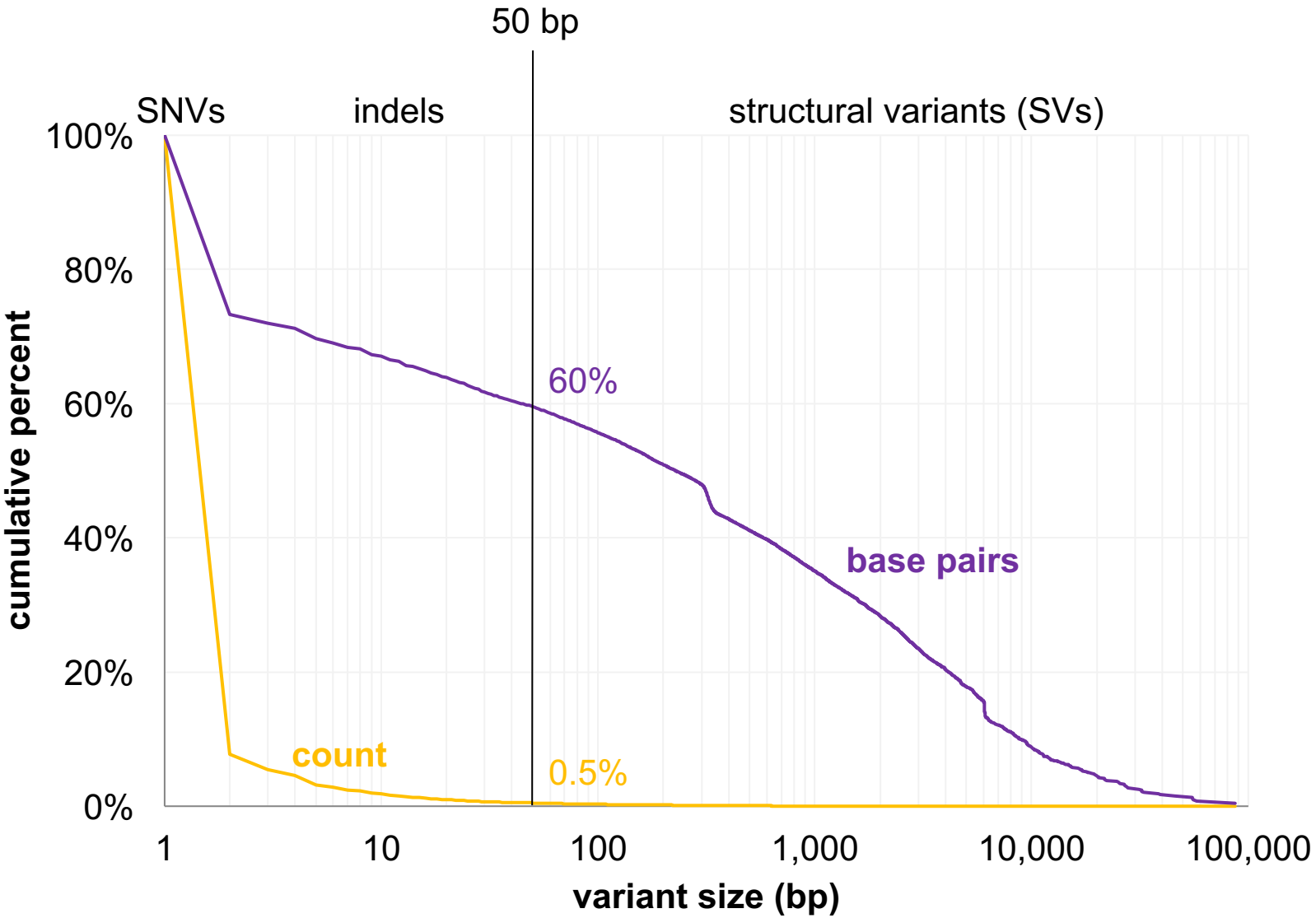
Identifying Structural Variants in Individuals and Populations with PacBio Long Reads

Aaron Wenger

2017-01-17

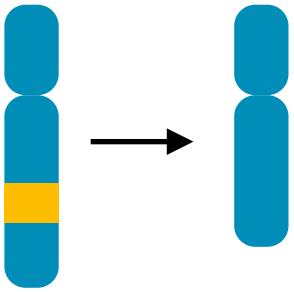
VARIATION IN A HUMAN GENOME – HG00733



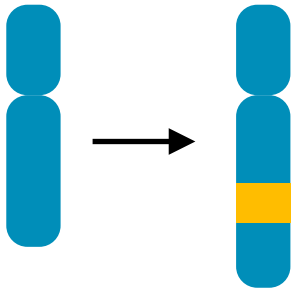
Chaisson et al. (2017) *bioRxiv*. doi:10.1101/193144.

TYPES OF STRUCTURAL VARIATION

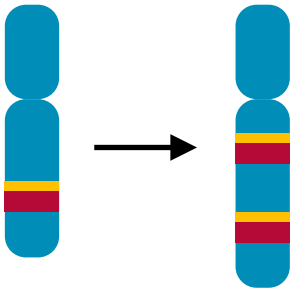
deletion



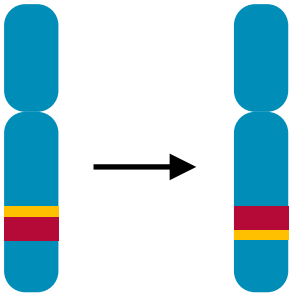
insertion



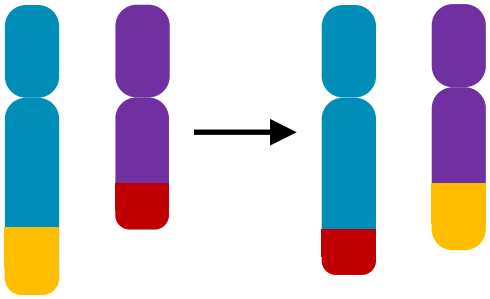
duplication



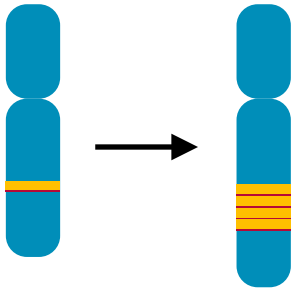
inversion



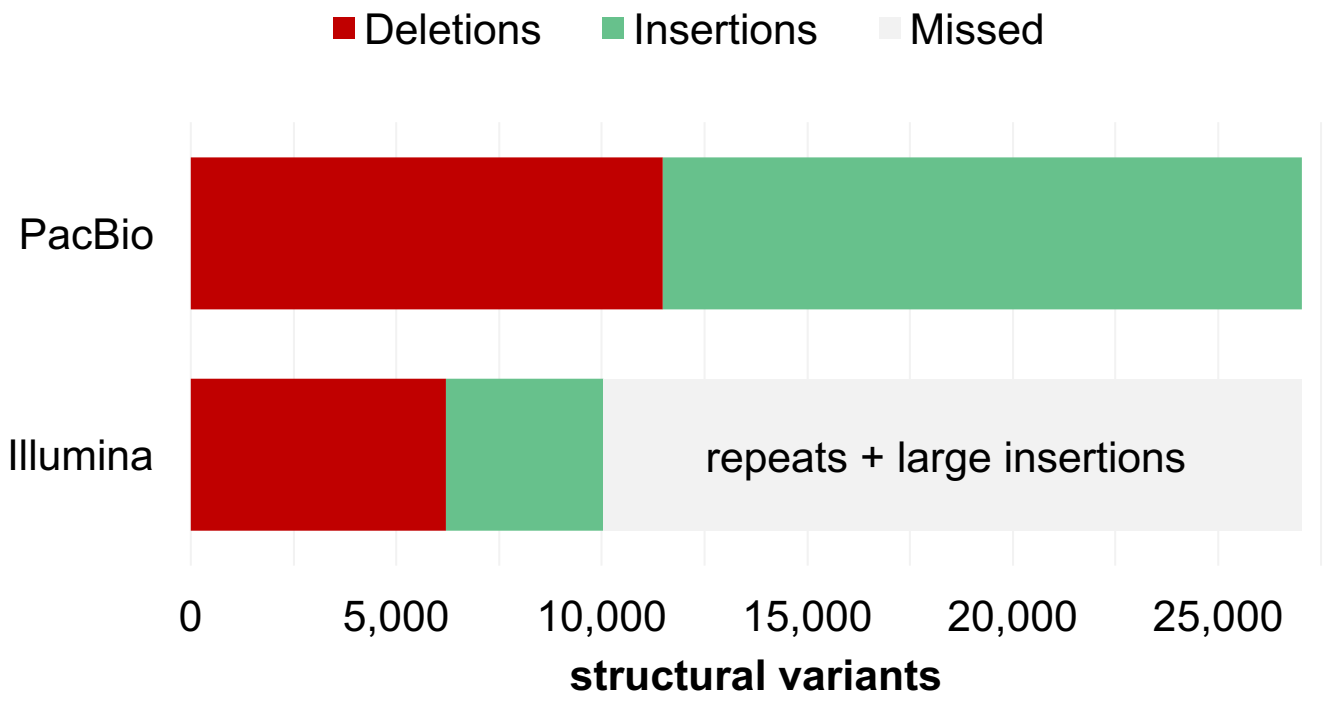
translocation



repeat expansion

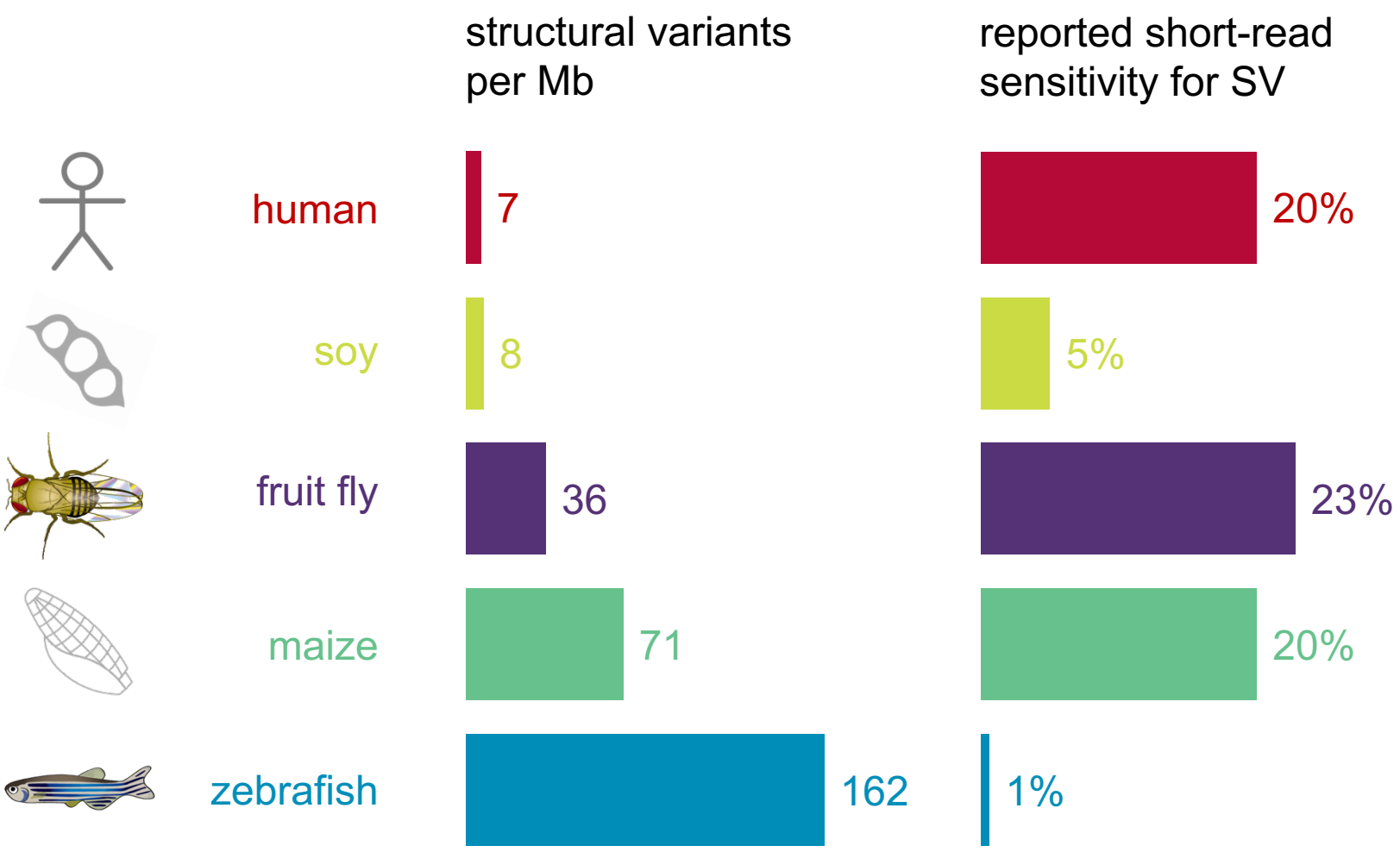


TECHNOLOGY TO DETECT STRUCTURAL VARIANTS



Chaisson et al. (2017) *bioRxiv*. doi:10.1101/193144.

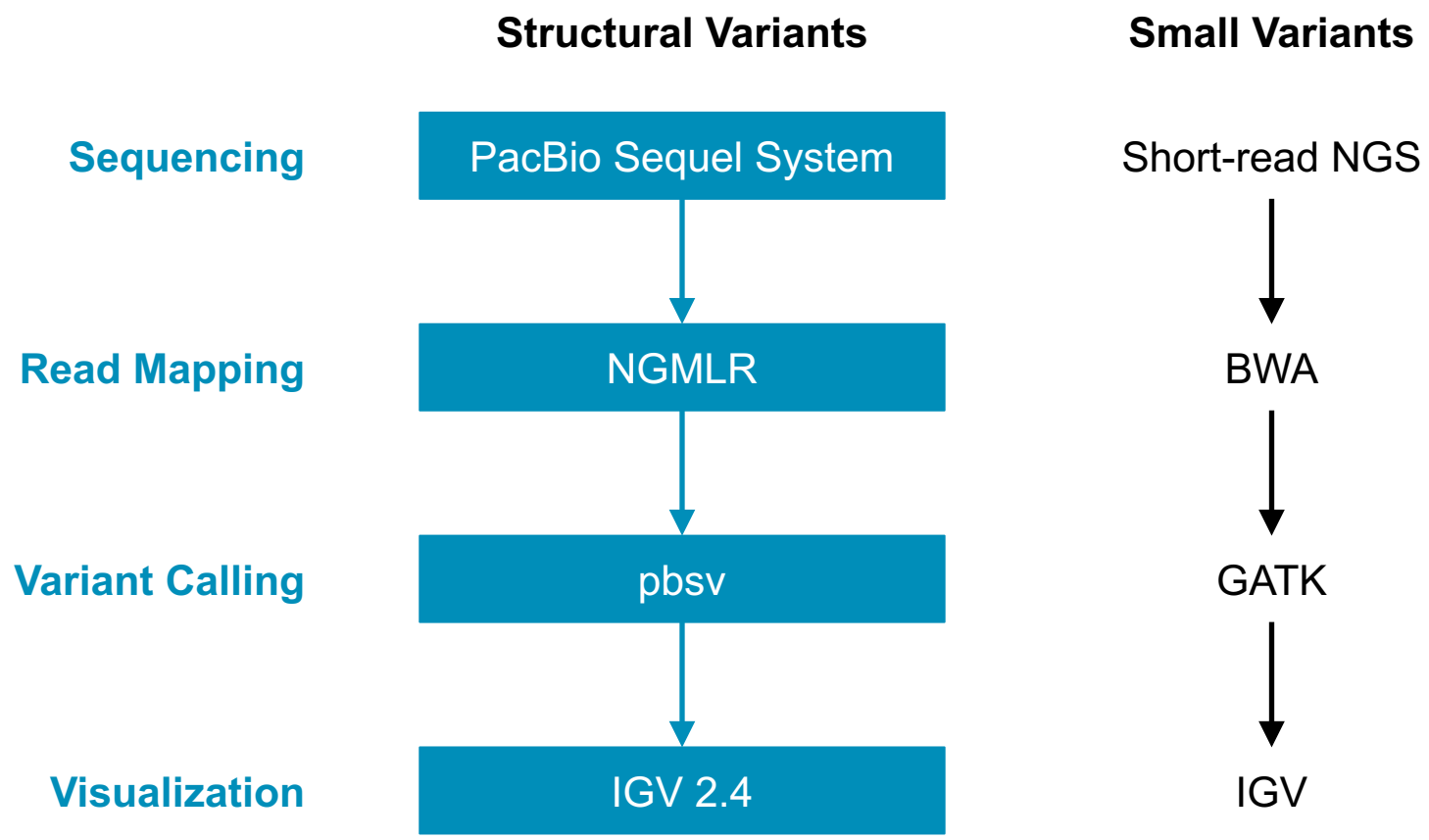
THE HUMAN GENOME IS COMPARATIVELY EASY



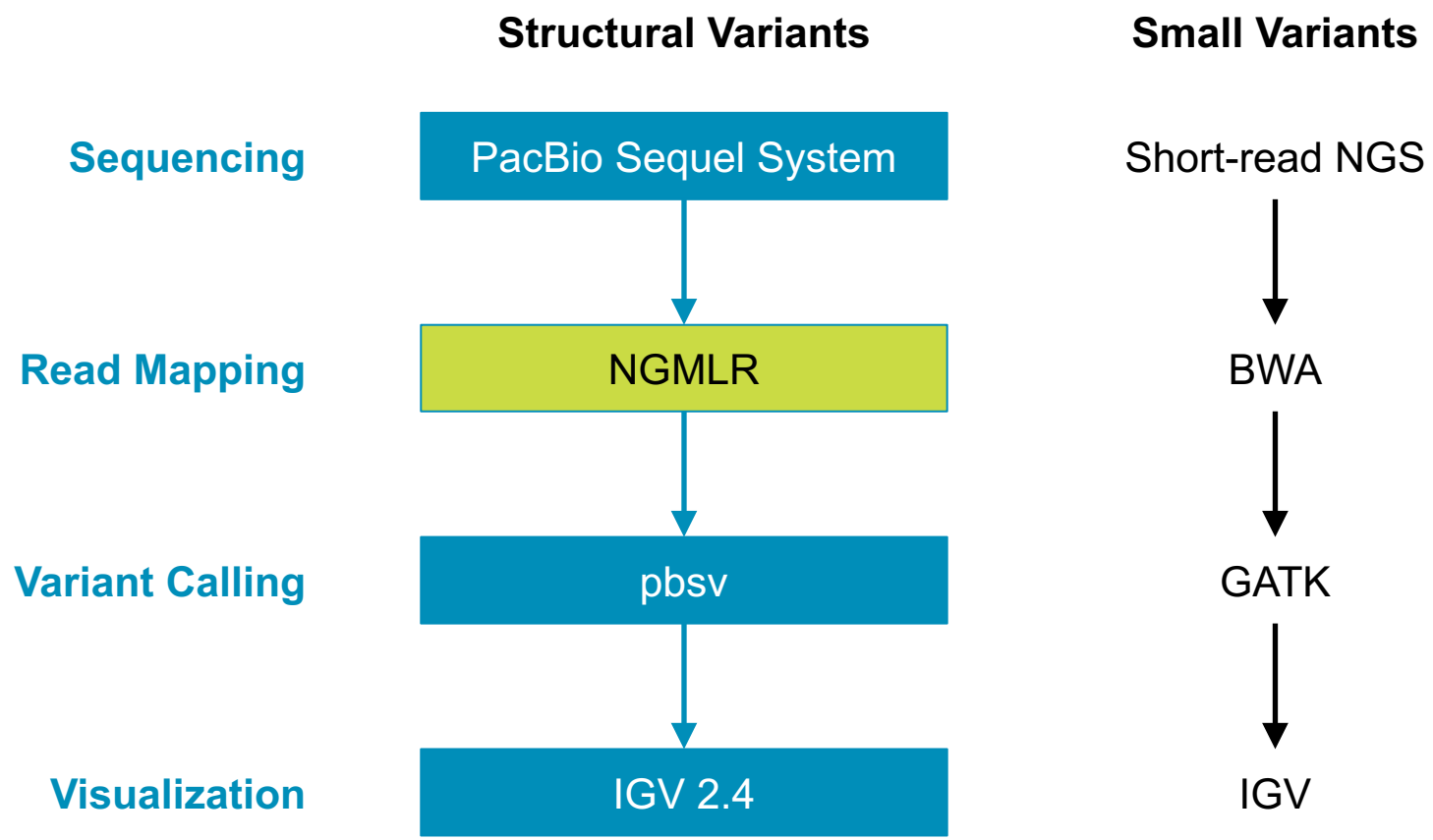
Huddleston et al. (2017) *Genome Research* 27(5):677-85.
Zichner et al. (2013) *Genome Research* 23(3):568-79.
Patowary et al. (2013) *Zebrafish* 10(1):15-20.

Concepcion et al. (2018) *PAG.*
Zebrafish image courtesy of Lizzy Griffiths

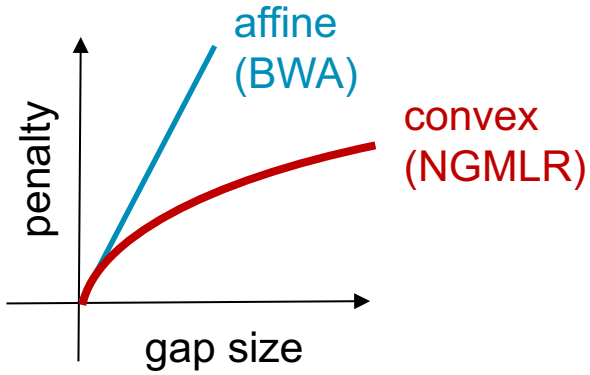
WGS FOR STRUCTURAL VARIANT DISCOVERY



WGS FOR STRUCTURAL VARIANT DISCOVERY

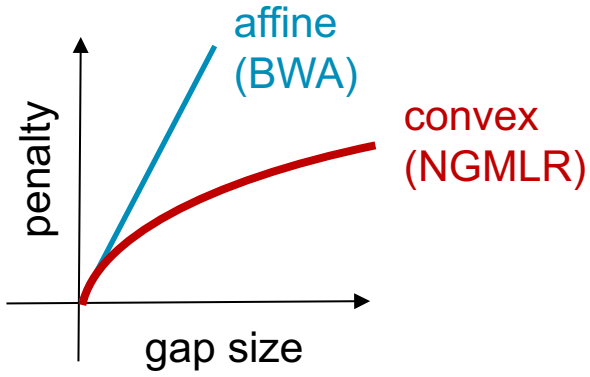


READ MAPPING

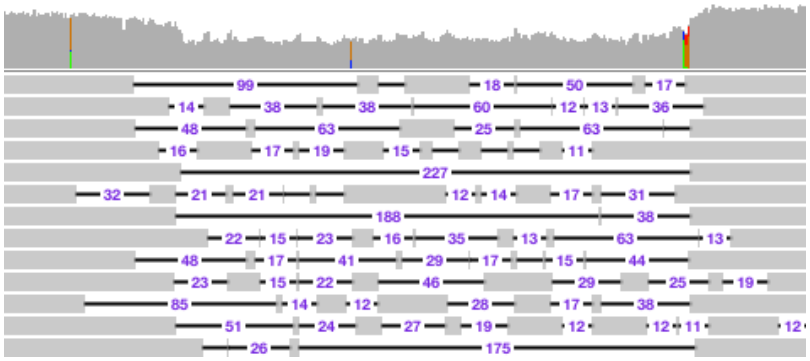


Sedlazeck et al. (2017) *bioRxiv*. doi:10.1101/169557.

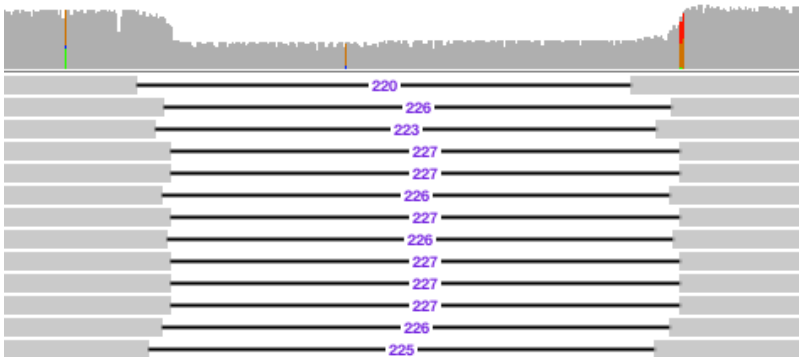
READ MAPPING



BWA

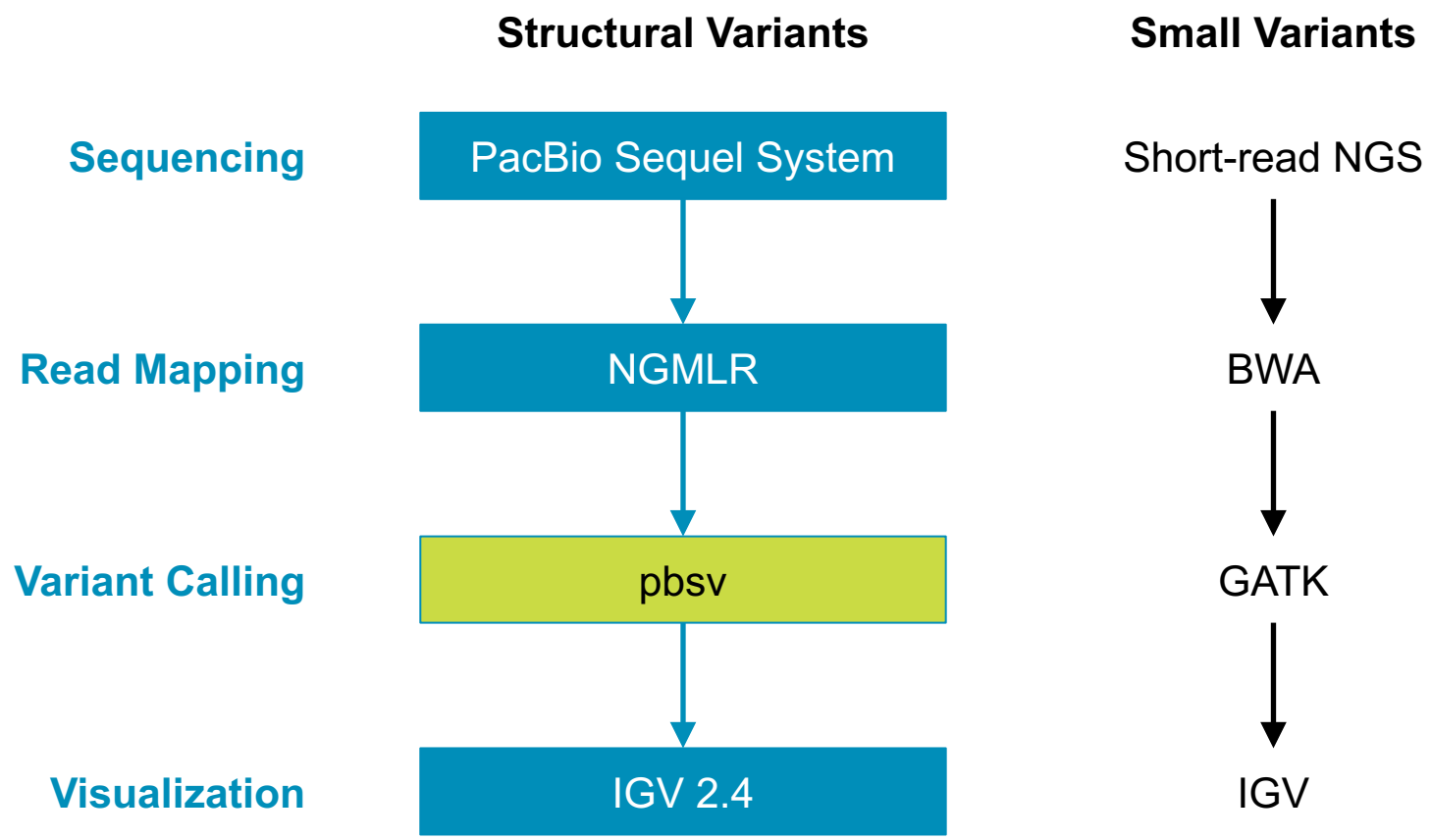


NGMLR



Sedlazeck et al. (2017) *bioRxiv*. doi:10.1101/169557.

WGS FOR STRUCTURAL VARIANT DISCOVERY



VARIANT CALLING

FIND SV SIGNATURES

CIGAR D & I ≥50 bp

CLUSTER SV SIGNATURES

nearby with similar sequence

FILTER

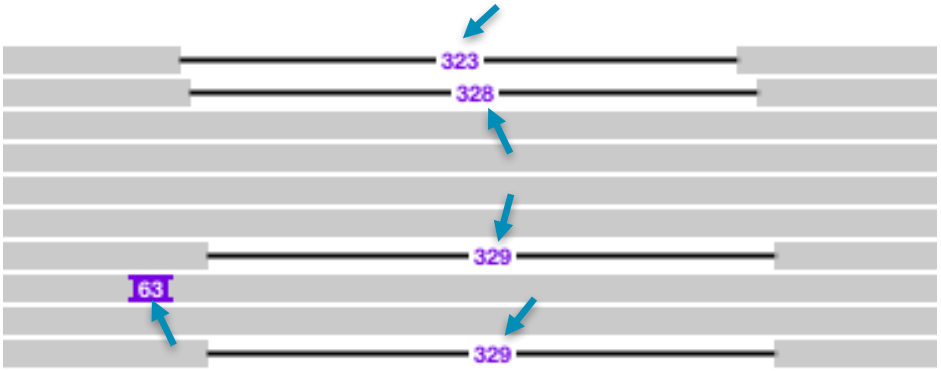
≥2 and ≥20% reads support

SUMMARIZE INTO SV

consensus of supporting reads

GENOTYPE

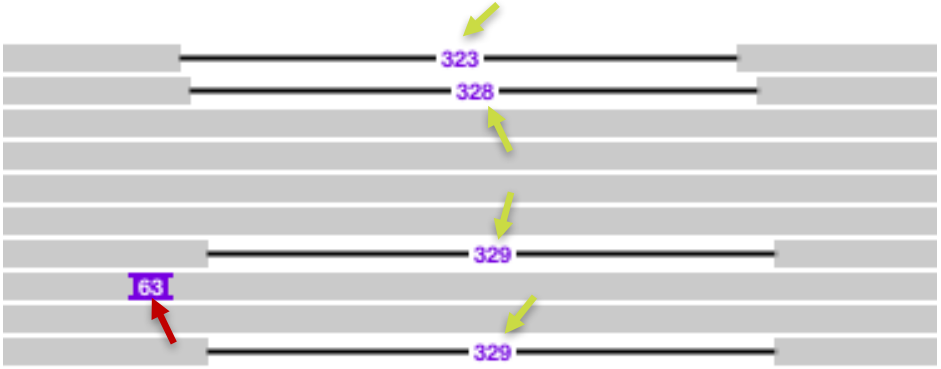
supporting reads / covering reads



http://pacb.com/sv

VARIANT CALLING

FIND SV SIGNATURES	CIGAR D & I ≥50 bp
CLUSTER SV SIGNATURES	nearby with similar sequence
FILTER	≥2 and ≥20% reads support
SUMMARIZE INTO SV	consensus of supporting reads
GENOTYPE	supporting reads / covering reads



VARIANT CALLING

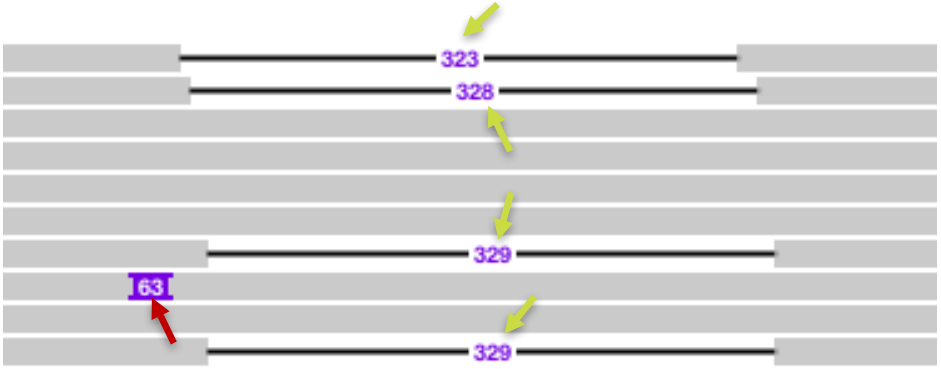
FIND SV SIGNATURES — CIGAR D & I ≥50 bp

CLUSTER SV SIGNATURES — nearby with similar sequence

FILTER — ≥2 and ≥20% reads support

SUMMARIZE INTO SV — consensus of supporting reads

GENOTYPE — supporting reads / covering reads



1 of 10 ✗

4 of 10 ✓

VARIANT CALLING

FIND SV SIGNATURES

CIGAR D & I ≥50 bp

CLUSTER SV SIGNATURES

nearby with similar sequence

FILTER

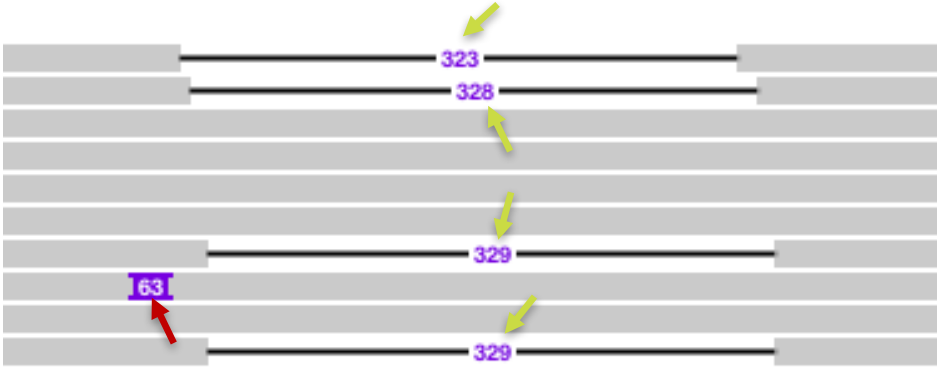
≥2 and ≥20% reads support

SUMMARIZE INTO SV

consensus of supporting reads

GENOTYPE

supporting reads / covering reads



1 of 10

4 of 10

329 bp deletion

http://pacb.com/sv

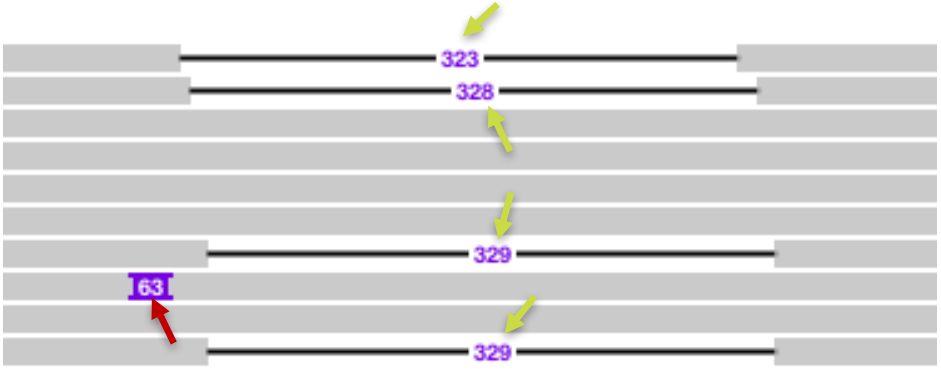
VARIANT CALLING

FIND SV SIGNATURES

CIGAR D & I ≥50 bp

CLUSTER SV SIGNATURES

nearby with similar sequence



FILTER

≥2 and ≥20% reads support

1 of 10

4 of 10

SUMMARIZE INTO SV

consensus of supporting reads

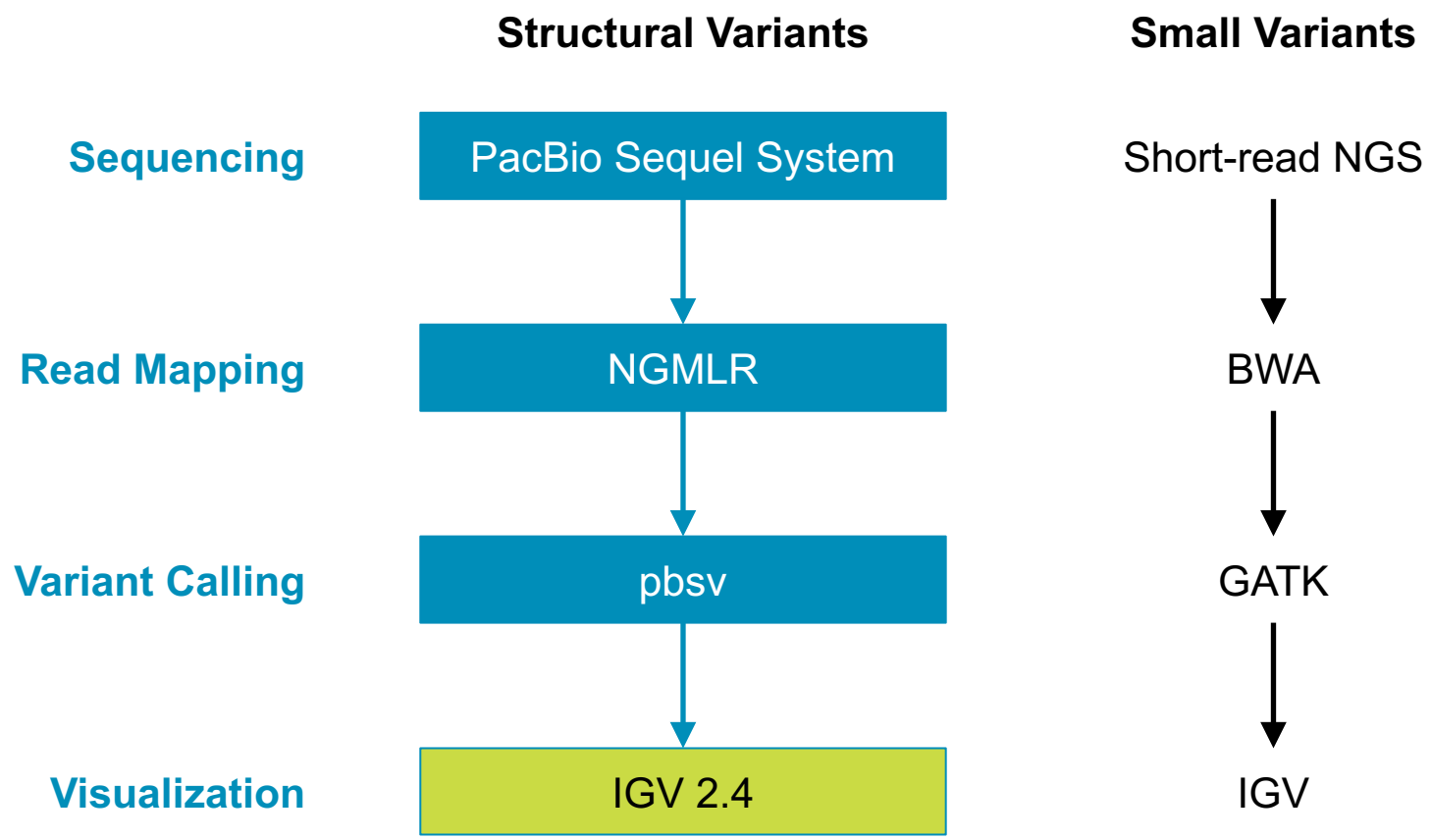
329 bp deletion

GENOTYPE

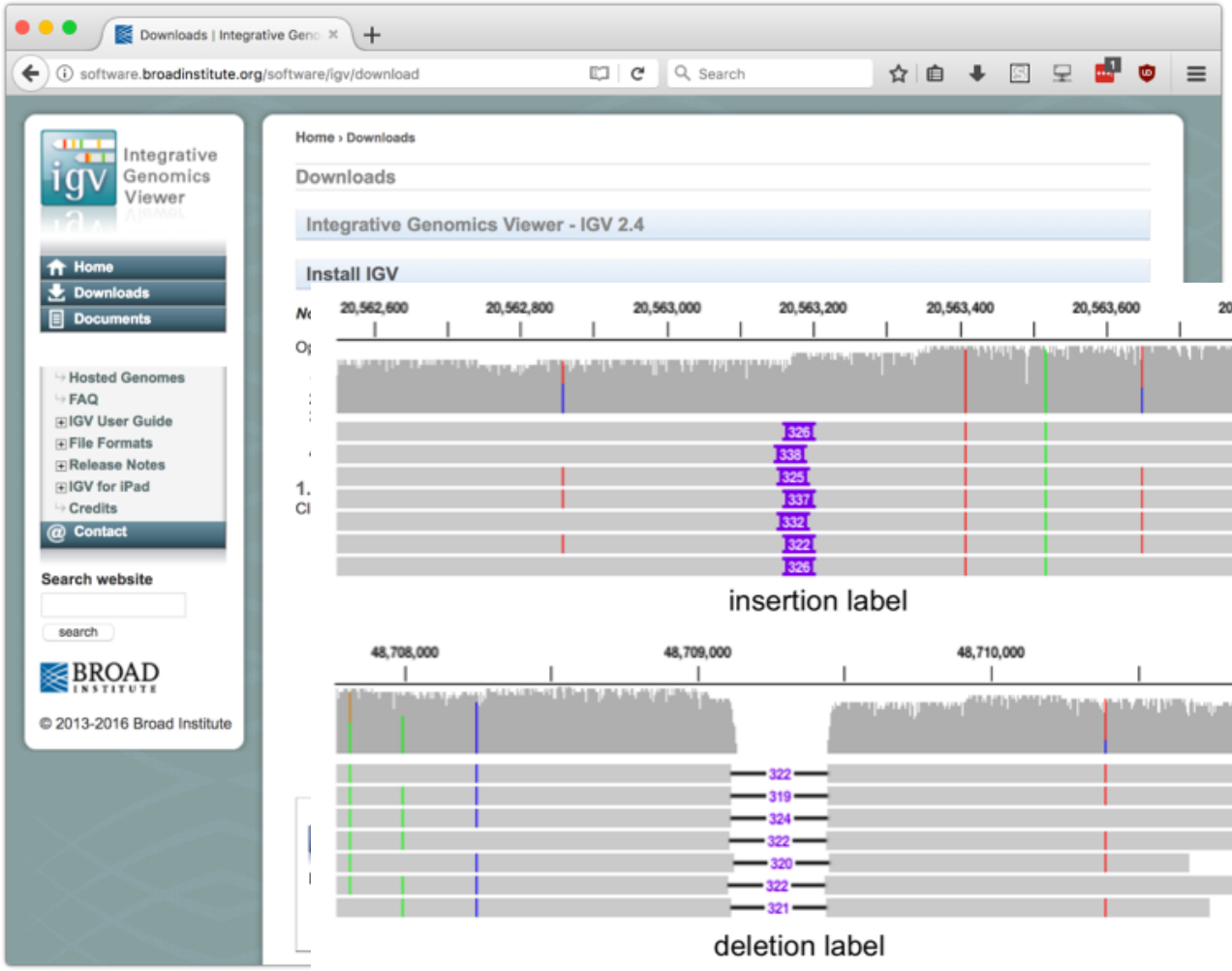
supporting reads / covering reads

heterozygous (4 of 10)

WGS FOR STRUCTURAL VARIANT DISCOVERY



VISUALIZATION – IGV 2.4 (FORMERLY IGV 3 BETA)

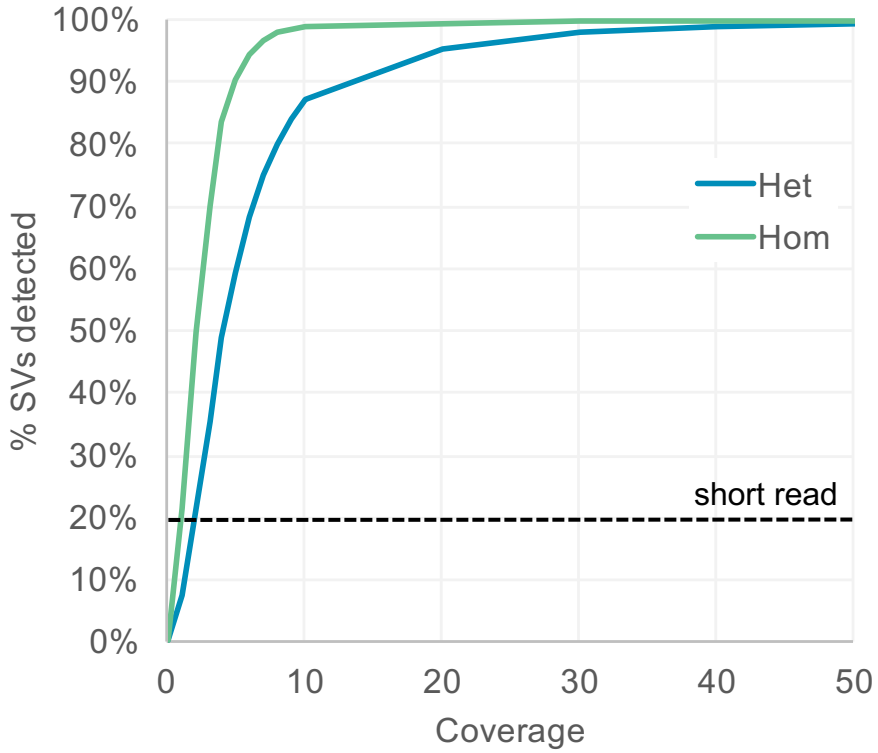


insertion

deletion

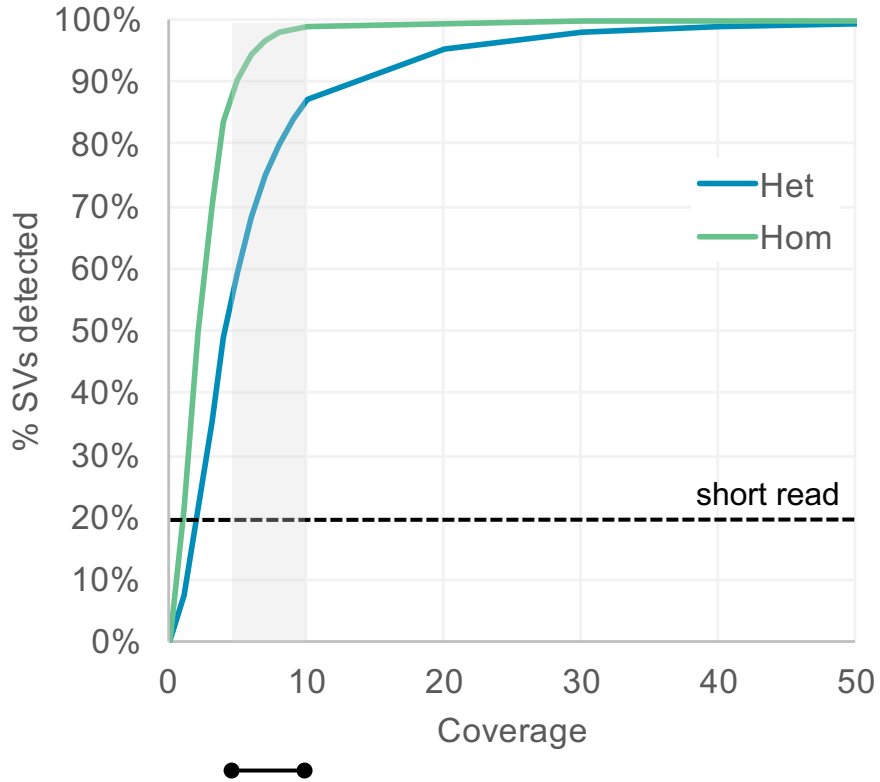
Robinson et al. (2011) *Nature Biotechnology*. doi:10.1038/nbt.1754.

HOW MUCH TO SEQUENCE?



Human HG00733
Sequel System
211 Gb (70-fold)

HOW MUCH TO SEQUENCE?

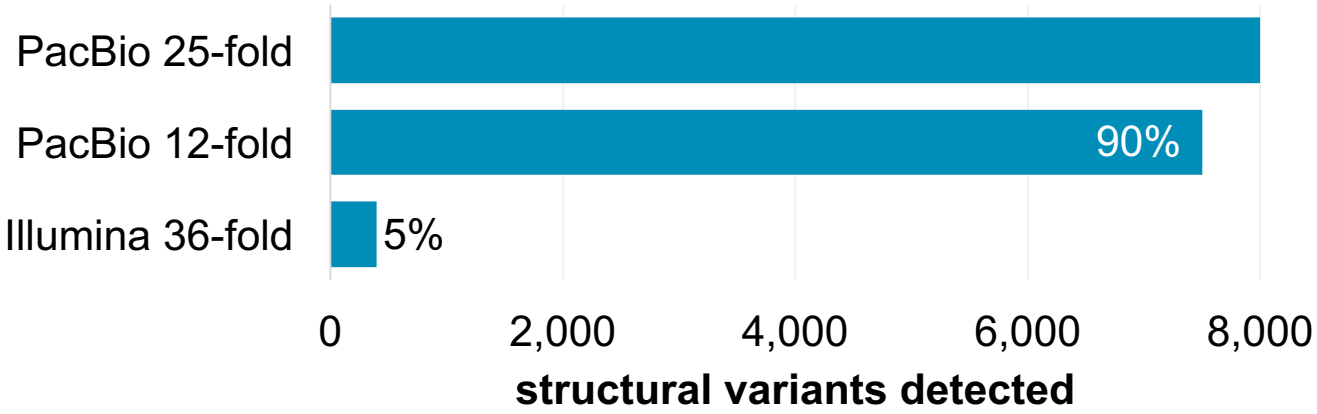


5- to 10-fold
optimal tradeoff of
cost vs. performance

Human HG00733
Sequel System
211 Gb (70-fold)

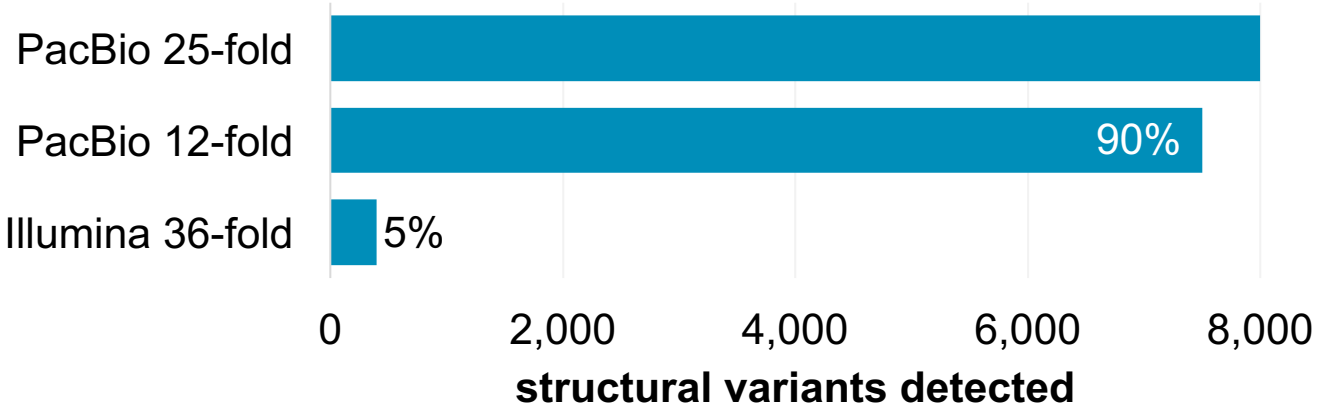
STRUCTURAL VARIANT SENSITIVITY IN PLANT GENOMES

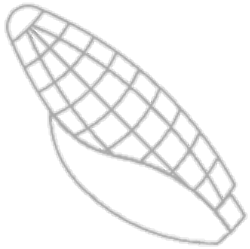
Soy Wm82
against Wm82a2



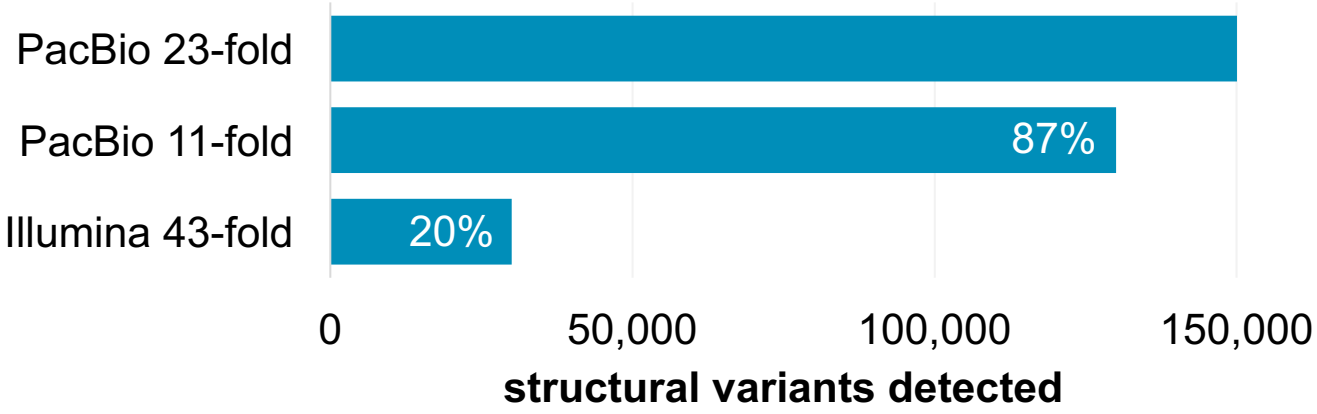
STRUCTURAL VARIANT SENSITIVITY IN PLANT GENOMES

Soy Wm82
against Wm82a2





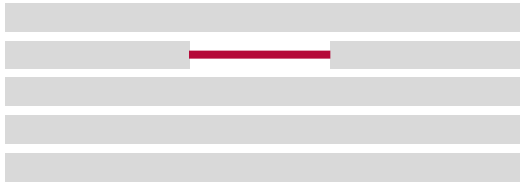
Maize Mo17
against AGPv4



JOINT VARIANT CALLING IN PBSV

Solo Calling

Individual #1

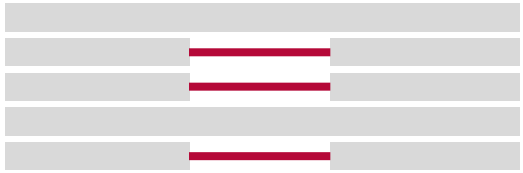


-

Individual #2

-

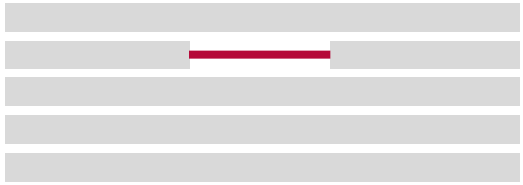
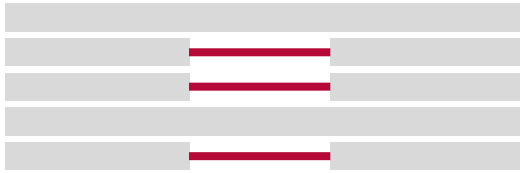
Individual #3



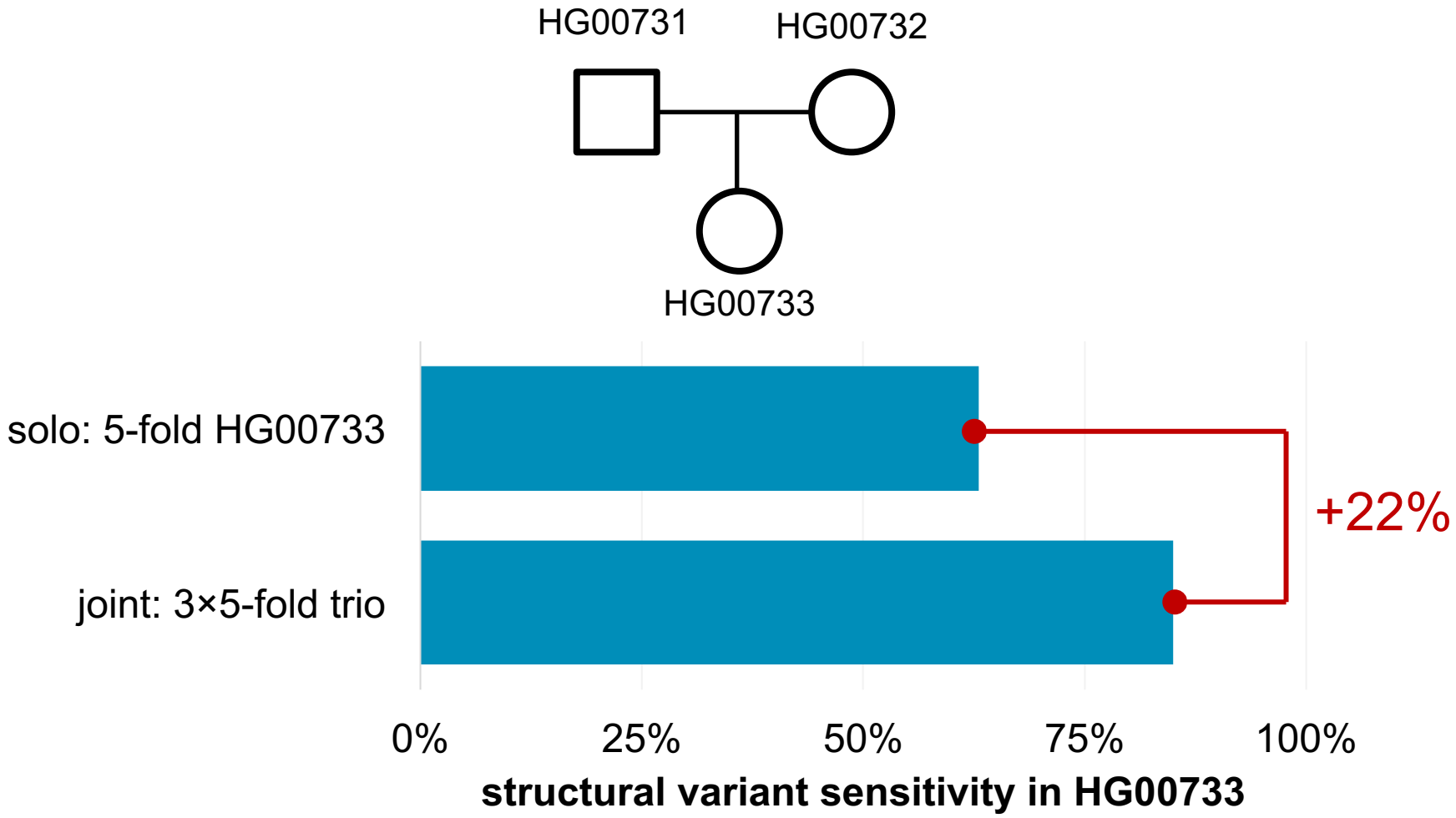
REF/ALT

JOINT VARIANT CALLING IN PBSV

available in next release of SMRT Link

		Solo Calling	Joint Calling
Individual #1		-	REF/ALT
Individual #2		-	REF/REF
Individual #3		REF/ALT	REF/ALT

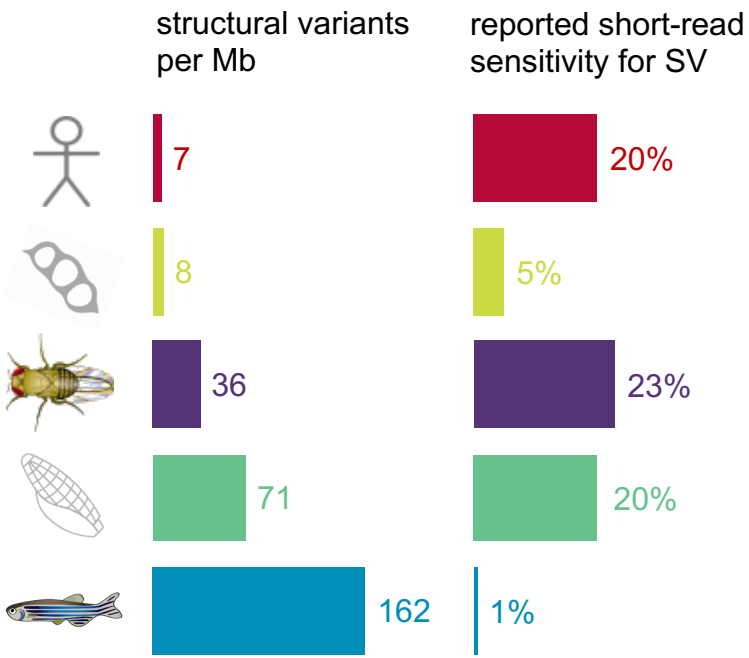
JOINT VARIANT CALLING IN A HUMAN TRIO



SUMMARY

Most base pairs that differ between individuals lie in structural variants.

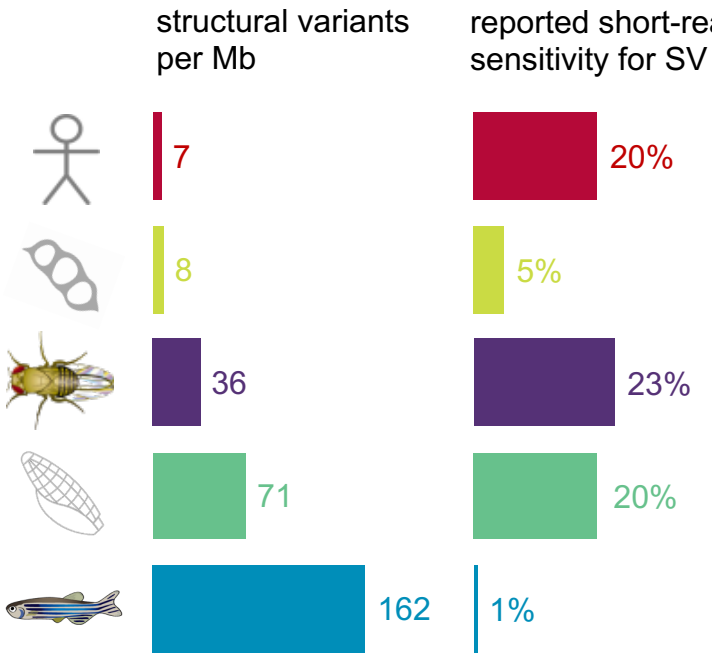
Most structural variants are missed by short reads.



SUMMARY

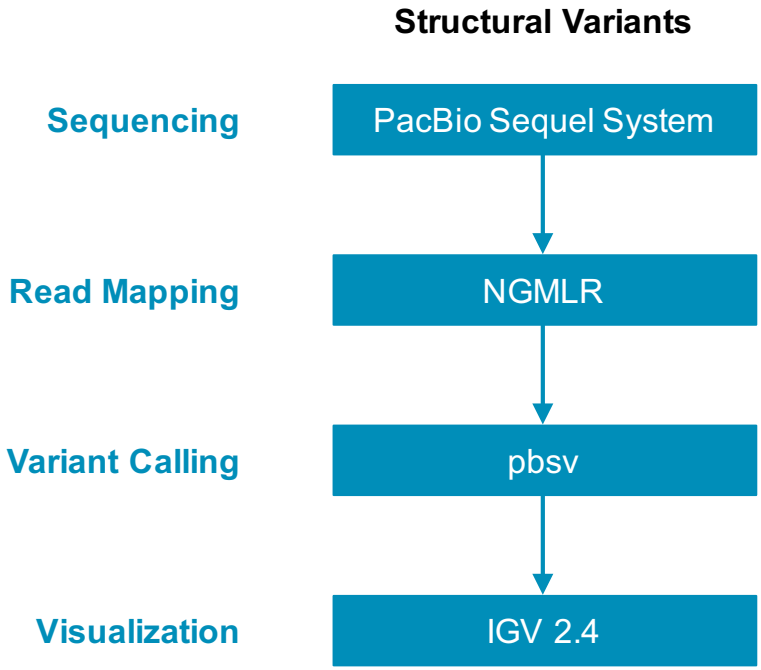
Most base pairs that differ between individuals lie in structural variants.

Most structural variants are missed by short reads.



NGMLR and pbsv effectively detect structural variants in diploids.

Joint calling increases sensitivity for shared variants.



ACKNOWLEDGMENTS

Schatz Lab

Michael Schatz
Philipp Rescheneder
Fritz Sedlazeck

PacBio

Mary Budagyan
Greg Concepcion
James Drake
Sarah Kingan
Aaron Klammer
Yuan Li
Paul Peluso
Joan Wilson
Janet Ziegle

www.pacb.com