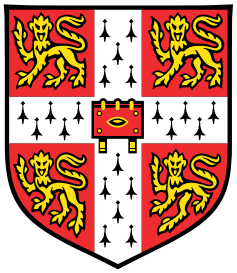
The Genetics of IBD: From Susceptibility to Drug Response and Patient Outcome



Aleksejs Sazonovs

Wellcome Sanger Institute
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

Pembroke College

September 2019

Declaration

This thesis 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, is being concurrently submitted 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 thesis 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. It does not exceed the prescribed word limit for the relevant Degree Committee.

Aleksejs Sazonovs
September 2019

The Genetics of IBD: From Susceptibility to Drug Response and Patient Outcome

Aleksejs Sazonovs

Abstract

Inflammatory bowel disease (IBD) is a group of immune-mediated autoinflammatory disorders, primarily manifesting in the gastrointestinal tract. Affecting millions of people around the world, IBD has a severe impact on patients' quality of life. Several pharmacologic treatments have been available since the 1950s. However, the majority of patients either do not respond to a given therapy or lose response to a previously effective treatment and thus require therapeutic escalation.

In the first research chapter of my thesis, I describe the results of the Personalised Anti-TNF Therapy in Crohn's disease study. Immunogenicity to anti-TNF therapy is a major cause of loss of response, hypersensitivity reactions, and discontinuation of treatment in patients. Currently, immunogenicity cannot be predicted prior to treatment. My analysis has identified a strong dominant association in the HLA region on chromosome 6 (HLA-DQA1*05, P=5.9x10^{-13}; HR=1.90; 95% CI, 1.60 to 2.25). Around 40% of individuals of European ancestry carry HLA-DQA1*05, and the data suggest that around 95% of these would develop immunogenicity within the first year of infliximab monotherapy treatment (a common anti-TNF treatment regime).

In the second research chapter of my thesis, I describe a genome-wide association study of thiopurine-induced liver damage (TILI). Ultimately, the study was underpowered to detect

any associations of moderate effect size and did not detect any associations of high effect size amongst the common genetic variants. Interestingly, I was not able to replicate the association in *PTPN22*, which was reported to be a risk factor for drug-induced liver damage by Cirulli et al. [39] – suggesting that its effect might be heterogeneous depending on the therapy.

Finally, the third research chapter describes the initial analysis of the IBD 15x dataset – a whole-genome sequencing association study of around 7,000 IBD patients paired with 12,000 matching controls. I provide an overview of the sample quality control procedures and describe some of the novel challenges that sequencing studies bring in comparison to standard GWAS (e.g., sample cross-contamination due to index mismatching). Finally, I also provide the results of the initial meta-analysis of the exome-sequencing dataset produced by the Broad Institute. The results demonstrate that rare coding genetic variants play a role in IBD pathogenesis.

I would like to dedicate this thesis to my parents – Gaļina Sazonova and Vadims Sazonovs

Acknowledgements

First of all, I would like to thank my supervisor Carl Anderson who has guided this work. I could not have wished for a more compassionate and brilliant mentor. Jeff Barrett, with whom I started my PhD, has shown me how passion for a topic can coincide with relentless rigour. The union of teams 143 and 152 – you have made me feel welcome here and I have learnt so much from each of you. The computational work (i.e. all) was enabled by the Human Genome Informatics team – Pavlos Antoniou, Chris Harrison, Colin Nolan, Alan Daly, Vivek Iyer, and Josh Randall. I've broken so many things and yet you seem to tolerate me. Nicole and other members of the Soranzo team, especially Klaudia Walter and Kousik Kundu for stimulating discussions and lively debates about 15x. Paris Litterick, Eloise Stapleton, Sally Bygraves, without whom these projects would have ground to a halt. The PANTS and TILI projects would not have been possible without the Exeter team – Nick Kennedy, Tariq Ahmad, Gareth Walker, Claire Bewshea and others. The last three years would not have been so joyful without my colleagues, who later became friends, then stopped being colleagues, but remained friends: Dan Rice, Loukas Moutsianas, Liu He, Mari Niemi, Fernando Riveros Mckay Aguilera, Scott Shooter, Tejas Shah, Arthur Gilly, and Sophie Hackinger. A heartfelt thank you to Emma Molloy for putting up with me during the final stretch of the PhD. Finally, I would like to express my gratitude to my family who have selflessly supported me on this journey. I owe so much to you.

Table of contents

List of figures

List of tables

Nomenclature

Acronyms / Abbreviations

ADA anti-drug antibody

ADR adverse drug reaction

EWAS exome-wide association study

GWAS genome-wide association study

HLA human leukocyte antigen

CD Crohn's disease

IBD inflammatory bowel disease

IBDU unclassified IBD

UC ulcerative colitis

LD linkage disequilibrium

MAF minor allele frequency

MHC major histocompatibility complex

TNF tumour necrosis factor

VEO very early onset

WES whole-exome sequencing

WGS whole-genome sequencing