Bionano Solve v3.6 Release Notes

Document Number: 30380

Document Revision: B

Table of Contents

Legal Notice

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

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®, Irys®, IrysView®, IrysChip®, IrysPrep®, IrysSolve®, Saphyr®, Saphyr Chip®, Bionano Access®, and Bionano EnFocus™ 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.

Revision History

Revision	Notes
A	Initial release of document.
B	Revision for Bionano Solve 3.6.1 release.

Updates

In the Bionano Solve 3.6.1 release, the hg19 SV mask BED file and the hg19 variant annotation control database for the *de novo* assembly pipeline have been updated. The database in Solve 3.6 version were incorrectly generated. The other mask BED files and control databases were not impacted. A comparison of the contents of the Solve 3.5 and Solve 3.6.1 masks is included in the Bionano Solve Theory of Operation: Structural Variant Calling document (PN 30110). In general, the differences are minor.

Introduction

This document describes the release of Bionano Solve 3.6. We provide an overview of the fixes and improvements of the Bionano Solve analysis tools and pipelines so that you may better understand the impact of moving to this version of our software. Should you have any questions, please contact support@bionanogenomics.com.

Bionano Tools and Bionano Solve are now combined together and branded as Bionano Solve. Bionano Solve is installed on Saphyr Compute, Bionano Compute, and Bionano Access Servers before server shipment and installation. IrysSolve servers, which have been in use with Bionano Access, also have Bionano Solve installed during an upgrade.

Bionano Solve (folder “tools”) is located at the /home/bionano directory on the compute server. The folder contains a collection of tools and scripts. Each individual tool is versioned independently. These tools together perform bioinformatics analyses on the compute server.

Compatibility

Bionano Solve 3.6 is compatible with Bionano Access 1.6 only.

Use of IrysSolve servers

If you are using an IrysSolve server to perform computational analysis on data generated by the Saphyr instrument, contact Bionano Support to reconfigure the IrysSolve prior to running any samples labeled with the DLS chemistry.

References

Visit https://bionanogenomics.com/support-page/data-analysis-documentation/ for file format specifications and Theory of Operation documents.

Bionano EnFocus™ FSHD Analysis Pipeline

- Added more detailed QC data in pipeline output package, in case they are needed for troubleshooting purposes.

Copy number analysis pipeline

- Added flexibility to perform non-human copy number analysis. Previously, copy number analysis was restricted to human datasets. We added pre-processed data for mouse (processed with reference build mm10/GRCm38.p6). As described in the Bionano Solve Theory of Operation: Structural Variant Calling (PN 30110), we also provided command-line based tools for pre-processing control data. The pipeline would output the raw coverage signal if no control data are available. These changes have been integrated with the *de novo* assembly pipeline and RVP.
    - Note: The mouse samples we used were of the C57BL/6 strain (commonly referred to as “Black 6” or “B6”). When analyzing of other strains, use of custom control data is encouraged. If such data are not available, unexpected copy number results may be due to strain differences.
- Added new region-based confidence scoring to improve handling of complex regions in the genome.
- Added new coverage variation metrics that assess impact of factors unrelated to genuine copy number changes for troubleshooting purposes.
- Improved memory utilization when reading in input data.
- Fixed an issue with aneuploidy events not being detected.
- Updated the copy number variant types from “deletion” and “duplication” to “loss” and “gain”, respectively.

De novo assembly pipeline

- Added filtering of chrY calls for samples inferred to be female, where chrY is assumed to be absent.
- Added automatic VCF conversion.
- Improved auto-noise parameters for better memory utilization.
- Improved logging for annotated runs, when VAP is enabled.

- Increased stringency of merging overlapping assembled maps during extension-and-merge stages to avoid generating chimeric maps. This might have the side effect of reducing the assembly N50 in certain cases (for example, if the data quality is suboptimal).
- Updated informatics report for clarity.

Hybrid Scaffolding

- Added option to trim overlapping sequence contigs when exporting the final FASTA to avoid having redundant sequences.
    - Note: sequence alignment is not performed. The trimming is done based on map alignment and may be imprecise.
- Improved alignment of contigs that had significant overhangs when aligned to the genome maps. The issue previously caused some contigs to be left out of the hybrid scaffolds.
- Fixed an issue with inconsistent counts between the summary report and the FASTA output.

Molecule Quality Report (MQR)

- Fixed an issue with the Molecule Integrity Number (MIN) values not being computed properly.

Rare Variant Pipeline (RVP)

- Added filtering of chrY calls for samples inferred to be female, where chrY is assumed to be absent.
- Added automatic VCF conversion.
- Improved logging for annotated runs, when VAP is enabled.
- Improved auto-noise parameters for better memory utilization.
- Improved translocation calling by adjusting parameters to minimize the number of false positive calls.
- Improved memory utilization during the molecule-based SV calling (“alignmolvrefsv”) stage.
- Updated informatics report for clarity.
- Fixed an issue with missing copy number data when RVP output is imported into Access.
    - Note: Import of pre-Solve 3.6 runs would result in visualization issues due to the missing copy number data. Users are encouraged to re-run those datasets using the de novo or RVP pipeline.

SV confidence

- Added new models for scoring translocation and inversion breakpoint confidence for human and non-human datasets. The new models were developed using simulated and recently published SV datasets, where there were orthogonal SV data.

SVMerge

- SVMerge has been deprecated in this release. The single-enzyme DLE-1 workflow is recommended for human SV analysis.

Variant Annotation Pipeline (VAP)

- Added new pre-processed DLE-1 control datasets for use with the *de novo* assembly pipeline and RVP. The number of DLE-1 datasets is now 180 in total.
    - Note: we incorporated data from 30 COVID-positive samples. They had no known severe genetic conditions otherwise. Contact Bionano Support if you require custom control databases that do not include the COVID-positive samples.
- Added flexibility to perform non-human variant annotation. Previously, variant annotation was restricted to human datasets. We added pre-processed data for mouse (processed with reference build mm10/GRCm38.p6). As described in Bionano Solve Theory of Operation: Variant Annotation Pipeline (PN 30190), we also provided command-line based tools for pre-processing control data. Users can input custom gene BED files and custom control databases. Also, VAP can proceed without those inputs.
    - Note: we incorporated data from 11 B6 mice, only one of which was considered a true control. Other mice had various phenotypes. Users need to exercise caution when using the mouse control data and when interpreting the annotation results.
- Added region-based UCSC links for annotated SVs (only available for hg19, hg38, and mm10).
- Added comparison of variants in Database of Genomic Variants (DGV; http://dgv.tcag.ca/dgv/app/home).
- Added SV output in BED format; the translocation breakpoint calls are output in Interact format (https://genome.ucsc.edu/goldenPath/help/interact.html). They can be uploaded to UCSC Genome Brower.
- Updated existing gene BED files. Now, we provide both the canonical gene annotations and Gencode (https://www.gencodegenes.org/) gene annotations. Also, we provide gene BED files for various common species such as rat, zebrafish, and sorghum.

VCF conversion

- Added option to include CNV calls in the VCF output.
- Added VCF export to SV operations.
- Fixed an issue with the input parameters not being correctly propagated.

Miscellaneous

- Added new reference CMAP files for common species such as rat, zebrafish, and sorghum.
- Added command-line based tools for generating custom SV mask BED files. These tools also provide auxiliary functions to enable application of the masks to SMAP output. The usage of the tools is described in the Bionano Solve Theory of Operation: Structural Variant Calling (PN 30110),
- Converted tar.gz output to .zip output for various pipelines for better import efficiency.

Other known issues and limitations

- Users with non-standard cluster or server configurations may experience suboptimal runtime performance.
- Hybrid Scaffold output FASTA/AGP files may contain header lines with whitespaces and would not pass NCBI AGP validation. Manual editing of the FASTA/AGP files may be needed.
- Haplotype-aware refinement on non-human datasets is not a supported feature, and its use may have unintended consequences.
- Large heterozygous duplications may be called as homozygous when the allele with a single copy is not successfully assembled, and if too few molecules span across the entire duplication region.
- We observed that with the Rare Variant Pipeline, PPV was slightly lower (~80%) for deletions under 25 kbp. This was not observed with the *de novo* assembly pipeline.
- 300X effective coverage is recommended for typical low variant allele frequency applications using RVP, where events are expected at as low as 5% allele fraction. To achieve that performance with most samples, we recommend 1.5 Tbp to be collected on Saphyr. RVP has been validated with up to 1000X effective coverage (equivalent of 5 Tbp of data collected on Saphyr). We do not recommend running this pipeline with more than 1000X.

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) 888-7663**
Website	**www.bionanogenomics.com/support**

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