Bold Prediction #8: A person's complete genome sequence along with informative annotations can be securely and readily accessible on their smartphone

Michael Schatz

Oct 4, 2021

NHGRI Bold Predictions Seminar Series

@mike_schatz

Perspective

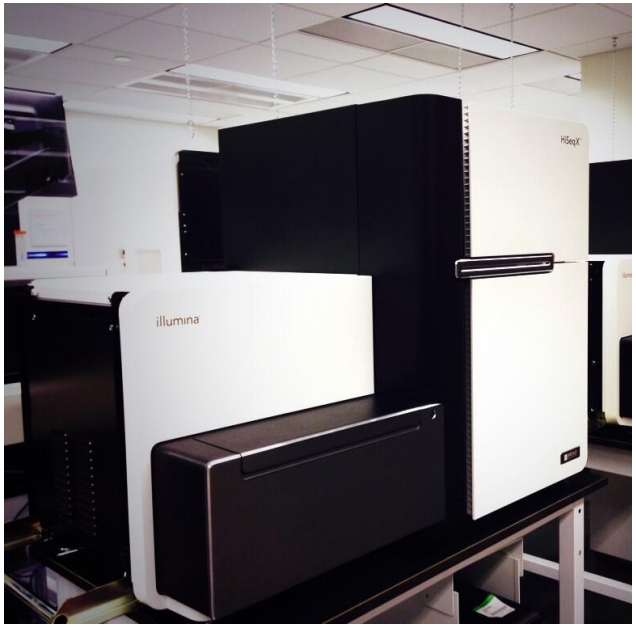
Genomic Futures?

Miniaturization of Sequencing

1st Generation Sequencing
TIGR Sequencing Lab
1995-2010

Miniaturization of Sequencing

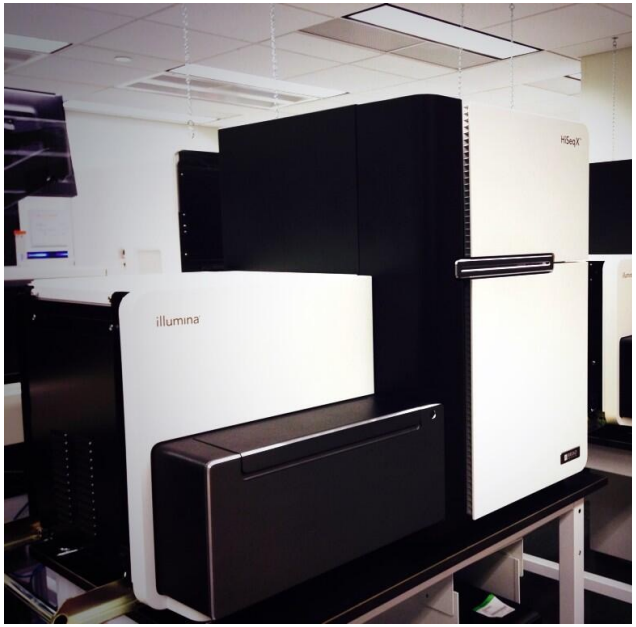
1st Generation Sequencing
TIGR Sequencing Lab
1995-2010



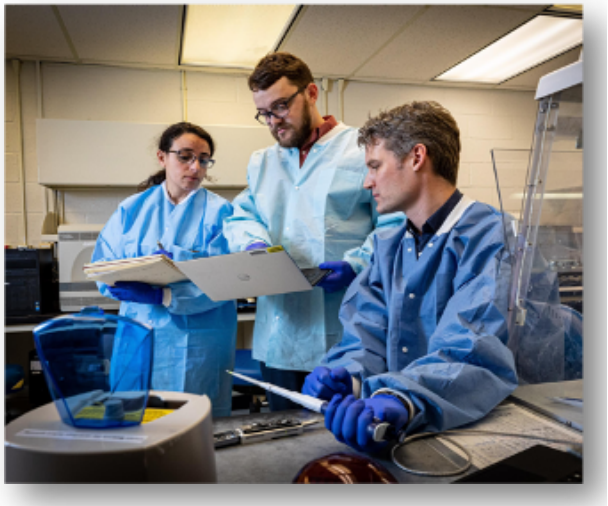
2nd Generation Sequencing
NYGC Sequencing Lab
2010-

Miniaturization of Sequencing

1st Generation Sequencing
TIGR Sequencing Lab
1995-2010



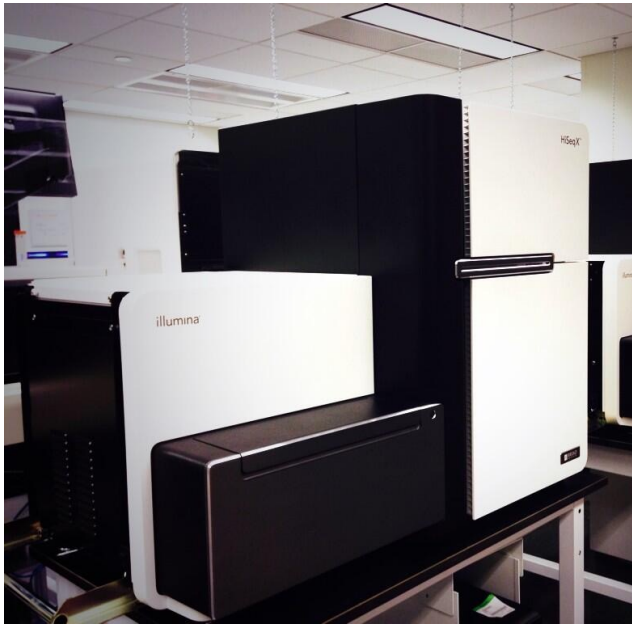
2nd Generation Sequencing
NYGC Sequencing Lab
2010-



3rd Generation Sequencing
JHU Pathology Lab
2020

Miniaturization of Sequencing

1st Generation Sequencing
TIGR Sequencing Lab
1995-2010

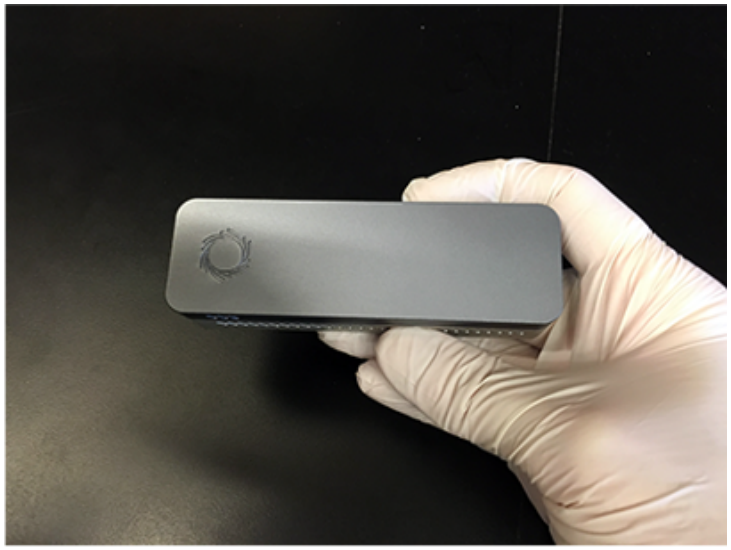


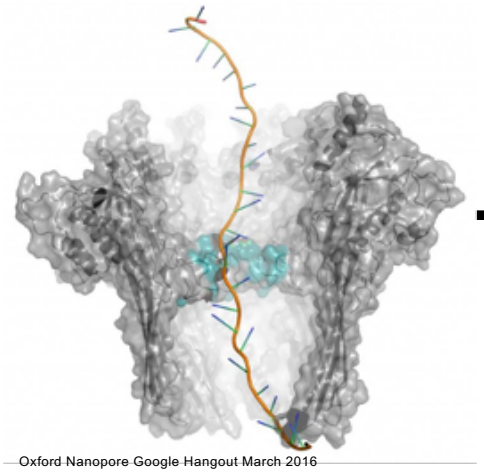
2nd Generation Sequencing
NYGC Sequencing Lab
2010-

3rd Generation Sequencing
JHU Pathology Lab
2020

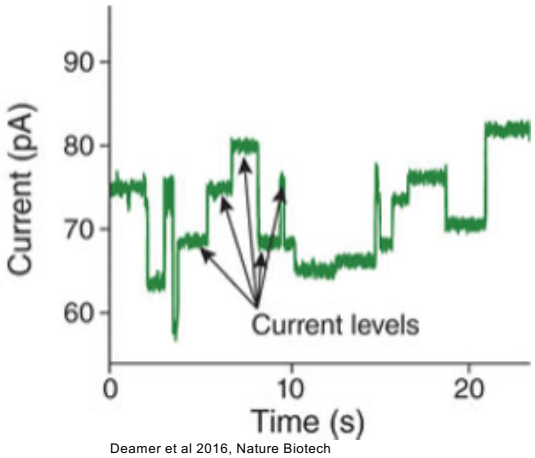
3rd Generation Sequencing: Nanopore Sequencing

- Oxford Nanopore Technologies
- No theoretical upper limit to sequencing read length, practical limit only in delivering DNA to the pore intact
- Highly mobile, palm sized sequencer
- Typical sequencing output ~20Gb





Oxford Nanopore Google Hangout March 2016



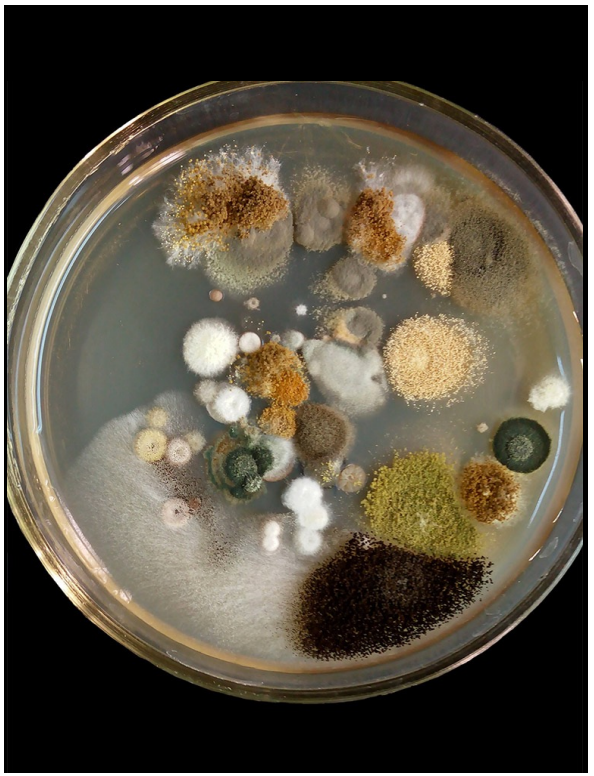
Deamer et al 2016, Nature Biotech

ATCGATCGATAGTAT
TAGATACGACTAGCG
ATCAG…

3rd Generation Sequencing

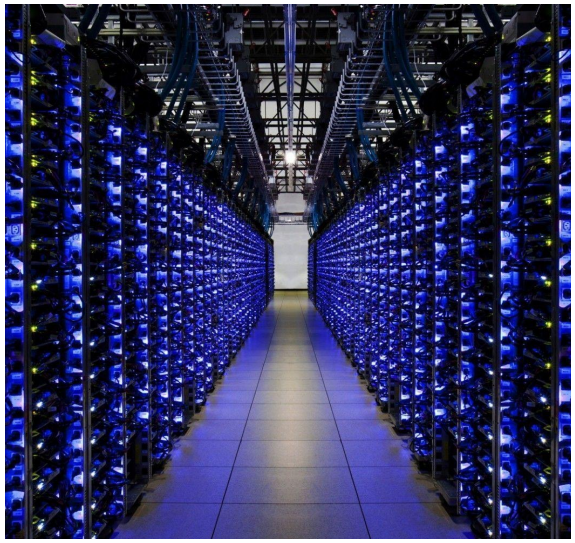
(Nattestad *et al.*, 2018, *Genome Research*)

(Alonge *et al.,* 2020,*Cell*)



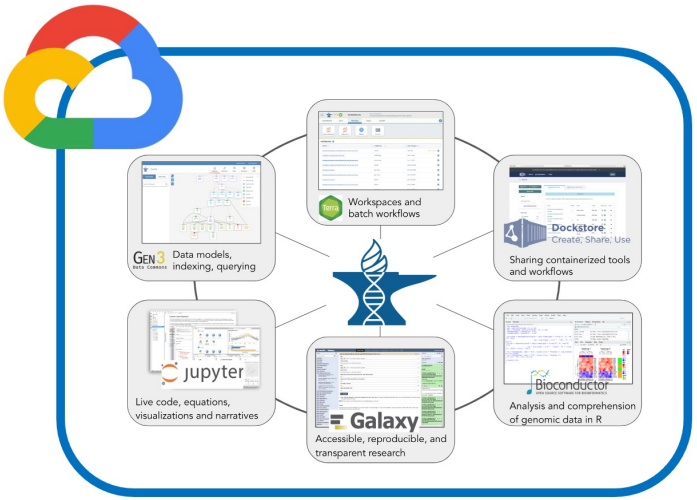
(Kovaka *et al.,* 2020, Nature Biotechnology)

Democratization of computing

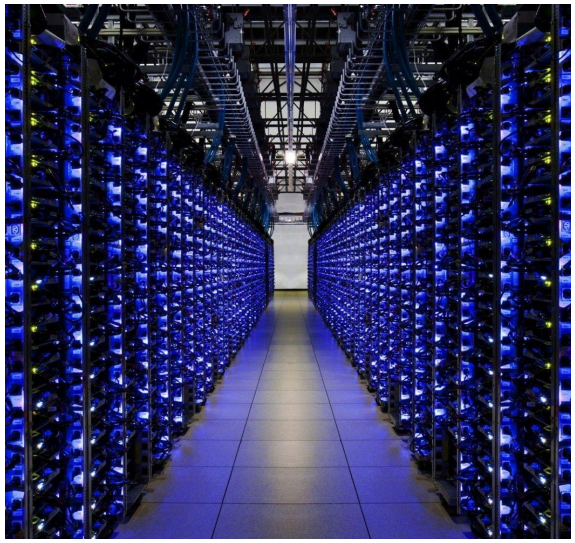


Institutional Computing
JHU Data Center

Democratization of computing

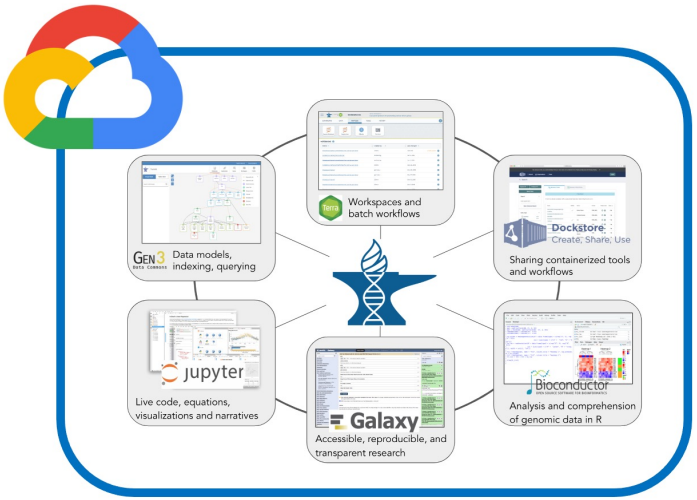


Cloud Computing
NHGRI AnVIL

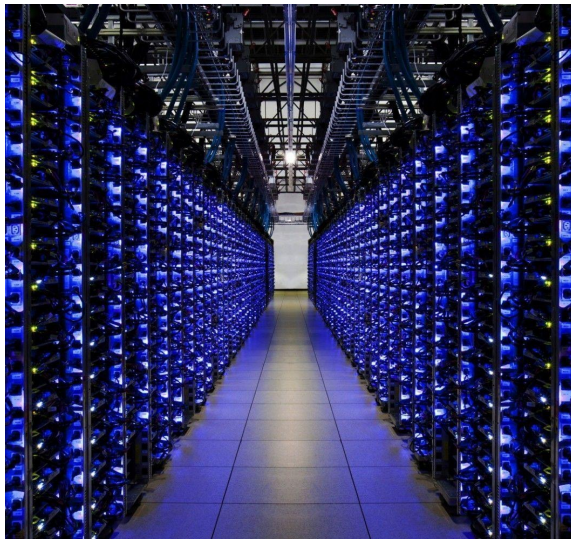


Institutional Computing
JHU Data Center

Democratization of computing



Cloud Computing
NHGRI AnVIL

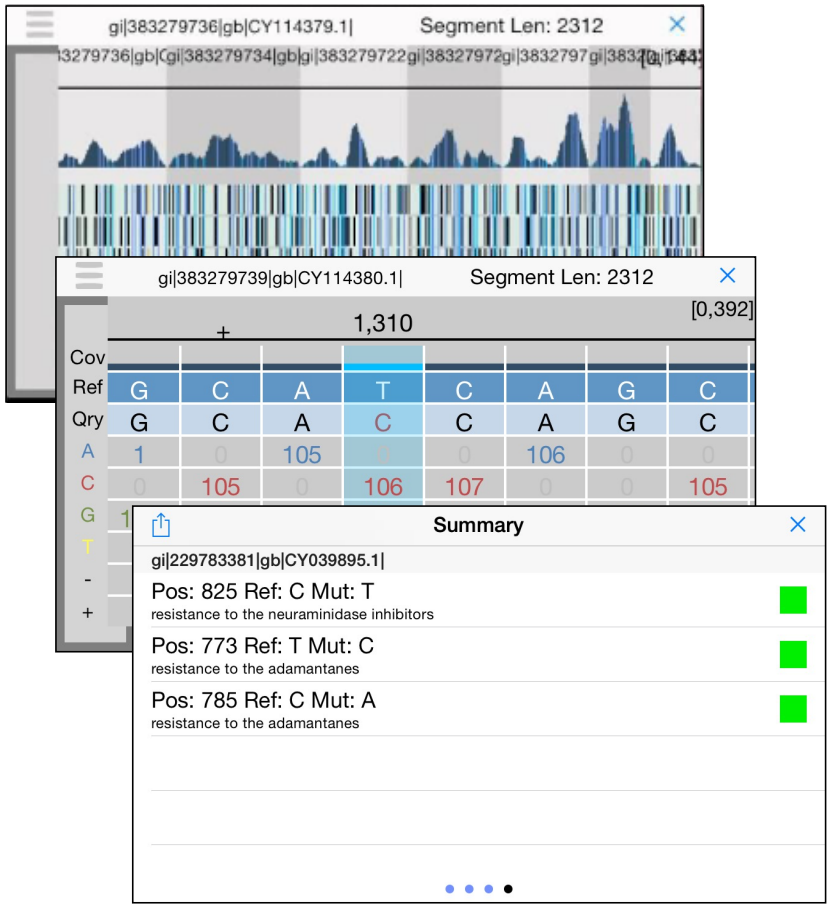


Institutional Computing
JHU Data Center

Mobile Computing
Aspyn Palatnick ~ iGenomics

iGenomics: Comprehensive DNA sequence analysis on your Smartphone

Aspyn Palatnick, Bin Zhou, Elodie Ghedin, Michael Schatz (2020) GigaScience. doi: 10.1093/gigascience/giaa138



The worlds first genomics analysis app for iOS devices

BWT + Dynamic Programming + UI

First application:

- Handheld diagnostics and therapeutic recommendations for influenza infections

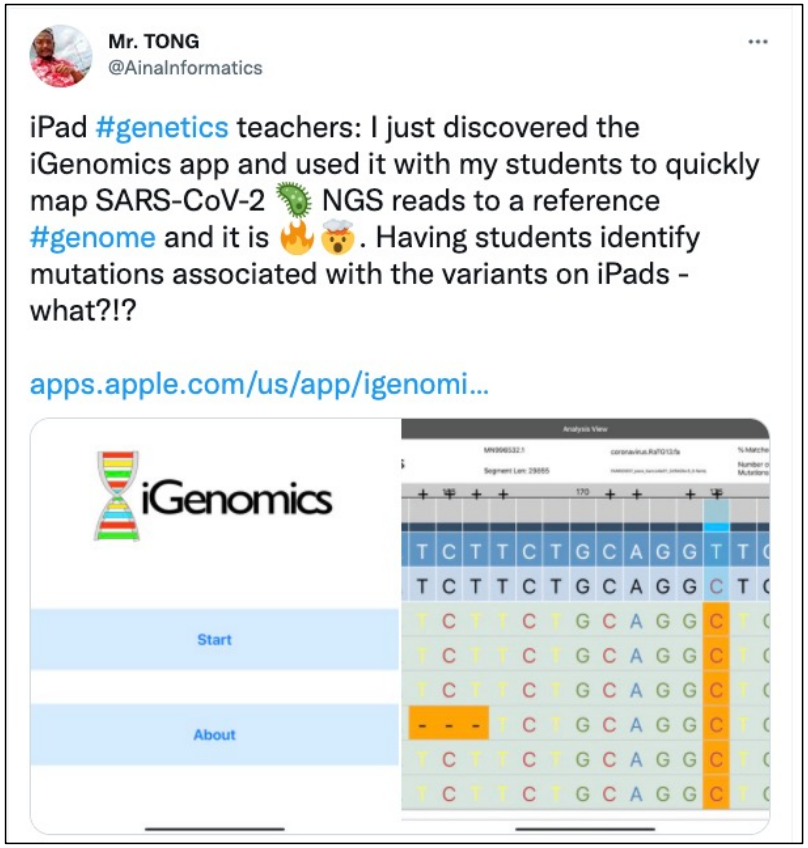
- Available in the iOS App Store
  - https://apple.co/2HCplzr

Future applications

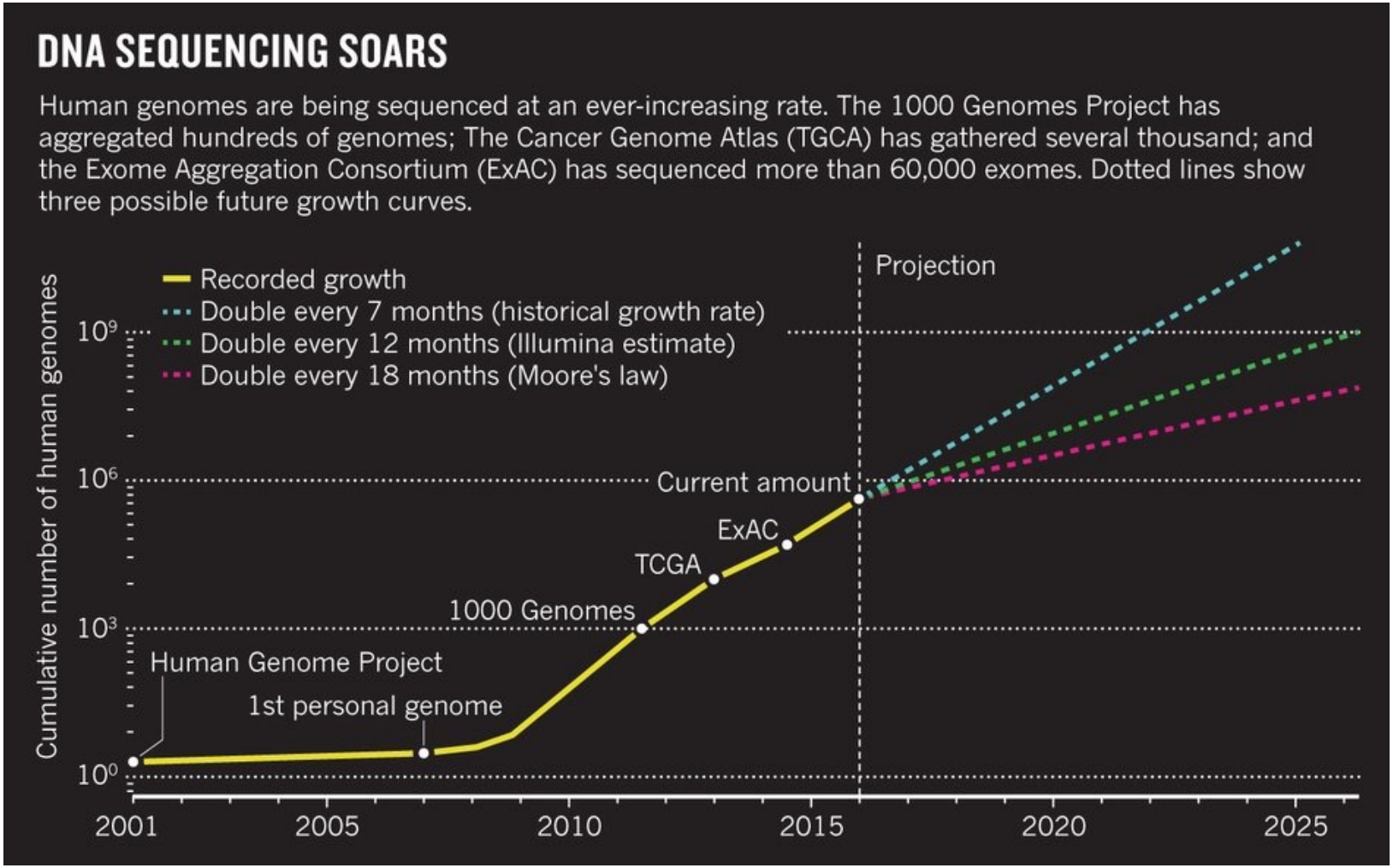
- Pathogen detection
- Food safety
- Biomarkers
- etc..

iGenomics: Comprehensive DNA sequence analysis on your Smartphone

Aspyn Palatnick, Bin Zhou, Elodie Ghedin, Michael Schatz (2020) GigaScience. doi: 10.1093/gigascience/giaa138



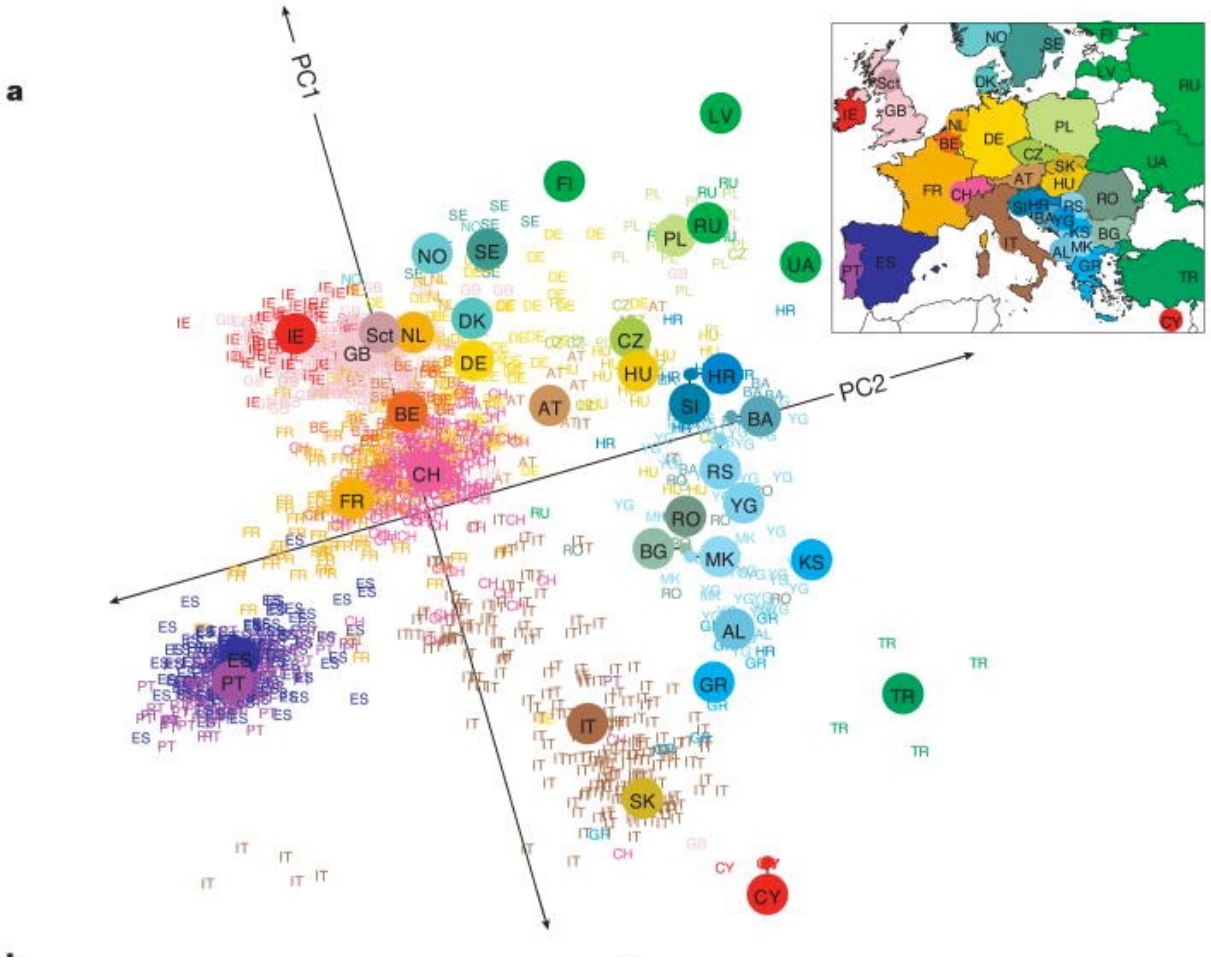
Research



Big Data: Astronomical or Genomical?

Stephens, Z, et al. (2015) PLOS Biology DOI: 10.1371/journal.pbio.1002195

What will we do with all this sequencing power?



Genes mirror geography within Europe
Novembre et al (2008) Nature. doi: 10.1038/nature07331

MyHeritage

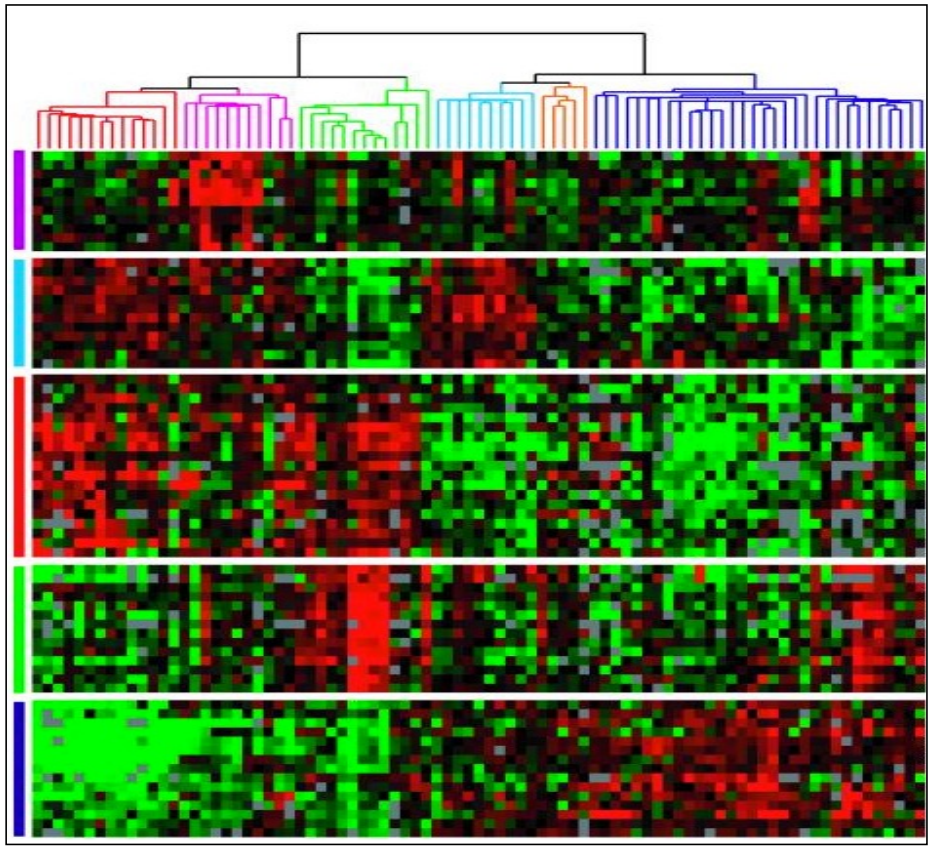
>50M customers
https://en.wikipedia.org/wiki/MyHeritage

ancestryDNA™

>15M customers
https://bit.ly/ancestryDNACustomers

23andMe

>12M customers
https://medical.23andme.com/

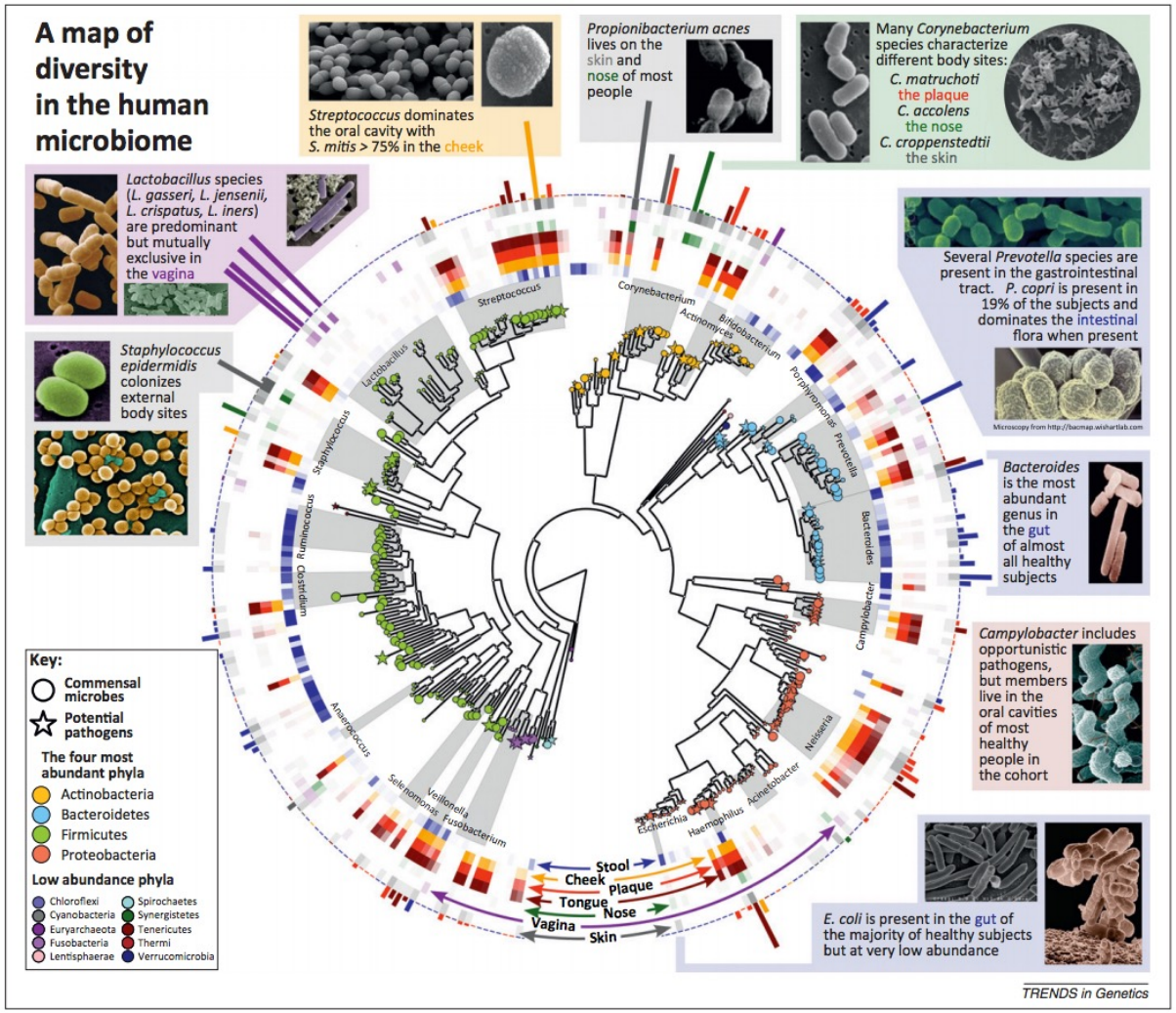


Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications.
Sørlie et al (2001) *PNAS*. 98(19):10869-74.

oncotype DX®
Breast Recurrence Score

mammaprint
decoding breast cancer.

prosigna™ Breast cancer gene signature assay

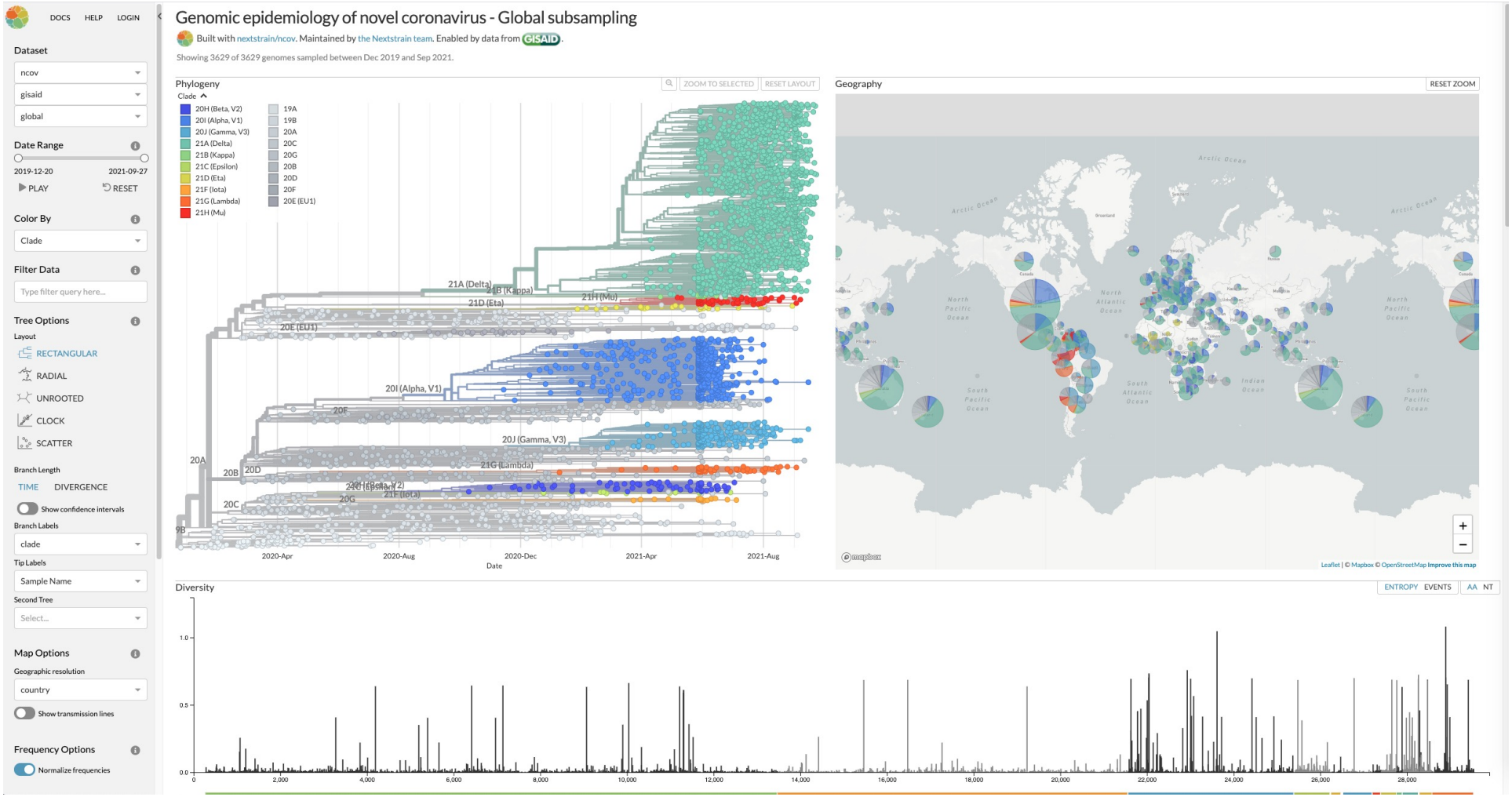


Biodiversity and functional genomics in the human microbiome
Morgan et al (2013) Trends in Genetics. http://doi.org/10.1016/j.tig.2012.09.005

NirvanaBiome

VIOME

BiomeFx™
Functional Microbiome Analysis



Genomic epidemiology of novel coronavirus - Global subsampling

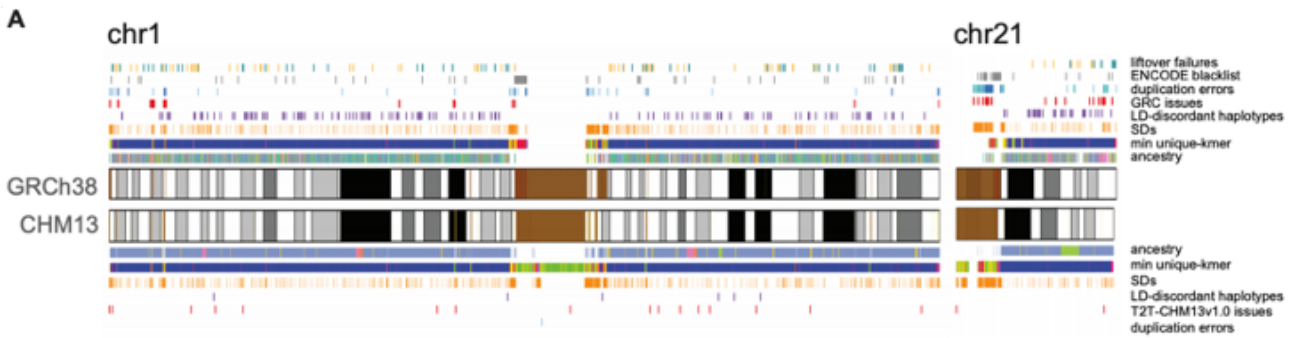
https://nextstrain.org/ncov/gisaid/global

What are the challenges?

The rise of a digital immune system
Schatz & Phillippy (2012) GigaScience 1:4 doi: 10.1186/2047-217X-1-4

Challenge 1: Fast, accurate, and ubiquitous sequencing

The complete sequence of a human genome

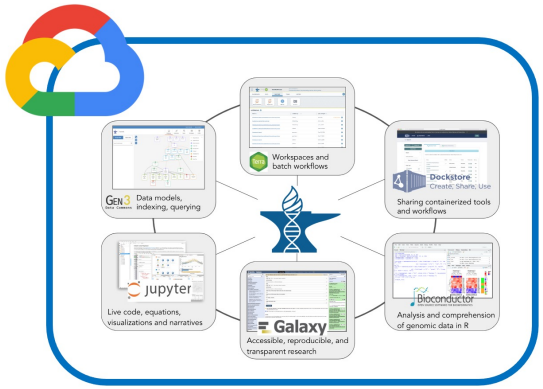


- **T2T-CHM13 is 3.057 Gbp with zero Ns**
  - Every chromosome is telomere-to-telomere, quality estimated >Q70
  - Universally improves variant calling across globally diverse cohorts

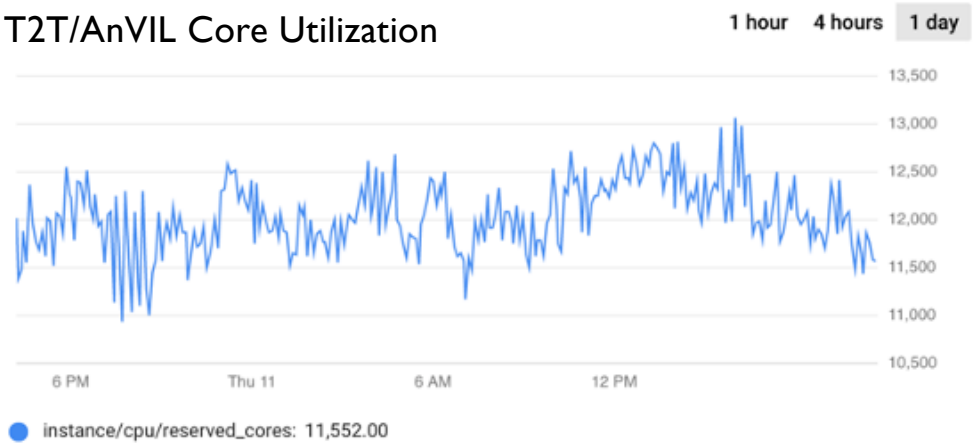
- **Need to develop and integrate other data modalities**
  - Health care data, imaging data, exposures, environmental, etc
  - Wearables as a continuous physicians test

The complete sequence of a human genome
Nurk et al (2021) bioRxiv doi: https://doi.org/10.1101/2021.05.26.445798

Challenge 2: World-wide analysis platform



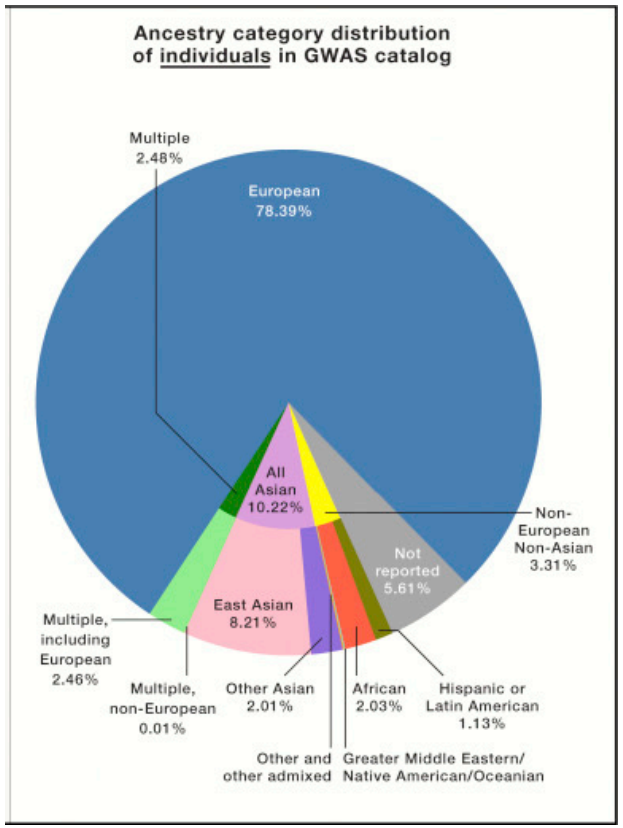
Cloud Computing
NHGRI AnVIL



- **Cloud computing is globally accessible and highly elastic**
  - Equal access for all researchers without huge IT resources; Scale up and scale down as needed
- **But there will never be a single cloud/system for all of genomics**
  - Legal, privacy considerations restrict data to certain facilities, e.g. European GDPR have strict guidelines
  - Different clouds offer different capabilities, e.g. infectious disease has different analysis needs than genetic medicine
  - Need: Continued research on interoperability technologies to enable distributed systems to work together

Inverting the model of genomics data sharing with the NHGRI Genomic Data Science Analysis, Visualization, and Informatics Lab-space (AnVIL)
Schatz, MC, Philippakis, AA, *et al.* (2021) bioRxiv doi: https://doi.org/10.1101/2021.04.22.436044

Challenge 3: Genomics equality for all



- **AI/ML and statistics are incredibly important in genomics**
  - Approaches are only as robust and fair as the training data
  - Current datasets are heavily skewed towards Europeans
  - Life-or-death genomics decisions may be based on biased results and not generalize

- **Need multiple interventions to improve equality in genomics**
  - ***Data:*** Additional sequencing on underrepresented groups
  - ***Software:*** Improved training and validation to detect biases
  - ***Experimental Design:*** Continued research on developing mechanistic understanding of an association
  - ***Training and Education:*** Provide resources and support to those with the greatest needs – trust & consent is essential

The Missing Diversity in Human Genetic Studies
Sirugo, Williams, Tishkoff (2019) *Cell*. https://doi.org/10.1016/j.cell.2019.02.048

Realizing our genomics future

- **We are on the cusp of a biotechnology revolution: tricorders for all**
  - We all carry professional digital cameras in our pocket with quality unimaginable 20 years ago
  - Endless applications to learn more about ourselves, improve our health, and study the world around us

- **Genomics is hard**
  - Interpreting genomics results requires very deep knowledge that is currently out of reach for many
  - We must immediately take action to ensure genomics equality for all
  - In 1000+ years, we will still be learning new meaning in the genome

Acknowledgements

Schatz Lab

Mike Alonge
Sergey Aganezov
Arun Das
Katie Jenike
Melanie Kirsche
Sam Kovaka
Bohan Ni
Gautam Prabhu
Srividya Ramakrishnan
T. Rhyker Ranallo-Benavide
Rachel Sherman
Margaret Starostik
Samantha Zarate
Your Name Here

JHU

Battle Lab
Langmead Lab
Leek Lab
McCoy Lab
Salzberg Lab
Taylor Lab
Timp Lab

Cold Spring Harbor Laboratory

McCombie Lab
Lippman Lab

Baylor College of Medicine

Sedlazeck Lab

Dana Farber Cancer Institute

Van Allen Lab

Telomere-to-Telomere Consortium

Adam Phillippy & Karen Miga et al.

National Human Genome Research Institute

NIH NATIONAL CANCER INSTITUTE

The Mark Foundation® for Cancer Research

NSF

Bloomberg Professors

Thank you!

@mike_schatz

http://schatz-lab.org