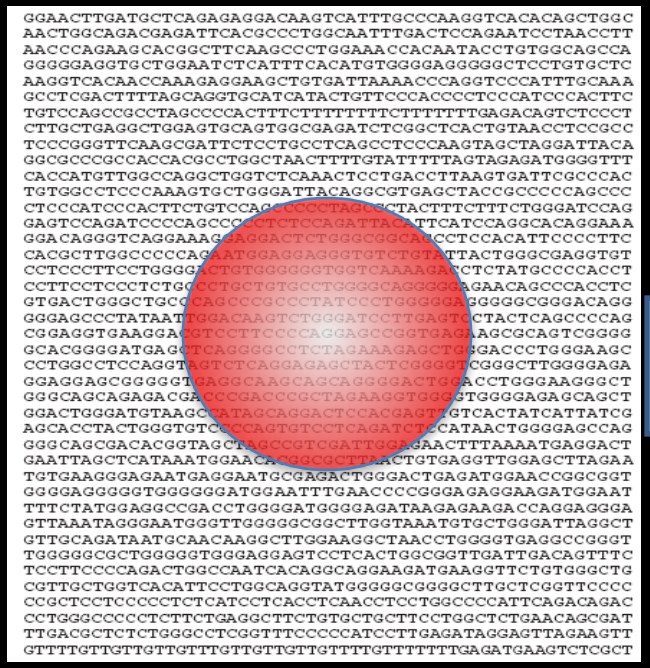
Sequencing approaches and power

Christopher E. Mason, Ph.D.
Associate Professor
Director, 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
March 26, 2019

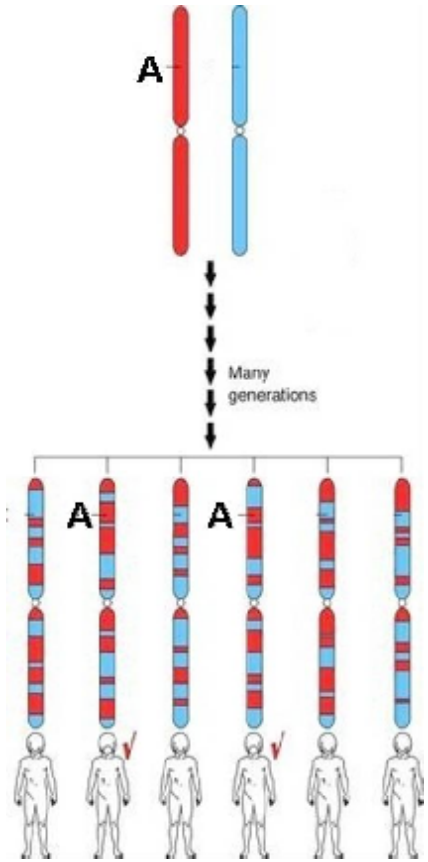
Start

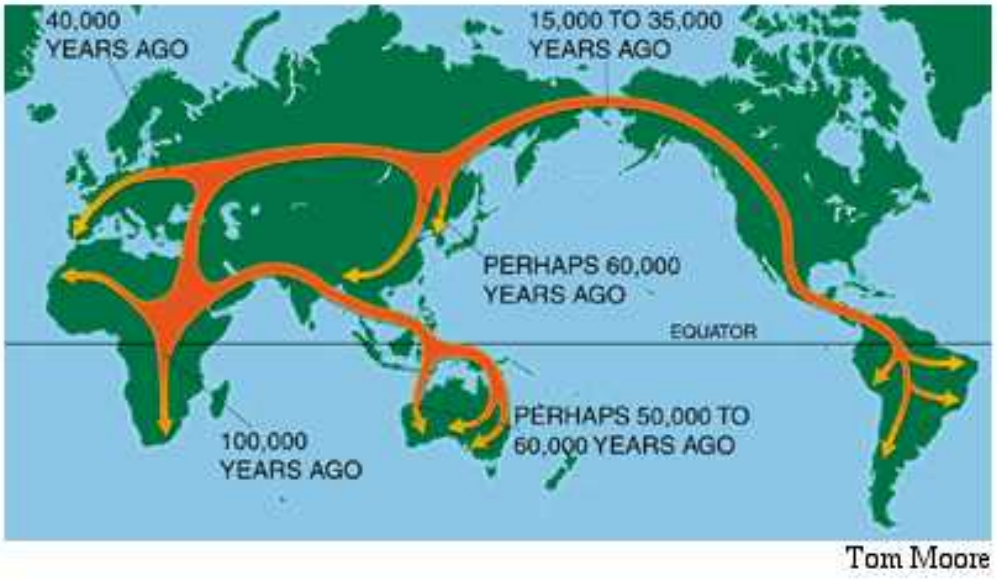
Finish



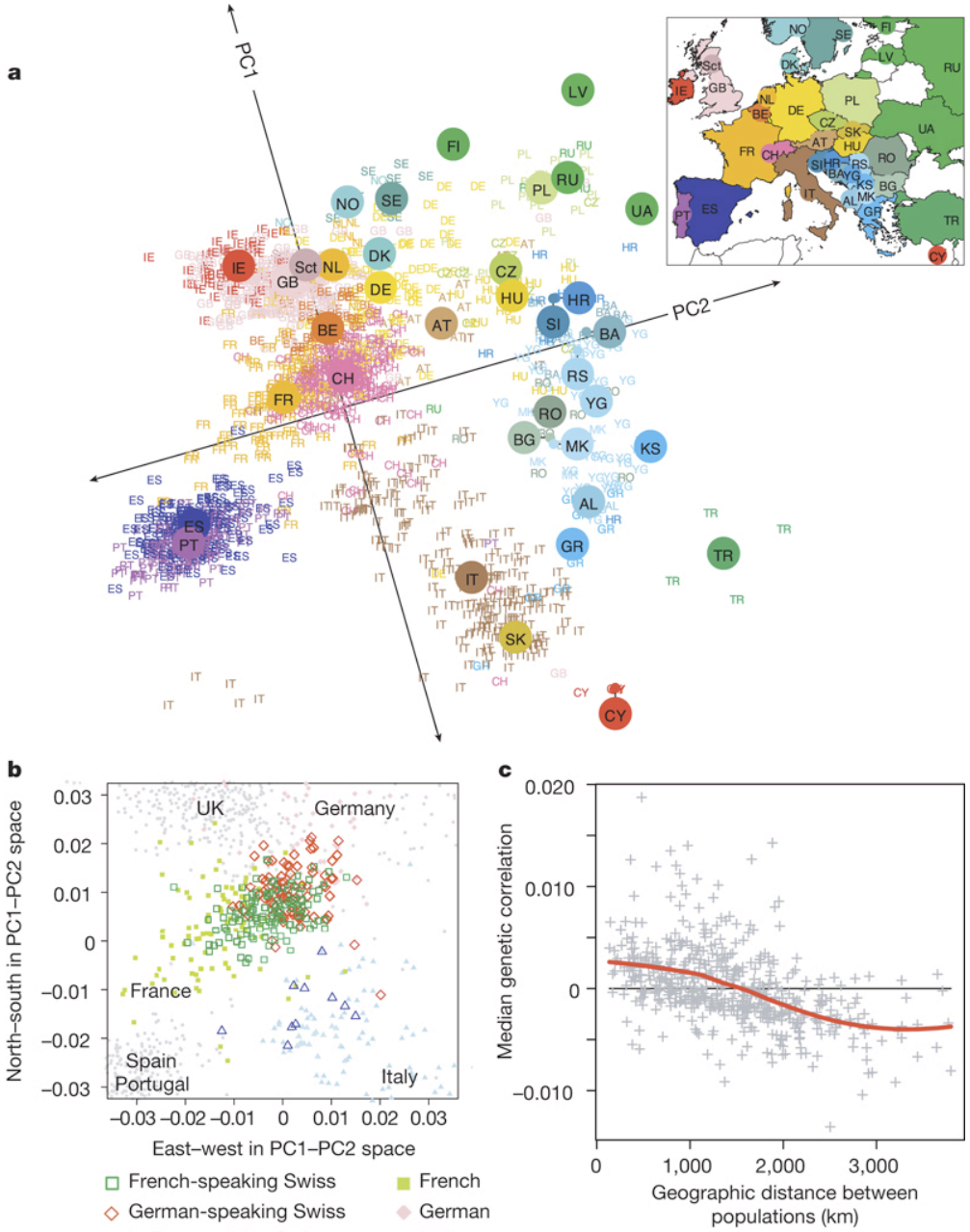
Time

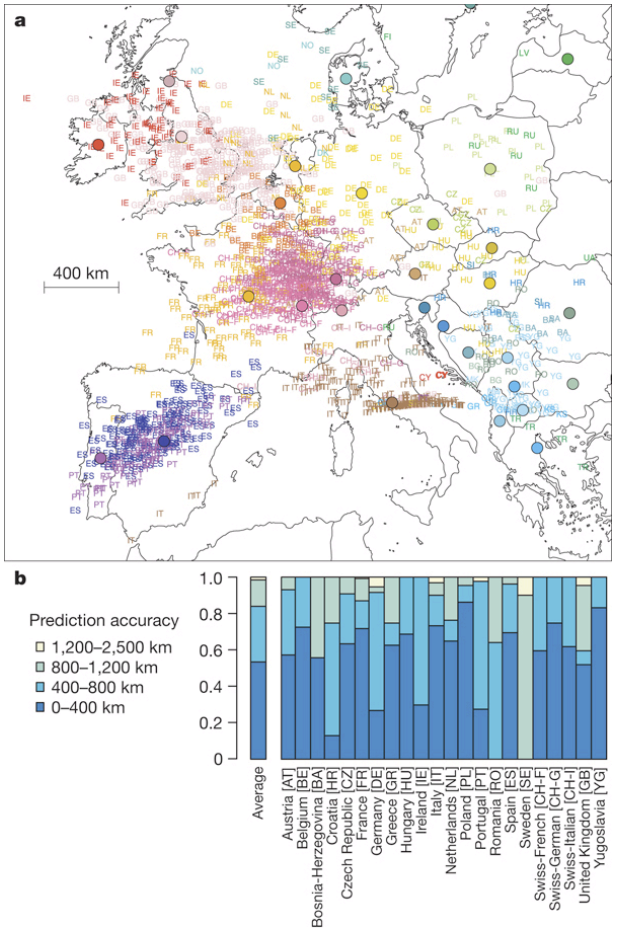
Our genes come from the migration patterns of haplotypes throughout human history ("Population Stratification")





Genotype data can even predict your birthplace





Genes mirror geography within Europe
Novembre *et al.*, 2008

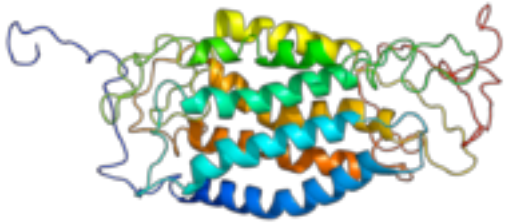
Specific genes can have significant impact

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



Carcinogenesis vol.29 no.4 pp.762–765, 2008
doi:10.1093/carcin/bgn044
Advance Access publication February 14, 2008

Constitutional CHEK2 mutations are associated with a decreased risk of lung and laryngeal cancers

Cezary Cybulski*, Bartłomiej Masojć, Dorota Oszutowska, Ewa Jaworowska[1], Tomasz Grodzki[2], Piotr Waloszczyk[2], Piotr Serwatowski[2], Juliusz Pankowski[2], Tomasz Huzarski, Tomasz Byrski, Bohdan Górski, Anna Jakubowska, Tadeusz Dębniak, Dominika Wokołorczyk, Jacek Gronwald, Czesława Tarnowska[1], Pablo Serrano-Fernández, Jan Lubiński and Steven A.Narod[3]

International Hereditary Cancer Center, Department of Genetics and Pathology, Pomeranian Medical University, ul. Połabska 4, 70-115 Szczecin, Poland, [1]Department of Otolaryngology and Laryngological Oncology, Pomeranian Medical University, ul.Unii Lubelskiej, 71–252 Szczecin, Poland, [2]Lung Diseases Hospital, ul. Sokołowskiego 11, 70–891 Szczecin, Poland and [3]Women's College Research Institute, Toronto, Ontario M5G IN8, Canada

*To whom correspondence should be addressed. Tel: +48 91 466 1532; Fax: +48 91 466 1533;
Email: cezarycy@sci.pam.szczecin.pl

Mutations in the *CHEK2* gene have been associated with increased risks of breast, prostate and colon cancer. In contrast, a previous report suggests that individuals with the I157T missense variant of the *CHEK2* gene might be at decreased risk of lung cancer and upper aero-digestive cancers. To confirm this hypothesis, we genotyped 895 cases of lung cancer, 430 cases of laryngeal cancer and 6391 controls from Poland for four founder alleles in the *CHEK2* gene, each of which has been associated with an increased risk of cancer at several sites. The presence of a *CHEK2* mutation was protective against both lung cancer [odds ratio (OR) = 0.3; 95% confidence interval (CI) 0.2–0.5; $P = 3 \times 10^{-8}$] and laryngeal cancer (OR = 0.6; 95% CI 0.3–0.99; $P = 0.05$). The basis of the protective effect is unknown, but may relate to the reduced viability of lung cancer cells with a *CHEK2* mutation. Lung cancers frequently possess other defects in genes in the DNA damage response pathway (e.g. *p53* mutations) and have a high level of genotoxic DNA damage induced by tobacco smoke. We speculate that lung cancer cells with impaired *CHEK2* function undergo increased rates of cell death.

Introduction

Germ line mutations in *CHEK2* have been associated with a range of cancer types, in particular of the breast and the prostate, but cancers of

of Brennan *et al.* We have extended our series of lung cancer cases from 272 to 895 and our control sample from 4000 to 6391. We have also identified a fourth deleterious *CHEK2* allele (a large deletion of exons 9 and 10). Because smoking is the principal risk factor for lung cancer in Poland and elsewhere, we asked whether the protective effect of *CHEK2* might extend to laryngeal cancer patients as well.

Materials and methods

We studied 895 unselected cases of lung cancer (226 women and 669 men) diagnosed in the Lung Diseases Hospital in Szczecin, Poland, between 2004 and 2006. We also ascertained 430 consecutive, unselected patients with squamous cell carcinoma of the larynx (70 women and 360 men) at Department of Otolaryngology and Laryngological Oncology of the Pomeranian Medical University, Szczecin, Poland, during the period 2001–2004. Patients were recruited from the oncology services of the contributing hospitals and were unselected for age or family history. Patients were approached by a member of the study team during an outpatient visit to the oncology clinic and were asked if they wished to participate. Patient acceptance rates exceeded 80% for both cancer sites. Patients provided written informed consent. A blood sample of 10 cc was then drawn for DNA extraction. Two hundred and seventy-two of the lung cancer patients have been included in our previous study (5). The mean age of diagnosis of the lung cancer patients was 61.4 years (range 29–88 years) and of the laryngeal cancer patients was 58.2 years (range 30–84). Patients completed a questionnaire about their smoking habits at the time of cancer diagnosis. Smoking histories were available for 818 of 895 (91%) lung cancer cases and for 387 of 430 (90%) laryngeal cancer cases. The study was approved by the Ethics Committee of the Pomeranian Medical University in Szczecin.

Unmatched analysis

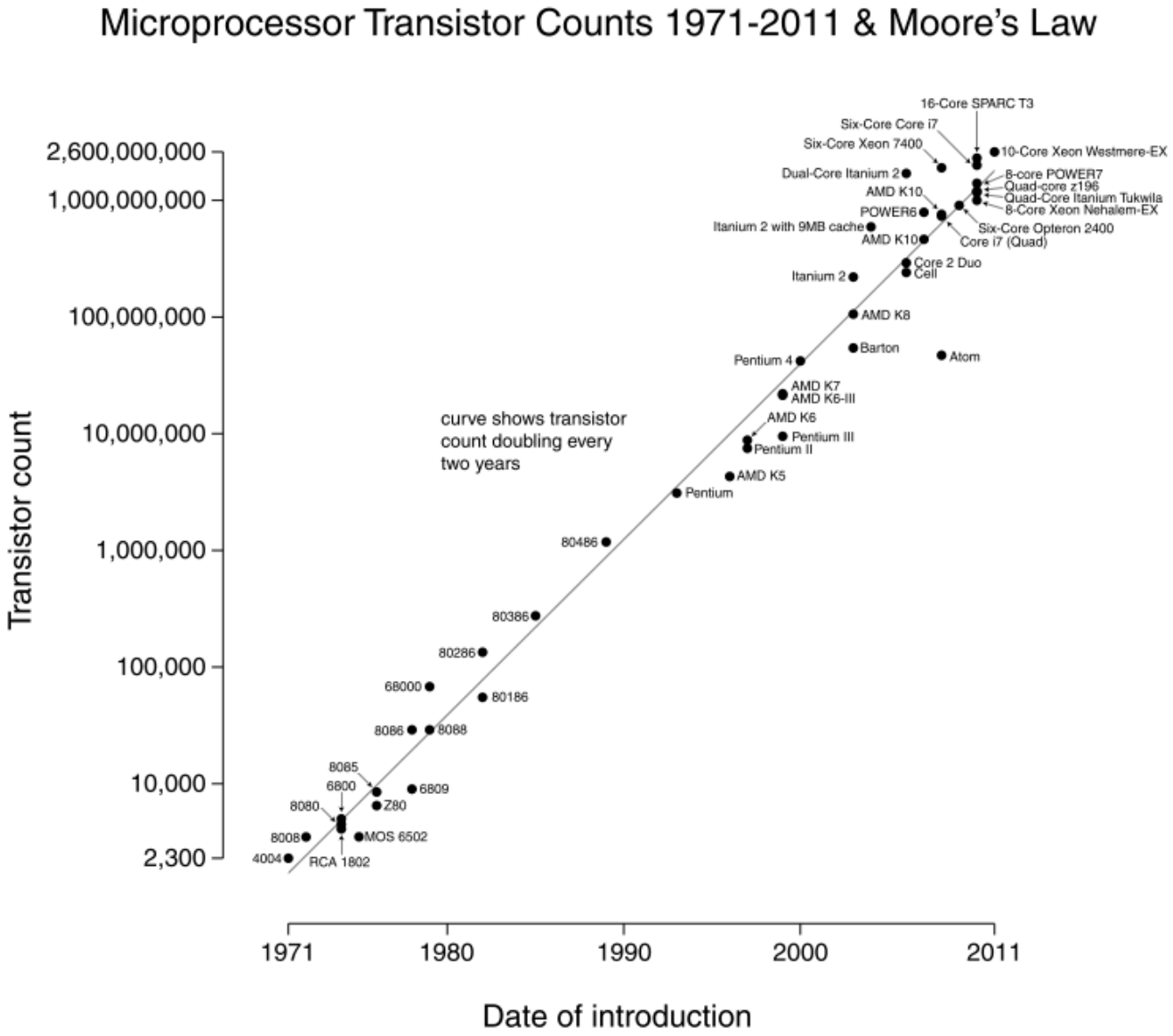
In the unmatched analysis, four non-overlapping control groups were combined in order to maximize the number of controls.

The first control group of 1896 healthy adults, including 1079 women (age range 15–91, mean 58.3) and 817 men (age range 23–90, mean 59.4). These controls were selected at random from the computerized patient lists of five large family practices located in the region of Szczecin. These healthy adults were invited to participate by mail and participated in 2003 and 2004. Participation rates for this group exceeded 70%. During the interview, the goals of the study were explained, informed consent was obtained, genetic counselling was given and a blood sample was taken for DNA analysis. A detailed family history of cancer was taken (first- and second-degree relatives included). Probands were included regardless of their cancer family history status. Individuals affected with any malignancy were excluded from the study.

The second control group consisted of 1417 unselected young adults (705 women and 712 men; age range 18–35, mean 24.3) from Szczecin metropolitan region who submitted a blood sample for paternity testing between 1994 and 2001.

The third control group consisted of 2183 children from nine cities in Poland

The effects from Moore's Law ushered in a whole new era of technology



By Wgsimon

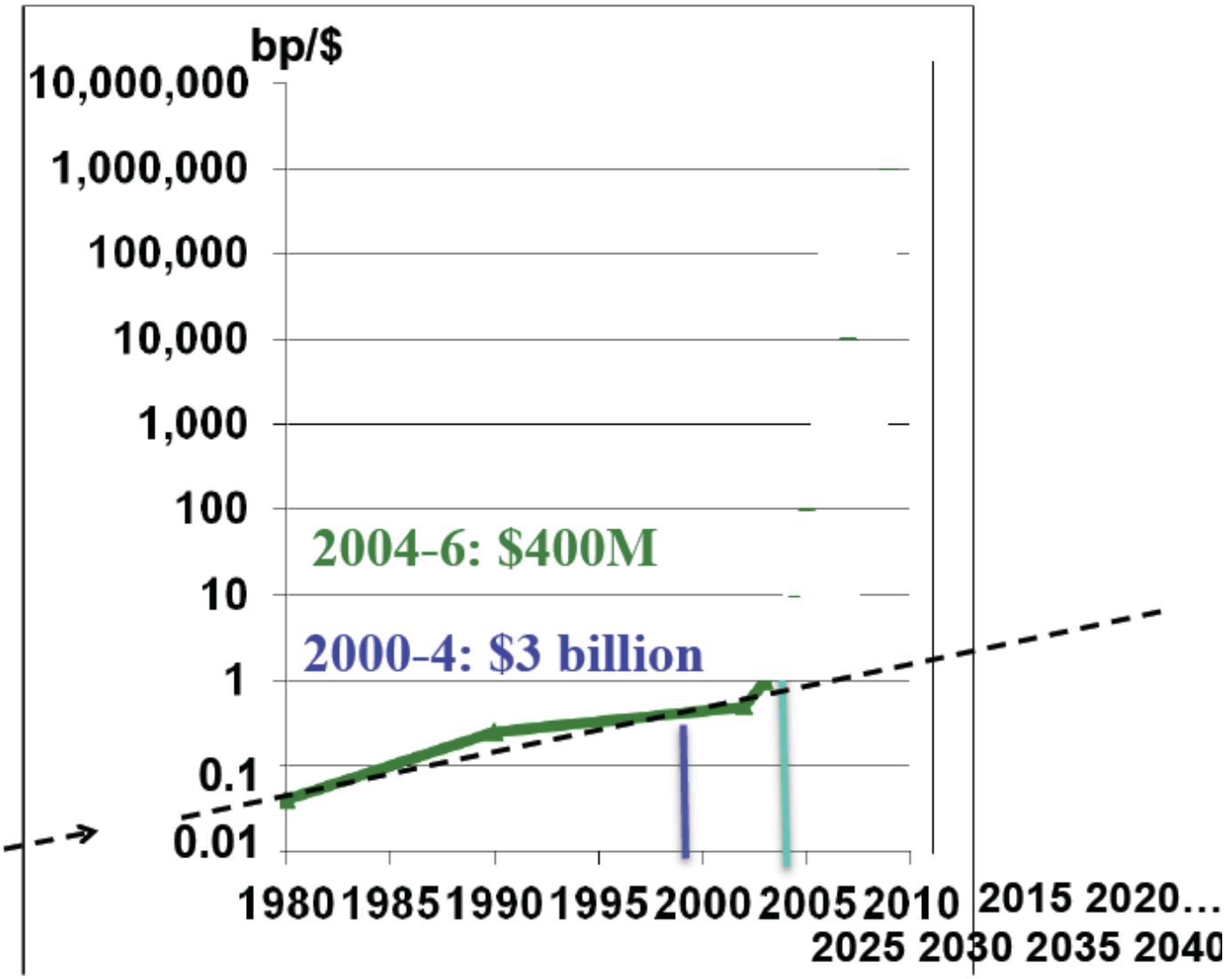
Initially we expected a $1K Genome in 2040

$1000 Genome

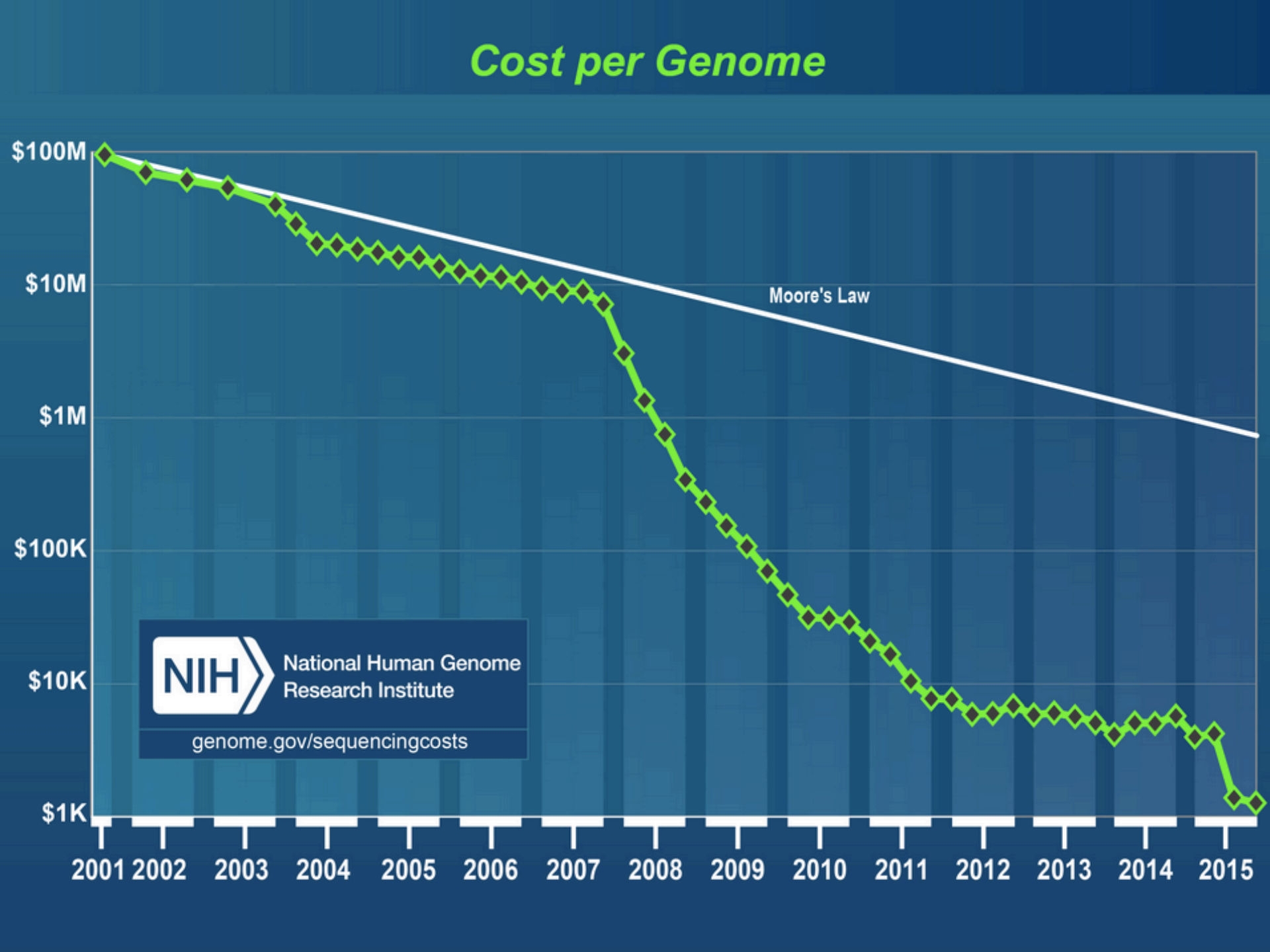
When?

2040

Moore's law 1.5x/yr for electronics



George Church


Cost per Genome
$100M
$10M
$1M
$100K
$10K
$1K
Moore's Law
NIH
National Human Genome Research Institute
genome.gov/sequencingcosts
2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015

nature International weekly journal of science

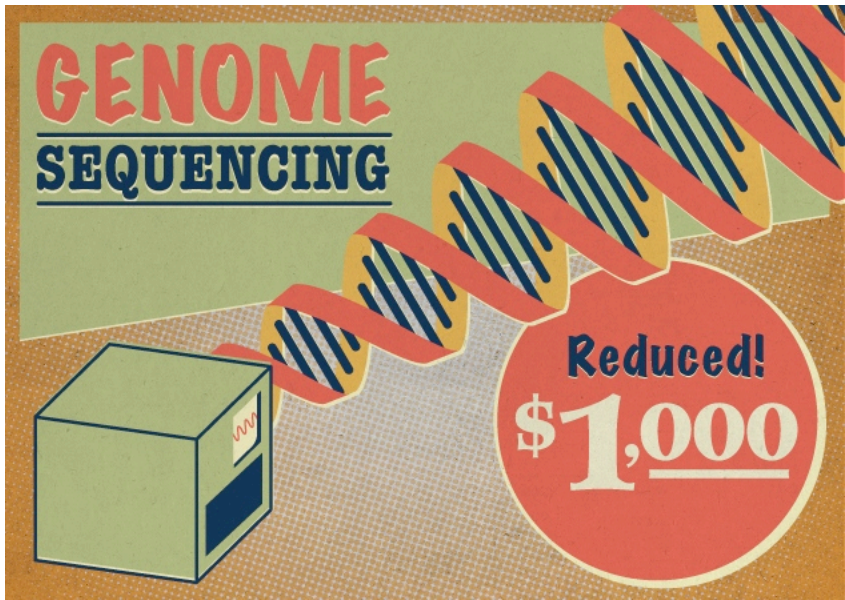
NATURE | NEWS FEATURE

Technology: The $1,000 genome

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

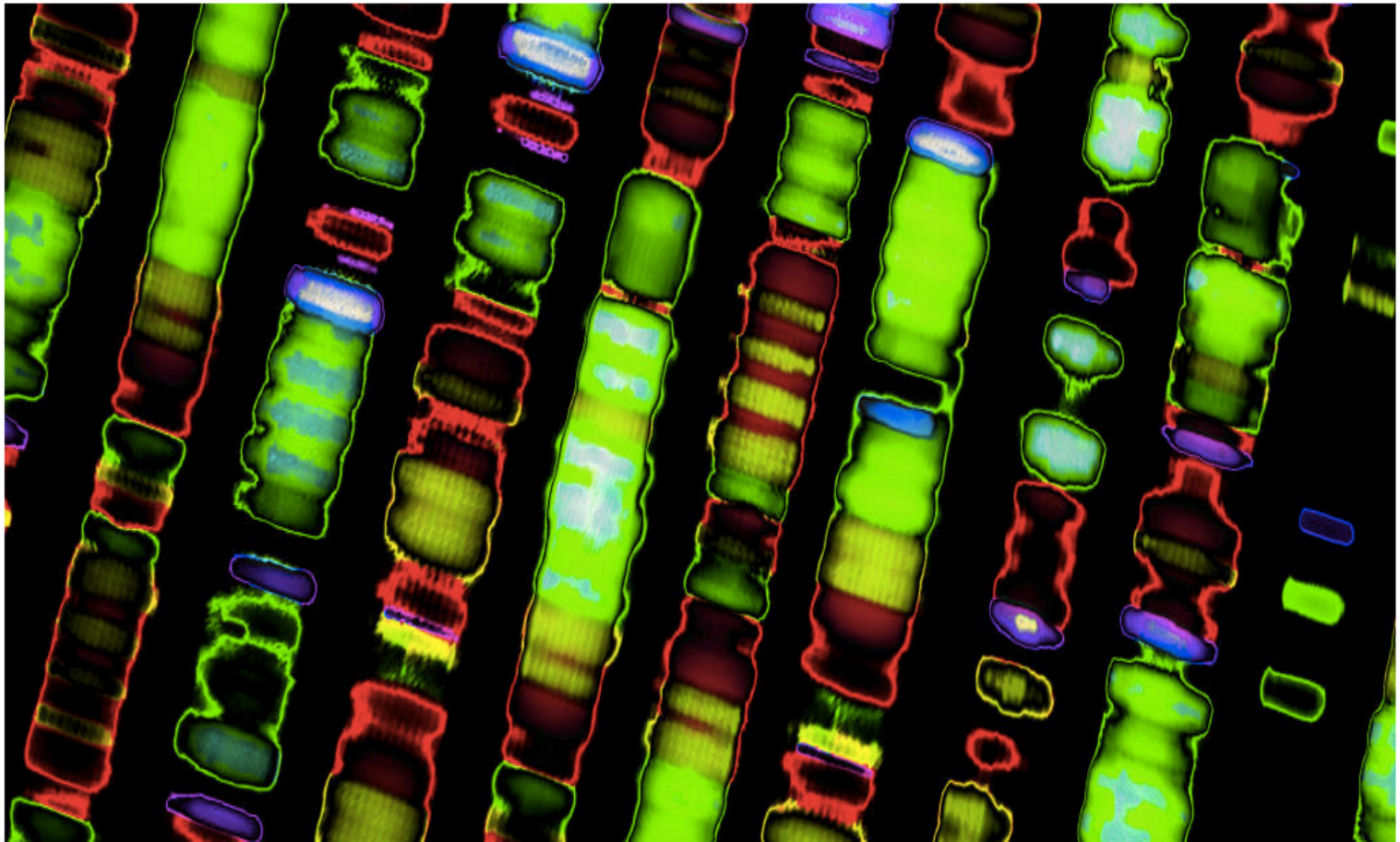
Erika Check Hayden

19 March 2014

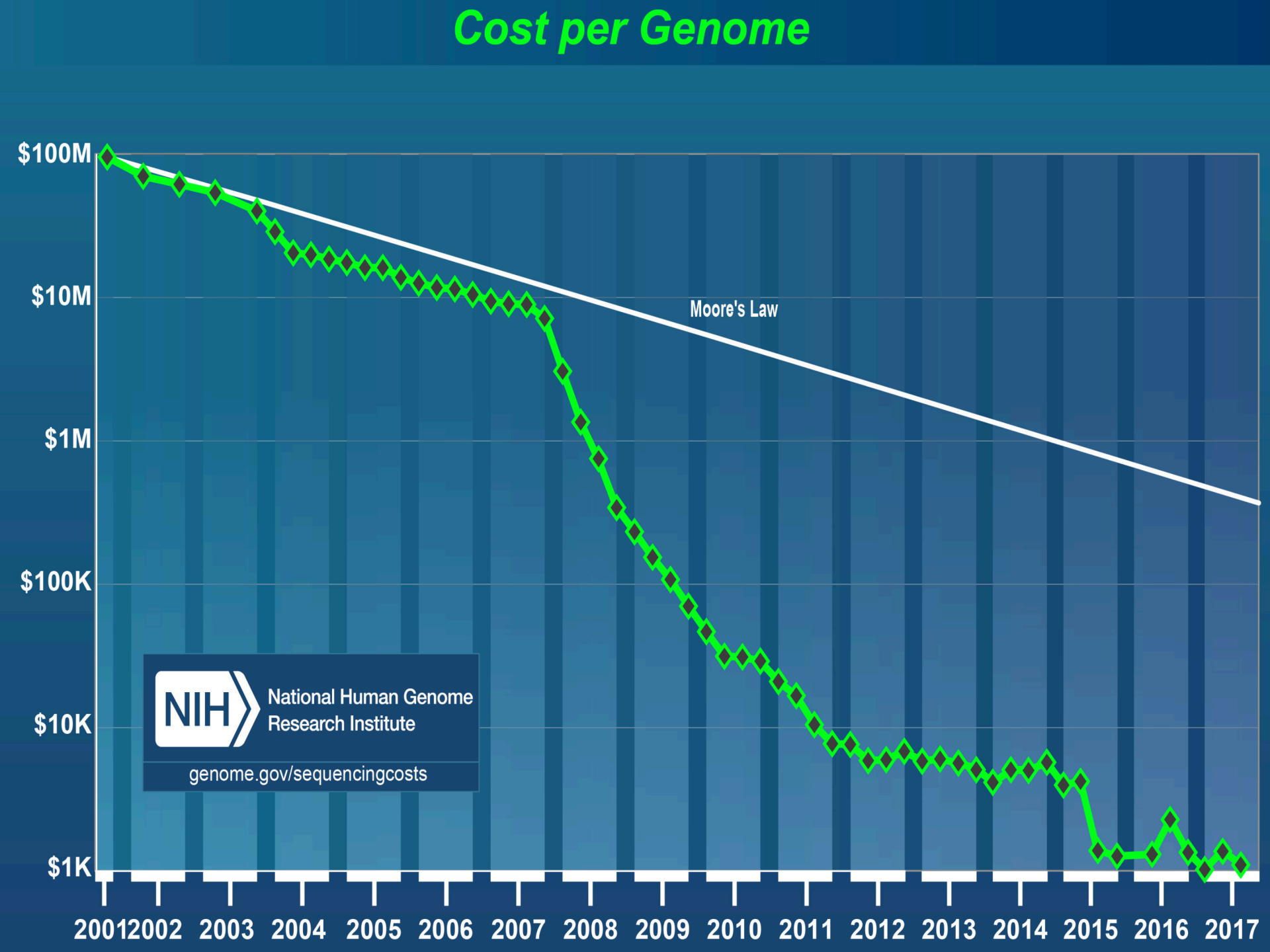


BUSINESS

Illumina says it can deliver a $100 genome — soon



Cost per Genome



The future is already here;
it's just not evenly distributed.

—William Gibson

NGS has also enabled a democratization of the genomes by 2009, making it personal and ubiquitous

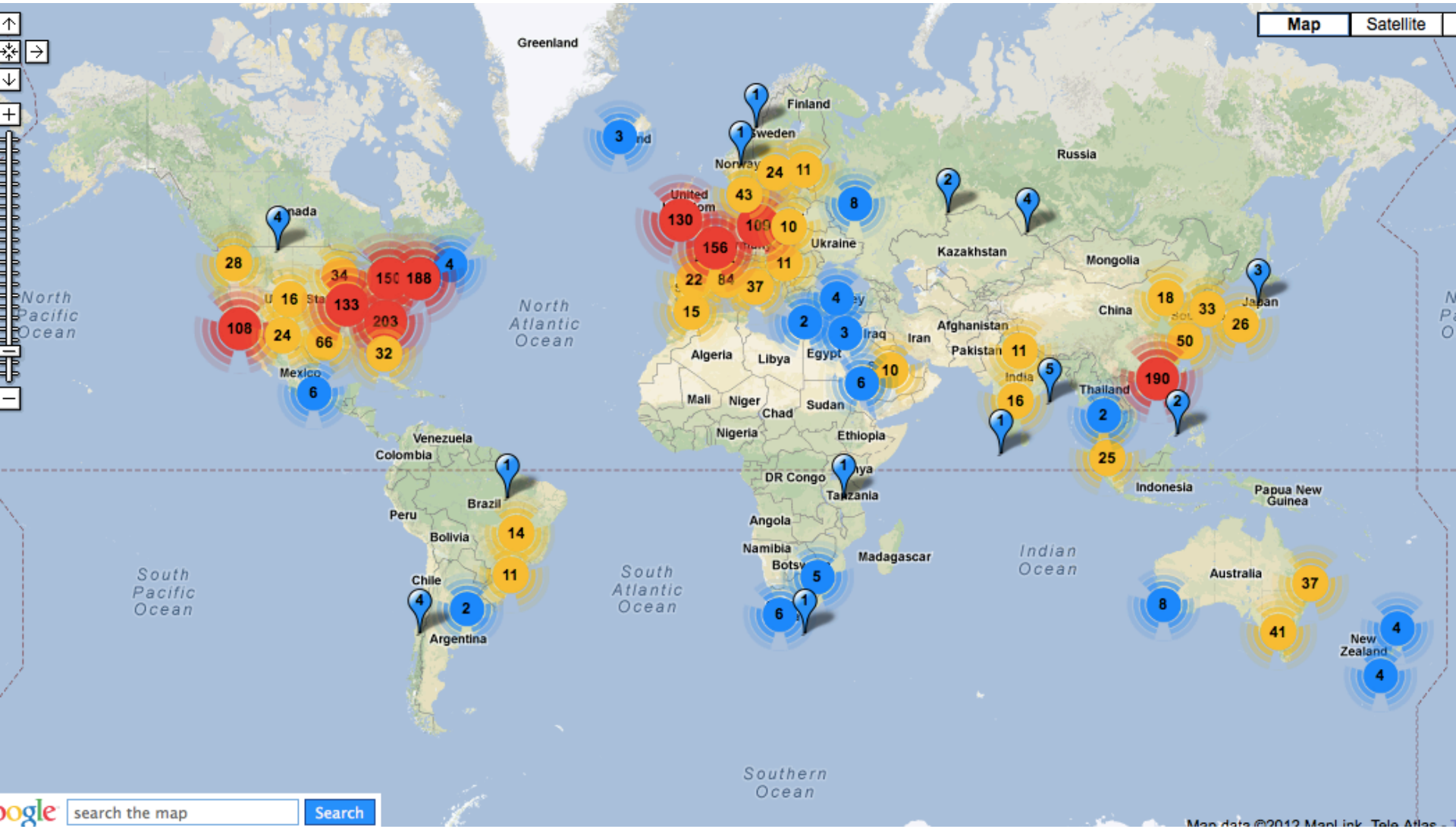
FAQ #3: What is the cost of human genome sequencing?

Pushkarev et al 2009

Year	Estimated cost	Technology	Ref.	Machine runs	Authors	Coverage
2001	$300,000,000	Sanger (ABI)	1	?	251	4
2001	$100,000,000	Sanger (ABI)	2	100,000	274	5
2007	$10,000,000	Sanger (ABI)	3	100,000	31	7
2008	$2,000,000	Roche(454)	4	234	27	7
2008	$1,000,000	Illumina	5	98	48	33
2008	$500,000	Illumina	6	35	77	36
2008	$250,000	Illumina	7	40	196	30
2009	$48,000	Helicos	This work	4	3	28

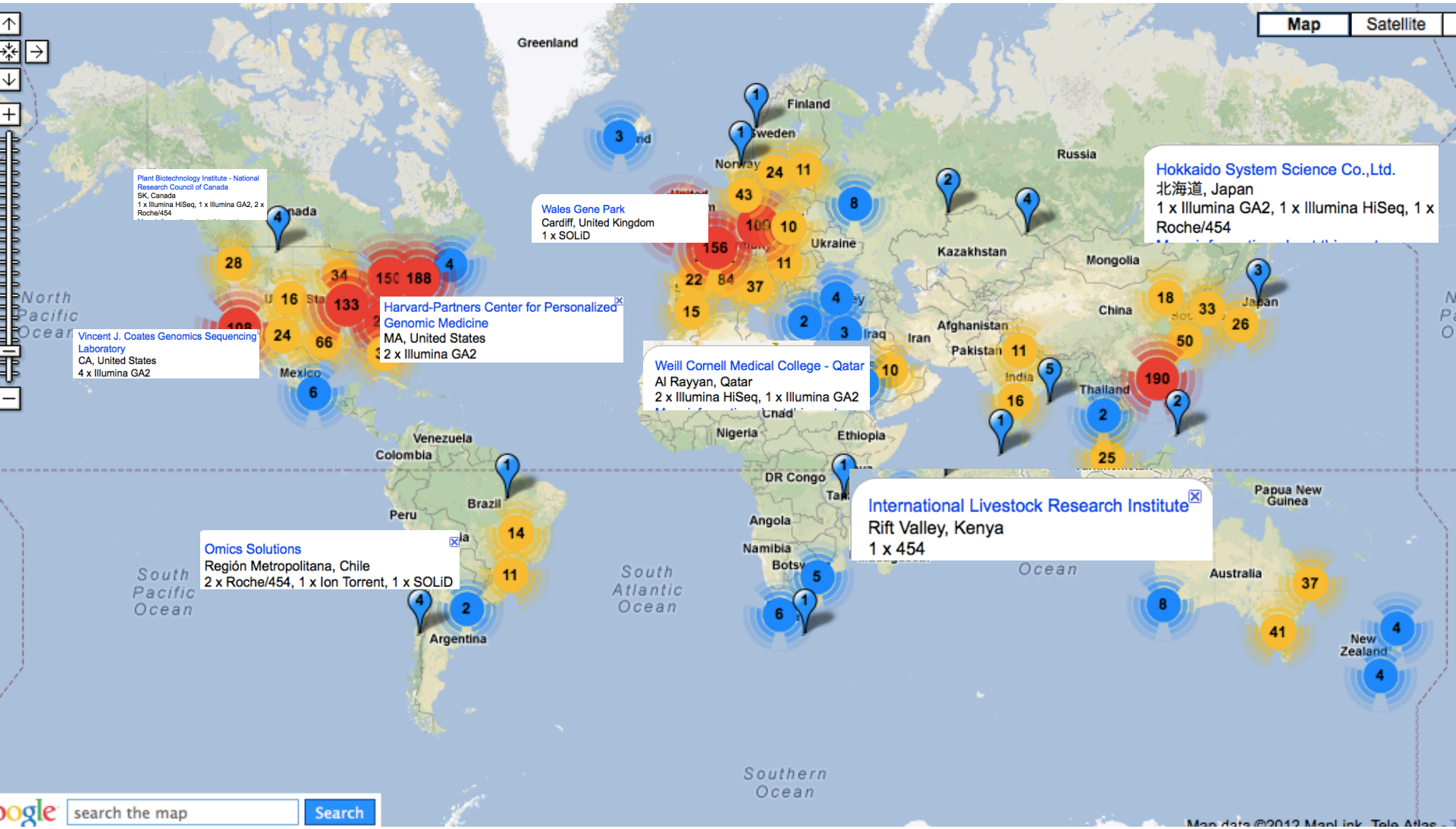
http://arep.med.harvard.edu/gmc/nexgen.html

NGS sites are globally distributed



http://omicsmaps.com/

And cover a wide range of applications in academia, government, and industry



Global Alliance
for Genomics & Health

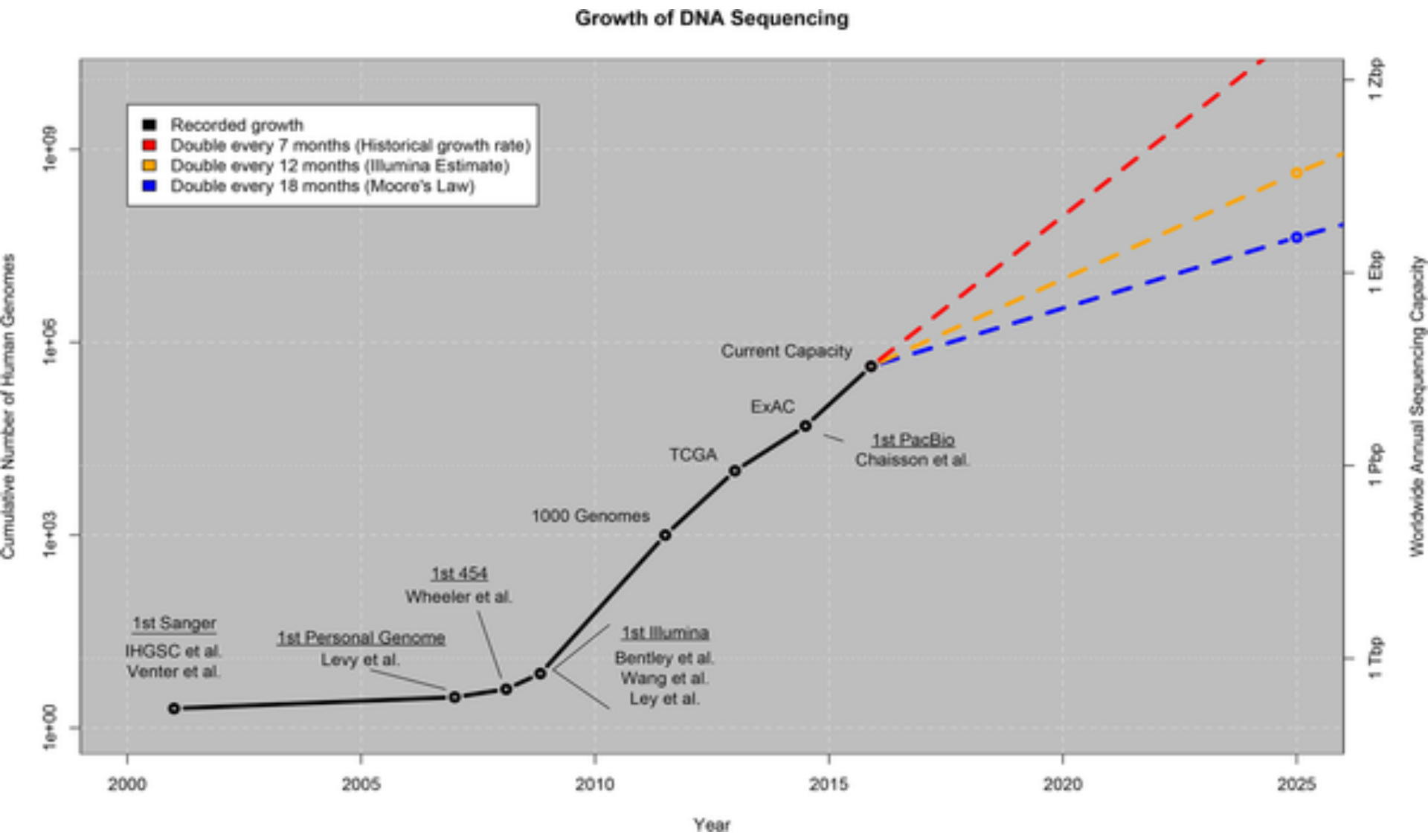
Dear Colleagues,

Last week, GA4GH leaders and key participants came together at the Wellcome Trust Genome Campus in Hinxton, UK to discuss a strategic plan for the organization over the next five years. During the past five months, this group has developed a vision for genomic data sharing in 2022 and identified key milestones necessary to get there. At the meeting, we agreed on a new set of priority deliverables and working processes for GA4GH that we believe will be most likely to help us develop the standards needed for responsible genomic and health-related data sharing.

By 2022, we expect tens of millions of genome sequences to be available from research and healthcare around the globe, providing an incredible opportunity to realize a true international learning health system. To accomplish this, we must ensure that the necessary means to share information across national and institutional boundaries are available to the many disparate organizations and entities producing and analyzing data around the world. Attendees of the meeting broadly agreed that GA4GH is uniquely capable of delivering on this need.


SAY BIG DATA
ONE MORE TIME
memegenerator.net

Big Data: Genomical or Astronomical?

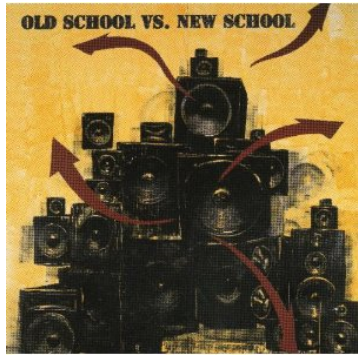


Stephens ZD, Lee SY, Faghri F, Campbell RH, Zhai C, et al. (2015)
http://journals.plos.org/plosbiology/article?id=info:doi/10.1371/journal.pbio.1002195

PLOS | BIOLOGY

Sequencing Technologies

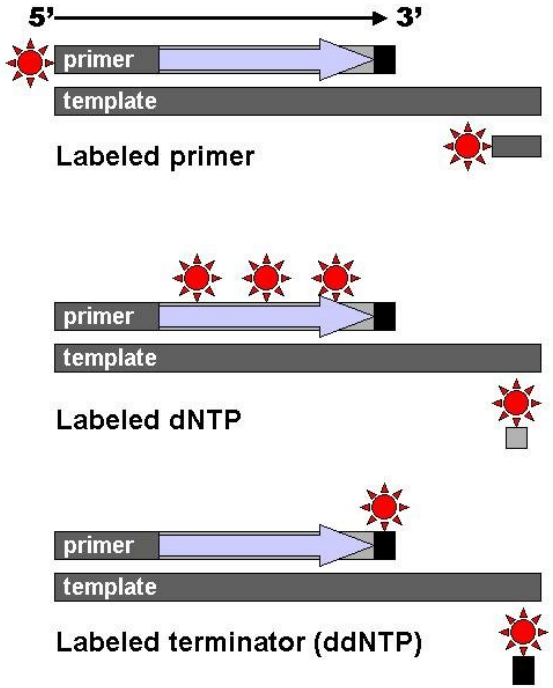
1. "Old School" dye-terminator sequencing (Sanger). 300-1000bp

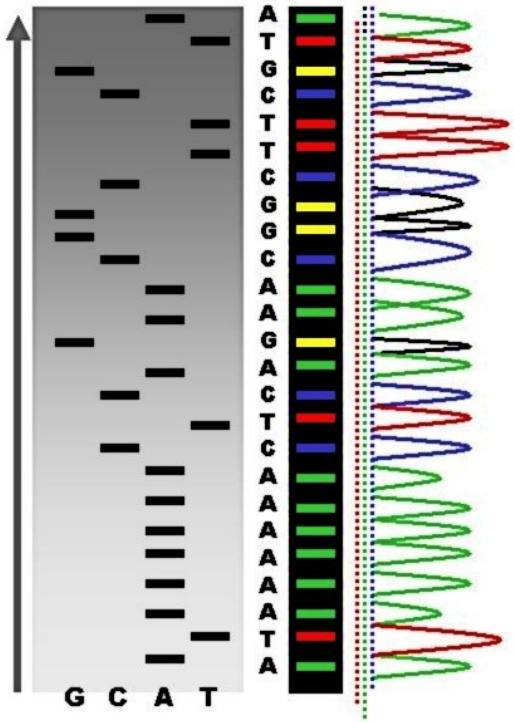


2. "New School" methods
   a. Emulsion PCR Pyrosequencing
   b. Solid-phase amplification sequencing by synthesis (clonal or single molecule)
   c. Sequencing by ligation
   d. Single-molecule, real-time (SMRT) sequencing
   e. Electrical sequencing

Sequencing Technologies

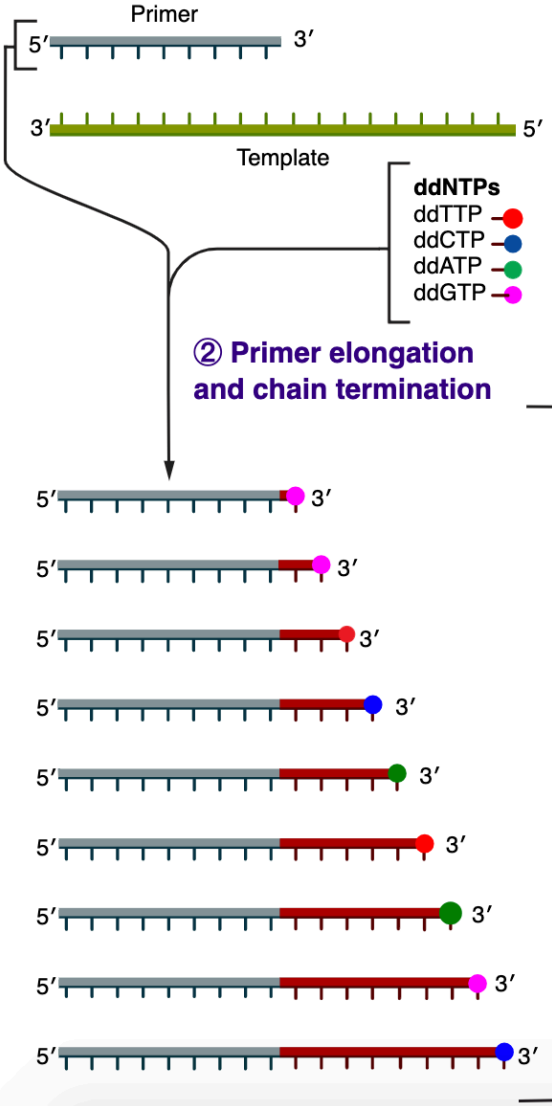
1. "Old School" dye-terminator sequencing (Sanger). 300-1000bp

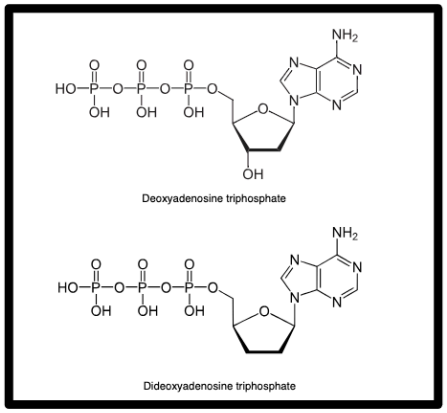




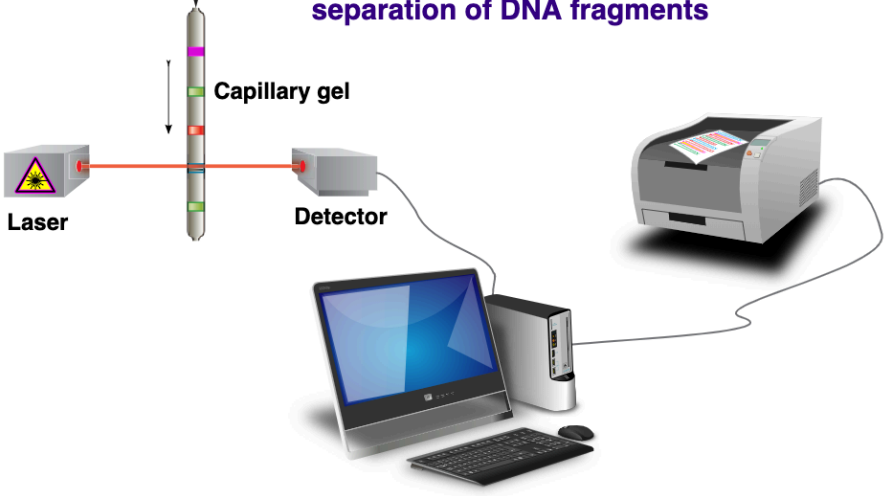
① Reaction mixture

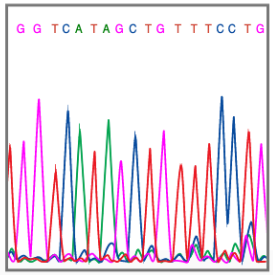
- Primer and DNA template
- DNA polymerase
- ddNTPs with flourochromes
- dNTPs (dATP, dCTP, dGTP, and dTTP)





③ Capillary gel electrophoresis separation of DNA fragments





Chromatograph

④ Laser detection of flourochromes and computational sequence analysis

Estevezj

By 2009, many options emerged

Platform	Library/ template preparation	NGS chemistry	Read length (bases)	Run time (days)	Gb per run	Machine cost (US$)	Pros	Cons	Biological applications	Refs
Roche/454's GS FLX Titanium	Frag, MP/ emPCR	PS	330*	0.35	0.45	500,000	Longer reads improve mapping in repetitive regions; fast run times	High reagent cost; high error rates in homo-polymer repeats	Bacterial and insect genome *de novo* assemblies; medium scale (<3 Mb) exome capture; 16S in metagenomics	D. Muzny, pers. comm.
Illumina/ Solexa's GA_{II}	Frag, MP/ solid-phase	RTs	75 or 100	4‡, 9§	18‡, 35§	540,000	Currently the most widely used platform in the field	Low multiplexing capability of samples	Variant discovery by whole-genome resequencing or whole-exome capture; gene discovery in metagenomics	D. Muzny, pers. comm.
Life/APG's SOLiD 3	Frag, MP/ emPCR	Cleavable probe SBL	50	7‡, 14§	30‡, 50§	595,000	Two-base encoding provides inherent error correction	Long run times	Variant discovery by whole-genome resequencing or whole-exome capture; gene discovery in metagenomics	D. Muzny, pers. comm.
Polonator G.007	MP only/ emPCR	Non-cleavable probe SBL	26	5§	12§	170,000	Least expensive platform; open source to adapt alternative NGS chemistries	Users are required to maintain and quality control reagents; shortest NGS read lengths	Bacterial genome resequencing for variant discovery	J. Edwards, pers. comm.
Helicos BioSciences HeliScope	Frag, MP/ single molecule	RTs	32*	8‡	37‡	999,000	Non-bias representation of templates for genome and seq-based applications	High error rates compared with other reversible terminator chemistries	Seq-based methods	91
Pacific Biosciences (target release: 2010)	Frag only/ single molecule	Real-time	964*	N/A	N/A	N/A	Has the greatest potential for reads exceeding 1 kb	Highest error rates compared with other NGS chemistries	Full-length transcriptome sequencing; complements other resequencing efforts in discovering large structural variants and haplotype blocks	S. Turner, pers. comm.

Michael Metzker, 2010

Then, by 2014, an ecosystem of options erupted

Table 1: Types of High-Throughput Sequencing Technologies

Optical Sequencing					
Platform	**Instrument**	**Template Preparation**	**Chemistry**	**Avearge Length**	**Longest Read**
Illumina	HiSeq2500	BridgePCR/cluster	Rev. Term., SBS	100	150
Illumina	HiSeq2000	BridgePCR/cluster	Rev. Term., SBS	100	150
Illumina	MiSeq	BridgePCR/cluster	Rev. Term., SBS	250	300
GnuBio	GnuBio	emPCR	Hyb-Assist Sequencing	1000*	64,000*
Life Technologies	SOLiD 5500	emPCR	Seq. by Lig.	75	100
LaserGen	LaserGen	emPCR	Rev. Term., SBS	25*	100*
Pacific Biosciences	RS	Polymerase Binding	Real-time	1800	15,000
454	Titanium	emPCR	PyroSequencing	650	1100
454	Junior	emPCR	PyroSequencing	400	650
Helicos	Heliscope	adaptor ligation	Rev. Term., SBS	35	57
Intelligent BioSystems	MAX-Seq	Rolony amplification	Two-Step SBS (label/unlabell)	2x100	300
Intelligent BioSystems	MINI-20	Rolony amplification	Two-Step SBS (label/unlabell)	2x100	300
ZS Genetics	N/A	Atomic Lableing	Electron Microscope	N/A	N/A
Halcyon Molecular	N/A	N/A	Direct Observation of DNA	N/A	N/A

Electical Sequencing					
Platform	**Instrument**	**Template Preparation**	**Chemistry**	**Avearge Length**	**Longest Read**
IBM DNA Transistor	N/A	none	Microchip Nanopore	N/A	N/A
NABsys	N/A	none	Nanochannel	N/A	N/A
Bionanogenomics	N/A	anneal 7mers	Nanochannel	N/A	N/A
Life Technologies	PGM	emPCR	Semi-conductor	150	300
Life Technologies	Proton	emPCR	Semi-conductor	120	240
Life Technologies	Proton 2	emPCR	Semi-conductor	400*	800*
Genia	N/A	none	Protein nanopore (a-hemalysin)	N/A	N/A
Oxford Nanopore	MinION	none	Protein Nanopore	10,000	10,000*
Oxford Nanopore	GridION 2K	none	Protein Nanopore	10,000	500,000*
Oxford Nanopore	GridION 8K	none	Protein Nanopore	10,000	500,000*

*Values are estimates from companies that have not yet released actual data

Mason, Porter, Smith, 2014

Coming of age: ten years of next-generation sequencing technologies

Sara Goodwin[1], John D. McPherson[2] and W. Richard McCombie[1]

Abstract | Since the completion of the human genome project in 2003, extraordinary progress has been made in genome sequencing technologies, which has led to a decreased cost per megabase and an increase in the number and diversity of sequenced genomes. An astonishing complexity of genome architecture has been revealed, bringing these sequencing technologies to even greater advancements. Some approaches maximize the number of bases sequenced in the least amount of time, generating a wealth of data that can be used to understand increasingly complex phenotypes. Alternatively, other approaches now aim to sequence longer contiguous pieces of DNA, which are essential for resolving structurally complex regions. These and other strategies are providing researchers and clinicians a variety of tools to probe genomes in greater depth, leading to an enhanced understanding of how genome sequence variants underlie phenotype and disease.

Company	Release	Instrument	Notes	Instrument	Run Time (h)	wells / pores / clusters / channels	active wells / pores / cluster	PassFilter Reads ---------- ----- Active Pores	Output / Sequence Site or Pore	Mean Read Length	Mb / Run	Gb / Run	Raw Cost / Run ($)	Reagent Cost /Gb ($)	Cost / 30X Human Genome ($)	# Genomes per Run (30X
Illumina	Q1 2019	NovaSeq6000	6Tb run (dual FC S4)	$950,000	72	20,000,000,000	100%	20,000,000,000	1.00	300	6,000,000	6,000	$ 61,000.00	$ 10.17	$ 915	64.52
Illumina	Q1 2017	NovaSeq6000	6Tb run (dual FC S4)	$950,000	48	20,000,000,000	100%	20,000,000,000	1.00	300	6,000,000	6,000	$ 61,000.00	$ 10.17	$ 915	64.52
Illumina	Q1 2017	NovaSeq5000	2Tb run (dual FC S2)	$850,000	60	6,800,000,000	95%	6,460,000,000	1.00	300	1,938,000	1,938	$ 29,900.00	$ 15.43	$ 1,389	20.84
Illumina	Q1 2014	X10	1Tb run	$1,000,000	72	6,200,000,000	95%	5,890,000,000	1.00	302	1,778,780	1,779	$ 12,750.00	$ 7.17	$ 645	19.13
Illumina	Q1 2015	X5	1Tb run	$1,000,000	72	6,200,000,000	95%	5,890,000,000	1.00	302	1,778,780	1,779	$ 19,200.00	$ 10.79	$ 971	19.13
Illumina	Q1 2015	HiSeq4000	Regular (v4, 1TB)	$900,000	144	5,200,000,000	97%	5,044,000,000	1.00	300	1,513,200	1,513	$ 29,900.00	$ 19.76	$ 1,778	16.27
Illumina	Q1 2012	HiSeq2500	Regular	$740,000	__	4,000,000,000	95%	3,800,000,000	1.00	250	950,000	950	$ 29,900.00	$ 31.47	$ 2,833	10.22
Illumina	Q1 2012	HiSeq2500	RapidGenome	$740,000	__	600,000,000	95%	570,000,000	1.00	300	171,000	171	$ 6,972.00	$ 40.77	$ 3,669	1.84
Illumina	Q1 2015	NextSeq	2x150bp run	$225,000	30	520,000,000	95%	494,000,000	1.00	300	148,200	148	$ 4,000.00	$ 26.99	$ 2,429	1.59
Illumina	Q1 2015	NextSeq	2x75bp run	$225,000	30	520,000,000	95%	494,000,000	1.00	150	74,100	74	$ 2,500.00	$ 33.74	$ 3,036	0.80
Illumina	Q1 2015	NextSeq	1x75bp run	$225,000	30	520,000,000	95%	494,000,000	1.00	75	37,050	37	$ 1,300.00	$ 35.09	$ 3,158	0.40
Illumina	Q1 2013	MiSeq	v2	$125,000	24	25,000,000	95%	23,750,000	1.00	500	11,875	12	$ 1,000.00	$ 84.21	$ 7,579	0.13
Illumina	Q1 2016	MiniSeq	v1	$49,500	24	26,000,000	95%	24,700,000	1.00	300	7,410	7	$ 1,000.00	$ 134.95	$ 12,146	0.08
Illumina	Q3 2017	Firefly	v1	$19,900	4	5,000,000	95%	4,750,000	1.00	300	1,425	1	$ 400.00	$ 280.70	$ 25,263	0.02
Genia	2019?	UNK	v1	unk	48	8,000	50%	4,000	500	5,000	10,000	7	$ 1,000.00	$ 100.00	$ 9,000	
BGI	Q1 2018	MGISEQ-2000	2x100	UNK	24					200	600,000.00	600	$ 5,000.00	$ 8.33		
MGI	Q1 2016	BGISEQ-500	2x100 (1x50, 1x300)	UNK	96	2,000,000,000	95%	1,900,000,000	1.00	200	380,000	380	$ 2,500.00	$ 6.58	$ 592	4.09
BGI	Q1 2018	MGISEQ-200	2x100	$150,000	48					200		60				
MGI	Q12019	MGISEQ-T7	2x150	UNK	24	5,000,000,000	99%	4,950,000,000	1.00	300	1,485,000	1,485	$ 10,000.00	$ 6.73	$ 606	15.97
BGI	Q1 2018	MGIFLP	2x100?	UNK												
LifeTech	Q1 1995	3730xl	capillary/Sanger	$300,000	2	96	100%	96	1.00	750	0.07	0.000072	$ 90.00	$ 1,250.00	$ 112,500	0.00
LifeTech	Q1 2010	PGM	318chip	$75,000	2	11,000,000	50%	5,500,000	1.00	400	2,200	2	$ 1,100.00	$ 500.00	$ 45,000	0.02
LifeTech	Q3 2012	Proton	Proton 1	$225,000	2	100,000,000	65%	65,000,000	1.00	120	7,800	8	$ 1,525.00	$ 195.51	$ 17,596	0.08
LifeTech	Q3 2015	S5 / S5XL	520 chip	$50,000	2	5,000,000	95%	4,750,000	1.00	400	1,900	2	$ 300.00	$ 157.89	$ 14,211	0.02
LifeTech	Q3 2015	S5 / S5XL	530 chip	$50,000	2	20,000,000	95%	19,000,000	1.00	400	7,600	8	$ 300.00	$ 39.47	$ 3,553	0.08
LifeTech	Q3 2015	S5 / S5XL	540 chip	$50,000	2	80,000,000	95%	76,000,000	1.00	200	15,200	15	$ 300.00	$ 19.74	$ 1,776	0.16
LifeTech	Q1 2015	Proton	Proton 2	$225,000	6	300,000,000	80%	240,000,000	1.00	120	28,800	29	$ 1,000.00	$ 34.72	$ 3,125	0.31
Oxford Nanopore	Q1 2019	Flongle	126	$1,000	6	126	75%	95	778	15000	1,102	1.10	$ 500.00	$ 453.62	$ 40,826	0.01
Oxford Nanopore	Q2 2015	MinION	Min500	$500	6	512	75%	384	778	15000	4,479	4.48	$ 500.00	$ 111.63	$ 10,047	0.05
Oxford Nanopore	Q2 2017	GridIONx5	5 pores	$125,000		2,560	75%	1,920	3,888	15,000	111,974	111.97	2,500	$ 22.33	$ 2,009	
Oxford Nanopore	Q2 2017	PrOmethION	100,000 pores	$75,000	6	98,304	60%	58,982	6,221	10000	3,669,177	3,669.18	$ 36,000.00	$ 9.81	$ 883	39.45
PacBio	Q1 2014	RSII	C2XL (120 min)	$700,000	6	150,000	45%	67,500	1.00	11000	743	0.74	$ 150.00	$ 202.02	$ 18,182	0.01
PacBio	Q1 2016	Sequel	C2XL (360 min)	$350,000	8	1,000,000	60%	600,000	1.00	11000	6,600	6.60	$ 700.00	$ 106.06	$ 9,545	0.07
PacBio	Q2 2019	Sequel	v3 (P6-c4)	$350,000	8	8,000,000	100%	8,000,000	0.90	10000	72,000	72.00	$ 750.00	$ 10.42	$ 938	0.77
QIAGEN	Q1 2015	GeneReader	150 bp run	$225,000	33	16,000,000	95%	15,200,000	1.00	150	2,280	2.28	$ 500.00	$ 219.30	$ 19,737	0.02
Roche	Q1 2007	454	FLX	$100,000	8	1,600,000	65%	1,040,000	1.00	500	520	0.52	$ 1,200.00	$ 2,307.69	$ 207,692	0.01

The $1000 genome is here!

- More often ~$1100 per genome. Coming down.
- Exome sequencing costs also are dropping
- Certain platforms are better suited for certain tasks:
  - Counting applications (ChIP-Seq, RNA-Seq) need more reads
  - *De novo* assembly work needs longer reads
  - Whole genome re-sequencing requires lower errors rate and high processivity

Genomics england

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

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

https://www.genomicsengland.co.uk/

UK Plans to Sequence 5M Genomes in Next Five Years

Oct 03, 2018 | staff reporter

Save for later

NEW YORK (GenomeWeb) – The UK's Department of Health and Social Care has announced its plans to sequence five million genomes in the UK over the next five years.

From 2019 onward, the National Health Service will offer "all seriously ill children" whole-genome sequencing as part of their standard care, as well as to adults with certain rare diseases or "hard-to-treat" illnesses, the UK government said. In addition, patients will be asked to give consent for their genomic data to be analyzed in an effort to develop new tests and treatments for cancer and rare diseases.

The NHS Genomic Medicine Service will also expand on existing projects, such as the 100,000 Genomics Project. The NHS and the charity UK Biobank will sequence 1 million whole genomics within five years. The multi-year project will bring together industry experts such as UK Research and Innovation, the NHS, and other partners, the government noted

https://www.genomeweb.com/sequencing/uk-plans-sequence-5m-genomes-next-five-year

October 12, 2016

PMI Cohort Program announces new name: the All of Us Research Program

The Precision Medicine Initiative® (PMI) Cohort Program will now be called the *All of Us* Research Program and will be the largest health and medical research program on precision medicine. A set of core values is guiding its development and implementation:

- Participation is open to all.
- Participants reflect the rich diversity of the U.S.
- Participants are partners.

1 million U.S. Veterans too!

U.S. Department of Veterans Affairs

Search ORD

VA SITE MAP [A-Z]

Health | Benefits | Burials & Memorials | About VA | Resources | News Room | Locations | Contact Us

VA » Health Care » Office of R&D » Mvp » Million Veteran Program (MVP)

Print

Office of Research & Development

- ORD Home
- About Us
- Services
- Programs
  - See All Programs
  - Animal Research
  - Biosafety & Biosecurity
  - Cooperative Studies Program (CSP)
  - Health Disparities & Minority Health
  - Million Veteran Program (MVP)

Million Veteran Program (MVP)

MVP is a national, **voluntary** research program funded entirely by the Department of Veterans Affairs Office of Research & Development. The goal of MVP is to partner with Veterans receiving their care in the VA Healthcare System to study how genes affect health. To do this, MVP will build one of the world's largest medical databases by safely collecting blood samples and health information from one million Veteran volunteers. Data collected from MVP will be stored anonymously for research on diseases like diabetes and cancer, and military-related illnesses, such as post-traumatic stress disorder. Learn more.

Frequently Asked Questions

- How do I participate?
- Do I need to schedule an appointment to participate?

Text size: + -

CONTACT MVP

Contact the MVP Information Center toll-free at:

866-441-6075

INFORMED CONSENT

A lot of genomic and medical data coming

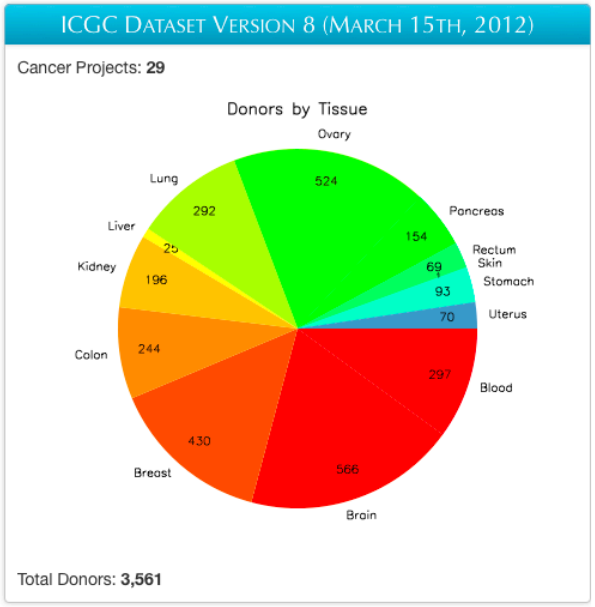
Announcements of Large Genome Consortia:

- AllOfUs – 1M U.S. Patients with medical data
- Netherlands GoNL– 250trios – preclinical (http://www.nlgenome.nl/)
- Faroer islands 100k –pre-clinical
- Qatar 300k – pre-clinical
- Iceland 2.5k – pre-clinical
- UK 100k and now 5M – clinical
- Genomics Medicine Ireland (GMI) with AbbVie - 100K
- Finland, number unknown – clinical (https://www.fimm.fi/en/research/grand-challenge-programs/finnish-genome-sequencing-and-preventive-health-care)
- Poland 100K
- Swiss Genome 100K
- Geisinger Health 100K (with Regeneron)
- Astrozenica (2M with HLI)
- 1 million U.S. Veterans Project
- Newfoundland 100K
- Estonia 1.3M

Large impact for normal genomes and diseases, especially cancer

1000 Genomes
A Deep Catalog of Human Genetic Variation

The Cancer Genome Atlas
Data Portal
Understanding genomics to improve cancer care



International Cancer Genome Consortium

ICGC Goal: To obtain a comprehensive description of genomic, epigenomic, and transcriptomic (GET) changes in 50 different tumor types and/or subtypes which are of clinical and societal importance across the globe.

Memorial Sloan-Kettering Cancer Center

cBio Cancer Genomics Portal

Visualize, analyze, discover.

HOME | TUTORIALS | NEWS | FAQ | DATA SETS | ABOUT | WEB API | R/MATLAB | NETWORKS

The cBio Cancer Genomics Portal provides **visualization**, **analysis** and **download** of large-scale **cancer genomics** data sets.

Please adhere to the TCGA publication guidelines when using any TCGA data in your

ltered in 66 (48%) of cases.

Total altered 66 cases with alter

Data Sets

The Portal contains data for **10410 tumor samples from 31 cancer studies.** [Details.]

National Cancer Institute | National Human Genome Research Institute

TCGA Home | Contact Us | For the Media

The Cancer Genome Atlas Data Portal

Understanding genomics to improve cancer care

Home | Query the Data | Download Data | Tools | About the Data | Publication Guidelines

Home

TCGA Data Portal Overview

We provide 3 ways to download data: The Cancer Genome Atlas (TCGA) Data Portal provides a platform for researchers to search, download, and analyze data sets generated by TCGA. It contains clinical information, genomic characterization data, and high-throughput sequencing analysis of the tumor genomes.

The TCGA Data Portal does not host lower levels of sequence data. NCI's Cancer Genomics Hub (CGHub) is the new secure repository for storing, cataloging, and accessing sequence related data. New users must still apply for authorized access through NCBI's Database of Genotypes and Phenotypes (dbGaP).

Query the Data ▸

Search summarized data for genes, patients and pathways

Download Data ▸

Choose from three ways to download data

Available Cancer Types	# Patients with Samples	# Downloadable Tumor Samples	Date Last Updated (mm/dd/yy)
Acute Myeloid Leukemia [LAML]	202	200	02/15/13
Bladder Urothelial Carcinoma [BLCA]	171	153	03/07/13
Brain Lower Grade Glioma [LGG]	232	222	03/08/13
Breast invasive carcinoma [BRCA]	956	940	03/08/13

Announcements

03/06/2013 - DCC Software Released

The software release scheduled for today has been successfully completed and the TCGA Data Portal has been returned to operation. A complete list of the issues addressed in this release can be found on the TCGA Wiki release notes and for those with JIRA access the tickets covered in this release can be found on the wiki here. Please note the release notes have been updated since they were published.

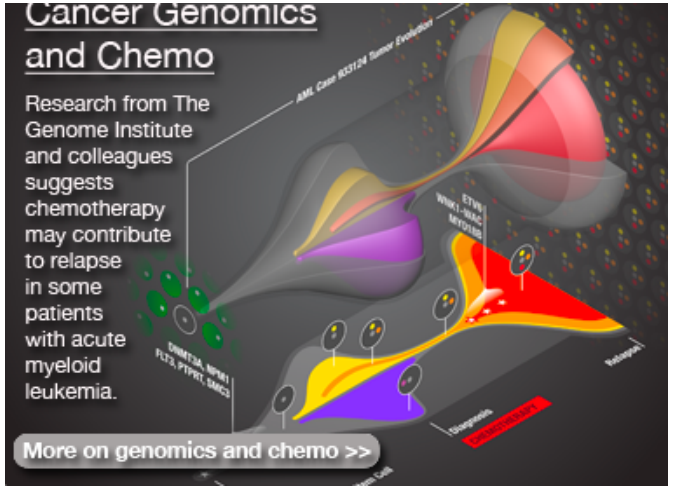
If you have any questions or concerns about this release, contact tcga-dcc-binf-l@list.nih.gov.

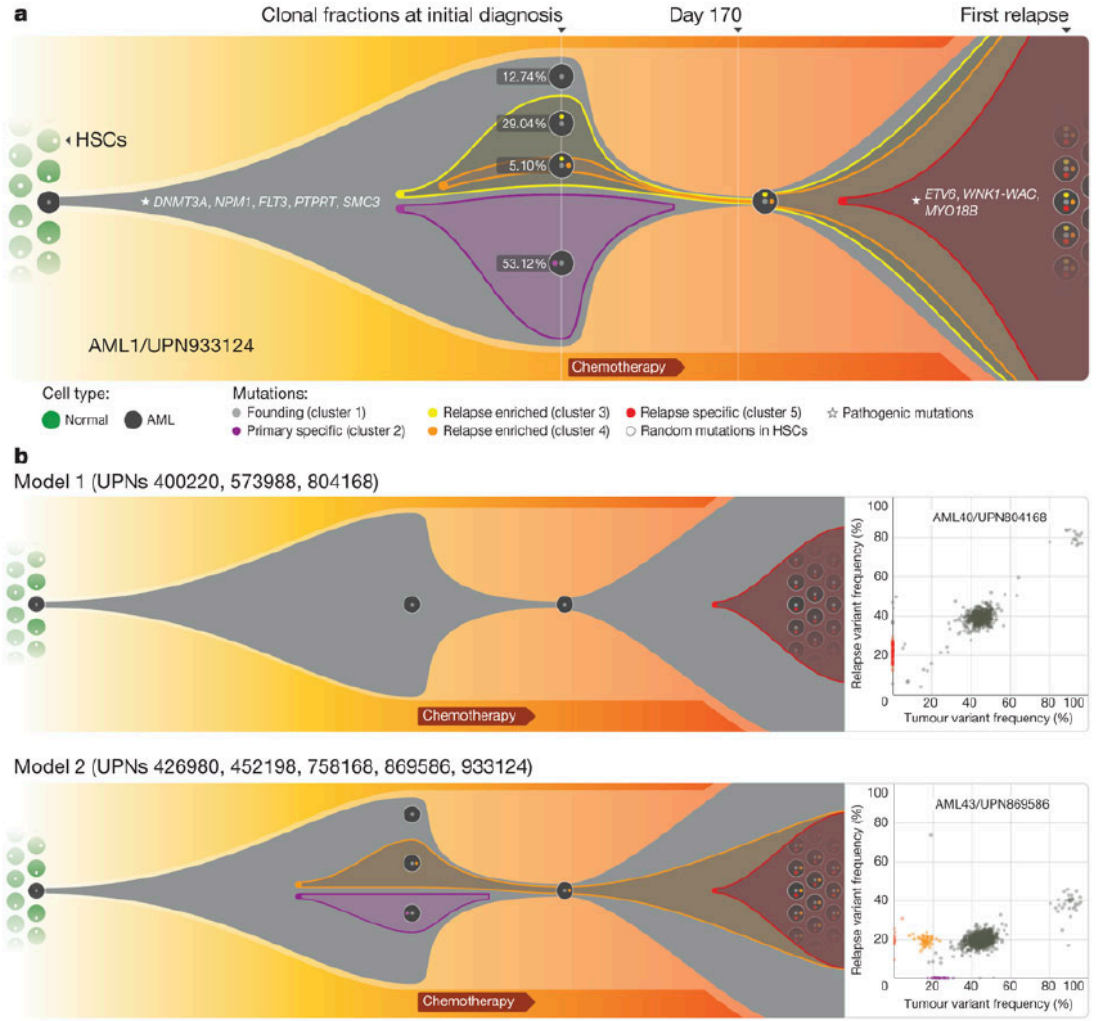
02/25/2013 - DCC Software Released

The software release scheduled for today has been successfully completed and the TCGA Data Portal has been returned to operation. A complete list of the issues addressed in this release can be found on the TCGA Wiki Release Notes and for those with JIRA access the tickets covered in this release can be found on the wiki here

If you have any questions or concerns about this release, contact tcga-dcc-

We can also observe the dynamics and evolution of cancers



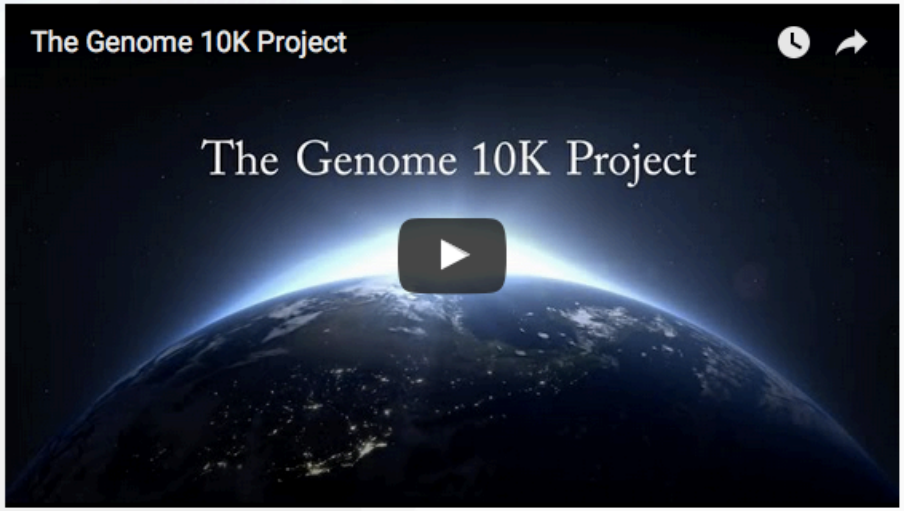


Ding L, et.al, Clonal evolution in relapsed acute myeloid leukemia revealed by whole-genome sequencing. Nature. 2012 Jan 11;481(7382):506-10.

And look beyond just humans

Genome 10K Project

To understand how complex animal life evolved through changes in DNA and use this knowledge to become better stewards of the planet



The Genome 10K project: Assembling a "Noah's Ark" of genomic data to save dying species.

GENOME 10K.

https://genome10k.soe.ucsc.edu/

https://www.hgsc.bcm.edu/i5k-pilot-project-summary

i5k

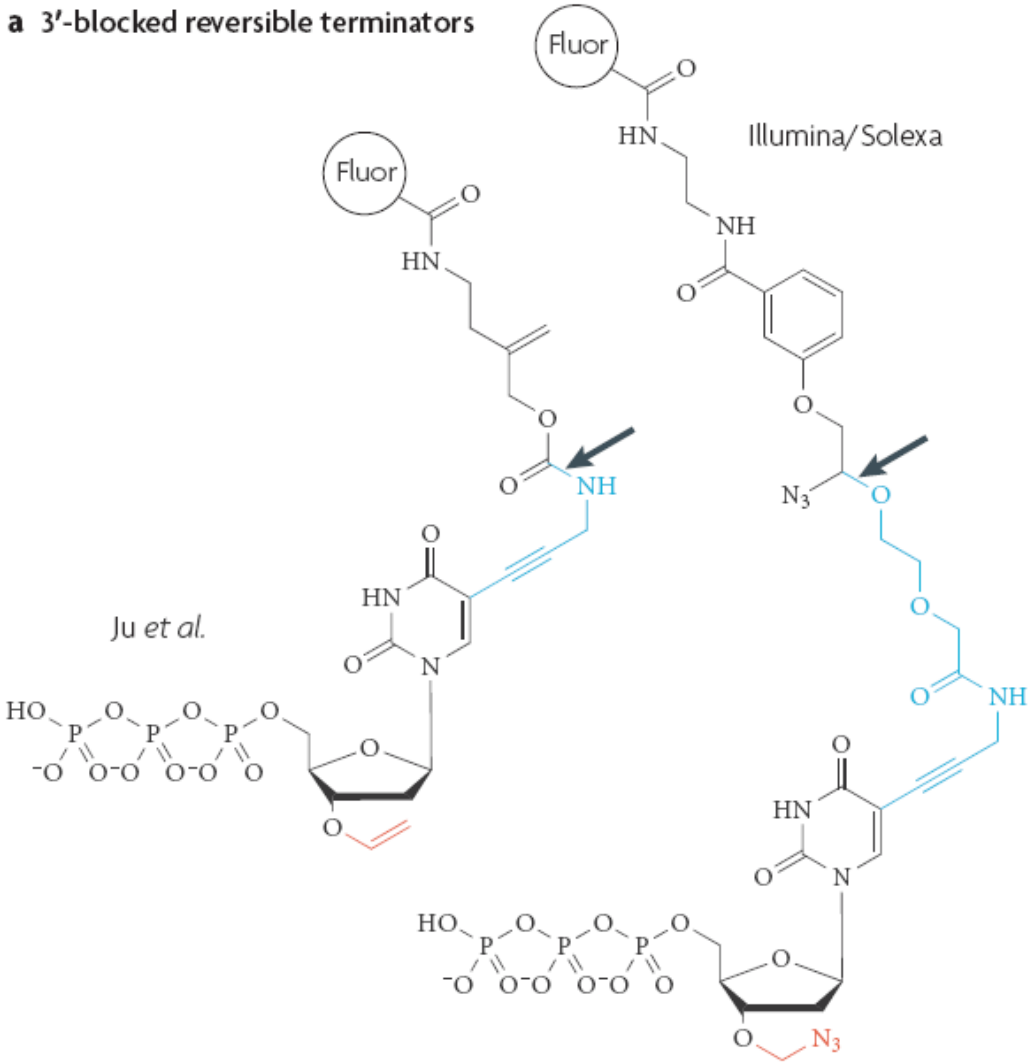
Plants as well!

华大基因
BGI

http://ldl.genomics.cn/page/pa-research.jsp

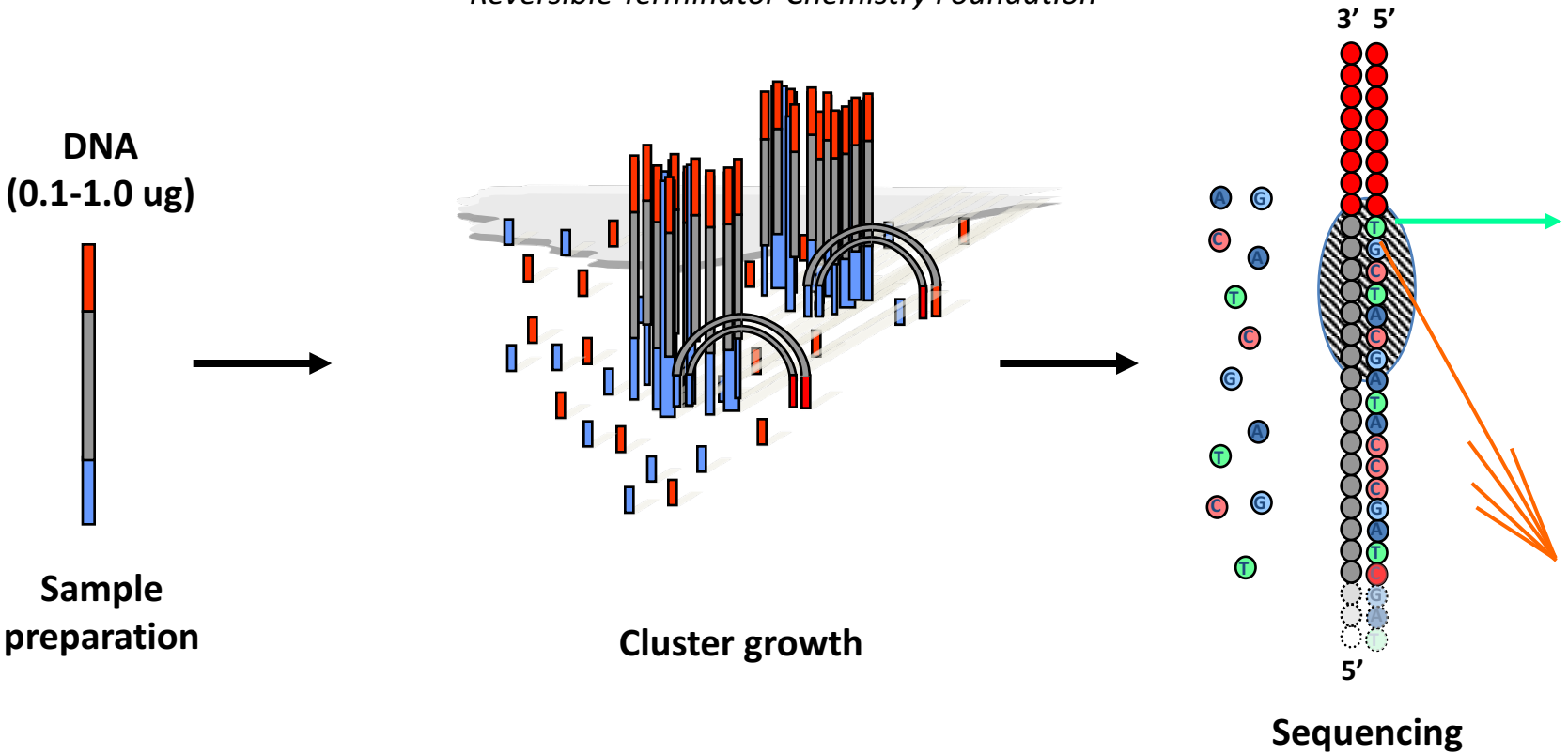
Consideration of WGS for each platform

Reversible Terminator Bases are Essential Technology Used in Many Chemistries

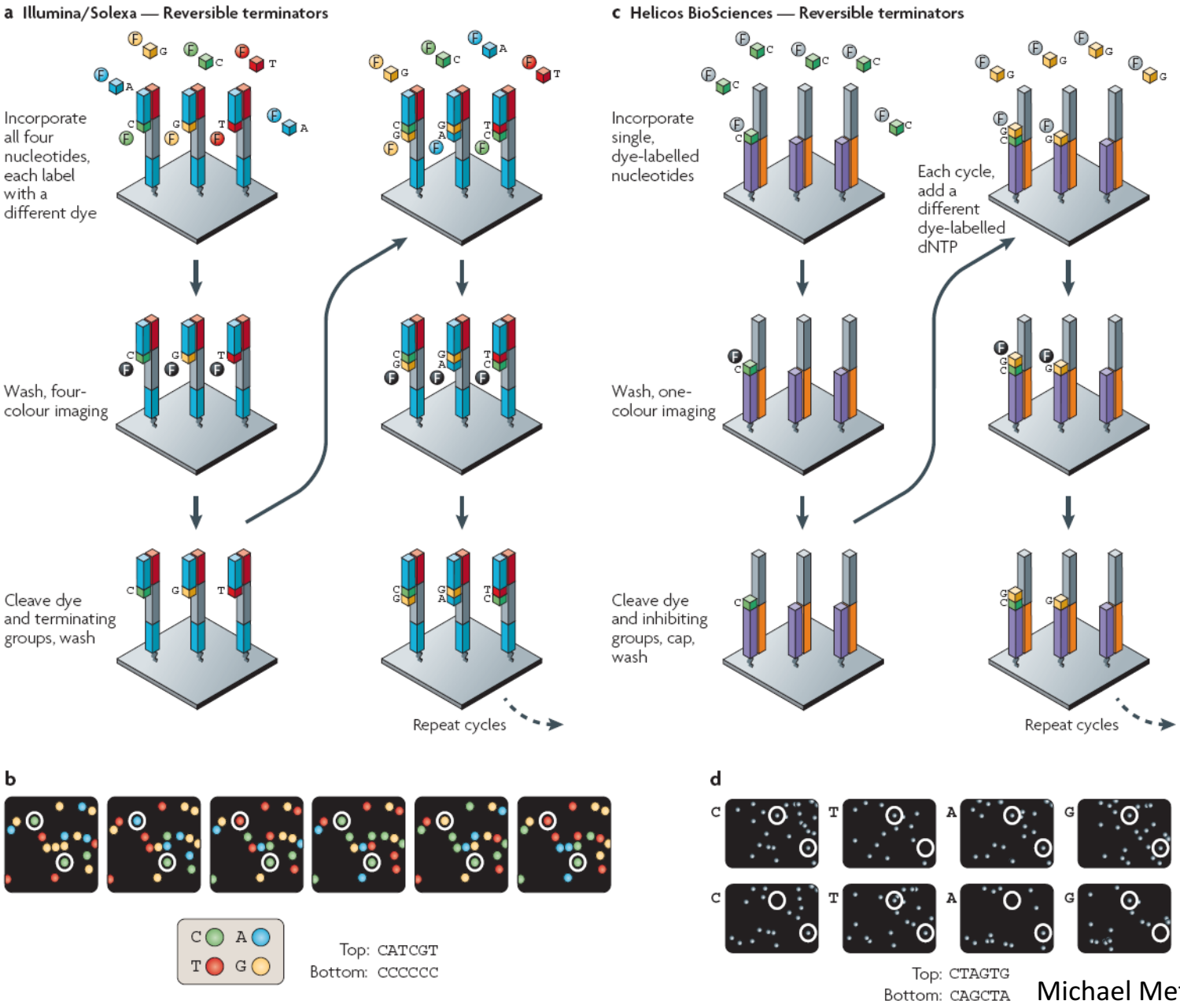


Illumina SBS Technology

Reversible Terminator Chemistry Foundation



Sequencing by Synthesis (SBS)



Michael Metzker, 2010

Now three kinds of chemistry

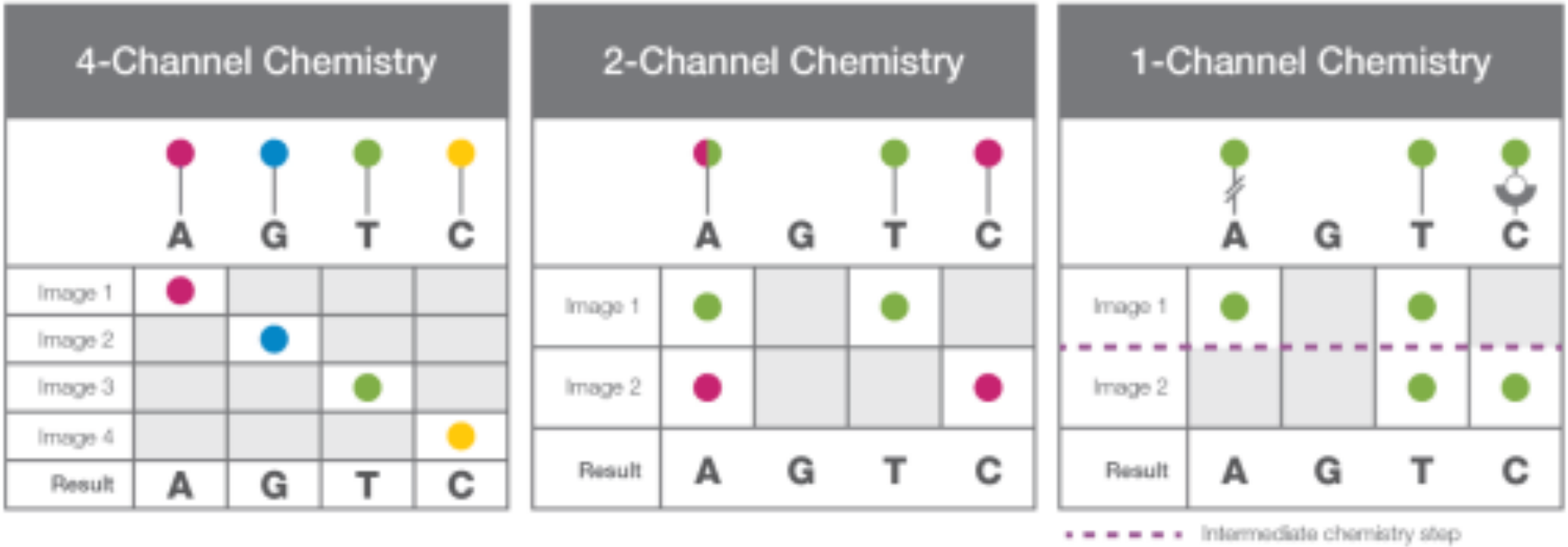
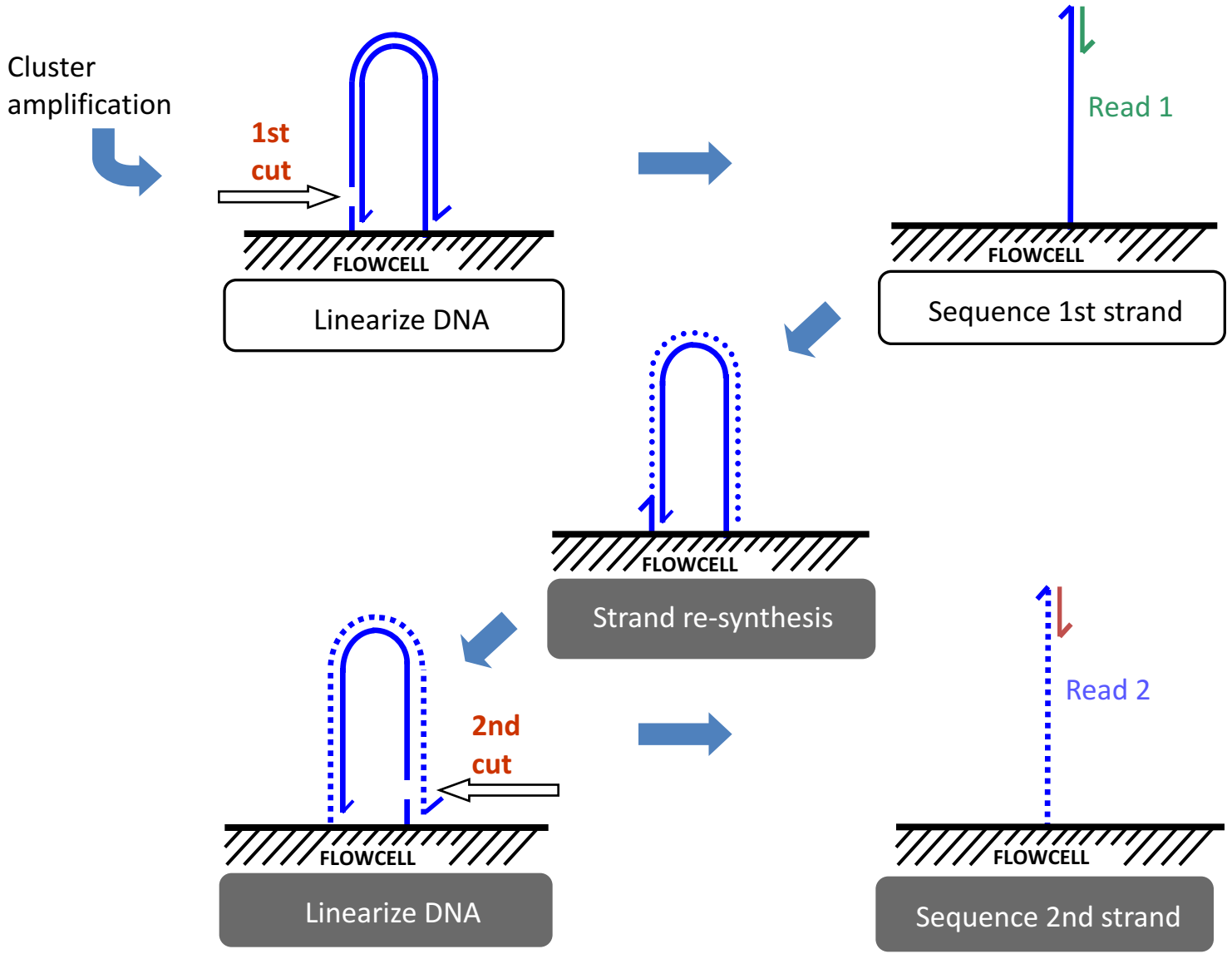
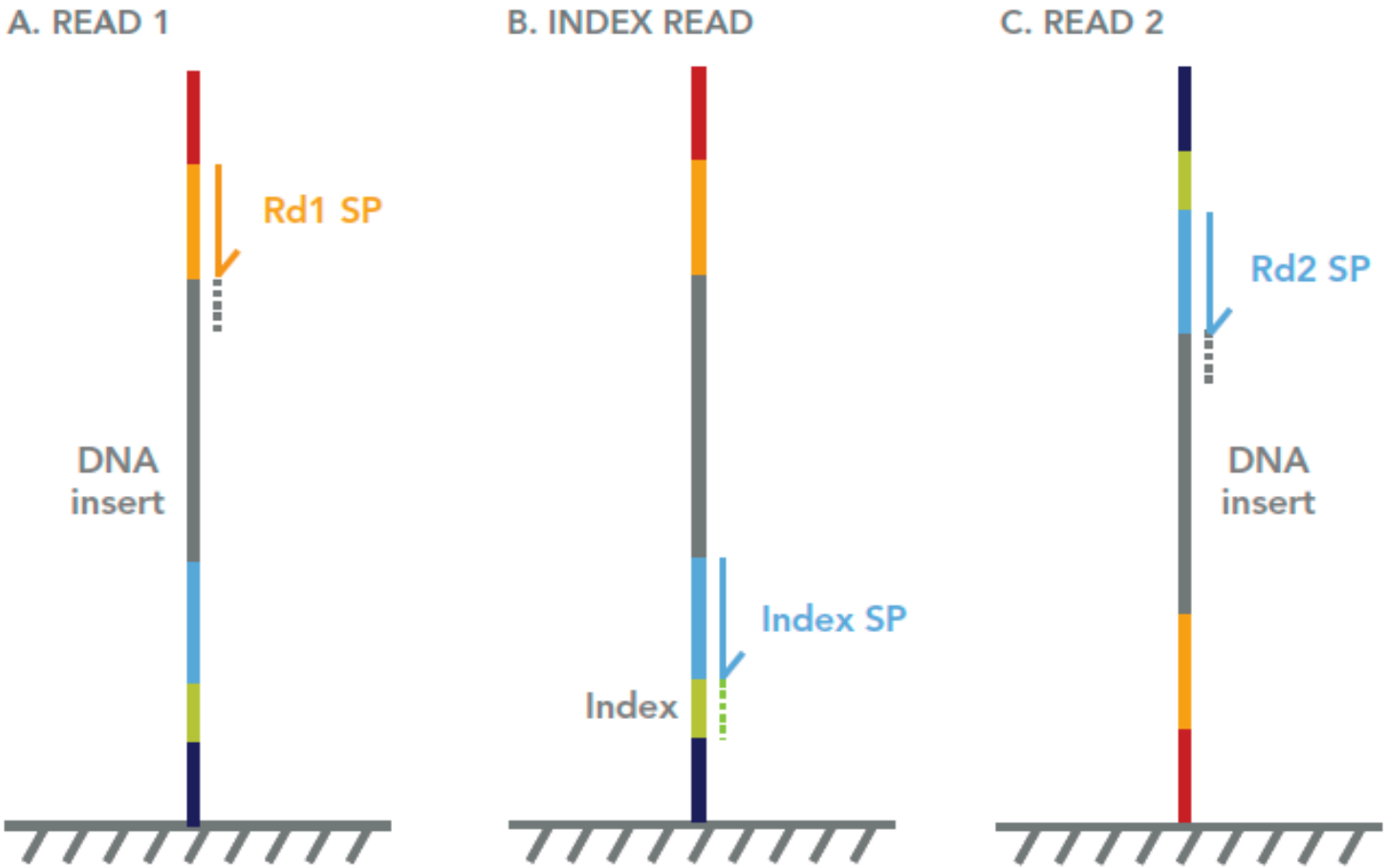


Figure 2: Four-, Two-, and One-Channel Chemistry—Four-channel chemistry uses a mixture of nucleotides labeled with four different fluorescent dyes. Two-channel chemistry uses two different fluorescent dyes, and one-channel chemistry uses only one dye. The images are processed by image analysis software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

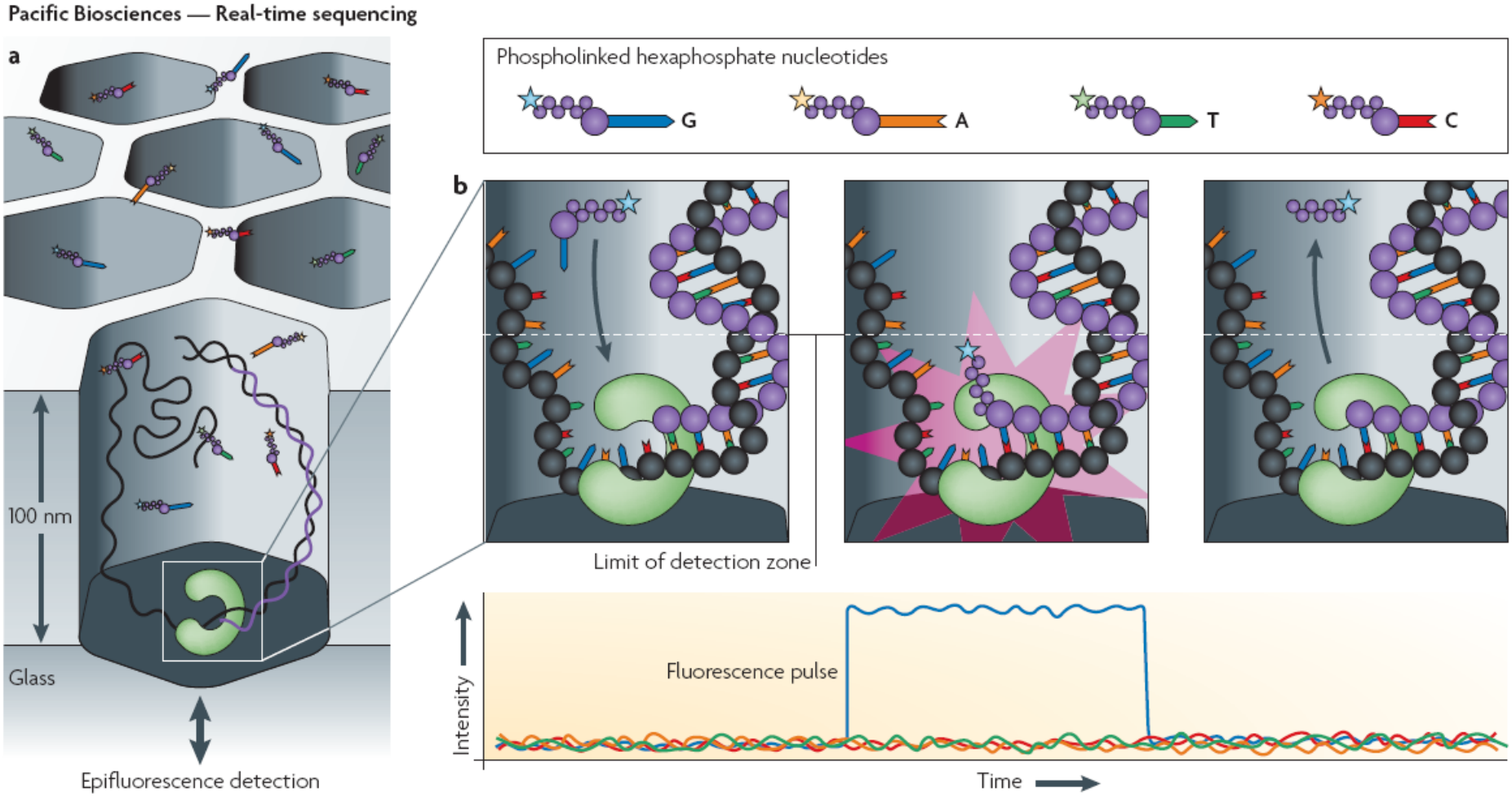
Paired-End Sequencing allows for two looks at a sequence



Indexed sequencing method is now standard for single and paired reads



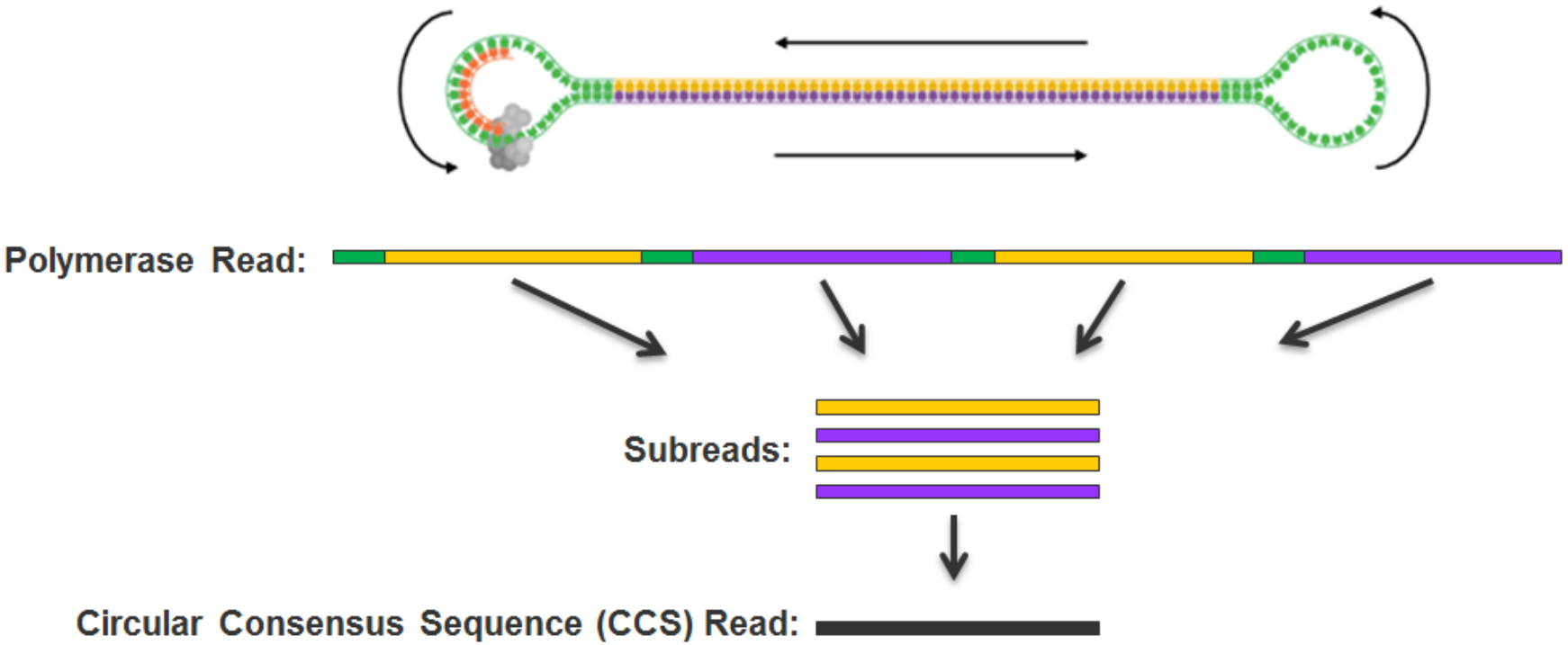
Pacific Biosciences
Single Molecule Real-Time (SMRT) Sequencing



https://www.pacb.com/videos/video-overview-of-smrt-technology/

Metzker, 2010

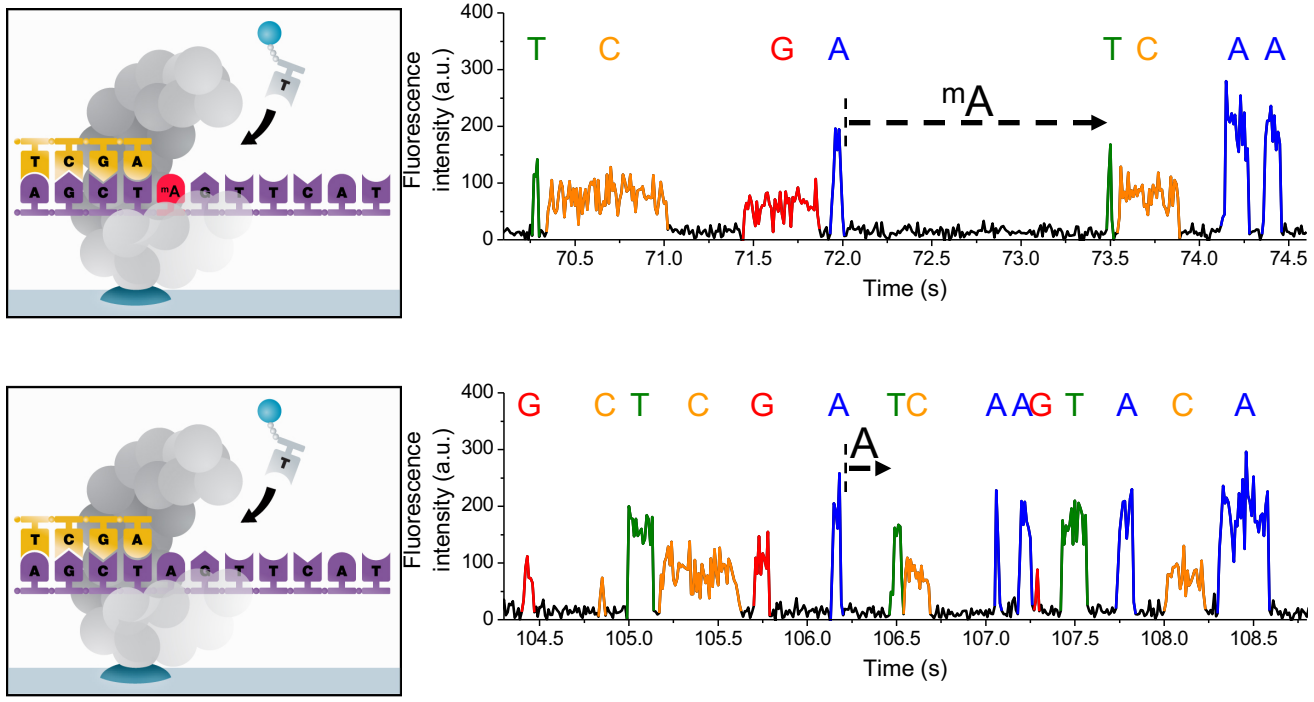
Circular Consensus Sequencing (CCS)



Single Molecule Kinetics Allow for the Direct Detection of Methylation

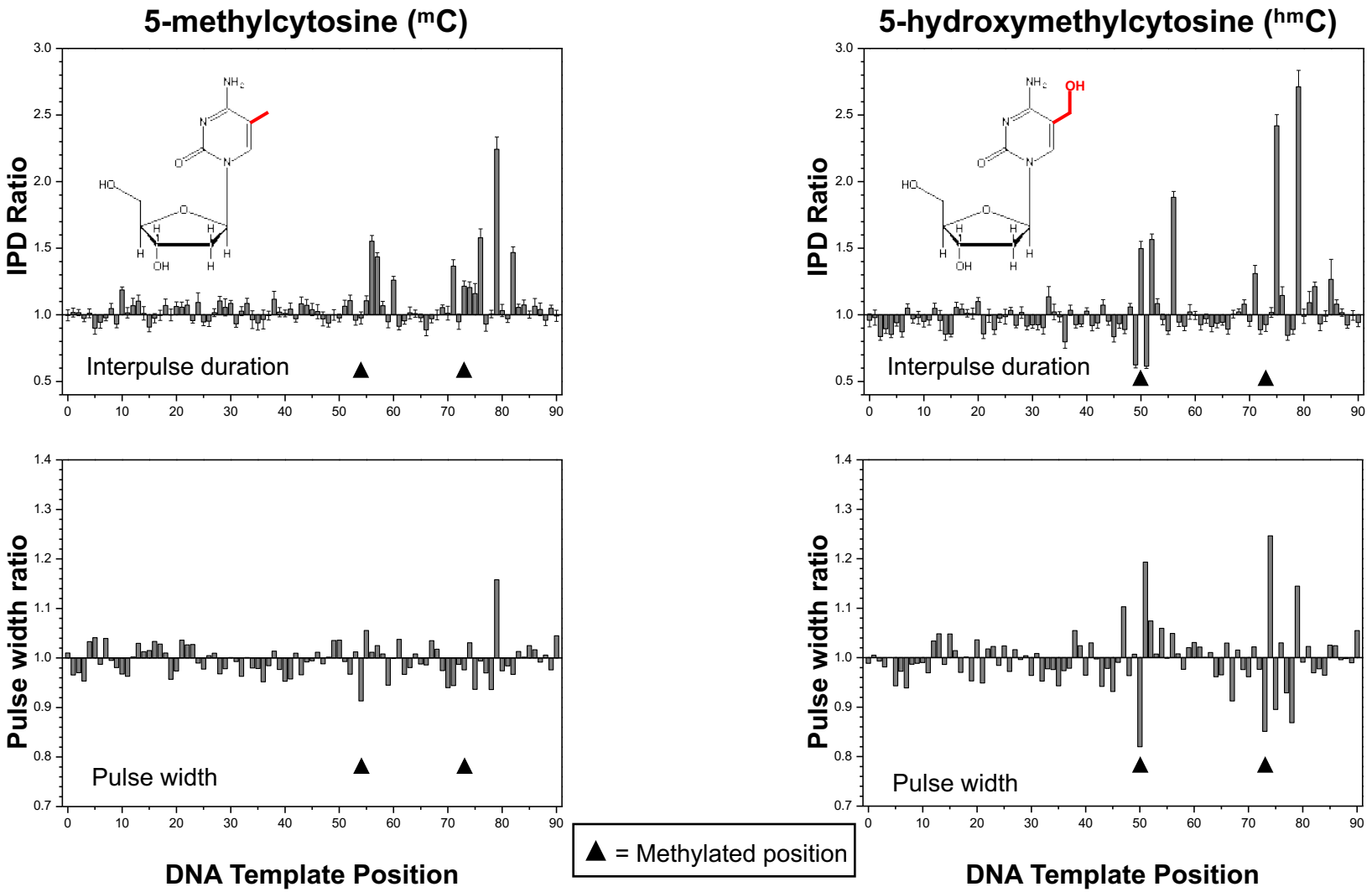
Approach: Kinetic detection of methylated bases during SMRT DNA sequencing

Example: N^6-methyladenosine (^{m}A)

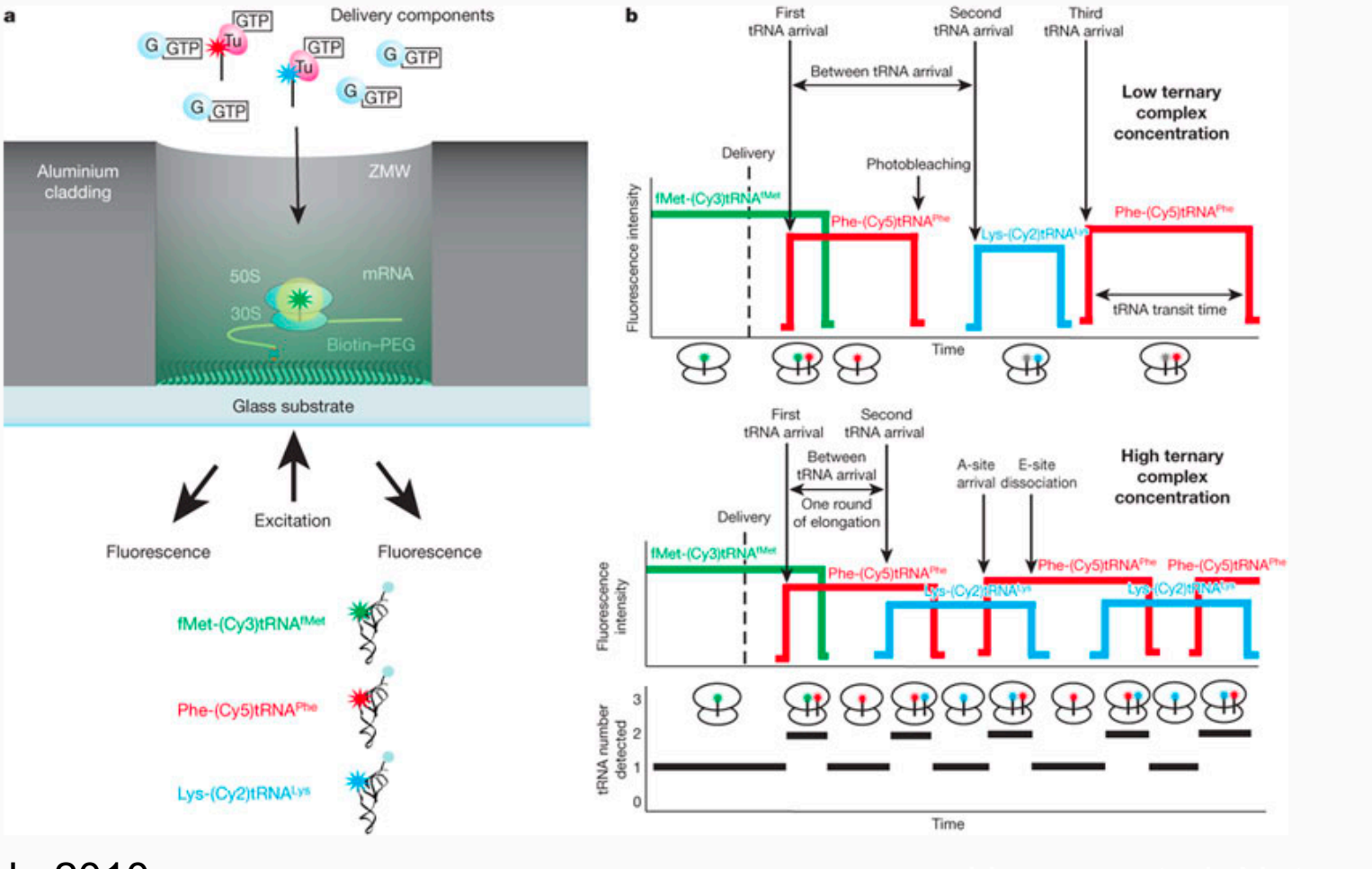


Flusberg et al., 2010.

Kinetics can detect other base modifications



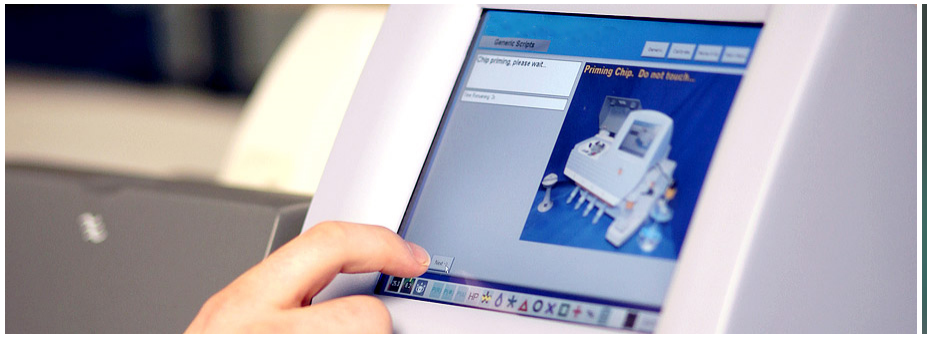
Kinetics allow one to watch protein translation as it occurs

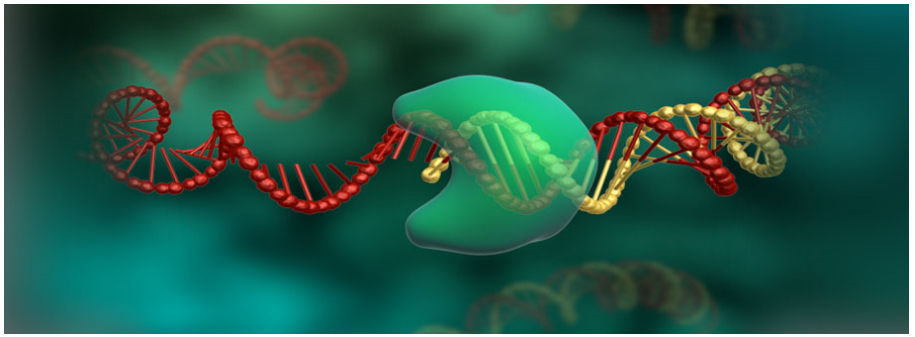


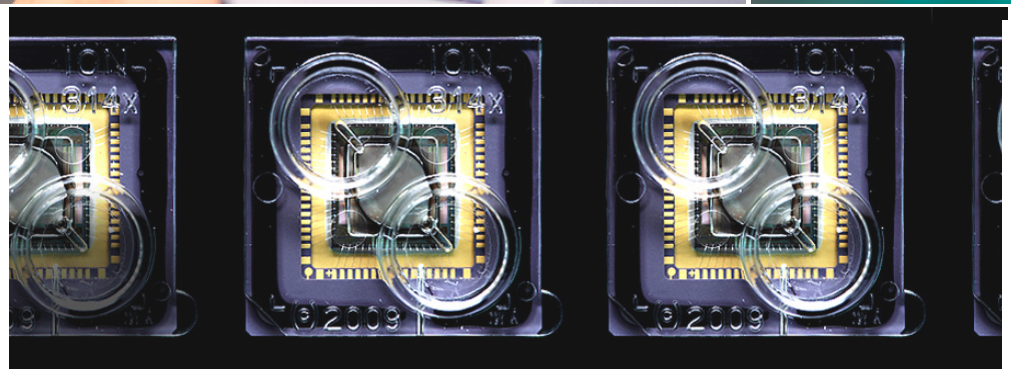
Uemura et al., 2010

“Post-Light,”
Semi-Conductor Sequencing:

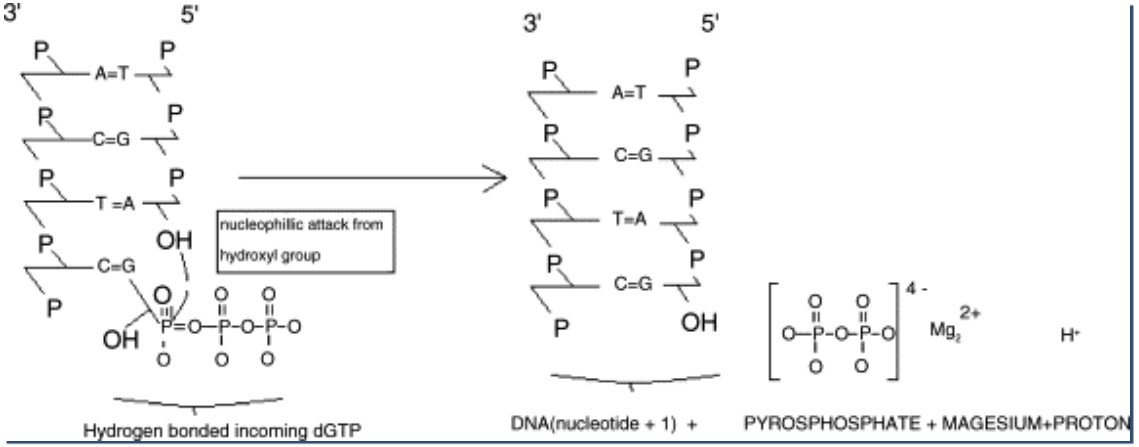
Thermo Fisher’s Personal Genome Machine (PGM), the Proton I and Proton II, and S5







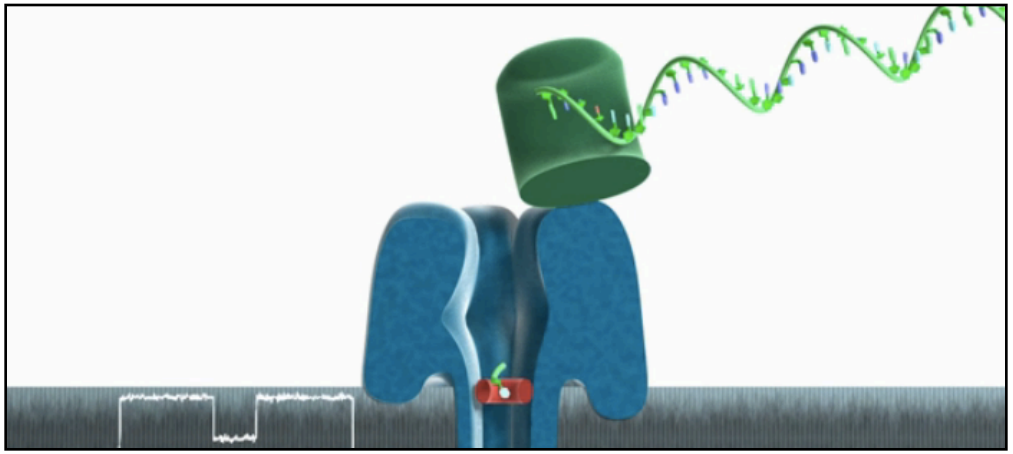
Essentially, Millions of very small pH meters



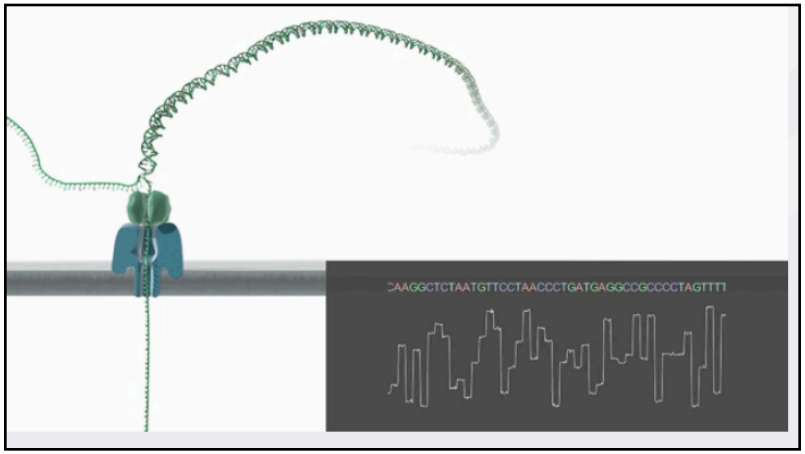
Purushothaman *et al*, 2005
IonTorrent, Inc.

Latest Ion Platforms
Thermo Fisher's Ion S5 & S5 XL

DNA Sequencing with a protein nanopore

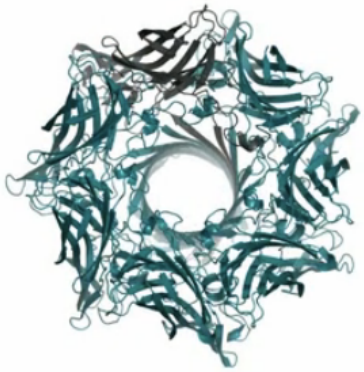


Exonuclease-Seq



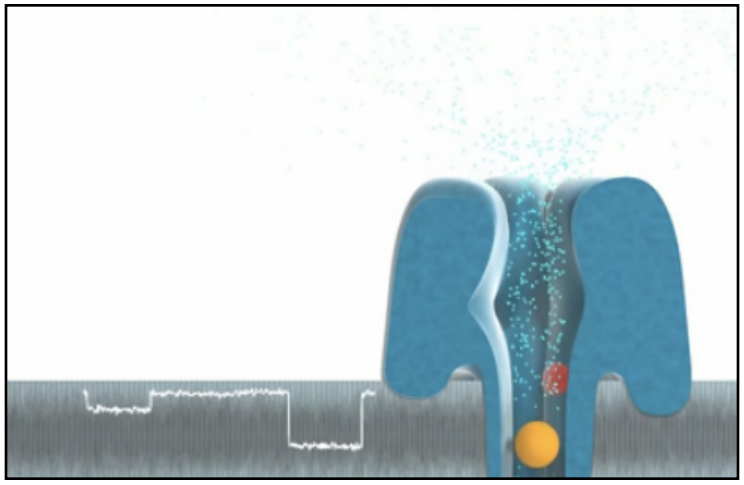
Strand-Seq

MinION

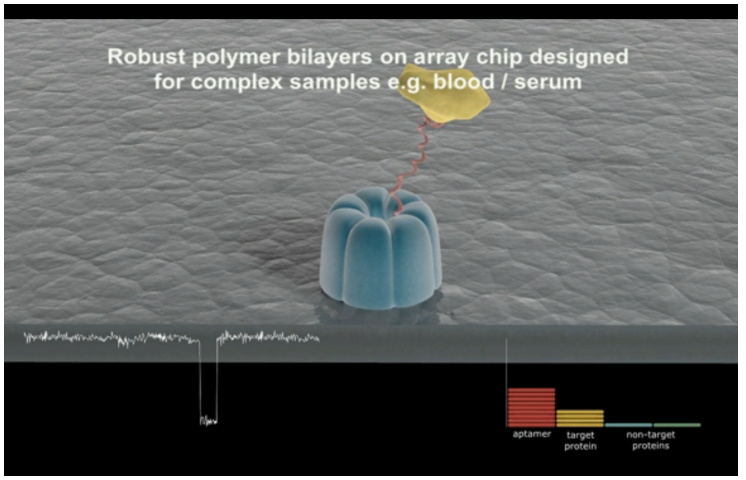


PromethION

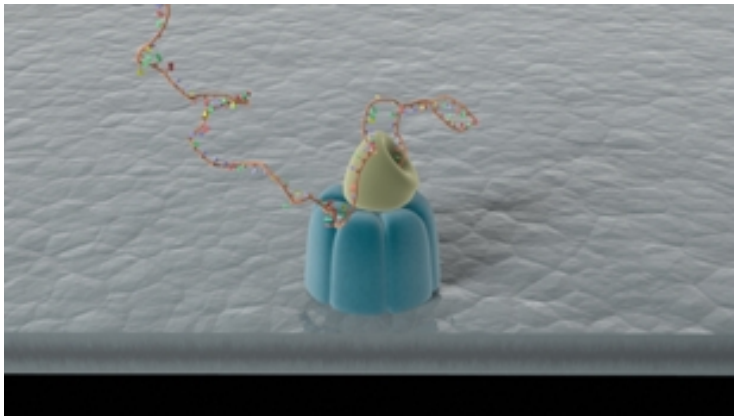
Other (Maybe Killer) Apps



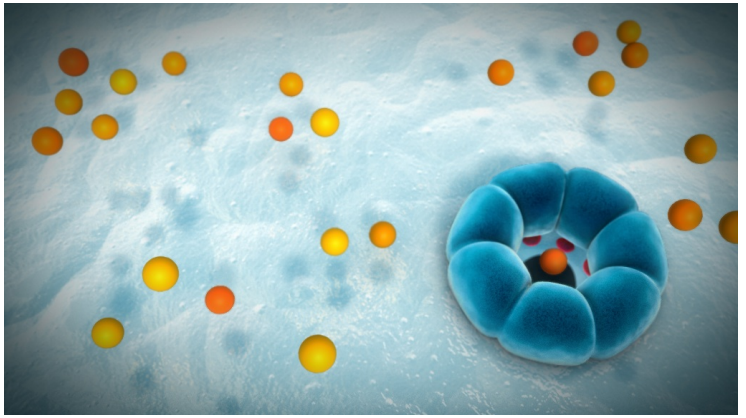
Analyte



Protein Aptamer

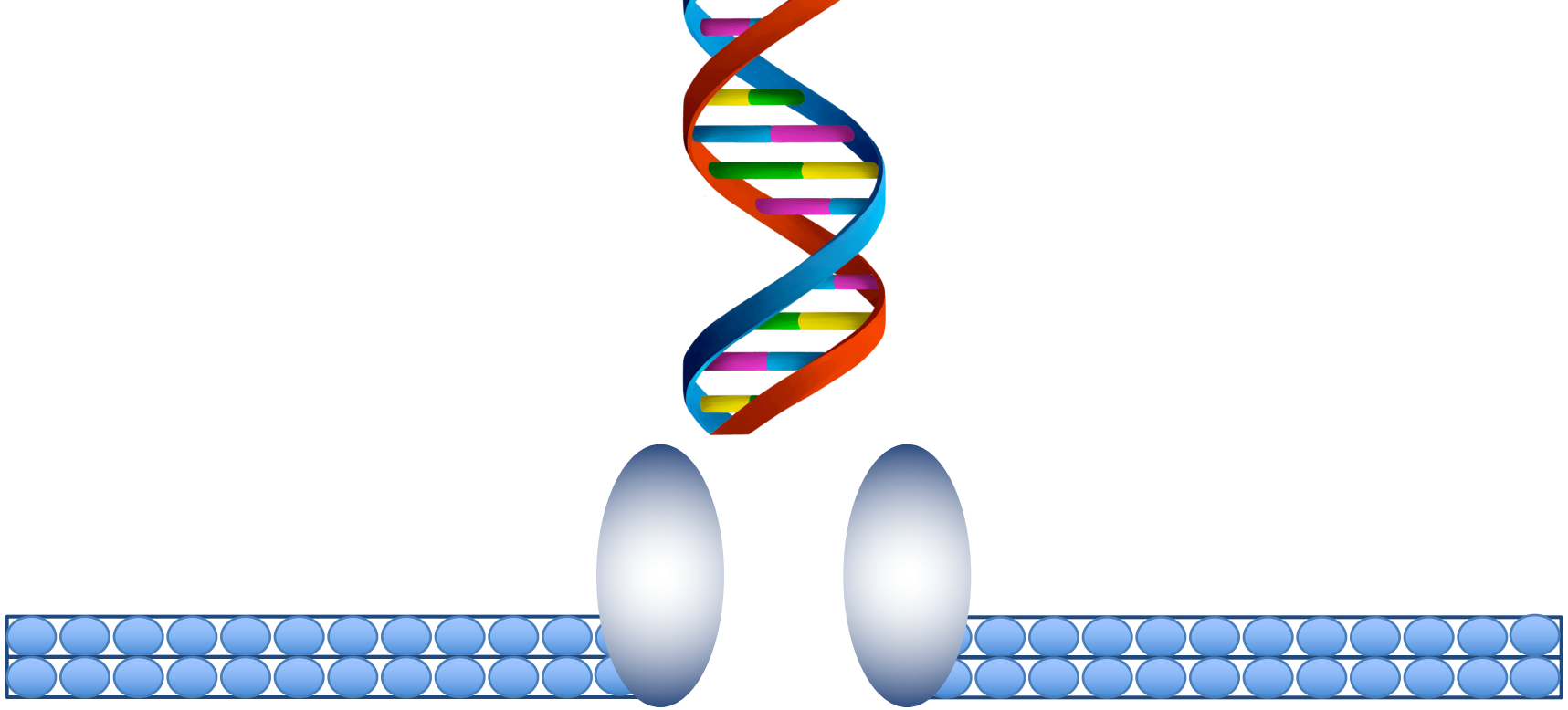


Direct RNA Sequencing



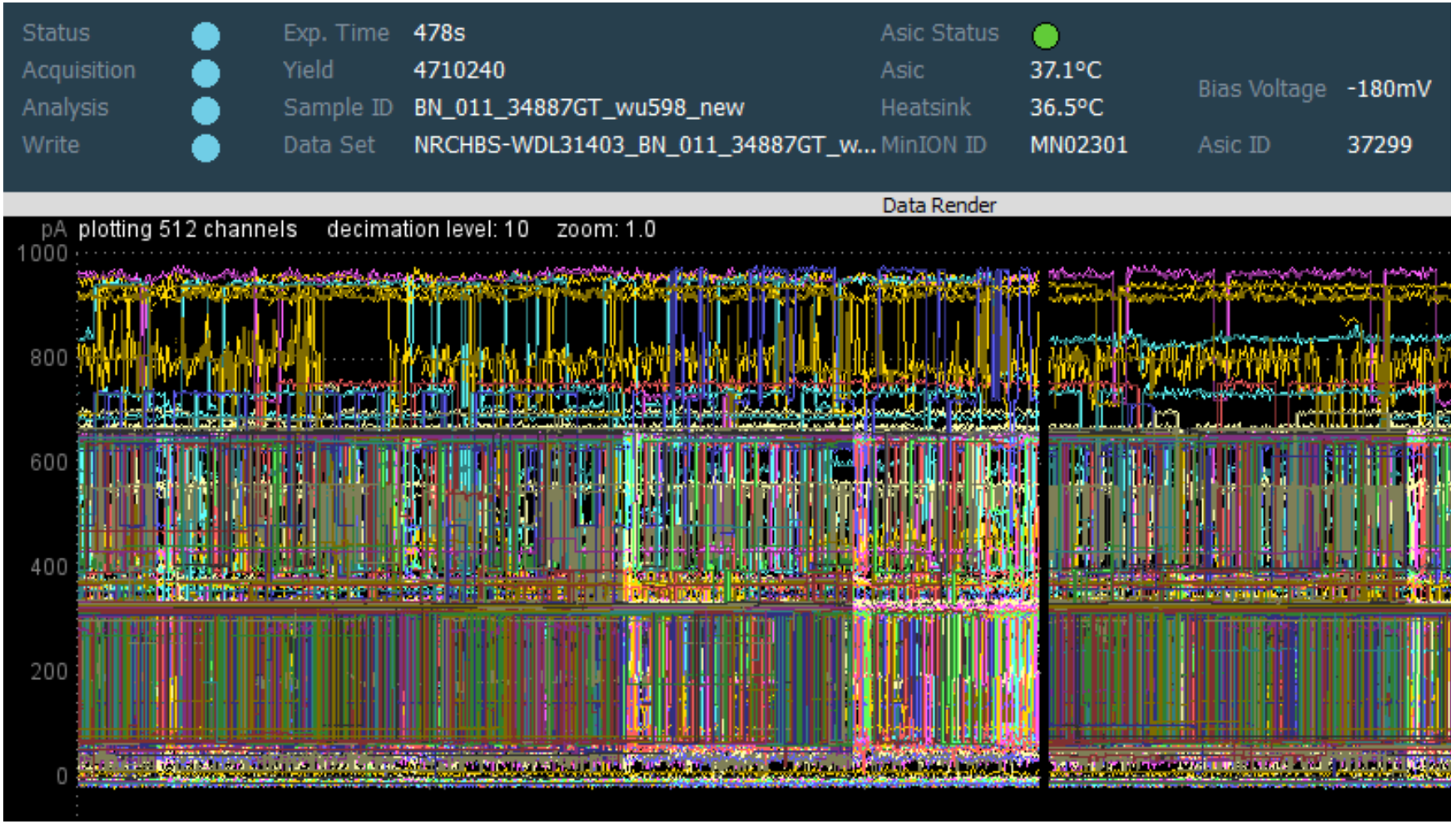
Small molecule

They are small



Meyer *et al*., *Cell*, 2012 | Saletore *et al*., *Genome Biology*, 2012 | McIntyre *et al*., 2015

Base space is now “squiggle space”



You can do it anywhere

nature
International journal of science

Letter | Published: 03 February 2016

Real-time, portable genome sequencing for Ebola surveillance

Joshua Quick, Nicholas J. Loman [...] Miles W. Carroll

Nature **530**, 228–232 (11 February 2016) | Download Citation

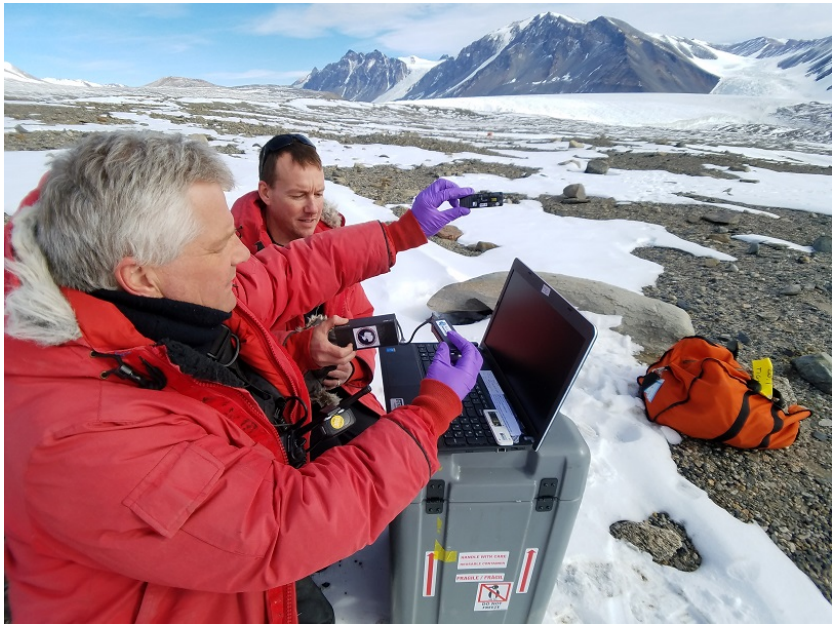
https://www.nature.com/articles/nature16996

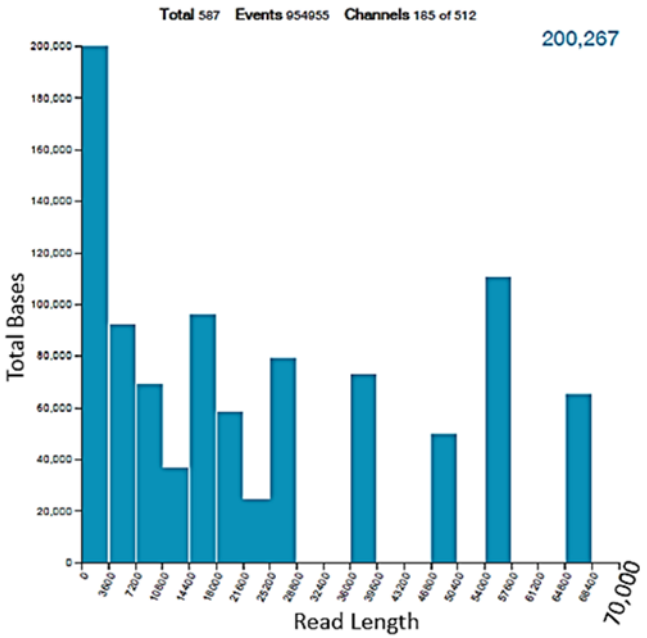

N403IL
NATIONAL SCIENCE FOUNDATION

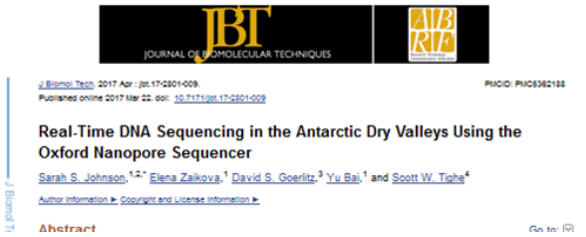

Scott Tighe

Lake Fryxell, Antarctica
Scott Tighe

Sequencing HW DNA in the field with the Oxford Nanopore
Sarah Johnson (PI) expedition G062 team






JOURNAL OF BIOMOLECULAR TECHNIQUES

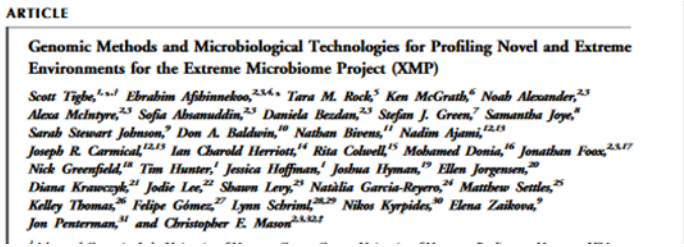
J Biomol Tech. 2017 Apr ; jbt.17-2801-009. PMCID: PMC5382188
Published online 2017 Mar 22. doi: 10.7171/jbt.17-2801-009

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

Sarah S. Johnson,[1,2,*] Elena Zaikova,[1] David S. Goerlitz,[3] Yu Bai,[1] and Scott W. Tighe[4]

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,[1,2,†] Ebrahim Afshinnekoo,[2,3,4,†] Tara M. Rock,[5] Ken McGrath,[6] Noah Alexander,[2,3] Alexa McIntyre,[2,3] Sofia Ahsanuddin,[2,3] Daniela Bezdan,[2,3] Stefan J. Green,[7] Samantha Joye,[8] Sarah Stewart Johnson,[9] Don A. Baldwin,[10] Nathan Bivens,[11] Nadim Ajami,[12,13] Joseph R. Carmical,[12,13] Ian Charold Herriott,[14] Rita Colwell,[15] Mohamed Donia,[16] Jonathan Foox,[2,3,17] Nick Greenfield,[18] Tim Hunter,[1] Jessica Hoffman,[1] Joshua Hyman,[19] Ellen Jorgensen,[20] Diana Krawczyk,[21] Jodie Lee,[22] Shawn Levy,[23] Natàlia Garcia-Reyero,[24] Matthew Settles,[25] Kelley Thomas,[26] Felipe Gómez,[27] Lynn Schriml,[28,29] Nikos Kyrpides,[30] Elena Zaikova,[9] Jon Penterman,[31] and Christopher E. Mason[2,3,32,†]

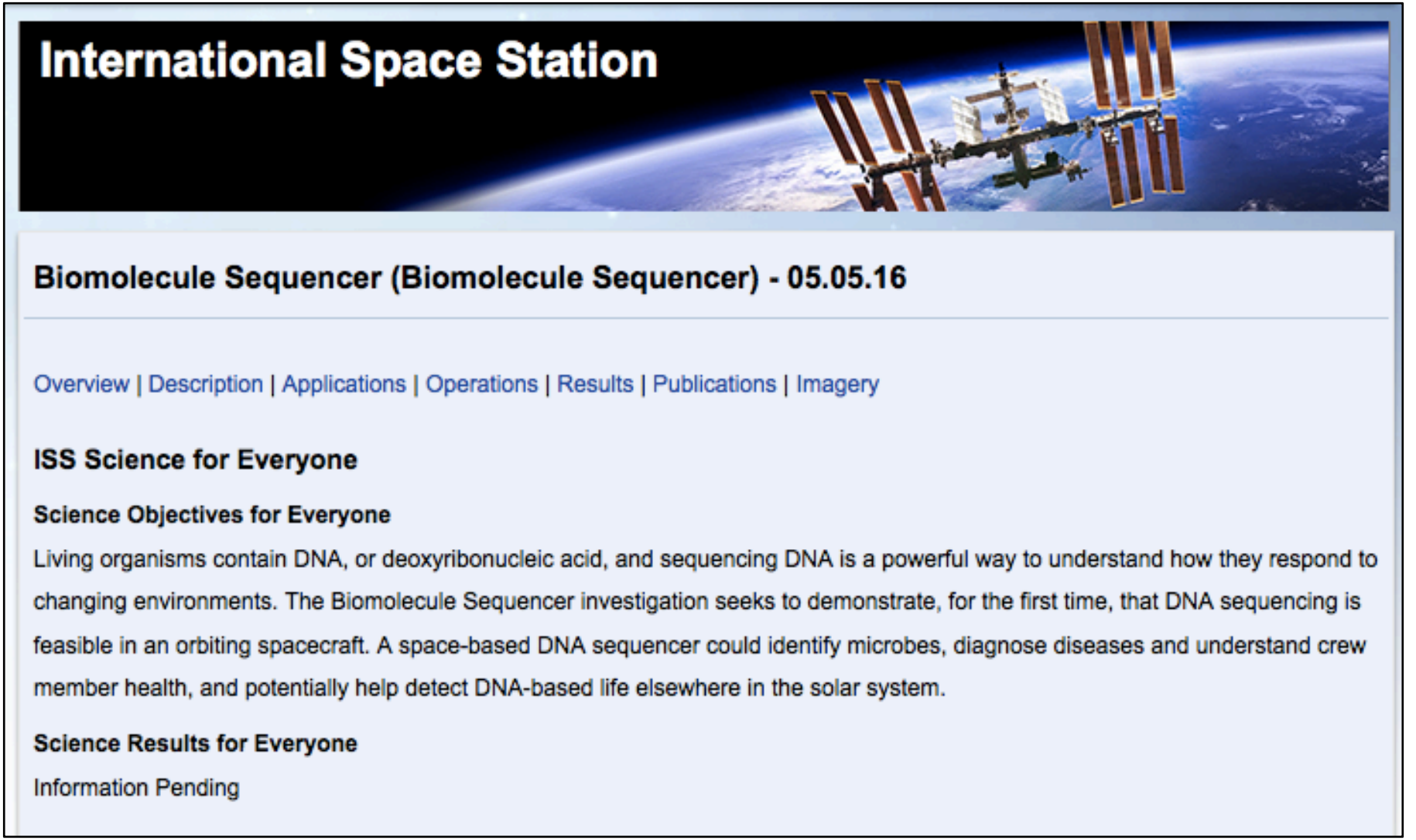
Integrative medicine with twin astronauts

(and maybe one future Senator)


THE TWINS STUDY
CSU JHU NASA NU SU UCSD UPenn WCMC wyle

< time
for measures

It would more convenient on the ISS and Mars to just sequence there!



http://www.nasa.gov/mission_pages/station/research/experiments/2181.html

Can we sequence in space?

OpNom: Biomolecule Sequencer

Principal Investigator(s)

Aaron Burton, Ph.D., NASA JSC, Houston, TX, United States

Co-Investigator(s)/Collaborator(s)

Sarah Castro-Wallace, Ph.D., NASA JSC, Houston, TX, United States

Kristen John, Ph.D., NASA JSC, Houston, TX, United States

Sarah Stahl, M.S., NASA Johnson Space Center, Houston, TX, United States

Douglas Botkin, Ph.D., Johnson Space Center, Houston, TX, United States

Jason Dworkin, Ph.D., NASA GSFC, Greenbelt, MD, United States

Mark Lupisella, Ph.D., NASA GSFC, Greenbelt, MD, United States

David Smith, Ph.D., NASA Ames, Moffett Field, CA, United States

Christopher Mason, Ph.D., Weill Cornell Medical College, New York, NY, United States

James Brayer, Oxford Nanopore Technologies Inc., Cambridge, MA, United States

Developer(s)

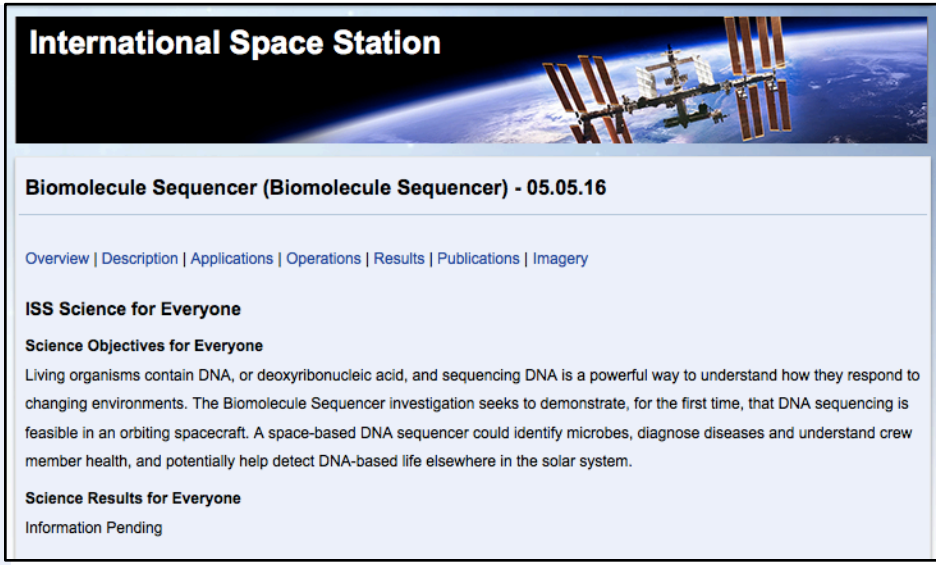
NASA Johnson Space Center, Houston, TX, United States

Sponsoring Space Agency

National Aeronautics and Space Administration (NASA)

Sponsoring Organization

Technology Demonstration Office (TDO)



- Operational environmental monitoring of microorganisms
  - Allow for in-flight identification of microbes, which is currently not possible but is essential for travel beyond our moon.
  - Inform real-time decisions and remediation strategies.
- Medical operations
  - Real-time analysis can impact medical intervention and define countermeasure efficacy.
- Research
  - DNA from any organism can be sequenced to assist any scientific investigation on the ISS.
- Astrobiology
  - ISS demonstration serves as functional testing for integration into robotics for Mars exploration missions.
  - This technology is superiorly suited for the detection of life based on DNA and DNA-like molecules.

http://www.nasa.gov/mission_pages/station/research/experiments/2181.html

Zero-G Pipetting:
Hardest Lab Job Ever

Dr. Andrew Feinberg

nature International weekly journal of science

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

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

www.nature.com/npjmgrav/

Search

EDITOR-IN-CHIEF Dr. Cheryl A. Nickerson, Ph.D.

nature.com > npj microgravity

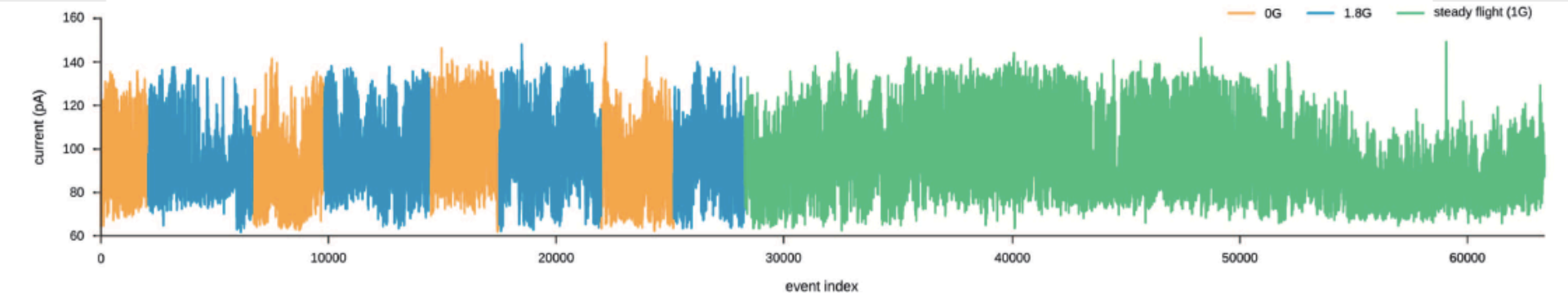
MENU

npj | Microgravity

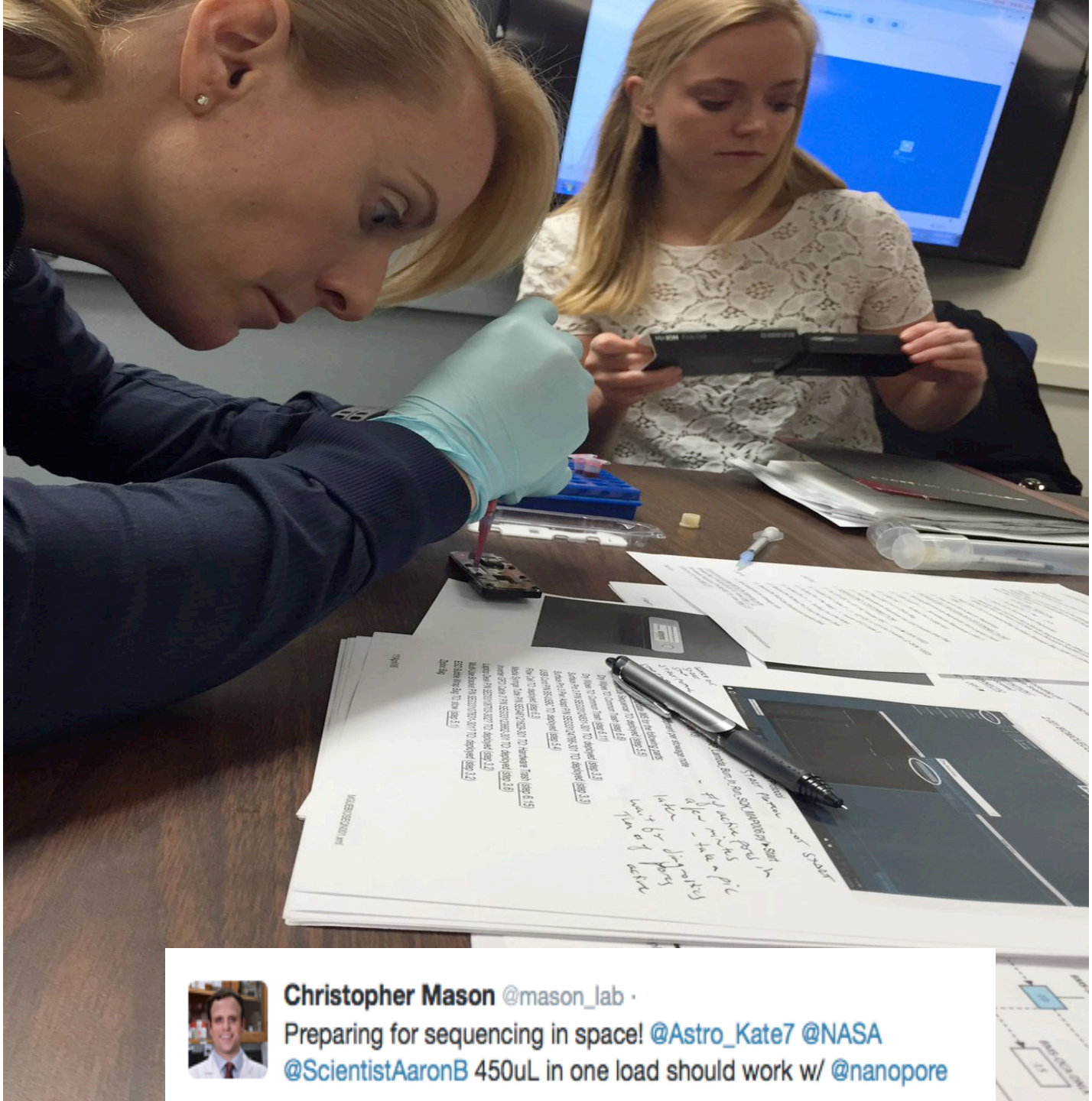
Search E-alert Submit Login

DNA sequencing in space: Nanopores ready for liftoff

Results from the first DNA sequencing experiments performed in microgravity reveal a promising future for portable 'nanopore' devices in space missions. Read the paper in full.



McIntyre ABR et al., *Nature Microgravity, 2016.*



Christopher Mason @mason_lab ·

Preparing for sequencing in space! @Astro_Kate7 @NASA @ScientistAaronB 450uL in one load should work w/ @nanopore

SpaceX CRS-7 blows up

National Aeronautics and Space Administration

Office of the Administrator
Washington, DC 20546-0001

Dr. Christopher Mason
Weill Cornell Medical College
1300 York Ave.
New York, NY 10065

Dear Dr. Mason:

As NASA astronaut Scott Kelley tweeted on Sunday, June 28, 2015, "space is hard."

Speaking as a fellow researcher, I can only imagine how devastated you must be feeling right now with the loss of SpaceX's CRS-7. I am saddened and disappointed too. I am sure that the tremendous honor of being selected to have your experiment flown on the International Space Station is of little solace after the loss of months, and perhaps even years, of hard work.

I am writing to encourage you – and in fact, to urge you – to continue your inquiry. The story of space exploration is the story of people just like you who meet adversity, head on, with determination and scientific and technological advancement. If you think about it, virtually every major innovation and technological breakthrough in human history has been the product of many different stops and starts; learning and being better because of failures and setbacks and, ultimately, enhanced knowledge and moving forward.


ВНИМАНИЕ


SpaceX CRS-9: perfect launch
and booster return
July 18, 2016


TO NOD 2

Latest | Related


Weekly Recap From the Expedition Lead Scientist
5 days ago


Weekly Recap From the Expedition Lead Scientist
13 days ago


Biological Sciences on the International Space Station
19 days ago

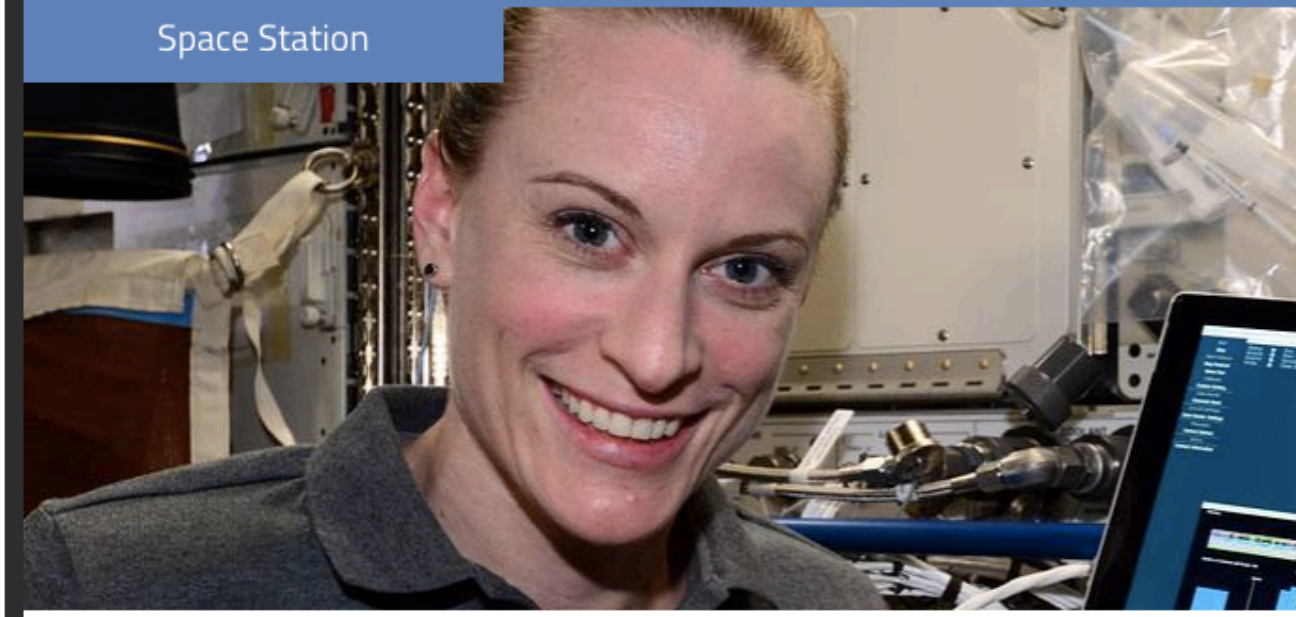

SAGE III to Look Back at Earth's Atmospheric 'Sunscreen'
19 days ago


Weekly Recap From the Expedition Lead Scientist
19 days ago


Weekly Recap From the Expedition Lead Scientist
24 days ago


Weekly Recap From the Expedition Lead Scientist
a month ago

Space Station



Aug. 29, 2016

First DNA Sequencing in Space a Game Changer

For the first time ever, DNA was successfully sequenced in microgravity as part of the **Biomolecule Sequencer** experiment performed by NASA astronaut Kate Rubins this weekend aboard the **International Space Station**. The ability to sequence the DNA of living organisms in space opens a whole new world of scientific and medical possibilities. Scientists consider it a game changer.

DNA, or deoxyribonucleic acid, contains the instructions each cell in an organism on Earth needs to live. These instructions are represented by the letters A, G, C and T, which stand for the four chemical bases of DNA, adenine, guanine, cytosine, and thymine. Both the number and arrangement of these bases differ among organisms, so their order, or sequence, can be used to identify a specific organism.

spasmunkey
@spasmunkey

Following

Great to see this team at work from training to operations at "the dawn of genomics...in space" #AstroKate

RETWEETS 4 LIKES 12

9:40 PM - 29 Aug 2016

Houston, TX

You, Aaron Burton, Kristen John and 3 others

4 12

From zero to one billion: sequencing the one billionth base pair of DNA in space. go.nasa.gov/2bV2UnD

sequencing the one billionth base pair of DNA

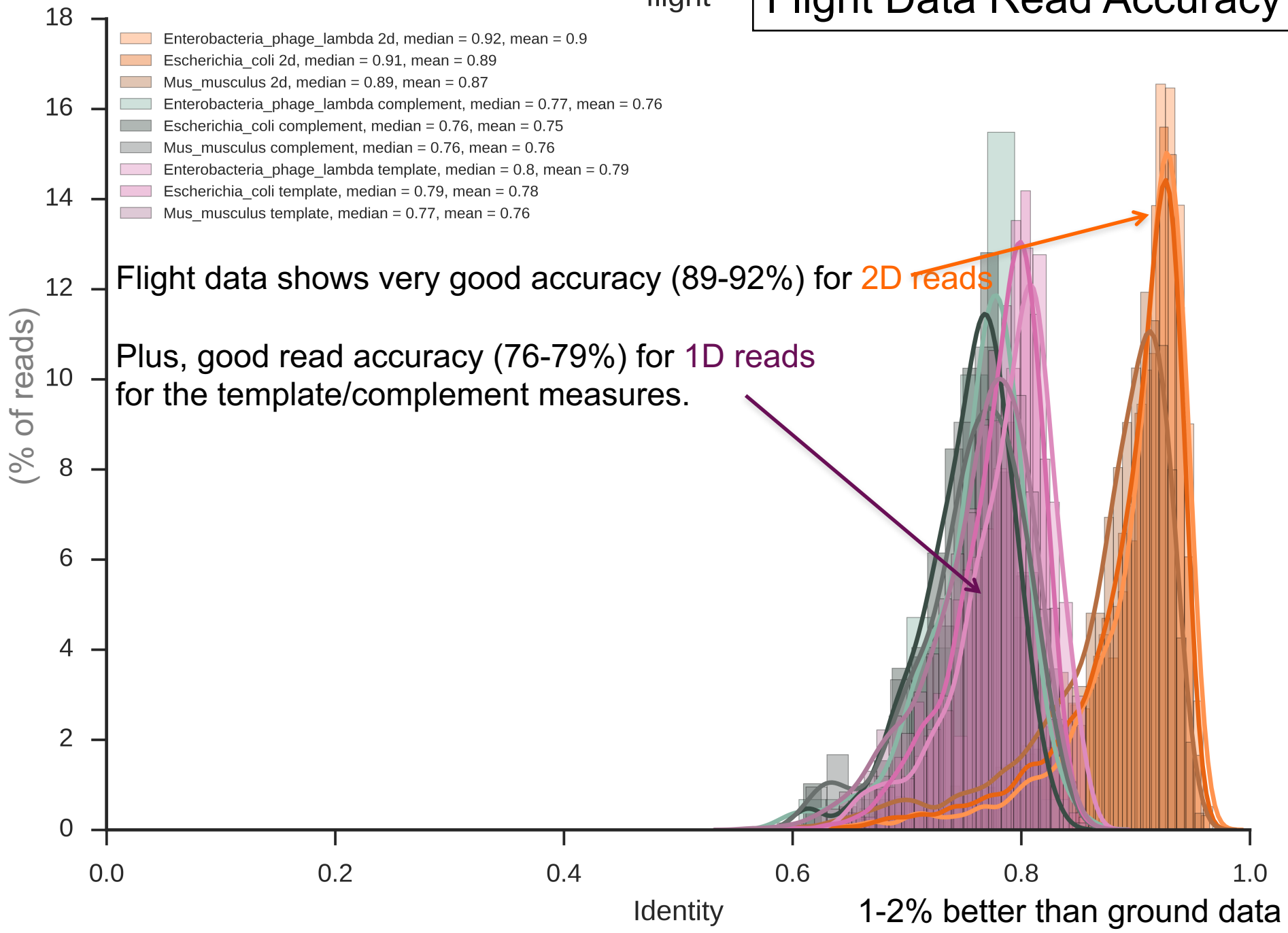
Clip from NASA TV

RETWEETS	LIKES
123	185

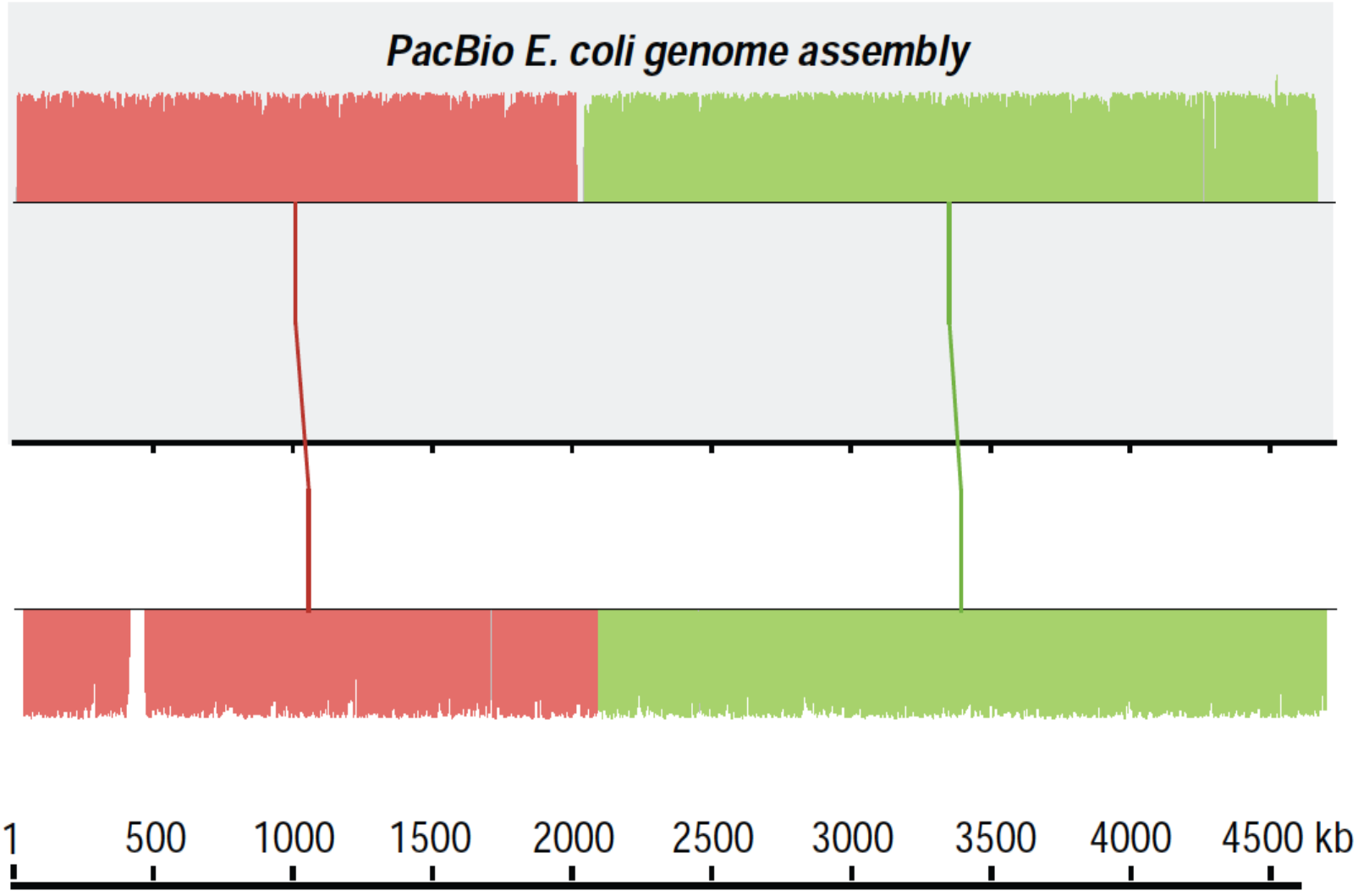
Bus Lon Dor Elai Alfc Oliv Jes  Lita

3:28 PM - 14 Sep 2016

Flight Data Read Accuracy



Almost perfect when compared to PacBio



The first genome sequence, assembly, and AMR detection off Earth

SCIENTIFIC REPORTS

Altmetric: 171 More detail »

Article | OPEN

Nanopore DNA Sequencing and Genome Assembly on the International Space Station

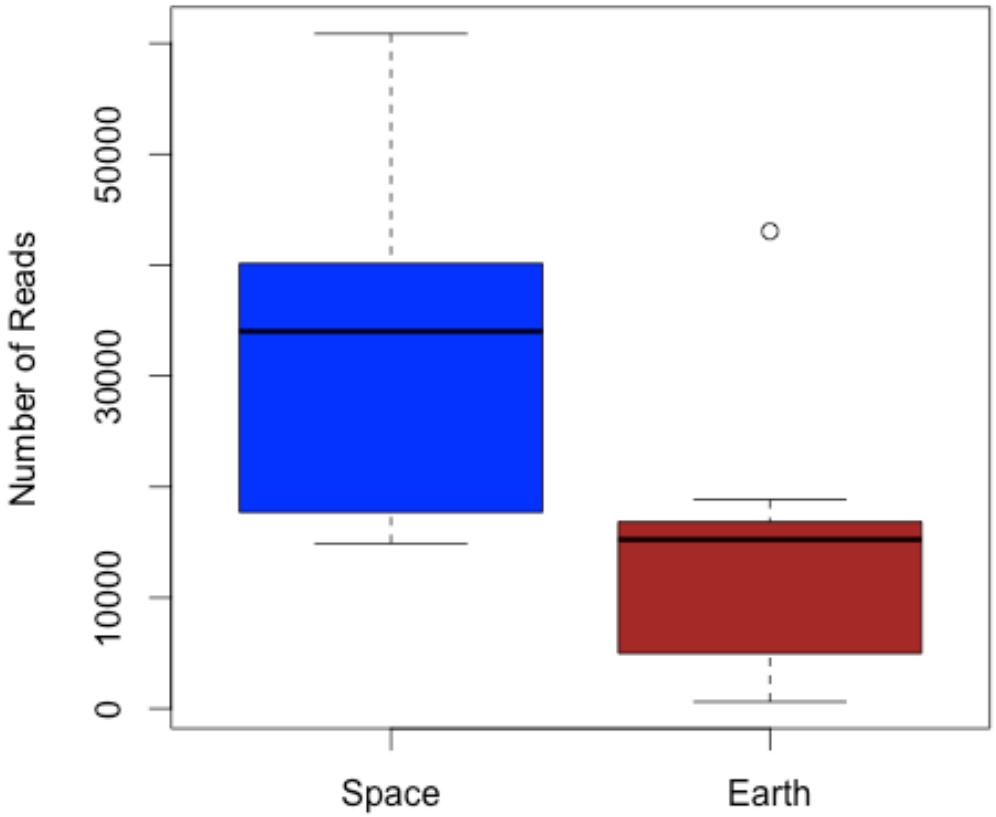
Sarah L. Castro-Wallace, Charles Y. Chiu, Kristen K. John, Sarah E. Stahl, Kathleen H. Rubins, Alexa B. R. McIntyre, Jason P. Dworkin, Mark L. Lupisella, David J. Smith, Douglas J. Botkin, Timothy A. Stephenson, Sissel Juul, Daniel J. Turner, Fernando Izquierdo, Scot Federman, Doug Stryke, Sneha Somasekar, Noah Alexander, Guixia Yu, Christopher E. Mason & Aaron S. Burton

Scientific Reports **7**, Article number: 18022 (2017)
doi:10.1038/s41598-017-18364-0

Received: 01 August 2017
Accepted: 11 December 2017
Published online: 21 December 2017

https://www.nature.com/articles/s41598-017-18364-0

As good, or better (8/9) data in space



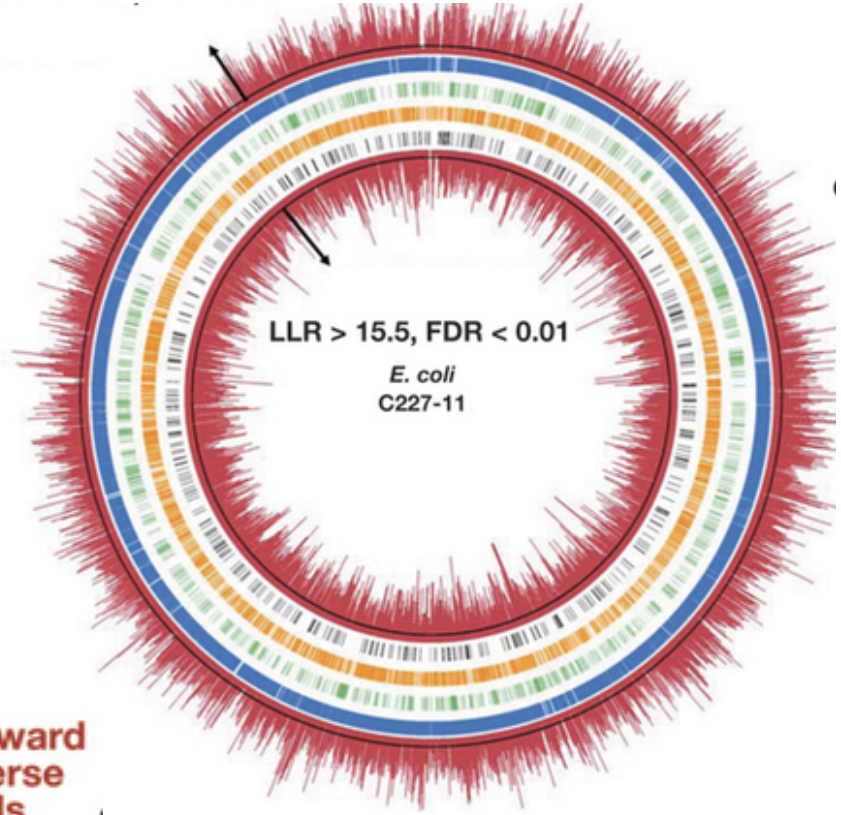
Bacteria are splattered with epigenetic marks

Genome-wide mapping of methylated adenine residues in pathogenic *Escherichia coli* using single-molecule real-time sequencing

Gang Fang, Diana Munera, David I Friedman, Anjali Mandlik, Michael C Chao, Onureena Banerjee, Zhixing Feng, Bojan Losic, Milind C Mahajan, Omar J Jabado, Gintaras Deikus, Tyson A Clark, Khai Luong, Iain A Murray, Brigid M Davis, Alona Keren-Paz, Andrew Chess, Richard J Roberts, Jonas Korlach, Steve W Turner, Vipin Kumar, Matthew K Waldor & Eric E Schadt

Affiliations | Contributions | Corresponding authors

Nature Biotechnology **30**, 1232–1239 (2012) | doi:10.1038/nbt.2432



LLRs, forward and reverse strands

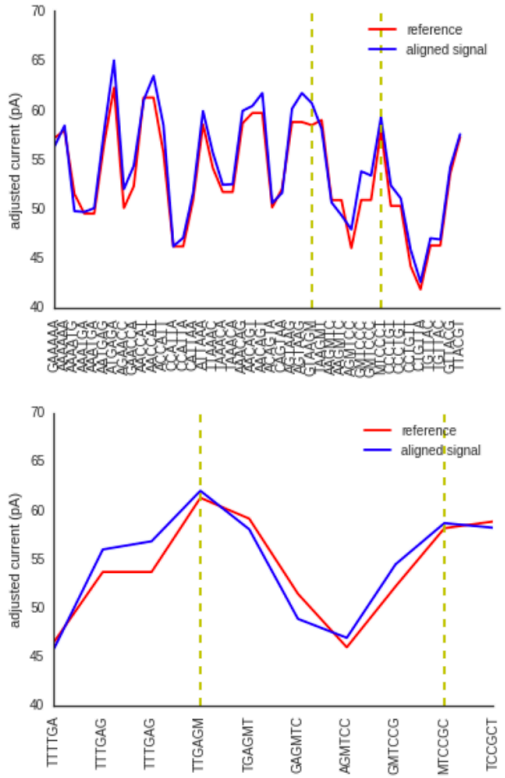
GATC

CTGCAG

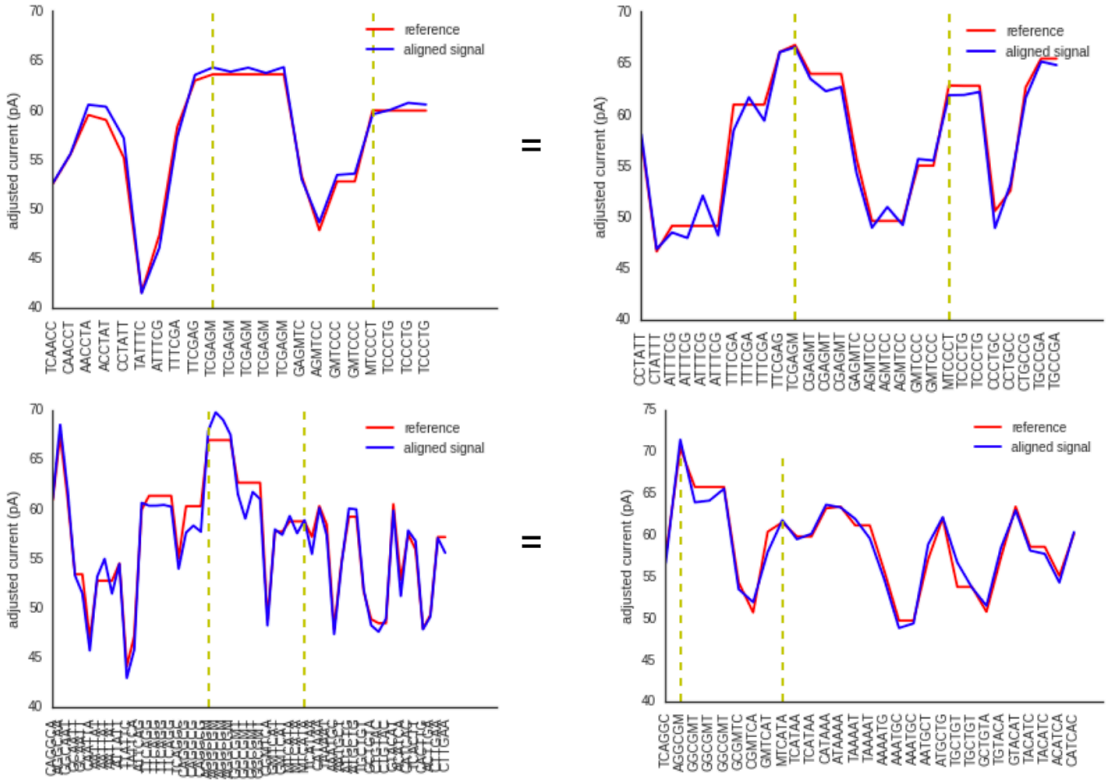
ACCACC

CCACN$_8$TGAY/R
TCAN$_8$GTGG

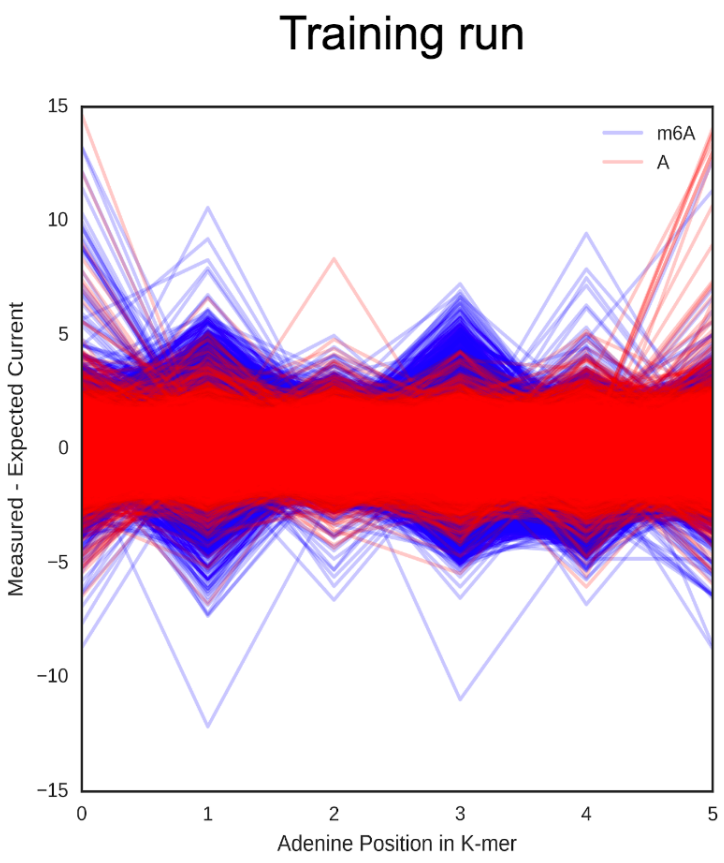
Calling current (pA) differences, similar to PacBio

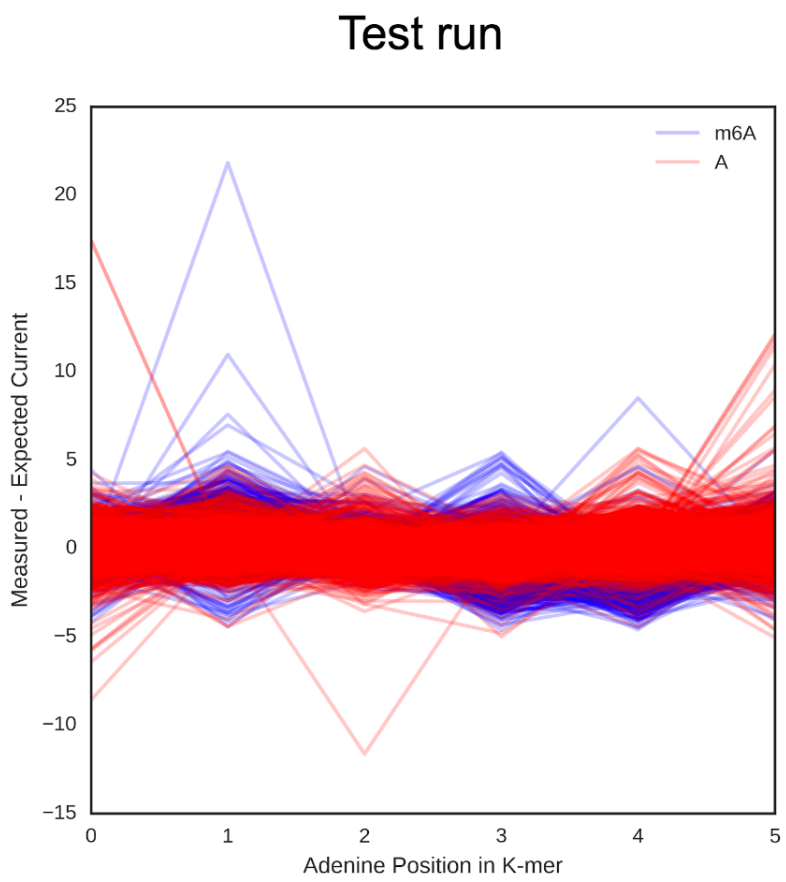


Reads aligned to same positions



Certain positions of the pore and more informative then others





GATTACA

THERE IS NO GENE FOR THE HUMAN SPIRIT

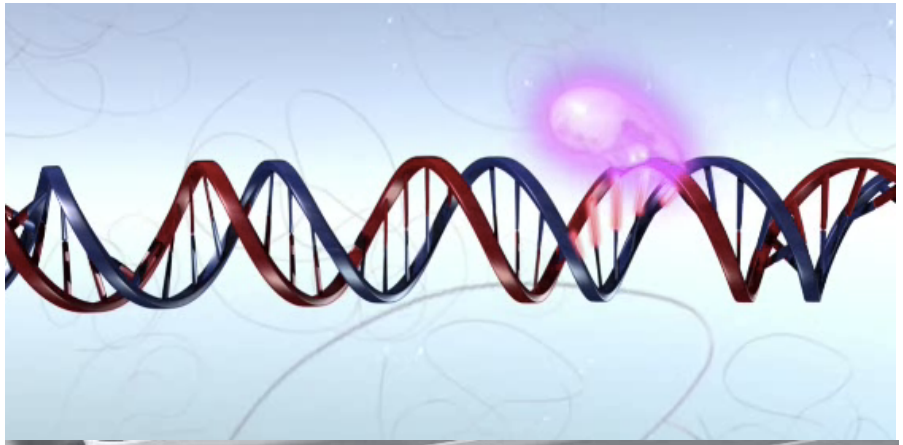
Is a 2.6 minute genome possible?
No today, but if the physics holds up…

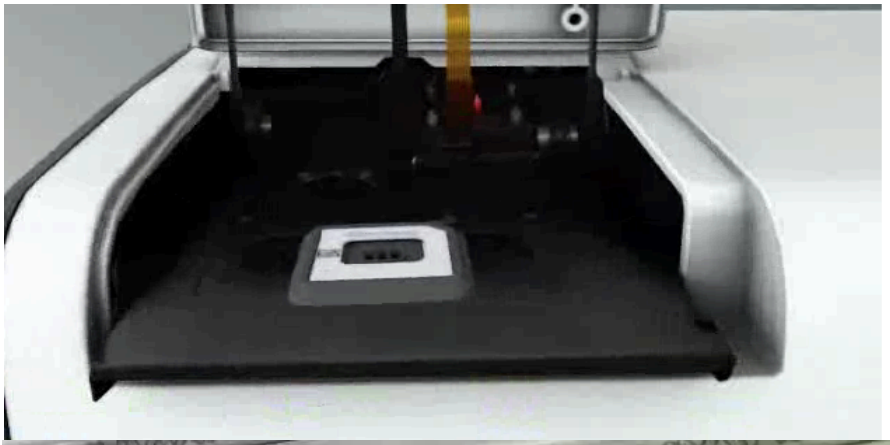
Table 2: Nanopore and Nanochannel Sequencing Considerations

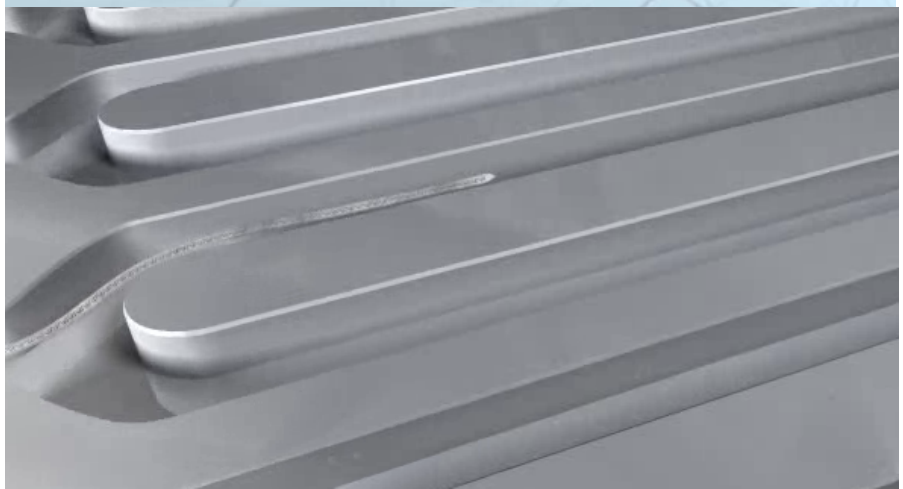
Parameter	DNA fragment (aveargebp)	Pore Speed (bp/s)	# nanopores	% of Pores Functional	transit time (seconds)	transit time (minutes)	run time (hours)	max # molecules / pore / run	% of time pores have DNA	actual # molecules/ pore/run	# of bases sequenced per device	Run Cost ($)	$ / Mb	$ / Gb	Hours for 30X WGS of 3.1Gb	Model
Time	10,000	100	512	0.5	100	1.67	6	216	80%	172.8	442,368,000	$ 1,000	$ 2.26	$ 2,260.56	1261.4	T1
	10,000	100	512	0.5	100	1.67	24	864	80%	691.2	1,769,472,000	$ 1,000	$ 0.57	$ 565.14	1261.4	T2
	10,000	100	512	0.5	100	1.67	48	1728	80%	1382.4	3,538,944,000	$ 1,000	$ 0.28	$ 282.57	1261.4	T3
Size	10,000	100	512	0.5	100	1.67	6	216	80%	172.8	442,368,000	$ 1,000	$ 2.26	$ 2,260.56	1261.4	S1
	100,000	100	512	0.5	1000	16.67	6	21.6	80%	17.28	442,368,000	$ 1,000	$ 2.26	$ 2,260.56	1261.4	S2
	1,000,000	100	512	0.5	10000	166.67	6	2.16	80%	1.728	442,368,000	$ 1,000	$ 2.26	$ 2,260.56	1261.4	S3
Size & Time	10,000	100	512	0.5	100	1.67	6	216	80%	172.8	442,368,000	$ 1,000	$ 2.26	$ 2,260.56	1261.4	S&T1
	100,000	100	512	0.5	1000	16.67	24	86.4	80%	69.12	1,769,472,000	$ 1,000	$ 0.57	$ 565.14	1261.4	S&T2
	1,000,000	100	512	0.5	10000	166.67	48	17.28	80%	13.824	3,538,944,000	$ 1,000	$ 0.28	$ 282.57	1261.4	S&T3
Pores	10,000	100	50000	0.5	100	1.67	6	216	80%	172.8	43,200,000,000	$ 1,000	$ 0.023	$ 23.15	12.9	P&T1
	10,000	100	100000	0.5	100	1.67	6	216	80%	172.8	86,400,000,000	$ 1,000	$ 0.012	$ 11.57	6.5	P&T2
	10,000	100	150000	0.5	100	1.67	6	216	80%	172.8	129,600,000,000	$ 1,000	$ 0.008	$ 7.72	4.3	P&T3
Pores & Time	10,000	100	50000	0.5	100	1.67	6	216	80%	172.8	43,200,000,000	$ 10,000	$ 0.23	$ 231.48	12.9	P&T1
	10,000	100	100000	0.5	100	1.67	24	864	80%	691.2	345,600,000,000	$ 20,000	$ 0.06	$ 57.87	6.5	P&T2
	10,000	100	150000	0.5	100	1.67	48	1728	80%	1382.4	1,036,800,000,000	$ 30,000	$ 0.03	$ 28.94	4.3	P&T3
Pores, Speed, & Time	10,000	100	50000	0.5	100	1.67	6	216	80%	172.8	43,200,000,000	$ 10,000	$ 0.23	$ 231.48	12.9	PS&T1
	10,000	1000	100000	0.5	10	0.17	24	8640	80%	6912	3,456,000,000,000	$ 20,000	$ 0.01	$ 5.79	0.6	PS&T2
	10,000	10000	150000	0.5	1	0.02	48	172800	80%	138240	103,680,000,000,000	$ 30,000	$ 0.00	$ 0.29	0.04	PS&T3

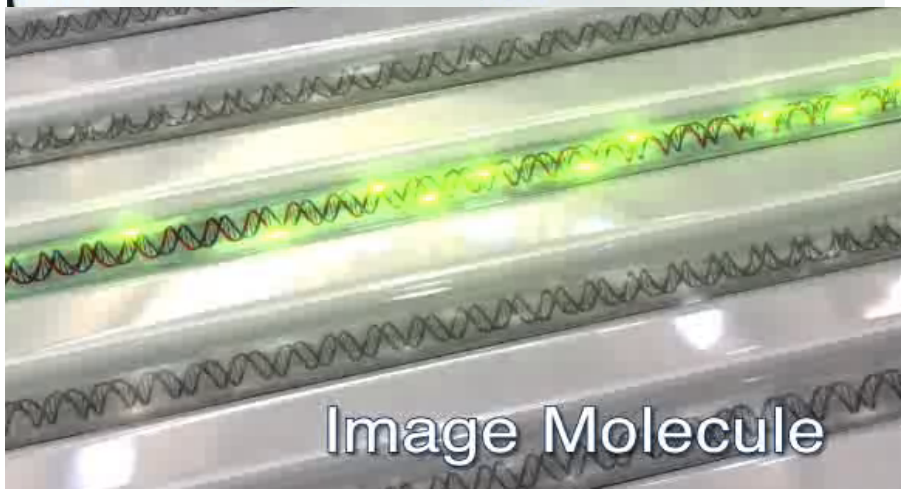
Bionanogenomics - Irys System

BIONANO GENOMICS



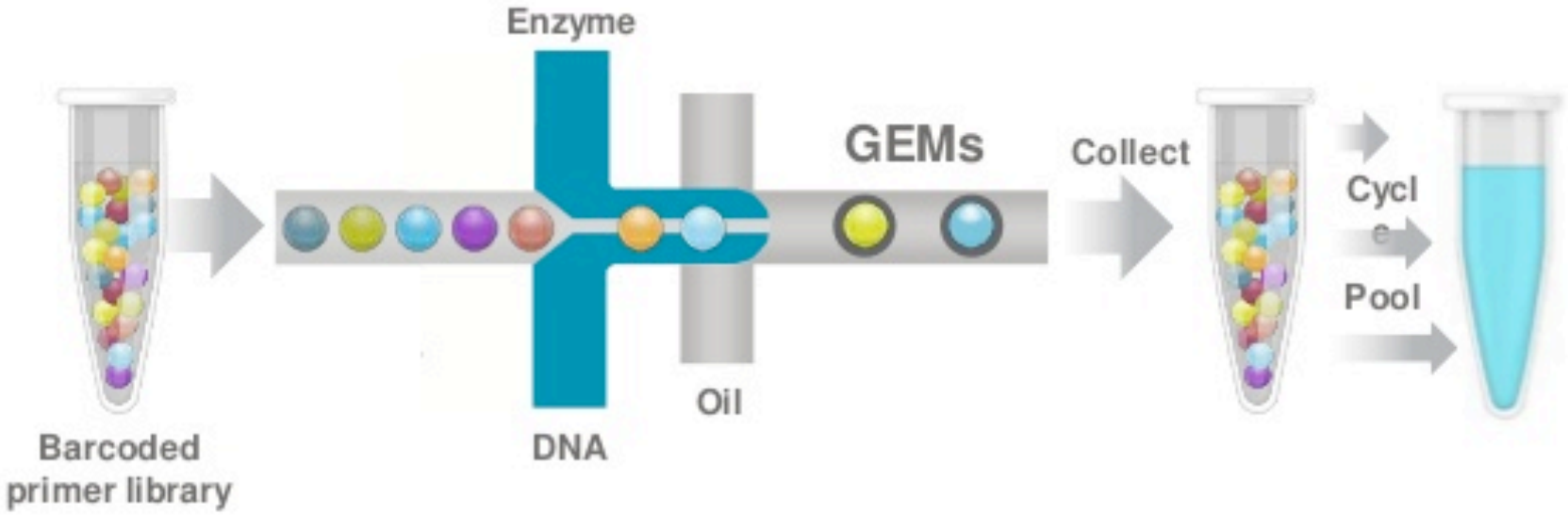






QIAGEN GeneReader

>100,000 Reactions Assembled in < 5 min



Solid phase reagent delivery → Fluid partitioning → Liquid phase biochemistry

10X GENOMICS

NATURE BIOTECHNOLOGY | RESEARCH | ARTICLE

日本語要約

Haplotyping germline and cancer genomes with high-throughput linked-read sequencing

Grace X Y Zheng, Billy T Lau, Michael Schnall-Levin, Mirna Jarosz, John M Bell, Christopher M Hindson, Sofia Kyriazopoulou-Panagiotopoulou, Donald A Masquelier, Landon Merrill, Jessica M Terry, Patrice A Mudivarti, Paul W Wyatt, Rajiv Bharadwaj, Anthony J Makarewicz, Yuan Li, Phillip Belgrader, Andrew D Price, Adam J Lowe, Patrick Marks, Gerard M Vurens, Paul Hardenbol, Luz Montesclaros, Melissa Luo, Lawrence Greenfield, Alexander Wong, David E Birch, Steven W Short, Keith P Bjornson, Pranav Patel, Erik S Hopmans, Christina Wood, Sukhvinder Kaur, Glenn K Lockwood, David Stafford, Joshua P Delaney, Indira Wu, Heather S Ordonez, Susan M Grimes, Stephanie Greer, Josephine Y Lee, Kamila Belhocine, Kristina M Giorda, William H Heaton, Geoffrey P McDermott, Zachary W Bent, Francesca Meschi, Nikola O Kondov, Ryan Wilson, Jorge A Bernate, Shawn Gauby, Alex Kindwall, Clara Bermejo, Adrian N Fehr, Adrian Chan, Serge Saxonov, Kevin D Ness, Benjamin J Hindson & Hanlee P Ji – Show fewer authors

Affiliations | Contributions | Corresponding authors

Nature Biotechnology **34**, 303–311 (2016) | doi:10.1038/nbt.3432

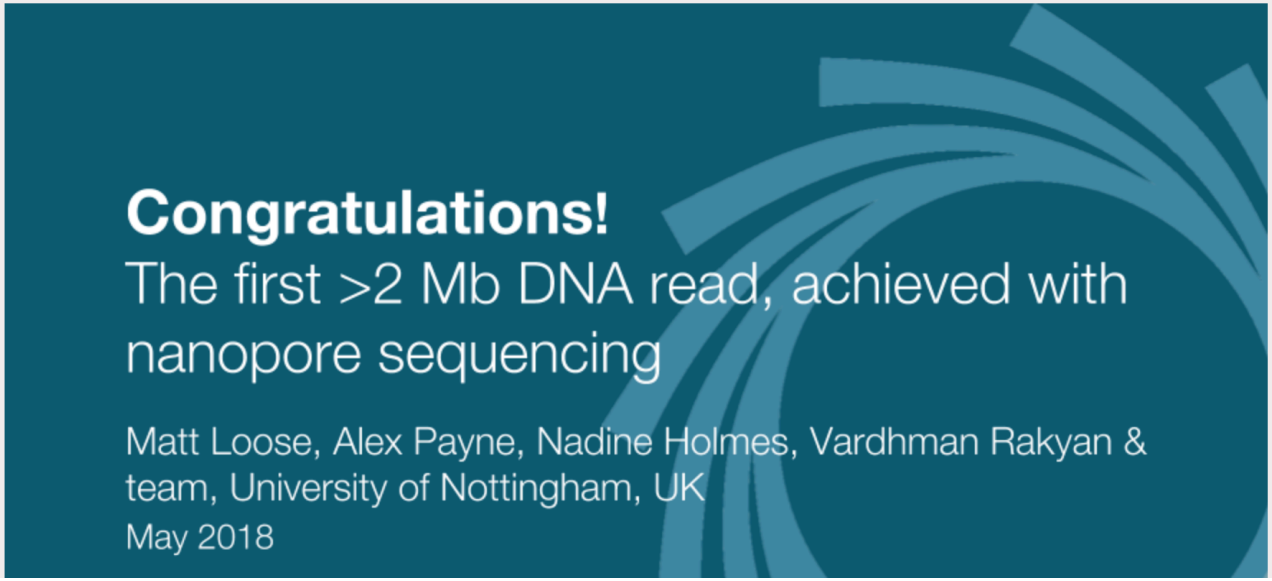
Received 16 May 2015 | Accepted 12 November 2015 | Published online 01 February 2016

Emerging Technologies

The race for long is on

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

Fri 4th 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

News

10/31/2018

BGI Unveils New High-Throughput Sequencing System.

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

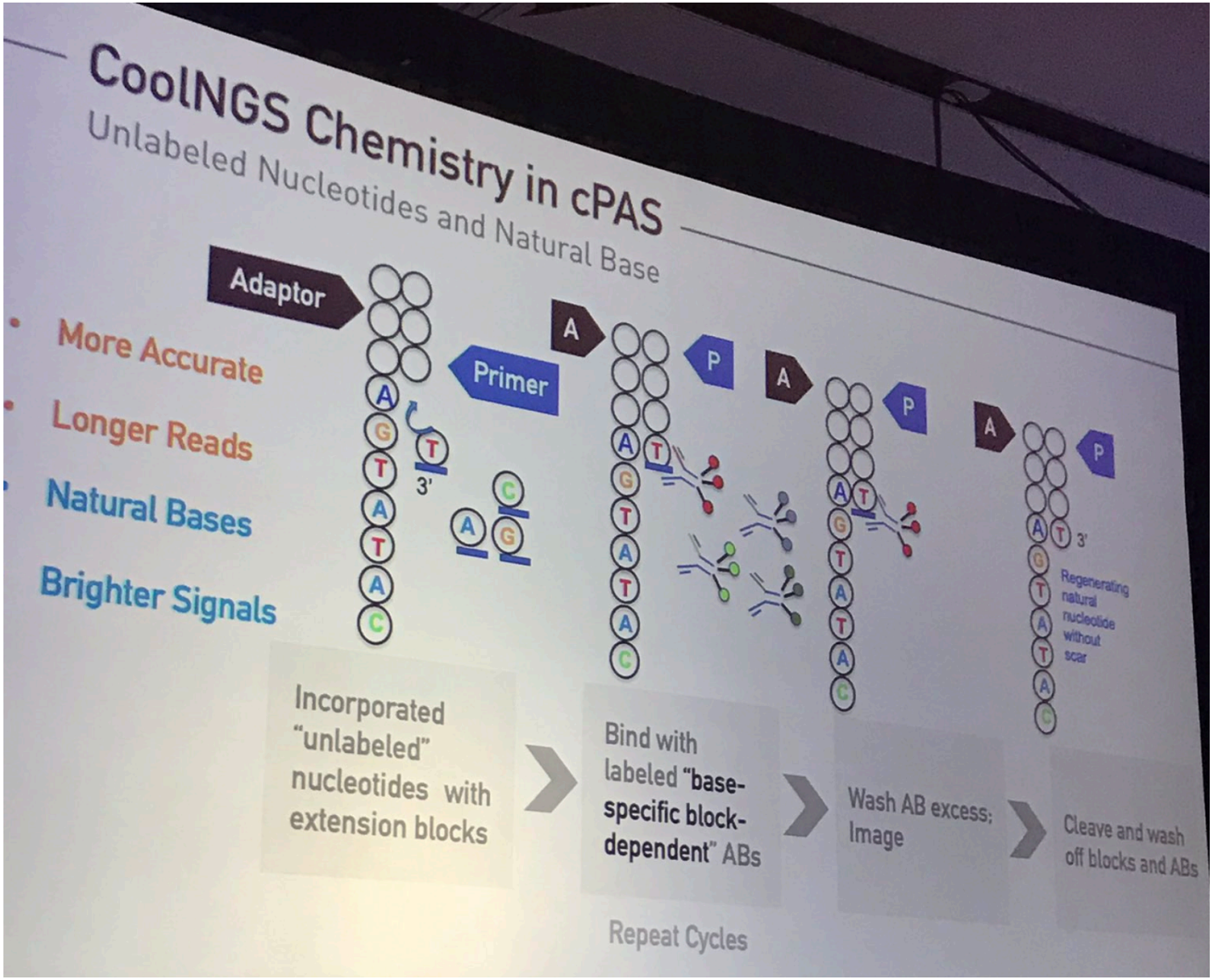
The newly unveiled MGISEQ-T7 is the most powerful sequencing system from BGI's MGI subsidiary, with a daily output capability of 6Tb of data.

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.

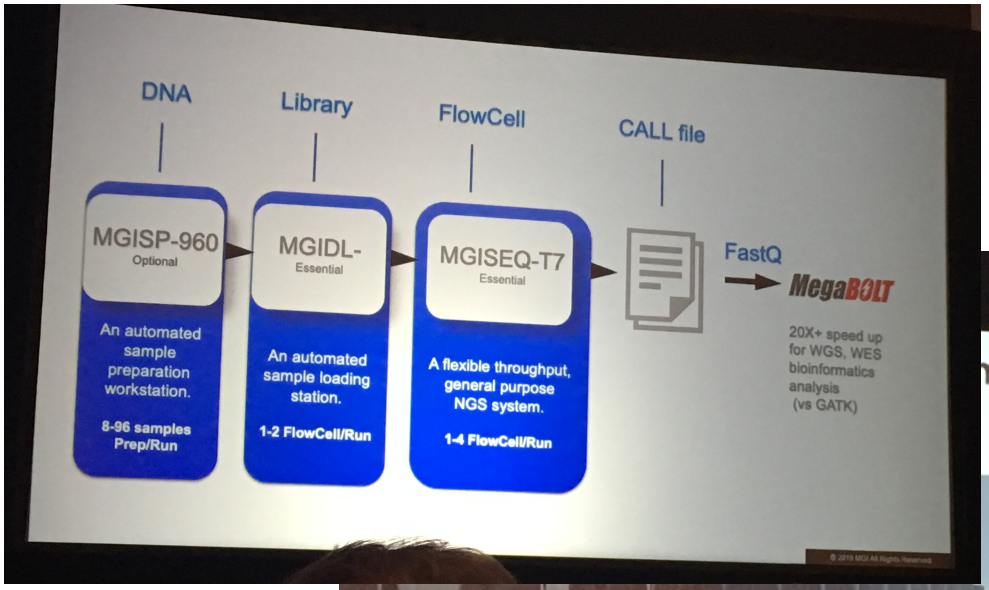
https://www.bgi.com/us/company/news/bgi-unveils-new-high-throughput-sequencing-system/

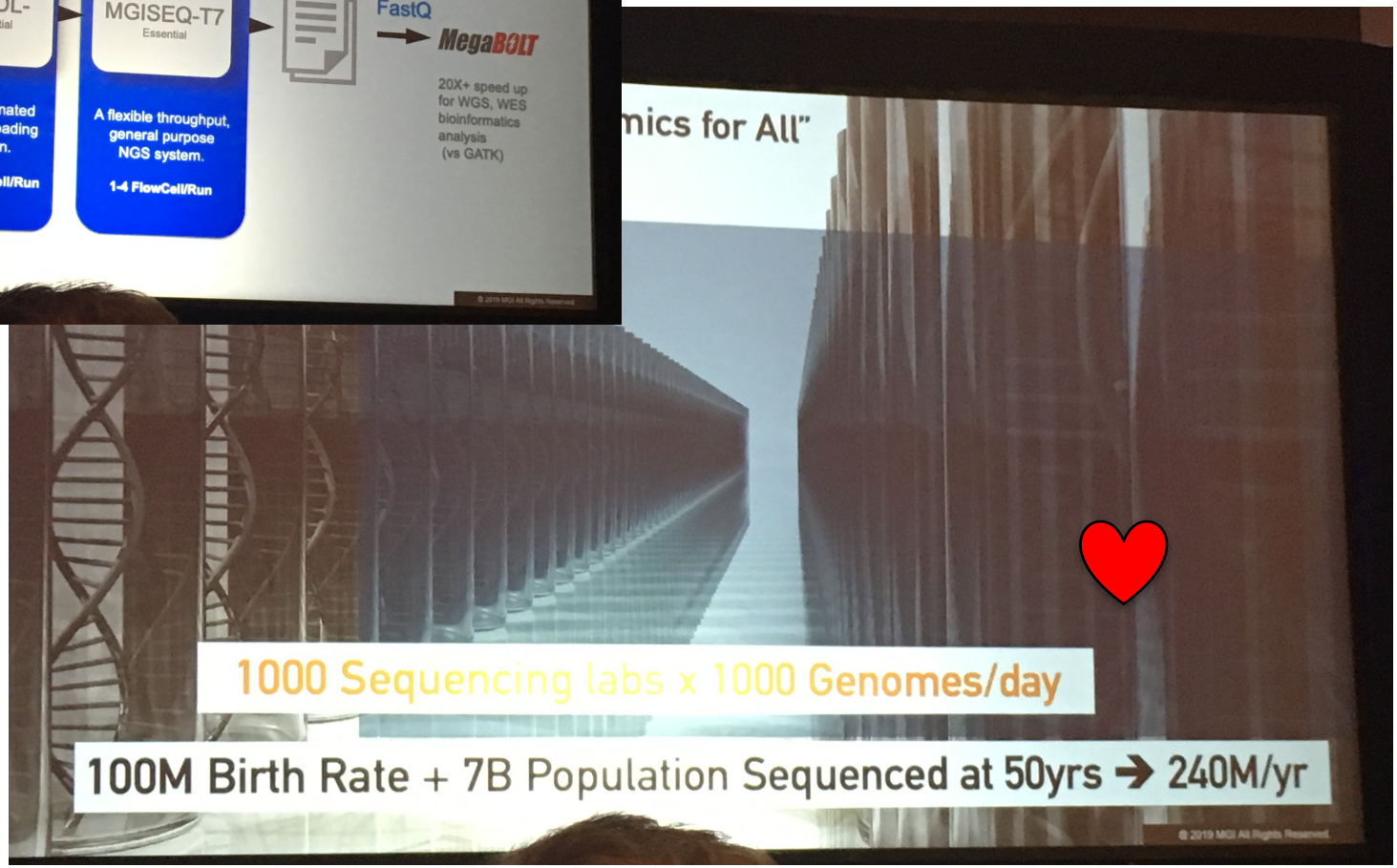
T-1000?

BGI - CoolNGS

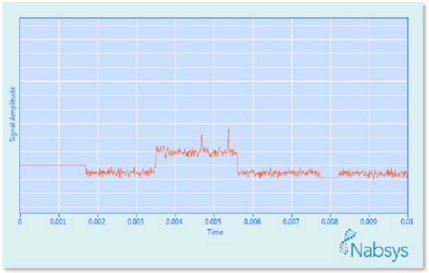

CoolNGS Chemistry in cPAS
Unlabeled Nucleotides and Natural Base
More Accurate
Longer Reads
Natural Bases
Brighter Signals
Adaptor
Primer
3'
A
P
Regenerating natural nucleotide without scar
Incorporated "unlabeled" nucleotides with extension blocks
Bind with labeled "base-specific block-dependent" ABs
Wash AB excess; Image
Cleave and wash off blocks and ABs
Repeat Cycles

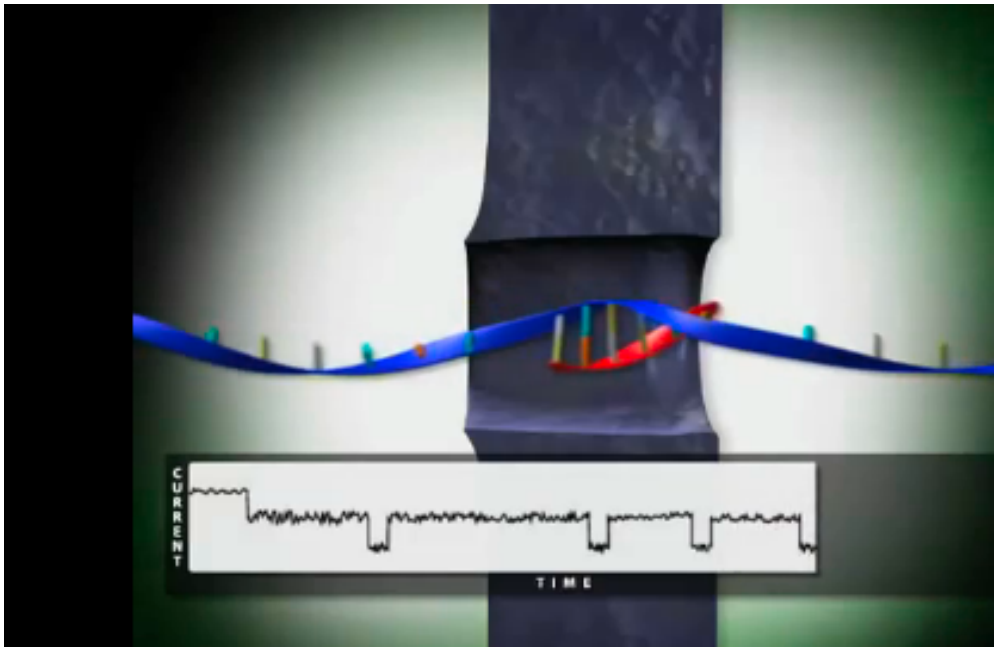
BGI – NGS streets

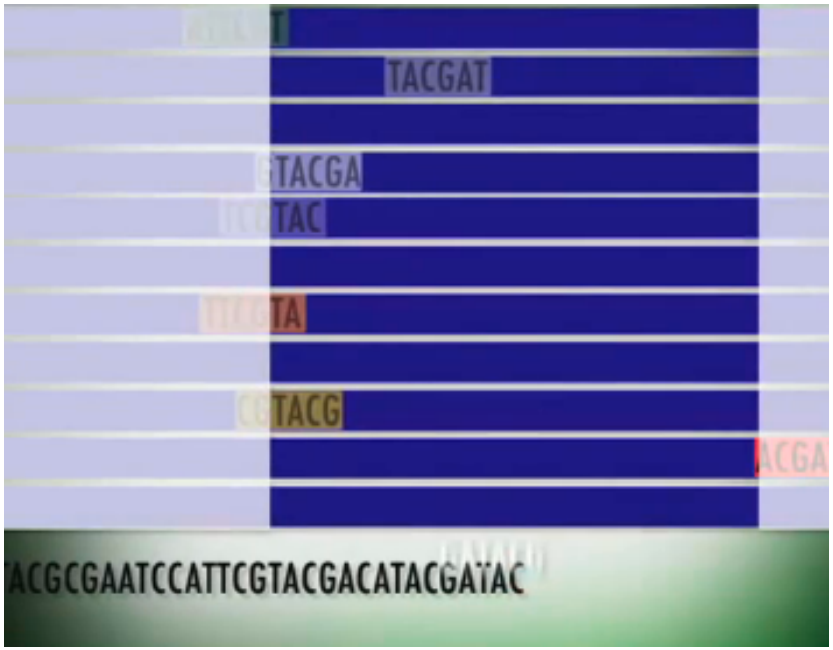




Nabsys







Hybridization -Assisted Nanopore Sequencing (HANS):

-1 million bases per second
-Variable probe length can be used for HANS
-Long Reads (100kb)
-Single molecule

ZS Genetics, Inc.

Working At The Scale Of Life

Single-atom labeling and then visualization with EM

-Long Reads (20kb)
-Single molecule

The new Illumina Firefly (iSeq100) can sequence in <6h.

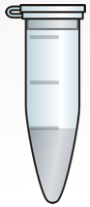
Nanostring's Hyb & Seq

Simple Workflow

No library preparation or amplification required

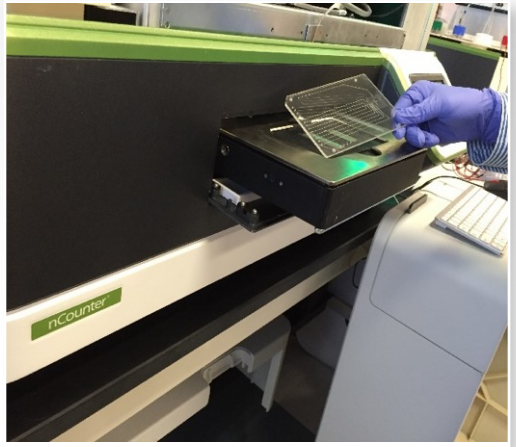
<30 minutes of hands on time Flexible input type (tissue, swabs, cells, etc.)

Single Tube Assay



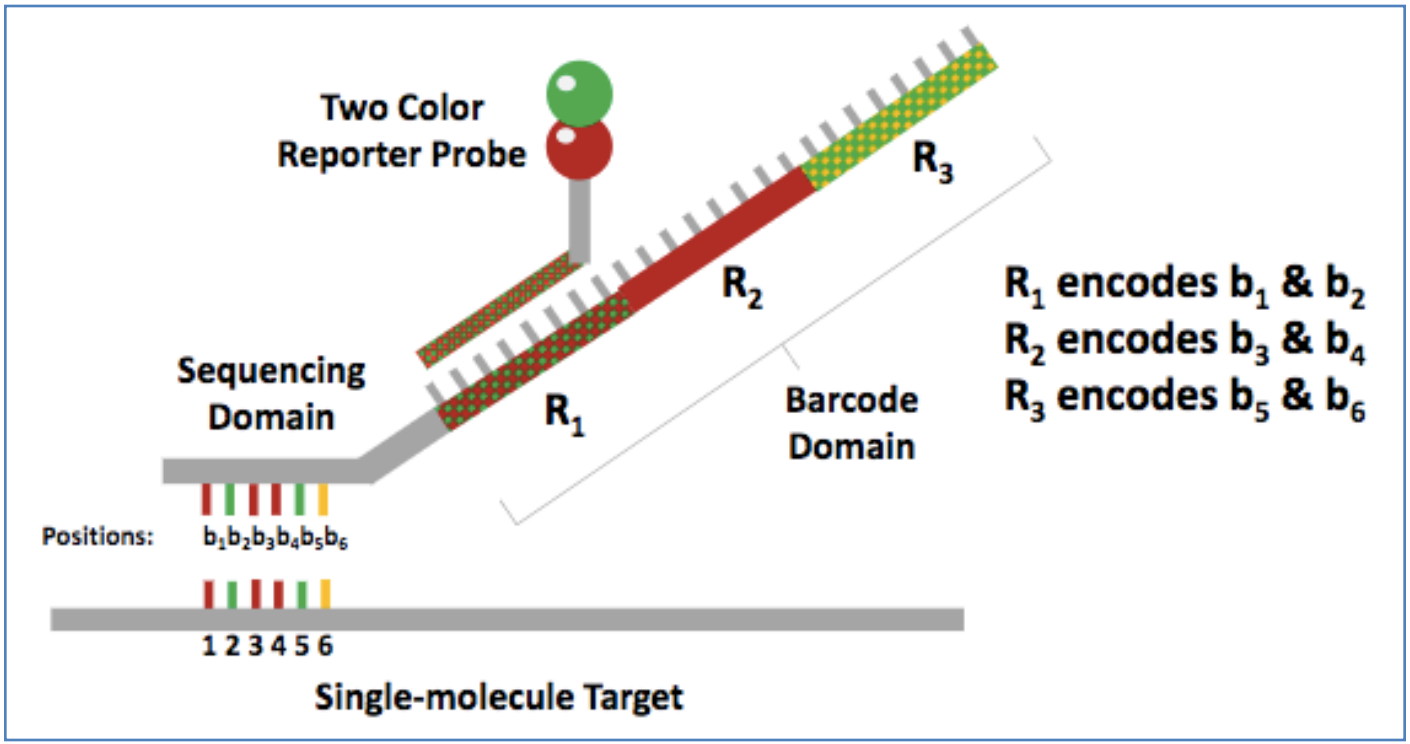
Enabling simultaneous and direct DNA and DNA sequencing

Clinically-relevant Timeframe



Sample-to-results in 4 hrs

Hyb & Seq



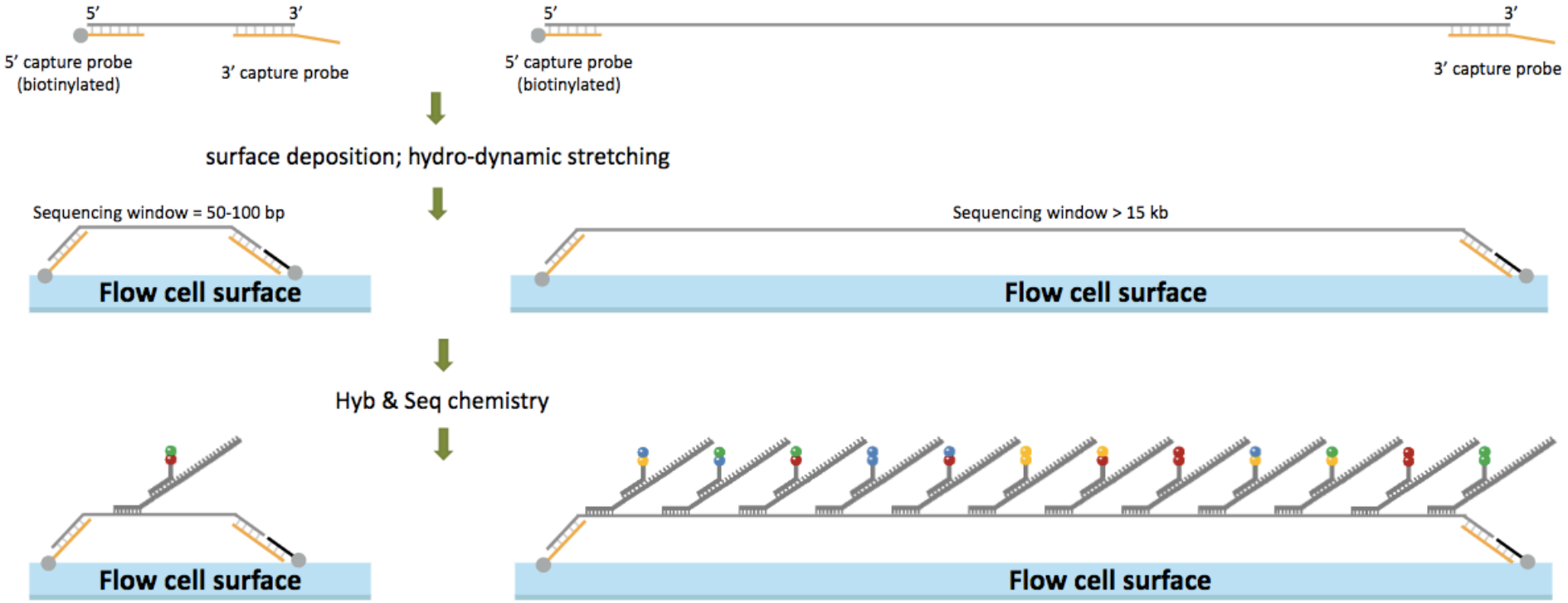
Sequencing Probes:

- Sequencing domain base-pairs with single-molecule target
- Barcode domain has three regions (R_1, R_2, R_3) encoding hexamer sequence
- Set of 4096 sequencing probes enables sequencing of any target sequence

Two-color Reporter Probes:

- Three reporter probes bind sequentially to barcode domain (R_1, R_2 and R_3)
- Each reporter probe represents a dinucleotide sequence

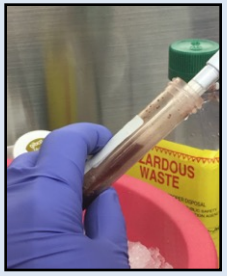
Long and Short reads possible (up to 33kb)



Clinical Sample Processing

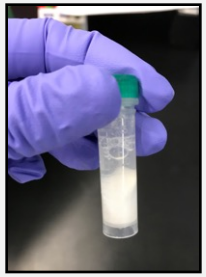
START

Clinical sample



STEP 1

Lyse sample (bead beat in lysis solution)



Total time:
15 min
Hands-on time:
5 min

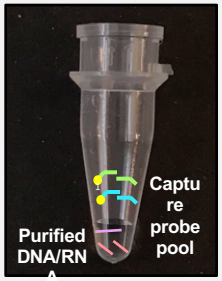
STEP 2

Cell debris removal, Nucleic acid concentration

Total time:
45 min
Hands-on time:
30 min

STEP 3

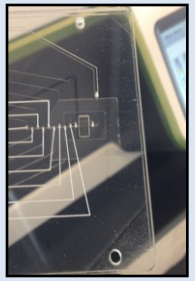
Multi-plex capture of DNA+RNA



Total time:
30 min
Hands-on time:
15 min

FINISH

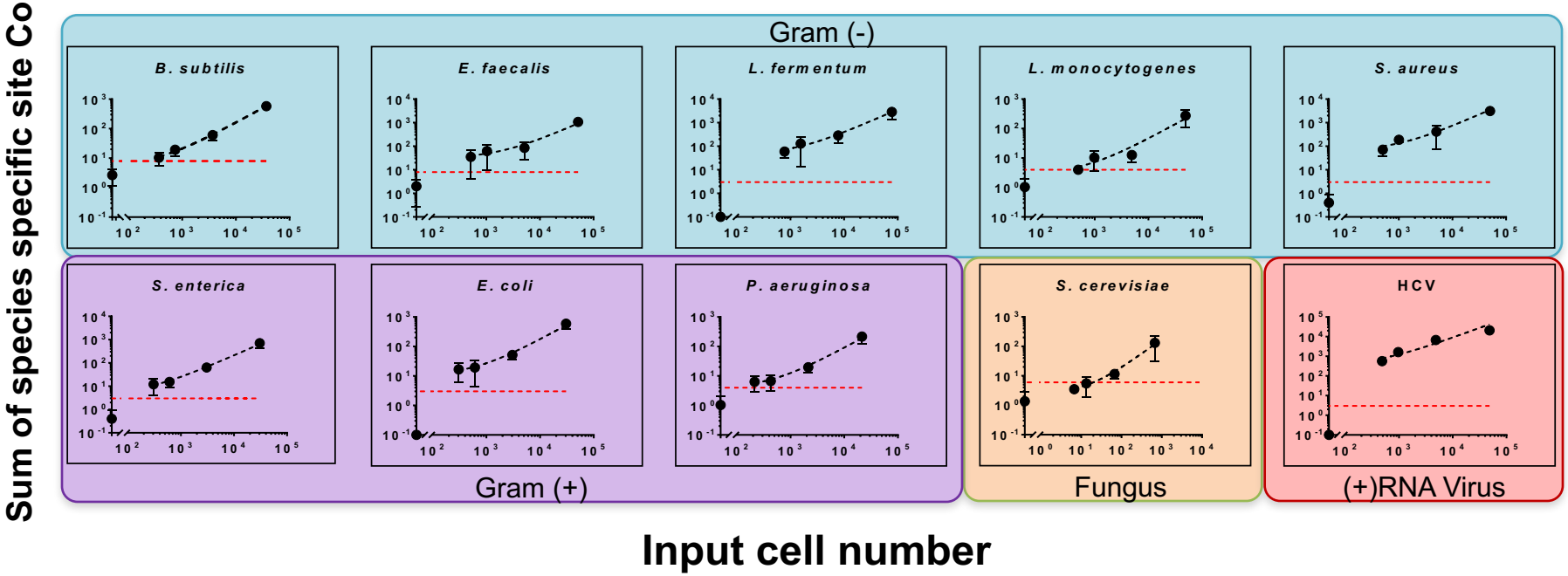
Purify and bind DNA and RNA target on flow cell



Load on to sequencer

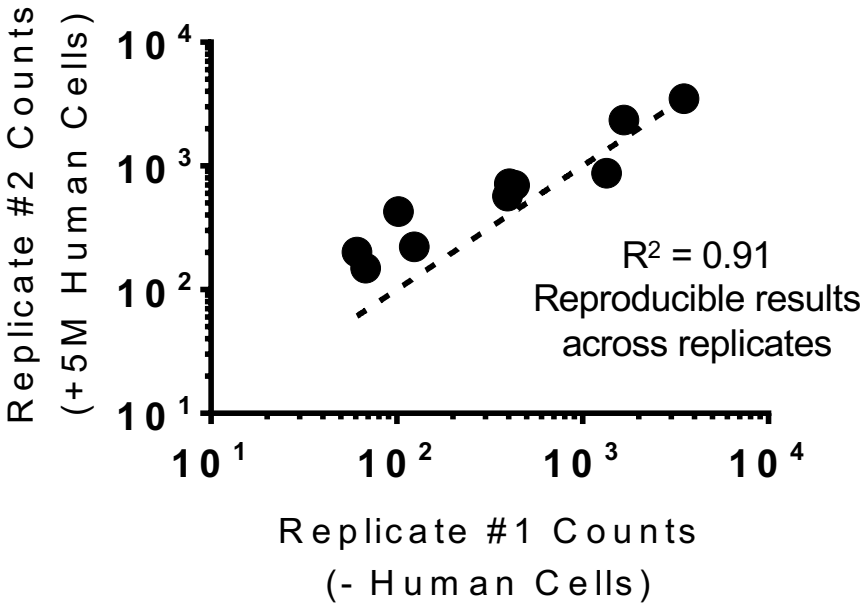
Completed in 90 min
No amplification, No library preparation

Assay Validation: Limit of Detection



Hyb & Seq simultaneously detected 10 pathogens at ≤ 1000 cells/ml from a same sample using a single tube assay

Assay Validation: No cross reactions with human DNA



- Amplification-free sequencing of pathogens even in the presence of human cell background (5 million cells, cell line GM19240/NA12878)
- High concordance of sequencing results with or without excess of human cells background
- Same workflow regardless of sample background (swab, cells, tissue, etc)
- Eliminates reads waste due to carrier human DNA/RNA

Clinical samples from WCM

Sample Name	Site	Final microbiology report
WCM300	Head Epidural Fluid	Sparse P. aeruginosa, Sparse Enterococcus faecalis
WCM301	Spleen	Sparse E. coli, Sparse Proteus mirabilis, Sparse Lactobacillus sp. (no final speciation)*
WCM302	R tibia	Sparse MRSA
WCM303	R leg wound	MSSA
WCM304	R 3rd metatarsal	Sparse Proteus mirabilis, Few Staphylococcus agalactiae, Sparse MSSA
WCM305	L thigh wound	MSSA
WCM306	Lung	Many Pseudomonas aeruginosa

With Lars Westblade

Precision Clinical Metagenomics
IRB#: 1606017347

Hyb & Seq Sequencing Results

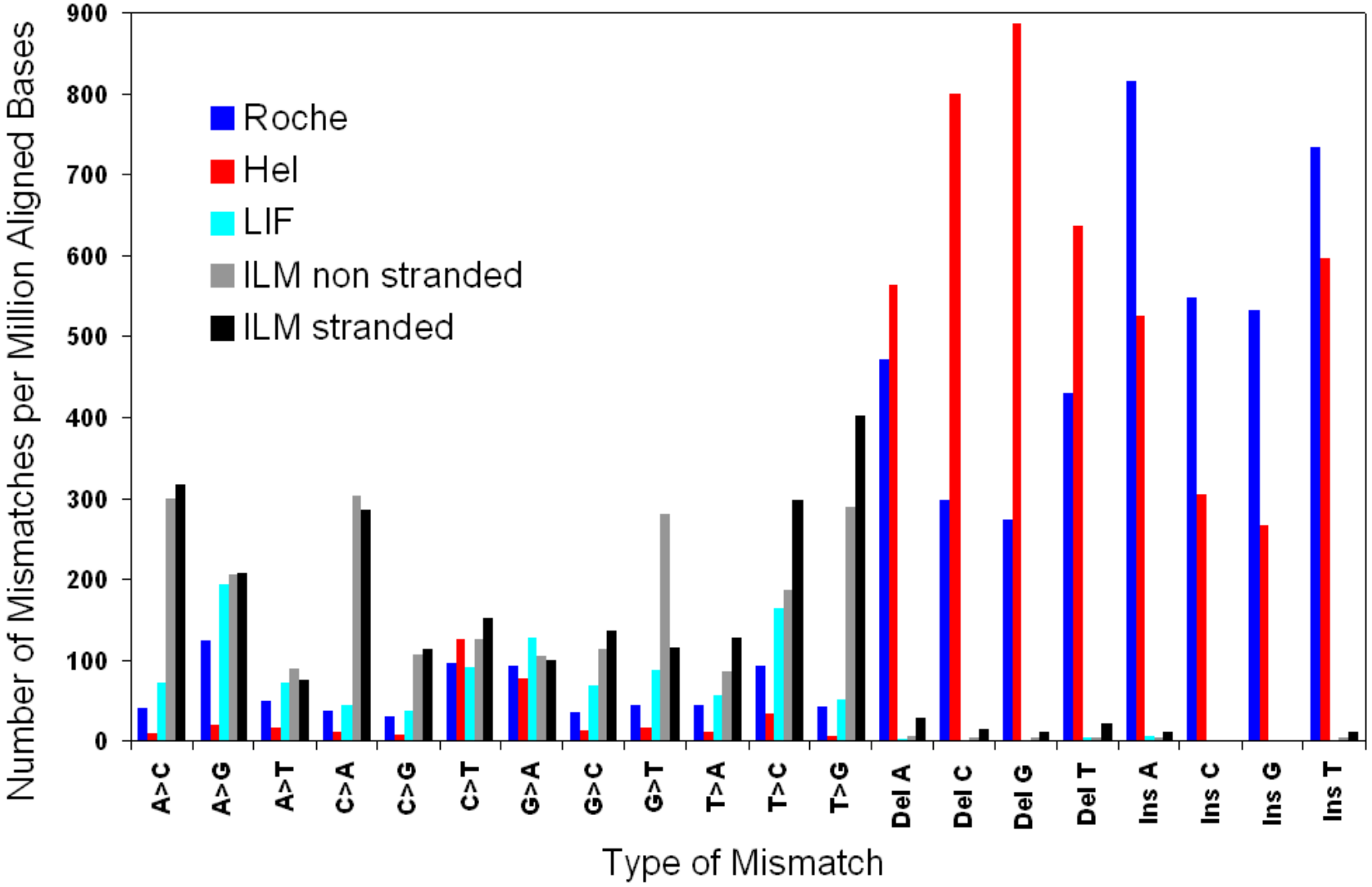
	WCM301 Spleen			WCM302 Tibia			WCM303 Leg Wound			WCM304 3rd Metatarsal			WCM305 Thigh Wound			WCM306 Lung		
	Lab	qPCR	H&S	Lab	qPCR	H&S	Lab	qPCR	H&S	Lab	qPCR	H&S	Lab	qPCR	H&S	Lab	qPCR	H&S
Lactobacillus fermentum	+		+	-		-	-		-	-		-	-		-	-		-
Pseudomonas aeruginosa	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+
Staphylococcus aureus	-	-	-	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-

- Six different clinical samples were analyzed
- Five positive calls across three kingdom of organisms
- High concordance with **pathology lab analysis (98%; 65/66)** and **100% concordance with PCR** analysis
- Simultaneously detected intra- and inter-species DNA and RNA
- *One discordant same was only found in the broth and flagged as ambiguous

Each Platform has various sources of noise, and thus Error

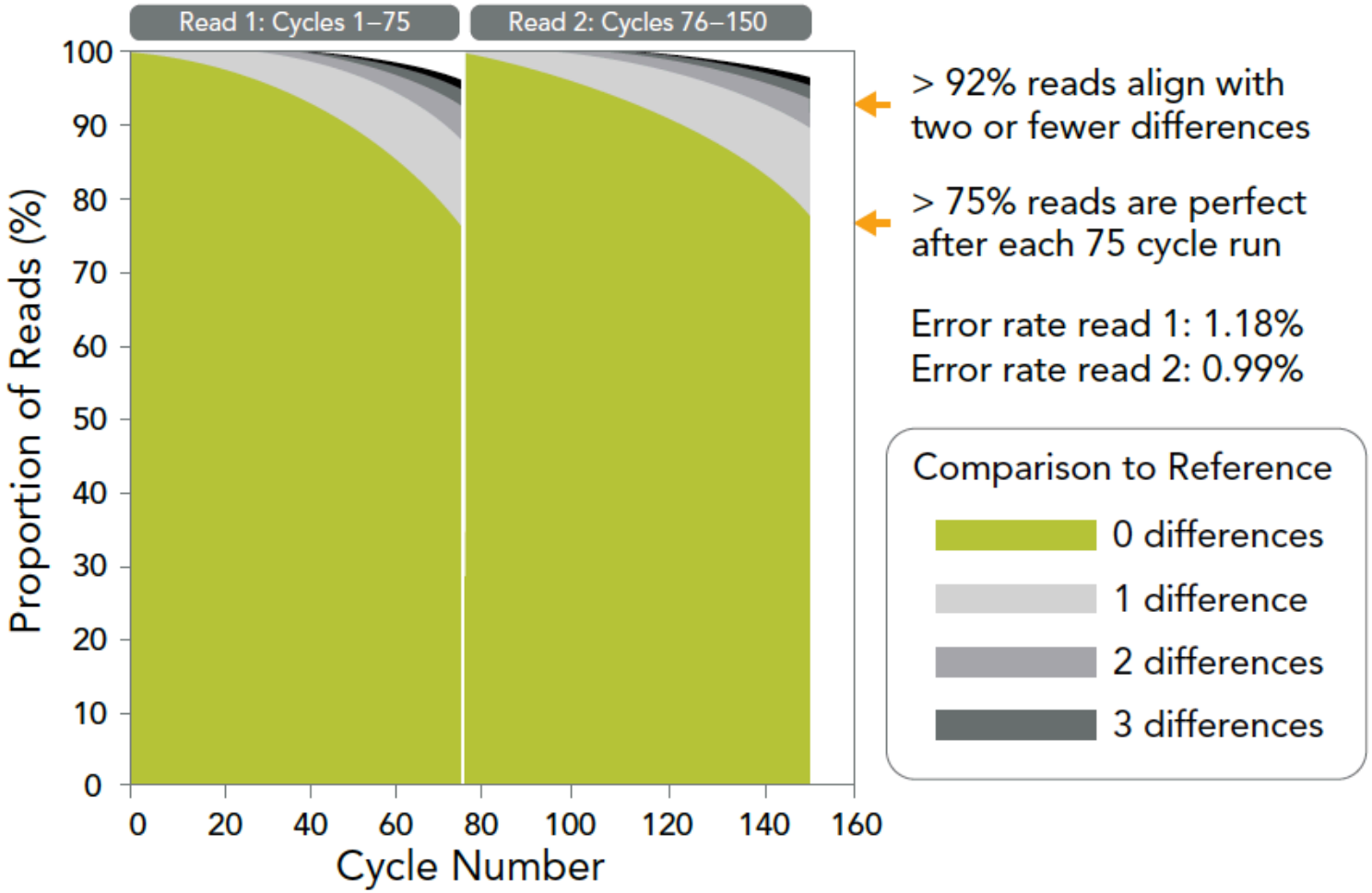
- De-Phasing
  - Lagging strand dephasing from incomplete extension
  - Leading strand dephasing from over-extension
- Dark Nucleotides
- Polymerase errors (10^{-5} to 10^{-7})
- Single molecule challenges
  - High noise
  - Polymerase “wiggling” from tail
- Platform-specific errors
  - Illumina more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

Each platform is slightly different, and so intrinic errors are different

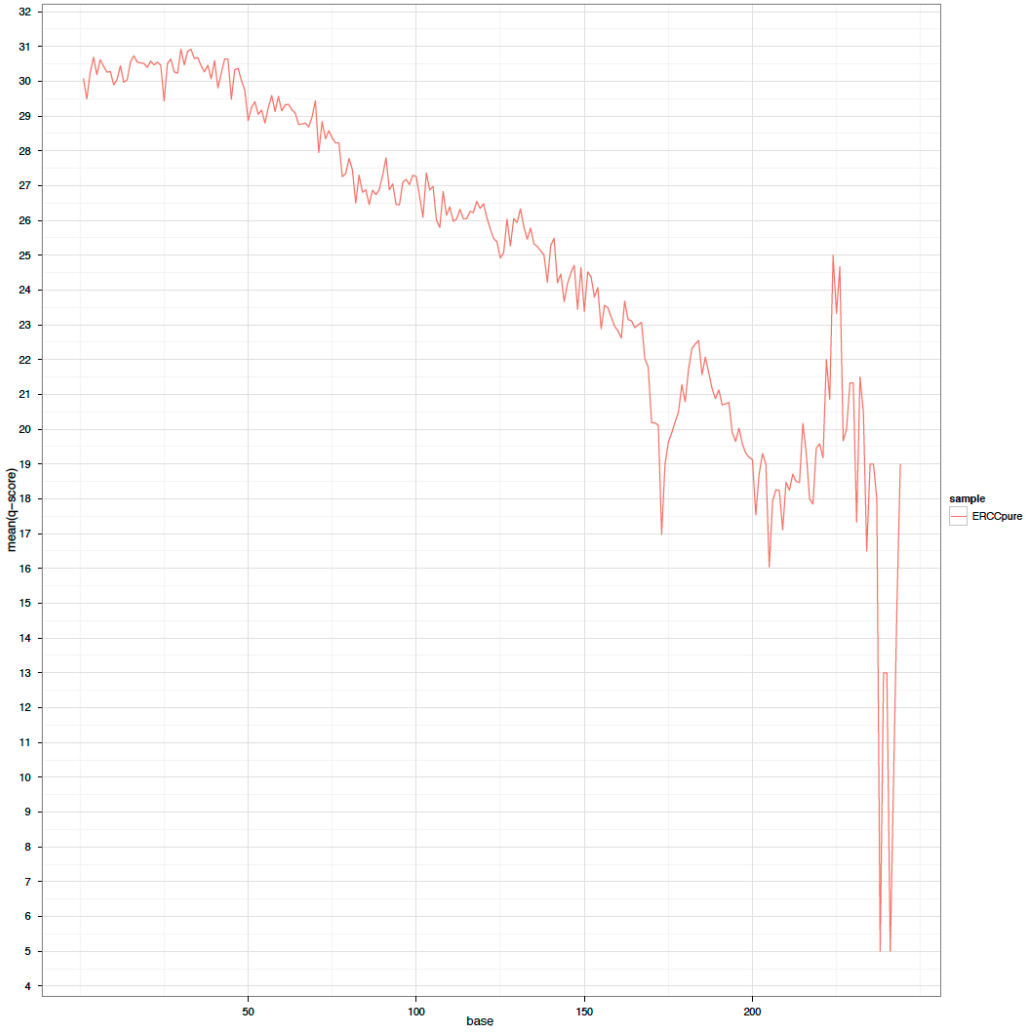


SeQC Consortium

Many platforms are cycle-dependent on error rate - ILMN

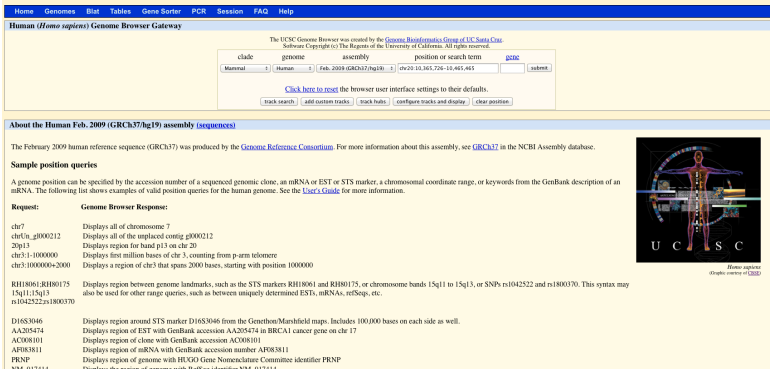


Many platforms are cycle-dependent on error rate - ION



Plan ahead for all genomes to be sequenced and available

However, your internet browser home page will likely change:



Single cells

Used to be very hard to look at individual cells

Proceedings of the National Academy of Sciences of the United States of America

PNAS

> Current Issue > vol. 110 no. 2 > Xinghua Pan, 594–599, doi: 10.1073/pnas.1217322109

CrossMark click for updates

Two methods for full-length RNA sequencing for low quantities of cells and single cells

Xinghua Pan[a,1], Russell E. Durrett[b,2], Haiying Zhu[a,c,2], Yoshiaki Tanaka[a,2], Yumei Li[a,d], Xiaoyuan Zi[a], Sadie L. Marjani[a], Ghia Euskirchen[e], Chao Ma[f,g], Robert H. LaMotte[f], In-Hyun Park[a], Michael P. Snyder[e], Christopher E. Mason[b], and Sherman M. Weissman[a,1]

Author Affiliations

Contributed by Sherman M. Weissman, October 8, 2012 (sent for review August 22, 2012)

This Issue

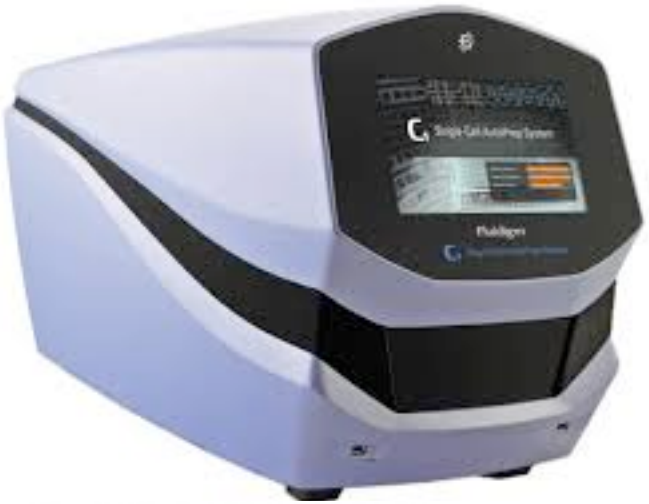
January 8, 2013
vol. 110 no. 2
Masthead (PDF)
Table of Contents

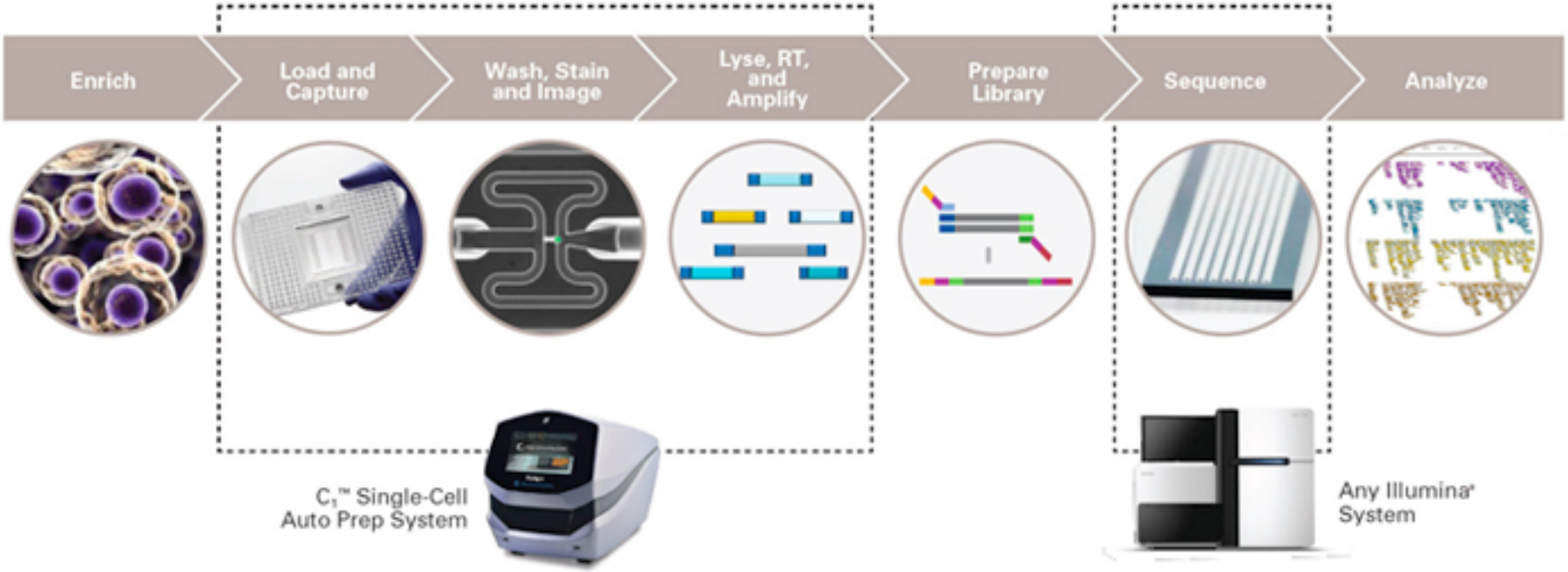
PREV ARTICLE NEXT ARTICLE

Don't Miss

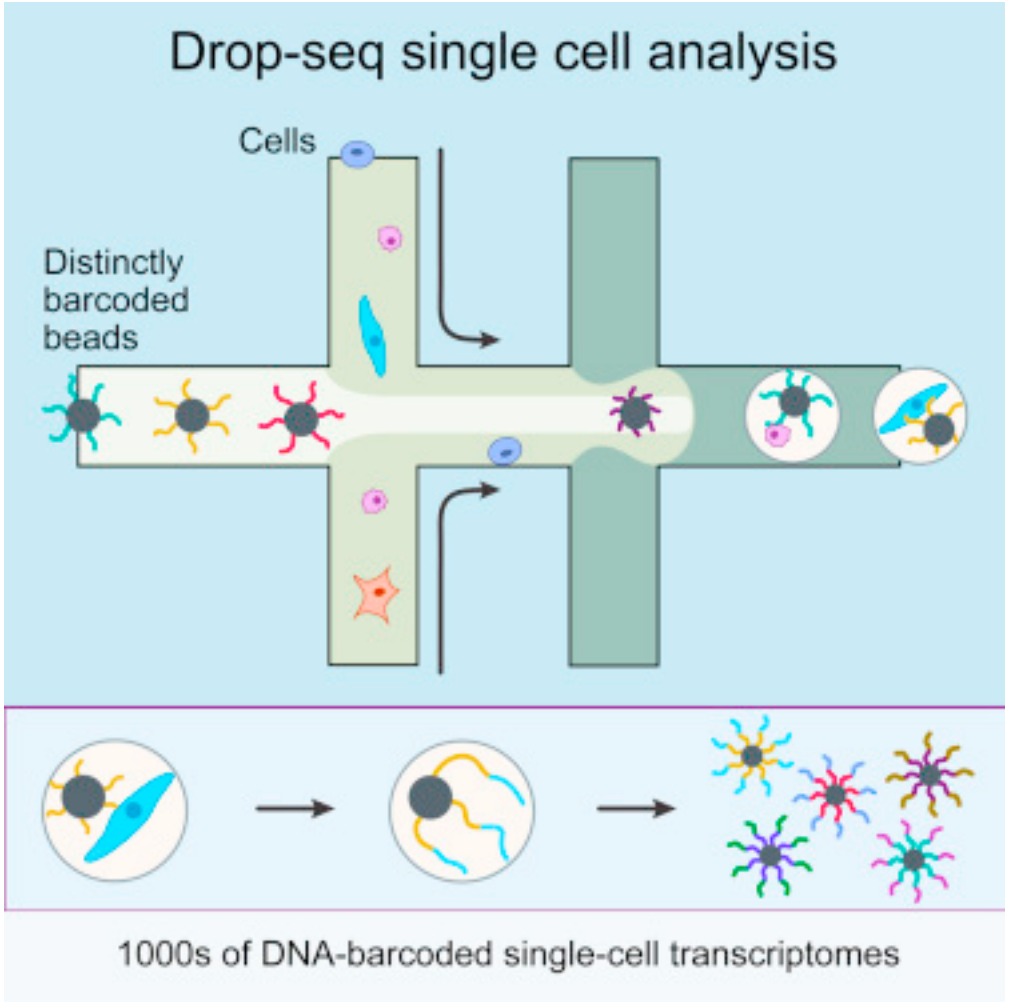
Customize your free PNAS alerts to receive a notification as new content becomes available.

But now it’s very easy – Fluidigm C1





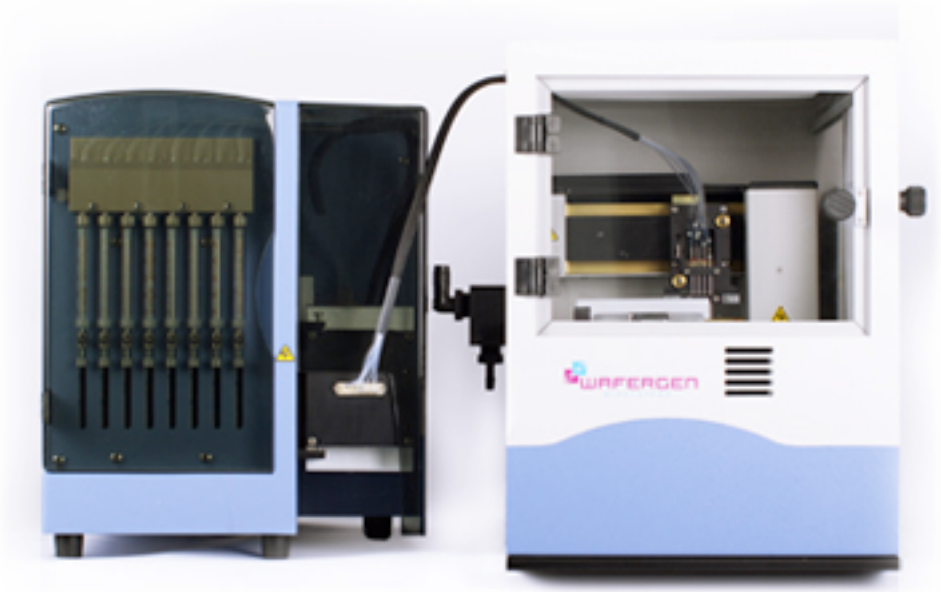
Drop-Seq



http://mccarrolllab.com/dropseq/

http://www.cell.com/abstract/S0092-8674%2815%2900549-8

WaferGen iCell8



BioRad QX200 & ILMN system

QX200™ Droplet Digital™ PCR System

Chromium NGS

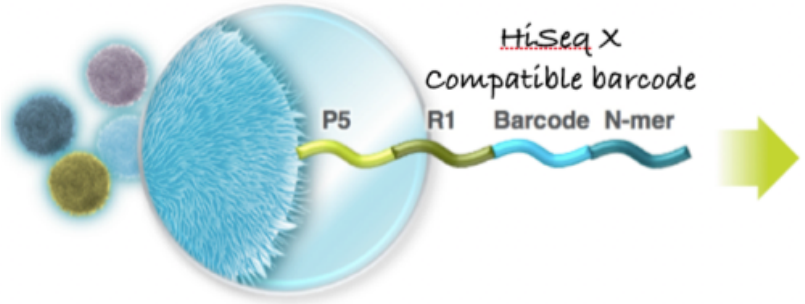
CHROMIUM™

Whole Genome Sequencing

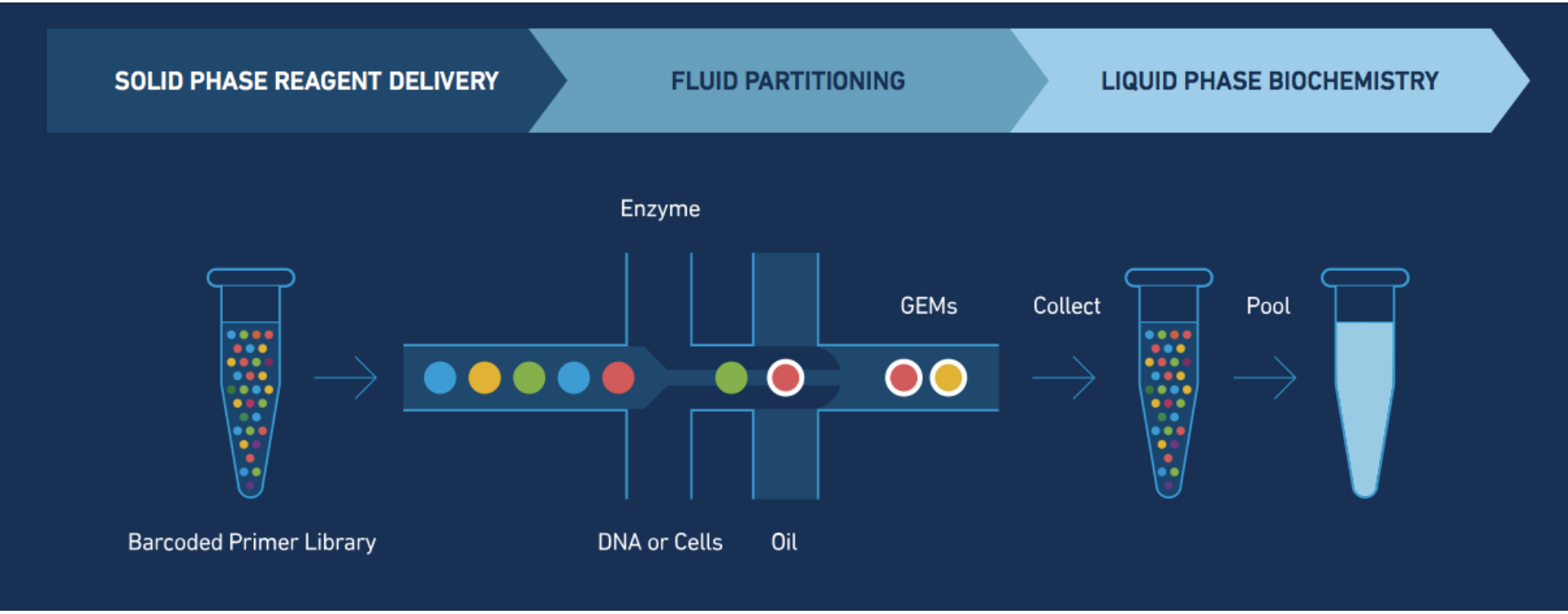
The upgraded Chromium product suite includes solutions for whole genome sequencing, exome sequencing and single-cell transcriptomics. Resolve phasing, structural variants and variants in previously inaccessible parts of the genome using the Chromium Whole Genome Sequencing Kit.

Product Features:

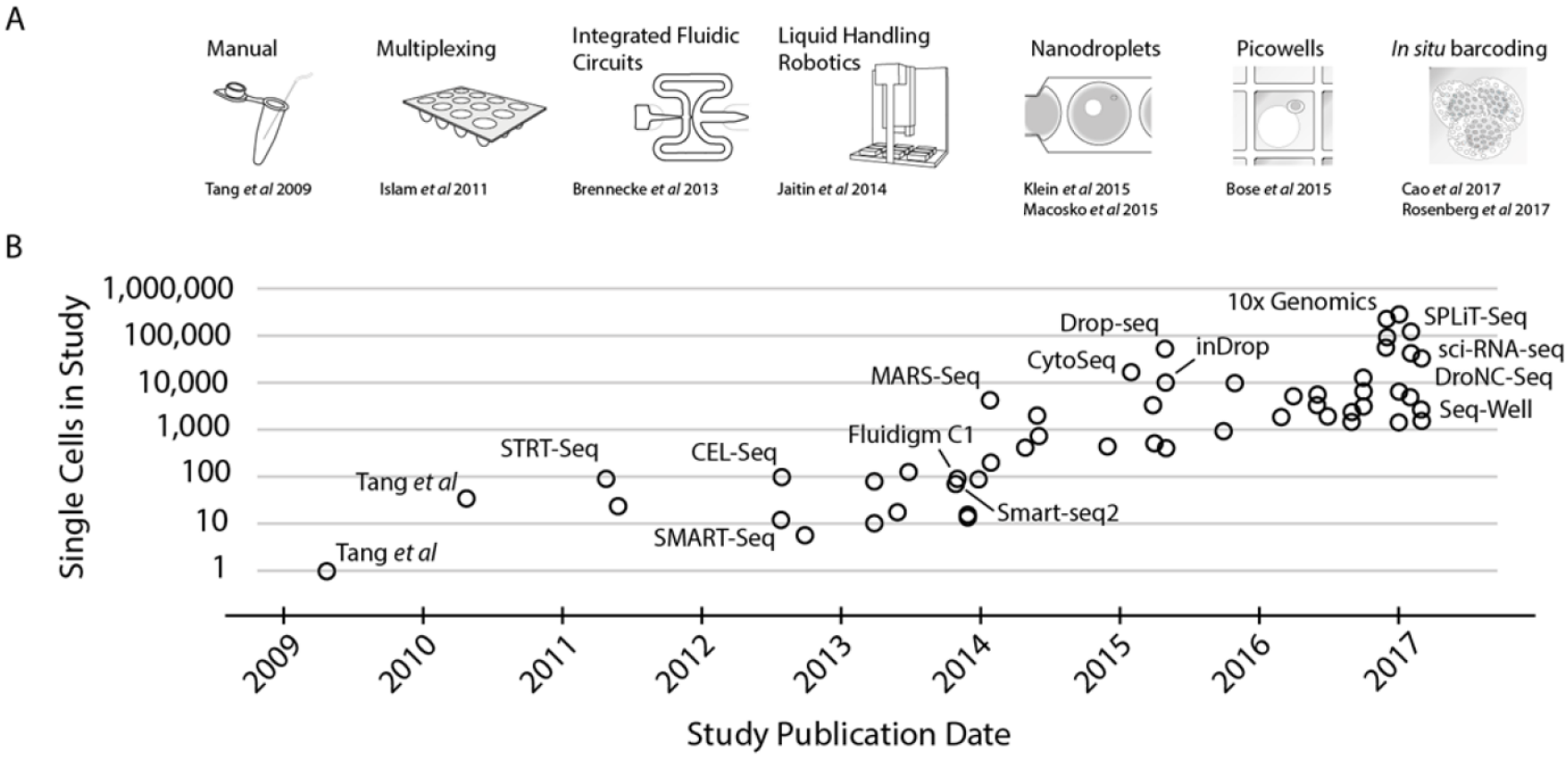
Reagent Kit Contents:



10X Genomics Single-Cell



The explosion of scRNA-seq experiments



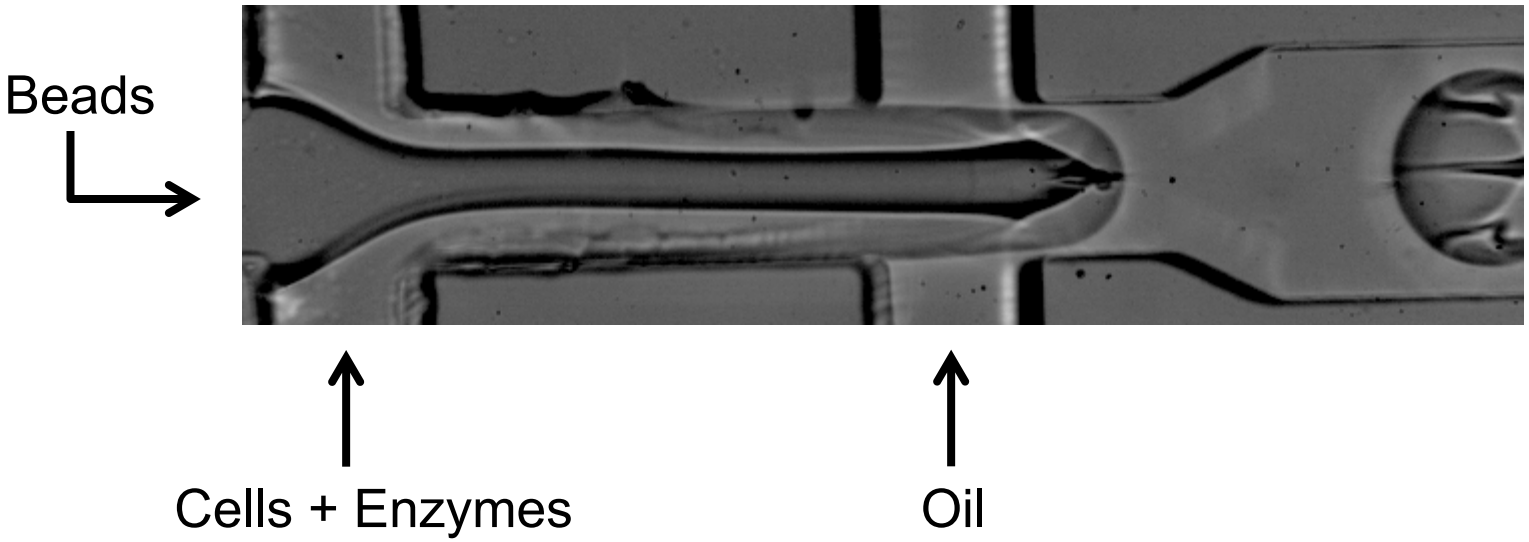
Svennson *et al.*, arXiv 2017

Many options today for single-cell sequencing

Source	Instrument	Number of Cells	input cells	est. cost per run	est. cost per cell	UMIs	Cell Phenotype	DNA	RNA	ATAC	3'	full cDNA	Size Rangeμ(m)
10X Genomics	Chromium	5,000	100,000	$ 1,290	$ 0.26	yes	no	yes	yes	yes	yes	no	1_60
Becton Dickinson	CSseq / BDPrecis	96	unk	$ 10,000	$ 104.17	yes	no	unk	unk	unk	unk	unk	5-100
Becton Dickinson	Resolve	10,000	50,000	$ 10,000	$ 1.00	yes	yes	unk	unk	unk	unk	unk	5-100
BioRad-ILMN	ddSeq	1,200	10,000	$ 1,200	$ 1.00	unk	no	no	yes	unk	unk	unk	unk
Drop-Seq	DropSeq	10,000	100,000	$ 1,000	$ 0.10	yes	no	no	yes	yes	yes	no	1-100
Fluidigm	C1	96	5,000	$ 1,900	$ 19.79	yes	no	yes	yes	yes	no	yes	5-10, 11-17, 17-24
Fluidigm	scRRBS	96	5,000	$ 1,900	$ 45.00	yes	no	yes	yes	yes	no	yes	5-10, 11-17, 17-24
Fluidigm	C1- high throughp	800	5,000	$ 4,000	$ 5.00	yes	no	yes	yes	yes	yes	no	5-10, 11-17, 17-24
Fluidigm	Polaris	800	5,000	$ 10,000	$ 12.50	no	yes	no	yes	no	yes	yes	5-10, 11-17, 17-24
In-Drop	custom	10,000	100,000	$ 5,000	$ 0.50	yes	no	no	yes	no	yes	no	5-100
Raindance	RainDrop	unk	unk	unk	unk	yes	no	unk	unk	unk	unk	unk	unk
QIAGEN	CellRaft	44,000	unk	unk	unk	unk	no	unk	unk	unk	unk	unk	unk
WaferGen	iCell8	1,800	40,000	$ 2,750	$ 1.53	yes	limited	soon	yes	unk	unk	maybe	5-100

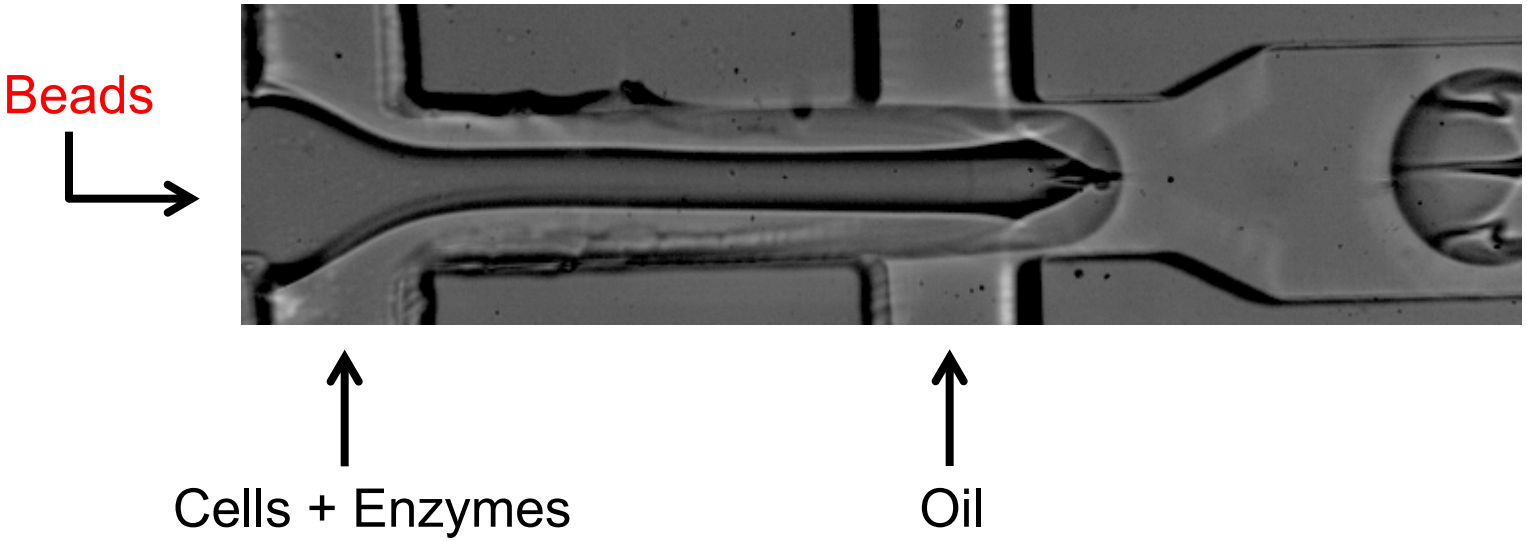
Single cell capture and RNA chemistry using nanodroplets

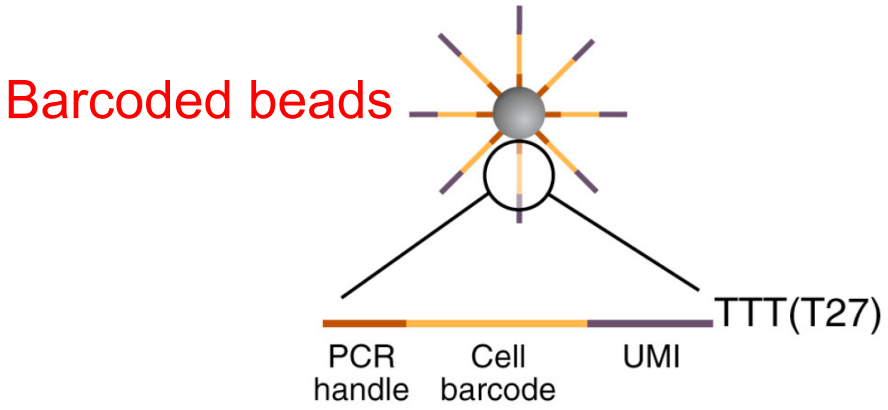
- Drop-seq



http://mccarrolllab.com/dropseq/
Macosko *et al.*, Cell 2015

Single cell capture and RNA chemistry using nanodroplets

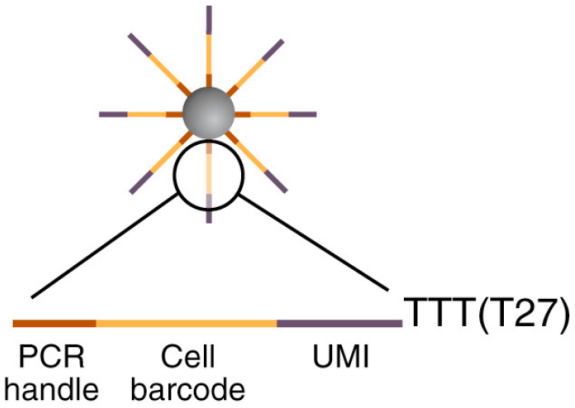


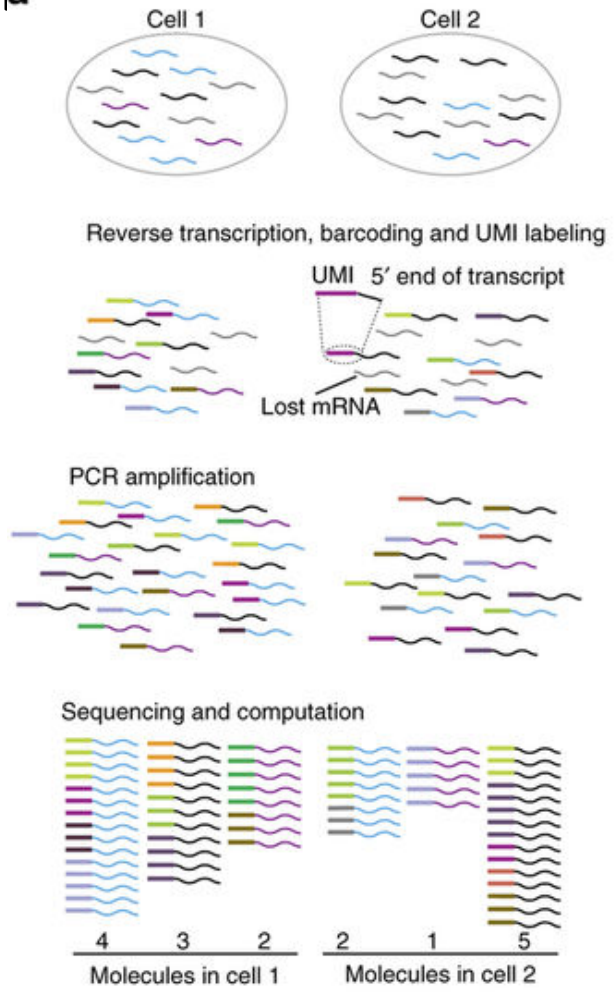


http://mccarrolllab.com/dropseq/
Macosko *et al.*, Cell 2015

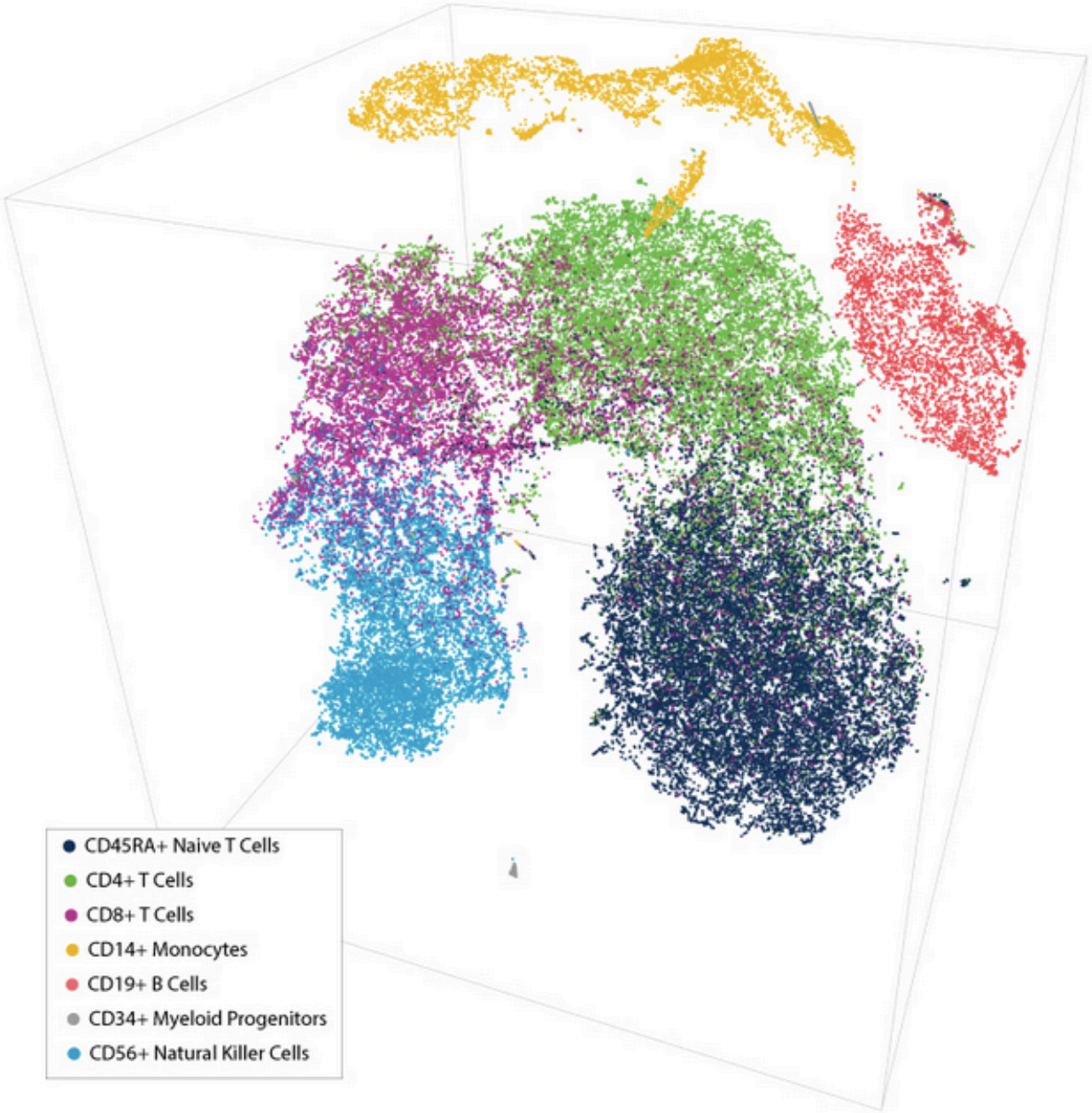
Unique Molecular Identifiers (UMIs)

Barcoded beads

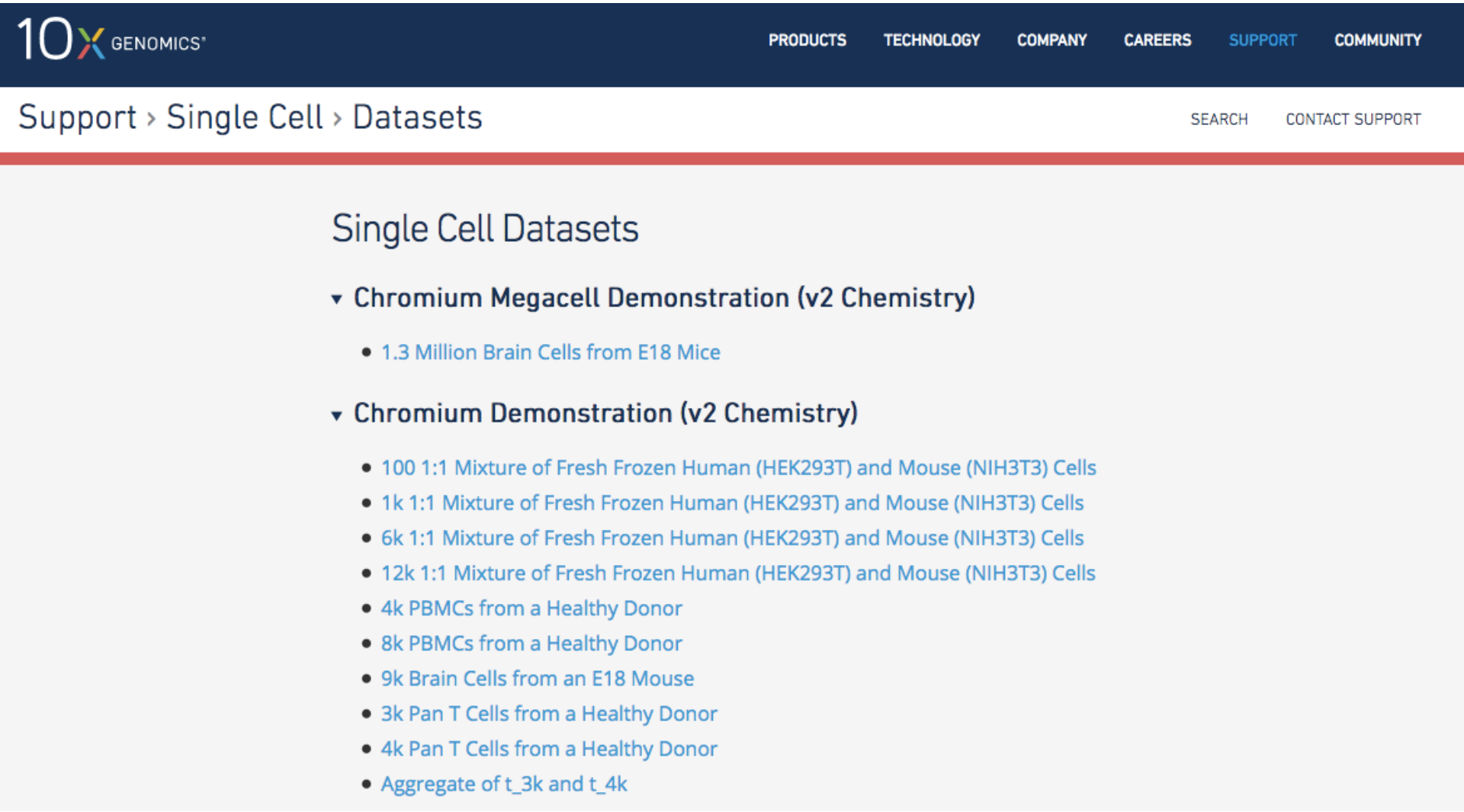




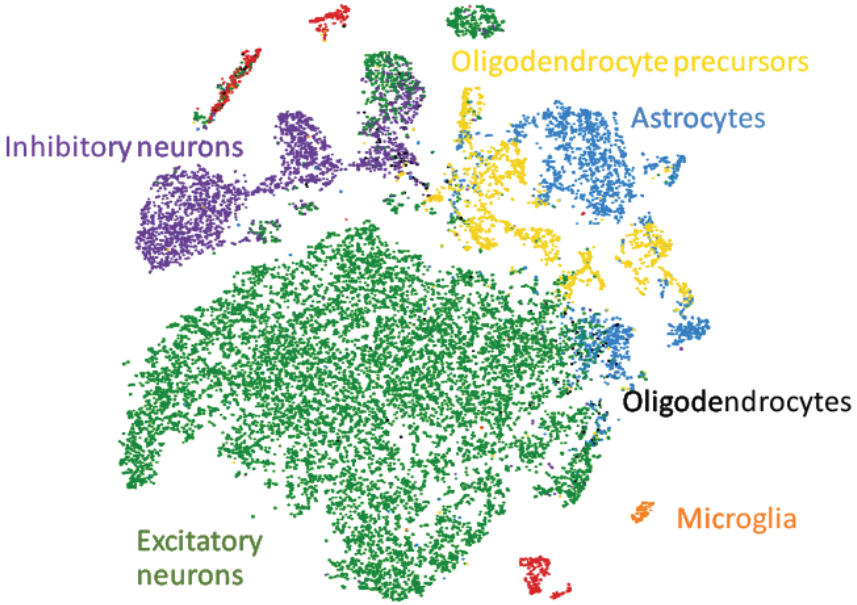
Islam *et al.*, Nature Methods 201

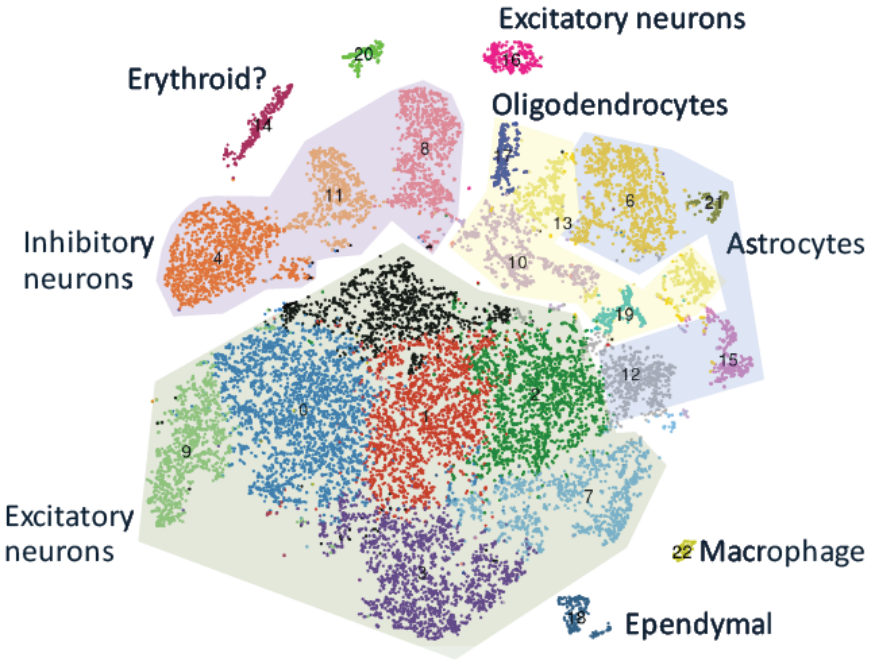

CD45RA+ Naive T Cells
CD4+ T Cells
CD8+ T Cells
CD14+ Monocytes
CD19+ B Cells
CD34+ Myeloid Progenitors
CD56+ Natural Killer Cells

1.3 million neurons catalogued



1.3 million mouse embryonic brain cells, 10X Chromium





MISSION

To create comprehensive reference maps of all human cells—the fundamental units of life—as a basis for both understanding human health and diagnosing, monitoring, and treating disease.

Beyond single cell RNA-seq

Single nuclei sequencing	scNuc-seq
Epigenomics	scBS-seq, scRRBS-seq, scCHIP-seq, scATAC-seq, scDNase-seq
Genomics	Whole genome, exome
Multiple simultaneous measurements	
RNA + DNA	DR-seq, G&T-seq
RNA + methylation	scM&T-seq, scMT-seq
RNA + DNA + methylation	scTrio-seq
RNA + protein	index sorting, CITE-seq
RNA + genome editing	Perturb-seq, CRISP-seq, CROP-seq

nature methods
Techniques for life scientists and chemists

ARTICLE PREVIEW

view full access options ▸

NATURE METHODS | BRIEF COMMUNICATION

G&T-seq: parallel sequencing of single-cell genomes and transcriptomes

Iain C Macaulay, Wilfried Haerty, Parveen Kumar, Yang I Li, Tim Xiaoming Hu, Mabel J Teng, Mubeen Goolam, Nathalie Saurat, Paul Coupland, Lesley M Shirley, Miriam Smith, Niels Van der Aa, Ruby Banerjee, Peter D Ellis, Michael A Quail, Harold P Swerdlow, Magdalena Zernicka-Goetz, Frederick J Livesey, Chris P Ponting & Thierry Voet

Affiliations | Contributions | Corresponding authors

Nature Methods **12**, 519–522 (2015) | doi:10.1038/nmeth.3370
Received 18 November 2014 | Accepted 27 March 2015 | Published online 27 April 2015

nature methods
Techniques for life scientists and chemists

NATURE METHODS | BRIEF COMMUNICATION

Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity

Christof Angermueller, Stephen J Clark, Heather J Lee, Iain C Macaulay, Mabel J Teng, Tim Xiaoming Hu, Felix Krueger, Sébastien A Smallwood, Chris P Ponting, Thierry Voet, Gavin Kelsey, Oliver Stegle & Wolf Reik

Affiliations | **Contributions** | **Corresponding authors**

Nature Methods **13**, 229–232 (2016) | doi:10.1038/nmeth.3728
Received 29 October 2015 | Accepted 09 December 2015 | Published online 11 January 2016

We report scM&T-seq, a method for parallel single-cell genome-wide methylome and transcriptome sequencing that allows for the discovery of associations between transcriptional and epigenetic variation. Profiling of 61 mouse embryonic stem cells confirmed known links between DNA methylation and transcription. Notably, the method revealed previously unrecognized associations between heterogeneously methylated distal regulatory elements and transcription of key pluripotency genes.

nature International weekly journal of science

ARTICLE PREVIEW

NATURE | LETTER

日本語要約

Single-cell chromatin accessibility reveals principles of regulatory variation

Jason D. Buenrostro, Beijing Wu, Ulrike M. Litzenburger, Dave Ruff, Michael L. Gonzales, Michael P. Snyder, Howard Y. Chang & William J. Greenleaf

Affiliations | Contributions | Corresponding authors

Nature **523**, 486–490 (23 July 2015) | doi:10.1038/nature14590

Received 12 January 2015 | Accepted 26 May 2015 | Published online 17 June 2015

nature International weekly journal of science

ARTICLE PREVIEW

NATURE | LETTER

日本語要約

The DNA methylation landscape of human early embryos

Hongshan Guo, Ping Zhu, Liying Yan, Rong Li, Boqiang Hu, Ying Lian, Jie Yan, Xiulian Ren, Shengli Lin, Junsheng Li, Xiaohu Jin, Xiaodan Shi, Ping Liu, Xiaoye Wang, Wei Wang, Yuan Wei, Xianlong Li, Fan Guo, Xinglong Wu, Xiaoying Fan, Jun Yong, Lu Wen, Sunney X. Xie, Fuchou Tang & Jie Qiao

Affiliations | Contributions | Corresponding authors

Nature **511**, 606–610 (31 July 2014) | doi:10.1038/nature13544

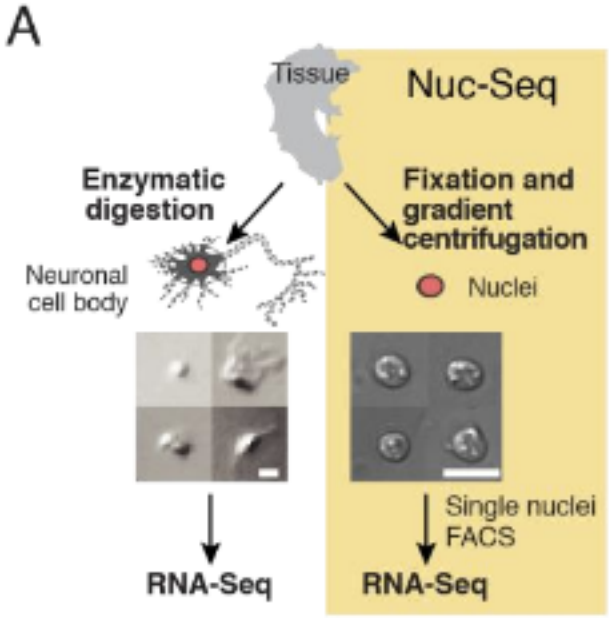
Received 10 November 2013 | Accepted 30 May 2014 | Published online 23 July 2014

CSH PRESS GENOME RESEARCH

Single-cell methylome landscapes of mouse embryonic stem cells and early embryos analyzed using reduced representation bisulfite sequencing

Hongshan Guo[1,3], Ping Zhu[1,2,3], Xinglong Wu[1], Xianlong Li[1], Lu Wen[1] and Fuchou Tang[1,4]

Other methods also emerging

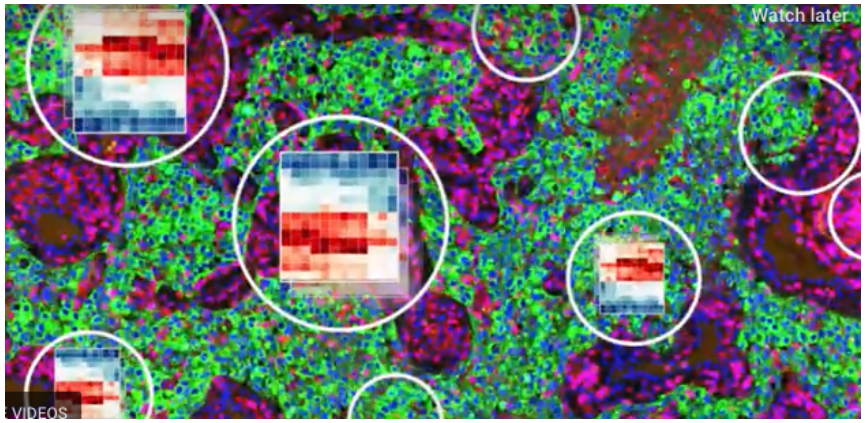


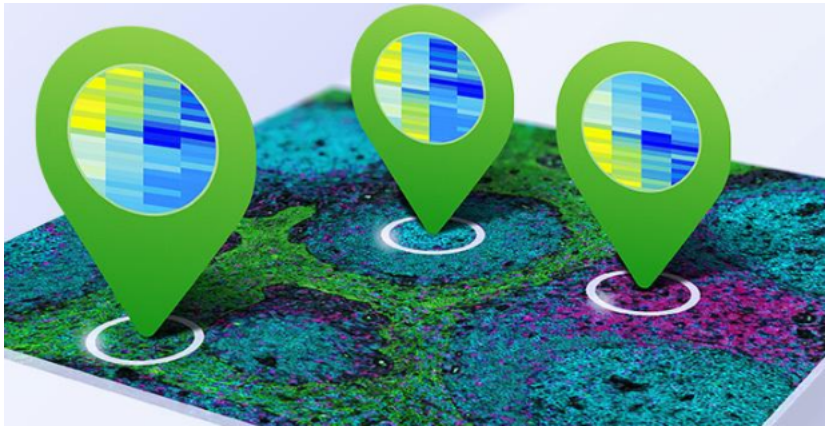
Div-Seq: A single nucleus RNA-Seq method reveals dynamics of rare adult newborn neurons in the CNS

http://biorxiv.org/content/biorxiv/early/2016/03/27/045989.full.pdf

THE TOPIC: SPATIAL GENOMICS

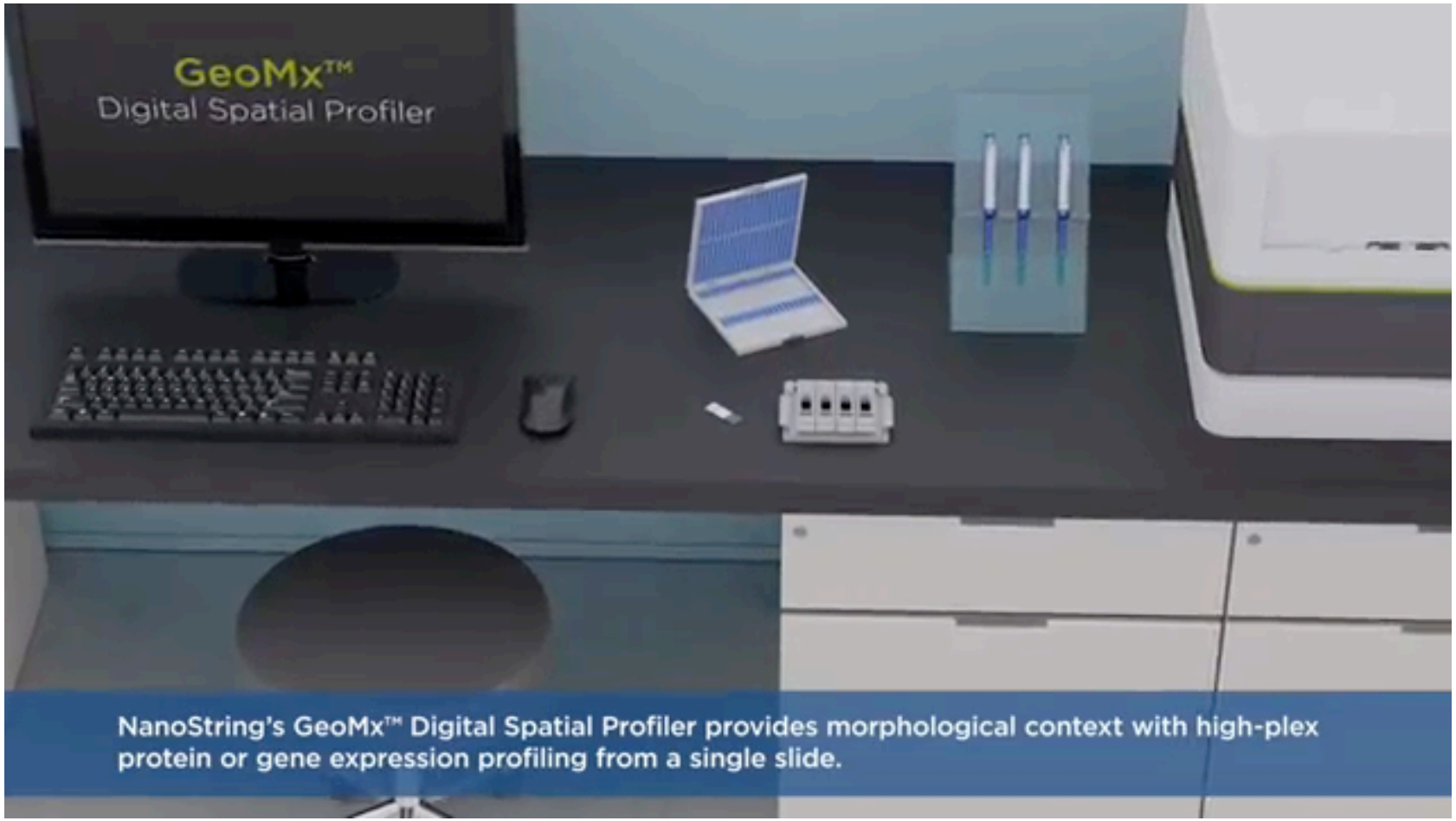
KEEPING SPATIAL INFORMATION – SPATIAL GENOMICS



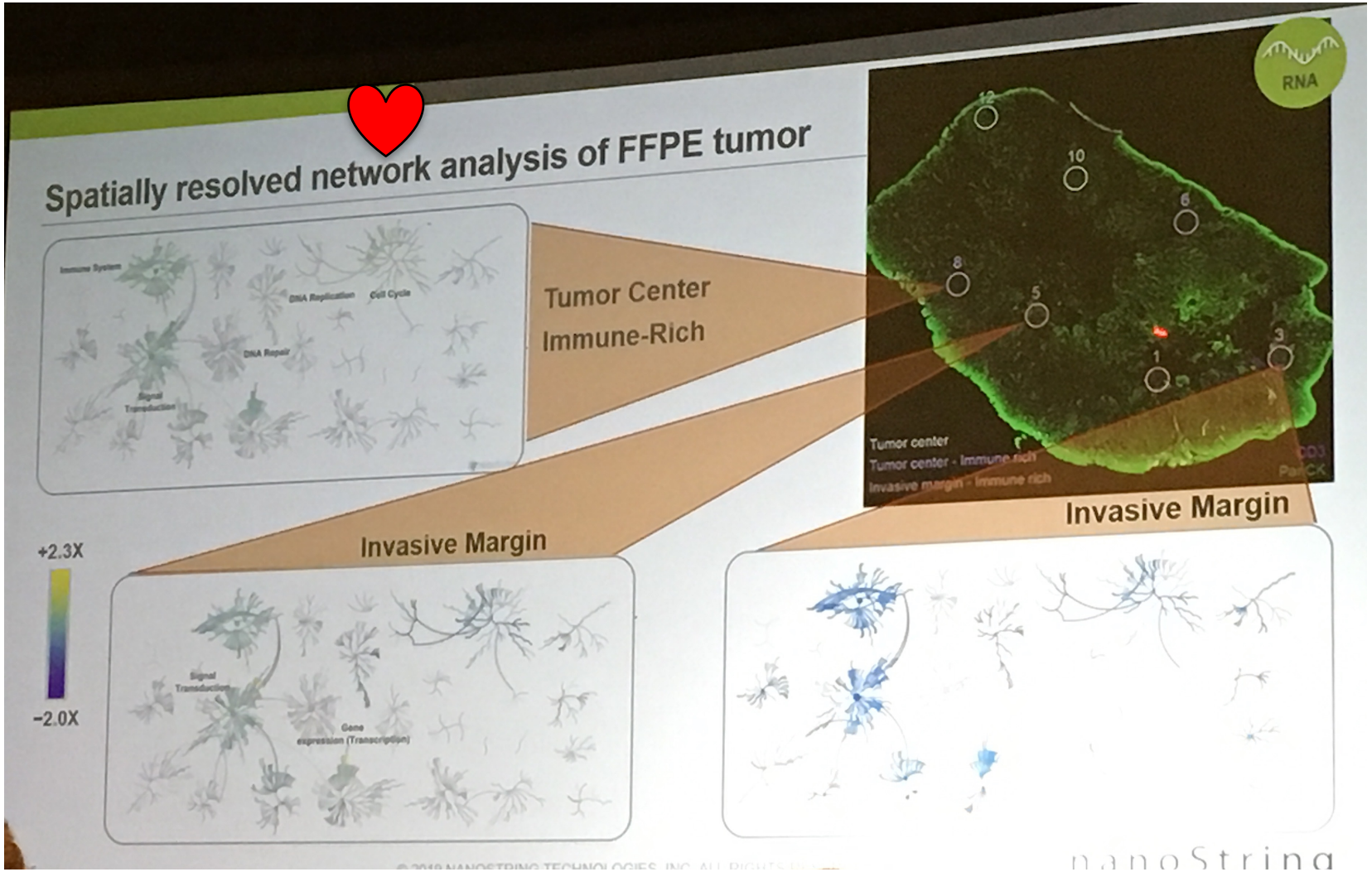


Weill Cornell Medicine

Spatial-genomics



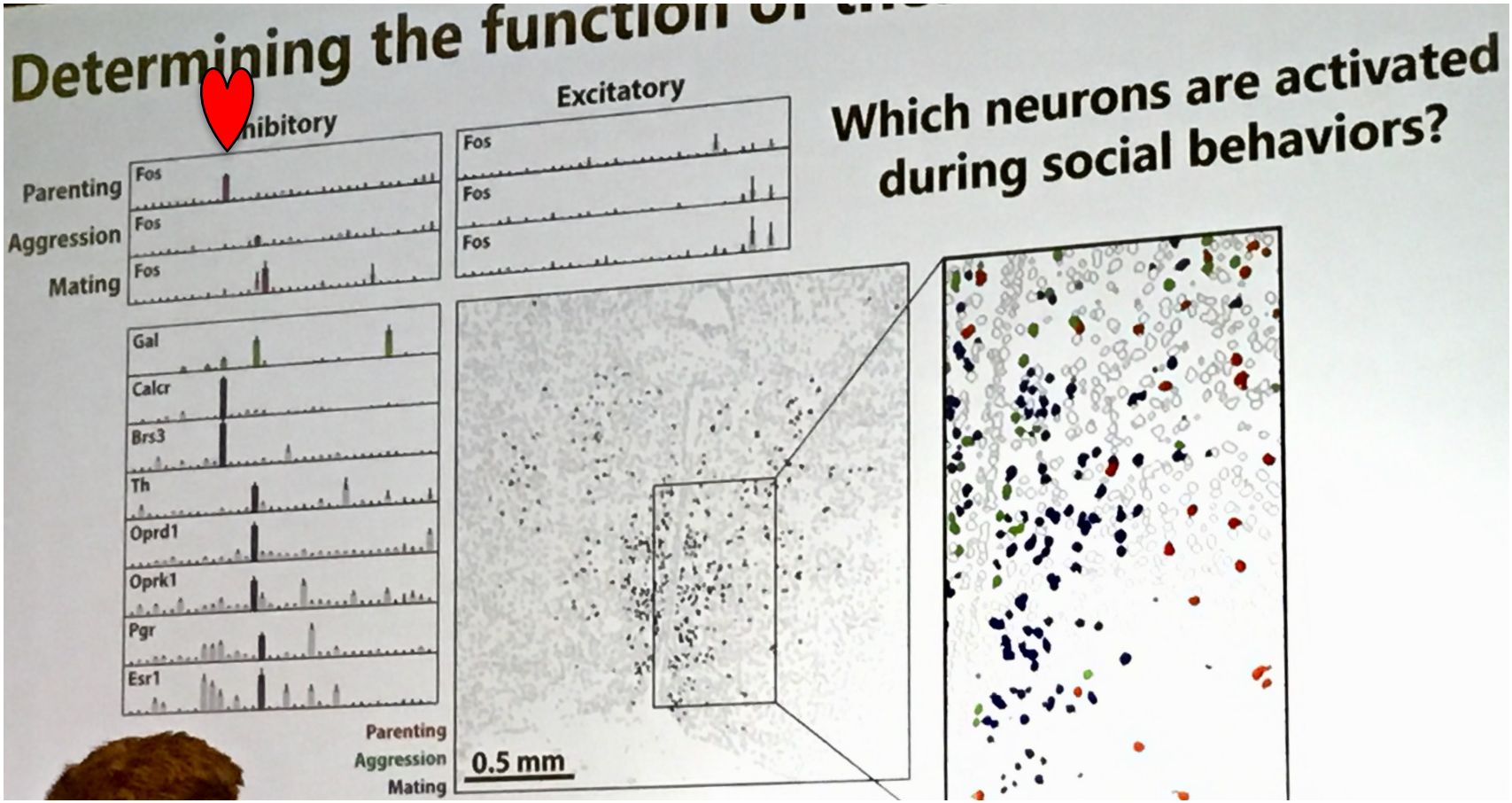
Spatial genomics – Network analysis



Weill Cornell Medicine

SPATIAL-GENOMICS - MERFISH

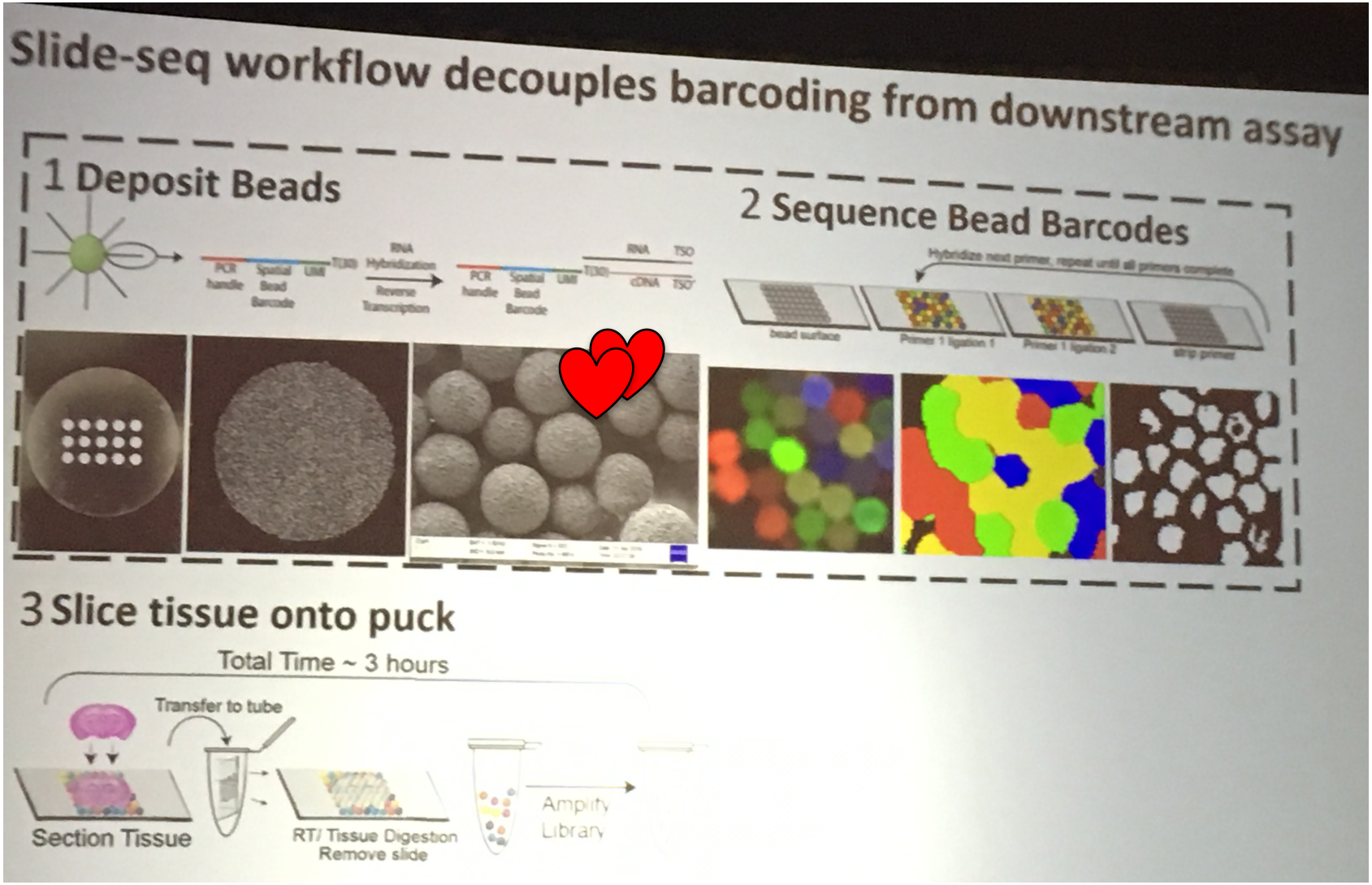
RNA Imaging with Multiplexed Error-Robust Fluorescence In Situ Hybridization (MERF



Jeffrey Moffitt

Weill Cornell Medicine

SPATIAL GENOMICS - Slide-Seq



FEI CHEN

Deep Gratitude to Many People:

@mason_lab

Mason Lab
Ebrahim Afshinnekoo
Noah Alexander
Pradeep Ambrose
Daniela Bezdan
Daniel Butler
Eugene Carragee
Chou Chou
David Danko
Tim Donahoe
Jonathan Foox
Matthew MacKay
Alexa McIntyre
Cem Meyden
Jake Reed
Heba Shabaan
Delia Tomoiaga
David Westfall
Craig Westover
WorldQuant Fellows

HudsonAlpha
Shawn Levy

Cornell/WCM
Selina Chen-Kiang
Iwijn De Vlaminck
Olivier Elemento
Samie Jaffrey
Iman Hajirasouliha
Ari Melnick
Margaret Ross
Epigenomics Core

Biotia
Niamh O'Hara (Biotia)
Rachid Ounit (Biotia)

NYU
Martin Blaser
Jane Carlton
Chris Park
Elizabeth Hénaff

Rockefeller
Jeanne Garbarino
Charles Rice

UVA
Francine Garrett-Bakelman

NASA/JPL/Ames
Kate Rubins
David Smith
Kasthuri Venkateswaran

UCSF
Charles Chiu

Miami
Maria E Figueroa

MIT Media Lab
Kevin Slavin
Devora Najjar

FDA/SEQC/Fudan.
Leming Shi

ABRF/Vermont
George Grills
Scott Tighe
Don Baldwin

Univ. Chicago
Yoav Gilad

Icahn/MSSM
Eric Schadt,
Joel Dudley,
Bobby Sebra

Illumina
Gary Schroth
Marc Van Oene

Duke
Stacy Horner
Nandan Gokhale

Jackson Labs
Sheng Li

Baylor
Jeff Rogers

MSKCC
Alex Kentsis
Christina Leslie
Ross Levine

AMNH
George Amato
Mark Sidall