Available online at www.sciencedirect.com

SciVerse ScienceDirect

Current Opinion in Immunology

Human genetics offers an emerging picture of common pathways and mechanisms in autoimmunity

Benjamin F Voight[1] and Chris Cotsapas[2]

In genetic studies of autoimmune and inflammatory diseases, one clear finding that has emerged from genome-wide association studies is that a substantial fraction of variation modifying risk in one disease also contributes mediate risk to multiple, additional autoimmune and inflammatory diseases. The unexpected magnitude of this overlap presents the unique opportunity to dissect the pathogenic mechanisms underlying multiple disease states in the expectation that this may lead to both more sensitive diagnostics and novel therapies. Here, we review the current evidence for this shared genetic architecture and, based on these data, outline models for shared pathways, the underlying hypotheses for them, how these models can be tested and validated.

Addresses
[1] Departments of Pharmacology and Genetics, The University of Pennsylvania – Perelman School of Medicine, Philadelphia, PA, United States
[2] Departments of Neurology and Genetics, Yale School of Medicine, New Haven, CT, United States

Corresponding authors: Voight, Benjamin F (bvoight@upenn.edu) and Cotsapas, Chris (chris.cotsapas@yale.edu)

Current Opinion in Immunology 2012, **24**:552–557

This review comes from a themed issue on **Autoimmunity: Insights from human genomics**

Edited by **Soumya Raychaudhuri** and **Stephen S Rich**

For a complete overview see the Issue and the Editorial

0952-7915/$ – see front matter,

http://dx.doi.org/10.1016/j.coi.2012.07.013

Introduction

Epidemiologic observation data for autoimmune and inflammatory diseases (AID) have long supported the hypothesis that AID share an underlying genetic basis in common. First, AID have been observed to cluster in families [1,2] and second, they co-occur in the same individuals at a rate higher than expected by chance [3•]. An obvious component of this co-occurrence is owing to the influence of shared HLA haplotypes; because variation at this locus cannot explain all of familial risk, other genetic determinants must also be shared in complex patterns of overlap [4]. However, these data have not completely explained the underlying shared biological basis of disease; the recent large-scale identification of genetic AID risk variants promises to change this.

Over the past five years, the genetics of AID have significantly advanced, with association at convincing levels of statistical significance [5] at over 200 common or rare variants [6]. These discoveries have been driven largely by genome-wide association studies (GWAS), whose aim is to test association between common genetic variation at hundreds of thousands of markers and a disease endpoint; in these studies, thousands of individuals are typically characterized [7]. Across the spectrum of AID, application of a range of informatics algorithms has generated plausible hypotheses about the causal genes and tissues underlying disease pathogenesis, motivating functional experiments designed to test these hypotheses [8•,9,10].

One clear observation is that a large fraction of associations at genome-wide levels of significance in one AID also relate to additional AID [11••]. While some shared biology might have been expected [3], the magnitude (>40% of loci) and precision of commonality was not. Several interesting and compelling observations have emerged implicating several genes with shared etiology: beyond the well-described examples at the MHC [12] and PTPN22 (see below) loci, associations near STAT3 (discussed below) and SH2B3 [13–22], amongst others appear to harbor shared risk variants. The latter is particularly intriguing as it is thought to be a T cell receptor adaptor (if indeed it is the relevant gene in the region), there are associations to many traits beyond immune disease such as coronary artery disease and blood pressure measurements and there is strong evidence of positive selection in the region [23].

A credible hypothesis that follows from the genetic data is that etiological *mechanisms* and *pathways* contributing susceptibility to multiple AID are shared among them. To address the critical aim of identifying pathways that are either shared across diseases or are unique to specific ones, it is essential to develop network models for commonalities, and consider how genetic association data might be used to distinguish between such models.

Here, we first review the evidence for a shared genetic architecture for AID, outline models for shared pathways, the underlying hypotheses for them, how such models can be generated and validated, and the implication for our understanding of disease pathogenesis.

Genetic studies of AID highlight common pathways underlying disease

GWA studies have been very successful in defining genomic regions harboring disease risk alleles. Studies in cohorts comprising tens of thousands of cases with epidemiologically matched controls have identified associations to common genetic markers across the entire

human genome, which satisfy stringent statistical criteria and have been replicated across multiple cohorts. This now includes >420 associations to >35 common, complex autoimmune or inflammatory diseases or quantitative measures related to immune function [6]. A biological interpretation and hypothesis generation is only now beginning (see accompanying reviews by Graham; Xavier; and Gregersen and Plenge in this issue).

As more associations have been reported, a substantial overlap between AID has emerged [22], with independent studies in individual diseases finding associations to the same genomic regions (for example, the region encoding TNFAIP3 on human chromosome 6 [24–26]). This has motivated several meta-analyses across pairs of diseases to establish their shared genetic basis: for example, celiac disease has been compared to rheumatoid arthritis [27], type 1 diabetes [28] and Crohn's disease [29] and in each case substantial locus overlap has been observed suggesting shared underlying pathogenesis. These observations very probably underestimate the actual extent of commonality, as individual studies lack statistical power to detect *all* true associations at genome-wide significance: Stahl *et al.* [30•] estimate that in several diseases where GWAS has been successful hundreds of further associations remain to be found. This problem compounds when considering independent discoveries across diseases, as power is multiplicative across studies. This means that, if an association is true in two diseases, the power to find it in *both* is the product of the power of each individual study. Therefore, estimates of the true extent of genetic sharing are probably underestimated by either simple overlaps or pairwise meta-analyses.

One common goal in genetic mapping studies is to identify the *causal* variant(s) and haplotypes driving genetic susceptibility in each region of association. The most associated marker in a locus may simply be tightly linked to an unassayed, causal variant, so maximal association is insufficient evidence of causality for variants. Another goal is to identify the causal gene implicated in disease pathogenesis: regions found to associate with traits often harbor multiple genes, requiring refinement before inferences on genetic mechanisms and etiological causes of disease can be made [31]. To accomplish both goals, the AID community has developed shared resources, including the *Immunochip*, a common platform for fine-mapping AID-associated loci [32,33] as well as computational approaches to select the likeliest candidate genes from regions of association [34,35••,36••].

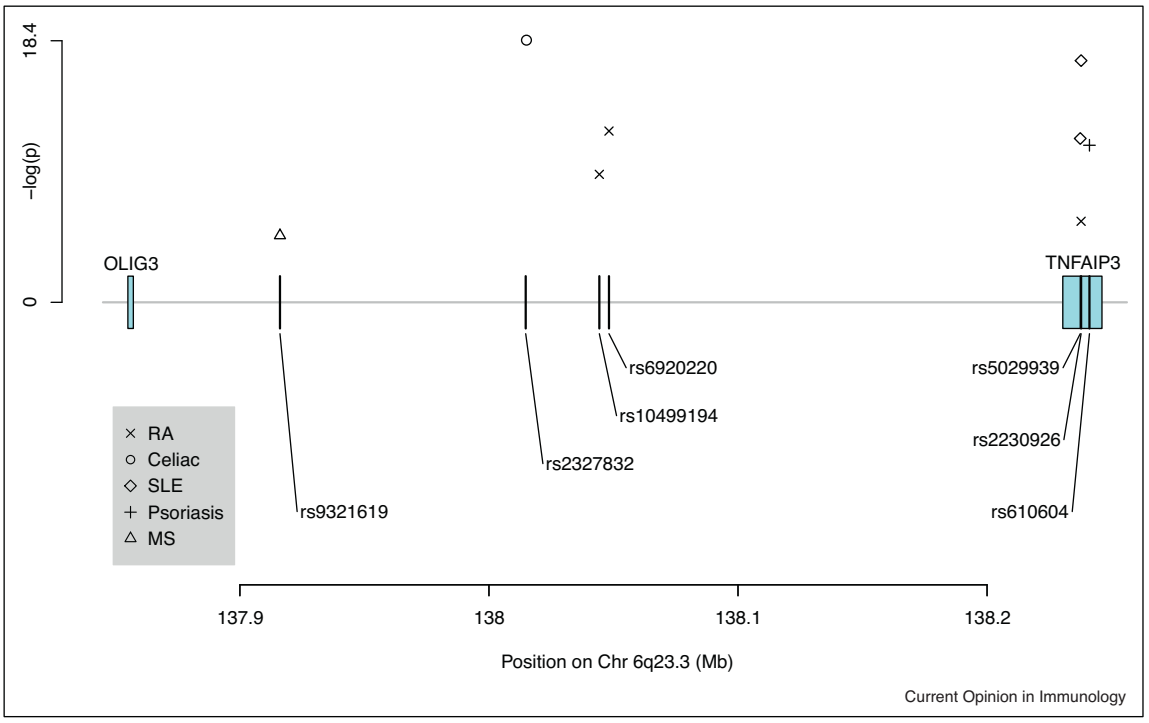
To circumvent this limitation, we developed a statistical approach to detect such simultaneous associations [11••]. Using GWAS data contributed by consortia studying celiac disease, Crohn's disease, multiple sclerosis, rheumatoid arthritis, systemic lupus erythematosus and type 1 diabetes, we have shown that 47 of 107 markers (44%) known to be associated to one AID are enriched for association to multiple – but not all seven – diseases. Grouping these loci by *which* diseases they mediate risk to revealed significant structure: groups of loci appear to influence susceptibility to the same subsets of diseases. Furthermore, the genes in these loci encode proteins which interact more often than expected by chance [8•] generating the hypothesis that risk variants perturb biological processes which mediate risk to several diseases.

Three obstacles hamper inferring pathogenic mechanisms shared across diseases from GWAS data. First, as discussed above, the level of precision in genetic mapping, in many cases, has not yet unambiguously identified causal variants. Thus, it is challenging to distinguish between genetic effects genuinely shared between diseases and independent risk variants residing in the same region of the genome. We illustrate this in Figure 1, where we present the strongest associations for multiple diseases to the locus on 6q23 encoding the candidate gene TNFAIP3. To resolve this first issue, the AID community is freely exchanging genetic data across disease areas, which now allows the development of new statistical approaches to resolve this ambiguity.

Second, increasingly precise functional maps of immune cell subsets are required. The maps will provide a resource by which the genes to pathways mapping can be inferred, and in which specific cells susceptibility to disease is determined. Genomic technologies are particularly attractive for such maps as they are amenable to automation, throughput and industrial process standards. The limiting factor is the careful acquisition of samples: rigorous standards in statistical experimental design, cell isolation, flow sorting and sample preparation are required to avoid the limitations which have hampered previous such efforts. Such experiments in highly parallel gene expression have begun, with several large-scale projects aiming to characterize lymphocyte populations currently underway (for example, GTEx [commonfund.nih.gov/GTEx/Publications.aspx] and ImmVar [www.immvar.org]). Other measurements (DNse I hypersensitivity [37], nucleotide methylation status by sequencing [38], chromatin immunoprecipitation and sequencing [39] and many others) are equally valuable and capture different types of information. With gene expression, for example, one can assess whether disease risk variants alter gene expression suggesting a mechanism of action [40•]. Similar inferences can be made from other technologies. Far more exciting is the prospect of using these data in bulk to infer gene regulatory programs for groups of risk variants that may cumulatively perturb pathways, a topic of much current research interest.

The third issue is one of interpretation: several risk alleles appear to impart risk to some diseases but are protective of others [41••,42] and it remains unclear how this evidence should be incorporated into a pathway view of disease. A well-known example in the literature is the R620W variant in PTPN22, known to confer risk to T1D [43] and the

Figure 1



Multiple disease associations at 6q23 do not clearly identify shared genetic effects. We show the location and magnitude of reported associations for RA [26], celiac disease [52], SLE [24], psoriasis [25] and multiple sclerosis [53]. It is impossible to infer shared vs. distinct associations from these data: the variants associated to RA, SLE and psoriasis in the TNFAIP3 (also known as A20) coding region itself may point to a shared effect, as may the RA and celiac associations in the center of the region. Whether these two blocks are distinct is also unclear. These ambiguities – common to many loci – can only be resolved by global analysis of all diseases.

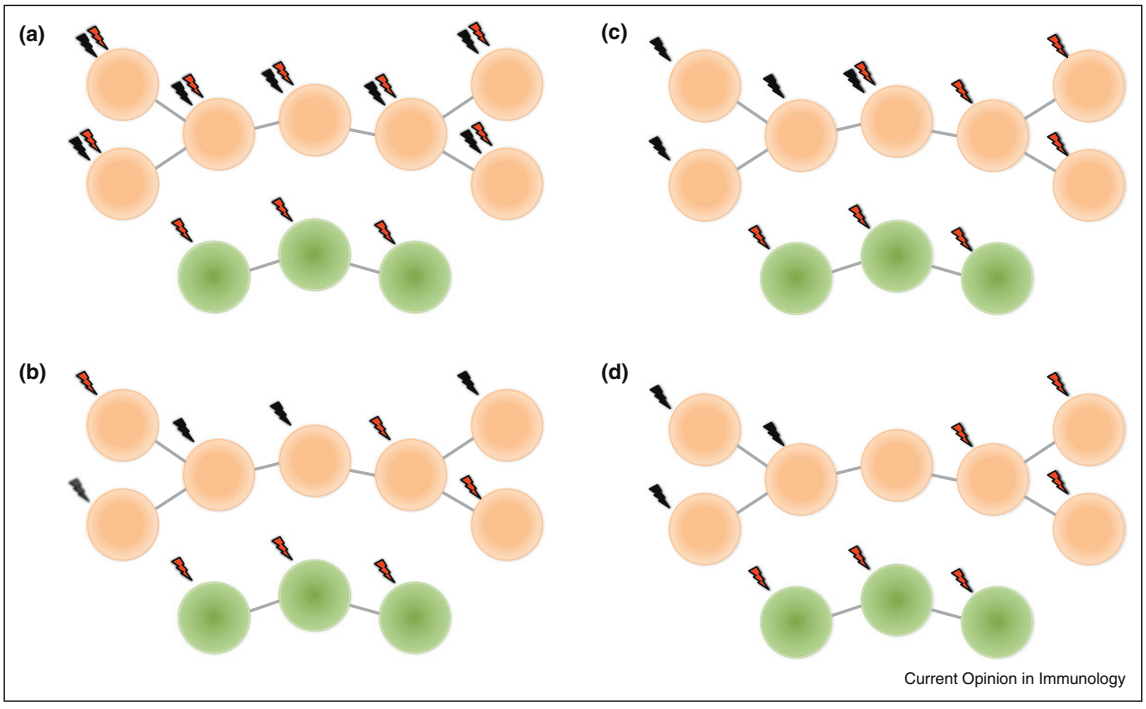
development of autoantibodies [20], RA [44] and vitiligo [45], but to be protective for Crohn's disease [46]. Another emerging example is the rs744166 polymorphism in an intron of STAT3 on chromosome 17, where the A allele confers risk to Crohn's disease [46] and the G allele to multiple sclerosis [47]. It is at present unclear whether this is a feature specific to inflammatory bowel disease. We discuss some possible explanations for these opposing risk effects in the following section. However, we can begin to outline shared pathway architecture from extant information. Under the assumption that, in each disease, causal variants perturb a limited number of cellular processes, pathway enrichment analysis [48,49] and functional genomics datasets can be queried to reveal pathway components that are preferentially encoded in associated loci. These approaches have shown that this *cumulative burden hypothesis* is true [36••], which argues that susceptibility alleles accumulate and perturb pathways to influence susceptibility (see Cotsapas and Hafler [50] for more discussion on this point). These approaches can be expanded to incorporate genetic data from multiple diseases in much the same fashion [11••]. We next discuss various models of how risk variants may congregate in shared pathways. These have implications not only for the development of methods to detect such architectures, but also the design of experiments to uncover the precise biological activities being perturbed.

Models for common AID pathways

Given the cumulative burden hypothesis, we consider models for how common pathways might be organized biologically. This is crucial because patterns of association in human genetic data across diseases could be used directly to evaluate which models (if any) apply, and the biological contexts in which they operate in practice. This will be required to devise precise, functional experiments aiming to uncover the biological processes and mechanisms that contribute to one or multiple AID.

Figure 2 depicts several models for how the underlying shared (and distinct) networks could be organized. Both shared and disease-specific mechanisms are represented by sets of interacting genes (circles connected by lines of interactions in Figure 2), perturbed by risk variants. The simplest arrangement is that a network is either private to a single disease or common to more than one. In this model, genetic variants perturbing a private network will only modify risk to one disease; those perturbing a shared network will associate to all those diseases (Figure 2a). At least some AID networks appear to have this discrete structure [11••]: the IL23R signaling network appears to mediate risk to multiple diseases in this fashion, although this has yet to be conclusively shown [51].

Figure 2



Models of shared pathway architecture. Consider two simple networks (orange and green), represented by spherical genes connected by lines if they interact in the network. A risk variant may perturb a gene to alter risk to one of two diseases (red *or* black bolt) or both (red *and* black double bolt). **(a)** The orange network is completely shared, with variants predisposing to both diseases affecting all genes. The functional consequences of perturbation should be similar for both diseases. **(b)** The network mediates risk to both diseases, but risk variants from each disease affect genes without detectable pattern. It is not clear how perturbations result in shared risk or whether the cellular context is the same across diseases. **(c, d)** Subsets of the network are disease specific and feed into a common core, which is itself either perturbed by shared risk variants or is not targeted by risk at all.

Alternative arrangements are where a shared pathway is modified by risk variants for multiple diseases essentially at random (Figure 2b) or where a modular architecture connects disease-specific pathways to a shared hub (Figure 2c,d). In the first view, any variant may predispose to one or more diseases, but the cumulative load on the pathway influences multiple outcomes. In the others, disease-specific variants act on peripheral modules of the pathway to further elevate risk to that disease, but not to others. In this configuration, it is possible that each of these modules are active in different cells or respond to different stimuli, but the higher-order structure is required for pathogenesis. This poses great challenges for experimentation and will possibly require development of new, systems-level immunological approaches to decipher correctly.

A consideration of susceptibility and protective effects across diseases as discussed above can neatly operate in the above framework. Biologically, there is no requirement that pathways contributing susceptibility to multiple diseases do so in the same place at the same time. For example, higher expression of an important gene in one tissue could contribute susceptibility for one disease, but this increase in expression protects for another in a different tissue. This may be mediated either by the same pathways in both tissues, or by different ones if the gene participates in multiple processes. Contextualizing these observations will be particularly important to address experimentally for opposite effect variants and the genes they act on. In reality, we expect multiple pathways and networks to contribute to pathogenesis and no particular model to dominate. Delineating the structure of biological networks contributing shared and distinct liability to AID provides one opportunity by which the key cellular components and mechanisms implicated in disease can be identified. After those processes are identified, this approach will also help to target the most tractable to take forward for functional experimentation in order to verify the underpinnings that contribute to AID susceptibility.

Conclusion

Has the investment in AID GWAS been a good one? We answer unambiguously in the affirmative: the landscape of spectacular shared genetic liability across multiple AID could not have been expected based on previous genetic, epidemiological, or clinical information. Because of this unexpectedly clear observation supporting shared mechanisms contributing to disease, and because of the accessibility of many relevant cellular models, immune-mediated traits are uniquely poised to uncover numerous biological and mechanistic insights. Furthermore, if the

biological underpinnings are in fact shared, discoveries in one model or system have the potential to translate quickly into others, thereby exponentially increasing the pace of discovery. We believe that the systematic dissection of multiple trait associations using genetic, genomic and immunological tools will result in rapid translation of discoveries from GWAS into biologically actionable information for functional studies and therapeutic efforts.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Crisswell LA, Pfeiffer KA, Lum RF, Gonzales B, Novitzke J, Kern M, Moser KL, Begovich AB, Carlton VE, Li W *et al.*: **Analysis of families in the Multiple Autoimmune Disease Genetics Consortium (MADGC) Collection: the PTPN22 620W allele associates with multiple autoimmune phenotypes**. *Am J Hum Genet* 2005, **76**:561-571 http://dx.doi.org/10.1086/429096.

2. Torfs C, King MC, Huey B, Malmgren J, Grumet FC: **Genetic interrelationship between insulin-dependent diabetes mellitus, the autoimmune thyroid diseases, and rheumatoid arthritis**. *Am J Hum Genet* 1986, **38**:170-187.

3. • Eaton W, Rose N, Kalaydjian A, Pedersen MG, Mortensen PB: **Epidemiology of autoimmune diseases in Denmark**. *J Autoimmun* 2007, **29**:1-9.

A comprehensive review of >5 000 000 medical records showing remarkable overlaps between AID. Of note is that no clear patterns emerge beyond the general observation of overlap as numbers are very small, even in a country-sized cohort.

4. Risch N: **Assessing the role of HLA-linked and unlinked determinants of disease**. *Am J Hum Genet* 1987, **40**:1-14.

5. Altshuler D, Daly M: **Guilt beyond a reasonable doubt**. *Nat Genet* 2007, **39**:813-815 http://dx.doi.org/10.1038/ng0707-813.

6. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA: **Potential etiologic and functional implications of genome-wide association loci for human diseases and traits**. *Proc Natl Acad Sci USA* 2009, **106**:9362-9367 http://dx.doi.org/10.1073/pnas.0903103106.

7. McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, Hirschhorn JN: **Genome-wide association studies for complex traits: consensus, uncertainty and challenges**. *Nat Rev Genet* 2008, **9**:356-369 http://dx.doi.org/10.1038/nrg2344.

8. • Dendrou CA, Plagnol V, Fung E, Yang JH, Downes K, Cooper JD, Nutland S, Coleman G, Himsworth M, Hardy M *et al.*: **Cell-specific protein phenotypes for the autoimmune locus IL2RA using a genotype-selectable human bioresource**. *Nat Genet* 2009, **41**:1011-1015 http://dx.doi.org/10.1038/ng.434.

One of the first detailed descriptions of the effect of risk alleles on cellular biology, with a T1Dassociated variant altering functionality of multiple immune cell subtypes. This complexity of subtype effects will make unravelling pathobiology affected by susceptibility variants more complex.

9. Kofler DM, Severson CA, Mousissian N, De Jager PL, Hafler DA: **The CD6 multiple sclerosis susceptibility allele is associated with alterations in CD4+ T cell proliferation**. *J Immunol* 2011, **187**:3286-3291 http://dx.doi.org/10.4049/jimmunol.1100626.

10. Xavier RJ, Huett A, Rioux JD: **Autophagy as an important process in gut homeostasis and Crohn's disease pathogenesis**. *Gut* 2008, **57**:717-720 http://dx.doi.org/10.1136/gut.2007.134254.

11. •• Cotsapas C, Voight BF, Rossin E, Lage K, Neale BM, Wallace C, Abecasis GR, Barrett JC, Behrens T, Cho J *et al.*: **Pervasive sharing of genetic effects in autoimmune disease**. *PLoS Genet* 2011, **7**:e1002254 http://dx.doi.org/10.1371/journal.pgen.1002254.t002.

Here we show that ~45% of AID genetic associations are shared between at least two diseases and that clusters of loci predisposing to the same diseases encode interacting genes, suggesting underlying shared pathways.

12. Fernando MM, Stevens CR, Walsh EC, De Jager PL, Goyette P, Plenge RM, Vyse TJ, Rioux JD *et al.*: **Defining the role of the MHC in autoimmunity: a review and pooled analysis**. *PLoS Genet* 2008, **4**:e1000024 http://dx.doi.org/10.1371/journal.pgen.1000024.g003.

13. Barrett JC, Clayton DG, Concannon P, Akolkar B, Cooper JD, Erlich HA, Julier C, Morahan G, Nerup J, Nerras C *et al.*: **Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes**. *Nat Genet* 2009, **41**:703-707 http://dx.doi.org/10.1038/ng.381.

14. Ehret GB, Munroe PB, Rice KM, Bochud M, Johnson AD, Chasman DI, Smith AV, Tobin MD, Verwoert GC, Hwang SJ *et al.*: **Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk**. *Nature* 2011 http://dx.doi.org/10.1038/nature10405.

15. Erikkson N, Tung JY, Kiefer AK, Hinds DA, Francke U, Mountain JL, Do CB: **Novel associations for hypothyroidism include known autoimmune risk loci**. *PLoS ONE* 2012, **7**:e34442 http://dx.doi.org/10.1371/journal.pone.0034442.

16. Ganesh SK, Zakai NA, van Rooij FJ, Soranzo N, Smith AV, Nalls MA, Chen MH, Kottgen A, Glazer NL, Dehghan A *et al.*: **Multiple loci influence erythrocyte phenotypes in the CHARGE Consortium**. *Nat Genet* 2009, **41**:1191-1198 http://dx.doi.org/10.1038/ng.466.

17. Gudbjartsson DF, Bjornsdottir US, Halapi E, Helgadottir A, Sulem P, Jonsdottir GM, Thorleifsson G, Helgadottir H, Steinthorsdottir V, Stefansson H *et al.*: **Sequence variants affecting eosinophil numbers associate with asthma and myocardial infarction**. *Nat Genet* 2009, **41**:342-347 http://dx.doi.org/10.1038/ng.323.

18. Ikram MK, Luijendijk HJ, Hofman A, de Jong PT, Breteler MM, Vingerling JR, Tiemeier H: **Retinal vascular calibers and risk of late-life depression: The Rotterdam Study**. *Am J Geriatr Psychiatry* 2010, **18**:452-455.

19. Levy D, Ehret GB, Rice K, Verwoert GC, Launer LJ, Dehghan A, Glazer NL, Morrison AC, Johnson AD, Aspelund T *et al.*: **Genome-wide association study of blood pressure and hypertension**. *Nat Genet* 2009, **41**:677-687 http://dx.doi.org/10.1038/ng.384.

20. Plagnol V, Howson JM, Smyth DJ, Walker N, Hafler JP, Wallace C, Stevens H, Jackson L, Simmonds MJ *et al.*: Type 1 Diabetes Genetics Consortium: **Genome-wide association analysis of autoantibody positivity in type 1 diabetes cases**. *PLoS Genet* 2011, **7**:e1002216 http://dx.doi.org/10.1371/journal.pgen.1002216.t005.

21. Schunkert H, König IR, Kathiresan S, Reilly MP, Assimes TL, Holm H, Preuss M, Stewart AF, Barbalic M, Gieger C *et al.*: **Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease**. *Nat Genet* 2011, **43**:333-338 http://dx.doi.org/10.1038/ng.784.

22. Zhernakova A, van Diemen CC, Wijmenga C: **Detecting shared pathogenesis from the shared genetics of immune-related diseases**. *Nat Rev Genet* 2009, **10**:43-55 http://dx.doi.org/10.1038/nrg2489.

23. Soranzo N, Spector TD, Mangino M, Kuhnel B, Rendon A, Teumer A, Willenborg C, Wright B, Chen L, Li M *et al.*: **A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium**. *Nat Genet* 2009, **41**:1182-1190 http://dx.doi.org/10.1038/ng.467.

24. Graham RR, Cotsapas C, Davies L, Hackett R, Lessard CJ, Leon JM, Burtt NP, Guiducci C, Parkin M, Gates C *et al.*: **Genetic variants near TNFAIP3 on 6q23 are associated with systemic lupus erythematosus**. *Nat Genet* 2008, **40**:1059-1061 http://dx.doi.org/10.1038/ng.200.

25. Nair RP, Duffin KC, Helms C, Ding J, Stuart PE, Goldgar D, Gudjonsson JE, Li Y, Tejasvi T, Feng BJ *et al.*: **Genome-wide scan reveals association of psoriasis with IL-23 and NF-κB pathways**. *Nat Genet* 2009, **41**:199-204 http://dx.doi.org/10.1038/ng.311.

26. Plenge RM, Cotsapas C, Davies L, Price AL, de Bakker PI, Maller J, Pe'er I, Burtt NP, Blumenstiel B, DeFelice M *et al.*: **Two independent alleles at 6q23 associated with risk of rheumatoid arthritis**. *Nat Genet* 2007, **39**:1477-1482 http://dx.doi.org/10.1038/ng.2007.27.

27. Zhernakova A, Stahl EA, Trynka G, Raychaudhuri S, Festen EA, Franke L, Westra HJ, Fehrmann RS, Kurreeman FA, Thomson B *et al.*: **Meta-analysis of genome-wide association studies in celiac disease and rheumatoid arthritis identifies fourteen**

non-HLA shared loci. *PLoS Genet* 2011, **7**:e1002004 http://dx.doi.org/10.1371/journal.pgen.1002004.

28. Smyth DJ, Plagnol V, Walker NM, Cooper JD, Downes K, Yang JH, Howson JM, Stevens H, McManus R, Wijmenga C *et al.*: **Shared and distinct genetic variants in type 1 diabetes and celiac disease**. *N Engl J Med* 2008, **359**:2767-2777 http://dx.doi.org/10.1056/NEJMoa0807917.

29. Festen EA, Goyette P, Green T, Boucher G, Beauchamp C, Trynka G, Dubois PC, Lagace C, Stokkers PC, Hommes DW *et al.*: **A meta-analysis of genome-wide association scans identifies IL18RAP, PTPN2 TAGAP, and PUS10 as shared risk loci for Crohn's disease and celiac disease**. *PLoS Genet* 2011, **7**:e1001283 http://dx.doi.org/10.1371/journal.pgen.1001283.

30. • Stahl EA, Wegmann D, Trynka G, Gutierrez-Achury J, Do R, Voight BF, Kraft P, Chen R, Kallberg HJ, Kurreeman FA *et al.*: **Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis**. *Nat Genet* 2012, **44**:483-489 http://dx.doi.org/10.1038/ng.2232.
Thoughtful modeling of the complex genetic architecture of RA based on current GWAS data. Results suggest hundreds more GWAS signals remain to be found and this will be a fruitful activity for the future.

31. Cantor RM, Lange K, Sinsheimer JS: **Prioritizing GWAS results: a review of statistical methods and recommendations for their application**. *Am J Hum Genet* 2010, **86**:6-22 http://dx.doi.org/10.1016/j.ajhg.2009.11.017.

32. Cortes A, Brown MA: **Promise and pitfalls of the immunochip**. *Arthritis Res Ther* 2011, **13**:101 http://dx.doi.org/10.1186/ar3204.

33. Trynka G, Hunt KA, Bockett NA, Romanos J, Mistry V, Szperl A, Bakker SF, Bardella MT, Bhaw-Rosun L, Castillejo G *et al.*: **Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease**. *Nat Genet* 2011, **43**:1193-1201 http://dx.doi.org/10.1038/ng.998.

34. Wang K, Li M, Bucan M: **Pathway-based approaches for analysis of genomewide association studies**. *Am J Hum Genet* 2007, **81**:1278-1283 http://dx.doi.org/10.1086/522374.

35. •• Raychaudhuri S, Plenge RM, Rossin EJ, Ng AC, International Schizophrenia Consortium, Purcell SM, Sklar P, Scolnick EM, Xavier RJ, Altschuler D *et al.*: **Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions**. *PLoS Genet* 2009, **5**:e1000534 http://dx.doi.org/10.1371/journal.pgen.1000534.t003.
Robust literature mining approach (GRAIL) to inferring connections between genes in regions of association is able to predict further associations to Crohn's disease and schizophrenia, suggesting risk alleles probably perturb pathogenic pathways.

36. •• Rossin EJ, Lage K, Raychaudhuri S, Xavier RJ, Tatar D, Benita Y, International Inflammatory Bowel Disease Genetics Consortium, Cotsapas C, Daly MJ: **Proteins encoded in genomic regions associated with immunemediated disease physically interact and suggest underlying biology**. *PLoS Genet* 2011, **7**:e1001273 http://dx.doi.org/10.1371/journal.pgen.1001273.
Shows that proteins encoded around risk alleles form significant interacting networks reminiscent of pathways. This approach (named DAPPLE) offers one robust way to generate biological hypotheses about pathobiology underlying association and suggests that risk variants load onto a limited set of processes.

37. Hesselberth JR, Chen X, Zhang Z, Sabo PJ, Sandstrom R, Reynolds AP, Thurman RE, Neph S, Kuehn MS, Noble WS *et al.*: **Global mapping of protein-DNA interactions in vivo by digital genomic footprinting**. *Nat Methods* 2009, **6**:283-289 http://dx.doi.org/10.1038/nmeth.1313.

38. Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, Zhang X, Bernstein BE, Nusbaum C, Jaffe DB *et al.*: **Genome-scale DNA methylation maps of pluripotent and differentiated cells**. *Nature* 2008, **454**:766-770 http://dx.doi.org/10.1038/nature07107.

39. Robertson G, Hirst M, Bainbridge M, Bilenky M, Zhao Y, Zeng T, Euskirchen G, Bernier B, Varhol R, Delaney A *et al.*: **Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing**. *Nat Methods* 2007, **4**:651-657 http://dx.doi.org/10.1038/nmeth1068.

40. • Nica AC, Montgomery SB, Dimas AS, Stranger BE, Beazley C, Barroso I, Dermitzakis ET: **Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations**. *PLoS Genet* 2010, **6**:e1000895 http://dx.doi.org/10.1371/journal.pgen.1000895.t002.
Highlights problems of simplistically overlaying association results across traits and assuming that the underlying variants and genes are the same, rather than independent effects in the same genomic region.

41. •• Sirota M, Schaub MA, Batzoglou S, Robinson WH, Butte AJ: **Autoimmune disease classification by inverse association with SNP alleles**. *PLoS Genet* 2009, **5**:e1000792 http://dx.doi.org/10.1371/journal.pgen.1000792.t002.
Description of opposite allelelic events across AID (where the allele is protective for one disease and confers risk for another).

42. Wang K, Baldassano R, Zhang H, Qu HQ, Imielinski M, Kugathasan S, Annese V, Dubinsky M, Rotter JI, Russell RK *et al.*: **Comparative genetic analysis of inflammatory bowel disease and type 1 diabetes implicates multiple loci with opposite effects**. *Hum Mol Genet* 2010, **19**:2059-2067 http://dx.doi.org/10.1093/hmg/ddq078.

43. Bottini N, Musumeci L, Alonso A, Rahmouni S, Nika K, Rostamkhani M, MacMurray J, Meloni GF, Lucarelli P, Pellecchia M *et al.*: **A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes**. *Nat Genet* 2004, **36**:337-338 http://dx.doi.org/10.1038/ng1323.

44. Begovich AB, Carlton VE, Honigberg LA, Schrodi SJ, Chokkalingam AP, Alexander HC, Ardlie KG, Huang Q, Smith AM, Spoerke JM *et al.*: **A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis**. *Am J Hum Genet* 2004, **75**:330-337 http://dx.doi.org/10.1086/422827.

45. Jin Y, Birlea SA, Fain PR, Gowan K, Riccardi SL, Holland PJ, Mailloux CM, Sufit AJ, Hutton SM, Amadi-Myers A *et al.*: **Variant of TYR and autoimmunity susceptibility loci in generalized vitiligo**. *N Engl J Med* 2010, **362**:1686-1697 http://dx.doi.org/10.1056/NEJMoa0908547.

46. Barrett JC, Hansoul S, Nicolae DL, Cho JH, Duerr RH, Rioux JD, Brant SR, Silverberg MS, Taylor KD, Barmada MM *et al.*: **Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease**. *Nat Genet* 2008, **40**:955-962 http://dx.doi.org/10.1038/ng.175.

47. Jakkula E, Leppa V, Sulonen AM, Varilo T, Kallio S, Kemppinen A, Purcell S, Koivisto K, Tienari P, Sumelahti ML *et al.*: **Genome-wide association study in a high-risk isolate for multiple sclerosis reveals associated variants in STAT3 gene**. *Am J Hum Genet* 2010, **86**:285-291 http://dx.doi.org/10.1016/j.ajhg.2010.01.017.

48. Segre AV, Groop L, Mootha VK, Daly MJ, Altshuler D: **Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits**. *PLoS Genet* 2010, **6** http://dx.doi.org/10.1371/journal.pgen.1001058.

49. Eleftherohorinou H, Wright V, Hoggart C, Hartikainen AL, Jarvelin MR, Balding D, Coin L, Levin M: **Pathway analysis of GWAS provides new insights into genetic susceptibility to 3 inflammatory diseases**. *PLoS ONE* 2009, **4**:e8068 http://dx.doi.org/10.1371/journal.pone.0008068.

50. Cotsapas C, Hafler DA: **Immune-mediated disease genetics: from risk variants to immune alterations and the shared basis of pathogenesis**. *Trends Immunol* 2012, http://dx.doi.org/10.1016/j.it.2012.09.001, in press.

51. Cho JH: **The genetics and immunopathogenesis of inflammatory bowel disease**. *Nat Rev Immunol* 2008, **8**:458-466 http://dx.doi.org/10.1038/nri2340.

52. Dubois PC, Trynka G, Franke L, Hunt KA, Romanos J, Curtotti A, Zhernakova A, Heap GA, Adany R, Aromaa A *et al.*: **Multiple common variants for celiac disease influencing immune gene expression**. *Nat Genet* 2010, **42**:295-302 http://dx.doi.org/10.1038/ng.543.

53. Patsopoulos NA, Bayer Pharma MS Genetics Working Group, Steering Committees of Studies Evaluating IFNB-1b and a CCR1-Antagonist, ANZgene Consortium, GeneMSA, International Multiple Sclerosis Genetics Consortium: Esposito F, Reischl J, Lehr S, Bauer D *et al.*: **Genome-wide meta-analysis identifies novel multiple sclerosis susceptibility loci**. *Ann Neurol* 2011, **70**:897-912 http://dx.doi.org/10.1002/ana.22609.