REVIEW

Genetics of autoimmune diseases: insights from population genetics

Paula S Ramos[1], Andrew M Shedlock[2,3] and Carl D Langefeld[4]

Human genetic diversity is the result of population genetic forces. This genetic variation influences disease risk and contributes to health disparities. Autoimmune diseases (ADs) are a family of complex heterogeneous disorders with similar underlying mechanisms characterized by immune responses against self. Collectively, ADs are common, exhibit gender and ethnic disparities, and increasing incidence. As natural selection is an important influence on human genetic variation, and immune function genes are enriched for signals of positive selection, it is thought that the prevalence of AD risk alleles seen in different population is partially the result of differing selective pressures (for example, due to pathogens). With the advent of high-throughput technologies, new analytical methodologies and large-scale projects, evidence for the role of natural selection in contributing to the heritable component of ADs keeps growing. This review summarizes the genetic regions associated with susceptibility to different ADs and concomitant evidence for selection, including known agents of selection exerting selective pressure in these regions. Examples of specific adaptive variants with phenotypic effects are included as an evidence of natural selection increasing AD susceptibility. Many of the complexities of gene effects in different ADs can be explained by population genetics phenomena. Integrating AD susceptibility studies with population genetics to investigate how natural selection has contributed to genetic variation that influences disease risk will help to identify functional variants and elucidate biological mechanisms. As such, the study of population genetics in human population holds untapped potential for elucidating the genetic causes of human disease and more rapidly focusing to personalized medicine.
Journal of Human Genetics (2015) **60**, 657–664; doi:10.1038/jhg.2015.94; published online 30 July 2015

INTRODUCTION

Autoimmune diseases (ADs) are a family of more than 80 chronic, and often disabling, illnesses broadly characterized by immune system dysfunction leading to the loss of tolerance to self-antigens, presence of increased level of autoantibodies, inflammatory and mediatory cells and resulting chronic inflammation. The family of ADs is remarkable for its complexity and similar underlying mechanisms despite the involvement of multiple organ systems.[1] Patients often endure lifelong debilitating symptoms, loss of organ function, reduced productivity at work and high medical expenses. Importantly here, as many ADs present before or during a woman's reproductive years, they can have effects on fetal[2,3] and maternal outcomes, such as pregnancy loss in women with systemic lupus erythematosus,[3,4] vasculitis[3] and type 1 diabetes,[5] and infertility in women with rheumatoid arthritis.[4] Collectively, ADs are common, affecting between five and nine percent of the population and are of considerable personal and public health burdens.[1,6] The reasons for their high prevalence, gender and ethnic disparities and rising incidence and prevalence remain unclear.[1]

The genetic basis of disease is influenced by individual and population variation. Population-level phenomena such as mutation, migration, genetic drift and natural selection, have left an imprint on genetic variation that is likely to influence phenotypic expression in specific populations.[7] Given its role in driving genetic variation, population genetics can help elucidate human genetic diversity and, consequently, disease etiology.

Left untreated, most ADs affect the ability to raise offspring that successfully reproduce and result in reduced reproductive fitness. Thus, alternative forces must exist that permit the relative high frequency of risk alleles. As immune and inflammatory responses can be highly sensitive to environmental change,[8] evolutionary adaptation to specific environments might have driven selection on immune-related genetic variants, impacting variant frequencies and leaving signatures of selection in the genome. Given that infectious organisms are strong agents of natural selection,[9,10] it is plausible that alleles selected for protection against infection confer increased risk of autoimmune and allergic diseases, or their complications,[11] as the 'hygiene hypothesis'[12] postulates. It is thought that the adaptation to pathogen pressure through functional variation in immune-related genes conferred a specific selective advantage for host survival, including protection from pathogens and tolerance to microbiota.[13]

[1]Division of Rheumatology and Immunology, Department of Medicine, and Department of Public Health Sciences, Medical University of South Carolina, Charleston, SC, USA; [2]Department of Biology, College of Charleston, Charleston, SC, USA; [3]Hollings Marine Laboratory Center for Marine Biomedicine and College of Graduate Studies, Medical University of South Carolina, Charleston, SC, USA and [4]Division of Public Health Sciences, Department of Biostatistical Sciences; and Center for Public Health Genomics, Wake Forest School of Medicine, Winston-Salem, NC, USA
Correspondence: Dr PS Ramos, Division of Rheumatology and Immunology, Department of Medicine, Medical University of South Carolina, 96 Jonathan Lucas Street, Suite 912, Charleston, SC 29425, USA.
E-mail: ramosp@musc.edu

However, the emergence of such variation conferring resistance to pathogens is also influencing AD risk in specific populations.

In the past decade, multiple genome scans for signatures of selection on common variation have identified many immune-related loci.[14–18] Similarly, over 130 genome-wide association studies have established AD-associated alleles. The evidence that AD-associated variants are also under selection is growing.[18–23] This review summarizes the evidence for AD-associated loci under selection and the candidate selective pressures. Collectively, these results support the hypothesis that AD is, in part, a pleiotropic consequence of alleles conferring pathogen resistance[24] and help explain the number of shared auto-immune variants conferring both disease risk and protection, as well as the health disparities in individuals and populations. Given that genomic variation can have clinically important consequences,[7] elucidating the patterns of variation and the functional role of the selective pressure might contribute to a better understanding of disease etiology and the development of new therapies for improved disease management.

GENETIC ETIOLOGY OF ADs

The family of ADs is remarkable for its complexity and similar underlying mechanisms despite the involvement of multiple organ systems.[1] Most ADs exhibit marked gender and ethnic disparities. They disproportionately afflict women and are among the leading causes of death for young and middle-aged women.[1] African Americans are at higher risk than European Americans for systemic lupus erythematosus and scleroderma (systemic sclerosis), which they tend to develop earlier in life and experience more severe disease, but are at lower risk for type 1 diabetes, thyroiditis and multiple sclerosis.[1] Although most of these diseases are individually rare, in the aggregate they affect between five and nine percent of the population,[1,6] and, for reasons that are poorly understood, their incidence and prevalence is rising.[1] Despite the variation in prevalence, incidence and disease severity that are known to vary among ethnic groups, little is known about the genetic etiology of these diseases in the different populations and the reasons for the ethnic disparities remain elusive.

Multiple lines of evidence suggest some degree of common genetic etiology in ADs, including clustering of multiple ADs in families and in individuals, and the number of confirmed genetic regions predisposing to several ADs. This genetic overlap is exemplified by the well-known associations of the human leukocyte antigen (HLA) region with all ADs, as well as other loci associated with multiple AD, such as *IL23R*, *TNFAIP3* and *IL2RA*.[25] Nevertheless, the genetic heritability of ADs is extremely variable, ranging from very high in Crohn's disease or ankylosing spondylitis to almost negligible in systemic sclerosis.[26] Genome-wide association studies have proved particularly powerful for ADs.[27] Table 1 summarizes the ADs with published genome-wide association studies and the number of disease-associated loci uncovered from these genome-wide association studies.

It is the general consensus that there is a common genetic background predisposing to autoimmunity, and that further combinations of more disease-specific variation at HLA and non-HLA genes, in interaction with epigenetic and environmental factors, contribute to disease and its clinical manifestations. Results from the most comprehensive evaluation of shared genetic autoimmune loci to date suggest that instead of resulting from common risk factors, auto-immunity may result from specific and multiple different pleiotropic effects.[25] This suggests that different population genetic factors (for example, natural selection with coevolution with pathogens, random mutation, isolations, migrations and interbreeding) in similar or distinct environments led to the establishment of the current plethora

Table 1 Autoimmune diseases with published GWAS and respective number of associated loci

| Autoimmune disease | Number of | |
	GWAS	loci
Alopecia areata	1	8
Ankylosing spondylitis	3	21
Behçet's disease	4	10
Celiac disease	5	60
Crohn's disease	15	110
Graves' disease	3	13
Granulomatosis with polyangiitis	1	6
Inflammatory bowel disease	4	117
Juvenile idiopathic arthritis	3	5
Kawasaki disease	6	16
Multiple sclerosis	15	22
Myasthenia gravis	1	5
Primary biliary cirrhosis	4	25
Primary sclerosing cholangitis	1	2
Psoriasis	7	23
Psoriatic arthritis	2	4
Rheumatoid arthritis	18	117
Sarcoidosis	2	3
Sjögren's syndrome	1	4
Systemic lupus erythematosus	11	73
Systemic sclerosis	3	10
Type 1 diabetes	9	60
Ulcerative colitis	10	73
Vitiligo	7	32

Abbreviation: GWAS, genome-wide association studies.
Numbers compiled from the National Human Genome Research Institute's Catalog of Published Genome-Wide Association Studies (http://www.genome.gov/gwastudies) accessed on 27 March 2015.[94]

of loci that predispose to autoimmunity.[25] As the authors conclude, these results generate an appreciation for potential interplay between population genetic factors and environmental factors.[25] This appreciation has been renewed in a recent systematic analysis of loci shared among multiple ADs, confirming that SNP or haplotype sharing is complex, that often the most associated variant at a given locus differs and, even when shared, the same allele often has opposite associations.[28] Population-level phenomena are the likely reason behind this complexity of gene effects in different ADs.

NATURAL SELECTION AND ADAPTATION

Human genome variation at the population level is typically thought to be shaped by five evolutionary processes: sexual reproduction, mutation, migration, random genetic drift and natural selection. In addition, recombination is often an underappreciated driver of genomic diversity and genome evolution,[29] as recombination events (the exchange of genetic material between homologous chromosomes) are accompanied by biased gene conversion, which causes allele frequency shifts that mimic the effects of natural selection,[30,31] shapes patterns of diversity in human genomes and contributes to substantially increased risks of hereditary disease.[31] Natural selection is the process by which a trait becomes either more or less common in a population depending on the differential reproductive success of those with the trait. Natural selection drives adaptation, the evolutionary process whereby over generations the members of a population become better suited to survive and reproduce in that environment.

Negative (or purifying) selection decreases the prevalence of traits by purging deleterious alleles, and is the most common mechanism of selection. Negatively selected variants are usually associated with rare Mendelian disorders and have low population frequencies. As an example, endosomal Toll-like receptors (TLR3, TLR7, TLR8 and TLR9), which are involved in the sensing of nucleic acids, mostly from viruses, have been subject to negative selection.[32]

Positive selection increases the prevalence of adaptive traits by increasing the frequency of favorable alleles. Positively selected variation is often associated with common complex traits and can achieve higher population frequencies.[33] A classic example of positive selection is the lactase (LCT) gene, in which independent variants associated with lactase persistence raised to high frequency in different populations owing to the strong selective force of adult milk consumption.[34] The enrichment for signals of positive selection among genes associated with complex traits is well documented.[15,35–37]

Balancing selection favors genetic diversity by retaining variation in the population as a result of heterozygote advantage and frequency-dependent advantage. Few loci show the signature of this type of selection. One classical example is the hemoglobin-β locus and the high prevalence of the hemoglobin-β mutation that causes sickle cell anemia in regions where malaria is endemic: while individuals homozygous for the wild-type hemoglobin-β are susceptible to malaria, and individuals homozygous for the hemoglobin-β mutation have sickle cell anemia, heterozygous individuals are resistant to malaria and otherwise healthy. Another pertinent example of balancing selection is the HLA (also known as major histocompatibility complex (MHC)) region,[38,39] where highly polymorphic loci have a central role in the recognition and presentation of antigens to the immune system. The high levels of polymorphism are the results of pathogen-driven balancing selection.[40] The heterozygote advantage against multiple pathogens contributes to the evolution of HLA diversity, which confers resistance against multiple pathogens and explains the persistence of alleles conferring susceptibility to AD.[41] Nevertheless, there is also recent evidence that positive selection might be acting on specific HLA alleles in a deme due to unique environmental pressures.[42]

Natural selection leaves a distinctive molecular signature in the targeted genomic region. Different statistical methods have been developed to detect signatures of selection. Quintana-Murci and Clark[13] provide a comprehensive description of popular statistical methods to detect the different types of selection. There are three main classes of tests that detect regions under selection: (i) site frequency spectrum tests summarize the allele frequency distribution of polymorphisms, (ii) linkage disequilibrium tests uncover high-frequency haplotypes with extended linkage disequilibrium and (iii) population differentiation statistics detect local increases in the magnitude of population structure due to geographically restricted selection.

Allele frequency spectrum tests (for example, Tajima's D, Fay and Wu's H-test) can detect recent selective events ($< 250\,000$ years) and are powerful for detecting completed or quasi-completed selective sweeps; haplotype-based tests (for example, integrated haplotype score (iHS), cross population-extended haplotype homozygosity (XP-EHH) test) are very powerful to detect very recent events of positive selection ($< 30\,000$ years); population structure statistics (for example, F_{ST} test) are powerful for detecting selective differences among population, especially those that occurred after the out-of-Africa dispersals ($< 60\,000$ years).[13] As recently reviewed,[43] allele frequency spectrum and haplotype-based tests are powered to detect classic selective sweeps, in which a novel adaptive variant arises de novo in a population and

rapidly increases in frequency to (or near) fixation, purging variation at linked sites as they spread. Soft sweeps in which selection arises from variants already polymorphic in a population (standing variation) are more difficult to detect using these tests, especially when an adaptive trait is influenced by multiple loci.[43] However, in a geographically restricted population being subjected to adaptive pressures, selection is likely to favor a weaker signature of adaptation, which may include subtle shifts in allele frequencies at multiple loci, resulting in an excess of allele frequency differentiation.[43] Recent studies suggest that such soft sweeps may have been the most common mechanism of adaptation in recent human evolution.[44,45] Interestingly, immune response pathways show evidence of polygenic adaptation,[44] the process in which adaptation occurs by simultaneous selection on variants at many loci. Hence, methods based on population structure may be more useful in studies of recent human natural selection as they can detect more subtle changes in allele frequency.[46]

As revised, it has been hypothesized that, in addition to genetic (sequence) variation, heritable epigenetic modifications can affect rates of fitness increase, as well as patterns of genotypic and phenotypic change during adaptation.[47] However, the role of epigenetic variation in the response to natural selection has not been formally assessed, as the methodology to test signatures of natural selection on epigenetic variation is just emerging.[48]

NATURAL SELECTION AND ADs

Given that, if untreated, ADs can diminish reproductive potential and impair the ability to raise offspring that successfully reproduce, some evolutionary process must sustain the relative high frequency of risk alleles seen in current populations around the world. As the human genome is shaped by adaptation to environmental pressures at the population level, one plausible reason for the higher frequency of disease risk alleles may be the direct effect of population-specific natural selection. This hypothesis is supported by the experimental evidence for MHC heterozygote superiority against multiple pathogens, a mechanism that would contribute to the evolution of HLA diversity and explain the persistence of alleles conferring susceptibility to disease.[41]

There is compelling evidence that natural selection is acting on a significant fraction of the human genome.[16,49–53] Immune function genes and pathways are consistently reported in tests for natural selection. As a result of several genome-wide scans, over 300 immune-related genes have been suggested as putative targets of positive selection.[14–18] Although the challenge in validating the true signals remain,[54] several genes involved in immune-related functions have been shown to be under selection.[22,55]

Over 40 regions with evidence for selection and associated with at least one AD have been reported (Table 2). Although slightly over half of these regions seem associated with a single disease, close to half are shared among ADs, including the HLA region, PTPN22, TNFSF4, ARHGAP31-CD80, TNIP1 and TYK2. A few ADs have reported loci with signatures of positive selection: inflammatory bowel disease (Crohn's disease and ulcerative colitis),[19] celiac disease[18,20] and systemic lupus erythematosus.[21] Several of the genes in Table 2 show patterns of genetic variation that are consistent with evidence for recent positive selection, such as the PTPN22, ITPR3 and BLK regions,[22] or SNPs in CLEC16A and UHRF1BP1.[23] As discussed in the next section, resistance to protozoa and tuberculosis infection have been implicated as the selective pressures for PTPN22 and UHRF1BP1, respectively. Interestingly, the systemic lupus erythematosus susceptibility allele in UHRF1BP1 is associated with decreased UHRF1BP1

Table 2 Autoimmune disease regions with the evidence for selection and implicated agents of selection

Gene region	Position	Autoimmune disease association	References for evidence of natural selection	Selective pressure	References for pathogen-driven selection
IL23R	1p31.3	IBD, CD, UC, PS, AS and BeD	19	Protozoa	95
PTPN22	1p13.2	SLE, RA, CD, T1D, VT, MG, AITD and UC	21,22	Protozoa	95
TNFRSF18	1p36.33	IBD	19		
ARHGEF2	1q21-q22	CD	95	Protozoa	95
SCAMP3	1q22	CD	95	Protozoa	95
FCGR2B	1q23.3	SLE	77	*Plasmodium falciparum*	77
TNFSF18	1q24.3	IBD	19		
TNFSF4	1q25.1	SLE, MS, RA, CD, CelD and SS	21		
CR1	1q32	SLE, SA	96	*Plasmodium falciparum*	96
TLR5	1q41-q42	SLE	22	*Salmonella enterica* ser. Typhimurium and other exposures	22
TET3-DGUOK	2p13	SLE	21		
IL18RAP	2q12.1	CelD, CD and IBD	20		
IFIH1	2q24.2	PS, VT, T1D and IBD	76,97,98	Antiviral response	76
IL8RA, SLC11A1	2q35	IBD, UC	19	Mycobacterial infection	19
BTLA	3q13.2	RA	22		
ARHGAP31, CD80	3q13.33	MS, CelD, PBC, JIA, SLE and VT	23		
CD86	3q13.33	MS	23		
LEKR1	3q25.31	MS	23		
IL12A	3q25.33	CelD, BeD, MS and PBC	20,23		
IL2, IL21	4q27	CelD, RA, UC, IBD, T1D and AA	23		
PTGER4	5p13.1	CD, UC, IBD, MS and AS	95	Protozoa	95
SLC22A5, IRF1	5q31.1	CD, IBD	23,95	Protozoa	95
TNIP1	5q33.1	SLE, SScl, PS, PSA, IBD and MG	21		
PTTG1	5q33.3	SLE	21		
ITPR3	6p21.31	SLE, T1D, GD and CD	22		
UHRF1BP1	6p21.31	SLE	21,23	*Mycobacterium tuberculosis*	80
HLA	6p22.1-21.32	All ADs	23,99–102	Bacterial infection	40,74,75
POPDC3	6q21	MS	23		
IKZF1	7p12.2	SLE, CD and IBD	21		
HIP1	7q11.23	SLE	23		
BLK	8p23.1	SLE, RA and KA	21,22		
ABO	9q34.2	GD	103	*Plasmodium falciparum, Norwalk virus, Campylobacter jejuni, Helicobacter pylori,* and *Vibrio cholerae*	104–107
VAV2	9q34.2	MS	91		
KLF6	10p15.1	CD	95	Protozoa	95
CD5	11q12.2	RA, IBD and MS	108		
HEBP1	12p13.1	CD	95	Protozoa	95
VDR	12q13.11	IBD	19	Mycobacterial infection	19
SH2B3, NAA25	12q24.12-q24.13	CelD, VT, RA, T1D	20,23	Bacterial infection	20
ITGAM- ITGAX	16p11.2	SLE	21,22		
CLEC16A, CIITA	16p13.13	SLE, MS, T1D, RA and PBC	21,23,109		
NOD2	16q12.1	IBD, CD	19,81,82	Protozoa; mycobacterial infection	19
IRF8	16q24.1	RA, MS, SScl, IBD and UC	110		
NLRP1	17p13.2	AdD, T1D and VT	81		
NSF	17q21	CD	95	Protozoa	95
TYK2	19p13.2	CD, RA, PS, T1D, IBD, MS and SLE	95	Protozoa	95
SAE1	19q13.32	MS	23		
FUT2	19q13.33	CD, MS	23		
TLR7-TLR8	Xp22.2	CelD	32		

Abbreviations: AA, alopecia areata; AD, autoimmune disease; AdD, Addison's disease; AITD, autoimmune thyroid disease; AS, ankylosing spondylitis; BeD, Behçet's disease; CD, Crohn's disease; CelD, celiac disease; GD, Graves' disease; IBD, inflammatory bowel disease; JIA, juvenile idiopathic arthritis; KA, Kawasaki disease; MG, myasthenia gravis; MS, multiple sclerosis; PBC, primary biliary cirrhosis; PS, psoriasis; PsA, psoriatic arthritis; PSC, primary sclerosing cholangitis; RA, rheumatoid arthritis; SA, sarcoidosis; SLE, systemic lupus erythematosus; SS, Sjögren's syndrome; SScl, systemic sclerosis; T1D, type 1 diabetes; UC, ulcerative colitis; VT, vitiligo.
Autoimmune disease associations were reported in the 'references for evidence of natural selection', and/or in the NHGRI GWAS catalog accessed on 27 March 2015 (see legend for Table 1).

RNA expression in different cell subsets, suggesting that the disease risk allele under selection has a regulatory effect.[23] It is noteworthy that alleles under selective pressure can confer increased risk of AD manifestations, as shown by the evidence that variants within the *APOL1* gene associated with resistance to human African trypanosomiasis (aka sleeping sickness) in some African populations predispose to end-stage kidney disease in systemic lupus erythematosus.[11]

AGENTS OF SELECTION

The wide variety of environments inhabited by human populations is likely exerting different selective pressures that lead to adaptation through natural selection. Climatic factors such as altitude, latitude, ultra-violet radiation levels, temperature, as well as diet and pathogens have been reported as agents of selection driving adaptations to these environments and lifestyles.

For example, several correlations between temperature, body mass and metabolic rates have been reported,[56–58] correlations between latitude with skin pigmentation are well-known,[59,60] and latitude combined with specific genetic variation have been reported to increase susceptibility to hypertension.[61,62] Signals of natural selection driven by annual photoperiod variation have been reported for schizophrenia, bipolar disorder and restless leg syndrome risk variants.[63] In a comprehensive genome-wide test for adaptations to continuous climate variables, Hanckock et al.[24] found correlation between SNPs and climate variables, including an enrichment of individual SNPs involved in pigmentation and immune response, as well as for pathways related to UV radiation, infection and immunity, and cancer. They note that their enrichment for pigmentation and immune response phenotypes contrasts with the metabolic traits with common evidence for adaptations to diet, subsistence and ecoregion.[24] In a recent analysis of correlations between genetic risk of multiple diseases and worldwide migration trajectories, the authors report that variants associated with primary biliary cirrhosis, alopecia areata, inflammatory bowel disease, systemic lupus erythematosus, systemic sclerosis, ulcerative colitis and vitiligo, have undergone genetic risk differentiation associated with migration.[64]

In addition to these general correlations between climatic variables and particular traits, examples of selective pressures driving natural selection of specific loci include malaria resistance and β-globin gene mutation,[65] lactose tolerance and the lactase (*LCT*) gene variation,[66,67] skin pigmentation and the *SLC24A5* gene variation,[68] high-altitude tolerance and the *EPAS1* gene variation,[69] and adaptation to cold resistance and the uncoupling protein UCP3.[70] Interestingly, expression QTLs (eQTLs) from immune function and metabolism genes are enriched in signals of environmental adaptation,[71] which highlights the importance of regulatory variations in local adaptation.

Nevertheless, the strongest effect of climate is in shaping the spatial pattern and species diversity of human pathogens,[9] which is directly relevant to AD predisposition. As recently reviewed,[72] in the constant co-evolutionary battle between host and pathogen, pathogens that diminish reproductive potential, either through death or poor health, drive selection on genetic variants that affect pathogen resistance. According to the 'hygiene hypothesis' first proposed by Strachan in 1989,[12] the increased disease prevalence of autoimmune and allergic diseases in industrialized countries may be due to modern society's limited pathogen exposure. The hygiene hypothesis posits that humans have adapted to infectious exposures that were the norm in the past and that exposure was protective against AD. Over many generations environmental pressure may have favored alleles that affect resistance to pathogenic microorganisms, allowing humans to respond to immune system challenges differently but resulting in an increased risk of ADs. As Hancock et al.[24] suggested, it is likely that selection signals in immune-related loci may implicate variants evolving under a model of antagonistic pleiotropy, where the selective pressure was pathogen resistance, and the autoimmune disorder is a pleiotropic consequence of the resistance allele. This could hence be a mechanism explaining the prevalence of AD risk alleles that are common in the population.

Indeed, pathogens have been the main selective pressure through human evolution.[10] In an analysis that included climate, diet regimes and pathogen loads, Fumagalli et al.[10] showed that the diversity of the local pathogenic environment is the predominant driver of local adaptation, and that climate conditions only had a relatively minor role. In addition, they reported an enrichment of genes associated to ADs, such as celiac disease, type 1 diabetes and multiple sclerosis, which supports the hypothesis that some susceptibility alleles for ADs may be maintained in human population due to past selective processes.[10] The enrichment for signals of positive selection in inflammatory-disease susceptibility loci has been recently corroborated.[23]

Recent reviews of selection signatures left by pathogen-exerted pressure, including immune-related genes, can be found elsewhere.[72,73]

Genetic regions associated with susceptibility to different ADs and evidence of selection that has been attributed to host–pathogen coevolution are shown in Table 2. In slightly less than half of the regions with evidence for selection and AD-association, known pathogens have been implicated as the selective pressure. Variation in the *HLA* and *SH2B3* has been reported as a protective factor against bacterial infection.[20,40,74,75] Variants in the *IFIH1* gene, whose protein is a cytoplasmic helicase that recognizes RNA of picornaviruses and mediates induction of interferon response to viral RNA, have been shown to affect IFIH1 function and host antiviral response.[76] In the context of systemic lupus erythematosus (SLE) predisposing loci, Clatworthy et al.[77] has shown that FCGR2B is important in controlling the immune response to *Plasmodium falciparum*, the parasite responsible for the most severe form of malaria, and suggests that the higher frequency of human *FCGR2B* polymorphisms predisposing to SLE in Asians and Africans may be maintained because these variants reduce susceptibility to malaria. Machado et al.[78] has suggested that helminth infection has driven positive selection of *FCGRs* variation. Grossman et al.[22] implicated *Salmonella typhimurium* and other exposures that directionally drive selection of the toll-like receptor 5 (*TLR5*) gene,[79] which is involved in recognition of flagellated bacteria. Unlike endosomal TLRs, such as TLR7 and TLR8, that have been subject to purifying selection, cell-surface TLRs involved in pathogen recognition (TLR1, TLR2, TLR4, TLR5, TLR6 and TLR10) experienced more relaxed constraints.[32] *UHRF1BP1* has been shown to be significantly differentially expressed in primary dendritic cells upon *Mycobacterium tuberculosis* infection,[80] suggesting that response to tuberculosis has shaped genetic variation at this locus. *NOD2* shows complex signatures of selection, with some variation consistent with neutrality[81] and other with balancing or recent positive selection,[19,82] probably reflecting its association with multiple traits (Crohn's disease, ulcerative colitis and mycobacterial infections).[19] Another complex pattern is that observed for the nonsynonymous variant in PTPN22 that increases the risk of SLE, rheumatoid arthritis, type 1 diabetes, vitiligo, autoimmune thyroid disease and ulcerative colitis, but is protective against Crohn's disease, despite ulcerative colitis and Crohn's disease being closely related phenotypes.[28] Karlsson et al.[83] have recently reported that cholera has exerted strong selective pressure on proinflammatory pathways.

It is important to highlight examples where there is strong statistical evidence of selection plus an experimentally validated phenotypic effect likely to predispose to autoimmunity or its complications, as these examples provide the most convincing evidence of positive selection increasing AD susceptibility. These include the SLE-associated polymorphism of the inhibitory receptor *FCGR2B*, which is found at higher frequency in African and Asian populations where malaria is endemic and enhances phagocytosis of *Plasmodium falciparum*-infected erythrocytes, demonstrating that *FCGR2B* is important in controlling the immune response to malarial parasites.[77] Experimental characterization of a nonsynonymous variant in TLR5, another SLE-associated gene, showed that it leads to altered NF-κB signaling in response to bacterial flagellin.[22] Functional analysis revealed that cells from individuals carrying the *SH2B3* risk allele display an increased proinflammatory cytokine production in response to muramyl dipeptide, a component in all bacterial cell walls, indicating that SH2B3 has a role in protection against bacteria infection.[20] As previously noted, it is noteworthy that variants within the *APOL1* gene known to be under strong selective pressure in some African populations predispose to a manifestation of SLE, end-stage kidney disease.[11] *APOL1* has a role in innate immunity by protecting against *Trypanosoma brucei* infection, the parasitic protozoa transmitted by the tsetse fly that causes human African trypanosomiasis (a.k.a. sleeping sickness). There is compelling evidence that the kidney disease risk alleles' frequencies increased rapidly because of their potential protective effects against human African trypanosomiasis.[84] Given that these disease risk adaptive alleles have been functionally validated, these examples support the hypothesis that the increased prevalence of ADs may result, at least partially, from past events of selection that increased host resistance to infection.[85]

DISCUSSION

This review summarizes the genetic regions associated with susceptibility to different ADs and concomitant evidence for selection, including known agents of selection when known. Uncovering these AD-associated loci under selection underscores the importance of population genetics and how the understanding of human genetic diversity is crucial to understanding disease etiology or treatment response at both the population and individual levels.

Progress achieved in recent years is a direct consequence of large-scale projects such as the HapMap[86] and HGDP.[87] However, the common variation analyzed by these projects does not capture the novel, deleterious or functional variants that, together with common variation, distinguish global human populations.[7] Now, with genome sequence information available on a population scale from projects like the 1000 Genomes,[88] the role of selection in shaping AD risk can be assessed with unprecedented detail in an unbiased fashion.[89,90] Recent studies using whole-genome sequence data helped identify the established functional SNPs in known loci and report novel candidate regions for positive selection.[23,91] In the future, analyses of heritable epigenetic variation will unveil signatures of natural selection on epigenetic variation.

The complexity of gene effects in different ADs is remarkable, as shown, for example, by the complex signatures of selection seen at *NOD2*,[19,81,82] and the plethora of shared variants that increase AD predisposition concordantly or discordantly, such as for *PTPN22*.[28] These observations illustrate that the model whereby AD risk alleles are positively selected due to their protective effects against infections is naive. A combination of population-level phenomena, including possibly bottlenecks, migration, admixture, natural selection and random genetic drift, are likely contributors to this complexity of gene effects. Given the complex history of selective pressures acting on humans, unequal selective pressures and a diverse spectrum of plausible evolutionary models are expected to be exerted on susceptibility loci for ADs.[33] It is likely that several pathogens have exerted pressure on the same loci and that selection can vary in form, intensity, time and space, which is consistent with the observation that both risk and protective alleles for ADs increased in frequency due to selection.[18] For most regions, the exact selective pressure leaving the signature of selection is unclear. Clearly, these signatures are not necessarily the result of adaptation, but might be a consequence of random genetic drift. In any case, regardless of the population phenomenon shaping current human genetic diversity, this genetic variation is the basis of clinically relevant traits at both the individual and population levels.[7]

An important next step to delineate the selective advantage conferred by these AD risk variants are functional studies using *in vitro* experiments and model organisms to identify the underlying functional variants and quantify the phenotypic consequences of the candidate adaptive alleles. Human–pathogen coevolution is ongoing and, despite the emergence of new pathogens (for example, HIV), potential pathogens driving these host-specific adaptations are expected to have long-standing relationships with humans, including those that cause malaria, smallpox, cholera, tuberculosis and leprosy,[92] as well as the human microbiome.[93] Regardless of the agent of selection and the reasons for the emergence of both common and rare AD-causing alleles, incorporating population genetics to understand human genetic diversity will lead to a better understanding of the causes of health disparities, identification of functional variants and discovery of cellular mechanisms and contribute to the development of new therapies.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

1   NIH Progress in Autoimmune Diseases Research. in National Institute of Health Publication No. 05-514 (2005).
2   Carvalheiras, G., Faria, R., Braga, J. & Vasconcelos, C. Fetal outcome in autoimmune diseases. *Autoimmun. Rev.* **11**, A520–A530 (2012).
3   Ostensen, M., Andreoli, L., Brucato, A., Cetin, I., Chambers, C., Clowse, M. E. *et al.* State of the art: reproduction and pregnancy in rheumatic diseases. *Autoimmun. Rev.* **14**, 376–386 (2015).
4   Clowse, M. E., Chakravarty, E., Costenbader, K. H., Chambers, C. & Michaud, K. Effects of infertility, pregnancy loss, and patient concerns on family size of women with rheumatoid arthritis and systemic lupus erythematosus. *Arthritis Care Res.* **64**, 668–674 (2012).
5   McGrogan, A., Snowball, J. & de Vries, C. S. Pregnancy losses in women with type 1 or type 2 diabetes in the UK: an investigation using primary care records. *Diabet. Med.* **31**, 357–365 (2014).
6   Cooper, G. S., Bynum, M. L. & Somers, E. C. Recent insights in the epidemiology of autoimmune diseases: improved prevalence estimates and understanding of clustering of diseases. *J. Autoimmun.* **33**, 197–207 (2009).

7 Torkamani, A., Pham, P., Libiger, O., Bansal, V., Zhang, G., Scott-Van Zeeland, A. A. et al. Clinical implications of human population differences in genome-wide rates of functional genotypes. *Front Genet.* **3**, 211 (2012).

8 Okin, D. & Medzhitov, R. Evolution of inflammatory diseases. *Curr. Biol.* **22**, R733–R740 (2012).

9 Guernier, V., Hochberg, M. E. & Guegan, J. F. Ecology drives the worldwide distribution of human diseases. *PLoS Biol.* **2**, e141 (2004).

10 Fumagalli, M., Sironi, M., Pozzoli, U., Ferrer-Admetlla, A., Pattini, L. & Nielsen, R. Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. *PLoS Genet.* **7**, e1002355 (2011).

11 Freedman, B. I., Langefeld, C. D., Andringa, K. K., Croker, J. A., Williams, A. H., Garner, N. E. et al. End-stage kidney disease in African Americans with lupus nephritis associates with APOL1. *Arthritis Rheum.* **66**, 390–396 (2014)

12 Strachan, D. P. Hay fever, hygiene, and household size. *B.M.J.* **299**, 1259–1260 (1989).

13 Quintana-Murci, L. & Clark, A. G. Population genetic tools for dissecting innate immunity in humans. *Nat. Rev. Immunol.* **13**, 280–293 (2013).

14 Pickrell, J. K., Coop, G., Novembre, J., Kudaravalli, S., Li, J. Z., Absher, D. et al. Signals of recent positive selection in a worldwide sample of human populations. *Genome. Res.* **19**, 826–837 (2009).

15 Sabeti, P. C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C. et al. Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**, 913–918 (2007).

16 Voight, B. F., Kudaravalli, S., Wen, X. & Pritchard, J. K. A map of recent positive selection in the human genome. *PLoS Biol.* **4**, e72 (2006).

17 Barreiro, L. B., Laval, G., Quach, H., Patin, E. & Quintana-Murci, L. Natural selection has driven population differentiation in modern humans. *Nat. Genet.* **40**, 340–345 (2008).

18 Barreiro, L. B. & Quintana-Murci, L. From evolutionary genetics to human immunology: how selection shapes host defence genes. *Nat. Rev. Genet.* **11**, 17–30 (2010).

19 Jostins, L., Ripke, S., Weersma, R. K., Duerr, R. H., McGovern, D. P., Hui, K. Y. et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491**, 119–124 (2012).

20 Zhernakova, A., Elbers, C. C., Ferwerda, B., Romanos, J., Trynka, G., Dubois, P. C. et al. Evolutionary and functional analysis of celiac risk loci reveals SH2B3 as a protective factor against bacterial infection. *Am. J. Hum. Genet.* **86**, 970–977 (2010).

21 Ramos, P. S., Shaftman, S. R., Ward, R. C. & Langefeld, C. D. Genes associated with SLE are targets of recent positive selection. *Autoimmune Dis.* **Vol. 2014**, Article ID 203435 (2014).

22 Grossman, S. R., Andersen, K. G., Shlyakhter, I., Tabrizi, S., Winnicki, S., Yen, A. et al. Identifying recent adaptations in large-scale genomic data. *Cell* **152**, 703–713 (2013).

23 Raj, T., Kuchroo, M., Replogle, J. M., Raychaudhuri, S., Stranger, B. E. & De Jager, P. L. Common risk alleles for inflammatory diseases are targets of recent positive selection. *Am. J. Hum. Genet.* **92**, 517–529 (2013).

24 Hancock, A. M., Witonsky, D. B., Alkorta-Aranburu, G., Beall, C. M., Gebremedhin, A., Sukernik, R. et al. Adaptations to climate-mediated selective pressures in humans. *PLoS Genet.* **7**, e1001375 (2011).

25 Ramos, P. S., Criswell, L. A., Moser, K. L., Comeau, M. E., Williams, A. H., Pajewski, N. M. et al. A comprehensive analysis of shared loci between systemic lupus erythematosus (SLE) and sixteen autoimmune diseases reveals limited genetic overlap. *PLoS Genet.* **7**, e1002406 (2011).

26 Selmi, C., Lu, Q. & Humble, M. C. Heritability versus the role of the environment in autoimmunity. *J. Autoimmun.* **39**, 249–252 (2012).

27 Hu, X. & Daly, M. What have we learned from six years of GWAS in autoimmune diseases, and what is next? *Curr. Opin. Immunol.* **24**, 571–575 (2012).

28 Parkes, M., Cortes, A., van Heel, D. A. & Brown, M. A. Genetic insights into common pathways and complex relationships among immune-mediated diseases. *Nat. Rev. Genet.* **14**, 661–673 (2013).

29 Duret, L. & Arndt, P. F. The impact of recombination on nucleotide substitutions in the human genome. *PLoS Genet.* **4**, e1000071 (2008).

30 Duret, L. & Galtier, N. Biased gene conversion and the evolution of mammalian genomic landscapes. *Annu. Rev. Genomics Hum. Genet.* **10**, 285–311 (2009).

31 Lachance, J. & Tishkoff, S. A. Biased gene conversion skews allele frequencies in human populations, increasing the disease burden of recessive alleles. *Am. J. Hum. Genet.* **95**, 408–420 (2014).

32 Barreiro, L. B., Ben-Ali, M., Quach, H., Laval, G., Patin, E., Pickrell, J. K. et al. Evolutionary dynamics of human Toll-like receptors and their different contributions to host defense. *PLoS Genet.* **5**, e1000562 (2009).

33 Di Rienzo, A. Population genetics models of common diseases. *Curr. Opin. Genet. Dev.* **16**, 630–636 (2006).

34 Tishkoff, S. A., Reed, F. A., Ranciaro, A., Voight, B. F., Babbitt, C. C., Silverman, J. S. et al. Convergent adaptation of human lactase persistence in Africa and Europe. *Nat. Genet.* **39**, 31–40 (2007).

35 Bustamante, C. D., Fledel-Alon, A., Williamson, S., Nielsen, R., Hubisz, M. T., Glanowski, S. et al. Natural selection on protein-coding genes in the human genome. *Nature* **437**, 1153–1157 (2005).

36 Blekhman, R., Man, O., Herrmann, L., Boyko, A. R., Indap, A., Kosiol, C. et al. Natural selection on genes that underlie human disease susceptibility. *Curr. Biol.* **18**, 883–889 (2008).

37 Torgerson, D. G., Boyko, A. R., Hernandez, R. D., Indap, A., Hu, X., White, T. J. et al. Evolutionary processes acting on candidate cis-regulatory regions in humans inferred from patterns of polymorphism and divergence. *PLoS Genet.* **5**, e1000592 (2009).

38 Andres, A. M., Hubisz, M. J., Indap, A., Torgerson, D. G., Degenhardt, J. D., Boyko, A. R. et al. Targets of balancing selection in the human genome. *Mol. Biol. Evol.* **26**, 2755–2764 (2009).

39 Gineau, L., Luisi, P., Castelli, E. C., Milet, J., Courtin, D., Cagnin, N. et al. Balancing immunity and tolerance: genetic footprint of natural selection in the transcriptional regulatory region of HLA-G. *Genes Immun.* **16**, 57–70 (2015).

40 Prugnolle, F., Manica, A., Charpentier, M., Guegan, J. F., Guernier, V. & Balloux, F. Pathogen-driven selection and worldwide HLA class I diversity. *Curr. Biol.* **15**, 1022–1027 (2005).

41 McClelland, E. E., Penn, D. J. & Potts, W. K. Major histocompatibility complex heterozygote superiority during coinfection. *Infect Immun.* **71**, 2079–2086 (2003).

42 Kawashima, M., Ohashi, J., Nishida, N. & Tokunaga, K. Evolutionary analysis of classical HLA class I and II genes suggests that recent positive selection acted on DPB1*04:01 in Japanese population. *PLoS One* **7**, e46806 (2012).

43 Scheinfeldt, L. B. & Tishkoff, S. A. Recent human adaptation: genomic approaches, interpretation and insights. *Nat. Rev. Genet.* **14**, 692–702 (2013).

44 Daub, J. T., Hofer, T., Cutivet, E., Dupanloup, I., Quintana-Murci, L., Robinson-Rechavi, M. et al. Evidence for polygenic adaptation to pathogens in the human genome. *Mol. Biol. Evol.* **30**, 1544–1558 (2013).

45 Hernandez, R. D., Kelley, J. L., Elyashiv, E., Melton, S. C., Auton, A., McVean, G. et al. Classic selective sweeps were rare in recent human evolution. *Science* **331**, 920–924 (2011).

46 Bhatia, G., Patterson, N., Pasaniuc, B., Zaitlen, N., Genovese, G., Pollack, S. et al. Genome-wide comparison of African-ancestry populations from CARe and other cohorts reveals signals of natural selection. *Am. J. Hum. Genet.* **89**, 368–381 (2011).

47 Klironomos, F. D., Berg, J. & Collins, S. How epigenetic mutations can affect genetic evolution: model and mechanism. *Bioessays* **35**, 571–578 (2013).

48 Wang, J. & Fan, C. A neutrality test for detecting selection on DNA methylation using single methylation polymorphism frequency spectrum. *Genome. Biol. Evol.* **7**, 154–171 (2015).

49 Eberle, M. A., Rieder, M. J., Kruglyak, L. & Nickerson, D. A. Allele frequency matching between SNPs reveals an excess of linkage disequilibrium in genic regions of the human genome. *PLoS Genet.* **2**, e142 (2006).

50 Sabeti, P. C., Reich, D. E., Higgins, J. M., Levine, H. Z., Richter, D. J., Schaffner, S. F. et al. Detecting recent positive selection in the human genome from haplotype structure. *Nature* **419**, 832–837 (2002).

51 Smith, J. M. & Haigh, J. The hitch-hiking effect of a favourable gene. *Genet. Res.* **23**, 23–35 (1974).

52 Williamson, S. H., Hubisz, M. J., Clark, A. G., Payseur, B. A., Bustamante, C. D. & Nielsen, R. Localizing recent adaptive evolution in the human genome. *PLoS Genet.* **3**, e90 (2007).

53 Gulko, B., Hubisz, M. J., Gronau, I. & Siepel, A. A method for calculating probabilities of fitness consequences for point mutations across the human genome. *Nat. Genet.* **47**, 276–283 (2015).

54 Akey, J. M. Constructing genomic maps of positive selection in humans: where do we go from here? *Genome Res.* **19**, 711–722 (2009).

55 Fumagalli, M., Cagliani, R., Pozzoli, U., Riva, S., Comi, G. P., Menozzi, G. et al. Widespread balancing selection and pathogen-driven selection at blood group antigen genes. *Genome Res.* **19**, 199–212 (2009).

56 Katzmarzyk, P. T. & Leonard, W. R. Climatic influences on human body size and proportions: ecological adaptations and secular trends. *Am. J. Phys. Anthropol.* **106**, 483–503 (1998).

57 Roberts, D. F. Body weight, race and climate. *Am. J. Phys. Anthropol.* **11**, 533–558 (1953).

58 Leonard, W. R., Sorensen, M. V., Galloway, V. A., Spencer, G. J., Mosher, M. J., Osipova, L. et al. Climatic influences on basal metabolic rates among circumpolar populations. *Am. J. Hum. Biol.* **14**, 609–620 (2002).

59 Jablonski, N. G. & Chaplin, G. The evolution of human skin coloration. *J. Hum. Evol.* **39**, 57–106 (2000).

60 Loomis, W. F. Skin-pigment regulation of vitamin-D biosynthesis in man. *Science* **157**, 501–506 (1967).

61 Young, J. H., Chang, Y. P., Kim, J. D., Chretien, J. P., Klag, M. J., Levine, M. A. et al. Differential susceptibility to hypertension is due to selection during the out-of-Africa expansion. *PLoS Genet.* **1**, e82 (2005).

62 Thompson, E. E., Kuttab-Boulos, H., Witonsky, D., Yang, L., Roe, B. A. & Di Rienzo, A. CYP3A variation and the evolution of salt-sensitivity variants. *Am. J. Hum. Genet.* **75**, 1059–1069 (2004).

63 Forni, D., Pozzoli, U., Cagliani, R., Tresoldi, C., Menozzi, G., Riva, S. et al. Genetic adaptation of the human circadian clock to day-length latitudinal variations and relevance for affective disorders. *Genome Biol.* **15**, 499 (2014).

64 Corona, E., Chen, R., Sikora, M., Morgan, A. A., Patel, C. J., Ramesh, A. et al. Analysis of the genetic basis of disease in the context of worldwide human relationships and migration. *PLoS Genet.* **9**, e1003447 (2013).

65 Currat, M., Trabuchet, G., Rees, D., Perrin, P., Harding, R. M., Clegg, J. B. et al. Molecular analysis of the beta-globin gene cluster in the Niokholo Mandenka population reveals a recent origin of the beta(S) Senegal mutation. *Am. J. Hum. Genet.* **70**, 207–223 (2002).

66 Bersaglieri, T., Sabeti, P. C., Patterson, N., Vanderploeg, T., Schaffner, S. F., Drake, J. A. et al. Genetic signatures of strong recent positive selection at the lactase gene. *Am. J. Hum. Genet.* **74**, 1111–1120 (2004).

67 Enattah, N. S., Sahi, T., Savilahti, E., Terwilliger, J. D., Peltonen, L. & Jarvela, I. Identification of a variant associated with adult-type hypolactasia. *Nat. Genet.* **30**, 233–237 (2002).

68 Lamason, R. L., Mohideen, M. A., Mest, J. R., Wong, A. C., Norton, H. L., Aros, M. C. *et al.* SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. *Science* **310**, 1782–1786 (2005).

69 Yi, X., Liang, Y., Huerta-Sanchez, E., Jin, X., Cuo, Z. X., Pool, J. E. *et al.* Sequencing of 50 human exomes reveals adaptation to high altitude. *Science* **329**, 75–78 (2010).

70 Hancock, A. M., Clark, V. J., Qian, Y. & Di Rienzo, A. Population genetic analysis of the uncoupling proteins supports a role for UCP3 in human cold resistance. *Mol. Biol. Evol.* **28**, 601–614 (2011).

71 Ye, K., Lu, J., Raj, S. M. & Gu, Z. Human expression QTLs are enriched in signals of environmental adaptation. *Genome Biol. Evol.* **5**, 1689–1701 (2013).

72 Karlsson, E. K., Kwiatkowski, D. P. & Sabeti, P. C. Natural selection and infectious disease in human populations. *Nat. Rev. Genet.* **15**, 379–393 (2014).

73 Cagliani, R. & Sironi, M. Pathogen-driven selection in the human genome. *Int. J. Evol. Biol.* **2013**, 204240 (2013).

74 Hughes, A. L. & Nei, M. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* **335**, 167–170 (1988).

75 Qutob, N., Balloux, F., Raj, T., Liu, H., Marion de Proce, S., Trowsdale, J. *et al.* Signatures of historical demography and pathogen richness on MHC class I genes. *Immunogenetics* **64**, 165–175 (2012).

76 Nejentsev, S., Walker, N., Riches, D., Egholm, M. & Todd, J. A. Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. *Science* **324**, 387–389 (2009).

77 Clatworthy, M. R., Willcocks, L., Urban, B., Langhorne, J., Williams, T. N., Peshu, N. *et al.* Systemic lupus erythematosus-associated defects in the inhibitory receptor FcgammaRIIb reduce susceptibility to malaria. *Proc. Natl Acad. Sci. USA* **104**, 7169–7174 (2007).

78 Machado, L. R., Hardwick, R. J., Bowdrey, J., Bogle, H., Knowles, T. J., Sironi, M. *et al.* Evolutionary history of copy-number-variable locus for the low-affinity Fcgamma receptor: mutation rate, autoimmune disease, and the legacy of helminth infection. *Am. Hum. Genet.* **90**, 973–985 (2012).

79 Hawn, T. R., Wu, H., Grossman, J. M., Hahn, B. H., Tsao, B. P. & Aderem, A. A stop codon polymorphism of Toll-like receptor 5 is associated with resistance to systemic lupus erythematosus. *Proc. Natl Acad. Sci. USA* **102**, 10593–10597 (2005).

80 Barreiro, L. B., Tailleux, L., Pai, A. A., Gicquel, B., Marioni, J. C. & Gilad, Y. Deciphering the genetic architecture of variation in the immune response to *Mycobacterium tuberculosis* infection. *Proc. Natl Acad. Sci. USA* **109**, 1204–1209 (2012).

81 Vasseur, E., Boniotto, M., Patin, E., Laval, G., Quach, H., Manry, J. *et al.* The evolutionary landscape of cytosolic microbial sensors in humans. *Am. J. Hum. Genet.* **91**, 27–37 (2012).

82 Nakagome, S., Mano, S., Kozlowski, L., Bujnicki, J. M., Shibata, H., Fukumaki, Y. *et al.* Crohn's disease risk alleles on the NOD2 locus have been maintained by natural selection on standing variation. *Mol. Biol. Evol.* **29**, 1569–1585 (2012).

83 Karlsson, E. K., Harris, J. B., Tabrizi, S., Rahman, A., Shlyakhter, I., Patterson, N. *et al.* Natural selection in a Bangladeshi population from the cholera-endemic ganges river delta. *Sci. Transl. Med.* **5**, 192ra86 (2013).

84 Genovese, G., Friedman, D. J., Ross, M. D., Lecordier, L., Uzureau, P., Freedman, B. I. *et al.* Association of trypanolytic ApoL1 variants with kidney disease in African Americans. *Science* **329**, 841–845 (2010).

85 Sironi, M. & Clerici, M. The hygiene hypothesis: an evolutionary perspective. *Microbes Infect* **12**, 421–427 (2010).

86 International HapMap 3 Consortium, Altshuler, D. M., Gibbs, R. A., Peltonen, L., Altshuler, D. M., Gibbs, R. A. *et al.* Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52–58 (2010).

87 Li, J. Z., Absher, D. M., Tang, H., Southwick, A. M., Casto, A. M., Ramachandran, S. *et al.* Worldwide human relationships inferred from genome-wide patterns of variation. *Science* **319**, 1100–1104 (2008).

88 1000 Genomes Project Consortium, Abecasis, G. R., Altshuler, D., Auton, A., Brooks, L. D., Durbin, R. M. *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).

89 Clark, A. G., Hubisz, M. J., Bustamante, C. D., Williamson, S. H. & Nielsen, R. Ascertainment bias in studies of human genome-wide polymorphism. *Genome Res.* **15**, 1496–1502 (2005).

90 Hider, J. L., Gittelman, R. M., Shah, T., Edwards, M., Rosenbloom, A., Akey, J. M. *et al.* Exploring signatures of positive selection in pigmentation candidate genes in populations of East Asian ancestry. *BMC Evol. Biol.* **13**, 150 (2013).

91 Colonna, V., Ayub, Q., Chen, Y., Pagani, L., Luisi, P., Pybus, M. *et al.* Human genomic regions with exceptionally high levels of population differentiation identified from 911 whole-genome sequences. *Genome Biol.* **15**, R88 (2014).

92 Anderson, R. M. & May, R. M. Coevolution of hosts and parasites. *Parasitology* **85**, 411–426 (1982).

93 Honda, K. & Littman, D. R. The microbiome in infectious disease and inflammation. *Annu. Rev. Immunol.* **30**, 759–795 (2012).

94 Hindorff, L. A., MacArthur, J., Morales, J., Junkins, H., Hall, P. N., Klemm, A. K. *et al.* A Catalog of Published Genome-Wide Association Studies. Available at www.genome.gov/gwastudies. Accessed 27 March 2015.

95 Cagliani, R., Pozzoli, U., Forni, D., Cassinotti, A., Fumagalli, M., Giani, M. *et al.* Crohn's disease loci are common targets of protozoa-driven selection. *Mol. Biol. Evol.* **30**, 1077–1087 (2013).

96 Cockburn, I. A., Mackinnon, M. J., O'Donnell, A., Allen, S. J., Moulds, J. M., Baisor, M. *et al.* A human complement receptor 1 polymorphism that reduces *Plasmodium falciparum* rosetting confers protection against severe malaria. *Proc. Natl Acad. Sci. USA* **101**, 272–277 (2004).

97 Li, Y., Liao, W., Cargill, M., Chang, M., Matsunami, N., Feng, B. J. *et al.* Carriers of rare missense variants in IFIH1 are protected from psoriasis. *J. Invest. Dermatol.* **130**, 2768–2772 (2010).

98 Fumagalli, M., Cagliani, R., Riva, S., Pozzoli, U., Biasin, M., Piacentini, L. *et al.* Population genetics of IFIH1: ancient population structure, local selection, and implications for susceptibility to type 1 diabetes. *Mol. Biol. Evol.* **27**, 2555–2566 (2010).

99 Cagliani, R., Riva, S., Pozzoli, U., Fumagalli, M., Comi, G. P., Bresolin, N. *et al.* Balancing selection is common in the extended MHC region but most alleles with opposite risk profile for autoimmune diseases are neutrally evolving. *BMC Evol. Biol.* **11**, 171 (2011).

100 Liu, X., Fu, Y., Liu, Z., Lin, B., Xie, Y., Liu, Y. *et al.* An ancient balanced polymorphism in a regulatory region of human major histocompatibility complex is retained in Chinese minorities but lost worldwide. *Am. J. Hum. Genet.* **78**, 393–400 (2006).

101 Tan, Z., Shon, A. M. & Ober, C. Evidence of balancing selection at the HLA-G promoter region. *Hum. Mol. Genet.* **14**, 3619–3628 (2005).

102 Black, F. L. & Hedrick, P. W. Strong balancing selection at HLA loci: evidence from segregation in South Amerindian families. *Proc. Natl Acad. Sci. USA* **94**, 12452–12456 (1997).

103 Calafell, F., Roubinet, F., Ramirez-Soriano, A., Saitou, N., Bertranpetit, J. & Blancher, A. Evolutionary dynamics of the human ABO gene. *Hum. Genet.* **124**, 123–135 (2008).

104 Fry, A. E., Griffiths, M. J., Auburn, S., Diakite, M., Forton, J. T., Green, A. *et al.* Common variation in the ABO glycosyltransferase is associated with susceptibility to severe *Plasmodium falciparum* malaria. *Hum. Mol. Genet.* **17**, 567–576 (2008).

105 Rowe, J. A., Handel, I. G., Thera, M. A., Deans, A. M., Lyke, K. E., Kone, A. *et al.* Blood group O protects against severe *Plasmodium falciparum* malaria through the mechanism of reduced rosetting. *Proc. Natl Acad. Sci. USA* **104**, 17471–17476 (2007).

106 Boren, T., Falk, P., Roth, K. A., Larson, G. & Normark, S. Attachment of *Helicobacter pylori* to human gastric epithelium mediated by blood group antigens. *Science* **262**, 1892–1895 (1993).

107 Lindesmith, L., Moe, C., Marionneau, S., Ruvoen, N., Jiang, X., Lindblad, L. *et al.* Human susceptibility and resistance to Norwalk virus infection. *Nat. Med.* **9**, 548–553 (2003).

108 Carnero-Montoro, E., Bonet, L., Engelken, J., Bielig, T., Martinez-Florensa, M., Lozano, F. *et al.* Evolutionary and functional evidence for positive selection at the human CD5 immune receptor gene. *Mol. Biol. Evol.* **29**, 811–823 (2012).

109 Swanberg, M., Lidman, O., Padyukov, L., Eriksson, P., Akesson, E., Jagodic, M. *et al.* MHC2TA is associated with differential MHC molecule expression and susceptibility to rheumatoid arthritis, multiple sclerosis and myocardial infarction. *Nat. Genet.* **37**, 486–494 (2005).

110 Choudhury, A., Hazelhurst, S., Meintjes, A., Achinike-Oduaran, O., Aron, S., Gamieldien, J. *et al.* Population-specific common SNPs reflect demographic histories and highlight regions of genomic plasticity with functional relevance. *BMC Genomics* **15**, 437 (2014).