

Bioinformatics: Life Sciences on Your Computer

Single Nucleotide Polymorphism (SNP)

This is supplemental reading for the video lecture on Single Nucleotide Polymorphisms.

SNP

A polymorphism is a DNA variation in which each possible sequence is present in at least 1% of the population. In other words, perhaps not all humans at a particular location have exactly the same sequence. The majority of these differences differ by just one nucleotide, and are thus named **single nucleotide polymorphisms (SNPs)**.

Discussion topic: How do SNPs develop? In what time frame?

Functional consequences of SNPs

Coding area, changing amino acid sequence: physical variation, cause of some genetic disorders (e.g. CF)

Coding area, not affecting protein sequence: may affect splicing

Promoter or regulatory region: may affect gene expression

Other regions: no known impact, useful as a chromosomal marker

Medical implications

The obvious medical consequence of an SNP is when a single nucleotide change causes a disease. That can include a change to a premature stop codon. In some cases, an SNP can be associated with medical risks. An SNP in the LDL receptor reduces its efficiency and is associated with a risk high cholesterol in the bloodstream. An SNP can also be associated with differential responses to drugs. For instance, one SNP in the CYP2D8 gene disrupts the breakdown of debrisoquine, which is used to treat high blood pressure.

Three categories of SNPs

Based on their locations, SNPs can be divided into two broad categories: coding and noncoding. Since coding SNPs can either change the amino acid or leave it the same, coding SNPs can be divided into two categories, making three major types of SNPs:

Synonymous SNPs: coding region, amino acid stays the same

Nonsynonymous SNPs: coding region, amino acid changes

Noncoding SNPs: noncoding region

dbSNP

NCBI hosts the largest database of SNPs with dbSNP. When accessing dbSNP at NCBI, you are accessing a primary database of submitted SNPs and a curated database of reference SNPs. Although it is called dbSNP, it does include **MNPs** (multiple nucleotide) and **short indels**. The mini-lecture will cover how the three genome browsers deal with SNPs.

SNPEffect, SNPedia

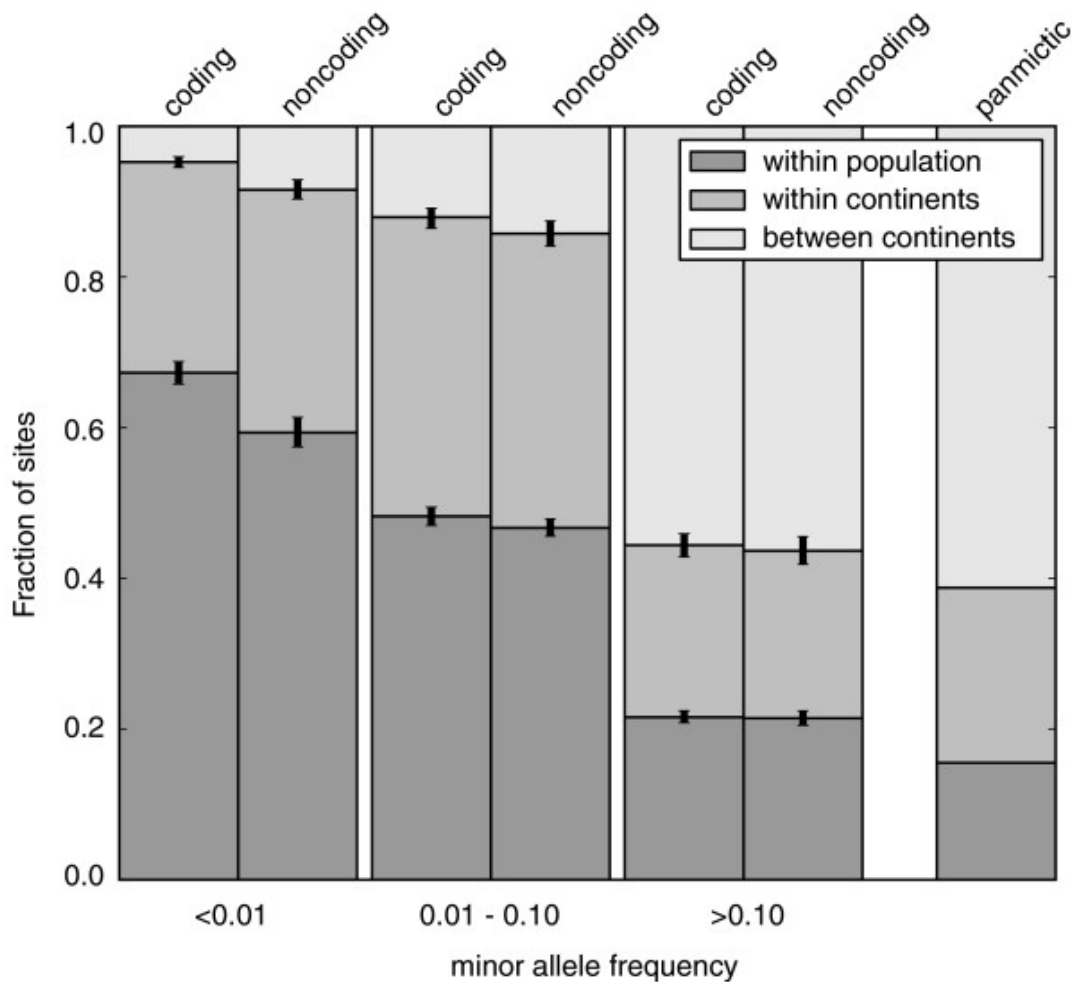
The [SNPEffect database](#) debuted in 2012. The focus of this database is on the effect of SNPs on proteins. Tools include predictors of protein stability and folding, including chaperone effect predictors.

[SNPedia](#) is a wiki that collects information from publications on known SNPs and their clinical effects.

1000 Genomes Exon Pilot Project

Next-gen sequencing has led to an explosion in new SNPs, as the sequencing process itself helps find SNPs. The 1000 Genomes Project ran an Exon Pilot Project several years ago and some results are shown in [this paper](#). An integrated map of variation [was published](#) in *Nature* in 2012. Of the 12,000+ SNPs discovered, 70% were claimed to be novel and 74% were seen at an allele frequency of less than 1%.

In addition to SNP discovery, the Exon Pilot helped reveal some insights on allele frequency and geography. The figure below relates allele frequency to allele distribution. Notably, coding SNPs tend to be less widespread when the allele frequency is low.



SNPchip

High-density single nucleotide polymorphism microarray technology is commonly called SNPchip. The idea is to use a microarray quickly identify SNPs carried in a particular genome. Version 6.0 of the [Affymetrix SNP chips](#) can quickly analyze 906,600 SNPs, almost twice as many as were in version 5.0. The Affymetrix arrays read both SNP genotype and copy number variation.

Discussion topic: What is copy number variation (CNV)?

