Biology notes

Stephen D. Coleman

April 10, 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) [10] have shown that the majority of these variants are located within the non-coding regions of the genome [11] 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 transformed the study of genetics. They provide a comprehensible, 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: immediately 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 [11].

trans-eQTL are hard to find. They have weaker effects than cis-eQTL and thus require greater power in the experiment [4]. For some context, Burgess [3] 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 [6]. Previous results suggested cis-eQTL would be have tissue-specific effects, but the increase in experimental power revealed that this is not the case [5]. The current power present in many genetic experiments is enough to observe trans-eQTLs and indicates these have tissue-specific properties [5][6]. It is

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

Nica and Dermitzakis [11] recommend 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. [15] who found that associations between *polygenic risk scores* and gene expression (this association is referred to as "expression quantitative trait score" (eQTS) in [15]) 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) [13]. 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.

3 Tissue specificity

Cell-type specific gene pathways are pivotal in differentiating tissue function, implicated in hereditary organ failure, and mediate acquired chronic disease [7]. More and more evidence is being accrued to highlight the cell-type specific level of gene expression [5][12][9]. As many gene set databases are summaries of multiple experiments across many different tissues, we attempt to create tissue specific gene sets. This seems particularly pertinent in the application of immunology where many diseases are tissue-specific and have strong associations to genetic pre-disposition [14][8][1][2]. Previous attempts to achieve this have used the Genotype Tissue Expression (GTEx) database [6], but this is a database that has a heavy focus on brain tissues and is also exclusively from tissues of dead people. We suspect that the data derived from these cells may not contain the same information as that collected from living, active cells.

References

[1] Thomas M. Aune, Joel S. Parker, Kevin Maas, Zheng Liu, Nancy J. Olsen, and Jason H. Moore. Co-localization of differentially expressed genes and shared susceptibility loci in human autoimmunity. *Genetic Epidemiology*, 27(2):162–172, September 2004. ISSN 0741-0395, 1098-2272. doi: 10.1002/gepi.20013. URL http://doi.wiley.com/10.1002/gepi.20013.

[2] David Botstein and Neil Risch. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nature Genetics*, 33(S3):228–237, March 2003. ISSN 1061-4036, 1546-1718. doi: 10.1038/ng1090. URL http://www.nature.com/articles/ng1090z.

[3] Darren J. Burgess. Gene expression: Principles of gene regulation across tissues. *Nature Reviews Genetics*, 18(12):701–701, November 2017. ISSN 1471-0056, 1471-0064. doi: 10.1038/nrg.2017.94. URL http://www.nature.com/doifinder/10.1038/nrg.2017.94.

[4] Anna L Dixon, Liming Liang, Miriam F Moffatt, Wei Chen, Simon Heath, Kenny C C Wong, Jenny Taylor, Edward Burnett, Ivo Gut, Martin Farrall, G Mark Lathrop, Gonçalo R Abecasis, and William O C Cookson. A genome-wide association study of global gene expression. *Nature Genetics*, 39(10):1202–1207, October 2007. ISSN 1061-4036, 1546-1718. doi: 10.1038/ng2109. URL http://www.nature.com/articles/ng2109.

[5] 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, James Nisbett, Magdalena Sekowska, Alicja Wilk, So-Youn Shin, Daniel Glass, Mary Travers, Josine L Min, Sue Ring, Karen Ho, Gudmar Thorleifsson, Augustine Kong, Unnur Thorsteindottir, Chrysanthi Ainali, Antigone S Dimas, Neelam Hassanali, Catherine Ingle, David Knowles, Maria Krestyaninova, Christopher E Lowe, Paola Di Meglio, Stephen B Montgomery, Leopold Parts, Simon Potter, Gabriela Surdulescu, Loukia Tsaprouni, Sophia Tsoka, Veronique Bataille, Richard Durbin, Frank O Nestle, Stephen O'Rahilly, Nicole Soranzo, Cecilia M Lindgren, Krina T Zondervan, Kourosh R Ahmadi, Eric E Schadt, Kari Stefansson, George Davey Smith, Mark I McCarthy, Panos Deloukas, Emmanouil T Dermitzakis, and Tim D Spector. Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nature Genetics*, 44(10):1084–1089, October 2012. ISSN 1061-4036, 1546-1718. doi: 10.1038/ng.2394. URL http://www.nature.com/articles/ng.2394.

[6] GTEx Consortium. Genetic effects on gene expression across human tissues. *Nature*, 550 (7675):204–213, October 2017. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature24277. URL http://www.nature.com/articles/nature24277.

[7] Wenjun Ju, Casey S. Greene, Felix Eichinger, Viji Nair, Jeffrey B. Hodgin, Markus Bitzer, Young-suk Lee, Qian Zhu, Masami Kehata, Min Li, Song Jiang, Maria Pia Rastaldi, Clemens D. Cohen, Olga G. Troyanskaya, and Matthias Kretzler. Defining cell-type specificity at the transcriptional level in human disease. *Genome Research*, 23 (11):1862–1873, November 2013. ISSN 1088-9051. doi: 10.1101/gr.155697.113. URL http://genome.cshlp.org/lookup/doi/10.1101/gr.155697.113.

[8] K. Maas, S. Chan, J. Parker, A. Slater, J. Moore, N. Olsen, and T. M. Aune. Cutting Edge: Molecular Portrait of Human Autoimmune Disease. *The Journal of Immunology*, 169(1):5–9, July 2002. ISSN 0022-1767, 1550-6606. doi: 10.4049/jimmunol.169.1.5. URL http://www.jimmunol.org/cgi/doi/10.4049/jimmunol.169.1.5.

[9] T Maniatis, S Goodbourn, and J. Fischer. Regulation of inducible and tissue-specific gene expression. *Science*, 236(4806):1237–1245, June 1987. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.3296191. URL http://www.sciencemag.org/cgi/doi/10.1126/science.3296191.

[10] Teri A. Manolio. Genomewide Association Studies and Assessment of the Risk of Disease. *New England Journal of Medicine*, 363(2):166–176, July 2010. ISSN 0028-4793, 1533-4406. doi: 10.1056/NEJMra0905980. URL http://www.nejm.org/doi/10.1056/NEJMra0905980.

[11] A. C. Nica and E. T. Dermitzakis. Expression quantitative trait loci: present and future. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1620):20120362–20120362, May 2013. ISSN 0962-8436, 1471-2970. doi: 10.1098/rstb.2012.0362. URL http://rstb.royalsocietypublishing.org/cgi/doi/10.1098/rstb.2012.0362.

[12] Chin-Tong Ong and Victor G. Corces. Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nature Reviews Genetics*, 12(4):283–293, April 2011. ISSN 1471-0056, 1471-0064. doi: 10.1038/nrg2957. URL http://www.nature.com/articles/nrg2957.

[13] The International IBD Genetics Consortium, Yukihide Momozawa, Julia Dmitrieva, Emilie Théâtre, Valérie Deffontaine, Souad Rahmouni, Benoît Charloteaux, François Crins, Elisa Docampo, Mahmoud Elansary, Ann-Stephan Gori, Christelle Lecut, Rob Mariman, Myriam Mni, Cécile Oury, Ilya Altukhov, Dmitry Alexeev, Yuri Aulchenko, Leila Amininejad, Gerd Bouma, Frank Hoentjen, Mark Löwenberg, Bas Oldenburg, Marieke J. Pierik, Andrea E. vander Meulen-de Jong, C. Janneke van der Woude, Marijn C. Visschedijk, Mark Lathrop, Jean-Pierre Hugot, Rinse K. Weersma, Martine De Vos, Denis Franchimont, Severine Vermeire, Michiaki Kubo, Edouard Louis, and Michel Georges. IBD risk loci are enriched in multigenic regulatory modules encompassing putative causative genes. *Nature Communications*, 9(1):2427, December 2018. ISSN 2041-1723. doi: 10.1038/s41467-018-04365-8. URL http://www.nature.com/articles/s41467-018-04365-8.

[14] Timothy J Vyse and John A Todd. Genetic Analysis of Autoimmune Disease. *Cell*, 85 (3):311–318, May 1996. ISSN 00928674. doi: 10.1016/S0092-8674(00)81110-1. URL https://linkinghub.elsevier.com/retrieve/pii/S0092867400811101.

[15] Urmo Võsa, Annique Claringbould, Harm-Jan Westra, Marc Jan Bonder, Patrick Deelen, Biao Zeng, Holger Kirsten, Ashis Saha, Roman Kreuzhuber, Silva Kasela, Natalia Pervjakova, Isabel Alvaes, Marie-Julie Fave, Mawusse Agbessi, Mark Christiansen, Rick Jansen, Ilkka Seppälä, Lin Tong, Alexander Teumer, Katharina Schramm, Gibran Hemani, Joost Verlouw, Hanieh Yaghootkar, Reyhan Sönmez, Andrew A. Andrew, Viktorija Kukushkina, Anette Kalnapenkis, Sina Rüeger, Eleonora Porcu, Jaanika Kronberg-Guzman, Johannes Kettunen, Joseph Powell, Bernett Lee, Futao Zhang, Wibowo Arindrarto, Frank Beutner, BIOS Consortium, Harm Brugge, i2QTL Consortium, Julia Dmitrieva, Mahmoud Elansary, Benjamin P. Fairfax, Michel Georges, Bastiaan T Heijmans, Mika Kähönen, Yungil Kim, Julian C. Knight, Peter Kovacs, Knut Krohn, Shuang Li, Markus Loeffler, Urko M Marigorta, Hailang Mei, Yukihide Momozawa, Martina Müller-Nurasyid, Matthias Nauck, Michel Nivard, Brenda Penninx, Jonathan Pritchard, Olli Raitakari, Olaf Rotzschke, Eline P. Slagboom, Coen D.A. Stehouwer, Michael Stumvoll, Patrick Sullivan, Peter A.C. 't Hoen, Joachim Thiery, Anke Tönjes, Jenny van Dongen, Maarten van Iterson, Jan Veldink, Uwe Völker, Cisca Wijmenga, Morris Swertz, Anand Andiappan, Grant W. Montgomery, Samuli Ripatti, Markus Perola, Zoltan Kutalik, Emmanouil Dermitzakis, Sven Bergmann, Timothy Frayling, Joyce van Meurs, Holger Prokisch, Habibul Ahsan, Brandon Pierce, Terho

Lehtimäki, Dorret Boomsma, Bruce M. Psaty, Sina A. Gharib, Philip Awadalla, Lili Milani, Willem H. Ouwehand, Kate Downes, Oliver Stegle, Alexis Battle, Jian Yang, Peter M. Visscher, Markus Scholz, Gregory Gibson, Tõnu Esko, and Lude Franke. Unraveling the polygenic architecture of complex traits using blood eQTL meta-analysis. preprint, Genomics, October 2018. URL http://biorxiv.org/lookup/doi/10.1101/447367.