X-SVLen

Input and results

Input:

ONT based SV callset using cuteSV

hg37

TrioCanu assemblies.

Method:

For each SV defined by chrom, pos, variant:
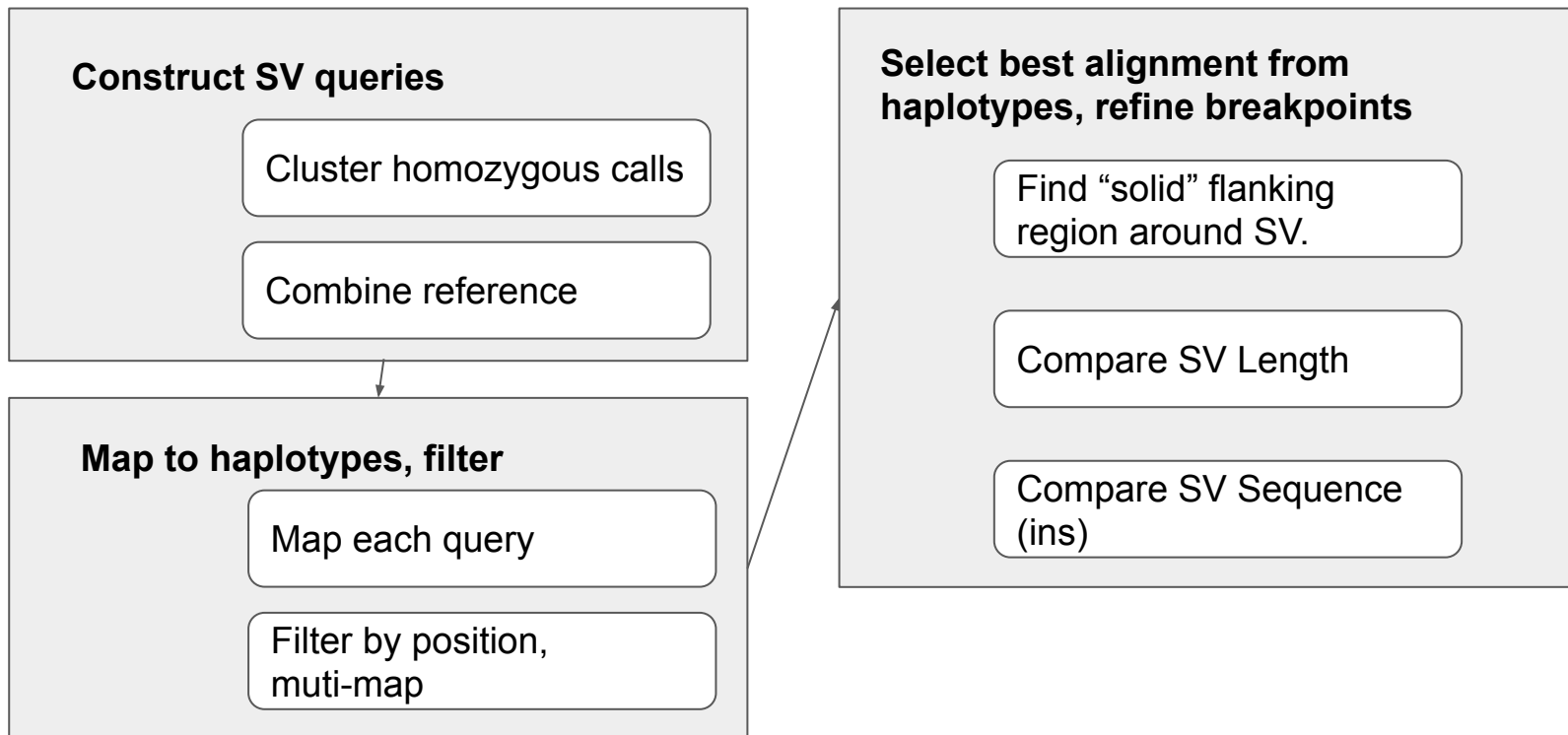
    Construct a query as: ref[chrom][pos-prefix:pos] + SV + ref[chrom][pos:pos+suffix]

Query:	prefix	SV	suffix

Map all queries to hap1 and hap2 assemblies.

Workflow

Construct SV queries

Cluster homozygous calls

Combine reference

Map to haplotypes, filter

Map each query

Filter by position, muti-map

Select best alignment from haplotypes, refine breakpoints

Find "solid" flanking region around SV.

Compare SV Length

Compare SV Sequence (ins)

Assembly calls that overlap contigs.

SVType	INS		DEL	
SVLen	<50	>=50	<50	>=50
number of these calls that overlap assembly contigs	16054	17481	30494	6475

Example difficult to validate call.
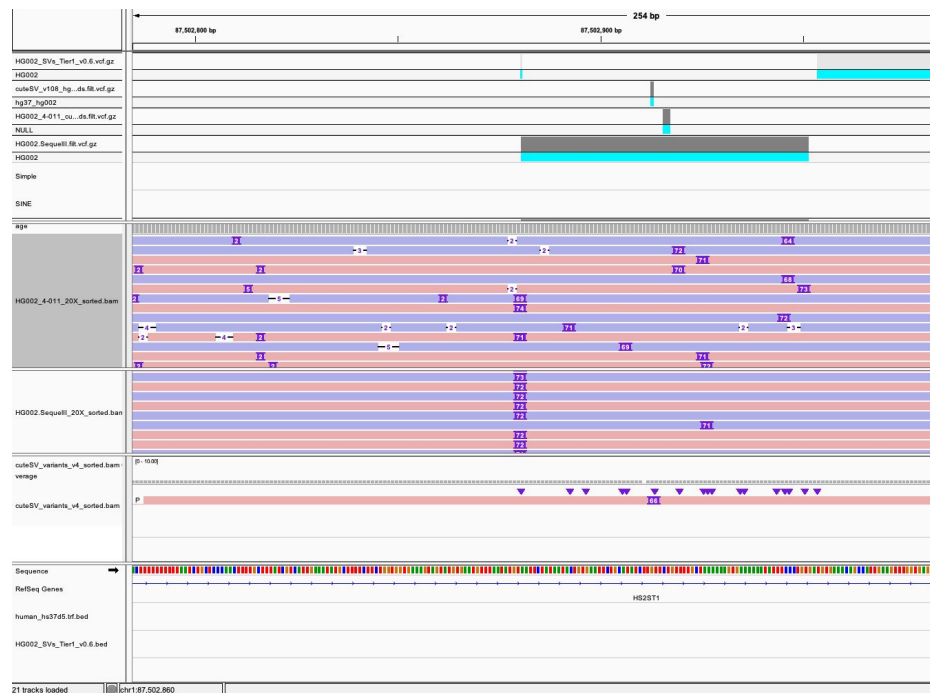
Annotation of true positives
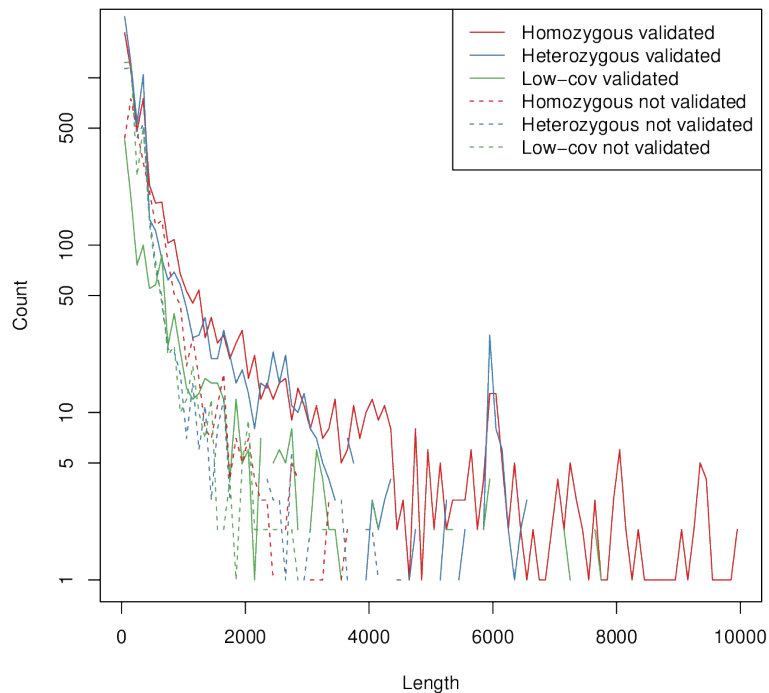
Comparison against Truvari

SVType	INS		DEL	
SVLen	<50	>=50	<50	>=50
number of truvari TP/FP/no-call	3351 FP 7200 TP 15974 not_processed	2835 FP 4095 TP 80 not_processed	880 FP 1925 TP 30196 not_processed	1571 FP 2099 TP 298 not_processed
number of assembly TP/FP/no-call	28422 FP	12409 FP		

Thank you!!

Rocio Esteban

Debapriya Sinha

Chunxiao Liao

And all the organizers of the 2020 eHackathon YOU ARE AWESOME