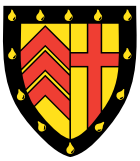
UNIVERSITY OF CAMBRIDGE

Primary sclerosing cholangitis: from genetic risk to disease biology

Elizabeth Claire Goode



Clare College

This dissertation is submitted for the degree of Doctor of Philosophy, May 2020

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or am concurrently submitting, for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or is being concurrently submitted, for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. This dissertation does not exceed the prescribed limit of 60 000 words.

Elizabeth Claire Goode
May, 2020

Abstract

Primary sclerosing cholangitis: from genetic risk to disease biology
Elizabeth Claire Goode

One in 10,000 people in the Western world lives with Primary Sclerosing Cholangitis (PSC), an immune-mediated, inflammatory disease of the bile ducts that is highly co-morbid with inflammatory bowel disease (IBD). PSC confers risk of serious disease sequelae including hepatobiliary malignancy and progression to end-stage liver failure, for which the only treatment option is liver transplantation. The absence of effective medical therapies for PSC reflects our current limited understanding of the disease's aetiology and pathogenesis.

Our DNA, laid down at conception, gives us an unrivalled opportunity to understand the underlying causal biology of disease. This is because the genetic variants associated with disease susceptibility perturb genes and biological pathways that contribute to disease causality. Twenty-two regions of the genome, outside of the HLA, have been associated with PSC risk. These loci offer the potential for huge insight into the causal biology of PSC, if only we can robustly identify the true causal variants driving these loci and the genes they perturb. However, this is complicated by several scientific challenges. Firstly, the majority of disease-associated risk loci occur within non-coding regions of the genome. Secondly, patterns of correlation between variants within a risk locus means that the true causal variant driving the signal could be any of those highly correlated with the variant with the smallest p-value.

In this thesis, I present analyses aimed at identifying the true genes and causal variants underlying each of the twenty-two PSC risk loci. Many non-coding risk variants associated with complex disease exert a quantitative affect upon gene expression i.e. are expression quantitative trait loci (eQTLs). Colocalisation assesses the evidence that a single shared causal variant is responsible for driving PSC risk and gene expression via an eQTL. In order to assign dysregulated genes to PSC risk loci, I perform colocalisation with eQTLs mapped in multiple cell-types and tissues mechanistically relevant to PSC. Because PSC is rare, eQTLs have not previously been mapped in all cell-types most relevant to this disease. In addition, I therefore map eQTLs in six peripheral blood T-cell subsets (including the rare CCR9+ gut-homing T-cells) from ∼80 patients with PSC and IBD. With colocalisation, I assign causal genes to five PSC risk loci, and assign other epigenetic regulatory features

including methylation or histone modification, to six risk loci. Statistical fine-mapping of each risk locus in both the GWAS and eQTL data enables me to resolve three PSC risk loci to a single causal variant and nine loci to 95% credible sets containing ten or fewer variants.

The results presented in this thesis identify three genes (*PRKD2, ETS2* and *UBASH3A*), causal in the pathogenesis of PSC, which are currently the target of existing or experimental therapeutic agents. Firstly, reduced expression of *PRKD2* causes excessive cell-autonomous T-follicular helper cell development and B-cell activation, and is associated with increased risk of PSC. Several studies are investigating the therapeutic effects of increasing the kinase activity of PRKD2. *ETS2* is involved in the induction of pro-inflammatory cytokine release from macrophages and IL-2 regulation in Th to Th0 transition. ETS2 inhibitors are currently the subject of early therapeutic trials. Finally, *UBASH3A* attenuates the NF-$\varkappa$B/I-KKβ pathway, an inflammatory pathway that is already targeted by proteasome inhibitors and acetylsalicylic acid, both of which could be potentially therapeutic in PSC.

PSC is a debilitating disease with serious disease sequelae, for which new therapeutic options are urgently needed. In this thesis, I elucidate multiple genes with a causal role in PSC pathogenesis, several of which are potential candidates for future therapeutic target.

Acknowledgements

As I began to write these final, but perhaps most important, few paragraphs I realise how fortunate I have been to be assisted by so many wonderful individuals along the road to completing this thesis. There a great number of people to whom I owe thanks and would like to make special mention to just some of them below.

In May of 2016, I began the work presented in this thesis under the supervision of Dr Carl Anderson. I wish to express my sincere gratitude to Carl for welcoming me to his research group and for his support and guidance in the years since. It is only with his enthusiasm, encouragement, scientific brilliance and humour that the work in this thesis has been made possible. I would also like to express my sincere appreciation to my secondary supervisors, Professor Nicole Soranzo, Dr Tim Raine and Dr Simon Rushbrook who have provided additional scientific clarity, wisdom and support. I would like to pay special regards to Simon for introducing me to this field of research back in 2012, and his mentorship ever since.

There are three talented scientists, whose assistance has been integral to the work presented in thesis and all of whom have shown me extensive kindness and patience. Firstly, I would like to thank Dr Loukas Moutsianas for teaching a naïve medic to code and for his guidance with fine-mapping and colocalisation. Secondly, I am grateful to Dr Laura Fachal who guided me through the functional annotation analyses and was always there to answer every genetics question, big or small. Thirdly, I would like to express my gratitude to Dr Nikos Panousis, whose direction was invaluable in the RNA-seq and eQTL mapping component of this thesis. To all three of you, I am indebted.

For the past four years I have been fortunate to be supported by a fellowship from the University of Cambridge Wellcome Trust Clinical PhD program. The Wellcome Sanger Institute has been an incredible source of support throughout my fellowship and a wonderful place to spend the last few years. I would also like to thank each and every member of the Anderson Team at the Wellcome Sanger Institute. This team has evolved and expanded during my time at Sanger and I have been lucky to spend time with a group of such kind, talented and good-humoured individuals. I would like to pay special thanks to Dr Rebecca McIntyre who has been a source of personal support and who also performed all of the DNA extraction from my study samples. I would also like to thank Mr

Ben Bai for his guidance with mashR *('Nobody really understands how mashR works...')*, Dr Alex Sazonovs for his help with Latex *('It's us against Latex!')*, Dr Carla Jones for her patient explanations of T-cell biology and Mr Sigurgeir Olafsson, Alex and Ben for helping solve my innumerable coding issues!

I would also like to extend my gratitude to Dr Norihito Kawasaki and Dr Alex Wittmann at the Quadram Institute, for their guidance in the preparation of samples for my T-cell eQTL study, alongside the many patients from the Norfolk and Norwich University Hospital who donated their time (and blood!) to this study. In addition I would like to thank the patient support charity *PSC Support*, who generously funded the sample acquisition for the T-cell study.

To my esteemed examiners, Dr Richard Sandford and Professor Heather Cordell, I would like to express my gratitude for their invaluable comments and for a thoroughly enjoyable viva (albeit via Zoom)!

Undoubtedly, my proudest achievement during my PhD has been the arrival of my daughter, Alexandra. Thank you for being a source of such joy, for reminding me of the most important things in life, and for sleeping well so that Mummy could write her thesis! I would also like to express my sincere gratitude to Dr Pat Tate whose unwavering positivity encouraged me to put confidence in myself, and has kept me (mostly) sane. My final and most special debt of gratitude goes to my wonderful husband, Grégory for his great love and support. Je tiens à exprimer ma profonde gratitude pour ton amour et ton soutien et pour m'avoir aidée à prendre confiance en moi. To him, I am wholeheartedly grateful.

Finally, to my parents and the many family, friends and colleagues who I have not named individually, but who have also given me their support. I hope that along the way I have shown you my appreciation, and to you all I give a heart-felt thank you.

Contents

List of Figures