nature genetics

Common variants at *PVT1*, *ATG13–AMBRA1*, *AHI1* and *CLEC16A* are associated with selective IgA deficiency

Paola G Bronson[1], Diana Chang[1], Tushar Bhangale[2], Michael F Seldin[3], Ward Ortmann[1], Ricardo C Ferreira[4], Elena Urcelay[5], Luis Fernández Pereira[6], Javier Martin[7], Alessandro Plebani[8,9], Vassilios Lougaris[8,9], Vanda Friman[10], Tomáš Freiberger[11,12], Jiri Litzman[13], Vojtech Thon[13,14], Qiang Pan-Hammarström[15], Lennart Hammarström[15], Robert R Graham[1] & Timothy W Behrens[1]

Selective immunoglobulin A deficiency (IgAD) is the most common primary immunodeficiency in Europeans. Our genome-wide association study (GWAS) meta-analysis of 1,635 patients with IgAD and 4,852 controls identified four new significant ($P < 5 \times 10^{-8}$) loci and association with a rare *IFIH1* variant (p.Ile923Val). Peak new variants (*PVT1*, $P = 4.3 \times 10^{-11}$; *ATG13–AMBRA1*, $P = 6.7 \times 10^{-10}$; *AHI1*, $P = 8.4 \times 10^{-10}$; *CLEC16A*, $P = 1.4 \times 10^{-9}$) overlapped with autoimmune markers (3/4) and correlated with 21 putative regulatory variants, including expression quantitative trait loci (eQTLs) for *AHI1* and *DEXI* and DNase hypersensitivity sites in FOXP3$^+$ regulatory T cells. Pathway analysis of the meta-analysis results showed striking association with the KEGG pathway for IgA production (pathway $P < 0.0001$), with 22 of the 30 annotated pathway genes containing at least one variant with $P \leq 0.05$ in the IgAD meta-analysis. These data suggest that a complex network of genetic effects, including genes known to influence the biology of IgA production, contributes to IgAD.

IgAD is the most common primary immunodeficiency and is defined by serum IgA levels <0.07 g/l (ref. 1). Its prevalence in Europeans is 1:600 (ref. 1). Secretory IgA (sIgA) is important for mucosal immunity and gut commensalism[2,3], and clinical features of IgAD include recurrent mucosal infections. In IgAD, B cells fail to terminally differentiate into IgA$^+$ plasma cells; however, IL-21 can restore B cell IgA production *in vitro*[4,5]. IgAD is strongly associated with the human leukocyte antigen (HLA) locus[6] and aggregates in families with autoimmunity[7]. The prevalence of celiac disease is 35 times higher in patients with IgAD, whereas the prevalence of systemic lupus erythematosus (SLE) and type 1 diabetes (T1D) is 10 times higher[8].

Ferreira *et al.*[6] previously imputed *HLA-B*, *HLA-DRB1* and *HLA-DQB1* in a study of IgAD cases and controls (>2,700 individuals)[9]. The primary signal mapped to HLA-DQB1*02 (odds ratio (OR) = 2.8, $P = 7.7 \times 10^{-57}$), because of combined independent effects of the HLA-B*08:01–HLA-DRB1*03:01–HLA-DQB1*02 and HLA-DRB1*07:01–HLA-DQB1*02 haplotypes[9]. There was a secondary signal at HLA-DRB1*01:02 (OR = 4.28, $P = 5.86 \times 10^{-17}$) and a protective effect for HLA-DRB1*15:01 (OR = 0.13, $P = 2.24 \times 10^{-35}$)[9]. HLA-DQB1*02:01 is protective for IgA nephropathy (OR = 0.71, $P = 2.61 \times 10^{-13}$)[10]. None of the IgAD loci outside of the major histocompatibility complex (MHC) region overlapped loci for IgA nephropathy[10].

A previous GWAS identified the common allele *IFIH1* p.Thr946Ala (OR = 0.62, control allele frequency = 0.39) as the first and, thus far, only non-HLA genome-wide significant IgAD locus[9]. The p.Thr946Ala variant is also protective for T1D[11], SLE[12], psoriasis[13] and vitiligo (see URLs). *IFIH1* encodes MDA5, a cytosolic receptor that recognizes double-stranded RNA and initiates interferon pathway activation.

To expand understanding of IgAD risk, we studied four new IgAD cohorts and performed a GWAS meta-analysis of ~9.5 million SNPs in 1,635 cases and 4,852 controls (**Table 1**). Genotypes for untyped markers were imputed for each cohort separately (1000 Genomes Project) and genotypes for variants fully typed in the entire cohort. Up to four controls per new case were iteratively selected on the basis of ancestry eigenvectors[14] to minimize population substructure

[1]Department of Human Genetics, Genentech, Inc., South San Francisco, California, USA. [2]Department of Bioinformatics and Computational Biology, Genentech, Inc., South San Francisco, California, USA. [3]Department of Biochemistry, School of Medicine, University of California, Davis, Davis, California, USA. [4]Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammation Laboratory, Cambridge Institute for Medical Research, Cambridge, UK. [5]Department of Immunology, Instituto de Investigación Sanitaria del Hospital Clínico San Carlos, IdISSC, Madrid, Spain. [6]Department of Immunology, Hospital San Pedro de Alcántara, Cáceres, Spain. [7]Instituto de Parasitología y Biomedicina López-Neyra, CSIC, Granada, Spain. [8]Pediatrics Clinic, Department of Clinical and Experimental Sciences, University of Brescia, Spedali Civili di Brescia, Brescia, Italy. [9]Institute for Molecular Medicine, A. Nocivelli, Department of Clinical and Experimental Sciences, University of Brescia, Spedali Civili di Brescia, Brescia, Italy. [10]Department of Infectious Diseases, University of Gothenburg, Gothenburg, Sweden. [11]Molecular Genetics Laboratory, Centre for Cardiovascular Surgery and Transplantation, Brno, Czech Republic. [12]Central European Institute of Technology, Masaryk University, Brno, Czech Republic. [13]Department of Clinical Immunology and Allergy, Faculty of Medicine, Masaryk University, St. Anne's University Hospital, Brno, Czech Republic. [14]Research Centre for Toxic Compounds in the Environment, Faculty of Science, Masaryk University, Brno, Czech Republic. [15]Division of Clinical Immunology and Transfusion Medicine, Karolinska Institutet, Stockholm, Sweden. Correspondence should be addressed to P.G.B. (paola.bronson@biogen.com) or L.H. (lennart.hammarstrom@ki.se).

Received 4 August 2015; accepted 24 August 2016; published online 10 October 2016; doi:10.1038/ng.3675

Table 1 Case–control cohorts for IgAD GWAS

Cohort	Cases (*n*)	Controls (*n*)	Total (*n*)	Significant ancestry eigenvectors (*n*)	Genomic inflation factor (λ_{GC})[a]	SNPs genotyped (*n*)	SNPs after imputation (*n*)	Median info score
Meta	1,635	4,852	6,487			556,344[b]	9,464,381[b]	
2016 (new)								
Swedish	483	1,932	2,415	2	1.03	423,694	9,464,381	0.988
Spanish	150	230	380	0	1.04	339,552	8,534,763	0.974
Italian	91	364	455	4	1.03	218,770	8,365,602	0.948
Czech	151	602	753	4	1.05	112,822	6,282,267	0.929
2010								
Swedish	421	1,080	1,501	5	1.03	289,843	8,765,152	0.979
Finnish	86	344	430	3	1.01	314,756	8,723,593	0.984
Spanish	253	300	553	3	1.00	538,800	8,794,151	0.985

[a]The genomic inflation factor listed was calculated from genotyped variants. However, genomic inflation factors for each cohort were also estimated separately for imputed SNPs, and results did not differ (for example, the maximum genomic inflation factor for imputed variants was 1.06). [b]Variants genotyped in more than one cohort.

(Online Methods). Genomic inflation factor values were minimal (**Table 1**), indicating that population substructure was adequately addressed. Association analyses were conducted with logistic regression (additive model), accounting for genotype uncertainty and using ancestry eigenvectors as covariates (Online Methods).

The strongest association was with the MHC region, 2.7 kb upstream of *HLA-DQA1* ($P = 3.3 \times 10^{-92}$) (**Table 2**). Protective association between IgAD and *IFIH1* p.Thr946Ala[9] was confirmed ($P = 3.7 \times 10^{-15}$), and association with the rare loss-of-function *IFIH1* p.Ile923Val variant ($P = 2.6 \times 10^{-8}$) was subsequently identified (**Supplementary Table 1**). Val923 has previously been shown to be protective for T1D[11] and psoriasis[13] and abrogates interferon signaling[15]. A rare gain-of-function *IFIH1* p.Arg779His variant has also been reported in an IgAD case with SLE and a type I interferon signature[16].

Our GWAS identified four new significant ($P < 5 \times 10^{-8}$) loci: *PVT1*, *ATG13–AMBRA1*, *AHI1* and *CLEC16A* (**Fig. 1** and **Table 2**). The peak new IgAD variants were (i) a protective 1-bp deletion 91 kb downstream of *PVT1* ($P = 4.3 \times 10^{-11}$); (ii) a variant 289 kb upstream of *ATG13* ($P = 6.7 \times 10^{-10}$); (iii) an intronic *AHI1* variant ($P = 8.4 \times 10^{-10}$); and (iv) a protective 1-bp intronic insertion in *CLEC16A* ($P = 1.4 \times 10^{-9}$) (**Table 2**, **Supplementary Figs. 1–6** and **Supplementary Table 2**). Twenty-eight additional loci contained at least one variant with $5 \times 10^{-8} < P < 1 \times 10^{-5}$, and half (denoted with an asterisk) were known autoimmune loci: *FAS**, *GATA3**, *MYO9B*, *CD86**, *BCL6*, *IKZF2*, *PTGER4**, *IKZF3**, *EGR2**, *FCRL3**, *IL2RA**, *PTPN2**, *IL2**–*IL21**, *CDH23**, *MAST4*, *FOXP1*, *MIR605*, *LINC01098*, *DRMT1*, *LINC00299*, *C22orf42*, *FAM171A1*, *BACH1*, *ZCCHC24*, *NCKAP5*, *LOC100505887*, *TMEM72* and *CD28** (**Supplementary Fig. 7** and **Supplementary Table 3**).

The heritability of IgAD in the Swedish cohort based on genome-wide imputed variants was 0.49 (standard error (SE) = 0.047). This estimate was reduced from 0.49 to 0.39 (SE = 0.059) when we conditioned on the five peak non-MHC variants, the peak MHC variant (rs116041786), and the peak MHC variant after conditioning on the peak MHC variant (rs116350876). When we excluded the MHC region altogether from the input, the estimated heritability dropped to 0.14 (SE = 0.054).

As one approach to compile a list of potential causative gene(s) and allele(s) in the new loci[17], we cross-referenced the peak variants plus 160 correlated variants ($r^2 \geq 0.7$) (44 variants in *PVT1*; 21 variants in *ATG13–AMBRA1*; 65 variants in *AHI1*; and 30 variants in *CLEC16A*; total $n = 164$) against epigenetic data, eQTLs, DNase I–hypersensitive sites (DHSs), transcription factor binding sites (ChIP–seq), consensus motifs, GWAS loci and promoter regions using RegulomeDB[18], Encyclopedia of DNA Elements (ENCODE)[19], Roadmap Epigenomics, Gene Expression Omnibus (GEO), Washington University EpiGenome Browser, ImmunoBase and the GWAS Catalog (see URLs). The RegulomeDB catalog lists regulatory evidence for ~60 million variants, and about 5% (~3 million) are categorized as likely to affect binding (score = 2) or likely to affect binding and linked to expression of a gene target (score = 1).

In our data set, 13.1% (21/160) of the correlated variants were likely to affect binding (RegulomeDB score <3) and were located in regions of open chromatin in an immune cell (as determined by DNase hypersensitivity) and/or in transcription factor binding sites (as determined by DNase footprints, ChIP–seq and/or binding site motifs) (**Supplementary Table 4**). Many of these variants (15/21) were also located in histone marks for active transcription (H3K4me3, H3K27ac or H3K4me1) in regulatory T (T_{reg}) cells, and all of the new loci, except for *AHI1*, encompassed at least one T_{reg} enhancer and active transcription start site. A comparison of the overlap between eQTL data from lymphoblastoid B cells and IgAD meta *P* values using the Bayesian statistical framework Sherlock identified three *cis*-eQTLs in *AHI1* and *DEXI* associated with IgAD (**Supplementary Table 4**).

The *PVT1* locus contains no protein-coding genes, and the long noncoding RNA (lncRNA) *PVT1* appears to be the most likely causative genetic element. *PVT1* has been shown to be important for expression and copy number increase of *MYC* in tumors[20]. The peak *PVT1* variant was in moderate linkage disequilibrium (LD) with four potential regulatory variants. The most interesting of these was rs7001706 ($r^2 = 0.70$), an intronic variant that lies in a FOXP3 transcription factor binding motif and in an H3K4me1 histone mark in T_{reg} cells. When we conditioned the regression analyses on either the peak variant or the potential regulatory variant rs7001706, they accounted for part but not all of the effect of the locus (data not shown), indicating that *PVT1* contains more than one independent signal.

AHI1 is highly expressed in hematopoietic cells and overexpressed in blood cancers[21]. *AHI1* stabilizes BCR-ABL in leukemia cells by recruiting JAK2 and can regulate phosphorylation of JAK2–STAT5 (ref. 22). JAK inhibitors are useful in treating rheumatoid arthritis, psoriasis and alopecia areata. The peak variant was in LD with eight eQTLs, all on the same haplotype. rs2064430 ($r^2 = 0.85$) had the strongest evidence for regulatory binding (RegulomeDB score = 1d) (**Supplementary Table 4**). A comparison of the overlap between eQTL data from lymphoblastoid B cells and IgAD meta *P* value using the Bayesian statistical framework Sherlock highlighted two *cis*-eQTLs in *AHI1* (rs2179781 and rs9647635) with strong positive log Bayes factor (LBF) scores (6.83 and 6.63, respectively) (**Supplementary Table 4**). Of the variants highlighted in **Supplementary Table 4**, rs2179781 had the smallest meta *P* value. rs2179781[A] was associated with reduced risk for IgAD and reduced *AHI1* expression (**Supplementary Fig. 8**).

The peak SNP upstream of *ATG13* (autophagy 13) and *AMBRA1* (autophagy/beclin-1 regulator 1) was correlated with two likely regulatory

Table 2 Genome-wide significant results ($P < 5 \times 10^{-8}$) for IgAD GWAS

Closest gene(s)	Variant[a]	Position (hg19)	Minor allele	Frequency in Swedish sample (%)		*P* value	FDR *q* value	OR	Type	Immune diseases sharing this locus[b]
				Cases	Controls					
HLA-DQA1[c]	rs116041786	6:32602396	C	14.4	38.7	3.3×10^{-92}	1.0×10^{-87}	0.38	Intergenic	Celiac, Graves', IBD, Sjögren's[d]
IFIH1	rs1990760	2:163124051	G	29.3	37.3	3.7×10^{-15}	2.7×10^{-12}	0.70	Missense	T1D, SLE, psoriasis, vitiligo[e]
PVT1	rs11299600	8:129204573	1-bp deletion	18.9	25.3	4.3×10^{-11}	2.4×10^{-8}	0.73	Intergenic	RA[f]
ATG13 – AMBRA1	rs4565870	11:46349869	C	28.9	23.6	6.7×10^{-10}	3.5×10^{-7}	1.38	Intergenic	
AHI1	rs7773987	6:135707486	C	51.3	44.6	8.4×10^{-10}	4.3×10^{-7}	1.30	Intronic	MS
CLEC16A	rs34069391	16:11161214	1-bp insertion	14.3	19.6	1.4×10^{-9}	6.9×10^{-7}	0.71	Intronic	T1D, MS, PBC[g]

IBD, inflammatory bowel disease; MS, multiple sclerosis; SLE, systemic lupus erythematosus; PBC, primary biliary cirrhosis; RA, rheumatoid arthritis; T1D, type 1 diabetes; FDR, false discovery rate.
[a]Peak variants were imputed, except for those in *IFIH1* and *CLEC16A*, which were genotyped in all but the Czech cohort. Three loci had suggestive association in our previous GWAS: *PVT1* ($P = 4 \times 10^{-6}$), *DGKZ* ($P = 2 \times 10^{-6}$) and *CLEC16A* ($P = 2 \times 10^{-7}$). [b]Autoimmune diseases reporting a genome-wide significant variant ($P < 5 \times 10^{-8}$) in ImmunoBase and the GWAS Catalog (see URLs) with at least a modest effect (OR ≥ 1.1) in the same direction as the IgAD results and on the same haplotype as a peak IgAD variant ($r^2 > 0.45$)[12]. [c]Genetic association of IgAD with HLA class II loci is well established, particularly for HLA-DQB1*02:01 and HLA-DRB1*01:02 and for the highly protective HLA-DRB1*15:01 allele[6]. Few of the cases (2.1%) were carrying two copies of rs116041786[C] (versus 15.7% of controls). [d]Multiple sclerosis, rheumatoid arthritis, SLE, T1D, vitiligo, allergy and asthma also have some associations reported at this locus. [e]IBD is also associated with this locus. [f]Multiple sclerosis, celiac disease, eczema and allergy also have some associations reported at this locus. [g]Celiac disease and asthma also have some associations reported at this locus.

variants (**Supplementary Table 4**). *ATG13* and *AMBRA1* physically interact[23] and share the same chromatin interaction domain. Autophagy has a role in autoimmunity[24], plasmablast differentiation[25] and immunoglobulin production in plasma cells[26]. *ATG13* and *AMBRA1* are the most likely risk-associated genes in this interval, but the causal variant remains to be determined.

CLEC16A contains several eQTLs for *DEXI*[27], and T1D-associated *CLEC16A* variants[28] are located in an intron that shows chromatin interaction with the *DEXI* promoter, as determined by 3C[27] and Hi-C[29,30] mapping. The peak variant was in LD with seven likely regulatory variants (**Supplementary Table 4**), including rs35300161, rs34972832 and the *DEXI* eQTL rs17806299 ($r^2 = 0.80$). rs35300161 and rs34972832 are present in open chromatin in germinal center and naive B cells (**Supplementary Table 4**) and in an enhancer in various T cell types and tissues. A comparison of the overlap between eQTL data from lymphoblastoid B cells and IgAD meta *P* value with Sherlock highlighted one *cis*-eQTL in *DEXI* (rs741175) with a strong positive LBF score (6.16) (**Supplementary Table 4**).

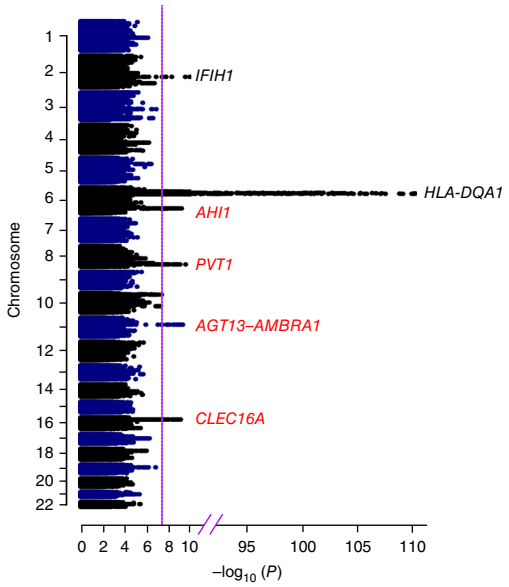

Figure 1 Genome-wide significant loci in the IgAD meta-analysis. The Manhattan plot shows *P* values from a genome-wide meta-analysis of ~9.5 million variants in 1,635 patients with IgAD and 4,852 controls. The vertical line corresponds to the genome-wide significance threshold ($P < 5 \times 10^{-8}$). Loci with genome-wide significant results are labeled; red labels highlight new genome-wide significant loci.

Mice with *Clec16a* knockdown show reduced numbers of B cells[31] and are protected from autoimmunity[32]. Further work is required to determine whether *CLEC16A* or *DEXI* is the relevant gene in this interval.

We next applied pathway analysis to the meta-analysis results using genomic randomization (PARIS)[33] to estimate empirical significance (n = 10,000 randomizations) and an alternate pathway algorithm based on LD-independent intervals for replication (INRICH)[34]. PARIS reduces bias from gene size, pathway size, LD blocks and SNP coverage. Preliminary testing with PARIS v1.1 using imputed data has not been completed and, thus, only the GWAS chip content was selected for this experiment. Using PARIS, we grouped variants into LD features and single SNPs into linkage equilibrium (LE) features. The LD and LE features were grouped by KEGG pathways and also by our user-defined Treg signature pathway. The features in each pathway were permuted 10,000 times with a randomly selected set of features of similar size. The empirical *P* value is based on the number of features with $P < 0.05$ in a pathway, in comparison to the number of features with $P < 0.05$ in the permuted pathway. IgAD was associated (empirical $P < 0.0001$) with 7 of 221 (3%) KEGG pathways (total n genes queried = 5,701) (**Supplementary Table 5**).

Of particular interest, one of the top pathways identified was the intestinal immune network for IgA production (empirical $P < 0.0001$), with 22 of 30 genes in the pathway containing variants with $P < 0.05$ in the IgAD meta-analysis (**Fig. 2** and **Supplementary Table 6**). These include genes encoding inflammatory and immune regulatory cytokines (*IL2*, *IL4*, *IL6*, *IL10*, *IL15*, *APRIL* and *BAFF*), key cell surface molecules on B and T cells (*TACI*, *CD40*, *IL15R*, *CD28*, *ICOS* and *ICOSLG*), molecules essential for the homing of lymphocytes to the gut (*ITGA4*, *CCR9*, *MADCAM1*, *CXCL12*, *CCL28*, *LTBR* and *MAP3KI4*) and DNA deaminase required for B cell class-switch recombination (*AICDA*). These data suggest that the genetic contribution to IgA deficiency includes many genes known to regulate IgA production.

Similar to rheumatoid arthritis and T1D GWAS SNPs, which are enriched for active histone marks (H3k4me3) in T_{reg} cells[17] and lymphoid gene enhancers[35], respectively, IgAD-associated variants were enriched for active histone marks and enhancers in T_{reg} cells. When we applied a set of T_{reg} signature genes (27 autosomal genes that are

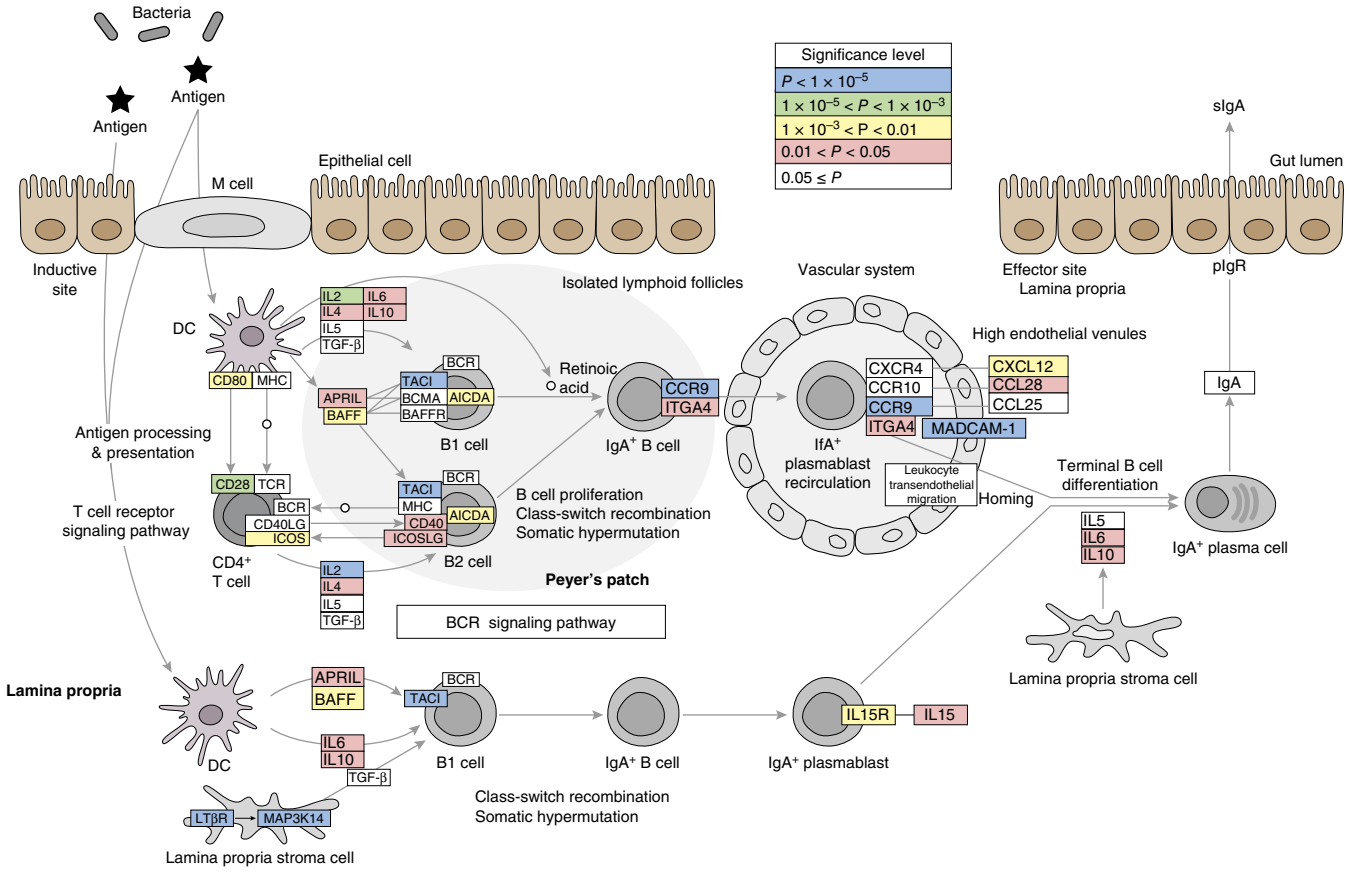

Figure 2 KEGG pathway for IgA production associated with IgAD. GWAS-based pathway analysis of IgAD meta-analysis *P* values was performed for all 221 pathways in the KEGG database using PARIS and INRICH software (Online Methods). A top identified pathway ($P < 0.0001$, based on 10,000 permutations) was the 'intestinal immune network for IgA production', where 22 of the 30 annotated pathway genes contained variants with $P < 0.05$ in the IgAD meta-analysis. The colors of the boxes around gene names correspond to the smallest *P* value observed in the IgAD meta-analysis within that gene.

significantly upregulated in FOXP3$^+$ T_{reg} cells versus conventional FOXP3$^-$ CD4$^+$ T cells[36]) as a user-defined pathway in PARIS pathway analysis, we found that the T_{reg} signature pathway was significantly enriched for association with IgAD (empirical $P < 1 \times 10^{-5}$) (**Supplementary Table 7**). *FOXP3* is an important transcriptional regulator of T_{reg} differentiation[37], and germline mutations of *FOXP3* cause IPEX[38], a disease characterized by impaired T_{reg} function and life-threatening autoimmunity. T_{reg} cells in B cell the germinal centers of Peyer's patches have been shown to influence the ability of sIgA to maintain microbial diversity[29].

Using GWAS summary statistics for T1D and rheumatoid arthritis, we tested for shared genetics between these two diseases and IgAD. Applying GPA[30], we found significant shared genetics both between T1D and IgAD and between rheumatoid arthritis and IgAD. GPA uses a likelihood-ratio test to compare a model of the pairwise summary statistics with no shared genetics to a model with shared genetics. Results indicate significant shared genetics, even when variants in the MHC region were omitted ($P < 1.6 \times 10^{-7}$).

In conclusion, we identified four new susceptibility loci for IgAD—*PVT1*, *AHI1*, *AMBRA1–ATG13* and *CLEC16A*. Interrogation of peak IgAD variants and SNPs in LD identified 21 putative regulatory variants. Pathway analyses highlighted enrichment for association in the intestinal immune network for IgA production and in T_{reg} signature genes. Further work is needed to validate new associations via genotyping, to identify causal variants in the new loci, to explore the regulatory role of associated variants through functional studies and to investigate autoimmune mechanisms that contribute to IgAD pathogenesis.

URLs. Hi-C mapping, http://promoter.bx.psu.edu/hi-c/index.html; LocusZoom, http://genome.sph.umich.edu/wiki/LocusZoom_Standalone; RegulomeDB, http://regulome.stanford.edu/; ImmunoBase, http://immunobase.org/; NHGRI-EBI GWAS Catalog, http://www.ebi.ac.uk/gwas; GenABEL, http://genabel.org/GenABEL/estlambda.html; GCTA-REML, http://cnsgenomics.com/software/gcta/reml.html; ENCODE, http://genome.gov/encode; Washington University EpiGenome Browser, http://epigenomegateway.wustl.edu/browser; Roadmap Epigenomics, http://roadmapepigenomics.org/; Sherlock, http://sherlock.ucsf.edu/; STRING, http://string-db.org/; R software, http://r-project.org/; rheumatoid arthritis GWAS summary statistics, http://plaza.umin.ac.jp/~yokada/datasource/software.htm; PLINK, http://pngu.mgh.harvard.edu/purcell/plink/; EIGENSOFT, http://hsph.harvard.edu/alkes-price/software; KEGG, http://www.genome.jp/kegg/pathway.html; PARIS, http://ritchielab.psu.edu/software/paris-download; INRICH, http://atgu.mgh.harvard.edu/inrich; SHAPEIT, http://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html; IMPUTE2, http://mathgen.stats.ox.ac.uk/impute/impute_v2.html; SNPTEST, http://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html; liftOver, http://genome.ucsc.edu/cgi-bin/hgLiftOver; gctaPower, http://cnsgenomics.com/shiny/gctaPower.

Data access. GWAS data generated in this study are accessible through Genentech, Inc., at http://research-pub.gene.com/bronson_et_al_2016.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank the individuals who participated in this study as case and control study subjects. We thank B. Yaspan, J. Kim, J. Sitrin and T. Hung for insightful discussion, J. Tom and O. Mayba for helpful feedback on a manuscript draft, and A. Bruce for sharing her graphical expertise. Genentech funded the current GWAS. A. Lee and P. Gregersen performed sample genotyping at the Laboratory of Genomics and Human Genetics at the Feinstein Institute for Medical Research. These studies were supported by the US National Institutes of Health (U19AI067152 and AR043274), the Swedish Research Council, the European Research Council (242551-ImmunoSwitch), EURO-PADnet grant 201549, CETOCOEN PLUS and the Fondazione C. Golgi, Brescia, Italy. Financial support was also provided through the regional agreement on medical training and clinical research (ALF) between the Stockholm County Council and Karolinska Institutet. Population allele and genotype frequencies were obtained from a data source funded by the Nordic Center of Excellence in Disease Genetics based on regional samples from Finland and Sweden. The Helsinki 4 study was supported by the Yale Center for Human Genetics and Genomics and the Yale Program on Neurogenetics and by US National Institutes of Health grants R01NS057756 and U24NS051869.

AUTHOR CONTRIBUTIONS

P.G.B. and D.C. carried out the analyses for this study. P.G.B., L.H., T.W.B., R.R.G. and T.B. conceived and directed this study. L.H., Q.P.-H., A.P., V.L., T.F., J.L., E.U., L.F.P., V.F. and V.T. performed subject diagnosis, coordinated the enrollment of subjects and provided access to genotyping data sets. M.F.S. and J.M. provided access to genotypes for healthy controls. M.F.S. provided guidance for addressing population structure due to ancestry. M.F.S., Q.P.-H., R.C.F., T.B., R.R.G. and W.O. contributed to data access and analysis. P.G.B., T.W.B., L.H. and R.R.G. wrote the manuscript with collaboration from coauthors. All authors discussed the results and commented on the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Pan-Hammarström, Q. & Hammarström, L. Antibody deficiency diseases. *Eur. J. Immunol.* **38**, 327–333 (2008).
2. Suzuki, K. *et al.* Aberrant expansion of segmented filamentous bacteria in IgA-deficient gut. *Proc. Natl. Acad. Sci. USA* **101**, 1981–1986 (2004).
3. Cong, Y., Feng, T., Fujihashi, K., Schoeb, T.R. & Elson, C.O. A dominant, coordinated T regulatory cell–IgA response to the intestinal microbiota. *Proc. Natl. Acad. Sci. USA* **106**, 19256–19261 (2009).
4. Borte, S. *et al.* Interleukin-21 restores immunoglobulin production *ex vivo* in patients with common variable immunodeficiency and selective IgA deficiency. *Blood* **114**, 4089–4098 (2009).
5. Cao, A.T. *et al.* Interleukin (IL)-21 promotes intestinal IgA response to microbiota. *Mucosal Immunol.* **8**, 1072–1082 (2015).
6. Ferreira, R.C. *et al.* High-density SNP mapping of the HLA region identifies multiple independent susceptibility loci associated with selective IgA deficiency. *PLoS Genet.* **8**, e1002476 (2012).
7. Oen, K., Petty, R.E. & Schroeder, M.L. Immunoglobulin A deficiency: genetic studies. *Tissue Antigens* **19**, 174–182 (1982).
8. Ludvigsson, J.F., Neovius, M. & Hammarström, L. Association between IgA deficiency & other autoimmune conditions: a population-based matched cohort study. *J. Clin. Immunol.* **34**, 444–451 (2014).
9. Ferreira, R.C. *et al.* Association of *IFIH1* and other autoimmunity risk alleles with selective IgA deficiency. *Nat. Genet.* **42**, 777–780 (2010).
10. Kiryluk, K. *et al.* Discovery of new risk loci for IgA nephropathy implicates genes involved in immunity against intestinal pathogens. *Nat. Genet.* **46**, 1187–1196 (2014).
11. Nejentsev, S., Walker, N., Riches, D., Egholm, M. & Todd, J.A. Rare variants of *IFIH1*, a gene implicated in antiviral responses, protect against type 1 diabetes. *Science* **324**, 387–389 (2009).
12. Cunninghame Graham, D.S. *et al.* Association of *NCF2*, *IKZF1*, *IRF8*, *IFIH1*, and *TYK2* with systemic lupus erythematosus. *PLoS Genet.* **7**, e1002341 (2011).
13. Li, Y. *et al.* Carriers of rare missense variants in *IFIH1* are protected from psoriasis. *J. Invest. Dermatol.* **130**, 2768–2772 (2010).
14. Gregersen, P.K. *et al.* Risk for myasthenia gravis maps to a (151) Pro→Ala change in TNIP1 and to human leukocyte antigen–B*08. *Ann. Neurol.* **72**, 927–935 (2012).
15. Shigemoto, T. *et al.* Identification of loss of function mutations in human genes encoding RIG-I and MDA5: implications for resistance to type I diabetes. *J. Biol. Chem.* **284**, 13348–13354 (2009).
16. Van Eyck, L. *et al.* Brief report: *IFIH1* mutation causes systemic lupus erythematosus with selective IgA deficiency. *Arthritis Rheumatol.* **67**, 1592–1597 (2015).
17. Okada, Y. *et al.* Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* **506**, 376–381 (2014).
18. Boyle, A.P. *et al.* Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* **22**, 1790–1797 (2012).
19. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
20. Tseng, Y.Y. *et al. PVT1* dependence in cancer with *MYC* copy-number increase. *Nature* **512**, 82–86 (2014).
21. Ringrose, A. *et al.* Evidence for an oncogenic role of AHI-1 in Sezary syndrome, a leukemic variant of human cutaneous T-cell lymphomas. *Leukemia* **20**, 1593–1601 (2006).
22. Zhou, L.L. *et al.* AHI-1 interacts with BCR-ABL and modulates BCR-ABL transforming activity and imatinib response of CML stem/progenitor cells. *J. Exp. Med.* **205**, 2657–2671 (2008).
23. Nazio, F. *et al.* mTOR inhibits autophagy by controlling ULK1 ubiquitylation, self-association and function through AMBRA1 and TRAF6. *Nat. Cell Biol.* **15**, 406–416 (2013).
24. Pierdominici, M. *et al.* Role of autophagy in immunity and autoimmunity, with a special focus on systemic lupus erythematosus. *FASEB J.* **26**, 1400–1412 (2012).
25. Clarke, A.J. *et al.* Autophagy is activated in systemic lupus erythematosus and required for plasmablast development. *Ann. Rheum. Dis.* **74**, 912–920 (2015).
26. Pengo, N. *et al.* Plasma cells require autophagy for sustainable immunoglobulin production. *Nat. Immunol.* **14**, 298–305 (2013).
27. Davison, L.J. *et al.* Long-range DNA looping and gene expression analyses identify *DEXI* as an autoimmune disease candidate gene. *Hum. Mol. Genet.* **21**, 322–333 (2012).
28. Todd, J.A. *et al.* Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nat. Genet.* **39**, 857–864 (2007).
29. Kawamoto, S. *et al.* Foxp3$^+$ T cells regulate immunoglobulin a selection and facilitate diversification of bacterial species responsible for immune homeostasis. *Immunity* **41**, 152–165 (2014).
30. Chung, D., Yang, C., Li, C., Gelernter, J. & Zhao, H. GPA: a statistical approach to prioritizing GWAS results by integrating pleiotropy and annotation. *PLoS Genet.* **10**, e1004787 (2014).
31. Li, J. *et al.* Association of *CLEC16A* with human common variable immunodeficiency disorder and role in murine B cells. *Nat. Commun.* **6**, 6804 (2015).
32. Schuster, C. *et al.* The autoimmunity-associated gene *CLEC16A* modulates thymic epithelial cell autophagy and alters T cell selection. *Immunity* **42**, 942–952 (2015).
33. Yaspan, B.L. *et al.* Genetic analysis of biological pathway data through genomic randomization. *Hum. Genet.* **129**, 563–571 (2011).
34. Lee, P.H., O'Dushlaine, C., Thomas, B. & Purcell, S.M. INRICH: interval-based enrichment analysis for genome-wide association studies. *Bioinformatics* **28**, 1797–1799 (2012).
35. Onengut-Gumuscu, S. *et al.* Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. *Nat. Genet.* **47**, 381–386 (2015).
36. Schmidl, C. *et al.* The enhancer and promoter landscape of human regulatory and conventional T-cell subpopulations. *Blood* **123**, e68–e78 (2014).
37. Fontenot, J.D. *et al.* Regulatory T cell lineage specification by the forkhead transcription factor Foxp3. *Immunity* **22**, 329–341 (2005).
38. Bennett, C.L. *et al.* The immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome (IPEX) is caused by mutations of *FOXP3*. *Nat. Genet.* **27**, 20–21 (2001).

ONLINE METHODS

The GWAS tested ~9.5 million variants in 1,635 patients with IgAD (clinical diagnosis defined as serum IgA ≤0.07 g/l with normal IgM and total IgG levels)[39] and 4,852 controls from seven independent case–control cohorts (**Table 1**).

Samples. Only IgAD cases without autoimmune diseases comorbidity (for example, type 1 diabetes or celiac disease) were included in this study. Furthermore, IgAD cases were screened for celiac disease (autoantibodies to tissue transglutaminase) and excluded from the study if positive.

Genotyping. Illumina developed all of the genotyping arrays. We genotyped cases on Omni1-Quad and Omni2.5 arrays, Swedish controls (Swedish TwinGene Project)[40] on OmniExpress arrays, Spanish controls on Omni1-Quad arrays, Italian controls on 1Mb arrays (Milan Hypergenes Study)[41] except for 42 Italian controls genotyped on the 550K array (New York Cancer Project; NYCP)[42], and Czech controls on HumanHap1M, Omni1-Quad, and Omni 2.5 arrays (NCI Division of Cancer Epidemiology and Genetics) (Imputation Reference Data set). Sixty-four Czech controls were genotyped on HumanHap500 arrays (NCI CGEMS Project). Ferreira *et al.* genotyped the Swedish cohort[9], which included 34 Icelandic cases, on Hap300 arrays (except for 175 cases genotyped on Human610 arrays), Finnish cases on 610-Quad arrays, Finnish controls on CNV370 arrays, and the Spanish cohort on 610-Quad v1 arrays. See the "SNPs genotyped (*n*)" column in **Table 1** for the number of SNPs that overlapped across the arrays for each of the seven cohorts. All SNPs were mapped to build hg19 coordinates using liftOver.

Quality control after genotyping and before imputation. Before imputation, we removed variants with a genotyping rate <98%, ambiguity (A/T SNPs and C/G SNPs), evidence of deviation from Hardy–Weinberg equilibrium in controls ($P < 1 \times 10^{-4}$), and MAF $< 1 \times 10^{-6}$. Subsequently, we removed individuals who were missing >1% of genotypes and who had heterozygosity ±6 s.d. from the mean. To estimate identity by descent (IBD) and conduct principal-components analysis (PCA) (EIGENSOFT v3.0; ref. 43), we randomly selected 100,000 LD-pruned variants (LD pruning was based on $r^2 > 0.2$; excluding insertions and deletions, variants in regions of long-range LD[44] or in the MHC region (chr. 6: 24–36 Mb) or in a large inversion on chromosome 8 (ref. 45), and variants with MAF <0.05) and included 878 ancestry-informative markers[46]. We estimated and plotted IBD for all pairs of samples and removed duplicate samples and cryptically related individuals ($P(\text{IBD} = 0) < 0.8$). For PCA, we used five outlier removal iterations and ten principal components along which to remove outliers during each outlier removal iteration. Any sample exceeding 6 s.d. along one of the principal components was considered a population outlier and was removed. We then reran PCA without the ancestry outliers.

For the four new cohorts, we iteratively selected up to four controls, similarly to a previously described approach (developed by M.F.S.), to minimize population substructure differences between cases and controls[14]. Each control could only be selected once. For matching, we sequentially selected a control that passed quality control and best matched the case, considering the eigenvalue results from principal components contributing at least 1% of variance determined from principal components with significant *P* values ($P < 0.05$) in the Tracy–Wisdom test[43]. We estimated the percentage of variance explained by subtracting the eigenvector value of the first non-significant principal component from each significant principal component, summing this value and dividing each eigenvector value by the sum. Eigenvalues for each principal component were adjusted for the eigenvector scale (adjusted for the relative contribution of each eigenvector in describing the data). The lowest cumulative sum of the absolute differences (Manhattan distance for each principal component after adjustment based on relative eigenvalues) between all cases and all controls was scored as the best match.

Subsequently, we reran PCA a third time for each cohort using the final set of case–control samples. Principal components contributing at least 1% of variance determined from principal components with significant *P* values ($P < 0.05$) in the Tracy–Wisdom test[43] were used as covariates in the association analyses. Significant principal components were also plotted with cases and controls color-coded differently for visual inspection. We tested for differential genotyping rates for cases and controls but did not observe any.

Statistical analysis before imputation. We conducted χ^2 tests of association on genotypes for each cohort separately, using only variants that overlapped across the arrays used to genotype the cohort, and estimated a genomic inflation factor (λ_{GC})[47] based on the median χ^2 value for non-MHC genotyped variants adjusted for ancestry (PLINK v1.07; ref. 47). Before meta-analysis and imputation, we plotted *P* values in a quantile–quantile plot (excluding the MHC region) and in a Manhattan plot in R.

Imputation. We imputed genotypes for each cohort separately using the 1000 Genomes Project Phase I integrated variant set release as a reference panel (SHAPEIT v2.r727 (ref. 48) and IMPUTE2 v2.3.0 (ref. 49)). We imputed cases and controls from the same cohort together, using only variants that overlapped across the arrays used to genotype the cohort. Genotypes were imputed in 5-Mb chunks, excluding centromeres. We used an effective population size of 20,000 and 80 haplotypes as templates when phasing observed genotypes, and we performed 30 Monte Carlo Markov chain (MCMC) iterations, where the first 10 MCMC iterations were discarded as burn-in.

Statistical analysis after imputation. We excluded variants with an imputation info score <0.6, MAF <0.01 in the controls or evidence of deviation from Hardy–Weinberg equilibrium ($P < 1 \times 10^{-6}$) and tested association for each case–control cohort separately with logistic regression (additive model), adjusting for ancestry (SNPTEST v2.5-beta4; ref. 50). The median info score for each cohort was as follows: Swedish, 0.988; Spanish, 0.974; Italian, 0.948; Czech, 0.929; Swedish 2010, 0.979; Finnish 2010, 0.984; and Spanish 2010, 0.985. We used a missing data likelihood score test to address genotype uncertainty. We estimated a genomic inflation factor (λ_{GC})[47] based on the distribution of *P* values for each cohort using the GenABEL estlambda package in R (see URLs). We plotted *P* values in a quantile–quantile plot (excluding the MHC region) and in a Manhattan plot in R.

Meta-analysis. We pooled results in a meta-analysis (PLINK v1.07; ref. 47) and report the *P* values and odds ratios for a random-effects model. We plotted genome-wide results with R v3.0.2 (see URLs) and regional results in LocusZoom v1.3, using LD estimated in PLINK with 1000 Genomes Project EUR data (release 20130502 v.5)[51]. Meta-analysis *P* values were adjusted for multiple testing with the FDR in R (ref. 52).

Pathway analysis of IgAD GWAS results. We conducted GWAS-based pathway analysis on IgAD GWAS meta-analysis *P* values after imputation for a subset of variants (~540,000 variants from the Illumina 550K array) using PARIS (v 1.1.0b)[33]. PARIS groups SNPs into LD features and single SNPs into LE features. The LD features can extend up to 50 kb beyond gene boundaries. PARIS then groups these features by pathway and permutes the genomic structure of the interrogated pathways to determine the significance of the pathway while accounting for differences in LD, gene size, pathway size, and SNP coverage between pathways. The total number of features with a significant *P* value (defined as $P < 0.05$) is compared with the number of significant features in the permuted pathway. The software was tested and validated using genotype data from the Illumina Human660W_Quad_v1 array. We used only a subset of variants because PARIS (v 1.1.0b) has been demonstrated to be robust when used on this SNP array (B. Yaspan, personal communication).

Variants with significant *P* values ($P < 0.05$ in the GWAS meta-analysis data) in Kyoto Encyclopedia of Genes and Genomes (KEGG)[53] pathways (v.3/2011) were identified, and the genome was permuted (10,000 permutations) to estimate pathway significance ($P < 0.0001$). Of 10,000 randomizations, none of the randomized pathways had more variants with $P < 0.05$ than the identified pathways. A follow-up pathway-based analysis was conducted with INRICH (v.1.0) using only pathways that were significant in the PARIS analysis ($P < 0.0001$)[34]. We identified clumps (LD-defined independent intervals) using $P < 0.0001$ as the significance threshold for index SNPs, $P < 0.05$ as the secondary significance threshold for clumped SNPs and $r^2 > 0.5$ as the LD threshold for clumping (PLINK). There were 72 non-overlapping independent genomic intervals of enriched association, based on the subset of meta-analysis *P* values used in the PARIS analysis. Pathways that contained at least two non-MHC genes and had both PARIS $P < 0.0001$ and INRICH $P < 0.05$ were considered statistically significant.

doi:10.1038/ng.3675

Regulatory variation. For each of the five non-MHC IgAD loci reaching genome-wide significance ($P < 5 \times 10^{-8}$), SNPs that were in LD ($r^2 > 0.5$, EUR) with the peak SNP were identified. The peak SNP was defined as the SNP with the lowest *P* value. LD was estimated using whole-genome sequence data (mean of 30× coverage) available for 583 European Americans (PLINK v1.90b20). Significant peak variants, and variants correlated with the peak variants, were queried for known and predicted regulatory DNA elements (RegulomeDB v1.1 and dbSNP141).

Overlap between eQTLs and GWAS signals. The overlap between IgAD meta-analysis *P* values and eQTL data from lymphoblastoid B cells[54,55] was assessed using the Bayesian statistical framework Sherlock (see URLs)[56].

Heritability estimates. The program GCTA-GREML (see URLs) was used to estimate the variance between the cases and controls by the entire genome and the variance explained by the IgAD genome-wide significant variants. The heritability of IgAD for the entire genome was estimated with GCTA in the Swedish IgAD cohort (n = 1,214,325 imputed genotypes)[57]. The other IgAD cohorts were underpowered for heritability estimates (gctaPower). We estimated the heritability of IgAD for the seven peak IgAD variants by including the seven peak variants as covariates. This analysis was then repeated without the MHC region (chr. 6: 24–36 Mb).

Shared genetics with T1D and rheumatoid arthritis. The program Genetic Analysis incorporating Pleiotropy and Annotation (GPA)[30] was used to estimate the extent of the genetic overlap between IgAD and half a million T1D variants[58] and between IgAD and 6 million rheumatoid arthritis variants[17]. Summary statistics for T1D were publicly available from dbGaP (accession phs000180.v3.p2). Summary statistics for rheumatoid arthritis were from the European subset of a recent meta-analysis of rheumatoid arthritis[17] (see URLs). There were 5,966,608 and 501,001 variants in the IgAD–rheumatoid arthritis and IgAD–T1D pairwise analyses, respectively. Using these genome-wide summary statistics, we estimated the proportion of shared variants (variants associated with both IgAD and rheumatoid arthritis or with both IgAD and T1D). We tested for the significance of shared genetics for pairs of diseases using a likelihood-ratio test[30]. GPA estimates the proportion of variants associated with both diseases and uses a likelihood-ratio test to assess statistical significance. This analysis was conducted across the genome and then repeated with the exclusion of the MHC region (chr. 6: 24–36 Mb) because of the extensive LD and highly significant variants in this region. After removing variants in the MHC region, 5,944,048 and 498,320 variants remained in the IgAD–rheumatoid arthritis and IgAD–T1D analyses, respectively.

39. Notarangelo, L.D. *et al.* Primary immunodeficiencies: 2009 update. *J. Allergy Clin. Immunol.* **124**, 1161–1178 (2009).
40. Magnusson, P.K. *et al.* The Swedish Twin Registry: establishment of a biobank and other recent developments. *Twin Res. Hum. Genet.* **16**, 317–329 (2013).
41. Salvi, E. *et al.* Genomewide association study using a high-density single nucleotide polymorphism array and case–control design identifies a novel essential hypertension susceptibility locus in the promoter region of endothelial NO synthase. *Hypertension* **59**, 248–255 (2012).
42. Mitchell, M.K., Gregersen, P.K., Johnson, S., Parsons, R. & Vlahov, D. The New York Cancer Project: rationale, organization, design, and baseline characteristics. *J. Urban Health* **81**, 301–310 (2004).
43. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
44. Price, A.L. *et al.* Long-range LD can confound genome scans in admixed populations. *Am. J. Hum. Genet.* **83**, 132–135, author reply 135–139 (2008).
45. Kosoy, R. *et al.* Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America. *Hum. Mutat.* **30**, 69–78 (2009).
46. Tian, C. *et al.* Analysis and application of European genetic substructure using 300 K SNP information. *PLoS Genet.* **4**, e4 (2008).
47. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
48. Delaneau, O., Zagury, J.F. & Marchini, J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat. Methods* **10**, 5–6 (2013).
49. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G.R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.* **44**, 955–959 (2012).
50. Marchini, J. & Howie, B. Genotype imputation for genome-wide association studies. *Nat. Rev. Genet.* **11**, 499–511 (2010).
51. Pruim, R.J. *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336–2337 (2010).
52. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).
53. Ogata, H. *et al.* KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **27**, 29–34 (1999).
54. Dixon, A.L. *et al.* A genome-wide association study of global gene expression. *Nat. Genet.* **39**, 1202–1207 (2007).
55. Duan, S. *et al.* Genetic architecture of transcript-level variation in humans. *Am. J. Hum. Genet.* **82**, 1101–1113 (2008).
56. He, X. *et al.* Sherlock: detecting gene–disease associations by matching patterns of expression QTL and GWAS. *Am. J. Hum. Genet.* **92**, 667–680 (2013).
57. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
58. Barrett, J.C. *et al.* Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat. Genet.* **41**, 703–707 (2009).