Absence/Loss of Heterozygosity Pipeline File Format Specification Sheet

Document Number: 30460

Document Revision: A

Table of Contents

Legal Notice

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

This material is protected by United States Copyright Law and International Treaties. Unauthorized use of this material is prohibited. No part of the publication may be copied, reproduced, distributed, translated, reverse-engineered or transmitted in any form or by any media, or by any means, whether now known or unknown, without the express prior permission in writing from Bionano Genomics. Copying, under the law, includes translating into another language or format. The technical data contained herein is intended for ultimate destinations permitted by U.S. law. Diversion contrary to U. S. law prohibited. This publication represents the latest information available at the time of release. Due to continuous efforts to improve the product, technical changes may occur that are not reflected in this document. Bionano Genomics reserves the right to make changes in specifications and other information contained in this publication at any time and without prior notice. Please contact Bionano Genomics Customer Support for the latest information.

BIONANO GENOMICS DISCLAIMS ALL WARRANTIES WITH RESPECT TO THIS DOCUMENT, EXPRESSED OR IMPLIED, INCLUDING BUT NOT LIMITED TO THOSE OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. TO THE FULLEST EXTENT ALLOWED BY LAW, IN NO EVENT SHALL BIONANO GENOMICS BE LIABLE, WHETHER IN CONTRACT, TORT, WARRANTY, OR UNDER ANY STATUTE OR ON ANY OTHER BASIS FOR SPECIAL, INCIDENTAL, INDIRECT, PUNITIVE, MULTIPLE OR CONSEQUENTIAL DAMAGES IN CONNECTION WITH OR ARISING FROM THIS DOCUMENT, INCLUDING BUT NOT LIMITED TO THE USE THEREOF, WHETHER OR NOT FORESEEABLE AND WHETHER OR NOT BIONANO GENOMICS IS ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

Patents
Products of Bionano Genomics® may be covered by one or more U.S. or foreign patents.

Trademarks
The Bionano Genomics logo and names of Bionano Genomics products or services are registered trademarks or trademarks owned by Bionano Genomics in the United States and certain other countries.

Bionano Genomics®, Saphyr®, Saphyr Chip®, and Bionano Access® are trademarks of Bionano Genomics, Inc. All other trademarks are the sole property of their respective owners.

No license to use any trademarks of Bionano Genomics is given or implied. Users are not permitted to use these trademarks without the prior written consent of Bionano Genomics. The use of these trademarks or any other materials, except as permitted herein, is expressly prohibited and may be in violation of federal or other applicable laws.

© Copyright 2021 Bionano Genomics, Inc. All rights reserved.

Revision History

Revision	Notes
A	Initial document release

Introduction

The Bionano Genomics® AOH/LOH detection pipeline detects regions with absence or loss of heterozygosity based on the zygosity of structural variants (SVs) detected by the de novo assembly pipeline. For more information about AOH/LOH detection, please refer to Bionano Solve Theory of Operation: Structural Variant Calling (PN 30110). The output of the pipeline consists of two tab-delimited, text-based files; one file describes each call, and one file provides detailed information about the structural variants that were used during detection. This file format specification sheet provides descriptions of the AOH/LOH output file headers and descriptions of the columns in each file.

When the data are imported into Bionano Access™, the AOH/LOH output files are automatically processed and visualized. AOH/LOH output files can also be opened in Excel for easy readability, or in any tab-delimited, text-based editor.

AOH/LOH Calls

Format

The Bionano Genomics Absence/Loss of Heterozygosity output file *loh_calls.txt* contains the following sections:

- AOH/LOH file header
    - # SMAP Entries From:
    - # Package:
    - # HMM parameters:
    - #h
    - #f
- AOH/LOH calls block (each row as defined by the column headers in #h
    - Unique ID for the AOH/LOH call [Id]
    - The position along the reference and width of each call [Chromosome, Start, End, Width]
    - Followed by the confidence score of the AOH/LOH call and their corresponding SV type [Confidence].

Header Specifications

Header rows are prefixed by the pound sign (#). “*” Denotes required header line tags.

Header Line Tag	Header Line Description
# Smap Entries From:	SMAP file used to make AOH/LOH calls*

# Package:	Package name, version and installation date*
# HMM Parameters:	Parameters used by Hidden Markov Model
#h	The columns for each data row
#f	The numerical data type for each data column

Header Specification Details

# Smap Entries From	
Header	# Smap Entries From:
Description	SMAP file used to make AOH/LOH calls, auto-generated.
Example	# Smap Entries From:<TAB>merged_smaps/exp_refineFinal1_merged_filter_inversions.smap

# Package	
Header	# Package:
Description	Package name, version and installation date, auto-generated.
Example	# Package: lohdetection. branch: master. Commit hash f47e05d21e761bda477230058f793. Installation date: 2021-10-01 09:56:38.621762

# HMM parameters	
Header	# HMM parameters:
Description	Parameters used for Hidden Markov Model.
Example	# HMM parameters: transm_bg2bg = 0.959011191677458; transm_loh2loh = 0.8042549108141759; startprob_bg = 0.8215713321165973; emissionprob_bg_hom = 0.31788854922897; emissionprob_loh_hom = 0.975137538627914

#h		
Header	#h	
Description	Defines the columns for each data row in #h rows:	
	Id	
	Chromosome	Map ID, ordinal number
	Start	Start position of AOH/LOH call on map [0-based from map start] in basepairs
	End	End position of AOH/LOH call on map [0-based from map start] in basepairs
	Width	AOH/LOH call width in basepairs

	Confidence	The probability that the AOH/LOH call is a true event. More specifically, this is the model's precision when detecting simulated AOH/LOH events of a similar width.
Example	#h Id<TAB>Chromosome<TAB>Start<TAB>End<TAB>Width<TAB>Confidence	

#f	
Header	#f
Description	Defines the numerical data type for each data column.
Example	#f<TAB>int<TAB>int<TAB>float<TAB>float<TAB>float<TAB>float

AOH/LOH Per SV Info

Format

The Bionano Genomics Absence/Loss of Heterozygosity output file *loh_per_sv_info.txt* contains details about the subset of SVs from the input SMAP file that were used in AOH/LOH detection, with a few additional columns containing information about the AOH/LOH calls. The file contains the following sections:

- AOH/LOH file header
    - # SMAP Entries From:
    - # Package:
    - # HMM parameters:
    - #h
    - #f
- AOH/LOH calls block (each row as defined by the column headers in #h
    - After the 2 IDs [SmapEntryID, RefcontigID1] are the positions along the reference of each SV [RefStartPos, RefEndPos].
    - Followed by the confidence scores of the SV calls, their corresponding SV type, and zygosity of the SV [Confidence, Type, Zygosity].
    - The final columns indicate whether the SV was determined to be in an AOH/LOH region, and the calculated probability that the SV is in an AOH/LOH region [In_AOH_LOH, AOH_LOH_prob].

Header Specifications

Header rows are prefixed by the pound sign (#). "*" Denotes required header line tags.

Header Line Tag	Header Line Description
# Smap Entries From:	SMAP file used to make AOH/LOH calls*
# Package:	Package name, version and installation date*
# HMM Parameters:	Parameters used by Hidden Markov Model
#h	The columns for each data row
#f	The numerical data type for each data column

Header Specification Details

# Smap Entries From	
Header	# Smap Entries From:
Description	SMAP file used to make AOH/LOH calls, auto-generated.
Example	# Smap Entries From:<TAB>merged_smaps/exp_refineFinal1_merged_filter_inversions.smap

# Package	
Header	# Package:
Description	Package name, version and installation date, auto-generated.
Example	# Package: lohdetection. branch: master. Commit hash f47e05d21e761bda477230058f793. Installation date: 2021-10-01

# HMM parameters	
Header	# HMM parameters:
Description	Parameters used for Hidden Markov Model.
Example	# HMM parameters: transm_bg2bg = 0.959011191677458; transm_loh2loh = 0.8042549108141759; startprob_bg = 0.8215713321165973; emissionprob_bg_hom = 0.31788854922897; emissionprob_loh_hom = 0.975137538627914

#h		
Header	#h	
Description	Defines the columns for each data row in #h rows:	
	SmapEntryID	A unique number for an entry in the input SMAP file
	RefcontigID1	Reference contig ID (XmapID1). Map ID of the reference map from the .cmap reference file (the .cmap file may contain multiple reference maps). Note: RefContigIDs must be integers, but they need not be sequential.
	RefStartPos	Coordinate of reference contig ID1 aligned site which borders this SV. This site is always either a start or end of XmapID1 and it matches the site at the query start position (QryStartPos).

	RefEndPos	Coordinate of reference contig ID2 aligned site which borders this SV. This site is always either a start or end of XmapID2 and it matches the site at the query end position (QryEndPos).
	Confidence	Estimate of probability of being correct for insertions and deletions, and a quality metric for inversion and translocation breakpoints. Other SVs are given a placeholder value of '-1.00'. See Bionano Solve Theory of Operation: Structural Variant Calling (document 30110).
	Type	Type of SV. For definitions, see SMAP File Format Specification Sheet (document 30041).
	Zygosity	One of 'homozygous', 'heterozygous' or 'unknown' based on overlap with other SVs and alignments.
	In_AOH_LOH	Whether this SV was determined to be in an AOH/LOH region. '1' for yes, '0' for no.
	AOH_LOH_prob	The probability of the SV being in an AOH/LOH region as calculated by the Hidden Markov Model.
Example	#h SmapEntryID<TAB>RefcontigID1<TAB>RefStartPos<TAB>RefEndPos<TAB>Confidence<TAB>Type<TAB>Zygosity<TAB>In_AOH_LOH<TAB>AOH_LOH_prob	

#f	
Header	#f
Description	Defines the numerical data type for each data column.
Example	#f<TAB>int<TAB>int<TAB>int<TAB>float<TAB>float<TAB>string<TAB>string<TAB>int<TAB>float

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