

GIABTR

Adam English
HGSC@BCM
September 6, 2022

Benchmarking Adotto

- Three replicates of HG002/NA24385
- Benchmark against CMRG with RTG and Truvari

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


RTG + CMRG smallvar

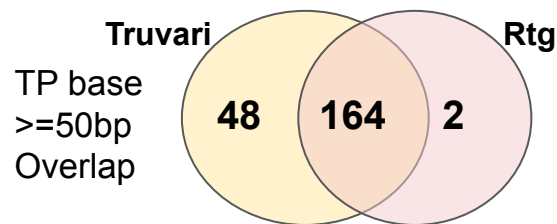
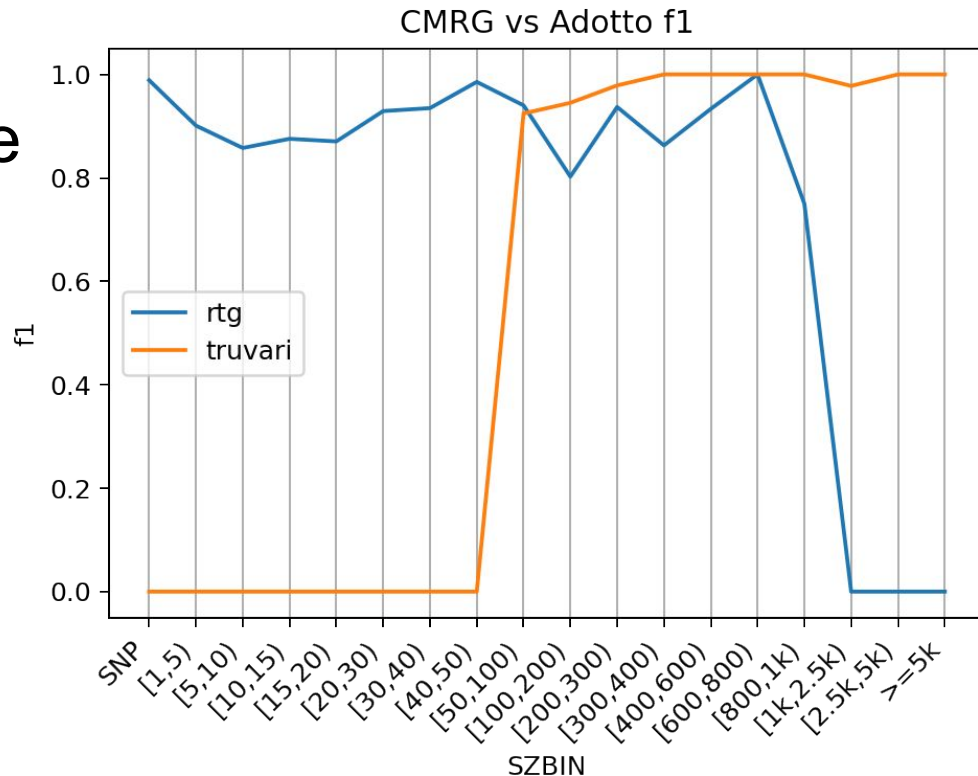
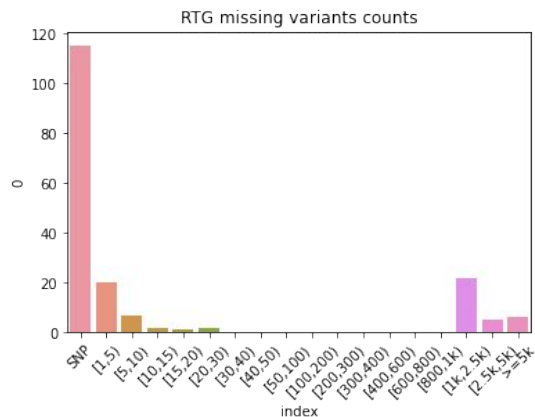
Replicate	True-pos baseline	True-pos call	False-pos	False-neg	Precision	Sensitivity	F-measure
eichler	20,271	22,103	9,599	960	0.697	0.955	0.806
hprc	21,131	22,986	479	100	0.980	0.995	0.987
li	20,288	22,117	8,221	943	0.729	0.956	0.827

Truvari + CMRG SV

eichler	209	209	11	7	0.950	0.968	0.959
hprc	213	213	7	3	0.968	0.986	0.977
li	210	210	17	6	0.925	0.972	0.948

CMRG performance by Size

- Truvari
 - doesn't analyze calls <50bp
- RTG
 - Doesn't analyze calls >=1k
 - Struggles to compare variants 50-1k
 - Reports uneven performance
 - Filters 'complex' regions 

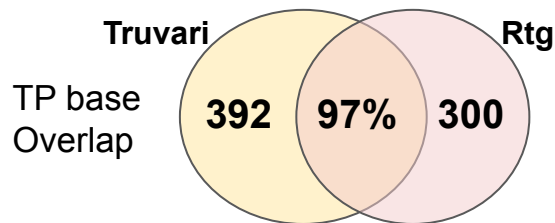
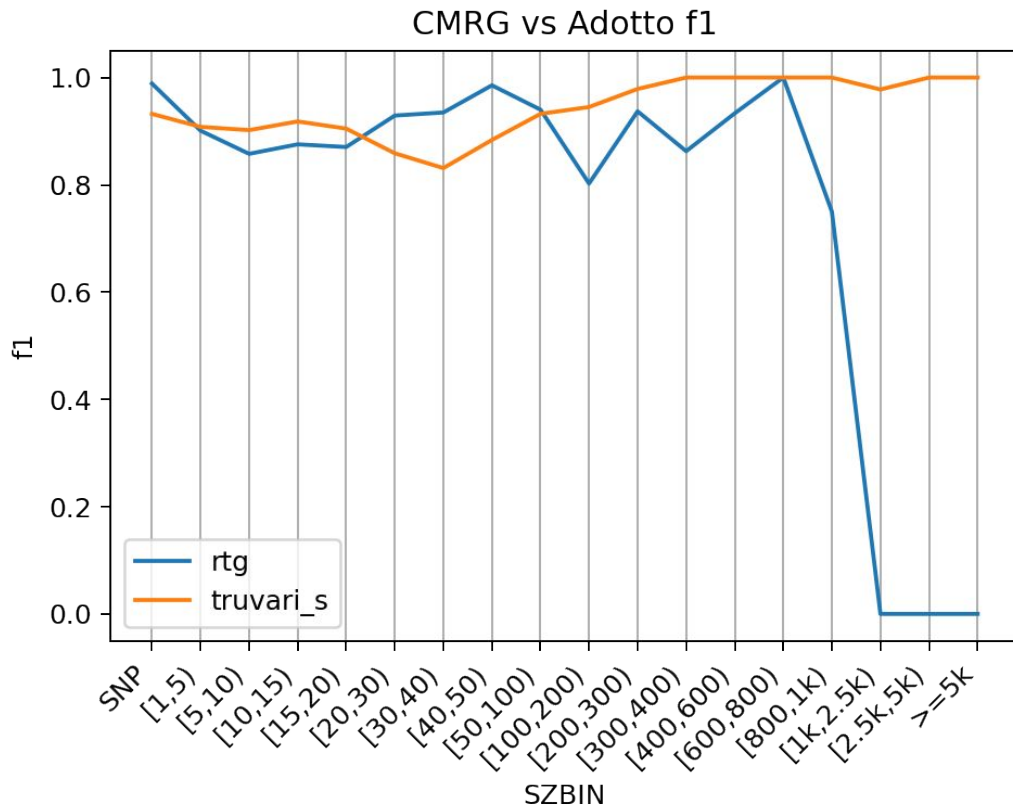


Truvari on small variants

- Set min size (-s) to 0
- New 'unroll' seqsim
- Still needs dynamic scoring
 - 70% similarity for small events might be weird.

[5,50)bp overall performance

prog	comp	precision	recall	f1
rtg	eichler	0.824	0.833	0.829
truvari	eichler	0.884	0.849	0.866
rtg	hprc	0.883	0.876	0.880
truvari	hprc	0.935	0.860	0.896
rtg	li	0.763	0.808	0.785
truvari	li	0.842	0.839	0.841

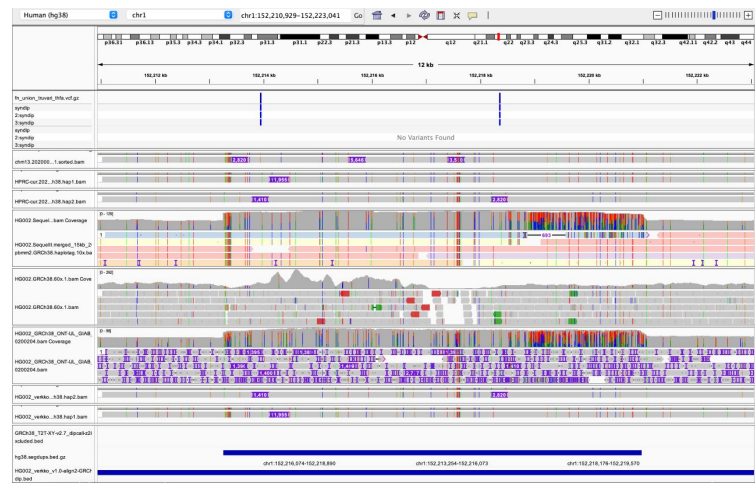
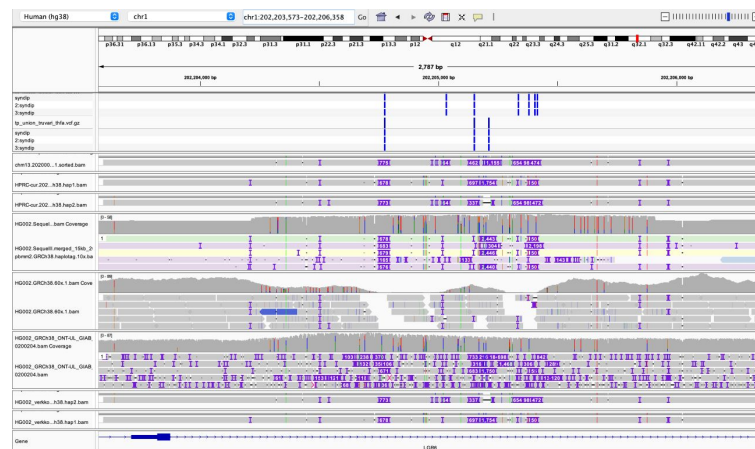
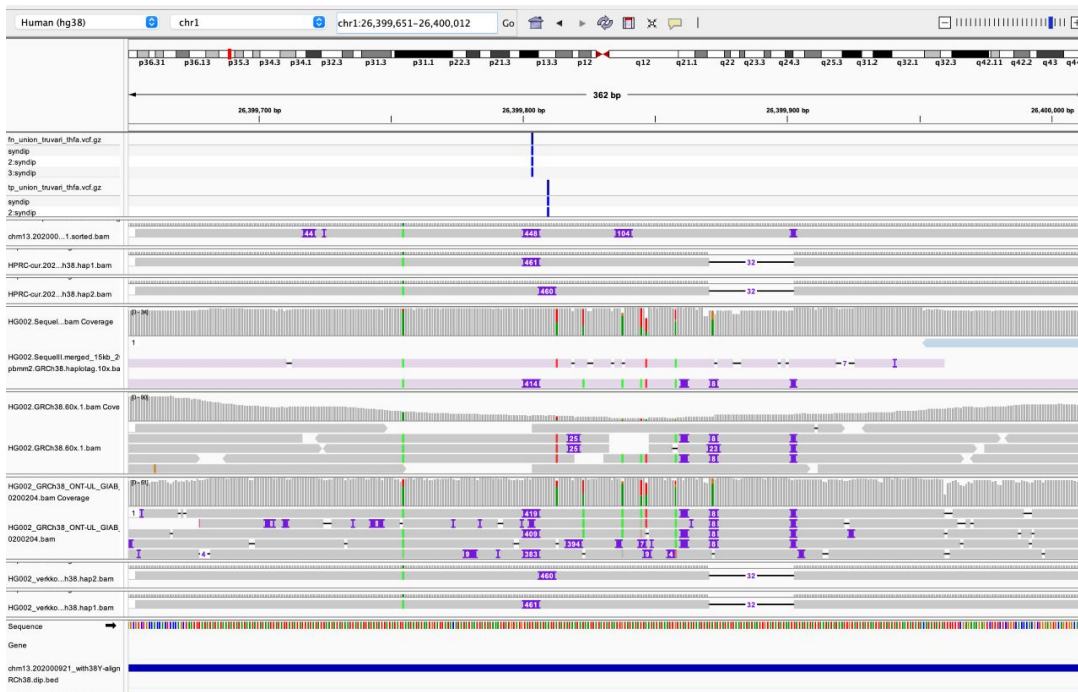


GIAB TrioHifiAsm Benchmarking

Program	Comp	True pos baseline	True pos-call	False-pos	False-neg	Precision	Sensitivity	F-measure
RTG	reichler	4,474,711	4,644,812	454,835	187,930	0.9108	0.9597	0.9346
Truvari	reichler	21,891	21,891	3,153	8,033	0.874	0.731	0.796
RTG	li	4,470,804	4,642,165	600,025	191,837	0.8855	0.9589	0.9207
Truvari	li	21948	21,948	3,261	7,976	0.870	0.733	0.796
RTG	hprc	4,476,465	4,658,725	119,799	186,176	0.9749	0.9601	0.9674
Truvari	hprc	22,384	22,384	2,768	7,540	0.889	0.748	0.812

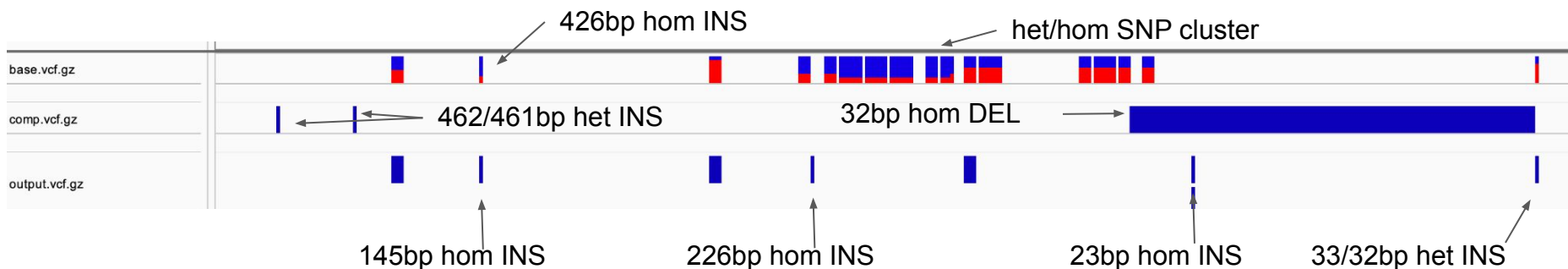
Why do the SVs have lower Sensitivity?

Manual inspection reveals non-matching variants (likely) comprise identical alleles through different representations



Truvari Phab

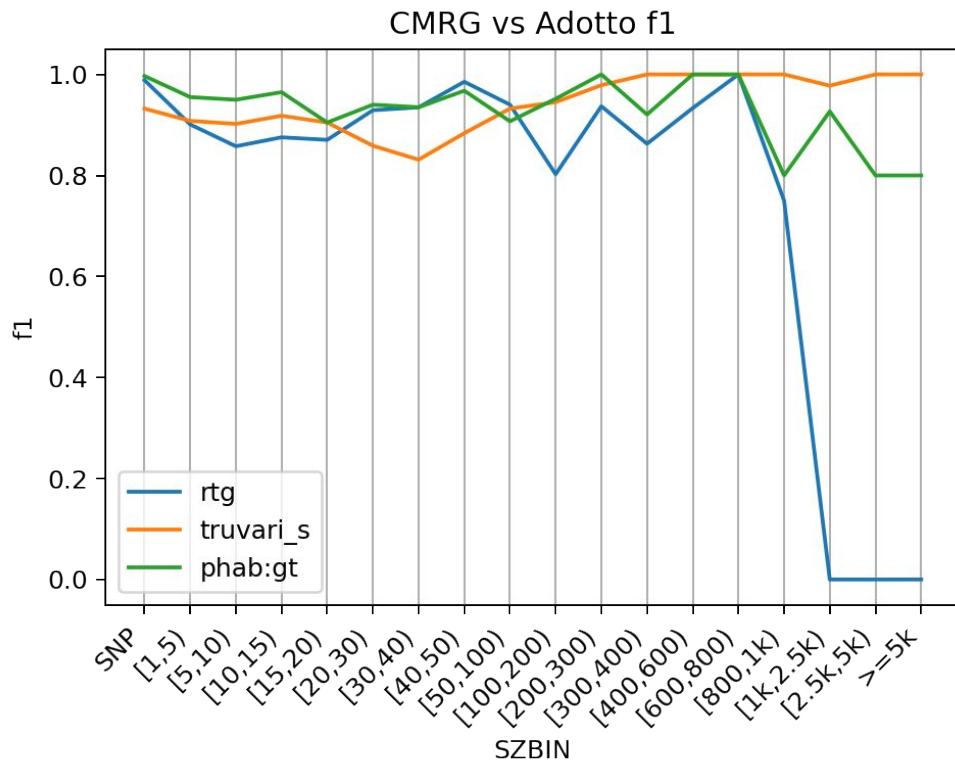
- The inconsistent representations between Adotto & GIAB on HPRC should be explained by the minimap2 parameters.
- Truvari Phab uses MSA to harmonize variants
 - Reconstruct haplotypes in region
 - Run MAFFT over haplotypes and reference sequence
 - Recall variants



Truvari Phab Result

- Phab:GT
 - TP = Present in both samples
 - FP/FN = Present in only one sample

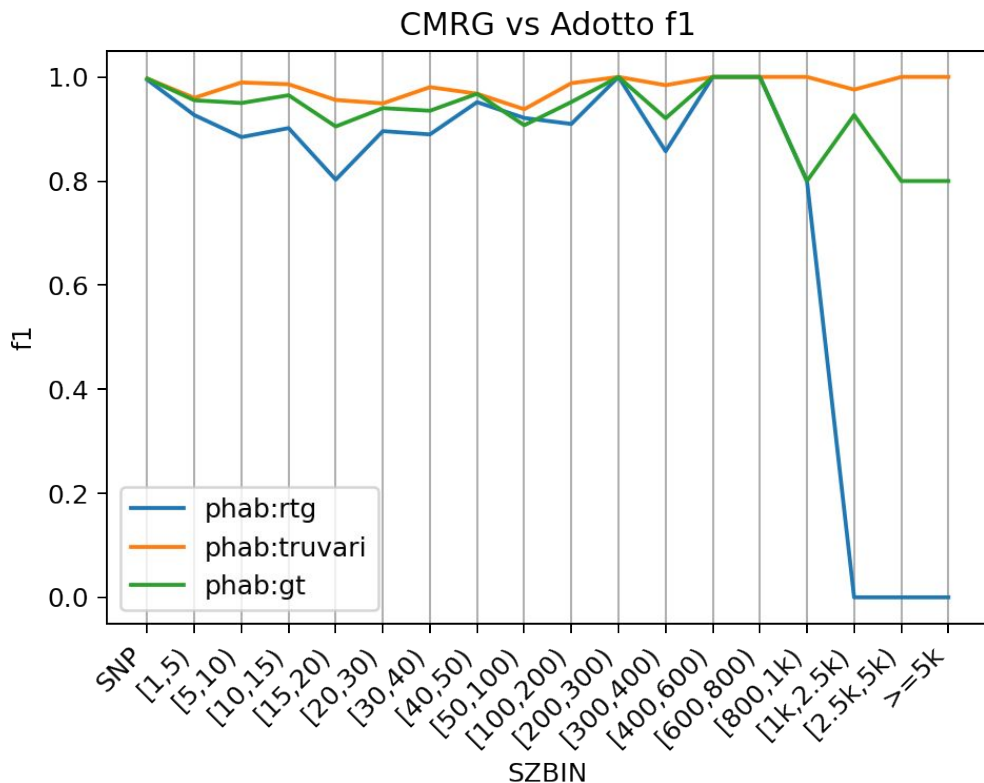
	RTG	Truvari	Phab:GT
tpbase	23,380	23,472	24,154
tp	25,452	23,472	24,154
fp	846	2,878	322
fn	736	820	375
recall	0.969	0.966	0.985
precision	0.968	0.891	0.987
f1	0.969	0.927	0.986



Truvari Phab Result

- Direct comparison of GT presence is strict.
- Use RTG/Truvari on harmonized variants to allow ambiguity

	Phab:GT	Phab:RTG	Phab:Truvari
tpbase	24,154	23,859	24,223
tp	24,154	23,903	24,223
fp	322	541	248
fn	375	639	303
recall	0.985	0.978	0.990
precision	0.987	0.974	0.988
f1	0.986	0.976	0.989



Phab:Truvari details

24,153 have identical sequence
24,164 have identical size
24,204 have identical position
~99.8% matching variant similarity

Need to understand remaining differences

```
#  
# Total 24240 calls across 3 VCFs
```

```
#  
#File      NumCalls  
Phab:GT      24154  
Phab:Truvari 24223  
Phab:RTG      23859  
#  
# Summary of consistency
```

```
#  
#VCFs      Calls      Pct  
3           23832      98.32%  
2           332       1.37%  
1           76        0.31%
```

```
#  
# Breakdown of VCFs' consistency  
#  
#Group      Total      TotalPct  PctOfFileCalls  
111         23832      98.32%    98.67% 98.39% 99.89%  
110         319       1.32%     1.32% 1.32% 0%  
010         61       0.25%     0% 0.25% 0%  
001         14       0.06%     0% 0% 0.06%  
011         11       0.05%     0% 0.05% 0.05%  
101         2        0.01%     0.01% 0% 0.01%  
100         1        0.00%     0.00% 0% 0%
```

Next Steps

- Fully understanding CMRG benchmarking will inform TR specific matching
- Improve Truvari bench --sizemin 0 with something like dynamic matching
- Truvari phab is only useful as a 'spot checker' due to slow speed
 - Find faster MSA(?)
 - Find patterns for when Phab is needed
- Start making Tiers based on representational ambiguities
 - Tier1 : RTG vcfeval & Truvari bench resolvable
 - Will need multiple call sets e.g. Adotto/THFA/CMRG/TRGT/GangSTR/HipSTR
 - Tier 2 :
 - Phab is probably necessary for accurate comparison