GIAB TR Project Deliverables

Deliverables

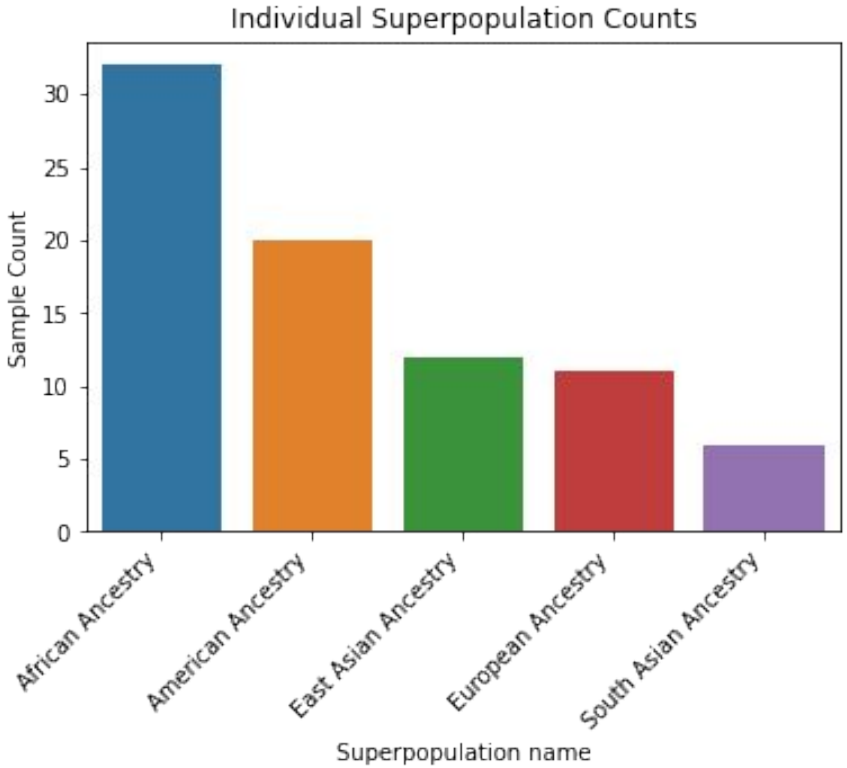
- pVCF using 86 assemblies
- TR Catalog
  - Truvari anno trf - annotates VCF entries with catalog
- HG002 Benchmark
  - HG002 Benchmark Regions
  - HG002 Separate VCF
- Benchmark Procedure
  - Tool or pipeline for how to benchmark TRs
  - RTG, Truvari bench, Truvari phab
- Biological Insights

pVCF

- 3 Projects
  - HPRC (47)
  - Eichler (34)
  - Li (4)
- 172 haplotypes
- 86 samples
- 78 individuals

Replicates	
HG00733	3
NA19240	2
NA24385	3
HG03486	2
HG02818	2
NA12878	2

Are there other assemblies we want to put into this pVCF?



TR Catalog

Columns:

- Position - chrom, start, end
- Repeats - TRF annotations of sub-regions (json)
- Pathogenic - Name of known pathogenic sites (e.g. ATXN3)
- Codis - Name of known Codis sites (e.g. CSF1PO)
- Score - How resolvable/consistently represented are variants in the region?

HG002 Benchmark

- We can pull HG002 (HPRC assembly) from the pVCF or use different VCF
  - Do we want the same VCF as what GIAB is building for WG benchmarking?
- Use TR Catalog to annotate the above VCF
- Subset TR Catalog to ‘Ranks’ or Tiers
  - Current beta version of the benchmark regions is ranked based on HPRC/Adotto and TrioHifiAsm/GIAB consistency as well as `truvari anno trf` annotateable entries >= 5bp

Benchmark Procedure

- Have a prototype of steps that can compare any (resolved) variants with truvari bench/phab
- Formalize the steps
  - How automated do the steps need to be? Single button?
- Build a report
  - Do we want a single precision/recall metric?
  - Currently ‘phab’ regions are FN/FP inside the truvari summary and then reevaluated separately.
  - A third step of ‘combine truvari bench/phab results’ could build the report once we figure out how to count True/False calls.
  - Stratifications(?!)

Biological Insights

- Amount of SV in TR regions
  - Inside/outside telomeric/centromeric
- How many TR regions intersect genes / regulatory elements
- Characteristics of STR and VNTR
- Observations of Variants in Pathogenic regions
- CODIS
- Other?

CHM13

- Can recreate pVCF and TR Catalog for CHM13
- Separate sub-section of paper where we say we did it
- Provide a few insights from comparing how CHM13 and GRCh38 results are different (Amount of variation, centromeres/telomeres)