Using GWAS to identify genetic predisposition in hepatic autoimmunity
Webb, Gwilym; Hirschfield, Gideon

DOI:
10.1016/j.jaut.2015.08.016

License:
Creative Commons: Attribution (CC BY)

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

Citation for published version (Harvard):

Link to publication on Research at Birmingham portal

Publisher Rights Statement:
Eligibility for repository : checked 16/12/2015

General rights
Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

•Users may freely distribute the URL that is used to identify this publication.
•Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
•User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
•Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy
While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

Download date: 06. Dec. 2018
Review article

Using GWAS to identify genetic predisposition in hepatic autoimmunity

G.J. Webb, G.M. Hirsch*
NIHR Birmingham Liver Biomedical Research Unit, University of Birmingham, Birmingham, UK

ABSTRACT

Primary biliary cirrhosis (PBC), primary sclerosing cholangitis (PSC) and autoimmune hepatitis (AIH) represent the three major hepatic autoimmune conditions. Patient morbidity and mortality remain high across these three diseases, and an unmet need for rational therapy exists. Disease understanding has focused on combining clinical and laboratory based science to provide better insights into the joint host and environmental factors necessary for the initiation, and perpetuation, of hepato-biliary inflammation. Twin studies, family studies, population studies and an inter-relationship with other autoimmune phenomena suggest a genetic component to risk for each disease. Until recently, understanding of this genetic risk has been limited to HLA haplotypes. Associations with risk-conferring and protective HLA haplotypes are present in all three diseases. Over the last few years, genome-wide association studies (GWAS), and related genetic association studies, have greatly increased understanding of the genetic risk signature of these three diseases and autoimmunity in general. Here we consider the rationale for GWAS in general and with specific reference to hepatic autoimmunity. We consider the process of GWAS, and highlight major findings to date. Potential functional implications of key findings are discussed including the IL-12/STAT4 pathway in PBC and the CD28/IL-2 pathway in PSC. We describe the marked pleiotropy demonstrated by PBC and PSC, which is consistent with other autoimmune diseases. Further, we focus on specific gene associations including SH2B3, which is common to all three diseases, and FUT2 in PSC, which represents a link between environment and genetics. We review attempts to translate GWAS findings into basic laboratory models including in vivo systems and highlight where clinical observations relate to genetics. Finally we describe deficiencies in GWAS to date and consider future study of genetics in hepatic autoimmunity.

© 2015 Published by Elsevier Ltd.
1. Introduction

Autoimmunity may be considered as a “clinical syndrome caused by the activation of T cells or B cells, or both, in the absence of an ongoing infection or other discernible cause” [1] and is typically associated with autoantibodies reactive against self antigens [2]. The prevalence of autoimmune in general is approaching 10% [3], and many of these individuals develop autoimmune phenomena affecting multiple organ systems simultaneously [4].

Hepatic autoimmunity comprises three major – and sometimes overlapping – clinico-pathological syndromes: primary biliary cirrhosis (PBC; also referred to as primary biliary cholangitis), autoimmune hepatitis (AIH) and primary sclerosing cholangitis (PSC) [5–7]. Each is associated with autoantibodies and with an increased likelihood of autoimmune affecting other organ systems. Notably, these are uncommon conditions with the prevalence of PBC quoted as 1.91–40.2 per 100,000 and PSC as 0–16.2 per 100,000 [8]. Data for AIH are more sparse, but a recent Danish study has quoted a prevalence of 23.9 per 100,000 [9]. These diseases can impact people of all ages, persist as chronic illnesses rather than acute illnesses, and remain challenging to treat. Of the three diseases, only AIH typically responds to immunosuppression. AIH may be separated into type 1 and type 2 with the former more responsive to immunosuppression. There are few data on the genetics of type 2 AIH and the remainder of this article discusses type 1 AIH.

In each of the three major hepatic autoimmune conditions, as in most major autoimmune diseases, there is evidence for a strong genetic component to risk [10]. Genome wide association studies (GWAS), and related high-density genotyping arrays, represent a powerful method for exploring the genetic makeup of this risk. This review for a Special Edition of Journal of Autoimmunity focuses on the process of applying such studies to hepatic autoimmunity. This is apt given the life-long contribution from Professors Diego and Giorgina Vergani in the field of hepatology, particularly the presentation and management, as well as immunologic phenotype, of autoimmune liver disease in children. The studies discussed herein parallel and complement such clinical and immunologic evaluation of these diseases, collectively driving forward our efforts to better treat patients.

2. Why study genetics in hepatic autoimmunity?

There is a clear genetic component to the risk of developing autoimmunity in general, and hepatic autoimmunity specifically. This inference is based on observations made on twin pairs, family studies, population-based studies, and clinical observations, and underpins the logic behind conducting GWAS. Together these observations strongly suggest a significant genetic component to the development of hepatic autoimmunity. A deeper understanding of genetic associations is hoped to guide investigation of underlying pathobiology and so translate to improved therapies, personalized medicine and more accurate prognostication [11].

2.1. Twin studies

Comparing disease concordance in mono- and dizygotic twins living in similar environments allows estimation of the relative contribution from environmental and genetic inputs to disease risk. Unfortunately, the low frequency of monozygotic twins (approximately 3.5 per 1000 live births) makes sufficient recruitment in rarer diseases such as hepatic autoimmunity challenging.

PBC has the highest reported concordance in monozygotic twins of any autoimmune condition, with the exception of celiac disease: 63% in one small series as compared to a risk in dizygotic twins close to the population level of <0.5% with no concordance between the 8 pairs studied [12]. Intriguingly, ages of presentation and presenting symptoms were similar among concordant twins and two of the discordant twins had other major autoimmune conditions. Similar data are not currently available for AIH or PSC, although a in a recent report there was no genetic concordance in five dizygotic twin pairs with concordance in the single set of monozygotic twins from a national Dutch AIH study [13].

2.2. Familial risk

A particularly increased risk within a family as compared to the surrounding population tends to suggest an increased genetic risk, although unrecognized environmental risks are always also potentially relevant. Such genetic risk can be measured by the relative sibling risk (denoted as AS) i.e. the increased risk conferred on the sibling of an individual affected with a certain trait as a multiple of the baseline population risk. For PBC AS has been calculated at ×10.5 in a large UK series [14] and ×10.7 in the USA [15]. For, PSC AS is approximately ×10 [16], whilst for AIH estimates
are lacking. However, in AIH it is estimated that 40% of individuals have a family history of autoimmunity in general [17]. In PBC, the risk for expressing AMA and for developing disease is increased in first-degree relatives [18,19] and a disproportionate number of PBC patients have relatives with the disease at the point of diagnosis [8,20].

2.3. Populations

When one subset of genetically admixed populations in a shared geographic area have an increased risk of a trait this may also delay a genetic component to risk. Examples of such phenomena in hepatic autoimmunity include reports of numbers of individuals affected by PBC are also abnormally high in ethnically defined sub-populations [21,22] and within selected wider families in characterized comprehensive healthcare systems [19,23]. Data on PSC sub-populations are more scarce, and although significant variation is seen between geographically separated populations, this may also indicate varying environmental exposure [8]. Data in AIH are similarly lacking with diagnostic variability between countries a key consideration when drawing comparisons [24].

2.4. Individuals developing multiple autoimmune conditions suggests shared risk

All three major hepatic autoimmune conditions are strongly associated with coincident extra-hepatic autoimmunity. The findings of three representative reports are summarized in Table 1. Such observations are consistent with autoimmune conditions in general, where development of one autoimmune condition increases the risk of another. However, this is not always the case and, for example, a diagnosis of rheumatoid arthritis is protective against a diagnosis of multiple sclerosis and vice versa [3].

3. The genetic risk profile of hepatic autoimmunity is challenging to assess

As suggested by the significant but relatively modest sibling relative risk and incomplete concordance in monozygotic twins, the major hepatic autoimmune conditions do not follow autosomal inheritance. In contrast, only a small minority of causes of hepatic autoimmunity are monogenetically inherited: Autoimmune polyendocrinopathy candidiasis ectodermal dystrophy [25], heterozygous CTLA4 mutations [26], immunodysregulation polyendocrinopathy enteropathy X-linked syndrome [27], and GATA2 mutation [28] may all present with AIH-like syndromes. Notably, all of these are associated with disruption of the regulatory CD4+ T cell pathway.

In addition, there is strong evidence of environmental influences on PBC [15], PSC [29] and AIH [17]. Similarly to their genetic profile, no currently identified environmental risk factor is sufficient to explain a large proportion of disease risk. This combination of an uncertain relative contribution to disease risk from multiple genetic and environmental variables – as well as probable interactions between these factors – marks these conditions as multifactorial, polygenetic and complex (Fig. 1) [30].

Unraveling the genetic contribution to disease risk is therefore a challenge. Prior to GWAS, a large number of candidate gene studies had been undertaken in hepatic autoimmune diseases [31–33]. However, these studies suffered from relatively weak statistical
power, a strong potential for type 1 statistical errors, lacked reproducibility and were criticized for their a priori approach based on an incomplete understanding of disease biology. Of the non-HLA candidate gene associations listed in the three pre-GWAS reviews cited above, none has subsequently shown significant association in subsequent GWAS.

4. The rationale behind GWAS

GWAS offers an approach that presupposes that there is a genetic component to the risk of developing a disease trait, but seeks evidence of association across the whole genome without assuming other prior knowledge. GWAS exploits the common genetic variability that exists between individuals. When considering a particular trait, GWAS works on the assumption that the risk of developing the trait of interest is at least partly derived from the cumulative effect of multiple variants, each of which confers low risk by itself. If two groups with and without a trait of interest are compared, certain variants may be disproportionately represented. GWAS works on the assumption that the risk of a particular trait, GWAS works on the assumption that the risk of developing the trait of interest is at least partly derived from the cumulative effect of multiple variants, each of which confers low risk by itself. If two groups with and without a trait of interest are compared, certain variants may be disproportionately represented. GWAS exploits the common genetic variability that exists between individuals.

GWAS has been made possible by leveraging advances in chip-based genotyping technology and incremental decreases in the cost of such techniques [34]. By generating data that are amenable to combination, the power of each individual GWAS may be increased by meta-analysis. This also permits combination of datasets to improve statistical power and facilitate replication analyses, especially where the same genotyping platforms or an overlapping selection of SNVs are used.

Considering fixed genetic information inherently focuses on the root rather than result of disease process. Potential gains from GWAS include improvements in predicting whether an individual will develop a disease, earlier prediction of disease severity or sub-phenotype, potential biomarker identification and improved diagnostics. Identifying specific genes and therefore protein targets guides the generation of pre-clinical models, informs the development of novel candidate drug treatments and may also reveal opportunities for drug re-purposing [11,35].

5. The process of GWAS

Conducting GWAS requires the assembly of cohorts of individuals with and without a trait of interest. Care must be taken to assess for the presence of over-lapping conditions, which may bias genetic risk associations and misdiagnoses, which may dilute risk associations. Relevant examples include the presence or absence of ulcerative colitis in PSC or in distinguishing Type 1 from Type 2 AIH [30,36]. Although, finding sufficient well-phenotyped cases in these relatively rare diseases can be challenging, over the last decade high-throughput genetic studies have transformed our understanding of the risk architecture of autoimmune conditions in general and hepatic autoimmune in particular (Table 2).

Once appropriate test and control cohorts have been assembled, each subject of both groups is genotyped for a wide array of candidate gene studies.

### Table 2

Genome-wide analyses in autoimmune liver disease.

<table>
<thead>
<tr>
<th>Study</th>
<th>Year</th>
<th>Patients</th>
<th>Controls</th>
<th>Replication cohort Pat/cont</th>
<th>Population</th>
<th>Loci implicated</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>PBC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hirschfeld et al. [111]</td>
<td>2009</td>
<td>536</td>
<td>1536</td>
<td>526/1206</td>
<td>White N American</td>
<td>HLA, IL12A, IL12RB2</td>
<td>Extension of [111] population</td>
</tr>
<tr>
<td>Hirschfeld et al.</td>
<td>2010</td>
<td>494</td>
<td>1502</td>
<td>857/3198</td>
<td>White N American</td>
<td>IRF-TNPO3, MMEL1, IRF3</td>
<td></td>
</tr>
<tr>
<td>Liu et al. [113]</td>
<td>2010</td>
<td>453</td>
<td>945</td>
<td>None</td>
<td>Italian and meta N American</td>
<td>SPB, IRF3, IRF5-TNPO3</td>
<td></td>
</tr>
<tr>
<td>Mellis et al. [114]</td>
<td>2011</td>
<td>1840</td>
<td>5163</td>
<td>620/2514</td>
<td>N European</td>
<td>DENND1B, STAT4, C4BD, NFKB, IL7R, CCR5, TNFRSF1A, RAD51L1, CLEC16A, MAP3K7P1, PLCL2, RPS6K4, TNFAIP3</td>
<td>in combination with Italian data [115]</td>
</tr>
<tr>
<td>Nakamura et al. [39]</td>
<td>2012</td>
<td>487</td>
<td>476</td>
<td>787/615</td>
<td>Japanese</td>
<td>HLA, TNFSF15, POU2AF1</td>
<td>Japanese Large meta-analysis</td>
</tr>
<tr>
<td>Cordell [98]</td>
<td>2015</td>
<td>2745</td>
<td>9802</td>
<td>3616/4261</td>
<td>N European, Italian, N American</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Karlson et al. [116]</td>
<td>2010</td>
<td>285</td>
<td>298</td>
<td>766/2935</td>
<td>Scandinavian, German</td>
<td>HLA</td>
<td>Fine-mapping study based on [117]</td>
</tr>
<tr>
<td>Melum et al. [117]</td>
<td>2011</td>
<td>715</td>
<td>2962</td>
<td>1025/2174</td>
<td>Scandinavian, German</td>
<td>HLA, BCL2L11, MST1</td>
<td></td>
</tr>
<tr>
<td>Foleva et al. [118]</td>
<td>2012</td>
<td>715</td>
<td>2962</td>
<td>1221/3508</td>
<td>Scandinavian and German discovery; Scandinavian, Central European and US replication</td>
<td>MMEL1-TNFRSF14</td>
<td></td>
</tr>
<tr>
<td>Srivastava et al. [119]</td>
<td>2012</td>
<td>992</td>
<td>5162</td>
<td>N/A</td>
<td>German, Scandinavian</td>
<td>MST1, IL2RA, GPR35, TCF4</td>
<td>Replication study Combined analysis with ulcerative cases</td>
</tr>
<tr>
<td>Ellinghaus et al. [120]</td>
<td>2013</td>
<td>392</td>
<td>2977</td>
<td>1012/2694</td>
<td>German, Scandinavian</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liu et al. [97]</td>
<td>2013</td>
<td>3789</td>
<td>25079</td>
<td>4343/1673</td>
<td>Pan-European, N America</td>
<td>HLA, MMEL1-TNFRSF14, CD28, MST1, IL12, IL21, BACH2, IL2RA, SIRK2, HDAC7, SH2B3, CD226, FUT2, SDM1G1</td>
<td></td>
</tr>
<tr>
<td>Type 1 AIH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Dutch (discovery); German/Swiss (replication)</td>
<td>HLA, SH2B3</td>
<td></td>
</tr>
</tbody>
</table>

Only loci associated at $p < 5 \times 10^{-8}$ are included with the exception of SH2B3 ($p = 7.7 \times 10^{-8}$).

* Total of all replication cohorts when several used. Where is not yet possible to distinguish between several candidate genes, a single example has been given.
individual single nucleotide variants (SNV). These SNV are pre-selected from those sufficiently prevalent in the general population, and sufficiently spaced across the genome, to power subsequent statistical association analysis. SNV selections are drawn from large haplotype mapping projects, particularly the Internationa HapMap project. In the case of GWAS, these traits are chosen from variations identified during sequencing efforts to dissect the genome and are spread across it. In more focused studies, a targeted selection of SNVs may be used. In the case of the Immunochip, SNVs implicated in immune pathways and other autoimmune diseases — including some rarer than those used in most GWAS—are used to allow finer-resolution analysis of loci of interest [37].

The analysis of large-scale GWAS-derived datasets enables identification of individual variants with estimates of the risk that they confer. Notably, far more common SNVs have been described than have been assessed in any one GWAS. Therefore over-representation of a given SNV in one cohort or another is likely to represent linkage disequilibrium rather than the causative mutation itself. Exploitation of linkage disequilibrium — the phenomenon by which genes located closer together on are more frequently inherited together — allows likely risk loci to be identified and therefore candidate genes within or near such loci inferred [38]. Thus, GWAS does not typically detect specific mutations and the resolution of a particular study may make it difficult to comment on the specific gene implicated. Several GWAS ‘hits’ in hepatic autoimmunity may still relate to one of several proposed genes.

The multiple comparisons made in GWAS make both statistical correction and the use of a conservative level of statistical significance mandatory. Typically, overall values of \( p < 5 \times 10^{-8} \) are taken as indicating significance in a genome-wide study. A key potential confounder in GWAS is the potential for false positives/type 1 errors because of the multiple comparisons inherent to its analysis. Thus, careful adjustment for the number of such comparisons is mandatory. Also mandatory is confirmation of associations by replication in separate validation cohorts.

An important component of GWAS is comparing the test cohort with a suitable control: i.e. one that would be expected to have a similar distribution of common SNVs other than those associated with the disease trait of interest. Typically this means a control selected from those sufficiently prevalent in the general population and geographically similar area. With the notable exception of the Japanese PBC GWAS [39], all of the studies performed in hepatic GWAS to date have concentrated on populations of predominantly white European origin. Two associations from the Japanese PBC cohort (PBL2AF1 and PTL1N1S51) are unique to that study. Also, recent work in a Chinese cohort failed to replicate an association in the IL-12 pathway that has proven reproducible in European populations [40].

Following the broad model described above, risk loci associated with PBC, PSC and AIH at genome-wide significance levels \( (p < 5 \times 10^{-8}) \) have been identified and these are summarized in Tables 3 and 4. This number will increase with larger, and higher resolution, studies e.g. there is an ongoing effort to perform a meta-analysis of PSC GWAS datasets. Notably, there is a high degree of overlap (or pleiotropy) between loci implicated in hepatic autoimmunity to date and those associated with other autoimmune conditions (Fig. 2).

6. Major observations in hepatic autoimmune resulting from GWAS

6.1. HLA associations

GWAS has given robust support to a number of associations with HLA haplotypes with each of the major diseases. Several of these associations were made prior to the GWAS era. HLA associations strongly implicate the adaptive immune system: the HLA region on chromosome 6 is highly variable encodes both the class I and II major histocompatibility complexes. Associations are summarized in Table 3. All three hepatic autoimmune diseases have associations with variants in both class I (encoded by HLA-A, HLA-B and HLA-C) and class II (HLA-D) major histocompatibility complexes, with more associations in HLA-D. HLA haplotype associations are seen in all major autoimmune diseases (Fig. 2).

SNVs affecting HLA can alter the MHC antigen presentation repertoire and have been associated with disease provoking responses in response to otherwise innocuous certain agents: for example HLA-A*02:05:01 and abacavir hypersensitivity [41]. In PBC much of the risk has been narrowed to HLA-DRB1 with much of these rest of the association signal attributed to linkage disequilibrium [42].

One theory that complements specific HLA associations with specific diseases alongside multiple shared other non-HLA variants is that HLA haplotypes may direct autoimmunity in an otherwise prone individual, perhaps after an environmental trigger. For

<table>
<thead>
<tr>
<th>Table 3</th>
<th>HLA associations of hepatic autoimmune diseases.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PBC</strong></td>
<td><strong>Risk conferring</strong></td>
</tr>
<tr>
<td>HLA-DQA1<em>04:01-DQB1</em>04:02-DRB1<em>08:01-DQB1</em>03:02</td>
<td>[58,96]</td>
</tr>
<tr>
<td>HLA-DRA1<em>04:04-DQB1</em>03:02</td>
<td>[58,96]</td>
</tr>
<tr>
<td>HLA-DRB1<em>14:01-DRB1</em>03:01</td>
<td>[42]</td>
</tr>
<tr>
<td>HLA-DQB1<em>08:03-DRB1</em>06:01</td>
<td>[64,69]</td>
</tr>
<tr>
<td>HLA-DQB1<em>04:05-DRB1</em>04:01</td>
<td>[69]</td>
</tr>
<tr>
<td><strong>PSC</strong></td>
<td><strong>Risk conferring</strong></td>
</tr>
<tr>
<td>HLA-B*08:01</td>
<td>[86], N. Am inc Afr-Am</td>
</tr>
<tr>
<td>HLA-DQA1<em>03:01-DQA1</em>05:01-DRB1*02:01</td>
<td>[121]</td>
</tr>
<tr>
<td>HLA-DQA1<em>13:01-DQA1</em>01:03:01-DRB1*06:03</td>
<td></td>
</tr>
<tr>
<td>HLA-DQB1<em>15:01-DQA1</em>01:02-DRB1*06:02</td>
<td></td>
</tr>
<tr>
<td>HLA-DRB1<em>01:01-DQA1</em>01:01</td>
<td></td>
</tr>
<tr>
<td>HLA-DQA4:01-03-DRB1<em>04:01-DQA1</em>03:02</td>
<td>[122]</td>
</tr>
<tr>
<td>HLA-DQB4<em>01:03-DRB1</em>07:01-DQA1<em>02:01-DRB1</em>03:03</td>
<td></td>
</tr>
<tr>
<td>HLA-DQA4<em>02:02-DRB1</em>01:11-DRB1<em>05:01-DRB1</em>03:01</td>
<td></td>
</tr>
<tr>
<td><strong>AIH (Type 1)</strong></td>
<td><strong>Risk conferring</strong></td>
</tr>
<tr>
<td>HLA-A1-HLA-B8-DRB3<em>01:01-DRB1</em>03:01-DQA1<em>05:01-DQB1</em>02:02:01</td>
<td>N Euro [123]</td>
</tr>
<tr>
<td>HLA-DQB4<em>04:04 and HLA-DRB4</em>04:05</td>
<td></td>
</tr>
<tr>
<td>HLA-DRB1*04:01</td>
<td></td>
</tr>
<tr>
<td>DRB1<em>13:01-01-DQB1</em>06</td>
<td></td>
</tr>
<tr>
<td>HLA-DRB1<em>01 and HLA-DRB1</em>14</td>
<td></td>
</tr>
<tr>
<td>HLA-DQA5<em>01:01-DRB1</em>15:01</td>
<td></td>
</tr>
<tr>
<td>HLA-DQB1<em>04-04:01-DQB1</em>03:01</td>
<td>N Euro [123]</td>
</tr>
<tr>
<td></td>
<td>Lat. Am. [125]</td>
</tr>
</tbody>
</table>
| [126]* Dense SNV analysis and subsequent conditional analysis of this HLA haplotype in the Italian PBC cohort has suggested that risk-conferring/protective effects are predominantly due to variants in HLA-DRB1 with associated linkage disequilibrium [42].
Table 4
Major non-HLA associations from GWAS in hepatic autoimmunity.

| Locus | PBC | PSC | AIH | Study | SNV | OR | P-value | Candidate gene(s) | Selected functional notes |
|-------|--|--|--|--|-----|--|--|---|---|---|
| 1p36 | x | [112] | rs3748816 | 1.33 | 3.15E-08 | MMEK1 |
| 1p36 | x | [97] | rs3748816 | 1.21 | 7.41E-12 | MMEK1, TNFSF14 |
| 1q31 | x | [114] | rs12134279 | 1.34 | 2.06E-14 | DENND1B |
| 2q12 | x | [98] | rs12712133 | 1.14 | 5.19E-09 | IL1R1, IL1R2, |
| 2q13 | x | [117] | rs6720394 | 1.6 | 4.10E-08 | BCL2L1 |
| 2q32 | x | [114] | rs10931468 | 1.5 | 2.35E-19 | STAT4, STAT1 |
| 2q33 | x | [97] | rs7426056 | 1.3 | 1.89E-20 | CD2D2 |
| 2q36 | x | [98] | rs4973341 | 1.22 | 2.54E-10 | CCL20 |
| 2q37 | x | [120] | rs4676410 | 1.38 | 2.43E-09 | |
| 2q12 | x | [39] | rs12372072 | 1.2 | 2.28E-08 | PLCL2 |
| 3p21 | x | [97] | rs1372072 | 1.2 | 2.28E-08 | |
| 3p24 | x | [114] | rs1724804 | 1.22 | 9.01E-12 | DGKQ |
| 4q24 | x | [114] | rs7650509 | 1.26 | 8.48E-14 | NFKB1 |
| 4q27 | x | [97] | rs13140464 | 1.3 | 8.87E-13 | IL2, IL21 |
| 5p13 | x | [96] | rs6871748 | 1.3 | 2.26E-13 | IL7R |
| 5q21 | x | [98] | rs526231 | 1.15 | 1.14E-08 | C5orf30 |
| 5q33 | x | [98] | rs52468890 | 1.15 | 1.06E-10 | IL12B, LOC285626 |
| 6q15 | x | [97] | rs56258221 | 1.23 | 8.36E-12 | BACH2 |
| 6q23 | x | [98] | rs6933404 | 1.18 | 1.27E-10 | OLIG3, TNFAIP3 |
| 7p14 | x | [114] | rs9749401 | 1.25 | 4.44E-08 | ELMO1 |
| 7q32 | x | [96] | rs35188261 | 1.52 | 6.52E-22 | IRF5 |
| 9q32 | x | [39] | rs49739462 | 1.57 | 1.85E-14 | TNFSF15 |
| 10p15 | x | [97] | rs1447359 | 1.24 | 8.19E-17 | IL2RA |
| 11q23 | x | [98] | rs49385354 | 1.38 | 3.27E-08 | POU2AF1 |
| 11q31 | x | [98] | rs538147 | 1.23 | 2.06E-10 | RPS6KA4 |
| 11q23 | x | [98] | rs7937682 | 1.17 | 3.17E-08 | SIK2 |
| 11q31 | x | [98] | rs80065107 | 1.39 | 7.20E-16 | CXCR5, DDX6 |
| 12p13 | x | [96] | rs1800693 | 1.27 | 1.18E-14 | TNFRSF1A, LTBR |
| 12q13 | x | [97] | rs11168249 | 1.15 | 5.49E-09 | HDACT7 |
| 12q24 | x | [97] | rs3184504 | 1.18 | 5.91E-11 | SH2B3, ATOX1 |
| 12q24 | x | [96] | rs10659379 | 1.2 | 2.87E-09 | SH2B3 |
| 13q14 | x | [96] | rs3862738 | 1.33 | 2.18E-08 | TNFSF11 |
| 14q24 | x | [96] | rs911263 | 1.26 | 9.05E-11 | RAD51B |
| 16p13 | x | [97] | rs12706715 | 1.29 | 2.19E-13 | CELT6A, SOC51 |
| 16q24 | x | [114] | rs11117432 | 1.31 | 4.66E-11 | IRF8 |
| 14q32 | x | [98] | rs8017161 | 1.22 | 2.61E-13 | TNFAIP2 |
| 17q12 | x | [96] | rs17564829 | 1.26 | 6.05E-14 | IKZF3 |
| 17q21 | x | [96] | rs17564829 | 1.25 | 2.15E-09 | MAPT |
| 18q21 | x | [120] | rs1452787 | 1.33 | 2.61E-08 | TCF4 |
| 18q22 | x | [97] | rs1788097 | 1.15 | 3.06E-08 | CD2D2 |

The transcription factor NFκB plays important roles in both the innate and adaptive immune system [133,134], influencing both B and T cell activation downstream of the antigen receptors and promoting inflammatory responses.

Important in lymphocyte development; loss of function mutations result in lymphopaenia and an autoimmune pre-disposition (product also known as CD127)

Regulatory protein with effects on apoptosis and key to antibody class switching [135]; represses effector lines of CD4+ T cells [136]

Transcription factor mediating effects of TLR ligation

Modulation of inflammatory cytokine production by murine macrophages [139]

CXCRI5: Major role in B cell and T follicular helper cell trafficking to germinal centres.

TNFRSF1A encodes a member of the tumor necrosis factor receptor family of receptors. It is predominantly expressed on antigen-presenting cells and represents a major receptor for tumor necrosis factor alpha (TNFα). Activation of this receptor can cause apoptosis through activation of NFκB and mutations leading to its constitutive activation are associated with periodic fever syndrome [140].

Histone deacetylase implicated in the suppression of myeloid lineage genes in lymphoid cells [141]

SH2B3: Also known as 'Lir'. Adaptor protein involved in multiple cell surface signaling pathways in both hematopoietic and non hematopoietic cells. Mutations are implicated in myeloproliferative conditions including malignancies [142].

AKA Rank-ligand or TNF-related activation-induced cytokine (TRANCE). TNFR ligand involved in control of osteoclast activity, and dendritic cell survival factor. Deficiency results in impaired lymphocyte differentiation [143].

Histone acetyltransferase implicated in the suppression of myeloid lineage genes in lymphoid cells [141]

SH2B3: Also known as 'Lir'. Adaptor protein involved in multiple cell surface signaling pathways in both hematopoietic and non hematopoietic cells. Mutations are implicated in myeloproliferative conditions including malignancies [142].

AKA Rank-ligand or TNF-related activation-induced cytokine (TRANCE). TNFR ligand involved in control of osteoclast activity, and dendritic cell survival factor. Deficiency results in impaired lymphocyte differentiation [143].

See Fig. 4

 sees Fig. 4

 Sees Fig. 4

 Sees Fig. 4

See Fig. 4

See Fig. 4

Sees Fig. 4

Sees Fig. 4

Sees Fig. 4

Sees Fig. 4

Sees Fig. 4

Sees Fig. 4

Sees Fig. 4

Sees Fig. 4

Sees Fig. 4

Sees Fig. 4

Sees Fig. 4

Sees Fig. 4

Sees Fig. 4

Sees Fig. 4

Sees Fig. 4
example, in the autoimmunity-prone NOD mouse typically develops insulinitis and diabetes, but altering its HLA alleles can redirect the autoimmune focus to thyroiditis [43]. Of note however, the haplotypes associated with each of the hepatic autoimmune diseases are variable between geographic and ethnic populations (Table 3). It remains unclear as to whether the overlapping gene variants contained within these haplotypes of the highly variable HLA region represent true divergent developments resulting in a risk for a similar phenotype, or whether the risk will in due course be narrowed down to very few SNVs as it has in rheumatoid arthritis [44].

6.2. Non-HLA associations

Multiple non-HLA associations have now been identified in both PBC and PSC by GWAS. These are summarized with brief notes as to the function of better-understood implicated genes in Table 4. Below, we consider associations forming patterns within pathways and of associations with genes of particular interest. Notably biological studies are needed to complete the understanding of proposed gene associations.

6.2.1. The IL-2 pathway in PSC

The combination of loci associated with PSC risk containing CD28, IL2 and IL2RA – which encodes the alpha subunit of the high affinity IL-2 receptor – suggest that the T lymphocyte focused IL-2 pathway has a major role. The multiple effects of IL-2 are summarized as Fig. 3 and the molecule has key roles in the development and maintenance of regulatory T cells but is also important in the adaptive effector response.

The IL2RA deficient mouse develops marked lymphadenopathy with ultimately fatal autoimmunity with lymphocytic infiltrates in skin, intestine, lung and liver in the context of a complete deficiency of regulatory T cells. The liver infiltrate is peri-portal but occurs without overt fibrosis and, owing to its development of anti-mitochondrial antibodies, the mouse has in fact been described as a model of PBC [45]. Indeed, a single case of a boy with a homozygous mutation in IL2RA developing a PBC-like disorder with Treg deficiency has been described [46]. By analogy to IL2RA mutations, mice deficient in the leucocyte anti-apoptotic protein Bcl2l11 also develop multi-system autoimmunity with lymphocytic infiltration and autoantibodies, but without a hepatic focus in a pattern that has been described as more like systemic lupus erythematosus [47].

6.2.2. The IL12-STAT4 pathway in PBC

PBC risk has been attributed to variants in a number of genes involved in the T cell activation pathway involving IL-12 and STAT4, which leads to Th1 T cell polarization (Fig. 4). IL-12 is a heterodimeric molecule made up of the two subunits, p35 and p40, encoded by the IL12A and IL12B genes respectively. The latter protein also heterodimerizes with IL-23p19 to form IL-23, a key signaling component in the Th-17 pathway. The IL-12 receptor is also encoded by two genes: IL12RB1, which is constitutively expressed, and IL12RB2 which is upregulated by interferon-γ (IFNγ) to act as a positive feedback loop in antigenic stimulation. STAT4 and tyrosine kinase 2 (TYK2) protein are important in both IL-12 and IL-23 receptor signaling. Each of IL12A, IL12RB2, STAT4 and TYK2 have risk associations with PBC. Deficiencies in IL12 and STAT4 confer susceptibility to intracellular pathogens such as tuberculosis [48,49], whilst STAT4 is required for developing autoimmune diabetes in a mouse model [50].

Using animal models of PBC, efforts have been made to see whether interruption of the IL-12 pathway ameliorates autoimmune liver disease. Results have been variable and all depended on models of PBC created by introducing profound defects in T cell regulatory pathways: deletion of the IL-12p40 subunit worsens peri-biliary inflammation and upregulates hepatic expression of gene products associated with fibrosis in IL2Rs deficient mice [51]; in dnTGFβRII mice, IL-12p40 deletion greatly reduced peri-biliary inflammation [52]; whereas IL-12p35 delayed but did not prevent disease in dnTGFβRII mice and was also associated with more marked fibrosis [53].

6.2.3. FUT2

The linkage of the gene FUT2 to PSC risk is of interest because it resonates with two other major potential etiological considerations in the disease; the gut microbiome and antigen presentation. FUT2 encodes fucosyl transferase 2 and is involved in the regulation of the expression of antigens including the ABO system and in the synthesis of the H antigen oligosaccharide. The latter acts as a binding moiety for a number of bacteria and viruses including the Norwalk virus and also a source of carbon for certain bacteria. In those homozygous for the disease-associated SNV demonstrate alterations in gut microbial community composition [54], FUT2 variants appear to affect carbohydrate metabolism in the gut, so affecting the microbiome. Such alterations, and FUT2 variations, have been associated with sub-clinical colitis in both mice and man [55]. FUT2 variants are also recognized as conferring risk in inflammatory bowel disease.

6.2.4. SH2B3 – Lnk

SH2B3 – also known as Lnk – is notable for being associated with all three of the hepatic autoimmune diseases. The association signal in the single AIH GWAS to date falls slightly short of the generally-accepted level for genome-wide significance, but it has been argued that the high likelihood of SH2B3’s involvement given its role in so many other autoimmune diseases means that it should be considered a significant association. SH2B3 encodes one a number of SH2B adaptor proteins and is within a widely shared autoimmune locus. It is involved in multiple growth factor and

<table>
<thead>
<tr>
<th>Locus</th>
<th>PBC</th>
<th>PSC</th>
<th>AIH</th>
<th>Study</th>
<th>SNV</th>
<th>OR</th>
<th>P-value</th>
<th>Candidate gene(s)</th>
<th>Selected functional notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>19p12</td>
<td>x</td>
<td></td>
<td></td>
<td>[96]</td>
<td>rs34536443</td>
<td>1.91</td>
<td>1.23E-12</td>
<td>TYK2</td>
<td>Signaling kinase in IL-12 and IL-23 receptor signaling. In systemic lupus erythematosus variants to influence IFNγ production [47].</td>
</tr>
<tr>
<td>1p31</td>
<td></td>
<td></td>
<td></td>
<td>[96]</td>
<td>rs72678531</td>
<td>1.61</td>
<td>2.47E-38</td>
<td>IL12RB2</td>
<td>See Fig. 4</td>
</tr>
<tr>
<td>19q13</td>
<td></td>
<td></td>
<td></td>
<td>[97]</td>
<td>rs60652743</td>
<td>1.25</td>
<td>6.51E-10</td>
<td>FUT2, PBK2, STRN4</td>
<td>FUT2: Antigen expression – see main text</td>
</tr>
<tr>
<td>19q13</td>
<td></td>
<td></td>
<td></td>
<td>[115]</td>
<td>rs3745516</td>
<td>1.46</td>
<td>7.97E-11</td>
<td>SPB</td>
<td>Transcription factor mediating B cell receptor signaling</td>
</tr>
<tr>
<td>21q22</td>
<td></td>
<td></td>
<td></td>
<td>[97]</td>
<td>rs2836883</td>
<td>1.28</td>
<td>3.19E-17</td>
<td>PSMG1</td>
<td>Embryonically lethal if functionally absent in mice; deficient hepatocytes demonstrated premature senescence [148]</td>
</tr>
<tr>
<td>22q13</td>
<td></td>
<td></td>
<td></td>
<td>[96]</td>
<td>rs2267407</td>
<td>1.29</td>
<td>1.29E-13</td>
<td>SYNGR1</td>
<td></td>
</tr>
</tbody>
</table>

Lowest p value shown where different studies have implicated the same variant. All candidate genes identified here have association signals with p < 5 × 10⁻⁸ with the exception of the SH2B3 signal in AIH (see text). Not all candidate genes represented for some loci.
cytokine signaling pathways and is a negative regulator of T cell activation, tumor necrosis factor and Janus kinase 2 and 3 (JAK2/3) signaling and is required for normal hematopoiesis. Mice deficient in SH2B3 have greater levels of activated T cells and a tendency to autoimmunity [56].

SH2B3 variants are also implicated in celiac disease. Investigation of the functional effect of SH2B3 genotype in response to lipopolysaccharide and muramyl dipeptide has demonstrated that the risk that carriers of the SH2B3 rs3184504*A risk allele — which is the same as that implicated in PSC and AIH — show stronger activation of the NOD2 pathogen recognition pathway [57]. This suggests that SH2B3 has a role in modulating the immune response to intestinal pathogens.

6.2.5. Epistasis

Epistasis is the process by which carrying a combination of two or more risk variants has a more than additive effect on total risk. One example of such interactions in PBC is the risk-conferring epistatic interaction between the 1p31 (IL12RB2) and 7q32 (IRF5) loci [58]. A potential risk-amplifying interaction between CTLA4 and TNFα variants has also been reported in the pre-GWAS era, but not replicated [59]. Numbers of epistatic gene–gene interactions are likely to grow as sample sizes increase and meta-analysis is carried out.

Fig. 2. Many genetic associations of PBC, PSC and AIH are shared with other autoimmune diseases. Circos plot [101] of gene variants associated with PBC, PSC and AIH alongside other autoimmune conditions in high-throughput genetics studies and their meta-analyses. Note that significant pleiotropy exists with some loci implicated in multiple conditions and variable HLA associations a universal feature. PBC = primary biliary cirrhosis; PSC = primary sclerosing cholangitis; AIH = autoimmune hepatitis; AA = alopecia areata; AS = ankylosing spondylitis; ATD = autoimmune thyroid disease; JIA = juvenile idiopathic arthritis; MG = myasthenia gravis; MS = multiple sclerosis; Narc = narcolepsy; Pso = psoriasis; RA = rheumatoid arthritis; ScL = scleroderma; Sj o = Sjögren's syndrome; SLE = systemic lupus erythematosus; T1DM = type 1 diabetes mellitus; UC = ulcerative colitis; Vit = vitiligo. Only validated associations at p < 5 × 10^-8 are included (with the exception of SH2B3 in AIH — see text); supporting citations are available at request.
The combination of TCR and CD28 signals upregulate IL-2 production in activated T cells, especially CD4⁺ T cells. T cell activation without co-stimulatory CD28 signal results in anergy with minimal IL-2 production. In general, IL-2 has pro-effector actions. However, in IL-2 deficiency or deficiency of IL-2 receptor alpha, there is widespread autoimmunity and lymphoproliferation ascribed to a failure to generate regulatory T cells. In developed immune systems however, exogenous IL-2 typically amplifies T cell responses. IL-2 receptor complex may exist as a dimer (from IL-2Rα and/or IL-2Rγ) or trimer (with IL-2Rβ). The addition of IL-2Rβ renders the receptor complex with much higher affinity to IL-2, and is particularly important for regulatory T cells. Cytokines secretion is shown as dashed arrows; positive effects are shown as solid arrows. Labels in red denote factors implicated by GWAS to date. Abbreviations: APC antigen-presenting cell; CD - cluster of differentiation; IL - interleukin; AICD activation-induced cell death; IL2Rα IL-2 receptor; IFNγ interferon gamma; TNFα tumor necrosis factor alpha; Treg regulatory T cell [102–105]. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

7. Translating GWAS findings to clinical cohorts

Translating findings from GWAS studies in general to changes of clinical care has proved challenging. Notable exceptions exist however, for example the utility of IL-12 and IL-23 blockade in clinical care has proved challenging. Notable exceptions exist.

7.1. HLA

There has been some success in predicting clinical characteristics from GWAS data, but prospective validation is required. For example, work arising form the Dutch AIH GWAS has suggested that HLA types influence clinical characteristics: HLA-DRB1*03:01 being associated with higher serum gamma immunoglobulin levels and HLA-DRB1*04:01 being associated with an older age of first presentation [62].

A number of pre-GWAS studies have related clinical characteristics to HLA haplotypes, although many have not been replicated. The well-established inter-population variability between HLA associations means that results must be applied between varying populations with caution. In PBC for example, HLA haplotypes correlate with anti-gp210 status in Japanese [63] and Chinese [64] patients. The presence of anti-gp210 antibodies is independently predictive of outcome. HLA-associated susceptibility to UC is different with and without the presence of concurrent PSC [65] and acute rejection after transplantation may also be associated with HLA-B*08, HLA-C*07 and HLA-DRB1*03 [66]. In AIH, HLA-DR4 may predict treatment failure [67]. Sequencing of DRB1*03-DRB1*04 amongst North American patients has shown a lysine-rich subset in which carriers with AIH more frequently develop cirrhosis and a requirement for liver transplantation [68]. Similarly, amongst Japanese with AIH, the DRB1*08:03-DQB1*06:01 haplotype was more frequent in patients who developed hepatic failure [69].

The mechanism by which given HLA variants may exert their effects in hepatic autoimmunity is unclear. However, in PBC recent important work has directly linked HLA-types to cellular phenotype. T cells with the risk-conferring HLA-DRB1*08:01 showed a high-affinity response to specific pyruvate dehydrogenase E2 subunit peptides; an analogous response was absent in cells expressing the protective HLA-DRB1*11:01 [70].

7.2. Non-HLA

The implication of the IL-2 pathway in PSC might logically suggest a role for immunosuppressants. However, no immunosuppressive regime, including those with particular activity against IL-2 has proven efficacy in PSC [71]. Similarly, despite a number of molecules implicated in activation of the adaptive immune system in PBC, no immunosuppressive therapy has been shown effective to date.

The alpha-subunit of the IL2 receptor complex (IL2RA; CD25) is also implicated by GWAS in PSC. This had lead to particular interest in regulatory T cells, which strongly express IL2Rα. Recent work attempting to address the role of regulatory T cells in PSC has
suggested a reduction in their number in PSC livers as compared with PBC [72]. Further, peripheral populations were reduced and this correlated with homozygosity for the risk IL2RA allele.

The proportion of liver-infiltrating lymphocytes expressing CD28 in PSC is reduced and these cells have a more activated phenotype and produce more of the effector cytokines tumor necrosis factor \( \alpha \) and interferon-\( \gamma \) than their CD28\( ^{+} \) counterparts [73]. Interestingly, the CD28-negative T cell compartment is also expanded in rheumatoid arthritis, in which the CD28 gene is similarly implicated by GWAS [74].

In follow-up studies of GWAS datasets, associations of STAT4 variations with ANA status have been identified in Japanese PBC patients [75] and TNFSF15 polymorphisms have been linked to functional changes in the gene product. Differences in mRNA and protein expression were also seen in both PBC and healthy subjects carrying this variant, although a link to outcome or phenotype remains unclear [76].

The risk association with CXCR5 is of particular interest because of CXCR5’s major role in lymphocyte trafficking, especially of B and T cells migrating to germinal centers. Subsequent to GWAS, two observations have strengthened the link to the clinical picture: firstly that numbers of CXCR5\( ^{+} \) T follicular helper cells are increased both peripherally and in PBC liver tissue [77,78] and secondly that the ligand for CXCR5, CXCL13, is upregulated in PBC as compared with PSC [79].

Fig. 4. Schematic representation of the IL-12 pathway, which is implicated in PBC by GWAS. Antigen activates APC through TLR. APC in turn produce IL-12 after phosphorylation of IRF5. Antigen is presented to CD4\( ^{+} \) T cells by HLA II with co-stimulation via CD80 and 86 to CD28. There is competitive inhibition of this co-stimulation by CTLA4. IL-12 activates a cascade of signaling factors including NFKB and STAT4 to promote the production of Th1-type cytokines including TNF\( \alpha \) and IFN\( \gamma \); the transcription factor IRF8 is involved. IL7R supports lymphocyte development. There is positive feedback from Th1 cytokines to APCs. Red text denotes confirmed risk associations with PBC. Arrows denote positive effects; barred lines denote negative effects. APC = antigen-presenting cell; CD = cluster of differentiation; CTLA4 = cytotoxic T lymphocyte antigen 4; HLA II = human leucocyte antigen class II; IFN-\( \gamma \) = interferon-\( \gamma \); IFN-R = interferon-\( \gamma \) receptor; IL-12 = interleukin-12; IL-12R(1/2) = IL-12 receptor 1 subunits 1 and 2; IL7R = interleukin-7 receptor; IRF5 and IRF8 = interferon response factors 5 and 8; JAK2 = Janus kinase 2; Lck = lymphocyte-specific protein tyrosine kinase; NFKB = nuclear factor kappa-light-chain-enhancer of activated B cells; PKC = protein kinase C; SOCS1 = suppressor of cytokine signaling 1; STAT4 = signal transducer and activator of transcription 4; TCR = T-cell receptor; TLR = Toll-like receptor; TNF\( \alpha \) = tumor necrosis factor alpha; TNF\( \alpha \)-induced protein 3; TCR = T-cell receptor; TLR = Toll-like receptor; TNFRSF1a = tumor necrosis factor receptor superfamily 1a; TNF-\( \alpha \) = tumor necrosis factor alpha; TYK2 = Tyrosine kinase 2 [106,107]. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
with other liver diseases [79].

A common feature of PBC, PSC and AIH is that they all recur after transplantation. This supports a key role for the immune system rather than an intrinsic biliary epithelial cell or hepatocyte defect. In PBC, data gathered as part of the UK-PBC cohort was re-analyzed according to individuals’ IL12A variant status and this has suggested a correlation between the AG variant and decreased time to diagnosis of PBC recurrence [80].

The IL12 pathway is also the target of the only GWAS-derived therapeutic intervention to date in hepatic autoimmunity. This study used ustekinumab, a monoclonal antibody that blocks IL-12 and IL-23 signaling. However, administration to PBC patients with established cholestatic, ursodeoxycholic-unresponsive disease did not significantly alter biochemical surrogates of disease activity [81]. However intriguingly in those patients in whom there was a fall in alkaline phosphatase noted, there was evidence of modulation of biological pathways (including Th17 lymphocytes) related to ustekinumab response.

Finally, work has been published attempting to calculate risk using aggregated GWAS data and applying these to cohorts. There has been some success, with a recent risk score generating a modest AuROC of 0.72. This is despite the same authors calculating a combinatorial disease risk from all extant PBC susceptibility alleles of only 5.3% of the observed heritable risk, suggesting the capacity for future refinement [82].

8. Deficiencies in GWAS to date

8.1. GWAS results do not account for all the observed risk

One reason for this is that these rare diseases are likely to have some of their risk conferred by rare genetic variants. The design of GWAS means that identified risk variants are, by definition, more frequent in the population than is the disease per se [83]. Further, rarer genes with smaller effect sizes require larger cohorts for association signals to reach the stringent levels of statistic significance: this effect has been especially apparent in more common autoimmune diseases. For example, there is an almost direct correlation between study size and the number of variants identified in inflammatory bowel disease [37] and studies of super-cohorts involving over 100,000 subjects have greatly clarified the genetic landscape of rheumatoid arthritis [84].

Importantly, it is apparent that GWAS will never account for all of the observed risk in complex diseases like those in hepatic autoimmunity: in PSC it has been commented that “more than 50% of the susceptibility to these diseases is still likely to be of an environmental origin.” [85]. Equally, in PBC, which has seen the greatest amount of GWAS investigation, calculation suggests that findings to date only account for a few percent of the observed heritability [82].

8.2. Genetic information cannot yet explain key aspects of disease phenotype

As yet, GWAS does not explain key phenotypic features of the hepatic autoimmune diseases. Examples of these include the marked differences in risk between genders, the co-incidence of other autoimmune phenomena, treatment response in PBC and AIH, and malignancy risk in PSC. Genetic risk also does not help explain the evolution of clinical phenotype that is particularly recognized in pediatric patients presenting with autoimmune hepatitis and sclerosing cholangitis.

8.3. GWAS does not reflect disease risk in varied ethnicities

As mentioned above, all major GWAS in hepatic autoimmunity to date have focused on populations of European Caucasian descent, with the exception of the Japanese PBC study [39]. The Japanese GWAS identified two associations not yet replicated in other populations. Similarly, a recent replication study from China in a Han population has suggested incomplete overlap with previous studies [40]. Observations like the lower rate of concurrent ulcerative colitis in African American PSC patients suggest genetic variation may be implicated [86]. The selection of SNVs assessed by Immunochip are also biased to white Caucasian populations [87]. It therefore remains possible that associations in those of non-European backgrounds will yield different associations.

8.4. Positive selection pressure for genes implicated by GWAS is not yet explained

Several associations uncovered in GWAS have putative links to resistance to microbial infection: IL-12/STAT4 pathway and intra-cellular/tubercular infection in PBC; FUT2 and Norwalk virus in PSC; SH2B3 and response via NOD2 in AIH, PSC and PBC [49,54,57]; However, for the most variants associated with hepatic autoimmune disease, no explanation exists for the selection force that explains their presence in the population. It is likely that responses to microbial agents will be implicated: most of human variation is yet explained

8.5. GWAS does not yet map to animal models nor therapies

As considered above, there is a current lack of either a direct animal model in which abnormalities in a gene identified by GWAS recapitulate the key phenotypic components of one of the major autoimmune hepatic disease. Equally, no targeted therapy based on GWAS findings has yet been successful.

8.6. GWAS may not capture epigenetic modifications, including heritable epigenetic traits

Epigenetic modifications are alterations to genes other than in their base-pair code that alter how, or whether, they are transcribed. They may be the result of environmental influences or a consequence of the activity of other genes. They may be inherited. However, epigenetic modifications are not captured by some current chip-based GWAS techniques nor sequencing technologies. Importantly, a high proportion of fine-mapped variants are seen in promoter regions or where regulatory proteins bind, and epigenetic variation has been identified in non-hepatic autoimmunity [90].

One epigenetic mechanism that has been proposed as explaining some of the gender differences in prevalence of PBC is biased X or Y chromosome inactivation. This process typically involves silencing by chromosome packaging in heterochromatin and has been reported as being biased in a number of autoimmune diseases [91,92]. However, it is notable that to date, no risk genes for hepatic autoimmune disease have been identified on either the X or Y chromosome.

In PBC further interest a role for epigenetic mechanisms has been given further support by two recent studies. First is the observation that the promoter region of the gene encoding CD40 ligand (CD40L) demonstrates reduced methylation in PBC patients as compared with controls. This is of particular interest given the key role CD40-CD40 ligand interactions play in B cells activation and because this observation correlates with serum IgM [93].
Second, recent reports have demonstrated differential DNA methylation of the promoter region of the gene CXCR3 in T lymphocytes from patients with PBC [94]. Demethylation of CXCR3 correlated with increased protein expression of CXCR3. Functionally, CXCR3 is present in many inflamed tissues and is associated with the generation of Th1 type responses and the subsequent migration of Th1 CD4+ T cells. Neither CXCR3 nor CD40L has yet been associated with risk for developing PBC.

9. The future of genetic investigation of hepatic autoimmune

Future investigation of the genetics of hepatic autoimmunity is likely to take advantage of the increasing efficiency and reducing cost of finer-resolution SNV chips and sparse or whole genome sequencing. Such “post-GWAS” strategies are starting to reveal new risk loci in other conditions [95]. Movses towards higher-resolution assessment of the genome that stop short of full sequencing may also yield extra information as suggested by ImmunoChip studies [96,97] (with their higher resolution coverage of targeted subsections of the genome) and recent work in other autoimmune conditions [90]. In addition, improved analyses of multiple SNVs associated with risk loci have improved the ability to identify specific risk-causing variants/mutations. Combining such information with knowledge of the epigenome of individual cell types has begun to direct researchers to highlight where individual variants may have particular effects [90]. For example, in PBC candidate SNVs are enriched in CD4+ T cells and B cells. Larger cohorts of patients will be valuable but may prove challenging to recruit and meta-analysis including smaller new cohorts as they become available is likely to continue to provide small incremental returns [98]. In addition, where patient — and importantly control — cohorts are well phenotyped there is the potential of linking genetic data to clinical outcomes and disease subtypes. Indeed, recent work has taken data on HLA haplotype associations from GWAS and associated them with presentation and outcome in AIH [62]. The three carefully phenotyped cohort projects known as UK-PBC, UK-PSC and UK-AIH are examples of current work that may provide insights in this direction.

Future genetic work will need to proceed hand-in-hand with closer examination of environmental influences, epigenetics and gene—environment interactions and will need to consider genomes other than in Caucasian populations.

10. Conclusion

In each of PBC, PSC and AIH, there is a clear genetic component to disease risk. Each also has a significant environmental input and it is uncertain as to the relative contribution of genes, environment and interactions within and between these influences. Until recently, although several risk-bearing HLA haplotypes in these diseases had been identified, these were insufficient to explain much disease heritability. GWAS and analogous focused chip-based studies have now greatly increased our knowledge of a number of variants associated with risk for PBC and PSC and begun to inform our knowledge in AIH. These variants have been mapped to a number of likely candidate genes. Of particular interest are links to the IL-12-STAT4 pathway in PBC and the CD28-IL2 pathway in PSC. However, despite the multiple genetic associations described by genetic studies to date, the observed genetic component to disease risk in hepatic autoimmunity is far from explained. Possible reasons for this include relatively small samples sizes, inherent deficiencies in GWAS methodology in detecting rare variants, and the uncertain extent of environmental and epigenetic influences. Further, GWAS does not yet explain a number of key phenotypic features of these diseases, nor have results yet mapped to definite clinical insights as they have in other autoimmune diseases. These problems are likely to be addressed by larger, sequencing-based or higher-resolution SNV-based studies alongside investigation of gene—environment interactions. Such efforts are starting to gain leverage in more common diseases. Finally, it is a great pleasure to contribute this paper to the special issue devoted to Diego and Giorgina Vergani, both of whom have excelled in every aspect of clinical medicine, from patient care to teaching and, of course, research. This paper is part of this dedicated issue which is a component of the yearly efforts of the Journal of Autoimmunity to recognize truly outstanding figures in autoimmunology. Previous recipients have included Abul Abbas, Michael Sela, Ruth Arnon and Ian Mackay [99,100]. Giorgina and Diego are fitting to include in this group.

Conflict of interest

None declared.

Acknowledgments

GJW is a UK Medical Research Council Clinical Research Training Fellow. GMH is a co-investigator for UK-PBC (www.uk-pbc.com) supported by a Stratified Medicine Award from the UK Medical Research Council, and Principal Investigator for UK-PSC (www.uk-psc.com), an NIH Rare Disease Translational Collaboration.

The views expressed are those of the author(s) and not necessarily those of the National Health System, the National Institute for Health Research, or the Department of Health.

References


