

### **What is a chromatin state?**

Epigenomic marks such as histone modifications and variants, as well as regions of open chromatin, are a great way to annotate genomes and study how these annotations hold across different cell types and diseases. However, further information can be derived if we start to study combinations of multiple marks in close spatial contexts. These groupings and patterns are termed **chromatin states** and can be extremely useful in notifying classic genomic elements such as promoters and enhancers. Recognizing these chromatin states within each cell type helps us to establish systematic annotation of various DNA elements across different cell types.

### **What is gene expression?**

Gene expression is the process where the information encoded in our DNA is translated into a set of functional observable phenotypes. A crucial problem to solve in bioinformatics is determining which DNA sequences lead to expression of which traits - for example, which DNA sequences are responsible for turning on or off the generation of a particular protein. Understanding the instructions that lead to the final expressed product will give us far more power in better understanding human biology and creating better treatments for diseases and ailments.

### **What is heritability?**

Heritability is a measure of how much of the phenotype (which can be considered a cross product of genotype and environmental factors) can be explained by the genotype - these can be physical characteristics, such as height and hair color, behavioral characteristics, such as autism or schizophrenia, or even intelligence. Computationally and statistically, heritability describes how much variation in some trait (such as height) can be attributed to genetic variation. These heritability values are both geographic and demographic specific.

### **How is Principal Component Analysis related to genetic ancestry?**

Principal Component Analysis can help reduce the dimensionality (and thereby complexity) of extremely complex high-dimensionality genotype data, and then help us to try and categorize different genes. In relation to genetic ancestry, PCA has shown to be capable of helping tackle the problem of geographic genetic ancestry. That is, cluster genotype data from the same country into distinct groups. However, there have been more accurate techniques that have been developed since, such as SPatial Ancestry Analysis (SPA), but that still leverage PCA.