

Genom projeleri 5N1H: ne, nerede, ne zaman, nasıl, neden ve hangi popülasyonda?

Genome projects 5W1H: what, where, when, why, how and in which population?

Pelin FİDANOĞLU¹, Nevin BELDER¹, Beyza ERDOĞAN², Özlem İLK³, Farid RAJABLI⁴, Hilal ÖZDAĞ¹

ÖZET

Genom projeleri yaşamın şifresi olarak tanımlanabilecek olan ve bir organizmanın genomunu oluşturan DNA 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 kimliği 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 Tekli 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 iki aylık bir süre içinde 1 milyon dolarlık bir bütçe ile dizilenmiştir. 2004 yılından bugüne yeni nesil dizileme teknolojisindeki gelişmeler ile

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 dollars. 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 6.600 USD. In this review the HGP which created big expectations especially in medicine will be explained from its start to the present. Then we will

¹ Ankara Üniversitesi, Biyoteknoloji Enstitüsü, Merkez Laboratuvarı, ANKARA

² Ankara Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, ANKARA

³ Orta Doğu Teknik Üniversitesi, İstatistik Bölümü, ANKARA

⁴ Turgut Özal Üniversitesi, Elektrik-Elektronik Mühendisliği Bölümü, ANKARA



İletişim / Corresponding Author : Hilal ÖZDAĞ

Ankara Üniversitesi, Biyoteknoloji Enstitüsü, Merkez Laboratuvarı, ANKARA

Tel : +90 312 222 58 26

E-posta / E-mail : hilalozdag@gmail.com

Geliş Tarihi / Received : 07.08.2013

Kabul Tarihi / Accepted : 19.09.2013

DOI ID : 10.5505/TurkHijyen.2014.14890

Fidanoğlu P, Belder N, Erdoğan B, İlk Ö, Rajabli F, Özdağ H. Genom projeleri 5N1H: ne, nerede, ne zaman, nasıl, neden ve hangi popülasyonda ? Turk Hij Den Biyol Derg, 2014; 71(1): 45-60.

insan genomunun dizilenme süresi bir güne ve maliyeti ise 6.600 dolara inmiştir. Bu derlemede 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: Haplotip, İGP, TNP, Varyasyon

summarize the studies paving the road to personalized medicine emphasizing the fact that to reveal the meaning of genomic information, it should become computable.

Key Words: Haplotype, HGP, SNP, Variation

GİRİŞ

1. İNSAN GENOM PROJESİ

Gregor Mendel'in bezelye bitkisi üzerinde yaptığı çalışmalar sonucunda kalıtımın kurallarını keşfetmesi ile başlayan bir çağ, kalıtımın doğasını bütünüyle anlayabilmek için başlatılan İnsan Genom Projesi (İGP) ile başka bir bilimsel boyut kazanmıştır. Böylece 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, James D. Watson ve Francis Crick'in DNA'nın çift sarmal yapısını keşfetmesinin 50. yılına denk gelmiştir. DNA temel baz dizisinin elde edilmesi, sadece onun başlangıcına işaret etmiştir (2).

3,8 milyar dolarlık bir bütçe ile 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 konulmuştur. 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 (Tablo 1) (3).

Tablo 1. İnsan genom projesine katkıda bulunan merkez ve ülkeler (3)

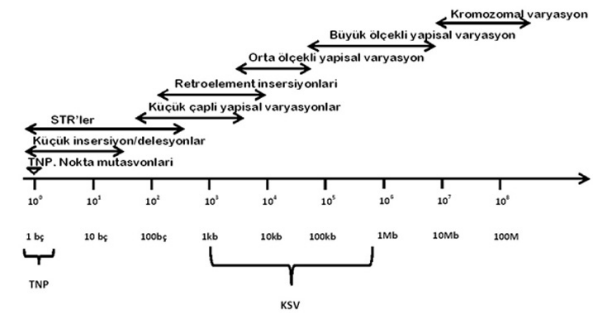
Merkez	Ülke
The Whitehead Institute/MIT Center for Genome Research	ABD
The Wellcome Trust Sanger Institute	İngiltere
Washington University School of Medicine Genome Sequencing Center	ABD
United States DOE Joint Genome Institute	ABD
Baylor College of Medicine Human Genome Sequencing Center	ABD
RIKEN Genomic 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 TOPOĞRAFYASININ BİLEŞENLERİ VE GENOMUN DİNAMİKLERİ

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 (Restriction Fragment Length Polymorphism-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 (Tekli Nükleotit Polimorfizmleri-TNP, dördüncü nesil) ile kopya sayısı çeşitliliği (Copy Number Variation-CNV) çalışmaları yoğunluk kazanmıştır (Şekil 1) (4, 5).



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 topografyasının 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 allele göre popülasyonda nadir 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 gelmemekte ve bu hastalığın gelişimi için kesin bir gösterge olarak ifade edilememektedir. Zira iki E4 alleleline sahip olan bireylerde hiçbir zaman Alzheimer gelişmediği, buna karşılık iki E2 alleleline sahip olan bireylerde ise Alzheimer hastalığının ortaya çıktığı durumlara da rastlanabilmektedir (9). Alzheimer, obezite, kanser, kardiyovasküler hastalıklar gibi kalıtsal özellik gösterebilmekle beraber, mendelyen kalıtım modelinin izlenmediği multifaktöryel ve multigenik karakter arz etmektedir. 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 kaldıradığı 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 bulunmaktadır. Kodlamayan bölgede, değişik motif ve uzunluklarda (2-10 baz çift) ardışık, orta sıklıkta tekrarlayan nükleotit dizileri mikrosatelit olarak adlandırılmaktadır. Genom boyunca dağılmış durumda olan mikrosatellitlerin bir DNA bölgesindeki tekrar sayısı bireyler arasında farklılık göstermektedir. Örneğin, insanlarda en yaygın mikrosatelit ikili nükleotit "CA"_n tekrarlarıdır ve tekrar sayısı genellikle 5 ve 50 arasında değişmektedir. Mikrosatellitlerin allel sayıları bu bölgelerin mutasyon oranları yüksek olduğu için çoğunlukla biallelik olan diğer yaygın polimorfizm çeşitlerinden üç kat daha fazla bilgilendirici olmaktadır. Bu nedenle mikrosatellitler genetik bağlantı ve evrimsel analizlerde sıklıkla kullanılmıştır (13).

1985'ler de Polimeraz zincir 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, mikrosatellitler bağlantı analizlerinde ve popülasyon genetiğinde başarılı bir şekilde kullanılmaya başlanmıştır. Heterozigotluk oranları fazla olan mikrosatellitler, paternal alleler için büyük oranda ayırtedici ve bağlantı analizlerinde oldukça bilgilendiricidirler.

Mikrosatellitler 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 olduğu, diğer bazı organizmalarda da gen ifadesinde yer aldığı ve bazı durumlarda ise kodladığı protein üzerinde etkili olduğu gösterilmiştir. Örneğin, Huntington hastalığında (HD) 10-35 olan CAG tekrarlarının 40'ı aşığı bilinmektedir. Normal FMR1'de 6 ve 35 GCC tekrarı bulunurken, hastalık durumunda 200 kez tekrar ettiği tespit edilmiştir (14).

2.1.c. Tekli Nükleotit Polimorfizmi (TNP)

(Single Nucleotide Polymorphism = SNP)

2000'li yılların başlarından itibaren TNP belirteç (dördüncü nesil) çalışmaları yoğunluk kazanmıştır. Genom boyunca her 200-300 baz çiftinde 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 tiplene 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 bireyler arasında bir kilobazdan birkaç megabaza kadar değişken bölgeler (segment) KSV'lar olarak tanımlanmıştır (17).

Kopya sayısı varyantlarını tanımlamak için başlatılan “İnsan Kopya Sayısı Varyasyonu Projesi” ile insan genomunun yaklaşık %12'sinin KSV olduğu ve bu KSV'lerin hastalıklara neden olduğu düşünülmüştür (18).

Bazı çalışmalarda insan genomunda KSV'ler nöropsikiyatrik, bağışıklık, enfeksiyon ve kardiyovasküler gibi yaygın hastalıklarla ilişkilendirilmiş, diğer çalışmalarda ise KSV'lerin yaygın hastalıklarla ilişkisi doğrulanamamıştır (19, 20). Yaygın hastalıklardaki KSV'lerin patogenezi tartışmalı olsa da, bazı farmakogenetik genlerin ilaç etkileşiminde ve toksisitede rol oynadığı bilinmektedir.

3. GENOM PROJE VERİTABANLARI

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 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 veri tabanlarından 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 yapabileceği de sunulmaktadır. Bu çerçevede kullanıcıların kendi özel betimlemelerini genom üstüne eklemesi mümkündür (Tablo 2).

University of California, Santa Cruz (UCSC)'daki 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 markır dizi aranmasına, belirli bir bölge veya tüm genom hakkında açıklama elde edilmesine olanak sağlamaktadır. Aynı zamanda ENCODE ve Neandertal projelerine bağlantı sağlamaktadır (Tablo 2).

4. GENOM PROJESİNDEN DOĞAN YENİ 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, 10 büyük farmakogenomik şirketi ve U.K.

Tablo 2. Genomik diziler için genom tarayıcıları

VERİ TABANLARI YAPISIAL	Databases (Örnek Veritabanları)	Database of Genomic Structural Variation (dbVar) http://www.ncbi.nlm.nih.gov/dbvar Bakınız-Genom Projesinden Doğan Yeni Projeler
		Database of Genotypes and Phenotypes (dbGaP) http://www.ncbi.nlm.nih.gov/gap Bakınız-Genom Projesinden Doğan Yeni Projeler
		Database of Short Genetic Variations (dbSNP) http://www.ncbi.nlm.nih.gov/snp Bakınız-Genom Projesinden Doğan Yeni Projeler
		Genome http://www.ncbi.nlm.nih.gov/genome 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.
		GenBank http://www.ncbi.nlm.nih.gov/genbank/ 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 işbirliğinin 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.
	Online Mendelian Inheritance in Man (OMIM) http://www.ncbi.nlm.nih.gov/omim 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.	
	Downloads (Örnek İndirme Dosyaları)	BLAST (Veri tabanından bağımsız kullanılabilir) http://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastDocs&DOC_TYPE=Download 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.
		FTP: BLAST Databases ftp://ftp.ncbi.nlm.nih.gov/blast/db/ Stand-alone BLAST programları ile kullanılması için dizi veritabanları. Bu klasördeki önceden formatlanmış veritabanlarıdır ve BLAST ile kullanılmak için hazırlanmıştır.
		FTP: SNP ftp://ftp.ncbi.nlm.nih.gov/snp/ İndirilebilir TNP verisi
	Submissions (Örnek Gönderiler)	FTP: Site http://www.ncbi.nlm.nih.gov/Ftp/ NCBI Veritabanları, araçları ve yardımcıları için ftp indirme sitesi
TNP Gönderi Araçları, dbGaP Veri Gönderi Politikaları,...		
Tools (Örnek Araçlar)	TNP Veri tabanına Özgü Arama Araçları http://www.ncbi.nlm.nih.gov/projects/SNP/ TNP veritabanlarını araştırmak için çeşitli araçlar mevcuttur. BLAST yardımıyla genotip, metod, popülasyon, gönderici (submitter), markır ve dizi benzerliği aramasına olanak sağlamaktadır.	
	1000 Genome Browseri http://www.ncbi.nlm.nih.gov/variation/tools/1000genomes/ 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.	
	Basic Local Alignment Search Tool (BLAST) http://blast.ncbi.nlm.nih.gov/ Biyolojik diziler arasındaki lokal benzerliklerin olduğu bölgeleri bulur. Nükleotid 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.	
	Map Viewer http://www.ncbi.nlm.nih.gov/mapview/ 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.	
	Genome BLAST http://www.ncbi.nlm.nih.gov/blast/Blast.cgi Genomik dizi veritabanlarından nükleotid ve protein dizilerinin karşılaştırılmasını sağlar ve BLAST algoritmasını kullanarak istatistiksel anlamlılığını hesaplamaktadır.	

Tablo 2. Genomik diziler için genom tarayıcıları (devam)

VERİ TABANLARI YAPISAL	UCSC	Genome Browser - Kromozom bölgelerini yakınlaştırmak, gezinmek, açıklamaları görebilmek.
		Blat - Araştırılan diziyi genoma hızlı bir şekilde haritalamak.
		Table Browser - Temel veri tabanına erişim sağlamak.
		Gene Sorter - Birbiriyle ilişkili genleri sıralar. Protein homolojisi, gen ekspresyon profili veya genomik yakınlığı da içeren farklı çeşitlerde ilişkiler olabilir.
		In - Silico PCR - PCR primer dizi çifti ile dizi araması yapar.
		VisiGene in situ - Görüntülerin taranması için sanal mikroskop.
	Session - Geçerli ayarları kaydetmenizi ve daha sonra çağırarak kaldığınız yerden devam etmenizi sağlar.	
	EBI (ENSEMBL)	İçerdiği organizma sayısına göre ve içerdiği bilgiye göre geniş kapsamlı bir veri tabanıdır.
	FONKSİYONEL	ENCODE
EPIGENOM		İnsan genomunda değişken metilasyon pozisyonlarını (Methylation Variable Positions (MVPs)) kataloglamayı amaçlamıştır.

Tablo 3. Genom varyasyon veritabanları

GENOM VARYASYON VERİ TABANLARI	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) 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%).

^aMinör Allel Frekansı (MAF) popülasyonda tanımlanan TNP'nin az yaygın allelin frekansına verilen addir. HapMap Projesinde MAF 0,05 ve üzeri seçilmiştir. 1000 Genom Projesinde MAF 0,05 ve altı seçilmiştir.

Wellcome Trust Philanthropy Arthur L. Holden'in liderlik ettiği 300.000 ortak TNP 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.

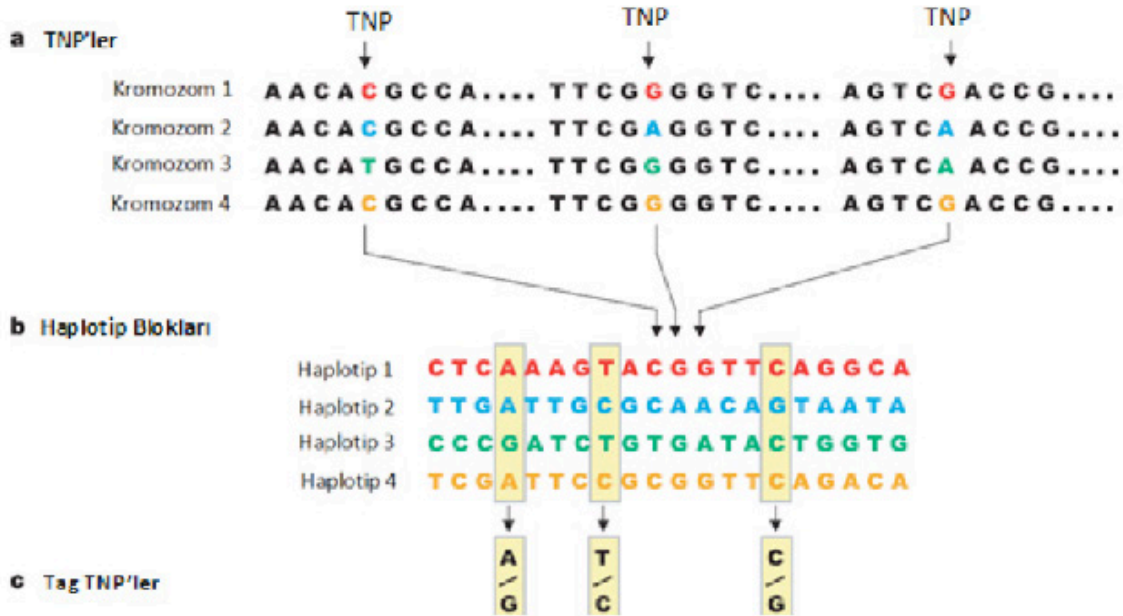
4.1. HapMap

2003 yılında İGP ile insan genomunun tamamının dizilenmesi sonucunda, genomun ancak %0,1'inin bireyler arası 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:

- Korelasyon ve frekans hesapları kullanılarak popülasyonlara özgü varyasyonlar ortaya konmuştur,
- Assosiyasyon çalışmaları ile bağlantı (linkage) analizleri için gerekli tüm genom taramasına imkan veren araçlar sağlanmıştır.



Şekil 2. TNP, Haplotip, Tag TNP¹

- 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'te C ve T alleleri mevcuttur.
- 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.
- 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).

4.2. 1000 Genom Projesi

İGP'nin 2003 yılında tamamlanmasından sonra 2004 yılında, sentez ile dizileme (sequencing by synthesis) esasına dayanan yeni nesil 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). Oluşturulan 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

Tablo 4. HapMap Projesi Aşamaları (21-23).

	Faz 1 (21)	Faz 2 (22)	Faz 3 (23)
Örnekler ve Popülasyon Çeşitleri	269 örnek (4 grup ^b)	270 örnek (4 grup ^c)	1,115 örnek (11 grup ^d)
Genotipleme merkezleri	HapMap International Consortium	Perlegen http://genome.perlegen.com	Broad & Sanger http://www.sanger.ac.uk/humgen/hapmap3/
Genotiplenen TNP	~1 M	~3.1 M (faz I+II)	1.6 M (Affy 6.0 & Illumina 1M)

^b 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).

^c 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).

^d 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 metropolitenindeki 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 Yorubalılardan (Batı Afrika) 180 birey.

Tablo 5. 1000 Genom pilot proje aşamaları

Pilot Proje Aşamaları	Örnek Sayısı	Dizileme	Kapsama (Coverage)	Durum
1	179 ^e	tüm genom	low 2-6X	Ekim 2008'de tamamlandı
2	2 aile ^f	tüm genom	high coverage - ortalama ~ 42X	Ekim 2008'de tamamlandı
3	7 popülasyondan 697 ^g	8,140 ekzon dizilemesi	(~50X, Haziran 2009'da tamamlandı)	Haziran 2009'da tamamlandı

^e 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).

^f 2 anne/baba/çocuk üçlüsü 6 kişi (Ibadan Nijeryadan bir Yorubalı (YRI); Utah'dan bir Avrupa kökenli (CEU)).

^g Afrikadan 7 Popülasyon (YRI, Webuye-Luhya, Kenya (LWK)), Avrupalı (CEU, İtalya-Toskana (TSI)) ve Doğu Asyalı (CHB), JPT, Colorado kökenli, Denver Çinlileri (CHD).

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. Ayrıca kodlama bölgelerinde %0,1'in altında olan allel frekansları da kataloglanmıştır (24).

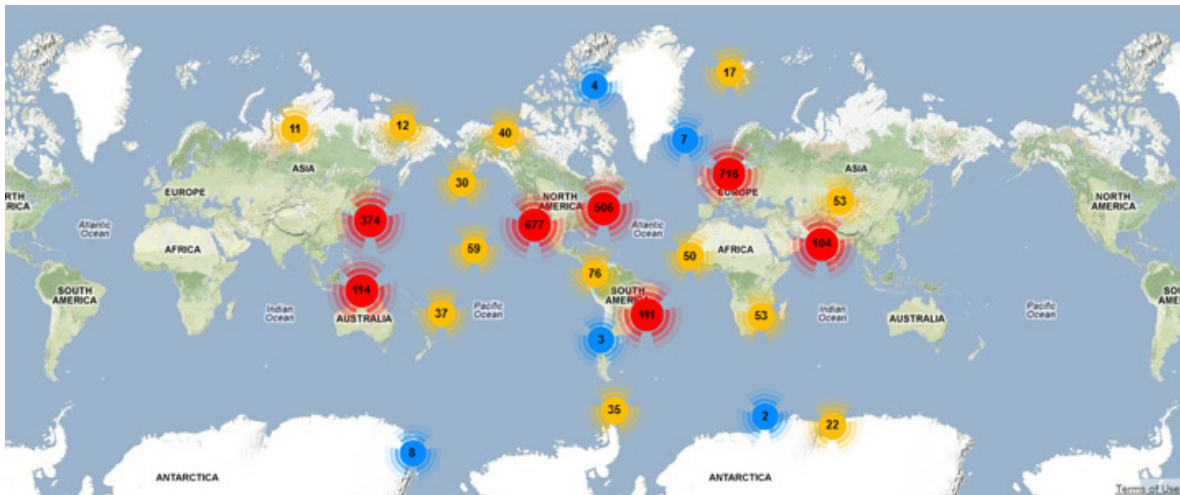
1000 Genom Projesi insan genetik varyasyonlarının coğrafi ve fonksiyonel spektrumunu betimleyerek genetiğin hastalara olan etkisini anlamamıza yardımcı olacak bir kaynak oluşturmayı hedeflemiştir. Projede 14 popülasyondan 1092 kişinin düşük kapsamlı bütün genom ve ekzom dizilemesinin kombinasyonundan oluşan genomlarını açığa çıkarılmıştır (Şekil 3) (25). Sonuçta bu projede çeşitli algoritmalar ve farklı veri kaynakları üzerinden bilgiyi entegre eden metodların geliştirilmesi suretiyle 38 milyon TNP, 1.4 milyon kısa in-del ve 14.000'den fazla büyük delesyon için doğrulanarak, haplotip haritası ortaya çıkarılmıştır (19). Farklı popülasyonlardan olan bireylerin nadir ve sık görünen varyantlarda farklı profiller taşıdığı ve düşük frekanslı varyantların dikkate değer oranda (negatif seleksiyonun etkisi ile daha da artan) coğrafik farklılaşma gösterdiği 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, her bir birey transkripsiyon faktörünün bağlanma bölgesindeki motifleri bozan değişimler gibi kodlayıcı olmayan korunmuş bölgelerde yüzlerce varyant taşıdığı tespit edilmiştir (26).

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



Şekil 3. GOLD Genom haritası (25).

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 veri yığınının var olan 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 oluşturmuştur.

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'lerini de içerecek şekilde Asya insanının popülasyon yapısını ortaya koyan değerli bir kaynaktır. Buna ilave olarak 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 itibarıyla 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 itibarıyla yaklaşık 5,8 milyar olan dünya nüfusunun resmi olarak tanınan 56 alt popülasyon ile beşte birini Çin oluşturmaktadır. Araştırmacılar CGHD Projesi ile mikrosatelitleri 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 = 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)

Hindistan, Nisan 2013 yılı itibarıyla 1 milyar 234 bin olan nüfusuyla dünyanın en kalabalık ikinci ülkesidir (36). Hindistan popülasyonu kabile içi evlenen yüzlerce grupta birlikte 4.693 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 açıdan 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 (EGK) anlaşması sonucunda 1999 Mart ayında planlanmaya başlanmıştır. Başarıya ulaşmanın ana unsurlarından biri olan kamu-özel sektör birlikteliği EGK ve US = Amerika 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 genomün toplum sağlığına getireceği sistematik avantajların pratikte uygulanabilmesi de projenin hedefleri 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'lularla 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, Afrikalı-Amerikalılar Vakfı tarafından desteklenen Afrikalı-Amerikalıların Afrika kalıtım miraslarının araştırıldığı Afrika Genom Projesi ve Türkiye Genom Projesi gibi genom projeleri de bulunmaktadır (40-42).

SONUÇ

İnsan Genom Projesi 1990 yılında başlatıldığında dünya kamuoyunda projenin bitiminde ortaya çıkarılacak bilgi ile yaşamın şifresinin çözüleceği ve bunun sonucunda da birçok hastalığın çaresinin bulunmuş olacağı beklentisi oluşmuştu. 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 = Genom, Encode vb.) yürütülmektedir. Bu durum bütün araştırmaların boşuna yapıldığı yanılgısını ortaya çıkarmamalıdır. 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 yaşam bilim modellerinin oluşturulabilmesinin ilk aşaması, bu verilerin uygun veri tabanlarında saklanması ve kategorize edilmeleri idi. Bu amaçla Genbank (genome, dbSNP, dbGaP vb.), HapMap, 1000 Genom, 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 veri tabanları 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 ederken 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 genom 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 tasarımından, uygun istatistiksel yöntemlerin seçilmesine veya yeni yöntemlerin geliştirilmesine, veri analizlerinden, sonuçların yorumlanmasına kadar bir çok konuda projeye dahil olması da büyük önem taşımaktadır. Örneklem sayısının, ana kütle 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 çalışmalarla kıyaslayarak genomik bilginin globalliğini araştırmaktadır. Aynı zamanda bu çalışmalar ile kendi popülasyonlarının ne ölçüde temsil edildiğini de incelenmektedir. Bu çalışmalar sonucunda ortaya konan hastalık genotip ilişkileri sayesinde klinik olarak ortaya konan tedavi yöntemlerinin seçimi, ilaç doz ayarlamaları ile ilgili çalışmalar yürütülebilmektedir. Sonuç olarak, bu yeni bilgiden 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>(Erişim tarihi: 07.08.2013)
2. <http://www.gpo.gov/fdsys/pkg/WCPD-2003-04-21/pdf/WCPD-2003-04-21-pg440.pdf> (Erişim tarihi: 07.08.2013)
3. <http://www.phgfoundation.org/news/8538> (Erişim tarihi: 07.08.2013)
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 th finished human chromosome 22 sequence. *Genome Biol*, 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 (Erişim tarihi: 07.08.2013)
9. <http://www.nia.nih.gov/alzheimers/publication/alzheimers-disease-genetics-fact-sheet> (Erişim tarihi: 07.08.2013)

10. White RL, Lalouel JM, Nakamura Y, Donis-Keller H, Green P, Bowden DW, Mathew CG, Easton DF, Robson EB, Morton NE, et al. The CEPH consortium primary linkage map of human chromosome 10. *Genomics*, 1990; 6(3): 393-412.
11. Botstein D, White RL, Skolnick M, Davis R. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Genet*, 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): 11501-3.
14. Klug WS, Cummings SMR. Kromozom Yapısı ve DNA Dizisinin Organizasyonu. In: Öner C. *Genetik Kavramlar*, 6. Baskı. Ankara: Palme Yayıncılık, 2003: 544-9.
15. Nakamura Y. DNA variations in human and medical genetics: 25 years of my experience. *J Human Genet*, 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 H L. 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 Genet*, 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: 1299-320.
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(Erişim tarihi: 07.08.2013)
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. *J Bioinformatics Comp Biol*, 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. Anonim (<http://www.trthaber.com/haber/dunya/iste-cinin-nufusu-71334.html>) (Erişim tarihi: 07.08.2013)
34. Qiu J, Hayden EC. Genomics sizes up. *Nature*, 2008; 451: 234.
35. <http://yh.genomics.org.cn/> (Erişim tarihi: 07.08.2013)
36. <http://www.indiastat.com/default.aspx> (Erişim tarihi: 07.08.2013)
37. The Indian Genome Variation Consortium. The Indian Genome Variation database (IGVdb): a project overview. *Hum Genet*, 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://iranges.com/how-to-participate>) (Erişim tarihi: 07.08.2013)
40. (<http://iranges.com/how-to-participate>) (Erişim tarihi: 07.08.2013)
41. (<http://dubois.fas.harvard.edu/african-genome-project>) (Erişim tarihi: 07.08.2013)
42. (<http://turkiyegenomprojesi.boun.edu.tr>) (Erişim tarihi: 07.08.2013)