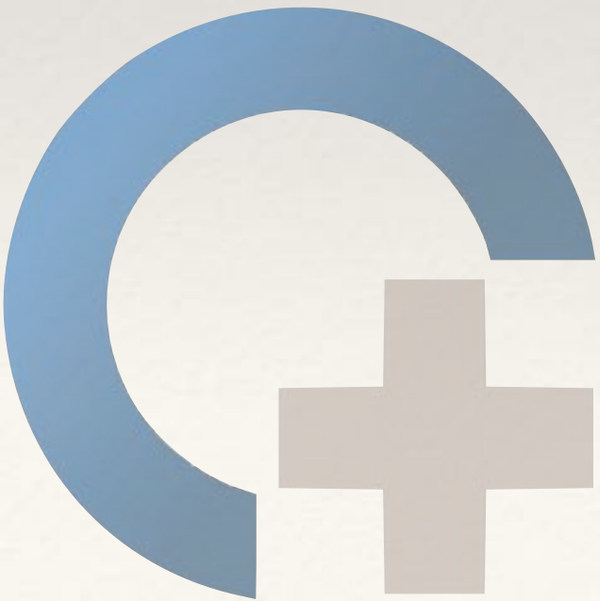
UiO IN-BIOS5000/9000

Illumina Technology

Arvind Sundaram
Oct 28, 2024
Norwegian Sequencing Centre
OUS, Ullevål, Oslo



High Throughput Sequencing

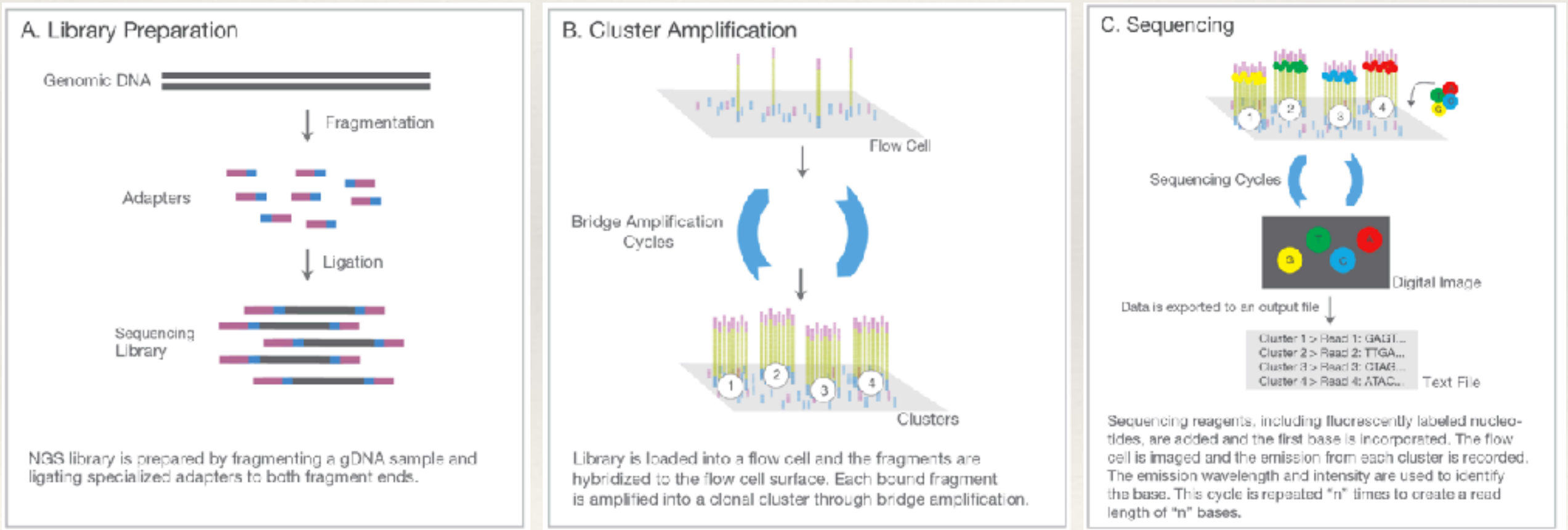
- First generation - past, present
  - Up to 1 kb; high quality data; multiplexed
  - SANGER; Highly automated (ABI Sanger 3730xl)
- Second generation - present
  - Shorter reads; Massive parallelisation and real high throughput
  - Illumina, BGISeq, Ion-torrent, [454, Solid]
  - RNA is reverse-transcribed to cDNA before sequencing
- Third generation - [present] future
  - Long-read sequencing; Single-molecule sequencing (without amplification)
  - PacBio, Oxford Nanopore, [more in development]
  - Potential to sequence RNA directly

Illumina sequencers

- Second generation sequencing technique
- Sequencing-by-synthesis aka SBS
  - https://www.youtube.com/watch?v=fCd6B5HRaZ8
- Mass parallelisation and real high throughput

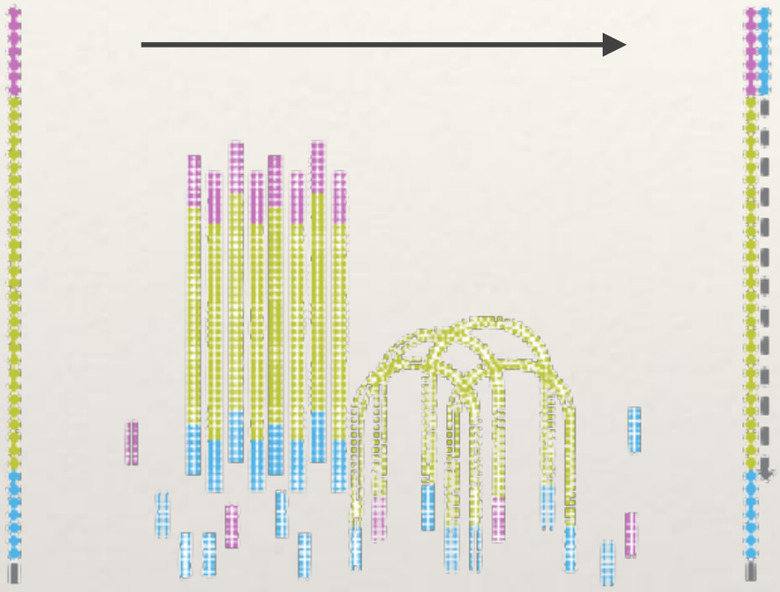
Library prep and sequencing

Fragment (DNA) sequenced: up to 800 bp



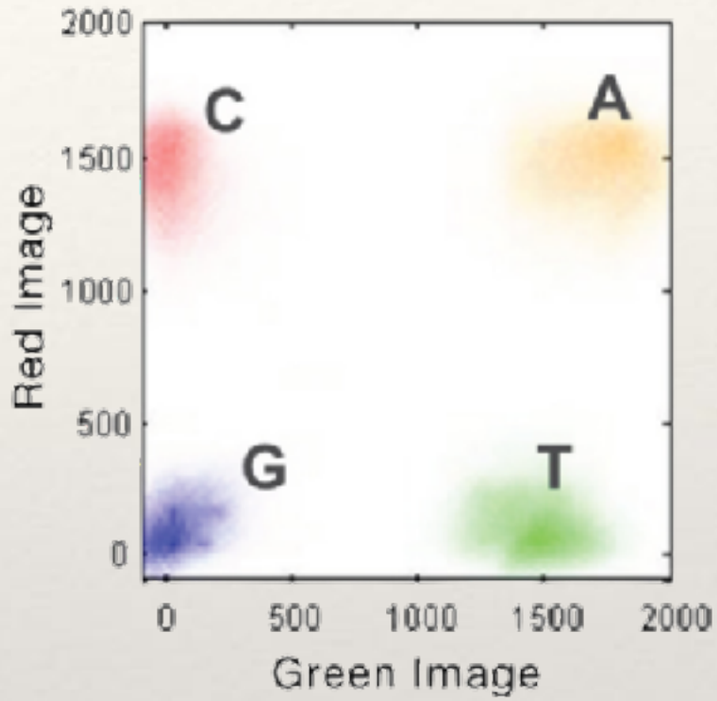
https://www.illumina.com/documents/products/illumina_sequencing_introduction.pdf

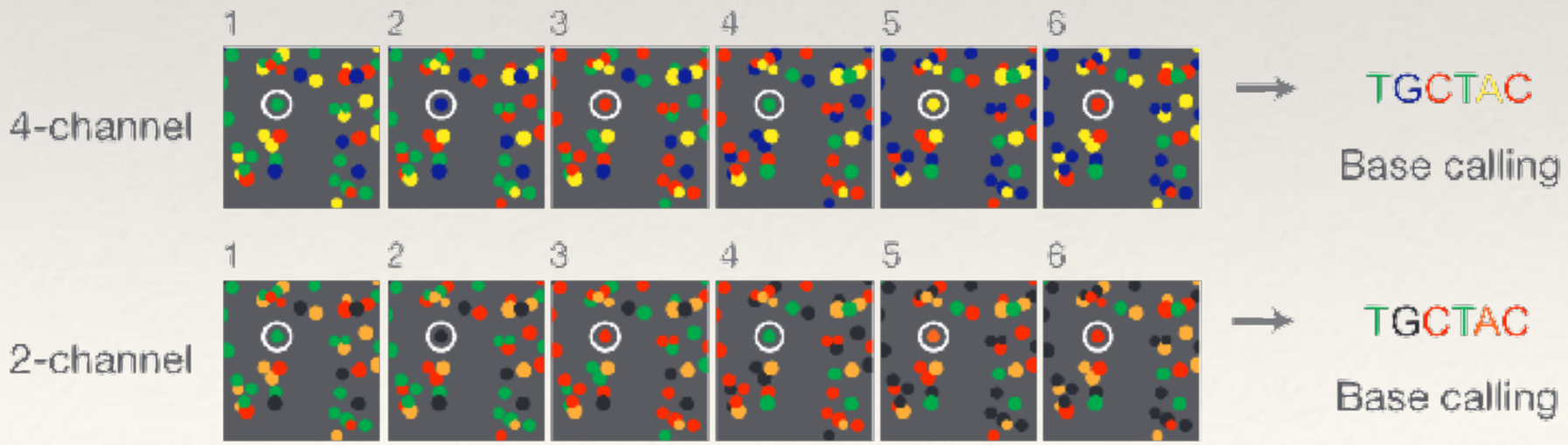
Sequencing by synthesis (SBS)



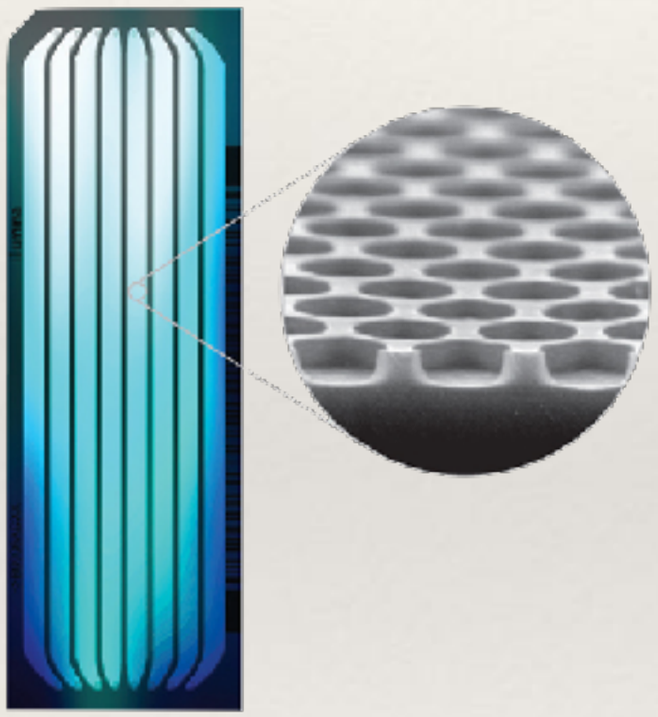
Old sequencers use 4 colors
Newer machines use 2 colors
iSeq uses 1 color

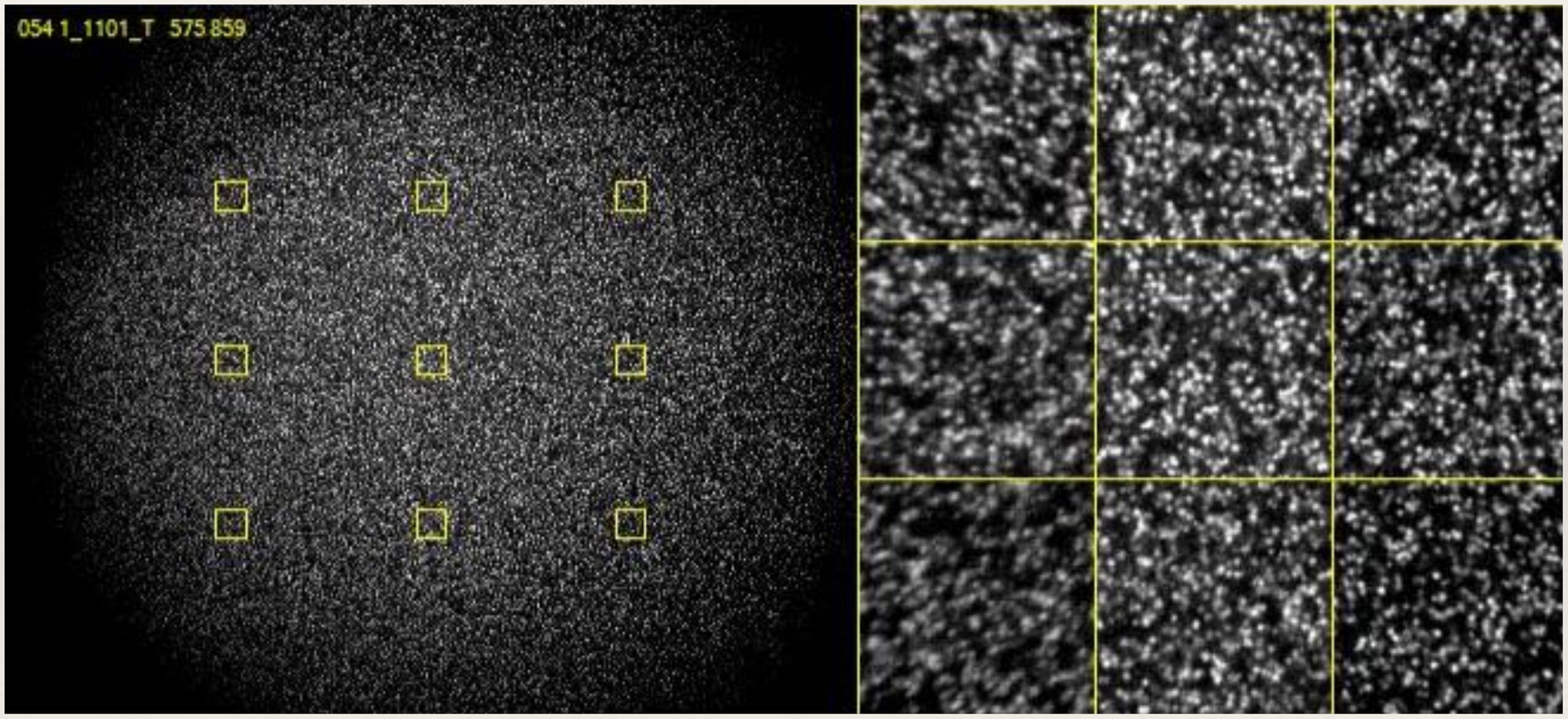
XLEAP-SBS (2022)
- Increased speed
- Greater fidelity
- Greater accuracy
- More robustness





Sequencing

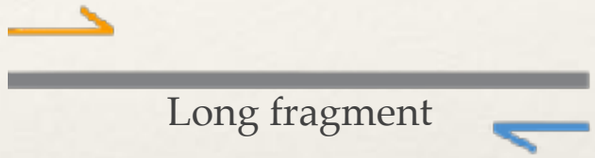




Library prep and sequencing

Fragment (DNA) sequenced: up to 800 bp

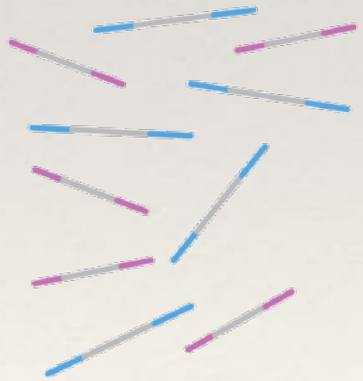
Read type: Mate pair



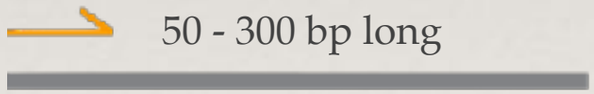
Add adapters during library preparation

Multiplexing: single/dual index

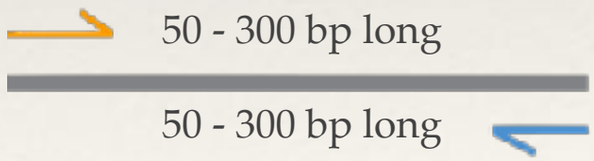
Multiplexing: pooling



Read type: Single end



Read type: Paired end



https://www.illumina.com/documents/products/illumina_sequencing_introduction.pdf

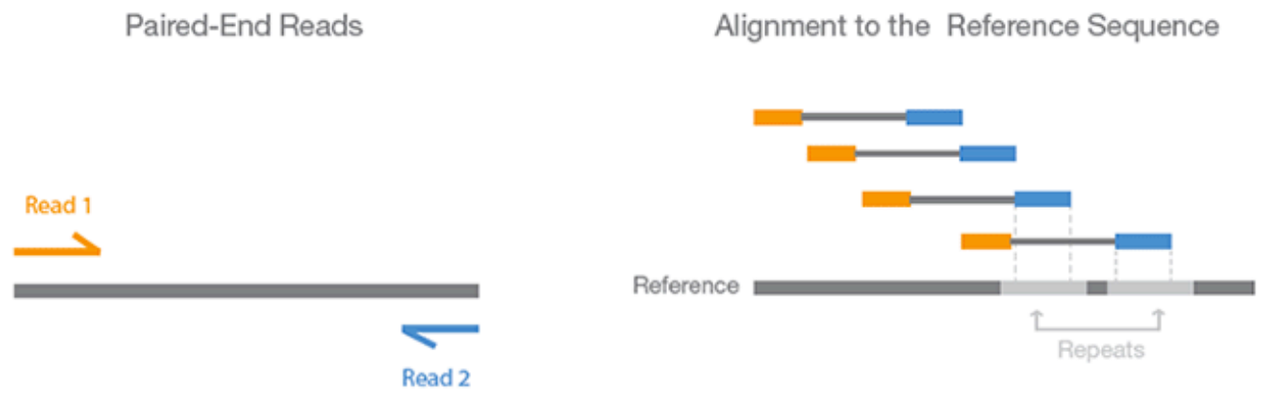
Single/Paired-end/Mate pair sequencing

- Paired-end reads or mate pair reads are pairs of reads known to come from two close regions in the genome.
- They are located with an approximate fixed distance from each other.
- Typically paired ends are a ~100-500bp apart, while mate pairs are ~2-10kb apart
- Allows short reads to have a larger ”effective” size
- Performed by sequencing fragments from both ends
- Often used with Illumina reads
  - Typically 2 x 150 bp separated by 300bp
- May also overlap (e.g. 2x250bp from 400bp fragments)

Paired-end sequencing

- Both ends of fragments will be sequenced
- Gives information on genomic distance between pairs of reads
- May be used to overcome some problems with short reads

Figure 4. Paired-End Sequencing and Alignment



Paired-end sequencing enables both ends of the DNA fragment to be sequenced. Because the distance between each paired read is known, alignment algorithms can use this information to map the reads over repetitive regions more precisely. This results in much better alignment of the reads, especially across difficult-to-sequence, repetitive regions of the genome.

Adpters and multiplexing

I2 sequencing order is sequencer dependent

https://biosails.github.io/pheniqs/tokenization

DRAGEN on-board/external

DRAGEN

- BCL Convert
- DRAGEN ORA Compression
- DRAGEN FASTQC + MultiQC
- Whole Genome
- Enrichment (including Exome)
- DRAGEN Amplicon
- RNA
- Single Cell RNA
- NanoString GeoMx NGS
- Methylation
- Metagenomics
- RNA Pathogen Detection
- COVID
- TSO 500 Portfolio
- Imputation
- ScATAC-Seq
- PGx Star Allele Caller
- Illumina Complete Long Reads
- DRAGEN secondary analysis for RIPIP and UPIP

FPGA based-architecture

30 minute* end-to-end analysis

Now open-source (2024)

Cloud based analysis

Stand-alone server

Integrated on-board

- NextSeq 1000/2000
- NovaSeq X
- MiSeq i100

https://doi.org/10.1038/s41587-024-02382-1

Illumina sequencers - benchtop

iSeq 100 System

MiniSeq System

MiSeq System

2024 XLEAP-SBS DRAGEN

MiSeq i100 and MiSeq i100 Plus Systems[a]

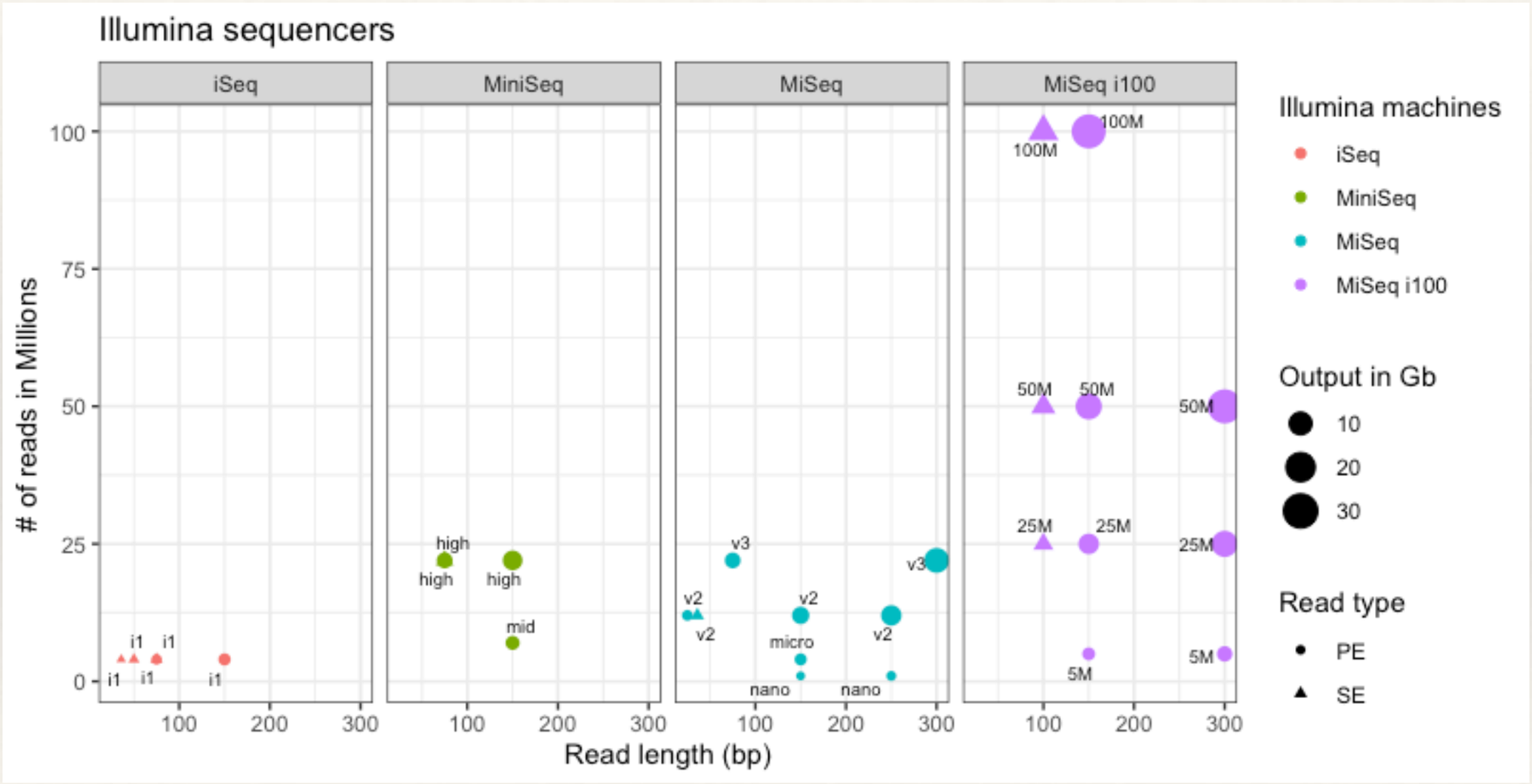
Flow cell	i1	Mid-output	Rapid	High-output	Nano	Micro	v2	v3	5M	25M	50M	100M
Output range	144 Mb–1.2 Gb	2.1–2.4 Gb	2 Gb	1.65–7.5 Gb	300–500 Mb	1.2 Gb	750 Mb–8.5 Gb	3.8–15 Gb	1.5–3 Gb	2.5–15 Gb	5–30 Gb	10–30 Gb
Single-end reads per run	4M	8M	20M	25M	1M	4M	15M	25M	5M	25M	50M	100M
Run time (hr)[b]	9–19	17	< 5	7–24	17–28	19	5.5–39	21–56	7–15	4–15	4–15	5–8
Maximum read length (bp)	2 × 150	2 × 150	1 × 100	2 × 150	2 × 250	2 × 150	2 × 250	2 × 300	2 × 300	2 × 300	2 × 300	2 × 150
Included data analysis	Local Run Manager	Local Run Manager			Local Run Manager				DRAGEN™ software			

a. The MiSeq i100 System supports the 5M and 25M flow cells only; the MiSeq i100 Plus System supports all four flow cells.

b. Listed run times are estimates.

https://www.illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/sequencing-platforms-brochure-m-gl-00451/sequencing-platforms-brochure-m-gl-00451.pdf

Data output


Illumina sequencers
iSeq
MiniSeq
MiSeq
MiSeq i100
# of reads in Millions
Read length (bp)
Illumina machines
iSeq
MiniSeq
MiSeq
MiSeq i100
Output in Gb
10
20
30
Read type
PE
SE

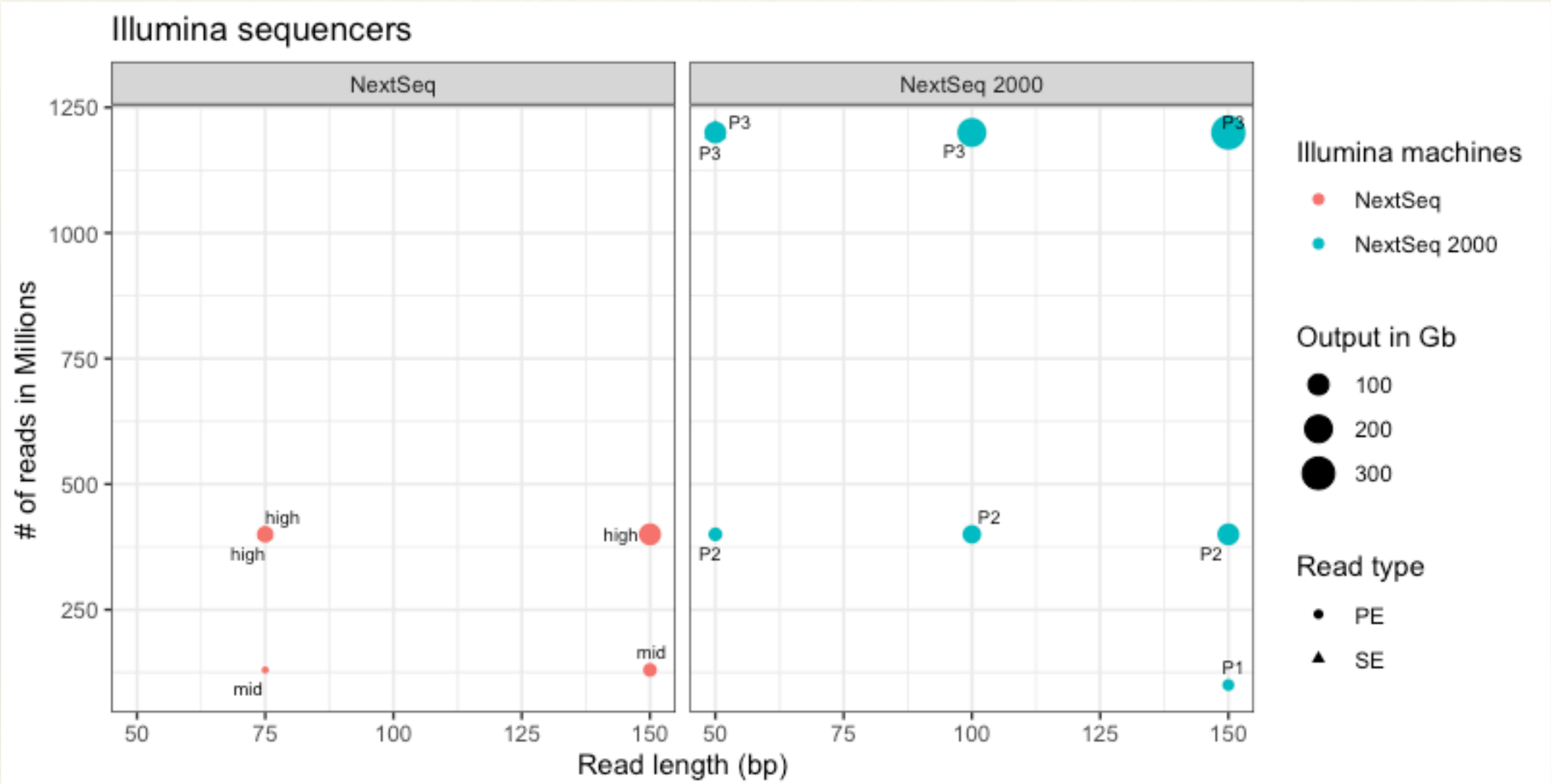
Illumina sequencers - slightly bigger

2020 XLEAP-SBS DRAGEN

	NextSeq 550 System[a]		NextSeq 1000 and NextSeq 2000 Systems			
Flow cell	Mid-output	High-output	P1[b]	P2[b]	P3[c]	P4[c]
Output range	16–39 Gb	25–120 Gb	10–60 Gb	40–240 Gb	120–360 Gb	90–540 Gb
Single-end reads per run	130M	400M	100M	400M	1.2B	1.8B
Run time (hr)	15–26	11–29	8–34	12–42	18–40	12–44
Maximum read length (bp)	2 × 150	2 × 150	2 × 300	2 × 300	2 × 150	2 × 150
Included data analysis	Local Run Manager		Onboard DRAGEN secondary analysis			

https://www.illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/sequencing-platforms-brochure-m-gl-00451/sequencing-platforms-brochure-m-gl-00451.pdf

Data output


Illumina sequencers
NextSeq
NextSeq 2000
# of reads in Millions
1250
1000
750
500
250
high
high
high
mid
mid
P3
P3
P3
P2
P2
P2
P1
50
75
100
125
150
Read length (bp)
Illumina machines
NextSeq
NextSeq 2000
Output in Gb
100
200
300
Read type
PE
SE

Illumina sequencers - production scale

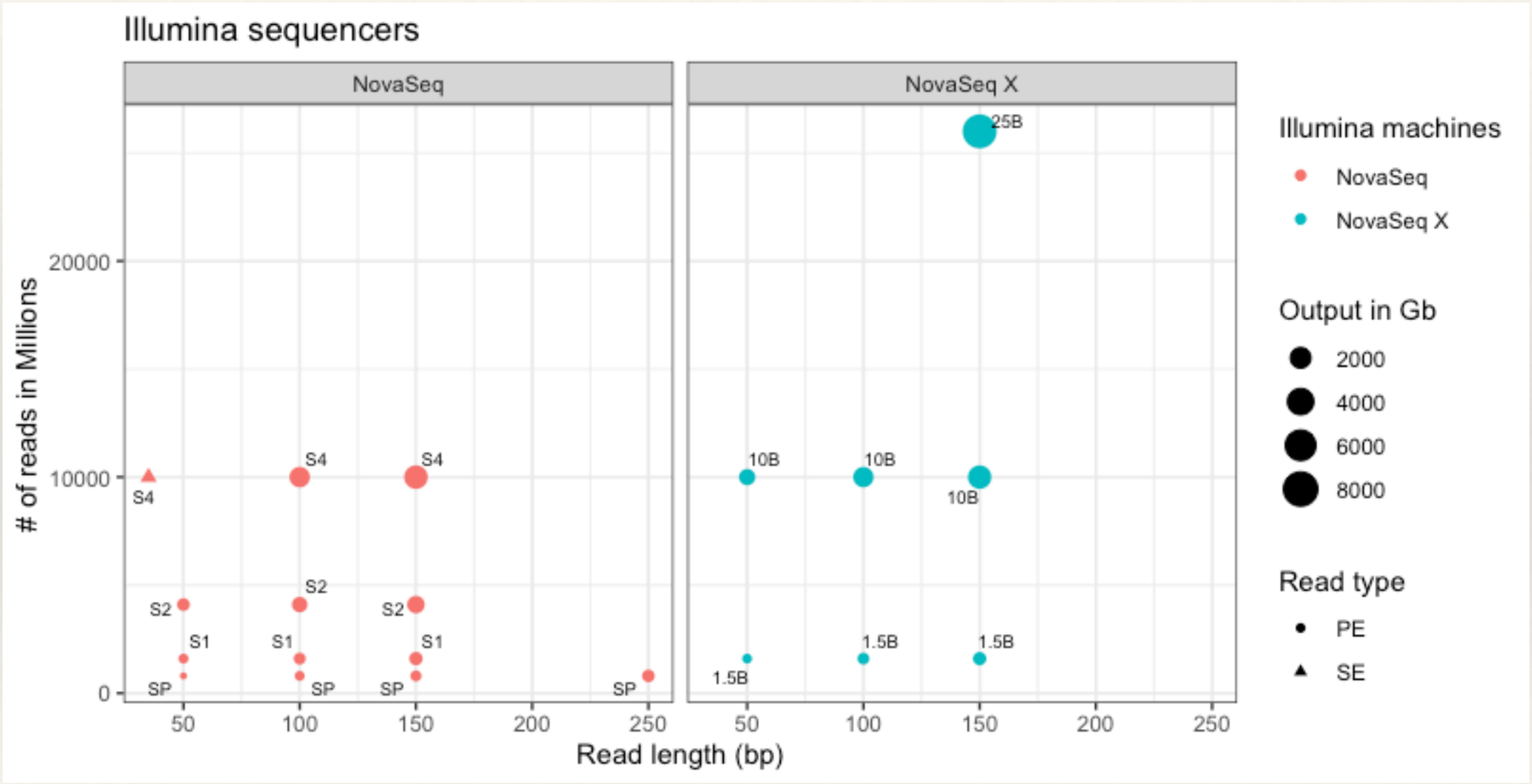
2022 XLEAP-SBS DRAGEN

2022 XLEAP-SBS DRAGEN

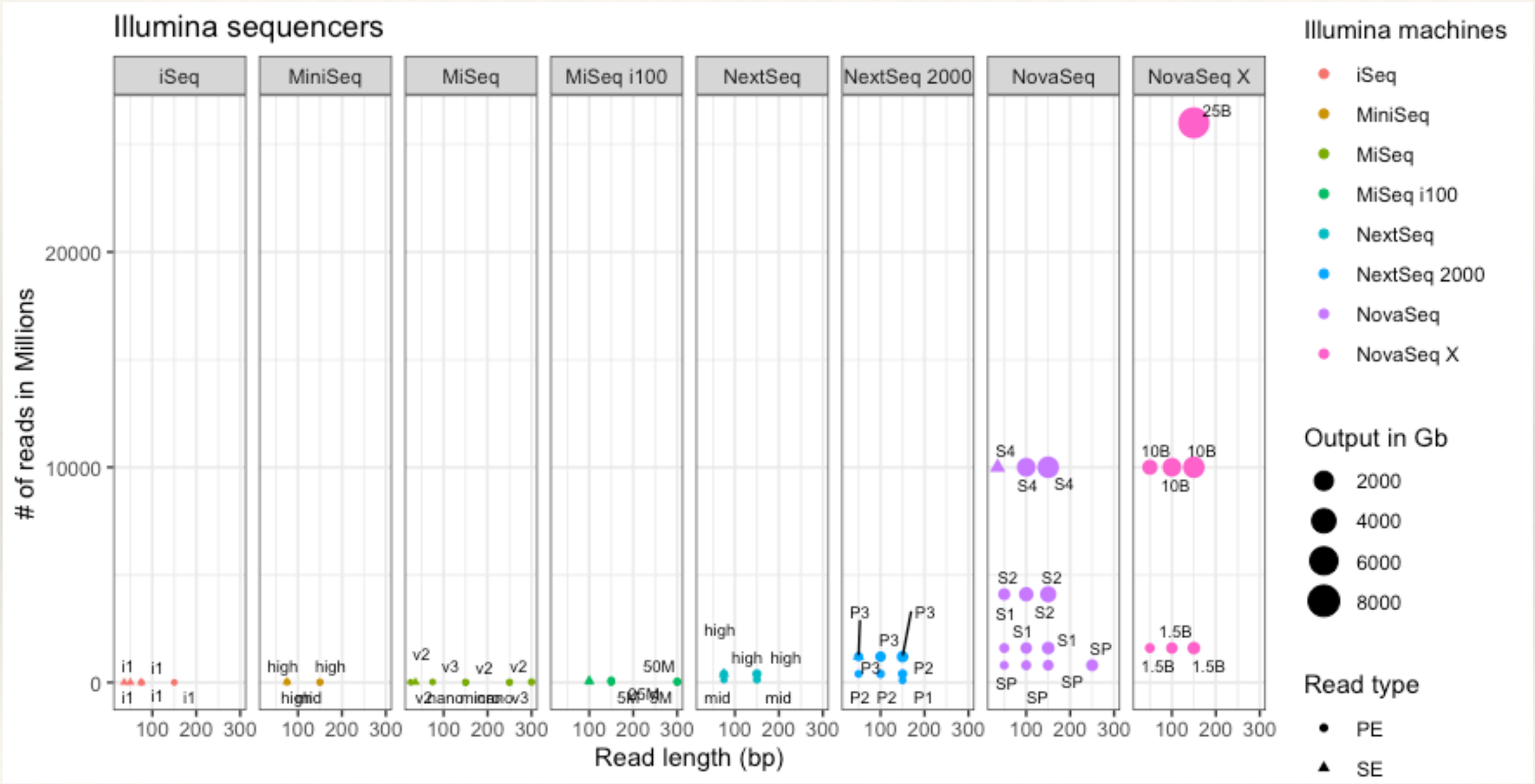
	NovaSeq 6000 System				NovaSeq X System			NovaSeq X Plus System		
Flow cell	SP	S1	S2	S4	1.5B	10B	25B	1.5B	10B	25B
Flow cells processed per run	1 or 2	1 or 2	1 or 2	1 or 2	1	1	1	1 or 2	1 or 2	1 or 2
Output range	65–800 Gb	134 Gb–1 Tb	333 Gb–2.5 Tb	280 Gb–6 Tb	165–500 Gb	1–3 Tb	8 Tb	165 Gb–1 Tb	1–6 Tb	8–16 Tb
Single-end reads per flow cell	800M	1.6B	4.1B	10B	1.6B	10B	26B	1.6B	10B	26B
Run time (hr)	13–38	13–25	16–36	< 44	17–23	18–25	~48	17–23	18–25	~48
Maximum read length (bp)	2 × 250	2 × 150	2 × 150	2 × 150	2 × 150	2 × 150	2 × 150	2 × 150	2 × 150	2 × 150
Included data analysis	—				Onboard DRAGEN secondary analysis					

https://www.illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/sequencing-platforms-brochure-m-gl-00451/sequencing-platforms-brochure-m-gl-00451.pdf

Data output

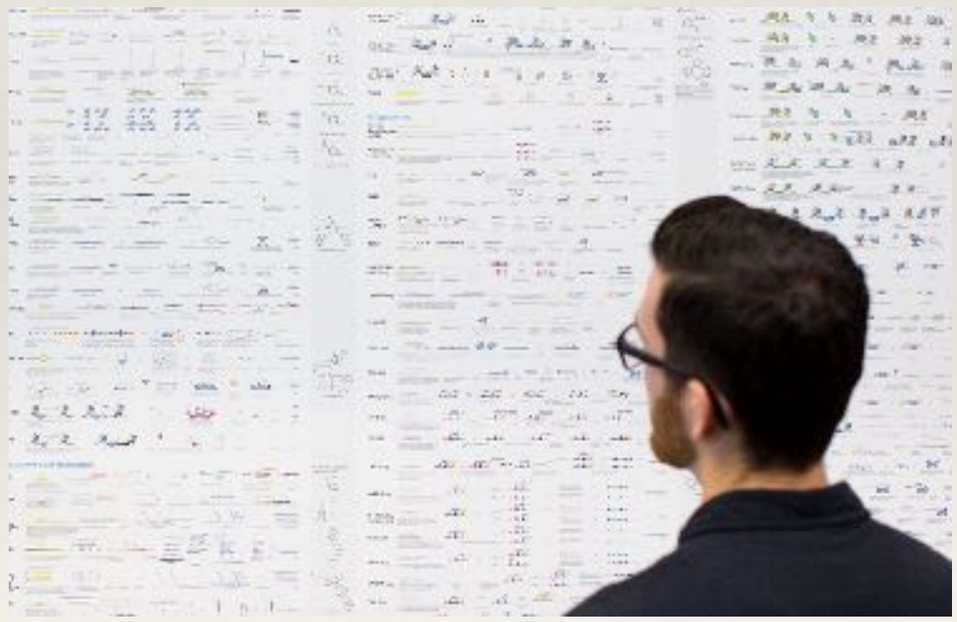

Illumina sequencers
NovaSeq
NovaSeq X
# of reads in Millions
20000
10000
0
50
100
150
200
250
Read length (bp)
25B
10B
1.5B
S4
S2
S1
SP
Illumina machines
NovaSeq
NovaSeq X
Output in Gb
2000
4000
6000
8000
Read type
PE
SE

Data output


Illumina sequencers
iSeq
MiniSeq
MiSeq
MiSeq i100
NextSeq
NextSeq 2000
NovaSeq
NovaSeq X
# of reads in Millions
20000
10000
0
Read length (bp)
100 200 300
25B
S4
S2
S1
SP
10B
1.5B
P3
P2
P1
i1
high
mid
v2
v3
50M
Illumina machines
iSeq
MiniSeq
MiSeq
MiSeq i100
NextSeq
NextSeq 2000
NovaSeq
NovaSeq X
Output in Gb
2000
4000
6000
8000
Read type
PE
SE

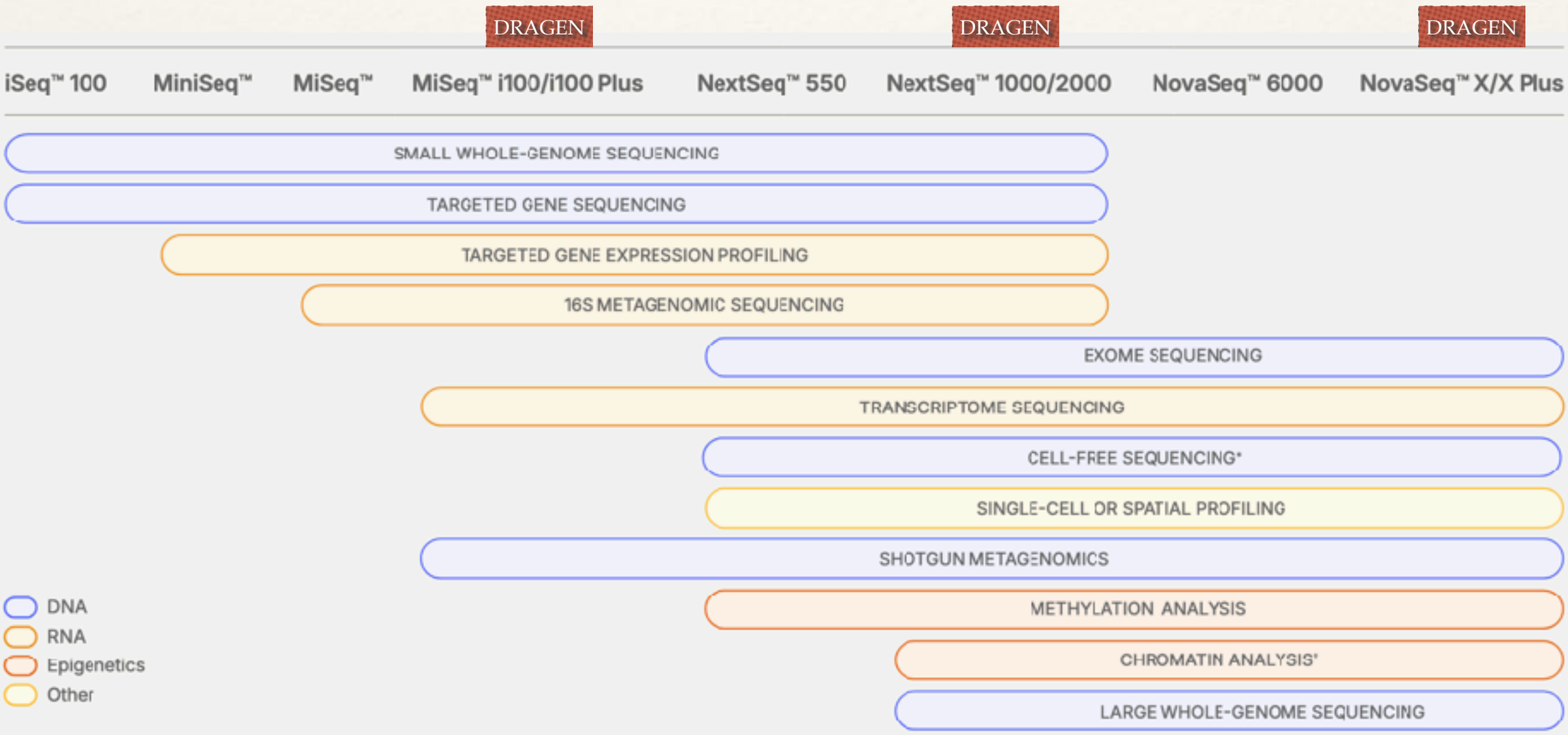
What can you sequence using Illumina

- DNA studies
  - Whole genome sequencing - short reads are a pitfall
  - Genome re-sequencing
  - Exomes and target re-sequencing…
  - ChiP seq and more…
- RNA studies
- modification studies
  - Methylation and more…



https://www.illumina.com/library-prep-array-kit-selector.html

Applications



https://www.illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/sequencing-platforms-brochure-m-gl-00451/sequencing-platforms-brochure-m-gl-00451.pdf