Structural Variations

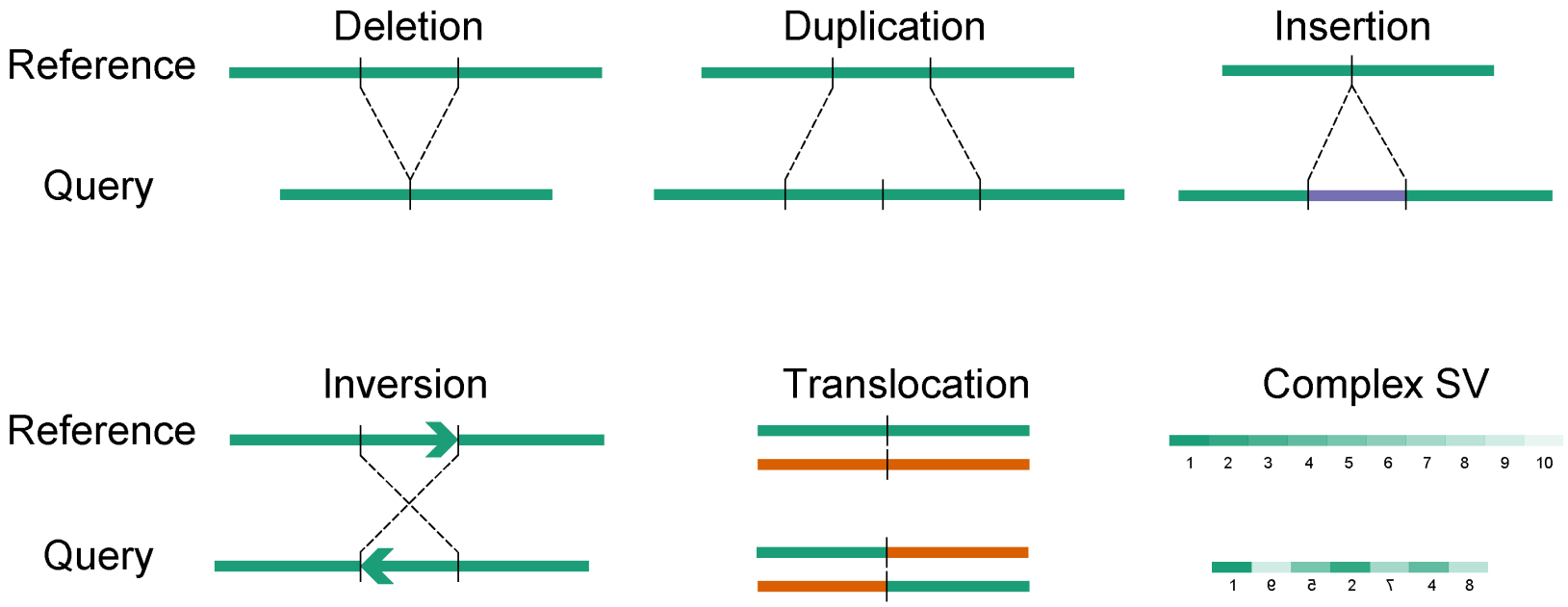
Lixing Yang, Ph.D.

Ben May Department for Cancer Research

Outline

- What are SVs?
- Complex SVs
- SV calling
- SV landscape
- SV signatures
- Oncogenic SVs

Structural Variations (SVs)/Genomic rearrangements



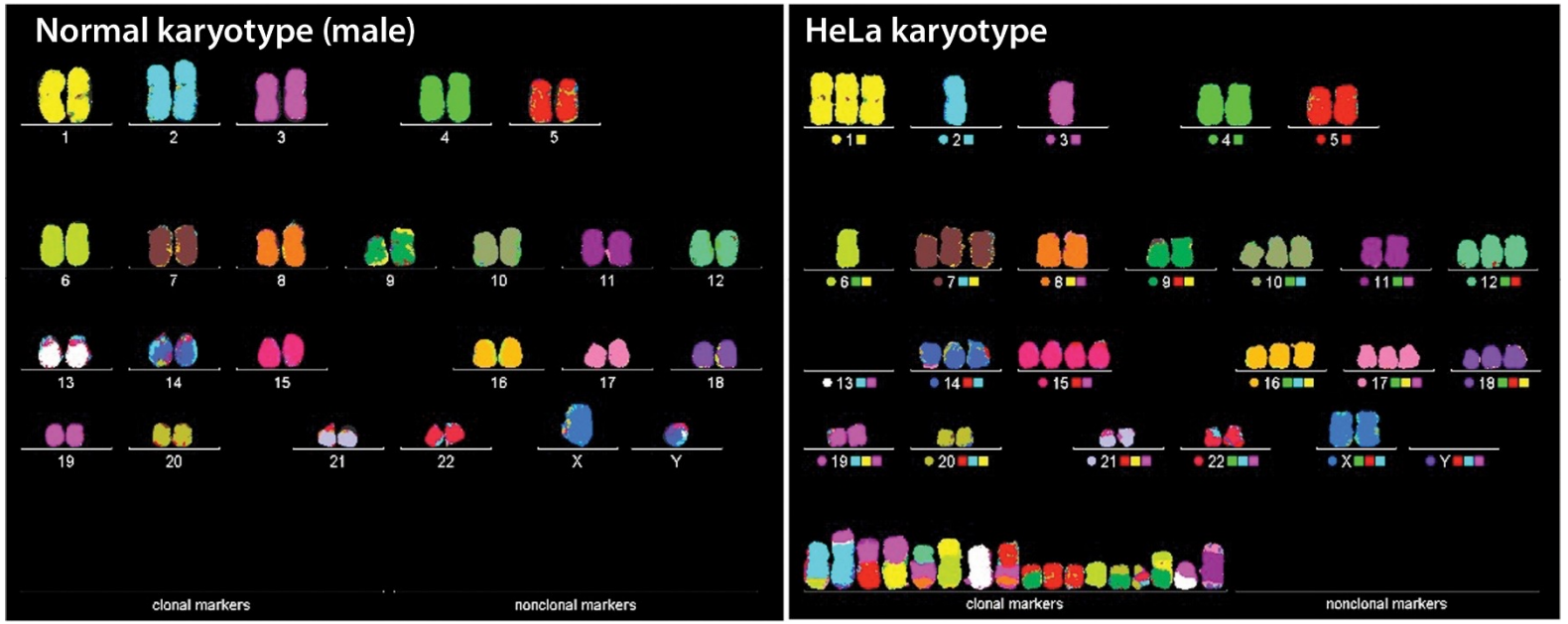
DNA Double Strand Break Repair

- HR: Homologous recombination
- NHEJ: Non-homologous end joining
- MMEJ/alt-EJ: Microhomology mediated end joining
- NAHR: Non-allelic homologous recombination
- FoSTeS/MMBIR: Fork stalling and template switching/Microhomology mediated break induced repair

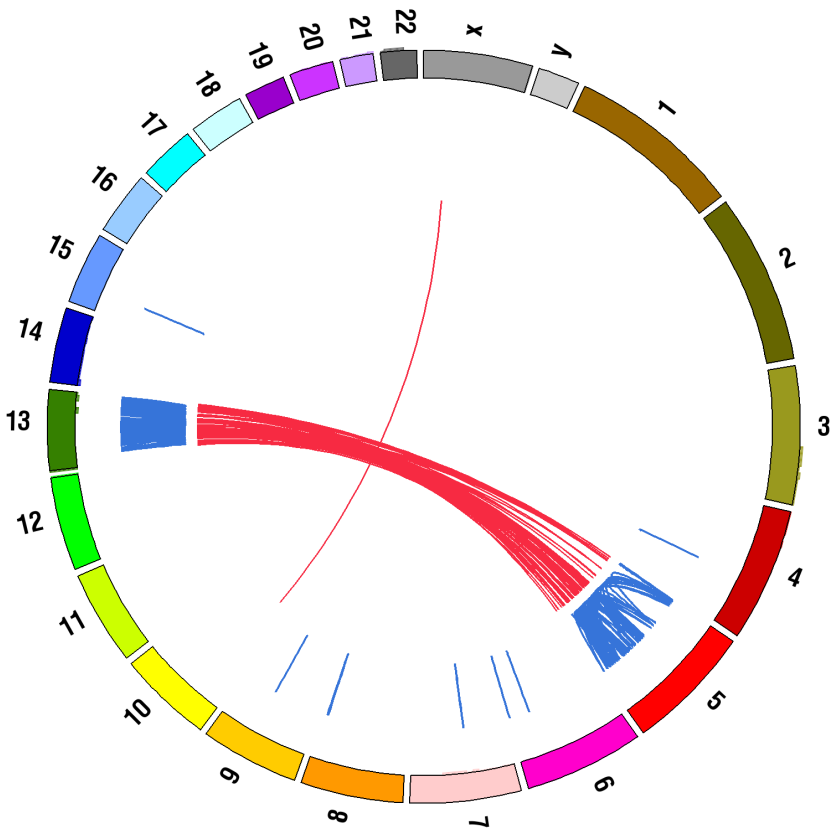
BCR-ABL Fusion

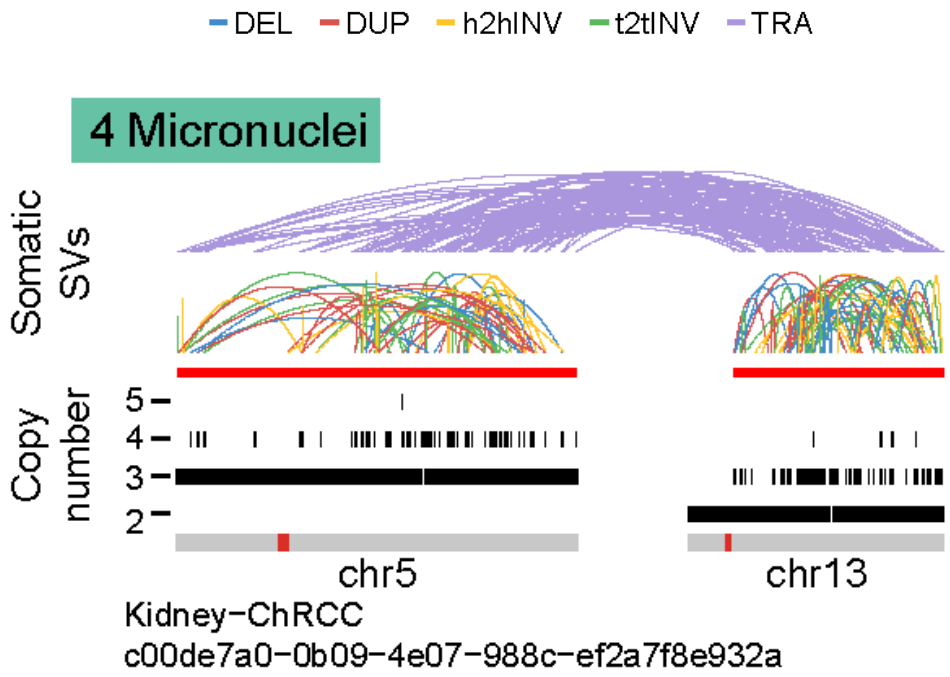
- Philadelphia chromosome (reciprocal translocation between chromosome 9 and 22) in CML
- *BCR-ABL1* fusion
- Imatinib, etc.

Complex karyotype

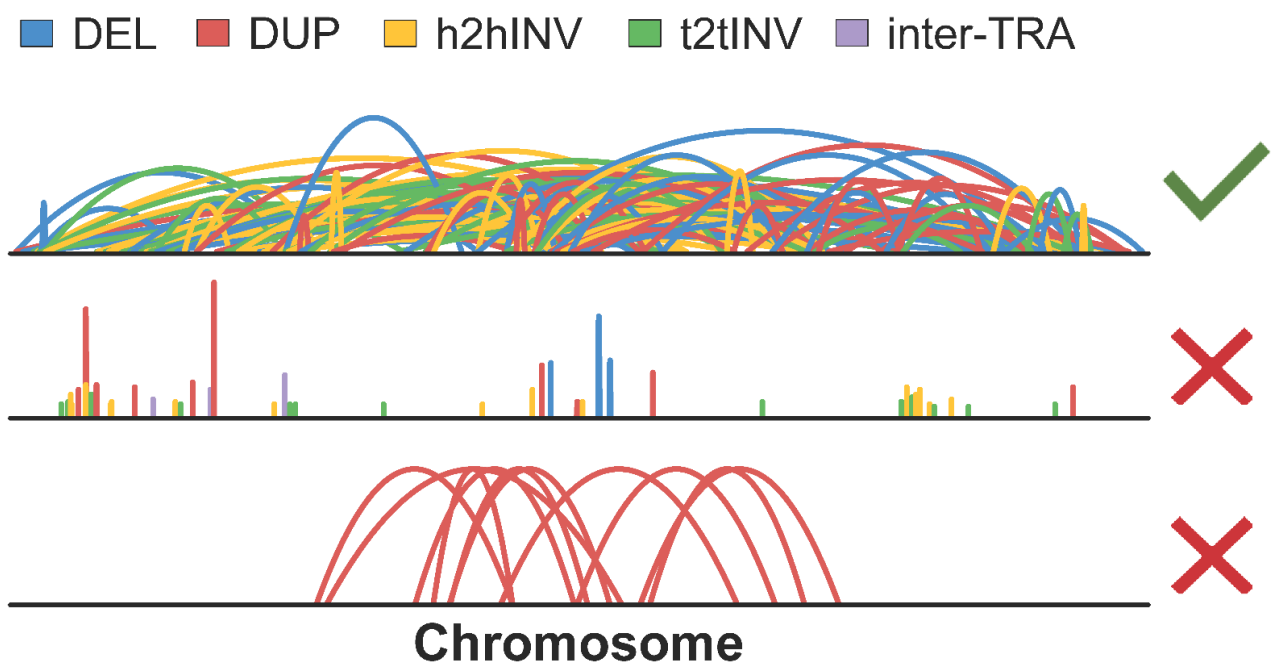


Chromothripsis





ShatterSeek



Complex SV mechanisms

Micronuclei

SVs

Copy number

Chromatin bridge

Extrachromosomal DNA (ecDNA)

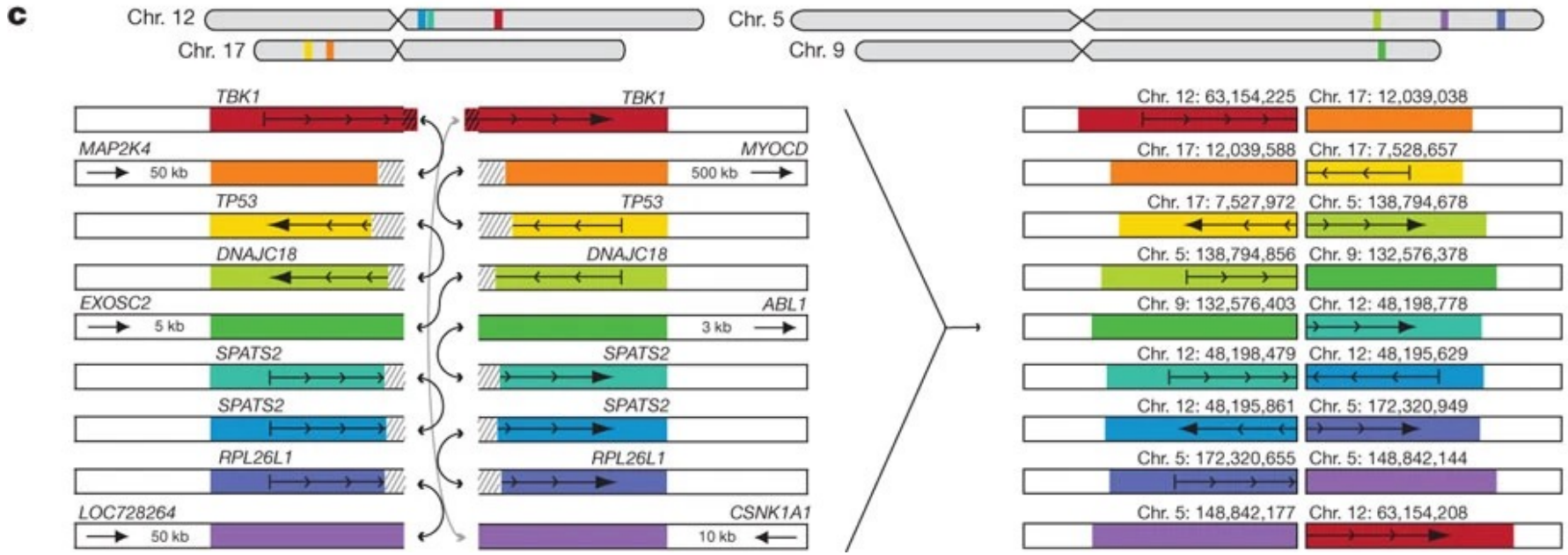
SVs

Copy number

BFB cycles

foldback inversion

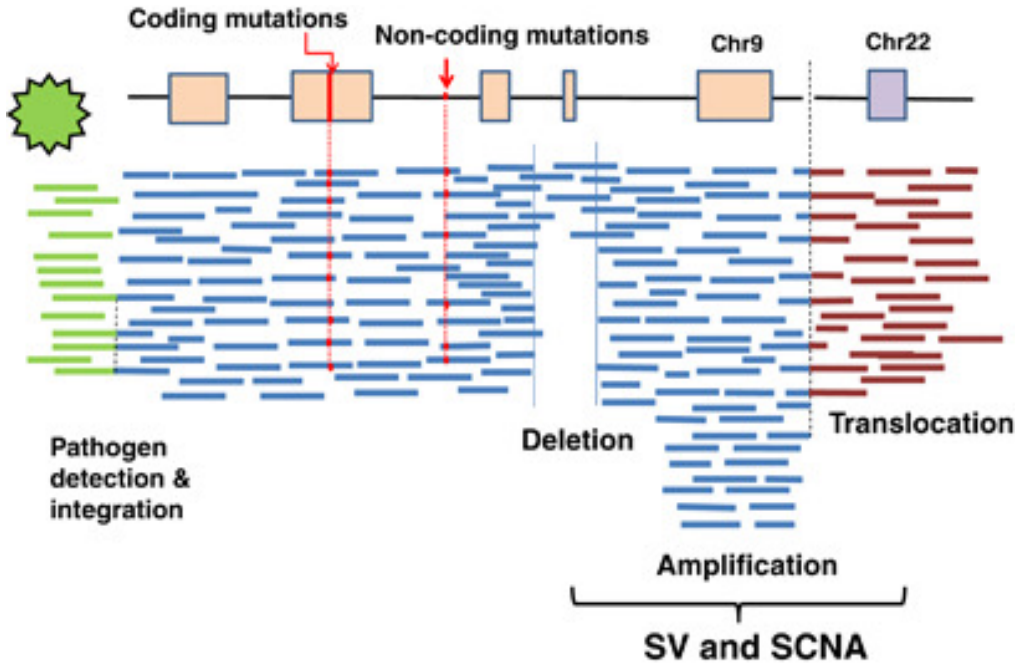
Chromoplexy

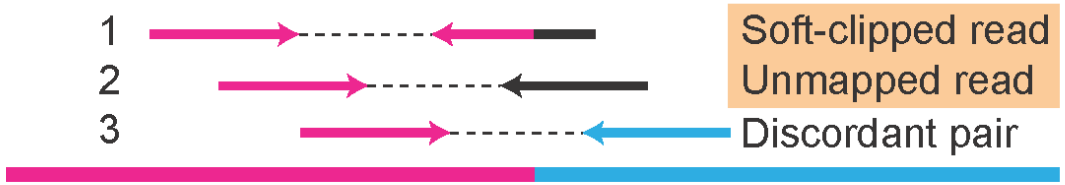


ChainFinder
Junction pattern

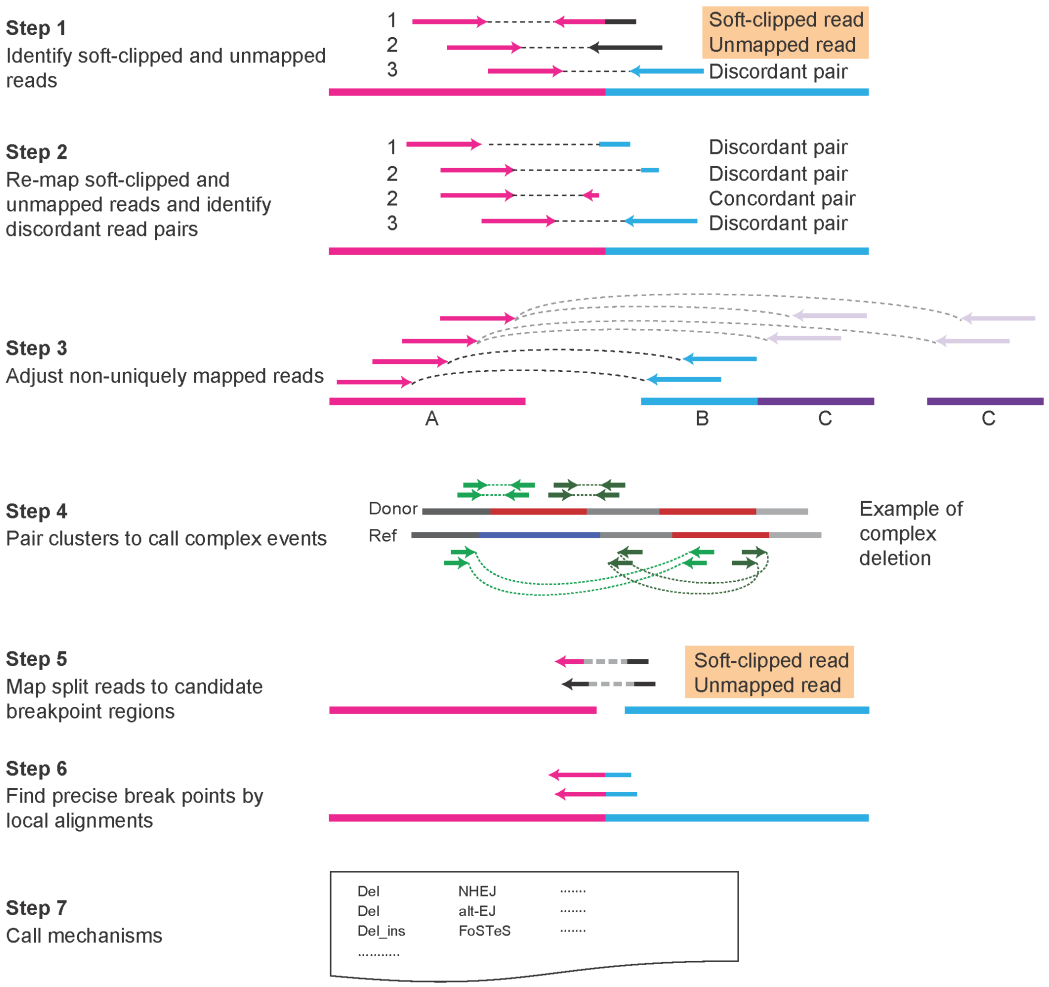
SV calling strategies

- Read depth
- Read pair
- Split read
- Genome assembly

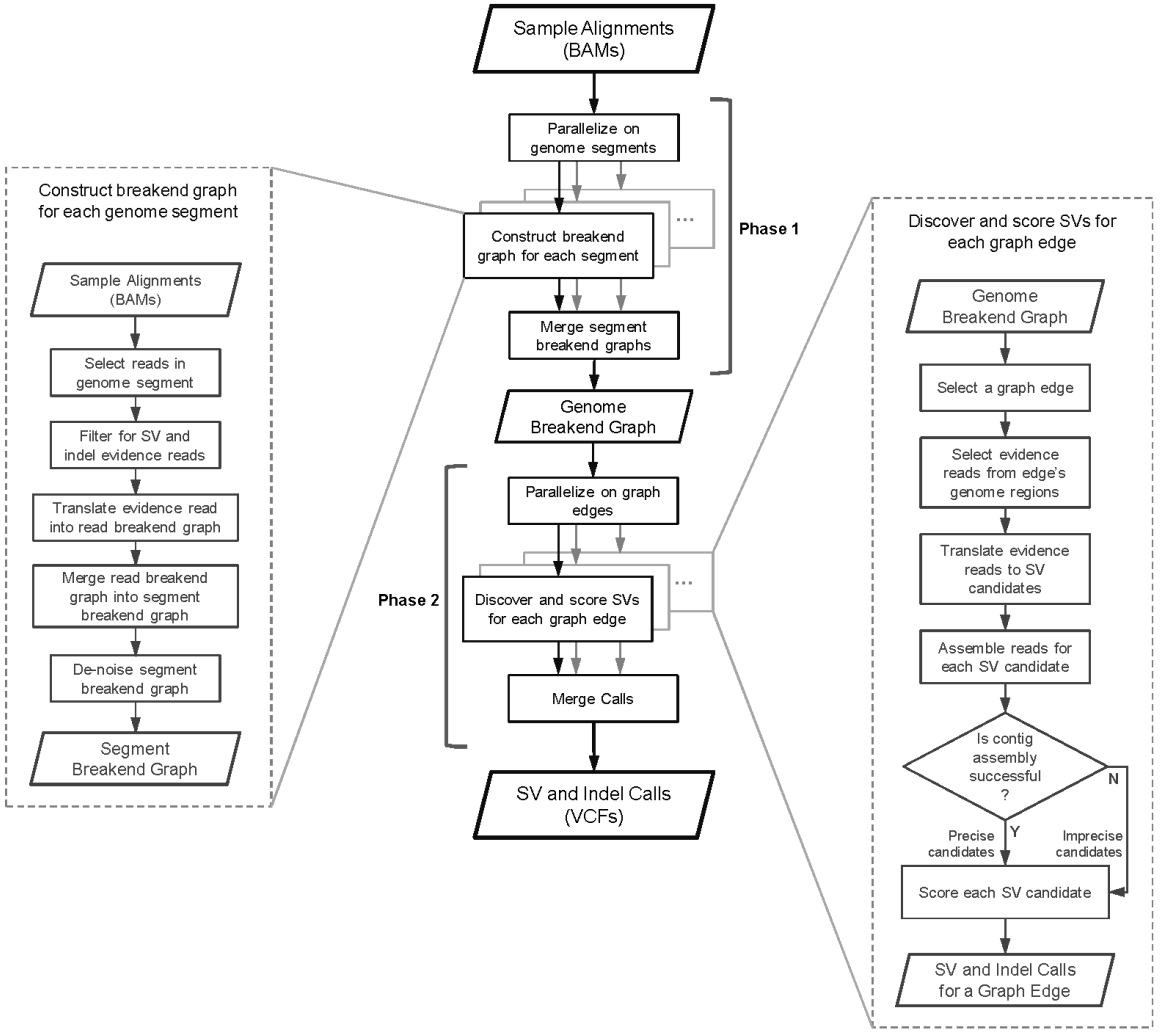




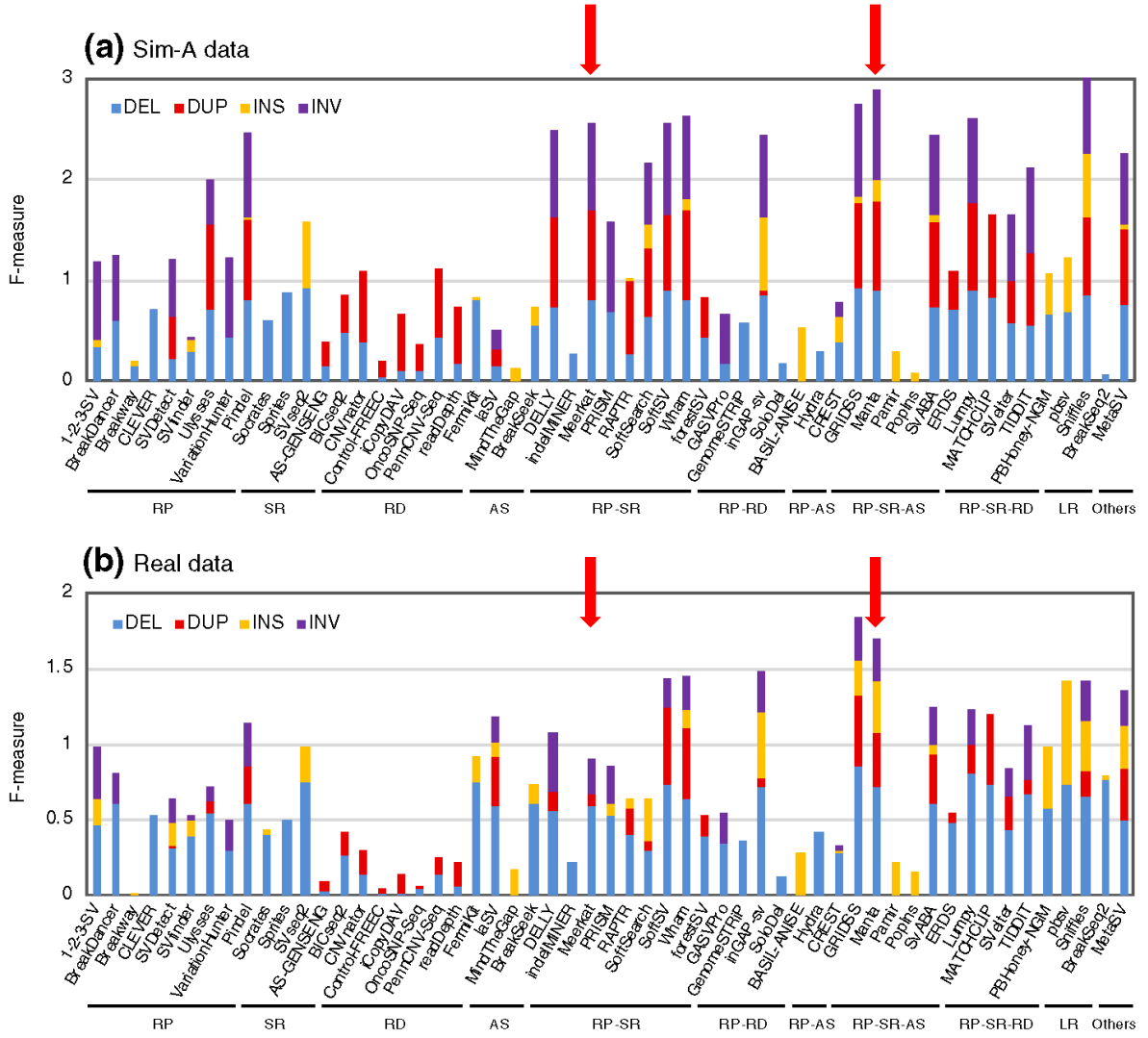
Meerkat (validation rate 80-90%)



Manta


Sample Alignments (BAMs)
Parallelize on genome segments
Construct breakend graph for each segment
Merge segment breakend graphs
Genome Breakend Graph
Parallelize on graph edges
Discover and score SVs for each graph edge
Merge Calls
SV and Indel Calls (VCFs)
Phase 1
Phase 2
Construct breakend graph for each genome segment
Sample Alignments (BAMs)
Select reads in genome segment
Filter for SV and indel evidence reads
Translate evidence read into read breakend graph
Merge read breakend graph into segment breakend graph
De-noise segment breakend graph
Segment Breakend Graph
Discover and score SVs for each graph edge
Genome Breakend Graph
Select a graph edge
Select evidence reads from edge's genome regions
Translate evidence reads to SV candidates
Assemble reads for each SV candidate
Is contig assembly successful ?
N
Y
Precise candidates
Imprecise candidates
Score each SV candidate
SV and Indel Calls for a Graph Edge

Benchmarking



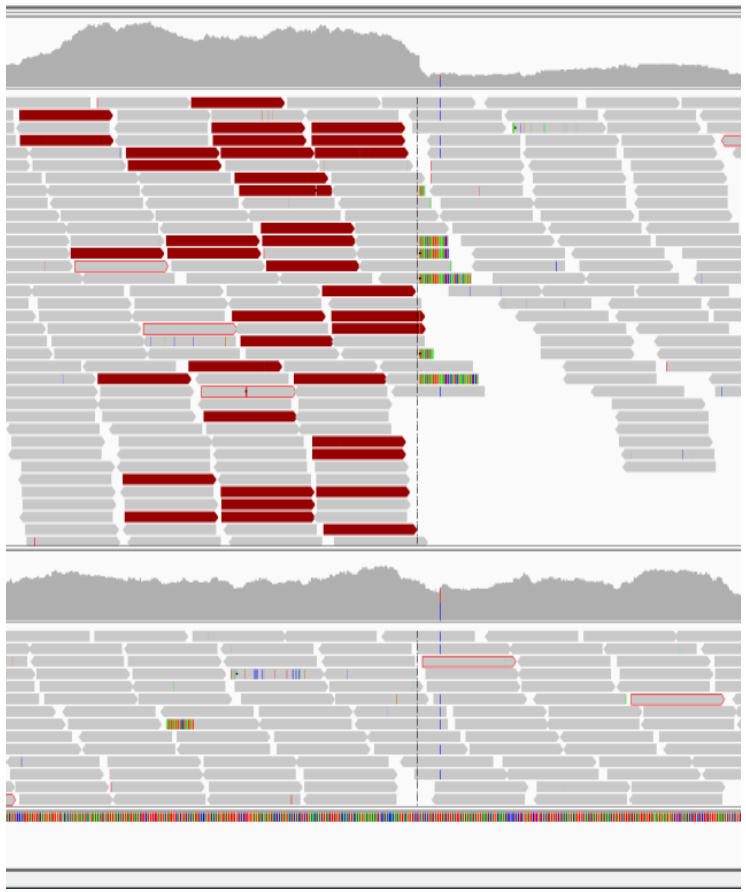
Good SV callers

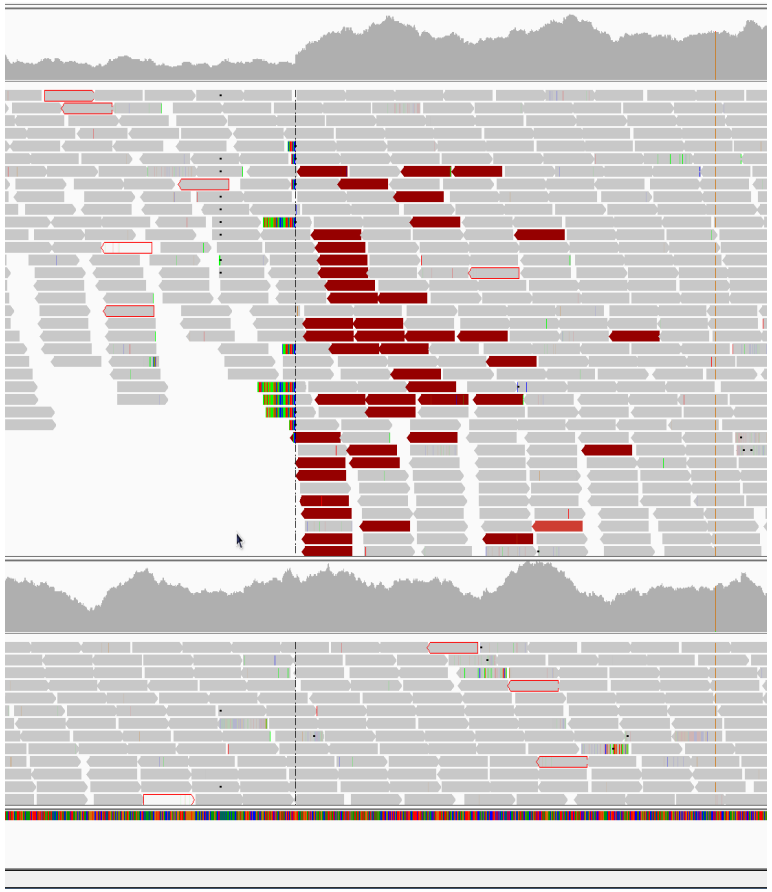
- Short reads:
  - Meerkat, Manta, GRIDSS, Delly, Breakseq, Lumpy, …
- Long reads:
  - Sniffles

True SV (300kb del, PTEN loss KICH)

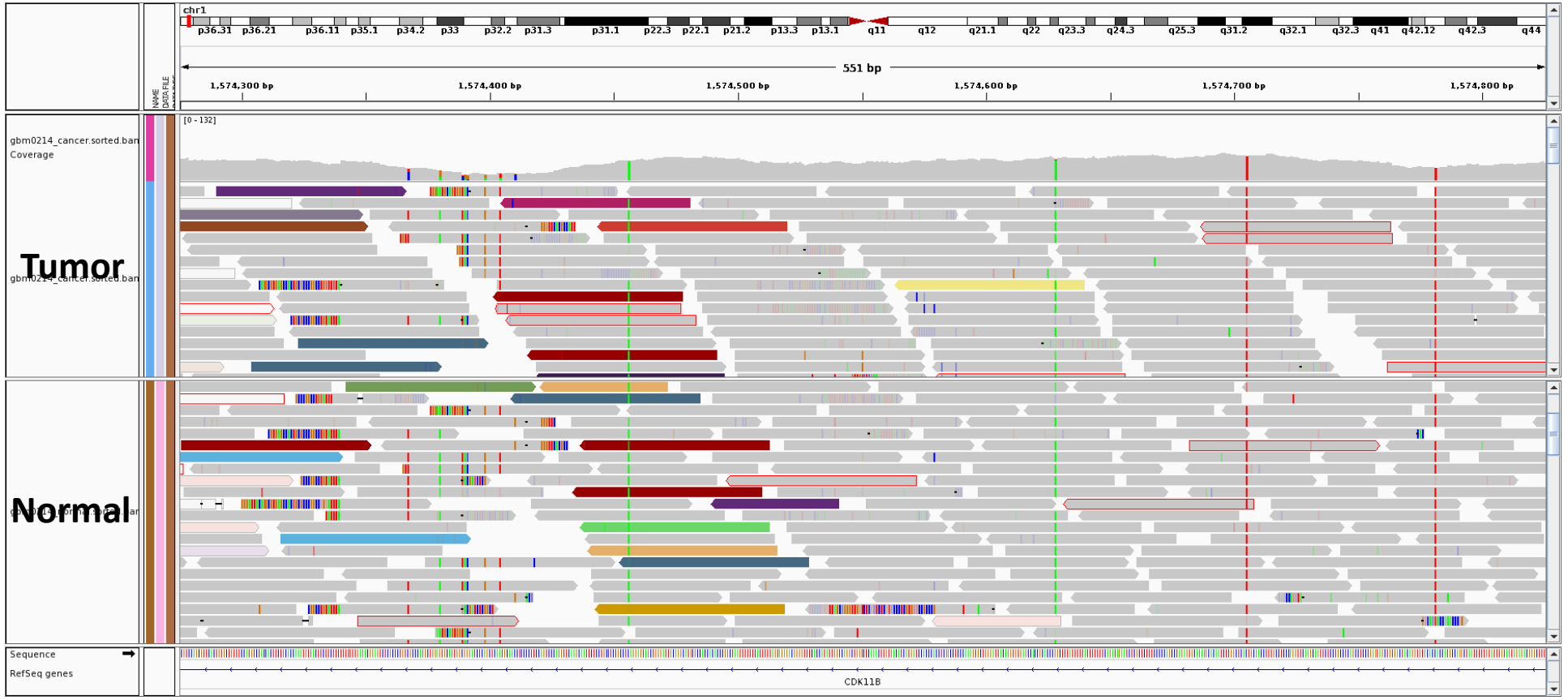
Tumor

Normal

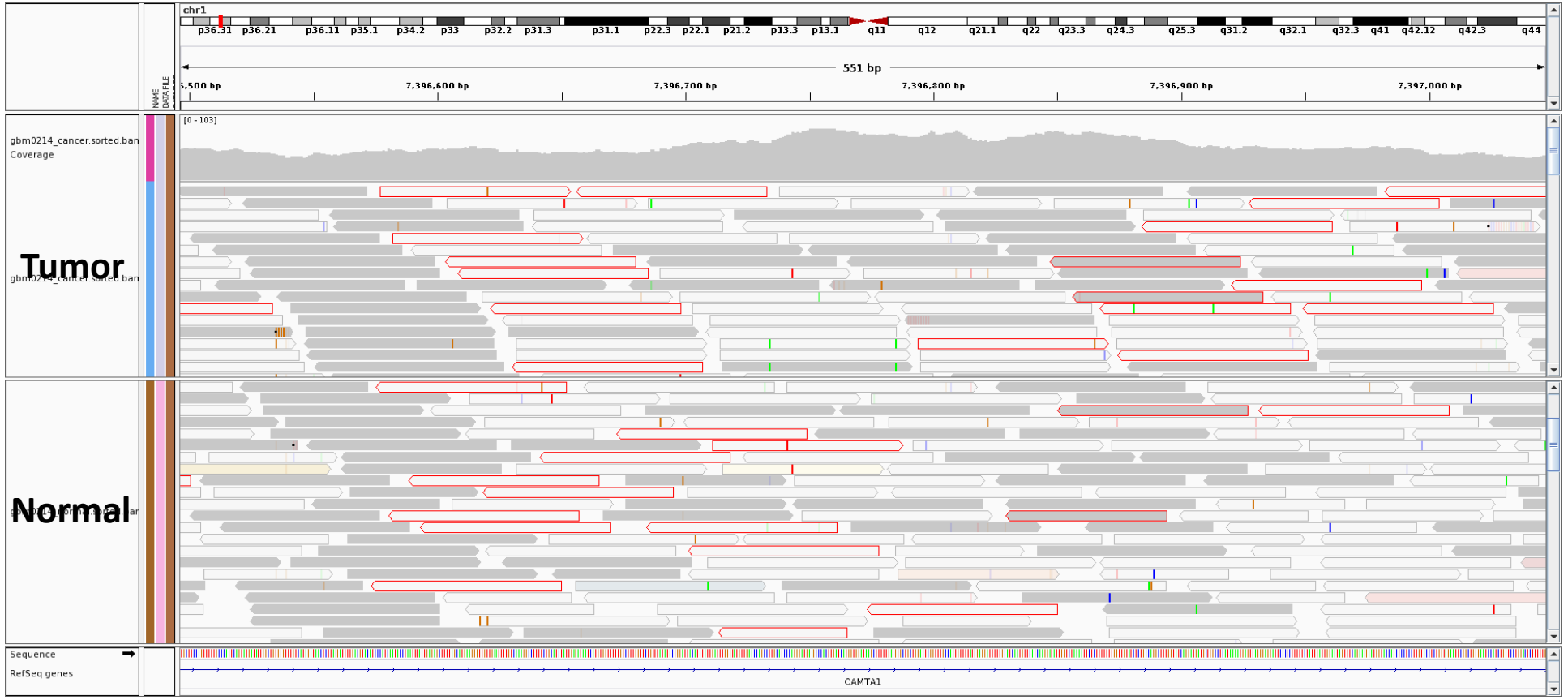




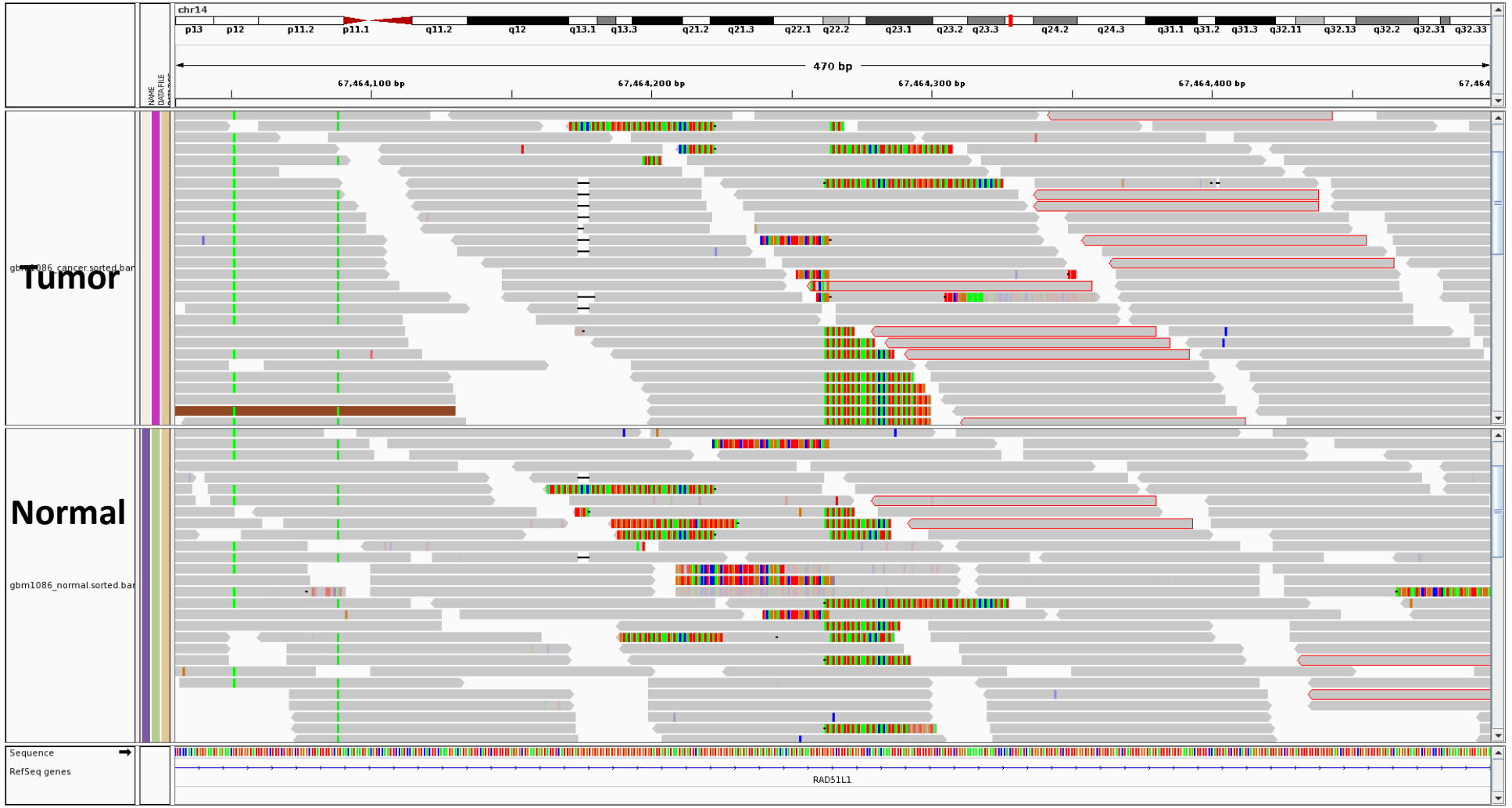
Too many discordant pairs in matched normal



Too many non-uniq mapped reads in matched normal



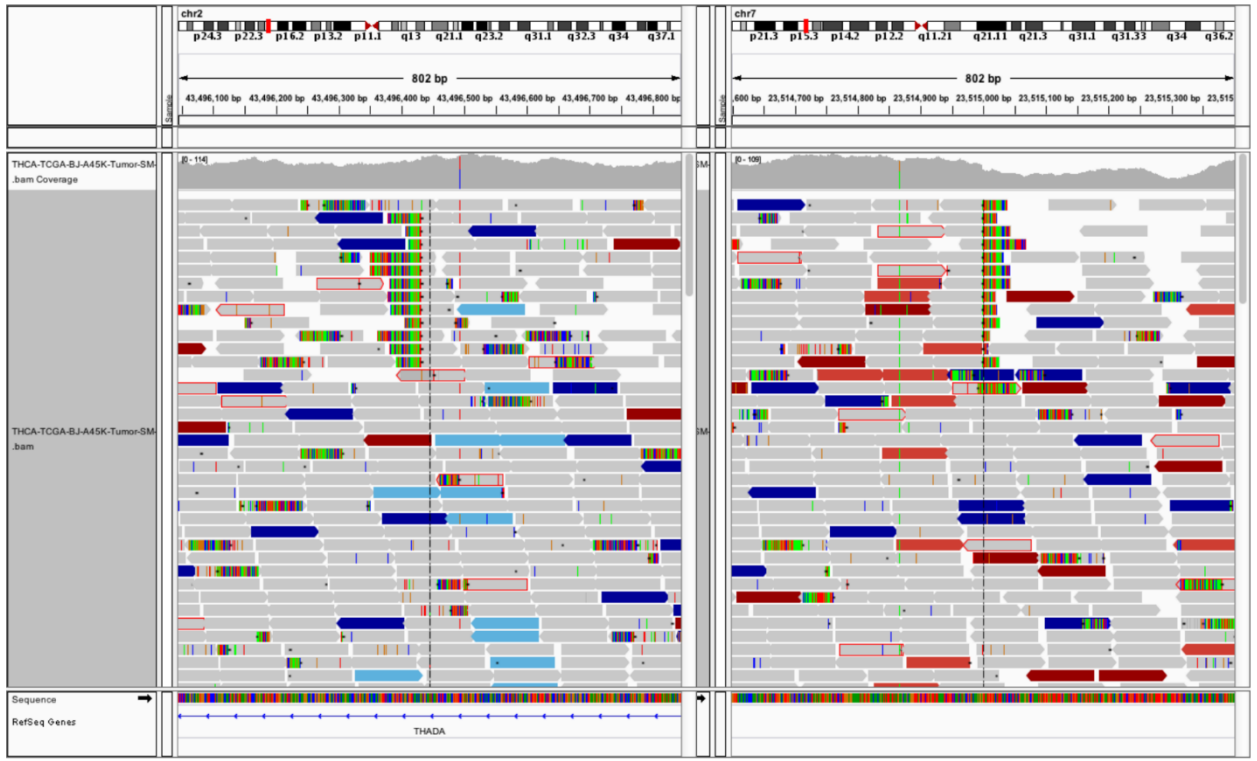
Too many soft-clipped reads in matched normal



True SV (chr2:chr7 THCA)

TCGA-BJ-A45K:
Intron of THADA(-): 10Kb after exon 36
IGF2BP3 IGR: 5Kb after IGF2BP3(-)

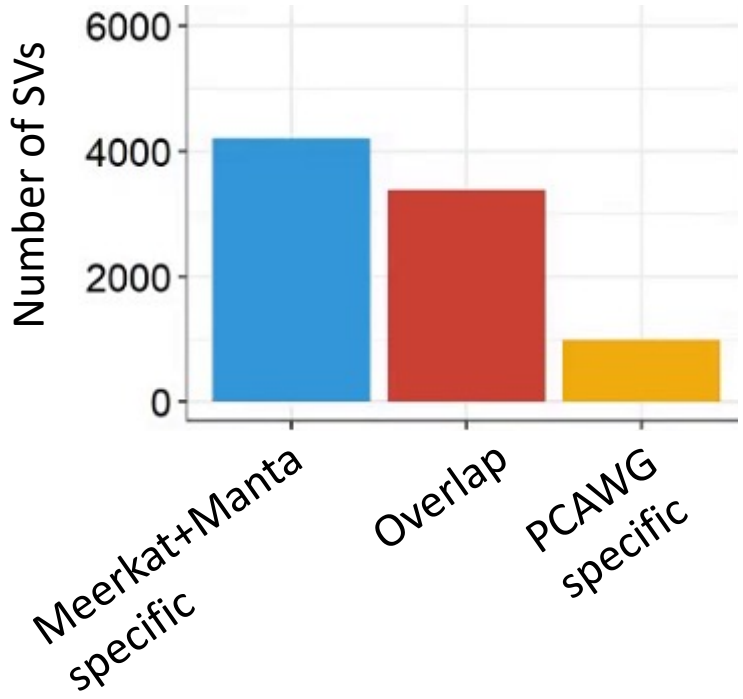
Tumor



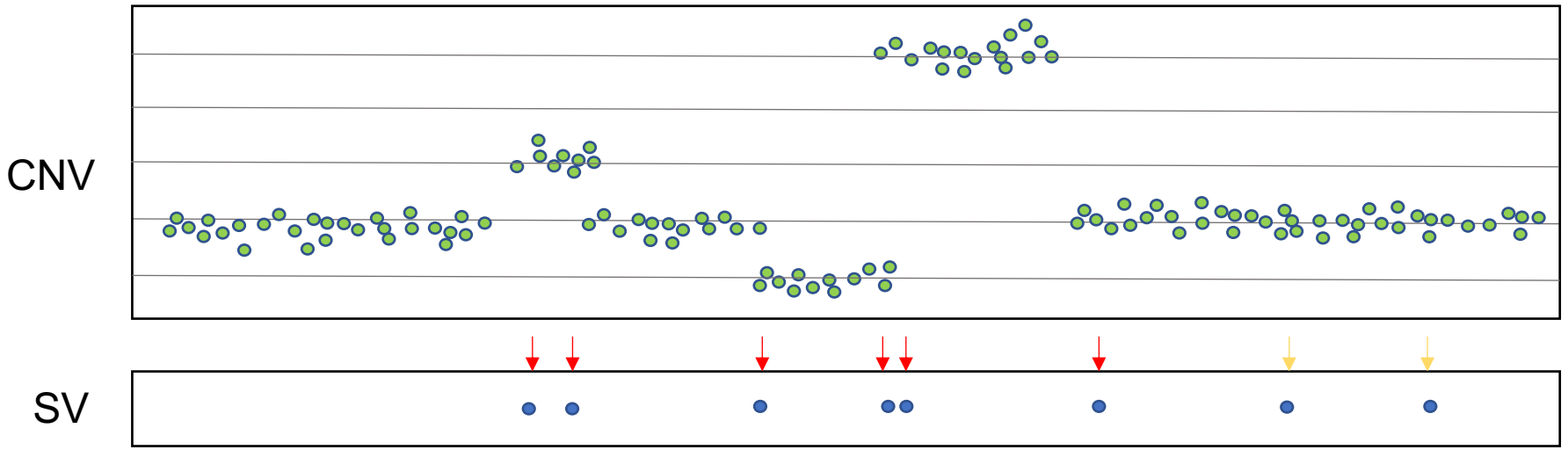
Sniffles (long reads)

Sherlock solution: union of Meerkat and Manta

- PCAWG: at least 2 out of 4 (SvABA, DELLY, BRASS, dRanger)
- Validation rate: 90%

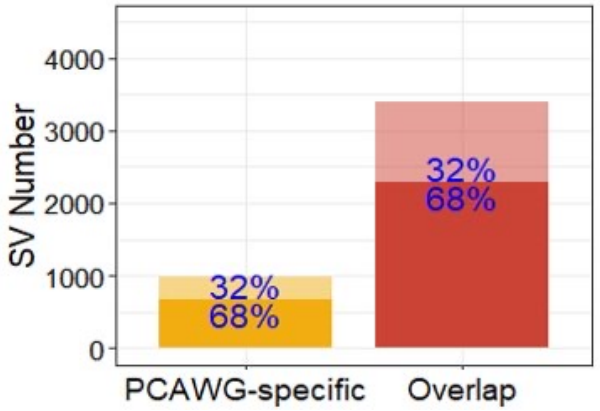


A significant portion of SV breakpoints should be supported by CNV breakpoints



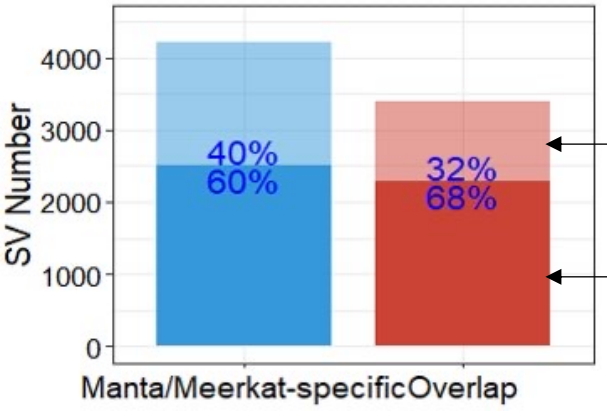
35 LUAD in PCAWG (hg19)

PCAWG



At least 2 out of 4:
SvABA, DELLY, BRASS, dRanger

Meerkat
Manta
union



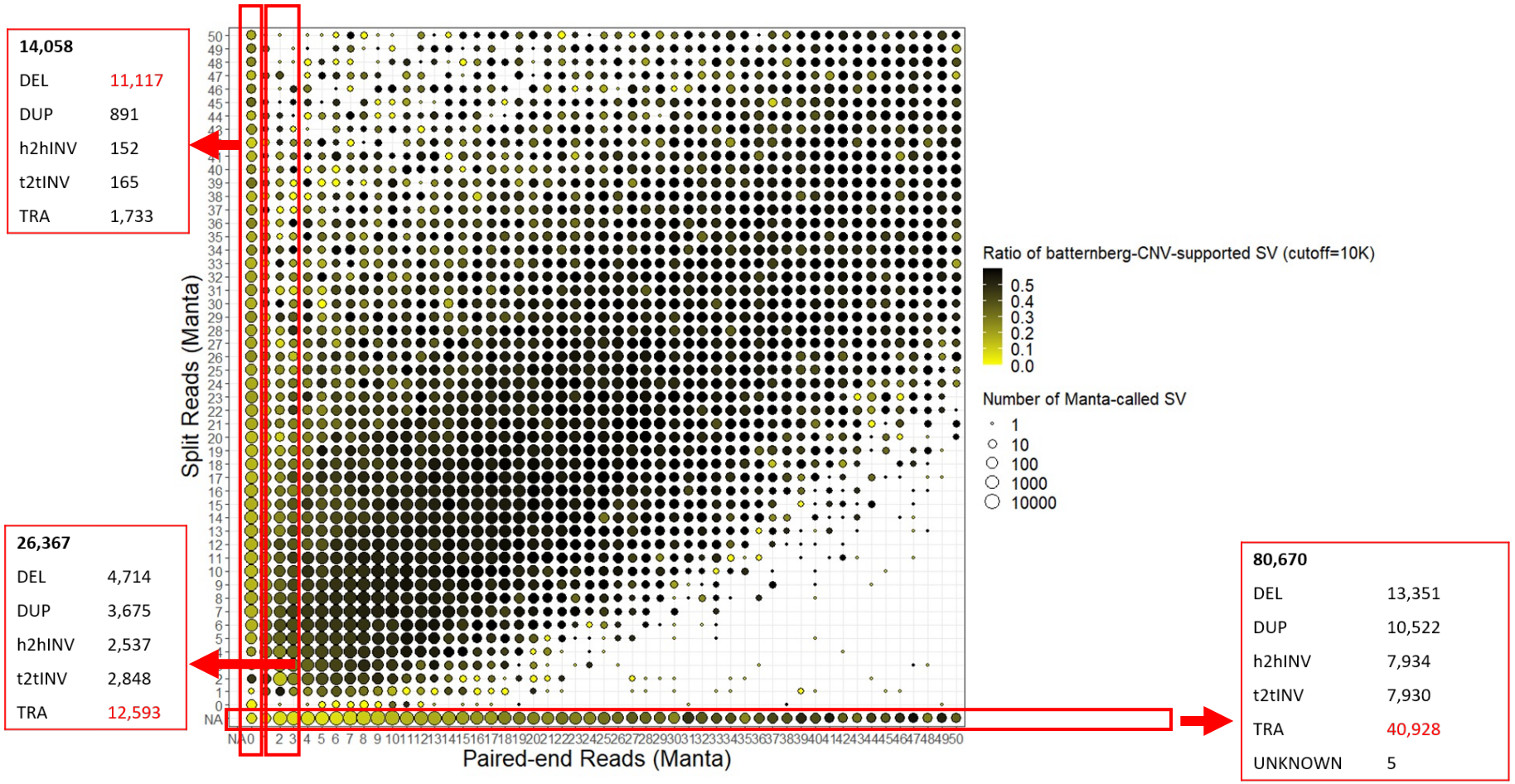
Without CNV support

With CNV support

Manta problematic with hg38

	hg19 Manta			hg38 Manta			hg19 Meerkat			hg38 Meerkat		
TCGA-05-4398	217	DEL DUP h2hINV t2tINV TRA	100 19 22 22 54	668	DEL DUP h2hINV t2tINV TRA	117 44 65 60 382	252	DEL DUP h2hINV t2tINV TRA	88 33 30 25 76	182	DEL DUP h2hINV t2tINV TRA	29 25 36 17 75
TCGA-50-5930	120	DEL DUP h2hINV t2tINV TRA	67 22 12 13 6	260	DEL DUP h2hINV t2tINV TRA	97 39 27 19 78	175	DEL DUP h2hINV t2tINV TRA	89 38 18 15 15	127	DEL DUP h2hINV t2tINV TRA	65 28 18 11 5
TCGA-50-6591	170	DEL DUP h2hINV t2tINV TRA	57 29 27 31 26	329	DEL DUP h2hINV t2tINV TRA	76 45 45 45 118	211	DEL DUP h2hINV t2tINV TRA	73 39 33 33 33	158	DEL DUP h2hINV t2tINV TRA	50 33 31 29 15

DEL and TRA in Manta



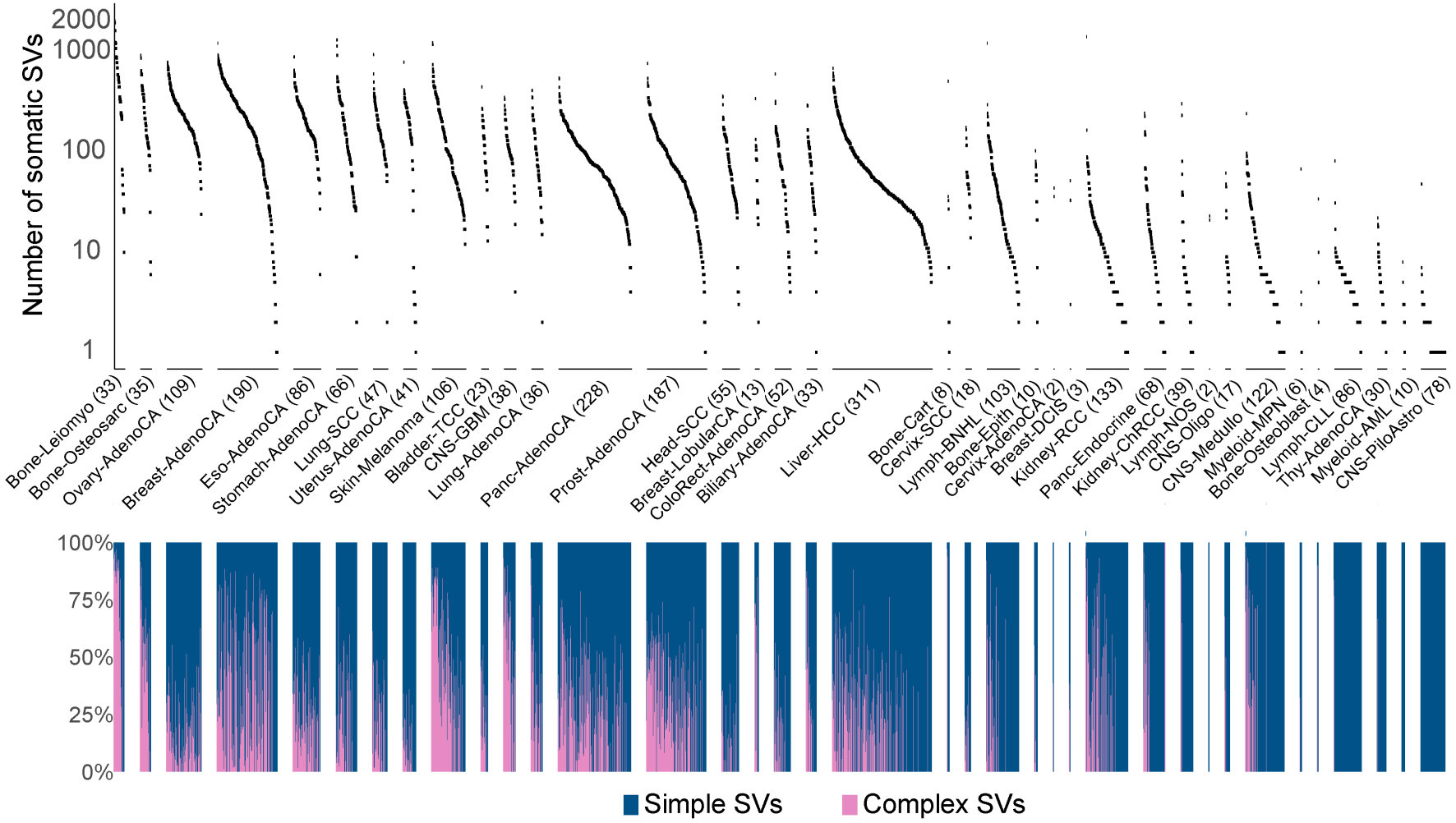
Sherlock solution

- Meerkat + Manta union
- Manta SVs: > 3 read pair support && >=1 split read support
- PCAWG missed half of the true SVs

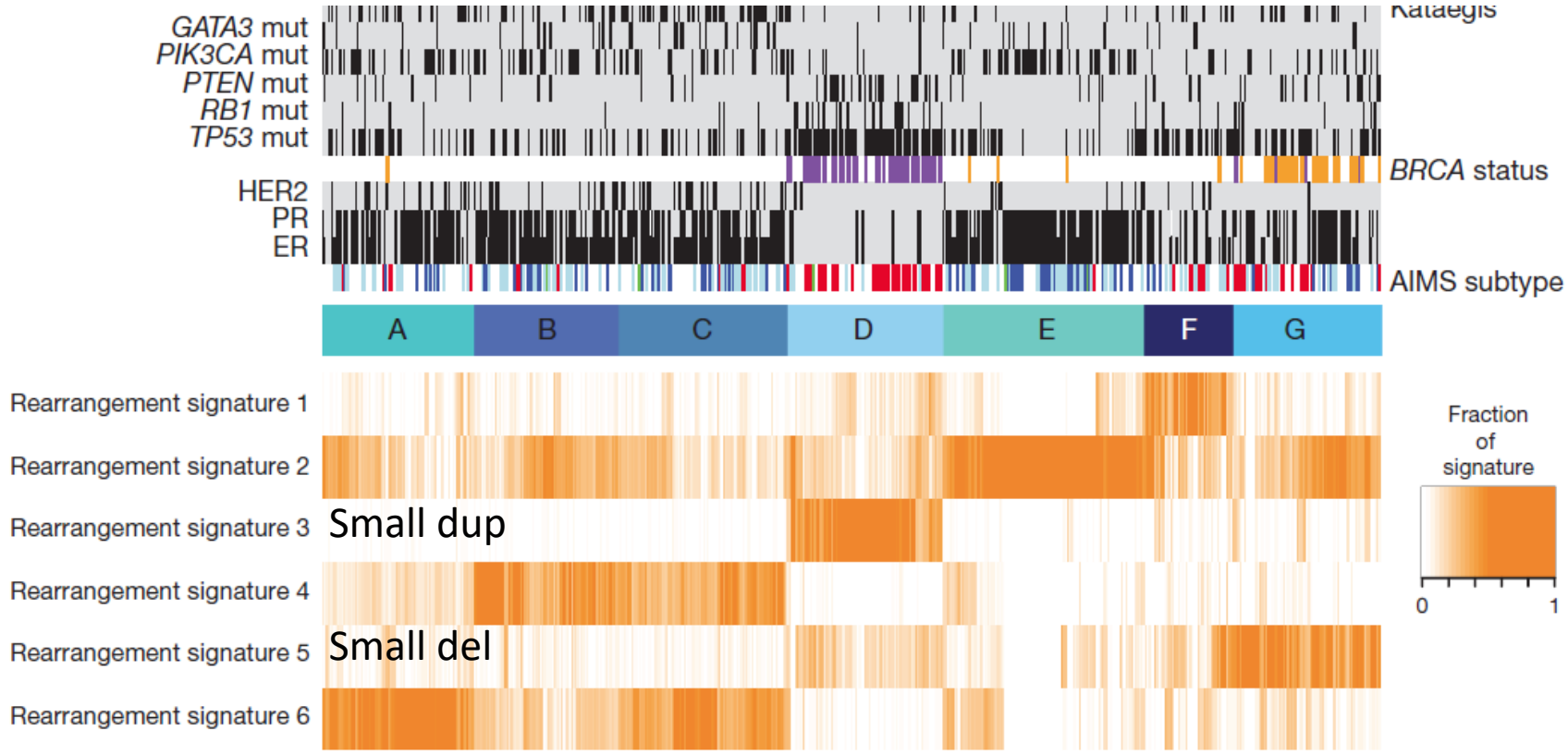
Types of variants

- Germline variants
  - ~20,000 per genome by long-read sequencing
  - 3,000-5,000 per genome by short-read sequencing
- De novo variants (parent-child trio)
  - Dozens per generation
  - 1/1000 newborns with translocations/large inversions
- Somatic variants (tumor-normal pair)

Landscape of somatic SVs



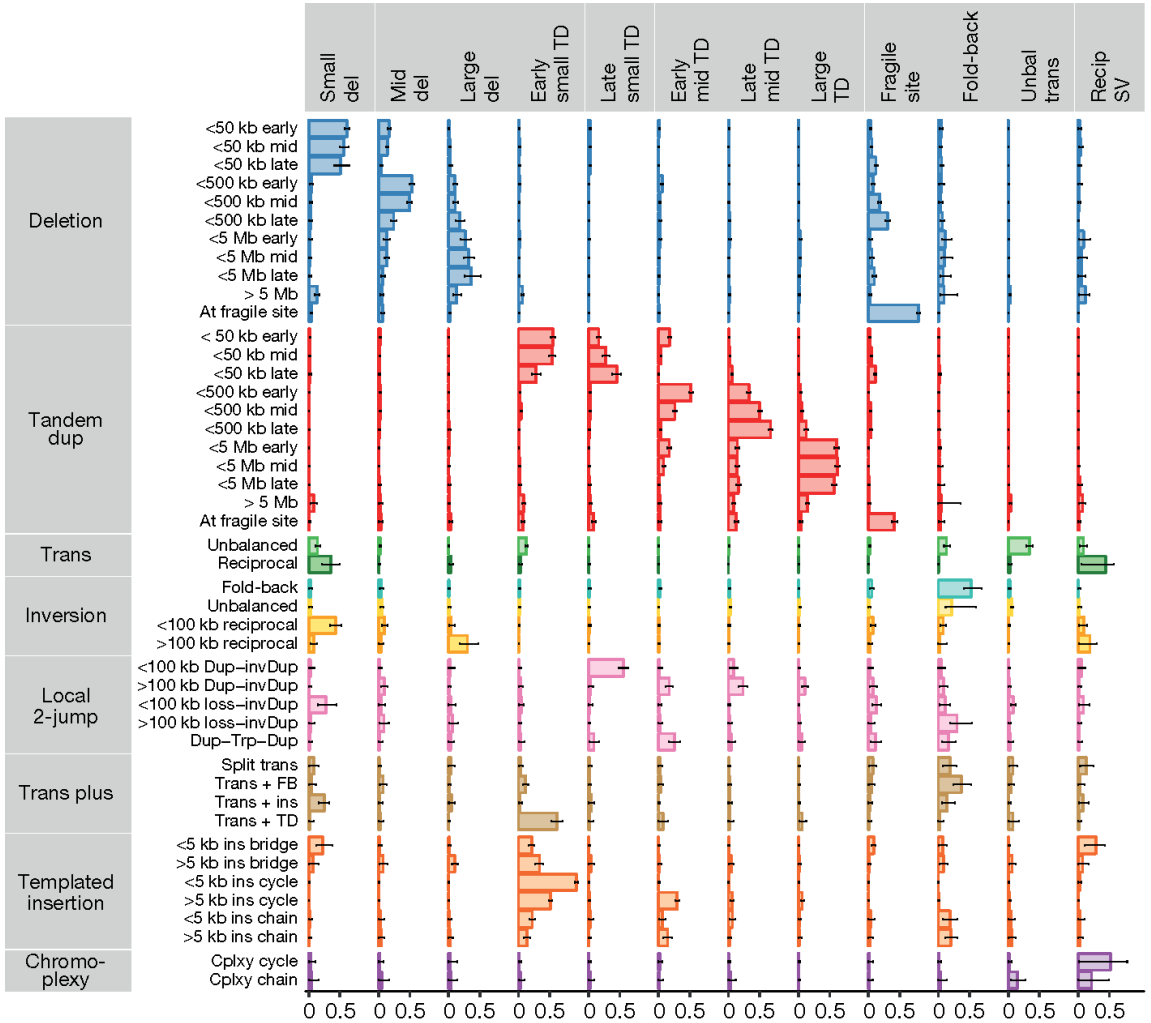
SV signatures in 560 breast cancers



BRCA Mutations

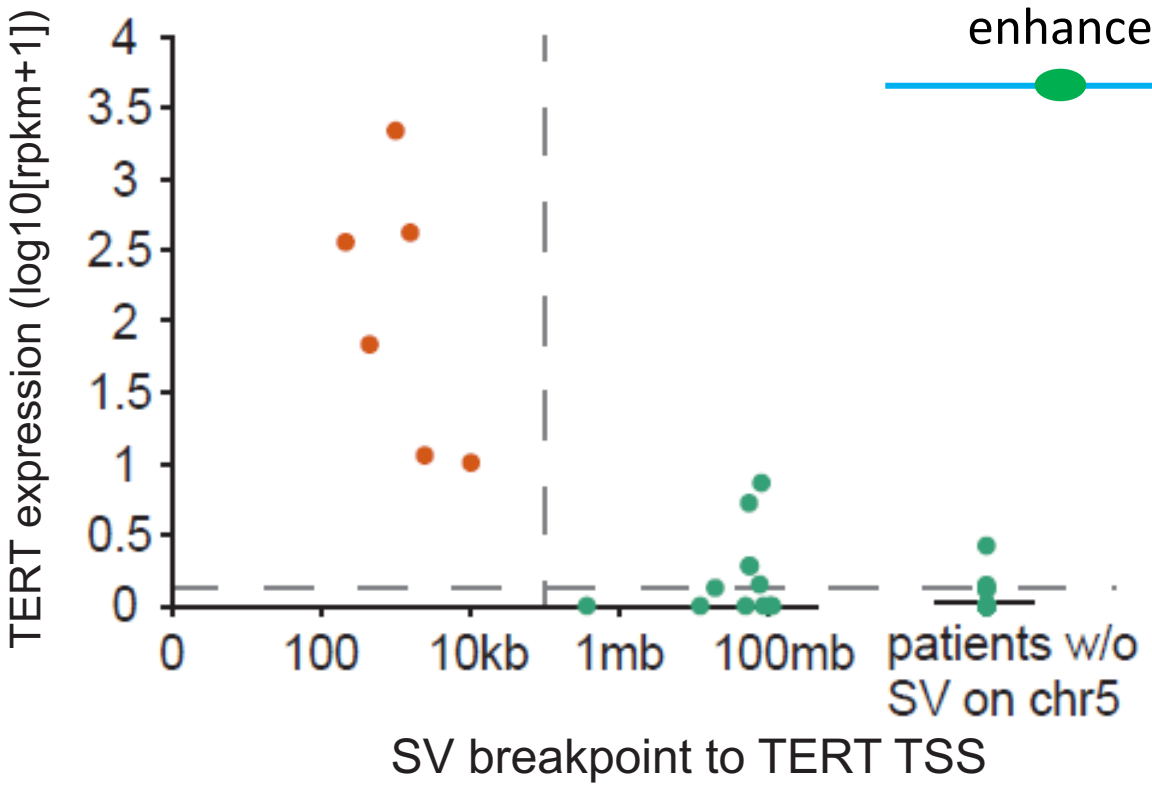
- *BRCA1* and *BRCA2*, DNA double strand break repair, 5-10% in breast cancer
- *PARP1*, DNA single strand break repair

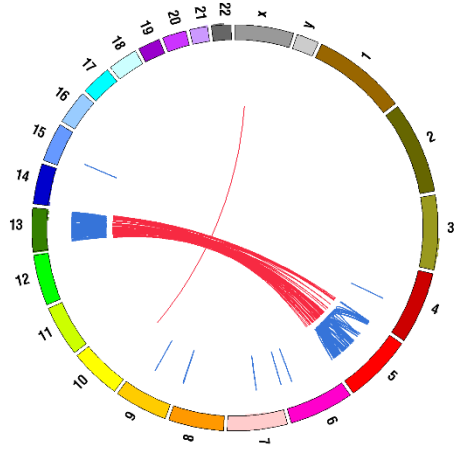
PCAWG SV signatures



TERT enhancer hijacking

- 6 out of 50 (12%) of chromophobe renal cell carcinomas





Davis*, Ricketts*, Wang*, Yang* *et al.* 2014 *Cancer Cell*