Beyond genome wide association studies in celiac disease by exploring the non-coding genome

de Almeida, Rodrigo

Beyond genome wide association studies in celiac disease by exploring the non-coding genome

Rodrigo Coutinho de Almeida

Rodrigo Coutinho de Almeida

Beyond genome wide association studies in celiac disease by exploring the non-coding genome

Thesis, University of Groningen, with summary in English, Dutch and Portuguese.

The research presented in this thesis was mainly performed at the Department of Genetics, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands, and was financially supported by FP6- 2005-FOOD-4B-36383–PREVENTCD, ERC Grant Agreement no. 2012-322698 and CAPES Foundation, an agency within the Ministry of Education of Brazil.

Printing of this thesis was financially supported by: Rijksuniversiteit Groningen, University Medical Center Groningen, Groningen University Institute for Drug Exploration (GUIDE).

Cover design and layout by Claudia Marcela Gonzaleza Arevalo (e-mail: argo1983@gmail.com).

Printed by NetzoDruk.

ISBN: 978-90-367-7718-6

university of groningen

Beyond genome wide association studies in celiac disease by exploring the non-coding genome

Proefschrift

to obtain the degree of PhD at the
University of Groningen
on the authority of the
Rector Magnificus Prof. E. Sterken
and in accordance with
the decision by the College of Deans..

This thesis will be defended in public on

Wednesday 25 March 2015 at 14.30 hours

by

Rodrigo Coutinho de Almeida

born on 9 November 1980
in Bom Jesus Da Lapa, Brazil

Promotor
Prof. C. Wijmenga

Co-promotor
Dr S. Withoff

Assessment committee
Prof. F. Koning
Prof. A.M.H. Boots
Prof. J.H.M. van den Berg

Paranimfs
Isis Ricaño Ponce
Olga Maria Alvarenga Fernandes Sin

Table of contents

Preface and outline of the thesis

Part I 9

Chapter 1 Shared genetics in coeliac disease and other immunemediated diseases 15
J Intern Med. 2011; 269(6): 591-603

Chapter 2 Systematic Annotation of Celiac Disease Loci Refines Pathological Pathways and Suggests a Genetic Explanation for Increased Interferon Gamma levels 37
Hum Mol Genet. 2014; pii: ddu453

Chapter 3 Fine mapping of the celiac disease-associated LPP locus reveals a potential functional variant 61
Hum Mol Genet. 2014; 23(9): 2481-9

Chapter 4 A Celiac Disease-associated SNP in the LPP locus affects expression of the long non-coding RNA LPP-AS1 79
Manuscript in preparation

Part II

Chapter 5	Genetic variation in the non-coding genome: involvement of micro-RNAs and long non-coding RNAs in disease *Biochim Biophys Acta. 2014; pii: S0925-4439(14)00071-4*	109
Chapter 6	Circulating miRNAs as potential biomarkers of early celiac disease *Manuscript in preparation*	147
Chapter 7	MicroRNA profiles by deep-sequencing in plasma and small intestinal biopsies of celiac disease patients *Manuscript in preparation*	159
Chapter 8	Discussion and future perspectives	175
Appendices	Summary	198
	Samenvatting	203
	Resumo	209
	Acknowledgements	215
	Curriculum Vitae	221

PART I

Preface and outline of the thesis

Preface

Over the last seven years, genome-wide association studies (GWAS) have identified hundreds of thousands of trait-associated genetic variants, providing novel insights into the mechanism of many diseases, including cancer and autoimmune diseases (1). Celiac disease (CeD) is an immune-mediated disease triggered by dietary gluten in genetically susceptible individuals. CeD is considered to be one of the most common food intolerances in the world, affecting ~1% of the population (2). The most dominant genetic component identified in CeD is the human leukocyte antigen (HLA) locus, and more specifically HLA-DQ2 and HLA-DQ8. GWAS and Immunochip studies have identified 40 non-HLA loci as associated with CeD and, together with HLA, these loci explain ~55% of the heritability of CeD (3, 4).

Because of linkage disequilibrium (LD) the identification of the true causal genetic variant or variants is challenging. LD can be defined as a non-random association of alleles of different single nucleotide polymorphisms (SNPs) near each other on the same chromosome, each of which could be the true causal variant. Recently, post-GWAS studies are performing custom-made genotype arrays to fine-map and replicate previously associated regions. The Immunochip is a great example of such an approach. This custom-made array covers 186 loci associated with autoimmune and inflammatory diseases; these loci cover 12 immune-related diseases (celiac disease, Crohn's disease, ulcerative colitis, autoimmune thyroid disease, ankylosing spondylitis, multiple sclerosis, primary biliary cirrhosis, psoriasis, rheumatoid arthritis, juvenile idiopathic arthritis, systemic lupus erythematosus and type 1 diabetes) (5). In particular, Trynka et al., used the Immunochip and successfully identified 39 non-HLA loci with 57 independent signals associated to celiac disease (3).

The majority of GWAS SNPs are located in non-coding regions, making it more difficult to hypothesize about their function. An explosion in the availability of new genomics datasets, such as from the ENCODE project (6), is helping to clarify the functionality of some disease-associated variants. It has recently been shown that ~90% of CeD-associated SNPs are located in non-coding regions (7) and the majority of these map to enhancers and promoters (8).

Despite the significant advances in understanding the molecular basis of CeD, the diagnosis of this disease remains challenging. It has been estimated that only 1 out of 8 patients is correctly diagnosed with CeD. The diagnosis is based on clinical symptoms, detection of CeD-specific antibodies in the blood, and having proof of villous atrophy from intestinal biopsies (9). Recently, it has been shown that circulating

microRNAs (miRNAs) in blood can provide biomarkers for disease and even disease-stage (10). If CeD-specific circulating miRNAs could be identified, these would provide welcome non-invasive biomarkers for easier and more accurate diagnosis of CeD.

Outline of this thesis

This thesis had two principal aims: (1) to perform fine-mapping of CeD loci identified by the Immunochip study; and (2) to investigate whether miRNA profiles in CeD patients can provide biomarker candidates. We have divided this thesis into two parts that both focus on the non-coding part of the genome. In the first part, we describe how we can use Immunochip results and publicly available data to fine map CeD-associated loci and to prioritize SNPs and genes associated with CeD. In the second part, we explore miRNA profiles in CeD patients and provide different miRNA candidates that may prove to be useful biomarkers in the future.

Part I

Chapter 1 gives an overview of the genetics factors identified by earlier studies and by GWAS that are involved in CeD and of the CeD-associated genes and pathways that are shared with other immune-mediated diseases.

In Chapter 2 we describe an integrative approach to annotate and prioritize functional SNPs, genes and pathways affected in CeD, using publicly available data. We found an enrichment of CeD genes in the Th1, Th2, and Th17 pathways, and also predicted a role for four associated genes in the intestinal barrier function. Furthermore, in this chapter we describe a transcriptional connection between Interferon- with CeD susceptibility genes, which sheds light on why there are no CeD-associated SNPs in the IFNG locus, even though IFN- is known to be dysregulated in CeD.

In Chapter 3 we describe a fine-mapping approach in a strongly associated CeD locus (LPP). For this purpose we applied haplotype analysis and integration of several types of data, such as imputation, re-sequencing and functional genomics data that were available to refine this locus. We successfully fine mapped the associated region down from a 70 kb region to a 2.8 kb region, and pinpointed a possible functional SNP in this small region.

In Chapter 4 we identified the first long non-coding RNA (lncRNA) that is regulated by a CeD-associated SNP. To explore possible SNPs affecting expression of nearby genes in the LPP locus, we performed expression quantitative locus (eQTL) mapping using RNA-seq data. This approach allowed us, for the first time, to determine an eQTL effect on a lncRNA (LPP-AS1) in CeD. Subsequently, we investigated the cell type specificity of this candidate and performed pathway enrichment analysis using co-expression data.

This analysis suggested that LPP-AS1 is involved in ubiquitination, which is known to be involved in CeD.

Part II

In Chapter 5 we describe how SNPs can affect non-coding RNAs such as miRNAs and lncRNAs. Moreover, we provide an overview of bioinformatics tools, high-throughput techniques and the available databases that help in understanding the functions of non-coding RNAs.

In Chapter 6 we describe the profiling of miRNAs circulating in the serum of CeD patients enrolled in the PreventCD project, a prospective study in which samples were collected at different time points after birth from individuals in families likely to have a genetic pattern to their CeD (11). By using next generation sequencing (NGS), we were able to determine a panel of 45 miRNAs, of which six displayed a suggestive pattern over different time points. In addition, these six candidate miRNAs were detectable before the reported time of CeD diagnosis. Our results indicated that these miRNAs are potential biomarker candidates for CeD.

Chapter 7 shows a profile of miRNAs in plasma and small intestine samples from CeD patients. We also used NGS to profile miRNAs. We provided a panel of 49 miRNAs in serum and 109 miRNAs in small intestine of CeD patients. Moreover, by comparing patients that were following a gluten-free diet with patients at diagnosis, we were able to find 11 miRNAs that can potentially be used to monitor gluten ingestion in CeD patients. Three out of the 109 miRNAs differentially expressed in the small intestine had already been validated in an independent study in CeD (12). Thus, we have pinpointed miRNAs differentially expressed in the small intestine showing that it is worthwhile to follow-up and that are likely to be involved in the pathogenesis of CeD.

Chapter 8 provides a discussion of how these fine-mapping and integrative approaches has helped to shed light on CeD. In addition, we discuss the most recent status of the microRNAs that might be involved in CeD. We discuss the agreements and contradictory indications of the miRNAs profiles and show how they might be used for future clinical applications. In the context of the rapid developments in the genomics field, we discuss the challenges arising in this field, which sees the data growing exponentially.

References

1. Hindorff, L., Sethupathy, P., Junkins, H., Ramos, E.M., Mehta, J.P., Collins, F.S. and Manolio, T. (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc. Natl. Acad. Sci. U. S. A., 106, 9362–7.

2. Guandalini, S. and Assiri, A. (2014) Celiac disease: a review. JAMA Pediatr., 168, 272–8.

3. Trynka, G., Hunt, K.A., Bockett, N.A., Romanos, J., Mistry, V., Szperl, A., Bakker, S.F., Bardella, M.T., Bhaw-rosun, L., Castillejo, G., et al. (2011) Articles Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. Nat. Genet., 43, 1193–1201.

4. Hunt, K., Mistry, V., Bockett, N. a, Ahmad, T., Ban, M., Barker, J.N., Barrett, J.C., Blackburn, H., Brand, O., Burren, O., et al. (2013) Negligible impact of rare autoimmune-locus coding-region variants on missing heritability. Nature, 498, 232–5.

5. Cortes, A. and Brown, M. (2011) Promise and pitfalls of the Immunochip. Arthritis Res. Ther., 13, 101.

6. Bernstein, B.E., Birney, E., Dunham, I., Green, E.D., Gunter, C. and Snyder, M. (2012) An integrated encyclopedia of DNA elements in the human genome. Nature, 489, 57–74.

7. Kumar, V., Wijmenga, C. and Withoff, S. (2012) From genome-wide association studies to disease mechanisms: celiac disease as a model for autoimmune diseases. Semin. Immunopathol., 34, 567–80.

8. Maurano, D., Humbert, M.T., Rynes, R., Thurman, E., Haugen, R.E., Wang, E., Reynolds, H., Sandstrom, A.P., Qu, R., et al. (2012) Systematic Localization of Common Variation. 337, 1190–1195.

9. Lundin, K.E. and Sollid, L.M. (2014) Advances in coeliac disease. Curr. Opin. Gastroenterol., 30, 154–62.

10. Zahm, A.M., Thayu, M., Hand, N.J., Horner, A., Leonard, M.B. and Friedman, J.R. (2011) Circulating microRNA is a biomarker of pediatric Crohn disease. J. Pediatr. Gastroenterol. Nutr., 53, 26–33.

11. Hogen Esch, C.E., Rosn, A., Auricchio, R., Romanos, J., Chmielewska, A., Putter, H., Ivarsson, A., Szajewska, H., Koning, F., Wijmenga, C., et al. (2010) The PreventCD Study design: towards new strategies for the prevention of coeliac disease. Eur. J. Gastroenterol. Hepatol., 22, 1424–30.

12. Magni, S., Comani, G. and Elli, L. (2014) miRNAs Affect the Expression of Innate and Adaptive Immunity Proteins in Celiac Disease. Am. J. Gastroenterol., 10.1038/ajg.2014.203 [Epub ahead of print].