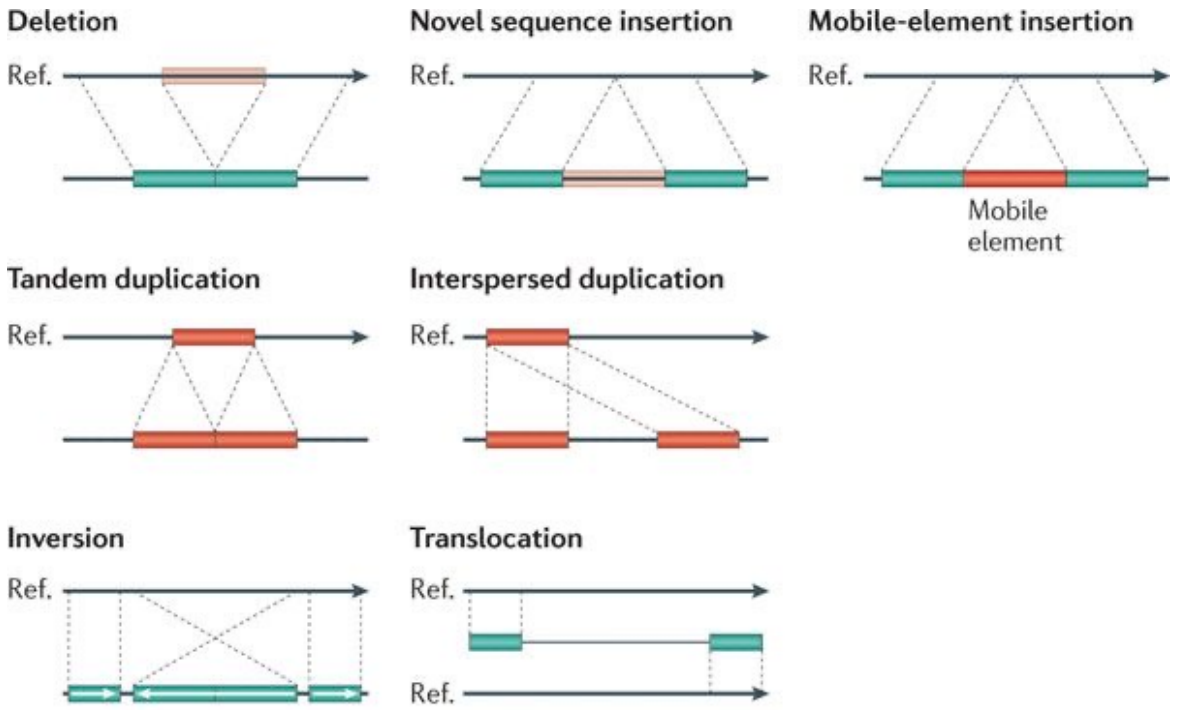
Other NGS applications: CNVs, Structural Variations

Erika Souche

GENOMICS CORELEUVEN

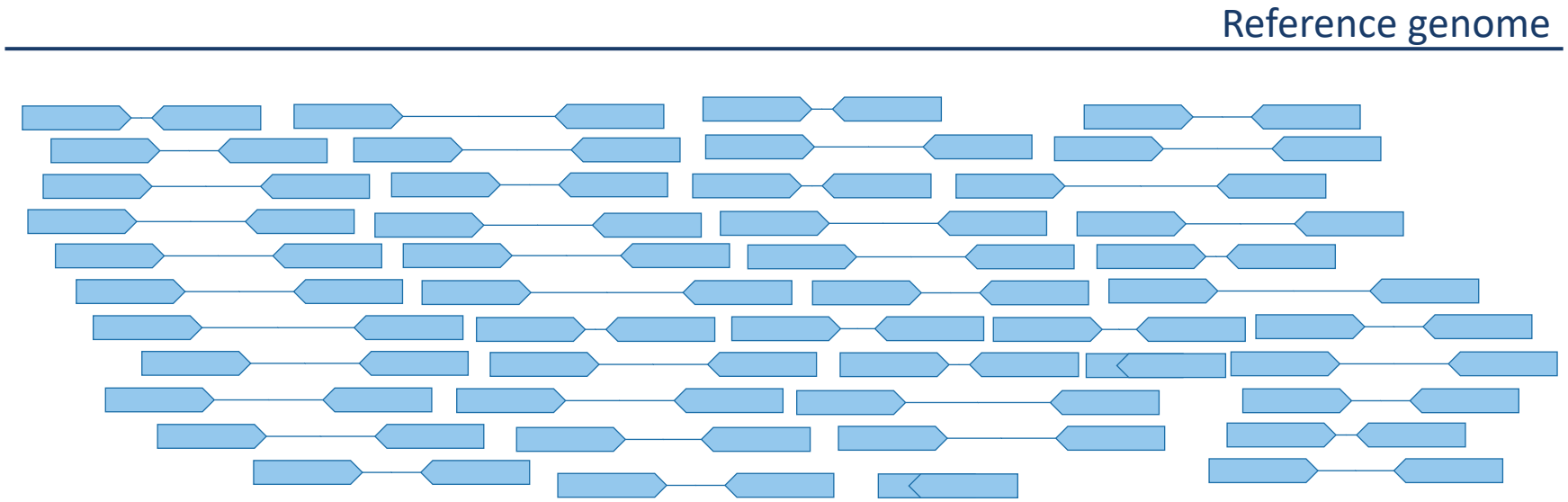
Definitions

- Structural Variant (SV)
  - Copy Number Variant (CNV)
    - Deletion
    - Duplication
    - Triplication
    - …
  - Insertion
  - Inversion
  - Translocation

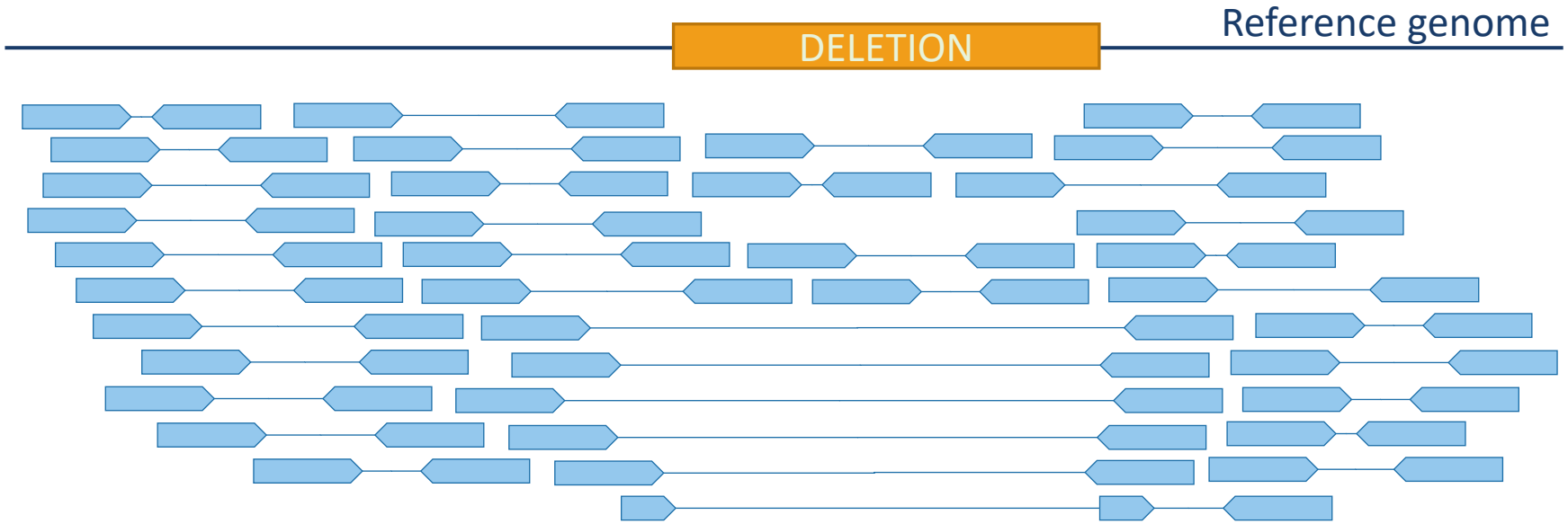


Alkan *et al.*, 2011 Nature Reviews | Genetics

CNV signatures

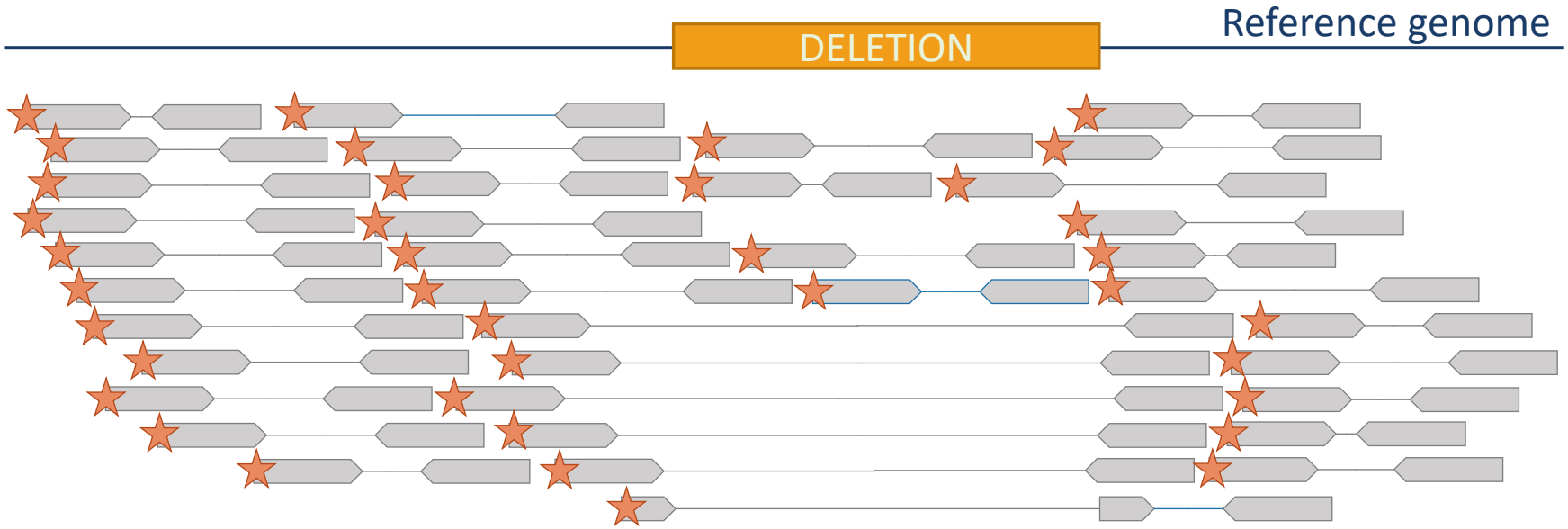


CNV signatures



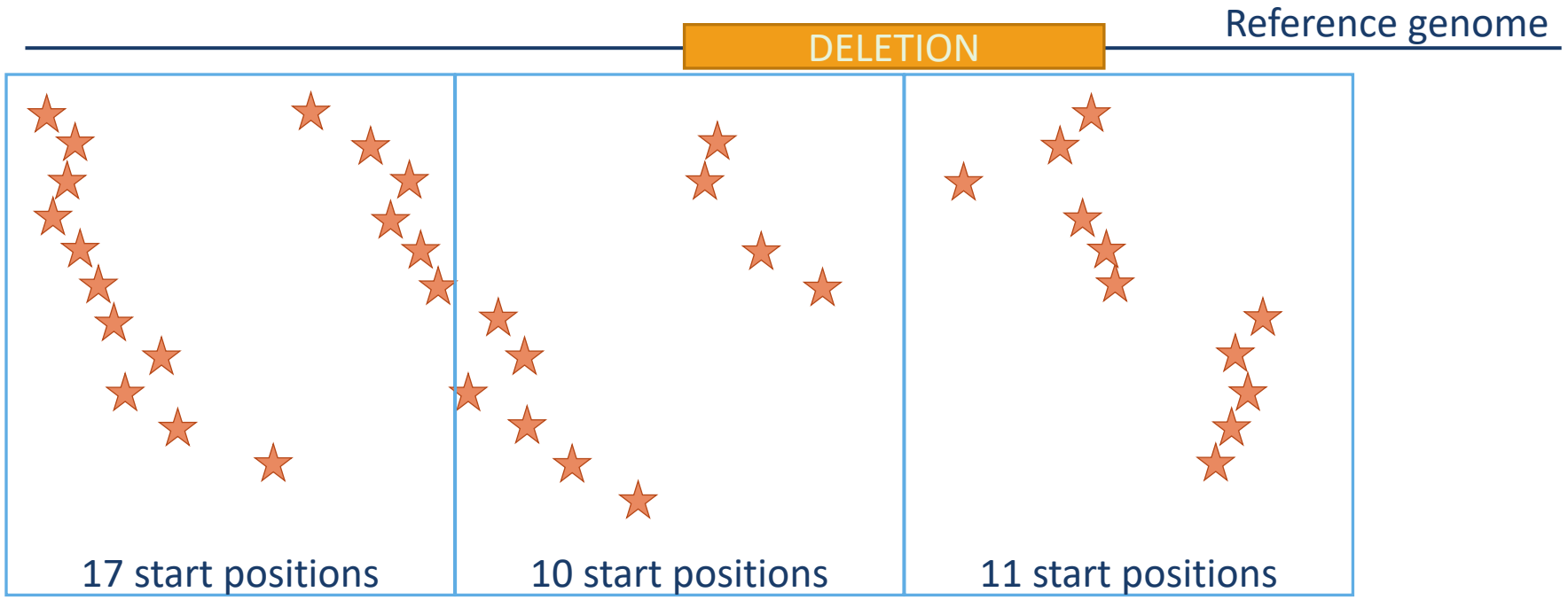
Detecting CNVs

- Based on Read Counts (RC)



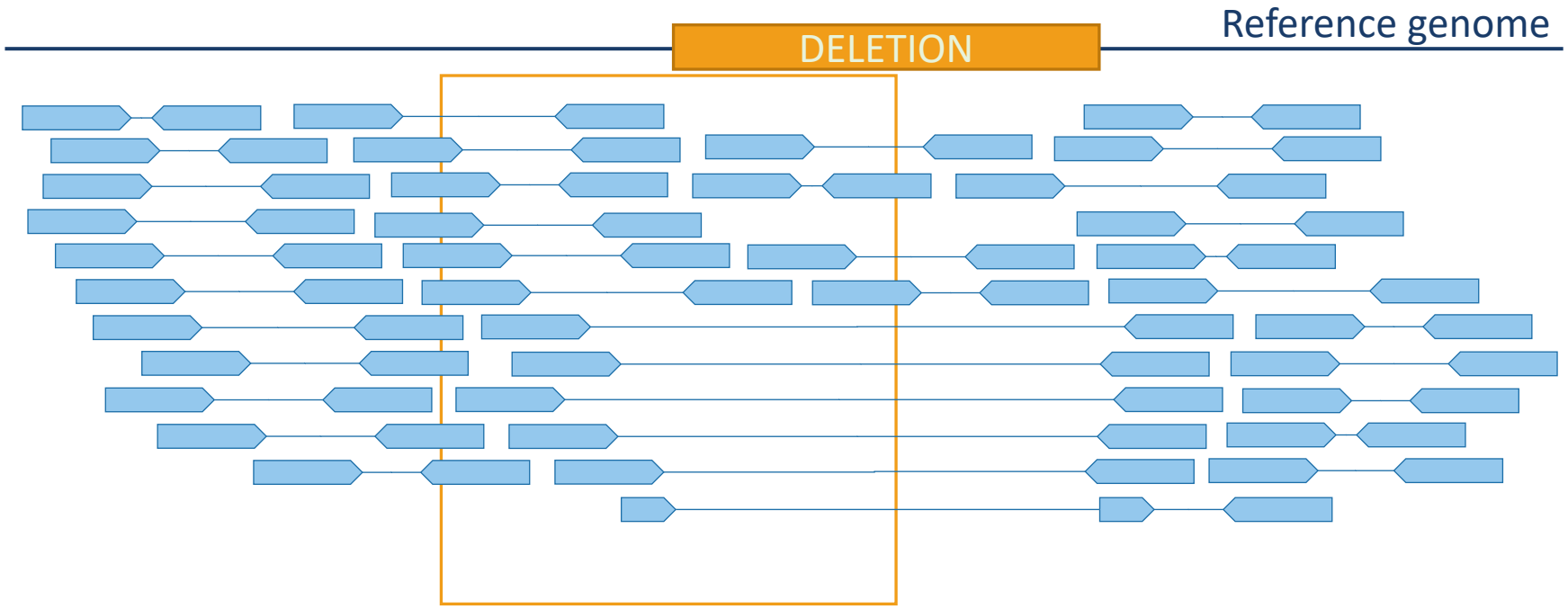
Detecting CNVs

- Based on Read Counts (RC)



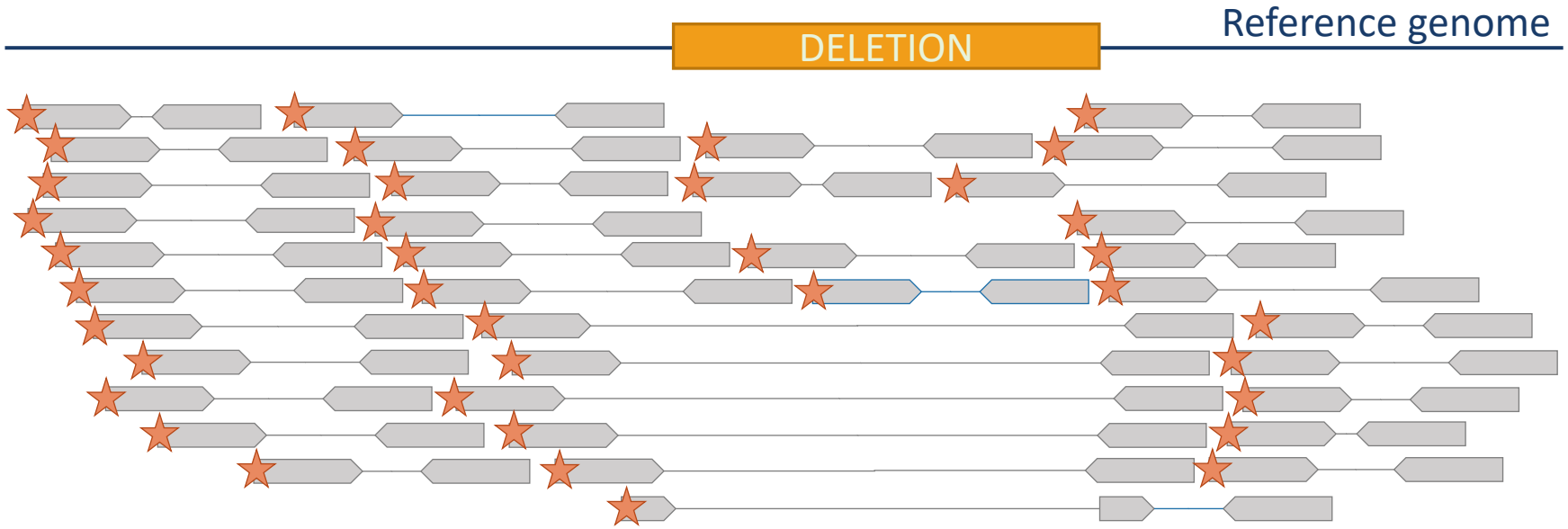
Detecting CNVs

- Based on Read Counts (RC)



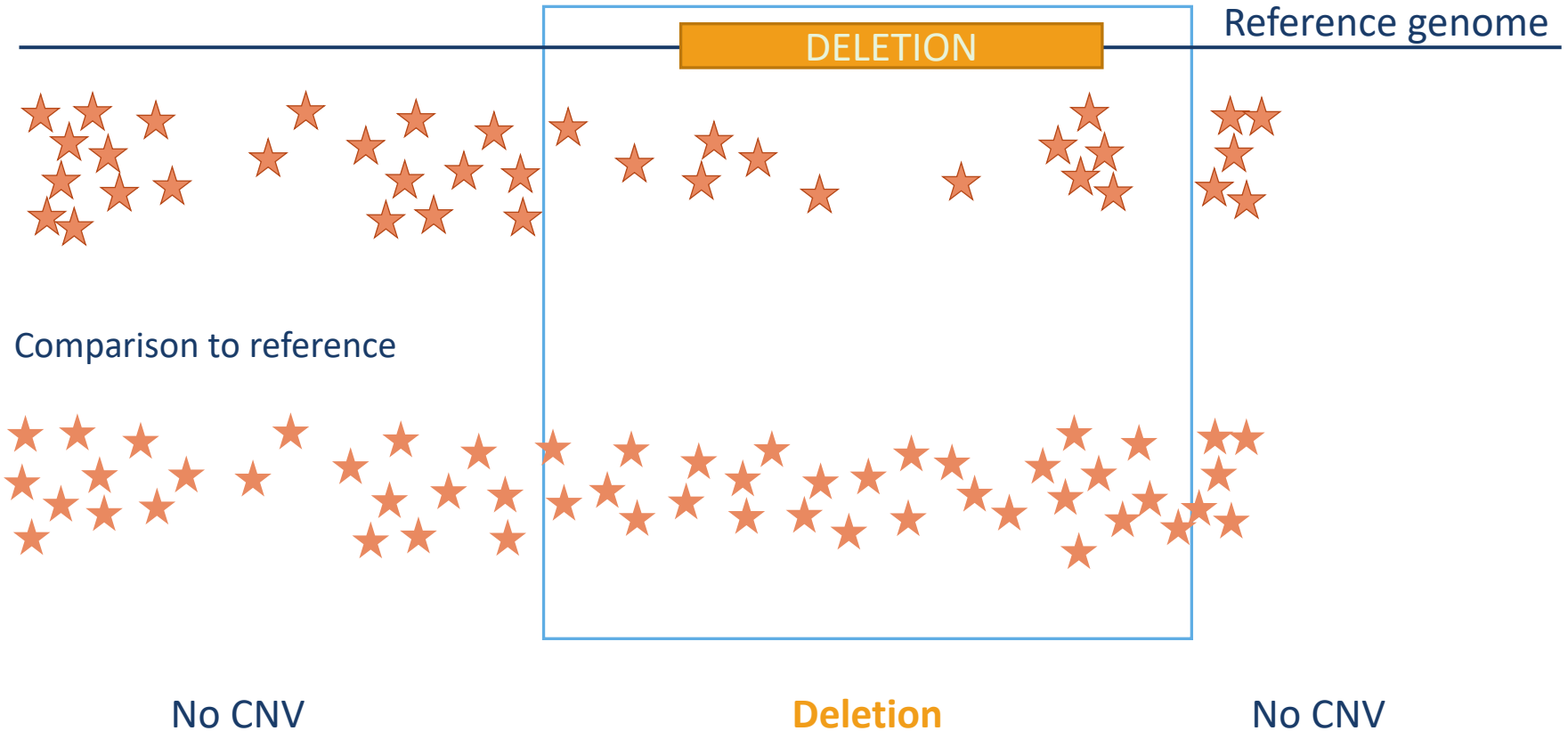
Detecting CNVs

- Based on Read Counts (RC)



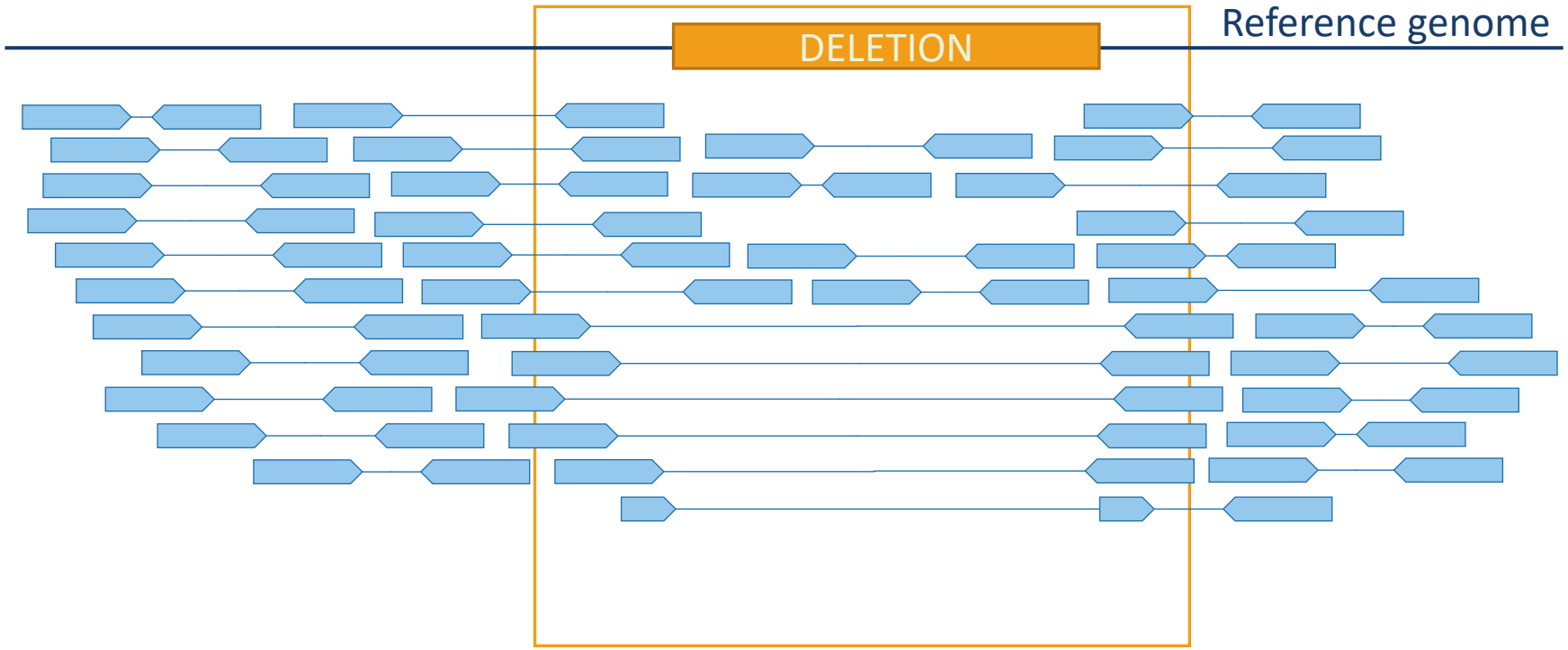
Detecting CNVs

- Based on Read Counts (RC)



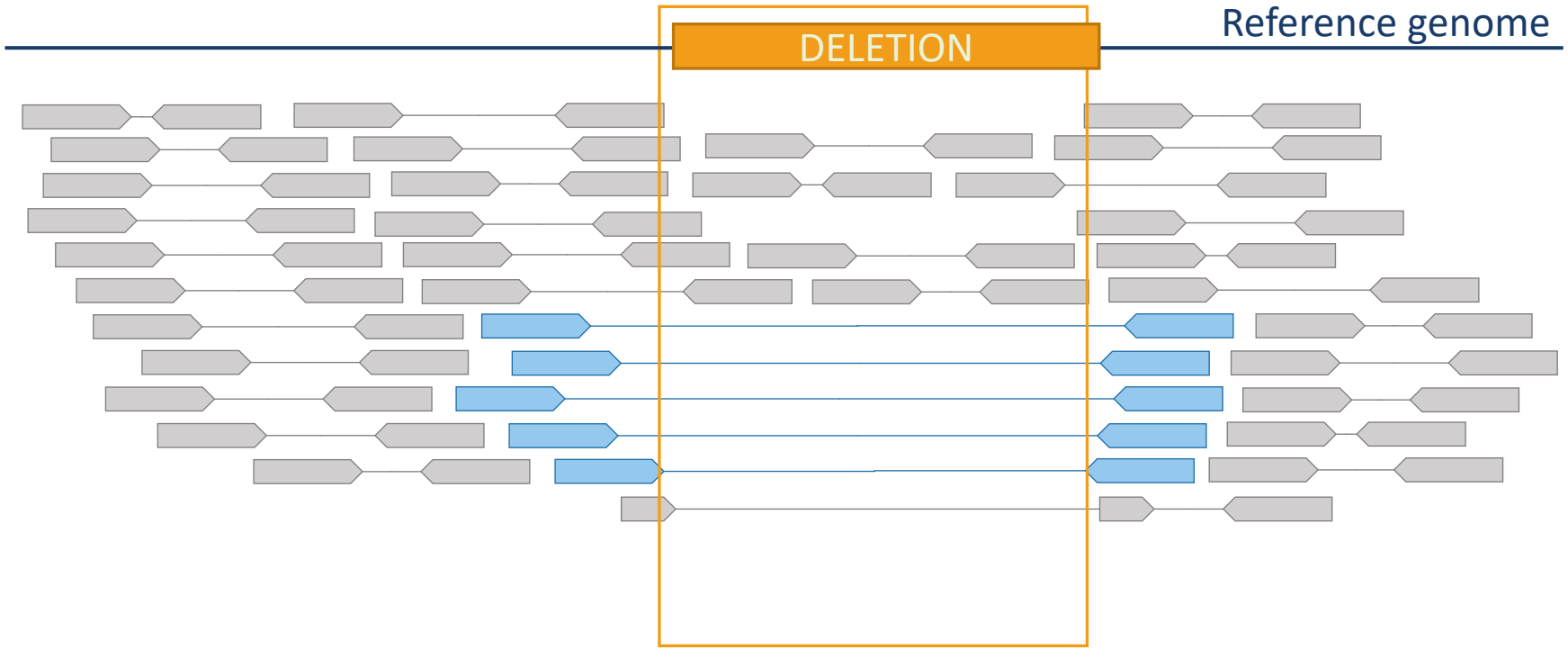
Detecting CNVs

- Based on Read Counts (RC)



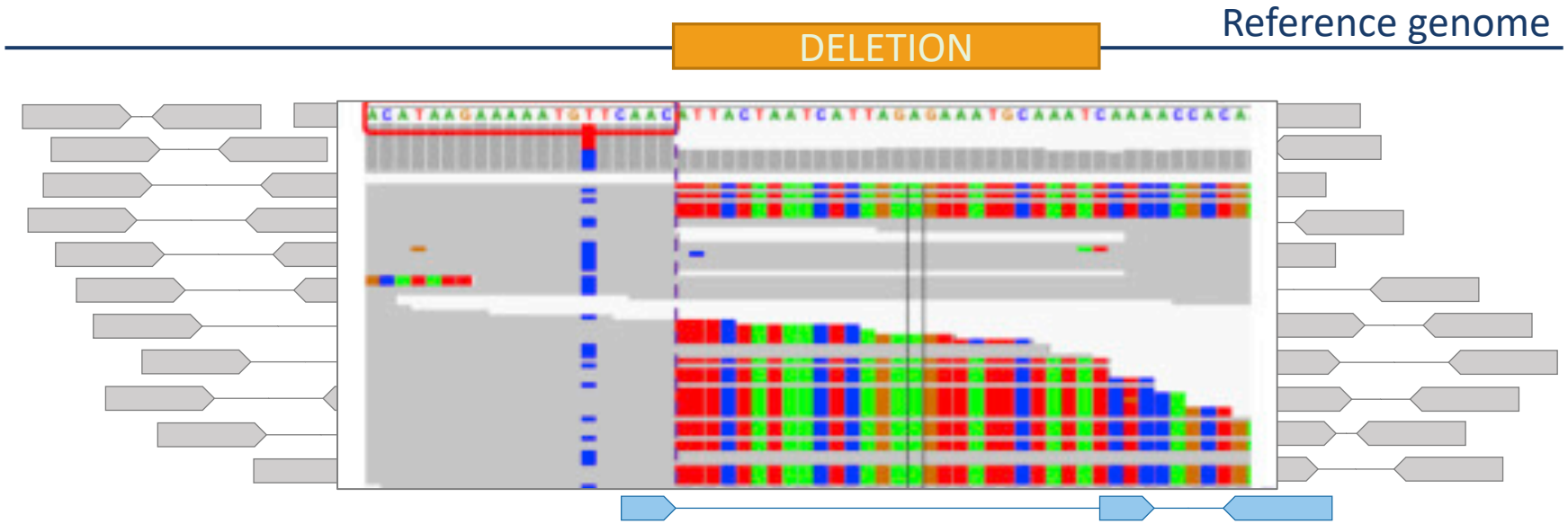
Detecting CNVs

- Based on Read Pairs (RP)



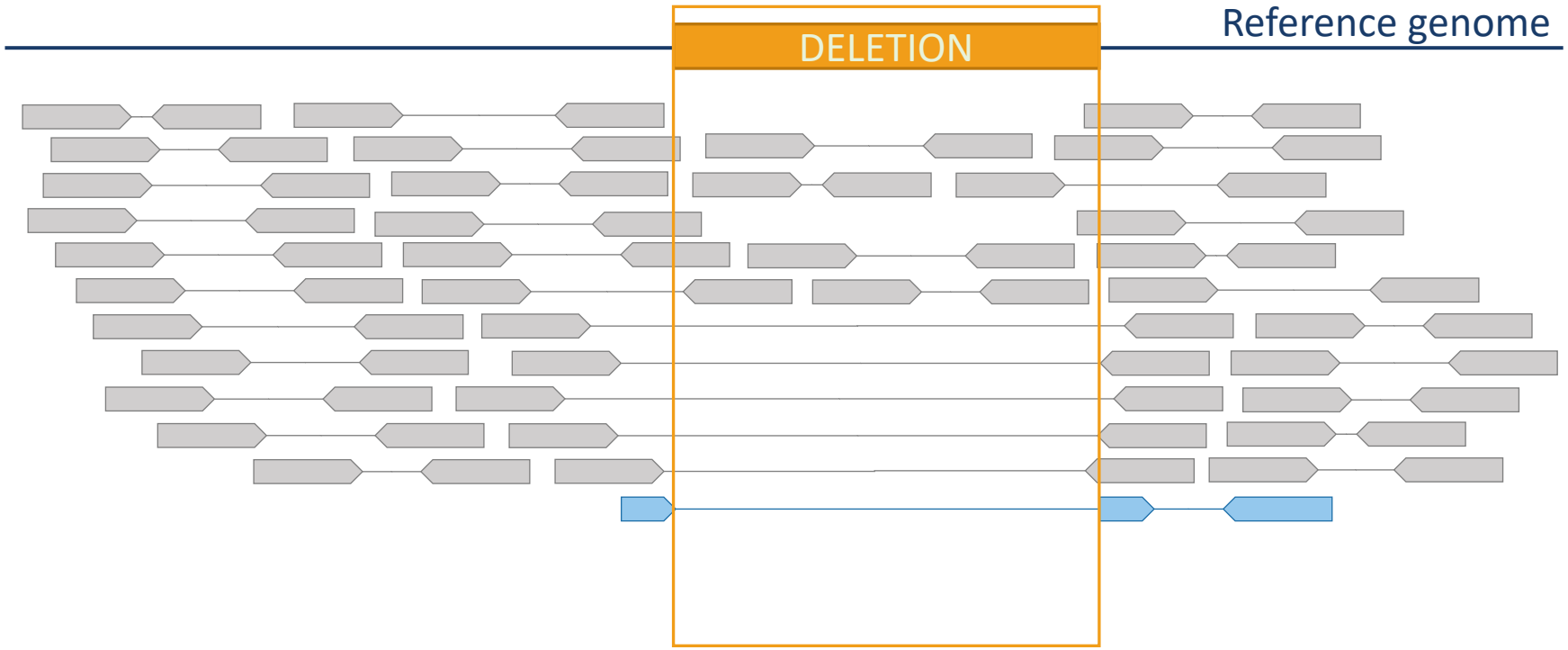
Detecting CNVs

- Based on Split Reads (SR)



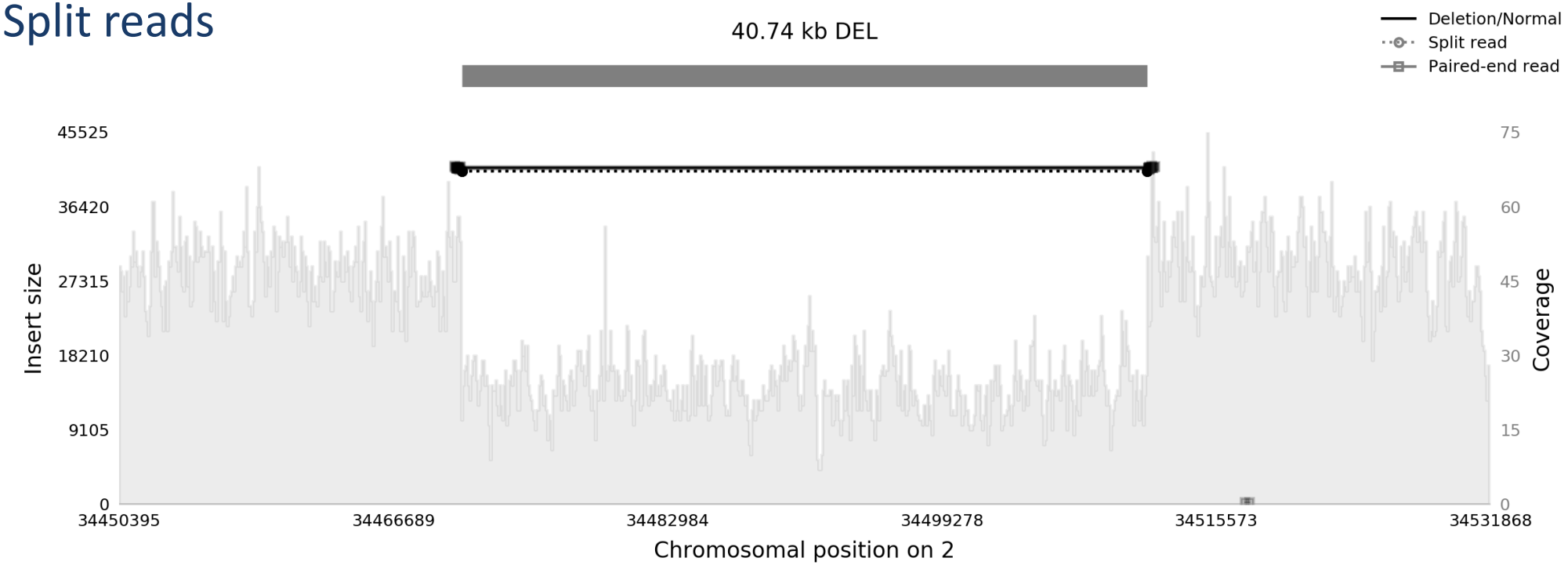
Detecting CNVs

- Based on Split Reads (SR)



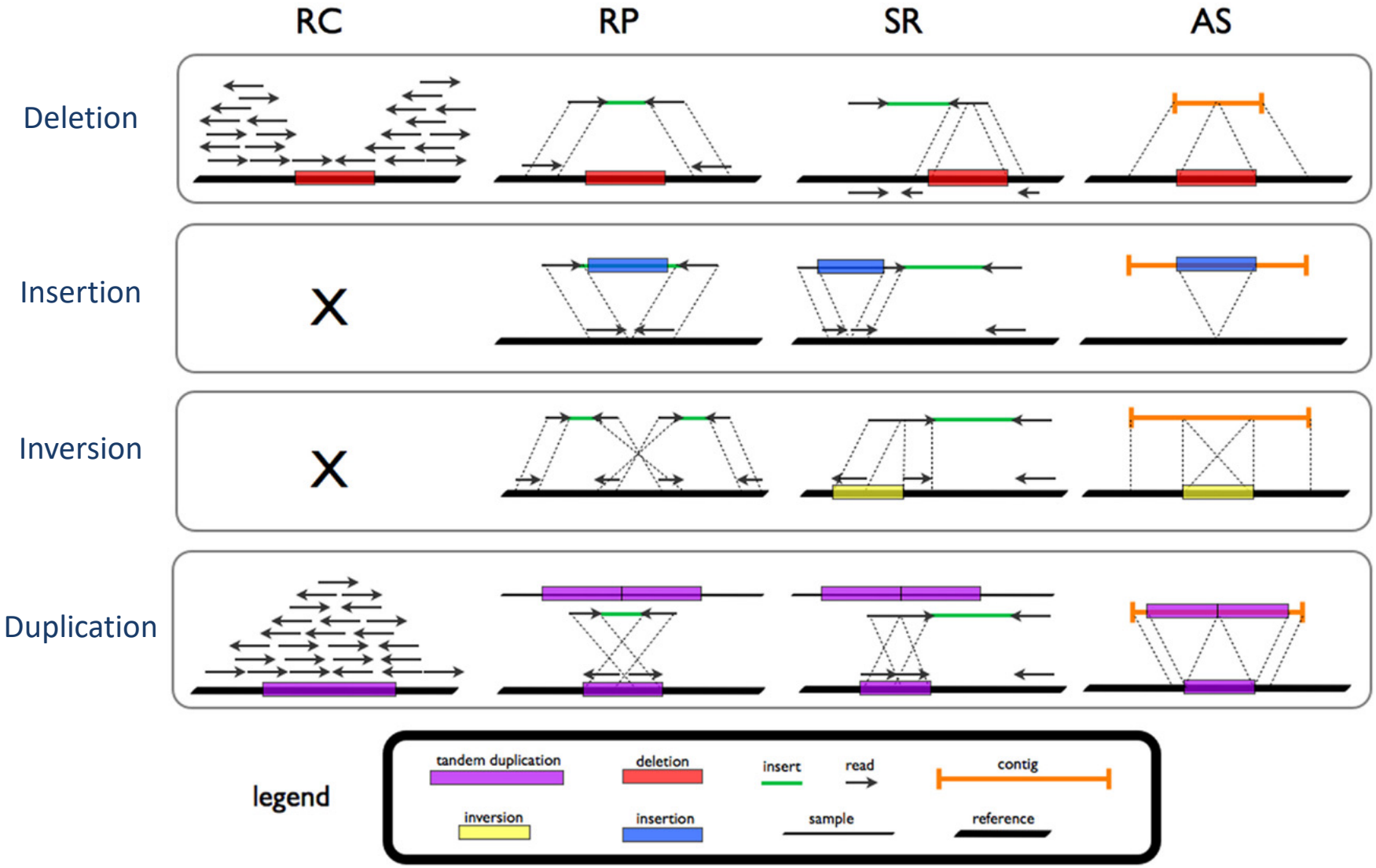
Detecting CNVs

- Deletion supported by
  - Read counts
  - Read pairs
  - Split reads



Plot generated by Samplot
Belyeu *et al.*, Genome Biology, 2021

Detecting SVs



Tattini *et al.*, Front. Bioeng. Biotechnol., 2015

Summary

- SV detection possible by WGS
- Resolution dependent on
  - Depth of coverage
  - Detection algorithm
- Limitations
  - Eventual need of reference
  - SVs in segmental duplications, repetitive regions
  - Complex SVs

SVs in targeted sequencing?

- Strategies
  - Read Counts
  - Read Pairs
  - Split Reads

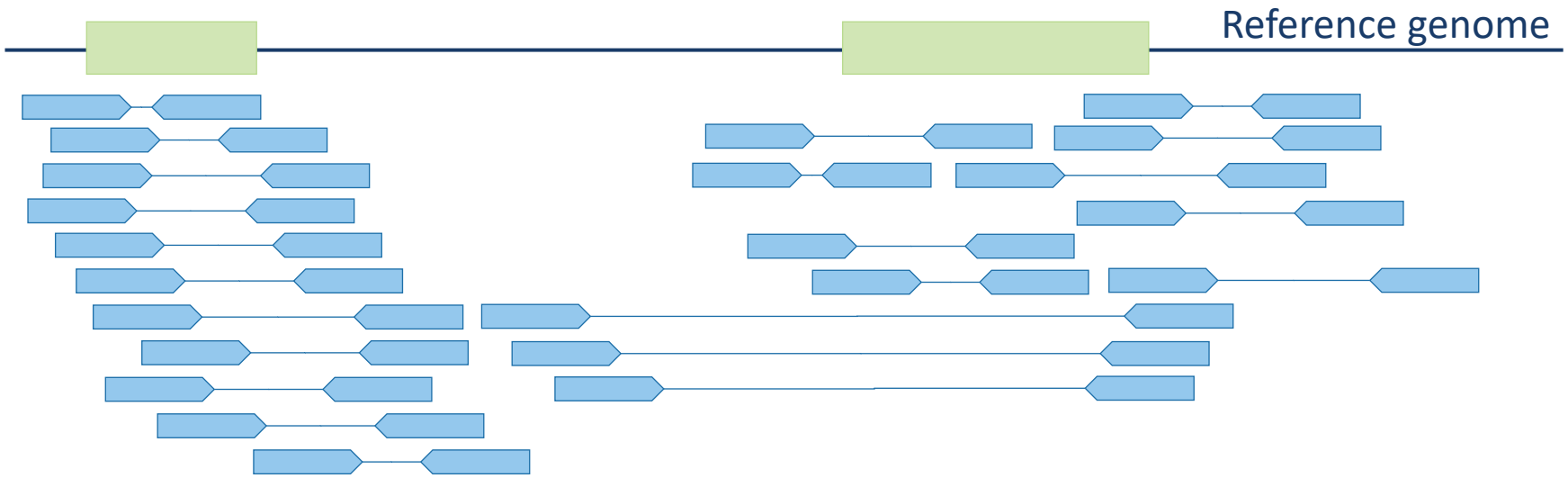
Read Counts: Requires reference set

Read Pairs, Split Reads: Only if breakpoints are part of/close to targeted regions

- Limitations
  - Bias introduced during capture/PCR
  - Resolution dependent on targeted regions
  - No information on breakpoints

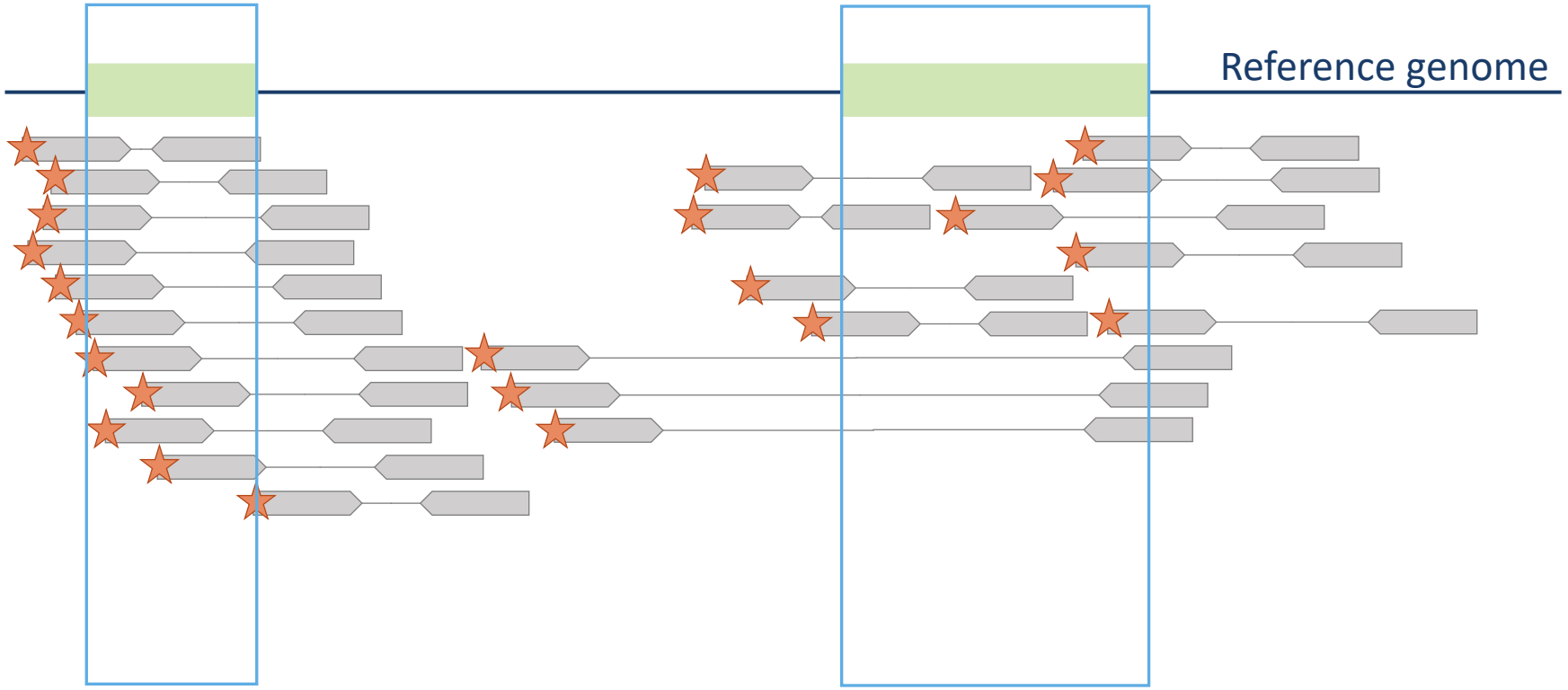
SVs in targeted sequencing

- Capture sequencing



SVs in targeted sequencing

- Capture sequencing

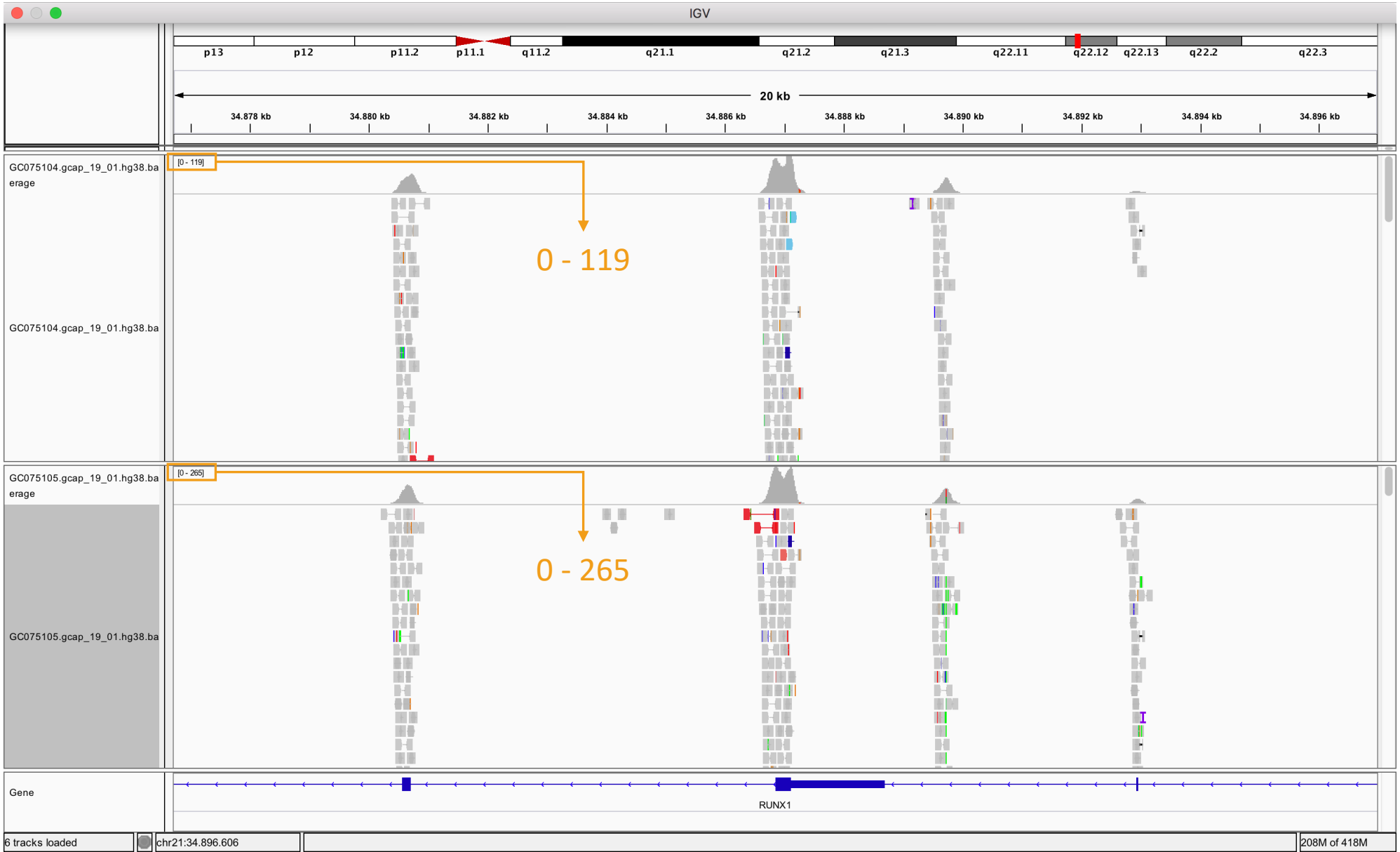


SVs in targeted sequencing

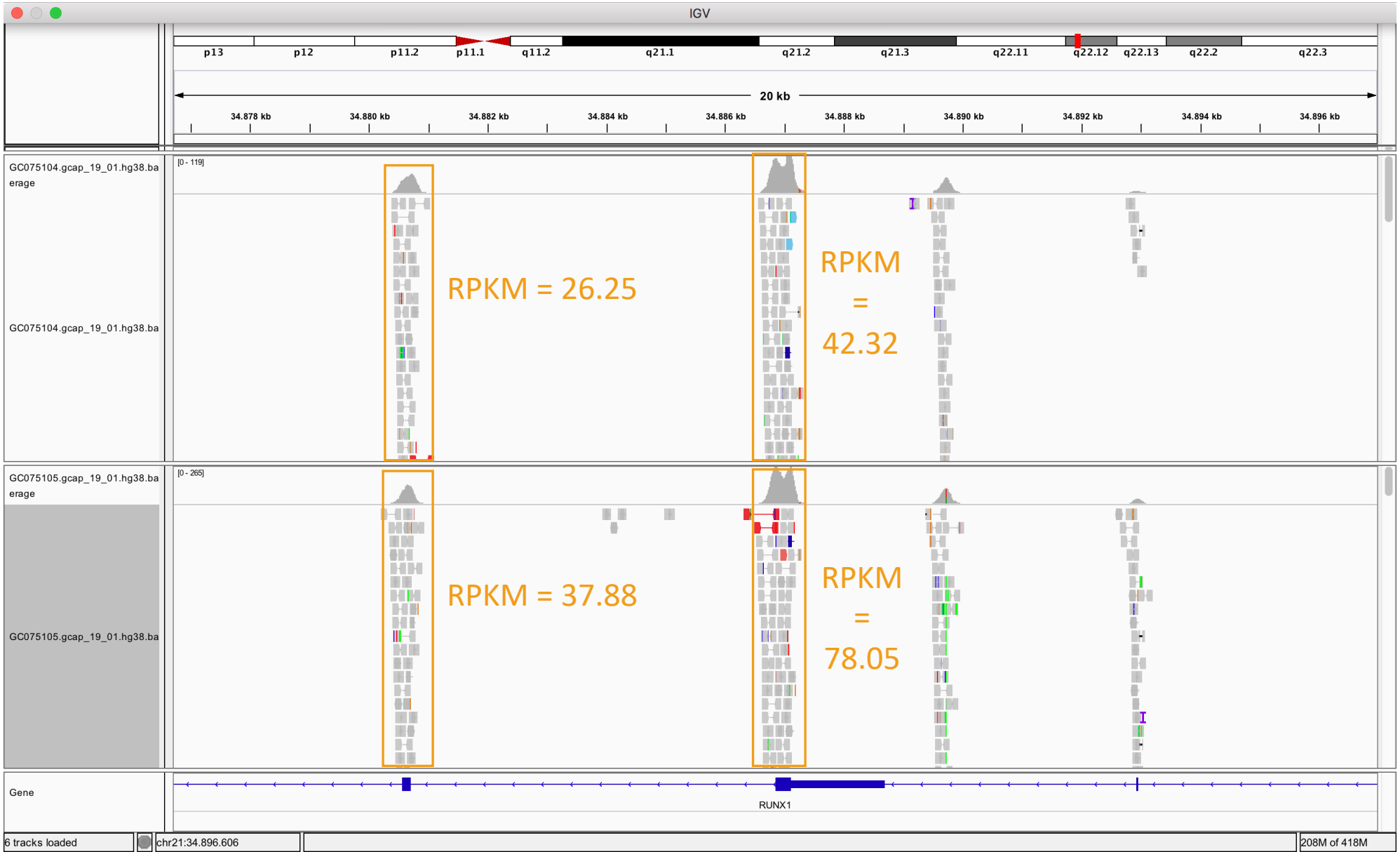
- Capture sequencing
  - RPKM
    Reads per thousand bases per million reads sequenced
  - 1 RPKM for each capture target/probe

$$RPKM = \frac{Read\ starts\ * 10^9}{Total\ reads\ *\ Target\ length}$$

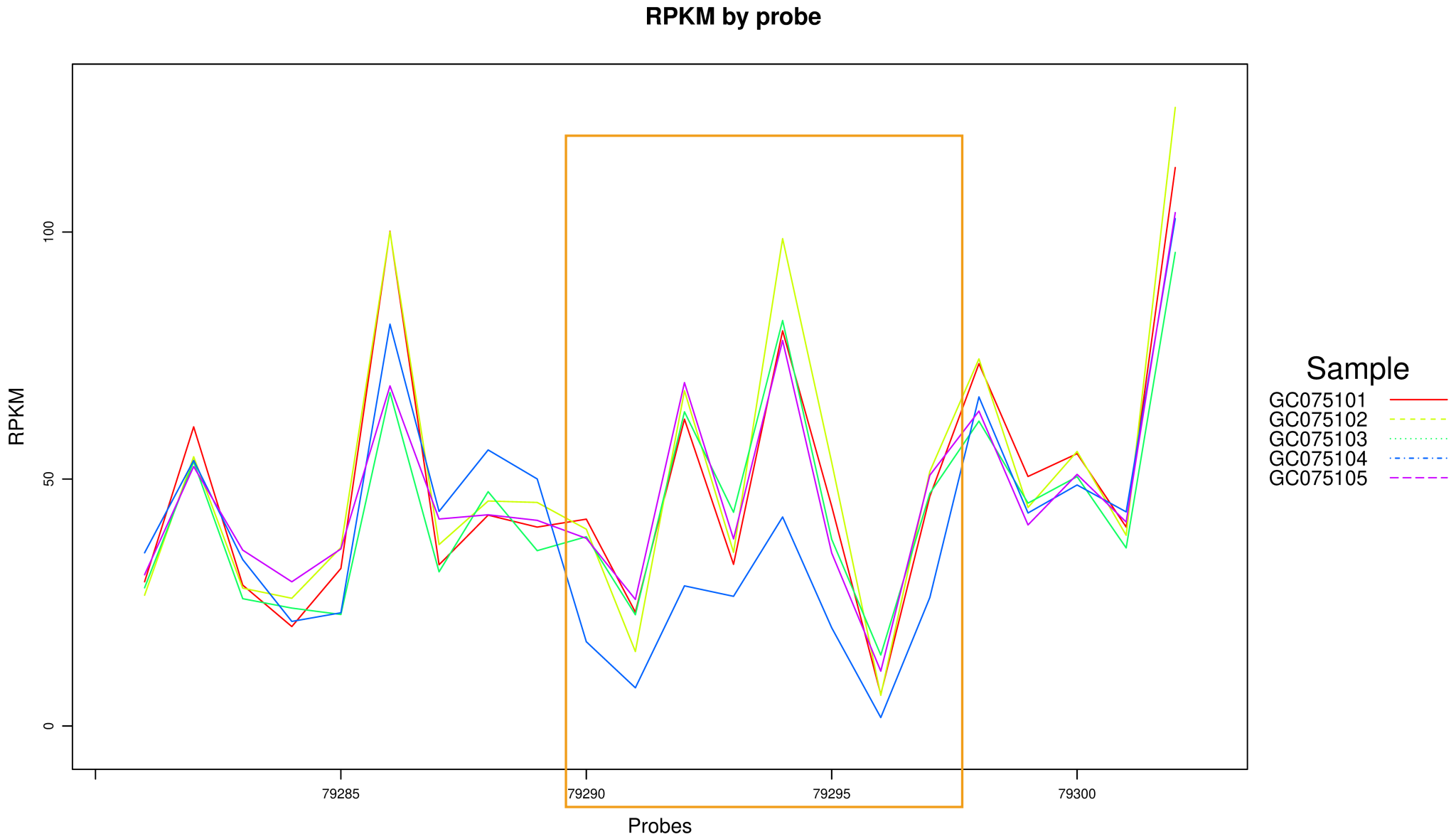
SVs in capture sequencing



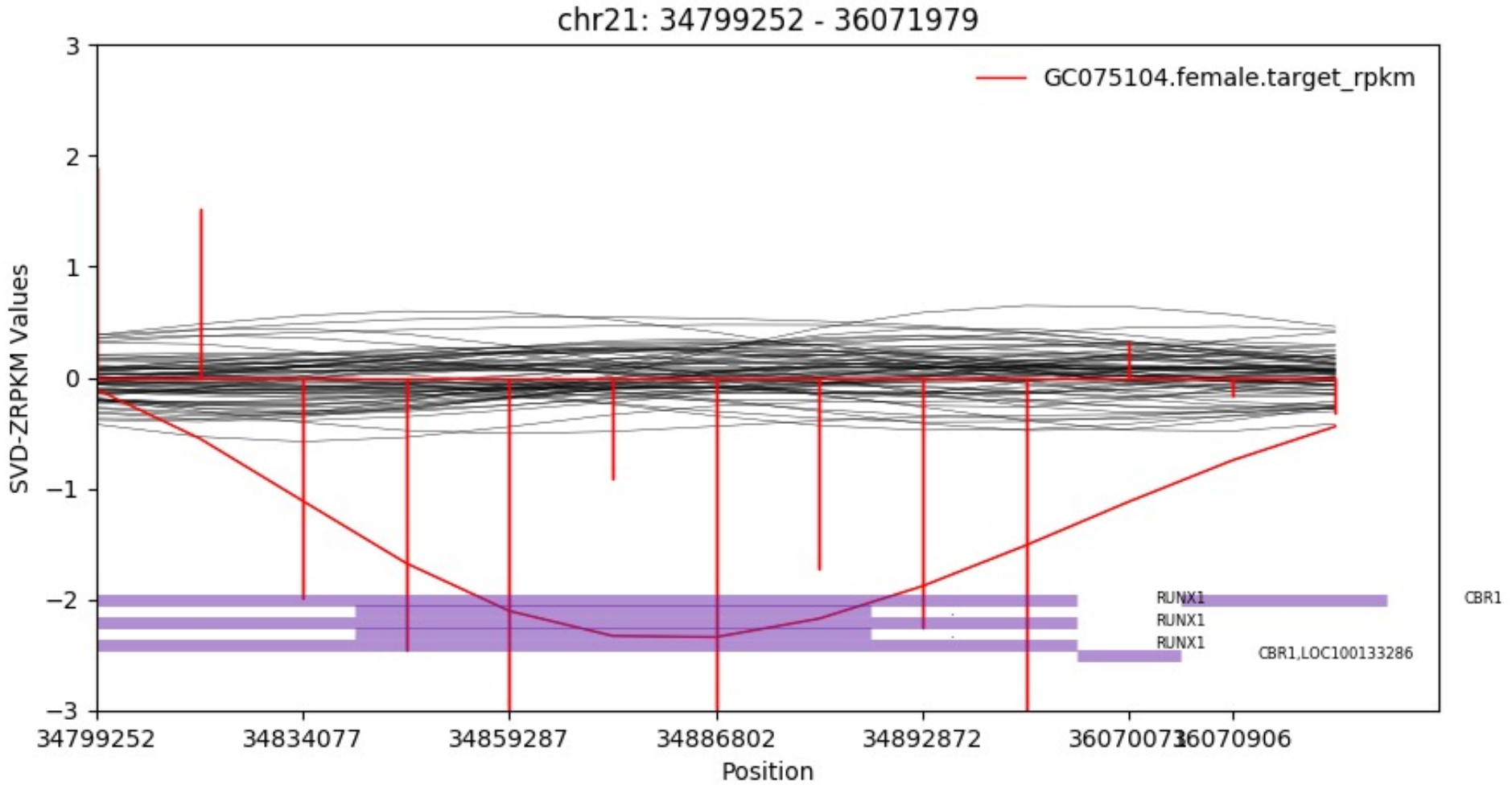
SVs in capture sequencing



SVs in capture sequencing

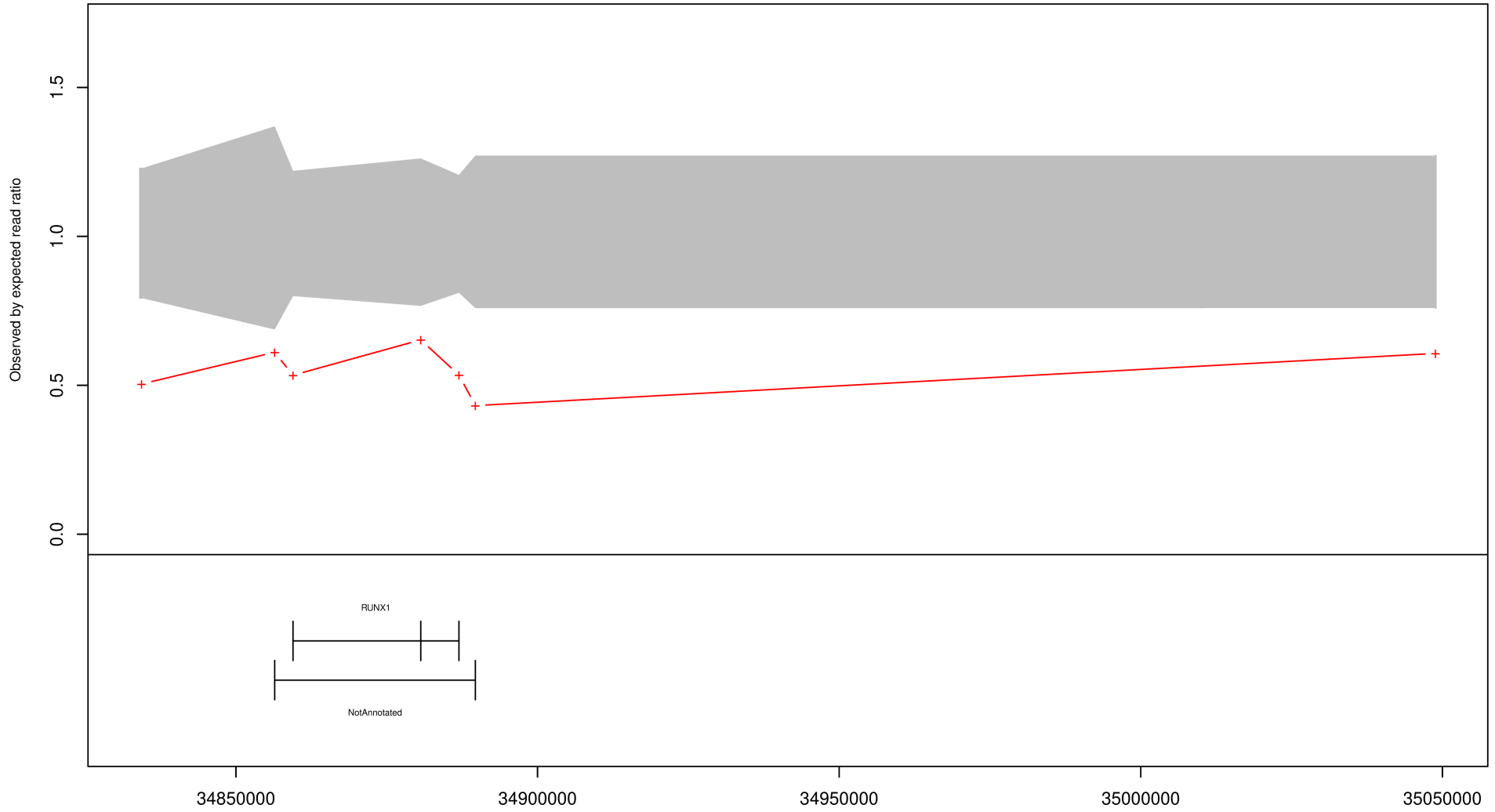


SVs in capture sequencing



Plot generated by CoNIFER
Krumm *et al.*, Genome Research, 2012

SVs in capture sequencing



Plot generated by ExomeDepth
Plagnol *et al.*, Bioinformatics, 2012

SVs in targeted sequencing

- Amplicon sequencing
  - RPKM
  - Deviations detected by
    - Z scores

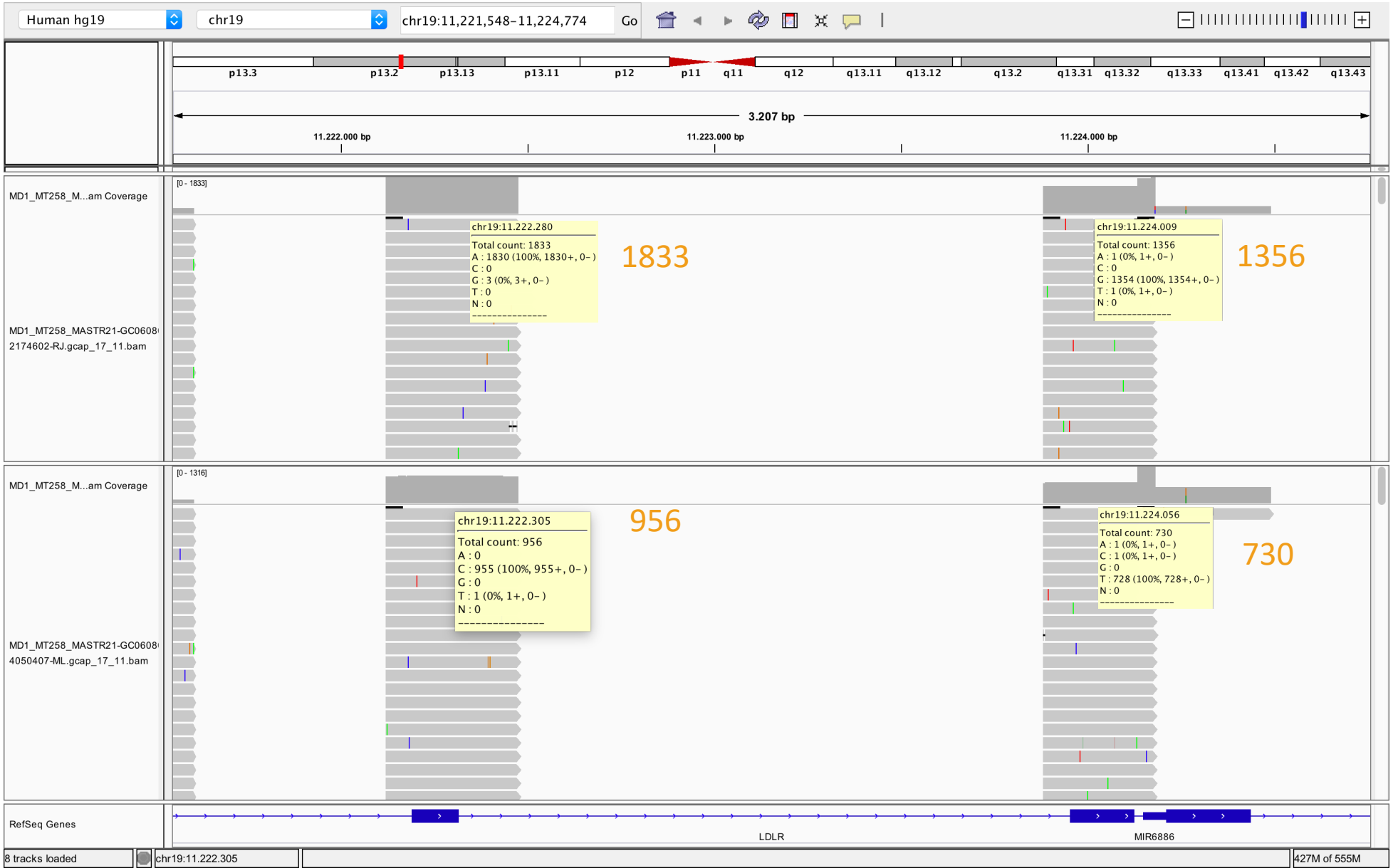
$$Z = \frac{RPKM - \text{Mean}}{\text{Standard deviation}}$$

    - Log2 ratios

$$log2 = \log\left(\frac{RPKM}{Mean}\right)/\log(2)$$

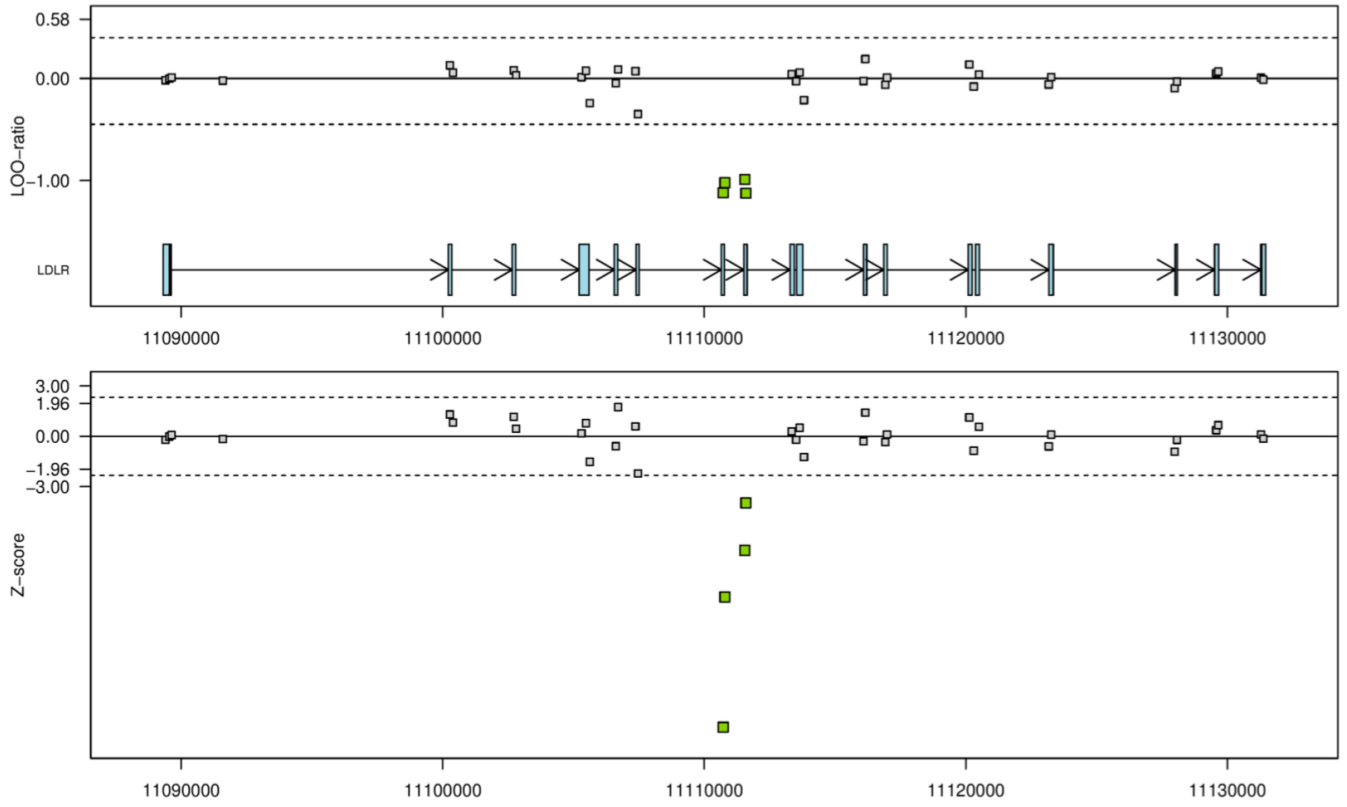
	Deletion	Duplication
Z score	< -3	> 3
Log2 ratio	< -0.7	> 0.5

SVs in amplicon sequencing



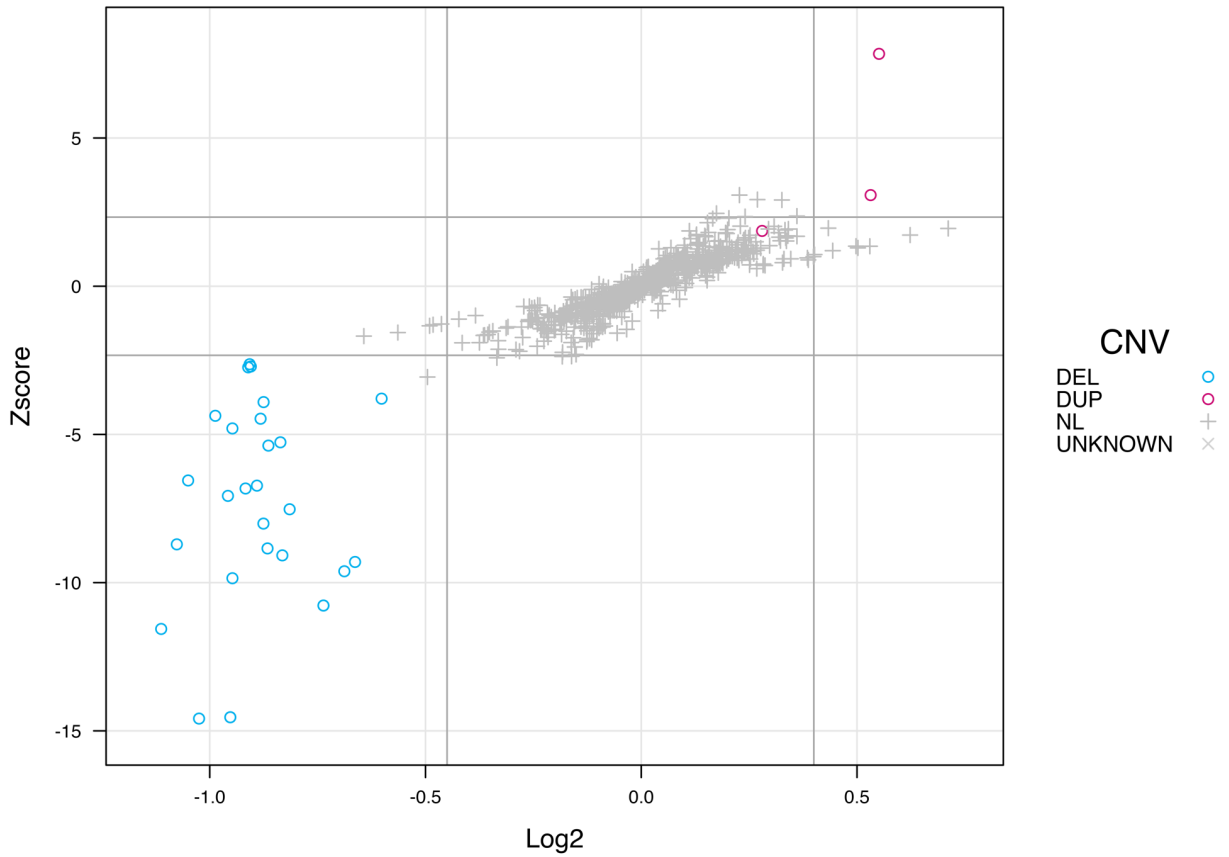
SVs in amplicon sequencing

Gene	Transcript	Exon	Amplicons/exon	Average log ratio	Average Z score
LDLR	NM 000527.4	7	2	-1.0713	-13.4763
		8	2	-1.0569	-5.3929



SV filtering

- Long lists of variants
  - Filtering required to discard False Negatives
  - Read ratio
  - CNV scores
  - CNV length
  - Etc.



Summary

- SVs can be called from short read data
- Some SVs are more difficult/impossible to detect from targeted data
- Other NGS strategies
  - *de novo* assembly
  - Long read sequencing

Questions?

Erika Souche