

Why genomics?

Steven Salzberg

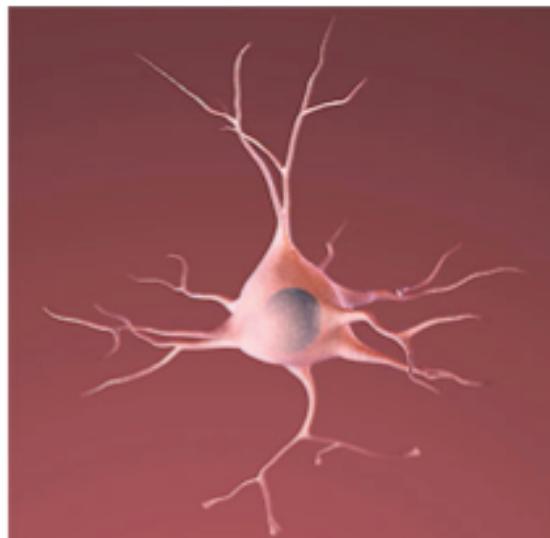
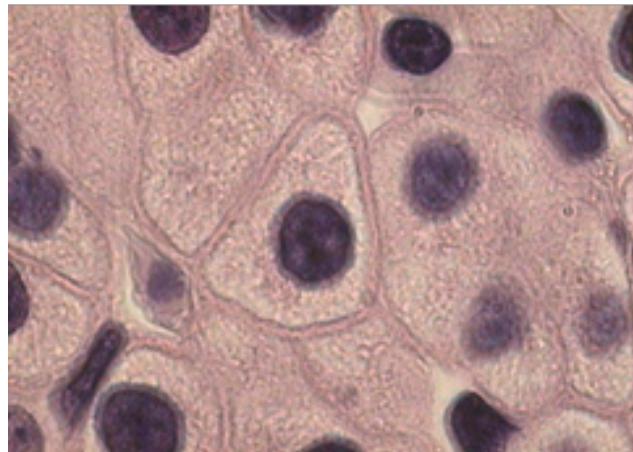
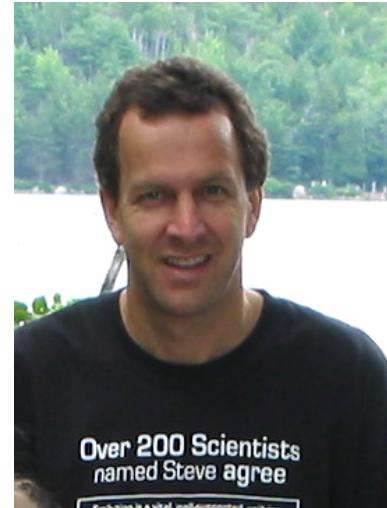


What makes us diverse?

A black and white photograph capturing a massive outdoor gathering of people. The crowd is dense, with many individuals facing towards the left side of the frame, suggesting they are watching a performance or event. In the background, a prominent city skyline with numerous skyscrapers rises against a clear sky. A body of water is visible between the crowd and the city. On the right side of the image, there's a dark truck or trailer with some text and graphics on it. The overall atmosphere is one of a public, daytime event.

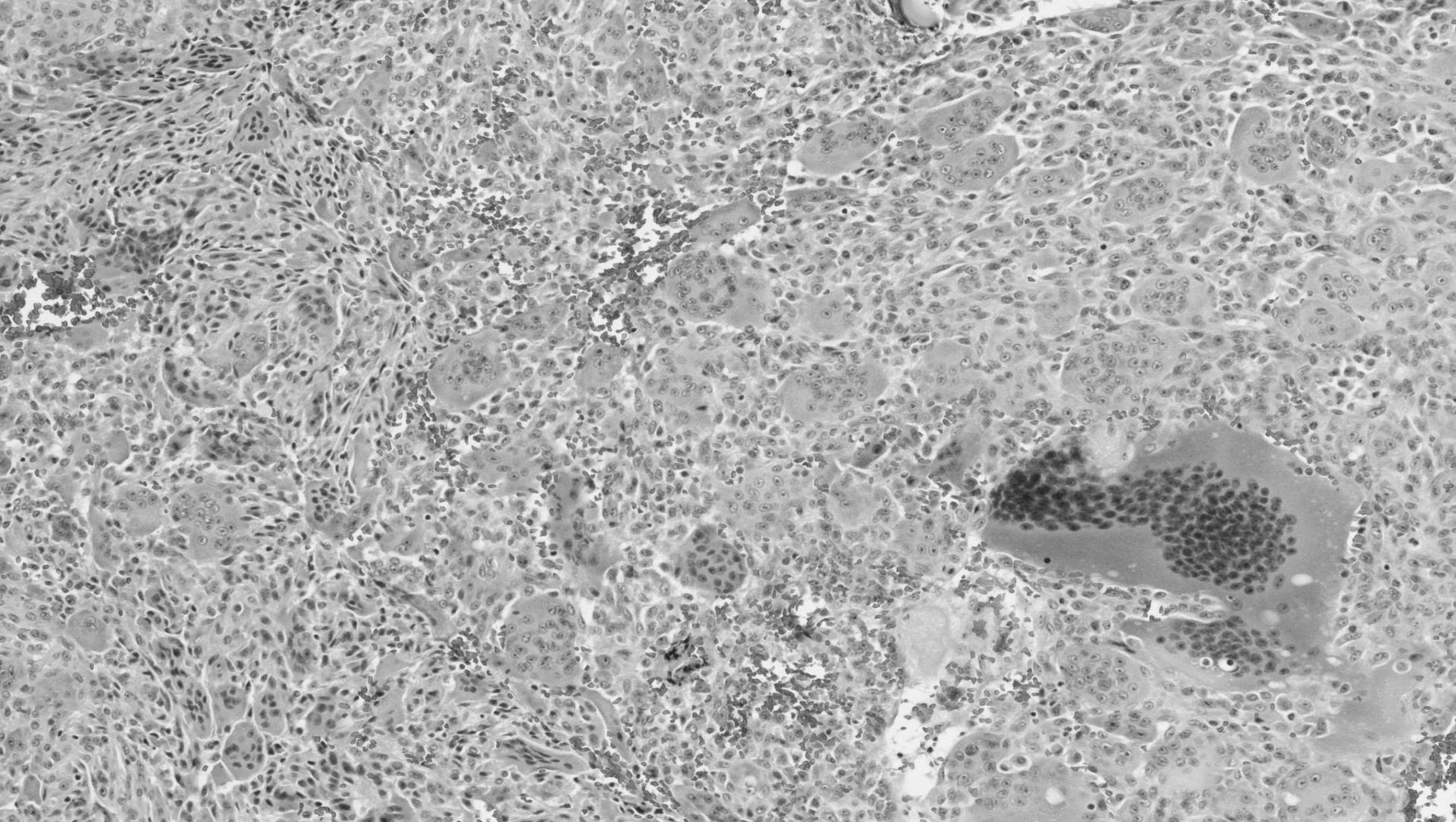
99.9% identical

How do our bodies develop?



slide adapted from Rafael Irizarry

How does a healthy cell become a
cancer cell?



What do our genes have to do with it?



The central dogma tells us how
information flows in a cell

DNA

ACTGACCTAGATCAGTGTAGCGATCGTATA CGAGACCGATT CATCGGCAT



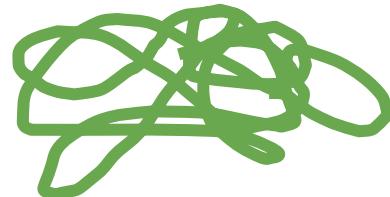
transcription

RNA

AUCAGUCGAUCACCGAU



translation



protein

But it is a little more complicated
than that

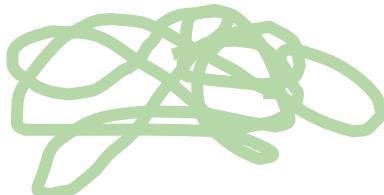
DNA

ACTGACCGTTCGATCGATCGTATA
GTCGATCGATCGTATAAGATTACAAAATCATCGGCAT

RNA

AUCAGUCGAUCACCGAU

protein



transcription



translation

How do we make genomic measurements?

Why genomics (part 2)?

Steven Salzberg



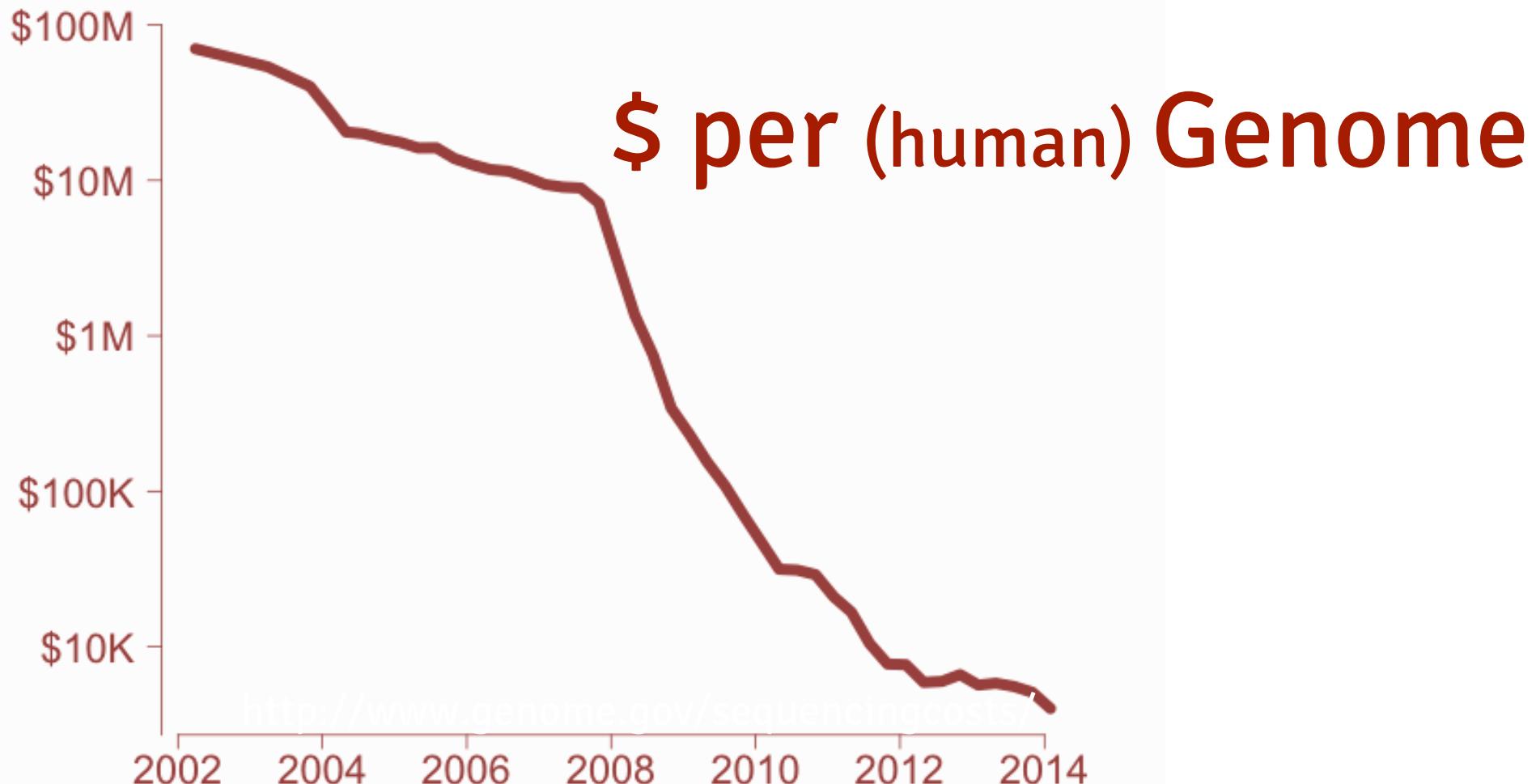


Image via Illumina: <http://goo.gl/y0M01y>

What does the data look like?

@HWI-EAS121:4:100:1783:550#0/1 http://brianknaus.com/software/srtoolbox/s_4_1_sequence80.txt
CGTTACGAGATCGGAAGAGCGGGTCAGCAGGAATGCCGAGACCGATCTCGTATGCCGTCTGCTTCCGTACAAACACACGGC
+HWI-EAS121:4:100:1783:550#0/1
aaaaaa`b_aa`aa`YaX]aZ`aZM^Z]YRa]YSG[[ZREQLHESDHNDDHNMEEDMPENITKFLFEEDDDHEJQMEDDD
@HWI-EAS121:4:100:1783:1611#0/1
GGGTGGGCATTCCACTCGCAGTATGGGTTGCCGCACCGACAGGCAGCGGTAGCCTGCGCTTGGCCTTCGGAAA
+HWI-EAS121:4:100:1783:1611#0/1
a``^_`_``^a``a``a_`_]a_]`a_____`_``^`]X]_XTV_\\]]NX_XVX]]_TTTG[VTHPN]VFDZ
@HWI-EAS121:4:100:1783:322#0/1
CGTTTATGTTTTGAATATGTCTTATCTAACGGTTATTTAGATGTTGGTCTTATTCTAACGGTCATATATTTCTA
+HWI-EAS121:4:100:1783:322#0/1
abaa``aaaaabbbaabbbbbbb`bbbb_bbbbbbbb`bbbaV^_a``a``]``at]a__V\\]]_``^a`]a_abbaV__
@HWI-EAS121:4:100:1783:1394#0/1
GGCTTTATTGGTCTGGTATCCCCCATATTCTCCGGTTGTGGTTAACCGATCATCCGCATTACTCCCCGCTGC
+HWI-EAS121:4:100:1783:1394#0/1
````[aa\b``^[]aabbb][`a\_abbb`a``bbbbbabaaaab\_Vza\_``bab\_X`[a\HV[\_]\_[^\_X\T\_VQQ  
@HWI-EAS121:4:100:1783:207#0/1  
CCCTGGAGATCGGAAGAGCGGGTCAGCAGGAATGCCGAGACCGATCTCGTATGCCGTCTGCTTGAACAAAAAAACA  
+HWI-EAS121:4:100:1783:207#0/1  
abba`Xa``\\`aa]ba\_\_bba[a\_O\_a`aa`aa`a]^V]X\_a^YS\R\_\H\_[]\ZTDUZZUSOPX]]POP\GS\WSHHD  
@HWI-EAS121:4:100:1783:455#0/1  
GGGTAATTCAAGGGACAATGTAATGGCTGCACAAAAAAATACATCTTCATGTTCCATTGCACCATTGACAAATACATATT  
+HWI-EAS121:4:100:1783:455#0/1  
abb\_babbabaabbbbbbbbbaabbba`b`\abbabbabbabbbaabbbbb`bb`ab\_O\_bab\_Q\_bbabaa\_a

# Why is genomic data science hot now?



# How much data is there?



# SRA

Sequence Read Archive (SRA) makes biological sequence data available to the research community to enhance reproducibility and allow for new discoveries by comparing data sets. The SRA stores raw sequencing data and alignment information from high-throughput sequencing platforms, including Roche 454 GS System®, Illumina Genome Analyzer®, Applied Biosystems SOLiD System®, Helicos Heliscope®, Complete Genomics®, and Pacific Biosciences SMRT®.

## Getting Started

[Understanding and Using SRA](#)[How to Submit](#)[Login to Submit](#)[Download Guide](#)

## Tools and Software

[Download SRA Toolkit](#)[SRA Toolkit Documentation](#)[SRA-BLAST](#)[SRA Run Browser](#)[SRA Run Selector](#)

## Related Resources

[dbGaP Home](#)[Trace Archive Home](#)[BioSample](#)[GenBank Home](#)