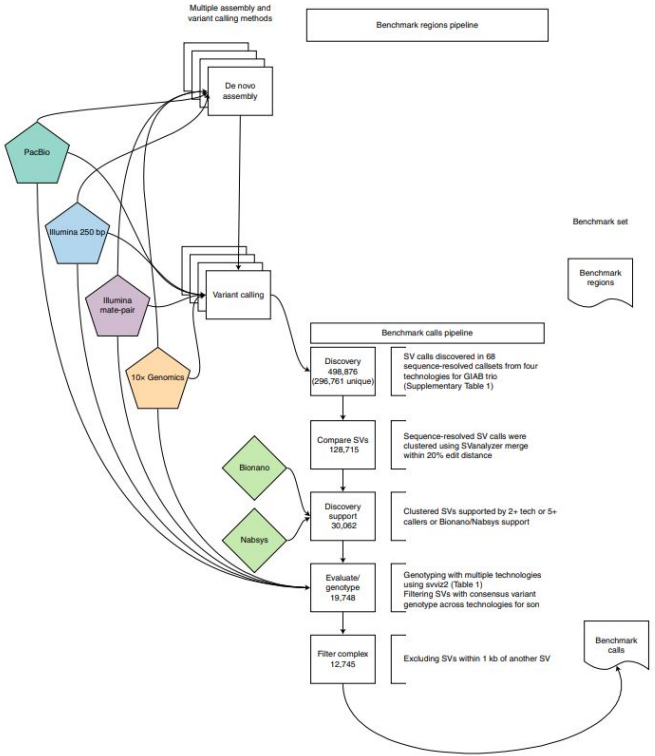
Assembly based validation of SV callsets

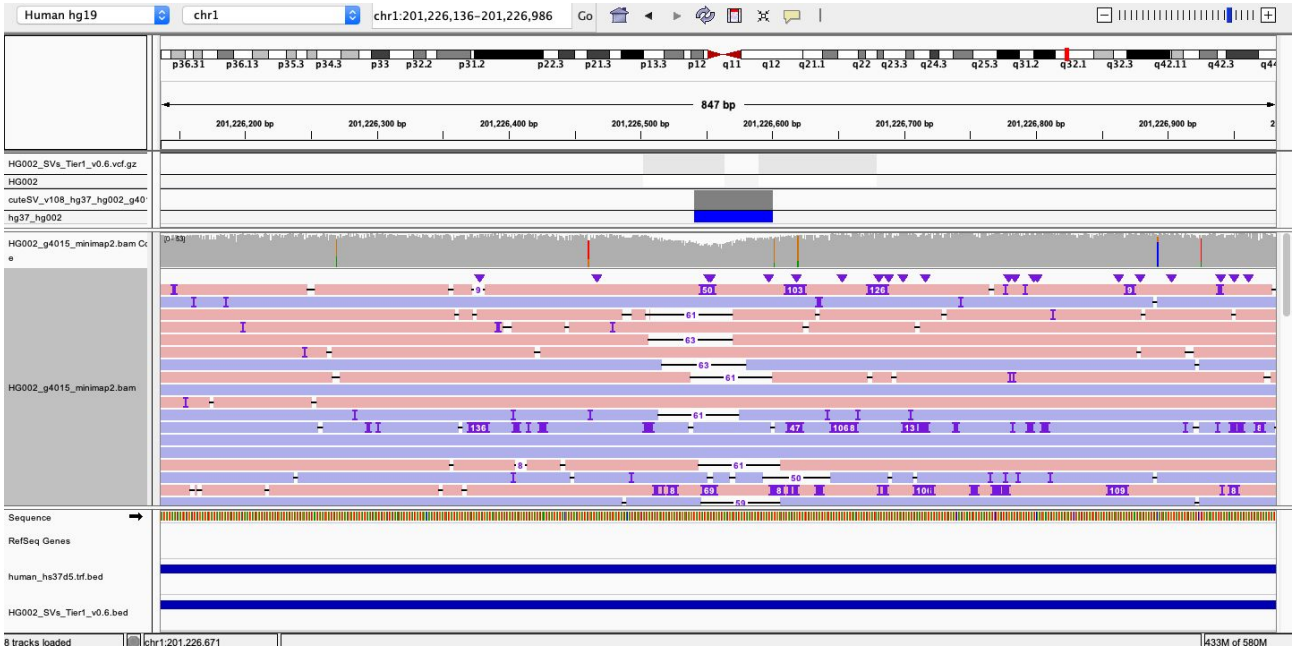
Team 10:

Current SV benchmarks are difficult to form

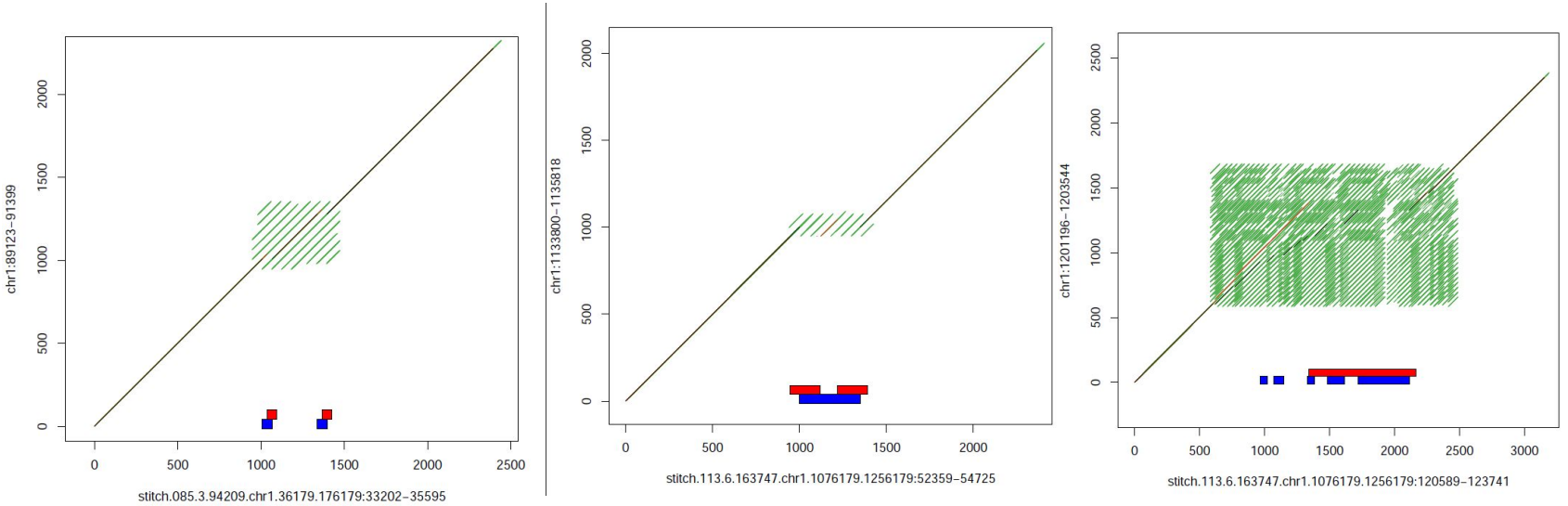
Genome in a Bottle uses multiple algorithms and data types to form a truth set.



Example SV detected from ONT reads missing from benchmark set

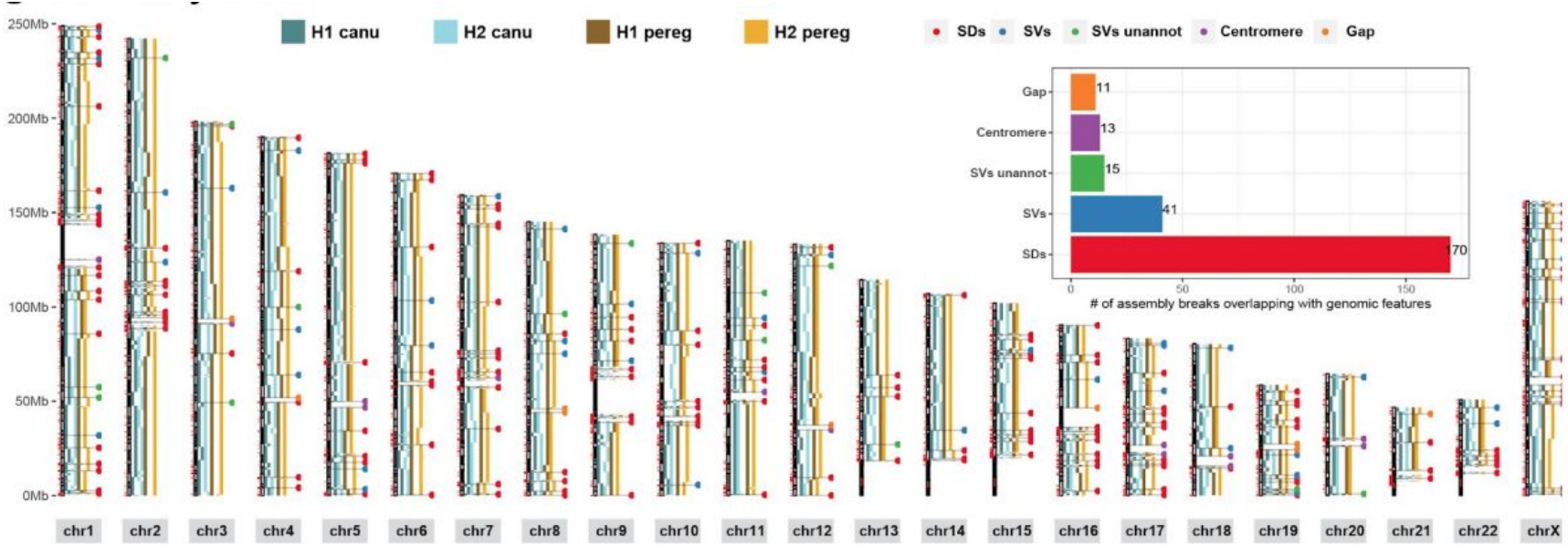


Variant calls can be subject to alignment methods/parameters



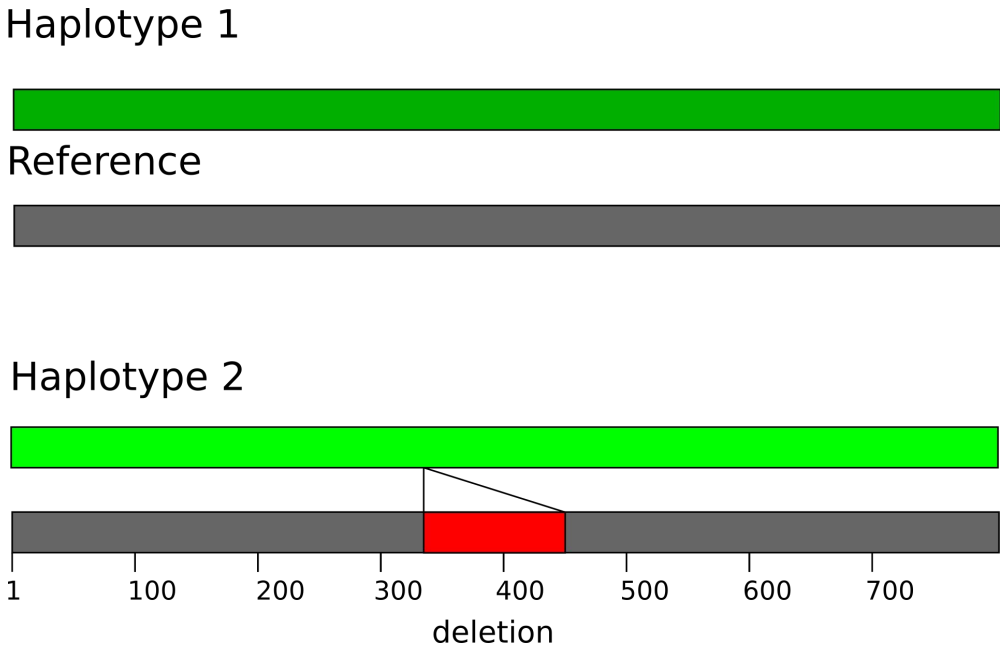
blasr based call, ngm-lr based call

Haplotype-resolved assemblies are becoming widely available



Can we use haplotype-resolved assemblies to annotate variant quality?

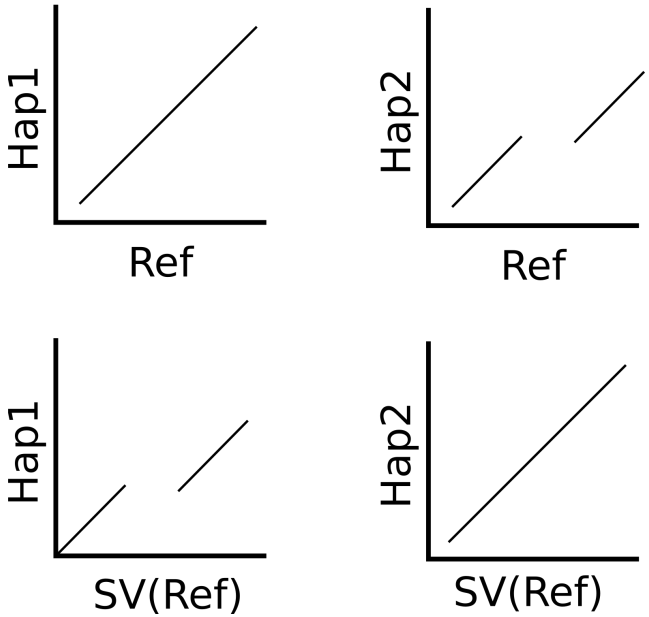
Toy example, heterozygous deletion



Comparison of reference locus and SV(Ref)

Variant call:

chr 1 POS=300,SVLEN=-120



Required steps/data

1. Develop a map to lift from reference to assembly.
2. Compare SV(ref) to assembly, determine a score threshold to validate call.
3. Determine which areas are possible to validate calls with reference.
4. Compare to benchmark datasets.