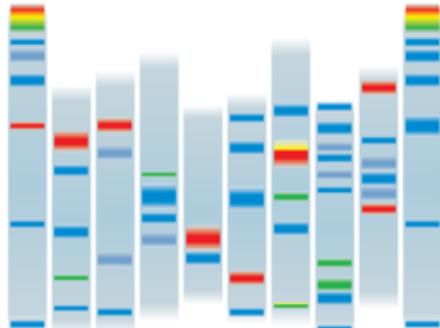


Introduction to ‘next-generation’ sequencing (NGS)

Dr David Studholme

d.j.studholme@exeter.ac.uk

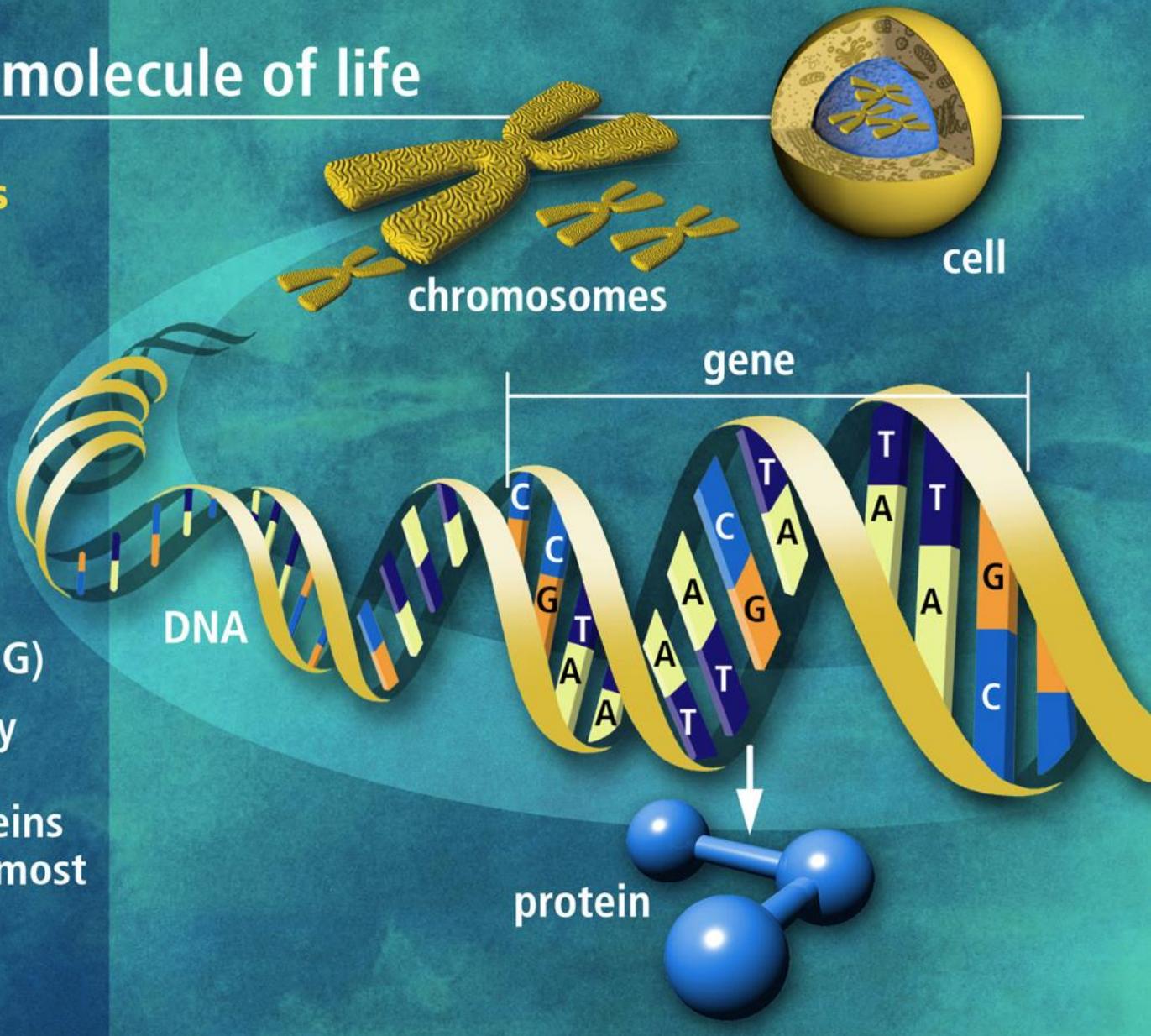


DNA the molecule of life

Trillions of cells

Each cell:

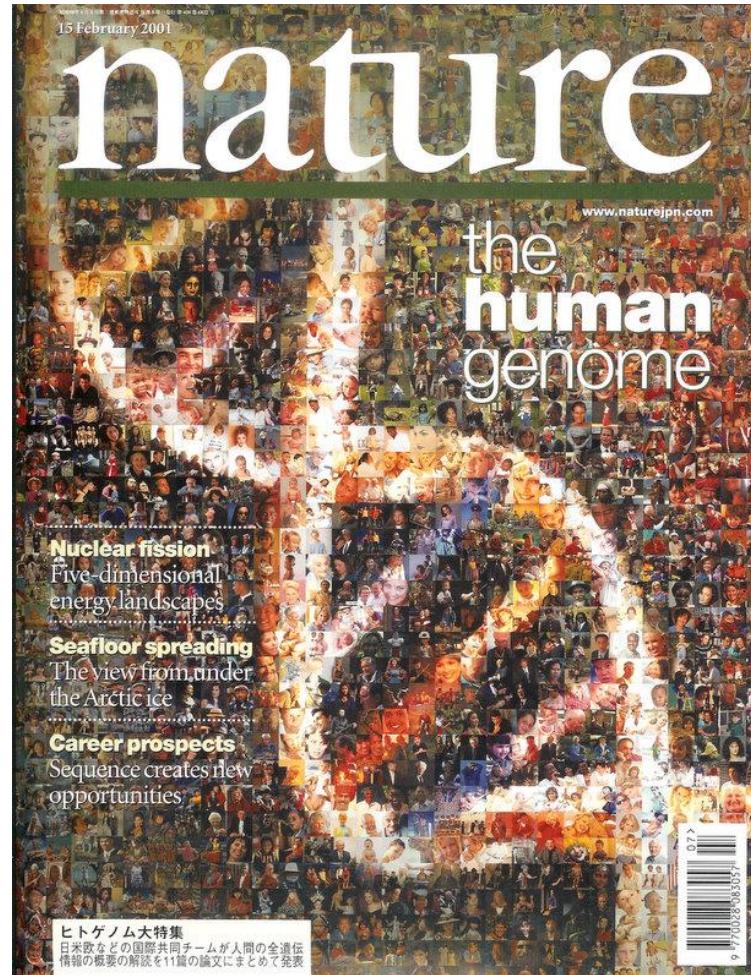
- 46 human chromosomes
- 2 meters of DNA
- 3 billion DNA subunits (the bases: A, T, C, G)
- Approximately 30,000 genes code for proteins that perform most life functions



<http://biocomicals.blogspot.com/2012/05/no-more-moores-law.html>

Credit: Alper Uzun www.biocomicals.com

2001: the end of the beginning



Genome sequencing before 2006



Sequencing today



Illumina NovaSeq 6000



Illumina MiSeq



Oxford Nanopore PromethION



Next-generation sequencing (NGS)

<https://biocomicals.blogspot.com/2011/11/squeezing-your-genome-is-cheaper-then.html>

Credit: Alper Uzun www.biocomicals.com

NGS can be highly portable



Portable DNA sequencing devices



Genomics
@GenomicsRR



Following

DNA Sequencing in the Final Frontier
rightrelevance.com/search/article ...



RETWEET
1 LIKES
7



6:47 PM - 22 Oct 2016



02/08/2022

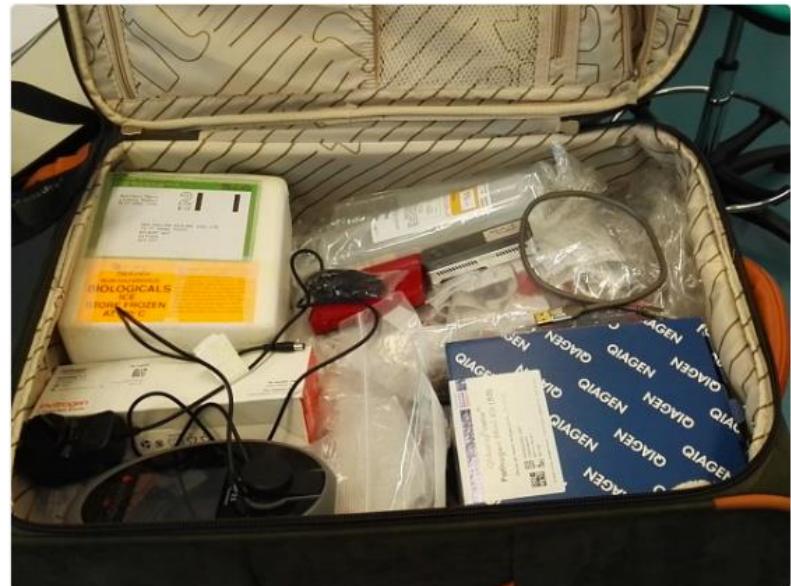


justin ogrady
@Justin_OGrady



Following

Lab in a suitcase for a trip to Zambia to diagnose childhood pneumonia using @nanopore sequencing.



RETWEETS
46 LIKES
102

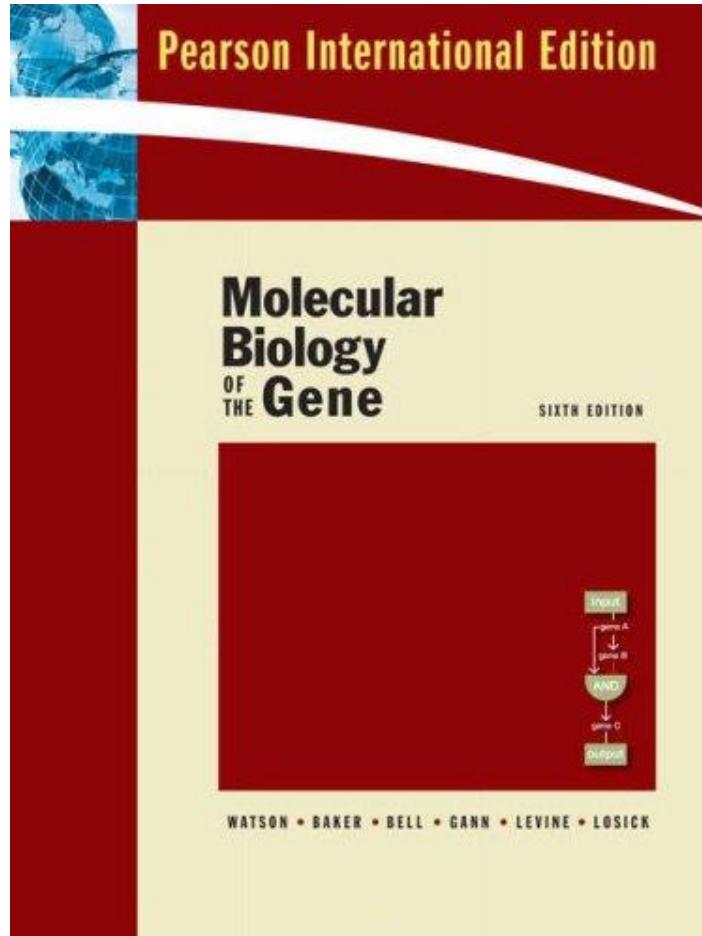


11:39 AM - 9 Sep 2016



12

NGS: Beyond genome sequencing

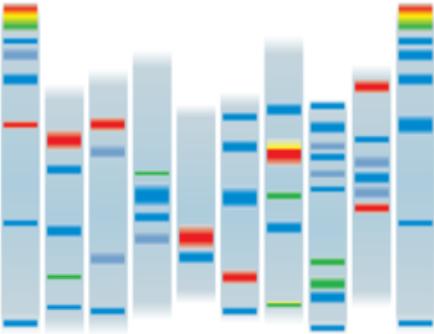


NGS-enabled study of epigenetic modifications in brain tumour genomes

Figure taken from: Etcheverry *et al.* (2010) DNA methylation in glioblastoma: impact on gene expression and clinical outcome. *BMC Genomics*. **11**:701. <http://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-11-701>

Questions?



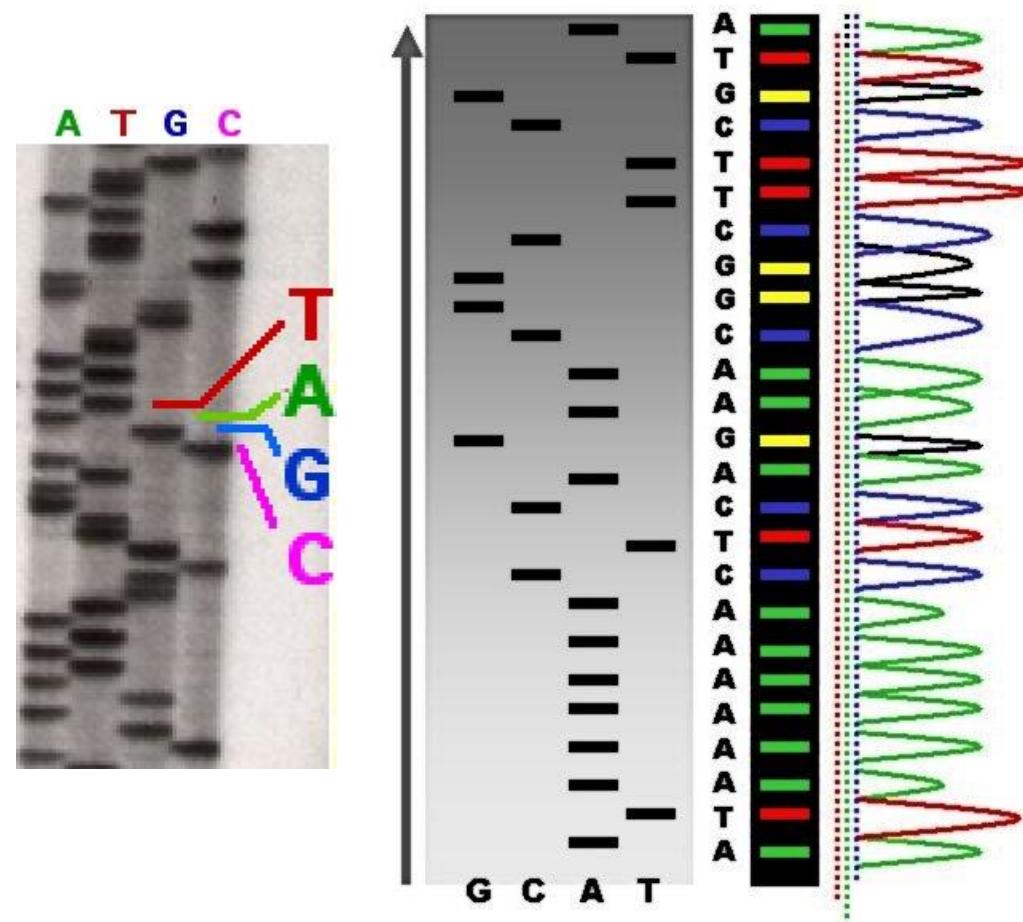


SANGER SEQUENCING

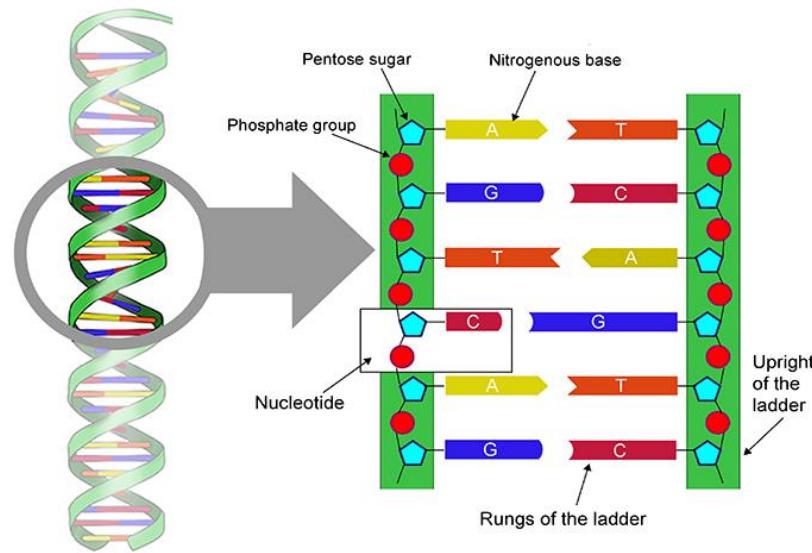
Sanger (dideoxy/chain termination) method is an example of DNA sequencing by synthesis



Frederick Sanger (1918 – 2013)
Inventor of dideoxy sequencing

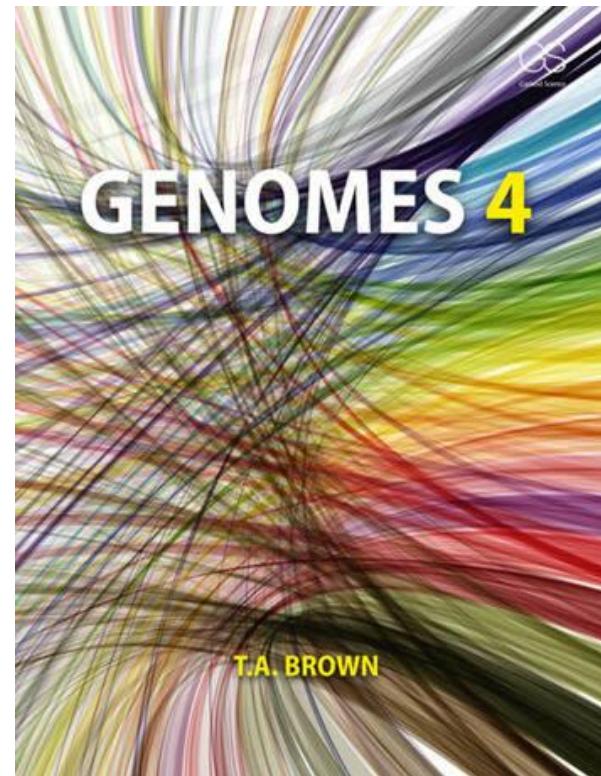


Sequencing by synthesis relies on: base pairing and DNA polymerase

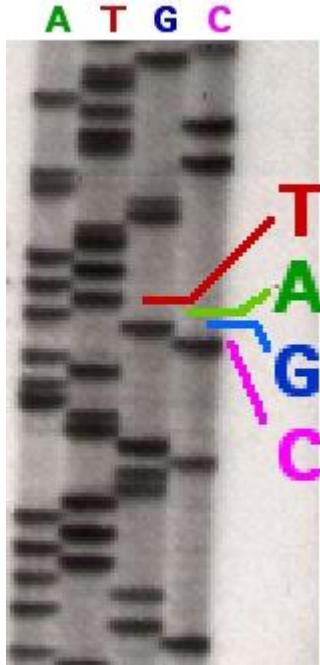


<http://mammothmemory.net/>

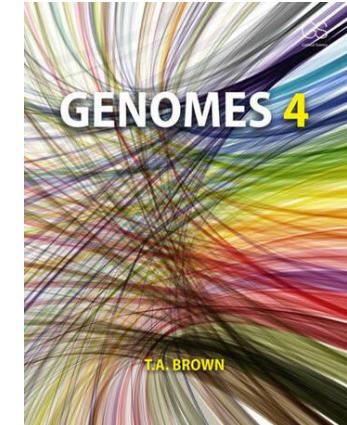
See Figure 4.2 of Genomes 4
By T. A. Brown
ISBN 9780815345084
Published June 21, 2017 by Garland Science



Reading the sequence: electrophoresis using gels or capillaries

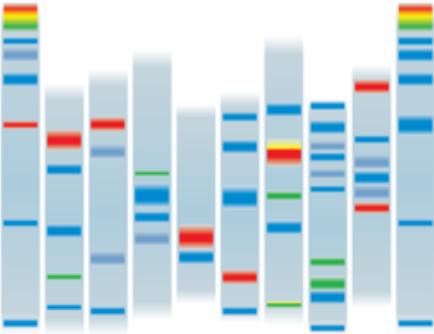


See Figure 4.3 of Genomes 4
By T. A. Brown
ISBN 9780815345084
Published June 21, 2017 by Garland Science



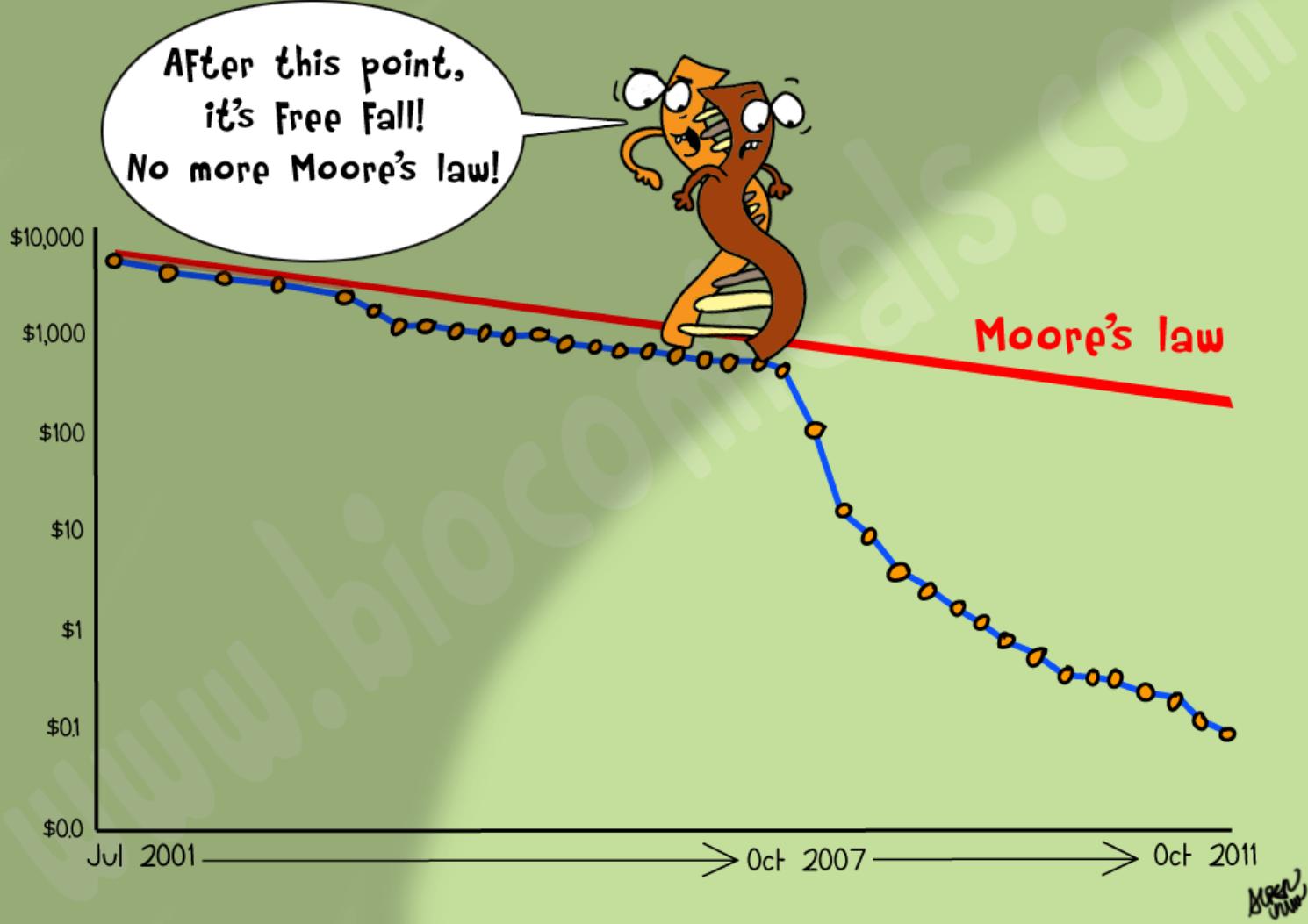
- Radioactive labelling of ddNTPs
- One ddNTP per lane of gel
- Fluorescent labeling of ddNTPs
- One ‘colour’ for each base (A, C, G, T)
- Four samples run together on capillary





SECOND-GENERATION SEQUENCING

Cost per raw Mb of DNA sequence



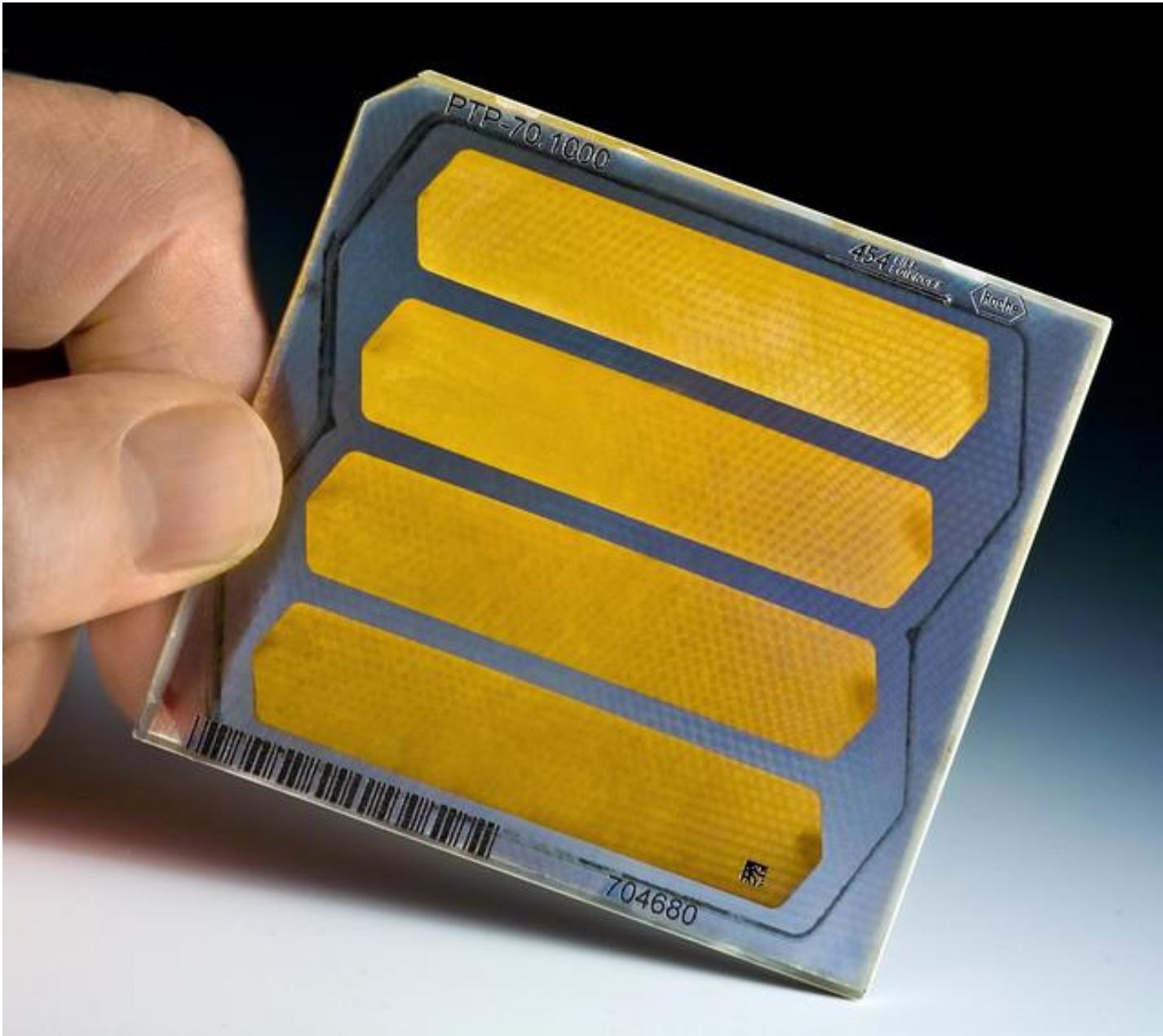
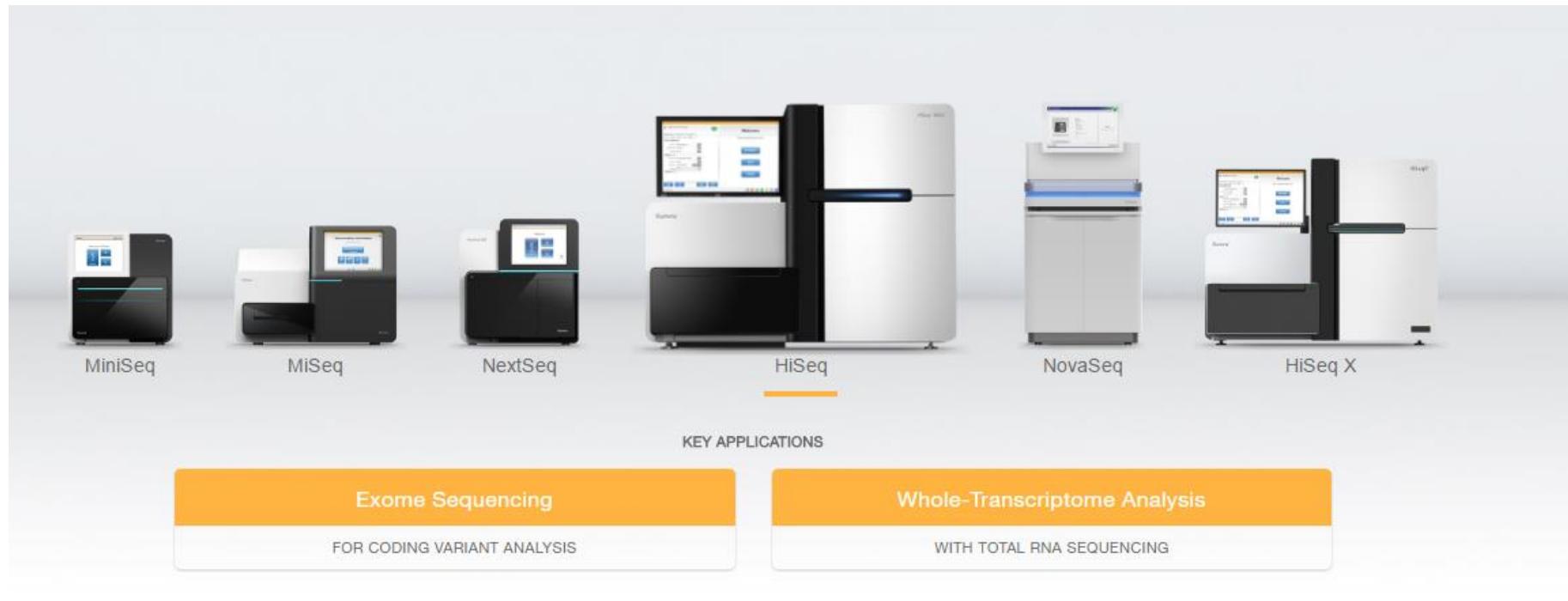


Image from: Mardis ER. The impact of next-generation sequencing technology on genetics. *Trends Genet.* 2008;24(3):133-141. doi:10.1016/j.tig.2007.12.007

**OB
SO
LE
TE
!**



Current market-leaders in second-generation sequencing: Illumina



The image displays the Illumina sequencing platform family, featuring six models arranged horizontally from left to right: MiniSeq, MiSeq, NextSeq, HiSeq, NovaSeq, and HiSeq X. Each model is shown with its respective control unit and reagent cartridge. Below the platforms, a horizontal bar indicates the 'KEY APPLICATIONS' for each model.

KEY APPLICATIONS

- Exome Sequencing**
FOR CODING VARIANT ANALYSIS
- Whole-Transcriptome Analysis**
WITH TOTAL RNA SEQUENCING

Power and efficiency for production-scale genomics.

[Explore HiSeq 2500](#) [Explore HiSeq 3000/HiSeq 4000](#) [Explore All Sequencing Platforms](#)

Illumina NovaSeq 6000



Illumina MiSeq



How does Illumina/Solexa sequencing work?



<https://youtu.be/fCd6B5HRaZ8>

Illumina: sequencing by synthesis

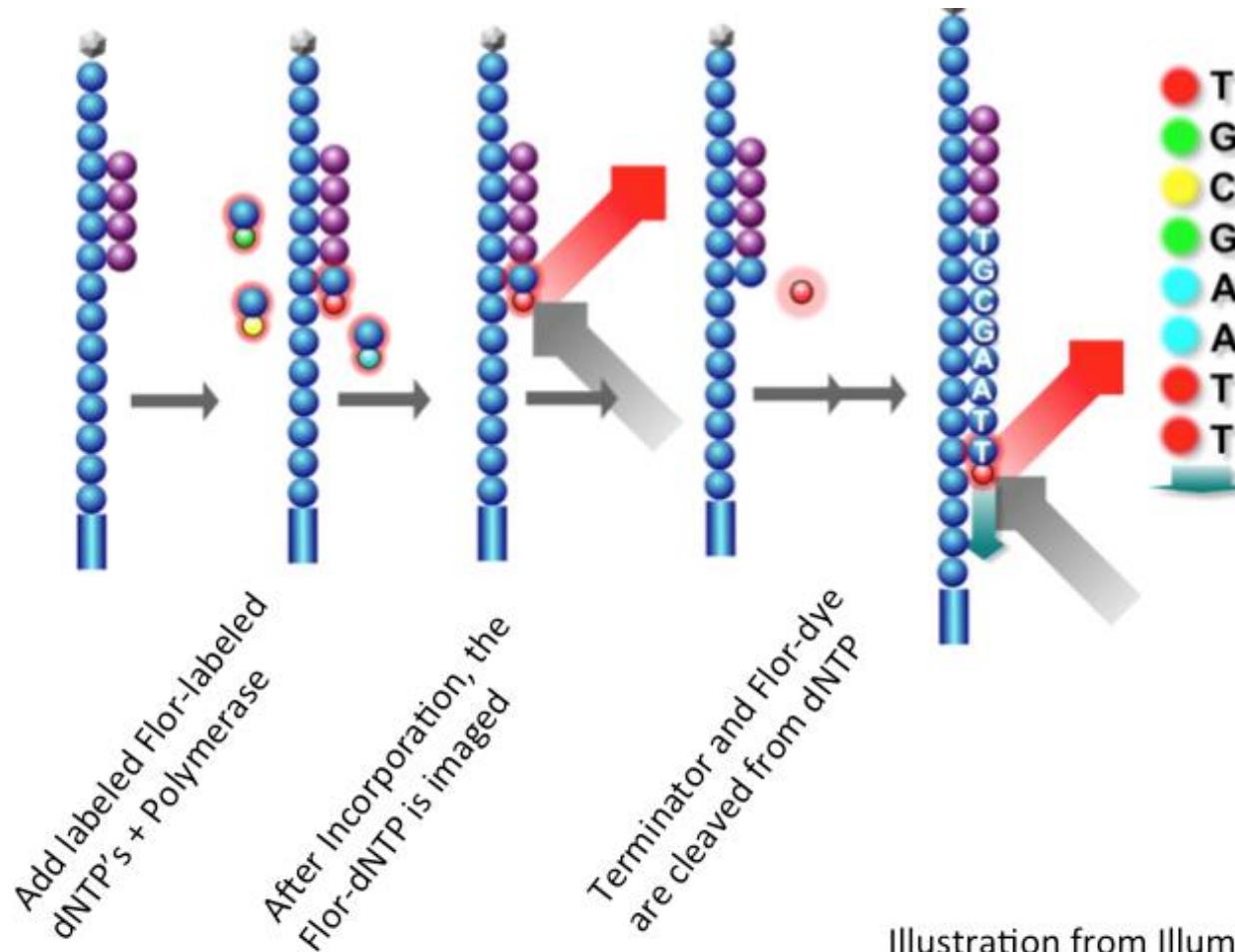
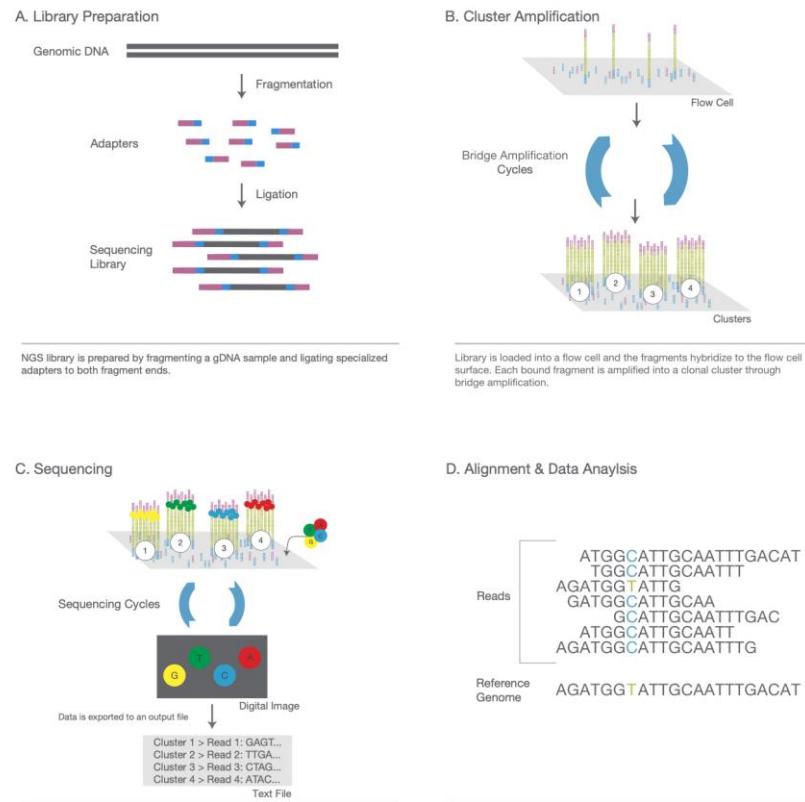
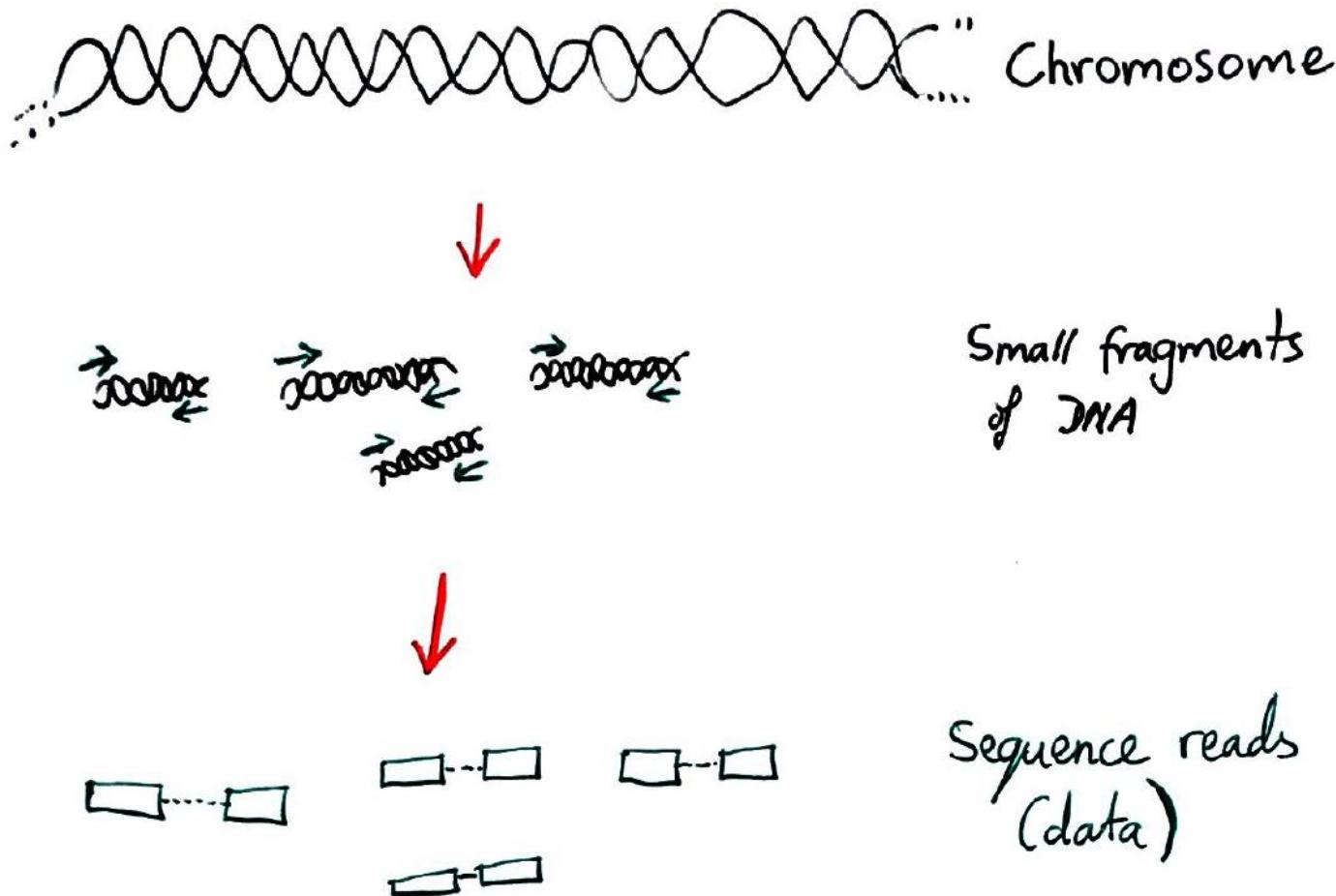


Illustration from Illumina, Inc



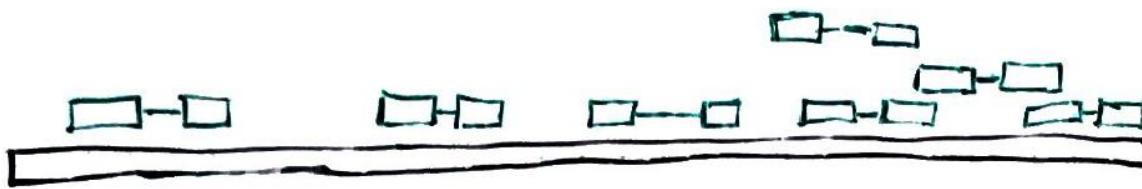
From: <https://emea.illumina.com/content/dam/illumina-marketing/documents/products/other/ivf-reproductive-genetic-health/ngs-primer-1570-2015-012.pdf>.

Shotgun sequencing approach



Two approaches to analysing NGS data

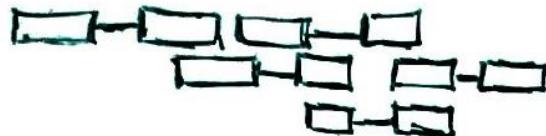
Align against reference genome



Sequence reads aligned

Reference genome
sequence

De-novo assembly



Sequence reads with
overlapping sequences

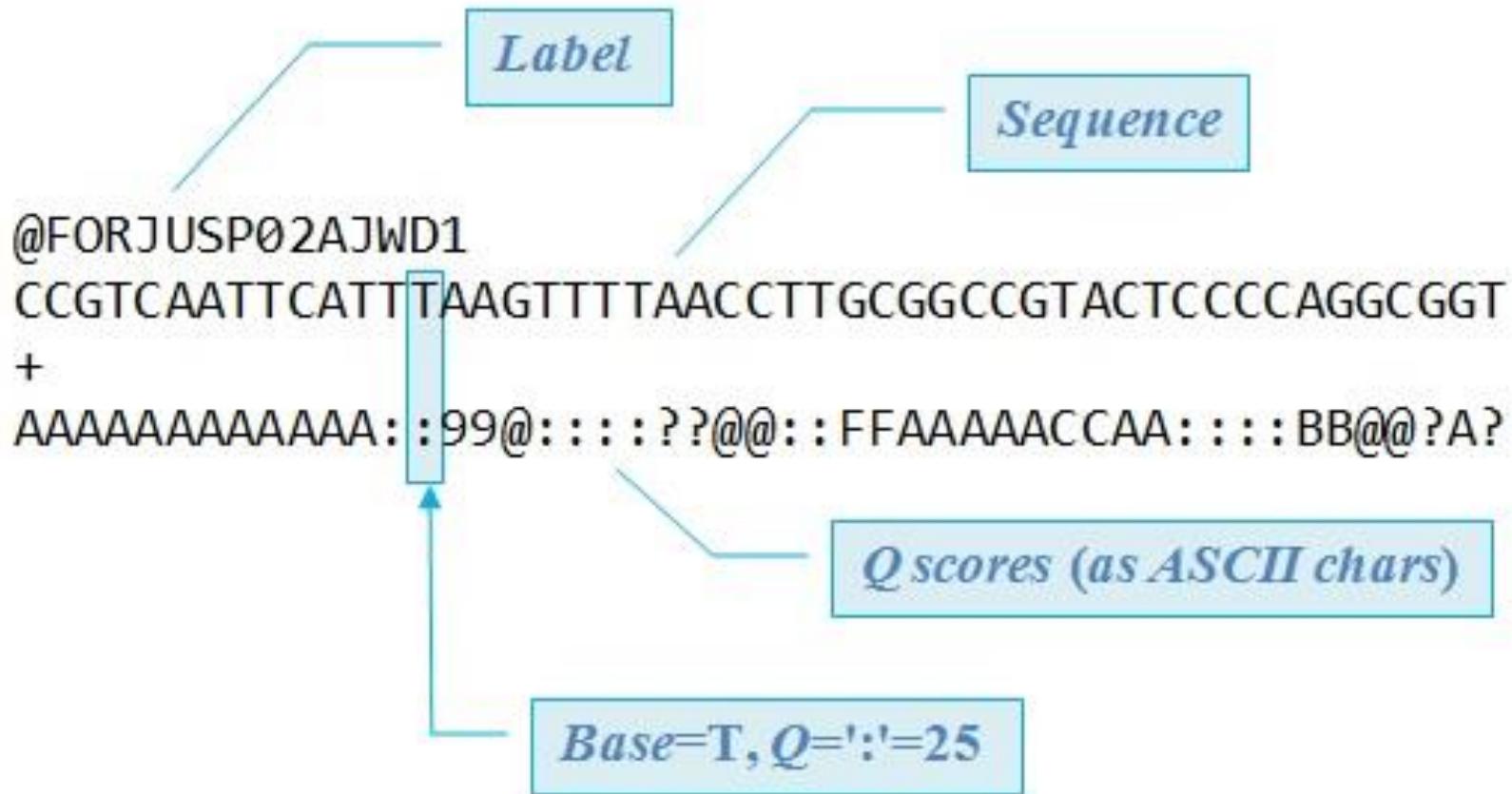
FastQ format

```
@A00283:59:HC7CWD$XX:1:1101:1904:1031 1:N:0
TNAAGTAATCGAGAACTGGCTACCCGAGGAACATAAAAACGCTACGCAC$TCAGTTAGCAGGCTACTACCGCAAGTCATGAAAAACTATGAAAGATCAGTGCACCCCTATTCTTGCTGAAAAAGATGTTTCAATG
+
F#FFFFFFFFFFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFF:FF:FFFF:FFFF
@A00283:59:HC7CWD$XX:1:1101:2157:1031 1:N:0
GNCTTTACTAATGGGTCAATTCTCCTCCCTCACAGGATTAGCCAGTCATTCTGTGCTACTGAGTCCA$ACTCTAATAGTTGTTAGCAAAGTCTGAGTTCTGGGTCATTCTGTCTTGATTTGTTCTTATTACTAAC
+
F#FFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFF:FFFF
@A00283:59:HC7CWD$XX:1:1101:5195:1031 1:N:0
CNTAGAGAGCATTCGCTGACTGTCAAGGCCAACTCTGAATCTCGTCAGAACTATAAATATTTGTAATACTATGATCAGTTGCTATACAAAGGTAGCTATAAAATCTGTCTTAAGTTGCTACATGCACGCACATATAA
+
F#FFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFF:FFFF
@A00283:59:HC7CWD$XX:1:1101:5321:1031 1:N:0
ANTAGAAAAGAAACTATACGATGGGAAGAAGTTAGTAACCTTGACGAGTATAGTAACAATATCAAAGTTCTGAATCAACAATTATGATAGGATAGCAACAGCTTGCAAGAACAGATTCTGGAATGATGGACTTTCTATA
+
F#FFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFF:FFFF
@A00283:59:HC7CWD$XX:1:1101:5990:1031 1:N:0
ANGACTTTGAAGGCATCCAATGATCATTTTGTATGAAGCAAATATATATATATACATATATCGTAGTATCATCTAAAGGAGATGAAGTATTCGACCATGATGTCATGTCGACAACTATACGAT
+
F#FFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFF:FFFF:FFFF:FFFF:FFFF:FF
@A00283:59:HC7CWD$XX:1:1101:7437:1031 1:N:0
CNTAGAAATCTCATTAGCCCTATGAGATAAATTGCAACCCCTACAAGATAGTAACACTGGTCAATCAGAACGAGCAAGAGAGTGAATTGACCATAAACCTTTCTTAGATAAAAGCTATTAACTGCCA
+
F#FFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFF:FFFFFFFFFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFF
@A00283:59:HC7CWD$XX:1:1105:31638:35744 1:N:0
TTGCTTCTGTTAATCAGATTGAACCAACATGGAACCTCTCATCAAAGAGCTCATCTTGCATGTCACTTGAGGATTGATTGTTCTGGAAACGACAAGTCAGACTGGAAACAGGTTCTGCTGCAAATTCTAAATCATT
+
F#FFFFFFFFFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FF:FF:FF:FF:FFFF:FFF
@A00283:59:HC7CWD$XX:1:1101:8250:1031 1:N:0
CNTTCTGATGCTTGAATGCTTGCATACAGTTTGGCGTATTCTGATCGATCCTGATGCCATTGACTTGTGTTCAAGACTTAGCATCGACAGTACTATATTGAGGGATTAGCGAATCTGGCCACCGTCTGAA
+
F#FFFFFFFFFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FF:FF:FF:FF:FFFF:FFF
```

Quality scores for sequence data

Phred Quality Score	Probability of incorrect base call	Base call accuracy
20	1 in 100	99%
30	1 in 1000	99.9%
40	1 in 10,000	99.99%
50	1 in 100,000	99.999%

FastQ format: quality scores

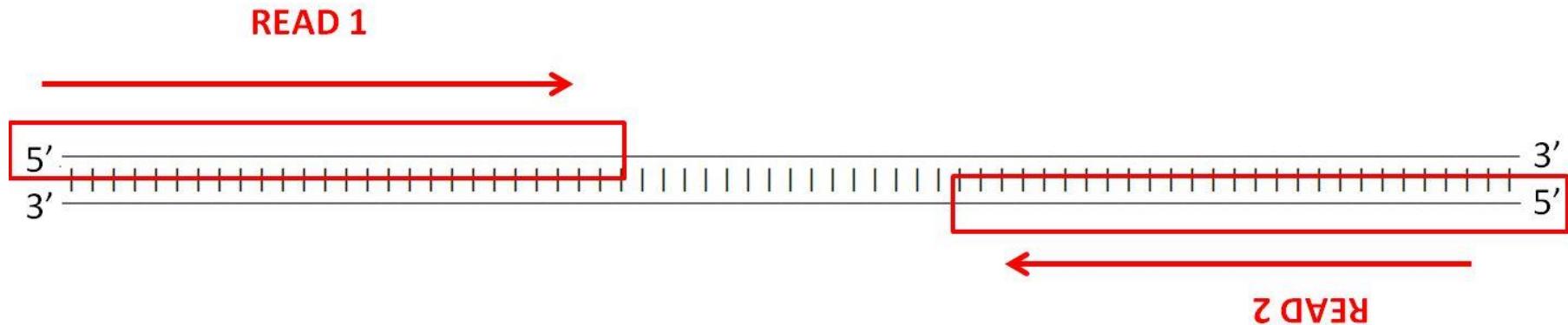


Data format: FastQ

```
@DRR000615.149 HWUSI-EAS505:1:1:15:14 length=51
GTAAGGGCACACCGTTCTCAAGGCCANNNNNNTNNNNNNNTNNNNNN
+DRR000615.149 HWUSI-EAS505:1:1:15:14 length=51
95?A/3@C@CC@A+ABCCBCCCC@#####!!!!#!!!!!!#!!!!!!
@DRR000615.1395 HWUSI-EAS505:1:1:123:13 length=51
CGACGACTGCCGTGAGCGTGTAGTCAGTCCGNNNNNNNNNNNNNNNNNN
+DRR000615.1395 HWUSI-EAS505:1:1:123:13 length=51
@B(92. (==9>2@=@#####!!!!#!!!!!!#!!!!!!#!!!!!!
@DRR000615.3018 HWUSI-EAS505:1:1:221:15 length=51
TAGGAACACTTCTCATTTATTCCCTGCCTATCNNNNNNNNNNNNNNNN
+DRR000615.3018 HWUSI-EAS505:1:1:22
7AA5->=9@75CA3>BB6BB;BA9;;5=#####!
sequence data
@DRR000615.3021 HWUSI-EAS505:1:1:221:13 length=51
GTAAAAGTACATCCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
+DRR000615.3021 HWUSI-EAS505:1:1:221:13 length=51
:,B*?4?3;BBB@A@@@BCCBCB#####!!!!!!#!!!!!!#!!!!!!
```

Paired sequence reads

(i.e. we sequence from both ends of the same molecule)



Data format: Paired FastQ

FastQ file 1

```
@ERR030887.1 HWI-BRUNOP16X_0001:8:1:7336:1073#0/1
TNTCGATTACATGTGGATCAGGTTGATTAATAATGGCGATAGGGNNCT
+
5#145555555A;A8445555555>>>.=@#####
@ERR030887.2 HWI-BRUNOP16X_0001:8:1:10288:1073#0/1
TNAGTCTTCCCAGCCTAACAAAGAAAGCAAGAATAATTGGCACNNNGA
+
5#156+43&4(0*55CFDAF#####
@ERR030887.3 HWI-BRUNOP16X_0001:8:1:13787:1073#0/1
ANGTTGCTATTCCCGGGCGTAAACCAAGACGTTCTGGCGTCTGTATGGACACTGATCNNNGA
+
5#55555554GGGG?FFFFFFGGGGEGGG
1 read data HWI-BRUNOP16X_0001:8:1:15389:1074#0/1
+
5#5555255555445EGGGGGGGGA@;>A>A<A>A#####

```

read name
("/1" means forward read.)

FastQ file 2

```
@ERR030887.1 HWI-BRUNOP16X_0001:8:1:7336:1073#0/2
ATGAANCTNTNNNGNAANNTNNNANGNGNNNNNNNCTTTNCANN
+
#####
@ERR030887.2 HWI-BRUNOP16X_0001:8:1:10288:1073#0/2
CAAAANTTNANNNGNNTNNNANNNCAGNTNNNNNNNNCTAGNTGNN
+
#####
@ERR030887.3 HWI-BRUNOP16X_0001:8:1:13787:1073#0/2
GGGTGNTANAGNNNTNAANNCNNNCNTNTN
+
#####
@ERR030887.4 HWI-BRUNOP16X_0001:8:1:15389:1074#0/2
CCCAAGACGTTCTGGCGTCTGTATGGACACTGATCNNNGA
+
#####

```

read name
("/2" means reverse read.)

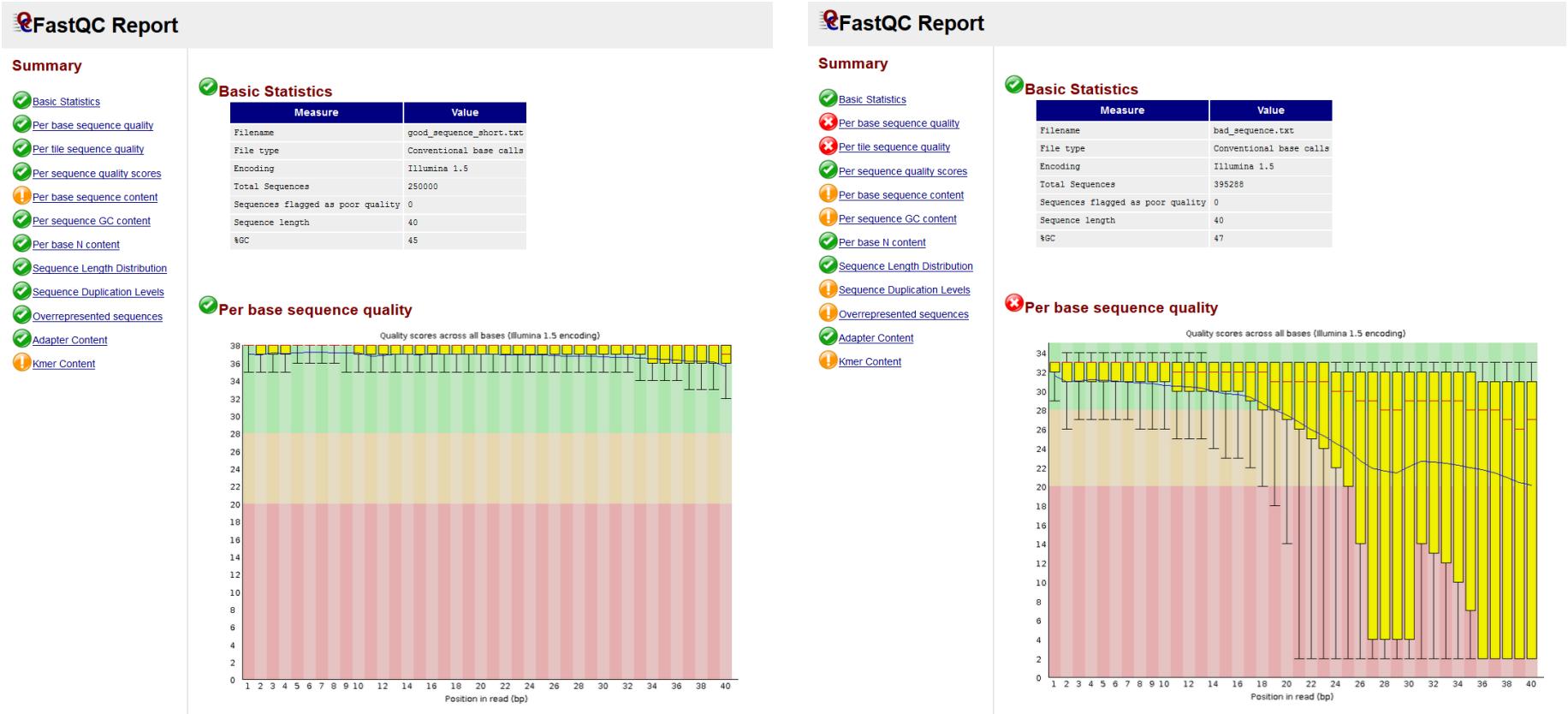
Character	ASCII	FastQ quality score (ASCII – 33)
!	33	0
"	34	1
#	35	2
\$	36	3
%	37	4
&	38	5
'	39	6
(40	7
)	41	8
*	42	9
+	43	10
,	44	11

ASCII TABLE

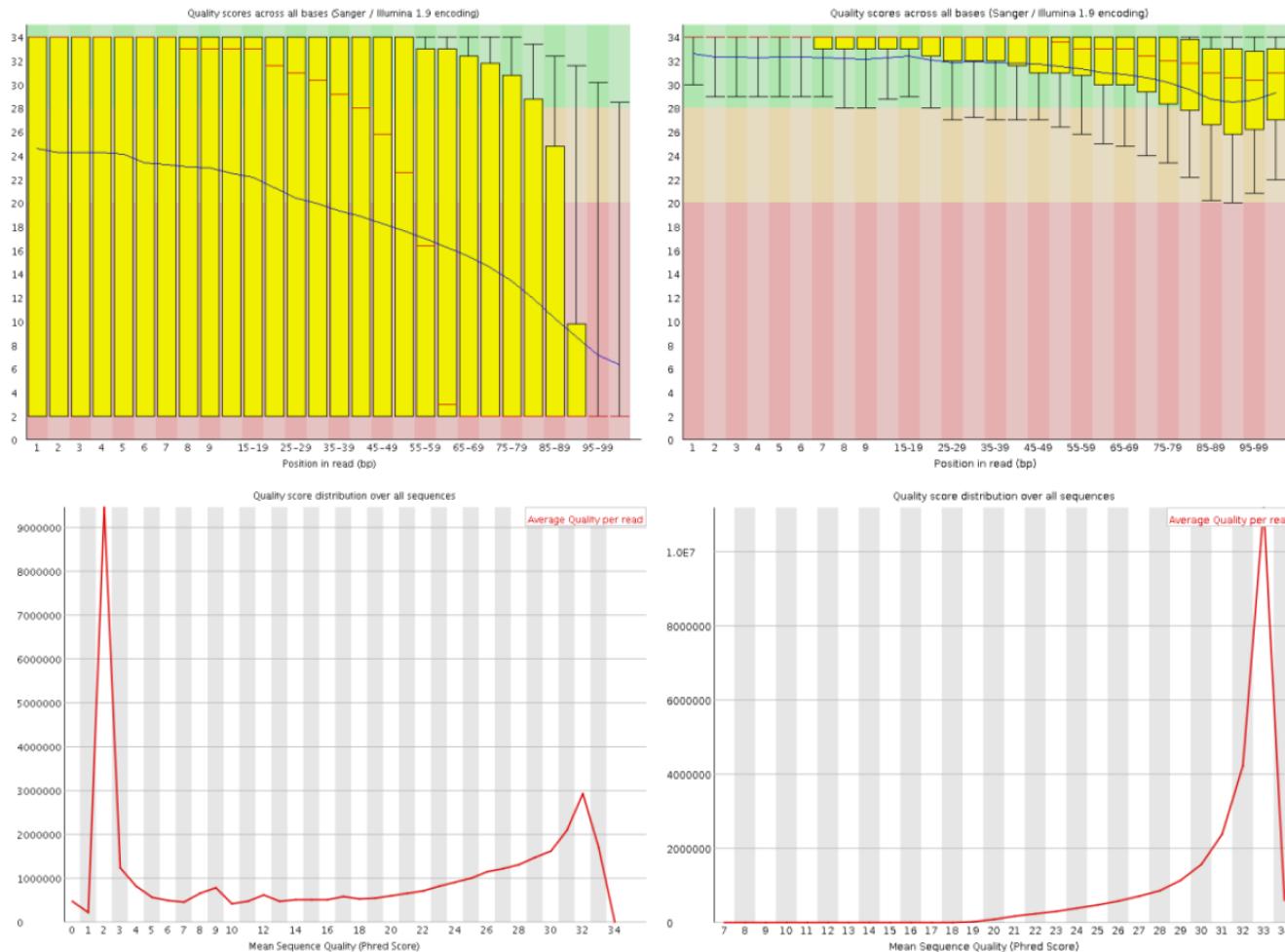
Decimal	Hex	Char	Decimal	Hex	Char	Decimal	Hex	Char	Decimal	Hex	Char
0	0	[NULL]	32	20	[SPACE]	64	40	@	96	60	`
1	1	[START OF HEADING]	33	21	!	65	41	A	97	61	a
2	2	[START OF TEXT]	34	22	"	66	42	B	98	62	b
3	3	[END OF TEXT]	35	23	#	67	43	C	99	63	c
4	4	[END OF TRANSMISSION]	36	24	\$	68	44	D	100	64	d
5	5	[ENQUIRY]	37	25	%	69	45	E	101	65	e
6	6	[ACKNOWLEDGE]	38	26	&	70	46	F	102	66	f
7	7	[BELL]	39	27	'	71	47	G	103	67	g
8	8	[BACKSPACE]	40	28	(72	48	H	104	68	h
9	9	[HORIZONTAL TAB]	41	29)	73	49	I	105	69	i
10	A	[LINE FEED]	42	2A	*	74	4A	J	106	6A	j
11	B	[VERTICAL TAB]	43	2B	+	75	4B	K	107	6B	k
12	C	[FORM FEED]	44	2C	,	76	4C	L	108	6C	l
13	D	[CARRIAGE RETURN]	45	2D	.	77	4D	M	109	6D	m
14	E	[SHIFT OUT]	46	2E	.	78	4E	N	110	6E	n
15	F	[SHIFT IN]	47	2F	/	79	4F	O	111	6F	o
16	10	[DATA LINK ESCAPE]	48	30	0	80	50	P	112	70	p
17	11	[DEVICE CONTROL 1]	49	31	1	81	51	Q	113	71	q
18	12	[DEVICE CONTROL 2]	50	32	2	82	52	R	114	72	r
19	13	[DEVICE CONTROL 3]	51	33	3	83	53	S	115	73	s
20	14	[DEVICE CONTROL 4]	52	34	4	84	54	T	116	74	t
21	15	[NEGATIVE ACKNOWLEDGE]	53	35	5	85	55	U	117	75	u
22	16	[SYNCHRONOUS IDLE]	54	36	6	86	56	V	118	76	v
23	17	[END OF TRANS. BLOCK]	55	37	7	87	57	W	119	77	w
24	18	[CANCEL]	56	38	8	88	58	X	120	78	x
25	19	[END OF MEDIUM]	57	39	9	89	59	Y	121	79	y
26	1A	[SUBSTITUTE]	58	3A	:	90	5A	Z	122	7A	z
27	1B	[ESCAPE]	59	3B	;	91	5B	[123	7B	{
28	1C	[FILE SEPARATOR]	60	3C	<	92	5C	\	124	7C	
29	1D	[GROUP SEPARATOR]	61	3D	=	93	5D]	125	7D	}
30	1E	[RECORD SEPARATOR]	62	3E	>	94	5E	^	126	7E	~
31	1F	[UNIT SEPARATOR]	63	3F	?	95	5F	-	127	7F	[DEL]

Good versus bad data

- http://www.bioinformatics.babraham.ac.uk/projects/fastqc/good_sequence_short_fastqc.html
- http://www.bioinformatics.babraham.ac.uk/projects/fastqc/bad_sequence_fastqc.html



Before and after trimming

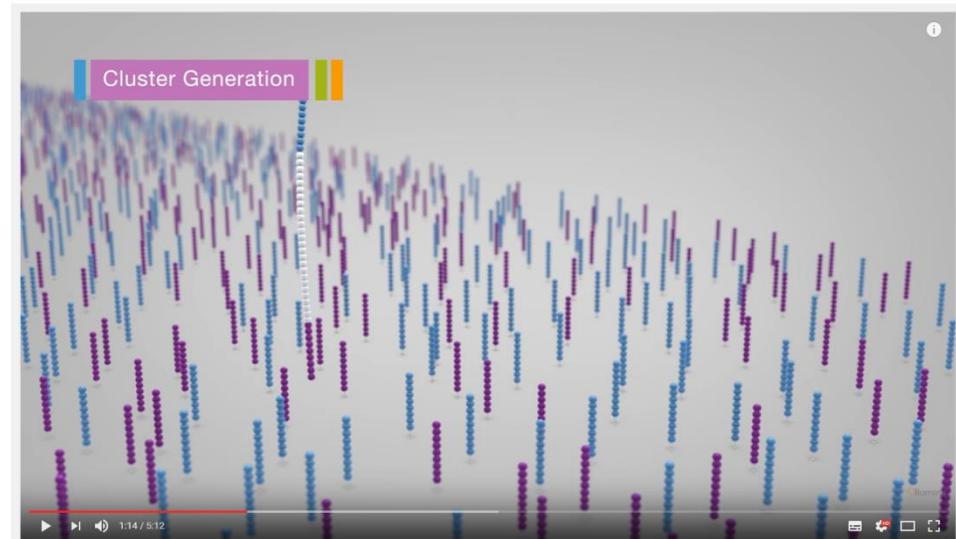
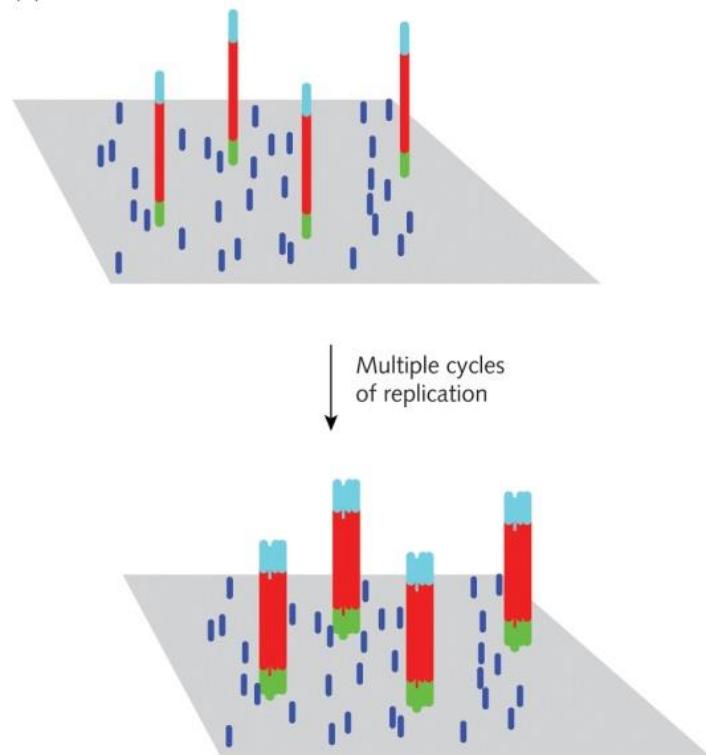


Challenges of second-generation methods: short sequence reads

```
@A00283:59:HC7CWD$XX:1:1101:1904:1031 1:N:0
TNAAGTAATGCAGAACTGGCTGACCCGAGGAACATAAAAAGCTACCGACTTCTCAGTTAGCAGGCTACTACCGCAAGTCATGAAAAACTATGAAAGATCAGTGCACCCCTATTCTTGCTGAAAAAAGATTTCCAATG
+
F#XXXXXXXXXXXXXXXXXXXXXXXXXXXX:XXXXXXXXXXXXXXXXXXXXXXXXXXXX:XXXXXXXXXXXXXXXXXXXXXXXXXXXX:XXXXXXXXXXXXXXXXXXXXXXXXXXXX:FF:FFFF:FFFFFF
@A00283:59:HC7CWD$XX:1:1101:2157:1031 1:N:0
GNCTTTACTAATGGGTCAATTCTCCCTCACAGGATTGCCAGTCATTCTGTGCTACTGAGTCCAACCTAATAGTTAGCAAAAGTCTGAGTTCTGGTTCAATTCTGTCTTTGATTGTTCTTACTAACCT
+
F#XXXXXXXXXXXXXXXXXXXXXXXXXXXX:XXXXXXXXXXXXXXXXXXXXXXXXXXXX:XXXXXXXXXXXXXXXXXXXXXXXXXXXX:XXXXXXXXXXXXXXXXXXXXXXXXXXXX:XXXXXXXXXXXXXXXXXXXXXXXXXXXX:FF:FFFF:FFFFFF
@A00283:59:HC7CWD$XX:1:1101:5195:1031 1:N:0
CNTAGAGAGCATTGTCTGCTGGACTGTCAAGGACCAAATCTGAATCTCGTCAGAACTATAAATATTGGAAATACTATGATCAGTTGCTATACAAAGTAGCTATAAACTCTGTCTTAAGTTGCTACATGCACGCACATAAA
+
F#XXXXXXXXXXXXXXXXXXXXXXXXXXXX:XXXXXXXXXXXXXXXXXXXXXXXXXXXX:XXXXXXXXXXXXXXXXXXXXXXXXXXXX:XXXXXXXXXXXXXXXXXXXXXXXXXXXX:XXXXXXXXXXXXXXXXXXXXXXXXXXXX:XXXXXXXXXXXXXXXXXXXXXXXXXXXX:FF:FFFF:FFFFFF
@A00283:59:HC7CWD$XX:1:1101:5321:1031 1:N:0
ANTAGAAAAGAAACTATACGATGGAAGAAGTTAGTAACCTTGACGAGTATAGTAACAATATCAAAGTTCTGAATCAACAATTATGATAGGATAGCAACAGCTTGTCAAGAACAGATTCTGGAATGATGGACTTTCTATA
+
F#XXXXXXXXXXXXXXXXXXXXXXXXXXXX:XXXXXXXXXXXXXXXXXXXXXXXXXXXX:XXXXXXXXXXXXXXXXXXXXXXXXXXXX:XXXXXXXXXXXXXXXXXXXXXXXXXXXX:XXXXXXXXXXXXXXXXXXXXXXXXXXXX:XXXXXXXXXXXXXXXXXXXXXXXXXXXX:FF:FFFF:FFFFFF
@A00283:59:HC7CWD$XX:1:1101:5990:1031 1:N:0
ANGACTTGAAGGCATCCAATGCATCATTCTGTTAGAAGCAAATATATATATATACATATACGTGAGTAATCATCTATAAGGAGATGAAGTATTCTGACCAGTATGTCATGTCGGACAACATATACGTAT
+
F#XXXXXXXXXXXXXXXXXXXXXXXXXXXX:XXXXXXXXXXXXXXXXXXXXXXXXXXXX:XXXXXXXXXXXXXXXXXXXXXXXXXXXX:XXXXXXXXXXXXXXXXXXXXXXXXXXXX:XXXXXXXXXXXXXXXXXXXXXXXXXXXX:FF:FFF:FFFF:FF
@A00283:59:HC7CWD$XX:1:1101:7437:1031 1:N:0
CNTAGAAATCTCTACATTATAGCCCCATGAGATAAATTGCAACCCCTACAAGATAGTAAACACTGGTCAATCACGAAGCAGGCAAGAGAGTGAATTGACCATAAACCTCTTCTTCTAGATAAAAGCTATTAAACTGCCA
+
F#XXXXXXXXXXXXXXXXXXXXXXXXXXXX:XXXXXXXXXXXXXXXXXXXXXXXXXXXX:XXXXXXXXXXXXXXXXXXXXXXXXXXXX:XXXXXXXXXXXXXXXXXXXXXXXXXXXX:XXXXXXXXXXXXXXXXXXXXXXXXXXXX:FF:FF:FFFF:FFF
@A00283:59:HC7CWD$XX:1:1105:31638:35744 1:N:0
TTGCTTGCTTAATCAGATTGAACCAACATGGAACATCTCATCAAAGAGCTCCATCATTCTGCAATGTCATTGAGGATTGATTGTTCTCGAAACGACAAGTCAGACTGGAAACAGGTTCTGCTGCAAATTCTAAATCATTG
+
FFFFFFFFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFF
@A00283:59:HC7CWD$XX:1:1101:8250:1031 1:N:0
CNTTCTGCATGCTGAAATGCTTCATACAGTTTTGGCGTGATTCTGTATATCGATCCTGATGCCATTGACTTGTGTTCAAGACTTAGCATCCGACAGCTACTATATTGAGGGATTAGCGAATCTGCCACCGTCTGAA
+
F#XXXXXXXXXXXXXXXXXXXXXXXXXXXX:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF:FFFF
```

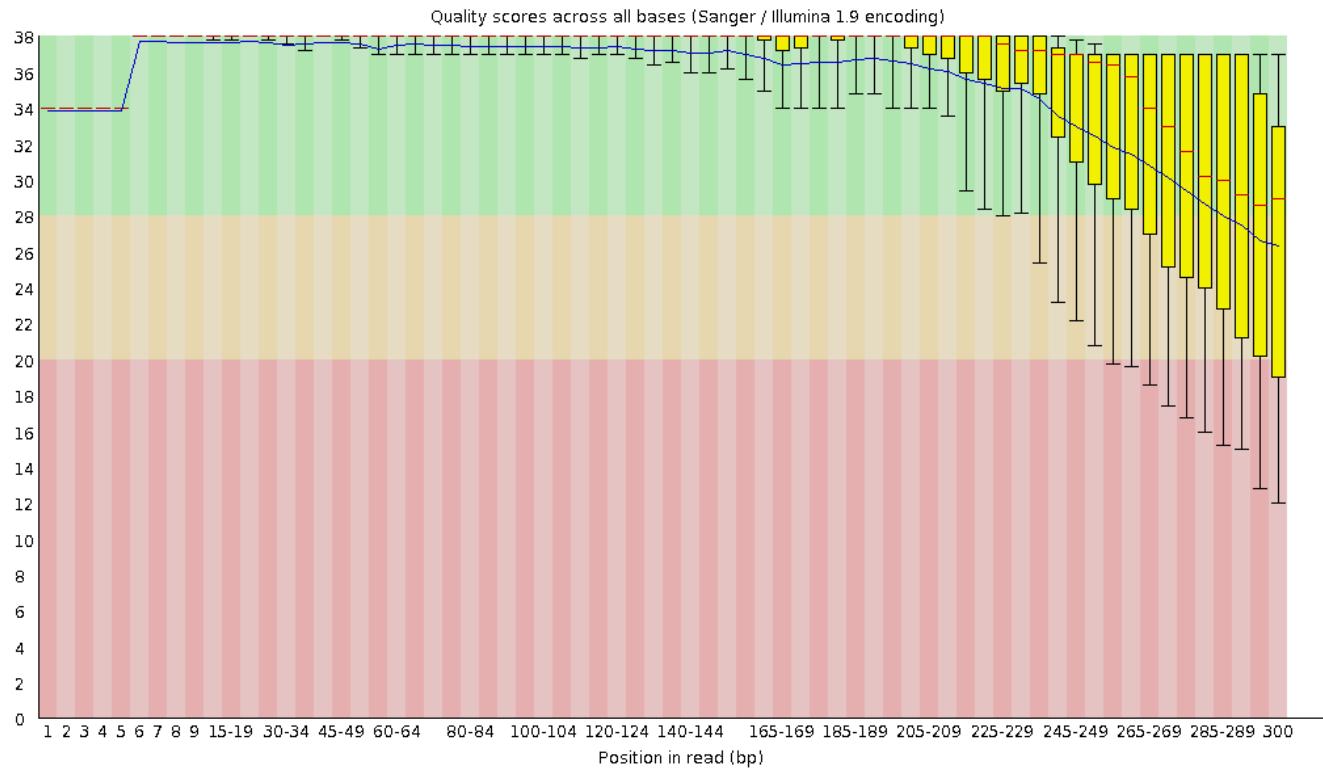
Shortness of reads: makes some analyses difficult

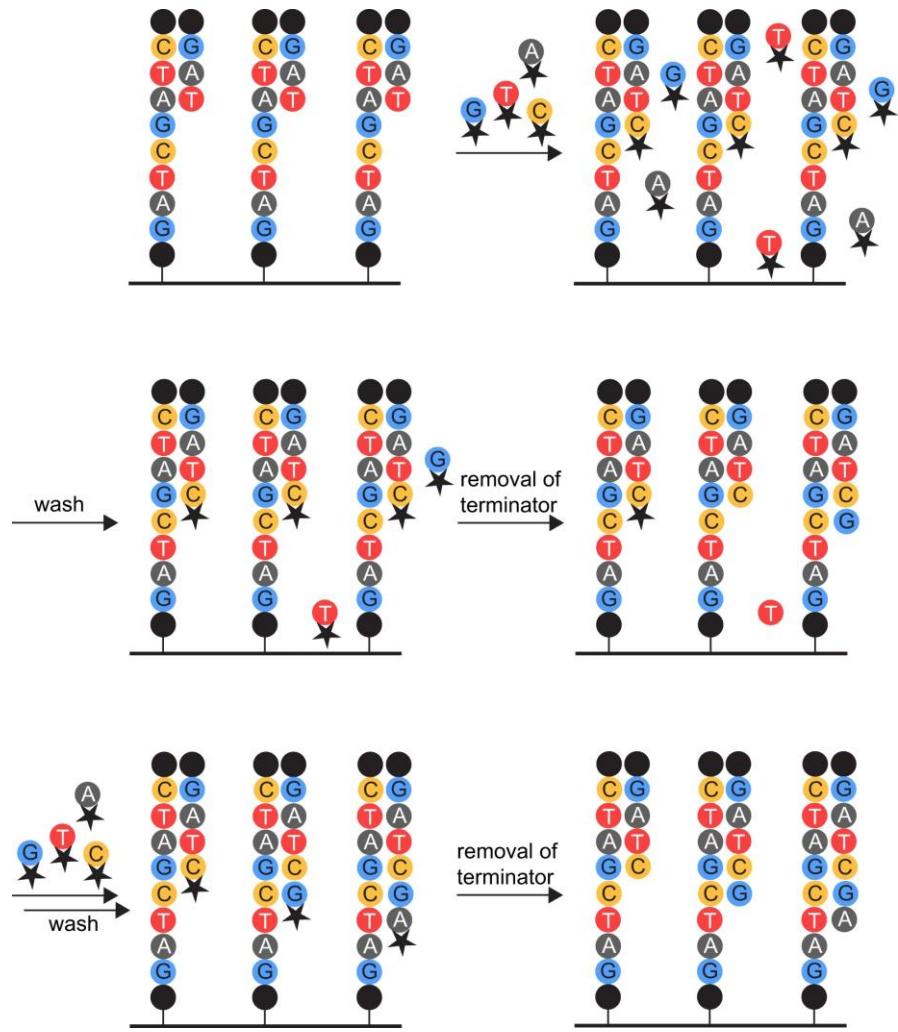
Challenges of NGS: DNA amplification



Amplification step: introduces bias and errors

Why does sequencing quality deteriorate along the Illumina read?





From: <https://doi.org/10.1038/s41598-018-29325-6>

Challenges posed by second-generation (Illumina) sequencing

- The output from the Illumina sequencing is an enormous text-based file containing sequence reads.
- Each sequence read is the result of one sequencing reaction, *i.e.* the data for one of the template DNA molecules in the sample.
- The **sequence reads are short**: typically 100– 300 nucleotides long.
- Each sequence read originates from a cluster of molecules that has undergone a **PCR-like amplification**; this amplification step can introduce **biases and errors** into the data.
- The **large number of sequences** presents bioinformatics challenges for actually using the data.



SINGLE-MOLECULE SEQUENCING (THIRD-GENERATION SEQUENCING)

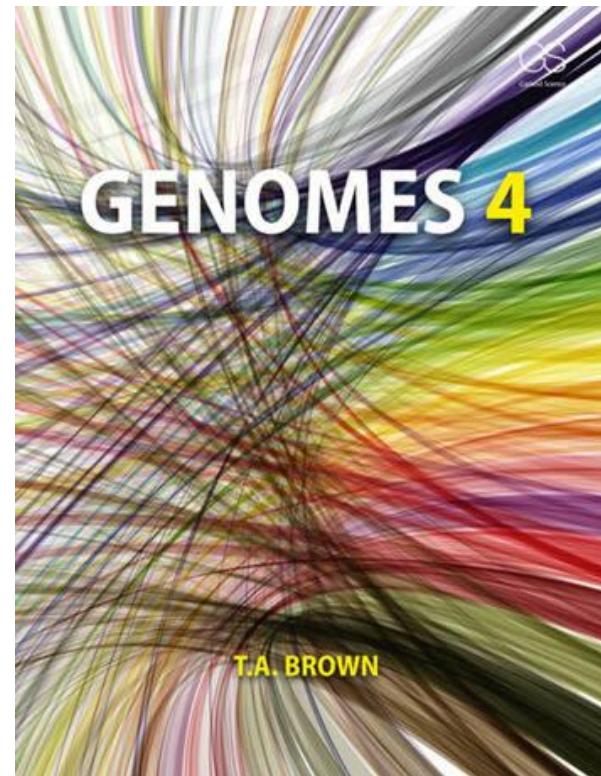
Solution I: nanopore technology

See Figure 4.16 of Genomes 4

By T. A. Brown

ISBN 9780815345084

Published June 21, 2017 by Garland Science



Solution I: Nanopore technology

The screenshot shows the Oxford Nanopore Technologies website. At the top, there's a navigation bar with links to PRODUCTS, HOW IT WORKS, APPLICATIONS, and GET STARTED. Below the navigation, there's a button labeled "Learn how nanopore sensing can scale".

MinION flow cell: An image of a rectangular grey device with a small screen and several ports.

PromethION flow cell: An image of a larger grey device with a circular inset showing a close-up of its internal components.

SmidgION: An image of a small, white, smartphone-compatible device.

MinION: A detailed description of the MinION system:

- Smartphone-compatible.
- Designed to be our smallest sequencing device so far
- Same nanopore sensing technology as MinION and PromethION
- In development

[Learn more about SmidgION](#)

MinION: A detailed description of the MinION system:

- Pocket-sized
- Up to 512 nanopore channels
- Simple 10-minute sample prep available
- Many publications illustrate broad usage
- Commercially available

[Learn more](#)

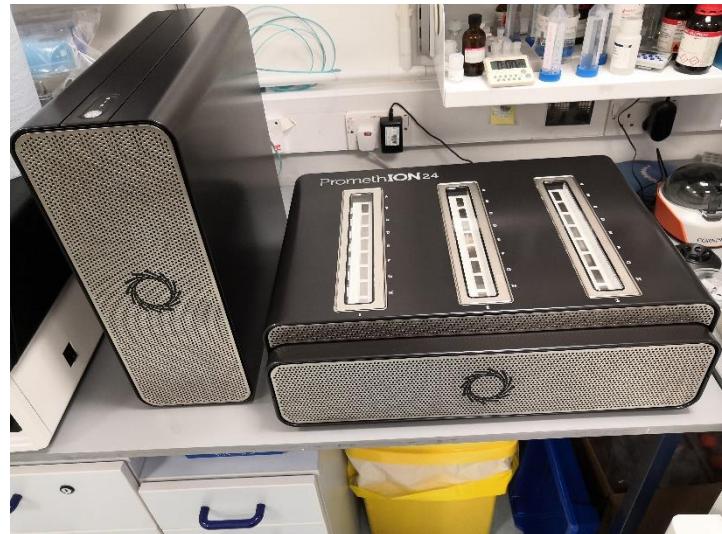
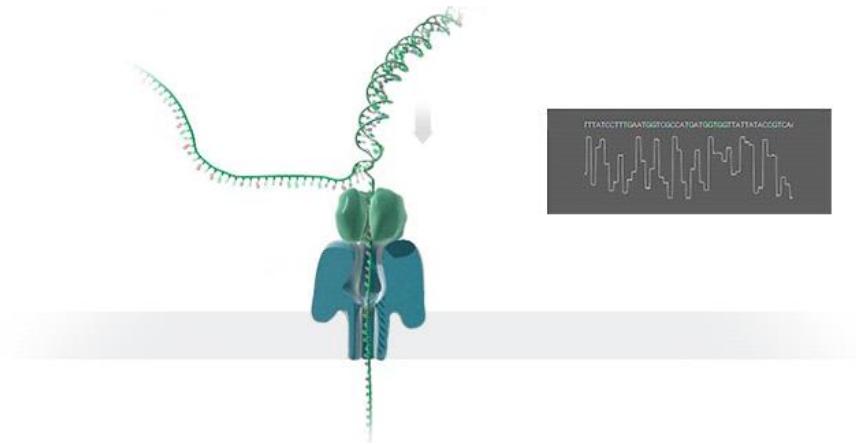
[Buy MinION](#)

PromethION: A detailed description of the PromethION system:

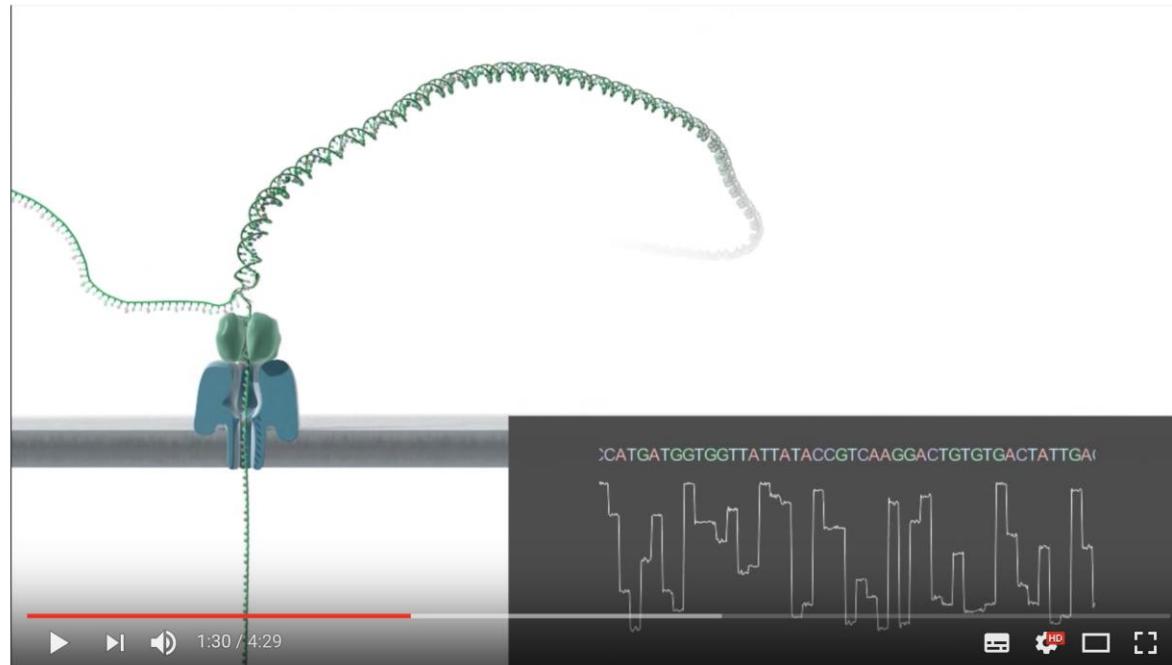
- Benchtop system
- Up to 48 flow cells of 3,000 nanopore channels each
- Simple 10-minute sample prep available
- Flow cells can be used together or separately, multiple samples may be applied to each flow cell

[Learn more](#)

[Register for early access](#)

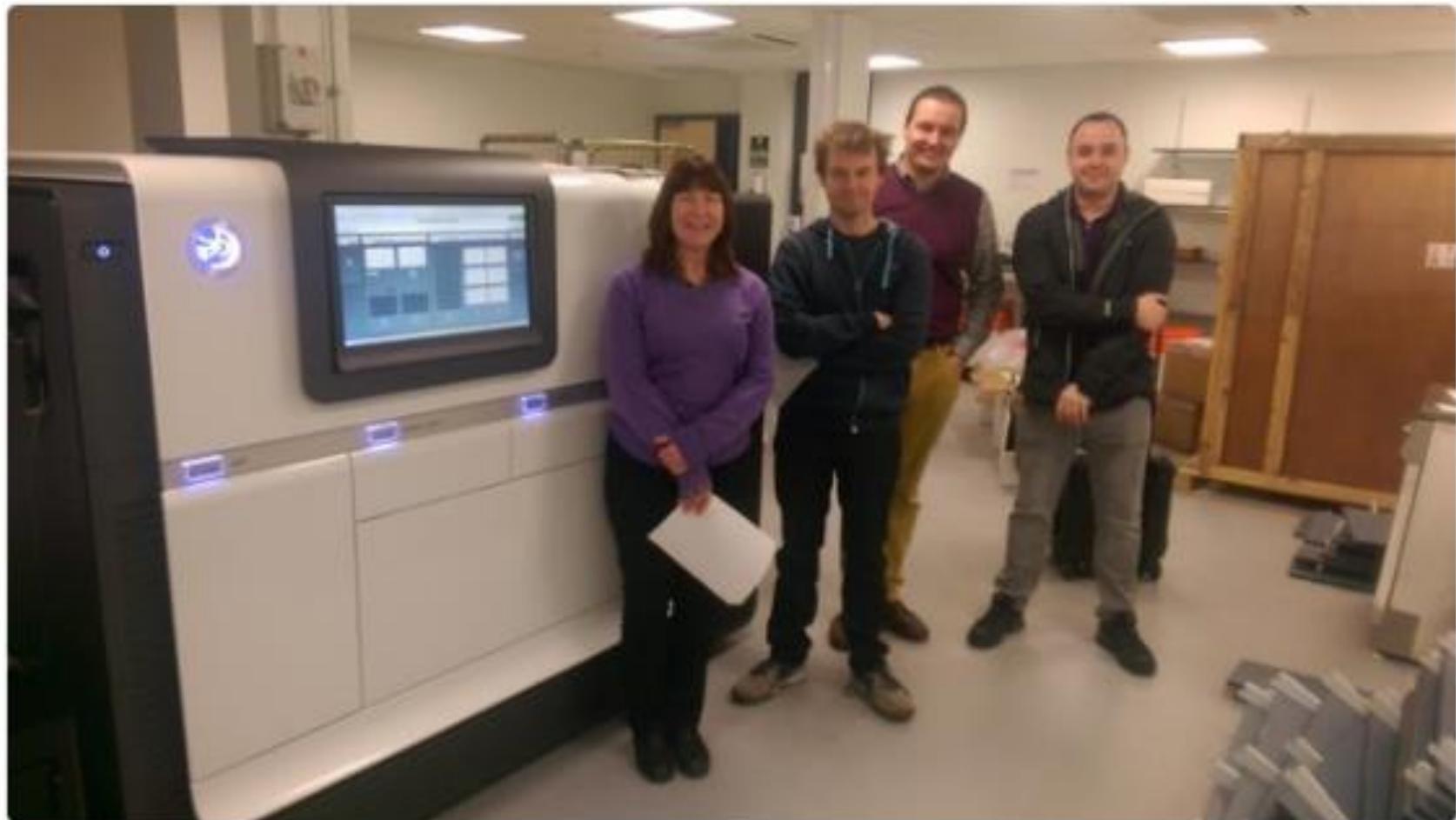


How does nanopore sequencing work?



<https://youtu.be/CE4dW64x3Ts>

Solution II: PacBio SMRT



How does PacBio SMRT sequencing work?



<https://youtu.be/v8p4ph2MAvI>

How does PacBio SMRT sequencing work?



Introduction to SMRT Sequencing

114,876 views • 6 Dec 2011

241 12 SHARE SAVE ...



PacBio

1.79K subscribers

SUBSCRIBE

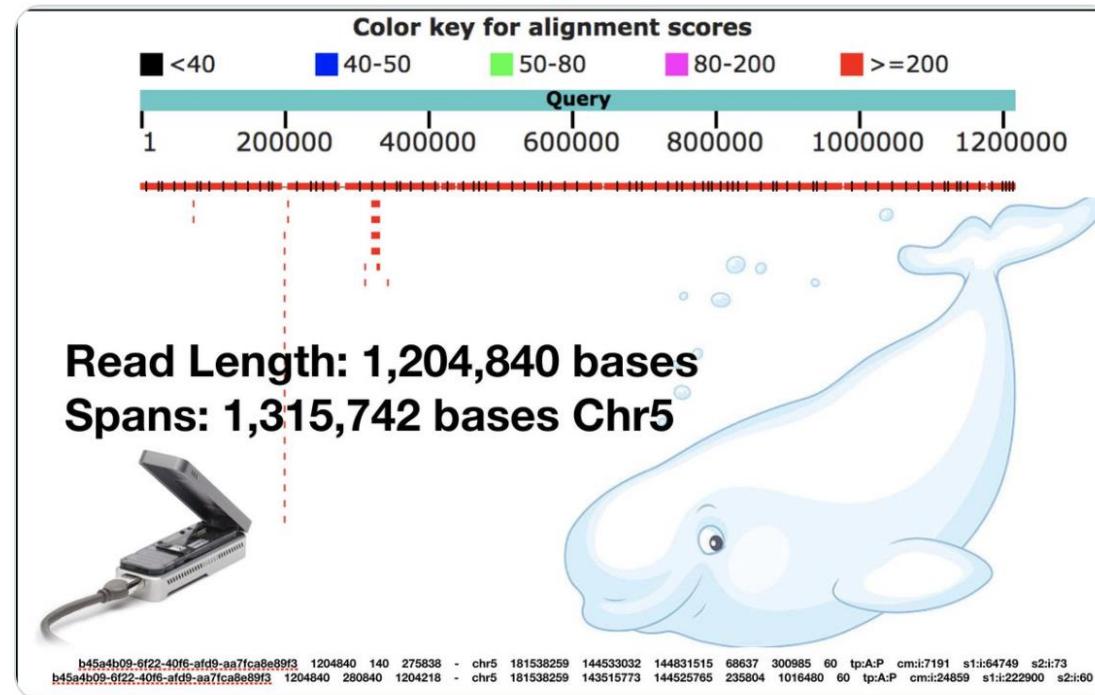
A brief animated introduction to Pacific Biosciences' Single Molecule, Real-Time (SMRT) Sequencing, including the SMRT Cell and zero mode waveguide (ZMW).

<https://youtu.be/NHCJ8PtYCFc>



Matt Loose
@mattloose

Amazing what you can see on an evenings whale watching with [@nanopore](#) - I saw a big one this tonight! Thanks to Nadine, [@alexomics](#) and of course [@scalene](#). Protocol tweaks will be available shortly on lab.loman.net/protocols

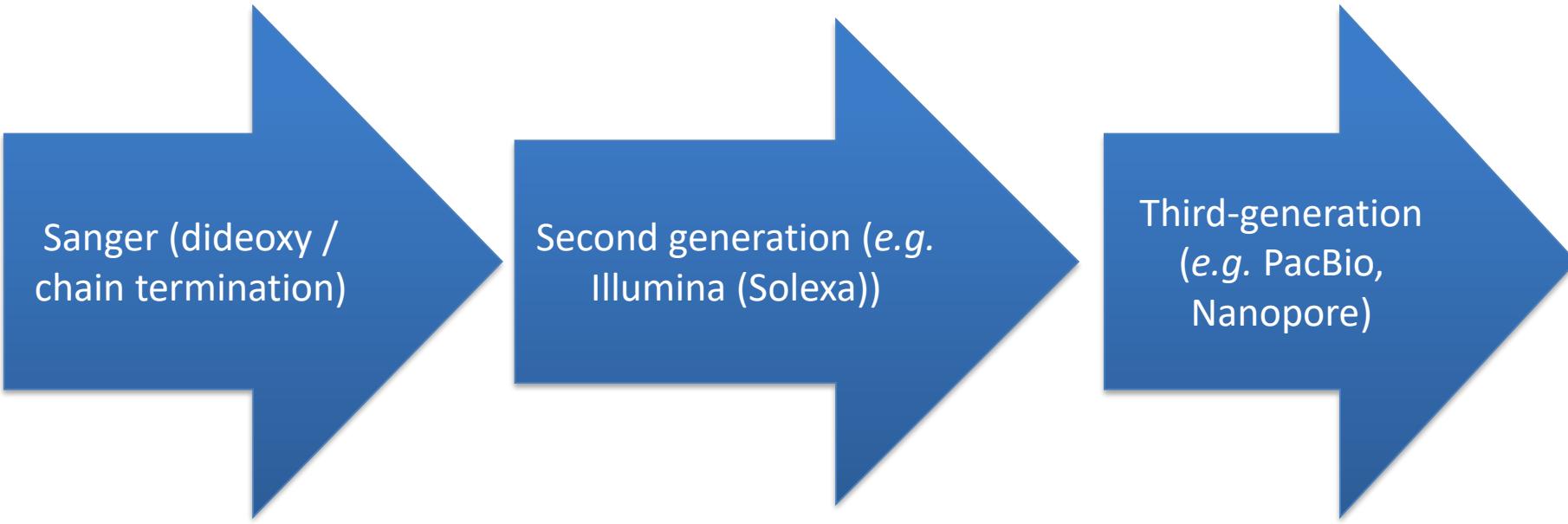


Summary / comparison

Method	Generation	Read length (bp)	Single pass error rate (%)	No. of reads per run	Time per run	Cost per million bases (US\$)
Sanger ABI 3730×1	First	600–1000	0.001	96	0.5–3 h	500
Ion Torrent	Second	200	1	8.2×10^7	2–4 h	0.1
454 (Roche) GS FLX+	Second	700	1	1×10^6	23 h	8.57
Illumina HiSeq 2500 (High Output)	Second	2 × 125	0.1	8×10^9 (paired)	7–60 h	0.03
Illumina HiSeq 2500 (Rapid Run)	Second	2 × 250	0.1	1.2×10^9 (paired)	1–6 days	0.04
SOLID 5500xl	Second	2 × 60	5	8×10^8	6 days	0.11
PacBio RS II: P6-C4	Third	1.0–1.5 × 10 ⁴ on average	13	$3.5\text{--}7.5 \times 10^4$	0.5–4 h	0.40–0.80
Oxford Nanopore MinION	Third	2–5 × 10 ³ on average	38	$1.1\text{--}4.7 \times 10^4$	50 h	6.44–17.90

From: Rhoads, A. and Au, K.F. (2015). PacBio Sequencing and Its Applications. *Genomics, Proteomics & Bioinformatics*, **13**, 278–289. (This article gives citations of sources of data.)

Summary: sequencing technologies

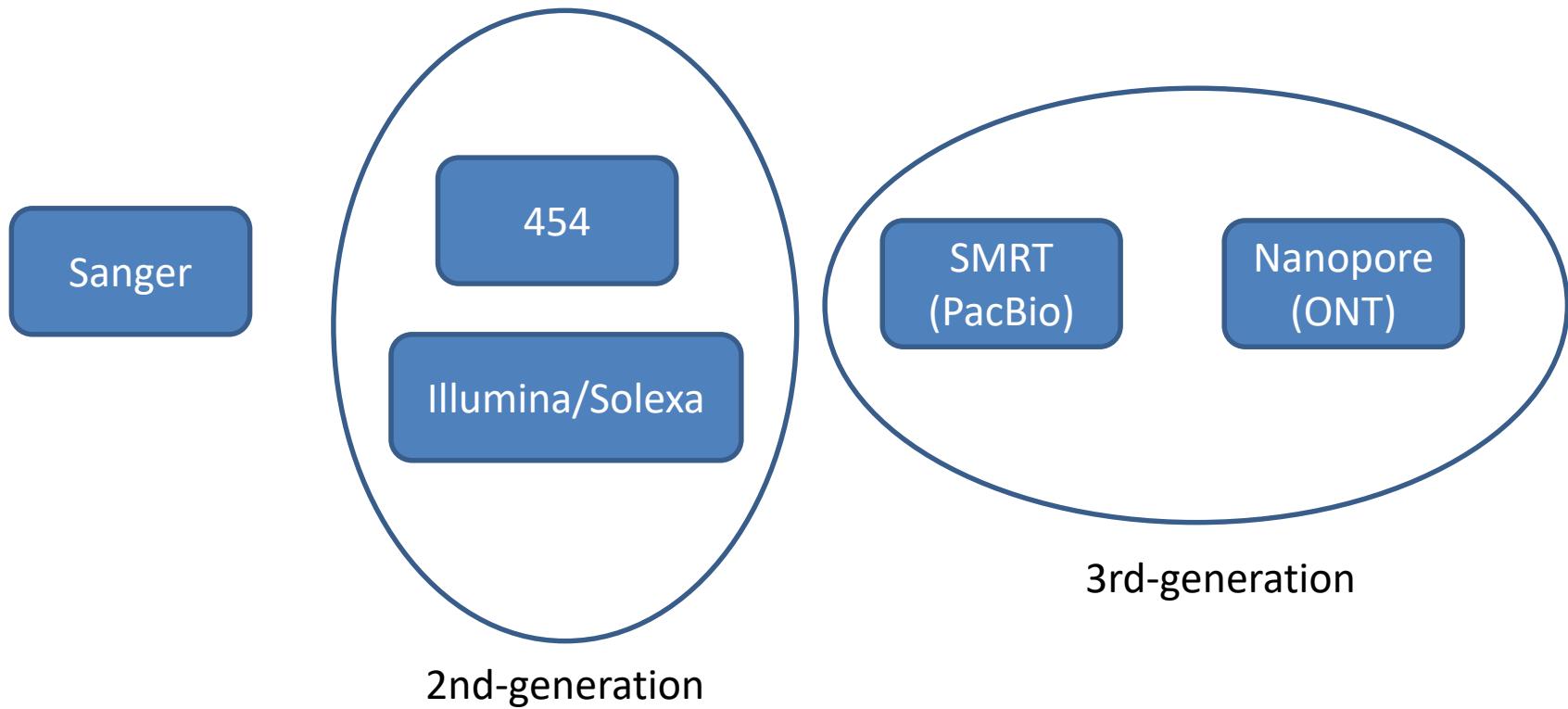


Sanger (dideoxy / chain termination)

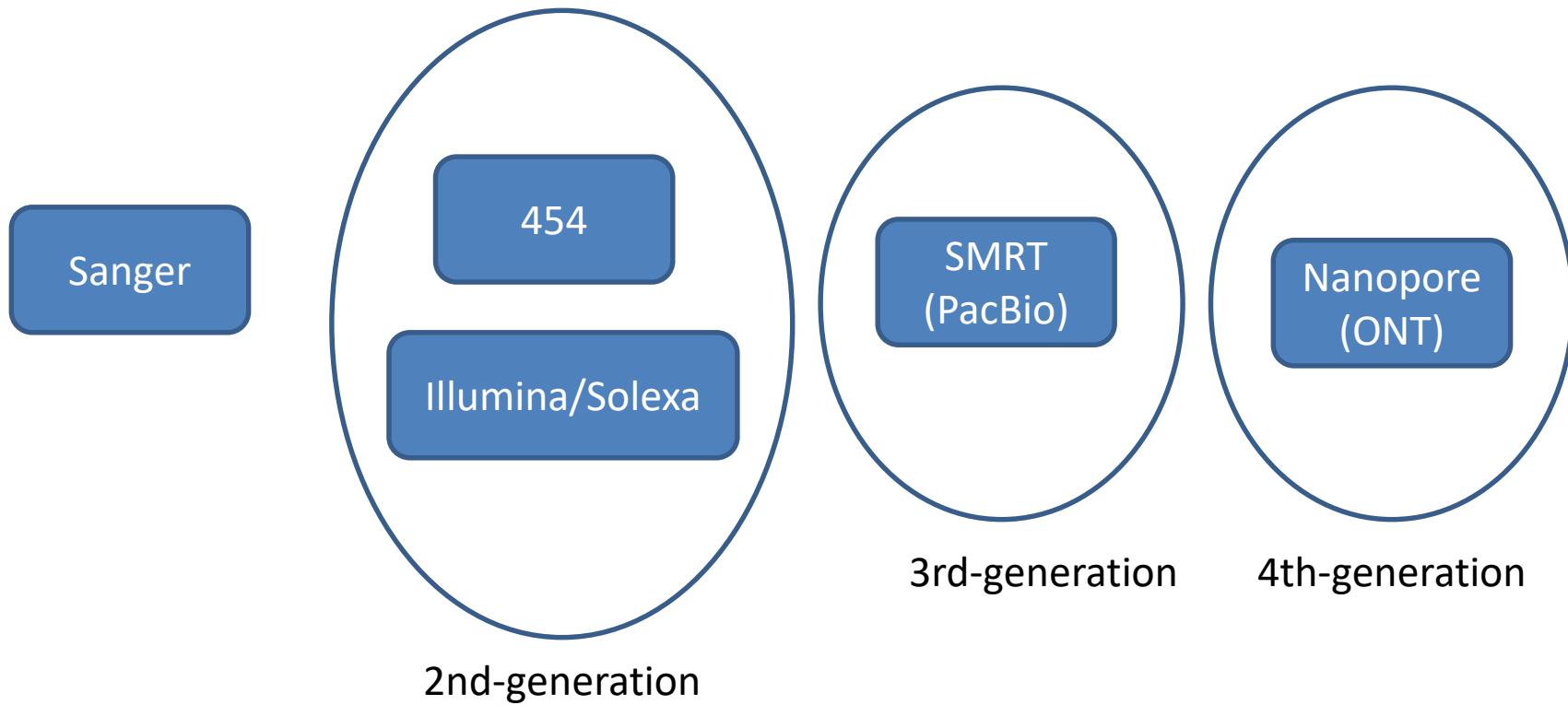
Second generation (*e.g.* Illumina (Solexa))

Third-generation (*e.g.* PacBio, Nanopore)

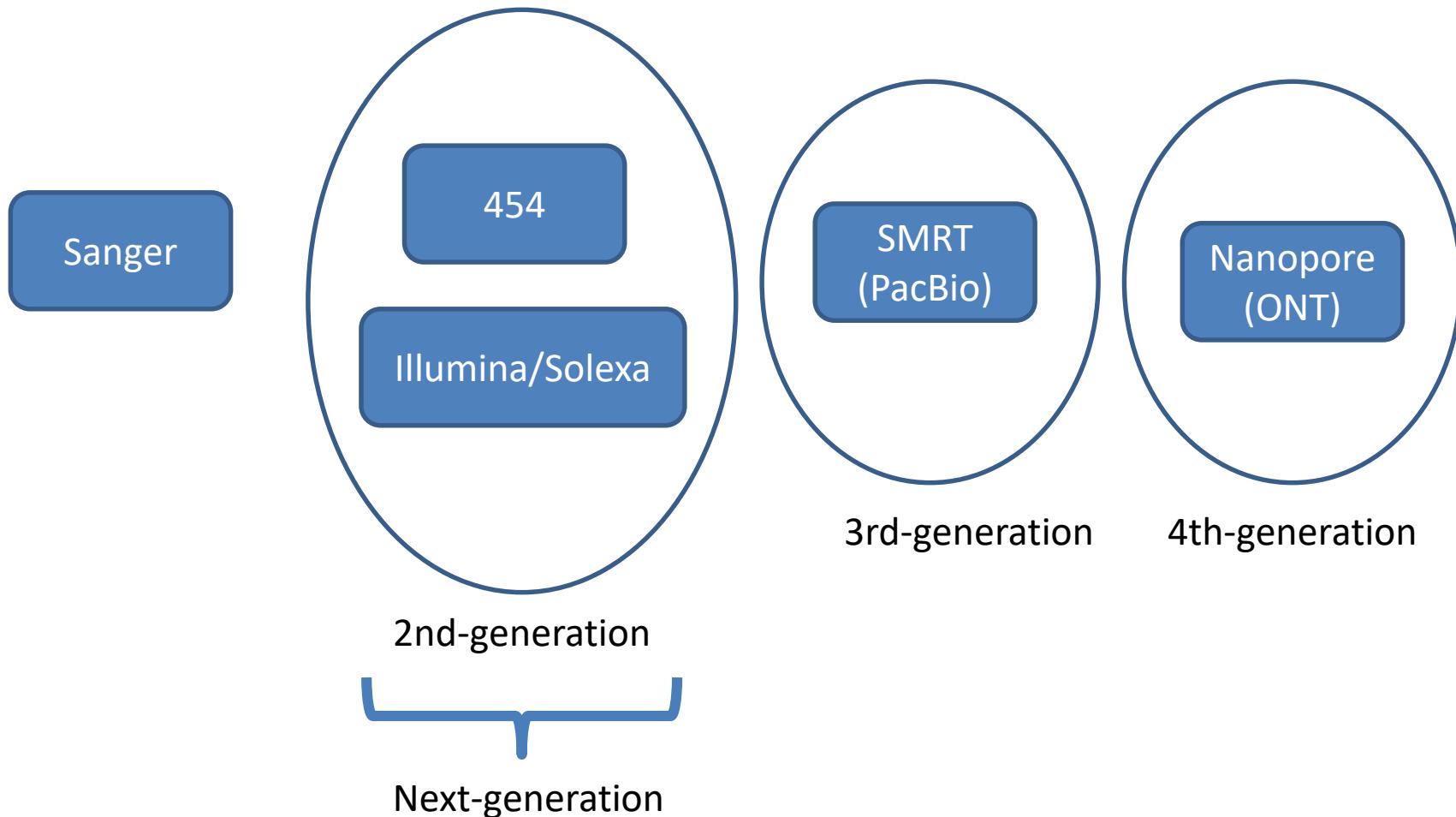
Ambiguous terminology I



Ambiguous terminology II



Ambiguous terminology III



Ambiguous terminology IV

