

# İnsan Genomunu Anlamlandırma Projesi

İnsan Genom Projesi, insanın genetik kodunu oluşturan DNA'daki 3 milyar baz çiftinin dizilimini ortaya koydu. Ancak proje tamamlandığında bu kodun büyük bir kısmının nasıl işlev gördüğü hâlâ bir sırdı. İnsan Genom Projesi'nin hemen ardından başlayan ENCODE Projesi, çok sayıda laboratuvarın ve araştırmacının uzun yıllar süren ortak çabaları sonucunda insan genomunun işleyişine dair bütünsel bir anlayış oluşturma yolundaki ilk verileri ortaya koydu.

2003 yılında tamamlanan İnsan Genom Projesi kamuoyunda büyük yankı uyandırmıştı. Çünkü insanın genetik kodunun ortaya çıkarılması söz konusuydu. Ancak genetik kodun ortaya çıkarılması yalnızca DNA'yı oluşturan baz çiftlerinin diziliminin ortaya çıkarılması anlamına geliyordu. Bilim insanlarını bekleyen asıl dev bulmacaysa bu dizilimin ne anlama geldiği ve genetik kodun nasıl işlediğiydi. İnsan Genom Projesi'nin hemen ardından başlatılan ve yine çok sayıda laboratuvarın ve araştırmacının dâhil olduğu çok kapsamlı bir proje olan ENCODE (Encyclopedia of DNA Elements) projesi işte bu bilginin peşine düştü.

İnsan Genom Projesi'nin tamamlanmasıyla üç milyar baz çiftinden oluşan insan genetik kodunun sadece % 1'inden biraz fazla bir kısmının protein kodladığı anlaşılmıştı. Bu da yaklaşık 20.000 gene karşılık geliyordu. Ancak pek çok bilim insanı, insanın sahip olduğu şaşırtıcı karmaşıklığın genomun o zamana kadar "çöp DNA" olarak anılan % 99'luk kısmında gizli olduğunu düşünüyordu. İşte ENCODE projesi bu kısmın işlevlerini ortaya çıkarmaya yönelik muazzam bir veri toplama çabası olarak tasarlandı. Projenin amacı, bu kısımdaki işlevsel DNA dizilerini tespit edip bunların hangi hücrelerde etkin olduğunu, genomun paketlenmesi ile genlerin düzenlenmesi ve ifadesi üzerinde nasıl etkileri olduğunu anlamak.

## "Çöp DNA"nın Sırları Çözülüyor

ENCODE projesi 2003 yılında hedeflenen sonuçları elde etmek için gerekli yöntemlerin ve stratejilerin geliştirildiği ve genomun sadece % 1'lik kısmının ele alındığı bir pilot aşamayla başladı. 2007'de araştırmacılar geliştirdikleri yöntemleri tüm genoma uygulamaya başladı. Bu aşamanın da sonuçları geçtiği-

miz yıl üç bilimsel dergide yayımlanan 30 makaleyle açıklandı. Projede genomun yaklaşık % 80'lik bir kısmına bir çeşit işlev atfedilmiş oldu. Bunlar arasında, DNA diziliminde genlerden hemen önce gelen ve belirli proteinlerin gen anlatımını kontrol etmek üzere bağlandığı 70.000 kadar promotör bölgesi ile kendilerinden uzak bölgelerde yer alan genlerin anlatımını düzenleyen 400.000 kadar hızlandırıcı (*enhancer*) bölge yer alıyor.

Projede toplamda 440'tan fazla araştırmacının yer aldığı 32 araştırma grubu 24 standart deney tipi kullandı. Araştırmacılar genomdan transkripsiyonla üretilen RNA moleküllerinin baz dizilimini belirledi ve protein kodlamayan pek çok RNA molekülü üretildiğini ortaya koydu. Promotörlere bağlanarak gen anlatımının başlamasını sağlayan 120 transkripsiyon faktörü için bağlanma bölgelerini belirlediler. Genellikle hangi genlerin pasif durumda olduğunun bir göstergesi olan metil kimyasal grubuyla örtülü genom bölgelerini haritaladılar. DNA'nın kromozomlar biçiminde paketlenmesini sağlayan histon proteinleri üzerinde oluşturulan ve gen anlatımının hızlandırıldığını mı yoksa baskılandığını mı gösteren kimyasal değişiklikleri incelediler. Ayrıca her ne kadar hemen hemen tüm hücrelerimiz aynı genoma sahip olsa da genom her hücrede aynı biçimde işlemiyor. Bu yüzden de araştırma grupları bu deneyleri en az 147 hücre üzerinde uyguladı ve sonuçta toplam 1648 deney gerçekleştirilmiş oldu. Yapılan deneyler, kısmen Ulusal İnsan Genomu Araştırma Enstitüsü'nün (NHGRI) teknoloji geliştirme programı sayesinde, sadece beş yıl önce kullanıma sunulan yeni nesil DNA dizi analizi teknolojilerine dayanıyor. ENCODE toplamda 15 trilyon baytlık ham veri ortaya koydu ve analizler için 300 yıla denk bilgisayar zamanı harcadı.

## ENCODE Neler Vaat Ediyor?

ENCODE projesi her şeyden önce NHGRI'deki program yöneticilerinden Dr. Elise Feingold'un deyişiyle genomun, proteinlerin nerede ne zaman üretilceğini belirleyen "açma kapama düğmeleri" olduğunu, canlı ve dinamik bir yapıya sahip olduğunu gösterdi ve genoma ilişkin anlayışımızı bir üst seviyeye taşıdı.

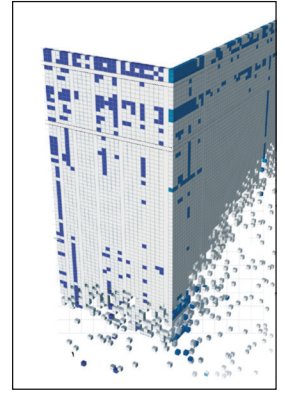
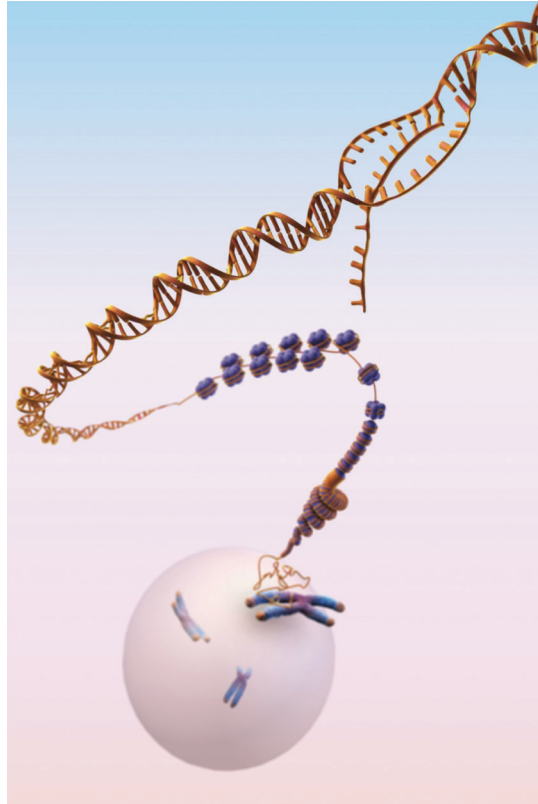
ENCODE genomun gizli kalmış yönlerini ortaya çıkararak genetik çeşitliliğin insan özelliklerini ve hastalıklarını nasıl etkilediğini daha iyi anlamak için bir fırsat yarattı. Bilim insanlarının, projenin genomda ortaya çıkardığı sayısız düzenleyici elemanı inceleyerek ve bunların dizilimlerini diğer memelilerdekiyle karşılaştırarak insanı diğer memelilerden farklı yapan şeyin ne olduğunu daha iyi anlayabileceği düşünüyor.

Projede elde edilen ve herkesin erişimine açılan veriler şimdiden araştırmacıların hastalık genetiğini daha iyi anlamasına katkıda bulunuyor. Genom ölçeğinde ilişkilendirme çalışmaları (GWAS) 2005 yılından bu yana, genomun DNA kodundaki tek harflik bir farklılığın hastalık riskiyle ilişkili görüldüğü binlerce nokta yakaladı. Ancak bu noktaların neredeyse % 90'ı protein kodlayan bölgelerin dışına düşüyor. Dolayısıyla araştırmacıların elinde bu noktaların hastalığa nasıl sebep olabileceğine ya da etki edebileceğine ilişkin fazla ipucu bulunmuyor. ENCODE projesinde oluşturulan harita, hastalıkla ilişkili bu bölgelerin pek çoğunun hızlandırıcı bölgeler ya da başka işlevsel diziler içerdiğini ortaya koydu. Ayrıca hücre tipi de önem taşıyor. Yani GWAS içerdiği tespit edilen düzenleyici bir bölge, belirli tipteki hücrelerde etkinken diğerlerinde etkin olmayabiliyor. Dolayısıyla ENCODE projesinin farklı hücre tiplerine ilişkin analiz sonuçları da GWAS sonuçlarının anlamlandırılmasında faydalı oluyor.

## Nereye Kadar ENCODE?

Bu kadar kapsamlı bir projenin maliyeti doğal olarak hayli yüksek. Pilot aşama 55 milyon dolara, tüm genom üzerindeki analiz aşamasıysa 130 milyon dolara mal olmuş. NHGRI sonraki aşama için 123 milyon dolar daha ayırabilecek. Bazı araştırmacılar projenin artık bu yüksek maliyetine değecek kazanımlar sağlaması gerektiğini ve projeye ayrılan bütçeyle hipoteze dayalı çok sayıda başka araştırma projesinin desteklenebileceğini savunuyor. Ayrıca projenin ucu açık gibi görünüyor. Çünkü ENCODE her ne kadar şimdiye kadar bu konuda yapılmış en kapsamlı

ve geniş ölçekli proje olsa da henüz hedeflenen verilerin küçük bir kısmına ulaşılmış durumda. Projenin ortaklarından California Üniversitesi'nin internet sitesinde, ENCODE'un ilerlemesini gösteren bir grafik yer alıyor (Yanda). 24 deney tipinden hangilerinin tamamlandığını ve 180 hücre tipinden hangilerinin incelendiğini gösteren grafiğin çok az bir kısmının dolu olduğu görülüyor. Ayrıca ENCODE, çeşitli teknik sebeplerden dolayı aslında genomun anlamlandırılması için gereken bilgilerin tamamını hedefleyemiyor. Örneğin projede 180 hücre tipi ele alınıyor oysa insan vücudunda birkaç bin kadar hücre tipi var. Ya da örneğin projede ChIP adı verilen bir teknik kullanılarak belirli proteinlerin bağlandığı DNA bölgeleri tespit ediliyor, ancak araştırmacıların elinde bu yöntemde kullanılacak sadece 2000 kadar protein var. Dolayısıyla projenin kapsamı genişledikçe genişleyebilir de. Bu da projenin hangi noktada tamamlanmış sayılması gerektiği konusunda tartışmalara neden oluyor. Belki de ENCODE projesinin sonuçları özellikle sağlık konusunda somut faydalar sağlamaya başladıkça projenin devam etmesinin gerektiği daha çok kabul görecektir. Görünüşe göre insan genomunun anlamlandırılması ve işleyişinin anlaşılması bu yüzyılın en önemli bilimsel meselelerinden biri olacak.



DNA'nın kromozomlar biçiminde paketlenmesini sağlayan histon proteinleri üzerinde oluşturulan ve gen anlatımının hızlandırıldığını mı yoksa baskılandığını mı gösteren kimyasal değişiklikler ENCODE projesinde incelenen özellikler arasında.

### Kaynaklar

Maher, B., "The Human Encyclopedia", Nature, Cilt 489, Sayı 7414, s.46-48, 2012. <http://www.genome.gov/27549810>