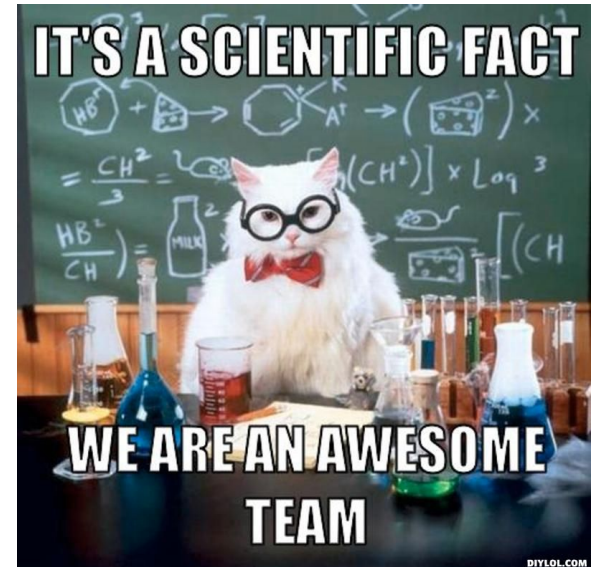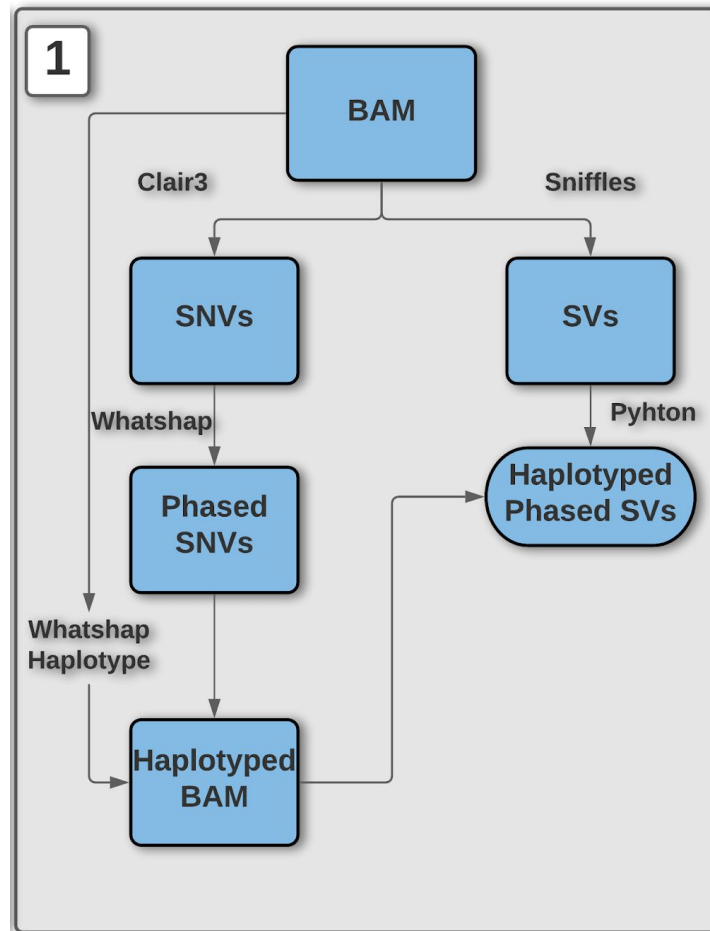# SniphlSV phasedes

Group 5

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

IT'S A SCIENTIFIC FACT
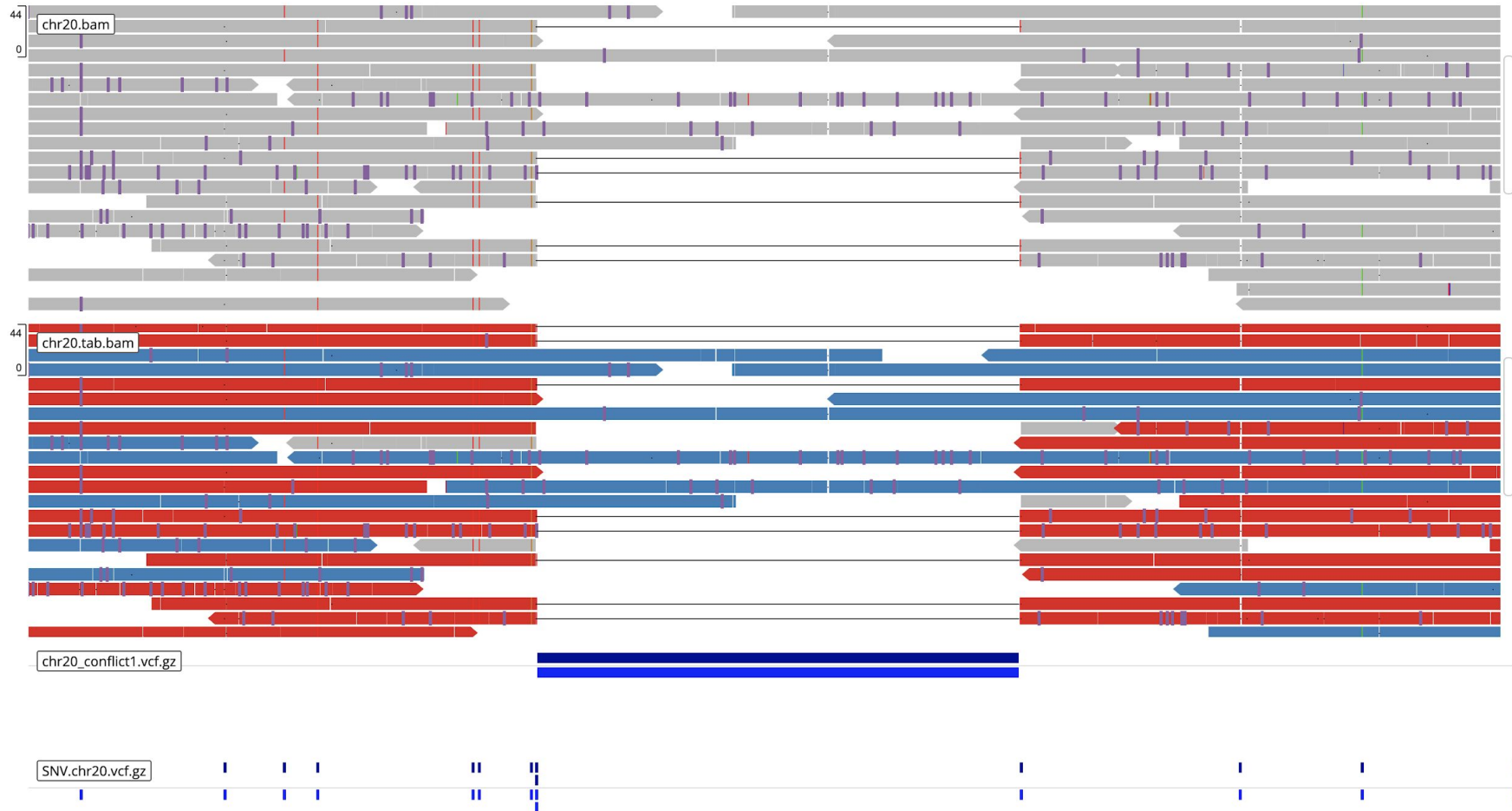WE ARE AN AWESOME TEAM
DIYLOL.COM

Step 1

- Used preexisting tool to phase SVs

- Analyse the output

- Enhance the algorithm
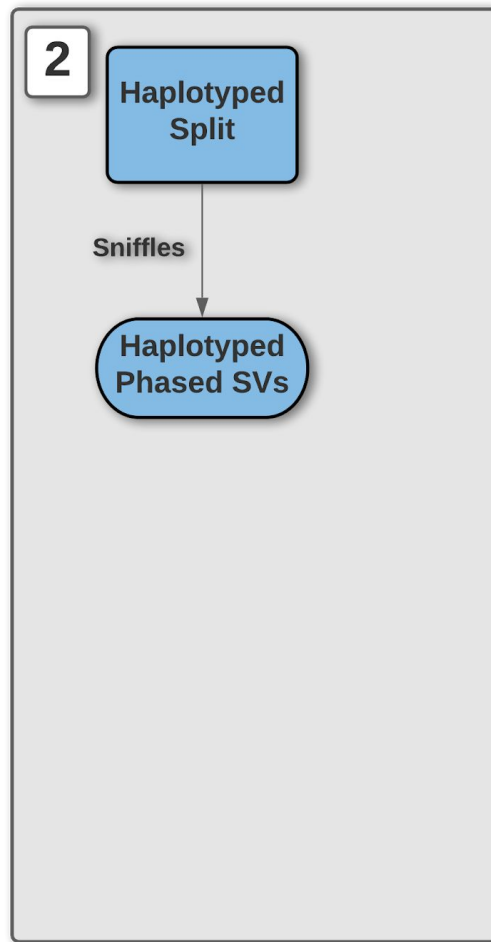
- 78.15% of heterozygous SV were phased
- The unphased SVs distributed as follow, where 14 % fill in centromere.

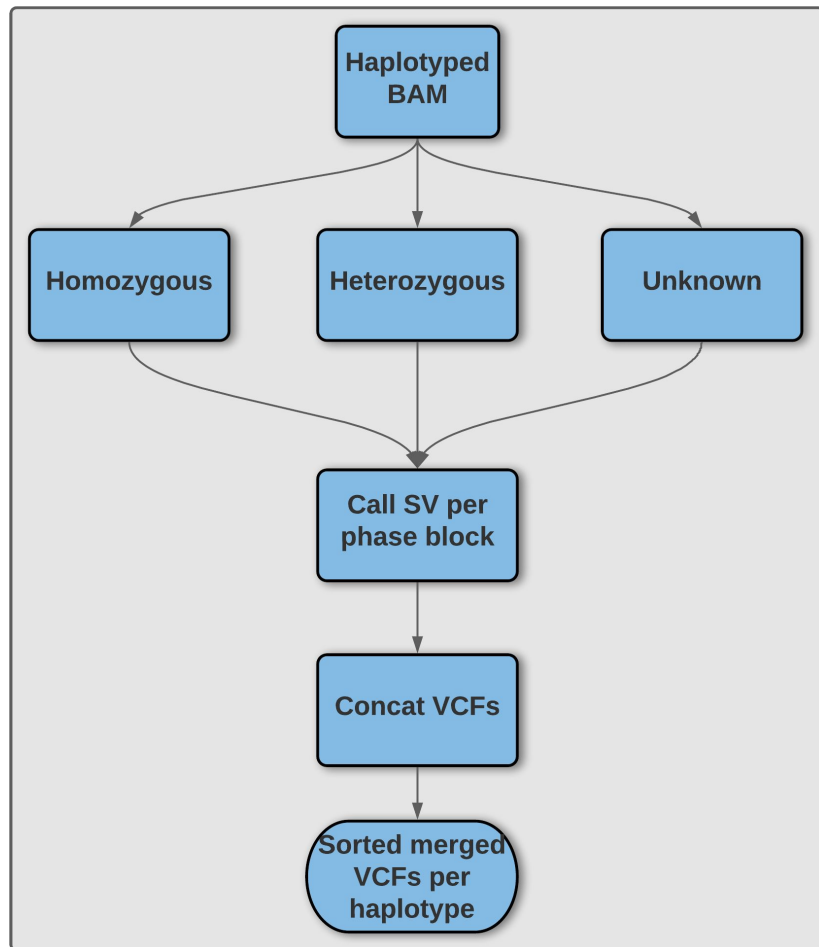| INS	DEL	BND	DUP	INV
66%	22%	7%	2%	1%

Step 2

- Begin with Haplotype file from Group 1
- Develop a divide and conquer tool
    - Extract reads based on phase block and haplotype
    - Recall Sniffles

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

- ❖ Prerequisites
  - ➢ Sniffles
  - ➢ Bcftools
  - ➢ Mosdepth
  - ➢ Survivor

```
usage: sniphles.py [-h] -b BAM -v VCF [-l LOG_FILE] [-s MINIMUM_SUPORT_READ] -n NAME
```
A

```
Use Sniffles on a phased bam to get phased SV calls

optional arguments:
  -h, --help            show this help message and exit
  -b BAM, --bam BAM     Phased bam to perform phased SV calling on (default: None)
  -v VCF, --vcf VCF     output VCF file (default: None)
  -l LOG_FILE, --log LOG_FILE
                        Log file (default: sniphles.log)
  -s MINIMUM_SUPORT_READ, --minimum_suport_read MINIMUM_SUPORT_READ
                        Minimum support read to call SV equals to -s in sniffles (default: 4)
  -n NAME, --name NAME  Sample name of output VCF (default: None)
```

```
root@7b1fbf0cf82b:/opt/notebooks# pytho
Working on chromosome 20
Working on block 20:0-63025520_Hu
```
B

```
Merging across chromosomes
Writing to /tmp/bcftools-sort.UgOGz6
Merging 1 temporary files
Cleaning
Done
Done

real    12m6.589s
user    13m50.216s
sys     0m18.956s
root@7b1fbf0cf82b:/opt/notebooks#
```
C

Thanks!

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