



HiSeq2000 / 2500



Ion torrent



Ion Proton



GS-FLX

MiSeq

# Sequencing Technologies

# Unraveling DNA: The evolution of Sanger sequencing



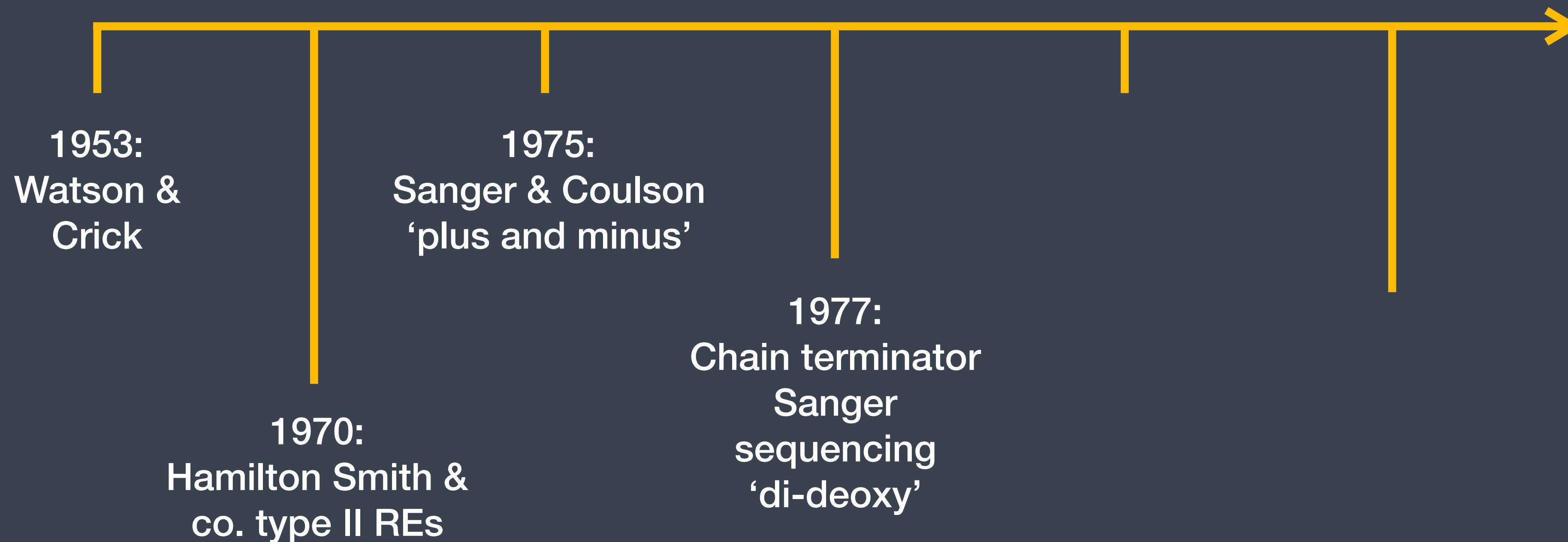
# Unraveling DNA: The evolution of Sanger sequencing



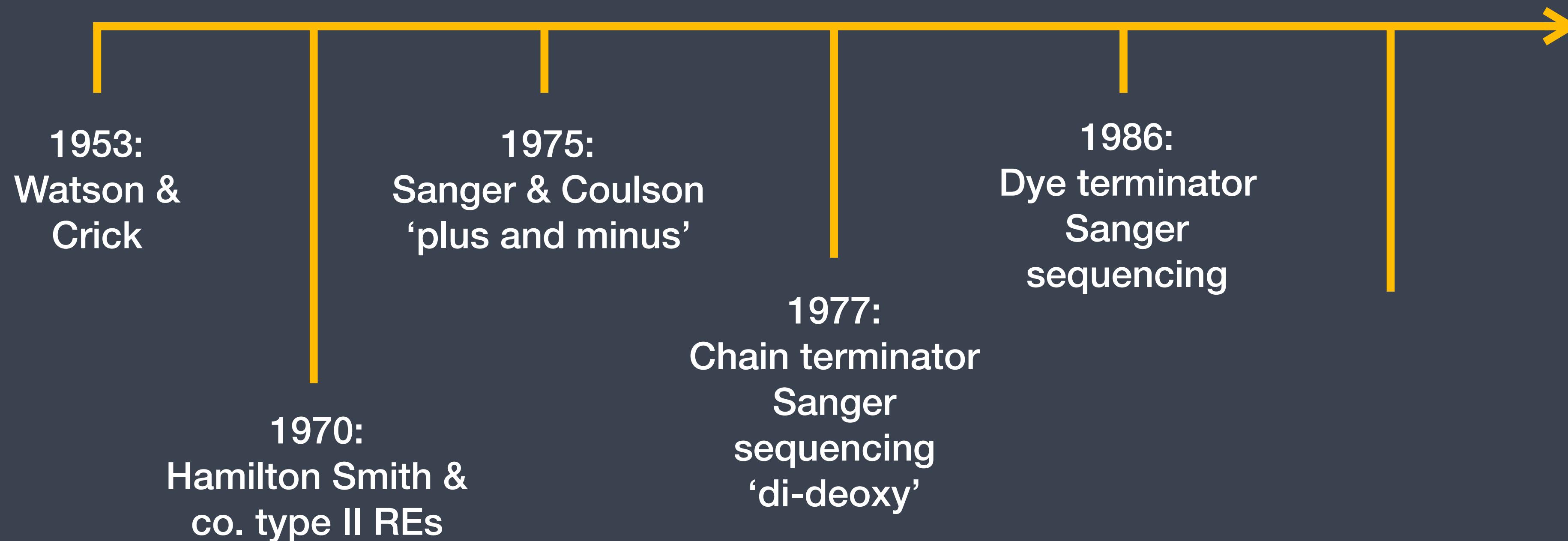
# Unraveling DNA: The evolution of Sanger sequencing



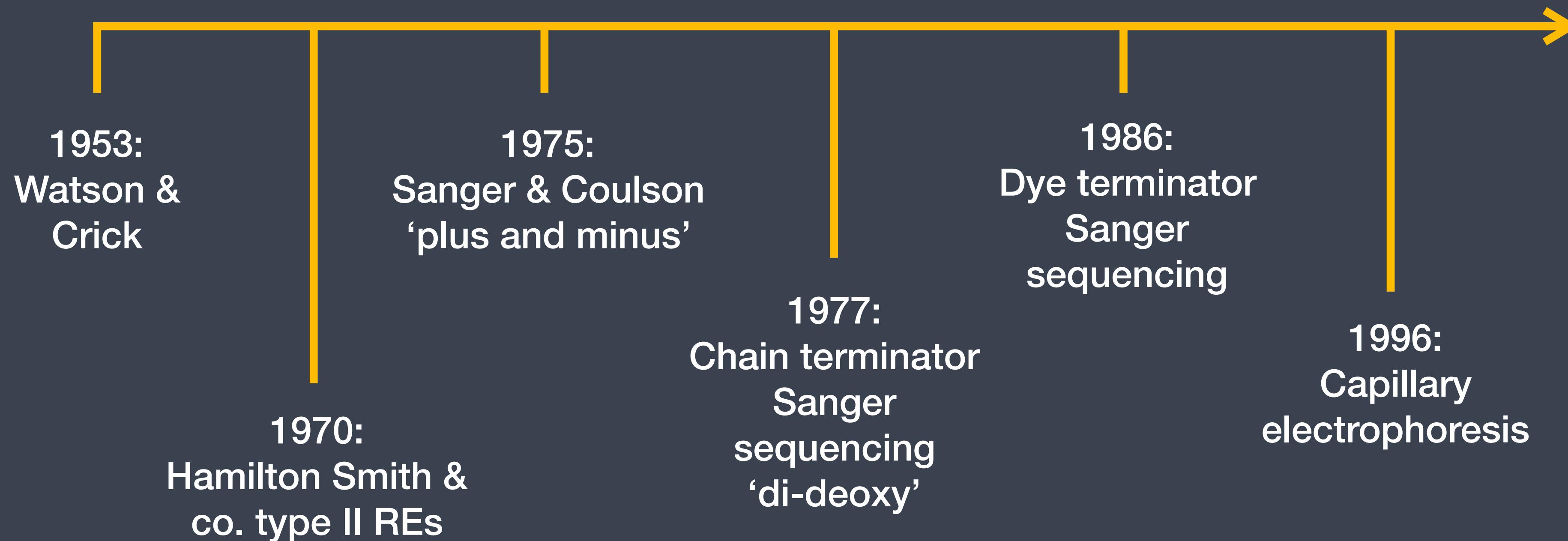
# Unraveling DNA: The evolution of Sanger sequencing

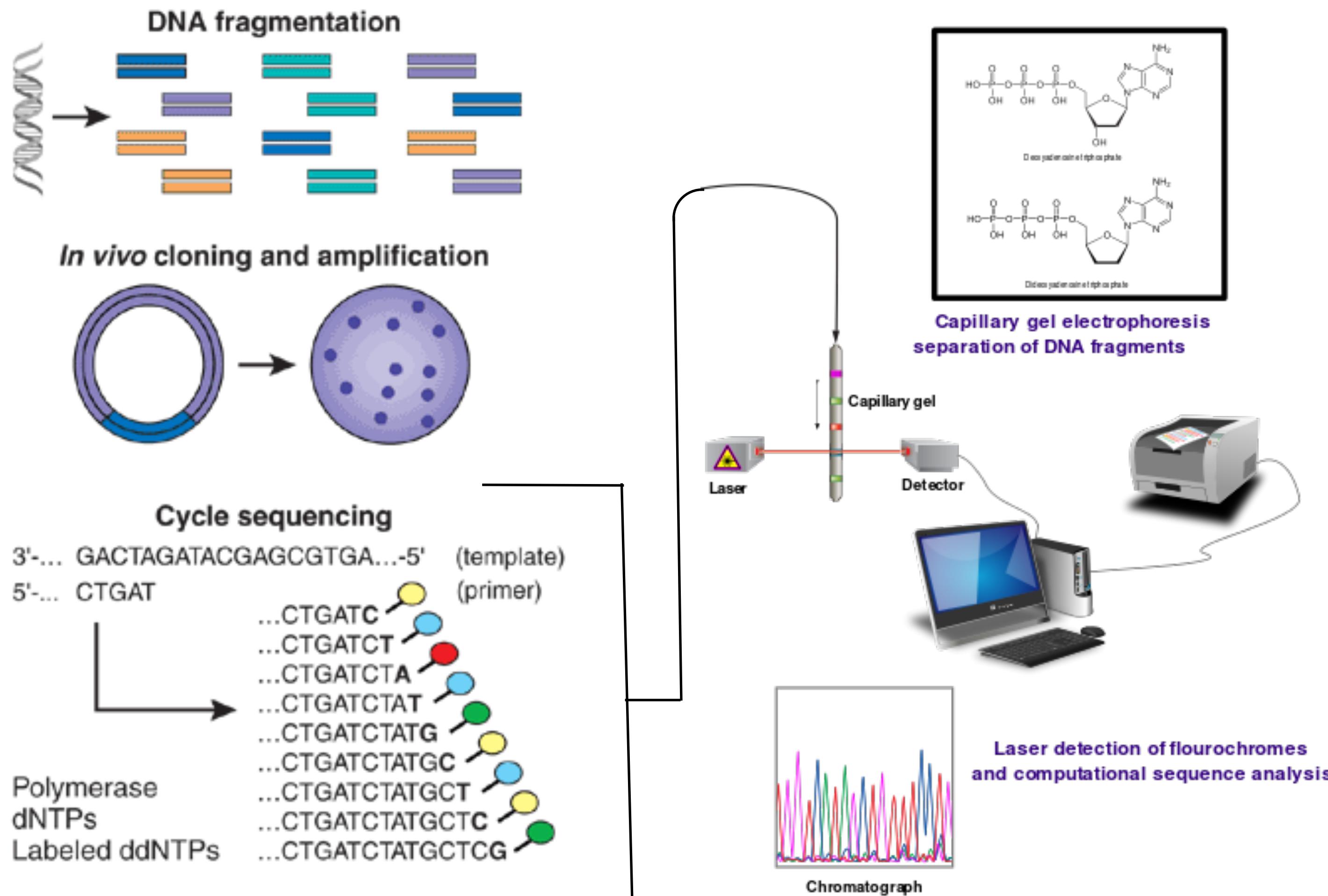


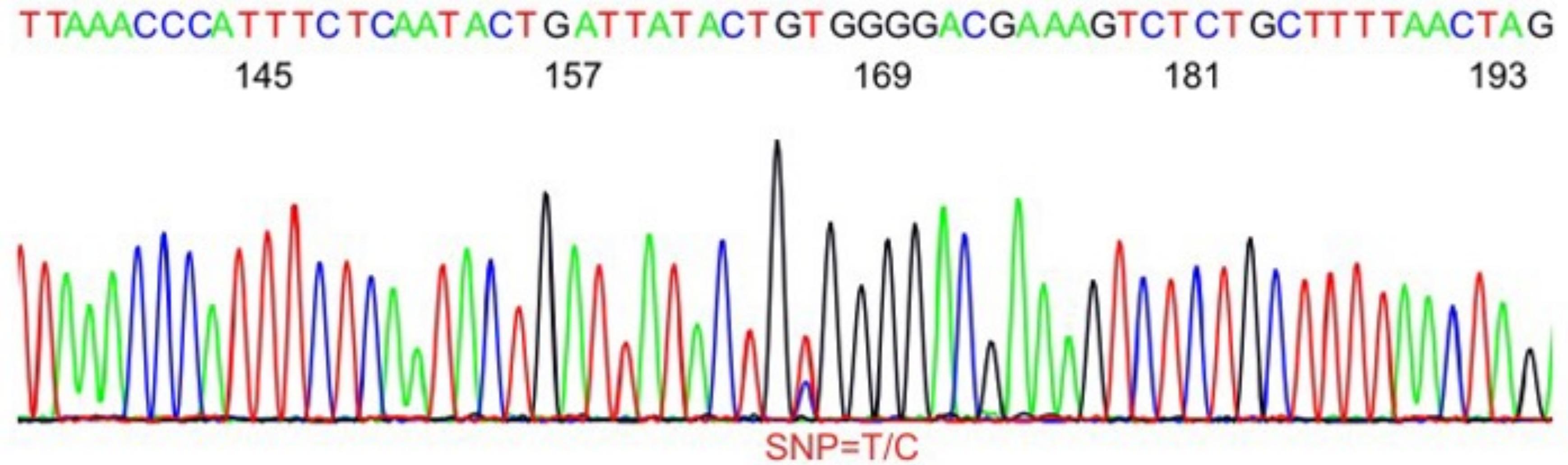
# Unraveling DNA: The evolution of Sanger sequencing



# Unraveling DNA: The evolution of Sanger sequencing

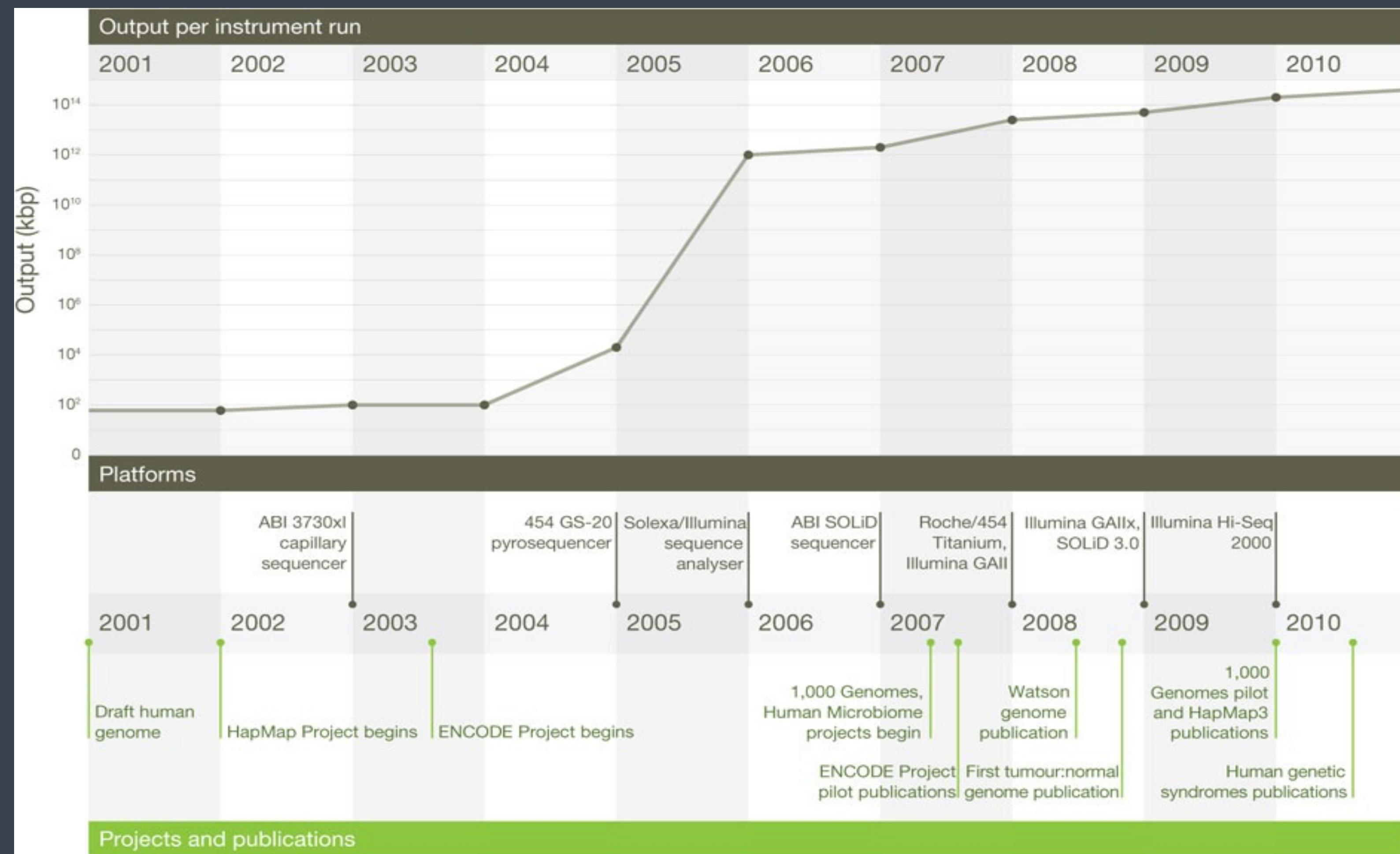


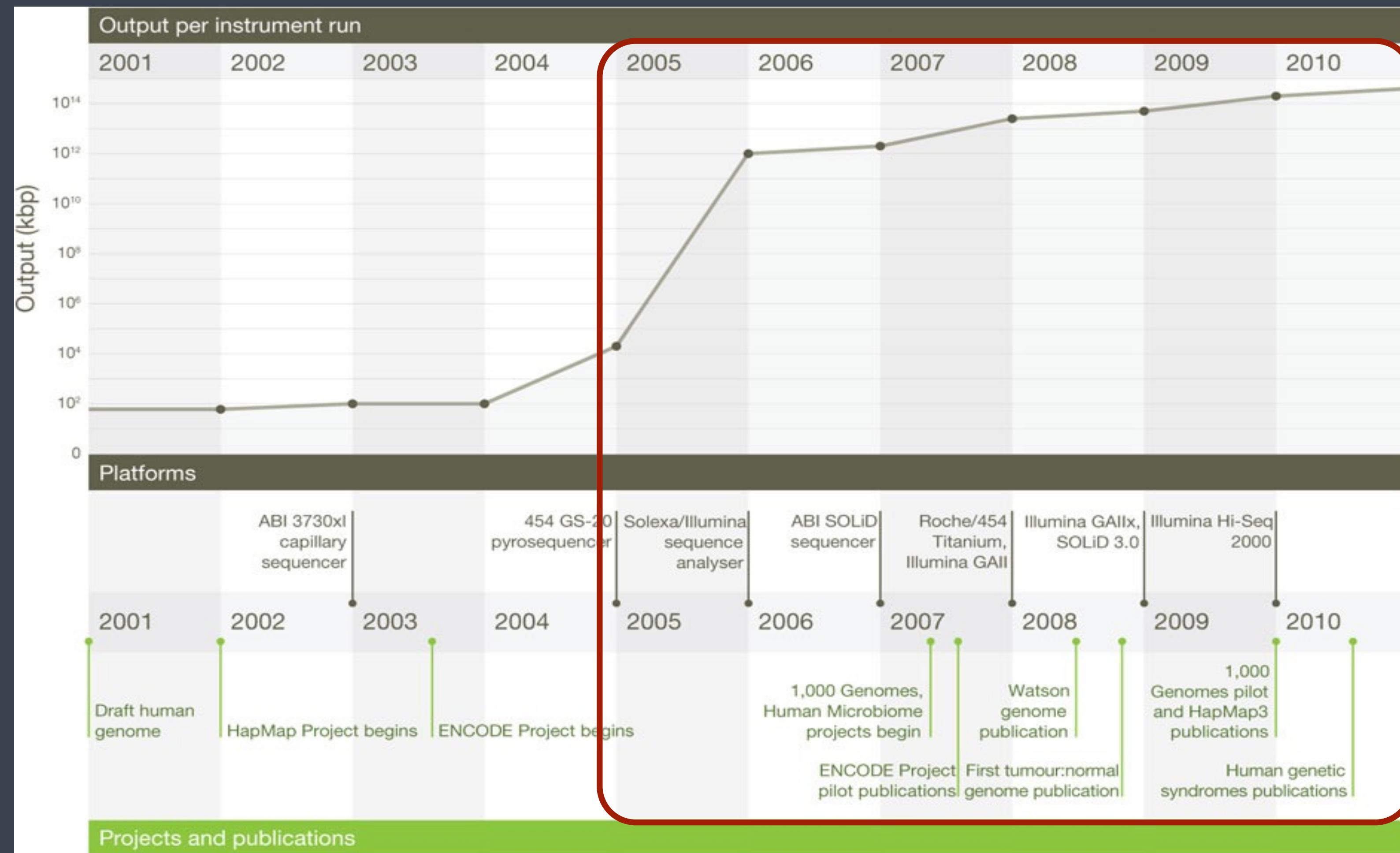




## Interpreting sequencing data

<http://www.xcelrisgenomics.com/SS-SNPGenotypingbySequencingonABI3730xl.html>

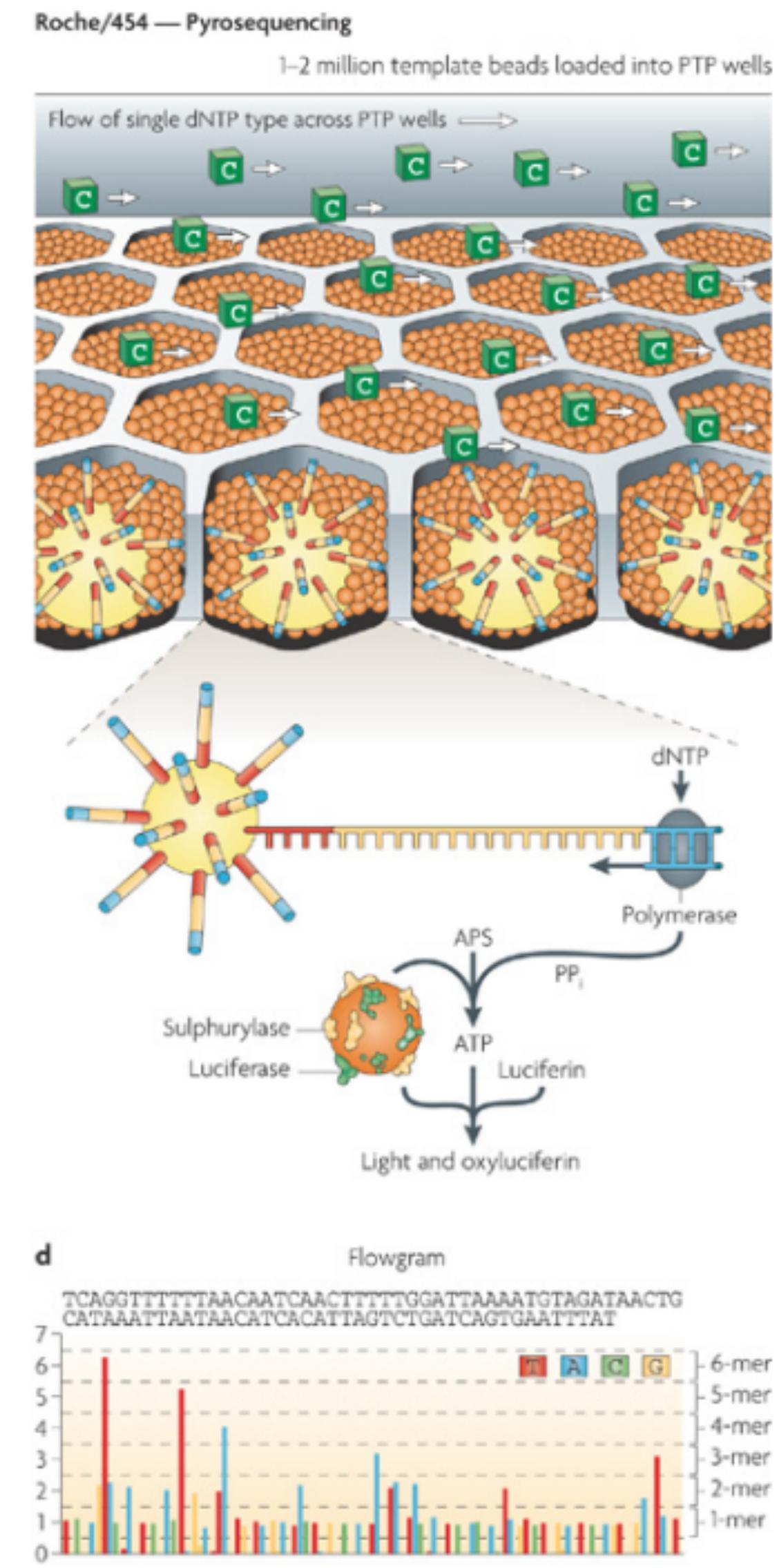


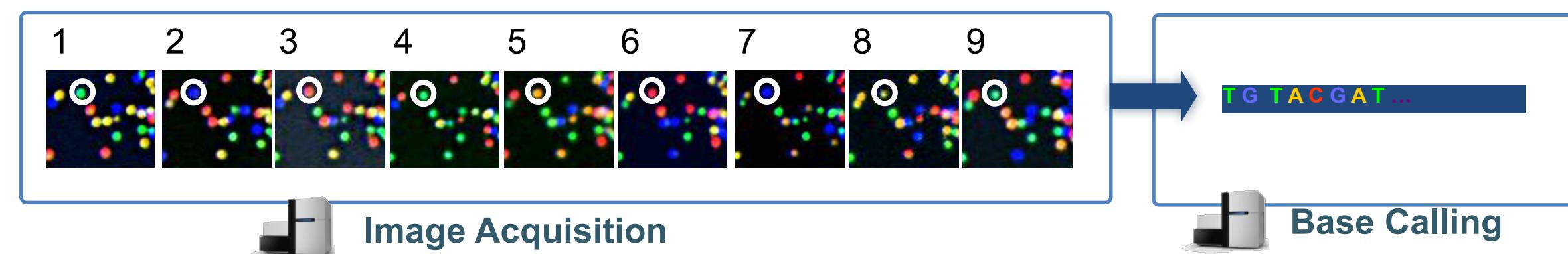
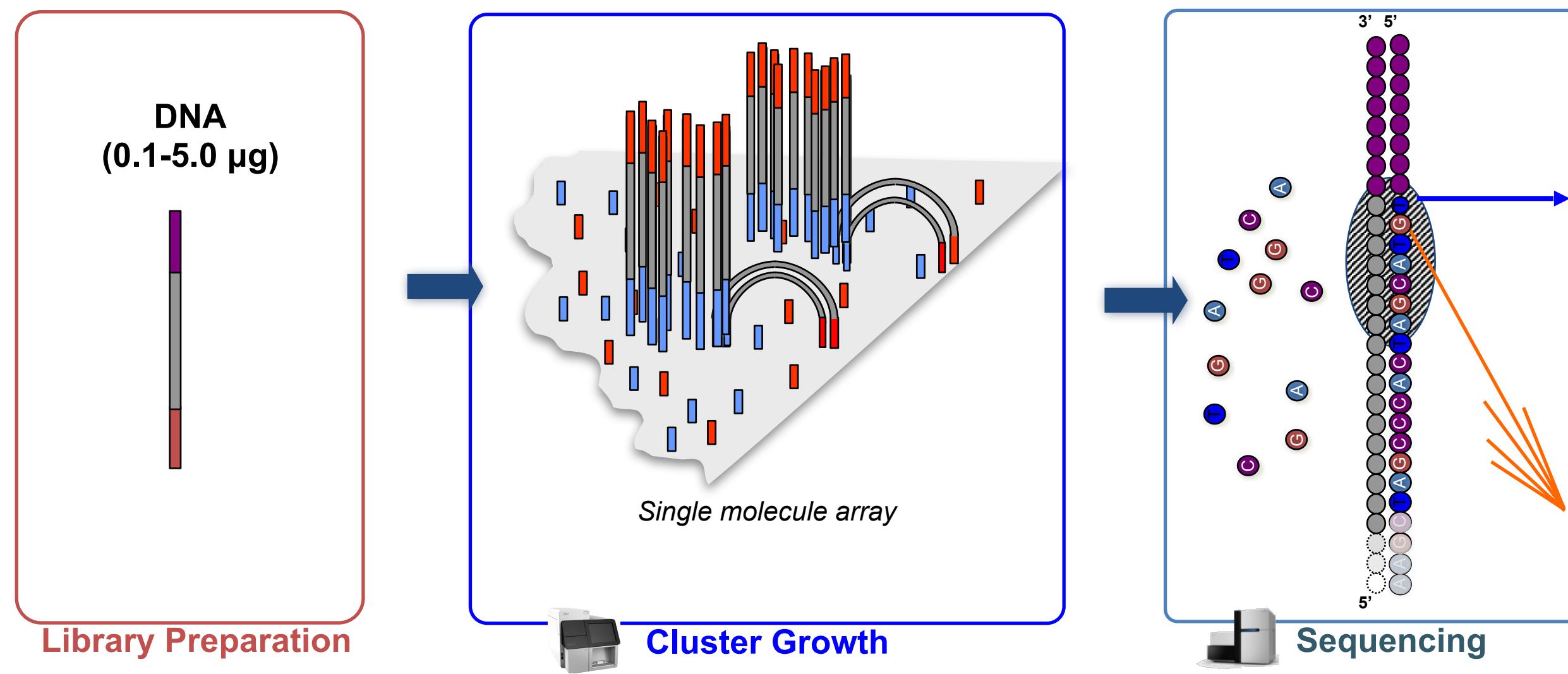


# Next-Generation Sequencing Technologies

# 454 Pyrosequencing

- ▶ DNA bound to capture beads for emulsion PCR (one strand per bead)
- ▶ Beads loaded onto plate (one bead per well)
- ▶ Addition of complementary base results in ATP production > activates luciferin > light production > CCD camera
- ▶ ‘Flowgram’ generated for each well





11

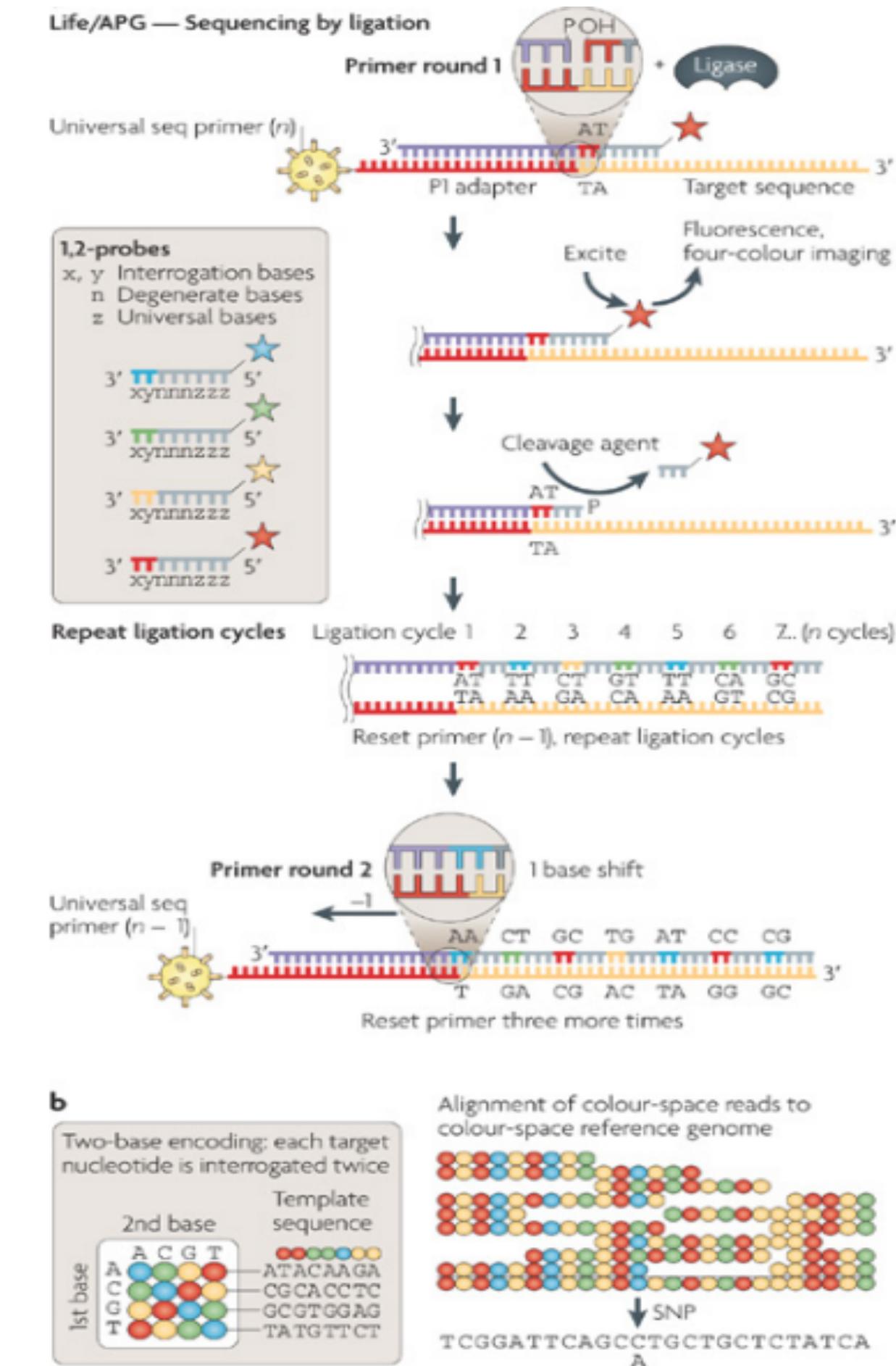


# Sequencing by Synthesis

# SOLID:

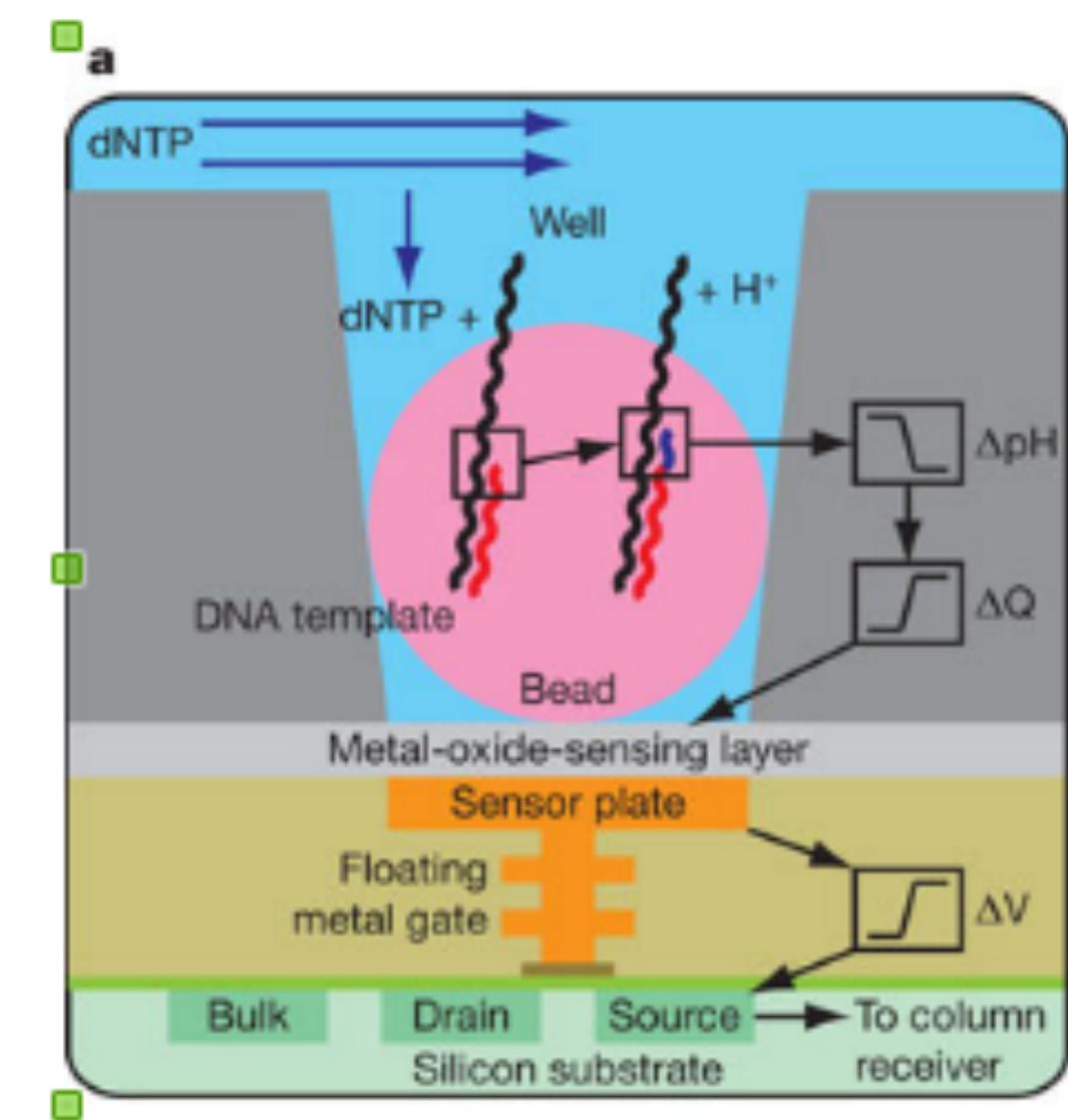
Sequencing by Oligonucleotide Ligation and Detection

- ▶ Emulsion PCR on beads which are attached to flow cell
- ▶ '2-base encoding' via oligonucleotides with degenerate sequence and fluorophore on 5' end
- ▶ Sequence extension by ligation
- ▶ Nine sets of primers, subsequently starting at n - 1,2,3....9
- ▶ All possible sequencing combinations need to be encoded

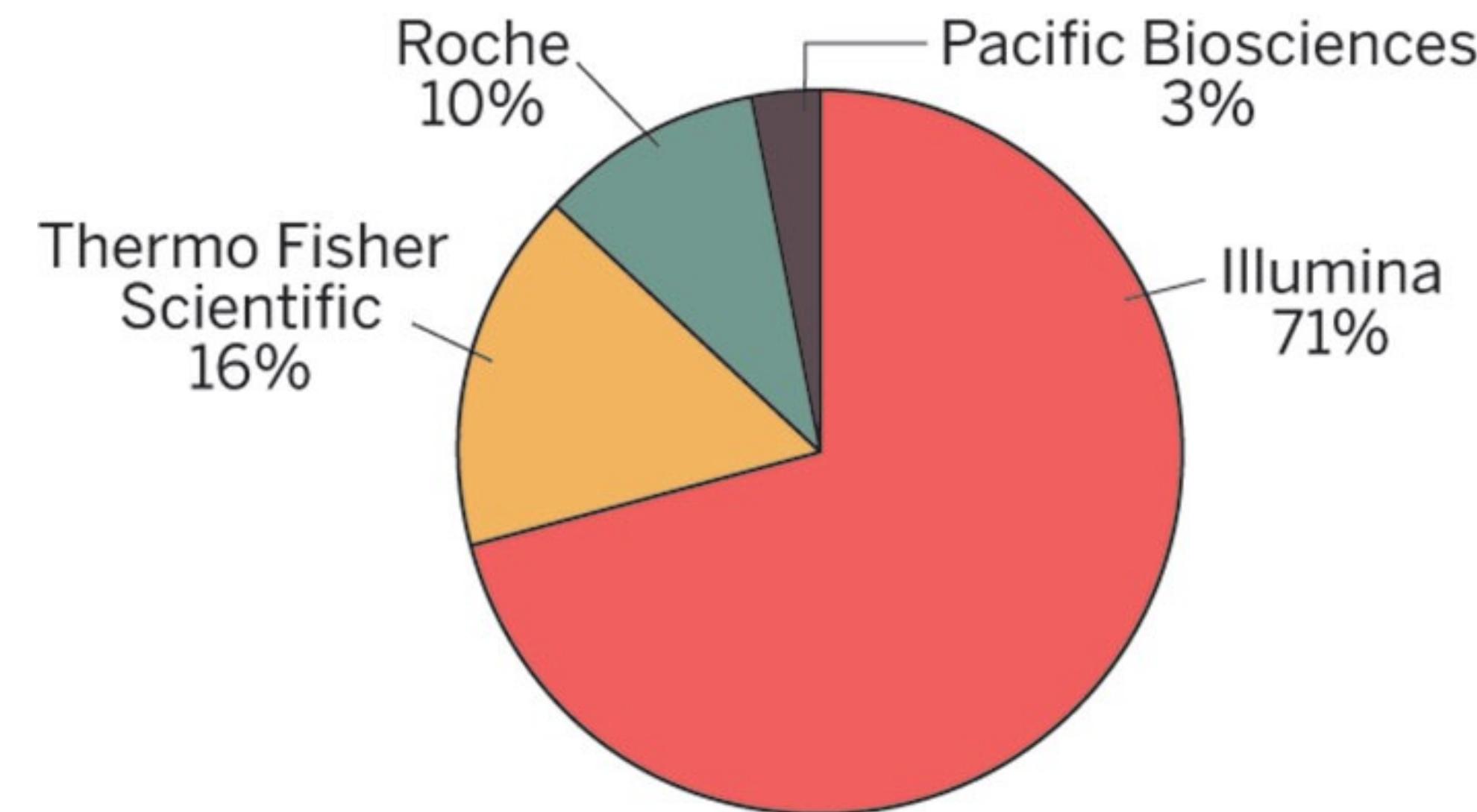


# IonTorrent

- ▶ Similar to 454 but uses semiconducting chip to detect dNTP incorporation
- ▶ Microwells with ion (pH) sensors
- ▶ Nucleotides flow sequentially over ion semiconductor → detecting hydrogen ions during polymerization
- ▶ IonProton and Ion PGM



# Illumina: the current ‘rulers’ of the throne



**World market in 2013 = \$1.3 billion**

SOURCES: Mizuho Securities USA, Frost & Sullivan

				
<b>MiniSeq System</b>	<b>MiSeq Series</b>	<b>NextSeq Series</b>	<b>HiSeq Series</b>	<b>HiSeq X Series*</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.

# Illumina's sequencing systems



Product	MiSeq	NextSeq 500*	HiSeq 2500	HiSeq 3000	HiSeq 4000	HiSeq X Five <sup>†</sup>	HiSeq X Ten <sup>†</sup>
Description	Focused power Speed and simplicity for targeted and small-genome sequencing	Flexible power Speed and simplicity for everyday genomics	Production power Power and efficiency for large-scale genomics	Production power Maximum throughput and lowest cost for production-scale genomics	Population power Maximum throughput and lowest cost for population- and production-scale human whole-genome sequencing		
Key methods	Small genome, amplicon, targeted gene panel sequencing	Everyday genome, exome, transcriptome sequencing, and more		Production-scale genome, exome, transcriptome sequencing, and more		Population-scale human whole-genome sequencing	
Run mode	—	Mid-output High-output	Rapid run High-output	— —	— —	— —	
Flow cells processed per run	1	1	1	1 or 2	1 or 2	1	1 or 2 1 or 2
Output range	0.3–15 Gb	20–39 Gb 30–120 Gb	10–300 Gb 50–1000 Gb 125–750 Gb 125–1500 Gb	900–1800 Gb 900–1800 Gb			
Run time	5–55 hours	15–26 hours	12–30 hours	7–60 hours	< 1–6 days	< 1–3.5 days	< 1–3.5 days < 3 days < 3 days
Reads per flow cell <sup>‡</sup>	25 million <sup>§</sup>	130 million 400 million	300 million 2 billion 2.5 billion 2.5 billion	3 billion 3 billion			
Maximum read length	2 × 300 bp	2 × 150 bp	2 × 150 bp	2 × 250 bp 2 × 125 bp	2 × 150 bp	2 × 150 bp	2 × 150 bp 2 × 150 bp

# Illumina's sequencing systems

<http://www.illumina.com/systems/sequencing.html>



Product	MiSeq	NextSeq 500*	HiSeq 2500	HiSeq 3000	HiSeq 4000	HiSeq X Five <sup>†</sup>	HiSeq X Ten <sup>†</sup>		
Description	Focused power Speed and simplicity for targeted and small-genome sequencing	Flexible power Speed and simplicity for everyday genomics	Production power Power and efficiency for large-scale genomics	Production power Maximum throughput and lowest cost for production-scale genomics	Population power Maximum throughput and lowest cost for population- and production-scale human whole-genome sequencing				
Key methods	Small genome, amplicon, targeted gene panel sequencing	Everyday genome, exome, transcriptome sequencing, and more		Production-scale genome, exome, transcriptome sequencing, and more		Population-scale human whole-genome sequencing			
Run mode	—	Mid-output High-output	Rapid run High-output	— —	— —	— —			
Flow cells processed per run	1	1	1	1 or 2	1 or 2	1	1 or 2 1 or 2		
Output range	0.3–15 Gb	20–39 Gb	30–120 Gb	10–300 Gb	50–1000 Gb	125–750 Gb	125–1500 Gb	900–1800 Gb	900–1800 Gb
Run time	5–55 hours	15–26 hours	12–30 hours	7–60 hours	< 1–6 days	< 1–3.5 days	< 1–3.5 days	< 3 days	< 3 days
Reads per flow cell <sup>‡</sup>	25 million <sup>§</sup>	130 million	400 million	300 million	2 billion	2.5 billion	2.5 billion	3 billion	3 billion
Maximum read length	2 × 300 bp	2 × 150 bp	2 × 150 bp	2 × 250 bp	2 × 125 bp	2 × 150 bp	2 × 150 bp	2 × 150 bp	2 × 150 bp

# Illumina's sequencing systems

<http://www.illumina.com/systems/sequencing.html>



Product	MiSeq	NextSeq 500*	HiSeq 2500	HiSeq 3000	HiSeq 4000	HiSeq X Five <sup>†</sup>	HiSeq X Ten <sup>†</sup>
Description	Focused power Speed and simplicity for targeted and small-genome sequencing	Flexible power Speed and simplicity for everyday genomics	Production power Power and efficiency for large-scale genomics	Production power Maximum throughput and lowest cost for production-scale genomics	Population power Maximum throughput and lowest cost for population- and production-scale human whole-genome sequencing		
Key methods	Small genome, amplicon, targeted gene panel sequencing	Everyday genome, exome, transcriptome sequencing, and more		Production-scale genome, exome, transcriptome sequencing, and more		Population-scale human whole-genome sequencing	
Run mode	—	Mid-output High-output	Rapid run High-output	— —	— —	— —	
Flow cells processed per run	1	1	1	1 or 2	1 or 2	1	1 or 2 1 or 2
Output range	0.3–15 Gb	20–39 Gb	30–120 Gb	10–300 Gb	50–1000 Gb	125–750 Gb	125–1500 Gb 900–1800 Gb
Run time	5–55 hours	15–26 hours	12–30 hours	7–60 hours	< 1–6 days	< 1–3.5 days	< 1–3.5 days < 3 days
Reads per flow cell <sup>‡</sup>	25 million <sup>§</sup>	130 million	400 million	300 million	2 billion	2.5 billion	2.5 billion 3 billion
Maximum read length	2 × 300 bp	2 × 150 bp	2 × 150 bp	2 × 250 bp	2 × 125 bp	2 × 150 bp	2 × 150 bp 2 × 150 bp

# Illumina's sequencing systems

<http://www.illumina.com/systems/sequencing.html>



Product	MiSeq	NextSeq 500*	HiSeq 2500	HiSeq 3000	HiSeq 4000	HiSeq X Five <sup>†</sup>	HiSeq X Ten <sup>†</sup>
Description	Focused power Speed and simplicity for targeted and small-genome sequencing	Flexible power Speed and simplicity for everyday genomics	Production power Power and efficiency for large-scale genomics	Production power Maximum throughput and lowest cost for production-scale genomics	Population power Maximum throughput and lowest cost for population- and production-scale human whole-genome sequencing		
Key methods	Small genome, amplicon, targeted gene panel sequencing	Everyday genome, exome, transcriptome sequencing, and more		Production-scale genome, exome, transcriptome sequencing, and more		Population-scale human whole-genome sequencing	
Run mode	—	Mid-output High-output	Rapid run High-output	— —	— —	— —	
Flow cells processed per run	1	1	1	1 or 2	1 or 2	1	1 or 2 1 or 2
Output range	0.3–15 Gb	20–39 Gb	30–120 Gb	10–300 Gb	50–1000 Gb	125–750 Gb	125–1500 Gb 900–1800 Gb
Run time	5–55 hours	15–26 hours	12–30 hours	7–60 hours	< 1–6 days	< 1–3.5 days	< 1–3.5 days < 3 days
Reads per flow cell <sup>‡</sup>	25 million <sup>§</sup>	130 million	400 million	300 million	2 billion	2.5 billion	2.5 billion 3 billion
Maximum read length	2 × 300 bp	2 × 150 bp	2 × 150 bp	2 × 250 bp	2 × 125 bp	2 × 150 bp	2 × 150 bp 2 × 150 bp

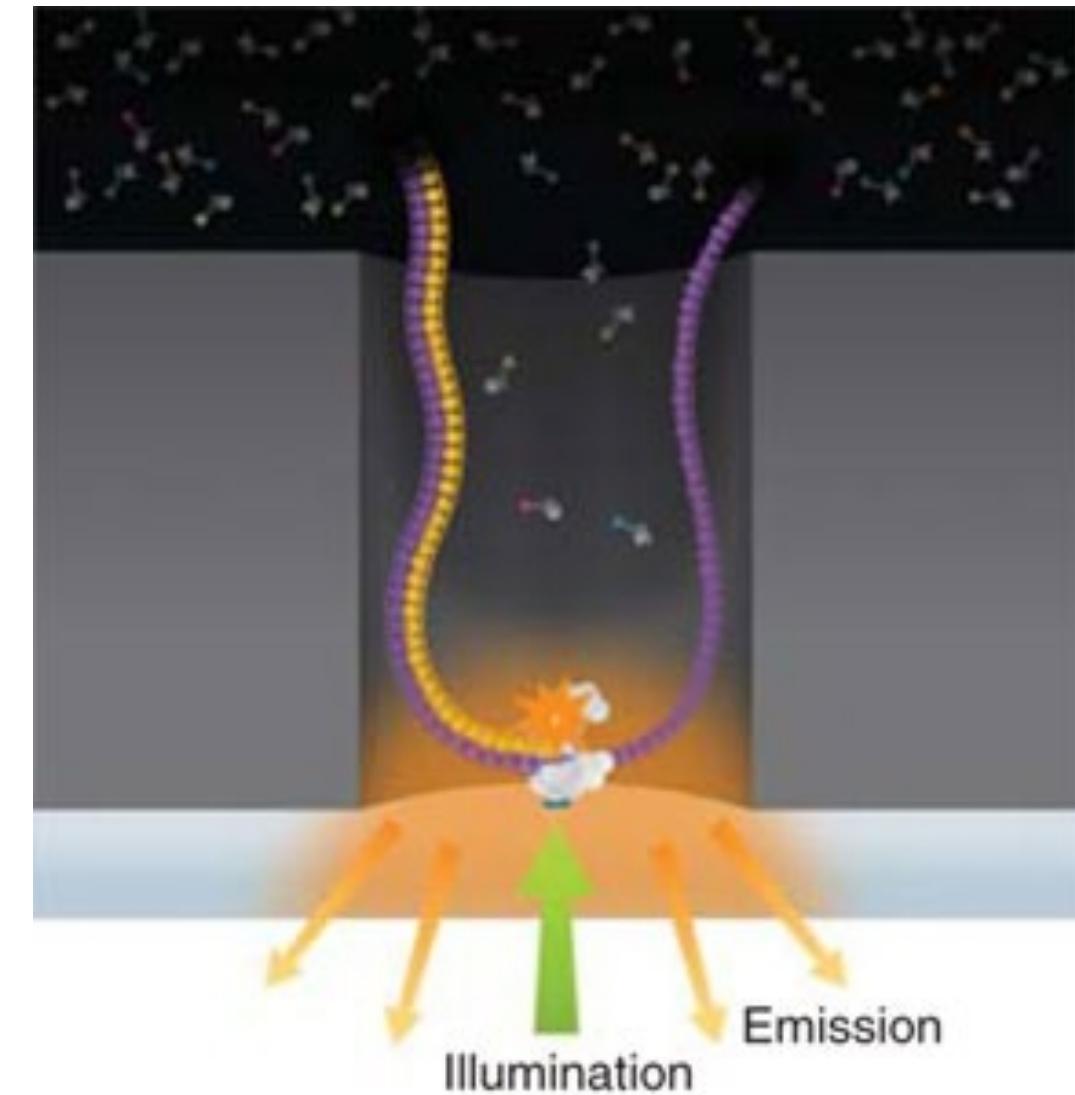
# Illumina's sequencing systems

# Third Generation Sequencing Technologies

# PacBio:

## Single molecule real-time sequencing (SMRT)

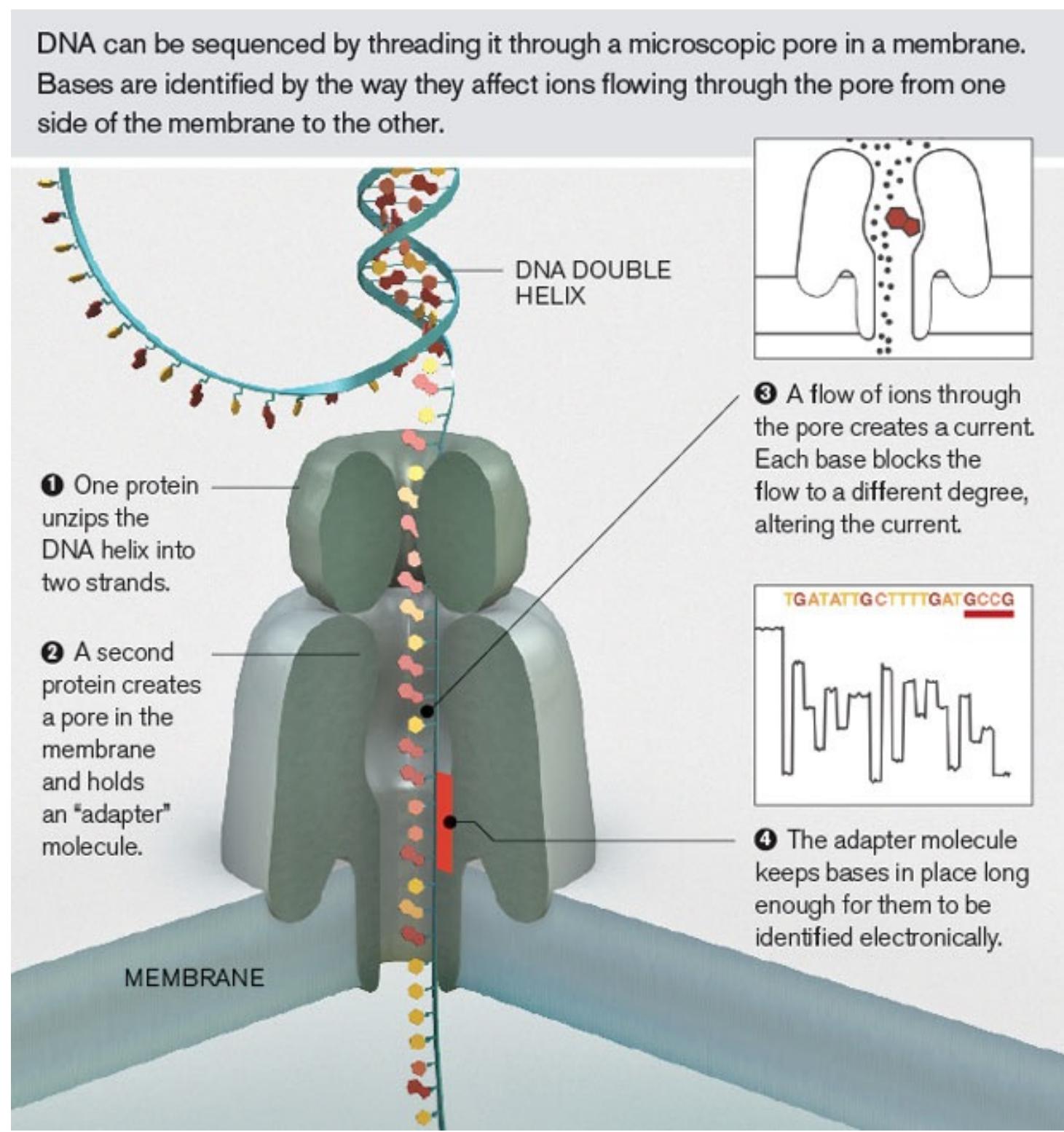
- ▶ DNA synthesis detected on a single strand and never halted
- ▶ Polymerase is affixed to the bottom of a tiny hole (~70 nm) in ZMW
- ▶ Only bottom portion of hole is illuminated allowing for detection of single dye-labeled nucleotide
- ▶ No amplification bias; least GC bias



# PacBio: RSII

- ▶ RS II released 2013; costly (~\$800K) and a behemoth of a machine
- ▶ Certified PacBio providers undergo rigorous training; many worldwide locations
- ▶ Long reads with reasonably high accuracy; ideal for whole genome assembly
- ▶ Low input sample amounts (10ng to 1 $\mu$ g)
- ▶ Up to 16 SMRT cells in one run





# Nanopore

- ▶ Single stranded DNA is passed through a protein pore created in a membrane (nanopore)
- ▶ Ions flow through the pore; measure the current as DNA goes through
- ▶ Bases are identified by the way they affect ions flowing through the pore (~6bp at a time)
- ▶ MinION, PromethION, GridION

Credit: John MacNeil

<http://www2.technologyreview.com/article/427677/nanopore-sequencing/>



## Nanopore: MinION

- ▶ Portable, real-time biological analyses; limited release launched May 2014
- ▶ Long(er) reads, with accuracies as good or better than PacBio with 2D reads
- ▶ Reliably sequence small genomes of bacteria and viruses for cheap (\$1K access fee)

# LETTER

doi:10.1038/nature16996

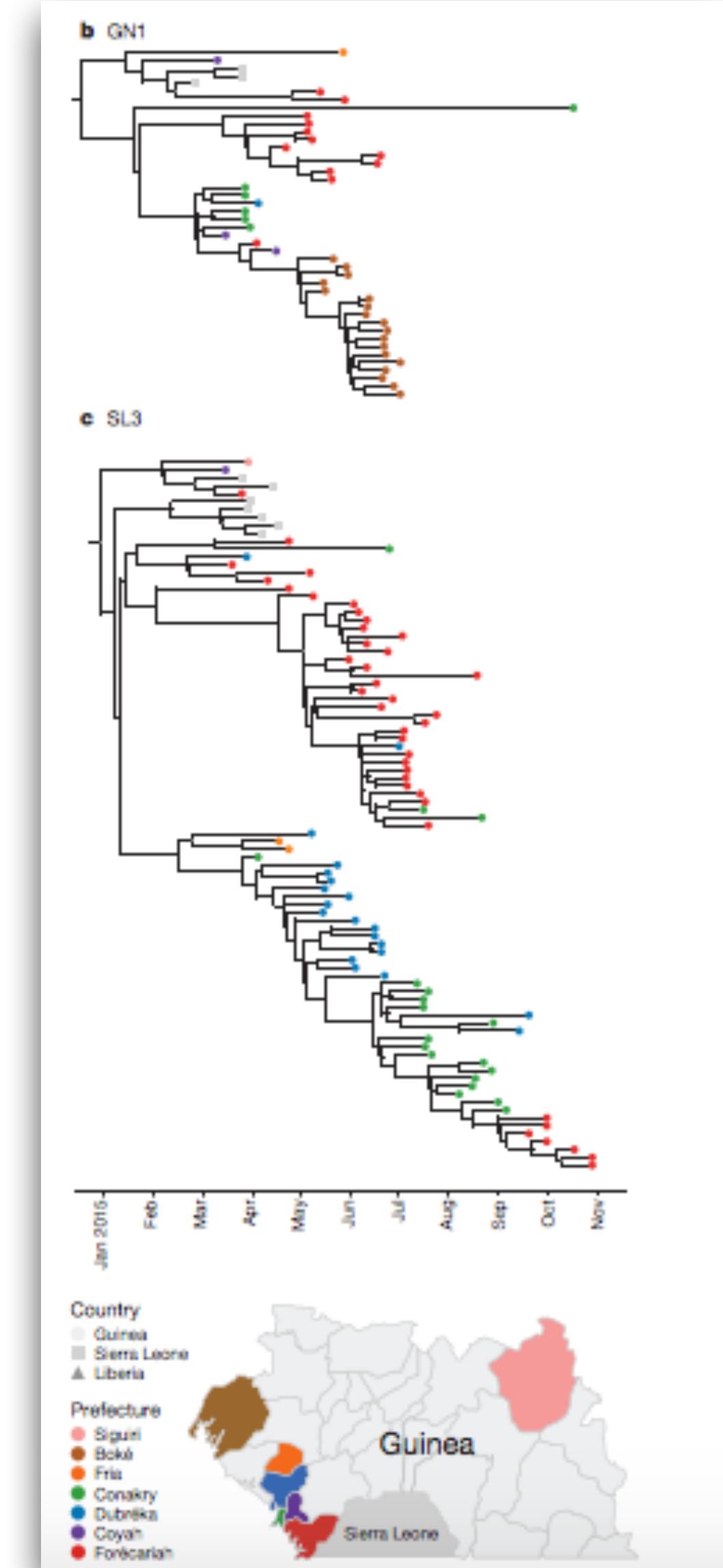
## Real-time, portable genome sequencing for Ebola surveillance

Joshua Quick<sup>1\*</sup>, Nicholas J. Loman<sup>1\*</sup>, Sophie Duraffour<sup>2,3\*</sup>, Jared T. Simpson<sup>4,5\*</sup>, Ettore Severi<sup>6\*</sup>, Lauren Cowley<sup>7\*</sup>, Joseph Akoi Bore<sup>2</sup>, Raymond Koundouno<sup>2</sup>, Gytis Dudas<sup>8</sup>, Amy Mikhail<sup>7</sup>, Nobila Ouédraogo<sup>9</sup>, Babak Afrough<sup>2,10</sup>, Amadou Bah<sup>2,11</sup>, Jonathan H. J. Baum<sup>2,3</sup>, Beate Becker-Ziaja<sup>2,3</sup>, Jan Peter Boettcher<sup>2,12</sup>, Mar Cabeza-Cabrero<sup>2,3</sup>, Álvaro Camino-Sánchez<sup>2</sup>, Lisa L. Carter<sup>2,13</sup>, Juliane Doerrbecker<sup>2,3</sup>, Theresa Enkirch<sup>2,14</sup>, Isabel García-Dorival<sup>2,15</sup>, Nicole Hetzelt<sup>2,12</sup>, Julia Hinzmman<sup>2,12</sup>, Tobias Holm<sup>2,3</sup>, Liana Eleni Kafetzopoulou<sup>2,16</sup>, Michel Koropogui<sup>2,17</sup>, Abigael Kosgey<sup>2,18</sup>, Eeva Kuismä<sup>2,10</sup>, Christopher H. Logue<sup>2,10</sup>, Antonio Mazzarelli<sup>2,19</sup>, Sarah Meisel<sup>2,3</sup>, Marc Mertens<sup>2,20</sup>, Janine Michel<sup>2,12</sup>, Didier Ngabo<sup>2,10</sup>, Katja Nitzsche<sup>2,3</sup>, Elisa Pallasch<sup>2,3</sup>, Livia Victoria Patrono<sup>2,3</sup>, Jasmine Portmann<sup>2,21</sup>, Johanna Gabriella Repits<sup>2,22</sup>, Natasha Y. Rickett<sup>2,15,23</sup>, Andreas Sachse<sup>2,12</sup>, Katrien Singethan<sup>2,24</sup>, Inês Vitoriano<sup>2,10</sup>, Rahel L. Yemanaberhan<sup>2,3</sup>, Elsa G. Zekeng<sup>2,15,23</sup>, Trina Racine<sup>25</sup>, Alexander Bello<sup>25</sup>, Amadou Alpha Sall<sup>26</sup>, Ousmane Faye<sup>26</sup>, Oumar Faye<sup>26</sup>, N'Faly Magassouba<sup>27</sup>, Cecelia V. Williams<sup>28,29</sup>, Victoria Amburgey<sup>28,29</sup>, Linda Winona<sup>28,29</sup>, Emily Davis<sup>29,30</sup>, Jon Gerlach<sup>29,30</sup>, Frank Washington<sup>29,30</sup>, Vanessa Montell<sup>31</sup>, Marine Joudain<sup>31</sup>, Marion Bererd<sup>31</sup>, Alimou Camara<sup>31</sup>, Hermann Somlare<sup>31</sup>, Abdoulaye Camara<sup>31</sup>, Marianne Gerard<sup>31</sup>, Guillaume Bado<sup>31</sup>, Bernard Bailler<sup>31</sup>, Déborah Delaune<sup>32,33</sup>, Koumpingnin Yacouba Nebie<sup>34</sup>, Abdoulaye Diarra<sup>34</sup>, Yacouba Savane<sup>34</sup>, Raymond Bernard Pallawo<sup>34</sup>, Giovanna Jaramillo Gutierrez<sup>35</sup>, Natacha Milhano<sup>6,36</sup>, Isabelle Roger<sup>34</sup>, Christopher J. Williams<sup>6,37</sup>, Facinet Yattara<sup>17</sup>, Kuiama Lewandowski<sup>10</sup>, James Taylor<sup>38</sup>, Phillip Rachwal<sup>38</sup>, Daniel J. Turner<sup>39</sup>, Georgios Pollakis<sup>15,23</sup>, Julian A. Hiscox<sup>15,23</sup>, David A. Matthews<sup>40</sup>, Matthew K. O'Shea<sup>41</sup>, Andrew McD. Johnston<sup>41</sup>, Duncan Wilson<sup>41</sup>, Emma Hutley<sup>42</sup>, Erasmus Smit<sup>43</sup>, Antonino Di Caro<sup>2,19</sup>, Roman Wölfel<sup>2,44</sup>, Kilian Stoecker<sup>2,44</sup>, Erna Fleischmann<sup>2,44</sup>, Martin Gabriel<sup>2,3</sup>, Simon A. Weller<sup>38</sup>, Lamine Koivogui<sup>45</sup>, Boubacar Diallo<sup>3</sup>, Sakoba Keïta<sup>17</sup>, Andrew Rambaut<sup>8,46,47</sup>, Pierre Formenty<sup>34</sup>, Stephan Günther<sup>2,3</sup> & Miles W. Carroll<sup>2,10,48,49</sup>

- Sequence data and analysis of 142 EBOV samples
- Generate results < 24h after receiving sample, with sequencing complete between 15-60 minutes



**Figure 1 | Deployment of the portable genome surveillance system in Guinea.** **a**, We were able to pack all instruments, reagents and disposable consumables within aircraft baggage. **b**, We initially established the genomic surveillance laboratory in Donka Hospital, Conakry, Guinea. **c**, Later we moved the laboratory to a dedicated sequencing laboratory in Coyah prefecture. **d**, Within this laboratory we separated the sequencing instruments (on the left) from the PCR bench (to the right). An uninterrupted power supply can be seen in the middle that provides power to the thermocycler. (Photographs taken by J.Q. and S.D.)



# Comparison of NGS technologies

	Max Read Length	Runtime	Output per run	Error rates
Illumina HiSeq2500	2x125 bp	6 days	450-500 Gb	0.26%
IonTorrent PGM	~200 bp	2 hours	20-50 Mb on 314 100-200 Mb on 316 1Gb on 318 chip	1.71%
PacBio RS II	Avg 10 Kb up to 30 Kb	2 hours	500Mb -1Gb	~10%*
MinION	Avg 10 Kb up to 50 Kb	no fixed run-time	200 - 300 Mb (varies)	~15%*

\*prior to error correction methods

# A rapidly evolving market

- ▶ Illumina's HiSeq 3000 and HiSeq 4000; an intermediate between the HiSeq X Ten and the well-loved HiSeq 2500

# A rapidly evolving market

- ▶ Illumina's HiSeq 3000 and HiSeq 4000; an intermediate between the HiSeq X Ten and the well-loved HiSeq 2500
- ▶ Complete Genomics introduced the Revelocity system; high parallelization can deliver 10,000 whole genomes a year with aims to triple that output with future upgrades (on hold)

# A rapidly evolving market

- ▶ Illumina's HiSeq 3000 and HiSeq 4000; an intermediate between the HiSeq X Ten and the well-loved HiSeq 2500
- ▶ Complete Genomics introduced the Revelocity system; high parallelization can deliver 10,000 whole genomes a year with aims to triple that output with future upgrades (on hold)
- ▶ Thermo Fisher announced Ion S5 series; an intermediate between Proton and PGM, easier to use and focus on targeted panels

# A rapidly evolving market

- ▶ Illumina's HiSeq 3000 and HiSeq 4000; an intermediate between the HiSeq X Ten and the well-loved HiSeq 2500
- ▶ Complete Genomics introduced the Revelocity system; high parallelization can deliver 10,000 whole genomes a year with aims to triple that output with future upgrades (on hold)
- ▶ Thermo Fisher announced Ion S5 series; an intermediate between Proton and PGM, easier to use and focus on targeted panels
- ▶ PacBio's new Sequel System; a more price effective version of RSII, a third the size and almost seven times more powerful

# A rapidly evolving market

- ▶ Illumina's HiSeq 3000 and HiSeq 4000; an intermediate between the HiSeq X Ten and the well-loved HiSeq 2500
- ▶ Complete Genomics introduced the Revelocity system; high parallelization can deliver 10,000 whole genomes a year with aims to triple that output with future upgrades (on hold)
- ▶ Thermo Fisher announced Ion S5 series; an intermediate between Proton and PGM, easier to use and focus on targeted panels
- ▶ PacBio's new Sequel System; a more price effective version of RSII, a third the size and almost seven times more powerful
- ▶ Oxford Nanopore released fast-mode early access of MinION (Jan 2016); DNA transit speeds substantially increased with sequence quality preserved (550 bp/sec), Rapid 1D prep using transposase like Nextera kit

*These materials have been developed by members of the teaching team at the [Harvard Chan Bioinformatics Core \(HBC\)](#). These are open access materials distributed under the terms of the [Creative Commons Attribution license \(CC BY 4.0\)](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.*

