Saphyr Molecule Quality Report Guidelines

Document Number: 30223

Document Revision: B

Table of Contents

Legal Notice

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

Saphyr Molecule Quality Report Guidelines

Important: The guidelines described herein are based on internal experiences at Bionano Genomics and are provided as-is. The purpose of this technical note is to provide guidelines to customers who want to evaluate the quality of data generated from the Saphyr System. For questions, please contact the Technical Support Team at support@bionanogenomics.com.

Molecule Quality Report (MQR)

The Molecule Quality Report provides a summary report on the molecule quality. The report is generated based on results from a molecule-to-reference alignment. The RefAligner tool aligns Bionano molecules to a given reference and identifies regions of similarity between Bionano molecules and reference. The input includes the molecule BNX file and the reference CMAP file.

The MQR identifies and outputs the best alignment of each molecule to the reference, provided that the alignment meets the minimum alignment quality criteria.

To determine if the data quality is sufficient to continue data collection and to run *de novo* assembly, the best indicators* are the following:

1) Map rate: What percentage of the Bionano molecules aligns to the reference (meeting minimum alignment quality criteria)?
2) Noise parameters: How different are the aligned Bionano molecules when compared to the reference?

***** The evaluation of the MQR results is highly dependent on the accuracy and completeness of the given reference and the identity of the sample with the reference. Many sequence assemblies, even at advanced stages, could have a high degree of structural inaccuracy that may compromise the use of the MQR. See the *Interpret Molecule Quality Report Results* section for details.

Molecule Quality Report Metrics

1. The metric results of MQR in Bionano Access are the following:

Metrics	Description
MQR Report	Format version.
Sample	Sample name given by the user when the experiment is created or the molecules object is imported.
Flowcell	Flowcell number of this sample.
Job ID	Job number of this object.
Job created on	Date and time of this job created on.
Total DNA (>= 20kbp)	Total amount of DNA from molecules that are 20 kbp or longer.
N50 (>= 20kbp)	N50 of the DNA molecules that are 20 kbp or longer.
Total DNA (>= 150kbp)	Total amount of DNA from molecules that are 150 kbp or longer.
N50 (>= 150kbp)	N50 of DNA molecules that are 150 kbp or longer.
Total DNA (>= 150kbp & minSites >= 9)	Total amount of DNA from molecules that are 150 kbp or longer and contain at least 9 labels.
N50 (>= 150kbp & minSites >= 9)	N50 of DNA molecules that are 150 kbp or longer and contain at least 9 labels.
Enzyme	Enzyme name for this sample.
Enzyme recognition sequence	Enzyme recognition sequence for this sample.
Label color	Color of the label.
Average label density (>= 150 kbp)	Average number of labels per 100 kbp for total DNA from molecules 150 kbp or longer.
Reference	Reference name for this sample.
Map Rate (>= 150 kbp)	Percentage of DNA molecules that are 150 kbp or longer which align to the reference.
Effective coverage	Total amount of aligned DNA divided by the size of the reference genome. Please see more details in *Data Collection Guidelines* (30173).
Molecule integrity number	Measure of molecule quality. Smaller number is better.
Positive label variance (PLV)	Percentage of molecule labels absent in reference labels.
Negative label variance (NLV)	Percentage of reference labels absent in molecule labels.
SiteSD	These parameters describe which inter-label distances in Bionano molecules match the reference. This is referred to as sizing error relative to reference. These parameters are components of the variance of the distance observed in the Bionano molecules for a given interval (distance between two labels) on the reference. variance(x) = sf^2 + x|sd|sd + x^2*sr^2 The **x** value is the interval length (kbp) and the other noise parameters are reported in the ERR file. SMin is not reported in the ERR file; it is either the minimum value of sqrt(variance) at x > 1 kbp OR the value of sqrt(variance) at x = 1 kbp, whichever is lower.
ScalingSD	
RelativeSD	
SMin	
Base pairs per pixel	Calculated base pairs per pixel in the alignment by comparing molecule intervals to reference intervals.

Interpret Molecule Quality Report Results

1. List of metrics with ranges based on Bionano internal human data (source: Saphyr instrument).

Table 1: NLRS Labeling

Metrics	**Range**
N50 (>= 20kbp)	> 130 kbp
N50 (>= 150kbp)	> 220 kbp
Map Rate (%)	60-90%
Average label density (>= 150 kbp)	9-12
Effective coverage	>80
Molecule integrity number	<20
Positive label variance	< 15%
Negative label variance	< 20%
SiteSD	< 0.25
ScalingSD	(-0.07)-0.05
RelativeSD	< 0.04
SMin	< 0.25
Base pairs per pixel	450-510

Table 2: DLS Labeling

Metrics	**Range**
N50 (>= 20kbp)	> 150 kbp
N50 (>= 150kbp)	> 230 kbp
Map Rate (%)	70-90%
Average label density (>= 150 kbp)	14-17
Effective coverage	>80
Molecule integrity number	<20
Positive label variance	3-10%
Negative label variance	6-15%
SiteSD	< 0.25
ScalingSD	(-0.07)-0.05
RelativeSD	< 0.04
SMin	< 0.25
Base pairs per pixel	450-510

2. Interpret the MQR Results

To interpret MQR results, check the molecule-to-reference map rate (%) first. The map rate is also closely tied to the completeness and accuracy of the reference (i.e. fraction of the genome that is assembled into large contigs or scaffolds in the reference, and how much error, such as repeat collapse, is in the reference assembly). Additionally, the map rate depends on the degree of identity of the Bionano sample with the reference sample (i.e. is the sample from the same individual as the reference?).

For example, the human reference is highly complete, so the map rate can be higher than 90% for a good molecule dataset. If the reference or sequence assembly is only 50% complete, then the expected map rate range may be half or 30-40%, even if the Bionano molecules are of good quality.

If the obtained map rate is significantly lower than the minimum desired map rate (i.e., < 60% for high quality reference or 30% for half-complete reference), check the noise parameters. If the noise parameters are within the recommended range, it could mean that the part of the Bionano data that does align to the reference is of good quality. In this case, these molecules can be used for *de novo* assembly; however, users may need to collect extra depth of the same data to compensate for low mapping rate.

When interpreting the results, it is important to consider the accuracy of the provided reference. However, evaluating the reference accuracy is often challenging. If the map rate is lower than expected based on the completeness of the reference, it is possible that the molecule quality is still good, but because of the inaccuracy of the reference, some molecules do not align. In this case, it is difficult to evaluate molecule quality using MQR.

Another way to evaluate alignment between the Bionano molecules and reference CMAP is to view alignments in Bionano Access. The aligned molecules should cover most of the reference genome (or reference contigs) relatively uniformly and without large errors (see Figure 1).

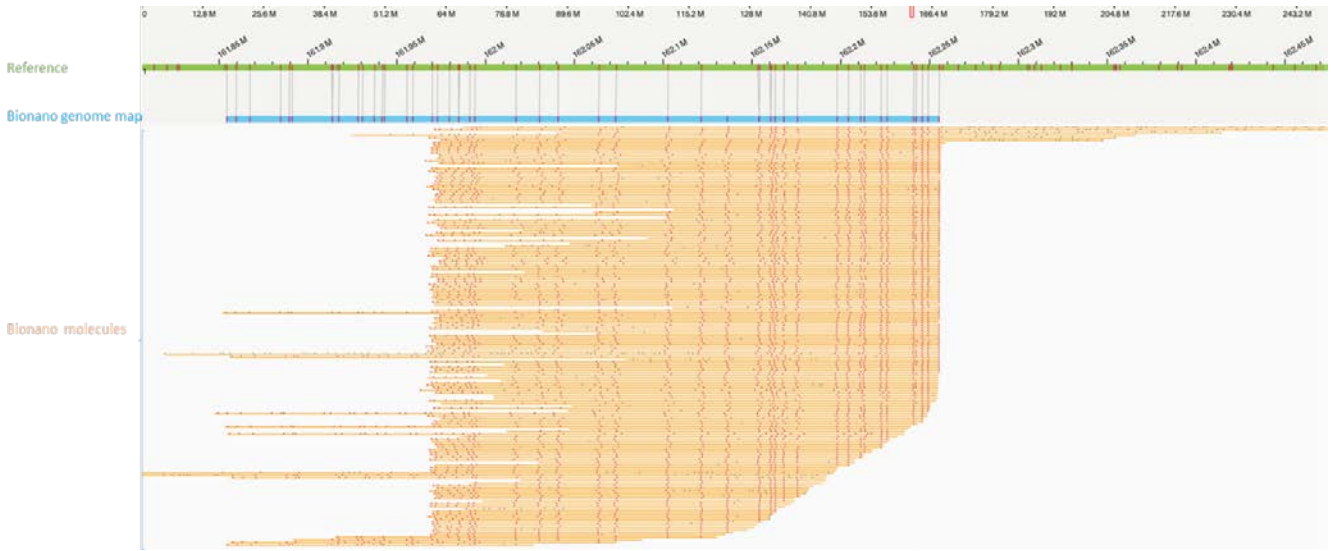

Figure 1*: An example of a good alignment.*

In cases where it is difficult to evaluate the reference completeness or accuracy or when it is difficult to obtain reliable noise parameters from MQR, we recommend that users perform *de novo* assembly using the pre-assembly option and default noise parameters starting with at least 100X coverage data (refer to the tutorial video *Assembly Objects* that describes how to do that). If most of the genome (> 80%) can be assembled with reasonable contiguity, the expected assembly length, and acceptable molecule alignments (i.e. a good alignment of the Bionano molecules to the assembled map as visualized; see Figure 1), then the data is likely to be of good quality.

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