Get a clue: CSI and the science of forensics: using bioinformatics to link genetics and traits

Whitehead Institute Bioinformatics and Research Computing (http://jura.wi.mit.edu/bio/)

Exercise 1

Introduction

Each row includes traits and genetic markers from one individual (plant or animal or person, for example). Columns B-G show the traits of about 50 individuals. Traits are described as H and L (so can be considered as High or Low) and are color coded. Columns H-AA show rows of genotypes at different positions in the genome. Each genotype is represented by 20 different tag SNPs. For simplification we assume that there are only 2 common alleles for each SNP. Each SNP is represented as two letters. One refers to a location on the chromosome coming from the mother and one refers to the location on the chromosome coming from the father. The order of the alleles is not significant, so a heterozygous SNP is always represented as "AB". The SNP notation includes "A", the allele with a sequence that matches the reference genome, or "B", the allele with a sequence that differs from the reference genome. The SNPs have been color coded too. **Your goal is to find the locus (SNP; region of the genome) that is associated with each trait.** In more detail, when the trait is H, is the SNP always AA or is it always BB? How about when the trait is L? In this simulated data, alleles AB may encode the same trait as alleles AA or as alleles BB. The color will help you find the relationships.

To do: Sheet "Demo_visual"

One trait at a time, sort the table by the trait column and look across the SNP columns. Which SNP column follows a similar pattern to the trait pattern? Record this genetic-trait connection. You will find it because the colors on the SNP column will be ordered too. If you don't see it right away try sorting the trait first A-Z and then Z-A.

Trait (number)	SNP most associated with trait
1	
2	
3	
4	
5	
6	