10x GENOMICS®

TECHNICAL NOTE

Chromium™ Genome Application Performance as a Result of DNA Quality and Sequencing Depth

INTRODUCTION

In this Note, we demonstrate the performance of the Chromium™ Genome Product as a function of both DNA quality and sequencing depth. For a thorough description of DNA Quality, please see Technical Note CG00045 on our Technical Support Website. In brief, we use the term "DNA Quality Level" as a simple classification to describe the average size of the DNA sample ***after*** running the Chromium™ Genome Workflow, which is impacted by both the length of the starting gDNA sample (which can be measured on a gel) as well as DNA damage including single-strand nicks (which cannot be measured on a gel). The Chromium™ Genome Protocols generate long-range information across the length of individual DNA molecules. Starting the process with High Molecular Weight Genomic DNA (HMW gDNA) will typically result in better application performance, such as increased haplotype phase block length and ability to call structural variants. This Note addresses the following questions:

1. Given a certain DNA Quality Level, what application performance can I expect? (To estimate the DNA Quality Level of a given sample, see Technical Note CG00045 on our Technical Support Website.
2. Given a certain application performance goal, what DNA Quality Level should I aim for? (For sample prep advice for achieving a particular DNA Quality Level, see Technical Note CG00045 on our Technical Support Website)
3. If I choose to sequence to a lower depth than the recommended 128 Gb, what performance difference can I expect?

Two examples are presented in Fig1, and a summary of performance categories is presented in Fig 2. Some application metrics, including Haplotype Block Size, are dependent on DNA Quality, and some application metrics, like Single Nucleotide Polymorphism (SNP) Sensitivity, are not dependent on DNA Quality. Detailed results are presented in the remainder of the Note. Importantly, all of the results in this Note should be viewed as a guide, rather than a guarantee of system performance. Many long-range phasing metrics are dependent on underlying biology, in addition to DNA quality and sequencing depth. For instance, individuals with higher heterozygosity (a biological parameter, rather than an experimental parameter) will show improved phasing performance, since heterozygous SNPs sites are critical for separating reads into different haplotypes.

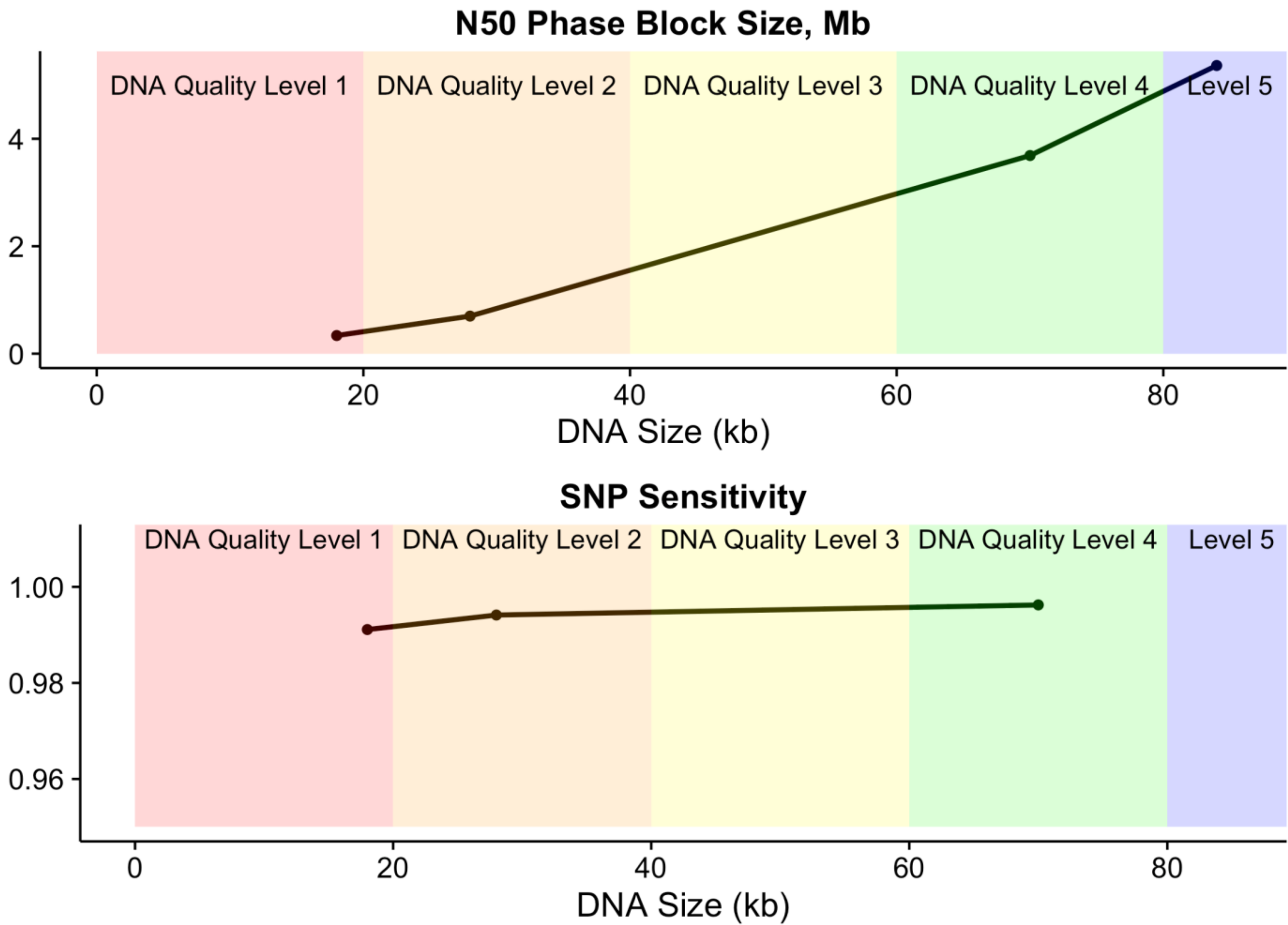

Fig 1. Examples of Chromium™ Genome Performance as a function of DNA Quality. As a general rule, gDNA extraction protocols that do not use heat or harsh buffers produce high quality samples. For more details, please download (Technical Note CG00045 from our Technical Support Website)

Performance Category	DNA Quality: Relative performance with reduction in quality	Sequencing depth: Relative performance with lower sequencing depth
Standard Variant Calling Sensitivity	Blue	Yellow
Standard Variant Calling Positive Predictive Value (PPV)	Blue	Blue
SV >30 kb Sensitivity	Yellow	Red
SV >30 kb PPV	Blue	Blue
SV 50 bp – 30 kb Sensitivity	Blue	Red
SV 50 bp – 30 kb PPV	Blue	Yellow
Phasing (SNP)	Blue	Blue
Phasing (Gene)	Blue	Yellow
Phasing (Phase Block Size)	Red	Red

Long Ranger metrics

Performance Category	DNA Quality: Relative Performance with reduction is quality
Scaffold Size	Red
Phase Block Size	Red
Contig Size	Red

Supernova metrics

Fig 2. Summary of performance as a function of both DNA quality and sequencing depth. Top table: Long Ranger metrics, Bottom table: Supernova metrics. Blue = Little/no dependence, Yellow = Minor-moderate dependence, red = significant dependence.

METHODS

Data in this document was generated using NA12878 cells (Coriell). gDNA was extracted using our modified QIAGEN® MagAttract® HMW protocol in the Chromium™ Genome User Guide (PN CG00022). Since this protocol generates DNA >200 kb, we intentionally sheared the sample to produce a range of shorter DNA sizes, and fractionated the sample using the SageELF™ instrument from Sage Science to create a set of DNA samples that span a wide size range. Samples were processed using the "Genome" Protocol in the User Guide, and sequenced to 128 Gb (and later computationally downsampled) on an Illumina® HiSeq® X Ten sequencer. The average DNA size of each sample is automatically calculated using 10x Genomics software; these values are reported in the following figures. SNP sensitivity was calculated against the Platinum Genome call set available from Illumina®. Software versions: Long Ranger™ 2.1, Supernova™ 2.1.

Seq Depth (Gb)
- 16
- 32
- 64
- 96
- 128

Fig 3. Sequencing depth used in this Note.

RESULTS

Standard Variant Calling (Reference-Based)

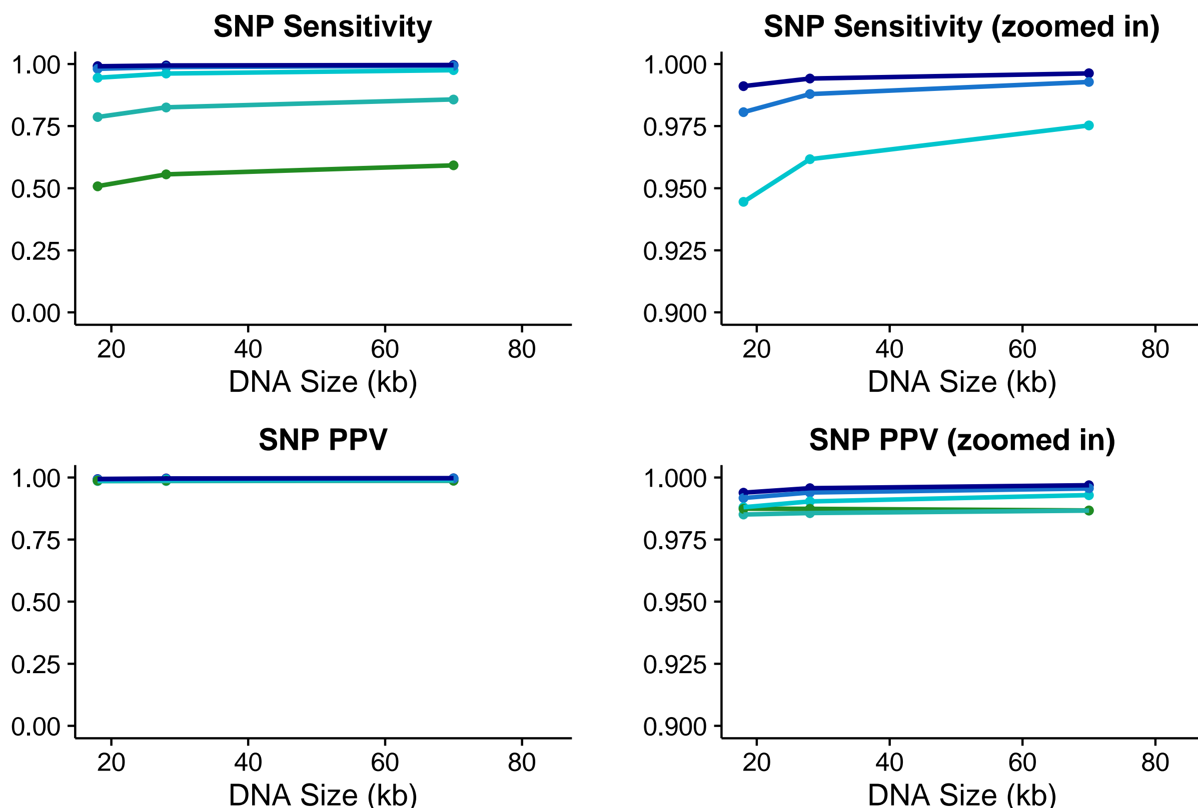

Fig 4. SNPs. SNP sensitivity and PPV were calculated against the Platinum Genome callset. Graphs on the Right Hand Side are Zoomed in versions of the graphs on the Left Hand Side.

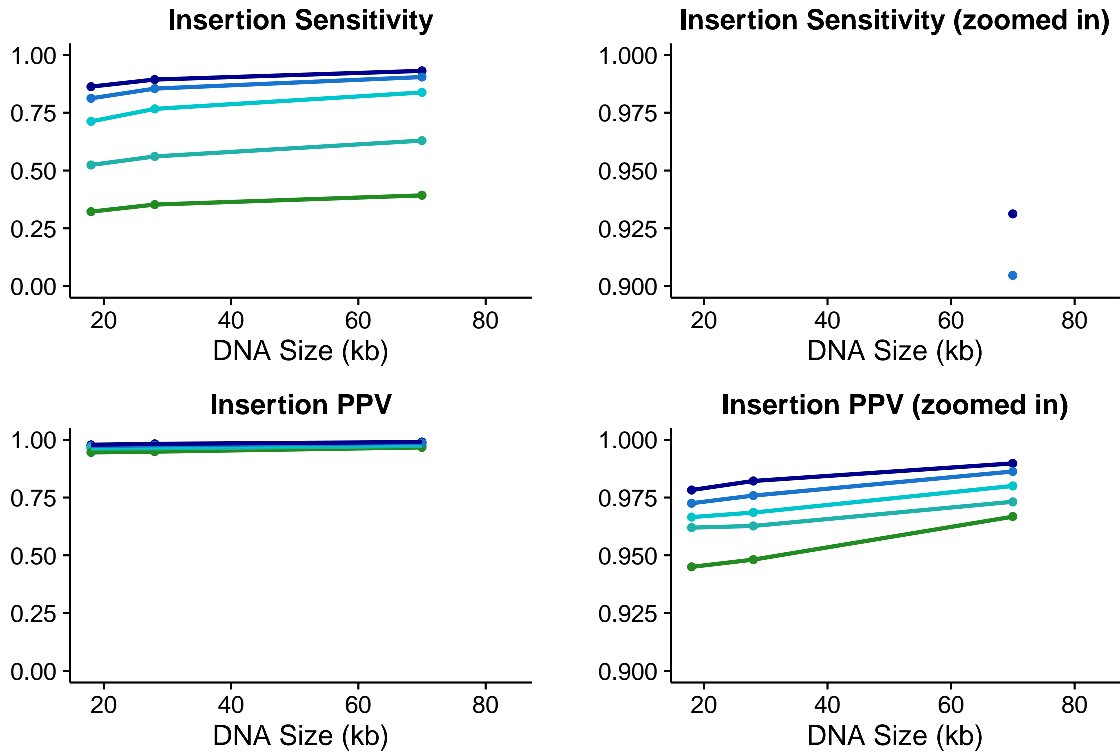

Fig 5. Insertions. Insertion sensitivity and PPV were calculated against the Platinum Genome callset. Graphs on the Right Hand Side are zoomed in versions of the graphs on the Left Hand Side.

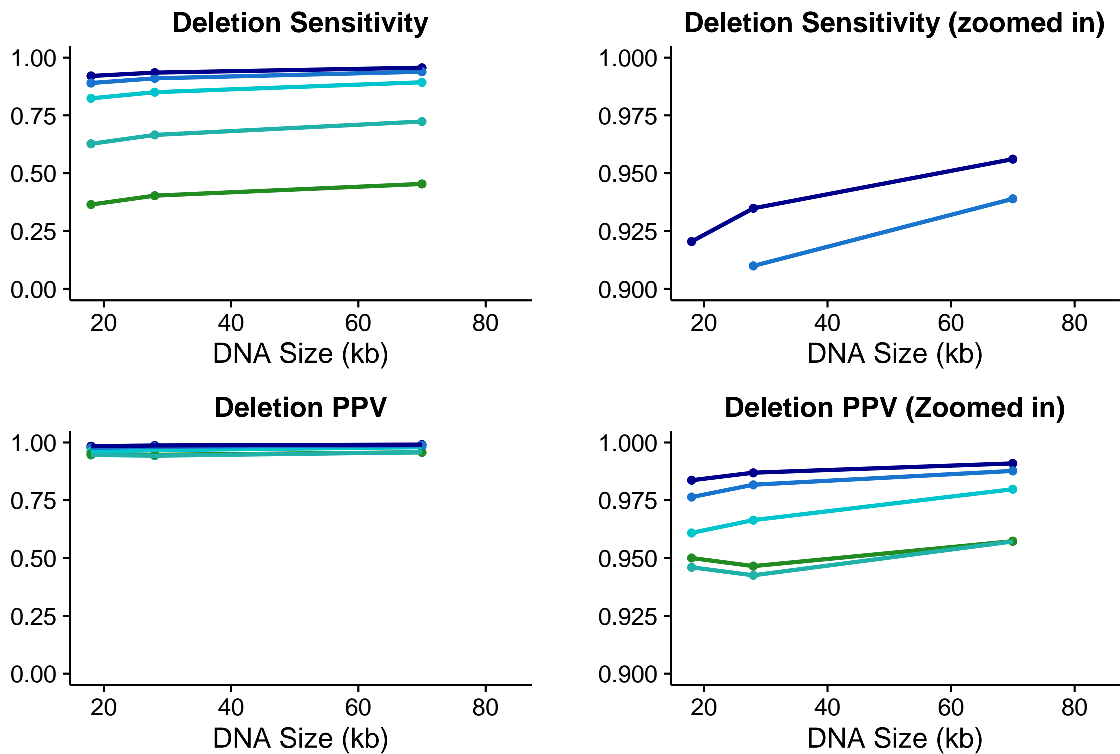

Fig 6. Deletions. Deletion sensitivity and PPV were calculated against the Platinum Genome callset. Graphs on the Right Hand Side are zoomed in versions of the graphs on the Left Hand Side.

Structural Variant Calling (Reference-Based)

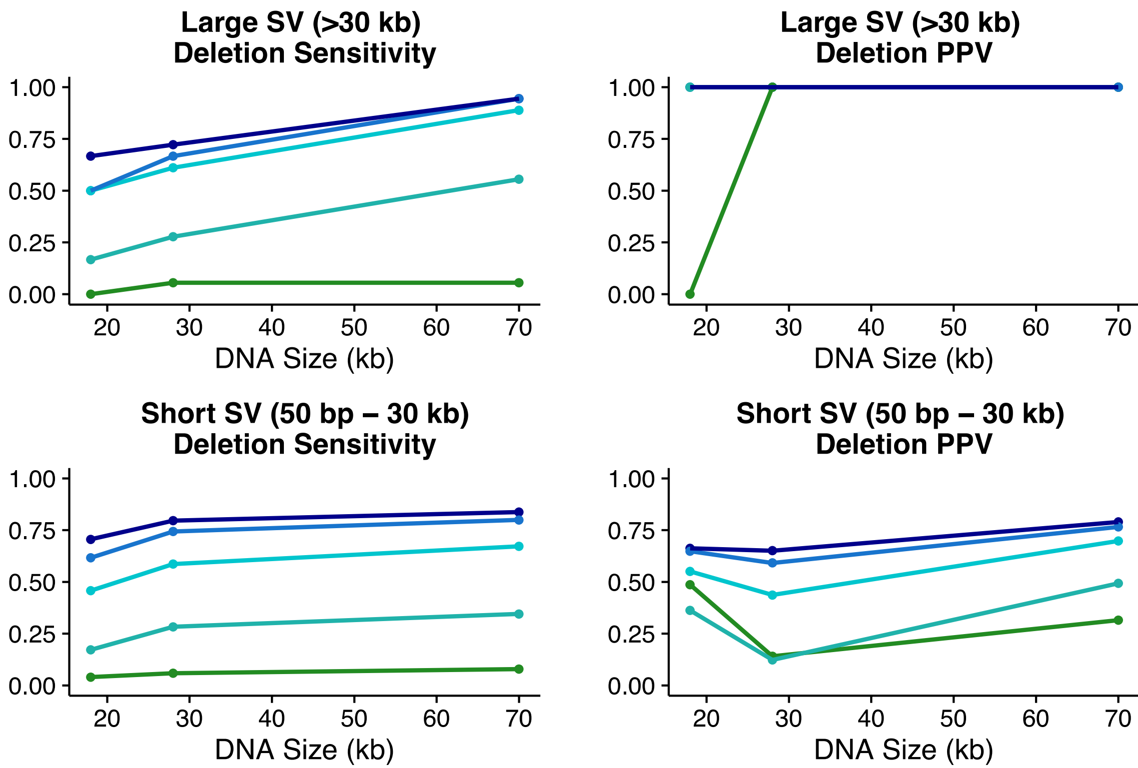

Fig 7. Large (>30 kb) and short (50 bp – 30 kb) structural variants. The ground truth set we used for this analysis consists of 18 large deletions for sensitivity measurement, 23 large deletions for PPV measurement, 2654 short deletions for sensitivity measurement, and 11,198 deletions for PPV measurement. The sensitivity ground truth sets are smaller, with high-confidence orthogonal calls. The PPV ground truth sets are larger, with a more relaxed orthogonal callset.

Phasing (Reference-Based)

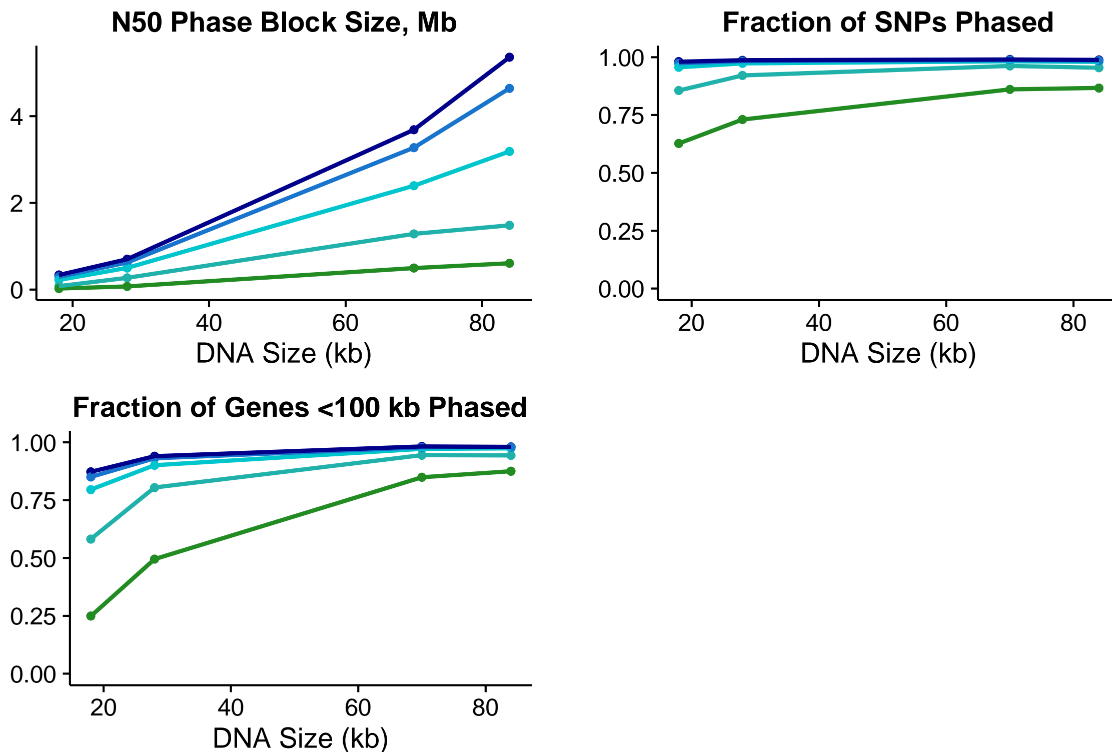

Fig 8. Phasing. Note that these values are dependent on both experimental parameters (DNA Quality, Sequencing Depth) and biological parameters (rate of heterozygosity).

Supernova™ Performance (*de novo* Assembly)

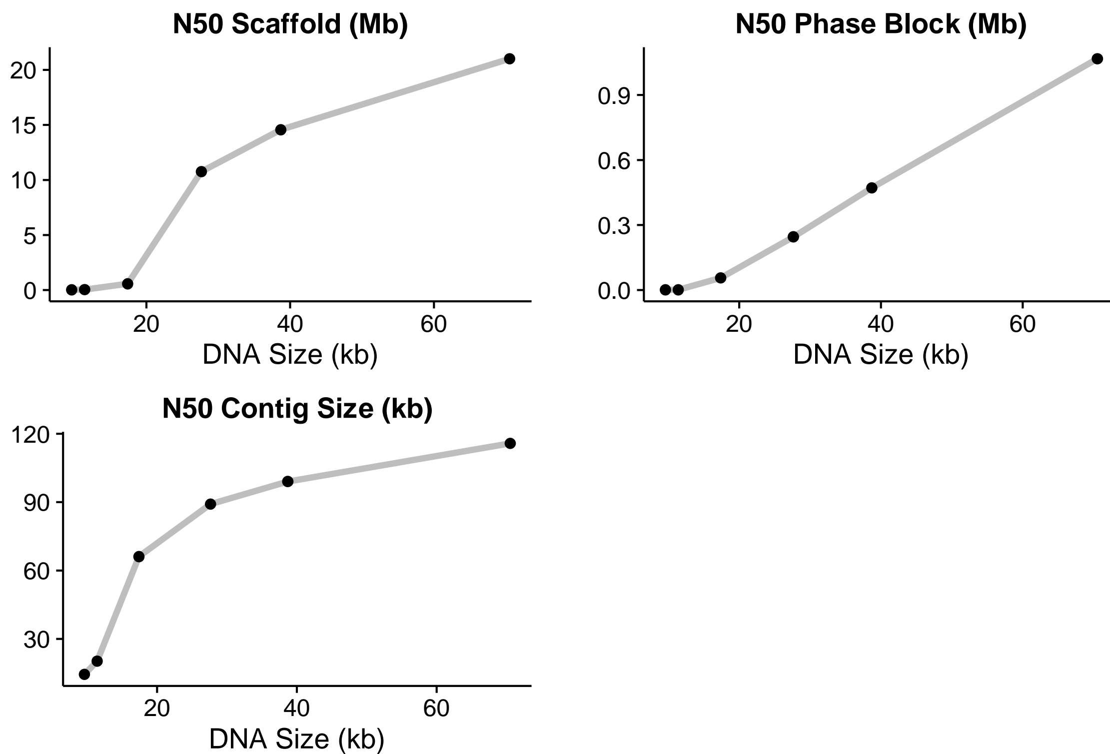

Fig 9. *de novo* assembly results for human genome analysis (800M reads per sample). For more details on our Supernova™ software, including system requirements and definitions of these performance metrics, please download our 10x Genomics Chromium™ *De Novo* Assembly Solution Application Note.

DISCUSSION

The Chromium™ Genome System offers an unprecedented view into genome structure, from Multi-Mb diploid scaffolds down to single nucleotide variants. High-quality genomic DNA will result in optimal system performance. However, lower quality genomic DNA and lower sequencing depth also provide significant value. For reference-based analysis using Long Ranger™ software, good performance is observed across a wide range of both DNA quality and depth (see Fig 2 for a summary and Figs 4-8 for details). For *de novo* assembly, there is a greater dependence on DNA quality (Fig 9). If using Supernova™ with a non-human genome, please visit our Technical Support Website (support.10xgenomics.com) for recommendations. The ideal sequencing depth required for non-human genomes may need to be empirically determined, particularly for organisms with a small genome size. In this Note we report 800M reads per human sample analyzed.

Notices

Document Number

CG00046 Rev A *Technical Note*

Legal Notices

© 2016 10x Genomics, Inc. All rights reserved. Duplication and/or reproduction of all or any portion of this document without the express written consent of 10x Genomics, Inc., is strictly forbidden. Nothing contained herein shall constitute any warranty, express or implied, as to the performance of any products described herein. Any and all warranties applicable to any products are set forth in the applicable terms and conditions of sale accompanying the purchase of such product. 10x Genomics provides no warranty and hereby disclaims any and all warranties as to the use of any third party products or protocols described herein. The use of products described herein is subject to certain restrictions as set forth in the applicable terms and conditions of sale accompanying the purchase of such product. “10x”, “10x Genomics”, “Changing the Definition of Sequencing”, “Chromium”, “GemCode”, “Loupe”, “Long Ranger”, “Cell Ranger” and “Supernova” are trademarks of 10x Genomics, Inc. All other trademarks are the property of their respective owners. All products and services described herein are intended FOR RESEARCH USE ONLY and NOT FOR USE IN DIAGNOSTIC PROCEDURES.

The use of 10x Product(s) in practicing the methods set forth herein has not been validated by 10x, and such non-validated use is NOT COVERED BY 10X STANDARD WARRANTY, AND 10X HEREBY DISCLAIMS ANY AND ALL WARRANTIES FOR SUCH USE.

Nothing in this document should be construed as altering, waiving or amending in any manner 10x Genomics, Inc., terms and conditions of sale for the Chromium™ Controller, consumables or software, including without limitation such terms and conditions relating to certain use restrictions, limited license, warranty and limitation of liability, and nothing in this document shall be deemed to be Documentation, as that term is set forth in such terms and conditions of sale. Nothing in this document shall be construed as any representation by 10x Genomics, Inc that it currently or will at any time in the future offer or in any way support any application set forth herein.

Customer Information and Feedback

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

Email: techsupport@10xgenomics.com

10x Genomics 7068 Koll Center Parkway

Suite 401

Pleasanton, CA 94566 USA