Novogene Product FAQs

Novogene US

Contents

1. Regular DNA products

1.1 Human Whole genome sequencing (hWGS)

1.1.1 Sample requirements

Q2. Client: Could you please recommend any kits for DNA extraction of WGS projects?

A2. TS: Based on our practical experience, Qiagen DNeasy series kits could be a better choice, which always help us to get DNA samples with good quality. You could also choose other kits on the market.

Q3. Client: Could we send the samples in 96-well plates or PCR stripe tubes?

A3. TS: We are so sorry that they are not acceptable. Samples should be shipped in 1.5 ml or 2.0 ml flip-top microcentrifuge tubes. Incorrect tubes will cause project delay and additional handling fee ($5/tube). Here I attached the file, Novogene-preship Checklist. Please follow the details as mentioned. They are essential.

How will Novogene perform sample QC on DNA samples?

Novogene utilizes 2 major QC methods for DNA sample qualification:

a) Agarose gel electrophoresis analysis for DNA purity and integrity;
b) Qubit® 2.0 flurometer quantitation for accurate measurement of DNA concentration.

What is the fragment size that gDNA samples will be sheared to?

The genomic DNA of each sample will be randomly sheared into short fragments of about 350 bp.

Can Novogene construct longer insert fragment libraries?

Yes, we have successfully construct libraries with up to 800 bp long.

What sequencing strategy does Novogene use for hWGS services?

WGS: NovaSeq 6000 or HiSeq X Ten platform, paired-end 150 bp, for 150~350bp insert fragment. Unless otherwise specified, this service is performed on NovaSeq 6000 platform by default.

What is the sequencing coverage do you recommend for each analysis?

For SNP and Indel calling, our bioinformatics team recommend at least 20X sequencing coverage.

For SV and CNV detection, 30X sequencing coverage is required.

1.1.8.3 Can Novogene provide any customized analysis based on client's requirement?

The customized analysis needs to be estimated the feasibility by the BI team before the project initiate.

1.1.8.4 What files do Novogene also provide along with the analysis report?

We also provide FASTQ, BAM, Annotation and VCF files with the report.

1.1.8.5 Can Novogene provide hWGS standard analysis for the data from other sequencing company?

The project needs to be estimated by the BI team. And for the evaluation, please inform us the sample quantity, sequencing platform used, reference genome, data output.

1.2 Plant and Animal re-sequencing (PAWGS)

1.2.1 Sample requirements

Q2. Client: Could you please recommend any kits for DNA extraction of WGS projects?

A2. TS: Based on our practical experience, Qiagen DNeasy series kits could be a better choice, which always help us to get DNA samples with good quality. You could also choose other kits on the market.

1.2.2 Sample shipping

Q1. Client: Could I send lyophilized-DNA power to your company?

A1. TS: Yes. However, it is worth to mention that lyophilized-DNA power requires tube transferring and dilution service, there will be additional cost.

Q2. Client: Since gDNA samples are relative stable, would we ship out samples under room temperature?

A2. TS: We are so sorry that we do not recommend you do it. Because the quality can still be affected by surrounding environment, even though DNA is more stable in comparison of RNA. Please submit your package with sufficient dry ice to allow the consumption (sublimation) of 5 kg per day, and ship the samples via FedEx second-day-delivery service.

What sequencing strategy do Novogene use for plant & animal WGS services?

a) Illumina platform, paired-end 150 bp.

b) PacBio Sequel I/II platform, SMRT sequencing.

c) Nanopore platform (PromethION).

What is the sequencing coverage do you recommend for each analysis?

a) Illumina platform based on variation detection

For SNP and Indel calling, our bioinformatics team recommend at least 20X sequencing coverage.

For SV and CNV detection, 30X sequencing coverage is required.

b) PacBio Sequel I/II or Nanopore platform base variation detection

For SV detection, 10X sequencing coverage is required.

Can Novogene provide CRISPR/Cas9 On-target and Off-target detection analysis?

We could perform the analysis by meeting the following four available conditions simultaneously.

a) The diploids and chromosome-level reference genome is required.

b) The gRNA sequence of case/control paired samples is required.

c) Results of standard analysis are required.

Please consult with the Novogene representative for more information.

What is the sequencing coverage do you recommend for CRISPR/Cas9 on-target and off-target detection analysis?

Our recommended sequencing depth of the on-target and off-target detection is over 30X.

Can Novogene provide genome integration & insertion detection analysis?

Yes, we do. The service includes:

a) Vector Evaluation (not required for commercial vector)

b) Insertion Calling: include SNP, Indel, SV, insertion point detection, and 2 Kb flank sequence

c) Copy Number Detection & Visualization of Recombination (optional)

Please consult with the Novogene representative for more information.

1.3 Whole Exome Sequencing (WES)

1.3.1 Sample requirements

Q2. Client: Does Novogene have enhanced kits or protocol for hWES samples in terms of sequencing those degraded samples?

A2. TS: Sorry, we do not have. According to the QC report, we do not recommend you process these samples in further steps. Here are reasons.

1) The quality of samples had been severely degraded. The main bands were around 500bp. 2) This size of bands was difficult to prepare the library, which may cause the failure of library or lead to lower data output.

Thus, we hope you could re-extract samples and send new batch to us.

Can Novogene construct libraries for low input gDNA samples from FFPE for hWES service?

Yes, our Tianjin lab has successful experience in library preparation for low input samples from FFPE, and the successful rate depends on the amount and the quality of the gDNA samples. For low input samples from FFPE, if the amount of gDNA samples is 400~800 ng and the main fragment size is

500~1000 bp, the success rate for library preparation is ~96%; if the amount of the gDNA samples is 200~400 ng and the main fragment size is 500~1000 bp, the success rate is ~89%. But if the amount is lower than 200 ng, we don't recommend library preparation at risk.

1.2.5.1 If the client would like to combine the sequencing data from different exome capture kits and perform data analysis altogether, can Novogene do this?

It is not recommended to perform the data analysis with the data from different exome capture kits. Due to the difference of capture region of kits, if the variants cannot be detected, we cannot know whether it is because of the difference of samples or the kits. The best way is to use the same kit for lib prep for all of the samples and do the sequencing and data analysis.

2. Regular RNA products

2.1 Eukaryotic mRNA sequencing

If I only have a NanoDrop to assess the quantity and quality of my samples, how will it compare to your in-house QC?

We always recommend QC using Qubit (concentration) and BioAnalyser2100 for RNA integrity (degradation) We always prefer samples of less amount and high quality, over more amount and bad quality.

260/280… A ratio of ~1.8 is generally accepted as "pure" for DNA; a ratio of ~2.0 is generally accepted as "pure" for RNA. If the ratio is appreciably lower in either case, it may indicate the presence of protein, phenol or other contaminants that absorb strongly at or near 280 nm.

260/230… This ratio is used as a secondary measure of nucleic acid purity. The 260/230 values for "pure" nucleic acid are often higher than the respective 260/280 values. Expected 260/230 values are commonly in the range of 2.0-2.2. If the ratio is appreciably lower than expected, it may indicate the presence of contaminants which absorb at 230 nm. If the 260/230 is not ideal, but the Qubit and the BioAnalyzer 2100 numbers are good, then a led then ideal 260/230 number most likely will not affect library prep or sequencing.

Q2. Client: I know my insect RNA is not degraded, why do I get such a low RIN number when I run my total RNA samples on the Bioanalyzer 2100?

A2. TS: For some special species (insects: silkworm, bee, cicada, locust, louse, etc.; Arthropoda： shrimp, crab, spider, fruit fly, mite, etc.; mollusk： scallop, oyster, roundworm, etc.), their 28S rRNA is thought to harbor a "hidden break" which cleaves under denaturing conditions to comigrate with 18S rRNA band to exhibit a degraded appearance on native agarose gels. The degraded appearance confounds determination of RNA

integrity in laboratories that rely on gel electrophoresis. As such, regardless of the quality of your total RNA, your RIN number will always be low.

Q2. Client: Can tissues or body fluids be sent for you to perform RNA extraction?

A2. TS: We provide RNA extraction services for tissues or body fluids. However, we recommend you send us purified RNA samples. If you have a need for RNA extraction, please contact your Novogene representative to discuss your specific needs.

2.1.3 Library preparation

What information will a "Strand-specific" library prep show to me, that a non-directional library prep would not?

Strand specific library prep will allow you to know what strand of DNA (+ or -) the gene is transcribed from.

Q2. Client: Would you please do RNA sequencing for low-input RNA samples?

A2. TS: 1) If total RNA>=400ug, we will proceed with regular lib prep protocol.

2) If total RNA<400ng, we will use the same kit, but use modified protocol for low input RNA (with more PCR cycles). An extra QC fee of highly sensitive Agilent 2100 will be changed in this case.

3) If total RNA<10ng, this is when we will recommend you to amplify the whole transcriptome RNA and prepare the libraries before sample submission (Low-Input RNA lib prep Recommendation).

2.1.4 Analysis

Can I perform whole transcriptome amplification on part of my samples, and compare the results with the rest of samples that went through regular mRNA library construction?

It's not advised to compare data obtained from different library construction methods. Different treatment on total RNA will introduce certain to bias to the sequencing data, and affect the validity of the comparison results.

Same applies to comparing data from stranded vs. non-stranded library; comparing data from mRNA library vs. lncRNA library etc.

Q1. Client: Can we only have mapping service for my project in analysis? If so, at what price?

A1. TS: Yes, you could choose only mapping analysis. And our sales will tell you the unit price for this service.

Q3. Client: I have done RNA-seq with a different vendor. Can I send the data to you for analysis? And can I also combine the data with my current project, and perform differential expression analysis?

A3. TS: Yes, it may be feasible. And could you please provide the accurate information for analysis on the dataset generated by another vendor? The following are details that we require.

Sample number, **Species** (Latin name and reference genome); **Lib prep method** (Kit used), **Sequencing platform & strategy**, **Data amount** (per sample),

Data format. Ex. Raw data in fastq format. BAM files etc. (If the data has been processed, please let us know the software and parameters used),

Detailed analysis requirement.

For comparison with a dataset generated by Novogene, the lib prep method and sequencing strategy should be the same.

2.2 Prokaryotic mRNA sequencing

2.2.3 Library preparation

Q1. Client: Could you please let me know the library preparation method that you will apply for prokaryotic RNA?

A1. TS: For prokaryotic RNA samples, they will be prepared for 250~300 bp insert strand specific library with rRNA removal.

3. Premade-library

Q2. Client: What are PhiX Control Libraries?

A2. TS: PhiX are control libraries generated from the PhiX virus. It composes of 45%AT and 55%GC bases. Adding PhiX library could enhance the library diversity for lane-sequencing and improve data quality. And Phix also serve as an effective control in sequencing runs.

Q3. Client: Why should we spike in PhiX while sequencing?

A3. TS**:** For most libraries, Illumina recommends using PhiX Control in a low-concentration spike-in (1%) to monitor sequencing quality control for cluster generation, sequencing, and alignment.

For unbalanced samples such as WGBS libraries, use a higher concentration PhiX spike-in to improve crosstalk and phasing calculations.

For samples with low diversity, use a high-concentration spike-in (5% or higher) of PhiX to create a more diverse set of clusters. Here I list the library types which require Phix.

Library Type	**Characteristics**
Restriction-site Associated DNA sequencing (RAD) Library	Low Diversity
Genotyping by Sequencing (GBS) Library	Low Diversity
10x Single Cell RNA Library	Low Diversity
Single-cell DNA /RNA Library	Low Diversity
Amplicon Library	Low Diversity
Microbe Re-sequencing Library	Unbalanced
Whole Genome Bisulfite Sequencing (WGBS) Library	Unbalanced

If my library has adapter contamination, how would this affect my data quality?

Library with adapter contamination would affect the cluster generation process and lead to high adapter rate. Novogene recommend to use AMPure beads for adapter removal before sequencing.

Q5. Client: What standard approach would be used for pooling?

A5. TS: It only depends on the molarity of each sample. And we would not arrange pooling based on library concentration quantified by Qubit.

3.2 Sequencing

Q1. Client: What kinds of sequencing strategy are available?

A1. TS: There are several types.

1) PE150 on HiSeq X, NovaSeq platform.
2) PE250 on NovaSeq SP
3) SE50/PE50 on NovaSeq SP platform.

Q2. Client: Can I use custom sequencing primer for sequencing?

A2. TS: Yes, you could. However, samples with custom primer would be shipped to China for sequencing, as our China lab has more experience to handle with custom sequencing primer.

Q3. Client: Could we order one lane sequencing with custom read1 primer?

A3. TS: Here are several cases.

*For HiSeq X,

Custom Read1 Sequencing Primer-order by lanes;

Custom Read2 Sequencing Primer-order by flow cells;

Custom Index Sequencing Primer-order by flow cells.

*For NovaSeq S2/S4

Custom Sequencing Primer-order by flow cells.

Note: The client will need to provide the sequence of the primer, form of delivery (lyophilized etc.), amount in nmoles, and desired volume of buffer to add for desired concentration.

Q5. Client: We will have 3 scATAC-Seq libraries (10k cells per lib) for sequencing. Is Novogene team experienced to do sequencing on them? Which sequencing platform is better? What's the optimal sequencing depth per cell? And what else I have to know? It is a little bit urgent. Thanks for your reply.

A5. TS: Thank you for contacting us. Yes, we have lots of experiences. Due to the special read length of 16bp i5 index, we need to change the settings of the machine for a whole flowcell. Thus, a whole flow cell is required to order. The recommended sequencing depth is 25,000 read pairs per nucleus. So, the total data output you need is 25,000 read pairs/nucleus*10,000 cells/lib*3lib=750M read pairs.

According to the output you need, NovaSeq SP platform is the best choose, the output of which is 800M/flow cell. The sequencing strategy is PE50. But the read2 will be 49bp, 1 bp shorter than the kit requires, because of the length of index5. It has been verified that it does not affect the results analysis. Besides the 1% PhIX spike-in is strongly recommended.

Here is the link for more details. https://support.10xgenomics.com/single-cell-atac/sequencing/doc/specifications-sequencing-requirements-for-single-cell-atac?tdsourcetag=s_pcqq_aiomsg

Sequencing Depth & Run Parameters

Sequencing Depth	25,000 read pairs per nucleus (25,000 reads for Read 1N; 25,000 reads for Read 2N)
Sequencing Type	Paired-end, dual indexing
Sequencing Read	Recommended Number of Cycles
Read 1N	50 cycles
i7 Index	8 cycles
i5 Index	16 cycles
Read 2N	50 cycles

3.3 Data analysis

Q1. Client: Can I get BCL file for my libraries?

A1. TS: BCL file could be provided only for flow cell sequencing. If only one lane is ordered, Novogene would not provide BCL file.

4. Epigenetic products

4.1 Whole Genome Bisulfite Sequencing (WGBS)

4.1.1 Sample requirements

Q2. Sales: Could our client perform bisulfite sequencing in the whole tissue (like placenta with different cell composition) is correct or not?

A2. TS: WGBS is helpful for investigating parental expression bias. However, we do not regard placenta as the appropriate tissue for this study because of its mixed origin of the parent and offspring. PBMCs (peripheral blood mononuclear cells) may be more appropriate.

4.1.3 Sequencing

Q1. Clients: Could you please let me know the appropriate sequencing depth?

A1. TS: We recommend you have 30X. If you would like to have lower data amount, 20X is also acceptable.

4.2 Reduced Representation Bisulfite Sequencing (RRBS)

4.2.1 Sample requirements

4.2.2 Sequencing

Q1. Client: I need to do an epigenome analysis of 58 human samples. We also need the bioinformatics. We have previously used EPIC bead chip of Illumina. Do you use this platform?

A1. TS: Here are my opinions.

1) We can provide the EPIC bead chip sequencing.

2） Here I would like to introduce a new product for methylation sequencing-RRBS (Reduced Representation Bisulfite Sequencing). RRBS was developed for animals primarily. RRBS is an accurate, efficient and economical method for DNA methylation research. Enrichment of promoter and CpG island regions by enzymatic cleavage (Mspl), combining with Bisulfite sequencing, achieved high resolution of DNA methylation status detection and high utilization of sequencing data.

4.3 Chromatin Immunoprecipitation Sequencing (CHIP-seq)

4.3.1 Sample requirements

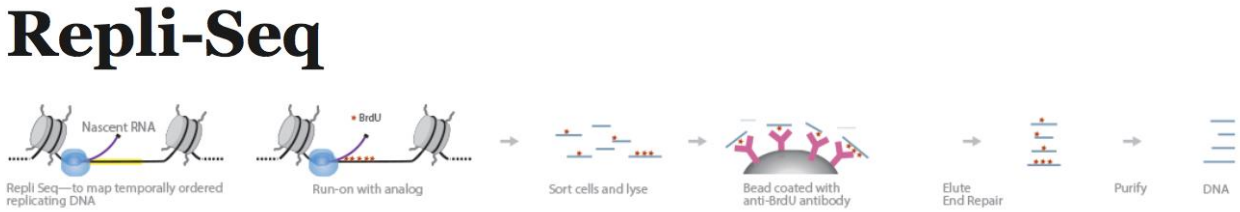
4.3.2 Sequencing

Q1. Client: Does Novogene have protocols for Repli-seq and what the analysis content for this sequencing?

A1. TS: We are so sorry that Novogene do not have related experience for Repli-seq. However, the principle of Repli-seq is similar to that of CHIP-seq. And we had performed a large number of ChIP-seq projects with successful cases. Would you mind us applying the workflow of CHIP-seq to process your project?

As for bioinformatic analysis, our team will construct a new mode for Repli-seq analysis based on your project. And some part of analysis contents for ChIP-Seq will be also used in Repli-seq, like strand cross correlation, peak calling, motif, differential analysis， annotation, enrichment and so on. If you agree with the analysis content which may satisfy their study purpose, we will generate the quote for you immediately. If you have special requirements for analysis content, please let us know and our team will evaluate the possibility that we could arrive at as well as the TAT & Price. And the evaluation time will be predicted longer than before.

Note: Cycling cells duplicate their DNA content during S phase, following a defined program called replication timing (RT). Early and late replicating regions differ in terms of mutation rates, transcriptional activity, chromatin marks and sub-nuclear position. Moreover, RT is regulated during development and is altered in disease. Exploring mechanisms linking RT to other cellular processes in normal and diseased cells will be facilitated by rapid and robust methods with which to measure RT genome wide.



Repli-Seq maps the sequences of nascent DNA replication strands throughout the whole genome during each of the six cell cycle phases. This is achieved by growing cells in bromouridine triphosphate (BrdU) media to replace thymidine. The cells are then sorted to their current state in cell division using

Fluorescent-activated cell sorting (FACS). BrdU-labelled DNA strands are immunoprecipitated by anti-BrdU antibodies on magnetic beads. These immunoprecipitated strands can be prepared for sequencing following TruSeq DNA library preparation protocol.

Pros: Maps sequences of newly replicated DNA to the phases of cell division;
Low sample input required (5000 cells) makes it suitable for studying rare cell populations;
Streamlined DNA library prep step.

Cons: The BrdU labeling requirement limits this approach to cultured cells

Q2. Sales: Would you provide the service for sequencing 24 yeast samples and 0.04Gb per sample for CHIP-seq?

A2. TS: We are so sorry that we generally do not accept such low data output per sample. However, to meet our client requirement, here is our recommendation. Since the minimum amount of data on the Tianjin Labs nova S4 PE150 strategy is 1G per sample, the nova SP SE50 strategy is recommended for the yeast Chip-seq project. It is worth to mention that our lab could not accept lib prep separately and sequence them together. Thank you.

4.3.3 Analysis

Q2. Client: we have opted for a "spike-in" approach which is basically a kit purchased from Active Motif that addresses normalization issues related to the target we have ChIP-ed (H3K27me3). It involves adding a small amount of Drosophila chromatin to each sample, which is immunoprecipitated along with our target of interest. For the analysis, we would need the sequences mapped to both the human/mouse and Drosophila genomes, and then the Drosophila reads can be used to normalize the human/mouse data. Does this sound feasible?

A2. TS: Thank you for your details. The Drosophila chromatin would act as the spike-in, and we can use the Drosophila reads to normalize the human/mouse data. We can include this normalization in the standard analysis.

Q3. Client: Our protein (Trib2) is a pseudokinase and its interaction with DNA is not much known. Our immediate aim is to find such interactions by ChIP-seq and then characterize them, identifying mechanism(s) of Trib2 regulation and both up-/ downstream.

A3. TS: ChIP-Seq provides genome-wide profiling of DNA targets for histone modifications, transcription factors, and other DNA-associated proteins. It can be used to find the binding sites of Trib2 on genome if it binds.

At least one negative control is indeed needed to eliminate the noise/background. Because the interaction of Trib2 with DNA is not much known. And the positive control could determine the effect of the ChIP assay.

Additionally, do you want to study the mechanisms of Trib2 in regulating the expression of the genes that it binds to or the expression of Trib2 is regulated?

1) If you want to learn how the Trib2 regulates expression of other genes by binding to their promoter, it is quite a challenging and innovative project because the Trib2 don't have the function as a pseudokinase as far as I know (could we ask a paper for reference with this). ChIP-seq can be exactly the essay they need.

2) If you want to study how the expression of Trib2 is regulated, ChIP-Seq of Trib2 cannot meet their goals.

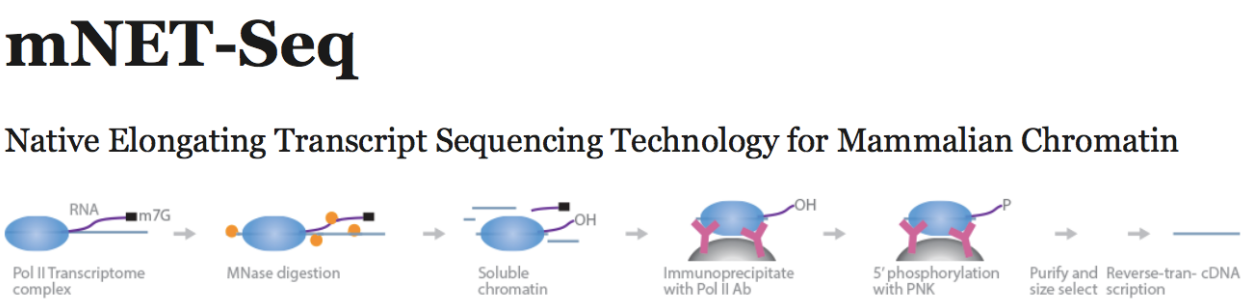
4.4 RNA Immunoprecipitation Sequencing (RIP-seq)

4.4.2 Analysis

Q1. Does Novogene have protocols for mNET-seq and what the analysis content for this sequencing?

A1. We are so sorry that Novogene do not have related experience for mNET-seq. However, the principle of mNET-seq is similar to that of RIP-seq. And we had performed some RIP-seq projects with successful cases.

Note:



mNET-Seq generates profiles of nascent RNA and cotranscriptional RNA processing associated with different C-terminal domain (CTD) phosphorylation states throughout the whole genome (Nojima et al., 2015). mNET-Seq is able to provide precise sequence reads of RNAPII active sites during transcript elongation and also RNA processing intermediates.First, elongating RNAPII complexes are isolated through chromatin fractionation. They are digested with MNase, breaking down all exposed DNA while leaving RNA strands protected by RNAPII or spliceosomes intact. The RNAPII complexes are immunoprecipitated using RNAPII antibodies and 5′ phosphorylated by T4 PNK. Next, 3′ linkers are ligated to the 3′ hydroxyl end of the RNA strand still embedded within RNAPII. They are also incubated with radioactive ATP to facilitate size selection. Nascent RNA strands are size-selected for 35_100 nt, processed into cDNA sequencing libraries, and sequenced. The use of various RNAPII antibodies during purification raises the versatility and specificity of the technique in targeting CTDs of RNAPII.

Advantages:

Maps nascent RNA strands and cotranscriptional RNA processing during RNAPII elongation with phosphorylated CTDs;

Able to detect sense and antisense transcripts at TSS (Nojima et al., 2016);

No crosslinkingãeliminates introduction of artifactual interactions;

MNase digestion is specific and efficient;

Various RNAPII-specific antibodies can be used to increase targeting accuracy.

Disadvantages:

Nascent RNAs shorter than 35 nt cannot be detected reliably;

RNA can degrade during RNAPII immunoprecipitation;

mNET-Seq peaks might be obscured by peaks from cotranscriptional RNA cleavage;

PCR amplification may give rise to peaks from amplification bias.

5.3 Whole Transriptome Sequencing (WTS)

5.3.2 Sequencing

Q1. Client: What is the recommended sequencing depth of whole trancriptome sequencing?

A1. TS: As WTS service is included two library type, lncRNA and small RNA, lncRNA-seq will be 12G raw data/sample, strand specific library with rRNA removal. And small RNA library will be 20M raw reads/sample.

5.3.3 Analysis

Q1. Client: What analysis contents do you provide for WTS?

A1. For WTS, we have the several analyses, including analysis for lncRNA & mRNA, analysis for circular RNA, analysis for small RNA and association analysis.

5.4.2 Analysis

Q1. Client: Do you have association analysis between ribo-seq data and mRNA-seq data?

A1. TS: Yes.

6. Microbial Sequencing

6.1 PCR product project

6.1.1 Sample requirements

6.1.2 Sequencing

Q1. Client: Could Novogene provide primer synthesis service for my PCR-product project, or use the previous one that had already be in your system for a repeat? Could you provide me only 100,000 raw reads per sample?

A1. TS: According to our policy, we can save the samples for 3 months after data release, I am sorry that the old samples are discarded, and we are not providing the service for primer synthesis anymore, since our current platform is incompatible with customized PCR products. If the primer could be synthesized by outsource company, we can provide two options for you in the further steps:

1) If you can send us the PCR product after amplification, we can perform it on Novaseq, with PE250 strategy, 1G data output is minimum.

2) Or you can add barcode on each of your amplicon and mix them into one or two tube, then demultiplex

by barcode after sequencing, this way will be more cost-effective.

As you only require 0.1M reads per sample, according to our new price list, we can do mixed library for your amplicons. In this case, each sample can get 100,000 raw reads, but you need to add 8bp barcode for each amplicon and provide us the barcode sequence and position.

Q2. Client: We are interested to survey variations in 10 target genes between 50 different samples/strains of wheat. We will be amplifying 10 different genomic regions from each sample using PCR and are considering whether to send us pre-prepared libraries for each sample or amplicons for library prep and sequencing at Novogene. The size of the amplicons/target regions we are interested in is around 4000bp. Here are my two questions:

1) For your PCR amplicon sequencing service, would it be possible for them to combine 10 separate amplicons (unique genomic regions) from each sample to a single tube for library prep at Novogene? (50 samples. 10 amplicons per sample -> pool amplicons; 50 library preps (one per sample))
2) For SNP/InDel calling in these target regions, what do you think would be an appropriate coverage?

A2. TS: Hope you everything goes well. First, I am writing to answer your question. Yes, those amplicons could be pooled in one single tube after adding 8bp barcode. And the size of each amplicon is no more than 450bp. In addition, the barcodes may not be terribly helpful if the client wants the full sequence for a PCR product that is 1kb+. Thus, we recommend the client to prepare library for each sample rather than to pool them together. Moveover, 30X should be sufficient for SNP/InDel calling in these target regions. As our client did not exactly require bioinformatic, here are our four strategies. The first two are for sequencing on Pacbio. And the last two are sequenced by Illumina.

1) PCR product **with barcode** sequence

   Sequencing strategy:

   Lib-pre - Pool PCR products together and construct one PacBio DNA library.

   Seq – Being sequenced on one Sequel SMRT cell.

2) PCR product **without barcode** sequence

Sequencing strategy:

Lib-pre – Do not pool the PCR product together; one PCR product per PacBio DNA library.

Seq - Be sequenced on one Sequel SMRT cell.

3) Premade library (Illumina platform)

Sequencing strategy:

Lib-pre - Pool of 50-amplicon libraries

Seq - 1 lane HiSeq

4) PCR-product (Illumina platform)

Sequencing strategy:

Lib-pre - one PCR product per library

Seq – Ilumina PE150

For Pacbio, it does not do well in SNP/Indel. For Illumina platform, it is too difficult to assembly with 4000bp. There will be several gaps or blankets. Our team cannot accept PCR product with this data output, as a result, we do not recommend samples being sequenced by Illumina over 2000bp.

However, if you requested Illumina seq and you did not prepared primers/barcodes for their existing amplicon preps. Solution 4 could meet you interest.

Q4. Client: We need to test the stability of recombinant virus obtained from chicken embryo eggs. Usually, we passage the recombinant virus up to 8-10 passages and check HA and NA gene sequence at each passage. What we need is to identify mutations at each nucleotide position of influenza HA and NA genes, also, we need to know the frequency of minor variant at each nucleotide position. It is virus PCR product lead for 1800 samples. And we will ship them in batches of 24 samples over the next 2 years. What kind of analysis could we have to meet our requirements?

A4. TS: As per your request, it will most likely go through our WGS pipeline--aligning the PCR product sequences to the reference genome sequence and then identify mutations such as SNPs and variant

frequencies.

Hence,

1) We need to know the length of the PCR products.

2) We need reference genome of the species.

3) Do you want to do assembly based on sequencing data?

Sales: The genes we amplified are HA and NA genes from influenza A. We will have different influenza A viruses each time. The length of HA is between 1.7-1.8kb and NA is between 1.5-1.6kb. The length may vary with different viruses but will be in that range. Out target sample size is 1800 in two years. We will try to run 24 samples each time. Yes, we do want to assemble sequence for each gene based on the sequencing data.

TS: We can perform this customized analysis.

6.2 Metagenomics

4.2.5.1 **What is the sufficient depth to identify a certain kind of genes, say, all the genes of antimicrobial resistance in the samples?**

Hard to say. The microbe content is different in different samples, thus need different depth of sequencing. Need to refer to papers for more information.

The short answer is there is no easy way to estimate read depth required for shotgun metagenomics sequencing. Environmental samples have a large distribution of species; each species would have to be accounted for individually. You would need to know the number of total species in the sample, the genome size and relative abundance for each species. In most cases this is not possible when you're sequencing a sample for the first time.

Let's assume you were dealing with a simple sample that had 10 bacterial species and wanted 100x coverage depth for de no assembly. If your 10 bacterial species had an estimated genome size of 2 Mb, you'd aim for around 2 Gb of sequencing data per sample.

10 dominant bacterial species * 100x * 2 Mb = 2 Gb
Reference https://genohub.com/shotgun-metagenomics-sequencing/
Our data demonstrate that D0.5 with ≥50 million reads would be a suitable compromise for sequencing bovine fecal samples and adequately inferring their resistome, considering that no further classes were discovered by the D1 sequencing depth and only a single unique mechanism was discovered as compared to the D0.5 level.

Reference https://www.nature.com/articles/s41598-018-24280-8

So 50M reads is the optimized data amount. We could see how that goes first.

4.1.1.1 Can we use metagenomics data to do amplicon analysis or vice versa?

NO, we cannot. From the perspective of sequencing regions, Meta sequences the whole genome, while the amplicon sequences the specific target region in the genome and this region is usually very short. Even if the result of Meta-sequencing contains the desired target region, the sequencing depth of meta and amplicon is not the same, and it is impossible to analyze the target region.
Secondly, for the purpose of sequencing, the purpose of meta is to study the genome and its function of a microorganism, while the purpose of amplicon sequencing is to study the microbial community structure in a specific environment. Those two purposes are different, and the intermediate process from sample acquisition to final data acquisition is different. Hence, meta data cannot be used for amplicon analysis and vice versa.

6.4 Small genome sequencing

6.4.1 Sample requirements

Q1. Client: What are your sample requirements for small genome sequencing?

6.4.2 Sequencing

Q1. Client: I am also very interested in small genome sequencing (mitochondria, chloroplast). Right now, I like to do 6 samples of chloroplast genome sequencing in mixed tomato and eggplant. And I would like to send you the fresh leaves for each sample as well and ask you for extracting the chloroplast DNA. Please let me know whether you can do and what is the cost for the extraction and the sequencing.

A1. TS: Thank you for your email. We are so sorry. Thus, we recommend you to extract the chloroplast DNA yourselves and send DNA samples to us. And then, we could proceed the library preparation with the chloroplast DNA. Thank you.

7. Third Generation Sequencing

7.1 Pacbio Sequel DNA sequencing

7.1.3 Sequencing

Q1. Client: Our goal of using long read sequencing is to be able not only get the whole genetic and transcriptomic information like SNPs and eQTL mapping but be able to do so in a haplotype dependent manner. Thus, we are hoping that long read sequencing will aid with the haplotype phasing – for example in case of allele specific expression for heterozygous variants? The samples for this particular project are human spinal cord samples and are not particularly disease specific. All of the samples will be human samples.

A1. TS: Thank you for your reply. Here I need to mention that the requirement of sample number is needed by the software, which is used for eQTLs analysis (MatirxeQT). It requires no less than 70 samples to be processed. Could you please increase your sample number to 70 samples?

Here I displayed four options for you.

1) Pacbio DNA sequel II (1 SMRT cell /sample) + Iso-seq (15G/sample)
2) Pacbio DNA sequel II (1 SMRT cell/ /sample) + RNA-seq (Illunima, 6G/sample)
3) Human WGS (Illumina, 30X/sample) + RNA-seq (Illunima, 6G/sample)

It is worth to mention that PacBio platform is not so accurate in SNP detection because of the mechine characterization. You could also choose the fourth option.

4) Pacbio DNA sequel II (1 SMRT cell/ /sample) + Human WGS (Illumina, 30X/sample) + RNA-seq (Illunima, 6G/sample)

Appendix

1. Service workflow:

Client: How is the project carried out?

TS: After we receive the PO for your project, your Novogene representative will send you the sample information form and shipping instruction. Please follow the instruction and choose a reliable carrier for sample submission. Once your samples arrive at our laboratory, sample quality control testing will be started immediately. However, library construction and sequencing will begin only after we receive your confirmation. After your project is completed, your Novogene representative will send you the data report and the invoice via email. Once the invoice is confirmed correct, your Novogene representative will release the data to you via FTP or agreed approach. Please follow the instructions on the invoice for payments.

2. Sample shipping address:

Client: Could you please let me know the sample shipping address as well as the recipient?

TS: Yes. Here is our address. And Kevin will be the recipient.

ATTN: Sample Receiving Department

Novogene Corporation Inc.

2921 Stockton Blvd.

Suite 1810

Sacramento CA 95817

Tel: 916-701-5130

3. Couriers for shipping:

Client: Which couriers do you recommend for shipping samples to you?

TS: For US domestic shipping, we recommend FedEx (fedex.com).

For international shipping, we recommend FedEx (fedex.com), UPS (ups.com), DHL (dhl.com), TNT (tnt.com). Please choose the "priority" option for international shipments. No matter which courier you choose, please select a reliable one and make sure that the courier can facilitate the import of DNA or RNA, and dry ice packaging (if applicable), to China. For more details, please see the Sample Preparation & Shipping page.

6. Data delivery:

Client: How could I obtain the data after the project is completed?

TS: We could offer three options for data delivery:

1) Data download through our secure FTP server for free;

2) Amazon cloud data download and will be charged.

3) Hard drive/U-disk sent via FedEx air transportation and will also be charged.

If you have made the decision, please contact our sales. They will provide the price for you.

7. Library return for non-premade projects:

Client: Could I get the remainder of the libraries to me, when the project is completed?

TS: We are so sorry. If the libraries have been constructed by Novogene, we do not return the remainder of the libraries.

8. Sample return:

Client: Could I get samples back, when the project is completed?

TS: Yes, of course. Please provide the information of your Fedex account, address and phone number.

9. Trimmed data requirement:

Client: Could I get trimmed data, when the project is completed?

TS: Yes. It is for free to get trimmed data, as long as you tell us your requirement before we release data to you. When you would like to get the trimmed data after you receive our project report, it is free for the project with no more than 20 samples.