10x GENOMICS®

TECHNICAL NOTE

Guidelines for *De Novo* Assembly of Genomes Smaller than ~3 Gb using 10x Genomics® Supernova ™ V1.2

INTRODUCTION

This Technical Note offers guidelines to ensure that read depth per molecule for non-human haploid and diploid genomes ranging from 100 Mb to ~3 Gb is standardized when using the 10x Genomics® Supernova™ v1.2 Assembler. Note, the proposed strategies are experimental and have not been validated extensively.

READ DEPTH PER MOLECULE

Because the Supernova algorithm was initially designed to assemble germline human genomes, it has been tested primarily on ~3.2 Gb genomes with ~1.25 ng of DNA loaded and sequenced, generating 1200 M reads. In this scenario, the coverage of a portion of a molecule in each Gel Bead-In-EMulsion (GEM) looks conceptually like this:

This is the standard read depth per molecule for 10x Genomics® *de novo* assembly: approximately 0.35x[1]. This gives the data strong connective power.

Note[1]*: How is this calculated? About 40%, or 0.5 ng, of the 1.25 ng make it into GEMs. That 0.5 ng is approximately 500 Gb and with 1200 M reads, the result is 1200 M reads / 500 Gb = 2.4 reads per kb, or ~0.35x, for 150 base reads. (Genome size is not a factor.)*

For example, if you had a ten-fold smaller genome (0.32 Gb), and sequenced proportionally fewer reads (120 M), but loaded the same amount (1.25 ng); because input mass is unchanged, fewer reads are distributed across the same molecules:

For this small genome, the result is that the read depth per molecule is ten-fold lower, now about 0.035x. The data lack connective power.

THE SOLUTION: LOAD LESS DNA AND SOMETIMES SEQUENCE DEEPER

For **genomes 1.6 to 3.2 Gb**, lower the DNA input amount and adjust the sequencing proportionately.

- Load 1.25 ng for a 3.2 Gb genome; load half (0.625 ng) for a 1.6 Gb genome. Adjust amounts proportionately for genomes in between.
- Sequence 800-1200 M reads for a 3.2 Gb genome; sequence half as many reads (400-600 M) for a 1.6 Gb genome. (In general, the higher the better within the range.)

For **genomes 0.1 to 1.6 Gb**, load less and sequence deeper.

- Load 0.625 ng for any genome in this smaller range. (Do not go lower as this could result in a low-complexity library.)
- Sequence deeply: 400-600 M reads for any genome in this smaller range.
- Then, use only a fraction of the barcodes by barcode subsampling (see below). This lowers read coverage of the genome to the optimal range for Supernova, 38-56x, while leaving read depth per molecule unchanged.

HOW TO SUBSAMPLE BARCODES

Barcode subsampling randomly selects barcodes at a specified rate, then removes all reads whose barcodes were not selected. To apply barcode subsampling, add the following arguments to the `supernova run` command:

Genome Size (Gb)	Added Args
0.8	`--bcfrac=0.5 --maxreads=300000000`
0.4	`--bcfrac=0.25 --maxreads=150000000`
Other	Adjust proportionally

<100 Mb GENOMES

Genomes less than 100 Mb have not been tested and are not recommended at this time. Such genomes may have issues with molecules from nearby genomic regions co-occurring too frequently within single GEMs.

EXCEPTIONS

Some genomes that are smaller than 1.6 Gb have been satisfactorily assembled simply by lowering loaded DNA to 0.625 ng. However, 10x Genomics® does not recommend this approach, because a drop in read depth per molecule could negatively impact the assembly. Small genomes can have other problems, including repeats or difficulty in preparing high molecular weight DNA. The approach here does not solve these problems.

CONCLUSION

The Chromium™ Genome workflow and Supernova™ assembler were developed for human germline genomes with the baseline input mass of 1.25 ng, and sequencing coverage of 38 to 56x (800-1200M reads). This document provides instructions for assembling smaller genomes, down to 100 Mb. These involve changes to both laboratory and computational protocols, and are designed to standardize read depth per molecule, so that approximately the same value is obtained, regardless of genome size. These instructions are experimental, but are believed to increase the likelihood of successful assembly with Supernova. The smaller the genome, the more barcode subsampling will help.

Supernova has been used successfully to assemble genomes with a range of sizes, however results vary based on the genome characteristics and input material quality. Starting with high molecular weight DNA is critical. 10x Genomics is working to expand the range of genomes over which Supernova performs well, and is actively collecting customer data to assist in this effort.

REFERENCES

- *Introduction to Linked-Read Technology for a More Comprehensive Genome & Exome Analysis* (CG00044)
- *Chromium™ Genome Reagent Kits v2 User Guide*
- Supernova™ Guidance: https://support.10xgenomics.com/de-novo-assembly/sample-prep/doc/technical note-supernova-guidance
- Weisenfeld *et al*. 2017. *Direct determination of diploid genome sequences*. *Genome Research*. 27:1-11.

Notices

Document Number

CG000100 Rev A *Technical Note: Guidelines for De Novo Assembly of Genomes Smaller than ~3 Gb using 10x Genomics® Supernova™ v1.2*

Legal Notices

Customer Information and Feedback

For technical information or advice, please contact our Customer Technical Support Division online at any time.

Email: support@10xgenomics.com

10x Genomics 7068 Koll Center Parkway

Suite 401

Pleasanton, CA 94566 USA