



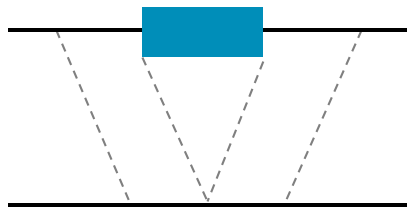
# Structural Variant Detection in SMRT Link 5 with pbsv

Aaron Wenger

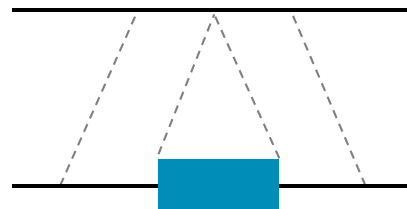
2017-06-27

## STRUCTURAL VARIANT = DIFFERENCE $\geq 50$ BP

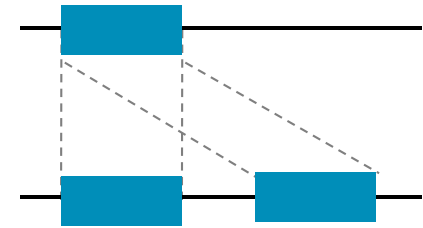
Deletion



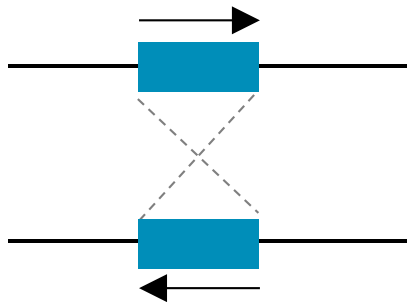
Insertion



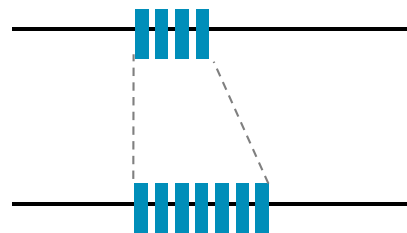
Duplication



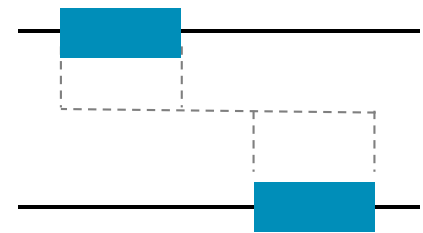
Inversion



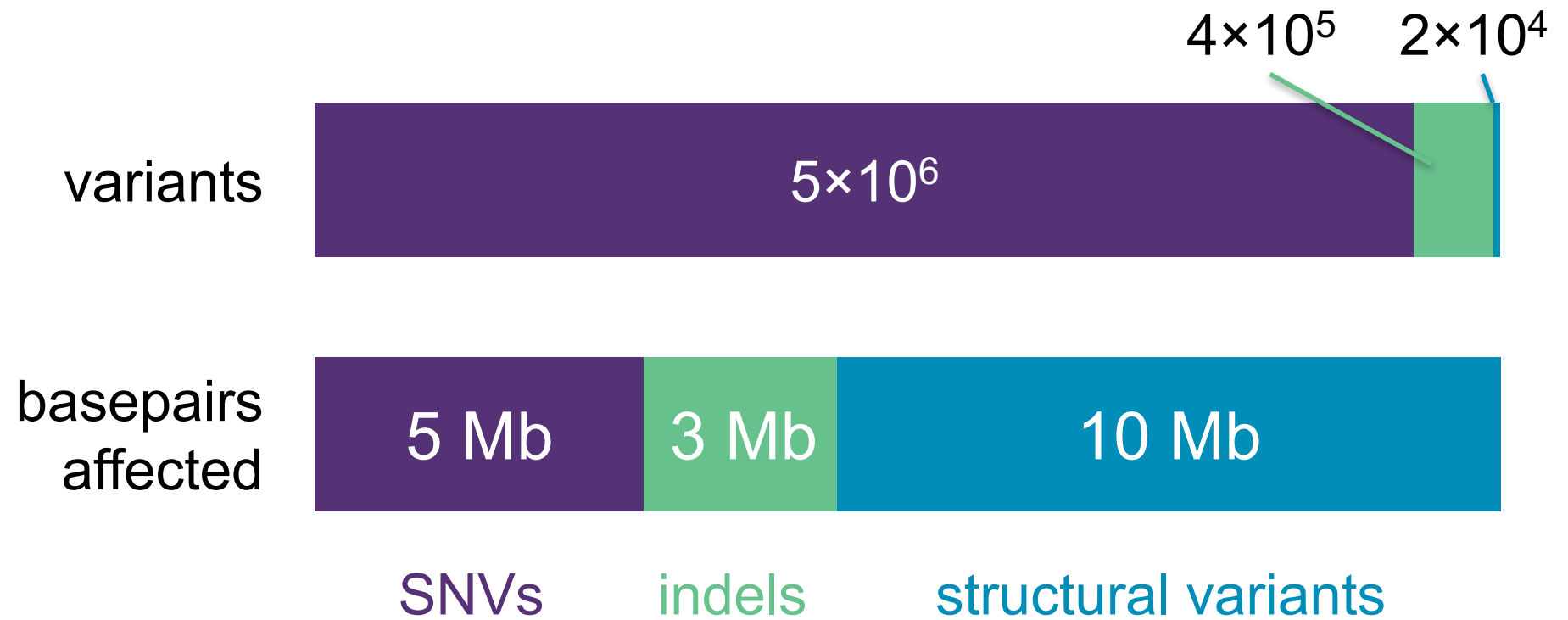
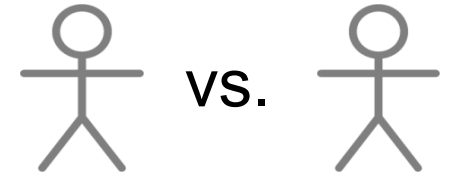
Tandem Repeat



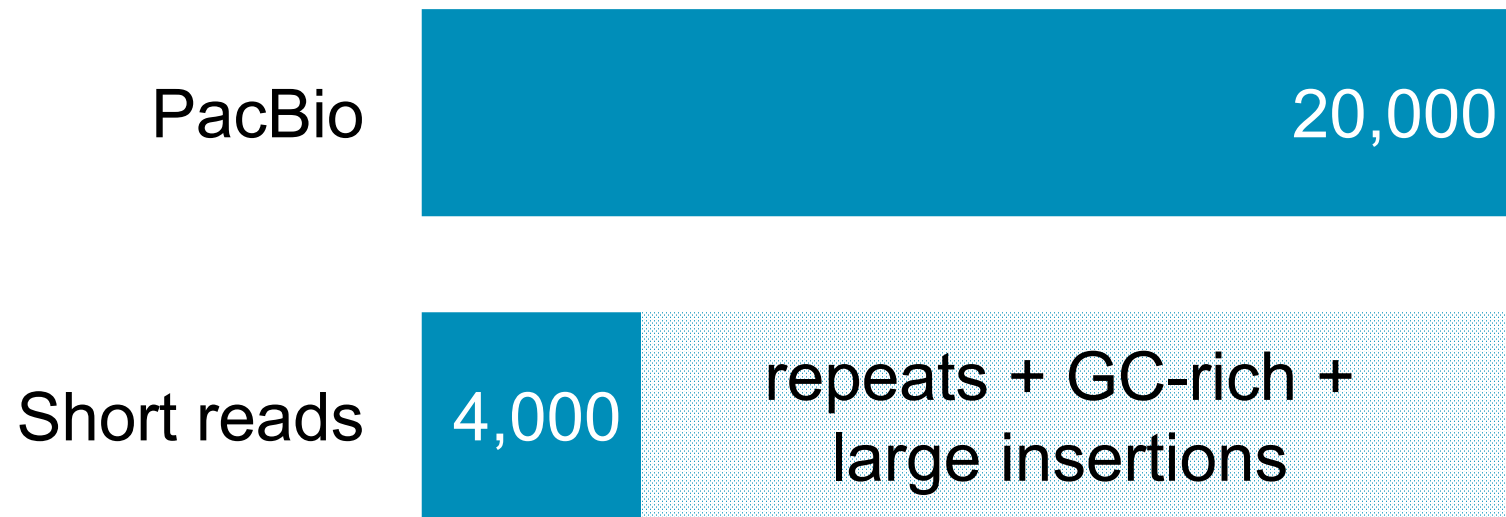
Translocation



# VARIATION BETWEEN TWO HUMAN GENOMES



# STRUCTURAL VARIANTS DETECTED IN A HUMAN GENOME



Huddleston et al. (2017) *Genome Research* 27(5):677-85.  
Seo et al. (2016) *Nature* 538:243-7.  
Sudmant et al. (2016) *Nature* 526:75-81.

# SEQUENCING + ANALYSIS



**pbsv**



Structural Variants

Short reads



**BWA**



SNVs + Indels

Li and Durbin (2009) *Bioinformatics* 25:1754-60.  
McKenna et al. (2010) *Genome Research* 20:1297-303.

## 3 COMPONENTS TO PBSV

**pbsv** command line utility for top-level commands

**pbsvutil** command line utility for detailed commands

**SMRT Link web interface**

## TOP-LEVEL PBSV COMMANDS

```
pbsv generate-config [-h] [-o sv.cfg]
```

(optional) Generate a configuration file to specify options for other stages.

```
pbsv align [-h] [--cfg_fn sv.cfg]  
ref.fa subreads.bam ref.align.bam
```

Map reads to a reference genome with a “structural variant aware” aligner.

```
pbsv call [-h] [--cfg_fn sv.cfg]  
ref.fa ref.align.bam ref.sv.bed|vcf
```

Call structural variants from aligned reads.

## TOP-LEVEL PBSV COMMANDS

```
pbsv generate-config [-h] [-o sv.cfg]
```

(optional) Generate a configuration file to specify options for other stages.

```
pbsv align [-h] [--cfg_fn sv.cfg]  
ref.fa subreads.bam ref.align.bam
```

Map reads to a reference genome with a “structural variant aware” aligner.

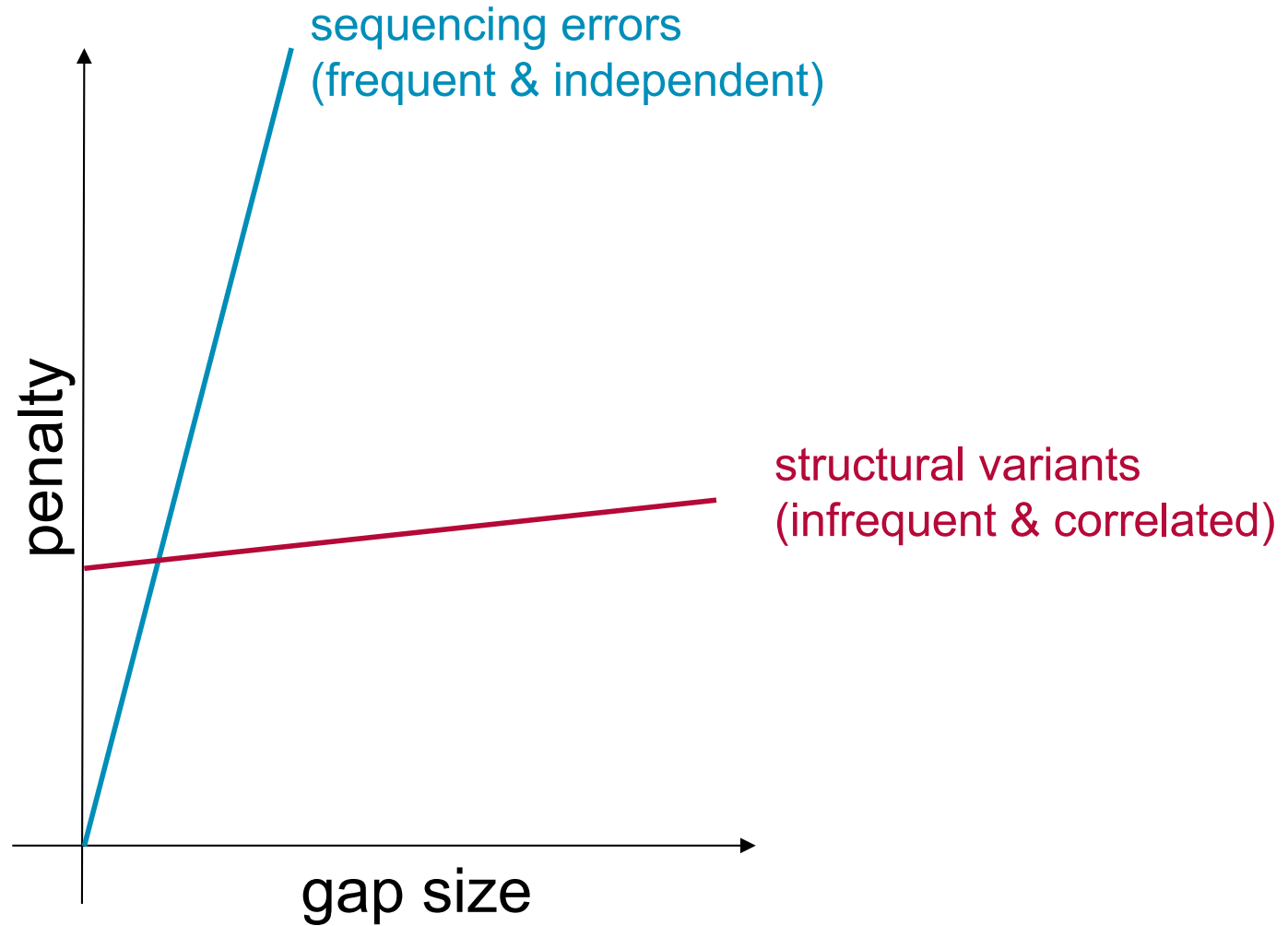
```
pbsv call [-h] [--cfg_fn sv.cfg]  
ref.fa ref.align.bam ref.sv.bed|vcf
```

Call structural variants from aligned reads.



# PBSV ALIGN UTILIZES NGM-LR

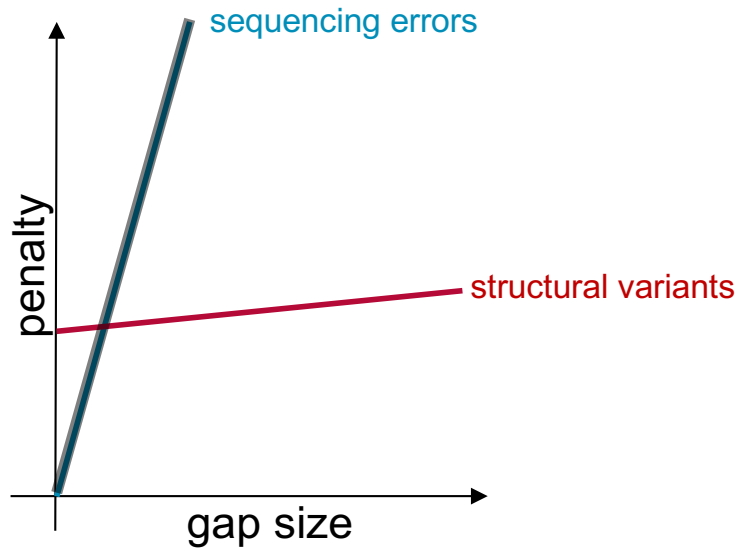
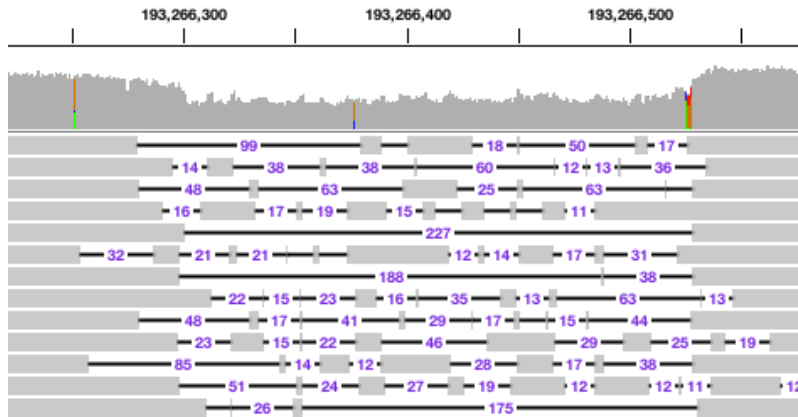
pbsvutil ngmlr



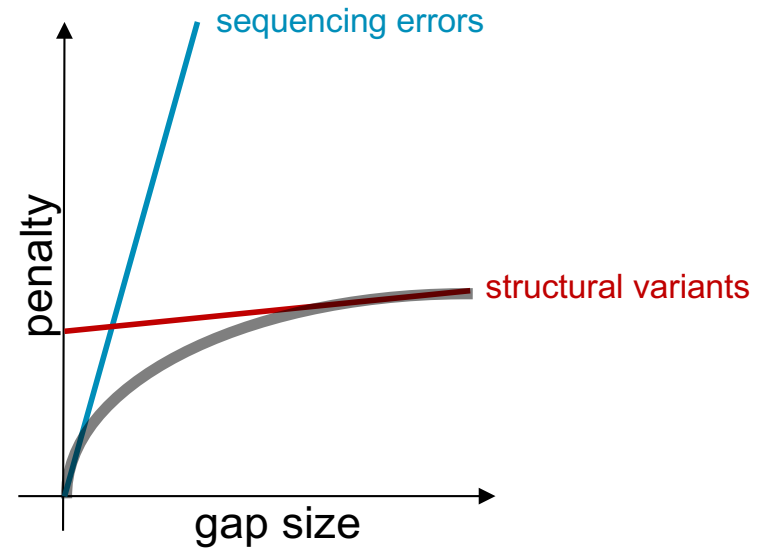
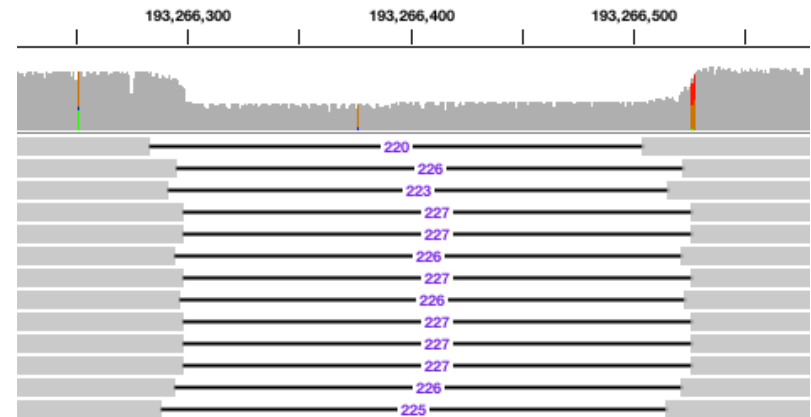
# PBSV ALIGN UTILIZES NGM-LR

pbsvutil ngmlr

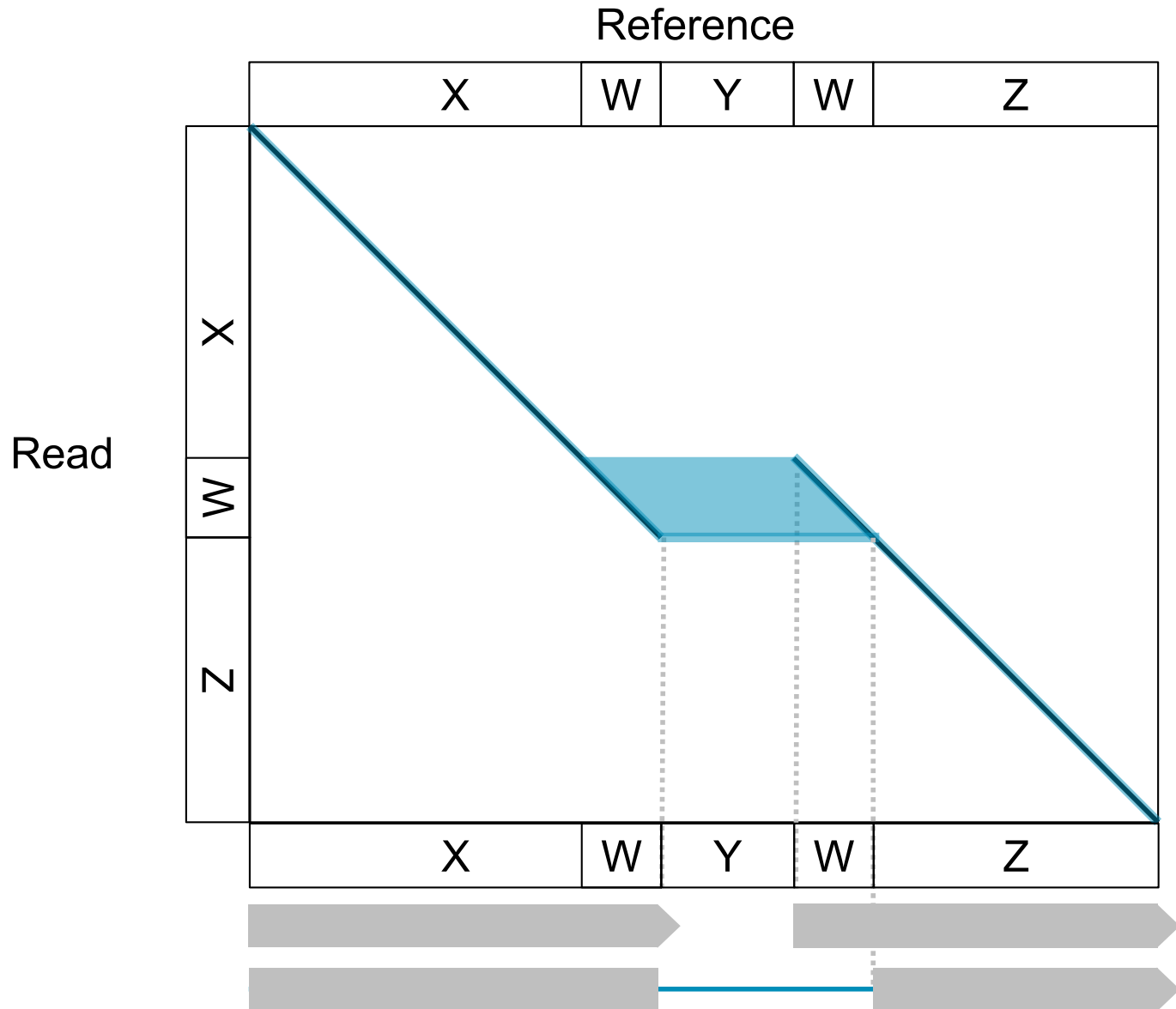
## BWA



## NGM-LR



# PBSV ALIGN CHAINS CO-LINEAR ALIGNMENTS `pbsvutil chain`



## TOP-LEVEL PBSV COMMANDS

```
pbsv generate-config [-h] [-o sv.cfg]
```

(optional) Generate a configuration file to specify options for other stages.

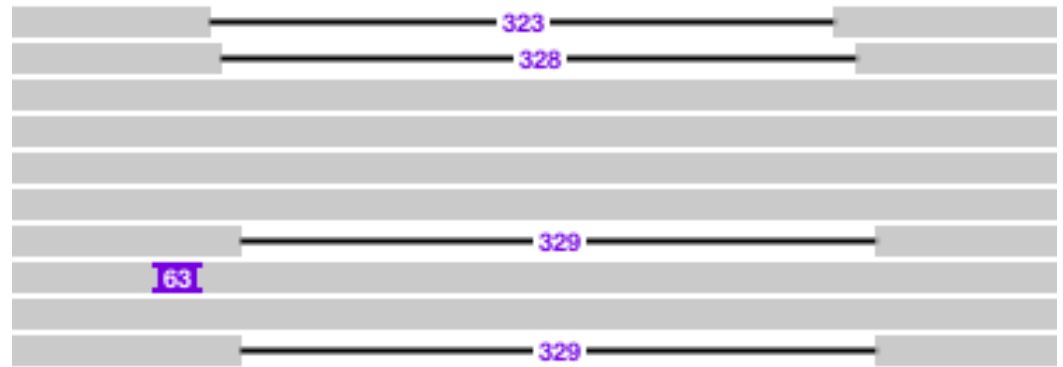
```
pbsv align [-h] [--cfg_fn sv.cfg]  
ref.fa subreads.bam ref.align.bam
```

Map reads to a reference genome with a “structural variant aware” aligner.

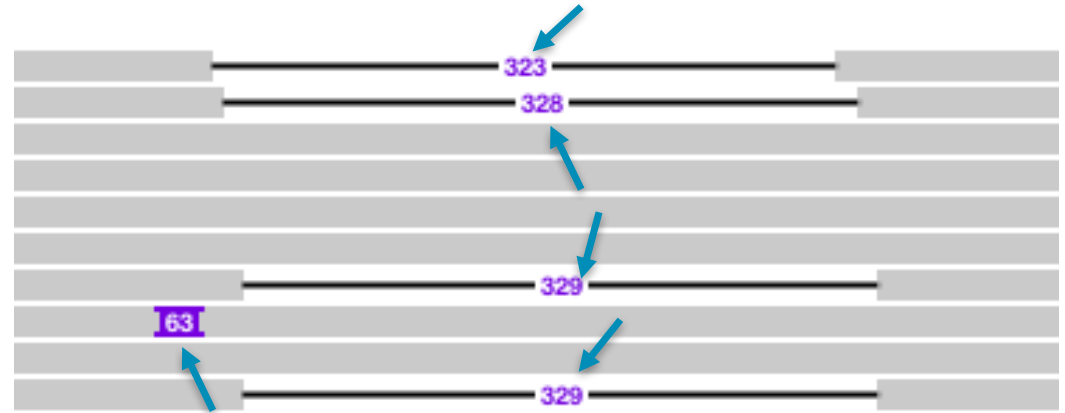
```
pbsv call [-h] [--cfg_fn sv.cfg]  
ref.fa ref.align.bam ref.sv.bed|vcf
```

Call structural variants from aligned reads.

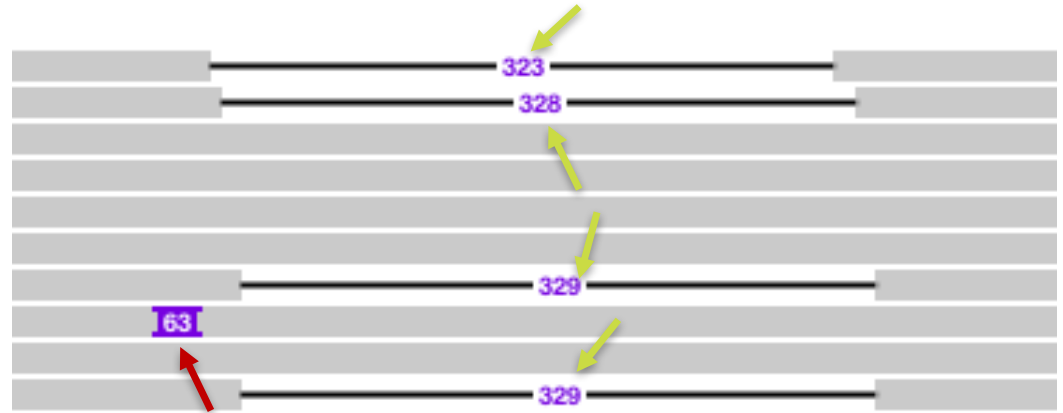
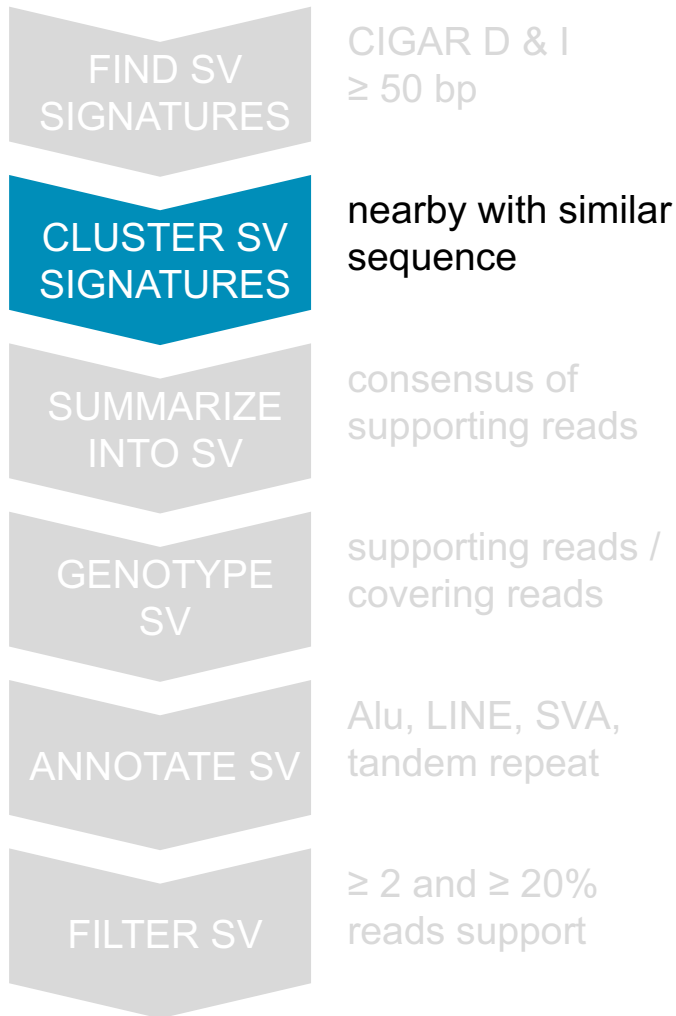
# PBSV CALL: STAGED STRUCTURAL VARIANT CALLER



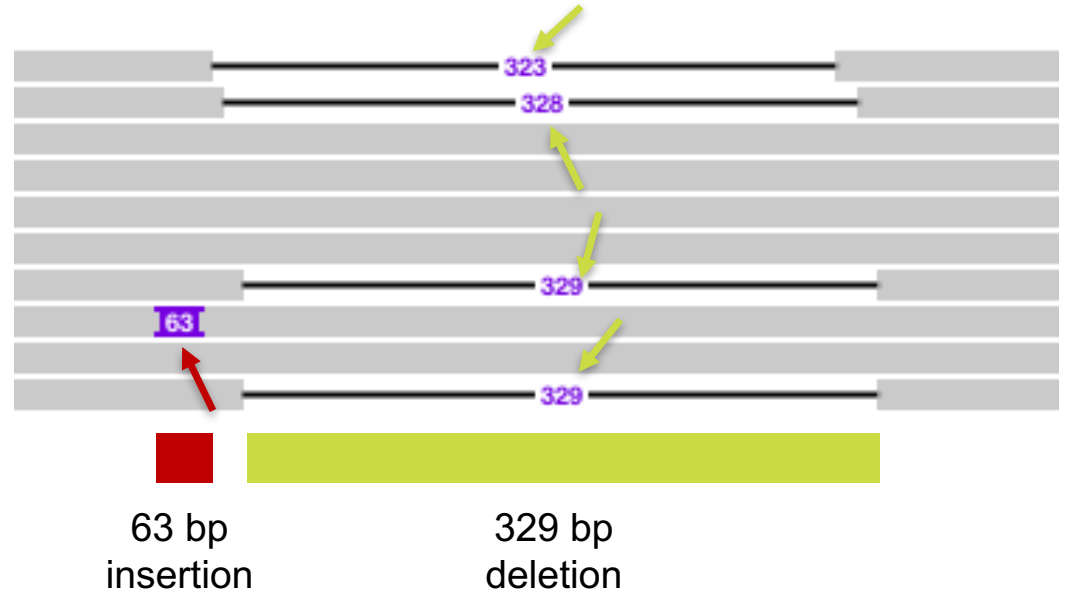
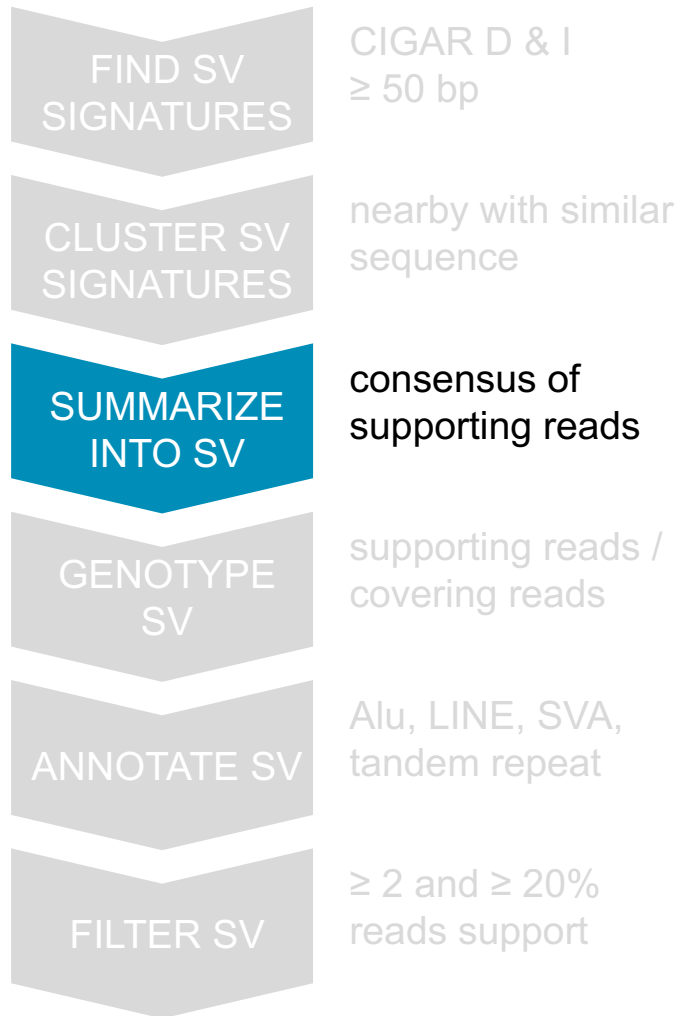
# PBSV CALL: STAGED STRUCTURAL VARIANT CALLER



# PBSV CALL: STAGED STRUCTURAL VARIANT CALLER

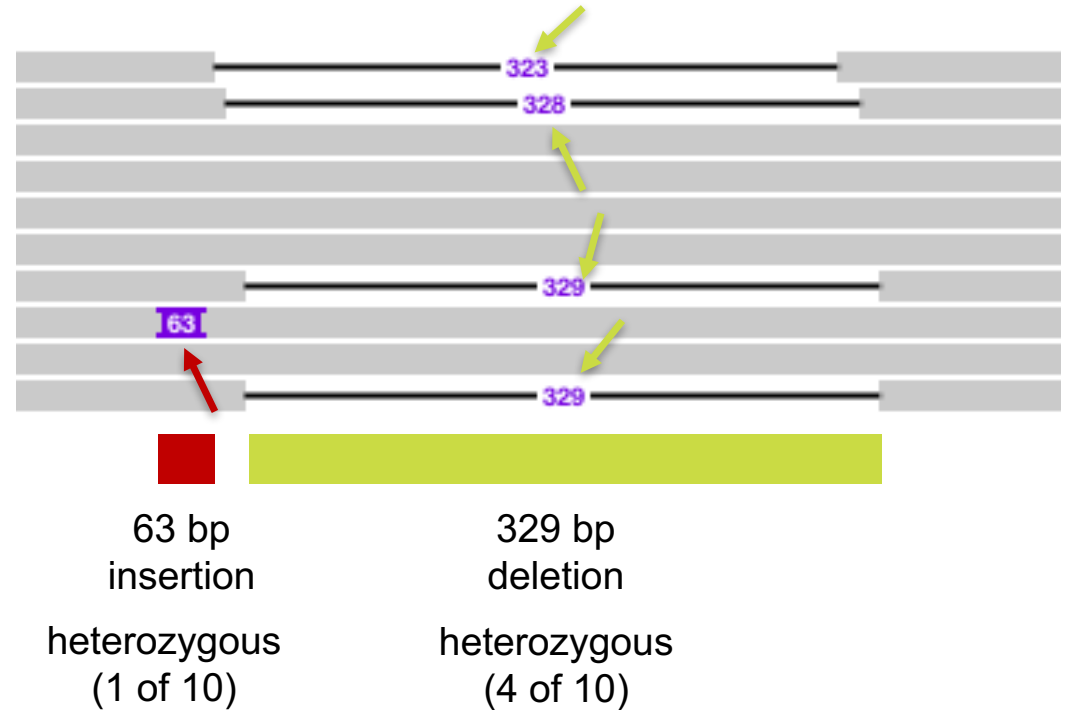
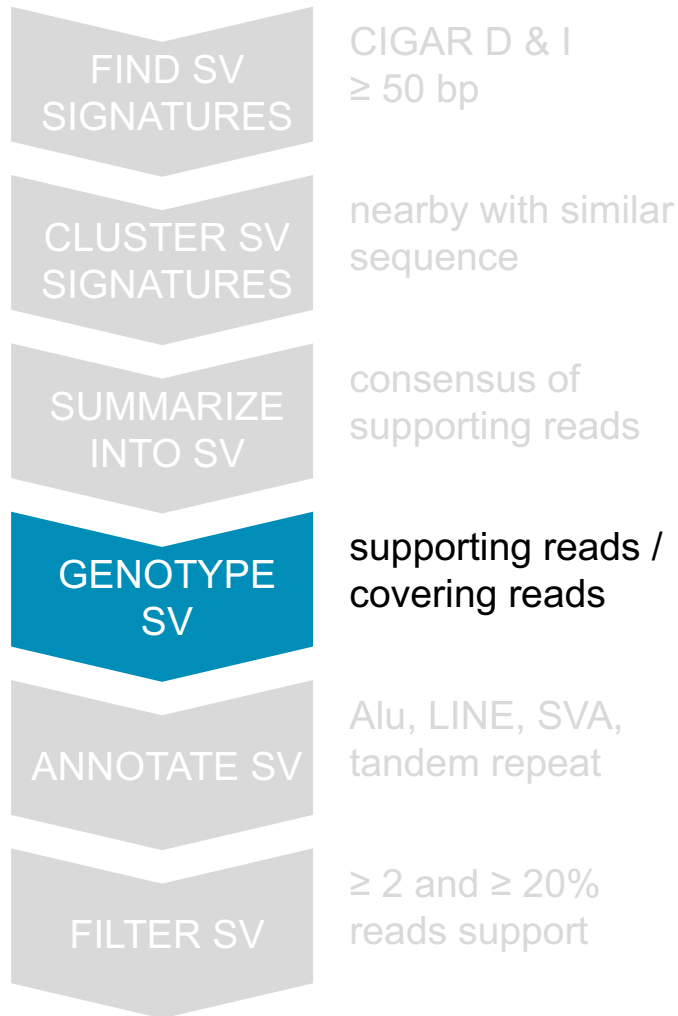


# PBSV CALL: STAGED STRUCTURAL VARIANT CALLER

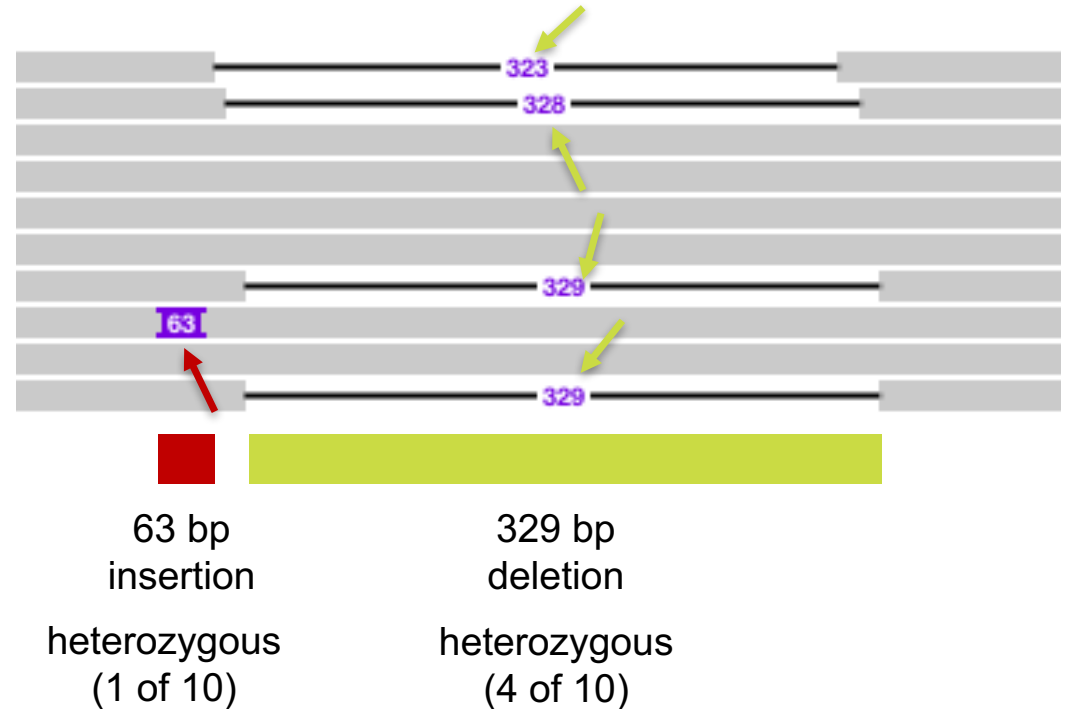




# PBSV CALL: STAGED STRUCTURAL VARIANT CALLER

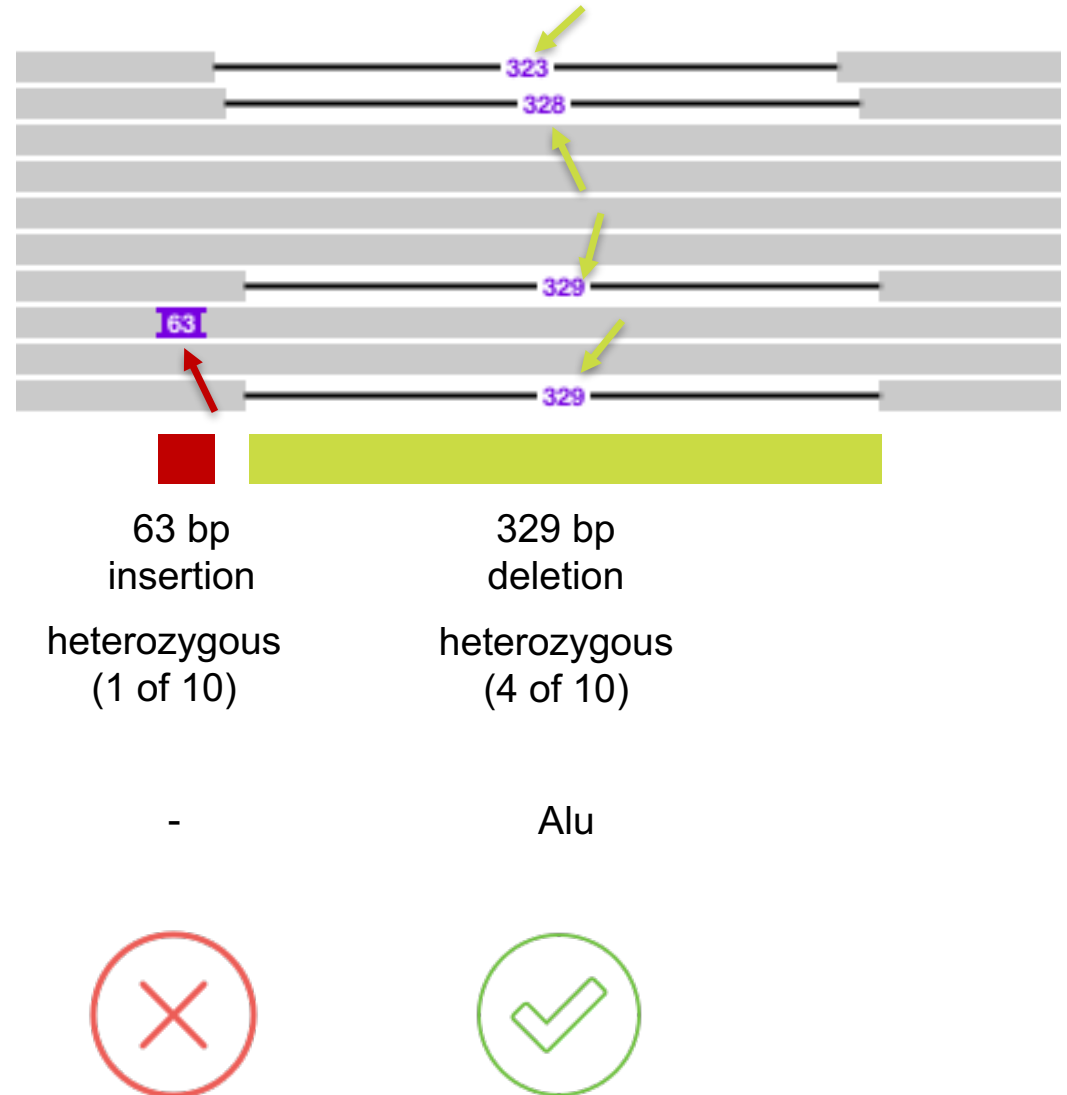


# PBSV CALL: STAGED STRUCTURAL VARIANT CALLER

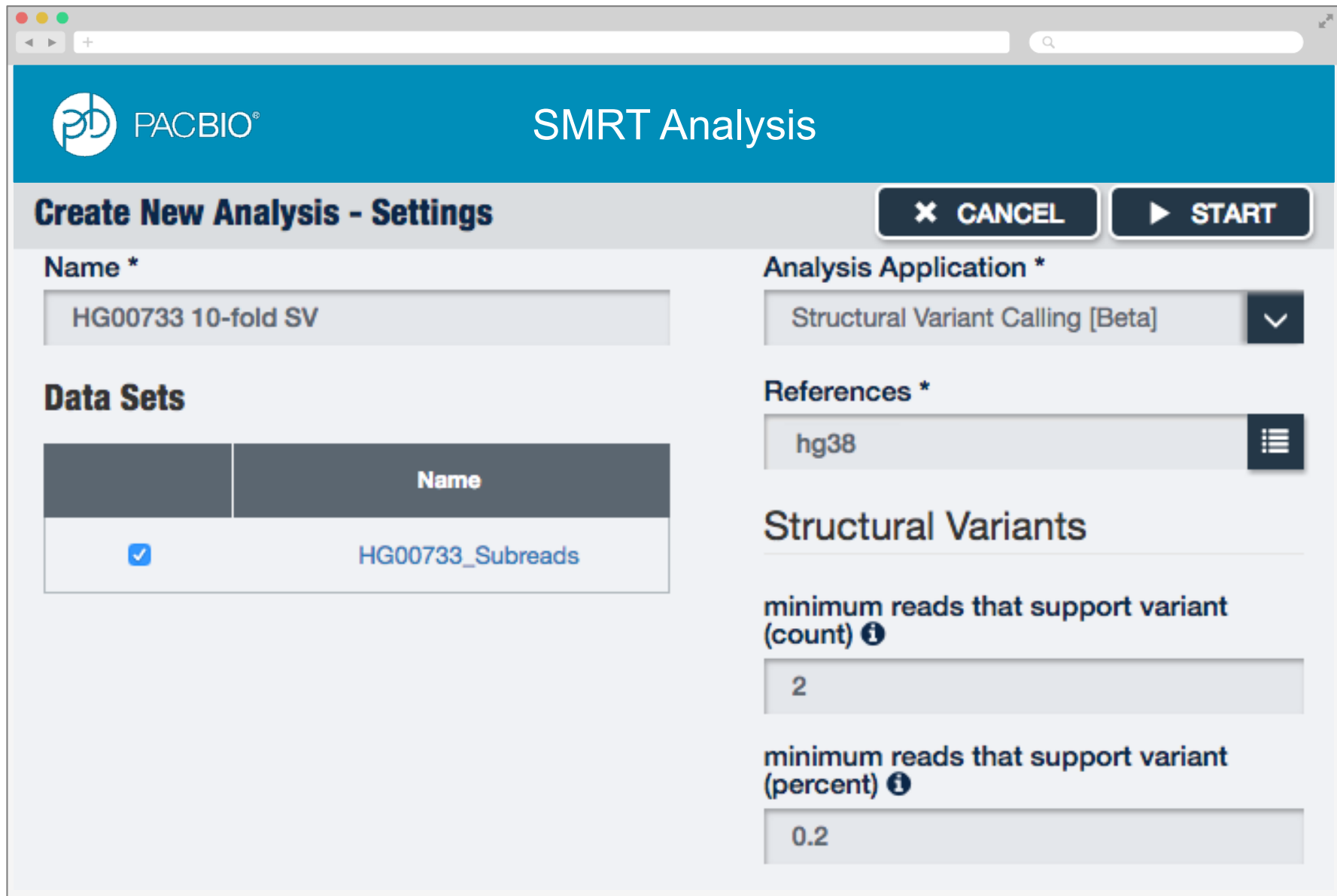


- Alu

# PBSV CALL: STAGED STRUCTURAL VARIANT CALLER



# PBSV: SMRT LINK STRUCTURAL VARIANT CALLER



**SMRT Analysis**

### Create New Analysis - Settings

**Name \***  
HG00733 10-fold SV

**Analysis Application \***  
Structural Variant Calling [Beta]

**References \***  
hg38

**Data Sets**

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

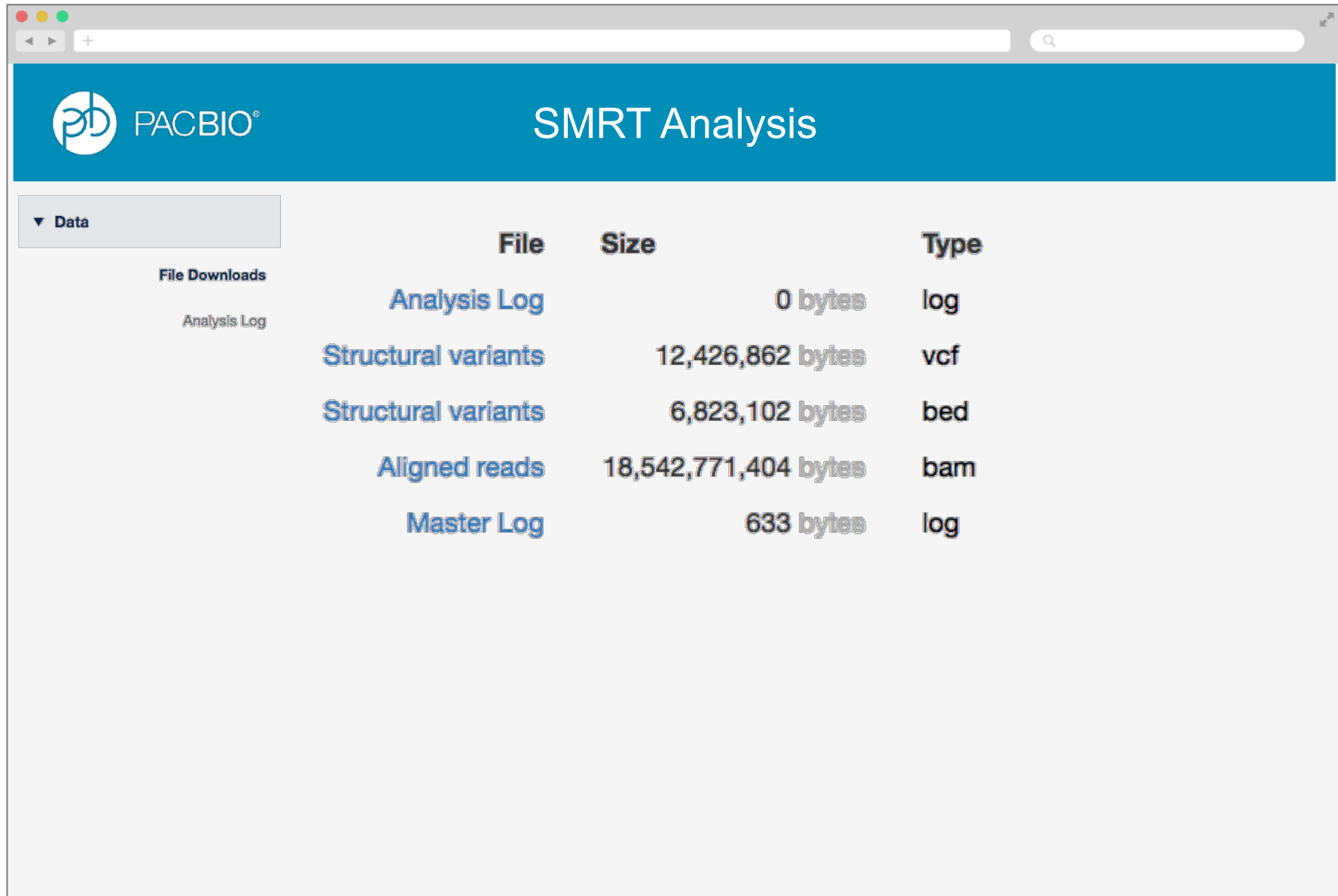
**Structural Variants**

minimum reads that support variant (count) ⓘ  
2

minimum reads that support variant (percent) ⓘ  
0.2

**CANCEL** **START**

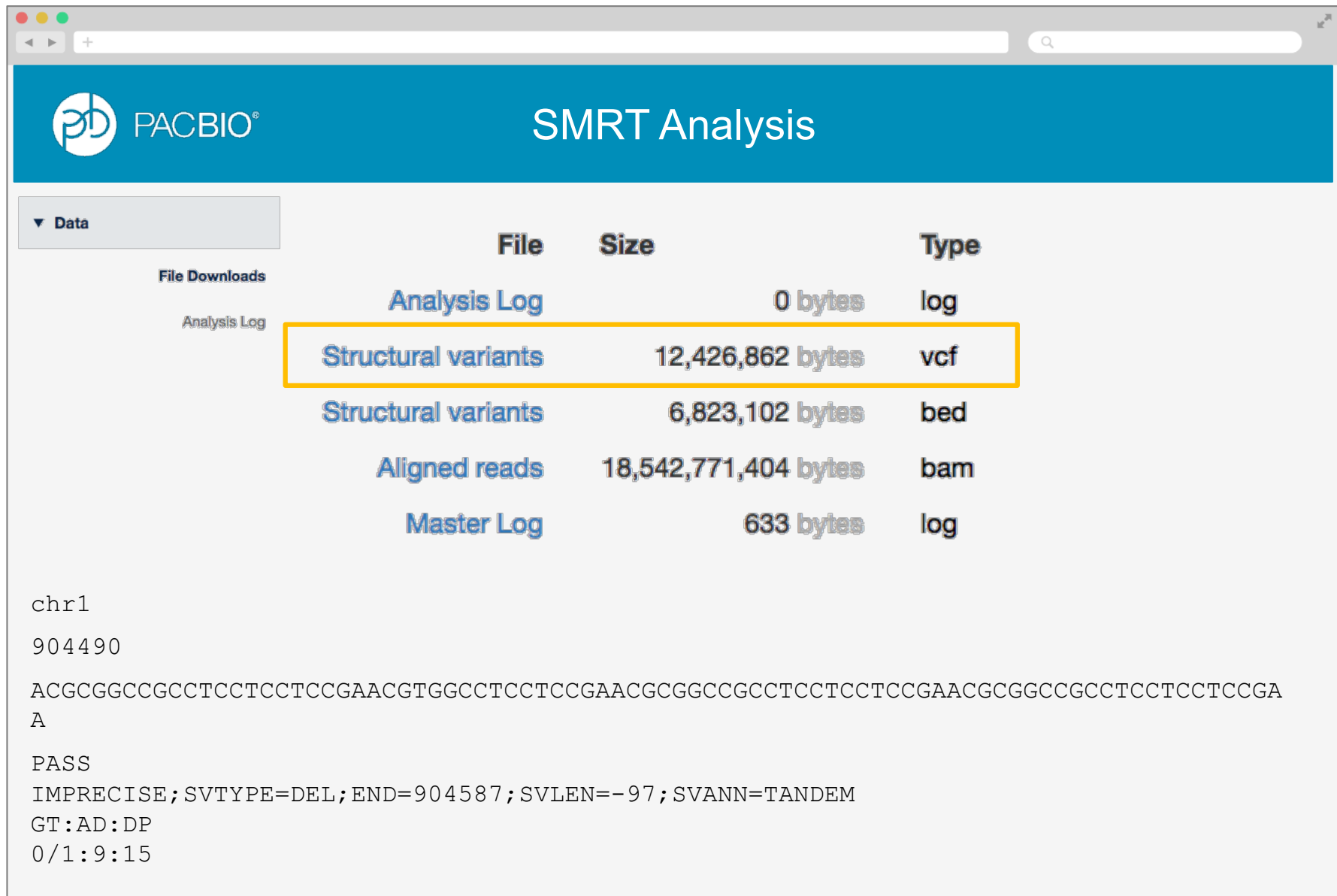
# PBSV: SMRT LINK STRUCTURAL VARIANT CALLER



The screenshot shows a web browser window displaying the 'SMRT Analysis' page. The page has a blue header with the PACBIO logo and the title 'SMRT Analysis'. Below the header, there is a 'Data' section with a 'File Downloads' sub-section. A table lists the following files:

File	Size	Type
Analysis Log	0 bytes	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

# PBSV: SMRT LINK STRUCTURAL VARIANT CALLER



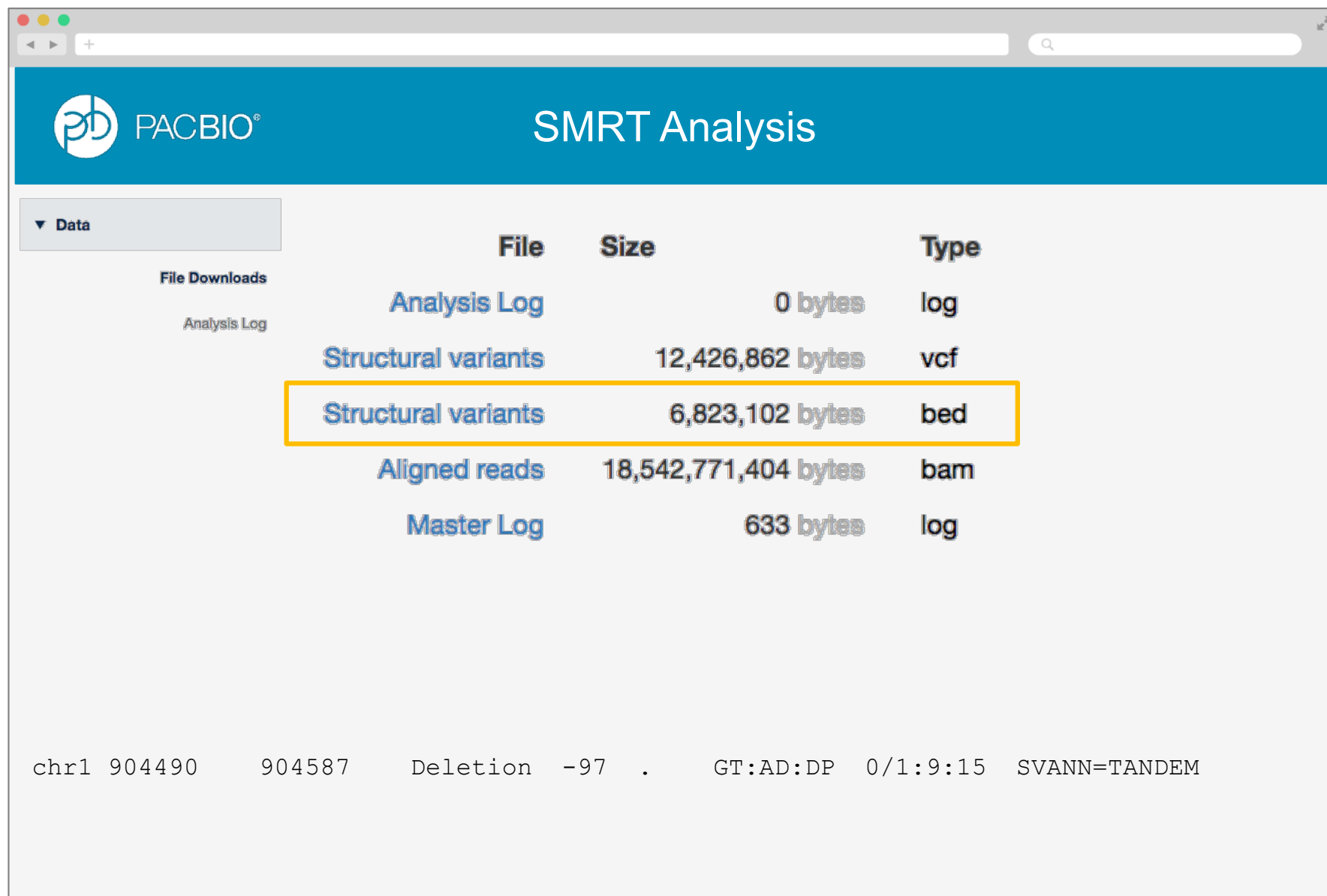
The screenshot shows the SMRT Analysis web interface. At the top, there is a blue header with the PACBIO logo and the text "SMRT Analysis". Below the header, there is a "Data" dropdown menu. Underneath, there is a "File Downloads" section with a sub-section for "Analysis Log". A table lists several files:

File	Size	Type
Analysis Log	0 bytes	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

The "Structural variants" file (12,426,862 bytes, vcf type) is highlighted with a yellow border. Below the table, there is a text block showing a variant call for chromosome 1 at position 904490:

```
chr1
904490
ACGCGGCCGCCTCCTCCTCCGAACGTGGCCTCCTCCGAACGCGGCCGCCTCCTCCTCCGAACGCGGCCGCCTCCTCCTCCGA
A
PASS
IMPRECISE;SVTYPE=DEL;END=904587;SVLEN=-97;SVANN=TANDEM
GT:AD:DP
0/1:9:15
```

# PBSV: SMRT LINK STRUCTURAL VARIANT CALLER



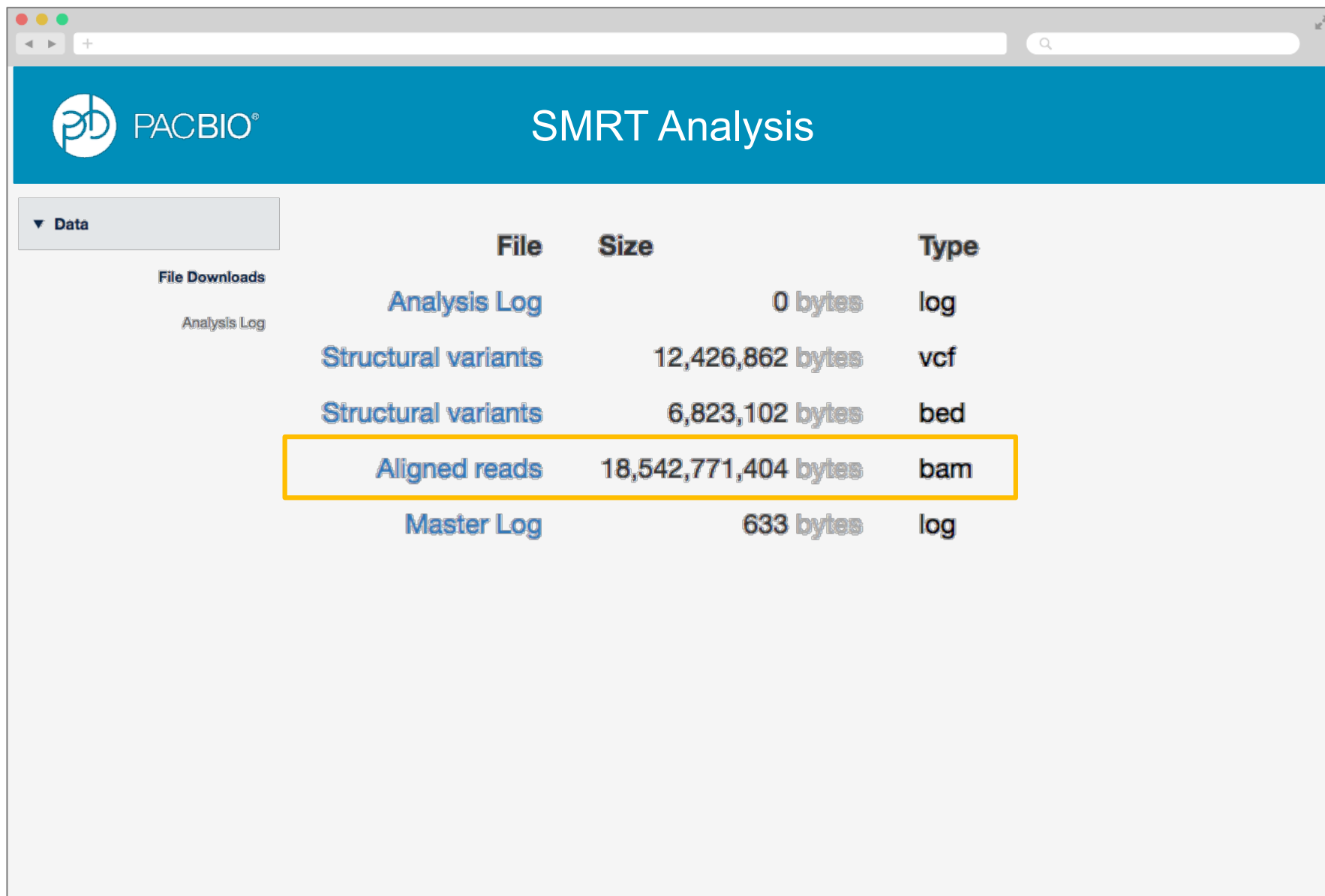
The screenshot shows the SMRT Analysis web interface. The top navigation bar includes the PACBIO logo and the text "SMRT Analysis". A sidebar on the left has a "Data" dropdown menu and a "File Downloads" section containing "Analysis Log". The main content area displays a table of files:

File	Size	Type
Analysis Log	0 bytes	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

Below the table, a structural variant call is displayed in a text format:

```
chr1 904490 904587 Deletion -97 . GT:AD:DP 0/1:9:15 SVANN=TANDEM
```

# PBSV: SMRT LINK STRUCTURAL VARIANT CALLER

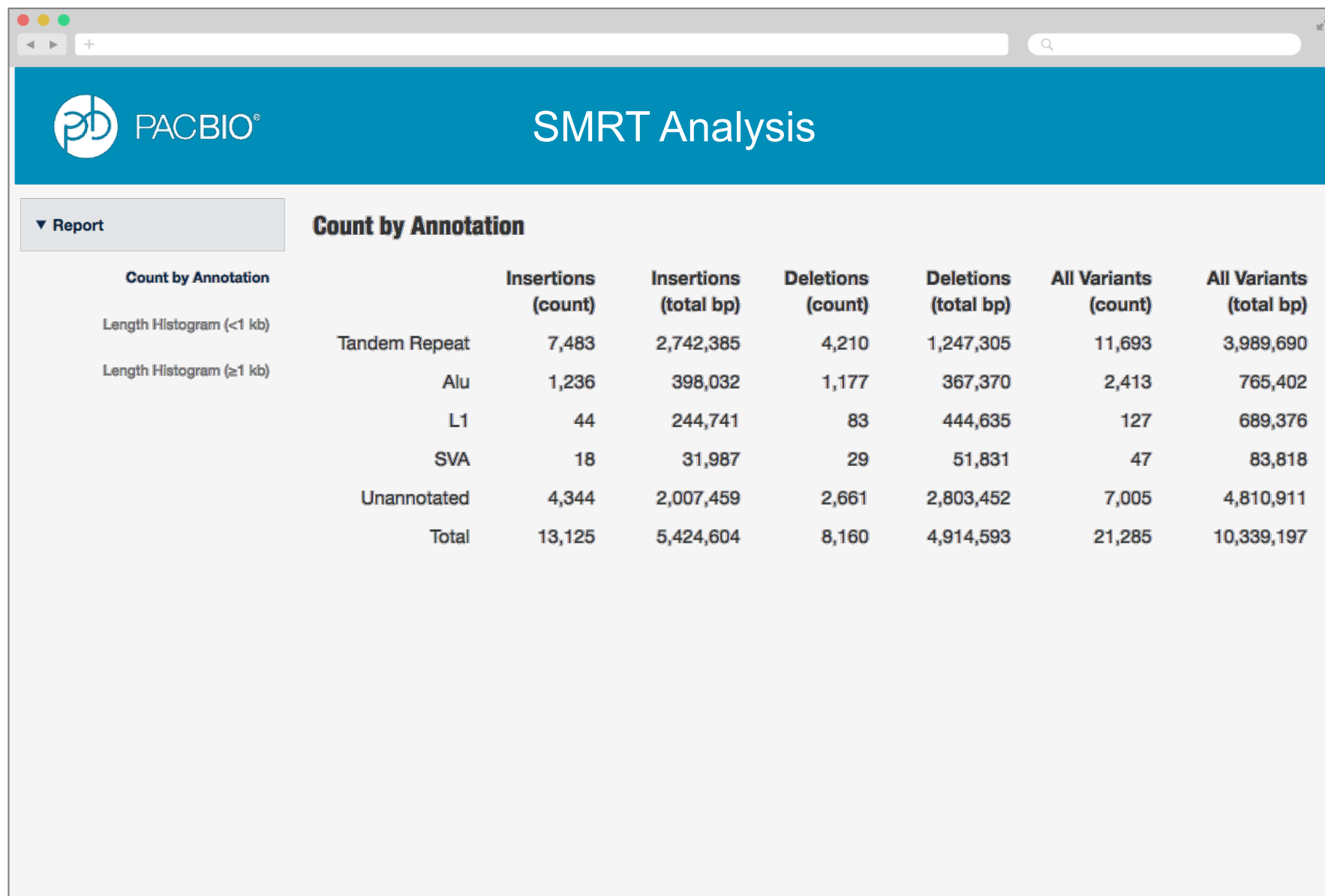


The screenshot shows the SMRT Analysis web interface. The header includes the PACBIO logo and the text "SMRT Analysis". A sidebar on the left has a "Data" dropdown menu and a "File Downloads" section containing "Analysis Log". The main content area displays a table of files:

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



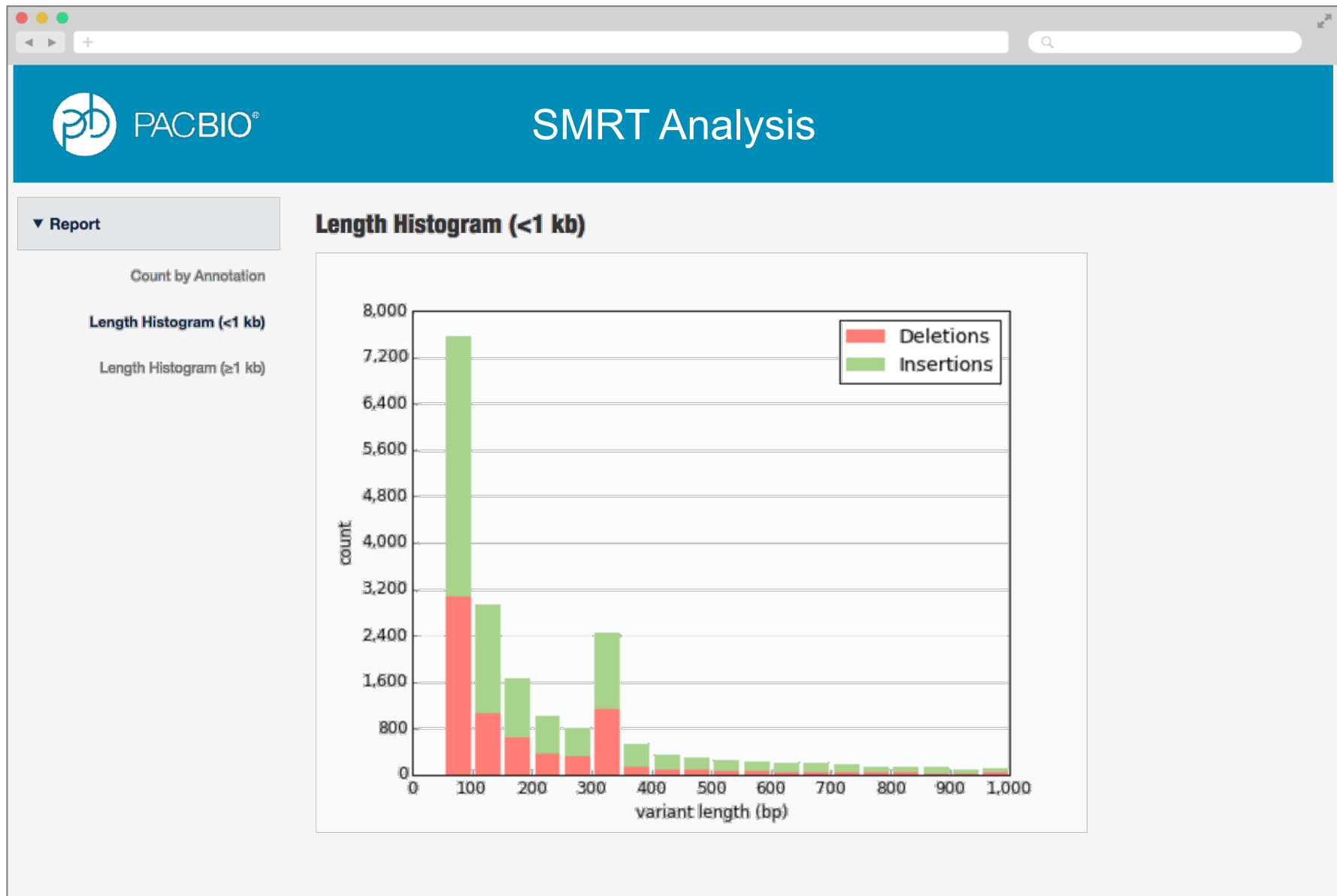
# PBSV: SMRT LINK STRUCTURAL VARIANT CALLER



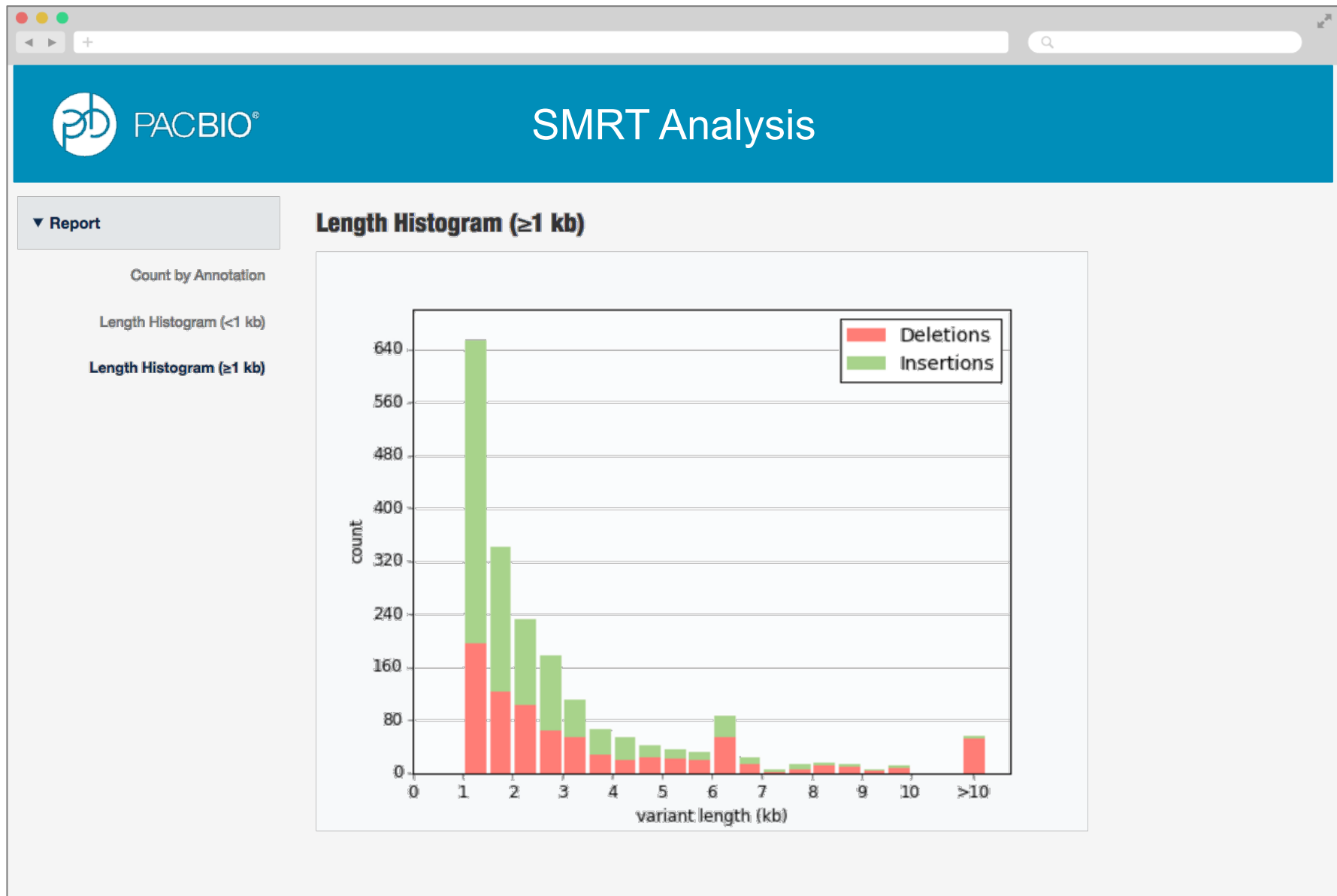
The screenshot shows a web browser window with the PACBIO logo and the title 'SMRT Analysis'. A sidebar on the left contains a 'Report' dropdown menu. The main content area displays a table titled 'Count by Annotation' with the following data:

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	11,693	3,989,690
Length Histogram (≥1 kb)	Alu	1,236	398,032	1,177	367,370	2,413	765,402
	L1	44	244,741	83	444,635	127	689,376
	SVA	18	31,987	29	51,831	47	83,818
	Unannotated	4,344	2,007,459	2,661	2,803,452	7,005	4,810,911
	<b>Total</b>	<b>13,125</b>	<b>5,424,604</b>	<b>8,160</b>	<b>4,914,593</b>	<b>21,285</b>	<b>10,339,197</b>

# PBSV: SMRT LINK STRUCTURAL VARIANT CALLER



# PBSV: SMRT LINK STRUCTURAL VARIANT CALLER



## 3 COMPONENTS TO PBSV

**pbsv** command line utility for top-level commands

**pbsvutil** command line utility for detailed commands

**SMRT Link web interface**

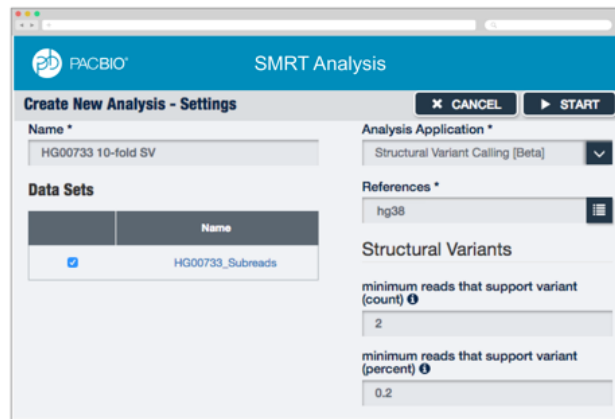
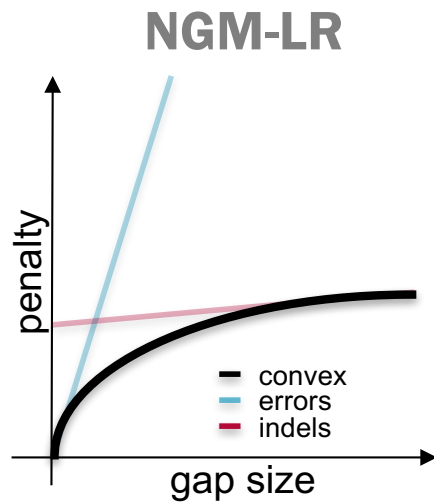
# ACKNOWLEDGMENTS

## Schatz Lab

Michael Schatz  
 Philipp Rescheneder  
 Fritz Sedlazeck

## PacBio

Yuan Li      Nat Echols  
 Chris Dunn    Aaron Klammer  
 Ben Lerch      Mary Budagyan  
 Jim Drake





PACBIO®

[www.pacb.com](http://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.