Week 5: Post-GWAS Knowledge Gap Questions

BIFX 504, Advanced Molecular Biology for Bioinformatics

Name: James Jedediah Smith

1. **According to the paper, what are the two types of valuable knowledge that researchers hoped to uncover using GWAS?**

First, they wanted to uncover the molecular mechanics that cause a disease, as well as relevant genes and variants. Second, they also hoped to use genetic associations as diagnostic markers for making informed health decisions.

1. **Which type of knowledge from question 1 have studies been most successful in producing?**

The second type, identifying risk of disease with genetic associations.

1. **What is a polygenic risk score and how is it used?**

Polygenic risk score is a measure of disease risk. It is used to inform treatment options and can help stratify disease subtypes.

1. **Why is that not helpful for Parkinson’s Disease.**

We do not know a lot about the disease and there is really only one treatment option.

1. **What example does the paper give of disease insights that were made because of intronic variants in a particular gene?**

Obesity. Intronic variants help alter the regulation of FTO1 and other nearby genes in the hypothalamus, which subsequently effects feeding behavior.

1. **Where do 90% of most PD variants lie and why does that make them difficult to interpret?**

They lie in non-coding DNA. This makes them harder to interpret because there are no protein products to be analyzed.

1. **What are risk variants in noncoding DNA likely to affect?**

They are most likely to effect the regulation of gene expression via regulatory elements such as enhancers, promoters, coding RNA, and more.

1. **What is an enhancer and why is it difficult to determine how variants in an enhancer might affect cellular function?**

Enhancers are short bits of DNA that bind activator or repressor proteins and regulate transcription of target genes through proximity to promoters. It is difficult to determine how they might effect cellular function because their target genes are not immediately obvious.

1. **Why are closely spaced SNPs that are in high LD difficult to interpret?**

It makes identifying their functional variants challenging. Any of the variants could be causal, maybe even more then one. And since risk loci are usually reported via the most significant index SNPs that tags a locus, the same point could be used to represent many other associated variants.

1. **How many SNPs of interest are in the 90 independent PD risk signals that have been discovered using GWAS?**

There are more than 6500 SNPs of interest for PD.

1. **What other types of data are typically merged with the SNP data in order to try to identify the ones that are most likely to be functional?**

The types of data include roadmap epigenomic data, GTEx eQTL data, gene expression data, and representative chromosomal conformation experiment data.

1. **Why are the “informative pathway enrichment” results based on gene PD risk gene sets that contain “bona fide” PD causal genes unreliable?**

The individual GWAS used for these data sets have very high levels of uncertainty, with many false positives and false negatives.

1. **According to the paper, what three things must be verified experimentally before a risk locus can really be considered valid as a driver gene for PD?**

Eacjh risk locus must be shown to function in a specific cell type, at a particular time with respect to the disease, and via a defied allele-dependent mechanism.

1. **What percentage of PD GWAS risk variants reside outside of exons?**

More than 95% of them.

1. **Why is it difficult to determine which gene is regulated by an enhancer?**

More than 40% of enhancers skip over the nearest gene, making it difficult to pin down which gene they are effecting. Some can even effect genes on other chromosomes.

1. **What makes eQTL studies have a lot of false positives and negatives?**

Bias for accessible tissues, cell type heterogeneity in samples, multiple hypotheses penalties, and population bias among donors.

1. **What two approaches can be taken in order to determine what tissue or condition risk loci might function?**

First, a locus-centric approach with an unbiased query of tissue or cell specific genomic activity. Second, guess a model system based on disease knowledge.

1. **For PD, what two types of risk may be identified via GWAS?**

The first type represents the likelihood for an increased toxicity load or insult level falling upon the nigrostriatal systems of the at-risk population. The second type represents a decreased capacity or robustness inherent to those nigrostriatal systems.

1. **Why do the authors hypothesize that at least some germline genetic risk loci are involved during developmental phases of dopaminergic differentiation of the substantia nigra?**

Dynamic regulation of gene expression occurs during cellular differentiation. So it is possible that genetic variation at the PD risk SNPs could impose functional consequences during neuron development. The germline is a key component of development, so variations there could result in neurological changes, such as the large variability of dopaminergicneuron counts that has recently been hypothesized as a potential PD risk factor.

1. **According to the authors, we cannot understand the genetic risk specifically or the genetic architecture generally unless we can answer \_\_\_\_, \_\_\_\_\_, and \_\_\_\_ a variant functionally manifests.**

Where, when, and how.