University of Groningen

HLA and other tales: The different perspectives of Celiac Disease

Gutierrez Achury, Henry Javier

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2015

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):
Gutierrez Achury, H. J. (2015). *HLA and other tales: The different perspectives of Celiac Disease*. University of Groningen.

Chapter 3.1

Appendix. Update in the genetics findings of autoimmune diseases

Isis Ricaño-Ponce[1], Cisca Wijmenga[1], Javier Gutierrez-Achury[1]

1. University of Groningen, University Medical Center Groningen, Department of Genetics, 9700 RB Groningen, the Netherlands

Abstract

New insights into the underlying molecular pathophysiology of celiac disease (CeD) over the last few years have been guided by major advances in the fields of genetics and genomics. The development and use of the Immunochip genotyping platform paved the way for the discovery of 39 non-HLA loci associated to CeD, and for follow-up functional genomics studies that pinpointed new disease genes, biological pathways and regulatory elements. By combining information from genetics with gene expression data, it has become clear that CeD is a disease with a dysregulated immune response, which can probably occur in a variety of immune cells. This type of information is crucial for our understanding of the disease and for providing leads for developing alternative therapies to the current gluten-free diet. In this review, we place these genetic findings in a wider context and suggest how they can assist the clinical care of CeD patients.

Introduction

During the last few years, the field of human genetics has benefited from an enormous gain of knowledge, specially due to the development of new technologies and techniques, growing disease cohorts, and new methods of data analysis and data integration to uncover new disease genes, pathways and regulatory networks for complex diseases.

Nowadays it is possible to analyze almost every kind of biological sample. Apart from DNA for genotyping, samples can cover individual cell types, RNA for gene expression, and proteins and metabolites from serum or plasma. The information extracted from these biological systems yields important insights into the complex biology of disease.

This technological leap initially allowed for the interrogation of hundreds of thousands of single nucleotide polymorphisms (SNPs) across the human genome, a process called genome-wide association studies (GWAS). With genotype information from a random sample of the population and from a group of patients, case-control association

studies are able to pinpoint the regions or genes potentially related to the pathophysiology of a disease, for example on celiac disease (CeD) [1, 2]. Two GWAS studies in CeD on some 4,918 patients and 5,684 controls led to the discovery of 26 loci outside the well-known HLA association. By comparing GWAS results between different diseases, we can detect regions and genes common to multiple diseases. Such pleiotropic effects can discover common pathways involved in phenotypically different, but biologically related pathologies, as shown for a group of autoimmune diseases [3]. By 2010, GWAS had found 186 loci to be associated to ten different autoimmune diseases and many of these loci showed association to more than one disease.

The genetic analysis of CeD represents an outstanding example of this development, with an important number of loci discovered not only in Caucasian populations but also in other ethnicities. New disease pathways have been linked to these loci [4].

Immunochip in celiac disease

The release of the dedicated Immunochip platform in 2010 gave a huge boost to the process of discovering new genetic regions linked to autoimmunity. This customized array contains 196,524 SNPs that are located in the 186 regions of immunologic interest and includes evidence of association based on previous GWAS analyses in ten different autoimmune diseases [5]. The Immunochip has become a popular genotyping platform because of its customized coverage and cost-efficiency. Its special design means it is suitable for use in European populations, but it is less informative for other ethnic groups. The chip contains SNPs that were known in the public domain by February 2010 [6], which means it lacks an important number of rare variants that have been discovered since then. It has been assumed that these rare variants have a stronger effect on disease susceptibility, but they are also more difficult to find as this requires large-scale sequencing studies. Probably one of the major weaknesses of the Immunochip is that it does not cover the entire genome, it is now apparent that

it eliminates potentially important regions for autoimmune diseases from analysis [5]. Nevertheless, the array has proven to be extremely efficient for deep replication of associations across a wide range of autoimmune diseases, as well as for the purpose of fine-mapping well-established and significant GWAS loci.

New findings from Immunochip

Using the Immunochip platform, in 2011 Trynka et al. [7] analyzed CeD cohorts from six different countries, encompassing 12,041 cases and 12,228 controls. They not only confirmed the loci discovered in previous analyses [1, 2], but also identified new associations, bringing the number of known CeD loci, including the MHC-HLA region, to 40 (Figure 1). These loci are represented by 57 independent SNP associations, of which 29 were localized to a single gene. With the use of proxies (i.e. closely correlated SNPs), three protein-altering SNPs were identified in the *MMEL1, SH2B3* and *IRAK1* genes, while other disease SNPs where localized in regulatory regions of the *RUNX3, RSG1, ETS1, TAGAP, ZFP36L1, IRF4, PTPRK* and *ICOSLG* genes. The remaining disease SNPs lie in intergenic regions [7]. The genes from these associated regions can be connected to multiple biological pathways like hematopoiesis, cell differentiation and selection, activation, co-stimulation and maturation of effector cells, or to regulation of the immune processes.

By linking genes and potential pathways it is possible to pinpoint genes that affect multiple independent autoimmune diseases (genes with pleiotropic effects), such as *CCR1/CCR2* (on chromosome 3p21.31) involved in cell differentiation, recruitment and signaling, or *FASLG* (on chromosome 1q24.3) involved in cell selection, survival, activation and co-stimulation of T cells. Other genes like *IL2, IL21* and *CTLA4* are involved in the creation of a pro-inflammatory environment. This is the final result of the interaction with cytokines (like IL-15, IL-21, IFN-g), autoantibodies and the activation of intraepithelial lymphocytes. The outcome of this complex interaction network is an impairment of
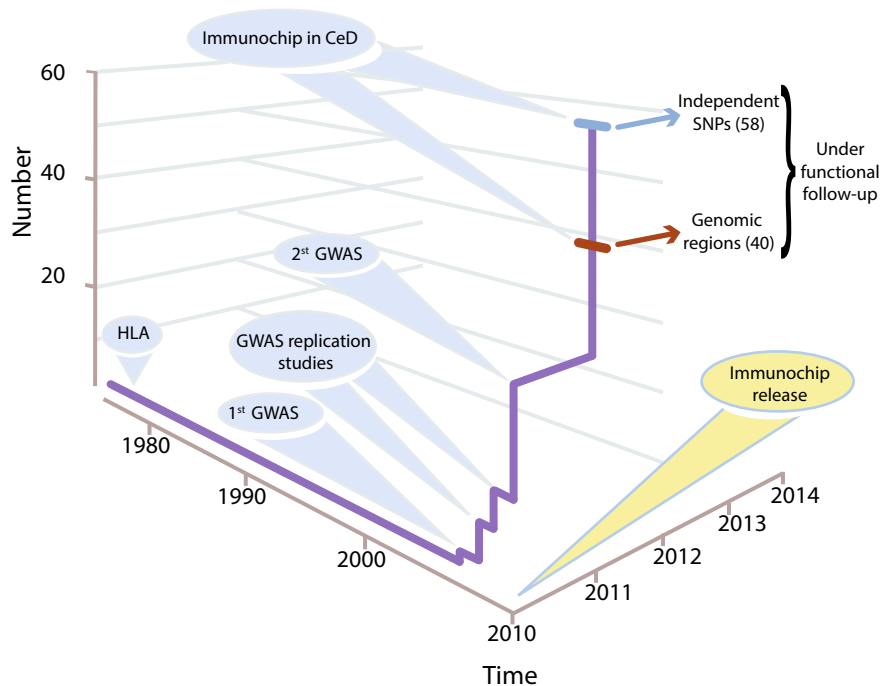
Figure 1. Chronological line representing how the number of discoveries has risen since 2007 when the first GWAS analysis was performed in 778 cases and the important increase in discoveries and number of samples analyzed since the arrival of the Immunochip in 2010.

the intestinal barrier function, which may contribute to the development of CeD [8].

Immunochip in other autoimmune diseases

To date (February 2015), the Immunochip platform has been applied to 14 different autoimmune diseases, including alopecia areata (AA) [9], atopic dermatitis (AD) [10], ankylosing spondylitis (AS) [11], autoimmune thyroid disease (ATD) [12], celiac disease (CeD) [7], inflammatory bowel disease (IBD) [13], juvenile idiopathic arthritis (JIA) [14], multiple sclerosis (MS) [15], primary biliary cirrhosis (PBC) [16], psoriasis (PS) [17], rheumatoid arthritis (RA) [18, 19], primary sclerosing cholangitis (PSCh) [20], Sjögren's syndrome (SJO) [21], and systemic scleroderma (SS) [22]. Although the replication and discovery of new loci via Immunochip

has been successful for many of these diseases, the results depend strongly on the prevalence of the disease, the population under study, the sample size, the relative risk conferred by each of the loci under analysis (the genetic architecture of the disease), and the proximity of the causal variant to the interrogated SNP markers on the Immunochip array (Figure 2).

Most of these autoimmune diseases show a varied prevalence, depending on the population (e.g. CeD in Brazil or Argentina is 0.4-0.6% compared to 2% in Finland) [23]. Most of the Immunochip analyses so far have been carried out in Caucasian populations where the prevalence of autoimmune diseases is relatively high, but some studies have been conducted in other ethnicities like East Asians and Latin Americans [7, 11, 19].

The combined Immunochip data of all 14 autoimmune diseases (Figure 3) shows that there are fewer than 60 associated loci for most of the diseases (from GWAS or Immunochip studies). Only three diseases have more associated loci: 110 for IBD, 97 for MS and 101 for RA. A common denominator for these three phenotypes is the large sample size analyzed in each of the studies. Another factor is the inherent genetic architecture of each of the diseases, which is also related to the statistical power of the analysis. The huge number of SNPs tested in an analysis like Immunochip means that an association must reach a stringent statistical significance ($P < 5x10^{-8}$) to be considered as a true positive [24]. Such a stringent threshold excludes many common SNPs (usually hundreds of them) with suggestive p-values and modest risk effects on disease from the current analyses, although together these SNPs are known to make a major contribution to the heritability of complex diseases [25].

A major challenge in the immunogenetics field has been to increase the power of the diseases under study in order to capture all the risk SNPs, notwithstanding that they also depend on the disease prevalence. There is, however, a trade-off between sample size and prevalence in the power to detect genetic associations: a higher
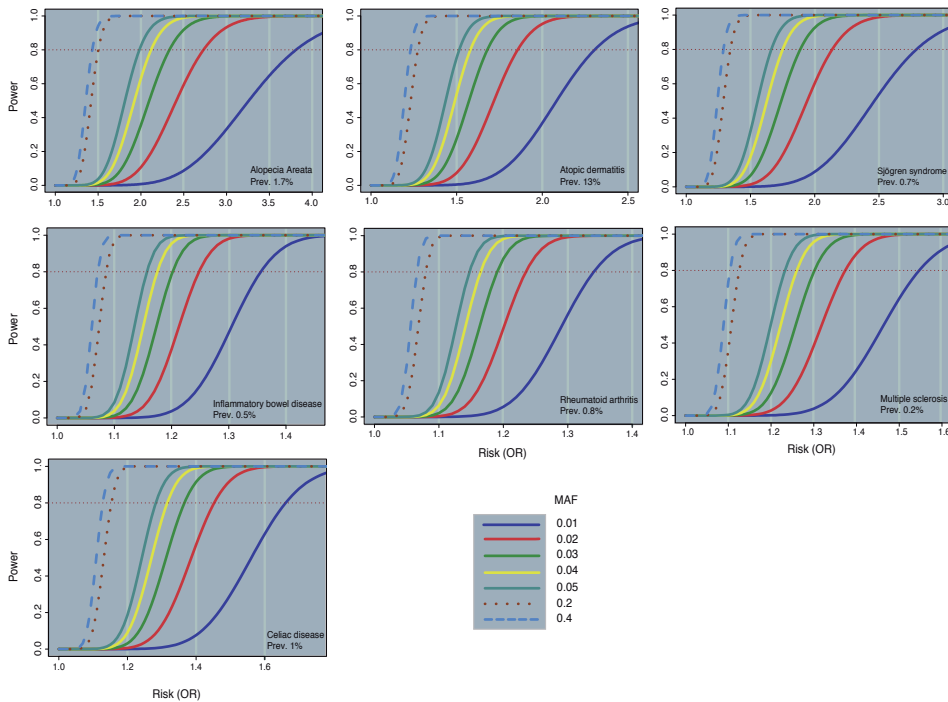
Figure 2. Power calculation for autoimmune diseases with a relatively high prevalence (alopecia areata, atopic dermatitis and Sjögren syndrome) with a low number of discovered loci compared with autoimmune diseases with a low prevalence but high number of discovered loci (inflammatory bowel disease, rheumatoid arthritis and multiple sclerosis). The power was calculated for rare variants (MAF ≤ 1%), low frequency variants (MAF 2-5%) and high frequency variants (MAF > 5%), using the last sample size reported for each disease. As can be observed, the low prevalence diseases have benefited for the important sample size used in it analysis, reaching a proper statistical power for all the range of frequencies, compared with the high prevalence diseases (which in theory could be more easy to obtain a proper sample size for analysis).

prevalence and larger sample sizes increase power. It is interesting that AA, AD and SJO show a relatively high disease prevalence (1.7%, 13% and 0.7%, respectively) while IBD, RA and MS show lower disease prevalences (0.5%, 0.8% and 0.2%, respectively). Yet studies on the latter diseases have been highly efficient in gene discovery because the sample size has lifted the study power substantially (Figure 2). The reason why one category of diseases is studied more than another might be related to the clinical symptoms and the detrimental effects of the condition for the patients. Hence, AA or AD could be seen as
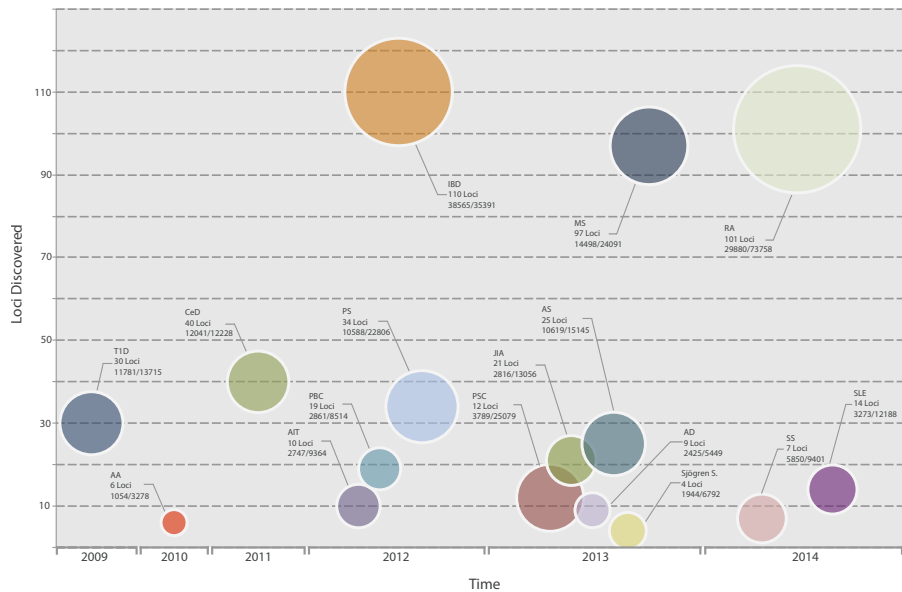
Figure 3. Number of loci discovered in different autoimmune diseases since 2009. The bubbles plot shows each of the diseases in which Immunochip has been applied to discover new genetic associations (with the exception of T1D and SLE). Each of the bubbles represents one disease and its size correlates with the number of samples analyzed. Note that none of the diseases, with the exception of RA, MS and IBD, reach the threshold of 50 associated loci (this is also related to the sample size). Alopecia areata (AA), atopic dermatitis (AD), ankylosing spondylitis (AS), autoimmune thyroid disease (ATD), celiac disease (CeD), inflammatory bowel disease (IBD), juvenile idiopathic arthritis (JIA), multiple sclerosis (MS), primary biliary cirrhosis (PBC), psoriasis (PS), rheumatoid arthritis (RA), primary sclerosing cholangitis (PSCh), Sjögren's syndrome (SJO) and systemic scleroderma (SS).

diseases with a mere cosmetic component and not life-threatening, whereas IBD, RA or MS are severe morbidities. In CeD, the increase in the sample size (and number of loci discovered so far) has been advancing since the first GWAS performed in 2007 [2]: with 778 cases in the first GWAS to 4,500 cases in the second [1], and 12,041 cases in the Immunochip analysis [7] (Figure 1).

Despite the difference in the number of loci discovered for each disease, it is possible to observe relationships among the biological and genetic factors involved in the individual diseases (Figure 4). Some loci are associated to five or more diseases, while others are associated only to a single disease. Loci in the first case might be involved in early or commons steps in the development of autoimmunity, whereas those

associated to a single disease could point to more specific disease processes.

It is important to realize that nearly all the genetic associations to autoimmune diseases that have been found so far correspond to common variations (frequency ≥ 5%) with very modest effects (odds ratio (OR < 1.5). Only in some exceptions have associations been seen to rare or low frequency variants (in CeD, IBD, PBC and RA), and these are often related to the number of samples analyzed and hence the study's power or methodology (e.g. the use of sequences instead of genotypes) [26]. Two examples serve to demonstrate the complex analysis leading to the discovery of rare variants. A recent exome sequencing study in CeD, using extended families (in a linkage analysis) and with re-sequencing of GWAS candidate genes, did not find any rare alleles in coding regions associated to the disease [27]. A similar scenario was presented by Hunt et al., who, after re-sequencing the exons of 25 candidate genes in 41,911 individuals, only identified a nearly "common" non-synonymous variant with very modest effect (rs17849502, minor allele frequency (MAF = 0.049, OR = 1.35) in the *NCF2* gene [26]. With hindsight, this variant was also found to be present in the CeD Immunochip study by Trynka et al., but it had been removed from further analysis by their quality control criteria [7].

The common variants explain a very small proportion of the phenotypic variation in the population attributable to the genetic variation among individuals, which is also known as heritability [28]. The explained heritability also depends on the type of study used for the analysis (family studies or case-control GWAS data), the prevalence of the disease, the allelic frequency of the variables included in the model (which will depend on the population under analysis) and the risk provided by those variables. According to our estimates based on case-control GWAS data, 41% of the heritability of CeD resides in the MHC-HLA region and only 6% in non-HLA significant associations [29]. However, based on family studies, 87% of the total heritability of CeD can be explained [30]. The 40% difference between the heritability
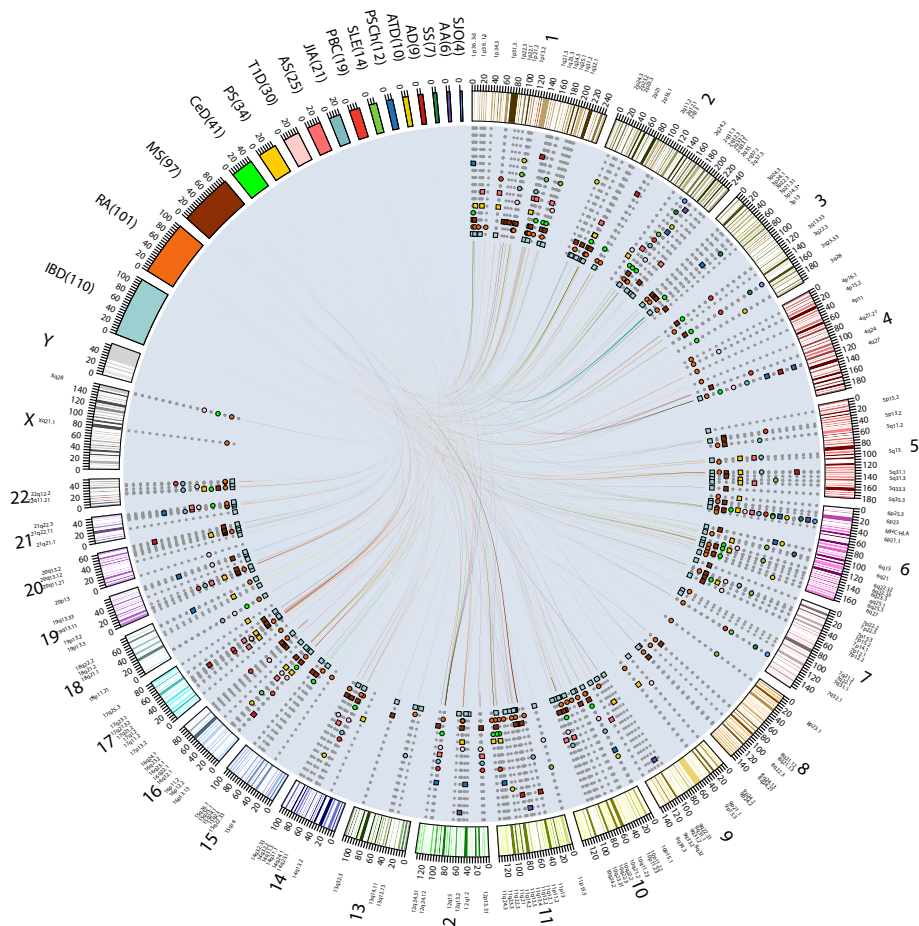
Figure 4. It is well known that phenotypically different autoimmune diseases are related biologically, although the amount of overlap between diseases is not completely clear yet. From the genetics findings, it is possible to determine that most of the diseases share at least one locus, even though some loci are associated to only one disease. Both scenarios are interesting, since those loci only associated to one disease could be related to specific biological pathways leading to the development of the specific phenotype, while, on the other hand, those loci shared by several diseases could be involved in the initial steps of autoimmune deregulation. The number of loci found so far (in parenthesis) is shown for each disease represented in the circus plot.

explained by known variants and by family studies is known as the "missing" heritability. It might be explained in part by thousands of common variants with very low effects – or rare variants with modest effects – which are not identified by association analysis because of power issues. A recent study suggested that another 2,550 common SNPs with modest effect may contribute to CeD susceptibility [25, 28].

The majority of CeD SNPS map to regulatory variations

One of the biggest challenges lies in the interpretation of the genetic associations identified so far. For all autoimmune diseases, including CeD, more than 90% of the disease SNPs are located outside protein-coding genes, which suggests they may have a regulatory role [31], although how this works is not yet clear.

One way to test if disease-associated SNPs are regulatory is to investigate if they regulate the expression of nearby protein-coding genes (so-called cis-eQTLs) (Figure 5). At the same time, expression quantitative trait loci (eQTL) analysis allows for prioritization of genes from regions of association that often contain more than one gene [8]. The first eQTL analysis in CeD, conducted in 1,469 whole blood samples reflecting primary leucocyte expression, showed that 38% of the disease SNPs were cis-eQTLs [1]. This is significantly higher than would be expected by chance (P = 9.3×10⁻⁵) and indicates that CeD-associated SNPs are greatly enriched for cis-eQTLs. Despite the importance of blood as an informative tissue in CeD due to the role of immune cells, one-third of the cis-eQTLs act in a tissue-specific manner, highlighting the importance of performing cis-eQTLs analyses in other relevant tissues (Table 1) [32].

Based on the hypothesis that the thymus plays a role in the deregulation of T cells in the development of CeD, Amundsen et al. analyzed thymic tissue for cis-eQTL effects of 50 SNPs located in the 39 non-HLA regions associated with CeD [33]. They found that 54% of the SNPs analyzed showed a cis-eQTL effect, of which 11 SNPs could represent potentially novel, thymus-specific, cis-eQTLs. These interesting results still need to be replicated due to the limited sample size studied (42 samples) and the lack of significant statistical evidence for some of the eQTLs they found.

The analysis of cis-eQTLs could be even more complex because some of them are also stimulus-dependent. Fairfax et al. found that more than 50% of the cis-eQTL overlapping GWAS loci were observed
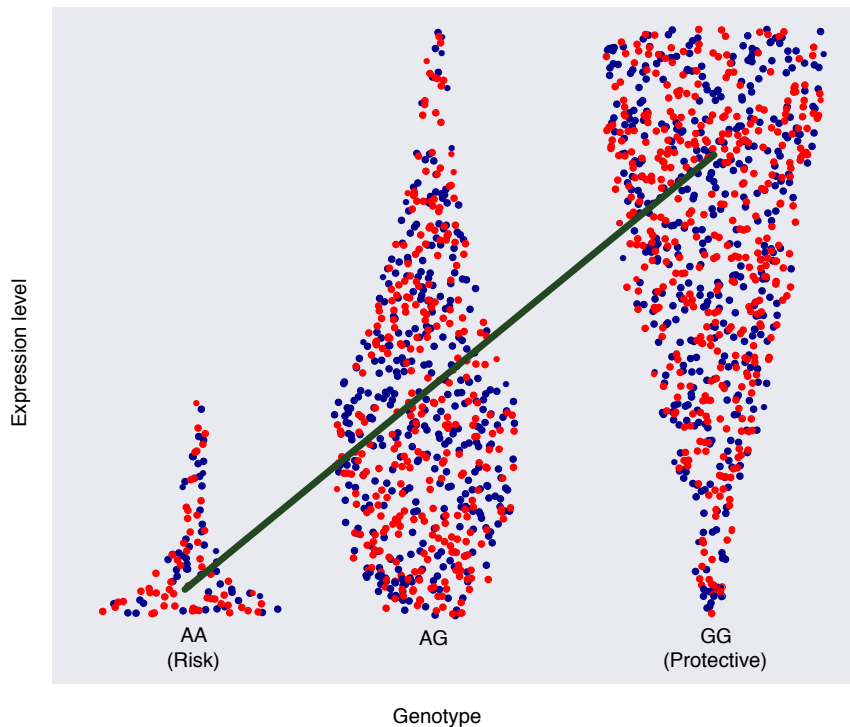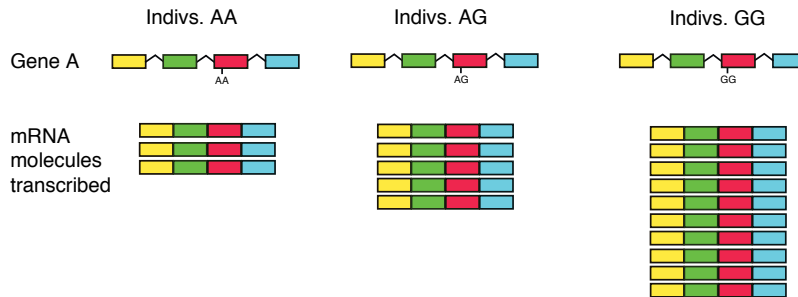
Figure 5. Some SNPs play a regulatory role in the expression of genes located nearby (cis-eQTL effect) or several thousands of nucleotides away, even in different chromosomes (trans-eQTL effect). This effect can be measured in nearly all kinds of tissues, thus yielding information about specific functional patterns, depending on the tissue analyzed.

only in cells after stimulation with IFN-gamma [34]. In a similar study [35], in which dendritic cells were stimulated with lipopolysaccharides (LPS), influenza virus, or IFN-beta, 38 SNPs associated with immune-mediated diseases were found to be significant cis-eQTL only after the

stimulation. In this case, Lee et al. observed that a SNP associated to CeD and RA affected the expression of TRAF1, a gene playing a role in cell survival and apoptosis in dendritic cells after stimulation with LPS and influenza virus [35].

A recent study [36] mapped cis-eQTLs using intestinal biopsies from CeD patients with active disease, from patients who had followed a gluten-free diet (GFD), and from healthy individuals. In this study, 44 SNPs and 45 candidate genes were investigated, which resulted in four cis-eQTL as well as multiple CeD SNPs that affected the expression of genes far away (often on other chromosomes, so-called trans-eQTLs). In agreement with the other studies mentioned above, they also found some stimulus-dependent eQTLs [36]. To obtain a full picture, more eQTL studies need to be done in disease-specific tissues and under different conditions, in order to prioritize the candidate genes from the associated loci.

Non-coding RNAs and CeD

Non-coding RNAs (ncRNAs) are functional molecules that are not translated into proteins [3], and their role in the immune system and immune diseases has been recently reviewed [37]. Long non-coding RNAs (lncRNAs) and microRNAs (miRNAs) are two important classes of ncRNAs: lncRNAs are transcribed RNA molecules longer than 200 nucleotides, while miRNAs are small (22 nucleotides on average). Both types of ncRNAs regulate gene expression in a sequence-specific manner.

In previous work we have shown that 10% of the SNPs associated to 11 immune-mediated diseases overlapped with lncRNAs [3], suggesting a possible role in the gene regulation of these diseases. To test if disease SNPs also can affect the expression of ncRNA genes, similar to that observed for protein-coding genes, Kumar et al. performed an eQTL analysis specifically aimed at lncRNAs. In this study, 125 cis-eQTLs were identified and some 10% overlapped with SNPs present in the GWAS catalog [38]. This number is low since they used a platform in which microarrays contained only a small fraction of all ncRNA genes.

Table 1. Overview of eQTLs in loci associated to celiac disease

Chromosome	Region	Number of genes	Reported candidate genes	Blood	Thymus	Monocytes	Dentritic cells	Celiac disease biopsies
1	1p36.32	57	C1orf93, MMEL1, TTC34	PLCH2, TNFRSF14, C1orf93, MMEL1	HES5, PANK4			
1	1p36.11	160	RUNX3		GRHL3, IL28RA, IL22RA1			
1	1q24.3	48	FASLG/ TNFSR18					TNFS18
1	1q31.2	35	RGS1	DDX59	RSG13, TROVE2		RGS1	
1	1q32.1	237	C1orf106	DDX59	GPR25, DDX59, KIF21B			
2	2p16.1	82	PUS10	AHSA2	AHSA2, C2orf74, XPO1			
2	2p16.1	82	PLEK and FBX048		PLEK, PPP3R1			
2	2q12.1	62	IL18R1, IL18RAP	IL18RAP	IL18RAP, TMEM182	IL18R1		
2	2q31.3	29	UBE2E3 and ITGA4					
2	2q32.3	45	STAT4					
2	2q33.2	40	CD28, CTLA4 and ICOS		RAPH1			ICOS
3	3p22.3	73	CCR4 and GLB1					
3	3p21.31	236	CCR1, CCR2, CCR3 and LTF	CCR3, CDCR6	TSP50, CCR1, LRRC2			
3	3q13.33	64	ARHGAP31		ARHGAP31			
3	3q25.33	30	SCHIP1 and IL12A				IL12A	
3	3q28	44	LPP					
4	4q27	42	KIAA1109, ADAD1, IL2, IL21		NUDT6			
6	6p25.3	30	IRF4 (NM_002460.3)					
6	6q15	75	BACH2 (NM_001170794.1)					
6	6q22.33	39	PTPRK (NM_002844.3)					
6	6q23.3	67	OLIG3 and TNFAIP3					
6	6q25.3	92	TAGAP (NM_152133.1)	TAGAP	EZR-AS1			
7	7p14.1	42	ELMO1	ELMO1	ELMO1			ELMO1
8	8q24.21	59	PVT1					
10	10p15.1	65	PFKFB3 and PRKCQ					
10	10q22.3	79	ZMIZ1	ZMIZ1	PPIF			
11	11q23.1	70	POU2AF1, C11orf93					
11	11q23.3	171	TREH and DDX6					
11	11q24.3	49	ETS1 (NM_001162422.1)					
12	12q24.12	20	SH2B3 and ATXN2	SH2B3, ALDH2, TMEM116	FAM109A, ERP29			
14	14q24.1	52	ZFP36L1					
15	15q24.1	76	CLK3, CSK					
16	16p13.13	71	CIITA			C16orf75		
16	16p13.13	71	SOCS1, PRM1 and PRM2					
18	18p11.21	142	PTPN2		SLMO1			
21	21q22.3	195	UBASH3A					
21	21q22.3	195	ICOSLG	RRP1	PDXK			
22	22q11.21	195	UBE2L3, YDJC	UBE2L3	UBE2L3, TOP3B, HIC2			
X	Xq28	244	HCFC1, TMEM187, IRAK1					TMEM187

RNA sequencing (RNA-seq) is a much better technique for studying ncRNAs, as it allows all the transcripts (protein-coding and ncRNA genes) in the genome to be quantified [38]. Using RNA-seq data did indeed show that many of the immune-mediated disease loci contained lncRNA genes: the loci of nine diseases (including CeD) were found to contain 240 lncRNAs and 626 protein-coding genes in 11 different immune cell types [39]. The study revealed a ratio of approximately 1:3 of lncRNA to protein-coding gene in almost all the associated disease loci. As expected, the loci shared by immune-mediated diseases, meant that the lncRNAs were also shared among diseases. Interestingly, the highest number of shared lncRNAs (11), as well as the highest number of protein-coding genes (51), were observed between RA and CeD (representing 31% of all RA lncRNAs and 30% of all CeD lncRNAs vs. 40% of all RA protein-coding genes and 40% of all CeD protein-coding genes) [39]. The lncRNAs in these loci showed, on average, a 2.5-fold higher expression than the remaining lncRNAs in the genome in the immune cell types analyzed, suggesting they do indeed play a role in immune processes.

RNA-seq data will also be crucial in implicating ncRNAs as causal genes. By performing a cis-eQTL analysis in peripheral mononuclear cells from 629 healthy individuals, we have shown that 42 risk SNPs associated to 14 immune-mediated diseases affect the expression of 47 lncRNAs (Ricaño-Ponce et al., manuscript submitted). We tested 40 SNPs associated to CeD and found 17 that regulate the expression of 28 genes, of which eight (i.e. 29%) correspond to ncRNAs. An interesting example occurs in the 2q31.3 locus, which is shared by CeD and AS. Previous eQTL analysis suggested UBE2E3 as a causal gene, however, using the RNA-seq approach, we determined a cis-eQTL effect on lncRNA AC104820.2. This lncRNA is strongly expressed in CD8+ T-cells and is suggested to function in alpha-beta T-cell proliferation and B-cell activation, both of which are crucial processes in autoimmunity. AC104820.2 was also found to be up-regulated in the intestinal biopsies of patients with active CeD, further confirming its role in the disease [36].

MiRNAs are also recognized as important in CeD. A recent study found that 20% of the miRNAs studied in intestinal biopsies from children with CeD and in healthy controls were differentially expressed [40]. For example, miR-449a was found to be overexpressed in patients. This miRNA is known to target and reduce the levels of NOTCH1 and KLF4, which were shown to correlate with a lower number of goblet cells in the small intestine of CeD patients [40]. The differences in the levels of miRNAs were not only related with the disease status (healthy or affected), but also with the phenotype expression: for example, miR-194-5p and miR-368 were found to be differentially expressed in anemic CeD patients [41].

Cellular and animal models in CeD

In order to understand how the different genes associated with CeD relate to the disease biology, in vitro or in vivo models can be useful. In addition, the limited number of phenotypes that can be safely measured in humans contrasts with the potentially exhaustive manipulation that can be performed in a model organism [42, 43]. The in vitro models of CeD include non-T cells (epithelial cells, macrophages, dendritic cells and monocytes) or T cell cultures propagated from small intestinal biopsies of patients and tumor cell lines (THP monocytic or Caco-2 cell lines) [44, 45]. These models have helped to evaluate the contribution of different cell types to disease, as key cell types relevant to a disease can be selectively isolated by using cell sorting or magnetic separation, then expanded with growth factors, cytokines and/or antigen cocktails, and finally manipulated to evaluate particular cell functions [46]. An example is the potential use of regulatory T cells (Tregs) as immunosuppressive agents to repress autoimmune diseases, or the ex vivo expansion of intraepithelial lymphocytes that are present as an infiltrating population in CeD to evaluate the mechanisms involved in their expansion [47].

Despite the importance of in vitro models, there are limitations in mimicking the environment, hormones, signals and cell-cell interactions in the organs. In vivo models for spontaneous or induced CeD have

been developed in dog, monkey, rabbit, rat and mouse. Although each animal model has a typical pathophysiology and both strengths and weaknesses, they have provided essential information towards the systematic understanding of diseases [48, 49]. For example, the most widely used animal models in CeD are mice that express the HLA genes DQ8 or DQ2, the major genetic risk factor contributing to CeD. These mice elicit a potent inflammatory T cell response against gliadin, but this trigger alone is not sufficient to develop a gluten-dependent enteropathy characterized by shortened villi. Thus, this stresses the importance of the non-HLA loci [48].

Prioritization of causal variants and genes by an integrative approach

The availability of huge amounts of functional data in the public domain and the integration of this information across different cell types have contributed to the development of new prioritization methods for causal variants and genes, revealing new pathways involved in disease. For example, Kumar et al. prioritized 41 SNPs and 49 genes as causal for CeD [8] after intersecting the SNPs associated to CeD with regulatory elements generated by the ENCODE consortium, such as DNase I hypersensitive sites (DHS), DNase I footprints, and transcription factor binding motifs. Interestingly, they observed that 60% of the SNPs were overlapping at least one regulatory element, thereby confirming earlier observations based on eQTLs. Looking at the enrichment of enhancers in different cell types, Kumar et al. confirmed an important role for T-cells in CeD and also discovered B-cells to be important players [8]. Using co-expression analysis, they implicated four genes in the intestinal barrier function, further substantiating a role for this pathway in CeD [50]. They also observed that 30% of the prioritized genes were co-expressed with INFG, confirming a genetic link with the strong, pro-inflammatory IFN-gamma directed response observed in CeD.

The future of CeD genetics

Studies of human genetics have been extremely successful over the last 10 years and led to major new insights into human disease. This success is reflected by the large number of GWAS discoveries that have furthered our understanding of the biological pathways predisposing to disease. However, translating this knowledge to the clinical and molecular phenotyping of disease has been slow. One key aspect has been the time required to collect large numbers of patient samples and the time needed to develop cost-effective molecular biology techniques. These important steps have so far mostly been implemented in studies of Caucasian populations. Some small steps have been taken in other ethnicities. Senapati et al. studied a cohort of North Indian CeD patients and compared the results with a similar sized Caucasian sample using the Immunochip platform [4]. Apart from the MHC-HLA region, five other loci previously associated in Caucasian populations were also replicated in the North Indian samples (corresponding to *FASLG/ TNFSF18, SCHIP1/IL12A, PFKFB3/PRKCQ, ZMIZ1* and *ICOSLG*). Two other loci, *PFKFB3/PRKCQ* and *PTPRK/THEMIS*, also showed association, but with different SNPs, which probably reflects the ethnic differences in genetic background. This type of cross-ethnic analysis has also been applied to other immune-related diseases, such as RA and asthma, and to non-immune related disease, such as type 2 diabetes, lipid levels, body mass index, and cancer [51-55]. These studies have demonstrated the effective use of this methodology in the process of discovering new variants and pathways involved in disease, as well as opening up the possibility of performing cross-population analyses for different diseases.

Is the current genetic information useful in clinical practice?

In silico analysis and experimental models of CeD may provide mechanistic insights which might lead to the development of novel treatments [19, 42, 56]. An interesting example of how genetic data can

be integrated into drug discovery in autoimmune diseases is a recent study in which – by combining genetic analyses and functional in silico annotation – 27 drug target genes for RA treatment were found to overlap with 98 biological RA risk genes (this number of risk genes overlapped with 871 drug target genes for other diseases). This shows the potential of genetics and the impact that these kinds of studies can have on translational medicine [19, 57].

Another exciting possibility lies in the use of ncRNAs as biomarkers for disease. Although miRNAs in CeD have so far only been investigated in intestinal biopsies, these molecules also appear to be stable in circulating and other body fluids, which means they carry the potential for developing into non-invasive tests. For example, circulating miRNA profiling in the plasma showed up-regulation of 11 miRNAs specifically in patients compared to controls [58].

Practice points:

• Celiac disease is a complex disease with a strong influence from genetic factors, especially from the MHC-HLA region.

• The integration of clinical, genetic and functional data has been possible as a result of close collaborations between clinicians and researchers.

• The characterization of loci shared by multiple diseases can be helpful in identifying new cross-disease therapies.

• The genetic factors associated to celiac disease will allow the construction of genetic risk profiles that may be used for stratification of patients.

Research agenda

• The translation of the genetics findings in celiac disease is challenging, but is slowly increasing our knowledge on the pathways involved in disease pathogenesis.

• Functional studies to better understand the effect of the genetic

variants associated to celiac disease should be performed in tissues and cells relevant to the disease, such as intestinal biopsies and gluten-restricted T cells.

• Further studies exploring the non-coding part of the genome will increase our understanding of the key regulators in celiac disease.

• Larger sample sizes and studies in different populations is needed to discover new loci contributing to celiac disease.

Summary

Just four years after the release of the Immunochip platform, the results of immunochip studies in a wide range of autoimmune diseases have greatly expanded our knowledge on disease etiology. We have seen a large number of loci discovered for all autoimmune disorders, an ever-increasing number of samples being studied, and the translation to functional studies to enhance our understanding of the biological roles of the individual disease genes from the associated loci. At the same time only a modest part of the genetic variation can be explained, suggesting that many more the genetic variants remain to be discovered. We also do not fully understand how genetics contributes to the diverse phenotype expression. We also need to determine how genetics can be used to predict disease susceptibility, or how genetic factors can be used to develop new treatments.

For CeD 41 different loci have been identified explaining 47% of the genetic variation. Most of the genetic variants have a role in gene regulation, presumably leading to a dysregulation of immune homeostasis in the intestinal immune system.

In CeD, both the development of novel statistics and functional disease models will help to further our understanding of the biological pathways underlying the disease phenotype, how these pathways are regulated, and to develop new therapies to treat patients.

Acknowledgments

References

1. Dubois PC, Trynka G, Franke L, Hunt KA, Romanos J, Curtotti A, et al. Multiple common variants for celiac disease influencing immune gene expression. Nat Genet. 2010;42(4):295-302.

2. van Heel DA, Franke L, Hunt KA, Gwilliam R, Zhernakova A, Inouye M, et al. A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21. Nat Genet. 2007;39(7):827-9.

3. Ricano-Ponce I, Wijmenga C. Mapping of immune-mediated disease genes. Annual review of genomics and human genetics. 2013;14:325-53.

4. Senapati S, Gutierrez-Achury J, Sood A, Midha V, Szperl A, Romanos J, et al. Evaluation of European coeliac disease risk variants in a north Indian population. Eur J Hum Genet. 2015;23(4):530-5.

5. Cortes A, Brown MA. Promise and pitfalls of the Immunochip. Arthritis Res Ther. 2011;13(1):101.

6. Genomes Project C, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, et al. A map of human genome variation from population-scale sequencing. Nature. 2010;467(7319):1061-73.

7. Trynka G, Hunt KA, Bockett NA, Romanos J, Mistry V, Szperl A, et al. Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. Nat Genet. 2011;43(12):1193-201.

8. Kumar V, Gutierrez-Achury J, Kanduri K, Almeida R, Hrdlickova B, Zhernakova DV, et al. Systematic annotation of celiac disease loci refines pathological pathways and suggests a genetic explanation for increased interferon-gamma levels. Hum Mol Genet. 2015;24(2):397-409.

9. Redler S, Angisch M, Heilmann S, Wolf S, Barth S, Basmanav BF, et al. Immunochip-based analysis: high-density genotyping of immune-related Loci sheds further light on the autoimmune genetic architecture of alopecia areata. J Invest Dermatol. 2015;135(3):919-21.

10. Ellinghaus D, Baurecht H, Esparza-Gordillo J, Rodriguez E, Matanovic A, Marenholz I, et al. High-density genotyping study identifies four new susceptibility loci for atopic dermatitis. Nat Genet. 2013;45(7):808-12.

11. International Genetics of Ankylosing Spondylitis C, Cortes A, Hadler J, Pointon JP, Robinson PC, Karaderi T, et al. Identification of multiple risk

variants for ankylosing spondylitis through high-density genotyping of immune-related loci. Nat Genet. 2013;45(7):730-8.

12. Cooper JD, Simmonds MJ, Walker NM, Burren O, Brand OJ, Guo H, et al. Seven newly identified loci for autoimmune thyroid disease. Hum Mol Genet. 2012;21(23):5202-8.

13. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature. 2012;491(7422):119-24.

14. Hinks A, Cobb J, Marion MC, Prahalad S, Sudman M, Bowes J, et al. Dense genotyping of immune-related disease regions identifies 14 new susceptibility loci for juvenile idiopathic arthritis. Nat Genet. 2013;45(6):664-9.

15. International Multiple Sclerosis Genetics C, Beecham AH, Patsopoulos NA, Xifara DK, Davis MF, Kemppinen A, et al. Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. Nat Genet. 2013;45(11):1353-60.

16. Liu JZ, Almarri MA, Gaffney DJ, Mells GF, Jostins L, Cordell HJ, et al. Dense fine-mapping study identifies new susceptibility loci for primary biliary cirrhosis. Nat Genet. 2012;44(10):1137-41.

17. Tsoi LC, Spain SL, Knight J, Ellinghaus E, Stuart PE, Capon F, et al. Identification of 15 new psoriasis susceptibility loci highlights the role of innate immunity. Nat Genet. 2012;44(12):1341-8.

18. Eyre S, Bowes J, Diogo D, Lee A, Barton A, Martin P, et al. High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. Nat Genet. 2012;44(12):1336-40.

19. Okada Y, Wu D, Trynka G, Raj T, Terao C, Ikari K, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature. 2014;506(7488):376-81.

20. Liu JZ, Hov JR, Folseraas T, Ellinghaus E, Rushbrook SM, Doncheva NT, et al. Dense genotyping of immune-related disease regions identifies nine new risk loci for primary sclerosing cholangitis. Nat Genet. 2013;45(6):670-5.

21. Lessard CJ, Li H, Adrianto I, Ice JA, Rasmussen A, Grundahl KM,

et al. Variants at multiple loci implicated in both innate and adaptive immune responses are associated with Sjogren's syndrome. Nat Genet. 2013;45(11):1284-92.

22. Mayes MD, Bossini-Castillo L, Gorlova O, Martin JE, Zhou X, Chen WV, et al. Immunochip analysis identifies multiple susceptibility loci for systemic sclerosis. Am J Hum Genet. 2014;94(1):47-61.

23. Cooper GS, Bynum ML, Somers EC. Recent insights in the epidemiology of autoimmune diseases: improved prevalence estimates and understanding of clustering of diseases. J Autoimmun. 2009;33(3-4):197-207.

24. Wellcome Trust Case Control C. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007;447(7145):661-78.

25. Stahl EA, Wegmann D, Trynka G, Gutierrez-Achury J, Do R, Voight BF, et al. Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. Nat Genet. 2012;44(5):483-9.

26. Hunt KA, Mistry V, Bockett NA, Ahmad T, Ban M, Barker JN, et al. Negligible impact of rare autoimmune-locus coding-region variants on missing heritability. Nature. 2013;498(7453):232-5.

27. Mistry V, Bockett NA, Levine AP, Mirza MM, Hunt KA, Ciclitira PJ, et al. Exome sequencing of 75 individuals from multiply affected coeliac families and large scale resequencing follow up. PLoS One. 2015;10(1):e0116845.

28. Witte JS, Visscher PM, Wray NR. The contribution of genetic variants to disease depends on the ruler. Nature Rev Genet. 2014;15(11):765-76.

29. Gutierrez-Achury J, Zhernakova A, Pulit SL, Trynka G, Hunt KA, Romanos J, et al. Fine mapping in the MHC region accounts for 18% additional genetic risk for celiac disease. Nat Genet. 2015;47(6):577-8.

30. Nistico L, Fagnani C, Coto I, Percopo S, Cotichini R, Limongelli MG, et al. Concordance, disease progression, and heritability of coeliac disease in Italian twins. Gut. 2006;55(6):803-8.

31. Farh KK, Marson A, Zhu J, Kleinewietfeld M, Housley WJ, Beik S, et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature. 2015;518(7539):337-43.

32. Fu J, Wolfs MG, Deelen P, Westra HJ, Fehrmann RS, Te Meerman GJ, et al. Unraveling the regulatory mechanisms underlying tissue-dependent genetic variation of gene expression. PLoS Genet. 2012;8(1):e1002431.

33. Amundsen SS, Viken MK, Sollid LM, Lie BA. Coeliac disease-associated polymorphisms influence thymic gene expression. Genes Immun. 2014;15(6):355-60.

34. Fairfax BP, Humburg P, Makino S, Naranbhai V, Wong D, Lau E, et al. Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. Science. 2014;343(6175):1246949.

35. Lee MN, Ye C, Villani AC, Raj T, Li W, Eisenhaure TM, et al. Common genetic variants modulate pathogen-sensing responses in human dendritic cells. Science. 2014;343(6175):1246980.

36. Plaza-Izurieta L, Fernandez-Jimenez N, Irastorza I, Jauregi-Miguel A, Romero-Garmendia I, Vitoria JC, et al. Expression analysis in intestinal mucosa reveals complex relations among genes under the association peaks in celiac disease. Eur J Hum Genet. 2015;23(8):1100-5.

37. Stachurska A, Zorro MM, van der Sijde MR, Withoff S. Small and Long Regulatory RNAs in the Immune System and Immune Diseases. Frontiers in immunology. 2014;5:513.

38. Kumar V, Westra HJ, Karjalainen J, Zhernakova DV, Esko T, Hrdlickova B, et al. Human disease-associated genetic variation impacts large intergenic non-coding RNA expression. PLoS Genet. 2013;9(1):e1003201.

39. Hrdlickova B, de Almeida RC, Borek Z, Withoff S. Genetic variation in the non-coding genome: Involvement of micro-RNAs and long non-coding RNAs in disease. Biochimica et biophysica acta. 2014;1842(10):1910-22.

40. Capuano M, Iaffaldano L, Tinto N, Montanaro D, Capobianco V, Izzo V, et al. MicroRNA-449a overexpression, reduced NOTCH1 signals and scarce goblet cells characterize the small intestine of celiac patients. PLoS One. 2011;6(12):e29094.

41. Vaira V, Roncoroni L, Barisani D, Gaudioso G, Bosari S, Bulfamante G, et al. microRNA profiles in coeliac patients distinguish different clinical phenotypes and are modulated by gliadin peptides in primary duodenal fibroblasts. Clinical science. 2014;126(6):417-23.

42. Merkle FT, Eggan K. Modeling human disease with pluripotent stem cells: from genome association to function. Cell stem cell. 2013;12(6):656-68.

43. Choy E, Yelensky R, Bonakdar S, Plenge RM, Saxena R, De Jager PL, et al. Genetic analysis of human traits in vitro: drug response and gene expression in lymphoblastoid cell lines. PLoS Genet. 2008;4(11):e1000287.

44. Junker Y, Zeissig S, Kim SJ, Barisani D, Wieser H, Leffler DA, et al. Wheat amylase trypsin inhibitors drive intestinal inflammation via activation of toll-like receptor 4. J Exp Med. 2012;209(13):2395-408.

45. Nanayakkara M, Lania G, Maglio M, Discepolo V, Sarno M, Gaito A, et al. An undigested gliadin peptide activates innate immunity and proliferative signaling in enterocytes: the role in celiac disease. Am J Clin Nutr; 2013;98(4):1123-35.

46. Campbell JD, Foerster A, Lasmanowicz V, Niemoller M, Scheffold A, Fahrendorff M, et al. Rapid detection, enrichment and propagation of specific T cell subsets based on cytokine secretion. Clin Exp Immunol. 2011;163(1):1-10.

47. Trzonkowski P, Szarynska M, Mysliwska J, Mysliwski A. Ex vivo expansion of CD4(+)CD25(+) T regulatory cells for immunosuppressive therapy. Cytometry Part A : the journal of the International Society for Analytical Cytology. 2009;75(3):175-88.

48. Stoven S, Murray JA, Marietta EV. Latest in vitro and in vivo models of celiac disease. Exp Opin Drug Discov. 2013;8(4):445-57.

49. Marietta EV, Schuppan D, Murray JA. In vitro and in vivo models of celiac disease. Exp Opin Drug Discov. 2009;4(11):1113-23.

50. Monsuur AJ, de Bakker PI, Alizadeh BZ, Zhernakova A, Bevova MR, Strengman E, et al. Myosin IXB variant increases the risk of celiac disease and points toward a primary intestinal barrier defect. Nat Genet. 2005;37(12):1341-4.

51. Al Olama AA, Kote-Jarai Z, Berndt SI, Conti DV, Schumacher F, Han Y, et al. A meta-analysis of 87,040 individuals identifies 23 new susceptibility loci for prostate cancer. Nat Genet. 2014;46(10):1103-9.

52. Bentley AR, Chen G, Shriner D, Doumatey AP, Zhou J, Huang H, et al. Gene-based sequencing identifies lipid-influencing variants with ethnicity-

specific effects in African Americans. PLoS Genet. 2014;10(3):e1004190.

53.Gong J, Schumacher F, Lim U, Hindorff LA, Haessler J, Buyske S, et al. Fine Mapping and Identification of BMI Loci in African Americans. Am J Hum Genet. 2013;93(4):661-71.

54.Grant SF, Thorleifsson G, Reynisdottir I, Benediktsson R, Manolescu A, Sainz J, et al. Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. Nat Genet. 2006;38(3):320-3.

55.Okada Y, Han B, Tsoi LC, Stuart PE, Ellinghaus E, Tejasvi T, et al. Fine Mapping Major Histocompatibility Complex Associations in Psoriasis and Its Clinical Subtypes. Am J Hum Genet. 2014;95(2):162-72.

56.Straub RH. Evolutionary medicine and chronic inflammatory state--known and new concepts in pathophysiology. J Mol Med (Berl). 2012;90(5):523-34.

57.Plenge RM, Scolnick EM, Altshuler D. Validating therapeutic targets through human genetics. Nat Rev Drug Discov. 2013;12(8):581-94.

58.Zahm AM, Thayu M, Hand NJ, Horner A, Leonard MB, Friedman JR. Circulating microRNA is a biomarker of pediatric Crohn disease. J Pediatr Gastroenterol Nutr. 2011;53(1):26-33.