Large-scale GWAS meta-analysis of serum antibody levels reveals distinct genetic architectures

Thomas W. Willis[1,4], Effrossyni Gkrania-Klotsas[2,3], Nicholas J. Wareham[2], Eoin F. McKinney[4,5], Paul A. Lyons[4,5], Kenneth G.C. Smith[6,7], and Chris Wallace[1,4,5]

[1]Medical Research Council Biostatistics Unit, University of Cambridge, Cambridge, UK
[2]Medical Research Council Epidemiology Unit, University of Cambridge, Cambridge, UK
[3]Department of Infectious Diseases, Cambridge University Hospital NHS Foundation Trust, Cambridge, UK
[4]Cambridge Institute of Therapeutic Immunology and Infectious Disease, University of Cambridge, Cambridge, UK
[5]Department of Medicine, University of Cambridge, Cambridge, UK
[6]Walter and Eliza Hall Institute of Medical Research, Melbourne, Australia
[7]Department of Medical Biology, University of Melbourne, Melbourne, Australia
* tom.willis@live.com; cew54@cam.ac.uk

Abstract

Antibodies are the principal effector proteins of humoral immunity. Dysregulated antibody production is a feature of a number of heritable immune-mediated diseases, such as the antibody deficiencies and IgA nephropathy. To characterise the common-variant architecture of serum antibody levels in the general population, we conducted the largest GWAS meta-analyses to date of serum IgA, IgM, and IgG, attaining sample sizes of 55,368, 58,777, and 85,204, respectively. We identified 82 novel associations across three isotypes. We found that isotype-specific genetic architecture was largely disjointed with few signals colocalising across isotypes in spite of their overall positive genetic correlation. We identified a large number of colocalisations between antibody phenotypes and immune-mediated diseases, and 16 between antibody phenotypes and a lymphocyte count phenotype. We report a greatly expanded catalogue of serum antibody-associated variants and characterise these in terms of their pathway context and relation to immune pathology.

Introduction

Antibodies are soluble immunoglobulins which form the basis of humoral immunity. They are antigen-binding effector proteins responsible for the neutralisation of pathogens and toxins, opsonisation, and the activation of complement (1). Immunoglobulins exist also as the membrane-bound antigen-binding moiety of cell-surface B-cell receptors (BCRs). In both guises, immunoglobulins are distinguished by the great diversity of antigen-binding affinities they provide to the adaptive immune system, together with T-cell receptors. This diversity depends upon extensive germline polygeny and polymorphism at the IGH, IGK, and IGL loci. This germline-encoded variation is supplemented by somatic hypermutation in the germinal centre.

In addition to its contribution to antibody repertoire diversity, germline genetic variation also contributes to interindividual variation in the quantity of serum antibody produced (2). Heritable influences on antibody production have long been studied in the setting of disorders of antibody production: these include antibody deficiencies, such as Bruton's agammaglobulinaemia (3) and common variable immunodeficiency (4), and hypergammaglobulinaemias, in which there is an excess of antibody production, such as the 'hyper IgM' syndromes (5). Family-based studies in healthy individuals have produced evidence of a substantial heritable component to antibody production (2,6) and the advent of genome-wide association studies (GWAS) has allowed discovery of common variants underlying this heritability (7–10).

Antibody dysfunction and the dysregulation of antibody production feature prominently in the pathogenesis of certain diseases such as IgA nephropathy, IgA vasculitis, and the aforementioned antibody deficiencies. Study of the common-variant architecture of antibody production in healthy individuals can inform understanding of the genetic basis of these diseases (7,11) as well as variation in the response to vaccination in the population at large. In this work we investigate this genetic architecture with two new GWAS of serum IgG and IgM,

and large-scale meta-analyses of IgG, IgM, and IgA. We greatly increase the number of common-variant associations, and report updated estimates of serum antibody level heritability and their genetic correlations with one another and immune-related traits. We find, contrary to our expectation and to positive phenotypic and genetic correlations overall, that the majority of significant signals detected in these large meta-GWAS are isotype-specific.

Material and methods

GWAS of serum IgG and IgM in the EPIC Cohort

We performed GWAS of serum IgG and IgM in 7,946 and 8,008 individuals, respectively. Participants were drawn from the European Prospective Investigation into Cancer (EPIC)-Norfolk Study (https://doi.org/10.22025/2019.10.105.00004). The study was carried out according to the principles of the Declaration of Helsinki, participants provided informed consent, and the study was approved by the Norwich Local Research Ethics Committee. The GWAS were performed using the methodology described previously in our study of serum IgA in EPIC (11). Briefly, IgG and IgM were quantified using immunosorbent assay; IgG was measured as the sum of the IgG1, IgG2, IgG3, and IgG4 subclasses. Serum antibody measurements were log-transformed and mean-centered. GWAS was performed for each isotype using an additive model in SNPTEST (v2.5.4-beta3) incorporating age, sex, and scores on the first ten genetic principal components.

Acquisition of public GWAS data sets

We downloaded immunoglobulin GWAS summary statistics (7,12–15) from the EBI GWAS Catalog where available or URLs provided in individual publications (Supplementary Table 1). Summary statistics from the Eldjarn study were available both with median normalisation and without it, and we chose to use datasets to which median normalisation had been applied to match the use of the same in the Pietzner datasets. Liu et al. published two IgA meta-analyses

distinguished by the inclusion or exclusion of data from a study of serum IgA in the deCODE cohort (16). To prevent the inclusion of samples overlapping with those in the Eldjarn study in our own meta-analysis, we used the Liu meta-analysis excluding the deCODE data.

GWAS summary statistics for immune-related traits and diseases (8,17–28) were downloaded from the URLs in Supplementary Table 2.

Harmonisation and standardisation of GWAS data sets

Prior to downstream analysis, we processed those GWAS datasets which were not available in a harmonised format from the GWAS Catalog with version 1.1.10 of the EMBL-EBI GWAS summary statistics harmoniser (29). This process lifts over variant coordinates to the GRCh38 assembly, infers the strand orientation of palindromic variants, and aligns alleles to the Ensembl 95 reference.

The phenotypic values used to produce each GWAS dataset were subject to transformation and standardisation by their respective authors, but the method chosen differed between studies (Supplementary Table 3). These differences produced phenotypic values, and therefore effect estimates and standard errors, on different scales. To correct for this we estimated the standard deviation of the phenotype in each dataset, exploiting the following relationship between phenotypic variance *var(Y)*, genotypic variance *var(X)*, the sample size *n*, and the variance of the effect estimate *var(β)*:

$$var(\beta) \cong \frac{var(Y)}{n * var(X)}$$

where genotypic variance is given by:

$$var(X) = 2 * f * (1 - f),$$

with *f* the minor allele frequency estimated using 1kGP European samples. We regressed *n* * *var(X)* on the reciprocal of *var(β)* and obtained an estimate of *var(Y)* as the regression slope. We conducted 10 replicates of this regression in each dataset using 10,000 SNPs randomly

selected after LD pruning using *PLINK* version 2.00a5.10 (30,31) with a 1,000 kb window, a step size of one, and $r^2 < 0.2$. We used the square root of the median of the estimates thus obtained as the estimated standard deviation, which we in turn used as a scaling factor to divide effect estimates and standard errors in each dataset prior to meta-analysis.

GWAS meta-analysis

We performed an outer join between the SNPs of each GWAS for each antibody isotype and then a fixed-effect, inverse variance-weighted meta-analysis.

We designated as candidate 'lead SNPs' those variants with a p-value $< 5x10^{-8}$ and also with the smallest p-value in the 1Mb interval centered upon the variant. We visually inspected the surrounding 4Mb interval using *locuszoomr* (32), and excluded lead SNPs which appeared to be in LD with a more significantly associated SNP. Estimates of r^2 reported between pairs of variants in the text were obtained from 1kGP European samples using the LDlink web application (33).

We assigned possible causal genes to each lead SNP using the Open Targets Platform (34) and Ensembl version 113 (35). Genes were assigned to a variant if the variant was missense in the gene or if the variant was in the fine-mapping credible set for any molecular QTL for the gene. Lead variants with no gene thus assigned were assigned to their nearest gene using metadata from Ensembl.

In order to identify novel associations, we downloaded known associations reported in the EBI GWAS Catalog on February 4th 2025[8]. We added associations from the GWAS performed by Gudjonsson et al. (36); (omitted our meta-analyses due to the possibility of sample overlap with the Icelandic deCODE cohort studied by Eldjarn et al). For IgA, we added associations reported in Table 2 of the study by Liu et al. (used in our meta-analysis) after lifting them over to

GRCh38, the associations we reported in our earlier meta-analysis of IgA (11). Lastly, we added associations present in the studies included in our meta-analyses except the studies of IgG and IgM in the EPIC Cohort. For each of our isotype-specific meta-analyses, we then merged the list of lead SNPs from signals attaining genome-wide significance with our compiled list of known associations using an 'overlap join': where no existing association's reported lead SNP lay within 500kb window up- or downstream of the lead SNP in our own meta-analysis, we designated the latter as a novel association.

We downloaded the 2024 edition of the IUIS' classification table of inborn errors of immunity (37) from the IUIS website (https://iuis.org/committees/iei/) and obtained the HGNC symbol and coordinates of the genes listed in this table from Ensembl using biomaRt (38,39). We mapped SNPs to IEI genes which lay within 200kb.

Estimation of trait heritability and the genetic correlation

We used the *SumHer* component of *LDAK* version 5.1 to estimate heritability and the genetic correlation between traits (40). We used the 'Human Default' heritability model when estimating the trait heritability, and the 'LDAK-Thin' heritability model and 'SumHer-GC' model of confounding when estimating the genetic correlation.

As described below, we identified highly significant and possibly spurious association signals in the *IGH* and *IGK* loci (Supplementary Text). We removed these loci (chromosome 14, positions 105,586,437 to 106,879,844, and chromosome 2, positions 88,857,361 to 90,235,368 in the GRCh38.p14 assembly, GenBank Gene IDs 3492 and 50802, respectively) and the *IGL* locus (chromosome 22, positions 22,026,076 to 22,922,913, GenBank Gene ID 3535) for the purposes of heritability and genetic correlation estimation. For the *IGH* and *IGK* loci, it was necessary to exclude a larger region, 103,000,000 to 107,00,000 and 88,700,000 to 90,235,368,

respectively to fully ablate the association signals.

We report heritability estimates with and without inclusion of the major histocompatibility locus (MHC), which we defined liberally as the interval on chromosome 6 between positions 24,000,000 and 45,000,000. The MHC was excluded from datasets for the purpose of genetic correlation estimation.

We estimated the phenotypic correlation of isotype-specific serum antibody levels using data from the EPIC Cohort using Pearson correlation for isotype pairs using both raw and log-transformed data, and calculated bootstrap confidence intervals for each estimate using 1,000 replicates.

We used SNP genotype data from 535 European ancestry samples from the Phase 3 release of the 1000 Genomes Project (1kGP) (41) for out-of-sample estimation of the squared correlation of SNP genotypes (*i.e.* r^2) and the calculation of *SumHer* SNP taggings. We downloaded data from the high-coverage GRCh38 release (42) and processed it using *PLINK* version 2.00a5.10 to meet the following criteria: SNPs only; MAF > 0.005; missingness per SNP < 0.01; missingness per sample < 0.01; exact test of Hardy-Weinberg equilibrium has p-value > 1e-50. We enforced alignment of reference and alternative alleles to the GRCh38 reference sequence.

Pathway analysis

To identify biological pathways highlighted by our genetic associations, we first created a test set of genes from the union of possible causal genes across our meta-analyses of IgA, IgG and IgM. We then used the *enrichKegg* function from the *clusterProfiler* package (43), and the *enrichPathway* function from the *reactomePA* package (44) to test for association with pathways annotated by KEGG and Reactome, respectively.

Colocalisation

We carried out colocalisation analysis using the *coloc::coloc.abf* function in *coloc* version 5.2.4 (45). We identified candidate serum antibody loci for colocalisation by searching for pairs of lead SNPs for distinct isotypes which lay within 1Mb of one another. For each pair of lead SNPs thus identified, we included in the analysis all SNPs lying in the interval between the lead SNPs as well as those in the 250kb flank regions lying up- and downstream of the same interval. For colocalisation of loci for serum antibody and non-serum antibody traits, we searched a 1Mb interval centered on each serum antibody lead SNP for SNPs reaching genome-wide significance ($p < 5x10^{-8}$) for the non-antibody trait. Where such SNPs were present, we carried out colocalisation using the SNPs in the 1Mb interval.

Our serum antibody meta-analyses incorporate multiple studies which differ in the panel of SNPs for which they report summary statistics. As a consequence, the sample size of each meta-analysis varies across SNPs. Approximate Bayes factors do not account for sample size imbalance of this kind (46) and the resulting miscalibration of fine-mapping (47) may propagate to colocalisation analysis. We conducted a sensitivity analysis, filtering the SNPs included in colocalisation analysis by their sample fraction, *i.e.* the SNP-specific sample size expressed as a proportion of the largest sample size in the meta-analysis. We retained only those SNPs for which the sample fraction in each meta-analysis differed by no more than 0.1. We found that filtering rarely affected the reported posterior probability of a shared causal variant, with one exception (Supplementary Figure 1).

Results

We conducted fixed-effect, inverse variance-weighted meta-analyses of IgG, IgM, and IgA datasets (Figure 1, Supplementary Table 1), attaining sample sizes of 55,368, 58,777, and

85,204, for these isotypes, respectively. We observed extremely strong associations between variants in the *IGH* and *IGK* loci and serum antibody levels in multiple studies included in our meta-analyses of each isotype (Supplementary Table 4). We excluded the *IGH*, *IGK*, and *IGL* loci from downstream analyses owing to the possibility of these associations being spurious (Supplementary Text). In total we identified 14 associations with serum IgG, of which 8 were novel (Supplementary Table 5), 82 associations with serum IgM (36 novel, Supplementary Table I6) and 77 associations with serum IgA (38 novel, Supplementary Table 7), making for a total of 82 novel associations across all three isotypes (Table 1). We obtained heritability estimates for each isotype using these data (Supplementary Table 8).

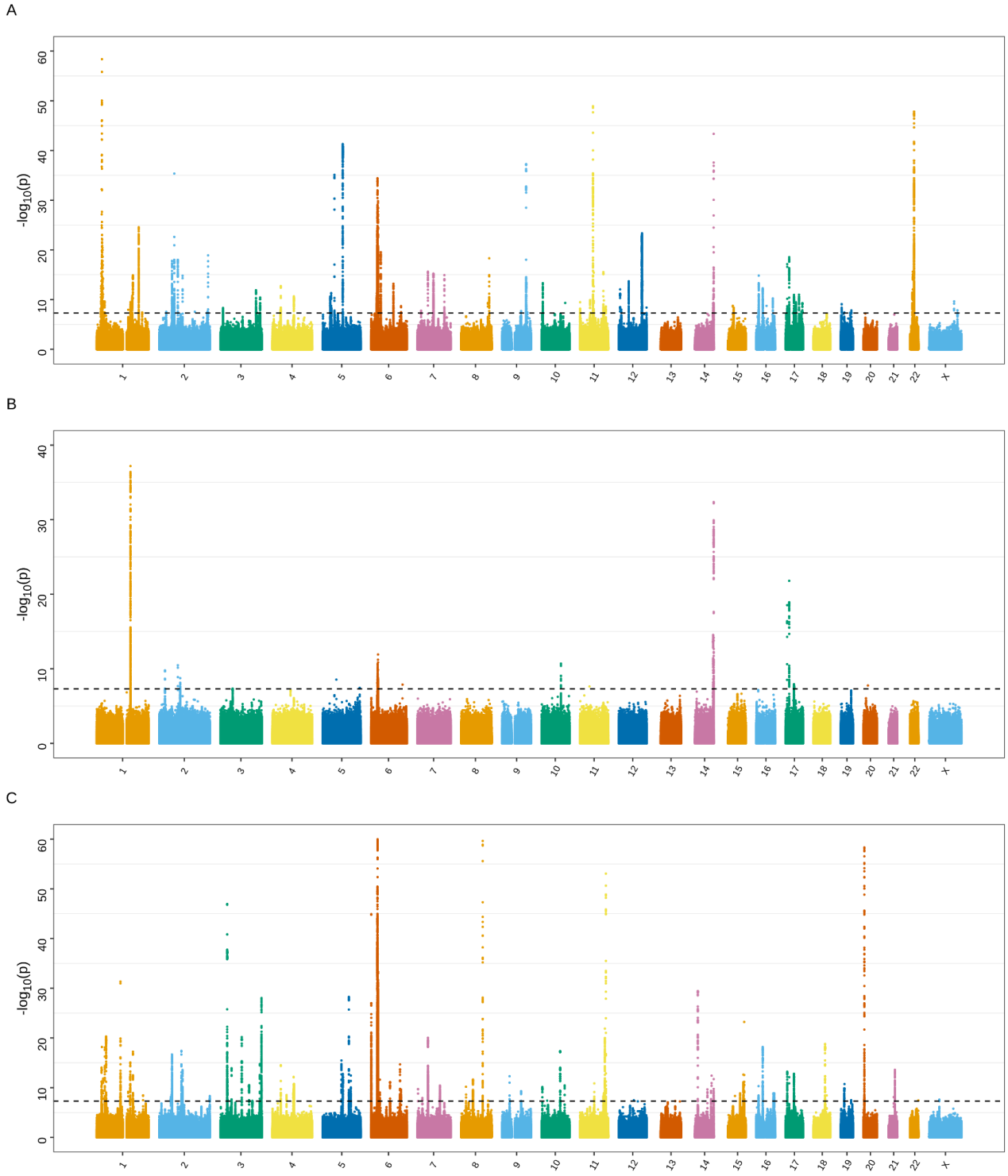


Figure 1. Manhattan plots of GWAS results for (A) IgA (B) IgG (C) IgM.

All serum immunoglobulin measures were positively correlated on the log scale (Supplementary Table 9). Concordant with this finding were significant positive genetic correlation (r_g) estimates

between serum IgA and IgG, and IgG and IgM (Table 2). In spite of these, however, we found only one locus at which there was strong evidence of a causal variant shared between IgG and IgA (Supplementary Figure 2), and none at which this was true for IgG and IgM (Supplementary Table 10). This may be explained in part by the much smaller number of IgG associations, consistent with its relatively lower heritability estimate (Supplementary Table 8).

Although IgA and IgM did not show significant genetic correlation overall, six associations with each isotype showed evidence of a causal variant common to both isotypes in colocalisation analysis (Supplementary Table 10, Supplementary Figures 2-10). For five of these colocalising associations, the direction of effect on serum IgA and IgM was concordant. The exception was on chromosome 4 near *RHOH,* for which associations the IgA and IgM lead SNPs had discordant effects on the level of each isotype. *RHOH* encodes the haematopoiesis-specific RhoH GTP-binding protein (48) and mutations in the gene cause an autosomal recessive combined immunodeficiency (49). RhoH-deficient mice show normal IgM production but impaired production of class-switched IgG1 in response to immunisation with T cell-dependent antigen (50), suggesting the *RHOH*-mapped variants affect serum antibody levels through an impact on the efficiency of class switch recombination. A combination of shared signals, some with the same effect direction (immunoglobulin production) and others with the opposite (class switching efficiency), could explain the lack of overall genetic correlation between these isotypes.

The colocalising associations with concordant direction of effect included a novel association with IgA (lead SNP rs2872516) which colocalised with a known association of the same locus with IgM (rs4795397). However, another novel association with IgG (rs17676923) in this region appeared to relate to a distinct causal variant. This was supported by an examination of LD in the region which showed that the lead SNP for IgA and IgG were not in LD (r^2 = 0.03), although

those for IgA and IgM were (r^2 = 0.83). All three SNPs belong to eQTL credible sets for *GSDMA*, *GSDMB*, and *IKZF3* in blood.

Given the overlap of association signals, we used pathway analysis to summarise the associations seen across all isotypes. This identified six and ten significant KEGG and Reactome pathways, respectively (q value < 0.05, Figure 2), the most prominent amongst them relating to B cell receptor and NF-kB signalling. Our gene selection strategy was deliberately broad, which sometimes leads to the nomination of multiple genes for a single association. By comparing this list of genes with the list of genes in associated pathways we can highlight more likely causal genes in some cases (Supplementary Figure 11). Of note, we see that the associated pathways typically contain associations with more than one isotype, highlighting the convergent biology of the GWAS associations in spite of the limited evidence of shared signals.

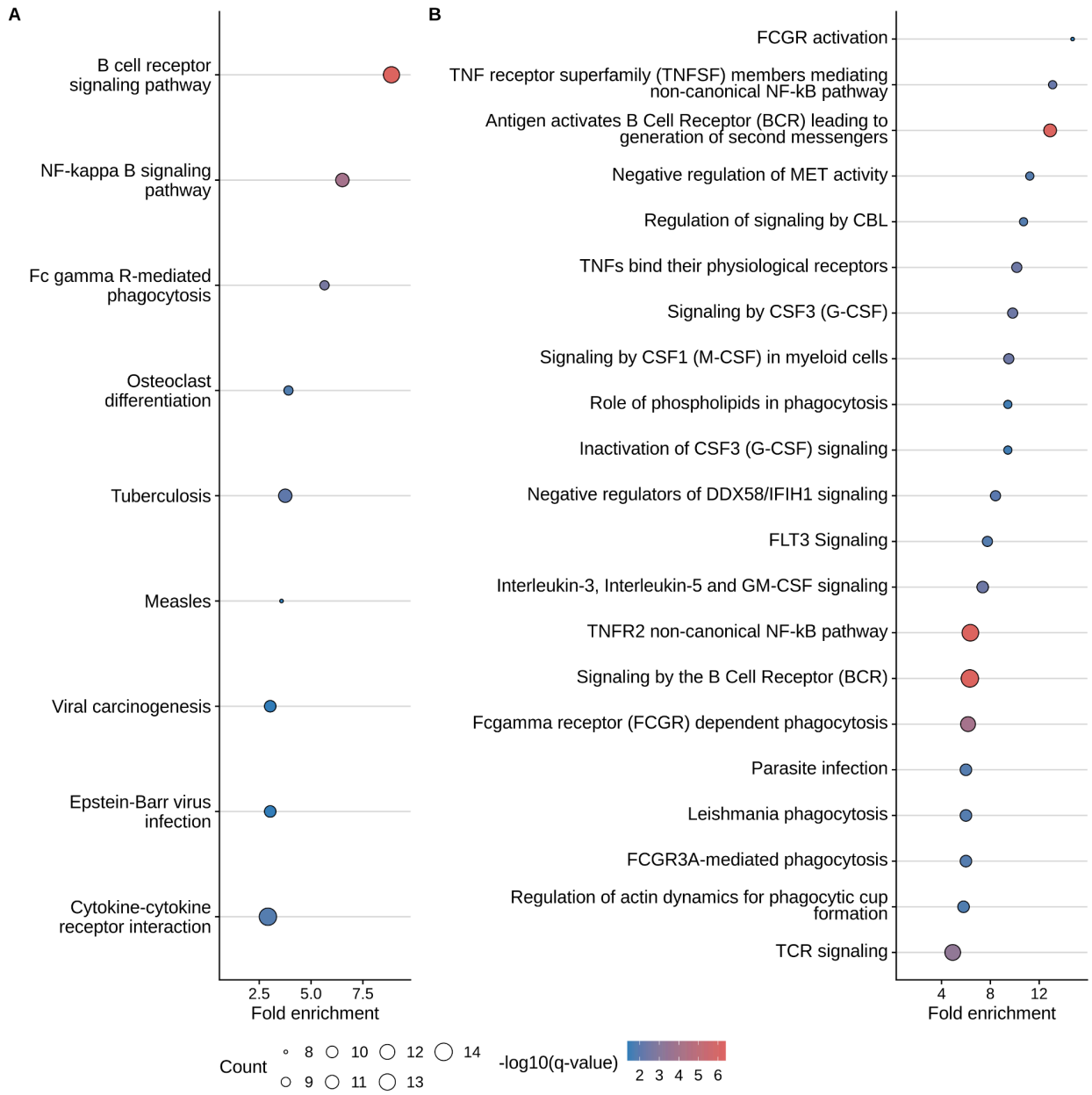


Figure 2. Significant (q-value < 0.05) pathways in (a) KEGG and (b) Reactome pathway analysis of the union of potential causal genes nominated for any isotype. The 'fold enrichment' is given by the 'ratio of the frequency of input genes annotated in a term to the frequency of all genes annotated to that term' (43).

As antibodies are secreted by plasma cells and plasmablasts derived from B cells, we hypothesised that some signals may relate to B cell abundance. While we could not find any sufficiently large GWAS of B cell count for use in colocalisation, total lymphocyte count has been studied in large cohorts, and we used one such GWAS as a proxy (28). We found a

modest but significant genetic correlation between serum IgA and lymphocyte count (r_g = 0.09, p = 3 x 10^{-4}). Whilst r_g estimates for IgM and IgG were not significant, we conducted colocalisation analyses for all three isotypes with lymphocyte counts as r_g had not proved a dispositive indicator of cross-isotype colocalisation.

We found 16 loci with strong evidence of a shared causal variant for serum antibody and lymphocyte count phenotypes (Supplementary Table 11): nine for IgA, six for IgM and one for IgG, including the above-mentioned signal near *GSDMB*, which colocalised with both IgA and IgM. However, the directions of effect were not consistent: higher immunoglobulin levels were associated with higher lymphocyte counts in eight cases and with lower lymphocyte counts in the other eight. This explains the lack of detectable genetic correlation, which is a genome-wide measure. Our interpretation of the ostensibly paradoxical relationship in the latter eight cases is hampered by the lack of B cell-specific data. Given lymphocyte count is the sum of B, T, and natural killer cell counts, an increased total lymphocyte count could arise from the production of more B cells, or fewer in the context of more T cells, or a change in T- or natural killer cell count absent any change in B-cell count.

All isotypes showed significant genetic correlation with two or more immune-mediated diseases (IMDs), amounting to nine significant estimates (Supplementary Table 11). These were positive with the exception of a negative genetic correlation between IgG and asthma (r_g = -0.1). We observed the strongest correlation between serum IgA and IgA nephropathy (r_g = 0.33), replicating published work (7). In spite of this genome-wide correlation, however, we observed no colocalisations of IgA and IgA nephropathy; the latter colocalised with IgG only, at a signal centered on the *TNFRSF13B* missense variant rs34562254.

Several loci well-known for their association with IMDs colocalised with serum antibody signals, bolstering the evidence for the biological salience of the associations identified in our meta-analyses (Supplementary Figure 12). IgA associations signals at the *IL7R*, *INAVA*, *UBE2L3*, and *IL2RA* loci colocalised with two to four IMDs each and at all loci SNP effects were positively correlated between the serum IgA and IMD phenotypes. We observed antibody-IMD signal colocalisation at the aforementioned *GSDMB* locus for six isotype-IMD pairs, including all three studied isotypes. The IgM association signal centered upon the *PTPN22* missense variant rs2476601 colocalised strongly with those of rheumatoid arthritis, type 1 diabetes, Crohn's disease, and hypothyroidism, IMDs whose association with this mutation is well-known (21,25,51). In general we observed that the direction of effect for serum antibody- and IMD-associated SNPs was neither consistently concordant nor discordant (Supplementary Figure 12).

68 of our associations were in or near known genes associated with inborn errors of immunity (IEIs); 32 of these associations were novel (Supplementary Table 13). With the exception of the more common antibody deficiencies, IEI phenotypes do not admit formal colocalisation analyses for want of summary statistics and evidence of common-variant architecture. For IgM, these associations included variants in CARD11, LYN, and CIITA as well as rs12554596, an intronic variant in and eQTL for *PAX5*, which encodes a transcription factor and 'master regulator' of B cell development (Medvedovic et al., 2011). A single case of PAX5 deficiency is known, featuring B-cell lymphopaenia and hypogammaglobulinaemia (Kaiser et al., 2022). In two cases, variants associated with IgA and IgM appeared to represent distinct causal variants (on the basis of low LD and *coloc* favouring the distinct causal variant hypothesis) but both variants were linked to the same gene. IgM-associated rs290243 and IgA-associated rs4744011 are not in LD (r^2=0.18), yet both belong to eQTL credible sets for *SYK* in NK cells and monocytes, respectively. SYK encodes a non-receptor tyrosine kinase which participates in BCR signalling

(Ackermann et al., 2015) and in which gain-of-function mutations cause an autoinflammatory disorder (L. Wang et al., 2021). Similarly, rs6589225 and rs4938518 were associated with IgM and IgA respectively. Again, we observed these to be in low LD (r^2=0.28) whilst belonging to eQTL credible sets for the same gene, *POU2AF1*. A loss-of-function mutation in *POU2AF1* has been observed to cause an autosomal recessive, CVID-like antibody deficiency, BOB1 deficiency, in one patient (Kury et al., 2021).

Other notable associations include IgA with rs1585213, an intronic variant in *NFKB1*, which encodes NF-κB, a critical immune regulator downstream of TLR and B-cell receptor (BCR) signalling and integral to development of the humoral response (Guldenpfennig et al., 2023). Loss-of-function mutations in *NFKB1* have been identified as the most common cause of common variable immunodeficiency (Tuijnenburg et al., 2018), an antibody deficiency characterised by IgG deficiency alongside IgA and/or IgM deficiency; IgA is absent in a majority of patients (Bonilla et al., 2016). Similarly, rs6890853 belongs to e-, s-, and pQTL credible sets for IL7R/IL7Rα, whose expression and signalling is important in the early stage of B-cell progenitor development. Loss-of-function mutations in IL7R cause the B cell-positive severe combined immunodeficiency IL7R deficiency (Kaiser et al., 2023).

Just as our pathway analyses related association-related genes to B-cell biology, we could relate many individual novel antibody associations to the same. We found three novel IgA associations featuring missense variants in B cell-related genes. rs13388394 is located in *RASGRP3*, which encodes a guanine nucleotide exchange factor and RAS activator. Its murine orthologue *Rasgrp3* contributes to B-cell receptor-induced Ras activation and the deletion of *Rasgrp3* leads to deficient production of certain IgG subclasses (Coughlin et al., 2005). rs113495277 is a missense variant in *TRAF1*. TNF receptor associated factor 1 (TRAF1) forms multimeric complexes with TRAF2 which bind to the cytoplasmic domain of CD40, the

cell-surface receptor for the costimulatory ligand CD40L, upon ligation and activate the canonical NF-κB signalling pathway (Elgueta et al., 2009; Fotin-Mleczek et al., 2004). rs3177243 is a missense variant in DERL3, a gene recently associated with B cell activation and differentiation, and the unfolded protein response (Lin et al., 2022). Its association with the latter may explain its expression in plasma cells given their burden of secretory protein synthesis; rs3177243 is also associated with an IgG N-glycan phenotype (Landini et al., 2022) and non-albumin serum protein levels (Sinnott-Armstrong et al., 2021). We also identified a novel IgM-associated missense variant, rs11666267, in *NIBAN3*, which encodes B cell novel protein 1, a regulator of BCR signalling (52).

Other variants were linked to genes involved in B cell-related malignancies. This included two IgA-associated variants, rs55901664 and rs17505102, which were intronic to the genes TBL1XR1 and TP63, respectively. Somatic mutations in TBL1XR1 cause impaired plasma cell differentiation and a bias towards the production of immature memory B cells in aggressive B cell lymphomas (Venturutti et al., 2020), while TP63 is known to participate in somatic gene fusion events with TBL1XR1 in a small minority of diffuse large B-cell lymphomas (Scott et al., 2012).The IgG-associated variant rs184552867 is an intronic variant in PRDM11 which has been identified as a tumour suppressor gene in B cells and which may be downregulated in diffuse large B-cell lymphoma (Asmar et al., 2010).

Discussion

We detected positive phenotypic and genetic correlations between antibody isotypes but limited overlap in their individual association signals. This suggests that our cross-isotype genetic correlation estimates are driven by shared variant effects which do not achieve genome-wide significance despite our large sample sizes. Our list of associations thus remains an incomplete catalogue of the common-variant determinants of serum antibody production; we also report a

SNP heritability estimate for serum IgA (0.18) which falls far short of the heritability estimate obtained from a past twin study (0.50) and the SNP heritability estimate (0.31) obtained in the same twin cohort (6). This disparity may be explained in part by our exclusion of the *IGH* and *IGK* loci. The effects we did detect reflect both known and novel biological mechanisms impacting immunoglobulin secretion, including B-cell progenitor development, antibody production and class switching, and the role of helper T cells in stimulating immunoglobulin production. Despite comparably sized sample sizes for IgG and IgM, we detected far fewer associations for the former and estimated its heritability to be around one third of the latter's.

We found evidence of common-variant architecture shared between the antibody phenotypes and immune-related traits. For the latter we observed multiple colocalisation signals indicative of shared causal variants in the absence of strong genetic correlation between lymphocyte count and the antibody phenotypes. Our expectation of straightforward concordance between antibody and lymphocyte count effect directions was refuted by the lack of a clear relationship. We observed the same inconsistency in effect direction across antibody signals colocalising with immune-mediated disease signals. These results suggest that immunoglobulin levels are not themselves on the causal path for any of these diseases, with the probable exception of IgA nephropathy (7). The shared associations instead reflect processes that lead to change in both immune-mediated disease risk and immunoglobulin levels. Our results also provide a cautionary tale for the use of genetic correlation more generally: weak or absent genetic correlation, as with lymphocyte counts, should not be taken as a dispositive indicator of genetic independence, and positive genetic correlation may be consistent with the presence of shared causal variants which nonetheless exert discordant effects on each phenotype.

We have presented an expanded view of the genetic architecture of serum antibody production. This view will be broadened still further by the adoption of modern high-throughput affinity-based

proteomics techniques, such as the 'SomaScan' DNA aptamer-based method used to generate the Pietzner and Eldjarn datasets included in our meta-analyses (53). These methods allow many proteins to be assayed in parallel and should make the measurement of serum proteins in large, biobank-sized cohorts affordable (54). Investigation of the pattern of shared genetic effects on lymphocyte abundance and antibody production, and the seemingly paradoxical disagreement in the direction of variant effects upon them, was hindered by the lack of large, publicly available B and T cell-specific phenotype GWAS. Should such datasets become accessible, their use in colocalisation analysis should prove insightful as to the extent to which variants mediate their effect on serum antibody levels through proximal effects on lymphocyte abundance.

As well as IMDs generally, our results are relevant to the study of the antibody deficiencies, particularly the two most common, selective IgA deficiency and common variable immunodeficiency (4). The genetic aetiology of these diseases remains obscure in most patients and there is evidence of a common-variant contribution in both (11,55,56). They are therefore good candidates for joint genetic analysis with the meta-analyses we publish here. Such an analysis could address the hypothesis of whether inheritance of a preponderance of common hypomorphic variants could account for clinical antibody deficiency in cases where a Mendelian pattern of disease or known monogenic causal variant are absent.

Data and code availability

GWAS summary statistics have been deposited at the following Zenodo URL: https://doi.org/10.5281/zenodo.17010403.

We have made available the code used to produce the results presented here in the form of a *snakemake* (57) pipeline: https://github.com/twillis209/serum_ig_pipeline. The standalone

snakemake pipeline used to process 1kGP data is also available: https://github.com/twillis209/1kGP_pipeline.

Acknowledgements

We would like to acknowledge the participants of the Fenland Study (doi: 10.22025/2017.10.101.00001). We thank Christopher Benson, Debora Lucarelli, and Jing Hua Zhao for their contribution to the generation of data used in this paper. We wish to acknowledge all GWAS participants, in particular those of FinnGen, for their contribution to the data used herein. We also acknowledge the investigators who carried out these GWAS and made their summary statistics publicly available. We are grateful to all the participants of the EPIC-Norfolk study and to the many members of the study teams at the University of Cambridge who have enabled this research.

Author contributions

Thomas W. Willis: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. Effrossyni Gkrania-Klotsas: Writing – review & editing, Methodology, Investigation, Funding acquisition, Data curation. Nicholas J. Wareham: Writing – review & editing, Funding acquisition, Data curation. Eoin F. McKinney: Writing – review & editing, Data curation. Paul A. Lyons: Writing – review & editing, Funding acquisition, Data curation. Kenneth G.C. Smith: Writing – review & editing, Funding acquisition. Chris Wallace: Writing – review & editing, Writing – original draft, Visualization, Supervision, Software, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Conflicts of Interest

CW is a part time employee of GSK. GSK had no role in this study or its publication.

Funding

This work was supported by the Wellcome Trust (WT107881, WT219506) and the Medical Research Council (MC_UU_00002/4, MC_UU_00040/01).

The EPIC-Norfolk study (DOI 10.22025/2019.10.105.00004) has received funding from the Medical Research Council (MR/N003284/1, MC-UU_12015/1, and MC_UU_00006/1) and Cancer Research UK (C864/A14136). The genetics work in the EPIC-Norfolk study was funded by the Medical Research Council (MC_PC_13048). We also acknowledge funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement 633964 (ImmunoAgeing).

For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

Tables

Isotype	rsID	Chromosome	Position	MAF	Gene(s)	Nearest gene	Distance to nearest gene	Missense gene	QTL genes	Beta	Standard error	p-value
IgA	rs7544568:G>A	1	38,173,536	0.27	POU3F1	POU3F1	126742		POU3F1	-0.04	0.01	2.28E-10
IgA	rs11204701:A>T	1	150,689,703	0.23	ENSA,ARNT,MCL1	GOLPH3L	0		ENSA,ARNT,MCL1	0.03	0.01	1.11E-08
IgA	rs3766151:A>G	1	169,886,435	0.32	FIRRM,KIFAP3,SCYL3	SCYL3	0		FIRRM,KIFAP3,SCYL3	0.03	0.01	1.74E-11
IgA	rs13388394:A>G	2	33,539,109	0.06	RASGRP3	FAM98A	0	RASGRP3		-0.06	0.01	2.80E-08
IgA	rs867529:G>C	2	88,613,755	0.27	EIF2AK3,EIF2AK3-DT,RPIA	EIF2AK3	0	EIF2AK3	EIF2AK3-DT,RPIA,EIF2AK3	0.05	0.01	9.71E-19
IgA	rs62193147:T>C	2	230,251,645	0.11	SP140,SP140L	SP140	0		SP140,SP140L	0.04	0.01	2.72E-08
IgA	rs709167:G>T	3	12,461,456	0.45	TSEN2,PPARG,MKRN2	TSEN2	22964		TSEN2,PPARG,MKRN2	0.03	0.01	4.70E-09
IgA	rs2293607:T>C	3	169,764,547	0.23	MYNN,ACTRT3,ENSG00000269984,LRRC34	ACTRT3	2373		MYNN,ACTRT3,ENSG00000269984,LRRC34	0.04	0.01	1.27E-12

IgA	rs55901664:A>T	3	177,042,914	0.32	TBL1XR1-AS1,TBL1XR1	TBL1XR1	0		TBL1XR1-AS1,TBL1XR1	0.03	0.01	4.83E-08
IgA	rs17505102:G>C	3	189,683,987	0.15	TP63	TP63	0			0.05	0.01	3.76E-11
IgA	rs2101521:G>A	4	38,809,930	0.25	RPL9,FAM114A1,TLR10,TLR1	TLR1	0		RPL9,FAM114A1,TLR10,TLR1	-0.03	0.01	1.12E-08
IgA	rs1585213:C>T	4	102,523,541	0.41	UBE2D3,ENSG00000260651,MANBA,NFKB1	NFKB1	0		UBE2D3,ENSG00000260651,MANBA,NFKB1	-0.03	0.00	2.16E-11
IgA	rs6890853:G>A	5	35,852,209	0.30	IL7R	IL7R	485		IL7R	-0.04	0.01	2.48E-11
IgA	rs10440635:G>A	5	40,490,688	0.40	PTGER4	TTC33	21644		PTGER4	0.03	0.00	4.68E-12
IgA	rs6570530:T>C	6	142,829,086	0.45	HIVEP2	HIVEP2	0			-0.03	0.00	2.08E-09
IgA	rs79715837:C>T	7	17,685,515	0.09	SNX13	SNX13	105245			-0.05	0.01	1.76E-08
IgA	rs1948915:T>C	8	127,210,176	0.33	POU5F1B	POU5F1B	112006			0.03	0.01	4.61E-10
IgA	rs4744011:C>T	9	90,792,925	0.44	SYK	SYK	8861		SYK	0.03	0.01	2.01E-08
IgA	rs113495277:A>G	9	120,913,617	0.03	TRAF1	TRAF1	0	TRAF1		-0.09	0.02	3.79E-08
IgA	rs1058935:G>C	11	308,363	0.48	IFITM3,PGGHG,IFITM2,EPS8L2,IFITM1,IFITM3-AS1,PTDSS2,ENSG00000288681	IFITM2	0		IFITM3,PGGHG,IFITM2,EPS8L2,IFITM1,IFITM3-AS1,PTDSS2,ENSG00000288681	-0.04	0.01	3.34E-10
IgA	rs1681630:T>C	11	47,947,600	0.36	F2,PTPRJ	PTPRJ	32824		F2,PTPRJ	0.03	0.01	1.25E-10
IgA	rs2714073:G>A	11	123,505,241	0.43	GRAMD1B	GRAMD1B	0			-0.03	0.00	2.46E-09
IgA	rs1008866:G>A	12	12,022,938	0.26	BCL2L14	BCL2L14	26905			-0.03	0.01	1.03E-08
IgA	rs11147102:G>A	12	132,469,613	0.30	ENSG00000256875,FBRSL1	FBRSL1	19937		ENSG00000256875,FBRSL1	-0.03	0.01	3.94E-09
IgA	rs539846:G>T	15	40,105,735	0.47	BMF	BMF	0		BMF	-0.03	0.01	1.72E-09
IgA	rs148076268:A>T	15	44,137,670	0.03	PATL2,MAPDA	FRMD5	0		PATL2,MAPDA	-0.09	0.02	4.23E-09
IgA	rs12149527:C>T	16	79,076,699	0.49	WWOX	WWOX	0			-0.03	0.01	5.80E-11
IgA	rs2872516:C>T	17	39,916,474	0.47	ZPBP2,GSDMA,GSDMB,IKZF3,ORMDL3	GSDMB	0		ZPBP2,GSDMA,GSDMB,IKZF3,ORMDL3	0.03	0.00	1.12E-11
IgA	rs11655056:C>T	17	49,286,745	0.43	ZNF652,PPP1R9B	ZNF652	2460		ZNF652,PPP1R9B	0.04	0.01	3.34E-10
IgA	rs2286564:A>G	17	63,932,631	0.40	FTSJ3,PSMC5,CD79B	CD79B	294		FTSJ3,PSMC5,CD79B	0.03	0.00	1.01E-11
IgA	rs11653912:T>C	17	75,324,509	0.24	GRB2	GRB2	0			-0.04	0.01	2.61E-08
IgA	rs3743:G>A	17	79,935,827	0.35	CBX2,TBC1D16,ENSG00000275516,SLC26A11,CBX4	TBC1D16	0		CBX2,TBC1D16,ENSG00000275516,SLC26A11,CBX4	0.03	0.01	5.32E-10
IgA	rs7245642:G>A	19	4,053,015	0.29	ZBTB7A	ZBTB7A	0		ZBTB7A	0.04	0.01	7.92E-10
IgA	rs754292:C>T	19	16,589,002	0.47	ENSG00000268790		0			-0.03	0.00	3.90E-08
IgA	rs8111826:G>A	19	45,053,960	0.47	ZNF234,CLASRP,ENSG00000267114	CLASRP	0		ZNF234,CLASRP,ENSG00000267114	0.03	0.00	4.33E-08
IgA	rs3810277:G>T	19	50,431,721	0.13	SPIB	SPIB	406			0.04	0.01	1.47E-08
IgA	rs2256609:A>G	22	21,570,728	0.20	PI4KA,CCDC116,PPIL2,PPM1F,UBE2L3,YDJC,ENSG00000273342	UBE2L3	0		PI4KA,CCDC116,PPIL2,PPM1F,UBE2L3,YDJC,ENSG00000273342	0.04	0.01	3.78E-12
IgA	rs3177243:G>C	22	23,837,735	0.16	DERL3,MMP11,SMARCB1	DERL3	0	DERL3	MMP11,SMARCB1,DERL3	0.05	0.01	2.49E-16
IgG	rs9864859:A>C	3	58,286,444	0.25	ENSG00000272182,RPP14,PDHB,ABHD6,FLNB,ENSG00000273493,PXK,HTD2	ABHD6	0		ENSG00000272182,RPP14,PDHB,ABHD6,FLNB,ENSG00000273493,PXK,HTD2	0.04	0.01	4.53E-08
IgG	rs537870777:C>T	5	65,146,194	0.00	ADAMTS6	ADAMTS6	2543			-1.38	0.23	2.83E-09
IgG	rs7449025:A>T	5	176,750,656	0.00	UNC5A	UNC5A	59862			1.75	0.32	3.69E-08
IgG	rs759314091:C>T	6	149,668,400	0.00	LATS1	LATS1	0			-1.94	0.34	1.33E-08
IgG	rs7918084:C>T	10	92,669,710	0.46	HHEX	KIF11	14314		HHEX	0.04	0.01	1.99E-11
IgG	rs184552867:A>G	11	45,133,431	0.00	PRDM11	PRDM11	0			-1.57	0.28	2.34E-08
IgG	rs17676923:C>T	17	39,893,872	0.06	MIEN1,GSDMB,IKZF3	GSDMB	10722		MIEN1,GSDMB,IKZF3	0.08	0.01	1.19E-08

IgG	rs562874417:C>G	20	20,892,169	0.00	RALGAPA2	RALGAPA2	179524			-1.41	0.25	1.75E-08
IgM	rs12026735:G>A	1	30,776,072	0.18	LAPTM5	LAPTM5	18297		LAPTM5	-0.06	0.01	1.54E-12
IgM	rs7536190:T>A	1	150,426,629	0.11	RPRD2	RPRD2	0			-0.06	0.01	6.98E-10
IgM	rs150412347:G>A	2	55,285,800	0.09	RPS27A,CLHC1,CCDC88A	CCDC88A	2041		RPS27A,CLHC1,CCDC88A	-0.06	0.01	7.59E-09
IgM	rs115979345:C>G	2	241,466,693	0.18	FARP2,MTERF4,SEPTIN2,STK25	FARP2	0		FARP2,MTERF4,SEPTIN2,STK25	-0.05	0.01	4.82E-09
IgM	rs28671336:C>G	3	136,795,300	0.33	NCK1,SLC35G2,NCK1-DT	SLC35G2	23346		NCK1,SLC35G2,NCK1-DT	-0.04	0.01	3.27E-11
IgM	rs1915948:T>C	3	152,504,354	0.33	ENSG00000243305,RAP2B,MBNL1,P2RY1,SUCNR1	MBNL1	38573		ENSG00000243305,RAP2B,MBNL1,P2RY1,SUCNR1	0.04	0.01	4.60E-08
IgM	rs1397934:A>G	4	40,185,976	0.31	RHOH	RHOH	5034			-0.05	0.01	3.08E-15
IgM	rs67960624:G>A	4	67,613,670	0.26	GNRHR,STAP1,YTHDC1,UBA6-DT,UBA6	UBA6	0		GNRHR,STAP1,YTHDC1,UBA6-DT,UBA6	-0.04	0.01	3.19E-09
IgM	rs7758383:A>G	6	142,848,586	0.49	HIVEP2	HIVEP2	0			-0.04	0.01	2.53E-10
IgM	rs12531349:A>T	7	3,037,424	0.30	CARD11	CARD11	0			-0.04	0.01	1.95E-10
IgM	rs73074879:G>A	7	23,551,590	0.09	TRA2A	TRA2A	19548			-0.06	0.01	1.23E-08
IgM	rs3823958:A>C	7	107,614,773	0.46	DUS4L,HBP1	BCAP29	0		DUS4L,HBP1	0.04	0.01	3.83E-11
IgM	rs2667978:A>G	8	55,985,392	0.17	LYN,RPS20	LYN	0		LYN,RPS20	-0.06	0.01	2.25E-12
IgM	rs12554596:A>G	9	37,002,145	0.19	ENSG00000250850,PAX5	PAX5	0		ENSG00000250850,PAX5	-0.04	0.01	4.12E-08
IgM	rs290243:G>A	9	90,809,055	0.21	SYK	SYK	0		SYK	-0.05	0.01	4.98E-10
IgM	rs7918238:G>A	10	3,876,889	0.46	LINC00705	KLF6	91607		LINC00705	-0.04	0.01	7.38E-11
IgM	rs7902629:G>C	10	110,423,817	0.49	DUSP5	DUSP5	74089			0.04	0.01	3.55E-11
IgM	rs36027301:C>T	11	68,041,801	0.06	TCIRG1	TCIRG1	0	TCIRG1	TCIRG1	-0.09	0.01	1.41E-11
IgM	rs6589225:T>C	11	111,376,136	0.24	POU2AF1,POU2AF3,COLCA1	POU2AF1	0		POU2AF1,POU2AF3,COLCA1	-0.04	0.01	5.56E-09
IgM	rs11215032:T>G	11	114,519,937	0.11	NXPE1,CADM1	NXPE1	0		NXPE1,CADM1	-0.06	0.01	1.23E-09
IgM	rs905688879:T>G	12	71,786,078	0.00	RAB21	RAB21	0			1.74	0.32	3.48E-08
IgM	rs56356298:A>T	13	86,109,332	0.00	SLITRK6	SLITRK6	302648			1.31	0.24	4.94E-08
IgM	rs7158350:T>C	14	75,480,111	0.23	YLPM1	JDP2	5999		YLPM1	-0.05	0.01	2.22E-10
IgM	rs11627032:T>C	14	92,637,727	0.26	RIN3	RIN3	0		RIN3	0.04	0.01	4.56E-10
IgM	rs146125856:T>C	15	50,492,793	0.10	USP8	USP8	0	USP8		0.11	0.02	4.66E-09
IgM	rs6495122:A>C	15	74,833,304	0.42	ULK3,CSK,FAM219B,MPI,SCAMP5,COX5A,RPP25	CPLX3	1501		ULK3,CSK,FAM219B,MPI,SCAMP5,COX5A,RPP25	-0.04	0.01	1.39E-09
IgM	rs34033173:G>A	16	10,889,076	0.12	CIITA	CIITA	0		CIITA	0.06	0.01	4.93E-11
IgM	rs7194131:C>A	16	81,831,630	0.11	PLCG2	PLCG2	0			-0.06	0.01	1.42E-09
IgM	rs305080:C>T	16	85,908,168	0.35	IRF8	IRF8	0		IRF8	0.04	0.01	1.38E-09
IgM	rs9907708:A>T	17	16,989,369	0.37	TNFRSF13B	TNFRSF13B	17250			-0.05	0.01	7.72E-13
IgM	rs7236090:T>C	18	63,279,878	0.47	BCL2	BCL2	0			-0.04	0.01	3.68E-09
IgM	rs10419696:T>G	19	16,309,861	0.31	OCEL1	KLF2	14964		OCEL1	-0.04	0.01	1.35E-10
IgM	rs11666267:G>A	19	17,549,491	0.45	COLGALT1,NIBAN3	NIBAN3	0	NIBAN3	COLGALT1,NIBAN3	0.04	0.01	1.90E-11
IgM	rs10412811:G>A	19	49,381,238	0.25	ENSG00000197813,ENSG00000268686,KASH5,CD37,TEAD2,SLC6A16,DKKL1	DKKL1	6121		ENSG00000197813,ENSG00000268686,KASH5,CD37,TEAD2,SLC6A16,DKKL1	-0.04	0.01	3.44E-08
IgM	rs192881351:C>T	21	29,697,533	0.00	GRIK1	GRIK1	0			1.65	0.29	7.97E-09
IgM	rs541130164:C>T	22	50,704,843	0.00	SHANK3	SHANK3	0	SHANK3		1.60	0.29	3.78E-08

Table 1. Lead SNPs from novel genome-wide significant associations in the IgA, IgG, and IgM meta-analyses. The ‘Variant' column gives the rsID of each SNP, and the reference and effect alleles separated by ‘>’. ‘MAF’ is the minor allele frequency computed from data in the component GWAS where this information was provided in the original studies. 'Gene(s)' gives the gene(s) with the most evidence linking it/them to the association signal; the procedure by which we assigned genes to variants, and identified quantitative trait loci (QTL) and missense variants is described in Methods. ‘Missense gene’ reports the gene (if any) in which the lead SNP was identified to be a missense variant.

First isotype	Second isotype	Heritability model	MHC	Genetic correlation estimate	Standard error	p-value
IgA	IgM	LDAK-Thin	FALSE	-0.02	0.04	0.62387
IgA	IgM	LDAK-Thin	TRUE	0.00	0.04	0.94618
IgA	IgG	LDAK-Thin	FALSE	0.19	0.05	0.00009
IgA	IgG	LDAK-Thin	TRUE	0.23	0.06	0.00009
IgG	IgM	LDAK-Thin	FALSE	0.14	0.05	0.00286
IgG	IgM	LDAK-Thin	TRUE	0.18	0.07	0.00857

Table 2. Estimates of the genetic correlation between serum antibody isotype levels. The ‘MHC’ column indicates whether variants lying within the major histocompatibility complex were included in the dataset for genetic correlation estimation. p-values were calculated from a chi-squared test of non-zero genetic correlation.

Supplementary Text

Associations in the *IGH* and *IGK* loci

We observed strong associations between serum antibody levels and variants in the IGH and IGK loci in one or more of the GWAS included in each isotype meta-analysis (Supplementary Table 4). The IGH, and IGK and IGL loci encode the heavy and light chain molecules which comprise immunoglobulin tetramers. We hypothesised that coding variation in these loci could modify the immunoglobulin epitopes to which bind the capture antibodies and aptamers used in affinity-based quantification, and hence induce spurious genotype-dependent variation in antibody measurement. The immunoglobulin allotypes defined by certain coding variants were originally characterised in terms of their differing reactivity to immunologic reagents (58) and recent work has shown IgG1 allotypes do indeed affect detection by commercially available

anti-IgG1 antibodies (59)

Coding variants featured in strong association signals at the IGH and IGK loci in the GWAS we examined: two signals in IGH and IGK had as their lead SNPs the missense variants rs117775520 and rs232230, respectively, whilst another *IGH* missense variant, rs1045853, occurred as part of association signals in the EPIC and Pietzner IgG GWAS. rs1045853 is located in *IGHG1*, encoding the heavy constant chain of IgG1, G1m, and its minor allele's substitution of glutamate for aspartate defines in part the nG1m1 allotype of IgG1 (60). To exclude the credible risk that coding variation at these loci could contribute to antibody measurement error in a genotype-dependent fashion and hence further error in downstream analysis, we removed the three immunoglobulin chain loci as described in Methods.

Supplementary Tables

Supplementary Table 1. The serum antibody GWAS subject to meta-analysis in this work.

Supplementary Table 2. The non-antibody GWAS used in this work. Where studies are not available on the EBI GWAS Catalog, we give publicly accessible URLs where available.

Supplementary Table 3. The transformation applied to each dataset included in our meta-analysis in its original publication.

Supplementary Table 4. Variant associations with serum antibody falling in the *IGH* and *IGK* loci. The 'Variant' column gives the rsID of each SNP, and the reference and effect alleles separated by '>'.

Supplementary Table 5. Lead SNPs from genome-wide significant associations in the serum IgG meta-analysis. The 'Variant' column gives the rsID of each SNP, and the reference and effect alleles separated by '>'. 'MAF' is the minor allele frequency computed from data in the component GWAS where this information was provided in the original studies. 'Gene(s)' gives the gene(s) with the most evidence linking it/them to the association signal; the procedure by which we assigned genes to variants, and identified quantitative trait loci (QTL) and missense variants is described in Methods. 'Missense gene' reports the gene (if any) in which the lead SNP was identified to be a missense variant. 'Novel' indicates whether an association with serum IgG has previously been reported for a variant.

Supplementary Table 6. Lead SNPs from genome-wide significant associations in the serum IgM meta-analysis. The ‘Variant' column gives the rsID of each SNP, and the reference and effect alleles separated by ‘>’. ‘MAF’ is the minor allele frequency computed from data in the component GWAS where this information was provided in the original studies. 'Gene(s)' gives the gene(s) with the most evidence linking it/them to the association signal; the procedure by which we assigned genes to variants, and identified quantitative trait loci (QTL) and missense variants is described in Methods. ‘Missense gene’ reports the gene (if any) in which the lead SNP was identified to be a missense variant. 'Novel' indicates whether an association with serum IgM has previously been reported for a variant.

Supplementary Table 7. Lead SNPs from genome-wide significant associations in the serum IgA meta-analysis. The ‘Variant' column gives the rsID of each SNP, and the reference and effect alleles separated by ‘>’. ‘MAF’ is the minor allele frequency computed from data in the component GWAS where this information was provided in the original studies. 'Gene(s)' gives the gene(s) with the most evidence linking it/them to the association signal; the procedure by which we assigned genes to variants, and identified quantitative trait loci (QTL) and missense variants is described in Methods. ‘Missense gene’ reports the gene (if any) in which the lead SNP was identified to be a missense variant. 'Novel' indicates whether an association with serum IgA has previously been reported for a variant.

Supplementary Table 8. Heritability estimates for serum antibody phenotypes. The ‘MHC’ column indicates whether variants lying within the major histocompatibility complex were included in the dataset for heritability estimation.

Supplementary Table 9. Cross-isotype phenotypic correlations in the EPIC cohort. ‘Scale’ indicates whether untransformed (‘raw’) or log-transformed antibody concentrations were used to compute correlations. ‘95% CI’ is 95% confidence interval.

Supplementary Table 10. The results of the antibody-antibody colocalisation analyses. ‘No. of SNPs’ is the number of SNPs included in the analysis. Columns ‘PP.H0.abf’ to ‘PP.H4.abf’ give the posterior probability of the *coloc* hypotheses H_0 to H_4. ‘Filtered’ indicates whether the analysis was performed on data subject to the filtering process described in the colocalisation section of Methods. ‘Min. locus p-value’ is the minimum p-value across all SNPs considered in the colocalisation analysis. The ‘Pearson correlation’ column gives the Pearson correlation of Z-scores between the antibody phenotypes. The ‘Effect ratio’ columns give the ratio of the first and second isotype’s effect estimates at the lead SNP of the former and latter, respectively; where an effect ratio is absent, this is due to the absence of one phenotype’s lead SNP in the other phenotype’s dataset. ‘r2’ gives the squared correlation coefficient of genotypes between the lead SNPs of the two isotypes.

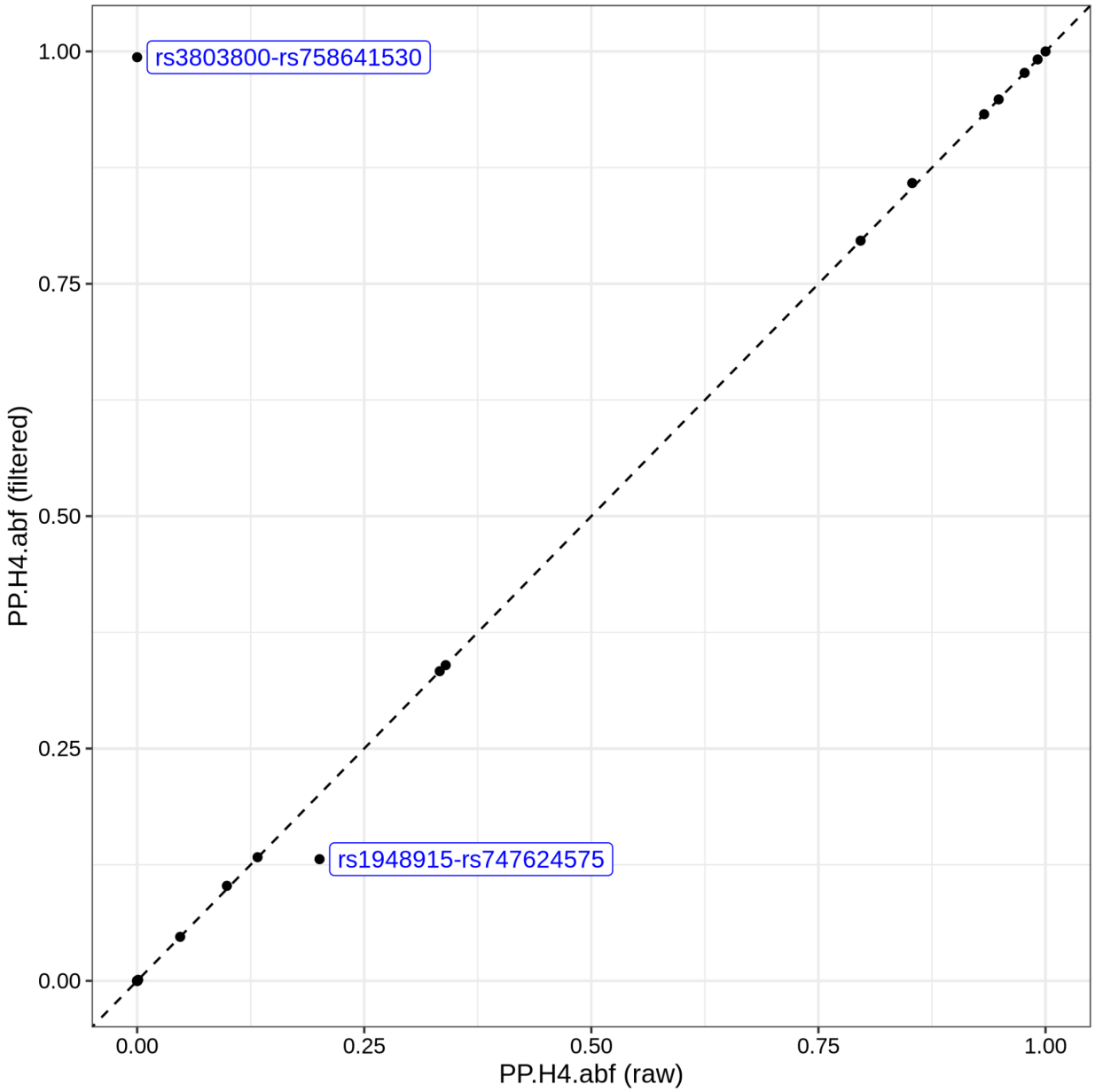
Supplementary Table 11. The results of the antibody-immune trait (‘non-Ig trait’) colocalisation analyses. ‘No. of SNPs’ is the number of SNPs included in the analysis. Columns ‘PP.H0.abf’ to ‘PP.H4.abf’ give the posterior probability of the *coloc* hypotheses H_0 to H_4. ‘Min. locus p-value’ is the minimum p-value across all SNPs considered in the colocalisation analysis. The ‘Pearson correlation’ column gives the Pearson correlation of Z-scores between the antibody and non-antibody phenotypes. The ‘Effect ratio’ columns give the ratio of the antibody and immune

trait effect estimates at the lead SNP of the former and latter, respectively; where an effect ratio is absent, this is due to the absence of one phenotype's lead SNP in the other phenotype's dataset. 'r2' gives the squared correlation coefficient of genotypes between the lead SNPs of the antibody and immune phenotypes.

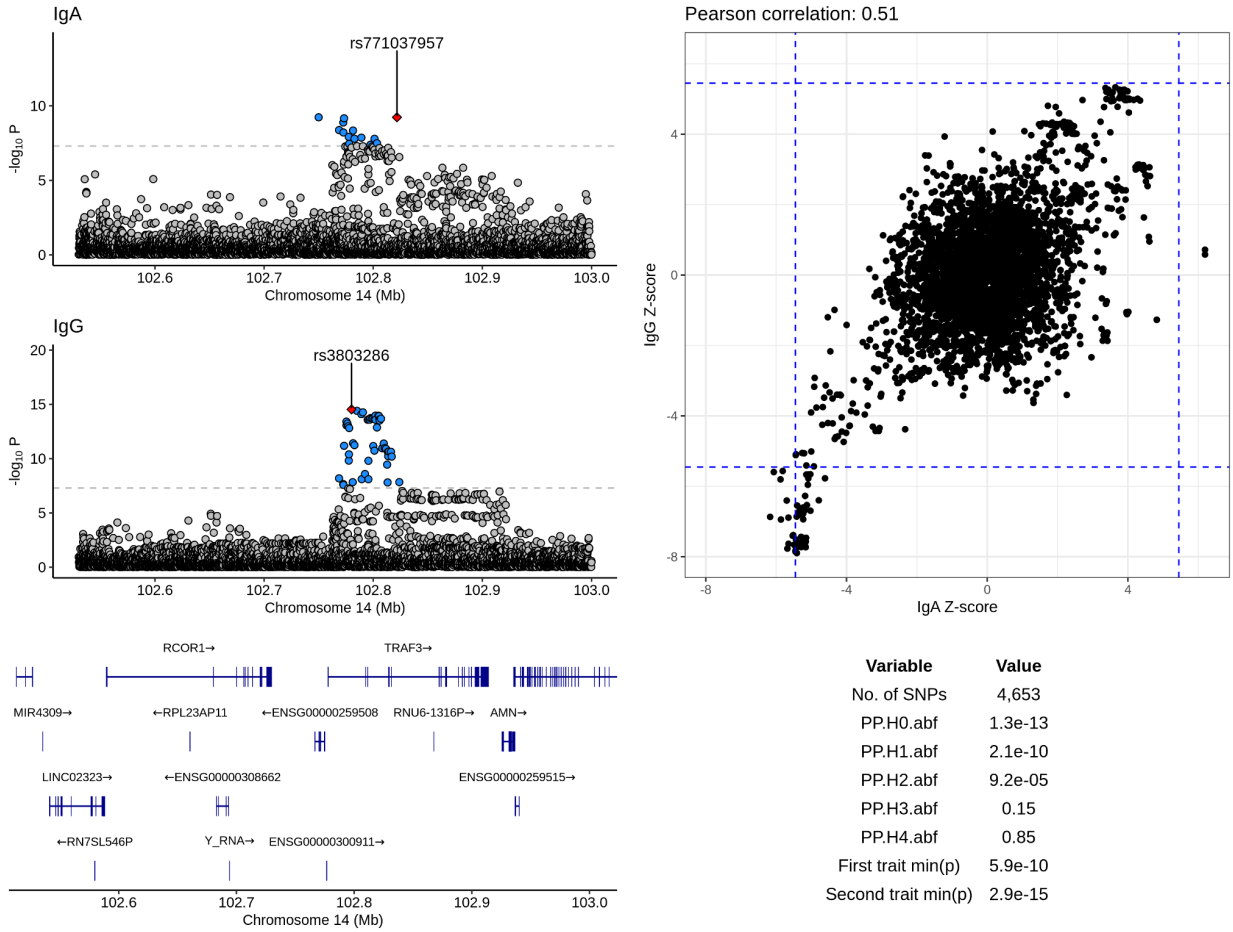
Supplementary Table 12. Estimates of the genetic correlation between antibody and immune phenotypes. p-values were calculated from a chi-squared test of non-zero genetic correlation. The 'FDR' column gives the minimum false discovery rate at which the null of no genetic correlation would be rejected by the Benjamini-Hochberg procedure.

Supplementary Table 13. Inborn errors of immunity (IEI) with causal variants occurring in genes within 200kb of the lead SNPs for antibody associations identified in this work.The 'Variant' column gives the rsID of each antibody lead SNP, and the reference and effect alleles separated by '>'. 'Novel' indicates whether an association with serum antibody has previously been reported for a variant. 'Distance to gene' gives the distance in base pairs from the antibody lead SNP to the IEI gene.

Supplementary Figures

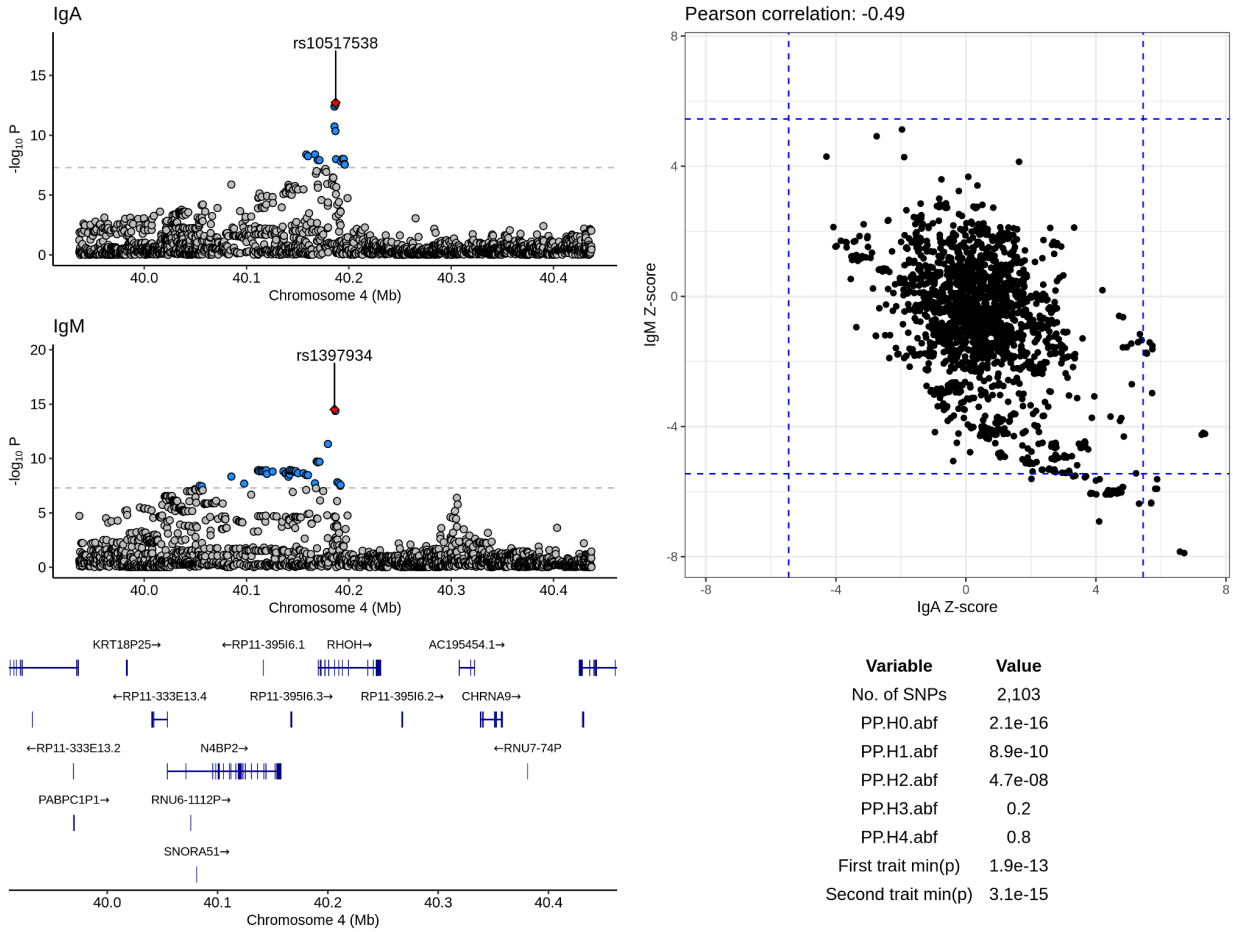


Supplementary Figure 1. The posterior probability of the 'H4' hypothesis of a shared causal variant ('PP.H4.abf') in *coloc* in antibody-antibody colocalisation analyses with ('filtered') and without ('raw') filtering of SNPs by sample fraction. The dashed line indicates y = x. The two labelled points are those lying an appreciable distance from the line y = x.



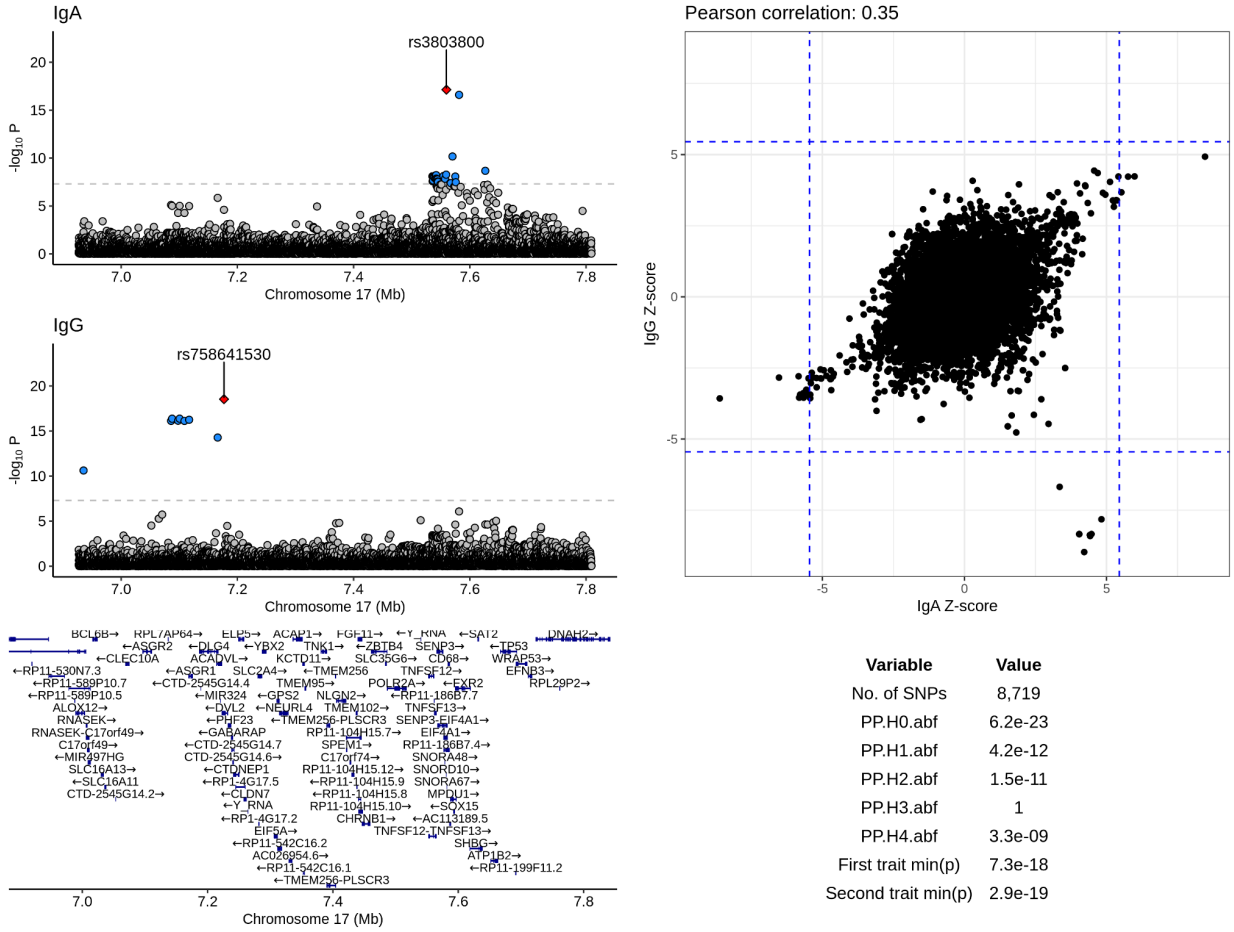
Variable	Value
No. of SNPs	4,653
PP.H0.abf	1.3e-13
PP.H1.abf	2.1e-10
PP.H2.abf	9.2e-05
PP.H3.abf	0.15
PP.H4.abf	0.85
First trait min(p)	5.9e-10
Second trait min(p)	2.9e-15

Supplementary Figure 2. Colocalisation analysis of the IgA and IgG association signals on chromosome 14 in the interval 102.5Mb to 103Mb. The left half of the Figure depicts meta-analytic p-values for the variants in the interval for each phenotype; lead SNPs for each phenotype's association are coloured red and labelled. SNPs with p-values below $5x10^{-8}$ are coloured blue. The panel below the Manhattan plots depicts the genes located within the same genomic interval. The top panel on the right half of the Figure depicts SNP Z-scores for each phenotype. The dashed blue lines give the threshold for genome-wide significance on the Z-score scale. The table in the lower half gives statistics relating to the colocalisation analysis. 'min(p)' is the smallest p-value in the genomic interval subject to colocalisation analysis. The 'PP.Hx.abf' statistics give the posterior probability of the colocalisation hypotheses H_0 to H_4.



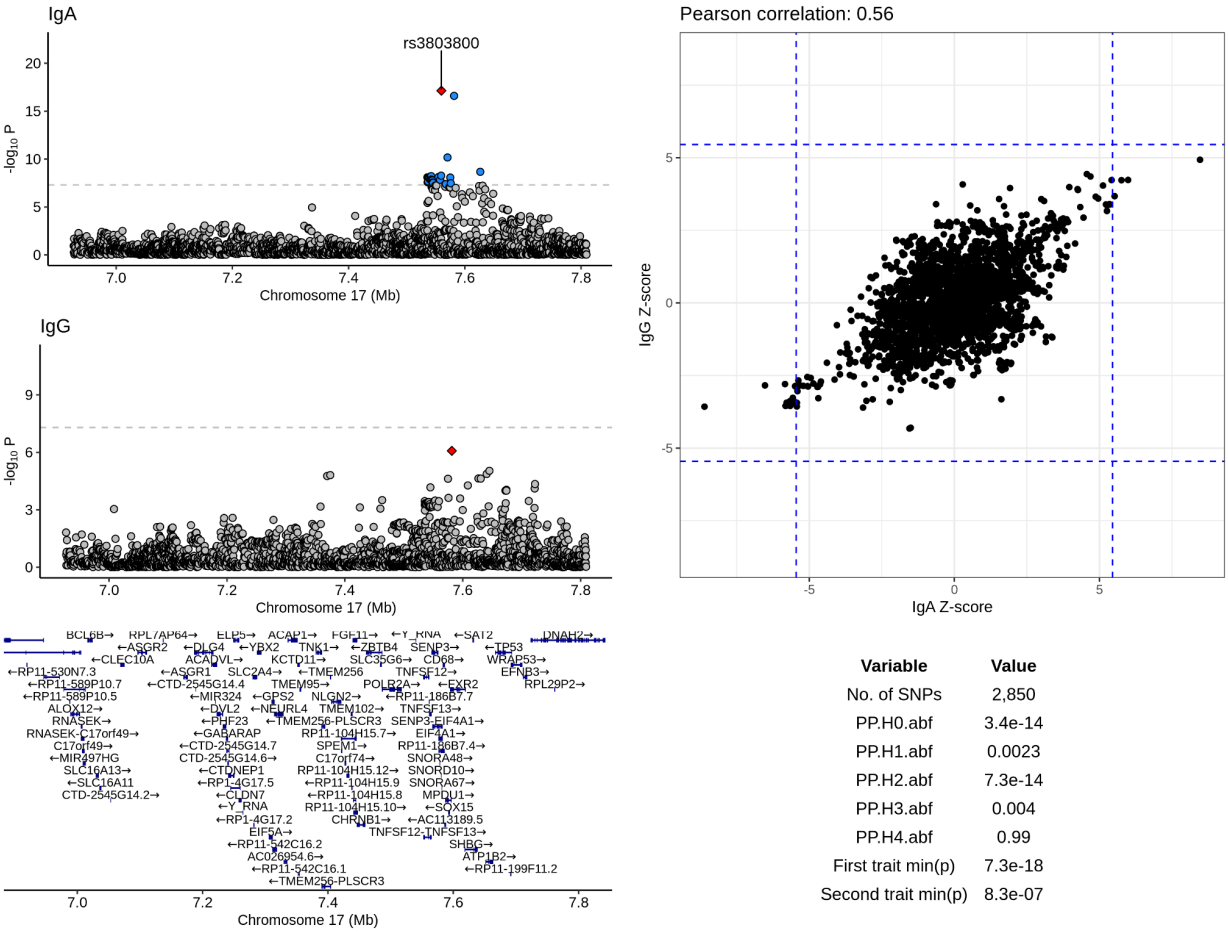
Variable	Value
No. of SNPs	2,103
PP.H0.abf	2.1e-16
PP.H1.abf	8.9e-10
PP.H2.abf	4.7e-08
PP.H3.abf	0.2
PP.H4.abf	0.8
First trait min(p)	1.9e-13
Second trait min(p)	3.1e-15

Supplementary Figure 3. Colocalisation analysis of the IgA and IgM association signals on chromosome 4 in the interval 39.9Mb to 40.5Mb. The left half of the Figure depicts meta-analytic p-values for the variants in the interval for each phenotype; lead SNPs for each phenotype's association are coloured red and labelled. SNPs with p-values below $5x10^{-8}$ are coloured blue. The panel below the Manhattan plots depicts the genes located within the same genomic interval. The top panel on the right half of the Figure depicts SNP Z-scores for each phenotype. The dashed blue lines give the threshold for genome-wide significance on the Z-score scale. The table in the lower half gives statistics relating to the colocalisation analysis. 'min(p)' is the smallest p-value in the genomic interval subject to colocalisation analysis. The 'PP.Hx.abf' statistics give the posterior probability of the colocalisation hypotheses H_0 to H_4.



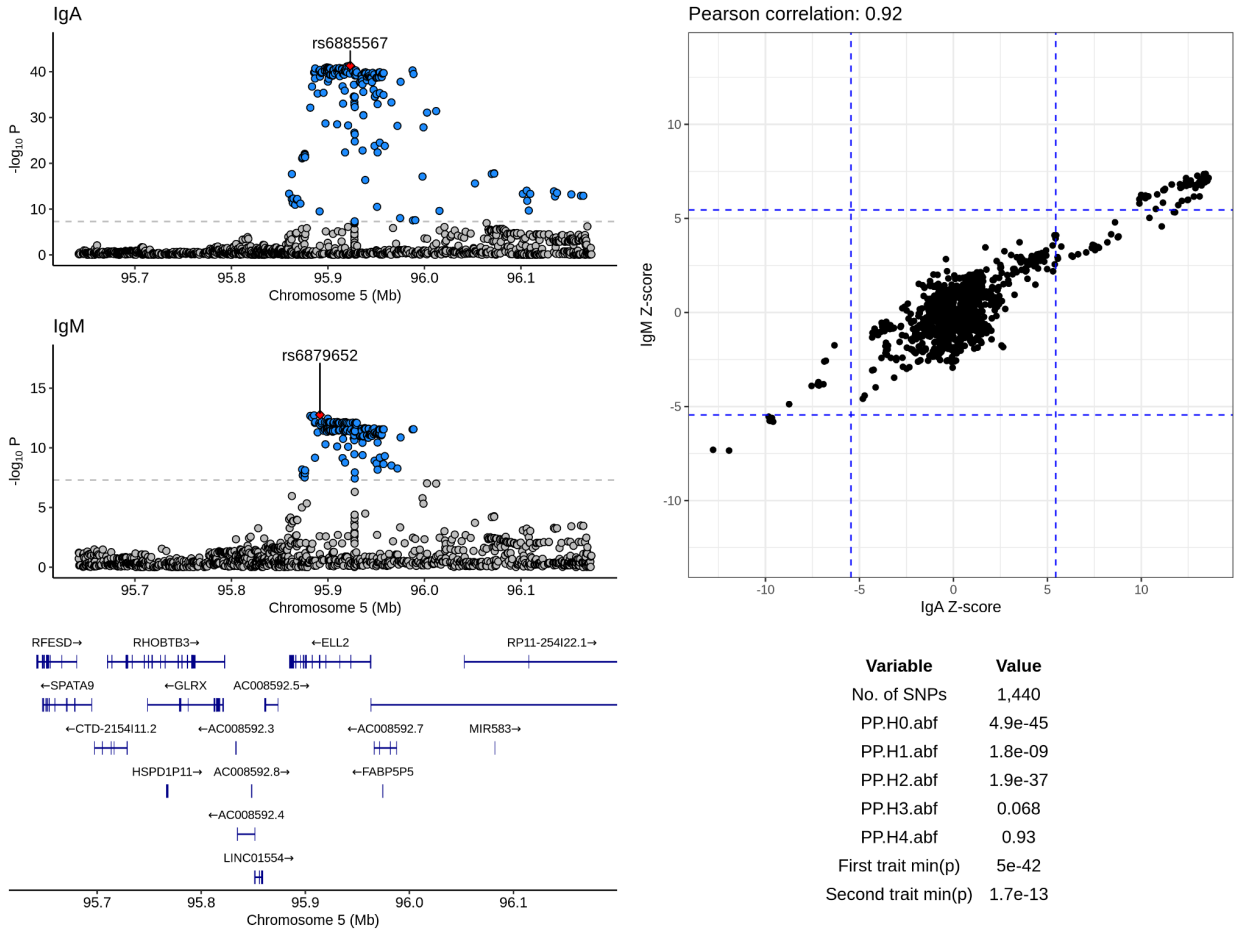
Variable	Value
No. of SNPs	8,719
PP.H0.abf	6.2e-23
PP.H1.abf	4.2e-12
PP.H2.abf	1.5e-11
PP.H3.abf	1
PP.H4.abf	3.3e-09
First trait min(p)	7.3e-18
Second trait min(p)	2.9e-19

Supplementary Figure 4. Colocalisation analysis of the IgA and IgG association signals on chromosome 17 in the interval 6.9Mb to 7.8Mb without filtering. The left half of the Figure depicts meta-analytic p-values for the variants in the interval for each phenotype; lead SNPs for each phenotype's association are coloured red and labelled. SNPs with p-values below $5x10^{-8}$ are coloured blue. The panel below the Manhattan plots depicts the genes located within the same genomic interval. The top panel on the right half of the Figure depicts SNP Z-scores for each phenotype. The dashed blue lines give the threshold for genome-wide significance on the Z-score scale. The table in the lower half gives statistics relating to the colocalisation analysis. 'min(p)' is the smallest p-value in the genomic interval subject to colocalisation analysis. The 'PP.Hx.abf' statistics give the posterior probability of the colocalisation hypotheses H_0 to H_4.



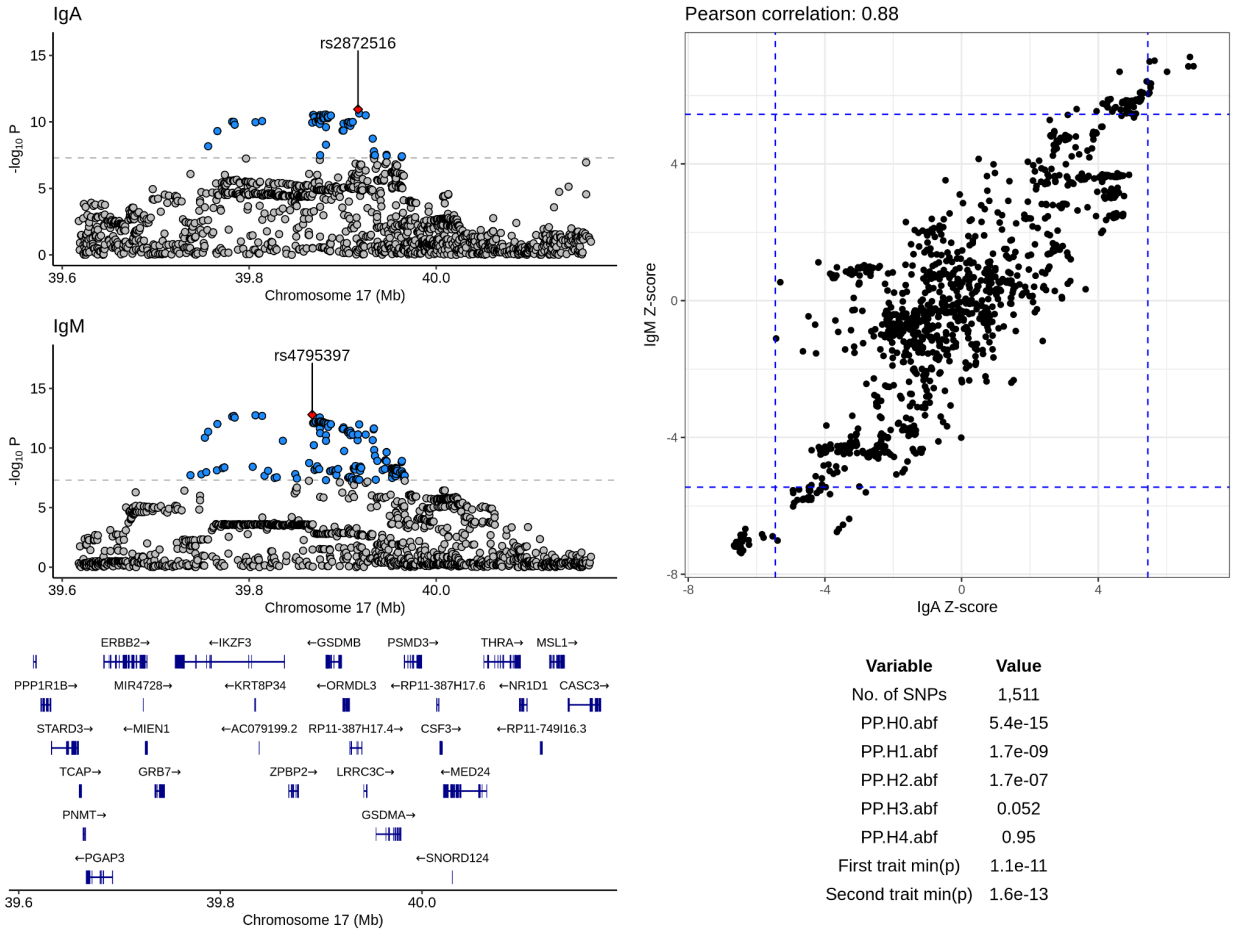
Variable	Value
No. of SNPs	2,850
PP.H0.abf	3.4e-14
PP.H1.abf	0.0023
PP.H2.abf	7.3e-14
PP.H3.abf	0.004
PP.H4.abf	0.99
First trait min(p)	7.3e-18
Second trait min(p)	8.3e-07

Supplementary Figure 5. Colocalisation analysis of the IgA and IgG association signals on chromosome 17 in the interval 6.9Mb to 7.8Mb with filtering. The left half of the Figure depicts meta-analytic p-values for the variants in the interval for each phenotype; lead SNPs for each phenotype's association are coloured red and labelled. SNPs with p-values below $5x10^{-8}$ are coloured blue. The panel below the Manhattan plots depicts the genes located within the same genomic interval. The top panel on the right half of the Figure depicts SNP Z-scores for each phenotype. The dashed blue lines give the threshold for genome-wide significance on the Z-score scale. The table in the lower half gives statistics relating to the colocalisation analysis. 'min(p)' is the smallest p-value in the genomic interval subject to colocalisation analysis. The 'PP.Hx.abf' statistics give the posterior probability of the colocalisation hypotheses H_0 to H_4.



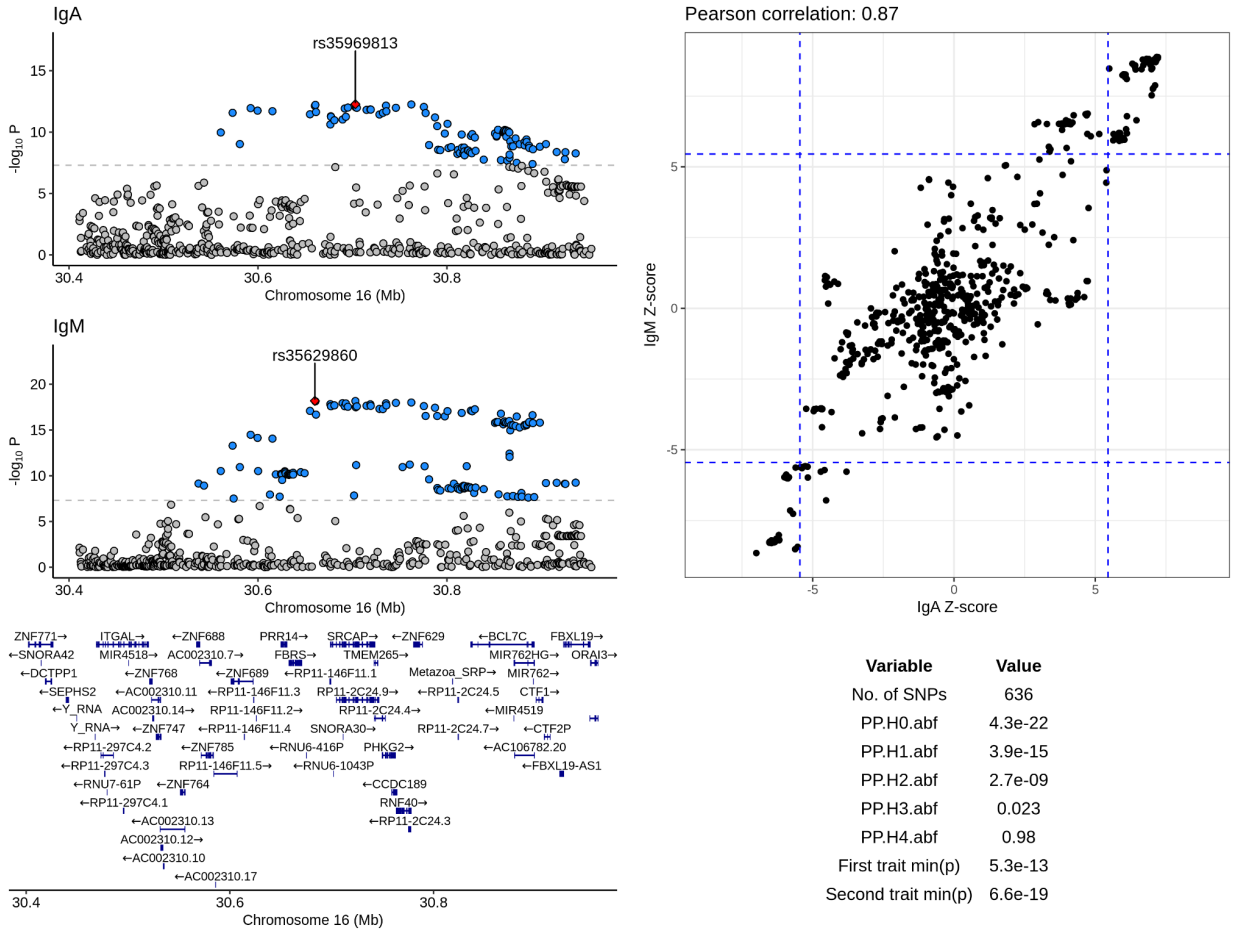
Variable	Value
No. of SNPs	1,440
PP.H0.abf	4.9e-45
PP.H1.abf	1.8e-09
PP.H2.abf	1.9e-37
PP.H3.abf	0.068
PP.H4.abf	0.93
First trait min(p)	5e-42
Second trait min(p)	1.7e-13

Supplementary Figure 6. Colocalisation analysis of the IgA and IgM association signals on chromosome 5 in the interval 95.6Mb to 96.2Mb. The left half of the Figure depicts meta-analytic p-values for the variants in the interval for each phenotype; lead SNPs for each phenotype's association are coloured red and labelled. SNPs with p-values below $5x10^{-8}$ are coloured blue. The panel below the Manhattan plots depicts the genes located within the same genomic interval. The top panel on the right half of the Figure depicts SNP Z-scores for each phenotype. The dashed blue lines give the threshold for genome-wide significance on the Z-score scale. The table in the lower half gives statistics relating to the colocalisation analysis. 'min(p)' is the smallest p-value in the genomic interval subject to colocalisation analysis. The 'PP.Hx.abf' statistics give the posterior probability of the colocalisation hypotheses H_0 to H_4.



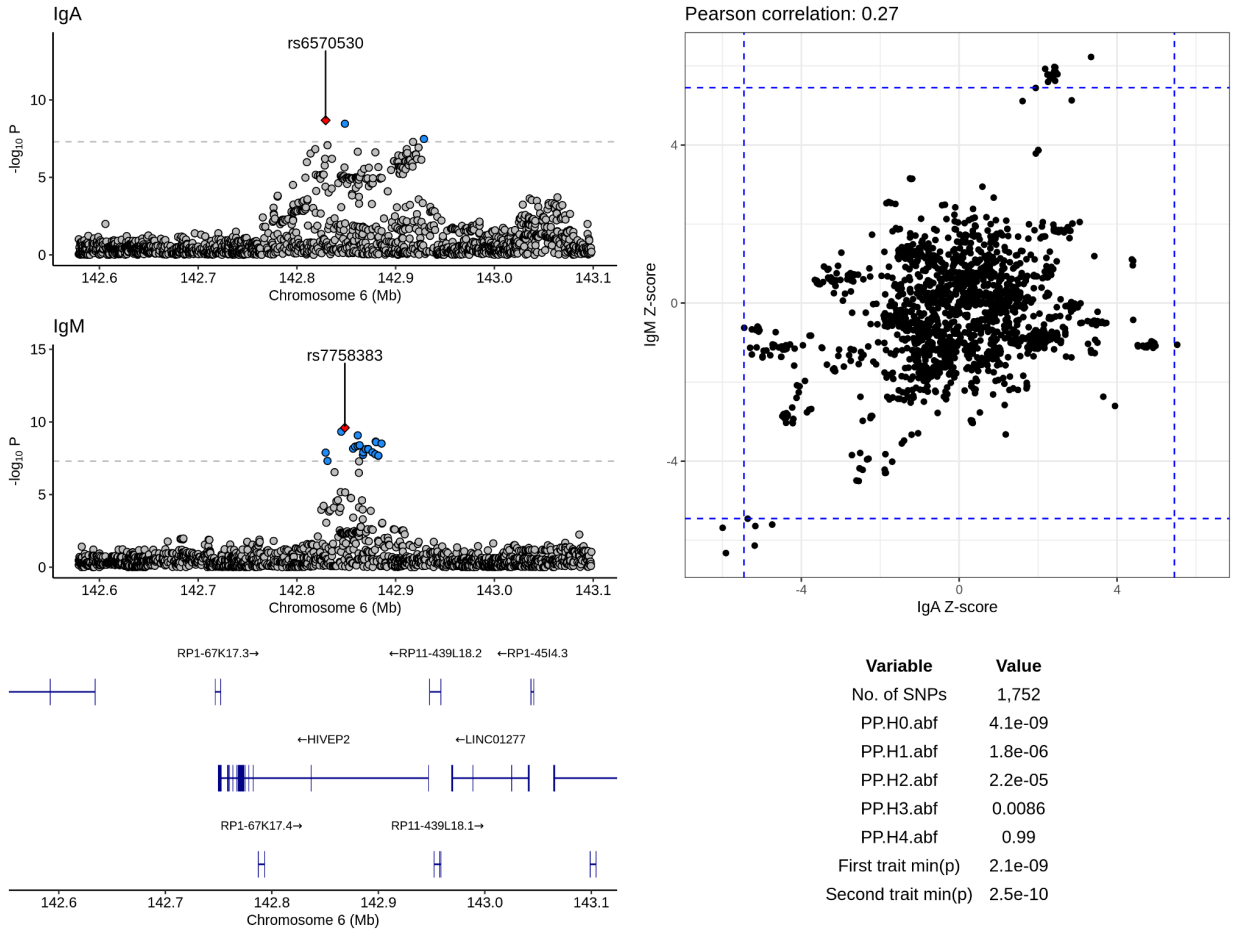
Variable	Value
No. of SNPs	1,511
PP.H0.abf	5.4e-15
PP.H1.abf	1.7e-09
PP.H2.abf	1.7e-07
PP.H3.abf	0.052
PP.H4.abf	0.95
First trait min(p)	1.1e-11
Second trait min(p)	1.6e-13

Supplementary Figure 7. Colocalisation analysis of the IgA and IgM association signals on chromosome 17 in the interval 39.6Mb to 40.2Mb. The left half of the Figure depicts meta-analytic p-values for the variants in the interval for each phenotype; lead SNPs for each phenotype's association are coloured red and labelled. SNPs with p-values below $5x10^{-8}$ are coloured blue. The panel below the Manhattan plots depicts the genes located within the same genomic interval. The top panel on the right half of the Figure depicts SNP Z-scores for each phenotype. The dashed blue lines give the threshold for genome-wide significance on the Z-score scale. The table in the lower half gives statistics relating to the colocalisation analysis. 'min(p)' is the smallest p-value in the genomic interval subject to colocalisation analysis. The 'PP.Hx.abf' statistics give the posterior probability of the colocalisation hypotheses H_0 to H_4.



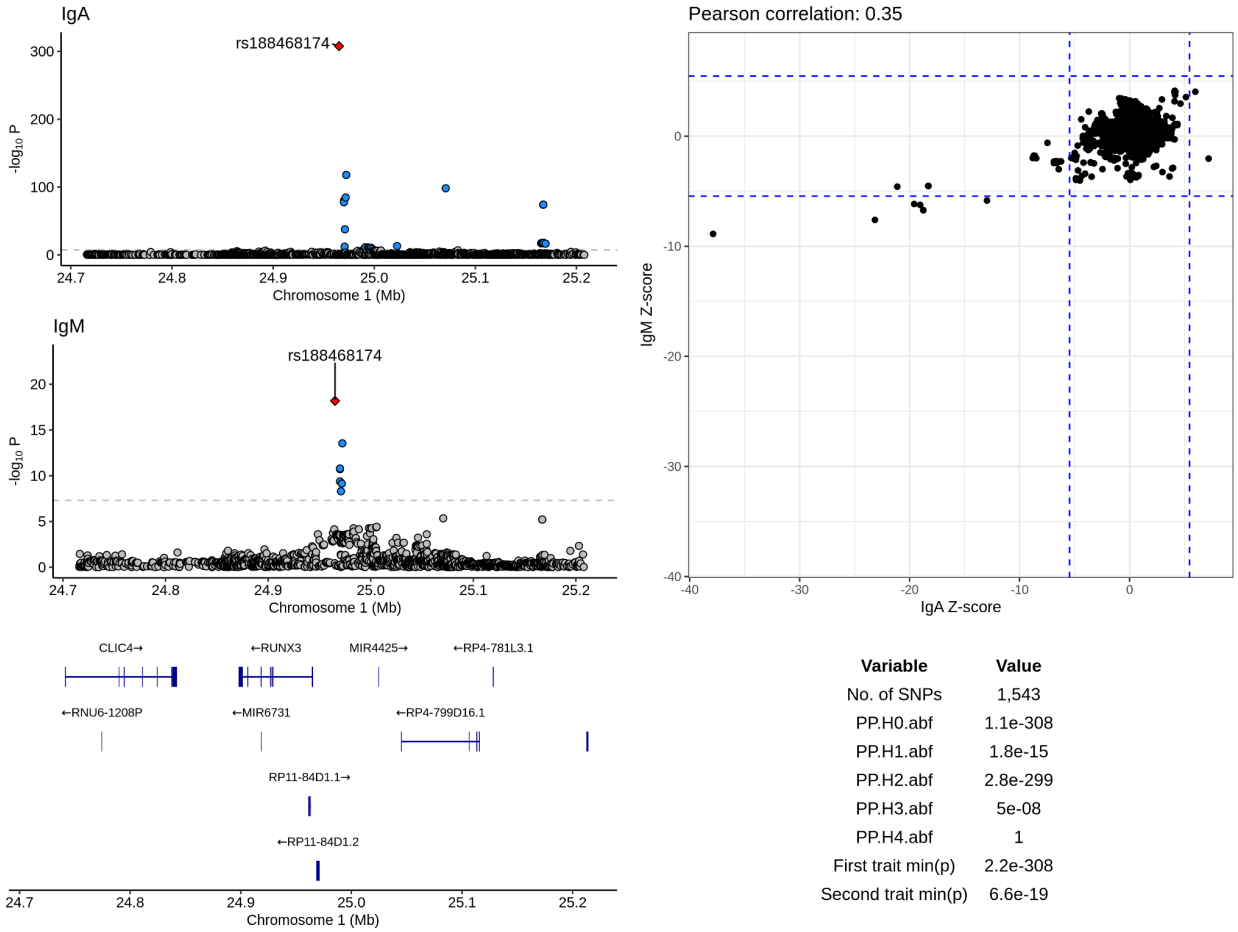
Variable	Value
No. of SNPs	636
PP.H0.abf	4.3e-22
PP.H1.abf	3.9e-15
PP.H2.abf	2.7e-09
PP.H3.abf	0.023
PP.H4.abf	0.98
First trait min(p)	5.3e-13
Second trait min(p)	6.6e-19

Supplementary Figure 8. Colocalisation analysis of the IgA and IgM association signals on chromosome 16 in the interval 30.4Mb to 31Mb. The left half of the Figure depicts meta-analytic p-values for the variants in the interval for each phenotype; lead SNPs for each phenotype's association are coloured red and labelled. SNPs with p-values below $5x10^{-8}$ are coloured blue. The panel below the Manhattan plots depicts the genes located within the same genomic interval. The top panel on the right half of the Figure depicts SNP Z-scores for each phenotype. The dashed blue lines give the threshold for genome-wide significance on the Z-score scale. The table in the lower half gives statistics relating to the colocalisation analysis. 'min(p)' is the smallest p-value in the genomic interval subject to colocalisation analysis. The 'PP.Hx.abf' statistics give the posterior probability of the colocalisation hypotheses H_0 to H_4.



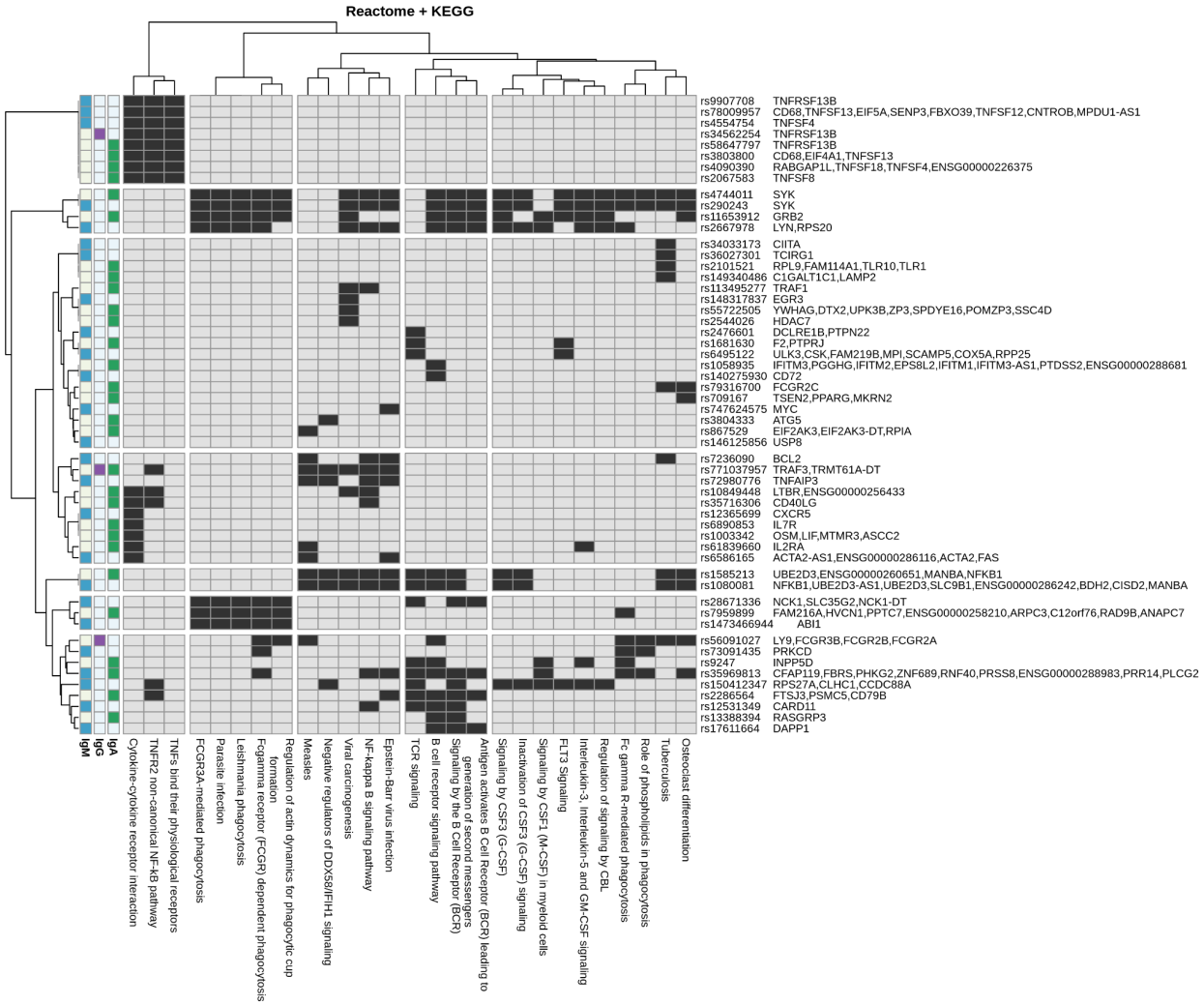
Variable	Value
No. of SNPs	1,752
PP.H0.abf	4.1e-09
PP.H1.abf	1.8e-06
PP.H2.abf	2.2e-05
PP.H3.abf	0.0086
PP.H4.abf	0.99
First trait min(p)	2.1e-09
Second trait min(p)	2.5e-10

Supplementary Figure 9. Colocalisation analysis of the IgA and IgM association signals on chromosome 6 in the interval 142.6Mb to 143.1Mb. The left half of the Figure depicts meta-analytic p-values for the variants in the interval for each phenotype; lead SNPs for each phenotype's association are coloured red and labelled. SNPs with p-values below $5x10^{-8}$ are coloured blue. The panel below the Manhattan plots depicts the genes located within the same genomic interval. The top panel on the right half of the Figure depicts SNP Z-scores for each phenotype. The dashed blue lines give the threshold for genome-wide significance on the Z-score scale. The table in the lower half gives statistics relating to the colocalisation analysis. 'min(p)' is the smallest p-value in the genomic interval subject to colocalisation analysis. The 'PP.Hx.abf' statistics give the posterior probability of the colocalisation hypotheses H_0 to H_4.

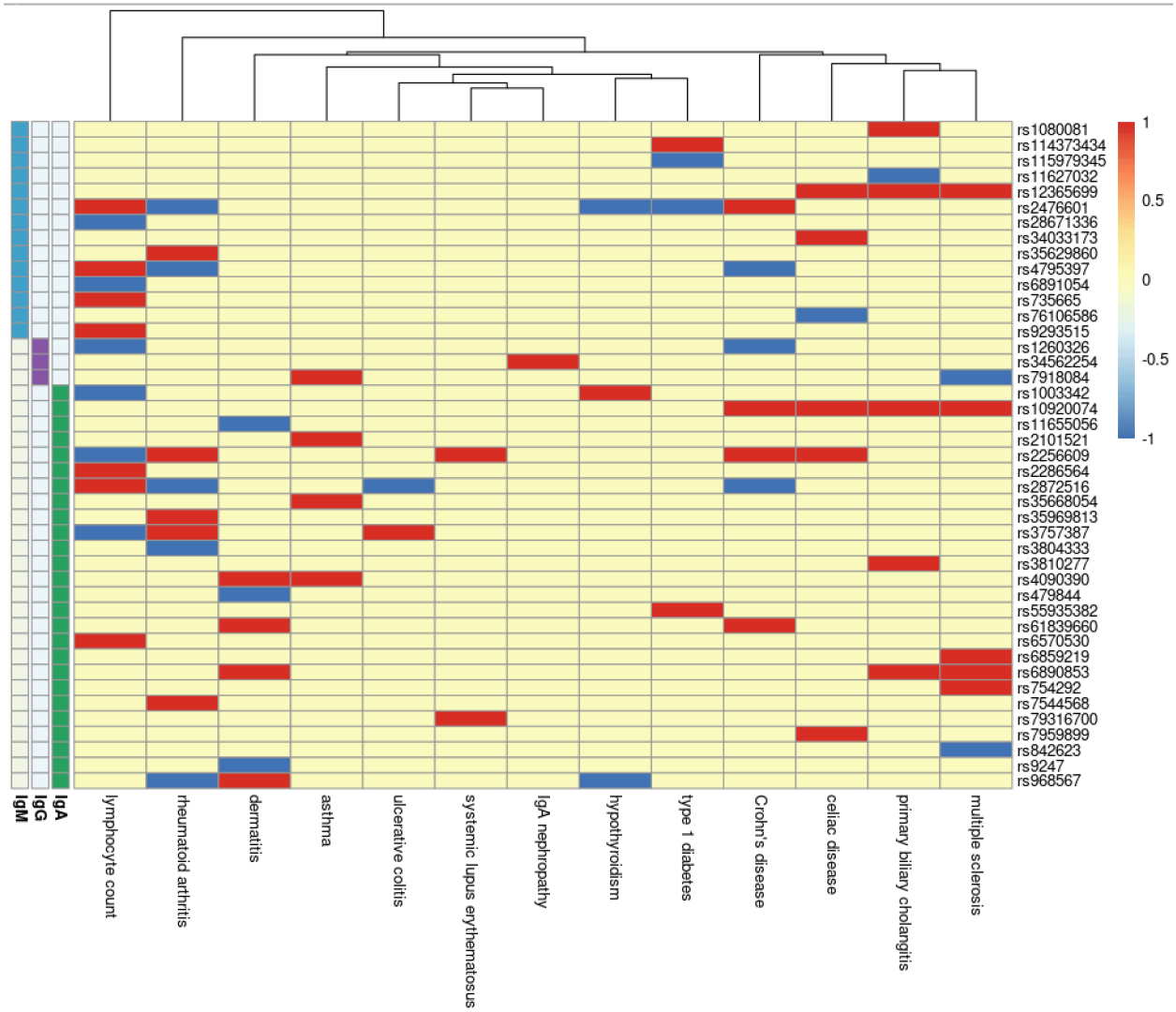


Variable	Value
No. of SNPs	1,543
PP.H0.abf	1.1e-308
PP.H1.abf	1.8e-15
PP.H2.abf	2.8e-299
PP.H3.abf	5e-08
PP.H4.abf	1
First trait min(p)	2.2e-308
Second trait min(p)	6.6e-19

Supplementary Figure 10. Colocalisation analysis of the IgA and IgM association signals on chromosome 1 in the interval 24.7Mb to 25.2Mb. The left half of the Figure depicts meta-analytic p-values for the variants in the interval for each phenotype; lead SNPs for each phenotype's association are coloured red and labelled. SNPs with p-values below $5x10^{-8}$ are coloured blue. The panel below the Manhattan plots depicts the genes located within the same genomic interval. The top panel on the right half of the Figure depicts SNP Z-scores for each phenotype. The dashed blue lines give the threshold for genome-wide significance on the Z-score scale. The table in the lower half gives statistics relating to the colocalisation analysis. 'min(p)' is the smallest p-value in the genomic interval subject to colocalisation analysis. The 'PP.Hx.abf' statistics give the posterior probability of the colocalisation hypotheses H_0 to H_4.



Supplementary Figure 11. Significant KEGG and Reactome pathways are shown together with the lead SNP and associated list of potential causal genes that drove the association. Indicators on the left show which isotype led to each association.



Supplementary Figure 12. Colocalisation of serum antibody and immune-mediated disease association signals. The left-hand annotations give the serum antibody isotype. The rsIDs on the right-hand side indicate the lead SNP in the serum antibody association signal. Cell colour gives the sign of the Pearson correlation of SNP effects for the serum antibody and IMD phenotype pair.

Bibliography

1. Murphy K, Weaver C. Janeway's Immunobiology. W.W. Norton; 2016.

2. Hatagima A, Cabello PH, Krieger H. Causal analysis of the variability of IgA, IgG, and IgM immunoglobulin levels. Hum Biol. 1999;71(2):219–29.

3. Bruton OC. AGAMMAGLOBULINEMIA. Pediatrics. 1952 Jun;9(6):722–8.

4. Pan-Hammarström Q, Hammarström L. Antibody Deficiency Diseases. Eur J Immunol. 2008;38(2):327–33.

5. Yazdani R, Fekrvand S, Shahkarami S, Azizi G, Moazzami B, Abolhassani H, et al. The hyper IgM syndromes: Epidemiology, pathogenesis, clinical manifestations, diagnosis and management. Clin Immunol. 2019 Jan 1;198:19–30.

6. Viktorin A, Frankowiack M, Padyukov L, Chang Z, Melén E, Sääf A, et al. IgA measurements in over 12 000 Swedish twins reveal sex differential heritability and regulatory locus near CD30L. Hum Mol Genet. 2014 Aug 1;23(15):4177–84.

7. Liu L, Khan A, Sanchez-Rodriguez E, Zanoni F, Li Y, Steers N, et al. Genetic Regulation of Serum IgA Levels and Susceptibility to Common Immune, Infectious, Kidney, and Cardio-Metabolic Traits. Nat Commun. 2022 Nov;13(1):6859.

8. Bian S, Guo X, Yang X, Wei Y, Yang Z, Cheng S, et al. Genetic determinants of IgG antibody response to COVID-19 vaccination. Am J Hum Genet. 2024 Jan 4;111(1):181–99.

9. Yang M, Wu Y, Lu Y, Liu C, Sun J, Liao M, et al. Genome-wide scan identifies variant in TNFSF13 associated with serum IgM in a healthy Chinese male population. PLoS One. 2012 Oct 31;7(10):e47990.

10. Kim JH, Cheong HS, Park JS, Jang AS, Uh ST, Kim YH, et al. A genome-wide association study of total serum and mite-specific IgEs in asthma patients. PLoS One. 2013 Aug 13;8(8):e71958.

11. Willis TW, Gkrania-Klotsas E, Wareham NJ, McKinney EF, Lyons PA, Smith KGC, et al. Leveraging pleiotropy identifies common-variant associations with selective IgA deficiency. Clin Immunol. 2024 Nov 1;268(110356):110356.

12. Dennis JK, Sealock JM, Straub P, Lee YH, Hucks D, Actkins K 'era, et al. Clinical Laboratory Test-Wide Association Scan of Polygenic Scores Identifies Biomarkers of Complex Disease. Genome Med. 2021 Jan;13(1):6.

13. Scepanovic P, Alanio C, Hammer C, Hodel F, Bergstedt J, Patin E, et al. Human genetic variants and age are the strongest predictors of humoral immune responses to common pathogens and vaccines. Genome Med. 2018 Jul 27;10(1):59.

14. Eldjarn GH, Ferkingstad E, Lund SH, Helgason H, Magnusson OT, Gunnarsdottir K, et al. Large-scale plasma proteomics comparisons through genetics and disease associations. Nature. 2023 Oct 4;622(7982):348–58.

15. Pietzner M, Wheeler E, Carrasco-Zanini J, Cortes A, Koprulu M, Wörheide MA, et al. Mapping the proteo-genomic convergence of human diseases. Science. 2021 Nov 12;374(6569):eabj1541.

16. Jonsson S, Sveinbjornsson G, de Lapuente Portilla AL, Swaminathan B, Plomp R, Dekkers G, et al. Identification of Sequence Variants Influencing Immunoglobulin Levels. Nat Genet. 2017 Aug;49(8):1182–91.

17. Cordell HJ, Fryett JJ, Ueno K, Darlay R, Aiba Y, Hitomi Y, et al. An International Genome-Wide Meta-Analysis of Primary Biliary Cholangitis: Novel Risk Loci and Candidate Drugs. J Hepatol. 2021 Sep;75(3):572–81.

18. Ji SG, Juran BD, Mucha S, Folseraas T, Jostins L, Melum E, et al. Genome-Wide Association Study of Primary Sclerosing Cholangitis Identifies New Risk Loci and Quantifies the Genetic Relationship with Inflammatory Bowel Disease. Nat Genet. 2017 Feb;49(2):269–73.

19. Ishigaki K, Sakaue S, Terao C, Luo Y, Sonehara K, Yamaguchi K, et al. Multi-Ancestry Genome-Wide Association Analyses Identify Novel Genetic Mechanisms in Rheumatoid Arthritis. Nat Genet. 2022 Nov;54(11):1640–51.

20. Bentham J, Morris DL, Cunninghame Graham DS, Pinder CL, Tombleson P, Behrens TW, et al. Genetic Association Analyses Implicate Aberrant Regulation of Innate and Adaptive Immunity Genes in the Pathogenesis of Systemic Lupus Erythematosus. Nat Genet. 2015 Dec;47(12):1457–64.

21. de Lange KM, Moutsianas L, Lee JC, Lamb CA, Luo Y, Kennedy NA, et al. Genome-Wide Association Study Implicates Immune Activation of Multiple Integrin Genes in Inflammatory Bowel Disease. Nat Genet. 2017 Feb;49(2):256–61.

22. Chiou J, Geusz RJ, Okino ML, Han JY, Miller M, Melton R, et al. Interpreting Type 1 Diabetes Risk with Genetics and Single-Cell Epigenomics. Nature. 2021 Jun;594(7863):398–402.

23. INTERNATIONAL MULTIPLE SCLEROSIS GENETICS CONSORTIUM. Multiple Sclerosis Genomic Map Implicates Peripheral Immune Cells and Microglia in Susceptibility. Science. 2019 Sep;365(6460):eaav7188.

24. Pasanen A, Sliz E, Huilaja L, FinnGen, Estonian Biobank Research Team, Reimann E, et al. Identifying atopic dermatitis risk loci in 1,094,060 individuals with subanalysis of disease severity and onset. J Invest Dermatol. 2024 Nov;144(11):2417–25.

25. Figuerêdo J, Krebs K, Pujol-Gualdo N, Haller T, Võsa U, Volke V, et al. Uncovering the shared genetic components of thyroid disorders and reproductive health. Eur J Endocrinol. 2024 Aug 5;191(2):211–22.

26. Trynka G, Hunt KA, Bockett NA, Romanos J, Mistry V, Szperl A, et al. Dense Genotyping Identifies and Localizes Multiple Common and Rare Variant Association Signals in Celiac Disease. Nat Genet. 2011 Dec;43(12):1193–201.

27. Kiryluk K, Sanchez-Rodriguez E, Zhou XJ, Zanoni F, Liu L, Mladkova N, et al. Genome-wide association analyses define pathogenic signaling pathways and prioritize drug targets for IgA nephropathy. Nat Genet. 2023 Jul;55(7):1091–105.

28. Chen MH, Raffield LM, Mousas A, Sakaue S, Huffman JE, Moscati A, et al. Trans-Ethnic and Ancestry-Specific Blood-Cell Genetics in 746,667 Individuals from 5 Global Populations. Cell. 2020 Sep;182(5):1198–213.e14.

29. Cerezo M, Sollis E, Ji Y, Lewis E, Abid A, Bircan KO, et al. The NHGRI-EBI GWAS Catalog: standards for reusability, sustainability and diversity. Nucleic Acids Res. 2025 Jan 6;53(D1):D998–1005.

30. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-Generation PLINK: Rising to the Challenge of Larger and Richer Datasets. Gigascience. 2015 Dec;4(1):s13742–015 – 0047–8.

31. Purcell S, Chang C. PLINK 2.0.

32. Lewis MJ, Wang S. locuszoomr: an R package for visualizing publication-ready regional gene locus plots. Bioinform Adv. 2024 Dec 26;5(1):vbaf006.

33. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics. 2015 Nov 1;31(21):3555–7.

34. Buniello A, Suveges D, Cruz-Castillo C, Llinares MB, Cornu H, Lopez I, et al. Open Targets Platform: facilitating therapeutic hypotheses building in drug discovery. Nucleic Acids Res. 2025 Jan 6;53(D1):D1467–75.

35. Dyer SC, Austine-Orimoloye O, Azov AG, Barba M, Barnes I, Barrera-Enriquez VP, et al. Ensembl 2025. Nucleic Acids Res. 2025 Jan 6;53(D1):D948–57.

36. Gudjonsson A, Gudmundsdottir V, Axelsson GT, Gudmundsson EF, Jonsson BG, Launer LJ, et al. A genome-wide association study of serum proteins reveals shared loci with common diseases. Nat Commun. 2022 Jan 25;13(1):480.

37. Poli MC, Aksentijevich I, Bousfiha AA, Cunningham-Rundles C, Hambleton S, Klein C, et al. Human inborn errors of immunity: 2024 update on the classification from the International Union of Immunological Societies Expert Committee. Journal of Human Immunity [Internet]. 2025 May 5 [cited 2025 Aug 10];1(1). Available from: http://dx.doi.org/10.70962/jhi.20250003

38. Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat Protoc. 2009 Jul 23;4(8):1184–91.

39. Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, Brazma A, et al. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. Bioinformatics. 2005 Aug 15;21(16):3439–40.

40. Speed D, Balding DJ. SumHer Better Estimates the SNP Heritability of Complex Traits from Summary Statistics. Nat Genet. 2019 Feb;51(2):277–84.

41. A Global Reference for Human Genetic Variation. Nature. 2015 Oct;526(7571):68–74.

42. Byrska-Bishop M, Evani US, Zhao X, Basile AO, Abel HJ, Regier AA, et al. High Coverage Whole Genome Sequencing of the Expanded 1000 Genomes Project Cohort Including 602 Trios. bioRxiv. 2021 Feb;2021.02.06.430068.

43. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation (Camb). 2021 Aug 28;2(3):100141.

44. Yu G, He QY. ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. Mol Biosyst. 2016 Feb 26;12(2):477–9.

45. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. PLoS Genet. 2014 May;10(5):e1004383.

46. Wakefield J. Bayes factors for genome-wide association studies: comparison with P-values. Genet Epidemiol. 2009 Jan 1;33(1):79–86.

47. Kanai M, Elzur R, Zhou W, Global Biobank Meta-analysis Initiative, Daly MJ, Finucane HK. Meta-analysis fine-mapping is often miscalibrated at single-variant resolution. Cell Genom. 2022 Dec 14;2(12):100210.

48. Li X, Bu X, Lu B, Avraham H, Flavell RA, Lim B. The hematopoiesis-specific GTP-binding protein RhoH is GTPase deficient and modulates activities of other Rho GTPases by an inhibitory function. Mol Cell Biol. 2002 Feb;22(4):1158–71.

49. Crequer A, Troeger A, Patin E, Ma CS, Picard C, Pedergnana V, et al. Human RHOH deficiency causes T cell defects and susceptibility to EV-HPV infections. J Clin Invest. 2012 Sep 4;122(9):3239–47.

50. Dorn T. Analysis of RhoH function in vivo [Internet] [Doctoral dissertation]. [Ludwig-Maximiliansuniversitaet Muenchen]: Ludwig-Maximiliansuniversitaet Muenchen; 2007. Available from: https://edoc.ub.uni-muenchen.de/7457/1/Dorn_Tatjana.pdf

51. Eyre S, Hinks A, Bowes J, Flynn E, Martin P, Wilson AG, et al. Overlapping Genetic Susceptibility Variants between Three Autoimmune Disorders: Rheumatoid Arthritis, Type 1 Diabetes and Coeliac Disease. Arthritis Res Ther. 2010 Sep;12(5):R175.

52. Hong R, Lai N, Ouchida R, Xiong E, Zhou Y, Min Q, et al. The B cell novel protein 1 (BCNP1) regulates BCR signaling and B cell apoptosis. Eur J Immunol. 2019 Jun;49(6):911–7.

53. Gold L, Walker JJ, Wilcox SK, Williams S. Advances in human proteomics at high scale with the SOMAscan proteomics platform. N Biotechnol. 2012 Jun 15;29(5):543–9.

54. Largest proteome study enlists 14 biopharmas. Nat Biotechnol. 2025 Feb 14;43(2):153.

55. Li J, Jørgensen SF, Maggadottir SM, Bakay M, Warnatz K, Glessner J, et al. Association of CLEC16A with Human Common Variable Immunodeficiency Disorder and Role in Murine B Cells. Nat Commun. 2015 Apr;6:6804.

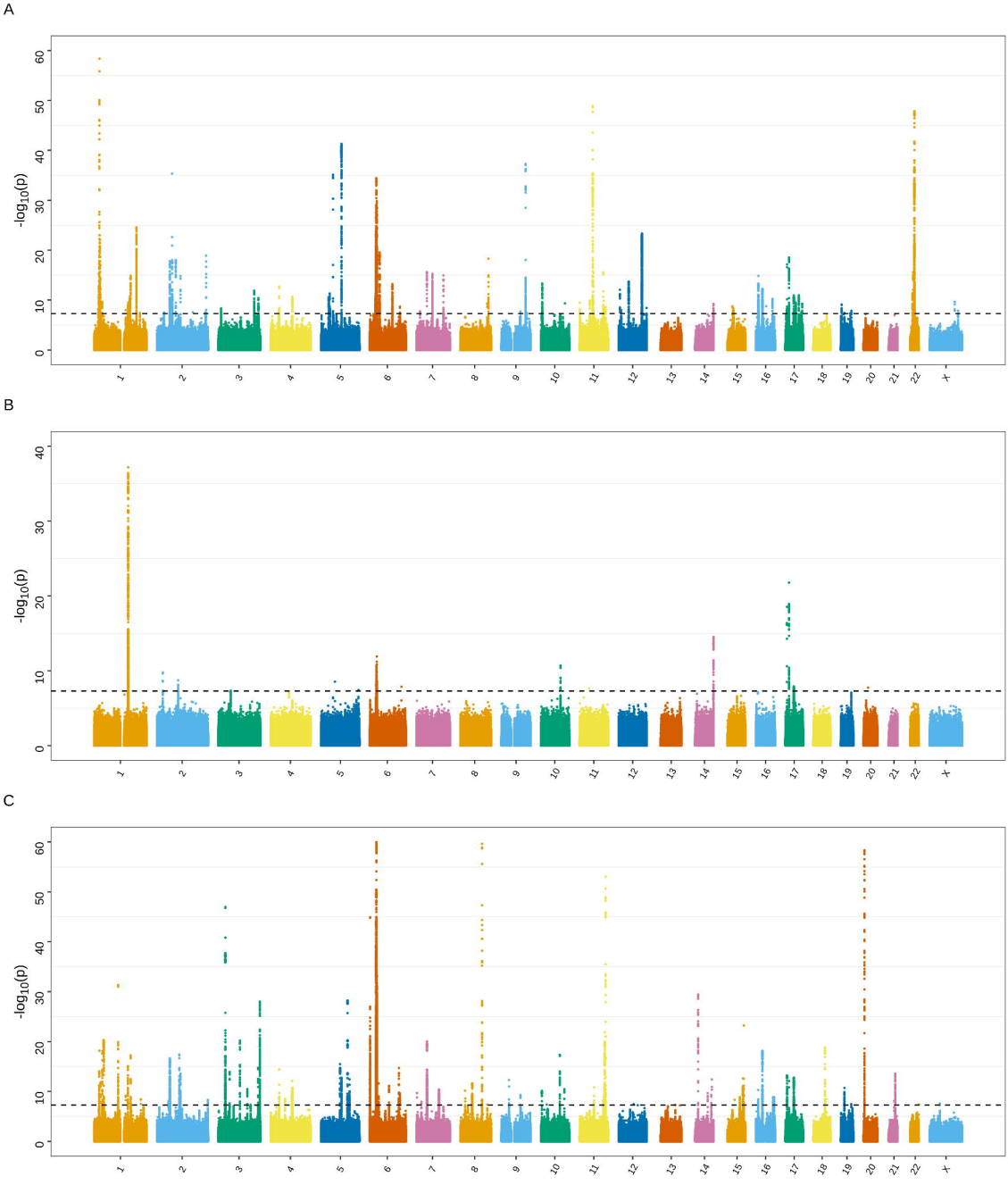
56. Bronson PG, Chang D, Bhangale T, Seldin MF, Ferreira RC, Urcelay E, et al. Common Variants at PVT1, ATG13-AMBRA1, AHI1 and CLEC16A Are Associated with Selective IgA Deficiency. Nat Genet. 2016 Nov;48(11):1425–9.

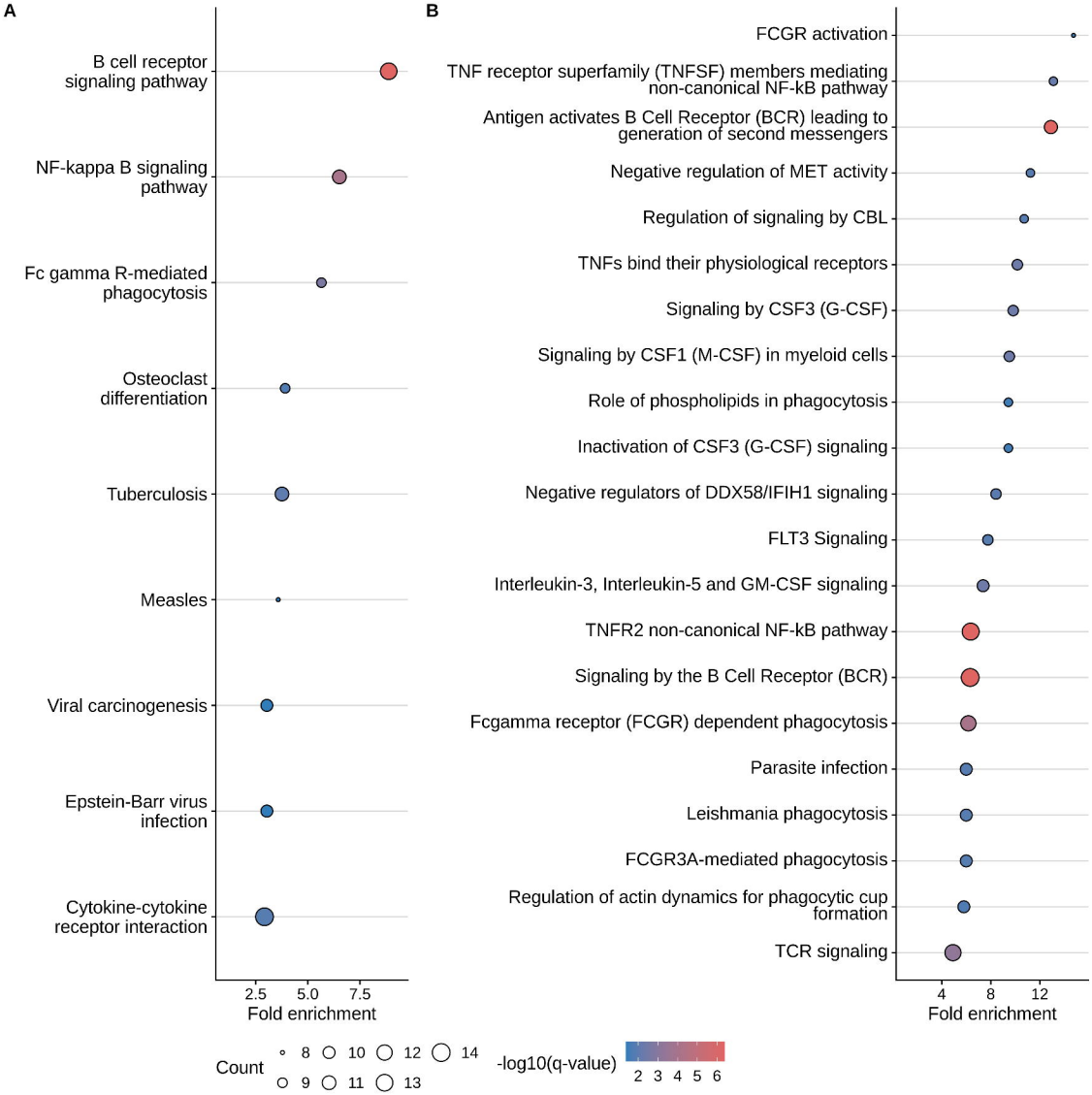
57. Mölder F, Jablonski KP, Letcher B, Hall MB, Tomkins-Tinch CH, Sochat V, et al. Sustainable Data Analysis with Snakemake. F1000Res [Internet]. 2021 Apr [cited 2022 Sep 27];10(33). Available from: http://dx.doi.org/10.12688/f1000research.29032.2

58. Jefferis R, Lefranc MP. Human immunoglobulin allotypes: possible implications for immunogenicity. MAbs. 2009 Jul;1(4):332–8.

59. Purcell RA, Aurelia LC, Esterbauer R, Allen LF, Bond KA, Williamson DA, et al. Immunoglobulin G genetic variation can confound assessment of antibody levels via altered binding to detection reagents. Clin Transl Immunology. 2024 Feb 29;13(3):e1494.

60. Calonga-Solís V, Malheiros D, Beltrame MH, Vargas L de B, Dourado RM, Issler HC, et al. Unveiling the diversity of immunoglobulin heavy constant gamma (IGHG) gene segments in Brazilian populations reveals 28 novel alleles and evidence of gene conversion and natural selection. Front Immunol. 2019 Jun 4;10:1161.


A
-log10(p)
0
10
20
30
40
50
60
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
B
-log10(p)
0
10
20
30
40
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
C
-log10(p)
0
10
20
30
40
50
60
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X


A
B cell receptor signaling pathway
NF-kappa B signaling pathway
Fc gamma R-mediated phagocytosis
Osteoclast differentiation
Tuberculosis
Measles
Viral carcinogenesis
Epstein-Barr virus infection
Cytokine-cytokine receptor interaction
2.5
5.0
7.5
Fold enrichment
B
FCGR activation
TNF receptor superfamily (TNFSF) members mediating non-canonical NF-kB pathway
Antigen activates B Cell Receptor (BCR) leading to generation of second messengers
Negative regulation of MET activity
Regulation of signaling by CBL
TNFs bind their physiological receptors
Signaling by CSF3 (G-CSF)
Signaling by CSF1 (M-CSF) in myeloid cells
Role of phospholipids in phagocytosis
Inactivation of CSF3 (G-CSF) signaling
Negative regulators of DDX58/IFIH1 signaling
FLT3 Signaling
Interleukin-3, Interleukin-5 and GM-CSF signaling
TNFR2 non-canonical NF-kB pathway
Signaling by the B Cell Receptor (BCR)
Fcgamma receptor (FCGR) dependent phagocytosis
Parasite infection
Leishmania phagocytosis
FCGR3A-mediated phagocytosis
Regulation of actin dynamics for phagocytic cup formation
TCR signaling
4
8
12
Fold enrichment
Count
8
9
10
11
12
13
14
-log10(q-value)
2
3
4
5
6