

# **Intro to Bioinformatics**

**Lecture 2  
Aug 31, 2016**

# ANNOUNCEMENTS

- Reading assignments
- Sign up for AWS

# **GENOMICS**

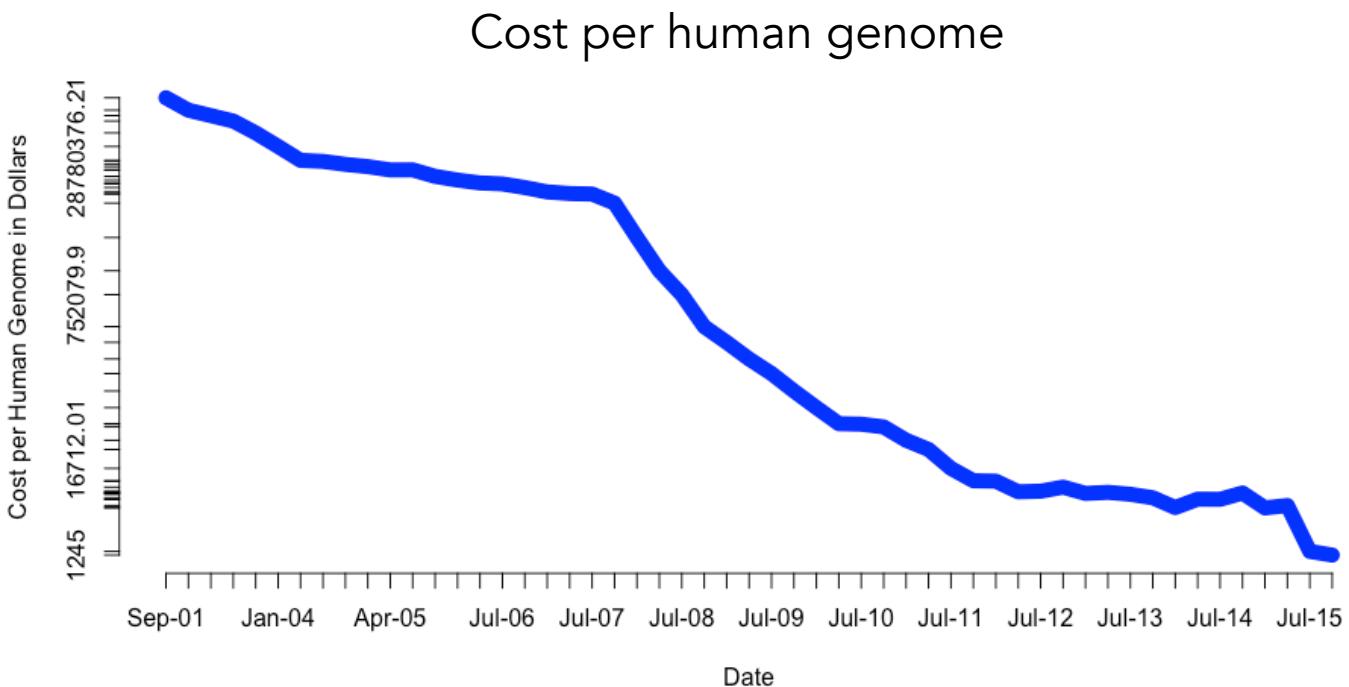
- What is genomics?

# **GENOMICS**

- What is bioinformatics?

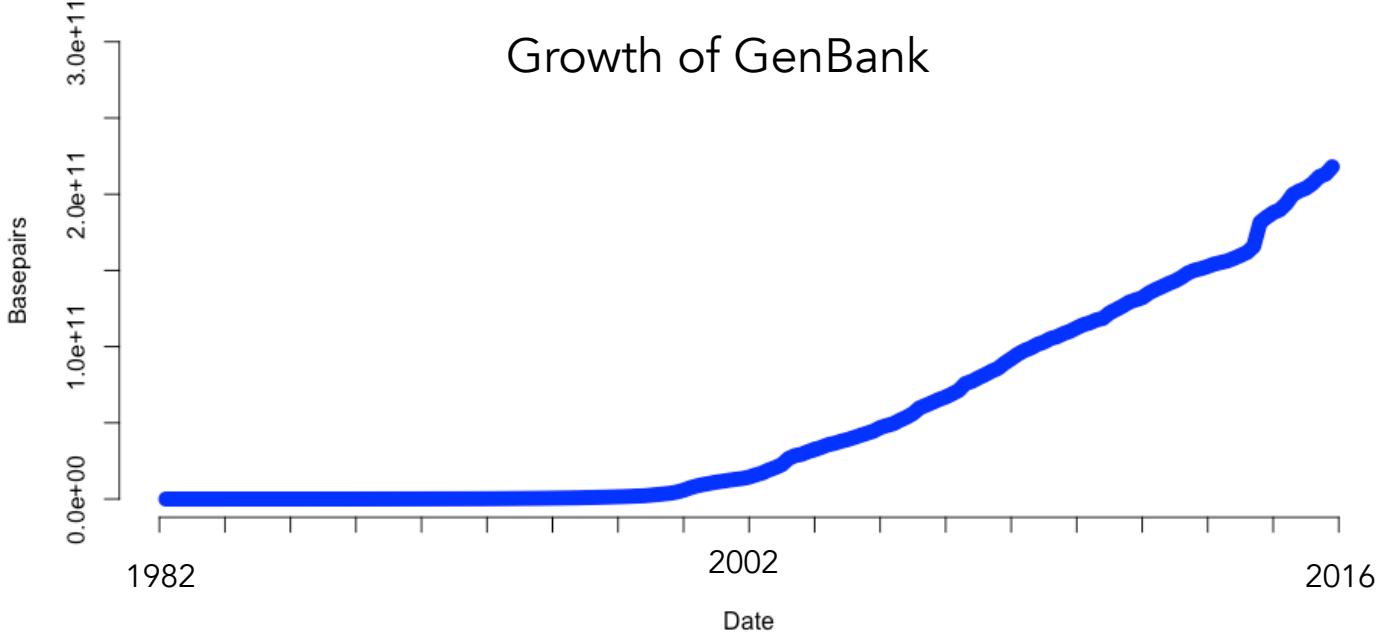
# BIOINFORMATICS

- Why now??



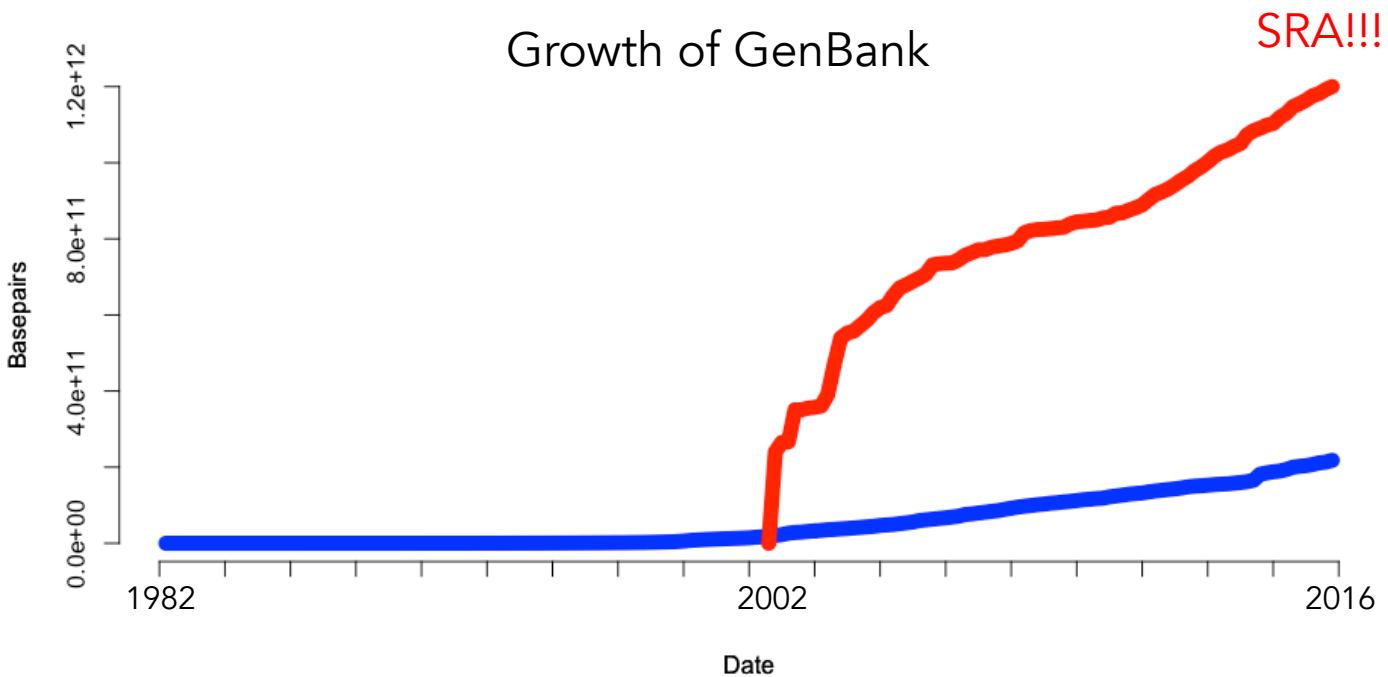
# BIOINFORMATICS

- Why now??



# BIOINFORMATICS

- Why now??



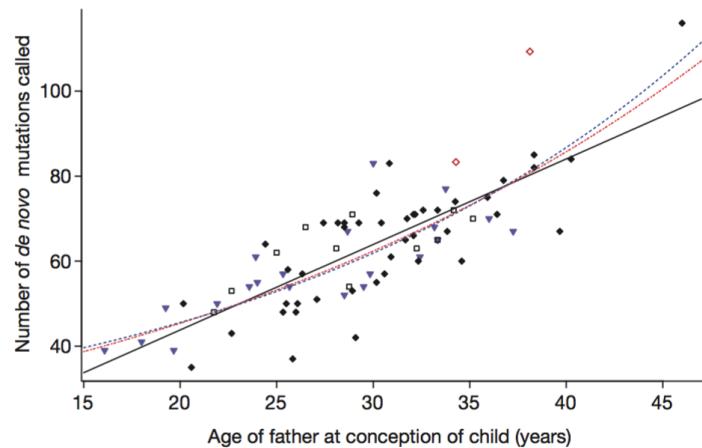
# BIOINFORMATICS

- Why now??



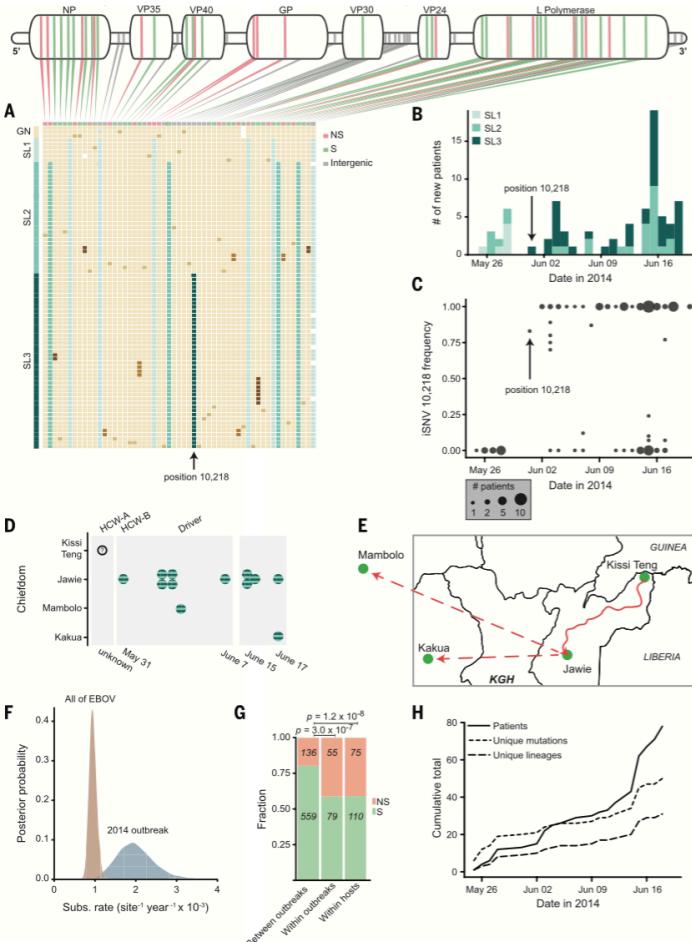
# Rate of *de novo* mutations and the importance of father's age to disease risk

Augustine Kong<sup>1</sup>, Michael L. Frigge<sup>1</sup>, Gisli Masson<sup>1</sup>, Soren Besenbacher<sup>1,2</sup>, Patrick Sulem<sup>1</sup>, Gisli Magnusson<sup>1</sup>, Sigurjon A. Gudjonsson<sup>1</sup>, Asgeir Sigurdsson<sup>1</sup>, Aslaug Jonasdottir<sup>1</sup>, Adalbjorg Jonasdottir<sup>1</sup>, Wendy S. W. Wong<sup>3</sup>, Gunnar Sigurdsson<sup>1</sup>, G. Bragi Walters<sup>1</sup>, Stacy Steinberg<sup>1</sup>, Hannes Helgason<sup>1</sup>, Gudmar Thorleifsson<sup>1</sup>, Daniel F. Gudbjartsson<sup>1</sup>, Agnar Helgason<sup>1,4</sup>, Olafur Th. Magnusson<sup>1</sup>, Unnur Thorsteinsdottir<sup>1,5</sup> & Kari Stefansson<sup>1,5</sup>



# Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak

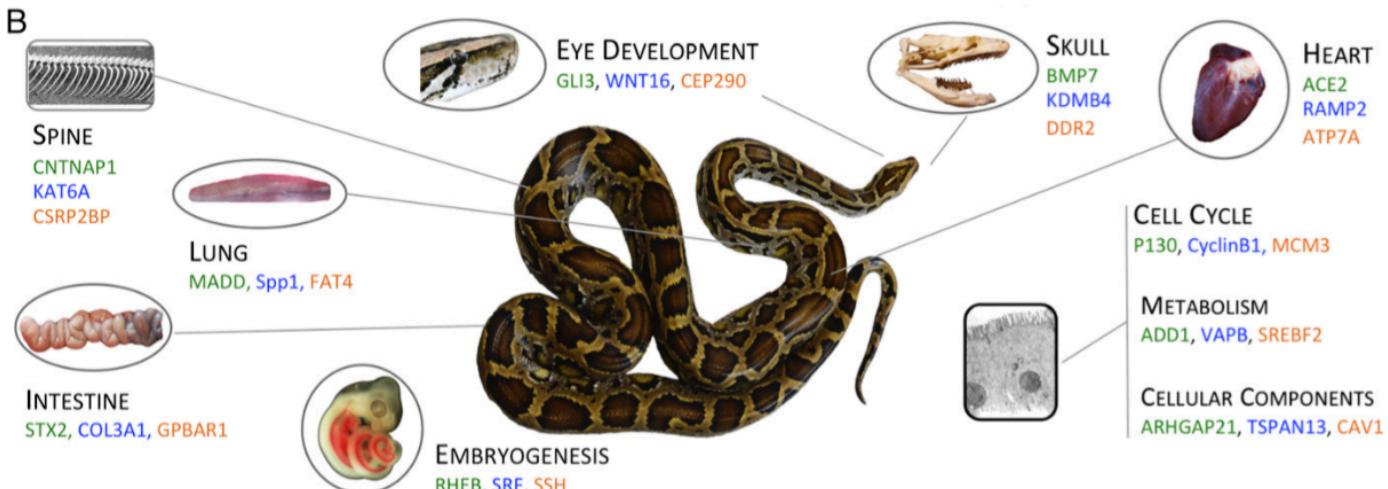
Stephen K. Gire,<sup>1,2\*</sup> Augustine Goba,<sup>3\*</sup>† Kristian G. Andersen,<sup>1,2,4†</sup> Rachel S. G. Sealton,<sup>2,4§</sup> Daniel J. Park,<sup>2§</sup> Lansana Kamneh,<sup>3</sup> Simbirie Jalloh,<sup>3</sup> Mambu Momoh,<sup>3,5</sup> Mohamed Fullah,<sup>3,5†</sup> Gyts Dudas,<sup>6</sup> Shirlee Wohr,<sup>1,2,7</sup> Lina M. Moses,<sup>8</sup> Nathan L. Yozwiak,<sup>1,2</sup> Sarah Winnicki,<sup>1,2</sup> Christian B. Matranga,<sup>2</sup> Christine M. Malboeuf,<sup>2</sup> James Qu,<sup>2</sup> Adrienne D. Gladden,<sup>2</sup> Stephen F. Schaffner,<sup>1,2</sup> Xiao Yang,<sup>2</sup> Pan-Pan Jiang,<sup>1,2</sup> Mahan Nekouli,<sup>1,2</sup> Andres Colubri,<sup>1</sup> Molanya Ruth Coomber,<sup>7</sup> Mbula Fomnie,<sup>9,‡</sup> Alex Moigboi,<sup>7,‡</sup> Michael Gbakpe,<sup>3</sup> Fatima K. Kamara,<sup>3</sup> Veronica Tucker,<sup>3</sup> Edwin Konuwa,<sup>3</sup> Sidiki Saffa,<sup>3,‡</sup> Josephine Sellu,<sup>3</sup> Abdul Azziz Jalloh,<sup>3</sup> Alice Kovoma,<sup>3,‡</sup> James Koninga,<sup>3</sup> Ibrahim Mustapha,<sup>3</sup> Kandeh Kargbo,<sup>3</sup> Momoh Foday,<sup>3</sup> Mohamed Yillah,<sup>3</sup> Franklyn Kamneh,<sup>3</sup> Willie Robert,<sup>3</sup> James L. B. Massally,<sup>2</sup> Sinéad B. Chapman,<sup>7</sup> James Bochicchio,<sup>2</sup> Cheryl Murphy,<sup>2</sup> Chad Nusbaum,<sup>2</sup> Sarah Young,<sup>2</sup> Bruce W. Birren,<sup>2</sup> Donald S. Grant,<sup>3</sup> John S. Scheiffelin,<sup>8</sup> Eric S. Lander,<sup>2,7,9</sup> Christian Happi,<sup>10</sup> Sahr M. Gevao,<sup>11</sup> Andreas Gnirke,<sup>2,§</sup> Andrew Rambaut,<sup>6,12,13§</sup> Robert F. Garry,<sup>8§</sup> S. Humarr Khan,<sup>3,‡§</sup> Pardis C. Sabeti<sup>1,2,†§</sup>



# The Burmese python genome reveals the molecular basis for extreme adaptation in snakes

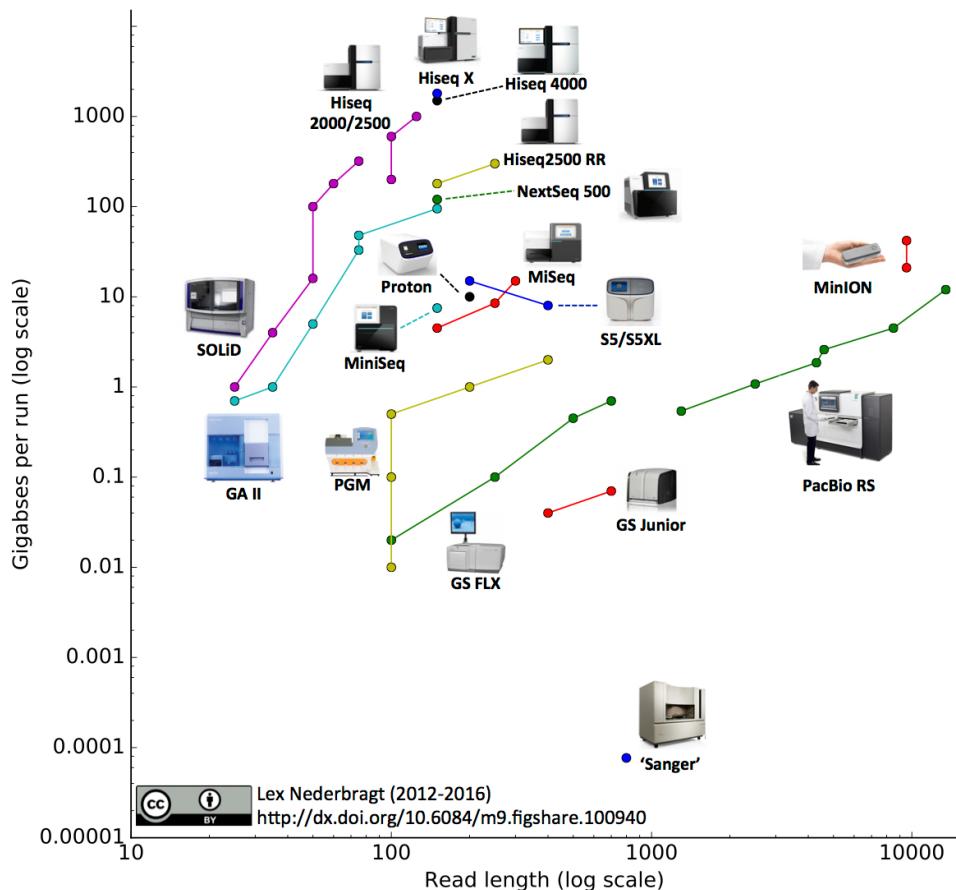
Todd A. Castoe<sup>a,b</sup>, A. P. Jason de Koning<sup>a,c</sup>, Kathryn T. Hall<sup>a</sup>, Daren C. Card<sup>b</sup>, Drew R. Schield<sup>b</sup>, Matthew K. Fujita<sup>b</sup>, Robert P. Ruggiero<sup>a</sup>, Jack F. Degner<sup>d</sup>, Juan M. Daza<sup>e</sup>, Wanjun Gu<sup>f</sup>, Jacobo Reyes-Velasco<sup>b</sup>, Kyle J. Shaney<sup>b</sup>, Jill M. Castoe<sup>a,b</sup>, Samuel E. Fox<sup>g</sup>, Alex W. Poole<sup>a</sup>, Daniel Polanco<sup>a</sup>, Jason Dobry<sup>h</sup>, Michael W. Vandewege<sup>i</sup>, Qing Li<sup>j</sup>, Ryan K. Schott<sup>k</sup>, Aurélie Kapusta<sup>l</sup>, Patrick Minx<sup>l</sup>, Cédric Feschotte<sup>l</sup>, Peter Uetz<sup>m</sup>, David A. Ray<sup>i,n</sup>, Federico G. Hoffmann<sup>i,n</sup>, Robert Bogden<sup>h</sup>, Eric N. Smith<sup>b</sup>, Belinda S. W. Chang<sup>k</sup>, Freek J. Vonk<sup>o,p,q</sup>, Nicholas R. Casewell<sup>q,r</sup>, Christiaan V. Henkel<sup>p,s</sup>, Michael K. Richardson<sup>p</sup>, Stephen P. Mackessy<sup>t</sup>, Anne M. Bronikowski<sup>u</sup>, Mark Yandell<sup>j</sup>, Wesley C. Warren<sup>l</sup>, Stephen M. Secor<sup>v</sup>, and David D. Pollock<sup>a,1</sup>

B



- Discover other stuff, too
  - Genes
  - Expression levels
  - Repetitive regions
  - Community composition
  - Methylation
  - ...

# SEQUENCING PLATFORMS



# SEQUENCING PLATFORMS

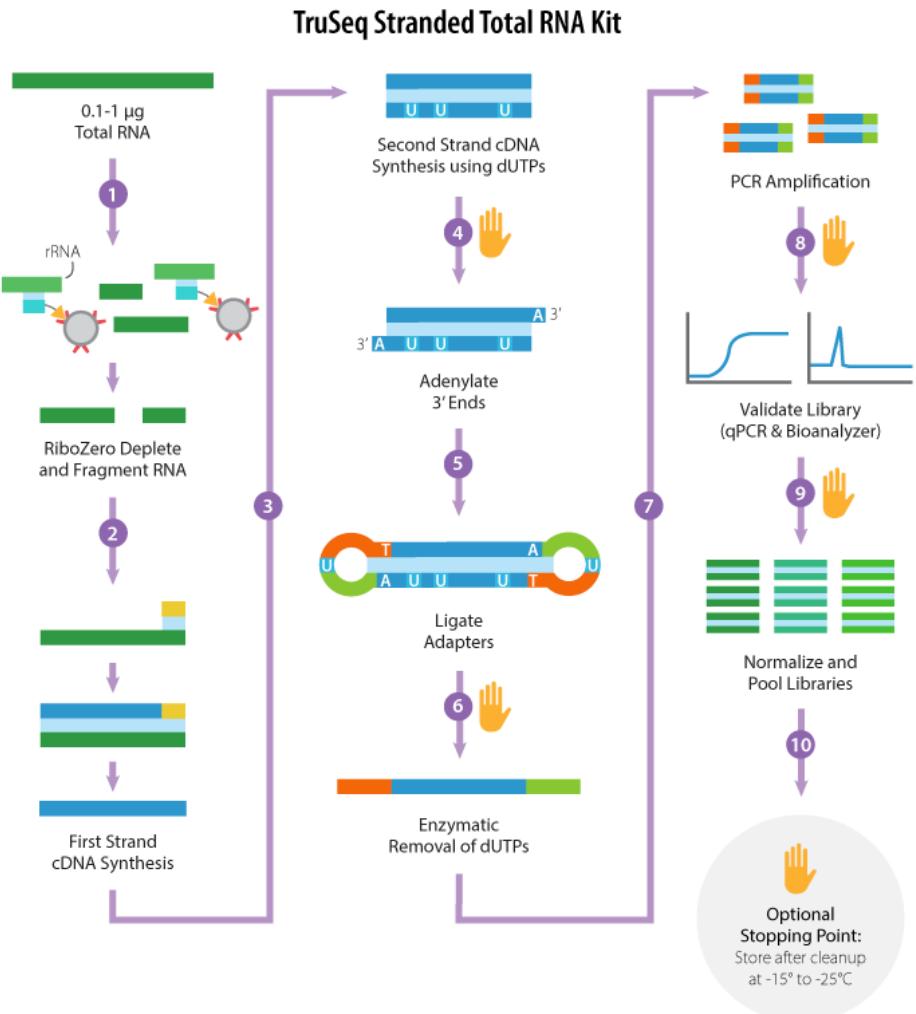
- Illumina (@illumina)

	 MiniSeq System	 MiSeq Series	 NextSeq Series	 HiSeq Series	 HiSeq X Series*
<b>Key Methods</b>	Amplicon, targeted RNA, small RNA, and targeted gene panel sequencing.	Small genome, amplicon, and targeted gene panel sequencing.	Everyday exome, transcriptome, and targeted resequencing.	Production-scale genome, exome, transcriptome sequencing, and more.	Population- and production-scale whole-genome sequencing.
<b>Maximum Output</b>	7.5 Gb	15 Gb	120 Gb	1500 Gb	1800 Gb
<b>Maximum Reads per Run</b>	25 million	25 million <sup>†</sup>	400 million	5 billion	6 billion
<b>Maximum Read Length</b>	2 × 150 bp	2 × 300 bp	2 × 150 bp	2 × 150 bp	2 × 150 bp
<b>Run Time</b>	4–24 hours	4–55 hours	12–30 hours	<1–3.5 days (HiSeq 3000/HiSeq 4000) 7 hours–6 days (HiSeq 2500)	<3 days

# SEQUENCING PLATFORMS

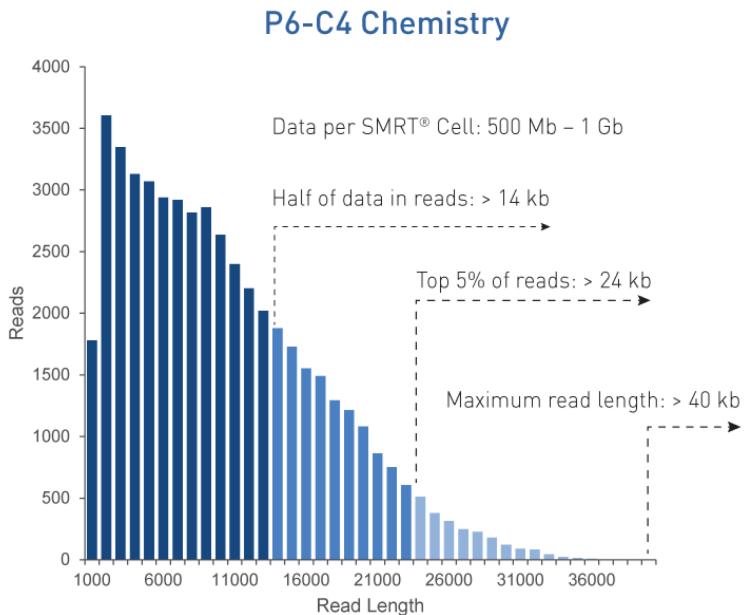
- Illumina (@illumina) <https://youtu.be/pfZp5Vgsbw0>

# How To MAKE AN ILLUMINA mRNA LIBRARY



# SEQUENCING PLATFORMS

- Pacific Biosystems (@PacBio )



Based on data from a 20 kb size-selected *E. coli* library using a 4-hour movie.

Each SMRT Cell yields ~ 50,000 reads.



# SEQUENCING PLATFORMS



- Pacific Biosystems (@PacBio )

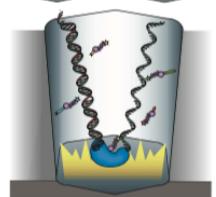
**SMRTbell template**  
Two hairpin adapters allow continuous circular sequencing



**ZMW wells**  
Sites where sequencing takes place



**Labelled nucleotides**  
All four dNTPs are labelled and available for incorporation



**Modified polymerase**  
As a nucleotide is incorporated by the polymerase, a camera records the emitted light



**PacBio output**  
A camera records the changing colours from all ZMWs; each colour change corresponds to one base

# SEQUENCING PLATFORMS

- Oxford Nanopore (@nanopore)



<https://www.youtube.com/watch?v=CE4dW64x3Ts>  
<https://youtu.be/btm9lrX1I-Q>