

1 **Supplementary Information**

2

3 **IBD risk loci are enriched in multigenic regulatory modules encompassing**
4 **causative genes.**

5 Momozawa et al.

6 **Supplementary note 1: Genes with strong DAP-EAP correlation**

7 ***IL18R1*** encodes the IL-18r1, the receptor of IL-18, a potent proinflammatory
8 cytokine governing host-microorganism homeostasis and is postulated to play a
9 role in IBD^{1,2}. However, IL-18/IL-18r1 precise contribution to the disease remains
10 controversial. Indeed, compared to wild-type mice, *Il18*^{-/-} and *Il18r1*^{-/-} full KO mice
11 are more susceptible to AOM/DSS-induced colitis and polyp formation³. However,
12 targeted deletion of *Il18*^{-/-} and *Il18r1*^{-/-} in intestinal epithelial cells confers
13 protection from colitis and mucosal damage in mice⁴. In human, several studies
14 have associated circulating or local IL-18 with IBD severity, suggesting that IL-18
15 could be an effector cytokine in IBD⁵.

16 ***IL6ST*** encodes the interleukin 6 signal transducer protein (IL6ST), also called IL6
17 beta, GP130 or CD130. IL6ST is a common transmembrane receptor for all family
18 members of IL6 that include IL-6, IL-11, ciliary neurotrophic factor (CNTF),
19 cardiotrophin-1 (CT-1), cardiotrophin like cytokine (CLC), leukaemia inhibitory
20 factor (LIF), oncostatin M (OSM), neuropoietin (NPN) and interleukin-27 (IL-27)⁶.
21 IL6 family members / IL6ST signaling pathways involve the activation of JAK
22 (Janus kinase) family members, leading to the activation of STAT (signal
23 transducers and activators of transcription) family, as well as the activation of
24 MAPK (mitogen-activated protein kinase) pathway. These pathways are involved
25 in cell survival, apoptosis, differentiation and proliferation⁶. The involvement of
26 IL6/IL6ST/STAT3 in the pathophysiology of IBD is well documented⁷. Indeed,
27 high circulating levels of IL6 is associated with increased severity of the disease⁷.
28 T cells from IBD patient show increased STAT3 activation with increased
29 expression of IL6ST and enhanced resistance to apoptosis⁸. A pilot clinical trial
30 (phase I) targeting of IL6/IL6ST pathway in patients with CD has shown that
31 blocking this pathway has effects similar to the inhibition of TNF^{9,10}.

32 ***THEMIS*** encodes the thymocyte-expressed molecule involved in selection
33 (THEMIS), the expression of which is limited to lymphoid tissues. In mice, THEMIS
34 is highly expressed in pre-TcR thymocytes and plays an important role in T-cell
35 development and TCR activation signaling^{11,12}. Its expression is reduced in
36 differentiated T lymphocytes¹². THEMIS deficiency in mice is associated with the

presence of higher percent of T_{reg} cells, with reduced TCR-mediated T cell response, increased proportion of memory CD4 and CD8 T cells and reduced proportions of naïve-phenotype populations¹². Interestingly, all these T cells associated features are implicated in the pathogenesis of IBD. Indeed, lamina propria T cells in IBD are hypo-responsive to TCR stimulation and high number of effector T cells are present in the inflamed bowel¹³. As for T_{reg}, only moderate expansion was seen in intestinal lesions of Crohn's patients suggesting that their suppressive activity is probably not sufficient against the overwhelming effector T cells activity¹³.

APEH encodes the acylpeptide hydrolase (APEH) enzyme that contributes to protein degradation processes in concert with the proteasome. It catalyzes the removal of *N*-acylated amino acids from acetylated peptides¹⁴. Its physiological role is not well understood. SNPs in APEH gene have been associated with both CD and UC¹⁵. Like other ubiquitin proteasome systems (UPS) such as USP40 or CYLD, APEH may also regulate the NF-κB pathway. Under this scenario, an alteration of NF-κB signaling may lead to aberrant immune response and inflammation.

ANKRD55 encodes an Ankyrin repeat domain-containing protein 55 with unknown function. Ankyrin repeats are composed of 33-34 aa and are the most abundant motifs in nature with highly diverse cellular functions¹⁶. SNPs at the ANKRD55 locus have also been associated with multiple sclerosis¹⁷ and RA¹⁸.

CISD1 gene encodes a highly conserved iron-sulfur domain-containing protein A, known as mitoNEET. This iron-containing protein is a dynamic redox-sensitive molecule that serves an important role in mitochondrial functions. It participates in critical process such as electron shuttling through the electron transport chain, regulation of enzymatic activity, and synthesis of heme and iron-sulfur clusters^{19,20}. Deregulation of iron metabolism and associated anemia has been associated with IBD²¹. The role that mitoNEET plays in the etiology of IBD remains to be determined.

CPEB4 gene encodes the cytoplasmic polyadenylation element-binding protein 4 (CEBP4), which belongs to a family of proteins that bind mRNAs and contain a cytoplasmic polyadenylation element (CPE) in their 3'-UTR. Binding results in 3'-

poly(A) tail extension and translational upregulation of target mRNAs. *Cpeb4* mRNA is rhythmically regulated in mouse liver, conferring temporal translational regulation. In the absence of CPEB4, a large number of mRNAs are transcribed, but remain untranslated until needed²². A recent study, using knockout mice models, showed that CPEB4 was required for translation of numerous proteins involved in ER homeostasis and CPEB4 loss resulted in mitochondrial dysfunction and defective lipid metabolism, two hallmarks of ER stress. *Cpeb4* KO livers were highly susceptible to ER stress-induced apoptosis and to development of NAFLD²³. In CD, reduced CPEB4 may also lead to ER stress and mitochondrial dysfunction.

DOCK7 encodes dedicator of cytokinesis 7 protein (Dock7), a member of Dock proteins family and an activator of Rac GTPases. DOCK7 plays an important role in axon outgrowth, Schwann cell migration, and axon myelination²⁴. Mutation in this gene in mice leads to hypopigmentation suggesting a non-redundant role in the distribution and function of dermal and follicular melanocytes. However, mutant mice show normal neuronal function despite the high expression of DOCK7 in the developing brain, suggesting redundancy with other Docks²⁵. The role of DOCK7 in IBD and immune cells function is totally unknown.

ERAP2 gene encodes an endoplasmic reticulum aminopeptidase (ERAP2), an enzyme involved in trimming of peptides for MHC-I loading. Aberrant ERAP2 function could influence peptide-HLA-B27 stability, formation of MHC-I free heavy chains and ER stress^{26,27,28}. SNPs in *ERAP2* gene have been associated with CD²⁹. Although the underlying mechanisms are not known, it is possible that ERAP2 modification contributes to the reported reduction of MHCI on CD4 T cells from CD patients³⁰. ERAP2 modification may also contribute to the epithelial ER stress associated with CD and UC.

GNA12 encodes Guanine nucleotide-binding protein subunit alpha-12 or $G\alpha_{12}$, which belongs to the heterotrimeric G proteins. $G\alpha_{12}$ is found in tight junctions (TJ) where it interacts with ZO-1³¹ and plays important roles in para-cellular permeability^{32,33}. $G\alpha_{12}$ is ubiquitously expressed and interacts, upon receptor-mediated activation, with certain Rho guanine nucleotide exchange factors (RhoGEFs) which in turn mediate activation of the small GTPase RhoA³⁴. Intestinal

permeability and barrier dysfunction is a hallmark of CD and UC. Several studies reported changes in the expression of several TJ proteins in both diseases³⁵. It is conceivable that modifications in the α_{12} pool leads to alteration of intestinal permeability. Tissue-specific α_{12} -deficient mice revealed important functions of this protein in modulating T cell trafficking and proliferation, as well as in the response to foreign and self antigens³⁶, important processes that may affect susceptibility for T cell-mediated diseases.

GPX1 encodes the glutathione peroxidase 1 (GPX1), a highly abundant and ubiquitously expressed cytosolic enzyme. Like all glutathione peroxidases family members, GPX1 catalyzes the reduction of H₂O₂ by glutathione and consequently, protects cells from oxidative damage. In IBD, it is believed that intestinal and colonic injuries and dysfunction is at least partially due to elevation of reactive metabolites of oxygen and nitrogen³⁷. Although the role of GPX1 is not known in IBD, deficiency of both GPX1 and GPX2 in mice lead to spontaneous ileo-colitis and intestinal cancer³⁸. A protective role of GPX1 and GPX2 against oxidative stress has also been suggested by studies reporting elevated *Gpx1/2* gene expression in gastric mucosa after *H. pylori* infection³⁹. Association of the elevated expression of *Gpx1/2* gene with tumorigenesis could be due to its anti-apoptotic activity⁴⁰.

GSDMB encodes Gasdermin-B protein (GSDMB) the function of which is largely unknown. The expression of *GSDMB* has been associated with differentiated epithelial cells and with regions containing proliferating cells or stem cells, respectively, of the esophagus and the gastric mucosa^{41,42}.

JAZF1, also known as *TIP27*, encodes a transcriptional repressor of *NR2C2*, also known as *TAK1* or *TR4*⁷⁶. Mice deficient in *NR2C2* show low IGF1 serum concentrations and perinatal and early postnatal hypoglycemia, as well as growth retardation⁷⁷. *JAZF1* also affects variation in human height⁷⁸. SNPs in *JAZ1F* have been associated with type II diabetes⁷⁹, prostate⁸⁰ and endometrial cancer⁸¹ and with systemic lupus erythematosus⁸². However, the role of *JAZF1* in immune response and autoimmunity remains to be elucidated.

LSP1 encodes a leukocyte-specific protein 1 (LSP1), a Ca²⁺-activated, intracellular

filamentous actin-binding protein that interacts with the cytoskeleton and is expressed in hematopoietic lineage and in endothelial cells⁷⁰. Evidence from mice model studies suggest that LSP1 plays a negative regulatory role on neutrophil and T cell migration^{71,72}. A recent study identified a novel *LSP1* deletion variant for RA susceptibility through CNV GWAS⁷³. The copy number of *LSP1* was found to be significantly lower in RA patients and was associated with increased T cell migration⁷³. We found a positive correlation of LSP1 expression (in CD14⁺ cells) with UC, but not with CD. UC, as well as CD, is characterized by an increased infiltration of immune cells in inflamed tissues. Our finding is therefore surprising if we consider the concept of an association between increased cell migration with *LSP1* CNVs and LSP1 insufficiency. It is possible that LSP1 plays an additional, yet unknown role in monocytes. On the other hand, if LSP1 participates actively in the cross-talk between leukocytes and endothelial cells during leukocyte transmigration, the physiological differences in microvasculature and the integrins involved may dictate organ-specific roles for LSP1 in leukocyte recruitment into the inflammatory sites.

NXPE1: Encodes Neuroexophilin and PC-esterase domain family member 1 (*NXPE1*). A human gastrointestinal tract (GIT) specific transcriptome and proteome study validate the expression pattern of this gene and protein in the intestine⁷⁴. *NXPE1* was recently identified as a novel target gene for IBD-associated variants⁷⁵. Its function remains largely unknown.

ORMDL3 encodes ORM1-like protein 3, a negative regulator of sphingolipid synthesis and a regulator of endoplasmic reticulum-mediated calcium signaling⁴⁵. *ORMDL3* is involved in the regulation of eosinophil and T cell functions^{46,47}. It also facilitate B cells survival and regulates autophagy through the ATF6 signaling pathway⁴⁸. Genetic variants regulating *ORMDL3* expression have been associated with susceptibility to asthma⁴⁹, T1D⁵⁰, atherosclerosis⁵¹, ankylosing spondylitis⁵² and IBD⁵³. *ORMDL3* might be associated with IBDs and other autoimmune and inflammatory diseases by activating ERS, inducing autophagy and/or promoting immune cells activation.

REXO2 encodes an oligoribonuclease protein. Its depletion, using RNAi, causes a significant decrease of mtDNA and mtRNA and impaired *de novo* mitochondrial protein synthesis⁸³. REXO2's function remains unknown but it may be involved in the well documented mitochondrial defects associated with IBD⁸⁴.

RNASET2 is the only RNase T2 family member in humans and is potentially involved in the inhibition of tumorigenesis, metastasis and angiogenesis^{85,86}. Loss-of-function of RNASET2 protects fibroblasts from oxidative stress⁸⁹ while its overexpression in melanocytes and keratinocytes sensitizes these cells to oxidative-stress-induced apoptosis⁹⁰. Interestingly, CD is characterized by an impaired immune cells apoptosis associated with elevated H₂O₂ in PBMC during the active phase of the disease⁹¹. Although speculative, it is possible that reduced **RNASET2** contributes to the altered oxidative stress in CD.

SKAP2 encodes the Src kinase-associated phosphoprotein 2 (Skap2), a cytosolic adaptor protein expressed in a variety of cell types including hematopoietic cells^{54,55,56}. Skap2 has been implicated in cell adhesion through association to integrins and cytoplasmic actin⁵⁵, and is required for global actin reorganization. It interacts with different molecules implicated in integrin signaling events^{54,56,57}. Loss of Skap2 in mice results in reduced inflammation in experimental autoimmune encephalomyelitis as well as defects in macrophage migration into tumor metastasis, suggesting a physiologically important role of Skap2 for leukocyte recruitment *in vivo*^{55,58}.

UBE2L3 gene encodes an atypical Ubiquitin E2 Conjugase (UBE2L3) the role of which has been recently uncovered. It is an indirect human and mouse Caspase-1 target and plays an important role in the maturation of IL-1 β . UBE2L3 depletion in mice increases pro-IL-1 β levels and mature-IL-1 β secretion by inflammasomes⁶¹. Several GWAS identified polymorphisms in the genomic locus of **UBE2L3** that are associated with multiple autoimmune diseases⁶² including CD²⁹. Decreased secretion of the inflammasome cytokine IL-1 β was noted in monocytes of Crohn's disease patients⁶³. It is therefore tempting to speculate that **UBE2L3** contributes to disease at least partially by modulating IL-1 β secretion.

ZMIZ1 encodes Zmiz1, a member of the protein inhibitor of activated STAT (PIAS)-like family of coregulators⁶⁴. Zmiz1 is widely and variably expressed⁶⁵. In GWAS, a SNP within ZMIZ1 gene was associated with early-onset Crohn's disease and IBD⁶⁶. *ZMIZ1* is co-expressed with activated *NOTCH1* across a broad range of T-ALL oncogenic subgroups. Its inhibition slows human T-ALL cell proliferation and/or sensitizes them to γ -Secretase inhibitors (GSI)⁶⁷. Evidence from Zmiz1-deficient mice demonstrated that Zmiz1 is a direct Notch1 cofactor that heterogeneously regulates Notch1 target genes and plays an important role in T cells development⁶⁸. Altered expression of *ZMIZ1* has been reported to affect Smad3-mediated transcription⁶⁹. Interestingly, our analysis shows that increased UC disease risk was associated with decrease of both *SMAD3* and *ZMIZ1* expression while no association was observed with *NOTCH1*. This association was observed in different tissues/cell types suggesting a possible trans effect of *ZMIZ1* on *SMAD3* expression.

Supplementary Table 1

Cell type	Naive (r^2 based)			Frequentist (Nica et al., 2010)			Theta-based		
	Overlaps observed	Overlaps expected	P value	Overlaps observed	Overlaps expected	P value	Overlaps observed	Overlaps expected	P value
CD4	12	3.3	< 0.01	14	4.9	< 0.01	17	8.4	< 0.01
CD8	12	3.5	< 0.01	18	4.3	< 0.01	16	6.9	< 0.01
CD14	8	3.3	0.061	9	4.7	0.211	10	7.1	0.720
CD15	4	1.9	0.646	4	2	0.720	7	5.1	0.909
CD19	7	2	0.010	7	3.6	0.410	12	5.8	0.044
PLA	4	0.9	0.010	3	0.9	0.475	5	1.8	0.119
IL	4	1.6	0.432	7	2.1	0.027	8	4.1	0.281
TR	6	2.6	0.211	5	3.5	0.928	11	6	0.086
RE	5	1.5	0.103	6	2.4	0.204	9	5.5	0.509

Enrichment of DAP-EAP matching in 63 of 97 CD risk loci covered by the Immunochip. For each cell type, we provide the number of matches (or overlaps) observed with the top disease-associated SNPs ($MAF > 0.05$), as well as the number of matches expected with the same number of SNPs ($MAF > 0.05$) sampled at random in the same 63 risk loci. The analyses were conducted using three "colocalisation" methods (Naive, Frequentists and Theta-based). The p-values were determined by simulation (1,000 sets of 63 randomly sampled SNPs) and Bonferroni corrected for the analysis of 9 cell types. < 0.01 means that the number of matches observed with the real disease-associated SNPs was never observed with any set (out of 1,000) of randomly sampled SNPs.

Supplementary Table 2

Tissue	Nr of samples	Nr of probes	Nr of PCs
CD4	303	13,466	38
CD8	294	13,317	35
CD19	282	12,648	40
CD14	286	13,170	36
CD15	289	11,069	27
PLA	251	6,565	23
IL	200	15,401	59
TR	271	15,082	50
RE	267	14,844	53

Number of usable samples, probes and PC for each tissue type.

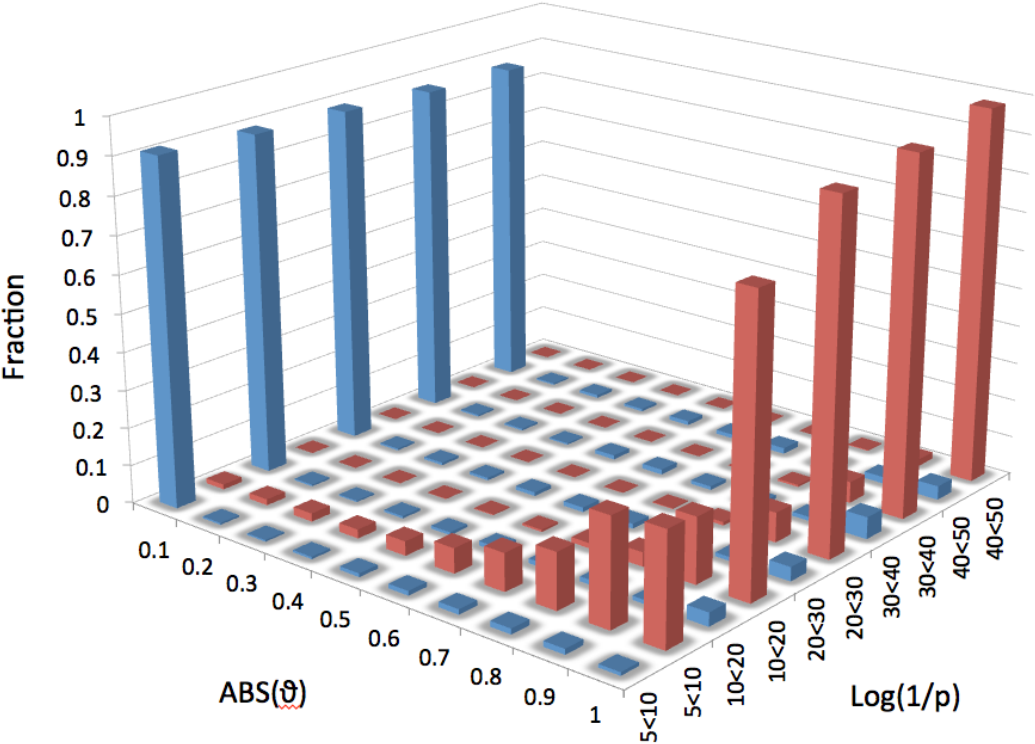
Supplementary Table 3

Tissue	Nr of probes	FDR≤0.25	FDR≤0.10	FDR≤0.05	FDR≤0.01
CD4	13,466	7,417	4,957	4,176	3,247
CD8	13,317	6,760	4,309	3,599	2,779
CD19	12,648	4,984	3,138	2,549	1,953
CD14	13,170	7,118	4,728	3,961	3,106
CD15	11,069	3,611	2,396	1,983	1,512
PLA	6,565	1,404	996	854	653
IL	15,401	2,769	1,728	1,426	1,031
TR	15,082	5,183	3,391	2,807	2,160
RE	14,844	4,180	2,726	2,295	1,731

Number of cis-eQTL found in the nine analyzed cell types for different FDR thresholds (see also Suppl. Figure 7).

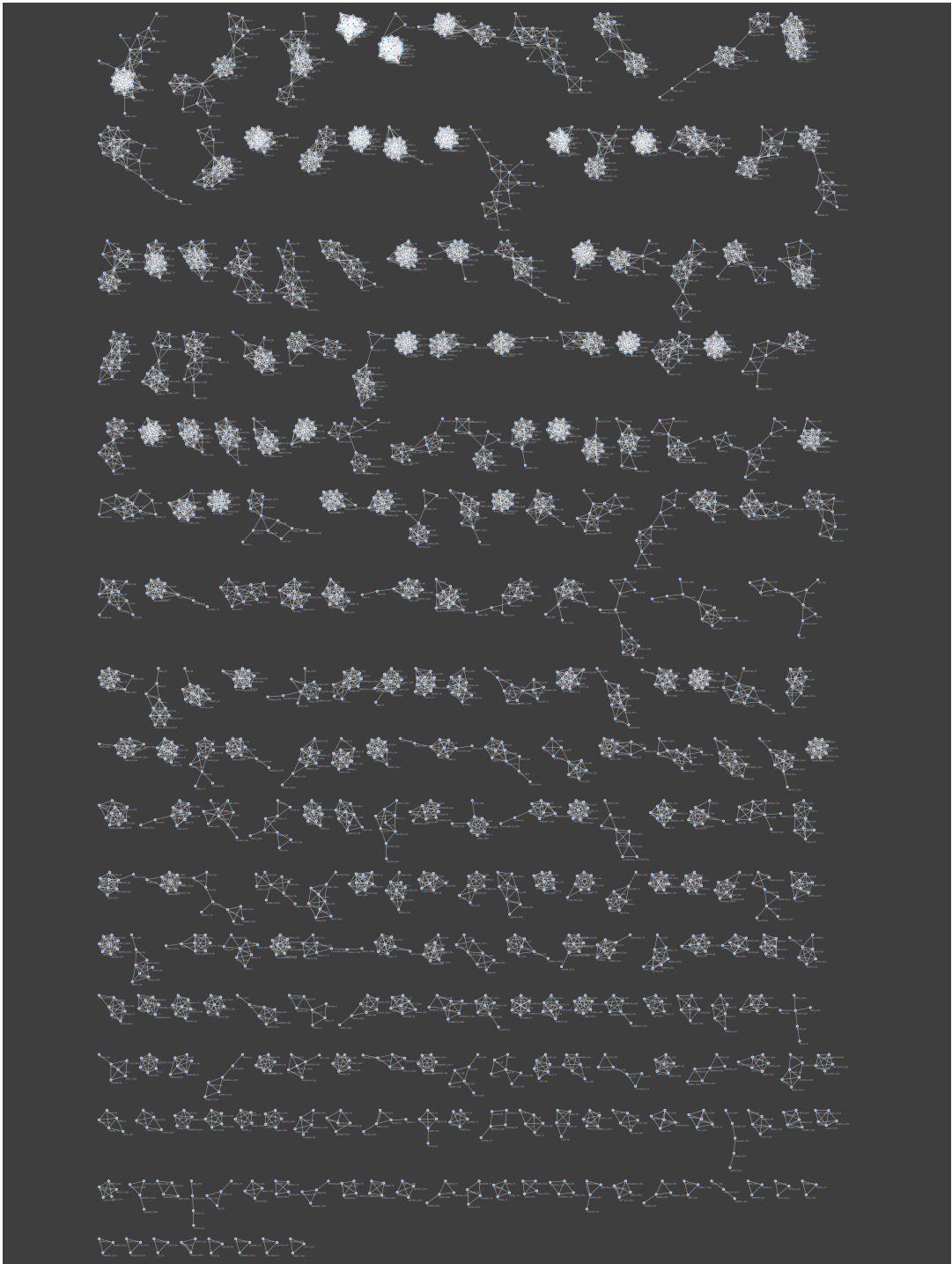
Supplementary Figures

1. Supplementary Figure 1



Absolute values of θ for pairs of eQTL driven by distinct regulatory variants (blue), and for pairs of eQTL driven by the same regulatory variants (red). The first (blue) were obtained by confronting real cis-eQTL with in silico simulated eQTL explaining the same variance as the real eQTL but driven by a randomly chosen SNPs in a 2Mb window centered around the probe. The second (red) were obtained by confronting eQTL obtained by reanalyzing two mutually exclusive halves of the CEDAR population separately in a region harboring a real cis-eQTL. It can be seen that θ very effectively discriminates between pairs of eQTL driven by distinct (blue) vs the same (red) regulatory variants. By choosing 0.6 as threshold value for θ , one captures most red pairs (~88%) with minimum contamination of blue pairs (~5%). Log(1/p): eQTL are sorted by the smallest log(1/p) value of the two eQTL being compared.

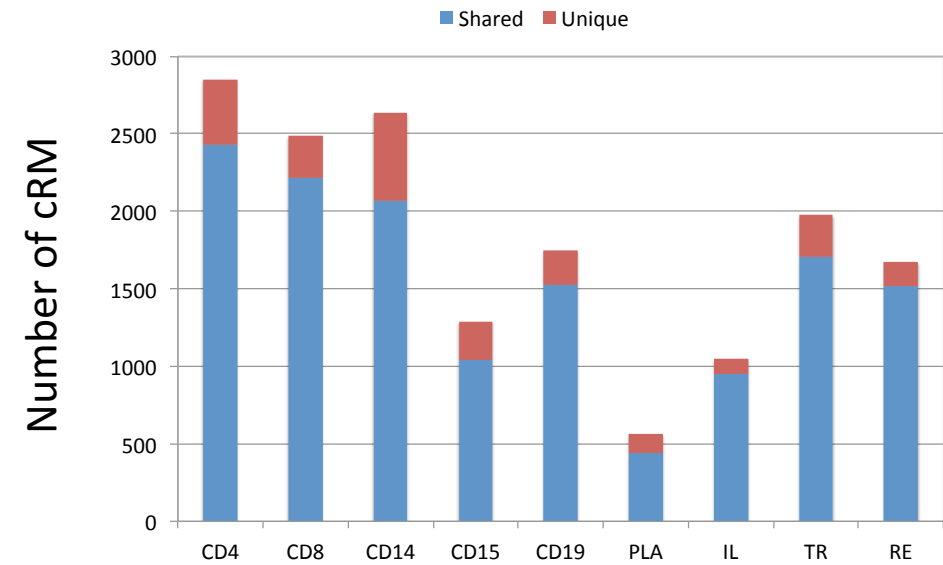
2. **Supplementary Figure 2**



Graphical representation (using Cytoscape¹) of 269 cis acting regulatory modules (cRM) including at least three genes (see Suppl. Table 2). Every node corresponds to a cis-eQTL involving a specific gene-tissue combination. Edges connect pairs of cis-eQTL for which $|var| \geq 0.6$.

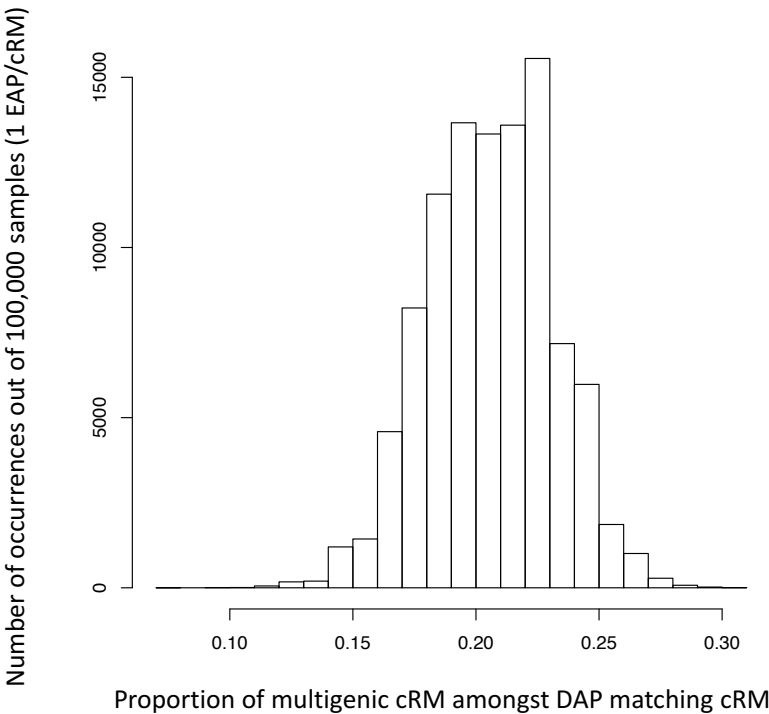
1. Shannon, P. et al. *Genome Res.* 13, 2498-2504 (2003).

3. Supplementary Figure 3



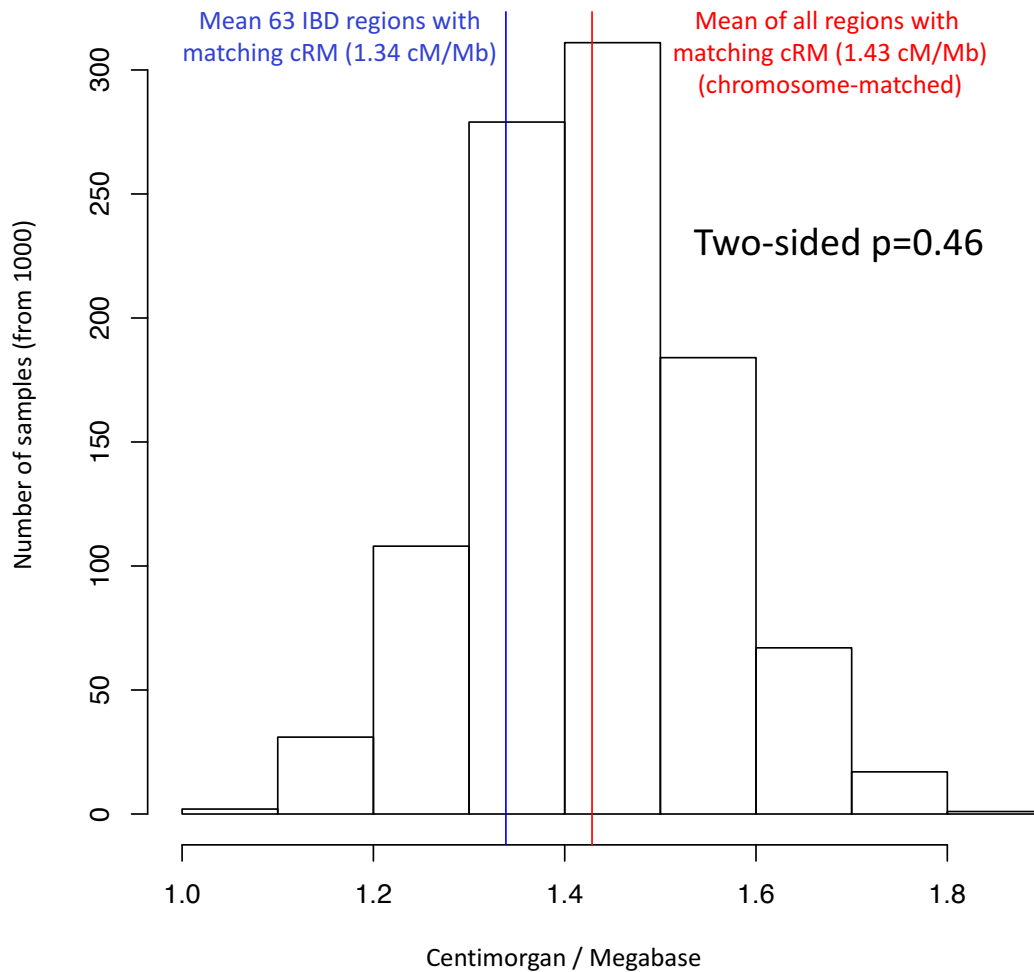
Number of cRM detected in each cell type. Blue: shared cRM (i.e. also detected in at least one other cell type). Red: Unique (i.e. detected only in that cell type).

4. Supplementary Figure 4



Across the entire genome, the proportion of multigenic cRM was shown to be 0.10 (see also main text, figure 1B). Amongst DAP matching cRM (mapping to 63 of 200 studied IBD risk loci; main text Table 1) this proportion was shown to be 0.33, hence a highly significant enrichment. To ensure that this enrichment was not only due to the fact that matching between DAP and EAP was de facto tested multiple times for multigenic cRM and only once for other cRM, we only tested one randomly sampled EAP per cRM (whether monogenic or multigenic). This was repeated 100,000 times and yielded the distribution of the proportion of multigenic cRM amongst DAP matching cRM shown above. The average was 0.22, and we never observed values ≤ 0.11 , i.e. the genome-wide average.

271 5. Supplementary Figure 5

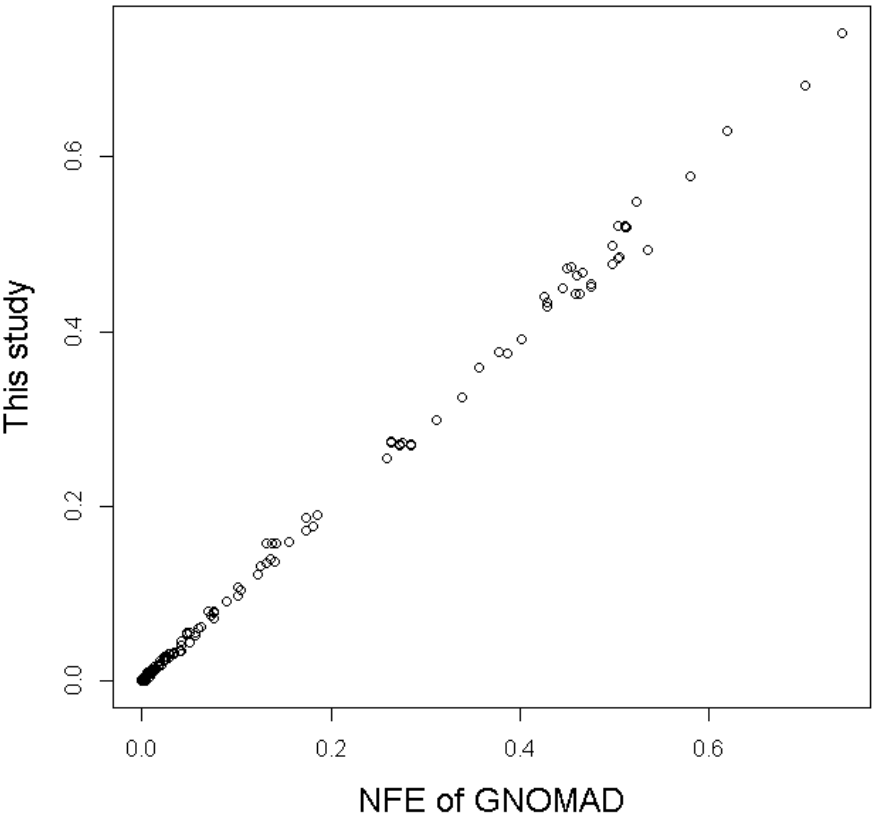


272

273

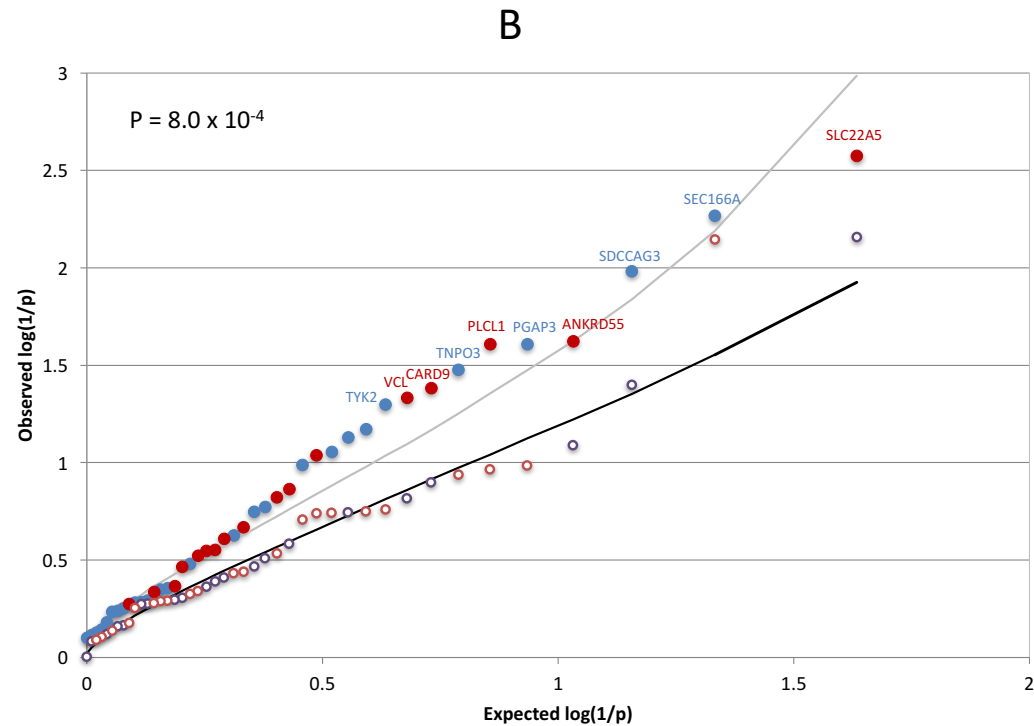
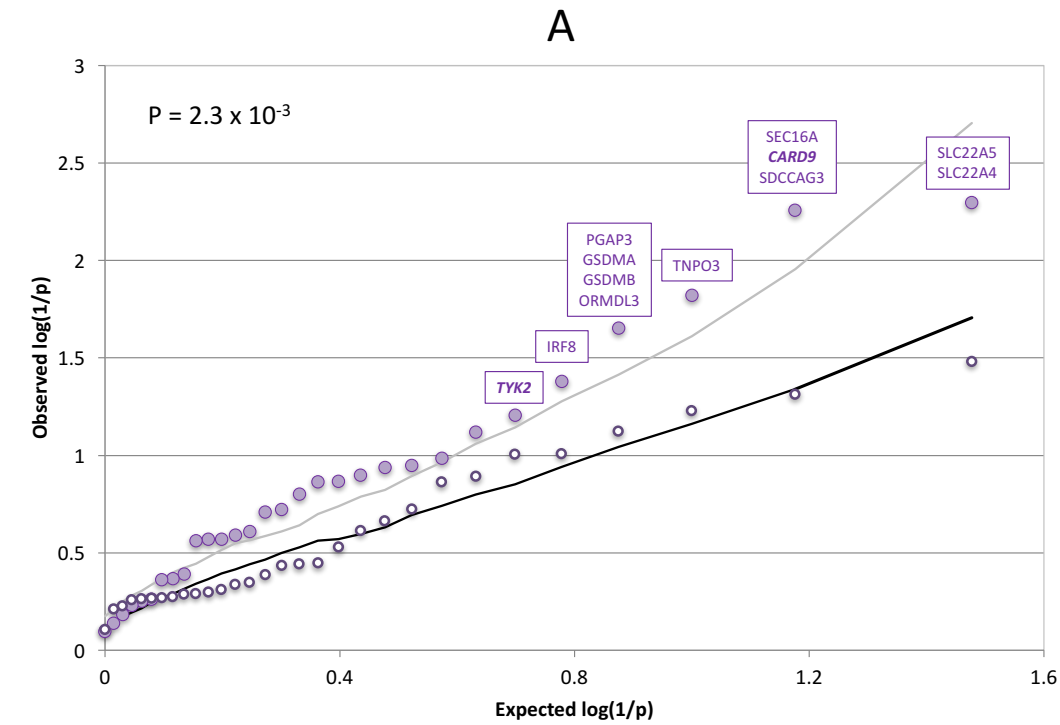
274 The 63 IBD risk loci with matching cRM are 2- to 3-fold enriched in multigenic
 275 cRM ($p \leq 10^{-5}$). This could be due to the fact that the LD is higher in IBD
 276 regions than in the rest of the genome. To test this, we downloaded LD-based
 277 recombination maps of the human genome from
 278 <https://github.com/joepickrell/1000-genomes-genetic-maps>. The average
 279 recombination rate across the human genome was 1.23 centimorgan per
 280 megabase (cM/Mb). The average recombination rate for the 63 IBD risk loci
 281 with matching cRM was 1.34 cM/Mb, i.e. less LD than in the rest of the genome.
 282 Regions encompassing eQTL (and hence cRM) may differ from the rest of the
 283 genome with regards to LD. Thus, we further sampled 1,000 sets of 63 loci
 284 centered on cRM (from our list of 9,720) that were matched for size and
 285 chromosomal location with the 63 cRM-matching IBD risk loci. The mean
 286 recombination rate for the cRM-centered genome was 1.43 cM/Mb. The figure
 287 shows the frequency distribution of the corresponding mean cRM/Mb per set
 288 (black), the mean of means of the 1,000 sets of 63 randomly drawn loci (red),
 289 and the mean of the 3 IBD risk loci (blue). The mean of the 63 IBD risk loci did
 290 not differ significantly from the rest of the cRM centered portion of the genome
 291 (two-tailed p -value: 0.46).

6. Supplemental Figure 6

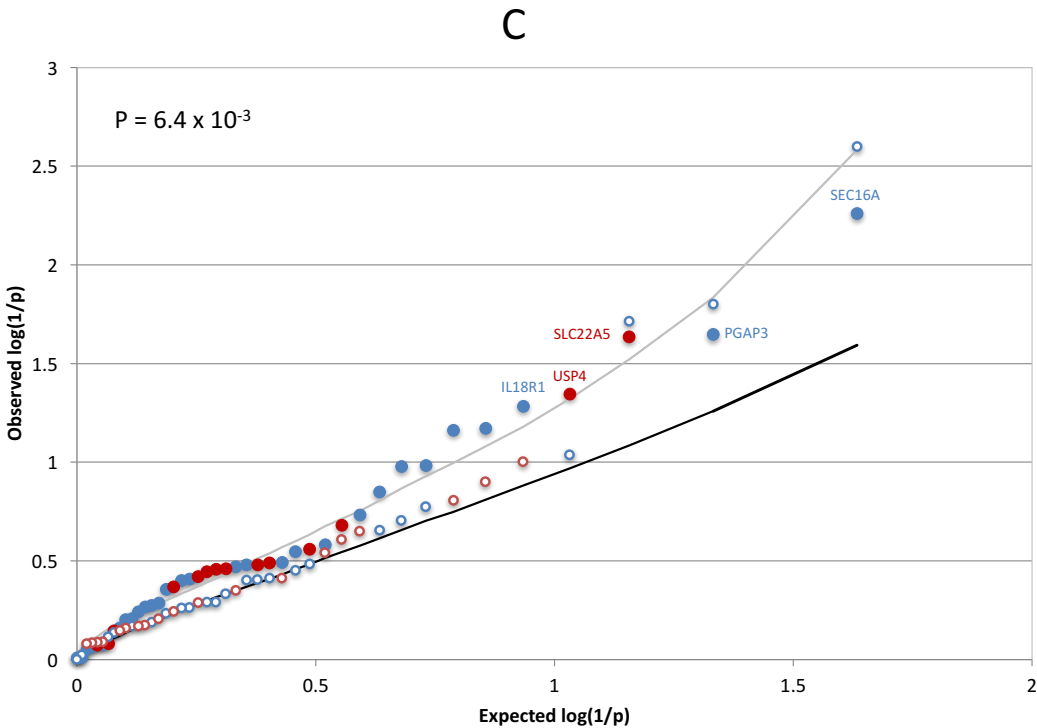


Comparison of the alternative allele frequency for 1,781 variants observed in this study and in 55,860 non-Finnish European samples from the GNOMAD study.

7. Supplementary Figure 7



302

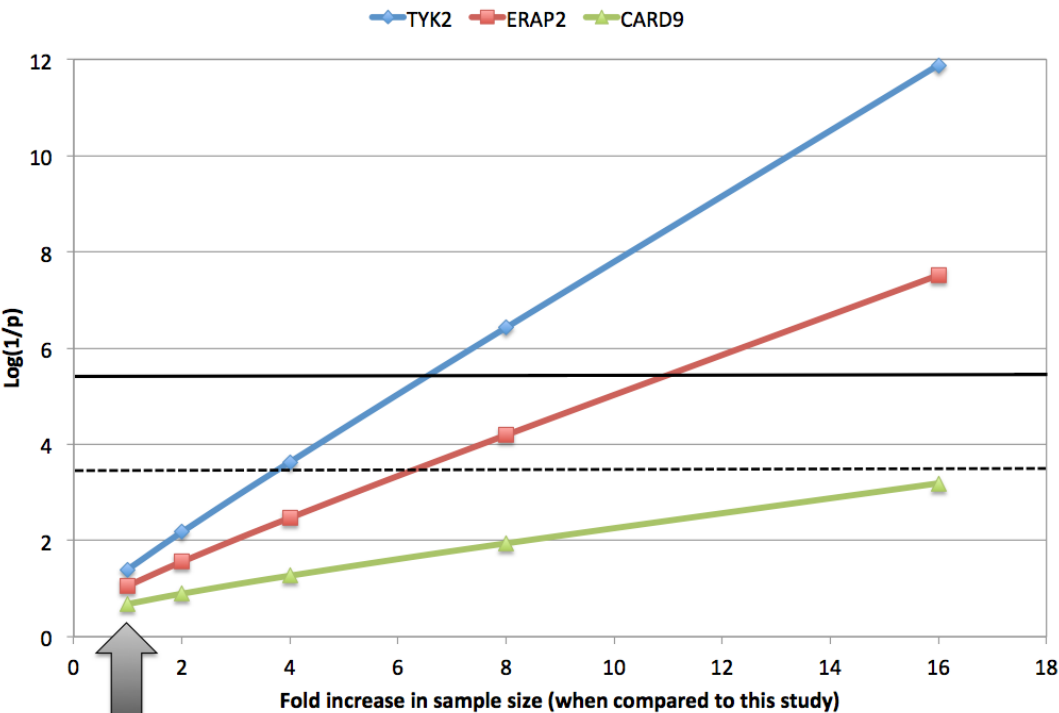


303

304 QQ-plot for the module-based burden test (A), disease plus age-of-onset-based
305 burden test (B), and disease plus familiarity-based burden test (C). Ranked
306 $\log(1/p)$ values obtained when considering LoF and damaging variants (full
307 circles), or synonymous variants (empty circles). The circles are labeled in blue
308 when the best p-value for that gene is obtained with CAST, in red when the best
309 p-value is obtained with SKAT, or in purple for the module-based test (as some
310 genes in the module may have their best p-value with CAST and other with
311 SKAT). The black line corresponds to the median $\log(1/p)$ value obtained (for
312 the corresponding rank) using the same approach on permuted data (LoF and
313 damaging variants). The grey line marks the upper limit of the 95% confidence
314 band. The name of the genes/modules exceeding the nominal p-value of 0.05
315 are given. The inset p-values correspond to the significance of the upwards
316 shift in $\log(1/p)$ values estimated by permutation.

317

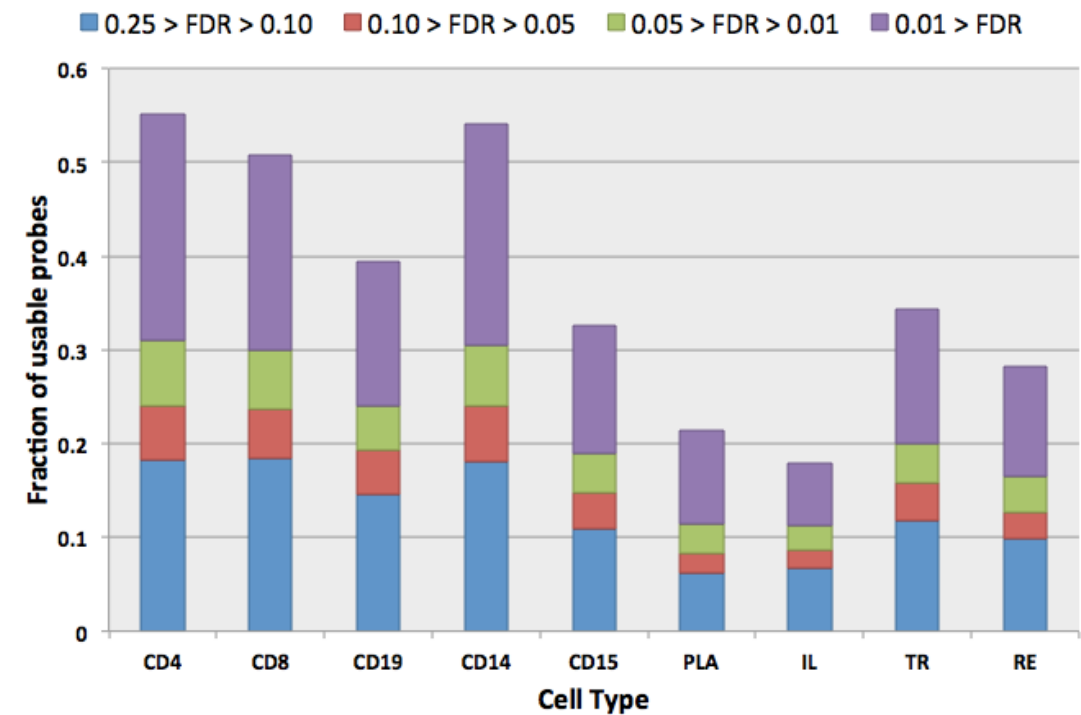
8. Supplementary Figure 8



Present study: 6,600 CD cases and 5,500 controls

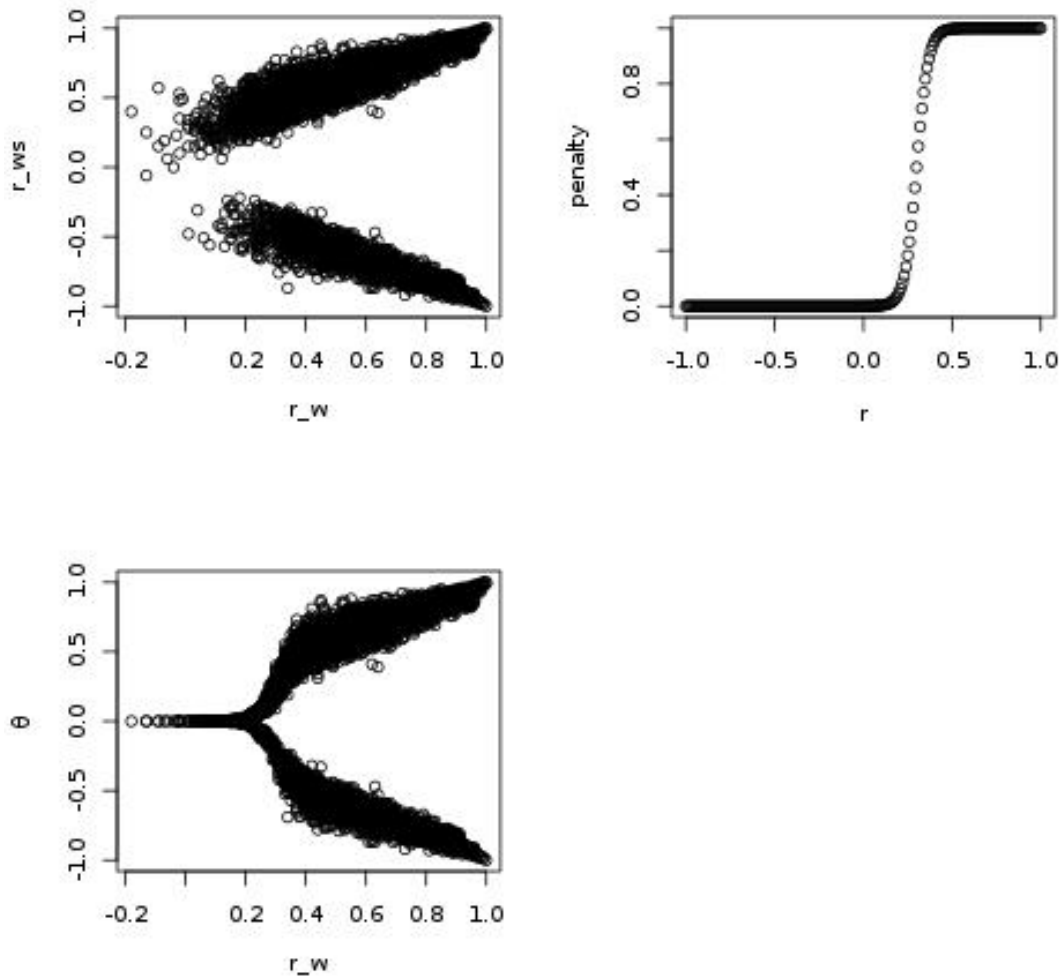
Effect of increasing sample size on the $\log(1/p)$ values of a one-sided burden test assuming that the effects observed for *TYK2* (blue), *ERAP2* (red) and *CARD9* (green) observed in this study are real unbiased. The dotted horizontal black line corresponds to an hypothetical experiment-wide significance threshold assuming the realization of 200 independent tests (targeting for instance 100-200 genes selected on the basis of coincident DAP-EAP patterns). The plain horizontal black line corresponds to an hypothetical genome-wide significance threshold assuming the realization of 20,000 independent tests (targeting all genes). It can be seen that an at least 4-fold increase in sample size is needed to achieve significance in the first scenario and at least 7-fold increase in the second scenario.

9. Supplementary Figure 9



Proportion of usable probes with cis-eQTL at various levels of FDR in the nine analyzed cell types.

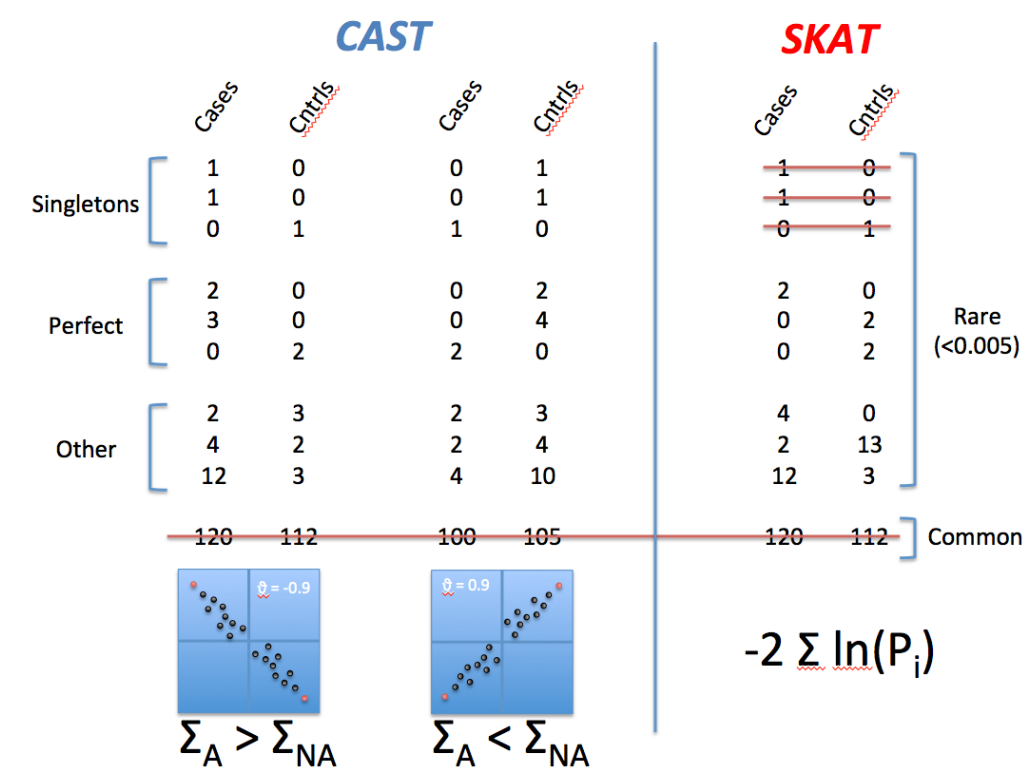
10. Supplementary Figure 10



Graphical illustration of the relationship between r_w , r_{ws} and ϑ . The penalty function applied to r_{ws} to generate ϑ , corresponds to $\frac{1}{1+e^{-k(r_w-T)}}$. The graph is shown for $k = 30$ and $T=0.3$, the values used in this study.

The point here is that if two association patterns are “similar” (driven by the same variants), the correlation (r_w in Suppl. Methods) between $-\log(1/p)$ values is expected to be positive. If two association patterns are different (driven by distinct variants) they may generate strong negative correlations (r_w). The first part of the method aims at weeding out such instances (negative r_w). One way to do this is to choose a simple threshold value for r_w . We herein propose an approach that offers more flexibility: it generates a penalty that increases when the correlation decreases with an adaptable rate. As shown in Suppl. Fig. 8, the values of $k=30$ and $T=0.3$ essentially correspond to a threshold value of 0.3. As can also be seen from Suppl. Fig. 8, there is (as expected) a strong linear relationship with slope 1 between r_w and $|r_{ws}|$ (and hence between r_w and $|\vartheta|$ for pairs with $r_w > 0.3$). Because we subsequently use a threshold value $|\vartheta| \geq 0.6$, the choice T has very little impact on the outcome unless one approaches 0.6.

11. Supplementary Figure 11

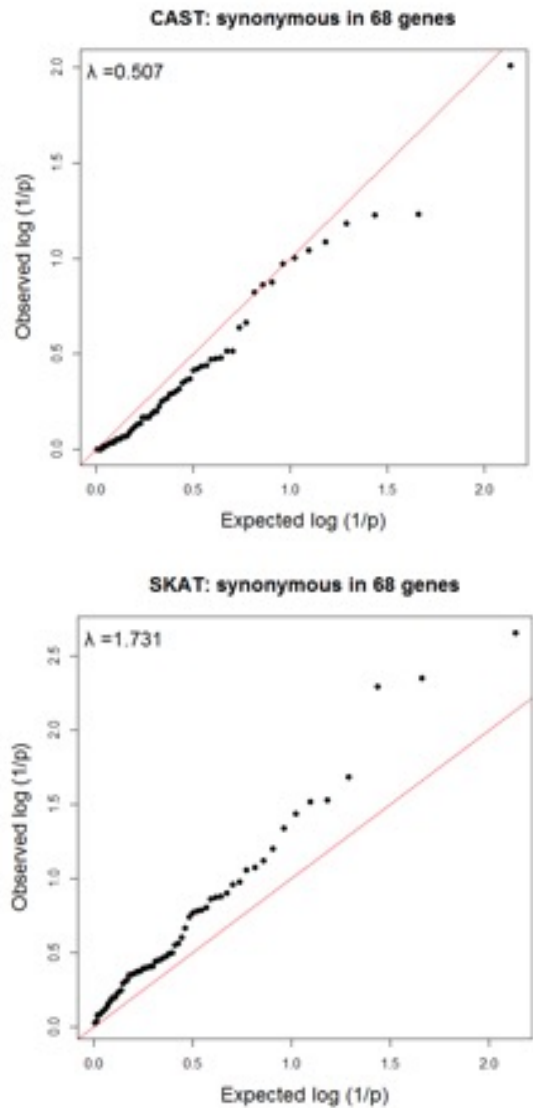


Schematic representation of the key features of the implemented “burden test”.

The analysis is restricted to rare variants with MAF < 0.005 to ensure that the new signal is independent of the one that lead to the identification of the corresponding risk loci by GWAS (based on common and low frequency variants). Variants can be sorted in (i) singletons (i.e. observed only ones in the analyzed samples), (ii) perfect (i.e. observed more than ones in the sample but perfectly associated with disease status), and (iii) other (i.e. observed more than ones in the sample in both cases and controls).

We test two hypotheses. The first assumes that disruptive variants are either enriched in cases or in controls as a function of the sign of the correlation between DAP and EAP (if decreased expression is associated with increased risk, disruptive “risk” variants are expected to be enriched in cases; if increased expression is associated with increased risk, disruptive “protective” variants are expected to be enriched in controls). The test is implemented with CAST and in essence performs a one-sided test of independence (what is the probability to observe the excess of disruptive variants in cases (respectively controls) by chance alone?). The second hypothesis tests whether the distribution of the variants in cases and controls is characterized by too many variants that tend to be overrepresented either in cases or in controls. Thus, this hypothesis allows some disruptive variants to increase risk and others to be protective. This hypothesis does not use information from singletons. Testing this hypothesis is implemented with SKAT. It can be seen in simplified form as combining the p-values (from a test of independence) across variants (without considering the sign of the effect) using for instance Fisher’s method.

12. Supplementary Figure 12



Distribution of permutation-based $-\log(p)$ values obtained for 68 analyzed genes with synonymous variants using CAST (A), and SKAT (B), indicating that CAST is conservative ($\lambda_{GC} = 0.507$), while SKAT is too permissive ($\lambda_{GC} = 1.73$). The 68 genes correspond to the 47 genes reported in this study, plus 21 genes sequenced in the same cohort as part of another study.

410

- 411 1. Dinarello, C. A., Novick, D., Kim, S. & Kaplanski, G. Interleukin-18 and IL-18
412 binding protein. *Frontiers in Immunology* **4**, (2013).
- 413 2. Elinav, E. *et al.* NLRP6 inflammasome regulates colonic microbial ecology and
414 risk for colitis. *Cell* **145**, 745–757 (2011).
- 415 3. Salcedo, R. *et al.* MyD88-mediated signaling prevents development of
416 adenocarcinomas of the colon: role of interleukin 18. *J. Exp. Med.* **207**, 1625–36
417 (2010).
- 418 4. Nowarski, R. *et al.* Epithelial IL-18 Equilibrium Controls Barrier Function in
419 Colitis. *Cell* **163**, 1444–1456 (2015).
- 420 5. Dinarello, C. A. Interleukin-18 and the Pathogenesis of Inflammatory Diseases.
421 *Semin. Nephrol.* **27**, 98–114 (2007).
- 422 6. Heinrich, P. C. *et al.* Principles of interleukin (IL)-6-type cytokine signalling and
423 its regulation. *Biochem. J.* **374**, 1–20 (2003).
- 424 7. Mitsuyama, K. *et al.* Therapeutic strategies for targeting the IL-6/STAT3
425 cytokine signaling pathway in inflammatory bowel disease. *Anticancer Res.* **27**,
426 3749–56
- 427 8. Atreya, R. *et al.* Blockade of interleukin 6 trans signaling suppresses T-cell
428 resistance against apoptosis in chronic intestinal inflammation: evidence in
429 Crohn disease and experimental colitis in vivo. *Nat. Med.* **6**, 583–588 (2000).
- 430 9. Ito, H. *et al.* A Pilot Randomized Trial of a Human Anti-Interleukin-6 Receptor
431 Monoclonal Antibody in Active Crohn's Disease. *Gastroenterology* **126**, 989–
432 996 (2004).
- 433 10. Jones, S. A., Scheller, J. & Rose-John, S. Therapeutic strategies for the clinical
434 blockade of IL-6/gp130 signaling. *J. Clin. Invest.* **121**, 3375–3383 (2011).
- 435 11. Lesourne, R. *et al.* Themis, a T cell-specific protein important for late thymocyte
436 development. *Nat. Immunol.* **10**, 840–7 (2009).
- 437 12. Fu, G. *et al.* Themis controls thymocyte selection through regulation of T cell
438 antigen receptor-mediated signaling. *Nat. Immunol.* **10**, 848–56 (2009).
- 439 13. Neurath, M. F. Cytokines in inflammatory bowel disease. *Nat. Rev. Immunol.* **14**,

- 440 329–342 (2014).
- 441 14. Palumbo, R. *et al.* APEH Inhibition Affects Osteosarcoma Cell Viability via
442 Downregulation of the Proteasome. *Int. J. Mol. Sci.* **17**, 1614 (2016).
- 443 15. Cleynen, I. *et al.* Genetic and microbial factors modulating the ubiquitin
444 proteasome system in inflammatory bowel disease. *Gut* **63**, 1265–74 (2014).
- 445 16. Li, J., Mahajan, A. & Tsai, M.-D. Ankyrin Repeat: A Unique Motif Mediating
446 Protein–Protein Interactions †. *Biochemistry* **45**, 15168–15178 (2006).
- 447 17. International Multiple Sclerosis Genetics Consortium (IMSGC) *et al.* Analysis
448 of immune-related loci identifies 48 new susceptibility variants for multiple
449 sclerosis. *Nat. Genet.* **45**, 1353–60 (2013).
- 450 18. Viatte, S. *et al.* Genetic markers of rheumatoid arthritis susceptibility in anti-
451 citrullinated peptide antibody negative patients. *Ann. Rheum. Dis.* **71**, 1984–90
452 (2012).
- 453 19. Wiley, S. E., Murphy, A. N., Ross, S. A., van der Geer, P. & Dixon, J. E.
454 MitoNEET is an iron-containing outer mitochondrial membrane protein that
455 regulates oxidative capacity. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 5318–5323
456 (2007).
- 457 20. Geldenhuys, W. J., Leeper, T. C. & Carroll, R. T. mitoNEET as a novel drug
458 target for mitochondrial dysfunction. *Drug Discovery Today* **19**, 1601–1606
459 (2014).
- 460 21. Goldberg, N. D. Iron deficiency anemia in patients with inflammatory bowel
461 disease. *Clinical and Experimental Gastroenterology* **6**, 61–70 (2013).
- 462 22. Kojima, S., Sher-Chen, E. L. & Green, C. B. Circadian control of mRNA
463 polyadenylation dynamics regulates rhythmic protein expression. *Genes Dev.* **26**,
464 2724–2736 (2012).
- 465 23. Maillo, C. *et al.* Circadian- and UPR-dependent control of CPEB4 mediates a
466 translational response to counteract hepatic steatosis under ER stress. *Nat. Cell*
467 *Biol.* **19**, 94–105 (2017).
- 468 24. Watabe-Uchida, M., John, K. A., Janas, J. A., Newey, S. E. & Van Aelst, L. The
469 Rac Activator DOCK7 Regulates Neuronal Polarity through Local
470 Phosphorylation of Stathmin/Op18. *Neuron* **51**, 727–739 (2006).

- 471 25. Blasius, A. L. *et al.* Mice with mutations of Dock7 have generalized
472 hypopigmentation and white-spotting but show normal neurological function.
473 *Proc. Natl. Acad. Sci.* **106**, 2706–2711 (2009).
- 474 26. Zhang, Z. *et al.* Functional interaction of ERAP2 and HLA-B27 activates the
475 unfolded protein response. *Arthritis Rheumatol.* (2016). doi:10.1002/art.40033
- 476 27. Andrés, A. M. *et al.* Balancing selection maintains a form of ERAP2 that
477 undergoes nonsense-mediated decay and affects antigen presentation. *PLoS*
478 *Genet.* **6**, 1–13 (2010).
- 479 28. Saveanu, L. *et al.* Concerted peptide trimming by human ERAP1 and ERAP2
480 aminopeptidase complexes in the endoplasmic reticulum. *Nat. Immunol.* **6**, 689–
481 97 (2005).
- 482 29. Franke, A. *et al.* Genome-wide meta-analysis increases to 71 the number of
483 confirmed Crohn's disease susceptibility loci. *Nat. Genet.* **42**, 1118–25 (2010).
- 484 30. Damjanovich, L., Volkó, J., Forgács, A., Hohenberger, W. & Bene, L. Crohn's
485 disease alters MHC-rafts in CD4 + T-cells. *Cytom. Part A* **81 A**, 149–164 (2012).
- 486 31. Dodane, V. & Kachar, B. Identification of isoforms of G proteins and PKC that
487 colocalize with tight junctions. *J. Membr. Biol.* **149**, 199–209 (1996).
- 488 32. Meyer, T. N., Schwesinger, C. & Denker, B. M. Zonula occludens-1 is a
489 scaffolding protein for signaling molecules: G α 12 directly binds to the Src
490 homology 3 domain and regulates paracellular permeability in epithelial cells. *J.*
491 *Biol. Chem.* **277**, 24855–24858 (2002).
- 492 33. Sabath, E. *et al.* G α 12 regulates protein interactions within the MDCK cell tight
493 junction and inhibits tight-junction assembly. *J. Cell Sci.* **121**, 814–824 (2008).
- 494 34. Fukuhara, S., Chikumi, H. & Gutkind, J. S. RGS-containing RhoGEFs: the
495 missing link between transforming G proteins and Rho? *Oncogene* **20**, 1661–
496 1668 (2001).
- 497 35. Landy, J. *et al.* Tight junctions in inflammatory bowel diseases and inflammatory
498 bowel disease associated colorectal cancer. *World Journal of Gastroenterology*
499 **22**, 3117–3126 (2016).
- 500 36. Herroeder, S. *et al.* Guanine Nucleotide-Binding Proteins of the G12 Family
501 Shape Immune Functions by Controlling CD4+ T Cell Adhesiveness and

- 502 Motility. *Immunity* **30**, 708–720 (2009).
- 503 37. Pavlick, K. P. *et al.* Role of reactive metabolites of oxygen and nitrogen in
504 inflammatory bowel disease 1,2 1This article is part of a series of reviews on
505 ‘Reactive Oxygen and Nitrogen in Inflammation.’ The full list of papers may be
506 found on the homepage of the journal. 2Gue. *Free Radic. Biol. Med.* **33**, 311–
507 322 (2002).
- 508 38. Chu, F. F. *et al.* Bacteria-Induced Intestinal Cancer in Mice with Disrupted Gpx1
509 and Gpx2 Genes. *Cancer Res.* **64**, 962–968 (2004).
- 510 39. Maeda, S. *et al.* cDNA microarray analysis of *Helicobacter pylori*-mediated
511 alteration of gene expression in gastric cancer cells. *Biochem. Biophys. Res.*
512 *Commun.* **284**, 443–9 (2001).
- 513 40. Hockenbery, D. M., Oltvai, Z. N., Yin, X. M., Millman, C. L. & Korsmeyer, S.
514 J. Bcl-2 functions in an antioxidant pathway to prevent apoptosis. *Cell* **75**, 241–
515 51 (1993).
- 516 41. Saeki, N. *et al.* GASDERMIN, suppressed frequently in gastric cancer, is a target
517 of LMO1 in TGF-beta-dependent apoptotic signalling. *Oncogene* **26**, 6488–98
518 (2007).
- 519 42. Saeki, N. *et al.* Distinctive expression and function of four GSDM family genes
520 (GSDMA-D) in normal and malignant upper gastrointestinal epithelium. *Genes,*
521 *Chromosom. Cancer* **48**, 261–271 (2009).
- 522 43. Pal, L. R. & Moul, J. Genetic basis of common human disease: Insight into the
523 role of missense SNPs from genome-wide association studies. *J. Mol. Biol.* **427**,
524 2271–2289 (2015).
- 525 44. Jostins, L. *et al.* Host-microbe interactions have shaped the genetic architecture
526 of inflammatory bowel disease. *Nature* **491**, 119–24 (2012).
- 527 45. Breslow, D. K. *et al.* Orm family proteins mediate sphingolipid homeostasis.
528 *Nature* **463**, 1048–1053 (2010).
- 529 46. Ha, S. G. *et al.* ORMDL3 promotes eosinophil trafficking and activation via
530 regulation of integrins and CD48. *Nat. Commun.* **4**, 2479 (2013).
- 531 47. Schmiedel, B. J. *et al.* 17q21 asthma-risk variants switch CTCF binding and
532 regulate IL-2 production by T cells. *Nat. Commun.* **7**, 13426 (2016).

- 533 48. Dang, J. *et al.* ORMDL3 Facilitates the Survival of Splenic B Cells via an
534 ATF6 α -Endoplasmic Reticulum Stress-Beclin1 Autophagy Regulatory Pathway.
535 *J. Immunol.* (2017). doi:10.4049/jimmunol.1602124
- 536 49. Zhai, W. H., Song, C. Y., Huang, Z. G. & Sha, H. Correlation between the
537 genetic polymorphism of ORMDL3 gene and asthma risk: a meta-analysis.
538 *Genet. Mol. Res.* **14**, 7101–12 (2015).
- 539 50. Saleh, N. M. *et al.* Genetic association analyses of atopic illness and
540 proinflammatory cytokine genes with type 1 diabetes. *Diabetes. Metab. Res. Rev.*
541 **27**, 838–43 (2011).
- 542 51. Ma, X. *et al.* ORMDL3 contributes to the risk of atherosclerosis in Chinese Han
543 population and mediates oxidized low-density lipoprotein-induced autophagy in
544 endothelial cells. *Sci. Rep.* **5**, 17194 (2015).
- 545 52. Laukens, D. *et al.* Evidence for significant overlap between common risk
546 variants for Crohn's disease and ankylosing spondylitis. *PLoS One* **5**, e13795
547 (2010).
- 548 53. McGovern, D. P. B. *et al.* Genome-wide association identifies multiple
549 ulcerative colitis susceptibility loci. *Nat. Genet.* **42**, 332–7 (2010).
- 550 54. Asazuma, N. *et al.* Interaction of linker for activation of T cells with multiple
551 adapter proteins in platelets activated by the glycoprotein VI-selective ligand,
552 convulxin. *J. Biol. Chem.* **275**, 33427–33434 (2000).
- 553 55. Togni, M. *et al.* Regulation of In Vitro and In Vivo Immune Functions by the
554 Cytosolic Adaptor Protein SKAP-HOM. *Mol. Cell. Biol.* **25**, 8052–8063 (2005).
- 555 56. Alenghat, F. J. *et al.* Macrophages require Skap2 and Sirp α for integrin-
556 stimulated cytoskeletal rearrangement. *J. Cell Sci.* **125**, 5535–5545 (2012).
- 557 57. Königsberger, S. *et al.* HPK1 associates with SKAP-HOM to negatively regulate
558 Rap1-mediated B-lymphocyte adhesion. *PLoS One* **5**, 1–9 (2010).
- 559 58. Tanaka, M. *et al.* SKAP2 Promotes Podosome Formation to Facilitate Tumor-
560 Associated Macrophage Infiltration and Metastatic Progression. *Cancer Res.* **76**,
561 358–369 (2016).
- 562 59. Sartor, R. B. Mechanisms of disease: pathogenesis of Crohn's disease and
563 ulcerative colitis. *Nat. Clin. Pract. Gastroenterol. Hepatol.* **3**, 390–407 (2006).

- 564 60. Ghosh, S. *et al.* Natalizumab for active Crohn's disease. *N Engl J Med* **348**, 24–
565 32. (2003).
- 566 61. Eldridge MJ, Sanchez-Garrido J, Hoben GF, Goddard PJ, S. A. The Atypical
567 Ubiquitin E2 Conjugase UBE2L3 Is an Indirect Caspase-1 Target and Controls
568 IL-1 β Secretion by Inflammasomes. *Cell Rep.* **18**, 1285–1297 (2017).
- 569 62. Alpi, A. F., Chaugule, V. & Walden, H. Mechanism and disease association of
570 E2-conjugating enzymes: lessons from UBE2T and UBE2L3. *Biochem. J.* **473**,
571 3401–3419 (2016).
- 572 63. Lamkanfi, M., Walle, L. Vande & Kanneganti, T. D. Deregulated inflammasome
573 signaling in disease. *Immunological Reviews* **243**, 163–173 (2011).
- 574 64. Shuai, K. & Liu, B. Regulation of gene-activation pathways by PIAS proteins in
575 the immune system. *Nat. Rev. Immunol.* **5**, 593–605 (2005).
- 576 65. Sharma, M. *et al.* hZimp10 is an androgen receptor co-activator and forms a
577 complex with SUMO-1 at replication foci. *EMBO J.* **22**, 6101–6114 (2003).
- 578 66. Imielinski, M. *et al.* Common variants at five new loci associated with early-
579 onset inflammatory bowel disease. *Nat. Genet.* **41**, 1335–40 (2009).
- 580 67. Rakowski, L. A. *et al.* Convergence of the ZMIZ1 and NOTCH1 Pathways at C-
581 MYC in Acute T Lymphoblastic Leukemias. *Cancer Res.* **73**, 930–941 (2013).
- 582 68. Pinnell, N. *et al.* The PIAS-like Coactivator Zmiz1 Is a Direct and Selective
583 Cofactor of Notch1 in T Cell Development and Leukemia. *Immunity* **43**, 870–
584 883 (2015).
- 585 69. Li, X., Thyssen, G., Beliakoff, J. & Sun, Z. The Novel PIAS-like Protein
586 hZimp10 Enhances Smad Transcriptional Activity. *J. Biol. Chem.* **281**, 23748–
587 23756 (2006).
- 588 70. Jongstra-Bilen, J. & Jongstra, J. Leukocyte-specific protein 1 (LSP1): a regulator
589 of leukocyte emigration in inflammation. *Immunol. Res.* **35**, 65–74 (2006).
- 590 71. Wang, C. *et al.* Modulation of Mac-1 (CD11b/CD18)-Mediated Adhesion by the
591 Leukocyte-Specific Protein 1 Is Key to Its Role in Neutrophil Polarization and
592 Chemotaxis. *J. Immunol.* **169**, 415–423 (2002).
- 593 72. Wang, J. *et al.* Accelerated wound healing in leukocyte-specific, protein 1-
594 deficient mouse is associated with increased infiltration of leukocytes and

- fibrocytes. *J. Leukoc. Biol.* **82**, 1554–63 (2007).
73. Hwang, S.-H. *et al.* Leukocyte-specific protein 1 regulates T-cell migration in rheumatoid arthritis. *Proc. Natl. Acad. Sci.* **112**, E6535–E6543 (2015).
74. Gremel, G. *et al.* The human gastrointestinal tract-specific transcriptome and proteome as defined by RNA sequencing and antibody-based profiling. *J. Gastroenterol.* **50**, 46–57 (2015).
75. Hulur, I. *et al.* Enrichment of inflammatory bowel disease and colorectal cancer risk variants in colon expression quantitative trait loci. *BMC Genomics* **16**, 138 (2015).
76. Nakajima, T. TIP27: a novel repressor of the nuclear orphan receptor TAK1/TR4. *Nucleic Acids Res.* **32**, 4194–4204 (2004).
77. Collins, L. L. *et al.* Growth retardation and abnormal maternal behavior in mice lacking testicular orphan nuclear receptor 4. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 15058–63 (2004).
78. Johansson, ??sa *et al.* Common variants in the JAZF1 gene associated with height identified by linkage and genome-wide association analysis. *Hum. Mol. Genet.* **18**, 373–380 (2009).
79. Cooper, J. D. *et al.* Meta-analysis of genome-wide association study data identifies additional type 1 diabetes risk loci. *Nat. Genet.* **40**, 1399–401 (2008).
80. Thomas, G. *et al.* Multiple loci identified in a genome-wide association study of prostate cancer. *Nat. Genet.* **40**, 310–5 (2008).
81. Koontz, J. I. *et al.* Frequent fusion of the JAZF1 and JJAZ1 genes in endometrial stromal tumors. *Proc. Natl. Acad. Sci.* **98**, 6348–6353 (2001).
82. Martin, J. E. *et al.* A systemic sclerosis and systemic lupus erythematosus pan-meta-GWAS reveals new shared susceptibility loci. *Hum. Mol. Genet.* **22**, 4021–4029 (2013).
83. Bruni, F., Gramegna, P., Oliveira, J. M. A., Lightowlers, R. N. & Chrzanowska-Lightowlers, Z. M. A. REXO2 is an oligoribonuclease active in human mitochondria. *PLoS One* **8**, e64670 (2013).
84. Matondo, A. & Kim, S. S. Targeted-mitochondria antioxidants therapeutic implications in inflammatory bowel disease. *J. Drug Target.* 1–8 (2017).

- doi:10.1080/1061186X.2017.1339196
85. Acquati, F. *et al.* Microenvironmental control of malignancy exerted by RNASET2, a widely conserved extracellular RNase. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 1104–9 (2011).
86. Acquati, F. *et al.* Loss of function of Ribonuclease T2, an ancient and phylogenetically conserved RNase, plays a crucial role in ovarian tumorigenesis. *Proc. Natl. Acad. Sci.* 1222079110- (2013). doi:10.1073/pnas.1222079110
87. Gabrielsen, I. S. M. *et al.* Genetic risk variants for autoimmune diseases that influence gene expression in thymus. *Hum. Mol. Genet.* ddw152 (2016). doi:10.1093/hmg/ddw152
88. Chu, X. *et al.* A genome-wide association study identifies two new risk loci for Graves' disease. *Nat. Genet.* **43**, 897–901 (2011).
89. Caputa, G. *et al.* RNASET2 is required for ROS propagation during oxidative stress-mediated cell death. *Cell Death Differ.* **23**, 347–57 (2016).
90. Wang, Q. *et al.* Stress-induced RNASET2 overexpression mediates melanocyte apoptosis via the TRAF2 pathway in vitro. *Cell Death Dis.* **5**, e1022 (2014).
91. Moret-Tatay, I. *et al.* Possible Biomarkers in Blood for Crohn's Disease: Oxidative Stress and MicroRNAs—Current Evidences and Further Aspects to Unravel. *Oxid. Med. Cell. Longev.* **2016**, 1–9 (2016).