

# Single-Molecule Sequencing: Methods, Training and Applications

## Lecture 1

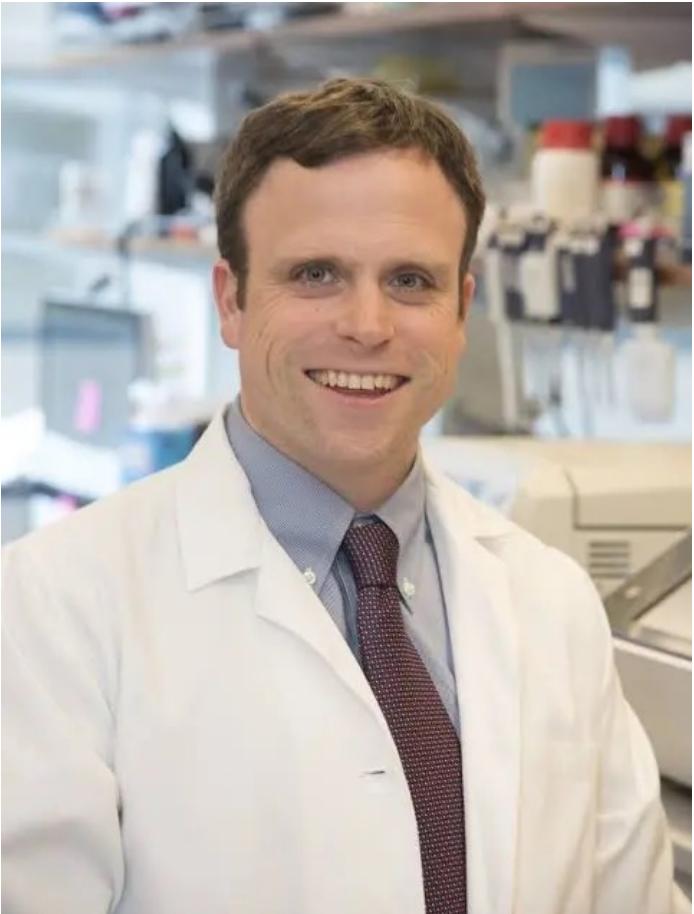
Theodore M. Nelson

Prof. Christopher E. Mason, Ph.D.

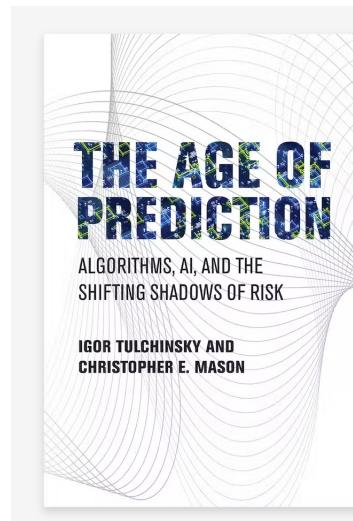
# Pre-Class Survey [Please do not use ChatGPT]



# Instructor Introduction: Christopher E. Mason



Professor  
Director of the WorldQuant Initiative for Quantitative Prediction  
Department of Physiology and Biophysics &  
The Institute for Computational Biomedicine (ICB),  
Meyer Cancer Center, Feil Family Brain and Mind Research Institute,  
at Weill Cornell Medicine,  
Fellow of the Information Society Project, Yale Law School



**The Age of Prediction**  
Algorithms, AI, and the Shifting Shadows of Risk  
By Igor Tulchinsky and Christopher E. Mason

232 pp., 6 x 9 in, 13 b&w illus.  
Hardcover  
ISBN: 9780262047739  
Published: August 22, 2023  
Publisher: The MIT Press

\$27.95 Hardcover

Penguin Random House    Amazon    Barnes and Noble    Bookshop.org    Indiebound  
Indigo    Books a Million



# Single-Molecule Sequencing & Christopher E. Mason

The screenshot shows a Google Scholar search results page for the query "single molecule chris mason". The results are filtered to show "Articles" and there are approximately 27,300 results found in 0.07 seconds. The results are listed in descending order of relevance.

**Search Query:** single molecule chris mason

**Filter:** Articles

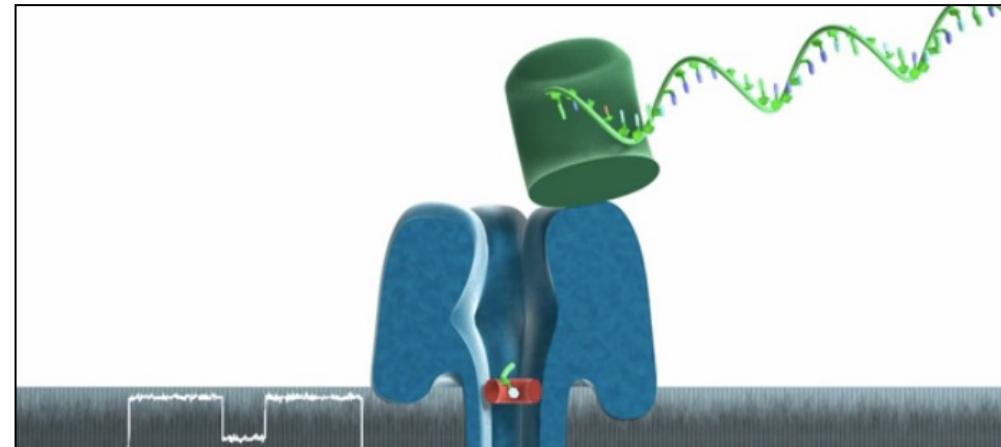
**Results:**

- Assembly and diploid architecture of an individual human genome via single-molecule technologies**  
... RB Darnell, WR McCombie, PY Kwok, CE Mason... - Nature ..., 2015 - nature.com  
... The draft map was further refined by mapping **single molecules** to it and recalculating the label positions. Next the consensus maps were extended by aligning overhanging **molecules** ...  
[PDF] nature.com  
Columbia e-link >>
- Single-molecule sequencing detection of N6-methyladenine in microbial reference materials**  
... D Bezdan, H Sichtig, CY Chiu, CE Mason - Nature ..., 2019 - nature.com  
... Prior to the development of **single molecule** sequencing, methods to detect m 6 A in DNA ...  
Alignment of data from **single molecule** sequencers and comparison of kinetic data to a ...  
[PDF] nature.com  
Columbia e-link >>
- The birth of the Epitranscriptome: deciphering the function of RNA modifications**  
... K Meyer, J Korlach, ID Vilfan, S Jaffrey, CE Mason - Genome biology, 2012 - Springer  
... We present a perspective on current work and new **single-molecule** sequencing methods ...  
In summary, the high-throughput and **single-molecule** methods described here represent the ...  
[PDF] springer.com  
Columbia e-link >>
- DNA methylation-calling tools for Oxford Nanopore sequencing: a survey and human epigenome-wide evaluation**  
... N Jillette, P Wang, A Taghbalout, J Fox, C Mason... - Genome biology, 2021 - Springer  
... Thus, to enable assessment of this critical epigenetic heterogeneity, we have evaluated the DNA methylation calling accuracy at **single-molecule** and **single-base** resolution, which is ...  
[PDF] springer.com  
Full View
- Characterizing multi-omic data in systems biology**  
CE Mason, SG Porter, TM Smith - Systems Analysis of Human Multigene ..., 2013 - Springer  
... for directly sequencing **single molecules** of RNA [44] and other **single-molecule** methods (...  
In early demonstrations, **single-molecule** RNA-Seq has shown that RNAs have extremely ...  
[PDF] academia.edu
- Extensive sequencing of seven human genomes to characterize benchmark reference materials**  
... L Vang, N Spies, A Sidow, Z Weng, Y Liu, CE Mason... - Scientific data, 2016 - nature.com  
... Genomic DNA from the Ashkenazi Jewish (AJ) son was prepared for sequencing via the Oxford Nanopore Technologies MinION **single molecule** sequencing instrument. Two libraries ...  
[PDF] nature.com  
Columbia e-link >>

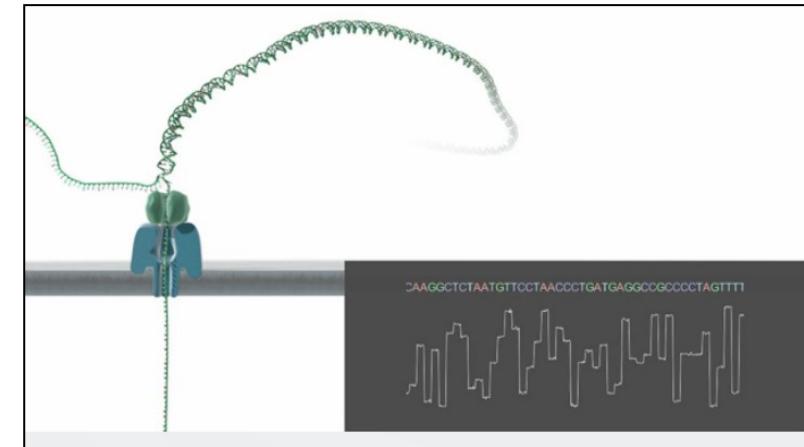
# Single-Molecule Sequencing in the Mason Lab



2014: Sequencing with  
a protein nanopore



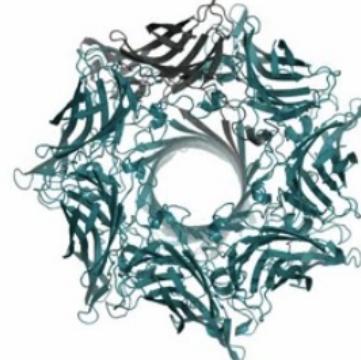
Exonuclease-Seq



Strand-Seq



MinION



## Nanopore Platforms



PromethION



# Really Long Reads!

Longer and longer: DNA sequence of more than two million bases now achieved with nanopore sequencing.

Fri 4th May 2018

## Congratulations!

The first >2 Mb DNA read, achieved with nanopore sequencing

Matt Loose, Alex Payne, Nadine Holmes, Vardhman Rakyan & team, University of Nottingham, UK

May 2018

Long read club

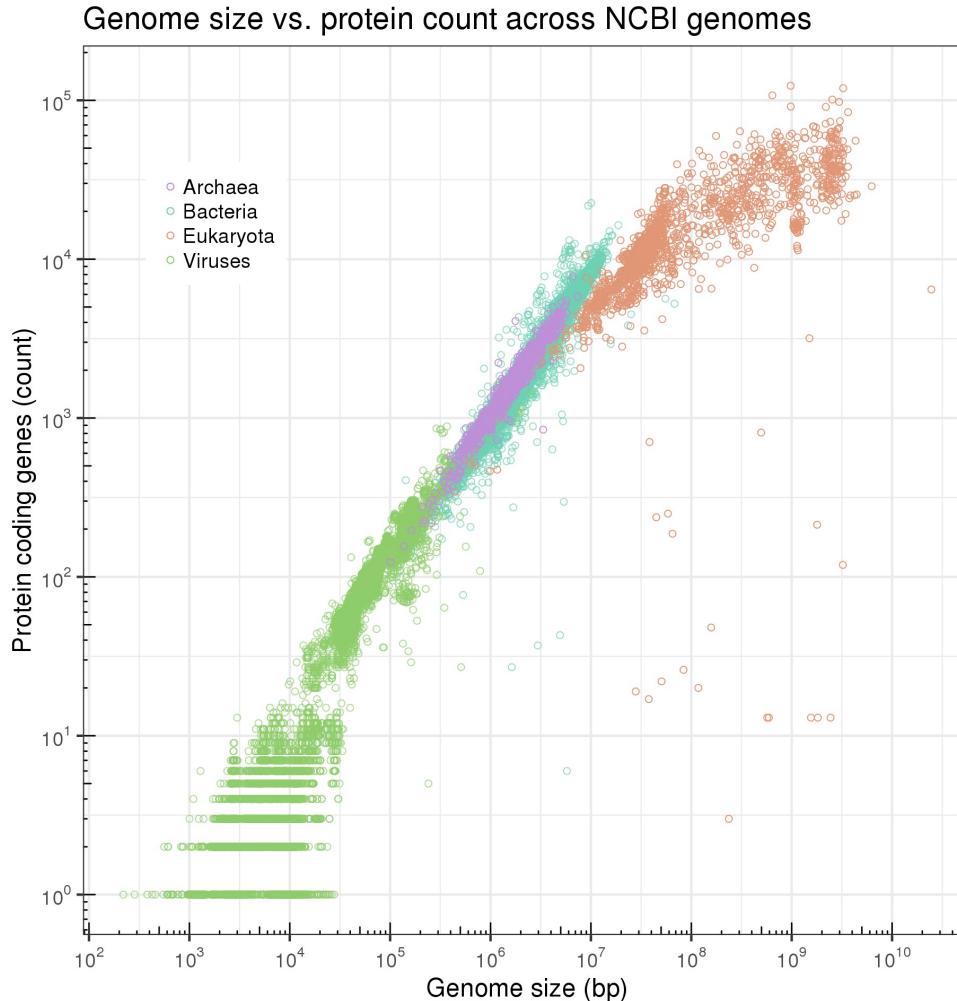


Really very long reads indeed

<http://longreadclub.org/>

<https://nanoporetech.com/about-us/news/longer-and-longer-dna-sequence-more-two-million-bases-now-achieved-nanopore>

# 4.2 Mb is the record so far, covers many genomes



<https://www.cgm.northwestern.edu/cores/nuseq/services/next-generation-sequencing/nanopore-sequencing.html>

# A Nanopore Revolution in Lake Fryxell, Antarctica!



# Sequencing HW DNA - Sarah Johnson (PI) expedition G062 team



**IBT**  
JOURNAL OF BIOMOLECULAR TECHNIQUES

J Biomol Tech. 2017 Apr; 10(17):2801-009  
Published online 2017 Mar 22. doi: 10.7171/jbt.17-2801-009  
PMCID: PMC5362181

**Real-Time DNA Sequencing in the Antarctic Dry Valleys Using the Oxford Nanopore Sequencer**

Sarah S. Johnson,<sup>1,2\*</sup> Elena Zaitkova,<sup>1</sup> David S. Goerlitz,<sup>3</sup> Yu Bai,<sup>1</sup> and Scott W. Tighe<sup>4</sup>

[Author Information](#) | [Copyright and License Information](#)

**Abstract**

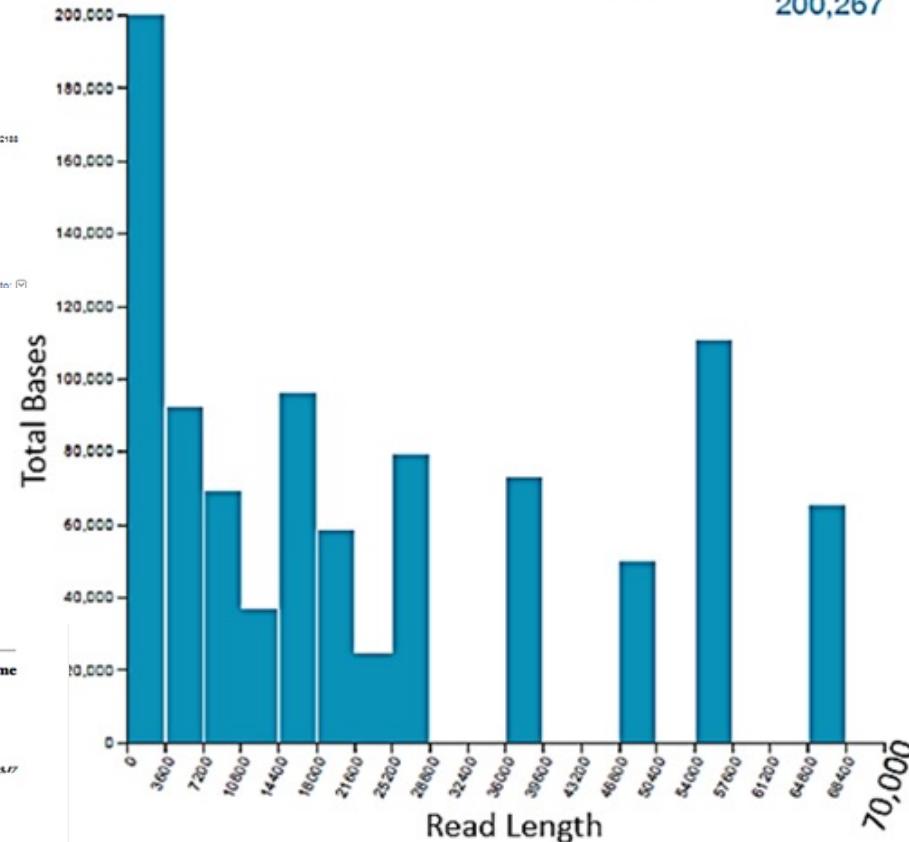
**ARTICLE**

**Genomic Methods and Microbiological Technologies for Profiling Novel and Extreme Environments for the Extreme Microbiome Project (XMP)**

Scott Tighe,<sup>1,4,5</sup> Ebrahim Afshinnekoo,<sup>2,5,6,8</sup> Tara M. Rock,<sup>5</sup> Ken McGrath,<sup>6</sup> Noah Alexander,<sup>2,3</sup> Alexa McIntyre,<sup>2,3</sup> Sofia Abusneffer,<sup>2,3</sup> Daniela Bezdán,<sup>2,3</sup> Stefan J. Green,<sup>7</sup> Samantha Joye,<sup>8</sup> Sarah Stewart Johnson,<sup>9</sup> Don A. Baldwin,<sup>10</sup> Nathan Bivens,<sup>11</sup> Nadim Ajami,<sup>12,13</sup> Joseph R. Carnical,<sup>12,13</sup> Ian Charold Herriott,<sup>14</sup> Rita Colwell,<sup>15</sup> Mohamed Noor,<sup>16</sup> Jonathan Fook,<sup>2,5,17</sup> Nick Greenfield,<sup>18</sup> Tim Hunter,<sup>1</sup> Jessica Hoffmann,<sup>1</sup> Joshua Hyman,<sup>19</sup> Ellen Jorgensen,<sup>20</sup> Diana Krawczyk,<sup>21</sup> Jodie Lee,<sup>22</sup> Shawn Levy,<sup>23</sup> Natália Garcia-Reyero,<sup>24</sup> Matthew Settles,<sup>25</sup> Kelley Thomas,<sup>26</sup> Felipe Gómez,<sup>27</sup> Lynn Schriml,<sup>28,29</sup> Nikos Kyrpides,<sup>30</sup> Elena Zaitkova,<sup>1</sup> Jon Pesterman,<sup>31</sup> and Christopher E. Mason<sup>2,3,8,32,33</sup>

Total 587 Events 954955 Channels 185 of 512

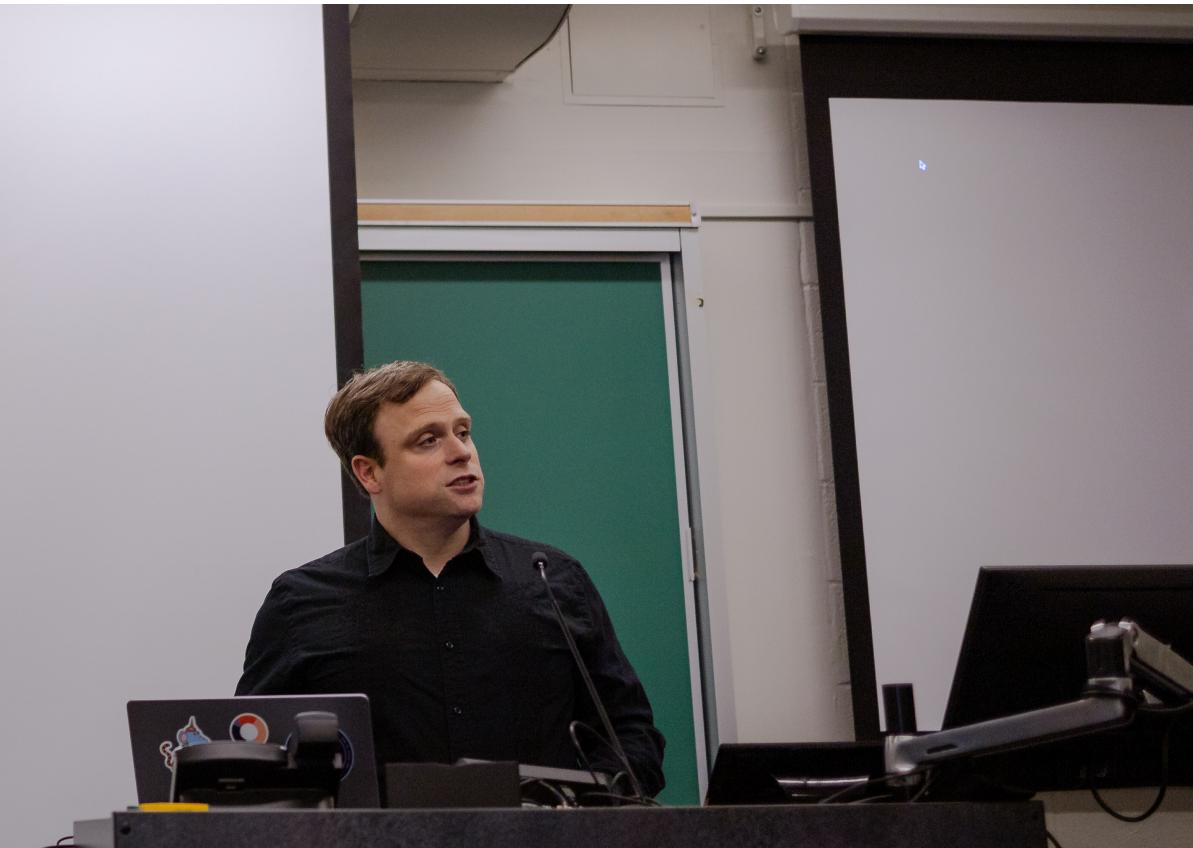
200,267



# Nanopore Sequencing in Space. Why?

More than ever,  
we are a  
space-faring  
species

Photo Credit: Claudio Solano



# Zero-G Pipetting: Hardest Lab Job Ever



Dr. Andrew Feinberg



# Nanopore Sequencing in Simulated Zero-Gravity Conditions



www.nature.com/news/zero-gravity-genomics-passes-first-test-1.18537

**nature** International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For Authors

News & Comment > News > 2015 > October > Article

NATURE | NEWS

## Zero-gravity genomics passes first test

Two experiments demonstrate sample transfer and sequencing in a low-gravity environment.

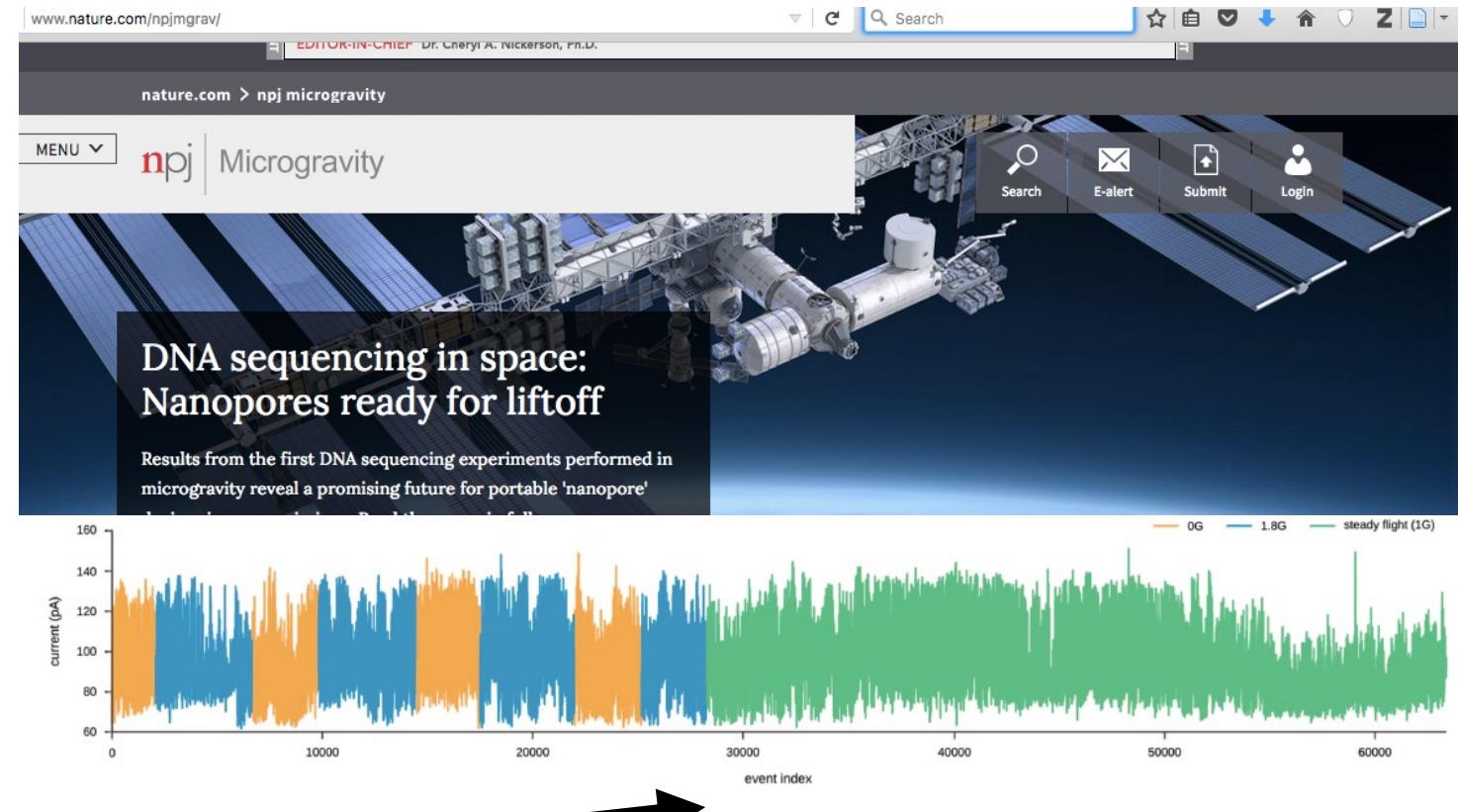
Chris Cesare

13 October 2015

Rights & Permissions

After 160 swoops in NASA's zero-gravity aeroplane, researchers have the first evidence that genetic sequencing can be done in space.

# Nanopore Sequencing in Simulated Zero-Gravity Conditions

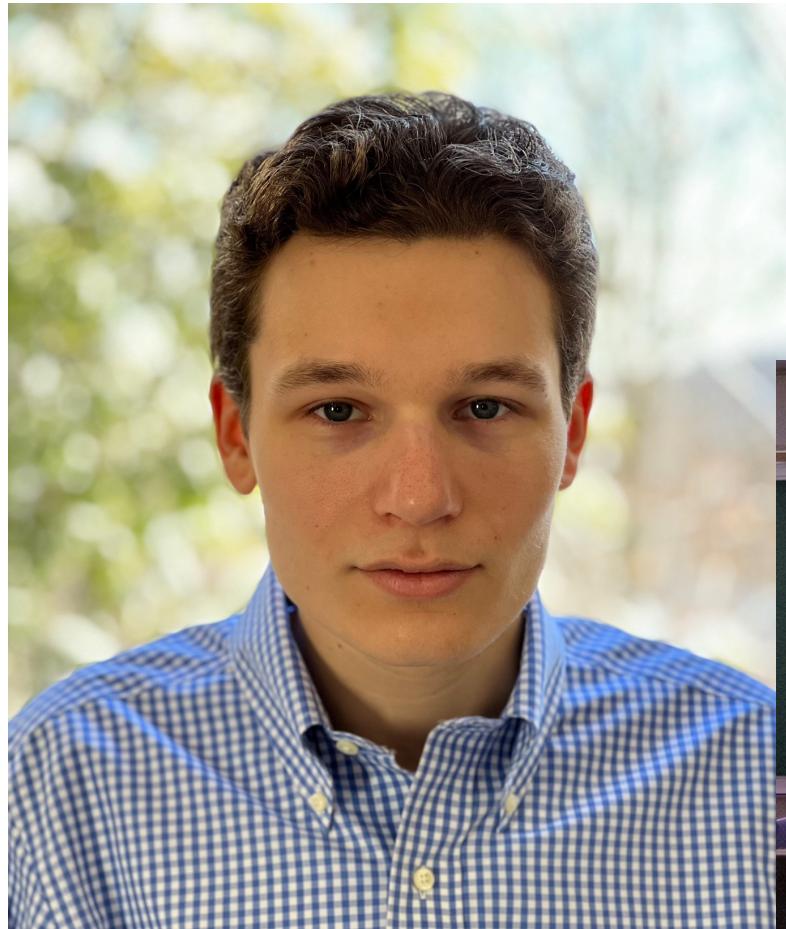


The nanopores are ready!

# Scientific Cliffhanger

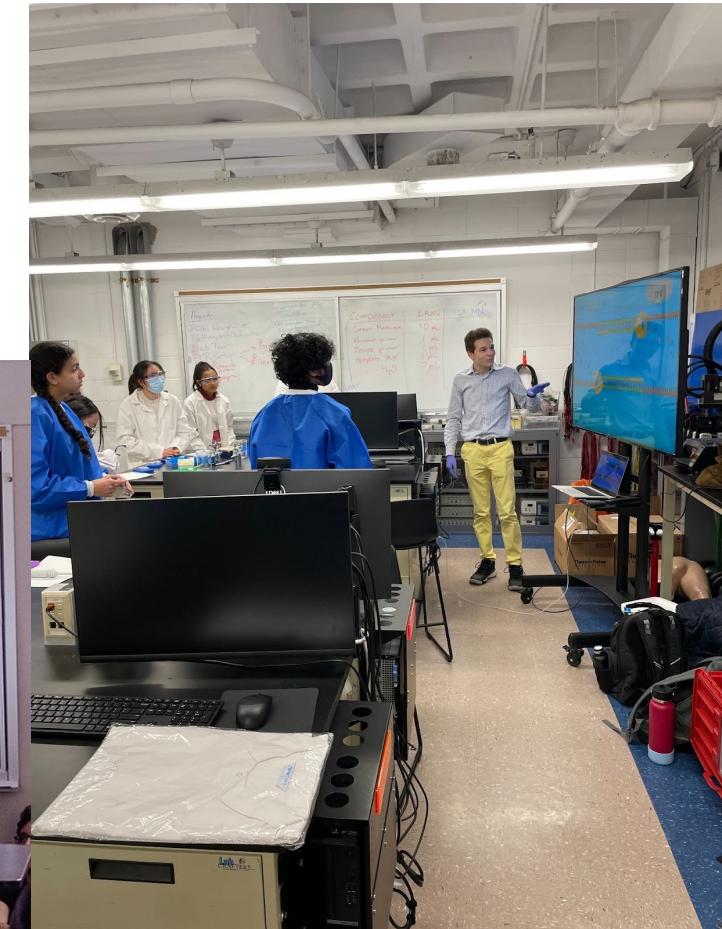
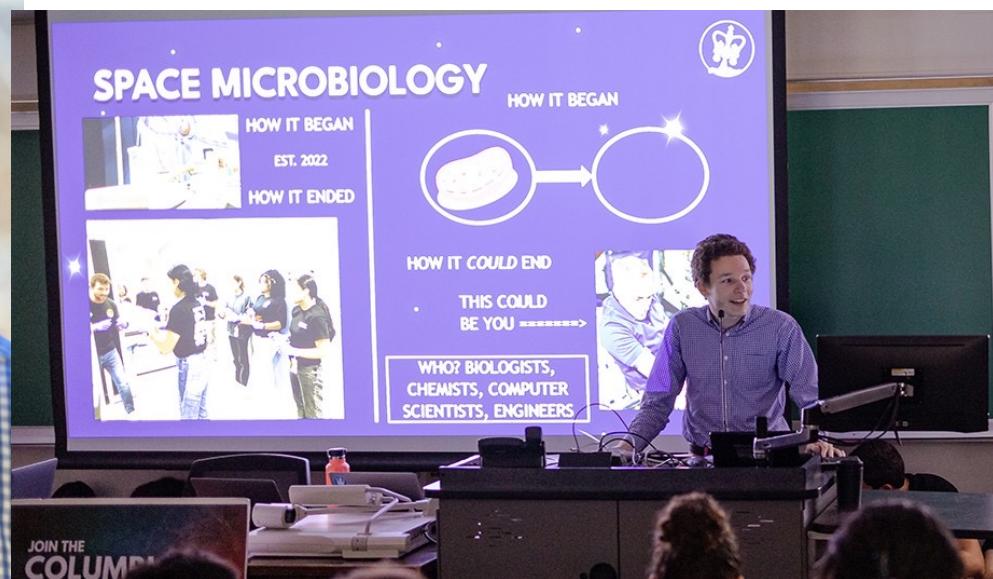
**</to be continued in Lecture 3>**

# Instructor Introduction: Theodore M. Nelson



Churchill Scholar

Undergraduate Student  
Columbia University



# Instructor Introduction: Theodore M. Nelson



Data is produced



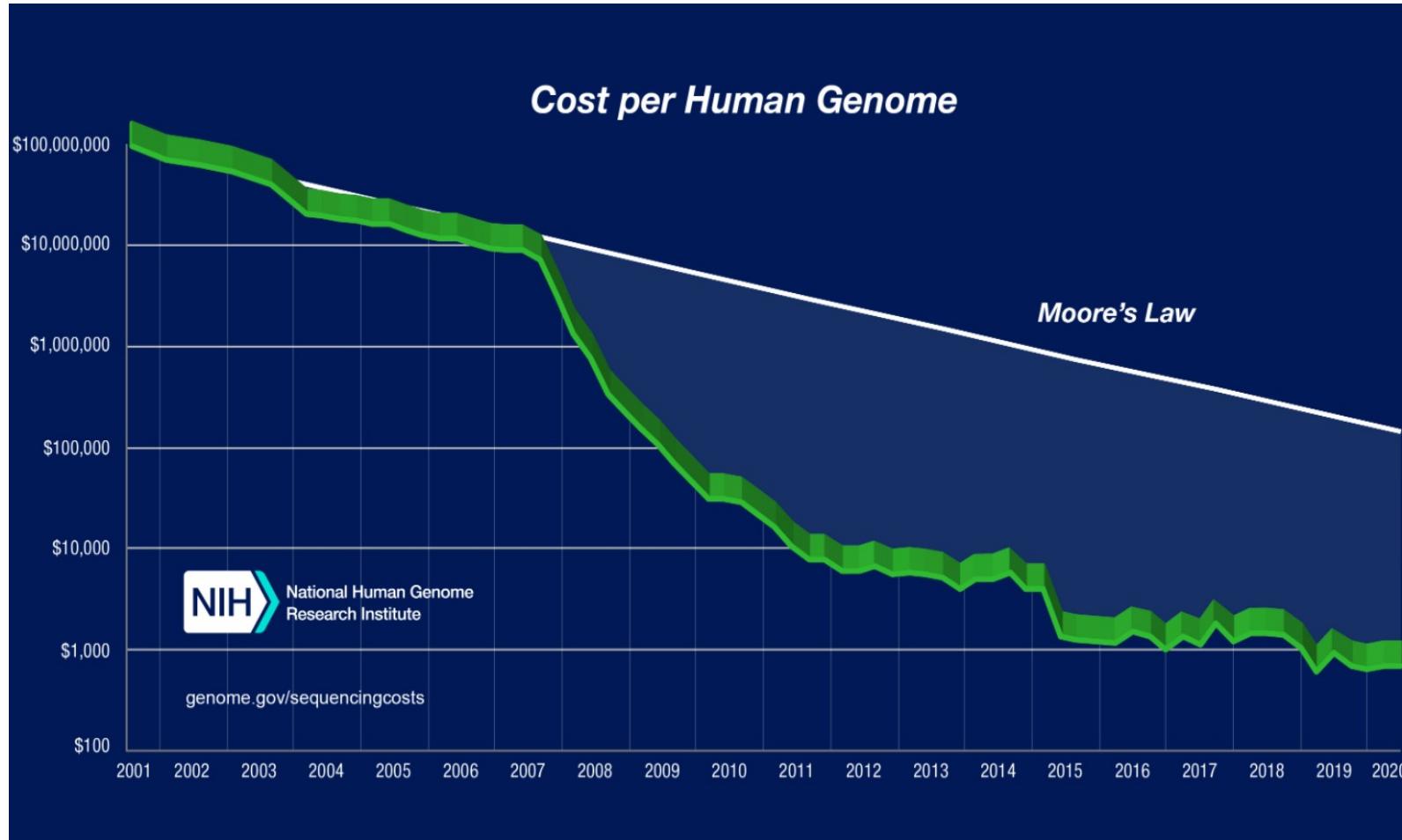
# Poll Question: What is not Single-Molecule Sequencing (SMS)?



# What is not Single-Molecule Sequencing (SMS)?

**<student answers here>**

# SMS did not drive the the last ten years of sequencing revolution



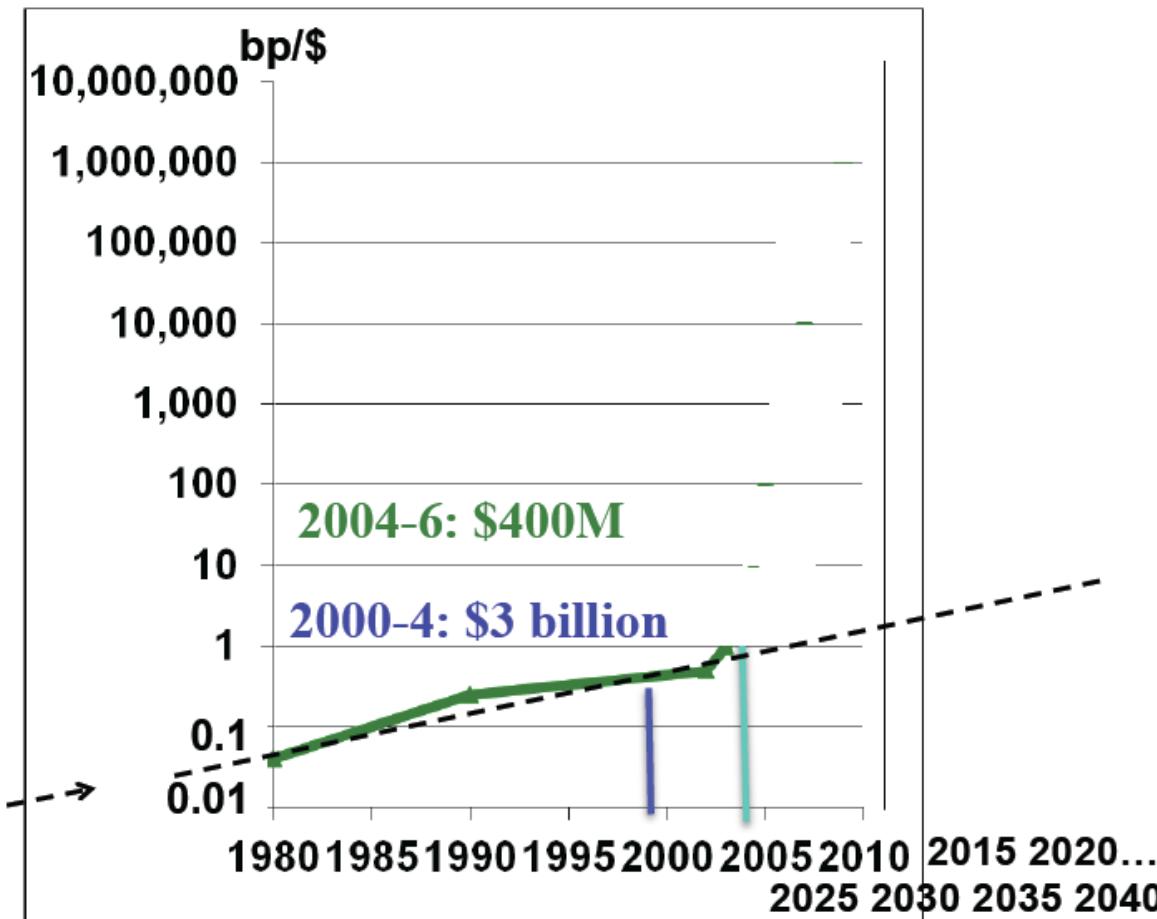
# Initially we expected a \$1K Genome in 2040

\$1000  
Genome

When?

2040

-----  
Moore's law  
1.5x/yr for  
electronics



# The \$1,000 Genome in 2014

**nature** International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For Authors

Archive > Volume 507 > Issue 7492 > News Feature > Article

NATURE | NEWS FEATURE

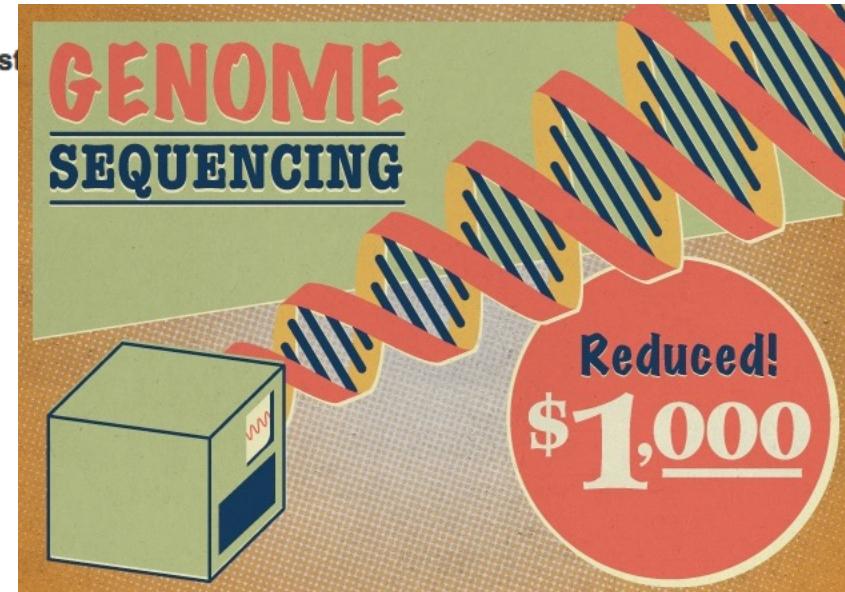


## Technology: The \$1,000 genome

With a unique programme, the US government has managed to drive the cost of sequencing down towards a much-anticipated target.

Erika Check Hayden

19 March 2014



# The \$100 Genome! 😊

<https://www.statnews.com/2017/01/09/illumina-ushering-in-the-100-genome/>

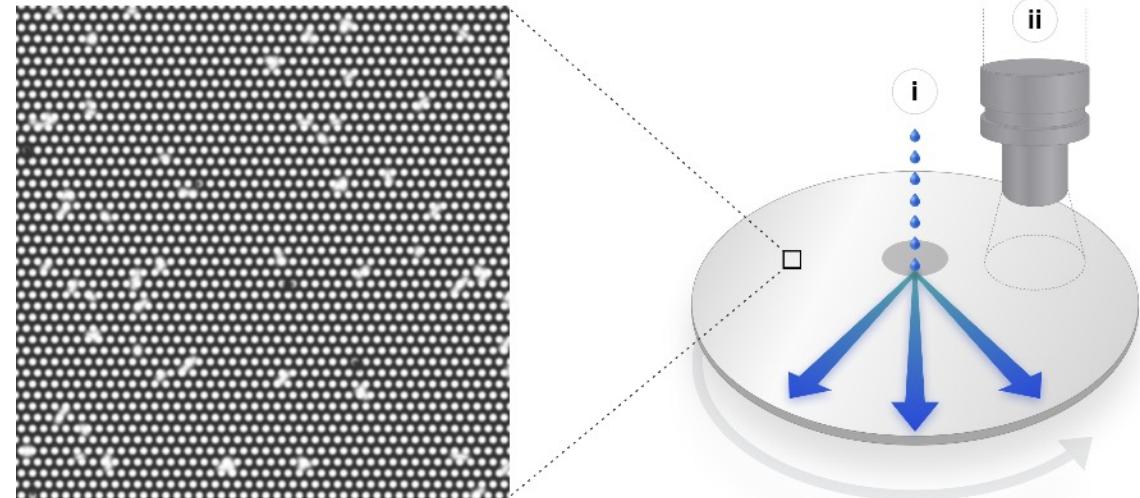
STAT Sections Topics Multimedia Popular STAT Plus Q

BUSINESS

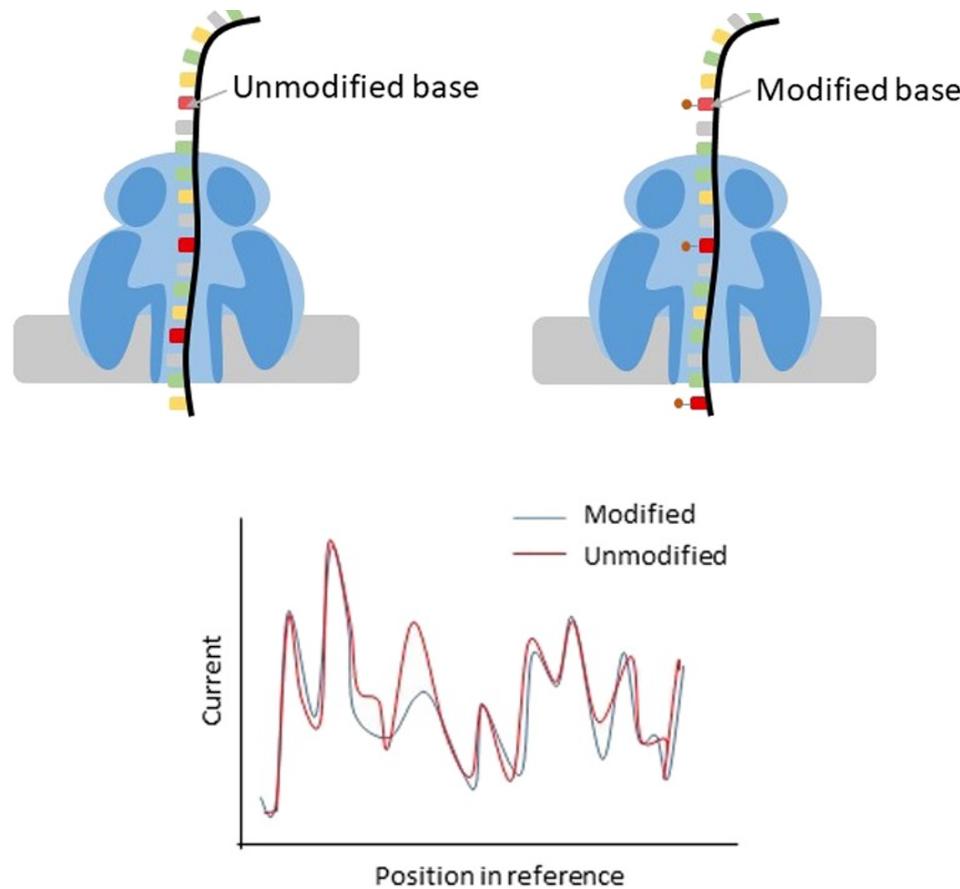
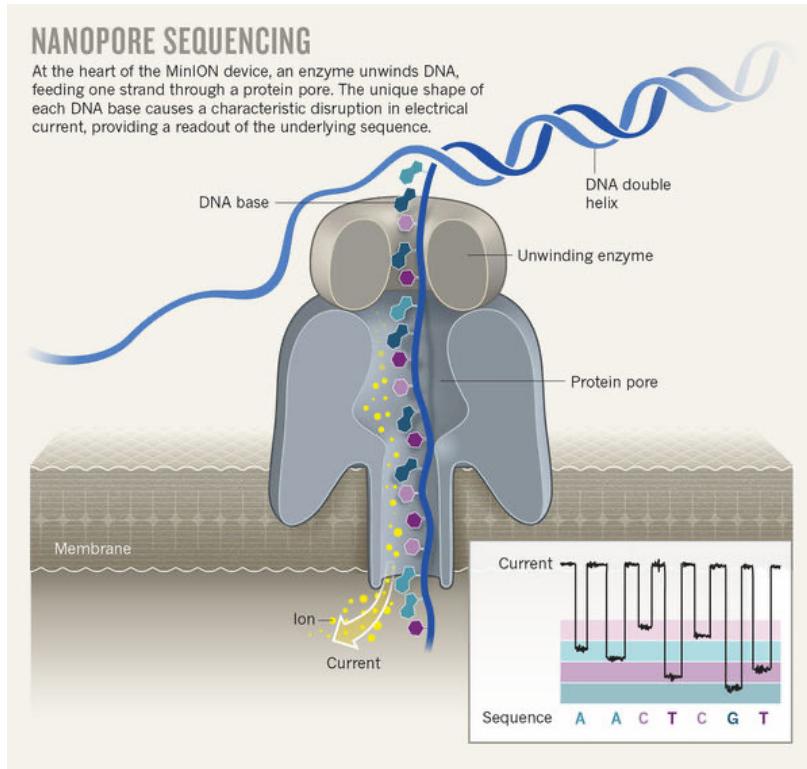
Illumina says it can deliver a \$100 genome — soon



ULTIMA  
GENOMICS

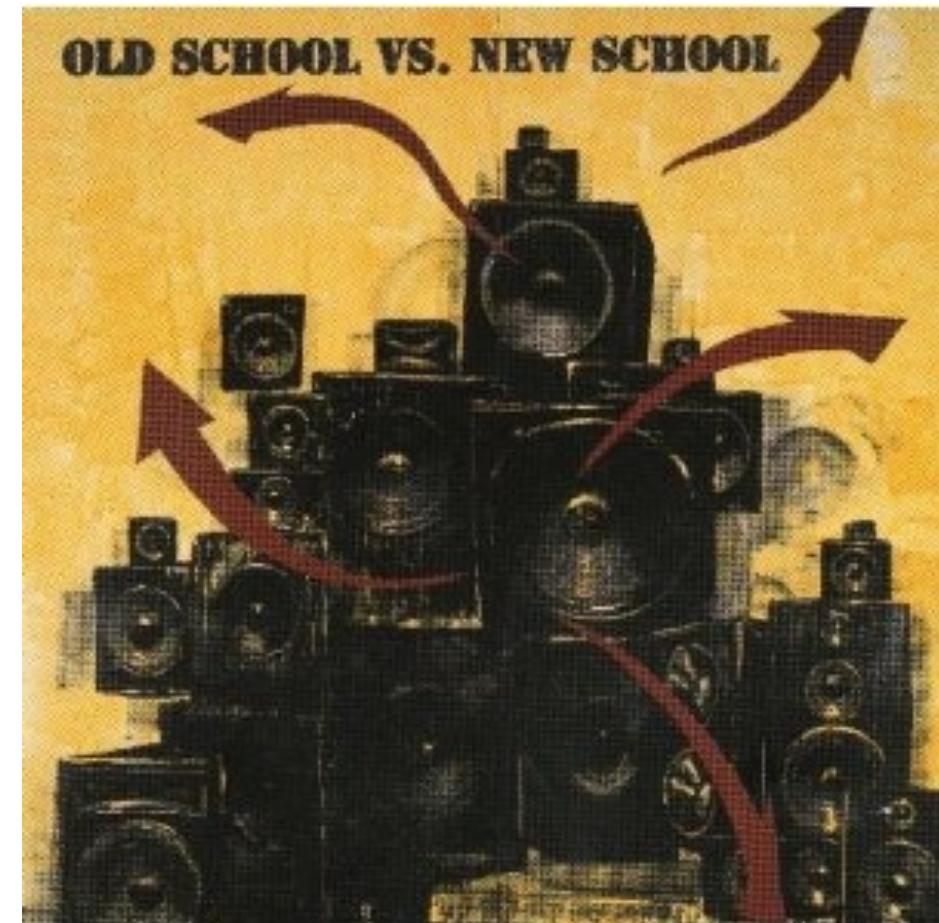


# But SMS likely will drive the next ten years!

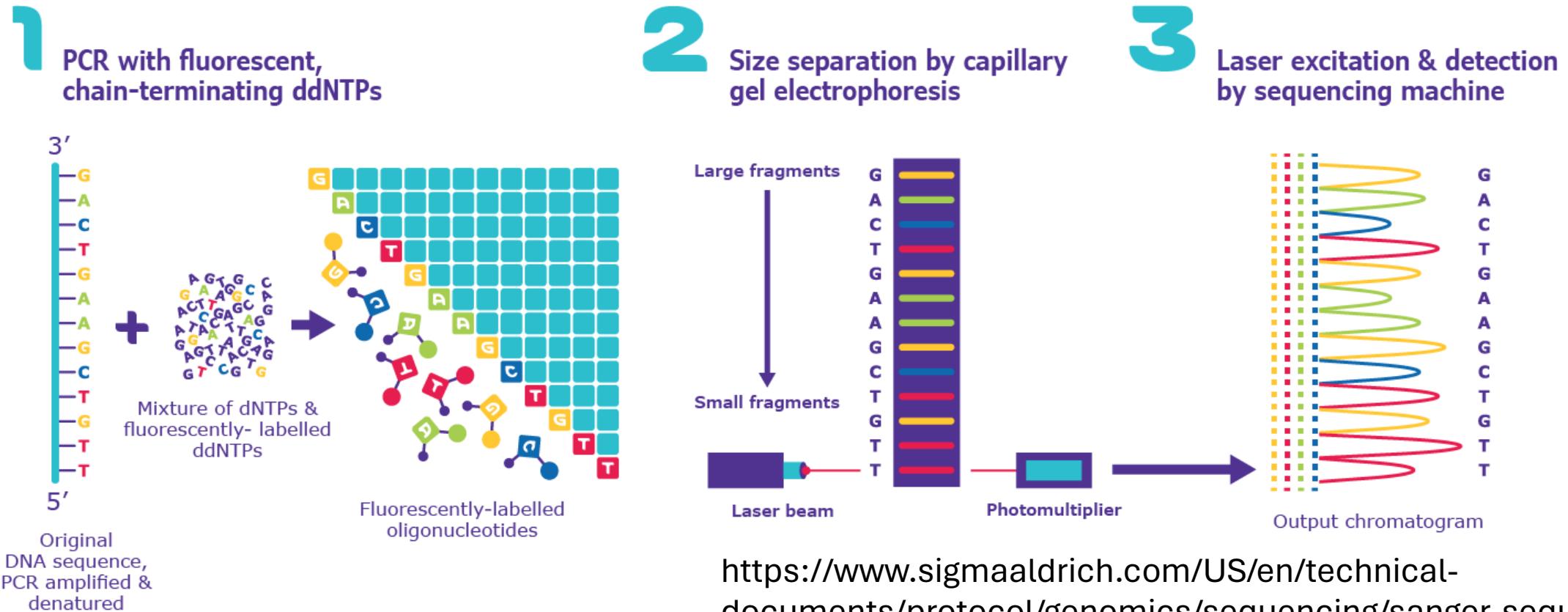


# Which sequencing technologies are not SMS?

1. “Old School” dye-terminator sequencing (Sanger). 300-1000bp
  
2. “New School” methods
  - a. Emulsion PCR Pyrosequencing
  - b. Solid-phase amplification sequencing by synthesis (clonal)
  - c. Sequencing by ligation
  - d. Electrical sequencing



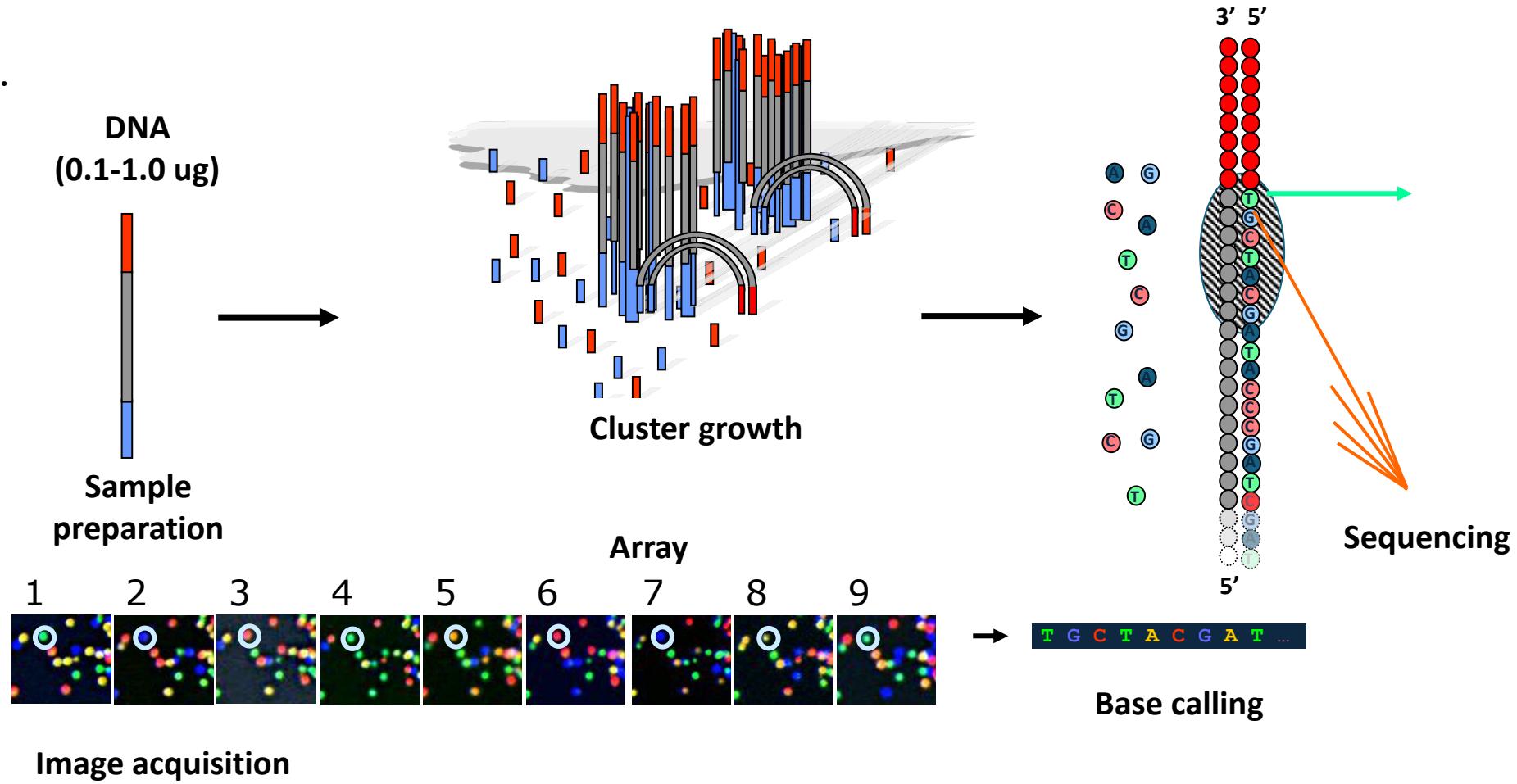
# Sanger sequencing relies on millions of oligonucleotides



<https://www.sigmaaldrich.com/US/en/technical-documents/protocol/genomics/sequencing/sanger-sequencing>

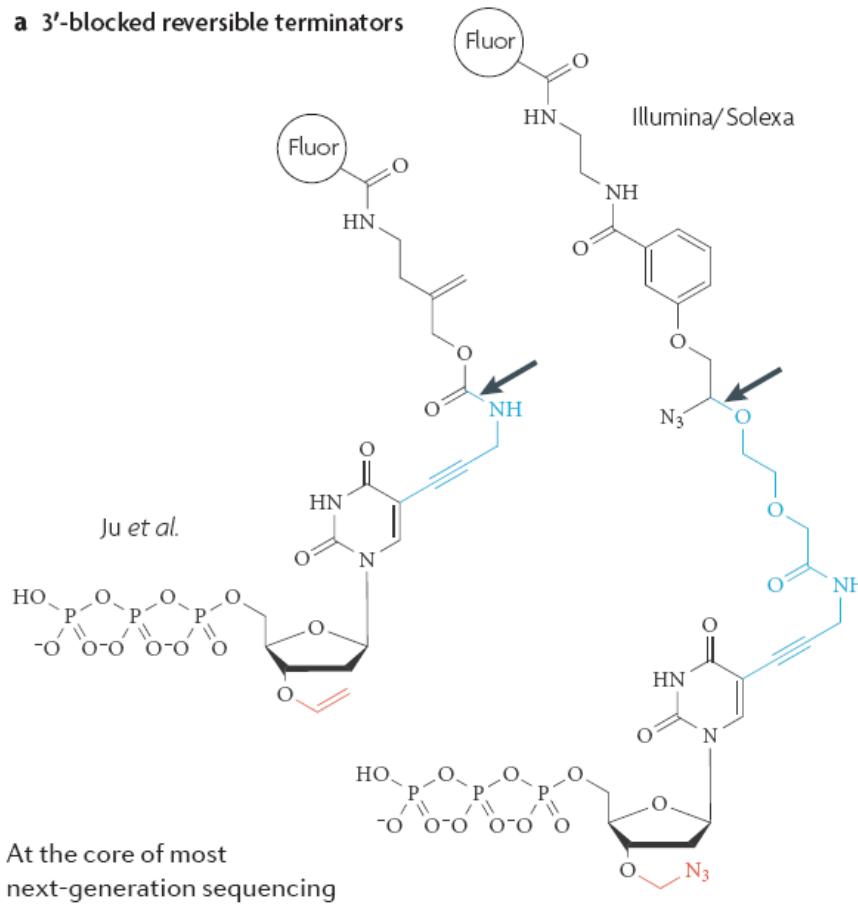
# Illumina Sequencing Relies on Cluster Amplification

© Illumina, Inc.

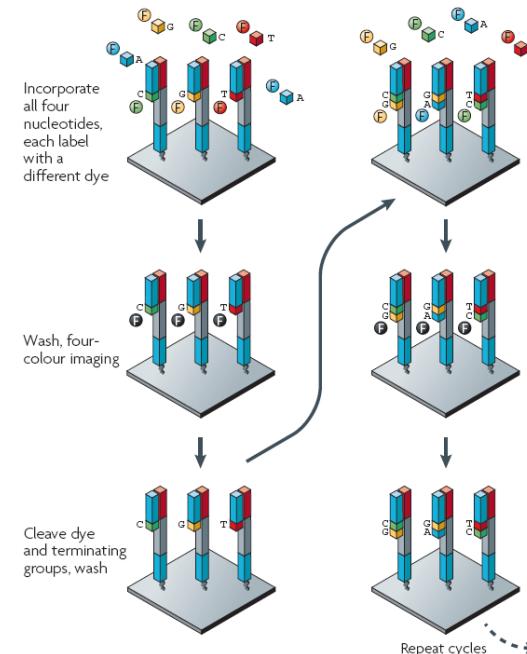


# These Technologies Rely on Reversible Terminators

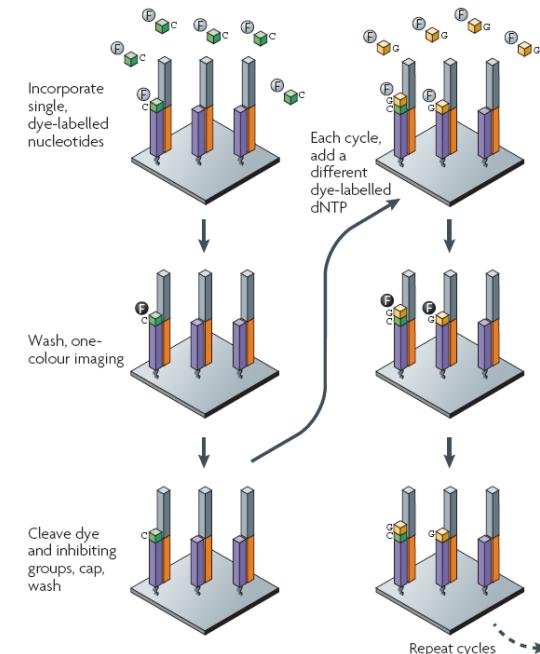
a 3'-blocked reversible terminators



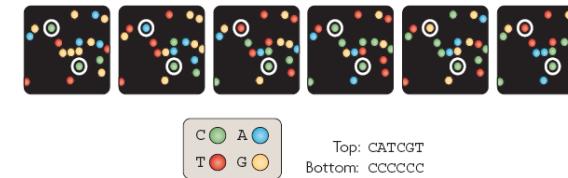
a Illumina/Solexa — Reversible terminators



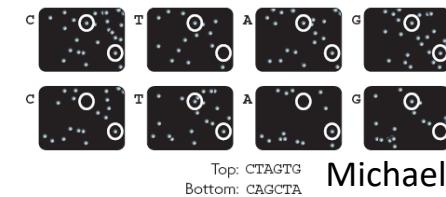
c Helicos BioSciences — Reversible terminators



b



d



Michael Metzker, 2010

# Different Fluorescence Chemistries in Illumina Sequencing

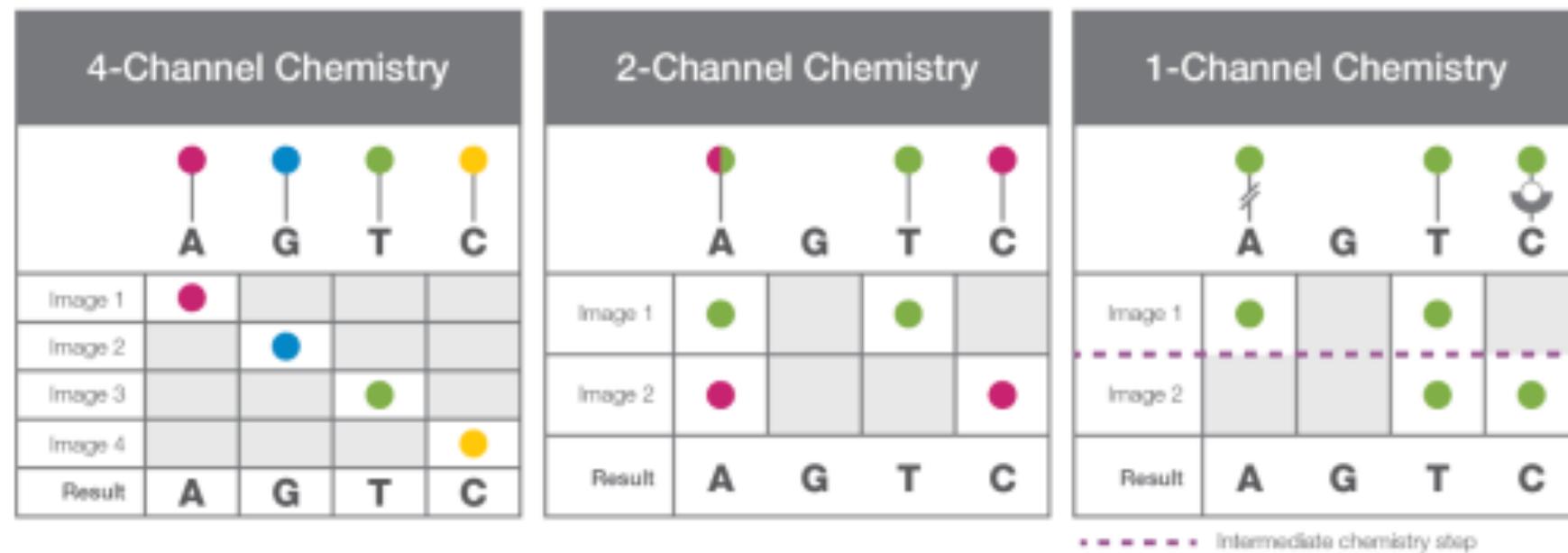
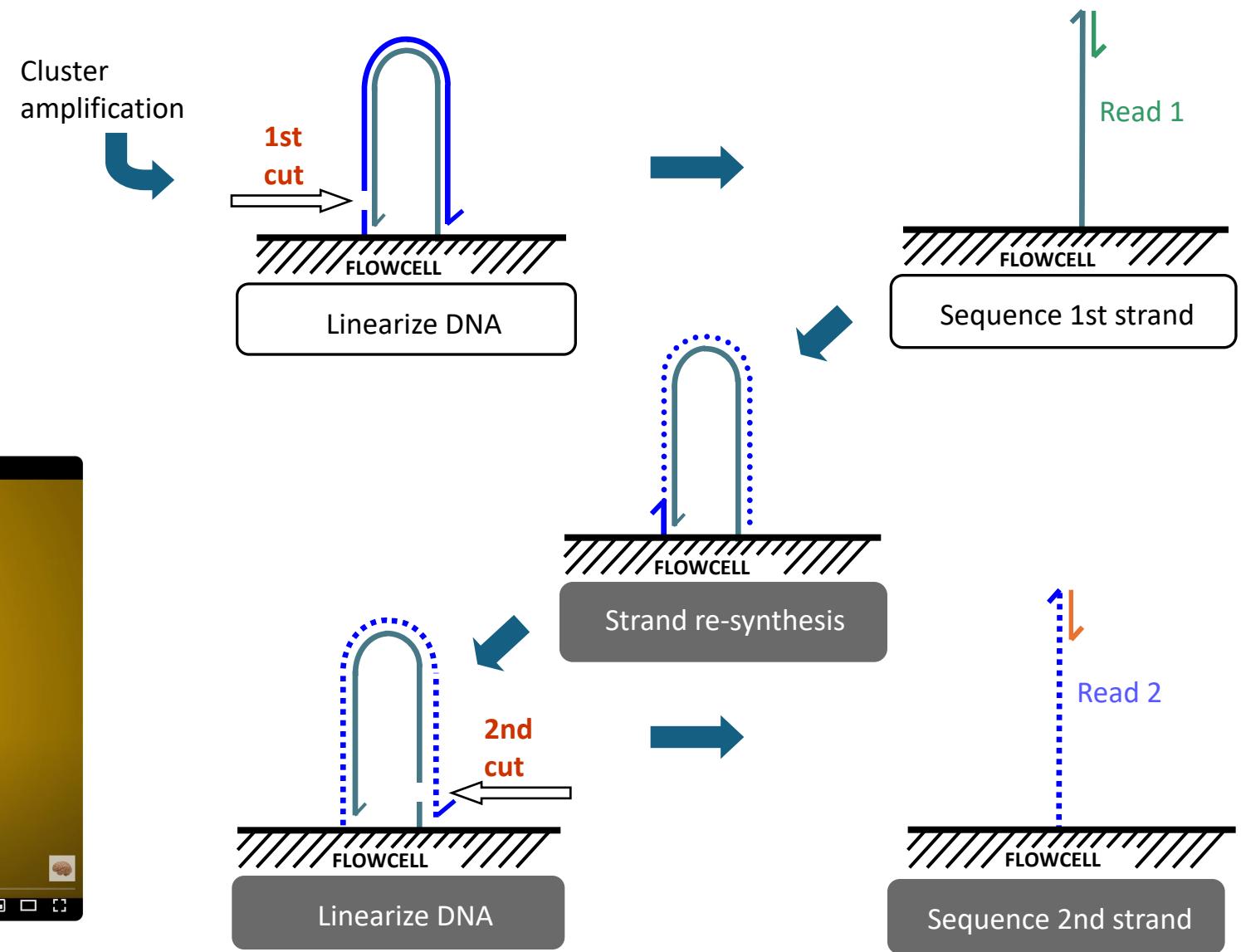
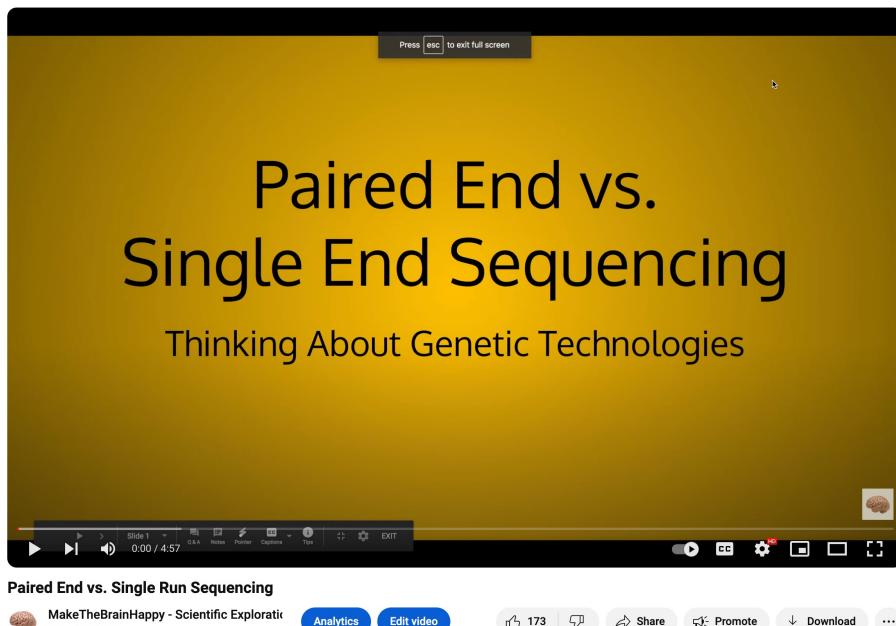
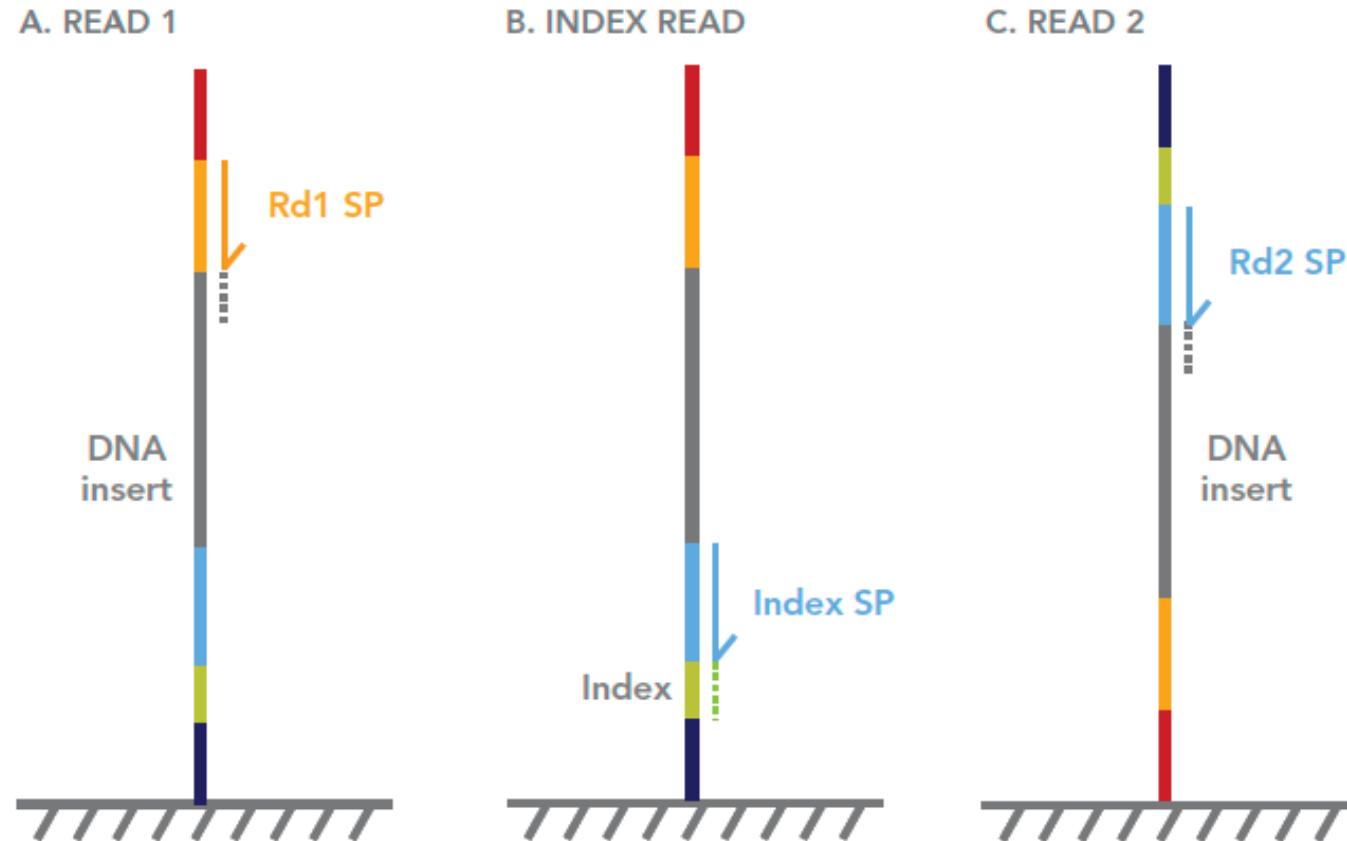


Figure 2: Four-, Two-, and One-Channel Chemistry—Four-channel chemistry uses a mixture of nucleotides labeled with four different fluorescent dyes. Two-channel chemistry uses two different fluorescent dyes, and one-channel chemistry uses only one dye. The images are processed by image analysis software to determine nucleotide identity.

You can read  
your cluster  
from both ends to  
a maximum of 300 bp



# Indexed sequencing method is now standard for single and paired reads



# Emerging Technologies in Next-Gen Seq

# Roche's nanopore tech



Sequencing by eXpansion (SBX)

<https://sequencing.roche.com/en/science-education/technology/nanopore-sequencing.html>

# BGI High-Throughput Sequencer



Sequencing Services

Pharma Solutions

Research Areas

Resources

## News

10/31/2018



### BGI Unveils New High-Throughput Sequencing System.

Last week at the 13<sup>th</sup> International Conference on Genomics (ICG-13) in Shenzhen, China, BGI announced a new sequencing system based on its DNBseq™ Technology.

The newly unveiled MGISEQ-T7 is the most powerful sequencing system from BGI's MGI subsidiary, with a daily output capability of 6Tb of data.

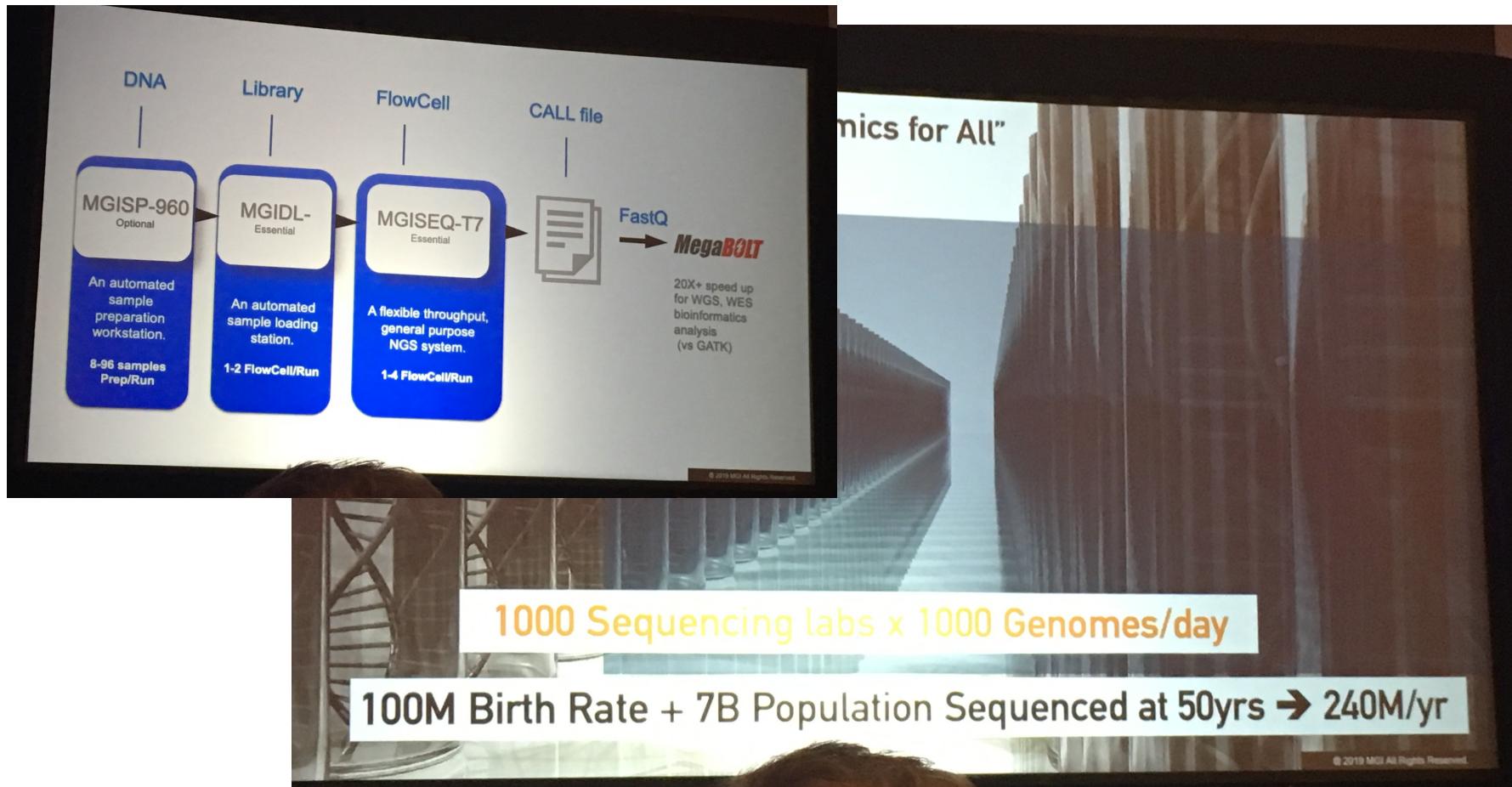
The MGISEQ-T7 is able to complete 60 human genomes in a single day, with essentially error-free sequencing from BGI's DNBseq sequencing technology.

<https://www.bgi.com/us/company/news/bgi-unveils-new-high-throughput-sequencing-system/>

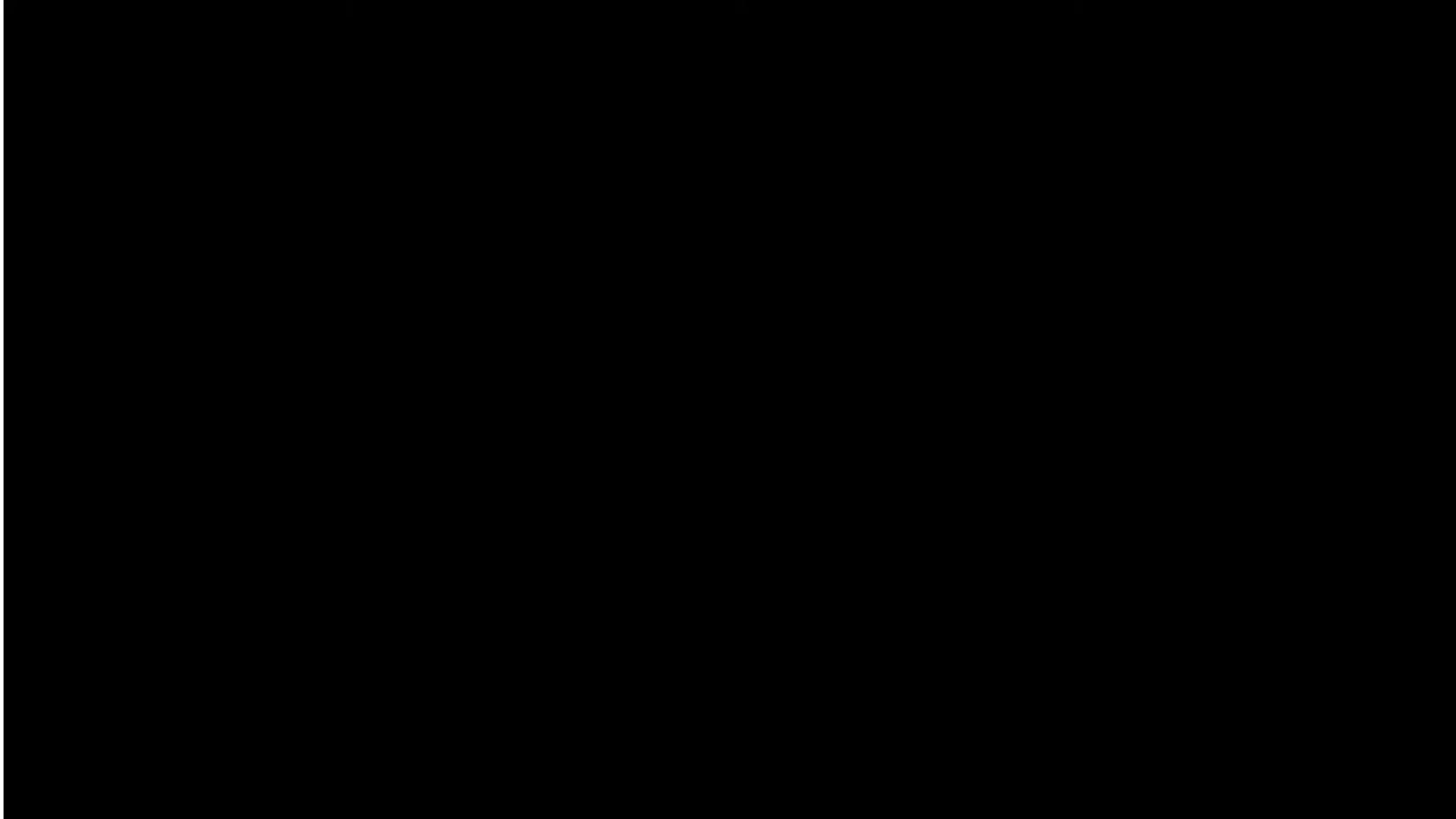
T-1000 Anyone?



# BGI – NGS streets



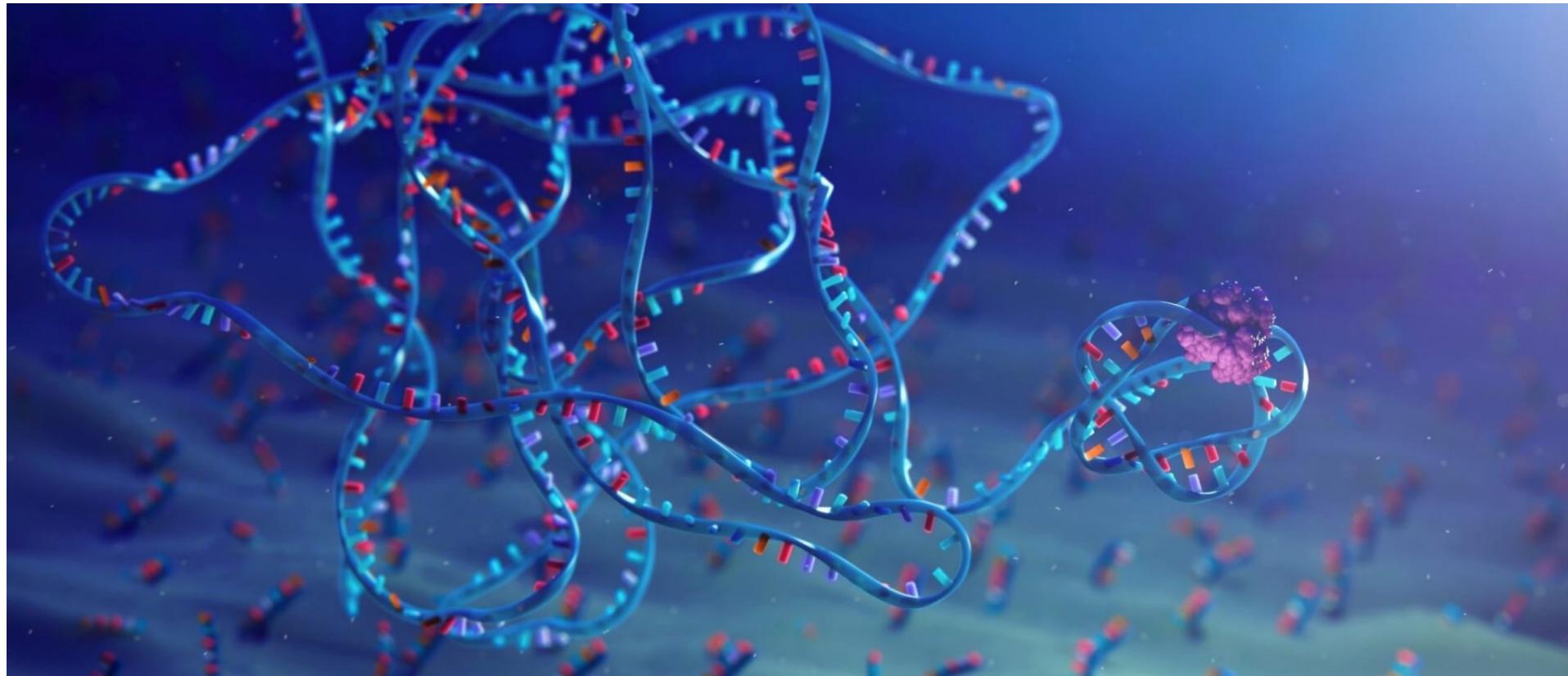
# New Kid on the Block: Element Biosciences



Element  
Biosciences

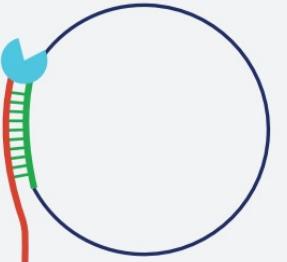
<https://www.elementbiosciences.com/technology>

# Surface-anchored amplification sites



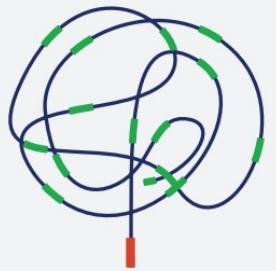
Element  
Biosciences

<https://www.elementbiosciences.com/technology>



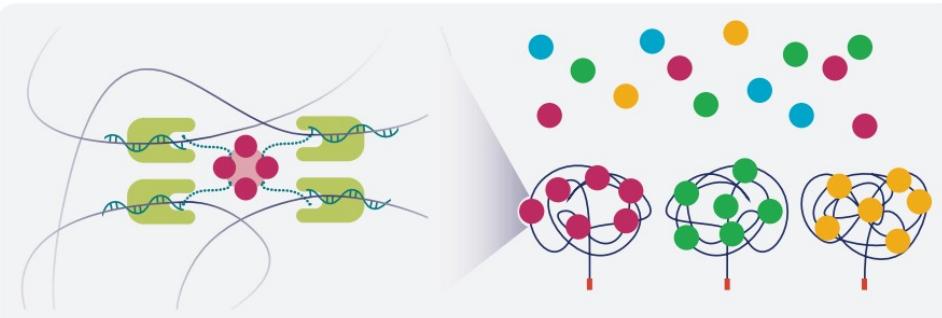
### Hybridize DNA

Circular DNA template attaches to **surface primers** coating the flow cell via an **adapter**.



### Generate Polonies

RCA copies the template DNA and rolls each strand into a tightly bound polony—no PCR, no copies of copies.

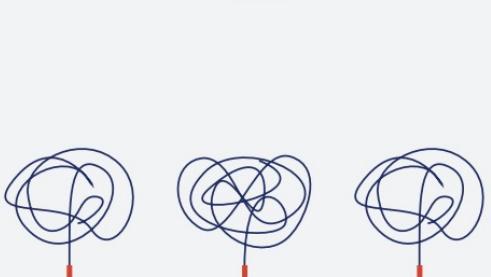


An **avidite** is a dye-labeled polymer with multiple nucleotide arms carrying the same nucleotide base.

### Bind Polonies and Avidites

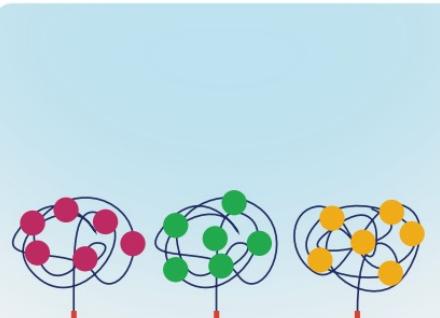
An engineered polymerase binds a primer-hybridized polony and an avidite nucleotide arm without incorporation or extension. Many arms bind to primer-hybridized regions in a polony, creating an ultrastable multivalent complex. Polymerase base-pair discrimination binds the correct avidite to each polony.

SBS combines base detection and strand extension, consuming micromolar reagent concentrations to complete the reaction while signal persists. In contrast, ABC separates the steps and leverages multivalent binding stability to require only nanomolar avidite concentrations. Multiple binding events for each avidite ensure persistent signal for base detection.



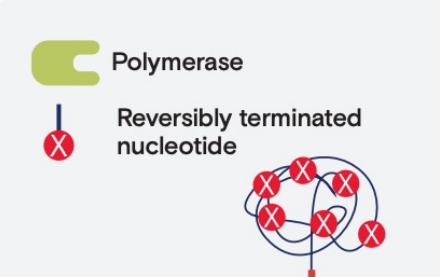
### Remove Blocks

Removing and converting blocking groups to terminal 3' hydroxyl groups allows the next cycle to begin on the primer strands, through ~600 cycles.



### Wash Avidites

A wash removes unbound avidites to leave only avidites bound to polonies. Ultratight binding enables a 100X reagent dilution, fundamentally decreasing cost.



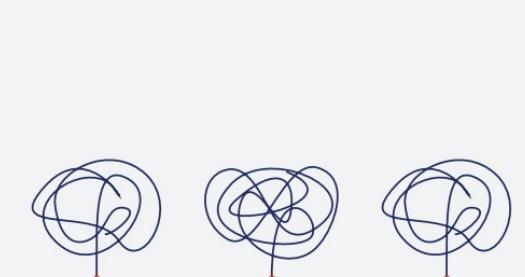
### Step and Block

An engineered polymerase incorporates an unlabeled, blocked nucleotide to extend hybridized primers by a base.



### Detect Bases

Florescent signals in four channels correlate with A, T, G, or C avidites. Low-binding surface chemistry makes the signals more prominent for highly accurate detection.



### Remove Avidites

Buffers remove avidites from the polonies.

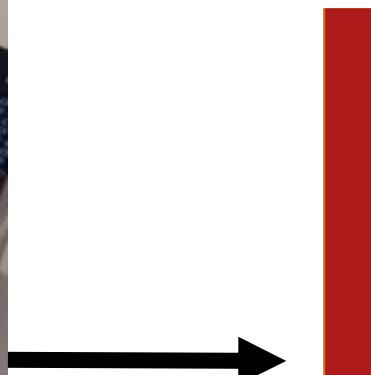
# Each Platform has various sources of noise, and thus Error

- De-Phasing
  - Lagging strand dephasing from incomplete extension
  - Leading strand dephasing from over-extension
- Dark Nucleotides
- Polymerase errors ( $10^{-5}$  to  $10^{-7}$ )
- Single molecule challenges
  - High noise
  - Polymerase “wiggling” from tail
- Platform-specific errors
  - Illumina more likely to have error after ‘G’
  - PCR-based methods miss GC- and AT-rich regions

Besides Nanopore, are you familiar with any other single-molecule seq (SMS)?

**<raise your hands>**

# Course Focus: Depth over Breadth [with good translational potential]



## Single-Molecule Sequencing: Methods, Training and Applications

### Lecture 1

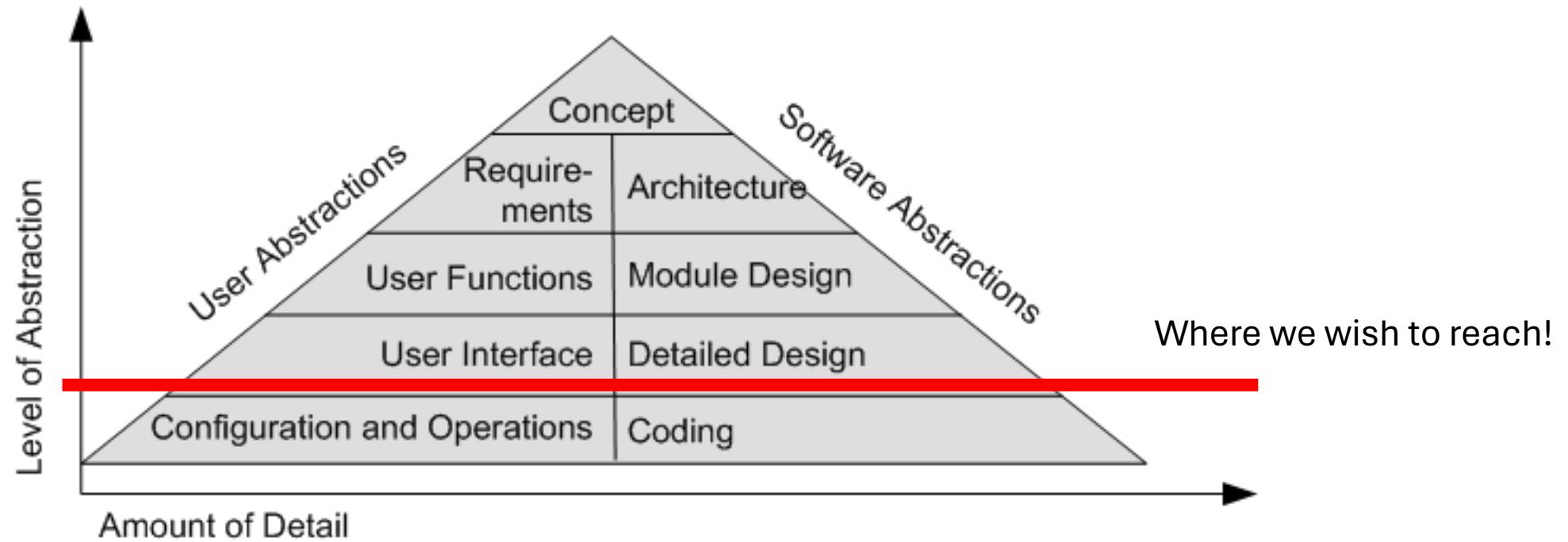
Theodore M. Nelson

Christopher E. Mason

Slides by Theodore M. Nelson

Weill Cornell Medicine

# Course Focus: Data, Programming, and Bioinformatic Analysis

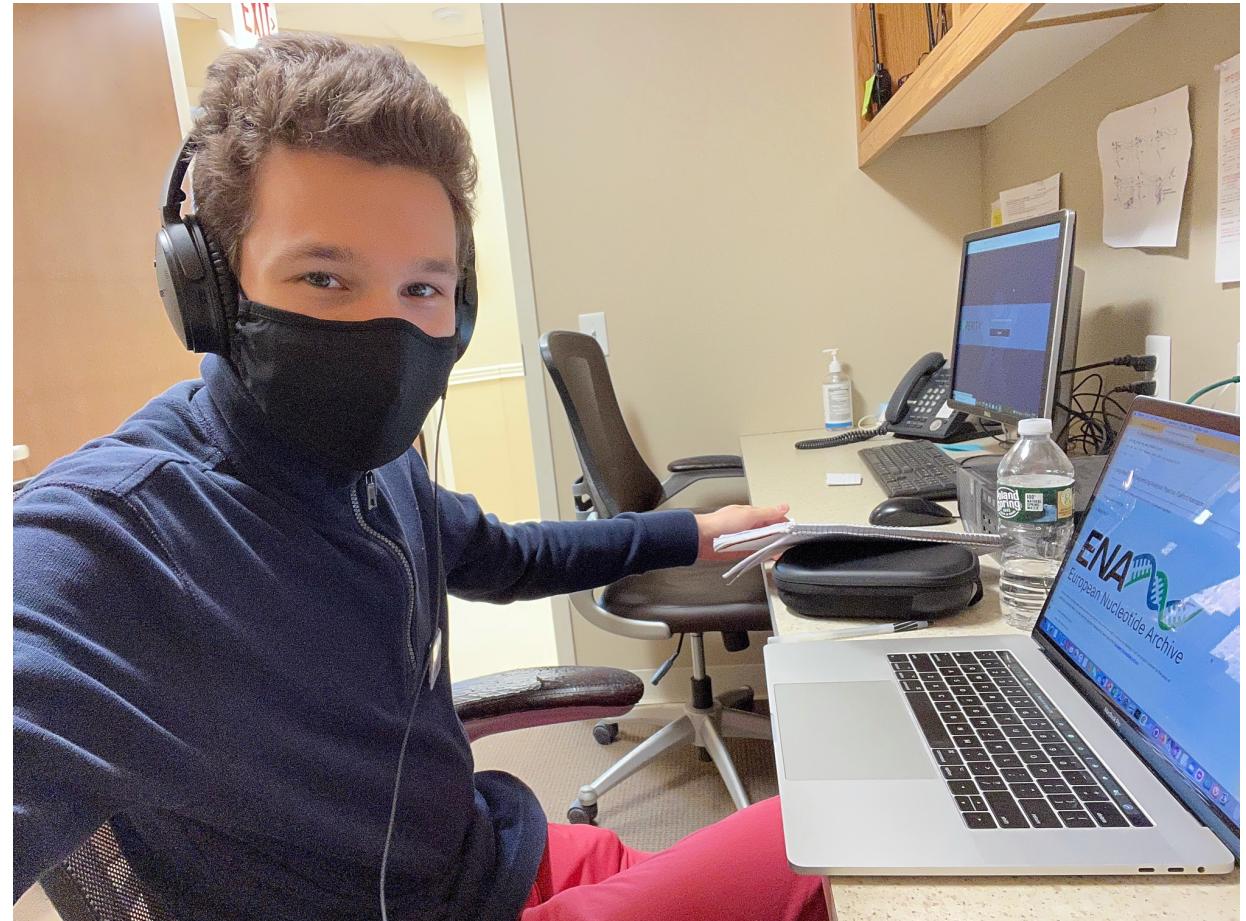


<https://pheugo.com/software/index.php?page=Guidebook.Design>

# Course Focus: five topics, two weeks each

Introduction Topics Course Computer	DNA What, Why Variations Extracting it!	RNA What, Why Variations Seq it!	Single-Cell & Proteomics What, Why Variations Research Talks	Analyze + Present Your Data! 
--	--	---	--	--

# Course has wet & dry lab components!



# ChatGPT can be a tool for in-context learning

 You  
generate a research superhero



Here's the "Cognitive Crusader," a superhero who embodies the essence of research, intellect, and the pursuit of truth. Surrounded by cutting-edge technology and data, this hero is ready to tackle any research challenge with precision and creativity.



 You  
E138: Can't write viminfo file /home/thn4005/.viminfo!  
Press ENTER or type command to continue

 ChatGPT  
The error message `E138: Can't write viminfo file /home/thn4005/.viminfo!` typically occurs in Vim (a text editor) due to issues with permissions or the `viminfo` file itself. The `viminfo` file is used by Vim to store information such as command history, buffers, global variables, etc., so that this information can persist between sessions. Here are some steps you can take to resolve the issue:

1. **Check File Ownership and Permissions:** The most common reason for this error is that Vim does not have the necessary permissions to write to the `viminfo` file. This could happen if the file is owned by another user or if the file permissions are set incorrectly.
  - Open a terminal.
  - Change to your home directory (if you're not already there) using `cd ~`.
  - Check the ownership and permissions of the `viminfo` file with `ls -l .viminfo`.
  - If the file is owned by another user or the permissions do not allow writing, you can change the ownership to your user (replace `thn4005` with your username) with `sudo chown thn4005:thn4005 .viminfo` and adjust the permissions to allow writing with `chmod u+rwx .viminfo`.

# Student Introductions

< 1.] name,  
2.] program,  
3.] decision to take the course,  
4.] a memorable magnet>

# Exploratory vs. Hypothesis-Driven Research



**Exploratory  
Research**



**Confirmatory  
Research**

**Wonky Stats**



**Sound Stats**

# **Cells** are a Single-Cell Discovery

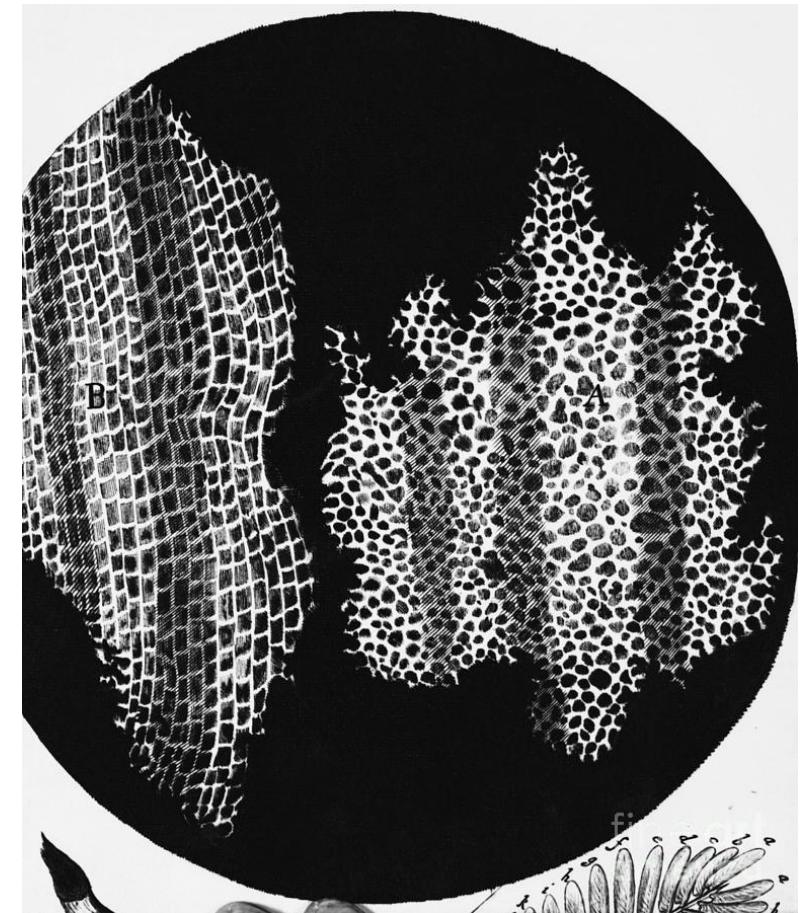
<https://www.independent.co.uk/2010/01/15/history-cell-part-i/>



Robert Hooke  
1665

Microscope  
Observations  
Of Single Cells

<https://pixels.com/featured/cork-from-micrographia-1665-biophoto-associates.html>

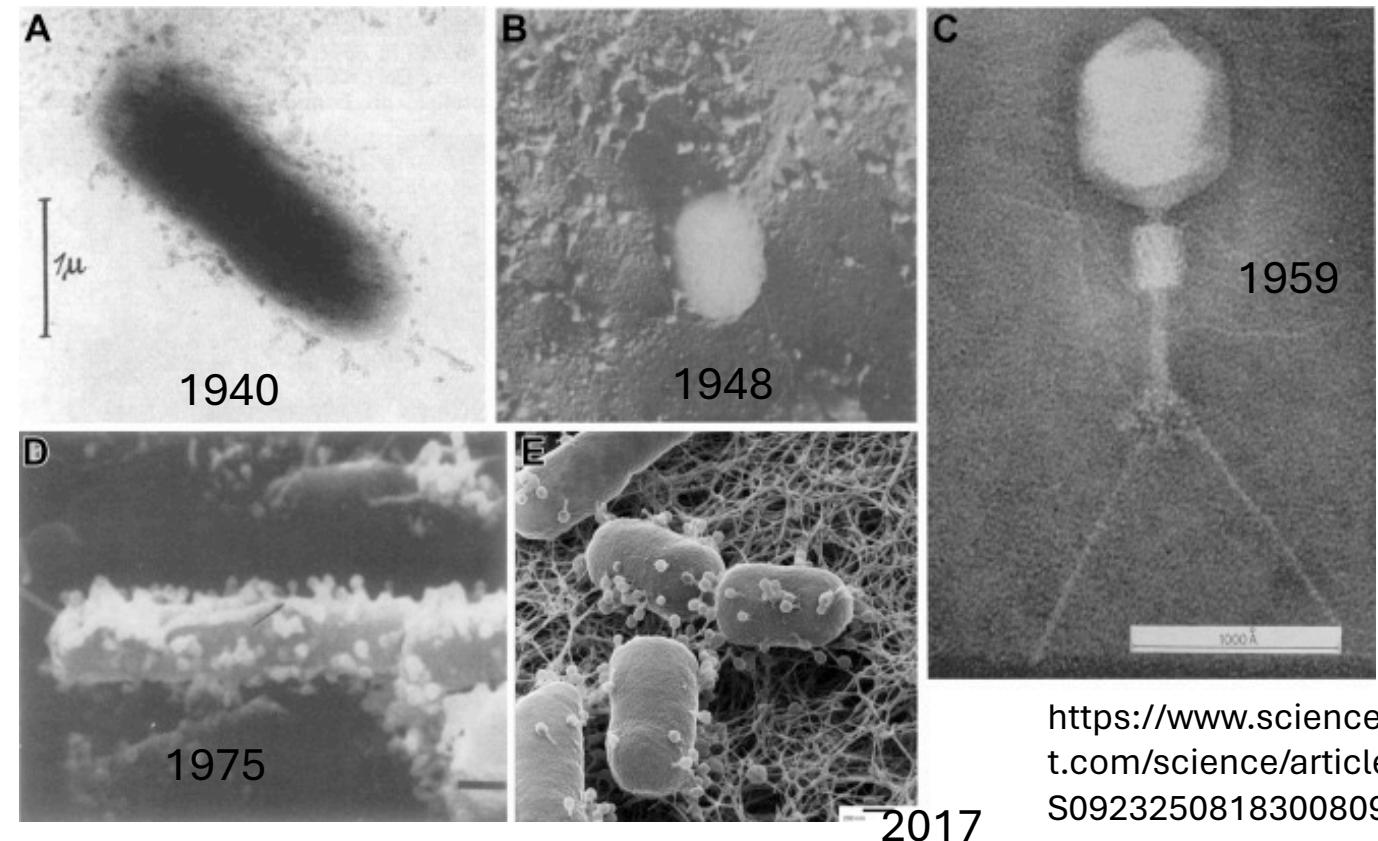


# **Viruses** are a Single-Particle Discovery



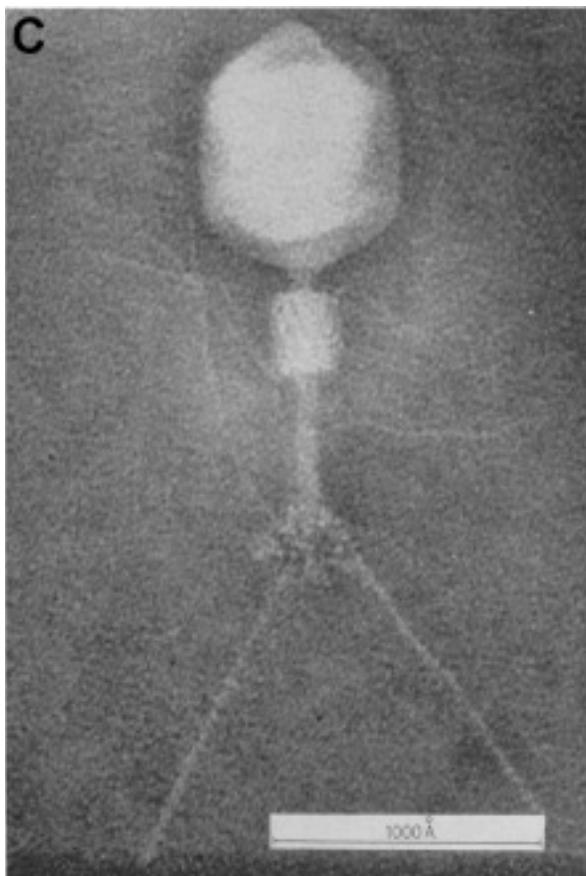
[http://wsm.wsu.edu/stories/2012/Spring/2012spring\\_1stelectron\\_popup.html](http://wsm.wsu.edu/stories/2012/Spring/2012spring_1stelectron_popup.html)

Electron  
Microscope  
Images  
  
Liquid  
or  
Particulate?



<https://www.sciencedirect.com/science/article/pii/S0923250818300809>

# Single-Cell & Single-Virus Questions



## Basic Questions:

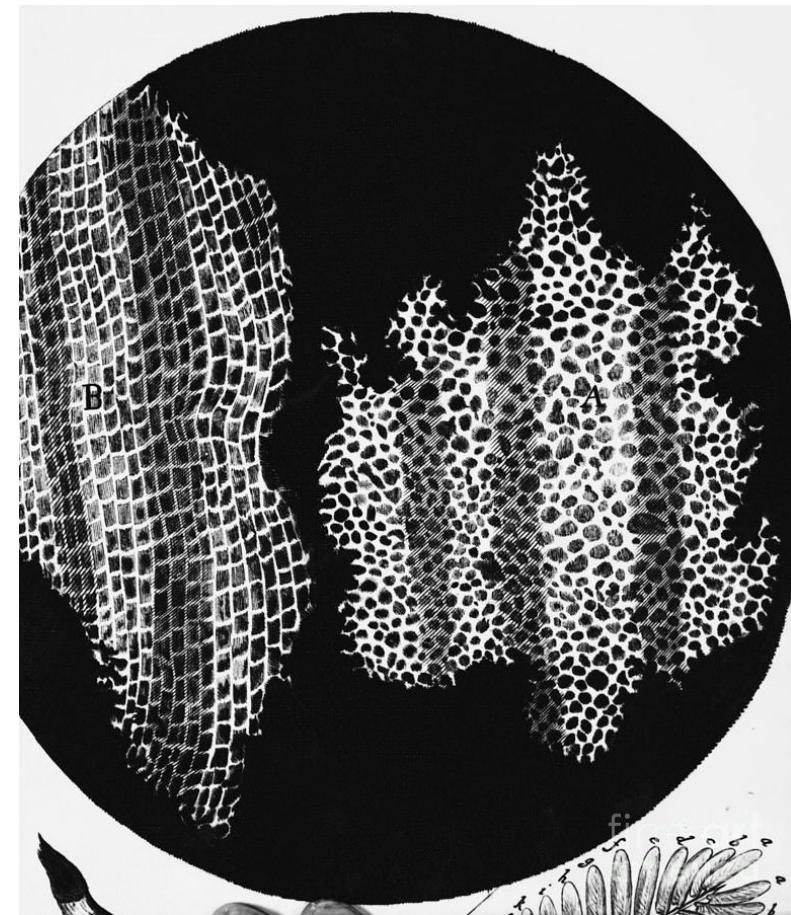
What is inside a virus / cell?

How do viruses / cells determine their structure?

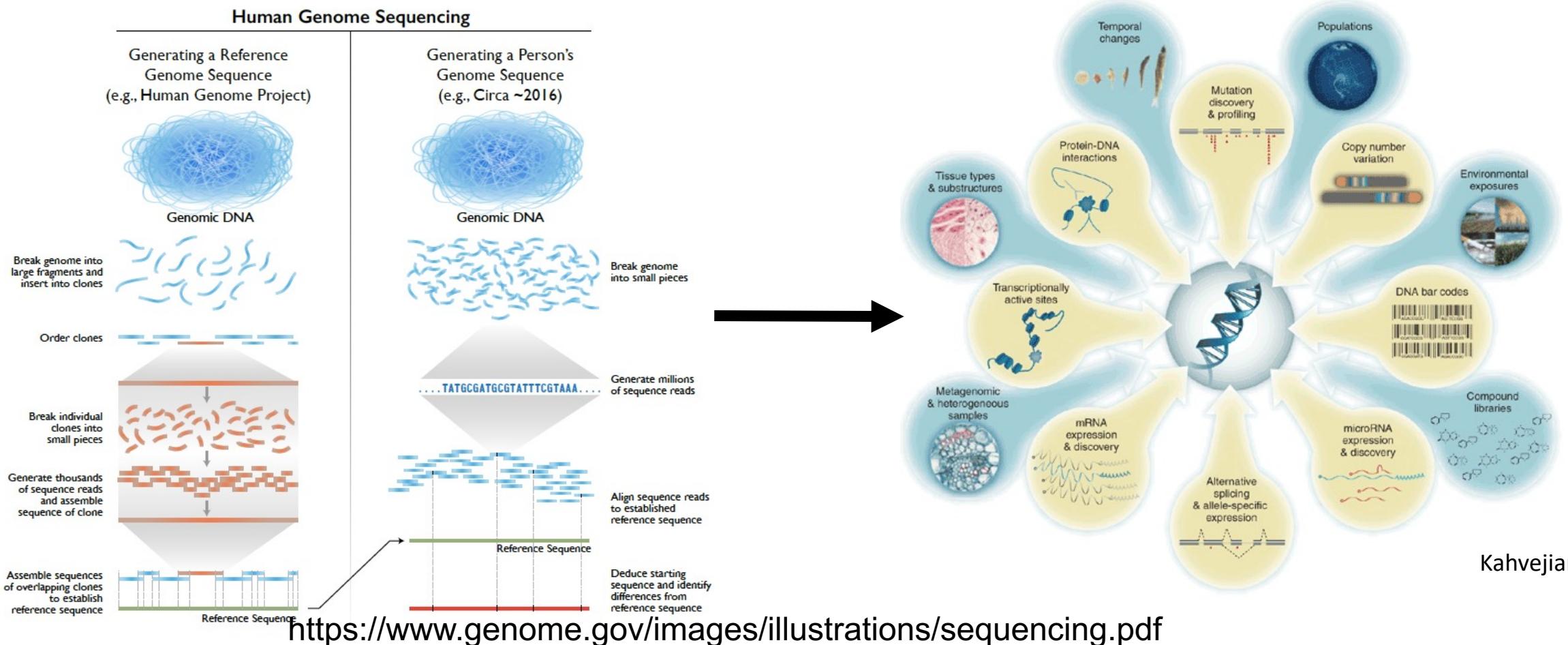
How do viruses / cells interact signal information between each other?

What differences are observed between different cell / virus particles?

What differences are observed between different cell / virus types based on origin?



# Dominant Questions in Sequencing Research



# Scaling Up to Deliver Novel Results



About Us ▾ 100,000 Genomes Project ▾ Taking Part ▾ For Healthcare Professionals ▾ Research ▾ Industry Partnerships ▾ News & E



## Genomics England is delivering the **100,000 Genomes Project**.

We are creating a new genomic medicine service with the NHS – to support **better diagnosis and better treatments** for patients. We are also enabling medical research.

[More information about the 100,000 Genomes Project](#)

News story

### **Genome sequencing project reaches the halfway mark**

50,000 human genomes have now been sequenced from patients with cancer or rare diseases, under the 100,000 Genomes Project.

Published 28 February 2018

# Scaling Up to Deliver Novel Results with Great Effect

## NHS to trial blood test to detect more than 50 forms of cancer

Researchers hopes Galleri trial will be a 'gamechanger' for early diagnosis and save many lives



▲ The Galleri blood test will be offered to 165,000 people in England from mid-2021, the vast majority of whom have no signs of the disease. Photograph: Jacqueline Larma/AP

Offered to 165,000 people in England from mid-2021 onward; no signs of disease.

Followed through 2023;  
If successful, move on to test 1M people in 2024-2025.

<https://www.theguardian.com/science/2020/nov/27/nhs-to-trial-blood-test-to-detect-more-than-50-forms-of-cancer>

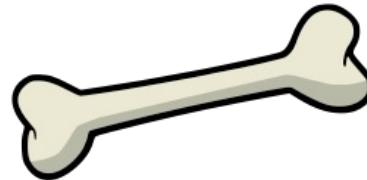
# Uncovering the Function of Specific Genes

Myostatin (MSTN) homozygous nulls (-/-) give lean and large muscles

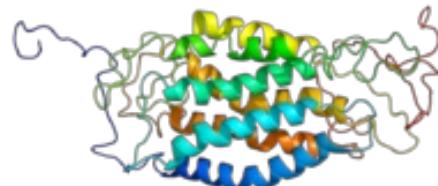


<http://thevoiceofnetizen.blogspot.com>

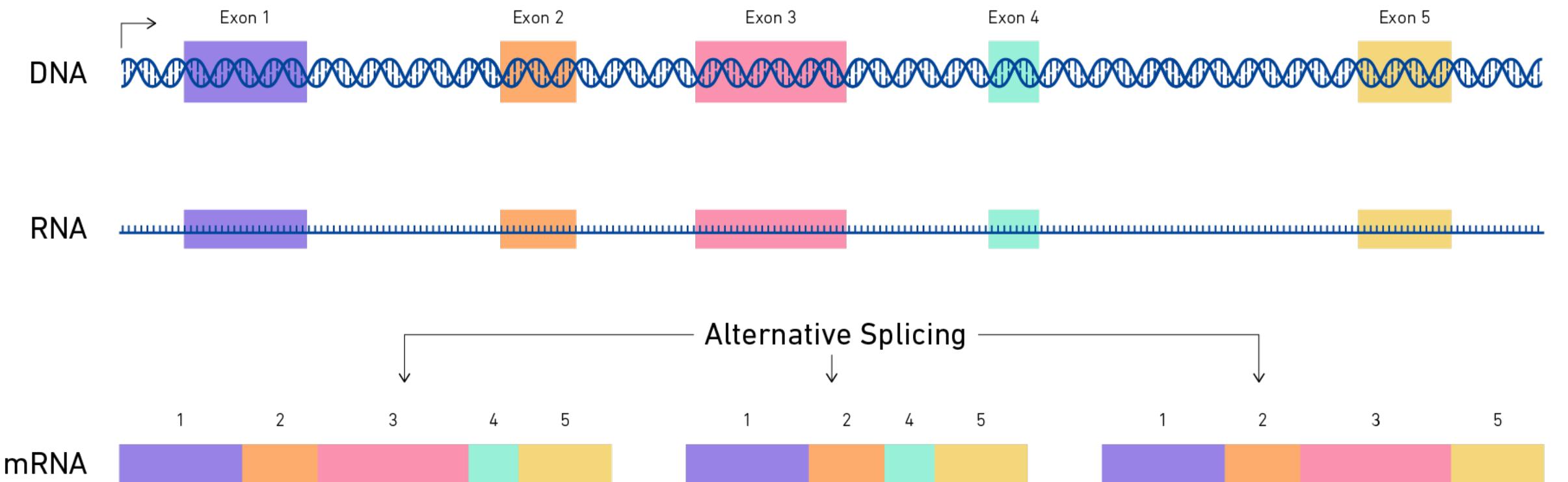
Low density lipoprotein receptor 5 (LRP5) heterozygotes (+/-) can have strong bones



C-C chemokine receptor type 5 (CCR5) homozygous nulls (-/-) have HIV protection

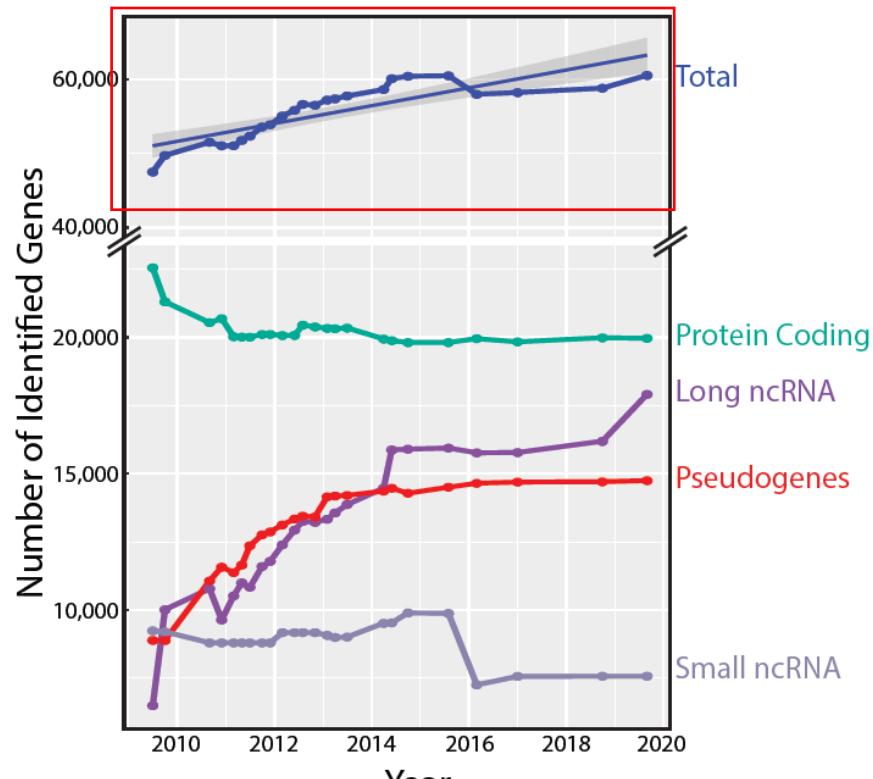


# Genes as Abstractions of Transcripts



<https://www.technologynetworks.com/genomics/articles/alternative-splicing-importance-and-definition-351813>

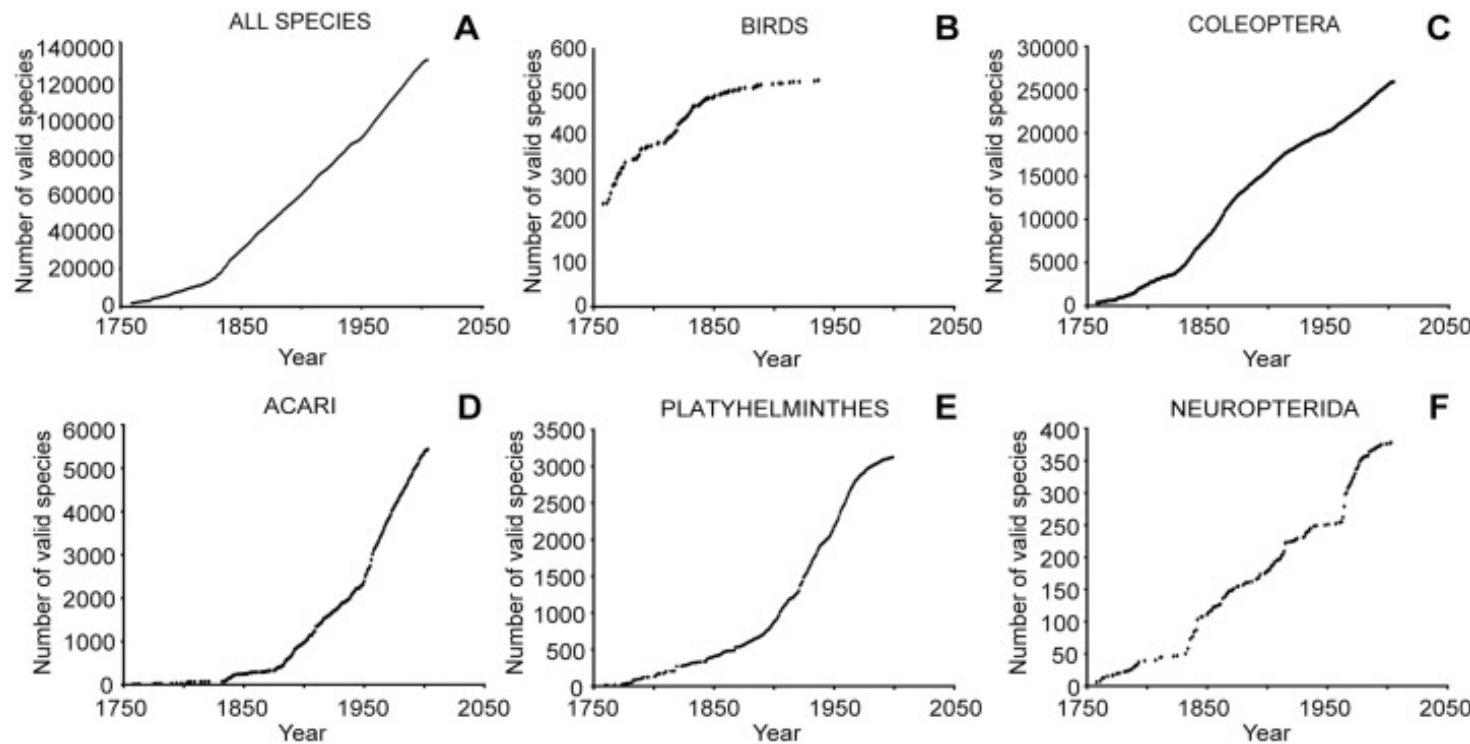
# *New Human Genes* are still being found



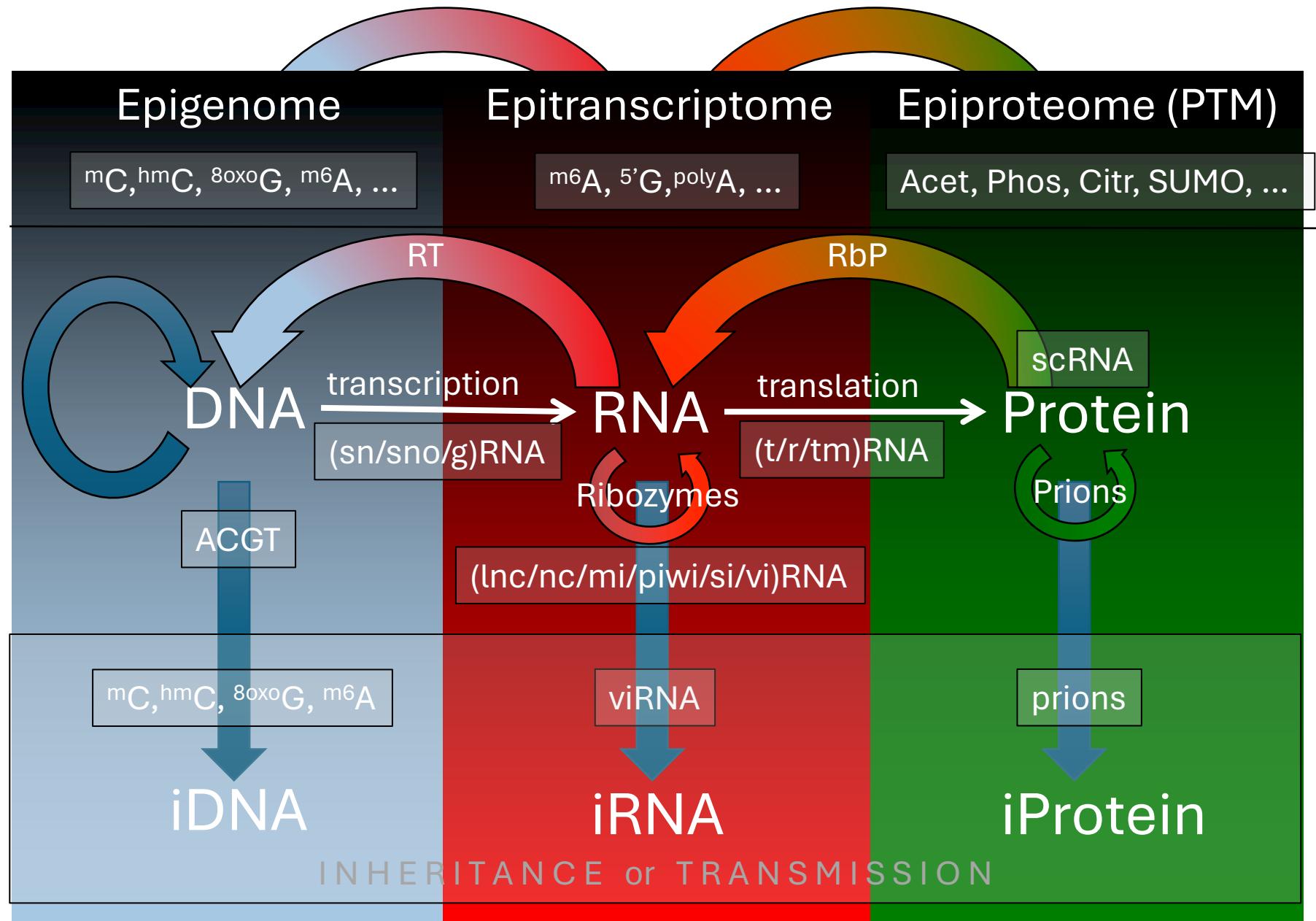
<https://mitpress.mit.edu/books/next-500-years>

<https://www.gencodegenes.org/>

# **New species** are being found & many more to come

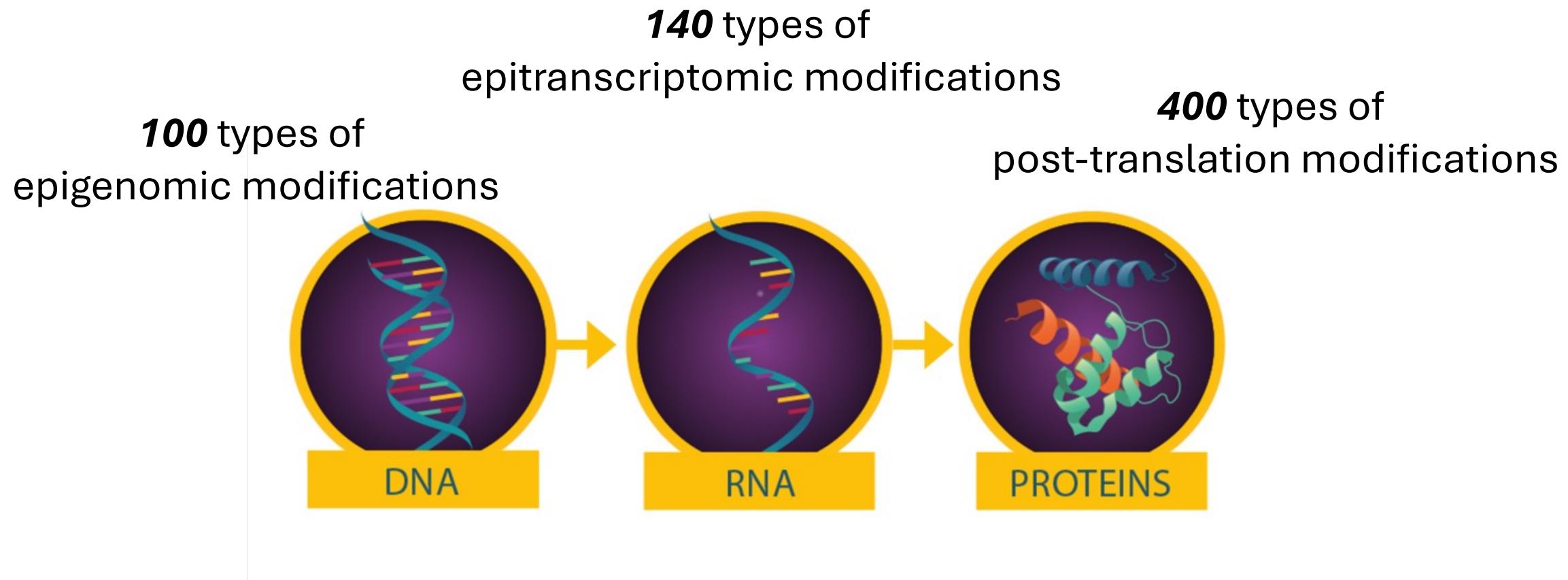


<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0036881>



from Salelore et  
al., Genome  
Biology, 2012

# **5,600,000** Permutations of a Biological Transcript



# What Questions Could You Ask of Single Molecules?



# What Questions Could You Ask of Single Molecules?

Sanne Boersma, Huib H. Rabouw,  
Lucas J.M. Bruurs, Tonja Pavlović,  
Arno L.W. van Vliet, Joep Beumer,  
Hans Clevers, Frank J.M. van  
Kuppeveld, Marvin E. Tanenbaum,

Translation and Replication  
Dynamics of Single RNA Viruses,

Cell,

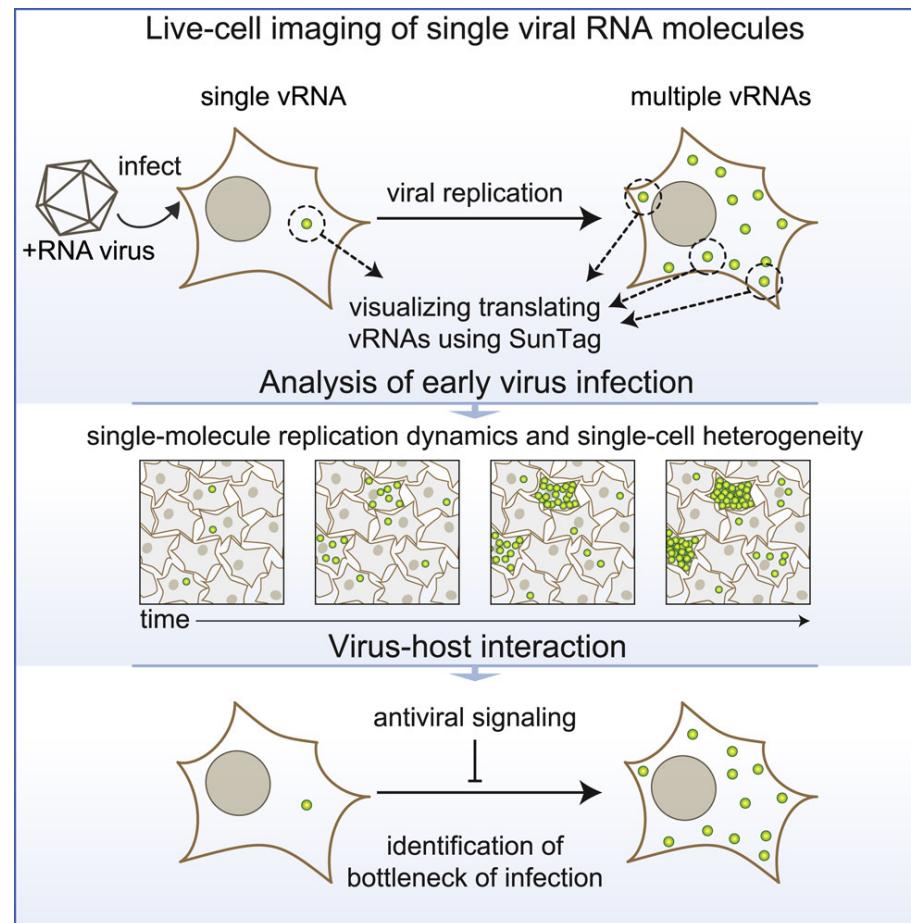
Volume 183, Issue 7,

2020,

Pages 1930-1945.e23,

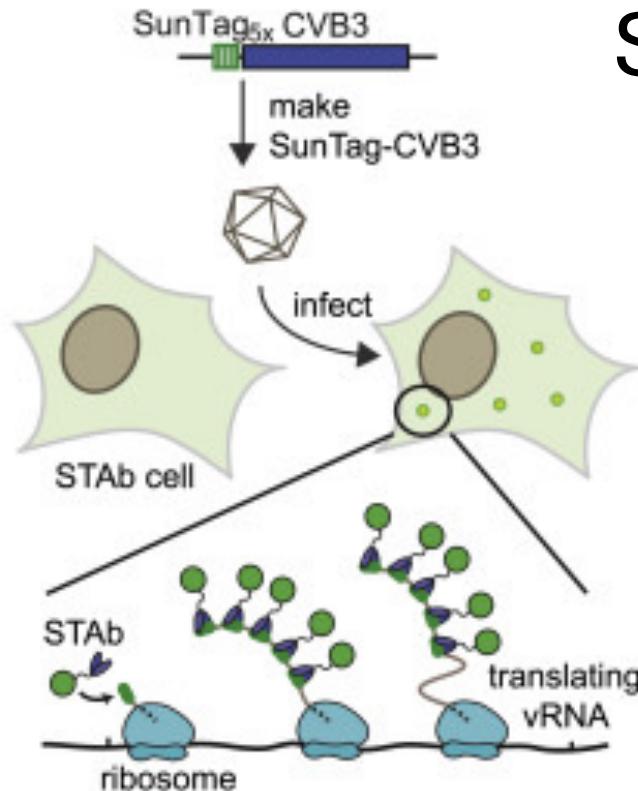
ISSN 0092-8674,

<https://doi.org/10.1016/j.cell.2020.1.019>.



00:00 (hh:mm)

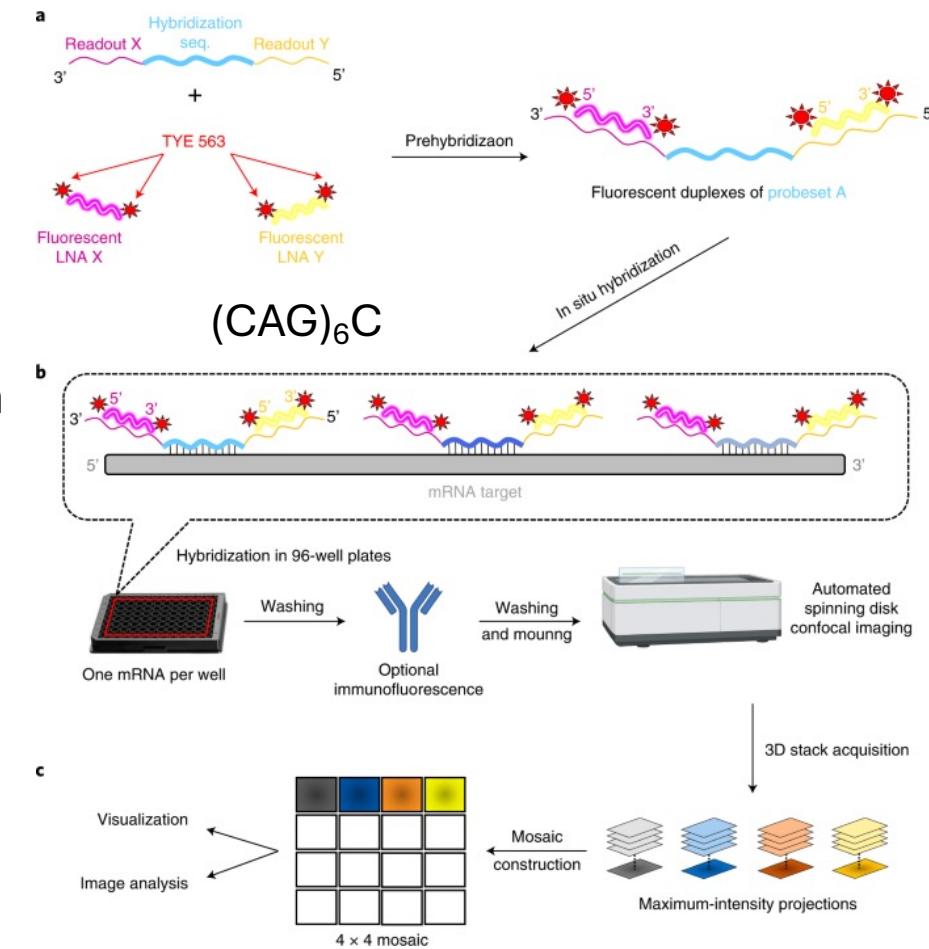
# Different Approaches to Single-Molecule RNA Visualization



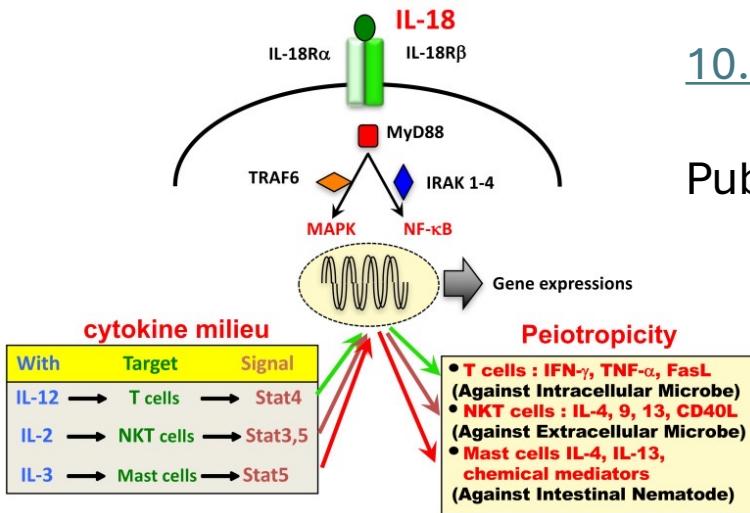
[https://doi.org/10.1016/j.cell.2020.10.019.](https://doi.org/10.1016/j.cell.2020.10.019)

## SunTag vs. smFISH

Safieddine, A., Coleno, E., Lionneton, F. et al. HT-smFISH: a cost-effective and flexible workflow for high-throughput single-molecule RNA imaging. *Nat Protoc* **18**, 157–187 (2023). <https://doi.org/10.1038/s41596-022-00750-2>



# Recontextualizing the Role of IL-18 with smFISH

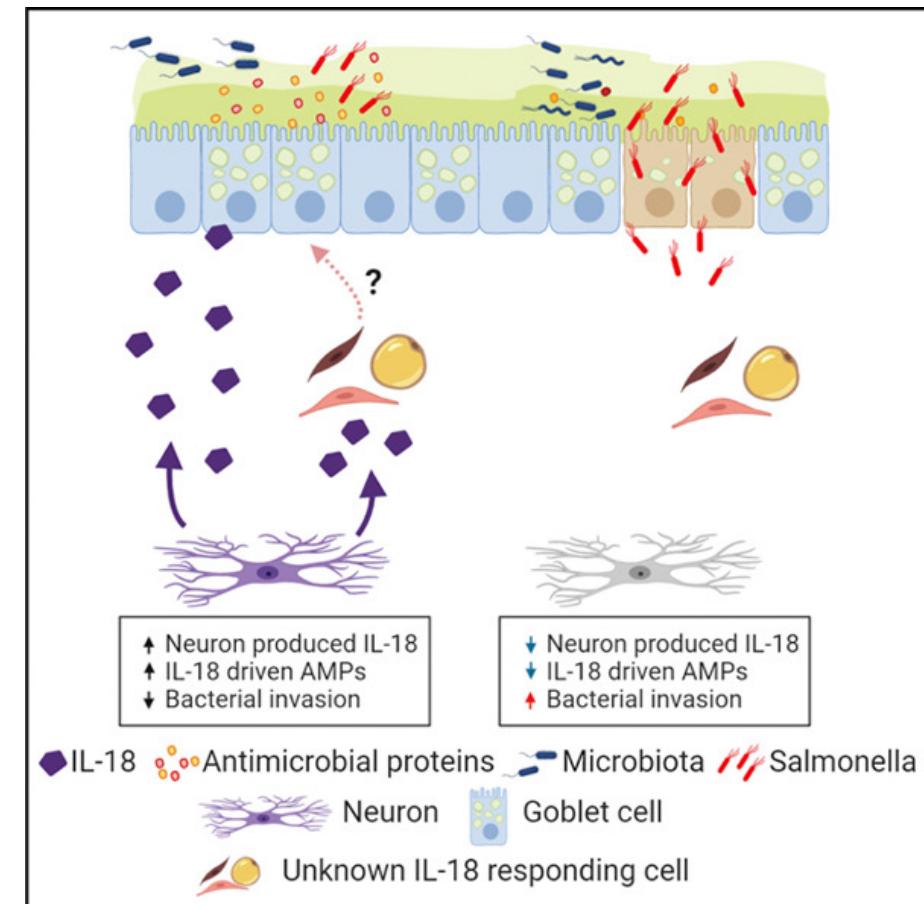


[10.3390/ijms20030649](https://doi.org/10.3390/ijms20030649)

Published: 2019

"In this study, we discovered that IL-18 produced by enteric neurons is necessary for protection against invasive bacterial infection. Unlike immune cell- and epithelial cell-derived sources, neuronal expression of IL-18 non-redundantly directs a program of AMP production in goblet cells, which, during homeostasis, enforces a sterile inner-mucus barrier and, during infection, facilitates killing of enteric pathogens."

[10.1016/j.cell.2019.12.016](https://doi.org/10.1016/j.cell.2019.12.016) | Published: 2020



# Class Summary

- 1.] Next-generation sequencing paved the way for the sequencing revolution, and allowed for a greater breadth of population-scale questions.
- 2.] Single-molecule sequencing has the potential to change the kinds of biological questions one can ask.
- 3.] The discovery of cells and viruses similarly generated their own biological fields.
- 4.] SunTag and smFISH are specific approaches to visualizing known RNA sequences, providing avenues for fluorescence-based confirmatory research.

# Thank you! Looking forward to a great semester.



Photo Credit:  
Claudio Solano

# Exit Ticket

