

Genomics

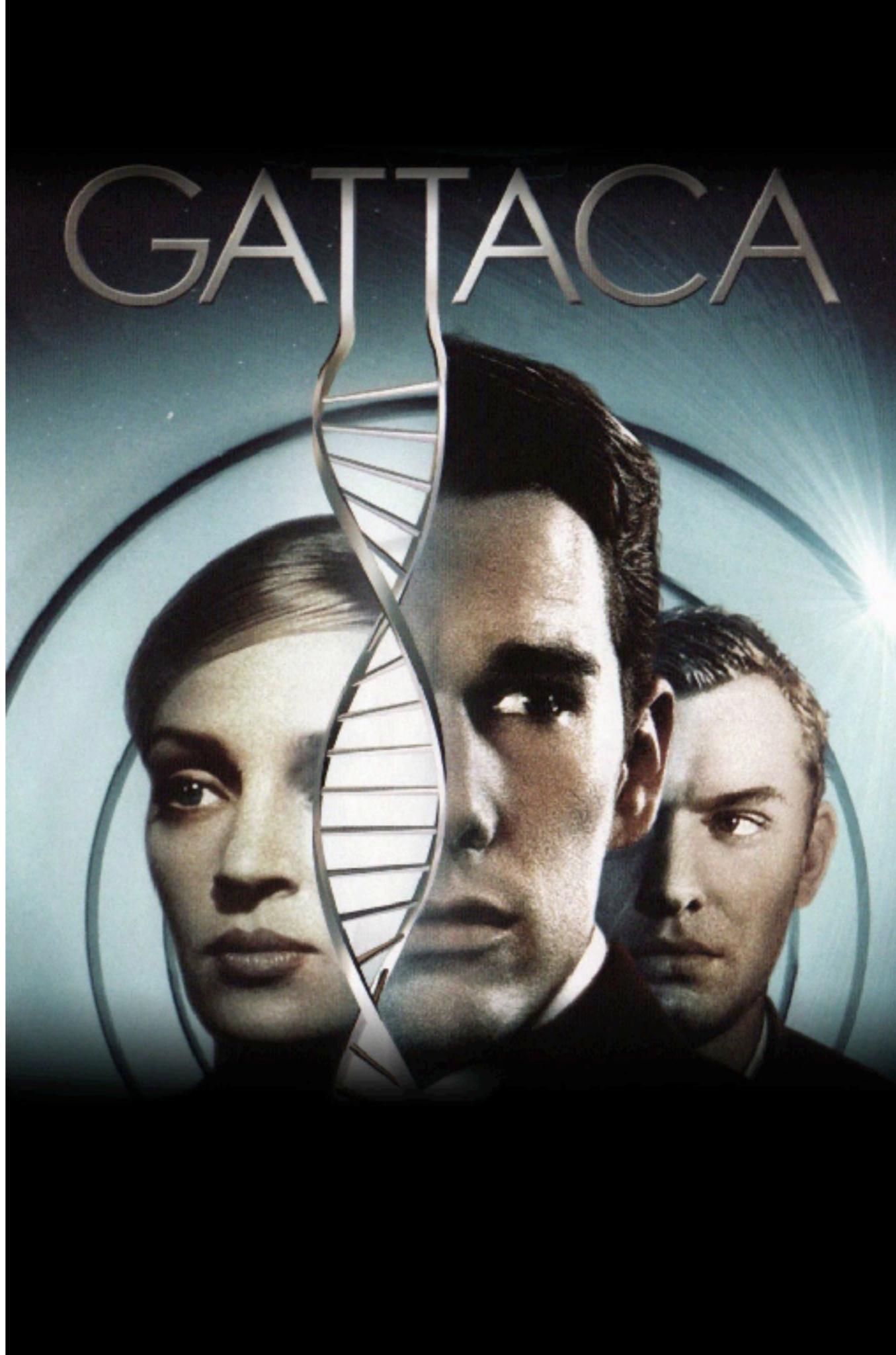
What do you know about genomes and genomics?

Where did you hear about them?

1993



1997





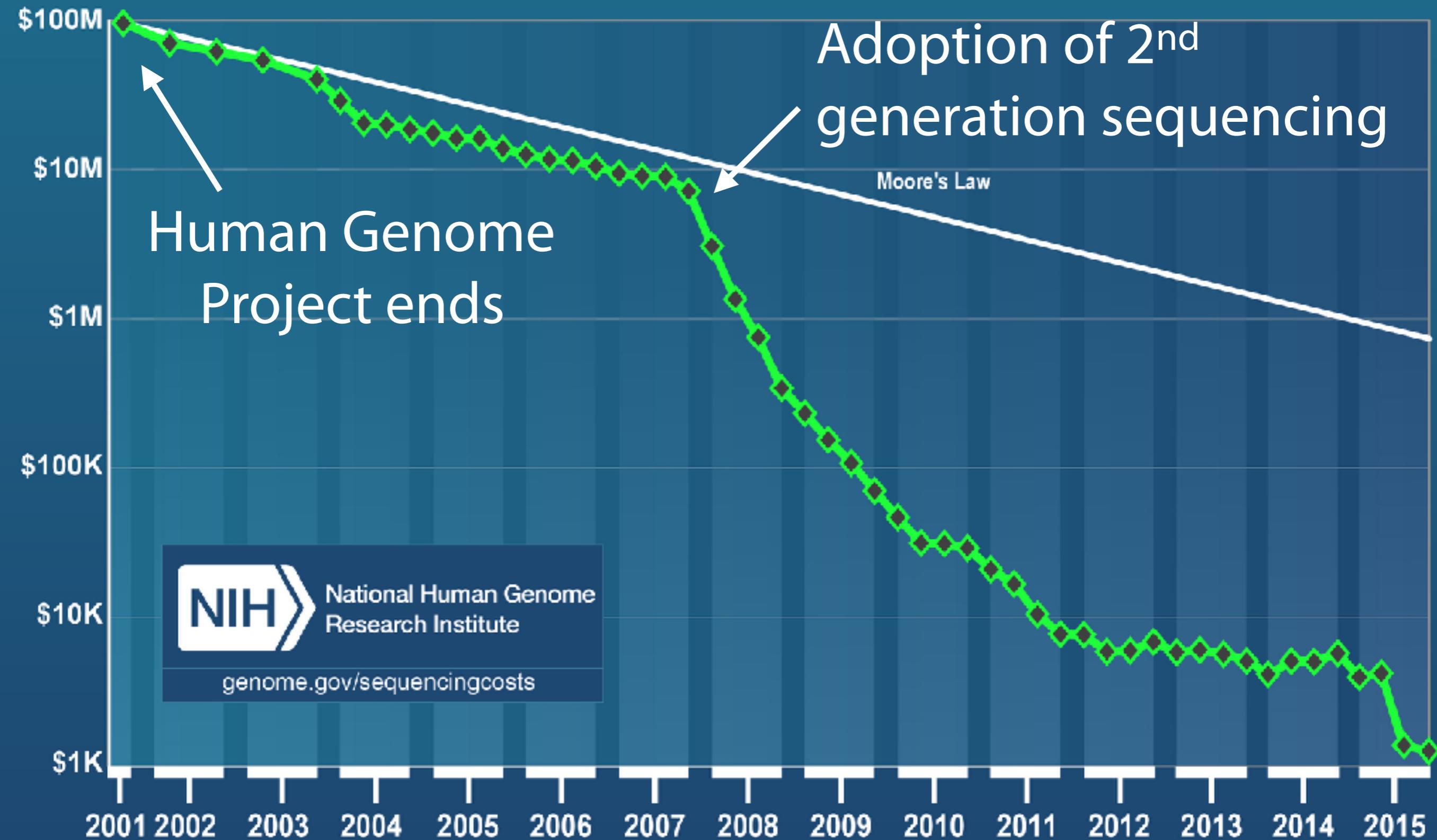
READING THE BOOK OF LIFE: THE OVERVIEW

READING THE BOOK OF LIFE: THE OVERVIEW; Genetic Code of Human Life Is Cracked by Scientists

By NICHOLAS WADE

Published: June 27, 2000

Cost per Genome



DNA sequencing instruments from Illumina: www.illumina.com



GA II
1.6 billion nt/day
(2008)



GA IIx
5 billion nt/day
(2009)



HiSeq 2000
75 billion nt/day
(2011)



HiSeq 2500
120 billion nt/day
(2012)

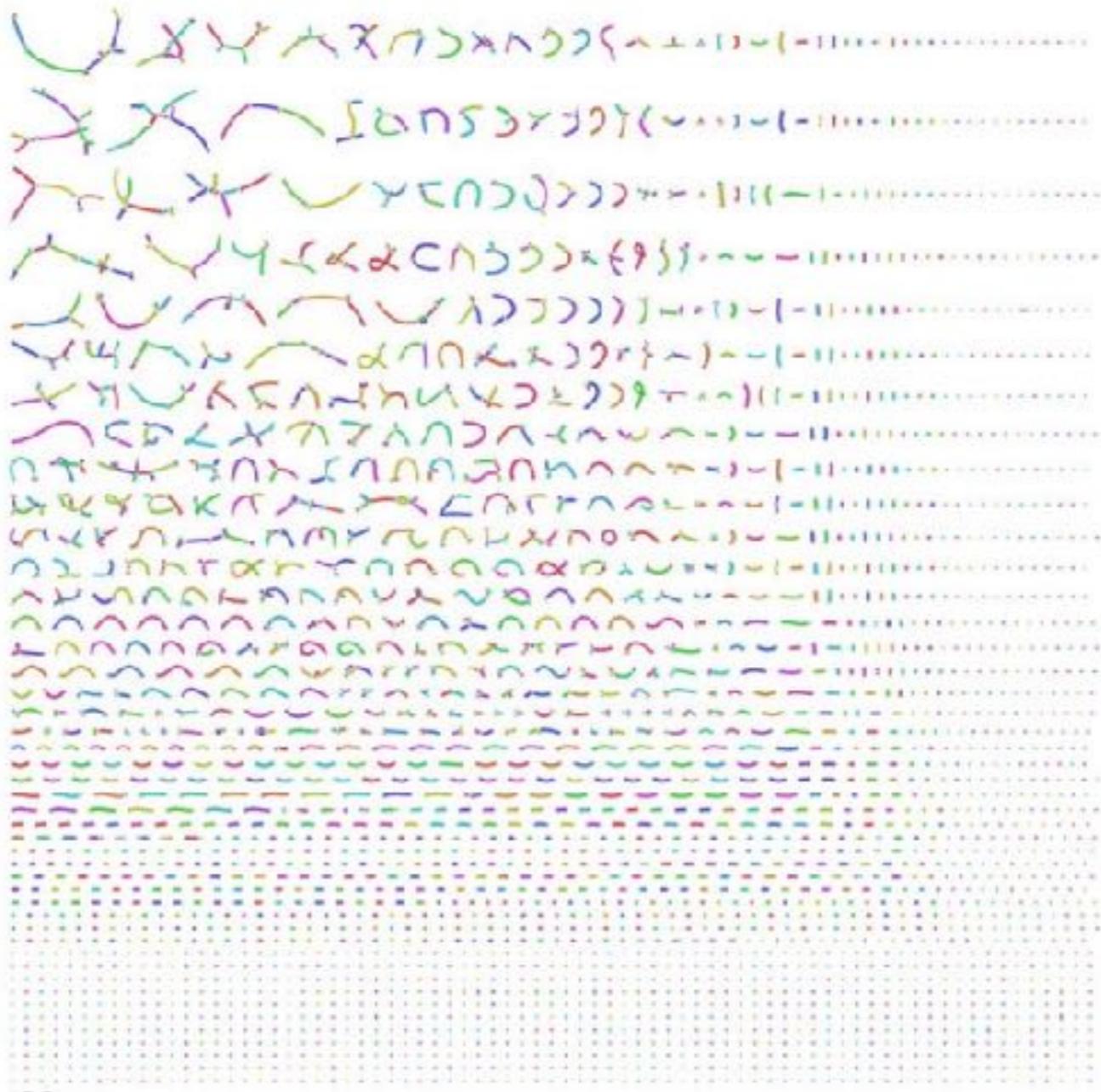
nt = nucleotide = **A, C, G or T**

HiSeq 3000/4000
200-400 billion nt/day
(2015)

NovaSeq 5000/6000
1-3 trillion nt/day
(2017)

Team of Rival Scientists Comes Together to Fight Zika

By AMY HARMON MARCH 30, 2016



A visualization of the recently sequenced *Aedes aegypti* genome. Each of the 3,752 colored lines is a fragment of its three chromosomes that could not be fit together without the additional information that the *Aedes* Genome Working Group hopes to produce. A 2007 genome map for *Aedes aegypti* is fragmented into about 10 times as many pieces. Mark Kunitomi

The New York Times

June 27, 2013

Studying Tumors Differently, in Hopes of Outsmarting Them

By CARL ZIMMER

THE NEW YORKER

MEDICAL DISPATCH | JULY 21, 2014 ISSUE

ONE OF A KIND

What do you do if your child has a condition that is new to science?

BY SETH MNOOKIN

SCIENCE

Man's Genome From 45,000 Years Ago Is Reconstructed

OCT. 22, 2014

Carl Zimmer

Sequencing

Sequencing is now a common tool for life scientists

The story echoes that of computing; once computers became fast & cheap, they were adopted everywhere





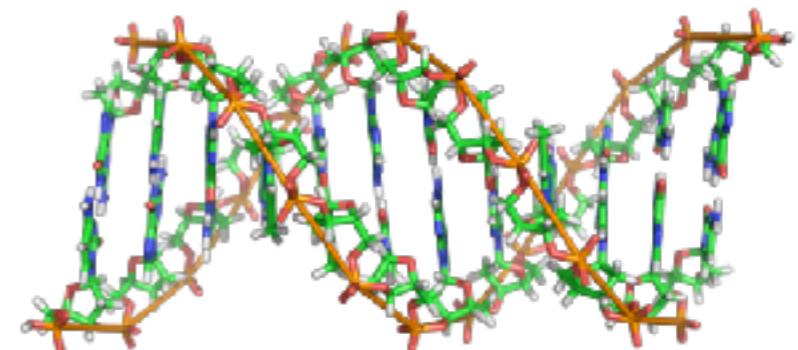
Genome

“The complete set of genes or genetic material present in a cell or organism.”

Oxford dictionaries

“Blueprint” or “recipe” of life

Self-copying store of read-only information about how to develop and maintain an organism



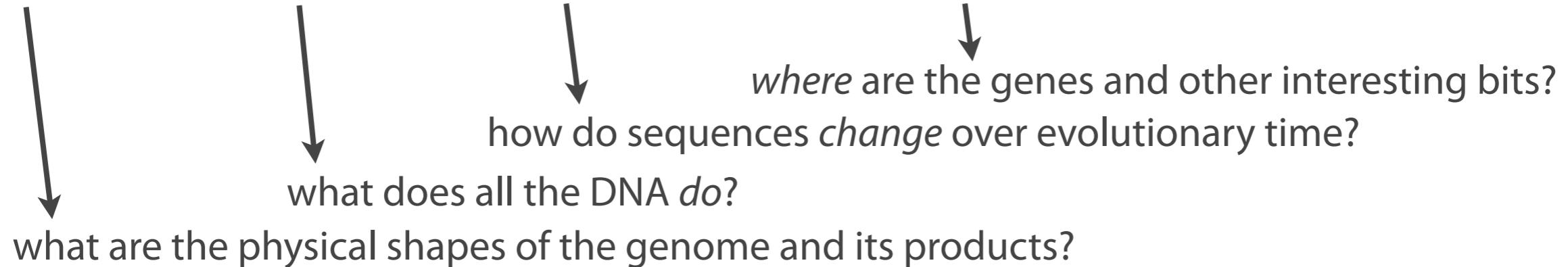
TAGCCCGACTTG

↓ ↓ ↓ ↓ ↓
↓ ↓ ↓ ↓ ↓ ↓ ↓
↑ ↑ ↑ ↓ ↓ ↓
↑ ↑ ↑ ↓ ↓ ↓

Genomics

Oxford dictionaries

“The branch of molecular biology concerned with the **structure, function, evolution, and mapping of genomes.**”



Collins English Dictionary

“The branch of molecular genetics concerned with the study of genomes, specifically the identification and sequencing of their constituent genes and the **application** of this knowledge in **medicine, pharmacy, agriculture, etc.**”

Genomics: contrast with biology & genetics*

* This slide has
gross generalizations

Biology & Genetics

Targeted studies of one
or a few genes

Targeted,
low-throughput
experiments

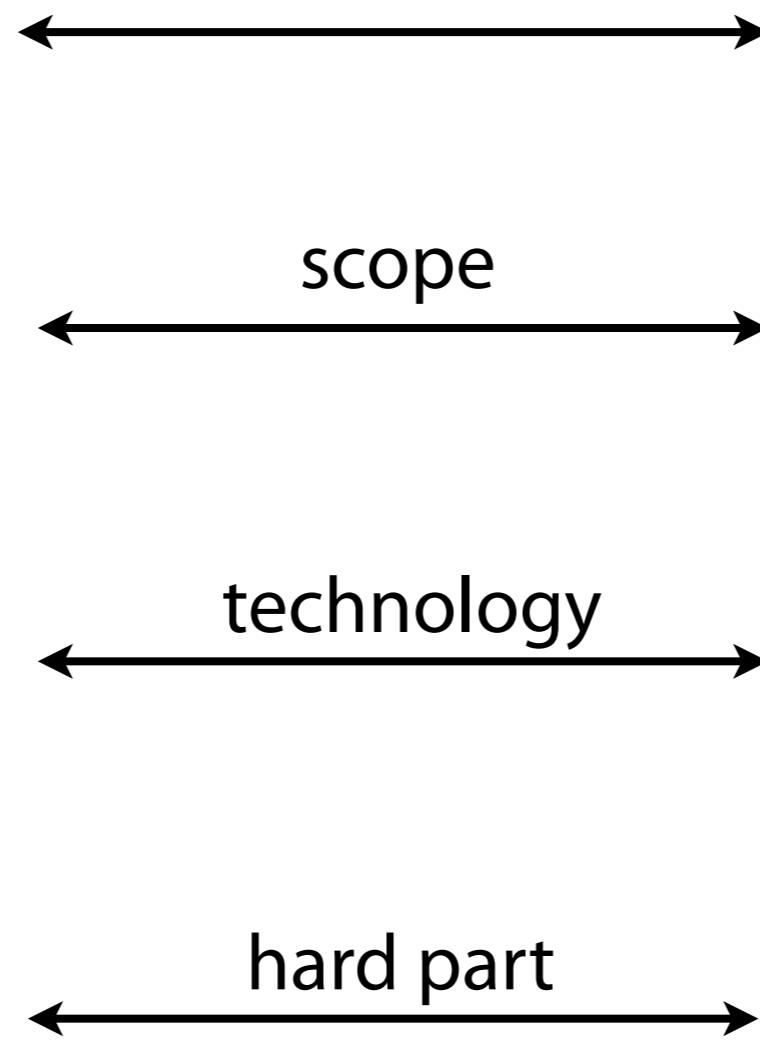
Clever experimental
design, painstaking
experimentation

Genomics

Studies considering all
genes in a genome

Global,
high-throughput
experiments

Tons of data,
uncertainty,
computation

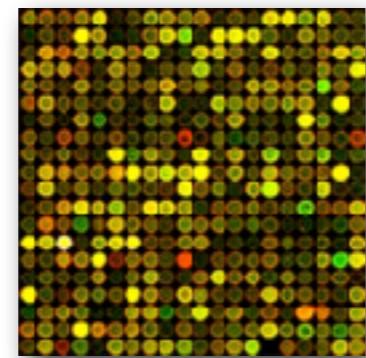


Genomics: shaped by technology



Sanger DNA sequencing

1977-1990s



DNA Microarrays

Since mid-1990s



2nd-generation DNA sequencing

Since ~2007



3rd-generation & single-molecule DNA sequencing

Since ~2010

These provide very high-resolution snapshots of the world of nucleic acids (not just DNA)

Genomics: tool for basic science

“The branch of molecular biology concerned with the **structure, function, evolution, and mapping** of genomes.”

Oxford dictionaries

Structure / mapping

What is the DNA sequence of the genome?

Where are the genes?

What is the genome’s three dimensional shape in the cell?

Function

What does all the DNA in the genome do?

What genes interact with what other genes?

How does the cell know what DNA is on/off?

Evolution

How did history shape our ethnicities and populations?

What big events shaped our current genetics?

Which portions of the genome are conserved by evolution?

Genomics: tool for medicine

“The branch of molecular genetics concerned with the study of genomes, specifically the identification and sequencing of their constituent genes and the **application** of this knowledge in **medicine, pharmacy, agriculture**, etc.”

Collins English Dictionary

How is genotype related to health phenotypes?

What's the difference between DNA in a tumor vs DNA in healthy tissue?

Can genomic data help predict what drugs might be appropriate for:

- a particular cancer patient?
- a particular genetic disorder?

Can genomic data reveal weaknesses in the defenses of pathogens?

Can genomic data help us predict what flu strains will prevail next year?

Computational Genomics

Addresses crucial problems at the intersection of genomics and computer science

The intersection:

Key biological models are straight out of computer science:
circuits and **networks** for molecular interactions, **trees** for evolution and pedigrees, **strings** for DNA, RNA and proteins

Thanks to sequencers and microarrays, research bottlenecks increasingly hinge on computational issues: **speed, scalability, energy, cost**

With large, noisy, biased high-throughput datasets comes a critical need for **machine learning** and **statistical reasoning**

Computational Genomics: computation

How to efficiently analyze the huge quantities of fragmentary evidence that come from DNA sequencers

How to model biological phenomena and make predictions

How to combine data from disparate datasets to reach new conclusions in the presence of error and systematic bias

How to store huge quantities of data economically and securely while also allowing it to be queried

How to visualize large, complicated datasets

Draws on: Algorithms, data structures, pattern matching, indexing, compression, information retrieval, distributed and parallel computing, cloud computing, machine learning, ...

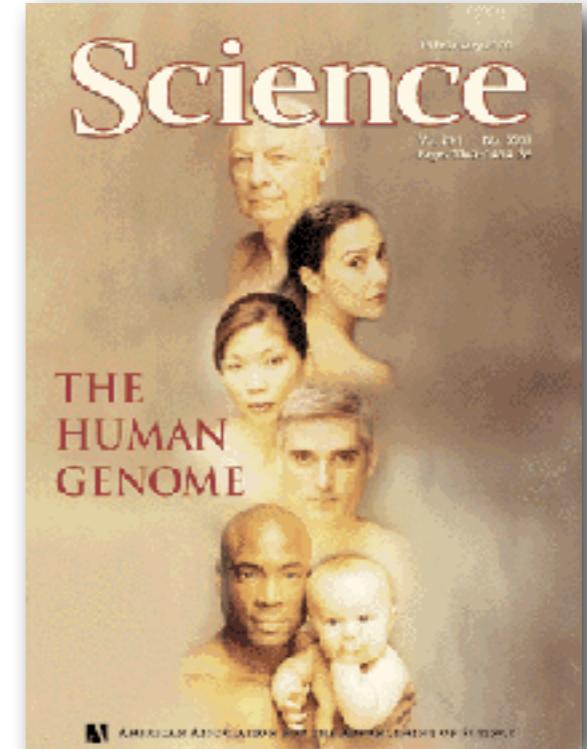
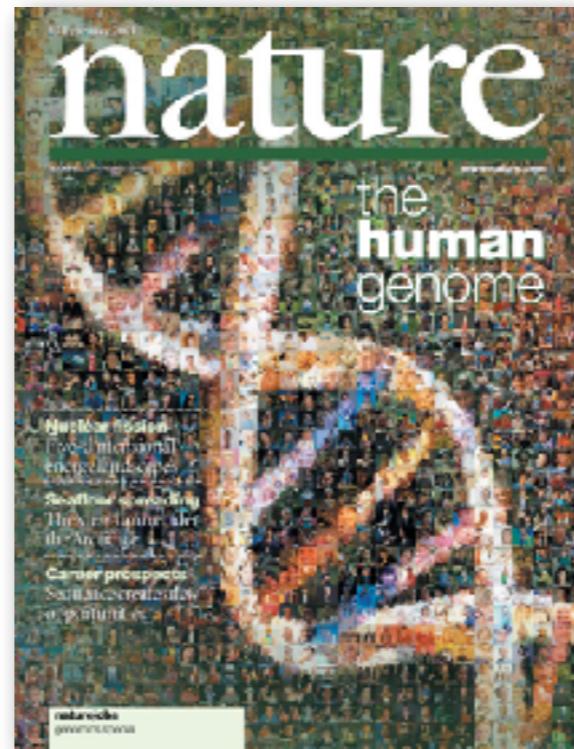
Computational Genomics: success stories

The screenshot shows the NCBI BLAST Standard Nucleotide BLAST search interface. At the top, there's a navigation bar with links for Home, Recent Results, Saved Strategies, Help, My NCBI (Sign In/Register), and a search bar. Below the navigation bar, it says 'NCBI/BLAST/blastn suite' and 'Standard Nucleotide BLAST'. There are tabs for blastn, blastp, blastx, tblastn, and tblastx, with 'blastn' selected. A sub-header states 'BLASTN programs search nucleotide databases using a nucleotide query.' with links for 'more...', 'Reset page', and 'Bookmark'. The main form area has a section for 'Enter Query Sequence' with a text input field and a 'Clear' button. To the right, there's a 'Query subrange' section with 'From' and 'To' input fields. Below that, there's an 'Or, upload file' section with a 'Choose File' button showing 'No file chosen'. There's also a 'Job Title' section with a text input field and a 'Choose Search Set' section with a 'Database' dropdown menu set to 'Nucleotide collection (nr/nt)'. Other options in the 'Choose Search Set' section include 'Human genomic + transcript', 'Mouse genomic + transcript', and 'Others (nr etc.)'.

http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome

The BLAST sequence alignment program is a hugely successful tool, a fixture of biological analysis and cited over 50,000 times

Computational Genomics: success stories



The Human Genome Project depended crucially on contributions by computer scientists, especially new methods for assembling DNA fragments into chromosomes.

Computational Genomics: success stories

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Whole-Genome Sequencing in a Patient
with Charcot–Marie–Tooth Neuropathy

NATURE REVIEWS | GENETICS

APPLICATIONS OF NEXT-GENERATION SEQUENCING

Advances in understanding
cancer genomes through
second-generation sequencing

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

The Origin of the Haitian Cholera
Outbreak Strain

the guardian

News > Science > Genetics

Mayo Clinic plans to sequence patients'
genomes to personalise care

Project will give doctors the genetic information they need to
choose drugs that work best and minimise side effects

The idea of using high-throughput DNA sequencing in medical settings
is only possible because of novel, extremely efficient software
developed in the years after second-generation sequencers arrived.