**Beyond disease susceptibility – leveraging GWAS for new insights into complex disease biology**

Short title: Genetics of prognosis in complex disease

Julia Bodmer Award invited review

**Author**:

James C. Lee

**Affiliation**:

Department of Medicine, University of Cambridge School of Clinical Medicine, Addenbrooke's Hospital, Cambridge, UK.

Email: [jcl65@cam.ac.uk](mailto:jcl65@cam.ac.uk)

Telephone: 01223 762639

Correspondence should be addressed to James C. Lee ([jcl65@cam.ac.uk](mailto:jcl65@cam.ac.uk))

**Abstract**

Genetic studies in complex diseases have been highly successful, but have also been largely one-dimensional: predominantly focusing on the genetic contribution to disease susceptibility. While this is undoubtedly important – indeed it is a pre-requisite for understanding the mechanisms underlying disease development – there are many other important aspects of disease biology that have received comparatively little attention. In this review, I will discuss how existing genetic data can be leveraged to provide new insights into other aspects of disease biology, why such insights could change the way we think about complex disease, and how this could provide opportunities for better therapies and/or facilitate personalized medicine. To do this, I will use the example of Crohn’s disease – a chronic form of inflammatory bowel disease that has been one of the main success stories in complex disease genetics. Indeed, thanks to genetic studies, we now have a much more detailed understanding of the processes involved in Crohn’s disease development, but still know relatively little about what determines the subsequent disease course (prognosis) and why this differs so considerably between individuals. I will discuss how we came to realise that genetic variation plays an important role in determining disease prognosis and how this has changed the way we think about Crohn’s disease genetics. This will illustrate how phenotypic data can be used to leverage new insights from genetic data and will provide a broadly applicable framework that could yield new insights into the biology of multiple diseases.

**Keywords**: GWAS, prognosis, Crohn’s disease, susceptibility, SNPs.

It is nearly 100 years since Ronald Fisher first proposed the genetic model that underpins our current understanding of complex disease genetics: that traits (or diseases) are likely to be determined by the inheritance of a sufficient number of “risk” alleles from across the genome, assuming an appropriate environmental context1. This model was in sharp contrast to traditional thinking of the time, which was based on Mendelian genetics, but proved to be a seminal insight since it explained why certain anthropometric features, such as height and hair colour, ran in families, even though they could not be explained by Mendelian genetics – a conundrum that had puzzled several eminent scientists since the late 19th century2. Interestingly, however, it then took nearly 90 years, and the advent of genome-wide association studies (GWAS), to definitively validate this model, and in so doing provide novel insights into the aetiology of complex diseases and phenotypic traits. For example, the role of autophagy in Crohn’s disease3, or of glutamate signaling and neuronal synaptic function in obesity4, had not been appreciated before single nucleotide polymorphism (SNP) associations were detected in, or near to, genes involved in these pathways. For this reason, GWAS studies of complex diseases and traits are rightly considered to have been a major success5 – especially considering the paucity of knowledge that had emerged from many years of largely unsuccessful linkage studies. Nonetheless, it is important to remember that by comparing the genetic profiles of patients who have a particular disease with those of healthy controls, conventional GWAS is designed to identify genetic variants that are involved in disease risk or susceptibility. Other aspects of disease biology, such as age at diagnosis, disease distribution, or the clinical course of disease are all ignored in conventional GWAS design. This has resulted in a notable imbalance in our understanding of complex disease genetics, for while we know much about the genetic contribution to disease risk, we understand very little about the contribution to other aspects of disease biology (or even whether there is any contribution at all). Here, I will review some insights that are beginning to emerge from genetic studies that go beyond simple analyses of disease susceptibility into other, clinically-important aspects of disease biology – particularly clinical outcome, or prognosis.

**Prognosis – a variable and unpredictable aspect of disease biology**

In almost all complex diseases, the clinical course and eventual outcome (or prognosis) of disease can vary substantially from patient to patient6-8. This variability is usually unpredictable, which makes assigning patients to the most appropriate treatment a major challenge in modern medicine9. Too potent a therapy will expose an affected individual to unnecessary risks and possible side-effects related to the medication, while a treatment that is not strong enough to control an individual’s disease will inevitably lead to disease-related morbidity and possible complications. Indeed, from a clinical point-of-view, disease prognosis is probably the most important aspect of disease biology, as it can make the difference between an excellent long-term outcome or a torrid, treatment-refractory course leading to progressive disability and/or death. Importantly, patients also prioritise disease prognosis over disease diagnosis, with questions such as “How will this disease affect my life?” being a much greater concern than “What caused this disease to develop?”. Despite this, the determinants of prognosis in most complex diseases remain poorly understood. It has been noted, however, that in families where more than one member is affected by a particular disease, the pattern of disease behavior is often similar10-12. For example, in a cohort of 262 co-affected siblings with multiple sclerosis, significant concordance was observed for measures of disability, disease progression, and eventual handicap12. This, together with similar observations in other diseases10,11, suggested that there might be a genetic contribution to disease course, and led investigators to begin to explore whether they could link specific genetic variants to disease outcome.

**The role of disease susceptibility variants in prognosis**

To date, most studies into the genetics of prognosis in autoimmune and inflammatory diseases have focused upon disease susceptibility variants. As such, the majority of these experiments have been conducted as candidate gene studies (investigating a limited number of variants rather than using a genome-wide approach) and the results have been highly variable. For example, in Crohn’s disease, individual studies have reported associations between variants in: *TLR10* and inflammatory behaviour and need for surgery13, *IRGM* and fistulating behaviour14, and *SMAD3* and need for recurrent surgery15. However, these results – and others from similar candidate gene studies – have often not been replicated in subsequent studies16-18.

One of the few loci that has been consistently linked with disease behaviour in Crohn’s disease is *NOD2*, which has been associated with need for surgery in several studies19-21. Based on these reports, it was concluded that genetic variation in *NOD2* influenced both the development of Crohn’s disease and its subsequent course – a finding that may have encouraged researchers to examine other susceptibility variants to see if they too influenced prognosis. However, it has since been shown that the link between *NOD2* variants and surgery is not an association with disease course, but rather a secondary phenomenon driven by the association of these variants with ileal Crohn’s disease (a form of Crohn’s disease that only affects the ileum – the final part of the small intestine). Crohn’s disease at this location typically only involves a short section of the intestine, and is therefore amenable to a relatively small surgical resection – an operation that carries a lower morbidity than other forms of intestinal surgery. For these reasons, surgical treatment is more commonly used for ileal disease than for other forms of Crohn’s disease (e.g. colonic Crohn’s disease) and thus the association of *NOD2* variants with surgery is actually due to the association with ileal disease, rather than any effect on disease course *per* *se*17,18,20,22. Indeed, when disease location is taken into account, no association between *NOD2* variants and disease behavior is detectable22.

This situation – where candidate gene studies have produced mixed results regarding the association of susceptibility variants with prognosis – is similar in other diseases. For example, in both multiple sclerosis and rheumatoid arthritis, several disease susceptibility variants have been reported to associate with clinical outcome, only for subsequent studies to fail to replicate these associations. Specifically, in multiple sclerosis an association was initially reported between HLA-DR15, a known susceptibility haplotype, and disease course23, only for a larger and better-powered study to fail to replicate this24. Similarly in rheumatoid arthritis, associations were originally reported between disease progression and susceptibility variants in *TRAF125* and *TNFAIP3*26 only for a larger study from the same group to fail to replicate either of these associations27. Interestingly, one exception is the “shared epitope”, a 5 amino acid sequence in HLA-DR1, which represents the strongest rheumatoid arthritis susceptibility haplotype and has also been repeatedly associated with disease severity, albeit with a much smaller effect size28. Nonetheless, as with the association of *NOD2* variants and surgery in Crohn’s disease, this association appears to be a secondary effect driven by a primary association between the shared epitope and anti-cirullinated peptide antibody-positive rheumatoid arthritis29 – a poor prognosis form of the disease.

Perhaps due to this general failure to identify prognostic associations at single variants, more recent studies have explored whether a polygenic risk score (calculated by summing the total number of risk alleles across all susceptibility SNPs) might show an association with clinical outcome. The rationale for this approach was based on the hypothesis that multiple weak effects at susceptibility SNPs would be difficult to identify individually – due to the constraints of study power – but might be detectable using an aggregate score. While this is both a reasonable hypothesis and approach, these studies have also failed to provide strong evidence for a link between susceptibility variants and disease course. For example, a study of approximately 700 Crohn’s disease patients and 400 ulcerative colitis patients did not identify any association between genetic burden (assessed using a polygenic risk score for 163 susceptibility variants) and development of complicated disease or need for bowel resection17. Similarly, a meta-analysis of 10 MS studies, including over 7,000 patients, did not detect any association between a 52 SNP polygenic risk score and disease severity30.

Collectively, therefore, the resounding message from these and other studies is that susceptibility variants do not seem to play a major role in disease prognosis. Importantly, however, this does not mean that associations would not be detected if a sufficiently large cohort were analysed (i.e. with enough power to detect very small effects). For example, a recent study by the International IBD Genetics Consortium which included approximately 20,000 Crohn’s disease patients and 15,000 ulcerative colitis patients did identify a small but significant association between a genetic risk score and the development of complicated disease22. However, the need for such large cohorts in order to detect any association at susceptibility variants highlights that if a genetic contribution to prognosis exists, it is probably mediated by non-susceptibility variants.

**How are disease susceptibility and prognosis related?**

Although susceptibility SNPs represent a convenient pool of variants to begin investigating the genetic contribution to disease prognosis, it is important to note that if these SNPs are hypothesized to affect disease course, then this implicitly suggests that common biological processes are expected to drive both the initial development and subsequent course of a disease (Figure 1A). In other words, this would imply that disease susceptibility and prognosis are not distinct entities, but rather related quantitative traits on a continuous spectrum, such that patients who carry more disease susceptibility variants would experience a worse prognosis (Figure 1A). If correct, this would have important implications for our understanding of disease and would also provide robust support for ongoing efforts to discover new therapies using GWAS data31 since it would confirm that pathways implicated in disease development remain relevant after diagnosis. Such a relationship between the total number of risk variants and the severity of the associated phenotype is certainly observed in quantitative traits, such as height or blood pressure. For example, most of the common height-associated SNPs have been reported to have an additive effect equivalent to a ~0.2 cm increase in height per allele32. Similarly, hypertension-associated SNPs have been associated with an increase in systolic blood pressure by ~1 mmHg per risk allele33. Indeed, the additive relationship between quantitative traits and their associated SNPs is so close that whole genome genetic data can be used to accurately predict traits such as height34.

In medicine, however, most diseases are not diagnosed in a quantitative fashion. Moreover, unlike genetic analyses of continuous traits, which include a quantitative measure for each individual (e.g. height in centimetres or blood pressure in mmHg), GWAS studies in complex disease only include an individual’s case / control status with no information regarding disease severity. Methodologically, it would therefore be surprising if this study design did identify prognosis-associated SNPs given that the disease cases will represent a highly admixed population with respect to prognosis. It is thus unsurprising that for most complex diseases, susceptibility variants have not shown a clear association with prognosis. Indeed, if disease development and prognosis are due to distinct biological processes, then any genetic contribution to prognosis would be expected to be driven by non-susceptibility variants (Figure 1B). As such, it is important to ensure that studies of disease prognosis – and potentially other aspects of disease biology – are not limited to disease susceptibility variants, but include other variants as well. Such a study design will also be important in order to answer the fundamental question of how disease susceptibility and prognosis are related.

**Looking beyond disease susceptibility variants**

To date, very few genetic studies of disease prognosis have considered non-susceptibility variants. This is probably because it is difficult to know which variants to include if a genome-wide approach is not used. Accordingly, the resulting situation is analogous to the state of disease susceptibility genetics prior to the advent of GWAS, when candidate gene linkage studies had provided mixed and sometimes contradictory results regarding the genetic contribution to disease development35. However, on the few occasions where a genome-wide approach has been applied to disease prognosis, the results have proven insightful, not least because the application of a genome-wide statistical threshold (*P* < 5 x10-8) has reduced concerns over false-positive associations. For example, two genetic loci were recently identified as disease modifiers of Huntington’s disease in a GWAS stratified by age at clinical onset. This result was particularly striking because Huntington’s disease is an autosomal dominant disorder where the genetic risk is entirely due to a CAG expansion in the Huntingtin gene36. As such, the realization that other genes and pathways could modify the onset of disease immediately presents new possibilities for therapeutic intervention. Similarly, a GWAS of survival in Amyotrophic Lateral Sclerosis identified two loci that influenced prognosis37, which had not previously been associated with disease development38. In Crohn’s disease, we recently performed a within-cases GWAS in which we compared the genetic profiles of patients with contrasting clinical courses of disease – defined using detailed phenotype data regarding the treatments that patients had required39. We identified four loci that were associated with prognosis in Crohn’s disease, including a haplotype in *FOXO3* that we had previously shown modulates inflammatory cytokine production in monocytes40 and an extended HLA haplotype (ancestral 8.1) that is known to lead to altered T cell responses41 – consistent with emerging evidence that the nature of T cell responses may influence prognosis in several diseases42-44. We also demonstrated that disease susceptibility loci showed no association with prognosis either individually or collectively – thus further emphasizing that the genetic architecture of prognosis is distinct from that of disease development. Intriguingly, however, we did find additional evidence that the genetic contribution to prognosis may be shared between distinct diseases. For example, the *FOXO3* haplotype that was associated with a milder course of Crohn’s disease, and reduced TNFα production, has also been associated with a milder course of rheumatoid arthritis40,45 and with more severe courses of malaria40,46 and tuberculosis47. Similarly, the *IGFBP1*/*IGFBP3* locus that was associated with more aggressive Crohn’s disease was previously shown to associate with anti-citrullinated peptide antibodies in rheumatoid arthritis48.

In addition to clarifying how genetic variation contributes to disease prognosis, these studies also highlight two important points regarding sub-phenotype studies in general. First, they demonstrate the absolute requirement for high quality, detailed phenotype data. Indeed, if the available phenotype data are insufficient to identify patient subgroups of interest, then any “within-cases” analysis are likely to produce erroneous results. This might occur because the phenotype data are insufficiently detailed, or if there are cryptic biases that affect the consistency of the data, such as differences in clinical practice between centres or inconsistencies in how certain parameters are recorded. Careful consideration must therefore be paid to subphenotype definitions in order to ensure that the resulting patient subgroups reflect the intended biological differences being studied. Second, these studies demonstrate how existing GWAS data can be re-purposed in order to yield new insights into disease biology. Since the advent of GWAS, thousands of studies have been performed resulting in the genome-wide genotyping of hundreds of thousands of individuals49. However, other than to perform ever larger meta-analyses, most of these datasets have not been re-examined since their initial publication. It would therefore appear that if sufficient phenotype data can be obtained, these historical datasets could be leveraged to obtain new insights into disease biology – and potentially avoid the costs associated with additional genotyping.

**Conclusions**

Despite the success of GWAS studies in complex disease, we are only beginning to understand how genetic variation influences aspects of disease beyond susceptibility. In Crohn’s disease, and probably many other diseases, the genetic contribution to prognosis appears to be distinct from the contribution to disease development. This finding has important implications, as the initiation of an immune-mediated disease – under the influence of susceptibility variants – is thought to occur years before the disease becomes clinically apparent50,51. The observed distinction between the genetics of disease susceptibility and prognosis therefore suggests that disease initiation is not only temporally distinct from active symptomatic disease, but also governed by separate biology. Indeed, it would appear that susceptibility variants, which are biologically relevant at the pre-clinical stage of disease, have little or even no role once disease is established. This is not entirely unexpected, as the factors influencing immunological tolerance, for example, might not be expected to remain important once tolerance is broken and an inflammatory disease has developed. Importantly, however, this does imply that using susceptibility GWAS data to prioritise candidate genes as therapeutic targets may not yield the most effective therapies. Instead, a more productive strategy might be to focus drug development efforts upon the genes and pathways that are relevant once the disease is established and that specifically determine disease outcome. Early evidence suggests that these pathways may be shared between different diseases and thus have effects on prognosis that are independent of a specific diagnosis40,44. Further elucidation of the extent of this sharing, and the biology of disease prognosis in general, would therefore appear to be an important step in the quest to deliver personalized medicine in immune-mediated disease.

**Acknowledgements**

J.C.L. is supported by a Wellcome Trust Intermediate Clinical Fellowship (105920/Z/14/Z).

**References**

1. Fisher RA. The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Philos Trans R Soc Edin*. 1918;52:399-433.

2. Galton F. Natural Inheritance. London: Macmillan; 1889.

3. Parkes M, Barrett JC, Prescott NJ, et al. Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. *Nat Genet*. 2007;39(7):830-832.

4. Locke AE, Kahali B, Berndt SI, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature*. 2015;518(7538):197-206.

5. Visscher PM, Wray NR, Zhang Q, et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am J Hum Genet*. 2017;101(1):5-22.

6. Jess T, Riis L, Vind I, et al. Changes in clinical characteristics, course, and prognosis of inflammatory bowel disease during the last 5 decades: a population-based study from Copenhagen, Denmark. *Inflamm Bowel Dis*. 2007;13(4):481-489.

7. Pincus T. Long-term outcomes in rheumatoid arthritis. *Br J Rheumatol*. 1995;34 Suppl 2:59-73.

8. Weinshenker BG, Bass B, Rice GP, et al. The natural history of multiple sclerosis: a geographically based study. I. Clinical course and disability. *Brain*. 1989;112 ( Pt 1):133-146.

9. Hamburg MA, Collins FS. The path to personalized medicine. *N Engl J Med*. 2010;363(4):301-304.

10. Satsangi J, Grootscholten C, Holt H, Jewell DP. Clinical patterns of familial inflammatory bowel disease. *Gut*. 1996;38(5):738-741.

11. Jawaheer D, Lum RF, Amos CI, Gregersen PK, Criswell LA. Clustering of disease features within 512 multicase rheumatoid arthritis families. *Arthritis Rheum*. 2004;50(3):736-741.

12. Chataway J, Mander A, Robertson N, et al. Multiple sclerosis in sibling pairs: an analysis of 250 families. *J Neurol Neurosurg Psychiatry*. 2001;71(6):757-761.

13. Morgan AR, Lam WJ, Han DY, Fraser AG, Ferguson LR. Genetic variation within TLR10 is associated with Crohn's disease in a New Zealand population. *Hum Immunol*. 2012;73(4):416-420.

14. Latiano A, Palmieri O, Cucchiara S, et al. Polymorphism of the IRGM gene might predispose to fistulizing behavior in Crohn's disease. *Am J Gastroenterol*. 2009;104(1):110-116.

15. Fowler SA, Ananthakrishnan AN, Gardet A, et al. SMAD3 gene variant is a risk factor for recurrent surgery in patients with Crohn's disease. *J Crohns Colitis*. 2014;8(8):845-851.

16. Jakobsen C, Cleynen I, Andersen PS, et al. Genetic susceptibility and genotype-phenotype association in 588 Danish children with inflammatory bowel disease. *Journal of Crohns & Colitis*. 2014;8(7):678-685.

17. Ananthakrishnan AN, Huang H, Nguyen DD, Sauk J, Yajnik V, Xavier RJ. Differential effect of genetic burden on disease phenotypes in Crohn's disease and ulcerative colitis: analysis of a North American cohort. *Am J Gastroenterol*. 2014;109(3):395-400.

18. Jung C, Colombel JF, Lemann M, et al. Genotype/phenotype analyses for 53 Crohn's disease associated genetic polymorphisms. *PLoS One*. 2012;7(12):e52223.

19. Helio T, Halme L, Lappalainen M, et al. CARD15/NOD2 gene variants are associated with familially occurring and complicated forms of Crohn's disease. *Gut*. 2003;52(4):558-562.

20. Onnie CM, Fisher SA, Prescott NJ, et al. Diverse effects of the CARD15 and IBD5 loci on clinical phenotype in 630 patients with Crohn's disease. *Eur J Gastroenterol Hepatol*. 2008;20(1):37-45.

21. Cleynen I, Gonzalez JR, Figueroa C, et al. Genetic factors conferring an increased susceptibility to develop Crohn's disease also influence disease phenotype: results from the IBDchip European Project. *Gut*. 2013;62(11):1556-1565.

22. Cleynen I, Boucher G, Jostins L, et al. Inherited determinants of Crohn's disease and ulcerative colitis phenotypes: a genetic association study. *Lancet*. 2016;387(10014):156-167.

23. Barcellos LF, Oksenberg JR, Begovich AB, et al. HLA-DR2 dose effect on susceptibility to multiple sclerosis and influence on disease course. *Am J Hum Genet*. 2003;72(3):710-716.

24. Barcellos LF, Sawcer S, Ramsay PP, et al. Heterogeneity at the HLA-DRB1 locus and risk for multiple sclerosis. *Hum Mol Genet*. 2006;15(18):2813-2824.

25. Kurreeman FA, Padyukov L, Marques RB, et al. A candidate gene approach identifies the TRAF1/C5 region as a risk factor for rheumatoid arthritis. *PLoS Med*. 2007;4(9):e278.

26. Scherer HU, van der Linden MP, Kurreeman FA, et al. Association of the 6q23 region with the rate of joint destruction in rheumatoid arthritis. *Ann Rheum Dis*. 2010;69(3):567-570.

27. Knevel R, de Rooy DP, Gregersen PK, et al. Studying associations between variants in TRAF1-C5 and TNFAIP3-OLIG3 and the progression of joint destruction in rheumatoid arthritis in multiple cohorts. *Ann Rheum Dis*. 2012;71(10):1753-1755.

28. Viatte S, Plant D, Han B, et al. Association of HLA-DRB1 haplotypes with rheumatoid arthritis severity, mortality, and treatment response. *JAMA*. 2015;313(16):1645-1656.

29. Klareskog L, Ronnelid J, Lundberg K, Padyukov L, Alfredsson L. Immunity to citrullinated proteins in rheumatoid arthritis. *Annu Rev Immunol*. 2008;26:651-675.

30. George MF, Briggs FB, Shao X, et al. Multiple sclerosis risk loci and disease severity in 7,125 individuals from 10 studies. *Neurol Genet*. 2016;2(4):e87.

31. Sanseau P, Agarwal P, Barnes MR, et al. Use of genome-wide association studies for drug repositioning. *Nat Biotechnol*. 2012;30(4):317-320.

32. Marouli E, Graff M, Medina-Gomez C, et al. Rare and low-frequency coding variants alter human adult height. *Nature*. 2017;542(7640):186-190.

33. Levy D, Ehret GB, Rice K, et al. Genome-wide association study of blood pressure and hypertension. *Nat Genet*. 2009;41(6):677-687.

34. Lello L, Avery SG, Tellier L, Vasquez A, de los Campos G, Hsu SDH. Accurate Genomic Prediction Of Human Height. *BioRXiv*. 2017.

35. Lee JC, Parkes M. Genome-wide association studies and Crohn's disease. *Brief Funct Genomics*. 2011;10(2):71-76.

36. Genetic Modifiers of Huntington's Disease C. Identification of Genetic Factors that Modify Clinical Onset of Huntington's Disease. *Cell*. 2015;162(3):516-526.

37. Fogh I, Lin K, Tiloca C, et al. Association of a Locus in the CAMTA1 Gene With Survival in Patients With Sporadic Amyotrophic Lateral Sclerosis. *JAMA Neurol*. 2016;73(7):812-820.

38. van Rheenen W, Shatunov A, Dekker AM, et al. Genome-wide association analyses identify new risk variants and the genetic architecture of amyotrophic lateral sclerosis. *Nat Genet*. 2016;48(9):1043-1048.

39. Lee JC, Biasci D, Roberts R, et al. Genome-wide association study identifies distinct genetic contributions to prognosis and susceptibility in Crohn's disease. *Nat Genet*. 2017;49(2):262-268.

40. Lee JC, Espeli M, Anderson CA, et al. Human SNP Links Differential Outcomes in Inflammatory and Infectious Disease to a FOXO3-Regulated Pathway. *Cell*. 2013;155(1):57-69.

41. Candore G, Cigna D, Todaro M, et al. T-cell activation in HLA-B8, DR3-positive individuals. Early antigen expression defect in vitro. *Hum Immunol*. 1995;42(4):289-294.

42. McKinney EF, Lyons PA, Carr EJ, et al. A CD8+ T cell transcription signature predicts prognosis in autoimmune disease. *Nat Med*. 2010;16(5):586-591, 581p following 591.

43. Lee JC, Lyons PA, McKinney EF, et al. Gene expression profiling of CD8+ T cells predicts prognosis in patients with Crohn disease and ulcerative colitis. *J Clin Invest*. 2011;121(10):4170-4179.

44. McKinney EF, Lee JC, Jayne DR, Lyons PA, Smith KG. T-cell exhaustion, co-stimulation and clinical outcome in autoimmunity and infection. *Nature*. 2015;523(7562):612-616.

45. Viatte S, Lee JC, Fu B, et al. Association Between Genetic Variation in FOXO3 and Reductions in Inflammation and Disease Activity in Inflammatory Polyarthritis. *Arthritis Rheumatol*. 2016;68(11):2629-2636.

46. Nguetse CN, Kremsner PG, Velavan TP. FOXO3A regulatory polymorphism and susceptibility to severe malaria in Gabonese children. *Immunogenetics*. 2015;67(2):67-71.

47. Lu Y, Zhu Y, Wang X, et al. FOXO3 rs12212067: T > G Association with Active Tuberculosis in Han Chinese Population. *Inflammation*. 2016;39(1):10-15.

48. Padyukov L, Seielstad M, Ong RT, et al. A genome-wide association study suggests contrasting associations in ACPA-positive versus ACPA-negative rheumatoid arthritis. *Ann Rheum Dis*. 2011;70(2):259-265.

49. MacArthur J, Bowler E, Cerezo M, et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res*. 2017;45(D1):D896-D901.

50. Arbuckle MR, McClain MT, Rubertone MV, et al. Development of autoantibodies before the clinical onset of systemic lupus erythematosus. *N Engl J Med*. 2003;349(16):1526-1533.

51. Israeli E, Grotto I, Gilburd B, et al. Anti-Saccharomyces cerevisiae and antineutrophil cytoplasmic antibodies as predictors of inflammatory bowel disease. *Gut*. 2005;54(9):1232-1236.

**Figure 1. Possible relationships between disease susceptibility and prognosis**

(A) The “continuous” model: disease susceptibility and prognosis are influenced by the same genetic variants – and thus biological pathways – on a quantitative continuum so that the burden of susceptibility variants determines the clinical course of disease (individuals who carry more susceptibility variants will experience a worse prognosis). (B) The “compartmentalised” model: disease susceptibility and prognosis are influenced by distinct sets of genetic variants – and thus distinct biological pathways – so that susceptibility variants, and the pathways they influence, play little or no role in determining disease prognosis.