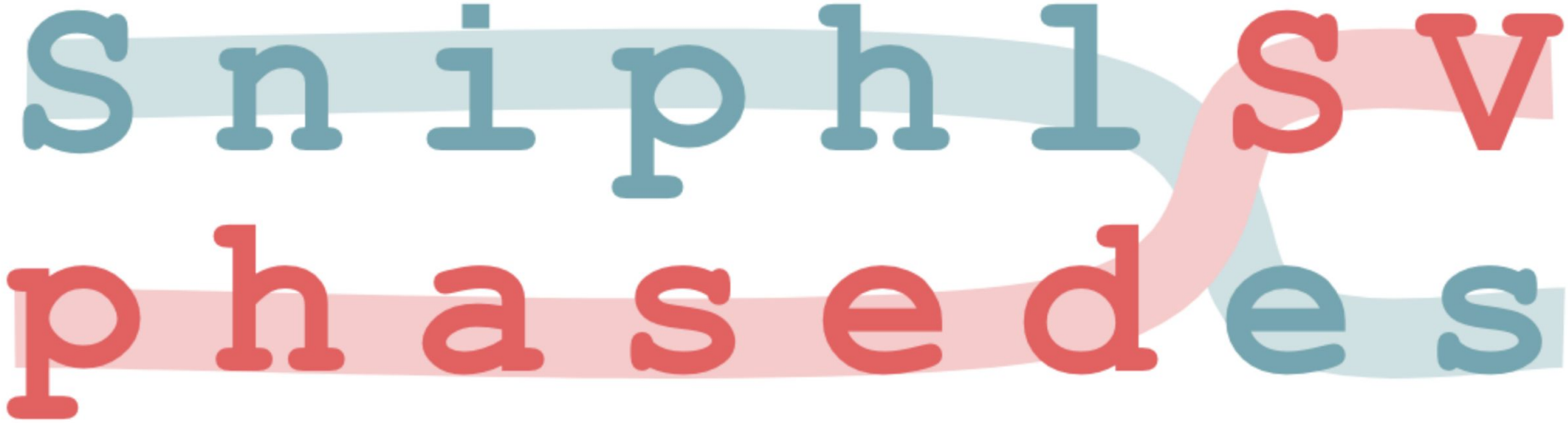
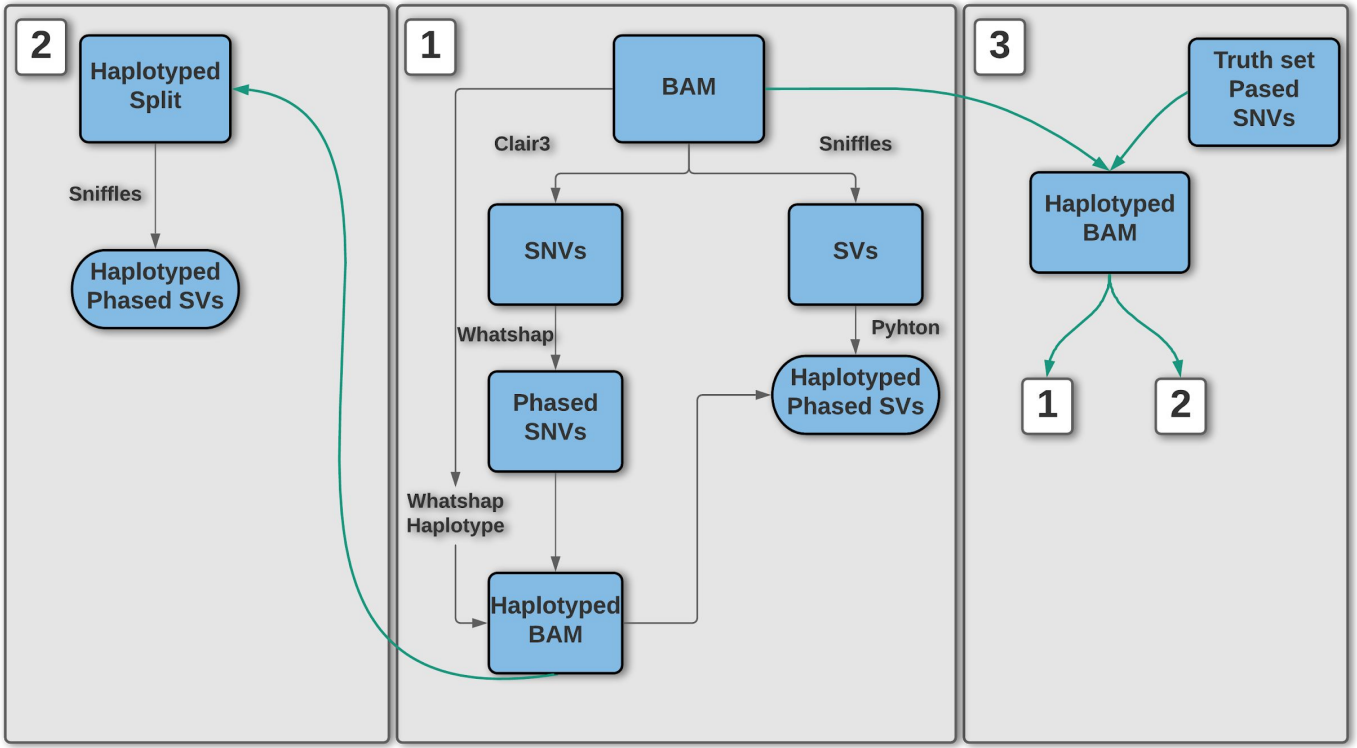


Group 5

Medhat Mahmoud
Elbey Aliyev
Kimberley Billingsley
Moez Dawood
Wouter De Coster
Claude Sinner
Arun Subramaniyan
Pamella Tater
Tony (Tsung-Yu)

Workflow


2
Haplotyped Split
Sniffles
Haplotyped Phased SVs
1
BAM
Clair3
Sniffles
SNVs
SVs
Whatshap
Pyhton
Phased SNVs
Haplotyped Phased SVs
Whatshap Haplotype
Haplotyped BAM
3
Truth set Pased SNVs
Haplotyped BAM
1
2

Chr20 Raw SV Count Breakdown of Phased vs Unphased SVs

	CONFLICT			SVTYPE						Phased		Zygosity	
	0	1	2	DEL	DUP	INV	INS	BND	TRA	UNK	Yes	No	HMZ	HET
CONFLICT=0	279	0	0	135	2	0	142	0	0	0	279	0	0	0
CONFLICT=1	0	59	0	14	1	1	43	0	0	0	59	0	0	59
CONFLICT=2	0	0	19	3	1	0	9	6	0	0	0	19	0	19
SVTYPE=DEL	135	14	3	379	0	0	0	0	0	0	149	230	227	17
SVTYPE=DUP	2	1	1	0	27	0	0	0	0	0	3	24	23	2
SVTYPE=INV	0	1	0	0	0	20	0	0	0	0	1	19	19	1
SVTYPE=INS	142	43	9	0	0	0	584	0	0	0	185	399	390	52
SVTYPE=BND	0	0	6	0	0	0	0	50	0	0	0	50	44	6
SVTYPE=TRA	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SVTYPE=UNK	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PS	279	59	0	149	3	1	185	0	0	0	338	0	0	59
Unphased	0	0	19	230	24	19	399	50	0	0	0	722	703	19
Homozygous	0	0	0	227	23	19	390	44	0	0	0	703	703	0
Heterozygous	0	59	19	17	2	1	52	6	0	0	59	19	0	78

Group 1 Focus:

- 76% of heterozygous SV were phased
  - 19 SVs were unphased
  - Majority are deletions and insertions

Sniphles: Using Sniffles - but phased

1. Start from a BAM with phase per read annotated (e.g. from WhatsHap, LongShot)
2. Identify phase blocks and split in monophasic (homozygous), diphasic (heterozygous) and unphased blocks
3. Loop over haplotype blogs, split BAM using temporary files and make phased bams => good opportunity for massive parallelization
4. Run sniffles
5. Filtering
6. Concatenating the parts
7. Merging haplotypes back together using SURVIVOR and force-calling Sniffles

BONUS:

8. Can handle cram input (Sniffles CAN'T)
9. Will output a correctly SORTED VCF (Sniffles doesn't care)
10. Homozygous SV phasing :P