1 What is genomics?

In a broad sense, "genomics" is an interdisciplinary field of biology that focuses on understanding genomes. This is a very broad definition, because the field is very broad! It includes:

1. **Medical genomics**– The study of the genome in the context of health. (examples: precision medicine of genetic, epigenetic, and infectious diseases, veterinary genomics)
2. **Evolutionary genomics**– The study of genomes in the context of evolution. (examples: phylogenomics, population genomics, comparative genomics)
3. **Structural genomics**– The study of gene product 3D structure across a genome.
4. **Genomics technology**– The study of how genomes can be sequenced.
5. **Ethical legal, and social implications [ELSI] genomics**– The study of how genomics impact society and individuals and how genomic science should be regulated.

Importantly, these sub-fields are not mutually exclusive. They can (and frequently do) interact with one another.

2 The diverse "-omics" of genomics

Potentially something you noticed from the sub-fields defined above– genomics isn't just about genomes. In other words, genomicists (people that study genomics) aren't only focused on the DNA contained within the genome of an organism. They also focus on other "-omes". For instance, genomics includes the study of (1) genomes, (2) epigenomes, (3) transcriptomes, and (4) proteomes.

2.1 Putting the "genome" in "genomics"

The DNA that makes up an organism is the **genome**. This is the one you think of when you think of genomics. Genomes can be sequenced using several technological approaches following DNA isolation; for the sake of simplicity these can be categorized as **long-read** and **short-read** approaches. The difference between these two approaches is intuitive. Short-read sequencing involves chopping the genome into short fragments (reads) which are all sequenced individually. Long-read sequencing can preserve the DNA in its intact state and sequence chromosome-length fragments (loooong reads). A few prominent sequencing companies include Illumina (primarily known for its short-read sequencing technology), Pacific Biosciences, and Oxford Nanopore (these last two companies are primarily known for their long-read sequencing technologies). Approaches using either of these technologies (or both of them together) to sequence a whole genome is known as **whole-genome sequencing** (WGS).

The timeline for the development of sequencing technologies has been divided into "generations". **First generation** sequencing was the slow, tedious process used by biologists to sequence loci and the first genomes (e.g., Sanger sequencing). **Second generation** sequencing was short-read sequencing, which first introduced the power of **high-throughput sequencing**. High-throughput sequencing is technology that can sequence large volumes of nucleic acid. **Third generation** sequencing introduced long-read sequencing. Upon its original introduction, long-read sequencing was plagued with the presence of sequencing errors that would be confused as mutations. However, as the technology has advanced the propensity to introduce sequencing errors has declined. Second-generation and third-generation sequencing is collectively called

Next-generation sequencing (NGS).

A subsample of the genome can also be examined in genomics. In other words, a whole genome doesn't have to be sequenced in order for a sequencing approach to be categorized as "genomics". The cutoff is basically arbitrary, but it There are several ways to examine a genome without sequencing every base pair. Examples include **reduced-representation sequencing** (restriction enzymes cut up the genome and then a sample of similarly-sized fragments are isolated for sequencing [e.g., Restriction-site Associated DNA sequencing [RAD-seq]]) and **targeted sequencing** (loci of interest are sequenced- this can be done using custom probe design [e.g., sequence capture] or using a already-designed microchip [e.g., microarray]). When you send off your spit to be sequenced by a company to understand associations your DNA has with a genetic disease or your ancestry, usually the sequencing company is using a microarray (commonly called a "SNP chip").

2.2 Epigenomics

Although DNA is the molecule of heredity, it is not the only entity that controls traits. The accessibility of genes to transcription machinery (e.g., transcription factors and RNA polymerase) also plays a role in determining phenotype. Changes in accessibility occur through alterations to the DNA that *are not* alterations to the nucleotide sequence itself; this is known as **epigenetics**. These alterations include modifications of the DNA directly or to histones (the proteins associated with DNA).

Histone modification can change the structure of chromatin from **euchromatin** (when histones are spread apart; DNA is more accessible and transcription is upregulated) to **heterochromatin** (when histones are tightly packed; DNA is less accessible and transcription is downregulated). Epigenetic histone modification includes methylation (the addition of a methyl group to a histone) and acetylation (the addition of an acetyl group to a histone). Epigenetic DNA modification includes methylation of an individual nucleotide itself (directly to the nitrogenous base). Methylation of histones/nucleotides decreases transcription of the genes involved. In both of these scenarios, the DNA sequence itself hasn't changed- but the epigenetic alterations are heritable. The epigenetic profile of a genome is known as the **epigenome**.

Epigenomic sequencing approaches include those that sequence loci wrapped up by histones (e.g., CHromatin ImmunoPrecipitation sequencing [CHIP-seq]), those that sequence loci that lack histones (e.g., Assay for Transposase-Accessible Chromatin using Sequencing [ATAC-seq]), and those that identify methylated nucleotides (e.g., Bisulfite sequencing- in this approach the methylated Cytosines are converted to Uracils so that downstream sequencing can identify them).

2.3 Transcriptomics

The **expression** of genes in the genome, which is *whether* (e.g., activated or silenced) / *how much* (e.g., upregulated or down regulated) / *in what way* (e.g., splicing) a gene is transcribed. The collective body of transcripts produced by a genome are known as the **transcriptome**. The transcriptome can be sequenced by isolating the RNA in a cell and using NGS to sequence the transcripts (RNA-seq).

Transcriptomics can be used to examine the effect of a treatment on genome-wide gene expression. For instance, to understand how a drug affects gene expression across the genes of the whole genome, individuals can be assigned to one of two groups: (1) treatment and (2) placebo.

After the treatment is administered, tissues can be collected and the transcripts sequenced to examine differences in gene expression across the genome. If differences are observed, and the individuals being tested all have the same genotype, then the agent responsible for the differences would be the treatment (i.e., a change in environment). Variation in some traits is due to differences in **environment** (in this experimental context, the treatment introduces a new environment for the genes). Alternatively, the effect of a mutation (naturally occurring or artificially introduced [e.g., via CRISPR-Cas9] on gene expression can be examined using a similar experimental design (i.e., where the groups are (1) wild type and (2) mutant). Variation in some traits is due exclusively to differences in **genes**. A combination of these two approaches can also be examined, where mutant and wild-type individuals can be exposed to a drug to see how it affects them and if there are differences in the interaction of genotype and treatment. This last approach is examining the gene-by-environment interaction (the environment in this scenario being the drug). Variation in many traits is due to variation in both genetics and environment- and the contribution of both these factors to phenotype is known as a **gene-by-environment** interaction. There are also scenarios where a trait is influenced by multiple genes (e.g., epistasis)- which are known as **gene-by-gene interactions**.

2.4 Proteomics

The collective examination of the protein products of a genome, which is known as the **proteome**.