High Throughput Sequencing Technology and Applications

Kap-Seok Yang

Macrogen Inc.

High throughput sequencing is opening the horizon for genomics research.

One of the most important issues for genomics field is to find out the factors affecting various of biological and medical phenomena. Most of the factors affecting life processes can be associated with genetic and epigenetic information, such as mutations, silencing control, SNPs, CNVs and expression patterns of messenger RNAs and proteins. As such, the sequence is considered to be the fundamental information governing many areas of life processes. However, the high cost and enormous amount of time needed for sequencing has limited the use of sequencing as a fundamental solution so far.

Instead, there have been active genomics approaches to address this fundamental question in various areas of research with various tools of research. Various genetic and epigenetic variations, gene expression variation and translational variations have been addressed with high-content and high-throughput technologies such DNA chip technologies and proteomics tools. However, there are certain limitations to chip technologies as it involves the qualitative nature of hybridization technologies and now we are entering the new era where scientists are learning to pair chip and sequencing technologies to ask even better questions.

With the advent of next generation sequencing technologies high throughput genome analysis is providing the new and fundamental tools for genomics research. The next generation sequencing technologies mean that the sequencing is becoming cheaper and faster, the amount of the sample needed is becoming smaller and the preparation cost of the sample is becoming cheaper.

Overview of high throughput sequencing technologies and its impact on research.

Arguably a single most important technological factor that opened the genomic era was Sanger-based sequencing technology. Composed of Sanger chemistry, capillary electrophoresis and fluorescence labeling and detection technologies, Sanger sequencing enabled Human Genome Project possible, and many area of genomics research are still relying on this chemistry. However, the cost and timeframe of this chemistry are the factors that limit its use in research and applications. In addition, it is reaching to the theoretical limit of extending read-length, which is around 1,000 bp for now. From the pioneers of technical revolution, new types of sequencing chemistry have been developed. Instead of Sanger chemistry, next generation sequencers uses chemistries such as sequencing by synthesis, sequencing by ligation and some other third generation sequencers uses single-molecule sequencing technologies.

The practical impact of NGS technologies is derived from its high-throughput nature. Compared to Sanger

technologies which produce sequencing throughput of Mb/day NGS platforms promise the level of throughput to Gb/day. As a consequence, the jobs thought to be possible only in huge genome centers which have enormous amount of Sanger sequencers and computing facilities can now be performed in a single lab or clinic.

Area of application

The information NGS technologies are providing is currently expanding. Since NGS is basically re-sequencing machine, the most promising area for NGS system is re-sequencing. The re-sequencing can be done for the whole genome or targeted region of interest. By the principle of assembly by mapping to reference genome sequence, re-sequencing can detect various genomic variations such as SNPs, small insertions and deletions, copy number variations and even larger scale genomic structural variations. NGS can be applied to the De Novo genome sequencing of the organisms with small genome size such as microorganisms. NGS can also be applied to the other areas of research such as transcriptomics, metagenomics and other functional studies, in which researchers can obtain different level of data from either tag-based digital array data to the full sequence of transcriptomes with detailed information on variation.

Limitation of NGS technologies

In some areas of research, NGS technologies are already offering the best solution compared to conventional technologies such as microarrays. However, in some area of research where a significant amount of sequencing is involved, the sequencing cost is still too high for ordinary laboratories and clinics. Second, the sheer amount of data from HTS has proved to be the major bottle neck for this area. NGS vendors and many third party bioinformatics solution providers are having a hard time catching up the technologies. Third, the experimental design of some area require a lot of pre-treatment and sample preparation process which, in some cases, proved to be more costlier than the sequencing itself.

Which technologies to choose?

One of the first challenges to the researchers face is to decide which technologies or machines to choose to best fit their research goal. Each technology has its advantages and drawbacks, and the technologies are at once very similar to each other but also drastically different from each other. Among the many factors the careful balancing of the specification such as read-length and throughput itself is the first factors to consider. Other factors such as support structure of the vendors and local distributors should also be considered. In this fast evolving technological field, the guideline for understanding and utilizing the technologies in user-perspective will be provided in the talk.