Biology notes

Stephen D. C.

February 15, 2019

Abstract

Some comments on relevant biology.

1 Expression quantitative trait loci

The last two decades have seen a huge body of research focused on genome variability due to its relevance in the risk of disease experienced by individuals. Fundamental to this study is understanding the effect different genome variants have; i.e. understanding how this change in genome translates to a different phenotype. This means we must investigate the change a variant effects within the cell. Ideally this information allows biological insight into the aetiology and nature of disease or of the phenotype. Genome-wide association studies (GWAS) [5] have shown that the majority of these variants are located within the non-coding regions of the genome implying that they are involved in gene regulation. These sites that explain some of the phenotypic variance are referred to as expression quantitative trait loci (eQTL).

eQTLs have transformated the study of genetics. They provide a comprehensuible, accessible and most importantly interpretable loecular link between genetic variation and phenotype. Standard eQTL analysis involves a direction association test between markers of genetic variation, typically using data collected from tens to hundreds of people.

This analysis can be proximal or distal.

- Proximal: immedietely responsible for causing some observed result;
- Distal: (also *ultimate*) higher-level than proximal. The true cause for an event or result.

Consider the example of a ship sinking. This could have a *proximate* cause such as the ship being holed beneath the waterline leading to water entering the ship; this resulted in the ship becoming denser than water and it sank. However, the *distal* cause could be the ship hit a rock tearing open the hull leading to the sinking.

In terms of eQTLs, we designate proximal effects as *cis-eQTL* and distal causes as *trans-eQTL*. We normally consider an eQTL to be cis-regulating if it is within 1MB of the gene transcription start site (TSS) and trans-regulating if it is more than 5MB upstream or downstream of the TSS or if found on a different chromosome.

trans-eQTL are hard to find. They have weaker effects than cis-eQTL and thus require greater power in the experiment [3]. For some context, Burgess [1] claims that 449 donors provide low power in terms of finding trans-eQTL. As the power of experiments increases more trans-eQTL are observed and cis-eQTL are shown to be generally tissue agnostic [2]. Preivous results suggested cis-eQTL would be have tissue-specific effects, but the increase in experimental power revealed that this is not the case [4]. The current power present in many genetic experiments is enough to observe trans-eQTLs and indicates these have tissue-specific properties [4] [2]. It is

possible that this result might be shown to be an artifact of insufficient power. However, for now we assume it is true and that trans-eQTL are are more likely to display tissue-specifc behaviour than cis-eQTL.

Nica and Dermitzakis [6] reccommend investigating groups of cis-eQTL affecting a gene network that when perturbed results in a disease state. They claim this is far higher powered than the classical approach. This claim is supported by the findings of Võsa et al. [7] who found that associations between *polygenic risk scores* and gene expression (this assoication is referred to as "expression quantitative trait score" (eQTS) in [7]) contained the most biological information about disease in a comparison of cis-eQTL, trans-EQTL and eQTS.

1.1 cis-eQTL

cis-eQTL play a direct role in gene regulation. Common examples are *enhancers* and *promoters*. In short:

- Enhancer: a piece of DNA that enhances (accelerates the rate of) gene transcription;
- Promoter: a piece of DNA which acts to start gene transcription.

1.2 Gene regulation

Honestly, the Gene regulation article from Wikipedia is a really nice introduction.

2 Data

The data is from the CEDAR cohort (CEDAR stands for Correlated Expression and Disease Association Research). This is data collected from 323 healthy individuals of European descent visting the University of Liège with samples across 9 tissue types. The cohort consists of 182 women and 141 men with an average age of 56 years (the total range is 19-86). To ensure the integrity of the data all of the individuals are not suffering from any autoimmune or inflammatory disease and were not taking corticosteroids or non-steroid anti-inflammatory drugs (with the exception of aspirin). Samples for six circulating immune cells types (CD4+ T lymphocytes, CD8+ T lymphocytes, CD14+ monocytes, CD15+ granulocytes, CD19+ B lymphocytes and platelets) and from intestinal biopsies from three distinct locations (the illeum, rectum and some other one) are present for each individual. We initially explored the gene expression data corrected for sex, age, smoking and batch effects.

References

[1] Darren J Burgess. Principles of gene regulation across tissues. *Nature Reviews Genetics*, 18, 2017.

[2] GTEx Consortium et al. Genetic effects on gene expression across human tissues. *Nature*, 550(7675):204, 2017.

[3] Anna L Dixon, Liming Liang, Miriam F Moffatt, Wei Chen, Simon Heath, Kenny CC Wong, Jenny Taylor, Edward Burnett, Ivo Gut, Martin Farrall, et al. A genome-wide association study of global gene expression. *Nature genetics*, 39(10):1202, 2007.

[4] Elin Grundberg, Kerrin S Small, Åsa K Hedman, Alexandra C Nica, Alfonso Buil, Sarah Keildson, Jordana T Bell, Tsun-Po Yang, Eshwar Meduri, Amy Barrett, et al. Mapping cis-and trans-regulatory effects across multiple tissues in twins. *Nature genetics*, 44(10):1084, 2012.

[5] Teri A Manolio. Genomewide association studies and assessment of the risk of disease. *New England journal of medicine*, 363(2):166–176, 2010.

[6] Alexandra C Nica and Emmanouil T Dermitzakis. Expression quantitative trait loci: present and future. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1620): 20120362, 2013.

[7] Urmo Võsa, Annique Claringbould, Harm-Jan Westra, Marc Jan Bonder, Patrick Deelen, Biao Zeng, Holger Kirsten, Ashis Saha, Roman Kreuzhuber, Silva Kasela, et al. Unraveling the polygenic architecture of complex traits using blood eqtl meta-analysis. *bioRxiv*, page 447367, 2018.