

Advanced Structural Variant Detection

Plugin

USER MANUAL

User manual for Advanced Structural Variant Detection (beta) 20.0

Windows, macOS and Linux

December 5, 2019

This software is for research purposes only.

QIAGEN Aarhus
Silkeborgvej 2
Prismet
DK-8000 Aarhus C
Denmark



Contents

1 Introduction	4
2 Advanced Structural Variant Detection	6
3 Install and uninstall plugins	10
3.1 Install	10
3.2 Uninstall	12
Bibliography	13

Chapter 1

Introduction

The Advanced Structural Variant Detection (beta) plugin includes a tool that identifies structural variants based on evidence from unaligned read ends and coverage information. **Note** that the functionality of the plugin described within this section is in beta. It is under active development and subject to change without notice.

The Advanced Structural Variant Detection (beta) tool is able to identify a greater number of variants with higher precision than the existing InDels and Structural Variants tool in the Workbench toolbox.

This tool comes with certain limitations:

- It is tested and developed for WGS (Whole Genome Sequencing) NGS data. Results may vary if used on targeted panel data or RNA-Seq data.
- It is developed for short read technologies (such as Illumina reads).
- It is intended for samples with a well-defined ploidy (e.g. for detecting germline variants, either haploid or diploid). Samples with somatic variants are not well-supported.
- The plugin detects the following types of structural variants: insertions (including tandem-duplications) and deletions. Inversions and inter-chromosomal rearrangements are not supported.

The tool processes each chromosome in a genome individually, through several steps:

Coverage and complexity estimation: Each chromosome is divided into bins, where each bin keeps track of the coverage and the complexity of the region the bin covers. The complexity is calculated using the Lempel-Ziv complexity measure and is used to avoid calling structural variants in low-complexity regions. The coverage information is used together with the predicated breakpoints to determine copy number variations.

Breakpoint estimation: The tool looks for unaligned read ends at each chromosome position. A consensus sequence is then made for the unaligned region and the aligned region (one sequence for each), based on a majority count in each column of nucleotides. Breakpoints are labeled either as a 'left' or 'right' breakpoint. This labeling is from the perspective of a deletion, where a left breakpoint is on the left side of a deletion (which means there is a right unaligned end)

and a right breakpoint is vice-versa on the right side of the deletion. The tool then applies a mathematical model, that based on the ploidy of the sample estimates how likely the breakpoint is to support a structural variant.

Copy number variation: The copy number variation (CNV) detection part of the algorithm looks for deletions and duplications. The algorithm uses the counts in each bin to determine if there is a statistically significant difference when compared with a normal distribution, where the normal distribution is modeled on the basis of the mean and standard deviation for the counts across all the bins (this part of the algorithm is based on [Yoon et al., 2009](#)). When there is a consecutive number of statistically significant bins, they are combined using Fisher's method to calculate a total significance value, which is then used to determine if there is a CNV.

Resolving structural variants: after the potential breakpoints have been established, consensus sequences for the aligned part of the reads, and for the unaligned ends for pairs of left-and right breakpoints are then aligned. The alignment scores for each possible pairing of left and right breakpoint alignments are stored in a matrix, and a dynamic programming algorithm is used to identify the most likely pairing of breakpoints. Breakpoints that were not matched in the previous step are then compared to the detected CNVs in order to resolve additional structural variants based on this coverage information.

Chapter 2

Advanced Structural Variant Detection

To run the Advanced Structural Variant Detection (beta) tool:

Toolbox | Resequencing Analysis (📁) | Variant Detection (📁) | Advanced Structural Variant Detection (▶️)

Once the tool wizard has opened (figure 2.1), choose the input you would like to use. The Advanced Structural Variant Detection tool accepts read mappings as either reads tracks or stand-alone read mappings.

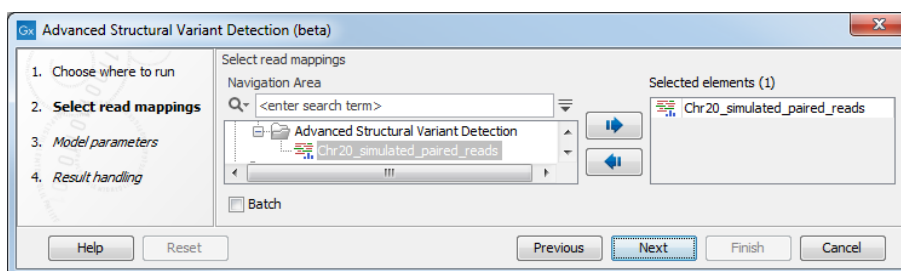


Figure 2.1: Select one or several reads tracks or stand-alone read mappings.

The tool has the following options (figure 2.2):

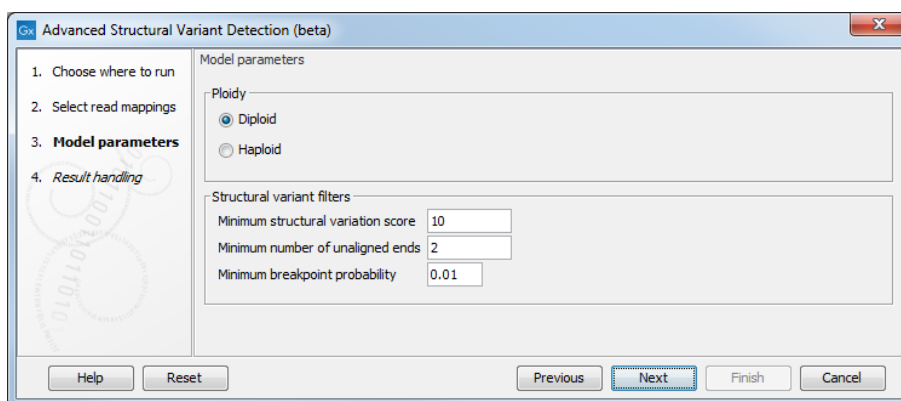


Figure 2.2: Set the model parameters for the tool.

- **Ploidy** Determines the ploidy of the sample. The value will be used in the statistical model that assign probabilities to breakpoints.

- **Minimum structural variation score** Measure of how likely this structural variant is. This value may be increased to reduce the number of structural variants called.
- **Minimum number of unaligned ends** Minimum number of reads with unaligned ends required to detect a breakpoint.
- **Minimum breakpoint probability** Probability of a breakpoint is based on a statistical model.

The tool outputs a report, a Breakpoint track (BP), and 2 Variant track (SV) and (SV annotations)

The report The report (figure 2.3) gives an overview of the numbers and types of structural variants found in the sample.

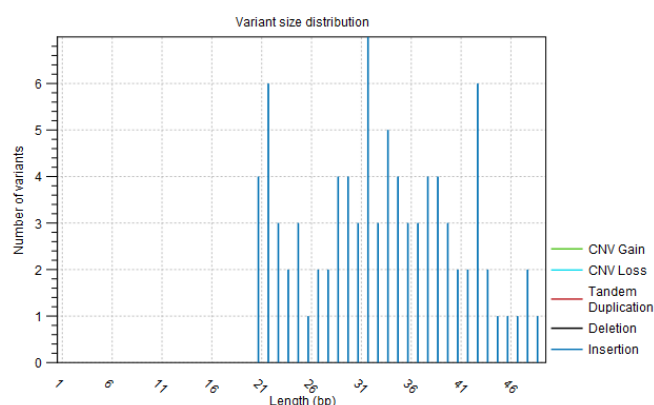
1 Variants

Chromosome	Length	Reads	Left breakpoints	Right breakpoints	Variants
20	64,444,167	8,302,333	2,214	2,227	2,089
Total	64,444,167	8,302,333	2,214	2,227	2,089

2 Summary

Chromosome	Total # variants	Insertion	Deletion	Tandem Duplication	CNV Loss	CNV Gain
20	2,089	234	1,560	266	20	9
Total	2,089	234	1,560	266	20	9

3 Length distribution <50 bp



4 Length distribution ≥ 50 bp

Figure 2.3: The Advanced Structural Variant Detection (beta) report.

It contains:

- **A 'Variants' table** with a row for each reference sequence, and information on the number of left and right unaligned breakpoint signatures and the resulting number of structural variants found.
- **A 'Summary' table** with a row for each reference sequence, and information on the total number of variants, stratified into the different variant categories (Insertion, Deletion, Tandem Duplication, CNV Loss, CNV Gain).
- **A length distribution plot for short (<50 bp) structural variants**
- **A length distribution plot for long (>50 bp) structural variants**

Breakpoint track (BP) The breakpoint track (figure 2.4) contains a row for each called breakpoint with the following information

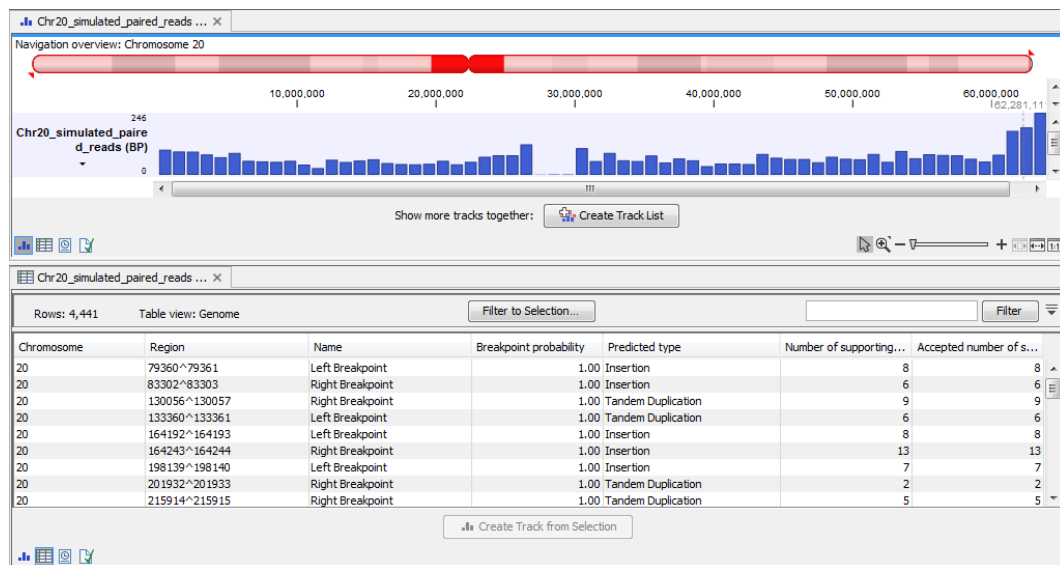


Figure 2.4: The Advanced Structural Variant Detection (beta) breakpoint track.

- **Chromosome** Chromosome on which the breakpoint is located.
- **Region** Location on the chromosome of the breakpoint.
- **Name** Type of the breakpoint ('left breakpoint' or 'right breakpoint').
- **Breakpoint probability** Estimate for how trustworthy the prediction is.
- **Predicted type** Whether the variant is a tandem duplication, or a deletion or insertion. Note that for tandem duplications only one duplication is reported, even in cases where a sequence appears in more than two copies in the reads.
- **Number of supporting unaligned ends** Number of reads at the breakpoint position with an unaligned end.
- **Accepted number of supporting unaligned ends** Number of reads at the breakpoint position with an unaligned end, after spurious unaligned ends have been filtered away.

Variant track (SV) and Variant as annotation track (SV annotations) The variant tracks (figure 2.5) contain a row for each of the called structural variations. In the case of the annotated track, the structural variations are stored as annotations: large deletions may be easier to visualize in this format.

In addition to the regular variant tracks information, the variant table contains the following information:

- **Zygosity probability** Estimate for how likely the zygosity prediction is.
- **Structural variation score** Measure of how likely this structural variant is.

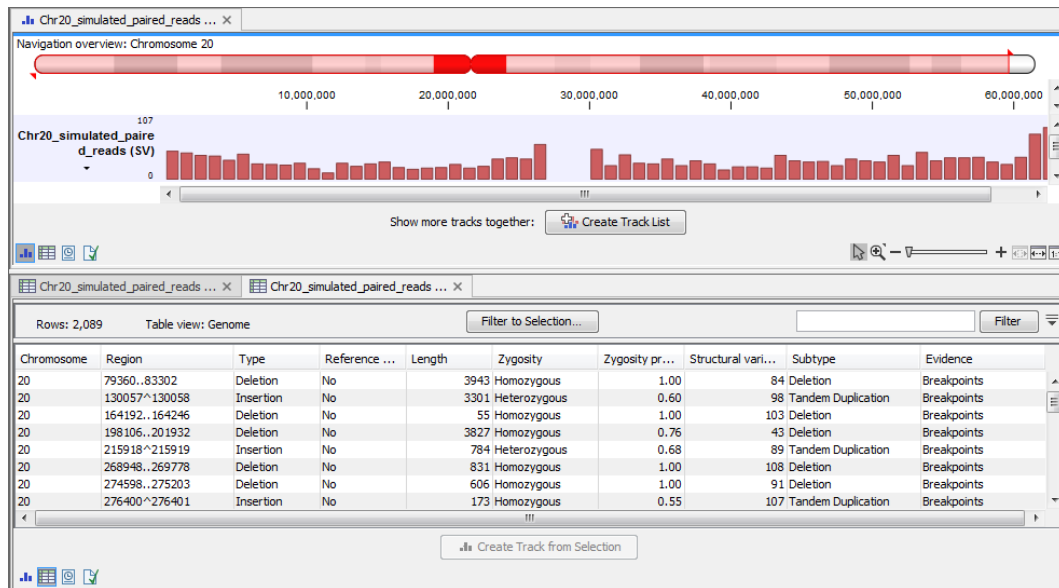


Figure 2.5: The Advanced Structural Variant Detection (beta) variant track.

- **Subtype** Available for the SV track only. This is a more specific categorization of the structural variant type: either Insertion, Deletion, Tandem Duplication, CNV Loss, or CNV Gain. Note that for Tandem Duplications only one duplication is reported, even in cases where a sequence appears in more than two copies in the reads.
- **Evidence** May be either Single Breakpoint, Paired pr Breakpoints, CNV + Breakpoint (i.e., based on coverage information and a single breakpoint), or Broken pairs (in which case the sequence may not be accurate). The broken pairs option is special since it is based on assembly of broken read pairs, where one of the reads in a pair maps at a different location in the genome. This allows for detection of insertions of Alu elements for example. Note that the sequence here may not be accurate since it is based on a homologous region of the genome.



Chapter 3

Install and uninstall plugins

Advanced Structural Variant Detection (beta) is installed as a plugin.

Note: In order to install plugins and modules, the Workbench must be run in administrator mode. On Linux and Mac, it means you must be logged in as an administrator. On Windows, you can do this by right-clicking the program shortcut and choosing "Run as Administrator".

Plugins are installed and uninstalled using the plugin manager.

Help in the Menu Bar | **Plugins...** () or **Plugins** () **in the Toolbar**

The plugin manager has two tabs at the top:

- **Manage Plugins.** This is an overview of plugins that are installed.
- **Download Plugins.** This is an overview of available plugins on QIAGEN Aarhus server.

3.1 Install

To install a plugin, click the **Download Plugins** tab. This will display an overview of the plugins that are available for download and installation (see figure 3.1).

Select Advanced Structural Variant Detection (beta) to display additional information about the plugin on the right side of the dialog. Click **Download and Install** to add the plugin functionalities to your workbench.

Accepting the license agreement

The End User License Agreement (EULA) must be read and accepted as part of the installation process. Please read the EULA text carefully, and if you agree to it, check the box next to the text **I accept these terms**. If further information is requested from you, please fill this in before clicking on the **Finish** button.

If Advanced Structural Variant Detection (beta) is not shown on the server but you have the installer file on your computer (for example if you have downloaded it from our website), you can install the plugin by clicking the **Install from File** button at the bottom of the dialog and specifying the plugin *.cpa file saved on your computer.

When you close the dialog, you will be asked whether you wish to restart the workbench. The

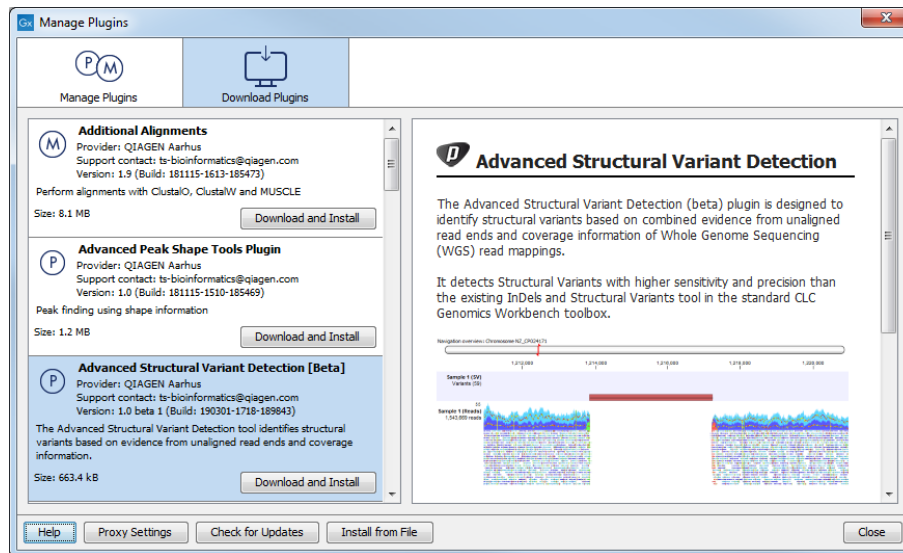




Figure 3.1: The plugins that are available for download.

plugin will not be ready for use until you have restarted.

3.2 Uninstall

Plugins are uninstalled using the plugin manager:

Help in the Menu Bar | Plugins... () or Plugins () in the Toolbar

This will open the dialog shown in figure 3.2.

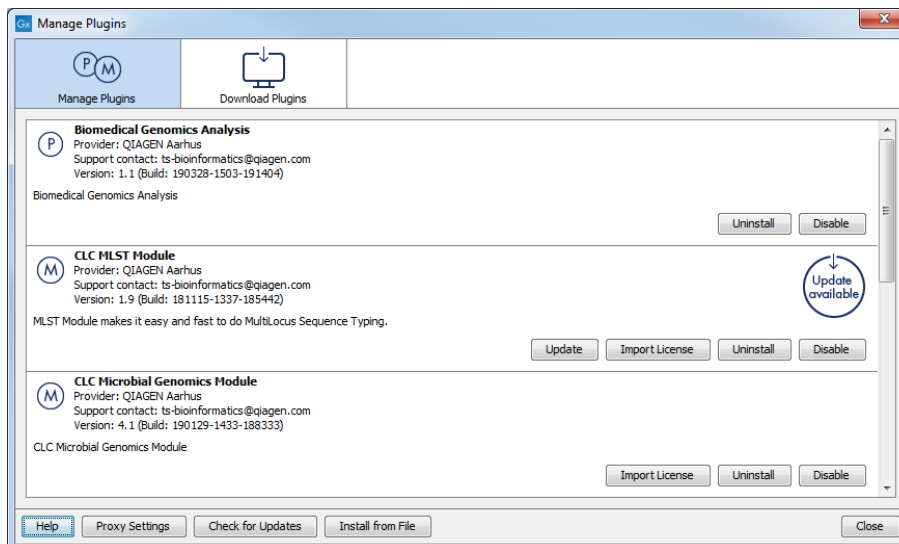


Figure 3.2: The plugin manager with plugins installed.

The installed plugins are shown in the **Manage plugins** tab of the plugin manager. To uninstall, select Advanced Structural Variant Detection (beta) and click **Uninstall**.

If you do not wish to completely uninstall the plugin, but you do not want it to be used next time you start the Workbench, click the **Disable** button.

When you close the dialog, you will be asked whether you wish to restart the workbench. The plugin will not be uninstalled until the workbench is restarted.

Bibliography

[Yoon et al., 2009] Yoon, S., Xuan, Z., Makarov, V., Ye, K., and Sebat, J. (2009). Sensitive and accurate detection of copy number variants using read depth of coverage. *Genome Res.*, 19:1586–1592.