cgatools Release Notes

Version 1.3.0

Related Documents

Customers should consult the *cgatools Methods* document for detailed information on specific tools. This document can be downloaded at: http://cgatools.sourceforge.net.

New Features and Enhancements Version 1.3

The following new features and enhancements are provided in this release by comparison with previous **cgatools** released by Complete Genomics:

1. Added generatemastervar (beta) tool to create a simple, integrated master variation (*masterVar*) file that contains one line per locus and combines variation, annotation, and coverage information. The *masterVar* file serves as an aggregated source of information that is formatted to be easily expanded to include custom annotations and processed with simple command-line tools, including searching and filtering. The *masterVar* file provides a structured content that can more easily be converted into other standard variation file formats. This tool can support Complete Genomics data from version 1.5 onward. For

versions prior to 1.10, it is necessary to exclude 'cnv' from the input to the –-annotations parameter since CNV data is not available in these versions.

2. Modified the calculation of the *SomaticScore* column in the calldiff somatic output file to use two new score ranking columns, *VarScoreARank* and *RefScoreBRank*. The new somatic score better handles somatic categories that only have a few variations. It is also able to capture potential somatic variations where the normal sample is no-called. Finally, with this new score, a selected cutoff will represent the same sensitivity when applied to different samples.

3. Changed the behavior of the listvariants tool such that for longer 'subs', the range of the loci is trimmed when edge sequence matches the reference. This, potentially, results in the assignment of a different *varType* for variations that can be more explicitly defined.

Addendum

1. We have made an addendum to Version 1.2 Release Notes and Version 1.0.0. Please refer to these Addendum sections for details of the changes.

Changes to Version 1.2

1. Added junctiondiff (beta) tool.

2. Changed LocusDiffClassification for calldiff LocusOutput from alt-consistent to ref-consistent and from alt-identical to ref-identical, wherever the allele is consistent with the reference. In previous versions, the alt- classification was used whenever reference-consistent calls were aligned to calls of the other genome that were reference-inconsistent.

Addendum

1. When using junctiondiff to identify structural variation junctions that exist in only one genome, if the --minlength parameter is specified, the report.txt statistics file does not correctly reflect the number of length-filtered junctions. However, the final junction output file does contain the correct set of unique, length-filtered junctions. As a result, the report.txt file does not accurately summarize the output junction file.

Changes to Version 1.1.1

1. Fixed crash in join tool when reading ***geneVarSummary*** file.

2. Added Mac OS X binary tarball.

Changes to Version 1.1.0

1. Enhanced the calldiff tool to allow users to identify somatic variants from a tumor/normal pair (beta). This tool takes two variant files—genome A (tumor) and genome B (normal)—as inputs and produces:

   □ a report ("SomaticOutput") that lists variants found only in genome A

   □ a score that indicates the likelihood that each variant is truly somatic

3. Added listvariants (beta) and testvariants (beta) tools for comparing variants across multiple genomes, allowing users to determine whether a variant was found in a given genome and the frequency of the variant across the set of tested genomes. listvariants generates a list of all fully called variants found in at least one genome within the tested set.

testvariants uses this list as input and reports for each variant whether the allele called in each genome is 1) inconsistent with the variant, 2) is fully called and is consistent with the variant, or 3) has no-calls and allele is consistent with the variant.

4. Added the join (beta) tool, allowing users to combine information from two tab-delimited files by specifying column(s) within the files to be used for determining overlap and column(s) from each file to be included in the merged file.

5. Changed parameters for snpdiff and calldiff tools. In previous releases of **cgatools**, separate parameters were required to output each report (such as Stats, Output, and SuperlocusOutput) to a specific location. These parameters were replaced with two new ones: a parameter that allows you to specify multiple reports to be output and a parameter that allows you to specify a path to the directory to which all output reports will be saved.

Changes to Version 1.0.0.15

1. Fixed snpdiff and calldiff failure processing male build 37 genomes.

Changes to Version 1.0.0.14

1. Fixed evidence2sam to be able to support genomes from assembly format version 1.0.

Changes to Version 1.0.0.13

1. Changed to dynamic linkage on Mac OS X.

Changes to Version 1.0.0

1. Renamed cgi2sam to map2sam.
2. Added evidence2sam (beta).

Addendum

1. Complete Genomics may call a variation that represents an insertion of sequence after the last base of the chromosome. In these rare cases, the evidence2sam tool will display the following warning message to alert the user that this event has occurred.

   ```
   Warning: offset position is behind the end of a non-circular chromosome
   ```

   The resulting SAM file will still be successfully generated. The impact of these alignments in the SAM/BAM file on downstream analysis tools is currently untested.

Changes to Version 0.5.0 (Initial Version)

1. The initial version of **cgatools** included:

   - Reference tools
   - snpdiff
   - calldiff

◻ cgi2sam