Genomics in Society
Forest Sciences Centre 2045 // 2424 Main Mall, Vancouver BC V6T 1Z4
https://genomics.forestry.ubc.ca // genomics.forestry@ubc.ca

Dr. Yeaman Clip #1: Definitions

Genetics is mainly dealing with classical things we can see varying; the classic things where it was discovered was Mendel's peas looking at tall versus short, looking at phenotypes and how phenotypes and different genetic variants—**alleles**—are segregating from the parents to the offspring. So, it's [genetics] the study of the transmission of **heritable variation** from one generation to the next and trying to figure out how that heritable variation is driven by the underlying genes.

Genomics is really more of a technology trying to use **sequencing** approaches to help us understand genetics. So genetics is kind of like the subject material and genomics is how we actually get at that with technology. Genetics can be seen through many different lenses: you can look at it at a trait level like Mendel's peas that are tall and short, you can look at it from particular **amino acids** or changes in the genome and whether one predominates over the other, and really anything in between that. And so genomics just provides a tool to really help us try to study how genetics works in a much more fine-scale detail.

In standard **Illumina sequencing** what we are doing is we are breaking the genome up into three or five hundred chunks of letters—so like base pairs, each base pair is a letter in the genome—and we'll look at three to five hundred base pairs in chunks and we'll read from either end of those chunks about a hundred base pairs or a hundred-fifty base pairs in and that will be our **sequence**, and we'll just do this over and over and over again billions of times and then have many, many, many of these little reads present in the sequencing file. And then the task is to put that puzzle together. And so often we do this by having a **reference genom**e and we do what is called mapping; so we take those small reads and map them against the [reference] genome and kind of stack them all up on each other and use that to infer what [base pairs] one individual has. So we can see if the individual is heterozygous or homozygous for a particular variant, based on those reads. So Illumina sequencing gives us the short reads where there is small amounts of DNA from either side of these chunks.

Other kinds of sequencing, like **PacBio sequencing**, is a longer read sequence. So you can get maybe ten thousand base pairs of sequence, but it has a different error profile, and it's more complicated to use in some ways, and it's sometimes more expensive. So, Illumina sequencing tends to be the one people use most commonly, but if you want really long molecules—so maybe ten thousand base pairs all in the same read instead of those little short ones—then you can use PacBio sequencing. And in my lab we sometimes we use PacBio to try and get at actually assembling a new genome because it is much easier to put the puzzle together if you have these big long puzzle pieces.

Sometimes you'll aim for the longer read technology when you want to make a new genome and the shorter read technology when you just want to compare different individuals you already have an idea of what the overall puzzle looks like.