

Frontiers of Metrology in Biology

26th CGPM 2018

Marc Salit
Joint Initiative for Metrology in Biology
NIST, Stanford University and SLAC

Frontiers of Metrology in Biology



**The Joint Initiative for
Metrology in Biology**

26th CGPM 2018

NIST

SLAC



**Stanford
University**

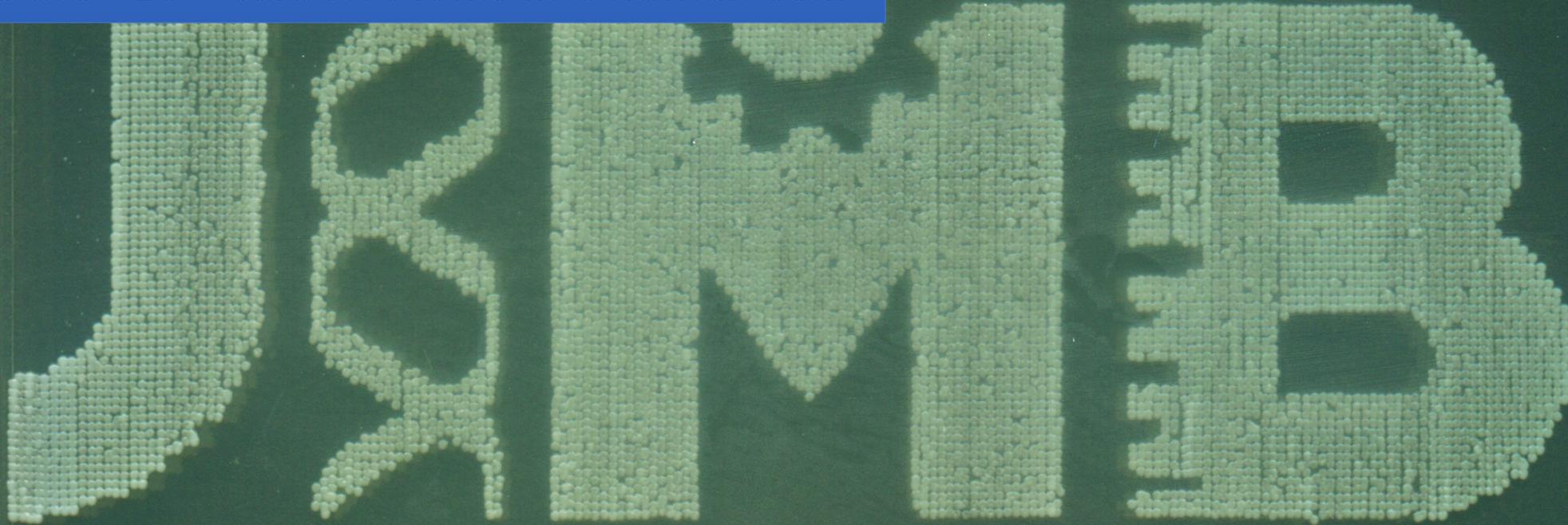
Marc Salit
Joint Initiative for Metrology in Biology
NIST, Stanford University and SLAC

DRUMS

There are 10^{13} cells in a human (human cells).

There are 10^{14} microbial cells in a human.

There are $\sim 10^{10}$ carbon atoms in a human cell.



This is a 192 x 128 grid at 18 droplets per cm.

It contains 7212 droplet transfers of 2.5 nl/droplet.

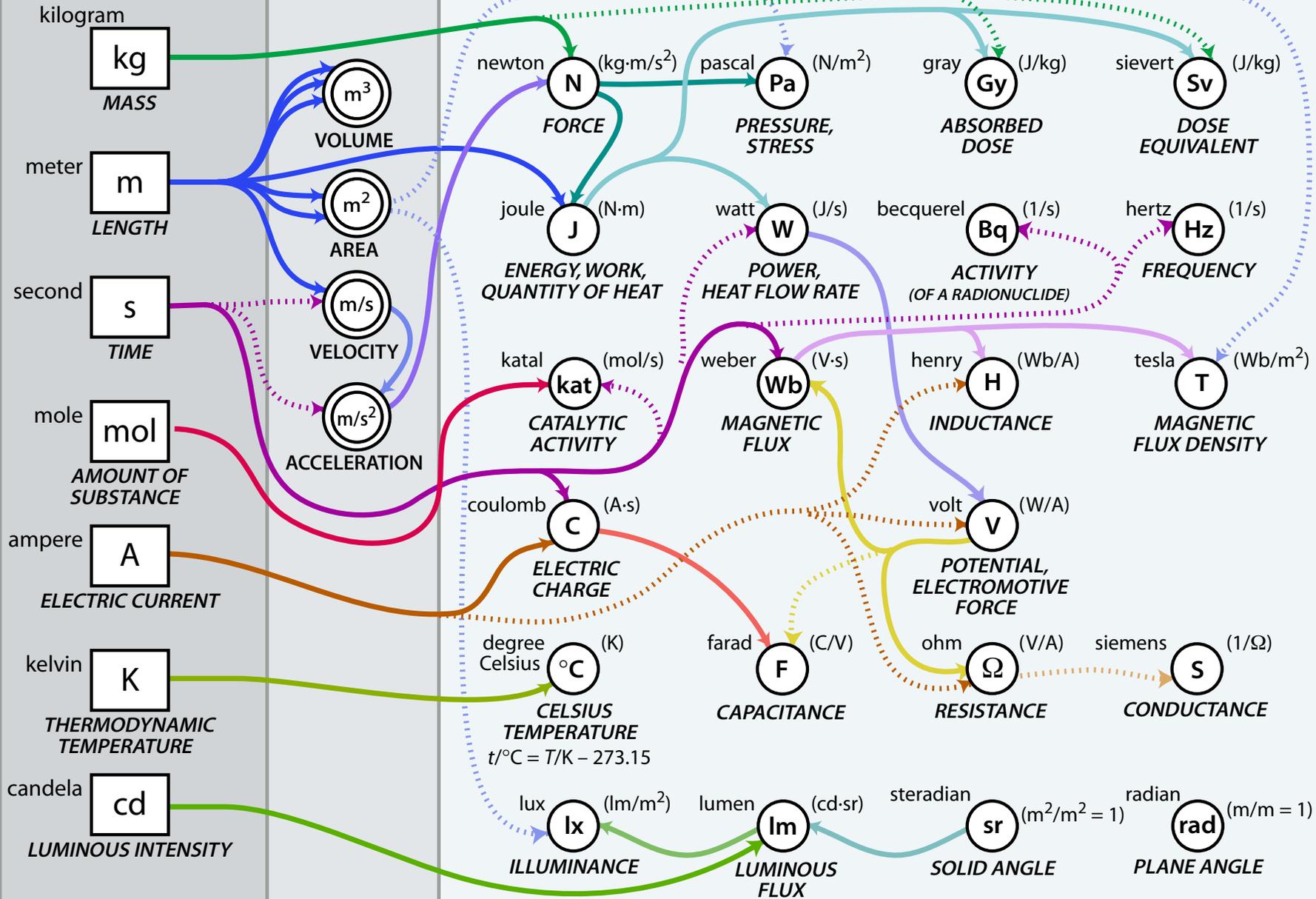
Each droplet contained about 1000 cells. Cells were grown for ~ 24 h.

SI BASE UNITS

Derived units without special names

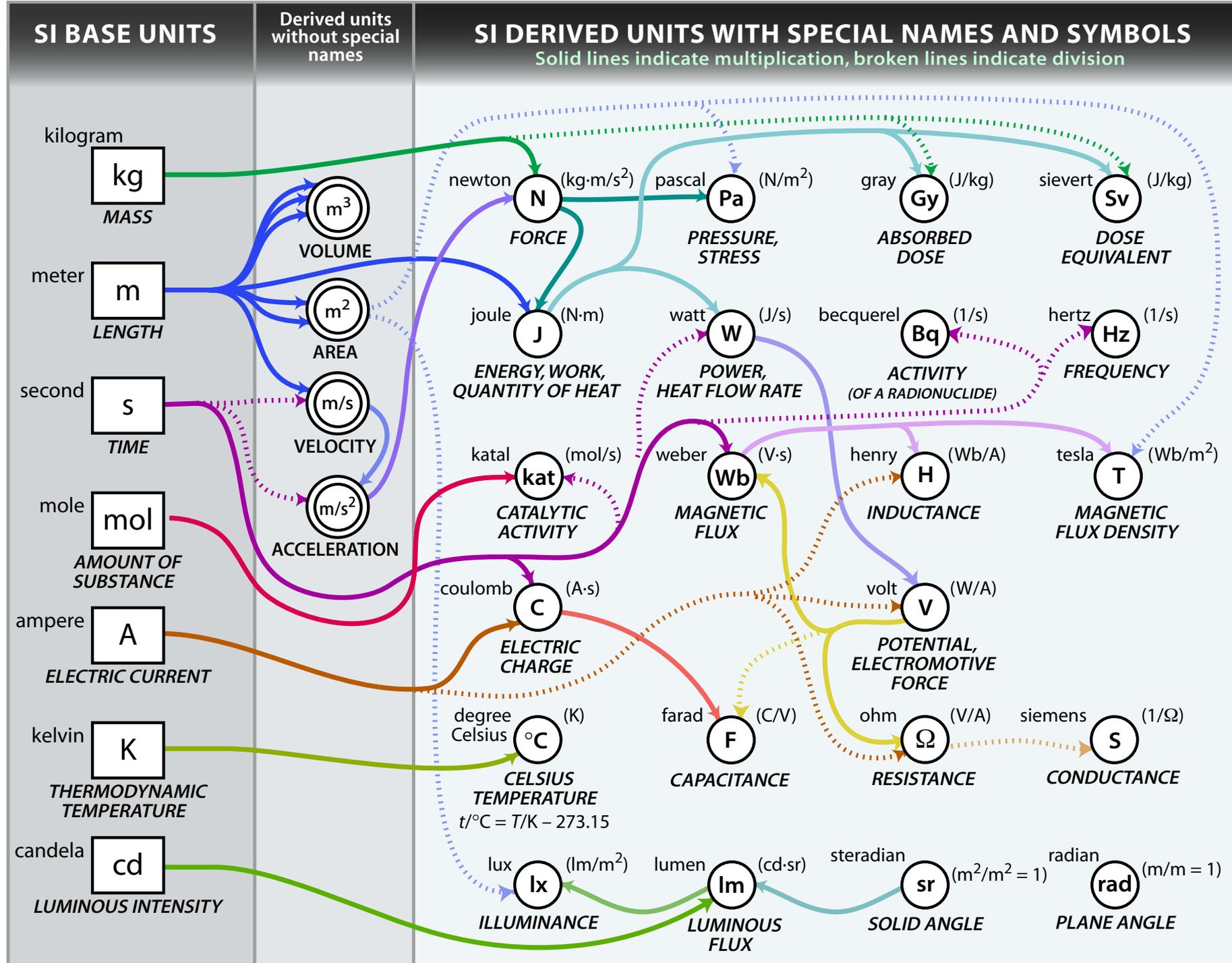
SI DERIVED UNITS WITH SPECIAL NAMES AND SYMBOLS

Solid lines indicate multiplication, broken lines indicate division



The Subway (in Paris, *The Metro*) Diagram

*“Wait... there’s
no stop for
Biology?!?”*



JIMB focusing on *Operational Mastery of Living Matter*

JIMB is focused on *Operational Mastery of living matter* at the cellular level.

- Organizing principle: **“Measure, Model, Make”**
- Through Genomics and Synthetic Biology
- measure everything inside the cell...

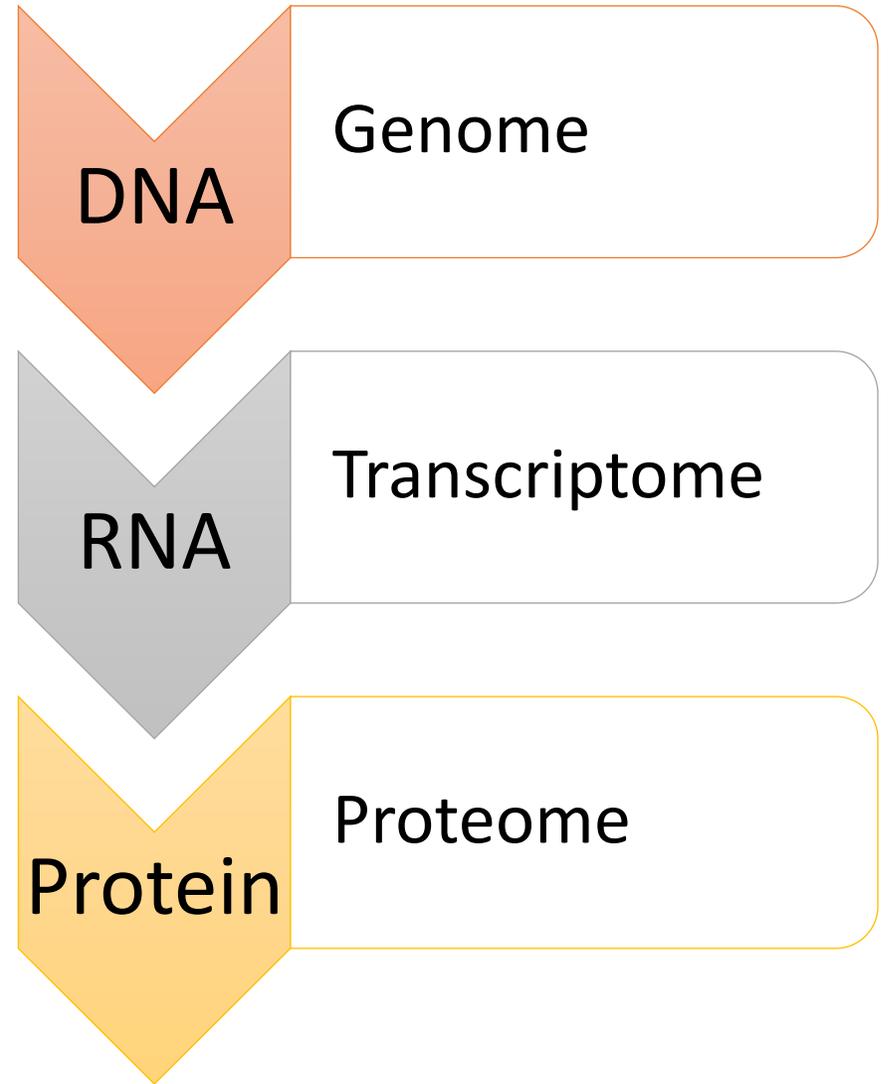
Not focusing on metrology of biomaterials properties, medical diagnostics, biotherapeutics, regenerative medicine, diagnostic imaging...



What's different about metrology in biology?

Characterizing living matter requires measuring massively multiplexed measurands of heterogeneous systems with complex dynamics and interactions.

- *A living cell is a dance of interacting chemical systems governed by biophysics.*
- The cell is the atom of biology.



There was a revolution in measuring biology in 2006.

Accurate Multiplex Polony Sequencing of an Evolved Bacterial Genome

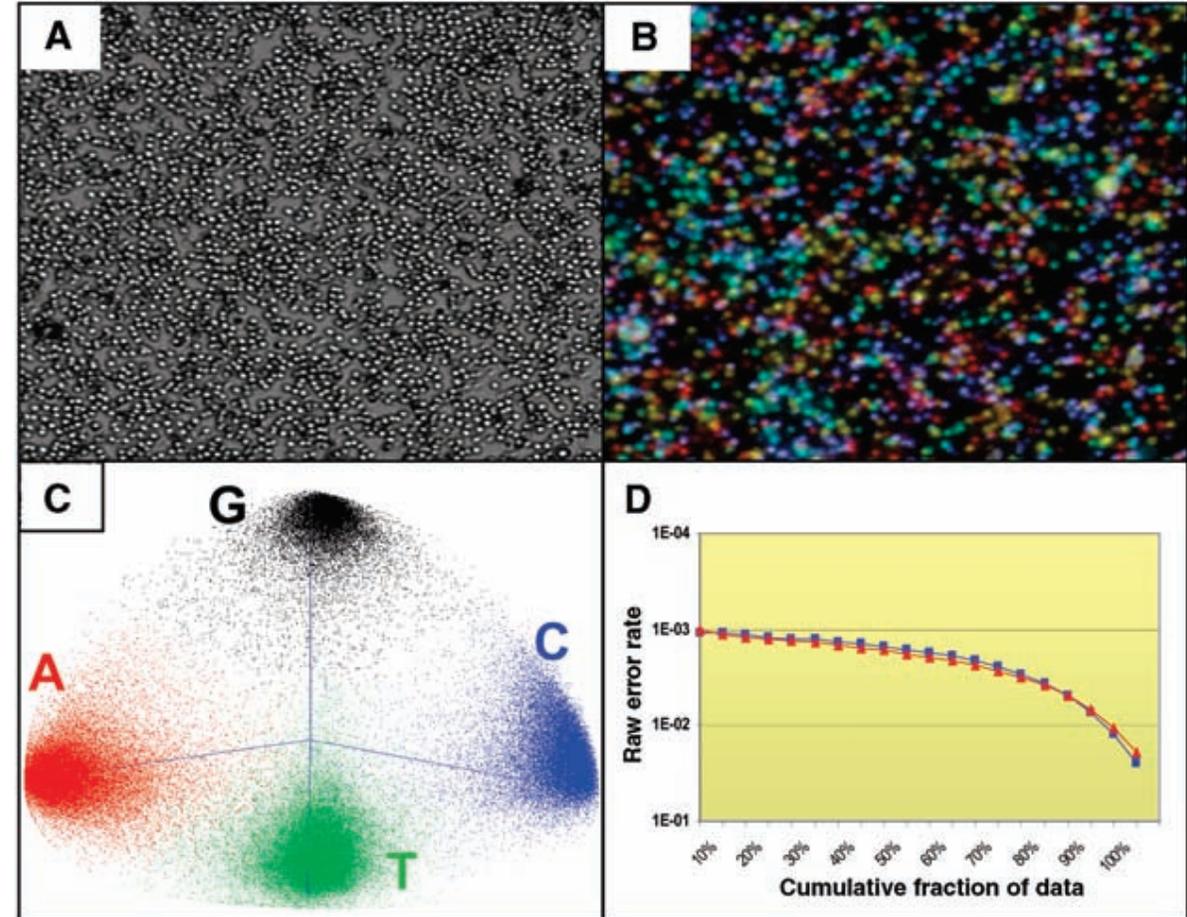
Jay Shendure,^{1*}† Gregory J. Porreca,^{1*}† Nikos B. Reppas,¹
Xiaoxia Lin,¹ John P. McCutcheon,^{2,3} Abraham M. Rosenbaum,¹
Michael D. Wang,¹ Kun Zhang,¹ Robi D. Mitra,² George M. Church¹

We describe a DNA sequencing technology in which a commonly available, inexpensive epifluorescence microscope is converted to rapid nonelectrophoretic DNA sequencing automation. We apply this technology to resequence an evolved strain of *Escherichia coli* at less than one error per million consensus bases. A cell-free, mate-paired library provided single DNA molecules that were amplified in parallel to 1-micrometer beads by emulsion polymerase chain reaction. Millions of beads were immobilized in a polyacrylamide gel and subjected to automated cycles of sequencing by ligation and four-color imaging. Cost per base was roughly one-ninth as much as that of conventional sequencing. Our protocols were implemented with off-the-shelf instrumentation and reagents.

The ubiquity and longevity of Sanger sequencing (1) are remarkable. Analogous to semiconductors, measures of cost and production have followed exponential trends (2). High-throughput centers generate data at a speed of 20 raw bases per instrument-second and a cost of \$1.00 per raw kilobase. Nonetheless, optimizations of elec-

trophoretic methods may be reaching their limits. Meeting the challenge of the \$1000 human genome requires a paradigm shift in our underlying approach to the DNA polymer (3).

Cyclic array methods, an attractive class of alternative technologies, are “multiplex” in that they leverage a single reagent volume to enzymatically manipulate thousands to mil-



Shendure, J., Porreca, G. J., Reppas, N. B., Lin, X., McCutcheon, J. P., Rosenbaum, A. M., ... Church, G. M. (2005). Accurate multiplex polony sequencing of an evolved bacterial genome. *Science*, 309(5741), 1728–1732. <https://doi.org/10.1126/science.1117389>

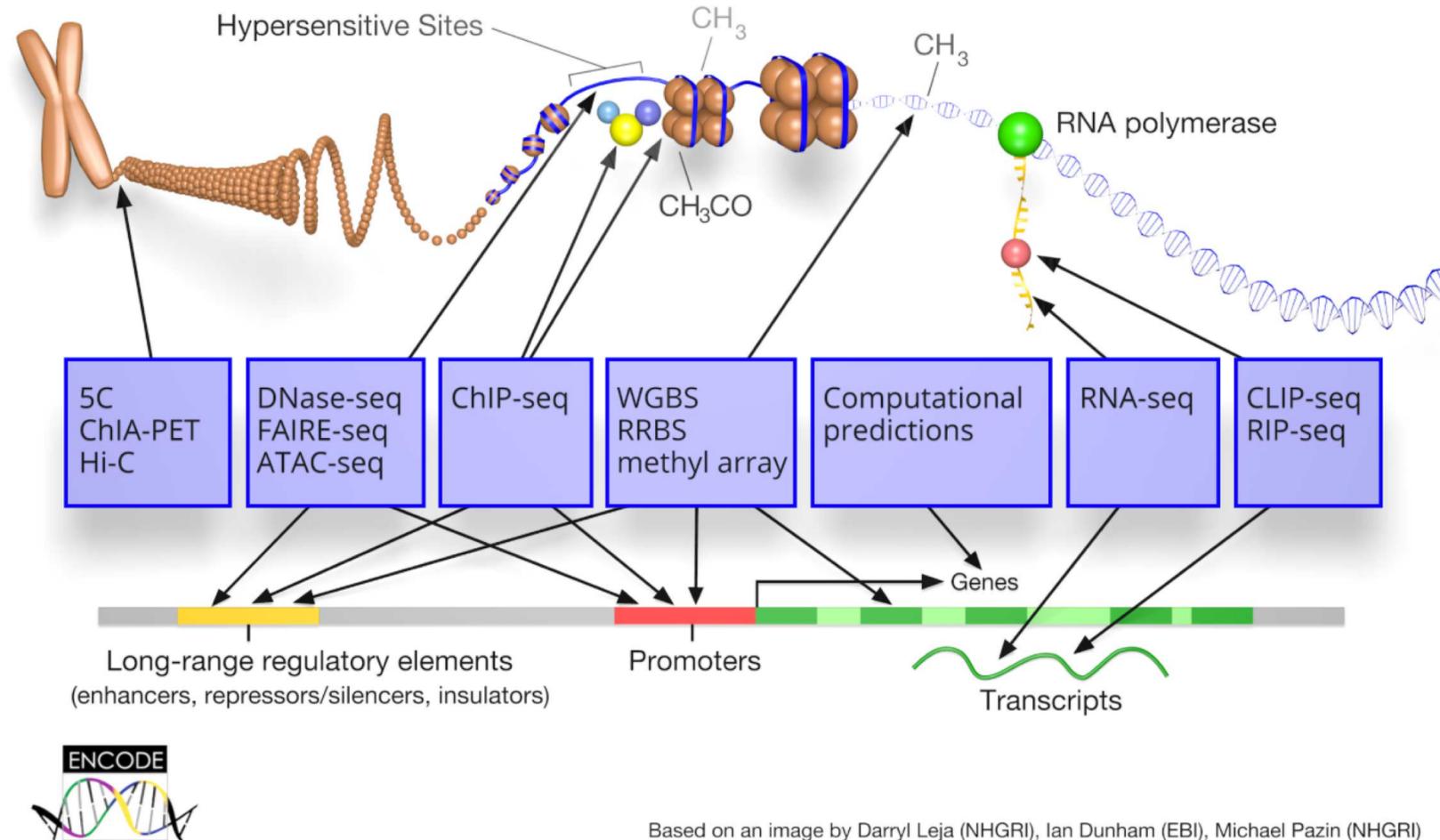
You can scan the landscape to frame a roadmap.

	Genome	Regulation	Transcriptome	Reg.	Proteome	Reg.	Metabolome
Organelle	<i>We're pretty good at sequencing, but sampling presents challenges.</i>	<i>Lots of methods, reasonably characterized</i>	<i>Pretty good at RNA-Seq, reasonably characterized, sampling challenges.</i>	<i>A couple of methods, still emerging</i>	<i>Variety of methods, technically challenging</i>	<i>Enzyme activity measures, not 'omics?</i>	<i>Variety of methods, still emerging</i>
Cell							
Tissue							
Organism							
System							
...							

It's more granular than this – there's work to do to roadmap our measurement capabilities.

Community is reaching out for Standards.

- Protocols
- Data Representation
- Data Exchange
- Requirements /Specifications
- Calibration Materials
- Validation/Benchmark Materials
- Validation/Benchmark Data



from <https://www.encodeproject.org>

based on an image from Darryl Leja (NHGRI), Ian Dunham (EBI), Michael Pazin (NHGRI)

COMMENT

AVIAN INFLUENZA Shift expertise to track mutations where they emerge **p.534**

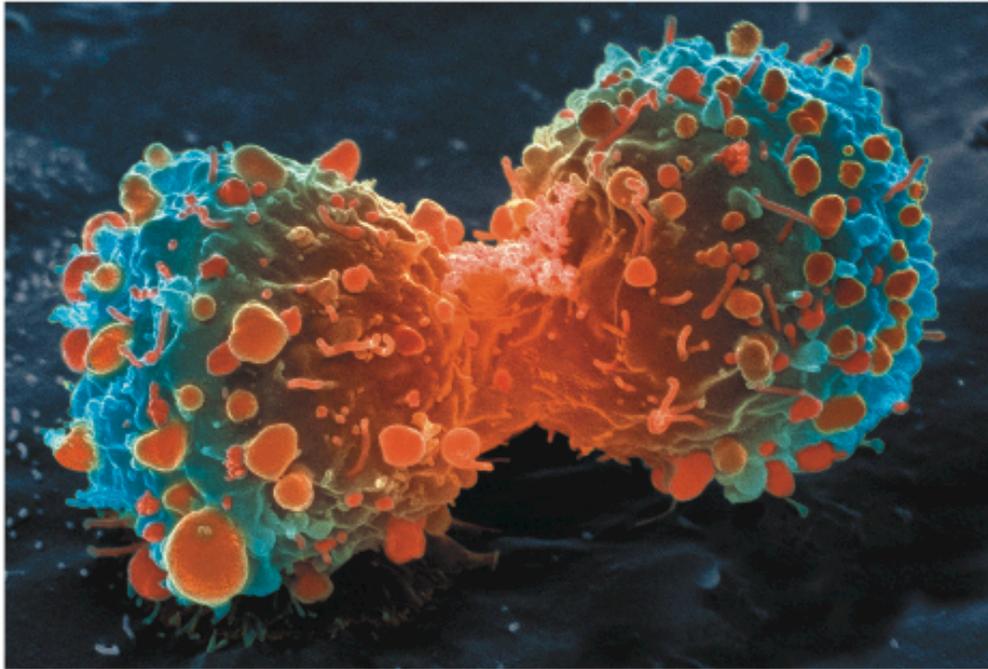
EARTH SYSTEMS Past climates give valuable clues to future warming **p.537**

HISTORY OF SCIENCE Descartes' lost letter tracked using Google **p.540**

OBITUARY Wylie Vale and an elusive stress hormone **p.542**



This 2012 *Nature* Comment triggered recognition of a “*Reproducibility Crisis*” in biomedical science...



Many landmark findings in preclinical oncology research are not reproducible, in part because of inadequate cell lines and animal models.

Raise standards for preclinical cancer research

C. Glenn Begley and Lee M. Ellis propose how methods, publications and incentives must change if patients are to benefit.

Efforts over the past decade to trials in oncology have the highest failure investigators must reassess their approach to

Taking a cue from Chemical Metrology... Reference Materials can work in Biology

- Both RMs depicted were created in consortium partnerships
- Both are widely adopted
- Both address needs in Genome-Scale Measurements

Quantitative

Qualitative

Transcriptome Spike-ins



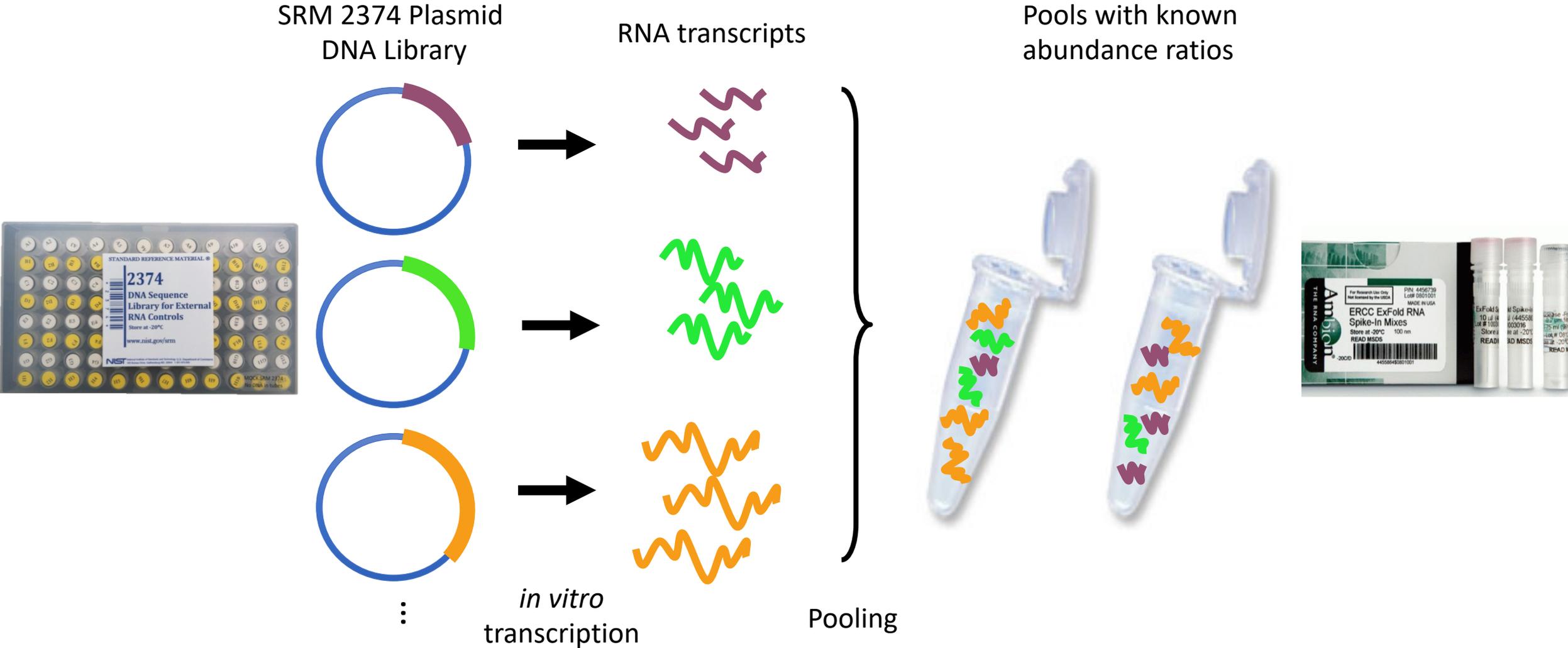
ERCC Controls

Human Genomes

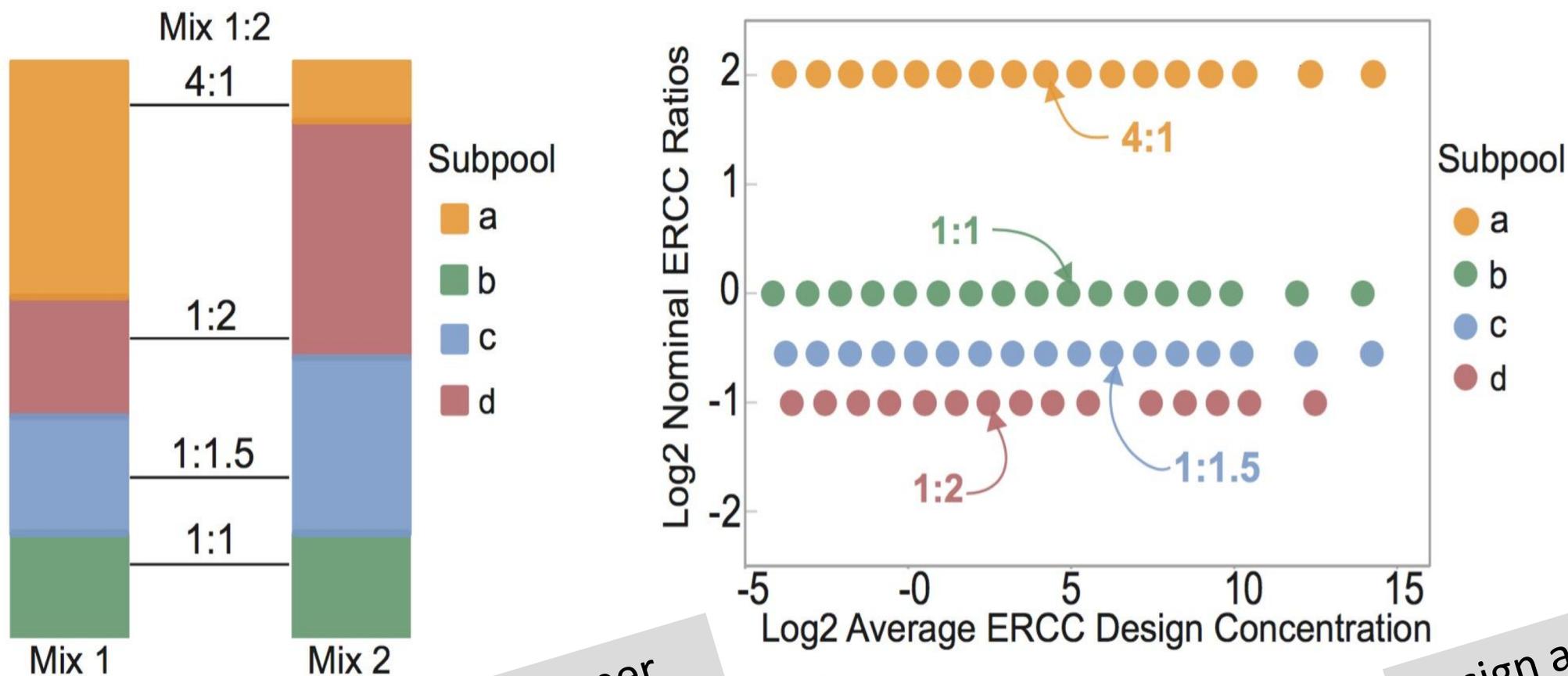


GIAB Genomes

Use the ERCC Reference Material plasmid library to make controls



Design of Ambion ERCC Spike-In Ratio Mixtures

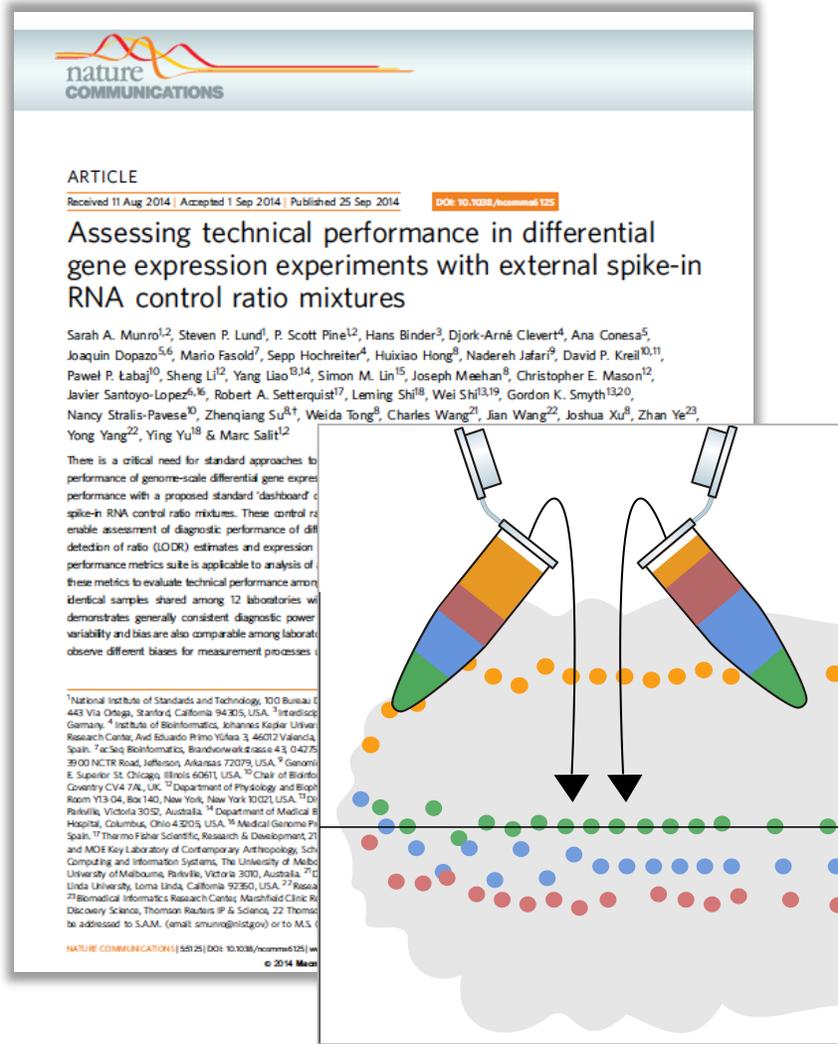


23 Controls per Subpool

Design abundance spans 2^{20} range within each Subpool

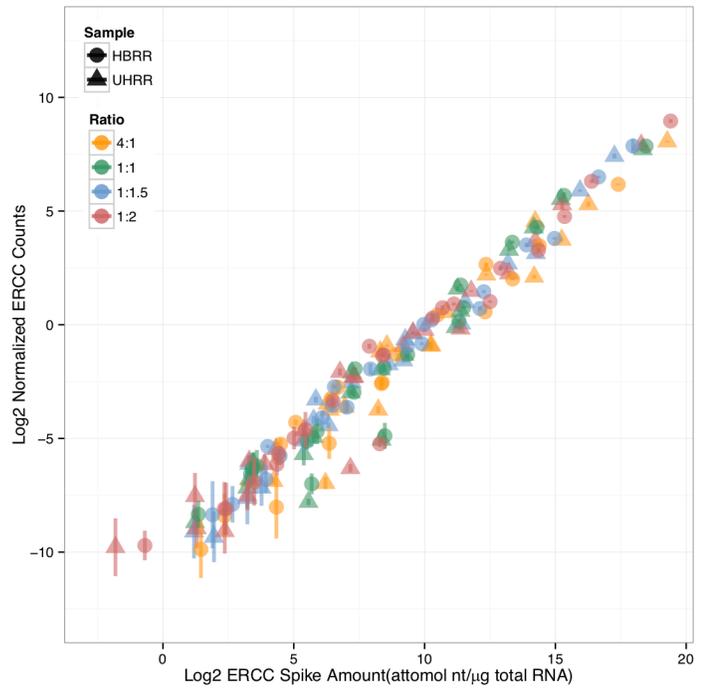
ercddashboard gives standard measures of technical performance

- Technology-independent ratio performance measures
- Shows differences in performance across
 - Experiments
 - Laboratories
 - Measurement processes

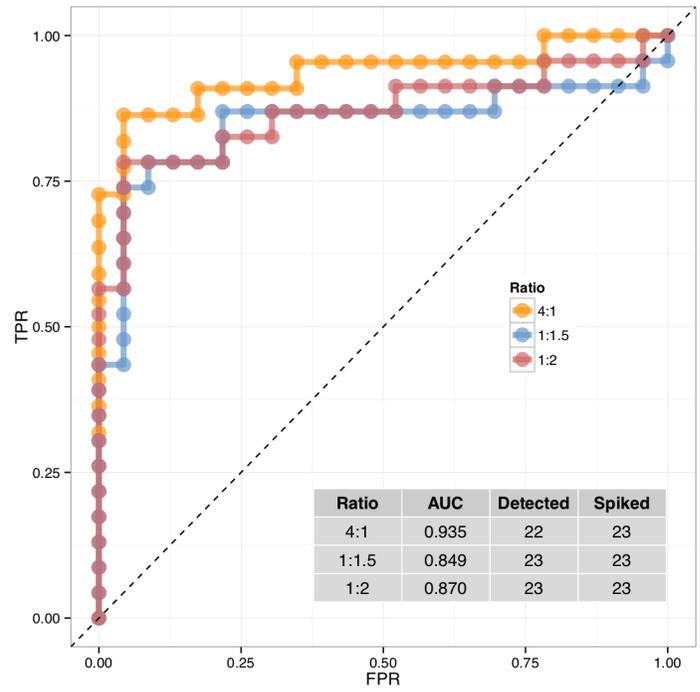


Munro, S. A. et al. Nat. Commun. 5:5125 doi: 10.1038/ncomms6125 (2014).

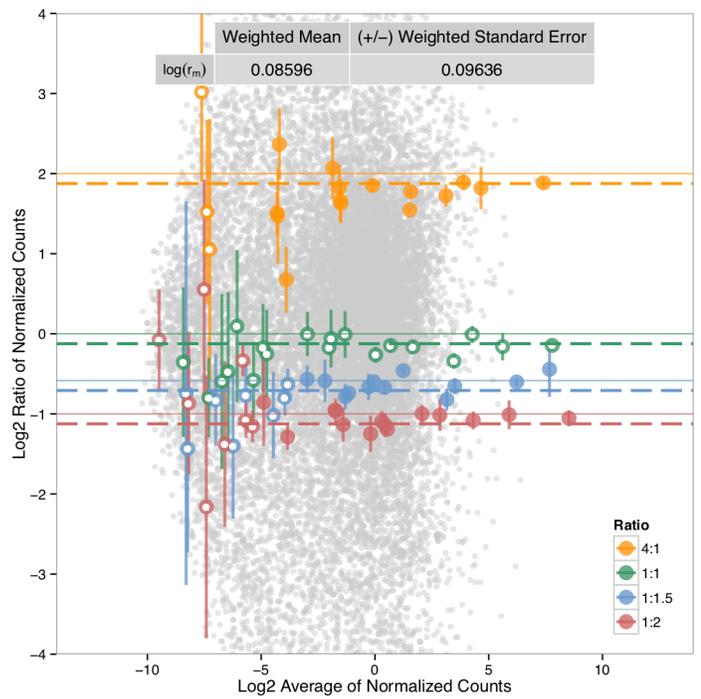
Evaluate Dynamic Range Performance



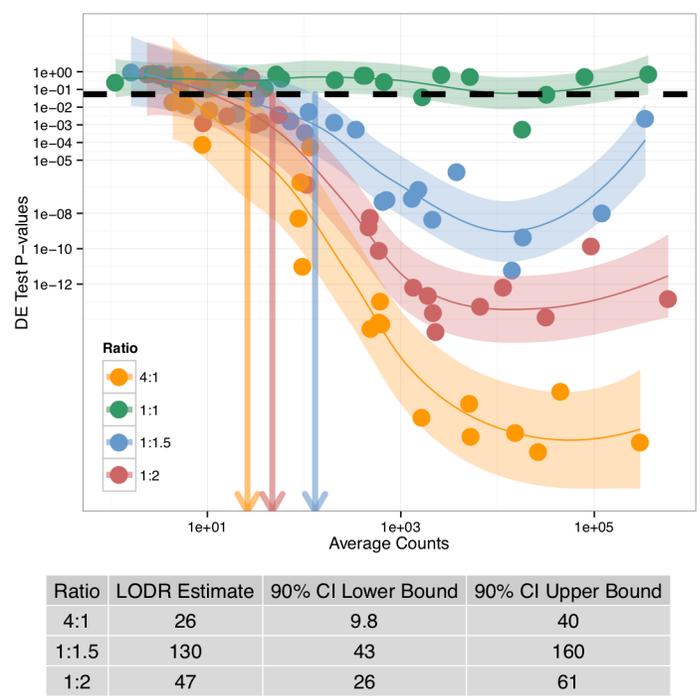
Evaluate Diagnostic Performance – “ROC Curve”



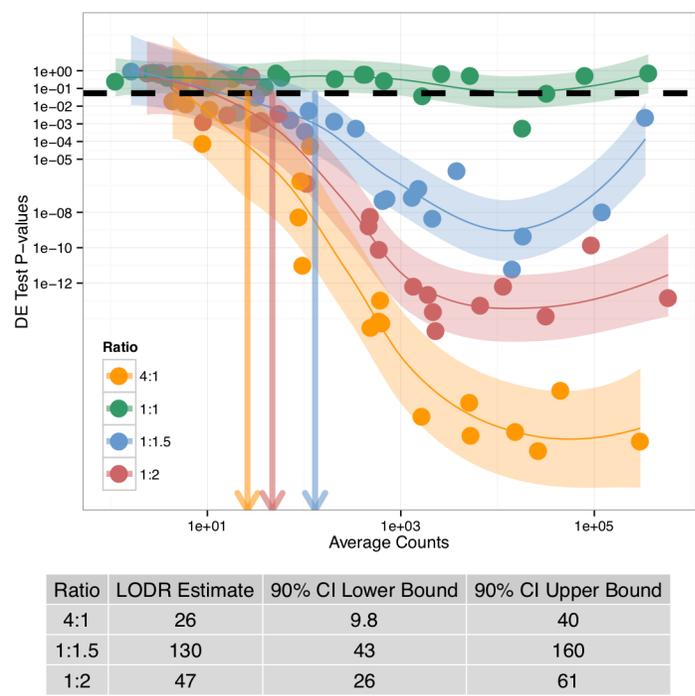
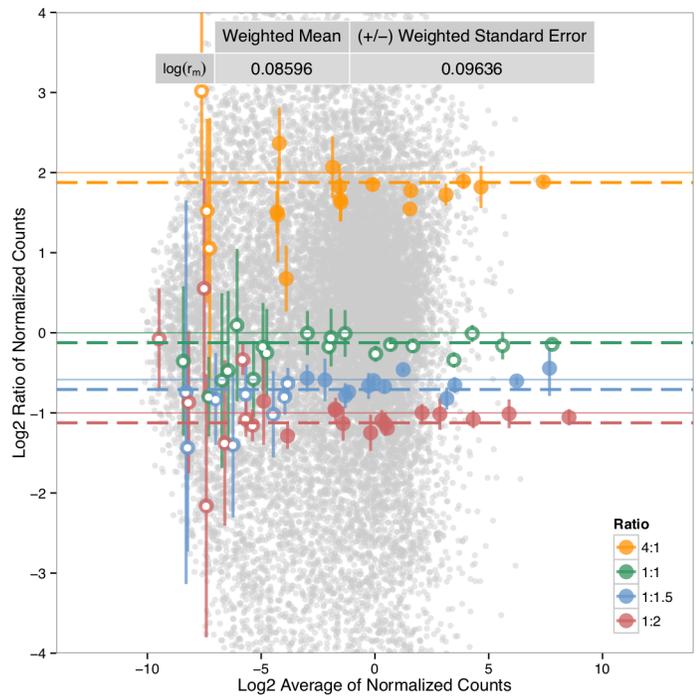
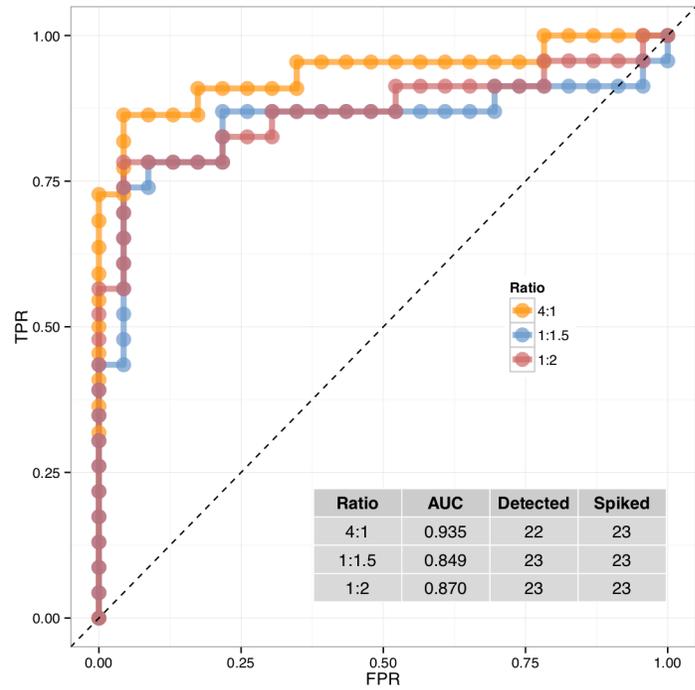
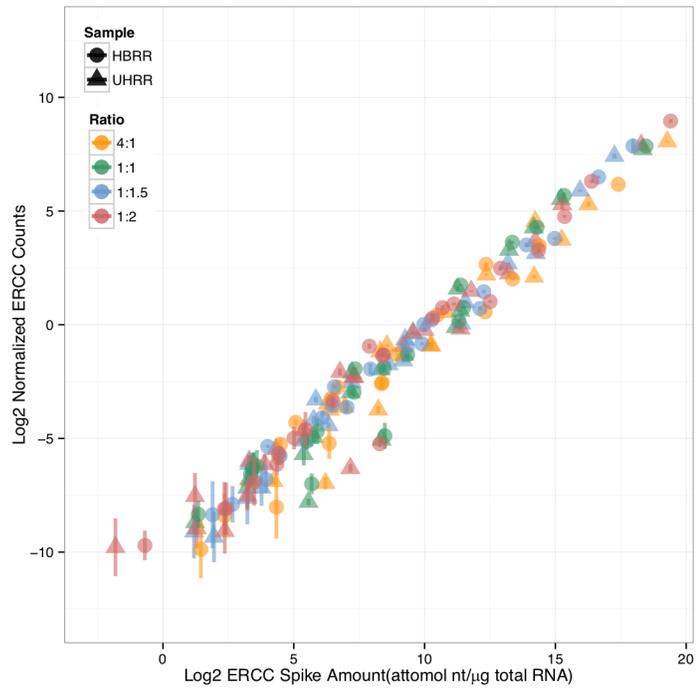
Evaluate Ratio Performance – “MA Plot”



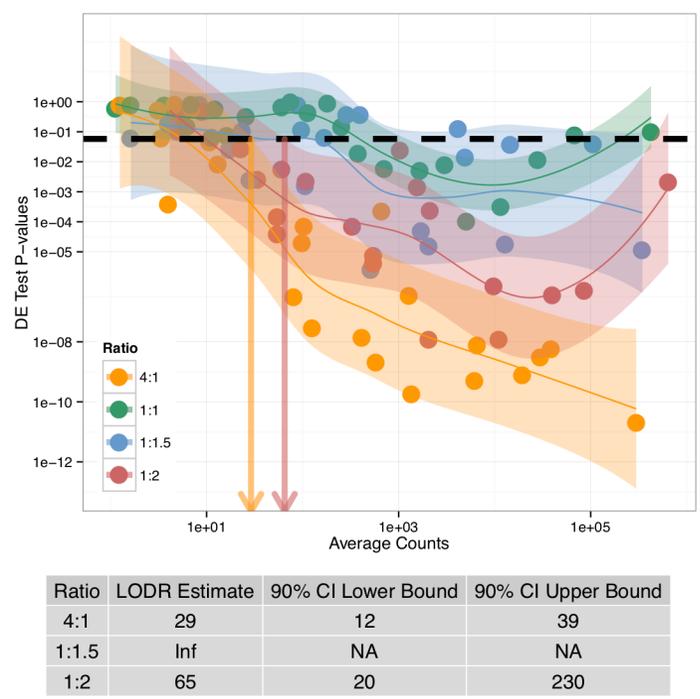
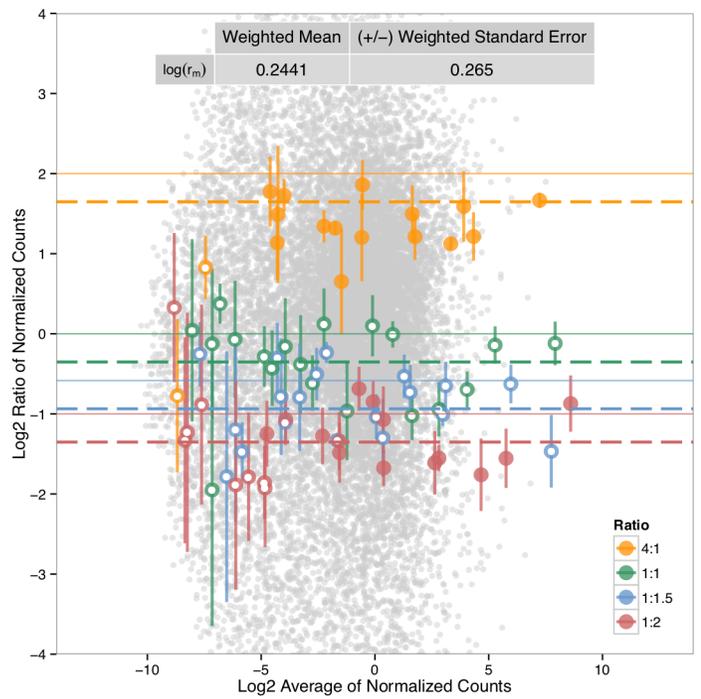
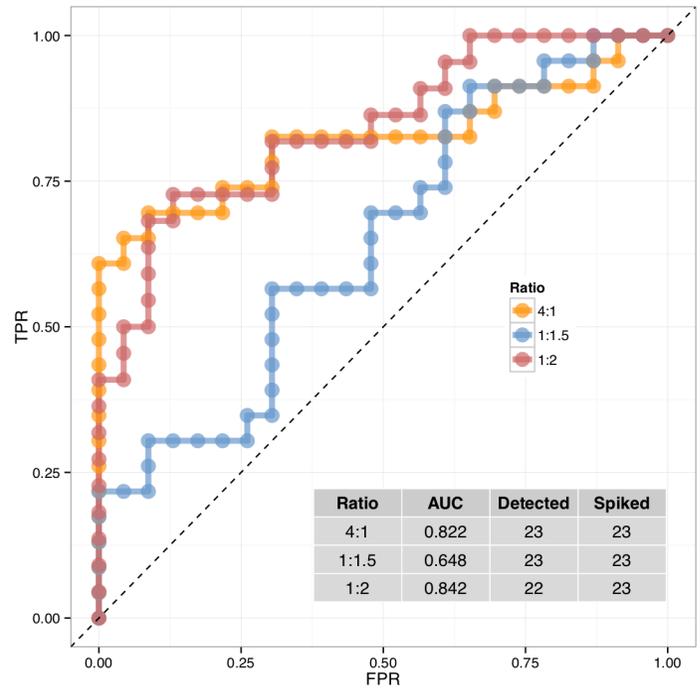
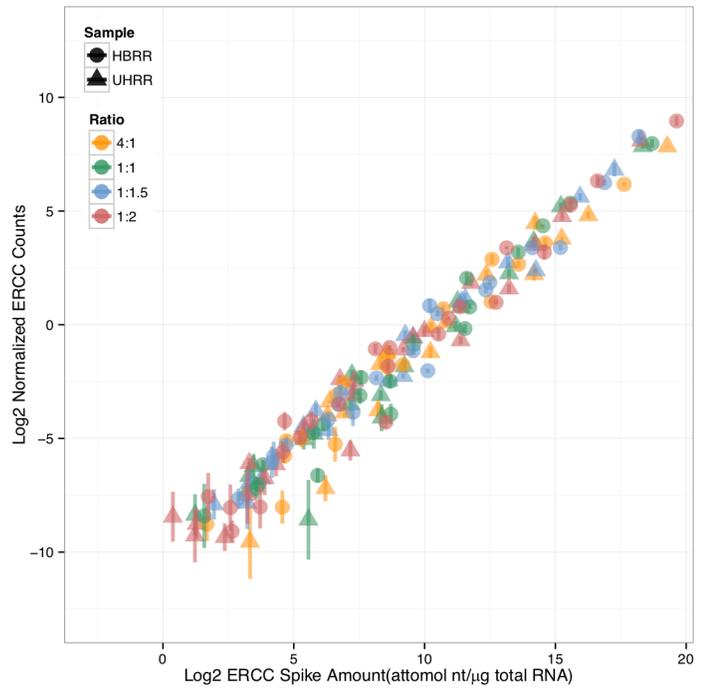
Establish Lower Limit of Detection for Differential Expression Detection – “LODR”



Good Lab

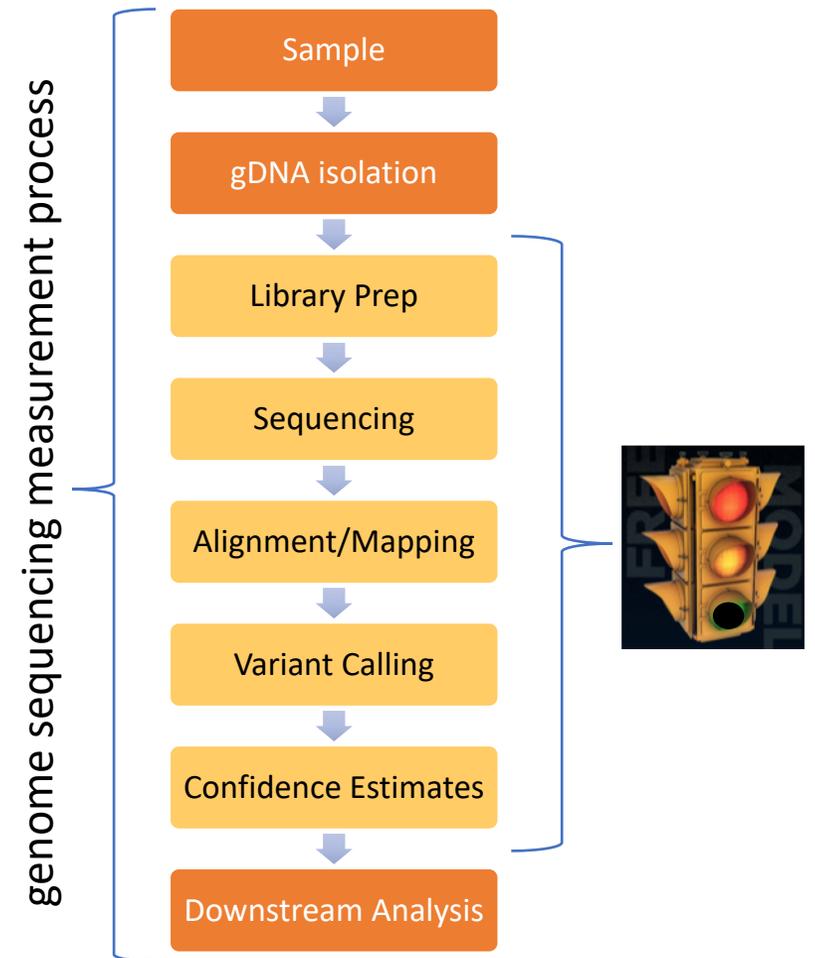


Bad Lab



Genome in a Bottle Consortium is making and disseminating human genome reference materials.

- create shared reference samples
- *validation materials* to evaluate, demonstrate, refine, optimize technologies
 - red light/yellow light...
- developed benchmarking dashboard with stakeholders @ GA4GH
- meeting needs for technology developers, regulators, clinical research teams



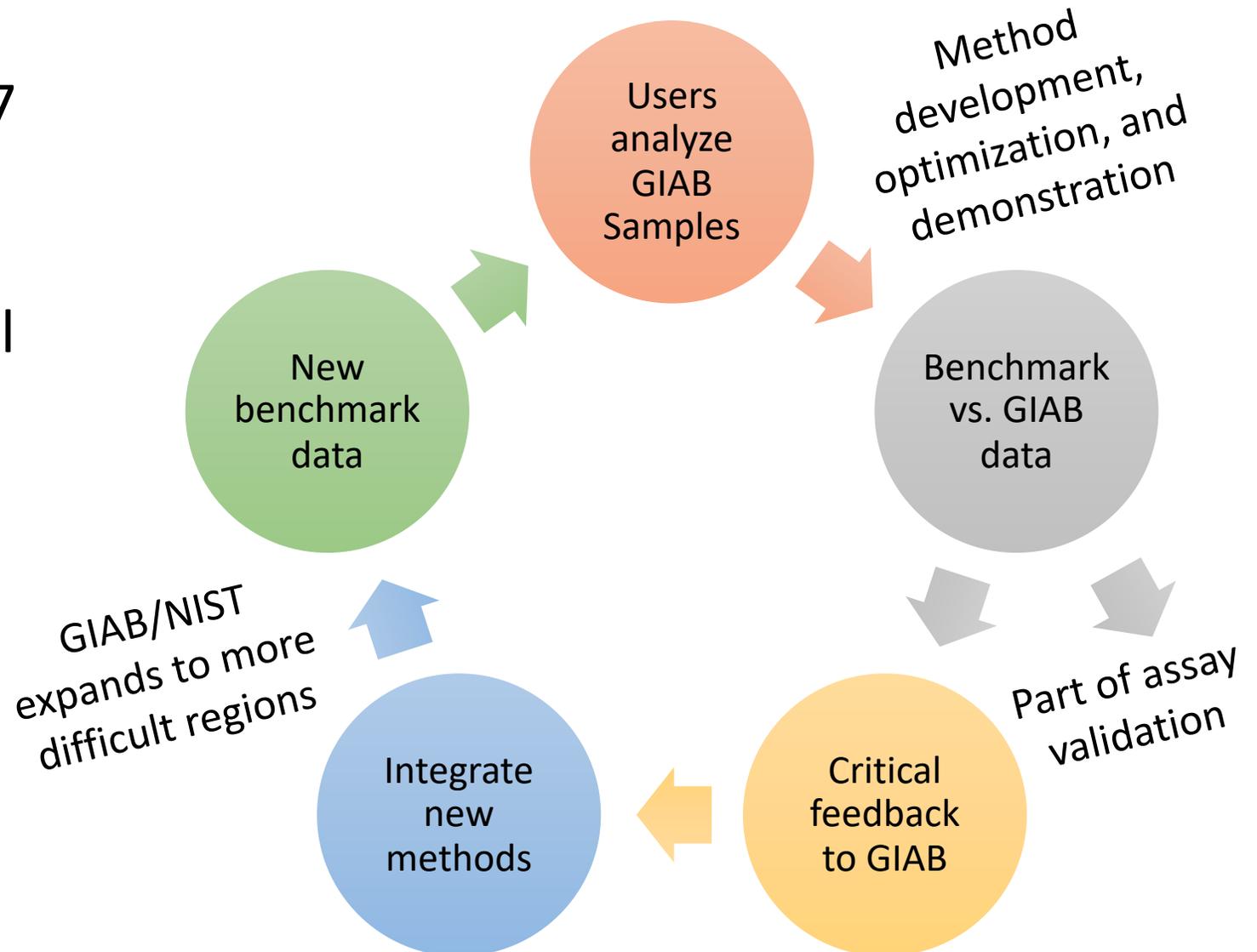
GIAB “Open Science” Virtuous Cycle

Reference data

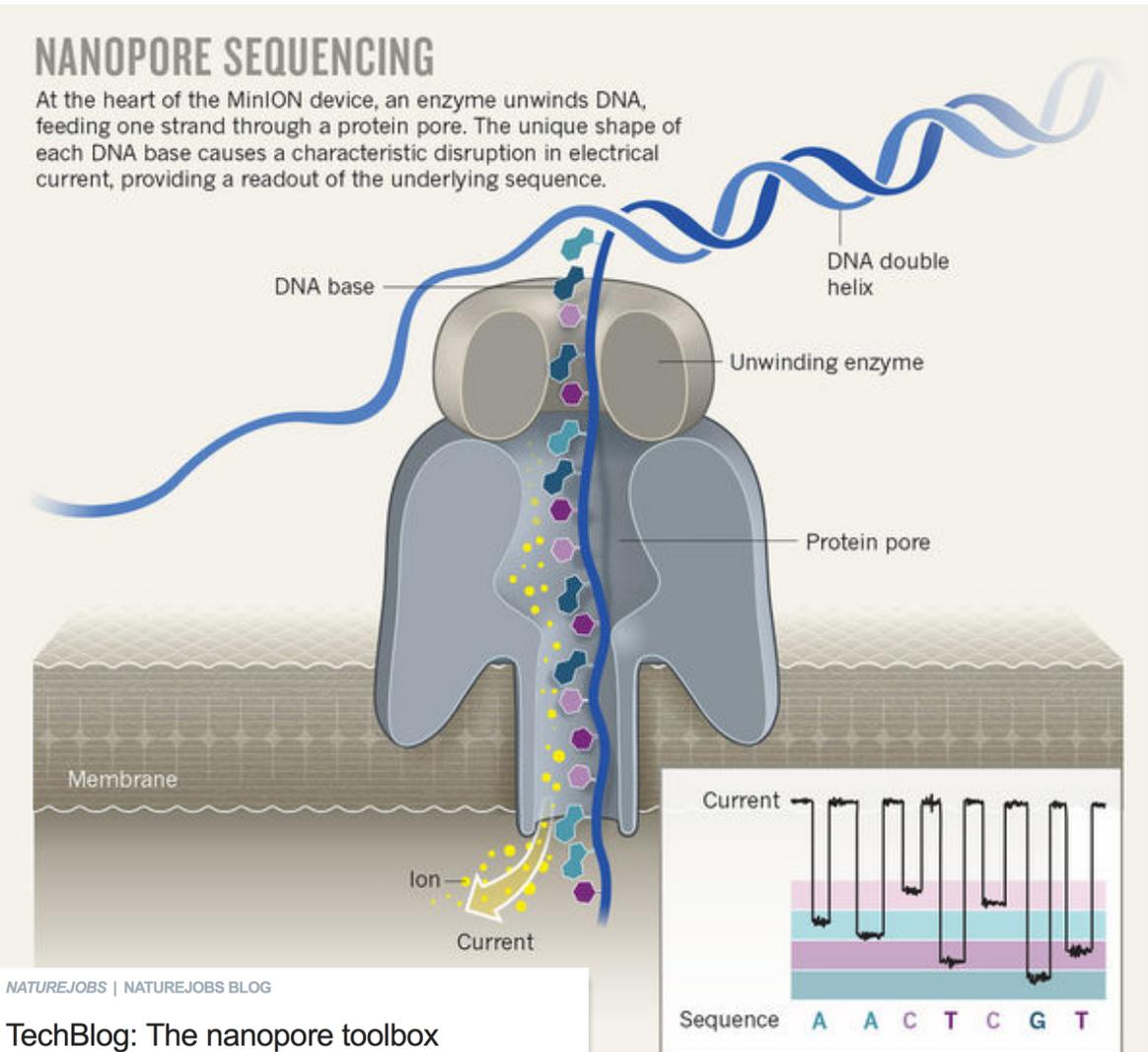
- phased ***variant calls*** across 7 human genomes
 - ~ 4M small variants
 - ~20,000 larger “structural variants”

All data available immediately without embargo

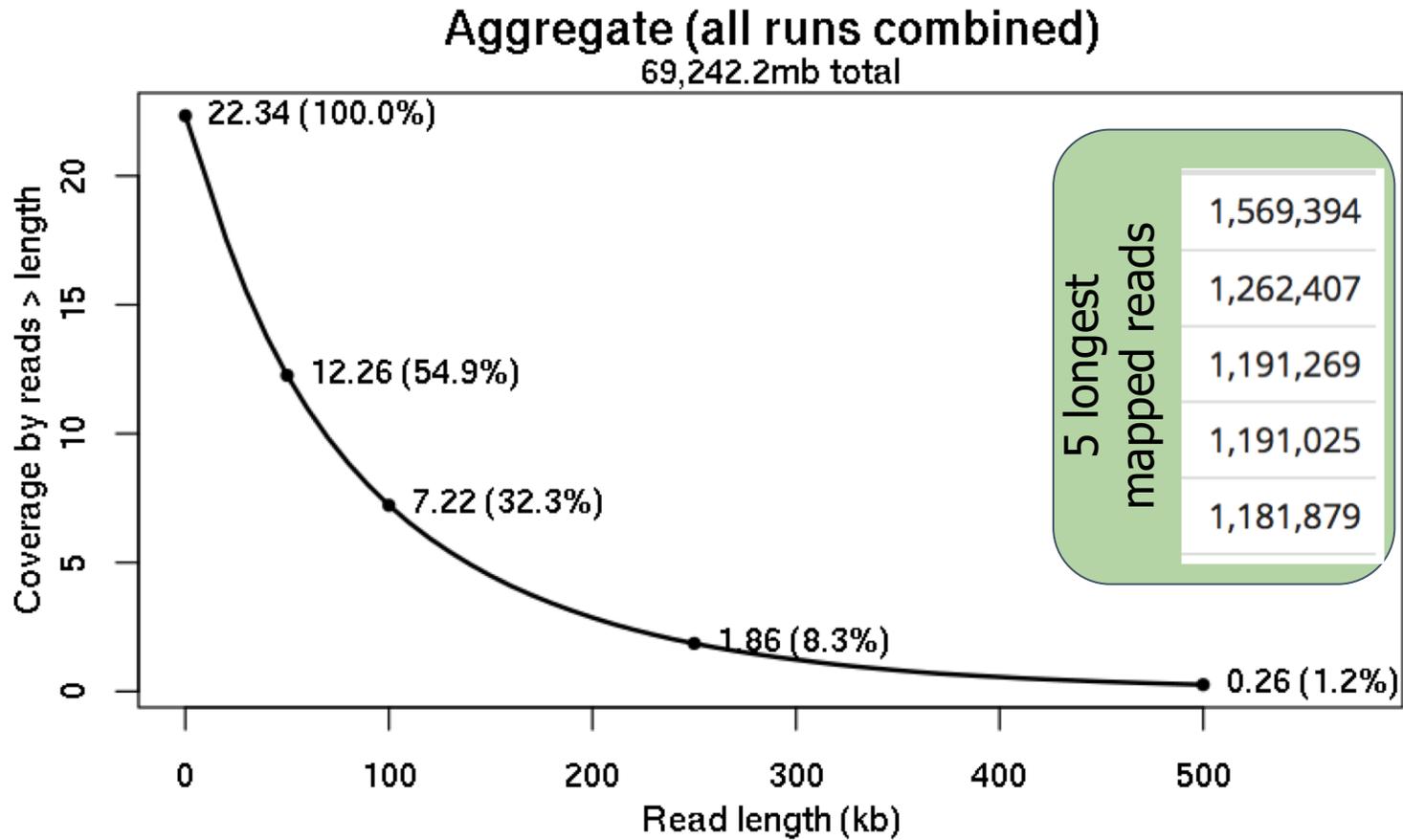
- consistent with transparency and metrology



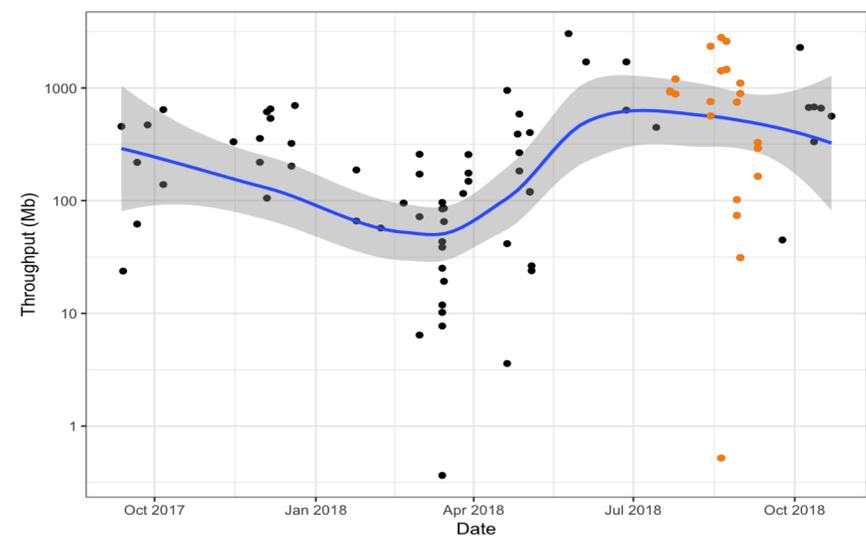
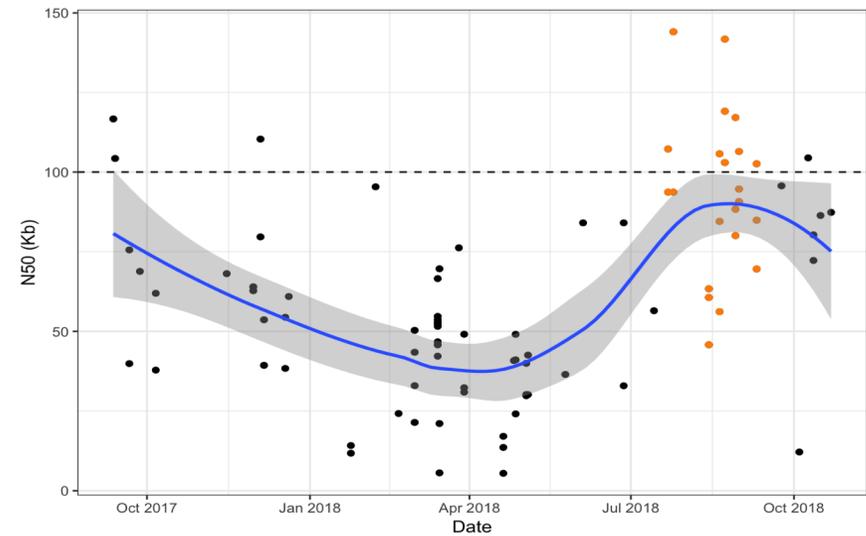
Evolving with Technologies: Single-molecule nanopore sequencer



We're Accumulating Nanopore Coverage



N50 and Coverage over time



Important characteristics of benchmark calls

What does “reference standard” mean?

Accurate

- high-confidence variants, genotypes, haplotypes, and regions
- ***compared to the benchmark, the majority of differences (FPs/FNs) are errors in the method***

Comprehensive characterization

- many examples of different variant types/genome contexts
- eventually, diploid assembly benchmarking

Representative examples

- different types of variants in different genome contexts

Important characteristics of benchmark calls

What does "reference standard" mean?

Accurate

- high-confidence variants, genotypes, haplotypes, regions
- **comparability**
majority
are essential

Comprehensive characterization

- different variant types
- assembly

Representative

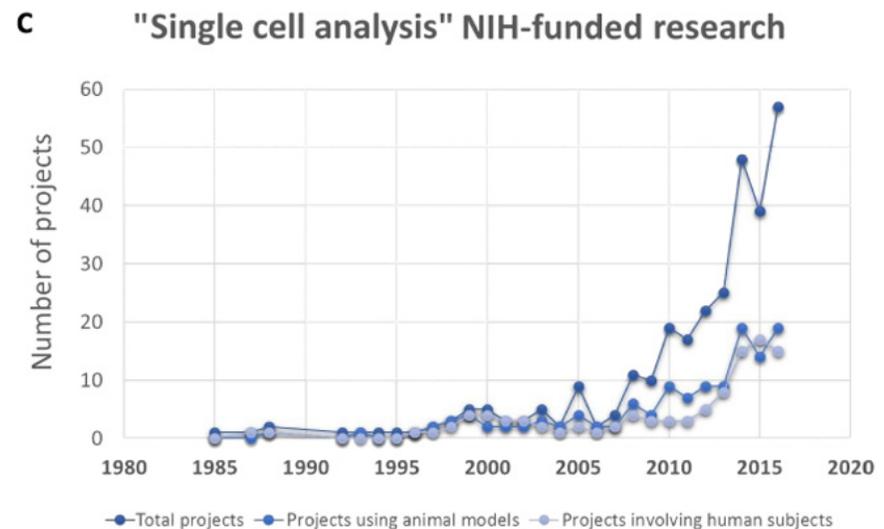
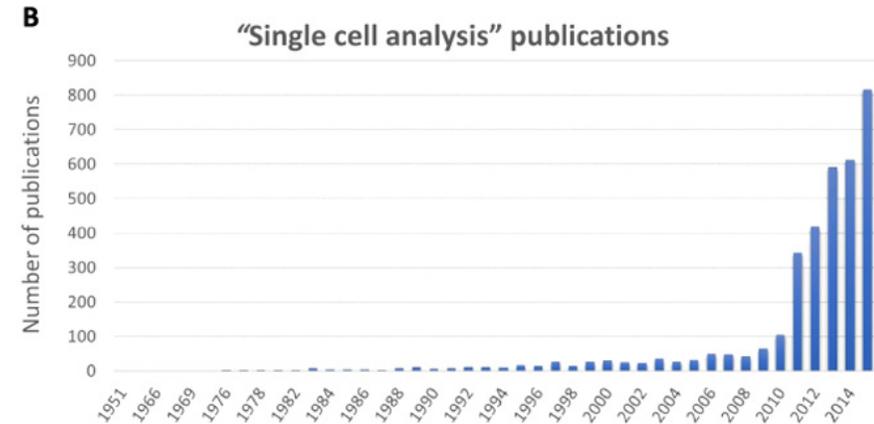
- different
- different

Establishing shared references, analytical approaches, performance metrics, and best practices for nominal properties is essential biometrology

Single-cells – the atoms of biology!

Single-cell genomics is being widely adopted.

- “Quantum” shift from measuring bulk populations of heterogeneous cells
- Innovation in “Wet” lab
 - tissue disaggregation
 - including spatial location
 - single-cell processing
- Innovation in “Dry” lab
 - data management
 - meaningful analysis



Building a human cell atlas with single-cell RNA-Seq

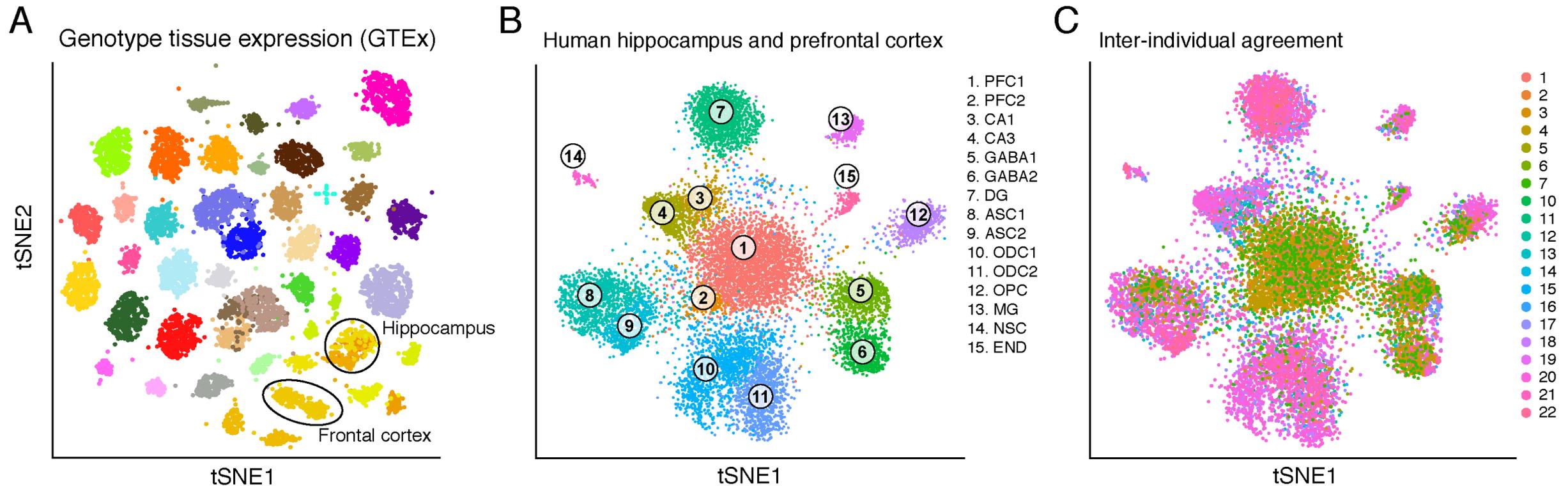
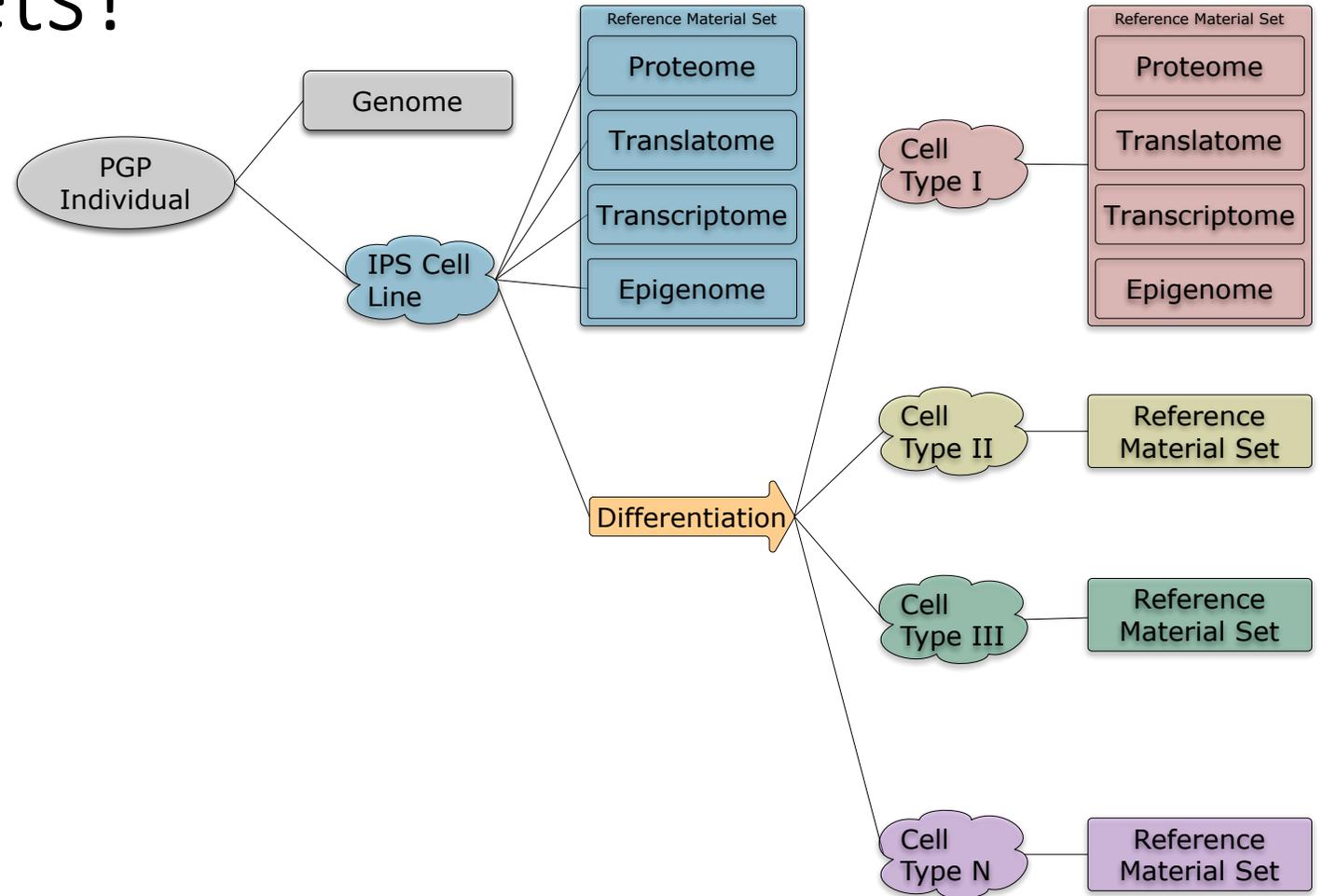


Figure 5. Retrospective samples from GTEx can be successfully profiled using single-nucleus RNA-Seq. (A) Bulk gene expression profiles from all GTEx tissues. Hippocampus and frontal cortex sample clusters, from which samples in (B) are obtained, are circled. (B) Single nucleus RNA-Seq (by DroNc-Seq) of hippocampus and frontal cortex samples from the GTEx collection. tSNE plots are colored by k-NN graph clustering and labeled *post hoc* by cell type. (C) Each cluster is supported by multiple individuals (from relevant tissue).

What could we do with iPS Reference Material Sets?

Develop reference sets from a single individual that represent a “body map” of functional ‘omes

- use for model development and validation
- use as substrate for technology development
- benchmark sets for biology



Predictive Systems Models, Iteratively Refined with Data

Organism Construction Coordination

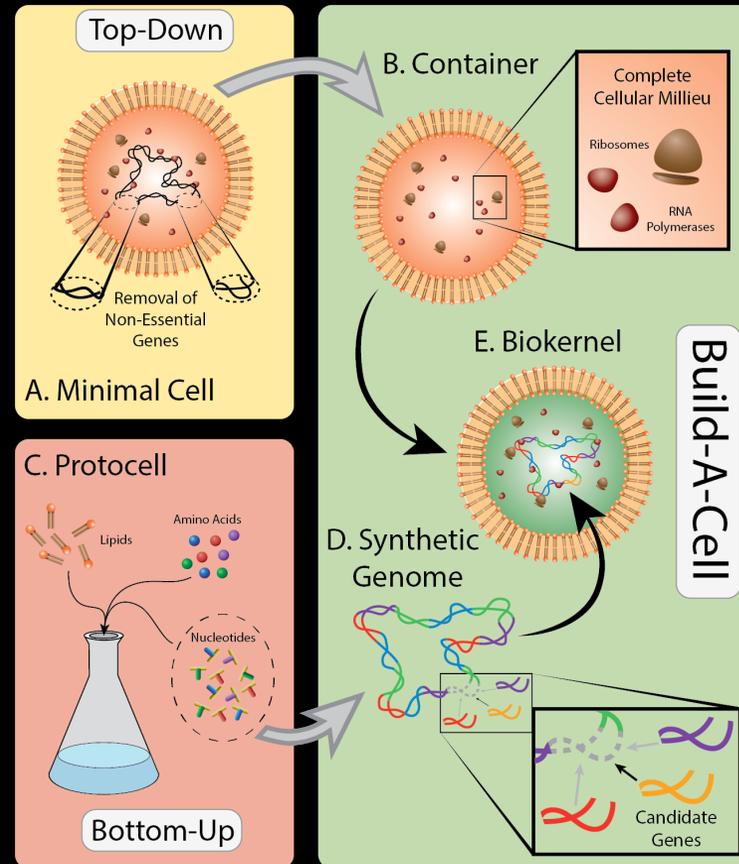
(enabling operational mastery of living matter)

Essential genes of
unknown function

*“what is naturally alive
I do not understand”*
— D. Endy

Unknown functions
that are essential

*“what I cannot create
I do not understand”*
— R. Feynman



1. Information

Essential gene sets
Abstracted functional modules
APIs to (2) and (3)

2. Operation

Cell-free & PURE
Expression architectures, from
gene to operon to genome

3. Measure & Model

Validated DNA via -omics
Molecular ensembles via Cryo-EM
Fluid physics ensemble dynamics

Frontiers of Metrology in Biology? What if...

- NMIs establish biometrology
 - coupled to emerging needs
- We figure out how to establish metrics and comparability for results from complex algorithms
 - bioinformatics is ***part of the measurement process***
 - this isn't new *per se*, but the degree of complexity is significant
- We develop more metrology of “nominal properties”
 - is traceability a useful concept?
 - measurement uncertainty?
 - are there analogues to yield compatibility/comparability?
- We consider metrology of “Completeness” of Knowledgebases...

The Joint Initiative for Metrology in Biology was built to work in this space.

- Collaborative home for measurement science and standards for 'omics and synthetic biology
- NIST, Stanford University, and private sector
 - operated by SLAC
- Watch for series of workshops to scope measurement science, measurement tool, and standards development

JIMB BRINGS TOGETHER

GOVERNMENT

INDUSTRY

ACADEMIA

Academia, Government, and Industry must work together to define a new type of partnership supporting 21st century life sciences and biotechnology. NIST, by establishing a West Coast presence where biotechnology was born, will help ensure the United States remains a global leader in bioscience and biotechnology.

Learn more and join us!
jimb.stanford.edu

SI

Tons of help from...

NIST Genome-Scale Measurements Group, MD & CA, and JIMB

Justin Zook
Jenny McDaniel
Lindsay Harris
David Catoe
Sarah Munro
Scott Pine
Noah Spies

Sasha Levy
Darach Miller

Arend Sidow
Drew Endy

and The Genome in a Bottle Consortium
and The External RNA Controls Consortium

