Data Collection Guidelines

Document Number: 30173

Document Revision: B

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™, 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.

Data Collection Guidelines

Introduction

The key factors for success in using any of the Bionano systems include:

- Design experiments to meet your research goals;
- Isolate high molecular weight gDNA from sample sources;
- Label DNA using proper nicking (NLRS) or direct labeling (DLS) enzymes;
- Collect high quality data for downstream data analysis.

Data Collection

For organisms that fall within the recommended site density (recommended density is 8 to 15 per 100 kbp using nickases or 8 to 25 per 100 kbp for direct labeling enzymes), the recommended coverage is detailed below. If your sample falls outside the recommendated site density range, please contact Bionano Support.

Coverage is used to estimate the amount of input data and is calculated from the predicted genome size and the amount of filtered input data (i.e. molecules with molecule length >150 kbp and proper label SNR filter).

$$\text{Coverage} = \frac{\text{Total amount of filtered data}}{\text{Predicted genome size}}$$

For example, 100 Gbp of filtered data collected for a sample with a genome size of 1.0 Gbp has 100x depth of coverage (100 Gbp ÷ 1.0 Gbp=100).

Generally, for the Saphyr™ system, we recommend collecting a minimum of 100x filtered data for structural variation (SV) analysis and 70x filtered data for hybrid scaffolding. For the Irys® system, we recommend collecting 120x filtered data for SV analysis and 80x for hybrid scaffolding. Furthermore, for both applications, we recommend that either 1 direct labeling enzyme or 2 nicking enzymes be used to achieve the best SV sensitivity. For the best hybrid scaffold contiguity, we recommend two enzymes be used, with one being a DLS enzyme. Structural variation analysis and hybrid scaffolding with 2 enzymes requires 2 separate Bionano *de novo* assemblies generated from 2 individually labeled samples processed with different enzymes.

To evaluate data quality, refer to the *Guidelines for Interpreting Bionano Genomics Molecule Quality Report (Document # 30175)* and *Guidelines for IrysView Run Metrics Report (Document # 30176),* which are for the Irys system only and can be accessed at https://bionanogenomics.com/support-page/irysview/

We used a human cell line sample as an example below to illustrate how to estimate the minimum amount of data to collect and the relevant consumables required. The example below takes into account a human cell line sample prepared with only one enzyme. To perform 2-enzyme SV or 2-enzyme hybrid scaffold, equal amounts of data for both labeled samples must be collected, doubling the consumables required for DNA labeling and data collection.

Table 1. Minimum amount of data to collect

(Human cell line sample prepared with single labeling enzyme; human genome size is ~3.2 Gbp)

	SV Analysis			Hybrid Scaffolding		
	Irys (NLRS)	Saphyr (NLRS)	Saphyr (DLS)	Irys (NLRS)	Saphyr (NLRS)	Saphyr (DLS)
Target raw coverage	120x	100x	100x	80x	70x	70x
Target effective coverage[a]	70x	70x	70x	50x	50x	50x
Target filtered throughput	384 Gbp	320 Gbp	320 Gbp	256 Gbp	224 Gbp	224 Gbp

Table 2. Throughput estimates of Irys and Saphyr systems

(Human cell line sample prepared with single labeling enzyme)

	Irys	Saphyr
Typical filtered throughput per flowcell per run	36 Gbp[b]	320 Gbp
Number of flowcells per chip	2	2
Typical filtered throughput per chip per run	72 Gbp	640 Gbp

Table 3. Estimate of Bionano consumable need for Irys system and Saphyr system

(Human cell line sample prepared with single labeling enzyme)

		Irys		Saphyr	
Procedure	Consumable	Amount	Target Yield	Amount	Target Yield
DNA isolation	Lysis Buffer	5 ml	DNA conc. 35-200 ng/µL; ~70 µl of solubilized DNA per plug	5 ml	DNA conc. 35-200 ng/µl; ~70 µL of solubilized DNA per plug
	5x Wash Buffer	30 ml		30 ml	
DNA Nick-Label-Repair-Stain (NLRS)	Bionano Prep NLRS Kit	3 reactions[c]	Final DNA conc. 3-10 ng/µL; 180 µl of NLRS prep containing ~900 ng labeled DNA	1 reaction[c]	Final DNA conc. 3-10 ng/µl; 60 µL of NLRS prep containing ~300 ng labeled DNA (enough for 3 flowcell runs)
DNA Direct-Label-Stain (DLS)	Bionano Prep DLS Kit	N/A	N/A	1 reaction	Final DNA conc. 4-12 ng/µl; 56 µL of DLS prep starting with ~750 ng labeled DNA (enough for 2 flowcell runs)
Data collection (for SV analysis)	Chip	10 flowcells (Irys Chip)	36 Gbp filtered throughput per flowcell per run, typical	1 flowcell (Saphyr Chip 2x320)	320 Gbp filtered throughput per flowcell per run, typical
Data collection (for hybrid scaffolding)	Chip	7 flowcells (Irys Chip)	36 Gbp filtered throughput per flowcell per run	1 flowcell (Saphyr Chip 2x320)	320 Gbp filtered throughput per flowcell per run

a) The human genome reference is considered relatively complete and accurate. Effective coverage is useful for estimating the amount of high quality data. For human SV detection, the minimum recommended effective coverage is 70x; for human hybrid scaffold application, we recommend at least 50x effective coverage. Effective coverage is calculated as below.

$$\text{Effective coverage} = \frac{\text{Total amount of filtered data} \times \text{Map rate}}{\text{Predicted genome size}}$$

For example, 320 Gbp of filtered data collected for a human sample (genome size= 3.2 Gbp) with 70% map rate has 70x effective coverage (320 Gbp x 70% ÷ 3.2 Gbp = 70).

Typically, we have seen an average map rate for human sample of 60% on the Irys system and an average map rate of 70% on the Saphyr system with NLRS samples, and slightly higher for DLS samples. The bottom row of the table 1 was calculated based on these values.

b) We recommend running only 30 cycles for both flowcells on an Irys Chip to maximize consistency and quality of the data. If users run more than 30 cycles per Irys Chip, the number of Irys Chips consumed for the project would be fewer than what is shown in Table 3.

c) One reaction using the Bionano Prep NLRS labeling kit requires 300 ng of DNA and generates 60 µl of labeled sample ready to load for data collection. Users can scale up the 300 ng reaction to a 900 ng reaction, to generate 180 µl of labeled sample.

When running an Irys Chip, each flowcell requires 16 µl of NLRS sample, which corresponds to 80 ng of NLRS labeled DNA per flowcell. When running a Saphyr Chip 2x320, each flowcell requires 19 µl of NLRS or DLS labeled sample, which corresponds to 95 ng of DNA per flowcell. Irys Chips are not compatible with DLS labeled samples.