



NUTRIGENOMICS
ERA CHAIR | WELCOME2

Measuring the Transcriptome

Carsten Carlberg, 29.05.2024, Maastricht



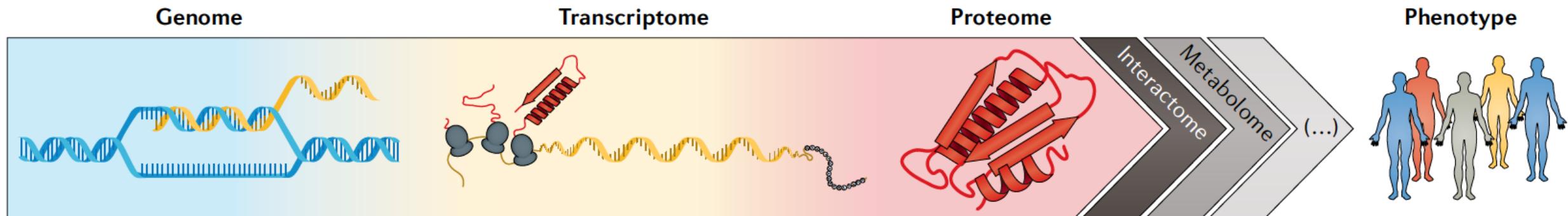
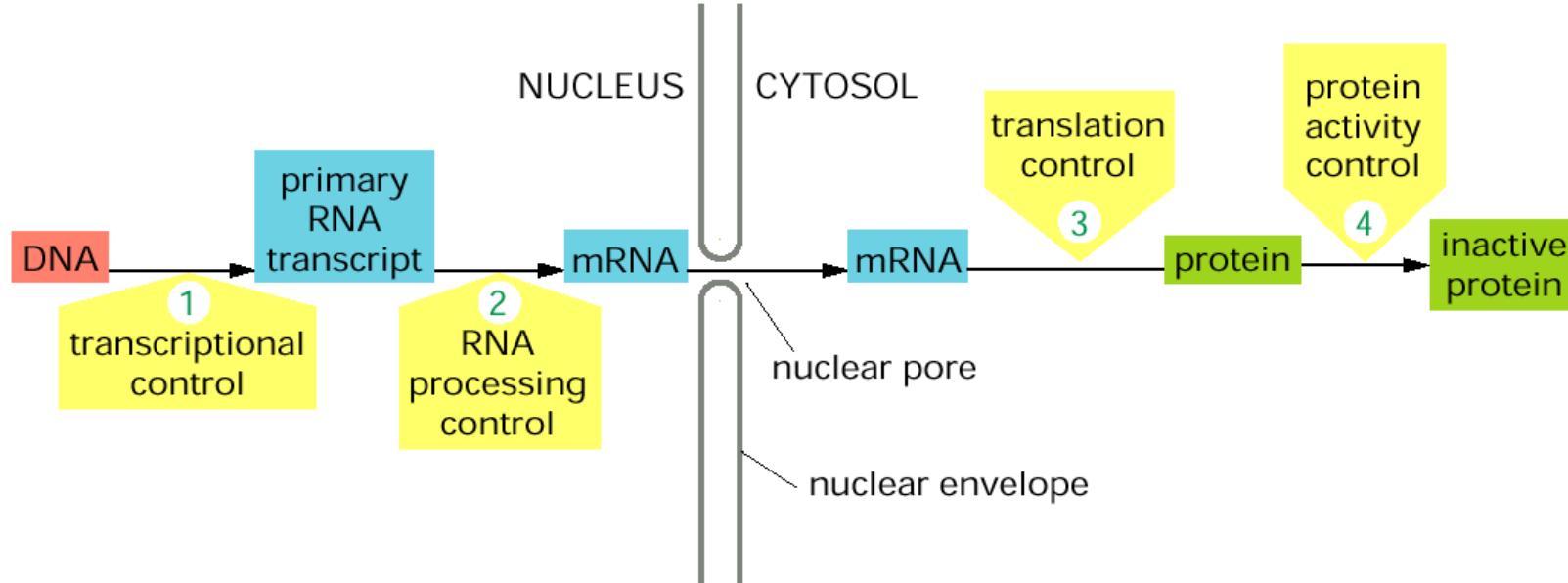
Institute
of Animal Reproduction and Food Research
Polish Academy of Sciences
in Olsztyn



This project has received funding from the
European Union's Horizon 2020 research
and innovation programme under grant
agreement No 952601

Processing information in living cells

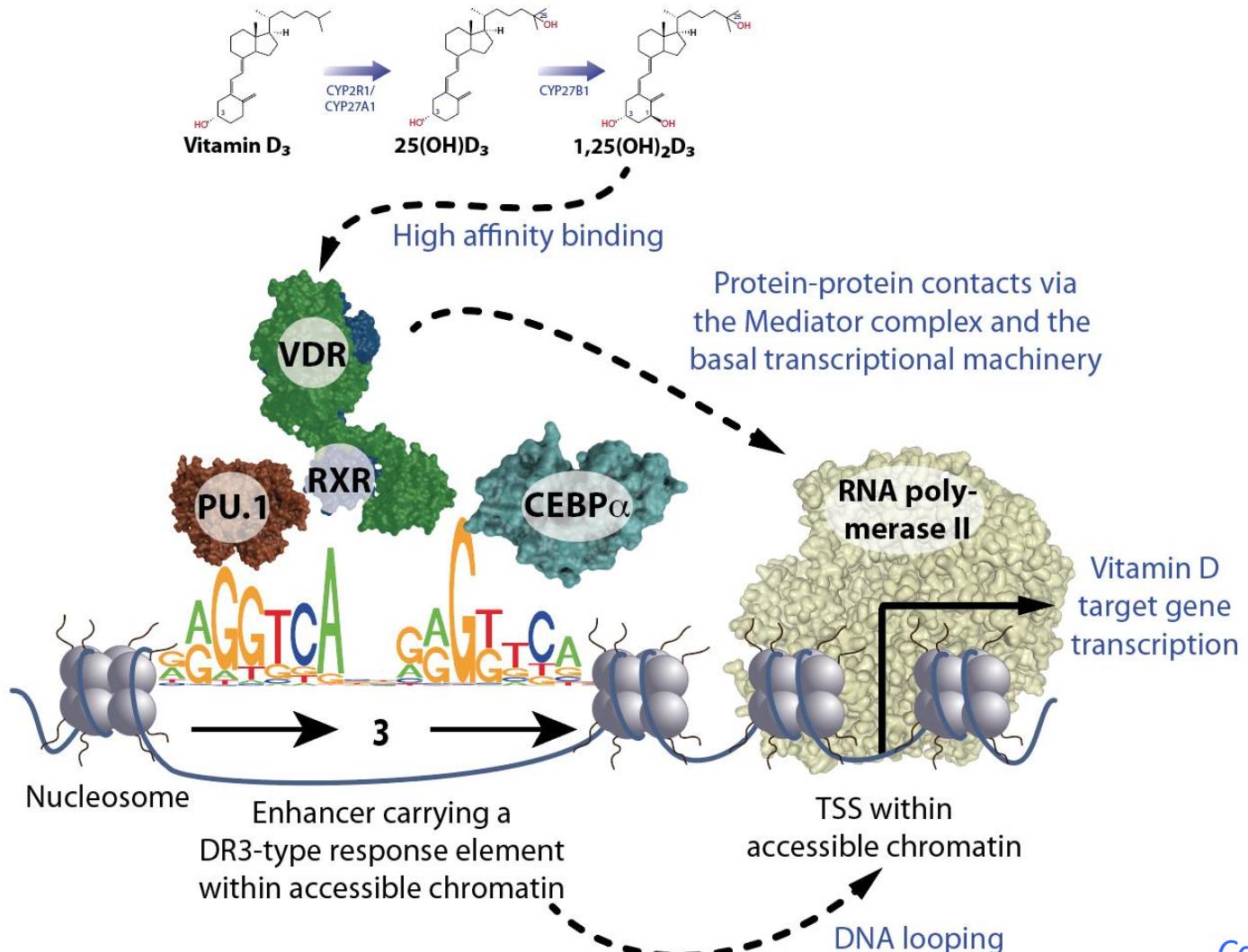
(Epi)genomics → Transcriptomics → (Functional) Proteomics



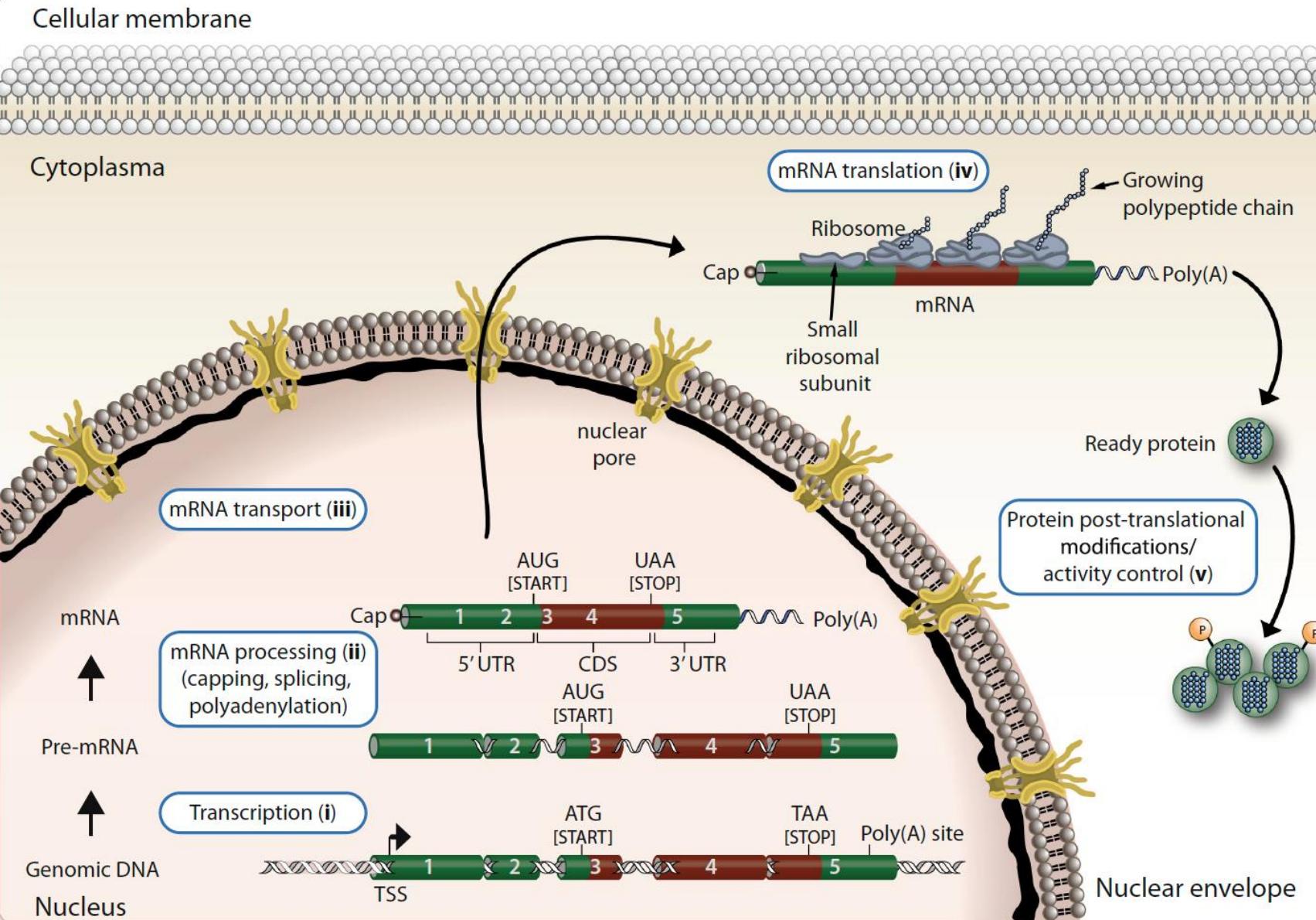
Genome architecture	Histone modification	Transcription	miRNA	Translation	Localization
		Splicing	Polyadenylation	Folding	Phosphorylation
Genetic variation	Chromosome topology	DNA methylation	mRNA transport	mRNA modification	Ubiquitylation
				mRNA degradation	Protein degradation



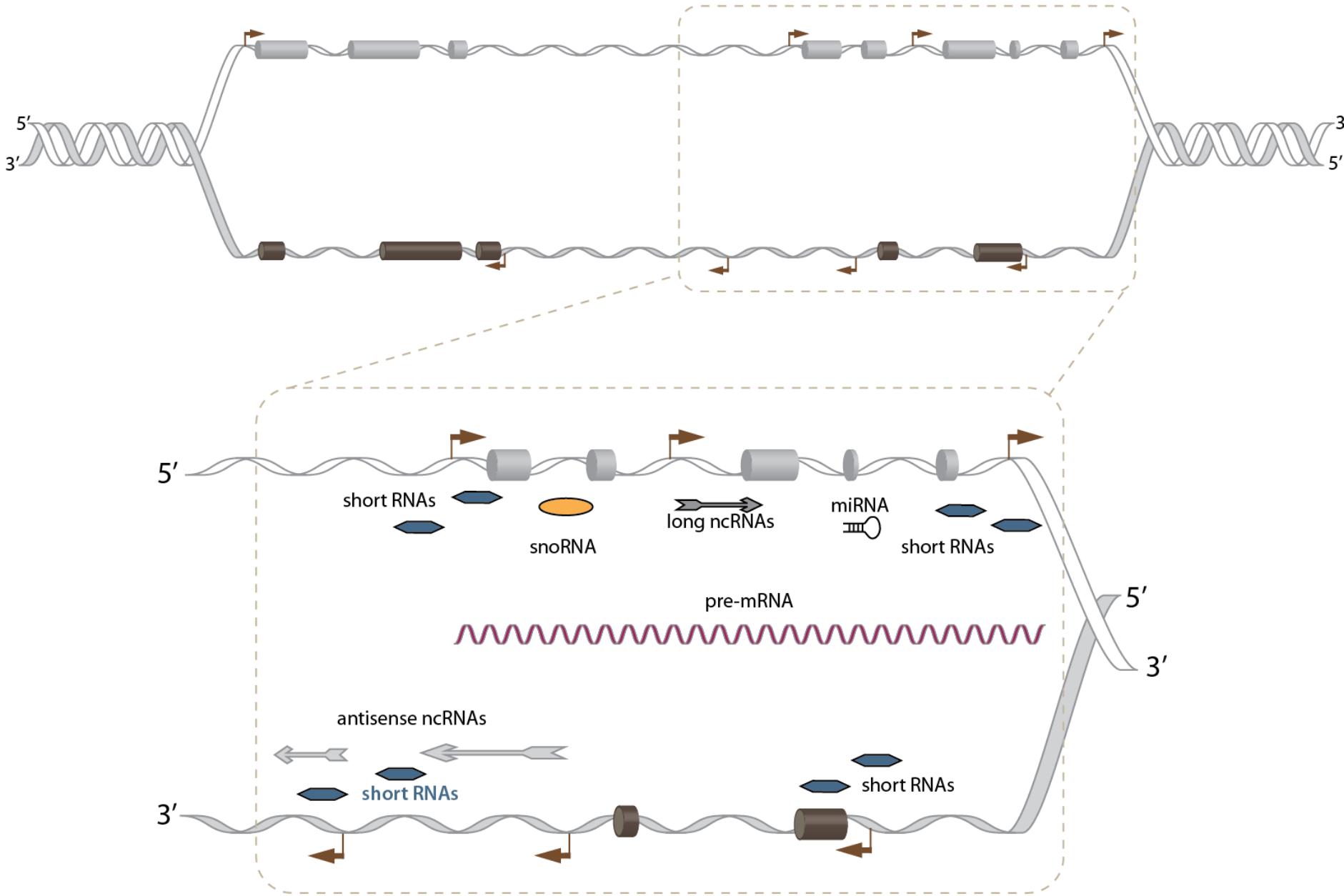
Principles of vitamin D signaling



The central dogma of molecular biology



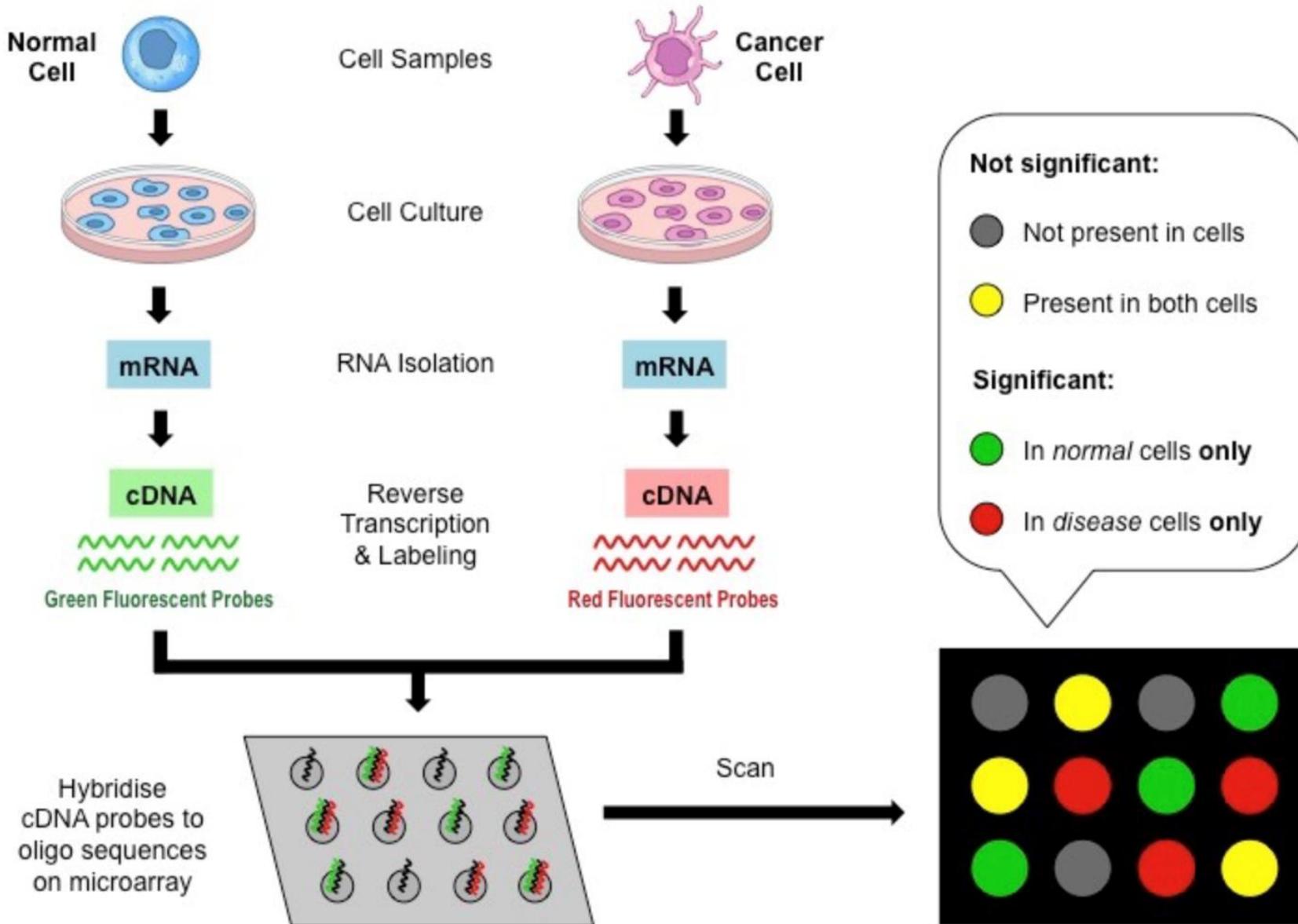
The “hidden” transcriptome

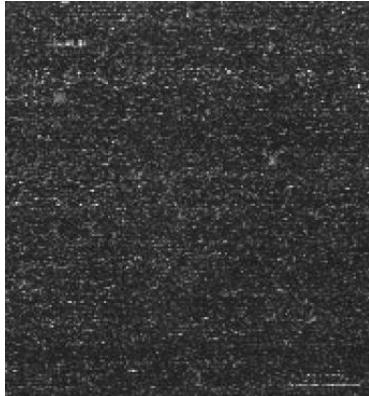


Reflection: The human genome in numbers

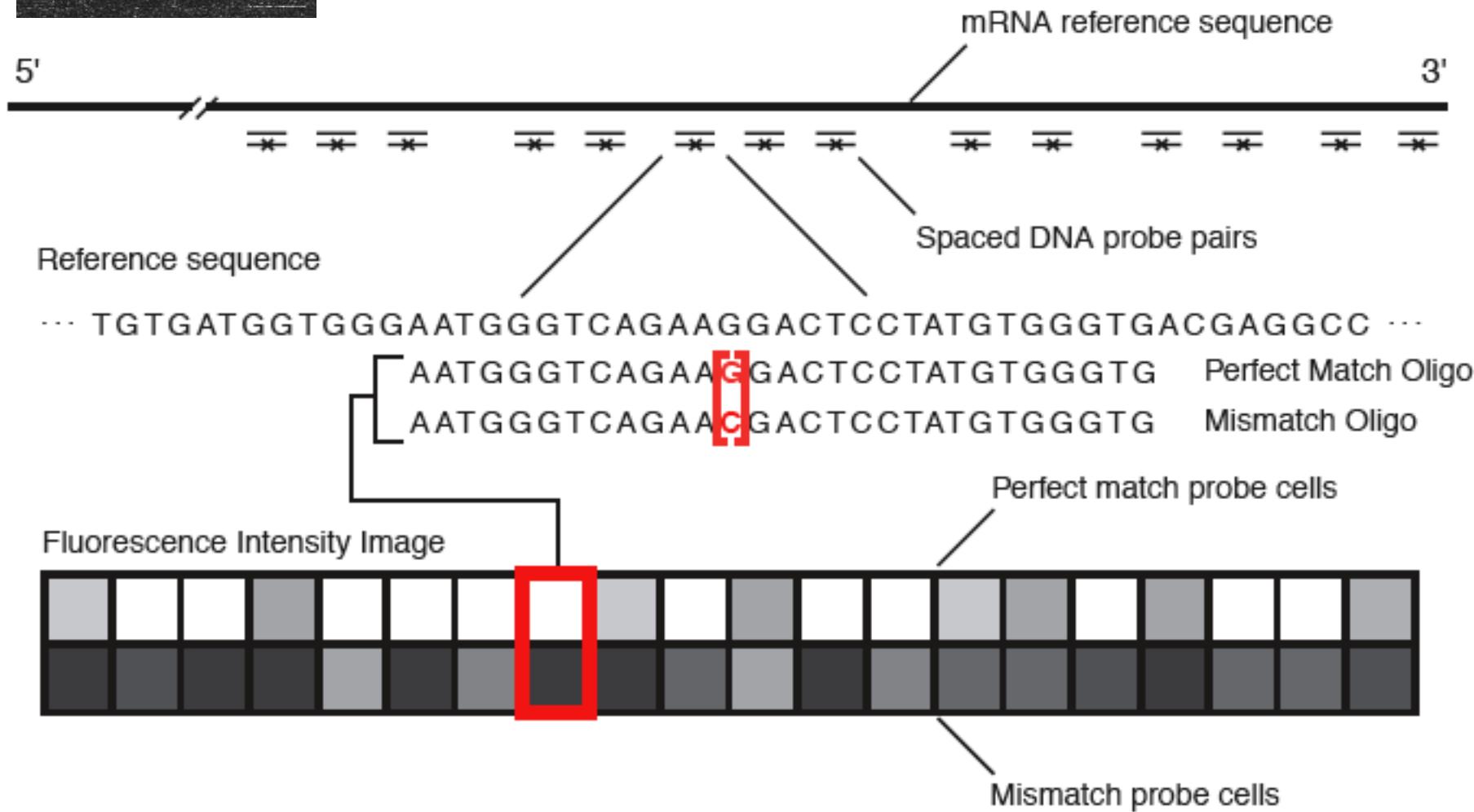
Number of chromosomes	22 + X + Y
Genome size (nt)	3,008,269,832
Number of genes	58,037
Number of transcripts	198,950
Number of protein-coding genes	19,950
Number of protein-coding transcripts	80,087
Number of long ncRNA transcripts	27,691
Number of pseudogenes	14,650
Number of small RNAs	7258
Number of miRNAs	2588
Number of tRNAs	631
Average size on human pre-mRNA (nt)	16,000
Average size of human protein (aa)	460
Average number of exons/gene	8

Microarrays: The principle

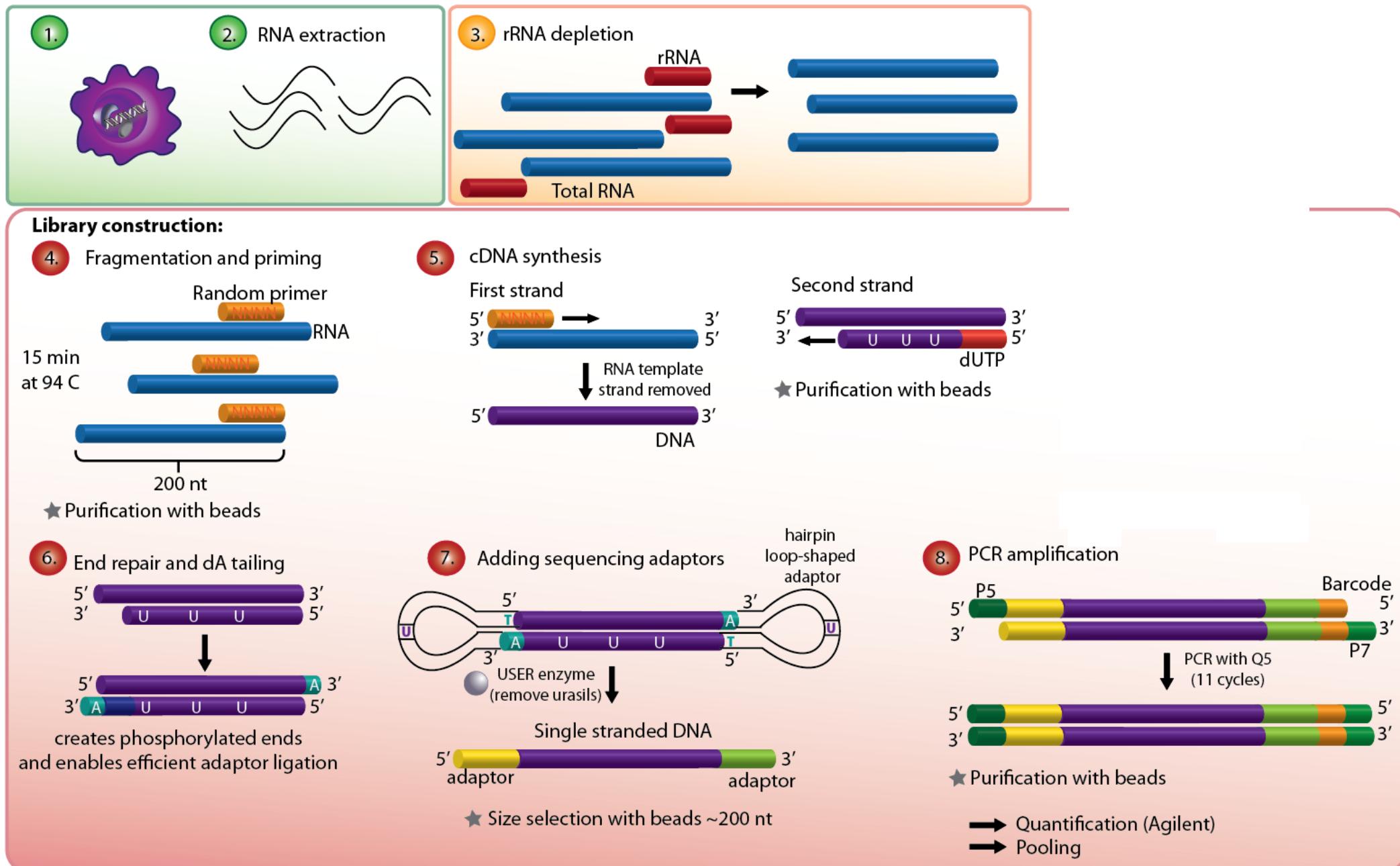




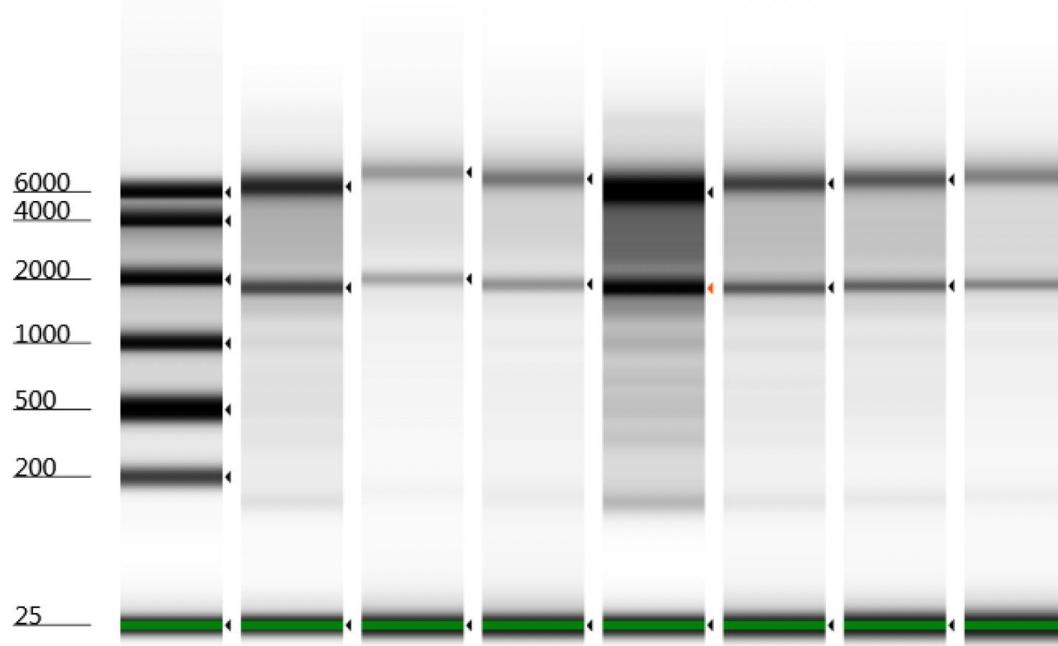
Affymetrixs chips



RNA sequencing (RNA-seq) method



RNA isolation and quality check

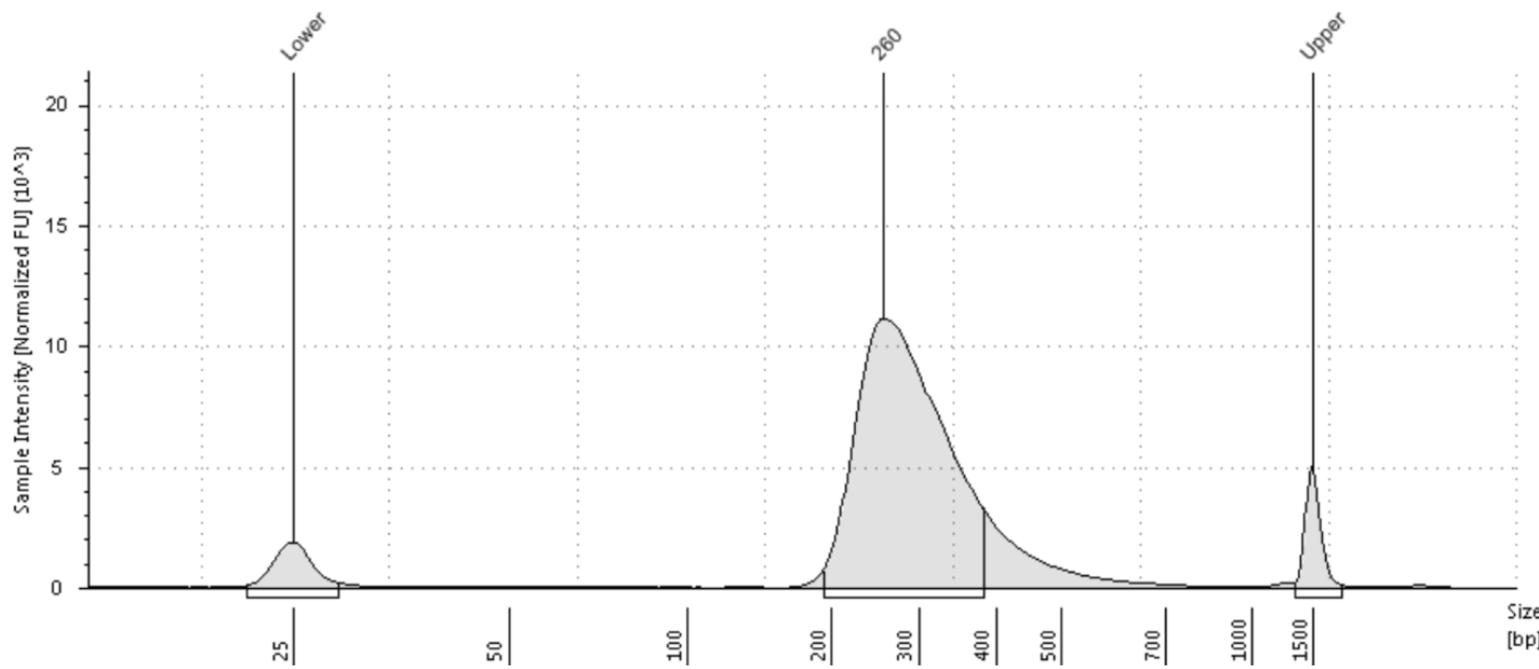


RIN^e RIN^e RIN^e RIN^e RIN^e RIN^e RIN^e
8.5 8.4 8.5 8.6 8.7 8.5 8.6

Well	RIN ^e	28S/18S (Area)	Conc. [ng/ μ l]
A1	-	-	182
B1	8.5	1.8	77.3
C1	8.4	1.8	27.1
D1	8.5	2.1	36.7
E1	8.6	2.0	153
F1	8.7	1.7	54.4
G1	8.5	1.6	43.3
H1	8.6	1.6	28.6

RIN = RNA integrity number

Pools of RNA-seq libraires



Sample Table

Well	Conc. [ng/ μ l]	Sample Description	Alert	Observations
G1	82.1	S4P6		

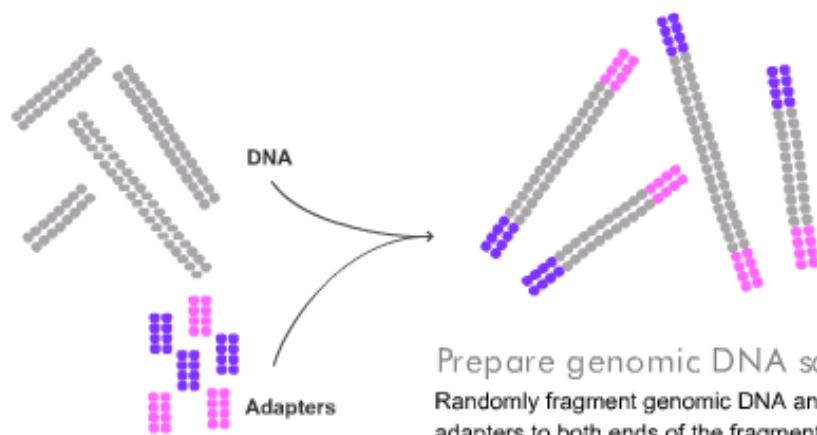
Peak Table

Size [bp]	Calibrated Conc. [ng/ μ l]	Assigned Conc. [ng/ μ l]	Peak Molarity [nmol/l]	% Integrated Area	Peak Comment	Observations
25	6.23	-	383	-		Lower Marker
260	82.1	-	485	100.00		
1500	6.50	6.50	6.67	-		Upper Marker

#	NEB Barcode	Sequence
53d84	oligo 1	ATCACG
11d84	oligo 2	CGATGT
33d84	oligo 3	TTAGGC
52d84	oligo 4	TGACCA
22d84	oligo 5	ACAGTG
24d84	oligo 6	GCCAAT
50d84	oligo 7	CAGATC
29d84	oligo 8	ACTTGA
39d84	oligo 9	GATCAG
1d84	oligo 10	TAGCTT
36d84	oligo 11	GGCTAC
48d84	oligo 12	CTTGTA
44d84	oligo 13	AGTCAA
12d84	oligo 14	AGTTCC
19d84	oligo 15	ATGTCA
26d84	oligo 16	CCGTCC
49d84	oligo 18	GTCCGC
10d84	oligo 19	GTGAAA
30d84	oligo 20	GTGGCC
51d84	oligo 21	GTTTCG

Massive parallel sequencing I

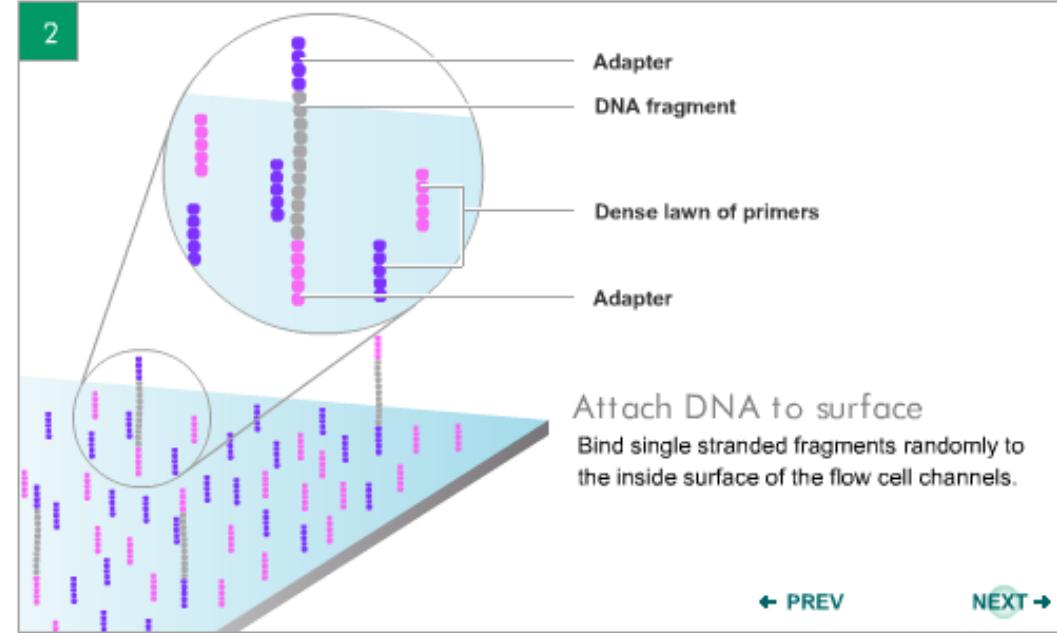
1



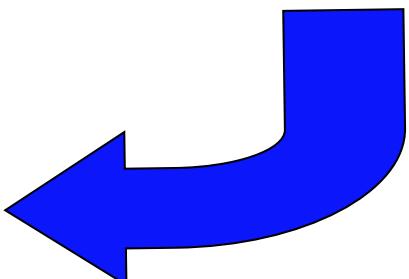
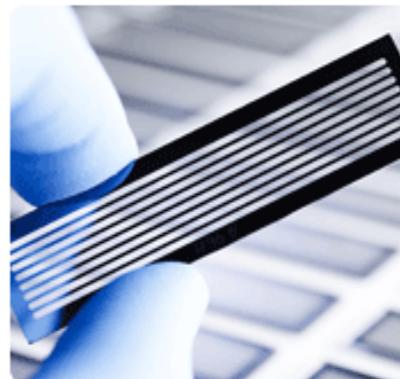
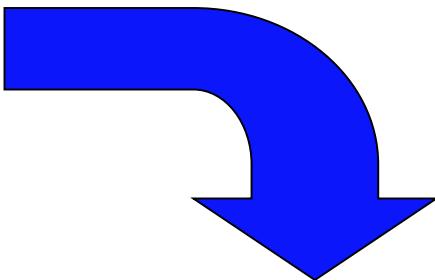
← PREV

NEXT →

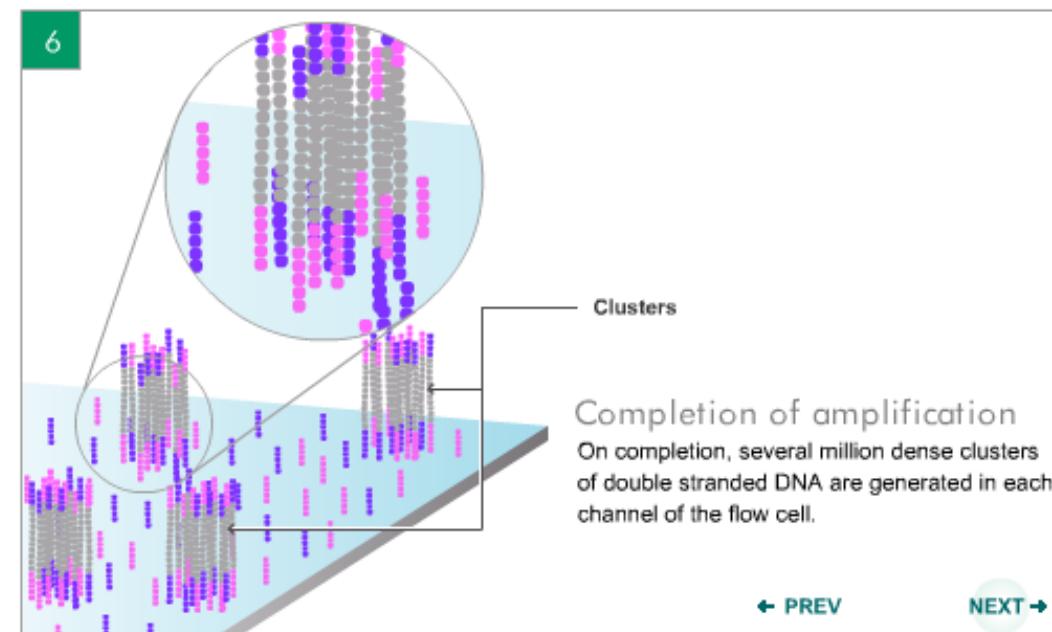
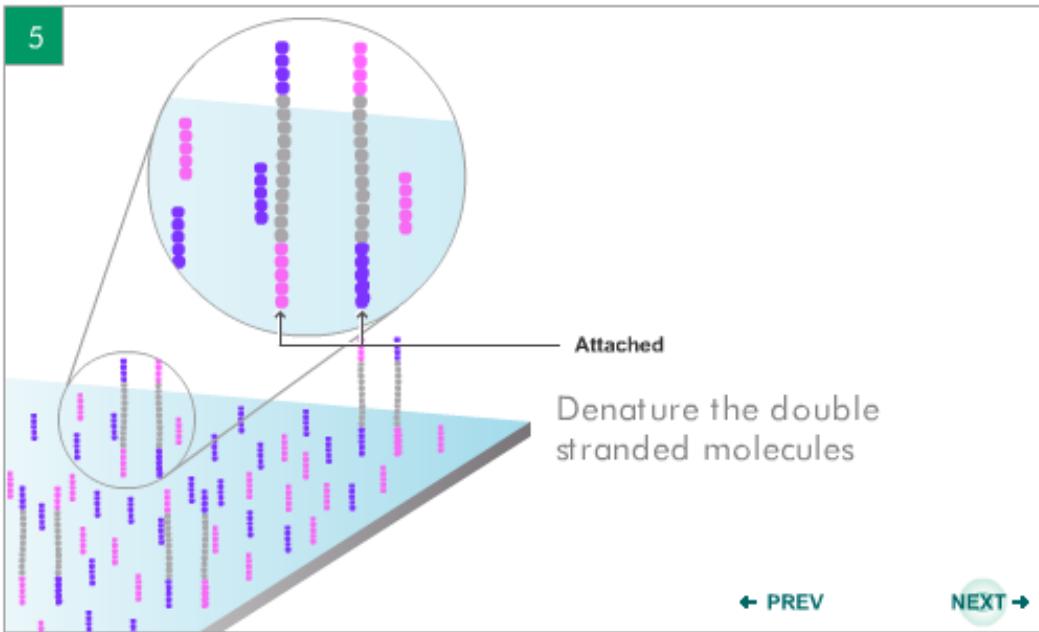
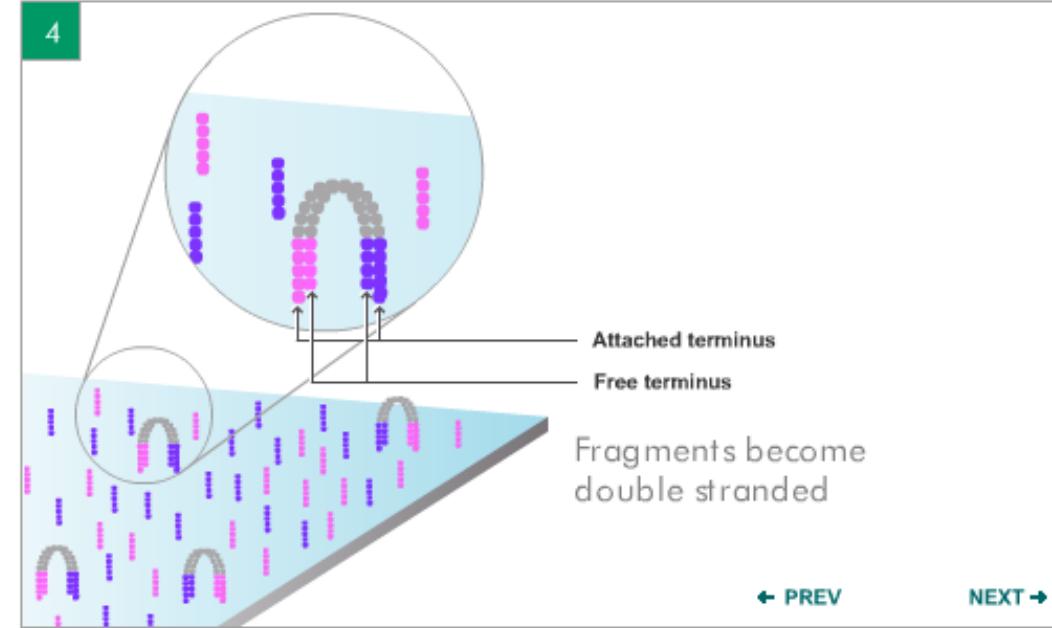
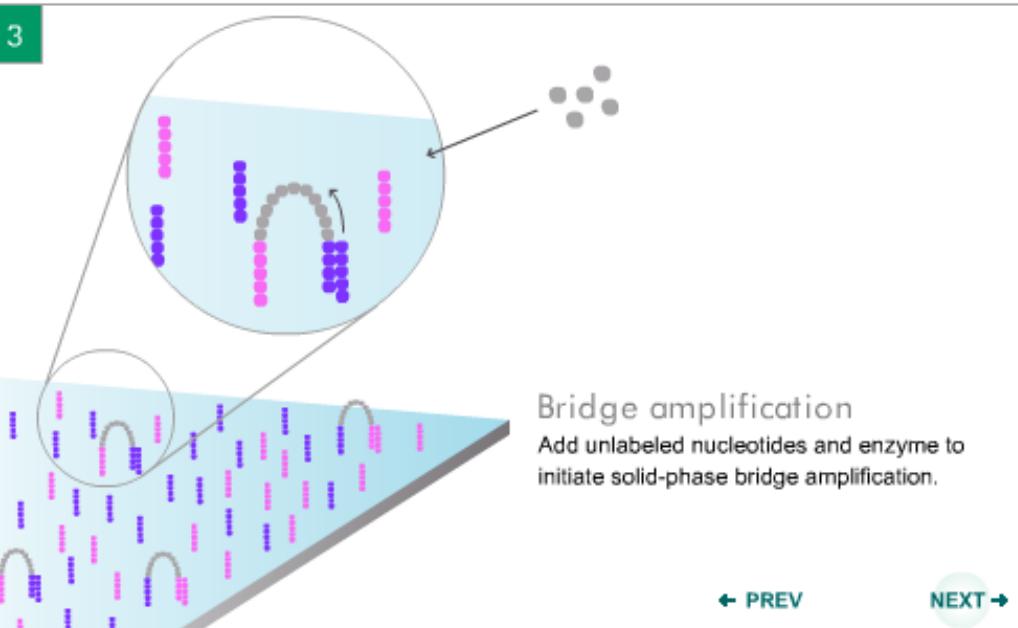
2



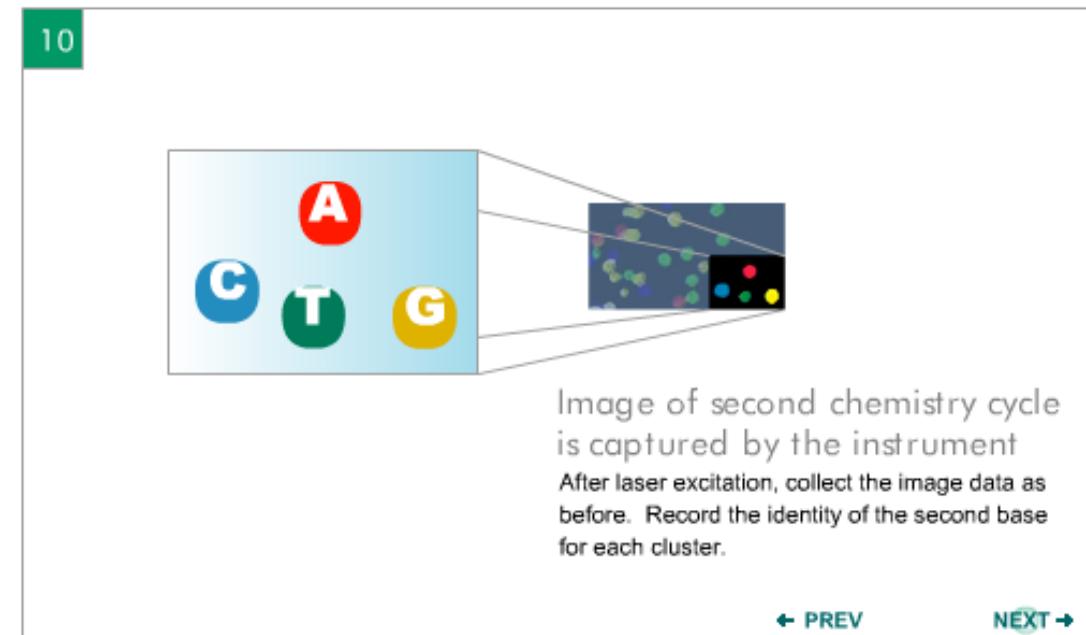
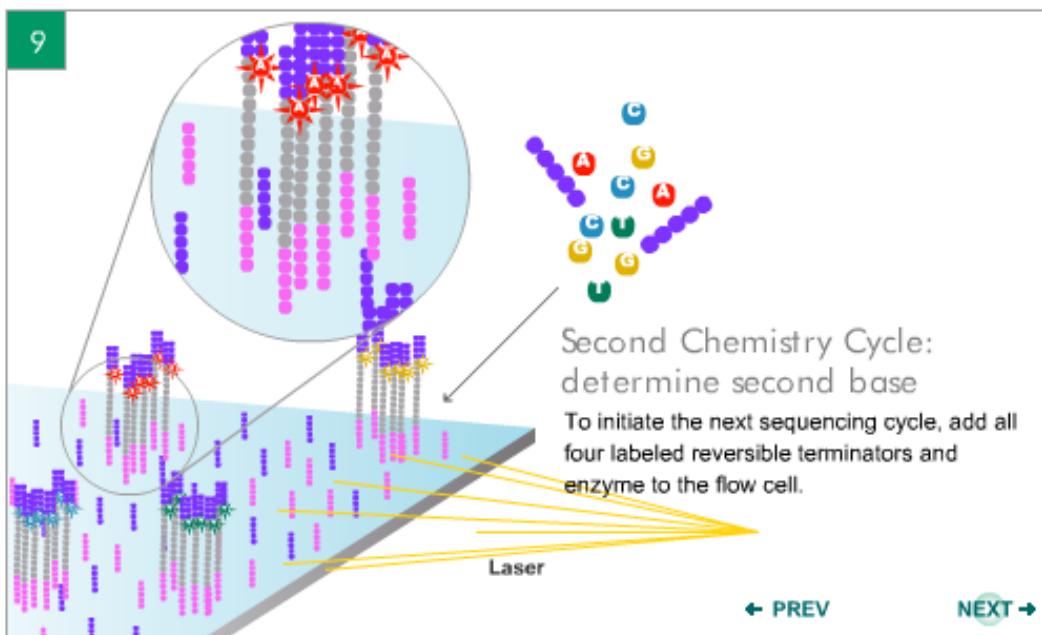
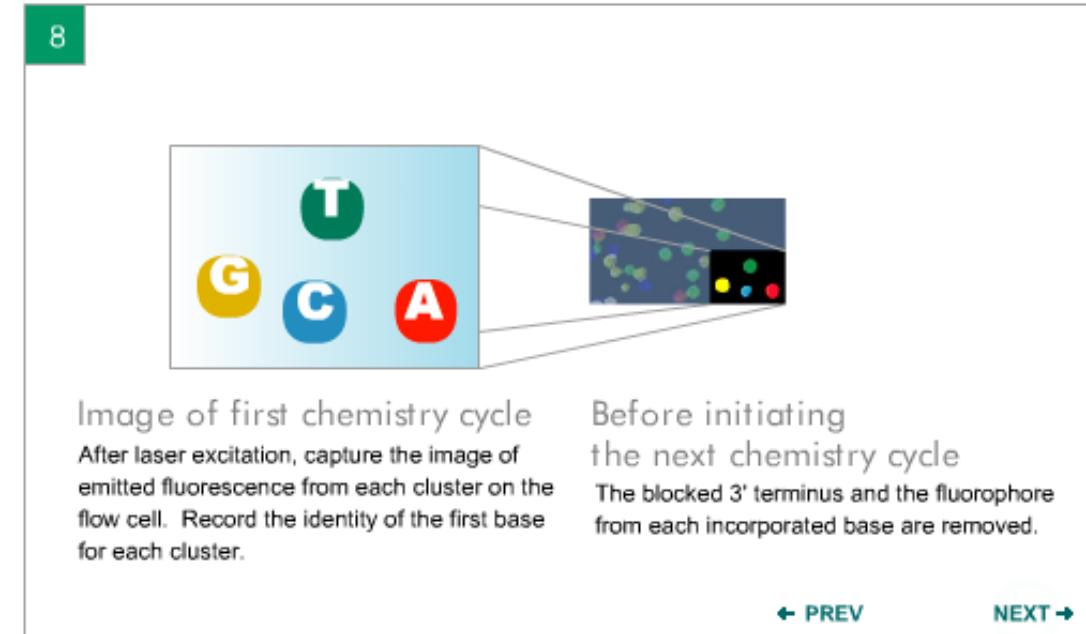
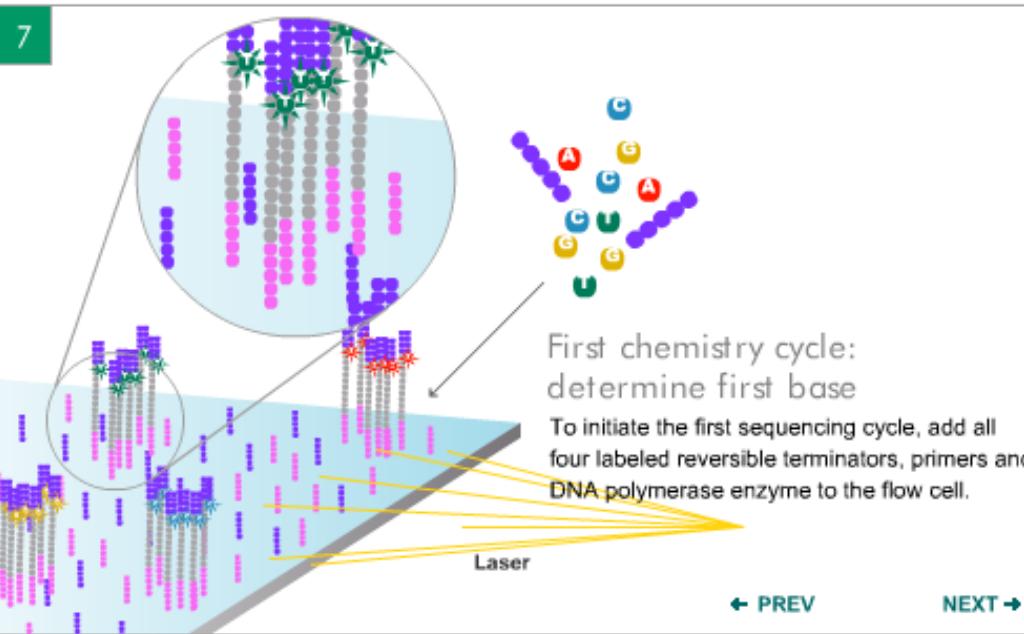
Add sample to flow cell



Massive parallel sequencing II

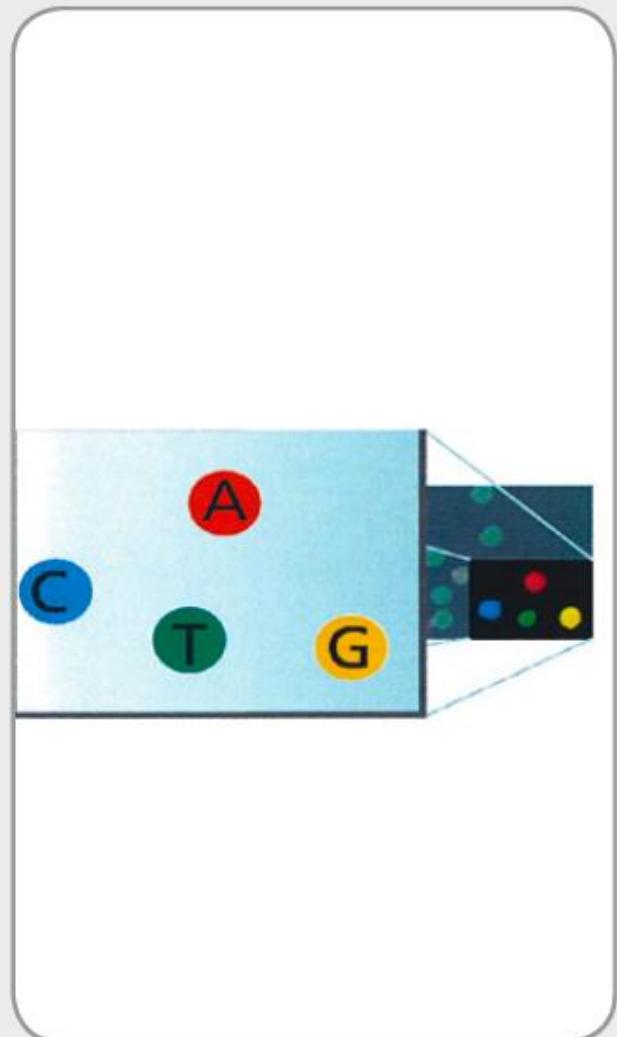


Massive parallel sequencing III



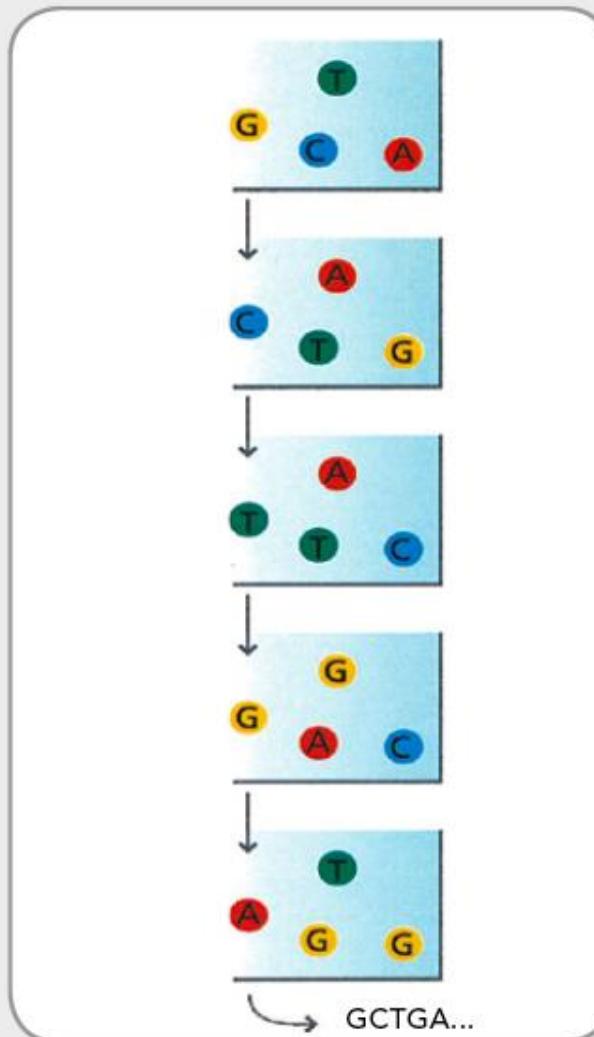
Massive parallel sequencing IV

10. IMAGE SECOND CHEMISTRY CYCLE



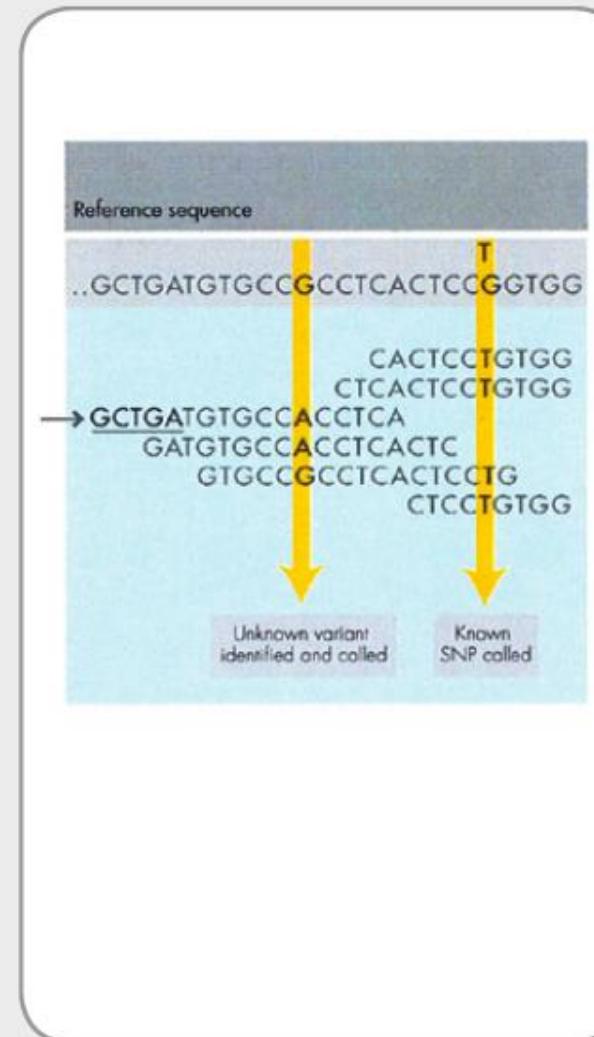
After laser excitation, collect the image data as before. Record the identity of the second base for each cluster.

11. SEQUENCE READS OVER MULTIPLE CHEMISTRY CYCLES



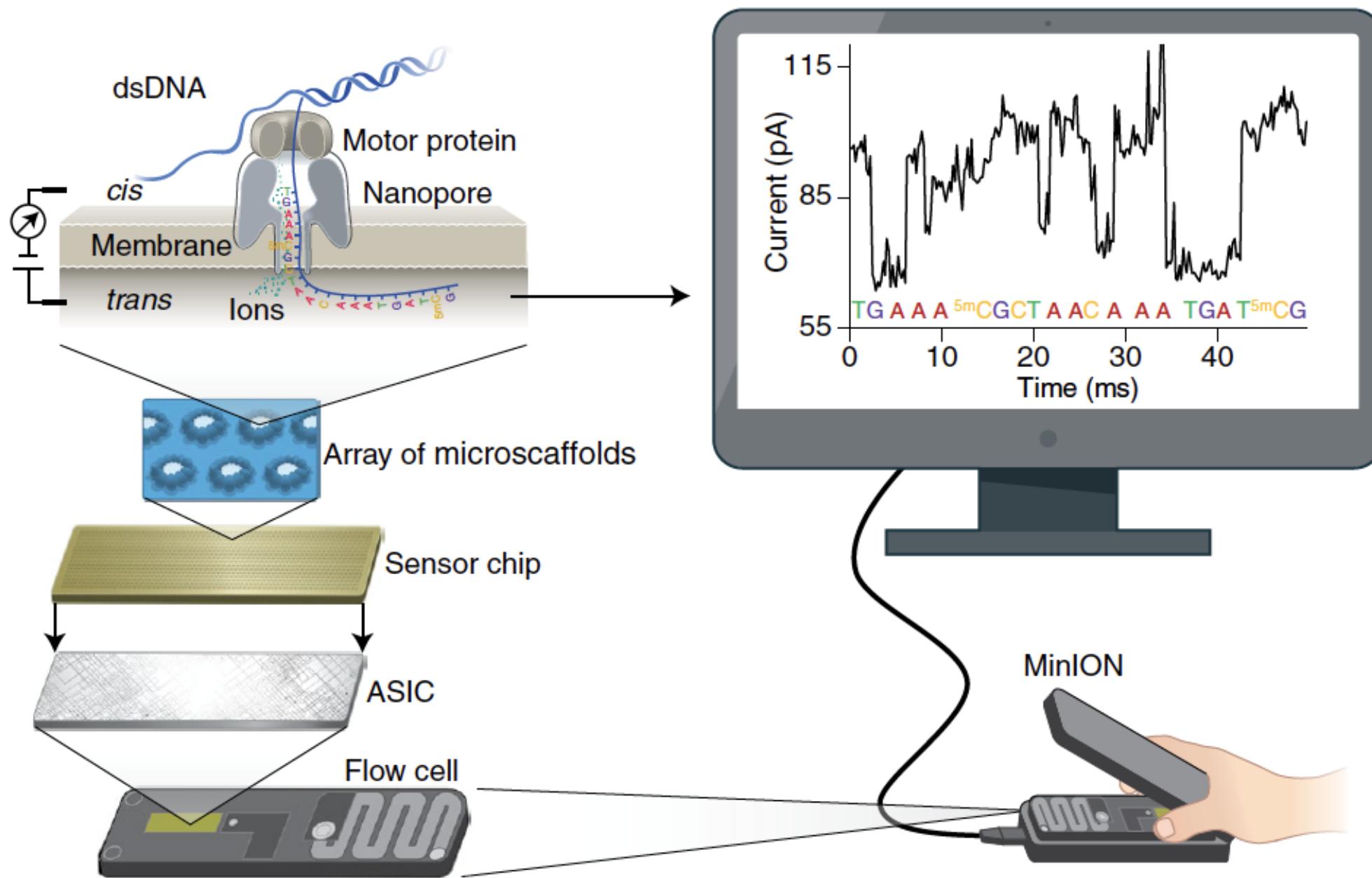
Repeat cycles of sequencing to determine the sequence of bases in a given fragment a single base at time.

12. ALIGN DATA

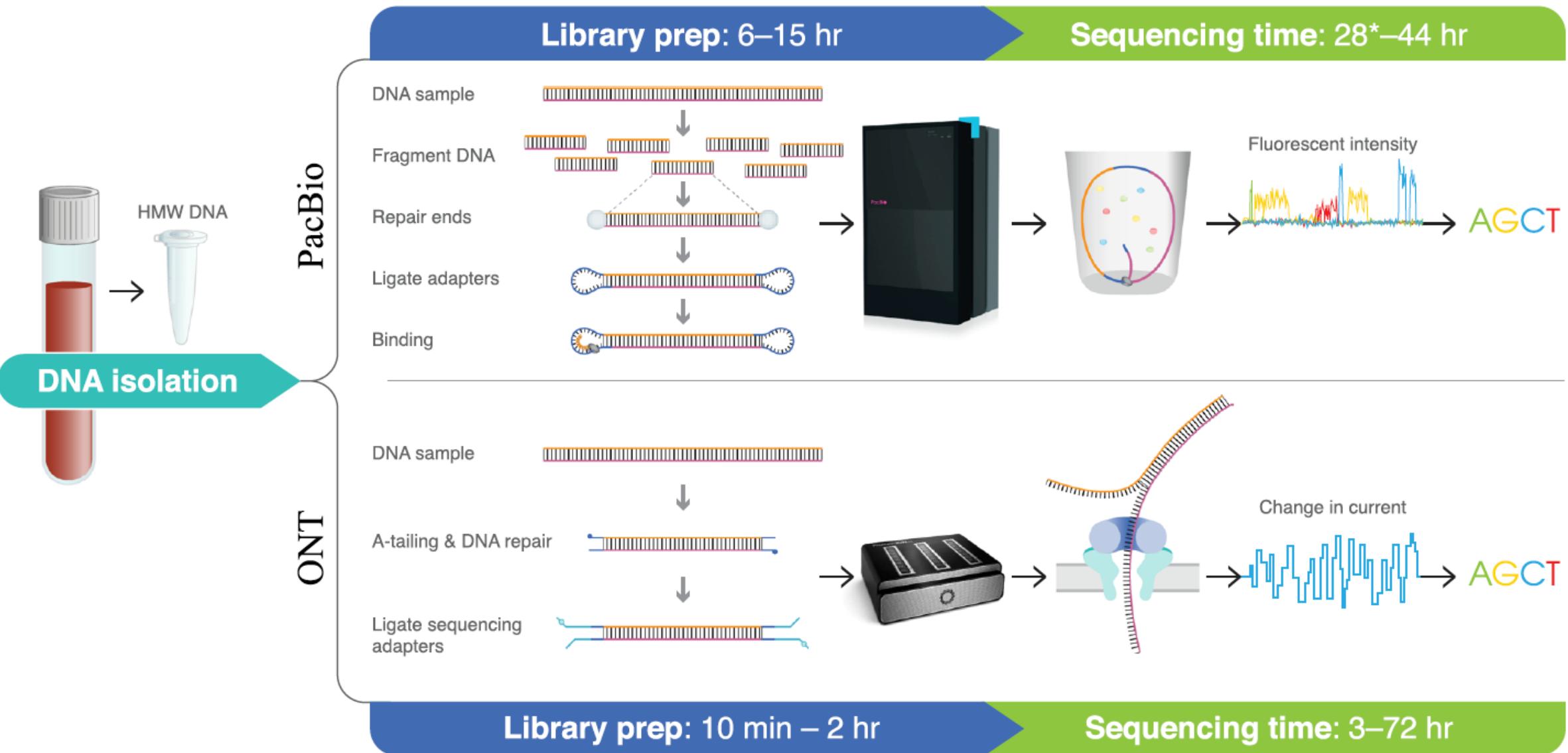


Align data, compare to a reference, and identify sequence differences.

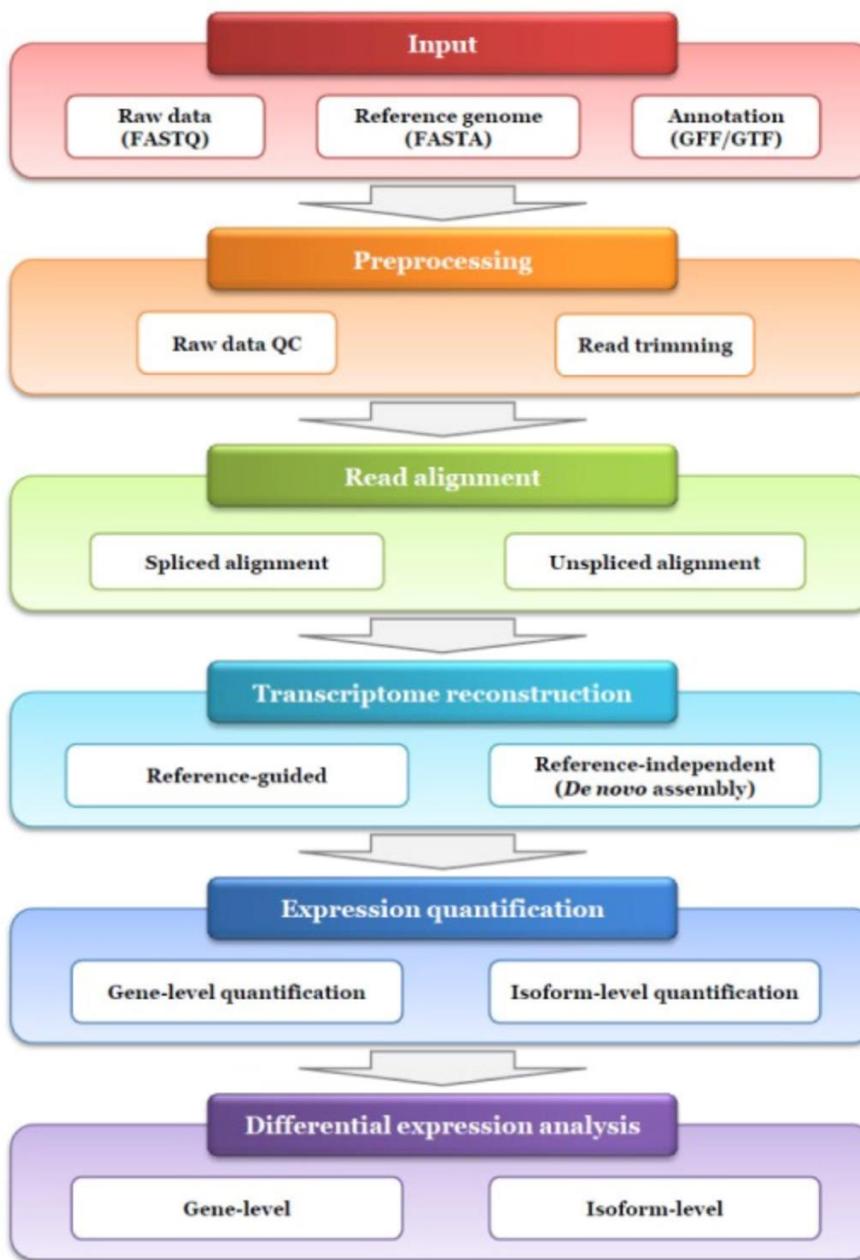
Principle of nanopore sequencing



Library preparation and sequencing workflow for both PacBio and ONT



Workflow of RNA-seq analysis



Workflow	Category	Package
Preprocessing of raw data	Raw data QC	FastQC HTQC
	Read trimming	FASTX-Toolkit FLEXBAR
Read alignment	Unspliced aligner	MAQ BWA Bowtie
	Spliced aligner	TopHat MapSplice STAR GSNAP
RNA-seq specific quality control		RNA-SeQC RSeQC
Transcriptome reconstruction		Qualimap 2 Cufflinks Scripture StringTie
Reference-independent		Trinity Oases transABYSS
Expression quantification		ALEXA-seq Enhanced read analysis of gene expression (ERANGE)
Isoform-level quantification		Normalization by expected uniquely mappable area (NEUMA) Cufflinks StringTie RSEM
Differential expression		Sailfish NOIseq edgeR DESeq
Gene-level		SAMseq Cuffdiff EBSeq
Isoform-level		Ballgown

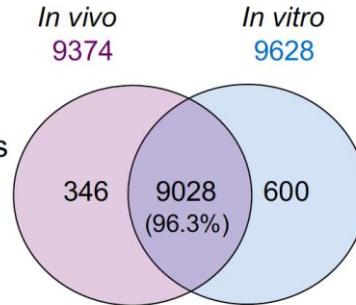


Example: Comparing *in vivo* and *in vitro* datasets of the VitDHiD study

25 individuals

A

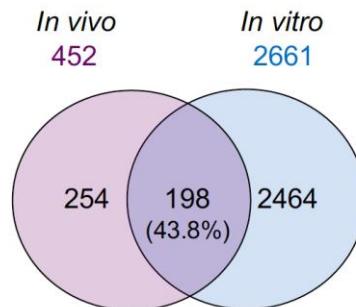
All commonly
expressed genes
(log CPM > 10)



14 individuals

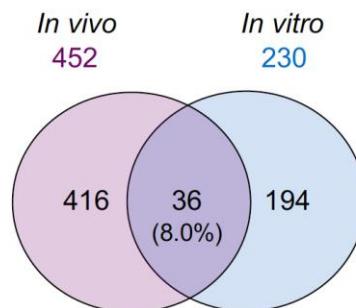
B

log CPM > 10
FDR < 0.05;



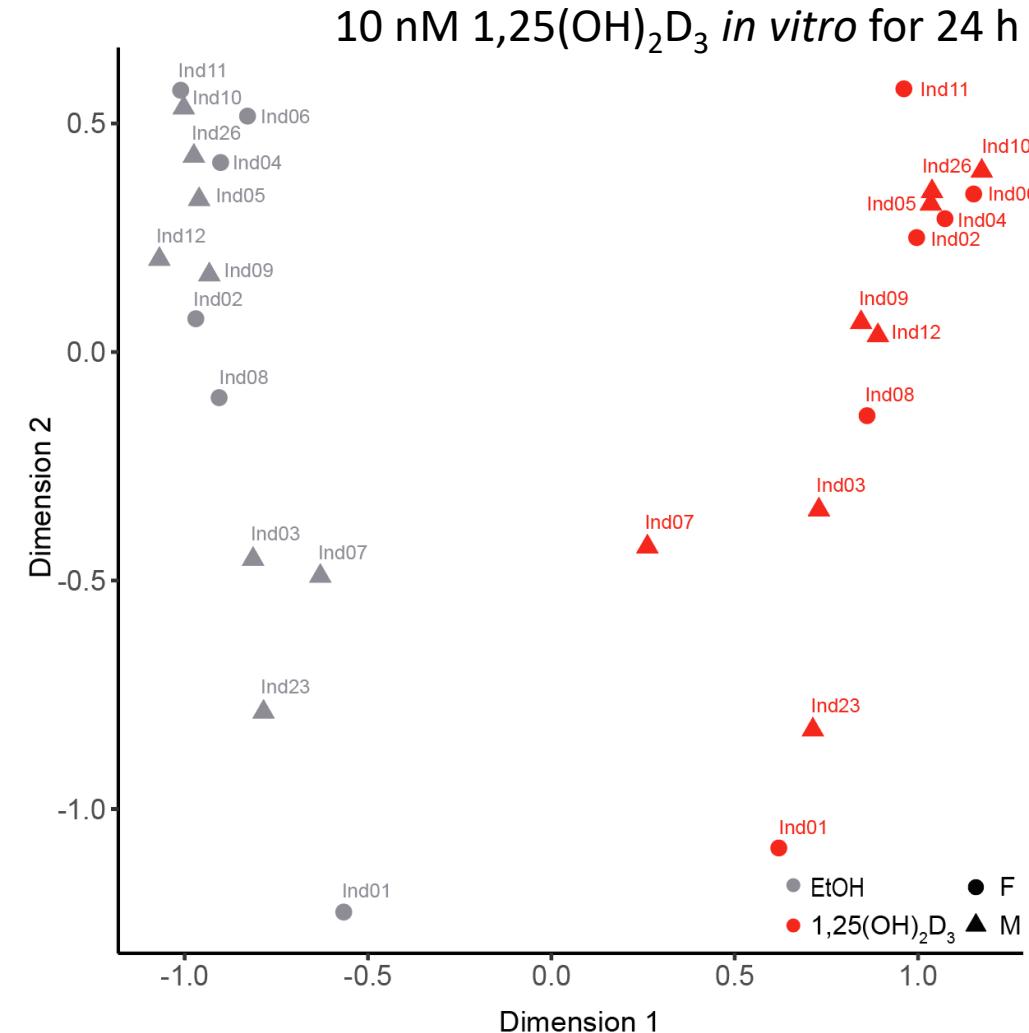
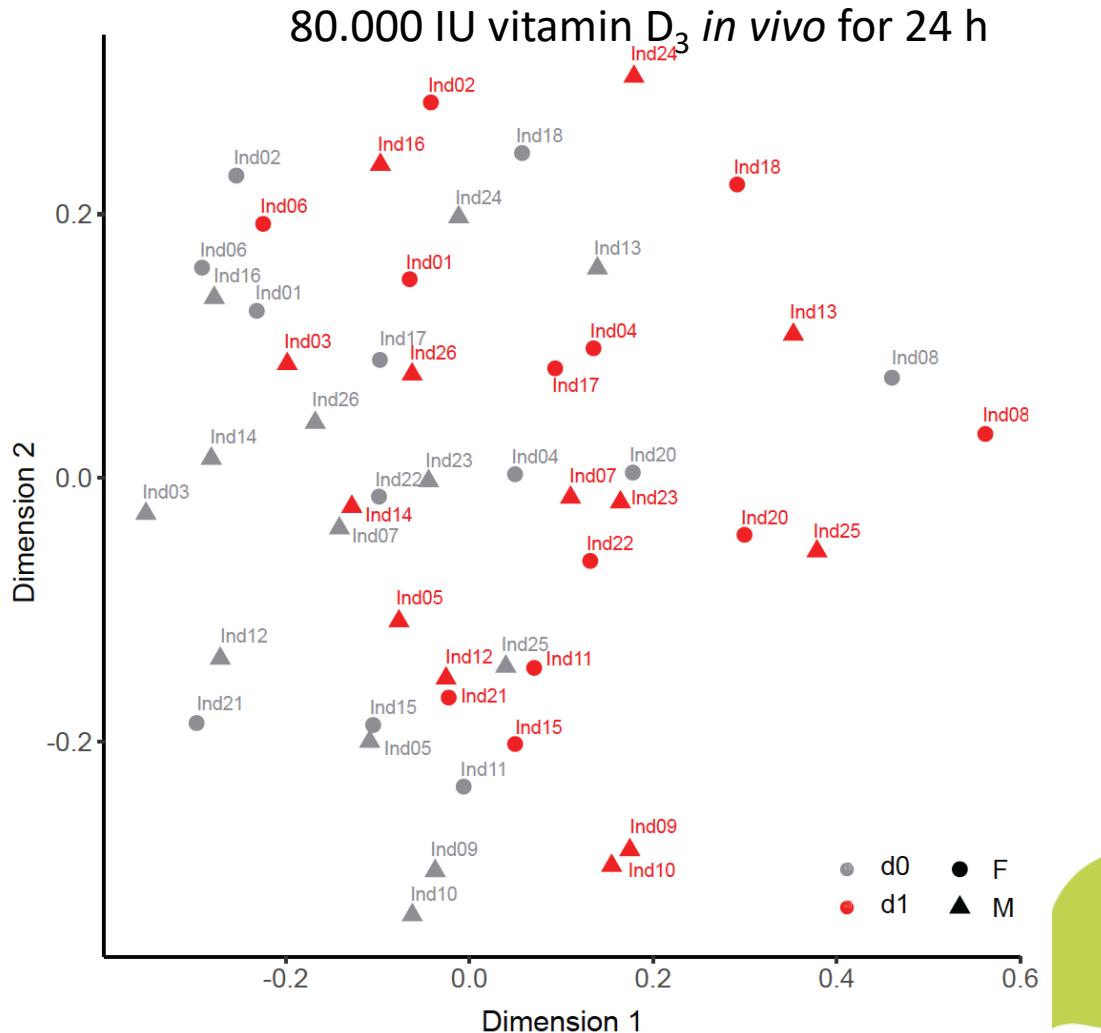
C

log CPM > 10
FDR < 0.05;
logFC > 1
(*in vitro* only)

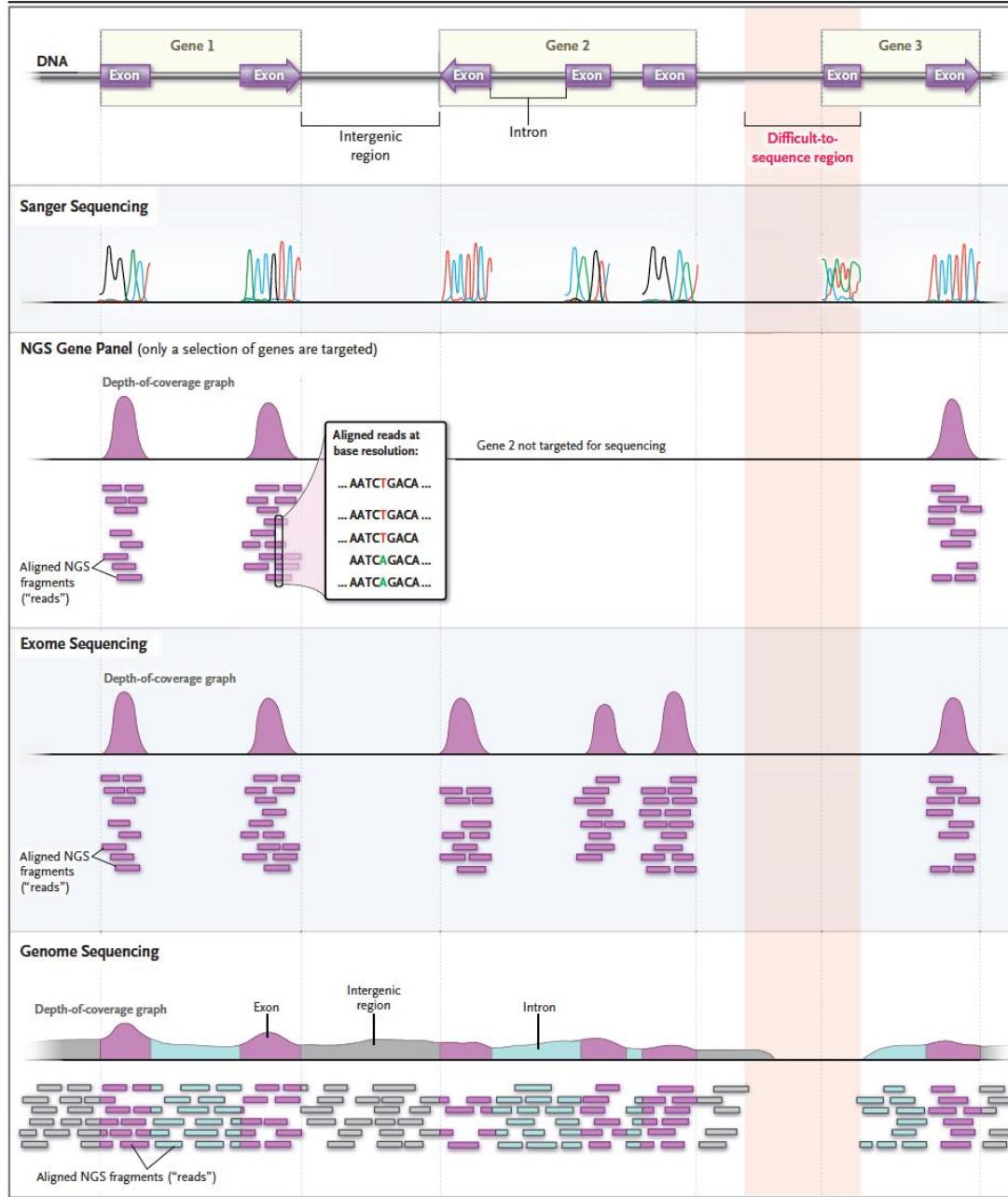




Bioinformatician's view: Principal component analysis



(Clinical) next-generation sequencing (NGS) types



UCSC Genome Browser on Human (GRCh38/hg38)

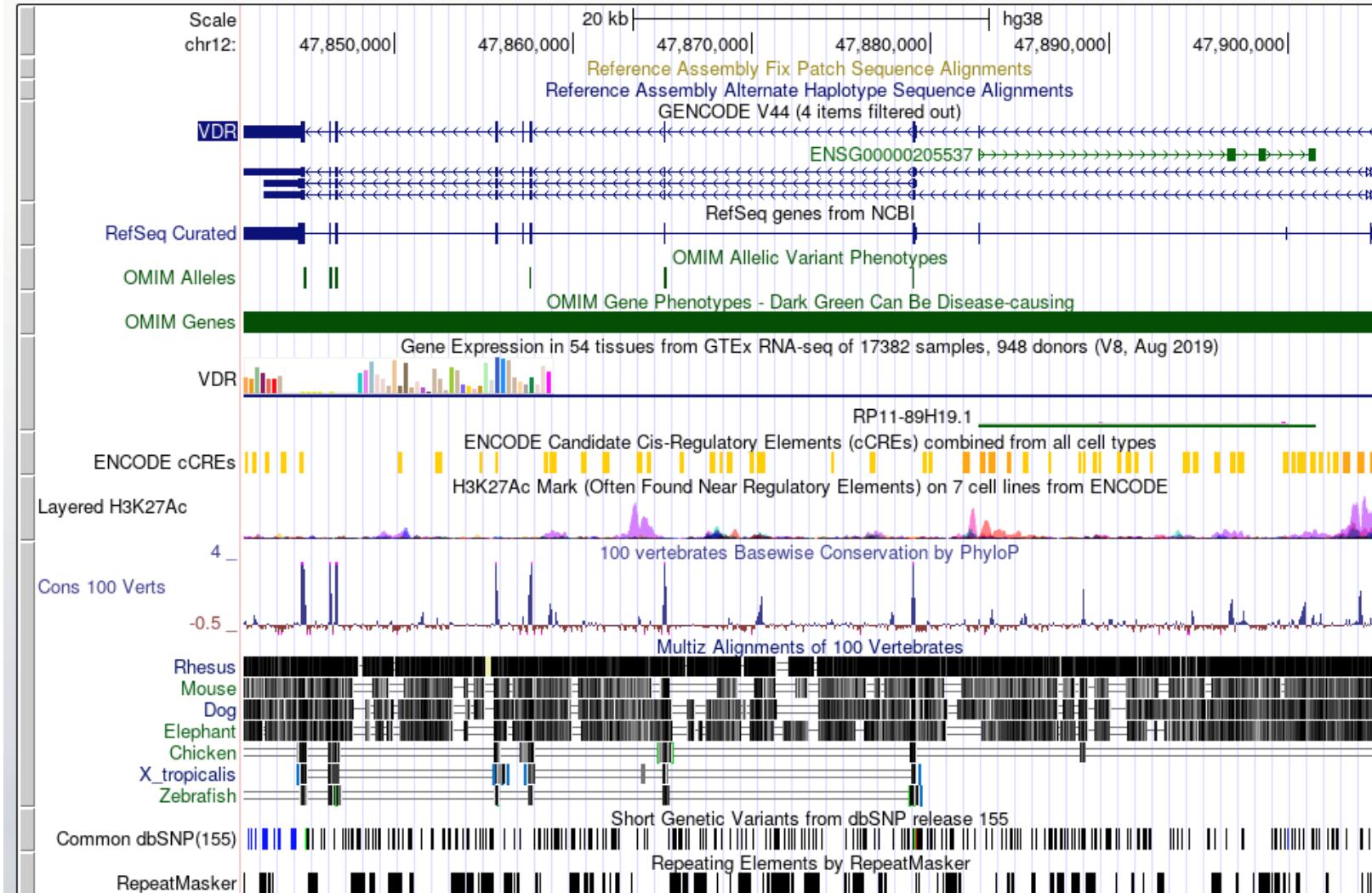
move [<<< << < > >>] zoom in [1.5x 3x 10x base] zoom out [1.5x 3x 10x 100x]

A guided tutorial is available for new users: [Start tutorial](#)

[Close](#) [Don't show again](#)

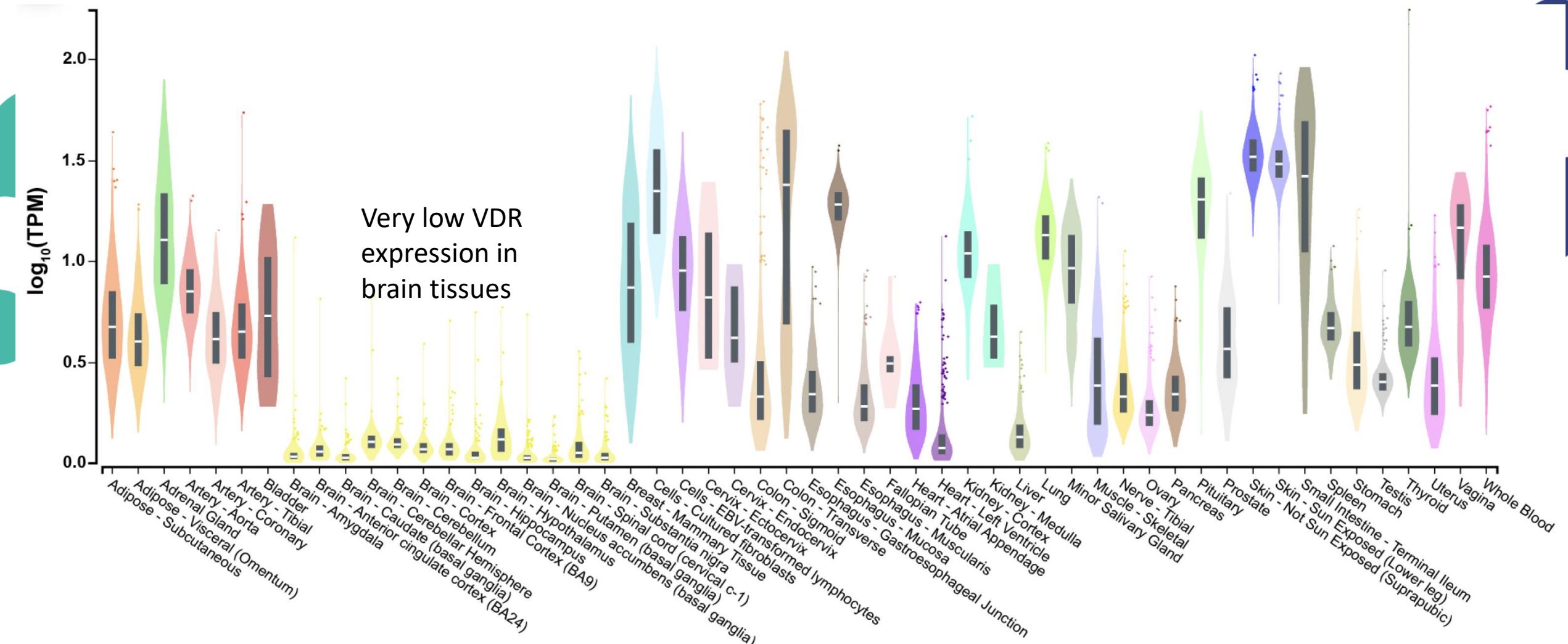
multi-region **chr12:47,841,537-47,904,994** 63,458 bp. [gene](#), [chromosome range](#), [search terms](#), [help pages](#), [see examples](#) go [Examples](#)

chr12 (q13.11) 12.3 12.1 12q12 14.1 15 21.31 22 23.1 23.3





Rather ubiquitous VDR expression: 54 primary human tissues



Based on 948 post-mortem donors

Carlberg, Nutrients 2023



NUTRIGENOMICS
ERA CHAIR | WELCOME2

Carsten Carlberg c.carlberg@pan.olsztyn.pl



Institute
of Animal Reproduction and Food Research
Polish Academy of Sciences
in Olsztyn



This project has received funding from the
European Union's Horizon 2020 research
and innovation programme under grant
agreement No 952601

Profiling transcription

