A Tour of Next Generation Sequencing:

Inside the Washington University Genome Center

Glossary

adapter – a short oligonucleotide with a known sequence that is attached to a template strand of unknown sequence.

bias – a problem area that exists in a genomic data based on an inherent problem in the sequencing technology. For example, the 454 technology has difficulty sequencing mononucleotide strings.

bioinformatics – use of computers to study biological data, often in the form of DNA or protein sequences.

cluster – many copies of the template DNA attached close to each other on a flowcell.

Cluster Generator – the first machine in the Illumina sequencing technology; bridge amplification occurs in this machine to make many copies of each template DNA molecule.

computationally intensive – requiring very large amounts of computer time and power to determine.

coverage – the number of times a genome has been sequenced by a particular method. To calculate, take the total sequencing output (in bases) and divide by the size of the genome being sequenced. For example, the human genome is about 3,000,000,000 bases. An Illumina sequencing machine can produce 18,000,000,000 bases of sequence data in one run. So one Illumina run produces 6x coverage of the human genome.

emulsion PCR – a PCR reaction that is done in a mixture of oil and water; each individual water droplet has a novel template strand and a PCR reaction occurs in each droplet of water to copy that template.

ePCR – emulsion PCR (polymerase chain reaction).

flowcell – a glass slide with sample lanes etched on it and a cover slip positioned on top; tiny volumes of liquid can be pumped over the flowcell by machines to automate reaction steps. In Illumina sequencing the flowcell has a lawn of primers that have sequences that match the adapter sequence.

Genome Analyzer – the second machine in the Illumina sequencing technology; in this machine the sequencing reaction proceeds and is observed.

GINA – Genetic Information Nondiscrimination Act – a 2008 law passed by Congress and signed by the President that forbids employers and health insurance companies from discriminating against individuals on the basis of their genome sequence

input DNA – DNA that will be used in a sequencing reaction; in most instances adapters need to be added to the ends of these DNA fragments.

luciferase – an enzyme originally discovered in fireflies; it uses luciferin and ATP as substrates to produce a flash of light.

luciferin – a substrate for the luciferase enzyme, required to create a flash of light.

massively parallel sequencing – a technique in which many sequencing reactions occur and are detected simultaneously.

metagenomics – the simultaneous study of a group of microbes from a particular environment, e.g., the human intestines, an ocean sample, etc. No attempt is made to obtain the DNA from a single species; rather DNA is prepared from the total sample for sequencing.

microbe – a small organism (bacterium, protist, or fungus).

microreactor – a small droplet of water that contains all of the necessary ingredients for a PCR reaction, suspended in a sea of oil (emulsion).

nebulizer – a small piece of equipment used to break input DNA into small fragments before use in Illumina sequencers.

Paired End Module – a machine that can be attached to the genome analyzer in Illumina sequencing technology to regenerate clusters for paired end sequencing.

personal genomics – the study of the genome of a single individual for the purpose of diagnosing or treating disease.

picotiter plate – a glass slide with tiny wells etched in it, used to spatially separate beads for 454 sequencing.

pyrophosphate – the two-phosphate group molecule that is released when a dNTP is added to a growing DNA strand.

pyrosequencing – sequencing DNA by observing the release of a pyrophosphate following incorporation of the next nucleotide into the growing chain.

Reference sequence – a fully assembled version of a genome that can be used for mapping short DNA sequence reads for comparisons of genomes from various individuals.

sequencing by synthesis – sequencing DNA by looking at the results of DNA synthesis (by DNA polymerase copying the template) as the reaction occurs (rather than looking at the products by gel electrophoresis).

single stranded template DNA – input DNA that has attached to a matrix and has been denatured so that it is single stranded, ready to serve as a template for emulsion PCR.

sulfyrlase – an enzyme that reacts with pyrophosphate and ADP to make ATP.

TCGA – The Cancer Genome Atlas – a consortium of researchers dedicated to studying genomes (DNA) from cancer cells to look for causative mutations.

template DNA – any DNA used by DNA polymerase to generate the complementary strand; here, input DNA with adapters attached that is ready to be used for sequencing.