

Structural Variants (SVs) and Copy Number Variants (CNVs) using long-read

May 11 2025

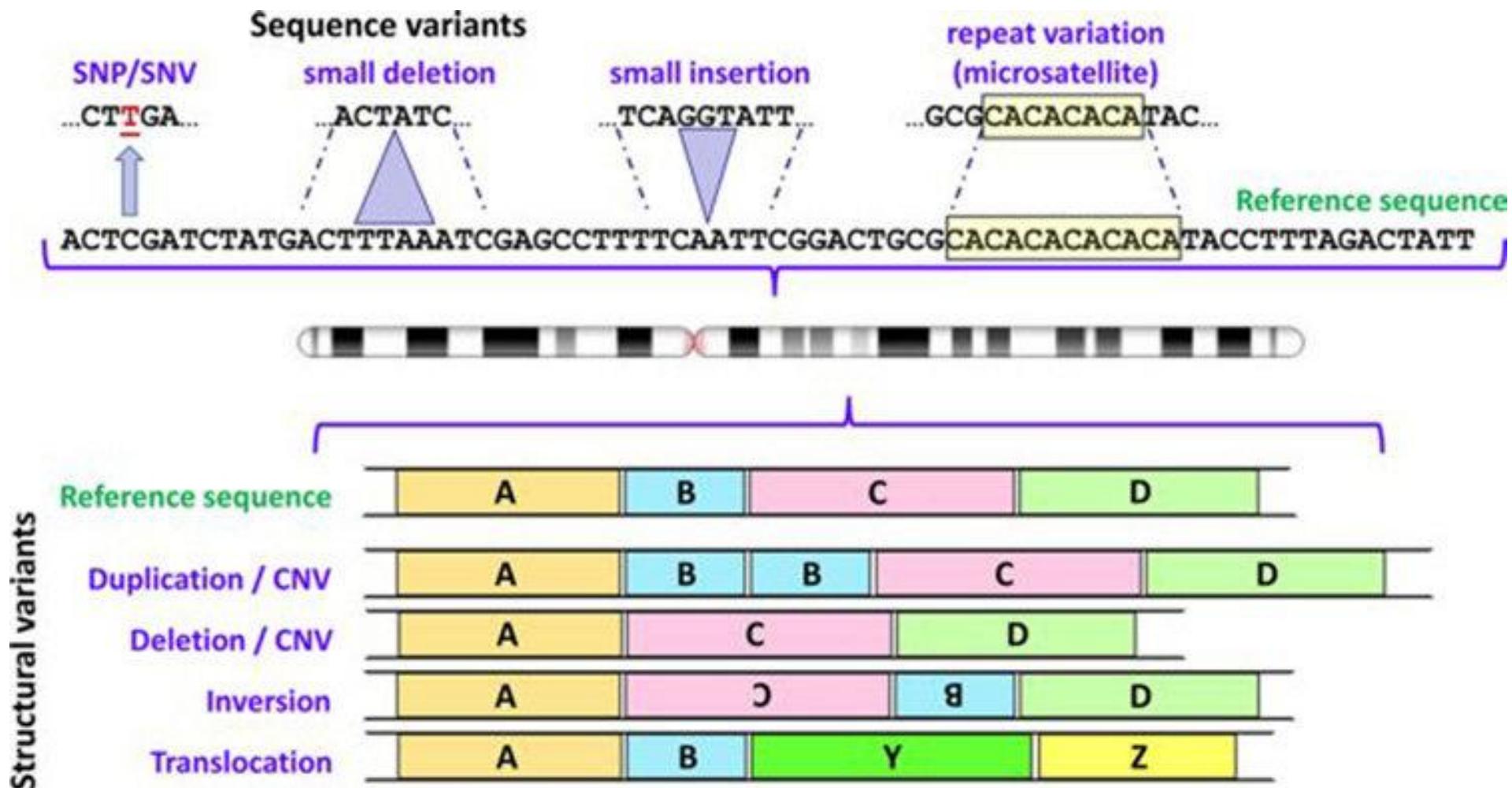
Slide adapted from Dr. Tobias Rausch - EMBL

Phuc Loi Luu, PhD

Email: Luu.p.loi@googlemail.com

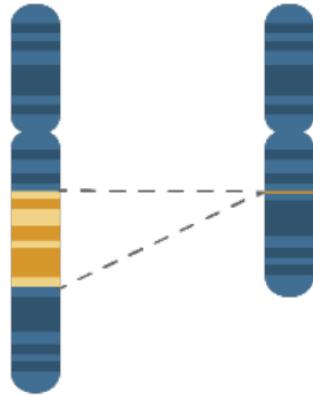
Zalo: 090102182

Type of variants

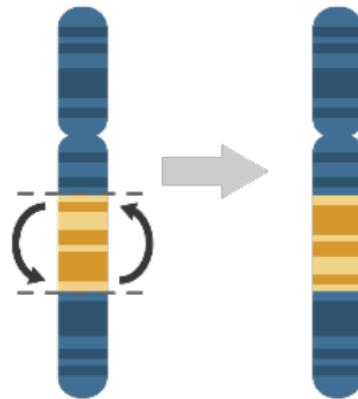


Somatic and Germline Structural Variants (SVs)

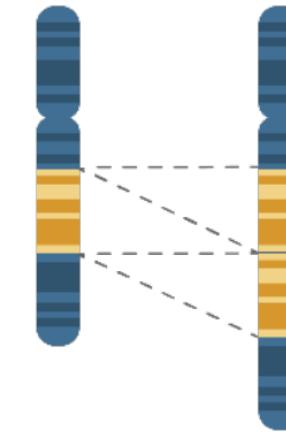
Deletion



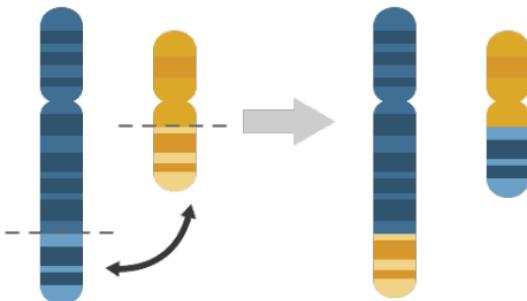
Inversion



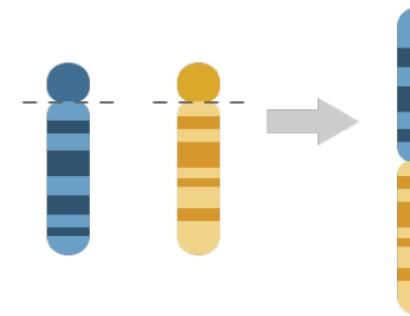
Duplication



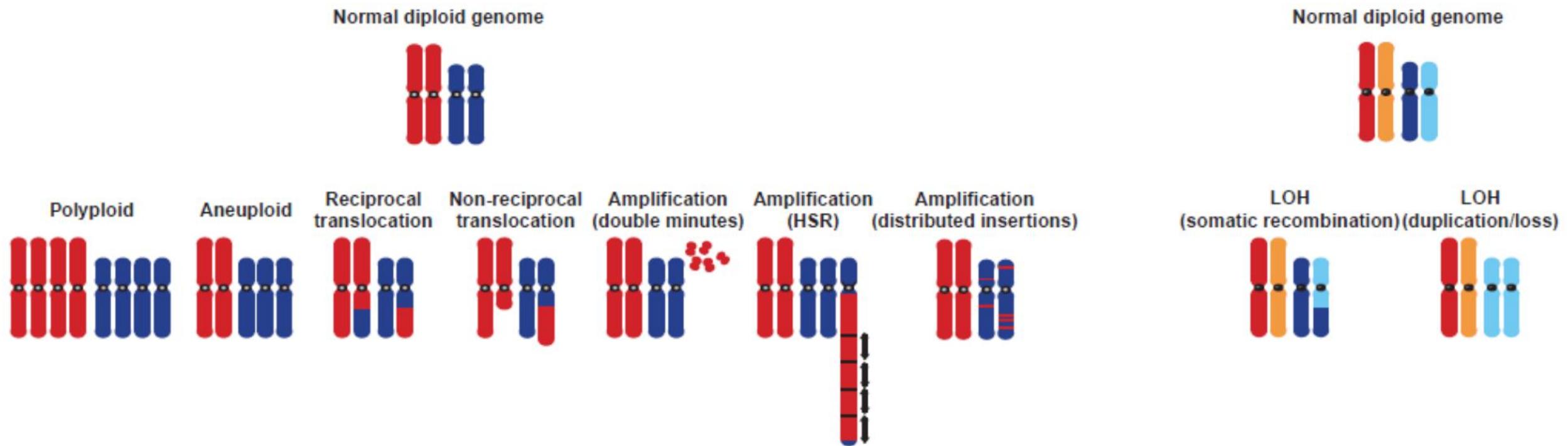
Reciprocal translocation



Robertsonian translocation



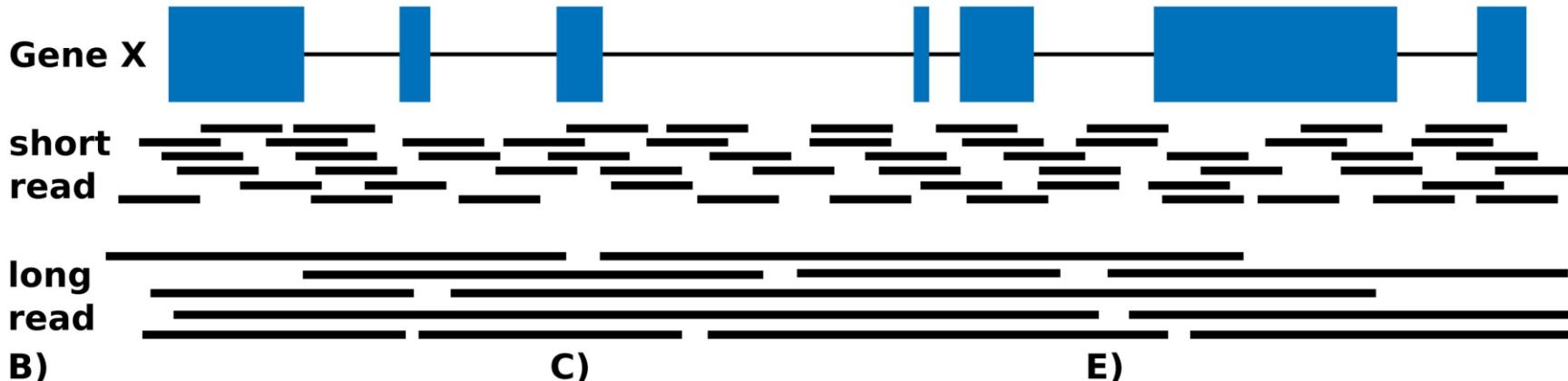
Cancers harbour a wide Range of Chromosome Aberrations



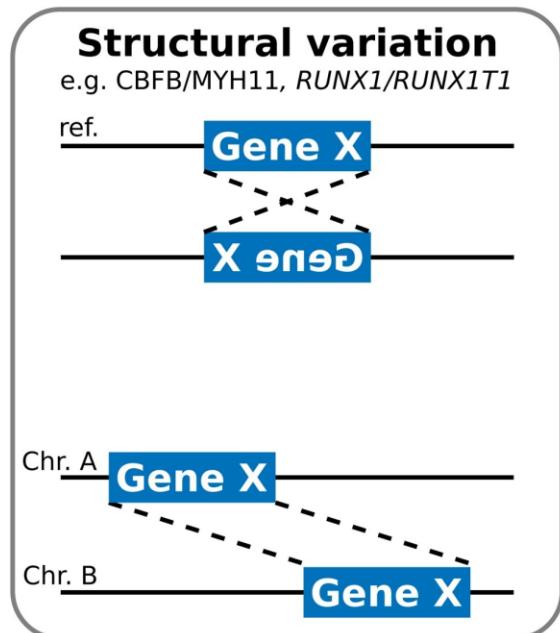
Chromosome aberrations in solid tumors. Albertson et al., Nat Genet. 2003 Aug;34(4):369-76.

Short Read vs Long Read

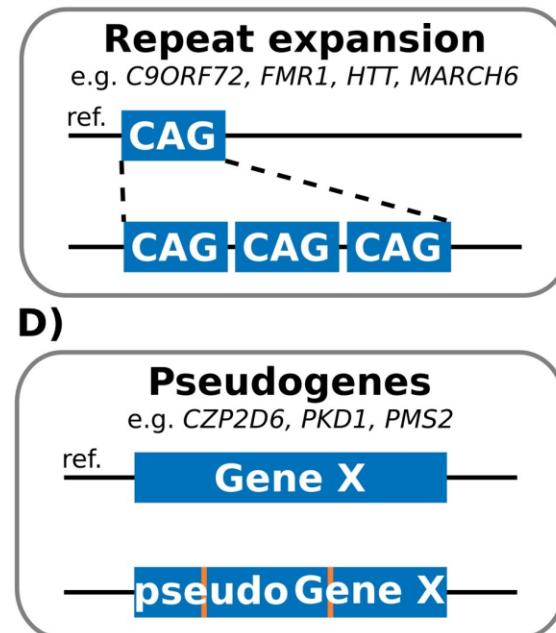
A)



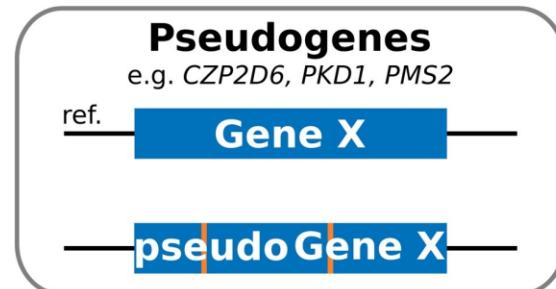
B)



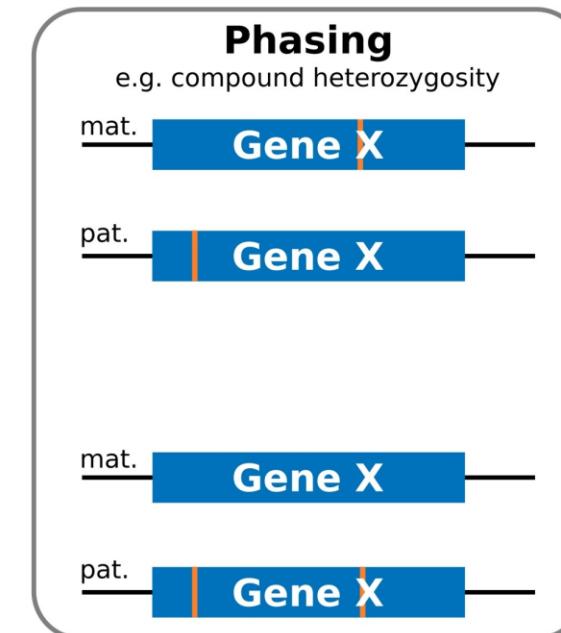
C)



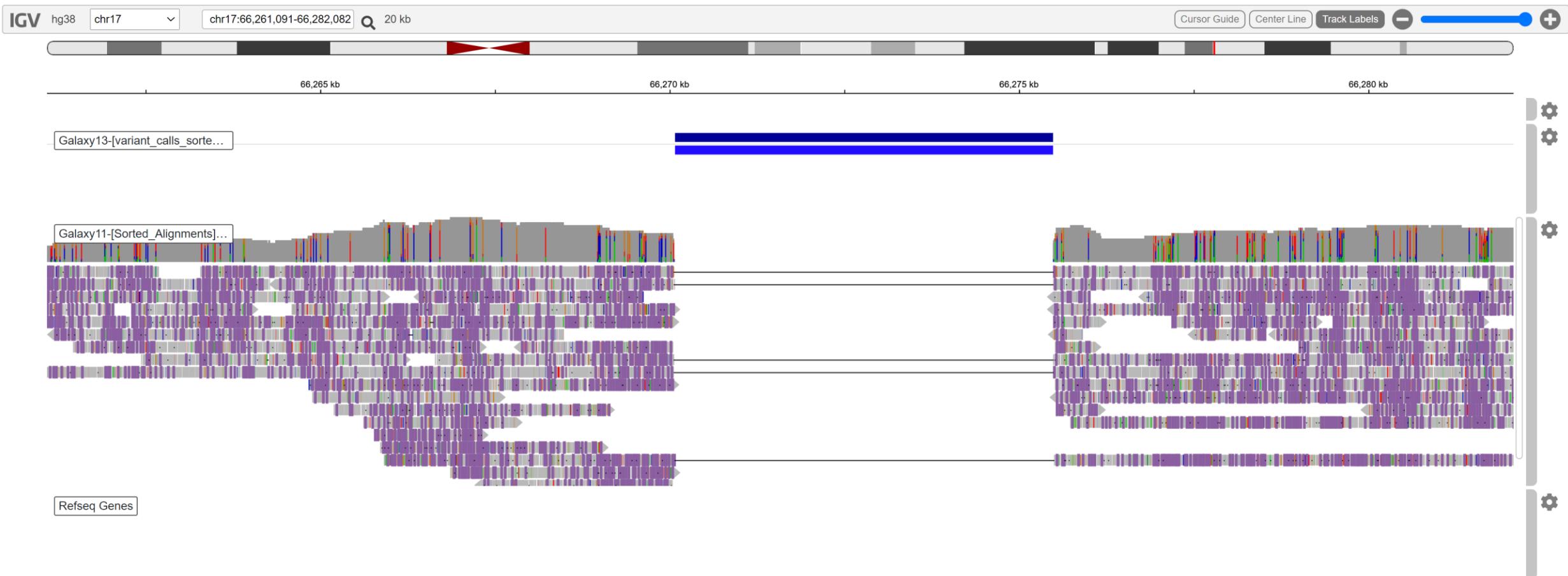
D)



E)

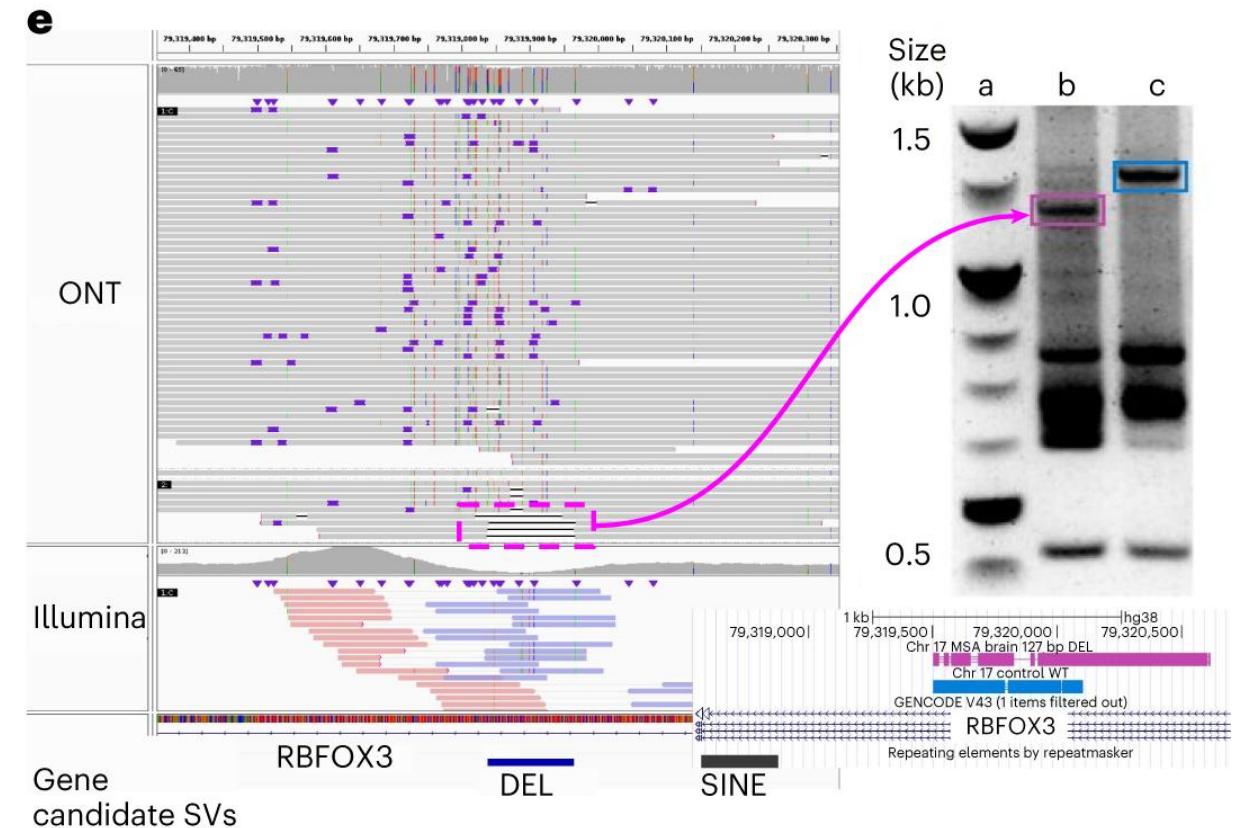
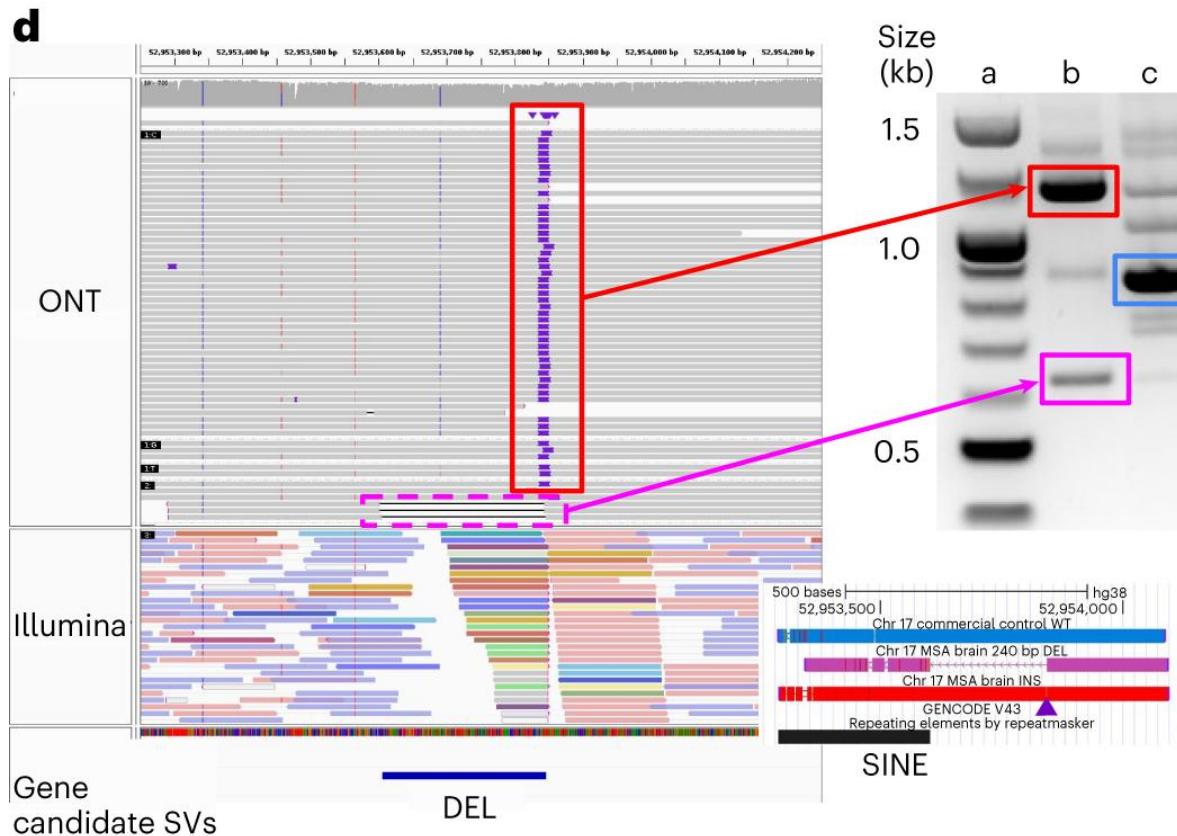


A large deletion in chromosome 17 from 66270K and 66276K



https://www.melbournebioinformatics.org.au/tutorials/tutorials/longread_sv_calling/longread_sv_calling/

Recovery of somatic SVs using the Sniffles2 mosaic mode



IGV alignments for a Structural Variant that was called by Sniffles2 but not represented in either the Bionano or Illumina call sets

ONT
cingulate
cortex

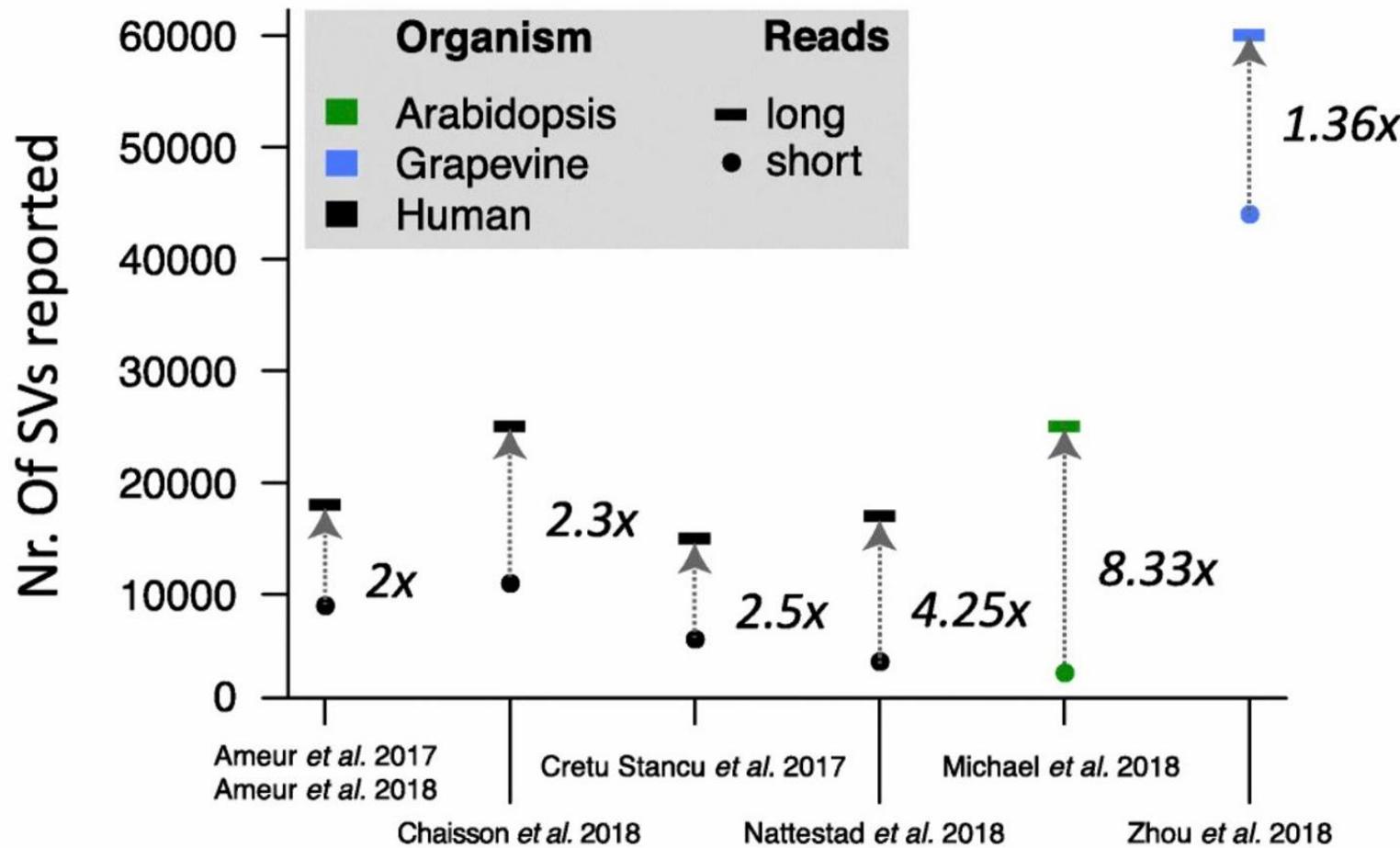
Illumina
cingulate
cortex

Illumina
white matter

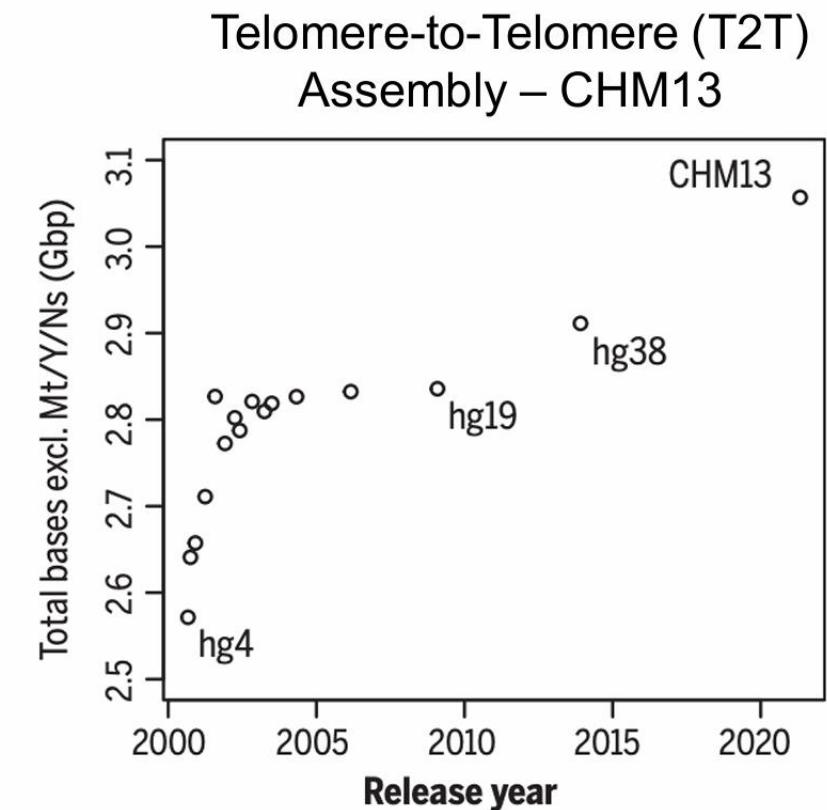
Gene
SV type
Repeat



Germline SV detection using short-reads is largely incomplete!



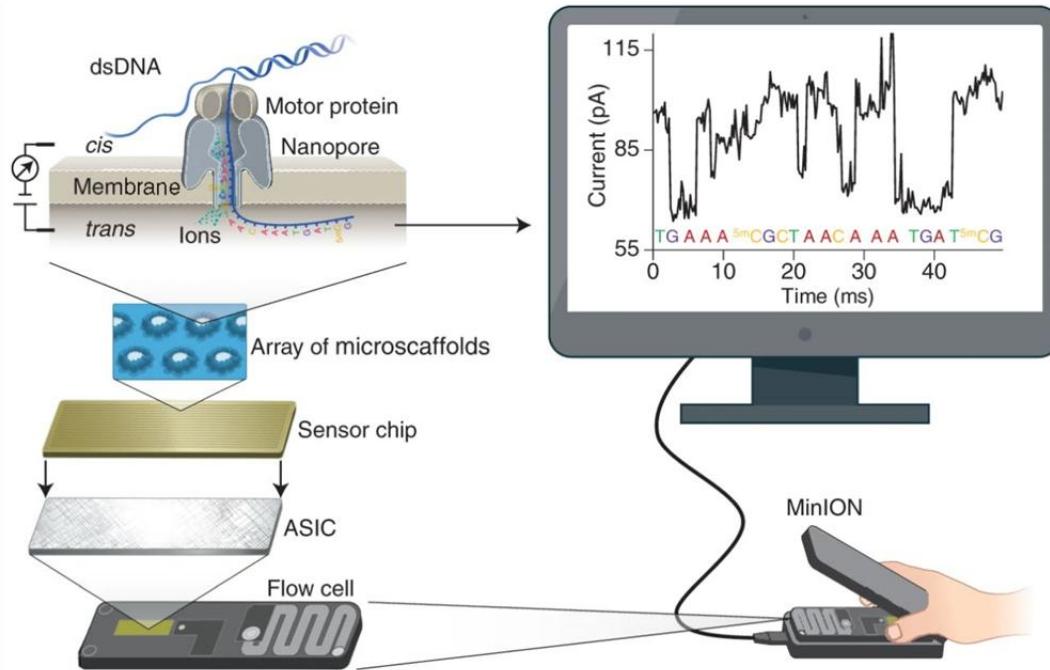
<https://doi.org/10.1186/s13059-019-1828-7>



<https://doi.org/10.1126/science.abj6987>

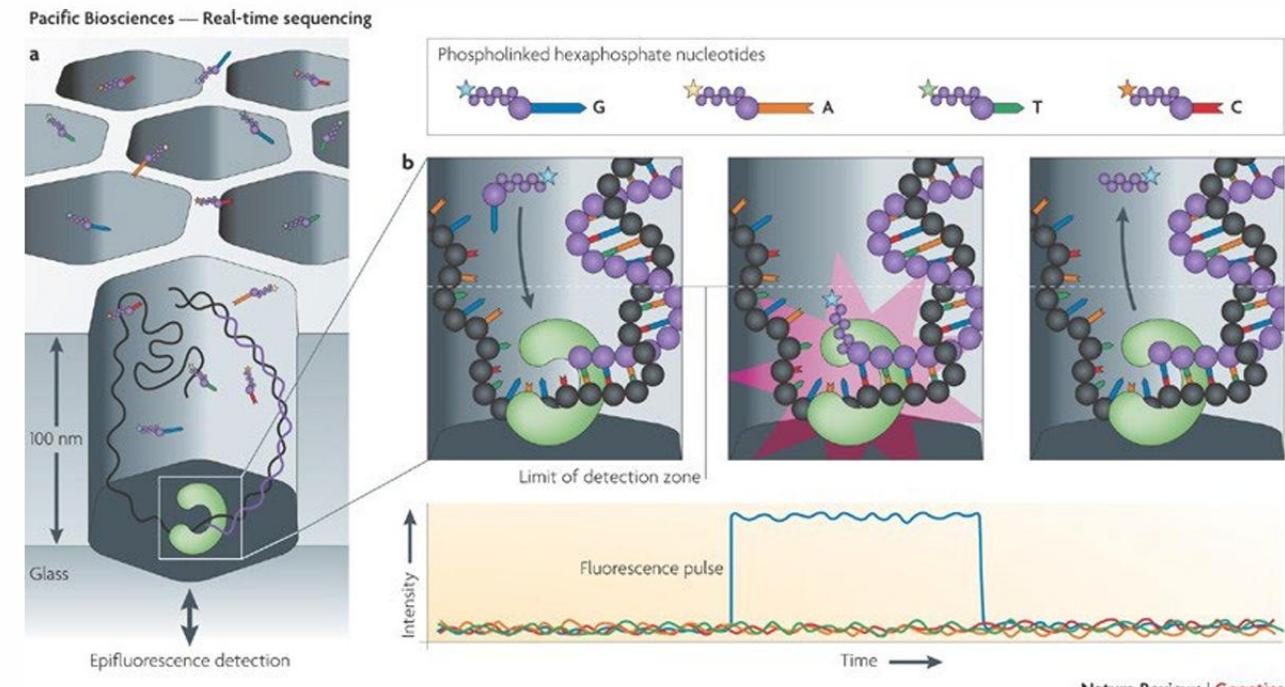
Oxford Nanopore Technologies (ONT) and Pacific Biosciences (PacBio)

Oxford Nanopore Sequencing



1,000bp – 20,000bp reads but some >>20Kbp
~1 error in 100 bases

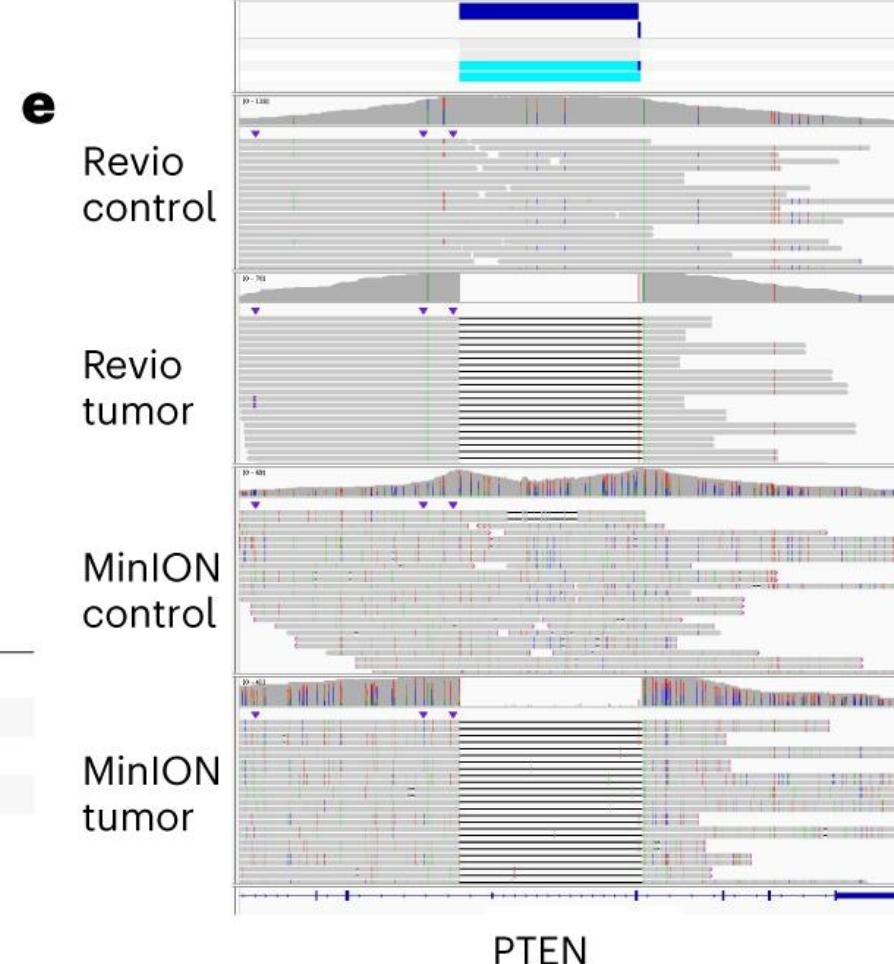
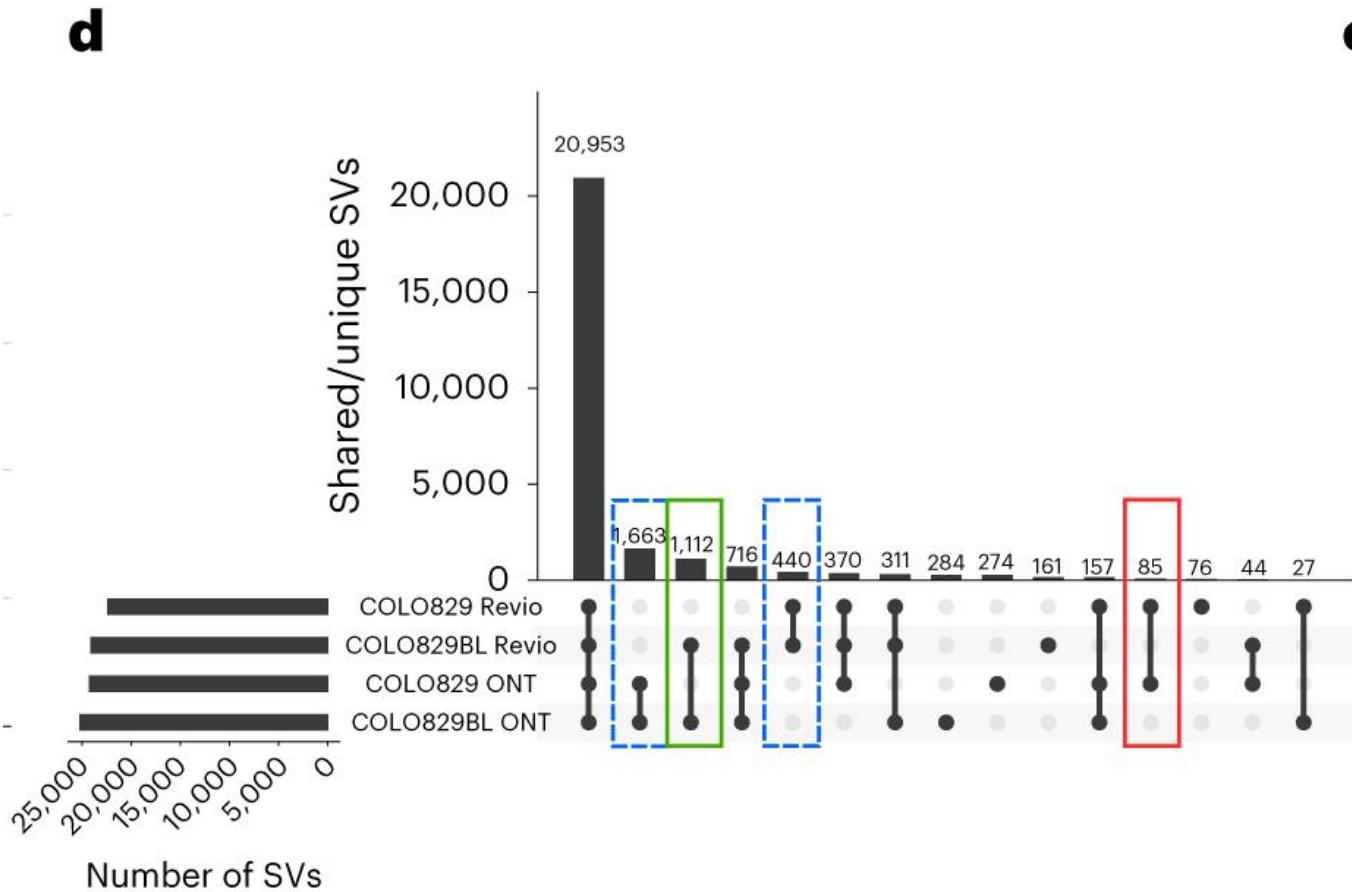
Pacific Biosciences Sequencing



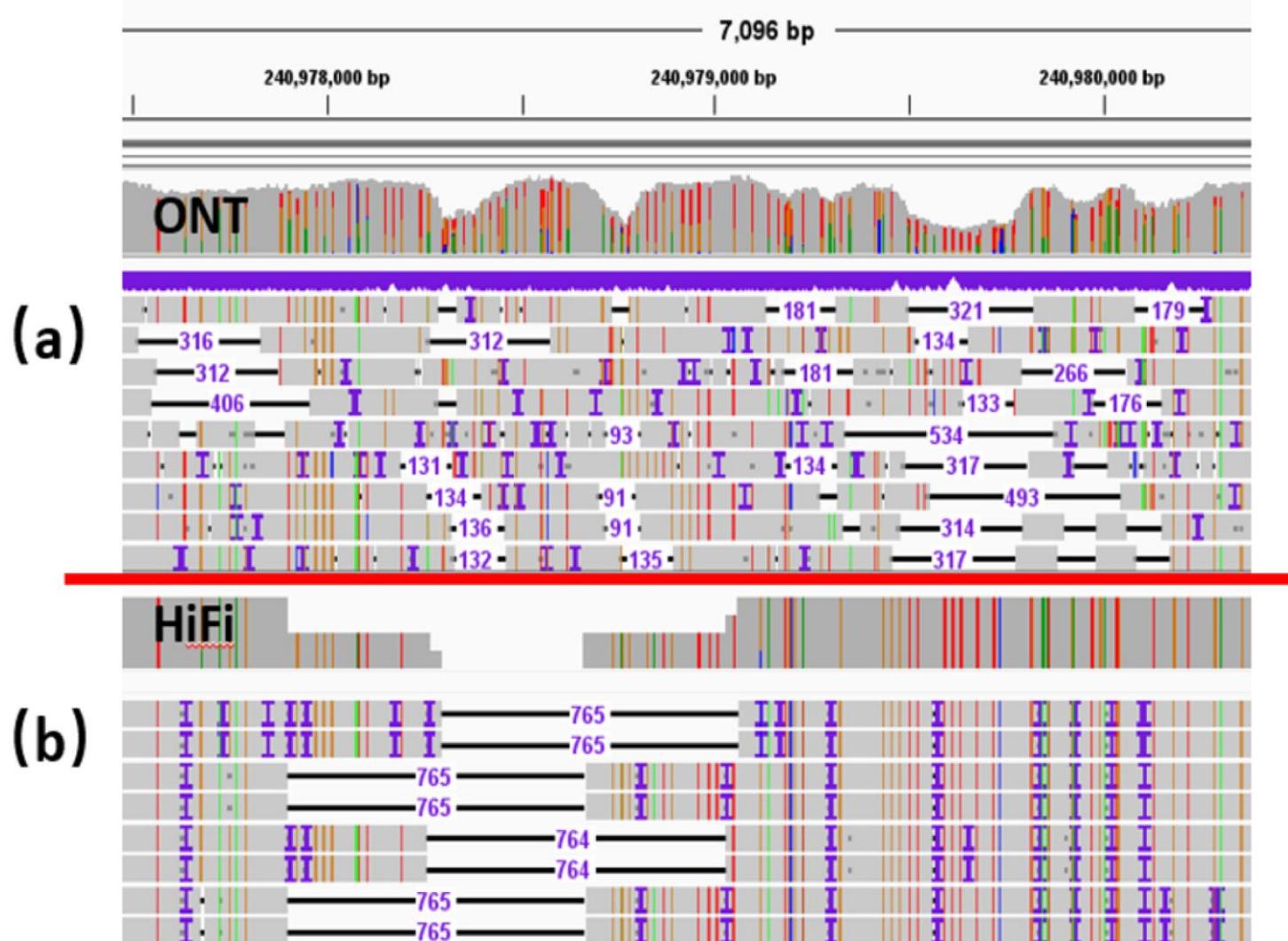
1,000bp – 20,000bp reads
~1 error in 10,000 bases

Nature Reviews | Genetics

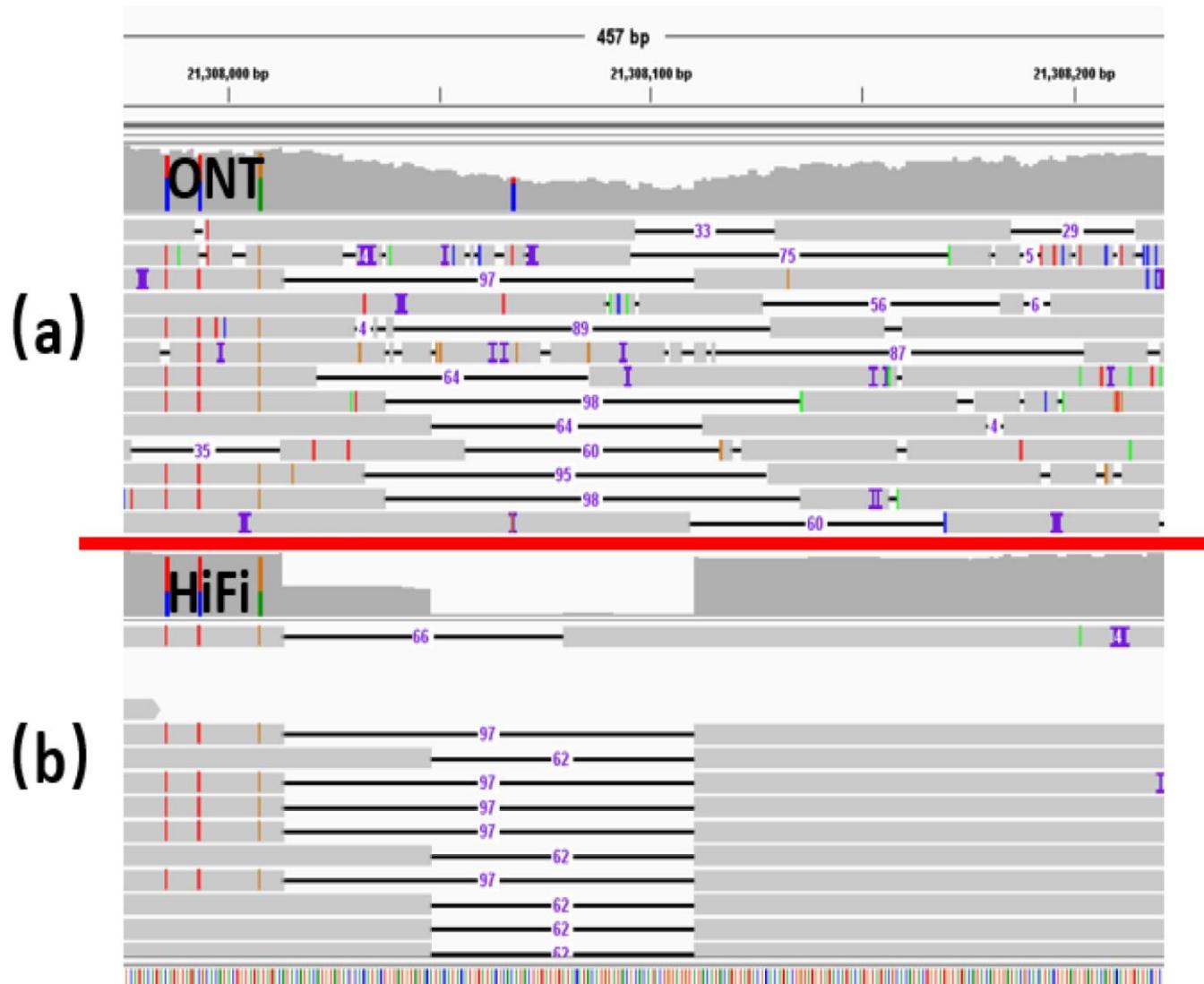
Recovery of somatic SVs using the Sniffles2 mosaic mode



The HiFi Pacbio long reads give the correct DEL of size 765 bp

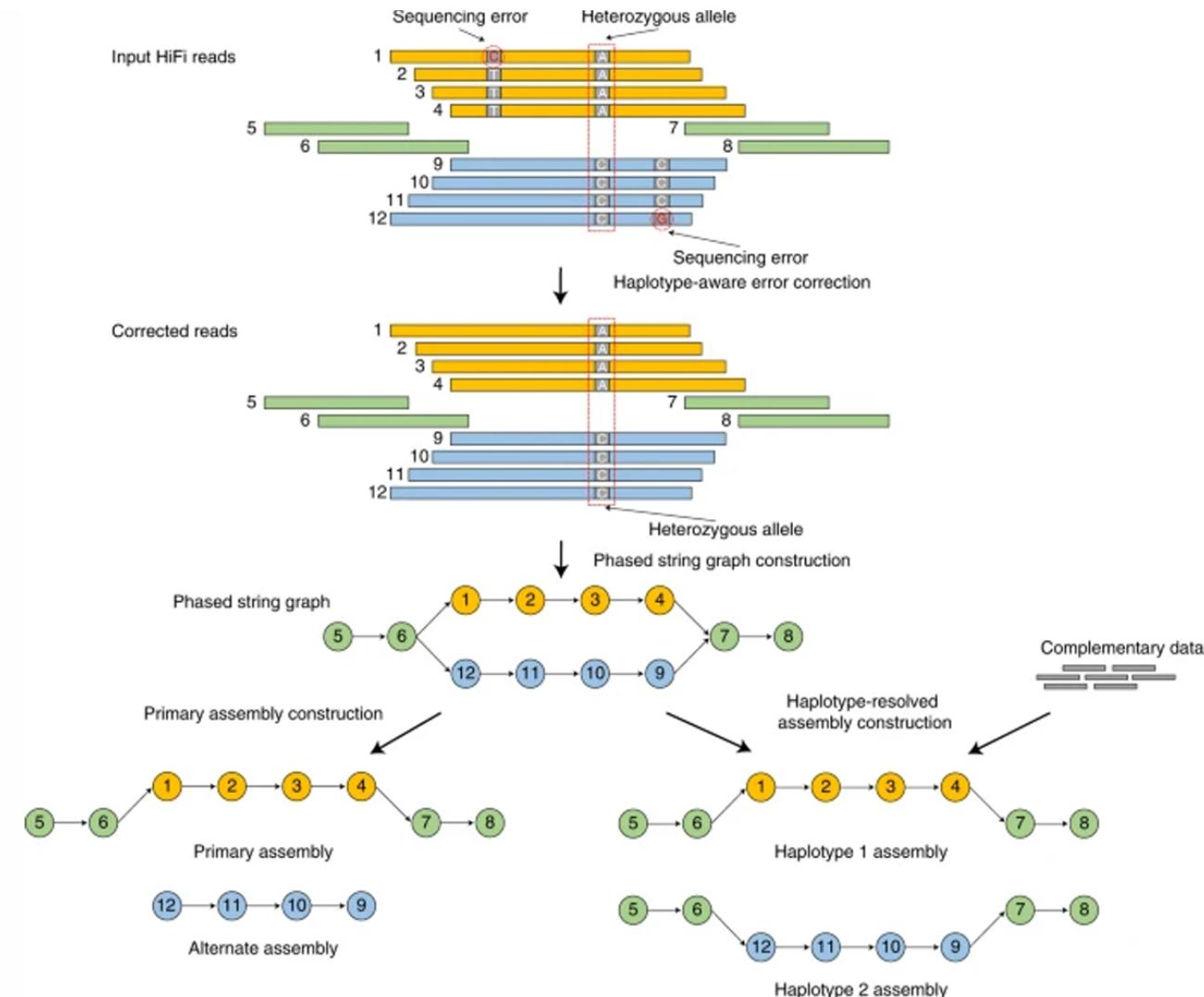
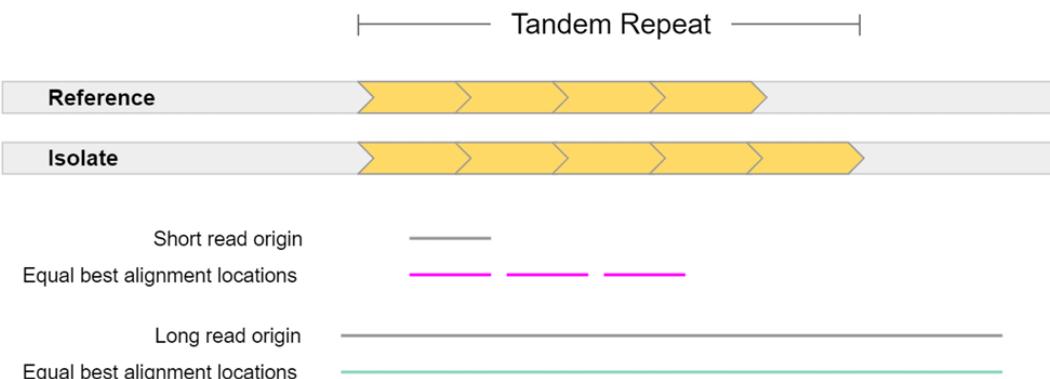


The HiFi Pacbio long reads give the correct two multi-allelic DELs of sizes 97 bp and 62 bp



Long read applications

- *De novo* genome Assembly
- Haplotype-resolved genome analysis
- Structural variant (SV) discovery
 - Repetitive SVs
 - Complex SVs
- Resolving genome structure
 - Derivative chromosomes in cancer



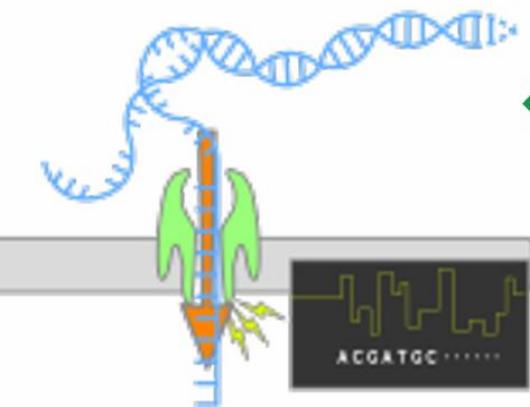
➤ Pan-genome for Structural Variants
(SVs)

Long-reads and T2T references for SV discovery

Short-reads: 100bp-300bp



Long-reads: 1,000bp-20,000Kbp, few >>20Kbp

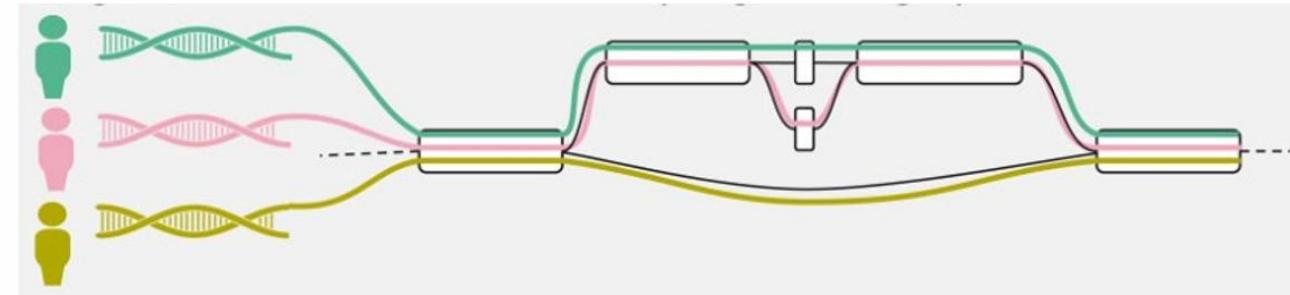


Nanopore sequencing

Linear reference genome (GRCh38)



Graph pan-genome



Human PanGenome
Reference Consortium



Pan-genomic Alignment Example

(a)

Ref.

```
ACGGTTAAGGGCGATCG--CTCGTTTT  
ACGGTTAAG--CGATCG--CTCGTTTT  
ACCGTTAA---GATCGAACTCG---  
ACCGTTAAGGGCGATCGAA---TTTT
```

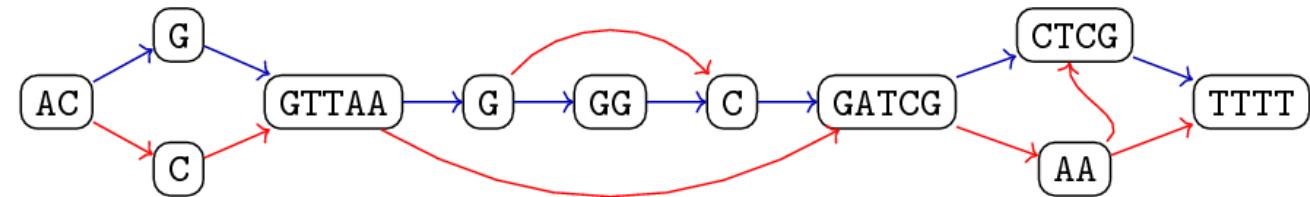
Reads:

```
ACCGTTAAGCGA  
TCGAATT
```

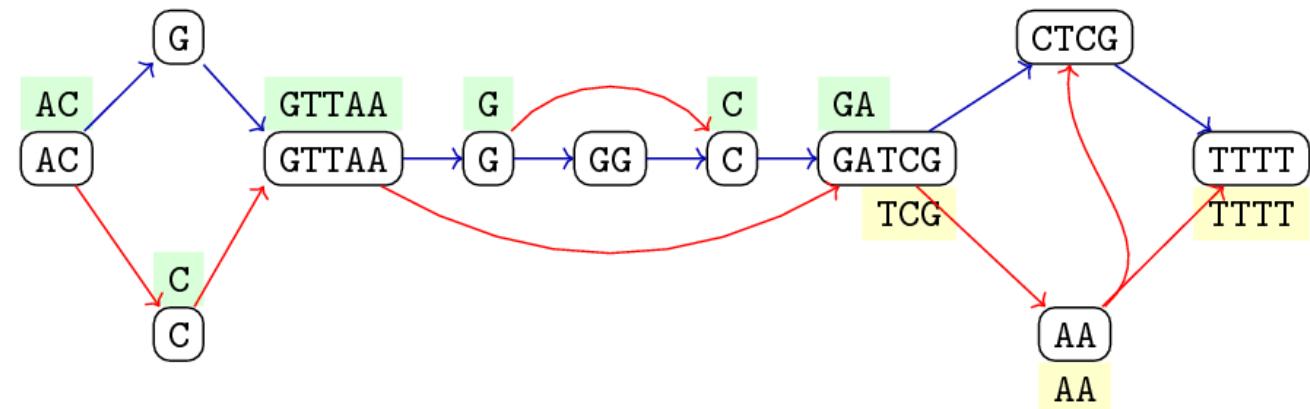
(c)

```
ACCGTTAAGCGA  
ACGGTTAAGGGCGATCGCTCGTTTT  
TCGAA--TTT
```

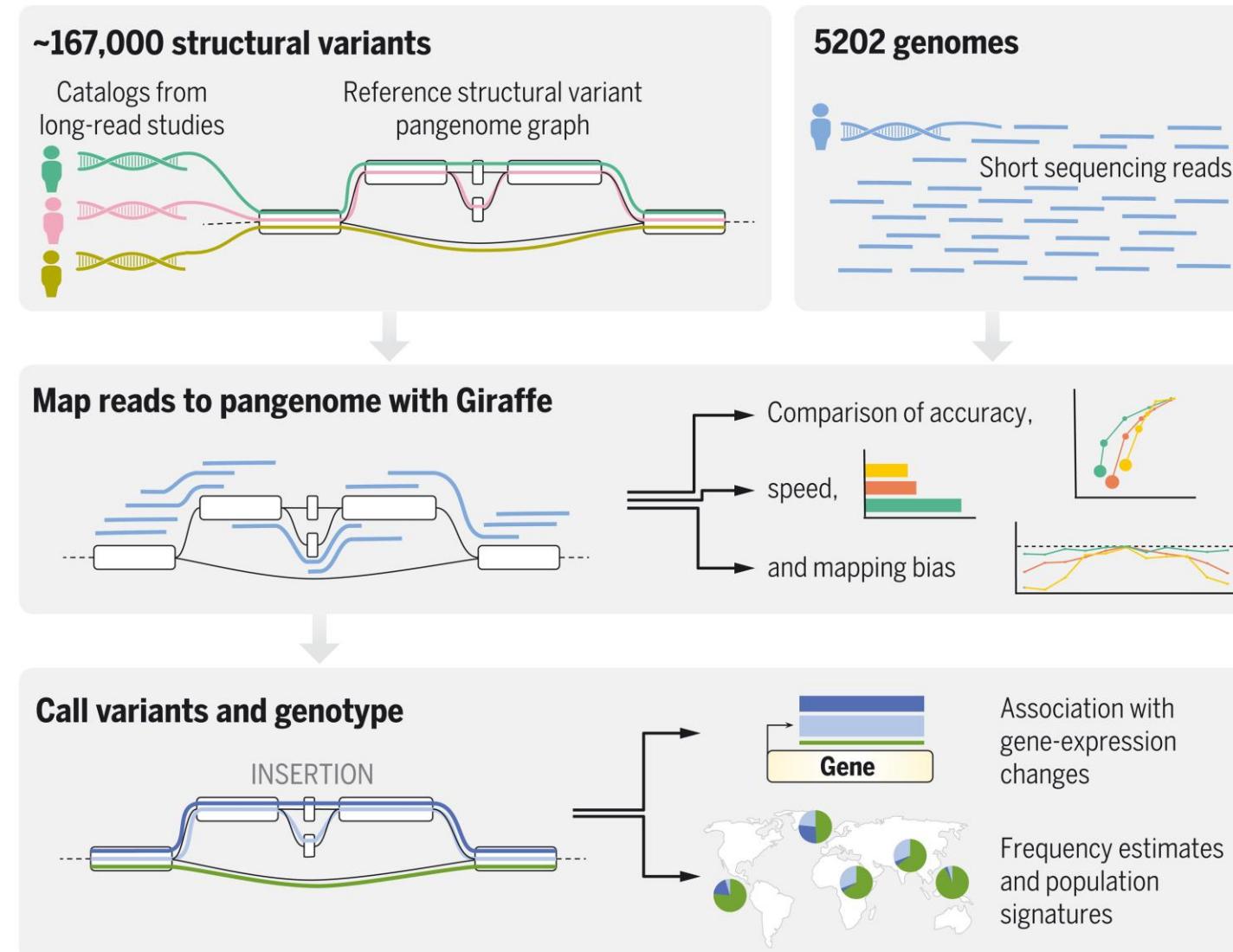
(b)



(d)

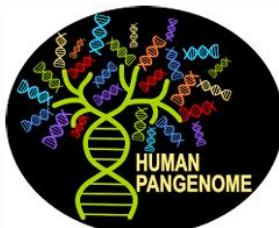
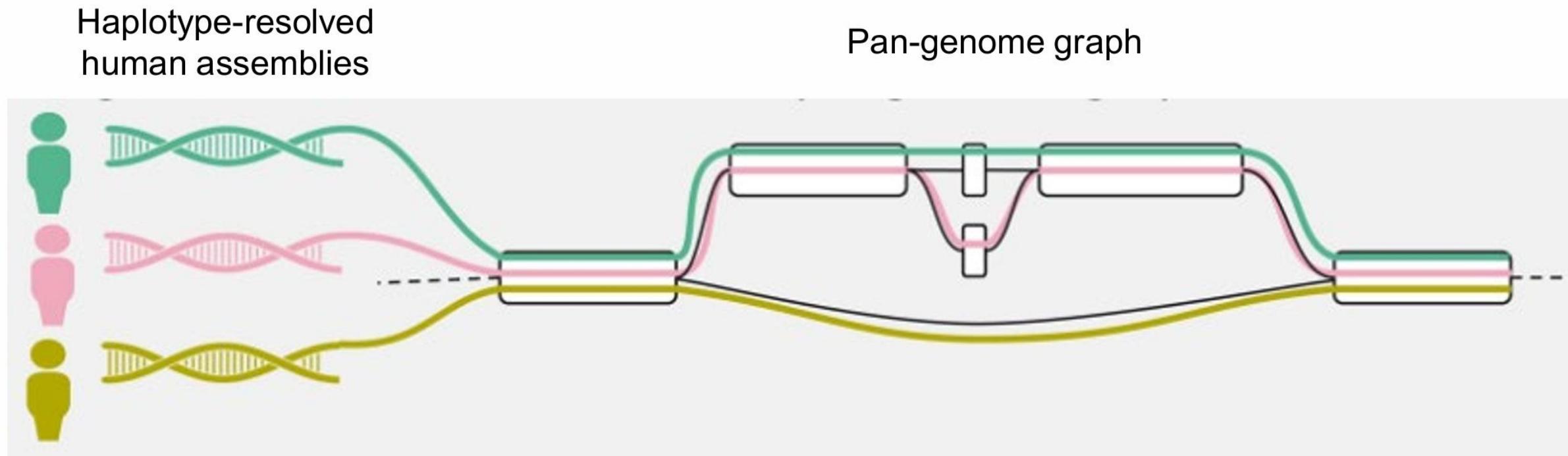


Pangenomics enables genotyping of known structural variants in 5202 diverse genomes



Pan-genome graphs

- A succinct representation of a set of reference genomes



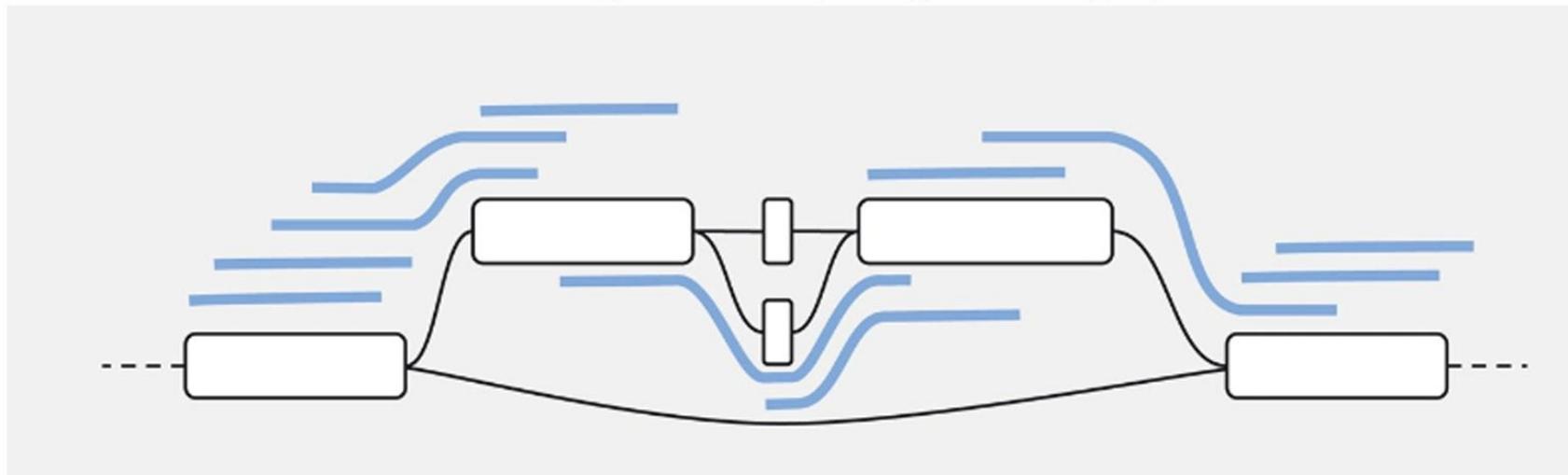
Human Pangenome Reference Consortium (HPRC)

- 44 samples (88 haplotypes) + GRCh38 + CHM13

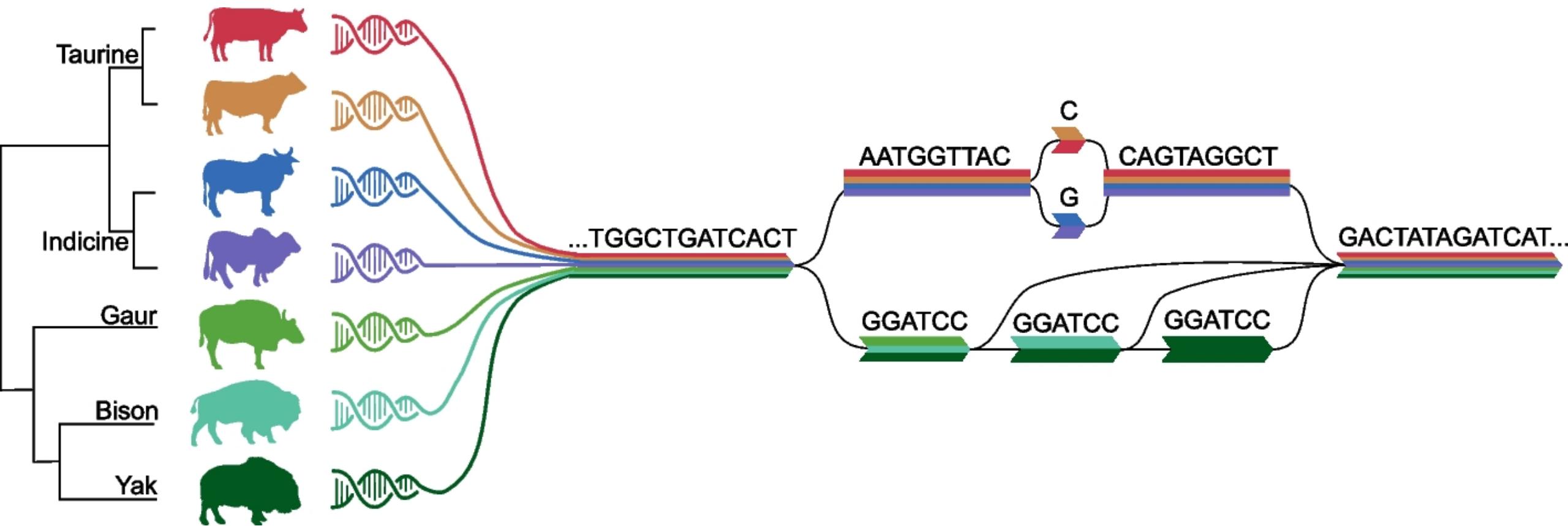
Pan-genome graphs

- HPRC pan-genome graph: 90 haplotypes (44 samples, GRCh38, CHM13)
- How to incorporate all types of variation?
 - Coarse-grained pangenome graph (structural variants only)
 - 751M on disk: **391,950 segments** (S); 566,204 links (L); 3,198,196,033bp
 - Fine-grained pangenome graph (including small variants)
 - 8.6G on disk: **81,415,956 segments** (S); 112,955,105 links (L); 3,287,932,785bp

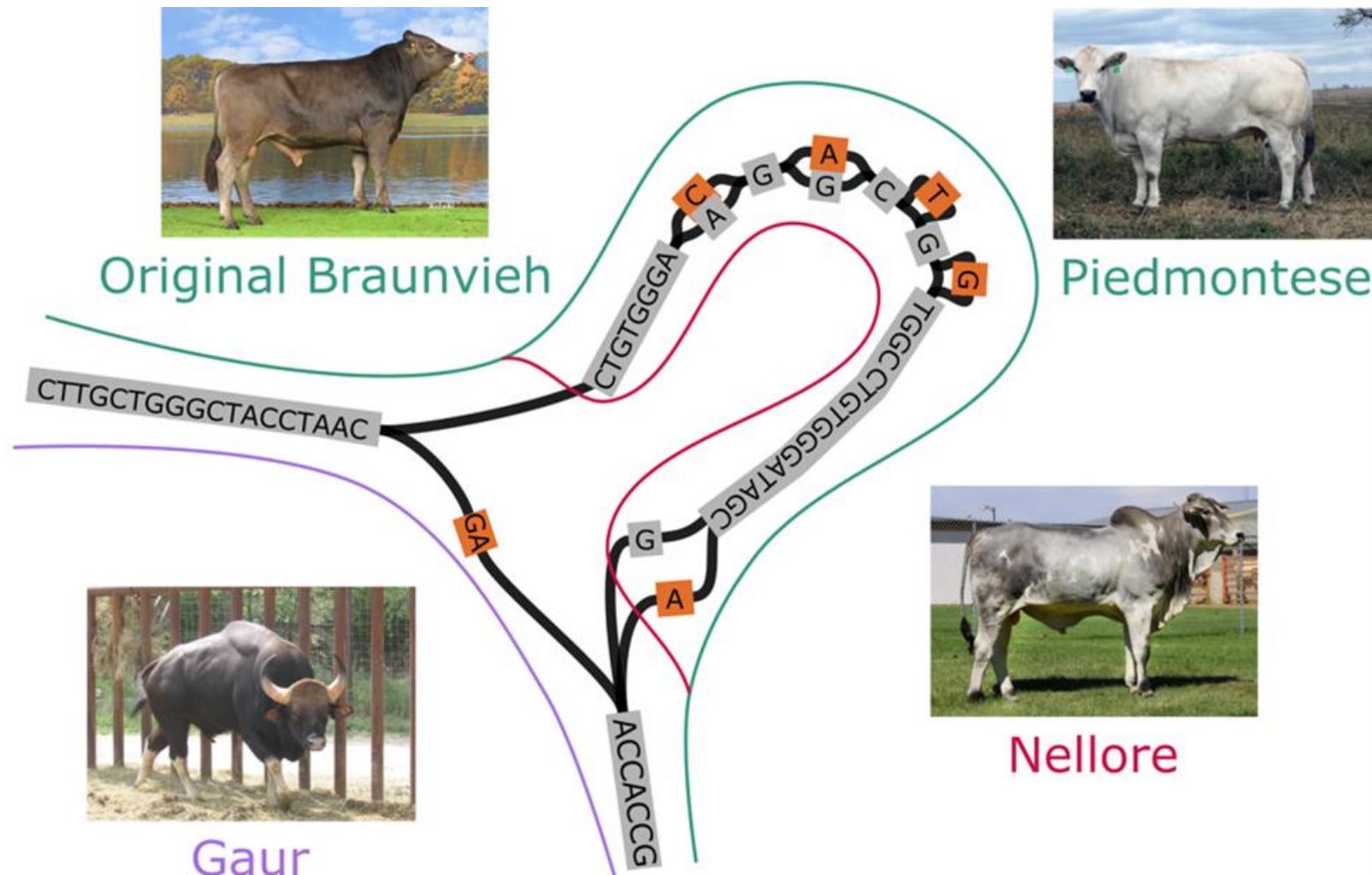
Alignment to pan-genome graph



Pan-genome graphs

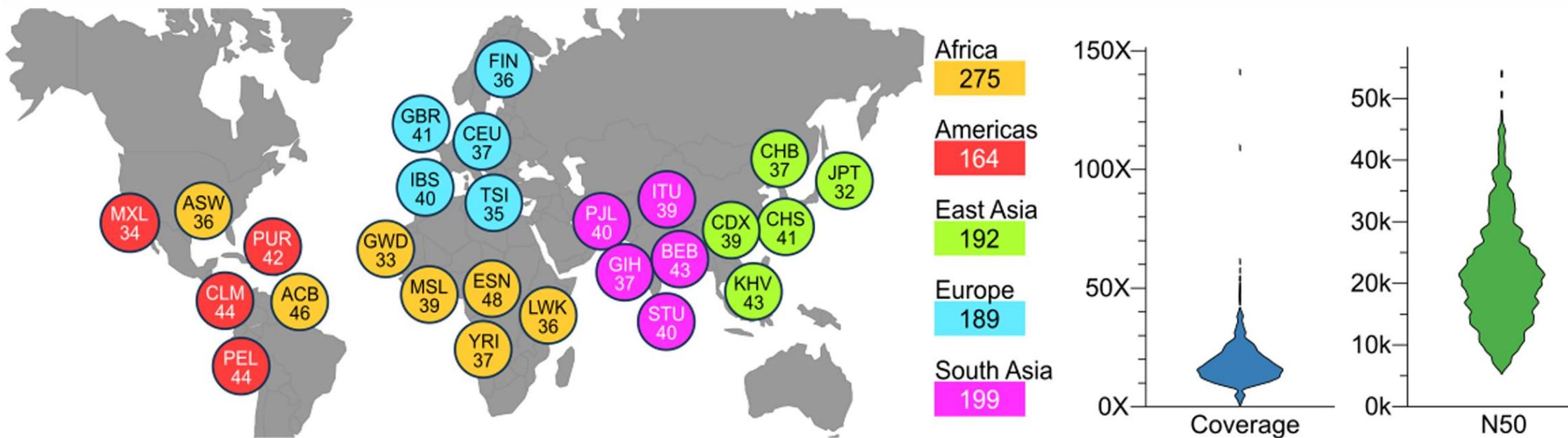


Pan-genome graphs



Genome variation discovery using long reads and graph genomes

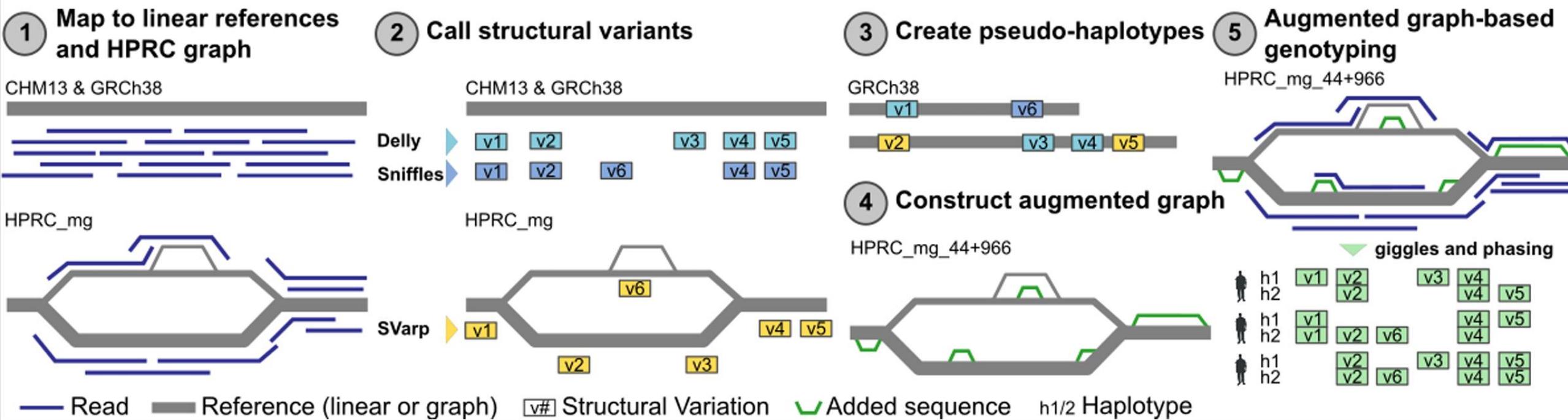
1000 Genomes ONT Project



1,019 samples sequenced with ONT

- ~15x coverage
- **Structural variant calling** using pan-genome graphs

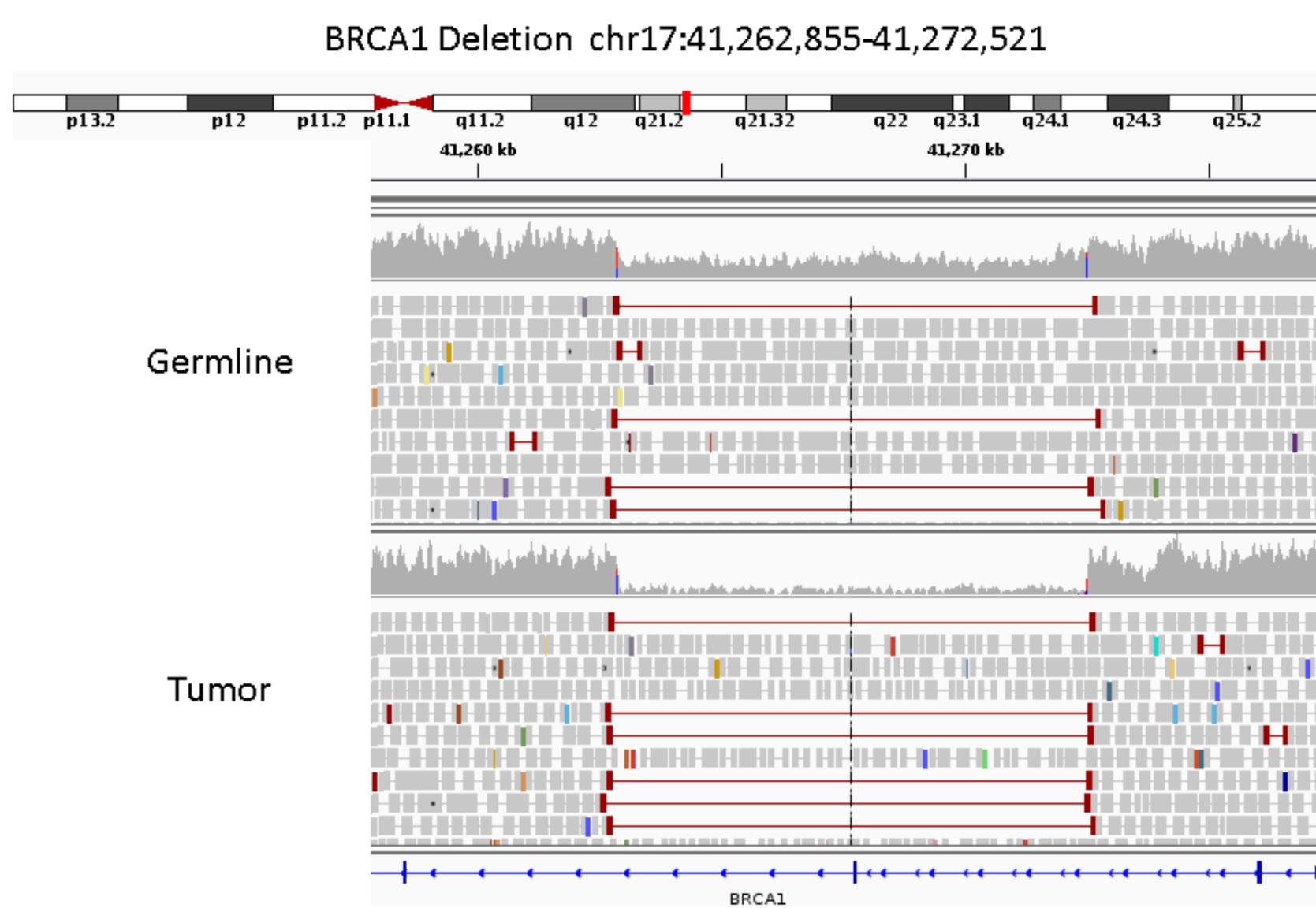
Variant Calling Strategy





Cancer Predisposing SVs

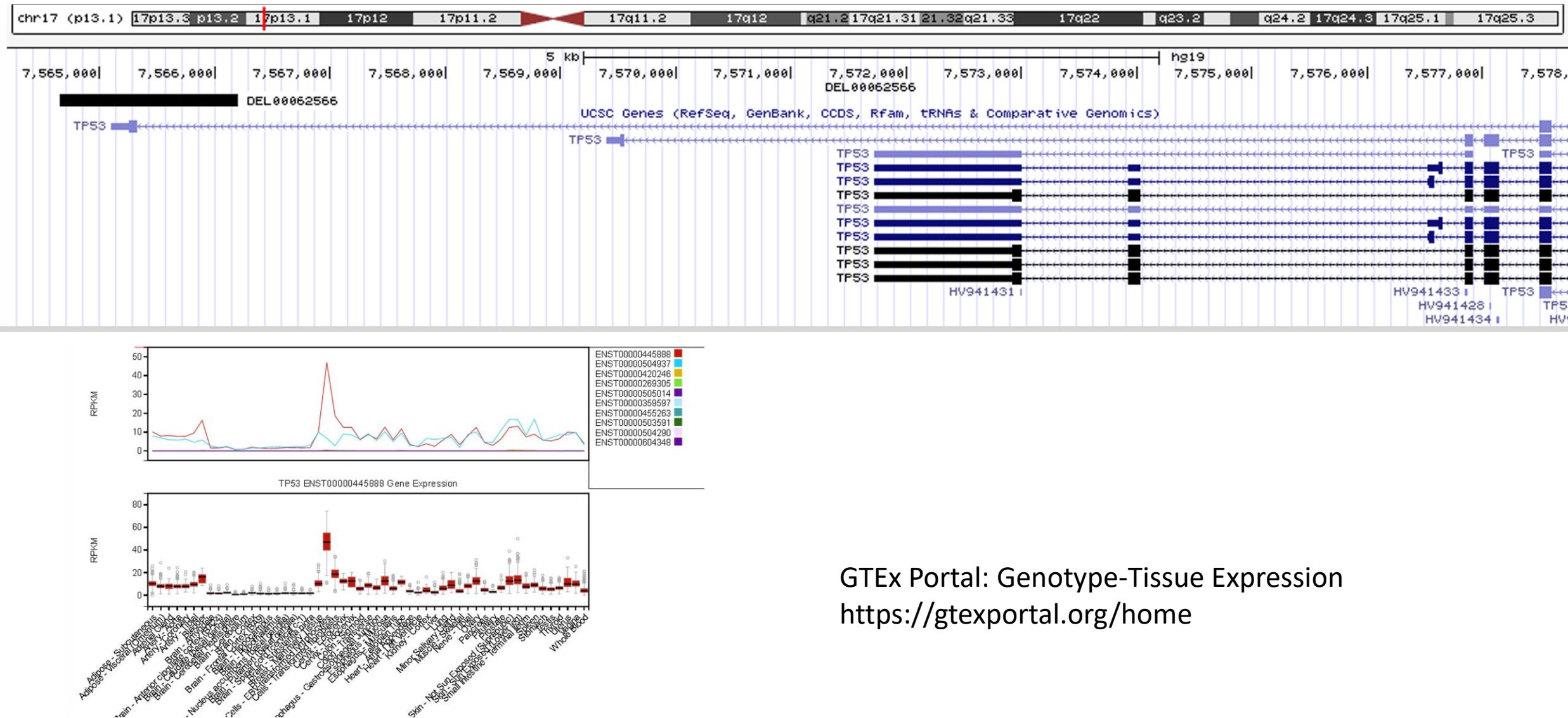
Copy Number Variants (CNVs)



Hypothetical Example: Deletion that affects geneX

- Pan-cancer cohort (e.g. 300 breast cancer samples)
 - 18 out of 300 samples are a carrier
 - Allele frequency: ~6%
- 1000 Genomes cohort (2504 samples)
 - 5 out of 2504 samples are a carrier
 - Allele frequency: ~0.2%
- SV may confer a higher risk for breast cancer but be aware of many possible confounders!
 - Sex, Related individuals, Population structure, ...
 - All 5 carriers have European ancestry and the cohort of Europeans is much smaller than 2504 samples
 - Technical confounders: Low vs. high-coverage, different insert size, error rate

Structural variants affecting genes



Thank you!

Phuc Loi Luu, PhD

Email: Luu.p.loi@googlemail.com

Zalo: 090102182