Welcome to Biomedical Research!

Michael Schatz

August 31, 2017 – Lecture 1
EN.601.452 Computational Biomedical Research
AS.020.415 Advanced Biomedical Research

Welcome!

The goal of this course is to prepare undergraduates to understand and perform state-of-the-art biomedical research. This will be accomplished through three main components:

1. **Lectures** on cross cutting techniques for biomedical research focusing on data visualization, statistical inference, and scientific computing
2. **Research presentations** from distinguished faculty on their active research projects
3. **A major research project** to be performed under the mentorship of a JHU professor.

Course Webpage:	https://github.com/schatzlab/biomedicalresearch
Course Discussions:	http://piazza.com
Class Hours:	Mon + Wed @ 3p – 3:50p Malone 107
Office Hours:	Wed @ 4-5p and by appointment Please try Piazza first!

Prerequisites and Resources

Prerequisites

- No formal course requirements
- Access to an Apple or Linux Machine, or Install VirtualBox
- Familiarity with a major programming language is nice but not essential
  - C/C++, Java, R, Perl, Python, JavaScript, others?

Primary Texts

- None! We will be studying primary research papers

Other Resources:

- Google, SEQanswers, Biostars, StackOverflow

- Applied Computational Genomics Course
  - https://github.com/schatzlab/appliedgenomics

- Ben Langmead's teaching materials:
  - http://www.langmead-lab.org/teaching-materials/

Grading Policies

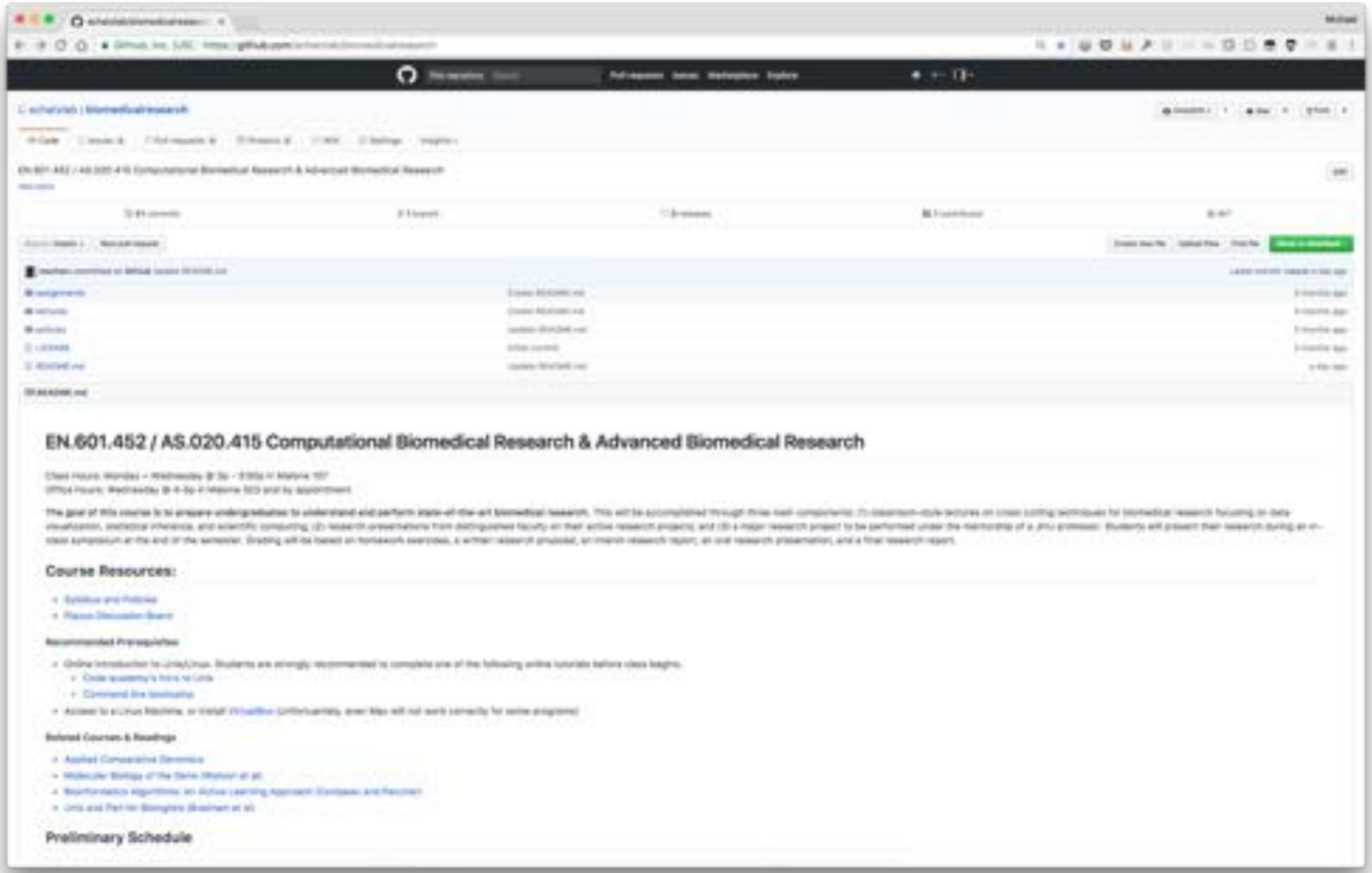
Assessments:

- ~6 HW Exercises: 10% Due at 11:59pm a week later
- Research Proposal: 10% ~1 page write up
- Interim Report: 10% ~3 page progress report
- Project Presentation: 30% Presented last week of class
- Final Report: 30% Due last week of semester
- In-class Participation: 10% Please ask questions!

Policies:

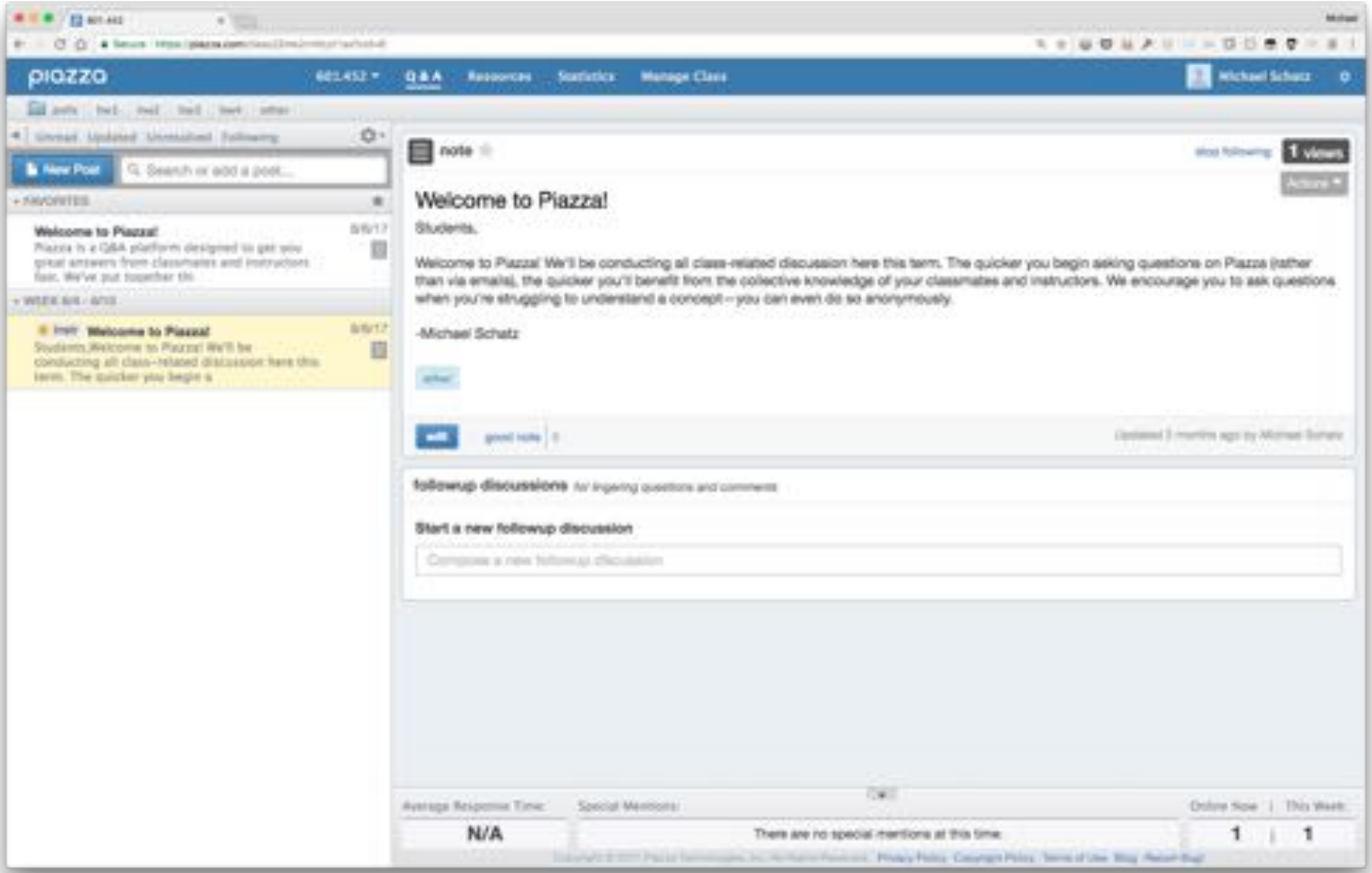
- Scores assigned relative to the highest points awarded
- Automated testing and grading of analysis assignments
- ***Late Days:***
  - Four (4) chances to extend the deadline for assignments by 24 hours without any penalty, after that 25% deduction per day

Course Webpage



https://github.com/schatzlab/biomedicalresearch

Piazza



http://piazza.com/jhu/fall2017/601452/home

A Little About Me

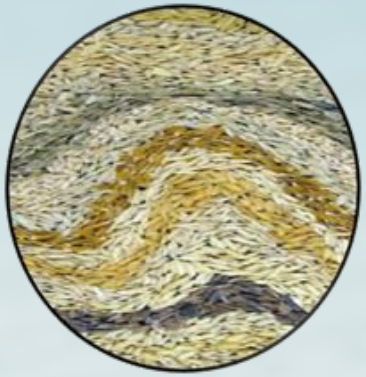
A Little About Me

Schatzlab Overview

Human Genetics

Role of mutations in disease

Feigin *et al.* (2017)
Fang *et al.* (2016)



Agricultural Genomics

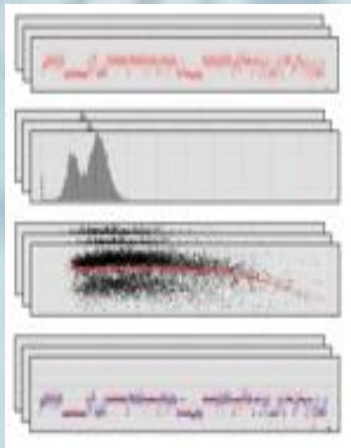
Genomes & Transcriptomes

Lemmon *et al.* (2016)
Ming *et al.* (2015)

Algorithmics & Systems Research

Ultra-large scale biocomputing

Stevens *et al.* (2015)
Marcus *et al.* (2014)



Biotechnology Development

Single Cell + Single Molecule Sequencing

Chin *et al.* (2016)
Garvin *et al.* (2015)

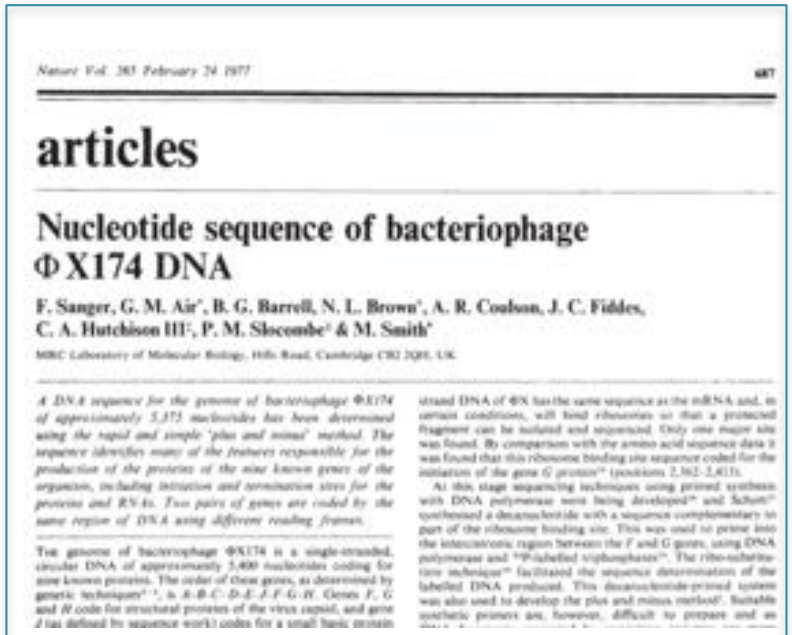
DNA: The secret of life

Your DNA, along with your environment and experiences, shapes who you are

- Height
- Hair, eye, skin color
- Broad/narrow, small/large features
- Susceptibility to disease
- Response to drug treatments
- Longevity and cognition

Physical traits tend to be strongly genetic, social characteristics tend to be strongly environmental, and everything else is a combination

The Origins of DNA Sequencing



articles

Nucleotide sequence of bacteriophage ΦX174 DNA

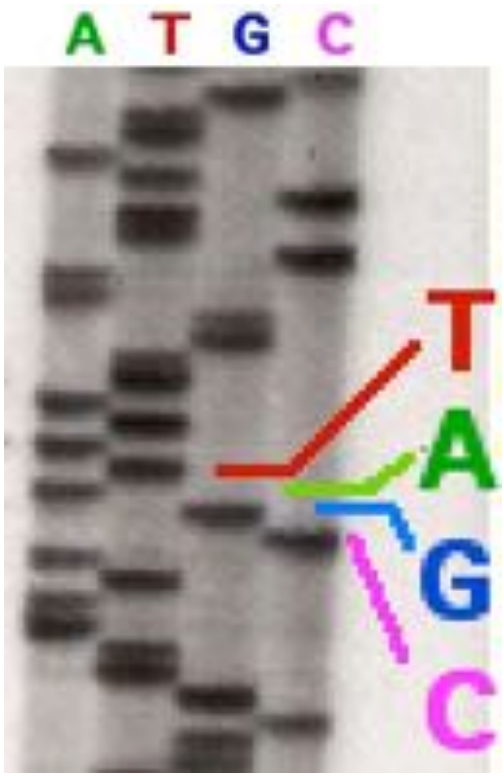
F. Sanger, G. M. Air, B. G. Barrell, N. L. Brown, A. R. Coulson, J. C. Fiddes, C. A. Hutchison III, P. M. Slocombe & M. Smith

1977

1st Complete Organism

Bacteriophage ϕX174

5375 bp



Radioactive Chain Termination

5000bp / week / person

http://en.wikipedia.org/wiki/File:Sequencing.jpg

http://www.answers.com/topic/automated-sequencer

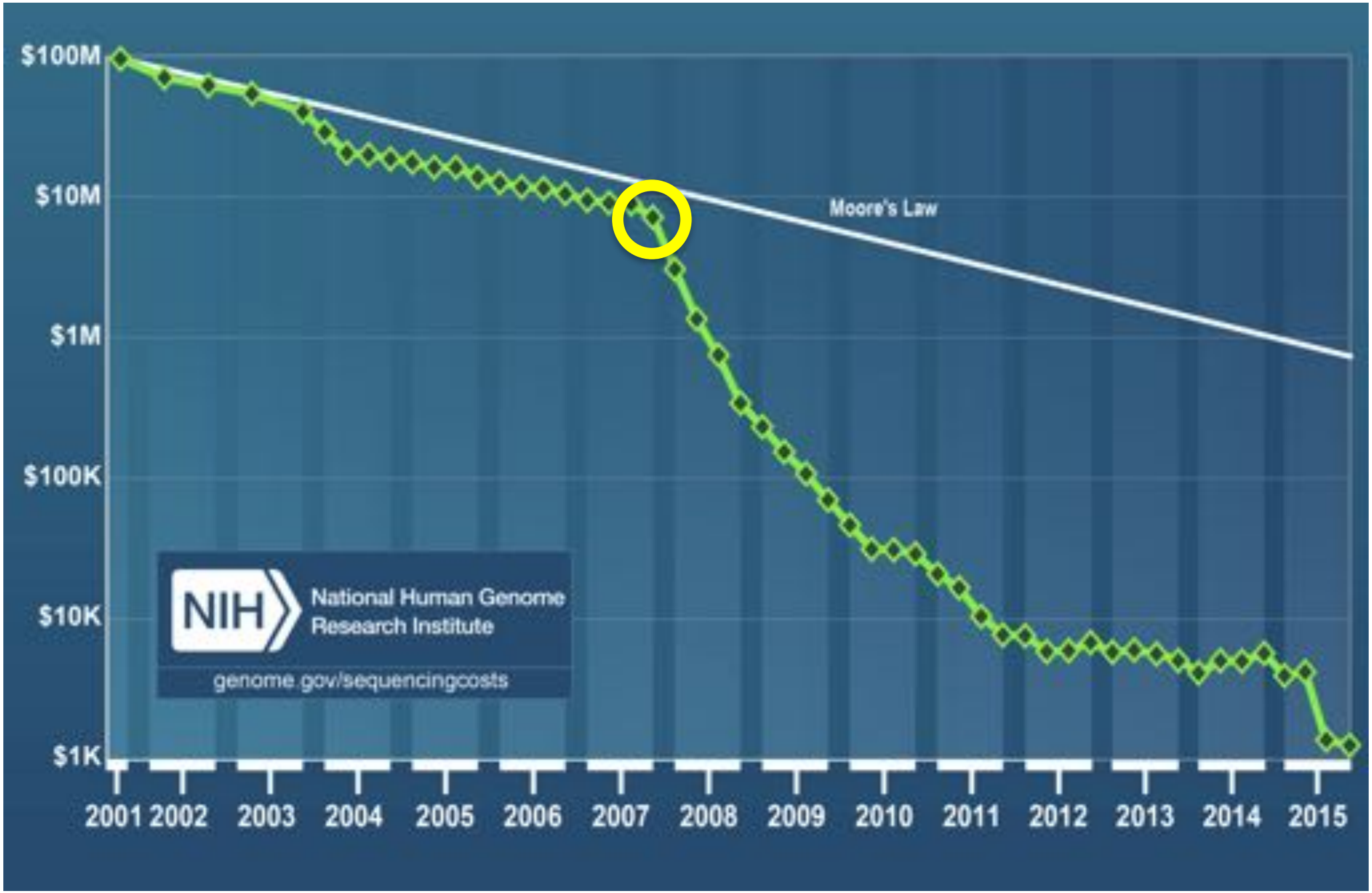
Nucleotide sequence of bacteriophage φX174 DNA

Sanger, F. et al. (1977) *Nature. 265: 687 - 695*

Milestones in DNA Sequencing

(TIGR/Celera, 1995-2001)

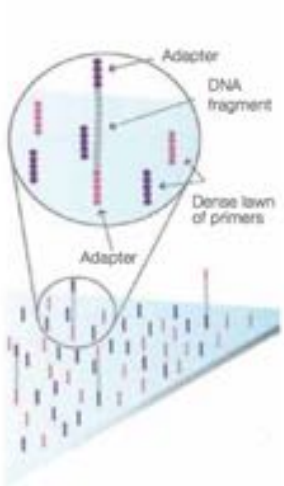
Cost per Genome



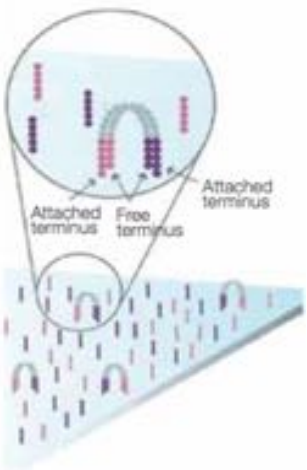
Massively Parallel Sequencing

Illumina HiSeq 2000
Sequencing by Synthesis

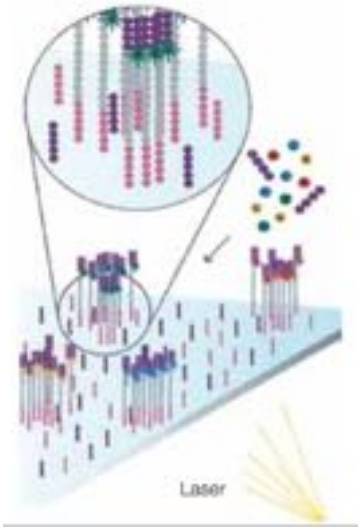
>60Gbp / day



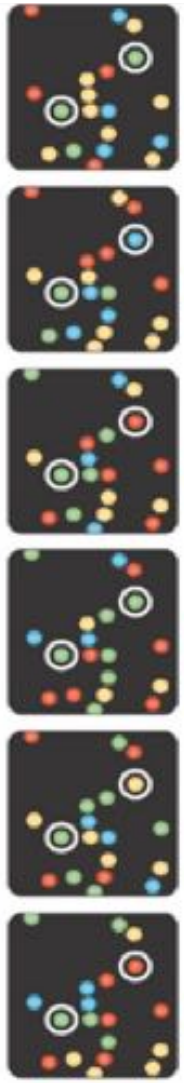
1. Attach



2. Amplify

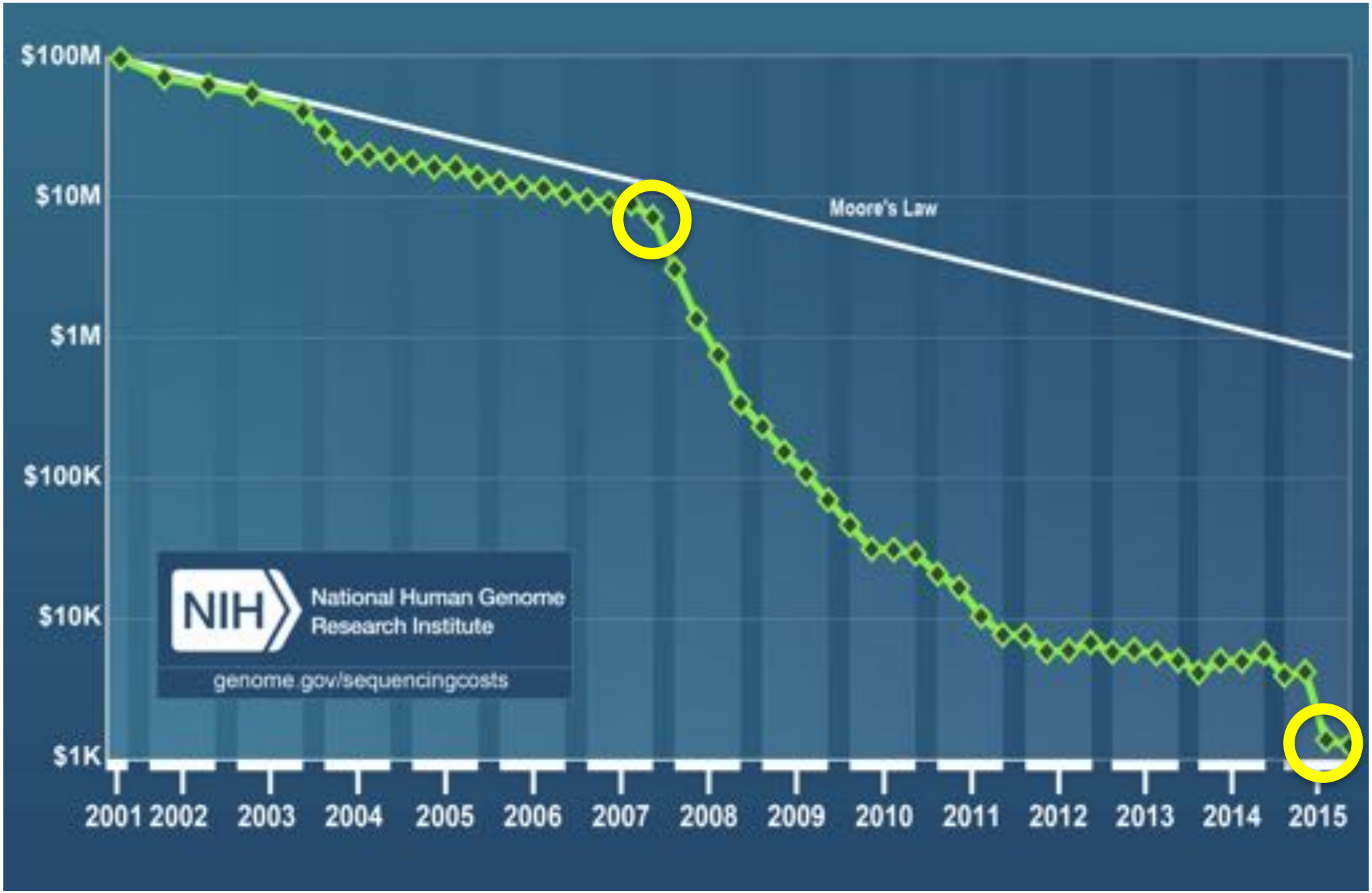


3. Image



Metzker (2010) Nature Reviews Genetics 11:31-46
http://www.youtube.com/watch?v=I99aKKHcxC4

Cost per Genome



HiSeq X Ten

BROAD INSTITUTE

GARVAN INSTITUTE

NEW YORK GENOME CENTER

deCODE genetics

MACROGEN Axeq Technologies

Novogene 诺禾致源

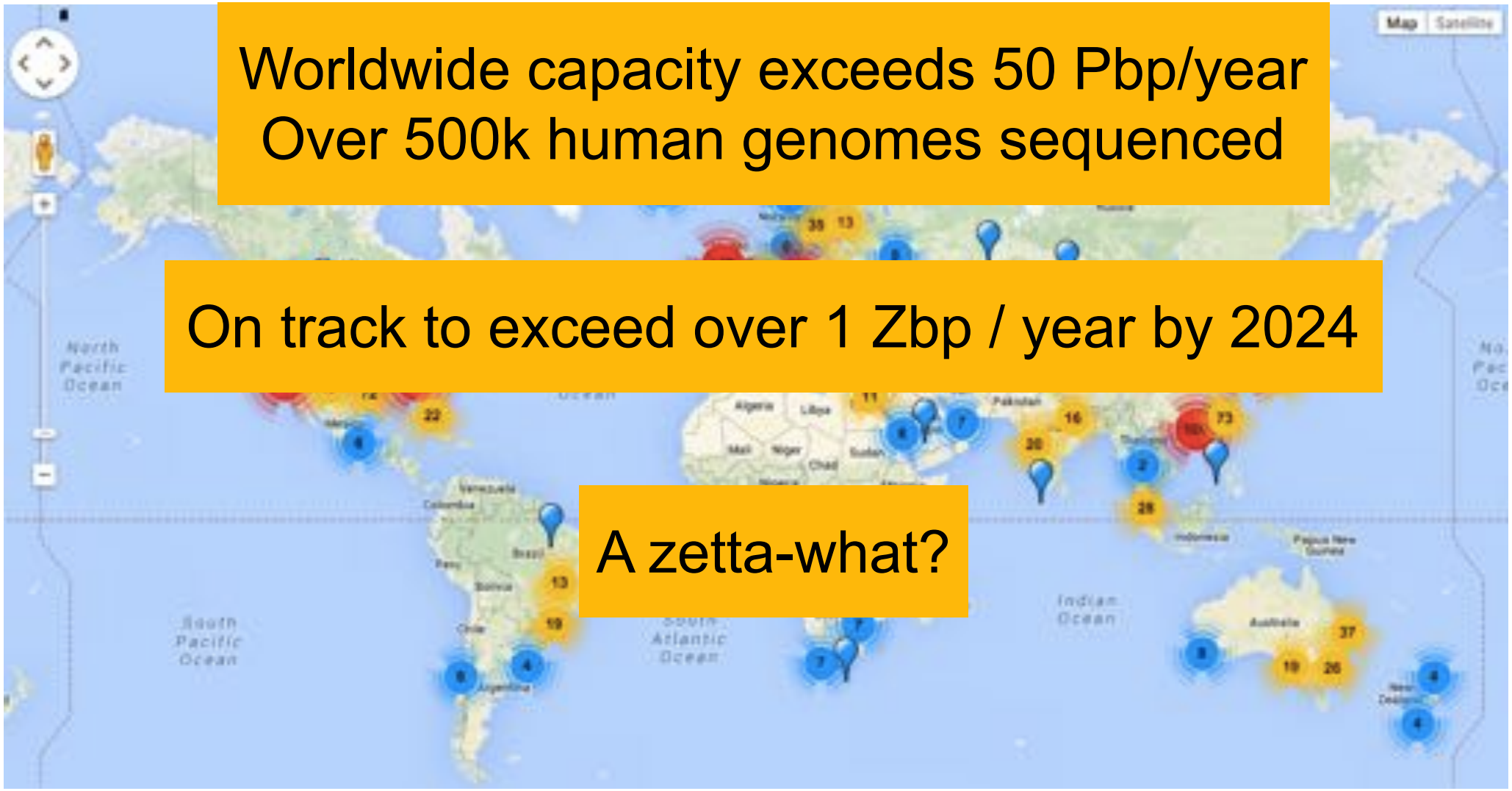
药明康德 WuXi PharmaTech

wellcome trust sanger institute

HUMAN LONGEVITY, INC.

320 genomes per week / 18,000 genomes per year
$1000 per genome / ~$10 M per instrument

Sequencing Centers



Worldwide capacity exceeds 50 Pbp/year
Over 500k human genomes sequenced

On track to exceed over 1 Zbp / year by 2024

A zetta-what?

Next Generation Genomics: World Map of High-throughput Sequencers
http://omicsmaps.com

How much is a zettabyte?

Unit	Size
Byte	1
Kilobyte	1,000
Megabyte	1,000,000
Gigabyte	1,000,000,000
Terabyte	1,000,000,000,000
Petabyte	1,000,000,000,000,000
Exabyte	1,000,000,000,000,000,000
Zettabyte	1,000,000,000,000,000,000,000

How much is a zettabyte?

100 GB / Genome
4.7GB / DVD
~20 DVDs / Genome

X

10,000,000,000 Genomes

=

1ZB Data
200,000,000,000 DVDs

150,000 miles of DVDs
~ ½ distance to moon

Both currently ~100Pb
And growing exponentially

Unsolved Questions in Biology

- What is your genome sequence?
-
-
-
-
-
-
-
-
-
-
-
- ***Plus thousands and thousands more***

The instruments provide the data, but none of the answers to any of these questions.

What software and systems will?

And who will create them?

Next Steps

1. Reflect on the magic and power of DNA ☺
2. Check out the course webpage
3. Register on Piazza & say hello!
4. Set up Dropbox for yourself!


Welcome to Advanced Biomedical Research
https://github.com/schatzlab/biomedicalresearch

Questions?