Analyzing structural variation in breast cancer with long read sequencing of patient-derived organoids

Rachel M Sherman

Johns Hopkins University

Schatz and Salzberg Labs

Biological Data Science 2018

Structural variation characterizes cancer

Prognostic indicator

Greater instability leads to worse patient outcomes

Copy number changes

Amplifications and deletions, especially of oncogenes and tumor suppressors

Gene fusions

Can drastically alter gene expression levels

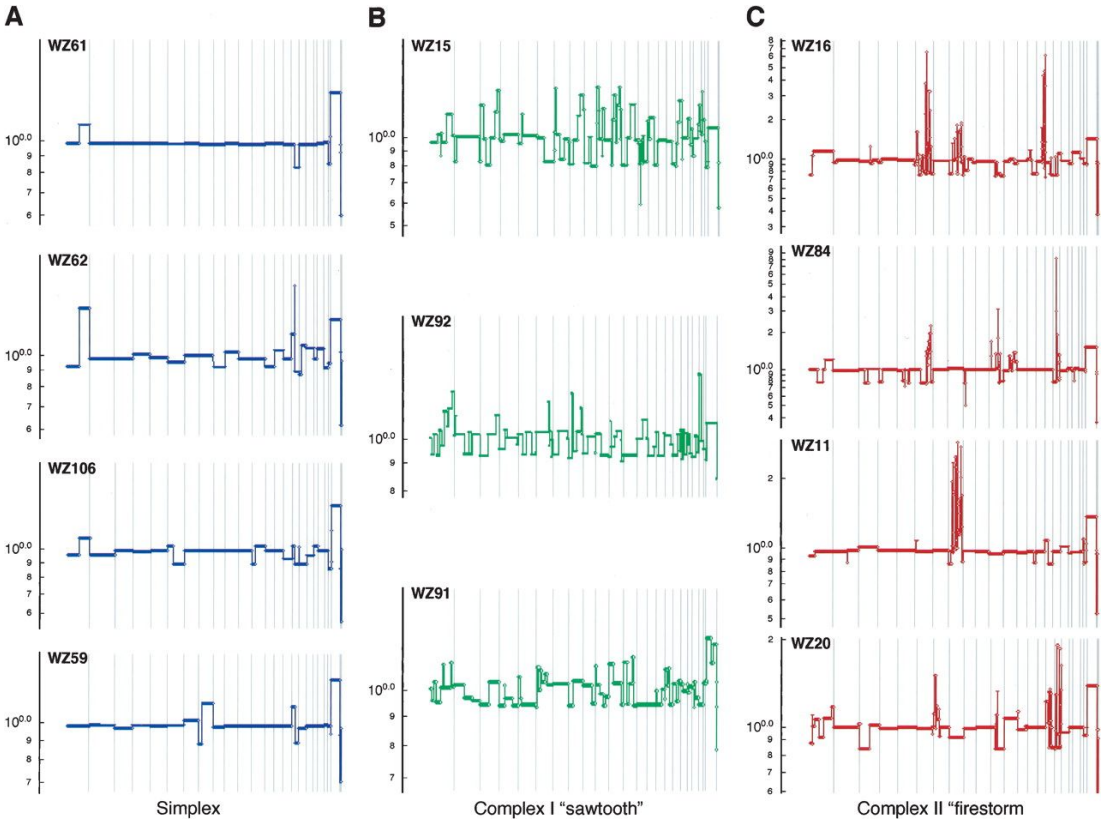


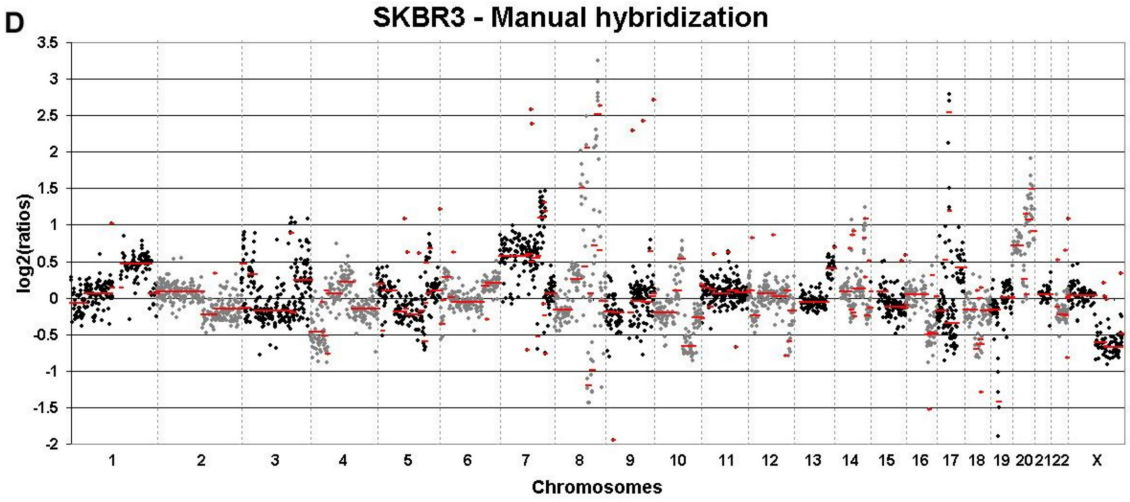
Figure 2. Major types of tumor genomic profiles. Segmentation profiles for individual tumors representing each category: (*A*) simplex; (*B*) complex type I or sawtooth; (*C*) complex type II or firestorm. Scored events consist of a minimum of six consecutive probes in the same state. The *y*-axis displays the geometric mean value of two experiments on a log scale. Note that the scale of the amplifications in *C* is compressed relative to *A* and *B* owing to the high levels of amplification in firestorms. Chromosomes 1–22 plus X and Y are displayed in order from *left* to *right* according to probe position.

(Hicks *et al*, 2006, Genome Research)

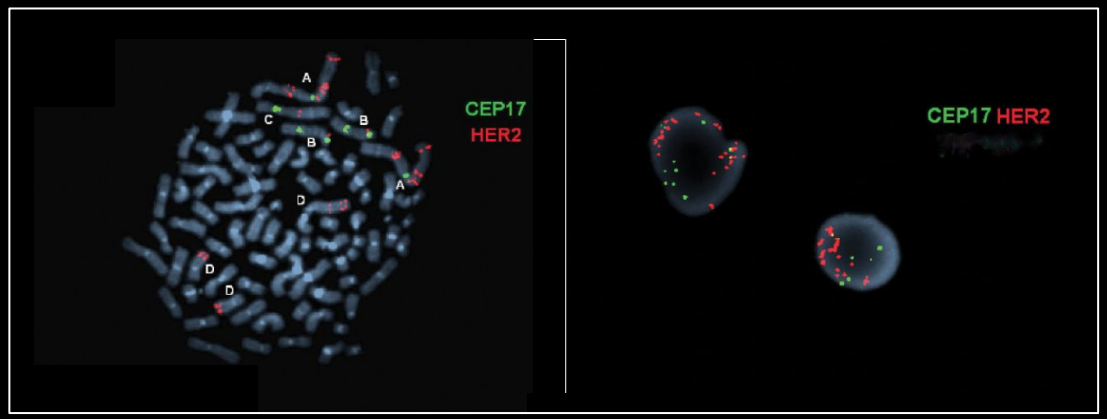
SK-BR-3 Her2 amplified cell line

Derived in 1970, most commonly used Her2 amplified cell line

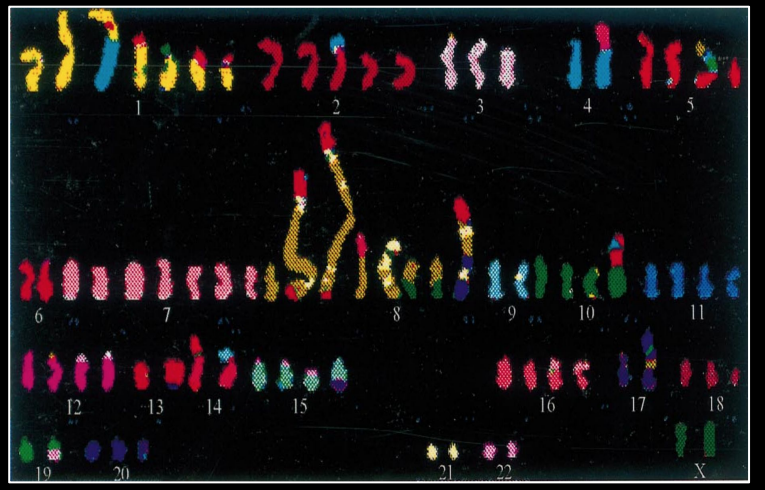
FISH, aCGH, and karyotyping show large-scale changes



Joosse, Beers, & Nederlof, 2007



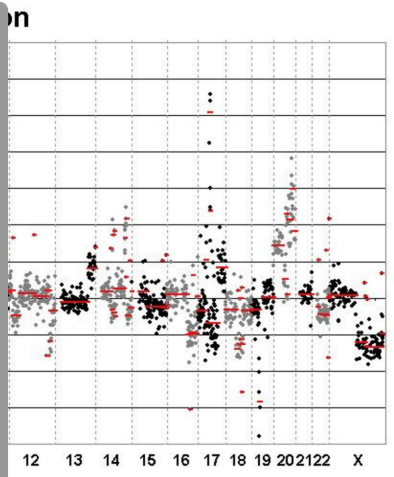
Mosoyan *et al*, 2013



Davidson *et al*, 2000

SK-BR-3 Her2 amplified cell line

Despite importance of structural variation, relatively little is known except large copy number variants



eers, & Nederlof, 2007

Mosyan *et al*, 2013

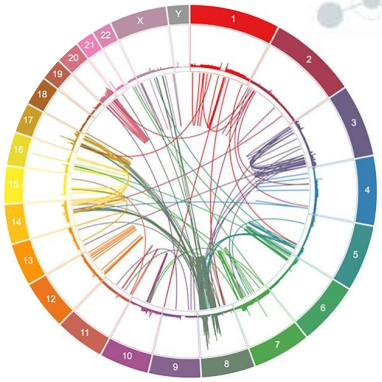
Davidson *et al*, 2000

Analyzing variation with long reads

Part 1 (SK-BR-3)

How prevalent are mid-sized variants?

How necessary are long reads?



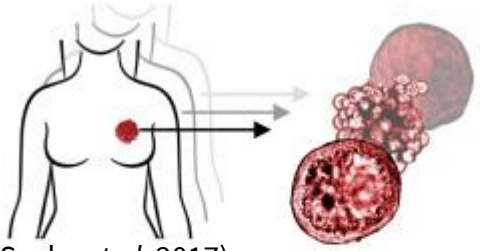
(Nattestad *et al,* 2018)

Part 2 (Patient-derived organoids)

What new variants have we found?

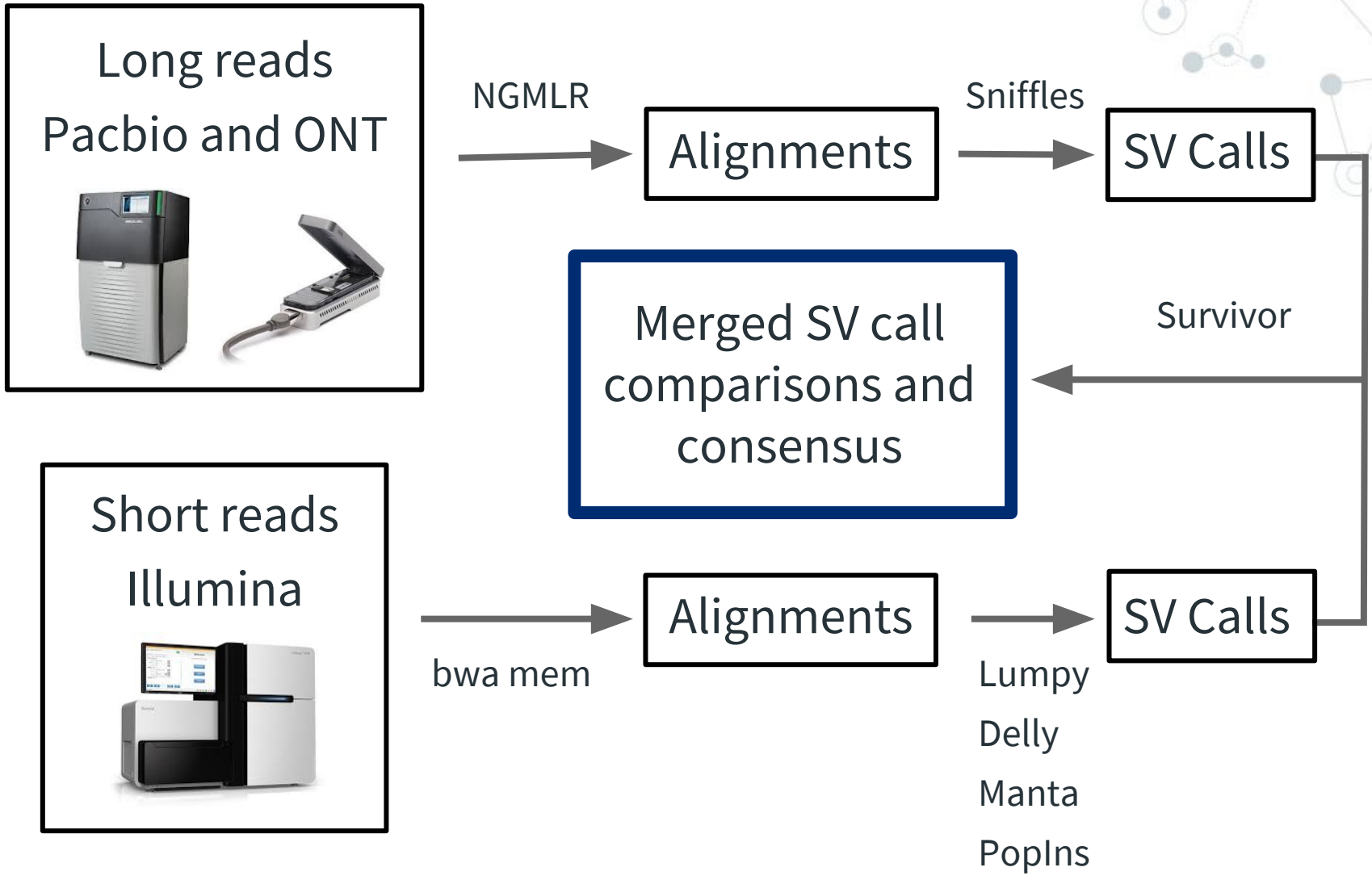
How do they compare to a cell line (SK-BR-3)?

What else can long reads tell us?



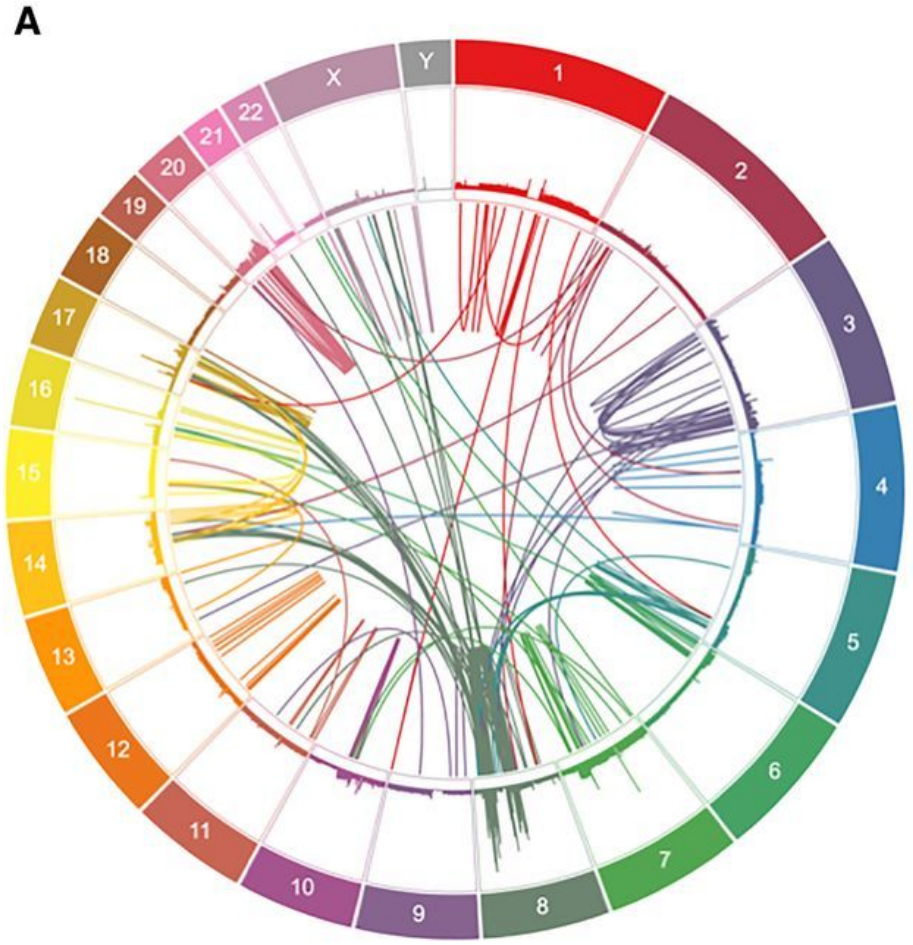
(Sachs *et al,* 2017)

SK-BR-3 sequencing and analysis

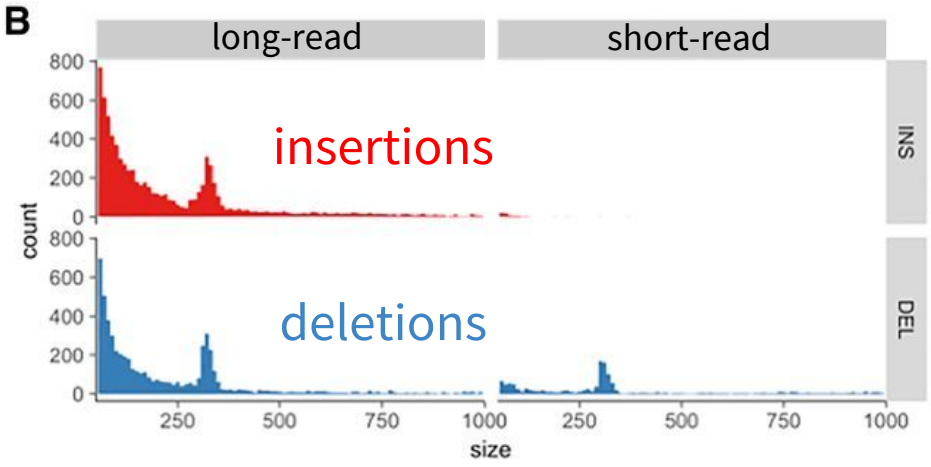


(Nattestad *et al*, Genome Research, 2018)

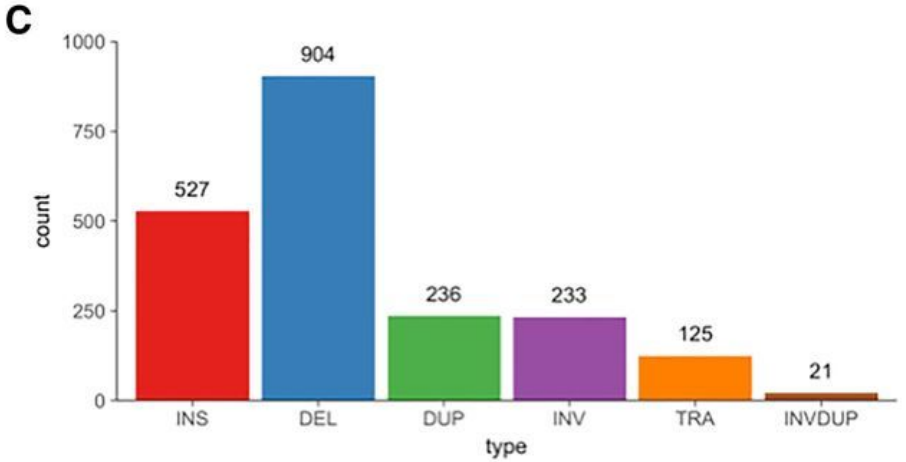
Inter-chromosomal or over 10 kb



50 bp to 1 kb insertions and deletions



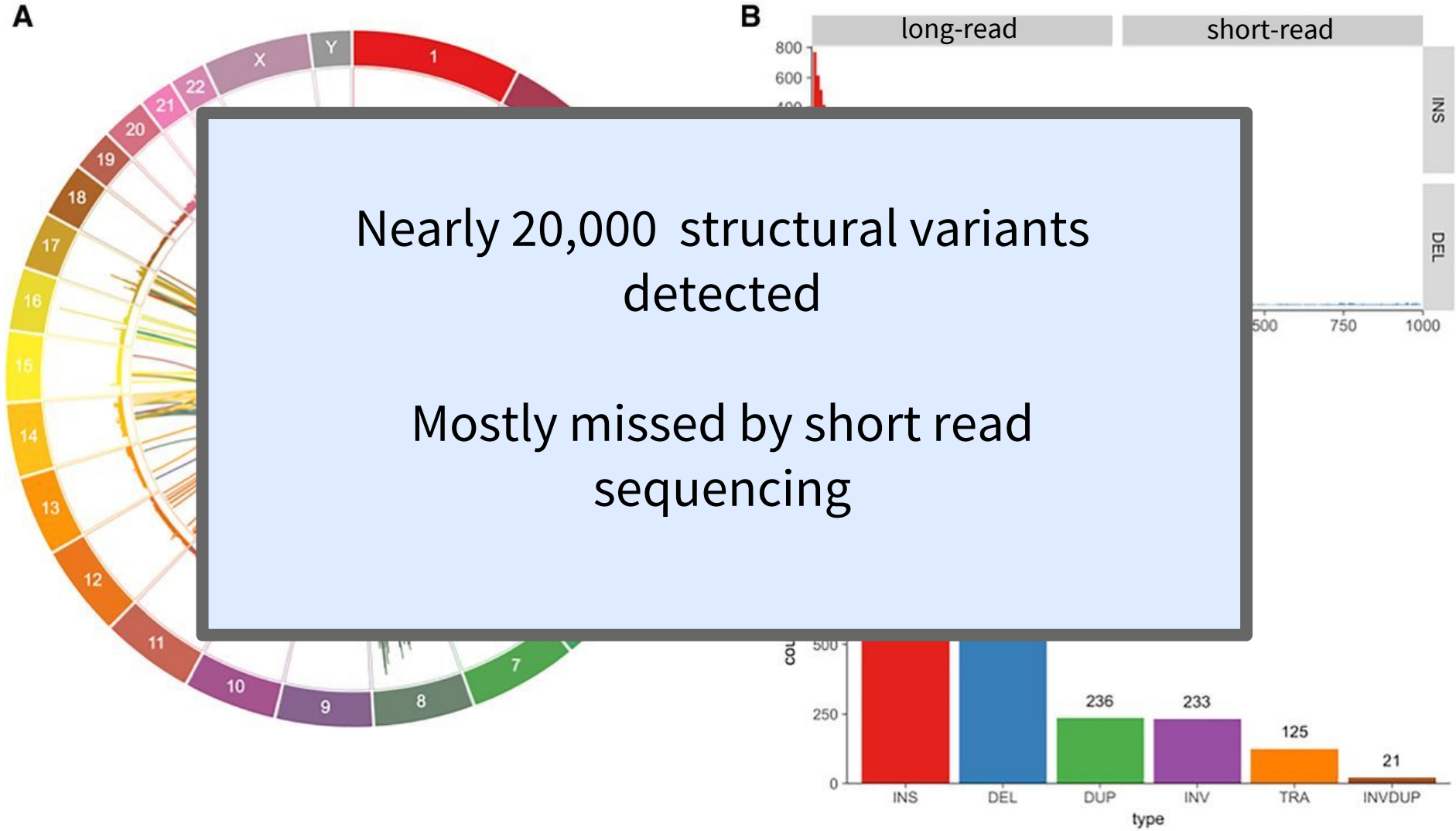
Variant counts over 1 kb



Structural variants detected with PacBio in SKBR3 (Nattestad *et al*, 2018).

Inter-chromosomal or over 10 kb

50 bp to 1 kb insertions and deletions

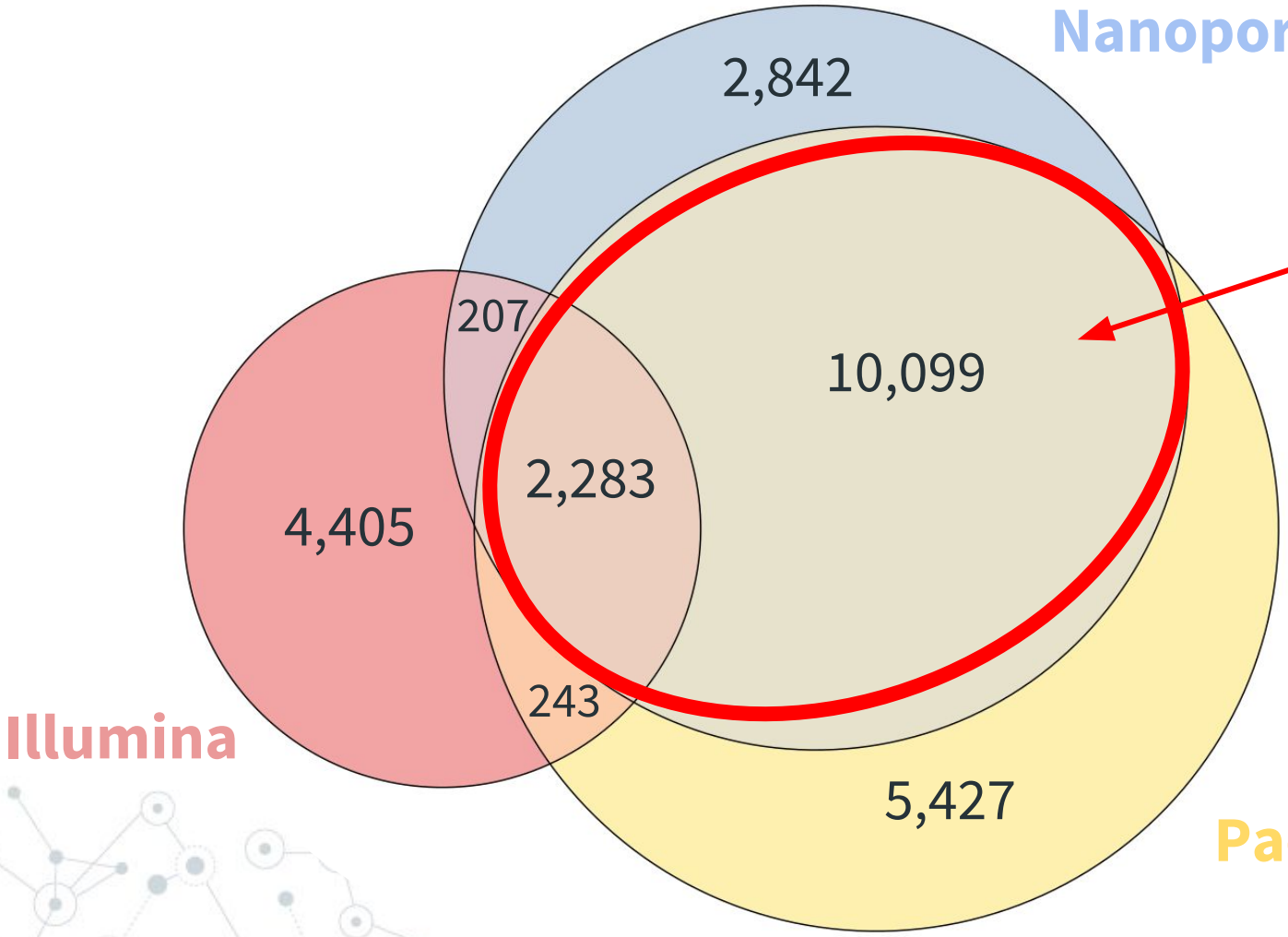


Nearly 20,000 structural variants detected

Mostly missed by short read sequencing

Structural variants detected with PacBio in SKBR3 (Nattestad *et al*, 2018).

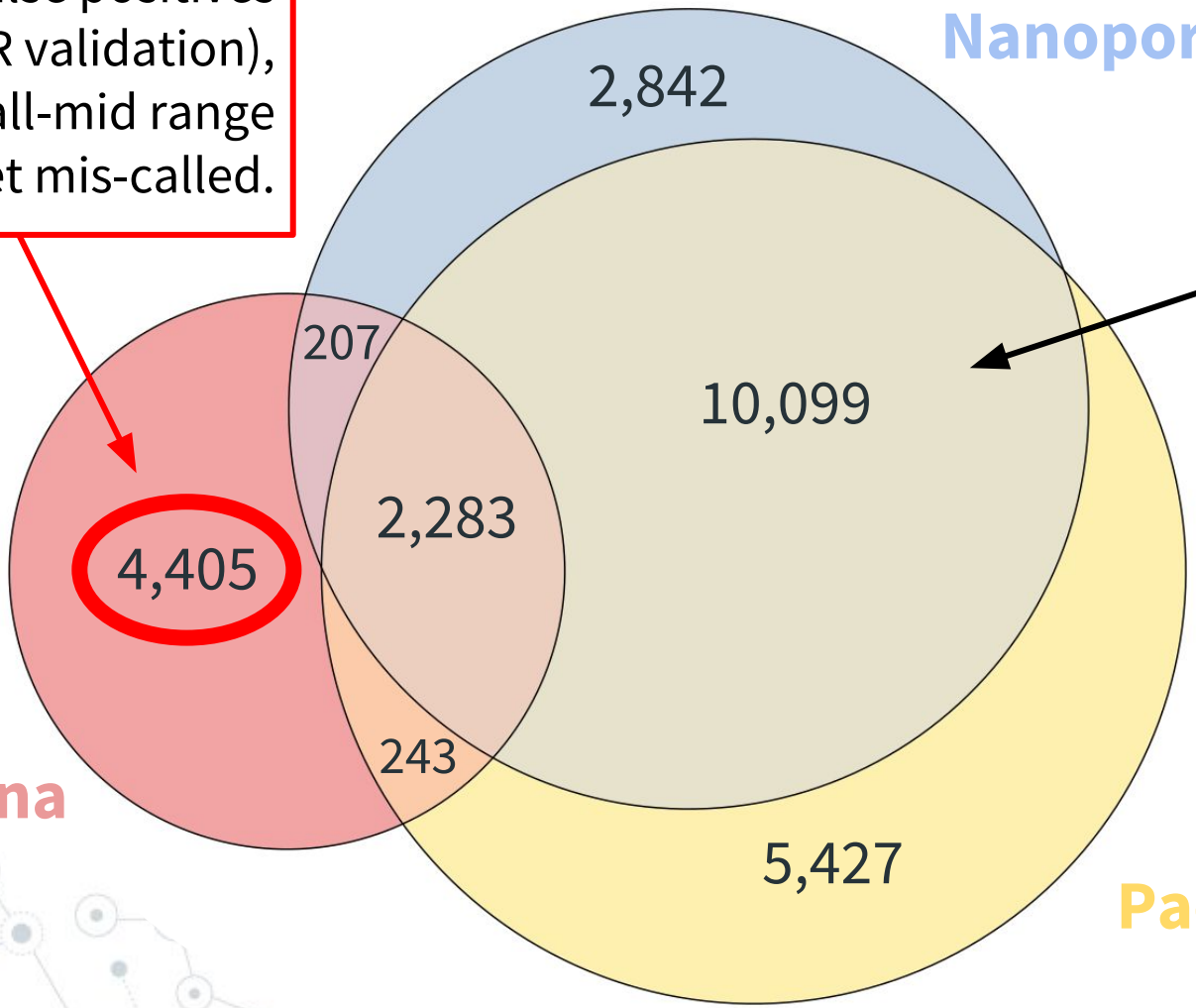
Variant calling across platforms



Good long read concordance and more SVs found with long reads. High rate of PCR validation.

Variant calling across platforms

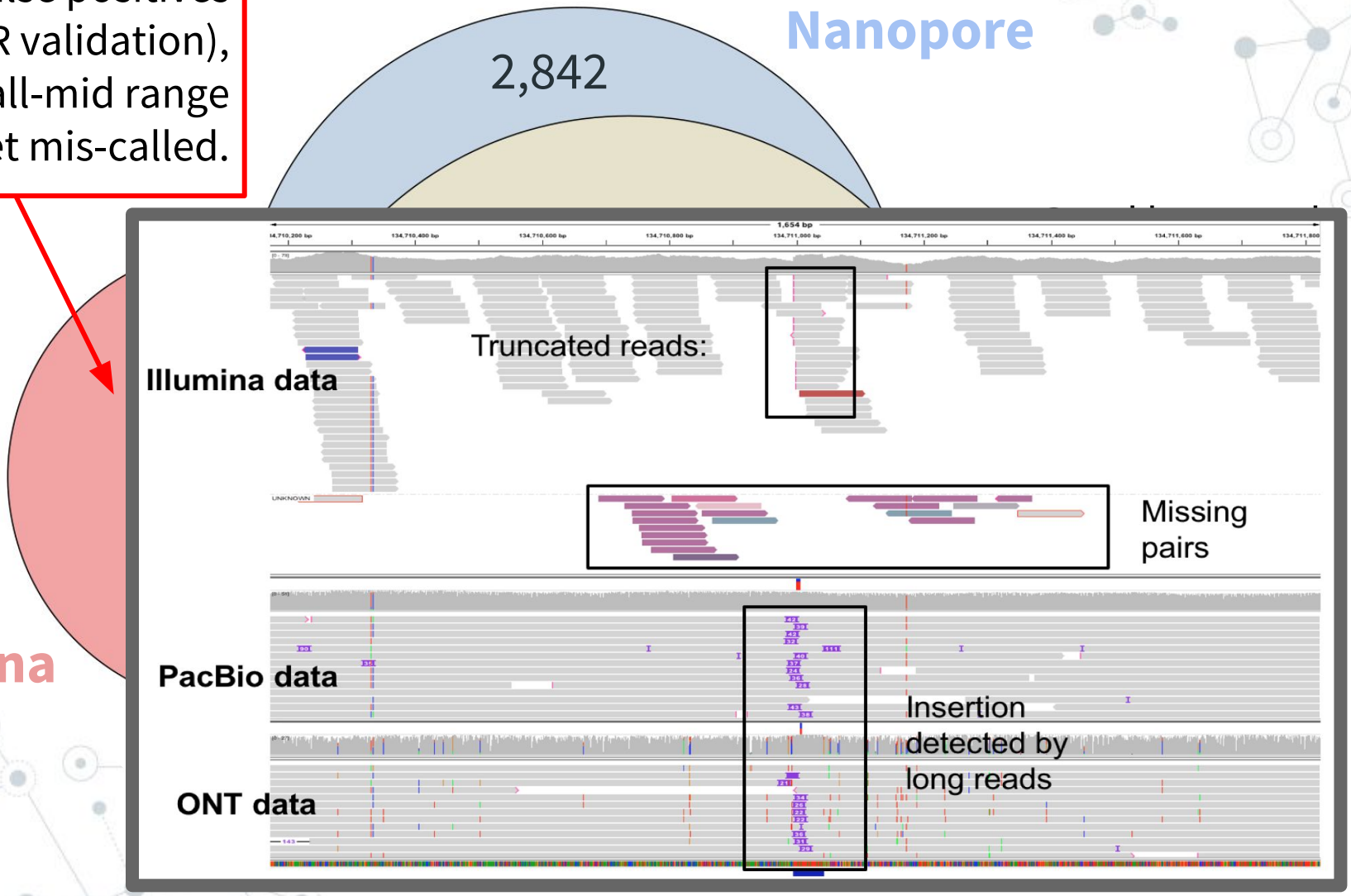
Mostly false positives (PCR validation), often small-mid range indels get mis-called.



Good long read concordance and more SVs found with long reads. High rate of PCR validation.

Variant calling across platforms

Mostly false positives (PCR validation), often small-mid range indels get mis-called.

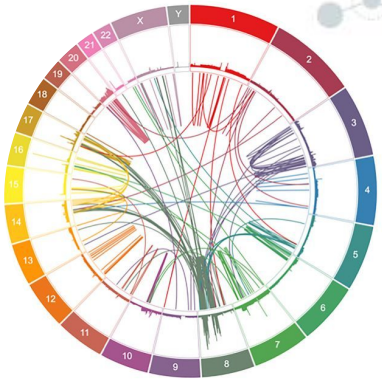


Analyzing variation with long reads

Part 1 (SK-BR-3)

How prevalent are mid-sized variants?

How necessary are long reads?

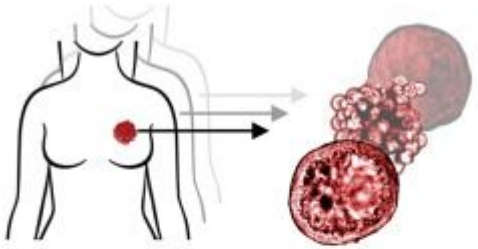


Part 2 (Patient-derived organoids)

What new variants have we found?

How do they compare to a cell line (SK-BR-3)?

What else can long reads tell us?

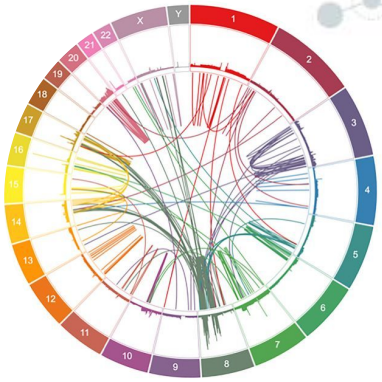


Analyzing variation with long reads

Part 1 (SK-BR-3)

How prevalent are mid-sized variants?

How necessary are long reads?

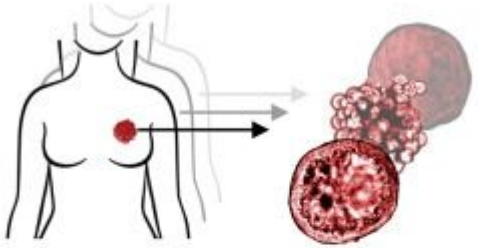


Part 2 (Patient-derived organoids)

What new variants have we found?

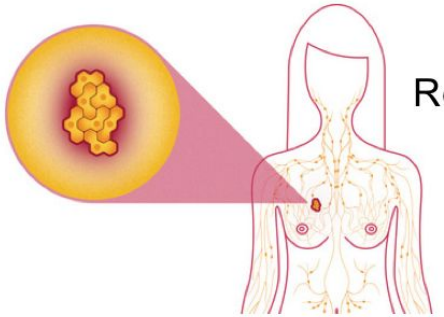
How do they compare to a cell line (SK-BR-3)?

What else can long reads tell us?



Patient-derived tumor organoids

Breast cancer patient tumor organoid, ER+, PR+, Her2-



Resected primary tumor

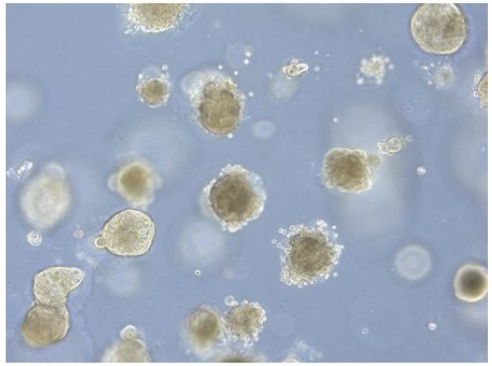
Mechanical/ enzymatic

Purification

Plating on Matrigel

Add growth factors

Tumor organoids in culture



- ✓ Stable Growth in 3D
- ✓ Recapitulate tumor pathology & treatment response
- ✓ Maintenance of tissue/tumor heterogeneity
- ✓ "2017 Method of the Year" - Nature Methods

CSH

David Spector

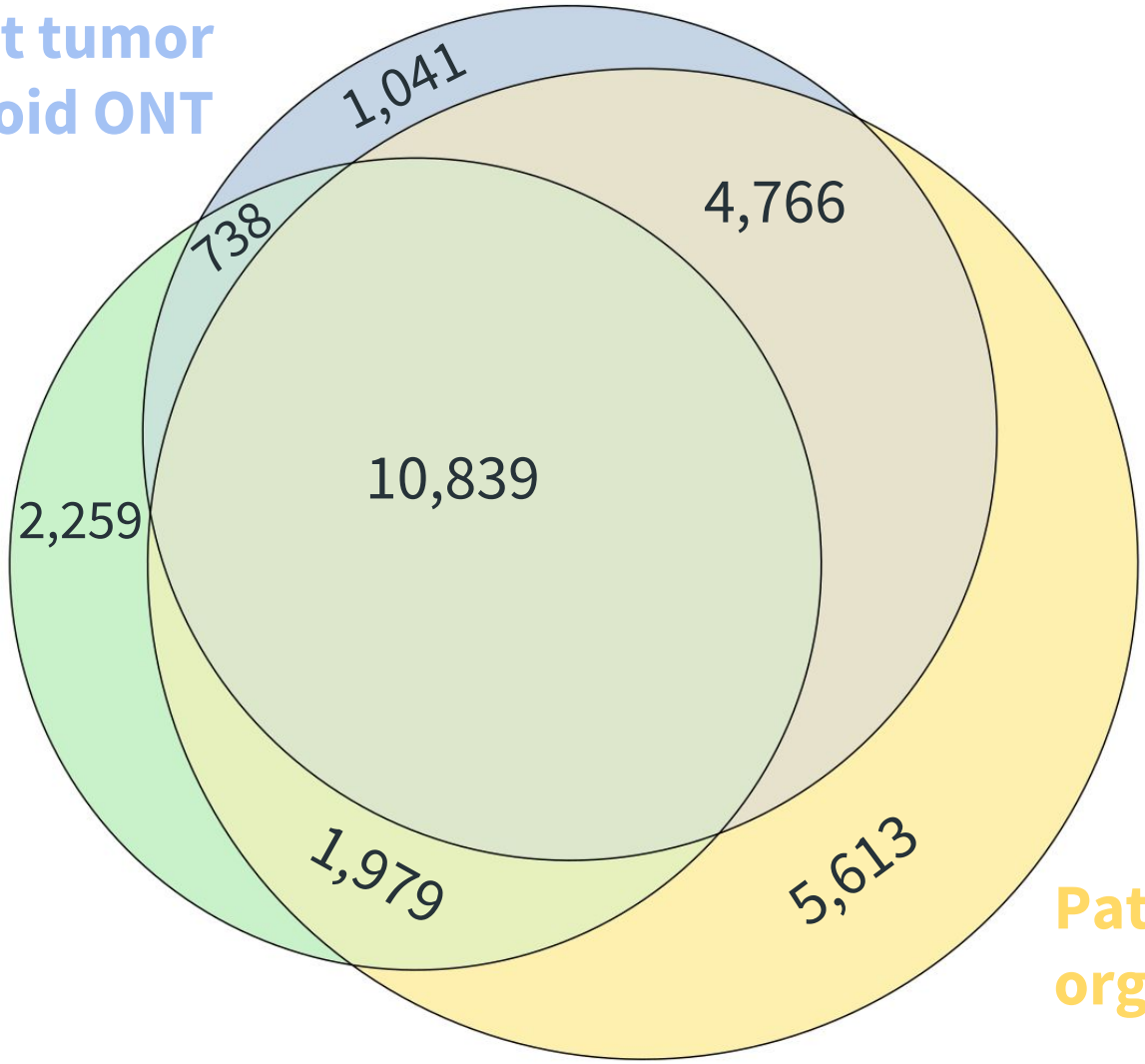
Northwell Health

Karen Kostroff

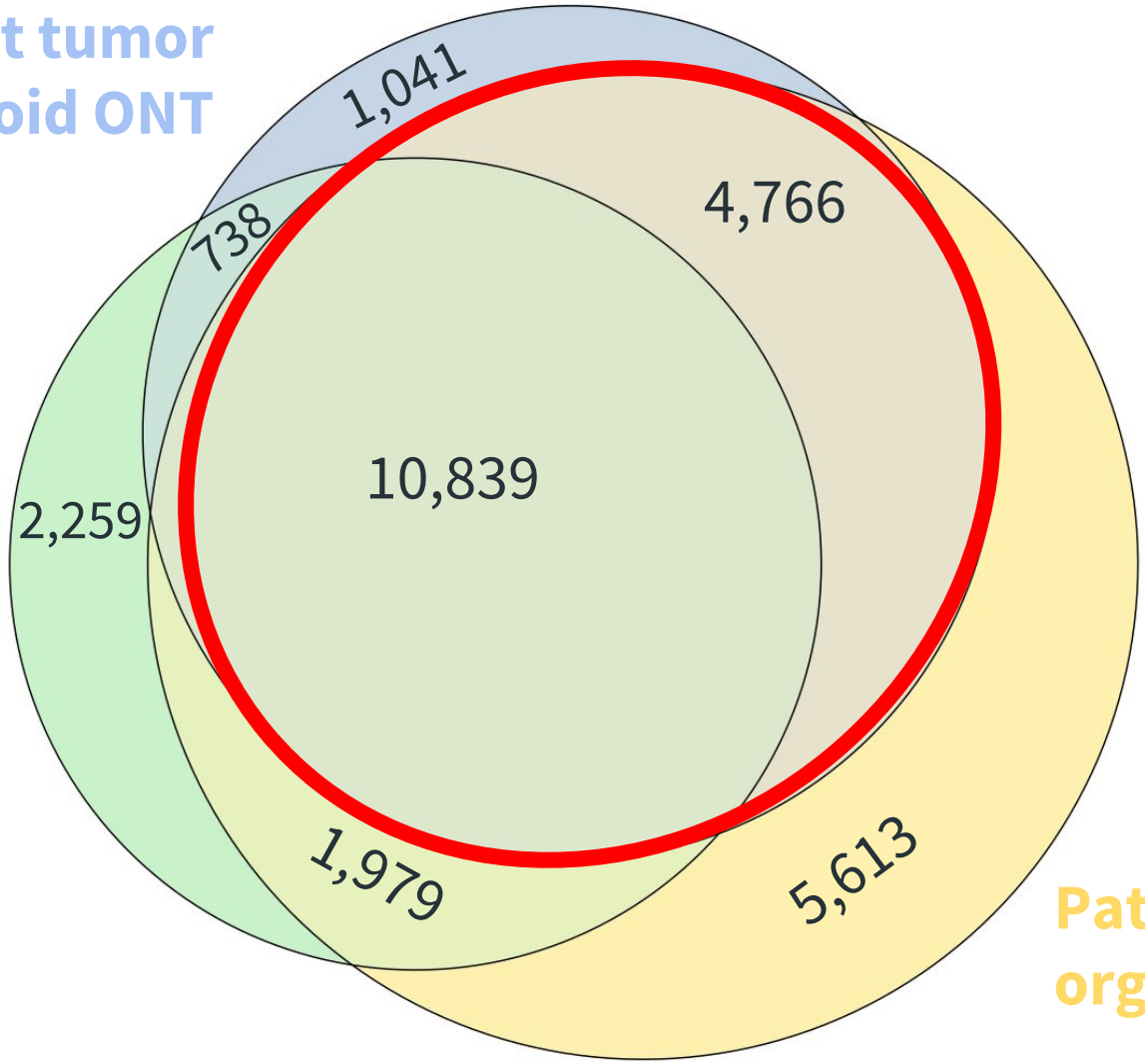
Samples and sequencing types

		ONT	Pacbio	Illumina
Breast Cancer	**Patient tumor (organoid, ER+, PR+, Her2-)**	✓	✓	✓
	SK-BR-3 (cell line, ER-, PR-, Her2+)	✓	✓	✓
	MCF-7 (cell line, ER+, PR+, Her2-)	✓		✓
	MDA-MB-231 (cell line, ER-, PR-, Her2-)	✓		✓
Normal	**Patient normal (breast tissue)**	✓		✓
	MCF-10A (normal breast cell line)	✓		✓
	NA12878	✓	✓	✓
	HG002 (Genome in a Bottle)	✓	✓	✓

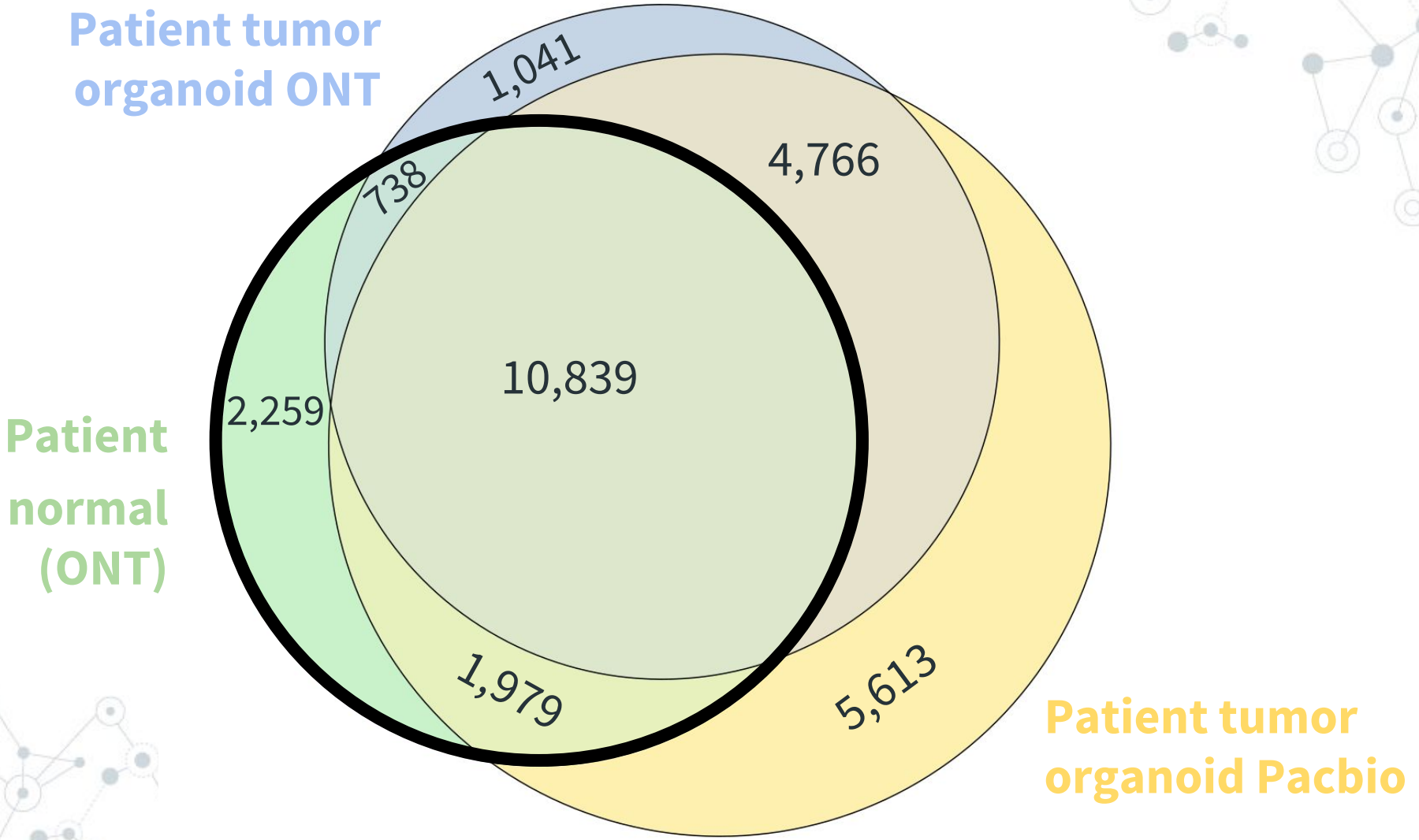
Tumor organoid vs normal variant calls



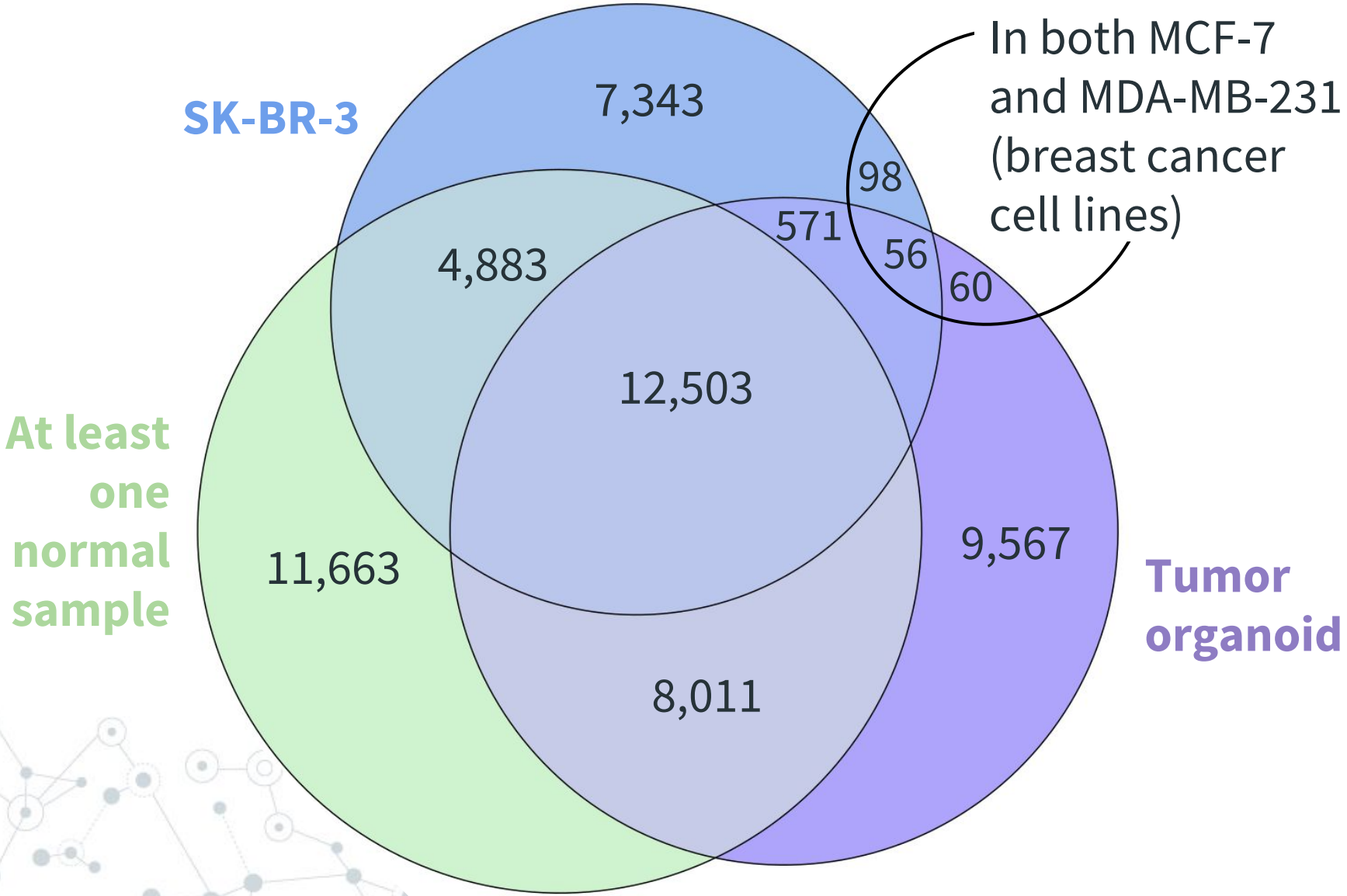
Tumor organoid vs normal variant calls



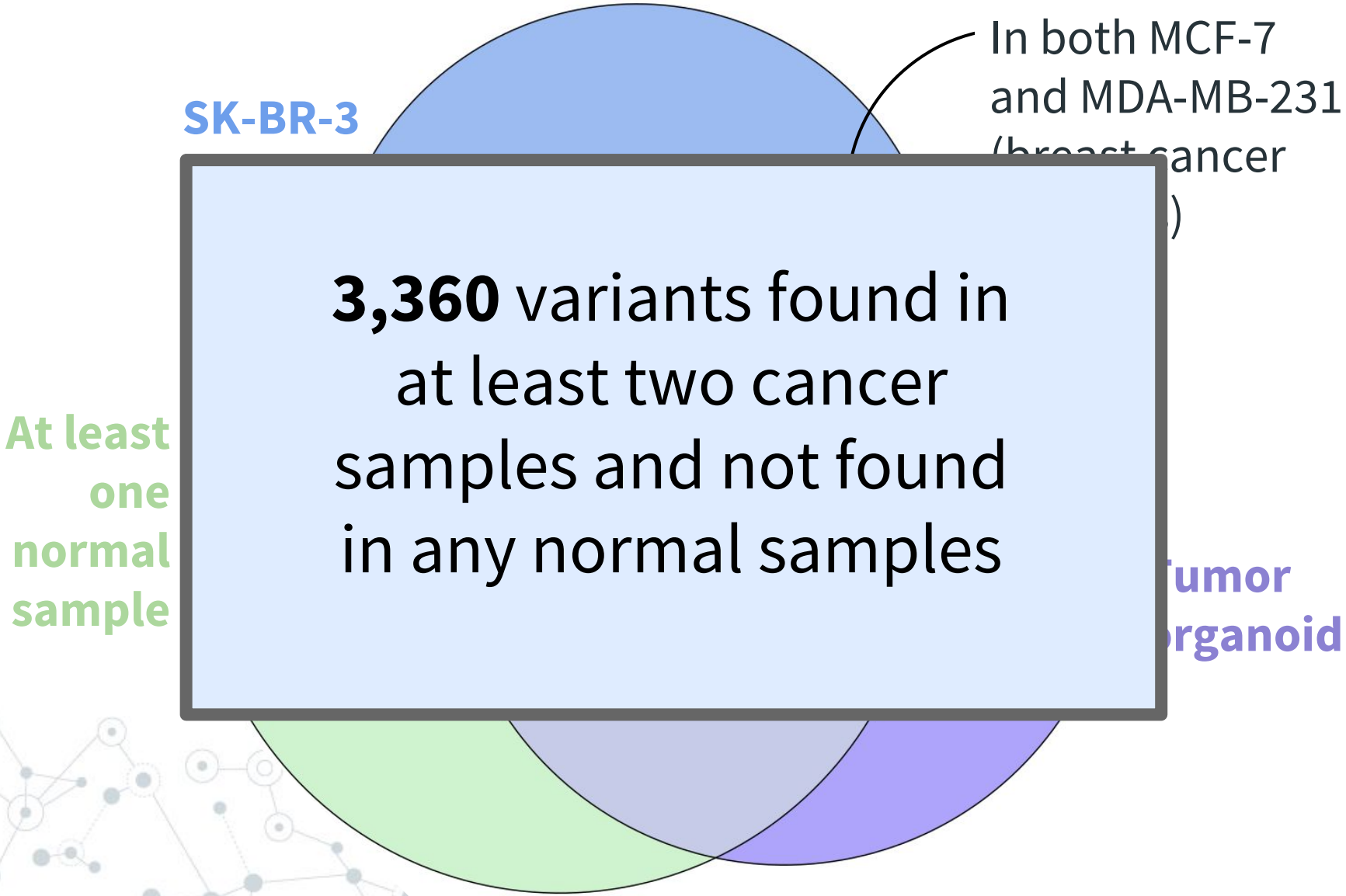
Tumor organoid vs normal variant calls



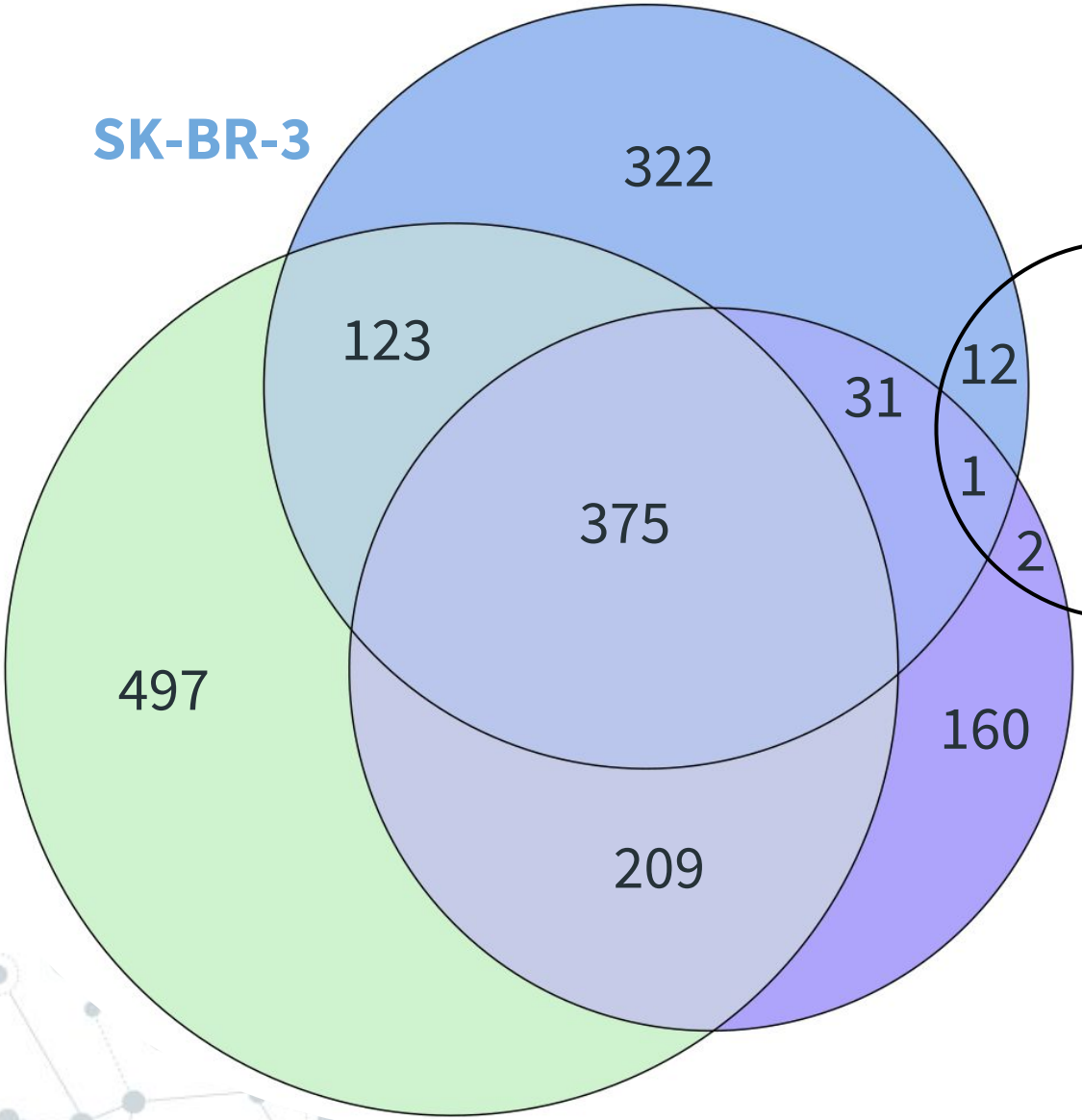
Cancer vs normal variant calls



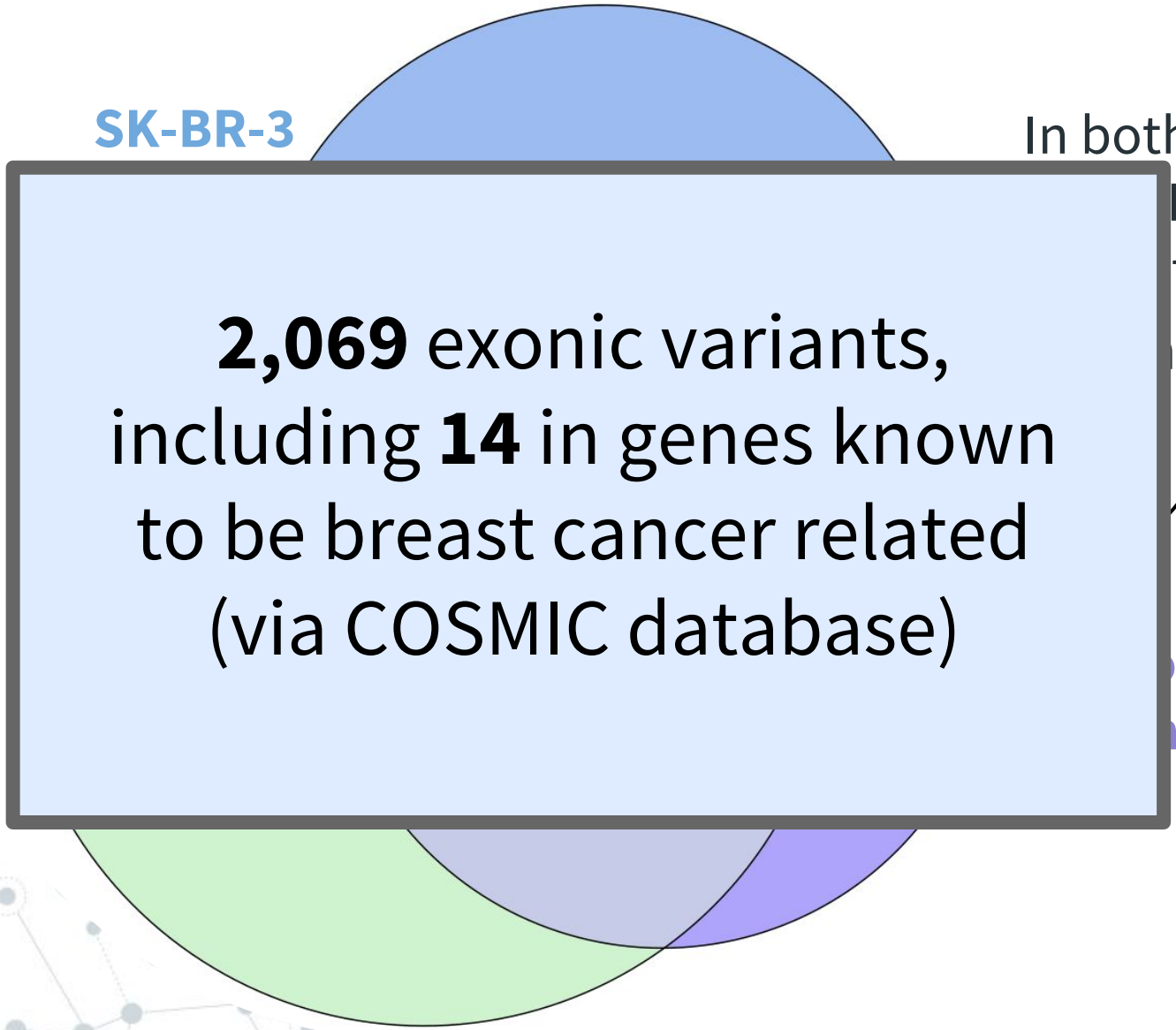
Cancer vs normal variant calls



Exonic variants only



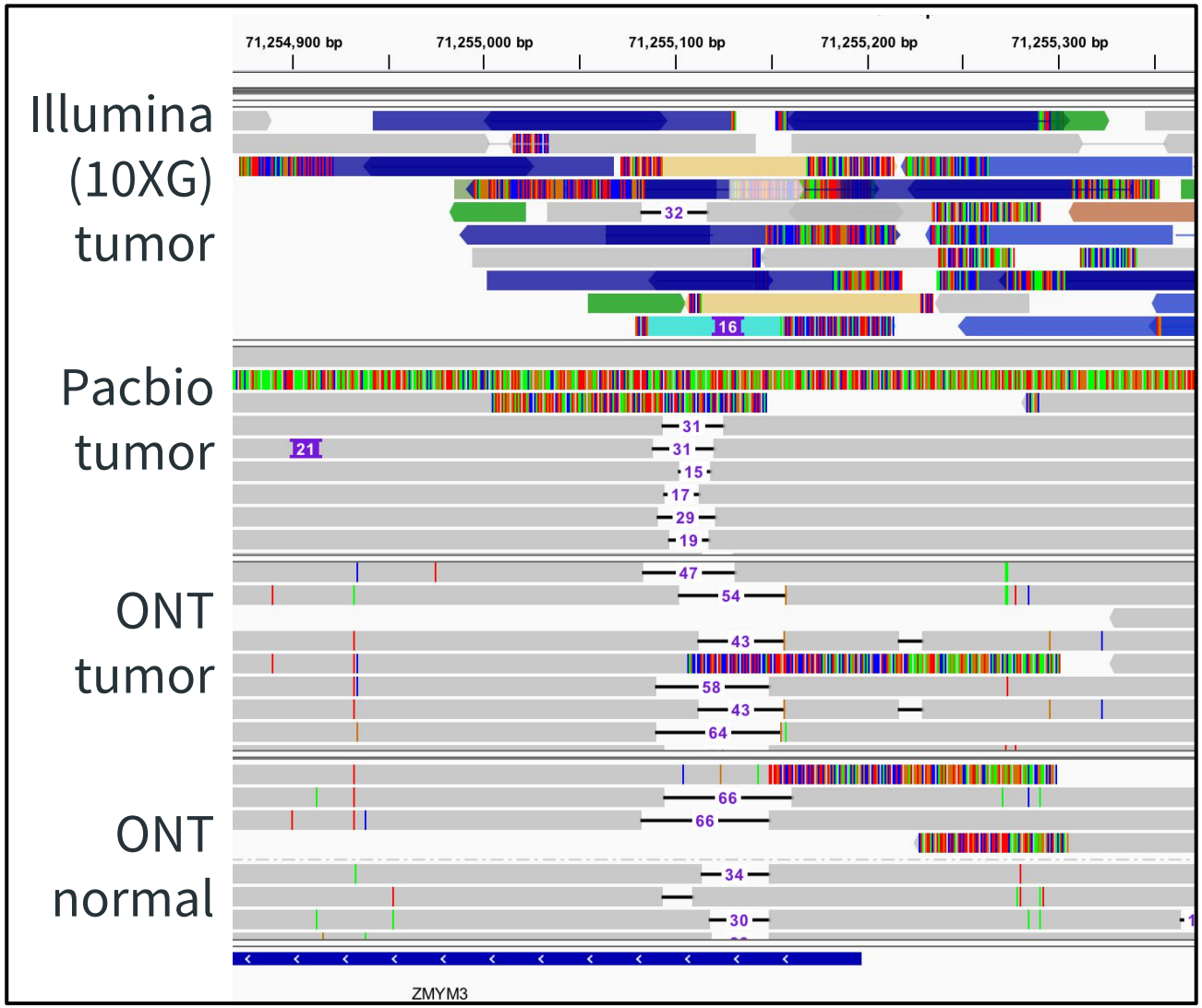
Exonic variants only



2,069 exonic variants, including **14** in genes known to be breast cancer related (via COSMIC database)

~50 bp deletion in ZMYM3 exon

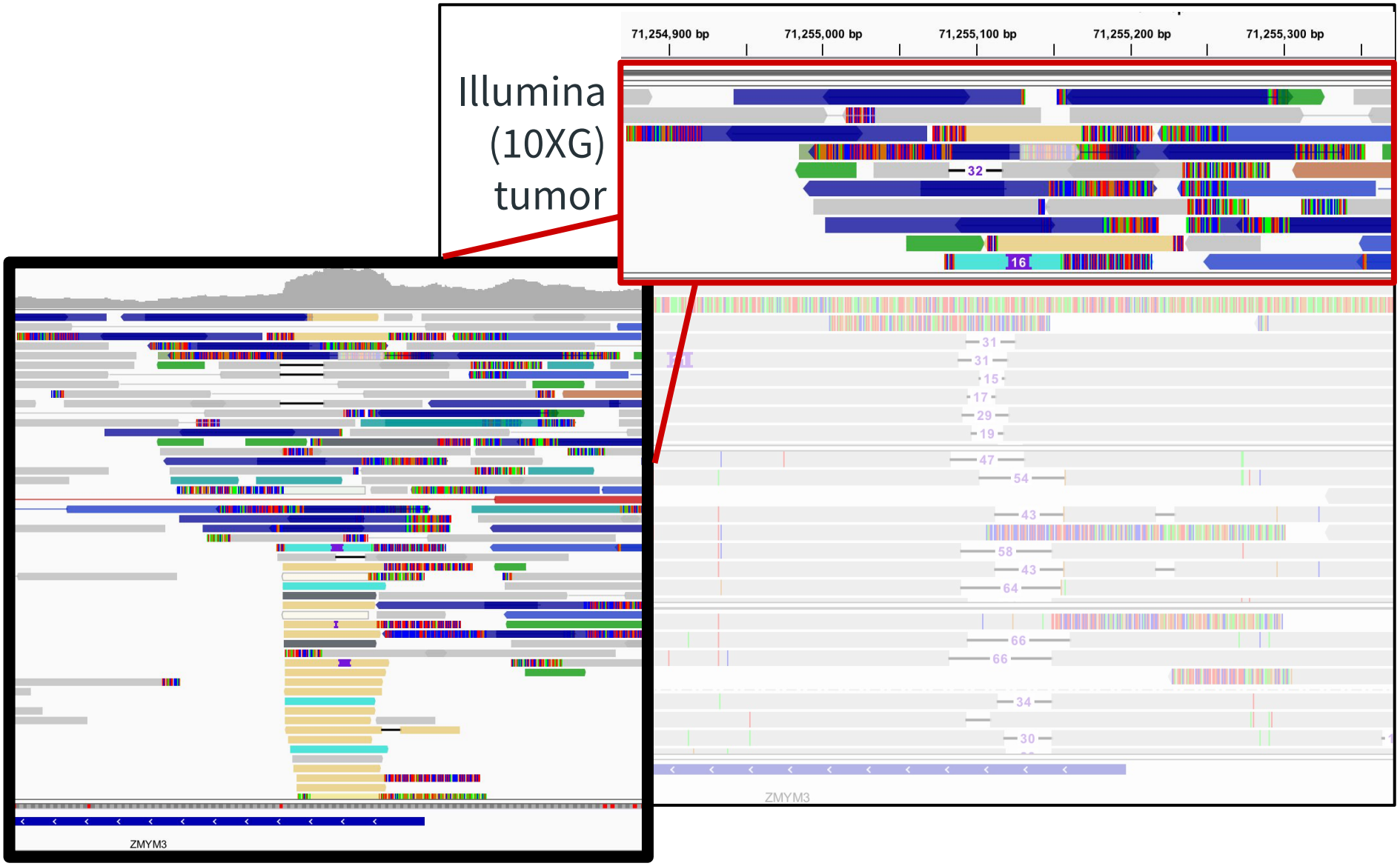
- Evidence it promotes DNA repair via BRCA1 recruitment [1]
- Encodes part of a histone deacetylase-containing complex involved in gene silencing *



* via Genecards

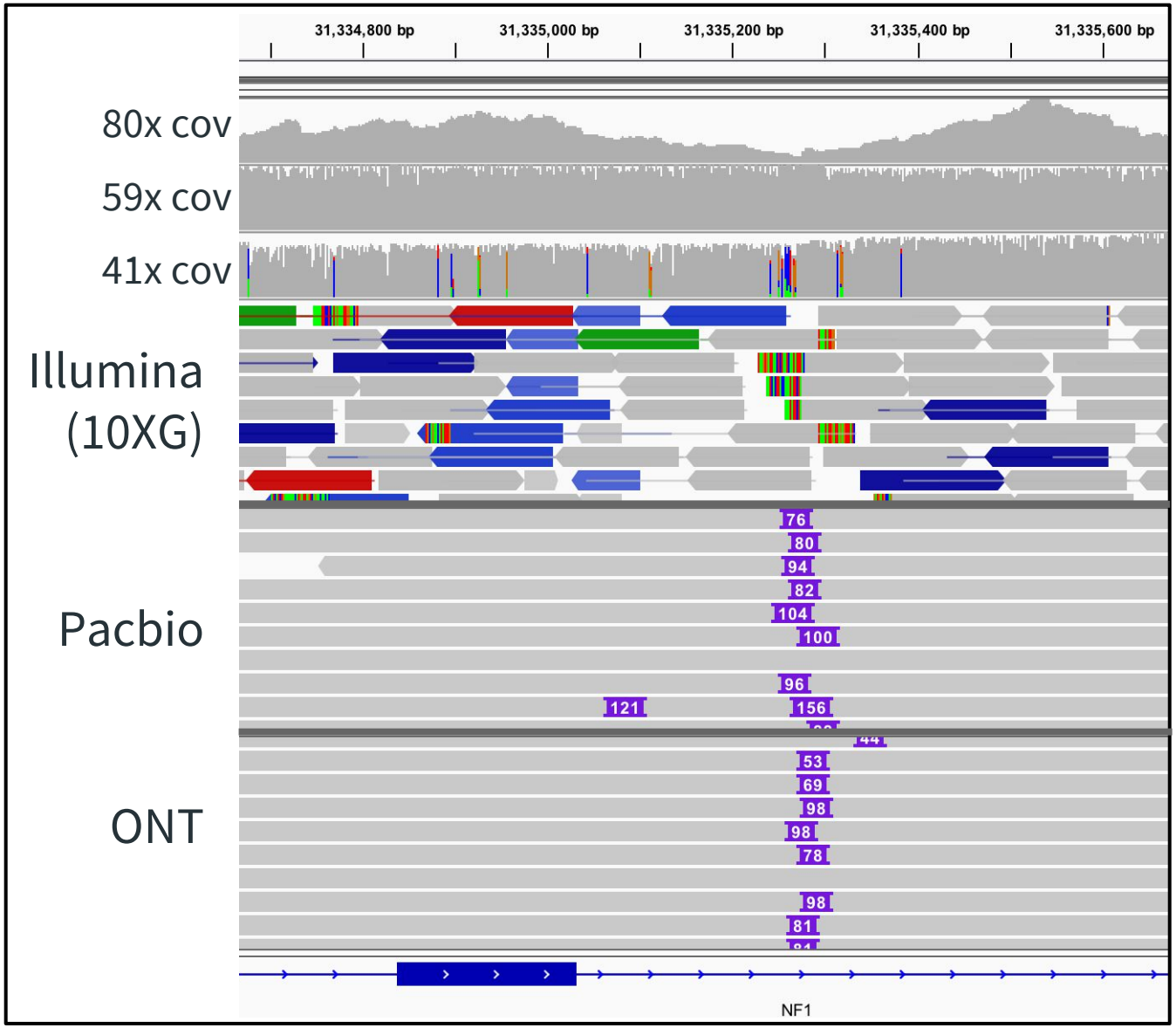
[1] Leung, J. W., *et al* (2017).

~50 bp deletion in ZMYM3 exon



~100 bp insertion in NF1 intron

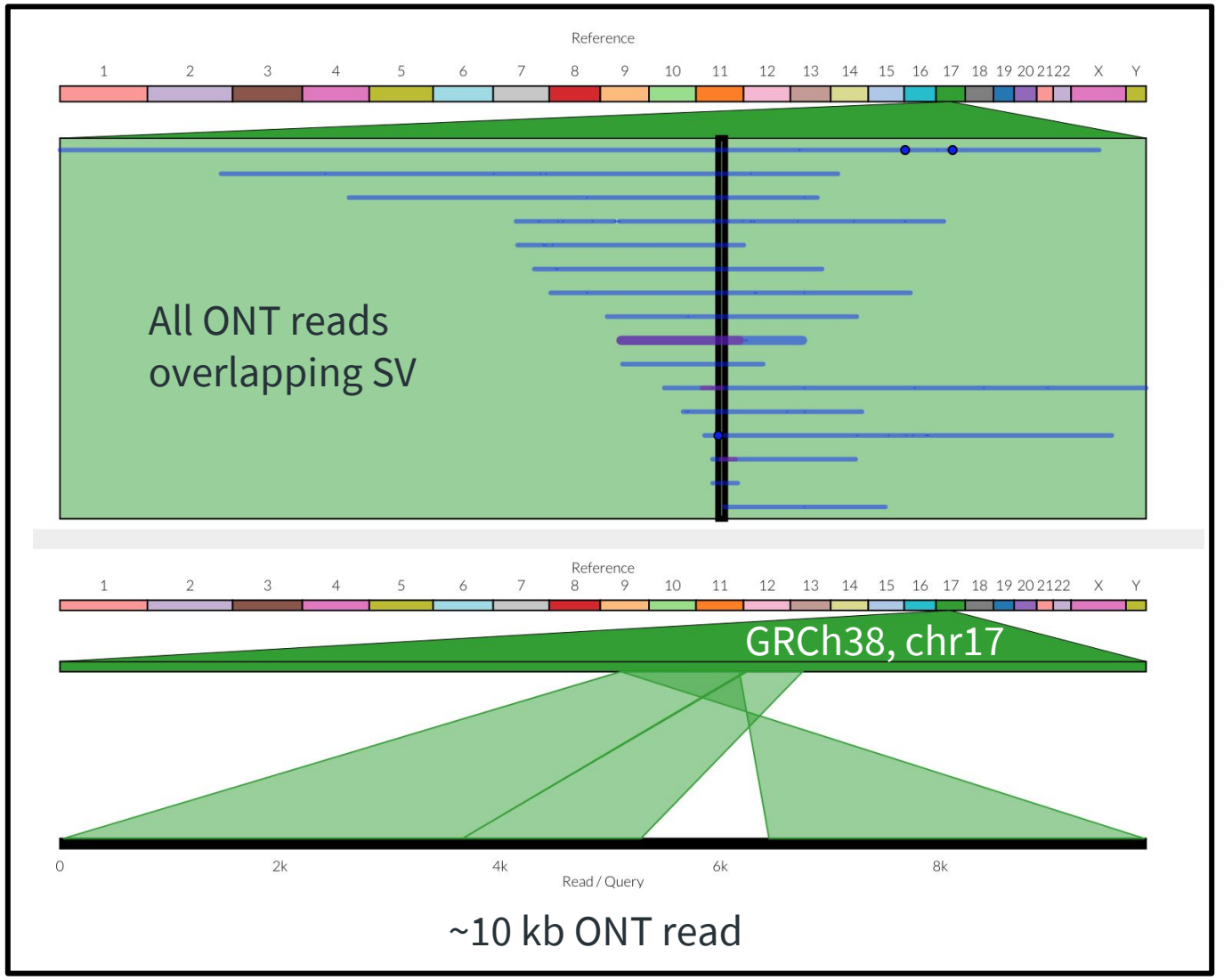
- Negative regulator of the ras signal transduction pathway *
- Known to be involved in the “Integrated Breast Cancer Pathway” *



* via Genecards

Inverted duplication in ERBB2 (Her2) exon

- Amplification/ overexpression detected in 20-30% of breast cancers *
- Involved in transcriptional regulation *



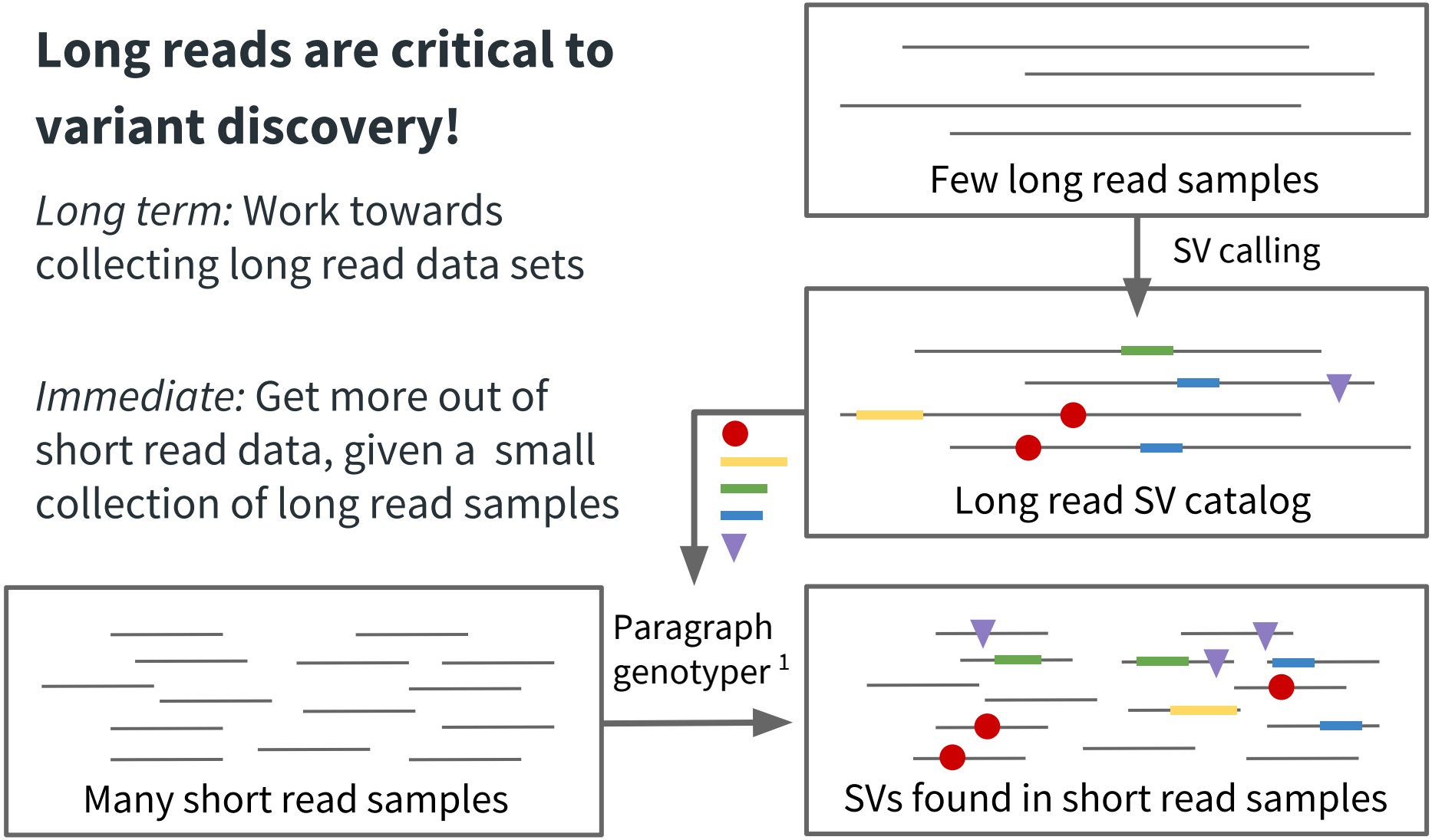
* via Genecards

Conclusions and future directions

Long reads are critical to variant discovery!

Long term: Work towards collecting long read data sets

Immediate: Get more out of short read data, given a small collection of long read samples



[1] https://github.com/Illumina/paragraph

Acknowledgments

Johns Hopkins University

Michael Schatz

Michael Kirsche

Sergey Aganezov

Sam Kovaka

Katie Jenike

Winston Timp

Isac Lee

Steven Salzberg

Baylor College of Medicine

Fritz Sedlazeck

Medhat Helmy

Northwell Health

Karen Kostroff

Cold Spring Harbor Lab

W. Richard McCombie

Sara Goodwin

Melissa Kramer

Robert Wappel

David L. Spector

Gayatri Arun

Illumina (Paragraph)

Peter Krusche

Sai Chen

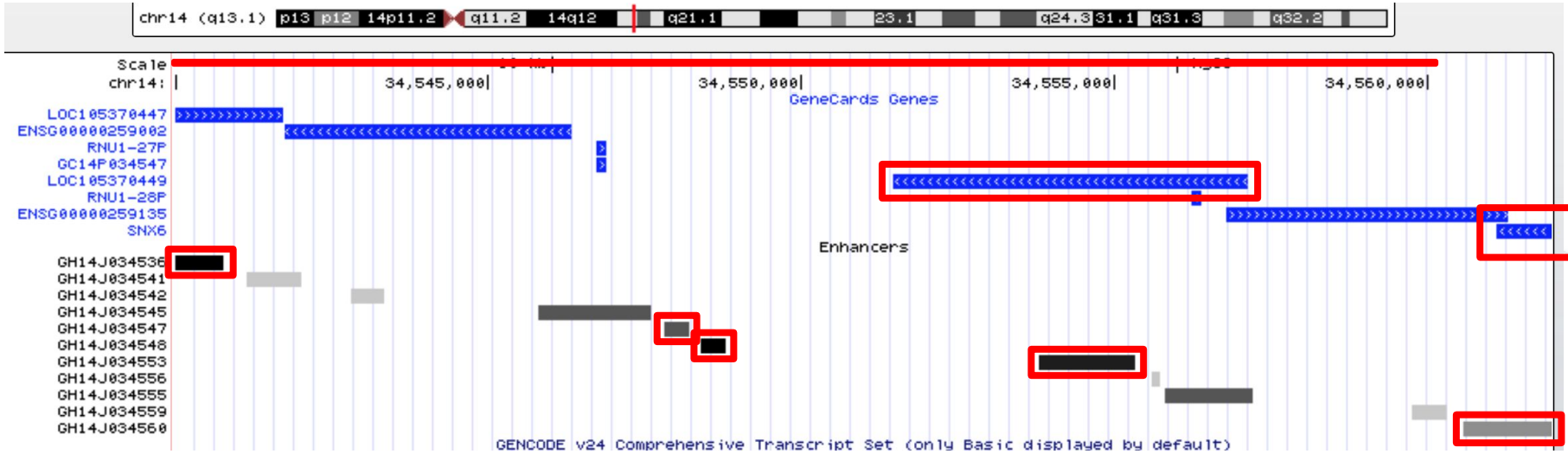
Michael Eberle

Felix Schlesinger

Egor Dolzhenko

Questions?

Exonic variant (all cancer, no normal samples)

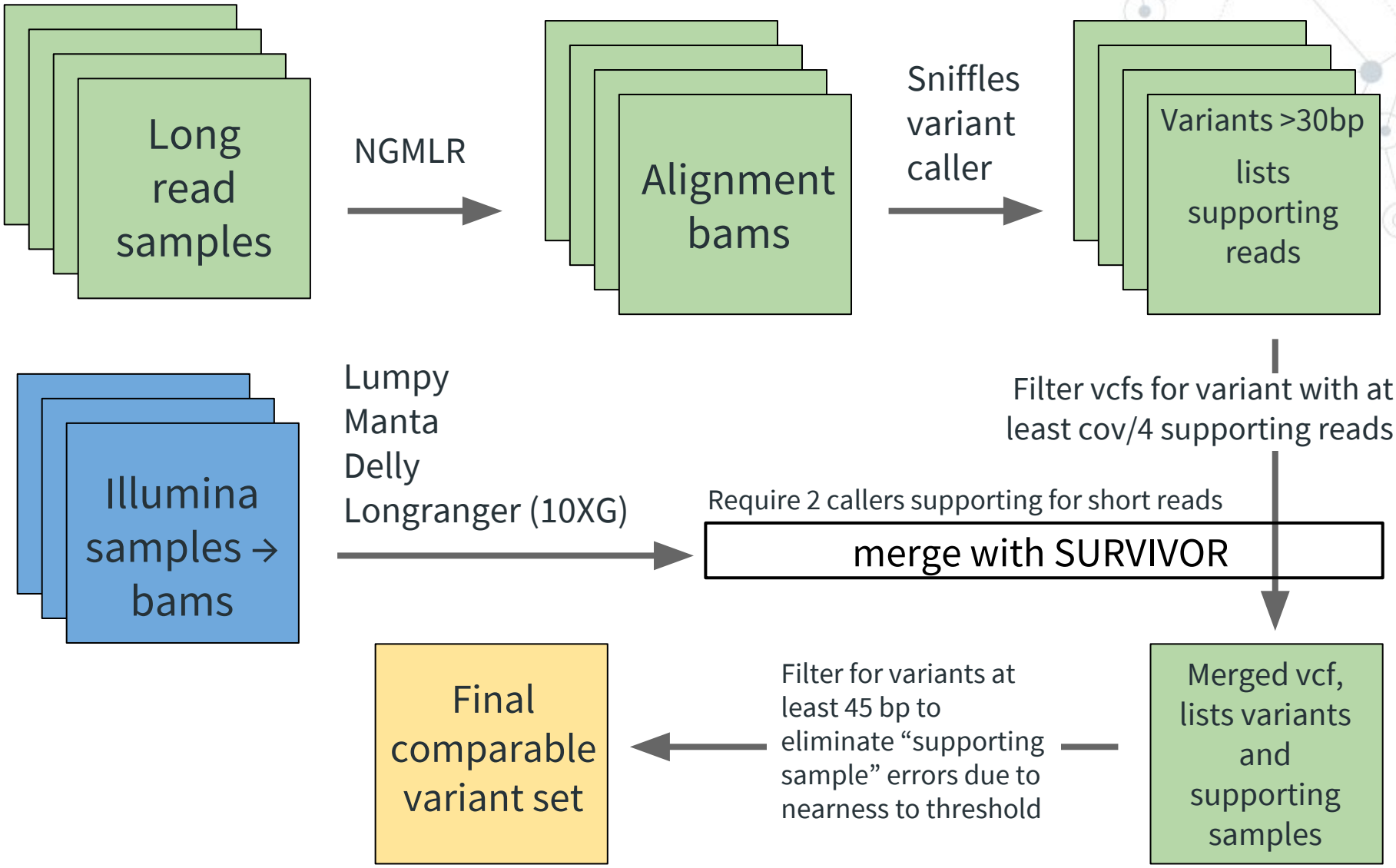


17 kb inversion

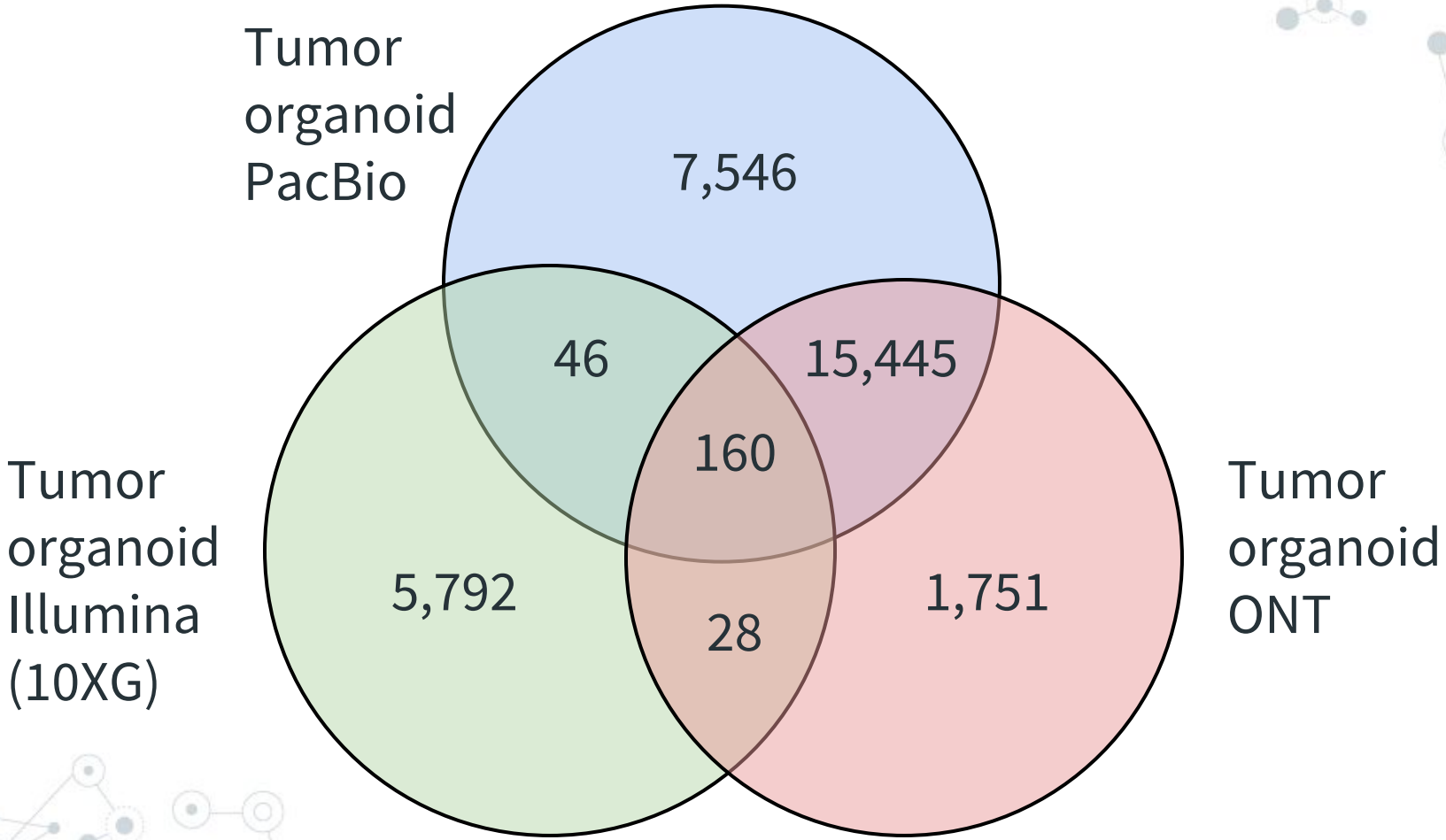
Encompases ncRNA LOC105370449

Just ahead of SNX6, aka TRAF4, a known oncogene

Protocol for variant comparison



Data statistics for all samples

Sample	Sequencing technology	Tumor or normal	Average coverage	Number of supporting reads required for Sniffles calls
SK-BR-3	ONT minion	Tumor	27.8	7
	ONT promethion		25	6
	Pacbio		52.9	10
	Illumina		20	N/A
MCF7	ONT minion	Tumor	16.5	4
MDA-MB-231	ONT minion	Tumor	17	4
Patient Tumor Organoid	10X Genomics	Tumor	28	N/A
	ONT minion		31.5	8
	Pacbio		55.3	10
Patient Breast Tissue	ONT minion	Normal	11.6	3
MCF10A	ONT minion	Normal	22.4	6
HG002 (GiaB son)	Pacbio	Normal	50.4	10
NA/GM12878	ONT minion	Normal	31.2	8
	Pacbio		41.4	10
	ONT minion		20	5

Platform concordance in tumor organoid