SINGULAR GENOMICS

G4™ BEST PRACTICES AND QUALITY CONTROL GUIDE

Table of Contents

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

Document # 600011 Rev. 0, February 2023

Introduction

There are several factors that influence a good sequencing experiment. These can be grouped in the following categories:

- Constructing the library.
- Designing the sequencing run.
- Preparing the sequencing run.
- Performing quality control.

This document lists many of the details you should consider when designing the experiment. Review the information below before planning a sequencing experiment.

Constructing a G4 Library

Use the information below to prepare a G4-compatible sample library, taking into consideration the type of experiment and analysis. Always refer to the library prep manufacturer's instructions for final guidance.

Many library prep manufacturers have partnered with Singular Genomics to make sure their library prep kits work with the G4. For more information, see https://singulargenomics.com/g4/workflow/.

Library Characteristics

There are several considerations when selecting and performing a library preparation method. The main considerations are listed below.

Sample Type

Libraries can be prepared using a variety of specimens, like whole blood, plasma, or tissue. Ensure that the sample preparation method is well-suited for the intended type of analysis and application. For example, for variant detection in FFPE samples, you may consider a preparation method to overcome the deamination of cytosine or adenosine.

Sample Quality

It's important to have a high-quality DNA or RNA sample before starting the library preparation protocol. Always follow the recommendations of your DNA or RNA isolation kit manufacturer or protocol to get the best quality possible.

Some types of input material are inherently challenging for the generation of high quality DNA or RNA, for example, DNA or RNA derived from Formalin-Fixed Paraffin-Embedded (FFPE) tissue specimens. You may need to do additional input controls:

- Evaluate the size distribution DV_{200} (the percentage of fragments >200 nucleotides) of your FFPE RNA samples. This can be done using a Bioanalyzer or Fragment Analyzer from Agilent.
- Calculate the Genomic Quality Number (GQN) for every sample using the Fragment Analyzer from Agilent.
- Perform qPCR-based methods for quantitation of the library, or for presence of genes of interest or control genes.
- Alternatively, for RNA samples, Customers can obtain the RNA Integrity Number (RIN) using a Bioanalyzer

Low quality samples may need higher input amounts in library prep or sequencing, or more cycles of amplification during library prep. Final libraries may also be shorter.

Indexing Strategy

When pooling unique libraries within a single lane, use an indexing strategy. Singular Genomics provides high-quality, empirically tested unique dual-indexed barcodes. The Singular Genomics Unique Dual Index (UDI) design consists of 12 bp sequences. Efficient demultiplexing can be achieved by sequencing the first 8 bp of the index, although we recommend sequencing the full index length when possible.

If you want to use your own custom indices, make sure your index design avoids misassignment of reads and tolerates several types of errors, including those introduced during index synthesis. The index sequences must be as divergent as possible while having similar GC content and avoiding extended homopolymer stretches. An index should best be at least 3 changes (insertions, deletions, and/or substitutions) away from another index sequence. This minimizes the possibility for misassignment.

The following provides some additional considerations you should evaluate when designing your own indices:

- Index Length
- Error Detection
- Error Correction
- Single or Dual Indexing
- Unique or Combinatorial Indexing
- Unique Molecular Identifiers

See the Adapters and Indices for the G4 Sequencing Platform Reference Guide for more information.

Fragmentation Method

Choose a fragmentation method according to your laboratory and workflow needs. Some considerations for each method are listed below.

Chemical Fragmentation	Suitable for fragmentation of RNA. Typically, heat is applied to RNA in the presence of a divalent metal cation (for example, magnesium or zinc). The length of the resulting fragments can be adjusted by modifying the incubation time.
Enzymatic Fragmentation	A cost-effective alternative to mechanical shearing methods. Depending on the approach, you may expect some amount of sequencing bias, as many enzymatic methods are influenced by sequence context.
Mechanical Fragmentation	Methods include sonication, acoustic shearing, needle shearing, nebulization, and others. Acoustic shearing generally has the tightest size distribution and introduces the least amount of bias compared to other methods. However, some investment in capital equipment is required. Mechanical methods can introduce DNA damage, and you might consider repair methods as part of your library construction process.

NOTE
Some sample types (for example, amplicons smaller than 300 bp) or applications (for example, small RNA, cfDNA) do not require fragmentation.

Insert Size

Generally, we recommend insert sizes between 25 bp to 500 bp, which means libraries range from 175 to 650 bp with adapters included. Be aware that the optimal insert size depends on a combination of factors, including the sequencing kit (refer to sections on read type and read length) and intended application.

Additional constraints, such as sample type, can prevent selection of an optimal insert size. For example, DNA derived from FFPE or from plasma may already be fragmented below your desired insert size. In these cases, it's important to choose compatible library preparation methods and to adjust other parameters, such as number of cycles and loading

concentration, to achieve a high-quality run.

Amplification Conditions

Choose amplification conditions according to the library prep manufacturer's instructions. In general, limit the number of PCR cycles to reduce the potential for additional bias and error associated with the PCR process. However, ensure that you have sufficient library for sequencing, as well as for any quality control (QC) processes. Quality control processes could use BioAnalyzer, Fragment Analyzer, Qubit, qPCR, or other processes.

Note for PCR-free library methods, the final product may be at such a low yield that you are unable to perform all desired quality control. qPCR quantification is recommended for PCR-free libraries.

Polymerase Fidelity

Polymerase fidelity can significantly affect the quality of your sequencing output. Choose a polymerase and amplification conditions to suit your experimental needs. Use a high-fidelity polymerase to minimize amplification errors.

Size Selection Using SPRI Beads

Small molecular species such as free adapters and adapter dimers are removed using SPRI beads (for example, AMPure XP or equivalent). The size selection cut-off is determined by the volume-to-volume ratio of bead suspension to DNA solution. A lower ratio results in a higher size cut-off, that means, a more stringent removal of the smaller size molecules.

Stringency of Fragment Removal

It is important to remove any excess adapters and adapter dimers from the library, to ensure the efficiency and quality of the sequencing reaction. Adapters may dimerize during ligation, which can negatively impact the quality and the usable throughput of the sequencing run. Clusters resulting from adapter dimers tend to generate high signal, resulting in increased crosstalk between adjacent nano wells. Dimers also tend to seed very efficiently due to their small size, further decreasing the number of nano wells available for library clusters.

Additionally, the presence of adapter dimers inflates the library quantitation, so the amount of library loaded does not correlate to the number of library molecules you want to sequence. This can also skew any sample pooling efforts, leading to undesirable differences in the relative representation of pooled libraries.

Choosing an Optimal Ratio

Several factors can influence the optimal ratio, including sample concentration, sample diluent, sample volume, and fragment size distribution. To determine the optimal size selection conditions, consider performing a titration series using different ratios (for example, 1.0×, 0.9×, 1×, 0.8×) for each new type of library.

The library size distribution should be checked using the Bioanalyzer DNA HS Chip or equivalent. Ensure the analysis shows the size on the x-axis and consider using tools to visualize and quantify the areas under the curve if any bimodality is observed.

Working with SPRI beads

When working with SPRI beads, always follow the manufacturer's instructions. In addition, make sure to pay attention to the following:

- Beads must be equilibrated to room temperature prior to use.
- Vortex beads well prior to each delivery.

- When adding beads to the sample, pipette up and down several times to ensure the sample and beads are well mixed, or vortex the sample and bead mix.
- For optimal results, working ethanol solution should be made fresh daily.
- Sample volume should be at least 50 µL to reduce variability.
- Sample should be suspended in molecular-grade water or standard buffers such as Tris or TE.
- A lower ratio of bead volume to sample volume retains longer fragments, and a higher ratio retains shorter fragments.
- When pipetting beads, aspirate slowly, ensuring there are no beads on the outside of the pipette tip.
- Ensure to remove all ethanol prior to elution. Residual ethanol can reduce the amount of recovered material.
- Do not over-dry the bead pellet. Excessive drying will reduce the final library yield. When you see cracks develop during drying, you may be over-drying the pellet.

Designing a Sequencing Run

There are several considerations when designing a new sequencing experiment. These include, but are not limited to:

- Read type
- Read length
- Depth of coverage or number of reads required
- Multiplexing
- Pooling

These considerations are described below.

Read Type

There are two established modes of reads.

Single Reads	The instrument reads from one end of a fragment to the other end. Single read runs can be more economical and are sufficient for less complex applications, like small RNA sequencing, ChIP-Seq, or differential expression analysis.
Paired Reads	The instrument reads from one end of a fragment to the other end, and then starts a second round of reading from the opposite end, resulting in a bidirectional sequence of a DNA molecule. Paired reads can attenuate many systematic, context-dependent sequencing errors, making it a good choice for *de novo* applications, as well as providing improved confidence in areas such as variant detection. Paired reads also help collecting more information by sequencing from each end if the insert is longer than the length of each read.

Read Length

The G4 Sequencing Platform provides flexibility in choosing the number of base pairs (cycles) you can sequence within a single read. The number of cycles chosen defines the output read length. Longer reads provide more accurate information in terms of relative position within the target genome and can be helpful when dealing with larger, more complex genomes. Shorter reads may be sufficient for simple mapping of reads to a small, lower complexity genome and for certain types of experiments (for example, RNA-Seq profiling or counting applications).

Shown below are examples of common applications that can be run on the G4 instrument. This table is only an example and is not an exhaustive list of all applications that are supported. Refer to https://singulargenomics.com/g4/reagents/ to see what flow cell is best for your application, and how they fit into a run configuration.

Table 1 Common applications for the G4 instrument.

Application	Typical Run
RNA Gene Expression	2 × 50 bp, 10 M reads
Single Cell RNA-Seq	130 cycles, 7 500 cells/sample and 20 000 reads/cell
Total RNA-Seq	2 × 100 bp, 50 M reads
Exome	2 × 100 bp, 35 Mb at 100× coverage
Target Enrichment	2 × 150 bp, 800 Kb at 4000× coverage

Depth of Coverage

In this guide, we define sequencing coverage as the number of unique reads that include a given nucleotide in the target region or genome of interest. Coverage and read requirements can depend on several of the following parameters:

- Application type
- Genome size, or in the case of target enrichment, region of interest/capture design size
- Genome complexity
- Read length
- Sequencing error rate
- Bioinformatics pipeline

NOTE
If you need more guidance regarding depth of coverage, contact Customer Care for consultation.

Multiplexing

Multiplexing means pooling several libraries and sequencing them simultaneously in the same lane. To determine the appropriate level of multiplexing, you need to know the output needed for a sample, and how the sequencing run is expected to achieve that. Considering the following:

- The sensitivity (depth of coverage) required per sample.
- The total size and/or the number of regions of interest.
- The expected number of reads and/or bases from the sequencing output, like number of reads per lane or per flow cell.
- Any additional filtering which is applied to the output, for example, more stringent quality filtering, removal of duplicates.

When you have determined the appropriate level of multiplexing, you can start designing the pooling scheme.

Pooling

When pooling libraries, it is critical to accurately quantify individual libraries to ensure equimolar pooling. Otherwise, you may see large variation in representation across samples, for example, sample 1 achieves 200× coverage while sample 2 achieves 50× coverage. Note that if individual products are over- or under-represented due to a multiplex PCR process, this cannot be overcome by any pooling strategy.

Other considerations when pooling libraries include the following:

- Choose libraries of identical size range for pooling.

- Choose libraries with indices that do not share sequence homology.
- Quantitate libraries using a functional assay, for example, qPCR.
- In the case that libraries of different size ranges must be pooled, employ a weighted strategy to achieve a more even representation.

  NOTE
  When planning a multiplex sequencing experiment, consider creating a single multiplex library pool and distributing the pool over the desired number of flow cell lanes to achieve the required read coverage per library. This strategy will mitigate variation in library coverage owing to differences in sequencing yield or quality across flow cell lanes.

Preparing a Sequencing Run

When preparing the sequencing run, consider the following:

- Loading concentration
- Titration
- Reagent preparation

These considerations are detailed below.

Loading Concentration

The loading concentration directly affects the observed clustering density, and therefore it is critical to choose the proper loading concentration. Factors that may influence optimal loading concentration include, but are not limited to:

- Average insert size.
- Number of functional molecules within the library, which means inserts with an adapter at each end.
- Quality of the library, for example, no contaminating short fragments.
- Accurate quantitation of the library; qPCR is ideal.

  NOTE
  PCR favors the amplification of shorter fragments over longer ones in a complex mixture of different length templates. This differential amplification needs to be taken into consideration to maximize sequencing yield. For example, if your library size range contains a large number of shorter fragments, which can yield extremely bright clusters, consider a lower loading concentration or perform additional size selection, where possible.

To help you determine your loading schedule, see the Loading Concentration Calculator at https://singulargenomics.com/loading-concentration-calculator/.

Titration

For larger experiments, or for previously uncharacterized library types, you consider performing a titration run to determine the optimal loading concentration for a library of interest. A typical titration series consists of 10 pM, 15 pM, 20 pM, 25 pM, and 30 pM loading concentrations.

Reagent Preparation

You can either thaw a cartridge for 24 hours in the refrigerator, or in a water bath at room temperature in 2.5 hours. Refer to the G4 Sequencing Platform User Guide for details regarding reagent preparation and loading.

Quality Control

Always adhere to the sample and library QC as described in the library prep manufacturer's instructions. Use the

information below in conjunction with those instructions and, in case of conflicts, follow the library prep manufacturer's instructions.

Quality Control Metric	Method	Interpretation
Quality of input DNA or RNA	Assess integrity (DIN or RIN score) and purity (OD 260/280 ratio)	Helps evaluate and control isolation method.
Size range before beginning library prep	BioAnalyzer, TapeStation, or equivalent method to visualize molecular weight	Helps determine an appropriate fragmentation method.
Size range after fragmentation	BioAnalyzer, TapeStation, or equivalent method to visualize size distribution	Assesses the efficiency of fragmentation and/or determine optimal SPRI bead: sample ratio.
Amount of library that is viable for sequencing	qPCR, QubIt, or equivalent method to assess the number of functional library molecules. Qubit is recommended to quantify libraries with different fragment size or broad fragment distribution.	Assesses ligation efficiency and determine the functionality of library.
Quality of the library	BioAnalyzer or TapeStation	Verify the absence of undesirable short fragments and determine optimal loading concentration.

Below are examples of what a trace for a clean library preparation should look like.

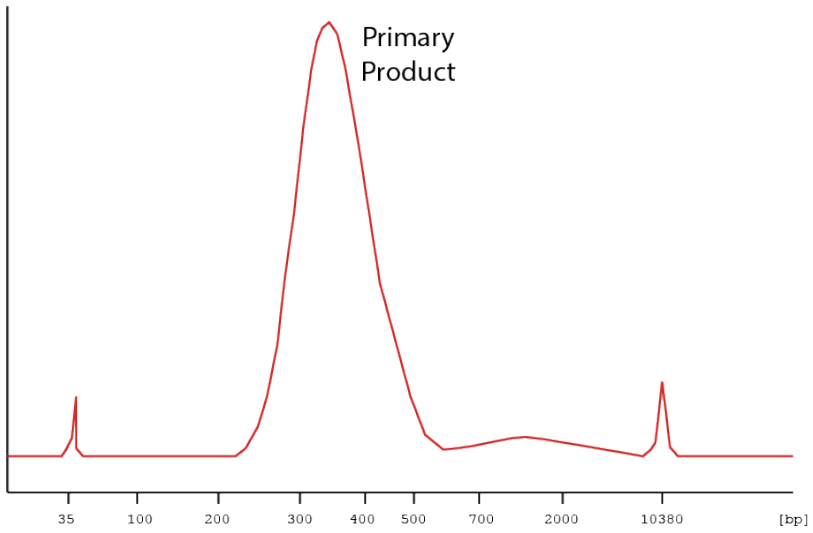

Figure 1 Example of a trace for a clean library preparation. Peaks at 35 and 10380 bp are size markers.

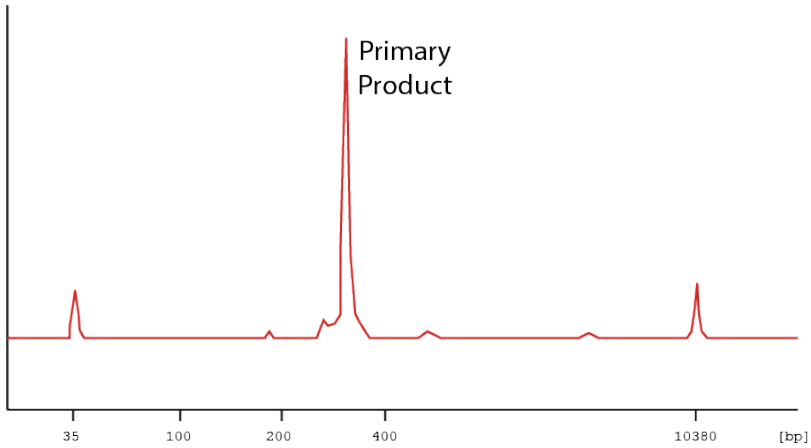

Figure 2 Example of a trace for a clean amplicon library preparation. Peaks at 35 and 10380 bp are size markers.

Adapter Dimers

The example below shows the presence of a sharp peak around 150 bp due to the presence of adapter dimers. For Singular Genomics adapters, sizes for adapter dimers range from about 125 bp for non-indexed adapters to about 151 bp for indexed adapters.

Generally, adapter dimers may be present for one or more of the following reasons:

- The size selection conditions were not stringent enough to remove the dimers.
- The size selection may have failed or was inefficient.
- The ligation conditions were not optimal, for example, low input or excessively high adapter to insert ratio.

If your library looks like the one below, consider an additional size selection using a lower ratio of beads to sample prior to sequencing.

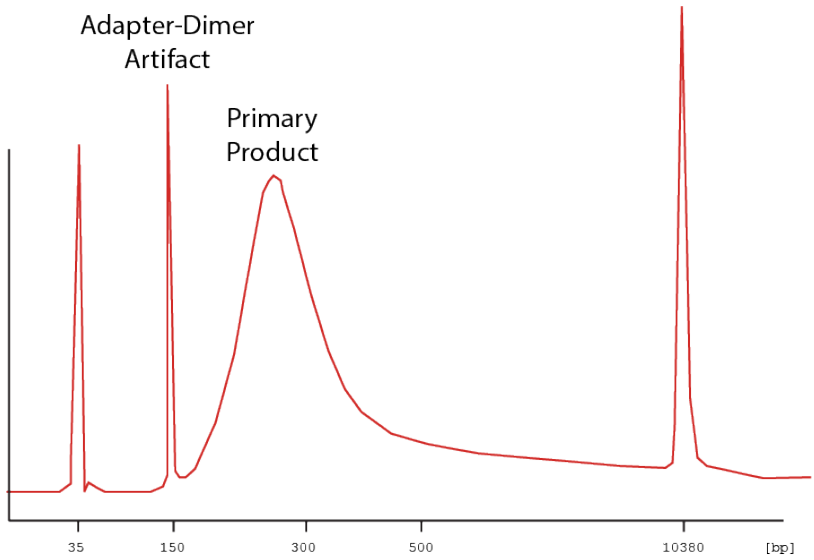

Figure 3 Example of a trace with adapter dimers. Peaks at 35 and 10380 bp are size markers.

Bubble Products

A bubble product contains regular adapter sequences on each end, but part of the insert is non-complementary. This results in a partially open bubble product that migrates slower on instruments for size determination, such as the Bioanalyzer (Figure 4).

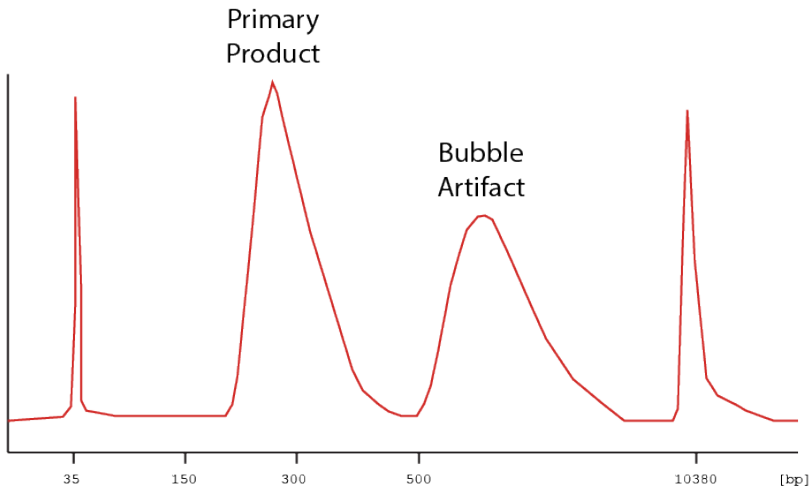

Figure 4 Example of a trace with bubble products. Peaks at 35 and 10380 bp are size markers.

Bubble products form when primers are not in excess anymore during PCR. In that case, instead of primers annealing to the library fragment, two library fragments may anneal to each other, forming a bubble product.

Bubble products may be caused by too much input, or too many cycles of PCR. Therefore, you can minimize bubble product formation the following ways:

- Quantify input using a fluorometric-based method as accurately as possible
- Use the recommended input amount for the library prep method.
- Use the recommended number of PCR cycles for the library prep method.

Bubble products can be sequenced, because once denatured, the single-stranded DNA products will behave the same as the ones from a clean library prep. However, be aware that fluorometric quantification will be inaccurate, it underestimates the amount of DNA present in bubble products. Use a qPCR-based method to quantify a library that contains bubble products.

Customer Care

Our team is committed to ensuring your success. Field application scientists, field service engineers, and customer care specialists are dedicated to responding quickly and effectively to your inquiries. From preparing your lab for the arrival of your instrument, to onsite training in system operation and maintenance and continued support, we'll be with you every step of the way.

Website	singulargenomics.com
Email	care@singulargenomics.com
Telephone	442.SG.CARES(442.742.2737)

SINGULAR GENOMICS SYSTEMS, Inc.

3010 Science Park Rd

San Diego, CA 92121 U.S.A.

Additional Resources

Go to https://singulargenomics.com/g4/support/ to find the following additional resources:

Resource	Description
G4 Sequencing Platform Safety and Compliance Guide	Information about operational safety considerations, compliance statements, and instrument labeling.
G4 Sequencing Platform Site Preparation Guide	Information for preparing your site for delivery, installation, and operation of the G4 instrument.
G4 Sequencing Kit Overview	Overview of available sequencing kit sizes, kit contents, and recommended use.
G4 Sequencing Platform Quick Reference Card	Dynamic Quick Reference Card for running the G4 Sequencing Platform for experienced users.
Loading Concentration Calculator	Calculator that helps you set up your denature and dilute steps.
Adapters and Indices for the G4 Sequencing Platform Reference Guide	Instructions for using Singular adapters with library prep protocols.
Networking, Security, and Remote Access for the G4 Sequencing Platform Technical Note	Specifications for connecting the instrument to a network and settings for antivirus software.
Training videos	Visual instructions for run setup and instrument operation.