Journal of Autoimmunity xxx (2015) 1–12

Contents lists available at ScienceDirect

Journal of Autoimmunity

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

Review article

The genetics of human autoimmune disease: A perspective on progress in the field and future directions

Michael F. Seldin [a, b, *]

[a] Department of Biochemistry and Molecular Medicine, University of California, Davis, Tupper Hall Room 4453, Davis, CA 95616, USA
[b] Division of Rheumatology and Allergy, Department of Medicine, University of California, Davis, Tupper Hall Room 4453, Davis, CA 95616, USA

ARTICLE INFO

Article history:
Received 23 June 2015
Accepted 23 August 2015
Available online xxx

Keywords:
Genetics
Autoimmune diseases
HLA
Susceptibility loci
Functional variants
Immunoregulation

ABSTRACT

Progress in defining the genetics of autoimmune disease has been dramatically enhanced by large scale genetic studies. Genome-wide approaches, examining hundreds or for some diseases thousands of cases and controls, have been implemented using high throughput genotyping and appropriate algorithms to provide a wealth of data over the last decade. These studies have identified hundreds of non-HLA loci as well as further defining HLA variations that predispose to different autoimmune diseases. These studies to identify genetic risk loci are also complemented by progress in gene expression studies including definition of expression quantitative trait loci (eQTL), various alterations in chromatin structure including histone marks, DNase I sensitivity, repressed chromatin regions as well as transcript factor binding sites. Integration of this information can partially explain why particular variations can alter proclivity to autoimmune phenotypes. Despite our incomplete knowledge base with only partial definition of hereditary factors and possible functional connections, this progress has and will continue to facilitate a better understanding of critical pathways and critical changes in immunoregulation. Advances in defining and understanding functional variants potentially can lead to both novel therapeutics and personalized medicine in which therapeutic approaches are chosen based on particular molecular phenotypes and genomic alterations.

Contents

* Department of Biochemistry and Molecular Medicine, University of California, Davis, Tupper Hall Room 4453, Davis, CA 95616, USA
E-mail address: mfseldin@ucdavis.edu.

http://dx.doi.org/10.1016/j.jaut.2015.08.015
0896-8411/

1. Introduction

Our understanding of which genes predispose to different autoimmune diseases has expanded rapidly over the last decade. This progress has been mostly due to genome-wide association studies (GWAS) and the development of various technical and analytic tools. However, despite this progress less than half of the heritability of most autoimmune diseases can be explained and nearly half of this identified genetic risk is due to variations within HLA. The actual functional variants that underlie statistically significant associations are with some notable exceptions are still largely unknown. In the following perspective, I will review some of the more salient advances in the field, provide examples to illustrate specific points, indicate where knowledge is sparse, and discuss the potential for future advances that I believe could further define the pathogenesis and perhaps enable application to diagnoses and therapy. More detailed aspects of the genetics for a variety of autoimmune diseases is presented by experts in the field in other sections of this special issue of the journal. A general paradigm for GWAS and sequence variant studies is shown in Fig. 1 and discussed in subsequent sections.

1.1. Heritability

Epidemiological studies of most autoimmune diseases including rheumatoid arthritis (RA), systemic lupus erythematosus (SLE), type 1 diabetes (T1D), multiple sclerosis (MS), and primary biliary cirrhosis (PBC) show that there is strong heritability. These include studies showing increased concordance in monozygotic compared to dizygotic twin as well as studies showing increased risk to siblings of proband cases compared to the general population.

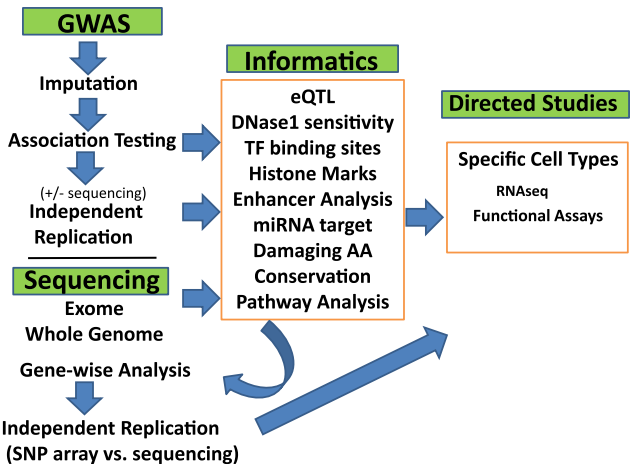

Fig. 1. Diagram of general scheme for genetic studies of complex autoimmune diseases. GWAS studies can greatly benefit from imputation and replication studies for loci suggested in the discovery phase. For some studies, replication is limited to those loci (genes) that are also part of pathways for genes previously identified as significantly associated with the disease. This is also proposed for sequencing studies to identify less common variants in which power issues may be partially addressed by limiting replication analyses based on prior pathway information.

Although most of these studies are not truly population based and may have biased results, there are some caveats worth noting. First, some autoimmune diseases have much higher sibling relative risk rates than other diseases (e.g. SLE [1,2], T1D [3], celiac disease [4], and PBC [5,6] compared to others e.g. RA [2], and MS [7]). Second, although concordance of disease is much higher in monozygotic compared with dizygotic twins for many autoimmune diseases, the overall monozygotic concordance of disease is usually substantially less than 50% [8]. This indicates that stochastic factors including environmental variables are a strong component and although genetics can be very useful in identifying important factors in etiopathogenesis it can only partially predict phenotype. Although some specific environmental factors have been identified (e.g. smoking and rheumatoid arthritis [9,10]) it is also possible that most of the incomplete concordance is simply chance or indefinable events.

1.2. General considerations for identifying genetic loci for autoimmune disease susceptibility

The major advance in identifying genetic loci that predispose to autoimmune diseases has been GWAS. Although some non-major histocompatibility (HLA in humans) loci were identified prior to GWAS using linkage or candidate gene studies, and other methodologies including admixture mapping have also enabled identification of a modicum of risk loci, the exponential increase in loci (over 200 for some autoimmune diseases) has been the direct result of GWAS. The basis of GWAS is the technology enabling efficient and accurate genotyping of single base polymorphisms (SNPs) and large collaborative studies such as HapMap [11,12] defining large numbers (hundreds of thousands) of SNPs in different populations. The success of GWAS is in large part due to practical advantage in conducting case/control design, namely the ability to recruit large numbers of cases and population controls as opposed to the difficulty in recruiting families: power for any association study is largely based on numbers. A critical aspect for these studies has been the ability to adequately control for population substructure differences using statistical methodology. Most commonly this is done by logistic regression using relevant principal components defined by principal component analyses or similar methods [13–15]. In some studies only continental population differences are accounted for, but for the most part type 2 errors (false positives) due to unrecognized stratification differences in case and control populations have been minimized. In fact, many studies have used publically available control genotypes rather than specific matched collections of controls. It is also worth noting that it may be possible to increase power (decrease Type 1 errors, false negatives) to ascertain risk variants by limiting studies to more homogenous populations and additional considerations of population substructure is discussed in subsequent sections (see sections 3.3 and 4). However, GWAS is largely applicable to those loci that fulfill the common variant (>0.05 minor allele frequency) common disease hypothesis since this methodology relies on linkage disequilibrium (LD) between the marker (SNP) detected and the actual disease causing variant(s). The commonly used genotyping platforms (Illumina and Affymetrix chip arrays) and

Please cite this article in press as: M.F. Seldin, The genetics of human autoimmune disease: A perspective on progress in the field and future directions, Journal of Autoimmunity (2015), http://dx.doi.org/10.1016/j.jaut.2015.08.015

clustering algorithms are also most accurate for minor allele frequencies (MAFs) > 0.05, although there have been some improvements in software to enable higher accuracy genotyping.

Another important aspect of current GWAS and other association studies is the ability to accurately impute much of the sequence difference that is not directly genotyped using current genotyping platforms. Imputation is based on sophisticated algorithms that enable under certain conditions the accurate estimation of the sequence variants between genotyped markers based on reference genomes that contain the missing information [16–19]. It relies on the shared loci (almost all of the SNPs in commonly used genotyping platforms are shared with the reference genomes) and the patterns of variation (http://www.1000genomes.org). With the completion of phase 3 of the 1000 genome sequencing study there is now sequence data for multiple populations that can be used to inform imputation(http://www.1000genomes.org/) [20]. This information can suggest candidate non-synonymous damaging variants and in others provides candidate non-coding variants that might be suggestive based on expression quantitative traits (eQTL), binding sites for transcription factors, DNase sensitivity or other features (e.g. histone marks) that have been elucidated in particular tissues or cell lines (discussed further in Section 6). Importantly, imputation methods can also be used to determine HLA classical determinants and amino acids.

1.3. Phenotype

A critical aspect for any association study is defining the phenotype of cases. Theoretically, the better the definition of phenotype the greater the likelihood that type 1 errors will be minimized since inclusion of unaffected individuals or those with a different autoimmune disease may increase heterogeneity and dilute the risk of variants for a particular phenotype. However, there may be a practical trade-off between including cases that may not have all the information (e.g. particular autoantibodies or even age of onset) that could increase the power to detect loci due to greater sample size as compared to more rigid criteria. For some autoimmune diseases particular features clearly increase the relative risk of many of the susceptibility loci. A strong example is for SLE where cases with anti-double strand (DS) DNA show substantially higher odds ratios for several lupus susceptibility alleles than those that are anti-DS DNA negative [21]. Similarly, some RA studies show stronger associations when the cases are limited to those with anti-cyclic citrullinated peptide antibodies (anti-CCP) [22]. However, in other studies when cases are limited to anti-CCP positive RA there are fewer susceptibility loci associated with disease, probably mostly due to decreased power as a function of a smaller sample size [23]. Similarly, if for example an SLE study was restricted to only those that manifested the same 4 out of 11 American College of Rheumatology criteria it is likely that many of the identified loci would have been missed due to weaker signals from decreased sample sizes.

The decision on phenotypic definition is analogous to the classic fight between the lumpers and the splitters in the definition of mammalian species [24] and was advanced many years ago with respect to the nosology of genetic disease [25]. However, for some autoimmune diseases the definition of “subtypes” may be crucial. In collaborative work by this author, the study of myasthenia gravis, it was critical to divide the disease by age-of-onset and the lack of thymomas. Here, there are major differences in which *HLA* genes are important in susceptibility when early onset (<45 years of age) compared to late onset (>50 years of age) are compared ([26,27] and Seldin, Gregersen, Hammarstrom, unpublished data). In addition, the highest risk non-*HLA* gene, *TNIP1*, observed in early onset myasthenia gravis (EOMG) [26] has no effect in late onset myasthenia gravis (LOMG) (Seldin, Gregersen, Hammarstrom, unpublished data).

As the field progresses there are likely to be more studies that examine endophenotypes including the most prominent disease manifestations, disease severity and morbidity. In SLE there is a study suggesting specific loci that predispose to nephropathy [28]. Such studies could also include those individuals that have very poor outcomes or more severe phenotypes.

2. The major histocompatibility complex region

With rare exception, genes in the human major histocompatibility complex (MHC), HLA, are the strongest risk genes for each autoimmune disease. Several general points are notable: 1) often there is more than one association signal from the MHC region with the second and even third signals sometimes stronger than any of the individual signals from loci outside this region; 2) recent studies suggest that most of the strongest signals are from classical HLA determinants (not from single SNPs); 3) in some studies epistasis between HLA determinants has been shown or suggested and 4) different HLA specificities are associated with different autoimmune diseases.

For most autoimmune diseases where pathology is characterized/driven by autoantibodies the strongest associations are with MHC class II genes (HLA D region antigens) (e.g. SLE, RA, T1D, and celiac disease [29–31]). For those that are not characterized by autoantibodies (e.g. ankylosing spondylitis, psoriasis, and Behcet's) most are strongly associated with class I genes (*HLA-A*, *HLA-B* and *HLA-C*) [32–34]. However, this is not always the case. The major exception for autoantibody associated disease is the early onset form of myasthenia gravis (EOMG), a disease characterized by anti-acetylcholine receptor antibodies. Here the association is clearly much stronger with the class I *HLA-B*08* than the *DRB1* determinant (*DRB1*03*) that is in linkage disequilibrium with *HLA-B*08*. For diseases not characterized by autoantibodies, converse examples are the finding that a class II gene (*DRB1*11*) has been strongly associated with ulcerative colitis and Crohn's disease [29]. Recently specific amino acid residues in DQA1 were found to be strongly associated with idiopathic achalasia, an enigmatic disease that is now thought to be an autoimmune process [35]. In this context, it may be useful to note that there is evidence for potential role for CD8 cells (that recognize antigen in the context of class I) in other processes than cell mediated cytotoxicity: CD8 cells producing IFN-γ are critical to inducing class switching in B cells or helper function in particular circumstances [36,37].

Although the specific associations are diverse, overall, these studies are highly suggestive that specific classical determinants or specific amino acids are associated primarily with specific diseases. Thus, the predominant association for RA have the “shared epitope” in which DRB1 determinants that share particular amino acid residues in the third hypervariable region explain the strongest component of the *HLA* association with this disease [38] but the same amino acids are not very strongly associated with most other autoimmune diseases. In contrast to the strong RA association with the shared epitope (mostly particular *DRB1*04* determinants), the strongest MS association is with *DRB1*1501*, SLE with *DRB1*0301*, and PBC with *DRB1*08* [29,39,40]. These associations are often stronger than specific MHC SNPs associated in GWAS and studies using conditional analyses have provided additional support for these specific genes and amino acid polymorphisms for some of these diseases. The overall genetics of *HLA*, even for the predominant *HLA* association is much more complicated at least for some autoimmune diseases including T1D. Extensive studies of T1D associated haplotypes show data consistent with a complex relationships among *DRB1*, *DQA1* and *DQB1* alleles [30]. Here specific

Please cite this article in press as: M.F. Seldin, The genetics of human autoimmune disease: A perspective on progress in the field and future directions, Journal of Autoimmunity (2015), http://dx.doi.org/10.1016/j.jaut.2015.08.015

combinations show very strong susceptibility or protective effects depending both on haplotypic combinations, and particular determinants can in some cases be part of both (conferring increased odds ratios in some contexts but decreased in others).

It is further notable that the D region amino acids residues that have been implicated (e.g. specific amino acids in DRB1 in RA and ulcerative colitis, and DQA1 in achalasia) are critical for either binding antigen or the peptide exchange chaperone HLA-DM in the context of antigen binding [35,41,42]. These findings further suggest that unknown and elusive antigens are driving these autoimmune diseases, similar to that involved in the pathogenesis of celiac disease where specific DQ amino acids participate in binding gluten peptides [43] or microbes in the *B*27* associated seronegative spondyloarthropathies such as ankylosing spondylitis and Reiter's syndrome [44–46]. Other postulates on the functional relationship of *HLA* class II genes have been advanced including molecular mimicry and in some studies have supported an alternative postulate that in RA the shared epitope amino acids bind and trigger signal transduction via cell surface calreticulin (CRT) an innate immune receptor [47].

As mentioned above, multiple different MHC associations can be observed in many autoimmune diseases including many secondary loci. These include both class 1 and *DPB1* determinants [40] as well as class III for some autoimmune diseases [30]. For RA, a study has attributed all MHC risk to 5 different amino acids including 3 DRB1 amino acids, and one HLA-B amino acid and one DPB1 amino acid [42]. It is also worth noting that the effect-size of the MHC associations varies greatly. For EOMG (*HLA-B*08*), celiac disease (*DRB1*03*) and ankylosing spondylitis (*HLA-B28*) and certain T1D haplotypes the odds ratio are >5. For most other autoimmune diseases the odds ratio is ~2.0. An odds ratio >1.6 is rare for non-MHC region associations. Thus, the especially strong associations of HLA determinants suggest that recognition of particular antigens is the likely driver of the initiating event(s) in autoimmune disease.

3. Non-*HLA* susceptibility loci

Several hundred non-*HLA* loci have been identified that contribute to the association of one or more autoimmune diseases. Most are relatively high frequency perhaps as a reflection of the sensitivity of the methods to detect these loci (i.e. GWAS as discussed above). The vast majority of these loci are located either within or close to genes that participate in immune system response. These include genes that are components of antigen processing, presentation, recognition, differentiation, signaling and/or some form of immune-regulation including particular cytokines, interleukins or interferons. The genes are often primarily expressed in T cells, B cells or dendritic cells. Most have relatively low odds ratios (<1.2) although some have odds ratios >1.5. Most of the SNP variants that show association are non-coding SNPs, many are located between genes, and a few are in genomic regions that are devoid of genes in what is sometimes termed as gene deserts. A few such as the *PTPN22* association are nonsynonymous amino acid differences (e.g. PTPN22 R620W) with functional effects [48–53]. Fine mapping, sequencing, expression studies and a modicum of functional tests have in some cases identified strong candidate variants and potential mechanisms but these are the exception. For some, even with strong associations, it is not clear whether the closest gene is necessarily the gene involved in the functional variation. A possible example of this is the association of SLE (and several other autoimmune diseases including RA) with a number of SNPs within the *STAT4* gene [54]. The strongest associated SNPs are all located in an intronic region of *STAT4* and thus far studies have not linked these SNPs to any function. Since *STAT1* is located only ~10 kb downstream from these SNPs it is certainly possible that the association is due to some regulatory function with respect to *STAT1* rather than *STAT4*. This is a reasonable hypothesis even though the association is much weaker with *STAT1*, since recombination distance (not directly related to physical distance) does not obviate a possible functional link which, in this example, can still be a cis regulatory mechanism. Although Occam's razor favors that most of the associations close to or within a reasonable candidate gene are due to a functional connection to this gene, it is possible that some of the identified variants regulate genes in disparate locations (i.e. trans-regulation could account for some of the associations).

For some genes multiple haplotypes are associated with disease including specific *TNFAIP3* and *TNIP1* haplotypes with SLE [55,56]. It is also possible that many associations with particular loci actually reflect multiple different functional lower frequency variants that are on the same haplotype. However, efforts to find such rare or uncommon variants within the high frequency haplotypes have thus far not proven to be generally successful and my personal viewpoint is that most of these associations are actually due to common variants, albeit not necessarily the sentinel variant identified in a GWAS study. Although fine mapping and sequencing of genomic regions is sometimes undertaken to further define the genomic variation that explain the association of common variants, much/most of the information on this variation can now probably be obtained using the imputation methods described above. The potential value of sequencing efforts or dense genotyping is further discussed in Section 3.3 and may yet be important for ascertaining rare variants that may also underlie susceptibility to autoimmune diseases.

3.1. Overlap and differences in non-HLA genetics of different autoimmune diseases

(For this section, I will confine the discussion to studies in European populations since there are substantial differences among different ancestries that will be discussed in a subsequent section.) Although some non-MHC loci appear unique or relatively unique to specific autoimmune diseases such as the *SpB* transcription factor for PBC [57], in general there are many non-MHC loci that are shared between different autoimmune diseases [54,58]. Although many of the loci overlap there may be substantial differences in effect size. For example among the genetic loci implicated for both SLE and RA are associations with *STAT4* and *PTPN22*. However, the odds ratio is substantially higher for *STAT4* in SLE than in RA [54] and the opposite appears to be the case for *PTPN22* [59,60]. *PTPN22* is also a shared susceptibility gene for many autoimmune diseases. This includes a high odds ratio (>2) in T1D [48], and lower odds ratios although still over 1.5 in several other diseases including EOMG [26] but only a minimal effect or no effect in PBC, celiac disease and MS. Interestingly, although the implicated (and in this case a nonsynonymous amino acid variant) also appears to be associated with Crohn's disease for this disease it has a protective rather than disease promoting effect [61]. There are multiple other examples of differences in apparently shared genetic loci. For EOMG the highest odds ratio for non-HLA genes is observed for *TNIP1* and although this gene is also implicated in SLE the odds ratio is much smaller and there are apparently two haplotypes (rather than one in EOMG) that are implicated [26,56]. *IRF5* has a high odds ratio in SLE but moderate odds ratio in PBC [57,62,63] and a low odds ratio RA [58]. *ITGAM* is another strong (high odds ratio) SLE susceptibility locus that is not associated with RA or other autoimmune diseases (including MS, T1D, and celiac disease) except systemic sclerosis where it has only a modest effect size [64,65].

Often for a given autoimmune diseases when sample sizes are increased additional loci are identified that are shared with other autoimmune diseases but many of these have very marginal effect

Please cite this article in press as: M.F. Seldin, The genetics of human autoimmune disease: A perspective on progress in the field and future directions, Journal of Autoimmunity (2015), http://dx.doi.org/10.1016/j.jaut.2015.08.015

sizes and odds ratios <1.1. Some examples of differences for gene associated loci with relatively high odds ratios in at least one autoimmune disease but show different effects with other autoimmune diseases is depicted in Fig. 2. Thus, for non-HLA variants there appears to be a complex pattern of gene effect sizes that is not readily discernible. Why are the odds ratios of these shared susceptibility genes so different between the different autoimmune diseases? Although some differences are due to variability/noise that can plague studies with relatively low sample sizes, it appears that many/most of these differences are real. Going forward it is important to determine how these differences relate to the interplay between the genetic risk factors and how these differential effects can alter the predisposition to different autoimmune diseases.

3.2. Epistasis between non-HLA variants and HLA

There are very few examples of epistasis between different susceptibility variants in any autoimmune disease. Perhaps the strongest examples are those between the non-*HLA* susceptibility gene *ERAP1* and *HLA* in three diseases: ankylosing spondylitis (*HLA-B*27*) [66]; psoriasis (*HLA-C*06*) [67], and Behcet's disease (*HLA-B*51*) [68]. *ERAP1* codes an aminopeptidase that functions to trim peptides for loading onto MHC class I determinants [69]. For these diseases the *ERAP1* variants are only important in the context of these particular HLA determinants and *ERAP1* variants have little or no effect in those patients that do not have these HLA determinants. Interestingly, while the same *ERAP1* variants show the same effect in ankylosing spondylitis and psoriasis, the opposite effect in seen in Behcet's disease (c) suggesting that different requirements for the *ERAP1* endoplasmic reticulum processing of the specific protein antigens that are part of triggering the immune responses that are critical to pathogenesis of these diseases.

3.3. Rare/uncommon variants

Only a small number of rare (minor allele frequency <0.01%) or uncommon (<0.05%) variants have been identified for common complex autoimmune diseases. Some of those identified for common complex autoimmune disease have been based on examining candidate genes that were known to be causal for rare Mendelian diseases with features of autoimmunity. This includes *TREX1* that causes autoimmune polyglandular syndrome type 1 [70]. For this gene, rare variants with high effect size has been identified in studies of SLE [71,72]. A modicum of other rare variants have been identified for autoimmune diseases based on either inclusion in the IMMUNOCHIP array (dense SNP variants for regions based mostly on genes identified by GWAS) including several for celiac disease [73]. However, in general these studies have been hampered by limited database of known variants (early versions of 1000 genome and other sequence information) and it is possible that more success will be realized in the near future for autoimmune diseases. Recent success identifying rare variants using exon chips has been reported for type 2 diabetes, a common non-autoimmune disease [74,75] and there are multiple ongoing studies using exon chips to try to uncover rare/uncommon variants in autoimmune diseases. A limited number of studies have also examined specific immune system genes by sequencing and some larger studies have used pooling approaches to look for whole exome variants. These and a single study that included a systematic sequencing of >800 immune related genes in psoriasis cases and controls have thus far met with limited success [76].

The overall status of efforts to identify rare variants in autoimmune disease has recently been reviewed [77]. As pointed out by these authors it is too early to make conclusions as to the role of rare variants. Systematic sequencing of exomes is limited and whole genome sequencing of large numbers of cases for specific autoimmune disease is just now on the horizon. Part of the difficulty in these studies is that the power to detect statistically significant associations for rare variants requires either prodigiously large numbers of samples (100s of thousands) or some method to limit the variants under consideration (i.e. those more likely to be functional) and/or applying some type of burden or collapsing algorithm to examine composite effects of multiple rare variants together. The numerous methods and algorithms for these analyses include several software packages that include different types of burden and collapsing tests (Variable Threshold, Madson Browning, KBAC, SKAT, SKAT0, VAAST) that use different algorithms to group together sequence variants to test for evidence of association [78–83]. Some of these methods have been assimilated in a variety of software packages many of which enable incorporation of

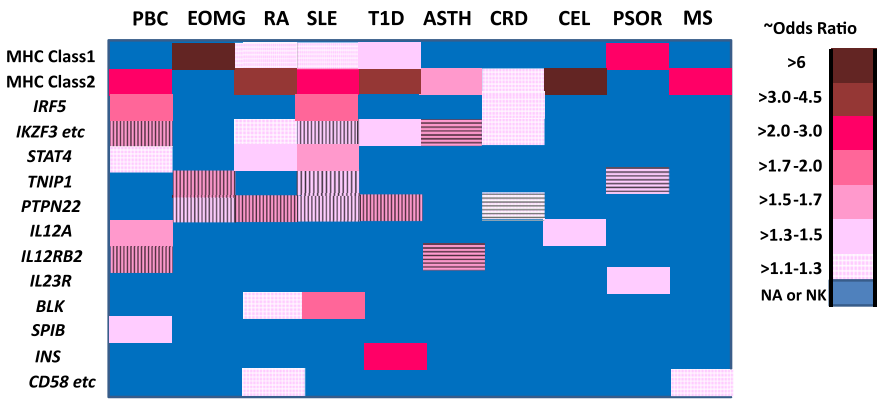

Fig. 2. Comparison of effect size for gene associated SNPs in multiple different autoimmune diseases. Disease abbreviations that are not clear from the main text include: asthma (ASTH), Crohn's disease (CRD), celiac disease (CEL) and psoriasis (PSOR). Asthma is included for comparison of autoimmune conditions with another immunologically mediated disease. The approximate odds ratios (ORs) are indicated by the color code legend and for each the OR is indicated in the positive direction (regardless of minor allele frequency). Horizontal or vertical lines are shown over the color coded ORs when the opposite SNP/haplotype is associated with different diseases (e.g. for the *IKZF3* etc. in which the asthma association is clearly the opposite of those for PBC and SLE. The ORs are derived from studies of European ancestry subject sets and are preferentially chosen from larger studies or meta-analyses [26,48,54,56,57,59–61,94,142,148,163–173]. For some of the genes for which the legend key indicates no association or not known (NA or NK) there may be limited data that suggests possible association that does not meet the standard criterion ($p < 5 \times 10^{-8}$). Where different SNPs show disparate ORs the association the highest OR is shown. The gene list was selected based on inclusion of most genes with at least moderate ORs (OR > 1.1) in at least one of the autoimmune diseases and is in part biased to emphasize the differences in effect sizes. *IKZF3* etc represents a gene cluster that includes *IKZF3*, *ORMDL3*, *ZPBP2*, and *GSDMB*. *CD58* etc. represents a gene cluster including *CD58*, *CD2*, and *IGSF2*. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Please cite this article in press as: M.F. Seldin, The genetics of human autoimmune disease: A perspective on progress in the field and future directions, Journal of Autoimmunity (2015), http://dx.doi.org/10.1016/j.jaut.2015.08.015

covariates including population substructure (http://varianttools.sourceforge.net/, http://csg.sph.umich.edu/kang/epacts/). Recent simulation studies suggest that SKAT0 and a generalized version of the C-alpha test may be more powerful than the other methods [84]. For exonic variants, algorithms such as SIFT, Polyphen and MutationTaster are popularly used to identify more likely causal variants and other algorithms use sequence conservation [85–87]. However, these are imperfect and some known causal variants for at least Mendelian diseases would be missed by these approaches. Compilations of functional and potentially functional coding variants can be found using different software [e.g. Annovar (http://annovar.openbioinformatics.org/en/latest/) [88]] and Ensemble (http://uswest.ensembl.org/Homo_sapiens/Info/Index)] [89].

For non-coding variation the situation is even more difficult although there are several search engines to look for critical variations in regulatory sequences including micro-RNAs, transcription binding sites, DNase1 hypersensitivity and expression quantitative trait loci (eQTL) [88–92]. (These issues are discussed further in Section 6.)

4. Ancestry makes a difference

There are substantial differences in the loci defined for autoimmune diseases in disparate population groups. These differences are most apparent comparing results from one continental population to another continental population. The largest amount available data are from studies of European populations and East Asian populations [93–96]. The differences are partially explained by the frequency of particular variants in these different populations. Thus, the *PTPN22* R602W variation that appears to account for most of the association of this gene with multiple different autoimmune diseases in European populations is virtually absent in other ethnicities including East Asian population groups [97,98]. A converse example is the strong association of *PADI4* with RA in East Asian populations that apparently plays a much smaller role in European populations consistent with the near absence of the strongest associated SNPs [94,99]. Examples showing comparisons of associations and odds ratios in European, Japanese and Chinese RA cohorts are shown in Fig. 3.

It should also be noted that when the frequency of an allelic variant is substantially lower in one population group, that even if the odds ratios are in reality similar, the sample sizes might need to be very large to detect statistically significant association for this population group. Thus, there is an association of *PADI4* and RA in European derived populations when the sample size is very large. However, the odds ratios are small (OR = ~1.1) compared to that observed in Japanese RA (1.36) and is apparently due to the absence of the more strongly associated haplotype [94,99].

There are also differences within continents. For example, the common variants of *PADI4* that are associated with RA in Japanese are not associated in Chinese cohorts [96]. Within Europe there is a substantial difference in frequency in the afore-noted *PTPN22* variant with a notable difference between northern and southern population groups: highest frequency in Finland (0.155) [100] and lowest in Italy (0.021) [48]. This emphasizes the importance of controlling for population substructure and using Cochran–Mantel–Haenszell corrections to account for the differences in allele frequencies when meta-analyses are performed in disparate population groups (reviewed in Ref. [15]).

It is also worth briefly noting that the striking geographic distribution of *PTPN22* suggests that it may have been under natural selection. In fact, other evidence suggests this is the case and suggest that many of the genes associated with SLE have natural selection signals [101]. Other autoimmune susceptibility genes also show evidence for natural selection. In celiac disease a non-

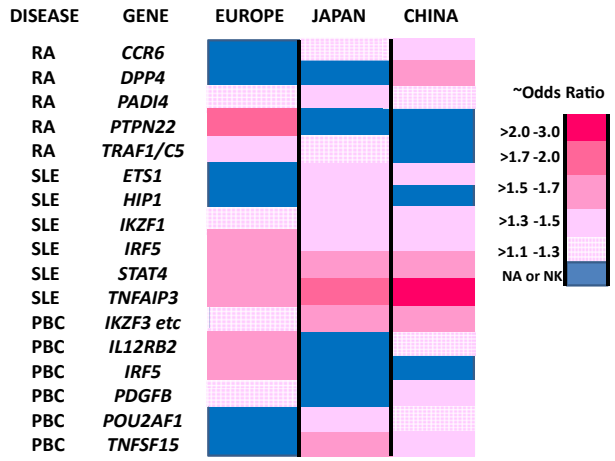

Fig. 3. Comparison of effect sizes for gene polymorphsims in three different population groups in three different autoimmune diseases. The approximate odds ratios (ORs) are indicated by the color code legend and for each the OR is indicated in the positive direction (regardless of minor allele frequency). The ORs were preferentially chosen from larger studies or meta-analyses ([57,94,96,98,99,148,170,171,174–179]. The genes selected were chosen on the bases of showing an OR of >1.3 in at least one of the populations chosen. For some of the genes for which the legend key indicates no association or not known (NA or NK) there may be limited data that suggests possible association that does not meet the standard genome-wide criteria for statistical significance. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

synonymous variant in *SH2B3* that shows a strong signal for positive selection in Europe and suggestive data that it might trace to the Justinan plague in the Eastern Roman Empire 541-542 AD [102]. Thus, some of the proclivity to autoimmunity may be the flip side of variants that are advantageous in the context of particular infectious agents and in part may underlie a portion of the differences in susceptibility gene variant frequencies between different population groups.

A difficulty in comparing studies in different ancestries is that for most susceptibility loci the actual causal variants are not known. Differences between populations can be the result of different haplotype structure rather than the absence or lack of effect of a particular causal variant (if unknown) or alternatively different causal variants of the same gene in different population groups. This is particularly a problem when candidate polymorphisms are selected based on a different population group. It can be addressed, at least with respect to common variant hypothesis, if sufficient SNPs marking any common haplotype in the study population are included in the analyses. Some have advocated studying multiple populations to enable finer mapping of causal variants. However, this pre-supposes that the same causal variants are operative in different population groups. Meta-analyses using trans-ethnic population samples have uncovered some novel variants in for example RA [94], presumably due to increased sample sizes but this does not negate the observation of differences in the strength of association in different population groups.

It is worth emphasizing that differences in disease phenotypes are also associated with ancestry. There are examples of both differences in the frequency of a particular autoimmune disease as well as differences in detected susceptibility associated variants. While some differences in disease frequency may be due to differences in environmental factors including presumed exposure to different infectious agents others are clearly due to genetics. One of the best examples is the low frequency of T1D in Native American population groups that has been attributed to the near absence of particular *HLA* determinants. More recently, in mixed population groups such as in the United States, ancestry informative markers

Please cite this article in press as: M.F. Seldin, The genetics of human autoimmune disease: A perspective on progress in the field and future directions, Journal of Autoimmunity (2015), http://dx.doi.org/10.1016/j.jaut.2015.08.015

(AIMs) or sufficient numbers of unselected independent markers (low intermarker LD) have been used to examine ancestry relationships between and within continents as well as admixture and autoimmune disease. Evidence for increased frequency of SLE has been associated with both West African ancestry and Native American ancestry [103–105]. Several studies of SLE have used this approach to examine whether endophenotypes of this disease are more likely to be associated with particular ancestry [106–109]. Some of these studies suggest substantial differences based on ancestry that cannot be accounted for by environmental covariates, albeit these covariates are imperfectly measured. Future studies may define the underlying variations that presumably can explain why particular ancestries predispose to particular disease manifestations (e.g. mucocutaneous manifestations were higher in northern compared with southern European ancestry and the opposite was found for anti-ds DNA in SLE) [106,107].

5. Missing heritability

For most autoimmune diseases, estimates suggest that over 50% of the genetic loci contributing to heritability are not yet elucidated. Many explanations that are not mutually exclusive are possible to account for this missing heritability including: 1) very large numbers of common low risk variants that can only be defined with huge numbers of cases and controls (e.g. sample sizes >>100,000); 2) large numbers of rare variants that have not been detected due to insufficient sequencing, sample size and appropriate models; 3) copy number repeats and/or small insertions or deletions that have not been adequately examined; 4) better definition of disease phenotypes (i.e. reduce disease heterogeneity) is necessary to better identify loci; 5) epistasis between two or more loci that may be difficult to ascertain; and 6) epistasis between different environmental factors and genetic loci (that can only be identified in the context of the environmental factors). As large scale sequencing becomes increasingly practical and perhaps applied systematically to large populations many of these issues should be addressable. In addition, some studies have suggested that a sequence driven search might identify additional genetic factors. This approach, PheWas [110] relies on looking at medical records and laboratory studies to see if particular variants are associated with any of the phenotypes. For example if there is a large database of sequence available it is possible that particular variants might be associated with a specific autoantibody. It is doubtful that ICD codes that are generally used in such efforts would be sufficiently informative to detect many novel associations although the method has shown some promise [110,111]. Alternatively, it might be possible to selectively study individuals who have sequence variants that are predicted to have a strong functional effect.

6. Towards functional studies: coding variants, gene expression, and epigenetics

Relatively few non-synonymous variants have been identified in non-MHC genetic loci contributing to the susceptibility to autoimmune diseases. These include *PTPN22*, *TNIP1*, *ITGAM*, and *SH2B3* [26,48,59,102,112]. Although Koch's postulates cannot be tested, for some variants there is compelling evidence that the variant changes a functional aspect(s) of the immune response. For *PTPN22* the R620W variant has been extensively studied and clearly has a major effect on increasing T cell and B cell expansion, dendritic cell hyper-responsiveness and B cell activation on engagement of the antigen receptor [49–53]. However, even here there is still substantial uncertainty with regards the actual mechanism although a gain of function appears to be most likely and there is evidence that the effect is likely to be on peripheral negative selection of T cells and an independent effect on B cells [113,114].

Most of the confirmed loci that contribute to the genetics of autoimmune disease are non-coding. To connect these to actual mechanisms is difficult and requires extensive circumstantial evidence. In contrast to Mendelian developmental defects that are usually coding, and where model organisms can frequently be used to easily provide evidence for functional defects, defining the physiologic role for non-coding variants in a complex disease may require multiple lines of evidence. As an initial step, many studies have examined gene expression and aspects of changes in DNA (methylation, histone marks) in various cell types. These studies can be either gene specific or genome-wide, and in general include: 1) studies comparing cells from cases compared to controls; 2) normal cells or tissues comparing those with a particular variant(s) to those without. Public databases can provide a valuable tool as will be discussed below.

Genome-wide expression studies have been performed in a variety of cells and tissues. Several approaches are possible. Many studies have used unseparated peripheral blood mononuclear studies may be difficult to interpret even when differential white blood cell counts are used in efforts to adjust based on cell type. Although some findings such as the type 1 (alpha) interferon signal in SLE (comparing cases with controls) are robust [115–119], it is likely that determining most of the functional variations will require finer examination of particular cell types and perhaps in the context of particular activating stimulation. Other studies have examined sorted cell populations and although these may be partially affected by cell separation procedures they are more likely to provide relevant information. For RA, synovial cells have also been studied [120–123], but for most diseases it is difficult to study regional lymphoid or parenchymal tissue that might enrich for particularly relevant subsets of cells.

There are also methodological issues that lead to substantial difficulty in obtaining accurate scans of gene expression. The use of chip gene arrays is likely to be replaced by RNA sequencing studies that may also have the potential to ascertain differences in splice form variants [124,125] and preferential allele expression when coupled with genomic sequence. Nevertheless, many of these studies have provided some clues to the pathways that are involved in the pathogenesis of individual autoimmune diseases. However, connecting the patterns of gene expression differences to the constellation of genetic factors that are defined by association studies (i.e. hereditary factors) is a critical step that may be hard to bridge.

One of the bridging approaches to look at the functional significance of sequence variants is to utilize eQTL. A number of studies have performed quantitative tests in cells or cell lines derived from different tissues to find polymorphisms (usually SNPs) that are associated with measurements of expression (eQTL). Other studies have examined methylation, DNase1 hypersensitivity and various histone modifications. These studies as well as eQTL can potentially identify specific variants that include those that are also identified by various association studies (GWAS or rare variant analyses) or are in strong linkage disequilibrium with these variants. Clues are often available in silico from public data bases of the larger projects, most notably the integrated encyclopedia of DNA elements in the human genome (Encode) project [126] and compilations of eQTL data. The Encyclopedia of DNA elements (Encode) data compiled and available via the UCSC Genome Browser provide data from multiple different cell lines and a number of tissues. This data provides Dnase sensitivity, information from sequences derived from chromatin immunoprecipitation (ChIP-seq), histone marks and transcription binding sites from a number of cell lines and tissues. A caveat for this approach is that for some genetic variants the right cell population at the right activation state may

Please cite this article in press as: M.F. Seldin, The genetics of human autoimmune disease: A perspective on progress in the field and future directions, Journal of Autoimmunity (2015), http://dx.doi.org/10.1016/j.jaut.2015.08.015

not be represented in the Encode or other available studies. Thus although this is a good starting point it may have both false negatives and even false positives.

Other tools also include Genevar that can be used to evaluate the measured effect of variants on DNA methylation (http://www.sanger.ac.uk/resources/software/genevar/) [127,128] and eQTLs (http://www.bios.unc.edu/research/genomic_software/seeQTL/) [129]. The University of Chicago eQTL browser (http://eqtl.uchicago.edu/cgi-bin/gbrowse/eqtl/) also provides another tool to identify expression quantitative trait loci (eQTLs) amongst candidate variants. It is worth noting that a good method for searching eQTL, and transcription factor binding sites including the Encode data is the RegulomeDB [90].

Other potentially useful approaches to finding putative important non-coding variations include searching for conserved motifs (e.g. SinBaD [92]) and using software that can ascertain whether variants affect splicing (e.g. Ensemble [89]). In addition, another resource, PolymiRTS [91], can be used to predict if 3′UTR variants are or are likely to alter micro-RNA (miR) target sites that may also be important in regulating genes or whether miR themselves have variants. Variants in other non-coding RNAs including long non-coding RNAs (lncRNAs) are also possible functional candidates as has been recently discussed [130,131].

It is also worth commenting that a recent study has developed and used a fine-mapping algorithm, Probabilistic Identification of Causal SNPs (PICs) to identify candidate functional variants at risk loci for autoimmune diseases (http://www.broadinstitute.org/pubs/finemapping/?q=pics) [132]. Together with data from mapping RNA and chromatin in both resting and stimulated CD4(+) T-cells, regulatory T cells, CD8(+) T cells, B cells, and monocytes this study provided data to suggest that most putative causal variants (~60%) map to immune-cell enhancers. However, only 10–20% directly altered recognizable TF binding motifs. Thus, much work remains to further understand the regulatory mechanisms. This reviewer would caution that although the PICs algorithm is useful especially in the context of meta-analyses, when primary data is available the use of conditional analyses can be extremely valuable in evaluating the potential importance of particular SNPs in accounting for association signals.

Another approach to further defining function is the use of animal models and in particular those in the mouse. Many studies have used mouse knockouts to help understand what genes do. This has evolved to using other manipulations in which specific base pairs are changed to study a particular variant (knockin mice). A mouse homologue of the *PTPN22* with a variant analogous to R620W variant (PEP R619W) has shown immune cell differences with respect to T cell expansion and transitional, germinal center, and age-related B cell expansion [113]. These mice developed autoantibodies and features of systemic autoimmunity. In addition the defective B cell selection and B lineage–restricted variant expression and was sufficient to promote autoimmunity. However, for studying human gene interactions the differences between mouse and human genes may make such knockin studies difficult or impossible for some/many variants.

Evolving methods to introduce human genes more efficiently into mouse stem cells could in the future enable more physiologic studies of putative human functional variants. Mice with the entire human immunoglobulin gene region have been generated [133] and introduction of multiple human genes in a particular pathway is not beyond the realm of feasibility. Of course such manipulations would have to include specific variants of these genes to mimic conditions similar to those in humans with particular autoimmune diseases. In addition, it is not clear that appropriate stimuli could be used to initiate a particular autoimmune phenotype that could depend on some specific antigenic challenge.

Finally, it should also be noted that there are also epigenetic changes and RNA editing variations that are not due to hereditary factors that may be important in the immunopathogenesis of some autoimmune diseases. Studies of identical twins concordant or discordant for particular autoimmune disease may help identify some of these non-hereditary changes that are part of the immunopathegenesis of autoimmune disease. RNA sequencing studies have the potential to define RNA editing variations but this may depend on identifying the right cell subtype in which critical (phenotype altering) somatic variations have occurred.

6.1. Pathways

A variety of databases can be queried to ascertain whether loci identified for an autoimmune disease are likely by some statistical measure to be members of particular biologic pathways.

In general, these approaches use independent signals (no LD) from association results, often from meta-analyses, to query membership in gene sets from a variety of databases often including BioCarta (http://www.biocarta.com/genes/index.asp), KEGG [134], Pathway Interaction Database (PID) (http://pid.nci.nih.gov/userguide/introduction.shtml), Gene Ontology [135] and Reactome [136]. An example would be using the MSigDB (http://www.broadinstitute.org/gsea/msigdb/index.jsp) curation of some of these gene sets. A subsequent step uses a program (examples include using the i-GSEA4GWAS web server [137]) to assess statistical significance. Different software including Paris [138,139], and Inrich [140] can be used to identified genes within some distance of the independent signals (e.g. 20 kb of the SNP or a border defined by a 2 log fall off from the peak association) and represent each gene by the strongest association. The gene sets (“pathway categories”) can then assessed for enrichment of signals using a permutation procedure that can also correct for bias from variations in gene size and gene set size. Often, a false discovery rate (FDR) is used to correct for multiple testing based on the distributions of enrichment scores generated by the permutations. Sometimes, the concordance of results between different methods is used as another way of filtering “significant” results. For autoimmune diseases, these studies are usually performed with and without the HLA signal to determine whether the pathways are dependent on HLA, since HLA is part of numerous gene pathways. A study of MS used a protein interaction data base to limit pathway analyses to only gene sets with evidence of physical interaction using a protein-interaction-network-based pathway analysis [141], although it is not clear if this approach is substantially better than those described above.

Such studies have been performed in many autoimmune diseases and including RA [142–144], SLE [144], MS [141], T1D [145], Celiac [146], PBC [147,148] and Crohn's disease [149]. These and other studies have highlighted particular pathways of which, not surprisingly, many are shared among different autoimmune diseases. Several of these studies have led to finding additional susceptibility loci by focusing additional genotyping on candidate genes which failed to reach statistical significance in previous studies but were “overrepresented” in pathways.

A caveat of these studies is of course that it is unknown how well the gene sets actually correspond to physiologic pathways. The improved curating and knowledge of these gene sets is ongoing that should facilitate these analyses. Future considerations might include ways of categorizing specific components of these pathways that currently are not distinguished including whether there is up or down regulation/signaling predicted for the different variants and perhaps further the ability to define diseases based on systems biology approaches.

Please cite this article in press as: M.F. Seldin, The genetics of human autoimmune disease: A perspective on progress in the field and future directions, Journal of Autoimmunity (2015), http://dx.doi.org/10.1016/j.jaut.2015.08.015

6.2. Implications for diagnosis and therapy

Theoretically, genetic risk factors can distinguish who is at high risk for a particular autoimmune disease. Studies of several diseases including RA, MS and PBC have used various genetic risk models to ask whether the combination of loci defined by GWAS studies can identify individuals at high risk for MS, RA, and PBC [150–152]. These studies have shown some promise, however, it is clear that our current state of knowledge of the genetics for most of these autoimmune diseases is still insufficient to meaningfully assist in diagnostic criteria. Furthermore, it is not clear whether incomplete penetrance will severely limit the usefulness of applying genetic information for complex diseases as opposed to Mendelian disorders. It is also possible that the genetic information could meaningfully complement other diagnostic information including serology, gene expression profiles and perhaps environmental risk factors. Theoretically, such information might also assist in determining who is more likely to have disease progression.

Another arena that will clearly receive more attention is studies to evaluate whether particular therapeutic agents are more likely to be effective in the context of different arrays of risk alleles. For SLE several directed studies and a GWAS have examined whether genotypes could in retrospective analyses account for responders or non-responders to TNF blockade. Although several loci have been suggested by such studies, none have replicated across studies where examined and at this time none have been confirmed [153–158]. Sample size issues have limited the power of such studies. Other efforts (e.g. in RA and PBC) have suggested using gene pathway information to identify potential therapeutic agents for autoimmune disease [94,148].

The identification of more risk loci and definition of actual causal variants together with the anticipated large scale implementation of individual sequencing and integration with medical records may enable the large amount of data necessary for evaluating whether more optimal therapeutic decisions can be developed based on genetics. However, actionable genetics for complex diseases such as most autoimmune disease is not yet on the horizon as compared with the beginning of implementation of such decisions for pediatric Mendelian conditions [159–162].

7. Final comments: what to make of it all

Some in the scientific community have questioned whether the myriad of loci identified in GWAS studies has been useful. While GWAS has not solved the problem of what causes autoimmunity it is my contention that it has had an important impact on the field and the identification of these loci will continue to be one of the major drivers in furthering our understanding of what causes these phenotypes. Like most investigative results, more questions are raised from these studies. What are the real functional variants? How do they really alter responses? In which cell populations and at what stage do the genetic variations matter? Why do some variants have a much larger impact in one autoimmune disease than another? These will not be easily solvable but as reviewed here, there is definite progress.

The extensive variation in the human genome has partly been shaped by natural selection including many variants that are theorized to have favored survival to various infections. In depth studies that define functional variation of the loci that predispose to or accelerate autoimmunity may also further our understanding of immune responses to infectious agents. Likewise, ongoing studies directed at defining host responses to infectious agents may also provide needed clues for further unraveling the enigmatic autoimmune diseases.

As discussed in this perspective, the increased understanding of the genetics of autoimmune disease may have an impact in early diagnoses and/or in a decision making tree for therapeutic intervention. At the least, our increased knowledge of the dictionary of relevant gene variations and pathways should focus functional studies. These may also lead to the development of novel therapeutics. Ongoing functional studies will undoubtedly, although perhaps slowly, increase our knowledge of not only pathogenesis but normal biology and immunoregulation.

Conflict of interest

The authors have no conflict of interest to declare.

Acknowledgments

This work was supported by R01DK091823.

References

[1] I.T. Harley, K.M. Kaufman, C.D. Langefeld, J.B. Harley, J.A. Kelly, Genetic susceptibility to SLE: new insights from fine mapping and genome-wide association studies, Nat. Rev. Genet. 10 (2009) 285–290.
[2] K. Sundquist, J.C. Martineus, X. Li, K. Hemminki, J. Sundquist, Concordant and discordant associations between rheumatoid arthritis, systemic lupus erythematosus and ankylosing spondylitis based on all hospitalizations in Sweden between 1973 and 2004, Rheumatol. Oxf. 47 (2008) 1199–1202.
[3] K.L. Mehers, K.M. Gillespie, The genetic basis for type 1 diabetes, Br. Med. Bull. 88 (2008) 115–129.
[4] I. Korponay-Szabo, J. Kovacs, M. Lorincz, E. Torok, G. Goracz, Families with multiple cases of gluten-sensitive enteropathy, Z. fur Gastroenterol. 36 (1998) 553–558.
[5] C. Selmi, P. Invernizzi, E.B. Keefe, R.L. Coppel, M. Podda, L. Rossaro, et al., Epidemiology and pathogenesis of primary biliary cirrhosis, J. Clin. Gastroenterol. 38 (2004) 264–271.
[6] D.E. Jones, F.E. Watt, J.V. Metcalf, M.F. Bassendine, O.F. James, Familial primary biliary cirrhosis reassessed: a geographically-based population study, J. Hepatol. 30 (1999) 402–407.
[7] N.M. Nielsen, T. Westergaard, K. Rostgaard, M. Frisch, H. Hjalgrim, J. Wohlfahrt, et al., Familial risk of multiple sclerosis: a nationwide cohort study, Am. J. Epidemiol. 162 (2005) 774–778.
[8] D.P. Bogdanos, D.S. Smyk, E.I. Rigopoulou, M.G. Mytilinaiou, M.A. Heneghan, C. Selmi, et al., Twin studies in autoimmune disease: genetics, gender and environment, J. Autoimmun. 38 (2012) J156–J169.
[9] L. Padyukov, C. Silva, P. Stolt, L. Alfredsson, L. Klareskog, A gene-environment interaction between smoking and shared epitope genes in HLA-DR provides a high risk of seropositive rheumatoid arthritis, Arth. Rheum. 50 (2004) 3085–3092.
[10] M. Heliovaara, K. Aho, A. Aromaa, P. Knekt, A. Reunanen, Smoking and risk of rheumatoid arthritis, J. Rheumatol. 20 (1993) 1830–1835.
[11] G.A. Thorisson, A.V. Smith, L. Krishnan, L.D. Stein, The International HapMap project web site, Genome Res. 15 (2005) 1592–1593.
[12] The International HapMap project, Nature 426 (2003) 789–796.
[13] A.L. Price, N.J. Patterson, R.M. Plenge, M.E. Weinblatt, N.A. Shadick, D. Reich, Principal components analysis corrects for stratification in genome-wide association studies, Nat. Genet. 38 (2006) 904–909.
[14] A.L. Price, N.A. Zaitlen, D. Reich, N. Patterson, New approaches to population stratification in genome-wide association studies, Nat. Rev. Genet. 11 (2010) 459–463.
[15] C. Tian, P.K. Gregersen, M.F. Seldin, Accounting for ancestry: population substructure and genome-wide association studies, Hum. Mol. Genet. 17 (2008) R143–R150.
[16] J. Marchini, B. Howie, Genotype imputation for genome-wide association studies, Nat. Rev. Genet. 11 (2010) 499–511.
[17] L. Huang, Y. Li, A.B. Singleton, J.A. Hardy, G. Abecasis, N.A. Rosenberg, et al., Genotype-imputation accuracy across worldwide human populations, Am. J. Hum. Genet. 84 (2009) 235–250.
[18] B.N. Howie, P. Donnelly, J. Marchini, A flexible and accurate genotype imputation method for the next generation of genome-wide association studies, PLoS Genet. 5 (2009) e1000529.
[19] B.L. Browning, S.R. Browning, A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals, Am. J. Hum. Genet. 84 (2009) 210–223.
[20] G.R. Abecasis, A. Auton, L.D. Brooks, M.A. DePristo, R.M. Durbin, R.E. Handsaker, et al., An integrated map of genetic variation from 1,092 human genomes, Nature 491 (2012) 56–65.
[21] S.A. Chung, K.E. Taylor, R.R. Graham, J. Nititham, A.T. Lee, W.A. Ortmann, et al., Differential genetic associations for systemic lupus erythematosus based on anti-dsDNA autoantibody production, PLoS Genet. 7 (2011) e1001323.
[22] S. Viatte, D. Plant, J. Bowes, M. Lunt, S. Eyre, A. Barton, et al., Genetic markers

of rheumatoid arthritis susceptibility in anti-citrullinated peptide antibody negative patients, Ann. Rheum. Dis. 71 (2012) 1984–1990.
[23] J. Cui, K.E. Taylor, Y.C. Lee, H. Kallberg, M.E. Weinblatt, J.S. Coblyn, et al., The influence of polygenic risk scores on heritability of anti-CCP level in RA, Genes Immun. 15 (2014) 107–114.
[24] G. Simpson, The Principles of Classification and a Classification of Mammals, Bull. AMNH 85 (1945) 23.
[25] V.A. McKusick, On lumpers and splitters, or the nosology of genetic disease, Perspect. Biol. Med. 12 (1969) 298–312.
[26] P.K. Gregersen, R. Kosoy, A.T. Lee, J. Lamb, J. Sussman, D. McKee, et al., Risk for myasthenia gravis maps to a (151) Pro–>Ala change in TNIP1 and to human leukocyte antigen-B*08, Ann. Neurol. 72 (2012) 927–935.
[27] A.H. Maniaol, A. Elsais, A.R. Lorentzen, J.F. Owe, M.K. Viken, H. Saether, et al., Late onset myasthenia gravis is associated with HLA DRB1*15:01 in the Norwegian population, PLoS One 7 (2012) e36603.
[28] S.A. Chung, E.E. Brown, A.H. Williams, P.S. Ramos, C.C. Berthier, T. Bhangale, et al., Lupus nephritis susceptibility loci in women with systemic lupus erythematosus, J. Am. Soc. Nephrol. 25 (2014) 2859–2870.
[29] J.D. Rioux, P. Goyette, T.J. Vyse, L. Hammarstrom, M.M. Fernando, T. Green, et al., Mapping of multiple susceptibility variants within the MHC region for 7 immune-mediated diseases, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 18680–18685.
[30] J.A. Noble, H.A. Erlich, Genetics of type 1 diabetes, Cold Spring Harb. Perspect. Med. 2 (2012) a007732.
[31] K. Kaukinen, J. Partanen, M. Maki, P. Collin, HLA-DQ typing in the diagnosis of celiac disease, Am. J. Gastroenterol. 97 (2002) 695–699.
[32] L. Schlosstein, P.I. Terasaki, R. Bluestone, C.M. Pearson, High association of an HL-A antigen, W27, with ankylosing spondylitis, N. Engl. J. Med. 288 (1973) 704–706.
[33] M. de Menthon, M.P. Lavalley, C. Maldini, L. Guillevin, A. Mahr, HLA-B51/B5 and the risk of Behcet's disease: a systematic review and meta-analysis of case-control genetic association studies, Arth. Rheum. 61 (2009) 1287–1296.
[34] R.P. Nair, P.E. Stuart, I. Nistor, R. Hiremagalore, N.V. Chia, S. Jenisch, et al., Sequence and haplotype analysis supports HLA-C as the psoriasis susceptibility 1 gene, Am. J. Hum. Genet. 78 (2006) 827–851.
[35] I. Gockel, J. Becker, M.M. Wouters, S. Niebisch, H.R. Gockel, T. Hess, et al., Common variants in the HLA-DQ region confer susceptibility to idiopathic achalasia, Nat. Genet. 46 (2014) 901–904.
[36] E. Maggi, M.G. Giudizi, R. Biagiotti, F. Annunziato, R. Manetti, M.P. Piccinni, et al., Th2-like CD8+ T cells showing B cell helper function and reduced cytolytic activity in human immunodeficiency virus type 1 infection, J. Exp. Med. 180 (1994) 489–495.
[37] E. Mohr, A.F. Cunningham, K.M. Toellner, S. Bobat, R.E. Coughlan, R.A. Bird, et al., IFN-{gamma} produced by CD8 T cells induces T-bet-dependent and -independent class switching in B cells in responses to alum-precipitated protein vaccine, Proc. Natl. Acad. Sci. U. S. A. 107 (2010) 17292–17297.
[38] P.K. Gregersen, J. Silver, R.J. Winchester, The shared epitope hypothesis. An approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis, Arth. Rheum. 30 (1987) 1205–1213.
[39] P. Invernizzi, Human leukocyte antigen in primary biliary cirrhosis: an old story now reviving, Hepatology 54 (2011) 714–723.
[40] P. Invernizzi, M. Ransom, S. Raychaudhuri, R. Kosoy, A. Lleo, R. Shigeta, et al., Classical HLA-DRB1 and DPB1 alleles account for HLA associations with primary biliary cirrhosis, Genes Immun. 13 (2012) 461–468.
[41] J.P. Achkar, L. Klei, P.I. de Bakker, G. Bellone, N. Rebert, R. Scott, et al., Amino acid position 11 of HLA-DRbeta1 is a major determinant of chromosome 6p association with ulcerative colitis, Genes Immun. 13 (2012) 245–252.
[42] S. Raychaudhuri, C. Sandor, E.A. Stahl, J. Freudenberg, H.S. Lee, X. Jia, et al., Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis, Nat. Genet. 44 (2012) 291–296.
[43] S.W. Qiao, R. Iversen, M. Raki, L.M. Sollid, The adaptive immune response in celiac disease, Semin. Immunopathol. 34 (2012) 523–540.
[44] R.W. Ebringer, D.R. Cawdell, P. Cowling, A. Ebringer, Sequential studies in ankylosing spondylitis. Association of Klebsiella pneumoniae with active disease, Ann. Rheum. Dis. 37 (1978) 146–151.
[45] P.L. Schwimmbeck, D.T. Yu, M.B. Oldstone, Autoantibodies to HLA B27 in the sera of HLA B27 patients with ankylosing spondylitis and Reiter's syndrome. Molecular mimicry with Klebsiella pneumoniae as potential mechanism of autoimmune disease, J. Exp. Med. 166 (1987) 173–181.
[46] T.K. Kvien, A. Glennas, K. Melby, K. Granfors, O. Andrup, B. Karstensen, et al., Reactive arthritis: incidence, triggering agents and clinical presentation, J. Rheumatol. 21 (1994) 115–122.
[47] D.E. de Almeida, S. Ling, J. Holoshitz, New insights into the functional role of the rheumatoid arthritis shared epitope, FEBS Lett. 585 (2012) 3619–3626.
[48] N. Bottini, L. Musumeci, A. Alonso, S. Rahmouni, K. Nika, M. Rostamkhani, et al., A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes, Nat. Genet. 36 (2004) 337–338.
[49] T. Vang, M. Congia, M.D. Macis, L. Musumeci, V. Orru, P. Zavattari, et al., Autoimmune-associated lymphoid tyrosine phosphatase is a gain-of-function variant, Nat. Genet. 37 (2005) 1317–1319.
[50] J. Zhang, N. Zahir, Q. Jiang, H. Miliotis, S. Heyraud, X. Meng, et al., The autoimmune disease-associated PTPN22 variant promotes calpain-mediated Lyp/pep degradation associated with lymphocyte and dendritic cell hyperresponsiveness, Nat. Genet. 43 (2011) 902–907.
[51] M. Rieck, A. Arechiga, S. Onengut-Gumuscu, C. Greenbaum, P. Concannon, J.H. Buckner, Genetic variation in PTPN22 corresponds to altered function of T and B lymphocytes, J. Immunol. 179 (2007) 4704–4710.
[52] A.F. Arechiga, T. Habib, Y. He, X. Zhang, Z.Y. Zhang, A. Funk, et al., Cutting edge: the PTPN22 allelic variant associated with autoimmunity impairs B cell signaling, J. Immunol. 182 (2009) 3343–3347.
[53] N. Bottini, E.J. Peterson, Tyrosine phosphatase PTPN22: multifunctional regulator of immune signaling, development, and disease, Annu. Rev. Immunol. 32 (2014) 83–119.
[54] E.F. Remmers, R.M. Plenge, A.T. Lee, R.R. Graham, G. Hom, T.W. Behrens, et al., STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus, N. Engl. J. Med. 357 (2007) 977–986.
[55] S.L. Musone, K.E. Taylor, T.T. Lu, J. Nititham, R.C. Ferreira, W. Ortmann, et al., Multiple polymorphisms in the TNFAIP3 region are independently associated with systemic lupus erythematosus, Nat. Genet. 40 (2008) 1062–1064.
[56] I. Adrianto, S. Wang, G.B. Wiley, C.J. Lessard, J.A. Kelly, A.J. Adler, et al., Association of two independent functional risk haplotypes in TNIP1 with systemic lupus erythematosus, Arth. Rheum. 64 (2012) 3695–3705.
[57] X. Liu, P. Invernizzi, Y. Lu, R. Kosoy, Y. Lu, I. Bianchi, et al., Genome-wide meta-analyses identify three loci associated with primary biliary cirrhosis, Nat. Genet. 42 (2010) 658–660.
[58] Y.H. Lee, S.C. Bae, S.J. Choi, J.D. Ji, G.G. Song, Associations between interferon regulatory factor 5 polymorphisms and rheumatoid arthritis: a meta-analysis, Mol. Biol. Rep. 40 (2013) 1791–1799.
[59] A.B. Begovich, V.E. Carlton, L.A. Honigberg, S.J. Schrodi, A.P. Chokkalingam, H.C. Alexander, et al., A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis, Am. J. Hum. Genet. 75 (2004) 330–337.
[60] S.A. Chung, L.A. Criswell, PTPN22: its role in SLE and autoimmunity, Autoimmunity 40 (2007) 582–590.
[61] L.M. Diaz-Gallo, L. Espino-Paisan, K. Fransen, M. Gomez-Garcia, S. van Sommeren, C. Cardena, et al., Differential association of two PTPN22 coding variants with Crohn's disease and ulcerative colitis, Inflamm. Bowel Dis. 17 (2011) 2287–2294.
[62] G.F. Mells, J.A. Floyd, K.I. Morley, H.J. Cordell, C.S. Franklin, S.Y. Shin, et al., Genome-wide association study identifies 12 new susceptibility loci for primary biliary cirrhosis, Nat. Genet. 43 (2011) 329–332.
[63] G.M. Hirschfield, X. Liu, Y. Han, I.P. Gorlov, Y. Lu, C. Xu, et al., Variants at IRF5-TNPO3, 17q12-21 and MMEL1 are associated with primary biliary cirrhosis, Nat. Genet. 42 (2010) 655–657.
[64] Y.H. Lee, S.C. Bae, Association between the functional ITGAM rs1143679 G/A polymorphism and systemic lupus erythematosus/lupus nephritis or rheumatoid arthritis: an update meta-analysis, Rheumatol. Int. 35 (2014) 815–823.
[65] J.M. Anaya, X. Kim-Howard, S. Prahalad, A. Chernavsky, C. Canas, A. Rojas-Villarraga, et al., Evaluation of genetic association between an ITGAM non-synonymous SNP (rs1143679) and multiple autoimmune diseases, Autoimmun. Rev. 11 (2011) 276–280.
[66] D.M. Evans, C.C. Spencer, J.J. Pointon, Z. Su, D. Harvey, G. Kochan, et al., Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility, Nat. Genet. 43 (2011) 761–767.
[67] A. Strange, F. Capon, C.C. Spencer, J. Knight, M.E. Weale, M.H. Allen, et al., A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between HLA-C and ERAP1, Nat. Genet. 42 (2010) 985–990.
[68] Y. Kirino, G. Bertsias, Y. Ishigatsubo, N. Mizuki, I. Tugal-Tutkun, E. Seyahi, et al., Genome-wide association analysis identifies new susceptibility loci for Behcet's disease and epistasis between HLA-B*51 and ERAP1, Nat. Genet. 45 (2012) 202–207.
[69] I.A. York, M.A. Brehm, S. Zendzian, C.F. Towne, K.L. Rock, Endoplasmic reticulum aminopeptidase 1 (ERAP1) trims MHC class I-presented peptides in vivo and plays an important role in immunodominance, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 9202–9207.
[70] K. Nagamine, P. Peterson, H.S. Scott, J. Kudoh, S. Minoshima, M. Heino, et al., Positional cloning of the APECED gene, Nat. Genet. 17 (1997) 393–398.
[71] B. Namjou, P.H. Kothari, J.A. Kelly, S.B. Glenn, J.O. Ojwang, A. Adler, et al., Evaluation of the TREX1 gene in a large multi-ancestral lupus cohort, Genes Immun. 12 (2011) 270–279.
[72] M.A. Lee-Kirsch, M. Gong, D. Chowdhury, L. Senenko, K. Engel, Y.A. Lee, et al., Mutations in the gene encoding the 3′-5′ DNA exonuclease TREX1 are associated with systemic lupus erythematosus, Nat. Genet. 39 (2007) 1065–1067.
[73] G. Trynka, K.A. Hunt, N.A. Bockett, J. Romanos, V. Mistry, A. Szperl, et al., Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease, Nat. Genet. 43 (2011) 1193–1201.
[74] J. Wessel, A.Y. Chu, S.M. Willems, S. Wang, H. Yaghootkar, J.A. Brody, et al., Low-frequency and rare exome chip variants associate with fasting glucose and type 2 diabetes susceptibility, Nat. Commun. 6 (2015) 5897.
[75] J.R. Huyghe, A.U. Jackson, M.P. Fogarty, M.L. Buchkovich, A. Stancakova, H.M. Stringham, et al., Exome array analysis identifies new loci and low-frequency variants influencing insulin processing and secretion, Nat. Genet. 45 (2013) 197–201.
[76] H. Tang, X. Jin, Y. Li, H. Jiang, X. Tang, X. Yang, et al., A large-scale screen for coding variants predisposing to psoriasis, Nat. Genet. 46 (2014) 45–50.

Please cite this article in press as: M.F. Seldin, The genetics of human autoimmune disease: A perspective on progress in the field and future directions, Journal of Autoimmunity (2015), http://dx.doi.org/10.1016/j.jaut.2015.08.015

[77] J. Massey, S. Eyre, Rare variants and autoimmune disease, Brief. Funct. Genom. 13 (2014) 392–397.
[78] S. Lee, M.J. Emond, M.J. Bamshad, K.C. Barnes, M.J. Rieder, D.A. Nickerson, et al., Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies, Am. J. Hum. Genet. 91 (2012) 224–237.
[79] D.J. Liu, S.M. Leal, A novel adaptive method for the analysis of next-generation sequencing data to detect complex trait associations with rare variants due to gene main effects and interactions, PLoS Genet. 6 (2010) e1001156.
[80] B.E. Madsen, S.R. Browning, A groupwise association test for rare mutations using a weighted sum statistic, PLoS Genet. 5 (2009) e1000384.
[81] A.L. Price, G.V. Kryukov, P.I. de Bakker, S.M. Purcell, J. Staples, L.J. Wei, et al., Pooled association tests for rare variants in exon-resequencing studies, Am. J. Hum. Genet. 86 (2010) 832–838.
[82] A.L. Price, N.A. Zaitlen, D. Reich, N. Patterson, New approaches to population stratification in genome-wide association studies, Nat. Rev. Genet. 11 (2010) 459–463.
[83] H. Hu, C.D. Huff, B. Moore, S. Flygare, M.G. Reese, M. Yandell, VAAST 2.0: improved variant classification and disease-gene identification using a conservation-controlled amino acid substitution matrix, Genet. Epidemiol. 37 (2013) 622–634.
[84] G.M. Clarke, M.A. Rivas, A.P. Morris, A flexible approach for the analysis of rare variants allowing for a mixture of effects on binary or quantitative traits, PLoS Genet. 9 (2013) e1003694.
[85] P. Kumar, S. Henikoff, P.C. Ng, Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm, Nat. Protoc. 4 (2009) 1073–1081.
[86] I. Adzhubei, D.M. Jordan, S.R. Sunyaev, Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2. Current Protocols in Human Genetics, editorial board, Jonathan L Haines [et al], 2013 (Chapter 7):Unit7 20.
[87] J.M. Schwarz, C. Rodelsperger, M. Schuelke, D. Seelow, Mutationtaster evaluates disease-causing potential of sequence alterations, Nat. Methods 7 (2010) 575–576.
[88] K. Wang, M. Li, H. Hakonarson, Annovar: functional annotation of genetic variants from high-throughput sequencing data, Nucleic Acids Res. 38 (2010) e164.
[89] P. Flicek, I. Ahmed, M.R. Amode, D. Barrell, K. Beal, S. Brent, et al., Ensembl 2013, Nucleic Acids Res. 41 (2013) D48–D55.
[90] A.P. Boyle, E.L. Hong, M. Hariharan, Y. Cheng, M.A. Schaub, M. Kasowski, et al., Annotation of functional variation in personal genomes using Regulomedb, Genome Res. 22 (2012) 1790–1797.
[91] A. Bhattacharya, J.D. Ziebarth, Y. Cui, PolymiRTS Database 3.0: linking polymorphisms in microRNAs and their target sites with human diseases and biological pathways, Nucleic Acids Res. 42 (2014) D86–D91.
[92] K.V. Lehmann, T. Chen, Exploring functional variant discovery in non-coding regions with SInBaD, Nucleic Acids Res. 41 (2013) e7.
[93] J.W. Han, H.F. Zheng, Y. Cui, L.D. Sun, D.Q. Ye, Z. Hu, et al., Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus, Nat. Genet. 41 (2009) 1234–1237.
[94] Y. Okada, D. Wu, G. Trynka, T. Raj, C. Terao, K. Ikari, et al., Genetics of rheumatoid arthritis contributes to biology and drug discovery, Nature 506 (2014) 376–381.
[95] H.S. Lee, T. Kim, S.Y. Bang, Y.J. Na, I. Kim, K. Kim, J.H. Kim, Y.J. Chung, H.D. Shin, Y.M. Kang, S.C. Shim, C.H. Suh, Y.B. Park, J.S. Kim, C. Kang, S.C. Bae, Ethnic specificity of lupus-associated loci identified in a genome-wide association study in Korean women, Ann. Rheum. Dis. 73 (2014) 1240–1245.
[96] L. Jiang, J. Yin, L. Ye, J. Yang, G. Hemani, A.J. Liu, et al., Novel risk loci for rheumatoid arthritis in Han Chinese and congruence with risk variants in Europeans, Arthr. Rheumatol. Hoboken NJ 66 (2014) 1121–1132.
[97] P.K. Gregersen, H.S. Lee, F. Batliwalla, A.B. Begovich, PTPN22: setting thresholds for autoimmunity, Semin. Immunol. 18 (2006) 214–223.
[98] K. Yamamoto, Y. Okada, A. Suzuki, Y. Kochi, Genetics of rheumatoid arthritis in Asia-present and future, Nat. Rev. 11 (2015) 375–379.
[99] K. Ikari, M. Kuwahara, T. Nakamura, S. Momohara, M. Hara, H. Yamanaka, et al., Association between PADI4 and rheumatoid arthritis: a replication study, Arth. Rheum. 52 (2005) 3054–3057.
[100] M.F. Seldin, R. Shigeta, K. Laiho, H. Li, H. Saila, A. Savolainen, et al., Finnish case-control and family studies support PTPN22 R620W polymorphism as a risk factor in rheumatoid arthritis, but suggest only minimal or no effect in juvenile idiopathic arthritis, Genes Immun. 6 (2005) 720–722.
[101] P.S. Ramos, S.R. Shaftman, R.C. Ward, C.D. Langefeld, Genes associated with SLE are targets of recent positive selection, Autoimmun. Dis. 2014 (2014) 203435.
[102] A. Zhernakova, C.C. Elbers, B. Ferwerda, J. Romanos, G. Trynka, P.C. Dubois, et al., Evolutionary and functional analysis of celiac risk loci reveals SH2B3 as a protective factor against bacterial infection, Am. J. Hum. Genet. 86 (2010) 970–977.
[103] M. Molokhia, P. McKeigue, Risk for rheumatic disease in relation to ethnicity and admixture, Arth. Res. 2 (2000) 115–125.
[104] M. Molokhia, C. Hoggart, A.L. Patrick, M. Shriver, E. Parra, J. Ye, et al., Relation of risk of systemic lupus erythematosus to west African admixture in a Caribbean population, Hum. Genet. 112 (2003) 310–318.
[105] M.F. Seldin, L. Qi, H.R. Scherbarth, C. Tian, M. Ransom, G. Silva, et al., Amerindian ancestry in Argentina is associated with increased risk for systemic lupus erythematosus, Genes Immun. 9 (2008) 389–393.
[106] S.A. Chung, C. Tian, K.E. Taylor, A.T. Lee, W.A. Ortmann, G. Hom, et al., European population substructure is associated with mucocutaneous manifestations and autoantibody production in systemic lupus erythematosus, Arth. Rheum. 60 (2009) 2448–2456.
[107] I.B. Richman, S.A. Chung, K.E. Taylor, R. Kosoy, C. Tian, W.A. Ortmann, et al., European population substructure correlates with systemic lupus erythematosus endophenotypes in North Americans of European descent, Genes Immun. 11 (2010) 515–521.
[108] I.B. Richman, K.E. Taylor, S.A. Chung, L. Trupin, M. Petri, E. Yelin, et al., European genetic ancestry protects against the development of renal disease in systemic lupus erythematosus, Arth. Rheum. 64 (2012) 3374–3382.
[109] E. Sanchez, A. Rasmussen, L. Riba, E. Acevedo-Vasquez, J.A. Kelly, C.D. Langefeld, et al., Impact of genetic ancestry and sociodemographic status on the clinical expression of systemic lupus erythematosus in American Indian-European populations, Arth. Rheum. 64 (2012) 3687–3694.
[110] J.C. Denny, M.D. Ritchie, M.A. Basford, J.M. Pulley, L. Bastarache, K. Brown-Gentry, et al., PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations, Bioinformatics 26 (2010) 1205–1210.
[111] S.J. Hebbring, S.J. Schrodi, Z. Ye, Z. Zhou, D. Page, M.H. Brilliant, A PheWAS approach in studying HLA-DRB1*1501, Genes Immun. 14 (2013) 187–191.
[112] S.K. Nath, S. Han, X. Kim-Howard, J.A. Kelly, P. Viswanathan, G.S. Gilkeson, et al., A nonsynonymous functional variant in integrin-alpha(M) (encoded by ITGAM) is associated with systemic lupus erythematosus, Nat. Genet. 40 (2008) 152–154.
[113] X. Dai, R.G. James, T. Habib, S. Singh, S. Jackson, S. Khim, et al., A disease-associated PTPN22 variant promotes systemic autoimmunity in murine models, J. Clin. Invest. 123 (2013) 2024–2036.
[114] D.J. Wu, W. Zhou, S. Enouz, V. Orru, S.M. Stanford, C.J. Maine, et al., Autoimmunity-associated LYP-W620 does not impair thymic negative selection of autoreactive T cells, PLoS One 9 (2014) e86677.
[115] E.C. Baechler, P.K. Gregersen, T.W. Behrens, The emerging role of interferon in human systemic lupus erythematosus, Curr. Opin. Immunol. 16 (2004) 801–807.
[116] E.C. Baechler, F.M. Batliwalla, G. Karypis, P.M. Gaffney, W.A. Ortmann, K.J. Espe, et al., Interferon-inducible gene expression signature in peripheral blood cells of patients with severe lupus, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 2610–2615.
[117] E.C. Baechler, J.W. Bauer, C.A. Slattery, W.A. Ortmann, K.J. Espe, J. Novitzke, et al., An interferon signature in the peripheral blood of dermatomyositis patients is associated with disease activity, Mol. Med. Camb. Mass 13 (2007) 59–68.
[118] T.B. Niewold, Interferon alpha as a primary pathogenic factor in human lupus, J. Interferon Cytokine Res. 31 (2011) 887–892.
[119] T.B. Niewold, J.E. Adler, S.B. Glenn, T.J. Lehman, J.B. Harley, M.K. Crow, Age- and sex-related patterns of serum interferon-alpha activity in lupus families, Arth. Rheum. 58 (2008) 2113–2119.
[120] V. Badot, C. Galant, A. Nzeusseu Toukap, I. Theate, A.L. Maudoux, B.J. Van den Eynde, et al., Gene expression profiling in the synovium identifies a predictive signature of absence of response to adalimumab therapy in rheumatoid arthritis, Arth. Res. Ther. 11 (2009) R57.
[121] C.L. Galligan, E. Baig, V. Bykerk, E.C. Keystone, E.N. Fish, Distinctive gene expression signatures in rheumatoid arthritis synovial tissue fibroblast cells: correlates with disease activity, Genes Immun. 8 (2007) 480–491.
[122] V. Devauchelle, S. Marion, N. Cagnard, S. Mistou, G. Falgarone, M. Breban, et al., DNA microarray allows molecular profiling of rheumatoid arthritis and identification of pathophysiological targets, Genes Immun. 5 (2004) 597–608.
[123] C.L. Verweij, Transcript profiling towards personalised medicine in rheumatoid arthritis, Neth. J. Med. 67 (2009) 364–371.
[124] S. Anders, A. Reyes, W. Huber, Detecting differential usage of exons from RNA-seq data, Genome Res. 22 (2012) 2008–2017.
[125] J.E. Hooper, A survey of software for genome-wide discovery of differential splicing in RNA-Seq data, Hum. Genom. 8 (2014) 3.
[126] Consortium EP, An integrated encyclopedia of DNA elements in the human genome, Nature 489 (2012) 57–74.
[127] T.P. Yang, C. Beazley, S.B. Montgomery, A.S. Dimas, M. Gutierrez-Arcelus, B.E. Stranger, et al., Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies, Bioinforma. Oxf. Engl. 26 (2010) 2474–2476.
[128] E. Grundberg, E. Meduri, J.K. Sandling, A.K. Hedman, S. Keildson, A. Buil, et al., Global analysis of DNA methylation variation in adipose tissue from twins reveals links to disease-associated variants in distal regulatory elements, Am. J. Hum. Genet. 93 (2013) 876–890.
[129] K. Xia, A.A. Shabalin, S. Huang, V. Madar, Y.H. Zhou, W. Wang, et al., seeQTL: a searchable database for human eQTLs, Bioinforma. Oxf. Engl. 28 (2012) 451–452.
[130] G.C. Wu, H.F. Pan, R.X. Leng, D.G. Wang, X.P. Li, X.M. Li, et al., Emerging role of long noncoding RNAs in autoimmune diseases, Autoimmun. Rev. 14 (2015) 798–805.
[131] B. Hrdlickova, V. Kumar, K. Kanduri, D.V. Zhernakova, S. Tripathi, J. Karjalainen, et al., Expression profiles of long non-coding RNAs located in autoimmune disease-associated regions reveal immune cell-type specificity,

Please cite this article in press as: M.F. Seldin, The genetics of human autoimmune disease: A perspective on progress in the field and future directions, Journal of Autoimmunity (2015), http://dx.doi.org/10.1016/j.jaut.2015.08.015

Genome Med. 6 (2014) 88.
[132] K.K. Farh, A. Marson, J. Zhu, M. Kleinewietfeld, W.J. Housley, S. Beik, et al., Genetic and epigenetic fine mapping of causal autoimmune disease variants, Nature 518 (2014) 337–343.
[133] L.E. Macdonald, M. Karow, S. Stevens, W. Auerbach, W.T. Poueymirou, J. Yasenchak, et al., Precise and in situ genetic humanization of 6 Mb of mouse immunoglobulin genes, Proc. Natl. Acad. Sci. U. S. A. 111 (2014) 5147–5152.
[134] H. Ogata, S. Goto, K. Sato, W. Fujibuchi, H. Bono, M. Kanehisa, KEGG: kyoto encyclopedia of genes and genomes, Nucleic Acids Res. 27 (1999) 29–34.
[135] M. Ashburner, C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, et al., Gene ontology: tool for the unification of biology. The gene ontology consortium, Nat. Genet. 25 (2000) 25–29.
[136] D. Croft, A.F. Mundo, R. Haw, M. Milacic, J. Weiser, G. Wu, et al., The reactome pathway knowledgebase, Nucleic Acids Res. 42 (2014) D472–D477.
[137] K. Zhang, S. Cui, S. Chang, L. Zhang, J. Wang, i-GSEA4GWAS: a web server for identification of pathways/gene sets associated with traits by applying an improved gene set enrichment analysis to genome-wide association study, Nucleic Acids Res. 38 (2010) W90–W95.
[138] B.L. Yaspan, W.S. Bush, E.S. Torstenson, D. Ma, M.A. Pericak-Vance, M.D. Ritchie, et al., Genetic analysis of biological pathway data through genomic randomization, Hum. Genet. 129 (2011) 563–571.
[139] B.L. Yaspan, O.J. Veatch, Strategies for Pathway Analysis from GWAS Data. Current Protocols in Human Genetics, editorial board, Jonathan L Haines [et al, 2011 (Chapter 1):Unit1 20.
[140] P.H. Lee, C. O'Dushlaine, B. Thomas, S.M. Purcell, INRICH: interval-based enrichment analysis for genome-wide association studies, Bioinformatics 28 (2012) 1797–1799.
[141] I.M. Sclerosis, Network-based multiple sclerosis pathway analysis with GWAS data from 15,000 cases and 30,000 controls, Am. J. Hum. Genet. 92 (2013) 854–865.
[142] S. Raychaudhuri, B.P. Thomson, E.F. Remmers, S. Eyre, A. Hinks, C. Guiducci, et al., Genetic variants at CD28, PRDM1 and CD2/CD58 are associated with rheumatoid arthritis risk, Nat. Genet. 41 (2009) 1313–1318.
[143] M.M. Zhang, Y.S. Jiang, H.C. Lv, H.B. Mu, J. Li, Z.W. Shang, et al., Pathway-based association analysis of two genome-wide screening data identifies rheumatoid arthritis-related pathways, Genes Immun. 15 (2014) 487–494.
[144] Y.H. Lee, S.C. Bae, S.J. Choi, J.D. Ji, G.G. Song, Genome-wide pathway analysis of genome-wide association studies on systemic lupus erythematosus and rheumatoid arthritis, Mol. Biol. Rep. 39 (2012) 10627–10635.
[145] M. Evangelou, D.J. Smyth, M.D. Fortune, O.S. Burren, N.M. Walker, H. Guo, et al., A method for gene-based pathway analysis using genomewide association study summary statistics reveals nine new type 1 diabetes associations, Genet. Epidemiol. 38 (2014) 661–670.
[146] V. Kumar, J. Gutierrez-Achury, K. Kanduri, R. Almeida, B. Hrdlickova, D.V. Zhernakova, et al., Systematic annotation of celiac disease loci refines pathological pathways and suggests a genetic explanation for increased interferon-gamma levels, Hum. Mol. Genet. 24 (2015) 397–409.
[147] S.P. Kar, M.F. Seldin, W. Chen, E. Lu, G.M. Hirschfield, P. Invernizzi, et al., Pathway-based analysis of primary biliary cirrhosis genome-wide association studies, Genes Immun. 14 (2013) 179–186.
[148] H. Cordell, Y. Han, Y. Li, G. Mells, G. Hirschfield, C. Greene, et al., An international genome-wide meta-analysis identifies new primary biliary cirrhosis risk loci and highlights pathogenic pathways for drug targeting, Nat. Genet. (2015) in press.
[149] P. Carbonetto, M. Stephens, Integrated enrichment analysis of variants and pathways in genome-wide association studies indicates central role for IL-2 signaling genes in type 1 diabetes, and cytokine signaling genes in Crohn's disease, PLoS Genet. 9 (2013) e1003770.
[150] J.H. Wang, D. Pappas, P.L. De Jager, D. Pelletier, P.I. de Bakker, L. Kappos, et al., Modeling the cumulative genetic risk for multiple sclerosis from genome-wide association data, Genome Med. 3 (2009) 3.
[151] A. Yarwood, B. Han, S. Raychaudhuri, J. Bowes, M. Lunt, D.A. Pappas, et al., A weighted genetic risk score using all known susceptibility variants to estimate rheumatoid arthritis risk, Ann. Rheum. Dis. 74 (2015) 170–176.
[152] R. Tang, H. Chen, Q. Miao, Z. Bian, W. Ma, X. Feng, et al., The cumulative effects of known susceptibility variants to predict primary biliary cirrhosis risk, Genes Immun. 16 (2015) 193–198.
[153] J. Cui, S. Saevarsdottir, B. Thomson, L. Padyukov, A.H. van der Helm-van Mil, J. Nititham, et al., PTPRC rheumatoid arthritis risk allele is also associated with response to anti-TNF therapy, Arth. Rheum. 62 (2010) 1849–1861.
[154] C. Liu, F. Batliwalla, W. Li, A. Lee, R. Roubenoff, E. Beckman, et al., Genome-wide association scan identifies candidate polymorphisms associated with differential response to anti-TNF treatment in rheumatoid arthritis, Mol. Med. 14 (2008) 575–581.
[155] M. Suarez-Gestal, E. Perez-Pampin, M. Calaza, J.J. Gomez-Reino, A. Gonzalez, Lack of replication of genetic predictors for the rheumatoid arthritis response to anti-TNF treatments: a prospective case-only study, Arth. Res. Ther. 12 (2010) R72.
[156] S.B. Krintel, G. Palermo, J.S. Johansen, S. Germer, L. Essioux, R. Benayed, et al., Investigation of single nucleotide polymorphisms and biological pathways associated with response to TNFalpha inhibitors in patients with rheumatoid arthritis, Pharmacogenetics Genomics 22 (2012) 577–589.
[157] D. Plant, J. Bowes, C. Potter, K.L. Hyrich, A.W. Morgan, A.G. Wilson, et al., Genome-wide association study of genetic predictors of anti-tumor necrosis factor treatment efficacy in rheumatoid arthritis identifies associations with polymorphisms at seven loci, Arth. Rheum. 63 (2011) 645–653.
[158] M. Umicevic Mirkov, J. Cui, S.H. Vermeulen, E.A. Stahl, E.J. Toonen, R.R. Makkinje, et al., Genome-wide association analysis of anti-TNF drug response in patients with rheumatoid arthritis, Ann. Rheum. Dis. 72 (2013) 1375–1381.
[159] R.C. Green, J.S. Berg, W.W. Grody, S.S. Kalia, B.R. Korf, C.L. Martin, et al., ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing, Genet. Med. 15 (2013) 565–574.
[160] Y. Yang, D.M. Muzny, J.G. Reid, M.N. Bainbridge, A. Willis, P.A. Ward, et al., Clinical whole-exome sequencing for the diagnosis of mendelian disorders, N. Engl. J. Med. 369 (2013) 1502–1511.
[161] H. Lee, J.L. Deignan, N. Dorrani, S.P. Strom, S. Kantarci, F. Quintero-Rivera, et al., Clinical exome sequencing for genetic identification of rare Mendelian disorders, Jama 312 (2014) 1880–1887.
[162] S.E. Soden, C.J. Saunders, L.K. Willig, E.G. Farrow, L.D. Smith, J.E. Petrikin, et al., Effectiveness of exome and genome sequencing guided by acuity of illness for diagnosis of neurodevelopmental disorders, Sci. Transl. Med. 6 (2014), 265ra168.
[163] F. Pociot, B. Akolkar, P. Concannon, H.A. Erlich, C. Julier, G. Morahan, et al., Genetics of type 1 diabetes: what's next? Diabetes 59 (2010) 1561–1571.
[164] G.M. Hirschfield, X. Liu, C. Xu, Y. Lu, G. Xie, Y. Lu, et al., Primary biliary cirrhosis associated with HLA, IL12A, and IL12RB2 variants, N. Engl. J. Med. 360 (2009) 2544–2555.
[165] G. Hom, R.R. Graham, B. Modrek, K.E. Taylor, W. Ortmann, S. Garnier, et al., Association of systemic lupus erythematosus with C8orf13-BLK and ITGAM-ITGAX, N. Engl. J. Med. 358 (2008) 900–909.
[166] I.A. Hoppenbrouwers, Y.S. Aulchenko, A.C. Janssens, S.V. Ramagopalan, L. Broer, M. Kayser, et al., Replication of CD58 and CLEC16A as genome-wide significant risk genes for multiple sclerosis, J. Hum. Genet. 54 (2009) 676–680.
[167] Y.H. Lee, Y.H. Rho, S.J. Choi, J.D. Ji, G.G. Song, S.K. Nath, et al., The PTPN22 C1858T functional polymorphism and autoimmune diseases–a meta-analysis, Rheumatol. Oxf. 46 (2007) 49–56.
[168] Y.F. Chen, J.S. Chang, PTPN22 C1858T and the risk of psoriasis: a meta-analysis, Mol. Biol. Rep. 39 (2012) 7861–7870.
[169] R.M. Plenge, M. Seielstad, L. Padyukov, A.T. Lee, E.F. Remmers, B. Ding, et al., TRAF1-C5 as a risk locus for rheumatoid arthritis–a genomewide study, N. Engl. J. Med. 357 (2007) 1199–1209.
[170] S. Sigurdsson, H.H. Goring, G. Kristjansdottir, L. Milani, G. Nordmark, J.K. Sandling, et al., Comprehensive evaluation of the genetic variants of interferon regulatory factor 5 (IRF5) reveals a novel 5 bp length polymorphism as strong risk factor for systemic lupus erythematosus, Hum. Mol. Genet. 17 (2008) 872–881.
[171] B.D. Juran, G.M. Hirschfield, P. Invernizzi, E.J. Atkinson, Y. Li, G. Xie, et al., Immunochip analyses identify a novel risk locus for primary biliary cirrhosis at 13q14, multiple independent associations at four established risk loci and epistasis between 1p31 and 7q32 risk variants, Hum. Mol. Genet. 21 (2012) 5209–5221.
[172] P.S. Ramos, L.A. Criswell, K.L. Moser, M.E. Comeau, A.H. Williams, N.M. Pajewski, et al., A comprehensive analysis of shared loci between systemic lupus erythematosus (SLE) and sixteen autoimmune diseases reveals limited genetic overlap, PLoS Genet. 7 (2011) e1002406.
[173] M. Cargill, S.J. Schrodi, M. Chang, V.E. Garcia, R. Brandon, K.P. Callis, et al., A large-scale genetic association study confirms IL12B and leads to the identification of IL23R as psoriasis-risk genes, Am. J. Hum. Genet. 80 (2007) 273–290.
[174] Y. Kochi, Y. Okada, A. Suzuki, K. Ikari, C. Terao, A. Takahashi, et al., A regulatory variant in CCR6 is associated with rheumatoid arthritis susceptibility, Nat. Genet. 42 (2010) 515–519.
[175] W. Yang, N. Shen, D.Q. Ye, Q. Liu, Y. Zhang, X.X. Qian, et al., Genome-wide association study in Asian populations identifies variants in ETS1 and WDFY4 associated with systemic lupus erythematosus, PLoS Genet. 6 (2010) e1000841.
[176] W. Yang, H. Tang, Y. Zhang, X. Tang, J. Zhang, L. Sun, et al., Meta-analysis followed by replication identifies loci in or near CDKN1B, TET3, CD80, DRAM1, and ARID5B as associated with systemic lupus erythematosus in Asians, Am. J. Hum. Genet. 92 (2013) 41–51.
[177] M. Dong, J. Li, R. Tang, P. Zhu, F. Qiu, C. Wang, et al., Multiple genetic variants associated with primary biliary cirrhosis in a Han Chinese population, Clin. Rev. Allerg. Immunol. 48 (2015) 316–321.
[178] Y. Okada, K. Shimane, Y. Kochi, T. Tahira, A. Suzuki, K. Higasa, et al., A genome-wide association study identified AFF1 as a susceptibility locus for systemic lupus erythematosus in Japanese, PLoS Genet. 8 (2012) e1002455.
[179] M. Nakamura, N. Nishida, M. Kawashima, Y. Aiba, A. Tanaka, M. Yasunami, et al., Genome-wide association study identifies TNFSF15 and POU2AF1 as susceptibility loci for primary biliary cirrhosis in the Japanese population, Am. J. Hum. Genet. 91 (2012) 721–728.

Please cite this article in press as: M.F. Seldin, The genetics of human autoimmune disease: A perspective on progress in the field and future directions, Journal of Autoimmunity (2015), http://dx.doi.org/10.1016/j.jaut.2015.08.015