Histones, Chromatin, and ChIP-Seq:

- Histone octamers are comprised of H2A and H2B dimers associated with tetramers formed from H3 and H4.
- Nucleosome: the protein-DNA complex formed when DNA wraps around the histone for compaction.
- Chromatin: the protein AND nucleic acids that comprise a chromosome. Includes RNA
- Variants: Histone variants involve different histone proteins incorporated into the histone octamer, which changes what regions DNA can wrap around, and changes accessibility to the DNA for transcription factors and other DNA binding proteins. Linked to development
- Histone Mods: Methylation, Acetylation, Phosphorylation, and some other more rare cases such as ubiquitination, sumoylation, and ADP ribosylation. Implicated in epigenetics
- Chromatin Immunoprecipitation: antibodies specifically bind to proteins, to isolate proteins in a sample that contains various different proteins, and involves chromatin when it targets histone proteins. Antibodies can be designed to specifically bind to histones with specific mods.
- ChIP-Seq: a chromatin immunoprecipitation procedure followed by next-generation sequencing of the bound DNA. Reads from the NGS can be aligned to a reference, with regions with many reads being associated with histones of specific modifications.

ChIP-seq data analysis:

- File Formats: typically FASTQ, many reads deposited into SRA and GEO. FASTQ from SRA can be uploaded to Galaxy directly.
- Alignment: The FASTQ files can be groomed, trimmed, and aligned to a reference genome. Alignment programs (Bowtie2/BWA) can be used in galaxy/command line.
- Peak Calling: the major part of the analysis that is specific for ChIP-seq is peak calling. ID's peaks from aligning to the reference to indicate which regions of the DNA are assoc. with the targeted protein. Input BAM, output BED/WIG. MACS2 and Sicer are two common peak callers.
- ChIP-Seq controls and determining noise from signal: need to determine difference between background noise and low/meaningful signals.
    - DNA Input control: Recovers DNA not from the target, to highlight which was associated with the target.
    - Immunoglobulin G antibody control is a mock-ChIP experiment using an antibody not binding to nuclear proteins, so the DNA should be random, which sometimes leads to not much DNA being collected.
    - Untagged Strains as an antibody control, using untagged proteins that should not be bound by the specifically designed antibody.
- Differential Peak calling: changes between tissues/cell types can be seen due to changes in DNA binding to the antibody targets in those differing conditions.