Published : 01-01-2018
Last revision : 07-01-2019

DNA Link Sequencing Lab

Sales Guidelines

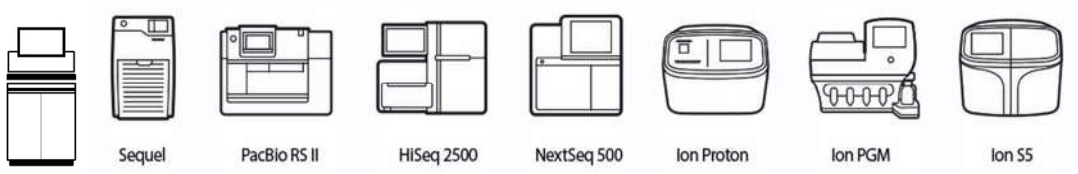
Global sales and marketing team

DNA Link Inc.

1.Introduction

This document is written by Jungsoo Park, Director of international marketing and sales, and intended for providing the basic knowledge and information about the guidelines for the distributors of DNA Link's sequencing services. The contents may be subject to change without prior notice and it is strongly recommended to check the version and revision date of this document every time this guideline should be referred to. This document is highly confidential and not to be duplicated or revealed to outside of the company without prior written confirmation of the author.

3 Sequencers



Our fleet of sequencers include both second and third generation platforms including NovaSeq, HiSeq and NextSeq from illumina , Sequel and RSII from Pacific Biosciences, and ion torrent and S5 from Thermo Fisher Scientific, Minion and Promethion from Oxford Nanopore Technologies. From genotyping by microarray to the most recent technologies of next and third generation sequencing, DNA Link Sequencing Lab is equipped with all diverse platforms with which it can provide a suited sequencing service to the researchers.

Platform selection guide

		PacBio		illumina		Thermo Fisher
	Application	Sequel	RSII	NovaSeq	NextSeq	Ion Torrent
DNA	Whole Genome Resequencing	V	V	V		
	De novo sequencing	V	V	V		
	Exome sequencing			V	V	V
	Targeted resequencing	V	V	V	V	V
RNA	RNA-Seq (Transcriptome)			V	V	V
	Expression profiling			V	V	
	Isoform sequencing	V	V			
Others	Epigenetic sequencing	V	V			
	Metagenome sequencing			V	V	

In brief, PacBio sequencers are the best for sequencing unknown samples to the highest level of sequence quality, in other words, the least number of contigs. For Bacterial genome we have seen countless projects which ended with less than a few contigs, or even one single contig. Also, it gives unprecedented quality of sequence results for RNA-Seq when the correct detection of splice variants is important. Thanks to its long readlength (over 20kb), it does not need assembly of transcripts which gives the best confidence about the structure of splice variants.

PacBio Sequencing services are sold per “SMRTcell”. Currently one SMRTcell is capable of generating around 10Gb of sequence data.

illumina sequencers are the best for resequencing of a known genome. By far it has the best cost efficiency in all types of sequencers which made its wide popularity possible. It is powerful especially when you have a reference genome, or sequence only a part of a genome (targeted sequencing, exome sequencing) It is also ideal for expression profiling is where the sequencing depth matters.

As of June 2018 we are equipped with NovaSeq, the latest version of illumina sequencers which produces up to 3Tb per flowcell. We are running the unit with 100PE mode in S2 (total 660Gb) and 150PE mode in S4 (3T). HiSeq service is discontinued as the adoption of NovaSeq.

We are offering NovaSeq HiSeq X, and NextSeq services.

4 Service portfolio

Whole Genome Sequencing (WGS, human)

As of the revision date of this document, human whole genome sequencing is the most competitive when done with NovaSeq systems. WGS by PacBio is not recommendable at the moment due to the high price and difficulty in the assembly.

The most common set up of WGS is 30X sequencing (=90Gb) per sample. With this coverage we can run about 8 sample per lane of NovaSeq S4. The cost of the data delivery(one HDD costs $250) and the data analysis is optional.

Based on this, when a new inquiry for WGS is in place, we may give a ballpark price range of $1,200-1,500 per sample, and then we need to figure out the following aspects of the project.

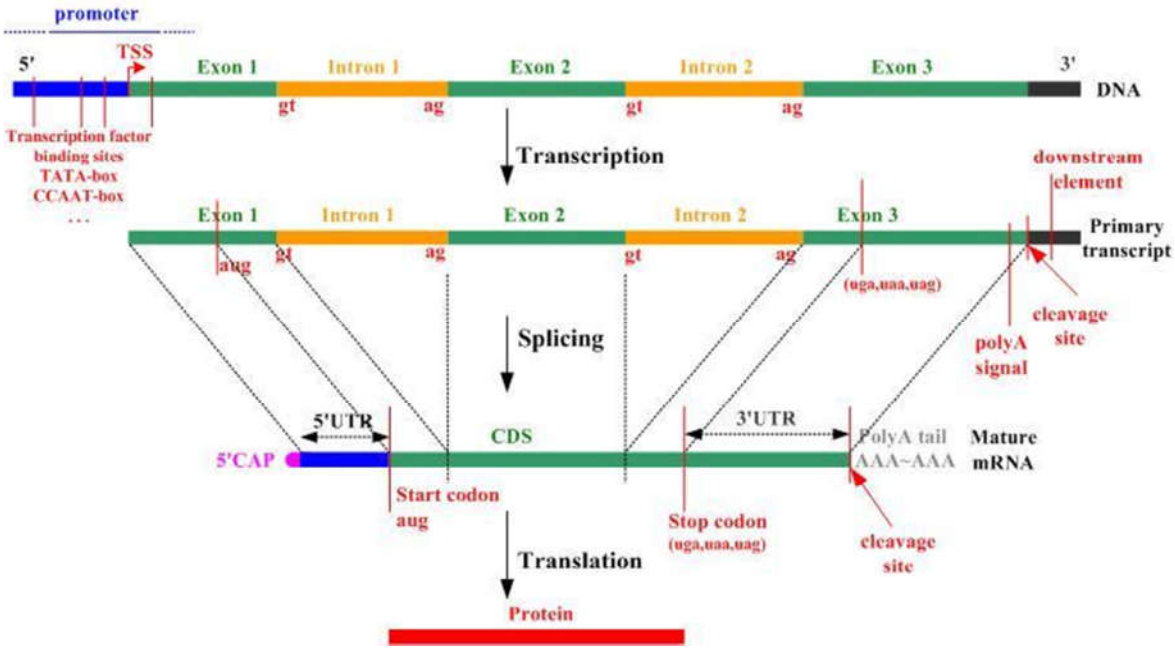
- The number of samples (Total number, batch size, frequency)
- Expected level of sequence data
- Project start time point
- Expected turnaround time

Once this information is in hand, we can contact our outsourcing partners and get the most fresh quote, and give As a company who does not have the key sequencing platform in our hand, we may be less competitive in pricing. Please arrange our marginal profit around 10%-20% of the supply price.

Whole genome sequencing (Bacteria, plant, animal, and so on)

While human whole genome is dominated by illumina’s sequencers, PacBio gives good advantage in the whole genome sequencing of other organisms which has weak or poor genome references. We provide two types of library options, 10kb and 20kb, followed by PacBio Sequel’s

Whole exome sequencing (WES, human)



A newly synthesized RNA molecule from a region on DNA (gene) is called a primary transcript. The primary transcript of an eukaryotic mRNA typically contains sequences encompassing one gene, although the sequences enconding the polypeptide(or protein) may not be contiguous. Nonconding tracts that break up the conding region of the transcript are called "**introns**", and the conding segments are called "**exons**". In a process called **splicing**, the introns are removed from the primary transcript and the exons are joined to for a continuous sequence that specifies a functional polypeptide. Spliced transcript is called “mature messenger RNA, ie, **mRNA**. mRNA further serves as a direct template for making protein, which process is called “Translation.”

Sequencing the DNA regions encompassing the exons is called exome sequencing. Sequencing the complementary copy of mRNA is called RNA-Seq or transcriptome sequencing. These shall be discussed further in **Service portfolio** section.

The word “Exome” is conjunction of two words, exon + chromosome. A gene consists of many regions including promoter, exons, introns and so on. Exome sequencing is sequencing only the exone regions of a whole set of chromosomes. Whole exome sequencing is one of the most common form of human genome sequencing, in which only the coding regions and UTR(untranslated regions) are captured and sequenced. Exons takes about 1% of a whole human genome (ie, 30Mb), and with UTR of 20Mb, a conventional exome sequencing sequences about 50Mb of regions. 100X on these region will require 5Gb of sequences, which makes it a very economical alternative to the whole genome sequencing which requires about 100Gb to cover 30X of the whole genome. Exome sequencing is ideal in that it only deals with the regions which are considered “important” with a higher probability of keeping meaningful genome biological information, but as it does not cover the other regions of intron, it is not a complete solution for the researchers who wish to see the entire genome.

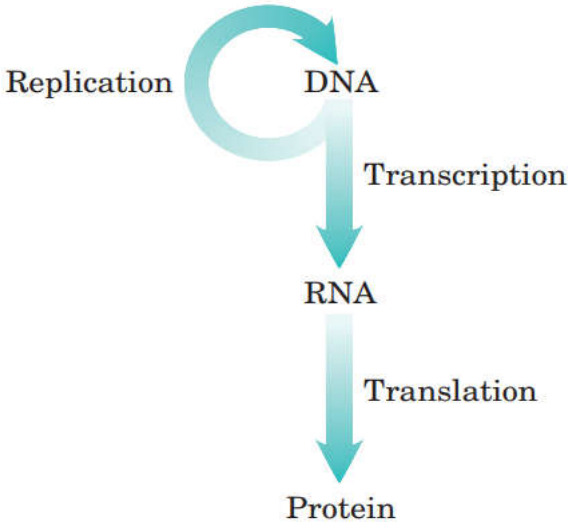
We use Agilents exome capture kit V5 which captures 50Mb region. Sequencing shall be done by HiSeq2500 or HiSeqX. In the former case, we will use our own sequencing unit, while the latter we use

our external partners. As of this revision date, the oursourcing partners for HiSeqX are Wuxi NextCode, Macrogen and Novogene.

Targeted sequencing

Targeted sequencing is ideal for sequencing only the region of interest. WES may be considered as a special type of targeted sequencing in which the region of interest happens to be the exon regions. Contrast to exome sequencing which are of high demand in the market, targeted sequencing is an application which needs customization based on the customer's need, and the capture kit should be customized and produced per customer's demand. Roche Nimblegene produces the customizable capture kits whose price is not insignificant. So this application is most widely used by the customers who want to sequence numerous samples to be captured by this custom-made DNA capture kit.

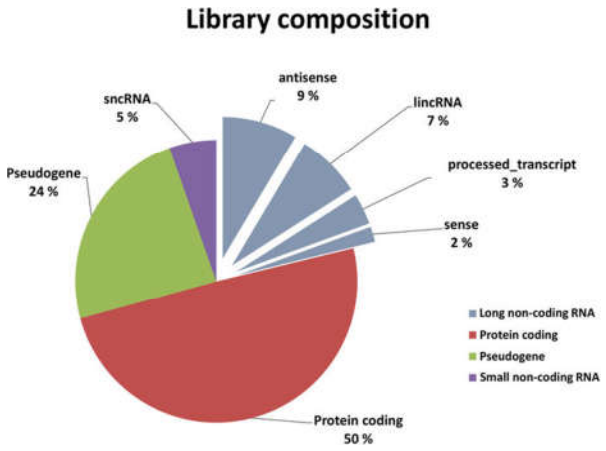
Introduction to RNA sequencing



The central dogma of moelcular biology, showing the general pathways of information flow via replication, transcription, and translation. The term "dogma" is a misnomer. Introduced by Francis Crick at a time when little evidence supported these ideas, the dogma has become a well-established principle.

RNA-Seq (transcriptome sequencing)

The word "transcriptome" is a conjunction of two words, transcript + chromosome, where transcript means the newly synthesized RNA strands via the process of transcription. Sequencing entire transcripts in a cell is called transcriptome sequencing, or RNA-Seq. Normally, RNA-Seq is done for the cells from a specific tissue, or the cells at a specific growth stage. Sequencing all transcripts from different types cells gives insights to which genes are highly (or low) expressed in which cells. Comparing more than two sets of different RNA-Seq data to contract this difference is called differential analysis.



A typical composition of RNA transcripts in a cell after rRNA depletion

RNA-Seq includes the following types of applications.

mRNA-Seq : Sequencing only the messenger RNA (mRNA) by capturing poly(A) tailed transcripts. It is the most widely used and common type of RNA-Seq. We use TruSeq stranded mRNA-Seq library prep kit followed by HiSeq or NextSeq sequencing.

Total RNA-Seq (whole transcriptome sequencing, WTS) : Instead of capturing mRNA, it depletes the most abundant but meaningless type of RNA, the rRNA. Depletion of rRNA leaves mRNA, long-non-conding RNA (lncRNA) and so on. Normally, the sequence reads from total RNA-Seq contains about 80-90% of mRNA sequences. We use TruSeq total RNA-Seq library prep kit which contains Epicentre ribo-depletion kit. Sequencing will be done by HiSeq or NextSeq.

Small (or micro) RNA-Seq : microRNA (miRNA), small RNA are the RNA which is about 20-25bp long. Previously, these RNA were considered as remaining from transcription process or degraded RNA molecules, but it was known that there are genes which are encoding these small RNAs, and they play a crucial role in the gene regulation and so on. We use TruSeq small RNA-Seq library prep kit for the service followed by HiSeq or NextSeq run. As the RNA are short, run mode with the shortest readlength is enough for this application.

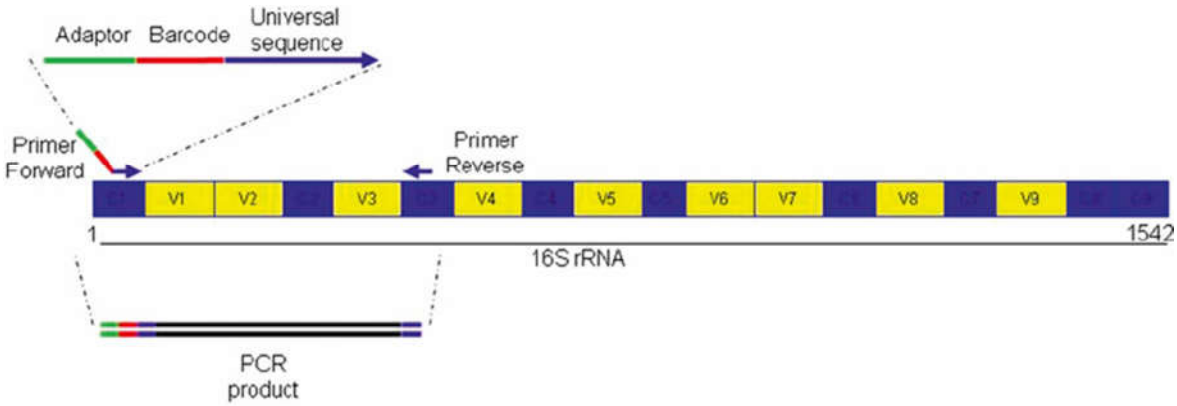
Single-cell RNA-Seq : While all the sequencing applications listed above are based on the cluster of cells which are considered to possess homogenous biological character, single-cell RNA-Seq is the application which sequences the RNA from just one single cell. For this purpose, there should be a cell sorting (or library prep per individual cell) should precede our services. If the research already made the library by 10X or fluidigm platforms, we can simply sequence them in HiSeq or NextSeq. If the customer only extracted RNA from the single cell, we can use Clontech's pico RNA-Seq library prep kit. (See the sample requirements manual for the amount of RNA required for this kit.)

Iso-Seq : The process of splicing gives diversity for the transcripts. Thus, one gene may product several different transcripts. The short reads from illumina is very weak in reconstituting the original sequence of each transcript, because they have all similar sequences, and unless the junctions of exons are sequenced properly. This results in many mis-aligned or mis-assembled

transcripts, making it difficult to see the correct alternative splicing. PacBio's long read provides the best solution for this. As one single read can read several kb, PacBio is capable of sequencing one single transcript till the end, eliminating the need for assembly of a transcript. Thus, more and more researchers are adopting PacBio technology for their study on alternative splicing. In our service, we provide three combination of libraries, 2kb, 3kb, and 6kb library, which are sequenced in each individual RSII cells. By this, the customer do not need to worry about mistakenly assembled transcripts, and get the most accurate landscape of the alternative splicing within a cell.

Metagenome sequencing

Metagenomics is the study of genetic material recovered directly from environmental samples. The broad field may also be referred to as environmental genomics, ecogenomics or community genomics. While traditional microbiology and microbial genome sequencing and genomics rely upon cultivated clonal cultures, early environmental gene sequencing cloned specific genes (often the 16S rRNA gene) to produce a profile of diversity in a natural sample. Such work revealed that the vast majority of microbial biodiversity had been missed by cultivation-based methods. Because of its ability to reveal the previously hidden diversity of microscopic life, metagenomics offers a powerful lens for viewing the microbial world that has the potential to revolutionize understanding of the entire living world. As the price of DNA sequencing continues to fall, metagenomics now allows microbial ecology to be investigated at a much greater scale and detail than before.



The most widely used strategy is to amplify 16s rRNA region of the organisms contained in the sample (normally from soil or water), and sequence the regions which are able to identify the organism. Sequencing platform is the most ideal with MiSeq, thanks to its relatively long read lengths (300PE) which can cover a couple of variable regions. Entire 16s rDNA is 1.5kb in length, but the most widely used region to sequence is the V1, V2, and V3 regions spanning in around 600bp region, which makes MiSeq's 300PE mode a suitable choice for the sequencing. PacBio can cover the entire region of 1.5kb with ease, but considering the low price of illumina sequencing, PacBio can be an overkill.

6. Marketing points

With plethora of sequencing service providers most of them are competing with low price. The lowest price for illumina sequencing can be achieved by using NovaSeq with S4 flowcell which generates up to 3T bases per flowcell. (4 lanes per flowcell = 750Gb per lane).

- Long read specialist

Currently the machines from PacBio and Oxford Nanopore Technologies (ONT) are the only two commercially successful third-generation sequencers which provides very long reads. These long read sequencing technologies uses DNA molecules without any amplification, so the library prep protocol is extremely sensitive and the quality of input DNA as well as hand-on experiences for these sequencers are the determining factors for successful sequencing. We adopted PacBio as early as 2014, and we are now (2018 Mar) considering to adopt ONT, and we are one of the companies with the most extensive experience in the long-read sequencing which gives added value to our services.

- RNA-Seq specialist

While WGS/WES market are experiencing oversupply by some sequencing companies with multiple units of NovaSeq, RNA-Seq market is still distributed among many smaller suppliers. What makes the future of RNA-Seq bright is the number of samples. WGS, for example, of a human genome, there's a need for only one sequencing for one individual, as the genome sequences are not changing over time, but in RNA-Seq people should see the change of a cell over a time, change of environment, and so on, so RNA-Seq never ends up with sequencing just one sample from an individual. With the advent of single-cell RNA-Seq it is expected that the number of sequencing will be more from RNA-Seq rather than DNA-seq(WGS or WES), so it is necessary to position ourselves as an expert in RNA-Seq.

- Flexibility in illumina sequencing

The majority of illumina market is for WGS and WES, but there's substantial and growing demand for the the applications such as RNA-Seq, ATAC-seq, ChIP-Seq and so on. These applications not only require high flexibility in the experiment set up, but also requires extensive bioinformatics analysis. Thus, what we promote ourselves as should be a company that walks through the experiment with the customer, not just a sequence data vending machine.

Appendix

Revision note

2018-1-15 : Address change of DNA Link Europe

2018-2-20 : Addition of new member in the team and related changes.

2018-3-1 : Novogene added as an outsourcing partner

2018-3-5 : Adjustment of term definitions

2018-09-17 : Change of team structure, Addition of new sequencing units

2019-01-07 : Amendment of information for Aurogene