Structural Variant Annotation Pipeline File Format Specification Sheet

Document Number: 30168

Document Revision: G

Table of Contents

Legal Notice

For Research Use Only. Not for use in diagnostic procedures.

Revision History

Revision	Notes
G	• Added description of zygosity and ethnicity control database columns. • Added description of ISCN notation column • Added description of molecule count columns

Introduction

The Variant Annotation Pipeline enables users to determine if a Bionano structural variant (SV) call is relevant to certain phenotypes or disease traits. For more information about the pipeline, please refer to Bionano Solve Theory of Operation: Variant Annotation Pipeline (PN 30190). The output file of the Variant Annotation Pipeline is an annotated SMAP file, with additional annotation columns appended. The SMAP file format is a general format to describe SVs detected by Bionano; please refer to SMAP File Format Specification Sheet (PN 30041).

This document describes only the additional annotation columns. Also, note that the last few columns can vary depending on whether a trio, dual or single analysis has been performed upon execution of the Variant Annotation Pipeline.

Annotation Columns – All Analyses

Annotation	Description
sample	The sample name.
algorithm	This is based on comparing the *de novo* assembly or the Rare Variant Pipeline SV calls of the sample with the reference; this is typically output as “assembly comparison.”
size	The size (in base pair) of insertion and deletion calls. It would be -1 for other variant types.
Present_in_%_of_BNG_control_samples	The percentage of samples in the Bionano control SV database that also carry the SV.
Present_in_%_of_BNG_control_samples_with_the_same_enzyme	Similar to above, but the percentage is calculated based on database samples having ***the same enzyme*** as the sample being annotated.
%_BNG_control_sample_with_homozygous_SV	**The percentage of samples in the Bionano control SV database that are homozygous for the SV**
%_of_BNG_control_sample with_heterozygous_SV	**The percentage of samples in the Bionano control SV database that are heterozygous for the SV**
%_of AFR_BNG_control_sample_with_SV	**The percentage of African samples in the Bionano**

	control SV database that carry the SV
%_of AMR_BNG_control_sample_with_SV	**The percentage of Admixed American samples in the Bionano control SV database that carry the SV**
%_of_EUR_BNG_control_sample_with_SV	**The percentage of European samples in the Bionano control SV database that carry the SV**
%_of_EAS_BNG_control_sample_with_SV	**The percentage of East Asian samples in the Bionano control SV database that carry the SV**
%_of_SAS_BNG_control_sample_with_SV	**The percentage of South Asian samples in the Bionano control database that carry the SV**
%_of_unknown_BNG_control_sample_with_SV	**The percentage of Unknown population samples in the Bionano control database that carry the SV**
Fail_assembly_chimeric_score	A flag used to denote whether there might be a potential chimeric join at the variant locus. This denotes whether a minimal chimeric quality score of 35 and coverage of 10X have been achieved around each SV breakpoint. A value of 'pass' means that the two criteria have been met; a 'fail' denotes the criteria not met; and a 'not_applicable' value denotes that the check has not been performed. Notice that this check is performed only for inversion and translocation calls. Note: a chimeric quality score of a label on a genome map is the percent of molecules that align to both sides of the label out of all molecules that align on either side near this label. See CMAP File Format Specification Sheet (PN 30039) for detail.
num_overlap_DGV_calls	**If the sample is human, then the SVs would be compared against the Database of Genomic Variants (DGV), and the number of DGV variants overlapping the call is outputted.**
OverlapGenes	A semi-colon separated list indicating which genes overlap with the SV.
NearestNonOverlapGene	The next closest gene to the SV.
NearestNonOverlapGeneDistance	The distance between the SV and the next closest
PutativeGeneFusion	The list of fusion genes that may be created by the SV.

Found_in_self_molecules	This column denotes whether there is a sufficient number of case/proband sample's molecules supporting the **case/proband** sample's genome map at the SV breakpoints. The possible values are 'yes' and 'no'. Since inversion_partial calls are not annotated, a value of '-' is shown for inversion_partial calls. Note that the minimum number of molecules required is defined as a parameter by the users upon running the variant annotation pipeline.
Self_molecule_count	**The number of molecules supporting the SV.**
UCSC_web_link1	If the sample is either human or mouse, then a weblink to the SV region in the UCSC genome browser would be created. If the SV is a translocation breakpoint, then the weblink goes to one side of the translocation breakpoint.
UCSC_web_link2	The weblink goes to the other side of the translocation or fusion breakpoint.
ISCN	International System for Human Cytogenomic Nomenclature notation for SV

Trio Analysis

Annotation	Description
Found_in_parents_assemblies	Whether the SV call is also identified in the father's or mother's assembly. The possible values are 'mother', 'father', 'both' and 'none'. Since inversion_partial calls are not annotated, a value of '-' is shown for any inversion_partial call.
Found_in_parents_molecules	This column shows whether there is a sufficient number of parents' molecules supporting the proband's genome map at the SV breakpoints. The possible values are 'mother', 'father', 'both' and 'none'. Since inversion_partial calls are not annotated, a value of '-' is shown for inversion_partial calls. Note that the minimum numbers of molecules required are defined as parameters by the users upon running the variant annotation pipeline.

Mother_molecule_count	**Number of molecules supporting the SV in the mother's assembly.**
Father_molecule_count	**Number of molecules supporting the SV in the father's assembly**

Dual Analysis

Statistic	Description
Found_in_control_sample_assembly	Whether the SV call is also identified in the control sample's assembly. The possible values are 'yes' or 'no'. Since inversion_partial calls are not annotated, a value of '-' is shown for any inversion_partial call. See Bionano Solve Theory of Operation: Variant Annotation Pipeline (PN 30190).
Found_in_control_sample_molecules	This column shows whether there is a sufficient number of control sample's molecules supporting the case sample's genome map at the SV breakpoints. The possible values are 'yes' and 'no'. Since inversion_partial calls are not annotated, a value of '-' is shown for inversion_partial calls. Note that the minimum number of molecules required is defined as a parameter by the users upon running the variant annotation pipeline. See Bionano Solve Theory of Operation: Variant Annotation Pipeline (PN 30190).
Control_molecule_count	The number of molecules supporting the SV in the control sample

Technical Assistance

For technical assistance, contact Bionano Genomics Technical Support.

You can retrieve documentation on Bionano products, SDS's, certificates of analysis, frequently asked questions, and other related documents from the Support website or by request through e-mail and telephone.

Type	Contact
Email	**support@bionanogenomics.com**
Phone	**Hours of Operation:** **Monday through Friday, 9:00 a.m. to 5:00 p.m., PST** **US: +1 (858) 888-7600**
Website	**www.bionanogenomics.com/support**

Bionano Genomics, Inc.
9540 Towne Centre Drive, Suite 100
San Diego, CA 92121