CELLECTA

USER MANUAL

NGS Prep Kit for sgRNA, shRNA, and DNA Barcode Libraries

v1 — Last update: 2019/11/27

Cellecta, Inc.

Copyright (c) 2018 Cellecta, Inc. All Rights Reserved.

Table of Contents

1. NGS Prep Kit for sgRNA, shRNA, and DNA Barcode Libraries

Background

Cellecta's NGS Prep Kit provides a protocol and reagents for the PCR amplification and NGS sequencing of effector or barcode inserts from genomic DNA isolated from cell populations screened with any of several Cellecta CRISPR sgRNA, dual-sgRNA, shRNA, and barcode libraries, as well as a few similar libraries available from other sources. Sufficient reagents are provided to process 12-48 DNA samples (depending on the amount of DNA/sample) with unique indexing primers to multiplex up to 12 samples run in a single lane. Depending on the specific library used for a screen, several versions of the NGS Prep Kit are available—each containing the appropriate primers for the target library. To multiplex more samples, Supplementary Primer Sets with an additional 12 indexed primers are also available for each Sequencing Kit.

Cat.#	Description
LNGS-101	NGS Prep Kit for shRNA Libraries in pRSI12 (DECIPHER)
LNGS-102	NGS Prep Kit for shRNA Libraries in pRSI16cb/17cb (HGW shRNA) (or pRSI16/17)
LNGS-120	NGS Prep Kit for sgRNA Libraries in pRSG16/17 (CRISPR KOHGW)
LNGS-130	NGS Prep Kit for sgRNA Libraries in pRSGT16/17
LNGS-200	NGS Prep Kit for Barcode Libraries in pRSI16/17 (CloneTracker™ 50M)
LNGS-300	NGS Prep Kit for Barcode Libraries in pScribe (CloneTracker XP™)
LNGS-350	NGS Prep Kit for CRISPR-Barcode Libraries in pRSGScribe (CloneTracker™ XP-CRISPR)
LNGS-400	NGS Prep Kit for Dual-sgRNA Libraries in pRSL10 (KADHGW/KIDHGW)
LNGS-900	NGS Prep Kit for sgRNA Libraries in LentiGuide (GeCKO, Brunello)
LNGS-905	NGS Prep Kit for sgRNA Libraries in LentiCRISPRv2 (GeCKO, Brunello Kinome)

If you are not sure which kit to use for your library, please contact Cellecta (tech@cellecta.com).

Supplementary Primer Sets

Cat.#	Description
LNGS-101-SP	Supplementary Primer Set for LNGS-101 (12 Additional Index Primers)
LNGS-102-SP	Supplementary Primer Set for LNGS-102 (12 Additional Index Primers)
LNGS-120-SP	Supplementary Primer Set for LNGS-120 (12 Additional Index Primers)
LNGS-130-SP	Supplementary Primer Set for LNGS-130 (12 Additional Index Primers)
LNGS-200-SP	Supplementary Primer Set for LNGS-200 (12 Additional Index Primers)
LNGS-300-SP	Supplementary Primer Set for LNGS-300 (12 Additional Index Primers)
LNGS-350-SP	Supplementary Primer Set for LNGS-350 (12 Additional Index Primers)
LNGS-400-SP	Supplementary Primer Set for LNGS-400 (12 Additional Index Primers)
LNGS-900-SP	Supplementary Primer Set for LNGS-900 (12 Additional Index Primers)
LNGS-905-SP	Supplementary Primer Set for LNGS-905 (12 Additional Index Primers)

Please read the entire user manual before proceeding with your experiment.

Click the ? Download as PDF link located at the bottom of the left menu to download the PDF version of this user manual.

PLEASE NOTE: **The purchase of all Cellecta Products is covered by Cellecta's standard Terms and Conditions of Sale as described on Cellecta's website, and selected Products containing particular technology or having certain features are also subject to restrictions on use as outlined in the Label License section of our website. Please review these Terms and Label License Restrictions before opening and using your Product and, if you are not able to abide by the restriction, contact Cellecta to return the item to Cellecta for a full refund.**

Last modified: 2019/09/10

2. NGS Prep Kit Required Materials

Last modified: 2017/03/10

2.1. Materials Provided

Box 1 Components

Description	Concentration	Quantity
Taq DNA Polymerase	50X	200 µl (2 vials)
Taq DNA Polymerase Buffer	10X	1,200 µl (2 vials)
dNTP Mix	50X (10 mM each)	200 µl
PCR-Grade Water	NA	7,320 µl (4 vials)

Box 2 Components

Description (Generic primer names: see Note below)	Concentration	Quantity
Forward Primer	10 µM	150 µl
Reverse Primer	10 µM	150 µl
NFwd Primer	20X (10 µM)	240 µl
NRev Index Primer A	20X (10 µM)	20 µl
NRev Index Primer B	20X (10 µM)	20 µl
NRev Index Primer C	20X (10 µM)	20 µl
NRev Index Primer D	20X (10 µM)	20 µl
NRev Index Primer E	20X (10 µM)	20 µl
NRev Index Primer F	20X (10 µM)	20 µl
NRev Index Primer G	20X (10 µM)	20 µl
NRev Index Primer H	20X (10 µM)	20 µl
NRev Index Primer I	20X (10 µM)	20 µl
NRev Index Primer J	20X (10 µM)	20 µl
NRev Index Primer K	20X (10 µM)	20 µl
NRev Index Primer L	20X (10 µM)	20 µl
Seq NGS Primer*	100 µM	40 µl
Index NGS Primer	100 µM	40 µl

* In the NGS Prep Kits for CRISPR-Barcode and Dual-sgRNA Libraries, two unique Seq NGS Primers are included and are required for sequencing.

> **Note:** The generic names for each of the primers are listed above for reference. However, each version of the NGS Prep Kit has primers specific for the vector of the target library. These specific primers names also include a suffix to indicate the vector or library they target (e.g., Forward-G, Forward-12, NRev-16, etc.). This Manual uses the generic primer name when, for example, it is describing which one to include in a reaction. For each generic primer reference, you should use the appropriate fully named specific primer included in your kit.

Storage Conditions:

Store kits and all components at -20°C until ready for use.

Last modified: 2019/09/10

2.2. Additional Materials Required

- QIAquick PCR Purification Kit (QIAGEN, Cat.# 28106)
- QIAquick Gel Extraction Kit (QIAGEN, Cat.# 28706)
- PhiX Control (Illumina, Cat.# FC-110-3001)
- NGS Reagent Kits (Illumina):

Platform	Kit Type	Illumina Cat.#	Description
NextSeq® Single Read (SR) or Paired End (PE)	Sequencing & Cluster Gen.	20024906	NextSeq 500 v2.5 Kit (75 cycles, High-Output)
HiSeq® 2000/2500 **Single Read (SR)**	Sequencing	GD-401-3001	TruSeq SBS Kit v3 – HS (50 cycles)
	Cluster Gen.	FC-401-3002	TruSeq SR Cluster Kit v3-cBot-HS
HiSeq® 2000/2500 **Paired End (PE)**	Sequencing	GD-401-3001	TruSeq SBS Kit v3 – HS (50 cycles)
	Cluster Gen.	PE-401-3001	TruSeq PE Cluster Kit v3-cBot-HS
HiSeq® 3000/4000 **Single Read (SR)**	Sequencing	FC-410-1001	HiSeq 3000/4000 SBS Kit (50 cycles)
	Cluster Gen.	GD-410-1001	HiSeq 3000/4000 SR Cluster Kit
HiSeq® 3000/4000 **Paired End (PE)**	Sequencing	FC-410-1001	HiSeq 3000/4000 SBS Kit (50 cycles)
	Cluster Gen.	PE-410-1001	HiSeq 3000/4000 PE Cluster Kit

> **Note:** See Illumina website for information on HiSeq 2500 Rapid Run kits.

> **Note:** We do not support the NGS Prep Kit for the MiSeq Instrument.

Other than the specific reagents and instruments listed above, the protocols assume the user has access to standard materials (e.g., polypropylene tubes, pipette tips), equipment (table top centrifuges, pipettes, scales), and common reagents (e.g., TE buffer, ethanol) and buffers used in a typical life science laboratory.

Last modified: 2019/05/03

2.3. Related Products and Services from Cellecta

Related Services

- NGS of DNA Samples from Genetic Screen
- Lentiviral Library Packaging Services
- Custom Pooled Lentiviral Libraries
- Custom Lentiviral Constructs

Last modified: 2019/01/10

3. Procedure Overview

Quantifying the representation levels of sgRNA, shRNA, or barcode sequences in experimental samples derived from genetic screens using pooled effector libraries first requires isolation of the genomic DNA. After DNA isolation, effector or barcode sequences are amplified by PCR (First Round PCR) from purified genomic DNA using library-specific primers. A Second Round of PCR amplification of the effector and barcode sequences is then performed with “nested” primers to introduce indexes and generate template compatible with Illumina NGS. Finally, the representation of sgRNA or barcode sequence in each sample are quantified by counting the number of specific reads generated by NGS on an Illumina Instrument (e.g., NextSeq 500 or HiSeq).

> **Note:** The protocol was optimized using a Life Technologies Veriti® Thermal Cycler. Use of other PCR enzymes and/or thermal cyclers may require additional optimization.

Last modified: 2017/05/16

3.1. Genomic DNA Extraction

To accurately measure the relative fraction of each sgRNA or barcode present in a specific cell population, it is important to isolate the whole amount of genomic DNA from the cells derived from genetic screen. Purification of genomic DNA from just a fraction of cells at a particular time point or treatment condition in a screen, may not provide full representation the effector constructs. There are several protocols that can be used for DNA isolation. Depending on the number and type of cells, some approaches may work better than others.

- Genomic DNA isolated from more than 10 million cells will overload most column-based DNA isolation kits and compromise yields. To avoid this loss of genomic DNA, which can distort representation of the guides or barcodes in the population, we highly recommend using the conventional Genomic DNA Extraction Protocol. With this procedure, you usually obtain 50 µg-100 µg of genomic DNA from 10 million cells.

This protocol is typically recommended to purify DNA from a “dropout viability” screen which often requires growth of 25-100M cells per time point or treatment variation, depending on the size of the library.

- For small and medium-sized populations of cells (from small libraries screens, positive selection screens where most of the cells are killed off, or FACS-based enrichment screens), we recommend using the following QIAGEN kits:
    - From 1 million to 10 million cells: Use the QIAGEN DNeasy Blood and Tissue Kit (QIAGEN, Cat.# 69504)
    - For fewer than 1 million cells: Use the QIAGEN QIAamp DNA Micro Kit (QIAGEN, Cat.# 56304)
- After purification, you should resuspend your DNA at a concentration of ca. 1-2 µg/µl. DNA samples can be stored at +4°C for a few weeks or at -20°C for an extended period of time.

Last modified: 2018/06/03

3.1.1. Large-Scale Genomic DNA Extraction Protocol

> **NOTE:** Use of disposable tubes is highly recommended in order to avoid contamination.

1. Resuspend cell pellet in 5 ml QIAGEN Buffer P1 (with RNase A) in a 15 ml polypropylene (phenol:chloroform resistant), Falcon screw-cap centrifuge tube (12,000 RCF rated).
2. Add 0.25 ml 10% SDS, mix, and incubate 5 minutes at room temperature.
3. Using an ultrasonic homogenizer, sonicate to shear DNA into 10-100 kb sized fragments. To prevent cross-contamination, thoroughly wash the ultrasound head with running water and dry with clean paper towels between samples.
4. Add 10 µl of Proteinase K, mix, and incubate 15 minutes at room temperature.
5. Add 5 ml Phenol:Chloroform:Isoamyl Alcohol solution, vortex hard, and spin down 60 min, 20°C at 8,000 rpm in JA-14 or equivalent rotor (Beckman).
6. At this point, there should be approximately 5 ml of clear upper phase. Transfer 4 ml of upper phase to a new 15 ml disposable screw cap tube (same as in Step 1).
7. Add 0.5 ml of 3M Sodium Acetate and 4 ml isopropanol, mix well, then spin down 30 min, 20°C at 8,000 rpm in a JA-14 or equivalent rotor.

> **NOTE:** To produce a more visible pellet that is compacted at the bottom of the tube, it is recommended to incubate overnight at room temperature before centrifugation.

8. Discard supernatant, add 10 ml of 70% ethanol, and spin down 5 min, 20°C at 8,000 rpm in a JA-14 or equivalent rotor.
9. Discard supernatant and air-dry pellet.
10. Dissolve DNA pellet in an appropriate volume of dH_2O to a concentration of ~2 mg/ml.
11. Incubate 30 minutes at 80°C before spectrophotometer reading.

> **NOTE:** Expected yield is about 10 µg genomic DNA per 1 million cells.

Last modified: 2017/03/10

3.2. Amplification of sgRNA or Barcode Sequences from Genomic DNA

The protocol is based on two rounds of PCR to amplify the integrated sgRNA, shRNA barcode sequences, or clonal barcodes from the genomic DNA. Also, the second round of PCR includes primers that add the sample-specific indexes and sequences (P5 and P7) complementary to primers immobilized in Illumina flow cells. With the PCR reactions, the optimization of cycles is required to avoid over-cycling. Using an optimal cycle number is necessary to generate a sufficient amount of amplicons to sequence while maintaining the relative representation of each sgRNA or barcode in the cell population.

- As an amplification/sequencing positive control, we recommend that you use the original shRNA/sgRNA/barcode plasmid library (10 ng) used in the screen. This positive control can be used as a baseline in NGS data analysis to measure the enrichment or depletion of effector constructs in the experimental samples.

- Although it is typically not necessary, you may use the packaged effector library as a positive control. In order to use the packaged library as a positive control, you can start with a volume containing sufficient viral particles equal to at least a few hundred-fold the complexity of the library. For example, for a 50,000-construct library, you should use a volume containing 10-25 million transduction units (TU). Also, before starting PCR, you will need to reverse transcribe the viral RNA sequence using a reverse primer that targets the sense (upper) strand of the insert site at a location where it will generate a full transcript of the target sgRNA/barcode region, including the PCR primer sites. For most Cellecta libraries containing the UbiC promoter, the RevUbiC1 primer (5'-AGGCAGCGCTCGCCGTGAGGA-3') adjacent to the cloning site can be synthesized and used for the reverse transcription step.

First Round PCR

The goal of the first PCR is to amplify shRNA barcodes, sgRNA inserts, or barcode library sequences from genomic DNA isolated from each experimental cell population. The amount of starting DNA can vary depending on the type of screen. For samples with up to 50 µg of genomic DNA, run a single 100 µl reaction for the each sample. For samples with more than 50 µg of DNA, we recommend dividing up the run into multiple 100 µl PCR reactions, each with a maximum of 50 µg DNA.

1. For each sample, prepare a PCR mix according to the table below, where x is the volume of genomic DNA (typically 50 µl at 1 µg/µl). Many samples will have more than 50 µg of DNA. For these samples, you will need to scale up the master reaction below as needed for each sample, then divide the reaction into multiple tubes with 100 µl in each. For example, if you have 180 µg of DNA in one sample, you should prepare a 400 µl master mix, then aliquot 100 µl in four PCR tubes for amplification (180 µg / 50 µg = 3.6, so round to 4 reactions).

Volume	Component
x µl	Genomic DNA (up to 50 µg)
3 µl	Forward 1st Round PCR Primer
3 µl	Reverse 1st Round PCR Primer
2 µl	dNTP Mix
10 µl	Taq Polymerase Buffer
80 – ***x*** µl	PCR-Grade Water

2 µl	Taq Polymerase
100 µl	**Total volume**

2. Mix gently, centrifuge briefly to collect droplets, divide samples with more than 50 µg of genomic DNA into 100 µl PCR reactions as necessary.

3. Perform PCR using the following cycling conditions.

95°C, 2 minutes	1 cycle
95°C, 30 seconds, 65°C, 30 seconds, 68°C, 2 minutes	16-18 cycles*
68°C, 2 minutes	1 cycle

> * **Note:** Depending on your library, the specific number of cycles may need some optimization. For barcode libraries (and shRNA libraries with barcodes), 16 cycles is typically sufficient. For sgRNA libraries and dual-sgRNA libraries, 18 cycles is recommended. If you are using less than 50 µg of genomic DNA in a 100 µl PCR reaction, increase the number of cycles (e.g., for 25 µg, use one extra cycle).

Second Round PCR

The Second Round of PCR is required in order to significantly reduce genomic DNA carryover into the NGS step. Additionally, the Second Round PCR primers add the P5 and P7 sequences that are complementary to the immobilized primers in the NGS Illumina Flow Cells.

- For Cellecta NGS Prep Kits and Supplementary Index Primer Sets, the NRev Index Primers (up to 24) each contain a different index sequence which will be used to deconvolute the sequencing results for each sample mixed together in the same Flow Cell or sequencing lane (i.e., multiplex sequencing). It is important, therefore, to use a different NRev Index Primer (each signified by a different letter) to amplify each sample that will be run together in the same Flow Cell or lane.

For the Second Round of PCR, start with an aliquot of 5 µl of the First-Round PCR reaction for each sample.

1. If you are starting from more than 50 µg of genomic DNA, it is important to combine all individual reactions for each sample together into a single tube.

2. Perform the Second Round of nested PCR using a 5 µl aliquot of the First Round PCR product for each sample:

Volume	Component
5 µl	First Round PCR Product
5 µl	NFwd Primer
5 µl	NRev (Index) Primer*
2 µl	dNTP Mix

10 µl	Taq Polymerase Buffer
71 µl	PCR-Grade Water
2 µl	Taq Polymerase
100 µl	**Total volume**

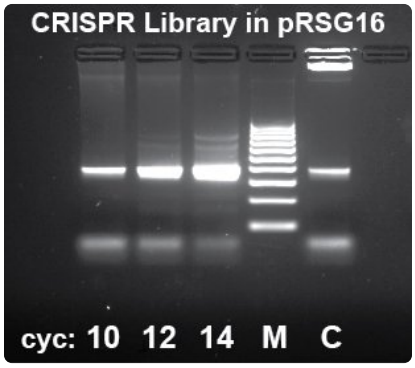
* **Note:** Each sample to be sequenced together in a single Flow Cell or lane needs to be amplified with a different NRev Index Primer in the NGS Prep Kit. Be sure to use a different lettered NRev Index for each sample that will be loaded on the same Flow Cell or lane.

3. Perform PCR under the following cycling conditions.

95°C, 2 minutes	1 cycle
95°C, 30 seconds, 65°C, 30 seconds, 68°C, 2 minutes	9 cycles**
68°C, 2 minutes	1 cycle

** **Note:** Additional cycles may be needed for samples as described below.

4. Analyze the amplified products from the Second Round of PCR. This analysis may be done on an Agilent® 2100 Bioanalyzer® instrument with the Agilent® High Sensitivity DNA Kit (Cat.# 5067-4626) or a Fragment Analyzer (Advanced Analytical) instrument using Standard Sensitivity NGS Analysis Kit (Cat.# DNF-473-1000) using the manufacturer's protocol, if you have access to this equipment. If not, then we recommend the following procedure using agarose gel electrophoresis to analyze the samples:
    - Run all samples including positive control on 3.5% agarose-1xTAE gel. Mix 5 µl of aliquot from the each sample with 2x loading buffer (e.g. 10% sucrose with Bromophenol Blue tracking dye).
    - Analyze the intensity of the bands after electrophoresis (i.e., Bromophenol Blue tracking dye about 80% down the gel). The typical results of agarose gel electrophoresis analysis are shown in the figure below. The expected size of amplified barcode, shRNA barcode, and sgRNA products is typically between 125 to 400 bp or dual-sgRNA product around 750bp and is provided in the NGS Cassette Design information specific for your library or NGS Prep Kit Product Certificate.



CRISPR Human Genome 80K Knockout Library NGS PCR product analysis by gel electrophoresis. In this example, 10 cycles is the recommended cycle number for the Second Round PCR step.

DECIPHER Library NGS PCR product analysis by gel electrophoresis

5. Quantify the amount of the amplified product for each sample using software tools provided with instrument or gel imager. If all samples show similar yield of amplified products (+/- 2-fold) no addition cycling is required and you should proceed to NGS Samples Purification. If there is a greater than 2-fold variance, you should put the samples with very weak or missing bands back into the PCR instrument for additional cycles. For example, if the amount of PCR product in the sample is 4, 8, or 16-fold less than in other samples, run that sample for an additional 2, 3, or 4 PCR cycles, respectively. For samples in the +/- 5-cycle range, don't run additional cycles.
    - If you run additional cycles on any samples, quantify the reaction products again by using one of the instruments mentioned above of gel electrophoresis to ensure that all samples have correctly amplified products.
    - Avoid overcycling! Try to use the minimum number of cycles which allows quantitation of the bands. Importantly, for dual-sgRNA libraries, overcycling induces a significant level of recombination between different dual-sgRNA molecules. Overcycling also often generates additional higher molecular weight bands. These bands do not normally interfere with NGS unless they are very strong. Also, overcycling has the potential to compress the representation of the targeted sequences if the reaction exceeds the exponential phase. Therefore, it is desirable to minimize the cycle number. If there are excessive higher molecular weight bands, you can rerun Second Round PCR for that sample using another 2 µl of DNA from the First Round PCR and a fewer number of cycles.

DNA samples can be stored at +4°C at this stage.

Last modified: 2019/11/27

3.3. NGS Sample Purification

To maximize sequencing of the targeted sgRNA or barcodes and minimize background sequences, we recommend purifying the amplified PCR products by preparative gel electrophoresis.

1. Based on quantitative analysis of PCR product yield for each of the samples, combine the PCR products from each sample at the same amount. For example, for samples with yields of correct PCR product of 20 ng/µl and 60 ng/µl, combine 6 µl and 2 µl of PCR reaction, respectively.

Purify and concentrate the PCR products from the combined sample using the QIAGEN QIAquick PCR Purification Kit, following the manufacturer's protocol. The combined sample should be eluted in a volume of approximately 30 µl.

2. After purifying and concentrating, run each purified combined PCR product with 10X loading buffer on a 3.5% agarose-1XTAE gel with well sizes that accommodate at least 50 µl.

3. Using a transilluminator and scalpel, excise the narrow band that corresponds to the correct target sgRNA or barcode amplicon size.

CAUTION! Be sure to use UV safety glasses to protect your eyes when viewing and excising the DNA from the gel on the transilluminator.

4. After excision, purify the combined PCR product from the gel fragment using a QIAquick Gel Extraction kit following the manufacturer's protocol. Elute the purified PCR product in 20 µl of elution buffer.

> **Note:** Be sure to centrifuge QIAquick columns at maximum speed for at least 3 minutes before eluting DNA to avoid ethanol contamination in the purified PCR product.

5. Quantify extracted DNA in the combined sample by A260nm OD measurement using a NanoDrop spectrophotometer (or equivalent), and then adjust the concentrations in all samples to 10 nM. For example, if the amplicon size is 200bp, the 10nM concentration corresponds to 1.42 ng/µl based on A260 OD measurement. For an amplicon size of 750bp (dual-sgRNA product), a 10nM concentration corresponds to 5.3 ng/µl.

DNA samples can be stored at -20°C at this stage.

Last modified: 2019/09/06

3.4. Sequencing sgRNA Inserts or Barcodes

The Second Round nested PCR primers contain sequences (P5 and P7) complementary to the sequences of the immobilized primers necessary for generating amplification clusters on Illumina's NextSeq or HiSeq Flow Cells. In addition, if you are using the Cellecta NGS Prep Kits, the NRev Index Primers add sample-specific index sequences which allow you to combine and multiplex up to 24 samples on a single Flow Cell or lane. The primers were designed with the NextSeq instrument in mind, but they are compatible with HiSeq instruments as well.

> **Note:** Please see the NGS Cassette Diagram for the specific Cellecta library or the Product Certificate for the NGS Prep Kit you purchased for information on compatibility with HiSeq Single-Read (SR) and Paired-End (PE) Reagent Kits.

The number of sequencing cycles (read length) required depends on the sgRNA or barcode length. Please refer to your library information (NGS Cassette Diagram for the specific Cellecta library or the Product Certificate for the NGS Prep Kit).

Guidelines for preparing samples for sequencing:

- Combine together equal amounts (each adjusted to 10 nM) of experimental and positive control samples to be sequenced together. Importantly, each sample in the combined pools needs to have a unique index sequence.

> **Note:** The number of samples which can be sequenced in one lane or flow cell is based on the complexity of effector

> library and expected total number of reads generated in NextSeq or HiSeq instrument. You will need to determine how many target reads you want per sample based on your experiment.

- Prepare the samples for NGS by following the Illumina “Denature and Dilute Libraries Guide” for the NextSeq or HiSeq System depending on which instrument you are using:
  - NextSeq 500: https://support.illumina.com/downloads/nextseq-500-denaturing-diluting-libraries-15048776.html
  - HiSeq: https://support.illumina.com/downloads/hiseq-denature-dilute-libraries-guide-15050107.html

- Spike in 5-15% of the PhiX Control to the pooled indexed sample to increase the nucleotide diversity.

- Mix Cellecta’s custom **Seq NGS Primer** with the Illumina primer mix (for the NextSeq, spike-in the primer at the appropriate cartridge position—usually at Read 1). For indexing multiple samples in a Flow Cell or lane, also add the **Index NGS Primer** directly to the appropriate well (i.e., spike in either with the Illumina index primer or the reverse sequencing primer, see Note). Each primer should have a final concentration of 300 nM for the NextSeq, or 500 nM for the HiSeq. We recommend measuring the final volume of the primer mix in each well to ensure accurate spiked-in primer concentrations.

> **Note:** Depending on the way the amplification and sequencing primers were designed, the position of the Seq NGS and Index NGS primers may vary. Refer to the supplementary information (NGS Cassette Design or Product Certificate) provided for the specifics on how to set up the NGS Primers for your specific library.

- Specify the appropriate number of cycles to sequence the complete sgRNA or barcode region (please refer to Product Certificate), and specify 7 cycles to read 6-base Index sequence in the NGS Prep Kit.

Last modified: 2019/09/06

3.5. Demultiplexing and Aligning Sequencing Reads

Cellecta provides NGS Demultiplexing and Alignment Software for most of its libraries. The program demultiplexes, aligns, and scores Illumina NextSeq platform sequence data generated from samples of Cellecta libraries prepared using the primers in Cellecta NGS Prep Kits (cat.#s LNGS-100 through LNGS-999). Below are the instructions for use of this Software.

Program Requirements and Installation

- The program runs on 64-bit Windows, such as Windows 7 or higher.

- If the option to generate FASTQ files is enabled, sufficient local disk space to store the resulting output is required; this will approximately equal the combined disk space of the input FASTQ files.

- From the Cellecta website, download the directory containing the program and your Cellecta Library Design File(s) to your computer.

- If you cannot find the Library Design File(s) for your library, you may need to download additional files. Contact a Cellecta representative for assistance.

Using the Alignment Software

The program requires the following input files:

1. The Illumina **FASTQ file** (.fastq or .fastq.gz) with the raw NGS data.
2. A **Sample Descriptions File** with data on which samples were sequenced and what indexes were used for each sample in TEXT (.txt or .tsv) or FASTA (.fa or .fna).
3. The **Library Design File(s)** which contain(s) the list(s) of target sequences from the Cellecta Library that were run on the Illumina Instrument in TEXT (.txt or .tsv) or FASTA (.fa or .fna).

The program outputs two files:

1. A table of aligned counts for each sample (in columns) and target (in rows).
2. A table of total aligned, unaligned, and ambiguous counts for each sample to assess the quality of the sequencing.

    - “Unaligned” sequences are those for which no sample or target could be identified. “Ambiguous” sequences are those for which multiple samples or targets could be identified; they are not expected to occur with most libraries, as library sequences are chosen to be unambiguous within the Hamming distance used.
    - In addition to the files above, there is an option to have the program output an individual FASTQ file for each sample when it demuliplexes the FASTQ file for the sequencing run.

Key Steps for Use of the Software:

1. Obtain the FASTQ format files from the Illumina NextSeq.
2. Fill out the Sample Description File and save it as .txt.
3. Pick the tab to select the correct version of the Alignment Software based on the NGS Prep Kit you are using.
4. Enter the locations of the FASTQ NGS File, Sample Descriptions File, and Library Design File(s) into the Program interface.

Set up the Sample Description File

1. Open the template file “**sample-description.xlsm**”. You will enter your sample names into this template, and then save the file with a name descriptive of your experiment.

> **Note:** Please make sure to enable macros in Excel, if prompted.

2. The ***Sample Description Form*** has 3 columns:
    - Column 1 lists the primer designation for each barcode.
    - Column 2 is reserved for the user to enter the sample descriptions.
    - Column 3 indicates which DNA index sequence will be used to demultiplex the samples.

3. Enter the sample descriptions in the column 2. Fill in a name for each sample at the appropriate location based on the primer designation used to PCR amplify the specific sample, and the DNA index sequence associated with that

sample.

> **Note:** Do not skip a description for any sample. No sequencing data will be extracted for blank rows.

4. Save the Sample table in tab-text format by either using the Tab-Text save button or manually saving the table in .txt format.

Running the Cellecta Alignment Software

1. Open the program and select the tab corresponding to the NGS Prep Kit used to prepare the library. The program uses different settings depending on the Library and NGS Kit used. Click on the tab corresponding to your experiment.

2. Select the folder with the FASTQ NGS data files. Click the button to navigate the file system and select the folder containing the FASTQ files from the sequencer. The program will report how many distinct lanes it detected in that location. If the FASTQ files are compressed in GZIP format (have the ".gz" extension), it is not necessary to de-compress them. Select the GZIP folder.

> **Note:** The program identifies FASTQ files by looking for file names following the Illumina NextSeq convention. Do not alter the Illumina-assigned file names.

3. Select the samples. Click the button to navigate the file system and select the Sample Descriptions File corresponding to the experiment that was prepared above.

4. Select the library sequences for alignment with the Library Design File(s). Click the button to navigate the file system and select the appropriate Library Design File(s). Some program versions require more than one Library Design File (depending on the design of the library being sequenced). The number of Library Design files needed will be obvious from the Program interface.

5. If you want to save the individual demultiplexed FASTQ files, select this option. The program will run more slowly if the individual FASTQ files option is selected. Also, ensure that there is adequate disk space (i.e., equal to the size of the input FASTQ files) to output these files.

> **Note:** When using the software with Cellecta shRNA libraries, you need to indicate if the library has *clonal barcodes* in addition to the standard shRNA barcodes. For the alignment work correctly, you need to indicate on the interface whether or not the library you have sequences was designed with these clonal barcodes. Please refer to the library Product Certificate if you are unsure whether your library has these or not. Cellecta's Human Genome-Wide shRNA does have clonal barcodes so the option is selected by default.

6. Begin the deconvolution/alignment run. Click the "RUN" button at the bottom of the tab, and choose the name and location of your output file in the file navigator window. If you are saving the demultiplexed files, they will be written to the same location, so please ensure that sufficient disk space is available.

- After pressing *return*, a new window will pop up describing the run. When the run is completed the "DONE" button at the bottom will become active. This may take several hours.

> **Note:** Additional runs can be started while waiting for the software to finish. To start an additional run, navigate to the

> main window and click any of the three buttons to select FASTQ files, samples, or library. The “RUN” button will re-activate and a new process can be initiated.

> **Note:** For CRISPR sgRNA Libraries sequenced with Cellecta’s NGS Prep Kit for sgRNA Libraries (LNGS-120), Cellecta’s Alignment Software can be used to just demultiplex the Illumina data to make individual FASTQ file for each sample.

Other Options Instead of Cellecta’s Alignment Software

If you are not using Cellecta’s Alignment software, you will need to demultiplex the run to into separate FASTQ (*.bcl files) files. This can be done using the Illumina BaseSpace Hub or, alternatively, demultiplexing and fastq file generation may be done faster using a UNIX server and the Illlumina bcl2fastq software. Using the bcl2fastq software also avoids the rather slow step of downloading FASTQ files from BaseSpace. You can obtain the bcl2fastq program at this link. Information on how to use this software is available at the following sites:

- The Bioinformatics I/O Blog on the University of Glasgow’s Centre for Viral Research Site (for instance, see blog articles How to generate a Sample Sheet from sample/index data in BaseSpace and How to demultiplex Illumina data and generate fastq files using bcl2fastq).

- The BaseSpace Sequence Hub Help Center from Illumina. In particular, see How to download data from Run.

Last modified: 2019/08/20

4. Troubleshooting

Last modified: 2017/03/16

4.1. No PCR Product

Possible reasons:

1. Incorrect primers or bad reagents used
2. Missing reagents
3. Low transduction of target cells
4. Poor gDNA purification prep contains PCR inhibitors

Solutions:

- Include 10 ng of plasmid library DNA as a positive control. If it produces the correct amplification product, the problem lies either with absent or low numbers of sgRNA (e.g., low MOI or problems with the transduction efficiency) or impurities in genomic DNA that block sgRNA amplification. Dilute the genomic DNA 2-5 fold and repeat the amplification step in several test tubes. If the positive control does not produce the correct product, confirm use of the correct primers and reagents.
- Verify that primer sequences are correct.

Last modified: 2018/05/10

4.2. No sgRNA/Barcodes Present in NGS Results

Possible reason #1:

Incorrect First or Second Round PCR Primers or Sequencing Primer used.

Solution:

Ensure that you use the correct First and Second Round PCR Primers and Cellecta Sequencing Primer for your library.

Check the Library Cassette Diagram provided with the kit.

Possible reason #2:

Incorrect NGS reagent kit used.

Solution:

Ensure that you or your NGS Core Facility uses the proper Illumina NGS reagent kit. For details, please check the kit descriptions in the Additional Materials Required section.

Last modified: 2019/01/10

4.3. NGS Data Includes Significant Percentage of Reads from a Single Sequence

Possible reason #1:

Cells used in screen may carry foreign DNA which contain NGS primer binding sites.

Solution:

Do not use cells for your screen which contain foreign DNA with NGS binding sites.

Last modified: 2018/05/10

4.4. Incorrect Band Pattern on Agilent Bioanalyzer

Possible reasons:

The Agilent Bioanalyzer is not a good tool to analyze these preps. The sizing is not as accurate as a 3% agarose gel for small fragments, and the preps tend to have a significant amount of genomic DNA which overloads the tiny Agilent capillary.

Solutions:

Use a 3% agarose gel to analyze your PCR products.

Last modified: 2017/03/16

5. Technical Support

If the Troubleshooting section has not helped, please provide answers to the questions below (where applicable) before contacting us.

Library Used:

1. Which library did you use?
2. What is the lot number?

Packaging the Library:

1. What was the lentiviral titer, and what was the total number of TU packaged?
2. How was the virus concentrated? (if applicable)

Transducing Target Cells:

1. What MOI did you use to transduce your target cells?
2. What target cells did you use?
3. How many replicates did you use? (i.e., duplicate, triplicate, etc.)
4. Did you use puromycin after transduction, and at what concentration?
5. For how long did you use puromycin on the cells?

Library Screen:

1. Could you briefly explain your experiment?
2. How many infected cells were used?

Sample Preparation & NGS:

1. Describe the protocol you used to amplify the barcodes.
2. What NGS system and which Illumina Kit did you use?
3. How much PCR product was used for NGS?
4. How many sequences were read per sample?

Email Addresses

Technical Support: tech@cellecta.com
General information: info@cellecta.com

Phone Numbers

Phone: +1 650 938-3910
Toll-free: (877) 938-3910

For the latest technical news and updates, visit Cellecta's blog at: https://www.cellecta.com/blog-news/

Last modified: 2018/05/10

6. Contact Us

Email Addresses

General information: info@cellecta.com
Sales: sales@cellecta.com
Orders: orders@cellecta.com
Technical Support: tech@cellecta.com

Mailing Address

Cellecta, Inc.
320 Logue Ave.
Mountain View, CA 94043
USA

Phone Numbers

Phone: +1 650 938-3910
Toll-free: (877) 938-3910
Fax: +1 650 938-3911

For the latest technical news and updates, visit Cellecta's blog at:
https://www.cellecta.com/blog-news/

CELLECTA
Discovery is *Yours*.

Copyright © 2019 Cellecta, Inc. All Rights Reserved.

Last modified: 2019/01/10