

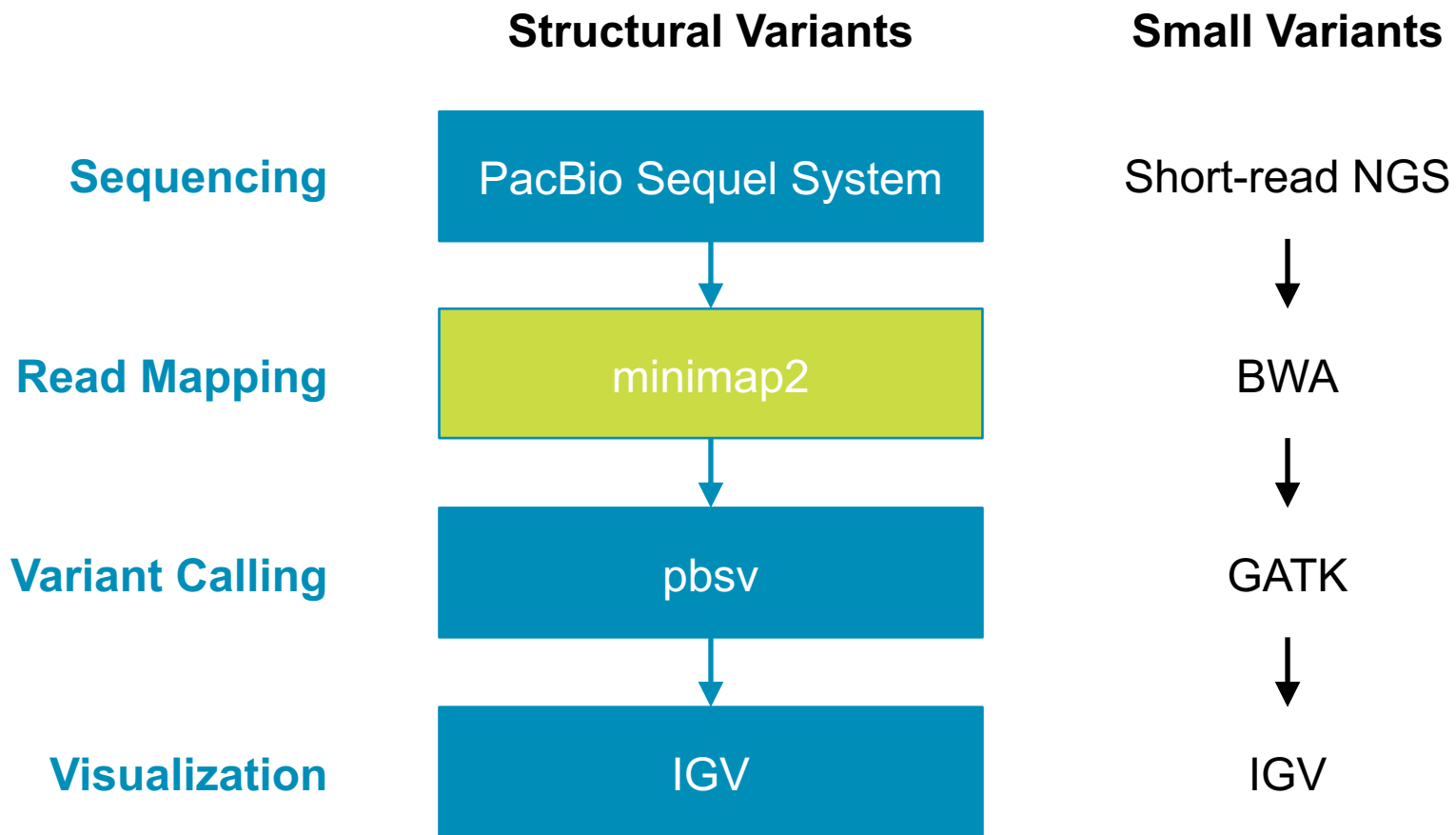


PACBIO®

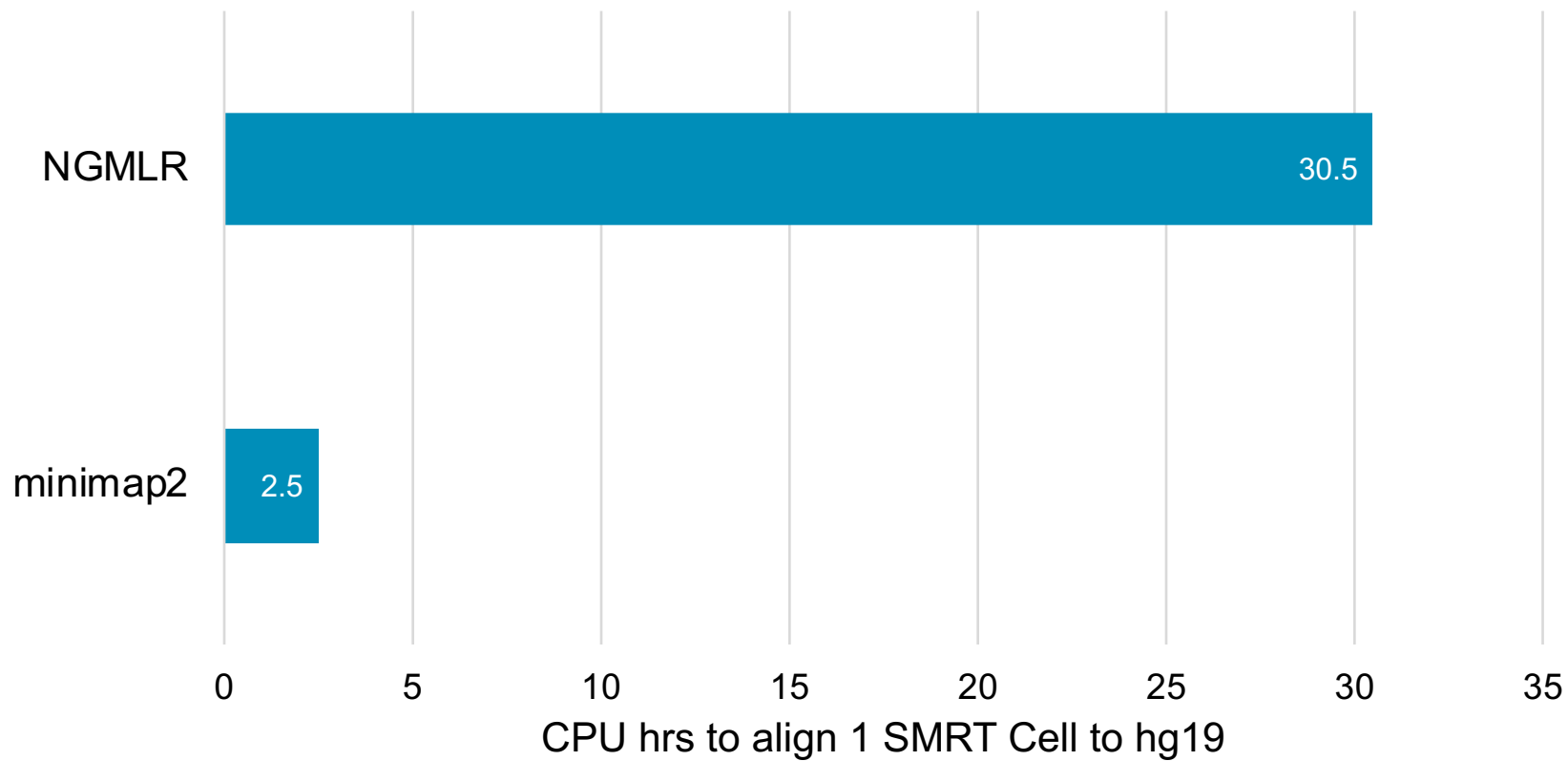
Calling all variants: fast, accurate, population-scale structural variant analysis

NA UGM 2018

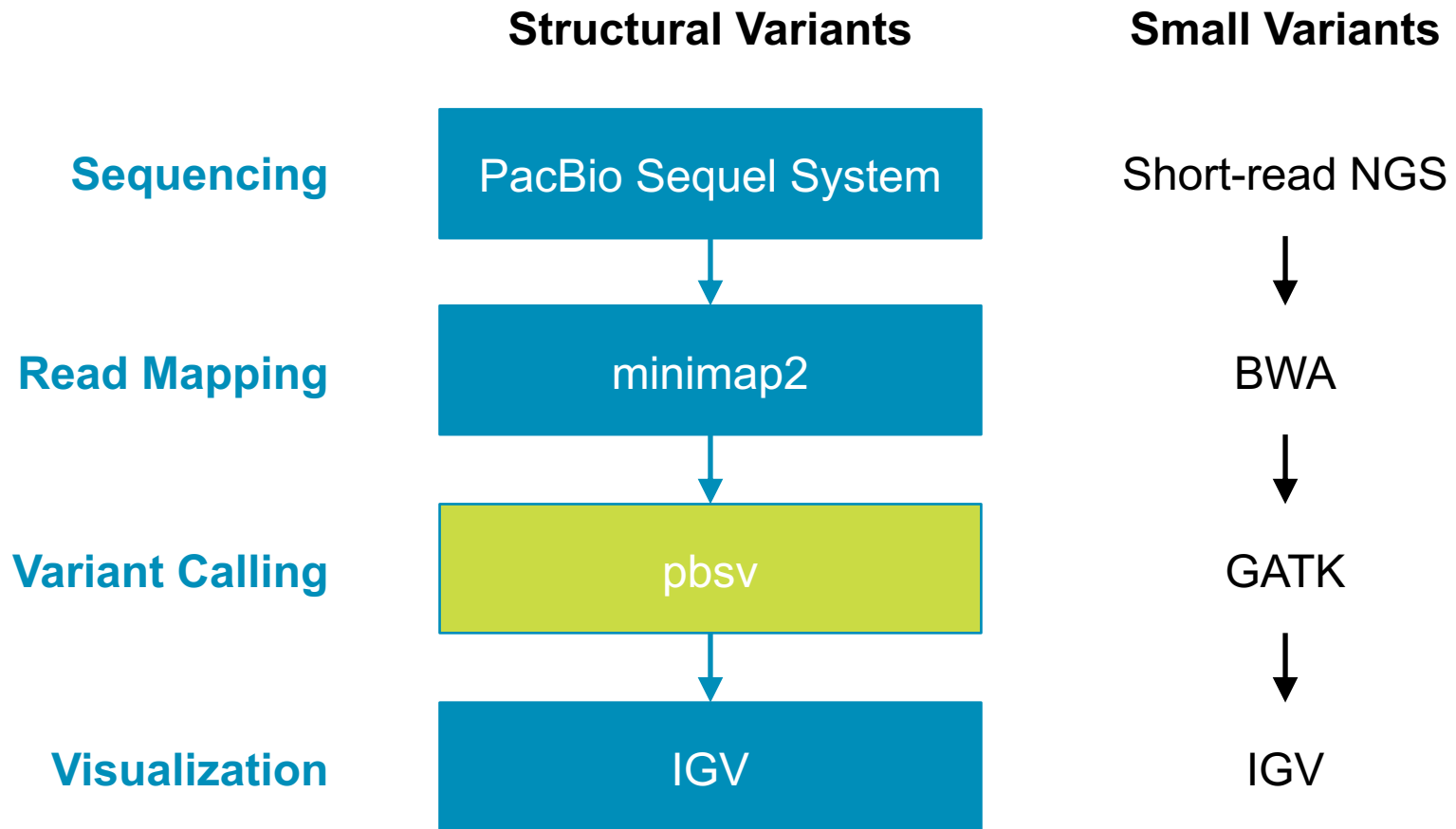
WORKFLOW TO DETECT VARIANTS (SMRT LINK 6.0)



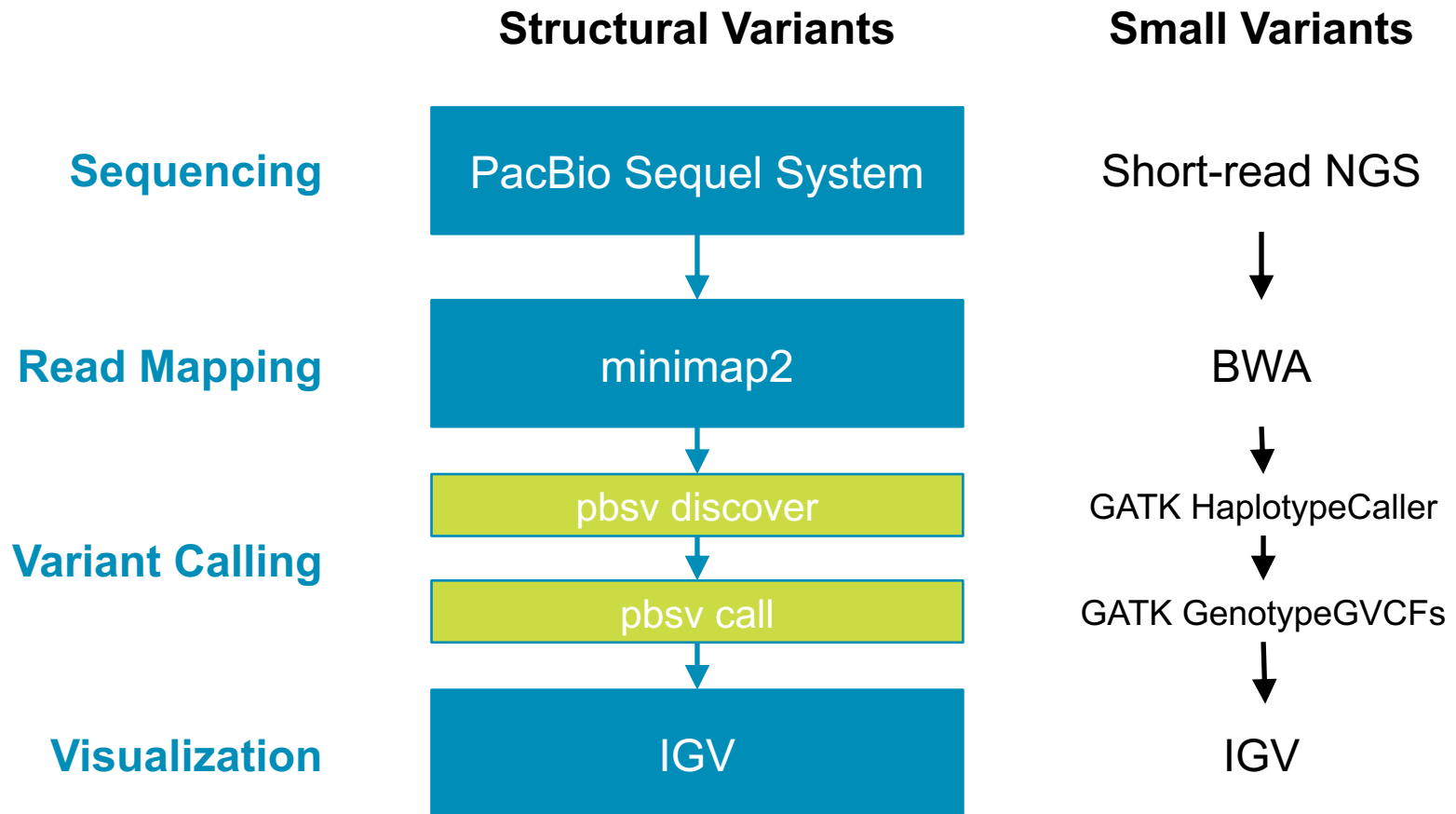
NGMLR VS MINIMAP2 RUN TIME PER SMRT CELL



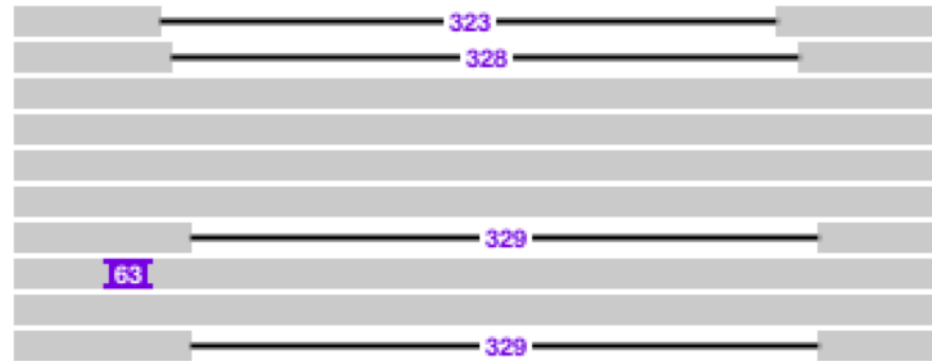
WORKFLOW TO DETECT VARIANTS (SMRT LINK 6.0)



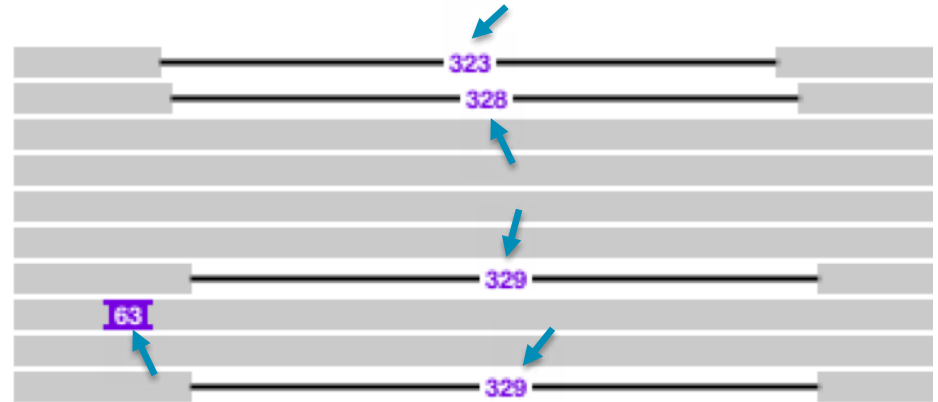
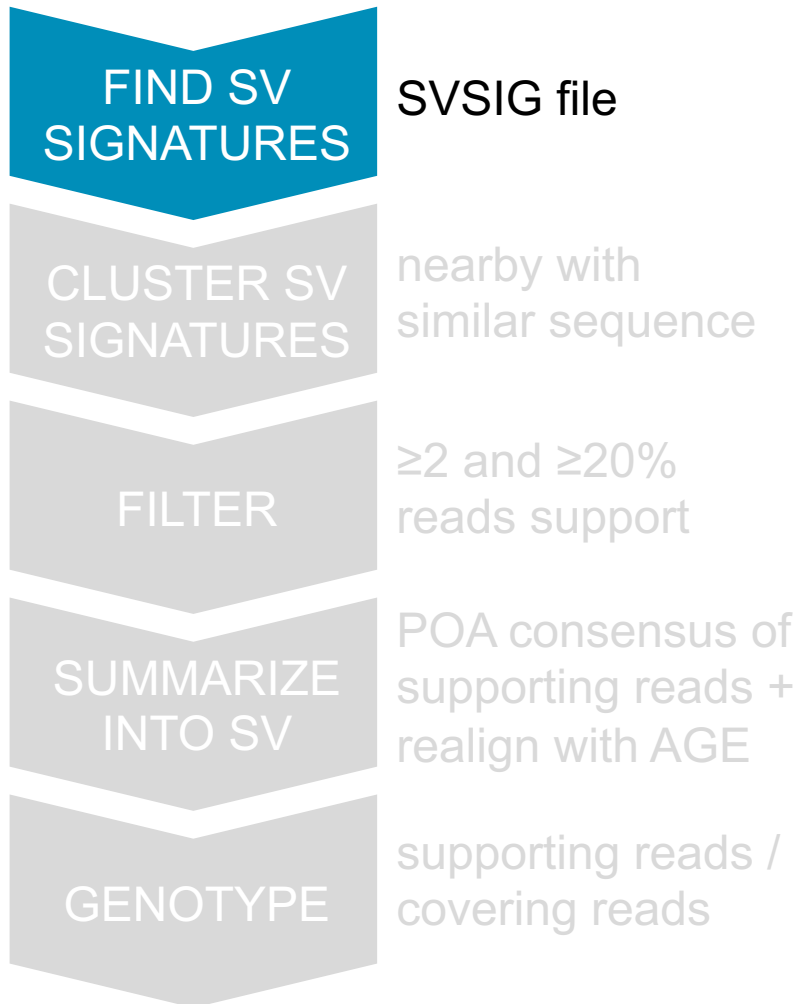
WORKFLOW TO DETECT VARIANTS (SMRT LINK 6.0)



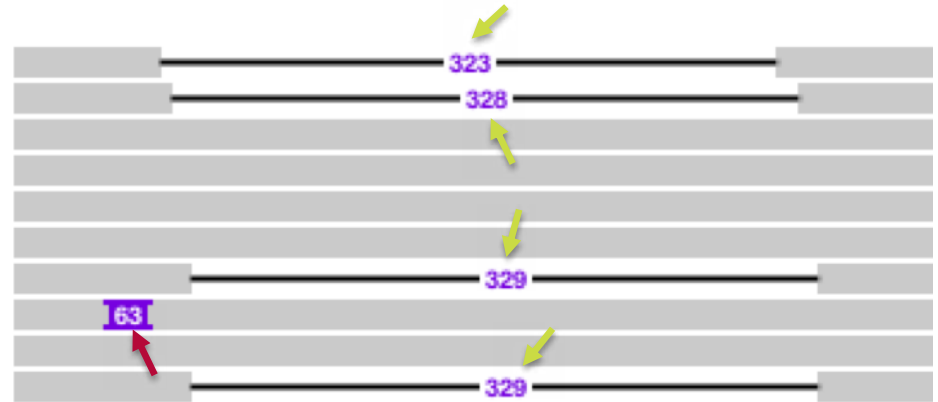
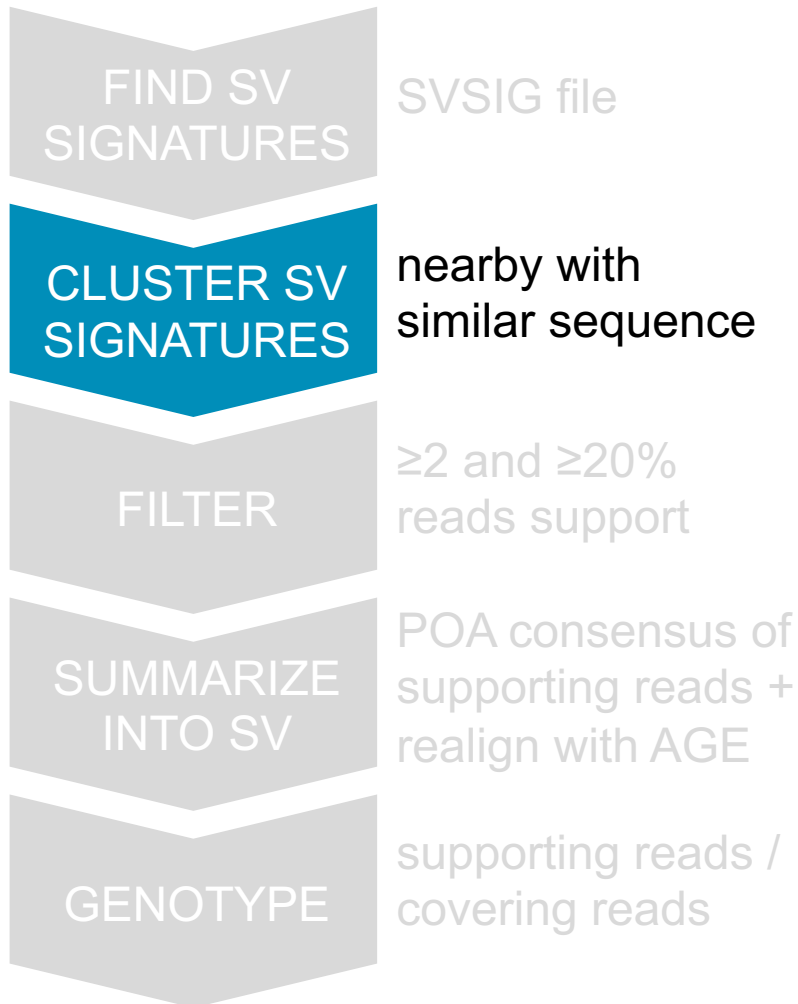
VARIANT CALLING



VARIANT CALLING

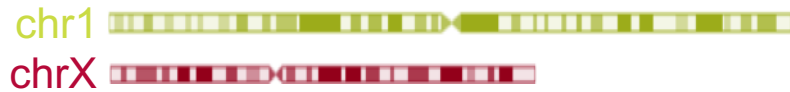


VARIANT CALLING

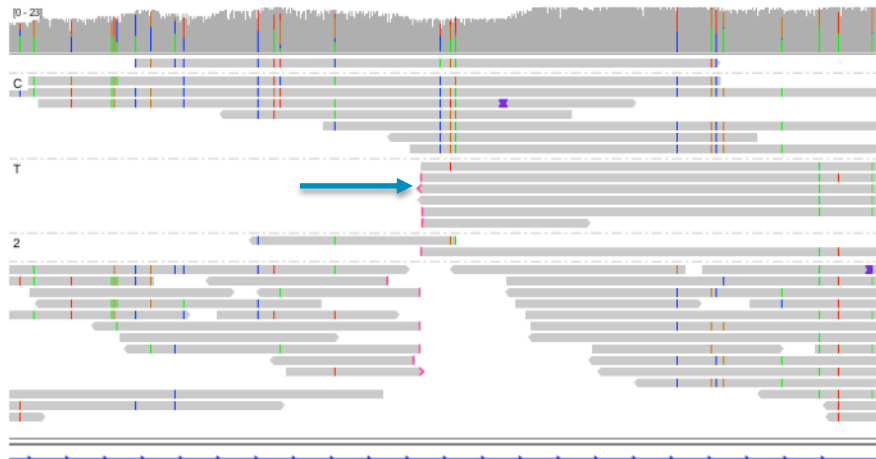


CLUSTERING TRANSLOCATIONS AND INVERSIONS

Human reference (hg19)



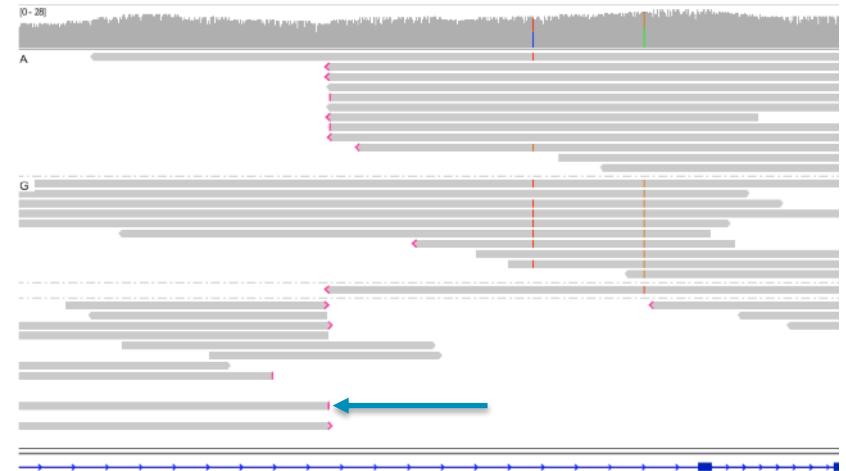
chr1:7,120,756-7,141,619 (20 kb)



Translocation



chrX:49,822,371-49,834,420 (12 kb)

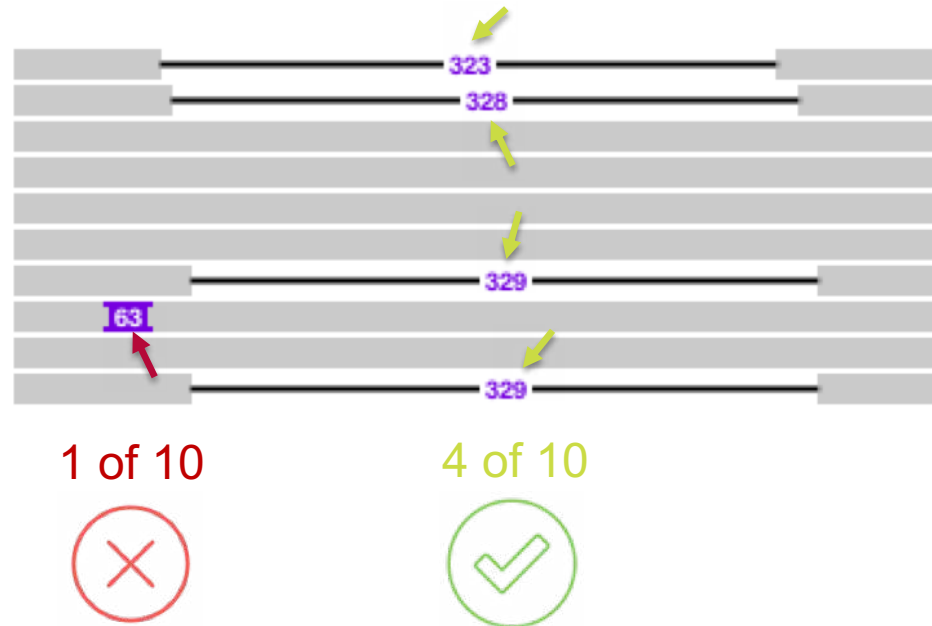


b m54086_180610_123120/69992912/0_35423 chr1 7130523 r chrX 49826914 l

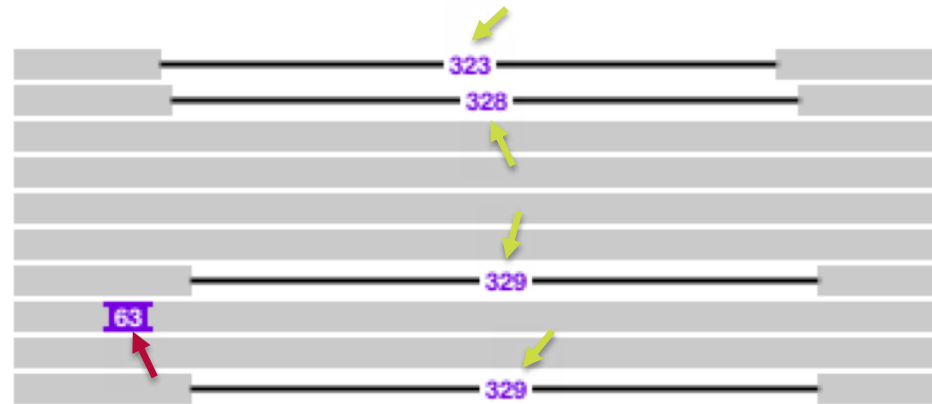
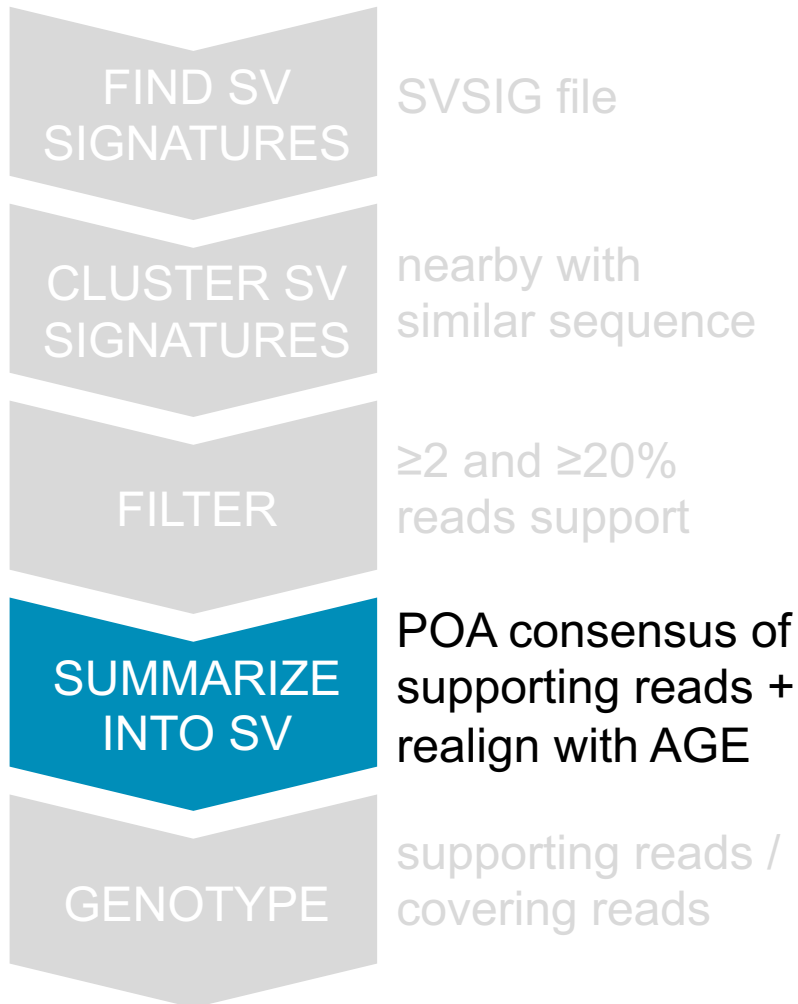
from right from left

Cluster if breakpoints are within 1,000 bp and orientation is the same.

VARIANT CALLING



VARIANT CALLING



1 of 10

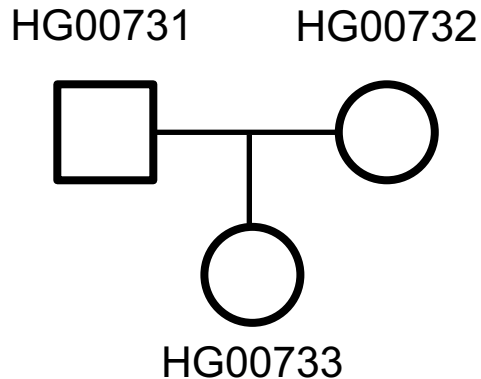


4 of 10



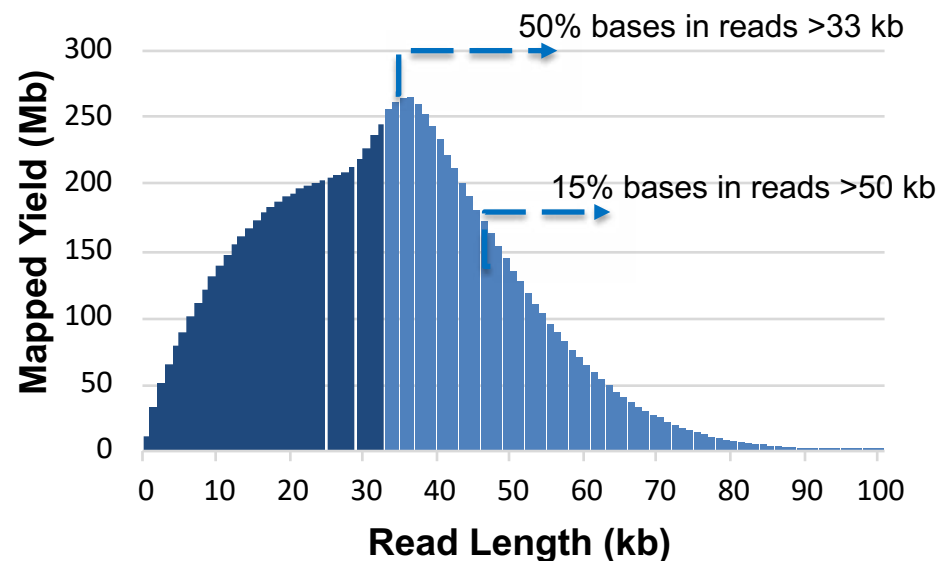
329 bp deletion

HG00733 – PUERTO RICAN CHILD



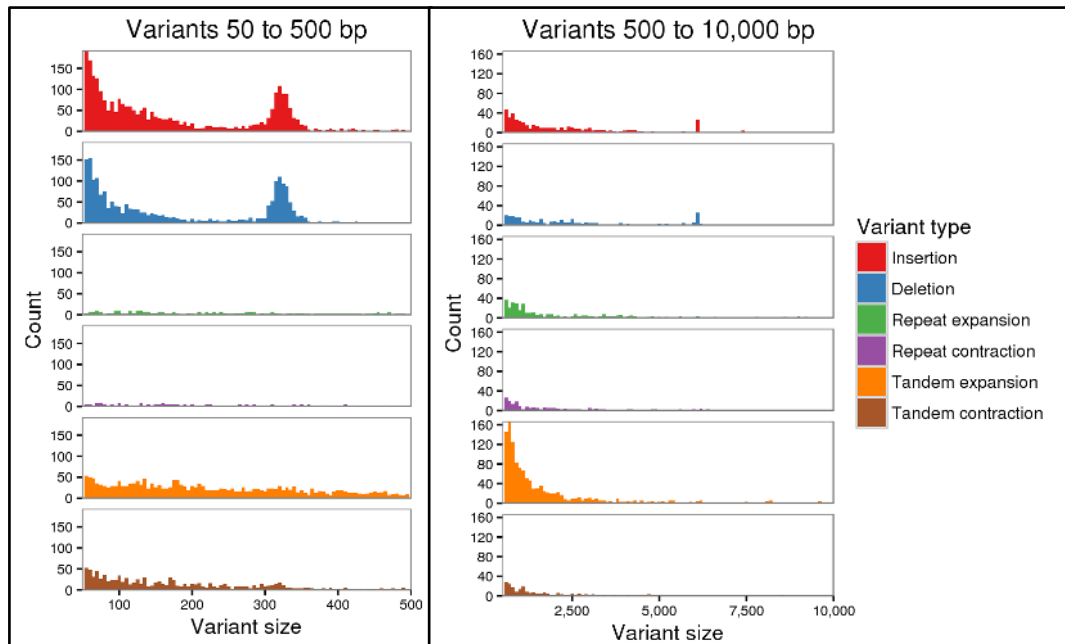
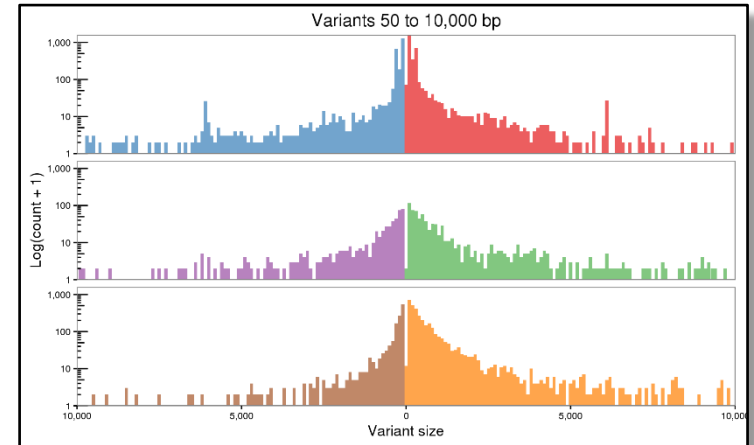
- ❖ Trio from 1000 Genomes Project and HGSC
- ❖ SMRTbell Express Template Prep Kit
- ❖ Sequel System 2.1 chemistry and 5.1 software
- ❖ 28 Sequel SMRT Cells 1M

- ❖ 263 Gb raw yield (82-fold human)
= 9.3 Gb / SMRT Cell
- ❖ 21 kb average read length



HG00733 – FALCON ASM REVEALS STRUCTURAL VARIATION

Reference	hg38
Sequences	415
Total Length	3.21 Gbp
Mean	7.73 Mbp
Max	248.96 Mbp
N50	145.14 Mbp



Query	HG00733
Sequences	947
Total Length	2.87 Gbp
Mean	3.03 Mbp
Max	86.08 Mbp
N50	31.43 Mbp

HG00733 – PBSV



28 servers each 16 cores (448c)

PBSV

Stage	CPU	Wall
Map to reference	9d	0h45m
Discover SV signatures	5h	0h5m
Joint call and polish variants	14h	0h23m
sum	9d19h	1h16m

16 servers each 64 cores (1024c)

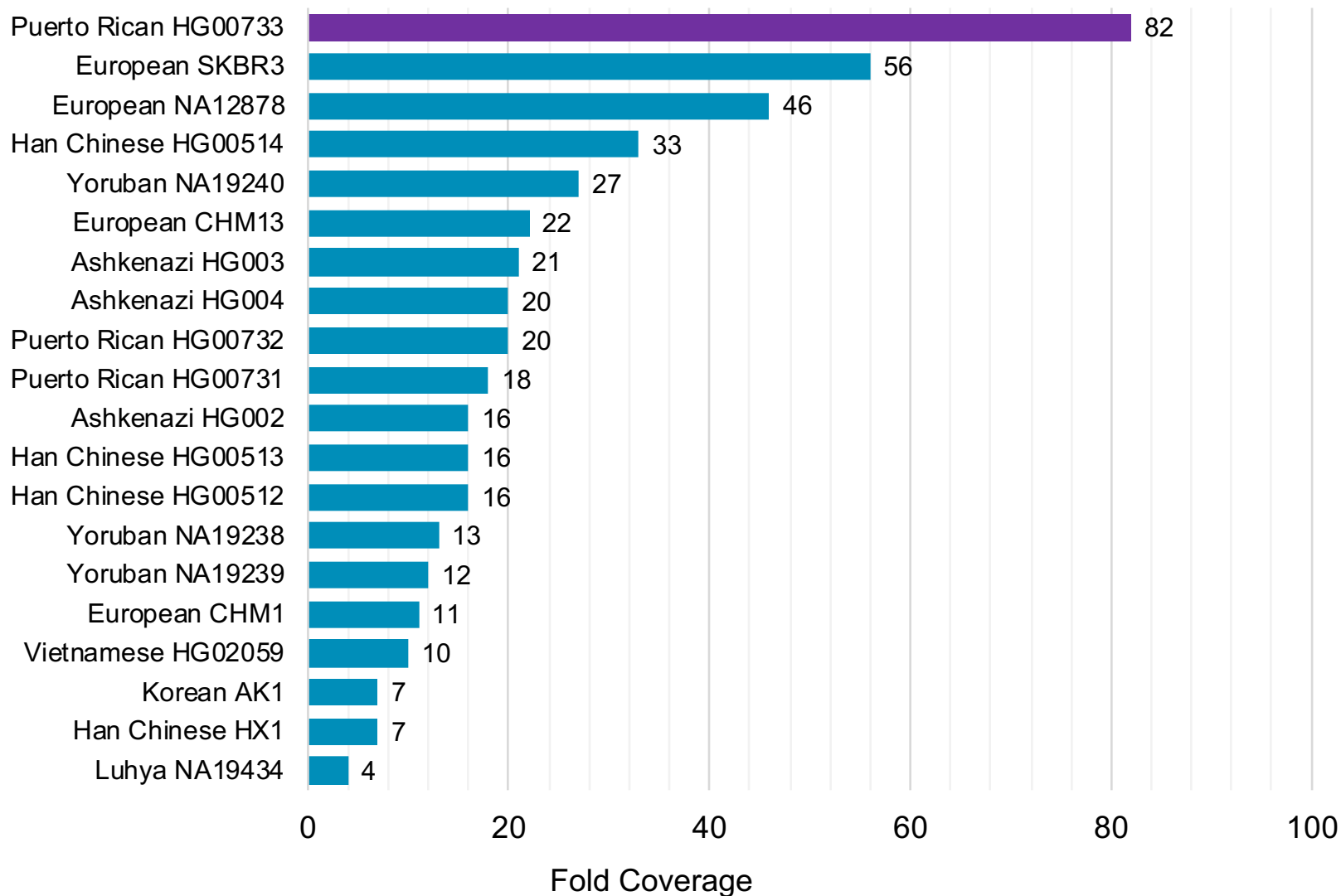
Assembly

Stage	CPU	Wall
Raw read overlap	862d	
Pread consensus	312d	
Pread overlap	845d	
sum	5y194d	2d12h

Low-fold calling

Fold	CPU	Wall
10-fold	6h	8m
5-fold	3h	6m

PBSV – HUMAN COHORT



PBSV – HUMAN COHORT



25 servers each 16 cores (400c)

Stage	CPU	Wall
Discover SV signatures	20h	0h17m
Joint call and polish variants	82h	0h51m
sum	4d6h	1h08m

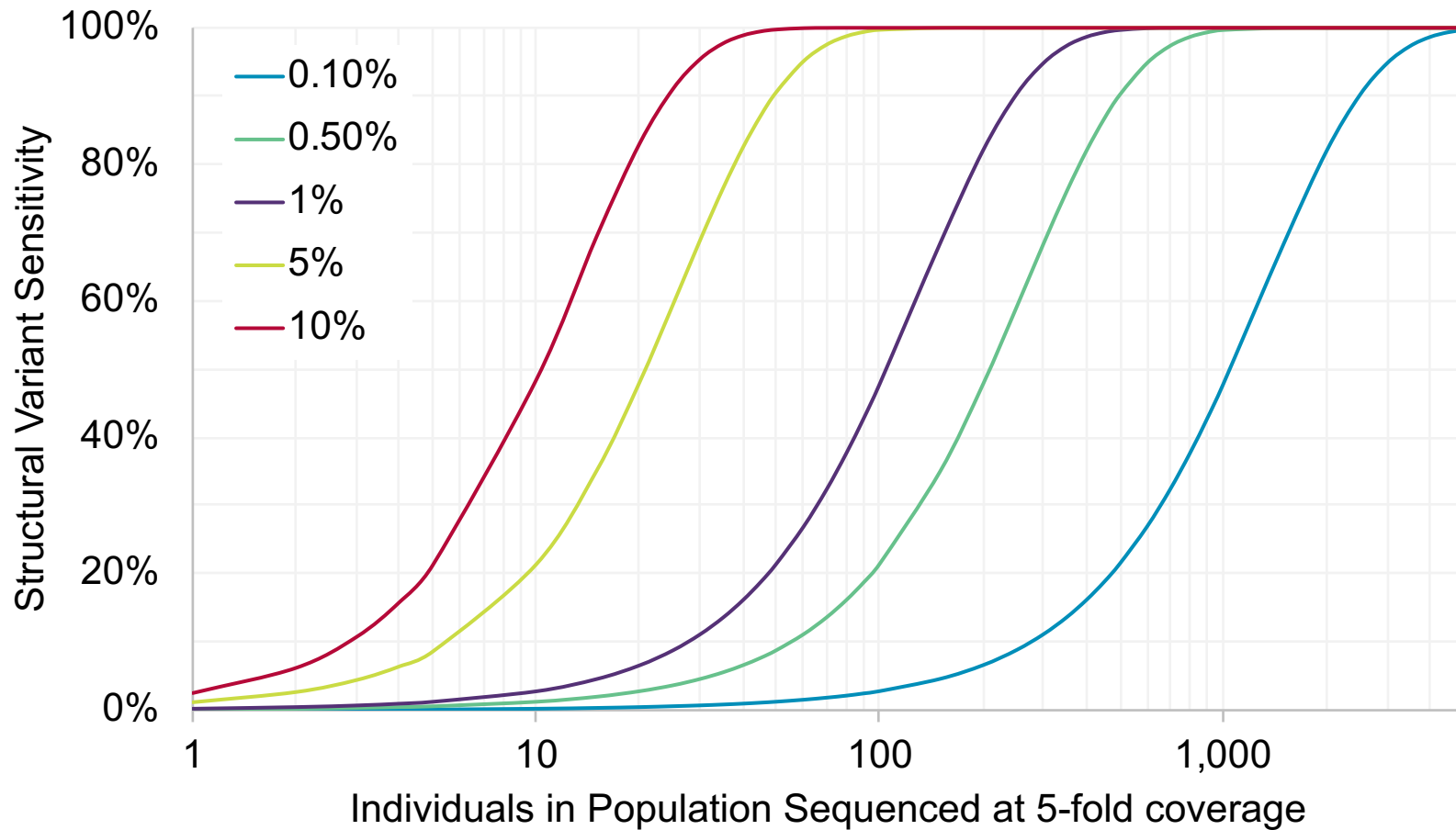


<https://www.tech-coffee.net/understand-failover-cluster-quorum/>



https://upload.wikimedia.org/wikipedia/commons/f/f1/lbm_pc_5150.jpg

5-FOLD COVERAGE FOR COMMON VARIANT DISCOVERY



Calculator: pacb.com/sv

NEW FEATURES IN SMRT LINK 6.0 + 3.0 CHEMISTRY

1. Cost reduction

10-fold human genome in 3 SMRT Cells instead of 4-5

2. Extension of PacBio's advantages over other techniques

- i. Calling of additional variant types: Translocations, inversions, indels 20-49 bp
- ii. Basepair-precise instead of approximate variant calls, consensus insert sequences

3. Support of large, population-scale studies

- i. Faster runtime
- ii. More scalable workflow beyond trio calling, effective for human cohorts
- iii. High sensitivity for shared variants even in low coverage samples
- iv. Bioconda support