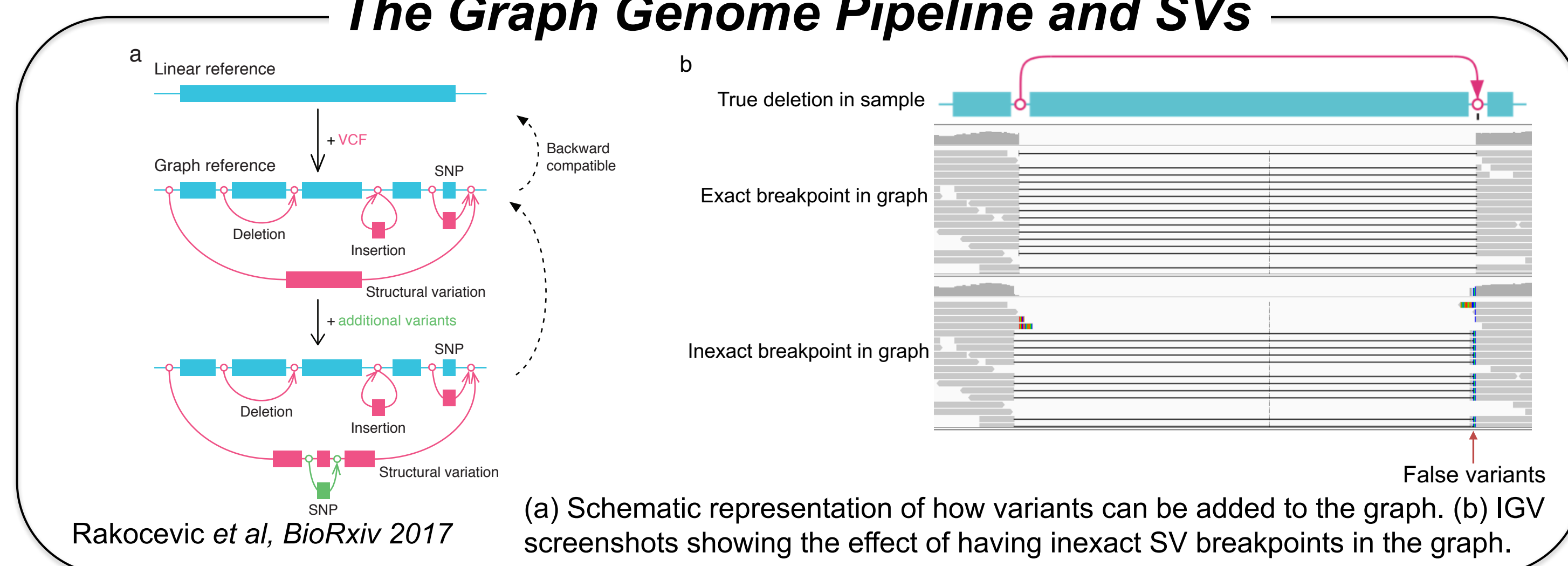


I. Background

The Graph Genome Pipeline and SVs



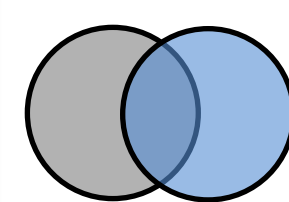
SV Breakpoints

- Not consistent across studies
 - Due to errors in current technology (biochemical, computational, or both)
 - Multiple SV events cluster together in complex regions of the genome (microhomology/repeats/etc.)
 - Overlapping SVs in multiple individuals can be a consequence of convergent evolution caused by different evolutionary events
- Goal: Identify SVs common in the population with consistent breakpoints (evolutionarily conserved SVs fixed in the population)
 - Filter for SVs in HWE that also have well-defined haplotypes

II. Methods

1. Conserved SVs from 1000 Genomes

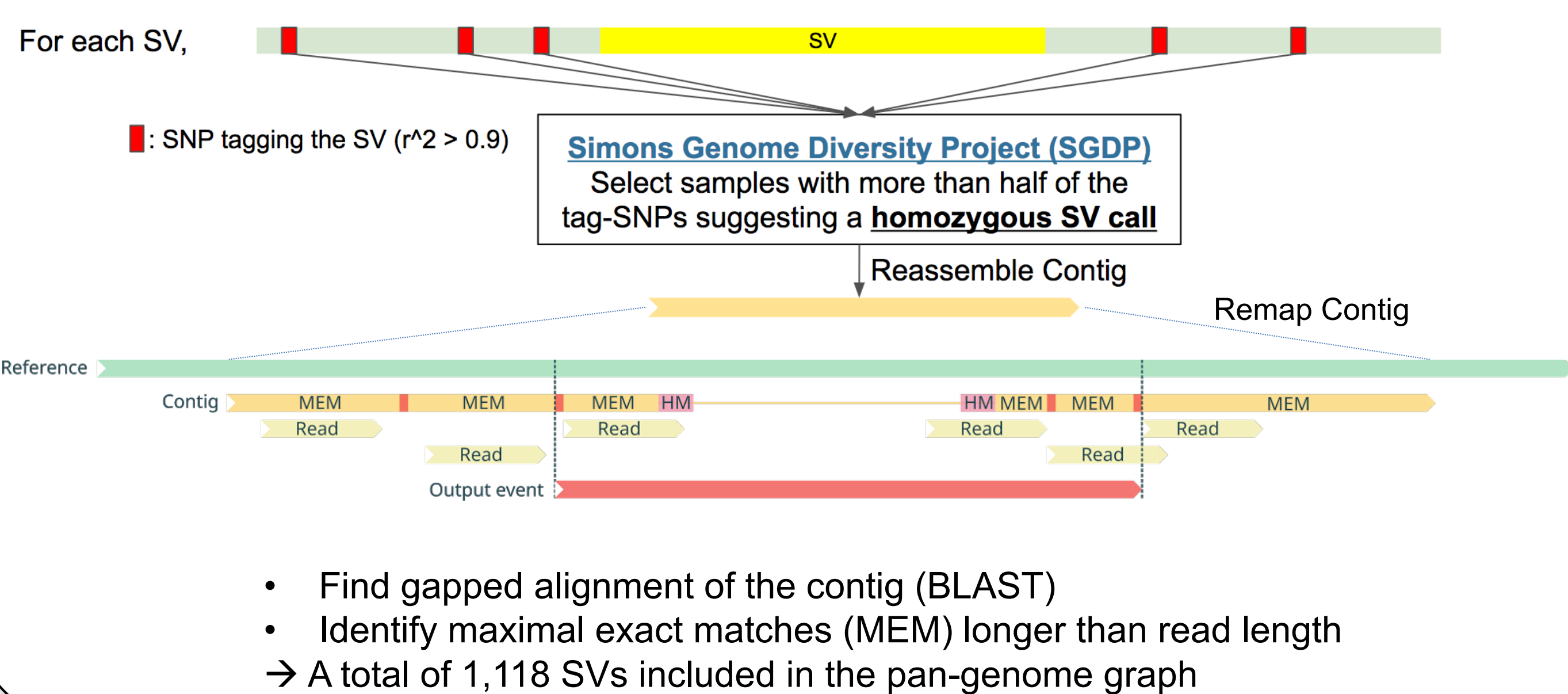
- 1000 Genomes SV (Sudmant et al, 2015)
- HWE $P > 0.05$ (bonferroni-corrected) in all five super population
 - At least one tag-SNP ($r^2 > 0.9$) across all individuals



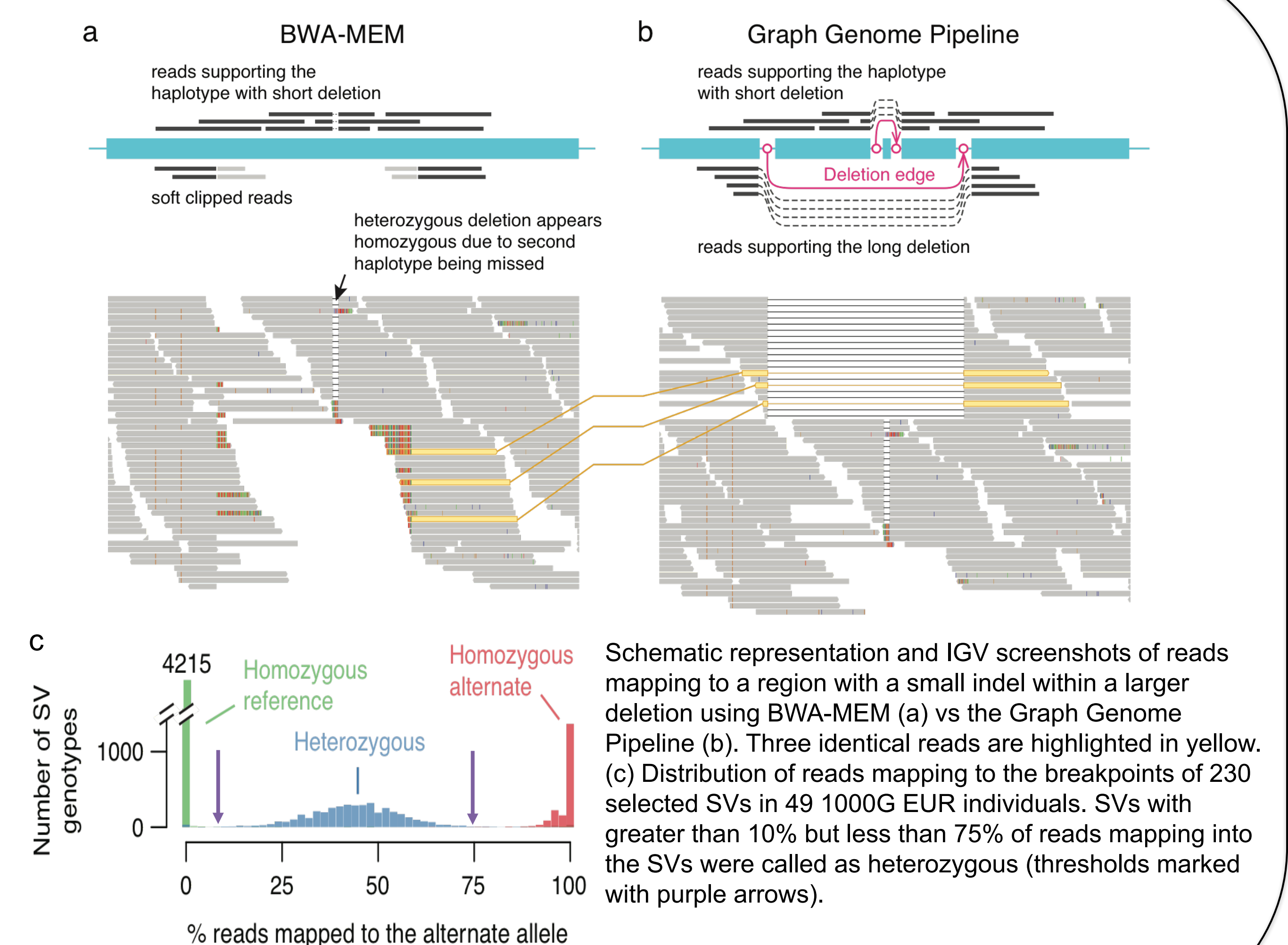
- > 2 studies
- > 2 algorithms
- > 2 technologies
- > 5% frequency
- > 250 unique samples

> 80% of selected SVs were present in at least one high-confidence human genome (CHM1, CHM13, AK1, Huref)

2. Reassembly and Remapping of Alternate Contig



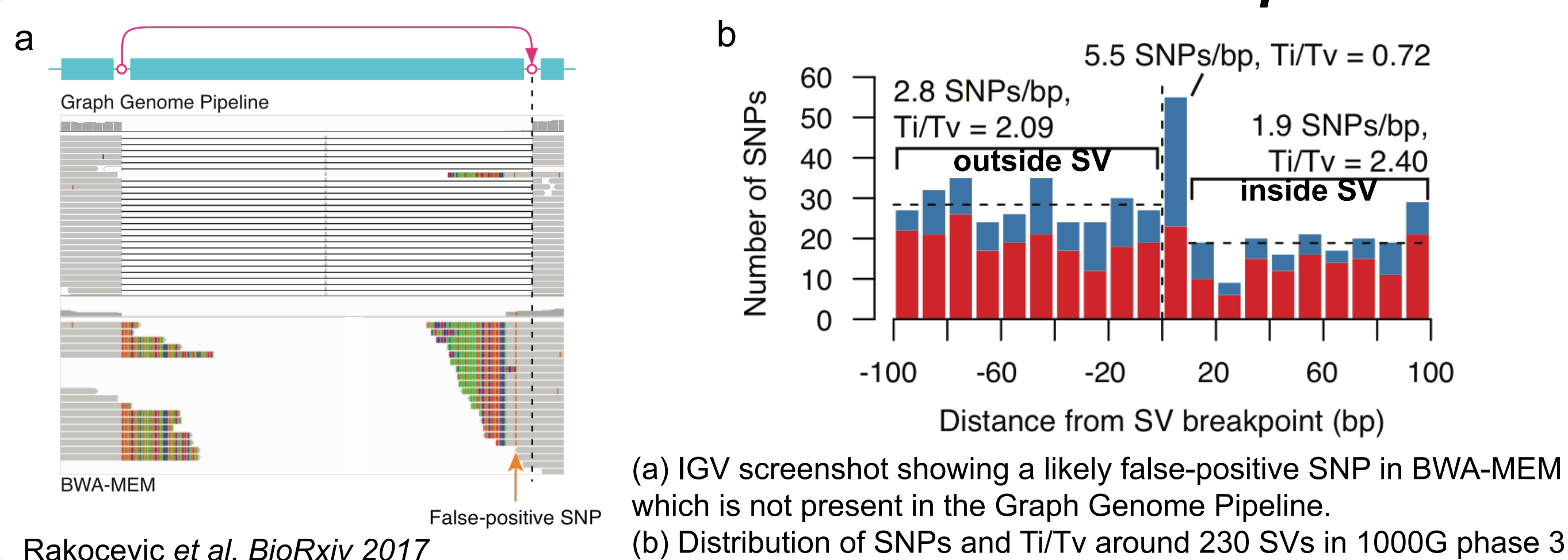
3. Direct Alignment and Genotyping of SVs



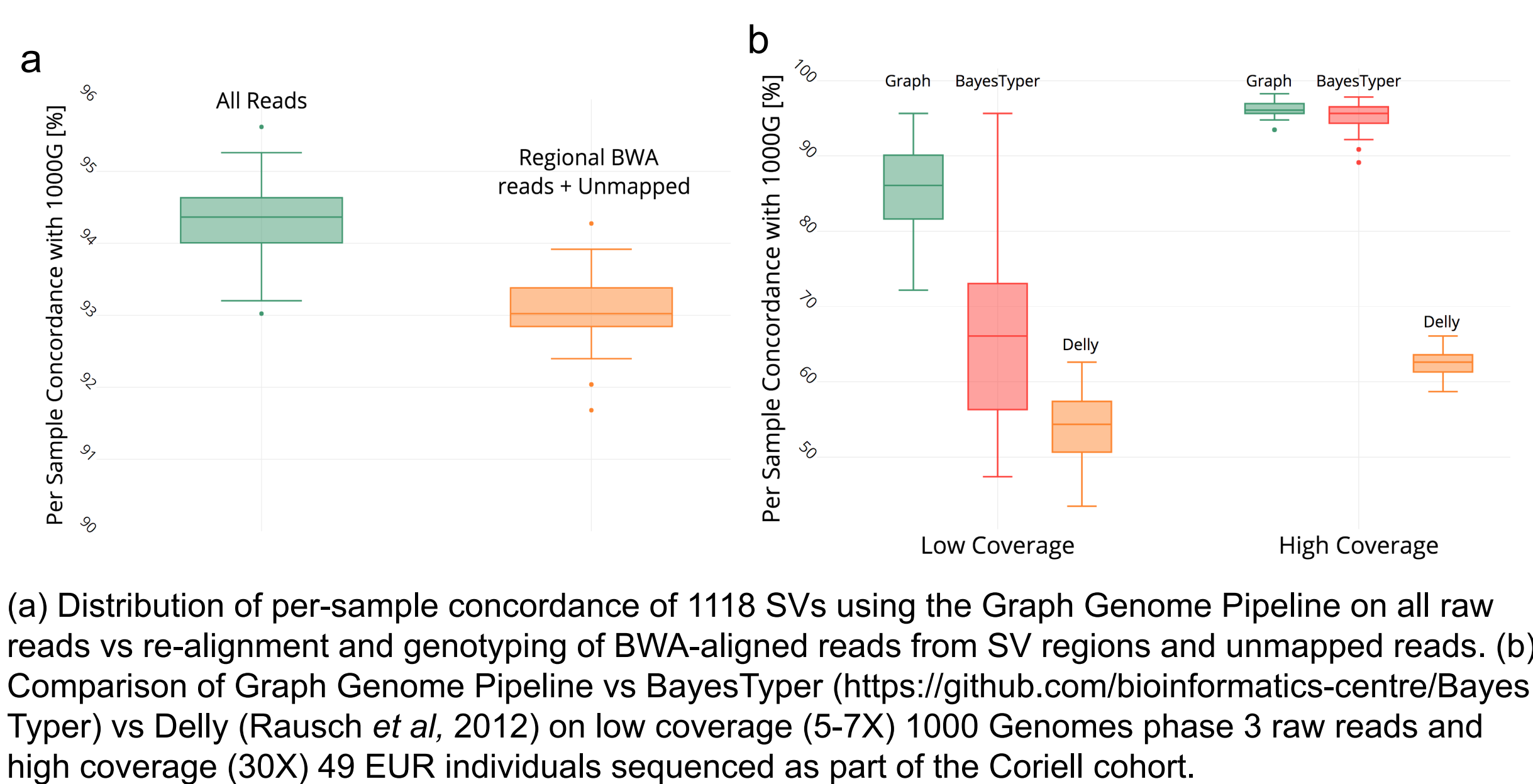
Rakocevic et al, BioRxiv 2017

III. Results

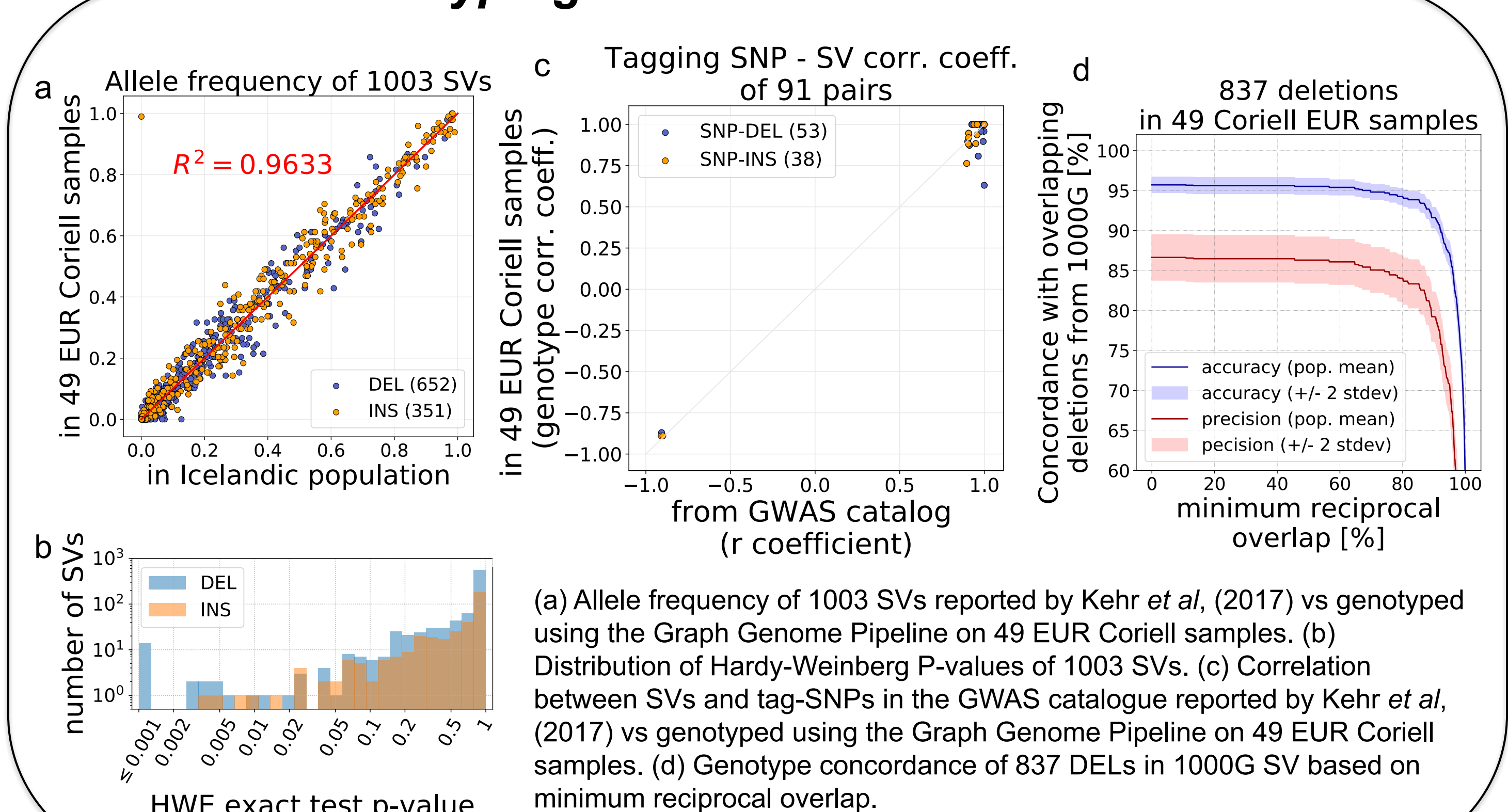
1. Potential False SNPs Around SV Breakpoints



2. SV Genotype Concordance with 1000 Genomes SV



3. Genotyping of deCODE SVs in 1000G EUR



4. SV Confidence Based on Genotype Concordance

Pre-alignment features

- Population evidence (DGV, 1000G, GoNL, Icelandic, etc.)
- Personal genome evidence (CHM1, HuRef, 1000G trio, etc.)
- Consistency of breakpoints in assembled contigs

Graph-alignment features

- Microhomology length/content around flanking regions
- K-mer content of new paths
- Simulation of reads and alignment

**SV genotype
confidence
score**

IV. Discussion

- **We can directly align reads to an SV-augmented pan-genome graph using the SBG graph pipeline**
 - No need for realignment
 - Straightforward genotyping of SVs
 - Rescue of potential FP SNPs/indels near SVs

Current limitations

- Limited to common SVs with consistent breakpoints in regions of high complexity
- SVs with true wobbly breakpoints will lead to FP SNPs

- **General information regarding the SBG graph pipeline can be found @ poster PgmNr 1263/F (Kallberg et al.)**