HTS

Arvind Sundaram
Sep 04, 2017

Norwegian Sequencing Centre
OUS, Ullevål, Oslo
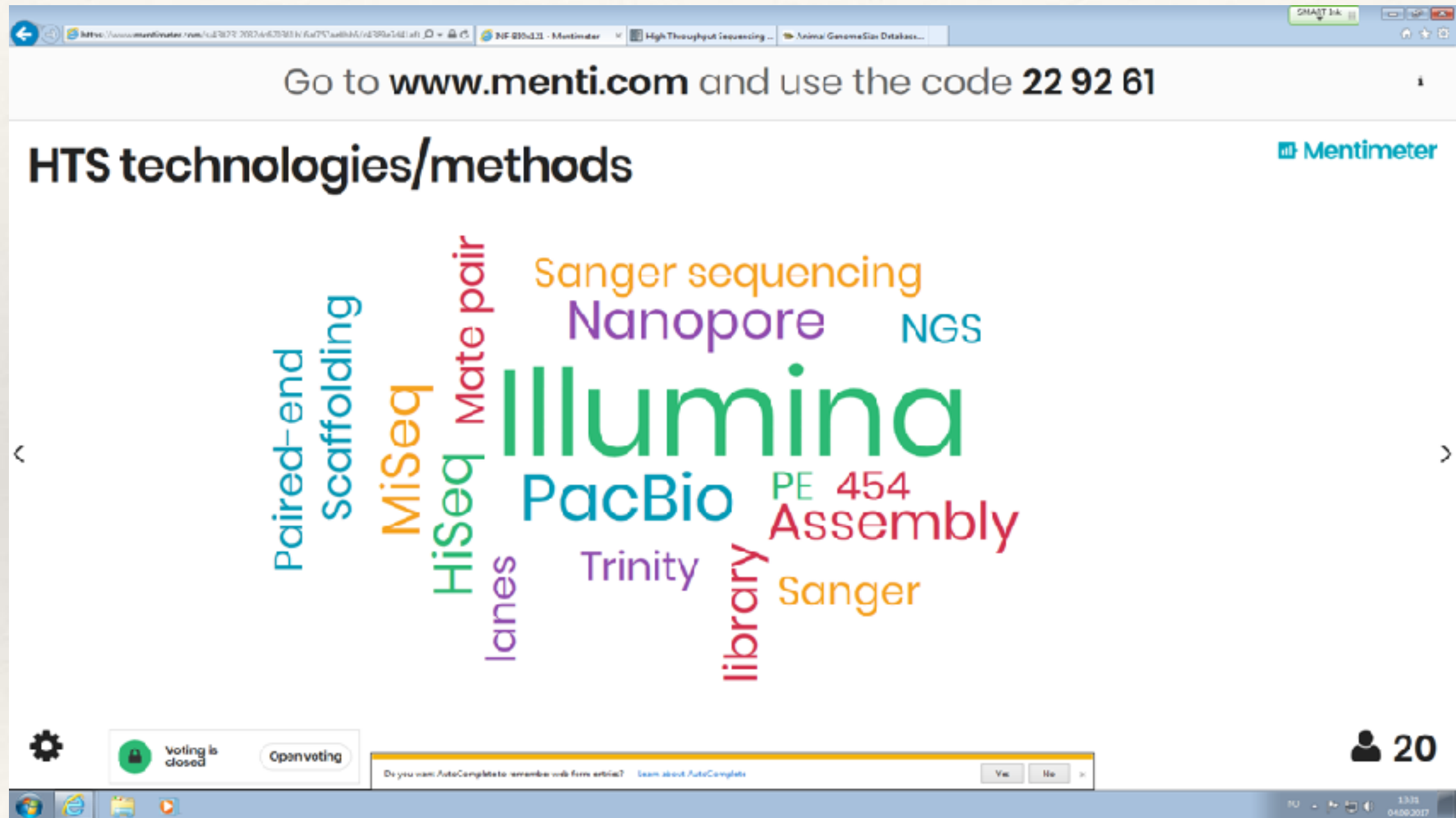
High throughput sequencing

High throughput sequencing

High throughput (DNA) sequencing
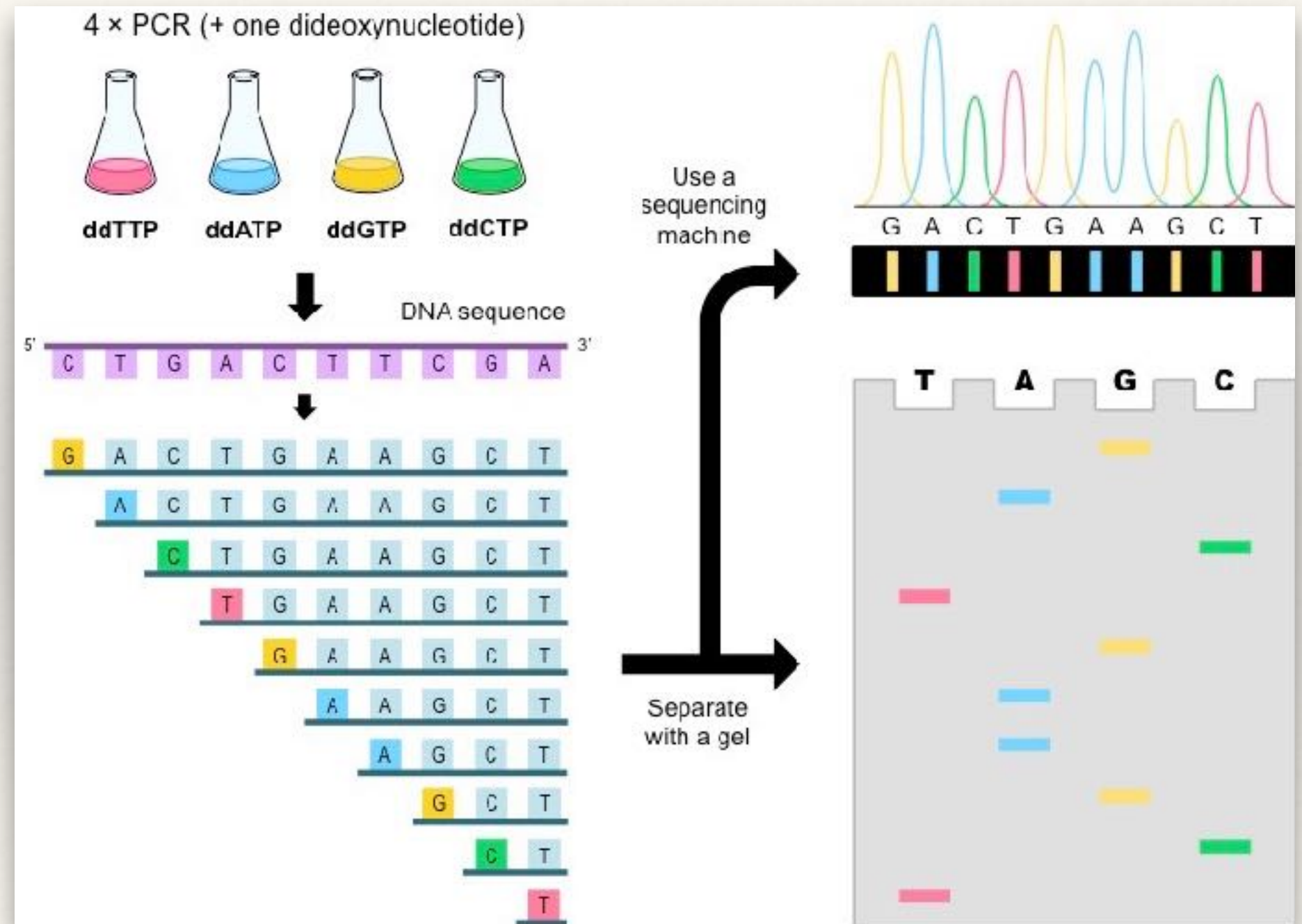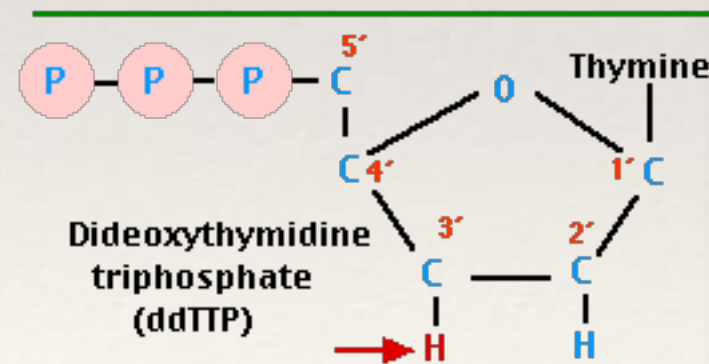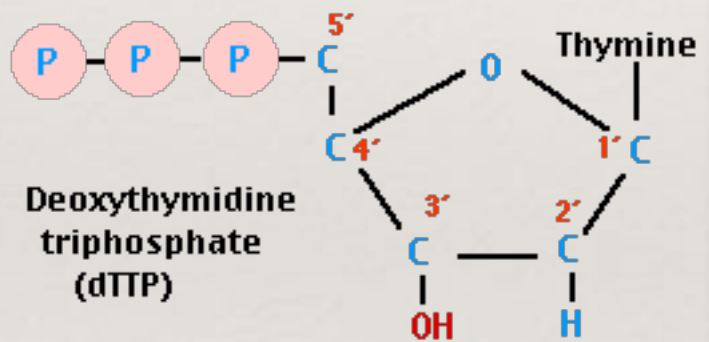
- First generation

- Second generation

- Third generation

- What about RNA???

  - unstable

  - amplification issues

- reverse-transcribed to cDNA before being sequenced

The sequence of sequencers: The history of sequencing DNA
https://doi.org/10.1016/j.ygeno.2015.11.003

Sequencing

Dideoxynucleotides
ddNTPs

First generation DNA sequencing

- ❖ SANGER

- ❖ Highly automated (ABI Sanger 3730xl)

- ❖ Up to 1 kb; high quality data; multiplexed

Second generation DNA sequencing

❖ Sequencing-by-synthesis aka SBS

    ❖ https://www.youtube.com/watch?v=fCd6B5HRaZ8

❖ Mass parallelisation and real high throughput

Third generation DNA sequencing

❖ **Single molecule sequencing**

❖ **Real time sequencing**

❖ **Non-SBS methods**

❖ Mostly under-development

❖ Few commercially available/successful methods

    ❖ PacBio (https://www.youtube.com/watch?v=NHCJ8PtYCFc)

    ❖ Oxford Nanopore (https://www.youtube.com/watch?v=hs0FdiTHMbc)

High throughput sequencing

illumina®

MiniSeq
MiSeq
NextSeq
HiSeq 2500
HiSeq 3000/4000
HiSeq X
NovoSeq

Roche 454
SOLiD
HeliScope

PACBIO®

RS II
Sequel

ion torrent

PGM
Proton
S5 / S5XL

BGISEQ
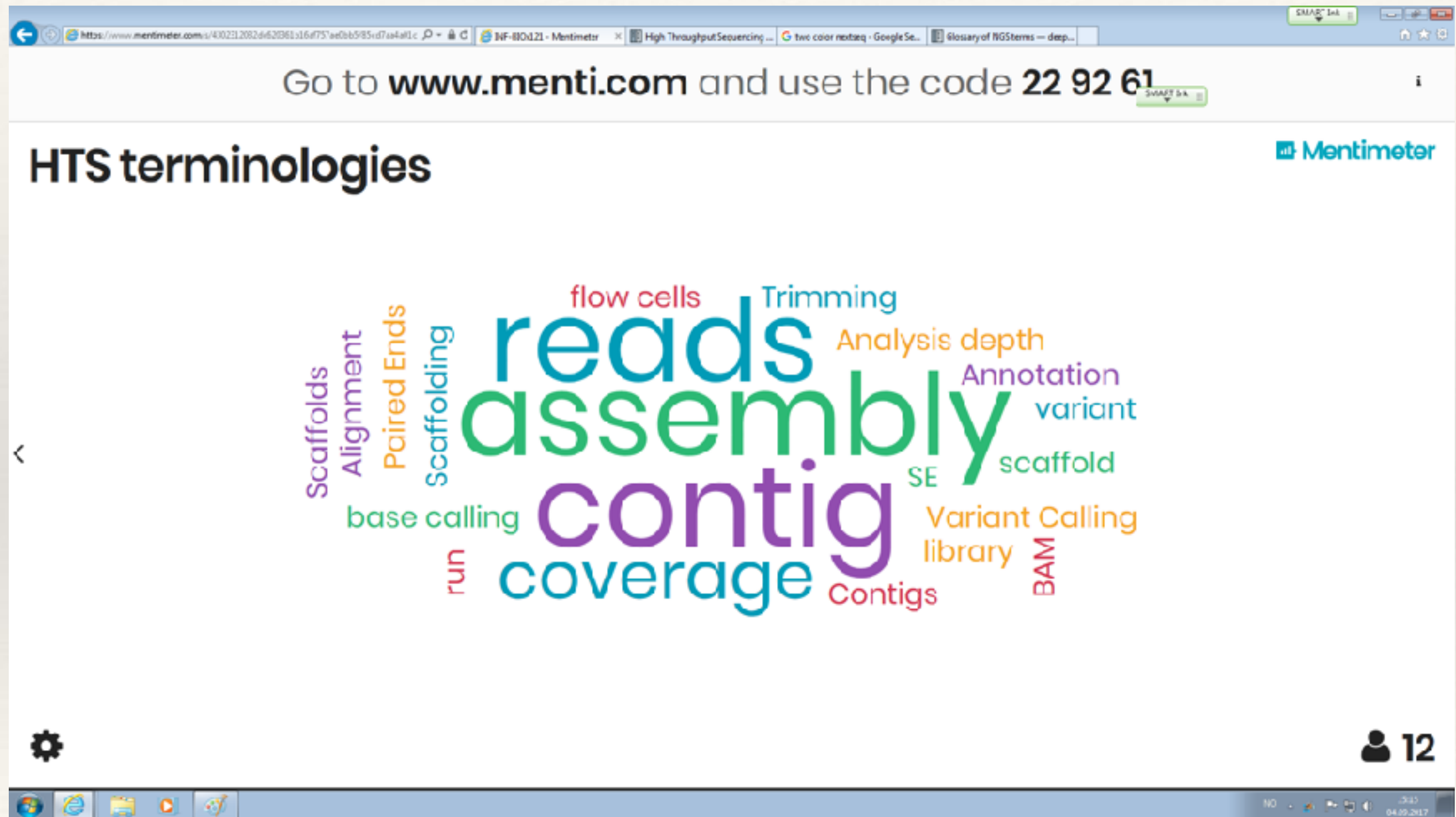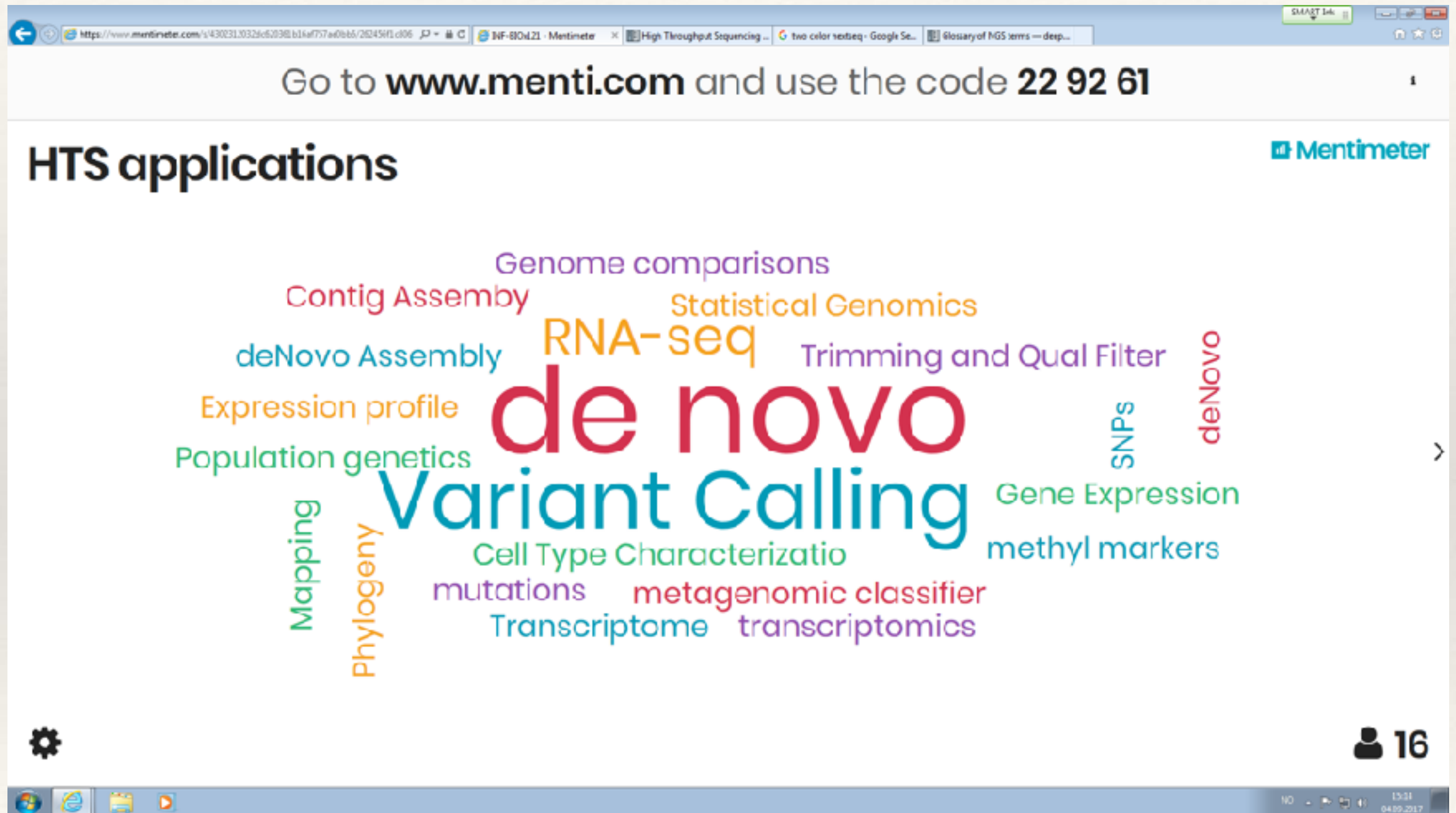
Special types

- 10X genomics

- Dovetail genomics

- BioNano genomics
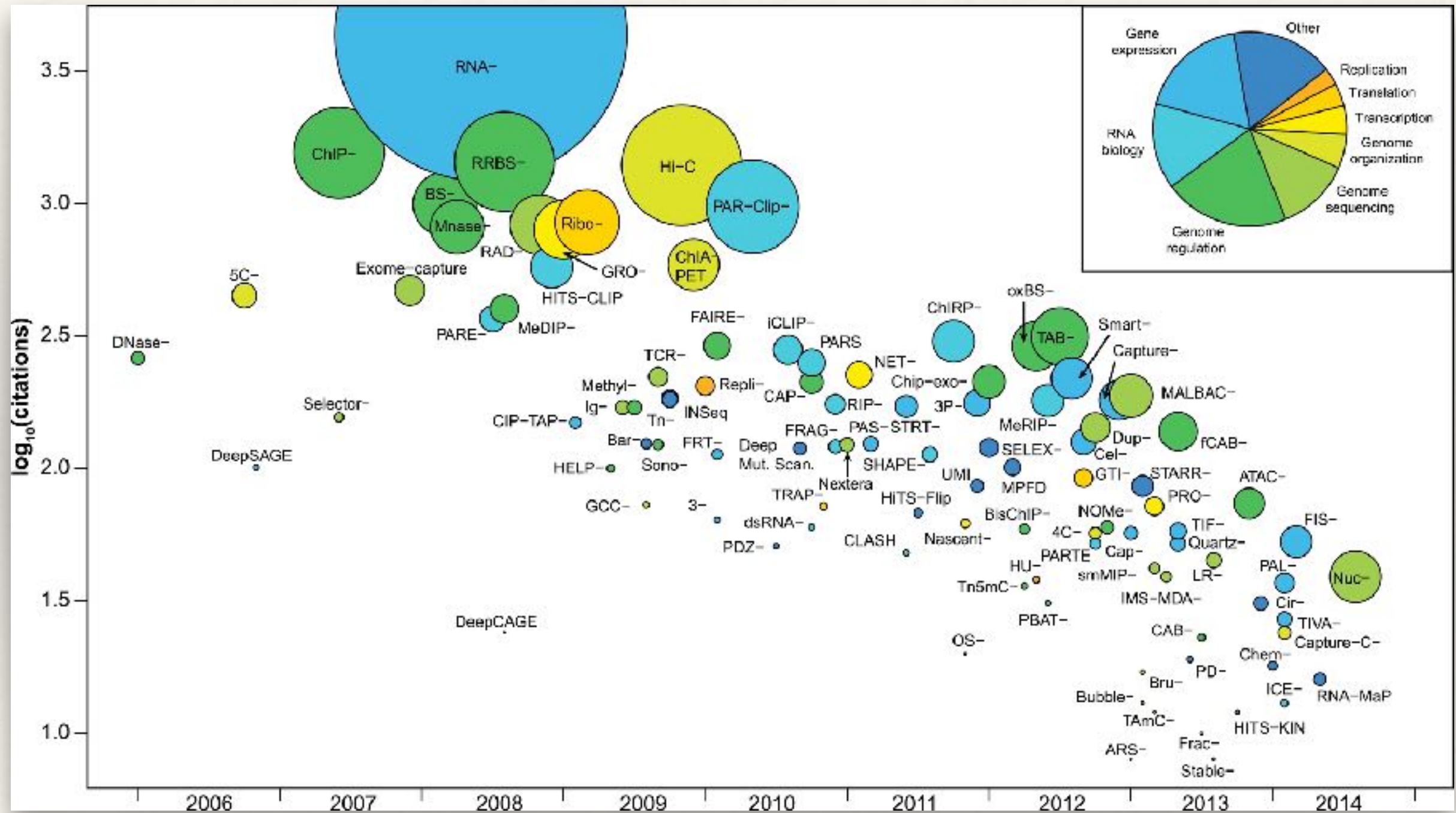
- Moleculo/TruSeq synthetic reads

High throughput sequencing

High throughput sequencing

High throughput sequencing

Experimental design

- Platform choice
- Technology variation
  - Technical bias
    - Lane bias
    - Index/barcode bias
    - Batch effect
    - Duplicates

- Biological question
- Sample variation
- Sequencing depth
- Data analysis
- Species-specific information
  - Is there a genome sequence available??
  - Genome size (c-value) (http://genomesize.com/)
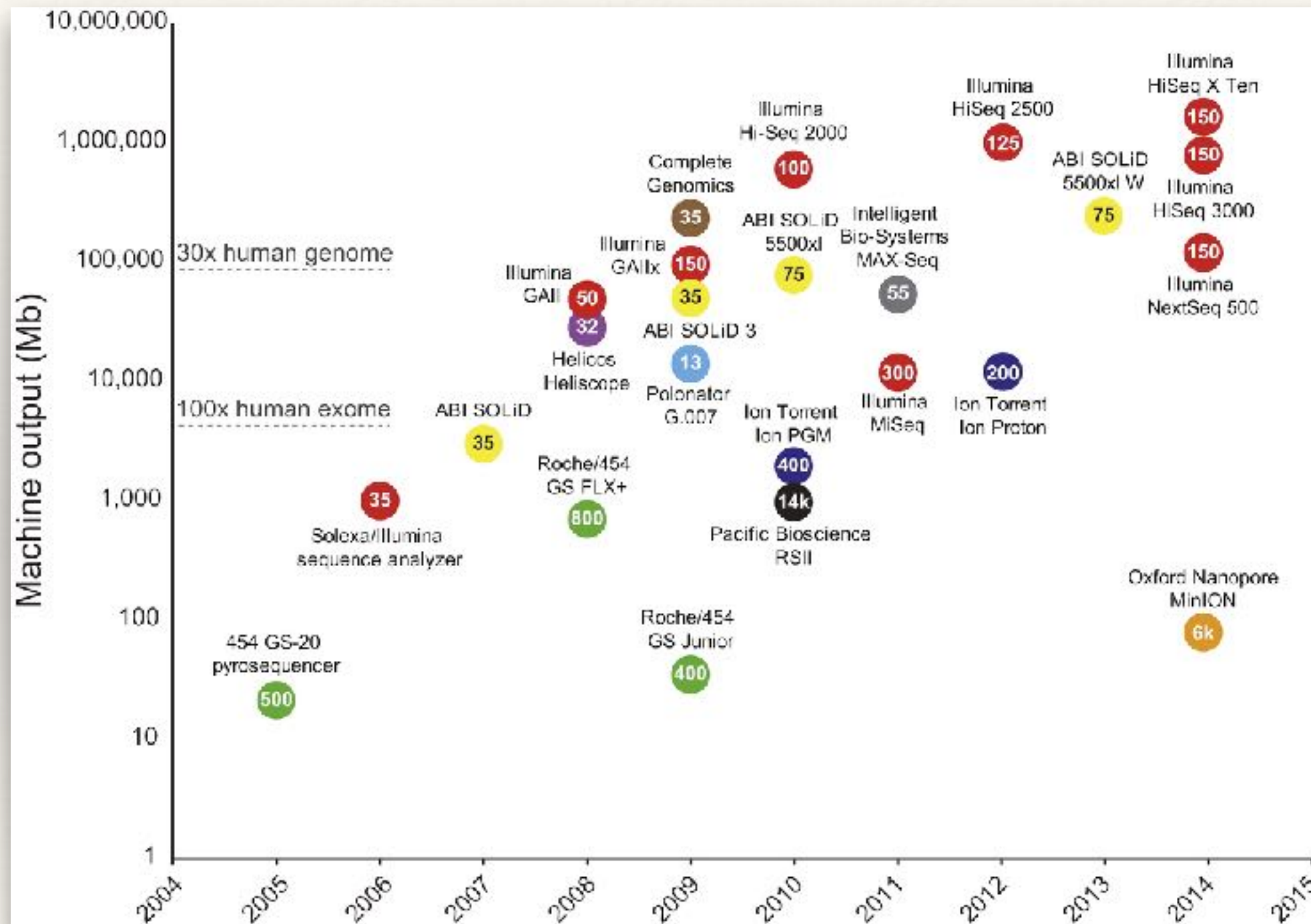
Experimental design
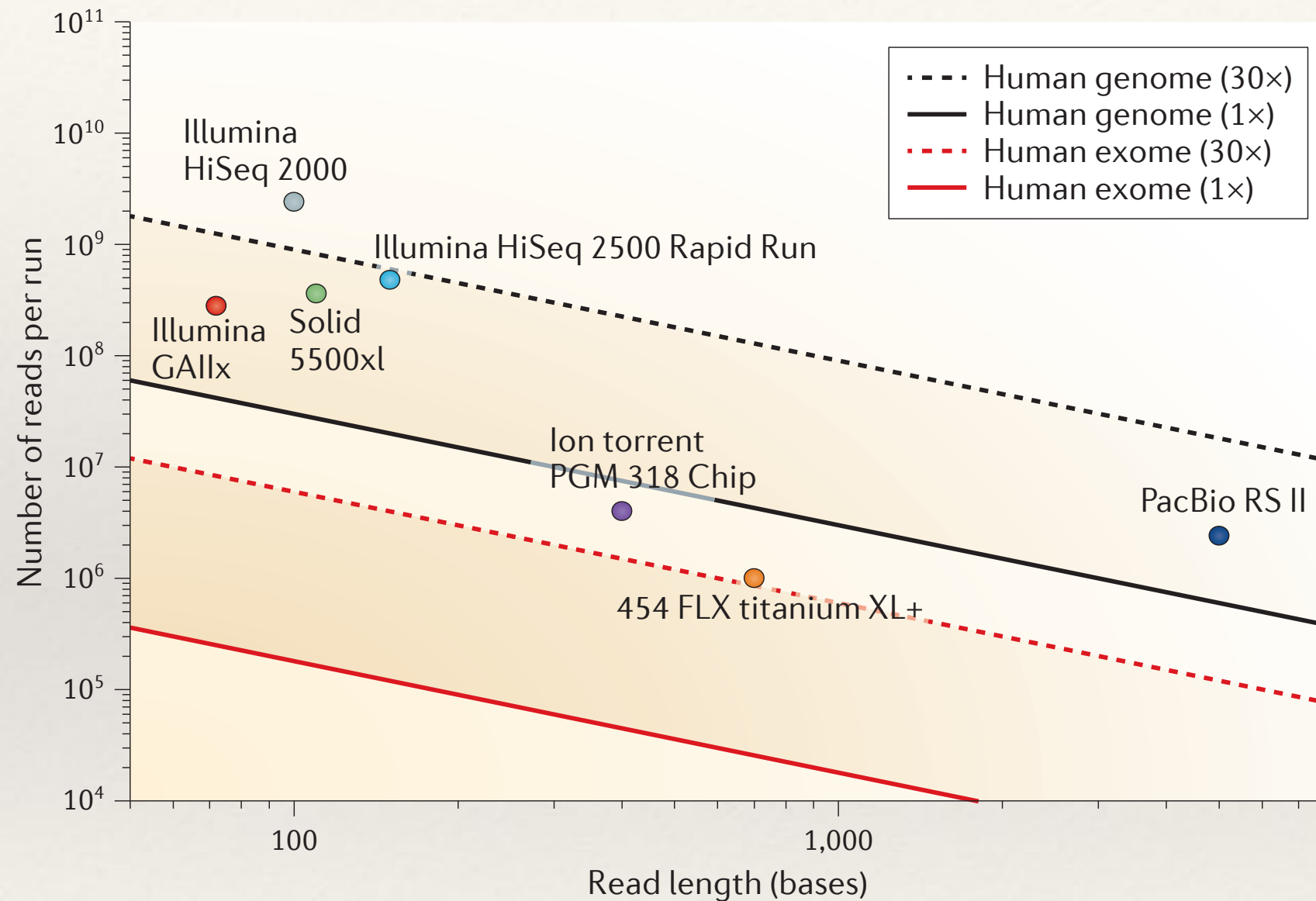
- Short or long fragments
- Single of paired end

- Depth required
- Coverage required

- Multiplexing
  - single index
  - dual index
  - or more??

- Replicates
  - Biological
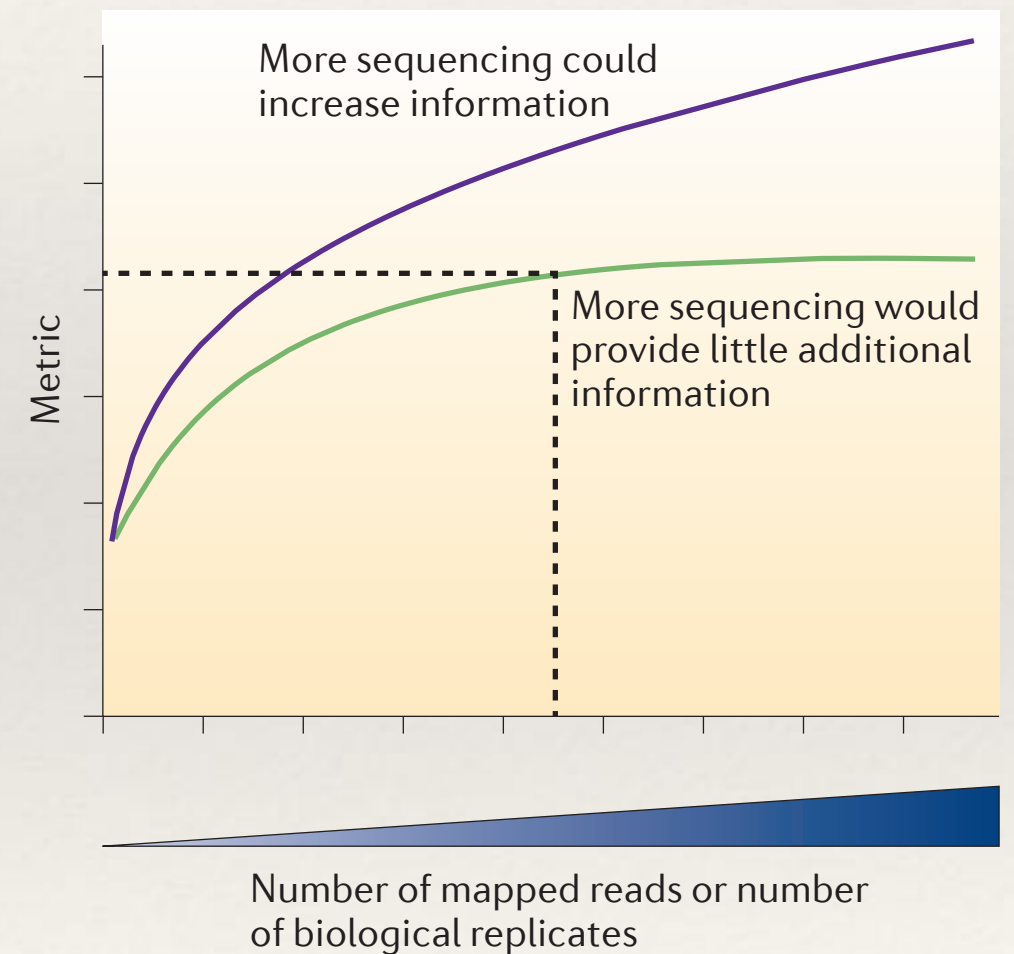 - Technical
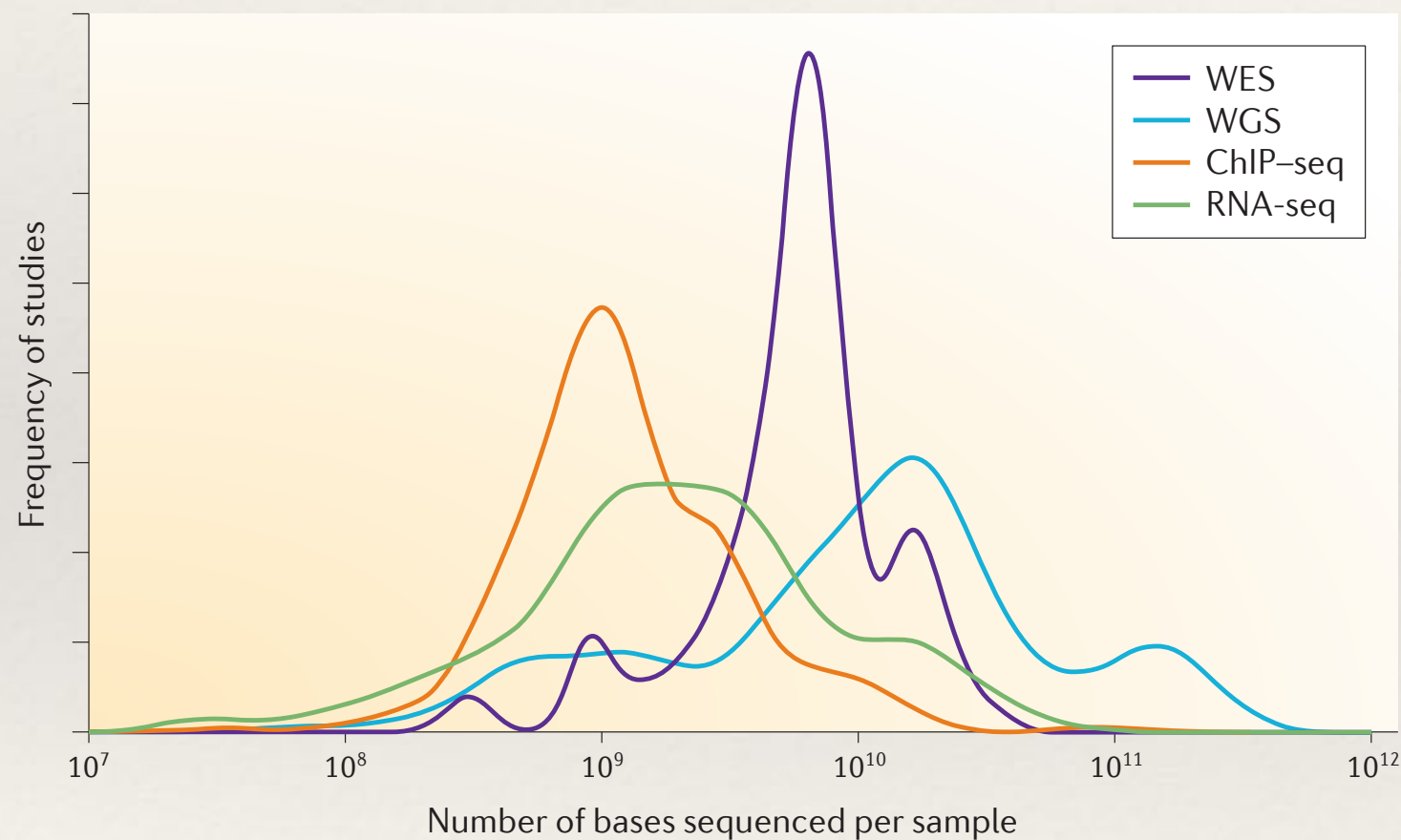
High throughput sequencing
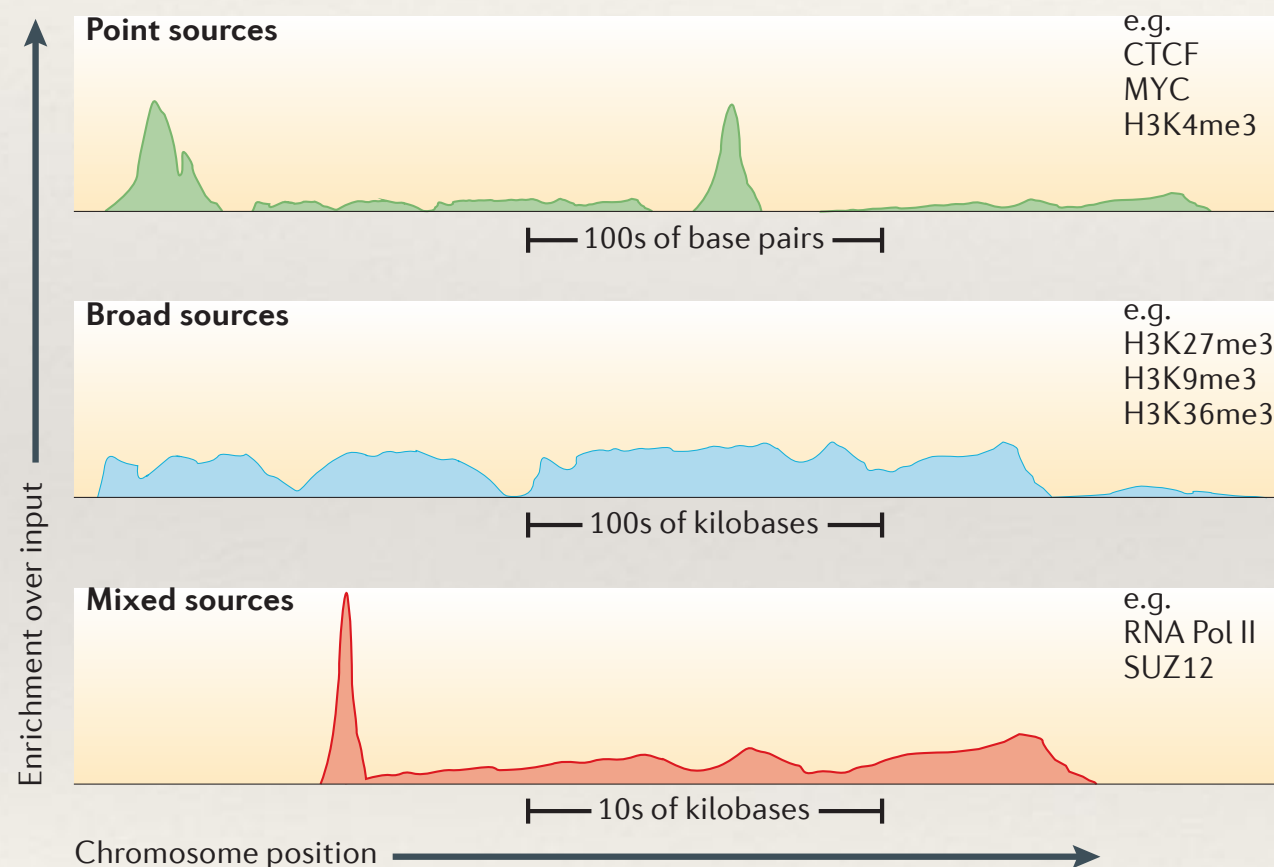
Sequencing depth and coverage

GAIIx, Genome Analyzer IIx; PacBio, Pacific Biosciences; PGM, personal genome machine.

Sequencing depth and coverage

Sequencing depth and coverage

Techniques	Read counts in representative studies
DNaseI-seq and FAIRE–seq	20–50 million
CLIP–seq	7.5 million; 36 million
iCLIP and PAR–CLIP	8 million; 14 million
CHiRP and CHART	26 million
4C	1–2 million
ChIA–PET	20 million
5C	25 million
Hi-C	>100 million
MeDIP–seq	60 million
CAP–seq	>20 million
ChIP–seq	>10 million per sample (point source); >20 million per sample (broad source)