

Test Tüpü Ya da Ram? İşte Asıl Sorun Bu

TEST TUBE OR RAM? THAT IS THE REAL QUESTION

Murat V. YAZICIOĞLU

Dokuz Eylül Tıp Fakültesi Dekanlık Bilgişlem Birimi

Murat YAZICIOĞLU
Dokuz Eylül Üniversitesi
Tıp Fakültesi
Dekanlık Bilgişlem Birimi
Tel: 0232 2595959 - 2260
e-mail: yazicimv@deu.edu.tr

ÖZET

Genom Projesi ile elde edilen data Internet üzerinden herkesin kullanımına açık veri tabanlarında sunulmakta. Bu verilerin toplanması için küçük bir kısmını oluşturdu. Şimdi bu verinin ne anlama geldiğini öğrenmemiz ve bu bilgiyi teknolojiye çevirmemiz gerek. Günümüzde bilgisayarlar bu aşamada test tüplerinden daha fazla yardımcı olabilecek yetenekler.

SUMMARY

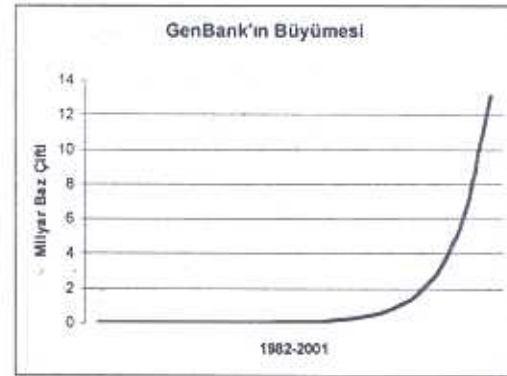
The data which collected with genome project is representing in anonymous databases on Internet. Collecting data is a small part of the task. Now we must learn what the data means and we must turn this knowledge into technology. At this point computers can help us more than test tubes.

Genom projesi sonuçlandığında elimize geçen data, "a", "g", "c" ve "t" lerden oluşmuş milyonlarca harflik bir metin oldu. Bu uzun metnin oluşturulabilmesi, biyologlar için uzun yıllar boyu sadece tatlı bir düşünmeydi. Saiki'nin PCR yöntemini geliştirmesi, bu düşünme gerçekleştirilmesi için önemli bir adım oldu. Böylece; Sanger'ın DNA dizileme yöntemi PCR ile birleştirilebilir ve DNA dizileri daha kolaylıkla okunabilirdi. Doksanlı yılların ortalarından itibaren, dünyanın bir çok yerinde araştırmacılar, eskisine oranla çok daha fazla DNA molekülünün nükleotid dizisini tanımladılar.

Aynı dönem; bilgisayarların günlük yaşamımızda daha fazla yer aldığı, internetin hızla yayıldığı ve bir çok saygın bilim insanının "kendi alanı dışında" kalan bu gelişime ayak uydurmakta zorlandığı bir dönem oldu. Bu iki alanı birleştiren bir bilim dalı olan biyobilişim (Bioinformatics); gelişmiş ülkelerde popülerliğini hızla artırırken, EMBL (European Molecular Biology Laboratory), DDJB (DNA DataBank of Japan), GenBank (NIH genetik Dizi Veritabanı) gibi

biyobilişimin önde giden merkezleri verilerini birleştirdiler (1).

Bilgisayarlar her 18 -24 ayda bir işlem yapabilme yeteneklerini ikiye katlıyorlar. Genbank'taki veriler ise, her 14 ayda bir iki katına çıkmaktadır (Şekil 1). Bu iki verinin paralelliği, bir rastlantıdan çok DNA dizilerinin elde edilmiş yöntemleriyle ilgilidir (1-3).



Şekil 1. GenBank'da biriken datanın yıllara göre dağılımı görülmekte. Bu data yaklaşık her 14 ayda ikiye katlanmaktadır (2). Bilgisayar işlemlerinde ki transistör sayısı ise her 18 -24 ayda ikiye katlanmaktadır (3).

Test tüpü ya da ram? İşte asıl sorun bu

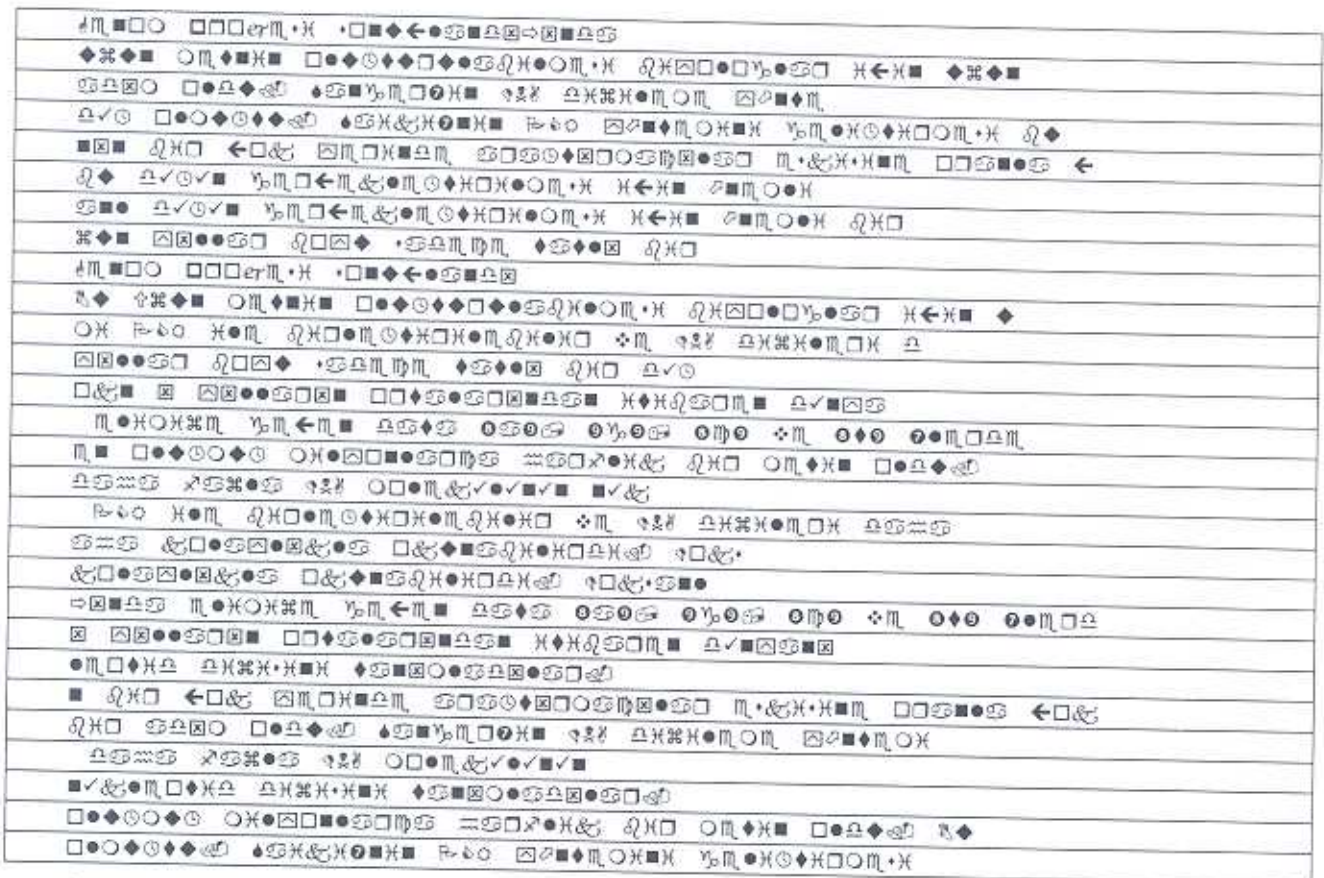
Günümüzde DNA dizilerinin elde edilmesinde kullanılan kimyasal yöntemler, bir çok laboratuvar da bilgisayar kontrollü robotlar tarafından uygulanmaktadır. Bu uygulama, Genom Projesini hızlandıran en önemli etkenlerden biri olmuştur (3).

DNA molekülündeki bazların sırası, çeşitli kimyasal yöntemler kullanılarak okunabilmektedir. Bu yöntemlerle bir defada okunabilen baz baz sayısı ise sınırlıdır. Bunun için DNA moleküllerinin her seferinde başka bir bölümü okunarak elde edilen verilerin birleştirilmesi gerekmektedir. İşin içine kişiden kişiye değişen (polimorfik) diziler de katılınca, iş daha da zorlaşmaktadır (3).

Şekil 2'de bir yap boz oyununun parçaları yer almakta. Oyunun parçaları oluşturulurken bu yazının bir

paragrafı farklı bir fontta iki kez yazılmış (her hücrede aynı genleri içeren DNA molekülünden iki tane olduğunu hatırlayın) ve birbirine yakın uzunluklarda (her iki allel farklı ebeveynlerden geldiği için küçük farklar içermeli) parçalara ayrılmıştır. Amaç, bu harfleri ilk sırasına getirmektir. Bu paragrafta yaklaşık 500 harf olduğu düşünülürse; GenBank'ta bulunan 100.000'den fazla türden elde edilmiş 13 milyar bazlık datanın değeri, daha iyi anlaşılacaktır.

Veri bankalarında bulunan datanın büyük bir kısmı, henüz tümüyle işlenmemiş ham data olarak saklanmaktadır. Farklı başlıklar altında bulunan data tümüyle işlendiğinde hata oranı %0,01 civarında gerçekleşmektedir (2,5).



Şekil 2. Yap Boz

Bu verinin ham data şeklinde depolanmış olan kısmının işlenmesi ile genom kitabının sayfaları yerlerine oturmuş olacak. 20 yüzyıl boyunca biyologlar, bu kitabı oluşturmak için büyük emek harcadılar. 21. yüzyılda ise biyobilimcilerin işi, matematik modeller ve bu modellerin *in vitro* ölçümlere uygunluğunu test etmek olacak. Bu modellerle, var olan data bilgiye dönüşecek ve elimizdeki genom kitabı "okunur" hale gelecektir (6).

Matematiksel modelleriniz ne kadar gerçeğe yaklaşırsa, bu modelleri kullanarak ulaştığınız sonuçlar da o denli gerçeğe yakın olacaktır. Modellerle ulaşabilecek sonuçlar, günümüz teknolojisi ile saptayamadığımız biyokimyasal yollar hakkında bilgi edinmemize olanak sağlayacaktır (6).

Hücrel fonksiyonların modellenmesi üzerine yapılan çalışmalar, biyolojik bilimlerle ilgilenenleri bun-

dan böyle yeni seçimler yapmaya zorlamaktadır. Eğer yöntem *in vivo* olmayacaksa, "*in vitro*"mu "*in silico*"mu olacak?

KAYNAKLAR

1. Yazıcıoğlu MV. Nükleik asit analizi yöntemlerinin bilgisayarda simüle edilmesine yönelik bir dizi program tasarımı. *Dokuz Eylül Üniversitesi Sağlık Bilimleri Enstitüsü*, 1998.
2. Anonim. www.ncbi.nlm.nih.gov/Database/index.html, 2001
3. Cohen D. *Umudun Genleri*, Kesit Yayıncılık, 1995.
4. Anonim: <http://developer.intel.com/update/archive/issue2/focus.htm>, 2001.
5. Anonim: <ftp://ncbi.nlm.nih.gov/repository/LiMB/>, 2001.
6. Milburn J. *Beyond the genome: turning data into knowledge*, DDT 2001;6:881-883.