
<b>EK96. X. Kromozomu Türkiye (7 Bölge) Hiyerarşik Kümeleme .....</b>	<b>176</b>
<b>EK97. 1. Kromozom TR Bölgeleri ve HapMap Popülasyonları Hiyerarşik Kümeleme .....</b>	<b>176</b>
<b>EK98. 2. Kromozom TR Bölgeleri ve HapMap Popülasyonları Hiyerarşik Kümeleme .....</b>	<b>177</b>
<b>EK99. 3. Kromozom TR Bölgeleri ve HapMap Popülasyonları Hiyerarşik Kümeleme .....</b>	<b>177</b>
<b>EK100. 4. Kromozom TR Bölgeleri ve HapMap Popülasyonları Hiyerarşik Kümeleme.....</b>	<b>178</b>
<b>EK101. 5. Kromozom TR Bölgeleri ve HapMap Popülasyonları Hiyerarşik Kümeleme.....</b>	<b>178</b>



<b>EK102. 6. Kromozom TR Bölgeleri ve HapMap Popülasyonları Hiyerarşik Kümeleme.....</b>	<b>179</b>
<b>EK103. 7. Kromozom TR Bölgeleri ve HapMap Popülasyonları Hiyerarşik Kümeleme.....</b>	<b>179</b>
<b>EK104. 8. Kromozom TR Bölgeleri ve HapMap Popülasyonları Hiyerarşik Kümeleme.....</b>	<b>180</b>
<b>EK105. 9. Kromozom TR Bölgeleri ve HapMap Popülasyonları Hiyerarşik Kümeleme.....</b>	<b>180</b>
<b>EK106. 10. Kromozom TR Bölgeleri ve HapMap Popülasyonları Hiyerarşik Kümeleme.....</b>	<b>181</b>
<b>EK107. 11. Kromozom TR Bölgeleri ve HapMap Popülasyonları Hiyerarşik Kümeleme.....</b>	<b>181</b>
<b>EK108. 12. Kromozom TR Bölgeleri ve HapMap Popülasyonları Hiyerarşik Kümeleme.....</b>	<b>182</b>
<b>EK109. 13. Kromozom TR Bölgeleri ve HapMap Popülasyonları Hiyerarşik Kümeleme.....</b>	<b>182</b>
<b>EK110. 14. Kromozom TR Bölgeleri ve HapMap Popülasyonları Hiyerarşik Kümeleme.....</b>	<b>183</b>
<b>EK111. 15. Kromozom TR Bölgeleri ve HapMap Popülasyonları Hiyerarşik Kümeleme.....</b>	<b>183</b>
<b>EK112. 16. Kromozom TR Bölgeleri ve HapMap Popülasyonları Hiyerarşik Kümeleme.....</b>	<b>184</b>
<b>EK113. 17. Kromozom TR Bölgeleri ve HapMap Popülasyonları Hiyerarşik Kümeleme.....</b>	<b>184</b>
<b>EK114. 18. Kromozom TR Bölgeleri ve HapMap Popülasyonları Hiyerarşik Kümeleme.....</b>	<b>185</b>
<b>EK115. 19. Kromozom TR Bölgeleri ve HapMap Popülasyonları Hiyerarşik Kümeleme.....</b>	<b>185</b>
<b>EK116. 20. Kromozom TR Bölgeleri ve HapMap Popülasyonları Hiyerarşik Kümeleme.....</b>	<b>186</b>
<b>EK117. 21. Kromozom TR Bölgeleri ve HapMap Popülasyonları Hiyerarşik Kümeleme.....</b>	<b>186</b>
<b>EK118. 22. Kromozom TR Bölgeleri ve HapMap Popülasyonları Hiyerarşik Kümeleme.....</b>	<b>187</b>

<b>EK119. X Kromozomu TR Bölgeleri ve HapMap Popülasyonları Hiyerarşik Kümeleme.....</b>	<b>187</b>
<b>EK120. 1. Kromozom Türkiye ve HapMap Popülasyonları Hiyerarşik Kümeleme.</b>	<b>188</b>
<b>EK121. 2. Kromozom Türkiye ve HapMap Popülasyonları Hiyerarşik Kümeleme.</b>	<b>188</b>
<b>EK122. 3. Kromozom Türkiye ve HapMap Popülasyonları Hiyerarşik Kümeleme.</b>	<b>189</b>
<b>EK123. 4. Kromozom Türkiye ve HapMap Popülasyonları Hiyerarşik Kümeleme.</b>	<b>189</b>
<b>EK124. 5. Kromozom Türkiye ve HapMap Popülasyonları Hiyerarşik Kümeleme.</b>	<b>190</b>
<b>EK125. 6. Kromozom Türkiye ve HapMap Popülasyonları Hiyerarşik Kümeleme.</b>	<b>190</b>
<b>EK126. 7. Kromozom Türkiye ve HapMap Popülasyonları Hiyerarşik Kümeleme.</b>	<b>191</b>
<b>EK127. 8. Kromozom Türkiye ve HapMap Popülasyonları Hiyerarşik Kümeleme.</b>	<b>191</b>
<b>EK128. 9. Kromozom Türkiye ve HapMap Popülasyonları Hiyerarşik Kümeleme.</b>	<b>192</b>
<b>EK129. 10. Kromozom Türkiye ve HapMap Popülasyonları Hiyerarşik Kümeleme ....</b> .....	<b>192</b>
<b>EK130. 11. Kromozom Türkiye ve HapMap Popülasyonları Hiyerarşik Kümeleme ....</b> .....	<b>193</b>
<b>EK131. 12. Kromozom Türkiye ve HapMap Popülasyonları Hiyerarşik Kümeleme ....</b> .....	<b>193</b>
<b>EK132. 13. Kromozom Türkiye ve HapMap Popülasyonları Hiyerarşik Kümeleme ....</b> .....	<b>194</b>
<b>EK133. 14. Kromozom Türkiye ve HapMap Popülasyonları Hiyerarşik Kümeleme ....</b> .....	<b>194</b>
<b>EK134. 15. Kromozom Türkiye ve HapMap Popülasyonları Hiyerarşik Kümeleme ....</b> .....	<b>195</b>
<b>EK135. 16. Kromozom Türkiye ve HapMap Popülasyonları Hiyerarşik Kümeleme ....</b> .....	<b>195</b>
<b>EK136. 17. Kromozom Türkiye ve HapMap Popülasyonları Hiyerarşik Kümeleme ....</b> .....	<b>196</b>
<b>EK137. 18. Kromozom Türkiye ve HapMap Popülasyonları Hiyerarşik Kümeleme ....</b> .....	<b>196</b>
<b>EK138. 19. Kromozom Türkiye ve HapMap Popülasyonları Hiyerarşik Kümeleme ....</b> .....	<b>197</b>

<b>EK139. 20. Kromozom Türkiye ve HapMap Popülasyonları Hiyerarşik Kümeleme ....</b>	<b>197</b>
<b>EK140. 21. Kromozom Türkiye ve HapMap Popülasyonları Hiyerarşik Kümeleme ....</b>	<b>198</b>
<b>EK141. 22. Kromozom Türkiye ve HapMap Popülasyonları Hiyerarşik Kümeleme ....</b>	<b>198</b>
<b>EK142. X. Kromozomu Türkiye ve HapMap Popülasyonları Hiyerarşik Kümeleme ...</b>	<b>199</b>
<b>EK143. NN (Okunamayan) rs Frekans Hesaplamaları .....</b>	<b>200</b>
<b>ÖZGEÇMİŞ .....</b>	<b>203</b>
<b>GENOM PROJELERİ 5N1P: .....</b>	<b>205</b>

## ŞEKİLLER DİZİNİ

Şekil 2.1. Genomik Varyasyonların Boyutları .....	6
Şekil 2.2. TNP, Haplotip, Tag TNP .....	9
Şekil 2.3. TNP1 ve TNP2 .....	15
Şekil 2.4. Uzayda koordinatları kodlanmış örnekler için iki adet hedef küme tanımlanır..	49
Şekil 2.5. Bu sınıf tanımlarına uzaklıklarına göre (örneğin öklit mesafesi) bütün örnekler sınıflandırılır (hangi renge daha yakınsa) .....	49
Şekil 2.6. Oluşan sınıfları ayıran bir hat çizilebilir .....	49
Şekil 2.7. Daha önceden sınıflandırılan örneklerin merkezleri bulunur.....	50
Şekil 2.8. Merkezleri hareket ettirdikten sonra örneklerden bazıları yeni merkezlere daha yakın olabilir.....	50
Şekil 4.1. tbl_hastalik .....	65
Şekil 4.2. tbl_kisi Örnek Tablosu .....	66
Şekil 4.3. tbl_ornek .....	67
Şekil 4.4. tbl_snp .....	68
Şekil 4.5. ornek_donusum .....	69
Şekil 4.6. Türk TNP Veri tabanı.....	71
Şekil 5.1. 1. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Yüksek Çözünürlük).....	82
Şekil 5.2. Düşük Çözünürlüklü Veri ile HapMap Popülasyonları 23 Kromozom Hiyerarşik Kümeleme.....	83
Şekil 5.3. 23 Kromozom Boyunca Türkiye 7 Coğrafi Bölgenin Hiyerarşik Kümelemesi..	84
Şekil 5.4. 23 Kromozom Boyunca Türkiye (Bölgesi Bilinmeyenler Dahil) Hiyerarşik Kümeleme.....	85
Şekil 5.5. 23 Kromozom TR Bölgeleri ve HapMap Popülasyonları Hiyerarşik Kümeleme .....	86
Şekil 5.6. 23 Kromozom Türkiye ve HapMap Popülasyonları Hiyerarşik Kümeleme.....	87

## ÇİZELGELER DİZİNİ

<b>Çizelge 2.1.</b> İnsan Genom Projesine Katkıda Bulunan Merkez ve Ülkeler .....	4
<b>Çizelge 2.2.</b> TNP1 ve TNP2 için Alleler .....	15
<b>Çizelge 2.3.</b> TNP1 ve TNP2 Allel Frekansları Tablo Gösterimi .....	16
<b>Çizelge 2.4.</b> TNP1 ve TNP2 Haplotipleri .....	16
<b>Çizelge 2.5.</b> TNP1 ve TNP2 Haplotip Frekansları .....	16
<b>Çizelge 2.6.</b> Dengede olma durumu .....	17
<b>Çizelge 2.7.</b> LD .....	17
<b>Çizelge 2.8.</b> D Çözümü .....	17
<b>Çizelge 2.9.</b> HapMap Projesi Aşamaları .....	25
<b>Çizelge 4.1.</b> Çalışmada Kullanılan Yüksek Çözünürlüklü HapMap Örnek ve Sayıları .....	54
<b>Çizelge 4.2.</b> Çalışmada Kullanılan Düşük Çözünürlüklü Türk Popülasyon Veri Sayıları ..	55
<b>Çizelge 4.3.</b> Çalışma Süresince Karşılaştırılan Düşük Çözünürlüklü HapMap ve Türk Popülasyonu Örnek ve Sayıları .....	56
<b>Çizelge 4.4.</b> Çalışma Süresince Karşılaştırılan Düşük Çözünürlüklü HapMap ve Türk Popülasyon Bölgelerinin Karşılaştırılan Örnek ve Varyasyon Sayıları .....	57
<b>Çizelge 4.5.</b> 10K-50K rs örtüşme yüzdesi .....	59
<b>Çizelge 4.6.</b> 10K-250K rs örtüşme yüzdesi .....	59
<b>Çizelge 4.7.</b> 50K-250K rs örtüşme yüzdesi .....	59
<b>Çizelge 4.8.</b> Veri Tabanına Aktarılan Türk Popülasyon Veri Sayıları ve Kaynakları .....	60
<b>Çizelge 4.9.</b> Veri Tabanına Aktarılan Türk Popülasyon Verilerinin Bölgesel Dağılımları	60
<b>Çizelge 5.1.</b> HapMap Popülasyonları Minör Allel Sayıları (Yüksek Çözünürlük) .....	76
<b>Çizelge 5.2.</b> HapMap Popülasyonları MAF Karşılaştırması (Yüksek Çözünürlük) .....	77
<b>Çizelge 5.3.</b> HapMap Popülasyon MAF'larının ikili ve üçlü Karşılaştırmaları (Yüksek Çözünürlük) .....	77
<b>Çizelge 5.4.</b> HapMap Popülasyonları ve TR popülasyonu MAF Karşılaştırmaları (Düşük Çözünürlük) .....	78

<b>Çizelge 5.5.</b> HapMap Popülasyonlarının MAF'larının ikili ve üçlü Karşılaştırmaları (Düşük çözünürlük) .....	79
<b>Çizelge 5.6.</b> HapMap Popülasyonları ve Türk popülasyonu MAF Karşılaştırmaları (Düşük Rezolüsyon) .....	79
<b>Çizelge 5.7.</b> Türkiye Coğrafi Bölgelerine Göre MAF Hesaplamaları .....	80
<b>Çizelge 5.8.</b> HapMap ve Türkiye Coğrafi Bölgelerine Göre MAF Karşılaştırmaları (Düşük Rezolüsyon) .....	81

## **SİMGELER DİZİNİ**

ANOVA	Analysis of Variance
ASD	Allel Sharing Distance
cDNA	complementary DNA
CGHDP	The Chinese Human Genome Diversity Project
CNV	Copy-Number Variation
cSNP	coding SNP
DAC	Data Access Committee
DAR	Data Access Request
dbGaP	The database of Genotypes and Phenotypes
dbProP	Protein Polimorfizm Veri tabanı
dbSNP	The Single Nucleotide Polymorphism Database
DUC	Data Use Certification
EBI	European Bioinformatics Institute
EGF	Estonian Genome Foundation
EGP	Estonya Genom Projesi
EMBL	European Molecular Biology Laboratory
EMBL	European Molecular Biology Laboratory
ENCODE	The Encyclopedia of DNA Elements
EST	Expressed Sequence Tag
FESD	The Functional Element SNP Database
GAD	The Genetic Association Database
GBAÇ	Genom Boyu Asosiyasyon Çalışmaları
GType	GeneChip Genotyping Analysis Software

GWAS	A Genome-wide Association Study
HapMap	Haplotype Map
HGBASE	Human Genic Bi-Allelic Sequences
HGC	Human Genome Center
HGVbase	Human Genome Variation Database
HTML	Hyper Text Markup Language
HUGO	Human Genome Organization
IGV	The Indian Genome Variation
IGVDB	The Indian Genome Variation Database
IMS	Institute of Medical Science
İGP	İnsan Genom Projesi
JSNP	The Japanese Database
JST	Japan Science and Technology Corporation
KI	Karolinska Institute
LD	Linkage Disequilibrium
MAF	Minor Allel Frequency
MDS	Multidimensional Scaling
mtDNA	Mitokontrial DNA
NCBI	National Center for Biotechnology Information
ncSNP	non-coding SNP
NHGRI	National Human Genome Research Institute
OMIM	Online Mendelian Inheritance in Man
PARSESNP	Project Aligned Related Sequences and Evaluate SNP
PC	Principal Component



PCA	Principal Component Analysis
PCoA	Principal Coordinate Analysis
PCR	Polymerase Chain Reaction
pSNP	promoter SNP@promoter
RFLP	Restriction Fragment Length Polimorphism
rs#	reference SNP ID number, refSNP cluster
SNP	Single Nucleotide Polymorphisms
ss#	submitted SNP ID number
STR	Short Tandem Repeat
TNP	Tek Nükleotit Polimorfizm
TR-Grid	TÜBİTAK ULAKBİM Yüksek Başarımlı ve Grid Hesaplama Merkezi
TSC	The SNP Consortium
UCSC	The University of California, Santa Cruz
VNTR	Variable Number of Tandem Repeat
WGAS	Whole Genome Association Studies
XML	Extensible Markup Language

## 1. GİRİŞ

İnsan Genom Projesinin (İGP) 2003 yılında tamamlanmasının ardından 2004 yılından itibaren rezolüsyonu artarak piyasaya çıkan yeni nesil dizileme teknolojileri genom boyunca analiz maliyetlerinin azalmasına, dolayısıyla veri artışına neden olmuştur. Günümüzde gelişen teknolojiler sayesinde her birey için tek tek analiz yapılması mümkün gibi görünse de bu hem zaman hem de maliyet açısından uygun olamamaktadır. Bu nedenle araştırmacılar, klinik uygulamalardaki hayati süreçleri en aza indirerek maksimum sayıda insanın genomik bilgiden faydalanabilmesi amaçlamaktadırlar. Bu amaca hizmet etmek üzere projelerin popülasyon düzeyinde yürütülebilmesi için oluşturulan uluslararası konsorsiyumlar genom boyunca varyasyon çalışmalarına başlamışlardır.

Genom boyu asosiyasyon, hastalık haritalaması ve evrim çalışmalarına destek olmak amacıyla varyasyonların %90'ını oluşturan ve en informatif belirteç olan TNP'lere yoğunlaşarak HapMap Projesi başlatılmıştır. Üç aşamada gerçekleştirilen proje ile farklı merkezlerde kimlikleri belirsiz kişi örneklerinin TNP taramaları yapılmıştır. Projenin 1. ve 2. aşamasında dünyayı temsil ettiği varsayılan 4 popülasyon, 270 örnek, 3. aşamasında ise 11 popülasyon, 1301 örneğin TNP taraması sonucunda 11 popülasyonun haplotip haritaları, phased haplotipleri, bağlantı dengesizliği analizleri (LD), tag TNP'leri, genotip frekansları saptanarak kantitatif karşılaştırmaları yapılmıştır. Proje ile ortaya çıkan veri yığınları kolay, hızlı ve güvenilir erişimin sağlanması amacı ile bir veri tabanında toplanmış ve web tabanlı olarak araştırmacıların kullanımına açılmıştır. HapMap projesi ile popülasyonlarda farklılığa neden olan varyasyonların tanımlanması amaçlanmıştır. HapMap popülasyonları kullanılarak yapılan popülasyon karşılaştırmaları, seçilen örneklerin aynı popülasyon grubuna ait olmasının ve grubun bütünü temsil ediyor olmasının önemini vurgulamaktadır.

Birçok çok ülke devlet destekli projeler ile kendi araştırma gruplarını kurarak varyasyon veri tabanlarını oluşturmaktadır. Böylece HapMap popülasyonlarının kendi popülasyonlarını ne derece temsil ettiğini araştırmak ve moleküler genetiğin avantajlarından yararlanmak için popülasyon yapılarını, varyasyonları yardımıyla, tanımlamaya çalışmaktadır. Özellikle bu amaca yönelik olarak henüz bir devlet desteği bulunmayan ülkemizde maliyeti yüksek olan bu tür popülasyon çalışmalarının önemi giderek farkedilmektedir.

Bu kapsamdaki çalışmaların önemine dikkat çekmek amacıyla deęişik kaynaklardan toplanan 250K TNP mikrodizin analiz sonuçlarından elde edilen verileri Türk popülasyon bilgisine dönüştürmek amacı ile yürütölen bu tez çalışması sayesinde Türk popülasyonu için hem ilk kez bir varyasyon veri tabanı oluşturulmuş hem de ilk defa popülasyonumuza ait veriler bu kapsamda analiz edilmiştir.

Bu tez çalışmasının kuramsal temeller bölümüne, bu çalışmanın olgunlaşmasına kaynak teşkil eden İnsan genom projesi, genomdaki varyasyonlar ve kullanım alanları hakkında bilgi verilerek başlanmıştır. Daha sonra informatif varyasyon olan TNP'lerin haritalarının çıkartılmasının amaçlandığı HapMap projesi ve bu projenin dięer popülasyonlarla karşılaştırılmalarının yapıldığı farklı genom projelerinden bahsedilmiştir. Kuramsal temellere, bu projeler sonucunda ortaya çıkan verilere araştırmacıların hızlı ve güvenilir erişimini sağlamak amacı ile oluşturulan veri tabanları incelenerek devam edilmiştir. Son olarak HapMap dışında yürütölen genom projeleri ve genom boyu asosiyasyon çalışmalarında kullanılan istatistiksel analizler üzerinde durulmuştur. Çalışmanın yürütölmesi için kullanılan materyal ve yöntem, ardından araştırma bulguları ve sonuç bölümü ile son verilmiştir.

## 2. KURAMSAL TEMELLER

1944 yılında DNA'nın genetik materyal olduğunun anlaşılması ve 1953 yılında Watson ve Crick'in DNA'nın yapısını keşfinden sonra genetik bilimi, pek çok bilim insanı için çok önemli bir alan olmuştur. Genom boyunca oluşan hataları ve hataların sebep olduğu hastalıkları bulabilmek için insan genom dizisinin tamamının okunarak genlerin işleyişinin ortaya konması amacıyla 1980'li yılların başlarında İGP konuşulmaya başlanmıştır.

### 2.1. İnsan Genom Projesi

İGP'nin 1986 yılında Amerika Birleşik Devletlerinde (ABD) başladığı kabul edilmektedir. 1988 yılında ABD Ulusal Enerji Bakanlığı ile Ulusal Sağlık Enstitüsü 5 yıllık bir plan oluşturmak üzere ortak bir komite kurmuş ve proje 1990 yılında yol almaya başlamıştır. Daha sonra, Fransa, İngiltere ve Japonya olmak üzere diğer ülkeler de, uluslararası bir organizasyon olan İnsan Genom Organizasyonu (HUGO) tarafından koordine edilen benzer projelere başlamışlardır.

İGP ile insan haploit genomuna ait 3,3 milyar nükleotitin baz dizisi belirlenerek; 50.000-140.000 arasında değiştiği düşünülen (gerçekte ~ 20.000-25.000) genin tespit edilmesi, elde edilen verilerin veri tabanında saklanması, gerekli araştırmaların yapılabilmesi amacıyla bilim insanlarının kullanımına açılması, çalışmalarla ilgili teknolojilerin özel sektöre aktarılması ve proje ile ortaya çıkabilecek, legal, etik ve sosyal durumlara dikkat çekilmesi amaçlanmıştır.

Proje aktif olarak 1995 yılında yol almaya başladığında, birbirlerinden yaklaşık 1 Mb (1 milyon baz çifti uzunluğundaki DNA) uzaklığında olan belirteçler (marker) kullanılarak her bir kromozoma ait genetik haritalar çıkartılmıştır (**Çizelge 2.1**). Genetik belirteçlerin kromozom üzerindeki yerlerinin belirlenmesi ve heterozigot ailelerdeki kalıtımın gözlenmesiyle aralarındaki uzaklıklar (genetik harita) ortaya konmuştur. Kromozomlardaki belirteçlerin yerleri tespit edilerek her bir kromozomun fiziksel haritası oluşturulmuştur. Maya yapay kromozomlarındaki (YAC) veya diğer vektörlerdeki, kromozomu kapsayan bölgelerde bir seri çakışan klonlar bulunarak, genomun tamamının nükleotit dizisi saptanmıştır.

## Çizelge 2.1. İnsan Genom Projesine Katkıda Bulunan Merkez ve Ülkeler

MERKEZ	ÜLKE
The Whitehead Institute/MIT Center for Genome Research	ABD
The Wellcome Trust Sanger Institute	İngiltere
Washington University School of Medicine Genome Sequencing Center	ABD
United States DOE Joint Genome Institute	ABD
Baylor College of Medicine Human Genome Sequencing Center	ABD
RIKEN Genomic Sciences Center	Japonya
Genoscope and CNRS	Fransa
GTC Sequencing Center, Genome Therapeutics Corporation	ABD
Department of Genome Analysis, Institute of Molecular Biotechnology	ABD
Beijing Genomics Institute/Human Genome Center, Institute of Genetics	Çin
Multimegabase Sequencing Center, The Institute for Systems Biology	ABD
Stanford Genome Technology Center	ABD
Stanford Human Genome Center and Department of Genetics	ABD
University of Washington Genome Center	ABD
Department of Molecular Biology, Keio University School of Medicine	Japonya
University of Oklahoma's Advanced Center for Genome Technology, Dept. of Chemistry and Biochemistry	ABD
Max Planck Institute for Molecular Genetics	Almanya
Cold Spring Harbor Laboratory, Lita Annenberg Hazen Genome Center	ABD
GBF - German Research Centre for Biotechnology	Almanya

Kamunun İGP çalışmaları bu seyirde devam ederken, 1998 yılında Craig Venter (Shotgun dizi saptama yöntemini geliştiren bilim insanı) insan genomunun baz dizisinin belirlenmesini 2001 yılında tamamlayacak özel bir firmanın kurulduğunu ilan etmiştir. Otomatik DNA dizi analizi makinelerinin imalatçısı olan firmanın işbirliği ile Craig Venter shotgun yöntemini kullanarak, insan genomunun dizilenmesini hükümet destekli projeden daha hızlı ve ucuz olarak tamamlayacağını duyurmuştur. Kamunun kullandığı klasik dizileme yönteminde genomik kütüphane önce bir seri kontigler şeklinde organize edilerek ve büyük klonların her biri, dizisi çıkartılmak ve doğru sıraya konmak üzere daha küçük klonlara bölünmüştür. Shotgun yönteminde ise gelişigüzel dizi analizleri yapılan klonlar bilgisayar analizleri yardımıyla örtüşen bölgelerinden kontigler halinde birleştirilmiştir. Bu şekilde yapılan dizi analizinde fiziksel harita hazırlama basamağı atlanmış olduğundan, insan genomu gibi büyük bir genomun dizilenmesi için gerekli zaman böylece büyük ölçüde kısaltılmıştır (1). 2001 yılında, ilan edilenden 2 sene önce, insan genomunun taslağı

yayınlanmış ve proje 2003 yılında iki farklı yöntemle tamamlanarak, sürekli güncellenmiştir.

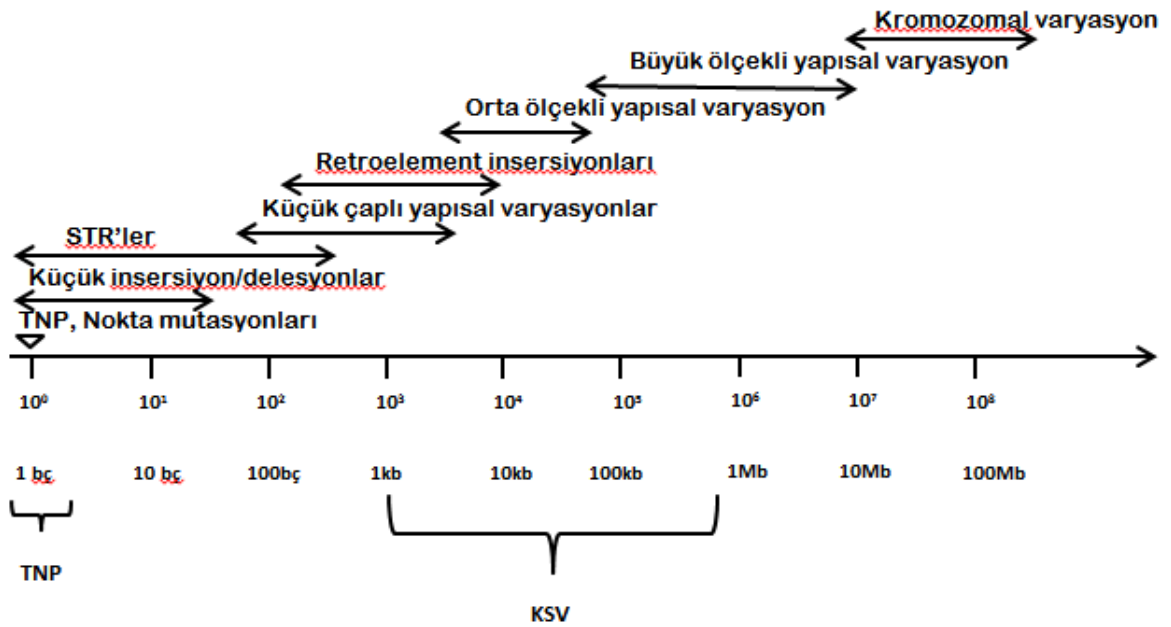
## 2.2. İnsan Genomundaki Varyasyonlar

İnsan genomunda varyasyonlar popülasyonda görülme sıklıkları % 1'den az olan mutasyonlar ve % 1'den fazla olan polimorfizmlerdir. Bir "bireysel farklılık" olan polimorfizmler bireylerin farklı fiziksel özelliklerinden, hastalıklara yatkınlık ve dirençten sorumludur. İnsülin direnci geliştirmeye direnç ve yatkınlık sağlayan bazı genler vardır. Kimi insanların bunlara sahip olması, kimilerinin olmaması bir polimorfizmdir.

Bir canlı organizmanın fenotip özellikleri, onun sahip olduğu genotip ile içinde yaşadığı çevrenin etkisine dayanmaktadır. Bir popülasyon içindeki fenotiplerdeki değişimin önemli kısmı, genotip farklılıklardan kaynaklanmaktadır (2). Bir genin allel frekansı, aynı genin diğer form ve tiplerine oranla, zaman içinde daha çok veya daha az yaygın hale gelmektedir. Yeni bir allel tipi sabitlendiğinde, ya ondan önce gelen allel tipinin yerini alarak ya da tamamıyla popülasyondan silinerek, bir varyasyonun kaybolmasına yol açmaktadır.

Hardy-Weinberg kuralına göre, yeterince büyük bir popülasyon içinde var olan allellerin frekansları (genlerdeki varyasyonlar), yumurta ve spermin oluşumu sırasında allellerin rastgele dağılması ve üreme hücrelerinin döllenmesi sırasında bu allellerin rastgele kombinasyonları gibi popülasyona etki eden güçler tarafından sabit kılınmaktadır (3). Bu anlamda varyasyonlar, genetik materyallerdeki mutasyon, eşeyli üreme yoluyla genlerin yeniden karıştırılması ve popülasyonlar arasındaki göçler (gen akışı) sonucu oluşmaktadır. Mutasyonlar ve gen akışı ile popülasyonlarda sürekli yeni varyasyonlar oluşsa da, bir türün genomunun büyük bölümü, o türün tüm bireylerinde benzer olup özdeştir (4). Buna rağmen, genotipteki nispeten küçük farklılıklar bile, fenotiplerde büyük farklılıklara yol açabilmektedir. Örneğin, şempanze ve insan genomu arasındaki farkların büyük bir kısmı, % 5,07 oranında delesyon ve insersiyonlardan kaynaklanırken, nükleotit farklılıklar ise sadece % 1.52 oranında olup böylece toplam sapma % 6,58 olarak tahmin edilmiştir. Hatta DNA tekrarları ile düşük karmaşıklıkta DNA'lar çıkarıldığında bu farklılık oranı % 2,37'ye kadar düşmektedir (5).

Genetikte popülasyon içindeki bireyler arasında ya da popülasyonlar arasında ortaya çıkabilen gen allellerinde gözlemlenen ve % 1' den daha sık gözlenen DNA farklılıkları çeşitli düzeylerde kendini gösterebilmektedir (**Şekil 2.1**). DNA, gen, kromozom, protein ve protein fonksiyonunda görülebilen varyasyonların çoğu anlamsız olup (örneğin, sessiz mutasyonlar) canlının hayatta kalma ve adaptasyon yeteneğini etkilememektedir. Bazen varyasyonlar hastalığa neden olmasa da varyasyonlardan biri bu duruma yatkınlık verebilmektedir. Farklı şekillerde etkilerini gösteren genetik varyasyonlardaki artış ve azalmanın sürekliliği, doğal seçilim, göç, iç eşleşme, genetik sürüklenme (genetic drift) gibi etkenlerle sağlanmaktadır (6; 7).



**Şekil 2.1.** Genomik Varyasyonların Boyutları

Çeşitli varyasyonlar için yaklaşık ölçüler **Şekil 2.1**'de gösterildiği gibidir. Sınırlar belirsiz olmasına rağmen tüm kromozomdan küçük ve 1 kilobazdan büyük dizi değişiklikleri yapısal varyantlar olarak tanımlanmaktadır. (8; 9)

### 2.2.1. Tek Nükleotit Polimorfizmleri

Genomdaki tek baz değişiklikleri çoğunlukla nötral değişiklikler olmasına rağmen bazıları gen aktivitesini değiştirmektedir. Bazı tek baz değişiklikleri direk olarak hastalığa katkısı olan gen fonksiyon veya regülasyonunda farklılığa yol açan DNA sekans varyantlarıdır ve

gen içerisinde kodlama bölgelerinde, promotor bölgede veya regülatör bölgede olabilir. Bu şekilde olan tek baz değişimleri bir proteinin fonksiyon, regülasyon ya da ekspresyonunu değiştirebilmektedir. Popülasyonda % 1'den az görülen bu değişikliklere mutasyon denir.

Tek Nükleotit Polimorfizmleri (TNP) (**Şekil 2.2.a**) yaygın olarak bireyler arasında DNA dizisinde tek bir nükleotitin (A, T, C veya G) farklı olma durumudur. Diğer varyasyon çeşitleri arasında en fazla rezolüsyona sahip olan TNP'lere insan genomunda ~300 bç de 1 rastlanmaktadır. Bu baz değişiklikleri genellikle haplotip blokları olarak kalıtlanmaktadır.

Sinonim olmayan (non-synonymous) TNP'ler en yaygın TNP tipleridir ve fonksiyonel tek baz değişimlerinin bulunmasında genetik belirteç (marker) olarak kullanılmaktadırlar. Belirteç TNP'lerin ve fonksiyonel tek baz değişiklikleri ile beraber kalıtlanması sebebiyle bu ikili etkin olarak kullanılmaktadır. Rezolüsyonları diğer varyasyonlardan daha fazla olduğu için mayozda crossing over sırasında diğer varyasyon çeşitlerinden parça değişimine uğrama olasılıkları daha düşüktür. Bu sebeple haplotip blokları olarak kalıtlanan TNP'ler hastalıkların takibinde hastalıkla ilişkili gen bölgesi ile hareket ettiklerinden genomdaki en önemli belirteçlerdir. Bu yaygın kullanım nedeniyle günümüzde birçok hastalığın ilişkilendirilmesi, tanı ve kişiye özgü tedavilerin geliştirilmesinde TNP'lerden sıklıkla bahsedilmektedir.

TNP'ler sözkonusu olduğunda Minör Allel Frekansı (MAF) ve Haplotip tanımları önem kazanmaktadır. TNP'lerin popülasyonda görülme sıklıkları frekans ile tanımlanır. Popülasyonda en az yaygın allelin frekansı "minör allel frekans" olarak adlandırılır ve HapMap projesi ile MAF'ları % 5'in üzerinde olan TNP'ler hedeflenmiştir. MAF aşağıdaki şekilde hesaplanır.

*MAF Hesaplanması:*

Hardy-Weinberg dengesi gözününde alındığında;

$f(AA)$ ,  $f(Aa)$  ve  $f(aa)$  bir lokusu temsil eden iki allelin üç farklı genotipi ise A allelinin p frekansı ve a allelinin q frekansı allelerin sayımı ile elde edilmektedir.

Bir popülasyon için A allelinin toplam p frekansı **Denklem 1** ve a allelinin q frekansı **Denklem 2**'deki gibi hesaplanmaktadır (10).



$$p = f(AA) + 1/2f(Aa)$$

### Denklem 1

$$q = f(aa) + 1/2f(Aa)$$

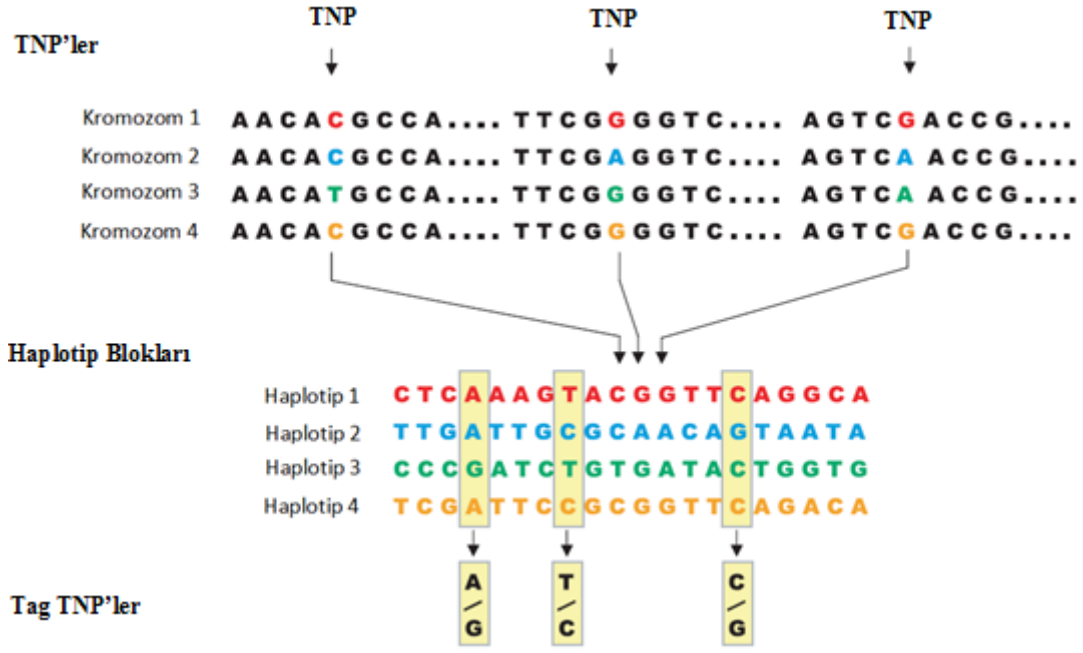
### Denklem 2

İGP'den günümüze kadar projelerden ortaya konan TNP'lerin kataloglandığı dbSNP (Bölüm 16.1'de bahsedildiği üzere) yaygın polimorfizmleri nadir varyantlardan ayır etmek için "evrensel MAF" tanımını kullanmaktadır. dbSNP yaygın (common) polimorfizmlerden nadir (rare) polimorfizmleri ayırt etmek için varsayılan popülasyonlardaki her rs için evrensel MAF'ı raporlamaktadır. Bu nedenle dbSNP'de MAF en sık gözlenen 2. Alleldir. Başka bir deyişle 0,50, 0,49 ve 0,01 frekanslarına sahip 3 allel için MAF 0,49 olarak raporlanmaktadır. Varsayılan popülasyon, farklı popülasyonlardan 1094 bireye ait Mayıs 2011'de yayınlanan 1000 Genome faz1 veriseti (11) kullanılmaktadır.

Örneğin rs222 için  $MAF/MinörAllelSayısı = 0,262/330$  olarak raporlandı ise buradan 1000 Genome faz1 popülasyonlarından Minör Allelin %26,2 frekansla "G" olduğu ve aynı popülasyonda 629 kişide (veya 1258 kromozomda) 330 kez gözlendiğini göstermektedir (12).

### Haplotip:

İnsan DNA dizisinin yalnızca %3-5'i protein (~30,000 gen) kodlanmaktadır. TNP'ler ise çoğunlukla kodlanmayan bölgelerde bulunur. Kodlanan bölgelerde bulunan tek baz değişimleri protein değişikliklerine neden olabildiklerinden araştırma konusudurlar. Bir bireyin bütün TNP'ler için tiplendirmek pek çok kez tekrar, büyük bir işlem hacmi ve maliyet gerektirmektedir. Hem işlem hacminin azaltılması hem de birbiriyle ilişkide olan TNP gruplarının tanımlanması için çalışılmaktadır. Bundan dolayı araştırmalar tek bir kromozomda birbirine yakın lokuslarda bulunan ve "haplotip" olarak tanımlanan allel kompleksleri (**Şekil 2.2.b**) üzerine yoğunlaştırılmıştır. İşlem hacminin daha da düşürülmesi için bazı allellerin seçilimi (**Şekil 2.2.c**) ile haplotip bloğunun tamamının seçilimi üzerinde durulmaktadır. Bu tür bilgiler yaygın hastalıkların genetik temellerinin araştırılmasında önem taşıdığından **Bölüm 2.4.**'de bahsedilen uluslararası HapMap projesinin araştırma konusudur.



Şekil 2.2. TNP<sup>1</sup>, Haplotip<sup>2</sup>, Tag TNP<sup>3</sup>

## 2.2.2. Diğer Küçük Çaplı ve Orta Ölçekli Yapısal Varyasyonlar

### Restriksiyon Parça Uzunluk Polimorfizmi (Restriksiyon Fragment Length Polymorphism-RFLP)

İnsan genomu boyunca, özellikle kodlayıcı olmayan bölgelerde, her 200 nükleotitte bir dizi farklılığı görülmektedir. Bu dizi farklılıkları tek nükleotit değişimleri, bir veya birden fazla nükleotitin delesyonu veya insersiyonu şeklinde olabilir. Bu değişiklik bir restriksiyon enziminin kesim noktasını ortadan kaldırabildiği gibi yeni bir kesim bölgesi de oluşturabilmektedir. Bu şekilde restriksiyon enzimleri ile oluşturulan parça uzunluklarındaki farklılıklar RFLP olarak adlandırılmaktadırlar. RFLP'ler yaygın olup, insan genomunda binlercesi tanımlanmıştır. Bunların birçoğu kromozomlara özgüdür.

<sup>1</sup> **TNP'ler:** 4 farklı insana ait aynı lokasyonda dört kromozomal bölgedeki TNP'ler. Dizinin çoğunluğu aynı olmasına rağmen varyasyonun olduğu 3 farklı baz gösterilmiştir. Her TNP 2 allel olasılığına sahiptir. 1. TNP'de C ve T alleleri mevcuttur.

<sup>2</sup> **Haplotip Blokları:** Yan yana TNP'lerin kombinasyonudur. Örnek olarak 20 TNP bloğu temsili olarak gösterilmektedir. Panel a'daki 3 temsili TNP b'de işaretlenmiştir. Bu panelde popülasyon üyeleri genelde 1-4 haplotipini göstermektedir.

<sup>3</sup> **Tag TNP'ler:** Bu 20 TNP içerisinde 3 tanesi 4 Haplotipi benzersiz olarak tanımlamaktadır. Eğer bir kromozomda A-T-C deseni var ise bu örnek haplotip1 olarak saptanır. (The International HapMap Project, December 2003)

Kodominant Mendel kalıtım modeli gösteren RFLP'ler genomda belli bir lokusa ait ailesel allellerin anne (maternal) veya babaya (paternal) ait olduğunu ayırt etmeye yardımcı olurlar. Kromozoma özgül bölgelerde haritalandıkları için bir ailede genetik hastalıkların nesilden nesile geçişini takip etmek amacıyla belirleyici olarak kullanılan RFLP'lerin medikal arařtırmalardaki önemi genomun dizilimine dair verinin henüz çok kısıtlı olduđu yıllarda ilk olarak Botstein ve arkadaşları tarafından ortaya konmuřtur (13).

### **Deęişken Sayılı Bitişik Tekrar (Variable Number of Tandem Repeat-VNTR) (Minisatelit)**

VNTR'ler RFLP'lerin bir çeşidi olarak ilk kez Nakamura ve arkadaşları tarafından ortaya konmuřtur (14). Polimorfik olan VNTR'ler çoęunlukla heterozigotluk gösterirler. VNTR'ler 15 veya 100 bç uzunluęunda genlerin içinde veya arasında olabilirler. Bu tür yapılar DNA kümeleri şeklinde genom boyunca daęılmış durumdadır.

VNTR belirteçleri, Jeffrey ve arkadaşlarının deęişken minisatelit dizileri isimli çalışması (15) sonucunda genomdan ayrılmıřtır. Polimorfik yapılarının fazla olması nedeniyle VNTR'ler adli tıp uygulamalarında belirteç olarak kullanılmaktadır. Bunun yanı sıra özellikle kanser arařtırmalarında anneden veya babadan gelen allel kaybı ile oluřan kromozom veya kromozom parça kayıplarının (heterozigotluęun kaybı) belirlenmesinde de kullanılmaktadır.

### **Kısa Bitişik Tekrarlar (Short Tandem Repeats-STR) (Mikrosatelit)**

İnsan genomunda sık rastlanan orta sıklıkta tekrarlanan DNA baz dizileri, genellikle ardışık tekrarlanan veya serpiřtirilen diziler olarak genomda bulunmaktadır. Kodlamayan bölgede (junk DNA), deęişik motif ve uzunluklardaki (2-10 baz çift) ardışık, orta sıklıkta tekrarlayan nükleotit diziler STR olarak adlandırılmaktadır. Genom boyunca daęılmış durumda olan STR'lerin bir DNA bölgesindeki tekrar sayısı bireyler arasında farklılık göstermektedir. Örneęin insanlarda en yaygın mikrosatelit ikili nükleotit (CA)<sub>n</sub> tekrarlarıdır. Tekrar sayısı genellikle 5 ve 50 arasında deęişir ve mutasyon oranları ile allel sayıları fazladır. Bu nedenle STR'ler genetik baęlantı analizlerinde ve evrimsel analizlerde kullanılmaktadırlar (16).

STR'ler ilk keřfedildiklerinde bazı bilim adamları tarafından işe yaramayan diziler olarak nitelendirilmiřlerdir. Fakat son yıllarda yapılan çalışmalar, bu dizilerdeki deęişikliklerin

özellikle ileri yaşlarda sinir sistemi hastalıklarında etkin olduğu, bazen kodladığı protein üzerinde önemli etkileri olduğunu göstermiştir. Örneğin; DRPLA (Dentatorubropallidoluysian atrofisi) 12. kromozomdaki DRPLA geninde oluşmaktadır. Normal gen 6-35 kopya CAG taşırken, hastalıkta durumunda kopya sayısı 90'lara ulaştığı görülmüştür. Huntington hastalığında (HD) 10-35 olan CAG tekrarlarının 40'ı aştığı bilinmektedir. Fragile X sendromundan (FRAXA) :X kromozomundaki FMR1 genindeki tekrar sayılarının sorumlu olduğu bulunmuştur. Normal FMR1'de 6 ve 53 CGG tekrarı, hastalık durumunda ise 230 tekrar bulunmuştur. Fragile XE mental anormali (FRAXE):X kromozomundaki FMR2 genindeki GCC tekrar sayılarının farklılığından oluştuğu bulunmuştur. Normal FMR1'de 6 ve 35 GCC tekrarı bulunurken, hastalık durumda 200 tekrar bulunmuştur. Myotonik distrofi ise (DM); DMPK geninde oluşmaktadır. Normal DMPK geninin 5-37 CTG kopyası varken, hastalıkta: 50 CTG kopyası bulunmuştur (1).

### **Kopya Sayısı Varyantları (KSV) (Copy Number Variations-CNV)**

Gen kopya sayısı kopya sayısı varyantları, bir diğer önemli varyasyon çeşididir. İnsan genomu anadan ve babadan kalıtsal olarak gelen 6 milyar bç'den oluşmaktadır. Her atadan bir set 23 kromozom gelecek şekilde çekirdekte paketlenmiştir. Her DNA'nın ~30,000 gen kodlaması ve genel olarak teoride genomda bir genden 2 kopya olması beklenirken DNA'nın büyük bir kesiminde genler farklı kopya sayısında bulunmaktadır. KSV bir bireyin genomunda bulunan bir genin kopya sayılarında bulunan farklılıklardan oluşan genetik karakter özelliklerini ifade etmektedir. Farmakogenomik için değerlidir.

Bazı genlerin kopya sayılarındaki farklılıkların genetik varyasyona katkıda bulunduğu İGP'nin tamamlanmasıyla ortaya konmuştur. Bazı insanlarda yalnızca bir kromozomda bazı genlerin delesyonu görülürken, bazılarında aynı genlerin birçok kopyası olduğu İGP ile ortaya konmuştur (17). Diğer genetik varyasyonlar gibi, bazı gen KSV'lerinin hastalığa yatkınlık ya da direnç gösterme gibi özelliklerini beraberinde getirdiği görülmüştür. Gen kopya sayıları kanser hücrelerinde yüksek olabilmektedir. Örneğin, İnsanda CCL3L1 geninin yüksek kopya sayısının HIV enfeksiyonuna yakalanmada daha az duyarlı olduğu gösterilmiştir.

### 2.3. Varyasyonların Kullanım Alanları

Bir genetik hastalığın mekanizması bilinmese de hastalığı gösteren yeterli sayıda birey varsa, genetik hastalığa neden olan lokus bağlantı analizi ile haritalanabilir (13). Yapılan çalışmalar kuşaklarboyu genetik göstergenin kalıtımının taranması yönteminin hastalıktan sorumlu olan genlerin bulunmasında güçlü bir araç olduğu göstermiştir.

1980'lerde tanımlanan ilk olarak tanımlanan RFLP'ler insan kromozomlarının genetik bağlantı haritalarının oluşturulmasında kullanılan birinci nesil genetik varyasyonlardır. White ve arkadaşları bu polimorfik belirteçleri kullanarak insan kromozomlarının genetik haritasını oluşturmuş ve DNA belirteçlerini araştırmacıların kullanımına sunmuşlardır (18).

Polimorfik DNA belirteci kromozomun maternal veya paternal kökeninin ayırt edilmesine olanak sağlamaktadır. Bu nedenle polimorfik belirtecin hastalık kalıtımıyla birlikte seyri araştırılmaktadır. Örneğin Gusella ve çalışma arkadaşları 4. kromozomun kısa kolundaki genetik lokusla Huntington hastalığı arasındaki ilişkiyi saptamıştır (19). 1980'lerin sonuna doğru polimorfik belirteçlerin artmasıyla diğer yaygın genetik hastalıklardan sorumlu olan başka genler de bulunmuştur.

1980'lerin başında PCR'ın geliştirilmesi RFLP (Southern teknolojisi) analizlerini yerini mikrosatelit (PCR teknolojisi) analizlerine bırakmıştır. Çünkü genomda mikrosatelit lokusları 100,000'den fazla genomik bölgeyi kapsamaktadır. Mikrosatelit belirteçlerle yapılan çalışmalarda çok az miktarda DNA'ya ihtiyaç duyulması ve yüksek işlem hacimli sistemlere adaptasyon kolaylığından dolayı, mikrosatelitler bağlantı analizlerinde ve popülasyon genetiğinde başarılı şekilde kullanılmaya başlanmıştır. Mikrosatelitlerin heterozigotluk oranları fazla, paternal alleler için büyük oranda ayırt edici ve bağlantı analizlerinde bilgilendirici olmaktadır.

İkinci nesil varyasyonlar (VNTR) sayesinde çok fazla sayıda genetik hastalık haritalanıp, hastalıktan sorumlu olan gen izole edilmiştir. 1990'larda STRP çalışmaları (üçüncü nesil) ve 2000'li yıllara gelindiğinde TNP belirteç (dördüncü nesil) çalışmaları yoğunluk kazanmıştır. TNP'ler genomda ortak olan bir varyasyon çeşididir. Her 500-1000 bç de 1 tane TNP olması beklenirken İGP ile ~300 bç de 1 tane olduğu ortaya konmuştur. Uluslararası TNP veri tabanlarının geliştirilmesi ve yüksek işlem hacimli TNP tipleme

platformlarının devreye girmesiyle genom boyunca yapılan arařtırmalar ilerlemiş, yaygın hastalıklarla ve ilaca cevapla TNP iliřkilendirme alıřmaları hız kazanmıřtır (20).

Hastalıkları oluřturan genlerin kromozom yerleřimlerinin bulunmasına ynelik alıřmalara “hastalık geni haritalanması” denir. Hastalıđın fenotipik zellikleri hekim tarafından tam olarak ortaya konduktan sonra aile ađacı analizleri yapılarak hastalıkla ilgili genetik bir model geliřtirilmeye alıřılmaktadır.

Hastalık modellemesi iin gerekleřtirilen molekler analizler sonrasında seilen istatistiksel yaklařımlar da nemlidir. Arařtırmacılar analizlerde bađlantı ve asosiyasyon olmak zere iki tip istatistiki yaklařım kullanmaktadırlar.

### **2.3.1. Bađlantı (Linkage)**

*Model bađımlı analizler (Parametrik Analizler):* Kromozom yerleřimi kesin olarak bilinen bir genetik gsterge kullanarak bu gstergenin kuřaklar boyunca birlikte kalıtılıp kalıtılmadıđını arařtırılmaktadır. Mayozda eř kromozomlar karřı karřıya gelerek birbirleriyle para deđiřimine uđramaktadır. Bu deđiřimler sonucu oluřan yavru kromozomlar ana ve baba kromozomuna tamamen benzeyen bir genetik yapıya sahip olabildikleri gibi her ikisine de benzemeyen yeni bir genetik yapı ierebilmektedir (rekombinasyon). Aynı kromozom zerinde birbirine ok yakın yerleřimli genlerin ya da polimorfik belirtelerin mayoz sırasında para deđiřimine uđraması (crossing over) ve birbirinden bađımsız olarak bir sonraki kuřađa kalıtılması zor olduđundan birbirine yakın yerleřimli genler kuřaklar boyunca bir arada kalıtılırken yeni yapılanmalar (rekombinant rnler) ortaya ıkmamaktadır. Yani bir kromozom zerinde birbirine ok yakın yerleřimli genler mayozda yeniden řekillenmeden yavru kuřađa daima birlikte aktarılmaktadırlar. Bu olaya bađlantılı kalıtım (linkage) adı verilmektedir. Aynı kromozom zerinde fakat birbirine uzak olan ya da farklı kromozomlarda bulunan genlerin ise yavru kuřaklarda bir arada grlme olasılıkları tamamen řansa bađlıdır (iki farklı geni alma ya da almama olasılıđı = %50 řans). Arařtırmacılar bu analizin uygulanabilmesi iin birlikte kalıtımın birkaç kuřak boyunca test edilebileceđi geniř ailelere ihtiya duymaktadır. Belirte allel ile iliđilenilen niteliđin kuřaklar arası beraber kalıtımı test edildiđi iin her ailede aynı allel ya da aynı haplotipin gzlenmesi beklenmemektedir (21).

### 2.3.2. Asosiyasyon

*Modelden bağımsız analizler (Parametrik olmayan analizler):* Model bağımlı analizler psikiyatrik hastalıklar gibi hastalık modellemesinin tam yapılamadığı durumlarda başarılı olamayabilmektedir. Bu durumlarda modelden bağımsız analizler (ilişkilendirme=association) kullanılmaktadır. Bu tip analizlerde kalıtım kalıbından bağımsız olarak hastalık ile genetik gösterge arasında anlamlı bir ilişki bulunmaya çalışılmaktadır. İlişkilendirme analizlerinde hastalıkla genetik göstergenin kuşaklararası birlikte kalıtılması araştırılmamaktadır. Bunun yerine hastalığın ortaya çıkmasında rolü olabilecek genler belirlenmekte ve bu aday genler içinde yer alan polimorfik DNA değişiklikleri saptanmakta (özellikle TNP'ler) ve bu değişikliği oluşturan allel ile hastalık alleli arasında ilişki olup olmadığı araştırılmaktadır.

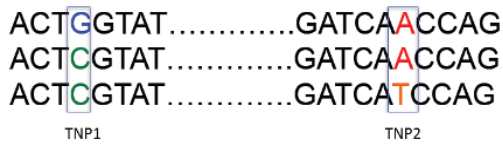
İlişkilendirme çalışmalarında bağlantı analizlerinden farklı olarak genetik yerleşim ile hastalığın birlikte kalıtılması değil, ilgili gene ait bir allelin varlığı ya da yokluğu araştırılmaktadır. Genetik gösterge alleleri ile hastalığın birlikte kalıtılması durumu “tam bağlantı” olarak tanımlanmıştır (Linkage). Hastalık için tek bir allelin seçilimi ise tam ilişki durumunda görülmektedir; örneğin daima 2 numaralı allel veya daima Guanin bazının Adenine değiştiği durum gibi. Bu durum “bağlantı dengesizliği” (Linkage Disequilibrium-LD) olarak adlandırılmaktadır ve ilişkilendirme çalışmalarının temelini oluşturmaktadır. Tam bağlantı ya da bağlantı dengesizliğinde; bir allelin yakınındaki lokusla arasında var olan ve rastgele olmayan bir ilişki olduğundan tüm bireylerde aynı allel ya da aynı haplotipin seçilimi aranmaktadır. LD; yeniden şekillenme olmaksızın (rekombinasyon) TNP allel çiftleri için tesadüfi ilişkilendirilme ölçüsüdür ve  $D'$ ,  $r^2$ , LOD ile ölçülmektedir. Asosiyasyon çalışmalarını yürütebilmek için aday olduğu düşünülen bir gene ve bu genin içindeki TNP dağılımlarına ihtiyaç duyulmaktadır. LD, gen haritalanmasında, kompleks hastalıkların haritalanmasında ve genom boyunca ilişkilendirme çalışmalarında önemli bir araçtır. Aynı zamanda popülasyonların evrimini göstermektedir. LD çalışmaları biyolojik rekombinasyon konusunda daha fazla bilgi edinmemize de ışık tutmaktadır (21).

Toplum tabanlı vaka-kontrol tipi çalışmalarda çok sayıda belirteç üzerinden LD varsayımının test edilmesindeki zorluk genellikle ana ve babalar mevcut olmadığından hangi allelin hangi ebeveynden geçtiğinin bilinmemesinden kaynaklanmaktadır. Aile çalışmalarında olduğu kadar rahatlıkla haplotip oluşturulamadığı için tahmini haplotipler

üzerinden en olası haplotip bulunmaya çalışılmakta ve daha sonra bu haplotipin ilgili nitelik ile ilişkisinin anlamlı olup olmadığı test edilmektedir. Özellikle HapMap projesi ile popülasyon polimorfizmlerine özgü haplotip bloklarının ve LD haritalarının çıkarılmasına yönelik ortaya konan veriler araştırmacılar için faydalı olmaktadır. Popülasyon verileri kullanılarak yapılan LD hesapları birbiriyle tam ilişkili TNP haplotiplerin seçilimini sağlamaktadır ve bir haplotipi tanımlayan minimum TNP seti tag TNP olarak adlandırılmaktadır,  $r^2=1$  olduğunda iki TNP'in bolluk içerisinde olduğuna bir TNP'in varlığının diğer TNP'in varlığına işaret ettiği durum olarak tanımlanmaktadır.

LD analizleri için kullanılan farklı bilgisayar programları (örneğin Haploview, unphased vs.) bulunmaktadır. Çalışma modellerinde vaka-kontrol yerine ana-baba-çocuk üçlüleri; kardeş çiftleri gibi modeller kullanmak daha kesin haplotipler oluşturmaya yardımcı olabilmektedir (22).

### Bağlantı Dengesizliği aşağıdaki şekilde hesaplanmaktadır (LD)



Şekil 2.3. TNP1 ve TNP2

İki TNP için (Şekil 2.3) Çizelge 2.2'deki alleler var ise

Çizelge 2.2. TNP1 ve TNP2 için Alleler

Alleller	TNP1	TNP2
Allel 1	G	A
Allel 2	C	T

1. aşamada allel frekansları hesaplanır

Eğer  $p_1$  ve  $p_2$  = TNP1 allel frekansları

$q_1$  ve  $q_2$  = TNP2 allel frekansları ise tablo gösterimi Çizelge 2.3'deki gibidir



**Çizelge 2.3.** TNP1 ve TNP2 Allel Frekansları Tablo Gösterimi

SNP1		SNP2	
Allele	Frequency	Allele	Frequency
G	p1	A	q1
C	p2	T	q2

2. aşamada haplotip frekansları hesaplanır

Örnekteki 2 allelli 2 TNP'nin bütün olası haplotipleri **Çizelge 2.4'**deki gibidir

**Çizelge 2.4.** TNP1 ve TNP2 Haplotipleri

	Allele	TNP2	
		A	T
TNP1	G	GA	GT
	C	CA	CT

Haplotip frekanslarında **Çizelge 2.5'**ki gibidir

**Çizelge 2.5.** TNP1 ve TNP2 Haplotip Frekansları

Haplotype	Frequency	Haplotype	Frequency
GA	p11	GT	p12
CA	p21	CT	q22

3. aşamada Bağlantı dengesi hesaplanır (*Linkage equilibrium*)

Eğer haplotip frekansları ilişkili olduğu ürünün allel frekansına **Çizelge 2.6**'da gösterildiği gibi eşit ise, o lokus dengedir.

**Çizelge 2.6.** Dengede olma durumu

Haplotip frekansı	Eşittir	Ürünün Allelik Frekansı
p11	=	p1q1
p12	=	p1q2
p21	=	p2q1
p22	=	p2q2

4. aşamada *Bağlantı Dengesizliği hesaplanır (Linkage disequilibrium)*

Denge halinde her haplotip için LD **Çizelge 2.7**'deki şekilde hesaplanmıştır

Beklenen allelik frekansı-gözlenen haplotip frekans sapması (D)

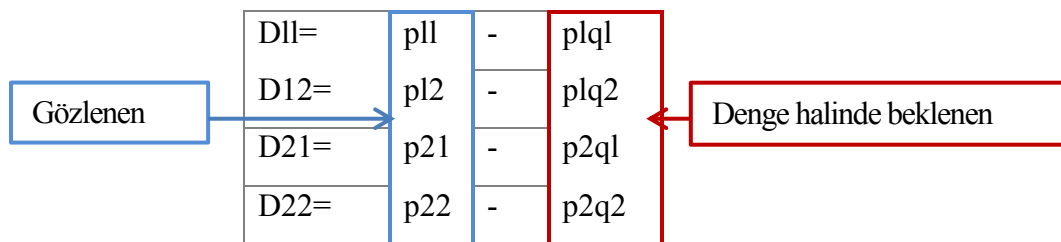
**Çizelge 2.7.** LD

SNP1	SNP2		
	1	2	
1	p1q1+D	p1q2-D	p1
2	p2q1-D	p2q2+D	P2
	q1	q2	1

Yukarıdaki şekilde D yi çözersek aşağıdaki sonucu alırız

D çözüldüğünde: **Çizelge 2.8**'deki sonuç alınır:

**Çizelge 2.8.** D Çözümü



### 5. aşamada D cinsinden Bağlantı Dengesizliği

Bağlantı dengesizliği genellikle D ile gösterilir ve  $D = p_{11}p_{22} - p_{12}p_{21}$  eşittir ve **Denklem 3-Denklem 5** çözerek ispatlanabilir.

$$a) \quad p_{11}p_{22} = (p_1q_1 + D)(p_2q_2 + D) = p_1q_1p_2q_2 + p_1q_1D + p_2q_2D + D^2$$

**Denklem 3**

$$b) \quad p_{12}p_{21} = (p_1q_2 - D)(p_2q_1 - D) = p_1q_1p_2q_2 - p_2q_1D - p_1q_2D + D^2$$

**Denklem 4**

$$c) \quad p_{11}p_{22} - p_{12}p_{21} = (p_1q_1p_2q_2 + p_1q_1D + p_2q_2D + D^2) - (p_1q_1p_2q_2 - p_2q_1D - p_1q_2D + D^2) = D(p_1q_1 + p_2q_1 + p_2q_2 + p_1q_2) = Dx(1) = D \quad \text{Denklem 5}$$

### Bağlantı Dengesi halinde D tahmini

**Denklem 6-Denklem 11**'da olduğu gibi  $p_1$  ve  $q_1$  'in allel frekanslarının ikisinde 0,5 ise, popülasyonda denge sadece Ab ve aB var ise oluşur.

$$p_{11} = p_1q_1 = 0,5 \times 0,5 = 0,25$$

**Denklem 6**

$$p_{22} = p_2q_2 = 0,5 \times 0,5 = 0,25$$

**Denklem 7**

$$p_{12} = p_1q_2 = 0,5 \times 0,5 = 0,25$$

**Denklem 8**

$$p_{21} = p_2q_1 = 0,5 \times 0,5 = 0,25$$

**Denklem 9**

$$D = (p_{11})(p_{22}) - (p_{12})(p_{21})$$

**Denklem 10**

$$D = (0,25)(0,25) - (0,25)(0,25) = 0$$

**Denklem 11**

### **D'nin Bağlantı dengesizliği durumunda tahmini**

Eğer **Denklem 12-Denklem 17**'de olduğu gibi  $p_1$  ve  $q_1$  'in her ikisinde de allel frekansı 0,5 ise bütün lokuslarda eşit allel frekanslı rastlantısal olmayan tam bağlantı durumu vardır (popülasyonda sadece AB ve ab var ise)

$$p_{11} = p_1q_1 + D = 0,25 + D = 0,5$$

**Denklem 12**

$$p_{22} = p_2q_2 + D = 0,25 + D = 0,5$$

**Denklem 13**

$$p_{12} = p_1q_2 - D = 0,25 - D = 0$$

**Denklem 14**

$$p_{21} = p_2q_1 - D = 0,25 - D = 0$$

**Denklem 15**

$$D = (p_{11})(p_{22}) - (p_{12})(p_{21})$$

**Denklem 16**

$$D = (0,5)(0,5) - (0)(0) = 0,25$$

**Denklem 17**

### **D'nin standardizasyonu**

Bazen iki lokusdaki allel frekanslarına bağlı olarak D negatif olabilir fakat gametik frekanslar negatif olamaz. Bunun üstesinden gelmek için standardizasyon metodu uygulanır.

Yaygın standardizasyon metodunda dengesizliğin (Disequilibrium-D) göreceli ölçüsü maksimum kullanımı ile karşılaştırılır (**Denklem 18-Denklem 20**).

$$D' = D / D_{max}$$

**Denklem 18**

*D pozitif olduğunda*

$$D_{\max} = \min [ (p_1q_2) \text{ veya } (p_2q_1) ]$$

**Denklem 19**

*D negatif olduğunda*

$$D_{\max} = \min [ (p_1q_1) \text{ veya } (p_2q_2) ]$$

**Denklem 20**

Bu standardizasyon D değerini 0-1 arasında değiştirir.

**LD ölçüsü olarak correlation coefficient**

Lokuslar arasında LD'yi hesaplamak için bir diğer yaygın kullanım Pearson coefficient correlation (r) (**Denklem 21**)

$$r = \frac{D}{(p_1q_2p_1q_2)^{1/2}}$$

**Denklem 21**

Fakat coefficient correlation'ın karesi ( $r^2$ ) genellikle negatifliği kaldırmak için kullanılır

**LD anlamlılığının test edilmesi**

LD'nin anlamlı olup olmadığını test etmek için  $\chi^2$  testi uygulanır.

$$\chi^2 = \sum \frac{(\text{obs} - \text{exp})^2}{\text{exp}}$$

**Denklem 22**

**Denklem 22** aleller arasında beklenen rastlantısal ilişkilerdir. Fakat r ki kare testi için kullanılabilir. **Denklem 23**. Burada N örnekteki kromozom numarasıdır.

$$\chi^2 = r^2N$$

**Denklem 23**

### **2.3.3. TNP Polimorfizmleri ve Farmakogenomik**

İnsan genom dizilemesinin 2003 yılında tamamlanacağını duyurulmasından kısa süre sonra birçok hastalığın tarihte kalacağı düşünölmekteydi. İnsan genomunu anlamada dizilemenin tamamlamasının sadece bir başlangıç olduđu ve genomdaki karmaşık mekanizmaları anlamanın bu kadar basit olmayacağı göröldü. Daha sonra insan genomu üzerinde yapılan çalışmalar ise bilim dünyasına her geçen gün birçok yeni bilgi ve kavram kazandırdı. Bu yeni kavramlardan biri olan farmakogenomik; kişilerin genetik farklılıklarından ötürü ilaçların farmakokinetik ve farmakodinamik özelliklerinin farklı olması ve buna bağılı olarak da bireylerin ilaçlara farklı şekillerde tepki vermesidir. Bu genetik farklılıklar ilaçların emiliminden, taşınımına, metabolize olmasına, dokulardaki dağılımına ve son organlardaki reseptör duyarlılıklarının farklarına kadar geniş bir çerçeveyi kapsamaktadır.

Farmakogenomik ilk ürününü Stanford Üniversitesi Tıp Fakültesi, Loyola Üniversitesi Tıp Merkezi ve Washington Üniversitesi Tıp Fakültesi ve St. Louis gibi yerlerdeki araştırmacıların ortak çalışmaları sonucunda vermiştir. Bu çalışma ile Afrika kökenli Amerikalılarda yüksek tansiyondan sorumlu gen bölgelerin bulunarak, sonuçları Şubat 2005’de Nature Genetics’de yayınlanmıştır.

### **2.3.4. Genetik Antropoloji**

Gen akışının az olduđu izole haldeki popölasyonlarla, fazla olan popölasyonların karşılaştırılması, insan evrimi ve göçlerinin anlaşılmasına ışık tutmaktadır. Soyağaçlarındaki mutasyonların incelenmesi, insana çok yakın, az yakın ya da uzak diđer canlıların genomlarının incelendiđi çalışmalar sonucunda yapılacak araştırmalar yaşamın gizinin ve canlıların moleküler evriminin anlaşılmasına büyük katkıda bulunmaktadır.

İnsan bilimi olan Antropoloji insan anlamına gelen Anthropos ile bilim anlamındaki logos kelimelerinden oluşmuştur. İnsanlar arasındaki benzerlikleri ve farklılıkları göz önüne alarak insanları karşılaştıran bir görüşle incelemektedir. Antropolojinin ilgisi grup halinde yaşayan insanlar, yaptıkları ve davranışları olup birey olarak insanla ilgilenmemektedir. Genetik antropoloji ise insanoğlunun nereden geldiđi ve nasıl geldiđi sorularına DNA teknolojilerini de kullanarak cevap arayan yeni bir bilim dalıdır. Genetik antropoloji

insanlığın atasal göç tarihine ulaşmak için DNA bilgisi ile fiziksel kanıtları birleştirmektedir.

Bireylerin birbirlerinden farklılığının nedeni genetik varyasyonlar olup anne ve babadan çocuklara kalıtlanmaktadır. Varyasyonlar bazen mutasyona uğrayarak kalıtlanabildikleri gibi, atasal da olabilmektedir. Bu atasal gen parçaları, popülasyon kökenini gösteren ve kalıtımın izleri olan genetik belirteçleri taşımaktadır. Bu nedenle bireyler atalarının doğduğu topraklarda yaşıyor olmasalar da genomlarındaki genetik belirteçler (ait oldukları coğrafi bölgeye özel taşıdıkları) kökenlerini araştırabilmemize olanak tanımaktadır.

Eski çağlardan günümüze aktarılan genetik materyal atalarımıza ait birçok genetik özelliği barındırmaktadır. Bu nedenle insanlığın son ortak atası DNA incelemesiyle belirlenebilir. Çünkü ortak ataya ait bir DNA parçası modern insan DNA'sının bir bölümünde yer almaktadır ve bu DNA parçası soydan soya değişmeden aktarılmaktadır. Konu ile ilgili araştırmalar modern insanın Sahraaltı Afrika'da ortaya çıktığına, 65,000 yıl önce göç ederek önce güney Asya, Çin ve Java'ya daha sonra da Avrupa'ya yerleştiğine işaret etmektedir ve halen bütün insanların atalarına ait birer DNA parçasını taşıdığı görülmektedir.

Bir nesilden diğerine DNA geçerken, eş kromozomlar mayozda karşı karşıya gelerek birbirleriyle parça değişimine uğramaktadır. Bu değişimler sonucu oluşan yavru kromozomlar ana ve baba kromozomuna tamamen benzeyen bir genetik yapıya sahip olabildikleri gibi her ikisine de benzemeyen yeni bir genetik yapıda içerebilmektedir (rekombinasyon). DNA'daki bazı özel kısımlar ana-babadan çocuğa geçerken değişikliğe uğramamaktadır. Bu parçalardan bir tanesi Y kromozomu üzerinde taşınmakta olup babadan oğula geçmektedir (paternal kalıtım). Diğer bir parça olan mitokondrial DNA (mtDNA) ise birkaç istisna dışında anneden çocuğa geçmektedir (maternal kalıtım). Y kromozomu başka bir DNA parçasıyla karışmadığından erkeğin kökenini takip etmek için bir nevi genetik soyadıdır. Aynı şekilde mtDNA da ana kökenini takip etmeye olanak sağlamaktadır (23) (24).

Hem Y kromozomu hem de mtDNA genetik belirteç olarak kalıtılan zararsız mutasyonların odağı olabilmektedir. Birkaç nesil sonra bölgede yaşayan çoğu erkek ve kadınlar bu belirleyici genetik belirteçleri taşıyor ve yaşadıkları bölgeyi terk ettikten sonra da bu genetik belirteçleri genomlarında taşımaya devam ederler. Günümüzde bilim insanları

farklı yerli popülasyonlarda görülen genetik belirteçlerin nerede oluştuğunu araştırmaktadır. DNA üzerinde var olan her belirteç o kişinin atalarına ait lokasyon ve göç yollarını işaret etmektedir. Örneğin İngilizlerin ~%70'i, İspanyolların ~%95'i ve İrlandalıların ~%95'inde ayırt edici M173 Y kromozomu mutasyonu bulunmaktadır. Çalışmalar son buzul çağı sonrasında ortak atanın İspanyanın kuzeyinden İngiltere ve İrlanda'ya göç ettiklerine işaret etmektedir (24).

## 2.4. HapMap

İnsan genetik kodlarını oluşturan 3 milyar harfin dizilimlerini bir haritada birleştiren İGP, herhangi iki insanın, birbirine genetik olarak % 99,9 benzediğini ortaya koymuştur. Yani insanlar % 99,9 oranında aynı özellikleri taşısa da % 0,1'lik farklılık insanların astım, diyabet, hafıza kaybı, kalp hastalıkları ya da kanser gibi Mendel kalıtım modeli göstermeyen hastalıklara yakalanmasına neden olmaktadır. Günümüzde insan genomundaki bu genetik farklılığın % 90'dan fazlasını TNP'lerin meydana getirdiği ve insan genomundaki en yaygın DNA farklılığını TNP'lerin oluşturduğu bilinmektedir. Bu nedenle İGP'den sonra başlatılan projelerin önemli bir hedefi genom bilgileri ile klinik bilgiler arasında bağ kurabilmek ve bu bağı hastalıklar ve genetik diziler (genotip-fenotip) arasındaki ilişkinin daha iyi anlaşılmasında kullanmak olmuştur. Genotip-Fenotip ilişkileri, popülasyon yapısı veya tabakalanması (stratification) doğru olarak ortaya konmadığı sürece yanlış fikir verebildiğinden, tüm genom ilişkilendirme (Whole Genome Association Studies- WGAS) çalışmalarının ilgililenilen fenotipik varyasyondan sorumlu olan genetik varyantların bulunmasında güçlü bir araç olduğu görülmüştür. Bu nedenle 2002 Ekim ayında İngiltere, ABD, Kanada, Japonya, Nijerya ve Çin'den 200 bilim adamının çalıştığı, bir özel-kamu ortaklığı olan Uluslararası HapMap Konsorsiyumu, "HapMap" denilen genetik varyasyonların belirleyicisi olan TNP'leri gösteren daha detaylı bir harita hazırlamaya başlamıştır. Bu amaca ulaşmak için geliştirilen projelerden biri olan HapMap ile insan genomunun haplotip haritasının çıkartılması ve bu sayede DNA dizi varyasyonlarındaki ortak modellerin tanımlanmasını amaçlanmıştır.

HapMap Projesi 2002 Ekim ayında genetik çalışmalardaki klinik fenotiplere ışık tutabilmek için tüm genomdaki yaygın dizi varyasyonlarına ait veri tabanının oluşturulmasıyla başlamıştır. Böylelikle genomun insan bireyleri arasındaki varyasyonunu daha iyi anlayarak, varyasyon, hastalık ve hastalık hassasiyetleri arasındaki ilişkilerin belirlenmesi hedeflenmiştir. Yeni haritada TNP'lerin bölge bölge gruplandırıldığı



(haplotipler) ve enformasyon blokları olarak insandan insana kalıtım yoluyla geçtiği görülmüştür (25).

HapMap Projesi: (1) İnsan genom dizisinin kullanılabilir olması, (2) yaygın TNP'lerin veri tabanı (bu proje ile genişletildi), (3) insan LD'sinin anlaşılması, (4) ucuz, yüksek işlem hacimli, doğru sonuç veren genotipleme teknolojilerin geliştirilmesi, (5) elde edilen verinin depolanması ve araştırmacıların erişimi için web tabanlı araçlar, (6) etik ve kültürel yasal çerçevenin oluşumunun tamamlanmasıyla gerçekleştirilmesi mümkün olabilmektedir.

Proje 3 aşamadan oluşmuştur (**Çizelge 2.9**): 1. aşamada; MAF 0,05 veya daha büyük olan 269 DNA örneğinden, her 5 kilo bazda (kb) en az 1 yaygın TNP genotiplenmesi amaçlanmıştır. 500 kb uzunluğunda 10 DNA bölgesi temsili olarak ENCODE Projesinden seçilerek, örneklerdeki yaygın ve nadir TNP'ler 269 DNA örneğinde genotiplenmiştir. 1 aşamanın sonucunda yaklaşık 1 milyon TNP genotipleme yapılmıştır (26).

2. aşamada ise 1. aşamadaki toplam 270 DNA örnek kullanılarak genom boyunca daha yüksek çözünürlükte TNP analizi gerçekleştirilmiştir. 2. fazda kilobaz başına ortalama 1.14 polimorfik TNP genotiplenmiş (ortalama boşlukları 875 baz çiftidir (bç)) ve birleştirilen genomun % 98,6'sının en yakın polimorfik TNP'e uzaklığı 5 kb olmuştur. Genotiplenen bütün verilere [www.hapmap.org](http://www.hapmap.org) ve [www.ncbi.nlm.nih.gov/SNP](http://www.ncbi.nlm.nih.gov/SNP) sitelerinden ulaşılabilmektedir. HapMap Faz 2, Faz 1'den sadece TNP aralıkları bakımından değil, aynı zamanda minör allel frekans dağılımı ve bağlantı dengesizliği modeli bakımından da farklılık göstermektedir. 2. Faz sonucunda yaklaşık 3,1 milyon TNP genotipleme tamamlanmıştır (27).

3. aşamada ise HapMap koleksiyonu orijinal 270 örneği de içerecek şekilde 11 farklı popülasyondan ve 1301 örnekten oluşturulmuştur. Araştırmanın ilk sonuçları Mayıs 2008'de (Draft Release 1, Eylül 2008'de Draft Rel. 2) yayınlanmıştır (28).

HapMap verileri ilk olarak medikal genetik çalışmaların, analiz ve tasarlanmasına rehberlik etmek için üretilmiştir. Başlangıçta genom boyunca ilişkilendirme çalışmalarının analizi ve tasarlanması için metod geliştirilmesinde ana yapıyı oluşturmuştur. Halen çok yönlü karmaşık hastalıklarda yer alan yeni lokusların tanımlanmasında kullanılmaktadır. HapMap verileri yapısal varyasyonun devamlılığını sağlayan, yeniden şekillenme bölgelerinin belirlenmesi sayesinde yeniden şekillenme (rekombinasyon) olaylarına da farklı bir bakış açısı getirmiştir. Aynı zamanda doğal popülasyonlarda evrimsel güçlerin

neden olduğu genom boyunca oluşan varyasyonların kaynaklarının araştırılması popülasyon genetikçileri için önem arz etmektedir.

### Çizelge 2.9. HapMap Projesi Aşamaları

	Faz 1 <sup>4</sup>	Faz 2 <sup>5</sup>	Faz 3 <sup>6</sup>
<b>Örnekler ve Popülasyon Çeşitleri</b>	269 örnek (4 grup*)	270 örnek (4 grup**)	1301 örnek (11 grup***)
<b>Genotipleme merkezleri</b>	HapMap International Consortium	Perlegen <a href="http://genome.perlegen.com">http://genome.perlegen.com</a>	Broad & Sanger <a href="http://www.sanger.ac.uk/humgen/hapmap3/">http://www.sanger.ac.uk/humgen/hapmap3/</a>
<b>TNP</b>	~1 M	~3.1 M (faz I+II)	(Affy 6.0 & Illumina 1M)
<b>Referans</b>	Nature 437, 1299–1320 (2005).	Nature 449; 851-861 (2007)	Draft Rel. 1 (Mayıs 2008) Draft Rel. 2 (Eylül 2008) Nature 467, 52-58. (2010)

Uluslararası HapMap Projesinin amacı; genomda DNA dizi varyasyonuna özgü ortak desenlerin korelasyon ve frekans hesapları yardımıyla, Afrika, Asya, ABD ve Avrupalı örnekler üzerinde araştırılması olmuştur. Proje, indirek asosiyasyon çalışmaları için genomda herhangi bir fonksiyonel gen, aile bazlı bağlantı analizlerinin ortaya koyduğu herhangi bir bölge veya hastalık risk faktörü için tüm genom taramasına araçlar sağlamaktadır.

İndirek asosiyasyon yaklaşımında; genomda bir set dizi varyantı, belirli genomik bölge ve hastalık ile (belirtecin genomda fonksiyonel etkisi olması gerekmez) ilişkilendirilerek genetik marker olarak kullanılmaktadır. Bu yaklaşım ile hastalıkların saptanması direk yaklaşımdan daha etkin olmaktadır.

4 1. Faz: (1) Nijerya, Ibida Yoruba - 90 birey 30 ana-baba-çocuk üçlüsü (2) USA, Utah - 90 birey (30 üçlü) [the Centre d'Etude du Polymorphisme Humain collection (CEU)] (3) 45 Çin, Beijing - Han Çinlileri (CHB) (4) 44 Japonya, Tokyo'dan Japonlar (JPT)

5 2. Faz: (1) Nijerya, Ibida Yoruban - 90 birey 30 ana-baba-çocuk üçlüsü (2) USA - Utah 90 birey (30 üçlü) (CEU'dan) (3) 45 birbiriyle ilişkisiz Çin, Beijing'deki, Han Çinlileri (CHB) (4) 45 birbiriyle ilişkisiz Japonya, Tokyo'dan Japonlar (JPT)

6 3. Faz: (1)ASW (A): Güneybatı ABD'de Afrika kökenliler 90 birey (2) CEU: CEPH koleksiyonundan Kuzey ve Batı Avrupa kökenli Utah sakinlerinden 180 birey (3) CHB: Çin, Beijing'deki Han Çinlilerinden 90 birey (4) CHD: Denver Colorado metropolitindeki Çinlilerden 100 birey (5)GIH: Teksas, Houston'da yaşayan Gujarati Hintlilerden 100 birey (6) JPT: Japonya, Tokyo Çinlilerden 91 birey (7) LWK: Webuye, Kenya'da yaşayan Luhya kabilesinden 100 birey (8) MEX: Kaliforniya, Los Angeles'da Meksika kökenli 90 birey (9) MKK: Kinyawa, Kenya'da yaşayan Maasai'lerden 180 birey (10) TSI: İtalya'daki Toskanalılarından 100 birey (11)YRI: Nijerya, Ibida'da yaşayan Yorubalılarından (Batı Afrika) 180 birey

## 2.5. HapMap Projesi ve Diğer Popülasyonlar

Genom boyunca ilişkilendirme çalışmalarına, kompleks hastalıkların haritalanmasına ve evrim çalışmalarına destek olmak amacı ile yürütülen HapMap projesi sonucunda; 1. ve 2. aşamada 4 popülasyon 270 bireyin, 3. aşamada ise 11 popülasyon 1301 bireyin LD, tag TNP, Phased Haplotipler, Genotip verileri ve frekansları ortaya konmuştur.

HapMap popülasyonlarının diğer popülasyon gruplarını ne derece temsil ettiği hala bilinmediğinden karşılaştırmaları yapabilmek adına başta Asya olmak üzere birçok toplum kendi genom varyasyon çalışmalarını başlatmıştır.

### 2.5.1. Çin İnsanı Genom Varyasyon Projesi

Çin 2007 sonu itibariyle 1 milyar 320 milyon insanıyla dünyanın en fazla nüfus yoğunluğuna sahip olan ülkesidir (29). CGHD Projesiyle insan türünün beşte birini oluşturan ve 56 alt popülasyon içeren Çin popülasyonundaki tabakalanma, genetik belirteçlerin bir çeşidi olan, mikrosatellitlerle ortaya konmak istenmiştir. Bu 56 alt popülasyonun birincisi olan Han grubu çoğunluğu oluşturmakta olup toplam nüfusun bir milyar 100 milyonunu kapsamaktadır. 55 azınlık grubu ise kalan 100 milyon kişi ile temsil edilmektedir (1998). Azınlıklar bütün Çine ama özellikle güney bölümlere dağılmıştır ve çalışma grubu ise azınlıkların yarısını oluşturmaktadır.

CGHDP Projesinde, 14 Doğu Asya popülasyonunun ağaç yapısının yeniden yapılandırılması amacıyla (dünyanın kalanını temsilen standart set olarak 11 popülasyonla) 30 mikrosatelit ve arkadaşları tarafından test edilmiştir. Aynı mikrosatellitlerin 15'lik bir alt seti 32 doğu Asya popülasyonundan ikinci bir ağaç oluşturmak için kullanılmıştır. Bunlar ilk 14 taneyi içermekte olup, standart setle karşılaştırılmışlardır.

Karşılaştırmalı araştırmalar sonucunda büyük genetik farklılaşmalara işaret edilmiştir. Doğu Asya popülasyonları beklendiği üzere küçük genetik farklılaşmalar gösterilmiştir. Coğrafi yakınlıktan dolayı farklı popülasyonlardan fazla gen akışı sağlanmıştır. Bu nedenle her iki ağaç yapısında da Doğu Asya popülasyon kümesi beraber kalmıştır. Bu grubun en yakın genetik komşuları ise Amerika yerlileri olarak saptanmıştır. Avusturalya yerlileri ve Yeni Gineliler tarafından oluşturulan küçük bir küme, genetik olarak biraz daha az yakın olanları oluşturmuştur (16).

### **2.5.2. Hindistan Varyasyon Projesi (IGVDB)**

1 milyar 148 milyona yaklaşan tahmini (2008 Haziran) nüfusuyla dünyanın en kalabalık ikinci ülkesi olan Hindistan popülasyonu, kabile içerisinde evlenen yüzlerce grupla birlikte 4693 komüniteden oluşan, 325 dilin ve 25 alfabenin kullanıldığı bir çeşitliliğe sahiptir. IGVDB, bu çeşitliliğin genom düzeyindeki etkilerinin araştırmak ve bazı kompleks hastalıklara ilişkin sorulara farmakogenomik açılarından cevap bulabilmek için Hindistan Genom Çeşitlilik (IGV) Konsorsiyumu tarafından yapılan bir çalışmadır. IGV Projesi tahmini ilaç seçimi için Hindistan alt popülasyonuna ait gen tekrar bölgelerini ve TNP'leri kullanarak bilgilendirici belirteçler oluşturmayı amaçlamıştır.

Proje başlangıcında amaçlara uygun olarak farklı alt popülasyonlardan 15.000 bireye ait örnekler toplanarak, yaygın hastalıklar ve ilaca cevapla ilişkili ~1.000 gen seçilmiş ve çalışmalara faydalı olabilmesi açısından gen başına yaklaşık 5-10 bilgilendirici belirteç saptanmıştır.

Projenin amaçları arasında (1) yeni ve daha önceden saptanmış TNP ve mikrosatelit allel frekanslarının tahmin edilmesi, (2) haplotiplerin oluşturulması, (3) Hindistan alt popülasyonları arasında, gen içinde ve genler arasındaki LD mesafesinin belirlenmesi bulunmaktadır. Projenin nihai hedefi ise Hindistan halkının DNA varyasyon veri tabanını oluşturmaktır. Veri tabanının araştırmacıların hizmetine sunulması ile hastalıklara yatkınlık, beklenmeyen yan etkiler ve popülasyon göçleri gibi durumlar için insan biyolojisinin anlaşılmasını sağlamak amaçlanmıştır

Proje aynı zamanda Hindistan popülasyonu içindeki varyasyonu temsil eden ve iyi tanımlanmış alt popülasyon gruplarına ait 15.000 bireydeki TNP'lerin doğrulanmasını hedeflemiştir. Popülasyon çeşitliliği göz önüne alınarak çıkartılması amaçlanan iki sonuç bulunmaktadır: (1) Tüm genetik varyasyonun yakalanabildiği popülasyon altyapısına ait kompozisyonu tanımlamak (2) Hindistan popülasyonunun tamamını temsil eden TNP'leri temsilen küçük bir TNP paneli kompozisyonu belirlemek (30).

### **2.5.3. Japon Popülasyon Yapısının TNP Profilleri ile Belirlenmesi**

Popülasyonlarda tabakalanma belli hastalık allel frekanslarını etkilemektedir. Birçok genetik bozukluk az sayıdaki atasal mutasyonların, muhtemelen belli popülasyonlarda sık görülmesi ve tabakalanmadan dolayı popülasyonun tümüne yayılmaması nedeniyle farklı

populasyonlarla karakterize olduğu görülmektedir. Populasyona özgü mutant allellerin varlığı genetik hastalıkların kökenini anlamak için önemlidir. Aynı zamanda risk altındaki toplumlarda tanı imkânı sağlamaktadır.

Vaka kontrol çalışmaları farklı populasyon yapılarında var olmayan ilişkilendirmelere işaret edebilmektedir. Bu nedenle populasyon yapılarının anlaşılması önem kazanmaktadır. Ancak bu sayede farmakogenomik çalışmalar belli bir populasyon için uygulanan bir tedaviye alınacak yanıtın önceden tahmin edilebilmesini mümkün kılabilir. Bu gelişmelerin ışığında; Japon bilim adamları 7003 bireyden alınan TNP genotiplerine dayanan Japon populasyon yapısını diğer alt populasyonlarla karşılaştırılmış, populasyona dayalı birlik (cemiyet) çalışmaları üzerinde etkilerini incelemişlerdir. 60 Avrupalı, 60 Afrikalı ve 90 Asyalı bireylerin genotip verileri HapMap veri tabanından elde edilen veriler Biobank Japon Projesinden elde edilen 7003 Japon hastanın genotip verileriyle kıyaslanmıştır. Genotipleme aşamasında 7 bölgeden bütün Japon DNA örnekleri hastalık türüne göre 272,844 TNP için Perlegen platformunda genotiplenmiştir. Seçilen otozomal TNP'lerin (1-22 kromozom) ileriki analizler için yeterli olabilmesi için verilerin 4 kriteri sağlanmasına dikkat edilmiştir: (1) Japon populasyonunda polimorfik olması, (2) call rate'in yeterince yüksek olması ( $\geq 90\%$ ), (3) Genotip frekanslarının Hardy-Weinberg dengesiyle uyumlu olması ve (4) HapMap Projesinde genotiplenmiş olması.

Bireyler arasındaki ilişkileri araştırmak için otozomal TNP genotipleri kullanılmıştır. Analizler "eigenanalysis" ile yürütülmüştür. Çalışma sonucunda Japon bireyler tarafından iki ana küme (cluster) (Hondo ve Ryukyu) oluşmuştur ve üçüncü küme Han-Çinlileri tarafından oluşturulmuştur. İki ana küme k-means metodu kullanılarak tanımlanmıştır. Aynı zamanda bireyler arasındaki ilişkileri saptayabilmek için PLINK'de (Whole genome association analysis toolset) multidimensional scaling (MDS) metodu kullanılmıştır. Hondo kümesine düşen bireylerin çoğu Japonya'nın başlıca adalarında yaşayanları ve Ryuku kümesine düşenlerin çoğu Okinowa'dan bireyleri içermektedir. Hondo ve Ryuku kümelerinde, en belirgin olarak, maksimum frekans farklılığı 6. Kromozomda HLA bölgesinde bulunmuştur. Hondo ve Ryuku kümelerindeki en fazla nonsynonymous TNP frekans farklılığı *EDAR* geninde saç kalınlığından sorumlu Val/Ala(rs3827760) polimorfizmi ve *ABCC11* geninde kulak vaks tipinden sorumlu Gly/Ala polimorfizmi (rs17822931) olarak bulunmuştur (31).

#### 2.5.4. Estonya Genom Projesi

Genetik bilginin hastalığa nasıl sebep olduğunu anlamak için bütünü kavramanın (hastalık veya fenotip), olayın bir bölümünü kavramaktan daha önemli olduğunu vurgulayan Tartu Üniversitesi Biyoteknoloji Profesörü Andres Metspalu tarafından öne sürülmüş bir projedir. Hastalık belirtilerini göstermiş olan bir grup hastanın genotiplerinin ortaya konmasıyla hastalığa ait genlerin tanımlanması, Estonyalıların kendi kişisel verilerine ulaşabilmeleri ve ileride gerçekleşmesi muhtemel kişiselleştirilmiş ilaç tedavilerinden faydalanabilmeleri için sağlıklı insanların veri tabanının oluşturulması amaçlanmıştır. Bu nedenle Avrupa gen havuzu içerisinde küçük bir bölümü temsil eden Estonyalıların kendi ulusal gen havuzlarını oluşturma ihtiyacında olduğunu söylemiştir.

Estonya Genom Projesi (EGP), Estonya hükümeti ve Estonya Genom Kurumunun (EGF) anlaşması sonucunda 1999 Mart ayında planlanmaya başlanmıştır. Başarıya ulaşmanın ana unsurlarından biri olan kamu-özel sektör birlikteliği EGF ve US kökenli EGeen arasında kurulmuştur. Proje kapsamında 2000 yılında 1,4 milyon olan toplam nüfusun %75'i olan 1 milyon kişinin genotiplenmesi amaçlanmıştır.

Projenin hedefleri arasında: Estonya popülasyonuna ait fenotip-genotip verilerini içeren veri tabanına ait Gen Bankasının EGP adı altında kurulması ile veri tabanının genetik, sağlık çalışmalarında ve yaygın hastalıklarda etkisinin araştırılmasında kullanılması bulunmaktadır. Ayrıca Estonya popülasyonunun sağlık durumunun büyük ölçüde tanımlanması, gönüllülerden doku örnekleri toplanması, LD haritalarının çıkartılması, veri erişiminin sağlanacağı ve pazara sunulacak genom projesi ürünlerine ait yazılımların geliştirilmesi özel amaçlardır. Son olarak büyük ölçekte genomun toplum sağlığına getireceği sistematik avantajların pratikte uygulanabilmesi uzun soluklu amaç olmuştur.

Araştırmalar Estonya popülasyonunun tüm Avrupa popülasyonunun iyi bir temsilcisi olduğunu göstermiştir. Çalışmaların sonucunda Estonya ve diğerlerinin (Caucasians) genotip verileri ve insan 22. Kromozomuna ait LD haritası Estonyalılarla Avrupa popülasyonları arasında küçük farklılıklar gözlenmiştir (32).

### 2.5.5. Pasifik Pan-Asya TNP Girişimi

HapMap Projesi ile farklı popülasyonlara özgü insan genom varyasyonlarının araştırılması amaçlamıştır ve araştırma farklı popülasyonları temsil ettiği düşünülen 1,115 örnek verisine dayanmaktadır. Çalışmada kullanılan veri kümesinin var olan bütün farklı popülasyonları temsil etme olasılığının neredeyse yok denilecek kadar azdır. Bu nedenle Pasifik Pan-Asya girişimi adındaki altındaki araştırmacılar HapMap'den elde edilen verilerin kendi popülasyonlarını ne kadar temsil ettiğini araştırma ihtiyacı duymuşlardır (33).

Çalışmada Asya'daki genetik varyasyonun araştırılması ve genetik haritaların oluşturulması için, 11 farklı Asya ülkesinden katılan bilim adamları ile işbirliğine gidilmiştir. Yapılacak araştırma için farklı popülasyonlara ait kan örnekleri toplanarak, TNP analizlerinin yapılması planlanmıştır. Aynı zamanda örneklerin bölge haritalarından (göç ve ikametgah gibi) yola çıkarak, belli başlı hastalıkların genetik karakteristiklerinin belirlenmesinde bu özelliklerin ilk adım olması amaçlanmıştır.

Pasifik Pan-Asya TNP Girişimine Çin, Hindistan, Endonezya, Kore, Malezya, Nepal, Filipinler, Singapur, Tayland ve Tayvan merkezli enstitülerdeki bilim adamları tarafından 18 Kasım 2004 yılında resmîyet kazandırılmıştır. Projenin 2 yıldan fazla ve 3 milyon dolara mal olacağı öngörülmüştür. Projenin sonucunda elde edilecek tüm verilerin kullanıma açık olması ve uluslararası HapMap projesinde olduğu gibi hastalıklar ve ilaca cevapla ilgili genlerle ilişkili veri tabanı geliştirilmesine benzer çalışmalar için proje grubunun sorumluluğu vurgulanmıştır (34).

Öncelikle Asyalıların birbirinden ne kadar farklı olduğunun araştırıldığı projede 11 ülkeden 30 enstitüden ve 73 farklı alt popülasyon grubundan örnekler alınarak otozomal belirteçler üzerinde çalışılmış ve Asya'ya göç olayları araştırılmıştır. Elde edilen sonuçlar dilbilim ile gen kümelerinin korele olduğu hipotezini doğrulamıştır. Asya'daki gen havuzunun çoğunluğunu ilk modern insanın kıtaya girişinden olduğu vurgulanarak, araştırılan bütün güney ve doğu Asya popülasyonunun aynı ortak atadan geldikleri ileri sürülmüştür. Çalışmaya ait yayın hazırlanma safhasındadır (35).

### **2.5.6. Singapur Genom Varyasyon Projesi**

Singapur Genom Varyasyon Projesi (SGVP) Güneydoğu Asya'daki Çin, Malaya ve Hint popülasyonundan 268 bireyin 1.6 milyon TNP'sini kullanıma sunmaktadır. Çevirim içi veri tabanı HapMap projesine benzer olarak allel frekansları, linkage disequilibrium (LD) değerlendirmeleri ve rekombinasyon oranlarını içeren genotip ve phased haplotip veri bilgilerini ve özetlerini kataloglamaktadır (36).

### **2.5.7. HapMap TNP verilerinin Singapur Çinli Popülasyonuna Transfer Edilebilirliğinin Değerlendirilmesi**

Uluslararası HapMap Projesi insan genomu varyasyon verisi olarak değerli bir kaynak olmasına rağmen diğer popülasyonlara uygulanabilirliği halen etraflı bir şekilde araştırılmamıştır. Bu nedenle Andiappan ve arkadaşları yaptıkları çalışmada yüksek yoğunlukta genotipleme çipleri ve tekrar dizileme stratejileri kullanarak Singapur Çinlileri ile HapMap popülasyonunu karşılaştırmışlardır. İlk olarak 1028 ve 114 akrabalık ilişkisi bulunmayan Singapur Çinli örneklerini Illumina Human HapMap 550K Chip ve Affymetrix 500K Çiplerini kullanarak 270 HapMap örneği ile karşılaştırmışlardır. Daha sonra astım aday geni için 5q31-33 bölgesindeki 20 hedef geni tekrar dizileyerek analizlerde kullanmışlardır.

Çalışma sonucunda tekrar dizileme yöntemiyle 237 TNP elde edilmiştir. Elde edilen TNP'lerin sadece 95'i (%40) HapMap'de bulunmuş, doğrudan genotiplenemeyen 56'sının (%24) ise HapMap'deki haplotipleri bulunmuştur. HapMap Singapur Çinlileri ile Han Çinlileri arasında belirlenen yüksek korelasyon GBAÇ'lerde HapMap kaynaklı çiplerin kullanılabilirliğini doğrulamıştır (37).

### **2.5.8. UK10K - Sağlık ve Hastalıkta Nadir Genetik Varyantlar**

Birleşik Krallık (UK) araştırmacılarının nadir genetik değişimlerle düşük frekans arasındaki bağlantının ortaya konması ve böylece protein değişimlerinin hastalıklara nasıl sebep olduğunun anlaşılması için geliştirilen bir projedir. Projede İngiliz ve Finlandiyalı 10.000 kişinin genetik kodunun detaylı olarak çalışılması planlanmıştır. Şizofren, otistik ve obez hasta grubu ile sekiz nadir hastalık grubundan toplamda 6.000 kişinin ekzon ve 4.000 kontrolün tüm genomunun dizilenmesi amaçlanmıştır. Bu kadar geniş örnek hacmi ile



hastalıklara etki eden yeni genetik varyantların ve nadir varyantların anlaşılması hedeflenmiştir. Proje Wellcome Trust'dan Mart 2010'da 10,5 milyon £ bütçe olarak, 2010 sonunda dizilenmeye başlanmıştır.

Bütün genetik varyasyonlar zararlı veya hastalık yapıcı olmadığından, nadir varyantların ve etkilerinin saptanabilmesi için iki yöntem kullanılmıştır. Projenin ilk aşamasında; fiziksel karakteristikleri iyi tanımlanmış 4.000 kişinin DNA'sı detaylı olarak çalışılarak karşılaştırılacaktır. Projenin bu aşamasında etkisiz olan DNA değişimleri ile hastalığa neden olabilecek değişimlerin tanımlanması amaçlanmıştır.

Projenin ikinci aşamasında ise; ileri derecede sağlık problemi olan 6.000 kişinin protein kodlayan DNA bölgesindeki değişiklikler çalışılarak birinci grupta farkları kıyaslanacaktır. Projede yapılan bu karşılaştırmalar sonucunda; DNA'daki hangi değişimlerin belirgin hastalık problemlerinden sorumlu olduğunu bulunması amaçlanmıştır (38).

#### **2.5.9. 1000 Genom Projesi**

Fenotip ve genotip arasındaki ilişkileri anlamak açısından insan genom dizilimini bilmek araştırmacıların genetik çalışmaları için tek başına yeterli olmasa da varyasyonların sistematik araştırılması söz konusu olduğunda, DNA ve allel frekanslarındaki varyasyonların tüm yelpazedeki yerinin tam olarak bilinmesine ihtiyaç duyulmaktadır. Bu sebeple araştırmacıların 2008 yılına kadar yaptığı çalışmaların sonuçları dbSNP'de (dbSNP 129) ~11 milyon TNP ve 3 milyon insersiyon ve delesyon kataloglanmıştır (39; 27). Bunun dışında büyük genomik varyantların lokasyonlarının kataloglandığı yapısal varyant veri tabanı (örneğin dbVAR) ve farklı popülasyonlara özgü 35 milyon TNP'in ve bağlantı dengesizliği (LD) olarak bilinen, allel frekansları ile varyant ilişkilerini kataloglayan HapMap projesinin web sitesi bulunmaktadır. Sözü geçen kaynaklar hastalık geni keşfetme çalışmalarında birkaç yüz bin varyant olan sitelerde LD yapısı bilgi birikimi ile birlikte ortak varyantların (>%5 MAF) çoğunluğunu kullanarak hastalık ve varyasyon ilişkisinin test edildiği (27) birinci nesil genom boyunca ilişkilendirme (GWAS) kaynakları olmuştur. Son 5 yılda ilişkilendirme çalışmaları ile hastalığa yatkınlık ve diğer ortak özellikler ile ilişkili 1000 genomik bölge ilişkilendirilmiştir (40).

Düşük frekanslı nadir varyantlar (%0,5-%5 ve <0.5) yaygın varyantlara göre fazla olup hastalıkların genel mimarisini anlamlı olarak tanımlamaktadır. Fakat henüz sistematik

olarak çalışılmaları mümkün görülmemektedir (41) (42). Bunun yanında DNA dizileme tekniklerinde olan gelişmeler sayesinde çoğunluğu yaygın varyantların olduğu birinci nesil veri tabanlarındaki bilinmeyenleri aydınlatmaya olanak sağlamaktadır. DNA varyasyonunun daha gelişmiş bir kataloğu oluşturmak yaygın ve nadir varyantları tamamen anlamının bir ön şartı olmuştur. Bu sebeplerden yola çıkarak planlanan 1000 Genom Projesi TNP, yapısal türevleri ve haplotipleri de dahil olmak üzere insan genetik varyasyonun kapsamlı bir kataloğunu oluşturmak için oluşturulan uluslararası işbirliğidir. Oluşturulacak katalog ile genom boyunca ilişkilendirme çalışmaları ve diğer medikal araştırma çalışmalarının desteklemesi amaçlanmıştır (43).

1000 Genom Projesi'nin temel amacı farklı popülasyonlara ait her çeşit DNA polimorfizmlerine özgü yeterli haplotip bilgisi sağlamak olmuştur. Fonksiyonel alleller düşük allel frekansı ile kodlama bölgesinde olduğundan bugünkü yüksek işlem hacimli dizileme teknolojileri ile beş ana popülasyon grubunda (Avrupa, Doğu Asya, Batı Afrika ve Amerika kökenli popülasyonlar) özellikle genomik bölgede yer alan % 95'in üzerinde ve allel frekansı %1 ve üzerinde olan (polimorfizm) varyantların karakterize edilmesi olmuştur. Ayrıca kodlama bölgelerinde % 0,1'in altında olan allel frekansları da kataloglanmıştır.

Ucuz, hızlı ve etkin metot seçilimi için 2008 yılında başlatılan pilot proje 3 aşamada yürütülmüştür. (1) 179 örneğin<sup>7</sup> tüm genom dizilemesi (low coverage 2-6X, Ekim 2008'de tamamlandı), (2) 2 ailenin<sup>8</sup> tüm genom dizilemesi (high coverage- ortalama ~42X, Ekim 2008'de tamamlandı), (3) 7 popülasyondan 697 örnekte<sup>9</sup> 8.140 ekzon dizilemesi gerçekleştirilmiştir. (~50X, Haziran 2009'da tamamlandı) (43).

Pilot projede uygulanacak metotlar geliştirilerek değerlendirerek, %95'den fazla yaygın varyantın keşfedilmesine rağmen özellikle ekzon kodlama bölgesi dışında kalan düşük frekanslı varyantların karakterizasyonu zayıf olduğu görülmüştür.

---

<sup>7</sup> akrabalığı olmayan 59 YRI'lı örnek, akrabalığı olmayan 60 CEU, akrabalığı olmayan Beijing'den 30 Han Çinlisi (CHB) ve akrabalığı olmayan Tokyo'dan 30 Japon örnek(JPT)

<sup>8</sup> 2 anne/baba/çocuk üçlüsü 6 kişi (Ibadan Nijeryadan bir Yorubalı (YRI); Utah'dan bir Avrupa kökenli (CEU))

<sup>9</sup> 7 Afrika popülasyonu (YRI, Webuye'deki Luhyalılar, Kenya (LWK)), Avrupalı (CEU, İtalyada yaşayan Toskanalılar (TSI)) ve Doğu Asyalılar (CHB, JPT, Denverda yaşayan Çinlikler, Kolorado (CHD))

Pilot proje sonrasında low-coverage ve ekzon dizilemesi yöntemlerinin kombinasyonu kullanılarak 14 popülasyondan<sup>10</sup> 1.092 örneğin, genom boyu dizilemesi (low-coverage 2-6X), hedef bölgeler için ayrıntılı (deep) ekzon dizilemesi (50-100X) ve TNP genotip verileri kullanılarak analiz edilmiştir. Çeşitli algoritmalar ve farklı veri kaynakları kullanılarak geliştirilen metotların entegrasyonu ile 38 milyon TNP, 1,4 milyon kısa insersiyon, delesyon ve 14.000'den fazla büyük delesyonun haplotip haritası elde edilmiştir. Çalışma sonucunda farklı popülasyonlara özgü örneklerin nadir ve yaygın varyantlar bakımından farklı profilleri olduğu, düşük frekanslı varyantların coğrafi farklılıklarla önemli değişiklik gösterdiği ve seçilimin saflaştırılması (purify) ile daha fazla arttığı ortaya konmuştur (44).

Evrimsel korunma ve kodlama sonuçları (özellikle biyolojik yollarda nadir varyantlarla değişkenlik gösteren) saflık seçiminde anahtar güçlerdir. Tüm bireyler korunan bölgelerde yüzlerce kodlanmayan nadir varyant içermektedir. Örneğin, transkripsiyon bağlanma bölgelerinde motif bozan değişiklikleri. Bu kaynak ile karışık popülasyonlar da dâhil olmak üzere farklı popülasyonlara özgü % 1 frekansta yaygın ve düşük frekanslı TNP'lerin % 98'ini yakalanabilmiştir (44; 45).

## **2.6. Araştırmalar için TNP veri tabanları**

Sistemik erişim imkânı olan, yönetilebilir, güncellenebilir, taşınabilir, birbirleri arasında tanımlı ilişkiler bulunabilen, veri tekrarının olmadığı, bilgiler kümesi veri tabanı olarak adlandırılmaktadır. Başka bir deyişle veri tabanı, bilgisayarda sistemik şekilde saklanan programlar yardımıyla işlenebilen veri yığınıdır. İlişkisel Veri tabanı Yönetim Sistemleri (Relational Database Management Systems - RDBMS) büyük miktarlardaki verilerin güvenli bir şekilde tutulabildiği, bütünlük içerisinde tutulan bilgilere hızlı erişimin sağlandığı ve birden fazla kullanıcıya aynı anda erişim imkânının sağlandığı SQL MySQL gibi programlardır. İlişkisel veri tabanı yönetim sistemleri ile veri tabanlarının yapısı oluşturulmaktadır. Bir veri tabanı oluşturulurken önce mantıksal daha sonra fiziksel tasarım yapılmaktadır. Veri tabanının mantıksal tasarımında çalışmalar veri tabanı işletim sisteminden bağımsız olarak yürütülmektedir. Fiziksel tasarımda ise mantıksal tasarımda ortaya çıkan nesnelere SQL gibi bir veri tabanı işletim sistemi kullanılarak gerçekleştirilmektedir.

---

<sup>10</sup> Avrupa, Doğu Asya, Sahra altı Afrika ve Amerikalılar

Gelişen yüksek işlem hacimli teknolojilerle birlikte ortaya çıkan veri artışı nedeniyle verilerin tekrar etmeden saklanabilmesi, saklanan verilerin içerisinde istenilen bilgilerin kısa zaman içerisinde özet halde alınabilmesi, analizlerin rutin göz taramasıyla veya basit hesaplarla yapılamaması nedeniyle genom projelerinin en başından itibaren veri tabanları yerlerini almışlardır. Bunun ilk göze çarpan örneği 1982 yılında Genbank veri tabanının kurulması olmuştur.

Günümüzde, kamuya ait TNP veri tabanlarında, genom boyunca ham olarak beş milyondan fazla tek nükleotit değişimi (TNP) tutulmakta ve çeşitli genom tarayıcıları bulunmaktadır (**EK1, EK2**).

TNP veri tabanları dışında insan gen ve genetik bozukluklarını kataloglayan önemli bir veri tabanı da OMIM (Online Mendelian Inheritance in Man)'dir (Mah & Chia, 2007).

### **2.6.1. Tek Nükleotit Polimorfizm Veri Tabanı (Short Genetic Variations-dbSNP)**

dbSNP en sık kullanılan veri tabanlarından birisidir (46). dbSNP günümüze kadar yapılan araştırmalar sonucunda bulunan genomik ve cDNA dizi varyasyonlarını içermektedir. dbSNP çoğunlukla tek baz mutasyonlarını (TNP), mikrosatelit ve herhangi bir türdeki kısa delesyon ve insersiyonları kapsamaktadır. Aynı zamanda dbSNP'de bu varyasyonlara ait deneme sayıları, keşif yöntemleri, bilginin doğrulanması, "flanking dizilerin" polimeraz zincir reaksiyon (PCR) şartları ve popülasyon istatistikleri gibi detaylı veriler bulunmaktadır. Kısıtlama olmaksızın herkese açık olan dbSNP, GenBank, Pubmed, Locuslink ve NCBI sitesindeki İnsan Genom Projesi bilgileriyle ilişkilendirilmiş bir veri tabanıdır. dbSNP'de TNP detayları fazla olmasına rağmen, hastalık türü ile TNP'lerin direk ilişkilendirebileceği bağlantı bilgisi bulunmamaktadır. Araştırmacı bu bilgiye ulaşmak için başka sitelerdeki bağlantıları kullanmak zorundadır. Ortalama TNP yoğunluğunun yaklaşık kilo bazda 1,1 TNP olduğu dbSNP'de yaklaşık 300 kaynaktan toplanan ~4 milyon "nonredundant" insan TNP'i bulunmaktadır ve bu sayı her geçen ay yaklaşık 90 TNP artış göstermektedir (47). 2008 yılında 11 milyon 129. versiyon Ocak 2013 yılı itibariyle 137. Versiyonu yayınlanan dbSNP'de ~ 63 milyon Nisan 2013 yılı itibariyle ise ~ 73 milyon TNP kaydı tutulmaktadır (48; 49).

### **2.6.2. TNP Konsorsiyum - Uluslararası HapMap Projesi (The SNP Consortium–TSC-International HapMap Project)**

TSC, 1999 yılında şirketlerin ve enstitülerin işbirliği ile farklı toplumlara özgü TNP'lerin tanımlanmasını amaçlamıştır. TSC 2 yıl içerisinde 300,000 TNP keşfi amaçlanmış fakat gelişen teknoloji ile 2001 yılı sonunda 1,4 milyon TNP yayınlamıştır.

TSC'nin web sitesi (<http://snp.cshl.org>) kullanıcılara gen veya TNP ile arama/göz atma veya metin dosyası olarak araştırma alanında kullanabilme olanağı sağlamaktadır. Grafiksel genom tarama arayüzü genom üzerinde TNP haritalanmasının yanı sıra harici olarak uygun olan gen tahminlerini de getirmektedir. Web sitesinden genoma ait açıklamalara ulaşılabilmekte, görüntü yakınlaştırma/uzaklaştırması yanı sıra, özelliklerin kişiselleştirilebilmesi de mümkün olmaktadır. Web sitesinden TNP allel sıklıkları ve genotip verilerinin FTP yoluyla indirilmesi, TNP'in bağlantı (linkage) haritasının indirilmesi ve karşılaştırmalı harita görüntüleme taraması da yapılabilmektedir. Fakat web sitesinde hastalık patolojisi ile ilgili bir arama özelliği bulunmamaktadır. Web sitesinin Japon veri tabanına bağlantısı bulunmaktadır (50).

### **2.6.3. Japon Veri Tabanı (Japanese Single Nucleotide Polimorphisms Database-JSNP)**

JSNP veri tabanı, Japon popülasyonu TNP verilerini tutmak amacıyla 2000 yılında Başbakanın mileniyum projesi olarak, İnsan Genom Merkezi (Human Genome Center (HGC)), Sağlık Bilimleri Enstitüsü (IMS) Tokyo Üniversitesi ve Japonya Fen ve Teknoloji (JST) işbirliği ile başlatılmıştır (<http://snp.ims.u-tokyo.ac.jp/> tarih yok). Projenin amacı gen bölgelerinde veya düzenleyici gen bölgelerinde, Japon popülasyonuna özgü, 150.000 TNP tanımlanması, iki sene içerisinde bir araya getirilerek; kullanıma açılması ve polimorfizmler için analitik araçlar geliştirilmesi olmuştur. Projenin esas amacı polimorfizmler ile yaygın hastalıklar ve/veya ilaç reaksiyonları arasındaki ilişkiyi tanımlayabilecek temel veri setlerini oluşturmaktır. Bu nedenle fenotipi etkileyen fakat hastalığa neden olmayan aday TNP'lerin tanımlanmasında detaylı olarak çalışılmıştır. JSNP web sitesi, PCR profilleri, ilaç reaksiyonları, özgün dizi arama ve karşılaştırma (BLAST) TNP, dbSNP, OMIM hastalık ve haritalama bilgi sorgusu yapabilmesine rağmen protein dizi analiz bilgisi araması bulunmamaktadır (51).

2002 yazında 190,562 genetik varyasyon keşfedilerek, veri yönetim ve veri dağıtım kısımlarından oluşan web sitesi araştırmacıların kullanımına açılmıştır (52).

Günümüzde Japon TNP veri tabanı Geriyatrik Araştırmalar için Japon TNP veri tabanı (JG-SNP) (Tokyo Metropolitan Geriatric Medical Center), İnsan Mitokondrial genomu TNP veri tabanı (mtSNP), (Gifu International Institute of Biotechnology) Protein Polimorfizm Veri tabanı (dbProP), (National Institute of Radiological Sciences), çalışmalarını bir arada tutan bir Japon veri tabanı web sitesi de oluşturulmuştur (53).

#### **2.6.4. Pan-Asya TNP Genotip Veri tabanı (the Pan-Asian SNP genotyping database-PanSNPdb)**

HapMap'de kullanılan veri yığınının var olan bütün farklı popülasyonları temsil etme olasılığının neredeyse yok denilecek kadar az olması varsayımından yola çıkarak Pasifik Pan-Asya girişimi adı altında araştırmacılar HapMap'den elde edilen verilerin kendi popülasyonlarını ne kadar temsil ettiğini araştırma ihtiyacı duymuşlardır (33).

Pasifik Pan-Asya TNP Girişimine Çin, Hindistan, Endonezya, Kore, Malezya, Nepal, Filipinler, Singapur, Tayland ve Tayvan merkezli enstitülerdeki bilim adamları tarafından 18 Kasım 2004 yılında resmiyet kazandırılmıştır. Projenin 2 yıldan fazla ve 3 milyon dolara mal olacağı öngörülmüştür (34; 35).

HUGO Pan Asya TNP konsorsiyumu Asyalıların genetik varyasyonu ile ilgili bugüne kadar olan en fazla örnekle<sup>11</sup> yapılmış olan en geniş çalışmadır. Bu örnek verilerini ve verilere özgü çeşitli analizleri içeren PanSNPdb veri tabanı <http://www4a.biotech.or.th/PASNP> oluşturulmuştur. PanSNPdb LD modeli, haplotip dağılımı ve KSV'leride içerecek şekilde Asya insanının popülasyon yapısını ortaya koyan analizler için bir araştırma kaynağıdır. Buna ilaveten PanSNPdb, HapMap3, JSNP, dbSNP, DGV (Database of Genomic Variants) gibi diğer TNP ve CNV veri tabanları ile interaktif bir karşılaştırma sağlamaktadır. Bunun yanında genetik varyasyonlar açısından kapsamlı bir kaynak sağlamaktadır (54).

---

<sup>11</sup> 71 popülasyondan (Çin, Hindistan, Endonezya, Japonya, Malezya, Filipinler, Singapur, Güney Kore, Tayvan ve Tayland alt popülasyonları) akrabalık ilişkisi olmayan 1,719 örnek

### **2.6.5. İnsan Genom Varyasyonları Veri Tabanı (Human Genome Variation Database-HGVbase)**

HGVbase resmîyette İnsan Geni Biallelik Dizileri (HGBASE) ismi ile kullanılmaktadır. HGVbase bilinen insan gen bölgelerindeki (intragenic) dizi varyasyonlarını kataloglayan gen temelli bir veri tabanıdır. Veri tabandaki bilgiler, Karolinska Enstitüsü (KI), Avrupa Biyoinformatik Enstitüsü (EBI) ve Avrupa Moleküler Biyoloji Laboratuvarını (EMBL) kapsayan Avrupa Akademik konsorsiyumu Interactiva GmbH (Almanya), tarafından sağlanmıştır. Bu veri tabanı çoğu TNP'ler olmak üzere farklı varyasyon tiplerini içermektedir. Veri tabanında insan geniyle ilişkili (promoter, ekzonik ve intronik) TNP detayları verilmekte ve TNP-gen fonksiyon ilişkisi online olarak sağlanmaktadır (55).

### **2.6.6. Genotip ve Fenotip Veri Tabanı (The database of Genotypes and Phenotypes-dbGaP)**

Genotip/Fenotip etkileşimlerini araştıran çalışmaların sonuçlarını yayınlamaktadır. Genom boyunca ilişkilendirme çalışmaları, tıbbi dizi analizi, moleküler tanı testlerinin yanı sıra, genotip ve klinik olmayan özellikteki ilişkilendirme çalışmalarını kapsamaktadır. Genotipleme için yüksek işlem hacimli, uygun maliyetli yöntemlerin geliştirilmesi ile bu analizleri mümkün kılacak güçlü araçlar sunmuştur (56).

### **2.6.7. Varyasyon Veri Tabanı (Database of Genomic Structural Variation-dbVar)**

dbVar büyük ölçekli genomik varyantların kataloglandığı insersiyon, delesyon ve inversiyon olaylarını içeren veri tabanıdır (57).

## **2.7. Araştırmalar için diğer veri tabanları**

### **2.7.1. Kaliforniya Üniversitesi, Santa-Cruz Genom Web Tarayıcısı Veri Tabanı (UCSC Genome Browser Database)**

Günümüzde gittikçe artan gen dizi bilgisinin kamu veri tabanlarında toplanarak, doğru ve tam açıklamalı sunulması, veri analizi ve yorumlanması açısından önem arz etmektedir. Bu nedenle UCSC'de geçerli ve arşivlenmiş genom bilgileri, anotasyonları ile araştırmacıların

kullanımına sunulmuştur (58). USCS meta veriler, mRNA ve Expressed Sequence Tag (EST), dizi sıralaması (alignment), gen tahminleri, farklı tür homolojileri ve TNP bilgilerini içermektedir. UCSB web tarayıcısında hızlı görüntüleme, verinin grafik ve metin tabanlı sorgulanması; dizi ve ek açıklama verilerinin MySQL ilişkisel veri tabanında tutulmasıyla sağlanmıştır.

### 2.7.2. SNPper

SNPper (<http://snpper.chip.org/>) farklı veri tabanlarından TNP verilerini otomatik olarak çeken web tabanlı bir uygulamadır. Analiz sonucunda uygulama, kullanıcının tanımladığı TNP'leri çeker ve sonuçları XML veya HTML gibi formatlarda ekrana getirir. Araştırmacı TNP sıklıkları, TNP bilgisinin güvenilirliği, TNP lokasyonu (intergenik veya intragenik bölgeler), TNP kaynaklı amino asit değişikliği ve bunların düzenleyici karakteristikleri web sayfasında tanımlayabilir (promoter, intron, ekzon). Sonuçlar web tabanlı arayüzde grafiksel olarak da görüntülenebilmektedir. Bunun dışında veri setinin ayıklanması, arama mesafesinin kısaltılması, bunun yanısıra analiz için daha fazla TNP girişi eklenmesi gibi ek kriterlerin girilmesi TNP filtreleme fonksiyonu ile yapılabilmektedir.

SNPper'in ilişkisel veri tabanının GoldenPath ile eş zamanlı bağlantısı bulunmaktadır. Veri tabanı dbSNP 'in ve GoldenPath'in ayrıştırılmasından derlenmiş olup yaklaşık 2 milyon TNP ve 15.000 genin üzerinde bilgi üretmektedir. Veri tabanında işleme sonrası TNP ilgili gen bölgesine bağlanmaktadır. Eğer TNP gen kodlayan bölgeye denk geliyorsa, amino asit değişikliğini de saptayabilmektedir.

SNPper'in kullanıcı erişimini sağlayan web tabanlı arayüzü Common Lisp dili ile yazılmıştır (Linux işletim sisteminde). SNPper'in temel amacı TNP seti olarak adlandırılan, genomun belirlenmiş bölgelerinde istenilen özelliklere sahip TNP koleksiyonu yaratmaktır. SNPper belirlenen TNP setini benzersiz şekilde işler, süzer, kaydeder ve karışık sorgular sonucunda TNP setini oluşturarak ekrana getirir (örneğin bir tanesi gen ontoloji sınıfını kullanır). Kullanıcıya uzak işlem arama arayüzü (Remote Procedure Call Interface) aracılığıyla xml dosyalarına tam erişim hakkının verildiği yegane TNP veri tabanıdır. Kullanıcı; yükleme, analiz ve TNP veri setlerini bu arayüz yardımıyla görüntüleyebilir. dbSNP ve GoldenPath web sitelerindeki analiz araçları karşısında SNPper diğer uygulamalarla iyi işlerliğe izin vermektedir. Fakat SNPper hastalık ve patolojik işlem aramalarında kolaylık sağlamamakta olup veri tabanı bilgisi insanla sınırlıdır. Bu yönüyle



kompleks hastalıklara neden olan TNP yerleşimleriyle ile ilgilenen klinik araştırmacılar için çok yararlı değildir (59).

### **2.7.3. TNP İlişkili Proje Dizilerin Sıralanması ve Değerlendirilmesi (Project Aligned Related Sequences and Evaluate SNP-PARSESNP)**

PARSESNP (<http://www.proweb.org/parsesnp/>) hem forward hem de reverse genetiğin kullandığı web tabanlı veri analiz aracıdır. Varyantlar, İnsan Geni Mutasyon Veri tabanı (Human Gene Mutation Database- (<http://www.hgmd.org/>)), dbSNP ve Swissprot, gibi veri tabanlarından okunabilir. Gendeki missense mutasyonları ve ilgili mutasyona özgü protein ürünlerini, restriksiyon enzim polimorfizmleri ve ilgili restriksiyon enzim bölgeleri gibi nükleotit değişikliklerinin etkilerini belirleyebilir. Fakat hastalık ve patolojik işlem araması gibi kolaylıklar bulunmamakta olup buda hastalıklara yol açan TNP'ler ile ilgilenen klinik araştırmacılar için çok yararlı olmamaktadır (60).

### **2.7.4. Fonksiyonel Elemanların TNP Veri Tabanı Functional Element SNPs Database in human (FESD)**

FESD gen bölgelerinde bulunan TNP setlerini araştırmacıya sunmasının yanısıra genotipleme deneyleri için ihtiyaç duyulan "flanking bölgelerin" ve insan genomundaki fonksiyonel elemanların kategoriler halinde organizasyonu için tasarlanmış bir veri tabanıdır. İnsan gen bölgeleri FESD'de on ayrı fonksiyonel elemana ayrılmıştır. Bu bölgeler; promoter bölgeler, CpG adacıkları, 5' transle olmayan bölgeler (5'-UTR), translasyonun başladığı bölgeler, splice bölgeleri, kodlayan ekzonlar, intronlar, translasyonun durduğu bölgeler, poliadenilasyon sinyal bölgeleri ve 3'-UTR bölgeleridir. Özgün bölgelerdeki her fonksiyonel eleman için bilinen bütün TNP'lerin yeri tayin edilmektedir. FESD'nin web arayüzüne <http://combio.kribb.re.kr/ksnp/resd/> adresinden ücretsiz erişim sağlanmaktadır. Araştırmacılar "flanking diziler" boyunca belirli fonksiyonel elemanlardaki özel ilgilendikleri TNP gruplarını manuel olarak seçebilirler. Seçilen TNP veri setleri kompleks hastalıklara neden olan veya katkıda bulunan mutasyonların bulunabilmesi için gen tabanlı haplotip veya bağlantı analizlerinde ve dizileme deneylerinde kullanılmaktadır. FESD'in diğer kamu veri tabanlarından farkı kamunun web sitesindeki uygulamalara aşına olmayanlar için tasarlanmış kullanıcı dostu arayüzüdür. FESD'in yazılım uygulaması Perl script kullanarak NCBI'nin FTP sitesinden veriyi ayrıştırır ve MySQL ilişkisel veri tabanına

aktarır. Site hastalık geninin tanımlanmasında, klinik arařtırmacılar için düzensizlik (disorder) bilgilerine ulařılmasında veya klinik özet (synopsis) bilgilerine ulařmakta yararlı olmaktadır. Web sitesinde gen ve hastalık baęlantısı saęlanabilmesine raęmen, direk olarak TNP ve hastalık iliřkisi kurulamamaktadır. Bu nedenle kullanıcının gerekli baęlantıyı saęlayarak, bahsi geen hastalıkla ilgili hangi TNP'nin hangi gende bulunduęu ile ilgili gerekli ıkarımlarda bulunması gerekmektedir (61).

### **2.7.5. Genetik Asosiyasyon Veri Tabanı (the Genetic Association Database-GAD)**

Kamu tarafından organize edilmiř olan GAD (<http://geneticassociationdb.nih.gov/>) 5000'den fazla insan asosiyasyon alıřmasını kapsayan etraflı bir veri tabanıdır. Günümüzde kullanılan yüksek iřlem hacimli sistemlere uygun olarak yaygın genetik hastalıklarının sistematik analizine olanak tanımaktadır. Web tabanlı moleküler ve klinik arama parametreleri sayesinde, hem moleküler biyoloji hem de klinik disiplin alıřanlarına eriřimi kolaylařtırmaktadır. GAD karıřıklıklara neden olmamak için veri tabanında resmi İnsan Genom Organizasyonunun (HUGO) resmi gen isim ve sembollerini kullanmaktadır. GAD her kayıt için açıklama ve HapMap ve PubMed gibi dięer veri tabanlarına baęlantı saęlamaktadır (62).

### **2.8. Genom Boyunca İliřkilendirme alıřmaları**

Farklı popülasyonlara ait bireyler farklı genleri tařımazlar yine de aynı genin farklı Őekillerde ifadenmesi söz konusudur. DNA metabolizmasında temel mekanizma aynı olduęu halde genlerin anlatım Őekilleri deęiřmektedir. Bu nedenle, farklı popülasyonlardaki genotip-fenotip iliřkilerinin belirlenmesi önemlidir. Alt popülasyonlardaki gen anlatım farklılıklarının arařtırılmasında genlere komřu olan bazı düzenleyici bölgeler önem tařmaktadır. Dolayısıyla, genlerin yakınındaki bölgelerdeki TNP farklılıklarına bakılarak bunların gen anlatımına olan etkileri üzerinde arařtırmalar önemli noktayı oluřturmaktadır.

İGP ve HapMap verilerinin daha detaylı arařtırılması ve gen ifadenmesindeki farklılıkların nedeninin ortaya konarak özellikle bazı alt popülasyonların bazı hastalıklara dięerlerinden neden daha yatkın olduęu anlařılabilecektir. Örneęin sistik fibröz hastalığın beyazlarda neden daha fazla etkili olduęunun sebebi bulunabilecek veya tay-sach genetik hastalığının, Kuzey Avrupa Kafkaslarında, Güney Avrupa Kafkaslarında,

Askenazi Yahudilerinde, Afrikalı Amerikalılarda, İspanyollarda ve Asya kökenli Amerikalılarda yüksek oranda görülmesine rağmen diğer popülasyonlarda ya hiç görülmediği ya da daha az sıklıkta görüldüğünün nedeni ortaya konabilecektir (63).

## 2.9. GBAÇ Çalışmalarının Değerlendirilmesi

DNA dizilimindeki farklılıkların bireyler arasındaki genetik farklılıklara neden olduğu, 1970'lerden itibaren genetikçiler tarafından yapılan araştırmalarla ortaya konmaktadır. GBAÇ çalışmalarının temel amacı; hastalıkların varyantla ilişkilendirilmesinin popülasyonlara uygulanabilirliğinin araştırılması ve ortak varyasyonların hastalıklarla, ilişkisinin ortaya konulmasına çalışmaktır. Araştırmalar; bazı popülasyonlarda bilinen milyonlarca dizi farklılığının bir kısmının % 5'den fazla olduğu ve farklı coğrafik ve bölgesel etkiler göz önüne alındığında bu varyasyonların bazılarının coğrafik bölgeye bağlı olarak daha da fazla görülebildiğini ortaya koymuştur. Ayrıca günümüze kadar yapılan GBAÇ araştırmaları hastalıkların sadece % 5-% 50'sinin kalıtsal olduğunu da göstermiştir (Carlos D. Bustamante, 14 JULY 2011).

Araştırmalar genomun büyük çoğunluğunun haplotip blokları şeklinde olduğunu ve araştırmalarının etkinliğinin haplotiplerin belirlenmesi ile arttırılabildiğini göstermektedir. Bu tür ilişkilerin ortaya konması, araştırmalardaki veri yoğunluğunun azalmasına ve araştırmacının daha fazla bilgi almasını sağlamasına olanak sağlamaktadır. Barret ve Cardon'a göre, günümüzde birbiriyle ilişkili TNP'ler (tag TNP) genotip-fenotip ilişkilerinin araştırılmasında ve genom boyunca yapılan araştırmalarda artan bir öneme sahiptir. Fakat gerek genotip-fenotip ilişkilerinin ortaya konmasında yapılan çalışmalarda gerekse popülasyon çalışmalarında her bir uygulama için farklı tag TNP çip oluşturmanın pratik olmadığını görülmüştür. Bu nedenle araştırmalar, tanımlanmış TNP genotipleri ile tasarlanan ve ticari olarak piyasaya sürülen marker çipler kullanılarak gerçekleştirilmeye başlanmıştır. Günümüzde ise bu hazır çipler aracılığıyla araştırmalar hızlı ve doğru olarak yürütülebilmektedir (64).

TNP mikrodizinler bireylerde ve popülasyonlarda genetik varyasyonu belirlemek için kullanılan özel bir çeşit DNA mikrodizin türüdür. Genetik varyasyon ve hastalıklara duyarlılıktan sorumlu olduğu düşünülen TNP'yi belirlemek için kısa nükleotit çipleri kullanılabilir. Genel olarak genotip belirleme olarak adlandırılan bu tip DNA mikroçip uygulamaları adli tıp çalışmalarında, hastalıklara genetik yatkınlığın hızlı bir

şekilde bulunması ya da ölçülmesi çalışmalarında ve DNA temelli ilaç adaylarının tanımlanmasında kullanılabilir.

GBAÇ çalışmaları sonucunda, popülasyona özgü fenotip-genotip ilişkilerinin ortaya konulabilmesi için popülasyona özgü TNP veri tabanlarının oluşturulması farmakogenomiğin de devreye girmesi ile büyük önem kazanmıştır. Araştırmacıların veri tabanları aracılığı ile yaptıkları GBAÇ analizlerinin sonucunda bölgesel farklılıkların neden olduğu şaşırtıcı popülasyon tabakalanmalarından bahsedilmektedir.

Genotipleme maliyetlerin azalması ve uluslararası HapMap projesinin tamamlanması ile GBAÇ çalışmaları doğru ve hızlı bir şekilde yürütülmeye başlanmıştır. Bu çalışmalar 4 milyondan fazla saptanan genetik varyant arasından seçilerek 100.000–500.000 marker ile sürdürülmektedir. Bugünkü imkanlar bazı durumlarda bütün varyantların genotiplenmesine olanak sağlasa da uygulamada, genetik varyant sayısının azaltılarak, verdiği bilginin maksimize edilmesi açısından daha büyük önem arz etmektedir.

Minimum sayıda varyantla, maksimum bilgi alma yaklaşımlarından bir tanesi genomda fonksiyonel bölgelere odaklanmak olmuştur. Bir diğeri ise büyük örnek gruplarında geçerli olan TNP varyantları arasından en doğru ve fazla bilgi verebilecek TNP gruplarına odaklanılarak bu bölgelerdeki araştırmalara devam edilmesi şeklindedir. Bu şekilde yapılan özgün tasarımlar, genom üzerindeki küçük bölgeler ve hedef genler içindir. Fakat GBAÇ çalışmalarında bu tür tasarımların yapılması pratik olmamakla birlikte maliyetli bir yöntemdir. Bu nedenle, genom boyunca yapılan araştırmalarda, yaklaşıma özgü tasarım yapmak yerine, önceden tasarlanmış ve ticari olarak üretilmiş olan hazır çipleri kullanmak doğru ve hızlı sonuçlara ulaşmamıza yardım etmektedir. Yalnız bu durumda piyasada bulunan araçlarla yapılacak sebep-sonuç ilişkisi araştırmalarında, doğru araç seçimi de önemli olmaktadır (64).

Hazır GBAÇ marker setlerinin pratik kullanımı belirgin avantajlara sağlamaktadır. Hastalık araştırması yapan laboratuvarlarda son 10 senede bağlantı çalışmalarında yapıldığı gibi, ortak çiplerdeki veri setlerinin bir araya getirilerek, istatistiksel metotlarla analiz edilebilmeleri bu avantajlardandır. Bu geniş kullanım amaçları popülasyon çeşitliliğinden allel sıklıklarının temsil edilmesine olanak sağlayan farklı belirteçlerin seçimin stratejisinin önemini göstermektedir (64).

## 2.10. GBAÇ Çalışmalarında TNP Seçim Metotları

**Tag:** LD esasına dayalı tag TNP seti; TNP başına yakalanan varyasyon miktarını maksimize etmek için dikkatlice seçilmiştir. Barrett ve Cardon tarafından yapılan çalışmada teorik olarak en etkin tag TNP seti araştırılmıştır. Daha sonra, bu teorik seti en iyi temsil eden örnekler dikkate alınmıştır. Bunlar ticari olarak mevcut ürünlerden 317 belirteçle Illumina HapMap-300 seti ve ~500.000 TNP'yi temsil eden tam HapMap500 setidir (64).

Illumina'nın HumanHap300 BeadChip'i uluslararası HapMap Projesinden ([www.hapmap.org](http://www.hapmap.org)) türetilen 317.000 tag TNP marker ile tüm genom genotiplemesine olanak tanımaktadır. BeadChip tag TNP'lerden faydalanarak akıllı TNP seçimini mümkün kılan "Infinium Assay"ler tarafından güçlendirilmiştir. HumanHap300 aynı zamanda iki örnek BeadChip (HumanHap300-Duo) veya ek bireysel TNP'ler ile 60.800'e kadar kişiselleştirilebilmektedir (HumanHap300-Duo+) (65).

**Rastgele (random):** LD modelini göz ardı ederek genom boyunca takriben rastgele dağılmış TNP setleridir. Barrett ve Cardon tarafından yapılan çalışmada Affymetrix'de 111.000 ve 500.000 marker ile mevcut olan iki "agnostik" set değerlendirilmiştir. (Affymetrix 500K, Affymetrix 111K) (64).

**Kombine (combined):** Rastgele seçilmiş TNP'ler ve dikkatlice seçilmiş "fill-in" setlerden oluşmuş, iki metodu da kapsayan bir kombinasyondur. Genotipleme hızı ve yüksek maliyet-verim sebebiyle, çeşitli hastalıkların araştırılmasında genomun rastgele temsil edilmesinin getirdiği açığı kapatmak için Affymetrix'in ısmarlama kombine tag setleri random set olarak hazırlanabilmektedir (Affymetrix 500k + 175K tag) (64).

**Fonksiyonel:** Bunlar ya hepsi bilinen polimorfik sinonim olmayan (nonsynonymous) TNP çiplerdir (Affymetrix tarafından 12.000 sinonim olmayan TNP içeren MegAllele sistemi) ya da genellikle gen bölgelerini içeren daha geniş setler üzerine yoğunlaşan, Illumina'nın Human-1 BeadChip'leri gibi çiplerdir.

## 2.11. GBAÇ Çalışmalarında Kullanılan İstatistiksel Analizler

### 2.11.1. TNP Frekanslarının Hesaplanması

Uluslararası HapMap Projesi fenotip farklılıklarından genetik determinantları tanımlamak amacıyla asosiyasyon çalışmalarına TNP genotip data kaynağı sağlamaktadır. Asosiyasyon çalışmalarında veri yığınının azaltılması, işleme süresinin kısaltılması ve araştırma sorularına kısa zamanda yanıt bulunabilmesi amacıyla, veriler bilgisayarda önişleme tabii tutulmaktadır. Veri işleme sonrasında düşük genotipleme yüzdesiyle az bilgi veren ve %5'e kadar olan minör allel frekans yüzdesi olanlar popülasyonlarda göz ardı edilmektedir.

Eğer  $f(AA)$ ,  $f(Aa)$  ve  $f(aa)$  bir lokusu temsil eden iki allelin üç farklı genotipi ise A allelinin p frekansı ve a allelinin q frekansı allelerin sayımı ile elde edilir. Bir popülasyon için A allelinin toplam p frekansı **Denklem 24** ve a allelinin q frekansı **Denklem 25**'deki gibi hesaplanmaktadır.

$$p = f(AA) + 1/2f(Aa)$$

**Denklem 24**

$$q = f(aa) + 1/2f(Aa) \quad (10)$$

**Denklem 25**

Popülasyonda en az yaygın allel frekansına MAF denir ve uluslararası HapMap projesi ile MAF %5'den küçük olan TNP'ler hedeflenmiştir.

### 2.11.2. R Project ile İstatistiksel Analizler

Mesafe-bazlı kümeleme metodu (distance-based clustering) ve model-bazlı kümeleme metotları popülasyon yapısının belirlenmesinde iki ana yaklaşımdır. Mesafe-bazlı uzaklık tabanlı yaklaşımlar bireyler arasında uzaklık olarak allel paylaşım oranını kullanmaktadır ve hesaplama açısından çok çaba gerektirmemektedir. Allel frekansları gibi bazı veriler göz ardı edildiğinden gelen kritikler, birçok TNP lokusu kullanılsa da mesafe bazlı metotların popülasyon yapısını iyi şekilde belirlemek için uygun olmadığını söylemektedir. Model bazlı metotlar ise genellikle her popülasyon için Hardy-Weinberg dengesini kabul ederek, popülasyon parametrelerini tahmin etmek için standart istatistiksel yöntemler kullanmaktadır. Örnek miktarının azlığında hatalı allel frekans tahmini, sonuçların iyi olmamasına neden olabilmektedir (Gao & Starmer, 2007).

## **Hiyerarşik Kümeleme**

Hiyerarşik kümeleme metodu veri setinin birimlerinin birbirlerine olan uzaklık değerlerini kullanarak, veri setindeki birimlerin hiyerarşik ayrıştırmasını yapmaktadır. Hiyerarşik ayrıştırma sırasında, “ağaç veri yapısı” olarak da bilinen dendogram kullanılır. Dendogram, hiyerarşik kümeleme tekniğiyle elde edilen kümelerin görselleştirilmesini sağlar. Bir dendogramın yapısı kökler, iç düğüm ve yapraklardan oluşmaktadır (66).

**R** (Ihaka R., Gentleman R., 1996) istatistiksel hesaplama ve grafikleri için bilgisayar programı olup aynı zamanda programlama dilidir ve farklı işletim sistemleri için versiyonları bulunmaktadır (<http://www.r-project.org/>). **S** programlama dilinin açık kaynak kodlu versiyonu olan **R** bundan dolayı bazen **GNU S** olarak da anılmaktadır.

**R**'ın çok geniş istatistiksel analiz yelpazesi içerisinde klasik istatistik testleri, zaman serileri analizi, sınıflandırma, kümeleme gibi analizler ve grafik çizim teknikleri bulunmaktadır. **R** de **S** gibi tam anlamıyla programlama dili olarak tasarlanmış olup kullanıcıya yeni fonksiyonlar tanımlayarak ilave özellikler geliştirme imkânı sunmaktadır. **R**, kullanıcıların eklediği özel fonksiyonlar veya çok özel araştırma alanlarına ait paketlerle geliştirilebilir. **S** dilinin mirasından dolayı **R** diğer birçok istatistiki hesaplama dilinden daha kuvvetli bir nesneye yönelik programlama kabiliyetine sahiptir (67).

**R** kullanıcılar tarafından eklenen paketlerle geliştirilmektedir. Bu paketler belli uzmanlık alanlarıyla ilgili teknikler, gelişmiş grafik özellikleri, birçok harici dosya biçimini okuma/yazma yeteneği (SPSS, Matlab, Excel, MySQL bağlantıları vb.) gibi birçok özelliği **R**'ye kazandırmaktadır. **R**'nin standart kurulumunda temel paketler bulunuyor olsa da 4517 paket CRAN'den (Geniş Kapsamlı **R** Yazılım Ağı) erişilebilir (68).

**R** programının içerisinde farklı hiyerarşik kümeleme metodları bulunmaktadır. Bunlardan birtanesi olan öklid uzaklığı popülasyonların birbirlerinden uzaklıklarının hesaplandığı kümeleme yöntemidir. **R** programı cluster paketi içindeki daisy fonksiyonunda bulunan öklid uzaklığı; farklarının karelerinin toplamının karekökü olarak tanımlanmaktadır (69).

### **2.11.3. Haploview ile Tag TNP analizi**

Haploview genetik haplotip çalışmaları için geliştirilmiş java tabanlı açık kaynak programdır. Haploview haplotip analizlerini hızlı ve basit bir şekilde yürütebilmek

amacıyla geliştirilmiş kullanıcı arayüzü olan bir programdır. Haploview aşağıdaki fonksiyonları desteklemektedir.

LD ve haplotip blok analizleri;

- haplotip populasyon frekans tahminleri,
- tek TNP ve haplotip ilişkilendirme testleri,
- ilişkilendirme anlamlandırması için permutasyon testi,
- Paul de Bakker'ın Tagger uygulamalı tag TNP seçim algoritması,
- HapMap'den faz (phased) genotype verisinin otomatik indirilmesi.

İleri filtreleme seçeneklerini içeren PLINK'in genom boyunca veri yığınları ile tamamen uyumludur. Binlerce TNP'yi (komut satırı modunda on binlerce) binlerce bireyde analiz edebilir (70).

#### **2.11.4. GBAÇ Çalışmalarında Kullanılan Diğer İstatistiksel Analizler**

Popülasyon çalışmalarında popülasyon yapısını tespit edilmesinde ve ilişkilendirme çalışmalarında tabakalanmanın düzeltilmesinde güçlü bir araç olarak öne çıkan istatistiksel analizlerden birisi Partial Component Analizidir (PCA). Kümeleme algoritmaları ile birlikte uygulandığında popülasyon kümelerini anlamada ve bireyleri alt popülasyonlara atamada kullanılabilir. (70).

#### **2.11.5. Partial Component Analizi**

Popülasyon yapısının belirlenmesinde yaygın kullanılan model bazlı algoritmalarından biri "Structure"dır. Structure bireyleri başarılı bir şekilde kümelere ayırmaktadır. Fakat binlerce birey ve binlerce belirtecin olduğu verilerde çalışma zamanı çok uzun süre aldığından algoritma pratik olmamaktadır. Son zamanlarda klasik parametrik olmayan doğrusal hacimli (non-parametrik linear dimensionality) PCA tercih edilmektedir. Büyük hacimli veri setlerinde çalışmayı daha etkin kılabilmek için, özellik azaltım yoluyla verinin daha küçük hacimli olmasının amaçlandığı, veri madenciliğinden görüntü işleme kadar birçok alanda kullanılan Karhunen-Loève transform olarak da bilinen bir yöntemdir. Cavalli Sforza PCA ile ilk olarak 30 yıl önce varyasyon çalışması yürütmüştür (71). Daha sonra popülasyon yapılarının anlaşılmasında, ilişkilendirme çalışmalarında, tabakalanmaların düzeltilmesinde güçlü bir araç olduğu görülmüştür. Kümeleme



algoritmasıyla beraber kullanıldığında popülasyon kümelerinin anlaşılmasında ve alt popülasyonlara bireylerin atanmasında kullanılmaktadır (72). PCA popülasyon yapısının belirlenebilmesi için mikrosatelit, TNP yanı sıra haplotip sıklıklarına ve alu insersiyon polimorfizmlerine uygulanabilmektedir (73).

PCA'nın iki özelliği GBAÇ için önemlidir: (1) Büyük veri setleriyle hızlı çalışması (birkaç saat içerisinde yüz binlerce marker ve binlerce örnek ile çalışabilir). (2) Bütün bireyleri ayrı popülasyonlar veya lineer kombinasyonlar şeklinde sınıflandırmaya çalışmaz (bu şekilde olan sınıflandırmalar popülasyon tarifi açısından her zaman doğru bir model oluşturmamaktadır).

Veri madenciliğinde kullanılan PCA ile veri setinin boyutu düşürüldüğü için veri hassasiyetinin azalabilmektedir. Fakat yöntem, yapılan azaltımların veri üzerinde istenilen bir değere kadar mümkün olabileceği durumlarda işe yaramaktadır. Örneğin eğer veri temsili % 90'lara kadar yapılabiliyorsa, verinin azaltılması, setin küçültülmesi, kabul edilebilir düzeyde olmaktadır.

#### **2.11.6. K-Means (K-Ortalama)**

En eski kümeleme metotlarından biri olan k-means algoritmasının genel mantığı n adet veri nesnesinden oluşan bir veri setini, giriş parametresi olarak verilen k adet kümeye bölümlenektir. Amaç, gerçekleştirilen bölümlenme işlemi sonunda elde edilen kümelerin, küme içi benzerliklerinin maksimum ve kümeler arası benzerliklerinin minimum olmasını sağlamaktır. Küme benzerliği, kümenin ağırlık merkezi olarak kabul edilen bir nesne ile kümedeki diğer nesnelere arasındaki uzaklıkların ortalama değeri ile ölçülmektedir (Han ve Kamber, 2001; Berkhin, 2002).

K-means algoritmasının işlem basamakları şöyledir:

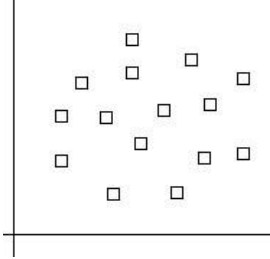
*1. Adım:* İlk küme merkezleri belirlenir. Bunun için iki farklı yol vardır. Birinci yol nesnelere arasından küme sayısı olan k adet rastgele nokta seçilmesidir. İkinci yol ise merkez noktaların tüm nesnelere ortalaması alınarak belirlenmesidir (**Şekil 2.4**).

*2. Adım:* Her nesnenin seçilen merkez noktalara olan uzaklığı hesaplanır, elde edilen sonuçlara göre tüm nesnelere k adet kümeden kendilerine en yakın olan kümeye yerleştirilmektedir (**Şekil 2.5**).

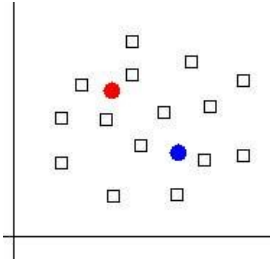
3. *Adım*: Oluşan kümelerin yeni merkez noktaları o kümedeki tüm nesnelerin ortalama değeri ile değiştirilmektedir (**Şekil 2.6**).

4. *Adım*: Merkez noktalar değişmeyene kadar 2. ve 3. adımlar tekrarlanmaktadır (**Şekil 2.7**).

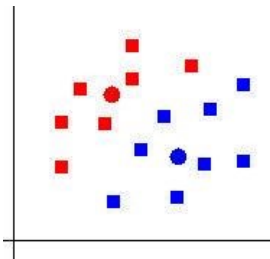
5. *Adım*: **Şekil 2.8**'deki şekilde örnek kümelerin sınıflandırılması güncellenir.



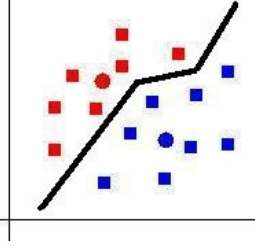
**Şekil 2.4.** Uzayda koordinatları kodlanmış örnekler için iki adet hedef küme tanımlanır (iki sınıf ve bu sınıfların karakterleri tanımlanmış olur)



**Şekil 2.5.** Bu sınıf tanımlarına uzaklıklarına göre (örneğin öklit mesafesi) bütün örnekler sınıflandırılır (hangi renge daha yakınsa)

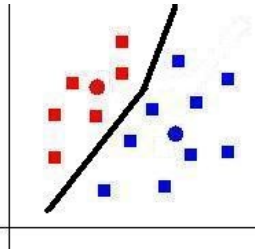


**Şekil 2.6.** Oluşan sınıfları ayıran bir hat çizilebilir



**Şekil 2.7.** Daha önceden sınıflandırılan örneklerin merkezleri bulunur

(yuvarlak ile gösterilen ve sınıf karakteristiğini temsil eden ilk örneklerin yerini değiştirmek şeklinde düşünülebilir).



**Şekil 2.8.** Merkezleri hareket ettirdikten sonra örneklerden bazıları yeni merkezlere daha yakın olabilir

K-means algoritmasının en büyük eksikliği  $k$  değerini tespit edememesidir. Bu nedenle başarılı bir kümeleme elde etmek için farklı  $k$  değerleri için deneme yanılma yönteminin uygulanması gerekmektedir (74).

### 3. GEREKÇE VE AMAÇ

İnsan genetik varyasyonları arasında en yüksek çözünürlüğe sahip olan TNP'lerin haplotipler oluşturmak suretiyle genomun global olarak taranması için önemli bir araç olduğu düşüncesinden yola çıkılarak, HapMap projesi ile farklı popülasyonları temsil ettiği düşünülen 1,301 örneğe ait TNP'ler haritalanmıştır. Halen devam eden çalışmada popülasyonlara özgü farklı haplotipler çıkartılmıştır. HapMap projesindeki popülasyon sayısının artması ve paralelinde devam eden projeler ile dbSNP'de kataloglanan TNP sayısı 2008 yılında (build 129 ) 11 milyon, Ocak 2013'de (build 137) ~63 milyon Nisan 2013 (build 138) itibariyle ise ~73 milyon olmuştur.

Yapılan araştırmalar ve dbSNP'deki kataloglanan varyasyonlardaki muazzam artış popülasyonlar arasındaki varyasyon farklılıklarının bir yansıması şeklindedir. Bugüne kadar yapılan uluslararası GBAÇ araştırmalarına konu olan örneklerin %96'sı Avrupalıdır ve sonuçlar Avrupa popülasyonunun kıtasal bir topluluk olduğuna işaret etmektedir. Çalışılan bu örneklem grupları göz önüne alındığında popülasyon profillerinin ortaya konmasında istatistiksel sapmaların devam edeceği öngörülmektedir. Bu hatadan kaçınmak için öncelikle örneklem gruplarının en küçük alt birime ayrılması daha sonra daha sonra grubun genişletilmesini gerekmektedir. (75; 76).

Günümüzde halen HapMap Proje verilerinin dünya popülasyonunun ne kadarını kadar temsil ettiği çalışma konusudur. Bu nedenle popülasyonlar kendi ülkelerindeki kompleks hastalıkların haritalanmasına, GBAÇ'larına, evrim çalışmalarına destek olmak ve kendi popülasyonlarının HapMap popülasyonları tarafından ne kadar temsil edildiklerini araştırmaya başlamışlardır. Bu amaçla, farklı popülasyonlar kendi varyasyon veri tabanlarını oluşturmak, haplotip haritalarını çıkartmak için kendi genom projelerini başlatmışlardır.

Günümüzde halen farklı çalışmalar ışığında ortaya konan popülasyon yapılarının Türk popülasyonunu ne derece temsil edildiği bilinmemektedir. Türk popülasyonuna özgü önemli varyantların seçilimi, Türk GBAÇ'larına destek olmak, Türk popülasyonunun HapMap popülasyonları tarafından ne derece temsil edildiğini görmek amacıyla memleket ve sağlık öyküleri bilinen daha önceki çalışmalardan elde edilmiş 227 sağlıklı 40 hasta örneğinin, Affymetrix GeneChip 250K TNP tarama sonuçlarının bir veri tabanında toplanması amaçlanmıştır. Böylece Türk popülasyonunun özgü ilk kez bir varyasyon veri

tabanı oluşturulması hedeflenmiştir. Biraraya getirilen ve Türk popülasyonunu temsil ettiği düşünülen verilerin HapMap Popülasyonları ile MAF farklılıklarını ortaya koymak suretiyle ağaç yapıları oluşturulacak ve popülasyonlar arasındaki tabakalanmalar ortaya konacaktır. Böylece Türk popülasyonunun HapMap popülasyonları tarafından ne derece temsil edildiğinin kantitatif olarak ortaya konması amaçlanmıştır.

Karşılaştırmaların yapılabilmesi için ilk olarak yüksek çözünürlüğe sahip olan HapMap popülasyon verilerinin kendi aralarında MAF ve ağaç yapıları ortaya konacaktır. İkinci olarak HapMap verileri Türk popülasyon verisi ile karşılaştırılabilecek şekilde düşük çözünürlüğe getirilecektir. Hapmap verilerinin Türk popülasyon verisi ile kıyaslanmasının yapılmasından önce yüksek çözünürlüklü HapMap verileri ile düşük çözünürlüklü HapMap verileri birbirleri ile hem MAF bakımından hem de ağaç yapılanması bakımından kıyaslanacaktır. Böylece veri çözünürlüğünün düşmesi ile verilerin gösterecekleri değişim izlenecektir. Böylece çalışmanın bu aşaması HapMap verileri ile Türk popülasyon verilerinin karşılaştırmaları için doğrulama basamağı olarak düşünülebilir. Bu tez çalışmasının son 3. aşamasında ise önce Türk popülasyon verisinin kendi içerisinde MAF ve ağaç yapısı değerlendirmeleri yapılacaktır. Çalışmanın 4. ve son aşamasında düşük çözünürlüklü HapMap verileri Türk popülasyon verileri ile kıyaslanacaktır. Bu sayede Türk popülasyonunun HapMap popülasyonları tarafından ne kadar temsil edildiği ortaya konmaya çalışılacaktır.

## 4. MATERYAL VE YÖNTEM

### 4.1. Materyal

Türk Popülasyon TNP verileri 267 unphased Affymetrix 250K Nsp ile analiz edilen genotip verilerini kapsamaktadır. Karşılaştırma yapılan HapMap popülasyon verileri HapMap verileri Broad, Sanger, Perlegen ve Sanger gibi genotipleme merkezleri tarafından Affymetrix: GenomeWideSNP 6.0:3, illumina:Human 1M BeadChip:3, Perlegen: Genotyping 1.0.0:2 gibi farklı çipler. ile yapılan çalışmalar trio (anne/baba/çocuk) örnek grupları ile yürütüldüğünden phased TNP genotiplerini içermektedir. Bu nedenle Affymetrix 250K Nsp çipinde taranan her referans dizi HapMap verisi ile örtüşmemektedir.

#### 4.1.1. HapMap Popülasyonları Yüksek Çözünürlüklü Verileri

HapMap popülasyon verileri HapMap projesi phases I+II (rel #24) ve III (hapmap3 r2) verisi hapmap web sitesinden indirilerek çalışma süresince kullanılmıştır ([http://hapmap.ncbi.nlm.nih.gov/downloads/genotypes/2009-02\\_phaseII+III/](http://hapmap.ncbi.nlm.nih.gov/downloads/genotypes/2009-02_phaseII+III/)). İndirilen 11 popülasyon TNP tarama sonuçları excel ortamında açılarak örnek ve varyasyon sayıları araştırılmıştır. HapMap popülasyonlarının birbirleri ile karşılaştırmalarının yapılabilmesi için hepsinde ortak olan varyasyonlar **Bölüm 4.2.7.**'de seçilmiştir. Sonuçta karşılaştırılan HapMap popülasyonlarının örnekleri ve varyasyon sayıları **Çizelge 4.1**'de verilmiştir.

**Çizelge 4.1.** Çalışmada Kullanılan Yüksek Çözünürlüklü HapMap Örnek ve Sayıları

HapMap Popülasyon Örnekleri	Popülasyon Grubu	Örnek Sayıları	Karşılaştırılan Toplam rs	Karşılaştırılan Toplam rs
ASW <sup>12</sup>	AFRİKA	83	1.560.998	<b>1.096.193</b>
LWK <sup>13</sup>		90	1.529.326	
MKK <sup>14</sup>		171	1.537.213	
YRI <sup>15</sup>		176	3.983.435	
JPT <sup>16</sup>	ASYA	89	4.048.499	
CHB <sup>17</sup>		86	4.051.413	
CHD <sup>18</sup>		85	1.305.798	
GIH <sup>19</sup>	ABD+AVRUPA	88	1.412.442	
MEX <sup>20</sup>		77	1.409.874	
CEU <sup>21</sup>		174	4.029.840	
TSI <sup>22</sup> : İtalya yaşayan Toskanalılar		88	1.419.526	
<b>Genel Toplam</b>		<b>1207</b>		

12 Güneybatı ABD'de Afrika kökenliler

13 Webuye'da Luhya kabilesi, Kenya

14 Kinyawa'da yaşayan Maasai'ler, Kenya

15 Ibida'da yaşayan Yorubalılar, Nijerya

16 Tokyodaki Japonlar, Japonya

17 Beijing'li Han Çinlileri, Çin

18 Denver metropolündeki Çinliler, Kolarado

19 Houston'da yaşayan Gujarati Hintlileri, Teksas

20 Los Angeles'da yaşayan Meksika kökenliler, Kaliforniya

21 CEPH koleksiyonundan Kuzey ve Batı Avrupa kökenli olup Utah'a yerleşenler

22 İtalya yaşayan Toskanalılar

#### 4.1.2. Türk Popülasyonu Düşük Çözünürlüklü Verileri

Bu tez çalışması için toplanarak Türk TNP veri tabanında biraraya getirilen 267 Türk popülasyon örneği bulunmaktadır. Memleket bilgilerine göre Türkiye (TR) coğrafyasında örneklerin sayıları **Çizelge 4.2**'de verilmiştir. Memleket bilgisine ulaşamayan 79 örnek ayrıca analiz edilmiştir.

**Çizelge 4.2.** Çalışmada Kullanılan Düşük Çözünürlüklü Türk Popülasyon Veri Sayıları

Bölge	Örnek Sayısı
Akdeniz 1B	32
Doğu Anadolu 2B	11
Ege 3B	8
Güneydoğu Anadolu 4B	26
İç Anadolu 5B	54
Karadeniz 6B	49
Marmara 7B	8
Bilinmiyor	79
Türk popülasyon verisinin tamamı	267



#### 4.1.3. HapMap ve Türk Popülasyonu Karşılaştırılabilir Düşük Çözünürlüklü Örnek ve Varyasyon Sayıları

Yüksek Çözünürlükteki HapMap verileri ile düşük çözünürlükteki TR popülasyon verilerinin karşılaştırılabilir varyasyonları **Bölüm 4.2.8**'de açıklandığı şekilde seçilerek, karşılaştırılan varyasyon sayıları **Çizelge 4.3**'de verilmiştir.

**Çizelge 4.3.** Çalışma Süresince Karşılaştırılan Düşük Çözünürlüklü HapMap ve Türk Popülasyonu Örnek ve Sayıları

	Örnek Sayısı	Grupların Örnek Sayısı	Toplam Örnek sayısı	Karşılaştırılan rs sayısı		
ASW	83	520	1207	1.096.193	HapMap	168.993
LWK	90					
MKK	171					
YRI	176					
JPT	89	260				
CHB	86					
CHD	85					
GIH	88	427				
MEX	77					
CEU	174					
TSI	88					
TR	267	267		267	261,961	

#### 4.1.4. Türk Popülasyonu Bölgelerinin kendi aralarında ve HapMap ile Karşılaştırılabilir Düşük Çözünürlüklü Örnek ve Varyasyon Sayıları

Yüksek Çözünürlükteki HapMap verileri ile düşük çözünürlükteki Türk popülasyon verilerinin karşılaştırılabilir varyasyonları yöntem bölümünde seçilmiş ve karşılaştırılabilecek varyasyon sayıları **Çizelge 4.4**'de verilmiştir.

**Çizelge 4.4.** Çalışma Süresince Karşılaştırılan Düşük Çözünürlüklü HapMap ve Türk Popülasyon Bölgelerinin Karşılaştırılan Örnek ve Varyasyon Sayıları

	Örnek Sayısı	Grupların Örnek Sayısı	Toplam Örnek sayısı	Karşılaştırılan rs sayısı		rs
ASW	83	520	1207	1.096.193	HapMap	168.531
LWK	90					
MKK	171					
YRI	176					
JPT	89	260				
CHB	86					
CHD	85					
GIH	88	427				
MEX	77					
CEU	174					
TSI	88					
Akdeniz-1B	32	188				
Doğu Anadolu-2B	11					
Ege-3B	8					
Güneydoğu Anadolu-4B	26					
İç Anadolu-5B	54					
Karadeniz-6B	49					
Marmara-7B	8					
Bilinmeyen	79		79			

## 4.2. Yöntem

### 4.2.1. Veri Toplama

Bu tez çalışmasında gerekli olan arařtırmaların yapılabilmesi için öncelikle Türkiye'ye ait bir TNP veri tabanının oluşturulması planlanmıřtır. Bu nedenle, Türkiye'de yürütölen ve kısmen devam eden farklı arařtırma projeleri sonucunda Affymetrix GeneChip Human Mapping ile yürütölen ve ulařılabilen 10K, 50K ve 250K ip seti TNP tarama sonuçları toplanmıřtır.

Bu tez çalışması için veri seilim kriterleri ařağıdaki řekilde belirlenmiřtir:

- Bireylerin hibirinin birbiriyle akrabalık iliřkisinin bulunmaması
- Analiz güven aralıęının %75'den fazla olması
- Yapılacak istatistiksel analizlerin anlamlılıęı aısından veri tabanına ilave edilen verilerin olabildięince erkek: kadın oranının 1:1 olması

Söz konusu alıřmalara dahil olan örneklerin TNP analizleri;

- Ankara Üniversitesi Merkez Laboratuvarında
- Hacettepe Üniversitesi Moleküler Pediatrik genetik laboratuvarında yapılmıřtır.

Toplanan verilerin iřlenebilir ve karřılařtırılabilir hale getirilmesi gerekmektedir. Bu nedenle \*.cel dosyası olarak mikrodizin cihazından alınan 10, 50 ve 250K TNP analiz sonuçları GeneChip Genotyping Analysis Software (GType) yardımıyla önce \*.chp dosyasına arkasından \*.txt formatına dönüřtürölmüř ve kaydedilmiřtir.

### 4.2.2. Veri Tabanında Saklanacak Verilerin İřlenmesi ve Veri Yıęınından Seilimi

Türk TNP veri tabanının oluşturulması için arařtırmalardan elde edilen veriler eřitli GBA arařtırmaları sonucunda elde edilmiřtir. Örneklerin TNP taramaları Affymetrix 10K, 50K ve 250K TNP ipler ile gerekleřtirilmiřtir. Affymetrix 10K (10,204 TNP), 50K (58,959 TNP), 250K (262,264 TNP) ile farklı sayılarda yaygın TNP taraması yapılmaktadır. Bu nedenle, tüm iplerin bir arada toplanmasının ne derece anlamlı olduęunu görmek için iplerde ortak taranan TNP'lerin, ip etiketleri ve dbSNP referans dizilerin (rs) sırasıyla karřılařtırılmıřtır.

Genom Boyu TNP analizlerinde kullanılan 3 farklı çipin (10K, 50K, 250K) referans dizileri de (rs) karşılaştırılmıştır. Çiplerde; 50K-10K'da %12,344 (7278) 250K-10K'da %0,429 (1126), 250K-50K'da ise % 2,915 (7650) TNP ortak olarak tarandığı görülmüştür (Çizelge 4.5, Çizelge 4.6, Çizelge 4.7 ).

**Çizelge 4.5.** 10K-50K rs örtüşme yüzdesi

Çip Tipi	Çipdeki Toplam rs sayısı	10K-50K Çiplerde Örtüşen rs sayısı	Yüzde kapsam
10K	10,204	7278	%12,344
50K	58,959		

**Çizelge 4.6.** 10K-250K rs örtüşme yüzdesi

Çip Tipi	Çipdeki Toplam rs sayısı	10K-50K Çiplerde Örtüşen rs sayısı	Yüzde kapsam
10K	10,204	1126	%0,429
250K	262,264		

**Çizelge 4.7.** 50K-250K rs örtüşme yüzdesi

Çip Tipi	Çipdeki Toplam rs sayısı	10K-50K Çiplerde Örtüşen rs sayısı	Yüzde kapsam
50K	58,959	7650	%2.915
250K	262,264		

Affymetrix 10K, 50K ve 250K çiplerinin karşılaştırılması sonucunda genom boyunca çalışmalarda araştırılan ortak TNP'lerin azlığı istatistiksel sapmaya neden olacaktır. Bu nedenle veri tabanında sadece en çok TNP taraması yapılan 267 örneğin 250K TNP tarama sonuçlarının tutulmasına karar verilmiştir.

#### 4.2.3. Veri Tabanında Saklanacak Veri Kaynakları ve Sayıları

Veri tabanında biraraya getirilmesi uygun olan 250K örneklerinin kaynak ve sayıları **Çizelge 4.8'** de verilmiştir. Örneklerin memleket bilgisine göre bölgesel dağılımları ise **Çizelge 4.9'**de verilmiştir. Veri tabanı tasarım aşamasında detaylı olarak kodlanan veri bölgeleri [EK3]'de verilmiştir. Çalışma süresince bir araya getirilen 267 örneğin 79 tanesinin memleket bilgisine ise ulaşılamamıştır.

**Çizelge 4.8.** Veri Tabanına Aktarılan Türk Popülasyon Veri Sayıları ve Kaynakları

Proje Merkezi	Proje Sorumlusu	Toplam Örnek Sayısı
Ankara Üniv. Biyoteknoloji Enst. Merkez Lab.	Prof. Dr. Hilal ÖZDAĞ	138 (sağlıklı)
Hacettepe Üniv. Tıp Fak. Çocuk Sağlığı ve Hastalıkları ABD	Prof. Dr. Ayşe Nurten AKARSU	97 (73 sağlıklı/24 hasta)
Çukurova Üniv., Tıp Fakültesi	Prof. Dr. Ali Kemal TOPALOĞLU	25 (10 sağlıklı/15 hasta)
Ankara Üniv., Tıp Fakültesi	Prof. Dr. Mustafa TEKİN	7 (6 sağlıklı/1 hasta)
Toplam Türk Popülasyon Verisi		267

**Çizelge 4.9.** Veri Tabanına Aktarılan Türk Popülasyon Verilerinin Bölgesel Dağılımları

Bölge	Örnek Sayısı
Akdeniz 1B	32
Doğu Anadolu 2B	11
Ege 3B	8
Güneydoğu Anadolu 4B	26
İç Anadolu 5B	54
Karadeniz 6B	49
Marmara 7B	8
Bilinmiyor	79
TÜRK POPÜLASYON VERİSİNİN TAMAMI	267

#### 4.2.4. Veri Tabanı Mantıksal Tasarımı

##### *Veri tabanı Kayıt Boyu Hesaplanması*

Veri tabanının; hastalık tablosu, kişi tablosu, snp tablosu, örnek tablosundan oluşmasına karar verilmiştir. Tablo yapıları **Bölüm 4.2.5** veri tabanı fiziksel tasarımında detaylı olarak açıklanmıştır. Örnek verilerinin tamamen anonim olmasını sağlamak için örnek kodlarında örnek dönüşüm tablosu aracılığı ile tekrar değişimi de sağlanmıştır.

Eklenen kayıtların veri tabanında ve disk üzerinde ne kadar yer tuttuğu veri tabanı planlama etkinliğini görmek açısından hesaplanmıştır. Aynı zamanda disk üzerinde kaplanan alan veri tabanı performansının da bir göstergesidir. Bu nedenle veri tabanı kayıt boyundan yola çıkılarak veri tabanı tablo yapılarının disk üzerinde kaplayacağı alan hesaplanmıştır.

Veri tabanı kayıt boyu; tablodaki tanımlanan alanların (sayısal ve sayısal olmayan) uzunlukları toplamıdır.

*Tbl\_ornek tablosunun veri tabanında kapladığı alan:*

```
CREATE TABLE `tbl_ornek` (  
  
  `RefDizi` int(6) NOT NULL,  
  
  `ornek` varchar(6) NOT NULL,  
  
  `call` char(2) ALLOW NULL,  
  
  PRIMARY KEY (`RefDiz`,`ornek`),  
  
  KEY `fk_ornek` (`ornek`)  
  
  ) ENGINE=MyISAM DEFAULT CHARSET=latin5
```

***Tbl\_örnek tablosunun kayıt boyu;***

Tablosunun Kayıt boyu =6+6+2=14 karakterdir <sup>23</sup>.

***Tbl\_örnek tablosunun veri tabanında kaplayacağı alan;***

Kayıt Boyu X Kayıt Sayısı = Toplam Boyut

14X(262264)=3671696 Byte Yani hard disk üzerinde yaklaşık 3,5 Mb yer tutacaktır

***Tbl\_örnek tablosundaki kayıtların disk üzerinde kapladığı alan;***

267 kayıt'ın veri tabanında tuttuğu yer=örnek sayısı X toplam boyut kadardır.

267X3,5=934 Mb yaklaşık 1 Gb

***Tbl\_hastalık tablosunun veri tabanında kapladığı alan:***

```
CREATE TABLE `tbl_hastalik` (  
  `hastalik_id` varchar (8) NOT NULL,  
  `hastalik_adi` varchar(60) NOT NULL,  
  PRIMARY KEY (`hastalik_id`),  
  ) ENGINE=MyISAM DEFAULT CHARSET=latin5
```

***Tbl\_hastalık tablosunun kayıt boyu;***

Tablosunun Kayıt boyu=8+60=68 karakterdir.

***Tbl\_hastalık veri tabanında ve disk üzerinde kaplayacağı alan;***

Toplam Boyut = Kayıt Boyu X Kayıt Sayısı

68X(71) = 4828 Byte yani hard disk üzerinde yaklaşık 4,7 Kb yer tutacaktır.

---

<sup>23</sup> (1karakter=1 Byte , 8 bit = 1 Byte, 1karakter=1 Byte, 1024 Byte = 1 Kilo Byte, 1024 KB = 1 Mega Byte , 1024 MB =1 Giga Byte, 1024 GB = 1 Tera Byte)

***Tbl\_kisi tablosunun veri tabanında kapladığı alan:***

```
CREATE TABLE `tbl_ornek` (  
  
  `ornek` varchar (6) NOT NULL,  
  
  `hastalik_id` varchar(8) ALLOW NULL,  
  
  `ikametgah` varchar (6) ALLOW NULL,  
  
  `d_yeri` varchar (6) NOT NULL,  
  
  `d_yeri` int (9) ALLOW NULL,  
  
  `cinsiyet` int (2) ALLOW NULL,  
  
  `call_rate` decimal (6) ALLOW NULL,  
  
  PRIMARY KEY (`ornek`),  
  
  KEY `fk_ornek` (`ornek`, hastalik_id)  
  
) ENGINE=MyISAM DEFAULT CHARSET=latin5
```

***Tbl\_kisi tablosunun kayıt boyu;***

Tablosunun Kayıt boyu =6+8+6+6+9+2+6=43 karakterdir.

***Tbl\_kisi tablosunun veri tabanında ve disk üzerinde kaplayacağı alan;***

Kayıt Boyu X Kayıt Sayısı = Toplam Boyut

43X(268)=11524 Byte. Yani hard disk üzerinde yaklaşık 11 Kb yer tutacaktır.



***Tbl\_snp tablosunun veri tabanında kapladığı alan:***

```
CREATE TABLE `tbl_snp` (  
  
  `RefDizi` int(6) NOT NULL,  
  
  `snp_id` varchar(15) NOT NULL,  
  
  `allele` char(10) ALLOW NULL,  
  
  `phys_pos` int(10) NOT NULL,  
  
  `db_snp` varchar(12) NOT NULL,  
  
  `chrom` char(2) ALLOW NULL,  
  
  PRIMARY KEY (`RefDiz`),  
  
  KEY `fk_RefDizi` (`ornek`)  
  
  ) ENGINE=MyISAM DEFAULT CHARSET=latin5
```

***Tbl\_snp tablosunun kayıt boyu;***

Tablosunun Kayıt boyu =6+15+10+10+12+2=55 karakterdir.

***Tbl\_snp tablosunun veri tabanında ve disk üzerinde kaplayacağı alan;***

Kayıt Boyu X Kayıt Sayısı = Toplam Boyut

55X(262264)=14424520 Byte. Yani hard disk üzerinde yaklaşık 13 Mb yer tutacaktır.

Veri tabanında 267 veri tablo yapıları ile birlikte ~1 Gb'lık yer tutmaktadır. Bu da veri tabanının etkin bir şekilde tasarlandığının göstergesidir.

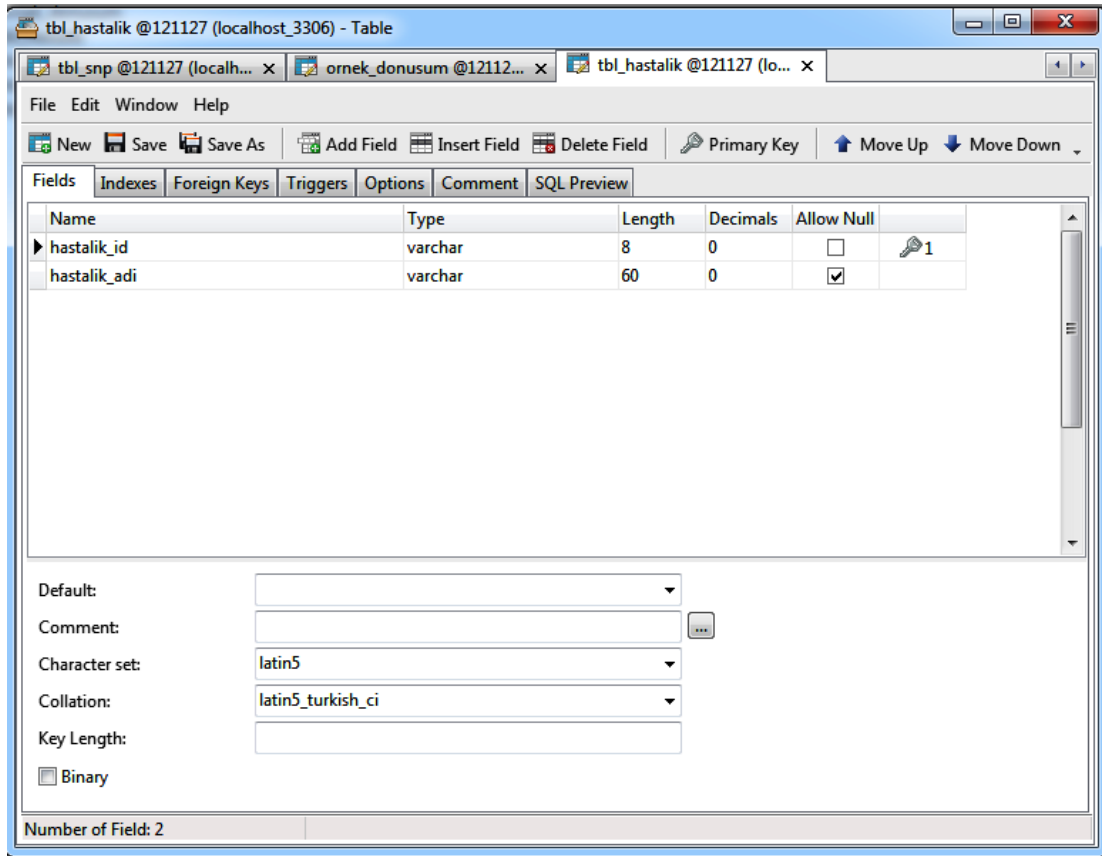
#### 4.2.5. Veri Tabanı Fiziksel Tasarımı

Uluslararası arařtırmalarda kullanılan veri tabanı aracı MySQL olduđundan bu tez alıřmasında; Materyal kısmında aıka tanımlanan TNP verilerinin bir araya getirilmesinde “MySQL 5.5” kullanılmıřtır.

##### *Hastalık Tablosu*

tbl\_hastalık rneklerin sađlık durumlarının tutulduđu ve tbl\_kisi tablosu ile iliřkili olan tablo yapısıdır (**řekil 4.1**). tbl\_hastalık tablo yapısında rneđin sađlık ve hastalık durumları heterozigotluk durumunu ierecek řekilde kodlanmıřtır.

Tabloda hastalık kodu “H” ile bařlar ve sonraki 2 sayısal basamak bireyin hastalıđının homozigotmu heterozigotmu olduđunu tanımlar. Eđer Kodlamanın ilk basamađı H0 ile bařlıyorsa ve arkasından 0’dan bařka bir sayı geliyor ise rnek sahibinin sađlıklı fakat ailesinde bilinen bir hastalık olduđunu gstermektedir. Eđer kodlamanın her iki basamađı da tekrar eden (0’dan farklı) sayılardan oluřuyorsa, rnek sahibinin hasta olduđunu gstermektedir. Kod H00 ise rnek sahibi tamamen sađlıklıdır.



**řekil 4.1.** tbl\_hastalik

## Kişi Tablosu

Örneklerin doğum yeri ve cinsiyet gibi temel bilgilerinin tutulduğu tablo yapısıdır (tbl\_kisi). Affymetrix 250K çiplerle analiz edilen her örnek için verilen bir isimdir (örnek), birincil anahtar (PK) olarak kişi tablosunda tutulmaktadır (**Şekil 4.2**). Kişi tablosunda yer alan ikametgah ve doğum yeri bilgileri fazla yer tutulmaması için kodlanmıştır. Kodlama Türkiye genelinde il/ilçe kodları kullanılarak, verilerin nüfus ve ikametgah bilgileri bölge/il/ilçe kodları şeklinde sırasıyla kodlanmıştır.

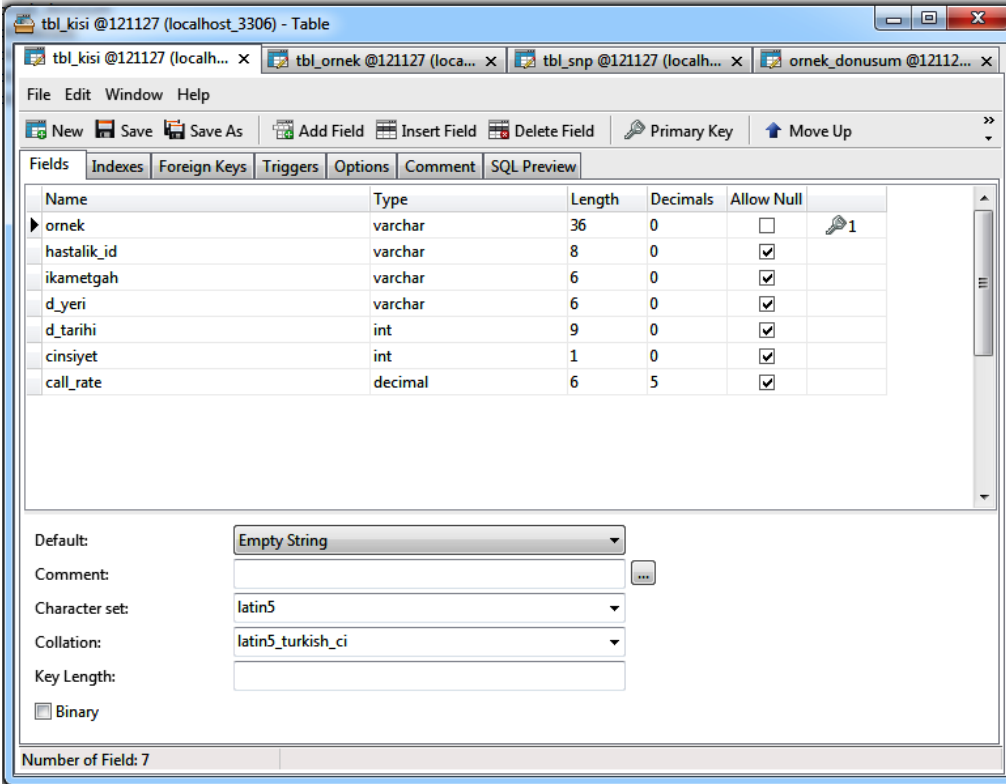
Örneğin;

İlk 2 karakter (Örn: 1B Akdeniz) Bölge ismini;

3-4. karakter il kodunu (örn: 01 (Adana));

5-6. karakter ise hastanın ilçesi biliniyor ise ilçeyi tanımlamaktadır.

Örneklerin ikametgah ve nüfus bilgilerinin ayrıca bir tabloda tutulmasının veri tabanının performansını düşüreceği ve veri tabanının disk üzerinde kaplayacağı alanı arttıracığı gerekçesiyle il/ilçe bilgilerinin veri tabanında bir tablo halinde kodlanmasına gerek olmadığına karar verilerek, gerektiğinde ulaşılabilmesi için Excel dosyasında saklanmıştır



Name	Type	Length	Decimals	Allow Null	
ornek	varchar	36	0	<input type="checkbox"/>	1
hastalik_id	varchar	8	0	<input checked="" type="checkbox"/>	
ikametgah	varchar	6	0	<input checked="" type="checkbox"/>	
d_yeri	varchar	6	0	<input checked="" type="checkbox"/>	
d_tarihi	int	9	0	<input checked="" type="checkbox"/>	
cinsiyet	int	1	0	<input checked="" type="checkbox"/>	
call_rate	decimal	6	5	<input checked="" type="checkbox"/>	

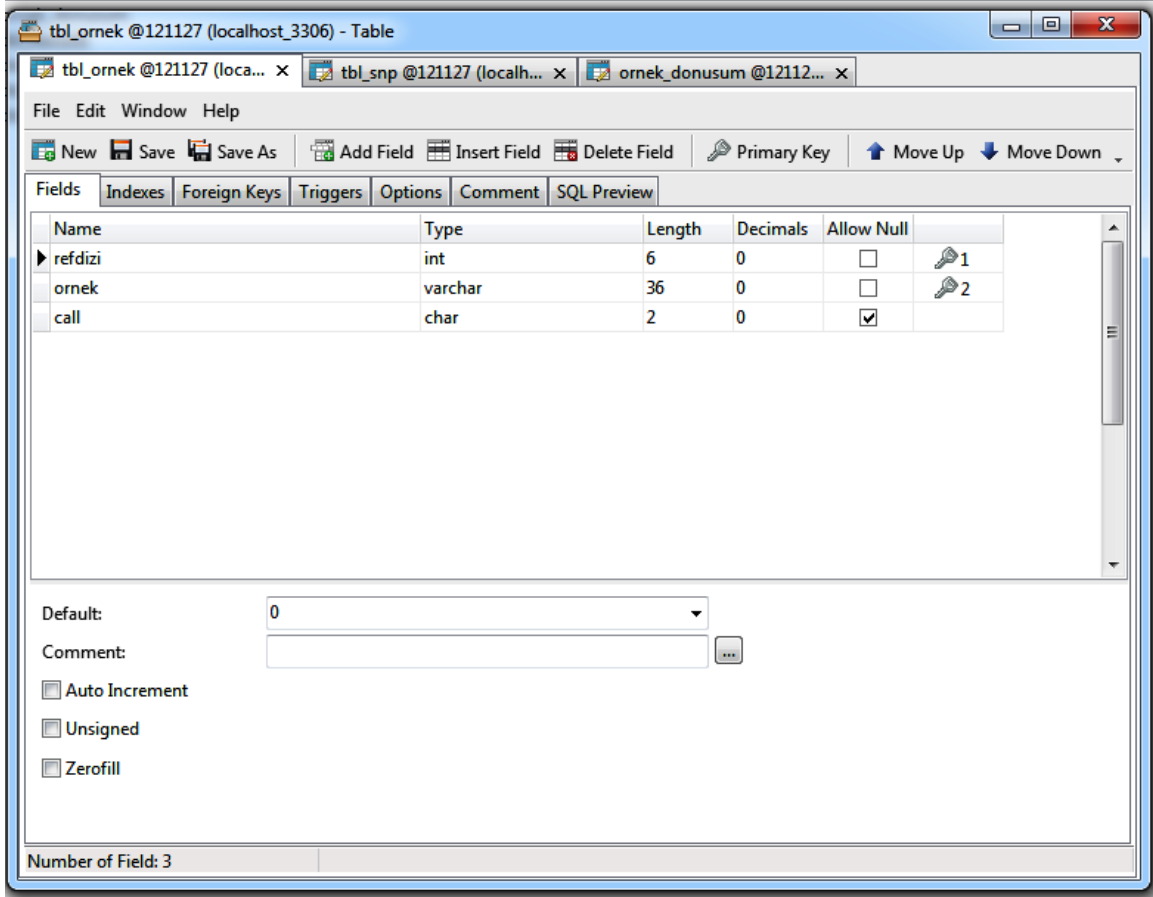
Default: Empty String  
Comment:   
Character set: latin5  
Collation: latin5\_turkish\_ci  
Key Length:   
 Binary

Number of Field: 7

Şekil 4.2. tbl\_kisi

## Örnek Tablosu

Örneklerin TNP analiz sonuçlarının tutulduğu tablo yapısıdır (tbl\_ornek). Kişi tablosuyla ilişkili tbl\_kisi.ornek ve TNP tablosuyla ilişkili tbl\_snp.refdizi bilgileriyle birlikte TNP analiz sonuçlarının tutulduğu tablo yapısıdır (Şekil 4.3).



Name	Type	Length	Decimals	Allow Null	
refdizi	int	6	0	<input type="checkbox"/>	1
ornek	varchar	36	0	<input type="checkbox"/>	2
call	char	2	0	<input checked="" type="checkbox"/>	

Default: 0

Comment:

Auto Increment

Unsigned

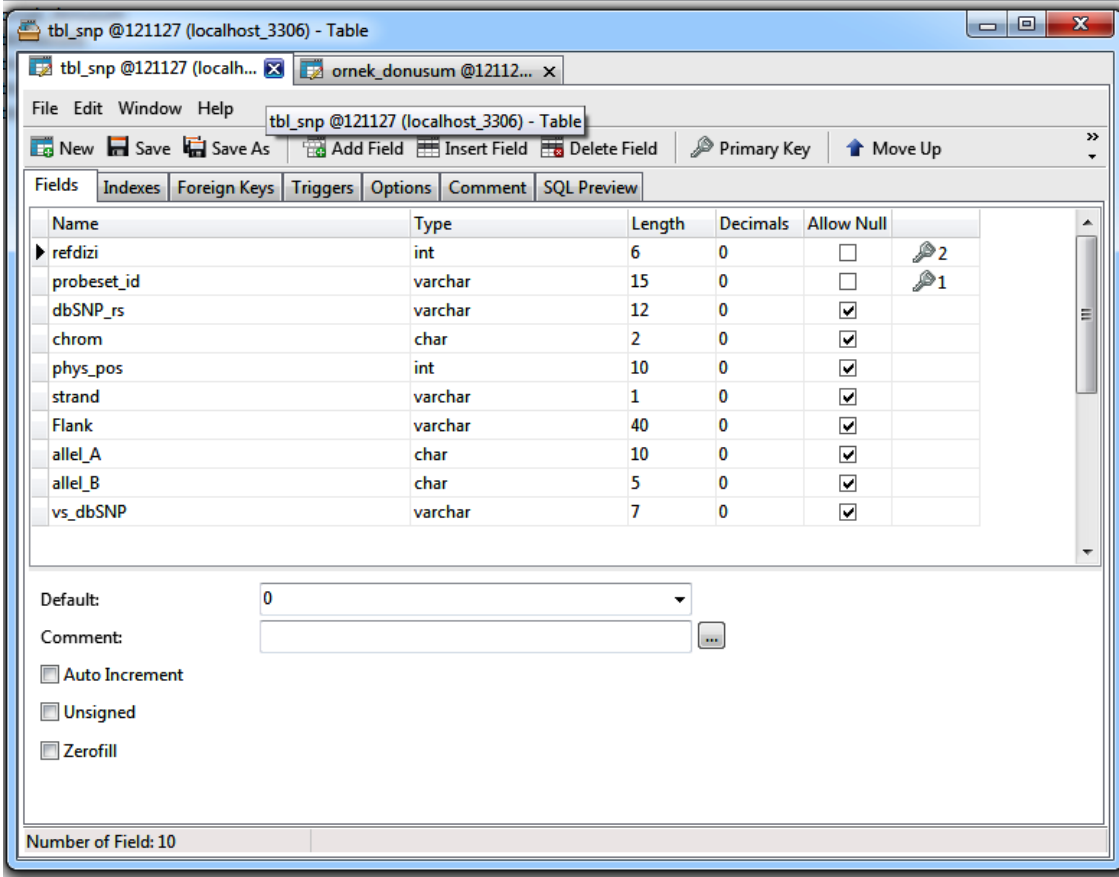
Zerofill

Number of Field: 3

Şekil 4.3. tbl\_ornek

## TNP Tablosu

Affymetrix 250K TNP çip verilerinin tutulduğu tablo yapısıdır (tbl\_snp). dbSNP’de etiketlenen her rs’e karşılık gelen AllelA, AllelB ve okuma değerinin tutulduğu tablo yapısıdır (Şekil 4.4).



Name	Type	Length	Decimals	Allow Null	Keys
refdizi	int	6	0	<input type="checkbox"/>	2
probeset_id	varchar	15	0	<input type="checkbox"/>	1
dbSNP_rs	varchar	12	0	<input checked="" type="checkbox"/>	
chrom	char	2	0	<input checked="" type="checkbox"/>	
phys_pos	int	10	0	<input checked="" type="checkbox"/>	
strand	varchar	1	0	<input checked="" type="checkbox"/>	
Flank	varchar	40	0	<input checked="" type="checkbox"/>	
allele_A	char	10	0	<input checked="" type="checkbox"/>	
allele_B	char	5	0	<input checked="" type="checkbox"/>	
vs_dbSNP	varchar	7	0	<input checked="" type="checkbox"/>	

Default: 0

Comment:

Auto Increment

Unsigned

Zerofill

Number of Field: 10

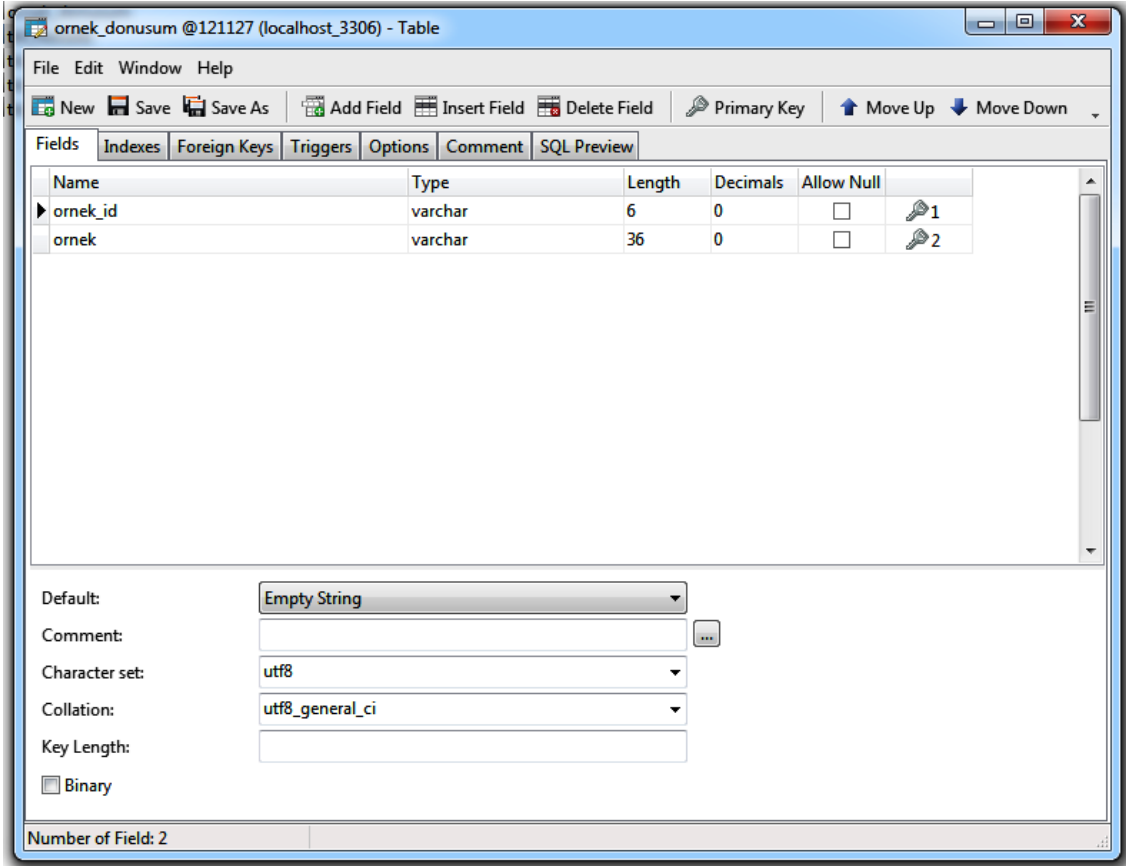
Şekil 4.4. tbl\_snp

Çalışmanın erken safhalarında, TNP tablosunda birincil anahtar olarak 1..n e kadar rastgele bir değer atanmıştır. Bunun yanında, çipte tanımlı TNP kimliği (SNP\_ID), Referans Dizi (rs), bu rs’e karşılık gelen AllelA/B, Fiziksel Pozisyon (PhysPos), dbSNPrs ve kromozom bilgisinin veri tabanında tutulması yeterli görüldü. Bu amaçla çipte okunan her rs’e karşılık gelen en güncel allelin veri tabanında saklanması için php program kodu yazıldı. Çalışma süresi internet erişimi ile sınırlı olan programın amacı ncbi veri tabanında bulunan en güncel rs değerlerine karşılık gelen allelleri arayıp bulmak ve snp.txt dosyasına yazmaktır. Açık php kodu verilen (index.php) program php 5.3.6 sürümünde kesintisiz olarak yaklaşık 5 gün boyunca çalıştırılarak ncbi veri tabanındaki en güncel Allel bilgileri çekildi [EK4].

Öngörüldüğü üzere ncbi web sayfasından kod ile indirilen AllelA/B bilgisinin ncbi tarafından sürekli güncellendiği için daha önceden tasarlanmış olan Affymetrix 500K chip ile %100 uyumlu olmadığı saptandı. Bu nedenle ileride yapılacak çalışmalara ışık tutabilmesi ve karışıklığı engellemek için Affymetrixin çipte basılı olan AllelA/B dizi bilgisinin de önemli olduğu kanısına varıldı. Bu sebeple çalışmanın ilk safhalarında tasarlanan tablo yapısındaki bilgilere ek olarak çipteki flanking dizi bilgisi, probesetid, dbSNP'deki karşılığı, dizi bilgisi eklenerek GeneChip® Human Mapping 500K array set çip bilgileri kullanıldı (77).

### ***Dönüşüm Tablosu***

Veri tabanında verilerin aktarılması esnasında oluşabilecek veri tekrarını engellemek için öncelikle veriler alındıkları analiz kodları ile veri tabanına eklenmiştir. Daha sonra veri kaynaklarına erişimi ortadan kaldırmak, örnekleri anonimleştirmek ve veri tabanında tutacağı yerin en aza indirgenmesi amacıyla analiz kodları için tekrar kodlama yapılmıştır (Şekil 4.5).



Şekil 4.5. ornek\_donusum

Bu amaçla, veri tabanında her analiz koduna O ile başlayan ve veri sayısı kadar devam eden (O1, O2...O267) ornek\_donusum tablosu oluşturulmuştur. Daha sonra tbl\_kisi ve tbl\_ornek tablolarında yer alan analiz kodlarının yenileri ile değişmesi için iki farklı sorgu yazılmıştır. Bu sayede, aşağıda açık kodu verilen **[Sorgu 1]** ve **[Sorgu 2]** kimlikleri bilinmeden veri tabanında analiz isimleri ile tutulan veri etiketleri, tbl\_kisi, tbl\_ornek tabloları için ayrı ayrı çalıştırılarak tekrar değiştirilmiştir. Daha sonra örnek dönüşüm tablosunun da veri tabanından silinerek veri tabanına aktarılan veriler anonimleştirilerek güvenliği sağlanmıştır.

**Sorgu 1.** tbl\_kisi tablosundaki örnek isimlerini değiştirir.

```
> update tbl_kisi, ornek_donusum  
> set tbl_kisi.ornek=ornek_id where tbl_kisi.ornek=ornek_donusum.ornek [sorgu 1:]
```

**Sorgu 2.** tbl\_ornek tablosundaki örnek isimlerini değiştirir.

```
> update tbl_ornek, ornek_donusum  
> set tbl_ornek.ornek=ornek_id where tbl_ornek.ornek=ornek_donusum.ornek [sorgu 2]
```

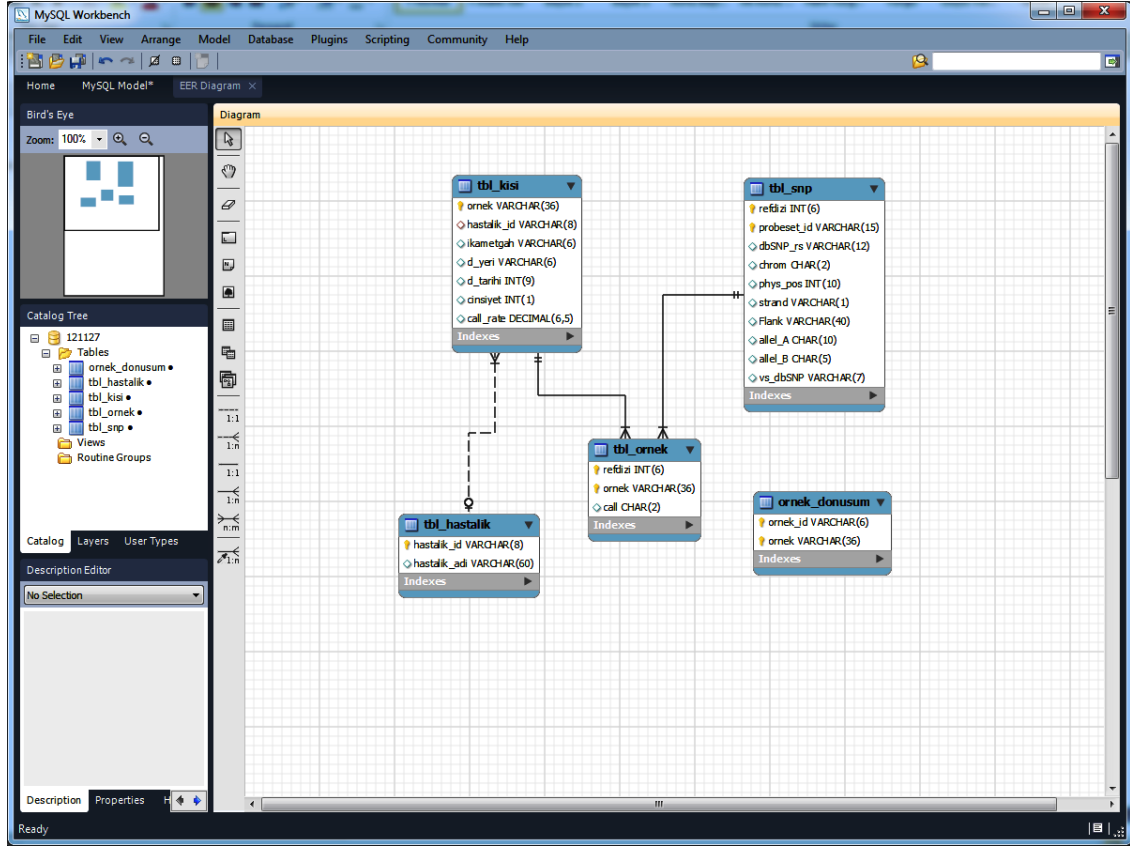
#### 4.2.6. Veri tabanı Araçları ile Türk TNP Veri Tabanı Oluşturulması

Mantıksal ve Fiziksel tasarımı tamamlanan ilişkisel Türk TNP Veri Tabanı yapısı **Şekil 4.6'**deki gibidir.

Veri işleme aşamasında toplanan verilerin veri tabanına aktarılabilmesi için veri tabanı ile uygun formata getirilmesi gerekmektedir. Materyal bölümünde detayları verilen 250K Nsp analizi sonucunda ortaya konan unphased TNP<sup>24</sup> verileri excel ortamında tablo yapılarına uygun olarak işlenerek Türk TNP veri tabanına aktarılmıştır.

---

<sup>24</sup> Phased veri; kromozom boyunca düzenli bir şekilde sıralanan veridir. Bu veriden haplotip okunabilir. Unphased veri ise basit anlamda bir çift kromozomdaki genotiplerdir, allel çiftlerinin hangi kromozomda olduğu bilinmemektedir.



Şekil 4.6. Türk TNP Veri tabanı

#### 4.2.7. Yüksek Çözünürlüklü HapMap Popülasyon Verilerinin Minör Allel Frekanslarının Hesaplanması

Allel frekanslarının hesaplanması ve ortak varyasyonların seçilimi için Matlab 7.8.0 (R2009A) programında çalıştırılmak üzere çeşitli algoritmalar yazılmıştır. Matlab’da HapMap popülasyonlarının allel frekansları Hardy-Weinberg dengesi gözününde bulundurulmuştur. Eğer  $f(AA)$ ,  $f(Aa)$  ve  $f(aa)$  bir lokusu temsil eden iki allelin üç farklı genotipi ise A allelinin  $p$  frekansı ve a allelinin  $q$  frekansı allelerin sayımı ile elde edilmiştir (**Bölüm 2.11.1**). Bir popülasyon için A allelinin toplam  $p$  frekansı **Denklem 26** ve a allelinin  $q$  frekansı **Denklem 27**’deki gibi hesaplanmaktadır.

$$p = f(AA) + 1/2f(Aa)$$

**Denklem 26**

$$q = f(aa) + 1/2f(Aa) \quad (10)$$

**Denklem 27**

11 HapMap popülasyonunun 23 kromozom için allel frekanslarının hesaplanması için **EK5** algoritması Matlab’da çalıştırılarak AllelA/B frekansları hesaplanmış ve Minör Allel



Frekansları seçilmiştir ( $MAF < 0,5$ ). Popülasyonlar arası allel frekanslarının karşılaştırmasının yapılabilmesi için 11 HapMap popülasyonunun bütün popülasyonlarda ortak olan varyasyonların seçilimi ile mümkündür. **EK6** algoritması Matlab'da çalıştırılarak ortak varyasyonların seçilimi gerçekleştirilmiştir. Sonuç olarak 11 HapMap popülasyonu için karşılaştırılabilecek 1.096.193 varyasyon olduğu ortaya konmuştur. 1.096.193 ortak varyasyon arasından  $MAF < 0,5$  seçilmiştir.

#### **4.2.8. HapMap ve Düşük Çözünürlüklü Türkiye Popülasyon Verilerinin Minör Allel Frekanslarının Hesaplanması**

Tez çalışmasının bu aşamasında 267 Türk popülasyon verisi ile **EK7** algoritması Matlab'da çalıştırılarak **Bölüm 4.2.7**'de açıklandığı şekilde allel frekansları hesaplanmıştır. TNP taramaları düşük çözünürlüklü olduğundan yüksek çözünürlüklü HapMap verileri ile karşılaştırılabilecek varyasyonlar **EK8** algoritması Matlab'da çalıştırılarak ortak varyasyonlar seçilmiştir. Sonuç olarak 11 HapMap popülasyonu ile 267 Türkiye verisinin karşılaştırılabilir 168.993 varyasyon olduğu ortaya konmuştur. 168.993 ortak varyasyon arasından  $MAF < 0,5$  seçilmiştir.

HapMap yüksek çözünürlüklü frekansları ile düşük çözünürlüklü frekansları karşılaştırılarak ortak Minör allel frekanslarındaki değişim araştırılmıştır.

#### **4.2.9. Türkiye'nin 7 Coğrafi Bölgesinin Minör Allel Frekanslarının Hesaplanması**

267 Türk popülasyon örneğinin 189 tanesinin Türkiye coğrafyasında memleket öyküsüne ulaşılmıştır (**Çizelge 4.2**). Türkiye 7 coğrafi bölgeye bölünerek 250K TNP taramaları için 7 coğrafi bölge verisi ile **EK7** algoritması Matlab'da çalıştırılarak **Bölüm 4.2.7**'de açıklandığı şekilde allel frekansları hesaplanmıştır. 250K verisinde karşılaştırılabilir çipteki kontroller çıkartıldığında toplam 261.961 varyasyon bulunmaktadır. 261.961 ortak varyasyon arasından  $MAF < 0,5$  seçilmiştir.

#### **4.2.10. HapMap Popülasyonlarının ve Türkiye'nin 7 Coğrafi Bölgesinin için Minör Allel Frekanslarının Hesaplanması**

Tez çalışmasının bu aşamasında memleket bilgisi bilinen 188 örneğin allel frekansları **EK7** algoritması Matlab'da çalıştırılarak **Bölüm 4.2.7**'de açıklandığı şekilde hesaplanmıştır.

HapMap verileri ile 7 coğrafik bölgenin karşılaştırılabilecek varyasyonları **EK8** algoritması Matlab’da çalıştırılarak ortak varyasyonlar seçilmiştir. Sonuç olarak 11 HapMap popülasyonu ile 7 coğrafi bölgenin karşılaştırılabilir 168.531 varyasyon olduğu ortaya konmuştur. 168.531 ortak varyasyon arasından  $MAF < 0,5$  seçilmiştir.

#### **4.2.11. HapMap Popülasyonları Yüksek Çözünürlüklü Verileri ile Hiyerarşik Kümeleme Analizi**

Verilerinin Hiyerarşik Kümeleme (Cluster) Analizi için araştırmacılar tarafından istatistiksel analizlerde kullanılan “R Project” programı seçilmiştir (78). R programı içinde hiyerarşik kümeleme analizinde kullanılan cluster paketi indirilerek çalışma için p frekansları ile kümeleme yapılmıştır. Kümeleme için “öklid uzaklığı” seçilmiştir. Öklid uzaklığını hesaplayabilmesi Öklid mesafesine göre hiyerarşik kümeleme yapmak için R Programı 2.15.3 (2013-03-01) versiyonu kullanılmıştır.

R programı cluster paketi içindeki daisy fonksiyonunda öklid uzaklığı; farklarının karelerinin toplamının karekökü olarak tanımlanmaktadır (69). Formül popülasyon verisine uyarlanırsa; iki popülasyon arası öklid uzaklığı, iki popülasyonun rs değerleri arasındaki farkların kareleri toplamının kareköküdür.

P1: popülasyon1

P2: popülasyon2 ise k tane “rs” olmak üzere

P1 ve P2 arası öklid uzaklığı formüldeki şekilde hesaplanmaktadır:

$$\text{öklid}_{(P1,P2)} = \sqrt{(rs_{p1,1} - rs_{p2,1})^2 + (rs_{p1,2} - rs_{p2,2})^2 + \dots + (rs_{p1,k} - rs_{p2,k})^2}$$

1.096.193 ortak HapMap varyasyonu (**Bölüm 4.2.7**) yüksek çözünürlüklü verilerle Hiyerarşik Kümeleme yapılabilmesi için Allel A frekansı kullanılmıştır. **EK9** algoritması R’de çalıştırılarak HapMap popülasyonlarının kendi içlerindeki ağaç yapısı ortaya konmuştur.

#### **4.2.12.HapMap Popülasyonları Düşük Çözünürlüklü Verileri ile Hiyerarşik Kümeleme Analizi**

HapMap Popülasyonları yüksek çözünürlükte ve düşük çözünürlükte karşılaştırmalarının yapılabilmesi amaçlanmıştır. 1.096.193 ortak HapMap varyasyonu için (**Bölüm 4.2.7**) Allel A frekansları kullanılmıştır. **EK9** algoritması **R**'da çalıştırılarak yüksek frekanslı veriler ile 11 HapMap popülasyonunun Bölüm 3.2.11'de açıklandığı şekilde hiyerarşik kümeleme analizi ortaya konmuştur. 11 HapMap ve Türk popülasyon karşılaştırması için seçilen 168.993 ortak varyasyonun (**Bölüm 4.2.8**) Allel A frekansları ile hiyerarşik kümeleme yapılmıştır. Yüksek frekanslı hiyerarşik kümeleme sonuçları ile düşük frekanslı hiyerarşik kümeleme sonuçları karşılaştırılmıştır.

#### **4.2.13. 7 Coğrafi Bölge Verileri (188) ile Hiyerarşik Kümeleme Analizi (Düşük Çözünürlük)**

Türk popülasyon verisinin 7 coğrafi bölgeye göre ağaç yapısının ortaya konması amaçlanmıştır. Bunun için 188 örneğin (**Bölüm 4.2.9**) 261.961 ortak varyasyonu için Allel A frekansları kullanılarak hiyerarşik kümeleme yapılmıştır. Bunun için **EK9** algoritması **R**'da çalıştırılarak Bölüm 4.2.11'de açıklandığı şekilde 7 coğrafi bölgenin ağaç yapısı ortaya konmuştur.

#### **4.2.14. 7 Coğrafi Bölgenin Verileri (188) ve Memleketi Bilinmeyen (79) veri ile Hiyerarşik Kümeleme Analizi (Düşük Çözünürlük)**

Memleket bilgisine ulaşılamayan 79 örneğin hiyerarşik kümelemede göstereceği davranışın araştırılması amacı ile coğrafi bölgesi bilinen 188 örnekle birlikte ağaç yapısının ortaya konması amaçlanmıştır. Bunun için 267 örneğin (**Bölüm 4.2.9**) 261.961 ortak varyasyonunun Allel A frekansları kullanılarak hiyerarşik kümelemesi yapılmıştır. Bunun için **EK9** algoritması **R**'da çalıştırılarak **Bölüm 4.2.11**'de açıklandığı şekilde 7 coğrafi bölgenin ağaç yapısı ortaya konmuştur.

#### **4.2.15. 11 HapMap Popülasyonu ve 7 Coğrafik Bölge Verisi (188) ile Hiyerarşik Kümeleme Analizi (Düşük Çözünürlük)**

11 HapMap ve 7 coğrafi bölgenin karşılaştırabilmesi için 168.531 ortak varyasyonun (**Bölüm 4.2.10**) Allel A frekansları ile hiyerarşik kümeleme yapılmıştır. Bunun için **EK9** algoritması **R**'da çalıştırılarak **Bölüm 4.2.11**'de açıklandığı şekilde HapMap popülasyonları ve 7 coğrafi bölgenin ağaç yapısı ortaya konmuştur.

#### **4.2.16. 11 HapMap Popülasyonu ve 267 Türk Popülasyon Verisinin ile Hiyerarşik Kümeleme Analizi (Düşük Çözünürlük)**

11 HapMap ve 7 coğrafi bölgenin karşılaştırabilmesi için 168.993 ortak varyasyonun (**Bölüm 4.2.10**) Allel A frekansları ile hiyerarşik kümeleme yapılmıştır. Bunun için **EK9** algoritması **R**'da çalıştırılarak **Bölüm 4.2.11**'de açıklandığı şekilde HapMap popülasyonları ve 7 coğrafi bölgenin ağaç yapısı ortaya konmuştur.

## 5. ARAŞTIRMA BULGULARI

### 5.1. Yüksek Çözünürlüklü HapMap Popülasyon Verilerinin MAF Karşılaştırması

Yüksek Çözünürlüklü HapMap popülasyon verilerinin Allel A/B frekansları **Bölüm 4.2.7**'de açıklandığı şekilde [EK5] algoritması kullanılarak  $MAF < 0,5$  seçilmiştir. 11 HapMap popülasyonunun web sitesinden indirildikten sonra MAF'ları hesaplanmış ve **Çizelge 5.1**'de gösterilmiştir. CHD popülasyonunda HapMap tarafından saptanan toplam varyasyon 1.305.798'dir. Bu varyasyonların seçilen MAF'lar **Çizelge 5.1**'de gösterilmiştir. En az toplam MAF 1.299.318 rs ile CHD popülasyonundadır. CHB popülasyonunda ise HapMap tarafından saptanan toplam varyasyon 4.051.413'dir. En fazla toplam MAF 4.033.107 ile CHB popülasyonundadır. Diğer HapMap popülasyonlarının saptanan varyasyon sayıları ve toplam minör allel bilgileri de **Çizelge 5.1**'dedir. Kromozoma göre yüksek çözünürlükteki MAF detayları ise **EK10-EK20**'de gösterilmiştir.

**Çizelge 5.1.** HapMap Popülasyonları Minör Allel Sayıları (Yüksek Çözünürlük)

	toplam rs	Allel A/B = 0,5 (Minör Allel Yok)	Minör Allel A	Minör Allel B	Toplam Minör Allel
ASW	1.560.998	7.129	777.780	776.089	1.553.869
LWK	1.529.326	7.017	761.550	760.759	1.522.309
MKK	1.537.213	3.463	766.971	766.779	1.533.750
YRI	3.983.435	9.045	1.989.137	1.985.253	3.974.390
JPT	4.048.499	18.337	2.016.741	2.013.421	4.030.162
CHB	4.051.413 (max)	18.306	2.016.626	2.016.481	4.033.107 (max)
CHD	1.305.798 (min)	6.480	649.850	649.468	1.299.318 (min)
GIH	1.412.442	6.882	705.154	700.406	1.405.560
MEX	1.409.874	7.378	700.914	701.582	1.402.496
CEU	4.029.840	9.693	2.010.437	2.009.710	4.020.147
TSI	1.419.526	6.934	705.349	707.243	1.412.592

HapMap popülasyonlarının MAF karşılaştırmalarının yapılabilmesi için ortak varyasyonlar **Bölüm 4.2.7**'de seçilmiştir. Bu seçim 11 HapMap popülasyonu karşılaştırılabilir toplam 1.096.193 varyasyonu olduğunu göstermiştir. HapMap popülasyonları dünya coğrafyasındaki konumlarına göre 1.096.193 varyasyonunun benzer MAF'ları karşılaştırıldığında **Çizelge 5.2**'deki gösterildiği üzere en az benzerliği %79,1 ile ABD+Avrupa grubu, en fazla benzerliği ise %90,13 ile Asya grubu göstermiştir. Popülasyonların coğrafi konumlarına göre neden bu şekilde dağıtıldığı **Bölüm 5.5**'de gösterilmiştir.

**Çizelge 5.2.** HapMap Popülasyonları MAF Karşılaştırması (Yüksek Çözünürlük)

	Karşılaştırılabilir rs	Minör Allel A	Minör Allel B	Toplam Minör Allel	% MAF
<b>AFRİKA (ASW, LWK, MKK, YRI)</b>	1.096.193	473.021	471.210	944.431	%86,16
<b>ASYA (JPT, CHB, CHD)</b>		494.388	493.558	987.946 (max)	%90,13
<b>ABD+AVRUPA (GIH, MEX, CEU, TSI)</b>		433.629	433.147	866.776 (min)	%79,1

HapMap popülasyon gruplarının 1.096.193 ortak varyasyonun MAF'larının benzerliği **Çizelge 5.3**'de verilmiştir. **Çizelge 5.3**'de görüldüğü üzere popülasyon gruplarını 2'li olarak değerlendirecek olursak en az MAF benzerliği %54,54 ile Afrika/Asya grubunda, en fazla MAF benzerliği ise %67,01 ile Asya/ABD+Avrupa gruplarında görülmektedir. Üç popülasyon grubu (Afrika/Asya/ABD+Avrupa-4A) beraber değerlendirildiğinde ise ortak MAF'lar %53,85'dir.

**Çizelge 5.3.** HapMap Popülasyon MAF'larının ikili ve üçlü Karşılaştırmaları (Yüksek Çözünürlük)

	Karşılaştırılabilir rs	Minör Allel A	Minör Allel B	Toplam Minör Allel	% MAF
<b>AFRİKA/ASYA</b>	1.096.193	354.403	353.109	707.512	%64,54
<b>AFRİKA/ ABD+AVRUPA</b>		337.954	337.048	675.002	%61,56
<b>ASYA/ ABD+AVRUPA</b>		367.580	366.991	734.571 (max)	%67,01
<b>AFRİKA/ASYA/ ABD+AVRUPA (4A)</b>		295.577	294.682	590.259 (min)	%53,85

## 5.2. Düşük Çözünürlüklü HapMap ve Türkiye Popülasyon Verilerinin Minör Allel Frekans Karşılaştırmaları

Düşük çözünürlüklü HapMap ve TR popülasyon verilerinin MAF karşılaştırmalarının yapılabilmesi için ortak rs'ler **Bölüm 4.2.8**'de seçilmiştir. Bu seçim verileri yüksek çözünürlükten (1.096.193 rs) düşük çözünürlüğe (168.993rs) getirmiştir. 11 HapMap popülasyonunun kendi içlerinde coğrafi bölgelerine göre **Çizelge 5.4**'deki düşük çözünürlüklü MAF'lar ile **Çizelge 5.1**'deki yüksek çözünürlüklü MAF'ları arasında belirgin bir farklılık olmadığı görülmüştür (**Çizelge 5.4**). Yapılan karşılaştırma veri çözünürlüğü azalsa da MAF benzerliklerinin değişmediğini göstermiştir. Yüksek çözünürlük verisine benzer şekilde ABD+Avrupa MAF benzerliği %79,1 ile en az, Asya grubunun MAF %90,13 ile benzerliği en fazladır.

Karşılaştırılabilir varyasyonların seçilimi TR verisi üzerinden yapıldığından TR popülasyonu kendi içinde benzer %99.81 MAF göstermektedir.

**Çizelge 5.4.** HapMap Popülasyonları ve TR popülasyonu MAF Karşılaştırmaları (düşük çözünürlük)

	Karşılaştırılabilir rs	Minör Allel A	Minör Allel B	Toplam Minör Allel	% MAF (Düşük çözünürlük)	% MAF (Yüksek çözünürlük)
AFRİKA (ASW, LWK, MKK, YRI)	168.993	72.881	73.073	145.954	% 86,37	%86.16
ASYA (JPT, CHB, CHD)		76.319	76.914	153.233	% 90,67	%90,13
ABD+AVRUPA (GIH, MEX, CEU, TSI)		67.074	67.878	134.952 (min)	% 79,86	%79,1
TR		82.090	86.587	168.677	%99,81	-

HapMap popülasyon gruplarının 168.993 varyasyon ile ikili ve üçlü MAF karşılaştırmaları için çözünürlük temel alındığında **Çizelge 5.5**'de görüldüğü gibi MAF değişimlerinde belirgin bir farklılık görülmemiştir. Sonuç olarak çözünürlük değişse de MAF'larda belirgin bir değişim olmaması Türk Popülasyonu verisi ile yapılan çalışmaların doğruluğunun göstergesidir. Düşük çözünürlükte benzer seyreden MAF'larda en az

benzerlik %62,69 ile Afrika/ ABD+Avrupa grubundadır. En fazla benzerlik ise %68,23 ile Afrika/ABD+Avrupa ikili grubundadır. Bu sonuçlar düşük çözünürlük sonuçları ile paraleldir. 4A grubunda ise ortak seyreden MAF'lar %55.21'dir.

**Çizelge 5.5.** HapMap Popülasyonlarının MAF'larının ikili ve üçlü Karşılaştırmaları (Düşük çözünürlük)

	Karşılaştırılabilir rs	Minör Allel A	Minör Allel B	Toplam Minör Allel	% MAF (Düşük çözünürlük)	% MAF (Yüksek çözünürlük)
<b>AFRİKA/ASYA</b>	168.993	54.283	54.647	108.930	%64,45	%64,54
<b>AFRİKA/ ABD+AVRUPA</b>		52.693	53.251	105.944 (min)	%62,69	%61,56
<b>ASYA/ ABD+AVRUPA</b>		57.348	57.962	115.310 (max)	%68,23	%67,01
<b>AFRİKA/ASYA/ ABD+AVRUPA (4A)</b>		46.342	46.884	93.226(max)	%55,21	% 55,17

Düşük çözünürlükteki HapMap popülasyon verileri ile TR verileri karşılaştırıldığında tüm karşılaştırma gruplarında benzer %MAF 4A'nın altına düşmüştür. MAF'ları karşılaştırılan gruplar arasında en az benzerlik %39.83 ile ABD+Avrupa/Türkiye, en fazla benzerlik % 44,95 ile Asya/Türkiye popülasyon grupları arasındadır. 4 popülasyon ortak analiz edildiğinde ise ortak MAF'lar 4A popülasyonlarının yaklaşık yarısı kadar azalmaktadır (Çizelge 5.6).

**Çizelge 5.6.** HapMap Popülasyonları ve Türk popülasyonu MAF Karşılaştırmaları (Düşük Rezolüsyon)

	Karşılaştırılabilir rs	Minör Allel A	Minör Allel B	Toplam Minör Allel	% MAF (Düşük rez)
<b>4A ORTAK</b>	168.993	46.342	46.884	93.226 (max)	% 55,17
<b>AFRİKA/TR</b>		35.339	37.315	72.654	% 43
<b>ASYA/TR</b>		36.807	39.152	75.959	% 44,95
<b>ABD+AVRUPA/TR</b>		32.512	34.794	67.306	% 39,83
<b>4 POPÜLASYON ORTAK</b>		22.495	23.983	46.478 (min)	%27,50



### 5.3. Türkiye'nin 7 Coğrafi Bölgesinin Minör Allel Frekanslarının Hesaplanması

Coğrafi bölgesi bilinen 188 örneğin 261.961 karşılaştırılabilir varyasyon için benzer MAF'lar karşılaştırıldığında (**Çizelge 5.7**) en az benzerlik % 95,38 ile Marmara bölgesindedir. En fazla benzerlik ise % 99,43 ile İç Anadolu bölgesindedir. Memleketi bilinmeyen örneklerin MAF benzerlikleri %99,66 ile örneklerin çoğunluğunun İç Anadolu bölgesine ait olduğunu düşündürmüştür (**Çizelge 5.7**).

**Çizelge 5.7.** Türkiye Coğrafi Bölgelerine Göre MAF Hesaplamaları

	Karşılaştırılabilir rs	Veri Sayıları	Minör Allel A	Minör Allel B	Toplam Minör Allel	% MAF
<b>Akdeniz-1B</b>	261.961	32	126.678	132.396	259.074	% 98,90
<b>Doğu Anadolu-2B</b>		11	123.353	129.800	253.153	% 96,64
<b>Ege-3B</b>		8	122.225	127.827	250.052	% 95,45
<b>Güneydoğu Anadolu-4B</b>		26	126.322	132.146	258.468	% 98,67
<b>İç Anadolu-5B</b>		54	127.382	133.072	260.454	% 99,43
<b>Karadeniz-6B</b>		49	127.107	133.073	260.180	% 99,32
<b>Marmara-7B</b>		8	122.158	127.707	249.865	% 95,38
<b>Bilinmeyen</b>		79	127.507	133.554	261.061	% 99,66

### 5.4. HapMap Popülasyonlarının ve Türkiye'nin 7 Coğrafi Bölgesinin MAF Karşılaştırmaları

11 HapMap popülasyonu ve TR coğrafi bölgesi bilinen 188 örnek için karşılaştırılacak ortak varyasyonlar Bölüm 3.2.15'de seçilmiştir 168.531 karşılaştırılabilir varyasyon için benzer MAF'ların karşılaştırma sonuçları **Çizelge 5.8**'de görüldüğü gibi TR bölgeleri birbirlerine göre oldukça yakın dağılım göstermiştir. TR bölgelerinde en fazla benzerliği % 99,07 ile İç Anadolu bölgesi göstermiştir. En az benzerlik ise % 93,78 ile Marmara bölgesindedir. HapMap popülasyonlarında ise en az benzerlik % 79,97 ile ABD+Avrupa popülasyon grubundadır. En fazla benzerlik ise % 90,68 ile Asya popülasyon grubundadır.

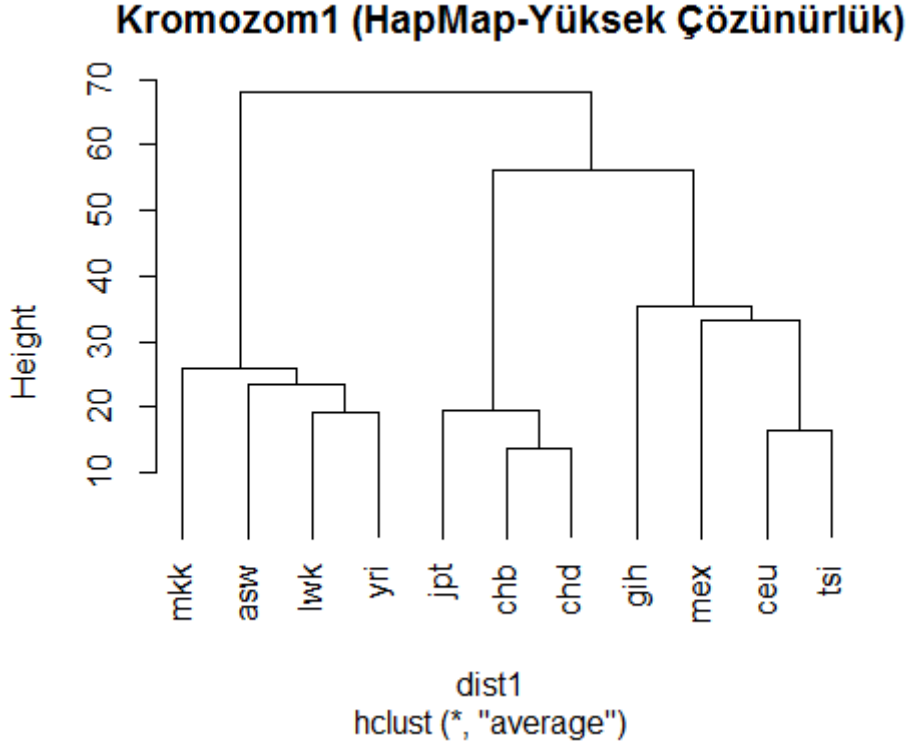
**Çizelge 5.8.** HapMap ve Türkiye Coğrafi Bölgelerine Göre MAF Karşılaştırmaları (Düşük Rezolüsyon)

		Minör Allel A	Minör Allel B	Toplam Minör Allel	% MAF
<b>AFRİKA</b>	168.531	72.710	72.812	145.522	% 86,35
<b>ASYA</b>		76.227	76.591	152.818	% 90,68
<b>ABD+AVRUPA</b>		66.998	67.775	134.773	% 79,97
<b>Akdeniz-1B</b>		80.726	85.050	165.776	% 98,37
<b>Doğu Anadolu-2B</b>		77.912	88.853	166.765	% 98,95
<b>Ege-3B</b>		77.014	81.172	158.186	% 93,86
<b>Güneydoğu Anadolu-4B</b>		80.398	84940	165.338	% 98,11
<b>İç Anadolu-5B</b>		81.383	85.574	166.957	% 99,07
<b>Karadeniz-6B</b>		81.108	85610	166.718	% 98,92
<b>Marmara-7B</b>		76.944	81.112	158.056	% 93,78

### 5.5. HapMap Popülasyonları Yüksek Çözünürlüklü Verileri ile Hiyerarşik Kümeleme Analizi

11 HapMap Popülasyonu yüksek çözünürlükte kümeleme sonrasında HapMap popülasyonlarının **Şekil 5.1**'de görüldüğü gibi göre Afrika (ASW, LWK, MKK, YRI), Asya (JPT, CHB, CHD) ve ABD+Avrupa (GIH, MEX, CEU, TSI) grupları oluşturduğu görülmüştür. Bu nedenle HapMap popülasyonlarının coğrafi bölgeleri kümeleme analizi sonucunda oluşturulmuştur. MAF karşılaştırmaları da bu coğrafi küme yapıları ile oluşturulmuştur.

Yüksek çözünürlüklü HapMap hiyerarşik kümeleme sonuçları Microsost Excel 2010 programının satır kapasitesini aşmıştır. Bundan dolayı 23 kromozomun ortak varyasyonları tek bir dosyada biraraya getirilememiş ve 23 kromozomun tamamının bütün olarak gösterildiği hiyerarşik kümelemesi yapılamamıştır. Kromozomlara göre yüksek çözünürlüklü hiyerarşik kümeleme sonuçları **EK28-EK50**'de gösterilmiştir.

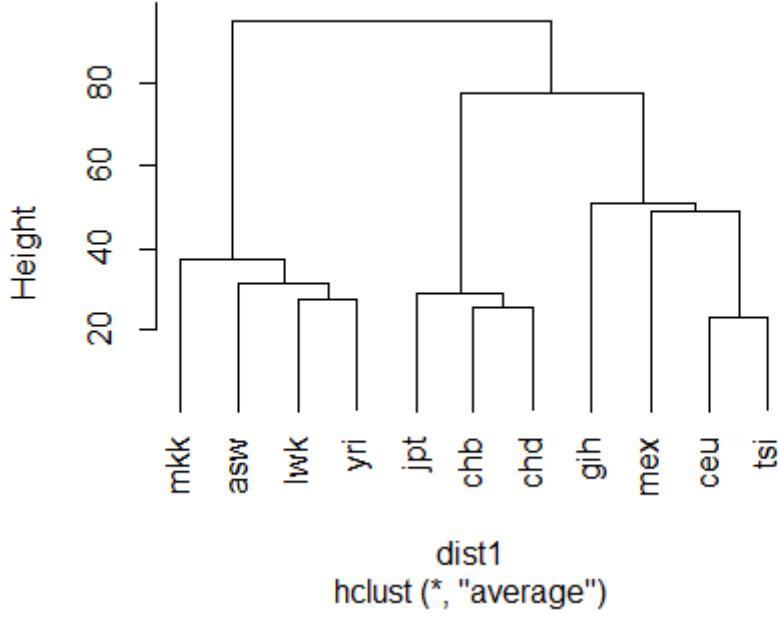


**Şekil 5.1.** 1. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Yüksek Çözünürlük)

### 5.6. HapMap Popülasyonları Düşük Çözünürlüklü Verileri ile Hiyerarşik Kümeleme Analizi

11 HapMap Popülasyon için 23 kromozom bütün olarak araştırılmıştır. **Şekil 5.2'**de görüldüğü gibi düşük çözünürlüklü kümeleme sonucunda popülasyonlar Afrika (ASW, LWK, MKK, YRI), Asya (JPT, CHB, CHD) ve ABD+Avrupa (GIH, MEX, CEU, TSI) grupları oluşturmuştur. HapMap yüksek ve düşük çözünürlüklü kümeleme sonuçları arasında farklılık oluşmamıştır. Kromozomlara göre düşük çözünürlüklü hiyerarşik kümeleme sonuçları **EK51-EK73'**de gösterilmiştir. Kromozoma göre yapılan analizler 23 kromozoma göre yapılan analizlerden belirgin farklılık göstermemiştir.

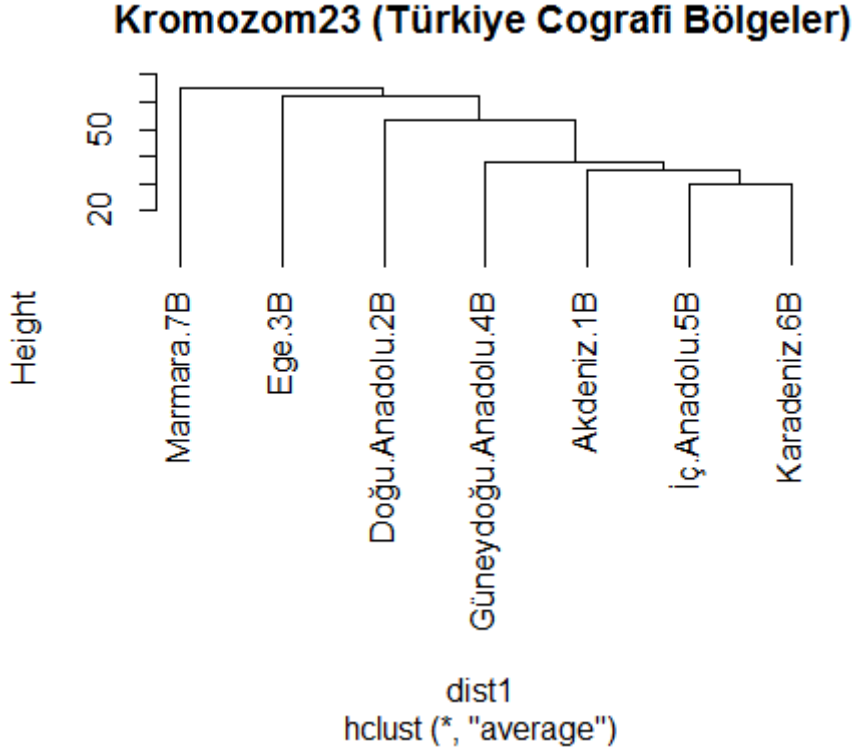
## Kromozom23 (HapMap Düşük Çözünürlük)



**Şekil 5.2.** Düşük Çözünürlüklü Veri ile HapMap Popülasyonları 23 Kromozom Hiyerarşik Kümeleme

### 5.7.7 Coğrafik Bölge Verileri (188) ile Hiyerarşik Kümeleme Analizi (Düşük Çözünürlük)

Düşük çözünürlükteki 23 kromozom için yapılan memleketi bilinen 188 örneğin hiyerarşik kümelemesi sonucunda **Şekil 5.3**'de görüldüğü üzere Karadeniz ve İç Anadolu Bölgesi aynı ağaç yapısında yer almıştır. Kalan 5 bölge basamak şeklinde bir ağaç yapısı oluşturmuştur. Kromozomlara göre Türkiye'nin coğrafi bölgelerine ait düşük çözünürlüklü hiyerarşik kümeleme sonuçları da **EK74-EK96**'de verilmiştir. Kromozoma göre yapılan analizler 23 kromozoma göre yapılan analizlerden farklılık göstermemiştir.

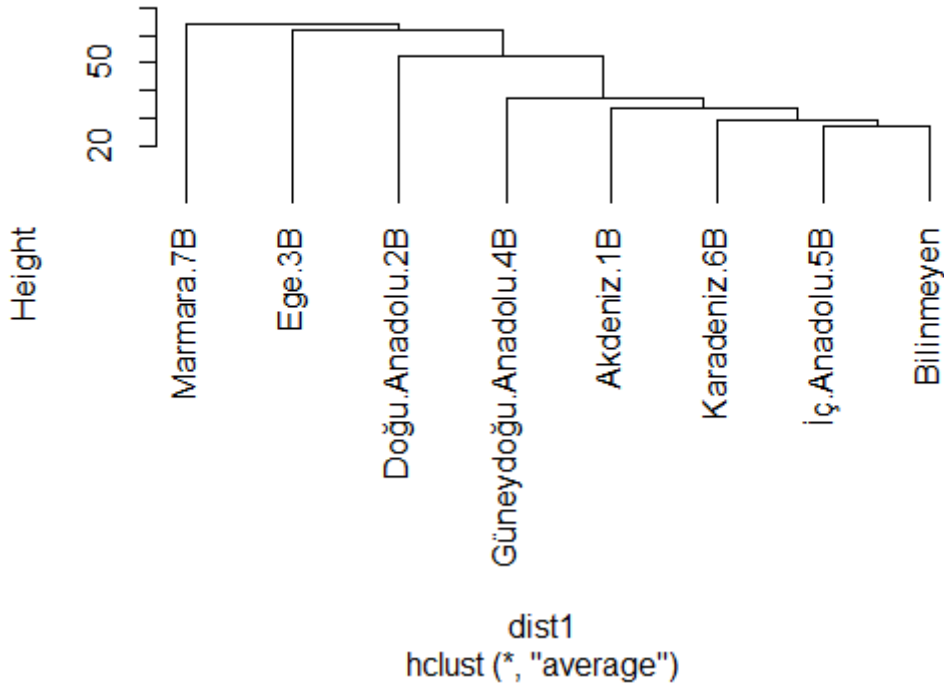


**Şekil 5.3.** 23 Kromozom Boyunca Türkiye 7 Coğrafi Bölgenin Hiyerarşik Kümelemesi

### **5.8. 23 Kromozom Boyunca Türkiye (Bölgesi Bilinmeyenler dahil) Hiyerarşik Kümeleme**

Bölüm 4.7’de TR coğrafi bölge verilerinin (188) hiyerarşik kümelemesine memleketi bilmeyen örnekler (79) eklenerek bilinmeyenler ayrı bir grup olarak araştırılmıştır. **Şekil 5.4**’de görüldüğü üzere memleketi bilinmeyen 79 örnek ayrı bir grup olarak eklendiğinde Türkiye’nin kümelenme yapısı değişmemiş ve memleketi bilinmeyen örnekler İç Anadolu bölgesi ile aynı ağaç yapısında kalmıştır. Bilinmeyen veri grubu yok iken **Şekil 5.3**’de İç Anadolu Bölgesi’ne en yakın Karadeniz Bölgesi olduğu görülmektedir.

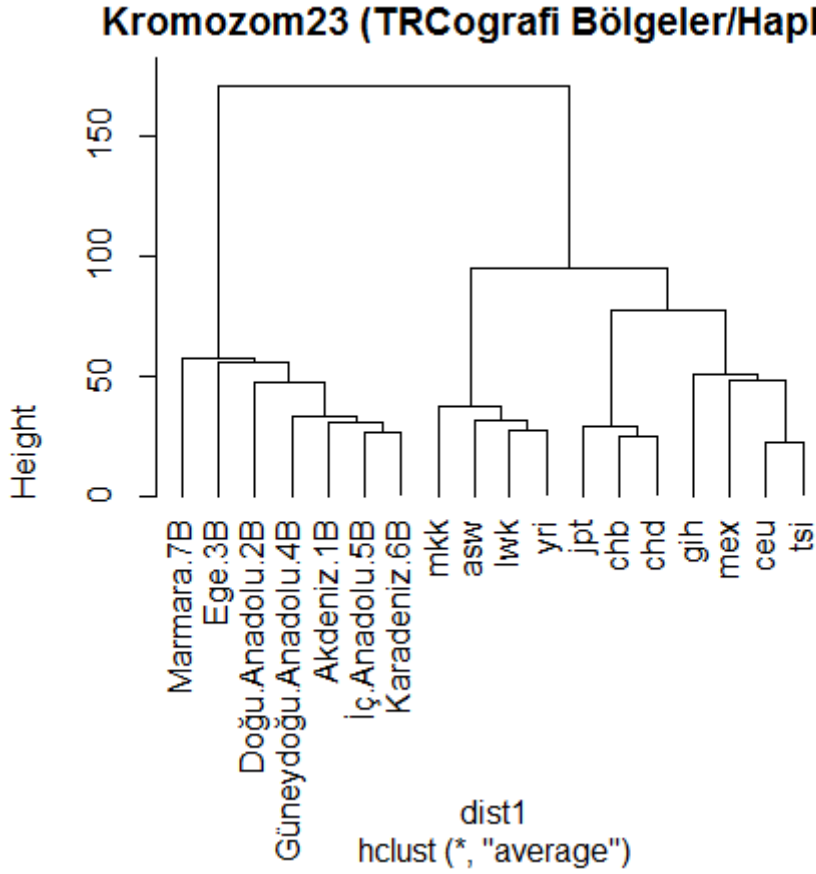
## Kromozom23 (Tr Coğrafi Bölgeler-Bilinmeyen dahil)



Şekil 5.4. 23 Kromozom Boyunca Türkiye (Bölgesi Bilinmeyenler Dahil) Hiyerarşik Kümeleme

### 5.9. 11 HapMap Popülasyonu ve 7 Coğrafi Bölge Verisi (188) ile Hiyerarşik Kümeleme Analizi (Düşük Çözünürlük)

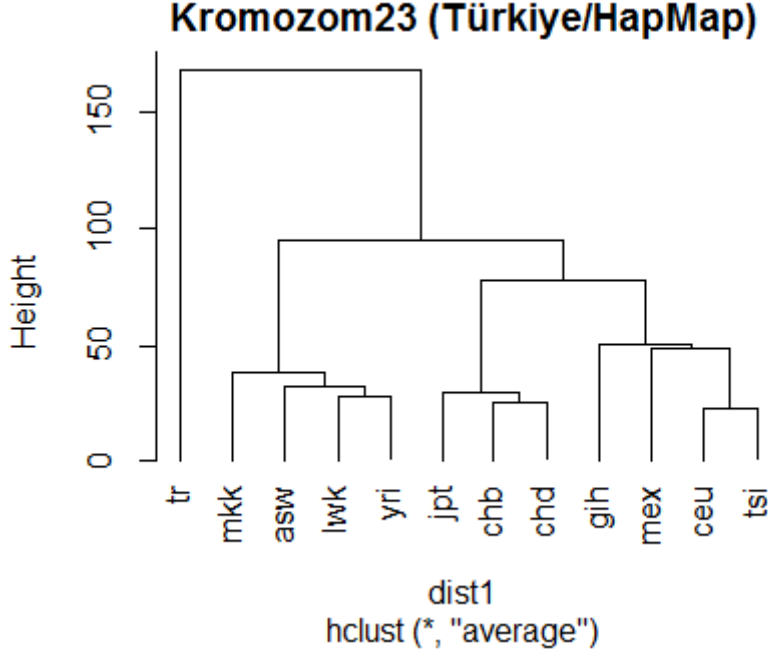
Türkiye 7 coğrafi bölgesi ve 11 HapMap popülasyonu verisinin 23 kromozom için kümeleme yapılmıştır. Şekil 5.5'deki düşük çözünürlüklü kümeleme sonucunda HapMap popülasyonları Afrika, Asya ve ABD+Avrupa gruplarını oluşturmuştur. Oluşan popülasyon grupları yüksek çözünürlüklü veriler ile yapılan hiyerarşik kümeleme sonucuyla aynıdır. TR bölgeleri kendi içlerinde ayrı bir küme oluşturmuştur. HapMap popülasyon gruplarında değişiklik olmaması Türkiye'nin oluşturduğu ayrı kümeyi doğrulamıştır. Kromozomlara göre Türkiye'nin coğrafi bölgeleri ve HapMap popülasyonlarının düşük çözünürlüklü hiyerarşik kümeleme sonuçları **EK97-EK119**'de verilmiştir.



**Şekil 5.5.** 23 Kromozom TR Bölgeleri ve HapMap Popülasyonları Hiyerarşik Kümeleme

### 5.10. 11 HapMap Popülasyonu ve 267 Türk Popülasyon Verisinin ile Hiyerarşik Kümeleme Analizi (Düşük Çözünürlük)

267 Türk popülasyon verisi ve HapMap verilerinin düşük çözünürlüklü analizinde Türk popülasyonu coğrafi bölgelerinde olduğu şekildedir. Türk popülasyonu HapMap popülasyonlarından ayrı bir küme yapısı oluşturmuştur. **Şekil 5.6**'da 23 kromozom bütün olarak araştırıldığında hem HapMap popülasyonlarının hem de TR'nin kümeleme davranışını değişmediği görülmüştür. Kromozomlara göre HapMap popülasyonlarının ve TR bölgelerinin hiyerarşik kümeleme sonuçları **EK120-EK142**'de verilmiştir.



**Şekil 5.6.** 23 Kromozom Türkiye ve HapMap Popülasyonları Hiyerarşik Kümeleme

Hapmap Popülasyonlarının Türkiye verisi ile kıyaslanmak üzere çözünürlüğü düşürüldüğünde küme davranışlarını değiştirmemiş, kümelerini korumuşlardır. Bu da çözünürlüğü düşük de olsa Türkiye verisinin genel itibari ile hiyerarşik kümelemede çözünürlük değişse de kümelerinin değişmeyebileceği düşüncesini vermiştir.



## 6. TARTIŞMA ve SONUÇ

İGP'den sonra gerçekleştirilen çözünürlüğü en yüksek olan varyasyonlardan TNP'lerin araştırıldığı HapMap merkezli çalışmalar popülasyon genetiğinin dünyadaki önemini gözler önüne sermiştir. Popülasyon çalışmalarından geniş kitlelerin fayda sağlayabilmesi için seçilen popülasyonun bütününün temsil ediliyor olmasının önemi araştırmacılar tarafından vurgulanmaktadır. Bu nedenle farklı popülasyonlar kendi popülasyonlarının HapMap tarafından ne derece temsil edildiklerini görmek, kendi popülasyonlarının varyasyon haritalarını çıkarmak böylece GBAÇ'larına destek olmak ve varyasyon veri tabanlarını araştırmacılarının kullanımına sunmak amacı ile genom projelerini başlatmışlardır.

Uluslararası platformlarda her geçen gün büyüyen devam eden bu çalışmalar ışığında birçok ülke devlet destekli projeler ile kendi araştırma gruplarını oluşturarak moleküler genetiğin avantajlarından yararlanmak amacıyla kendi popülasyonlarını, en küçük alt grupların seçilimi ile tanımlamaya çalışmaktadırlar. Maliyeti yüksek olan bu tür popülasyon çalışmalarının, özellikle bu amaca yönelik olarak henüz bir devlet desteği bulunmayan ülkemizde, önemi giderek farkedilmektedir.

Bu kapsamdaki çalışmaların önemine dikkat çekmek, değişik kaynaklardan toplanan veriyi popülasyon bilgisine dönüştürmek amacı ile yürütülen bu tez çalışması sayesinde Türk Popülasyonu için hem ilk kez bir varyasyon veri tabanı oluşturulmuş, hem de popülasyonumuza ait veriler bu kapsamda analiz edilmiştir. Bu tez çalışmasının ana yapısı üç aşamadan oluşmuştur. TNP veri tabanının oluşturulması, HapMap ve TR MAF'larının hesaplanması ve MAF'lar yardımı ile ağaç yapılarının oluşturulması yoluyla popülasyonlar arası ve popülasyon içi karşılaştırmalar yapılmıştır.

### 6.1. Çalışmada Kullanılan Verinin Yapısı ve Veri Tabanının Oluşturulması

Bu tez çalışmasının veri toplama aşamasında 4 farklı araştırma projesi kapsamında elde edilen 10K, 50K ve 250K verileri toplanmıştır. Toplanan 3 farklı çip verisinin Türk TNP veri tabanında saklanması amaçlanmış ve farklı çiplerin bilgilerinin bir algoritma ile seçiliminin yapılması planlanmıştır. Fakat 3 çipteki benzer varyasyonların karşılaştırılması sonucunda (**Çizelge 4.5-Çizelge 4.7**) taranan ortak TNP'lerin çok az olduğu görülmüştür.

Bu nedenle aynı veri tabanında 3 farklı çip bilgisinin saklanması istatistiksel sapmaya neden olacağı için bu çalışmadan vazgeçilmiştir.

Farklı çiplerin aynı veri tabanında saklanması uygun olmadığına karar verildikten sonra çözünürlüğü en yüksek olan 250K çip verilerinin veri tabanında saklanması karar verilmiştir. Çalışmanın sonucunda 4 farklı çalışmadan toplanan anonim 267 örneğin Affymetrix 250K TNP verisi Türk TNP veri tabanında saklanmıştır. 250K TNP veriseti veri tabanı HapMap'e göre çok düşük çözünürlüktedir. Fakat veri tabanının tasarımı yüksek çözünürlüklü verinin saklanması uygundur. Performans testleri yapılan TNP veri tabanı sorgulara hızlı cevap verebilmekte ve web arayüzü ile kullanılabilir. Aynı zamanda veri tabanı bilimsel gelişmelerdeki ihtiyaçlara cevap verecek şekilde tasarlanmıştır. Bugüne kadar Türkiye'de yapılan ilk TNP veri tabanıdır.

Çalışmanın veri tabanı tasarımı aşamasında Türk TNP veri tabanında TNP tablo yapısında saklanacak TNP bilgilerinin en güncel halinin kullanılması düşünülmüştür. Bu amaçla [EK4] çipte okunan her rs'ye karşılık gelen en güncel allelin veri tabanında saklanması için php program kodu yazılmıştır. Php 5.3.6 sürümünde kesintisiz olarak yaklaşık 5 gün boyunca çalıştırılan algoritma ile ncbi veri tabanındaki en güncel Allel bilgileri çekilmiştir. Fakat öngörüldüğü üzere ncbi web sayfasından kod ile indirilen AllelA/B bilgisinin NCBI tarafından sürekli güncellendiği için daha önceden tasarlanmış olan Affymetrix 500K chip ile %100 uyumlu olmadığı görülmüştür. Bu nedenle çalışmanın ilk safhalarında tasarlanan tablo yapısındaki bilgilere ek olarak çipteki flanking dizi bilgisi, probesetid, dbSNP'deki karşılığı, dizi bilgisi eklenerek GeneChip® Human Mapping 500K array set çip bilgileri kullanılmıştır (77).

## **6.2. Tag TNP Analizleri**

Planlanan çalışma sonucunda haplotiplerin ortaya konmasının haricinde tag TNP'lerin de belirlenmesi amaçlanmıştır. Bunun için HapMap popülasyonları için Hapview programı kullanılarak 23 kromozom için tag TNP'ler seçilmiştir. Fakat program deneme aşamasında olduğu için doğruluğu kontrol edilememiştir. Bu nedenle programın kullanımından vazgeçilmiştir. Tag TNP hesaplama çalışmasının 2. aşamasında yazılacak bir algoritma kullanılması planlanmıştır. Fakat bu sefer de çalışma verilerinde anne baba çocuk üçlüsüne (trio) ihtiyaç duyulmuştur. Çünkü TNP çiplerinde heterozigot okumalarda (AB veya BA) aynı allel sıraları seçilememektedir. Ancak tag TNP seçilimi için kromozom üzerinde

TNP'ler arasında AB ve BA arasındaki mesafenin belirlenmesi gerekmektedir. Bunun içinde A ve B allelerinin sırası önem kazanmaktadır. Örneğin araştırılan alleler A/B için A/T ise fazın AT veya TA olduğunun belirlenmesi sadece ebeveynlerin varlığı ile mümkün olmaktadır. Kısaca toplanan verilerin anne baba çocuk üçlüsü (trio) olmaması nedeni ile unphased veriler phased hale dönüştürülemediği. Sonuç olarak tag TNP hesaplaması expectation maximization algoritması ile yapılmaktadır. Bunun yanında bir de allel fazının tahmininin yapılacak olması tag TNP sonucunu gerçeklikten daha da uzaklaştıracağı için algoritma ile tag TNP hesaplamasından vazgeçilmiştir.

Bu tez çalışmasının bir diğer aşamasında okunamayan varyasyonların değerlendirilebilmesi için NoCall (NN) 3. Allel gibi kabul edilerek (A/B) frekansı hesaplanmıştır. Sonuçta 250K verisinde 8008 varyasyonun okunamadığı (NN) bulunmuştur. Bu okunamama durumunun veri kalitesinin ötesinde ilgili TNP'lerin poliallelik olmasına bağlı olduğu düşünülmektedir.

### 6.3. Popülasyonlar içi ve Popülasyonlar arası MAF karşılaştırmaları

HapMap Popülasyonlarının **EK5** ile hesaplanan ve **Çizelge 5.1**'de sonuçlarının verildiği MAF karşılaştırmalarına göre varyasyonların ~1.500.000 den ~4.000.000'a kadar değiştiği görülmüştür (**Çizelge 5.1**). 11 HapMap popülasyonuna özgü varyasyonların birbiri ile karşılaştırılabilmesi için ortak varyasyonların seçilimi **EK6** ile yapılmıştır. Sonuç olarak HapMap popülasyonları arasında web (79) sayfasından indirilen yüksek çözünürlüklü verilerden 1.096.193 varyasyonun karşılaştırılabilir olduğu görülmektedir (**Çizelge 5.2**).

Hiyerarşik kümeleme ile yapılan kümeleme analizi sonucunda HapMap popülasyonlarının **Şekil 5.1, Şekil 5.2**'de görüldüğü gibi göre Afrika (ASW, LWK, MKK, YRI), Asya (JPT, CHB, CHD) ve ABD+Avrupa (GIH, MEX, CEU, TSI) grupları oluşturduğu görülmüştür. Bu nedenle MAF karşılaştırmalarında hiyerarşik kümeleme grupları esas alınarak karşılaştırmalar yapılmıştır. Oluşan bu HapMap gruplarının ikili ve üçlü MAF karşılaştırmalarına göre 1.096.193 ortak varyasyonda (**Çizelge 5.3**) (yüksek çözünürlük) en az benzerliği %54.54 ile Afrika/Asya grubu, en fazla benzerliği ise %67.01 ile Asya/ABD+Avrupa grupları göstermiştir. Üç popülasyon grubu (Afrika/Asya/ABD+Avrupa-4A) birlikte değerlendirildiğinde ortak MAF %53.85'dir.

Yüksek çözünürlüklü HapMap verilerinin karşılaştırmalarının yapılmasından sonra 250K olan 267 TR popülasyon verisi için çalışmaya devam edilmiştir. 250K verisi içerisinde

kontroller hariç toplam 261.961 varyasyon araştırılmaktadır. HapMap verilerine göre düşük çözünürlüktedir. Yüksek ve düşük çözünürlükteki verilerin karşılaştırılabilmesi için HapMap ve TR MAF'ları **EK7** ile hesaplanmıştır. Daha sonra HapMap ve TR popülasyonlarının (12 popülasyon) **EK8** ile ortak varyasyonları seçilmiştir. **Çizelge 5.4**'de verilen 12 popülasyonun MAF karşılaştırmalarına göre ABD+Avrupa %79,1 ile en az, Asya grubunun %90,13 ile MAF benzerliği en fazladır.

TR ile kıyaslanabilmesi için düşürülen HapMap veri çözünürlüğü ile HapMap popülasyon gruplarının 168.993 varyasyon (düşük çözünürlük) için ikili ve üçlü MAF karşılaştırmaları yapılmıştır. Düşük çözünürlükte benzer MAF'lar en az %62,69 ile Afrika/ ABD+Avrupa grubundadır. En fazla benzerlik ise %68,23 ile Afrika/ABD+Avrupa ikili grubundadır (**Çizelge 5.5**). 4A grubunda ise ortak seyreden MAF'lar %55,21'dir. Bu sonuçlar yüksek çözünürlüklü verilerin MAF karşılaştırmaları ile benzer seyrettiğini göstermiştir. Sonuç olarak çözünürlüğün değişimi ile MAF'larda belirgin bir değişim olmaması Türk Popülasyonu verisi ile yapılan çalışmaların doğruluğunun göstergesidir. Bu sonuçlar düşük çözünürlük sonuçları ile paraleldir.

267 TR örneğinden 188 tanesinin memleketi bilinmektedir. Türkiye 7 coğrafi bölgeye ayrılarak coğrafi bölgeye göre MAF benzerlikleri **EK7** ile hesaplanmıştır. Daha sonra 7 TR coğrafi bölgenin (**EK3**) HapMap popülasyonları ile karşılaştırmasının yapılması içinde HapMap (11 popülasyon) ve TR 7 coğrafik bölge (7 popülasyon) için ortak varyasyonları **EK8** ile seçilmiştir. TR 7 coğrafik bölge analizine memleketi bilinmeyen 79 örnek bir grup olarak ilave edilerek örneklerin hangi coğrafi bölgeden olduğu tahmin edilmeye çalışılmıştır.

Coğrafi bölgesi bilinen 188 örneğin 261.961 ortak varyasyonu için benzer MAF'lar karşılaştırıldığında % 95,38 ile Marmara Bölgesi'nde en az ise % 99,43 ile İç Anadolu Bölgesi'nde en fazladır (**Çizelge 5.7**). Memleketi bilinmeyen örneklerde ayrı bir grup olarak araştırıldığında % 99,66 MAF benzerliği ile örneklerin çoğunluğunun İç Anadolu Bölgesi'ne ait olduğu sonucuna varılmıştır. Çalışmanın bir sonraki aşamasında 7 TR bölgesinin MAF'ları HapMap ile kıyaslanmıştır. TR bölgeleri birbirlerine göre oldukça yakın dağılım göstermiştir (**Çizelge 5.8**). TR bölgelerinde en fazla benzerliği % 99,07 ile İç Anadolu Bölgesi, en az benzerliği ise % 93,78 ile Marmara Bölgesi göstermiştir. HapMap popülasyonlarında ise en az benzerliği % 79,97 ile ABD+Avrupa popülasyon grubu en fazla benzerliği ise % 90,68 ile Asya popülasyon grubu göstermiştir.

#### 6.4. Kümeleme Analizleri

Kümeleme analizleri MAF karşılaştırmalarında olduğu gibi yüksek ve düşük çözünürlük olmak üzere iki basamaklı olarak gerçekleştirilmiştir. Bu sayede HapMap popülasyon gruplarının değişim gösterip göstermediği araştırılmıştır.

[EK9] algoritması kullanılarak kümeleme analizleri yapılmıştır. Çalışmada yapılan istatistiksel araştırmalar verinin gücüne göre olabildiğince yüksek çözünürlükte yürütülmüştür. 11 HapMap popülasyonu için EK6 ile seçilen 1.096.193 karşılaştırılabilir varyasyonun frekansları kullanılarak HapMap popülasyonlarının yüksek çözünürlüklü ağaç yapısı oluşturulmuştur. 11 HapMap popülasyonu Afrika (ASW, LWK, MKK, YRI), Asya (JPT, CHB, CHD) ve ABD+Avrupa (GIH, MEX, CEU, TSI) grupları oluşturmuştur. HapMap popülasyonlarının 267 TR verisinin varyasyonlarının karşılaştırılabilmesi için EK8 ile seçilen 168.993 varyasyon ile 11 HapMap popülasyonunun düşük çözünürlüklü ağaç yapısı da oluşturulmuştur. Ağaç yapıları yüksek çözünürlük için EK28-EK50'de düşük çözünürlük için EK51-EK73. 23 kromozom için ayrı ayrı incelendiğinde ağaç yapılarında farklılaşma görülmemiştir.

7 coğrafi TR bölgesinin 188 örnek ve memleketi bilinmeyen 79 örnek için EK8 ile seçilen 261.961 ortak varyasyon ile yapılan hiyerarşik kümeleme sonucunda İç Anadolu ve Karadeniz Bölgesi aynı ağaç yapısında kalmıştır (Şekil 5.4). Bilinmeyen veri grubu yokken İç Anadolu Bölgesi'ne en yakın bölge Karadeniz Bölgesi görülmektedir (Şekil 5.3).

Türkiye 7 coğrafik bölgesi ve 11 HapMap popülasyonu verisinin 23 kromozom için küme analizleri sonucunda (Şekil 5.5) HapMap popülasyonları Afrika, Asya ve ABD+Avrupa gruplarını korumuştur. TR bölgeleri kendi içinde basamak şeklindeki küme yapısını korumuş ve HapMap popülasyonlarından ayrı bir küme oluşturmuştur. HapMap popülasyon gruplarında değişiklik olmaması Türkiye'nin oluşturduğu ayrı kümeyi doğrular şekildedir. Kromozomlara göre Türkiye'nin coğrafi bölgeleri ve HapMap popülasyonlarının düşük çözünürlüklü hiyerarşik kümeleme sonuçları EK97-EK119'da verilmiştir.

267 Türk popülasyonu ile HapMap verilerinin ile 168.993 varyasyon kullanılarak yapılan kümeleme analizinde Türk popülasyonu HapMap popülasyonlarından 23 kromozomun

tamamı için ayrı bir küme yapısı oluşturmuştur (**Şekil 5.6**) Kromozomlara göre HapMap popülasyonlarının ve TR bölgelerinin hiyerarşik kümeleme sonuçları **EK120-EK142**'de verilmiştir. Sonuç olarak çalışma boyunca yapılan tüm kümeleme analizlerinde HapMap ve Türk popülasyonu kümeleme davranışlarını değiştirmemiştir.

## 6.5. Sonuç

Elde edilen kümeleme yapıları MAF dökümlerinin bir yansımasıdır ve popülasyonumuzun farklı bir genomik varyasyon yapısına sahip olduğunu göstermektedir. Bu çerçevede gelişmekte olan genomik yaklaşımlar ve kişisel tıp uygulamalarından yararlanılabilmesi için popülasyonumuza ait yüksek çözünürlüklü ve yüksek işlem hacimli genom verisine ihtiyaç olduğu açıktır.

Bu tez çalışmasında yapılamayan tag TNP analizleri trio verisinin varlığı ile yapılabilir. Türk popülasyonunun hazır çipler ile ne derece temsil edildiğinin bulunması açısından tag TNP analizleri çok önemlidir. Bu çalışmanın devamı olarak erişilebilen trio verisi ile tag TNP analizlerinin yürütülmesidir.

Çalışmada kullanılan genotipleme verileri incelendiğinde genotiplenen 267 bireyin %25 ve üstü genotiplenmemiş (NoCall) TNP sayısı toplam ~8000 TNP gözlenmiştir. Bu durum kullanılan çiplerdeki bir kısım TNP'lerin poliallelik olabileceği ihtimaline de işaret etmektedir. Bu denli yüksek bir genotiplenememe durumunun örnekler arası değerlendirme yapıldığında yalnızca hibridizasyon başarısı ile açıklanması mümkün görünmemektedir.

Çalışmanın son aşamasında Türk Popülasyonu HapMap popülasyonları ile öklid mesafesi kullanılarak hiyerarşik kümeleme analizleri **R** programında yapılmıştır. Türk Popülasyonunun hem bölgelere ayrılarak hem de bütün olarak yapılan genom boyunca hiyerarşik kümelemeler sonucunda HapMap'de analiz edilen Afrika Asya ve ABD+Avrupa gruplarından farklı bir grup oluşturduğu görülmüştür (**Şekil 5.5**, **Şekil 5.6**). **EK97-EK142**'de görüldüğü üzere bu durum kromozomdan kromozoma veya kromozomlar bir bütün olarak değerlendirildiğinde değişim göstermemektedir.

Sonuç olarak bu tez çalışmasında genomboyu bağlantı analizleri ve GBAÇ çalışmaları için kullanılan TNP mikrodizinlerinde kullanılan ve HapMap veriseti içinde yer alan 168.993 TNP için Türk Popülasyonu allel ve minör allel frekanslarını ortaya koyarken bu kapsamda çalışılan diğer popülasyonlar ile de karşılaştırılması gerçekleştirilmiştir. Çalışma sonunda

ortaya konan durum Türk Popülasyonunun kapsamlı olarak olarak çalışılan diğer popülasyonlara göre farklılıklar arzettiğini göstermektedir.

Bu sonuçlar Türk popülasyonunda yürütülecek gerek genomboyu bağlantı analizleri ve GBAÇ çalışmaları, gerekse farmakogenomik çalışmalar için popülasyona ait daha yüksek çözünürlükte, çok daha yüksek sayıda bireyin temsil edildiği ve daha yüksek işlem hacimli analizlerin gerekliliğini ilk kez bu denli kantitatif ve analitik olarak ortaya koymuştur.

## KAYNAKLAR

1. William, Klug S. Cummings, Micheal R. Genetik Kavramlar. Ankara, Palme Yayıncılık, c2003.288p.
2. Wu R, Lin M. Functional mapping – how to map and study the genetic architecture of dynamic complex traits. *Nat. Rev Genet.* March 2006; 7(3): 229–37.
3. Ewens, WJ. *Mathematical Population Genetics (2nd Edition)*. New York: Springer-Verlag.c2004.417p.
4. Butlin RK, Tregenza T. Levels of genetic polymorphism: marker loci versus quantitative traits. *Philos Trans R Soc Lond B Biol Sci.* 1998 Feb 28; 353(1366): 187–198.
5. Wetterbom A, Sevov M, Cavelier L, Bergström TF. Comparative genomic analysis of human and chimpanzee indicates a key role for indels in primate evolution; *J Mol Evol.* 2006 Oct 29;63(5): 682–90.
6. Scitable by Nature Education. Available at: <http://www.nature.com/scitable/topicpage/the-genetic-variation-in-a-population-is-6526354>. Accessibility verified September 13, 2013.
7. National Human Genome Research Institute. Available at: <http://www.genome.gov/Pages/Education/Modules/GeneticVariation.pdf>. Accessibility verified September 15, 2013.
8. Cole CG, McCann OT, Collins JE, et al. Finishing th finished human chromosome 22 sequence, *Genome Biology*, 2008 May 13; 9(5):R78.
9. Kidd JM Cooper GM, Donahue WF, et al. Mapping and sequencing of structural variation from eight human genomes. *Nature.* 2008 May 1:453(7191).
10. Zhang W, Duan S, Dolan ME. HapMap filter 1.0: A tool to preprocess the HapMap genotypic data for association studies. *Bioinformatics.* 2008 May 13; 2(8):322-324.
11. 1000 Genomes A Deep Catalog of Human Genetic Variation. Available at: <http://www.1000genomes.org/node/506>. Accessibility verified September 19, 2013.
12. The National Center for Biotechnology Information. Available at: [http://www.ncbi.nlm.nih.gov/projects/SNP/docs/rs\\_attributes.html](http://www.ncbi.nlm.nih.gov/projects/SNP/docs/rs_attributes.html). Accessibility verified September 28, 2013.
13. Botstein D, White RL, Skolnick M, Davis RW. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Genetics.* 1980 May;32(3):314-31.
14. Nakamura Y, Leppert M, O'Connell P, et al Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science.* 1987 Mar 27;235(4796):1616-1622.
15. Jeffrey A.J, Wilson V, Thein SL. Hypervariable 'minisatellite' regions in human DNA. *Nature.* 1985 Mar 07; 314: 67-73.
16. Cavalli-Sforza LL. The Chinese Human Genome Diversity Project. *Proc Natl Acad Sci U S A.* 1998 Sep 29;95(20):11501-1153.
17. Sebat J, Lakshmi B, Troge J, et al. Large-scale copy number polymorphism in the human genome. *Science.* 2004 Jul 23;305(5683):525-8.
18. Nakamura Y, Lathrop M, Bragg T, et al. An extended linkage map for human chromosome 10. *Genomics.* 1988 Nov;3(4):389-92.
19. Gusella JF, Wexler NS, Conneally PM, et al. A polymorphic DNA market genetically linked to Huntington's disease. *Nature.* 1983 Nov 17-23;306(5940):234-8.



20. Nakamura Y. .DNA variations in human and medical genetics:25 years of my experience. *J Hum Genet.* 2009 Jan;54(1):1-8.
21. Akarsu A N, Çakır B. Psikiyatrik genetik arařtırmalarda kullanılabilir genetik yöntemler: IV-A. Hastalık geni haritalanması. *3P Dergisi.* 2004 Mar;12 (Ek.1):31-49.
22. Aynacıođlu, AŞ. Farmakogenetik/Genomik arařtırmalar ve tıpta uygulamaları. Türk Farmakoloji Derneđi Eđitim Sempozyumu. Yüzüncü Yıl Üniversitesi Tıp Fakültesi Van. 2 Haziran 2006. Türk Farmakoloji Derneđi Eđitim Sempozyumu. Available at: [http://www.tfd.org.tr/eski/van2006\\_%20kitapcik.pdf](http://www.tfd.org.tr/eski/van2006_%20kitapcik.pdf). Accessibility verified September 27, 2013.
23. Schwartz M, Vissing J. Paternal Inheritance of Mitochondrial DNA. *N Engl J Med.* 2002 Aug 22;347(8):576-80.
24. Oak Ridge National Laboratory. Available at: [http://www.ornl.gov/sci/techresources/Human\\_Genome/elsi/humanmigration.shtml](http://www.ornl.gov/sci/techresources/Human_Genome/elsi/humanmigration.shtml). Accessibility verified February 16, 2009.
25. International HapMap Project. International HapMap Project Organization. Available at:<http://www.hapmap.org>. Accessibility verified February 16, 2009.
26. The International HapMap Consortium. A haplotype map of the human genome. *Nature.* 2005 Oct 27;437:1299–1320.
27. International HapMap Consortium, Frazer KA, Ballinger DG. A second generation human haplotype map of over 3.1 million SNPs. *Nature.* 2007 Oct 18;449(7164):851-61.
28. International HapMap 3 Consortium, Altshuler DM, Gibbs RA. Integrating common and rare genetic variation in diverse human populations. *Nature.* 2010 Sep 2;467(7311):52-8.
29. CNN Turk. Available at: <http://www.cnnturk.com/2008/yasam/diger/12/12/dunya.nufusu.beklenenden.400.milyon.daha.az/504423.0/index.html>. Accessibility verified September 25, 2013.
30. The Indian Genome Variation Consortium. The Indian Genome Variation database (IGVdb): a project overview. *Hum Genet.* 2005 Oct;118(1):1-11.
31. Yamaguchi-Kabata Y, Nakazono K, Takahashi A, et al. Japanese Population Structure, Based on SNP Genotypes from 7003 Individuals Compared to Other Ethnic Groups: Effects on Population-Based Association Studies. *Am J Hum Genet.* 2008 Oct;83(4):445-56.
32. Tartu Biotechnology Park. Available at: [http://www.biopark.ee/documents/Estonian\\_genome\\_project.pdf](http://www.biopark.ee/documents/Estonian_genome_project.pdf). Accessibility verified September 28, 2013.
33. Pan-Asian SNP Consortium. Available at: <http://www4a.biotec.or.th/PASNP>. Accessibility verified September 29, 2013.
34. Normile D. Consortium hopes to map human history in Asia. *Science.* 2004 Dec 3;306(5702):1667.
35. David C. Genomics takes hold in Asia. *Nature.* 2008 Nov 6;(456);12.
36. Teo Y-Y, Sim X, Ong RTH, et al. Singapore Genome Variation Project: A haplotype map of three South-East Asian populations. *Genome Res.* 2009 Aug 21;19: 2154-2162.
37. Andiappan AK, Anantharaman R, Nilkanth PP, et al. Evaluating the transferability of Hapmap SNPs to a Singapore Chinese population. *BMC Genetics.* 2010 May 07; 11:36.
38. Wellcome Trust Sanger Institute. Available at:<http://www.uk10k.org/>. 2013. Accessibility verified September 28, 2013.

39. Sachidanandam R, Weissman D, Schmidt SC. et al. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*. 2001 Feb 15;409(6822):928-33.
40. Wellcome Trust Case Control Consortium, Craddock N, Hurles ME, et al. Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature*. 2010 Apr 1;464(7289):713-20
41. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nature*. 2009 Oct 8;461(7265):747-53.
42. Cohen JC, Boerwinkle E, Mosley TH Jr, et al. Sequence variations in PCSK9, low LDL, and protection against coronary heart disease *N Engl J Med*. 2006 Mar 23;354(12):1264-72.
43. 1000 Genomes Project Consortium, Abecasis GR, Altshuler D, et al. A map of human genome variation from population-scale sequencing. *Nature*. 2010 Oct 28;467(7319):1061-73.
44. 1000 Genomes Project Consortium, Abecasis GR, Auton A, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012 Nov 1;491(7422):56-65.
45. 1000 Genomes A Deep Catalog of Human Genetic Variation. Available at: <http://www.1000genomes.org/about>. Accessibility verified September 29, 2013.
46. The National Center for Biotechnology Information. Available at: <http://www.ncbi.nlm.nih.gov/SNP/> . Accessibility verified September 30, 2013.
47. Mizrahi I. The NCBI Handbook. GenBank: The Nucleotide Sequence Database. Available at: <http://www.ncbi.nlm.nih.gov/books/NBK21105/>. Accessibility verified September 30, 2013.
48. The National Center for Biotechnology Information. Available at: [http://www.ncbi.nlm.nih.gov/projects/SNP/snp\\_summary.cgi?view+summary=vi](http://www.ncbi.nlm.nih.gov/projects/SNP/snp_summary.cgi?view+summary=vi). Accessibility verified September 30, 2013.
49. The National Center for Biotechnology Information. Available at: <http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Info&id=9606>. Accessibility verified September 30, 2013.
50. International HapMap Project. International HapMap Project Organization. Available at: <http://snp.cshl.org>. Accessibility verified September 30, 2013.
51. Mah JT, Chia KS. A Gentle Introduction TO SNP Analysis. *J Bioinform Comput Biol*. 2007 Oct;5(5):1123-38.
52. Hirakawa M, Tanaka T, Hashimoto Y, et al. JSNP: a database of common gene variations in the Japanese population,. *Nucleic Acids Res*. 2002 Jan 1;30(1):158-62.
53. Japanese Single Nucleotide Polymorphisms. Available at: <http://snp.ims.u-tokyo.ac.jp/mission.html>. Accessibility verified September 30, 2013.
54. Ngamphiw C, Assawamakin A, Xu S, et al. PanSNPdb: The Pan-Asian SNP Genotyping Database. *PLoS One*. 2011;6(6):e21451.
55. GWAS Central. Available at: <http://www.gwascentral.org/info/about/history/>. Accessibility verified September 30, 2013.
56. The National Center for Biotechnology Information. Available at: <http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/about.html>. Accessibility verified September 30, 2013.
57. The National Center for Biotechnology Information. Available

- at:<http://www.ncbi.nlm.nih.gov/dbvar/>. Accessibility verified September 30, 2013.
58. UCSC Genome Bioinformatics. Available at:<http://genome.ucsc.edu>. Accessibility verified September 30, 2013
  59. CHIP Bioinformatics Tool. Available at: <http://snpper.chip.org/bio/snpper-faq>. Accessibility verified July 18, 2009.
  60. Project Aligned Related Sequences and Evaluate SNPs. Available at: [http://www.proweb.org/parsesnp/parsesnp\\_help.html](http://www.proweb.org/parsesnp/parsesnp_help.html). Accessibility verified September 30, 2013
  61. Functional Element SNPs Database (FESD). Available at: <http://www.oxfordjournals.org/nar/database/summary/669>. Accessibility verified September 30, 2013
  62. Becker KG, Barnes KG, Bright TJ, et al. The Genetic Association Database. *Nature Genetics*. 2004; 36:431-432.
  63. Lee P, Shulman, Cystic Fibrosis Screening. *The Journal of Midwifery & Women's Health*; 2005 Jun. 50(3); 205-210.
  64. Barrett JC, Cardon LR. Evaluating coverage of genome-wide association studies. *Nat Genet*. 2006 Jun;38(6):659-62.
  65. Illumina. Available at: <http://investor.illumina.com/phoenix.zhtml?c=121127&p=irol-newsArticle&ID=938987&highlight>. Accessibility verified September 30, 2013
  66. R Tutorial. Available at: <http://www.r-tutor.com/gpu-computing/clustering/hierarchical-cluster-analysis>. Accessibility verified September 30, 2013
  67. Fox J, Andersen R. Using the R Statistical Computing Environment to Teach Social Statistics Courses. Available at: <http://www.unt.edu/rss/Teaching-with-R.pdf>. Accessibility verified September 30, 2013.
  68. Available CRAN Packages By Date of Publication. Available at: [http://cran.r-project.org/web/packages/available\\_packages\\_by\\_date.html](http://cran.r-project.org/web/packages/available_packages_by_date.html). Accessibility verified September 30, 2013.
  69. Dissimilarity Matrix Calculation. Available at: <http://stat.ethz.ch/R-manual/R-devel/library/cluster/html/daisy.html>. Accessibility verified September 30, 2013.
  70. Broad Institute. Available at: <http://www.broadinstitute.org/scientific-community/science/programs/medical-and-population-genetics/haploview/haploview.Haploview>. Accessibility verified August 18, 2013.
  71. Menozzi P, Piazza A, Cavalli-Sforza L. Synthetic maps of human gene frequencies in Europeans. *Science*. 1978 Sep 1;201(4358):786-92.
  72. Paschou P, Ziv E, Burchard EG, et al. PCA-Correlated SNPs for structure Identification in Worldwide Human Populations. *Plos Genetics*. 2007 Sep;3(9):1672-1686.
  73. Patterson N, Price AL, Reich D. Population Structure and Eigenanalysis. *Plos Genetics*, 2006 Dec;2(12): 2074-2093.
  74. Demiralay M, Çamurcu AY. Cure, Agnes ve K-Means algoritmalarındaki kümeleme yeteneklerinin karşılaştırılması. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, Güz 2005/2;4(8):1-18.
  75. Bustamante CD, Burchard EG, De la Vega FM. Genomics for the world. *Nature*. 2011 Jul 13;475(7355):163-5.
  76. Bauchet M, McEvoy B, Pearson LN, et al. Measuring European Population Stratification

- with Microarray Genotype Data. May 2007. Am J Hum Genet. 2007 May;80(5):948-56.
77. Affymetrix. Available at:  
<http://www.affymetrix.com/analysis/netaffx/xmlquery.affx?netaffx=mapping>.  
Accessibility verified September 30, 2013.
  78. R Project. Available at: <http://www.r-project.org>. Accessibility verified September 30, 2013
  79. International HapMap Project. International HapMap Project Organization Available at:  
[http://hapmap.ncbi.nlm.nih.gov/downloads/frequencies/2009-02\\_phaseII+III/](http://hapmap.ncbi.nlm.nih.gov/downloads/frequencies/2009-02_phaseII+III/).  
Accessibility verified September 30, 2013

## EKLER

### EK1. Genomik diziler için Genom Tarayıcıları

VERİ TABANLARI	YAPISAL	NCBI
Databases (Örnek Veri tabanları)	<b>Database of Genomic Structural Variation (dbVar)</b> <a href="http://www.ncbi.nlm.nih.gov/dbvar">http://www.ncbi.nlm.nih.gov/dbvar</a> <i>Bakınız-Genom Projesinden Doğan Yeni Projeler</i>	
	<b>Database of Genotypes and Phenotypes (dbGaP)</b> <a href="http://www.ncbi.nlm.nih.gov/gap">http://www.ncbi.nlm.nih.gov/gap</a> <i>Bakınız-Genom Projesinden Doğan Yeni Projeler</i>	
	<b>Database of Short Genetic Variations (dbSNP)</b> <a href="http://www.ncbi.nlm.nih.gov/snp">http://www.ncbi.nlm.nih.gov/snp</a> <i>Bakınız-Genom Projesinden Doğan Yeni Projeler</i>	
	<b>Genome</b> <a href="http://www.ncbi.nlm.nih.gov/genome">http://www.ncbi.nlm.nih.gov/genome</a> 1000'in üzerinde organizmanın tüm genom verisini ve dizisini içermektedir. Hem dizilenmesi tamamlanmış hem de dizilenmesi devam eden organizmaların genomlarını temsil etmektedir. Yaşamın üç domainininin yanısıra (bakteri, arke ve ökaryotlar) birçok virüs, faj, viroidler, plasmidler ve organelleri temsil etmektedir.	
	<b>GenBank</b> <a href="http://www.ncbi.nlm.nih.gov/genbank/">http://www.ncbi.nlm.nih.gov/genbank/</a> Sağlık Bakanlığı (National Institute of Health (NIH)) genetik dizi veri tabanı kamuya açık olan DNA dizilerinin anote edilmiş bir koleksiyonudur. Genbank Japon Veribankası (DDBJ), Avrupa Moleküler Biyoloji Laboratuvarı (EMBL) ve NCBI Genbank'dan oluşan Uluslararası Nükleotid Dizi veri tabanı İşbirliğinin bir parçasıdır. Bu üç organizasyon günlük olarak veri değişimi yapmaktadır. Genbank çoğunluğuna Nükleotid veri tabanı aracılığı ile ulaşılabilen birçok bölüm içermektedir. Expressed Sequence Tags (EST), Genome Survey Sequences (GSS) bölümlerine Nükleotid EST ve Nükleotid GSS veritabanları aracılığı ile erişilebilmektedir.	
	<b>Online Mendelian Inheritance in Man (OMIM)</b> <a href="http://www.ncbi.nlm.nih.gov/omim">http://www.ncbi.nlm.nih.gov/omim</a> İnsan gen ve genetik bozuklukları veri tabanı. NCBI; içerik desteğinin yanısıra, arama motorları ve farklı veri tabanları ile de entegrasyon desteği de sağlamaktadır. Fakat OMIM'in artık omim.org adında yeni bir adresi bulunmaktadır. Tüm kayıtları görebilmek için kullanıcı bu adrese yönlendirilmektedir.	
Downloads (Örnek İndirme Dosyaları)	<b>BLAST (Veri tabanından bağımsız kullanılabilir)</b> <a href="http://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&amp;PAGE_TYPE=BlastDocs&amp;DOC_TYPE=Download">http://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&amp;PAGE_TYPE=BlastDocs&amp;DOC_TYPE=Download</a> Solaris, LINUX, Windows, ve MacOSX sistemlerde Lokal kullanım için BLAST çalıştırılabilirler (executable). Nükleotid, protein BLAST ve transle aramaların (translated searches) indirilmesi db altklasörü altında mevcuttur.	
	<b>FTP: BLAST Databases</b> <a href="ftp://ftp.ncbi.nlm.nih.gov/blast/db/">ftp://ftp.ncbi.nlm.nih.gov/blast/db/</a> Stand-alone BLAST programları ile kullanılması için dizi veritabanları. Bu klasördeki önceden formatlanmış veritabanlarıdır ve BLAST ile kullanılmak için hazırlanmıştır.	
	<b>FTP: SNP</b> <a href="ftp://ftp.ncbi.nlm.nih.gov/snp/">ftp://ftp.ncbi.nlm.nih.gov/snp/</a> İndirilebilir SNP verisi	

			<p><b>FTP: Site</b> <a href="http://www.ncbi.nlm.nih.gov/Ftp/">http://www.ncbi.nlm.nih.gov/Ftp/</a> NCBI Veritabanları, araçları ve yardımcıları için ftp indirme sitesi</p>
		<b>Submissions (Örnek Gönderiler)</b>	TNP Gönderi Araçları, dbGaP Veri Gönderi Politikaları
		<b>Tools (Örnek Araçlar)</b>	<p><b>TNP Veri tabanına Özgü Arama Araçları</b> <a href="http://www.ncbi.nlm.nih.gov/projects/SNP/">http://www.ncbi.nlm.nih.gov/projects/SNP/</a> SNP veritabanlarını araştırmak için çeşitli araçlar mevcuttur. BLAST yardımıyla genotip, metod, popülasyon, gönderici (submitter), belirteç ve dizi benzerliği aramasına olanak sağlamaktadır.</p>
			<p><b>1000 Genome Browseri</b> <a href="http://www.ncbi.nlm.nih.gov/variation/tools/1000genomes/">http://www.ncbi.nlm.nih.gov/variation/tools/1000genomes/</a> 1000 Genom Projesinde ortaya çıkan varyant değerlerini (variant calls), genotip değerlerini (genotype calls) ve hizalanmış dizi okumaları gibi kanıtları interaktif grafiksel görüntüleyici yardımıyla araştırılmasını sağlamaktadır</p>
			<p><b>Basic Local Alignment Search Tool (BLAST)</b> <a href="http://blast.ncbi.nlm.nih.gov/">http://blast.ncbi.nlm.nih.gov/</a> Biyolojik diziler arasındaki lokal benzerliklerin olduğu bölgeleri bulur. Nükleotit veya protein dizilerini dizi veritabanları ile karşılaştırır ve eşleşmenin istatistiksel olarak anlamlılığını hesaplar. Bunların dışında BLAST gen ailelerini tanımlamaya yardımcı olmasının yanında, diziler arasında fonksiyonel ve evrimsel ilişkileri anlamlandırmak için kullanılmaktadır.</p>
			<p><b>Map Viewer</b> <a href="http://www.ncbi.nlm.nih.gov/mapview/">http://www.ncbi.nlm.nih.gov/mapview/</a> Bir kısım (subset of organisms) birleştirilen dizilerine ve haritalarına istenilen şekillerde göz atmayı sağlar. Bir organizmanın tüm genomunu, haritasını ve yakınlaştırma ve uzaklaştırma özelliği ile ayrıntılı inceleme sağlamaktadır.</p>
			<p><b>Genome BLAST</b> <a href="http://www.ncbi.nlm.nih.gov/blast/Blast.cgi">http://www.ncbi.nlm.nih.gov/blast/Blast.cgi</a> Genomik dizi veritabanlarından nükleotit ve protein dizilerinin karşılaştırılmasını sağlar ve BLAST algoritmasını kullanarak istatistiksel anlamlılığını hesaplamaktadır.</p>
			<b>UCSC</b>
		<b>EBI (ENSEMBL)</b>	İçerdiği organizma sayısına göre ve içerdiği bilgiye göre geniş kapsamlı bir veri tabanıdır.

	FONKSİYONEL	ENCODE	İnsan genomunda kodlanan fonksiyonel elemanların tamamının (transkripsiyon, transkripsiyon faktörü bağlanma bölgeleri, kromatin yapısı, histon modifikasyonları) ortaya konmasını amaçlamıştır.
		EPIGENOM	İnsan genomunda değişken metilasyon pozisyonlarını (Methylation Variable Positions (MVPs)) kataloglamayı amaçlamıştır.

## EK2. Genom Varyasyon Veritabanları

GENOM VARYASYON VERİTABANLARI	NCBI	<b>dbVAR</b> - Büyük insersiyonlar, delesyonlar, translokasyonlar, ve insersiyonlar da dahil olmak üzere büyük ölçekli genomik varyasyonların kataloglandığı veri tabanıdır. dbVAR aynı zamanda tanımlanan varyantların fenotip bilgileri ile ilişkilerini tutar.
		<b>dbGAP</b> – Genotip ve Fenotip ve fenotip ilişkilerinin interaksyonunu araştıran çalışmaların sonuçlarını tutar. Genom boyunca çalışmalar, tıbbi amaçlı dizileme moleküler teşhis için yapılan analizlerin yanısıra, genotip ve klinik olmayan özellikler arasındaki ilişkileri tutar.
		<b>dbSNP</b> - TNP, mikrosatelitler, küçük ölçekli insersiyon ve delesyonları içerir. Popülasyona özgü frekans ve genotip verileri, deneysel şartlar, moleküler içerik ve hem nötral hem de klinik mutasyonlar için haritalama bilgilerini içermektedir.
	HapMap	İnsan genomunda Minör Allel Frekansı %5'in üzerinde olan (MAF > 0,05) <sup>25</sup> varyasyonların tamamının ortaya konulması kromozomlar üzerinde birbirleriyle ilişkili lokuslardaki allel kombinasyonlarının (Haplotip) ortaya konularak Haplotip haritalarının ortaya çıkartılması amaçlanmıştır
1000 Genom	Genotip ve fenotip ilişkilerinin araştırıldığı proje ile farklı popülasyonlar özgü düşük frekans ve nadir varyantların araştırılması amaçlanmıştır (MAF 0,5%-5% ve MAF < 0,5%)	

<sup>25</sup> Minör Allel Frekansı (MAF) popülasyonda tanımlanan TNP'nin az yaygın allelin frekansına verilen addır. HapMap Projesinde MAF 0,05 ve üzeri seçilmiştir. 1000 Genom Projesinde MAF 0,05 ve altı seçilmiştir.

### EK3. Türkiye Bölge Kodları

<b>1B Akdeniz Bölgesi</b>																				
01	Adana	0100 Büyükşehir	0101 Seyhan	0102 Aladağ	0104 Ceyhan	0106 Feke	0107 İmamoğlu	0109 Karaisalı	0110 Karataş	0111 Kozan	0113 Pozantı	0114 Saimbeyli	0115 Tufanbeyli	0116 Yumurtalık	0117 Yüreğir					
46	Kahramanmaraş	4601 Merkez	4602 Afişin	4603 Andırın	4604 Çağlıyancerit	4605 Elbistan	4606 Göksun	4607 Pazarcık	4608 Türkoğlu	4609 Ekinözü	4610 Nurhak									
07	Antalya	0701 Merkez	0702 Akseki	0703 Alanya	0704 Elmalı	0705 Finike	0706 Gazipaşa	0707 Gündoğmuş	0708 Kale	0709 Kaş	0710 Korkuteli	0711 Kumluca	0712 Manavgat	0713 Serik	0714 İbradı	0715 Kemer				
15	Burdur	1501 Merkez	1502 Ağlasun	1503 Bucak	1504 Göhlisar	1505 Karamanlı	1506 Kemer	1507 Tefenni	1508 Yeşilova	1509 Altınyayla	1510 Cavdır	1511 Çeltikci								
31	Hatay	3100 Merkez	3101 Antakya	3102 Altınözü	3103 Dörtöyl	3104 Erzin	3105 Hassa	3106 İskenderun	3107 Kırıkhan	3108 Reyhanlı	3109 Samandağ	3110 Yayladağı	3111 Belen	3112 Kumlu						
32	Isparta	3201 Merkez	3202 Aksu	3203 Atabey	3204 Eğirdir	3205 Gelendost	3206 Keçiborlu	3207 Senirkent	3208 Sütçüler	3209 Şarkikaraağaç	3210 Uluborlu	3211 Yalvaç	3212 Gönen	3213 Yenisarbademli						
33	Mersin	3302 Anamur	3303 Aydıncık	3304 Bozyazı	3305 Erdemli	3306 Gülnar	3307 Mut	3308 Silifke	3309 Tarsus	3310 Çamlyayla										
80	Osmaniye	8001 Merkez	8002 Bahçe	8003 Düzici	8004 Kadirli	8005 Hasanbeyli	8006 Sumbas	8007 Toprakkale												
<b>2B Doğu Anadolu Bölgesi</b>																				
04	Ağrı	04 Ağrı	0401 Merkez	0402 Diyadin	0403 Doğubeyazıt	0404 Eleşkirt	0405 Hamur	0406 Patnos	0407 Taşlıçay	0408 Tutak										
75	Ardahan	7501 Merkez	7502 Çıldır	7503 Damal	7504 Göle	7505 Hanak	7506 Posof													
12	Bingöl	1201 Merkez	1202 Adaklı	1203 Genç	1204 Karlıova	1205 Kiğı	1206 Solhan	1207 Yayladere	1208 Yedisu											
13	Bitlis	1301 Merkez	1302 Adilcevaz	1303 Ahlat	1304 Güroymak	1305 Hizan	1306 Mutki	1307 Tatvan												
23	Elazığ	2301 Merkez	2302 Ağın	2303 Arıcak	2304 Baskil	2305 Karakoçan	2306 Keban	2307 Kovancılar	2308 Maden	2309 Palu	2310 Sivrice	2311 Alacakaya								
24	Erzincan	2401 Merkez	2402 Cayırli	2403 Iliç	2404 Kemah	2405 Kemaliye	2406 Refahiye	2407 Tercan	2408 Üzümlü	2409 Otlukbeli										
25	Erzurum	2501 Merkez	2502 Aşkale	2503 Çat	2504 Hınıs	2505 Horasan	2506 İspir	2507 Karaçoban	2508 Karayazı	2509 Narman	2510 Oltu	2511 Olur	2512 Pasinler	2513 Şenkaya	2514 Tekman	2515 Tortum	2516 Üzundere	2517 Pazaryolu	2518 Köprüköy	2519 İliça
30	Hakkâri	3001 Merkez	3003 Çukurca	3004 Şemdinli	3006 Yüksekova															
76	Iğdır	7601 Merkez	7602 Aralık	7603 Karakoyunlu	7604 Tuzluca															
36	Kars	3601 Merkez	3602 Akyaka	3605 Arpaçay	3607 Digor	3611 Kağızman	3613 Sarıkamış	3614 Selim	3615 Susuz											
44	Malatya	4401 Merkez	4402 Akçadağ	4403 Arapgir	4404 Arguvan	4405 Battalgazi	4406 Darende	4407 Doğanşehir	4408 Hekimhan	4409 Pötürge	4410 Yeşilyurt	4411 Doğanyol	4412 Kale	4413 Kuluçak	4414 Yazihan					
49	Muş	4901 Merkez	4902 Bulanık	4903 Hasköy	4904 Malazgirt	4905 Varto	4906 Korkut													
62	Tunceli	6201 Merkez	6202 Çemişkezek	6203 Hozat	6204 Mazgirt	6205 Nazimiye	6206 Ovacık	6207 Pertek	6208 Pülümür											
65	Van	6501 Merkez	6502 Bahcesaray	6503 Başkale	6504 Çaldıran	6505 Çatak	6506 Erciş	6507 Gevaş	6508 Gürpınar	6509 Muradiye	6510 Özalp	6511 Edremit	6512 Saray							



3B Ege Bölgesi																										
03	Afyonkarahisar	0301 Merkez	0302 Başmakçı	0303 Bayat	0304 Bolvadin	0305 Çay	0306 Dazkırı	0307 Dinar	0308 Emirdağ	0309 İhsaniye	0310 İncehisar	0311 Sandıklı	0312 Sincanlı	0313 Sultandağı	0314 Şuhut	0315 Çobanlar	0316 Evciler	0317 Hocalar	0318 Kızılören							
09	Aydın	0901 Merkez	0902 Bozdoğan	0903 Buharkent	0904 Çine	0905 Germencik	0906 Incirliova	0907 Karacasu	0908 Koçarlı	0909 Kuşadası	0910 Kuyucak	0911 Nazilli	0912 Söke	0913 Sultanhisar	0914 Yenipazar	0915 Karpuzlu	0916 Koşk	0917 Didim								
20	Denizli	2001 Merkez	2002 Acıpayım	2003 Babadağ	2004 Bekilli	2005 Buldan	2006 Çal	2007 Çameli	2008 Çardak	2009 Çivril	2010 Güney	2011 Honaz	2012 Kale	2013 Sarayköy	2014 Serinhisar	2015 Tavas	2016 Akköy	2017 Baklan	2018 Beyağaç	2019 Bozkurt						
35	İzmir	3501 İzmir	3502 Büyükşehir	3503 Konak	3504 Aliağa	3505 Bayındır	3506 Bergama	3507 Beydağ	3508 Bornova	3509 Buca	3510 Çeşme	3511 Dikili	3512 Foça	3513 Karaburun	3514 Karşıyaka	3515 Kemalpaşa	3516 Kınık	3517 Kiraz	3518 Menderes	3519 Menemen	3520 Ödemiş	3521 Seferihisar	3522 Selçuk	3523 Tire	3524 Torbalı	3525 Urla
		3524 Balçova	3525 Çiğli	3526 Gazimeci	3527 Narlıdere	3528 Güzelyazır																				
43	Kütahya	4301 Merkez	4302 Altıntaş	4303 Aslanapa	4304 Domaniç	4305 Dumlupınar	4306 Emet	4307 Gediz	4308 Hisarcık	4309 Simav	4310 Şaphane	4311 Tavşanlı	4312 Çavdarhisar	4313 Pazarlar												
45	Manisa	4501 Merkez	4502 Ahmetli	4503 Akhisar	4504 Alaşehir	4505 Demirci	4506 Gölmarımara	4507 Gördes	4508 Kırkağaç	4509 Kula	4510 Salihli	4511 Sarıgöl	4512 Saruhanlı	4513 Selendi	4514 Soma	4515 Turgutlu	4516 Köprübaşı									
48	Muğla	4801 Merkez	4802 Bodrum	4803 Dalaman	4804 Datça	4805 Fethiye	4806 Köyceğiz	4807 Marmaris	4808 Milas	4809 Ortaca	4810 Ula	4811 Yatağan	4812 Kavaklıdere													
64	Uşak	6401 Merkez	6402 Banaz	6403 Eşme	6404 Karahallı	6405 Sivasslı	6406 Ulubey																			

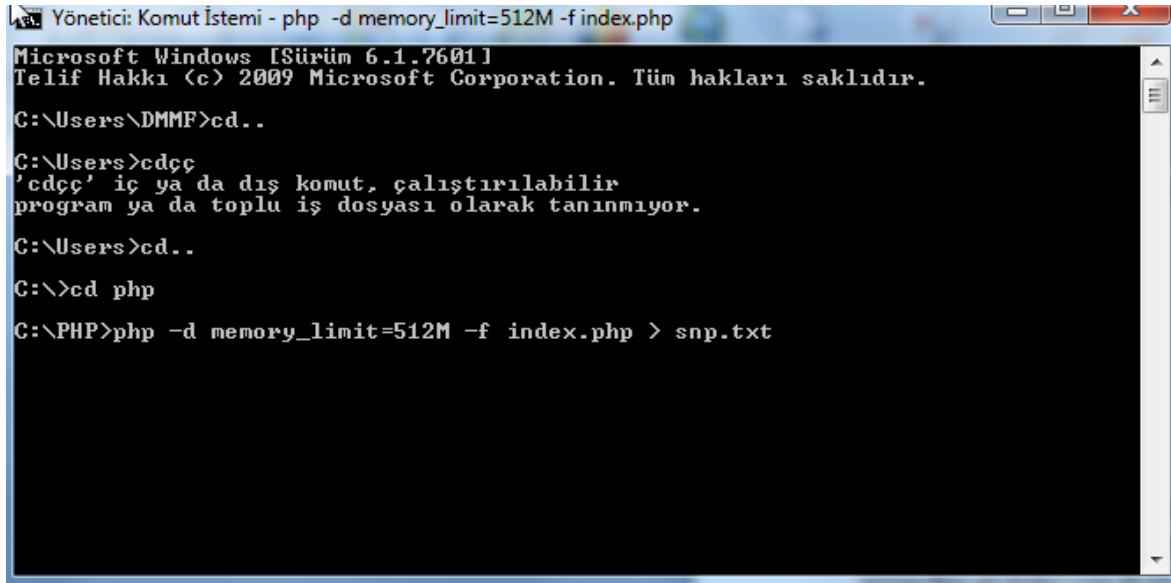
4B Güneydoğu Anadolu Bölgesi																	
02	Adıyaman	0201 Merkez	0202 Besni	0203 Çelikhan	0204 Gergeri	0205 Gölbaşı	0206 Kahta	0207 Samsat	0208 Sincik	0209 Tut							
72	Batman	7201 Merkez	7202 Beşiri	7203 Gercüş	7204 Hasankeyf	7205 Kozluk	7206 Sason										
21	Diyarbakır	2101 Merkez	2102 Bismil	2103 Çermik	2104 Çınar	2105 Çüngüş	2106 Diele	2107 Eğil	2108 Ergani	2109 Hani	2110 Hazro	2111 Kulp	2112 Lice	2113 Silvan	2114 Kocaköy		
27	Gaziantep	2700 Büyükşehir	2701 Şehitkamil	2702 Araban	2703 İslahiye	2705 Nizip	2706 Oğuzeli	2707 Yavuzeli	2708 Şahinbey	2709 Karkamış	2710 Nurdagi						
79	Kilis	7901 Merkez	7902 Elbeyli	7903 Musabeyli	7904 Polateli												
47	Mardin	4701 Merkez	4702 Dargeçit	4703 Derik	4704 Kızıltepe	4707 Mazıdağı	4708 Midyat	4709 Nusaybin	4710 Ömerli	4711 Savur	4712 Yeşilli						
56	Sıirt	5601 Merkez	5602 Baykan	5603 Eruh	5604 Kurtalan	5607 Pervari	5608 Sirvan	5611 Aydınlı									
63	Şanlıurfa	6301 Merkez	6302 Akcakale	6303 Birecik	6304 Bozova	6305 Ceylanpınarı	6306 Halfeti	6307 Harran	6308 Hilvan	6309 Siverek	6310 Suruç	6311 Viranşehir					
73	Şırnak	7301 Merkez	7302 Beytüşşebap	7303 Cizre	7304 Güçlükonak	7305 İdil	7306 Silopi	7307 Uludere									

5B İç Anadolu Bölgesi																										
68	Aksaray	6801 Merkez	6802 Aşağıören	6803 Güzel yurt	6804 Ortaköy	6805 Sarıyağlı	6806 Eskil	6807 Gülağaç																		
06	Ankara	0600 Büyükşehir	0601 Altındağ	0602 Ayaş	0603 Bala	0604 Beypazarı	0605 Çamlıdere	0606 Çankaya	0607 Çubuk	0609 Elmadağ	0610 Gölbaşı	0611 Güdül	0612 Haymana	0613 Kalecik	0614 Kazan	0615 Keçiören	0618 Kızılcahamam	0619 Mamak	0620 Nallıhan	0621 Polatlı	0622 Sincan	0624 Şereflikoçhisar	0625 Yenimahalle	0626 Akyurt	0627 Etimesgut	0628 Evren
18	Çankırı	1801 Merkez	1802 Atkaracalar	1803 Çerkeş	1804 Eldivan	1806 Ilgaz	1807 Kızılırmak	1808 Kursunlu	1809 Orta	1811 Şabanözü	1812 Yapraklı	1813 Bayramören	1814 Korgun													
26	Eskişehir	2601 Merkez	2602 Alpu	2603 Beylikova	2604 Çifteler	2605 İnönü	2606 Mahmudiye	2607 Mihaliccik	2608 Sarca	2609 Seyitgazi	2610 Sivrihisar	2611 Günyüzü	2612 Han	2613 Mihalgazi												
70	Karaman	7001 Merkez	7002 Ayrancı	7003 Ermenk	7004 KazımKarabekir	7005 Başyayla	7006 Sarveliler																			
38	Kayseri	3800 Büyükşehir	3802 Bünyan	3803 Develi	3804 Felahiye	3805 İncesu	3806 Pınarbaşı	3807 Sarıoğlan	3808 Sarız	3809 Tomarza	3810 Yahyalı	3811 Yeşilhisar	3812 Akkışla	3813 Talas	3814 Melikgazi	3815 Kocasinan	3816 Hacılar	3817 Özvatan								
71	Kırkkale	7101 Merkez	7102 Delice	7103 Keskin	7104 Sulakyurt	7105 Balışeyhan	7106 Çelebi	7107 Karakeçili	7108 Yahşihan	7109 Bahşili																
40	Kırşehir	4001 Merkez	4002 Akpınar	4003 Çiçekdağı	4004 Kaman	4005 Mucur	4006 Akçakent	4007 Boztepe																		
42	Konya	4200 Konya	4201 Selçuklu	4202 Akören	4203 Akşehir	4204 Altınekin	4206 Beyşehir	4207 Bozkır	4208 Cihanbeyli	4209 Çumra	4210 Derebucak	4211 Doğanhisar	4212 Ereğli	4214 Hadım	4215 Hüyük	4216 Ilgın	4217 Kadınhanı	4219 Karapınar	4220 Kulu	4221 Sarayönü	4222 Seydişehir	4223 Taşkent	4224 Yunak	4225 Karatay	4226 Meram	
		4227 Ahırlı	4228 Çeltik	4229 Derbent	4230 Emirgazi	4231 Güneysınır	4232 Halkapınar	4233 Tuzlukçu	4234 Yalıhüyük																	
50	Nevşehir	5001 Merkez	5002 Acıgöl	5003 Avanos	5004 Derinkuyu	5005 Gülşehir	5006 Hacıbektaş	5007 Kozaklı	5008 Ürgüp																	
51	Niğde	5101 Merkez	5103 Bor	5104 Çamardı	5106 Ulukışla	5107 Altunhisar	5108 Çiftlik																			
58	Sivas	5801 Merkez	5802 Divriği	5803 Gemerek	5804 Gürün	5805 Hafik	5806 İmranlı	5807 Kangal	5808 Koyulhisar	5809 Suşehri	5810 Şarkışla	5811 Yıldızeli	5812 Zara	5813 Akıncılar	5814 Gölova	5815 Ulaş	5816 Doğanşar	5817 Altınayla								
66	Yozgat	6601 Merkez	6602 Akdağmadeni	6603 Boğazlıyan	6604 Çayıralan	6605 Çekerek	6606 Sarıkaya	6607 Sorgun	6608 Şefaati	6609 Yerköy	6610 Aydıncık	6611 Çandır	6612 Kadışehri	6613 Saraykent	6614 Yenifakılı											

6B Karadeniz Bölgesi																					
05	Amasya	0501 Merkez	0502 Göynücek	0503 Gümüşhacıköy	0504 Merzifon	0505 Suluova	0506 Tasova	0507 Hamamözü													
08	Artvin	0801 Merkez	0802 Ardandere	0803 Arhavi	0804 Borçka	0805 Hopa	0806 Murgul	0807 Şavşat	0808 Yusufeli												
74	Bartın	7401 Merkez	7402 Amasra	7403 Kurucasıle	7404 Ulus																
69	Bayburt	6901 Merkez	6902 Aydıntepe	6903 Demirözü																	
14	Bolu	1401 Merkez	1405 Gerede	1407 Göynük	1408 Kıbrısçık	1409 Mengen	1410 Mudurnu	1411 Seben	1414 Dörtdivan	1415 Yeniçağa											
19	Çorum	1901 Merkez	1902 Alaca	1903 Bayat	1904 Boğazkale	1905 İskilip	1906 Kargı	1907 Mecitözü	1908 Ortaköy	1909 Osmancık	1910 Sungurlu	1911 Uğurludağ	1912 Dodurga	1913 Laçın	1914 Oğuzlar						
81	Düzce	8101 Merkez	8102 Akçakoca	8103 Cumayeri	8104 Çilimli	8105 Gölyaka	8106 Gümüşova	8107 Kaynaşlı	8108 Yiğilca												
28	Giresun	2801 Merkez	2802 Alucra	2803 Bulancak	2804 Dereli	2805 Espiye	2806 Eynesil	2807 Görele	2808 Keşap	2809 Piraziz	2810 Şebinkarahisar	2811 Tirebolu	2812 Yağlıdere	2813 Çanakçı	2814 Güce	2815 Çamoluk	2816 Doğankest				
29	Gümüşhane	2901 Merkez	2905 Kelkit	2906 Köse	2907 Şiran	2908 Torul	2909 Kurtun														
78	Karabük	7801 Merkez	7802 Eflani	7803 Eskipazar	7804 Ovacık	7805 Safranbolu	7806 Yenice														
37	Kastamonu	3701 Merkez	3702 Abana	3703 Araç	3704 Azdavay	3705 Bozkurt	3706 Cide	3707 Catalzeytin	3708 Daday	3709 Devrekani	3710 İhsangazi	3711 İnebolu	3712 Küre	3713 Taşköprü	3714 Tosya	3715 Pınarbaşı	3716 Şenpazar	3717 Ağlı	3718 Doğanıyurt	3719 Hanönü	3720 Seydiler
52	Ordu	5201 Merkez	5202 Akkuş	5203 Aybastı	5204 Fatsa	5205 Gülyalı	5206 Gürgentepe	5207 Gökçöy	5208 Korgan	5209 Kumru	5210 Mesudiye	5211 Perşembe	5212 Ulubey	5213 Ünye	5214 Çamas	5215 Çatalpınar	5216 Çaybaşı	5217 İkizce	5218 Kabadüz	5219 Kabataş	
53	Rize	5301 Merkez	5302 Ardeşen	5303 Çamlıhemşin	5304 Çayeli	5305 Fındıklı	5306 Güneşsu	5307 İkizdere	5308 Kalkandere	5309 Pazar	5310 Derepazarı	5311 Hemşin	5312 İyidere								
55	Samsun	5501 Merkez	5502 Alaçam	5503 Asarcık	5504 Bafra	5505 Çarşamba	5506 Havza	5507 Kavak	5508 Ladik	5509 Ondokuzmayıs	5510 Salıpazarı	5511 Tekkeköy	5512 Terme	5513 Vezirköprü	5514 Ayvacık	5515 Yakakent					
57	Sinop	5701 Merkez	5702 Ayancık	5703 Boyabat	5704 Durağan	5705 Erfelek	5706 Gerze	5707 Türkeli	5708 Dikmen	5709 Saraydüzü											
60	Tokat	6001 Merkez	6002 Almus	6003 Artova	6004 Erbaa	6005 Niksar	6006 Pazar	6007 Reşadiye	6008 Turhal	6009 Yeşilyurt	6010 Zile	6011 Sulusaray	6012 Başçiftlik								
61	Trabzon	6101 Merkez	6102 Akcaabat	6103 Araklı	6104 Arşin	6105 Beşikdüzü	6106 Çaykara	6107 Maçka	6108 Of	6109 Şalpazarı	6110 Sürmene	6111 Tonya	6112 Vakfikebir	6113 Yomra	6114 Çarşıbaşı	6115 Dernekpazarı	6116 Düzköy	6117 Hayrat	6118 Köprübaşı		
67	Zonguldak	6701 Merkez	6702 Alaplı	6705 Çaycuma	6706 Devrek	6708 Ereğli	6714 Gökçebe														

7B Marmara Bölgesi																										
10	Bahkesir	1001 Merkez	1002 Ayvalık	1003 Balya	1004 Bandırma	1005 Bigadiç	1006 Burhaniye	1007 Dursunbey	1008 Edremit	1009 Erdek	1010 Gönen	1011 Havran	1012 İvrindi	1013 Kepsut	1014 Manyas	1015 Marmara	1016 Savaştepe	1017 Sındırgı	1018 Susurlu	1019 Gömeç						
11	Bilecik	1101 Merkez	1102 Bozüyük	1103 Gölpaazarı	1104 Osmaniye	1105 Pazaryeri	1106 Söğüt	1107 Yenipazar	1108 İnhisar																	
16	Bursa	1600 Büyükşehir	1601 Yıldırım	1602 Büyükorhan	1603 Gemlik	1604 Harmancık	1605 İnegöl	1606 İznik	1607 Karacabey	1608 Keles	1609 Mudanya	1610 M.Kemalpaşa	1611 Orhaneli	1612 Orhangazi	1613 Yenisehir	1614 Nilüfer	1615 Osmangazi	1616 Gürsu	1617 Kestel							
17	Çanakkale	1701 Merkez	1702 Ayvacık	1703 Bayramiç	1704 Biga	1705 Bozcaada	1706 Çan	1707 Eceabat	1708 Ezine	1709 Gelibolu	1710 Gökçeada	1711 Lapseki	1712 Yenice	18 Çankırı												
22	Edirne	2201 Merkez	2202 Enez	2203 Havsa	2204 İpsala	2205 Keşan	2206 Lalapaşa	2207 Meriç	2208 Uzunköprü	2209 Süloğlu																
34	İstanbul	3400 Büyük Şehir	3401 Adalar	3402 Bakırköy	3403 Beşiktaş	3404 Beykoz	3405 Beyoğlu	3406 Büyükçekmece	3407 Çatalca	3408 Eminönü	3409 Eyüp	3410 Fatih	3411 Gaziosmanpaşa	3412 Kadıköy	3413 Kağıthane	3414 Kartal	3415 Küçükçekmece	3416 Pendik	3417 Sarıyer	3418 Silivri	3419 Şile	3420 Şişli	3421 Ümraniye	3422 Üsküdar	3423 Zeytinburnu	3424 Bayrampaşa
39	Kırklareli	3901 Merkez	3902 Babaeski	3903 Demirköy	3904 Kofçaz	3905 Lüleburgaz	3906 Pehlivanköy	3907 Pınarhisar	3908 Vize																	
41	Kocaeli	4101 İzmit	4102 Gebze	4103 Gölçük	4104 Kandıra	4105 Karamürsel	4106 Körfez																			
54	Sakarya	5401 Adapazarı	5402 Akyazı	5403 Geyve	5404 Hendek	5405 Karasu	5406 Kaynarca	5407 Kocaeli	5408 Pamukova	5409 Sapanca	5410 Taraklı	5411 Karapürçek	5412 Ferizli	5413 Söğüt												
59	Tekirdağ	5901 Merkez	5902 Çerkezköy	5903 Çorlu	5904 Hayrabolu	5905 Malkara	5906 Marmara Ereğlisi	5907 Muratlı	5908 Saray	5909 Şarköy																
77	Yalova	7701 Merkez	7702 Altınova	7703 Armutlu	7704 Çınarcık	7705 Çiftlikköy	7706 Termal																			

**EK4:** Çipte her rs'e karşılık gelen en güncel allelleri (AllelA/B) NCBI web sitesinden çeker



```
Yönetici: Komut İstemi - php -d memory_limit=512M -f index.php
Microsoft Windows [Sürüm 6.1.7601]
Telif Hakkı (c) 2009 Microsoft Corporation. Tüm hakları saklıdır.

C:\Users\DMMF>cd..

C:\Users>cdçç
'cdçç' iç ya da dış komut, çalıştırılabilir
program ya da toplu iş dosyası olarak tanınmıyor.

C:\Users>cd..

C:\>cd php

C:\PHP>php -d memory_limit=512M -f index.php > snp.txt
```

```
<?php
ini_set("display_errors","1");
error_reporting(E_ALL);
$datafile = 'data.csv';
$fp = fopen($datafile, 'r');
$fc = fread($fp, filesize( $datafile ) );
fclose($fp);
$lines = explode("\r", $fc );
$index = 0;
foreach ( $lines as $line ){
    if( $index < 1 ){
        $index++;
        continue;
    }
    $columns = explode(",",$line);
    $curlURL = "http://www.ncbi.nlm.nih.gov/sites/entrez";
```

```

$postdata ="db=snp&cmd=search&term=".$columns[0];

echo $postdata;

$browser_id = "Mozilla/5.0 (Macintosh; U; Intel Mac OS X 10.6; en-US; rv:1.9.2.3)
Gecko/20100401 Firefox/3.6.3";

$ch = curl_init($curlURL);

        curl_setopt($ch, CURLOPT_SSL_VERIFYPEER      ,      false
        );

        curl_setopt($ch, CURLOPT_USERAGENT      ,      $browser_id
);

        curl_setopt($ch, CURLOPT_POST
        );

        curl_setopt($ch, CURLOPT_POSTFIELDS
$postdata      );

        curl_setopt($ch, CURLOPT_FOLLOWLOCATION , 1
        );

        curl_setopt($ch, CURLOPT_HEADER      ,      0
        );

        curl_setopt($ch, CURLOPT_RETURNTRANSFER , 1
        );

$sonuc = curl_exec($ch);

if( $sonuc === false ){

echo 'Curl error: ' . curl_error($ch);

}

curl_close($ch);

preg_match_all('#<span style="color:red">(.*?)</span>#msi', $sonuc, $matches);

if(isset($matches[1][0])){

        echo $columns[0]." - ".$matches[1][0]."\r\n";

}

$index++;

}

?>

```

## EK5. Matlab ile HapMap Popülasyon Frekanslarının Hesaplanması

```
list_num=8;

fid = fopen('hap_map_excel_list_chr1.txt');
C = textscan(fid, '%s');

SNP_number=125164;

N_AA(SNP_number,list_num)=0;
N_AB(SNP_number,list_num)=0;
N_BB(SNP_number,list_num)=0;

for n=1:list_num

if n<8
k=10;
else
k=3;
end
A = importdata(C{1,1}{n,1});

for j=1:SNP_number
```

```

AC = strcmp(A.Sayfa1 {j,1}, 'A/C');
AG = strcmp(A.Sayfa1 {j,1}, 'A/G');
AT = strcmp(A.Sayfa1 {j,1}, 'A/T');
CG = strcmp(A.Sayfa1 {j,1}, 'C/G');
CT = strcmp(A.Sayfa1 {j,1}, 'C/T');
GT = strcmp(A.Sayfa1 {j,1}, 'G/T');

if AC==1
    for i=1:k
        AA = strcmp(A.Sayfa1 {j,i+1}, 'AA');
        AB = strcmp(A.Sayfa1 {j,i+1}, 'AC');
        BB = strcmp(A.Sayfa1 {j,i+1}, 'CC');

        if AA==1
            N_AA(j,n)=N_AA(j,n)+1;
        end

        if AB==1
            N_AB(j,n)=N_AB(j,n)+1;
        end

        if BB==1
            N_BB(j,n)=N_BB(j,n)+1;
        end
    end
end
end

```



```

if AG==1
    for i=1:k
        AA = strcmp(A.Sayfa1 {j,i+1}, 'AA');
        AB = strcmp(A.Sayfa1 {j,i+1}, 'AG');
        BB = strcmp(A.Sayfa1 {j,i+1}, 'GG');

        if AA==1
            N_AA(j,n)=N_AA(j,n)+1;
        end

        if AB==1
            N_AB(j,n)=N_AB(j,n)+1;
        end

        if BB==1
            N_BB(j,n)=N_BB(j,n)+1;
        end
    end
end
end
if AT==1
    for i=1:k
        AA = strcmp(A.Sayfa1 {j,i+1}, 'AA');
        AB = strcmp(A.Sayfa1 {j,i+1}, 'AT');
        BB = strcmp(A.Sayfa1 {j,i+1}, 'TT');

        if AA==1
            N_AA(j,n)=N_AA(j,n)+1;
        end
    end
end

```

```

end

if AB==1
    N_AB(j,n)=N_AB(j,n)+1;
end

if BB==1
    N_BB(j,n)=N_BB(j,n)+1;
end

end

end

if CG==1
    for i=1:k
        AA = strcmp(A.Sayfa1 {j,i+1}, 'CC');
        AB = strcmp(A.Sayfa1 {j,i+1}, 'CG');
        BB = strcmp(A.Sayfa1 {j,i+1}, 'GG');

        if AA==1
            N_AA(j,n)=N_AA(j,n)+1;
        end

        if AB==1
            N_AB(j,n)=N_AB(j,n)+1;
        end

        if BB==1
            N_BB(j,n)=N_BB(j,n)+1;
        end
    end
end

```

```

    end

    end

end

if CT==1

    for i=1:k

        AA = strcmp(A.Sayfa1 {j,i+1}, 'CC');

        AB = strcmp(A.Sayfa1 {j,i+1}, 'CT');

        BB = strcmp(A.Sayfa1 {j,i+1}, 'TT');

        if AA==1

            N_AA(j,n)=N_AA(j,n)+1;

            end

        if AB==1

            N_AB(j,n)=N_AB(j,n)+1;

            end

        if BB==1

            N_BB(j,n)=N_BB(j,n)+1;

            end

        end

    end

end

if GT==1

    for i=1:k

        AA = strcmp(A.Sayfa1 {j,i+1}, 'GG');

        AB = strcmp(A.Sayfa1 {j,i+1}, 'GT');

```

```

BB = strcmp(A.Sayfal {j,i+1}, 'TT');

if AA==1
N_AA(j,n)=N_AA(j,n)+1;
end

if AB==1
    N_AB(j,n)=N_AB(j,n)+1;
end

if BB==1
    N_BB(j,n)=N_BB(j,n)+1;
end
end
end
end
clear A
end

N_AA=sum(N_AA');
N_AB=sum(N_AB');
N_BB=sum(N_BB');

N_AA=N_AA';
N_AB=N_AB';
N_BB=N_BB';

for i=1:SNP_number

```

```

f_AA(i,1)=N_AA(i,1)/(N_AA(i,1)+N_AB(i,1)+N_BB(i,1));
f_AB(i,1)=N_AB(i,1)/(N_AA(i,1)+N_AB(i,1)+N_BB(i,1));
f_BB(i,1)=N_BB(i,1)/(N_AA(i,1)+N_AB(i,1)+N_BB(i,1));
end

for i=1:SNP_number
    p(i,1)=f_AA(i,1)+1/2*f_AB(i,1);
    q(i,1)=f_BB(i,1)+1/2*f_AB(i,1);
end

for i=1:SNP_number
    if p(i,1)<=q(i,1)
        MAF(i,1)=p(i,1)*100;
    else
        MAF(i,1)=q(i,1)*100;
    end
end
end

```

## EK6. Matlab ile HapMap Popülasyonlarının Ortak rs'lerin Seçilimi (Yüksek Çözünürlük)

```
fid = fopen('chr1_list.txt');
C = textscan(fid, '%s');
[rc cc]=size(C{1,1});

for j=1:rc-1
if j==1
HM=importdata(C{1,1}{1,1});
HM2=importdata(C{1,1}{j+1,1});
[rh ch]=size(HM2.textdata.Sayfa1);
boyut(1,j)=rh;

[rs_a, HMi, HM2i] = intersect(HM.textdata.Sayfa1, HM2.textdata.Sayfa1);

else
HM2=importdata(C{1,1}{j+1,1});
[rh ch]=size(HM2.textdata.Sayfa1);
boyut(1,j)=rh;
[rs_a, rs_ai, HM2i] = intersect(rs_a, HM2.textdata.Sayfa1);

end
end

[r c]=size(rs_a);
for i=1:rc
    HM3=importdata(C{1,1}{i,1});
```

```
[rs_ak, rs_ai, HM3i] = intersect(rs_a, HM3.textdata.Sayfa1);
```

```
for j=1:r
```

```
    compare_matrix(j,i)=HM3.data.Sayfa1(HM3i(j),1);
```

```
end
```

```
end
```

## EK7. Türk Popülasyon Frekanslarının Hesaplanması

```
N_data=267;
SNP_number=262264;
N_AA(SNP_number,1)=0;
N_AB(SNP_number,1)=0;
N_BB(SNP_number,1)=0;

fid = fopen('excel_list.txt');
C = textscan(fid, '%s');
for i=1:N_data
    S{1,1}{i,1} = strcat(C{1,1}{i,1},'.xlsx');
end

for i=1:N_data
    A = importdata(S{1,1}{i,1});

    for j=1:SNP_number

        AA = strcmp(A.textdata.Sayfa1{j,3}, 'AA');
        AB = strcmp(A.textdata.Sayfa1{j,3}, 'AB');
        BB = strcmp(A.textdata.Sayfa1{j,3}, 'BB');

        if AA==1
            N_AA(j,1)=N_AA(j,1)+1;
        end

        if AB==1
```



```

    N_AB(j,1)=N_AB(j,1)+1;

end

if BB==1

    N_BB(j,1)=N_BB(j,1)+1;

end

end

end

end

for i=1:262264

    f_AA(i,1)=N_AA(i,1)/(N_AA(i,1)+N_AB(i,1)+N_BB(i,1));
    f_AB(i,1)=N_AB(i,1)/(N_AA(i,1)+N_AB(i,1)+N_BB(i,1));
    f_BB(i,1)=N_BB(i,1)/(N_AA(i,1)+N_AB(i,1)+N_BB(i,1));

end

for i=1:262264

    p(i,1)=f_AA(i,1)+1/2*f_AB(i,1);
    q(i,1)=f_BB(i,1)+1/2*f_AB(i,1);

end

for i=1:262264

    if p(i,1)<=q(i,1)

        MAF(i,1)=p(i,1)*100;

    else

        MAF(i,1)=q(i,1)*100;

    end

end

end

```

## EK8. HapMap ve Türk Popülasyon rs'lerinin Ortak olanlarının Seçilimi

```
fid = fopen('chr1_list.txt'); ## chr1_list.txt açılacak veri dosyalarını içerir
C = textscan(fid, '%s');
[rc cc]=size(C{1,1});

for j=1:rc-1
if j==1
HM=importdata(C{1,1}{1,1});
HM2=importdata(C{1,1}{j+1,1});
[rh ch]=size(HM2.textdata.Sayfa1);
boyut(1,j)=rh;

[rs_a, HMi, HM2i] = intersect(HM.textdata.Sayfa1, HM2.textdata.Sayfa1);

else
HM2=importdata(C{1,1}{j+1,1});
[rh ch]=size(HM2.textdata.Sayfa1);
boyut(1,j)=rh;
[rs_a, rs_ai, HM2i] = intersect(rs_a, HM2.textdata.Sayfa1);

end
end

[r c]=size(rs_a);
for i=1:rc
HM3=importdata(C{1,1}{i,1});
```

```
[rs_ak, rs_ai, HM3i] = intersect(rs_a, HM3.textdata.Sayfa1);
```

```
for j=1:r
```

```
    compare_matrix(j,i)=HM3.data.Sayfa1(HM3i(j),1);
```

```
end
```

```
end
```

## EK9. R 2.15.3 (2013-03-01) Programı ile Öklid Uzaklığına Göre Hiyerarşik Kümeleme Kodu

```
setwd ("c:\\pelin")

chr1=read.csv("chr1_p.csv", header=T)  ## csv dosyasını okur

cn=chr1[,1]  ## satir isimlerini verinin ilk kolonundan ceker

chr1=chr1[,-1]  ## verinin ilk kolonunu siler

chr1=t(chr1) ## chr1'in transpozunu alır

#chr1 <- as.data.frame(lapply(chr1, as.factor))  ## elektronik tablo (dataframe) yapar ve her degiskeni
faktor yapar

colnames(chr1)=cn

#str(chr1)

### euclidean index

require("cluster")  ## "cluster" paketini yükler

dist1=daisy(chr1,metric="euclid")  ## euclidean indeksi hesaplar

detach("package:cluster")  ## "cluster" paketini kaldırır

### hierarchical clustering

require("fastcluster")

require("graphics")

hc <- hclust(dist1, "average")

plot(hc, hang = -1,main="Chr1_p")
```

**EK10.** Güneybatı ABD’de Afrika Kökenlilerin (ASW) Minör Allel Sayıları (Yüksek Çözünürlük)

ASW	Karşılaştırılabilir rs	Allel A <0,5 (Allel A Minör Allel)	Allel A>0,5 (Allel B Minör Allel)	Allel A/B = 0,5 (Minör Allel Yok)
1. Kromozom	125.164	62.592	62.045	527
2. Kromozom	126.472	62.934	62.978	560
3. Kromozom	104.858	52.209	52.144	505
4. Kromozom	94.828	47.184	47.235	409
5. Kromozom	95.689	47.840	47.375	474
6. Kromozom	99.332	49.279	49.631	422
7. Kromozom	82.175	41.064	40.770	341
8. Kromozom	82.406	41.026	40.949	431
9. Kromozom	69.344	34.618	34.423	303
10. Kromozom	79.805	39.785	39.666	354
11. Kromozom	76.819	38.376	38.101	342
12. Kromozom	74.290	36.879	37.053	358
13. Kromozom	57.082	28.405	28.405	272
14. Kromozom	49.241	24.631	24.418	192
15. Kromozom	45.561	22.811	22.522	228
16. Kromozom	48.073	23.833	24.027	213
17. Kromozom	40.775	20.343	20.253	179
18. Kromozom	44.892	22.206	22.472	214
19. Kromozom	27.754	13.827	13.791	136
20. Kromozom	38.747	19.157	19.409	181
21. Kromozom	21.186	10.590	10.501	95
22. Kromozom	21.785	10.867	10.813	105
X Kromozom	54.720	27.324	27.108	288
<b>Karşılaştırılan toplam rs</b>	<b>1.560.998</b>	<b>777.780</b>	<b>776.089</b>	<b>7.129</b>

**EK11. Webuye'da Luhya Kabilesi, Kenya (LWK ) Minör Allel Sayıları (Yüksek Çözünürlük)**

LWK	Karşılaştırılabilir rs	Allel A <0,5 (Allel A Minör Allel)	Allel A>0,5 (Allel B Minör Allel)	Allel A/B = 0,5 (Minör Allel Yok)
1. Kromozom	122.489	61.034	60.949	506
2. Kromozom	124.129	61.855	61.681	593
3. Kromozom	102.919	51.371	51.114	434
4. Kromozom	93.104	46.318	46.375	411
5. Kromozom	93.834	46.803	46.599	432
6. Kromozom	96.847	48.043	48.360	444
7. Kromozom	80.666	40.276	40.008	382
8. Kromozom	80.795	40.214	40.211	370
9. Kromozom	67500	33.572	33.597	331
10. Kromozom	77.929	38.686	38.917	326
11. Kromozom	75.040	37.524	37.134	382
12. Kromozom	72.400	35.881	36.202	317
13. Kromozom	56.138	28.029	27.838	271
14. Kromozom	48.274	24.106	23.942	226
15. Kromozom	44.792	22.433	22.138	221
16. Kromozom	47.213	23.365	23.630	218
17. Kromozom	39.988	19.946	19.870	172
18. Kromozom	44.090	21.838	22.047	205
19. Kromozom	27.495	13.711	13.655	129
20. Kromozom	38016	18.814	19.023	179
21. Kromozom	20.802	10.388	10.297	117
22. Kromozom	21.162	10.531	10.536	95
X Kromozom	53.704	26.812	26.636	256
<b>Karşılaştırılan toplamlar</b>	<b>1.529.326</b>	<b>761.550</b>	<b>760.759</b>	<b>7.017</b>

**EK12.** Kinyawa'da Yaşayan Maasai'lerin, Kenya (MKK) Minör Allel Sayıları (Yüksek Çözünürlük)

MKK	Karşılaştırılabilir rs	Allel A <0,5 (Allel A Minör Allel)	Allel A>0,5 (Allel B Minör Allel)	Allel A/B = 0,5 (Minör Allel Yok)
1. Kromozom	123.157	61.645	61.249	263
2. Kromozom	124.679	62.214	62.183	282
3. Kromozom	103.429	51.733	51.485	211
4. Kromozom	93.829	46.651	46.949	229
5. Kromozom	94.545	47.399	46.942	204
6. Kromozom	97.455	48.449	48.779	227
7. Kromozom	81.168	40.466	40.510	192
8. Kromozom	81.305	40.556	40.546	203
9. Kromozom	68.067	33.996	33.908	163
10. Kromozom	78.508	39.128	39.207	173
11. Kromozom	75.554	37.855	37.537	162
12. Kromozom	72.875	36.194	36.541	140
13. Kromozom	56.677	28.303	28.251	123
14. Kromozom	48.553	24.258	24.179	116
15. Kromozom	44.854	22.411	22.328	115
16. Kromozom	47.204	23.432	23.663	109
17. Kromozom	39.983	20.030	19.855	98
18. Kromozom	44.294	21.940	22.255	99
19. Kromozom	27.154	13.535	13.539	80
20. Kromozom	38.227	18.923	19.239	65
21. Kromozom	20.964	10.483	10.440	41
22. Kromozom	21.246	10.686	10.512	48
X Kromozom	53.486	26.684	26.682	120
<b>Karşılaştırılan toplam rs</b>	<b>1.537.213</b>	<b>766.971</b>	<b>766.779</b>	<b>3.463</b>

**EK13.** Ibida'da Yaşayan Yorubalı'ların, Nijerya (YRI) Minör Allel Sayıları (Yüksek Çözünürlük)

YRI	Karşılaştırılabilir rs	Allel A <0,5 (Allel A Minör Allel)	Allel A>0,5 (Allel B Minör Allel)	Allel A/B = 0,5 (Minör Allel Yok)
1. Kromozom	313.089	156.481	155.931	677
2. Kromozom	323.672	161.783	161.121	768
3. Kromozom	255.305	127.620	127.115	570
4. Kromozom	243.525	121.633	121.278	614
5. Kromozom	246.468	122.972	122.924	572
6. Kromozom	270.243	134.839	134.829	575
7. Kromozom	212.991	106.558	105.909	524
8. Kromozom	215.260	107.980	106.804	476
9. Kromozom	182.668	91.278	90.978	412
10. Kromozom	210.285	105.166	104.637	482
11. Kromozom	202.434	101.377	100.640	417
12. Kromozom	193.705	96.662	96.637	406
13. Kromozom	156.923	78.138	78.384	401
14. Kromozom	122.651	60.985	61.398	268
15. Kromozom	106.474	53.103	53.104	267
16. Kromozom	109.668	54.276	55.154	238
17. Kromozom	90.774	45.474	45.123	177
18. Kromozom	118.999	59.276	59.439	284
19. Kromozom	58.688	29.084	29.465	139
20. Kromozom	119.071	59.000	59.864	207
21. Kromozom	51.153	25.604	25.409	140
22. Kromozom	56.748	28.429	28.222	97
X Kromozom	122.641	61.419	60.888	334
<b>Karşılaştırılan toplam rs</b>	<b>3.983.435</b>	<b>1.989.137</b>	<b>1.985.253</b>	<b>9.045</b>



**EK14.** Tokyoda yaşayan Japon'ların, Japonya (JPT) Minör Allel Sayıları (Yüksek Çözünürlük)

JPT	Karşılaştırılabilir rs	Allel A <0,5 (Allel A Minör Allel)	Allel A>0,5 (Allel B Minör Allel)	Allel A/B = 0,5 (Minör Allel Yok)
1. Kromozom	317.064	157.927	157.884	1.253
2. Kromozom	330.336	164.263	164.558	1.515
3. Kromozom	259.233	129.267	128.738	1.228
4. Kromozom	248.212	123.616	123.546	1.050
5. Kromozom	251.068	125.175	124.642	1.251
6. Kromozom	275.745	136.995	137.359	1.391
7. Kromozom	217.030	108.204	107.824	1.002
8. Kromozom	218.728	108.879	108.643	1.206
9. Kromozom	185.112	92.028	92.202	882
10. Kromozom	213.855	106.702	106.166	987
11. Kromozom	207.943	103.813	103.318	812
12. Kromozom	195.814	97.536	97.414	864
13. Kromozom	159.479	79.035	79.698	746
14. Kromozom	125.270	62.164	62.431	675
15. Kromozom	108.836	54.184	54.158	494
16. Kromozom	111.611	55.064	56.091	456
17. Kromozom	92.030	46.178	45.511	341
18. Kromozom	120.938	59.936	60.380	622
19. Kromozom	56.350	29.437	26.639	274
20. Kromozom	121.007	59.884	60.672	451
21. Kromozom	52.363	26.079	26.038	246
22. Kromozom	57.496	28.798	28.506	192
X. Kromozom	122.979	61.577	61.003	399
<b>Karşılaştırılan toplam rs</b>	<b>4.048.499</b>	<b>2.016.741</b>	<b>2.013.421</b>	<b>18.337</b>

**EK15. Beijing’li Han Çinlilerinin, Çin (CHB) Minör Allel Sayıları (Yüksek Çözünürlük)**

<b>CHB</b>	<b>Karşılaştırılabilir rs</b>	<b>Allel A &lt;0,5 (Allel A Minör Allel)</b>	<b>Allel A&gt;0,5 (Allel B Minör Allel)</b>	<b>Allel A/B = 0,5 (Minör Allel Yok)</b>
<b>1. Kromozom</b>	317.029	157.950	157.806	1.273
<b>2. Kromozom</b>	330.327	164.272	164.512	1.543
<b>3. Kromozom</b>	259.203	129.129	128.746	1.328
<b>4. Kromozom</b>	248.219	123.556	123.496	1.167
<b>5. Kromozom</b>	251.038	125.068	124.735	1.235
<b>6. Kromozom</b>	275.906	137.304	137.485	1.117
<b>7. Kromozom</b>	217.044	108.192	107.890	962
<b>8. Kromozom</b>	218.713	109.078	108.556	1079
<b>9. Kromozom</b>	185.100	92.224	92.024	852
<b>10. Kromozom</b>	213.854	106.731	106.135	988
<b>11. Kromozom</b>	207.904	103.714	103.339	851
<b>12. Kromozom</b>	195.796	97.405	97.482	909
<b>13. Kromozom</b>	159.498	78.870	79.872	756
<b>14. Kromozom</b>	125.271	62.097	62.453	721
<b>15. Kromozom</b>	108.827	54.057	54.138	632
<b>16. Kromozom</b>	111.600	55.089	56.015	496
<b>17. Kromozom</b>	91.994	46.068	45.549	377
<b>18. Kromozom</b>	120.958	59.936	60.443	579
<b>19. Kromozom</b>	59.352	29.425	29.649	278
<b>20. Kromozom</b>	121.014	59.974	60.624	416
<b>21. Kromozom</b>	52.359	26.068	26.047	244
<b>22. Kromozom</b>	57.474	28.781	28.453	240
<b>X. Kromozom</b>	122.933	61.638	61.032	263
<b>Karşılaştırılan toplamlar</b>	<b>4.051.413</b>	<b>2.016.626</b>	<b>2.016.481</b>	<b>18.306</b>

**EK16. Denver Metropolünde Yaşayan Çinlilerin, Kolarado (CHD) Minör Allel Sayıları**  
(Yüksek Çözünürlük)

CHD	Karşılaştırılabilir rs	Allel A <0,5 (Allel A Minör Allel)	Allel A>0,5 (Allel B Minör Allel)	Allel A/B = 0,5 (Minör Allel Yok)
1. Kromozom	103.706	51.766	51.447	493
2. Kromozom	105.178	52.278	52.427	473
3. Kromozom	87.824	43.703	43.661	460
4. Kromozom	79.033	39.335	39.276	422
5. Kromozom	81.155	40.429	40.275	451
6. Kromozom	85.080	42.192	42.519	369
7. Kromozom	70.020	34.903	34.810	307
8. Kromozom	68.789	34.113	34.289	387
9. Kromozom	59.262	29.508	29.442	312
10. Kromozom	67.592	33.473	33.767	352
11. Kromozom	64.479	32.245	31.954	280
12. Kromozom	62.783	31.191	31.274	318
13. Kromozom	47.806	23.690	23.870	246
14. Kromozom	41.557	20.794	20.559	204
15. Kromozom	38.648	19.225	19.240	183
16. Kromozom	39.489	19.638	19.621	230
17. Kromozom	33.630	16.827	16.660	143
18. Kromozom	37.267	18.310	18.764	193
19. Kromozom	23.528	11.764	11.631	133
20. Kromozom	32.237	15.955	16.117	165
21. Kromozom	18.143	9.111	8.920	112
22. Kromozom	18.183	9.181	8.912	90
X. Kromozom	40.409	20.219	20.033	157
<b>Karşılaştırılan toplam rs</b>	<b>1.305.798</b>	<b>649.850</b>	<b>649.468</b>	<b>6.480</b>

**EK17. Houston’da Yaşayan Gujarati Hintlilerinin, Teksas (GIH) Minör Allel Sayıları**  
(Yüksek Çözünürlük)

<b>GIH</b>	<b>Karşılaştırılabilir rs</b>	<b>Allel A &lt;0,5 (Allel A Minör Allel)</b>	<b>Allel A&gt;0,5 (Allel B Minör Allel)</b>	<b>Allel A/B = 0,5 (Minör Allel Yok)</b>
<b>1. Kromozom</b>	111.741	55.766	55.419	556
<b>2. Kromozom</b>	113.345	56.340	56.472	533
<b>3. Kromozom</b>	94.383	46.965	46.983	435
<b>4. Kromozom</b>	85.351	42.416	42.466	469
<b>5. Kromozom</b>	87.120	43.355	43.350	415
<b>6. Kromozom</b>	90.708	44.762	45.526	420
<b>7. Kromozom</b>	75.158	37.450	37.355	353
<b>8. Kromozom</b>	74.807	37.343	37.063	401
<b>9. Kromozom</b>	68.740	36.648	31.819	273
<b>10. Kromozom</b>	72.893	36.206	36.288	399
<b>11. Kromozom</b>	69.181	34.487	34.352	342
<b>12. Kromozom</b>	67.738	33.730	33.669	339
<b>13. Kromozom</b>	51.852	25.906	25.710	236
<b>14. Kromozom</b>	44.698	22.312	22.193	193
<b>15. Kromozom</b>	41.086	20.456	20.445	185
<b>16. Kromozom</b>	42.710	21.180	21.313	217
<b>17. Kromozom</b>	36.357	18.126	18.034	197
<b>18. Kromozom</b>	40.082	19.851	20.066	165
<b>19. Kromozom</b>	25.035	12.512	12.406	117
<b>20. Kromozom</b>	34.992	17.179	17.621	192
<b>21. Kromozom</b>	19.391	9.708	9.572	111
<b>22. Kromozom</b>	19.752	9.947	9.712	93
<b>X. Kromozom</b>	45.322	22.509	22.572	241
<b>Karşılaştırılan toplam rs</b>	<b>1.412.442</b>	<b>705.154</b>	<b>700.406</b>	<b>6.882</b>

**EK18.** Los Angeles’da yaşıyan Meksika Kökenlilerin, Kaliforniya (MEX) Minör Allel Sayıları (Yüksek Çözünürlük)

MEX	Karşılaştırılabilir rs	Allel A <0,5 (Allel A Minör Allel)	Allel A>0,5 (Allel B Minör Allel)	Allel A/B = 0,5 (Minör Allel Yok)
1. Kromozom	113.046	56.434	56.041	571
2. Kromozom	113995	56.584	56.810	601
3. Kromozom	94.614	47.198	46.933	483
4. Kromozom	84.253	41.751	42.091	411
5. Kromozom	86.644	43.136	43.051	457
6. Kromozom	90.521	44.799	45.256	466
7. Kromozom	74.875	37.310	37.188	377
8. Kromozom	74.056	36.825	36.814	417
9. Kromozom	63.040	31.355	31.296	389
10. Kromozom	73.680	36.557	36.756	367
11. Kromozom	69.045	34.503	34.185	357
12. Kromozom	67.875	33.782	33.769	324
13. Kromozom	51.151	25.379	25.538	234
14. Kromozom	44.830	22.347	22.255	228
15. Kromozom	40.940	20.289	20.444	207
16. Kromozom	43.122	21.415	21.480	227
17. Kromozom	36.920	18.440	18.269	211
18. Kromozom	40.385	19.946	20.209	230
19. Kromozom	25.771	12.819	12.802	150
20. Kromozom	35.679	17.486	17.984	209
21. Kromozom	19.265	9.643	9.545	77
22. Kromozom	19.692	9.853	9.730	109
X. Kromozom	46.475	23.063	23.136	276
<b>Karşılaştırılan toplam rs</b>	<b>1.409.874</b>	<b>700.914</b>	<b>701.582</b>	<b>7.378</b>

**EK19.** CEPH Koleksiyonundan Kuzey ve Batı Avrupa Kökenli olup Utah'a Yerleşenlerin (CEU) Minör Allel Sayıları (Yüksek Çözünürlük)

CEU	Karşılaştırılabilir rs	Allel A <0,5 (Allel A Minör Allel)	Allel A>0,5 (Allel B Minör Allel)	Allel A/B = 0,5 (Minör Allel Yok)
1. Kromozom	314.024	156.804	156.546	674
2. Kromozom	329.831	164.486	164.499	846
3. Kromozom	259.412	129.548	129.262	602
4. Kromozom	248.215	123.852	123.755	608
5. Kromozom	250.737	125.160	124.924	653
6. Kromozom	272.260	135.553	135.987	720
7. Kromozom	216.504	108.008	107.902	594
8. Kromozom	215.909	108.040	107.302	567
9. Kromozom	183.866	91.883	91.556	427
10. Kromozom	211.711	105.797	105.461	453
11. Kromozom	206.952	103.343	103.066	543
12. Kromozom	195.155	97.677	97.035	443
13. Kromozom	157.463	78.136	78.937	390
14. Kromozom	124.726	61.937	62.470	319
15. Kromozom	108.464	54.011	54.190	263
16. Kromozom	111.845	55.366	56.216	263
17. Kromozom	92.358	46.256	45.865	237
18. Kromozom	120.207	59.852	60.097	258
19. Kromozom	59.607	29.659	29.803	145
20. Kromozom	121.069	60.067	60.810	192
21. Kromozom	50.983	25.461	25.389	133
22. Kromozom	55.941	28.114	27.699	128
X. Kromozom	122.601	61.427	60.939	235
<b>Karşılaştırılan toplam rs</b>	<b>4.029.840</b>	<b>2.010.437</b>	<b>2.009.710</b>	<b>9.693</b>

**EK20. İtalya Yaşayan Toskanalıların (TSI) Minör Allel Sayıları (Yüksek Çözünürlük)**

TSI	Karşılaştırılabilir rs	Allel A <0,5 (Allel A Minör Allel)	Allel A>0,5 (Allel B Minör Allel)	Allel A/B = 0,5 (Minör Allel Yok)
1. Kromozom	112.832	56.443	55.872	517
2. Kromozom	114.423	56.737	57.091	595
3. Kromozom	95.327	47.391	47.422	514
4. Kromozom	86.343	42.782	43.124	437
5. Kromozom	87.645	43.548	43.685	412
6. Kromozom	92.312	45.623	46.203	486
7. Kromozom	75.898	37.696	37.862	340
8. Kromozom	75.166	37.485	37.337	344
9. Kromozom	64.017	31.834	31.860	323
10. Kromozom	73.287	36.361	36.553	373
11. Kromozom	69.851	34.732	34.802	317
12. Kromozom	67.478	33.639	33.550	289
13. Kromozom	52.346	25.956	26.153	237
14. Kromozom	45.153	22.427	22.488	238
15. Kromozom	41.300	20.389	20.707	204
16. Kromozom	43.081	21.461	21.391	229
17. Kromozom	36.980	18.509	18.287	184
18. Kromozom	40.511	20.042	20.264	205
19. Kromozom	25.310	12.651	12.539	120
20. Kromozom	35.327	17.407	17.745	175
21. Kromozom	19.620	9.862	9.683	75
22. Kromozom	19.943	10.008	9.820	115
X. Kromozom	45.376	22.366	22.805	205
<b>Karşılaştırılan toplamlar</b>	<b>1.419.526</b>	<b>705.349</b>	<b>707.243</b>	<b>6.934</b>

**EK21. HapMap Afrika Popülasyonlarının Minör Allel Sayıları (Düşük Çözünürlük)**

AFRİKA	Karşılaştırılabilir rs	ASW / LWK / MKK / YRI		
		Allel A <0,5 (Allel A Minör Allel)	Allel A>0,5 (Allel B Minör Allel)	Allel A/B = 0,5 (Minör Allel Yok)
1. Kromozom	12.545	5.377	5.472	-
2. Kromozom	14.375	6.121	6.160	-
3. Kromozom	12.059	5.168	5.196	-
4. Kromozom	11.828	5.095	5.053	-
5. Kromozom	11.177	4.793	4.771	-
6. Kromozom	11.418	4.943	4.975	-
7. Kromozom	9.185	3.963	4.018	-
8. Kromozom	9.647	4.134	4.279	-
9. Kromozom	7.647	3.378	3.315	-
10. Kromozom	9.202	4.066	4.019	-
11. Kromozom	8.493	3.641	3.711	-
12. Kromozom	8.390	3.680	3.587	-
13. Kromozom	6.987	3.038	3.046	-
14. Kromozom	5.297	2.352	2.292	-
15. Kromozom	4.558	1.925	1.990	-
16. Kromozom	4.598	2.009	1.959	-
17. Kromozom	3.146	1.334	1.355	-
18. Kromozom	5.177	2.228	2.240	-
19. Kromozom	1.757	753	757	-
20. Kromozom	3.805	1.591	1.669	-
21. Kromozom	2.578	1.116	1.121	-
22. Kromozom	1.550	679	647	-
X Kromozom	3.574	1.497	1.441	-
<b>Karşılaştırılan toplam rs</b>	<b>168.993</b>	<b>72.881</b>	<b>73.073</b>	<b>-</b>



**EK22. HapMap Asya Popülasyonlarının Minör Allel Sayıları (Düşük Çözünürlük)**

ASYA	JPT / CHB / CHD			
	Karşılaştırılabilir rs	Allel A <0,5 (Allel A Minör Allel)	Allel A>0,5 (Allel B Minör Allel)	Allel A/B = 0,5 (Minör Allel Yok)
1. Kromozom	12.545	5.761	5.739	-
2. Kromozom	14.375	6.534	6.596	-
3. Kromozom	12.059	5.458	5.524	-
4. Kromozom	11.828	5.519	5.396	-
5. Kromozom	11.177	5.024	5.273	-
6. Kromozom	11.418	5.154	5.302	-
7. Kromozom	9.185	4.228	4.197	-
8. Kromozom	9.647	4.320	4.420	-
9. Kromozom	7.647	3.491	3.526	-
10. Kromozom	9.202	4.218	4.220	-
11. Kromozom	8.493	3.845	3.869	-
12. Kromozom	8.390	3.841	3.785	-
13. Kromozom	6.987	3.133	3.189	-
14. Kromozom	5.297	2.462	2.353	-
15. Kromozom	4.558	2.014	2.120	-
16. Kromozom	4.598	2.170	2.059	-
17. Kromozom	3.146	1.413	1.451	-
18. Kromozom	5.177	2.357	2.399	-
19. Kromozom	1.757	774	814	-
20. Kromozom	3.805	1.694	1.769	-
21. Kromozom	2.578	1.141	1.192	-
22. Kromozom	1.550	714	701	-
X. Kromozom	3.574	1.054	1.020	-
<b>Karşılaştırılan toplam rs</b>	<b>168.993</b>	<b>76.319</b>	<b>76.914</b>	

**EK23. HapMap ABD+Avrupa Popülasyonlarının Minör Allel Sayıları (Düşük Çözünürlük)**

ABD+AVRUPA	GIH / MEX / CEU / TSI			
	Karşılaştırılabilir rs	Allel A <0,5 (Allel A Minör Allel)	Allel A>0,5 (Allel B Minör Allel)	Allel A/B = 0,5 (Minör Allel Yok)
1. Kromozom	12.545	4.994	4.950	
2. Kromozom	14.375	5.749	5.840	-
3. Kromozom	12.059	4.698	4.828	-
4. Kromozom	11.828	4.778	4.693	-
5. Kromozom	11.177	4.500	4.666	-
6. Kromozom	11.418	4.494	4.658	-
7. Kromozom	9.185	3.704	3.725	-
8. Kromozom	9.647	3.780	3.828	-
9. Kromozom	7.647	3.080	3.095	-
10. Kromozom	9.202	3.684	3.794	-
11. Kromozom	8.493	3.357	3.420	-
12. Kromozom	8.390	3.503	3.376	-
13. Kromozom	6.987	2.819	2.883	-
14. Kromozom	5.297	2.147	2.101	-
15. Kromozom	4.558	1.782	1.872	-
16. Kromozom	4.598	1.838	1.826	-
17. Kromozom	3.146	1.226	1.224	-
18. Kromozom	5.177	2.125	2.121	-
19. Kromozom	1757	721	709	-
20. Kromozom	3.805	1.163	1.309	
21. Kromozom	2.578	1.037	1.070	-
22. Kromozom	1.550	603	591	-
X. Kromozom	3.574	1.292	1.299	-
<b>Karşılaştırılan toplam rs</b>	<b>168.993</b>	<b>67.074</b>	<b>67.878</b>	

**EK24. HapMap 4A Popülasyonları Minör Allel Sayıları (Düşük Çözünürlük)**

	<b>AFRİKA/ASYA/ABD+AVRUPA</b>			
	<b>Karşılaştırılabilir rs</b>	<b>Allel A &lt;0,5 (Allel A Minör Allel)</b>	<b>Allel A&gt;0,5 (Allel B Minör Allel)</b>	<b>Allel A/B = 0,5 (Minör Allel Yok)</b>
<b>1. Kromozom</b>	12.545	3.504	3.503	-
<b>2. Kromozom</b>	14.375	3.899	3.964	-
<b>3. Kromozom</b>	12.059	3.260	3.337	-
<b>4. Kromozom</b>	11.828	3.297	3.241	-
<b>5. Kromozom</b>	11.177	3.110	3.183	-
<b>6. Kromozom</b>	11.418	3.133	3.244	-
<b>7. Kromozom</b>	9.185	2.615	2.602	-
<b>8. Kromozom</b>	9.647	2.592	2.718	-
<b>9. Kromozom</b>	7.647	2.206	2.180	-
<b>10. Kromozom</b>	9.202	2.659	2.692	-
<b>11. Kromozom</b>	8.493	2.373	2.386	-
<b>12. Kromozom</b>	8.390	2.481	2.411	-
<b>13. Kromozom</b>	6.987	1.986	2.013	-
<b>14. Kromozom</b>	5.297	1.520	1.467	-
<b>15. Kromozom</b>	4.558	1.189	1.282	-
<b>16. Kromozom</b>	4.598	1.338	1.289	-
<b>17. Kromozom</b>	3.146	791	826	-
<b>18. Kromozom</b>	5.177	1.489	1.497	-
<b>19. Kromozom</b>	1.757	493	511	-
<b>20. Kromozom</b>	3.805	789	910	-
<b>21. Kromozom</b>	2.578	727	747	-
<b>22. Kromozom</b>	1.550	419	406	-
<b>X Kromozom</b>	3.574	472	475	-
<b>Karşılaştırılan toplam rs</b>	<b>168.993</b>	<b>46.342</b>	<b>46.884</b>	<b>-</b>

**EK25. HapMap Afrika ve Türk Popülasyonunun Ortak Minör Allel Sayıları (Düşük Çözünürlük)**

AFRİKA	Karşılaştırılabilir rs	ASW / LWK / MKK / YRI			
		Allel A < 0,5	Allel A >0,5	Allel A < 0,5	Allel A >0,5
TÜRKİYE		Allel A < 0,5	Allel A > 0,5	Allel A >0,5	Allel A < 0,5
1. Kromozom	12.545	2.682	2.767	2.688	2.698
2. Kromozom	14.375	3.040	3.239	3.072	2.909
3. Kromozom	12.059	2.435	2.687	2.725	2.503
4. Kromozom	11.828	2.458	2.508	2.629	2.535
5. Kromozom	11.177	2.300	2.453	2.487	2.308
6. Kromozom	11.418	2.421	2.555	2.512	2.417
7. Kromozom	9.185	1.940	2.023	2.014	1.986
8. Kromozom	9.647	2.007	2.228	2.122	2.044
9. Kromozom	7.647	1.656	1.707	1.719	1.607
10. Kromozom	9.202	1.932	2.004	2.130	2.006
11. Kromozom	8.493	1.800	1.851	1.831	1.850
12. Kromozom	8.390	1.815	1.827	1.860	1.753
13. Kromozom	6.987	1.439	1.538	1.593	1.506
14. Kromozom	5.297	1.102	1.159	1.246	1.129
15. Kromozom	4.558	931	986	986	999
16. Kromozom	4.598	972	1.006	1.035	953
17. Kromozom	3.146	626	725	703	630
18. Kromozom	5.177	1.102	1.194	1.124	1.042
19. Kromozom	1.757	392	382	359	375
20. Kromozom	3.805	740	860	845	806
21. Kromozom	2.578	536	563	575	556
22. Kromozom	1.550	316	322	363	322
X Kromozom	3.574	697	731	797	705
<b>Toplam rs</b>	<b>168.993</b>	<b>35.339</b>	<b>37.315</b>	<b>37.415</b>	<b>35.639</b>

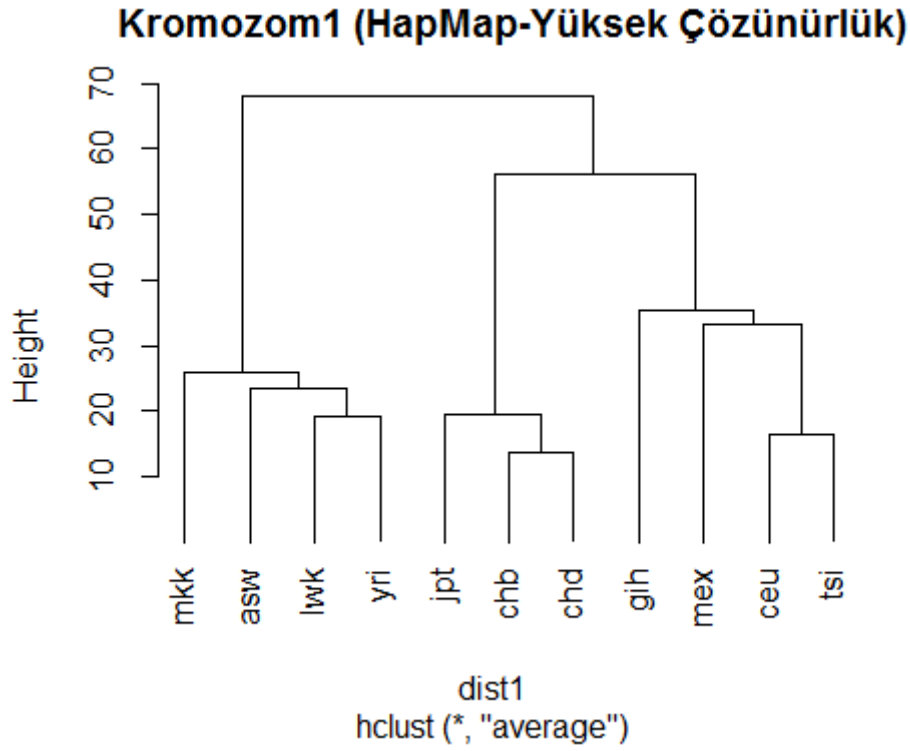
**EK26. HapMap Asya ve Türk Popülasyonunun Ortak Minör Allel Sayıları (Düşük Çözünürlük)**

	Karşılaştırılabilir rs	JPT / CHB / CHD			
		ASYA Allel A < 0,5	ASYA Allel A >0,5	ASYA Allel A < 0,5	ASYA Allel A >0,5
		TÜRKİYE Allel A < 0,5	TÜRKİYE Allel A > 0,5	TÜRKİYE Allel A >0,5	TÜRKİYE Allel A < 0,5
<b>1. Kromozom</b>	12.545	2.847	2.950	2.904	2.780
<b>2. Kromozom</b>	14.375	3.208	3.423	3.316	3.158
<b>3. Kromozom</b>	12.059	2.579	2.858	2.876	2.653
<b>4. Kromozom</b>	11.828	2.648	2.690	2.863	2.693
<b>5. Kromozom</b>	11.177	2.403	2.696	2.610	2.470
<b>6. Kromozom</b>	11.418	2.509	2.689	2.638	2.608
<b>7. Kromozom</b>	9.185	2.036	2.136	2.183	2.051
<b>8. Kromozom</b>	9.647	2.091	2.240	2.218	2.175
<b>9. Kromozom</b>	7.647	1.709	1.780	1.779	1.743
<b>10. Kromozom</b>	9.202	2.014	2.122	2.199	2.087
<b>11. Kromozom</b>	8.493	1.875	1.912	1.961	1.947
<b>12. Kromozom</b>	8.390	1.910	1.970	1.922	1.805
<b>13. Kromozom</b>	6.987	1.486	1.601	1.645	1.583
<b>14. Kromozom</b>	5.297	1.157	1.198	1.299	1.154
<b>15. Kromozom</b>	4.558	979	1.028	1.029	1.085
<b>16. Kromozom</b>	4.598	1.035	1.066	1.132	993
<b>17. Kromozom</b>	3.146	665	779	744	669
<b>18. Kromozom</b>	5.177	1.121	1.236	1.232	1.161
<b>19. Kromozom</b>	1.757	393	415	381	396
<b>20. Kromozom</b>	3.805	790	892	900	870
<b>21. Kromozom</b>	2.578	531	595	607	595
<b>22. Kromozom</b>	1.550	327	355	387	343
<b>23. Kromozom</b>	3.574	494	521	558	497
<b>Toplam rs</b>	<b>168.993</b>	<b>36.807</b>	<b>39.152</b>	<b>39.383</b>	<b>37.516</b>

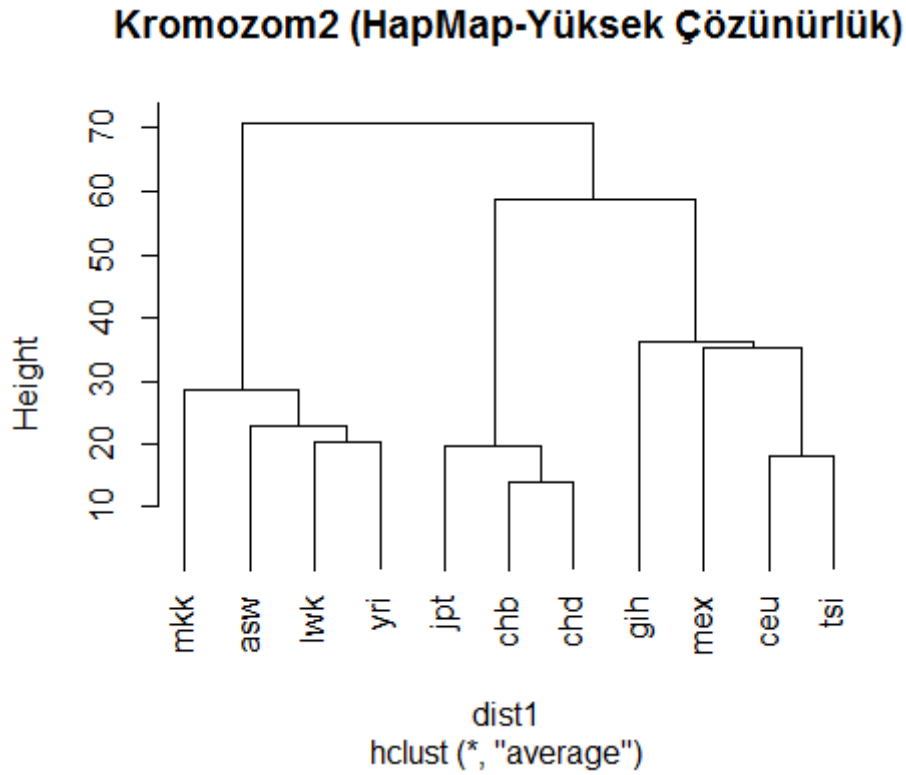
**EK27. HapMap ABD+Avrupa ve Türk Popülasyonunun Ortak Minör Allel Sayıları (Düşük Çözünürlük)**

ABD+AVRUPA	Karşılaştırılabilir rs	GIH / MEX / CEU / TSI			
		Allel A < 0,5	Allel A >0,5	Allel A < 0,5	AS YA Allel A >0,5
TÜRKİYE		Allel A < 0,5	Allel A > 0,5	Allel A >0,5	Allel A < 0,5
<b>1. Kromozom</b>	12.545	2.477	2.540	2.517	2.409
<b>2. Kromozom</b>	14.375	2.820	3.083	2.928	2.754
<b>3. Kromozom</b>	12.059	2.243	2.531	2.454	2.294
<b>4. Kromozom</b>	11.828	2.300	2.348	2.476	2.341
<b>5. Kromozom</b>	11.177	2.147	2.416	2.351	2.448
<b>6. Kromozom</b>	11.418	2.187	2.366	2.289	2.307
<b>7. Kromozom</b>	9.185	1.825	1.928	1.875	1.791
<b>8. Kromozom</b>	9.647	1.834	1.925	1.944	1.901
<b>9. Kromozom</b>	7.647	1.524	1.597	1.554	1.497
<b>10. Kromozom</b>	9.202	1.773	1.937	1.910	1.854
<b>11. Kromozom</b>	8.493	1.637	1.712	1.716	1.706
<b>12. Kromozom</b>	8.390	1.745	1.764	1.758	1.609
<b>13. Kromozom</b>	6.987	1.335	1.446	1.481	1.437
<b>14. Kromozom</b>	5.297	1.002	1.047	1.142	1.054
<b>15. Kromozom</b>	4.558	880	930	901	939
<b>16. Kromozom</b>	4.598	878	937	960	889
<b>17. Kromozom</b>	3.146	588	639	638	582
<b>18. Kromozom</b>	5.177	1.030	1.120	1.095	1.000
<b>19. Kromozom</b>	1.757	372	360	349	360
<b>20. Kromozom</b>	3.805	526	673	636	633
<b>21. Kromozom</b>	2.578	508	538	527	531
<b>22. Kromozom</b>	1.550	285	312	318	279
<b>X Kromozom</b>	3.574	596	645	695	653
<b>Toplam rs</b>	<b>168.993</b>	<b>32.512</b>	<b>34.794</b>	<b>34.514</b>	<b>33.268</b>

**EK28. 1. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Yüksek Çözünürlük)**

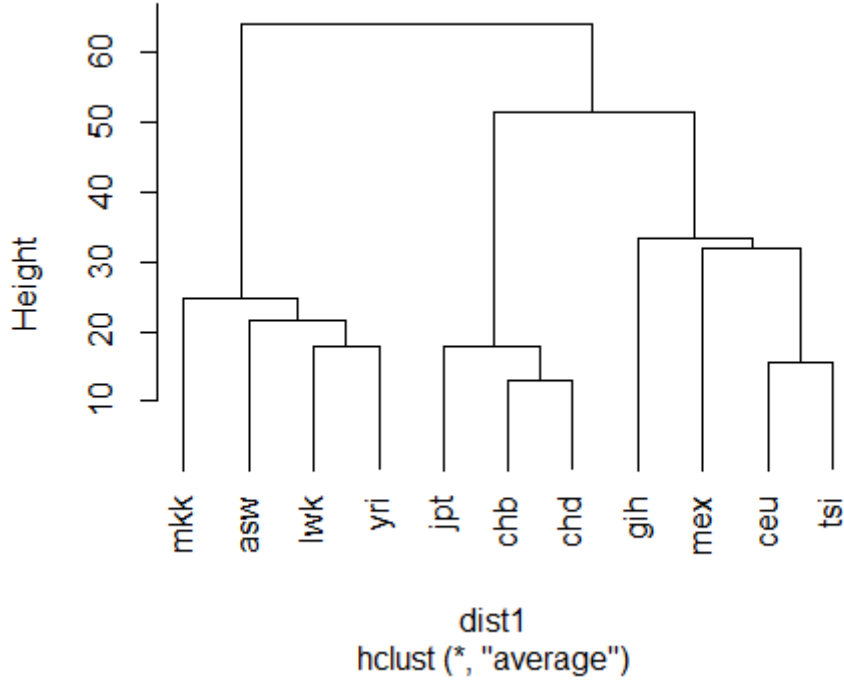


**EK29. 2. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Yüksek Çözünürlük)**



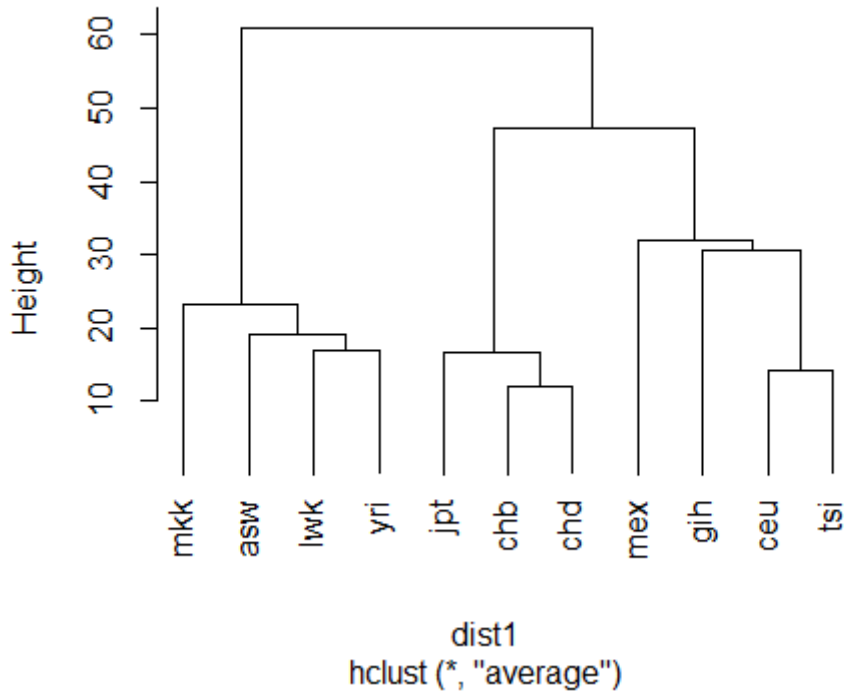
**EK30. 3. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Yüksek Çözünürlük)**

**Kromozom3 (HapMap-Yüksek Çözünürlük)**



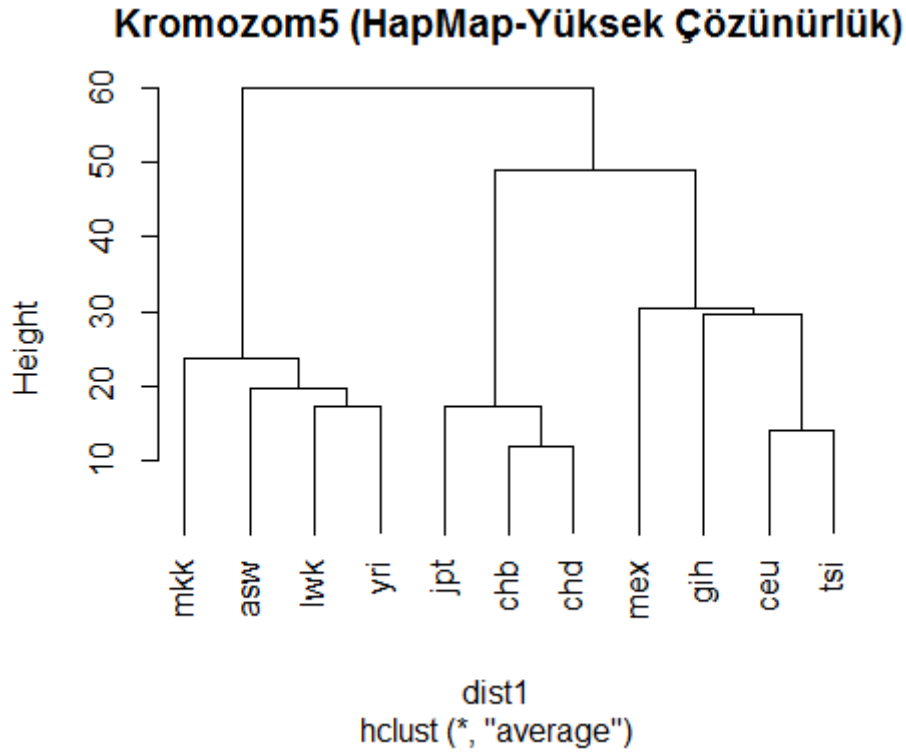
**EK31. 4. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Yüksek Çözünürlük)**

**Kromozom4 (HapMap-Yüksek Çözünürlük)**

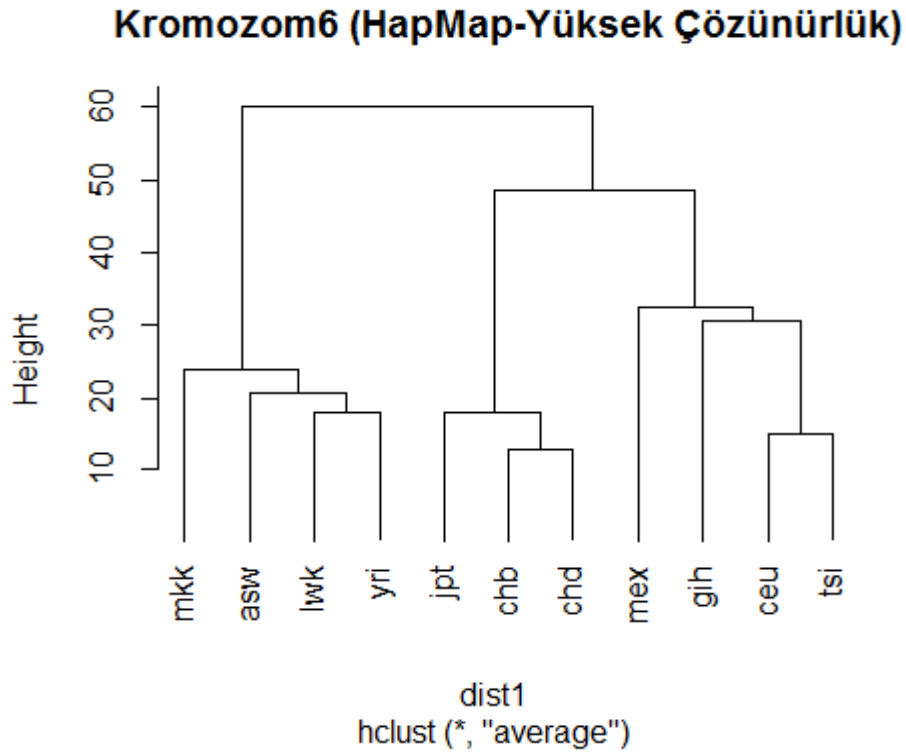




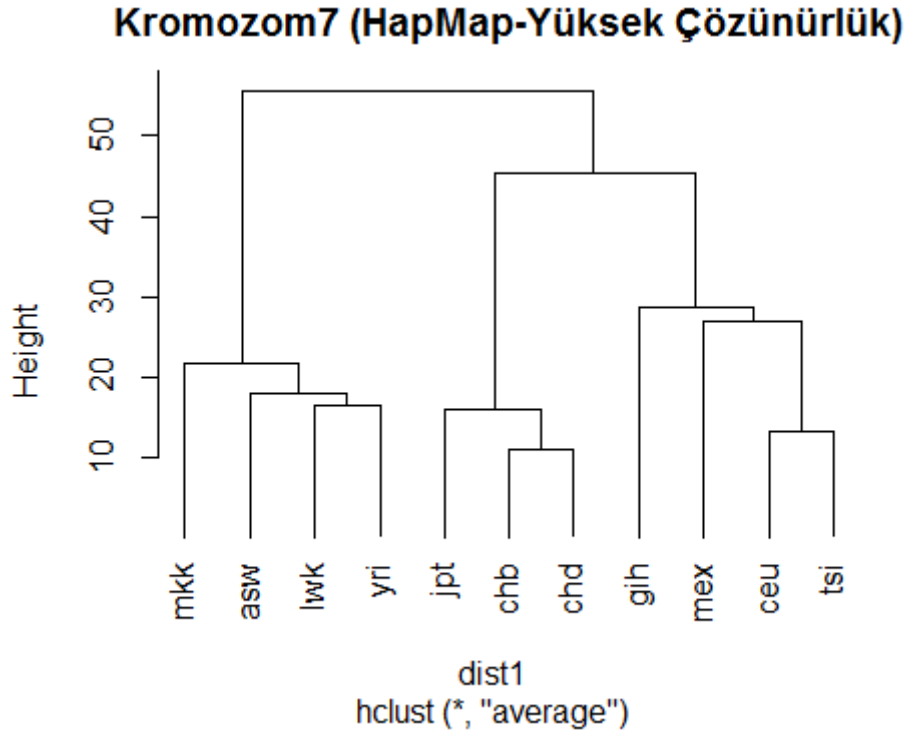
**EK32.** 5. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Yüksek Çözünürlük)



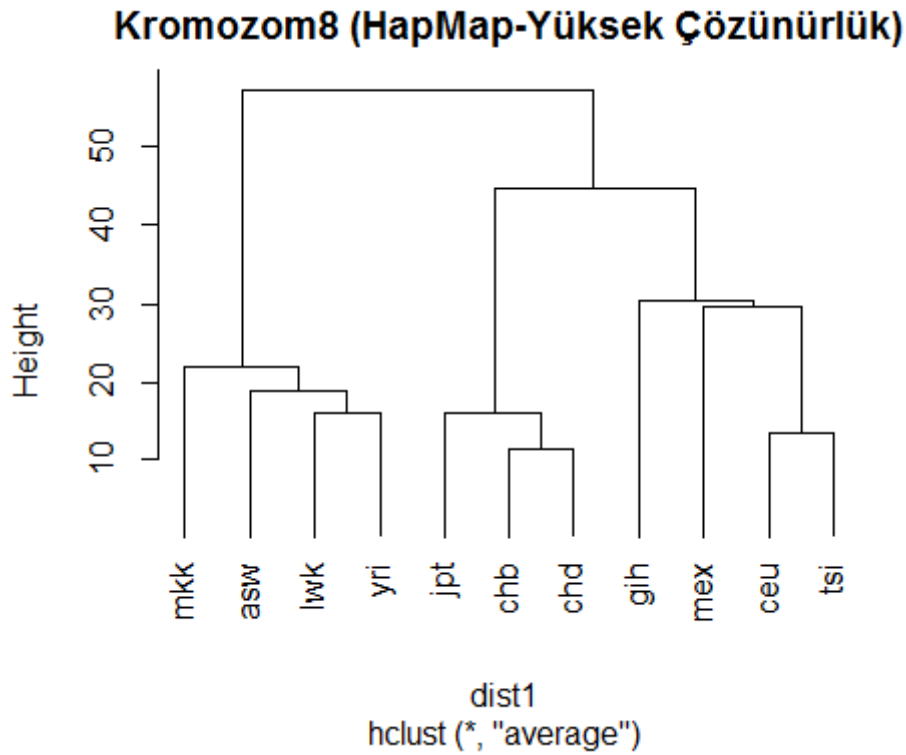
**EK33.** 6. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Yüksek Çözünürlük)



**EK34.** 7. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Yüksek Çözünürlük)

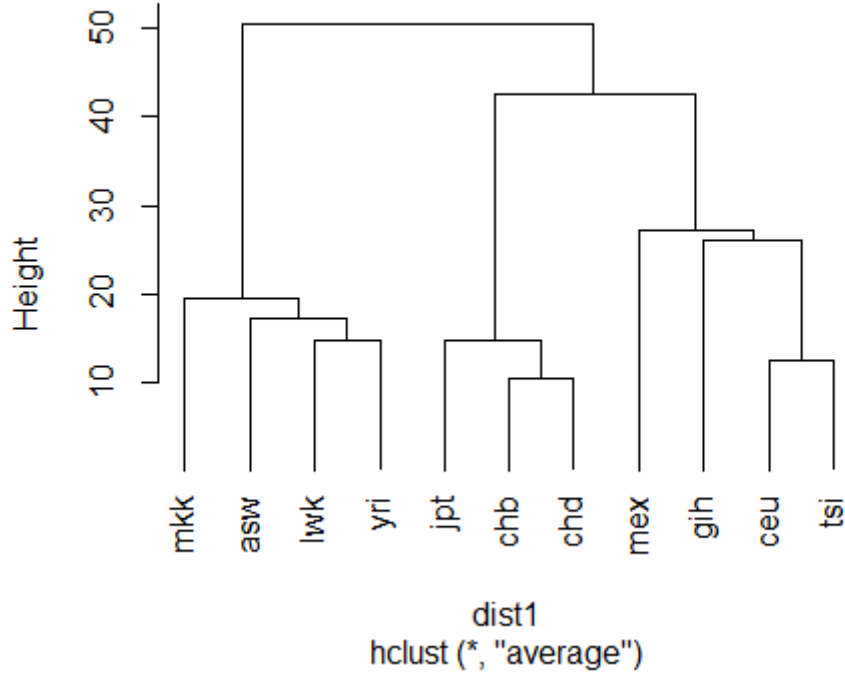


**EK35.** 8. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Yüksek Çözünürlük)



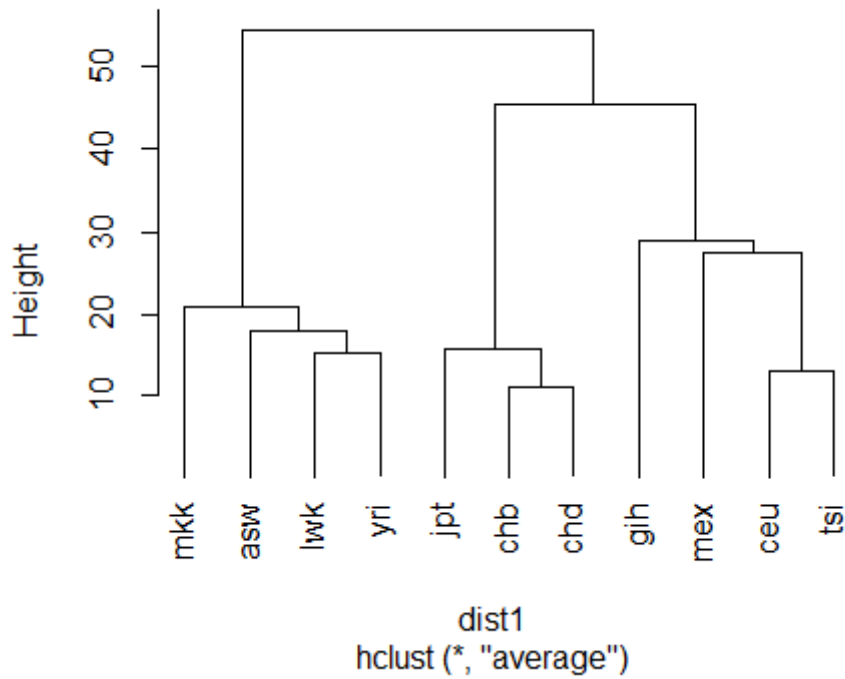
**EK36.** 9. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Yüksek Çözünürlük)

**Kromozom9 (HapMap-Yüksek Çözünürlük)**



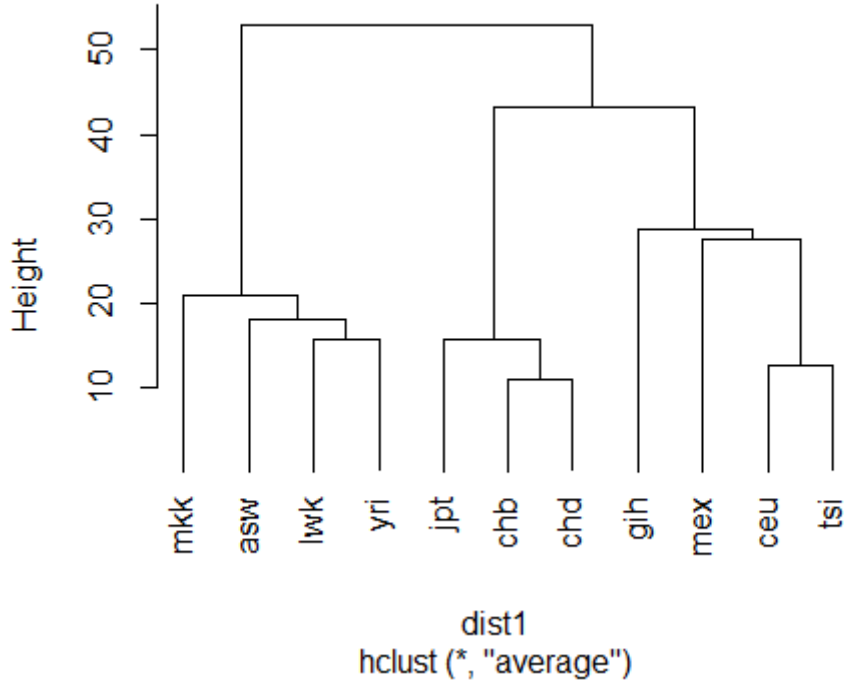
**EK37.** 10. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Yüksek Çözünürlük)

**Kromozom10 (HapMap-Yüksek Çözünürlük)**



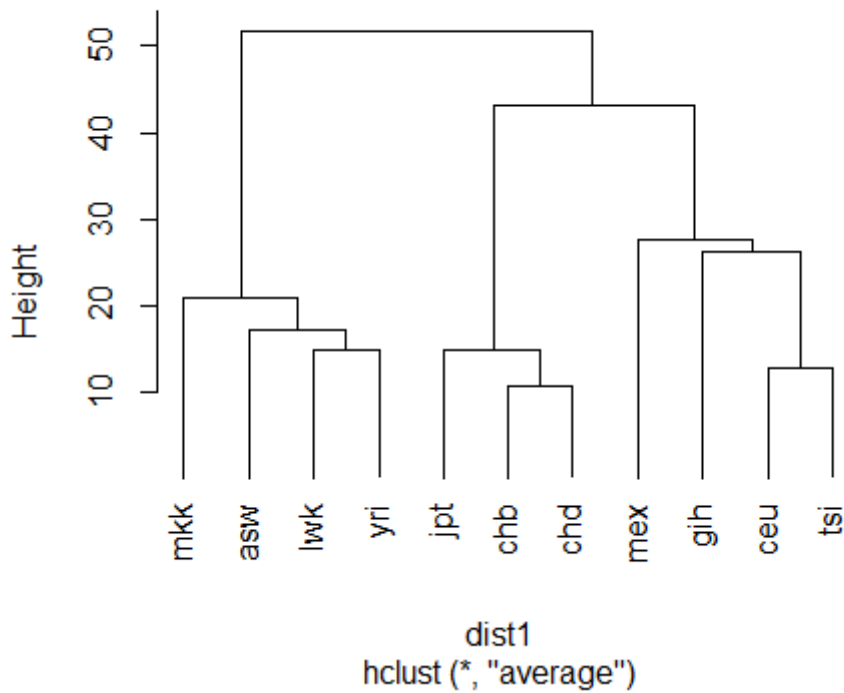
**EK38.** 11. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Yüksek Çözünürlük)

**Kromozom11 (HapMap-Yüksek Çözünürlük)**



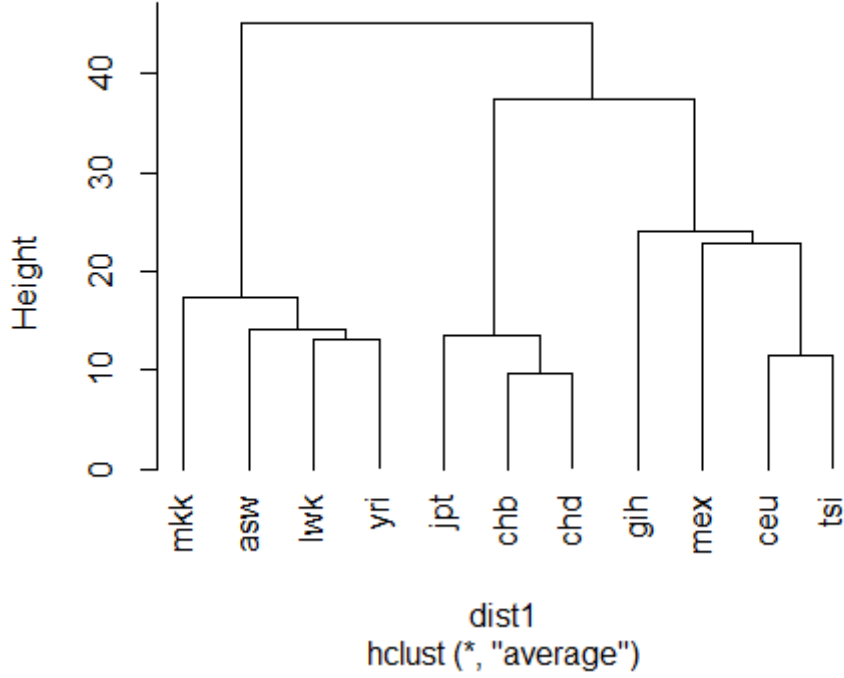
**EK39.** 12. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Yüksek Çözünürlük)

**Kromozom12 (HapMap-Yüksek Çözünürlük)**



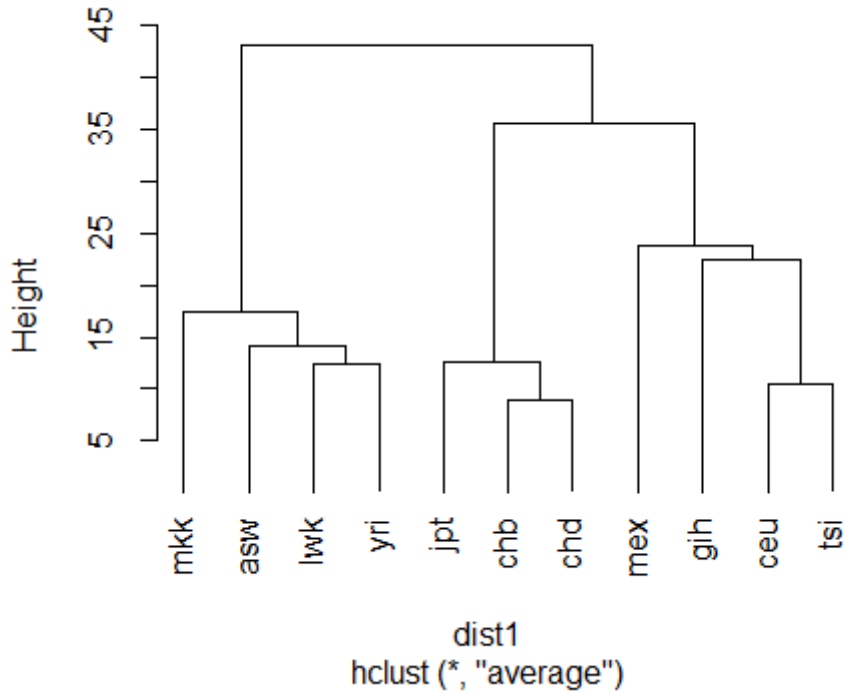
**EK40.** 13. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Yüksek Çözünürlük)

### Kromozom13 (HapMap-Yüksek Çözünürlük)

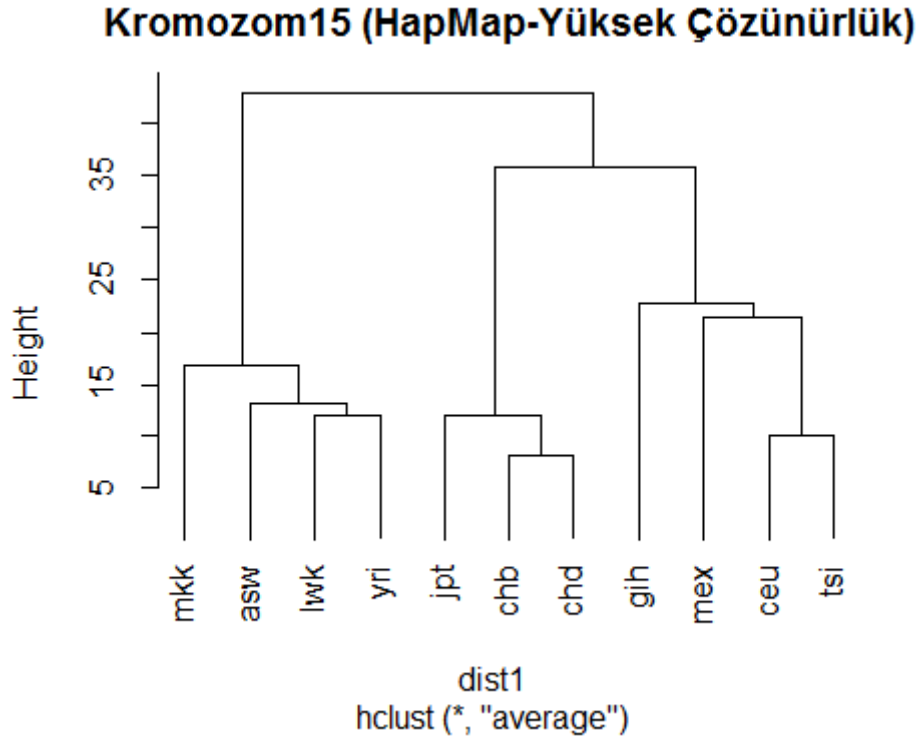


**EK41.** 14. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Yüksek Çözünürlük)

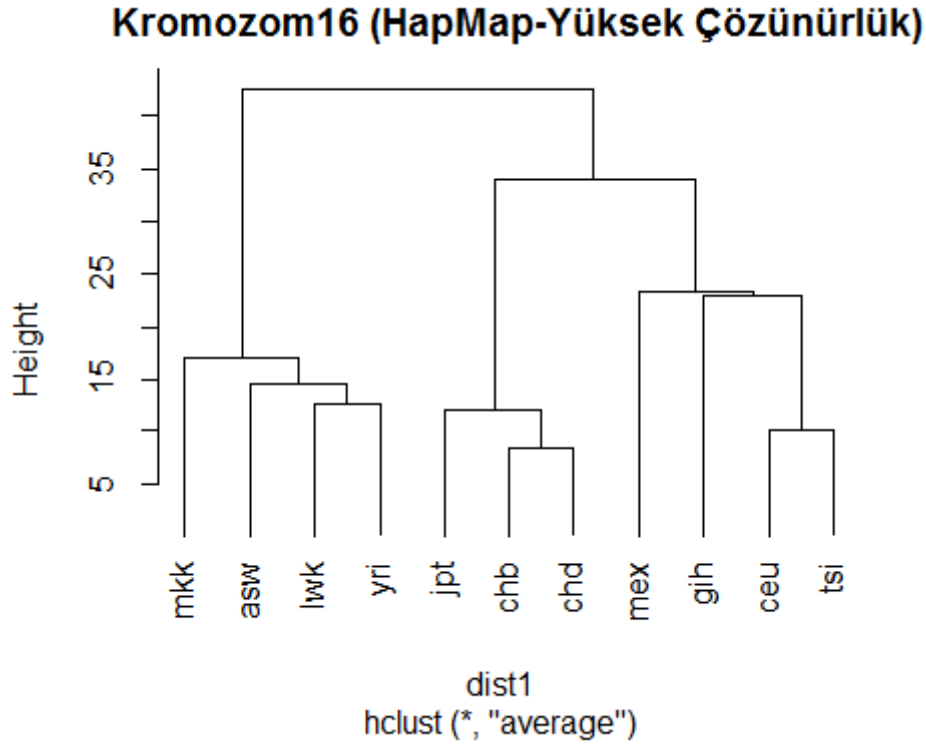
### Kromozom14 (HapMap-Yüksek Çözünürlük)



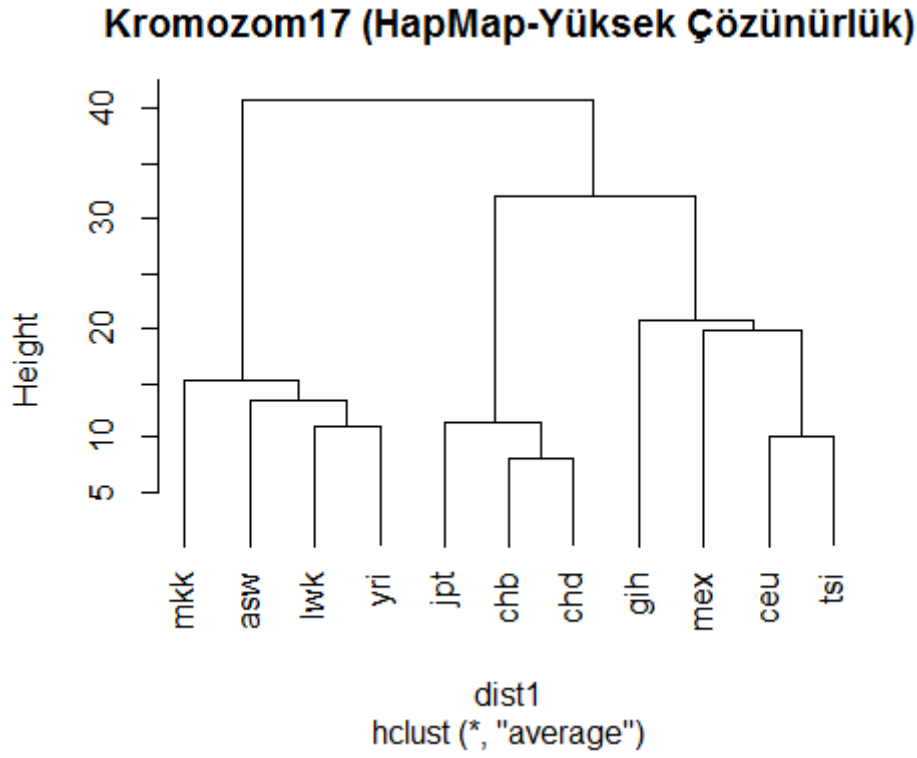
EK42. 15. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Yüksek Çözünürlük)



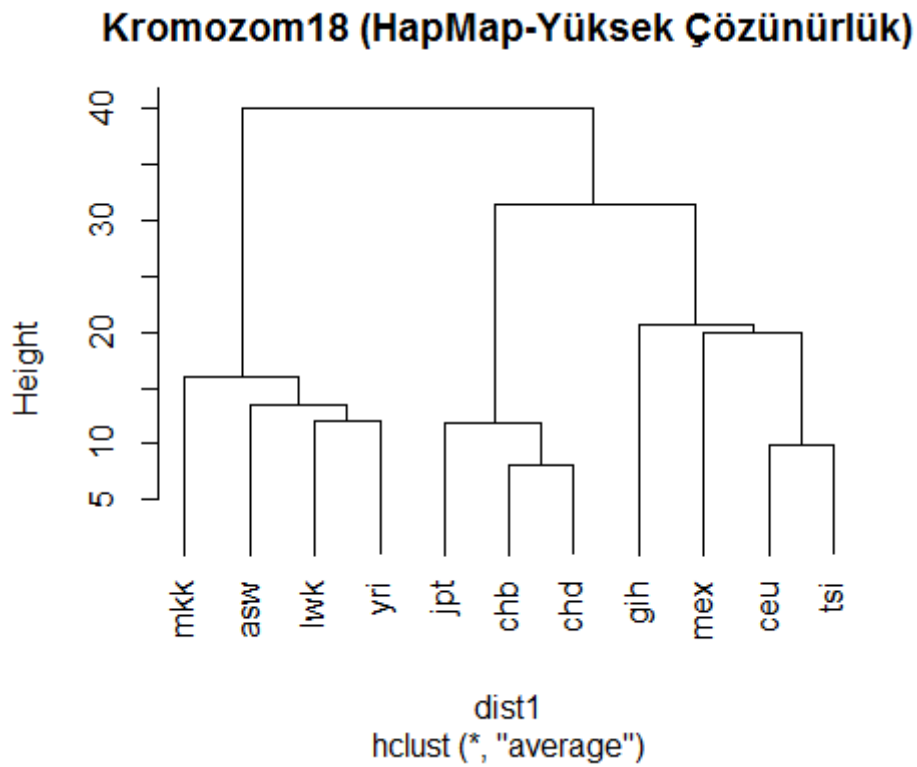
EK43. 16. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Yüksek Çözünürlük)



EK44. 17. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Yüksek Çözünürlük)

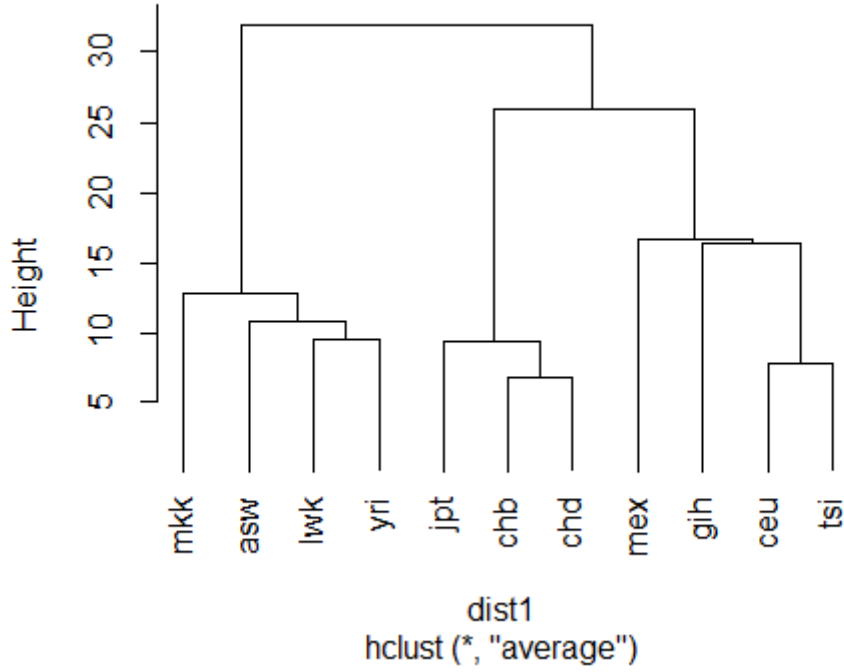


EK45. 18. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Yüksek Çözünürlük)



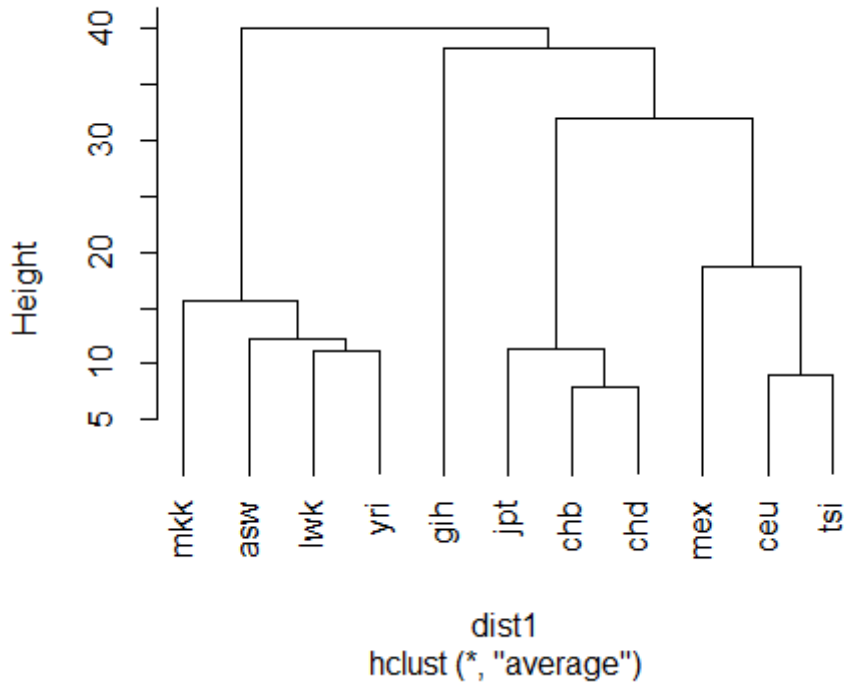
EK46. 19. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Yüksek Çözünürlük)

### Kromozom19 (HapMap-Yüksek Çözünürlük)



EK47. 20. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Yüksek Çözünürlük)

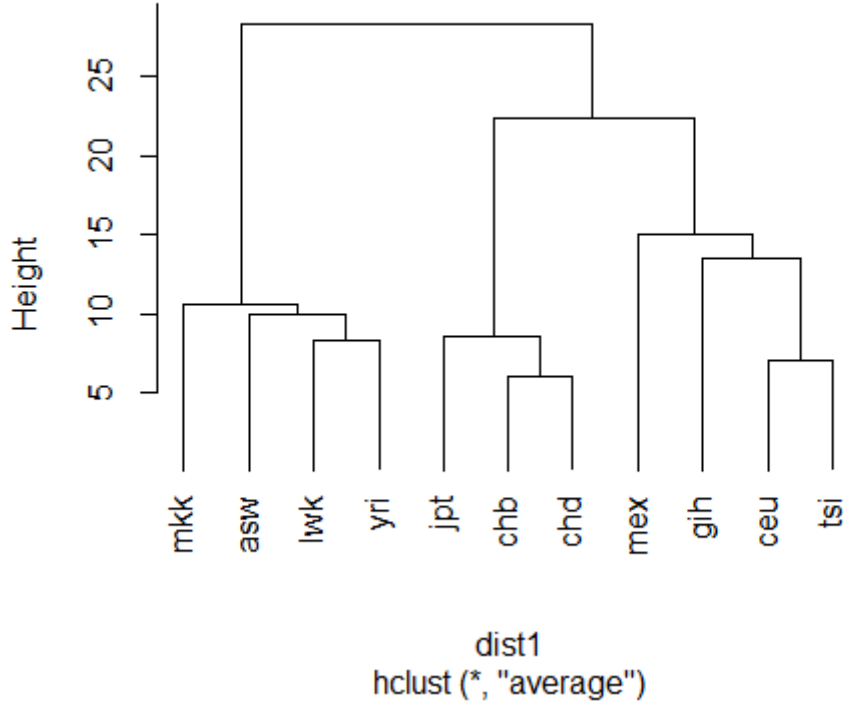
### Kromozom20 (HapMap-Yüksek Çözünürlük)





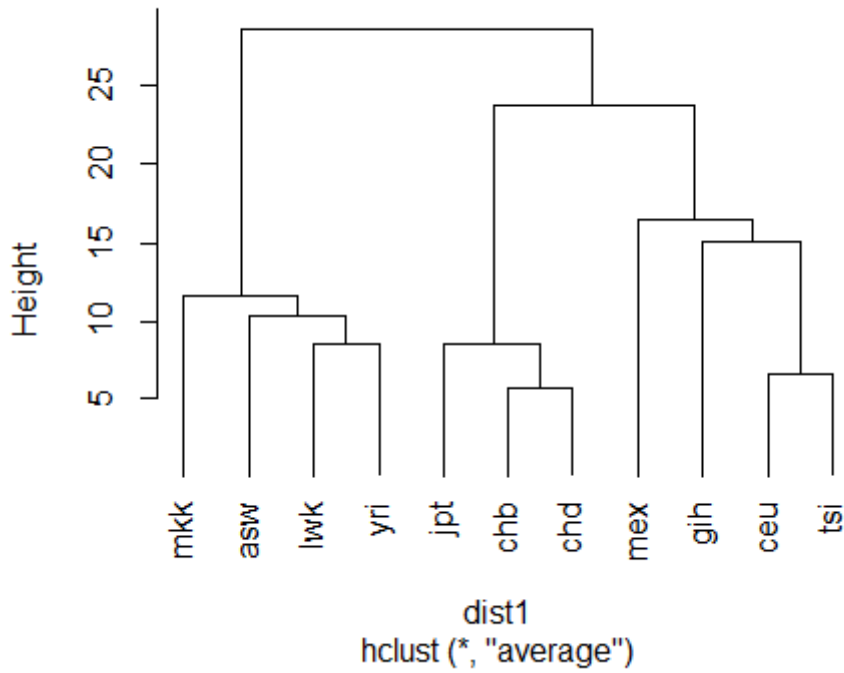
EK48. 21. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Yüksek Çözünürlük)

### Kromozom21 (HapMap-Yüksek Çözünürlük)



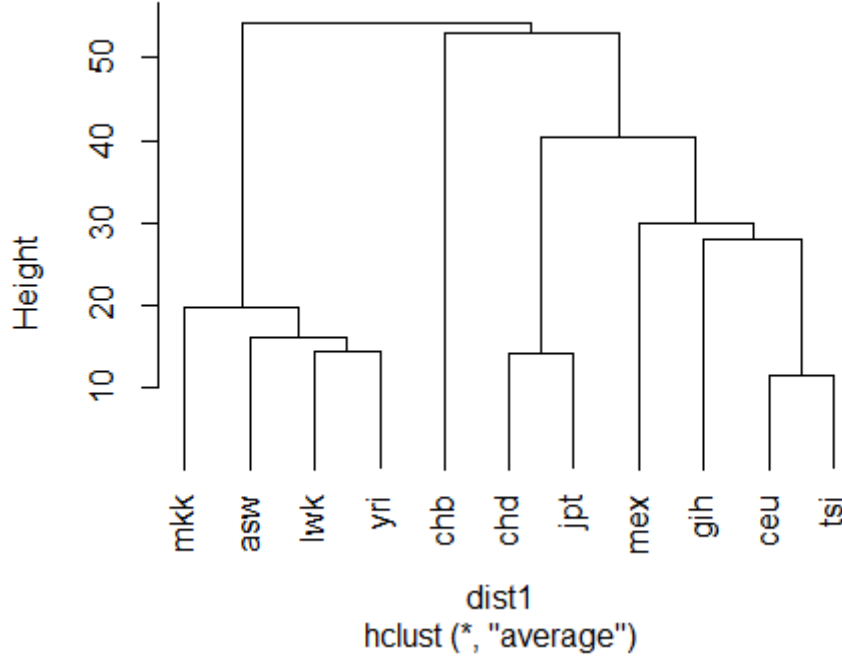
EK49. 22. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Yüksek Çözünürlük)

### Kromozom22 (HapMap-Yüksek Çözünürlük)



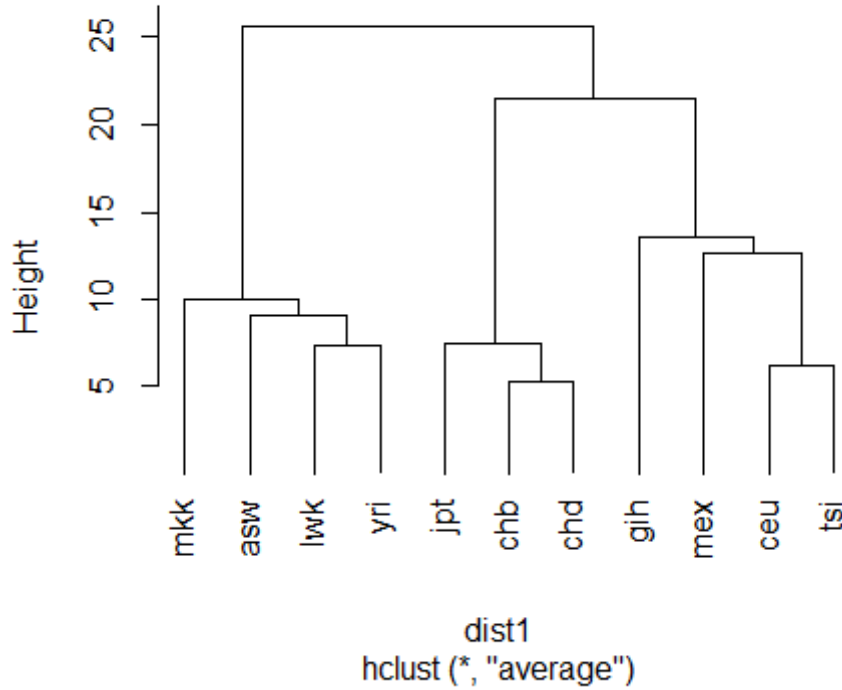
**EK50. X. Kromozomu HapMap Popülasyonları Hiyerarşik Kümeleme (Yüksek Çözünürlük)**

### KromozomX (HapMap-Yüksek Çözünürlük)

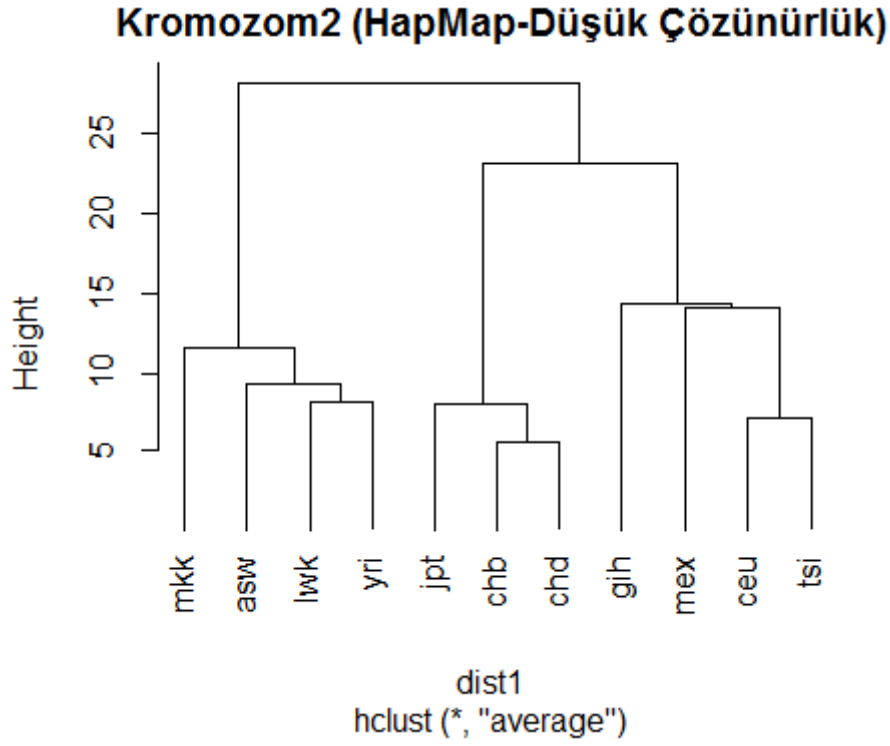


**EK51. 1. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Düşük Çözünürlük)**

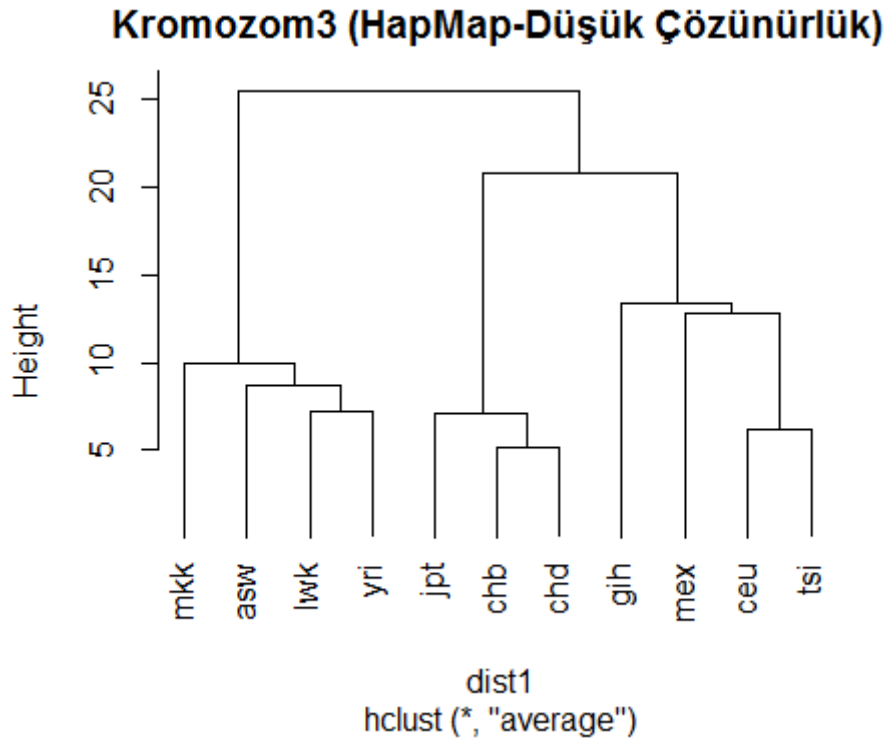
### Kromozom1 (HapMap-Düşük Çözünürlük)



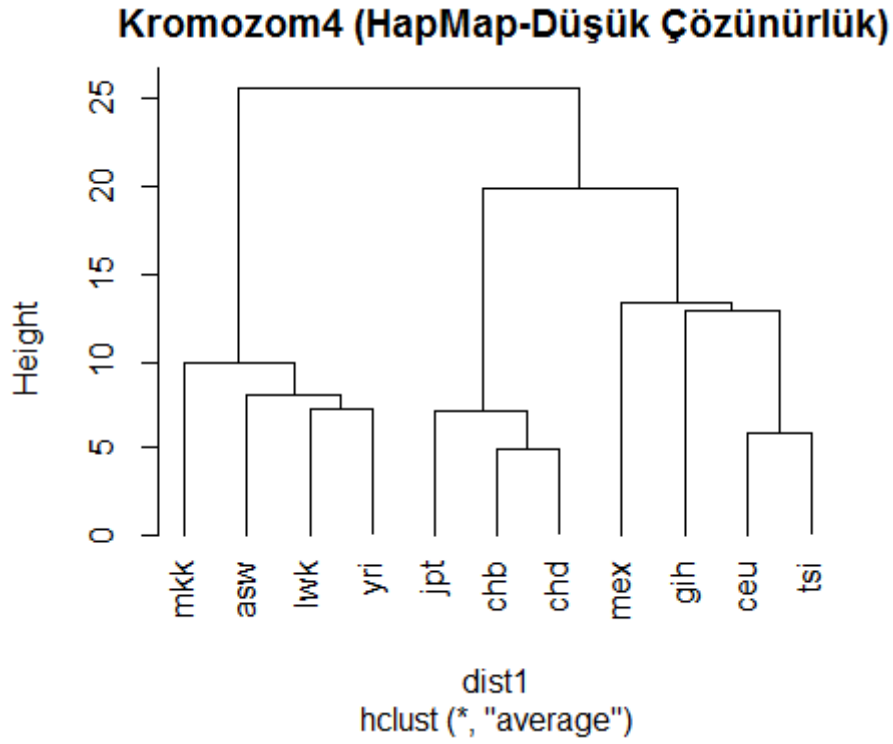
**EK52. 2. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Düşük Çözünürlük)**



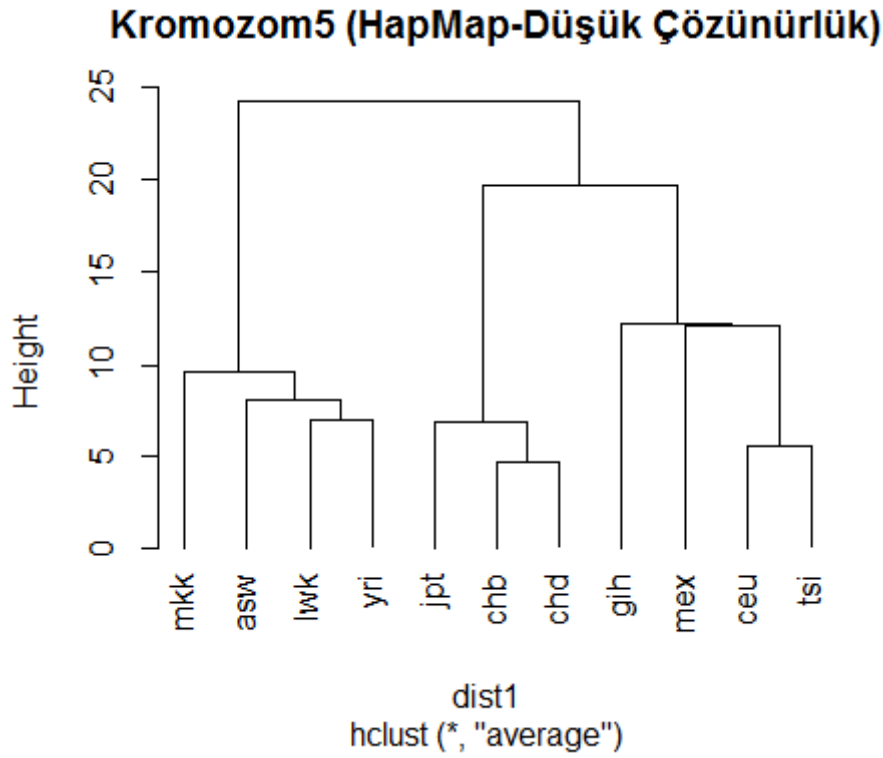
**EK53. 3. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Düşük Çözünürlük)**



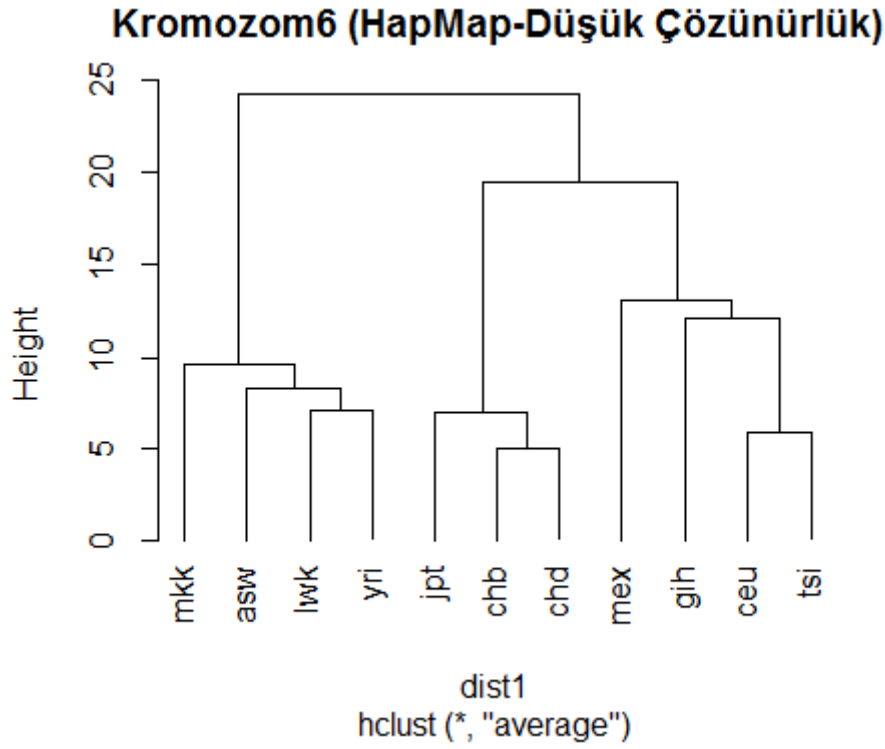
**EK54.** 4. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Düşük Çözünürlük)



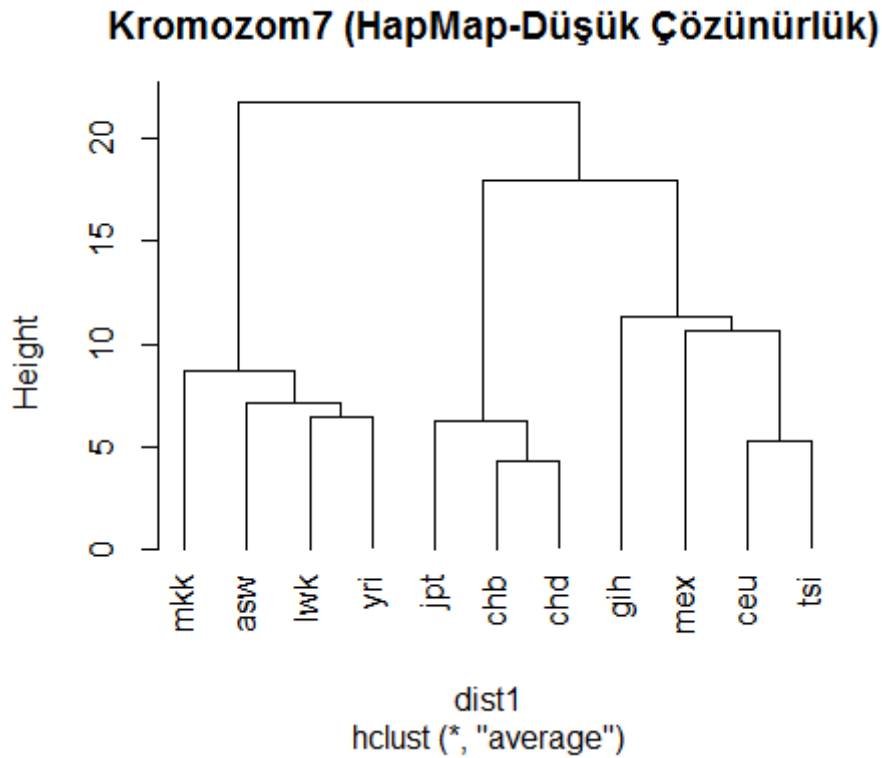
**EK55.** 5. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Düşük Çözünürlük)



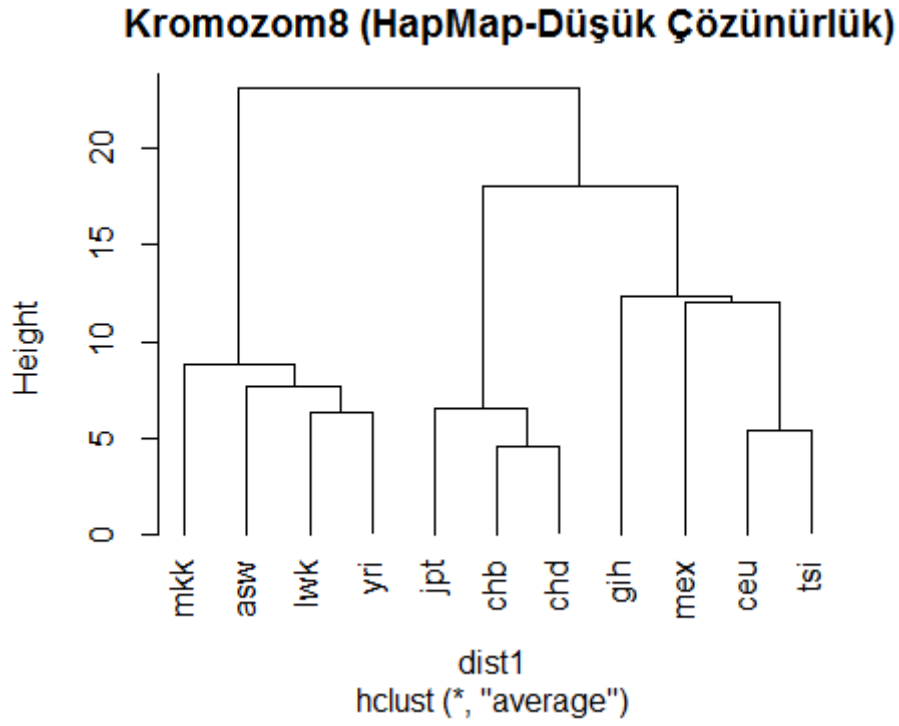
**EK56.** 6. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Düşük Çözünürlük)



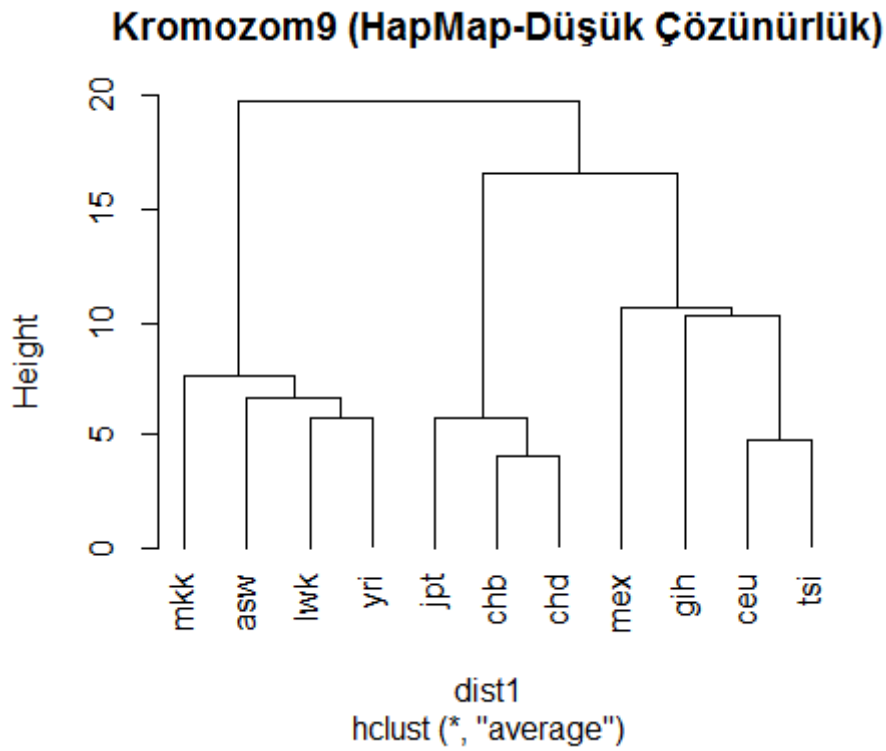
**EK57.** 7. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Düşük Çözünürlük)



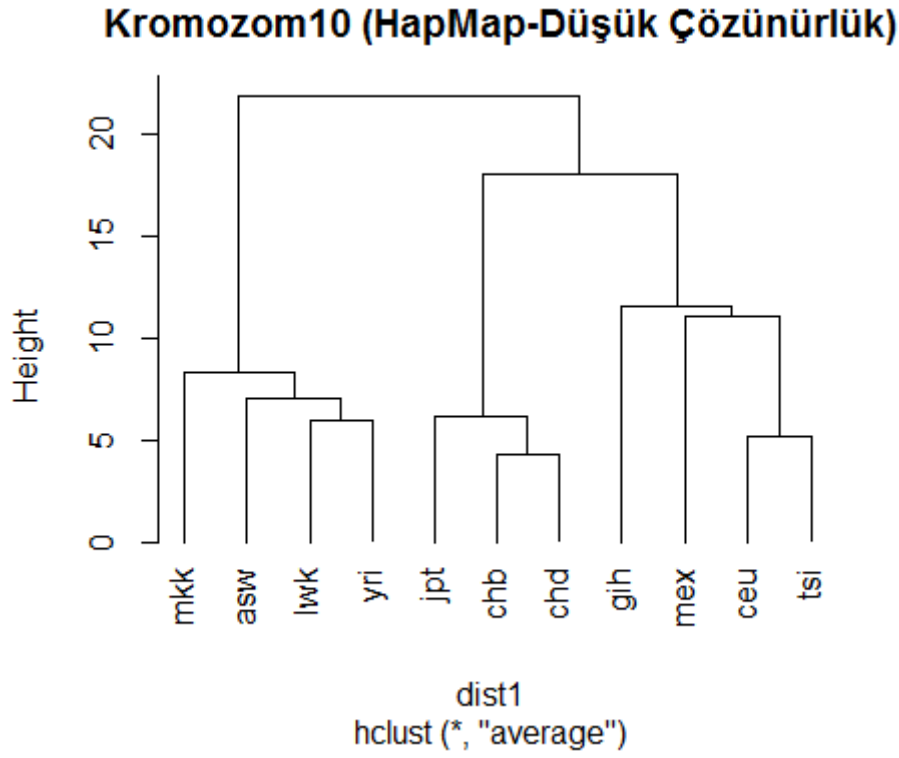
EK58. 8. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Düşük Çözünürlük)



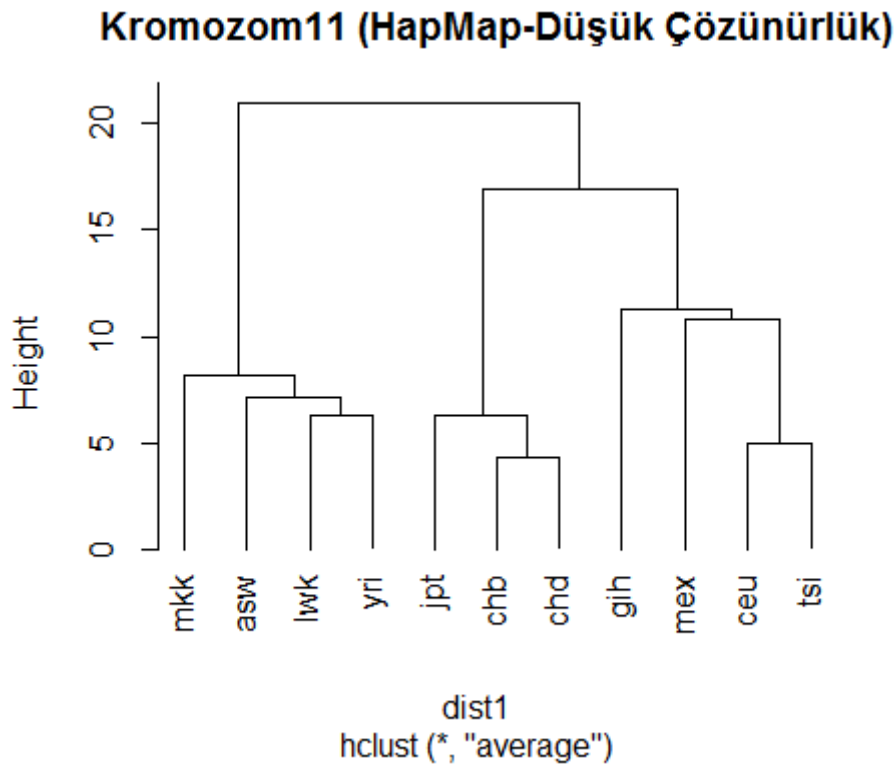
EK59. 9. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Düşük Çözünürlük)



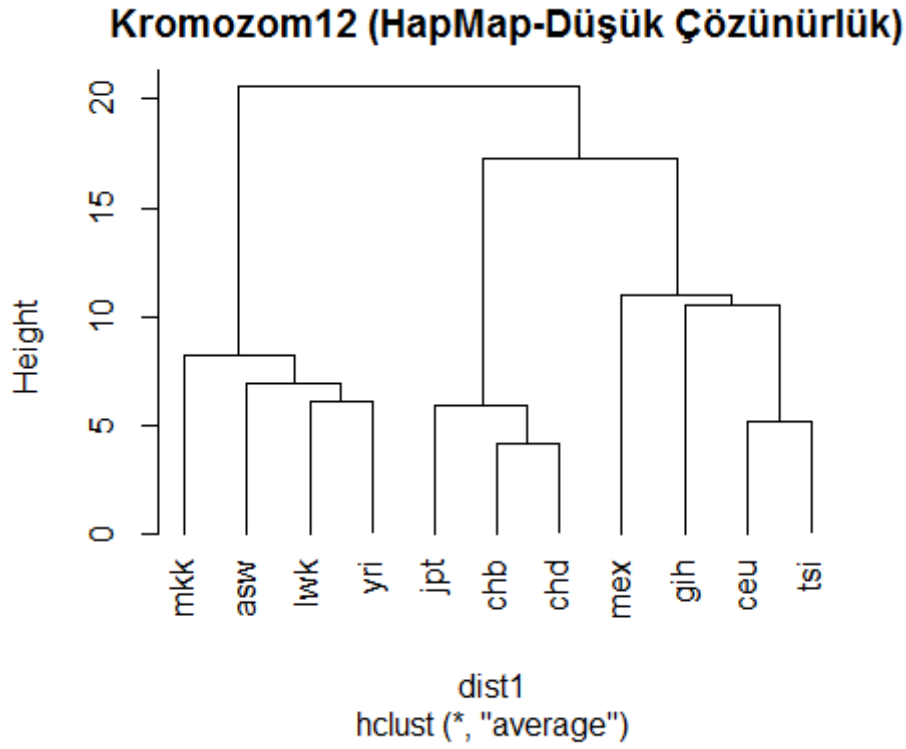
**EK60.** 10. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Düşük Çözünürlük)



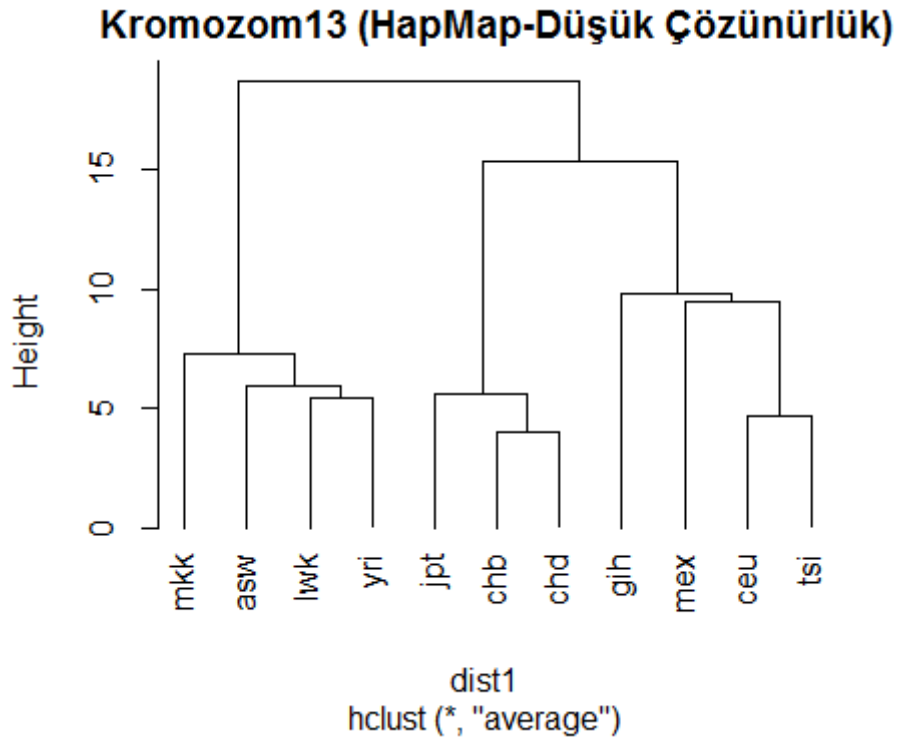
**EK61.** 11. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Düşük Çözünürlük)



**EK62.** 12. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Düşük Çözünürlük)

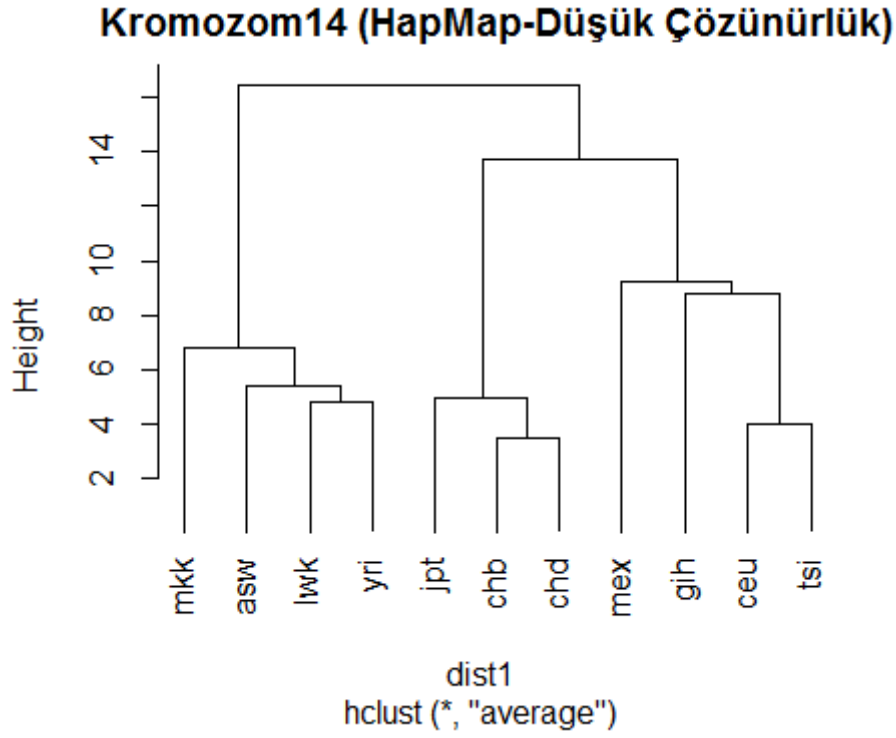


**EK63.** 13. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Düşük Çözünürlük)

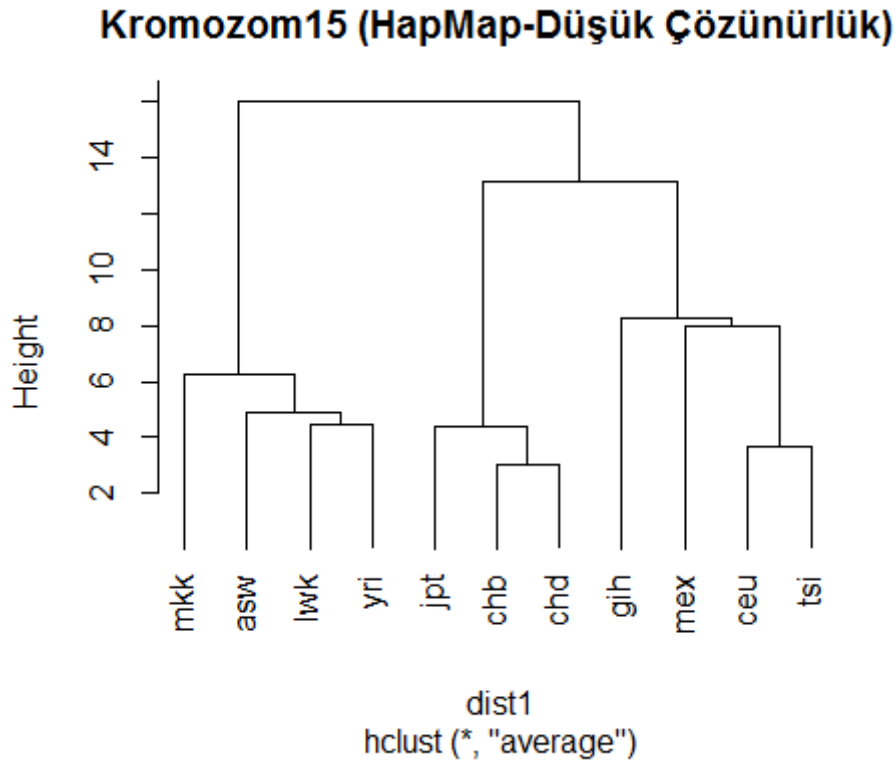




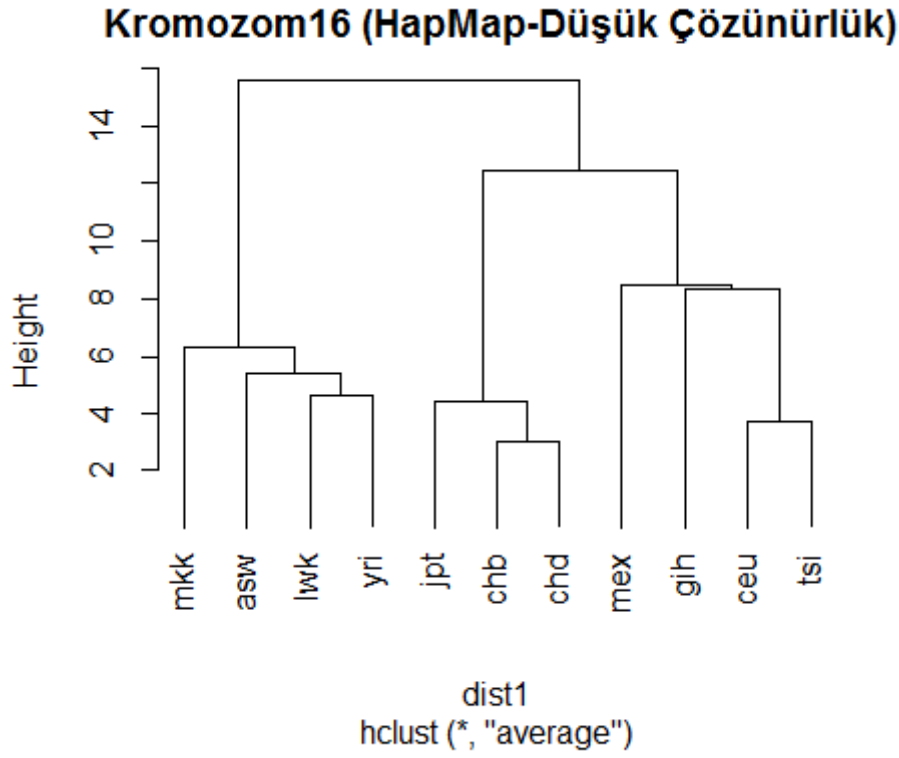
**EK64.** 14. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Düşük Çözünürlük)



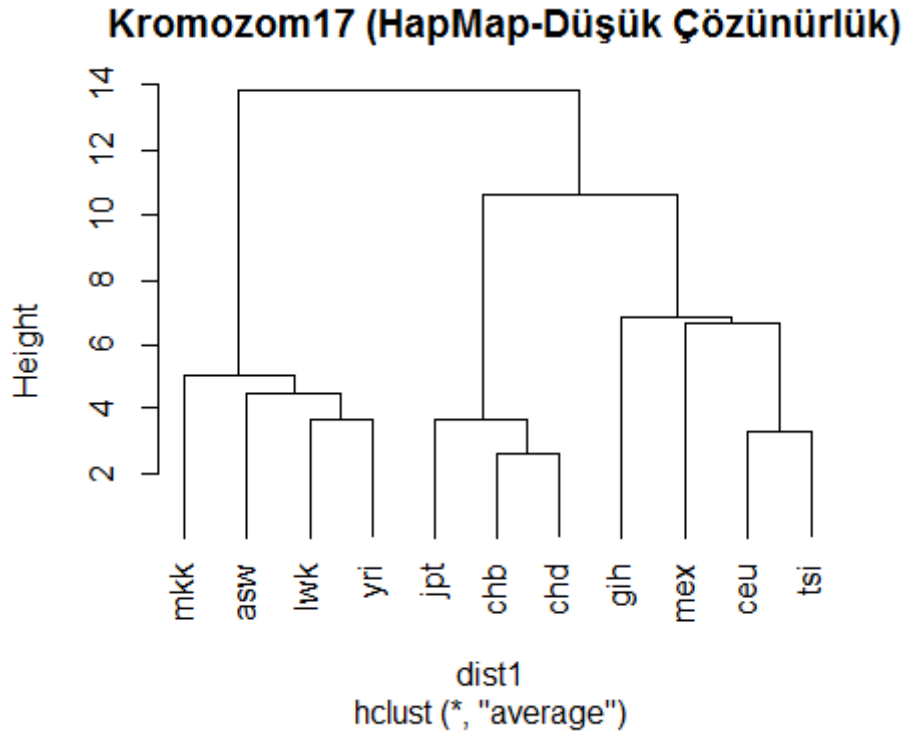
**EK65.** 15. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Düşük Çözünürlük)



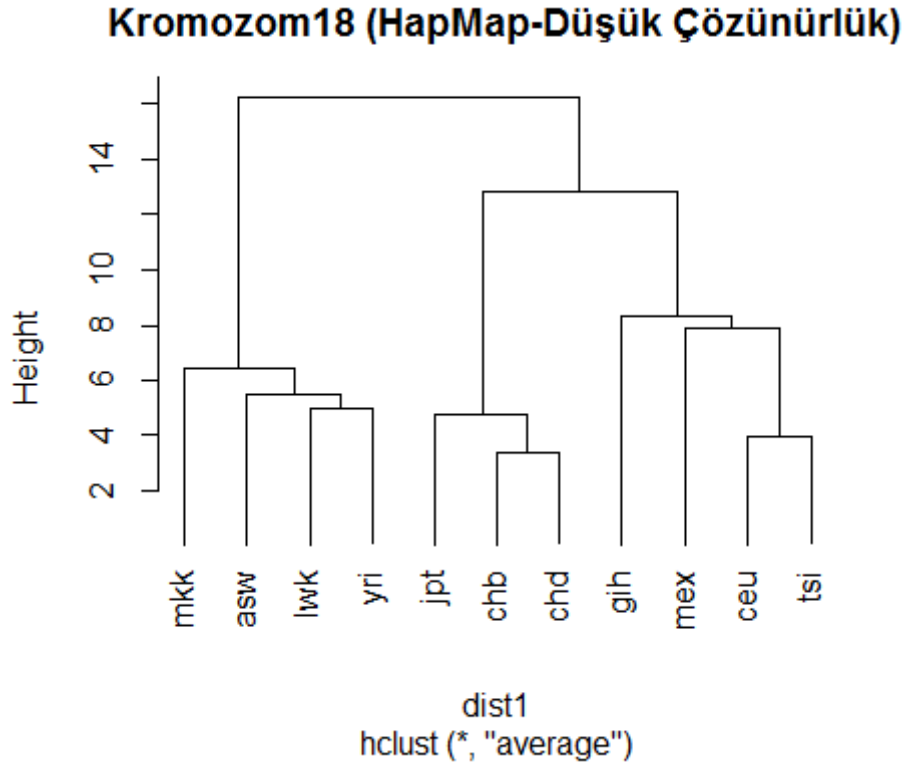
EK66. 16. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Düşük Çözünürlük)



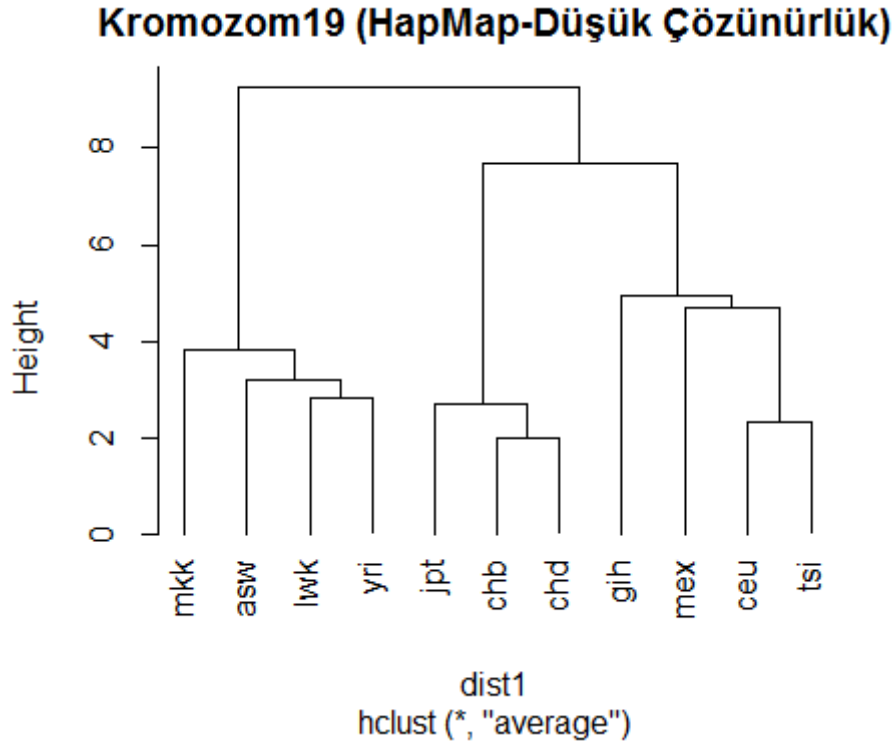
EK67. 17. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Düşük Çözünürlük)



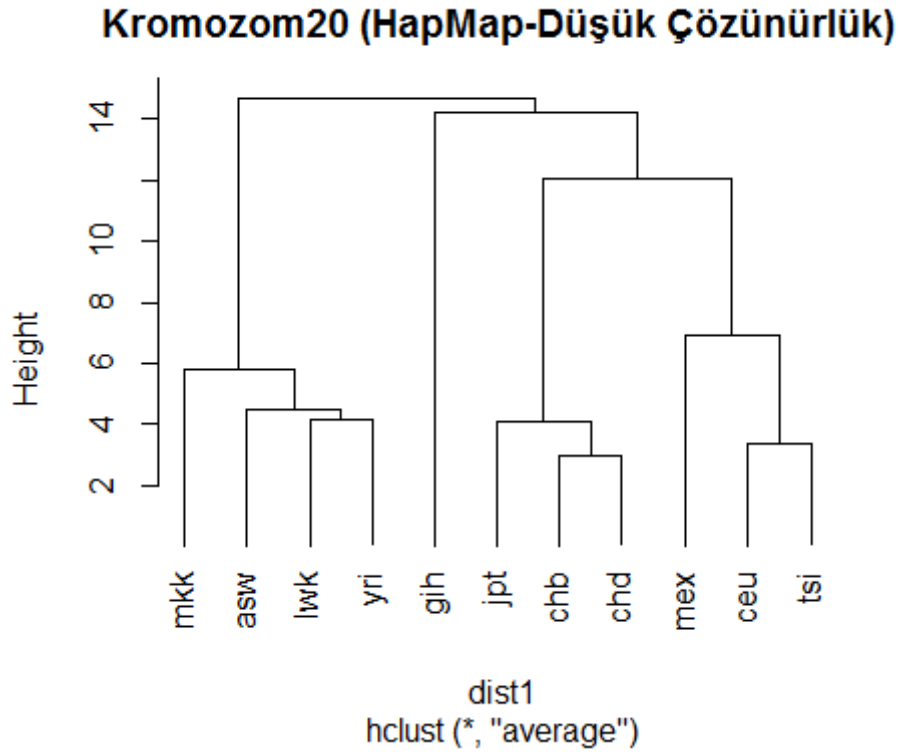
**EK68.** 18. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Düşük Çözünürlük)



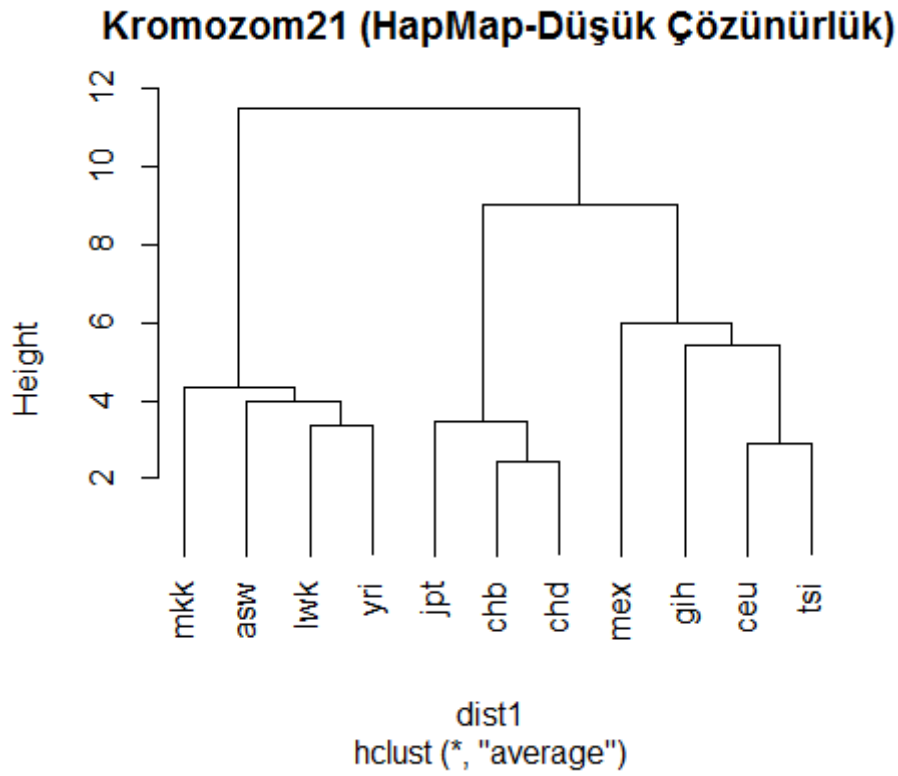
**EK69.** 19. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Düşük Çözünürlük)



EK70. 20. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Düşük Çözünürlük)

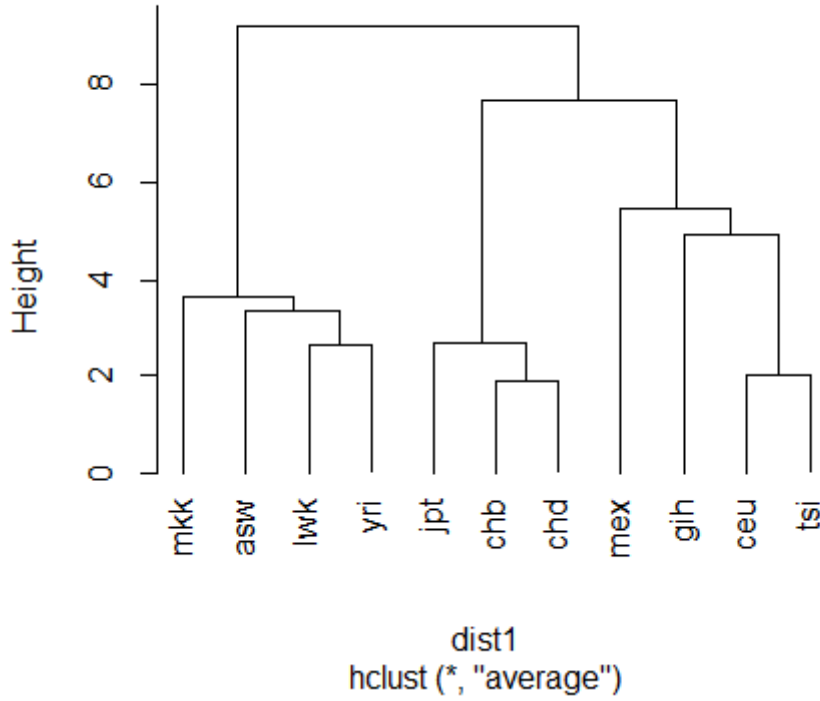


EK71. 21. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Düşük Çözünürlük)



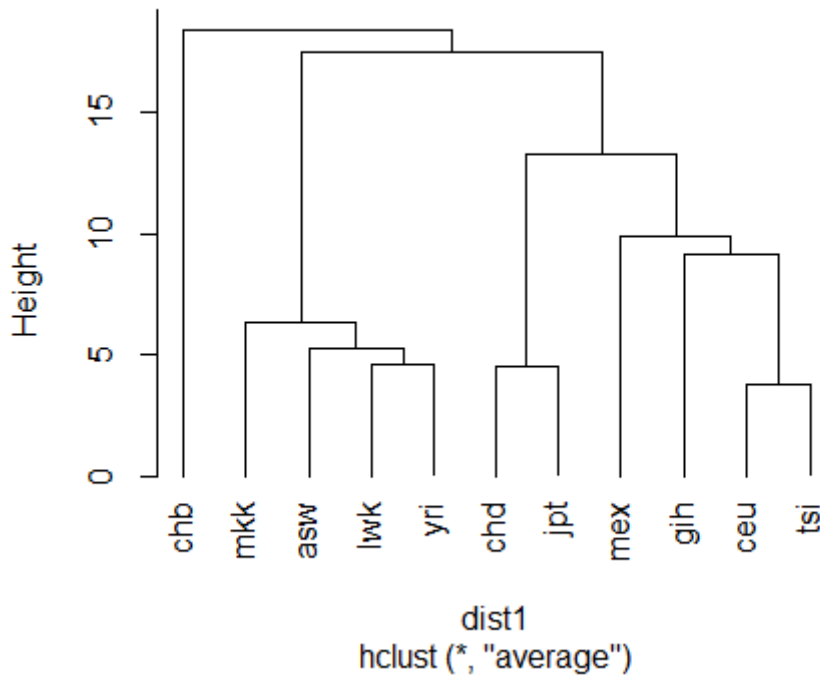
EK72. 22. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Düşük Çözünürlük)

### Kromozom22 (HapMap-Düşük Çözünürlük)

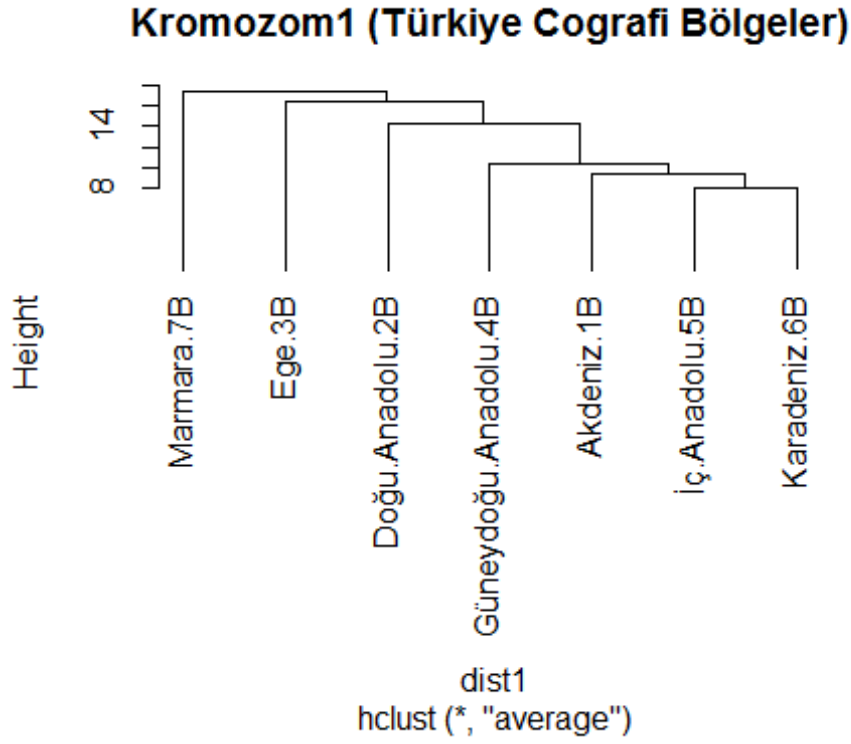


EK73. X. Kromozom HapMap Popülasyonları Hiyerarşik Kümeleme (Düşük Çözünürlük)

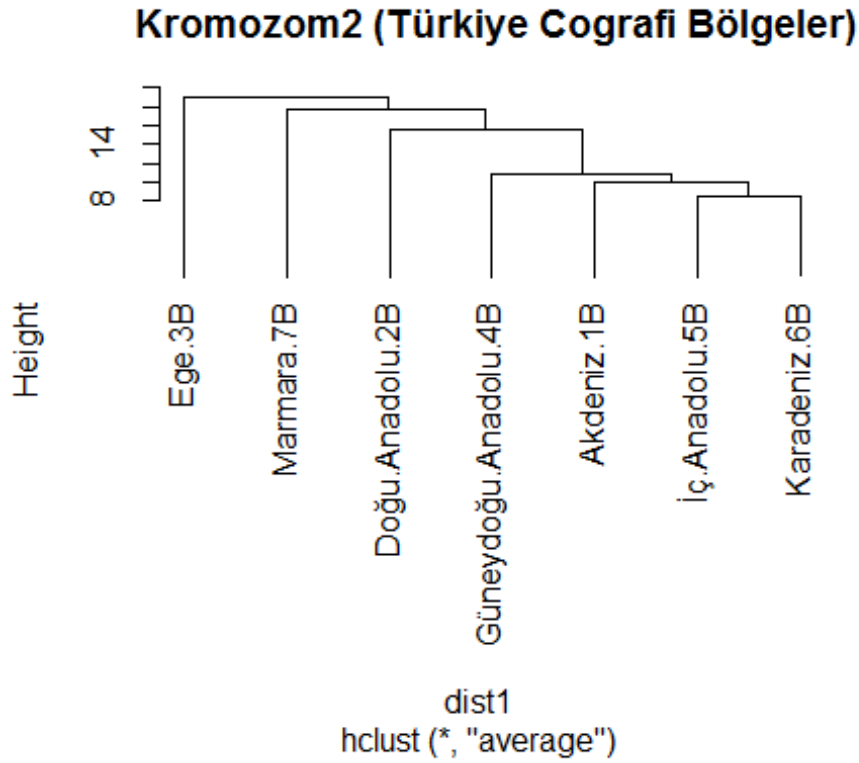
### KromozomX (HapMap-Düşük Çözünürlük)



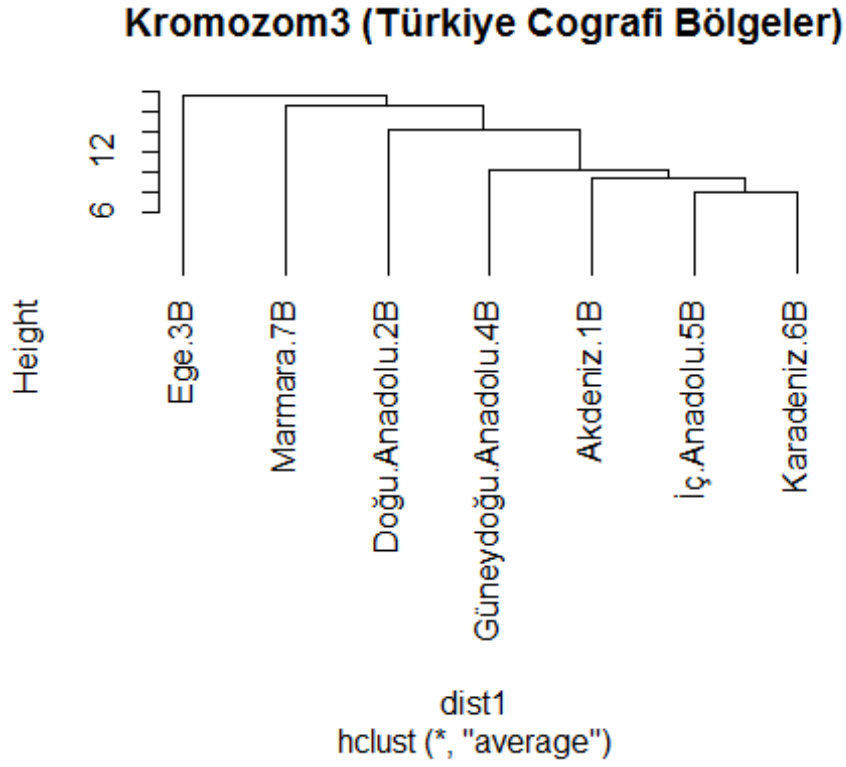
EK74. 1. Kromozom Türkiye (7 Bölge) Hiyerarşik Kümeleme



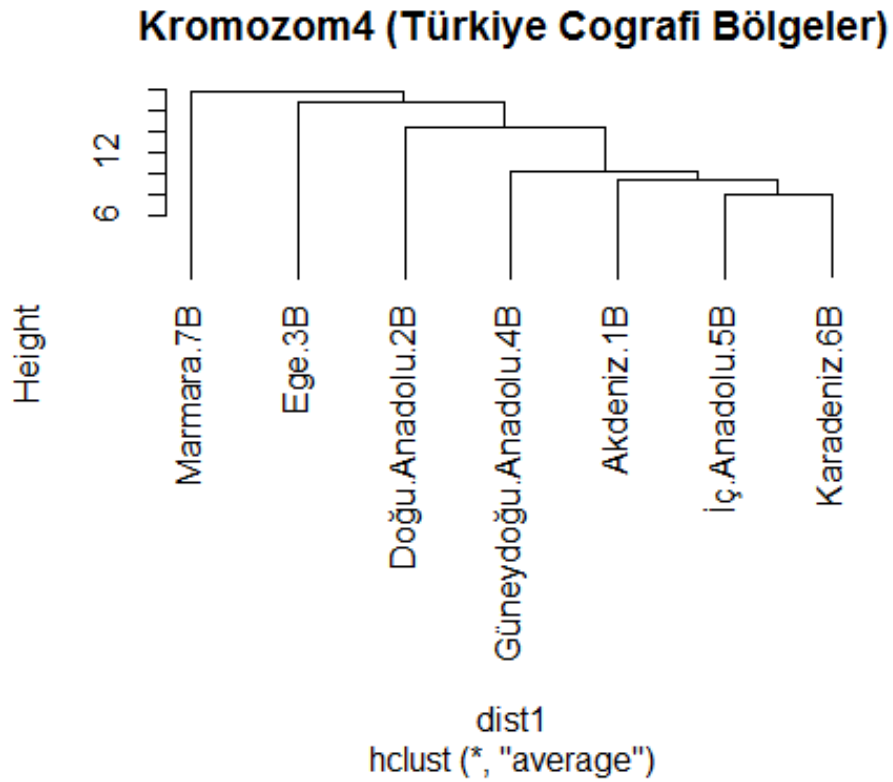
EK75. 2. Kromozom Türkiye (7 Bölge) Hiyerarşik Kümeleme



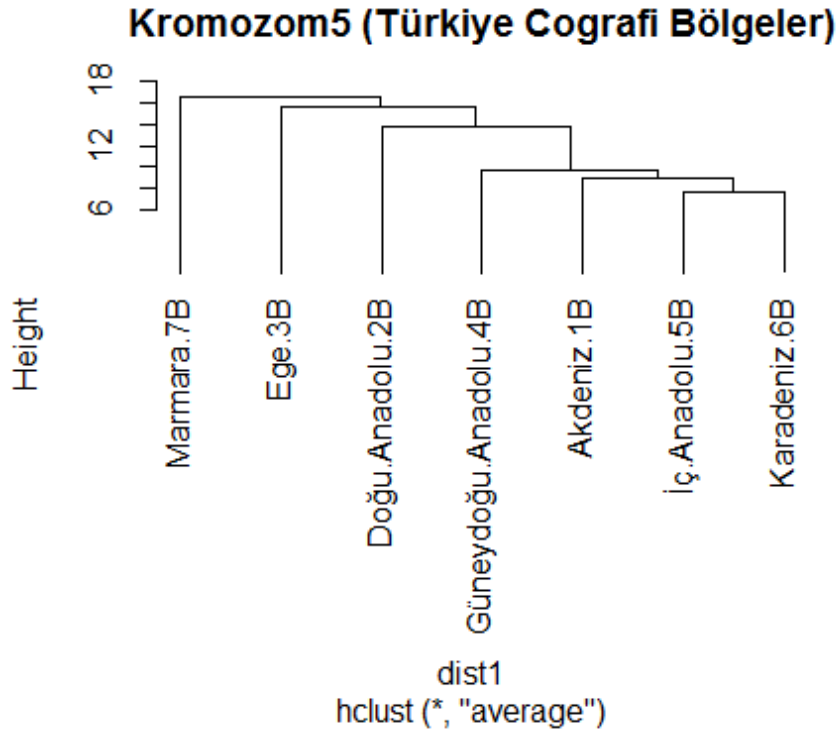
EK76. 3. Kromozom Türkiye (7 Bölge) Hiyerarşik Kümeleme



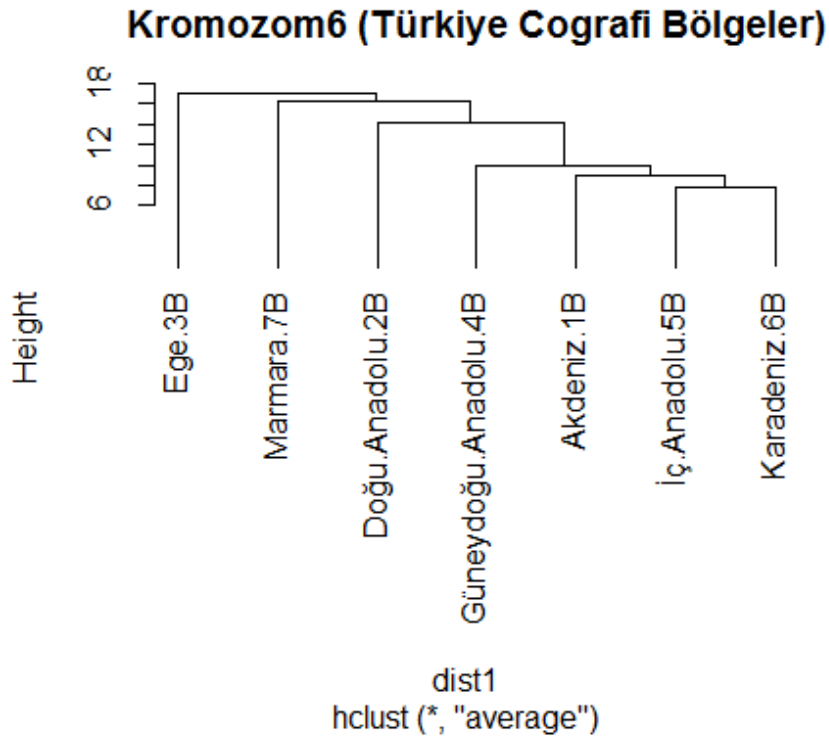
EK77. 4. Kromozom Türkiye (7 Bölge) Hiyerarşik Kümeleme



**EK78.** 5. Kromozom Türkiye (7 Bölge) Hiyerarşik Kümeleme

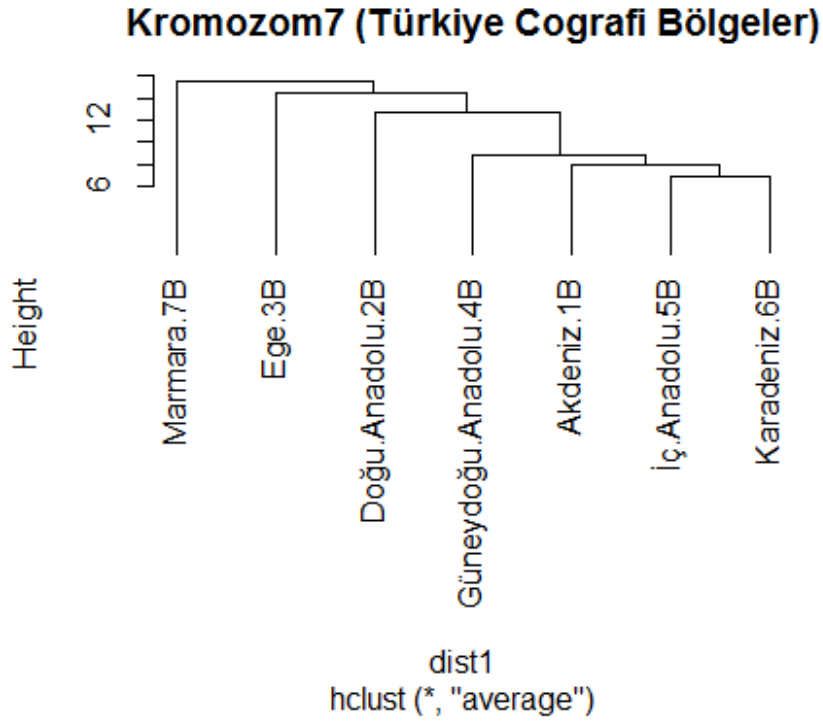


**EK79.** 6. Kromozom Türkiye (7 Bölge) Hiyerarşik Kümeleme

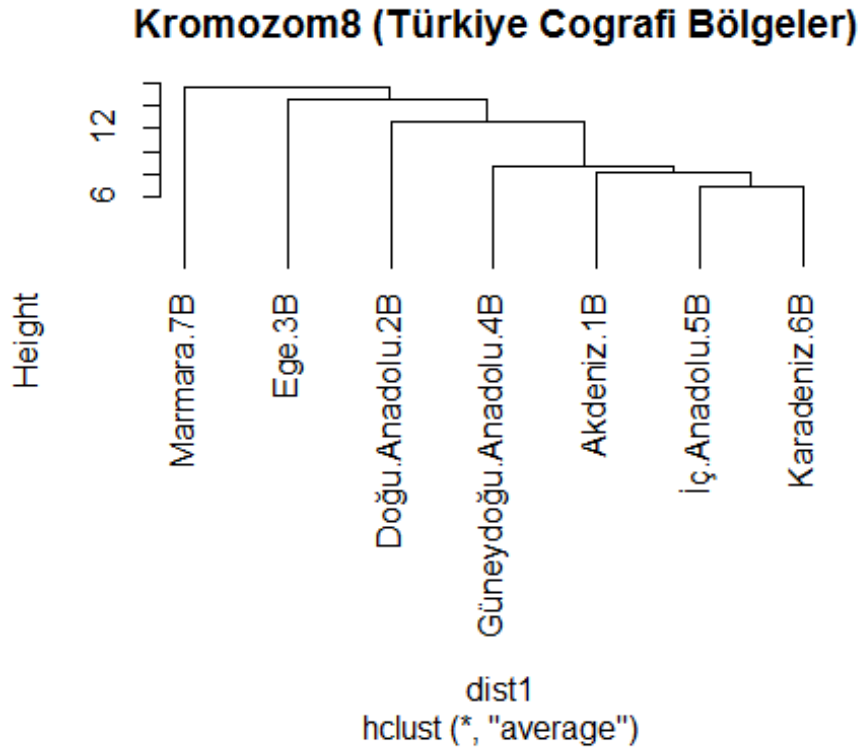




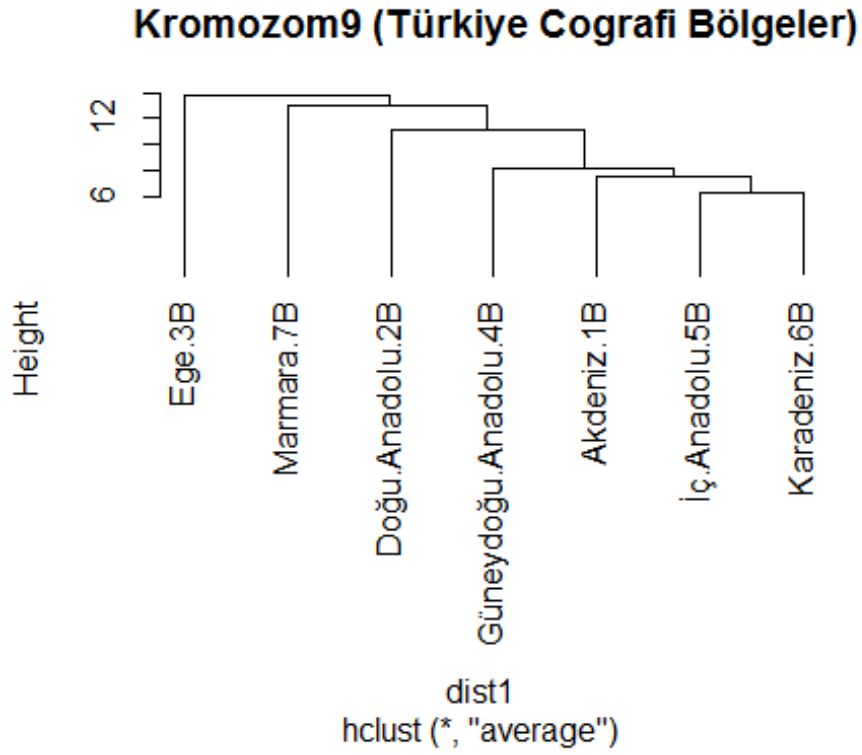
**EK80.** 7. Kromozom Türkiye (7 Bölge) Hiyerarşik Kümeleme



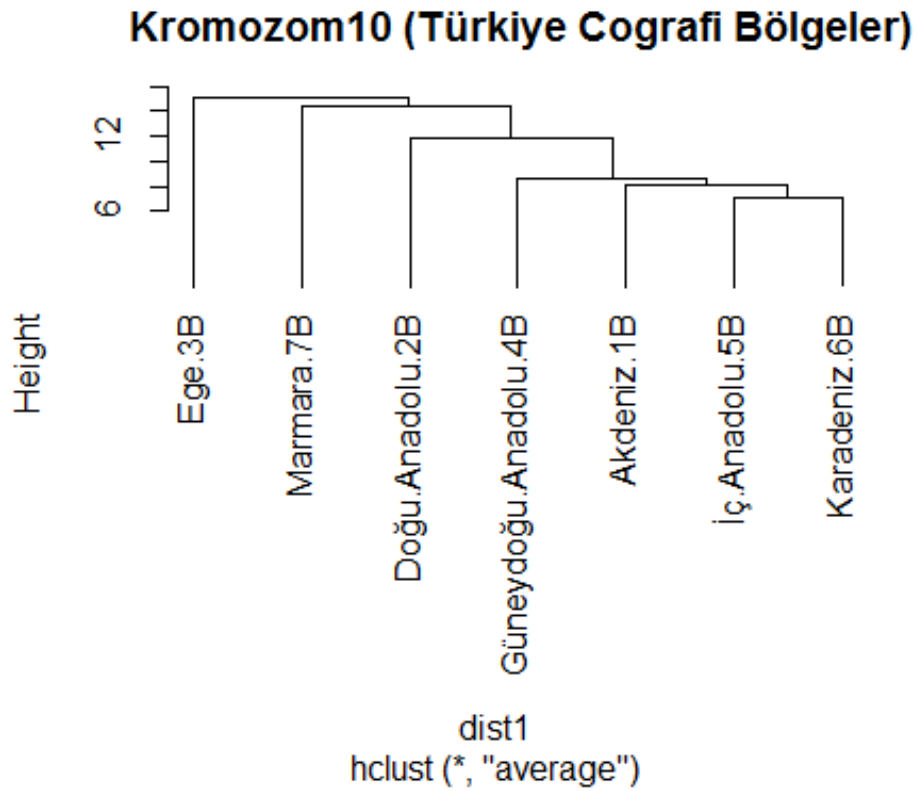
**EK81.** 8. Kromozom Türkiye (7 Bölge) Hiyerarşik Kümeleme



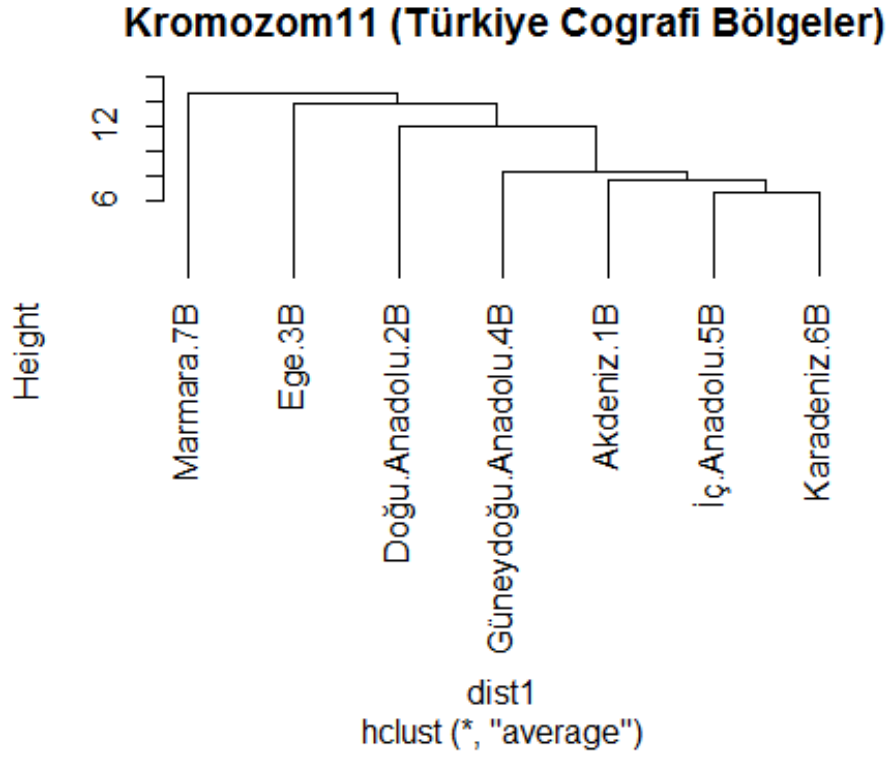
**EK82.** 9. Kromozom Türkiye (7 Bölge) Hiyerarşik Kümeleme



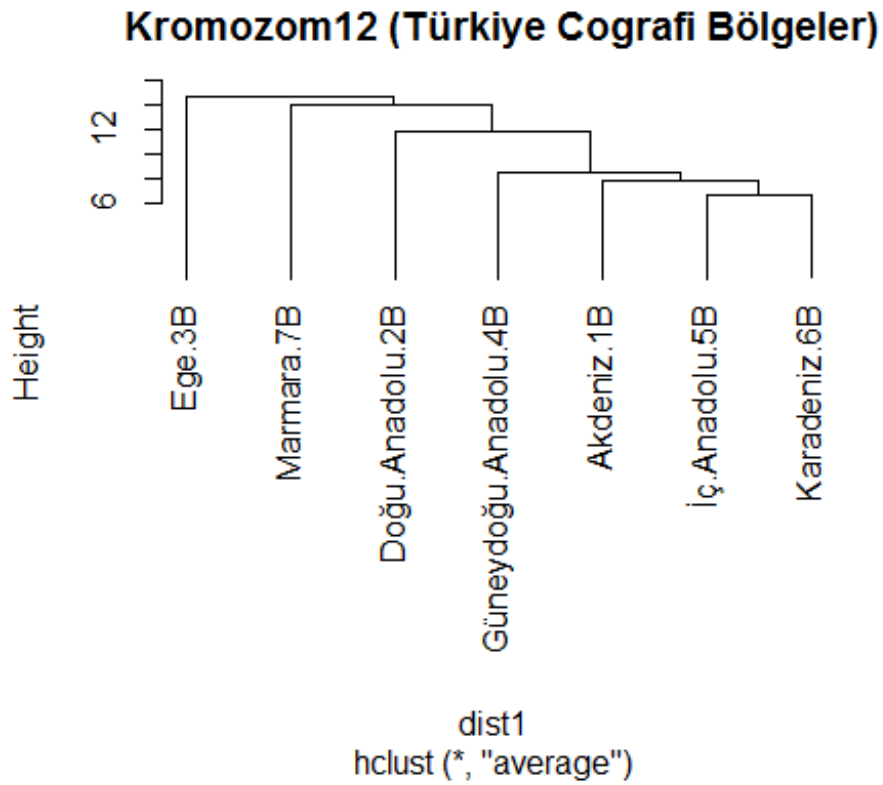
**EK83.** 10. Kromozom Türkiye (7 Bölge) Hiyerarşik Kümeleme



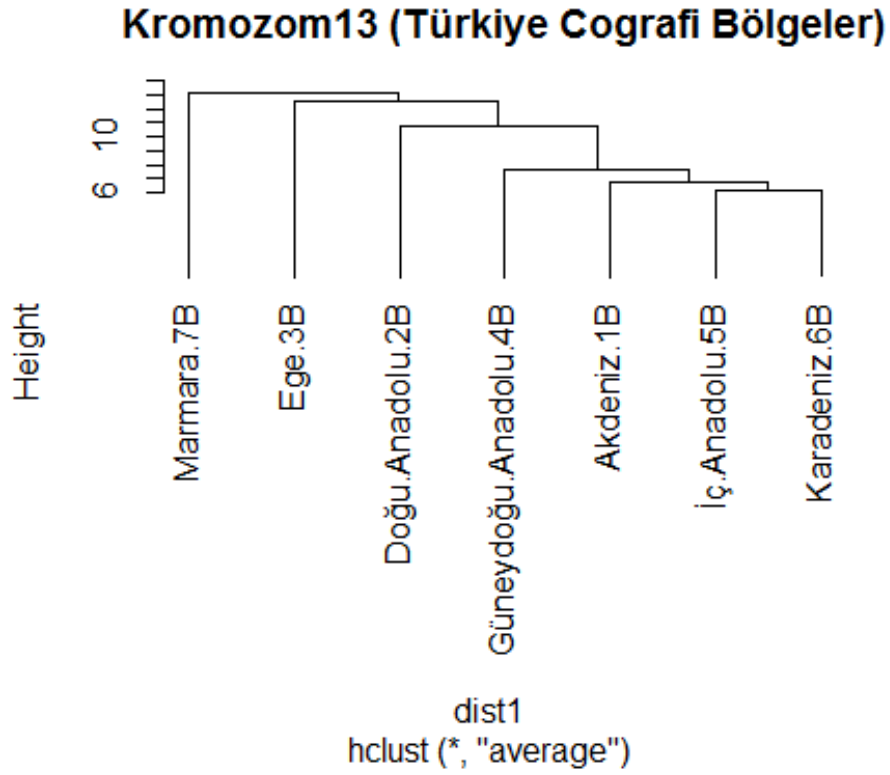
**EK84.** 11. Kromozom Türkiye (7 Bölge) Hiyerarşik Kümeleme



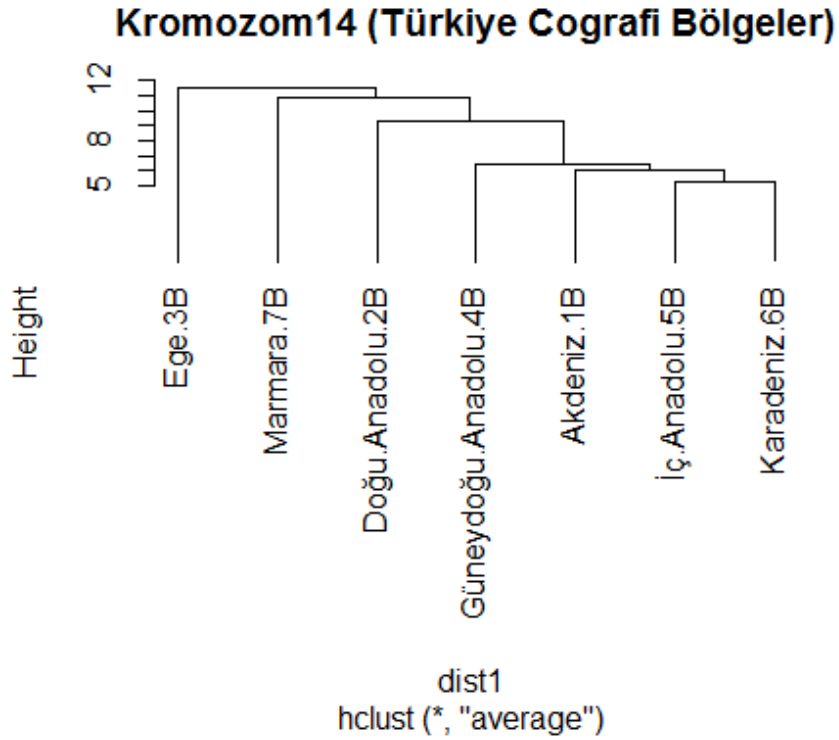
**EK85.** 12. Kromozom Türkiye (7 Bölge) Hiyerarşik Kümeleme



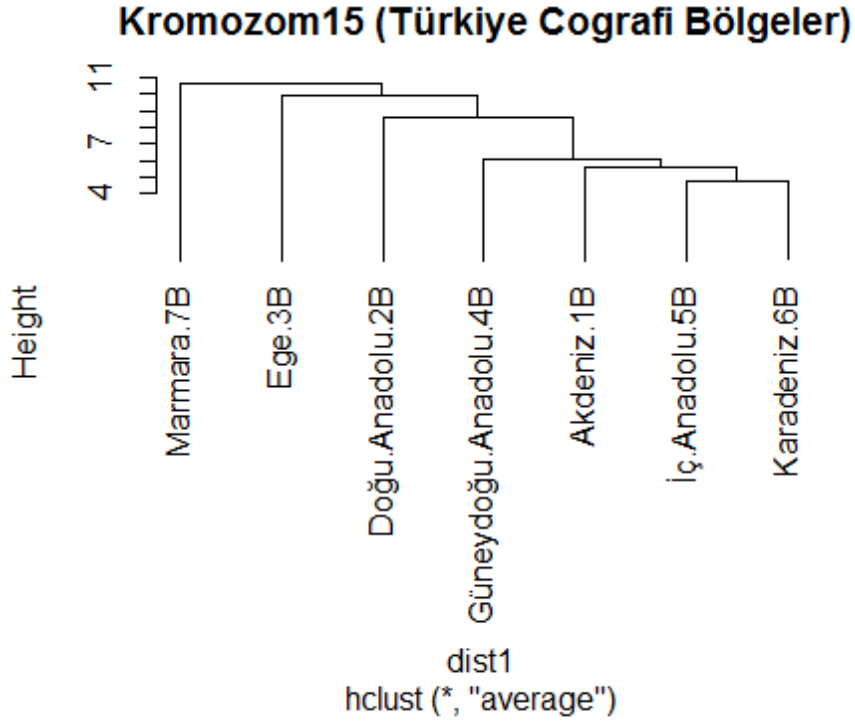
EK86. 13. Kromozom Türkiye (7 Bölge) Hiyerarşik Kümeleme



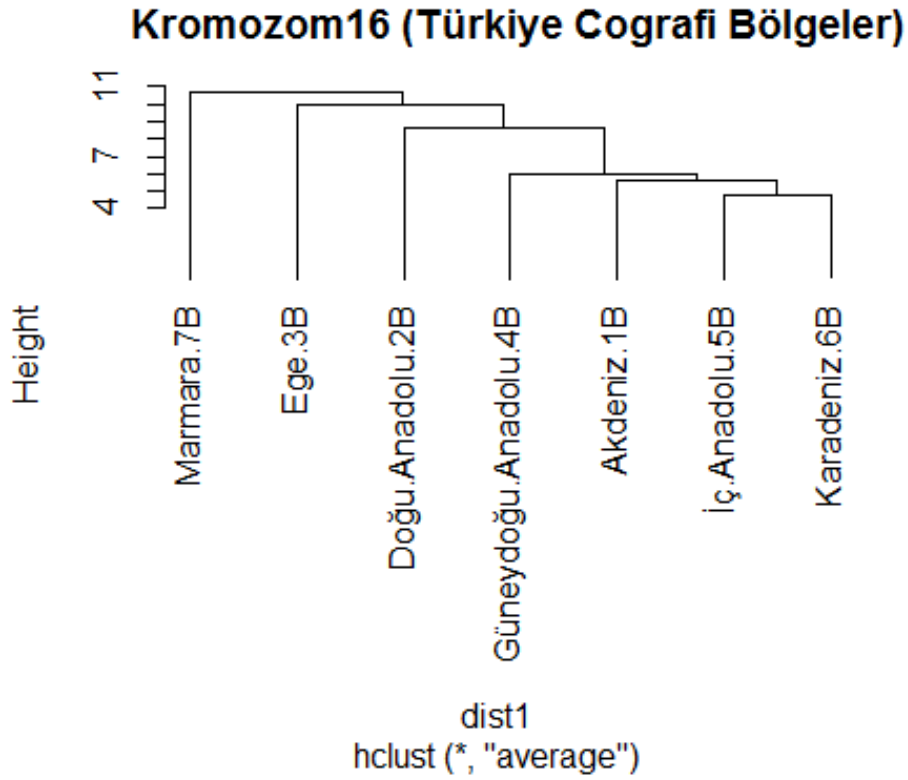
EK87. 14. Kromozom Türkiye (7 Bölge) Hiyerarşik Kümeleme



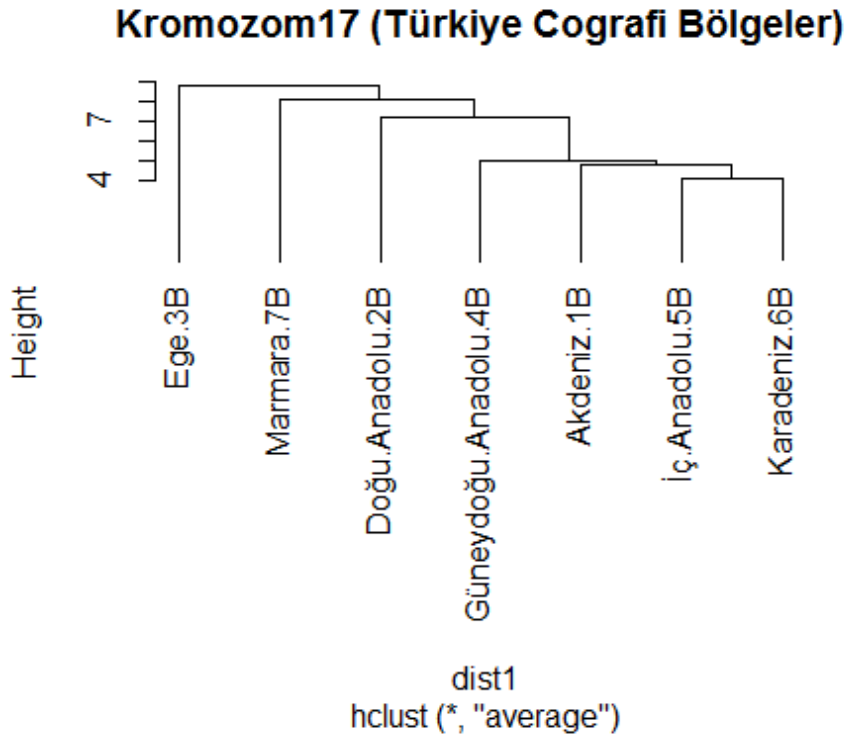
**EK88.** 15. Kromozom Türkiye (7 Bölge) Hiyerarşik Kümeleme



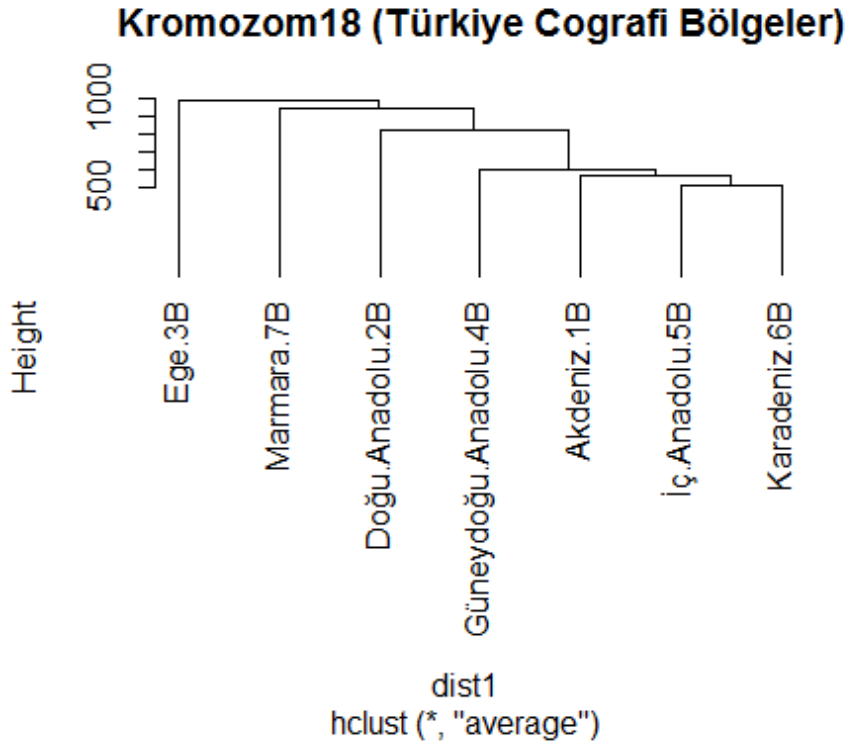
**EK89.** 16. Kromozom Türkiye (7 Bölge) Hiyerarşik Kümeleme



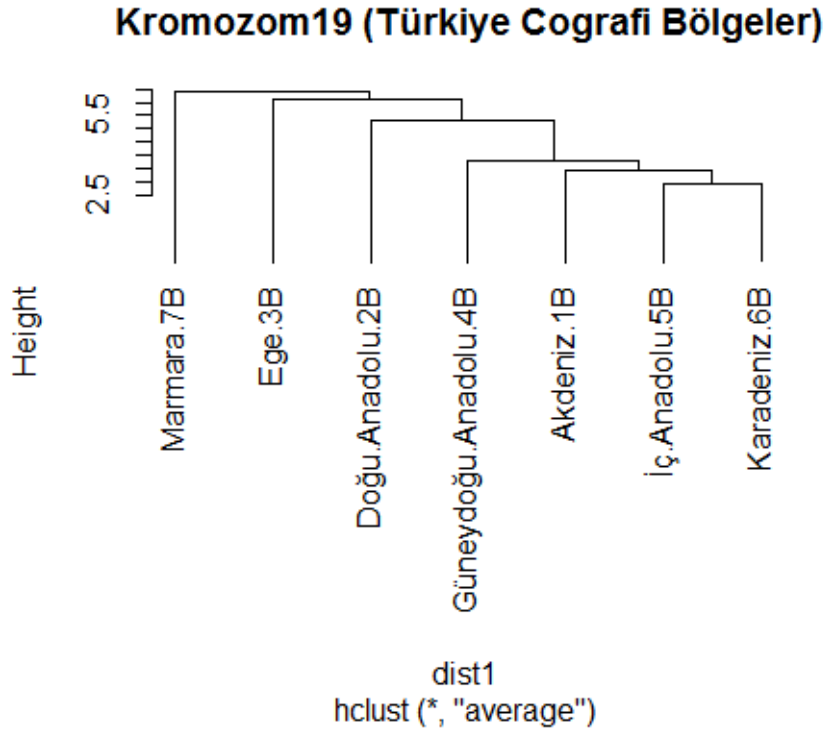
**EK90.** 17. Kromozom Türkiye (7 Bölge) Hiyerarşik Kümeleme



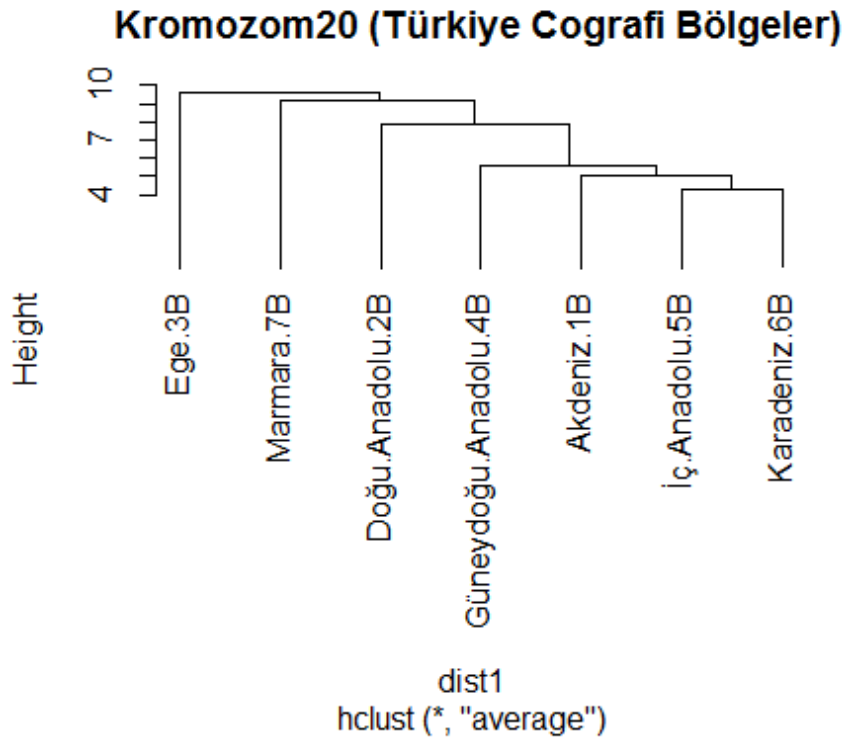
**EK91.** 18. Kromozom Türkiye (7 Bölge) Hiyerarşik Kümeleme



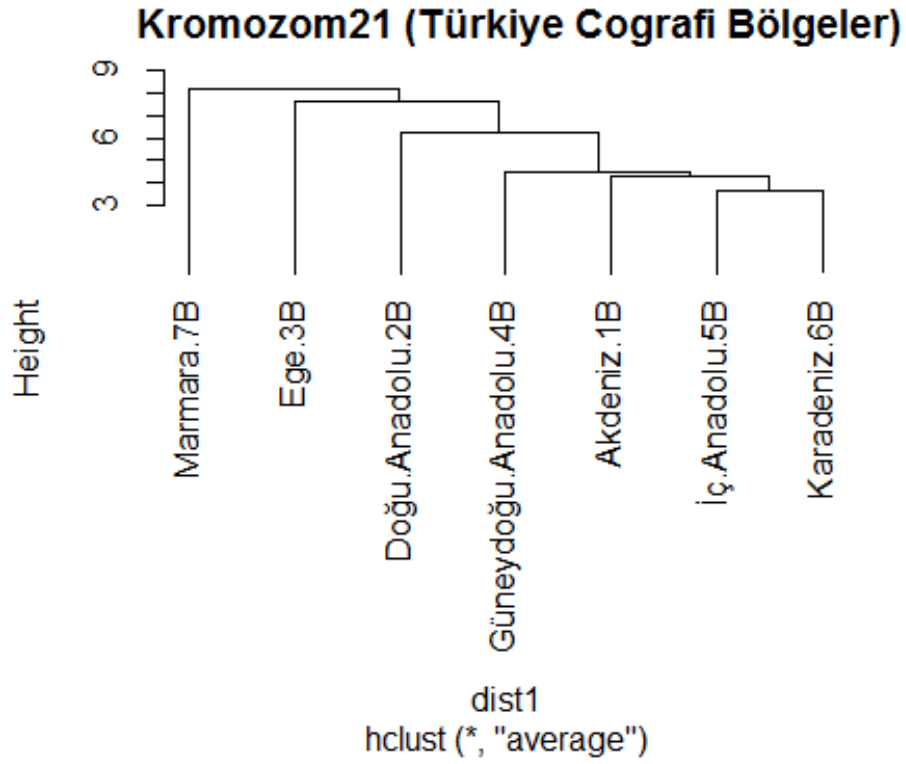
**EK92.** 19. Kromozom Türkiye (7 Bölge) Hiyerarşik Kümeleme



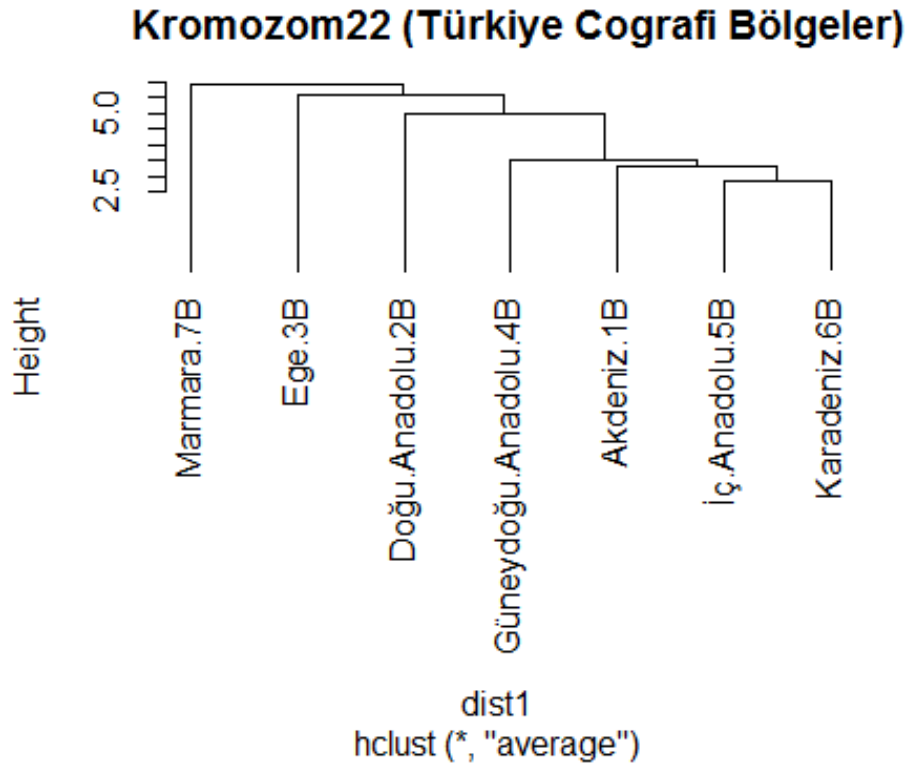
**EK93.** 20. Kromozom Türkiye (7 Bölge) Hiyerarşik Kümeleme



EK94. 21. Kromozom Türkiye (7 Bölge) Hiyerarşik Kümeleme

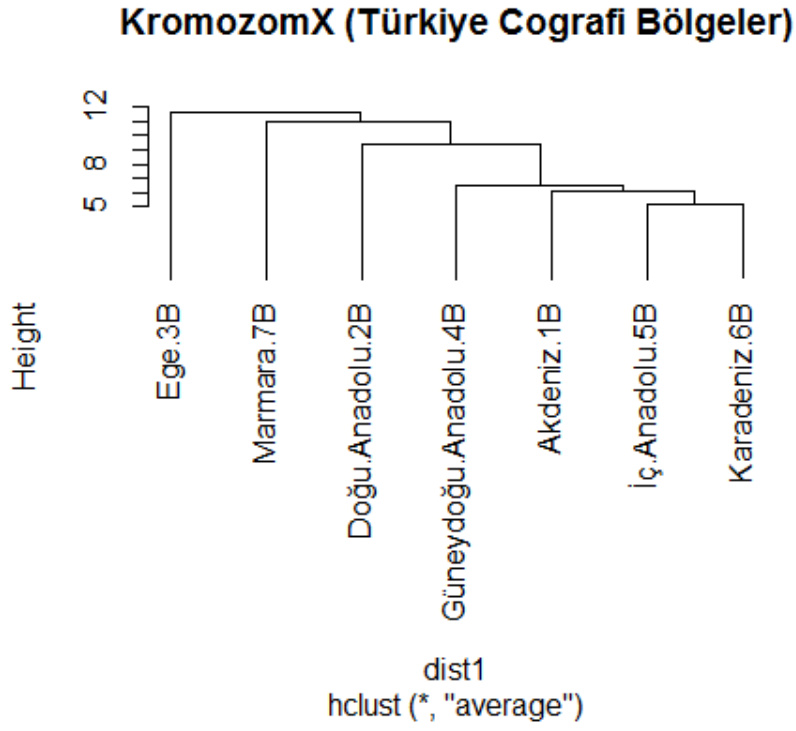


EK95. 22. Kromozom Türkiye (7 Bölge) Hiyerarşik Kümeleme

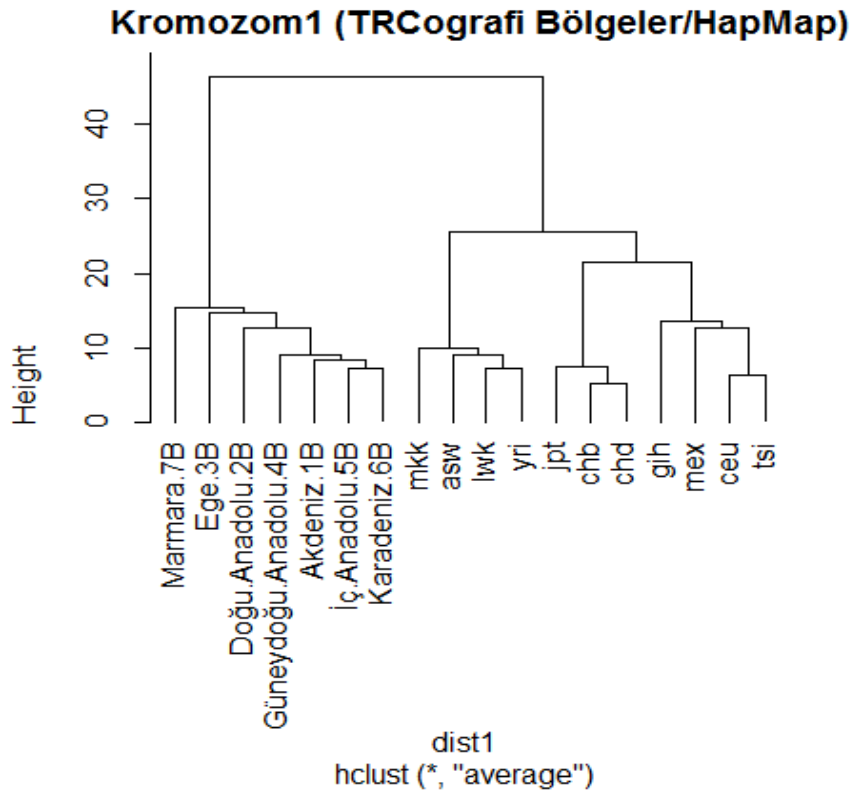




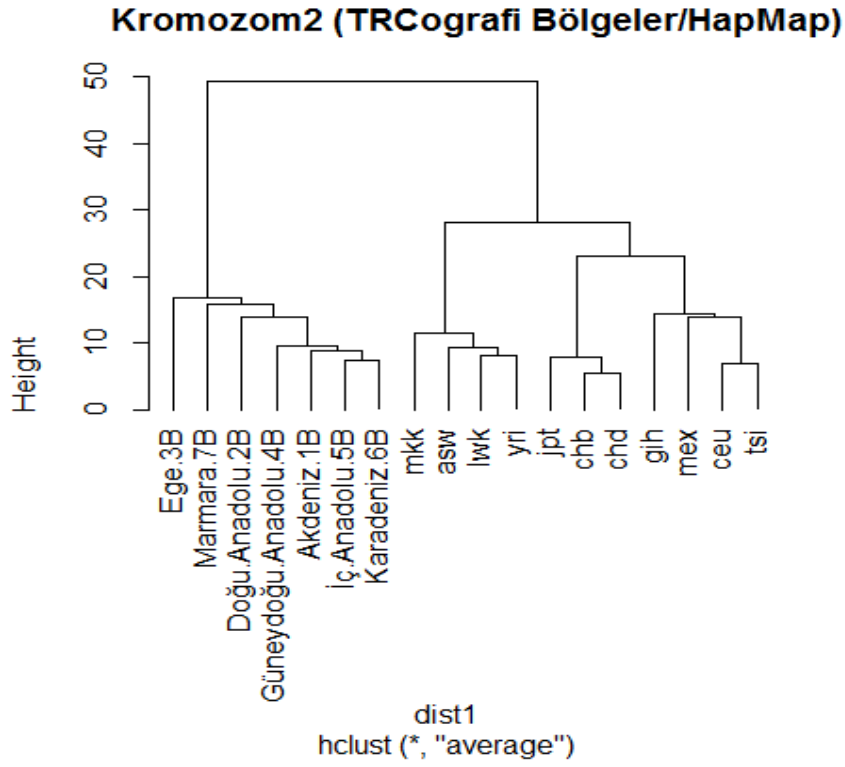
**EK96. X. Kromozomu Türkiye (7 Bölge) Hiyerarşik Kümeleme**



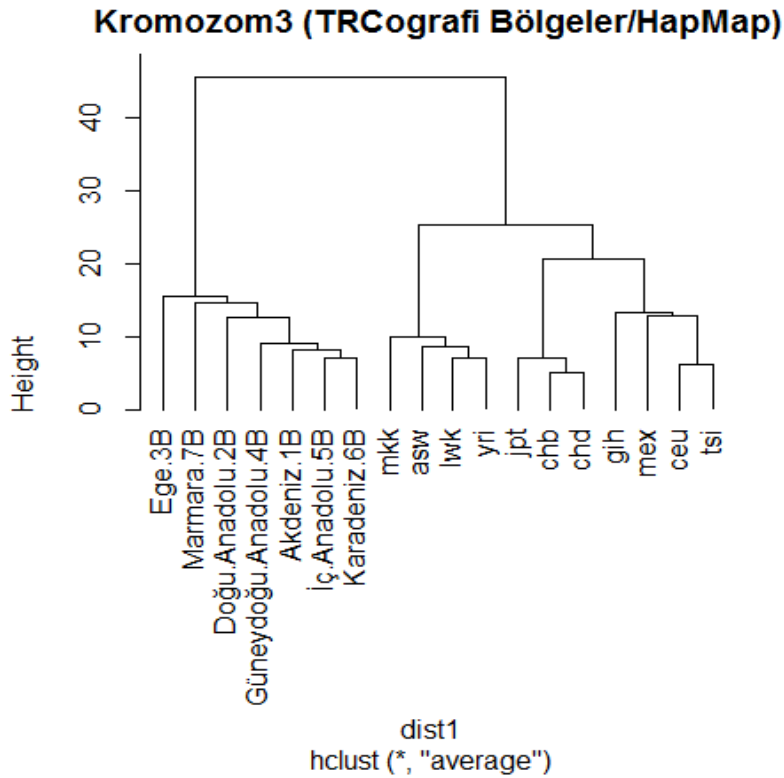
**EK97. 1. Kromozom TR Bölgeleri ve HapMap Popülasyonları Hiyerarşik Kümeleme**



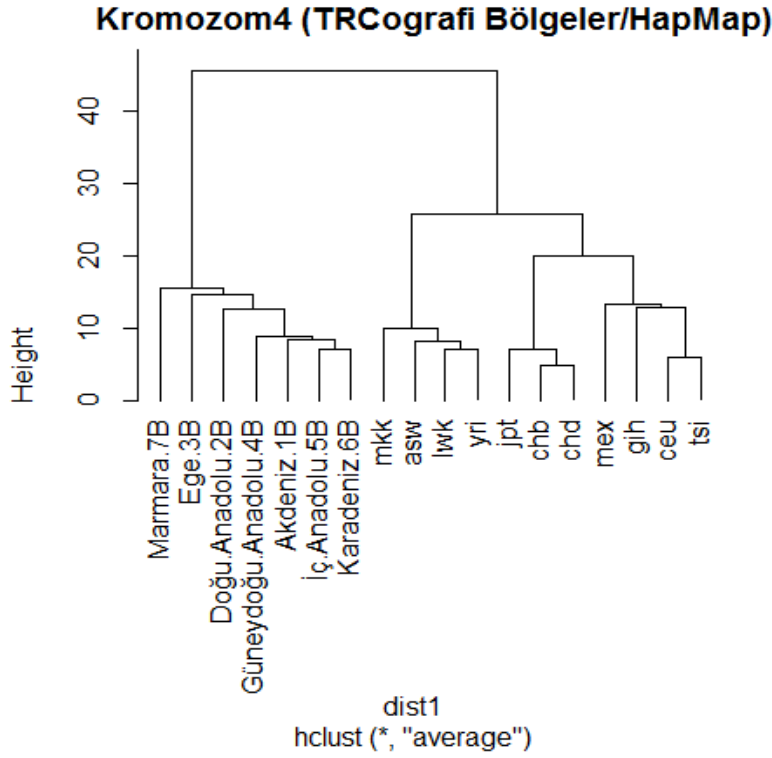
**EK98. 2. Kromozom TR Bölgeleri ve HapMap Popülasyonları Hiyerarşik Kümeleme**



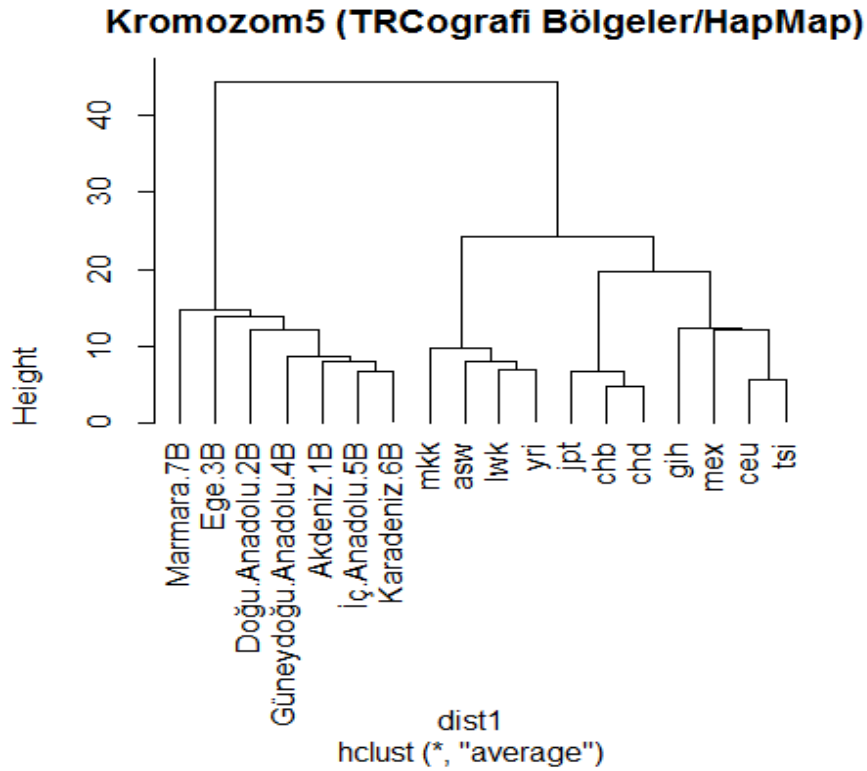
**EK99. 3. Kromozom TR Bölgeleri ve HapMap Popülasyonları Hiyerarşik Kümeleme**



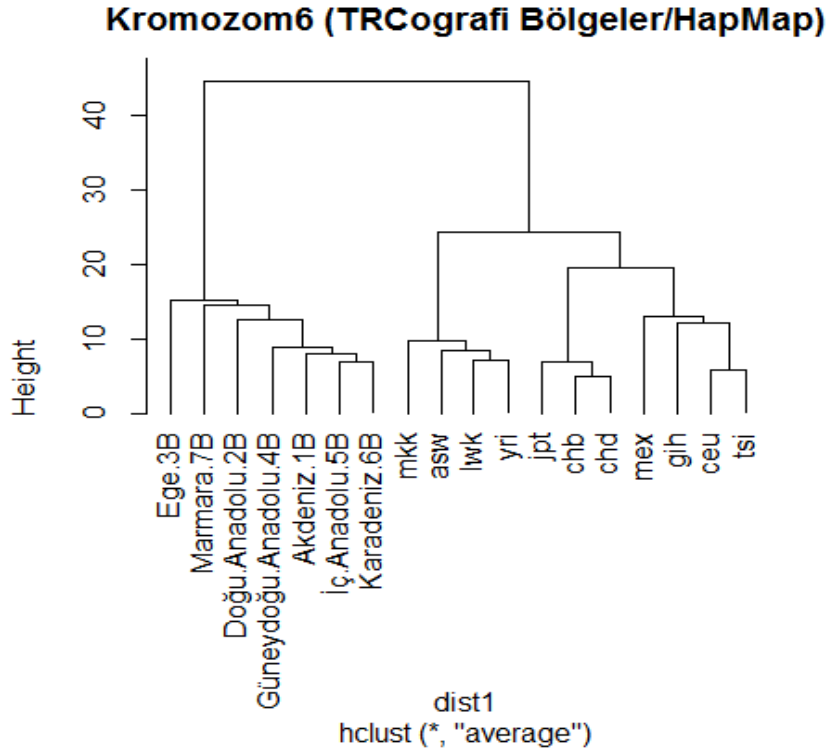
**EK100.** 4. Kromozom TR Bölgeleri ve HapMap Popülasyonları Hiyerarşik Kümeleme



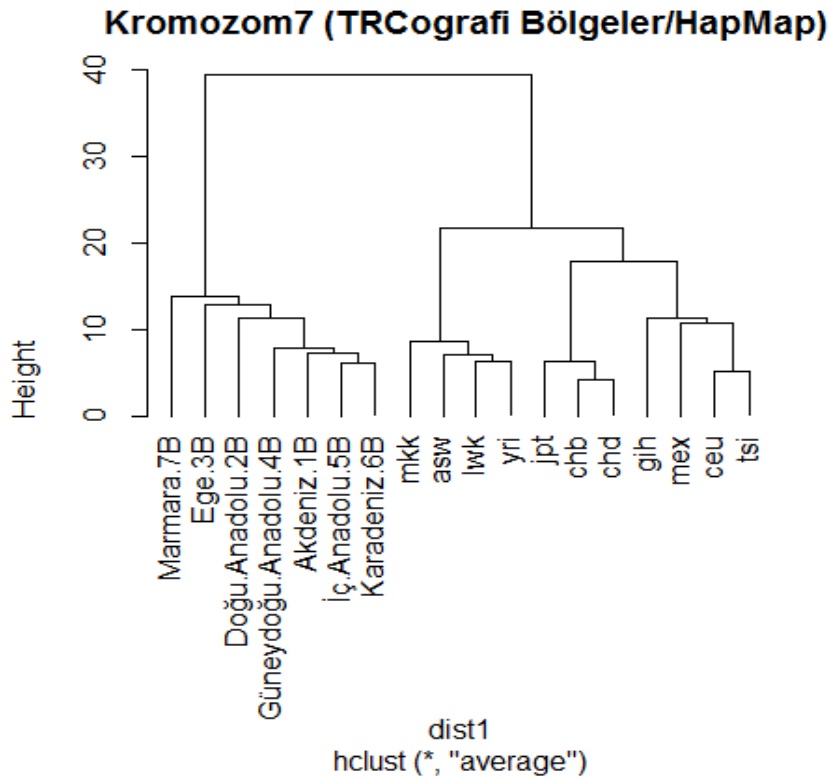
**EK101.** 5. Kromozom TR Bölgeleri ve HapMap Popülasyonları Hiyerarşik Kümeleme



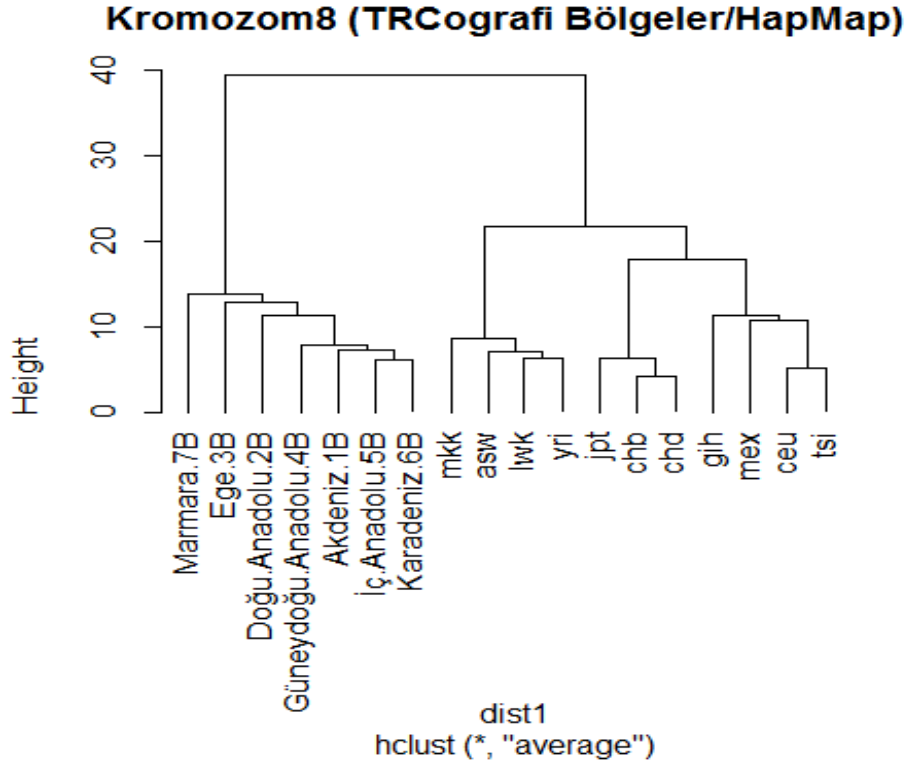
**EK102.** 6. Kromozom TR Bölgeleri ve HapMap Popülasyonları Hiyerarşik Kümeleme



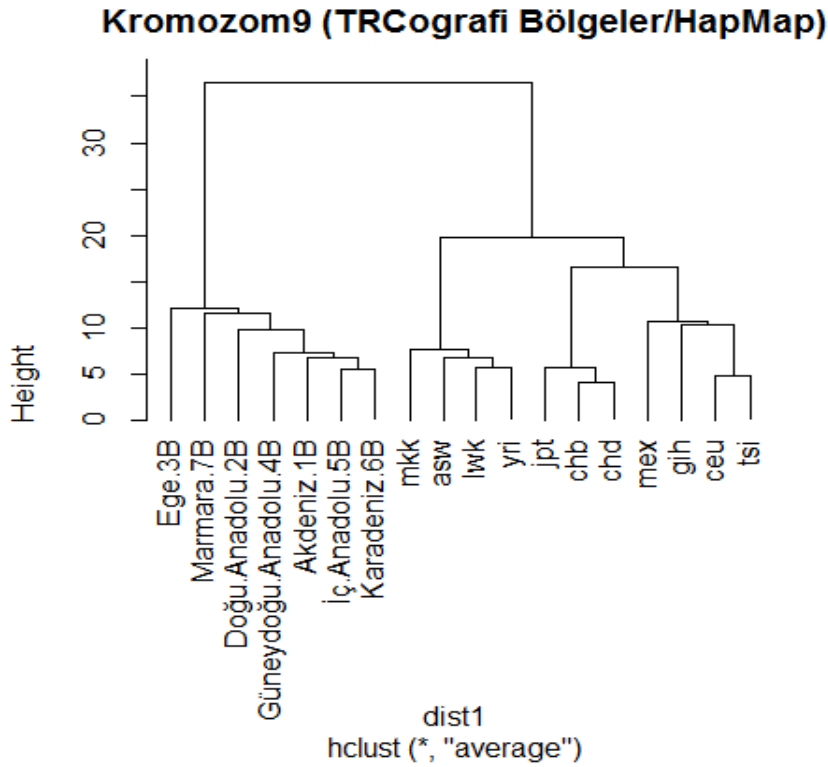
**EK103.** 7. Kromozom TR Bölgeleri ve HapMap Popülasyonları Hiyerarşik Kümeleme



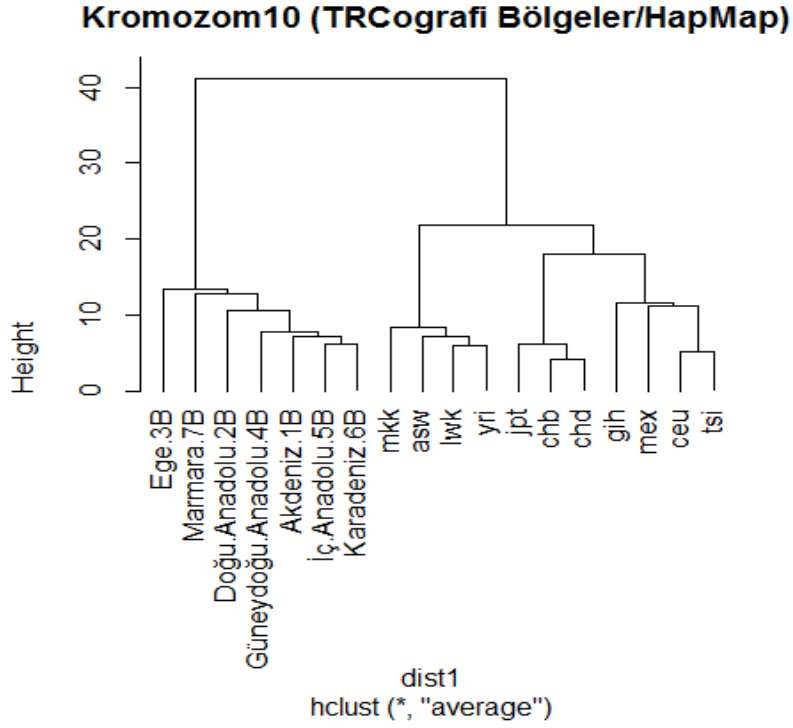
**EK104.** 8. Kromozom TR Bölgeleri ve HapMap Popülasyonları Hiyerarşik Kümeleme



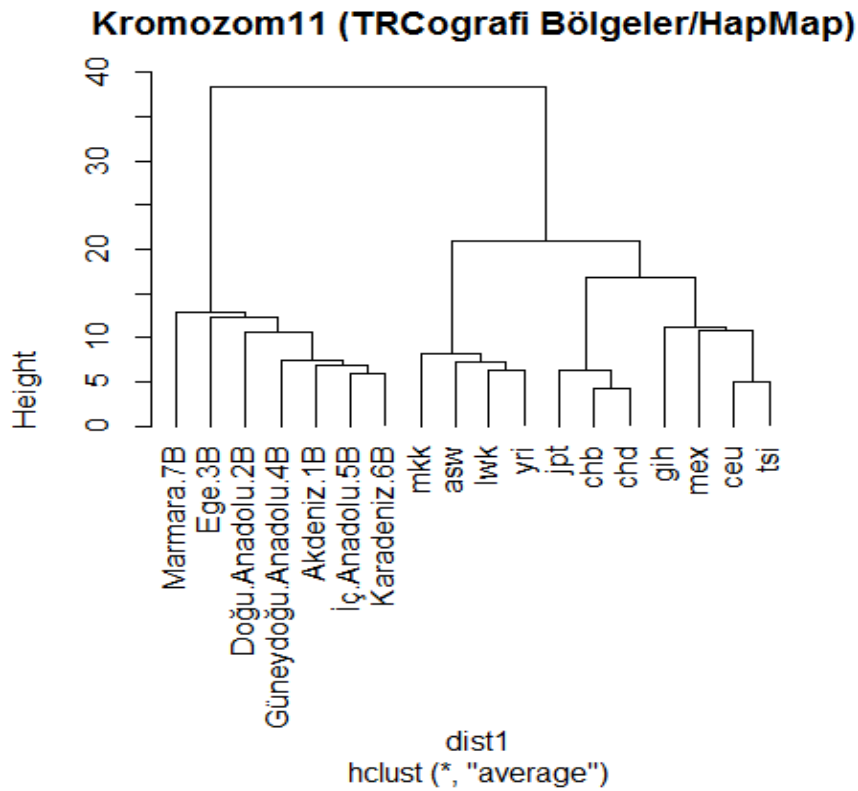
**EK105.** 9. Kromozom TR Bölgeleri ve HapMap Popülasyonları Hiyerarşik Kümeleme



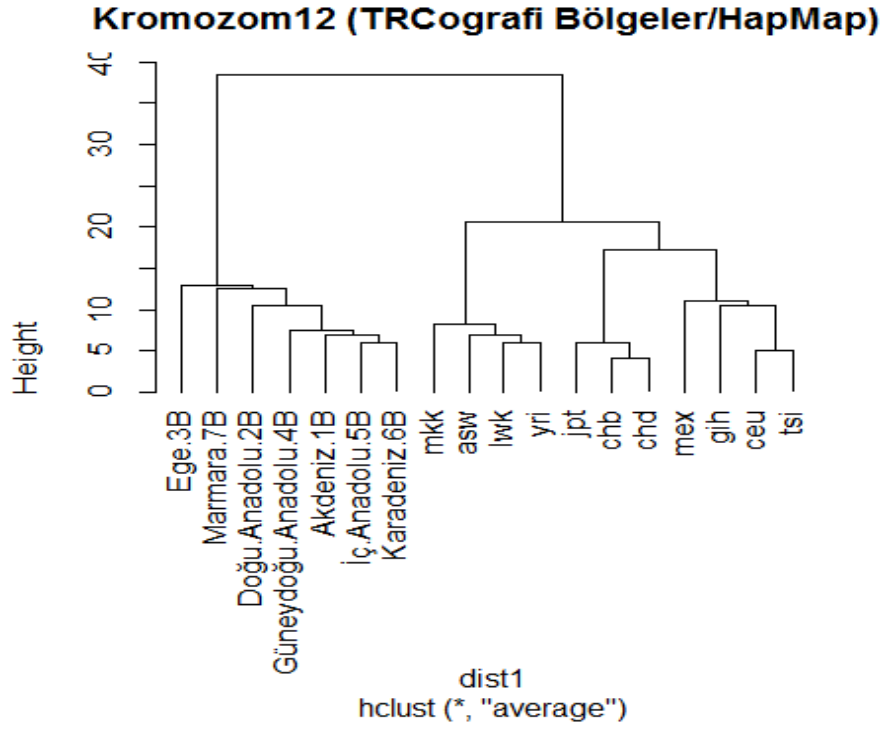
**EK106.** 10. Kromozom TR Bölgeleri ve HapMap Popülasyonları Hiyerarşik Kümeleme



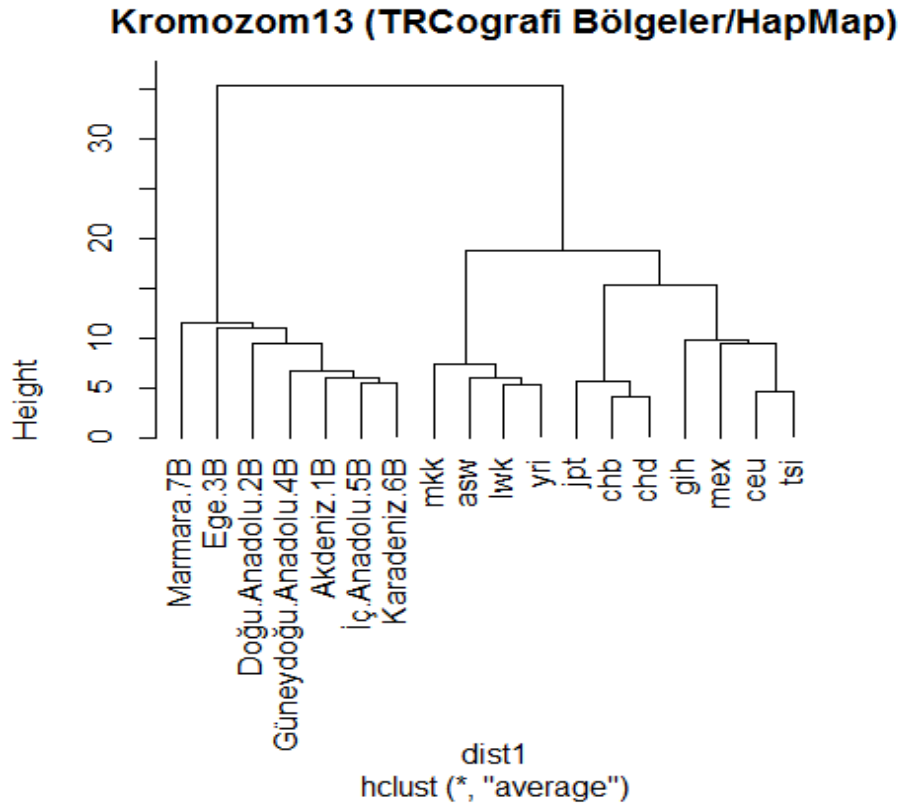
**EK107.** 11. Kromozom TR Bölgeleri ve HapMap Popülasyonları Hiyerarşik Kümeleme



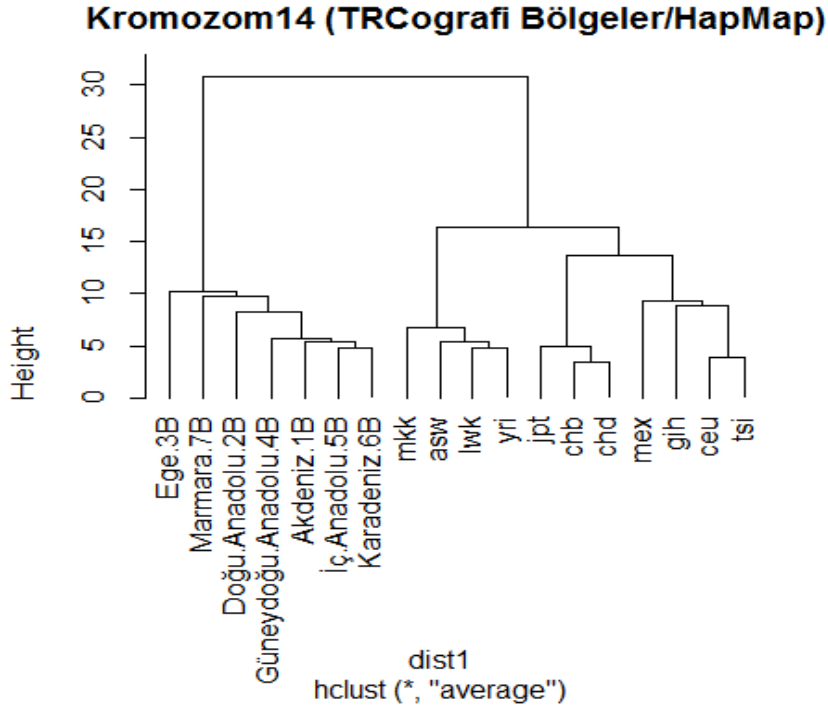
**EK108.** 12. Kromozom TR Bölgeleri ve HapMap Popülasyonları Hiyerarşik Kümeleme



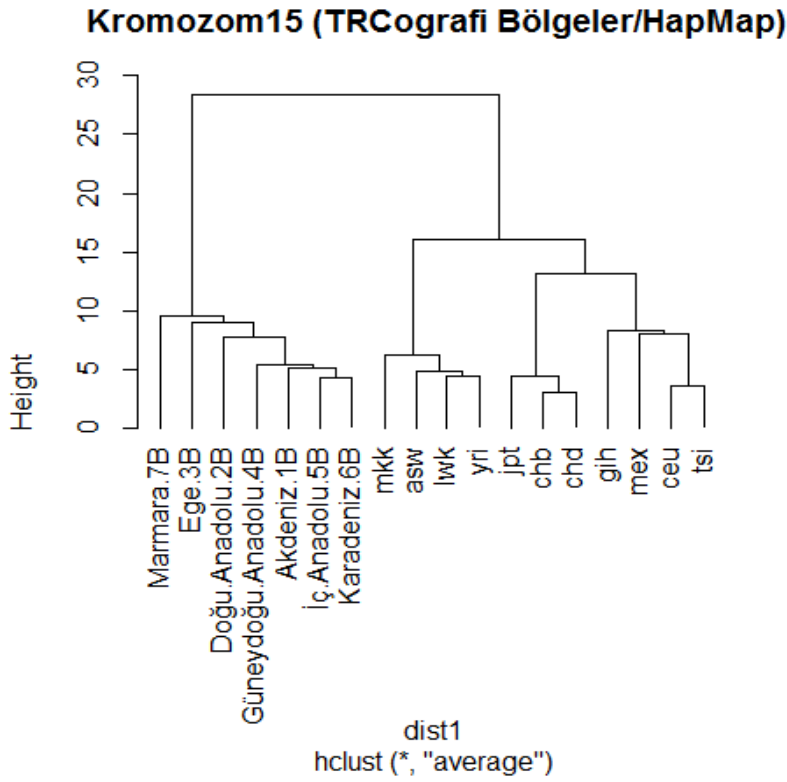
**EK109.** 13. Kromozom TR Bölgeleri ve HapMap Popülasyonları Hiyerarşik Kümeleme



EK110. 14. Kromozom TR Bölgeleri ve HapMap Popülasyonları Hiyerarşik Kümeleme

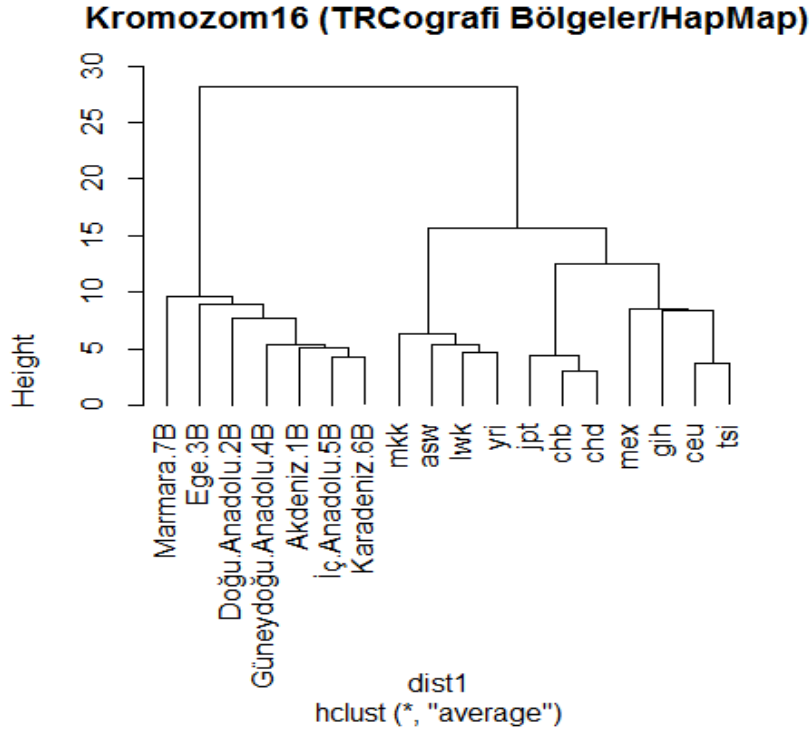


EK111. 15. Kromozom TR Bölgeleri ve HapMap Popülasyonları Hiyerarşik Kümeleme

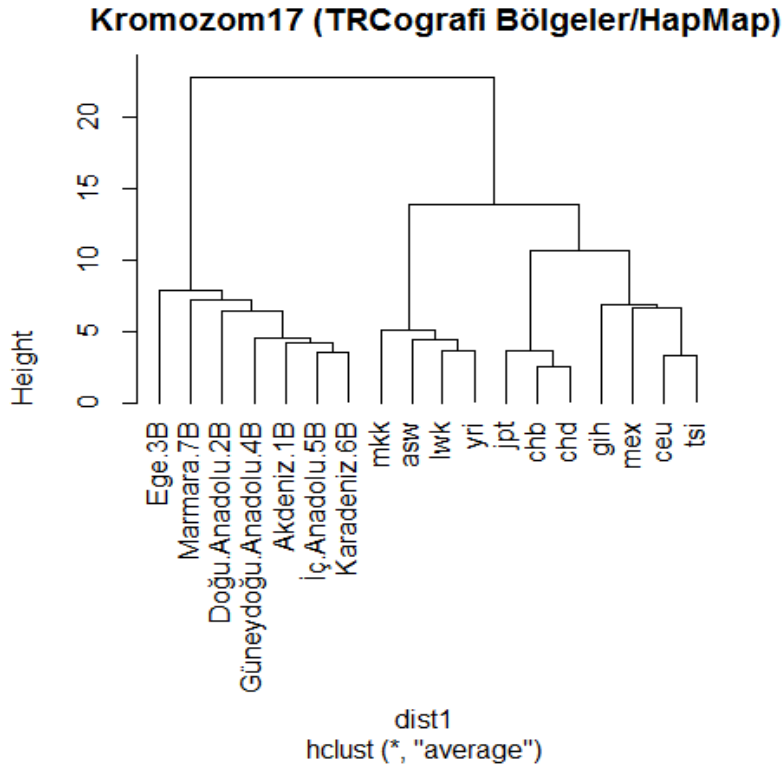




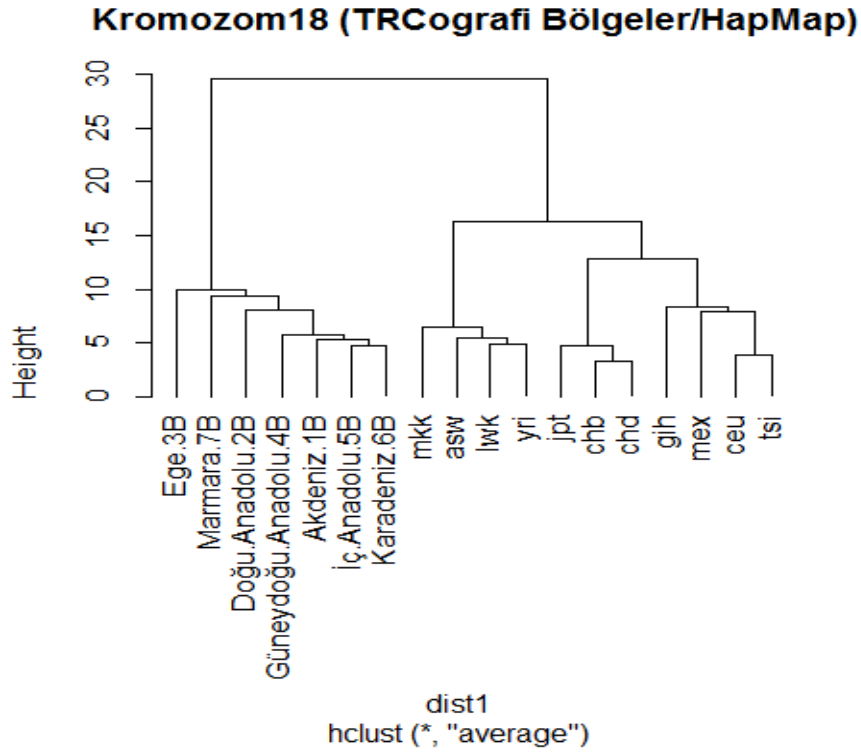
**EK112.** 16. Kromozom TR Bölgeleri ve HapMap Popülasyonları Hiyerarşik Kümeleme



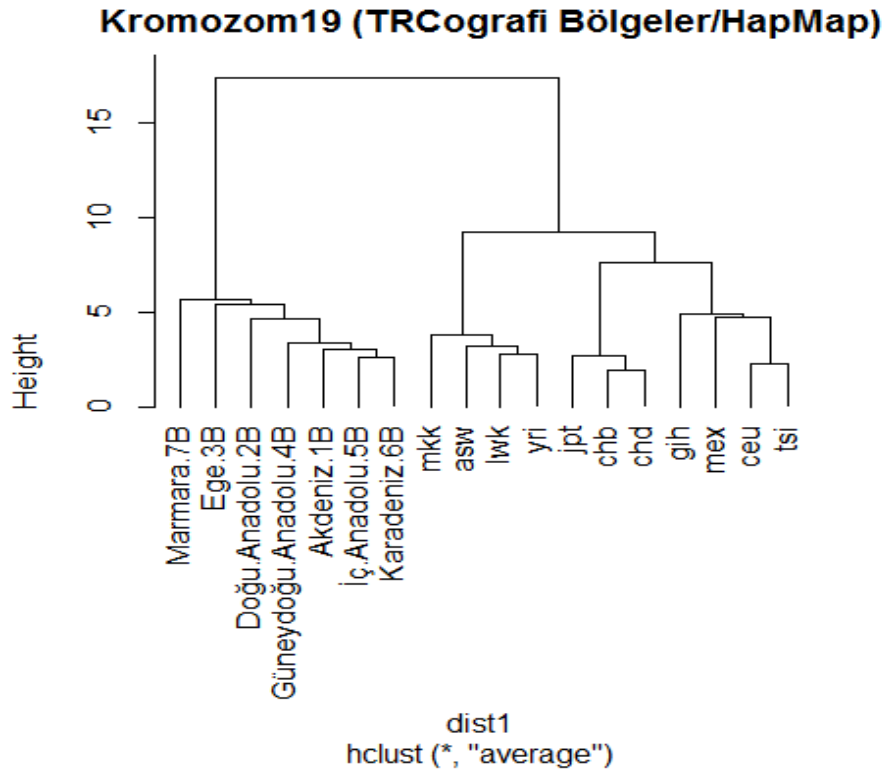
**EK113.** 17. Kromozom TR Bölgeleri ve HapMap Popülasyonları Hiyerarşik Kümeleme



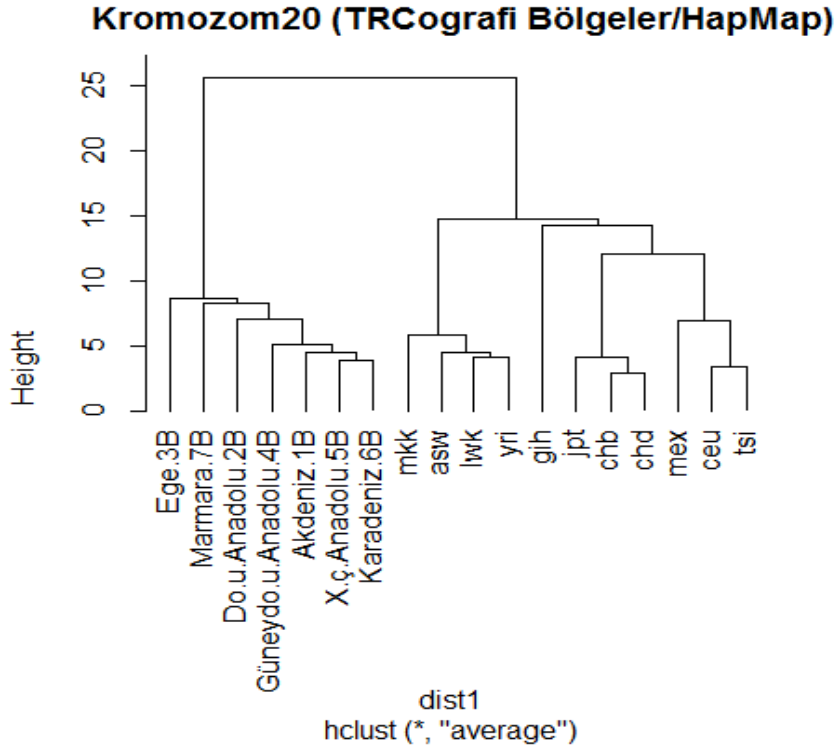
EK114. 18. Kromozom TR Bölgeleri ve HapMap Popülasyonları Hiyerarşik Kümeleme



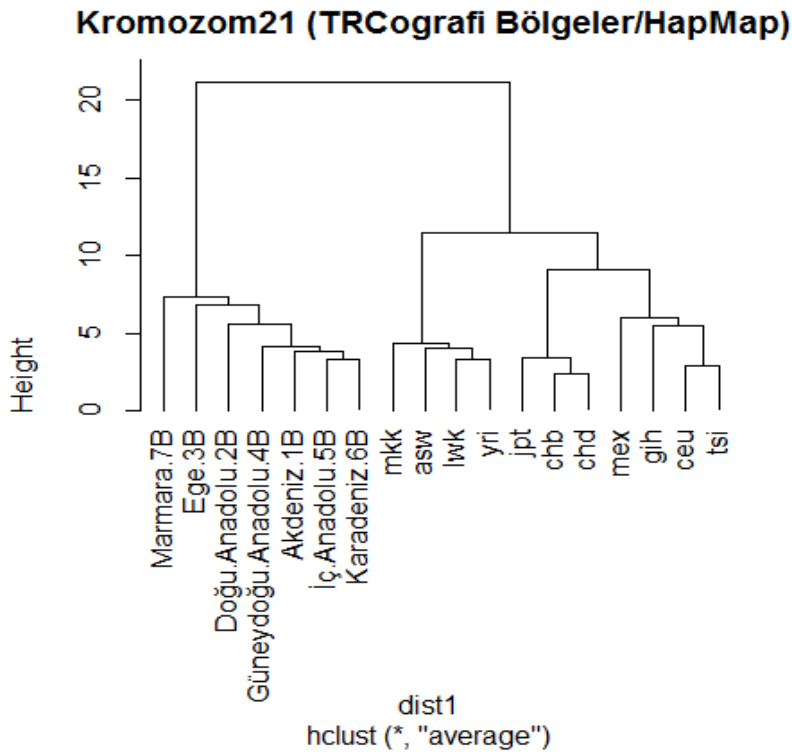
EK115. 19. Kromozom TR Bölgeleri ve HapMap Popülasyonları Hiyerarşik Kümeleme



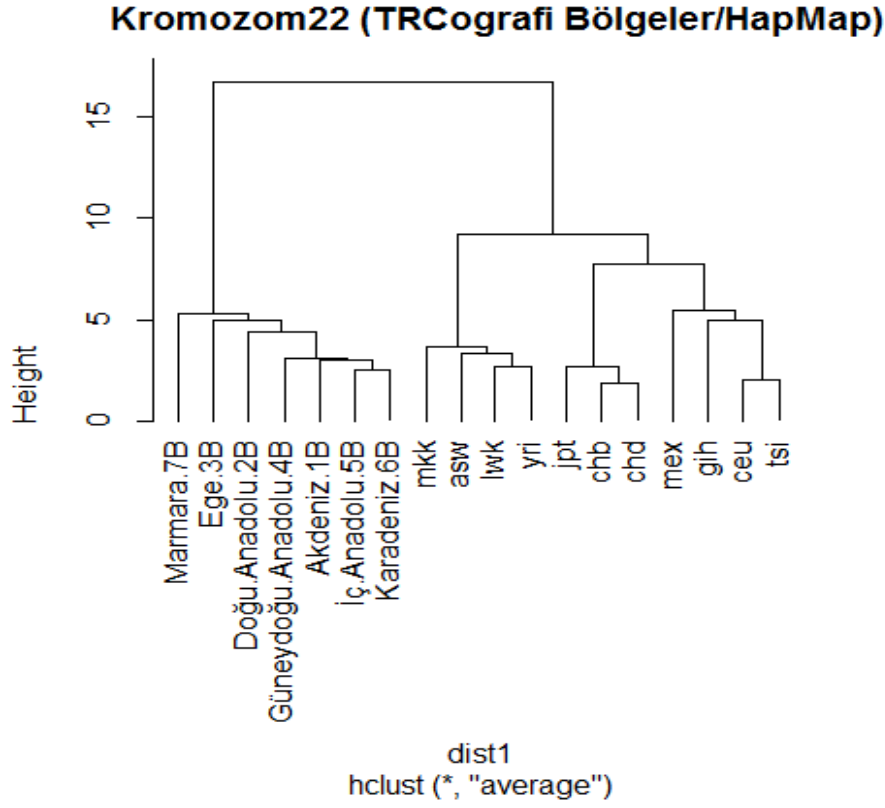
**EK116.** 20. Kromozom TR Bölgeleri ve HapMap Popülasyonları Hiyerarşik Kümeleme



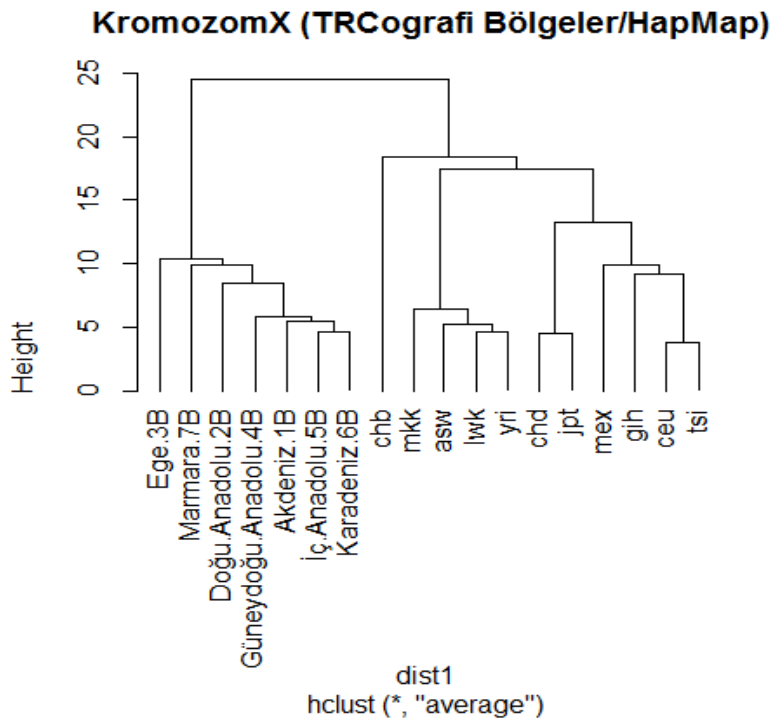
**EK117.** 21. Kromozom TR Bölgeleri ve HapMap Popülasyonları Hiyerarşik Kümeleme



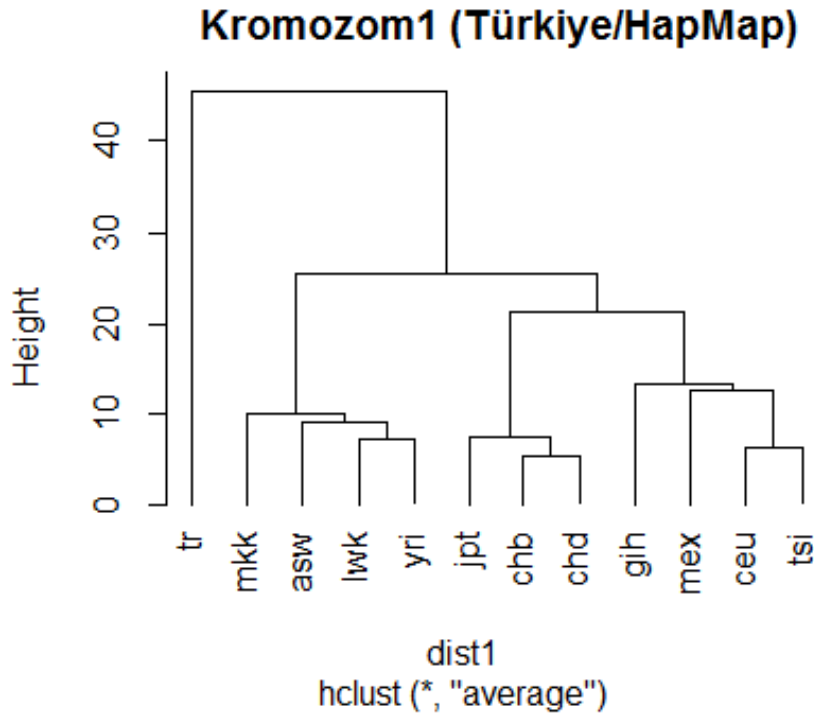
**EK118. 22. Kromozom TR Bölgeleri ve HapMap Popülasyonları Hiyerarşik Kümeleme**



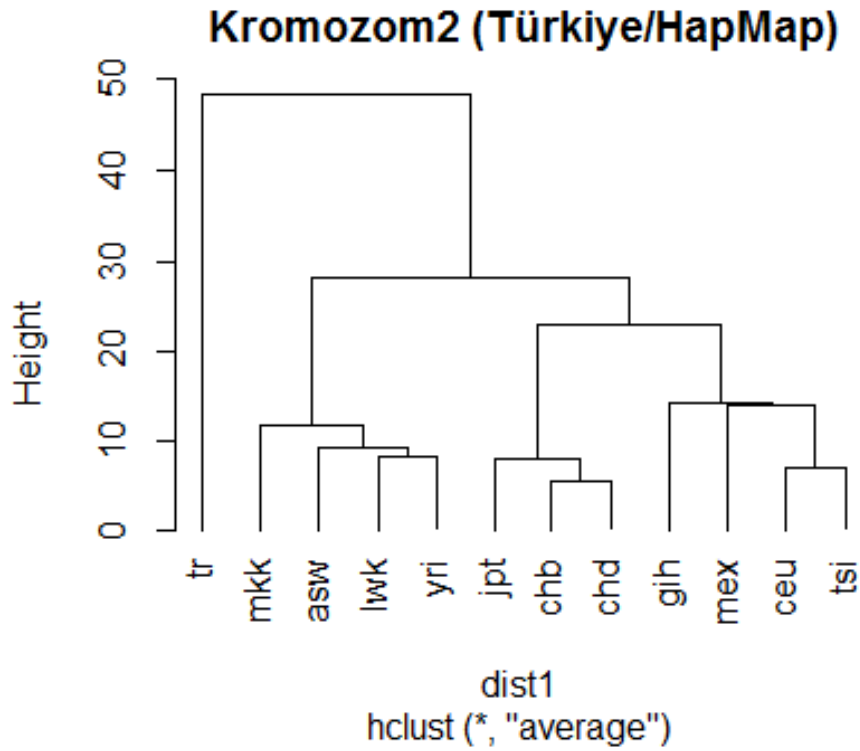
**EK119. X Kromozomu TR Bölgeleri ve HapMap Popülasyonları Hiyerarşik Kümeleme**



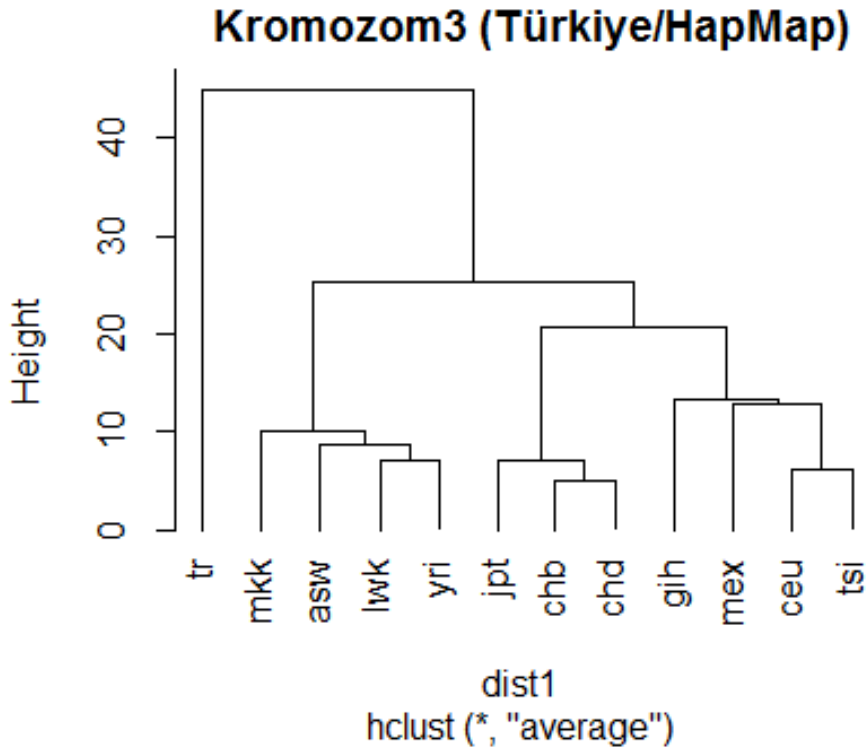
**EK120. 1. Kromozom Türkiye ve HapMap Popülasyonları Hiyerarşik Kümeleme**



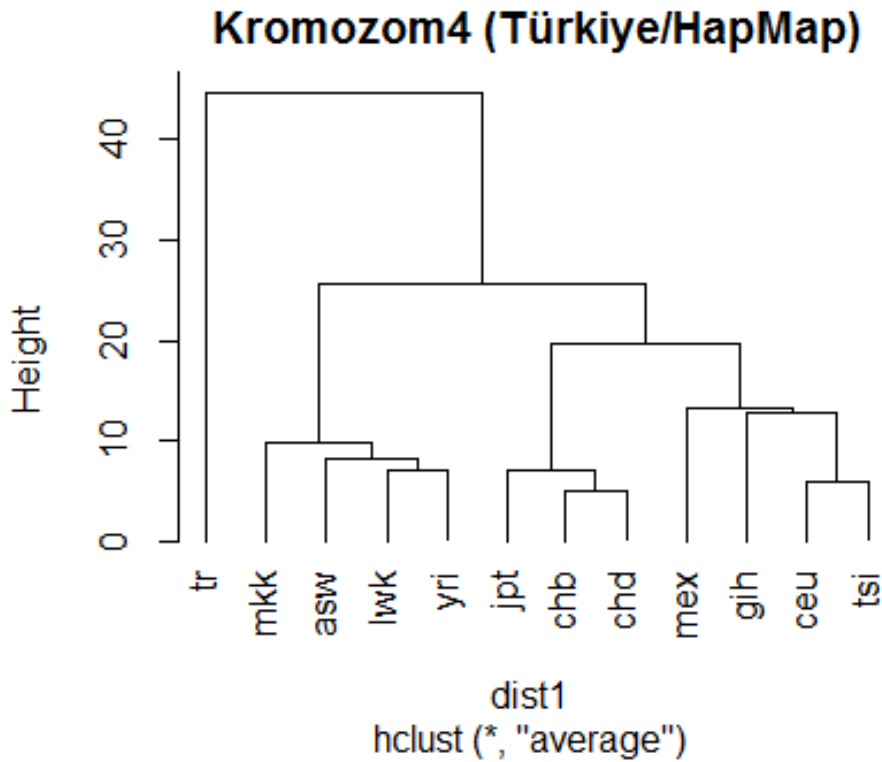
**EK121. 2. Kromozom Türkiye ve HapMap Popülasyonları Hiyerarşik Kümeleme**



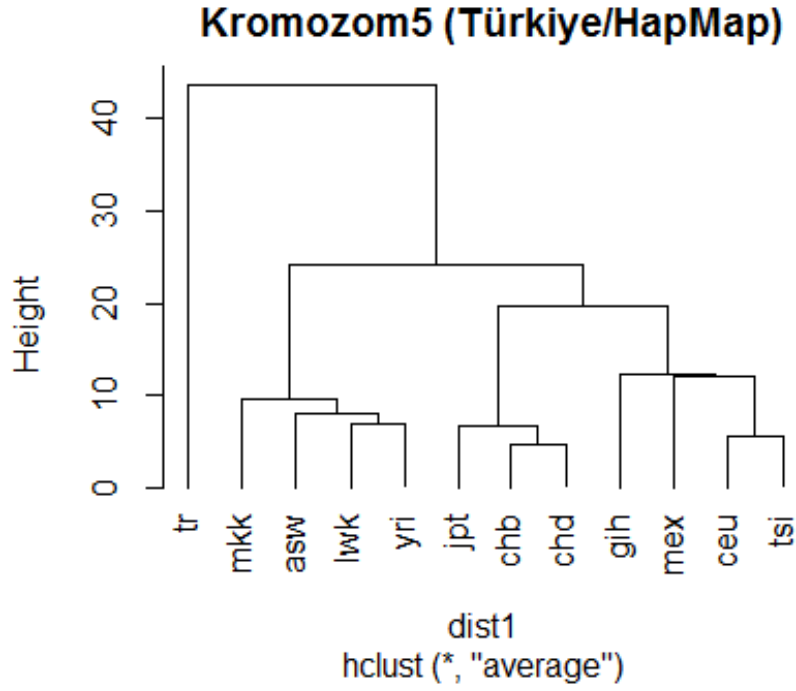
**EK122.** 3. Kromozom Türkiye ve HapMap Popülasyonları Hiyerarşik Kümeleme



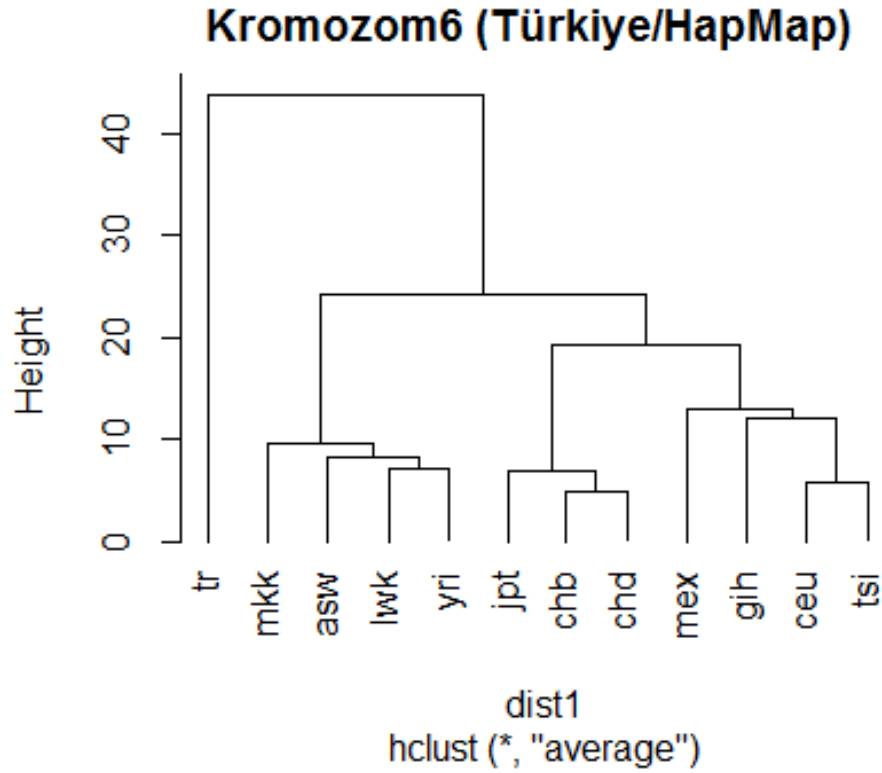
**EK123.** 4. Kromozom Türkiye ve HapMap Popülasyonları Hiyerarşik Kümeleme



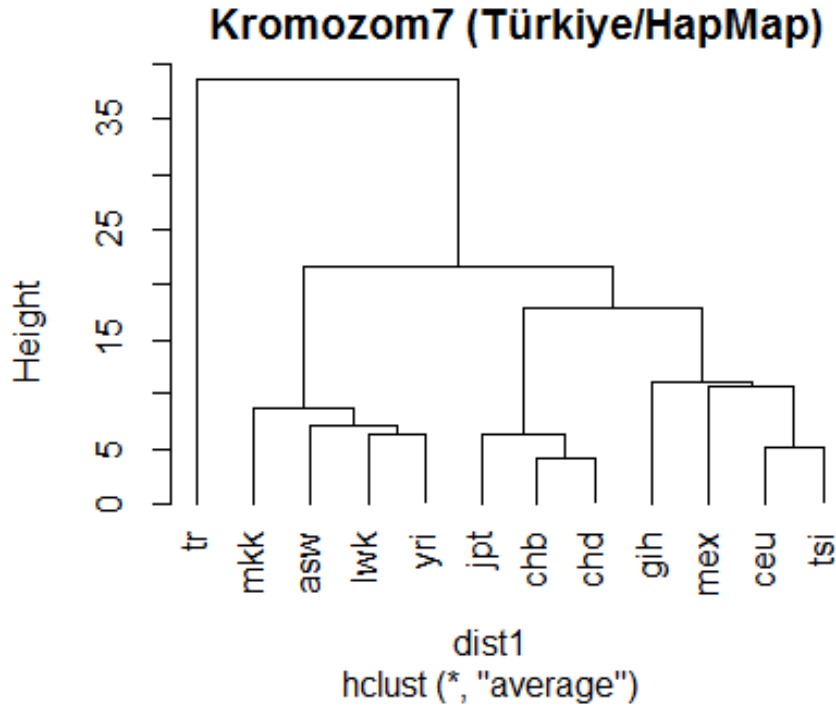
**EK124.** 5. Kromozom Türkiye ve HapMap Popülasyonları Hiyerarşik Kümeleme



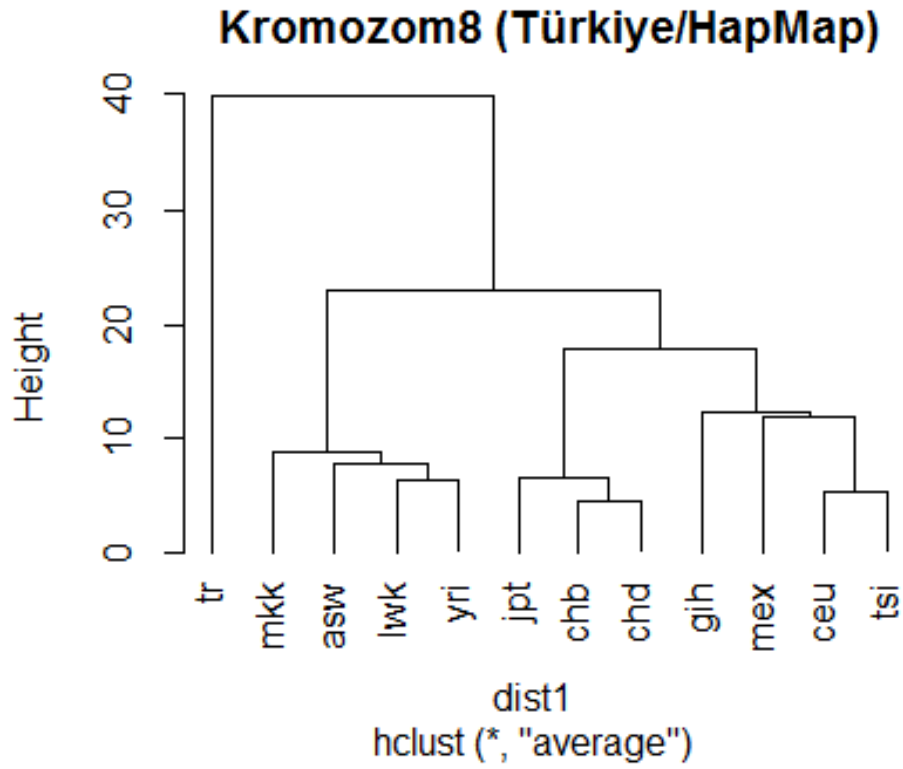
**EK125.** 6. Kromozom Türkiye ve HapMap Popülasyonları Hiyerarşik Kümeleme



**EK126.** 7. Kromozom Türkiye ve HapMap Popülasyonları Hiyerarşik Kümeleme

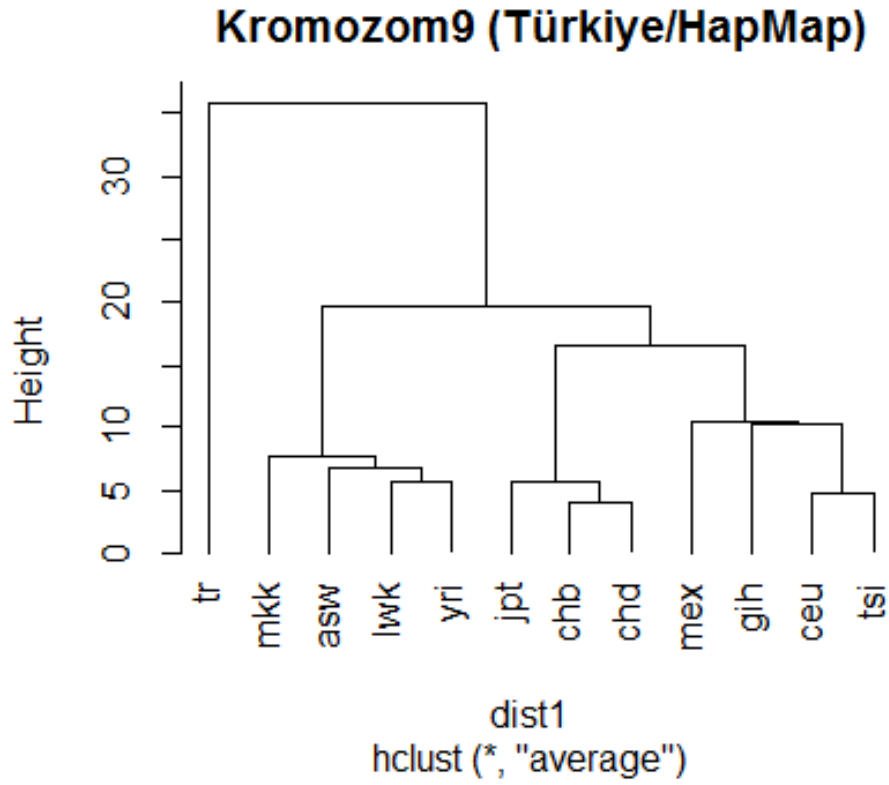


**EK127.** 8. Kromozom Türkiye ve HapMap Popülasyonları Hiyerarşik Kümeleme

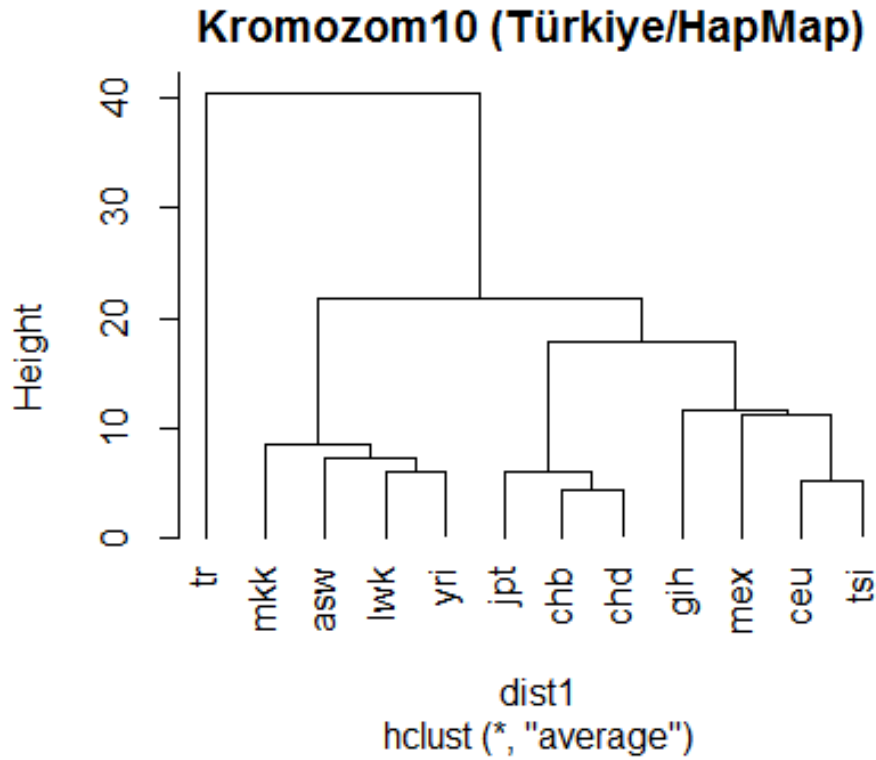




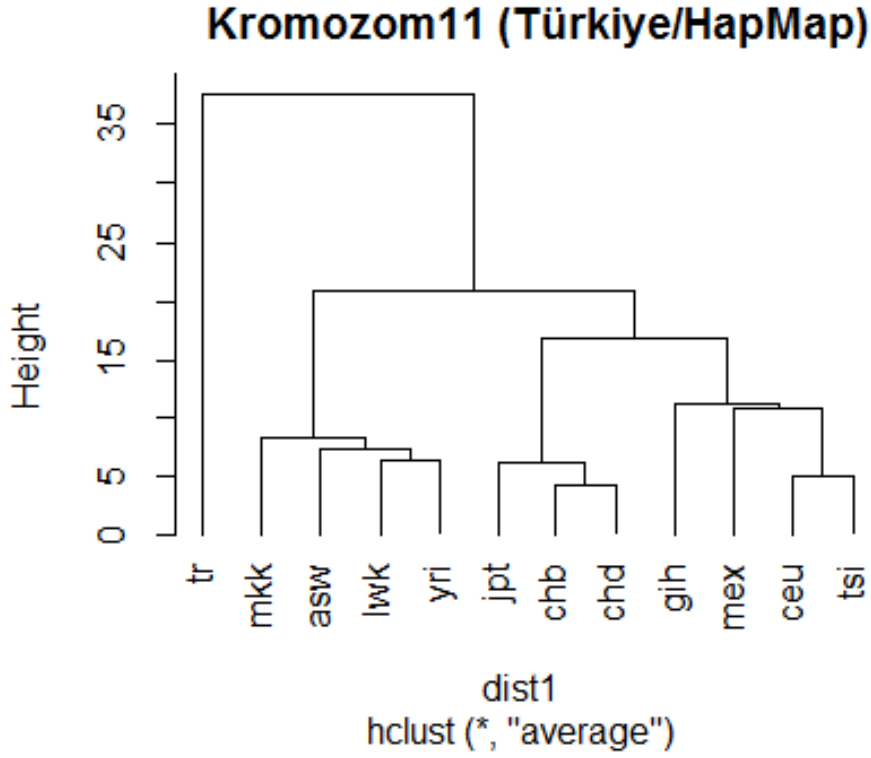
EK128. 9. Kromozom Türkiye ve HapMap Popülasyonları Hiyerarşik Kümeleme



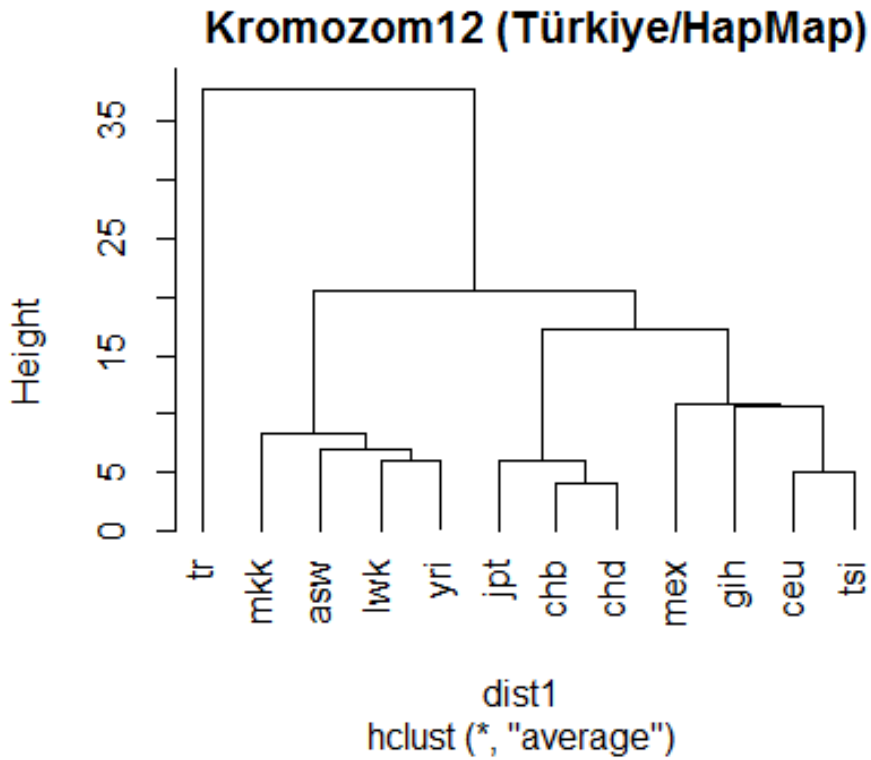
EK129. 10. Kromozom Türkiye ve HapMap Popülasyonları Hiyerarşik Kümeleme



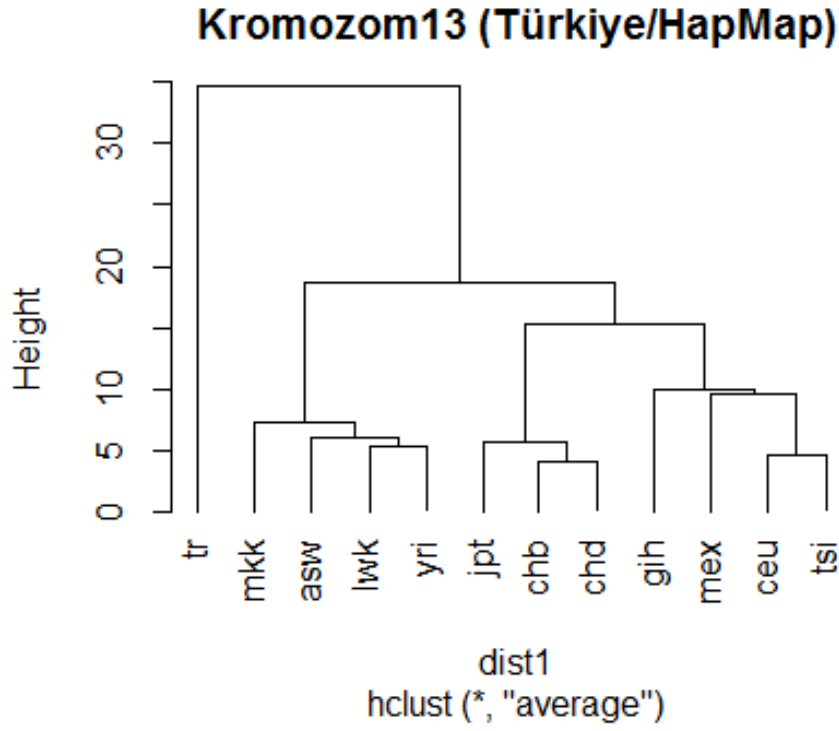
**EK130.** 11. Kromozom Türkiye ve HapMap Popülasyonları Hiyerarşik Kümeleme



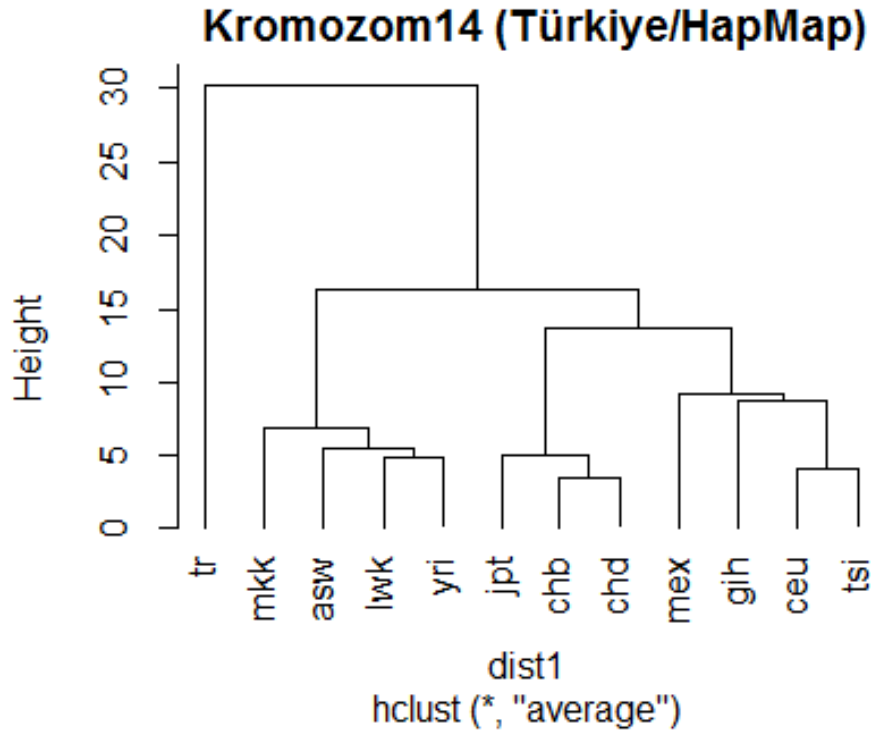
**EK131.** 12. Kromozom Türkiye ve HapMap Popülasyonları Hiyerarşik Kümeleme



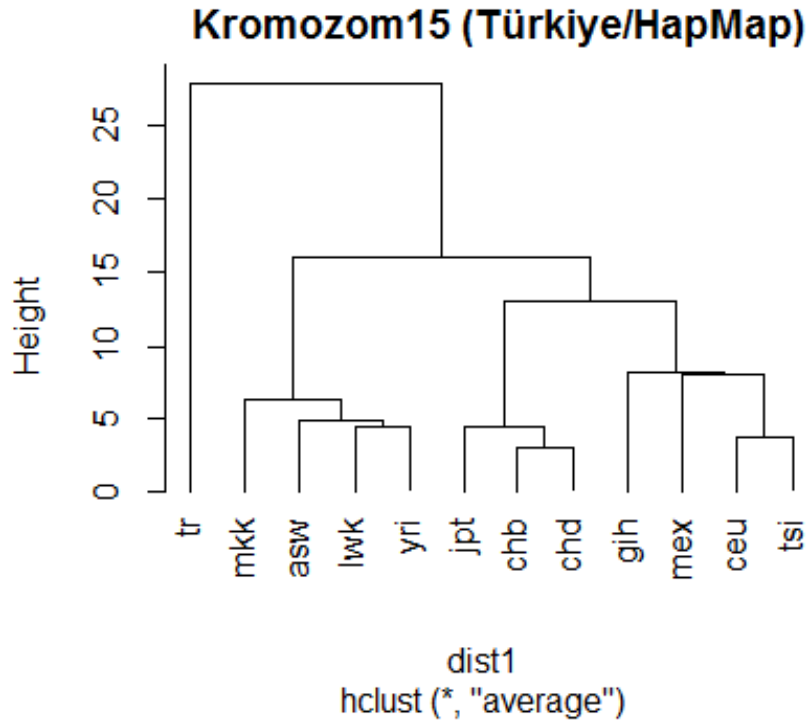
EK132. 13. Kromozom Türkiye ve HapMap Popülasyonları Hiyerarşik Kümeleme



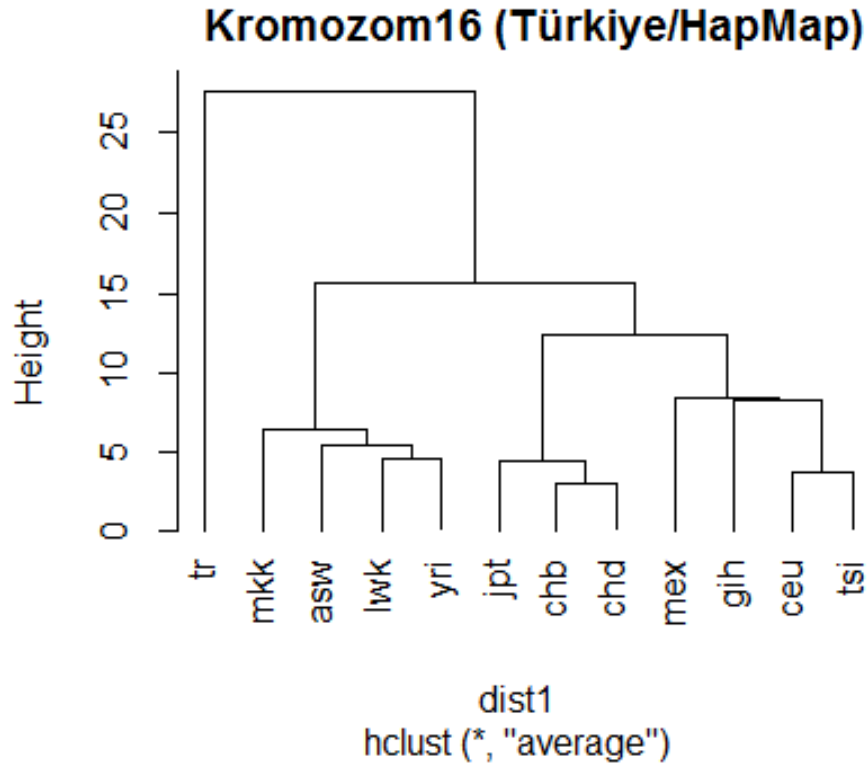
EK133. 14. Kromozom Türkiye ve HapMap Popülasyonları Hiyerarşik Kümeleme



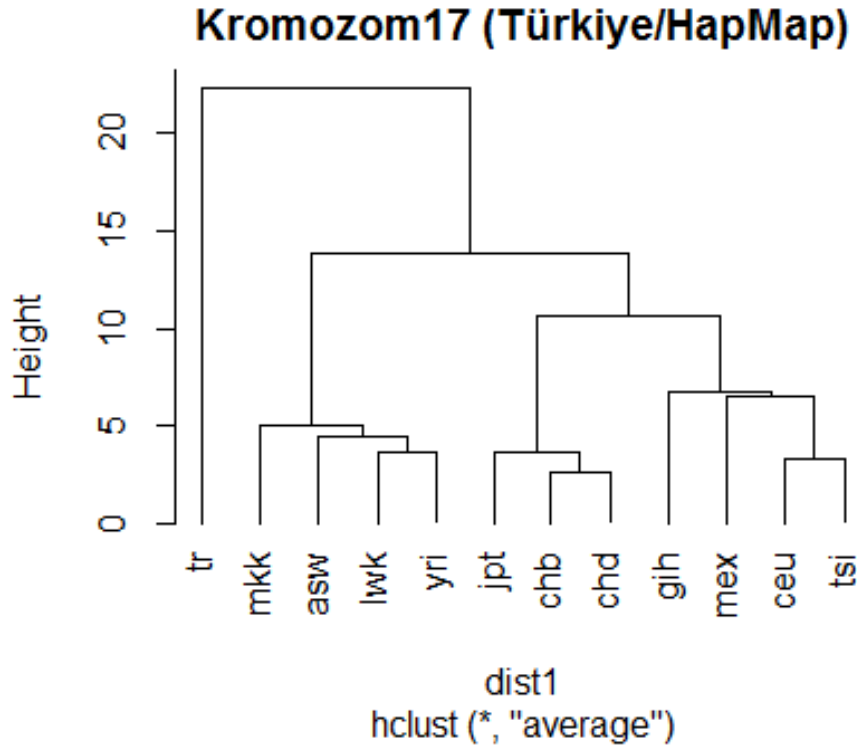
**EK134.** 15. Kromozom Türkiye ve HapMap Popülasyonları Hiyerarşik Kümeleme



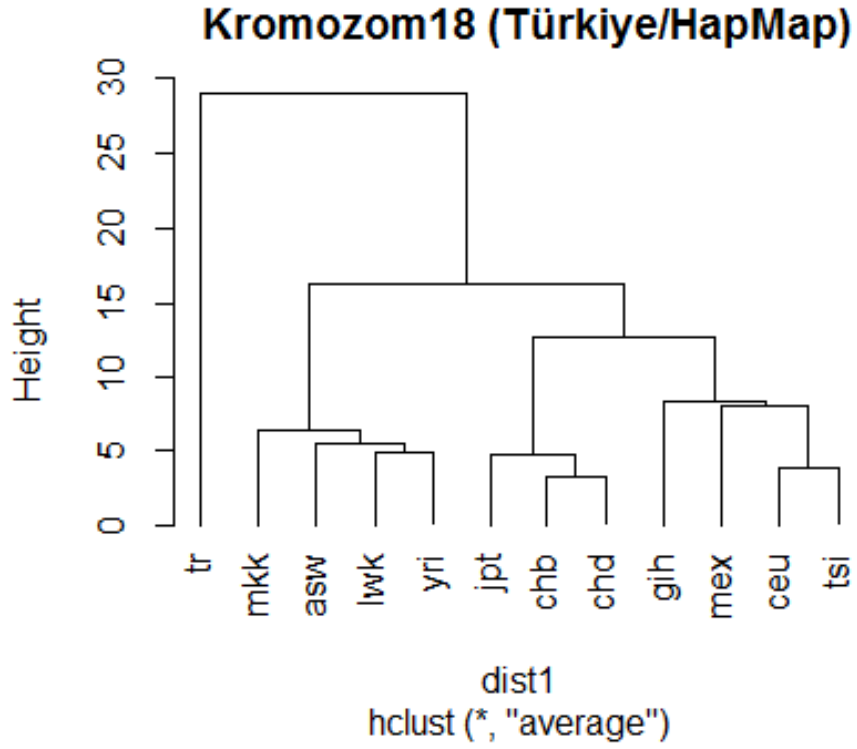
**EK135.** 16. Kromozom Türkiye ve HapMap Popülasyonları Hiyerarşik Kümeleme



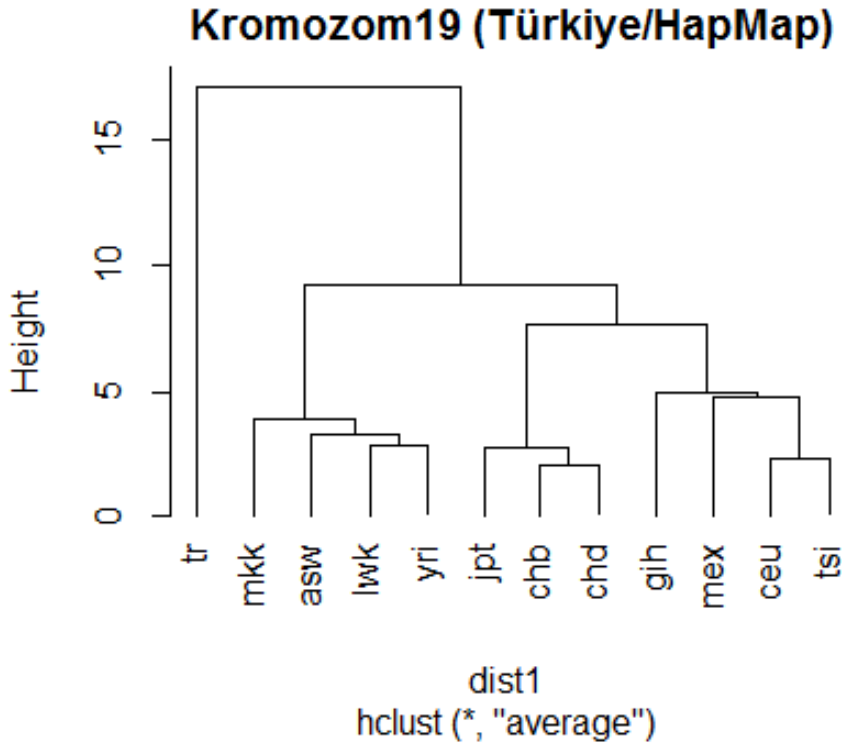
EK136. 17. Kromozom Türkiye ve HapMap Popülasyonları Hiyerarşik Kümeleme



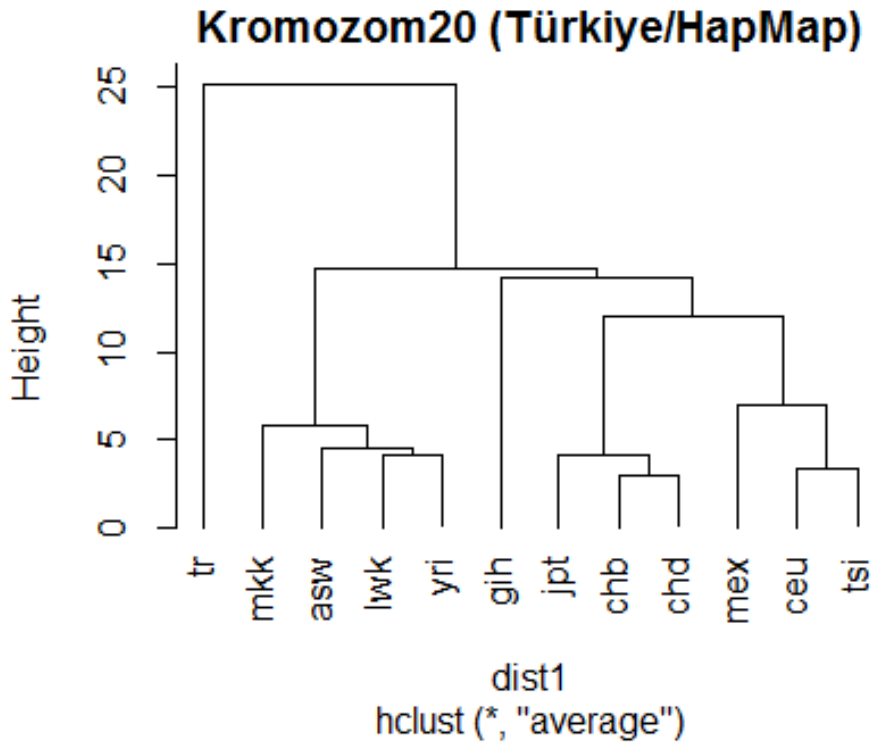
EK137. 18. Kromozom Türkiye ve HapMap Popülasyonları Hiyerarşik Kümeleme



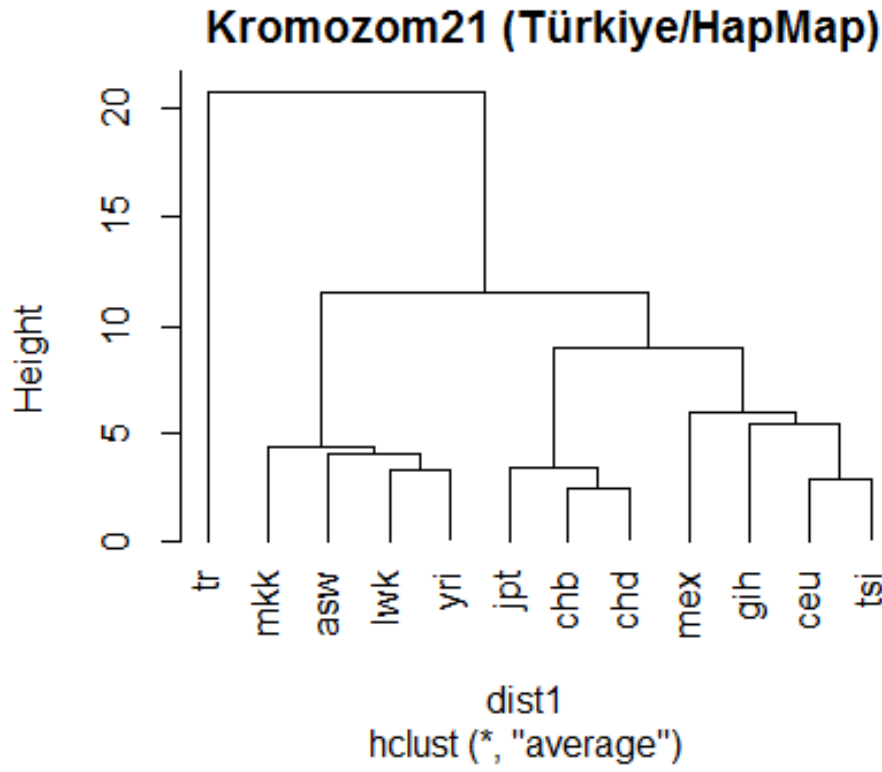
**EK138.** 19. Kromozom Türkiye ve HapMap Popülasyonları Hiyerarşik Kümeleme



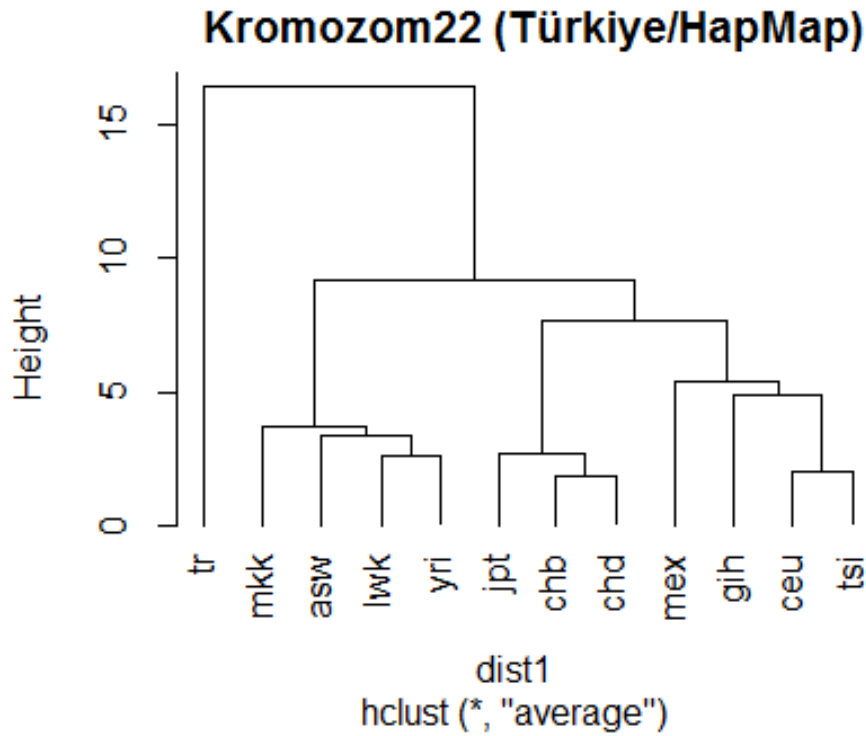
**EK139.** 20. Kromozom Türkiye ve HapMap Popülasyonları Hiyerarşik Kümeleme



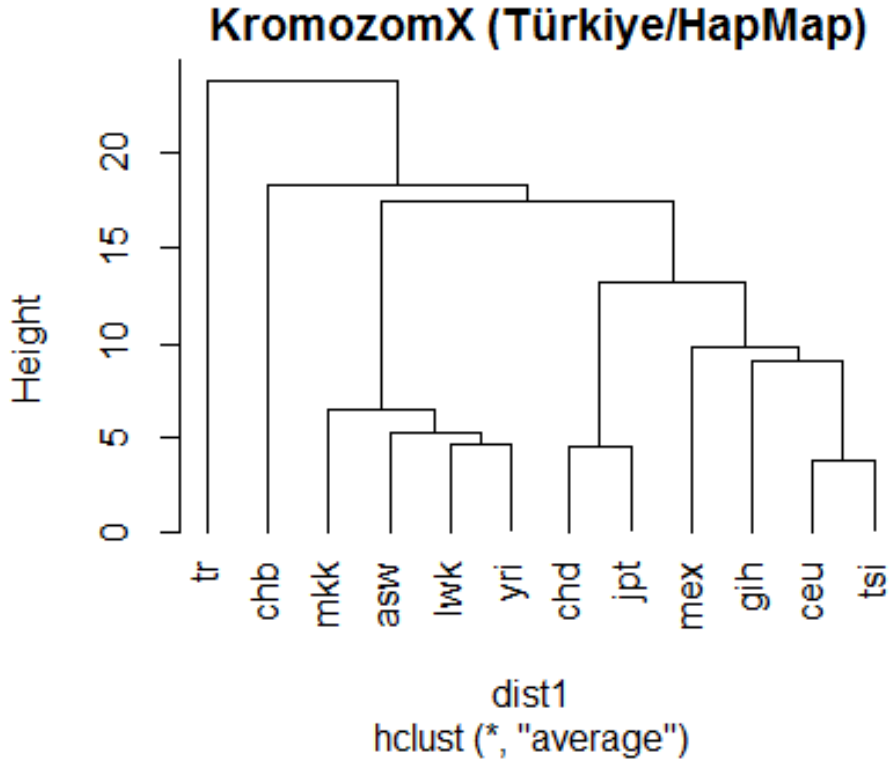
EK140. 21. Kromozom Türkiye ve HapMap Popülasyonları Hiyerarşik Kümeleme



EK141. 22. Kromozom Türkiye ve HapMap Popülasyonları Hiyerarşik Kümeleme



EK142. X. Kromozomu Türkiye ve HapMap Popülasyonları Hiyerarşik Kümeleme





### EK143. NN (Okunamayan) rs Frekans Hesaplamaları

```
N_data=267;

SNP_number=262264;

N_AA(SNP_number,1)=0;

N_AB(SNP_number,1)=0;

N_BB(SNP_number,1)=0;

N_NN(SNP_number,1)=0;

fid = fopen('excel_list.txt');

C = textscan(fid, '%s');

for i=1:N_data

    S{1,1}{i,1} = strcat(C{1,1}{i,1},'.xlsx');

end

for i=1:N_data

A = importdata(S{1,1}{i,1});

for j=1:SNP_number

AA = strcmp(A.textdata.Sayfa1{j,3}, 'AA');

AB = strcmp(A.textdata.Sayfa1{j,3}, 'AB');

BB = strcmp(A.textdata.Sayfa1{j,3}, 'BB');
```

```
NN = strcmp(A.textdata.Sayfa1{j,3}, 'NN');
```

```
if AA==1
```

```
    N_AA(j,1)=N_AA(j,1)+1;
```

```
end
```

```
if AB==1
```

```
    N_AB(j,1)=N_AB(j,1)+1;
```

```
end
```

```
if BB==1
```

```
    N_BB(j,1)=N_BB(j,1)+1;
```

```
end
```

```
if NN==1
```

```
    N_NN(j,1)=N_NN(j,1)+1;
```

```
end
```

```
end
```

```

end

for i=1:262264

    f_AA(i,1)=N_AA(i,1)/(N_AA(i,1)+N_AB(i,1)+N_BB(i,1));

    f_AB(i,1)=N_AB(i,1)/(N_AA(i,1)+N_AB(i,1)+N_BB(i,1));

    f_BB(i,1)=N_BB(i,1)/(N_AA(i,1)+N_AB(i,1)+N_BB(i,1));

end

for i=1:262264

    p(i,1)=f_AA(i,1)+1/2*f_AB(i,1);

    q(i,1)=f_BB(i,1)+1/2*f_AB(i,1);

end

for i=1:262264

    if p(i,1)<=q(i,1)

        MAF(i,1)=p(i,1)*100;

    else

        MAF(i,1)=q(i,1)*100;

    end

end

end

```

## ÖZGEÇMİŞ

Adı Soyadı : Pelin Fidanoğlu

Doğum Yeri : Ankara

Doğum Tarihi :1969

Medeni Hali : Bekar

Yabancı Dili : İngilizce

Eğitim Durumu (Kurum ve Yıl)

Lise : Ankara Lisesi, 1986

Lisans : ODTU Fen Edebiyat Fakültesi, Biyoloji, 1993

Yüksek Lisans :Gazi Üniversitesi Fen Bilimleri Enstitüsü, 2006

Çalıştığı Kurum/Kurumlar ve Yıl:

Gazi Üniversitesi Mühendislik Fakültesi, 2000, halen

İçişleri Bakanlığı, Mernis Projesi, 1996-2000

Milli Eğitim Bakanlığı, 1995-1996

Biotek Medikal, 1994-1995

Sesa Aş. 1993, 1994

### **SCI’da Yer Alan Makaleler:**

Murathan A, Fidanoğlu P. Diklorotriazin Reaktif Boyarmaddesi Sentezi ve Selülozik Elyafa Uygulanması. *Mühendislik Mimarlık Fakültesi Dergisi*, 24(2), 285-291, 2009

## **TEZDEN ÜRETİLEN YAYINLAR**

1. Genom Projeleri 5N1P:Ne, Nerede, Ne Zaman, Nasıl, Neden Ve Hangi Popülasyonda?  
Türk Hijyen ve Deneysel Biyoloji Dergisi, Kabul edildi, 2013

## GENOM PROJELERİ 5N1P:

### Ne, Nerede, Ne Zaman, Nasıl, Neden ve Hangi Popülasyonda?

Pelin Fidanođlu<sup>1</sup>, Nevin Belder<sup>1</sup>, Beyza Erdoğan<sup>2</sup>, Özlem İlk<sup>3</sup>, Farid Rajabli<sup>4</sup>, Hilal Özdađ<sup>1</sup>

#### ÖZET

Genom projeleri yaşamın şifresi olarak tanımlanabilecek olan ve bir organizmanın genomunu oluşturan DNA'nın dizisinin deşifre edilmesini hedeflemektedir. İnsan Genom Projesinin (İGP) fikri temelleri 1980'li yılların başlarında atılmıştır. 1990-2003 yılları arasında gerçekleştirilen ve 3,8 milyar dolara mal olan İGP ile sayısı ve kimliđi gizli tutulan gönüllülerden alınan örneklerden insan genom dizisi açığa çıkarılmıştır.

Genom verisinin anlamlandırılabilmesi için öncelikle genom "topoğrafyasının" ortaya konması, gen anatomisinin belirlenmesi gerekmiştir. Bu amaca ulaşabilmek için insan genom projesinin paralelinde birçok model organizmanın genom projesi gerçekleştirilerek bir genomun yapısına ait temel yapısal bileşenleri tanımlanmış ve genomun organizasyonel yapısı ile evrimsel gelişimine dair önemli bilgiler edinilmiştir.

2000'li yılların başlarından itibaren rezolüsyonu artarak gelişen mikrodizin teknolojisi ile genom topoğrafyasının en önemli bileşenleri olan Tek Nükleotit Polimorfizm (TNP) ve Kopya Sayısı Varyasyonlarının (KSV) geniş ölçekle taranması mümkün hale gelmiştir. Diğer yandan İGP'nin temelini 13 yılın sonunda tamamlanmasının ardından, 2004 yılında piyasaya çıkan yeni nesil dizileme teknolojisi ile James D. Watson'ın genomu yalnızca 2 aylık bir süre içinde 1 milyon dolarlık bir bütçe ile dizilenmiştir. 2004 yılından bugüne yeni nesil dizileme teknolojisindeki gelişmeler ile insan genomunun dizilenme süresi 1 güne ve maliyeti ise 6600 dolara inmiştir. Bu çalışmada özellikle tıp alanında büyük beklentiler yaratmış olan İGP'nin başlangıcından günümüze olan seyri anlatıldıktan sonra genom bilgisinin anlamlandırılabilmesi için modellenebilmesi ve hesaplanabilir hale gelmesinin gereğinin altı çizilecek, kişisel genetik tanı ve tedaviye giden yolda yapılan çalışmalar özetlenecektir.

**Anahtar Kelimeler:** İGP, Haplotip, Varyasyon, TNP

**Genome Projects:  
What, Where, When, How and in Which Population?**

**ABSTRACT**

Genome projects aim to decode an organism's complete set of deoxyribonucleic acid (DNA), which can be described as the living code of organism. The idea of the Human Genome Project (HGP) was conceived in the early 1980s. The project was started at 1990 and finished at 2003. The sequencing of the whole human genome derived from the DNA of several anonymous volunteers, costed 3.8 billion dollar.

In order to annotate the genome data, the "topography of the genome" and the anatomy of the genes should have been revealed. For this purpose, genome projects of several model organisms was carried out in parallel with HGP with the aim to identify basic structural components, organizational structure and evolutionarily development of the genome.

With the advent of microarray technology in the early 2000s, high-throughput screening of Single Nucleotide Polymorphisms (SNPs) and Copy Number Variations (CNVs) became feasible. After the completion of HGP in 13 years, James D. Watson's genome was sequenced with 1 million dollar budget in just 2 months using next generation sequencing technology. Today a human genome can be sequenced in just one day with the cost of 6600 USD.

In this paper the HGP which created big expectations especially in medicine will be explained from its start to the present. Then we will summarize the studies paving the road to personalized medicine emphasizing the fact that to reveal the meaning of genomic information, it should become computable.

**Keywords:** HGP, Haplotype, Variation, SNP

## GİRİŞ

### 1. İnsan Genom Projesi

Gregor Mendel'in bezelye bitkisi üzerinde yaptığı çalışmalar sonucunda kalıtımın kurallarını keşfetmesi ile başlayan bir çağ, kalıtımın doğasını bütünüyle anlayabilmek için başlatılan İGP ile başka bir bilimsel boyut kazanmıştır. Böylece İnsan Genom Projesi (İGP)'den önce çalışma alanı fizik ve kimya bilimleri ile sınırlı olan biyoloji bilimi, yanına matematik, istatistik, bilgisayar ve elektronik mühendislikleri gibi bilim dallarını da alarak disiplinler arası nitelik kazanmıştır.

1990 yılında resmi olarak başladığı kabul edilen İGP ile insan haploit genomuna ait 3,3 milyar nükleotit baz dizisinin belirlenmesi ile genomdaki mevcut bütün genlerin tespit edilmesi amaçlanmıştır. Proje kapsamında bilim adamları ve araştırmacıların çalışmalarını yürütebilmeleri için, elde edilecek verilerin veri tabanının oluşturulması ve kullanıcılara sunulması, ilgili teknolojilerin özel sektöre aktarılması ve ortaya çıkabilecek, legal, etik ve sosyal durumlara dikkat çekilmesi de yine bu projenin hedefleri arasında yer almıştır (1).

İGP Amerika merkezli bir proje olmakla beraber dünya üzerinde birçok laboratuvar 22 otozomal ve iki cinsiyet kromozomunu dizilemek ve haritalamak için projeye katkıda bulunmuştur (Tablo 1). Dizileme çalışmaları altı ülke başkanları (Amerika, İngiltere, Japonya, Fransa, Almanya ve Çin) tarafından desteklenmiş ve "insan yaşamı moleküler talimat" kitabı olarak adlandırılacak insan genom DNA'sına özgü üç milyar baz çifti temel dizisinin elde edildiği ortak bir bildiri ile yayınlanmıştır (Tablo 1). Projenin resmi olarak tamamlandığı 12 Nisan 2003 yılı tesadüfen James D. Watson ve Francis Crick'in DNA'nın çift sarmal yapısını keşfetmesinin 50. yılına denk gelmiştir. DNA temel baz dizisinin elde edilmesi sadece sonun başlangıcına işaret etmiştir (2).

3.8 milyar dolarlık büyük bir harcama sonucunda nihayete ulaşan İGP için yapılan bu harcama dev bir yatırım niteliği taşımaktadır. Bu potansiyeli öngören Çin, projenin yalnızca %1'ini yapabilmek için üç milyar dolar yatırım yapmıştır. İGP'nin yarattığı ekonomik yatırım hacmi toplamda 796 milyar dolar olarak hesaplanmıştır. Bu hesabın ayrıntıları Life Tech. Corp.'un sponsorluğunda bağımsız bilimsel ARGE organizasyonu Battle tarafından yürütülen modelleme çalışması ile ortaya konmuştur. Çalışma sonucuna göre İGP için ABD'nin yatırım



yaptığı her bir dolar ekonomiye 141 dolarlık kaynak sağlamıştır. Sadece 2010 yılında akademik ve ticari genom dizileme ve araştırma merkezleri 310.000 iş olanağı sağlamış ve ekonomik olarak ülkeye katkısı 67 milyar dolar olmuştur (3) (Tablo 1).

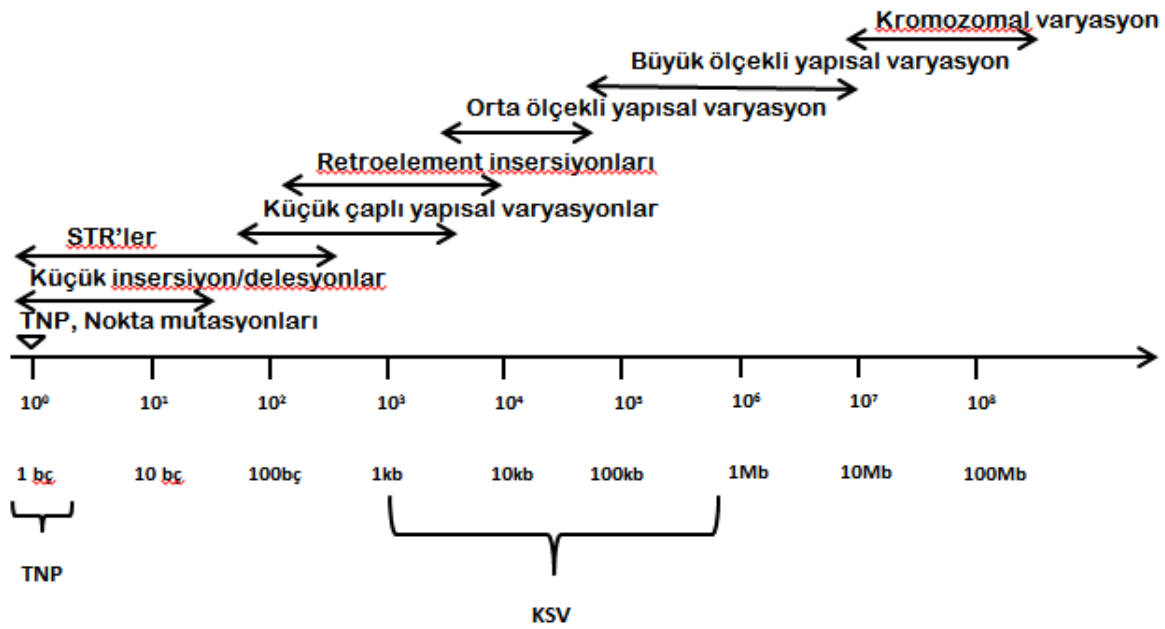
**Tablo 1.** İnsan Genom Projesine Katkıda Bulunan Merkez ve Ülkeler (3)

MERKEZ	ÜLKE
The Whitehead Institute/MIT Center for Genome Research	ABD
The Wellcome Trust Sanger Institute	İngiltere
Washington University School of Medicine Genome Sequencing Center	ABD
United States DOE Joint Genome Institute	ABD
Baylor College of Medicine Human Genome Sequencing Center	ABD
RIKEN Genomic Sciences Center	Japonya
Genoscope and CNRS	Fransa
GTC Sequencing Center, Genome Therapeutics Corporation	ABD
Department of Genome Analysis, Institute of Molecular Biotechnology	ABD
Beijing Genomics Institute/Human Genome Center, Institute of Genetics	Çin
Multimegabase Sequencing Center, The Institute for Systems Biology	ABD
Stanford Genome Technology Center	ABD
Stanford Human Genome Center and Department of Genetics	ABD
University of Washington Genome Center	ABD
Department of Molecular Biology, Keio University School of Medicine	Japonya
University of Oklahoma's Advanced Center for Genome Technology, Dept. of Chemistry and Biochemistry	ABD
Max Planck Institute for Molecular Genetics	Almanya
Cold Spring Harbor Laboratory, Lita Annenberg Hazen Genome Center	ABD
GBF - German Research Centre for Biotechnology	Almanya

## 2. Genom topoğrafyasının bileşenleri ve genomun dinamikleri

Genom en basit ifadesi ile bir organizmaya ait DNA dizi bilgisinin bütününe verilen addır. Ökaryot bir organizmanın genomu temel alındığında mesaj kodlayan ekzonlar ile kodlama fonksiyonu olmayan intronlardan oluşan genler ve genlerin ifadesini düzenlemekten sorumlu regülatör diziler bu genomun temel fonksiyonel yapısal birimleri olarak nitelendirilirler.

DNA dizisi yapısal olarak incelendiğinde genomun belli bir topoğrafyaya sahip olduğu gözlemlenmektedir. Genom topografyasının ortaya konabilmesi için 1980'lerden itibaren varyasyonlar üzerinde yoğunlaşmıştır. 1980'lerde ilk olarak Restriksiyon Parçaları Uzunluk Polimorfizm'leri (RFLP, birinci nesil), daha sonra Değişken Sayılı Bitişik Tekrarlar (Variable Number Tandem Repeats-VNTR, ikinci nesil) sayesinde çok sayıda genetik hastalık haritalanarak, hastalıktan sorumlu olan genler izole edilmiştir. 1990'larda mikrosatelit (Short Tandem Repeats-STR), üçüncü nesil) çalışmaları ve 2000'li yıllara gelindiğinde Tek Nükleotit Polimorfizmleri- TNP, dördüncü nesil) ile kopya sayısı çeşitliliği (Copy Number Variation- CNV) çalışmaları yoğunluk kazanmıştır (4, 5) (Şekil 1).



**Şekil 1.** Genomik Varyasyonların Boyutları. Çeşitli varyasyonlar için yaklaşık ölçüler belirtildiği şekildedir. Sınırların belirsiz olmasına rağmen tüm kromozomdan küçük ve bir kilobazdan büyük dizi değişiklikleri yapısal varyantlar olarak tanımlanmaktadır (4,5).

Genom dizileme teknolojilerinin gelişmesi ve genom projelerinin tamamlanmasıyla genom topoğrafyasının birer bileşeni olan bu varyasyonlar detaylı olarak tanımlanmış ve rezolüsyonları yani genomda ne sıklık ve aralıkta buldukları belirlenmiştir. Genom topografyasının bu bileşenleri yapısal varyasyon haritalarının üzerine işlenmiştir. Bir varyasyon haritası genomda değişken boyutlarda görülen varyasyon çeşitlerini (kromozomal,

yapısal ve dizi varyasyonları) göstermektedir. Genom boyunca farklı sıklıklarda ve boyutlarda görülen varyasyonlar, gen bölgelerinin yerlerini konumlandırmada güçlü birer araç olarak kullanılmakta ve bu nedenle genetik belirteç (markır) olarak adlandırılmaktadır.

Genomda ardışık olarak konumlanmış bahsi geçen bu varyasyonlar bir aile içinde veya popülasyonda nesiller boyu takip edildiğinde ilgili genomik bölgelerin birbirlerine göre olan konumları belirlenebilmektedir. Zira her eşey hücresinin oluşumu esnasında gerçekleşen genetik rekombinasyon süreci genomik bölgelerin birbirlerine göre olan uzaklığı ile doğru orantılı olarak gerçekleşmektedir. Dolayısıyla ilgili genomik bölgede hangi genin yer aldığı bilinmese de o bölgedeki genetik belirteçlerin birbirlerine göre olan konumları hesaplanabilmektedir. Genetik hastalıklara neden olan genlerin saptanmasına yardımcı olan genetik belirteçlerin birbirlerine göre konumlarının belirlenmesi genetik araştırmalarda verimli bir araç olarak kullanılmaktadır. Metot en genel anlamı ile lokalizasyonu aranan gen ile lokalizasyonu bilinen bir genetik belirtecin (markır) kuşaklar arasında birlikte kalıtılmasının test edilmesi esasına dayanmaktadır (6, 7).

Yukarıda bahsi geçen tüm varyasyon haritaları genoma farklı rezolüsyonlarda bakış sağlamaktadır. Hastalıklarla ilişkili olmayan bu varyasyonlar iki genom arasındaki %0,1 farklılığı oluşturmaktadır. Bu farklılıklar bireylerin fiziksel özelliklerinden sorumlu olabildikleri gibi hastalıklara yatkınlık veya direnç gibi özelliklerinden de sorumlu olabilmektedirler (8).

Genom varyasyonları arasında önemli bir yere sahip olan TNP'ler esas alındığında yukarıdaki ifade daha net açıklanabilir. Şöyle ki, TNP'ler, doğrudan hastalığa neden olmamakla beraber, bir kişinin belli bir hastalığa olan yatkınlığını belirleyebilmektedirler. Alzheimer hastalığı ile ilişkilendirilmiş olan ApoE (Apolipoprotein E) geni bu gelişimi açıklamak açısından örnek olarak verilebilir. ApoE geni dört ekson ve üç introndan oluşmakta ve E2, E3 ve E4 şeklinde üç olası alleli bulunmaktadır. Diğer allelere göre popülasyondanadır rastlanan ve hastalığa karşı koruma sağlayabildiği düşünülen ApoE2 allelinde C→T (Arg158Cys) nokta mutasyonu bulunmaktadır. Popülasyonda yaygın bulunan ApoE3 allelinin (Cys112, Arg158) hastalığa karşı nötral bir rol oynadığı düşünülmektedir. Popülasyonda görülme sıklığı %25-30 olan ApoE4 allelinde ise T→C (Cys112Arg) nokta mutasyonu bulunmaktadır. ApoE4 alleleline sahip olan bireylerin %40'ı yaşlılık dönemlerinde Alzheimer hastalığına yakalanmaktadır.

Ancak bireyin bu alleli taşıması kesin suretle Alzheimer hastası olacağı anlamına da gelmemektedir. Bu hastalığın gelişimi için kesin bir gösterge olarak ifade edilememektedir. Ziraiki E4 alleleline sahip olan bireylerde hiçbirzaman Alzheimer gelişmediği buna karşılık iki E2 alleleline sahip olan bireylerde ise Alzheimer hastalığının ortaya çıktığı durumlara rastlanabilmektedir (9). Alzheimer, obezite, kanser, kardiyovasküler hastalıklar gibi kalıtsal özellik gösterebilmekle beraber mendelyen kalıtım modelinin izlenmediği multifaktöryel ve multigenik karakter arzemektedir. TNP haritaları multigenik hastalıkların doğasının araştırılmasında önemli bilgi sağlamaktadır.

## **2.1. Genetik Belirteçler**

### **2.1.a. Restriksiyon Parçaları Uzunluk Polimorfizmi**

#### **(Restriction Fragment Length Polymorphism = RFLP)**

İnsan genomu boyunca, özellikle kodlayıcı olmayan bölgelerde, her 200 nükleotitte bir dizi farklılığı görülmektedir. Bu dizi farklılıkları tek nükleotit değişimleri olabildiği gibi, bir veya birden fazla nükleotitin delesyonu veya insersiyonu şeklinde de olabilmektedir. Bu değişimler bir restriksiyon enziminin kesim noktasını ortadan kaldıradıldığı gibi yeni bir kesim bölgesi de oluşturabilmektedir. Bu şekilde restriksiyon enzimleri kesim noktalarında oluşan varyasyonlar nedeniyle açığa çıkan parça uzunluklarındaki farklılıklar RFLP olarak adlandırılmaktadırlar. İnsan genomunda yaygın görülen RFLP'lerin binlercesi tanımlanmıştır. RFLP'ler kodominant mendelyen kalıtım modeli gösterdiklerinden genomda belli bir lokusa ait ailesel allellerin anneye (maternal) veya babaya (paternal) ait olduğunu ayırt etmede yardımcı olmaktadır. Bir ailede genetik hastalıkların nesilden nesile geçişini takip etmek amacıyla belirleyici olarak kullanılmaktadırlar. RFLP'ler aynı zamanda genetik bağlantı haritalarının oluşturulmasında da kullanılan birinci nesil genetik varyasyonlardır. White ve arkadaşları bu polimorfik belirteçleri kullanarak insan kromozomlarının genetik haritasını oluşturmuş araştırmacıların kullanımına sunmuşlardır (10-12).

### **2.1.b. Kısa Ardışık Tekrarlar (Mikrosatelit)**

#### **(Short Tandem Repeat =STR(Mikrosatellit))**

İnsan genomunda rastlanan orta sıklıkta tekrarlanan DNA baz dizileri, genellikle ya ardışık tekrarlanan ya da serpiştirilmiş diziler olarak genomda bulunmaktadırlar. Kodlamayan

bölgede, değişik motif ve uzunluklarda (2-10 baz çift) ardışık, orta sıklıkta tekrarlayan nükleotit dizileri mikrosatelit olarak adlandırılmaktadır. Genom boyunca dağılmış durumda olan mikrosatelitlerin bir DNA bölgesindeki tekrar sayısı bireyler arasında farklılık göstermektedir. Örneğin insanlarda en yaygın mikrosatelit ikili nükleotit (CA)<sub>n</sub> tekrarlarıdır ve tekrar sayısı genellikle 5 ve 50 arasında değişmektedir. Mikrosatelitlerin allel sayıları bu bölgelerin mutasyon oranları yüksek olduğu için çoğunlukla biallelik olan diğer yaygın polimorfizm çeşitlerinden üç kat daha fazla bilgilendirici olmaktadır. Bu nedenle mikrosatelitler genetik bağlantı ve evrimsel analizlerde sıklıkla kullanılmıştır (13).

1985'ler de Polimeraz zincirleme tepkimesinin (Polymerase Chain Reaction- PCR) geliştirilmesi DNA analiz işlemlerinin hacim ve boyutlarının artışına neden olmuştur. Southern blot teknolojisi temelli RFLP analizleri 1990'larda yerini PCR temelli mikrosatelit analizlerine bırakmıştır. Mikrosatelit belirteçlerle yapılan çalışmalarda çok az miktarda DNA'ya ihtiyaç duyulması ve yüksek işlem hacimli sistemlere adaptasyon kolaylığından dolayı, mikrosatelitler bağlantı analizlerinde ve popülasyon genetiğinde başarılı bir şekilde kullanılmaya başlanmıştır. Heterozigotluk oranları fazla olan mikrosatelitler, paternal alleler için büyük oranda ayırıcı ve bağlantı analizlerinde oldukça bilgilendiricidirler.

Mikrosatelitler ilk keşfedildikleri dönemde bazı bilim adamları tarafından işe yaramayan diziler olarak nitelendirilmişlerse de son yıllarda yapılan çalışmalar ile bu dizilerdeki değişikliklerin özellikle ileri yaşlarda görülen sinir sistemi hastalıklarında etkin olduğunu, diğer bazı organizmalarda da gen ifadesinde yer aldığını ve bazı durumlarda ise kodladığı protein üzerinde etkili olduğu gösterilmiştir. Örneğin Huntington hastalığında (HD) 10-35 olan CAG tekrarlarının 40'ı aştığı bilinmektedir. Normal FMR1'de 6 ve 35 GCC tekrarı bulunurken, hastalık durumunda 200 kez tekrar ettiği tespit edilmiştir (14).

### **2.1.c. Tek Nükleotit Polimorfizmi (TNP)**

#### **(Single Nucleotide Polimorfism = SNP)**

2000'li yılların başlarından itibaren TNP belirteç (dördüncü nesil) çalışmaları yoğunluk kazanmıştır. Genom boyunca her 200-300 bazçiftinde bir görülen ve sıklıkları sebebiyle genomda bulunan en informatif yapı olan TNP'ler popülasyon genetiği için, özellikle çeşitli hastalıklara ait bazı genlerin lokalizasyonlarının tespitinde etkin olarak kullanılmaktadır. Yüksek işlem hacimli TNP tipleme platformlarının geliştirilmesi ve elde edilen bilgilerin

Uluslararası TNP veri tabanlarına aktarılması ile TNP'lerin yaygın hastalıklarla, ilaçlara verilen cevaplarla ilişkilendirme çalışmaları ivme kazanmıştır (15). 2008 yılında (dbSNPBuild 129) kataloglanan TNP sayısı 11 milyon iken 2012 yılında (dbSNP Build 137)~54 milyon olmuştur (16).

#### **2.1.d. Kopya Sayısı Varyasyonu (KSV) (Copy Number Variation = CNV)**

İGP ile insan genomunun dizilenmesinin tamamlanmasıyla normal bireylerin genom topografyalarında yapısal farklılıklar olduğu ortaya konmuştur. Genomda (DNA) popülasyonlar arasında 0-13 gen kopyası içeren alleler rapor edilerek “Kopya Sayısı Varyantları” (KSV) olarak tanımlanmıştır. Delesyon, insersiyon, inversiyon, duplikasyon veya kompleks rekombinasyonlar sonucunda bireylerarasında bir kilobazdan birkaç megabaza kadar değişken bölgeler (segment) KSV'larolarak tanımlanmıştır (17).

Bu kopya sayısı varyantlarını tanımlamak için başlatılan İnsan Kopya Sayısı Varyasyonu Projesi ile insan genomunun yaklaşık %12'sinin KSV olduğu ve bu KSV'larının hastalıklara neden olduğu düşünülmüştür (18).

Bazı çalışmalarda insan genomunda KSV'ların öropsikiyatrik, bağışıklık, enfeksiyon ve kardiyovasküler gibi yaygın hastalıklarla ilişkilendirilmiş (19) diğer çalışmalarda ise KSV'larının yaygın hastalıklarla ilişkisi doğrulanamamıştır (20). Yaygın hastalıklardaki KSV'larının patogenezi henüz tartışmalı olsa da, bazı farmakogenetik genlerin ilaç etkileşiminde ve toksisitede rol oynadığı bilinmektedir.

### **3. Genom Proje Veritabanları**

1953 yılında James D. Watson ve Francis Crick'in DNA'nın yapısını çözmesinden sonra insanoğlu bu yapıyı oluşturan alfabenin şifreli dizilimlerinden meydana gelen kelime ve deyimleri çözmeye yönelmiştir. 1985'den sonra konuşulmaya başlanan İGP ile ortaya çıkacak veri yığınının bilgisayar ortamına taşınmasının önemini farkederek ABD Sağlık Bakanlığı (National Institute of Health-NIH) biyoteknolojik veri tabanlarının tutulması için 1988 yılında Ulusal Biyoteknoloji Veri Bankasını (National Center for Biotechnology Information-NCBI) kurmuştur. NCBI'nin bünyesinde Tek Nükleotit Polimorfizmleri ve diğer varyasyonların tutulduğu dbSNP (Database of Short Genetic Variations-dbSNP) veri tabanı, büyük ölçekli

varyasyonların kataloglandığı dbVAR (Database of Genomic Structural Variation-dbVar), genotip ve fenotip ilişkilerinin tutulduğu dbGAP (Database of Genotypes and Phenotypes-dbGaP) gibi varyasyon veritabanları bulunmaktadır (Tablo 2).

Genom verilerinin rafine edilerek birarada tutulduğu en önemli referans veritabanlarından biri olan Ensembl, EMBL–EBI (The European Molecular Biology Laboratory-The European Bioinformatics Institute) ve Wellcome Trust Sanger Enstitüsünün ortak bir projesidir. Ensembl (EBI) geniş kapsamlı bir veri tabanıdır. Organizmaların genetik özelliklerinin yanı sıra, birçok uygulamayı da içinde barındırmaktadır. Bu veri tabanında ileri seviyedeki kullanıcılara veri tabanı üzerinden kendi öngördükleri parametrelerle araştırma yapabilmeye olanağı da sunulmaktadır. Bu çerçevede kullanıcıların kendi özel betimlemelerini genom üstüne eklemesi mümkündür (Tablo 2).

University of California, Santa Cruz'daki (UCSC) insan genomu ve diğer birçok organizma genomu için resmi, referans ve taslak DNA dizilerini içermektedir. Araştırmacılar bu sayfayı bilinen gen dizilerine, tahmini gen dizilerine, ekspresyonları ile ilgili bilgilere, türler arası karşılaştırmalı bilgiye, tek nükleotit varyasyonlarına ve daha birçok bilgiye erişebilmek amacıyla kullanmaktadırlar. UCSC ayrıca yeni başlayan araştırmacılara markör dizi aranmasına, belirli bir bölge veya tüm genom hakkında açıklama elde edilmesine olanak sağlamaktadır. Aynı zamanda ENCODE ve Neandertal projelerine bağlantı sağlamaktadır (Tablo 2).

**Tablo 2.** Genomik diziler için Genom Tarayıcıları

<b>VERİTABANLARI</b>	<b>YAPISAL</b>	<b>NCBI</b>	<b>Databases (Örnek Veritabanları)</b>	<p><b>Database of Genomic Structural Variation (dbVar)</b> <a href="http://www.ncbi.nlm.nih.gov/dbvar">http://www.ncbi.nlm.nih.gov/dbvar</a> Bakınız-Genom Projesinden Doğan Yeni Projeler</p> <p><b>Database of Genotypes and Phenotypes (dbGaP)</b> <a href="http://www.ncbi.nlm.nih.gov/gap">http://www.ncbi.nlm.nih.gov/gap</a> Bakınız-Genom Projesinden Doğan Yeni Projeler</p> <p><b>Database of Short Genetic Variations (dbSNP)</b> <a href="http://www.ncbi.nlm.nih.gov/snp">http://www.ncbi.nlm.nih.gov/snp</a> Bakınız-Genom Projesinden Doğan Yeni Projeler</p> <p><b>Genome</b> <a href="http://www.ncbi.nlm.nih.gov/genome">http://www.ncbi.nlm.nih.gov/genome</a> 1000'in üzerinde organizmanın tüm genom verisini ve dizisini içermektedir. Hem dizilenmesi tamamlanmış hemde dizilenmesi devam eden organizmaların genomlarını temsil etmektedir. Yaşamın üç domainininin yanısıra (bakteri, arke ve ökaryotlar) birçok virüs, faj, viroidler, plasmidler ve organelleri temsil etmektedir.</p> <p><b>GenBank</b> <a href="http://www.ncbi.nlm.nih.gov/genbank/">http://www.ncbi.nlm.nih.gov/genbank/</a> Sağlık Bakanlığı (National Institute of Health (NIH)) genetik dizi veri tabanı kamuya açık olan DNA dizilerinin anote edilmiş bir koleksiyonudur. Genbank Japon Veritabanı (DDBJ), Avrupa Moleküler Biyoloji Laboratuvarı (EMBL) ve NCBI Genbank'dan oluşan Uluslararası Nükleotid Dizi veri tabanı İşbirliğinin bir parçasıdır. Bu üç organizasyon günlük olarak veri değişimi yapmaktadır. Genbank çoğunluğuna Nükleotid veri tabanı aracılığı ile ulaşılabilen birçok bölüm içermektedir. Expressed Sequence Tags (EST), Genome Survey Sequences (GSS) bölümlerine Nükleotid EST ve Nükleotid GSS veritabanları aracılığı ile erişilebilmektedir.</p> <p><b>Online Mendelian Inheritance in Man (OMIM)</b> <a href="http://www.ncbi.nlm.nih.gov/omim">http://www.ncbi.nlm.nih.gov/omim</a> İnsan gen ve genetik bozuklukları veri tabanı. NCBI; içerik desteğinin yanısıra, arama motorları ve farklı veritabanları ile de entegrasyon desteğinde sağlamaktadır. Fakat OMIM'in artık omim.org adında yeni bir adresi bulunmaktadır. Tüm kayıtları görebilmek için kullanıcı bu adrese yönlendirilmektedir.</p>
			<b>Downloads (Örnek İndirme Dosyaları)</b>	<p><b>BLAST (Veri tabanından bağımsız kullanılabilir)</b> <a href="http://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&amp;PAGE_TYPE=BlastDocs&amp;DOC_TYPE=Download">http://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&amp;PAGE_TYPE=BlastDocs&amp;DOC_TYPE=Download</a> Solaris, LINUX, Windows, ve MacOSX sistemlerde Lokal kullanım için BLAST çalıştırılabilirler (executable). Nükleotid, protein BLAST ve transle aramaların (translated searches) indirilmesi db altklasörü altında mevcuttur.</p> <p><b>FTP: BLAST Databases</b> <a href="ftp://ftp.ncbi.nlm.nih.gov/blast/db/">ftp://ftp.ncbi.nlm.nih.gov/blast/db/</a> Stand-alone BLAST programları ile kullanılması için dizi veritabanları. Bu klasördeki önceden formatlanmış veritabanlarıdır ve BLAST ile kullanılmak için hazırlanmıştır.</p> <p><b>FTP: SNP</b> <a href="ftp://ftp.ncbi.nlm.nih.gov/snp/">ftp://ftp.ncbi.nlm.nih.gov/snp/</a> İndirilebilir SNP verisi</p> <p><b>FTP: Site</b> <a href="http://www.ncbi.nlm.nih.gov/Ftp/">http://www.ncbi.nlm.nih.gov/Ftp/</a> NCBI Veritabanları, araçları ve yardımcıları için ftp indirme sitesi</p>





	FONKSİYONEL	ENCODE	İnsan genomunda kodlanan fonksiyonel elemanların tamamının (transkripsiyon, transkripsiyon faktörü bağlanma bölgeleri, kromatin yapısı, histon modifikasyonları) ortaya konmasını amaçlamıştır.
		EPIGENOM	İnsan genomunda değişken metilasyon pozisyonlarını (Metilation Variable Positions (MVPs)) kataloglamayı amaçlamıştır.

**Tablo 3.** Genom Varyasyon Veritabanları

GENOM VARYASYON VERİTABANLARI	NCBI	dbVAR- Büyük insersiyonlar, delesyonlar, translokasyonlar, ve insersiyonlar da dahil olmak üzere büyük ölçekli genomik varyasyonların kataloglandığı veri tabanıdır. dbVAR aynı zamanda tanımlanan varyantların fenotip bilgileri ile ilişkilerini tutar.
		dbGAP – Genotip ve Fenotip ve fenotip ilişkilerinin interaksyonunu araştıran çalışmaların sonuçlarını tutar. Genom boyunca çalışmalar, tıbbi amaçlı dizileme moleküler teşhis için yapılan analizlerin yanısıra, genotip ve klinik olmayan özellikler arasındaki ilişkileri tutar.
		dbSNP - TNP, mikrosatelitler, küçük ölçekli insersiyon ve delesyonları içerir. Popülasyona özgü frekans ve genotip verileri, deneysel şartlar, moleküler içerik ve hem nötral hem de klinik mutasyonlar için haritalama bilgilerini içermektedir.
	HapMap	İnsan genomunda Minör Allel Frekansı %5'in üzerinde olan (MAF> 0,05) <sup>2</sup> varyasyonların tamamının ortaya konulması kromozomlar üzerinde birbirleriyle ilişkili lokuslardaki allel kombinasyonlarının (Haplotip) ortaya konularak Haplotip haritalarının ortaya çıkartılması amaçlanmıştır
1000 Genom	Genotip ve fenotip ilişkilerinin araştırıldığı proje ile farklı popülasyonlar özgü düşük frekans ve nadir varyantların araştırılması amaçlanmıştır (MAF 0,5%-5% ve MAF<0,5%)	

<sup>2</sup>Minör Allel Frekansı (MAF) popülasyonda tanımlanan TNP'nin az yaygın allelin frekansına verilen addır. HapMap Projesinde MAF 0,05 ve üzeri seçilmiştir.1000 Genom Projesinde MAF 0,05 ve altı seçilmiştir.

#### 4. Genom Projesinden Doğan Yeni Projeler

Genom topoğrafyasının bileşenlerinden TNP'ler genom boyunca en sık görülen varyasyon türü olduklarından haritalanmaları 1998 yılından itibaren İGP'nin hedeflerinden birisi olmuştur (Tablo 3). Nisan 1999'da on büyük farmakogenomik şirketi ve U.K. Wellcome Trust Philanthropy Arthur L. Holden'in liderlik ettiği 300.000 ortak TNP'in haritalandığı bir konsorsiyum oluşturulduğunu ilan etmiştir (2012 itibarıyla ~54 milyon TNP- dbSNP Build 137). Tek gen hastalığına sebep olan geleneksel gen yakalama metotları ile kompleks hastalıkların çok azı yakalanabildiğinden, TNP haritaları kullanılarak TNP'ler arasındaki istatistiksel ilişkilerin çalışılması ve değerlendirilmesi kanser, diyabet gibi çoklu gen hastalıklarının tanımlanmasına katkı sağlayacağı düşünülmüştür.

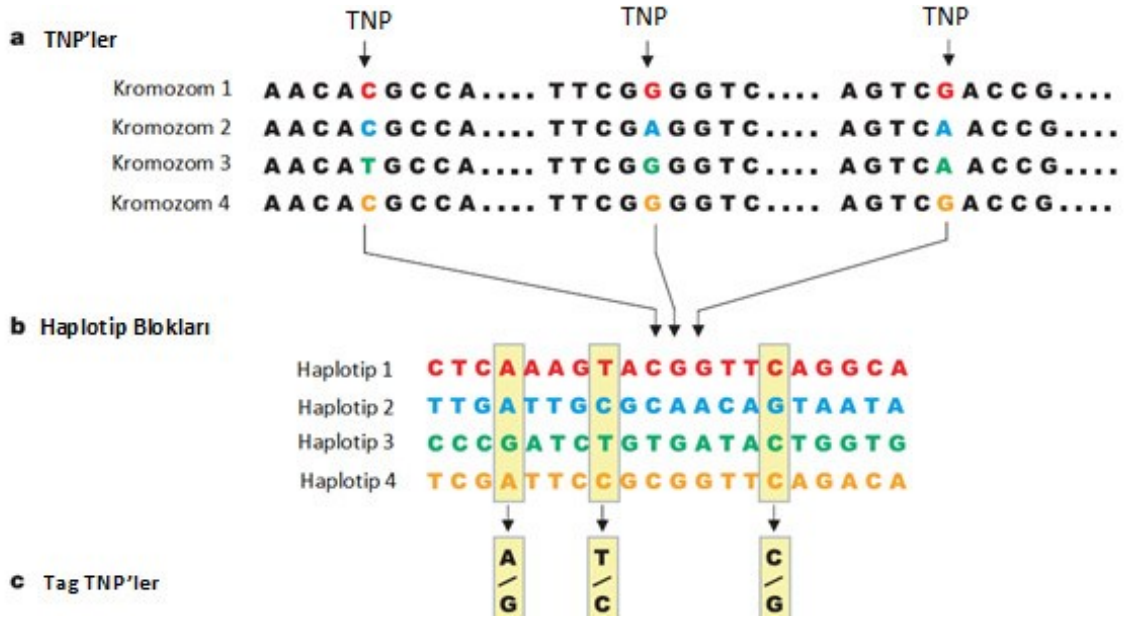
##### 4.1. HapMap

2003 yılında İGP ile insan genomunun tamamının dizilenmesi sonucunda genomun bireyler arası ancak %0,1'inin farklılık arz ettiği ortaya konmuştur. Sağlık, hastalık, ilaca cevap gibi konularda etkin olan ve çevresel faktörlerden etkilenen genlerin araştırmacılar tarafından belirlenebilmesi için 2002 Ekim ayında İngiltere, ABD, Kanada, Japonya, Nijerya ve Çin'den 200 bilim adamının çalıştığı, bir özel-kamu ortaklığı olan Uluslararası HapMap Konsorsiyumu "HapMap" oluşturulmuştur. Araştırmacılar için önemli bir kaynak olmayı hedefleyen HapMap Projesi ile yaygın DNA dizi varyasyonlarının modelinin çıkartılması suretiyle haplotip (Şekil 2) haritalarının ortaya konması amaçlanan proje üç aşamada (Tablo 4) yürütülmüştür (21-23).

HapMap projesi sonucunda ortaya konan veriler ilk olarak medikal genetik çalışmaların, analiz ve tasarlanmasına rehberlik etmek için üretilmiştir. Proje genom boyunca ilişkilendirme çalışmalarının tasarımı ve uygun analiz metodlarının geliştirilmesi için gerekli ana yapıyı oluşturmuştur. Ayrıca popülasyonların evrimsel ve tarihsel olarak geçirdikleri süreçlerin analizi için de önemli bir kaynak bilgiyi temsil eden HapMap projesi popülasyon genetikçileri için de büyük önem arz etmektedir.

Sonuç olarak HapMap Projesi ile (i) korelasyon ve frekans hesapları kullanılarak popülasyonlara özgü varyasyonlar ortaya konmuştur; (ii) Assosiyasyon çalışmaları ile

bağlantı (linkage) analizleri için gerekli tüm genom taramasına imkan veren araçlar sağlanmıştır.



- a. **TNP'ler:** 4 farklı insana ait aynı lokasyonda dört kromozomal bölgedeki TNP'ler. Dizinin çoğunluğu aynı olmasına rağmen varyasyonun olduğu 3 farklı baz gösterilmiştir. Her TNP 2 allele sahiptir. 1. TNP'te C ve T allelleri mevcuttur.
- b. **Haplotip Blokları:** Haplotip blokları yanyana TNP'lerin kombinasyonudur. Örnek olarak 20 TNP bloğu temsili olarak gösterilmektedir, panel a'daki 3 temsili TNP b'de işaretlenmiştir. Bu panelde popülasyon üyeleri genelde 1-4 haplotipini göstermektedir.
- c. **Tag TNP'ler:** Bu 20 TNP içerisinde 3 tanesi 4 Haplotipi benzersiz olarak tanımlamaktadır. Eğer bir kromozomda A-T-C deseni var ise bu örnek haplotip1 olarak saptanır. (The International HapMap Project, December 2003)

**Şekil 2.** TNP, Haplotip, Tag TNP

**Tablo 4.** HapMap Projesi Aşamaları (21-23).

	Faz 1 (21)	Faz 2 (22)	Faz 3 (23)
<b>Örnekler ve Popülasyon Çeşitleri</b>	269 örnek (4 grup <sup>27</sup> )	270 örnek (4 grup <sup>28</sup> )	1,115 örnek (11 grup <sup>29</sup> )
<b>Genotipleme merkezleri</b>	HapMap International Consortium	Perlegen <a href="http://genome.perlegen.com">http://genome.perlegen.com</a>	Broad & Sanger <a href="http://www.sanger.ac.uk/humgen/hapmap3/">http://www.sanger.ac.uk/humgen/hapmap3/</a>
<b>Genotiplenen TNP</b>	~1 M	~3.1 M (faz I+II)	1.6 M (Affy 6.0 & Illumina 1M)

## 4.2. 1000 Genom Projesi

İGP'nin 2003 yılında tamamlanmasından sonra 2004 yılında, sentez ile dizileme (sequencing by synthesis) esasına dayanan yeninesil dizileme teknolojisi sayesinde, dizilenen insan genomu sayısı artmaya başlamıştır. Yeni dizileme teknolojisi ile genom varyasyon boyutunun netleştirilebilmesi için büyük ölçekli genom dizilim verisine ihtiyaç olduğu anlaşılmıştır. Bu sebeple 1000 Genom Projesi olarak adlanan insan genetik varyasyonun düşük frekanslı nadir varyantların araştırılarak (MAF 0,5%-5% ve MAF<0,5%) detaylı bir kataloğunun oluşturulacağı yeni bir uluslararası işbirliği planlanmıştır (Tablo 5). Oluşturulacak katalog ile genom boyu ilişkilendirme çalışmaları ve diğer medikal araştırmaların desteklenmesi amaçlanmıştır (24).

1000 Genom Projesi'nin temel amacı farklı popülasyonlara özgü farklı DNA polimorfizm bilgisini sağlamak olmuştur. Bugünkü yüksek işlem hacimli dizileme teknolojileri sayesinde beş ana popülasyon grubunda (Avrupa, Doğu Asya, Batı Afrika ve Amerika kökenli popülasyonlar) özellikle genomik bölgede yer alan %95'in üzerinde ve allel frekansı %1 ve üzerinde olan (polimorfizm) varyantların karakterizasyonu çalışılmıştır.

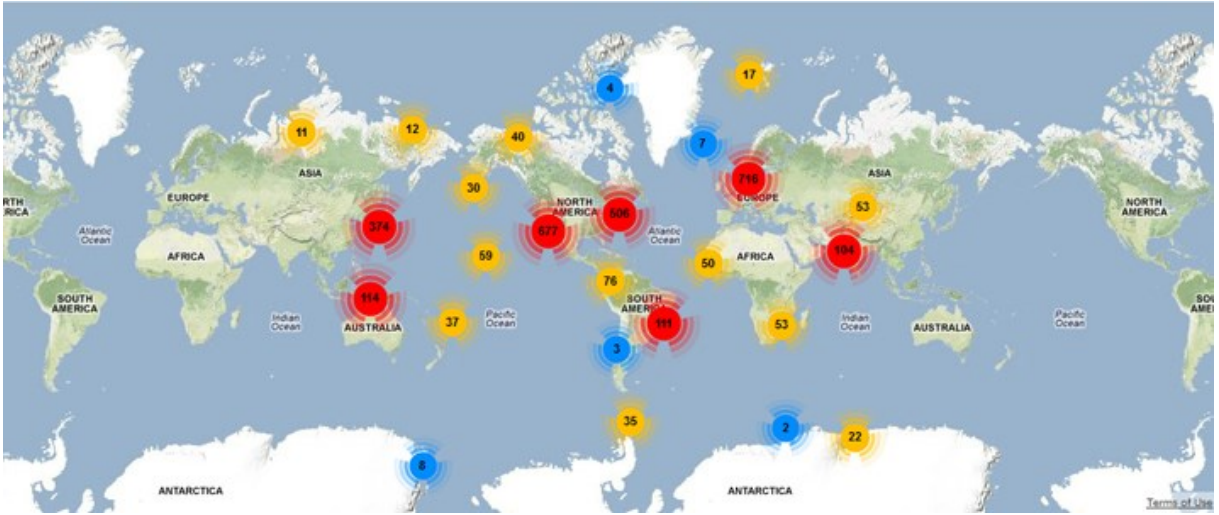
<sup>27</sup>**1. Faz:**(1) Nijerya, Ibida Yoruba'dan 90 birey 30 ana-baba-çocuk üçlüsü (2) USA, Utah'dan 90 birey (30 üçlü) [the Centre d'Etude du Polymorphisme Humain collection (CEU)'dan] (3) 45 Çin, Beijing'deki, Han Çinlileri (kısaltma CHB) (4) 44 Japonya, Tokyo'dan Japonlar (JPT)

<sup>28</sup>**2. Faz:**(1) Nijerya, Ibida Yoruba'dan 90 birey 30 ana-baba-çocuk üçlüsü (2) USA, Utah'dan 90 birey (30 üçlü) [the Centre d'Etude du Polymorphisme Humain collection (CEU)'dan] (3) 45 birbiriyle ilişkisiz Çin, Beijing'deki, Han Çinlileri (kısaltma CHB) (4) 45 birbiriyle ilişkisiz Japonya, Tokyo'dan Japonlar (JPT)

<sup>29</sup>**3. Faz:** (1)ASW (A): Güneybatı USA' de Afrika kökenliler 90 birey (2)CEU (C): CEPH koleksiyonundan Kuzey ve Batı Avrupa kökenli Utah sakinlerinden 180 birey (3)CHB (H): Çin, Beijing'deki Han Çinlilerinden 90 birey (4)CHD (D): Denver Colorado metropolitindeki Çinlilerden 100 birey (5)GIH (G): Teksas, Houston'da yaşayan Gujarati Hintlilerden 100 birey (6)JPT (J): Japonya, Tokyo Çinlilerden 91 birey (7)LWK (L): Webuye, Kenya'da yaşayan Luhya kabilesinden 100 birey (8)MEX (M): California, Los Angeles'da Meksika kökenli 90 birey (9) MKK (K): Kinyawa, Kenya'da yaşayan Maasai'lerden 180 birey (10)TSI (T): İtalya'daki Toskanalılardan 100 birey (11)YRI (Y): Nijerya, Ibida'da yaşayan Yorubalılarından (Batı Afrika) 180 birey

Ayrıca kodlama bölgelerinde % 0,1'in altında olan allel frekansları da kataloglanmıştır (24).

1000 Genom Projesi insan genetik varyasyonlarının coğrafi ve fonksiyonel spektrumunu betimleyerek genetiğin hastalara olan etkisini anlamamıza yardımcı olacak bir kaynak oluşturmayı hedeflemiştir. Projede 14 popülasyondan 1092 kişinin düşük kapsamlı bütün genom ve ekzom dizilemesinin kombinasyonundan oluşan genomlarını açığa çıkarılmıştır (25) (Şekil 3). Sonuçta bu projede çeşitli algoritmalar ve farklı veri kaynakları üzerinden bilgiyi entegre eden metodların geliştirilmesi suretiyle 38 milyon TNP, 1.4 milyon kısa indel ve 14.000'den fazla büyük delesyon için doğrulanarak, haplotip haritası ortaya çıkarılmıştır (19). Farklı popülasyonlardan olan bireylerin nadir ve sık görülen varyantlarda farklı profiller taşıdığını ve düşük frekanslı varyantların dikkate değer oranda (negatif seleksiyonun etkisi ile daha da artan) coğrafik farklılaşma gösterdiğini tespit edilmiştir. Evrimsel korunmanın ve kodlayıcı bölgede yer almanın negatif seleksiyonun gücünü belirleyen anahtar öğeler olduğu tespit edilmiştir. Bu çerçevede nadir varyant yükü biyolojik yollar arasında ciddi oranda değişim gösterirken, herbir birey transkripsiyon faktörünün bağlanma bölgesindeki motifleri bozan değişimler gibi kodlayıcı olmayan korunmuş bölgelerde yüzlerce varyant taşıdığı tespit edilmiştir (26).



Şekil 3. GOLDGenom Haritası (25).

### **4.3. Popülasyonlara Özgü Genom Projeleri**

#### **4.3.a.Japon Popülasyonuna Özgü Yaygın Genetik Varyasyon Veri tabanı (JSNP database)**

JSNP veri tabanı, Japon popülasyonuna ait TNP verilerini saklamak amacıyla 2000 yılında Japonya Başbakanının talimatı ile milenyum projesi olarak, İnsan Genom Merkezi (Human Genome Center (HGC)), Sağlık Bilimleri Enstitüsü (IMS) Tokyo Üniversitesi ve Japonya Fen ve Teknoloji (JST) işbirliği ile başlatılmıştır. Projenin amacı Japon popülasyonuna özgü gen bölgelerinde veya kodlama bölgelerini etkileyebilecek düzenleyici gen bölgelerindeki 150,000 TNP'lerin tanımlanması ile analitik araçların geliştirilmesi ve iki yıl içinde kullanıma açmak olarak belirlenmiştir (27). Burada polimorfizmler ile yaygın hastalıklar ve/veya ilaç reaksiyonları arasındaki ilişkiyi tanımlayabilecek temel veri setlerin oluşturulması hedeflenmiştir. Bu nedenle fenotipi etkileyen fakat hastalığa neden olmayan aday TNP'lerin tanımlanması için detaylı çalışılmıştır (28). 2002 yazında 190,562 genetik varyasyonun saklandığı, veri yönetim ve veri dağıtım kısımlarından oluşan web sitesi: <http://snp.ims.u-tokyo.ac.jp/> araştırmacıların kullanımına açılmıştır (29).

Günümüzde Japon TNP veri tabanı (JSNP database), geriyatrik araştırmalar için Japon TNP veri tabanı (JG-SNP), insan mitokondriyal genomu TNP veri tabanı (mtSNP) ve protein polimorfizm veri tabanını (dbprop) bir arada tutan bir Japon veri tabanı network internet sitesi de <http://snp.ims.u-tokyo.ac.jp/mission.html> olarak oluşturulmuştur.

#### **4.3.b.Pan-Asya TNP Genotip Veri tabanı (PanSNPdb)**

HapMap'te kullanılan veriyığının varolan bütün popülasyonları temsil etme olasılığının neredeyse yok denilecek kadar az olması varsayımından yola çıkarak Pasifik Pan-Asya girişimi adı altında araştırmacılar HapMap'den elde edilen verilerin kendi popülasyonlarını ne ölçüde temsil ettiğini araştıran bir grup oluşturulmuştur.

Pasifik Pan-Asya TNP Girişimi aracılığıyla Çin, Hindistan, Endonezya, Kore, Malezya, Nepal, Filipinler, Singapur, Tayland ve Tayvan merkezli enstitülerdeki bilim adamları tarafından 18 Kasım 2004 yılında resmîyet kazandırılan projenin, iki yıldan fazla süreceği, maliyetinin ise üç milyon dolar olacağı öngörülmüştür (30, 31).

HUGO Pan Asya TNP konsorsiyumu tarafından Asyalıların genetik varyasyonu ile ilgili bugüne kadar en fazla örnekle yapılmış olan bir çalışmadır. Çalışma sonucunda oluşturulan

PanSNPdb veri tabanı (<http://www4a.biotech.or.th/PASNP>) bu örnek verilerini ve bu verilere özgü çeşitli analizleri içermektedir. PanSNPdb LD modeli, haplotip dağılımı ve KSV'larını dairecek şekilde Asya insanının popülasyon yapısını ortaya koyan değerli bir kaynaktır. Buna ilaveten PanSNPdb genetik varyasyonlar açısından kapsamlı bir kaynak olmakla birlikte HapMap3, JSNP, dbSNP, DGV (Database of Genomic Variants) gibi TNP ve KSV veritabanları ile interaktif bir karşılaştırma sağlamaktadır (32).

#### **4.3.c. Çin İnsanı Genom Varyasyonu Projesi (CGHDP)**

Çin 2012 yılı sonu itibariyle 1,354 milyar insanı ile dünyanın en fazla nüfus yoğunluğuna sahip ülkesidir (33). Eylül 1998 “The Chinese Human Genome Diversity Project”de belirtildiği üzere 1998 yılı itibariyle yaklaşık 5,8 milyar olan dünya nüfusunun resmi olarak tanınan 56 alt popülasyon ile beşte birini Çin oluşturmaktadır. Araştırmacılar CGHD Projesi ile mikrosatellitleri kullanarak popülasyon tabakalanmasını ortaya koymaya çalışmışlardır. Çin’de yaşayan 56 alt popülasyonun en geniş olan Han grubu çoğunluğunun nüfusu bir milyardır (toplam nüfusun ~%90). 55 azınlık grup ise kalan 100 milyon kişi ile temsil edilmektedir.

Çin Genom Varyasyon Projesi kapsamında yapılan karşılaştırmalı araştırmalar önemli genetik farklılaşmaları işaret etmiştir. Coğrafi yakınlıktan dolayı farklı popülasyonlar arası fazla gen akışı sağlandığından Doğu Asya popülasyonları arasında beklendiği üzere küçük genetik farklılaşmalar görülmüştür. Bu grubun en yakın genetik komşuları ise Amerika yerlileri olarak saptanmıştır. Avustralya yerlileri ve Yeni Gine’liler tarafından oluşturulan küçük bir küme, genetik olarak biraz daha az yakın olanları oluşturmuştur (13).

#### **4.3.d. Asyalıların Diploid Genomlarının Dizilenmesi (Yanhuang Project)**

Yeni nesil dizileme teknolojisinin hayata geçmesiyle, ilk büyük hacimli tüm genom dizileme projesi olan, Yanhuang (YH) Projesi Çin’de başlatılmıştır. Beijing Genomik Enstitüsü tarafından 8 Ocak’da ilan edilen projede başlangıç olarak üç yıl içerisinde 100 Çinlinin tüm genomunun dizilenmesi, tamamlandıktan sonra popülasyon alt grupları dahil diğer Asya ülkelerinden binlerce insanın daha genomunun dizilenmesi ve daha uluslararası bir proje ile 1000 kişinin daha (1000 genom Projesi ve Multigenome Project) genomunun dizilenmesinin amaçlandığı duyurulmuştur (34).



11 Ekim 2007 tarihinde Shenzhen'deki Beijing Genomik Enstitüsü tarafından Asya popülasyonunun temsilcisi olan Han Çinlilerinin ilk kez diploid genom dizilenmesinin tamamlandığı duyurulmuştur (35).

#### **4.3.e. Hindistan Genom Varyasyon Veri tabanı (IGVDB)**

Nisan 2013 yılı itibariyle 1 milyar 234 bin olan nüfusuyla dünyanın en kalabalık ikinci ülkesidir (36). Hindistan popülasyonu kabile içi evlenen yüzlerce grupla birlikte 4693 komüniteden oluşan, 325 dilin ve 25 alfabenin kullanıldığı bir çeşitliliğe sahiptir. IGVDB, bu çeşitliliğin genom düzeyindeki etkilerini araştırmak ve bazı kompleks hastalıklara ilişkin sorulara farmakogenomik açılarından cevap bulabilmek için Hindistan Genom Çeşitlilik (IGV) Konsorsiyumu tarafından yapılan bir çalışmadır. IGV Projesi tahmini ilaç seçimi için Hindistan alt popülasyonuna ait gen tekrar bölgelerini ve TNP'leri kullanarak bilgilendirici belirteçler oluşturmayı amaçlamıştır (37).

Proje başlangıcında amaçlara uygun olarak farklı alt popülasyonlardan 15,000 örnek toplanarak, yaygın hastalıklar ve ilaca cevapla ilişkili yaklaşık 1000 gen seçilmiş ve gen başına yaklaşık 5-10 bilgilendirici belirteç belirlenmiştir (37).

Projenin amaçları arasında (i) yeni ve daha önceden saptanmış TNP ve mikrosatelit allel frekanslarının hesaplanması (ii) haplotiplerin oluşturulması (iii) Hindistan alt popülasyonları arasında, gen içinde ve genler arasındaki LD mesafesinin belirlenmesi bulunmaktadır. Projenin nihai hedefi ise Hindistan halkının DNA varyasyon veri tabanını oluşturarak hastalıklara yatkınlık, beklenmeyen yan etkiler ve popülasyon göçleri gibi durumlar için insan biyolojisinin anlaşılmasını sağlamak ve bu veri tabanını (<http://www.igvdb.res.in>) araştırmacıların hizmetine sunmaktır (37).

Hindistan popülasyonunun çeşitliliği göz önüne alınarak hedeflenen sonuçlar: (i) Tüm genetik varyasyonun yakalanabildiği popülasyon altyapısına ait kompozisyonu tanımlamak (ii) Hindistan popülasyonunun tamamını temsil eden TNP'leri temsilen küçük bir TNP paneli kompozisyonu belirlemektir (37).

#### **4.3.f. Singapur Genom Varyasyon Projesi: Üç Güneydoğu Asya Popülasyonunun Haplotip Haritası**

Singapur Genom Varyasyon Projesi (SGVP) Güneydoğu Asya'daki Çin, Malaya ve Hintli 268 örnekten dizilenen 1,6 milyon TNP veri tabanında toplanarak (<http://www.statgen.nus.edu.sg/cgi-bin/gbrowse/sgvp>) bilim insanlarının kullanımına sunulmuştur. Veri tabanı HapMap projesine benzer olarak allel frekansları, bağlantı dengesizliği (Linkage Disequilibrium (LD)) değerlendirmeleri ve rekombinasyon oranlarını içeren genotip ve haplotip veri bilgilerini ve özetlerini kataloglamaktadır (38).

#### **4.3.g. Estonya Genom Projesi**

Avrupa gen havuzu içerisinde oldukça küçük bir bölümü temsil eden Estonyalıların kendi ulusal gen havuzlarını oluşturma amacıyla başlatılan Estonya Genom Projesi (EGP), Estonya hükümeti ve Estonya Genom Kurumunun (EGF) anlaşması sonucunda 1999 Mart ayında planlanmaya başlanmıştır. Başarıya ulaşmanın ana unsurlarından biri olan kamu-özel sektör birlikteliği EGF ve US kökenli EGeen arasında kurulmuştur. Proje kapsamında 2000 yılında 1,4 milyon olan toplam nüfusun %75'i olan 1 milyon kişinin genotiplenmesi amaçlanmıştır.

Proje ile Estonya popülasyonuna ait fenotip-genotip verilerini içeren veri tabanının (<http://www.geenivaramu.ee/en>) EGP adı altında kurularak genetik, sağlık ve yaygın hastalıkların çalışılması için araştırmacıların kullanımına sunulması hedeflenmiştir. Ayrıca Estonya popülasyonunun sağlık taramasının büyük ölçüde tanımlanması, doku örneklerin toplanması, LD haritalarının çıkartılması, veri erişiminin sağlanacağı ve pazara sunulacak genom projesi ürünlerine ait yazılımların geliştirilmesi, büyük ölçekte genomun toplum sağlığına getireceği sistematik avantajların pratikte uygulanabilmesi de proje amaçları arasında ele alınmıştır.

Yapılan araştırmalar Estonya popülasyonunun tüm Avrupa popülasyonunun iyi bir temsilcisi olduğunu göstermiştir. Araştırmalar ile Estonya ve diğerlerinin (Caucasians) genotip verileri ve 22. kromozoma ait LD haritası Estonya'lılarla Avrupa popülasyonları arasında oldukça küçük farklılıklara işaret etmiştir (39).

Yukarıda bahsi geçen genom projeleri dışında İranlıların genetik geçmişini araştırmak üzere başlatılan ve halen devam eden İran Genom Projesi (40), Afrikalı-Amerikalılar Vakfı

tarafından desteklenen Afrikalı-Amerikalıların Afrika kalıtım miraslarının araştırıldığı Afrika Genom Projesi (41) ve Türkiye Genom Projesi (42) gibi genom projeleri de bulunmaktadır.

## SONUÇ

İnsan Genom Projesi 1990 yılında başlatıldığında dünya kamuoyunda projenin bitiminde ortaya çıkarılacak bilgi ile yaşamın şifresinin çözülmüş olacağı ve bunun sonucunda da birçok hastalığın çaresinin bulunmuş olacağı beklentisi oluştu. Proje genel hatları ile 2003 yılında bitmiş olmasına rağmen bu beklentiye yaklaşıldığını söylemek mümkün değildir. Bugün yeni ileri teknolojilerle daha kapsamlı, daha büyük, daha farklı içerikteki genom projeleri (Epigenom, 1000 genom, Encode vb.) yürütülmektedir. Bu durum bütün araştırmaların boşuna yapıldığı yanılgısını ortaya çıkarmamalıdır. Zira aslında bunun sırrı henüz biyolojinin kanunlarının bütünüyle yazılmamış olmasında saklıdır.

Yeryüzü şartlarında temel kanunları keşfedilmiş olan matematik, fizik ve kimya bilimlerine karşılık biyoloji halen birçok bilinmezi içermektedir. Zira elde biyolojiye ait ne bütün modelleri oluşturacak veri setleri ne de bu çerçevede bahsi geçen modellerin hesaplanabilir kuralları mevcuttur. İşte bu nedenledir ki yalnızca anonim birkaç insanın genom dizisini deşifre etmiş olan İGP tek başına yukarıda ifade edilen beklentiye karşılayamamıştır. Esas itibarıyla bu sonuç yaşam bilimciler için bir sürpriz olmamıştır. Onlar İGP'nin bitişini takiben daha sistematik, daha yüksek rezolüsyonlu, daha büyük kapsamlı veri üretecek teknolojilere ihtiyaç duyacaklarını öngörmüşlerdi. Nitekim 2000'li yıllardan itibaren hem genetik teknolojiler hem de ortaya çıkan dev veri yığını depolayabilen ve hesaplayabilen bilişim araç ve teknolojileri ivmesel bir hızla gelişti.

Genom projeleri ile üretilen verilerden yaşambilim modellerinin oluşturulabilmesinin ilk aşaması, bu verilerin uygun veritabanlarında saklanması ve kategorize edilmeleri idi. Bu amaçla Genbank (genome, dbSNP, dbGaP vb.), HapMap, 1000 genomes, Encode veritabanları oluşturulmuştur. Genom projeleri ve münferit genetik araştırmalar sonucunda açığa çıkan popülasyonlar arası genetik çeşitliliğin insanoğlunun yaşam kalitesi üzerindeki olası etkilerinin derinliğinin araştırılması amacı ile bu genel veritabanlarının dışında popülasyona özgü veritabanları da oluşturulmuştur.

Bütün bu veriler ve veritabanları analizlerinin bir hedefi de günümüzde “kişisel tıp” olarak isimlendirilen ve tüm sağlık sorunlarını bireysel düzeyde ve en etkin yolla çözmeyi amaçlayan analiz yöntemlerinin geliştirilmesidir. Bu doğrultuda yapılacak olan araştırmalar, hasta olan kişinin en kısa sürede, en etkin yöntemlerle tedavi edilmesini, hastalığa yakalanmadan, gelecekte kendisinde oluşabilecek hastalıkların saptanmasına, bu hastalıklara karşı “kişisel koruyucu tıbbın” geliştirilmesine olanak sağlayabilecektir.

Dünyada bölge bazlı nadir (%0.5-%5 ve <0.5) varyantlar çalışılmaya devam edilirken Türkiye’de 2010 yılında Boğaziçi Üniversitesi Moleküler Biyoloji ve Genetik Bölümü araştırmacılarının önderliğinde genom boyunca dizileme ve biyoinformatik analizlerini içeren bir proje başlatılmıştır. Boğaziçi Üniversitesi Türkiye Genom Araştırması tarafından sürdürülen bu proje kapsamında 16 örneğin genomunun tamamının yeni nesil DNA dizileme yöntemi ile belirlenmesi, TNP, KSV ve yapısal çeşitliliklerin saptanmasına çalışılmaktadır. Uluslararası çalışmalar ile kıyaslandığında görünen odur ki 16 kişinin genomlarının dizilenmesi ancak önemli bir başlangıç olarak kabul edilebilir.

Tüm bu sebeplerden dolayı öncelikle yaygın varyantların tanımlanabilmesi ve tüm genom dizilemesi tamamlanan 16 kişi ile karşılaştırılabilmesi için bir yandan bu sayının artırılması diğer yandan da Türkiye’de bugüne kadar analiz edilmiş yüksek işlem hacimli örneklerin veritabanlarında biraraya getirilerek Türk popülasyonuna özgü TNP haritalarının oluşturulabilmesi büyük önem arz etmektedir. Bu tip çalışmalarda, istatistikçilerin deney tasarımı, uygun istatistiksel yöntemlerin seçilmesine veya yeni yöntemlerin geliştirilmesine, veri analizlerinden, sonuçların yorumlanmasına kadar bir çok konuda projeye dahil olması da büyük önem taşımaktadır. Örneğin örneklem sayısının, anakütle ve örneklemin özelliklerinin, araştırma hipotezine uygun grupların belirlenmesi vb. konularda çalışmanın başından itibaren istatistikçilerin fikri alınmalıdır.

Küreselleşen dünyada bireyselleşen tedavi yöntemleri için yapılan çalışmalarda ilk olarak farklı genom projeleri ile genom boyunca yaygın varyantların tanımlanmasına, günümüzde ise nadir varyantların ortaya konmasına dair çalışmalar yoğun bir şekilde devam etmektedir. Dünya üzerinde bilgi ve teknoloji sahibi hemen her popülasyon kendi varyasyon haritalarını ABD ve/veya Avrupa merkezli olarak yapılan çalışmalarla kıyaslayarak genomik bilginin globalliğini araştırmaktadır. Aynı zamanda bu çalışmalar ile kendi popülasyonlarının ne ölçüde temsil edildiğini de araştırmaktadır. Bu çalışmalar sonucunda ortaya konan hastalık genotip ilişkileri sayesinde klinik olarak ortaya konan

tedavi yöntemlerinin seçilimi, ilaç doz ayarlamaları ile ilgili çalışmalar yürütülebilmektedir. Sonuç olarak, bu yeni bilgidен ülkemizin de faydalanabilmesi için yeterli örnek grupları ile çalışarak tüm genom dizileme çalışmalarının yürütülmesi, bugüne dek münferit genetik araştırmalar kapsamında tanımlanmış mutasyon ve varyantların bir veri tabanında bir araya getirilmesi, yaygın varyantların ve daha sonraki çalışmalar ile nadir varyantların tanımlanması fenotip genotip ilişkilerinin ortaya konması açısından önemli ve gereklidir.

## KAYNAKLAR

1. <http://www.genome.gov/10001772>
2. <http://www.gpo.gov/fdsys/pkg/WCPD-2003-04-21/pdf/WCPD-2003-04-21-Pg440.pdf>
3. <http://www.phgfoundation.org/news/8538/>
4. Kidd JM, Cooper GM, Donahue WF, Hayden HS, Sampas N, Graves T, et al. Mapping and sequencing of structural variation from eight human genomes. *Nature* 2008; 453: 56-64.
5. Cole CG, McCann OT, Collins JE, Oliver K, Willey D, Gribble SM, et al. Finishing the finished human chromosome 22 sequence. *Genome Biology* 2008; 9(5): 78.
6. Akarsu N, Lüleci G. Gene Mapping: How are genes mapped, What do these maps contain, How are they interpreted. *DEU Tıp Fakültesi Dergisi* 2002; 29-39.
7. Akarsu N. Alternative Approaches In Pediatric Ophthalmology. *Türkiye Klinikleri J Pediatr Sci* 2005; 1(6): 70-6.
8. [http://www.ornl.gov/sci/techresources/Human\\_Genome/faq/snps.shtml#goals](http://www.ornl.gov/sci/techresources/Human_Genome/faq/snps.shtml#goals)
9. <http://www.nia.nih.gov/alzheimers/publication/alzheimers-disease-genetics-fact-sheet>
10. White TJ, Bruns TD, Lee S, Taylor J. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis MA, Gelfand DH, Sninsky JJ, White eds. *PCR Protocols*, Academic Press, San Diego, 1990: 315-22.
11. Botstein D, White RL, Skolnick M, Davis R. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Genetics*, 1980; 32(3): 314-31.
12. Nakamura Y, Lathrop M, Bragg T, Jones C, O'Connell P., Leppert M. An extended linkage map for human chromosome 10. *Genomics* 1988; 3(4): 389-92.
13. Sforza C, Luca L. The Chinese Human Genome Diversity Project. *Proc. Natl. Acad. Sci. USA* 1998; 95(20): 11501-503.
14. Klug WS, Cummings SMR. Kromozom Yapısı ve DNA Dizisinin Organizasyonu. In: Öner C. *Genetik Kavramlar*, 6. Baskı. Ankara: Palme Yayıncılık, 2003: 544-49.
15. Nakamura Y. DNA variations in human and medical genetics: 25 years of my experience. *Journal of Human Genetics* 2009; 54(1): 1-8.
16. <http://www.ncbi.nlm.nih.gov/mailman/pipermail/dbsnpannounce/2012q2/000122.html>
17. Weinshilboum R. Inheritance and Drug Response. *N Engl J Med*. 2003; 348: 529-37.
18. He Y, Hoskins J, McLeod HL. Copy Number Variants in pharmacogenetic genes. *Trends Mol Medicine* 2011; 17(5): 244-51.

19. McCarroll SA, Hadnott TN, Perry GH, Sabeti PC, Zody MC, Barrett JC, et al. Common deletion polymorphisms in the human genome. *Nature Genetics* 2006; 38: 86–92.
20. The Wellcome Trust Case Control Consortium. Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature* 2010; 464: 713–20.
21. The International HapMap Consortium. A haplotype map of the human genome. *Nature* 2005; 437: 1299–1320.
22. The International HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. *Nature* 2007; 449: 851–61.
23. The International HapMap 3 Consortium. Integrating common and rare genetic variation in diverse human populations. *Nature* 2010, 467: 52–58.
24. The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* 2010; 467: 1061–73.
25. [http://www.genomesonline.org/cgi-bin/GOLD/index.cgi?page\\_requested=GenomeMap](http://www.genomesonline.org/cgi-bin/GOLD/index.cgi?page_requested=GenomeMap)
26. The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012; 491: 56–65.
27. Hirakawa M, Tanaka T, Hashimoto Y, Kuroda M, Takagi T, Nakamura Y. JSNP: a database of common gene variations in the Japanese population. *Nucleic Acids Research* 2002; 30(1): 158–62.
28. Mah J. T, Chia K. S. A Gentle Introduction TO SNP Analysis. *Journal of Bioinformatics and Computational Biology* 2007; 5(5): 1123–38.
29. Hirakawa M, Tanaka T, Hashimoto Y, Kuroda M, Takagi T, Nakamura Y. JSNP: a database of common gene variations in the Japanese population. *Nucleic Acids Res.* 2002; 30(1): 158–62.
30. Normile D. Consortium hopes to map human history in Asia. *Science* 2004; 306: 1667.
31. Cyranoski D. Genomics takes hold in Asia. *Nature* 2008; 456: 12.
32. Ngamphiw C, Assawamakin A, Xu S, Shaw P. J, Yang J. O, Ghang H, et al. PanSNPdb: The Pan-Asian SNP Genotyping Database. *PLoS ONE* 2011; 6(6): 1–7.
33. <http://www.trthaber.com/haber/dunya/iste-cinin-nufusu-71334.html>
34. Qiu J, Hayden EC. Genomics sizes up. *Nature* 2008; 451: 234.
35. <http://yh.genomics.org.cn/>
36. <http://www.indiastat.com/default.aspx>
37. The Indian Genome Variation Consortium. The Indian Genome Variation database (IGVdb): a project overview. *Human Genetics* 2005; 118: 1–11.
38. Teo YY, Sim X, Ong RT, Tan AK, Chen J, Tantoso E, Small KS, et al. Singapore Genome Variation Project: A haplotype map of three Southeast Asian populations. *Genome Res.* 2009; 19(11): 2154–62.
39. [http://www.biopark.ee/documents/Estonian\\_genome\\_project.pdf](http://www.biopark.ee/documents/Estonian_genome_project.pdf)
40. <http://irangen.com/how-to-participate>
41. <http://dubois.fas.harvard.edu/african-genome-project>
42. <http://turkiyegenomprojesi.boun.edu.tr>