

High-throughput DNA sequencing

Robert Lyle

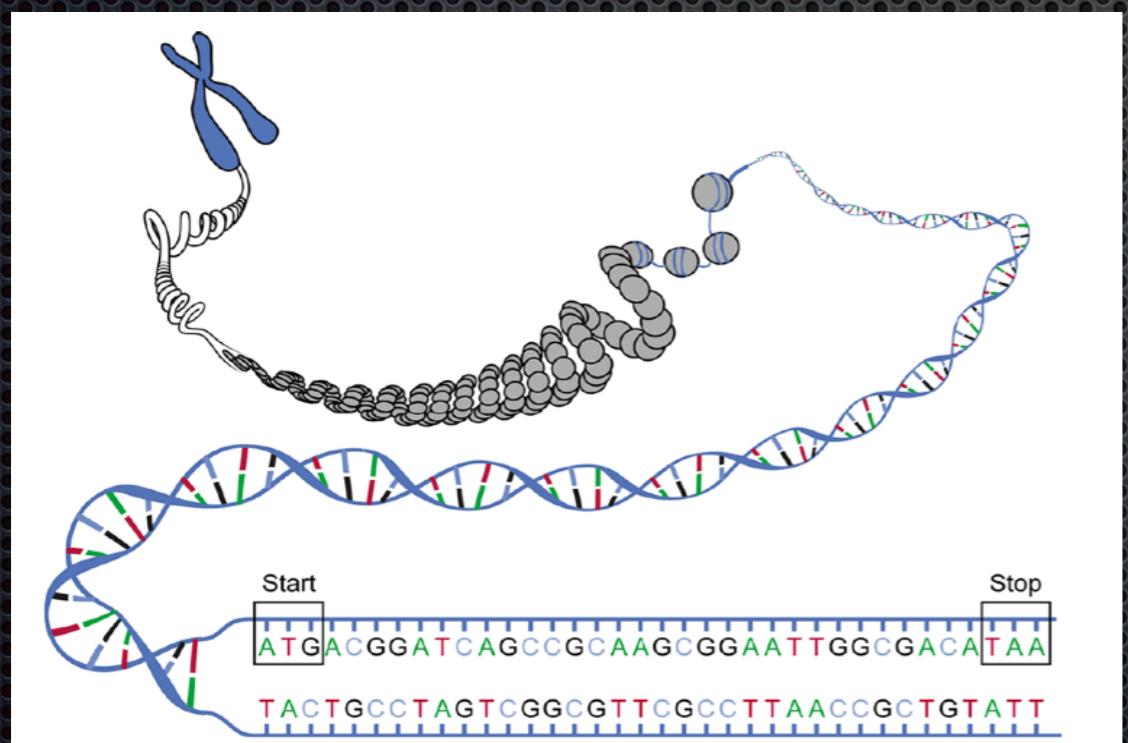
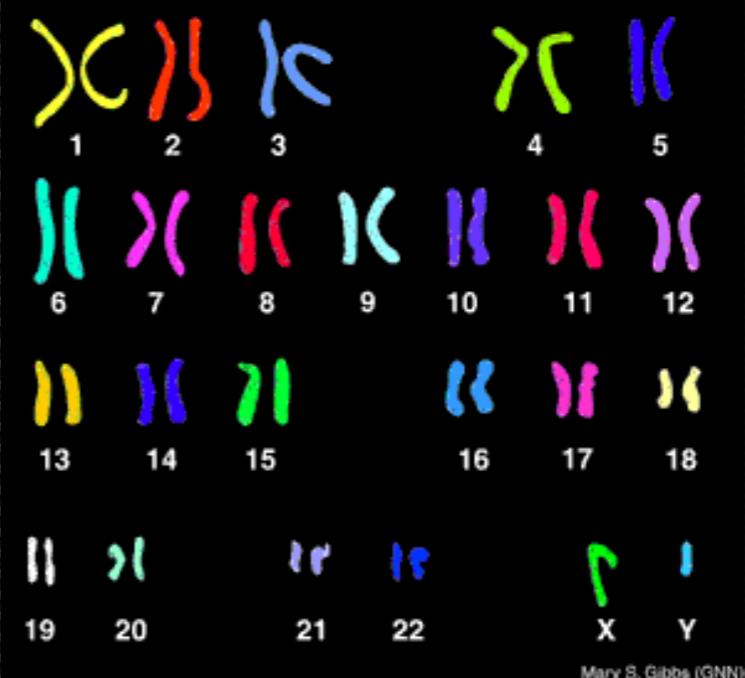
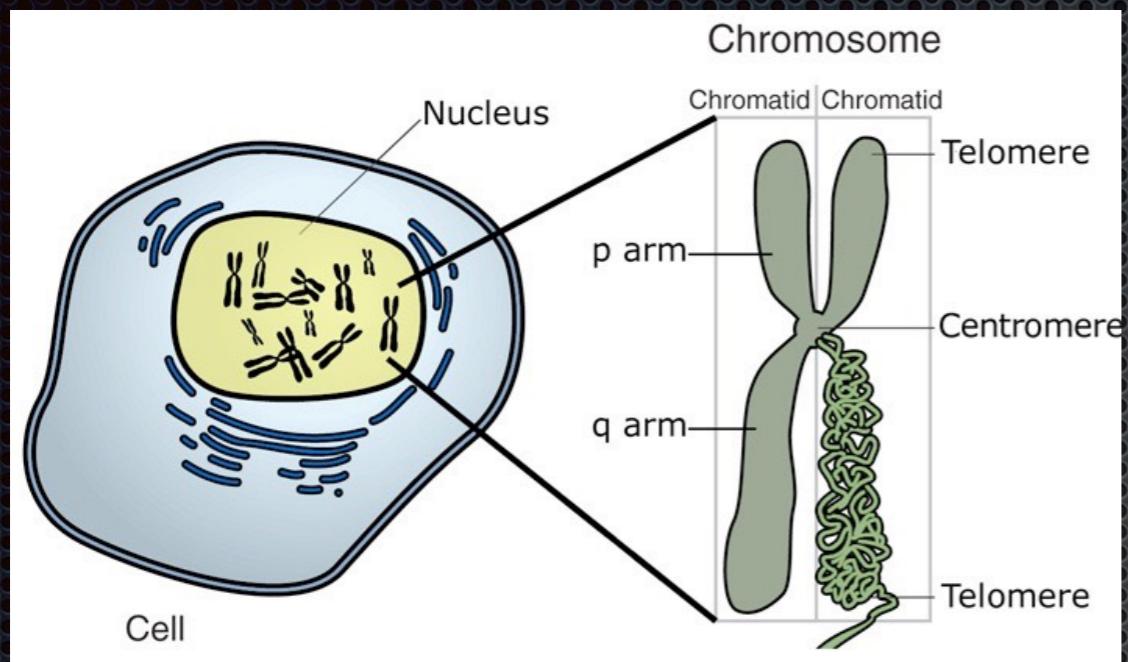
Department of Medical Genetics
Oslo University Hospital

Centre for Fertility and Health
Norwegian Institute of Public Health

Robert.Lyle@medisin.uio.no

DNA sequencing

DNA



- 4 bases - A, G , C, T G A T C
 C T A G
- Human **genome** ~3 billion bases
- How can we read the sequence of bases?

How many bases?

	bp	1
kilo	kb	1,000
mega	Mb	1,000,000
giga	Gb	1,000,000,000
tera	Tb	1,000,000,000,000

Human genome: 3 000 000 000 (3 Gb)

DNA replication

CGATGCTGTTGCATGATGCTAGTCGATGCTGTTGGGTTGATCGTGGGCTAGCTAGC
ACCTAACTAGCAATAGAATCACATGACAGCATGACACATGCAACAGCATCG



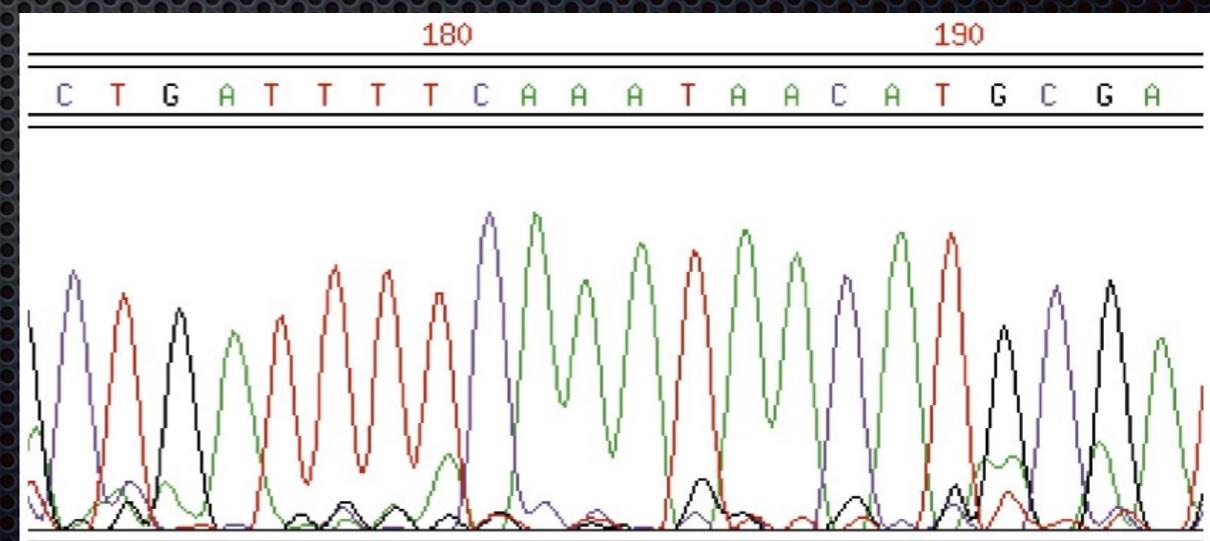
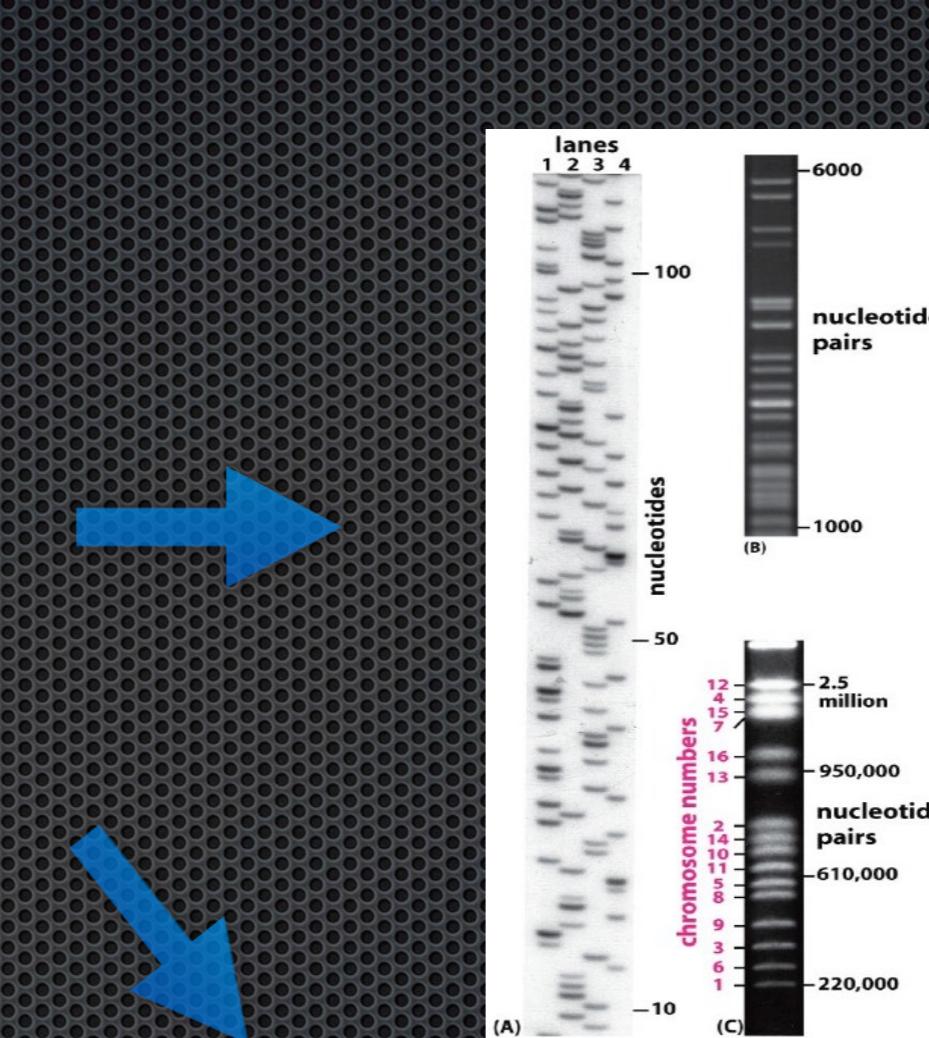
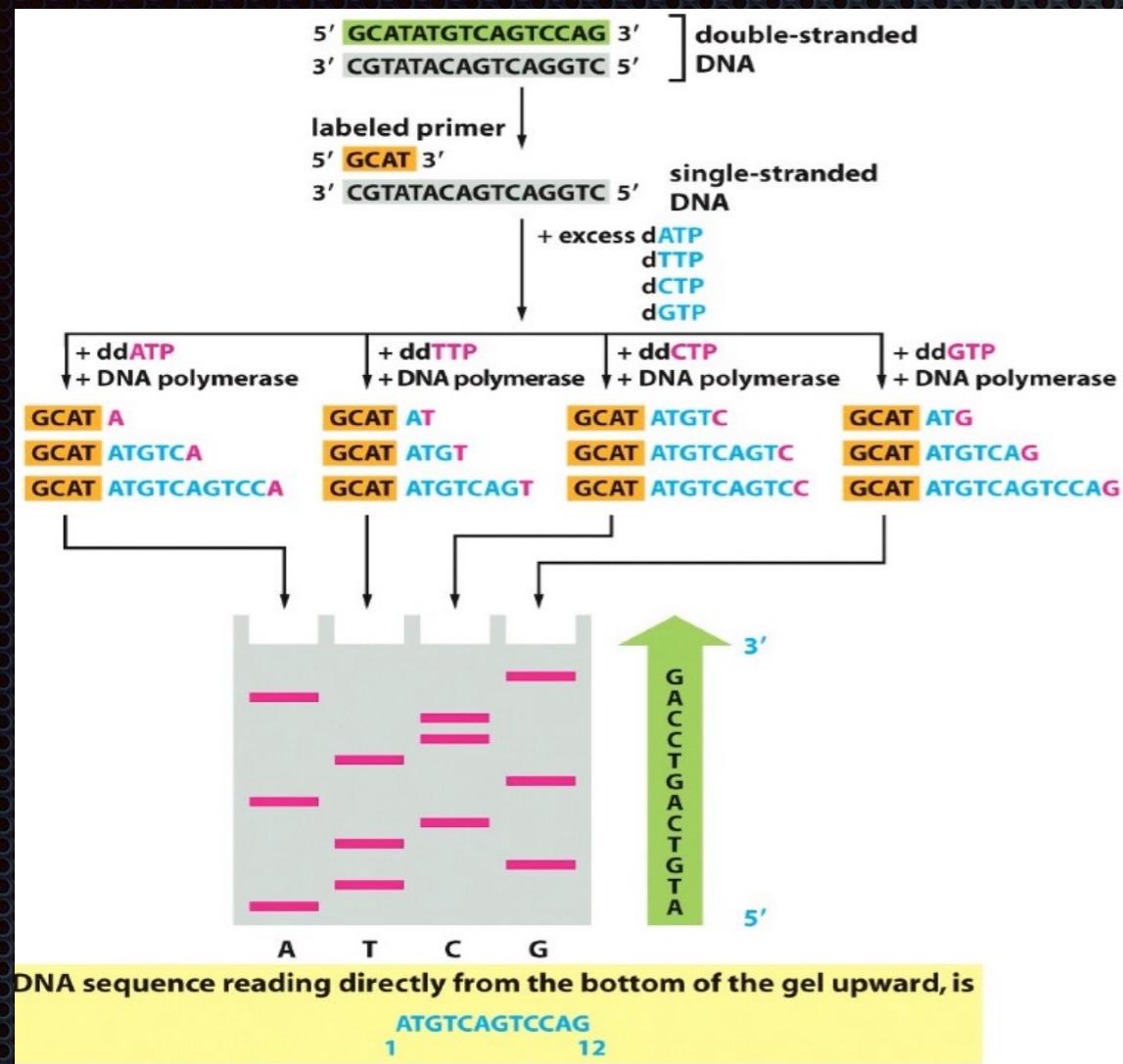
DNA polymerase

CGATGCT GTTGCATGATGCTAGTCGATGCTGTTG



- Denature DNA
- Prime with short DNA (oligonucleotide)
- Polymerase extends base by base

DNA sequencing

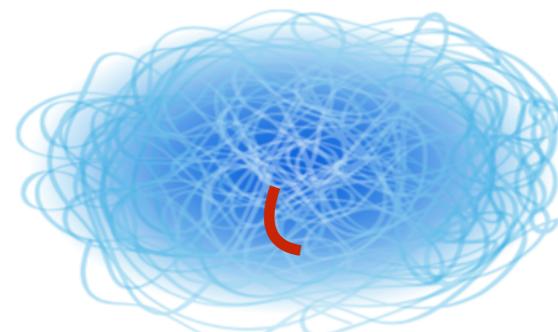


- Sanger sequencing
- Detect nucleotide extension with radioactivity or fluorescence
- Accurate but slow

High-throughput sequencing

Sanger and HTS

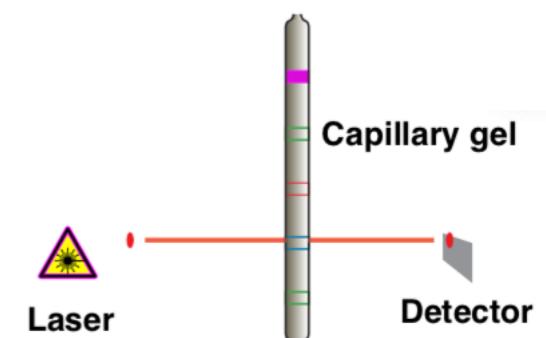
Sanger



Genomic DNA



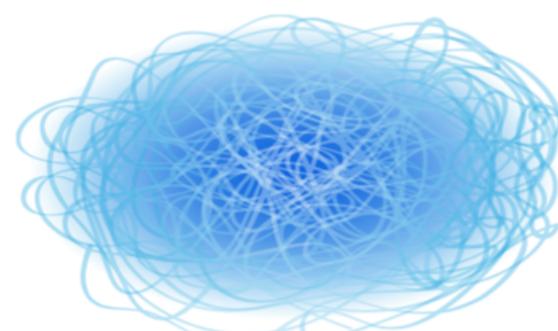
Locus of interest



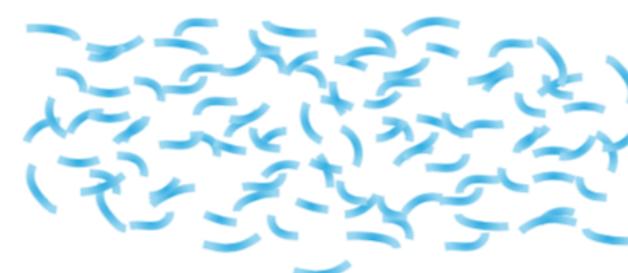
16, 48, 96 reads

- + low error rate
- very low throughput

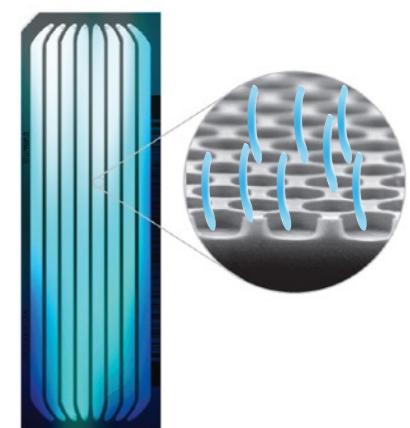
HTS



Genomic DNA



Flow cell

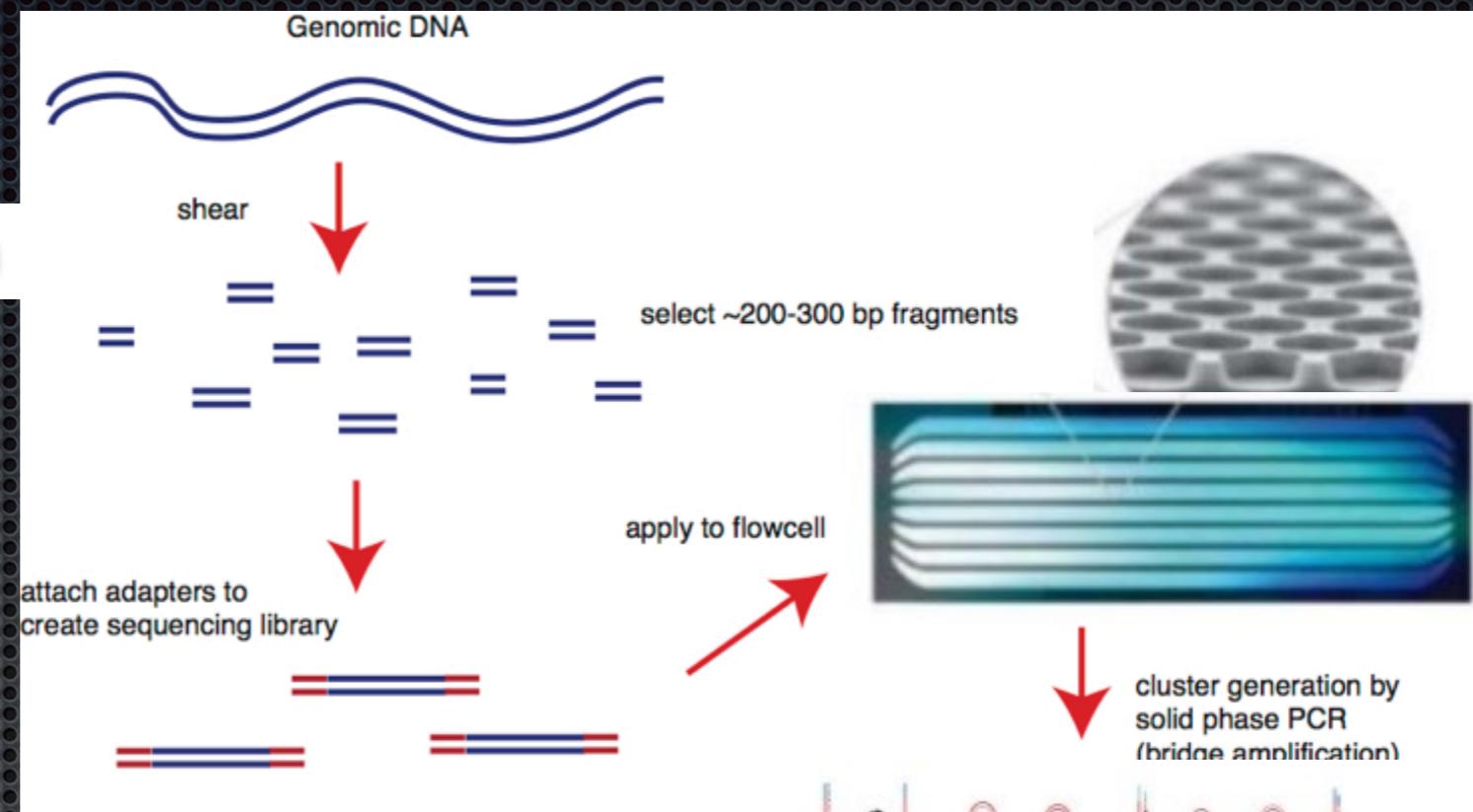


20×10^9 reads

- + very high throughput
- high error rate

Illumina sequencing technology

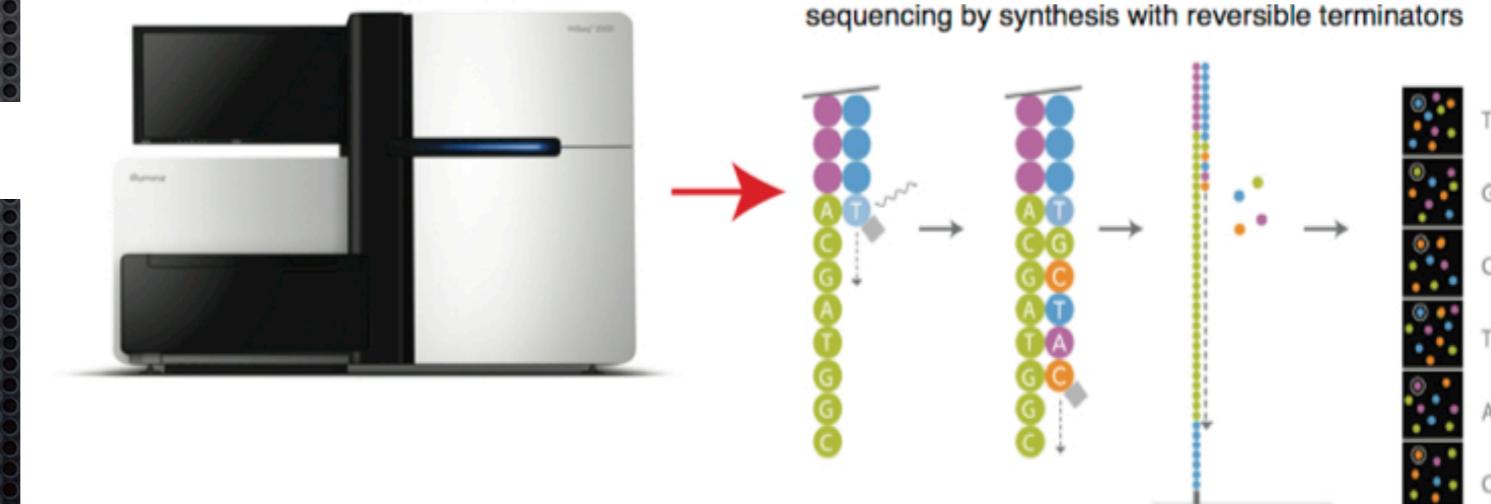
1. Library preparation



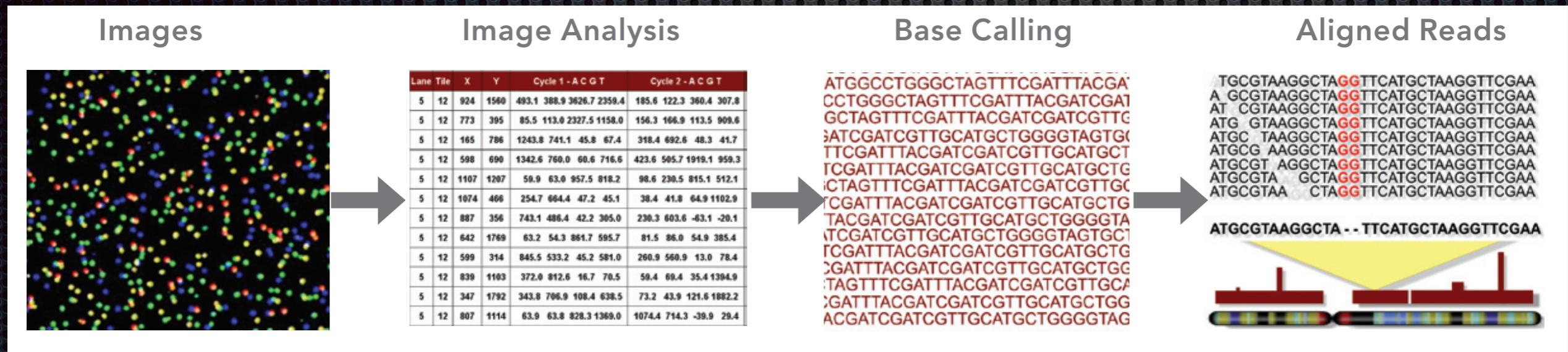
2. Cluster generation



3. Sequencing



Analysis pipeline



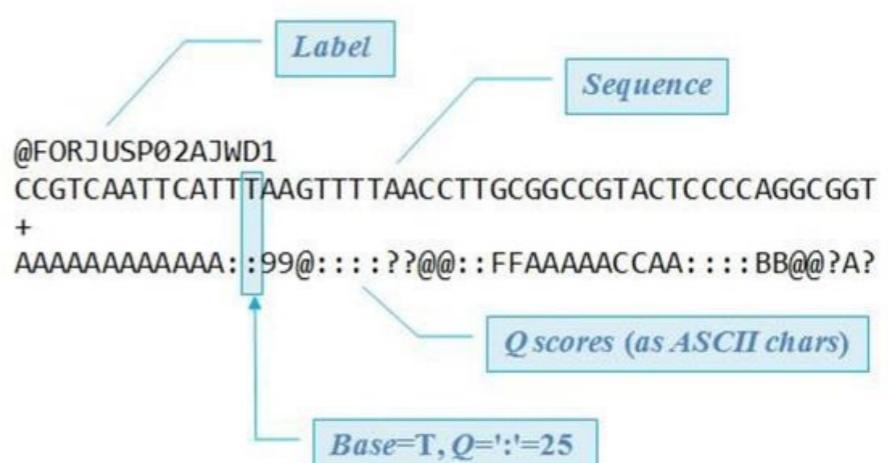
- Short fragments: 50...300 bp
- <20 billion DNA sequence reads**
- Single-end reads
- Paired-end reads

```

@HWI-M01141:63:A4NDL:1:1101:14849:1418 1:N:0:TATAGCGAGACACCGT
NACGAAGGGTGCAGCGTTACTCGGAATTACTGGCGTAAAGCGTGCAGCGCAGCGGGTTGTTA
+
#>>>A??AFAA1BGGEGGAAFGCA0BFF1D2BCF/EEG/DBEE/E?GAEEFAEAEFG1
@HWI-M01141:63:A4NDL:1:1101:13802:1421 1:N:0:TATAGCGAGACACCGT
NACGGAGGGTGCAGCGTTAACCGGAATTACTGGCGTAAAGCGCAGCGCAGCGGGTTGTTA
+
#>>AAABBBABBGGGGGGGG?FHGGGGGGHHHHHHHHGGGGH
@HWI-M01141:63:A4NDL:1:1101:15928:1426 1:I
NACGTAGGGTGCAGCGTTAACCGGAATTACTGGCGTAAAGCGCAGCGCAGCGGGTTGTTA
+
#>>AABFB@FBBGGGGGGGGGGHGGGGFHHHHHHHHGGGGH
@HWI-M01141:63:A4NDL:1:1101:14861:1431 1:I
NACGAAGGGTGCAGCGTTACTCGGAATTACTGGCGTAAAGCGCAGCGCAGCGGGTTGTTA
+
#>>AAAABBFABGGGGGECEHGGEFFHHHHHHGGGGH
@HWI-M01141:63:A4NDL:1:1101:15264:1465 1:I
NACGTAGGGTGCAGCGTTGTCCCGAACCGGAATTACTGGCGTAAAGCGCAGCGCAGCGGGTTGTTA
+

```

FASTQ format



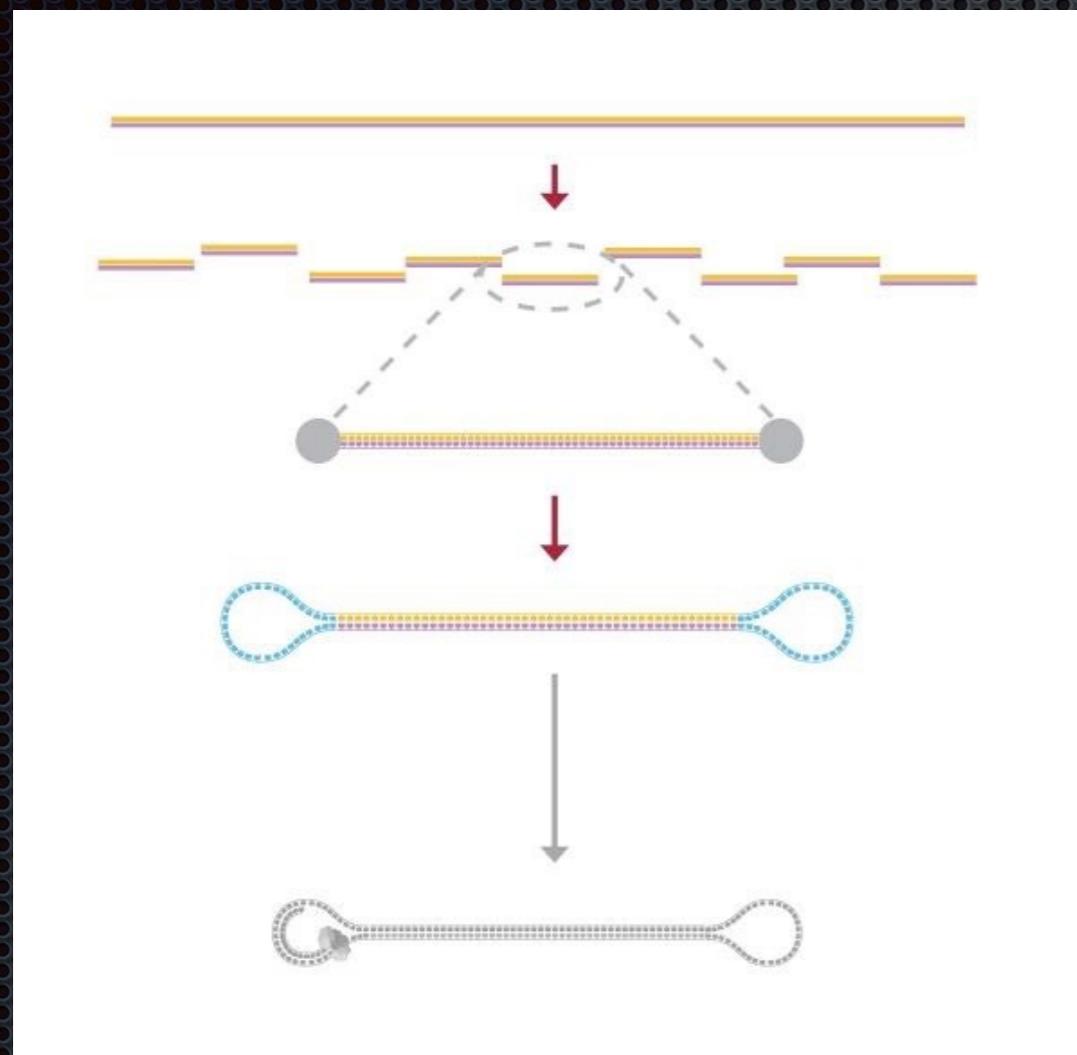
Pacific BioSciences

- Single Molecule Real Time
- Long Readlength
- Speed

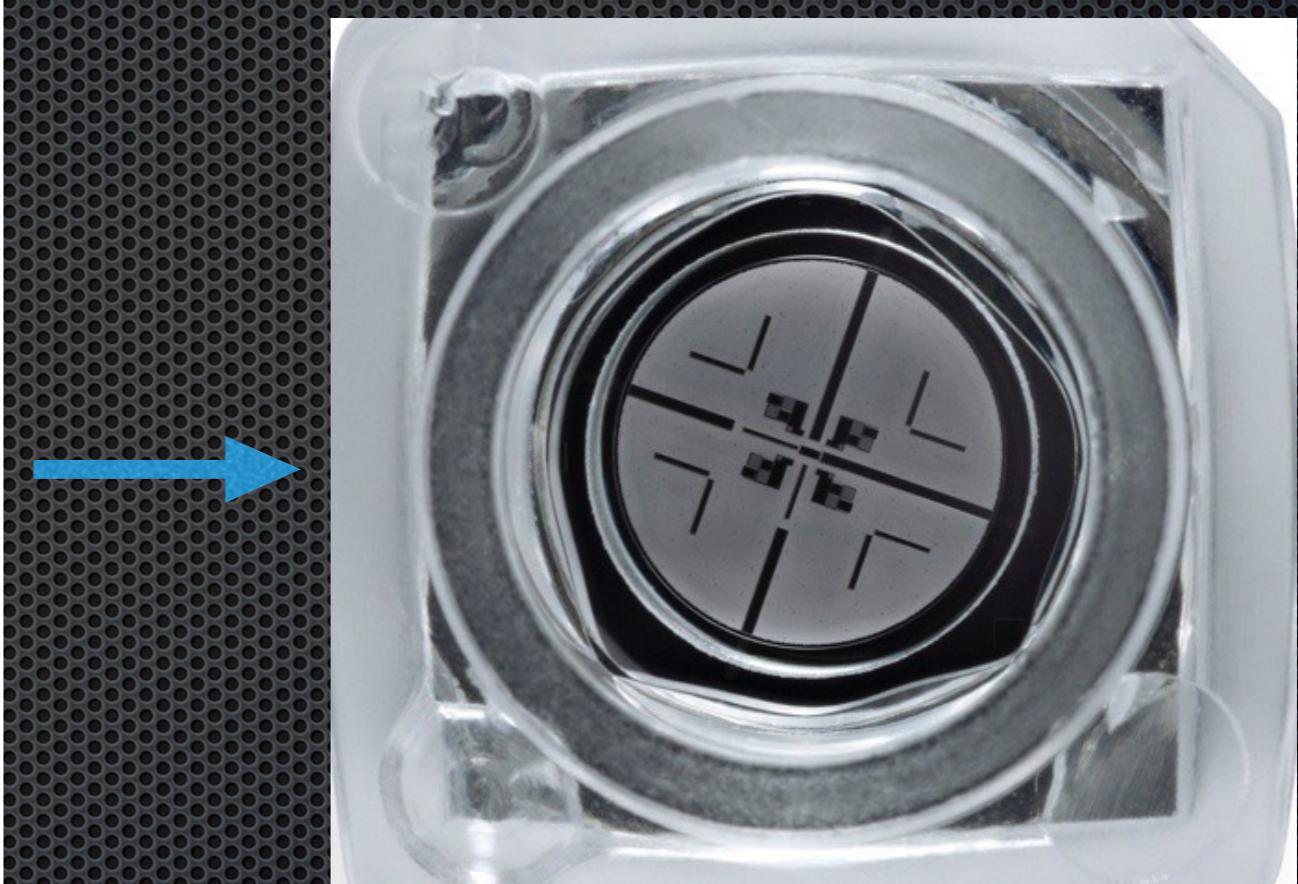


<http://www.pacificbiosciences.com>

PacBio sequencing I

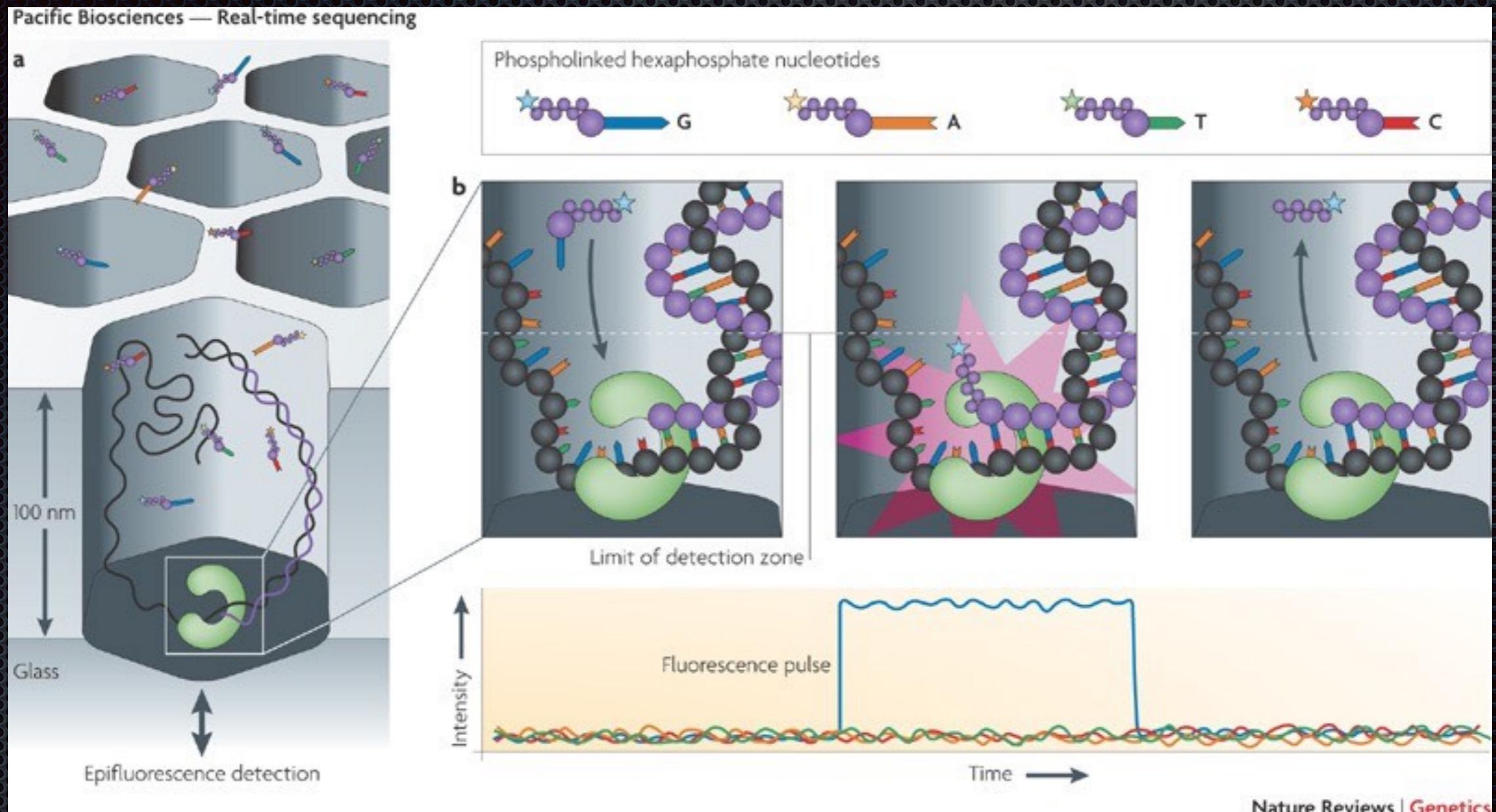


SMRT bell



SMRT cell

PacBio sequencing II

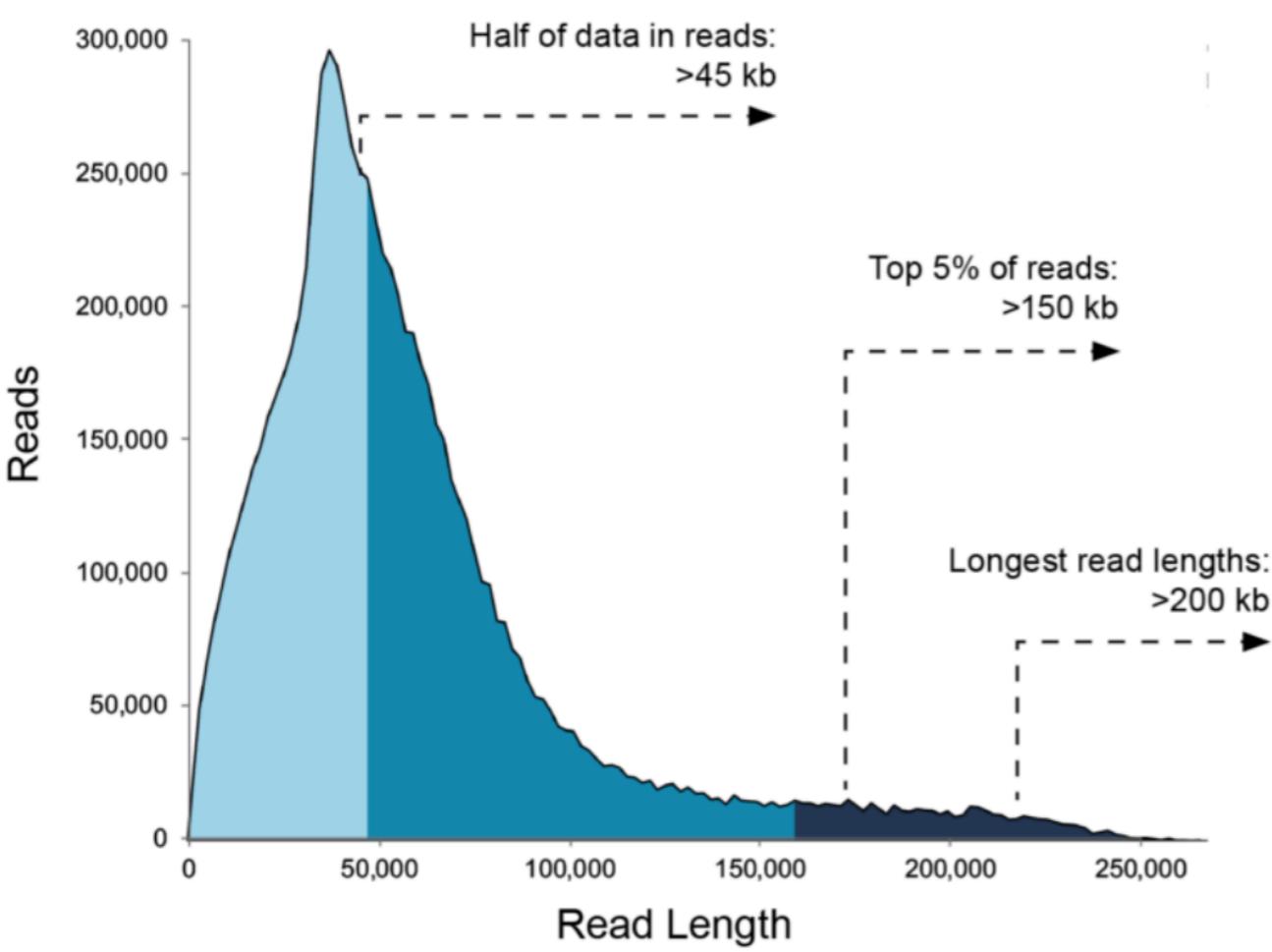


- Detection of base incorporation in single molecule

PacBio data

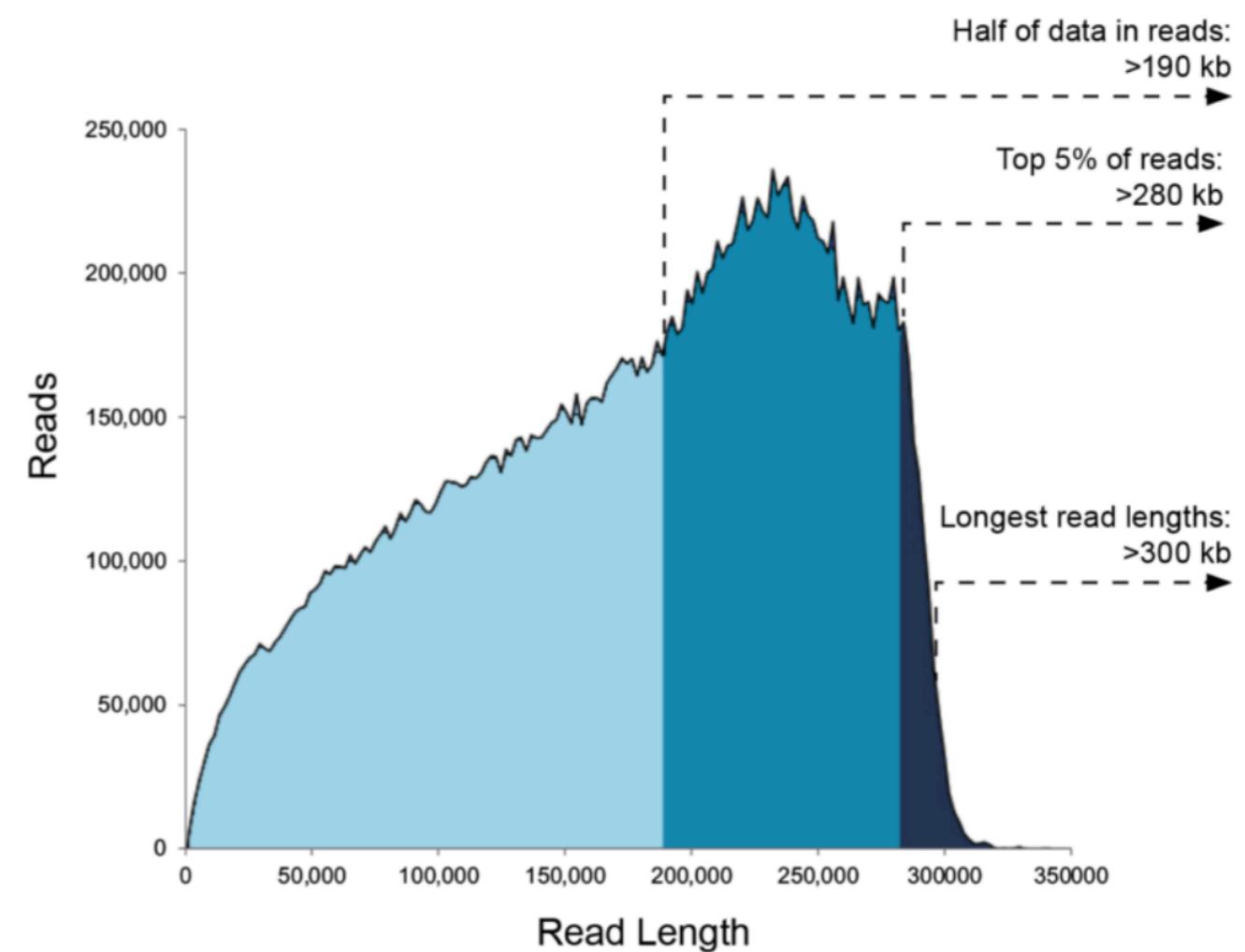
Long Read Lengths (Libraries >20 kb)

Half of data in reads: >45 kb
Data per SMRT Cell: Up to 20 Gb



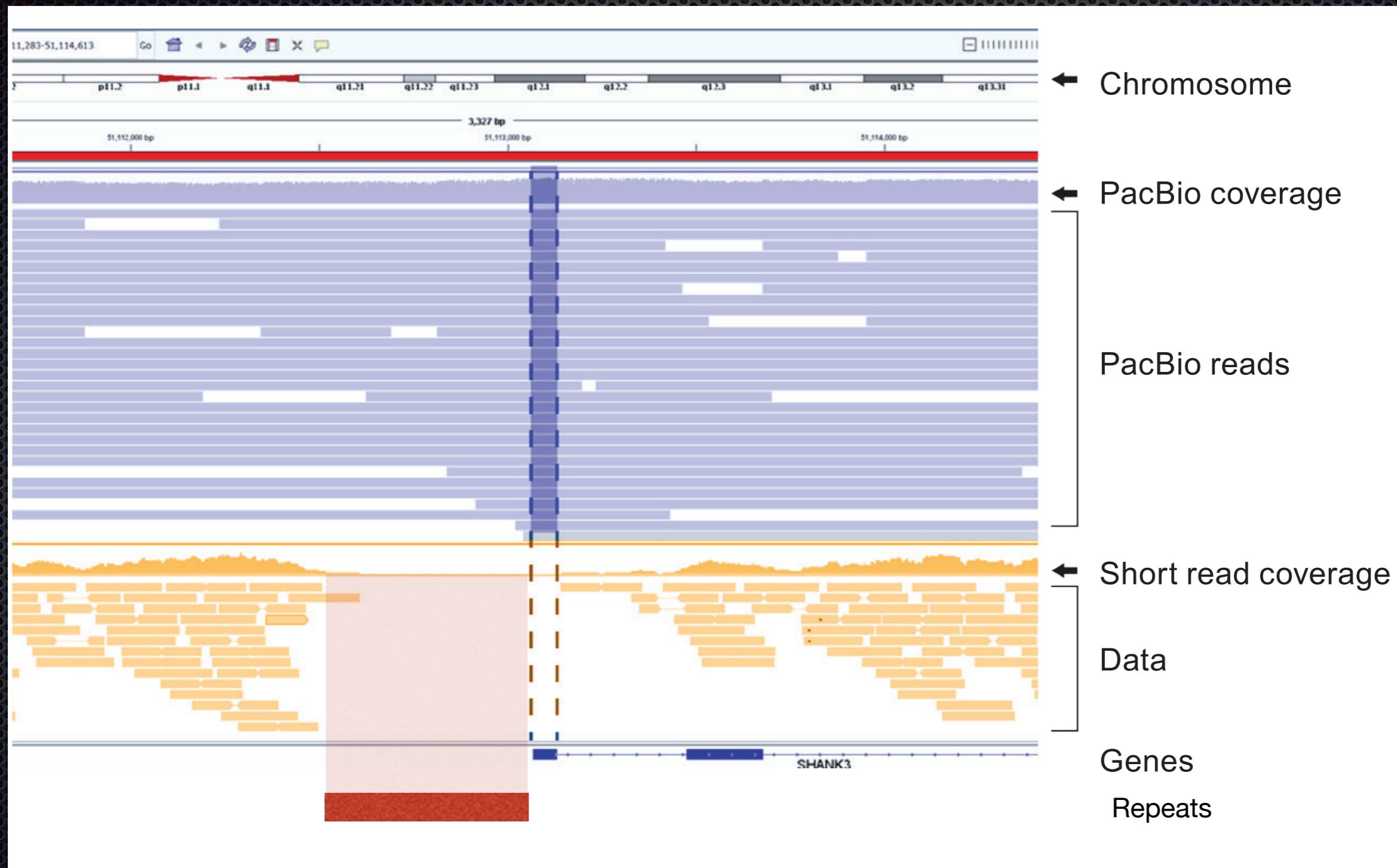
Long Read Lengths (Libraries <20 kb)

Half of data in reads: >190 kb
Data per SMRT Cell: Up to 50 Gb



The long read advantage

- Span repeats
- Improve mapping for gene families (paralogues)



Sequencing instruments

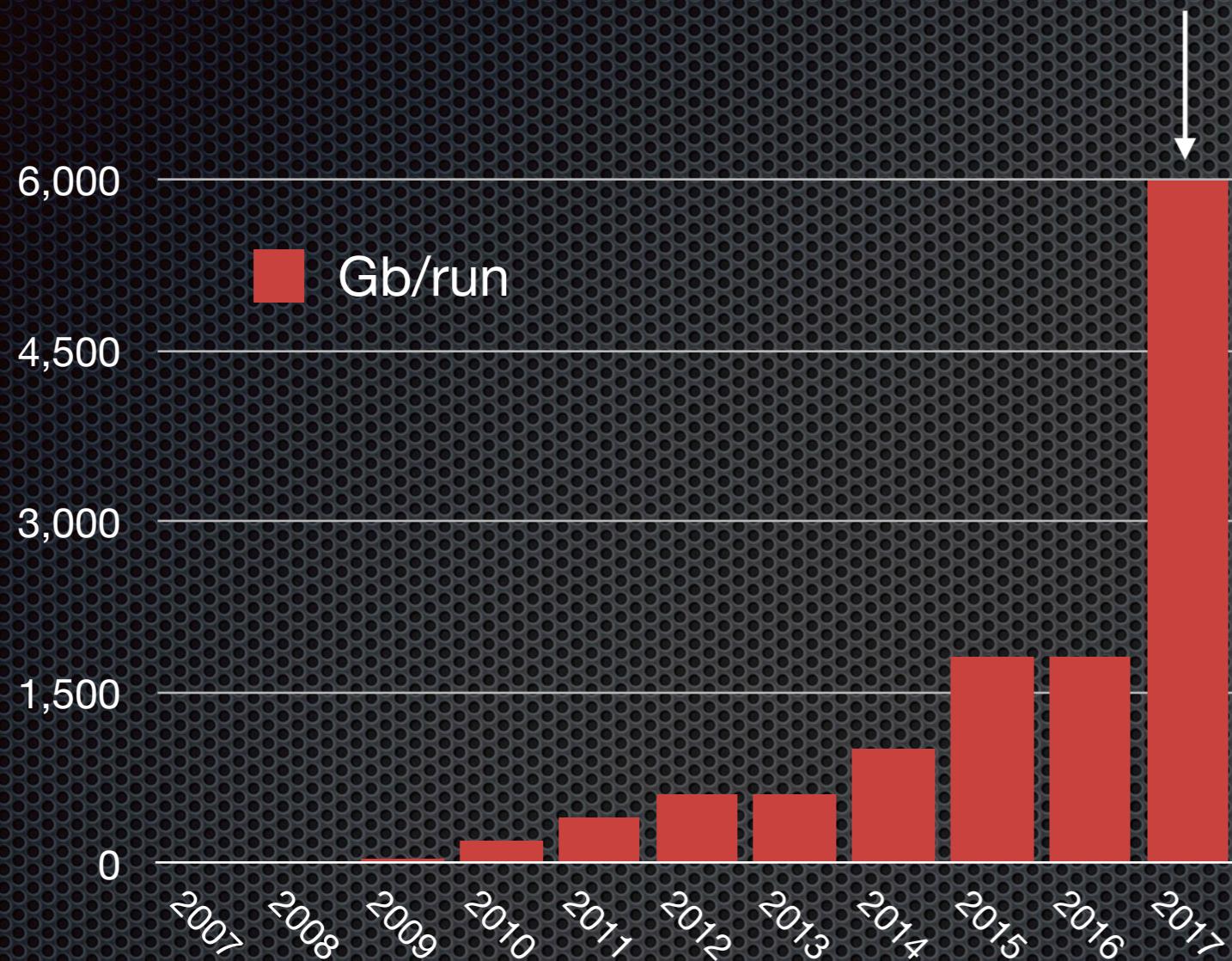


Platform	NovaSeq	HiSeq X	HiSeq 3000/4000	NextSeq	MiSeq	PacBio Sequel II
@NorSeq	2	4	1/4	5	4	1/1
Run time	1-2 days	1-3 days	1-5 days	29 hours	29 hours	0.5-30 hours
Read accuracy	99 %	99 %	99 %	99 %	99 %	99 %
Read number	20 10e9	6 10e9	4 10e9	400,000,000	20,000,000	4000000
Read length	2x150 bp	2x150 bp	2x125 bp	2x150 bp	2x300 bp	~20 kb
Output	6000 Gb	1800 Gb	1-1500 Gb	129 Gb	12 Gb	160 Gb

So what?

HTS throughput: data per run

2000x human genome

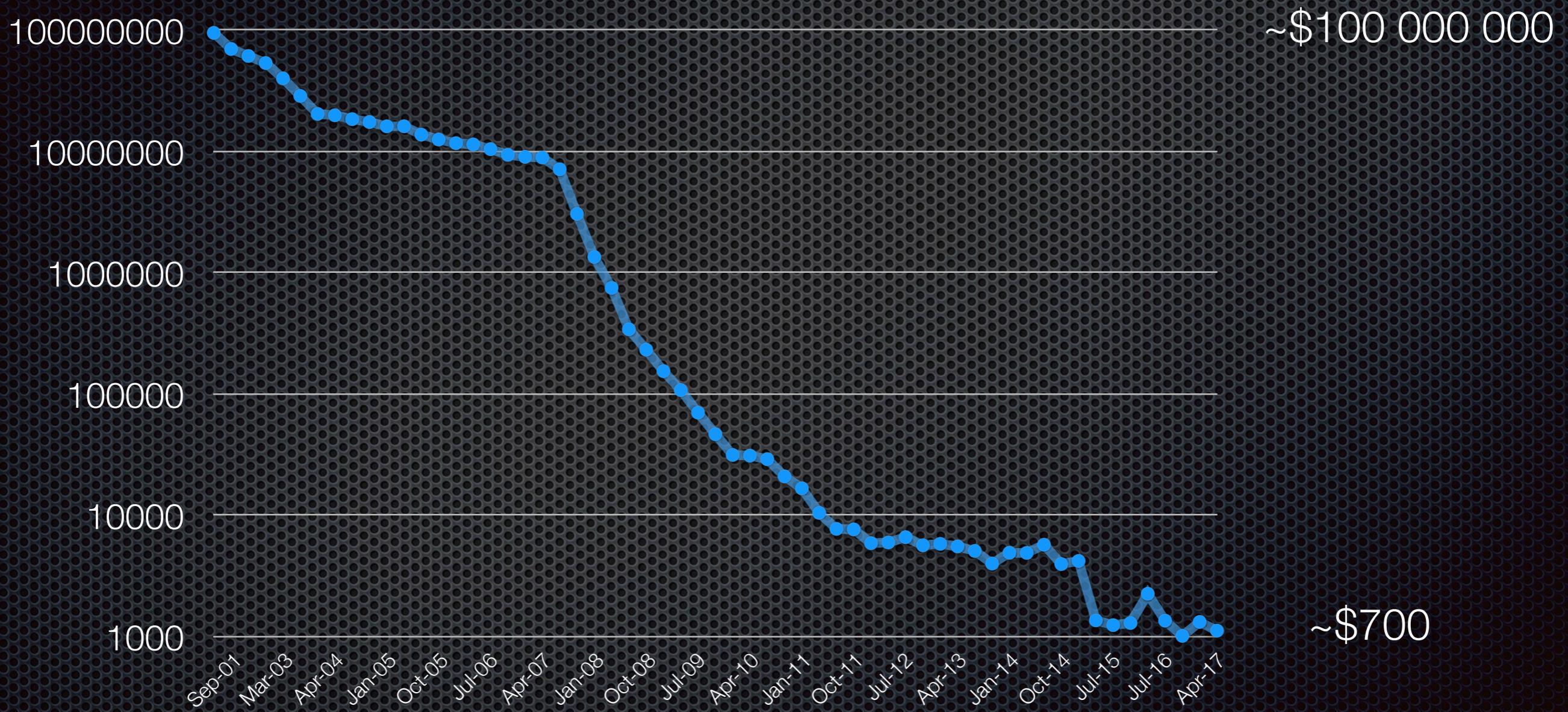


Illumina NovaSeq6000

- 6 Tb (6000 Gb)
- 2000x human genome

Sequencing costs

First human genome ~\$3 000 000 000



www.genome.gov/sequencingcosts/

Really, so what?



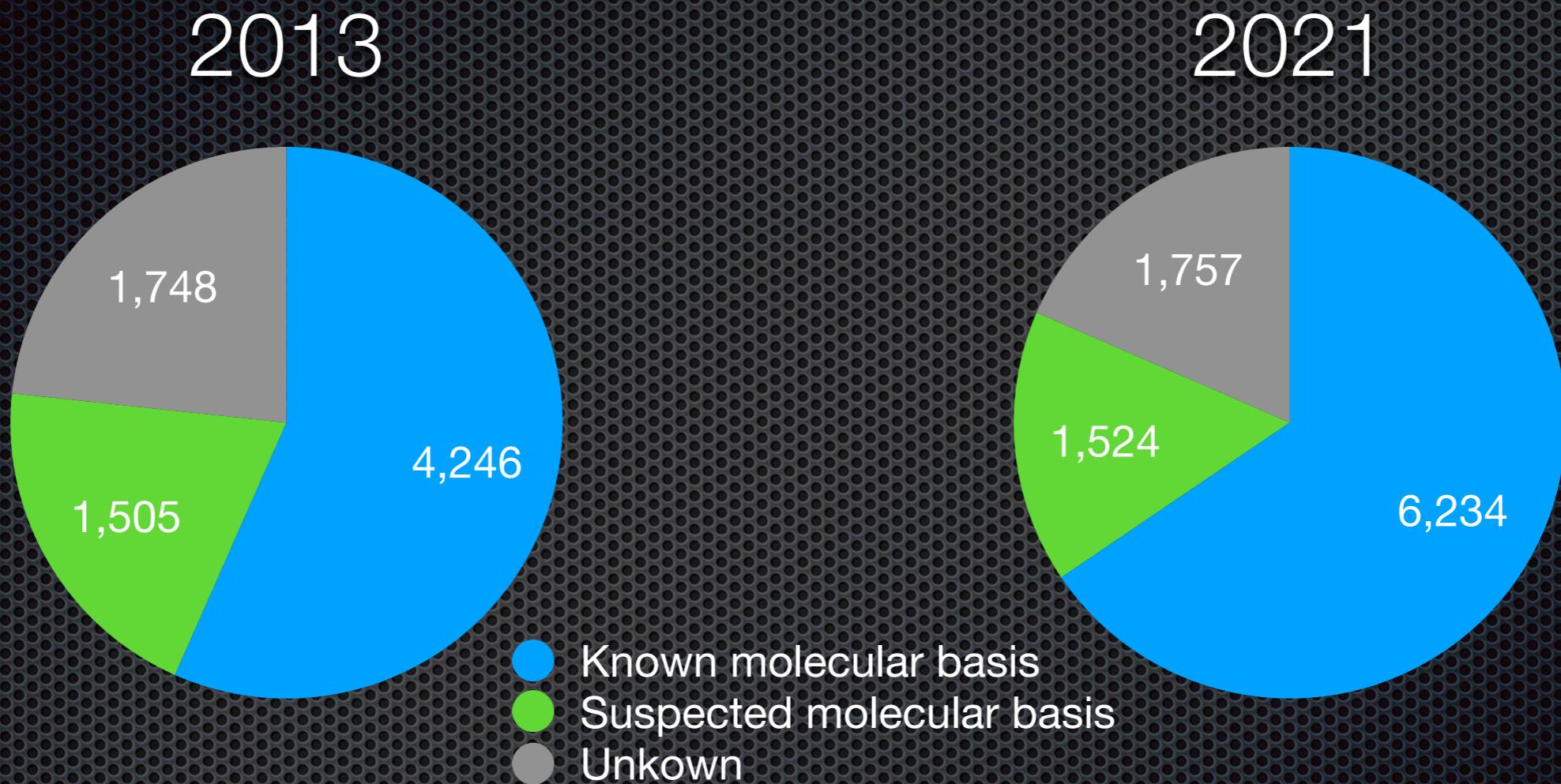
Parameter	ABI 3100	ABI 3730	NovaSeq6000
Read length	~700	~700	150 (x2)
Reads per run	16	96	20000000000
Run time	2 hours	30 minutes	2 days
Time for 1x human genome (3 Gb)	120 years	15 years	~90 seconds

Earth BioGenome Project

<https://www.earthbiogenome.org>

- Sequencing and annotate ~1.5 million known eukaryotic species
- “Sequencing life for the future of life”
- “...a moonshot for biology that aims to sequence, catalog, and characterize the genomes of all of Earth’s eukaryotic biodiversity over a period of 10 years.”
- Cost estimate: \$4.7 billion USD
- Less than the cost of creating the first draft human genome sequence

Impact of HTS on disease gene identification



2000 genetic disease genes in 8 years

NorSeq

The Norwegian Consortium for Sequencing and
Personalized Medicine



NorSeq-Tromsø

NorSeq-Trondheim



NorSeq-Bergen



NorSeq-Oslo

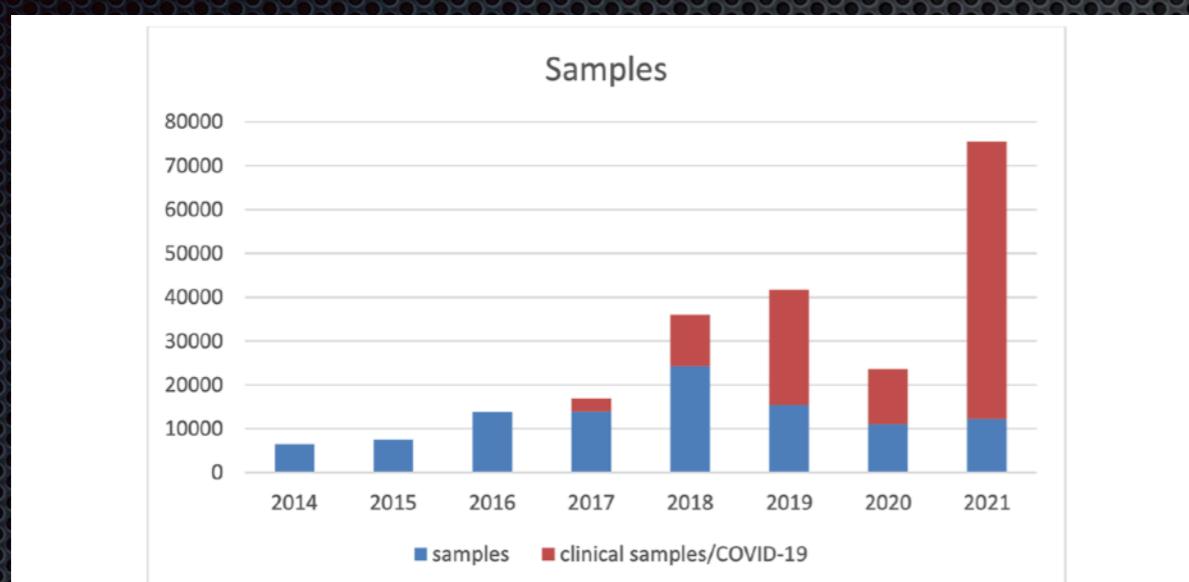
NorSeq-Cancer



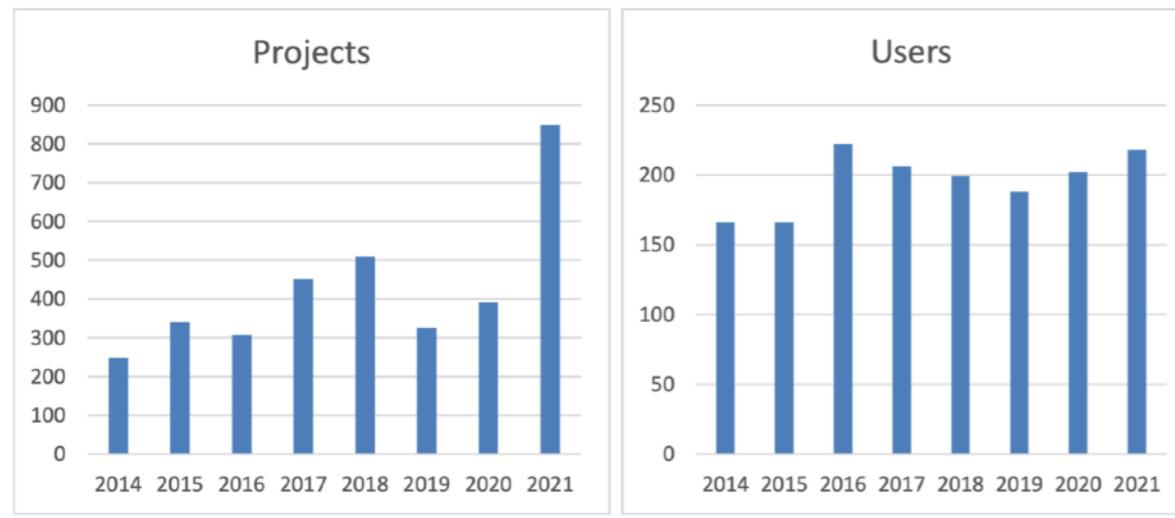
Coordinated national facility for HTS

- Services
 - Advice
 - DNA/RNA sample QC
 - DNA sequencing library prep
 - Sequencing
 - Data QC and delivery
 - Bioinformatics
 - Courses
- Planning and monitoring
 - Users and projects
 - Fast turnaround time
 - User experience
- Facilitate personalized medicine
 - Standardizing human diagnostic testing
 - Whole-genome sequencing for diagnostics
 - Bioinformatic pipelines

User statistics

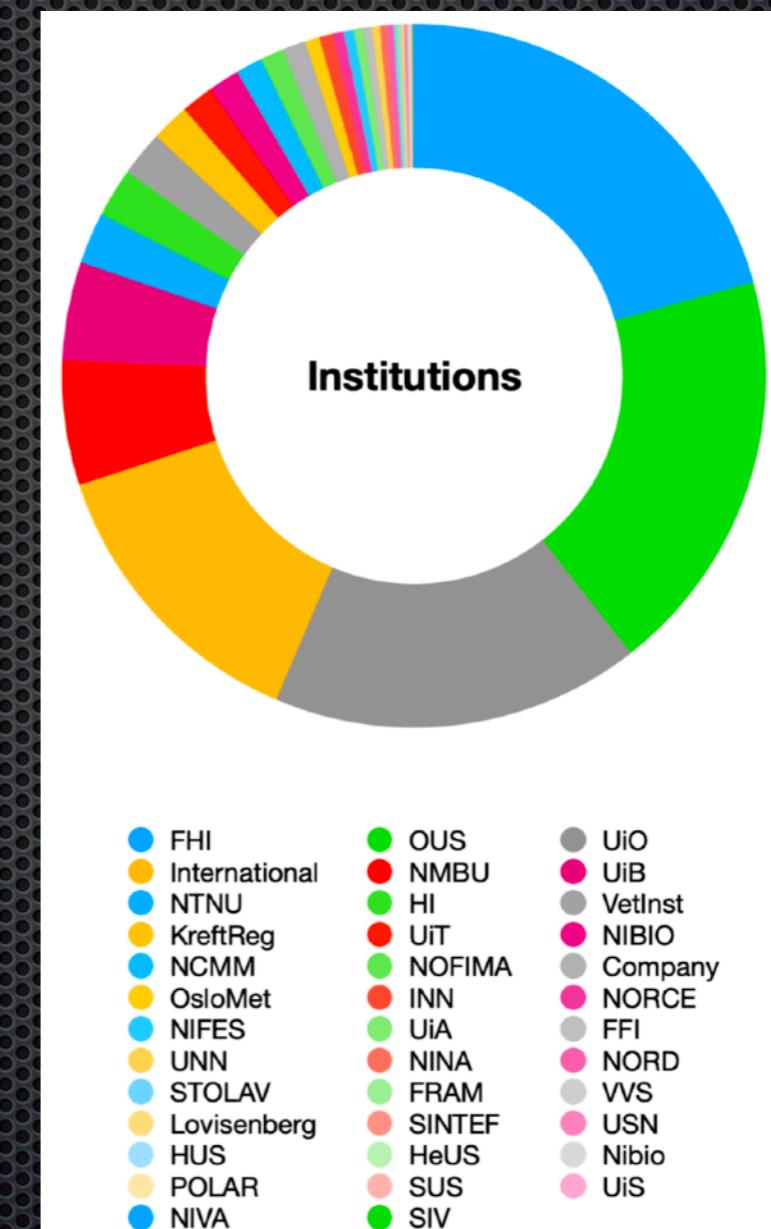


Figur 1. Antall prøver sekvensert per år under prosjektperioden

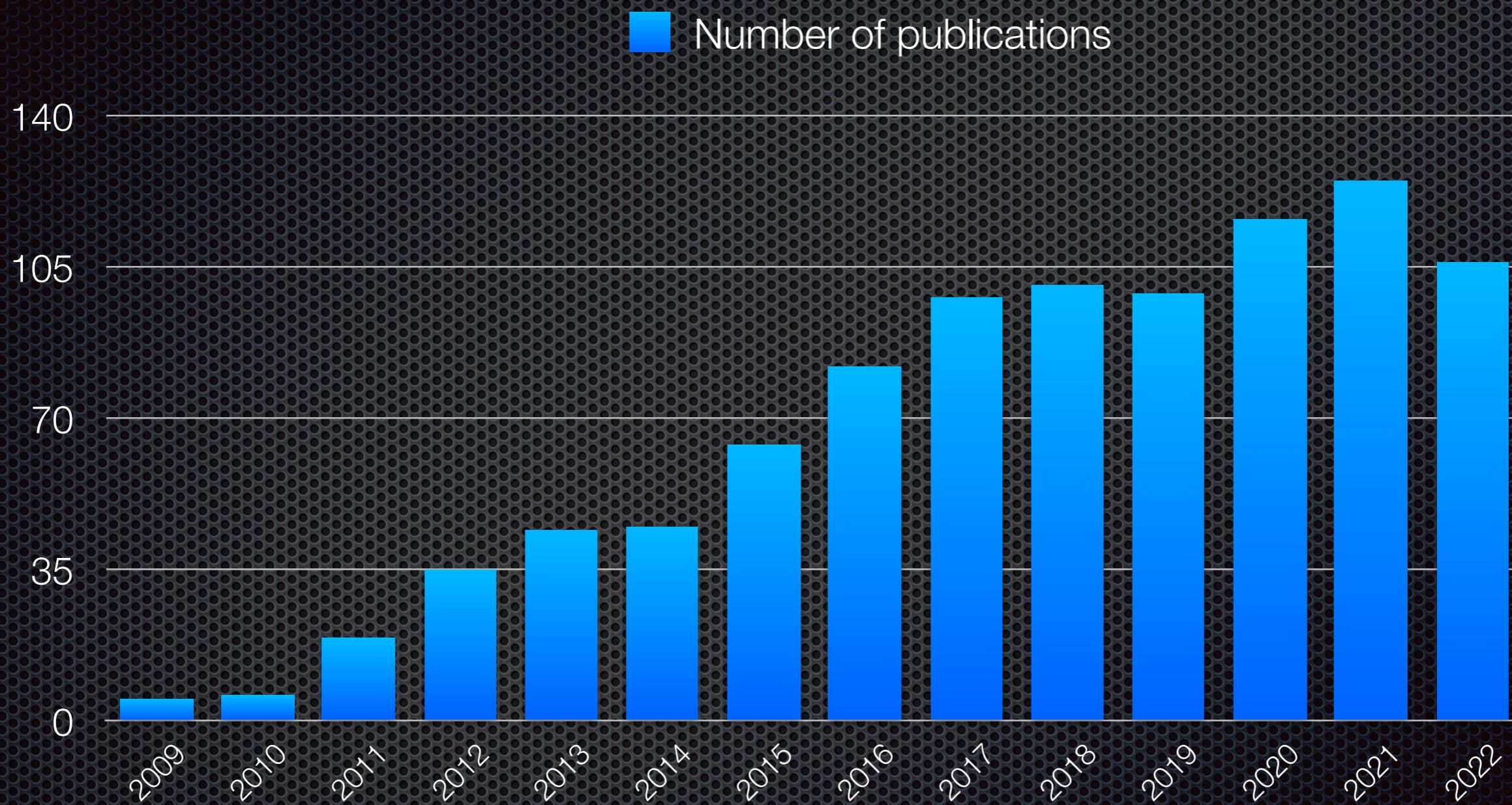


Figur 2. Antall prosjekter per år 2014-2021

Figur 3. Antall brukere per år 2014-2021



Publications from NSC-generated data



Total: 945

Norwegian Sequencing Centre (NSC)

The screenshot shows the homepage of the Norwegian Sequencing Centre (NSC) at <https://www.sequencing.uio.no/>. The page features a dark orange header bar with the NSC logo and a search bar. Below the header is a banner image of a DNA helix and laboratory glassware. The main content area includes a brief introduction, four service cards, and a news section.

The introduction states: "The Norwegian Sequencing Centre is a national technology core facility offering sequencing services on the HiSeq X, HiSeq 3000/4000, HiSeq 2500, NextSeq 500 & MiSeq instruments from Illumina, and also the Sequel from Pacific Biosciences."

The service cards are:

- Overview of our services** (Icon: computer monitor)
- Submit your sample(s) for sequencing** (Icon: hand holding a test tube)
- Before submitting your sample(s)** (Icon: question mark inside a circle)
 - learn about HTS technology
 - ask a specific question (FAQ)
 - need help planning your sequencing project?
- After sequencing** (Icon: DNA helix)
 - understanding the data delivered by NSC
 - analysing sequence data

The news section lists recent updates:

- Summer closure 2019** (Apr. 30, 2019 12:29 PM)
- Easter closure 2019** (Apr. 2, 2019 2:27 PM)
- 16S service launched!** (Jan. 11, 2019 4:46 PM)
- 10x Genomics is here!** (Jan. 11, 2019 4:45 PM)

The footer contains logos for The Research Council of Norway, Oslo universitetssykehus, HELSE SØR-ØST, and CEES.

<https://www.sequencing.uio.no/>

post@sequencing.uio.no

Amplicon sequencing service @ NSC

de Muinck et al. *Microbiome* (2017) 5:68
DOI 10.1186/s40168-017-0279-1

Microbiome

METHODOLOGY Open Access

 CrossMark

A novel ultra high-throughput 16S rRNA gene amplicon sequencing library preparation method for the Illumina HiSeq platform

Eric J. de Muinck^{1†}, Pål Trosvik^{1†}, Gregor D. Gilfillan², Johannes R. Hov³ and Arvind Y. M. Sundaram^{2*}

Abstract

Background: Advances in sequencing technologies and bioinformatics have made the analysis of microbial communities almost routine. Nonetheless, the need remains to improve on the techniques used for gathering such data, including increasing throughput while lowering cost and benchmarking the techniques so that potential sources of bias can be better characterized.

Methods: We present a triple-index amplicon sequencing strategy to sequence large numbers of samples at significantly lower cost and in a shorter timeframe compared to existing methods. The design employs a two-stage PCR protocol, incorporating three barcodes to each sample, with the possibility to add a fourth-index. It also includes heterogeneity spacers to overcome low complexity issues faced when sequencing amplicons on Illumina platforms.

Results: The library preparation method was extensively benchmarked through analysis of a mock community in order to assess biases introduced by sample indexing, number of PCR cycles, and template concentration. We further evaluated the method through re-sequencing of a standardized environmental sample. Finally, we evaluated our protocol on a set of fecal samples from a small cohort of healthy adults, demonstrating good performance in a realistic experimental setting. Between-sample variation was mainly related to batch effects, such as DNA extraction, while sample indexing was also a significant source of bias. PCR cycle number strongly influenced chimera formation and affected relative abundance estimates of species with high GC content. Libraries were sequenced using the Illumina HiSeq and MiSeq platforms to demonstrate that this protocol is highly scalable to sequence thousands of samples at a very low cost.

Conclusions: Here, we provide the most comprehensive study of performance and bias inherent to a 16S rRNA gene amplicon sequencing method to date. Triple-indexing greatly reduces the number of long custom DNA oligos required for library preparation, while the inclusion of variable length heterogeneity spacers minimizes the need for PhiX spike-in. This design results in a significant cost reduction of highly multiplexed amplicon sequencing. The biases we characterize highlight the need for highly standardized protocols. Reassuringly, we find that the biological signal is a far stronger structuring factor than the various sources of bias.

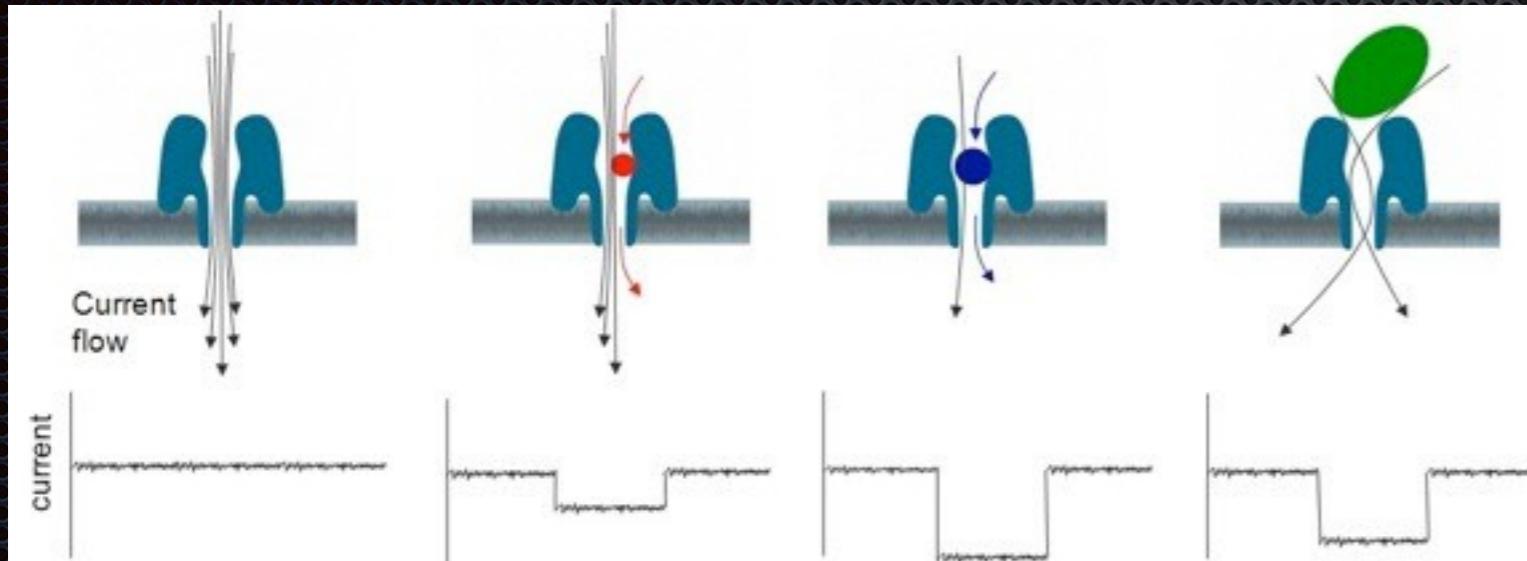
Keywords: 16S rRNA gene amplicon sequencing, Illumina library preparation, Indexed PCR, Mock community, Environmental sequencing, Benchmarking, PCR bias, Chimera formation

<https://tinyurl.com/2tayem2t>

Other technologies

Nanopores and single cells

Oxford Nanopore



- read lengths 100s kb
 - longest read 2.1 Mb
- G A T C mC hmC
- directly sequence RNA

<http://www.nanoporetech.com/>

Oxford Nanopore sequencers



MinION
Small-scale

GridION
Large genomes

SmidgION
?

LETTER

doi:10.1038/nature16996

Real-time, portable genome sequencing for Ebola surveillance

Joshua Quick^{1*}, Nicholas J. Loman^{1*}, Sophie Duraffour^{2,3*}, Jared T. Simpson^{4,5*}, Ettore Severi^{6*}, Lauren Cowley^{7*}, Joseph Akoi Bore², Raymond Koundouno², Gytis Dudas⁸, Amy Mikhail⁷, Nobila Ouédraogo⁹, Babak Afrough^{2,10}, Amadou Bah^{2,11}, Jonathan H. J. Baum^{2,3}, Beate Becker-Ziaja^{2,3}, Jan Peter Boettcher^{2,12}, Mar Cabeza-Cabrero^{2,3}, Álvaro Camino-Sánchez², Lisa L. Carter^{2,13}, Juliane Doerrbecker^{2,3}, Theresa Enkirch^{2,14}, Isabel García-Dorival^{2,15}, Nicole Hetzelt^{2,12}, Julia Hinzenmann^{2,12}, Tobias Holm^{2,3}, Liana Eleni Kafetzopoulou^{2,16}, Michel Koropogui^{2,17}, Abigail Kosgey^{2,18}, Eeva Kuisma^{2,10}, Christopher H. Logue^{2,10}, Antonio Mazzarelli^{2,19}, Sarah Meisel^{2,3}, Marc Mertens^{2,20}, Janine Michel^{2,12}, Didier Ngabo^{2,10}, Katja Nitzsche^{2,3}, Elisa Pallasch^{2,3}, Livia Victoria Patrono^{2,3}, Jasmine Portmann^{2,21}, Johanna Gabriella Repits^{2,22}, Natasha Y. Rickett^{2,15,23}, Andreas Sachse^{2,12}, Katrin Singethan^{2,24}, Inês Vitoriano^{2,10}, Rahel L. Yemanaberhan^{2,3}, Elsa G. Zekeng^{2,15,23}, Trina Racine²⁵, Alexander Bello²⁵, Amadou Alpha Sall²⁶, Ousmane Faye²⁶, Oumar Faye²⁶, N'Faly Magassouba²⁷, Cecelia V. Williams^{28,29}, Victoria Amburgey^{28,29}, Linda Winona^{28,29}, Emily Davis^{29,30}, Jon Gerlach^{29,30}, Frank Washington^{29,30}, Vanessa Monteil³¹, Marine Jourdain³¹, Marion Bererd³¹, Alimou Camara³¹, Hermann Somlare³¹, Abdoulaye Camara³¹, Marianne Gerard³¹, Guillaume Bado³¹, Bernard Baillet³¹, Déborah Delaune^{32,33}, Koumpingnin Yacouba Nebie³⁴, Abdoulaye Diarra³⁴, Yacouba Savane³⁴, Raymond Bernard Pallawo³⁴, Giovanna Jaramillo Gutierrez³⁵, Natacha Milhano^{6,36}, Isabelle Roger³⁴, Christopher J. Williams^{6,37}, Facinet Vattara¹⁷, Kuiama Lewandowski¹⁰, James Taylor³⁸, Phillip Rachwal³⁸, Daniel J. Turner³⁹, Georgios Pollakis^{15,23}, Julian A. Hiscox^{15,23}, David A. Matthews⁴⁰, Matthew K. O'Shea⁴¹, Andrew McD. Johnston⁴¹, Duncan Wilson⁴¹, Emma Hutley⁴², Erasmus Smit⁴³, Antonino Di Caro^{2,19}, Roman Wölfel^{2,44}, Kilian Stoecker^{2,44}, Erna Fleischmann^{2,44}, Martin Gabriel^{2,3}, Simon A. Weller³⁸, Lamine Koivogui⁴⁵, Boubacar Diallo³⁴, Sakoba Keïta¹⁷, Andrew Rambaut^{8,46,47}, Pierre Formenty³⁴, Stephan Günther^{2,3} & Miles W. Carroll^{2,10,48,49}

The Ebola virus disease epidemic in West Africa is the largest on record, responsible for over 28,599 cases and more than 11,299 deaths¹. Genome sequencing in viral outbreaks is desirable to characterise the evolution of the virus and to identify the source of new infections.

Owing to a lack of local sequencing capacity coupled with practical difficulties transporting samples to remote sequencing facilities², we address this problem here by devising a genomic surveillance system.

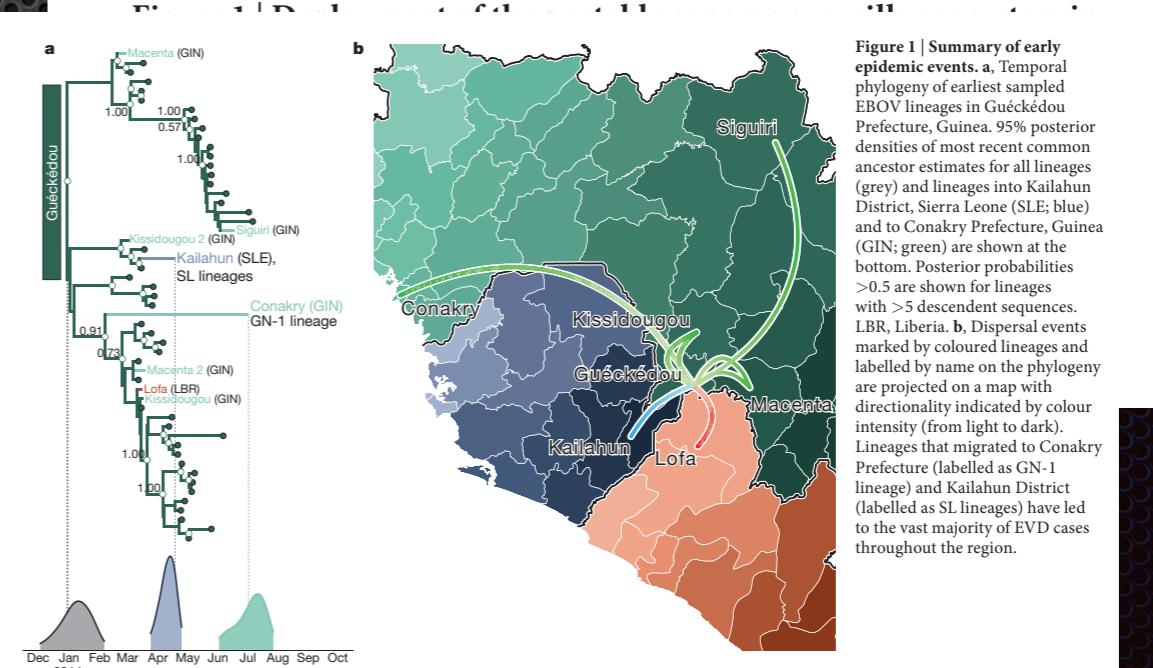
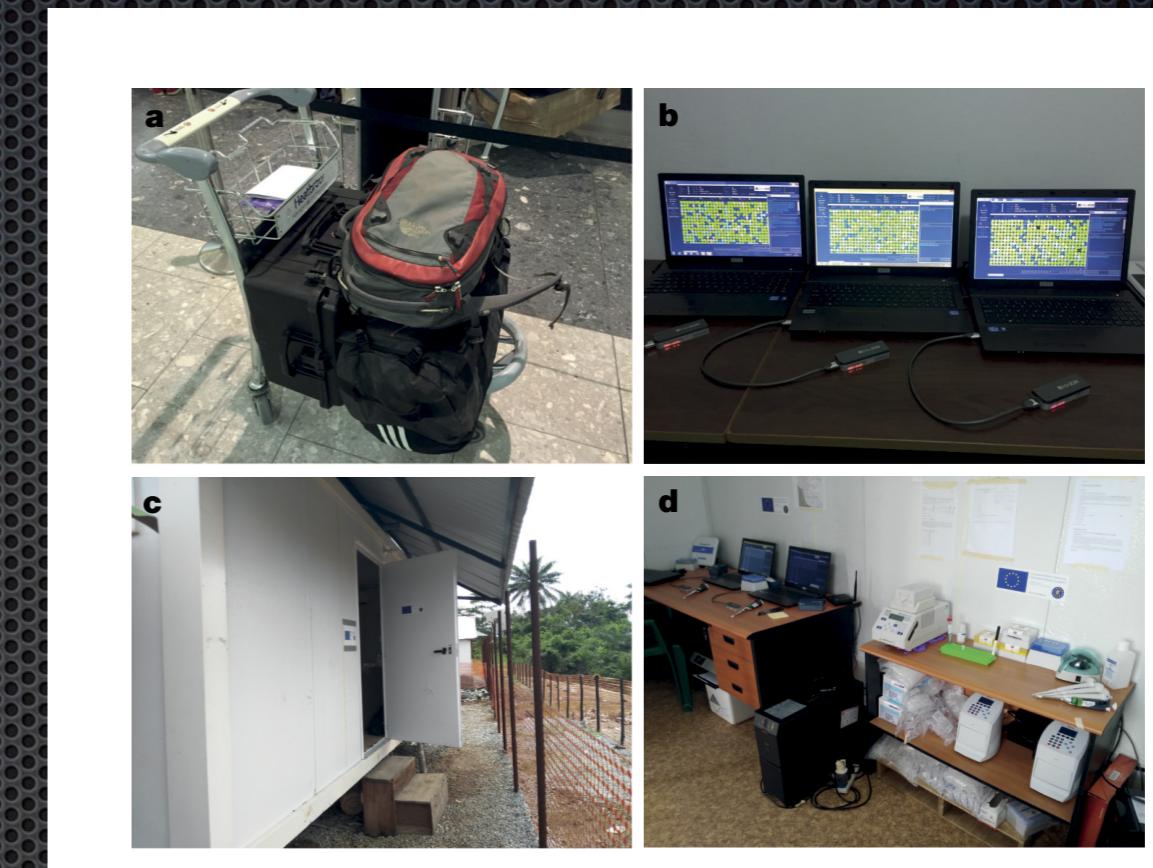
ARTICLE

doi:10.1038/nature22040

Virus genomes reveal factors that spread and sustained the Ebola epidemic

A list of authors and their affiliations appears at the end of the paper

The 2013–2016 West African epidemic caused by the Ebola virus was of unprecedented magnitude, duration and impact. Here we reconstruct the dispersal, proliferation and decline of Ebola virus throughout the region by analysing 1,610 Ebola virus genomes, which represent over 5% of the known cases. We test the association of geography, climate and demography with viral movement among administrative regions, inferring a classic ‘gravity’ model, with intense dispersal between larger and closer populations. Despite attenuation of international dispersal after border closures, cross-border transmission had already sown the seeds for an international epidemic, rendering these measures ineffective at curbing the epidemic. We address why the epidemic did not spread into neighbouring countries, showing that these countries were susceptible to substantial outbreaks but at lower risk of introductions. Finally, we reveal that this large epidemic was a heterogeneous and spatially dissociated collection of transmission clusters of varying size, duration and connectivity. These insights will help to inform interventions in future epidemics.



News

Mobile laboratories use LamPORE COVID-19 test, UK Government LamPORE evaluation report shows high accuracy

Fri 29th January 2021



Mobile COVID-19 testing laboratories containing Oxford Nanopore's LamPORE test are now being deployed in a pilot programme, to support testing efforts, including in remote locations.

Telomere-to-telomere (T2T)

RESEARCH ARTICLE

HUMAN GENOMICS

The complete sequence of a human genome

Sergey Nurk^{1†}, Sergey Koren^{1†}, Arang Rhee^{1†}, Mikko Rautiainen^{1†}, Andrey V. Bzikadze², Alla Mikheenko³, Mitchell R. Vollger⁴, Nicolas Altemose⁵, Lev Uralsky^{6,7}, Ariel Gershman⁸, Sergey Aganezov^{9†}, Savannah J. Hoyt¹⁰, Mark Diekhans¹¹, Glennis A. Logsdon⁴, Michael Alonge⁹, Stylianos E. Antonarakis¹², Matthew Borchers¹³, Gerard G. Bouffard¹⁴, Shelise Y. Brooks¹⁴, Gina V. Caldas¹⁵, Nae-Chyun Chen⁹, Haoyu Cheng^{16,17}, Chen-Shan Chin¹⁸, William Chow¹⁹, Leonardo G. de Lima¹³, Philip C. Dishuck⁴, Richard Durbin^{19,20}, Tatiana Dvorkina³, Ian T. Fiddes²¹, Giulio Formenti^{22,23}, Robert S. Fulton²⁴, Arkarachai Fungtammasan¹⁸, Erik Garrison^{11,25}, Patrick G. S. Grady¹⁰, Tina A. Graves-Lindsay²⁶, Ira M. Hall²⁷, Nancy F. Hansen²⁸, Gabrielle A. Hartley¹⁰, Marina Haukness¹¹, Kerstin Howe¹⁹, Michael W. Hunkapiller²⁹, Chirag Jain^{1,30}, Miten Jain¹¹, Erich D. Jarvis^{22,23}, Peter Kerpeljiev³¹, Melanie Kirsche⁹, Mikhail Kolmogorov³², Jonas Korlach²⁹, Milinn Kremitzki²⁶, Heng Li^{16,17}, Valerie V. Maduro³³, Tobias Marschall³⁴, Ann M. McCartney¹, Jennifer McDaniel³⁵, Danny E. Miller^{4,36}, James C. Mullikin^{14,28}, Eugene W. Myers³⁷, Nathan D. Olson³⁵, Benedict Paten¹¹, Paul Peluso²⁹, Pavel A. Pevzner³², David Porubsky⁴, Tamara Potapova¹³, Evgeny I. Rogaev^{6,7,38,39}, Jeffrey A. Rosenfeld⁴⁰, Steven L. Salzberg^{9,41}, Valerie A. Schneider⁴², Fritz J. Sedlazeck⁴³, Kishwar Shafin¹¹, Colin J. Shew⁴⁴, Alaina Shumate⁴¹, Ying Sims¹⁹, Arian F. A. Smit⁴⁵, Daniela C. Soto⁴⁴, Ivan Sovic^{29,46}, Jessica M. Storer⁴⁵, Aaron Streets^{5,47}, Beth A. Sullivan⁴⁸, Françoise Thibaud-Nissen⁴², James Torrance¹⁹, Justin Wagner³⁵, Brian P. Walenz¹, Aaron Wenger²⁹, Jonathan M. D. Wood¹⁹, Chunlin Xiao⁴², Stephanie M. Yan⁴⁹, Alice C. Young¹⁴, Samantha Zarate⁹, Urvashi Surti⁵⁰, Rajiv C. McCoy⁴⁹, Megan Y. Dennis⁴⁴, Ivan A. Alexandrov^{3,7,51}, Jennifer L. Gerton^{13,52}, Rachel J. O'Neill¹⁰, Winston Timp^{8,41}, Justin M. Zook³⁵, Michael C. Schatz^{9,49}, Evan E. Eichler^{4,53*}, Karen H. Miga^{11,54*}, Adam M. Phillippy^{1*}

Since its initial release in 2000, the human reference genome has covered only the euchromatic fraction of the genome, leaving important heterochromatic regions unfinished. Addressing the remaining 8% of the genome, the Telomere-to-Telomere (T2T) Consortium presents a complete 3.055 billion-base pair sequence of a human genome, T2T-CHM13, that includes gapless assemblies for all chromosomes except Y, corrects errors in the prior references, and introduces nearly 200 million base pairs of sequence containing 1956 gene predictions, 99 of which are predicted to be protein coding. The completed regions include all centromeric satellite arrays, recent segmental duplications, and the short arms of all five acrocentric chromosomes, unlocking these complex regions of the genome to variational and functional studies.

HUMAN GENOMICS

Epigenetic patterns in a complete human genome

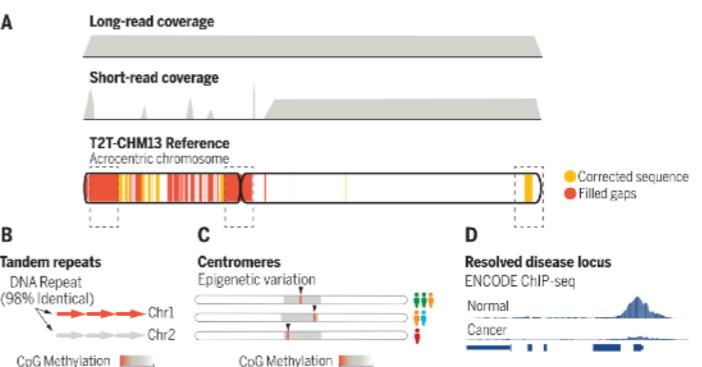
Ariel Gershman, Michael E. G. Sauria, Xavi Guitart, Mitchell R. Vollger, Paul W. Hook, Savannah J. Hoyt, Miten Jain, Alaina Shumate, Roham Razaghi, Sergey Koren, Nicolas Altemose, Gina V. Caldas, Glennis A. Logsdon, Arang Rhee, Evan E. Eichler, Michael C. Schatz, Rachel J. O'Neill, Adam M. Phillippy, Karen H. Miga*, Winston Timp*

INTRODUCTION: The human reference genome has served as the foundation for many large-scale initiatives, including the collective effort to catalog the epigenome, the set of marks and protein interactions that act to control gene activity and cellular function. However, for more than two decades, efforts to construct a complete epigenome have been hampered by an incomplete reference genome. With recent technological advances, we can now study genome structure and function comprehensively across a complete telomere-to-telomere human genome assembly, T2T-CHM13. As a result, we can now broaden the human epigenome to include 225 million base pairs (Mbp) of additional sequence.

RATIONALE: The epigenome refers to DNA modifications (e.g., CpG methylation), protein-DNA interactions, histone modifications, and chromatin organization that collectively influence gene expression, genome regulation, and genome stability. These epigenetic features are heritable upon cell division but dynamic

during development, generating profiles that are unique to different tissues and cell types. Here, we present an epigenetic annotation of the human genome in which we explore previously unresolved regions, including acrocentric chromosome short arms, segmentally duplicated genes, and a diverse collection of repeat classes, including human centromeres. Generating a complete epigenetic annotation of the previously missing 8% of the human genome provides a foundation for elucidating the functional roles of these genomic elements that are critical to our understanding of genome regulation, function, and evolution.

RESULTS: Completion of the human epigenome required that we develop approaches to profiling the previously unresolved regions. Using the T2T-CHM13 reference with existing short-read epigenetic data, we identified 3 to 19% more enrichment sites for epigenetic markers. However, even with the complete reference, these short-read epigenetic methods cannot correctly resolve regions of the genome of



Epigenetic characterization across a complete human genome. (A) The T2T-CHM13 reference contains filled gaps and corrected sequences. Using short- and long-read sequencing data, we functionally annotated these added regions. (B) Tandem repeats, which are nearly identical, vary in epigenetic state depending on genomic location. (C) The epigenetic basis of centromere identity is variable among diverse individuals. (D) In genes associated with disease, short reads mapped to T2T-CHM13 elucidate epigenetic dysregulation in human disease states.

Nurk, S. et al. The complete sequence of a human genome. Science 376, 44–53 (2022).

Gershman, A. et al. Epigenetic patterns in a complete human genome. Science 376, eabj5089 (2022).

Single-cell sequencing

Single-cell sequencing

10x Genomics Chromium system



Single Cell Genomics

Copy Number Variation

Single Cell Transcriptomics

Gene Expression Profiling **IMPROVED!**

Gene Expression CRISPR Screening **NEW!**

Gene & Cell Surface Protein **NEW!**

Immune Profiling

Immune Profiling & Cell Surface Protein **NEW!**

Immune Profiling & Antigen Specificity **NEW!**

Single Cell Epigenomics

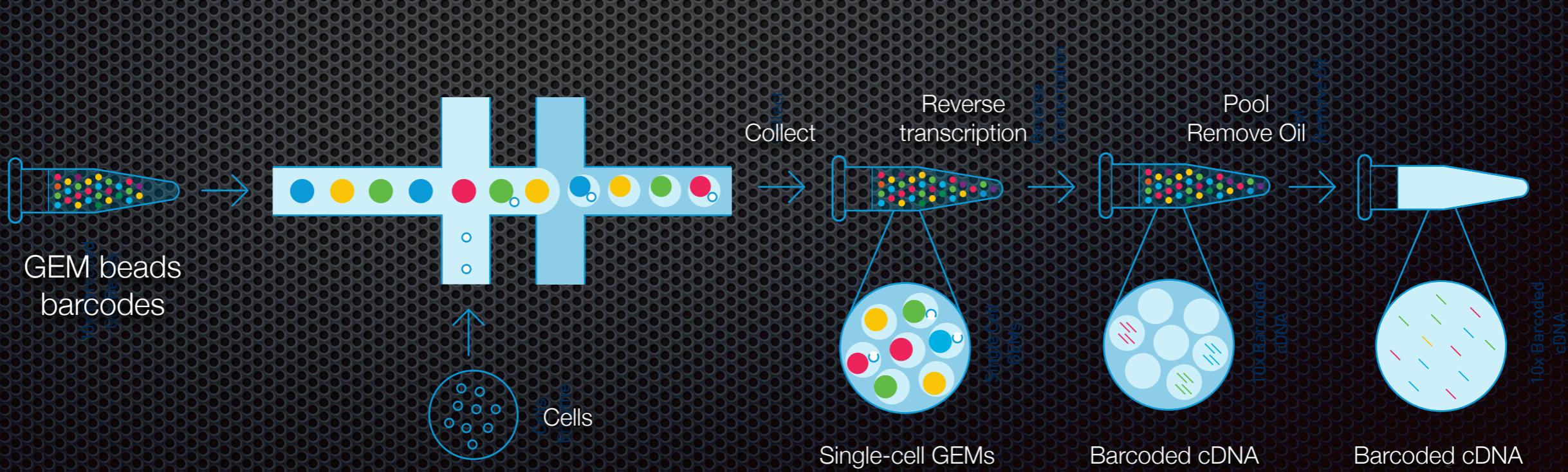
Chromatin Accessibility **NEW!**

Linked-Reads Genomics

Whole Genome Sequencing

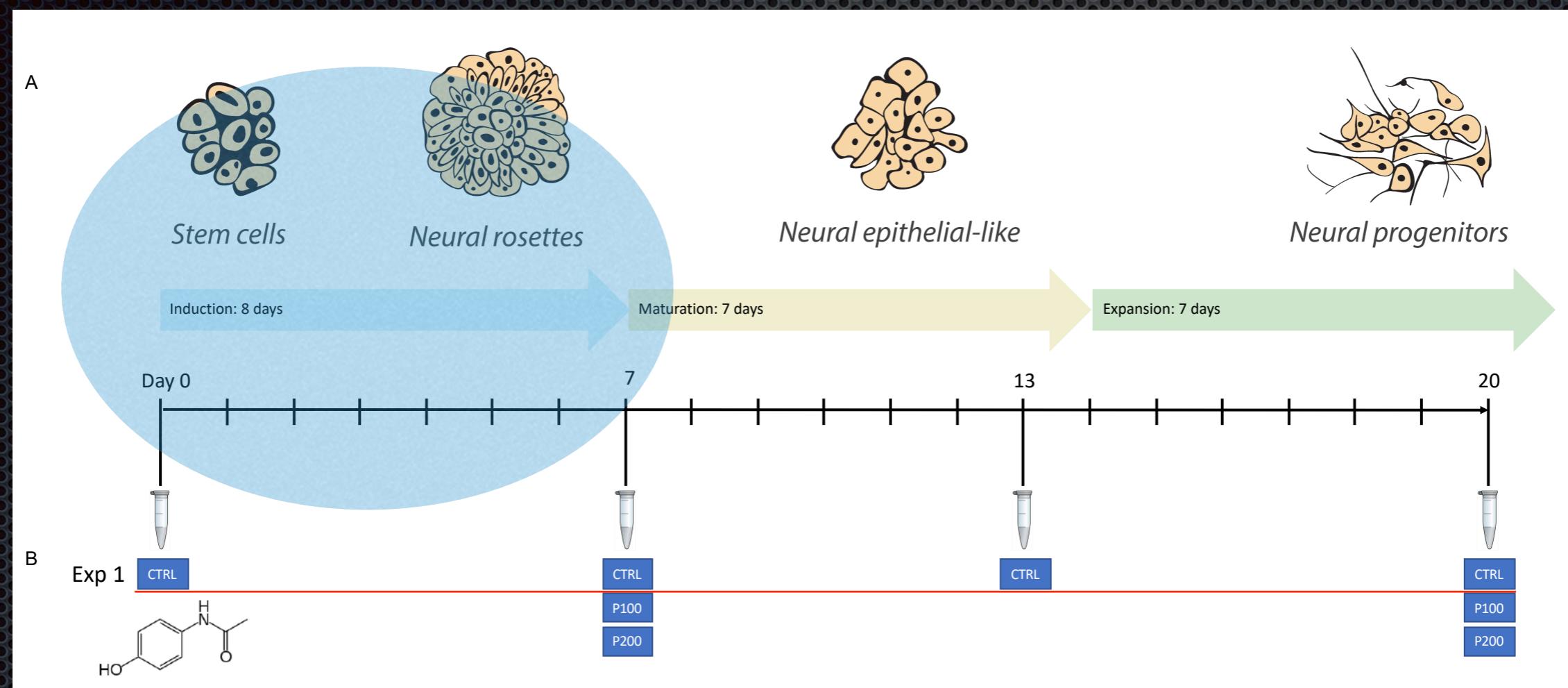
Exome Sequencing

de novo Assembly



Single-cell RNAseq

Embryonic stem cell model for drugs in pregnancy



- Protocol problems
- qPCR and FACs
- Variation in cell types during differentiation?
- Consistency of differentiation protocol?

Article

A multi-omics approach to visualize early neuronal differentiation from hESCs in 4D

Athina Samara,^{1,2,18,*} Mari Spildejorje,^{3,4,5,18} Ankush Sharma,^{6,7,8,14,18} Martin Falck,^{3,8,15} Magnus Leithaug,^{4,16} Stefania Modafferi,^{4,17} Pål Marius Bjørnstad,⁴ Ganesh Acharya,^{9,10} Kristina Gervin,^{3,11,12} Robert Lyle,^{3,4,13,*} and Ragnhild Eskeland^{3,7,19,*}

SUMMARY

Neuronal differentiation of pluripotent stem cells is an established method to study physiology, disease, and medication safety. However, the sequence of events in human neuronal differentiation and the ability of *in vitro* models to recapitulate early brain development are poorly understood. We developed a protocol optimized for the study of early human brain development and neuropharmacological applications. We comprehensively characterized gene expression and epigenetic profiles at four timepoints, because the cells differentiate from embryonic stem cells towards a heterogeneous population of progenitors, immature and mature neurons bearing telencephalic signatures. A multi-omics roadmap of neuronal differentiation, combined with searchable interactive gene analysis tools, allows for extensive exploration of early neuronal development and the effect of medications.

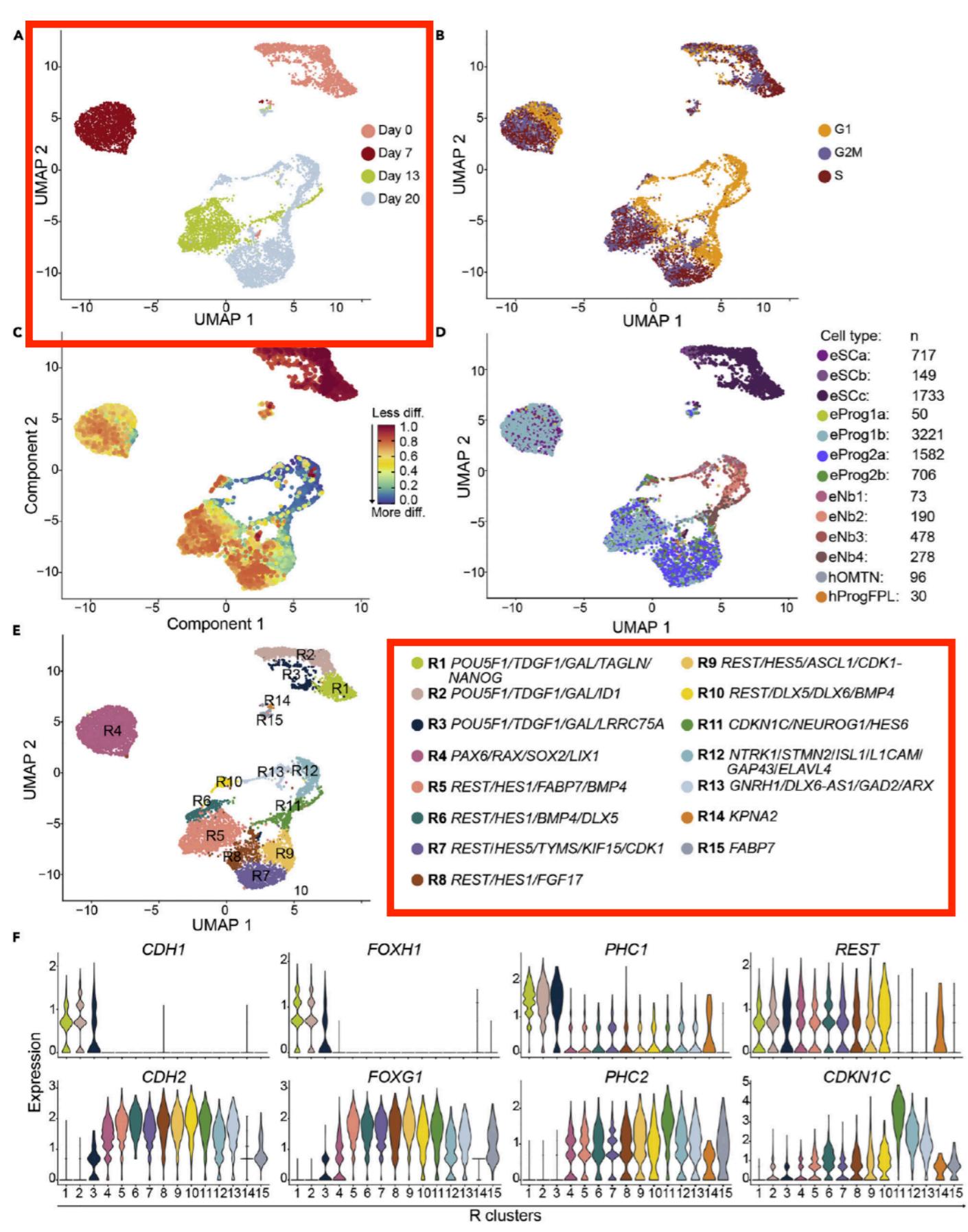
¹Division of Clinical Paediatrics, Department of Women's and Children's Health, Karolinska Institutet, Solna, Sweden

²Astrid Lindgren Children's Hospital Karolinska University Hospital, Stockholm, Sweden

³PharmaTox Strategic Research Initiative, Faculty of Mathematics and Natural Sciences, University of Oslo, Oslo, Norway

⁴Department of Medical Genetics, Oslo University Hospital and University of Oslo, Oslo, Norway

⁵Institute of Clinical



Cell clusters
change over
developmental
time

Clusters identified by
known genes

Samara, A. et al. A multi-omics approach to
visualize early neuronal differentiation from hESCs
in 4D. iScience 25, 105279 (2022).

Summary

- High-throughput sequencing
 - Dramatic increase in sequence production
 - Many applications on one platform
 - Field new and moving very quickly
 - Diagnostic (exome) sequencing in place
 - Huge impact on human/medical genetics
- Challenges/opportunities
 - Data storage/backup/distribution
 - Data analysis
 - Whole-genome sequencing?