The Genetic Architecture of Immune-Mediated Complex Diseases

Jimmy Zhenli Liu

Darwin College
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy
September 2014

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except where specifically indicated in the Contributions section within each chapter. It does not exceed the word limit set by the Degree Committee for the Faculty of Biology, and is not substantially the same as any work that has been, or is being, submitted to any other university degree, diploma or any other qualification.

Jimmy Liu

18 September, 2014

That’s the whole problem with science. You’ve got a bunch of empiricists trying to describe things of unimaginable wonder.

— Bill Watterson, Calvin and Hobbes

Abstract

Complex disease risk is characterised by a combination of multiple genetic factors along with the environment. Since 2005, genome-wide association studies have discovered thousands of genetic variants associated with hundreds of such diseases. Following on from these types of studies, custom genotyping arrays with dense SNP content have allowed for greater refinement across risk loci, while their low cost has enabled powerful locus discovery projects and cross-phenotype comparisons in very large sample sizes. Combining risk loci with disease-relevant functional genomic data allows for insights into the biology of disease. In this dissertation, I explore locus discovery, cross-phenotype comparisons and functional data integration across four immune-mediated complex diseases – primary biliary cirrhosis, primary sclerosing cholangitis, and the two forms of inflammatory bowel disease – Crohn's disease and ulcerative colitis.

In Chapter 1, I provide a historical background of our understanding of how genetic variation contributes to phenotypic variation, and the technological and theoretical advances in the last twenty years that have lead to the large-scale high-throughput locus discovery projects of today.

In Chapter 2, I describe a locus discovery project using the Immunochip custom genotyping array for primary biliary cirrhosis. In addition to identifying three new risk loci and refining associated variants within known risk loci, I explore how integrating association results with functional genomic annotations across various cell lines from the ENCODE Project can provide insights into the cell types and genomic features most relevant to disease.

In Chapter 3, I describe a similar locus discovery project using the Immunochip for primary sclerosing cholangitis (PSC), where nine novel risk loci were identified. Over 80% of PSC patients are also diagnosed with inflammatory bowel disease, the majority of which is ulcerative colitis. I explore genetic factors

that may explain this overlap, and show that despite this high comorbidity, around half of PSC risk loci appear unique to PSC.

In Chapter 4, I describe a trans-ethnic genome-wide association meta-analysis for inflammatory bowel disease (IBD) comprising individuals of European, East Asian, Indian and Iranian ancestry genotyped on a combination of genome-wide arrays and the Immunochip. Forty new IBD loci were discovered associated with Crohn's disease, ulcerative colitis or both. I show that there exists pervasive sharing of IBD risk loci between European and non-European populations, while also noting specific loci where effect sizes differ between populations. The study demonstrates the utility of performing large-scale GWAS meta-analyses across different populations to identify novel susceptibility loci.

I then move beyond locus discovery in Chapter 5, where I describe a simple method for integrating differential gene expression datasets with disease risk loci. I applied the method to two gene expression datasets reflecting the genes that are involved in maintaining intestinal T cell homeostasis, and those triggered in the gut in response to infection. I find that in both cases, genes that are differentially expressed between these conditions are significantly overrepresented among risk loci for a range of autoimmune disorders, allowing for the identification of additional candidate genes at these loci and the generation of hypotheses about the mechanism through which they mediate disease.

Finally, in Chapter 6, I discuss the major themes of the preceding chapters on unravelling the genetic architecture of complex diseases. I then look to the types locus discovery projects that will shape the field in the coming years, and the potential for these to be ultimately translated into better treatment outcomes for patients.

Acknowledgements

First and foremost, I thank my supervisor, Carl Anderson, for giving me the opportunity to pursue this PhD. It has been an absolute privilege. This dissertation would not have been possible without your continued guidance, enthusiasm and ceaseless faith in me throughout these years. With your stubborn attention to detail and unrelenting loyalty to rigour, you stand as a role model scientist for myself and no doubt many others to come.

I also thank my secondary supervisor, Jeff Barrett, and my degree committee, Stephen Sawcer and Ines Barroso for their guidance over these years. Thank you also to Christina Hedberg-Delouka, Annabel Smith, Alex Bateman and Julian Rayner for keeping the Sanger PhD program such a well-oiled machine.

To members of the Anderson Group (both past and present) - Tejas Shah, Eva Serra, Sun-Gou Ji and Jamie Floyd - I could not have asked for nicer folks to share an office with. It's been an absolute pleasure working with you all; thank you for putting up with me.

The work presented in this dissertation would not have been possible without the efforts of collaborators both at Sanger and around the world, of which there are far too many to list here. But for their hard work, dedication and willingness to share the spoils of research, I am especially indebted to Mohammed Al Marri, Luke Jostins, Daniel Gaffney, Richard Sandford, Trine Folseraas, Tom Hemming Karlsen, Johannes Roksund Hov, Eva Ellinghaus, Andre Franke, Tim Raine, Adam Reid, Suzanne van Sommeren, Rinse Weersma and Hailiang Huang. Thank you also to legions of doctors, nurses, researchers and administrators of the UK PBC Consortium, the International PSC Genetics Consortium and the International IBD Genetics Consortium for their tireless efforts in bringing together groups around the world towards the common noble goal of advancing disease research. None of this of course would have been possible without the >100,000 donors whose DNA were used in these projects, for which I will be forever grateful.

I would also like to thank the many friends and colleagues I've gotten to know during my time at Sanger. Chris Franklin, Yang Luo, James Morris, Scott Shooter, Isabelle Cleynan, Mari Niemi and everyone in Morgan N333 and N309 - cheers for the lunchtime banter, post-lunchtime strolls, post-stroll tea breaks, mingles and thoughts. To everyone I've enjoyed playing with (and competing against) in the Genome Campus volleyball, football and cricket leagues, I tried my best and can only apologise. Thank you to the fellow PhD students and Sanger and those I've gotten to know in Cambridge for making my time here so enjoyable.

I also thank the Wellcome Trust for generously funding my time at Sanger, as well as the various funding bodies for making these projects possible.

Lastly, I want to thank my family, especially my parents, Hua and Xiaoyu, for the unconditional love, support, understanding and encouragement that they have shown me in everything that I do.

Table of Contents

Publications

From this dissertation

Liu J.Z., van Sommeran S., Huang H., Ng S.C. *et al.*, Association study discovers 38 susceptibility loci for inflammatory bowel disease and shows pervasive sharing of genetic risk across diverse populations. *Under review.*

Raine T., Liu J.Z., Anderson C.A., Parkes M. and Kaser A., Generation of primary human intestinal T cell transcriptomes reveals differential expression at genetic risk loci for immune-mediated disease. Gut. *In press.*

Liu J.Z. and Anderson C.A., Genetic studies of Crohn's disease: past, present and future. Best Practice & Research: Clinical Gastroenterology, 28:373-386, 2014.

Foth B.J., Isheng J.T., Reid A.J., Bancroft A., *et al.*, Whipworm genomes and dual-species transcriptome analysis provide molecular insights into an intimate host-parasite interaction. Nature Genetics, 46:693-700, 2014.

Liu J.Z., Hov J.R., Folseraas T., Ellinghaus E., *et al.*, Dense genotyping of immune-related disease regions identifies nine new risk loci for primary sclerosing cholangitis. Nature Genetics 45:670-675, 2013.

Liu, J.Z., Almarri, M.A., Gaffney, D.J., Mells, G.F., Jostins, L., Cordell, H.J., Ducker, S.J., Day, D.B., Heneghan, M.A., Neuberger, J.M., *et al*. Dense fine-mapping study identifies new susceptibility loci for primary biliary cirrhosis. Nature Genetics, 44:1137-1141, 2012.

Arising elsewhere

Curtis J., Luo Y., Zenner H.L., Cuche-Lourenco D., *et al.,* Susceptibility to tuberculosis is associated with the ASAP1 gene that regulates dendritic cell migration. *Under Review.*

Houldcroft C.J., Petrova V., Liu J.Z., Frampton D., *et al.,* Host genetic variants and gene expression patterns associated with Epstein-Barr virus copy number in lymphoblastoid cell lines. PLoS One 9:e108384, 2014.

Robles-Espinoza C.D., Harland M., Ramsay A.J., Aoude L.G., *et al.,* POT1 loss-of-function variants predispose to familial melanoma. Nature Genetics, 46:478-481, 2014.

Shah T.S., Liu J.Z., Floyd J.A.B., Morris J.A., *et al.,* optiCall: A robust genotyping-calling algorithm for rare, low frequency and common variants. Bioinformatics 28:1598-1603, 2012.

List of tables

List of figures