Journal of Autoimmunity xxx (2015) 1–16

Contents lists available at ScienceDirect

Journal of Autoimmunity

journal homepage: www.elsevier.com/locate/jautimm

Review article

The genetics of celiac disease: A comprehensive review of clinical implications

Romina Dieli-Crimi [a], M. Carmen Cénit [b, c], Concepción Núñez [a, *]

[a] UGC de Inmunología, Hospital Clínico San Carlos, Instituto de Investigación Sanitaria del Hospital Clínico San Carlos (IdISSC), Madrid, Spain
[b] Departamento de Pediatría, Hospital Universitario Dr. Peset, Valencia, Spain
[c] Departamento de Ecología Microbiana, Nutrición y Salud. Instituto de Agroquímica y Tecnología de los Alimentos. Consejo Superior de Investigaciones Científicas (IATA-CSIC), Valencia, Spain

ARTICLE INFO

Article history:
Received 30 June 2015
Accepted 1 July 2015
Available online xxx

Keywords:
Celiac disease
Epigenetics
Gene expression
HLA
Microbiota

ABSTRACT

Celiac disease (CD) is a complex immune-related disease with a very strong genetic component. Multiple genetic findings over the last decade have added to the already known MHC influence numerous genetic variants associated to CD susceptibility. Currently, it is well-established that 6 MHC and 39 non-MHC loci, including a higher number of independent genetic variants, are associated to disease risk. Moreover, additional regions have been recently implicated in the disease, which would increase the number of involved loci. Together, the firmly described genetic variants account for roughly 31% of CD heritability, being 25% explained by the MHC influence. These new variants represent markers of disease risk and turn the identification of the causal genes and the causal variants inside the associated loci, as well as their precise biological role on the disease, into a major challenge in CD research. Numerous studies have been developed with this aim showing the high impact of risk variants on gene expression. These studies also indicate a central role of $CD4^+$ T cells in CD pathogenesis and point to B cells as important players, which is in accordance with the key steps highlighted by the immunological models of pathogenesis. We comprehensively summarize the current knowledge about the genetic architecture of CD, characterized by multiple low-risk variants located within diverse loci which are most likely affecting genes with immune-related functions. These findings are leading to a better understanding of CD pathogenesis and helping in the design of new treatments. The repertoire of potential drug targets for CD has largely broadened last years, bringing us closer to get alternative or complementary treatments to the life-long gluten-free diet, the only effective treatment so far. Epigenetics and microbiota are emerging as potent factors modulating disease risk and putatively affecting disease manifestation, which are also being explored as therapeutic targets.

1. Introduction

Described for the first time in the first century AD, the perception of celiac disease (CD) has highly evolved prompted by the advances in epidemiological, clinical and genetic research, that have led to a better understanding of the pathogenesis of the disease. Nowadays, CD is considered a systemic immune-mediated disease triggered by gluten ingestion in genetically susceptible individuals and further characterized by the presence of enteropathy, which can range from intraepithelial lymphocytosis (LIEs) to total villous flattening. The presence of serum antibodies, mainly those directed against type 2 transglutaminase (TG2) (anti-TG2 antibodies and anti-endomysium antibodies) and the HLA-DQ2 and/or HLA-DQ8-encoding alleles also constitute important hallmarks of CD [1].

The etiology of CD is still not completely understood, but the discovery in 1953 of some cereals as the environmental triggers of the disease [2] has facilitated its study. Currently, the term gluten is commonly used to refer to the cereal components involved in CD development. It denotes the proline and glutamine rich proteins present in wheat, barley and rye, although being strict, we should talk about gluten (gliadin and glutenin) in wheat, hordeins in barley and secalins in rye.

* Corresponding author. Servicio de Inmunología Clínica, Hospital Clínico San Carlos, C/ Profesor Martín Lagos s/n, 28040 Madrid, Spain.
E-mail addresses: rodielicrimi@gmail.com (R. Dieli-Crimi), mccenit81@gmail.com (M.C. Cénit), conchita.npardo@gmail.com (C. Núñez).

http://dx.doi.org/10.1016/j.jaut.2015.07.003
0896-8411/

However, gluten ingestion does not necessarily trigger CD. Indeed, only around 1% of individuals in Western populations, where a widespread consumption of cereals exists, develop the disease. Moreover, a wide spectrum of clinical manifestations as well as a high variability in age of onset, severity of mucosal lesion or antibodies levels are observed in CD. The individual genetic background, most likely interacting with other environmental triggers, is probably contributing to the observed variability. A better understanding of all these factors and their complex interaction could provide us with new clues for a better diagnosis and management of CD patients.

In this review, we summarize the current knowledge about the genetic architecture of CD and hypothesized about the putative implications of these findings in CD pathogenesis.

2. Epidemiology

CD is the most frequent food intolerance, with an increasing prevalence over the last three decades [3,4]. This has been partially prompted by globalization, which led to an increased consumption of gluten-containing food worldwide. The development of new sensitive and specific serological tests and their extended availability, as well as the best knowledge of the disease including a wide clinical spectrum have allowed to expand and facilitate CD diagnosis, being indirectly responsible for the rising prevalence. Several environmental factors have been also related to this increase [5]. Changes in breastfeeding practices were for a long time considered to influence CD development [6,7], but last year data demonstrated that the time and amount of gluten introduction do not seem to influence CD risk [8–10], dismissing what was previously thought. Instead, a longer period of breastfeeding seems to delay CD onset [9].

A large number of epidemiological studies have been performed aimed to establish CD prevalence. Fig. 1 shows the reported CD prevalence around the world. Slight differences can be observed between most countries, being most notorious the high prevalence observed in Western Sahara (5.6%) or the absence of the disease reported in Burkina Faso. In general, it is currently estimated that 1% of the general population develops CD, with a higher rate in female gender (ratio of 2:1) [11,12]. This differs from the first considerations of CD as a rare condition. It is important to consider that this prevalence corresponds to the percentage of real cases of CD, which is higher than the number of diagnosed cases. A ratio for diagnosed to undiagnosed cases of 1:3 to 1:5 for pediatric and adult population, respectively, has been reported [13]. One solid interpretation of this situation was given in 1991 by Richard Logan with his idea of a *celiac iceberg* [14], explaining that the whole iceberg represents the true prevalence of the disease, and the water line would be the ratio of diagnosed (visible part) to undiagnosed (under water part) cases. The wide spectrum of clinical symptoms that can appear at any age contributes to underdiagnosis, especially in countries with inadequate awareness, where sometimes common symptoms pass unnoticed or misunderstood. Moreover, some patients show subclinical CD and are often only detected if they are enrolled in screening programs.

Conversely, there may exist some overdiagnosis of the disease, since upper endoscopy and the presence of a qualified pathologist are not accessible in all countries and the diagnosis is made according to serological tests regardless of false positive results [15].

It is expected that genetic and pathogenic insights contribute to a better understanding of the disease and consequently to a better diagnosis, which is a major challenge of research in CD.

3. Genetic risk

CD shows a strong genetic component, which has been demonstrated by studies in siblings, showing a concordance around 80% in monozygotic twins and less than 20% among dizygotic twins [16,17]; and through familial aggregation studies, which calculate the risk of a patient's sibling to develop CD (sibling relative risk), being 20–60 in CD [18–20]. Nowadays, CD is considered to have a high heritability [17] and it is well characterized as a polygenic disease with a complex non-Mendelian pattern of inheritance, involving MHC and non-MHC genes, which jointly provide the genetic risk to develop the disease (Table 1).

3.1. MHC

The main genetic predisposing factors for CD lie on the major histocompatibility complex (MHC) region. This region, located on 6p21, contains hundreds of genes with immunological functions and is responsible for the strongest association signals observed in most immune-mediated diseases.

In CD, the most influent risk variants in this region, as well as their role in triggering the disease, are largely known. In the 1970s, the association between MHC and CD development was first described [21,22]. One decade later, the alleles-encoding HLA-DQ2 were identified as the main responsible of the genetic risk conferred by the MHC region [23,24], which had been previously attributed to the serological alleles HLA-B8 and HLA-DR3 [21,25]. HLA-DQ molecules are conformed by two subunits: α and β, which are encoded by two different genes of the class II MHC region: *HLA-DQA1* and *HLA-DQB1*, respectively.

Around 90% of CD patients carry the alleles encoding HLA-DQ2, specifically, HLA-DQ2.5: *DQA1*05* and *DQB1*02*, which can be inherited in *cis* (in presence of *HLA-DRB1*03*) or *trans* configuration. The resulting *cis* and *trans* HLA-DQ2.5 molecules differ in residues which are not involved in peptide recognition (one residue in the leader peptide in the α-chain and one residue in the membrane proximal domain in the β-chain) and are described as conferring similar risk. These are the main genetic risk factors for CD. In addition, there is a *HLA-DQB1*02* gene dosage effect. When both progenitors transmit this allele to the child, the risk to develop CD is higher than when the child receives this allele from only one of the two progenitors [26–28]. This is conditioned to the presence of at least one copy of the *HLA-DQA1*05* allele. Most of the remaining patients (without HLA-DQ2.5) carry the *DQA1*03* and *DQB1*03:02* alleles, which encode the HLA-DQ8 molecule [29]. A gene dosage effect for HLA-DQ8 has been also proposed [30]. In almost all the CD patients who carry neither HLA-DQ2.5 nor HLA-DQ8, one of the two alleles encoding HLA-DQ2.5 is present: most commonly *DQB1*02* (HLA-DQ2.2) and in a minority of cases, *DQA1*05* (HLA-DQ7.5) [30].

The genetic risk conferred by these described MHC variants ranges from the highest influence of HLA-DQ2.5 to the nil effect assigned to HLA-DQ7.5. However, the role of HLA-DQ7.5 in CD should not be disregarded, since this factor is present in almost all CD patients lacking the known HLA-associated risk variants [[30]; supplementary figure 1 in Ref. [31]].

Risk factors in *HLA-DQA1* and *HLA-DQB1* account for 22% of CD heritability [31]. The complexity of the MHC region, mainly characterized by harboring numerous genes, high polymorphism and broad linkage disequilibrium, made very difficult to identify additional risk variants in this region for many years. However, very recently, a fine mapping of the MHC region taking advantage of high-density imputation, identified five new independent variants in this region [31]. HLA-DPβ1 (position 9), *HLA-B* (the classical *HLA-B*08* and *HLA-B*39:06* alleles) and two SNPs,

Please cite this article in press as: R. Dieli-Crimi, et al., The genetics of celiac disease: A comprehensive review of clinical implications, Journal of Autoimmunity (2015), http://dx.doi.org/10.1016/j.jaut.2015.07.003