



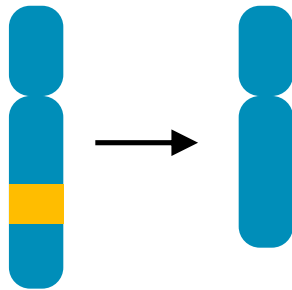
PACIFIC
BIOSCIENCES®



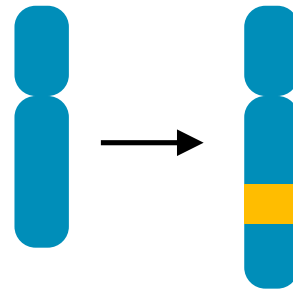
Structural Variant Detection in SMRT Link powered by pbsv

TYPES OF STRUCTURAL VARIATION

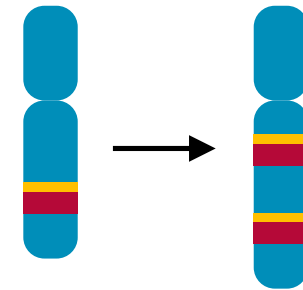
deletion



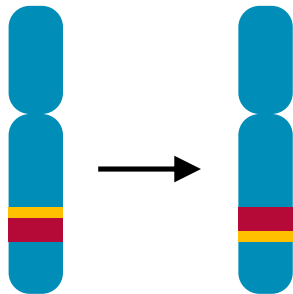
insertion



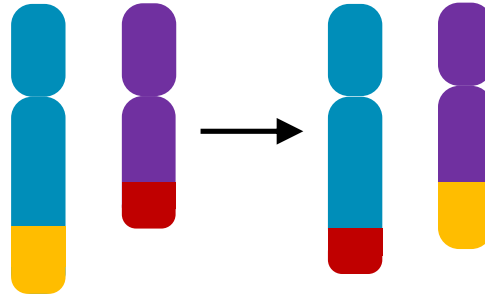
duplication



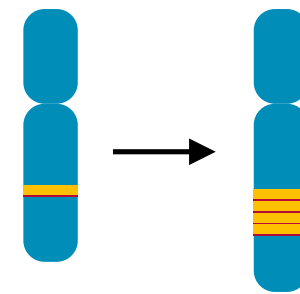
inversion



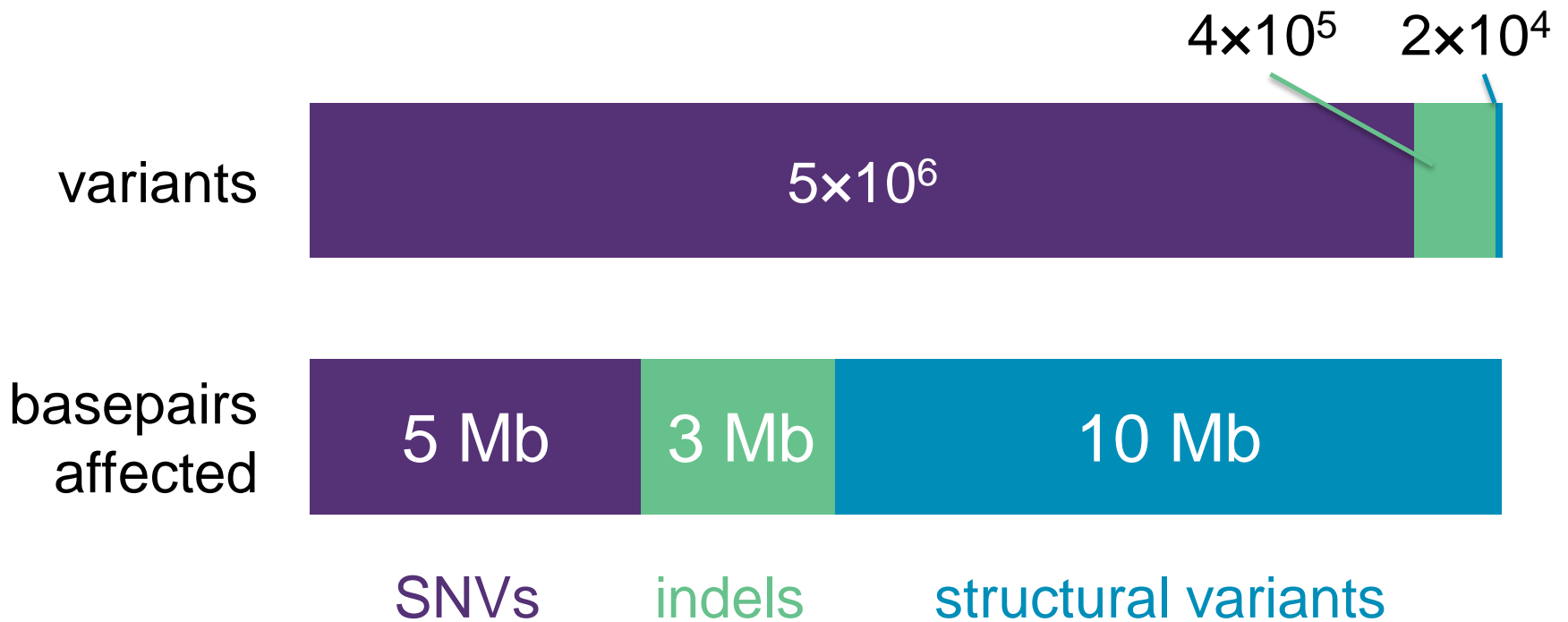
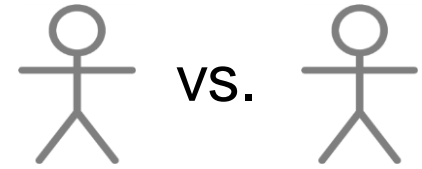
translocation



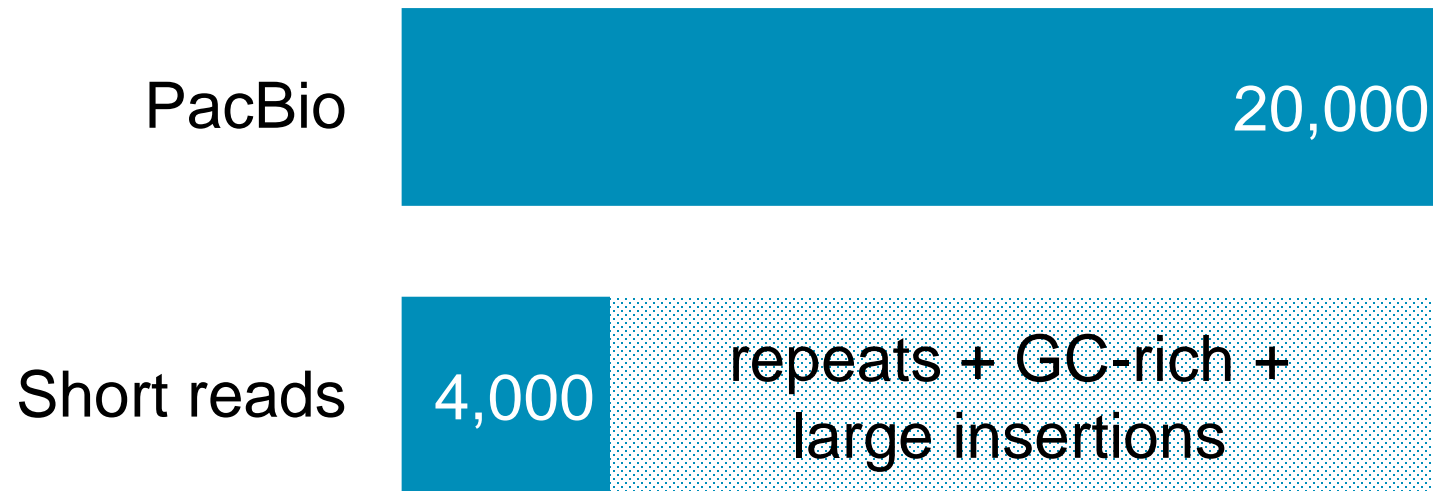
repeat expansion



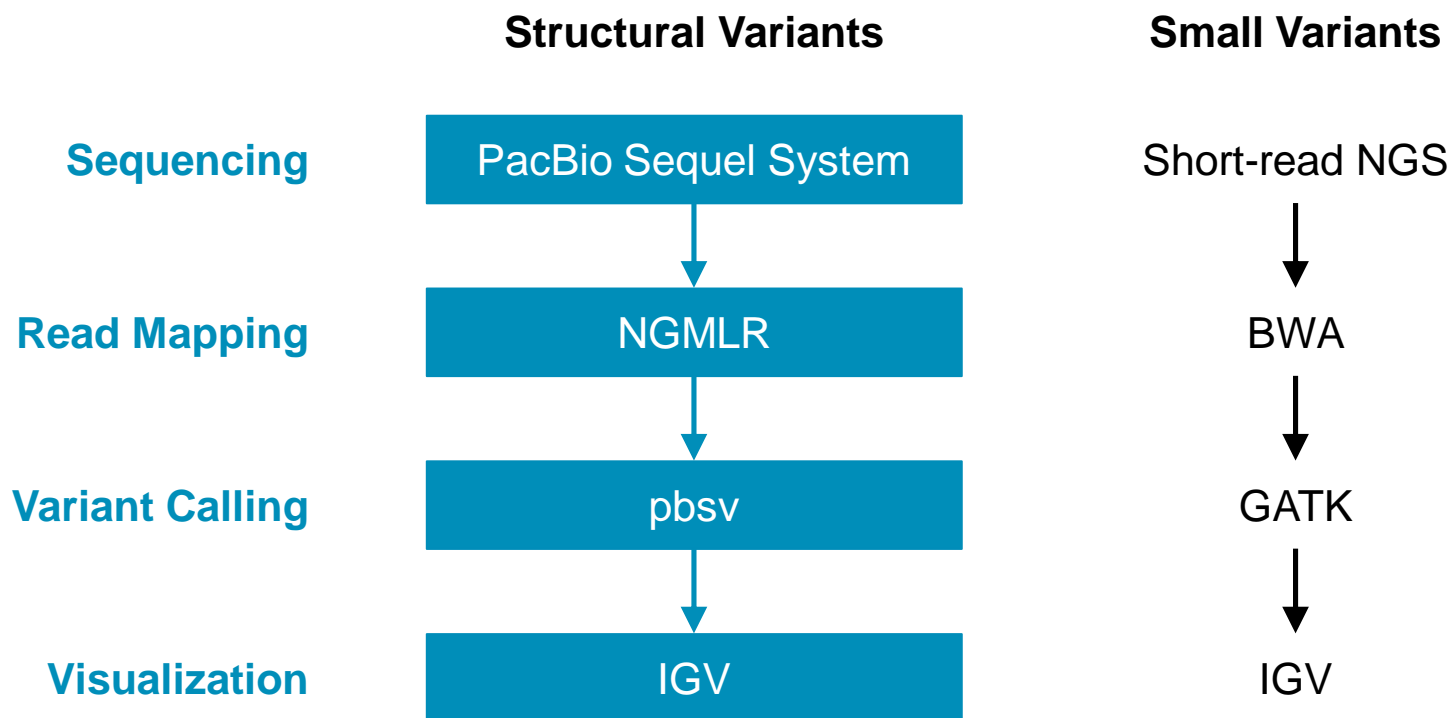
VARIATION BETWEEN TWO HUMAN GENOMES



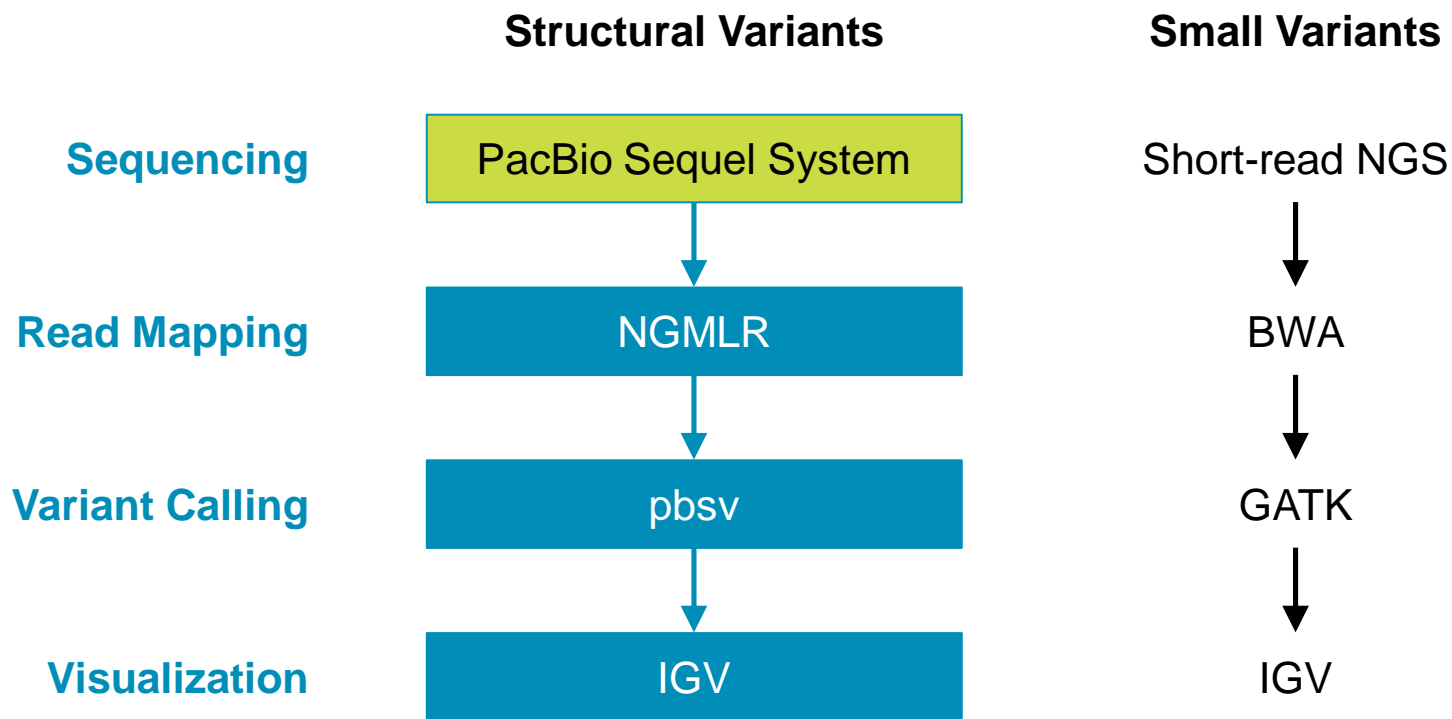
STRUCTURAL VARIANTS DETECTED IN A HUMAN GENOME



WGS FOR STRUCTURAL VARIANT DISCOVERY



WGS FOR STRUCTURAL VARIANT DISCOVERY



SEQUENCING

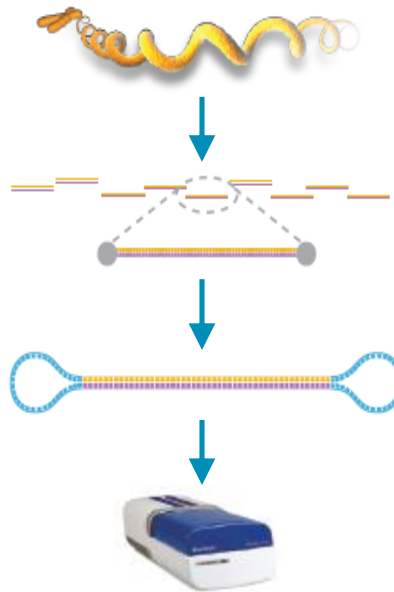
Library Preparation

5 μ g DNA

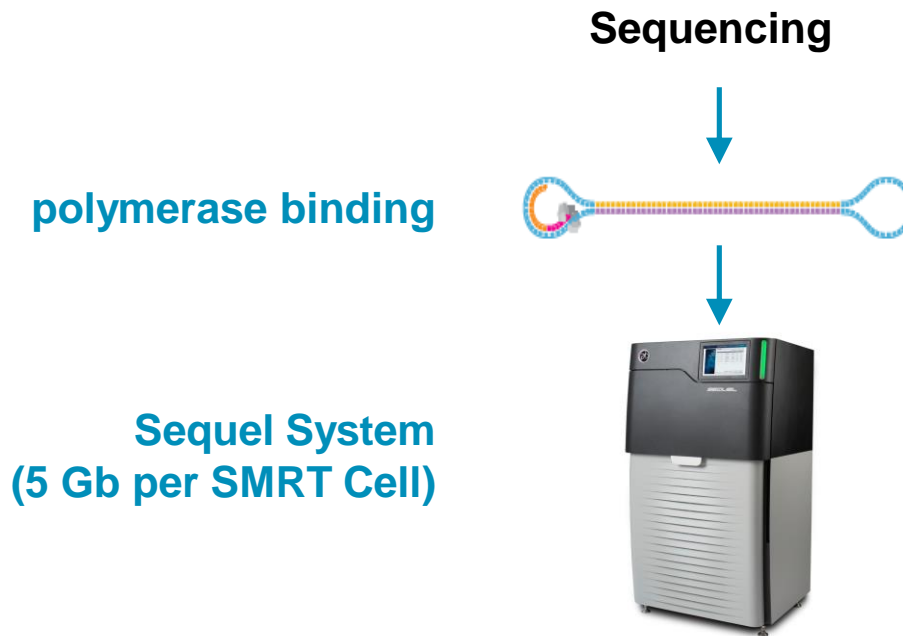
20 kb shear
+ damage repair

SMRTbell adapter ligation

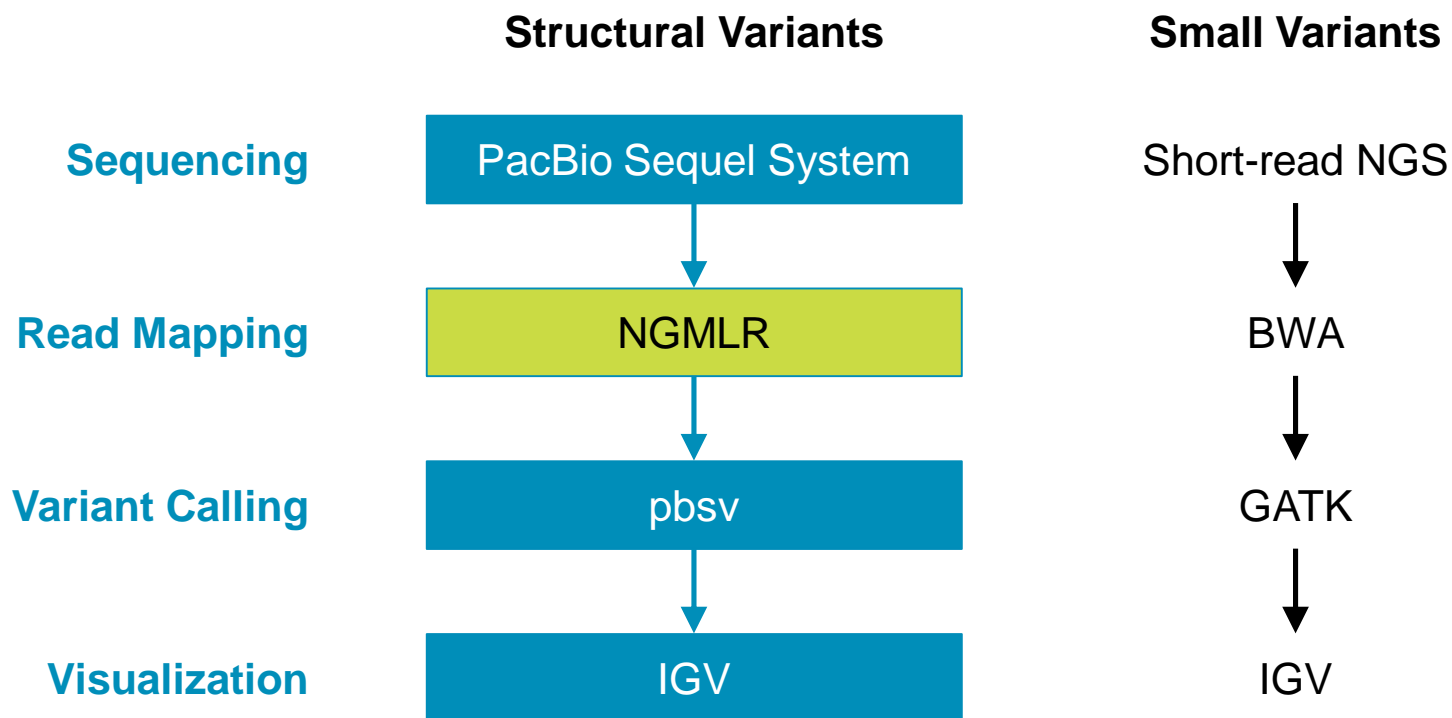
15 kb size selection



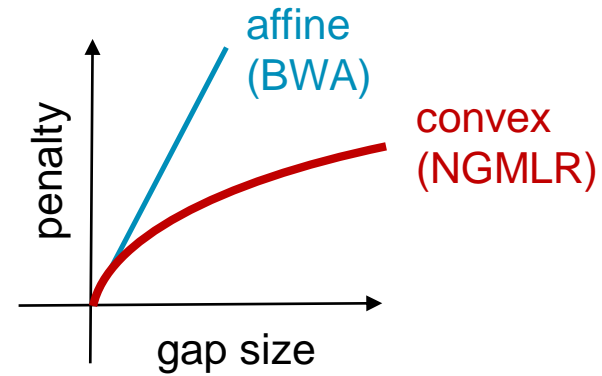
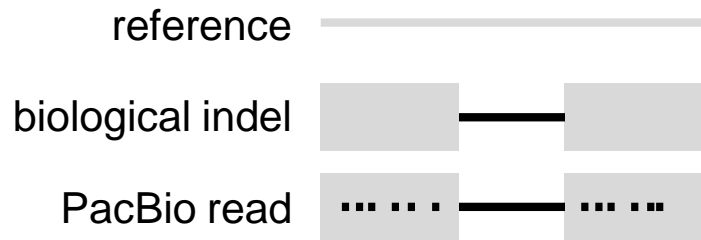
SEQUENCING



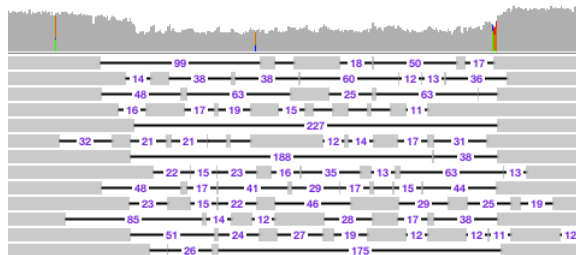
WGS FOR STRUCTURAL VARIANT DISCOVERY



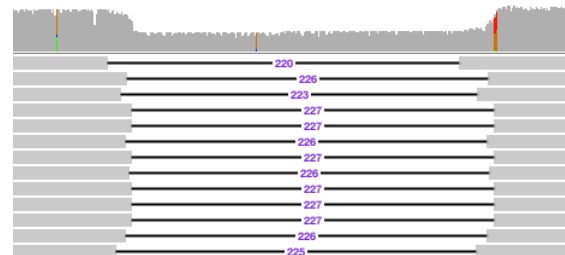
READ MAPPING



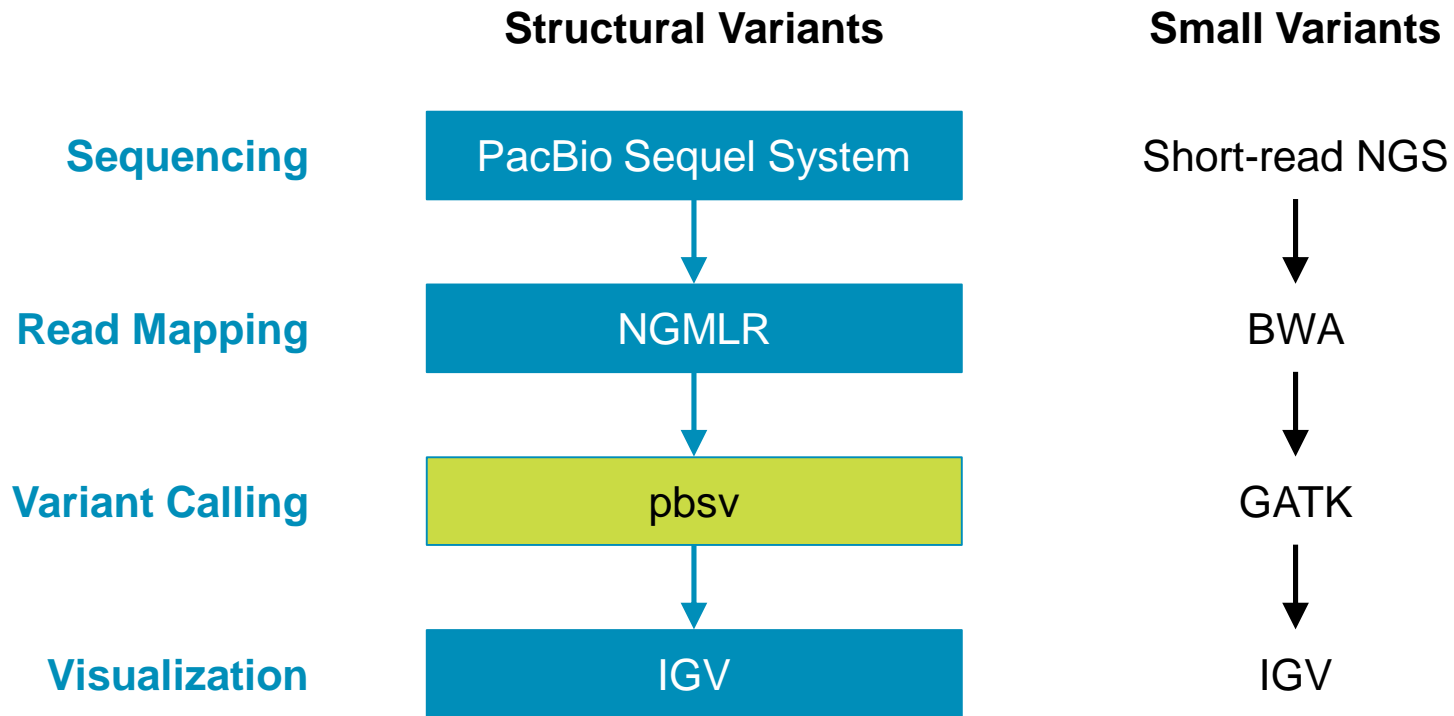
BWA



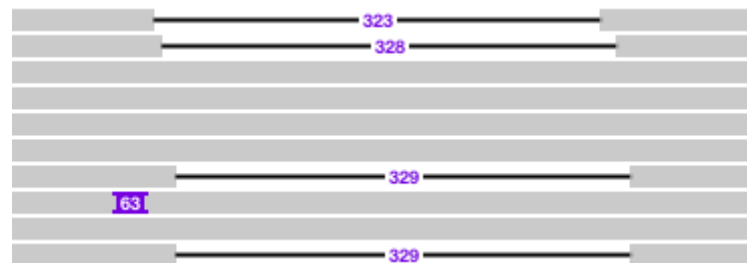
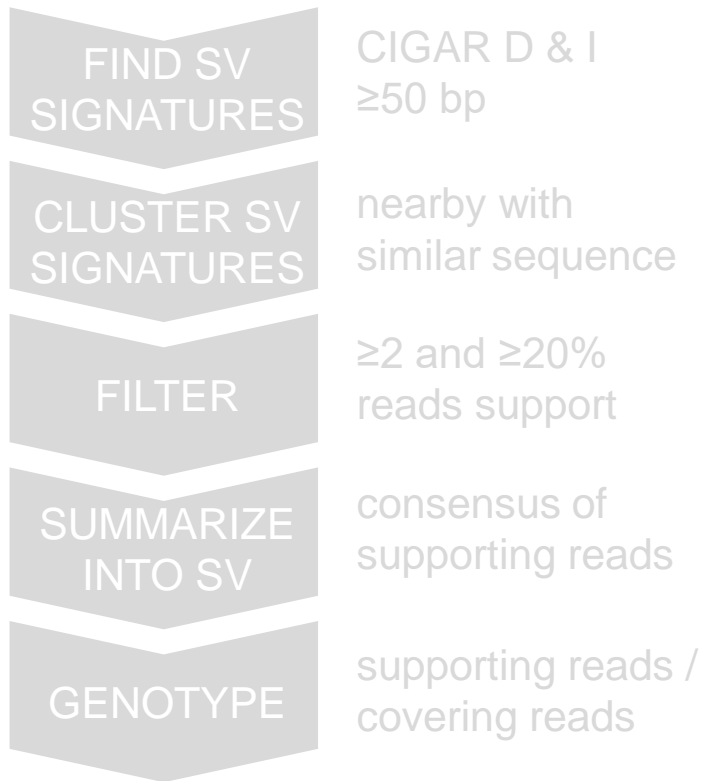
NGMLR



WGS FOR STRUCTURAL VARIANT DISCOVERY



VARIANT CALLING



VARIANT CALLING



FIND SV SIGNATURES
 CIGAR D & I
 ≥ 50 bp

CLUSTER SV SIGNATURES

nearby with similar sequence

FILTER

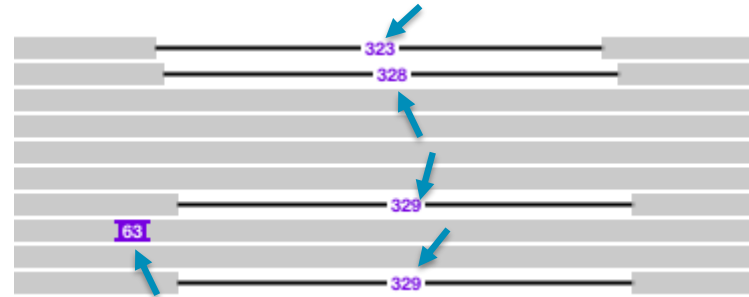
≥ 2 and $\geq 20\%$ reads support

SUMMARIZE INTO SV

consensus of supporting reads

GENOTYPE

supporting reads / covering reads



VARIANT CALLING



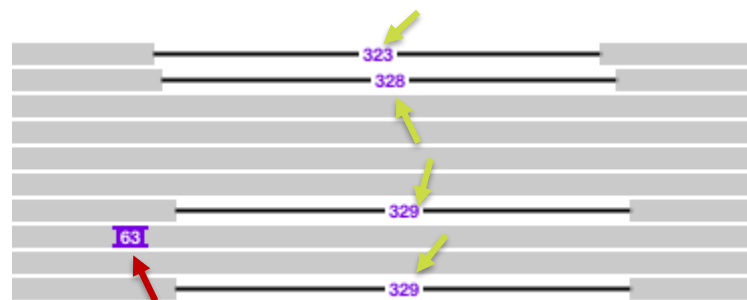
CIGAR D & I
 ≥ 50 bp

nearby with
 similar sequence

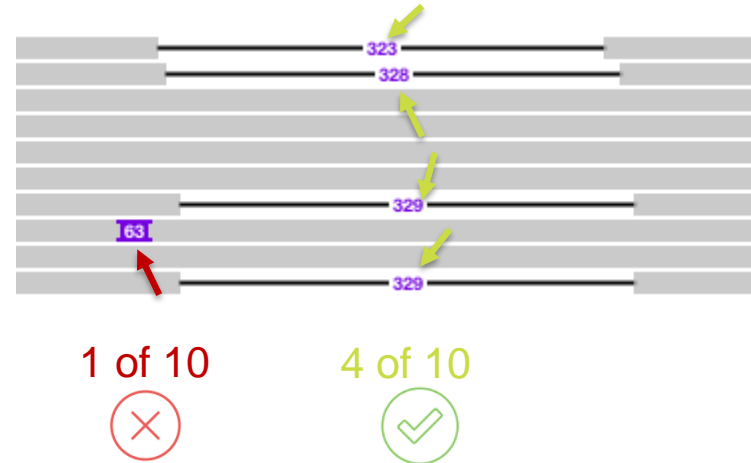
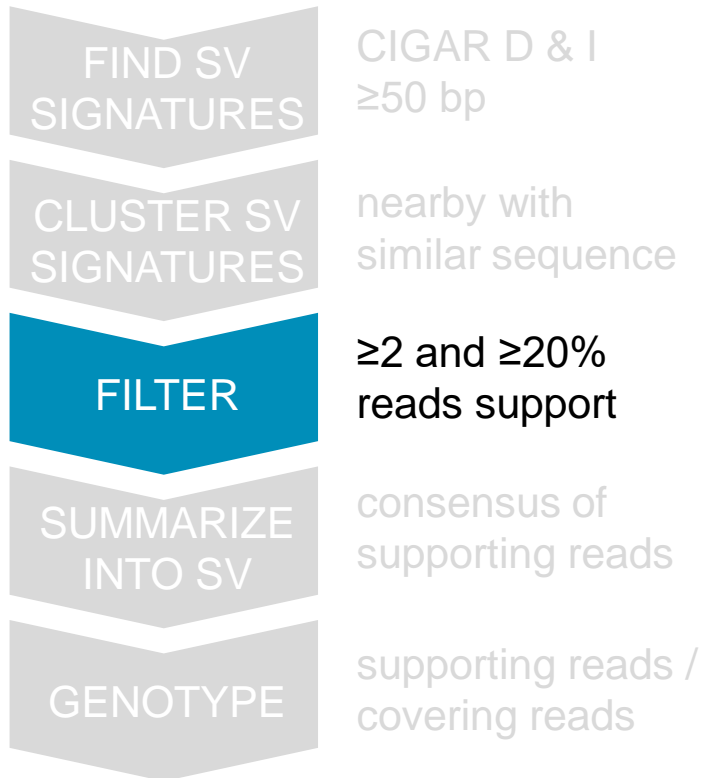
≥ 2 and $\geq 20\%$
 reads support

consensus of
 supporting reads

supporting reads /
 covering reads



VARIANT CALLING



VARIANT CALLING



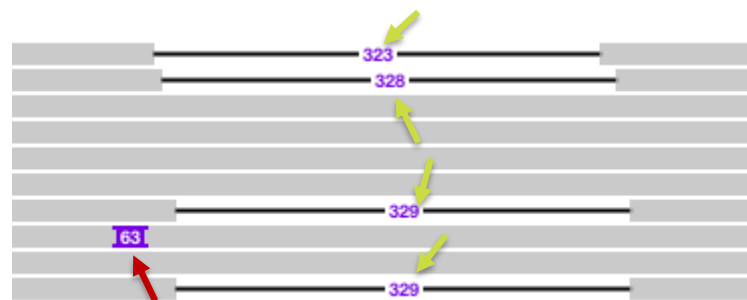
CIGAR D & I
≥50 bp

nearby with
similar sequence

≥2 and ≥20%
reads support

consensus of
supporting reads

supporting reads /
covering reads



1 of 10

4 of 10

329 bp deletion

VARIANT CALLING



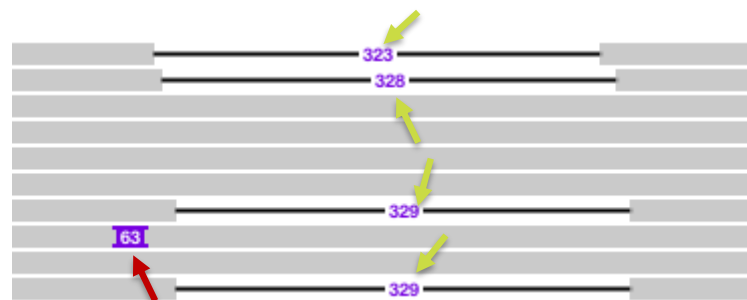
CIGAR D & I
≥50 bp

nearby with
similar sequence

≥2 and ≥20%
reads support

consensus of
supporting reads

supporting reads /
covering reads



1 of 10



4 of 10



329 bp deletion

heterozygous (4 of 10)

VARIANT CALLING

SMRT Analysis

Create New Analysis - Settings [CANCEL] [START]

Name *
HG00733 10-fold SV

Data Sets

| Name |
|--|
| <input checked="" type="checkbox"/> HG00733_Subreads |

Analysis Application *
Structural Variant Calling [Beta]

References *
hg38

Structural Variants

minimum reads that support variant (count) ⓘ
2

minimum reads that support variant (percent) ⓘ
0.2

SMRT Analysis

Data

| File Downloads | File | Size | Type |
|----------------|---------------------|----------------------|------|
| Analysis Log | Analysis Log | 0 bytes | log |
| Analysis Log | Structural variants | 12,426,862 bytes | vcf |
| | Structural variants | 6,823,102 bytes | bed |
| | Aligned reads | 18,542,771,404 bytes | bam |
| | Master Log | 633 bytes | log |

chr1
904490
ACGGGGCCGGCCTCC TCC TC CGA ACG TG GCC TCC TC CGA ACG CG GCC GGC TC CTC CTC CG AAC GCG GC CGC CTC CT OCT CCGA
A
PASS
IMPRECISE; SVTY PE= DEL; END= 904587; SVLEN= -97; SVANN=TANDEM
GT:AD:DP
0/1:9:15

SMRT Analysis

Report

Count by Annotation

| Count by Annotation | Insertions (count) | Insertions (total bp) | Deletions (count) | Deletions (total bp) | All Variants (count) | All Variants (total bp) |
|--------------------------|--------------------|-----------------------|-------------------|----------------------|----------------------|-------------------------|
| Length Histogram (<1 kb) | | | | | | |
| Tandem Repeat | 7,483 | 2,742,385 | 4,210 | 1,247,305 | 11,893 | 3,989,690 |
| Length Histogram (>1 kb) | | | | | | |
| Atu | 1,236 | 398,032 | 1,177 | 367,370 | 2,413 | 765,402 |
| L1 | 44 | 244,741 | 83 | 444,635 | 127 | 689,376 |
| SV4 | 18 | 31,887 | 29 | 51,831 | 47 | 83,718 |
| Unannotated | 4,344 | 2,007,499 | 2,661 | 2,803,452 | 7,005 | 4,810,911 |
| Total | 13,125 | 5,424,604 | 8,160 | 4,914,593 | 21,285 | 10,339,197 |

SMRT Analysis

Report

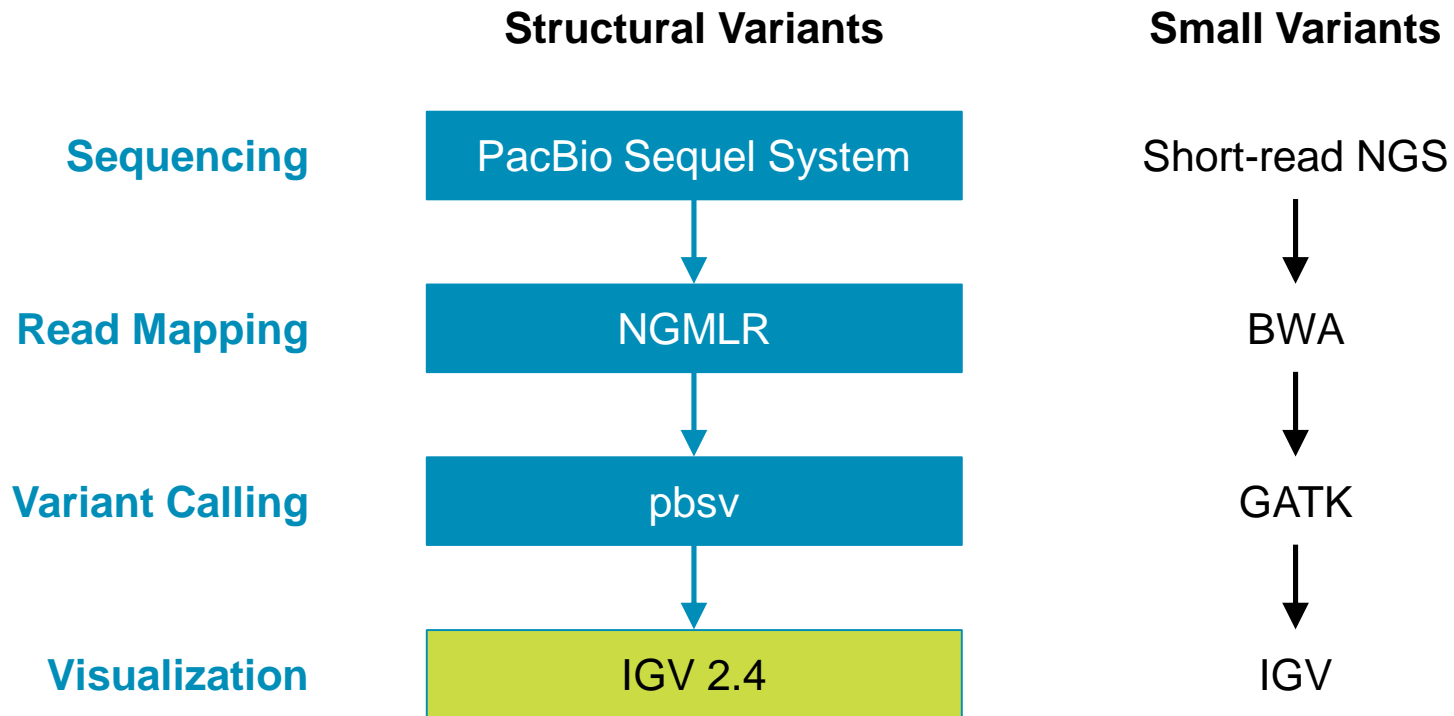
Length Histogram (<1 kb)

Count by Annotation

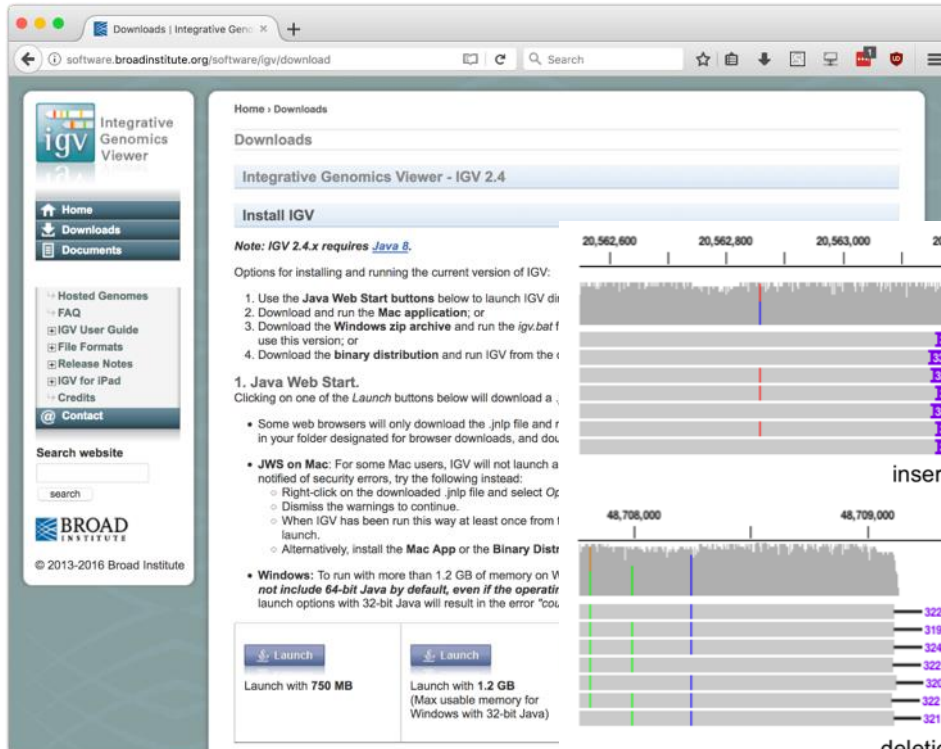
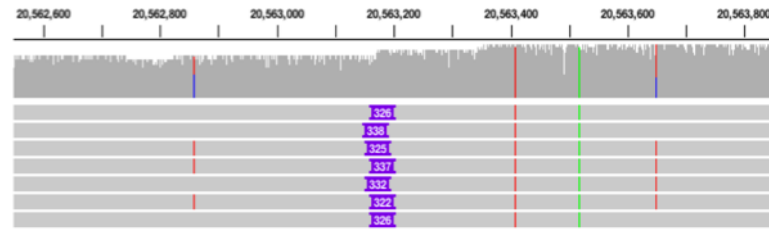
Length Histogram (<1 kb)

Length Histogram (>1 kb)

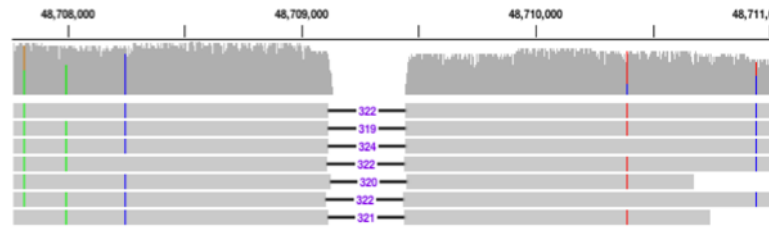
WGS FOR STRUCTURAL VARIANT DISCOVERY



VISUALIZATION

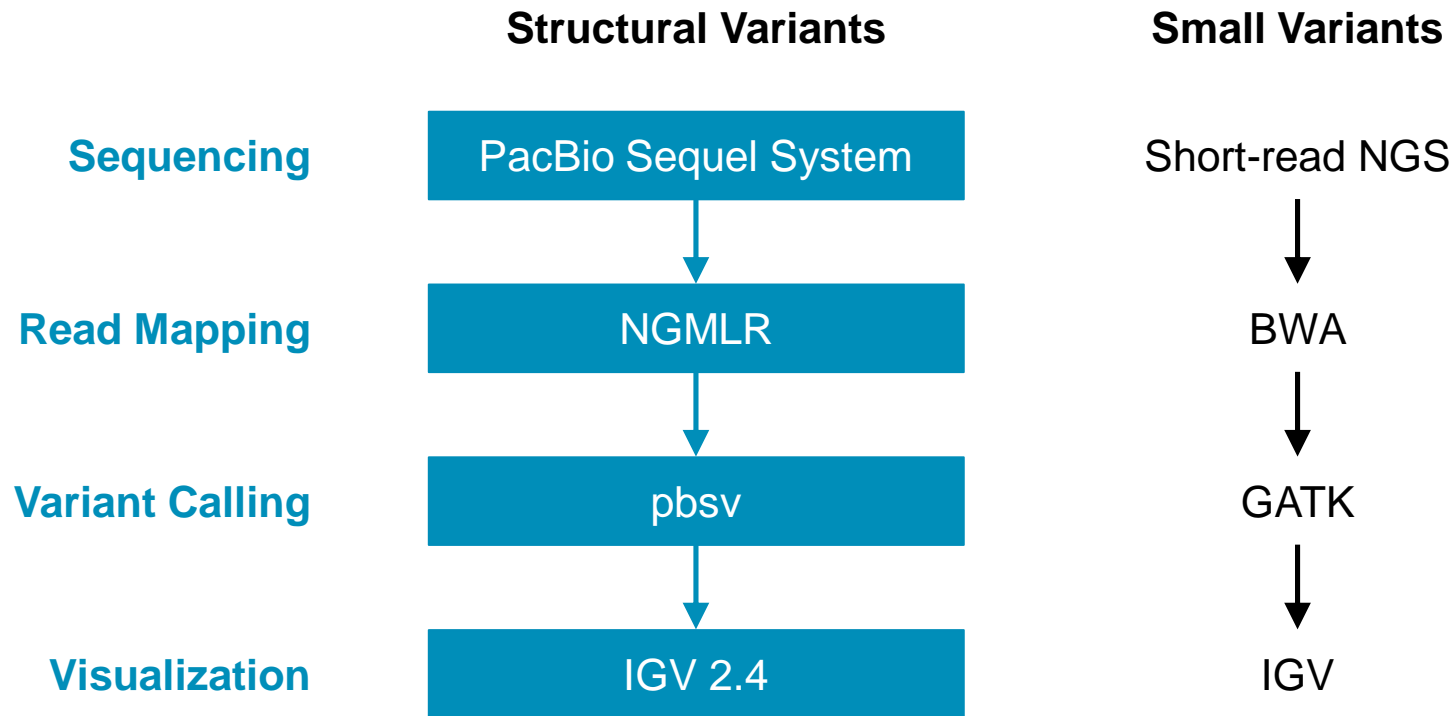



insertion



deletion

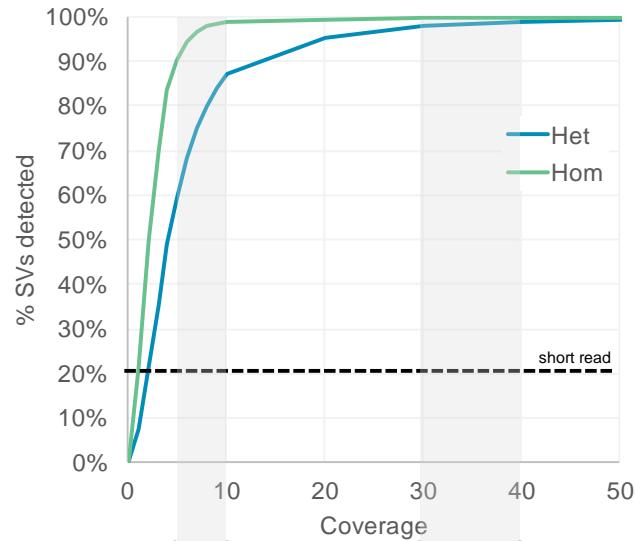
WGS FOR STRUCTURAL VARIANT DISCOVERY



HOW MUCH TO SEQUENCE?



Human HG00733
Sequel System
211 Gb (70-fold)



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

disease gene discovery;
population characterization

30- to 40-fold
saturate discovery

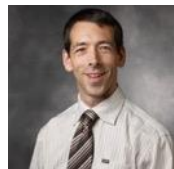
de novo variant discovery

CLINICAL CASE HISTORY

- 7 yrs left atrial myxoma resection, atrial repair
- 10 yrs testicular mass, right orchiectomy
- 13 yrs pituitary tumor
- 16 yrs recurrence of myxomata, resection, adrenal microadenoma
- 18 yrs recurrence of ventricular myxomata, resection, VT
- 19 yrs ACTH-independent Cushing's disease, thyroid nodules
- 21 yrs transphenoidal resection of pituitary
- present (26 yrs) recurrence of myxomata, consideration for heart transplant

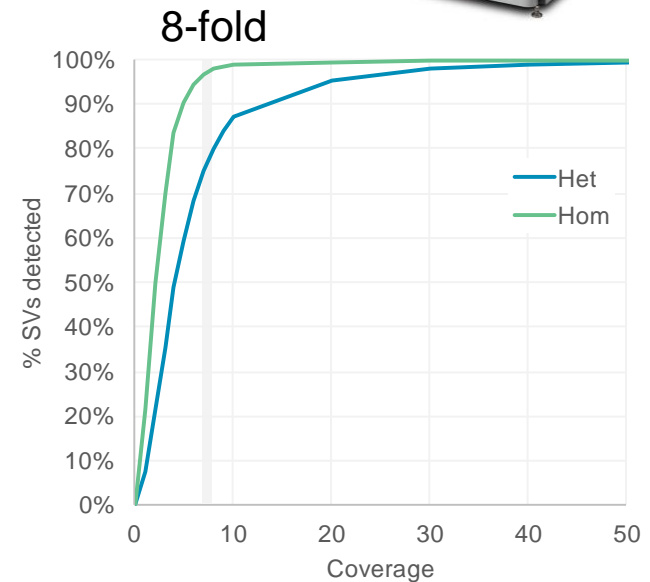
→ genetics suggests Carney complex
PRKAR1A testing negative

→ short-read whole genome sequencing negative



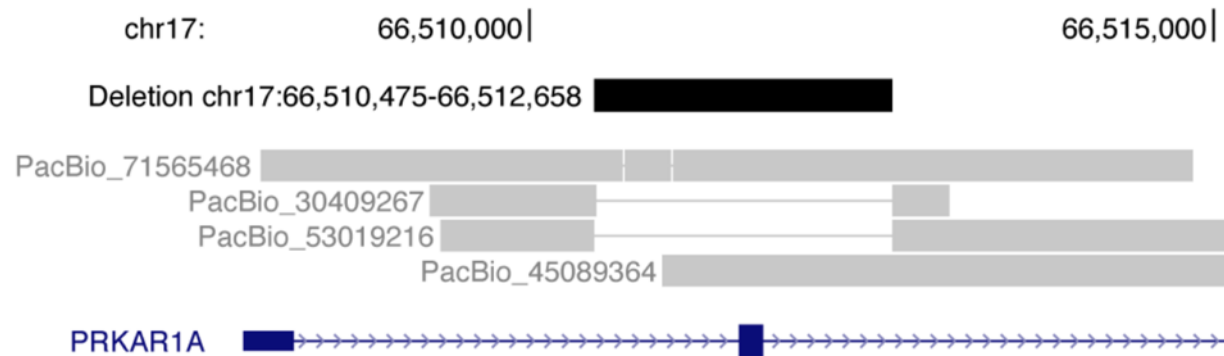
EVALUATING STRUCTURAL VARIANTS

| | Deletions | Insertions |
|--|-----------|------------|
| Initial call set | 6,971 | 6,821 |
| Not in segdup | 5,893 | 6,254 |
| Not in NA12878 "healthy" control | 2,476 | 3,171 |
| Overlaps RefSeq coding exon | 39 | 16 |
| Gene linked to some disease in OMIM | 3 | 3 |



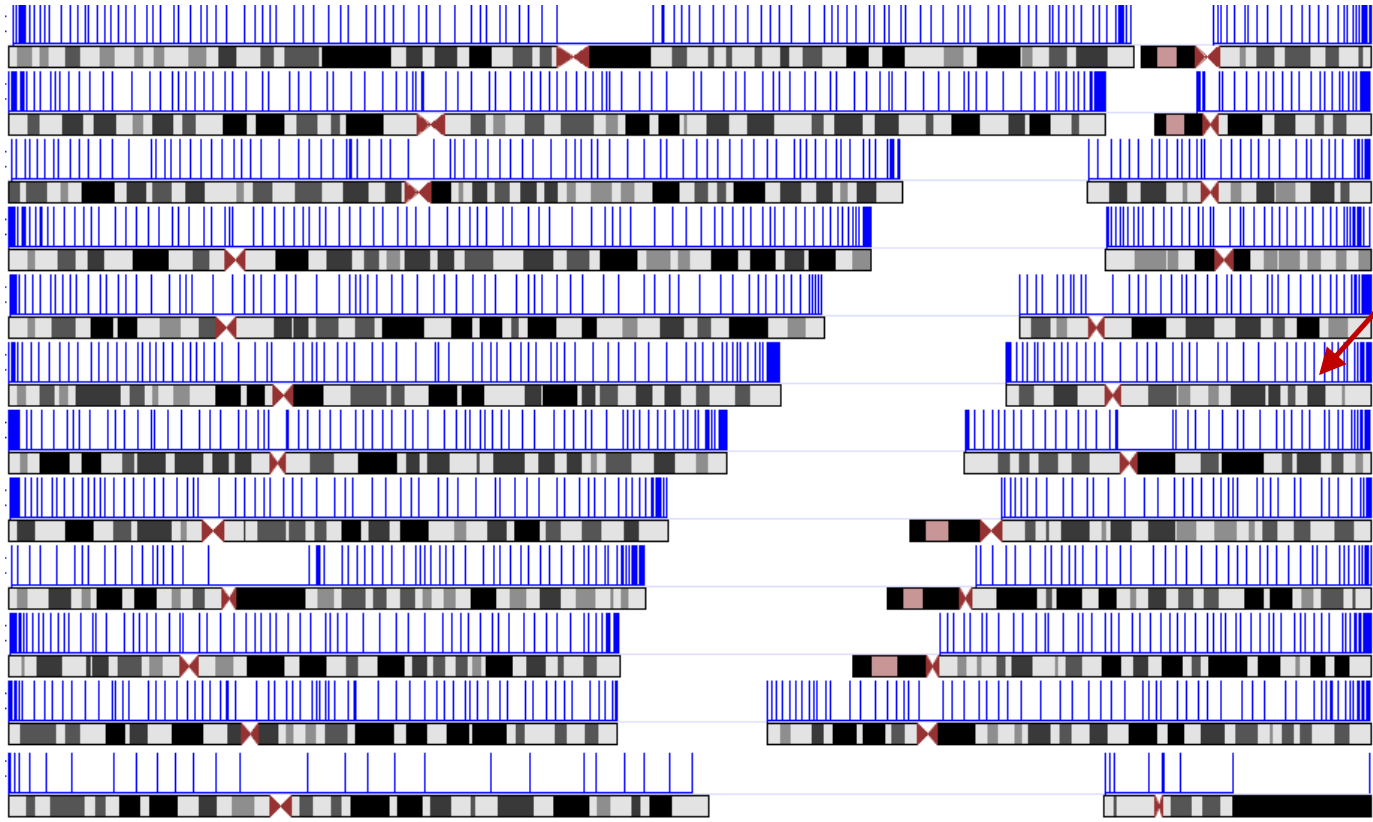
HETEROZYGOUS 2.2 KB DELETION IN *PRKAR1A*

PacBio
discovery



Sanger
confirmation







www.pacb.com

For Research Use Only. Not for use in diagnostics procedures. © Copyright 2017 by Pacific Biosciences of California, Inc. All rights reserved. Pacific Biosciences, the Pacific Biosciences logo, PacBio, SMRT, SMRTbell, Iso-Seq, and Sequel are trademarks of Pacific Biosciences. BluePippin and SageELF are trademarks of Sage Science. NGS-go and NGSengine are trademarks of GenDx. FEMTO Pulse and Fragment Analyzer are trademarks of Advanced Analytical Technologies.

All other trademarks are the sole property of their respective owners.