4.5.2: GWAS

Learning Objectives

- Explain how molecular markers and linkage can be used to identify genes that contribute to complex traits through GWAS.
- Interpret a Manhattan plot, understanding that GWAS identify chromosomal regions that are linked to phenotypes.
- Explain how scientists use this information from GWAS to perform experiments on genes in candidate loci to determine if those genes have a role in the phenotype.

What are genome wide association studies?

We saw that linkage could be used to map genes. However, this approach is not always appropriate for tackling important questions in genetics. Consider these factors:

- In some cases, especially human genetics, directed crosses (test crosses) are not practical or ethical!
- For complex traits, the phenotype may vary along a spectrum.
- For complex traits, alleles of genes on multiple chromosomes likely contribute to the phenotype.

Instead, **genome wide association studies (GWAS)** compare single nucleotide polymorphisms across the genome and use them as markers to compare many individuals with and without the phenotype to ask which **SNPs** (loci) are likely to be inherited with the phenotype.

Important terms for understanding GWAS

- SNP: single nucleotide polymorphisms are base pairs in particular positions (loci) that vary among individuals
• Linkage: loci are sufficiently close to each other along a chromosome such that homologous recombination occurs between them less often than predicted by independent assortment
• Haplotype: a set of SNPs that can be inherited together on the same chromosome

How to genotype SNPs on a genomic scale

Depending on the species being studied, there may be several options for technologies to genotype the samples. An example is the Illumina platform, in which DNA is attached to slides and probes adjacent to each SNP are provided. DNA polymerase can then extend the probe by adding a base complementary to the next base in the sample. The free nucleotides are tagged and visualized by fluorescence.

Video: Illumina Advances Genomic Research with the Infinium Assay. (Illumina via https://www.youtube.com/watch?v=lVG04dAAyvY)

P values and Manhattan Plots

The genotype data is analyzed to identify SNPs that have a statistically significant association with the trait of interest. Without covering the details of the calculations, in statistical analysis, a P value represents the likelihood that two events are associated by random chance. A low P value indicates a low likelihood that two things are associated by chance and therefore, the two things are correlated or connected one another. The datasets in GWAS are very large, for example, a study of human ear morphology genotyped approximately 670,000 SNPs in 4,919 individuals (Adhikari et al, 2015). For GWAS studies, a P value threshold may vary depending on the analysis, but it is often less than 0.00001 (1x10^-5) or
even less than 0.00000005 (5x10^{-8}). The differences between such small P values can be difficult to visualize graphically, so scientists convert the P values to negative logarithms (−log_{10} (P value)).

negative logs and P values

Remember that a log is an exponent, for example, log_{10} (100) = 2 because 10^2 = 100.

A significant P value for GWAS is P < 5x10^{-8} and the log_{10} (0.00000005) = -7.3, so the -log = 7.3, so any SNPs that are significantly associated should have a −log_{10} (P value) > 7.3.

The −log_{10} (P ) are plotted on a graph called a Manhattan plot because the peaks at various positions across the genome resemble buildings such as in a city skyline. The peaks identify SNPs that are associated with the trait. The significant SNPs are not necessarily the cause of the disease or phenotype, but due to linkage may be inherited with the causal mutation.

Research example: A genome-wide association study identifies multiple loci for variation in human ear morphology

Craniofacial features are examples of complex traits that are likely controlled by many genes. In a GWAS in 2015, researchers examined 10 traits associated with human ear shape: ear protrusion, lobe size, lobe attachment, tragus size, antitragus size, helix rolling, folding of antihelix, crus helix expression, superior crus of antihelix expression and Darwin’s tubercle (Adhikari et al, 2015).

This study identified SNPs on chromosomes 1 and 2 as the strongest candidates loci with potential to influence human ear morphology. A significant SNP on chromosome 2 is associated with a missense mutation in the EDAR protein, which has been implicated in human ectodermal derivatives including teeth and sweat glands. Additional studies performed in mice suggest that this protein has an additional role in ear formation. A significant SNP on chromosome 1 has been identified as a binding site for a transcriptional regulatory protein for a nearby gene, TBX15, which functions in cartilage development, suggesting this region is part of the regulatory environment that influences ear shape.
COVID GWAS research


Consider the following questions about the paper:

- How many patients were involved and what information about patients was collected?
- Were there any ethical considerations about the study? If so, explain why.
- How many SNPs were genotyped and what was the method (see video above) and how many SNPs (or variants) showed statistical significance?
- How do the scientists move from an associated SNP (or other variant) and potential genes?
- What should the next experiments be to examine the role of the candidate gene(s)?

References
