

REVIEW

# GWASs and the age of human as the model organism for autoimmune genetic research

Robert Plenge\*

## Abstract

Genetic studies have identified more than 150 autoimmune loci, and next-generation sequencing will identify more. Is it time to make human the model organism for autoimmune research?

Human genetics - linking inherited variation in DNA sequence with traits such as susceptibility to disease - provides *prima facie* evidence that a gene and a pathway are associated with a disease. The most recent wave of genomic technology has allowed human genomes to be scanned for variant DNA sequences (or alleles) in many people to determine which alleles are associated with a particular disease or phenotype of interest. Termed genome-wide association studies, or GWASs, this approach has identified hundreds of alleles that are associated with a variety of human traits [1,2]. By most accounts, the GWAS approach has been very successful at identifying new regions of the genome (or loci) that are important in disease, even though the effect sizes of most alleles are modest.

The GWAS approach has been particularly successful at uncovering risk alleles for autoimmune diseases. Collectively, autoimmune diseases are common, affecting more than 5% of the adult population [3]. These diseases include rheumatoid arthritis (RA), type 1 diabetes (T1D), inflammatory bowel disease (IBD), systemic lupus erythematosus (SLE), multiple sclerosis (MS), psoriasis and celiac disease (among others). RA is a chronic inflammatory disease that destroys free moving joints. T1D is a form of diabetes that results from the destruction of insulin-producing beta cells of the pancreas. IBD is a group of inflammatory conditions of the colon and small intestine; the two major types are Crohn's disease and ulcerative colitis. In SLE, the immune system attacks

a wide variety of organs, including the heart, joints, skin, lungs, blood vessels, liver, kidneys and nervous system. MS is an autoimmune disease in which the fatty myelin sheaths around the axons of the brain and spinal cord are damaged, leading to a broad spectrum of signs and symptoms. Psoriasis is a chronic disease in which the skin develops red, scaly patches, which is the result of areas of inflammation and excessive skin production. Celiac disease is an autoimmune disorder of the small intestine caused by a reaction to storage proteins (called gluteins) found in cereal grains; the ensuing excessive immune reaction leads to an attack on the intestinal villi and tissue damage, resulting in malabsorption of nutrients.

So far, approximately 150 loci have been identified that increase risk of these autoimmune diseases [4-14]. For each disease, the strongest genetic risk factors reside within the major histocompatibility complex (MHC) region on chromosome 6 [15]. Most associated alleles in other regions are common in the general population, but increase the disease risk by only 10 to 20% (corresponding to an odds ratio (OR) of 1.10 to 1.20 per copy of the risk allele). (The OR is a measure of the strength of association; it refers to the ratio of the odds of an event occurring in one group (such as cases) to the odds of it occurring in another group (such as controls).) For any given autoimmune disease, the known genetic risk alleles explain between 10 and 20% of variance in disease risk, whereas more than 50% of disease risk is estimated to be heritable. The remaining 30% or so of unexplained genetic disease risk is termed the missing heritability.

The challenges now are, first, to find the causal mutation responsible for the signal of association; second, to understand which gene is disrupted by the causal mutation and how it is disrupted (that is, whether the mutation results in gain of function, loss of function, or a new function altogether); third, to understand which cell type and biological pathways are altered by these mutations; and finally to find additional mutations that explain the missing heritability [16]. The next wave of genomic technology - next-generation sequencing - will be a powerful ally in this effort. In particular, next-generation sequencing will help localize the causal mutation, as well as help identify rare alleles that confer risk of autoimmune disease.

\*Correspondence: rplenge@partners.org  
Brigham and Women's Hospital, Division of Rheumatology, Immunology and Allergy, Boston, MA 02115, USA

Thus, an important question remains: what is the most appropriate scientific approach to understand function of risk alleles discovered in human genetics research? Is the mouse the most appropriate model organism, or do these genetic discoveries provide new resources to enable functional studies directly in human immune cells?

Here, I discuss the confluence of events that create a unique opportunity to use human subjects as the 'model organism' for the study of autoimmune disease pathogenesis. In addition to GWASs and next-generation sequencing, registries of blood draws from healthy, consenting human volunteers enable functional studies of genetic variants in a wide range of primary human immune cells, and human stem cell technology has advanced to the point at which induced pluripotent stem (iPS) cells can be derived from patients with specific mutations and differentiated into diverse immune lineages. These resources should allow investigators to understand the altered cellular state in diseases that are uniquely human, which should ultimately lead to new therapeutics to treat or prevent the devastating consequences of autoimmune disease.

### Common SNPs and risk of autoimmune disease

In general, 'common' variants are those present at a frequency of over 1% in any one continental population (such as Europeans, Asians and Africans), whereas 'rare' variants are those present at a frequency of less than 1% in these populations [17]. This simple categorical distinction has been made in order to frame the genetic approach to discovering and testing DNA variants for their role in disease. For common variants, it is possible to screen a reference population to identify a catalog of variants (the discovery phase), and then test these variants in case-control collections using high-throughput genotyping technologies (the testing phase). A variety of resources have been developed to catalog common single nucleotide polymorphisms (SNPs), including the International HapMap Project [17,18]. More recently, data from the 1000 Genomes Project [19] have begun to be used to catalog variants in the 1% frequency range.

In order to test whether these common SNPs are associated with risk of disease, commercial 'SNP chips' or arrays have been developed that capture most, although not all, common variation in the genome. These genotyping arrays can genotype hundreds of thousands of SNPs in a single experiment, at a cost of several hundred US dollars per sample. Contemporary GWASs use these arrays to measure the frequency of alleles in cases compared with controls. If the difference in allele frequency reaches a stringent level of statistical significance that corrects for the fact that there are about 1,000,000 independent common SNPs in the human genome (this significance level is about  $P < 5 \times 10^{-8}$ ), then the allele is said to be 'associated' with disease.

There are approximately 10 million common SNPs in the human genome. A fundamental challenge in human genetics is to systematically test each of these 10 million common SNPs for its role in disease. Contemporary GWASs test several hundred thousand SNPs across the entire human genome, most of which are common (minor allele frequency over 5%) in the general, healthy population. To test the remaining over 9 million common SNPs, the GWAS approach relies on the correlation structure of nearby SNPs. That is, nine out of ten SNPs are highly correlated, and testing one SNP serves to tag the remaining nine nearby SNPs. This concept is known as linkage disequilibrium (LD).

The underlying rationale for the GWAS approach is rooted firmly in population genetics, as most of the differences between any two chromosomes are due to common SNPs [20]. On the basis of the hypothesis that disease alleles reflect the allelic spectrum of diseases in the general population, the risk of common diseases will be attributable in part to allelic variants that are also common.

GWASs have discovered about 150 loci that harbor SNPs associated with risk of autoimmune diseases. Several of the earliest GWASs that successfully identified common risk alleles were done in autoimmune diseases. Crohn's disease is an illustrative example. Before GWASs, only two loci outside the MHC were known to be associated with Crohn's disease risk [21]. In 2006, a GWAS of about 1,000 case-control samples identified a coding variant in the interleukin 23 receptor (*IL23R*) gene locus [22]. The landmark Wellcome Trust Case Control Consortium GWAS, published in 2007, included three autoimmune diseases (of the seven diseases studied): Crohn's disease, T1D and RA [23]. Since these initial GWASs, over 30 Crohn's disease risk loci [7], over 40 T1D risk loci [6] and over 25 RA risk loci have been discovered [24].

From these GWASs, an important theme has emerged: the overlap among the loci that confer risk of autoimmune disease. In 2008, Smyth and colleagues [9] studied the overlap between celiac disease and T1D. The study [9] found that nearly half of the about 30 risk loci contributed to both diseases, whereas the others seemed to be disease-specific. Other studies have compared and contrasted risk confirmed alleles for a variety of autoimmune diseases [9,25-27]. There is clear overlap for many of the known risk alleles, consistent with epidemiological data of disease clustering within families [28]. A partial list of loci associated with multiple autoimmune diseases is shown in Table 1.

### Missing heritability: next-generation sequencing and the role of rare SNPs

Although the number of loci associated with autoimmune disease is impressive, these loci cannot explain a sizeable

**Table 1. Loci associated with multiple autoimmune diseases**

Chromosome	Position	Gene(s)	Disease*	References
1	67466790	IL23R	Psoriasis, CD, UC	[7,12,53]
1	114179091	PTPN22	SLE, CD, RA, T1D	[7,8,54,55]
1	116905738	CD58	MS, RA	[5,11]
1	205006527	IL10	SLE, T1D, UC	[6,8,14]
2	102437000	IL18RAP	T1D, celiac	[9]
2	191672878	STAT4	RA, SLE	[56]
2	204402121	CTLA4	RA, T1D, celiac	[6,13,57]
4	25694609	RBPJ	T1D, RA	[6]
4	123351942	IL2	T1D, Celiac, RA, UC	[6,25,58]
5	150437678	TNIP1	SLE, psoriasis	[8]
5	158650367	IL12B	Psoriasis, CD	[7,12]
6	106541962	PRDM1, ATG5	CD, RA, SLE	[5,7,8]
6	138014761	TNFAIP3	Celiac, RA, SLE, psoriasis	[12,13,59-61]
6	159385965	TAGAP	Celiac, RA	[5,10]
6	167357978	CCR6	CD, RA	[7]
7	128376236	IRF5	SLE, RA	[31,62]
8	11377591	BLK	SLE, RA	[35,63]
10	6138955	IL2RA	RA, MS, T1D	[6,64]
10	6430456	PRKCCQ	T1D, RA	[4,65]
16	11074189	CLEC16A	MS, T1D	[11,66]
18	12769947	PTPN2	CD, celiac, T1D	[7,13,66]

\*CD, Crohn's disease; MS, multiple sclerosis; RA, rheumatoid arthritis; SLE, systemic lupus erythematosus; T1D, type 1 diabetes; UC, ulcerative colitis.

fraction of disease risk. In fact, outside the MHC, common alleles can only explain 5 to 10% of disease risk associated with autoimmune disease. Considering that family studies have shown that more than 50% of autoimmune disease risk is thought to be genetic, the question arises as to why so much of the heritability is apparently unexplained by initial GWAS findings. One of the most frequently cited explanations for 'missing heritability' is that rare SNPs contribute substantially to disease risk, and contemporary GWAS arrays do not adequately capture rare variants [16].

There are two ways to test rare variants systematically for association with disease. First, it is possible to catalog low-frequency variants - those variants present in approximately 0.5 to 5% of control chromosomes - in a manner analogous to common variants. The only difference is that a greater number of subjects need to be included in the discovery effort. This is the main premise behind the 1000 Genomes Project [19]. Once discovered and catalogued, these low-frequency variants could be genotyped in a high-throughput manner using genotyping arrays.

The second approach is to couple the discovery and testing phases into a single experiment. That is, direct sequencing is done in case-control collections themselves, generating an unbiased catalog of DNA variants that are then tested for association with disease.

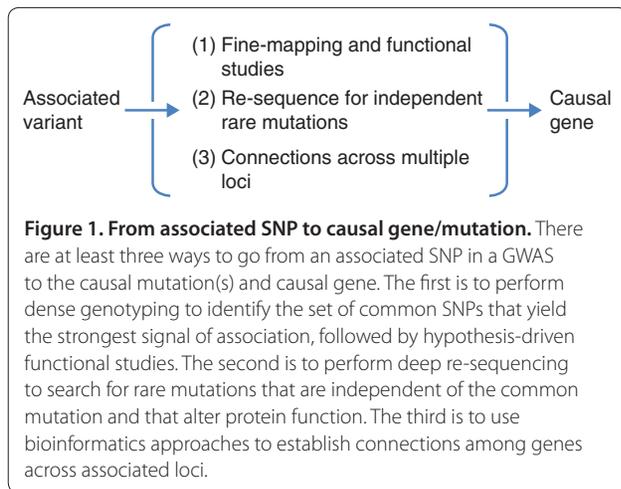
Until recently, direct sequencing in large patient samples was cost prohibitive. Next-generation sequencing has been developed to sequence large regions of DNA - with the ultimate goal of sequencing the complete genome - in a high-throughput and cost-effective manner. In the near future, next-generation sequencing will probably be the technical method of choice for conducting GWASs.

### From associated SNP to causal allele and causal gene

An important promise of human genetics is that GWASs offer an unbiased approach to discovering new pathways that cause disease. Towards this end, a major challenge is to take the expanding list of disease risk alleles and understand the effect on gene function. The first step is to identify which gene near the associated SNP has its function affected by the underlying causal mutation (which is rarely known). This step is critical, as the region of LD surrounding the associated SNP often contains more than one gene (although often there is one likely candidate gene from the known biology). A region of LD includes neighboring sequence in which a group of SNPs are highly correlated (for example, at a correlation coefficient of  $r^2 > 0.80$ ). Moreover, it is conceivable that the causal mutation exerts its effect at a distance (for example, by altering gene expression) or that the causal mutation is rare in the general population and located some distance from the associated SNP [29].

As shown in Figure 1, there are at least three general approaches to get from associated SNP to causal gene (and causal mutation). First, fine-mapping of the region of LD is performed using resequencing and dense genotyping. An allele is considered causal if it is predicted to alter function and if direct experimentation demonstrates altered function. An intriguing result from GWASs is that most associated SNPs lie outside coding regions, and most of the causal mutations probably also fall outside coding regions. It is likely that many causal mutations affect gene expression or mRNA splicing.

One of the best examples was fine-mapping and functional studies of *IRF5*, a gene associated with SLE and other autoimmune diseases [30,31]. *IRF5* encodes a member of the interferon regulatory factor (IRF) family, a group of transcription factors with diverse roles, including virus-mediated activation of interferon and modulation of cell growth, differentiation, apoptosis and immune system activity. Studies have revealed three functional alleles of *IRF5*: an exon 1 splice site variant, a



30-bp in-frame insertion/deletion variant of exon 6, and a variant in a conserved poly(A)<sup>+</sup> signal sequence that alters the length of the 3' untranslated region and stability of *IRF5* mRNAs [30]. Haplotypes of these three variants define at least three distinct levels of risk to SLE. There is an approximately twofold increase in the level of risk between carriers of the highest and lowest risk haplotypes.

Second, candidate genes from a region of LD can be resequenced to search for independent, rare protein-coding mutations. The underlying hypothesis is that a true causal gene will harbor multiple risk alleles; at least one of these might be common (and identified by GWAS), whereas many others will be rare. Precedence for this hypothesis comes from studies of Mendelian disorders, for which disease can be caused by many different mutations to the same gene (genetic heterogeneity). In a study published in 2009 [32], the coding exons of six genes identified by GWASs of T1D were resequenced to search for independent rare mutations. Two rare SNPs in the interferon-induced helicase C domain-containing protein 1 (*IFIH1*) gene were identified that conferred protection from T1D. *IFIH1* is a cytoplasmic protein that recognizes RNA of certain viruses and mediates immune activation. Following infection, the *IFIH1* protein senses the presence of viral RNA in the cytoplasm, triggers activation of nuclear factor (NF)- $\kappa$ B and IRF pathways and induces antiviral IFN- $\beta$  response. The non-synonymous SNP with the strongest association, rs35667974 (which causes the amino acid substitution Ile923Val), was observed on an estimated 3 out of 960 case chromosomes but 24 out of 960 control chromosomes ( $P = 0.00004$ ); another SNP, rs35337543 (which affects a splice donor site), was observed on 7 case chromosomes and 23 control chromosomes ( $P = 0.005$ ). Both SNPs were genotyped in more than 20,000 additional case-control samples: rs35667974 was present in about 1% of cases and 2% of controls ( $P = 2.1 \times 10^{-16}$ ) and

rs35337543 in 1% of cases versus 1.5% of controls ( $P = 1.4 \times 10^{-4}$ ). Both mutations are predicted to be loss-of-function mutations, although why these mutations influence risk of T1D remains unknown.

The third approach is less direct, but nonetheless very powerful, especially when there are many loci associated with risk of disease. The underlying hypothesis is that there are a limited number of biological pathways that are altered to confer risk of disease and that true causal genes will be restricted to those specific pathways. Examples of such pathways include known signaling pathways (such as the NF- $\kappa$ B pathway and risk of RA [33]) or catabolic pathways (such as autophagy and risk of Crohn's disease [20]). The challenge of this computational approach is to define categories of pathways, as our understanding of many biological processes is incomplete. One successful approach has been to use information contained in PubMed abstracts to establish connections between gene loci [34]. This approach has been used to identify putative causal genes for RA and celiac disease [5,13]. In the RA study, three loci were identified that contained the genes *CD28*, *CD2/CD58* and *PRDM1*, respectively [5]. Both CD2 and CD28 are co-stimulatory molecules on the surface of T cells. *PRDM1* (also known as *BLIMP-1*) is a transcription factor that regulates terminal differentiation of B cells into immunoglobulin-secreting plasma cells. Once these connections are established among risk loci, direct experimentation is still required to prove the pathways are critical to disease.

### Resources to validate the biological effects of causal mutations

Once the causal gene and causal mutation(s) have been identified, the next major challenge is to understand the underlying biological pathways that lead to autoimmune disease. New resources now make it possible to study the effects of mutations linked to autoimmune disease directly in relevant human tissue.

Registries have now been established at academic medical centers to study the functional consequences of common genetic mutations in blood cells from healthy control subjects [35]. Human immune cells (such as B and T lymphocytes) are easily accessible through a simple blood draw. These immune cells are of direct relevance to pathogenesis of autoimmune diseases, as indicated not only by recent genetic studies but also by previous studies in patients with autoimmune diseases [36]. Human immune cells derived from healthy control subjects have been used successfully to gain insight into function of common mutations at several autoimmune genes. A missense mutation in the *PTPN22* gene is associated with several autoimmune diseases. *PTPN22* encodes a protein tyrosine phosphatase that is expressed in lymphoid

tissues and implicated in T-cell activation [37]. Functional studies in T cells derived from healthy human participants have shown that the *PTPN22* risk allele alters secretion of IL2 from T cells stimulated through the T-cell receptor [38]. Other autoimmune risk alleles have been studied in a similar manner: a common multiple sclerosis risk mutation at *CD58* can explain about 40% of the variance of *CD58* cell surface expression on peripheral blood mononuclear cells (PBMCs) [39]; and a common T1D mutation in *IL2RA* alters *IL2RA* cell surface expression on CD4<sup>+</sup> memory T cells [40].

Another new approach is to generate iPS cells from patients who carry specific genetic mutations. First described in 2006 [41], several studies have shown that iPS cells can be derived from patients with Mendelian disorders [42]. By definition, iPS cells are pluripotent and can be differentiated into any human cell type. Specific protocols are required to direct differentiation into a specific cell lineage. In the case of immune lineages, protocols have been developed to differentiate human embryonic stem (ES) cells into B cells, T cells, natural killer cells, and other immune lineages [43-50]. Because of the similarities between ES and iPS cells, differentiation protocols developed in ES cells should be applicable to differentiation of iPS cells into these same immune lineages.

Whether iPS cells derived from patients with autoimmune disease will be useful for functional studies of human genetic mutations is a hypothesis that needs to be rigorously tested. Human iPS cells offer several theoretical advantages over primary human immune cells derived from healthy patients. First, although many immune lineages can be isolated from peripheral blood, many reside within lymph nodes and other privileged sites not accessible through the blood. Moreover, it is impractical to isolate more than a few immune lineages in the amount of blood drawn from a single individual at a single point in time. Second, in studies of primary human immune cells, it is important to investigate carriers and non-carriers of mutations on the same day to minimize technical variability. iPS cells have the theoretical advantage of repeated measurements under a set of controlled conditions. Finally, primary human cells have a limited lifespan in culture. As a consequence, it is difficult to manipulate primary cells with transfections and other cellular perturbations.

Most genetic discoveries have concerned the risk of disease overall, rather than relevant subsets of disease; this applies not just to autoimmunity but also to other diseases. As a consequence, a new challenge is to correlate genotype with clinically relevant phenotypes, such as response to therapy and disease severity. For genotype-phenotype correlation studies, the major bottleneck is setting up large registries of patients with biospecimens

for genomic studies and detailed clinical data. Traditional patient registries and clinical trials - the workhorse for sample collection over the past decades - are unlikely to achieve the size required to obtain thousands of autoimmune patient samples for these studies. New approaches - next-generation registries - will be required to break this bottleneck. In theory, it should be possible to collect data as part of routine patient care. Increased use of electronic medical records [51] and new approaches to mining clinical data from such records [52] is one exciting approach to expanding sample collections.

Contemporary GWASs of common variants have identified approximately 150 loci that confer risk of common autoimmune diseases. Once the causal genes and causal mutations have been identified, the next challenges will be to understand the underlying biological pathways and to correlate genotype with clinically relevant phenotypes. New resources are now available to enable these translational immunology studies in humans. Over the next few years, great strides should be made towards accomplishing these ambitious yet attainable goals.

Published: 5 May 2010

#### References

1. Altshuler D, Daly MJ, Lander ES: **Genetic mapping in human disease.** *Science* 2008, **322**:881-888.
2. Hindorf L, Junkins H, Mehta J, Manolio T: **A catalog of published genome-wide association studies.** [http://www.genome.gov/26525384]
3. Jacobson DL, Gange SJ, Rose NR, Graham NM: **Epidemiology and estimated population burden of selected autoimmune diseases in the United States.** *Clin Immunol Immunopathol* 1997, **84**:223-243.
4. Raychaudhuri S, Remmers EF, Lee AT, Hackett R, Guiducci C, Burt NP, Gianniny L, Korman BD, Padyukov L, Kurreeman FA, Chang M, Catanese JJ, Ding B, Wong S, van der Helm-van Mil AH, Neale BM, Coby J, Cui J, Tak PP, Wolbink GJ, Crusius JB, van der Horst-Bruinsma IE, Criswell LA, Amos CI, Seldin MF, Kastner DL, Ardlie KG, Alfredsson L, Costenbader KH, Altshuler D, et al.: **Common variants at CD40 and other loci confer risk of rheumatoid arthritis.** *Nat Genet* 2008, **40**:1216-1223.
5. Raychaudhuri S, Thomson BP, Remmers EF, Eyre S, Hinks A, Guiducci C, Catanese JJ, Xie G, Stahl EA, Chen R, Alfredsson L, Amos CI, Ardlie KG, BIRAC Consortium, Barton A, Bowes J, Burt NP, Chang M, Coby J, Costenbader KH, Criswell LA, Crusius JB, Cui J, De Jager PL, Ding B, Emery P, Flynn E, Harrison P, Hocking LJ, et al.: **Genetic variants at CD28, PRDM1 and CD2/CD58 are associated with rheumatoid arthritis risk.** *Nat Genet* 2009, **41**:1313-1318.
6. Barrett JC, Clayton DG, Concannon P, Akolkar B, Cooper JD, Erlich HA, Julier C, Morahan G, Nerup J, Nierras C, Plagnol V, Pociot F, Schuilenburg H, Smyth DJ, Stevens H, Todd JA, Walker NM, Rich SS, The Type 1 Diabetes Genetics Consortium: **Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes.** *Nat Genet* 2009, **41**:703-707.
7. Barrett JC, Hansoul S, Nicolae DL, Cho JH, Duerr RH, Rioux JD, Brant SR, Silverberg MS, Taylor KD, Barmada MM, Bittton A, Dassopoulos T, Datta LW, Green T, Griffiths AM, Kistner EO, Murtha MT, Regueiro MD, Rotter JJ, Schumm LP, Steinhardt AH, Targan SR, Xavier RJ, NIDDK IBD Genetics Consortium, Libioulle C, Sandor C, Lathrop M, Belaiche J, Dewit O, Gut I, et al.: **Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease.** *Nat Genet* 2009, **41**:995-996.
8. Gateva V, Sandling JK, Hom G, Taylor KE, Chung SA, Sun X, Ortmann W, Kosoy R, Ferreira RC, Nordmark G, Gunnarsson I, Svenungsson E, Padyukov L, Sturfelt G, Jönsen A, Bengtsson AA, Rantapää-Dahlqvist S, Baechler EC, Brown EE, Alarcón GS, Edberg JC, Ramsey-Goldman R, McGwin G Jr, Reveille JD, Vilá LM, Kimberly RP, Manzi S, Petri MA, Lee A, Gregersen PK, et al.: **A large-scale replication study identifies TNIP1, PRDM1, JAZF1, UHRF1BP1 and IL10 as**

- risk loci for systemic lupus erythematosus. *Nat Genet* 2009, **41**:1228-1233.
9. Smyth DJ, Plagnol V, Walker NM, Cooper JD, Downes K, Yang JH, Howson JM, Stevens H, McManus R, Wijmenga C, Heap GA, Dubois PC, Clayton DG, Hunt KA, van Heel DA, Todd JA: **Shared and distinct genetic variants in type 1 diabetes and celiac disease.** *N Engl J Med* 2008, **359**:2767-2777.
  10. Hunt KA, Zhernakova A, Turner G, Heap GA, Franke L, Bruinenberg M, Romanos J, Dinesen LC, Ryan AW, Panesar D, Gwilliam R, Takeuchi F, McLaren WM, Holmes GK, Howdle PD, Walters JR, Sanders DS, Playford RJ, Trynka G, Mulder CJ, Mearin ML, Verbeek WH, Trimble V, Stevens FM, O'Morain C, Kennedy NP, Kelleher D, Pennington DJ, Strachan DP, McArdle WL, et al.: **Newly identified genetic risk variants for celiac disease related to the immune response.** *Nat Genet* 2008, **40**:395-402.
  11. De Jager PL, Jia X, Wang J, de Bakker PI, Ottoboni L, Aggarwal NT, Piccio L, Raychaudhuri S, Tran D, Aubin C, Briskin R, Romano S; International MS Genetics Consortium, Baranzini SE, McCauley JL, Pericak-Vance MA, Haines JL, Gibson RA, Naegelin Y, Uitdehaag B, Matthews PM, Kappos L, Polman C, McArdle WL, Strachan DP, Evans D, Cross AH, Daly MJ, Compston A, Sawcer SJ, et al.: **Meta-analysis of genome scans and replication identify CD6, IRF8 and TNFRSF1A as new multiple sclerosis susceptibility loci.** *Nat Genet* 2009, **41**:776-782.
  12. Nair RP, Duffin KC, Helms C, Ding J, Stuart PE, Goldgar D, Gudjonsson JE, Li Y, Tejasvi T, Feng BJ, Ruether A, Schreiber S, Weichenthal M, Gladman D, Rahman P, Schrodi SJ, Prahalad S, Guthery SL, Fischer J, Liao W, Kwok PY, Menter A, Lathrop GM, Wise CA, Begovich AB, Voorhees JJ, Elder JT, Krueger GG, Bowcock AM, Abecasis GR; Collaborative Association Study of Psoriasis: **Genome-wide scan reveals association of psoriasis with IL-23 and NF-kappaB pathways.** *Nat Genet* 2009, **41**:199-204.
  13. Dubois PC, Trynka G, Franke L, Hunt KA, Romanos J, Curtotti A, Zhernakova A, Heap GA, Adány R, Aromaa A, Bardella MT, van den Berg LH, Bockett NA, de la Concha EG, Dema B, Fehrmann RS, Fernández-Arquero M, Fiala S, Grandone E, Green PM, Groen HJ, Gwilliam R, Houwen RH, Hunt SE, Kaukinen K, Kelleher D, Korponay-Szabo I, Kurppa K, MacMathuna P, Mäki M, et al.: **Multiple common variants for celiac disease influencing immune gene expression.** *Nat Genet* 2009, **41**:295-302.
  14. McGovern DP, Gardet A, Törkvist L, Goyette P, Essers J, Taylor KD, Neale BM, Ong RT, Lagacé C, Li C, Green T, Stevens CR, Beauchamp C, Flesher PR, Carlson M, D'Amato M, Halfvarson J, Hibberd ML, Lördal M, Padyukov L, Andriulli A, Colombo E, Latiano A, Palmieri O, Bernard EJ, Deslandres C, Hommes DW, de Jong DJ, Stokkers PC, Weersma RK, et al.: **Genome-wide association identifies multiple ulcerative colitis susceptibility loci.** *Nat Genet* 2010, **42**:332-337.
  15. Fernando MM, Stevens CR, Walsh EC, De Jager PL, Goyette P, Plenge RM, Vyse TJ, Rioux JD: **Defining the role of the MHC in autoimmunity: a review and pooled analysis.** *PLoS Genet* 2008, **4**:e1000024.
  16. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCauley SA, Visscher PM: **Finding the missing heritability of complex diseases.** *Nature* 2009, **461**:747-753.
  17. International HapMap Consortium: **A haplotype map of the human genome.** *Nature* 2005, **437**:1299-1320.
  18. International HapMap Consortium, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, Yu F, Yang H, Zeng C, Gao Y, Hu H, Hu W, Li C, Lin W, Liu S, Pan H, Tang X, Wang J, Wang W, Yu J, Zhang B, Zhang Q, et al.: **A second generation human haplotype map of over 3.1 million SNPs.** *Nature* 2007, **449**:851-861.
  19. Reich DE, Lander ES: **On the allelic spectrum of human disease.** *Trends Genet* 2001, **17**:502-510.
  20. Abraham C, Cho JH: **Inflammatory bowel disease.** *N Engl J Med* 2009, **361**:2066-2078.
  21. Duerr RH, Taylor KD, Brant SR, Rioux JD, Silverberg MS, Daly MJ, Steinhardt AH, Abraham C, Regueiro M, Griffiths A, Dassopoulos T, Bitton A, Yang H, Targan S, Datta LW, Kistner EO, Schumm LP, Lee AT, Gregersen PK, Barnada MM, Rotter JI, Nicolae DL, Cho JH: **A genome-wide association study identifies IL23R as an inflammatory bowel disease gene.** *Science* 2006, **314**:1461-1463.
  22. Wellcome Trust Case Control Consortium: **Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls.** *Nature* 2007, **447**:661-678.
  23. Raychaudhuri S: **Recent advances in the genetics of rheumatoid arthritis.** *Curr Opin Rheumatol* 2010, **22**:109-118.
  24. Sirota M, Schaub MA, Batzoglu S, Robinson WH, Butte AJ: **Autoimmune disease classification by inverse association with SNP alleles.** *PLoS Genet* 2009, **5**:e1000792.
  25. Zhernakova A, Alizadeh BZ, Bevova M, van Leeuwen MA, Coenen MJ, Franke B, Franke L, Posthumus MD, van Heel DA, van der Steege G, Radstake TR, Barrera P, Roep BO, Koeleman BP, Wijmenga C: **Novel association in chromosome 4q27 region with rheumatoid arthritis and confirmation of type 1 diabetes point to a general risk locus for autoimmune diseases.** *Am J Hum Genet* 2007, **81**:1284-1288.
  26. Coenen MJ, Trynka G, Heskamp S, Franke B, van Diemen CC, Smolonska J, van Leeuwen M, Brouwer E, Boezen MH, Postma DS, Platteeuw M, Zanen P, Lammers JW, Groen HJ, Mali WP, Mulder CJ, Tack GJ, Verbeek WH, Wolters VM, Houwen RH, Mearin ML, van Heel DA, Radstake TR, van Riel PL, Wijmenga C, Barrera P, Zhernakova A: **Common and different genetic background for rheumatoid arthritis and coeliac disease.** *Hum Mol Genet* 2009, **18**:4195-4203.
  27. Hemminki K, Li X, Sundquist K, Sundquist J: **Shared familial aggregation of susceptibility to autoimmune diseases.** *Arthritis Rheum* 2009, **60**:2845-2847.
  28. 1000 Genomes Project [http://www.1000genomes.org]
  29. Dickson SP, Wang K, Krantz I, Hakonarson H, Goldstein DB: **Rare variants create synthetic genome-wide associations.** *PLoS Biol* 2010, **8**:e1000294.
  30. Graham RR, Kyogoku C, Sigurdsson S, Vlasova IA, Davies LR, Baechehr EC, Plenge RM, Koeth T, Ortmann WA, Hom G, Bauer JW, Gillett C, Burtt N, Cunningham Graham DS, Onofrio R, Petri M, Gunnarsson I, Svenungsson E, Rönnblom L, Nordmark G, Gregersen PK, Moser K, Gaffney PM, Criswell LA, Vyse TJ, Syvänen AC, Bohjanen PR, Daly MJ, Behrens TW, Altschuler D: **Three functional variants of IFN regulatory factor 5 (IRF5) define risk and protective haplotypes for human lupus.** *Proc Natl Acad Sci USA* 2007, **104**:6758-6763.
  31. Sigurdsson S, Nordmark G, Göring HH, Lindroos K, Wiman AC, Sturfelt G, Jönsen A, Rantapää-Dahlqvist S, Möller B, Kere J, Koskenmies S, Widén E, Eloranta ML, Julkunen H, Kristjansdóttir H, Steinsson K, Alm G, Rönnblom L, Syvänen AC: **Polymorphisms in the tyrosine kinase 2 and interferon regulatory factor 5 genes are associated with systemic lupus erythematosus.** *Am J Hum Genet* 2005, **76**: 528-537.
  32. Nejentsev S, Walker N, Riches D, Egholm M, Todd JA: **Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes.** *Science* 2009, **324**:387-389.
  33. Plenge RM: **Recent progress in rheumatoid arthritis genetics: one step towards improved patient care.** *Curr Opin Rheumatol* 2009, **21**:262-271.
  34. Raychaudhuri S, Plenge RM, Rossin EJ, Ng AC; International Schizophrenia Consortium, Purcell SM, Sklar P, Scolnick EM, Xavier RJ, Altschuler D, Daly MJ: **Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions.** *PLoS Genet* 2009, **5**:e1000534.
  35. Gregersen PK: **Closing the gap between genotype and phenotype.** *Nat Genet* 2009, **41**:958-959.
  36. Gregersen PK, Behrens TW: **Genetics of autoimmune diseases - disorders of immune homeostasis.** *Nat Rev Genet* 2006, **7**:917-928.
  37. Hasegawa K, Martin F, Huang G, Tumas D, Diehl L, Chan AC: **PEST domain-enriched tyrosine phosphatase (PEP) regulation of effector/memory T cells.** *Science* 2004, **303**:685-689.
  38. Vang T, Congia M, Macis MD, Musumeci L, Orrú V, Zavattari P, Nika K, Tautz L, Taskén K, Cucca F, Mustelin T, Bottini N: **Autoimmune-associated lymphoid tyrosine phosphatase is a gain-of-function variant.** *Nat Genet* 2005, **37**:1317-1319.
  39. De Jager PL, Baecher-Allan C, Maier LM, Arthur AT, Ottoboni L, Barcellos L, McCauley JL, Sawcer S, Goris A, Saarela J, Yelensky R, Price A, Leppa V, Patterson N, de Bakker PI, Tran D, Aubin C, Pobywajlo S, Rossin E, Hu X, Ashley CW, Choy E, Rioux JD, Pericak-Vance MA, Ivinson A, Booth DR, Stewart GJ, Palotie A, Peltonen L, Dubois B, et al.: **The role of the CD58 locus in multiple sclerosis.** *Proc Natl Acad Sci USA* 2009, **106**:5264-5269.
  40. Dendrou CA, Plagnol V, Fung E, Yang JH, Downes K, Cooper JD, Nutland S, Coleman G, Himsforth M, Hardy M, Burren O, Healy B, Walker NM, Koch K, Ouwehand WH, Bradley JR, Wareham NJ, Todd JA, Wicker LS: **Cell-specific protein phenotypes for the autoimmune locus IL2RA using a genotype-selectable human bioresource.** *Nat Genet* 2009, **41**:1011-1015.
  41. Takahashi K, Yamanaka S: **Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors.** *Cell* 2006, **126**:663-676.
  42. Park IH, Arora N, Huo H, Maherali N, Ahfeldt T, Shimamura A, Lensch MW,

- Cowan C, Hochedlinger K, Daley GQ: Disease-specific induced pluripotent stem cells. *Cell* 2008, **134**:877-886.
43. Vodyanik MA, Bork JA, Thomson JA, Slukvin II: Human embryonic stem cell-derived CD34+ cells: efficient production in the coculture with OP9 stromal cells and analysis of lymphohematopoietic potential. *Blood* 2005, **105**:617-626.
44. Kaufman DS, Hanson ET, Lewis RL, Auerbach R, Thomson JA: Hematopoietic colony-forming cells derived from human embryonic stem cells. *Proc Natl Acad Sci USA* 2001, **98**:10716-10721.
45. Chadwick K, Wang L, Li L, Menendez P, Murdoch B, Rouleau A, Bhatia M: Cytokines and BMP-4 promote hematopoietic differentiation of human embryonic stem cells. *Blood* 2003, **102**:906-915.
46. Zambidis ET, Peault B, Park TS, Bunz F, Civin CI: Hematopoietic differentiation of human embryonic stem cells progresses through sequential hematendothelial, primitive, and definitive stages resembling human yolk sac development. *Blood* 2005, **106**:860-870.
47. Martin CH, Woll PS, Ni Z, Zuniga-Pflucker JC, Kaufman DS: Differences in lymphocyte developmental potential between human embryonic stem cell and umbilical cord blood-derived hematopoietic progenitor cells. *Blood* 2008, **112**:2730-2737.
48. Woll PS, Martin CH, Miller JS, Kaufman DS: Human embryonic stem cell-derived NK cells acquire functional receptors and cytolytic activity. *J Immunol* 2005, **175**:5095-5103.
49. Galic Z, Kitchen SG, Kacena A, Subramanian A, Burke B, Cortado R, Zack JA: T lineage differentiation from human embryonic stem cells. *Proc Natl Acad Sci USA* 2006, **103**:11742-11747.
50. Galic Z, Kitchen SG, Subramanian A, Bristol G, Marsden MD, Balamurugan A, Kacena A, Yang O, Zack JA: Generation of T lineage cells from human embryonic stem cells in a feeder free system. *Stem Cells* 2009, **27**:100-107.
51. Steinbrook R: Personally controlled online health data-the next big thing in medical care? *N Engl J Med* 2008, **358**:1653-1656.
52. Murphy S, Churchill S, Bry L, Chueh H, Weiss S, Lazarus R, Zeng Q, Dubey A, Gainer V, Mendis M, Glaser J, Kohane I: Instrumenting the health care enterprise for discovery research in the genomic era. *Genome Res* 2009, **19**:1675-1681.
53. Silverberg MS, Cho JH, Rioux JD, McGovern DP, Wu J, Annesse V, Achkar JP, Goyette P, Scott R, Xu W, Barmada MM, Klei L, Daly MJ, Abraham C, Bayless TM, Bossa F, Griffiths AM, Ippoliti AF, Lahaie RG, Latiano A, Paré P, Proctor DD, Regueiro MD, Steinhart AH, Targan SR, Schumm LP, Kistner EO, Lee AT, Gregersen PK, Rotter JJ, et al.: Ulcerative colitis-risk loci on chromosomes 1p36 and 12q15 found by genome-wide association study. *Nat Genet* 2009, **41**:216-220.
54. Begovich AB, Carlton VE, Honigberg LA, Schrodi SJ, Chokkalingam AP, Alexander HC, Ardlie KG, Huang Q, Smith AM, Spoerke JM, Conn MT, Chang M, Chang SY, Saiki RK, Catanese JJ, Leong DU, Garcia VE, McAllister LB, Jeffery DA, Lee AT, Batliwalla F, Remmers E, Criswell LA, Seldin MF, Kastner DL, Amos CI, Sninsky JJ, Gregersen PK: A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis. *Am J Hum Genet* 2004, **75**:330-337.
55. Bottini N, Musumeci L, Alonso A, Rahmouni S, Nika K, Rostamkhani M, MacMurray J, Meloni GF, Lucarelli P, Pellecchia M, Eisenbarth GS, Comings D, Mustelin T: A functional variant of lymphoid tyrosine phosphatase is associated with type 1 diabetes. *Nat Genet* 2004, **36**:337-338.
56. Remmers EF, Plenge RM, Lee AT, Graham RR, Hom G, Behrens TW, de Bakker PI, Le JM, Lee HS, Batliwalla F, Li W, Masters SL, Booty MG, Carulli JP, Padyukov L, Alfredsson L, Klareskog L, Chen WW, Amos CI, Criswell LA, Seldin MF, Kastner DL, Gregersen PK: STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus. *N Engl J Med* 2007, **357**:977-986.
57. Plenge RM, Padyukov L, Remmers EF, Purcell S, Lee AT, Karlson EW, Wolfe F, Kastner DL, Alfredsson L, Altschuler D, Gregersen PK, Klareskog L, Rioux JD: Replication of putative candidate-gene associations with rheumatoid arthritis in >4,000 samples from North America and Sweden: association of susceptibility with PTPN22, CTLA4, and PADI4. *Am J Hum Genet* 2005, **77**:1044-1060.
58. van Heel DA, Franke L, Hunt KA, Gwilliam R, Zhernakova A, Inouye M, Wapenaar MC, Barnardo MC, Bethel G, Holmes GK, Feighery C, Jewell D, Kelleher D, Kumar P, Travis S, Walters JR, Sanders DS, Howdle P, Swift J, Playford RJ, McLaren WM, Mearin ML, Mulder CJ, McManus R, McGinnis R, Cardon LR, Deloukas P, Wijmenga C: A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21. *Nat Genet* 2007, **39**:827-829.
59. Plenge RM, Cotsapas C, Davies L, Price AL, de Bakker PI, Maller J, Pe'er I, Burtt NP, Blumenstiel B, DeFelice M, Parkin M, Barry R, Winslow W, Healy C, Graham RR, Neale BM, Izmailova E, Roubenoff R, Parker AN, Glass R, Karlson EW, Maher N, Hafler DA, Lee DM, Seldin MF, Remmers EF, Lee AT, Padyukov L, Alfredsson L, Cobylin J, et al.: Two independent alleles at 6q23 associated with risk of rheumatoid arthritis. *Nat Genet* 2007, **39**:1477-1482.
60. Thomson W, Barton A, Ke X, Eyre S, Hinks A, Bowes J, Donn R, Symmons D, Hider S, Bruce IN; Wellcome Trust Case Control Consortium, Wilson AG, Marinou I, Morgan A, Emery P; YEAR Consortium, Carter A, Steer S, Hocking L, Reid DM, Wordsworth P, Harrison P, Strachan D, Worthington J: Rheumatoid arthritis association at 6q23. *Nat Genet* 2007, **39**:1431-1433.
61. Graham RR, Cotsapas C, Davies L, Hackett R, Lessard CJ, Leon JM, Burtt NP, Guiducci C, Parkin M, Gates C, Plenge RM, Behrens TW, Wither JE, Rioux JD, Fortin PR, Graham DC, Wong AK, Vyse TJ, Daly MJ, Altschuler D, Moser KL, Gaffney PM: Genetic variants near TNFAIP3 on 6q23 are associated with systemic lupus erythematosus. *Nat Genet* 2008, **40**:1059-1061.
62. Graham RR, Kozyrev SV, Baechler EC, Reddy MV, Plenge RM, Bauer JW, Ortmann WA, Koeuth T, González Escribano MF; Argentine and Spanish Collaborative Groups, Pons-Estel B, Petri M, Daly M, Gregersen PK, Martin J, Altschuler D, Behrens TW, Alarcón-Riquelme ME: A common haplotype of interferon regulatory factor 5 (IRF5) regulates splicing and expression and is associated with increased risk of systemic lupus erythematosus. *Nat Genet* 2006, **38**:550-555.
63. Hom G, Graham RR, Modrek B, Taylor KE, Ortmann W, Garnier S, Lee AT, Chung SA, Ferreira RC, Pant PV, Ballinger DG, Kosoy R, Demirci FY, Kamboh MI, Kao AH, Tian C, Gunnarsson I, Bengtsson AA, Rantapää-Dahlqvist S, Petri M, Manzi S, Seldin MF, Rönnblom L, Syvänen AC, Criswell LA, Gregersen PK, Behrens TW: Association of systemic lupus erythematosus with C8orf13-BLK and ITGAM-ITGAX. *N Engl J Med* 2008, **358**:900-909.
64. International Multiple Sclerosis Genetics Consortium, Hafler DA, Compston A, Sawcer S, Lander ES, Daly MJ, De Jager PL, de Bakker PI, Gabriel SB, Mirel DB, Ivinson AJ, Pericak-Vance MA, Gregory SG, Rioux JD, McCauley JL, Haines JL, Barcellos LF, Cree B, Oksenberg JR, Hauser SL: Risk alleles for multiple sclerosis identified by a genomewide study. *N Engl J Med* 2007, **357**:851-862.
65. Cooper JD, Smyth DJ, Smiles AM, Plagnol V, Walker NM, Allen JE, Downes K, Barrett JC, Healy BC, Mychaleckyj JC, Warram JH, Todd JA: Meta-analysis of genome-wide association study data identifies additional type 1 diabetes risk loci. *Nat Genet* 2008, **40**:1399-1401.
66. Todd JA, Walker NM, Cooper JD, Smyth DJ, Downes K, Plagnol V, Bailey R, Nejentsev S, Field SF, Payne F, Lowe CE, Szeszeko JS, Hafler JP, Zeitels L, Yang JH, Vella A, Nutland S, Stevens HE, Schuilenburg H, Coleman G, Maisuria M, Meadows W, Smink LJ, Healy B, Burren OS, Lam AA, Ovington NR, Allen J, Adlem E, Leung HT, et al.: Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nat Genet* 2007, **39**:857-864.

doi:10.1186/gb-2010-11-5-212  
Cite this article as: Plenge R: GWASs and the age of human as the model organism for autoimmune genetic research. *Genome Biology* 2010, **11**:212.