



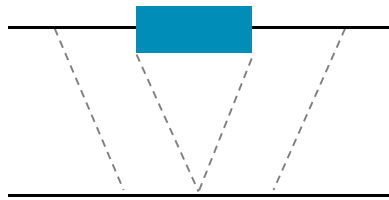
# Structural Variant Detection in SMRT Link 5 with pbsv

Aaron Wenger

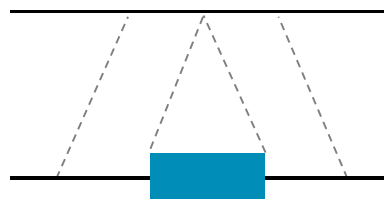
2017-06-27

# STRUCTURAL VARIANT = DIFFERENCE $\geq 50$ BP

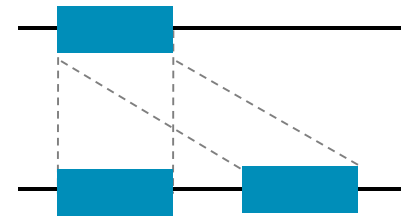
## Deletion



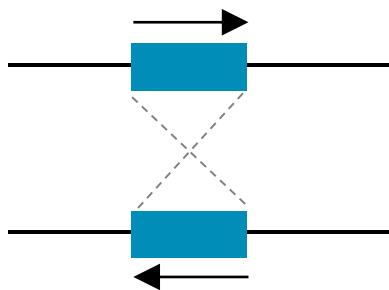
## Insertion



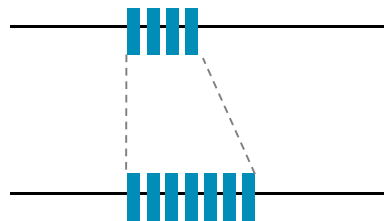
## Duplication



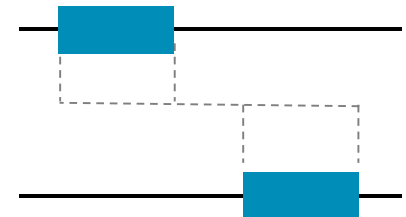
## Inversion



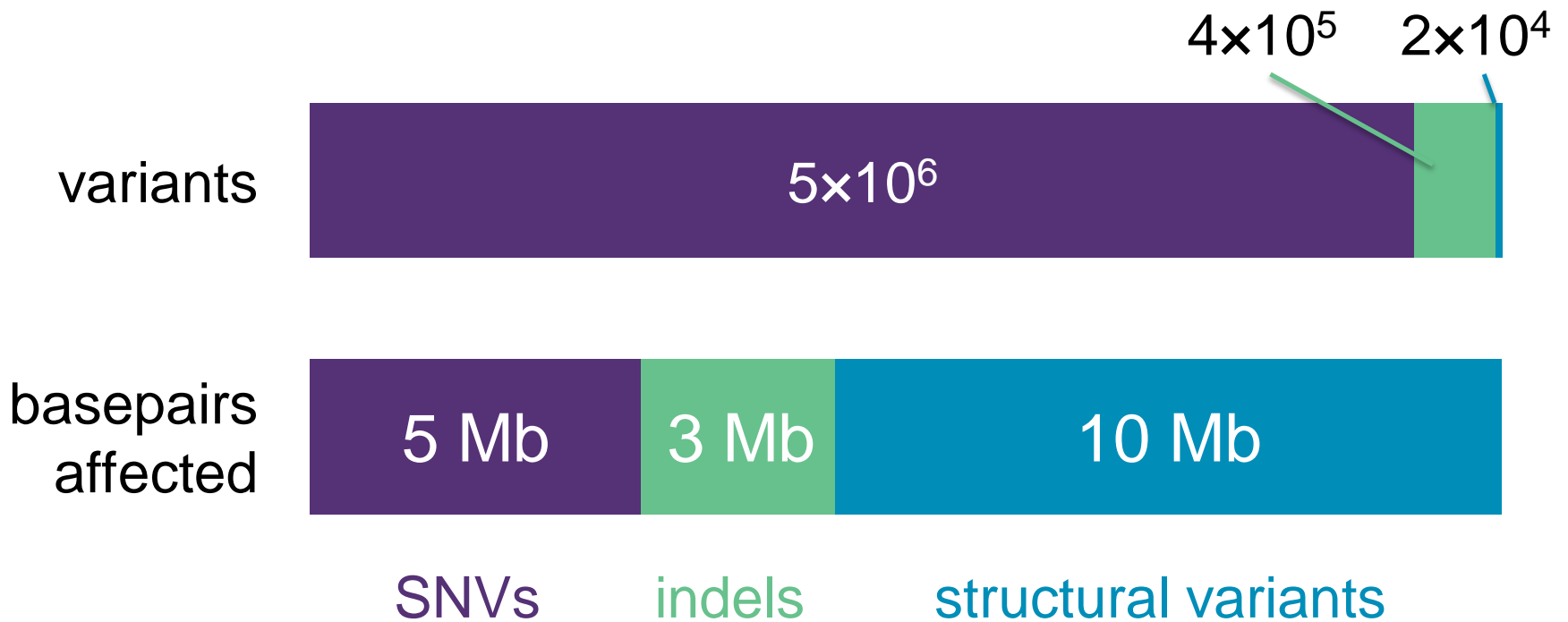
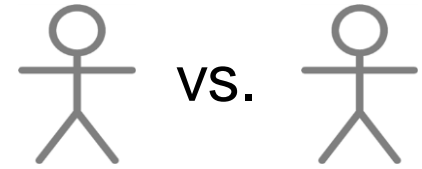
## Tandem Repeat



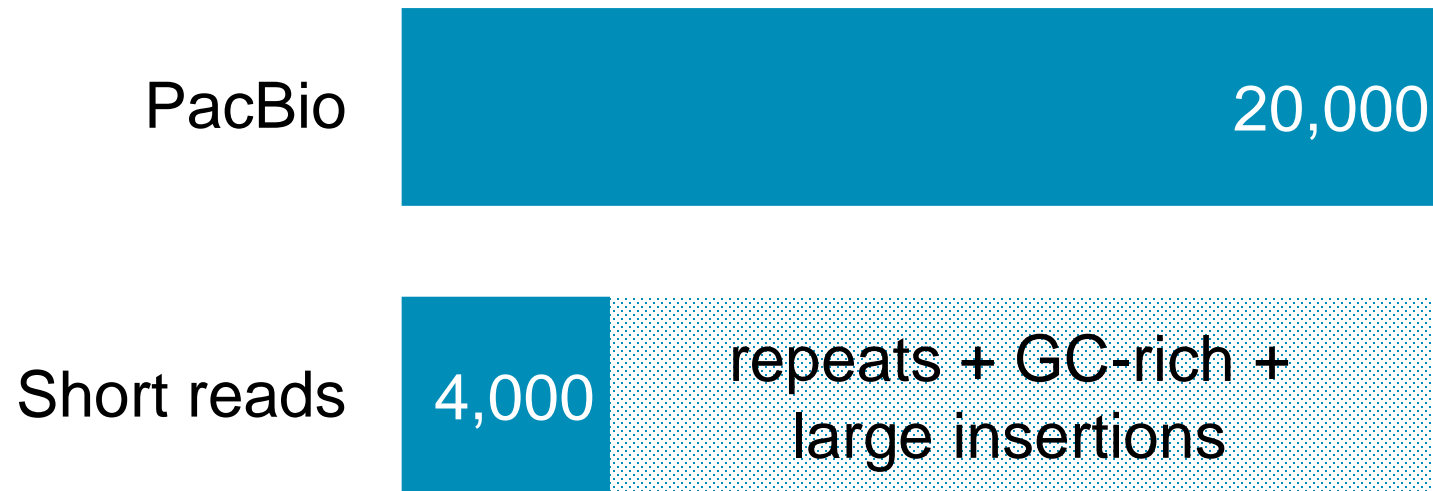
## Translocation



# VARIATION BETWEEN TWO HUMAN GENOMES



# STRUCTURAL VARIANTS DETECTED IN A HUMAN GENOME



## SEQUENCING + ANALYSIS



**pbsv**



Structural  
Variants

Short  
reads



**BWA**



SNVs +  
Indels

## 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.

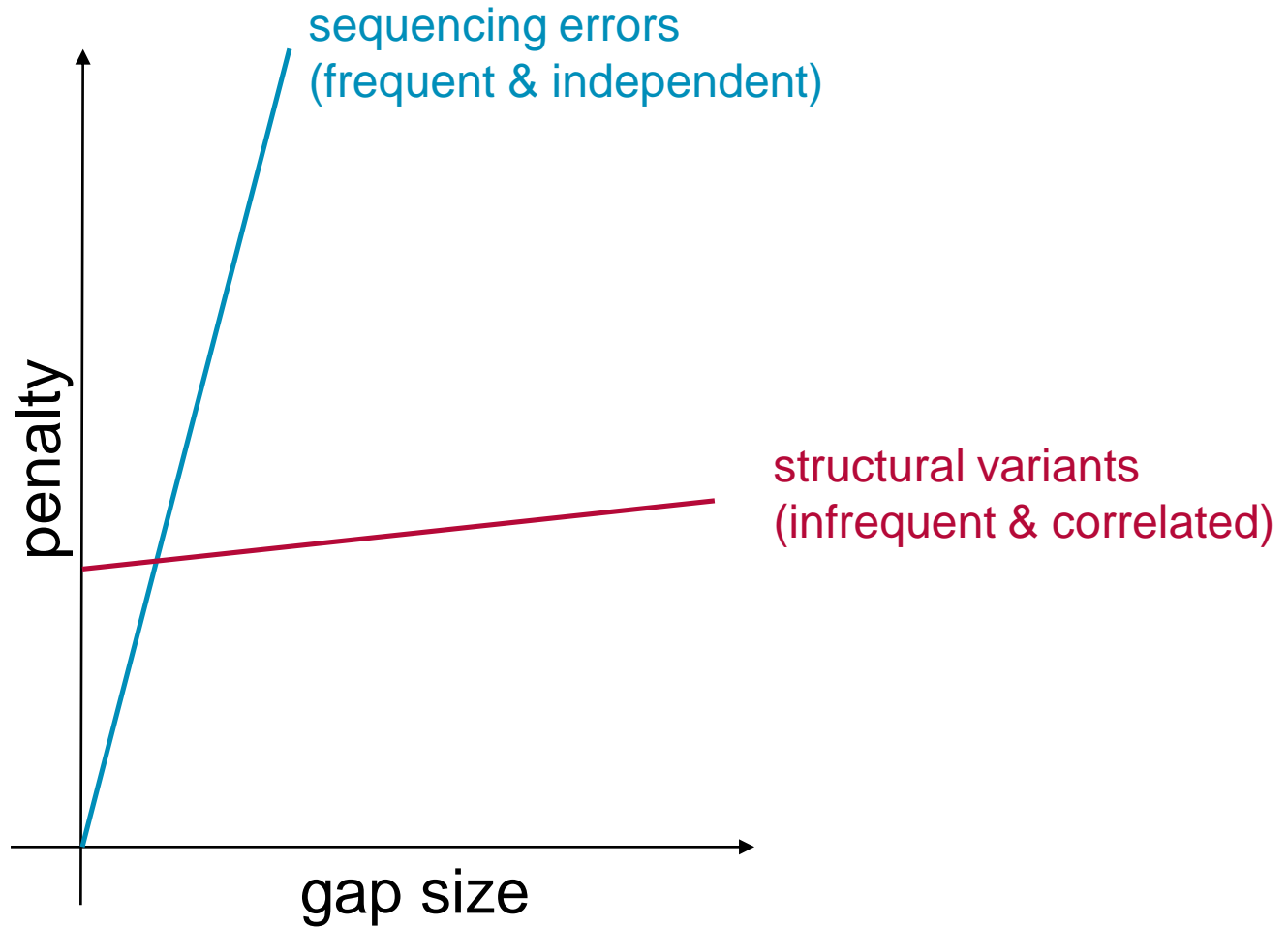
```
pbsv call [-h] [--cfg_fn sv.cfg]
  ref.fa ref.align.bam ref.sv.bed|vcf
```

Call structural variants from aligned reads.



# PBSV ALIGN UTILIZES NGM-LR

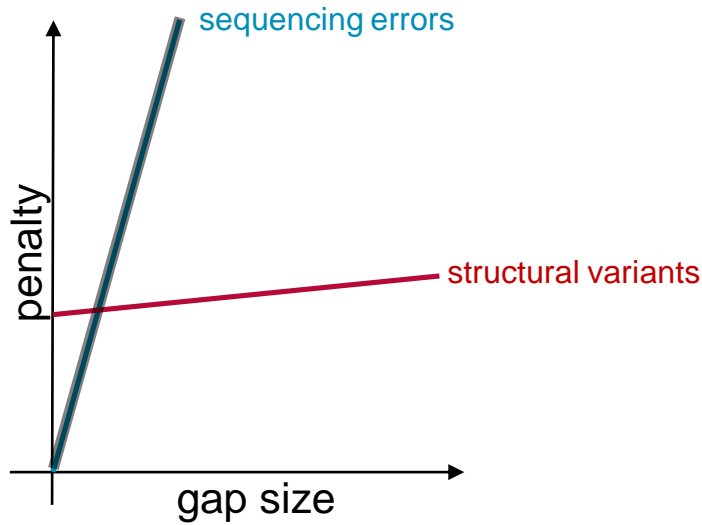
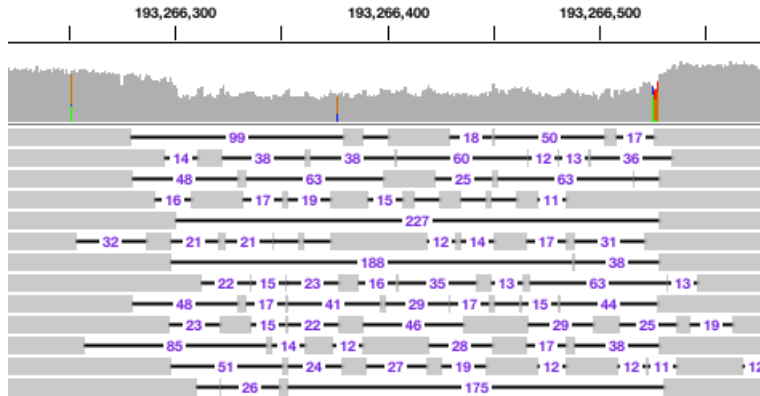
pbsvutil ngmlr



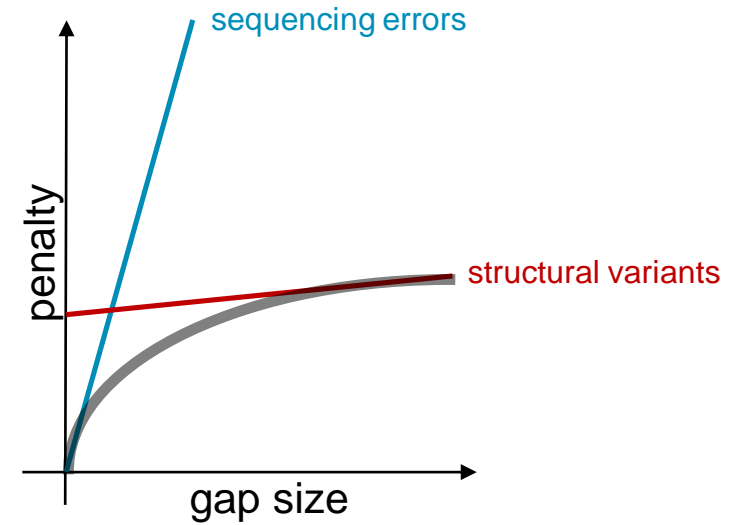
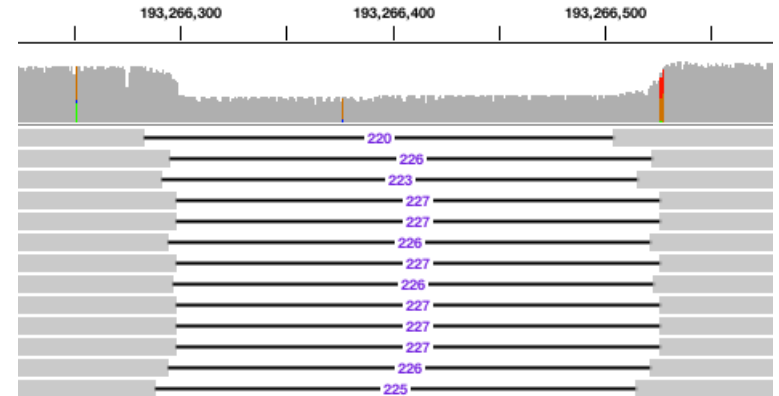
# PBSV ALIGN UTILIZES NGM-LR

pbsvutil ngmlr

## BWA

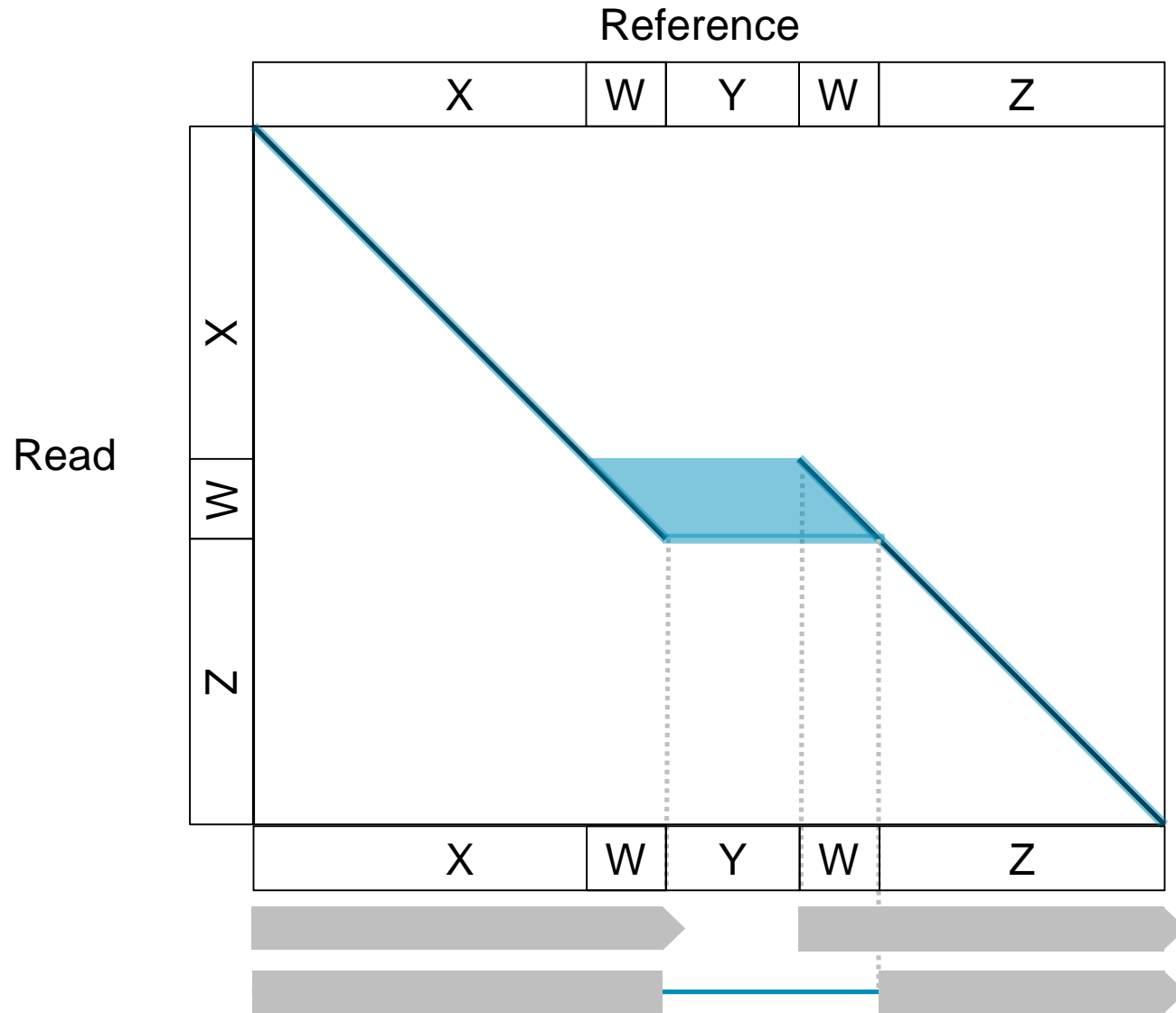


## NGM-LR



# 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.

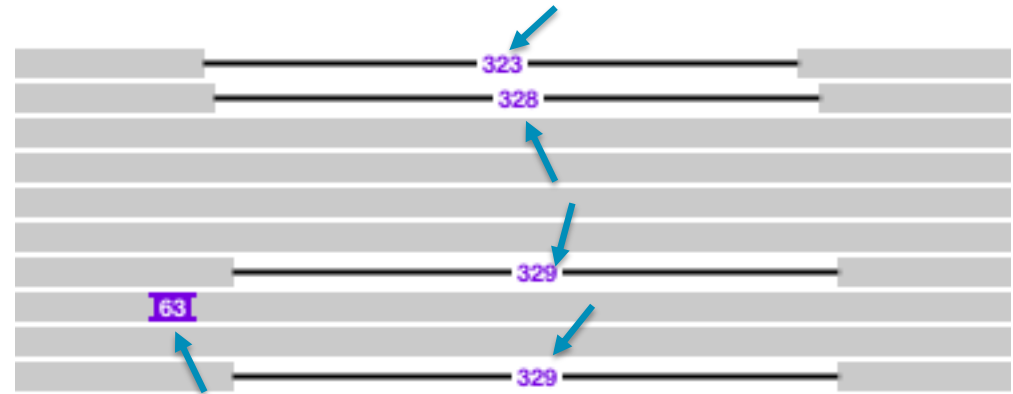
```
pbsv call [-h] [--cfg_fn sv.cfg]
    ref.fa ref.align.bam ref.sv.bed|vcf
```

Call structural variants from aligned reads.

# PBSV CALL: STAGED STRUCTURAL VARIANT CALLER



# PBSV CALL: STAGED STRUCTURAL VARIANT CALLER



# PBSV CALL: STAGED STRUCTURAL VARIANT CALLER

FIND SV SIGNATURES

CIGAR D & I  
 $\geq 50$  bp

CLUSTER SV SIGNATURES

nearby with similar sequence

SUMMARIZE INTO SV

consensus of supporting reads

GENOTYPE SV

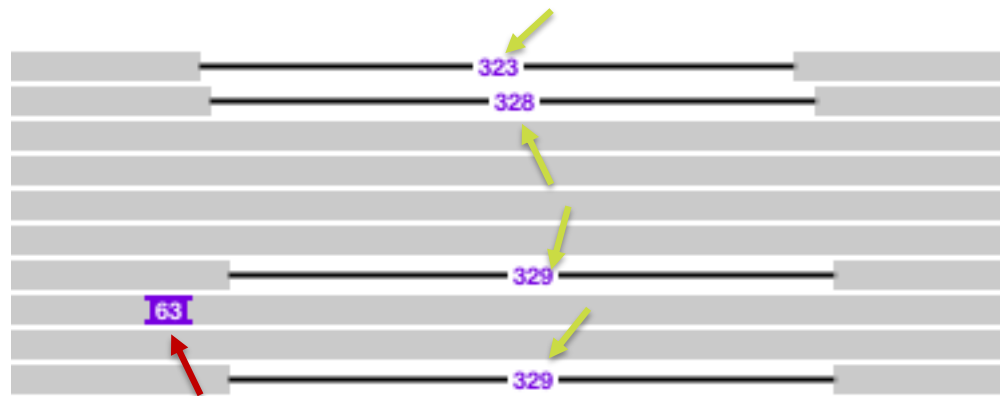
supporting reads / covering reads

ANNOTATE SV

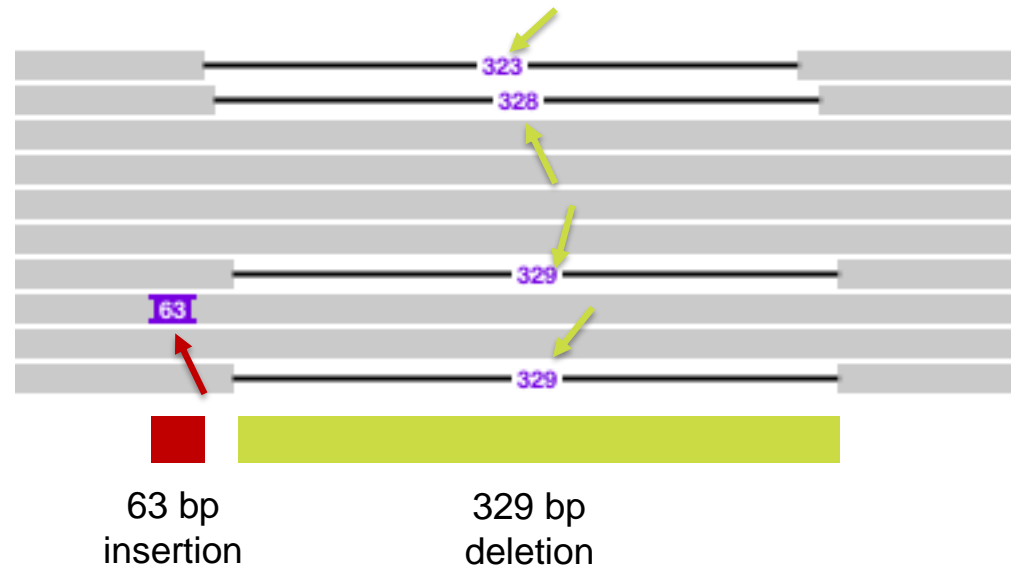
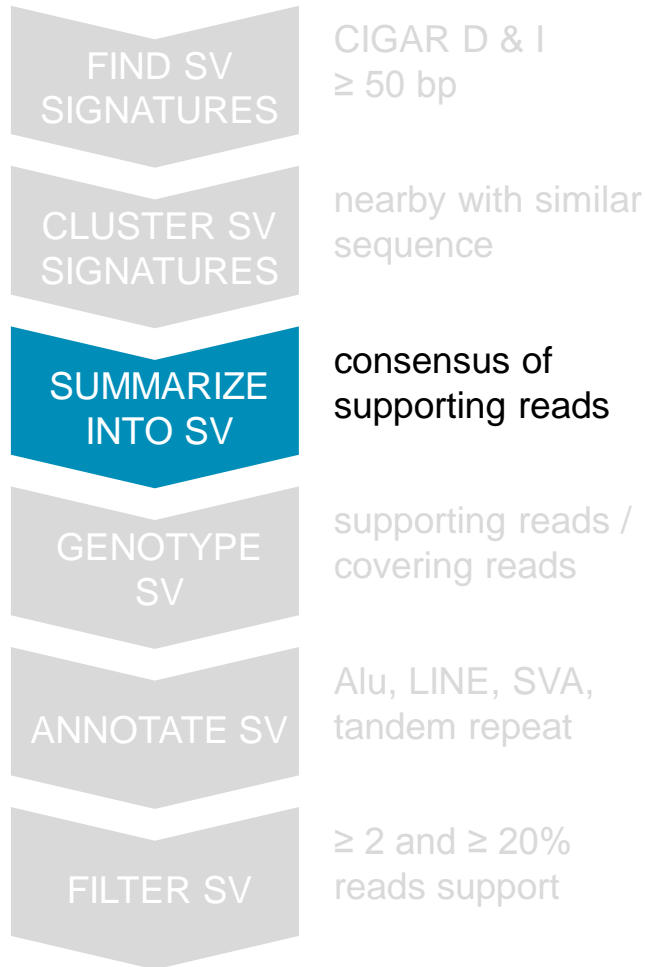
Alu, LINE, SVA, tandem repeat

FILTER SV

$\geq 2$  and  $\geq 20\%$  reads support

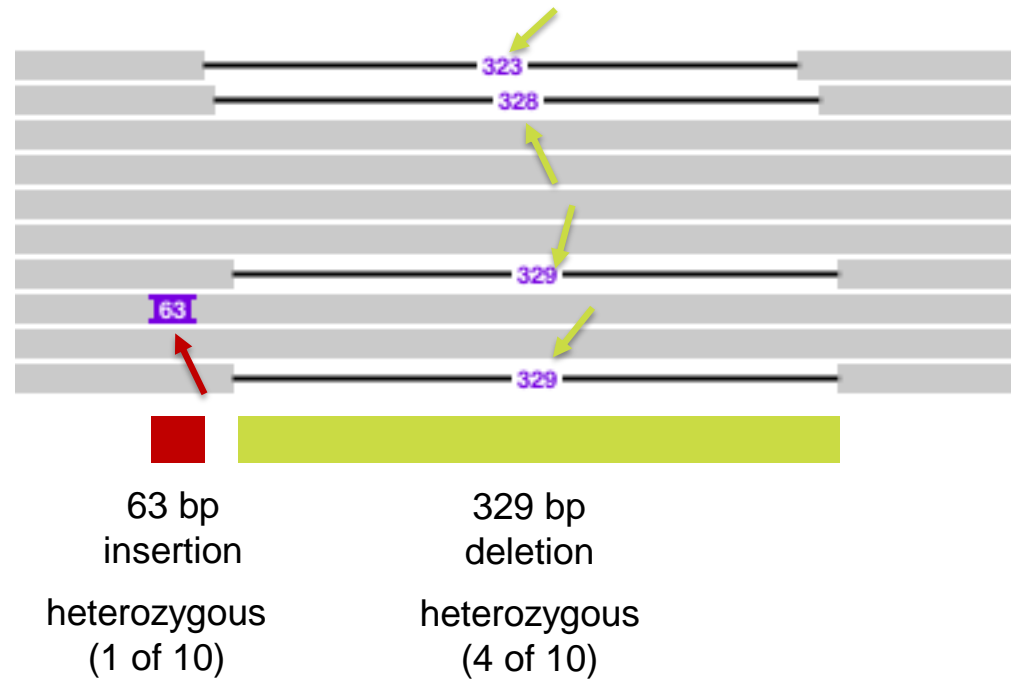
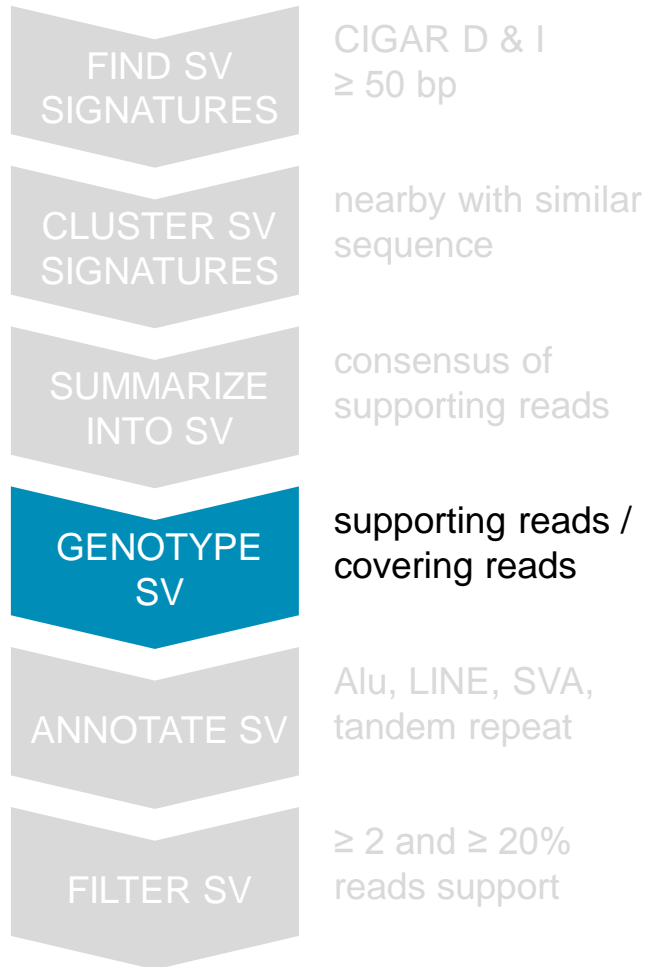


# PBSV CALL: STAGED STRUCTURAL VARIANT CALLER

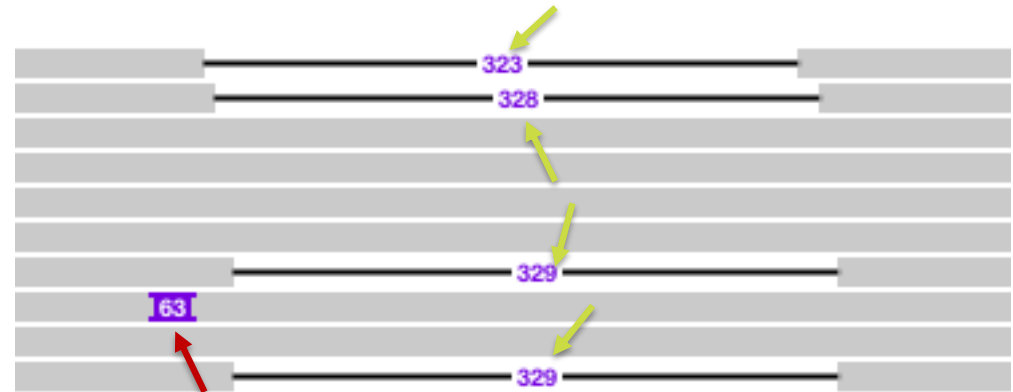
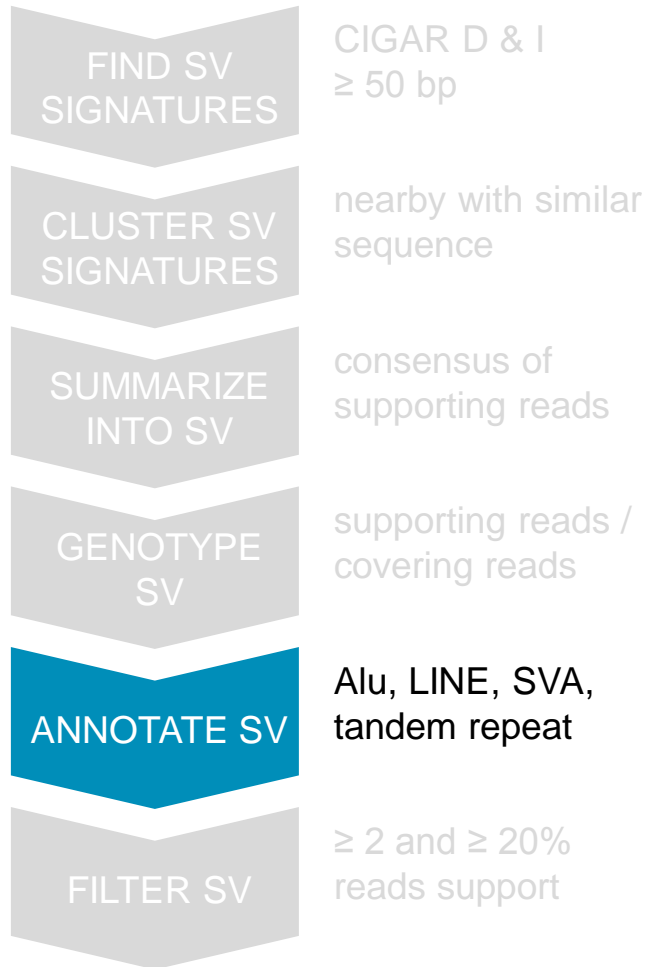




# PBSV CALL: STAGED STRUCTURAL VARIANT CALLER

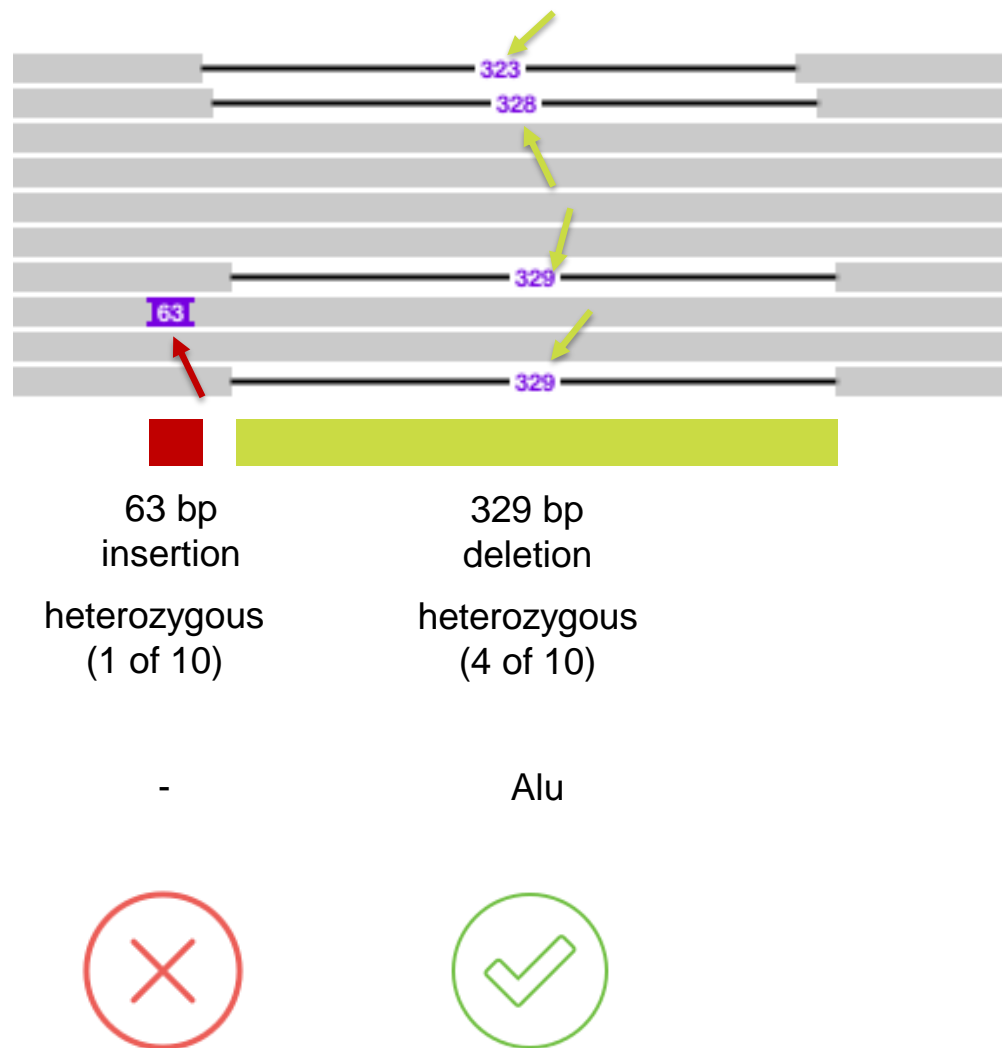


# PBSV CALL: STAGED STRUCTURAL VARIANT CALLER




63 bp insertion	329 bp deletion
heterozygous (1 of 10)	heterozygous (4 of 10)
-	Alu

# PBSV CALL: STAGED STRUCTURAL VARIANT CALLER



# PBSV: SMRT LINK STRUCTURAL VARIANT CALLER


SMRT Analysis

**Create New Analysis - Settings**

✕ CANCEL
▶ START

**Name \***

**Analysis Application \***

Structural Variant Calling [Beta]
▼

**Data Sets**

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

**References \***

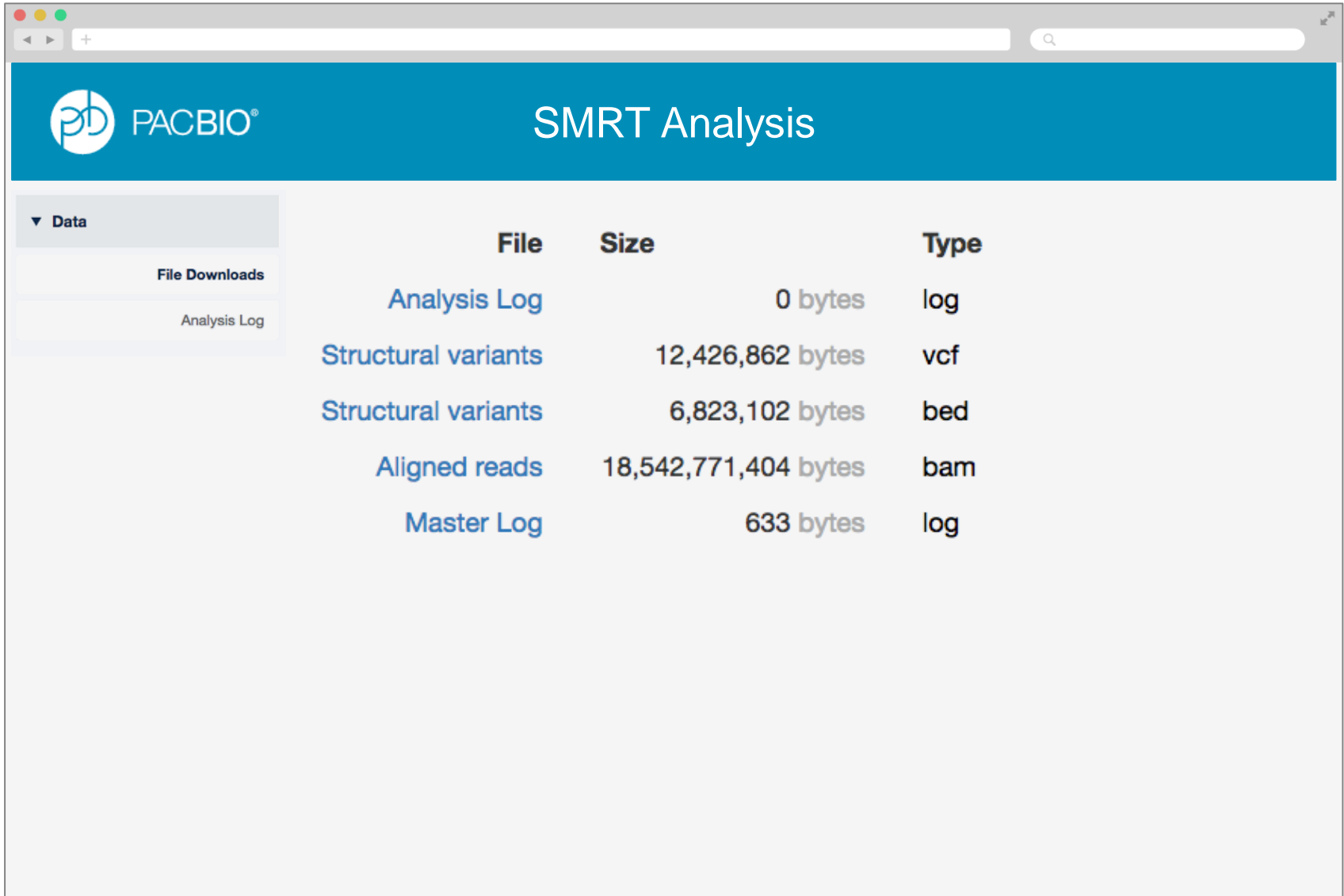
hg38
☰

**Structural Variants**

minimum reads that support variant (count) ⓘ

minimum reads that support variant (percent) ⓘ

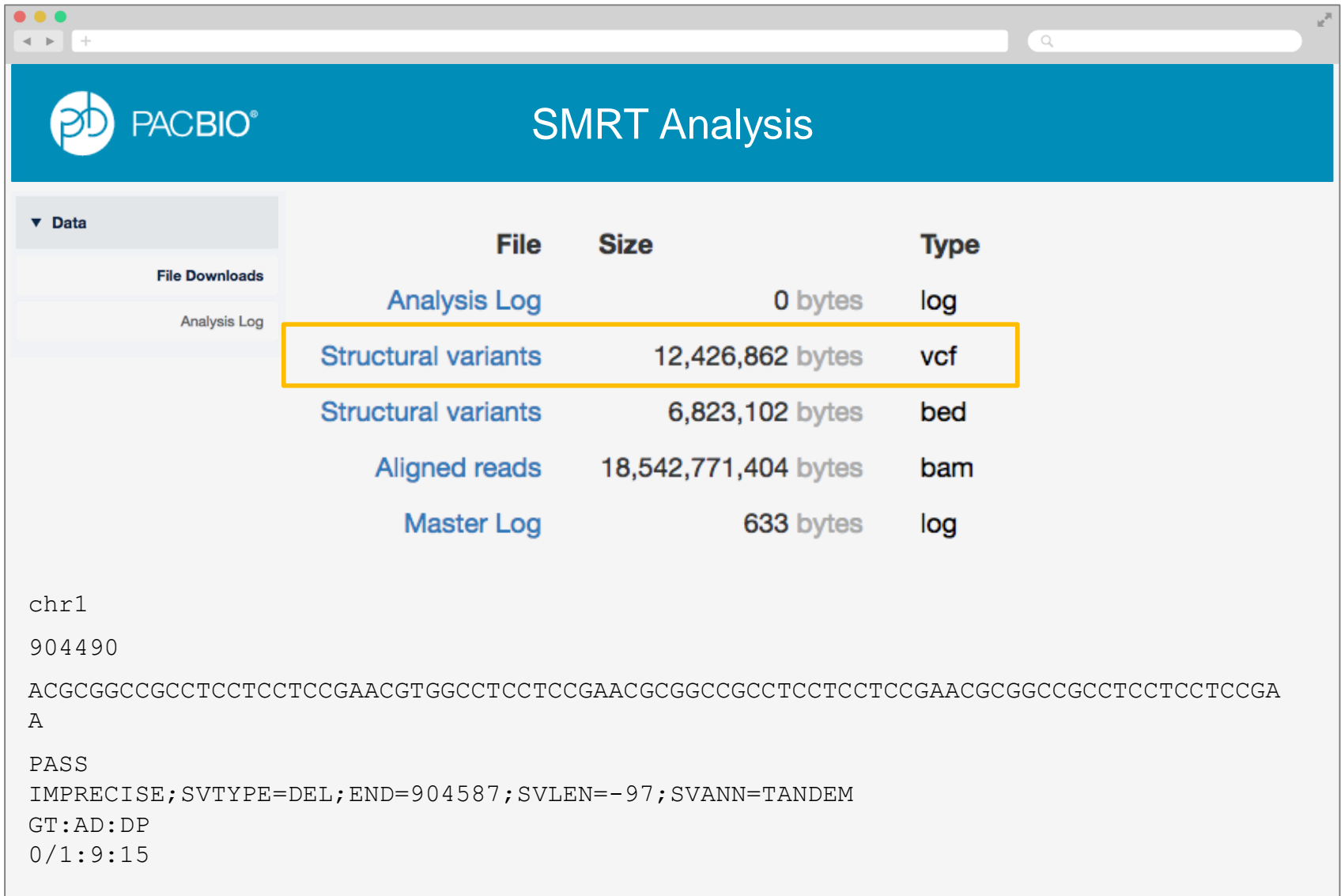
# PBSV: SMRT LINK STRUCTURAL VARIANT CALLER



The screenshot shows a web browser window displaying the 'SMRT Analysis' interface. The page has a blue header with the PACBIO logo and the title 'SMRT Analysis'. On the left, there is a sidebar with a 'Data' dropdown menu and a 'File Downloads' section containing an 'Analysis Log' link. The main content area features a table with the following data:

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

# PBSV: SMRT LINK STRUCTURAL VARIANT CALLER



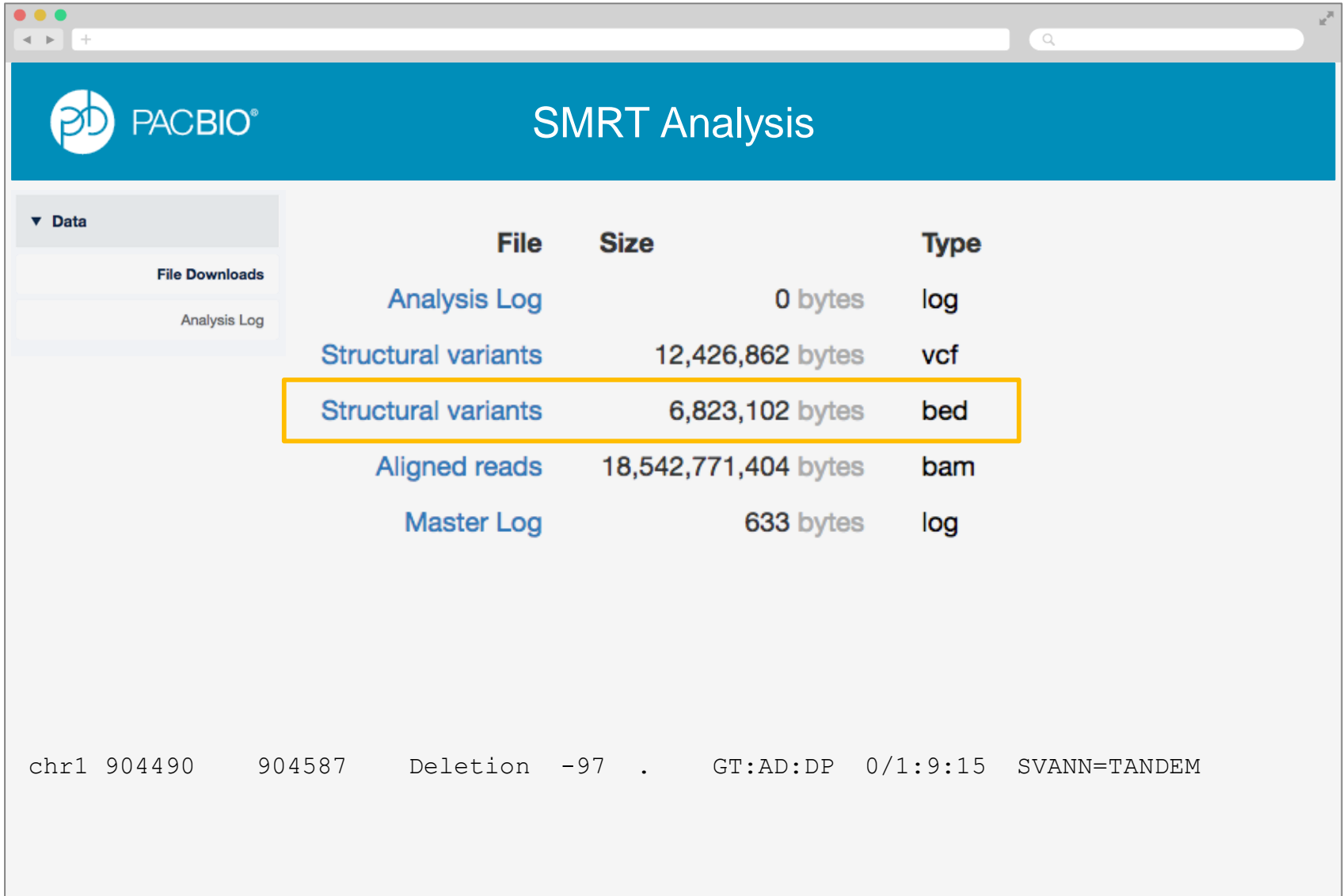
The screenshot shows the 'SMRT Analysis' web interface. On the left, there is a sidebar with a 'Data' dropdown menu containing 'File Downloads' and 'Analysis Log'. The main area displays a table of files:

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 shown in a text area:

```
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



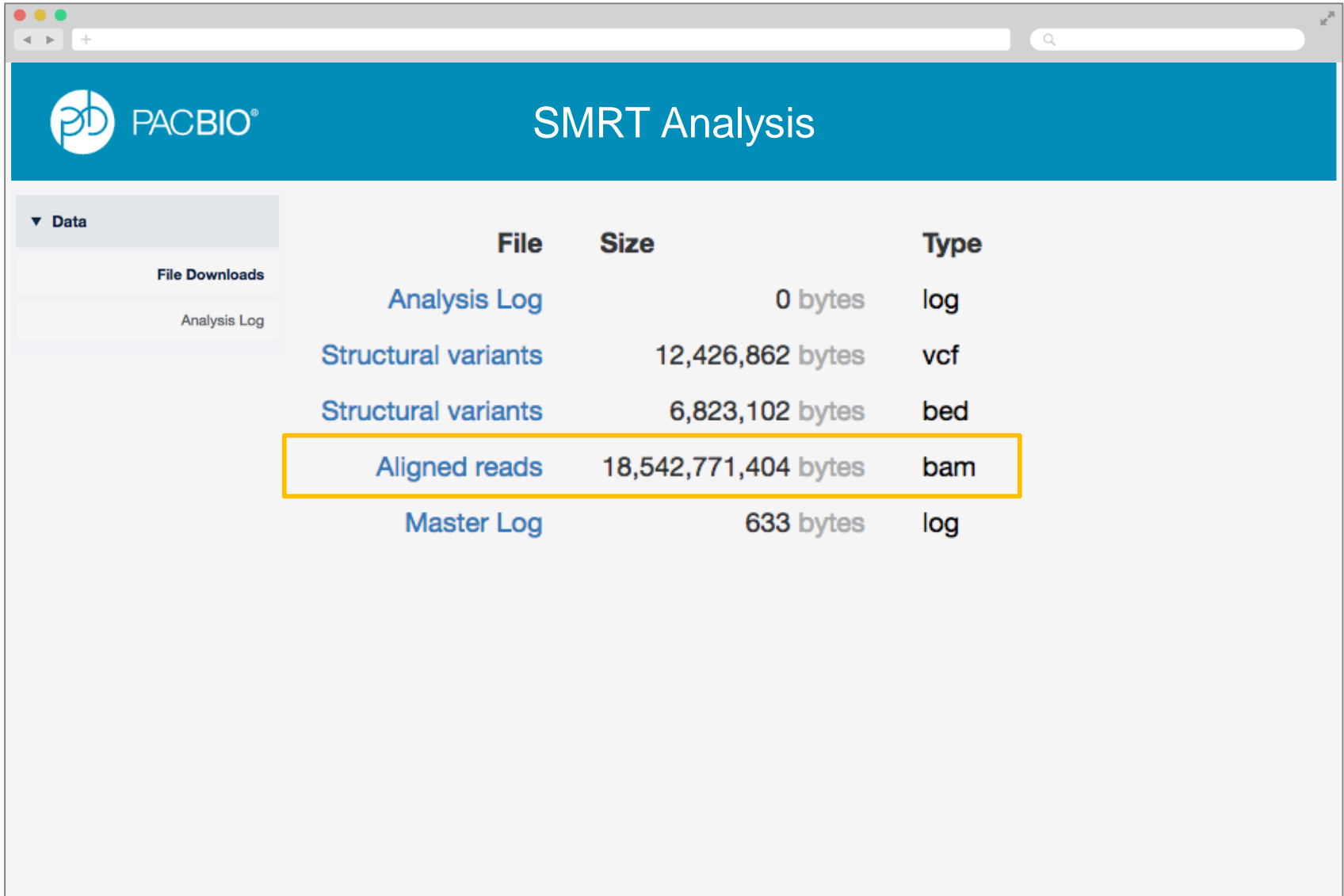
The screenshot shows the SMRT Analysis web interface. The top navigation bar includes the PACBIO logo and the title "SMRT Analysis". On the left, there is a sidebar with a "Data" dropdown menu containing "File Downloads" and "Analysis Log". The main content area displays a table of files:

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 monospaced font:

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

# PBSV: SMRT LINK STRUCTURAL VARIANT CALLER

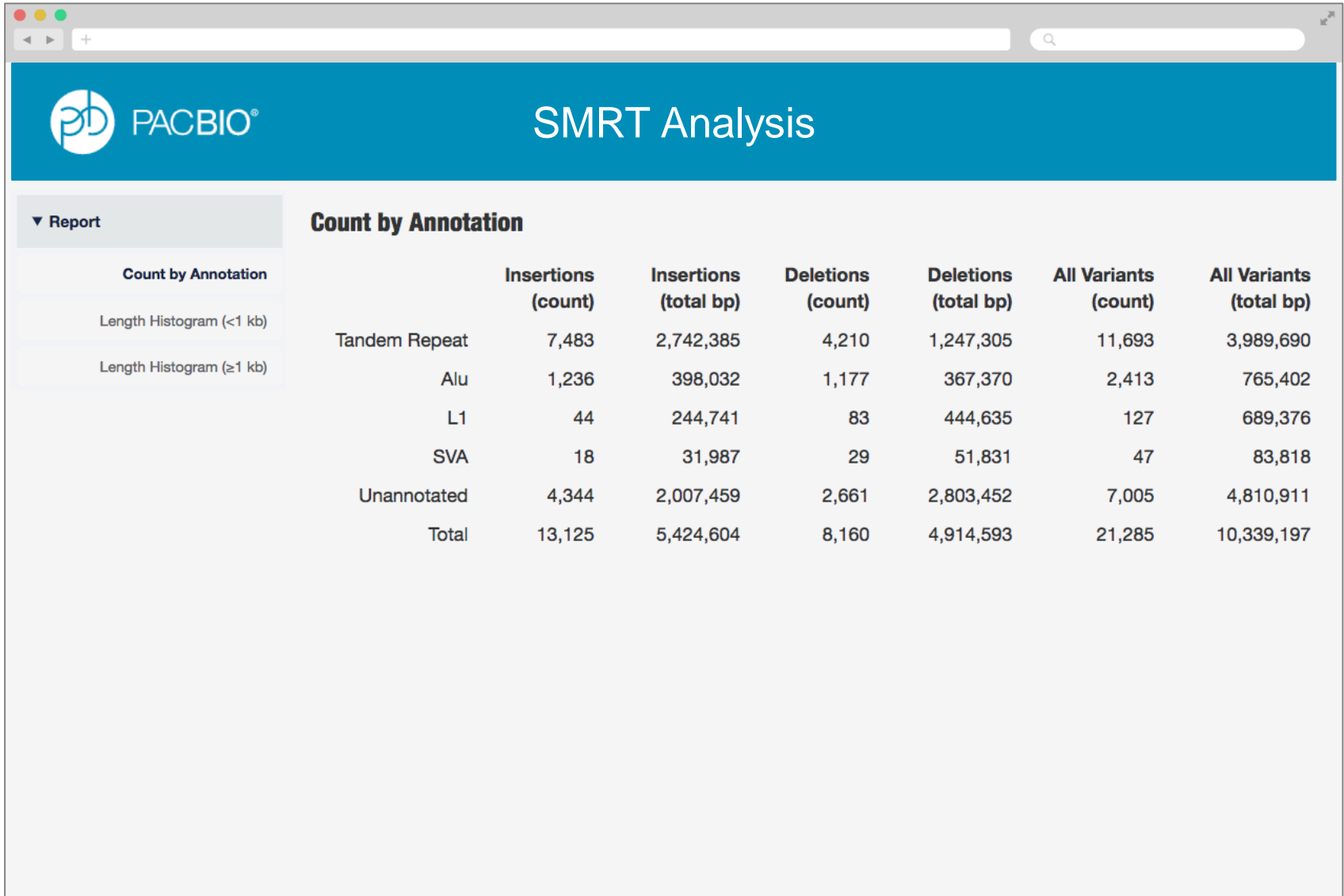


The screenshot shows a web browser window with the PACBIO logo and the title "SMRT Analysis". On the left, there is a sidebar with a "Data" dropdown menu containing "File Downloads" and "Analysis Log". The main content area displays a table of files with the following data:

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



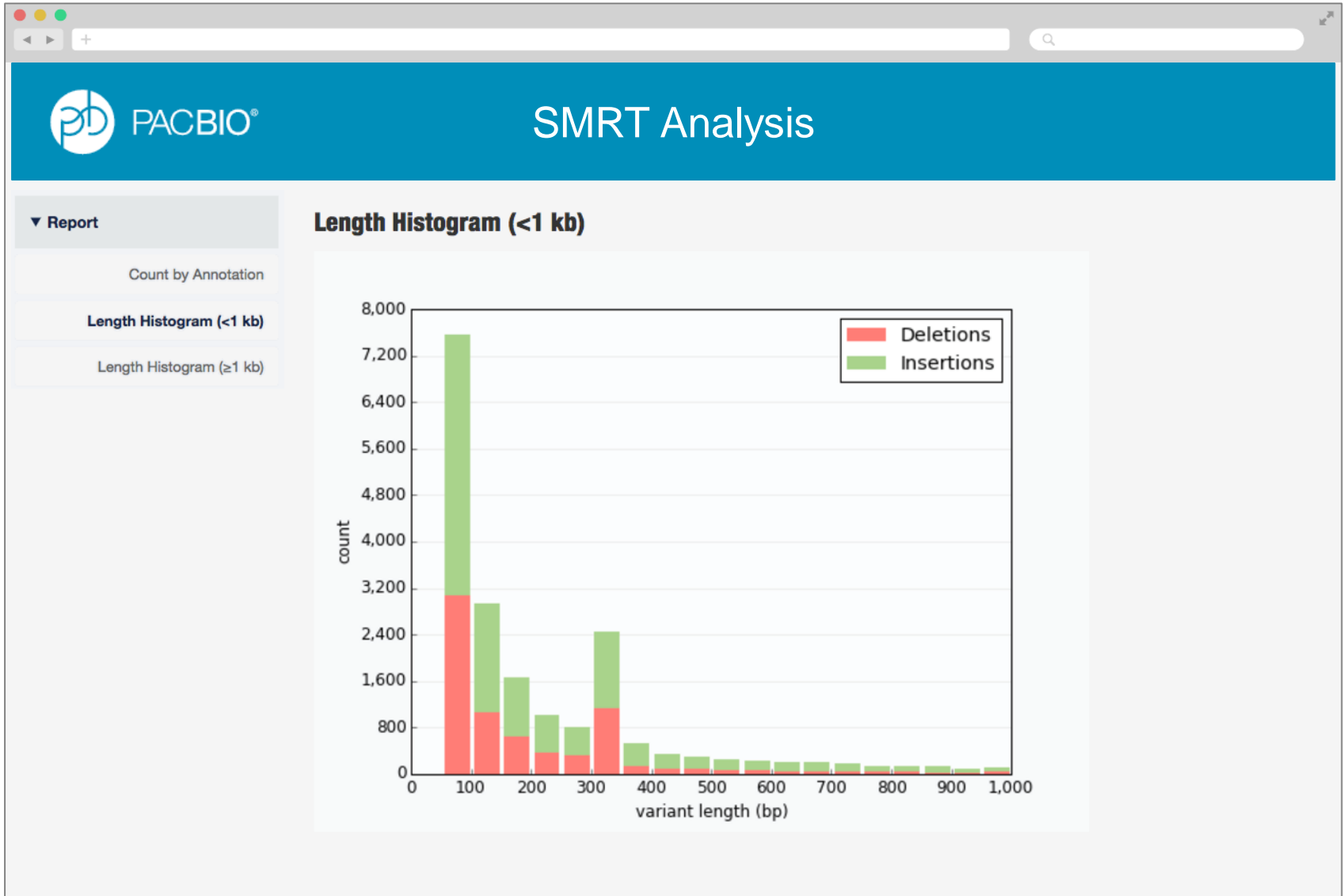
# PBSV: SMRT LINK STRUCTURAL VARIANT CALLER



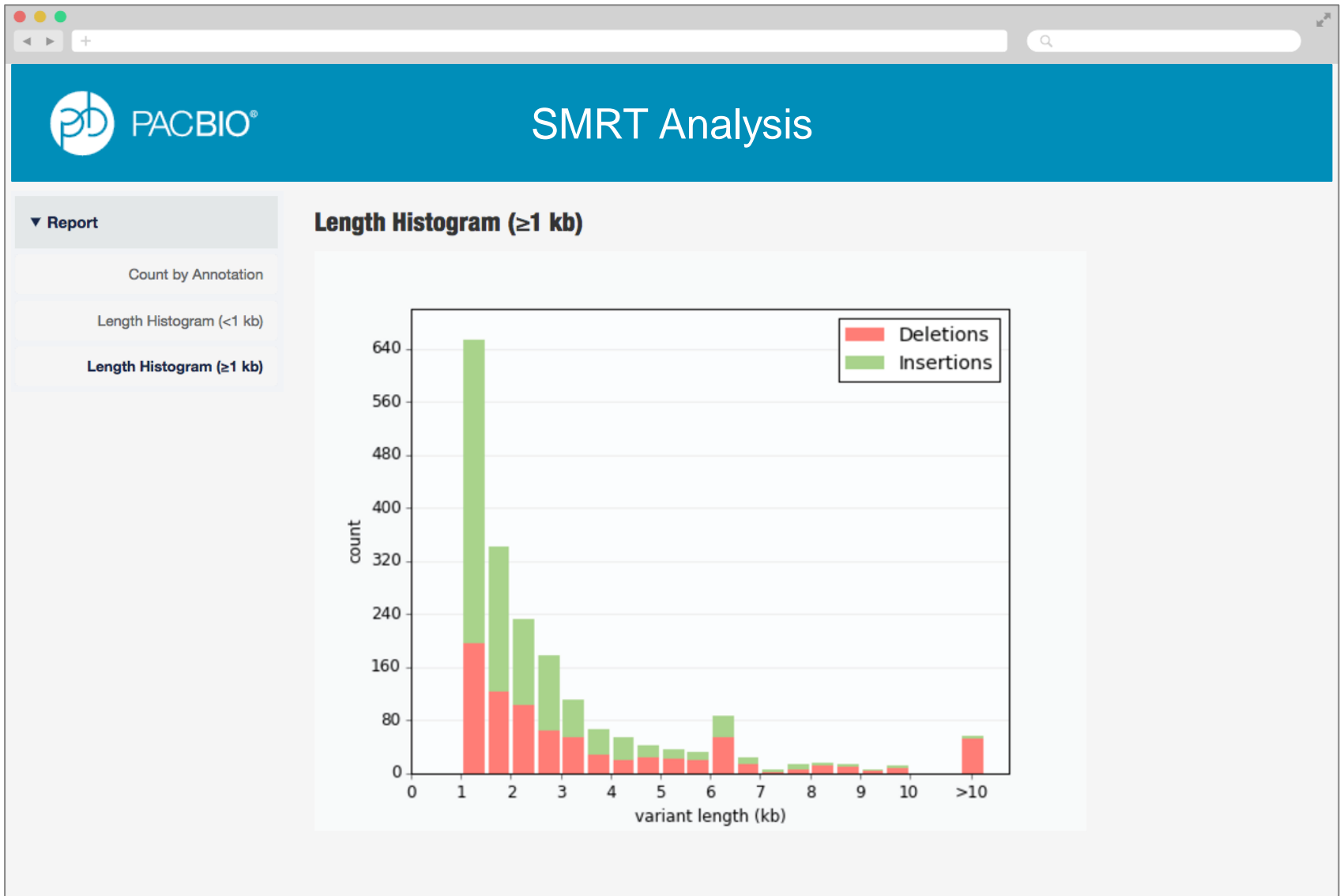
The screenshot shows a web browser window with the PACBIO logo and the title "SMRT Analysis". On the left, there is a "Report" dropdown menu with "Count by Annotation" selected. The main content area displays a table titled "Count by Annotation" with the following data:

	Insertions (count)	Insertions (total bp)	Deletions (count)	Deletions (total bp)	All Variants (count)	All Variants (total bp)
Tandem Repeat	7,483	2,742,385	4,210	1,247,305	11,693	3,989,690
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
Total	13,125	5,424,604	8,160	4,914,593	21,285	10,339,197

# 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**

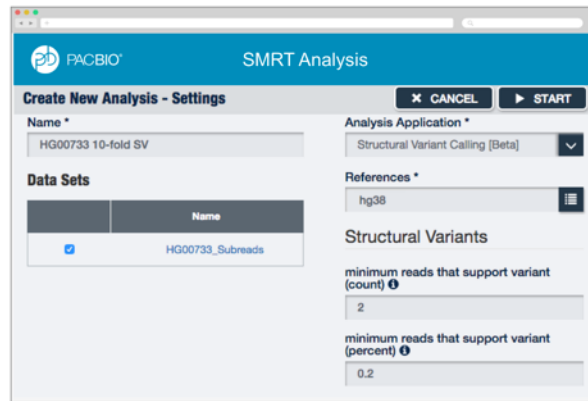
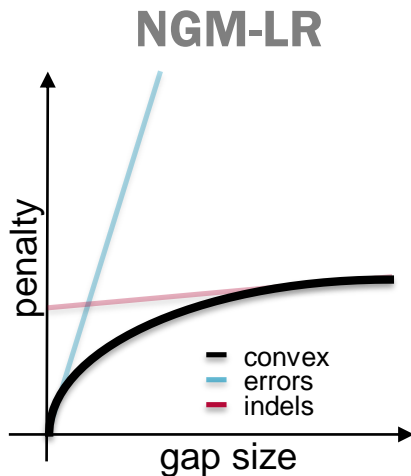
# ACKNOWLEDGMENTS

## Schatz Lab

Michael Schatz  
 Philipp Rescheneder  
 Fritz Sedlazeck

## PacBio

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





[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.