bionano™

Data Collection Guidelines

Table of Contents

Revision History

REVISION	NOTES
A	Initial release.
B	Revised
C	Revised
D	Add guidance for RVA, FSHD. Add example scenarios
E	Document updates for Solve 3.7 release • Clarify minSites >= 9 role in effective coverage calculations • Add guidance for Fragile X • Add file size table • Update informatics report formats
F	Changed heading in column C of Table 4 and added new column. Also added text regarding sample quality.
G	Revised to include guided assembly pipeline
H	Updated to include agnostic information intended for both Saphyr and Stratys instruments.

Introduction

The Saphyr® and Stratys™ Systems collect data by drawing ultra-high molecular weight DNA into nanochannel arrays, imaging the molecules in linear confinement, and detecting molecule and label positions from the images. Once a batch of molecules in the nanochannels have been imaged, the instrument clears the array and loads in new molecules. The system repeatedly cycles through this process (each cycle is one "scan"). The greater the number of scans that accumulate, the more data collects for a given dataset. Operational details are covered in the *Saphyr System User Guide* (CG-30247) and the Stratys System User Guide (CG-00041).

Per-flowcell limits on maximum accumulated output and runtime vary by chip part number and instrument. Within those limits, it is up to the user to determine how much data to collect. This document will provide guidance and discuss data collection strategies for various downstream applications.

The key factors for success in using the Saphyr or Stratys System include:

- Design experiments to meet research goals.
- Isolate high molecular weight gDNA from sample sources, then efficiently label at enzyme-specific recognition sequence.
- Collect enough high-quality data for downstream data analysis.

The Saphyr and Stratys Systems collect whole genome imaging data. As such, throughput and coverage targets discussed here are genome-wide in scope, regardless of any more focused downstream application. Furthermore, the guidance here assumes that data is of reasonably good quality. For a discussion of raw data quality, please refer to *Bionano Access Dashboard Guidelines* (CG-30304) and/or *Molecule Quality Report Guidelines* (CG-30223). For guidance on troubleshooting sample prep or quality from a particular sample type, please refer to the appropriate Bionano Prep™ documentation.

Throughput targets and coverage

Prior to collecting data, the user enters chip and flowcell information in Bionano Access™. A user is prompted to specify the Throughput target (in Gbp) for each flowcell. This throughput target should reflect an appropriate fold coverage (X) of the haploid genome size investigated, specific to the intended application.

Conceptually, coverage refers to the average depth of molecules representing a given locus in the genome. It can be calculated from the predicted haploid genome size and the amount of input data.

$$Coverage\ (X) = \frac{Total\ DNA}{Genome\ size}$$

For example, 310 Gbp total DNA throughput of a sample with a 3.1 Gbp haploid genome size equals 100X coverage (310 Gbp ÷ 3.1 Gbp = 100).

In practice, not every molecule in a raw dataset adequately represents a locus. Similarly, reference assemblies are not perfect approximations of the genome. It is useful to define <u>effective coverage of the reference</u>, which uses alignment (i.e., Map Rate) and genome size as in reference assembly as parameters.

$$Effective\ coverage\ of\ reference\ (X) = \frac{Total\ DNA \times [Map\ Rate]}{Reference\ size}$$

Throughput and coverage are further refined into specific terms found throughout Bionano software, outlined in **Table 1.**

Table 1. Glossary of terms relating Throughput and Coverage

Term	Meaning	Source	Comment
Pre-Analysis metrics: *prior to secondary bioinformatic analysis*			
Raw coverage (X)	[Total DNA (>=150 kbp)] ÷ [haploid genome size]		Useful baseline when reference and/or data quality unclear
Throughput target	Minimum total flowcell throughput (≥150 kbp) that the instrument targets to collect	Bionano Access; Instrument Control Software (ICS)	
Total DNA (>= 20 kbp)	Cumulative length of DNA molecules that are ≥ 20 kbp	Dashboard; Molecule Quality Report	Raw molecules detectable by the instrument
Total DNA (>=150 kbp)	Cumulative length of DNA molecules that are ≥ 150 kbp	Dashboard; Molecule Quality Report	Counts toward Throughput target
Est Effective coverage >= 150 kbp*	[Total DNA (>=150 kbp)] × [Map Rate %]* ÷ [reference size]	Dashboard	Approximates effective coverage of reference (X) prior to analysis
Total DNA (>= 150 kbp & minSites >= 9)	Cumulative length of DNA molecules that are ≥ 150 kbp and have ≥ 9 labels	Molecule Quality Report	Pass filters for downstream analyses
Effective coverage*	[Total DNA (>= 150 kbp & minSites >= 9)] × [Map Rate %]* ÷ [reference size]	Molecule Quality Report	Approximates effective coverage of reference (X) prior to analysis
Post-Analysis metrics: *output from secondary bioinformatic analysis pipelines*			
Effective coverage of reference (X)*	[Total DNA aligned to the reference in pipeline]* ÷ [reference size]	Informatics Summary	

| Effective coverage of assembly (X) | [Total DNA aligned to the assembly in pipeline] ÷ [assembly size] | Informatics Summary |

*Calculation of Map Rate and Effective Coverages relies on quality and completeness of the provided reference, as well as sequence similarity between reference and sample. Human genome references are considered relatively complete, so these metrics are meaningful and can be clearly benchmarked. With non-human organisms or custom references, molecule-reference alignment can be hindered by a myriad of factors, which may underestimate coverage and data quality. Please see *Assembly Report Guidelines* (CG-30255) for additional details.

In human samples, the effective coverage provided pre-analysis is generally higher than the effective coverage of reference (X) in the pipeline results. This stems from different stringencies in alignment parameters between pre-analysis mapping and eventually running the pipeline. Effective coverage recommendations outlined here refer to pre-analysis metric shown in the Molecule Quality Report.

Coverage recommendations for different analyses

Coverage recommendations vary considerably by downstream application. As of Bionano Solve 3.8.1, there are pipelines for Structural Variant (*de novo* Assembly, Guided Assembly, Rare Variant Analysis), Copy Number Analysis, Facioscapulohumeral muscular dystrophy (FSHD) Analysis, Fragile X Analysis, and Hybrid Scaffolding. **Table 2**, **Table 3**, **Table 4**, and **Table 5** provide a summary of minimum coverage recommendation for Direct Label and Stain (DLS) data for each application. For a more detailed view of rationale (such as SV sensitivity or CN calling performance at different coverages) please refer to the Theory of Operations guide appropriate to the analysis at https://bionano.com/software-and-data-analysis-support-materials/.

Beginning in Solve 3.8.1, Bionano offers three distinct pipelines for whole genome calling of structural variants, variant annotation, and copy number variants of human samples. For constitutional and germline applications, there are two pipeline options: *de novo* assembly pipeline and Guided Assembly parameterized for constitutional analyses. These two pipelines are intended for downstream analysis of constitutional, germline, and "typically" diploid SVs. It requires a modest amount of coverage (**Table 2**) to assemble large genome maps *de novo*, then aligns maps to a provided reference. Genome structure is elucidated, and SVs called are generally assumed to be homozygous or heterozygous.

Table 2. Recommended minimum input for constitutional human SV analysis

Pipelines	SV Target VAF	Minimum Pre-Analysis Effective Coverage (X)	hg19 / hg38 length (Gbp)	Example Map Rate	Min. DNA (Gbp)	Post-Analysis Effective Coverage of Reference(X)
de novo Assembly Guided Assembly	50%	80x	× 3.1	÷ 78%	= 318	≈70
de novo Assembly Guided Assembly	20%	160x	× 3.1	÷ 78%	= 636	≈140

Bionano has observed extremely high sensitivity and confidence (PPV) at 80X effective coverage pre-analysis for heterozygous and homozygous SVs in non-mosaic cases. 80X is a recommended minimum, which translates to

320Gbp of DNA collected for a typical run with good molecule metrics. More coverage will yield additional sensitivity for mosaic SVs. For human constitutional analyses inclusive of mosaicism to ~20% VAF, it is recommended that 640Gbp of DNA collected for a typical run with good molecule metrics. If a different effective coverage is desired, simply substitute a new coverage target in the calculation table above.

In contrast, analysis of cancer, somatic, or otherwise heterogenous samples is achieved with either the Rare Variant Analysis (RVA) or Guided Assembly parameterized for Low Allele Fraction (GA LAF) pipelines. Somatic analysis leverages deeper input coverage (**Table 3**) to target SVs at low allelic fractions. The pipeline directly aligns molecules to hg19, hg38, or T2T to capture SVs with a relatively low fraction of molecule support.

Table 3. Recommended minimum input for somatic analysis, targeting ≥5% Variant Allele Fraction

Pipelines	Minimum Pre-Analysis Effective Coverage (X)	hg19 / hg38 length (Gbp)	Example Map Rate	Min. DNA (Gbp)	Post-Analysis Effective Coverage of Reference(X)
Guided Assembly LAF Rare Variant Analysis	340†	× 3.1	÷ 78%	= 1351	≈300†

†Bionano has observed at least 90% sensitivity and PPV to structural variants at 5% variant allele fraction (VAF), with datasets at 300X effective coverage of reference post-analysis. Data input of lower coverage is acceptable, but VAF sensitivity will decrease. This performance is shown in detail in *Theory of Operation – Structural Variant Calling* (CG-30110).

To ensure adequate effective coverage, we generally recommend a Throughput target of 1500 Gbp for datasets intended for somatic analysis targeting ≥5% Variant Allele Fraction. The pipeline has been validated with datasets of up to 5000 Gbp Total DNA (>=150 kbp). This extremely high degree of coverage can improve sensitivity beyond typical low variant allele frequency applications (false discovery rate at 5000 Gbp has not been extensively examined). If a different effective coverage is desired, simply substitute a new coverage target in the calculation table above.

The Bionano EnFocus™ FSHD Analysis and EnFocus™ Fragile X Analysis use targeted assembly approaches. The pipelines define minimum acceptable criteria for data input and performs a PASS/FAIL assessment of molecule quality. The whole genome coverage recommendation provided is a safe minimum shown to confidently resolve targeted regions. Criteria are shown in **Table 4**. Of note, the post-auto noise correction effective coverage will likely be affected by the quality of the sample.

Table 4. EnFocus™ FSHD Analys is and EnFocus™ Fragile X Analysis Minimum Data Criteria

Analysis	Pipeline	Post Auto-noise Correction Effective Coverage (X)	Map Rate	N50 (>= 150 kbp)	Recommended Min. DNA (Gbp)
FSHD Analysis	EnFocus™ FSHD Analysis	≥75	≥70%	≥200 kbp	=400
Fragile X Analysis	EnFocus™ Fragile X Analysis	≥75	≥70%	≥200 kbp	=400

Finally, the hybrid scaffolding pipeline utilizes one or two *de novo* assembly output(s) to scaffold NGS contigs/scaffolds. It is normally used with plant and animal haploid NGS assemblies, frequently of genomes that are otherwise poorly characterized. The first step in a hybrid scaffolding workflow is collecting data to support *de novo* assembly to be used as input alongside NGS data. The assembly coverage recommendation is what is required to build contiguous and accurate consensus genome maps (**Table 5**).

Table 5. *De novo* Assembly recommended minimum input targeting downstream use with Hybrid Scaffolding, using Saphyr data and reference derived from sample matched haploid NGS with reasonable completeness and contiguity.

Pipeline	Minimum Pre-Analysis Effective Coverage (X)*	Haploid genome size (Gbp)	Example Map Rate*	Min. DNA (Gbp)	Post-Analysis Effective Coverage of Reference (X)
de novo Assembly	80	× <u>1.1</u>	÷ <u>55%</u>	= <u>160</u>	≥70

*Calculation of Map Rate and Effective Coverage of the Reference (X) relies on quality and completeness of the provided reference, as well as sequence similarity between reference and sample. For example, the Map Rate shown above (55%) would be considered low for human data against the highly complete hg19. It may be sufficient in a non-model system. Following calculation in **Table 5**, reduced map rates could cause an overestimation of Min DNA (Gbp) needed to meet target pre-analysis effective coverage. See example scenario 2 below.

If reference quality is poor or none is available, 100-120X raw haploid coverage can serve as minimum input target.

Genome researchers have enjoyed significant size and accuracy gains from scaffolding NGS contigs/scaffolds with *de novo* assembled Bionano maps. 80X is a recommended minimum for this application. When using a Direct Label and Stain (DLS) approach, effective coverage up to and beyond 100X has shown improved map contiguities for some plants and animals. Expect pipeline output to be highly genome specific. If a different effective coverage is desired, simply substitute a new coverage target in the calculation table above.

Resource considerations

Collecting and/or analyzing high volumes of data incurs costs. These costs include *i.* time a chip spends occupying the Saphyr or Stratys System, *ii.* data storage consumption, and *iii.* tokens (for Compute On Demand operations) and runtime costs of completing analysis pipelines. Runtime estimates provided by Bionano are based on our recommended coverage levels.

CG-30173 Rev.H, Data Collection Guidelines

For Research Use Only. Not for use in diagnostic procedures.   Page **8** of **17**

- As measured by running the Bionano control sample, throughput of G2.3 chips in a Saphyr System can support collecting three datasets of ≥320 Gbp (fit for human *de novo* assembly; **Table 2**) every 8 hours. For targeting ≥1500 Gbp (Rare Variant Analysis; **Table 3**), data collection for three samples generally can take around 24-36 hours. Applications requiring higher data collection will take longer. For example, Bionano has observed collection of full 5000 Gbp datasets assume up to 96 hours. Any time dedicated to collecting a data volume beyond necessary coverage poses some availability cost against collecting the next datasets.

- On the Stratys Instrument, as measured by running the Bionano control sample, throughput can support collecting 12 datasets of ≥ 400 Gbp (fit for Guided Assembly) every 6-10 hours. For targeting ≥1500 Gbp (LAF Guided Assembly), data collection of 12 samples can take around 18-22 hours. Applications requiring higher data collection will take longer.

- Larger volume datasets require larger files which consume storage space faster. Large file sizes may create difficulties with file transfer and data backup. Please see **Table 6** for an approximation of file sizes.

- Submitting higher volume datasets for bioinformatic analyses increases the computational burden. Higher data inputs increase job runtime, imposing an availability cost on analysis goals. For analyses through Compute On Demand, higher coverage than necessary needlessly increases token cost.

If a dataset is larger than necessary for a desired application, it is recommended to downsample it in Bionano Access. See *Bionano Access® Software User Guide* (CG-30142), and example scenario 1b.

Table 6. Representative sizes for different file types and volumes.

Output Type	Extension	Total DNA (>=150 kbp) input	Size* (compressed)
Molecules*	.bnx.gz	400 Gbp	1.1 GB
		800 Gbp	1.6 GB
		1500 Gbp	3.6 GB
Annotated *de novo* Assembly Annotated Guided Assembly	.zip	400 Gbp	2.4 GB
		800 Gbp	4.5 GB
Annotated Rare Variant Analysis	.zip	1500 Gbp	3.8 GB
Annotated Guided Assembly LAF	.zip	1500 Gbp	6.5 GB
EnFocus™ FSHD	.zip	400 Gbp	140 MB
EnFocus™ Fragile X	.zip	400 Gbp	133 MB

***Sizes assume good raw data quality, with reasonably low proportion of molecules < 150 kbp. Datasets with excessive DNA fragmentation may inflate file sizes.

CG-30173 Rev.H, Data Collection Guidelines

For Research Use Only. Not for use in diagnostic procedures.   Page **9** of **17**

Example Scenarios

Scenario 1. Somatic cancer analysis

Research goal: Guided Assembly LAF pipeline targeting ≥5% variant allele fraction.

In this scenario, a cancer researcher wishes to interrogate a sample for structural variants at ≥5% variant allele fraction. The researcher enters deidentified sample information into Bionano Access and sets the Throughput target (Gbp) to 1500. The flowcell completes in 22 hours and generates the following MQR (**Table 7**).

Table 7. MQR Report Details

label	value	description
Version	1	Report Version
Total DNA (≥ 150kbp)	1,527.11 Gbp	Total amount of DNA from molecules that are 150kbp or longer
Maximum molecule length	2.2 Mbp	The longest molecule detected during the chip run.
N50 (≥ 150kbp)	299.18 kbp	N50 of DNA molecules that are 150kbp or longer
N50 (≥ 20 kbp)	232.21 kbp	N50 of the molecules that are 20kbp or longer)
Total DNA (≥ 20kbp)	2,213.93 Gbp	Total amount of DNA from molecules that are 20 kbp or longer
Total DNA (≥ 150kbp and min sites ≥ 9)	1,483.58 Gbp	Same as other Total DNA fields, but molecules must have at least 9 labels
N50 (≥ 150kbp and min sites ≥ 9)	301.06 kbp	Same as other N50 fields, but molecules must have at least 9 labels
Label color	BNGFLGR001	Label color used for detection.
Site	CTTAAG	Recognition sequence of the enzyme used.
Average label density (≥ 150kbp)	15.93 /100kbp	Average number of labels per 100 kbp for the molecules that are 150kbp or longer
Enzyme	DLE-1	Name of the enzyme used in this sample.
Reference	hg38_DLE1_0kb_0labels_masked_YPARs.cmap	Name of the reference genome this sample was aligned to.
Reference Length	3,088,269,832 bp	Total length of reference sequence

CG-30173 Rev.H, Data Collection Guidelines

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

Page **10** of 17

label	value	description
Effective coverage	395.40	Total amount of aligned DNA divided by the size of the reference genome times the map rate.
Map rate	82.3 %	Percentage of molecules that are 150kbp or longer mapped to the reference
Label color	BNGFLGR001	Label color used for detection.
Scaling SD	0	Linear term in sizing error relative to reference
Site SD	0.11	Constant term in sizing error relative to reference
Relative SD	0.021	Quadratic term in sizing error relative to reference
Base pairs per pixel	487.7	Calculated base pairs per pixel in the alignment by comparing molecules to the reference.
Positive label variance (PLV)	6.78	Percentage of labels absent in reference
Negative label variance (NLV)	10.7	Percentage of reference labels absent in molecules
Molecule integrity number	0.11	Measure of molecule quality

Referring to **Table 3**, the researcher notes that pre-analysis effective coverage has reached the minimum for the application. The research team proceeds with Guided Assembly LAF or Rare Variant Analysis. When processing the analyzed dataset in VIA, she selects the processing settings and *in silico* reference file appropriate for her genome build and coverage tier (e.g., *hg38_female_300x_masked.MSR* or *hg38_male_300x_masked.MSR*) to deploy the VIA algorithm for CNV and AOH event calling with automated variant classification.

Scenario 2. Hybrid scaffolding a small plant genome

Research goal: *de novo* Assembly to support Hybrid Scaffolding a small plant genome.

An agricultural center wants to improve the reference genome of a novel plant species. The NGS assembly that will be scaffolded is not available at the time Bionano data is collected. However, reference sequences are available for related plants. The most similar one is chosen for *in silico* Digestion. See *Bionano Access™ Software User Guide* (CG-30142). The Files Summary produced is assessed for site density with DLE-1:

```
Channel 1 site density (sites/100kbp): 26.0

Channel 1 estimated label density (labels/100kbp) for Saphyr:    19.2

Channel 1 estimated label density (labels/100kbp) for Irys:      16.8
```

As site density falls within the recommended 8 – 31 / 100kbp range for direct labeling - the cmap is added to references in Access, and a technician labels plant DNA with DLE-1.

The researchers expect a haploid genome size of 850 Mbp and decide to target 110X raw coverage. They enter sample information into Bionano Access and set the Throughput target (Gbp) to 94. The flowcell completes in 5 hours and generates the following MQR. (**Table 8**):

Table 8. MQR Report Details

label	value	description
Reference	similar plant_DLE1_0kb_0labels.cmap	Name of the reference genome this sample was aligned to.
Reference Length	850,399,333 bp	Total length of reference sequence
Enzyme	DLE-1	Name of the enzyme used in this sample.
Site	CTTAAG	Recognition sequence of the enzyme used.
Maximum molecule length	1.98 Mbp	The longest molecule detected during the chip run.
N50 (>= 20 kbp)	124.44 kbp	N50 of the molecules that are 20kbp or longer)
Total DNA (>= 20kbp)	225.4 Gbp	Total amount of DNA from molecules that are 20 kbp or longer
N50 (>= 150kbp)	207.42 kbp	N50 of DNA molecules that are 150kbp or longer
Total DNA (>= 150kbp)	94.52 Gbp	Total amount of DNA from molecules that are 150kbp or longer
N50 (>= 150kbp and min sites >=9)	211.75 kbp	Same as other N50 fields, but molecules must have at least 9 labels
Total DNA (>= 150kbp and min sites >= 9)	94.0 Gbp	Same as other Total DNA fields, but molecules must have at least 9 labels
Map rate	29.9 %	Percentage of molecules that are 150kbp or longer mapped to the reference
Effective coverage	32.85	Total amount of aligned DNA divided by the size of the reference genome times the map rate.
Average label density (>= 150kbp)	18.19 /100kbp	Average number of labels per 100 kbp for the molecules that are 150kbp or longer
Site SD	0.1	Constant term in sizing error relative to reference
Relative SD	0.024	Quadratic term in sizing error relative to reference
Scaling SD	0	Linear term in sizing error relative to reference
Molecule integrity number	0.1	Measure of molecule quality

CG-30173 Rev.H, Data Collection Guidelines

For Research Use Only. Not for use in diagnostic procedures.   Page **12** of **17**

label	value	description
Negative label variance (NLV)	23.4	Percentage of reference labels absent in molecules
Base pairs per pixel	500.4	Calculated base pairs per pixel in the alignment by comparing molecule
Label color	BNGFLGR001	Label color used for detection.
Version	1	Report version
Positive label variance (PLV)	18.33	Percentage of labels absent in reference

The map rate of the sample to this reference is poor (<30%), with positive and negative label variance both rather high. The label density is within ∼15% of expectation. These factors suggest labeled DNA quality may be relatively good, with sample-reference mismatch causing low map rate and coverage estimate problems. Sequence dissimilarity between sample and reference, and/or poor reference assembly contiguity, can cause these issues in non-model organisms.

The operator may continue to collect data in the flowcell or begin analysis with what he has collected. To account for the uncertainty, the operator decides to do both. A *de novo* Assembly is launched on the dataset; here is output from the assembly report (**Table 9** and **Table 10**):

Table 9. Molecules aligned to the reference.

label	value	description
Total number of molecules aligned	310769	The number of molecules after filtering (≥ 150 kbp) that align either to the reference file (.cmap), (e.g., GRCh37 or GRCh38) or the assembly.
Fraction of molecules aligned	0.3	The proportion of filtered molecules that align to the reference or consensus genome maps (assembly only).
Effective coverage of reference	32.92	The total length of molecules divided by the length of the reference or consensus assembled maps after *de novo* assembly.
Average confidence	21.1	The average alignment score for all the molecules that align to the reference or assembly.

Table 10. Molecules aligned to the assembly

label	value	description
Total number of molecules aligned	857354	The number of molecules after filtering (≥ 150 kbp) that align either to the reference file (.cmap), (e.g., GRCh37 or GRCh38) or the assembly.
Effective coverage of assembly	82.489	The total length of molecules divided by the length of the reference or consensus assembled maps after *de novo* assembly.
Average confidence	36.2	The average alignment score for all the molecules that align to the reference or assembly.

The researchers refer to *Assembly Report Guidelines* (CG-30255) to evaluate their result and observe that molecule alignment against the assembly is much improved over alignment to the reference. Also, the assembly size was consistent with their expectation. They note that this reference is a poor guide for the sample's map rate and pre-analysis effective coverage. The completed assembly looks adequate, and more data is incoming from the running flowcell if added coverage is desired. In the absence of a good reference, they may also consider running the *de novo* assembly with the pre-assembly option enabled.

Scenario 3. Constitutional trio analysis

Research goal: *de novo* Assembly or Guided Assembly in human datasets for non-mosaic analysis to support Variant Annotation: trio. One dataset map rate is lower than expected.

In this example, a researcher wants to analyze germline SVs in a child with apparent genetic disease. The researcher plans a trio analysis of the proband and both parents. The lab draws three blood samples, which are labeled with DLE-1. The researcher adds a G2.3 chip to the experiment and populates the three flowcells with deidentified mother/father/proband information. Throughput target (Gbp) is set to <u>320</u> for each. The chip completes in 6 hours, and generates the following three map rates and pre-analysis coverages:

label	value	description
Map rate	85.8%	Percentage of molecules that are 150kbp or longer mapped to the reference
Effective coverage	92.52	Total amount of aligned DNA divided by the size of the reference genome times the map rate.

label	value	description
Map rate	83.3%	Percentage of molecules that are 150kbp or longer mapped to the reference

CG-30173 Rev.H, Data Collection Guidelines
For Research Use Only. Not for use in diagnostic procedures.

Page **14** of **17**

label	value	description
Effective coverage	89.9	Total amount of aligned DNA divided by the size of the reference genome times the map rate.

label	value	description
Map rate	71.9%	Percentage of molecules that are 150kbp or longer mapped to the reference
Effective coverage	75.21	Total amount of aligned DNA divided by the size of the reference genome times the map rate.

The Saphyr operator observes that one sample finished below the recommended 80X effective coverage minimum. Its map rate is somewhat lower than expectation, which reduces input coverage as per calculation in **Table 2**. During data collection, its map rate stabilized lower than the other two samples.

The Throughput target was adjusted accordingly. The flowcell was rehydrated and run by itself to a new total Throughput target (Gbp) of 400. The additional runtime completes in 90 minutes, and the run generates the following total dataset map rate and pre-analysis coverage:

label	value	description
Map rate	71.8%	Percentage of molecules that are 150kbp or longer mapped to the reference
Effective coverage	93.7	Total amount of aligned DNA divided by the size of the reference genome times the map rate.

With all three datasets now ≥80X effective pre-analysis coverage, the research team is satisfied all SVs annotated in the trio will have been generated with comparable sensitivity. They begin analyses with the *de novo* Assembly and Variant Annotation Pipelines.

CG-30173 Rev.H, Data Collection Guidelines
For Research Use Only. Not for use in diagnostic procedures.

Page **15** of **17**

Technical Assistance

For technical assistance, contact Bionano 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@bionano.com
Phone	Hours of Operation: Monday through Friday, 9:00 a.m. to 5:00 p.m., PST US: +1 (858) 888-7663 Monday through Friday, 9:00 a.m. to 5:00 p.m., CET UK: +44 115 654 8660 France: +33 5 37 10 00 77 Belgium: +32 10 39 71 00
Website	www.bionano.com/support
Address	Bionano, Inc. 9540 Towne Centre Drive, Suite 100 San Diego, CA 92121

CG-30173 Rev.H, Data Collection Guidelines

For Research Use Only. Not for use in diagnostic procedures.　　　　　　　　　　　　　　　　Page **16** of **17**

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, Inc. 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, Inc. 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, Inc. Customer Support for the latest information.

BIONANO GENOMICS, INC. 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, INC. 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, INC. 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 logo and names of Bionano products or services are registered trademarks or trademarks owned by Bionano Genomics, Inc. ("Bionano") in the United States and certain other countries.

Bionano™, Bionano Genomics®, Saphyr®, Saphyr Chip®, Bionano Access™, Stratys™, Stratys™ Compute, Stratys™ Chip, 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 is given or implied. Users are not permitted to use these trademarks without the prior written consent of Bionano. 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 2024 Bionano Genomics, Inc. All rights reserved.