Peer-reviewed version available at Genes 2018, 9, 377; doi:10.3390/genes9080377

Unravelling the roles of susceptibility loci for autoimmune diseases in the post-GWAS era

Jody Ye[1*], Kathleen Gillespie[1], Santiago Rodriguez[2]

[1]Diabetes and Metabolism, Translational Health Sciences, University of Bristol, UK,

[2]MRC Integrative Epidemiology Unit, Population Health Sciences, University of Bristol, UK

Corresponding author:

Dr. Jody Ye

Diabetes and Metabolism, Translational Heath Sciences, University of Bristol, Level2 Learning and Research Building, Southmead Hospital, Bristol, UK, BS10 5NB; jody.yi.ye@bristol.ac.uk, Tel: +44 (0)117 414 8041

*Present address: Room 702, Forchheimer Building, Division of Endocrinology, Department of Medicine, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, New York, USA, 10461, yi.ye@einstein.yu.edu, Tel: +1 718-430-8503

Abstract

Although genome-wide association studies (GWAS) have identified several hundred loci associated with autoimmune diseases, their mechanistic insights are still poorly understood. The human genome is more complex than common single nucleotide polymorphisms (SNPs) that are interrogated by GWAS arrays. Some structural variants such as insertions-deletions, copy number variations, and minisatellites that are not very well tagged by SNPs cannot be fully explored by GWAS. Therefore, it is possible that some of these loci may have large effects on autoimmune disease risk. In addition, other layers of regulations such as gene-gene interactions, epigenetic-determinants, gene and environmental interactions also contribute to the heritability of autoimmune diseases. This review focuses on discussing why studying these elements may allow us to gain a more comprehensive understanding of the aetiology of complex autoimmune traits.

Key words: complex loci, autoimmune diseases

Peer-reviewed version available at Genes 2018, 9, 377; doi:10.3390/genes9080377

Introduction

Autoimmune disease is a major human health burden affecting 5% to 8% of the world's population. To date more than 80 autoimmune diseases have been described (1). For some of the more common conditions such as Multiple Sclerosis, Crohn's disease, rheumatoid arthritis, and type 1 diabetes, increases in incidence and prevalence have been observed in westernized societies over the last several decades (2-5). Both genetic and environmental factors are thought to play roles in their initiations and progression of autoimmunity. Environmental factors such as viral infection (6), nutrition (7), gut dysbiosis (8) and *in utero* environment (9) have all been postulated to play a role but have been difficult to confirm due to variations among individuals, populations and geographical areas. In contrast, common genetic components of autoimmune diseases have been better characterised.

In the 1970s, studies of twins and first-degree relatives identified the most important genetic risk attributed by the HLA region (10-14). Subsequently, genome-wide association studies (GWAS) comparing affected cases with unrelated healthy individuals led to the discovery over 300 loci associated with autoimmune diseases (15). GWAS improved our knowledge of disease risk, but the specific design of GWAS does not allow consideration of other elements potentially involved in disease susceptibility. This is because GWAS are based on common genetic variants, specifically SNPs, but individuals show additional genetic variation at many other levels. For example, a phenotype can be influenced by many other types of genetic loci involving variation at more than one nucleotide. We call these 'complex loci'. In addition, phenomena such as gene-gene interactions (epistasis), and genetic-epigenetic interactions increase the complexity of the genetic basis of human common diseases and traits. Recent studies suggest that many causal variants trigger autoimmune responses in a cell-type specific

and/or cell-state specific manner. In this review, we discuss the roles of complex loci at genetic and epigenetic levels in the aetiology of autoimmune diseases.

The GWAS era

GWAS were facilitated by advancing technology to conduct high-throughput SNP analyses in very large case-control populations. For autoimmune diseases, by identifying susceptible genetic variants, one can (a) help understand the underlying biological pathways and inform the design of novel immune therapies; and (b) to predict the risk of individuals developing autoimmunity. One of the strengths of GWAS is its study design. The powerful hypothesis free, association mapping design of GWAS, has enabled the identification of hundreds of candidate genes strongly associated with human traits. Briefly, a GWAS is performed first by genotyping a set of SNPs using commercial microarrays. Subsequently, genotypes at the genome-wide level are imputed according to the haplotype structures provided by reference panels. For example, based on the early HapMap phase 1-3 reference panel, the Wellcome Trust Case Control Consortium (WTCCC) published a number of GWAS for common autoimmune diseases such as rheumatoid arthritis and type 1 diabetes (16). The 2.5 million SNPs from the HapMap reference panel only included most common variants with minor allele frequencies greater than 5% of the study population (17). The subsequent 1000 Genomes Project interrogated genotypes of nearly 40 million variants, including 1.4 million insertions / deletion variants (18) that dramatically improved GWAS coverage, allowing the identification of several novel autoimmune disease loci (19). One of the applications of GWAS results is the development of genetic risk score (GRS). GRS does not aim to detect individual SNPs, but instead it is an aggregate of genetic risk across the genome. It is calculated by combining the effect sizes of multiple SNPs, weighted by the strength of each SNP (20). Although GRS is not

Peer-reviewed version available at Genes 2018, 9, 377; doi:10.3390/genes9080377

powerful enough to make clinical diagnosis at individual level (21), it has been shown to be useful for patient stratification and risk prediction (22-25). For example, in type 1 diabetes, it was suggested that targeting the top 18% individuals of the general population with the highest ranked GRS would capture 80% of future cases (26). Given that the prevalence of type 1 diabetes is 0.4% in the general population, this will decrease the number of individuals that will need to be treated to prevent one T1D case from 250 to 50 (26).

Missing heritability in GWAS

Narrow sense heritability (h^2) is a term that defines the proportion of phenotypic variance that is contributed by additive genetic variance (27). It has been shown that GWAS tend to explain a smaller fraction of h^2 compared to those estimated from classical methods based on population data, such as from the analysis of offspring and parental phenotypes, siblings or monozygotic and dizygotic twins (28). This gap is known as missing heritability (28). Even in highly heritable conditions such as type 1 diabetes and juvenile idiopathic arthritis (h^2 up to 90%), SNP based h^2 still falls short of those reported from population estimates (28). Missing heritability was originally thought to be partly contributed by rare variants of the genome, but a recent study suggested that they have a negligible impact (29). Thus, the components of missing heritability remain largely unknown and may involve complex loci, gene-gene interactions and epigenetic regulations (30, 31).

The role of complex loci at genetic level

At the genetic level, complex loci may include insertion-deletion polymorphisms (indels), microsatellites (or short tandem repeats), minisatellites (or variable number of tandem repeats, VNTRs), long interspersed nuclear elements (32), short interspersed nuclear elements (SINES),

Alu repeats, Copy Number Variants (CNVs), cytogenetic abnormalities (including insertions, duplications, translocations, inversions), etc. There is extensive evidence demonstrating both functional significance and effect on human disease caused by genetic variation other than SNPs (33). Although the evidence linking complex loci and autoimmune diseases is relatively scarce, complex loci are under represented by the existing GWAS studies because not all of them are well tagged by SNP arrays used to conduct a GWAS and hence, they are not directly tested in GWAS. Given the large number of SNPs densely spaced through all human chromosomes, one could assume that genetic variation from complex loci could be tagged by SNPs. For example, a given SNP in linkage disequilibrium with another genetic variant, such as a VNTR or CNV, could capture the genetic variations and therefore be used as proxies. However, studies suggest that only 40% CNVs located in the segmental duplications were tagged by SNPs, leaving a large proportion unexplored (34). Similarly, only 23% rare CNVs with minor allele frequency <5% were tagged by SNPs (35). Therefore, we cannot exclude the possibilities that some poorly tagged complex loci may have large effect sizes for autoimmune conditions. One particular case relevant to Type 1 diabetes is the insulin VNTR (36). This VNTR shows wide genetic variation not only in terms of size, but also in terms of sequence within each size group (37). To our knowledge, tagging SNP for each of the different *INS* VNTR sequence variants have not been identified to date. This opens the possibility that some potential genetic effects conferred by the *INS* VNTR in relation to T1D cannot be captured by conventional GWAS. Another example is Addison's disease (AD). AD is caused by autoimmune destruction of the adrenal cortex. Apart from pan-autoimmune susceptible genes, no other genes that are specific to AD were detected (38). Bronstad et al., therefore investigated the possible susceptibility contributed by CNVs. In patients with AD, low copy number of *UGT2B28* and high copy number of *ADAM3A* were more frequent compared with healthy controls. In addition, eight novel rare CNVs were found in AD patients (38). In systemic lupus

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 May 2018 doi:10.20944/preprints201805.0160.v1

Peer-reviewed version available at Genes 2018, 9, 377; doi:10.3390/genes9080377

erythematosus (SLE), low copy number of gene *C4* encoding the complement component was strongly associated with increased disease risk, whereas high copy number of C4 was protective (39). Further analyses are required to determine if complex loci have a significant additional value to SNPs in association studies or if their potential effects are negligible.

The added complexity conferred by individual complex genetic loci increases even more when one considers combined effects of various genetic variants. If the effect of each complex locus is small, it is likely that GWAS cannot detect these complex associations. Figure 1 illustrates hypothetical scenarios to explain association with a complex trait. In Figure 1 a, genetic variation at a single locus is sufficient to cause an effect. In Figure 1b, the independent effect conferred by each of the three genetic variants is smaller than the minimum threshold that is required for the expression of the effect. However, in this case (Figure 1b), combination of the three variants is sufficient to cause an effect. GWAS is well placed to detect an association given the scenario shown in Figure 1a. However, in the absence of an appropriate tagging SNP, it is likely that GWAS would have limitations to detect the association in Figure 1b.

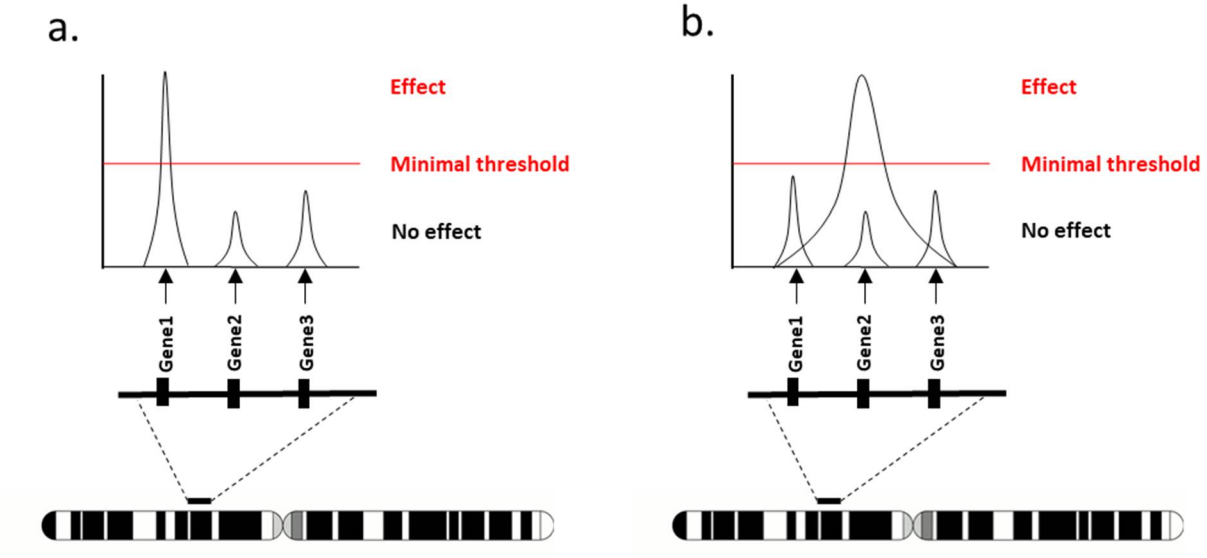

Figure 1. Schematic diagram with two potential scenarios to explain possible genetic effects of complex loci consisting three genes on a disease risk a) Genetic variation in only one of the genes in the cluster (Gene1) is sufficient to produce the effect. b) None of the genes on their own are sufficient to produce the effect. However, the joint contribution of them produces an effect on the complex trait. Figure 1 is adapted and modified from (40).

To detect multiple independent signals, various fine-mapping strategies have been developed and we chose *IL-2RA* as an example to illustrate how they have helped to define candidate causal variants for several autoimmune diseases. The *IL-2RA* (CD25) is located on chromosome 10p15.1; it encodes a subunit (IL-2Rα) of the receptor for the pro-inflammatory cytokine IL-2, which has been associated with a number of autoimmune diseases including multiple sclerosis, rheumatoid arthritis, autoimmune thyroid disease, and type 1 diabetes (2, 41, 42). Upon stimulation of IL-2, IL-2RA signals to maintain the suppressive functions of CD4+FOXP3+ regulatory T cells and to facilitate effector and memory T cell differentiation (43). The *IL-2RA* region was initially found to be associated with type 1 diabetes using a multi-locus genetic association test in 2005 (44). To better define regional associations, Lowe et al., (45) used stepwise regression to fine-map the region covering *IL-2RA* and its neighbouring gene *RBM17*, which led to the identification of two loci independently associated with type 1 diabetes (45), each comprising a number of indistinguishable SNPs. Group 1 is located in intron1 of *IL-2RA* (marked by rs12722495, previously marked by rs41295061) and Group 2 is located at the intergenic region between the 5' of *IL-2RA* and *RBM17* (marked by rs11594656) (45). Maier et al., later discovered an independent group 3 signal in intron 1 of *IL-2RA*, tagged by SNP rs2104286 (46). Apart from type 1 diabetes, Group 3 is also associated with multiple sclerosis. Functionally, the protective variant of the Group 1 SNP rs12722495 induces higher expression levels of CD25 on the surface of CD4+ memory T cells, potentially causing increased T cell activation in response to IL-2 stimulation. The protective variant of Group 3 SNP rs2104286 was counterintuitively associated with a lower percentage of CD25+ cells in CD4+ naïve fraction, suggesting a

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 May 2018 doi:10.20944/preprints201805.0160.v1

Peer-reviewed version available at Genes 2018, 9, 377; doi:10.3390/genes9080377

reduced likelihood of T cell activation (47). However, a limitation of stepwise logistic regression is that it assumes a single variant could best explain a trait (48), which potentially leads to the loss of signals that may contribute a joint effect.

The recent development of methods such as 'Bayesian stochastic search' proved to be more efficient in detecting multiple independent association signals. Bayesian stochastic search tests the question which sets of SNPs can best jointly explain type 1 diabetes association. Wallace et al., applied it to re-analyse the *IL-2RA* region in 2015. They showed that instead of three, there are four groups of SNPs independently associated with type 1 diabetes (48), with group A SNPs located in the intron 1 of *IL2-RA* (that is equivalent to group 1), group C SNPs located in the intergenic region between *IL-2RA* and *RBM17* (equivalent to group 2) that replaced the previous group 3 SNP, group E SNPs located at the 5' of *RBM17*, and group F SNPs reside in the 5' of *RBM17* to intron 2 of *PFKFB3* (Figure 2) (48). For multiple sclerosis, the risk could be explained either by group A and group D SNPs (tagged by rs56382813) jointly or by group B SNP (rs2104286) alone.

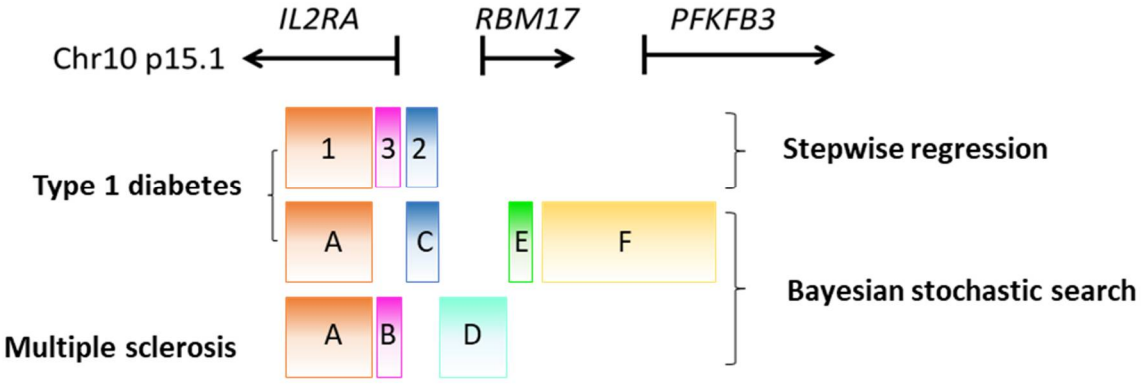

Figure 2: Schematic representation of multiple independent loci at the *IL2RA-RBM17-PFKFB3* region discovered using stepwise regression and Bayesian stochastic search that are associated with both type 1 diabetes and multiple sclerosis. Group 1-3 SNPs that contribute to type 1 diabetes were originally identified using stepwise regression. The later developed Bayesian stochastic search, however, identified four groups of SNPs (A, C, E, F) that jointly explain type 1 diabetes risk at the *IL-2RA* locus; whereas for multiple sclerosis, the risk at this

Peer-reviewed version available at Genes 2018, 9, 377; doi:10.3390/genes9080377

locus was explained by group A and D jointly, or by group B SNPs alone. Figure was adapted and modified from (48).

Epistasis in autoimmune diseases

Genes may not function in isolation. Gene-gene interactions (epistasis) are major contributors to autoimmune disease risk; a classical example is the HLA class II haplotypes. For example, *HLA-DRB1*1501-DQB1*0602* is the most susceptible haplotype for multiple sclerosis (49) and *HLA-DRB1*0301–DQB1*0201* is the most susceptible haplotype for type 1 diabetes (50). A summary of HLA class II haplotypes in autoimmune diseases is reviewed here (51). Class II and class I haplotypes also play a role in autoimmune disease risk. For instance, in type 1 diabetes patients, a combination of *HLA-A*24, DQA1*03,* and *DR9* has been associated with accelerated beta cell loss (52). The interaction between *HLA-DR3/DR4* and class I *A*03* allele demonstrated significant protective effect of clinical progression to type 1 diabetes, whereas *HLA-DR3/DR4* and class I *B*39* interaction contributed significantly to the progression from multiple islet autoantibody to type 1 diabetes (53). HLA also interact with non-HLA genes. For example, an increased risk of SLE risk (OR=1.19) was observed when *CTLA4* (cytotoxic T lymphocyte antigen 4), a negative regulator of T cell response interacts with SNPs rs3131379 and rs1270942 located in the HLA class III region (54). Another similar example was reported in Multiple Sclerosis (MS). A driver of increased MS risk is the soluble form of interleukin-7 receptor (sIL7R). Exon 6 of IL-7R interacts with many protein factors, one of them is encoded by the *DDX39B* gene that is located at the HLA region. Galarza-Munoz et al., recently identified that a SNP rs2523506 within the *DDX39B* region reduces *DDX39B* expression. More importantly, the authors showed a significant increased risk in MS when rs2523506 interacts with a risk variant of IL-7R (rs6897932). Carriers of risk alleles of both SNPs increased sIL7R expression, thereby increasing the risk of MS (55).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 May 2018 doi:10.20944/preprints201805.0160.v1

Peer-reviewed version available at Genes 2018, 9, 377; doi:10.3390/genes9080377

Epigenetic regulation

Epigenetic mechanism is another layer of regulation that influences a gene function. Although the Greek prefix 'epi' indicates an effect that is acting 'on' the genome, epigenotype can influence autoimmune susceptibility in several ways, illustrated in Figure 3.

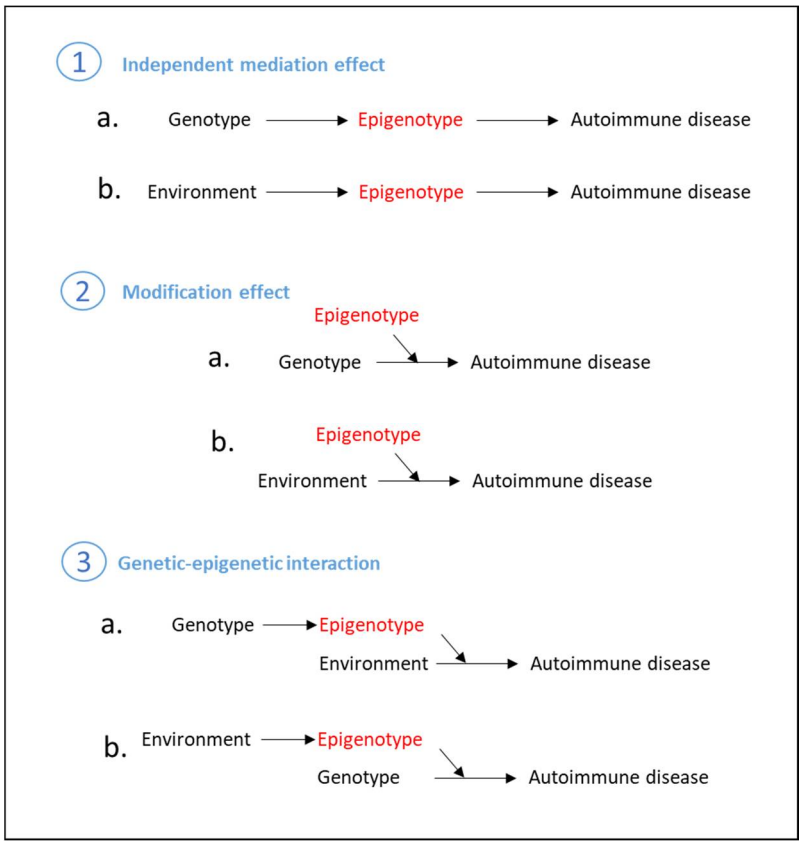

Figure 3: Schematic representation of three modes of actions where epigenetic regulations can take place to contribute to autoimmune disease risk. Figure was adapted and modified from (56).

As shown in the first scenario in Figure 3, epigenetic mechanisms can independently mediate genetic and environmental risk, which subsequently lead to autoimmune disease. Perhaps the most well-studied epigenetic mechanisms involve DNA methylation (DNAm), histone modification, and long non-coding RNAs. DNA methylation occurs at the cytosine-phosphate-guanine (57) residues and it has been shown that approximately 20 % of DNAm variance is explained by additive genetic variance (58). SNPs that are in close proximity to the CpG site

(cis, <1Mb centred around the SNP) as well as SNPs that are far away from the CpG site (trans, >1Mb centred around the SNP) can both influence DNA methylation levels (59). These SNPs are known as methylation quantitative trait loci, or mQTLs. Many early studies attempted to find mQTL associated DNAm changes by performing epigenome-wide association analyses (60-62). However, one needs to carefully interpret these results because it is difficult to discriminate causality (that DNAm is causally influencing a trait) from reverse causation (that DNAm is a consequence of changes in gene expression), LD confounding (that causal variants and mQTLs are simply in linkage disequilibrium) or horizontal pleiotropy (that mQTLs alters DNAm and T1D via different mechanisms) from these associations (63). Recent developments in statistical strategies such as causal inference test (CIT) (64), Mendelian randomization (MR) (65) (66), and co-localization fine mapping (63) have made such discriminations possible.

Liu and colleagues applied a CIT test on a cohort of 354 rheumatoid arthritis patients and 337 controls to investigate whether genetic risk of rheumatoid arthritis is mediated by DNAm. They identified a large number of differentially methylated CpG sites within the HLA region (535 SNP-CpG pairs) potentially exhibit mediatory effect. A similar observation was made at a non-HLA locus, within the *GSTA2* (Glutathione S-Transferase Alpha 2) gene (67). Using CIT test, DNAm was also found to potentially mediate the expression of a number of genes in human islets, including *HLA-DQB1*, the main predisposing gene to type 1 diabetes (68). In our recent work, we applied Mendelian Randomization together with genetic co-localization fine-mapping to study whether DNAm mediates the genetic risk of type 1 diabetes. We identified a number of loci, including *CTSH, PTPN2* and *AFF3*, where DNAm is potentially on the causal pathway to T1D (69). Richardson et al., further extended this statistical framework to systematically investigate the functional roles of DNAm in hundreds of traits including other autoimmune diseases (70). DNAm appeared to increase autoimmune disease risk in a number

of susceptible loci, where inflammatory bowel disease had the greatest number of DNAm mediated loci; followed by rheumatoid arthritis and Crohn's disease (70).

In the second scenario, an epigenotype is thought to modulate a genetic / environmental risk factor. A typical example of this is the regulation by long non-coding RNAs. Long non-coding RNAs (lncRNAs) are RNAs that exceed 200 nucleotides in length and they are broadly classified into five subclasses: stand-alone lncRNAs (or large intergenic non-coding RNAs, lincRNAs), natural antisense transcripts (NATs), pseudogenes-derived lncRNAs (71), long-intronic ncRNAs, and promoter / enhancer – associated ncRNAs (72). lncRNAs have been shown to involve in anti-viral responses (73), T cell differentiation (74), and NFkB signalling (75). Mirza et al., overlapped known lncRNAs with susceptible variants for inflammatory bowel disease and type 1 diabetes based on their physical locations of the genome. They identified over 2,000 inflammatory bowel disease associated SNPs physically located within 468 lncRNAs and over 1,000 type 1 diabetes SNPs within 247 lncRNAs; many of them potentially disrupt the secondary structure of lncRNAs (76). The authors therefore hypothesized that some of the autoimmune disease associated SNPs can alter the expression and function of lncRNAs, which subsequently influence disease related genes. This hypothesis was systematically investigated using statistical approaches by Kumar et al., where the authors found that disease associated expression quantitative trait loci (eQTLs) affect 112 out of 2,140 lncRNAs in whole blood (77). Hrdlickova et al., mapped lncRNAs expressed in seven immune cell types (granulocytes, monocytes, NK cells, B cells, memory T cells and naïve CD8+ cells) to susceptible loci in nine autoimmune diseases (78). They found that the proportion of lncRNAs expressed in autoimmune disease loci were significantly higher than the proportion of lncRNAs expressed genome-wide; similarly, the expression levels of lncRNAs in autoimmune disease loci were higher than that detected at the genome-wide level (78). In

Peer-reviewed version available at Genes 2018, 9, 377; doi:10.3390/genes9080377

addition, lncRNAs overlapping the disease regions tend to be tissue specific. For example, inflammatory bowel disease associated lncRNAs are preferentially expressed in NK cells, juvenile idiopathic arthritis associated lncRNAs are enriched in memory and CD8+ T cells (78).

The third scenario elucidates the most complicated situation, which is the genetic-epigenetic interaction. This type of interaction has been seen in many occasions involving enhancer regulations. Enhancers are defined as cis-acting DNA sequences that can increase the transcription of genes. 98% of enhancers are located in the non-coding regions of the genome, either upstream or downstream of genes, or in introns. Enhancers can be identified using high-throughput sequencing targeting specific markers such as H3 acetylated at lysine 27 (H3K27ac) and H3 monomethylated at K4 (H3K4me1), which are chemical modifications of the histone proteins that wrap around DNA. Their activities can also be specific to a tissue or a particular cell type, a time-point in life, or a unique physiological state (79). A recent study overlapped causal variants with histone marks in 21 common autoimmune diseases found that causal variants of a disease trait are enriched in enhancers specific to disease affecting tissues. For example, causal variants of Alzheimer's disease are enriched in enhancers in brain tissues; causal variants of type 1 diabetes are enriched in enhancers in lymphocytes as well as pancreatic islets, etc (80). In addition, a disproportionate number of enhancers respond to *ex vivo* stimulation, reflected by increased H3K27Ac (marks active promoter and enhancer) signals and non-coding RNA transcription upon immune cell activation (80).

A specific example of genetic-epigenetic interaction was described in Grave's disease. One hypothesis for the initiation of Grave's disease is viral infection. Infection can lead to the

recognition of auto-antigens via molecular mimicry, which further cause bystandard activation of auto-reactive T cells and global pro-inflammatory cytokine production (81, 82) (83). To mimic the consequence of viral infection, Stefan et al., treated human thyroid cells with pro-inflammatory cytokines (84). They observed that thyroid cells exerted significant changes of H3K4me1 signatures at the intron1 of *TSHR* gene, which harbours a previously predicted causal variant rs12101261 to Grave's disease. After pro-inflammatory cytokine treatment, rs12101261 was able to interact with histone deacetylase and a transcription repressor PLZF, resulting in reduced *TSHR* expression and breakdown of central tolerance (84).

An immediate problem of studying genetic-epigenetic interaction is that only 10-20% causal variants were predicted to disrupt transcription factor binding motifs at the enhancer sites, 80-90% causal variants function by modifying the non-classical regulatory sequence (80, 85). In addition, the nature of stimulus dependent enhancer interactions makes it challenging to robustly study them in un-stimulated cells. The advances in Clustered Regulatory Interspaced Short Palindromic Repeats (CRISPR) – Cas9 technologies enabled their screening and characterization more readily in un-stimulated cells. CRISPR activation (CRISPRa) utilises guide RNAs conjugated with a strong transcriptional activator (i.e. VP64) to induce the expression of endogenously weakly expressed genes (86). Simeonov et al., recently used this approach to scan enhancers surrounding the *CD69* and *IL-2RA* genes (87). They identified a CRISPR responsive enhancer at the intronic region of *IL-2RA*, and confirmed the ability of the candidate causal variant rs61839660 (group A SNP) to disrupt this enhancer activity in a stimulus dependent manner. Using mouse models, they subsequently showed that upon T cell stimulation, this enhancer controls CD4 naïve T cell polarization, as CD4+ naïve T cells in the enhancer deletion mouse strain tend to favour a pro-inflammatory Th17 cell

differentiation rather than T regulatory cell differentiation (87). Interestingly, enhancer disruption of *IL-2RA* delayed its expression, which was eventually recovered three days after T cell stimulation (87), implying that the induction of autoimmunity could happen in a transient and tissue specific manner.

Conclusion remarks

With GWAS studies, we gained significant knowledge to broadly define autoimmune diseases associated regions genome-wide. Although GWAS have their inherent limitations, it was a big step forward considering that the concept of autoimmunity was initially proposed during the 1940s (88) and the HLA associations were only first described in the 1970s. The challenges for the next decade is to precisely characterise the functions of disease risk loci. With fast-growing profiling of the genomic and epigenomic regulatory elements, as well as the cutting edge bioinformatic and molecular genetic tools, we are in a unique position to study complex loci. Only by understanding the complex and dynamic nature of autoimmune disease mechanisms, we can deliver truly translational research that impact future clinical care.

References

1. Campbell AW (2014) Autoimmunity and the gut. *Autoimmune Dis* 2014:152428.
2. Barrett JC, *et al.* (2009) Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat Genet* 41(6):703-707.
3. Kappelman MD, Moore KR, Allen JK, & Cook SF (2013) Recent trends in the prevalence of Crohn's disease and ulcerative colitis in a commercially insured US population. *Dig Dis Sci* 58(2):519-525.
4. Benito-Leon J (2011) Are the prevalence and incidence of multiple sclerosis changing? *Neuroepidemiology* 36(3):148-149.
5. Myasoedova E, Crowson CS, Kremers HM, Therneau TM, & Gabriel SE (2010) Is the incidence of rheumatoid arthritis rising?: results from Olmsted County, Minnesota, 1955-2007. *Arthritis Rheum* 62(6):1576-1582.

6. Ercolini AM & Miller SD (2009) The role of infections in autoimmune disease. *Clin Exp Immunol* 155(1):1-15.
7. Manzel A, *et al.* (2014) Role of "Western diet" in inflammatory autoimmune diseases. *Curr Allergy Asthma Rep* 14(1):404.
8. Vieira SM, Pagovich OE, & Kriegel MA (2014) Diet, microbiota and autoimmune diseases. *Lupus* 23(6):518-526.
9. Fronczak CM, *et al.* (2003) In utero dietary exposures and risk of islet autoimmunity in children. *Diabetes Care* 26(12):3237-3242.
10. Singal DP & Blajchman MA (1973) Histocompatibility (HL-A) antigens, lymphocytotoxic antibodies and tissue antibodies in patients with diabetes mellitus. *Diabetes* 22(6):429-432.
11. Cudworth AG & Woodrow JC (1974) Letter: HL-A antigens and diabetes mellitus. *Lancet* 2(7889):1153.
12. Stastny P (1976) Mixed lymphocyte cultures in rheumatoid arthritis. *J Clin Invest* 57(5):1148-1157.
13. Moser KL, Kelly JA, Lessard CJ, & Harley JB (2009) Recent insights into the genetic basis of systemic lupus erythematosus. *Genes Immun* 10(5):373-379.
14. Schlosstein L, Terasaki PI, Bluestone R, & Pearson CM (1973) High association of an HL-A antigen, W27, with ankylosing spondylitis. *N Engl J Med* 288(14):704-706.
15. Gutierrez-Arcelus M, Rich SS, & Raychaudhuri S (2016) Autoimmune diseases - connecting risk alleles with molecular traits of the immune system. *Nat Rev Genet* 17(3):160-174.
16. Wellcome Trust Case Control C (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447(7145):661-678.
17. Long T, *et al.* (2017) Whole-genome sequencing identifies common-to-rare variants associated with human blood metabolites. *Nat Genet* 49(4):568-578.
18. de Vries PS, *et al.* (2017) Comparison of HapMap and 1000 Genomes Reference Panels in a Large-Scale Genome-Wide Association Study. *PLoS One* 12(1):e0167742.
19. Li YR, *et al.* (2015) Meta-analysis of shared genetic architecture across ten pediatric autoimmune diseases. *Nat Med* 21(9):1018-1027.
20. Marees AT, *et al.* (2018) A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. *Int J Methods Psychiatr Res*.
21. Wray NR, *et al.* (2013) Pitfalls of predicting complex traits from SNPs. *Nat Rev Genet* 14(7):507-515.
22. Sorosina M, *et al.* (2015) Inverse correlation of genetic risk score with age at onset in bout-onset and progressive-onset multiple sclerosis. *Mult Scler* 21(11):1463-1467.
23. Patel KA, *et al.* (2016) Type 1 Diabetes Genetic Risk Score: A Novel Tool to Discriminate Monogenic and Type 1 Diabetes. *Diabetes* 65(7):2094-2099.
24. Oram RA, *et al.* (2016) A Type 1 Diabetes Genetic Risk Score Can Aid Discrimination Between Type 1 and Type 2 Diabetes in Young Adults. *Diabetes Care* 39(3):337-344.
25. Johnson MB, *et al.* (2018) A type 1 diabetes genetic risk score can discriminate monogenic autoimmunity with diabetes from early-onset clustering of polygenic autoimmunity with diabetes. *Diabetologia* 61(4):862-869.
26. Polychronakos C & Li Q (2011) Understanding type 1 diabetes through genetics: advances and prospects. *Nat Rev Genet* 12(11):781-792.
27. Zaitlen N & Kraft P (2012) Heritability in the genome-wide association era. *Hum Genet* 131(10):1655-1664.

28. Li YR, *et al.* (2015) Genetic sharing and heritability of paediatric age of onset autoimmune diseases. *Nat Commun* 6:8442.
29. Hunt KA, *et al.* (2013) Negligible impact of rare autoimmune-locus coding-region variants on missing heritability. *Nature* 498(7453):232-235.
30. Zuk O, Hechter E, Sunyaev SR, & Lander ES (2012) The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc Natl Acad Sci U S A* 109(4):1193-1198.
31. Kilpinen H & Dermitzakis ET (2012) Genetic and epigenetic contribution to complex traits. *Hum Mol Genet* 21(R1):R24-28.
32. Ricano-Ponce I, *et al.* (2016) Refined mapping of autoimmune disease associated genetic variants with gene expression suggests an important role for non-coding RNAs. *J Autoimmun* 68:62-74.
33. Strachan T, Read AP, & Strachan T (2011) *Human molecular genetics* (Garland Science, New York) 4th Ed pp xxv, 781 p.
34. Campbell CD, *et al.* (2011) Population-genetic properties of differentiated human copy-number polymorphisms. *Am J Hum Genet* 88(3):317-332.
35. Conrad DF, *et al.* (2010) Origins and functional impact of copy number variation in the human genome. *Nature* 464(7289):704-712.
36. Stead JD, Buard J, Todd JA, & Jeffreys AJ (2000) Influence of allele lineage on the role of the insulin minisatellite in susceptibility to type 1 diabetes. *Hum Mol Genet* 9(20):2929-2935.
37. Day IN, *et al.* (2006) Questioning INS VNTR role in obesity and diabetes: subclasses tag IGF2-INS-TH haplotypes; and -23HphI as a STEP (splicing and translational efficiency polymorphism). *Physiol Genomics* 28(1):113.
38. Bronstad I, Wolff AS, Lovas K, Knappskog PM, & Husebye ES (2011) Genome-wide copy number variation (CNV) in patients with autoimmune Addison's disease. *BMC Med Genet* 12:111.
39. Yang Y, *et al.* (2007) Gene copy-number variation and associated polymorphisms of complement component C4 in human systemic lupus erythematosus (SLE): low copy number is a risk factor for and high copy number is a protective factor against SLE susceptibility in European Americans. *Am J Hum Genet* 80(6):1037-1054.
40. Rodríguez S (2009) *Genetic endocrinology of the metabolic syndrome* (Nova Biomedical Books, New York) pp vi, 103 p.
41. International Multiple Sclerosis Genetics C, *et al.* (2011) Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* 476(7359):214-219.
42. Stahl EA, *et al.* (2010) Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nat Genet* 42(6):508-514.
43. Boyman O & Sprent J (2012) The role of interleukin-2 during homeostasis and activation of the immune system. *Nat Rev Immunol* 12(3):180-190.
44. Vella A, *et al.* (2005) Localization of a type 1 diabetes locus in the IL2RA/CD25 region by use of tag single-nucleotide polymorphisms. *Am J Hum Genet* 76(5):773-779.
45. Lowe CE, *et al.* (2007) Large-scale genetic fine mapping and genotype-phenotype associations implicate polymorphism in the IL2RA region in type 1 diabetes. *Nat Genet* 39(9):1074-1082.
46. Maier LM, *et al.* (2009) IL2RA genetic heterogeneity in multiple sclerosis and type 1 diabetes susceptibility and soluble interleukin-2 receptor production. *PLoS Genet* 5(1):e1000322.

47. Dendrou CA, *et al.* (2009) Cell-specific protein phenotypes for the autoimmune locus IL2RA using a genotype-selectable human bioresource. *Nat Genet* 41(9):1011-1015.
48. Wallace C, *et al.* (2015) Dissection of a Complex Disease Susceptibility Region Using a Bayesian Stochastic Search Approach to Fine Mapping. *PLoS Genet* 11(6):e1005272.
49. Oksenberg JR, *et al.* (2004) Mapping multiple sclerosis susceptibility to the HLA-DR locus in African Americans. *Am J Hum Genet* 74(1):160-167.
50. Noble JA, *et al.* (2008) Type 1 diabetes risk for human leukocyte antigen (HLA)-DR3 haplotypes depends on genotypic context: association of DPB1 and HLA class I loci among DR3- and DR4-matched Italian patients and controls. *Hum Immunol* 69(4-5):291-300.
51. Gough SC & Simmonds MJ (2007) The HLA Region and Autoimmune Disease: Associations and Mechanisms of Action. *Curr Genomics* 8(7):453-465.
52. Nakanishi K & Inoko H (2006) Combination of HLA-A24, -DQA1*03, and -DR9 contributes to acute-onset and early complete beta-cell destruction in type 1 diabetes: longitudinal study of residual beta-cell function. *Diabetes* 55(6):1862-1868.
53. Lipponen K, *et al.* (2010) Effect of HLA class I and class II alleles on progression from autoantibody positivity to overt type 1 diabetes in children with risk-associated class II genotypes. *Diabetes* 59(12):3253-3256.
54. Hughes T, *et al.* (2012) Evidence for gene-gene epistatic interactions among susceptibility loci for systemic lupus erythematosus. *Arthritis Rheum* 64(2):485-492.
55. Galarza-Munoz G, *et al.* (2017) Human Epistatic Interaction Controls IL7R Splicing and Increases Multiple Sclerosis Risk. *Cell* 169(1):72-84 e13.
56. Ladd-Acosta C & Fallin MD (2016) The role of epigenetics in genetic and environmental epidemiology. *Epigenomics* 8(2):271-283.
57. Jiang X, *et al.* (2018) Genome-wide association study in 79,366 European-ancestry individuals informs the genetic architecture of 25-hydroxyvitamin D levels. *Nat Commun* 9(1):260.
58. Gaunt TR, *et al.* (2016) Systematic identification of genetic influences on methylation across the human life course. *Genome Biol* 17:61.
59. Richardson TG, *et al.* (2016) Collapsed methylation quantitative trait loci analysis for low frequency and rare variants. *Hum Mol Genet* 25(19):4339-4349.
60. Imgenberg-Kreuz J, *et al.* (2016) Genome-wide DNA methylation analysis in multiple tissues in primary Sjogren's syndrome reveals regulatory effects at interferon-induced genes. *Ann Rheum Dis* 75(11):2029-2036.
61. Lemire M, *et al.* (2015) Long-range epigenetic regulation is conferred by genetic variation located at thousands of independent loci. *Nat Commun* 6:6326.
62. Belot MP, *et al.* (2013) CpG methylation changes within the IL2RA promoter in type 1 diabetes of childhood onset. *PLoS One* 8(7):e68093.
63. Richardson TG, *et al.* (2017) Mendelian Randomization Analysis Identifies CpG Sites as Putative Mediators for Genetic Influences on Cardiovascular Disease Risk. *Am J Hum Genet* 101(4):590-602.
64. Millstein J, Zhang B, Zhu J, & Schadt EE (2009) Disentangling molecular relationships with a causal inference test. *BMC Genet* 10:23.
65. Relton CL & Davey Smith G (2012) Two-step epigenetic Mendelian randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease. *Int J Epidemiol* 41(1):161-176.
66. Relton CL & Davey Smith G (2015) Mendelian randomization: applications and limitations in epigenetic studies. *Epigenomics* 7(8):1239-1243.

67. Liu Y, *et al.* (2013) Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nat Biotechnol* 31(2):142-147.
68. Olsson AH, *et al.* (2014) Genome-wide associations between genetic and epigenetic variation influence mRNA expression and insulin secretion in human pancreatic islets. *PLoS Genet* 10(11):e1004735.
69. Ye J, *et al.* (2018) Identification of loci where DNA methylation potentially mediate genetic risk of type 1 diabetes. *bioRxiv*.
70. Richardson TG, *et al.* (2017) Systematic Mendelian randomization framework elucidates hundreds of genetic loci which may influence disease through changes in DNA methylation levels. *bioRxiv*.
71. Milligan MJ & Lipovich L (2014) Pseudogene-derived lncRNAs: emerging regulators of gene expression. *Front Genet* 5:476.
72. Kung JT, Colognori D, & Lee JT (2013) Long noncoding RNAs: past, present, and future. *Genetics* 193(3):651-669.
73. Liu W & Ding C (2017) Roles of LncRNAs in Viral Infections. *Front Cell Infect Microbiol* 7:205.
74. Spurlock CF, 3rd, *et al.* (2015) Expression and functions of long noncoding RNAs during human T helper cell differentiation. *Nat Commun* 6:6932.
75. Pearson MJ & Jones SW (2016) Review: Long Noncoding RNAs in the Regulation of Inflammatory Pathways in Rheumatoid Arthritis and Osteoarthritis. *Arthritis Rheumatol* 68(11):2575-2583.
76. Mirza AH, Kaur S, Brorsson CA, & Pociot F (2014) Effects of GWAS-associated genetic variants on lncRNAs within IBD and T1D candidate loci. *PLoS One* 9(8):e105723.
77. Kumar V, *et al.* (2013) Human disease-associated genetic variation impacts large intergenic non-coding RNA expression. *PLoS Genet* 9(1):e1003201.
78. Hrdlickova B, *et al.* (2014) Expression profiles of long non-coding RNAs located in autoimmune disease-associated regions reveal immune cell-type specificity. *Genome Med* 6(10):88.
79. Pennacchio LA, Bickmore W, Dean A, Nobrega MA, & Bejerano G (2013) Enhancers: five essential questions. *Nat Rev Genet* 14(4):288-295.
80. Farh KK, *et al.* (2015) Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* 518(7539):337-343.
81. Carella C, Mazziotti G, Amato G, Braverman LE, & Roti E (2004) Clinical review 169: Interferon-alpha-related thyroid disease: pathophysiological, epidemiological, and clinical aspects. *J Clin Endocrinol Metab* 89(8):3656-3661.
82. Cabrera SM, Chen YG, Hagopian WA, & Hessner MJ (2016) Blood-based signatures in type 1 diabetes. *Diabetologia* 59(3):414-425.
83. Psarras A, Emery P, & Vital EM (2017) Type I interferon-mediated autoimmune diseases: pathogenesis, diagnosis and targeted therapy. *Rheumatology (Oxford)* 56(10):1662-1675.
84. Stefan M, *et al.* (2014) Genetic-epigenetic dysregulation of thymic TSH receptor gene expression triggers thyroid autoimmunity. *Proc Natl Acad Sci U S A* 111(34):12562-12567.
85. Lin JD & Cheng CW (2017) Causal variants in autoimmune disease: a commentary on a recent published fine-mapping algorithm analysis in genome-wide association studies study. *Ann Transl Med* 5(6):151.
86. Perez-Pinera P, *et al.* (2013) RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat Methods* 10(10):973-976.

87. Simeonov DR, *et al.* (2017) Discovery of stimulation-responsive immune enhancers with CRISPR activation. *Nature* 549(7670):111-115.
88. Anderson W & Mackay IR (2014) *Intolerant bodies : a short history of autoimmunity* (Johns Hopkins University Press, Baltimore) pp xii, 250 pages.