NGS Overview

Content

- 

Sequencing Methods

Before we discuss analysis of sequencing data, lets learn about the most commonly used sequencing techniques.

The following screencast, *Introduction to Next Generation Sequencing*, will provide an introduction (or review in some cases) of NGS terminology and its major applications, as well as a quick overview of the standard file format for NGS data, FASTQ. After the video, please read the short addendum I've added below the video.

Addendum to Video Above

The screen cast above is a little out of date in one aspect, regarding the details of the current Illumina instruments. The instruments that would run up to 12 days are no longer sold, and I think most that are still left in the field are probably subject to retirement. All of the current or recent instruments coming from Illumina will complete runs in 48 hours or less. As you can imagine, the newer instruments are not only faster, but some are capable of much greater output. From a purely bioinformatics standpoint, this doesn't have to change the way we approach analysis. In most cases it simply means we can sequence more samples, faster and generally for lower cost. The current state of the art from Illumina is the NovaSeq, which can output between 1 & 3 tera-bases in around 40 hours. This translates to 8 - 24 whole-genome sequencing samples every two days. This will naturally require greater computing capacity, but that's more of an IT issue.

- 

Whole Genome and Targeted Re-Sequencing

Now that we've discussed sequence assembly, we will continue our discussion by learning about re-sequencing, which is sequencing of a known genome or a subset of it.

While the *de novo* assembly of complete genomes is important and ongoing, the vast majority of DNA sequencing, at least NGS, is in the area of re-sequencing. This is where individual genomes (or subsets) are sequenced and mapped to their corresponding reference sequence. Once mapped, the reads are compared to the reference and analyzed to find variants (SNPs & INDELs). The power here is that both known and novel variants can be found, as opposed to genotyping assays such as 23 & me, which only genotype pre-determined locations. Once variants have been identified, qualified physicians or geneticists can make recommendations based on what is known about these variants. This is the path to so-called personalized medicine.

The following screencast, *NGS Alignment & Variant Calling,* will give you a brief overview of a typical re-sequencing pipeline, and explain how to use the typical outputs to identify variants.

Source: http://youtu.be/IBpzcxU73Gg