Regulatory variation and its role in disease

Alexandra Cristina Nica

Hughes Hall College

University of Cambridge

August 2010

This dissertation is submitted for the degree of Doctor of Philosophy

UNIVERSITY OF CAMBRIDGE

To my parents

Mihaela and Marian

Declaration

This thesis describes my work undertaken at the Wellcome Trust Sanger Institute under the supervision of Prof. Emmanouil Dermitzakis and Dr. Inês Barroso, in fulfilment of the requirements for the degree of Doctor of Philosophy at University of Cambridge. This dissertation is the result of my own work and includes nothing that is the outcome of work done in collaboration, except where indicated in the text. The work described here has not been submitted for a degree, diploma or any other qualification at any other university or institution. I confirm that this dissertation does not exceed the word limit specified by the Biology Degree Committee.

Alexandra Cristina Nica
Cambridge, August 2010

Abstract

The role of regulatory variation in shaping phenotypes became apparent once significant species differences could not be explained by differences at DNA sequence level. Since then, the control of gene expression emerged as an essential process at the heart of cell-type differentiation and determination of phenotypic variance across multiple populations and tissues. Concurrent with the identification of genetic variants affecting transcript levels (eQTLs) across the human genome, large-scale genome-wide association studies (GWAS) shed light into the genetics of complex traits by detecting a multitude of susceptibility loci of modest effect-size. The goal of this thesis is to explore the role of regulatory variation in explaining genetic associations with complex traits and assess how that role differs across tissues.

To address this aim, I first developed an empirical methodology called Regulatory Trait Concordance (RTC) that integrates eQTLs and GWAS results in order to reveal the subset of association signals due to proximal eQTLs (*cis* variants*)*. By simulating different genomic regions, I show that this method outperforms simple correlation metrics between single nucleotide polymorphisms (SNPs). I observe a significant enrichment of regulatory effects among currently known GWAS loci and I apply the RTC method to prioritize relevant genes for each of the tested complex traits. For this purpose, I use gene expression data measured in lymphoblastoid cell lines (LCLs) derived from HapMap 3 individuals and I detect several potential disease-causing regulatory effects, with a strong enrichment for immunity-related conditions. Furthermore, I present an extension of the method in *trans*, where interrogating the whole genome for downstream effects of the disease variant can be informative regarding its unknown primary biological effect.

Given that certain phenotypes manifest themselves only in certain tissues, I next explore the complexity of regulatory tissue-specificity in three human cell-types: LCLs, skin and fat. I discover an abundance of eQTLs in each of the three tissues derived from a sample set of well-phenotyped female twins and I make use of the unique study design (matched co-twins) to validate the discoveries. I highlight the challenges of comparing eQTLs between tissues and propose that continuous significance estimates and direct comparison of the magnitude of effect on the fold

change in expression are essential properties providing a biologically realistic view of tissue-specificity. Under this framework, I find evidence for extensive tissue-specificity: 30% of eQTLs are shared among the three tested tissues and of those, 10-20% have significant differences in the magnitude of fold change between homozygote genotypic classes across tissues.
Finally, I show that finding causal regulatory effects for complex disease associations is highly impacted by the tissue where expression is quantified and its relevance to the trait. I apply the RTC method on GenCord, a dataset where gene expression had been previously measured in LCLs, fibroblasts and primary T-cells derived from the same 75 individuals of Europeans descent. As expected, I find a large proportion of likely causal regulatory effects for GWAS signals to be tissue dependent (70% of all significant signals).

Altogether, my results support the informative value of gene expression in explaining a subset of GWAS signals and highlight the need to explore a variety of cell-types for enhancing our understanding of the biology behind these associations.

Publications

Publications arising during the course of the work described in this thesis:

Nica A.C., Parts L. et al. The architecture of regulatory gene variation across multiple human tissues: the MuTHER study. *In review.* (2010)

Nica A.C., Montgomery S.B. etal. Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. **PLoS Genetics** 6(4) (2010)

Xue Y, Zhang X, Huang N, Daly A, …, **Nica A.C.**, …, Tyler-Smith C. Population differentiation as an indicator of recent positive selection in humans: an empirical evaluation. **Genetics** 183(3): 1065-77 (2009)

Soranzo N, Rivadeneira F, Chinappen-Horsley U, Malkina I, … , **Nica A.C.**, …, Deloukas P. Meta-analysis of genome-wide scans for human adult stature identifies novel loci and associations with measures of skeletal frame size. **PLoS Genet** 5(4) (2009)

Loos RJ, Lindgren CM, Li S, Wheeler E, … , **Nica A.C**., … , Mohlke KL. Common variants near MC4R are associated with fat mass, weight and risk of obesity. **Nat Genet** 40, 768-75 (2008)

Nica A.C. & Dermitzakis E.T. Using gene expression to investigate the genetic basis of complex disorders. **Hum. Mol. Gen**. 17, R129-R134 (2008)

Stranger B.E**., Nica A.C.** et al. Population genomics of human gene expression. **Nat Genet** 39, 1217-24 (2007)

Acknowledgements

Firstly, I am very grateful to my supervisor Manolis Dermitzakis, whose guidance and support have been essential for my PhD. Thank you for the exciting discussions, the encouragement to take on challenging projects and the great help with relocating to Geneva. I have learned a lot professionally and personally from you. In particular, I would like to also thank my co-supervisor Inês Barroso for kindly proofreading my manuscripts and giving me helpful comments on the thesis. Thank you for always finding the time to see me when I needed advice.

Throughout the four years, I have been fortunate to have extraordinary colleagues who helped me with my projects and with whom I shared not only an office, but also memorable times at conferences, dinner parties or coffee breaks. For their valuable support I thank Antigone Dimas, Barbara Stranger, Catherine Ingle, Christine Bird, Claude Beazley, Daniel Jaffares, James Nisbet, Maria Gutierrez-Arcelus, Stephen Montgomery, Tuuli Lappalainen and Tsun-Po Yang.

I can hardly imagine a better PhD program than the one at the Wellcome Trust Sanger Institute and I am deeply thankful to the Wellcome Trust and the Sanger Committee of Graduate Studies for caring so well for us.

For the first time while being abroad, living in Cambridge stopped feeling temporary and I owe this unexpected feeling to my dear friends. PhD06 have been great company in the pub, on the tennis court, at concerts, formals and in every other Central or rather Eastern European home country we visited together. Great thanks to Naomi, Andrei, Dan and Mircea for the benefits of missing a boat in Yvoire, the gems on the bookshelf, the most tasty meals with a post-rock twist and caring Millroadian cycling.

My beginning in Geneva would have been even harder if not for Tamara, Ouarda and Djamel cheering me up so many times.

Carmen’s inspiring interludes since early school days make me believe now too in the astral hours of our encounter.

Finally, they say you don’t get to choose your relatives, but that feels inaccurate in my case. For as long as I can remember, my parents have given me confidence to aim high while holding the safety net with love, humour and honesty. This thesis is dedicated to them.

Table of Contents