Published in final edited form as:

Cell. 2013 September 26; 155(1): 242–256. doi:10.1016/j.cell.2013.08.041.

Genetic Variants Regulating Immune Cell Levels in Health and Disease

A full list of authors and affiliations appears at the end of the article.

SUMMARY

The complex network of specialized cells and molecules in the immune system has evolved to defend against pathogens, but inadvertent immune system attacks on "self" result in autoimmune disease. Both genetic regulation of immune cell levels and their relationships with autoimmunity are largely undetermined. Here, we report genetic contributions to quantitative levels of 95 cell

*Correspondence: edoardo.fiorillo@irgb.cnr.it (E.F.), serena.sanna@irgb.cnr.it (S.S.), francesco.cucca@irgb.cnr.it (F.C.).
12These authors contributed equally to this work

ACCESSION NUMBERS

Tables containing the entire set of genome-wide association results can be downloaded or visualized at http://www.irgb.cnr.it/facsdataexplorer. All raw flow cytometry data are deposited at the European Genome-phenome Archive (EGA, http://www.ebi.ac.uk/) under accession number EGAS00001000574.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, four figures, ten tables, and one data file and is available with this article online at http://dx.doi.org/10.1016/j.cell.2013.08.041.

AUTHOR CONTRIBUTIONS

M.D., S.L., M.G.P., M.L., M.P., and F.D. prepared blood samples; F.B., A.M., M.F.U., M. Marcelli, R.C., M.O., R.P., C.B., B.T., and R.L. performed sequencing; C.S., S.S., F.R., R.B., R.A., I.Z., E.P., A.K., H.M.K., and M.V. analyzed sequencing data; G.R.A. supervised analyses on sequencing data; G.R.A and F.C. designed the sequencing effort; A.A., C.M.J., G.R.A., S.S., and F.C. led the sequencing effort; M. Marongiu, L.P., L.L., and C.M.J. provided IT support; M.D., F.B., S.L., A.M., M.P., and M.Z. performed genotyping; S.U. provided genotyping support; F.V., M.D., S.L., and V.S. measured immunophenotypes; V.O., G.R., M.C., E.F., and F.C. designed the immunophenotyping experiments; D.F. clinically evaluated flow cytometry data; E.F. and V.O. supervised flow cytometric measurements and performed cytometric analysis; M.S., G.S., and S.S. performed statistical analyses; S.S. supervised statistical analyses; M.F., S.A.M.U., S.O., and W.I.M. performed bioinformatic analyses; M.S., G.S., M.Z., M.P., and S.N. performed region-specific analysis of the data and selected candidate genes; G.R.A., D.S., and F.C. provided funds and reagents; V.O., M.B.W., E.F., S.S., and F.C. wrote the paper; D.S., S.N., C.M.J., and G.R.A. revised the paper; and F.C. designed and supervised the overall study. All authors read the paper and contributed to its final form.

WEB RESOURCES

The URLs for data presented herein are as follows:
1000G Imputation Cookbook, http://genome.sph.umich.edu/wiki/Minimac:_1000_Genomes_Imputation_Cookbook
ARLEQUIN, http://cmpg.unibe.ch/software/arlequin/about/about.php
ChEMBL, https://www.ebi.ac.uk/chembl/
Clinical Trials, http://clinicaltrials.gov/
eQTL BROWSER, http://eqtl.uchicago.edu/cgi-bin/gbrowse/eqtl/
ExPASy, http://expasy.org/
GWAS CATALOG, http://www.genome.gov/gwastudies
Haploxt, http://genome.sph.umich.edu/wiki/Haploxt
heatmap.2 {gplots}, http://hosho.ees.hokudai.ac.jp/~kubo/Rdoc/library/gplots/html/heatmap.2.html
ImmunoBase, http://www.immunobase.org/
Minimac, http://genome.sph.umich.edu/wiki/Minimac
OMIM, http://www.ncbi.nlm.nih.gov/omim
PharmGKB, http://www.pharmgkb.org/
Pupasuite 3, http://pupasuite.bioinfo.cipf.es/
REACTOME, http://www.reactome.org/
UCSC Genome Bioinformatics, http://genome.ucsc.edu/
UMAKE, http://genome.sph.umich.edu/wiki/UMAKE
US Food and Drug Administration, http://www.fda.gov/
VcfCodingSnps, http://genome.sph.umich.edu/wiki/VcfCodingSnps

types encompassing 272 immune traits, in a cohort of 1,629 individuals from four clustered Sardinian villages. We first estimated trait heritability, showing that it can be substantial, accounting for up to 87% of the variance (mean 41%). Next, by assessing ~8.2 million variants that we identified and confirmed in an extended set of 2,870 individuals, 23 independent variants at 13 loci associated with at least one trait. Notably, variants at three loci (*HLA*, *IL2RA*, and *SH2B3*/*ATXN2*) overlap with known autoimmune disease associations. These results connect specific cellular phenotypes to specific genetic variants, helping to explicate their involvement in disease.

INTRODUCTION

The immune system must defend against a huge variety of microbes and remember them. To accomplish this and kill cancer-transformed and virus-infected cells while recognizing and tolerating our own untransformed components requires the formation and regulation of a wide range of both generalist and specialist white cell (leukocyte) types. A fluorescence-activated cell sorting (FACS) approach has facilitated highly sensitive, simultaneous analysis of levels of these leukocyte subpopulations and is being used by the Human Immunology Project to characterize the immunological profile of healthy and sick individuals (Davis, 2008; Maecker et al., 2012).

Despite methodological advances, searches for connections between genetic variants and cellular immune phenotypes have typically proceeded by examining broad classes of immune cells (Ferreira et al., 2010; Nalls et al., 2011; Okada et al., 2011), and even the extent to which variation in immune cell subtypes is heritable is still unknown.

Here, we use FACS to profile extensively the human immune cell repertoire for a large population sample. Applying state-of-the-art genotyping and sequencing technologies to the same individuals, we proceed to dissect the inherited phenotypic structure of the human immune cell repertoire. Importantly, our results demonstrate connections between known immune-related disease risk alleles and levels of particular immune cell types, thus representing an important extension of previous autoimmune disease GWAS. Our hypothesis-generating approach, using individuals from a general population, is also distinct from hypothesis-driven comparisons of immune cell types between cases and controls, which can be hampered by limited a priori knowledge and affected by second-order effects due to the disease process or its therapy.

RESULTS AND DISCUSSION

Profiling the Human Immune Cell Repertoire

By FACS analyses, we characterized a wide range of circulating cell subtypes in an initial sample of 1,629 individuals enrolled in the SardiNIA study population cohort. The cells comprise the major leukocyte populations in peripheral blood (Figure 1), including monocytes, granulocytes, circulating dendritic cells (cDCs), natural killer (NK), B cells, and T cells, with a more detailed characterization of T cell subsets. More specifically, because of their functional relevance and potential involvement in many autoimmune and inflammatory

diseases, we focused on T cells subdivided according to their maturation and activation status, including subsets of regulatory T cells (Tregs) (Shevach, 2000; Wing and Sakaguchi, 2010). Overall, we defined a total of 95 cell types that were further assessed with respect to their parental and grandparental cell lineages, resulting in 272 evaluated immunophenotypic traits (Experimental Procedures and Figure 1; Figure S1 and Tables S1A–S1C available online).

Heritability and Correlations between Traits

We estimated heritability of circulating immune cell counts in the first 1,629 phenotyped individuals (Experimental Procedures), observing values from 3% to 87% (mean 41%). The most heritable traits corresponded to Tregs and their subsets (mean 55%) (Figure S2; Table S2A provides descriptive statistics and impact of age and gender covariates). Remarkably, most cell populations with very high heritability (>60%) were positive for the CD39 marker (see below). Gender typically had negligible effects on phenotypic variation; age was important for a subset of cellular phenotypes, especially the previously characterized reduction in naive CD8 T cells that might explain reduced vaccination success in the elderly (Buchholz et al., 2011; Sansoni et al., 2008).

By their nature, many traits are hierarchically and functionally correlated, as the different immune cell types originate from a limited number of common progenitors and interact continuously. To examine these relationships, we performed a bivariate analysis to estimate phenotypic and genetic correlation coefficients, i.e., the proportion of variance between each pair of traits due to the combined contribution of genetic and environmental factors and the variance attributable to genetic causes only, respectively (Extended Experimental Procedures). A depiction of the genetic and phenotypic correlations between cell counts and CD4:CD8 and T:B cell ratios is presented in the heat map (Figure 2). Similarities but also important differences in the patterns of genetic and phenotype correlation coefficients—reported in the upper and lower triangles of the figure delimited by the central diagonal—are immediately apparent. On the one hand, two large squares in the upper part of the diagonal are indicative of conjoint genetic and phenotypic correlations and tend to involve cells with markers, such as CD39 and CD45RA, whose expression is under strong genetic control (see next section), suggesting that the extent of similarity between traits reflects intrinsic relations dictated mainly by ontogenesis and coordinated evolution of traits—and hence shared antigen expression. On the other hand, the overall phenotypic correlations tend to be stronger than genetic correlations, consistent with additional effects of nongenetic factors on cell levels. An example of strong positive phenotypic but not corresponding strong genetic correlation is observed between some DC and Treg subsets (corresponding to the lower-right red cluster), in line with a mechanism by which an increase of DCs is controlled by an increase of Tregs (Wing and Sakaguchi, 2010).

Genetic Changes Affecting Immune Cell Traits

To identify the genetic variation accounting for the inherited component of the 272 immunophenotypic traits, we next performed a sequencing-based GWAS, assessing ~8.2 million variants in the 1,629 phenotyped individuals (Experimental−Procedures). At the significance threshold of $p < 5.26 \times 10^{-10}$, we identified 21 signals at 11 loci, linking genetic

variation to multiple cellular immunophenotypes and resulting in a total of 180 SNP-trait associations. We also replicated ($p < 5 \times 10^{-8}$) two previously suggested associations (Ferreira et al., 2010), resulting in a total of 23 association signals at 13 loci, which were then assessed and unequivocally confirmed in the extended sample set of 2,870 individuals, including 1,241 additional volunteers (Figures 3 and 4, Data S1, and Tables 1, S4A, S4B, and S5A).

The amount of phenotypic variation explained was always >2%, consistent with the expected statistical power of our sample, with nine variants explaining >5% and three variants >15% (Table 1). Considering the 132 traits for which we observed at least one genome-wide significant signal, the heritability explained ranged from 3.7% to 90.3%, and the proportion of explained heritability was >50% for 35 traits and >80% for four traits (Figure 5 and Table S2A), showing relatively large effects for human quantitative traits (Teslovich et al., 2010; Lango Allen et al., 2010).

Among the largest genetic effects detected, a single intronic variant of *ENTPD1*, coding for CD39, accounted for 60.8% of phenotypic variation (and 72% of the heritability) of the levels of CD39+activated CD4+ Tregs (Table 1 and Figure 4). Thus, this association has an obvious candidate mechanism in which *cis*-acting variation regulates the expression of a key marker in individual cells and therefore determines the number of cells expressing this molecule. CD39 is an ectoenzyme, expressed on monocytes, neutrophils, B, T, and NK cells (Pulte et al., 2007), which hydrolyzes extracellular ATP and ADP to AMP. Notably, among T cells, CD39 is mainly expressed by activated CD4+ Tregs, where it has an anti-inflammatory function by reducing extracellular proinflammatory ATP (Borsellino et al., 2007).

Other clear biological candidates among our lead associations included a variant near *IL2RA*, a gene encoding the transmembrane protein CD25 and associated with variation of T cells expressing high CD25 levels (CD45RA− CD25hi CD4+ not Treg cells); a variant near the *CD8A* and *CD8B* genes, encoding the cell surface glycoprotein CD8 and associated with variation in the level of T cells expressing CD8 (CD4+ CD8 dim); a variant near the HLA class II transactivator (*CIITA*) gene associated with the levels of activated T cells (i.e., HLA DR+ T lymphocytes); and a variant in the *TNFSF13B* gene associated with the levels of B cells. Notably, *CIITA* encodes a transcription factor influencing HLA class II expression, whereas *TNFSF13B* encodes the B-cell-activating factor of the TNF family (BAFF), inactivation of which is specifically associated with loss of mature circulating B cells (Table 1) (Mackay and Schneider, 2009).

Overall, 19 of the 23 variants reported here were associated with multiple traits, often with divergent effects on different traits (Tables S4A and S5A). A further layer of complexity was added by instances of multiple independent associations with the same traits within a single associated region. For example, independent variants within a region encompassing the *GALM* and *HNRPLL* genes (Table 1) increased the percentage of naive and terminally differentiated T cell subpopulations (those that are CD45RA positive), with corresponding decreases in the percentage of the memory T cell subsets (which are CD45RA negative). Association with *HNRPLL* is fully concordant with its role as the master regulator of *CD45*

splicing, a hallmark of T cell maturation (Wu et al., 2010). By contrast, the biology underlying the associations with *GALM* is less clear, though variants in *GALM* may act in long-distance regulation of *HNRPLL* because they fall in DNA regions known to interact with its promoter (Table S6A) (Li et al., 2012).

Other examples of multiple independent signals clustered in the same gene regions and associated with several traits were found near *ENTPD1* and in the *HLA* region (Table 1), where *multilocus* and *multiallelic* associations with complex diseases have been extensively documented (Marrosu et al., 2001). These results illustrate a new role for *HLA* variants, modulating immune system function by affecting the level of specific immune cell types. Of note, several variants in *HLA* class I alleles were associated with variation in the levels of numerous distinct CD8+ T cell subtypes, consistent with the notion that self-class I MHC molecules support CD8+ T cell survival (Takada and Jameson, 2009).

In general, most of the associations reported in this work are new, though some are consistent with previously detected signals. Specifically, we confirmed the putative associations between NK cell levels and variants near the *Schlafen* gene cluster and the association of CD4+ T cells with variation in the *SH2B3/ATXN2* gene region (Ferreira et al., 2010).

In addition to associations with $p < 5.26 \times 10^{-10}$, we observed several additional signals at $p < 5 \times 10^{-8}$ (Table S5B) that require confirmation by further analyses. Most of them are likely to be genuine—for example, the association of a common nonsynonymous variant (N1639S) in the lactase gene (*LCT*) with pDCs. It is striking that the association of two independent missense variants at this locus with leukocyte count in African Americans was recently reported (Auer et al., 2012), further supporting an unanticipated role of coding variation within this locus in the regulation of immune cell levels.

Our results also highlight the benefit of imputation and sequencing-based GWAS, both in detection of association signals and in the identification of the causal genes and variants (so-called “fine mapping”), which is relevant for downstream functional studies. In fact, 3 of the 13 detected loci (*NCAM*, *CD4*, and *HLA-E*) reached significance only after imputation. Across all loci, 20 lead variants were imputed, and 2 of them were not present even in the HapMap data set or in the most recent 1000 Genomes release and thus were not directly accessible by imputation from external resources. One, rs58055840, has proxies in the 1000 Genomes panel, but the other, chr10:98088623, is not strongly correlated with other known markers. Further investigation is required to determine whether these variants are specific to Sardinians.

Functional Clues from the Associated Variants

The 23 lead variants are located in noncoding regions, although two of them are in strong linkage disequilibrium (LD) ($r^2 > 0.8$) with nonsynonymous coding variants (with features of variants detailed in Table S4C). Furthermore, seven variants fall within known elements with regulatory capacity, including repressors, enhancers, and promoter elements or transcription-factor-binding sites (Table S6A). To assess functional processes and pathways through which the variants exert their effects, we selected a set of candidate genes based on

physical position and biological features and surveyed Gene Ontology (GO) terms and pathway enrichment (Experimental Procedures and Table 1). As expected, even when genes located in the *HLA* region were excluded, the overrepresented pathways and GO categories were predominantly related to immune function (e.g., immune response, immune system process, primary immunodeficiency, hematopoietic cell lineage, antigen processing and presentation, T cell receptor complex, IgG binding, MHC protein binding, and IL12-mediated signaling events) (Table S6B).

Overlapping Associations between Immune Traits and Diseases

After identifying immune-cell-associated variants, we checked whether any of them correlated with known disease associations. After identifying immune-cell-associated variants, we systematically checked in public databases whether any of them were, or were highly correlated ($r^2 > 0.8$) with, a known disease-associated variant previously reported at $p < 5 \times 10^{-8}$. We identified overlaps at three genetic loci: *HLA*, *IL2RA*, and *SH2B3/ATXN2* (Table 2, Figure S3, and Extended Experimental Procedures). Such overlapping associations identify specific immune cell types that are unbalanced in disease status and also suggest mechanisms by which specific risk alleles might lead to disease susceptibility, as follows.

Variation downstream of the *HLA-DRA* gene decreased the levels of memory CD8+ cells not expressing the costimulatory molecule CD28 (CD45RA−CD28−CD8+ cells) and correlated with published associated risk alleles for ulcerative colitis, systemic sclerosis, Parkinson's disease, and Hodgkin's lymphoma (Barrett et al., 2009; Enciso-Mora et al., 2010; Gorlova et al., 2011; Hamza et al., 2010).

A variant in the *IL2RA* gene region, rs61839660, was associated with a memory T cell subset expressing high CD25 levels (CD45RA−CD25hi CD4+ not Treg cells) and is also the strongest type 1 diabetes (T1D)-associated variant in the region (Huang et al., 2012; Lowe et al., 2007). Moreover, association with the same immune cells was previously observed at a variant in moderate LD ($r^2 = 0.77$), which was, at the time, the strongest T1D-associated variant (Dendrou et al., 2009). The allele that is responsible for an increase in the CD45RA −CD25hi CD4+ not Treg cells reduces the risk for T1D, thus linking this specific cell type to protection against T1D. The results also suggest that anti-CD25 therapies might increase risk for T1D by reducing the number of this protective cell type. Consistent with this, clinical trials have suggested an increased risk of T1D in transplant patients treated with anti-CD25 antibody (Bayés et al., 2007; Vendrame et al., 2010).

Another overlap was seen for a variant in *ATXN2* that is highly correlated with a missense variant within the *SH2B3* gene (R262W). The W262 nonancestral allele increases the levels of T lymphocytes and the helper CD4+ T cell subset with similar effect sizes, and it is positively associated with many autoimmune diseases (such as type 1 diabetes and celiac disease), as well as with hypertension and related pathologies (i.e., coronary heart disease and chronic kidney disease). Additionally, this variant has been associated with several endophenotypes in the general population, including platelet and eosinophil levels as well as systolic and diastolic blood pressure (http://www.genome.gov/gwastudies/). *SH2B3* encodes the adaptor protein LNK, whose mouse ortholog was earlier shown to be a negative regulator of hematopoiesis, cytokine signaling, and inflammation (Devallière and Charreau, 2011). An

increase of total T cells (CD3+ lymphocytes) and particularly of CD4+ T cells resulting from the W262 allele may thereby result in loss of function. Furthermore, this observation is consistent with findings in mouse models and humans suggesting the potential efficacy of monoclonal antibodies against CD3 in T1D and other autoimmune diseases (Chatenoud, 2010).

Relevant to its function, the *SH2B3*-associated variant marks an extended haplotype spanning ~200 kb (Data S1, panel 22A), indicative of strong positive selection (Barreiro and Quintana-Murci, 2010). During human evolution, a lymphocytosis-associated variant may have been useful for thousands of years in resistance to pathogens but, in recent less septic environments, becomes a risk factor for autoimmunity.

Other genetic variants might also be enriched by balancing selection to maintain a high degree of variation in immune cell levels in a given population, increasing the chances for survival of groups of individuals under different and often opposite environmental pressures. Among our associated variants, clear evidence of balancing selection was found in the *HLA* region (Extended Experimental Procedures), consistent with its key role in host defense and disease susceptibility.

In addition to coincident associations clearly satisfying stringent criteria, other overlapping signals for variants affecting both levels of specific cells and disease risk are likely genuine. For example, the allele associated with a higher level of HLA-DR+ (activated) T lymphocytes at *CIITA* is in moderate LD ($r^2 = 0.44$) with the risk allele for Celiac disease (CD) (Trynka et al., 2011). In this case, the lack of full coincidence at the same SNP or a suitable proxy may be attributable to differences in map resolution in different studies (the coverage at this locus was low in the CD study). Furthermore, our top variant is in strong LD ($r^2 = 0.99$) with a variant showing suggestive association with ulcerative colitis (McGovern et al., 2010).

Overall, the coincident associations between diseases and immune traits have special potential to reveal sites for therapeutic intervention, and indeed some of those detected had already been selected as targets for pharmaceutical therapy (Table S6C). It is also noteworthy that our work does not support some previous claims—largely based on functional evidence— about the involvement of specific cell type levels in specific diseases. For instance, a protective role of CD39+-activated CD4+ Tregs in various autoimmune diseases has been suggested (Chalmin et al., 2012; Fletcher et al., 2009), but no overlapping association was observed between disease and the major genetic variants affecting the quantitative regulation of this cell type.

Conclusions and Prospects

As part of the dynamic mounting and control of immune reactions, our results reveal that DNA variation superimposes powerful programmed regulation on various subtypes of leukocytes. Interestingly, those showing the greatest estimated inherited control are implicated in the more sophisticated cellular functions, such as regulatory T cells, which were phylogenetically the last to evolve and are also the last to appear in ontogenesis.

A number of the genetic associations identified here explain an appreciable fraction of trait heritability and demonstrate the feasibility of genetic dissection of quantitative variation of specific immune cell types. At least three factors likely contribute to the unusually high degree of explained heritability, which contrasts sharply with typical observations in GWAS for quantitative traits, for which “missing heritability” is the norm. First, examining more restricted cell types avoids dilution and possible opposing effects in mixtures of leukocytes; this notion is consistent, for example, with findings of large effect size variants associated with fetal hemoglobin that have no detectable effects on total hemoglobin (Uda et al., 2008). Second, the large genetic effect sizes could be related to intrinsic properties of the immune response, which, confronted at the population level with an unpredictable and changing environment, must ensure optimal primed variability in the quantitative levels of immune cell types. Finally, the sequencing-based approach employed provides assessment of genomic variation at an unprecedented level of resolution, except for very rare SNPs and indels.

The 13 reported loci point to specific DNA polymorphisms and putative proteins and mechanisms involved in regulation of cellular immunity. They also identify specific molecules and cell subtypes involved in a range of diseases, particularly autoimmune diseases, reflecting the dramatically shifting evolutionary balance between the optimization of effective response to pathogens and the risk of autoimmunity. Given that several association signals may have not been captured in this study due to sample size restrictions —and in most previous disease GWAS due to their restriction to common ubiquitous variants—many more overlapping associations are likely to be forthcoming when the approach described here is extended to larger samples. These overlaps should include multiple associations for the same trait and disease, reinforcing evidence of causal relationships between them. Our survey also reveals primary candidate genes to be resequenced in searches for both germline mutations in patients with selective and combined immunodeficiencies and driver somatic mutations in patients with circulating hematopoietic malignancies. Some of the observations presented here also hint at previously undocumented involvement of the immune system in maladies such as Parkinson’s disease, though rigorous testing in appropriate cohorts is required to assess these possibilities further.

For some autoimmune pathologies, the mechanistic clues involving specific cell types suggest targets but also concomitant risks for therapeutic interventions, with some drugs already in use or under clinical experimentation targeting the associated protein products for a number of loci. Further functional studies to explicate the effects of the variants on identified cell types could foster therapies aimed at controlling the numbers of those cell types to help regulate the immune system safely, preventing occurrence or lessening severity of autoimmune diseases.

EXPERIMENTAL PROCEDURES

Study Population

The SardiNIA project is a longitudinal study that recruited and phenotyped 6,148 individuals, males and females, aged 14–102, from a cluster of four towns in the Lanusei Valley (Pilia et al., 2006), located on the central east coast of Sardinia, Italy. During clinic

visits, fresh blood samples were collected and used for both DNA extraction and flow cytometric measurements. Initially, 1,629 individuals were characterized for the immune-related phenotypes described below, followed by an additional 1,241 individuals from the same cohort, to extend the sample size and validate the identified association results. Ethical permission for this study was granted by the Regional Ethics Committee (No 2009/0016600).

Flow Cytometric Measurements

Immunophenotyping was carried out by flow cytometry on fresh blood samples, and cell phenotyping was performed within 2 hr after collection to avoid any time-dependent artifacts. We selected and tested a set of multiplexed fluorescent antibodies to characterize the major leukocyte cell populations in peripheral blood, including monocytes, granulocytes, circulating dendritic cells, and lymphocytes subdivided into NK, B, and T cells and their subsets (Extended Experimental Procedures, Figure S1, and Table S1A). In particular, we assessed regulatory T cells (CD25hi, CD127-), subdivided into resting, activated, and cytokine-secreting nonsuppressive cells (Miyara et al., 2009; Shevach, 2000). We also used the HLA-DR marker to assess the activation status of T and NK cells and both the chemokine receptor CCR7 and the phosphatase CD45RA antigens to distinguish between naive, central memory (CM), effector memory (EM), and terminally differentiated (TD) T cell subsets (Sallusto et al., 1999). Moreover, in selected T cell subpopulations, we assessed the positivity for the ectoenzyme CD39 and the CD28 costimulatory antigen (Keir and Sharpe, 2005). Finally, cDCs were separated into myeloid (mDCs) and plasmacytoid (pDCs) cells and were further subdivided by the expression of the adhesion molecule CD62L and the costimulatory ligand CD86 (Steinman and Banchereau, 2007; Ohnmacht et al., 2009).

For all cell populations, we measured both absolute counts (AC) and the proportion of each type with respect to their progenitor cell lineages, expressed as percentages of the levels of parent (%P) and grandparent (%GP) cell lineages (Figure 1 and Table S1B). For example, helper CD4+ T cells were evaluated relative to CD3+ cells (parent cell population representing all T cells) and to total lymphocytes (grandparent cell population). Percentages with respect to parental and grandparental cell populations lead to more robust measures of cell levels by reducing variability in measurements resulting from sample handling or fluctuations by transient environmental factors that affect the total leukocyte counts. These percentages may also reveal association with molecular changes that alter factors involved in feedback mechanisms responsible for maintaining a balance between cells. Finally, we assessed the specific ratios of cell types that are widely clinically used and that examine the balance between T and B cells and between helper (CD4) and cytotoxic (CD8) T cells.

Overall, we examined 95 absolute counts, 94 percentages with respect to parent cells, 80 percentages with respect to grandparent cells, and 3 ratios between cell subsets (Table S1B).

To ensure reproducible measures over time, we followed a rigorous standardization protocol (Extended Experimental Procedures). In brief, (1) we daily adjusted internal parameters of FACS using standardized fluorescent beads to check and correct for laser wear and fluidic instability, and (2) we weekly validated cell counts through suitable quality control of stabilized blood samples. To directly assess reproducibility, we repeated the FACS

measurements in 35 participants sampled at least 3 months after their initial enrollment, finding overall high reproducibility (median value for all traits 0.90, mean 0.85, standard deviation 0.13) (Table S2B).

Heritability Estimation and Bivariate Analysis

We estimated heritability for all inverse-normalized traits in the first 1,629 immunophenotyped individuals (comprising 211 unrelated individuals and 1,418 subjects grouped in 249 families, leading to 567 sib pairs, 30 half-sib pairs, 248 cousins pairs, 609 parent-child pairs, 32 grandparent-grandchild pairs, and 561 avuncular pairs for analysis), including age and gender as covariates. Furthermore, familial clustering of blood sampling (i.e., same-day sampling of closely related individuals), which could bias heritability estimates, was checked for and excluded (Extended Experimental Procedures).

We also performed a bivariate analysis to estimate the phenotypic and genetic correlations between traits. In particular, for each trait pair, the phenotypic correlation was computed as the Spearman coefficient, whereas the genetic correlation was estimated as the cross trait-cross individual additive genetic covariance between traits normalized by the geometric mean of the individual trait genetic variances and by the kinship coefficient of pairs of individuals. We then used a hierarchical clustering analysis that successively connected the most similar traits, based on the estimated phenotypic and genetic correlation coefficients (Extended Experimental Procedures).

Genotyping and Whole-Genome Sequencing

The entire SardiNIA cohort was characterized using two Illumina custom arrays: the Cardio-MetaboChip and the ImmunoChip. These arrays were designed by international consortia to genotype regions of prior interest in metabolic and immune-related traits and diseases, respectively (Cortes and Brown, 2011; Voight et al., 2012), and resulted in quality-controlled 284,722 SNPs derived from both arrays. We also whole-genome sequenced 1,146 Sardinians at low pass (average 4-fold coverage) (Extended Experimental Procedures).

Statistical and Bioinformatical Analyses

We performed a GWAS for each trait, analyzing ~8.2 million variants assembled from the integration of the two assessed arrays and markers imputed with the Sardinian sequencing reference panel (Table S3 and Extended Experimental Procedures) (Li et al., 2009). Association was evaluated by a variance component-based regression analysis to account for family structure, using the same covariates as in heritability estimation (Chen and Abecasis, 2007). Traits were normalized using inverse normal transformation.

We selected all independently associated variants for each trait ($r^2 < 0.1$ or those remaining significant in a stepwise conditional analysis), using a significance threshold of $p < 5.26 \times 10^{-10}$. This threshold corresponds to the standard genome-wide threshold of 5×10^{-8} after further adjustment for 95 independent tests (the number of absolute cell count measurements). Although this approach is conservative given the high interdependency of cell lineages, it ensures the robustness of our findings. We successively removed poorly imputed variants and then eliminated redundant trait-variant associations by prioritizing the

most strongly associated variants at each locus and removing those in LD. We also included two suggestive associations ($5.26 \times 10^{-10} < p < 5 \times 10^{-8}$) at the previously described *SH2B3/ATXN2* and *SLFN13* gene regions (Ferreira et al., 2010).

To validate findings, we measured the corresponding associated immunophenotypes in an additional 1,241 individuals from the same SardiNIA cohort and genotyped variants representing novel signals that were not supported by a directly genotyped variant ($r^2 > 0.85$). Variants showing an excess of discordant genotypes or less significant p values after addition of the extended sample were excluded from further analyses (Tables S4A and S4B and Extended Experimental Procedures).

To calculate the amount of phenotypic variance explained by genetic factors, for each trait, we fitted a linear model containing age, gender, and all of the independent SNPs associated with that specific trait (full model) and a linear model containing only age and gender (basic model). The variance explained was calculated as the difference of the r^2-adjusted quantity observed in the full and basic models (Table S2A).

To prioritize candidate gene(s) at each locus, we searched for correlated expression quantitative trait loci (eQTLs), coding variants, and nearby genes involved in immune-related disorders, as reported in OMIM (Online Mendelian Inheritance in Man) or implicated in immunity in previous studies (Tables S4C and S4D). Bioinformatic analyses were carried out to characterize variants and genes, including colocalization with regulatory features and their potential for pharmaceutical interest (Tables S6A, S6B, and S6C). Lastly, to assess possible impact of the detected variants on disease susceptibility, we searched for coincident associations in public repositories (Table 2), such as the GWAS catalog (http://www.genome.gov/gwastudies/) and ImmunoBase (http://www.immunobase.org/) (Extended Experimental Procedures).

A schematic overview of the overall study design is depicted in Figure S4.

The Extended Experimental Procedures provide details about the study design, genetic and immunophenotypic data collection, and statistical and bioinformatic analyses.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Valeria Orrù[1,12], Maristella Steri[1,12], Gabriella Sole[1], Carlo Sidore[1,2,3], Francesca Virdis[1], Mariano Dei[1], Sandra Lai[1], Magdalena Zoledziewska[1], Fabio Busonero[1], Antonella Mulas[1,3], Matteo Floris[4], Wieslawa I. Mentzen[1], Silvana A.M. Urru[4], Stefania Olla[1], Michele Marongiu[1], Maria G. Piras[1], Monia Lobina[1,3], Andrea Maschio[1,2], Maristella Pitzalis[1], Maria F. Urru[4], Marco Marcelli[4], Roberto Cusano[1,4], Francesca Deidda[1,4], Valentina Serra[1,3], Manuela Oppo[4], Rosella Pilu[1,4], Frederic Reinier[4], Riccardo Berutti[3,4], Luca Pireddu[4,5], Ilenia Zara[4], Eleonora Porcu[1,3], Alan Kwong[2], Christine Brennan[11], Brendan Tarrier[11], Robert Lyons[11], Hyun M. Kang[2],

Sergio Uzzau[3,6], Rossano Atzeni[4], Maria Valentini[4], Davide Firinu[7], Lidia Leoni[4], Gianluca Rotta[8], Silvia Naitza[1], Andrea Angius[1,4], Mauro Congia[9], Michael B. Whalen[1], Chris M. Jones[4], David Schlessinger[10], Gonçalo R. Abecasis[2], Edoardo Fiorillo[1,12,*], Serena Sanna[1,12,*], and Francesco Cucca[1,3,12,*]

Affiliations

[1]Istituto di Ricerca Genetica e Biomedica (IRGB), CNR, Monserrato 09042, Italy

[2]Center for Statistical Genetics, University of Michigan, Ann Arbor, MI 48109, USA

[3]Dipartimento di Scienze Biomediche, Università di Sassari, Sassari 07100, Italy

[4]CRS4, Parco Tecnologico della Sardegna, Pula, Cagliari 09010, Italy

[5]Università degli Studi di Cagliari, Cagliari 09010, Italy

[6]Laboratorio di Proteomica, Porto Conte Ricerche Srl, Tramariglio, Alghero 07041, Italy

[7]Dipartimento di Allergologia e Immunologia, Università di Cagliari, Cagliari 09124, Italy

[8]BD Biosciences Italia, Buccinasco, Milano 20090, Italy

[9]Dipartimento di Scienze Biomediche e Biotecnologie, Università di Cagliari, Cagliari 09124, Italy

[10]Laboratory of Genetics, NIA, Baltimore, MD 21224, USA

[11]University of Michigan Sequencing Core, University of Michigan Medical School, Ann Arbor, MI 48109, USA

Acknowledgments

This work is dedicated to the memory of Professor Antonio Cao, mentor and leader, who passed away while this manuscript was in preparation. We thank all of the volunteers who generously participated in this study and made this research possible. We thank John Todd and Marcella Devoto for critical revisions of the manuscript, Manuela Sironi and Rachele Cagliani for advice on balancing selection tests, Luiz Pereira for web server implementation, and Manuela Uda for her previous leadership on the ProgeNIA/SardiNIA study. This research was supported, in part, by the Intramural Research Program of the NIH, National Institute on Aging, with contracts N01-AG-1-2109 and HHSN271201100005C; by Italian grants FISM 2011/R/13 “Approccio razionale per la ricerca di composti per la cura della sclerosi multipla basato sull’analisi dei target biologici individuati dagli studi di associazione sull’intero genoma in Sardegna” and FaReBio2011 “Farmaci e Reti Biotecnologiche di Qualità”; by funds MIUR/CNR for rare diseases and molecular screening, PNR-CNR Aging Program 2012–2014, Giovani Ricercatori 2007 (D.lgs 502/92) from the Italian Ministry of Health to F.C.; and by National Human Genome Research Institute grants HG005581, HG005552, HG006513, and HG007022 to G.R.A.

References

Auer PL, Johnsen JM, Johnson AD, Logsdon BA, Lange LA, Nalls MA, Zhang G, Franceschini N, Fox K, Lange EM, et al. Imputation of exome sequence variants into population- based samples and blood-cell-trait-associated loci in African Americans: NHLBI GO Exome Sequencing Project. Am J Hum Genet. 2012; 91:794–808. [PubMed: 23103231]

Barreiro LB, Quintana-Murci L. From evolutionary genetics to human immunology: how selection shapes host defence genes. Nat Rev Genet. 2010; 11:17–30. [PubMed: 19953080]

Barrett JC, Lee JC, Lees CW, Prescott NJ, Anderson CA, Phillips A, Wesley E, Parnell K, Zhang H, Drummond H, et al. Genome-wide association study of ulcerative colitis identifies three new

susceptibility loci, including the HNF4A region. Nat Genet. 2009; 41:1330–1334. [PubMed: 19915572]

Bayés B, Pastor MC, Lauzurica R, Granada ML, Salinas I, Romero R. Do anti-CD25 monoclonal antibodies potentiate posttransplant diabetes mellitus? Transplant Proc. 2007; 39:2248–2250. [PubMed: 17889153]

Borsellino G, Kleinewietfeld M, Di Mitri D, Sternjak A, Diamantini A, Giometto R, Höpner S, Centonze D, Bernardi G, Dell'Acqua ML, et al. Expression of ectonucleotidase CD39 by Foxp3+ Treg cells: hydrolysis of extracellular ATP and immune suppression. Blood. 2007; 110:1225–1232. [PubMed: 17449799]

Buchholz VR, Neuenhahn M, Busch DH. CD8+ T cell differentiation in the aging immune system: until the last clone standing. Curr Opin Immunol. 2011; 23:549–554. [PubMed: 21664807]

Chalmin F, Mignot G, Bruchard M, Chevriaux A, Végran F, Hichami A, Ladoire S, Derangère V, Vincent J, Masson D, et al. Stat3 and Gfi-1 transcription factors control Th17 cell immunosuppressive activity via the regulation of ectonucleotidase expression. Immunity. 2012; 36:362–373. [PubMed: 22406269]

Chatenoud L. Immune therapy for type 1 diabetes mellitus-what is unique about anti-CD3 antibodies? Nat Rev Endocrinol. 2010; 6:149–157. [PubMed: 20173776]

Chen WM, Abecasis GR. Family-based association tests for genomewide association scans. Am J Hum Genet. 2007; 81:913–926. [PubMed: 17924335]

Cortes A, Brown MA. Promise and pitfalls of the Immunochip. Arthritis Res Ther. 2011; 13:101. [PubMed: 21345260]

Davis MM. A prescription for human immunology. Immunity. 2008; 29:835–838. [PubMed: 19100694]

Dendrou CA, Plagnol V, Fung E, Yang JH, Downes K, Cooper JD, Nutland S, Coleman G, Himsworth M, Hardy M, et al. Cell-specific protein phenotypes for the autoimmune locus IL2RA using a genotype-selectable human bioresource. Nat Genet. 2009; 41:1011–1015. [PubMed: 19701192]

Devallière J, Charreau B. The adaptor Lnk (SH2B3): an emerging regulator in vascular cells and a link between immune and inflammatory signaling. Biochem Pharmacol. 2011; 82:1391–1402. [PubMed: 21723852]

Enciso-Mora V, Broderick P, Ma Y, Jarrett RF, Hjalgrim H, Hemminki K, van den Berg A, Olver B, Lloyd A, Dobbins SE, et al. A genome-wide association study of Hodgkin's lymphoma identifies new susceptibility loci at 2p16.1 (REL), 8q24.21 and 10p14 (GATA3). Nat Genet. 2010; 42:1126–1130. [PubMed: 21037568]

Ferreira MA, Mangino M, Brumme CJ, Zhao ZZ, Medland SE, Wright MJ, Nyholt DR, Gordon S, Campbell M, McEvoy BP, et al. Quantitative trait loci for CD4:CD8 lymphocyte ratio are associated with risk of type 1 diabetes and HIV-1 immune control. Am J Hum Genet. 2010; 86:88–92. [PubMed: 20045101]

Fletcher JM, Lonergan R, Costelloe L, Kinsella K, Moran B, O'Farrelly C, Tubridy N, Mills KH. CD39+Foxp3+ regulatory T Cells suppress pathogenic Th17 cells and are impaired in multiple sclerosis. J Immunol. 2009; 183:7602–7610. [PubMed: 19917691]

Gorlova O, Martin JE, Rueda B, Koeleman BP, Ying J, Teruel M, Diaz-Gallo LM, Broen JC, Vonk MC, Simeon CP, et al. Identification of novel genetic markers associated with clinical phenotypes of systemic sclerosis through a genome-wide association strategy. PLoS Genet. 2011; 7:e1002178. [PubMed: 21779181]

Hamza TH, Zabetian CP, Tenesa A, Laederach A, Montimurro J, Year-out D, Kay DM, Doheny KF, Paschall J, Pugh E, et al. Common genetic variation in the HLA region is associated with late-onset sporadic Parkinson's disease. Nat Genet. 2010; 42:781–785. [PubMed: 20711177]

Huang J, Ellinghaus D, Franke A, Howie B, Li Y. 1000 Genomes-based imputation identifies novel and refined associations for the Wellcome Trust Case Control Consortium phase 1 Data. Eur J Hum Genet. 2012; 20:801–805. [PubMed: 22293688]

Keir ME, Sharpe AH. The B7/CD28 costimulatory family in autoimmunity. Immunol Rev. 2005; 204:128–143. [PubMed: 15790355]

Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F, Willer CJ, Jackson AU, Vedantam S, Raychaudhuri S, et al. Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature. 2010; 467:832–838. [PubMed: 20881960]

Li Y, Willer C, Sanna S, Abecasis G. Genotype imputation. Annu Rev Genomics Hum Genet. 2009; 10:387–406. [PubMed: 19715440]

Li G, Ruan X, Auerbach RK, Sandhu KS, Zheng M, Wang P, Poh HM, Goh Y, Lim J, Zhang J, et al. Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. Cell. 2012; 148:84–98. [PubMed: 22265404]

Lowe CE, Cooper JD, Brusko T, Walker NM, Smyth DJ, Bailey R, Bourget K, Plagnol V, Field S, Atkinson M, et al. Large-scale genetic fine mapping and genotype-phenotype associations implicate polymorphism in the IL2RA region in type 1 diabetes. Nat Genet. 2007; 39:1074–1082. [PubMed: 17676041]

Mackay F, Schneider P. Cracking the BAFF code. Nat Rev Immunol. 2009; 9:491–502. [PubMed: 19521398]

Maecker HT, McCoy JP, Nussenblatt R. Standardizing immunophenotyping for the Human Immunology Project. Nat Rev Immunol. 2012; 12:191–200. [PubMed: 22343568]

Marrosu MG, Murru R, Murru MR, Costa G, Zavattari P, Whalen M, Cocco E, Mancosu C, Schirru L, Solla E, et al. Dissection of the HLA association with multiple sclerosis in the founder isolated population of Sardinia. Hum Mol Genet. 2001; 10:2907–2916. [PubMed: 11741834]

McGovern DP, Gardet A, Törkvist L, Goyette P, Essers J, Taylor KD, Neale BM, Ong RT, Lagacé C, Li C, et al. Genome-wide association identifies multiple ulcerative colitis susceptibility loci. Nat Genet. 2010; 42:332–337. [PubMed: 20228799]

Miyara M, Yoshioka Y, Kitoh A, Shima T, Wing K, Niwa A, Parizot C, Taflin C, Heike T, Valeyre D, et al. Functional delineation and differentiation dynamics of human CD4+ T cells expressing the FoxP3 transcription factor. Immunity. 2009; 30:899–911. [PubMed: 19464196]

Nalls MA, Couper DJ, Tanaka T, van Rooij FJ, Chen MH, Smith AV, Toniolo D, Zakai NA, Yang Q, Greinacher A, et al. Multiple loci are associated with white blood cell phenotypes. PLoS Genet. 2011; 7:e1002113. [PubMed: 21738480]

Ohnmacht C, Pullner A, King SB, Drexler I, Meier S, Brocker T, Voehringer D. Constitutive ablation of dendritic cells breaks self-tolerance of CD4 T cells and results in spontaneous fatal autoimmunity. J Exp Med. 2009; 206:549–559. [PubMed: 19237601]

Okada Y, Hirota T, Kamatani Y, Takahashi A, Ohmiya H, Kumasaka N, Higasa K, Yamaguchi-Kabata Y, Hosono N, Nalls MA, et al. Identification of nine novel loci associated with white blood cell subtypes in a Japanese population. PLoS Genet. 2011; 7:e1002067. [PubMed: 21738478]

Pilia G, Chen WM, Scuteri A, Orrú M, Albai G, Dei M, Lai S, Usala G, Lai M, Loi P, et al. Heritability of cardiovascular and personality traits in 6,148 Sardinians. PLoS Genet. 2006; 2:e132. [PubMed: 16934002]

Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, Boehnke M, Abecasis GR, Willer CJ. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics. 2010; 26:2336–2337. [PubMed: 20634204]

Pulte ED, Broekman MJ, Olson KE, Drosopoulos JH, Kizer JR, Islam N, Marcus AJ. CD39/NTPDase-1 activity and expression in normal leukocytes. Thromb Res. 2007; 121:309–317. [PubMed: 17555802]

Sallusto F, Lenig D, Förster R, Lipp M, Lanzavecchia A. Two subsets of memory T lymphocytes with distinct homing potentials and effector functions. Nature. 1999; 401:708–712. [PubMed: 10537110]

Sansoni P, Vescovini R, Fagnoni F, Biasini C, Zanni F, Zanlari L, Telera A, Lucchini G, Passeri G, Monti D, et al. The immune system in extreme longevity. Exp Gerontol. 2008; 43:61–65. [PubMed: 17870272]

Shevach EM. Regulatory T cells in autoimmmunity*. Annu Rev Immunol. 2000; 18:423–449. [PubMed: 10837065]

Steinman RM, Banchereau J. Taking dendritic cells into medicine. Nature. 2007; 449:419–426. [PubMed: 17898760]

Takada K, Jameson SC. Self-class I MHC molecules support survival of naive CD8 T cells, but depress their functional sensitivity through regulation of CD8 expression levels. J Exp Med. 2009; 206:2253–2269. [PubMed: 19752186]

Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, Pirruccello JP, Ripatti S, Chasman DI, Willer CJ, et al. Biological, clinical and population relevance of 95 loci for blood lipids. Nature. 2010; 466:707–713. [PubMed: 20686565]

Trynka G, Hunt KA, Bockett NA, Romanos J, Mistry V, Szperl A, Bakker SF, Bardella MT, Bhaw-Rosun L, Castillejo G, et al. Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. Nat Genet. 2011; 43:1193–1201. [PubMed: 22057235]

Uda M, Galanello R, Sanna S, Lettre G, Sankaran VG, Chen W, Usala G, Busonero F, Maschio A, Albai G, et al. Genome-wide association study shows BCL11A associated with persistent fetal hemoglobin and amelioration of the phenotype of beta-thalassemia. Proc Natl Acad Sci USA. 2008; 105:1620–1625. [PubMed: 18245381]

Vendrame F, Pileggi A, Laughlin E, Allende G, Martin-Pagola A, Molano RD, Diamantopoulos S, Standifer N, Geubtner K, Falk BA, et al. Recurrence of type 1 diabetes after simultaneous pancreas-kidney transplantation, despite immunosuppression, is associated with autoantibodies and pathogenic autoreactive CD4 T-cells. Diabetes. 2010; 59:947–957. [PubMed: 20086230]

Voight BF, Kang HM, Ding J, Palmer CD, Sidore C, Chines PS, Burtt NP, Fuchsberger C, Li Y, Erdmann J, et al. The metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. PLoS Genet. 2012; 8:e1002793. [PubMed: 22876189]

Wing K, Sakaguchi S. Regulatory T cells exert checks and balances on self tolerance and autoimmunity. Nat Immunol. 2010; 11:7–13. [PubMed: 20016504]

Wu Z, Yates AL, Hoyne GF, Goodnow CC. Consequences of increased CD45RA and RC isoforms for TCR signaling and peripheral T cell deficiency resulting from heterogeneous nuclear ribonucleoprotein L-like mutation. J Immunol. 2010; 185:231–238. [PubMed: 20505149]

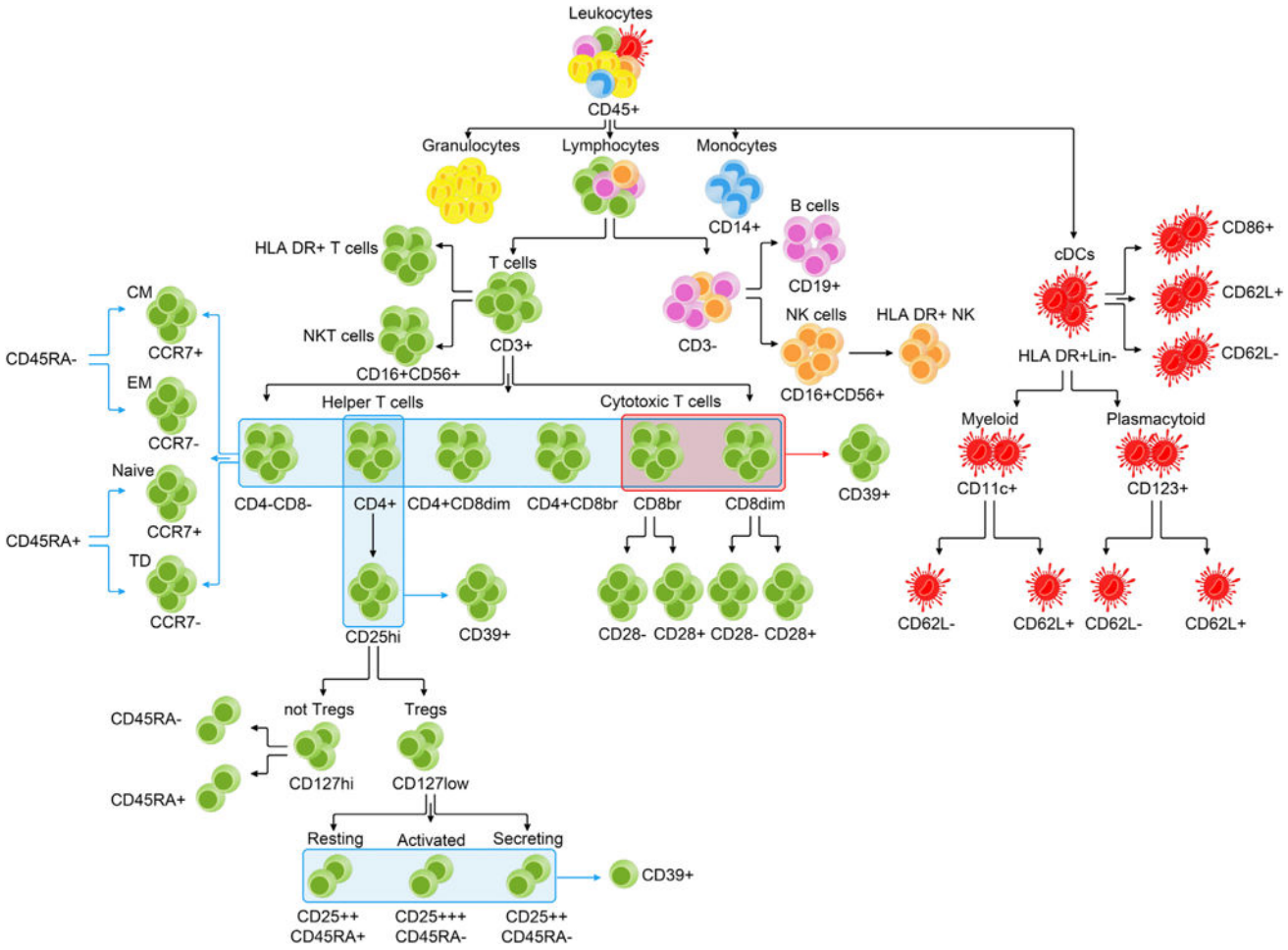

Figure 1. Studied Leukocyte Subpopulations

Color-coded diagram of the cell types analyzed by flow cytometry with arrows depicting the hierarchical levels of separation of circulating cell populations (leukocytes) and constituent subsets of the two main arms, innate and adaptive, of the immune system. Innate cell types, which provide prompt but generic responses to aggressors, include granulocytes (yellow), monocytes (pale blue), and dendritic cells (red). Adaptive cell types, which provide highly specific responses to microbial targets and may maintain a “memory” that enables a faster and greater response to previously encountered pathogens, include B cells (magenta) and T cells (green). The natural killer cells (orange) share features of both arms of the immune system. The name and, when relevant, the identifying marker are indicated beside each population. Cells inside a light-blue rectangle were phenotypically characterized with the antigen pointed to by the adjacent light-blue arrow; for example, the six CD3+ subsets (CD4− CD8−, CD4+, CD4+ CD8 dim, CD4+ CD8br, CD8br, and CD8 dim) are shown within a blue rectangle and were further subdivided into naive, central memory, effector memory, and terminally differentiated cells. The red rectangle indicates that the included cell populations have been jointly analyzed for CD39, the marker indicated by the red arrow. For simplicity, 45 of the 95 analyzed cell types, described in the full text, are shown. See also Figure S1 and Table S1.

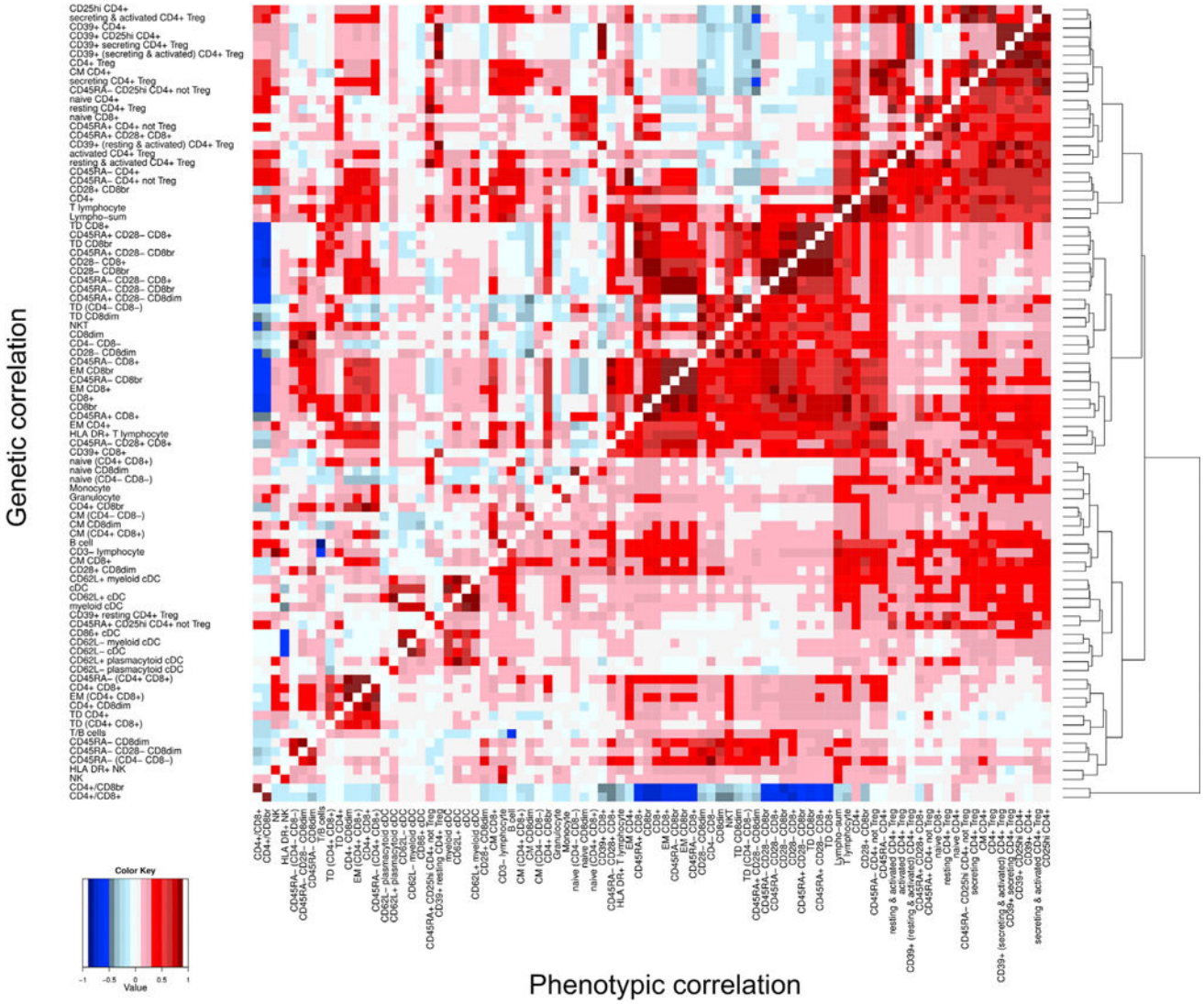

Figure 2. Phenotypic and Genetic Clustering

Heatmap of phenotypic (lower-right triangle) and genetic (upper-left triangle) correlations for cell counts and CD4:CD8 and T:B cell ratios. Traits with a phenotypic correlation ≥0.99 were excluded (Extended Experimental Procedures). Genetic and phenotypic triangles follow the same trait order, dictated by the clustering of phenotypic correlations, and the dendrogram at the right reflects the clustering. Traits connected by short branches share stronger phenotypic correlation, whereas traits that join near the root of the tree are weakly correlated. Color gradations indicate correlation strength, with red indicating direct correlation (from 0 to +1) and blue inverse correlation (from 0 to − 1). See also Figure S2 and Table S2 for further details.

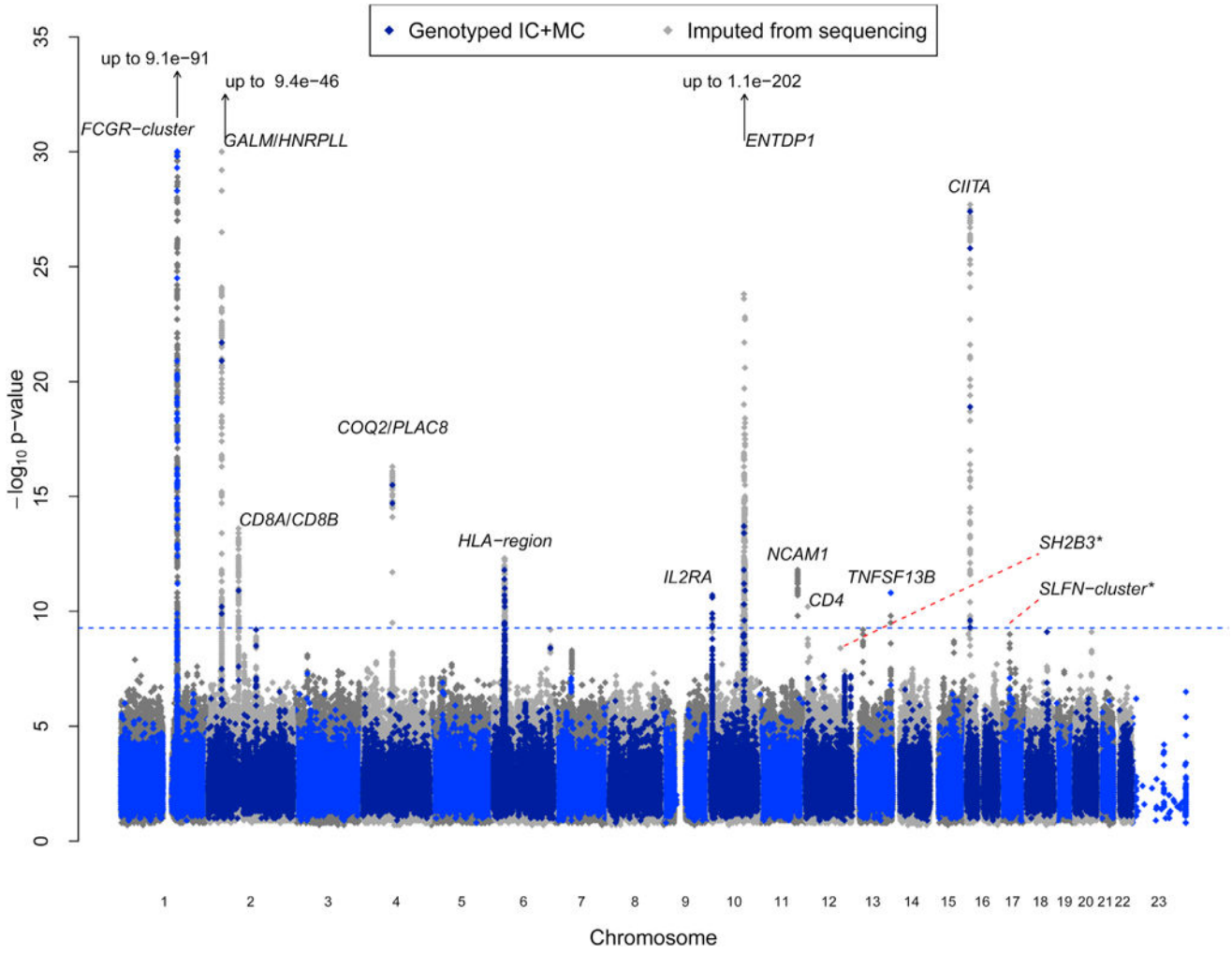

Figure 3. Manhattan Plot of Best p Values

For each SNP, the best p value observed among all assessed traits is plotted on a –log10 scale (y axis), according to its genomic coordinates (x axis). SNPs are colored in blue if the corresponding best p value was directly genotyped with ImmunoChip (IC) or Cardio-MetaboChip (MC) and in gray if imputed from genomic sequencing of Sardinians. The dotted horizontal line indicates the threshold for declaring a locus genome wide to be significant (5.26×10^{-10}). The best candidate gene is indicated near the peak. Loci below the significance threshold and previously described are marked with an asterisk.

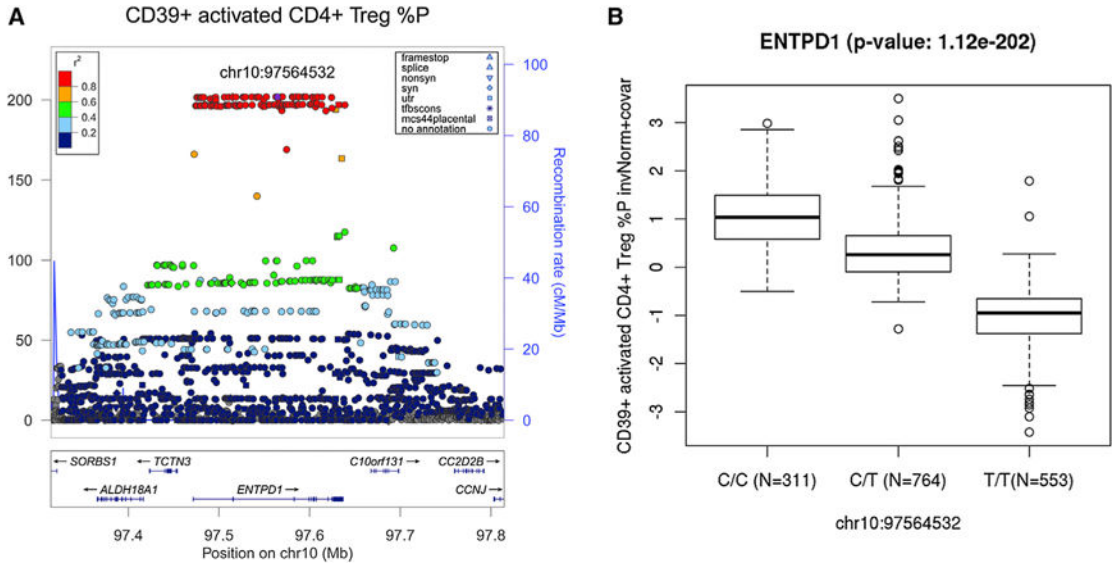

Figure 4. Regional Plot and Box Plot for the Top Signal in *ENTPD1*

(A and B) Representation of the association in the genomic context (A) and in the biological context (B) for the most strongly associated variant at the *ENTPD1* gene.

(A) Representation of the association strength (y axis shows the–log10 p value) versus the genomic positions (on hg19/GRCh37 genomic build) around the most significant SNP, which is indicated with a purple circle. Other SNPs in the region are color coded to reflect their LD with the top SNP, as in the left inset (taken from pairwise r^2 values calculated on Sardinian haplotypes), whereas symbols reflecting genomic functional annotation are indicated in the right inset. Genes and the position of exons, as well as the direction of transcription, are noted in lower boxes. This plot was drawn using the standalone version of the LocusZoom package (Pruim et al., 2010).

(B) The distribution of the immunophenotypic levels within each genotype class considering the normalized trait adjusted for age and gender in relation to the 1,629 initial samples, showing the additive effect that was statistically observed.See also Data S1.

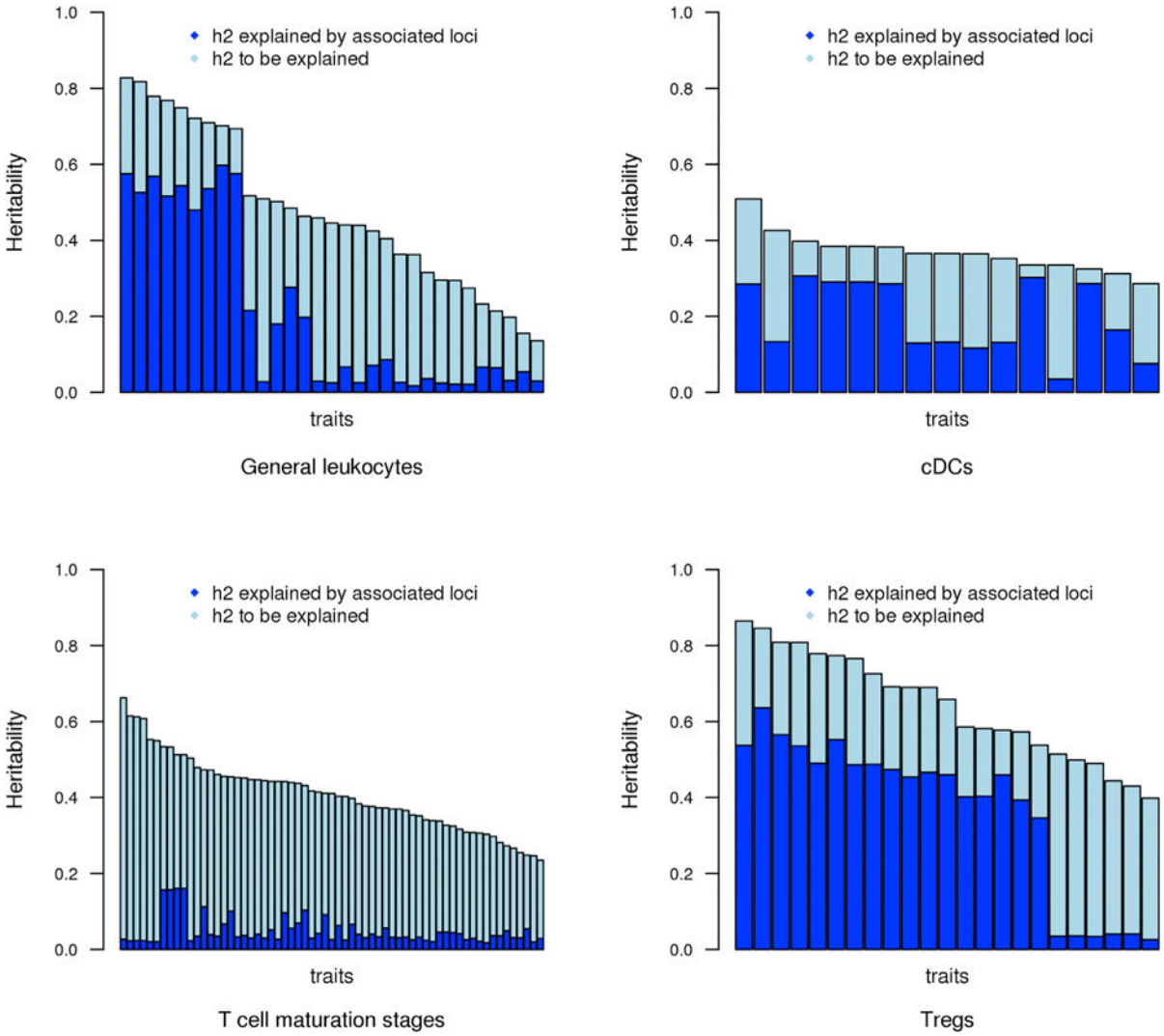

Figure 5. Proportion of Heritability Explained

The bar plots show the heritability of each trait (represented by a bar) for which genetic association was detected. The proportion of heritability explained by the detected loci is indicated in dark blue, and the proportion of heritability that remains to be explained is shown in light blue. Bars are grouped in their corresponding biological category, as specified in Table S1B. See also Tables S2A and S5.

Table 1

Twenty-Three Variants at the Thirteen Associated Loci

Locus	Candidate Genes	topSNP (chr:position/rsID)	A1/A2	Freq A1	Trait	Effect (SE)	Var. Expl.	p Value (n = 1,629)	SNP for Validation (chr:position/rsID)	r2 with topSNP	Validation p Value (n = 2,870)
1	*FCGR3A(p,c,o), FCGR2C(p,o), FCGR2A(e,c,o), FCGR2B(e,o), HSPA6(e), HSPA7(e)*	chr1:161536758/rs58055840	T/C	0.742	CD62L −myeloidcDC AC	−0.895 (0.044)	30.26	3.73×10^{-91}	chr1:161515326/rs55971447	0.937	6.83×10^{-129}
2	*HNRPLL(p)*	chr2:38792045/rs183949931	T/C	0.967	CD45RA −CD28 −CD8br %P	0.778 (0.105)	4.05	1.05×10^{-13}	chr2:38792045/rs183949931	same SNP	1.046×10^{-20}
2	*GALM(p,c,e), HNRPLL(b)*	chr2:38897074/rs13011383	G/A	0.730	TD CD4+ %GP	−0.371 (0.042)	5.52	6.05×10^{-19}	chr2:38886041/rs4670262	0.87	1.26×10^{-27}
2	*GALM(p), DHX57(e), HNRPLL(b)*	chr2:38921934/rs7583259	G/C	0.508	CD45RA −CD28 −CD8br %P	−0.548 (0.039)	15.09	9.40×10^{-46}	chr2:38932777/rs4670265	0.9	2.82×10^{-62}
3	*CD8A(p,c,o), RMND5A(p), CD8B(b), VPS24(e)*	chr2:87014377/rs2944254	C/T	0.810	CD4+ CD8dim AC	0.383 (0.05)	4.55	2.52×10^{-14}	chr2:87018547/rs3810831	0.943	1.3×10^{-22}
4	*COQ2(e), PLAC8(e), HPSE(e)*	chr4:84150313/rs4431216	T/C	0.633	CD62L −plasmacytoidcDC %P	0.337 (0.04)	5.19	4.96×10^{-17}	chr4:84179071/rs7667017	0.84	3.37×10^{-23}
5	*HLA-E(p,c,e), HCG27(e), GNL1(c), ABCF1(e), C2(e), PSORS1C3(e), RPP21(e), TRIM39(e), ZKSCAN2(e)*	chr6:30466505/rs117765619	G/T	0.516	CD45RA −CD8+ AC	−0.228 (0.037)	2.62	5.24×10^{-10}	chr6:30482993/rs2534812	0.974	1.34×10^{-11}
5	*HLA-B(p,c), VARS2(e), IER3(e), ZFP57(e)*	chr6:31327382/rs2395476	T/G	0.858	CD45RA −CD28+ CD8+ %P	0.352 (0.051)	3.21	3.69×10^{-12}	chr6:31327382/rs2395476	same SNP	1.827×10^{-19}
5	*HLA-DRA(p,e), BTNL2(p,c), HLA- DRB1(c,e), HLA-DQA1(e), HLA- DQB1(e), HLA-DRB5(e), HLA-DOB(e), LOC642073(e), VARS2(e), LST1(e), IER3(e), GTF2H4(e), HMGA1(e), RPL34(e)[a], AOAH(e)[a]*	chr6:32386433/rs113534101	G/A	0.776	CD4+ CD8dim %P	−0.299 (0.043)	3.07	5.68×10^{-12}	chr6:32383138/rs115615758	0.97	2.78×10^{-16}
5	*HLA-DRA(p), LOC642073(e), HLA-DOB(e), RPL34(e)[a], ARHGAP24(e)[a], AOAH(e)[a]*	chr6:32428186/rs6923504	G/C	0.618	CD45RA −CD28 −CD8+ AC	−0.249 (0.037)	3.01	2.81×10^{-11}	chr6:32428285/rs6903608	0.99	4.3×10^{-13}
6	*IL2RA(p,o)*	chr10:6094697/rs61839660	C/T	0.934	CD45RA −CD25hi CD4+ not Treg %P	−0.49 (0.073)	2.82	1.85×10^{-11}	chr10:6094697/rs61839660	same SNP	5.65×10^{-23}
6	*RBM17(p), IL2RA(p,o)*	chr10:6158412/rs8463	A/G	0.802	CD25hi CD4+ %P	−0.294 (0.046)	2.85	1.21×10^{-10}	chr10:6158412/rs8463	same SNP	2.02×10^{-15}
7	*SORBS1(p), C10orf61(e), ALDH18A1(c), ENTPD1(e)*	chr10:97331924/rs117568941	T/C	0.955	CD39+ CD8+ %GP	−0.650 (0.062)	6.68	1.45×10^{-25}	chr10:97331958/rs7099430	0.969	1.32×10^{-35}
7	*ALDH18A1(p), ENTPD1(b)*	chr10:97393678/rs1890187	A/G	0.975	CD39+ activated CD4+ Treg %P	−0.671 (0.073)	5.97	5.72×10^{-20}	chr10:97550405/rs11188485	0.97	2.97×10^{-32}
7	*ENTPD1(p,e)*	chr10:97564532/rs11517041	T/C	0.578	CD39+ activated CD4+ Treg %P	−1.113 (0.037)	60.81	1.12×10^{-202}	chr10:97515137/rs3814159	0.993	7.05×10^{-327}
7	*ZNF518A(p), BLNK(p,o), ENTPD1(b)*	chr10:97932006/rs117592294	C/T	0.955	CD39+ CD25hi CD4+ %P	0.497 (0.066)	4.33	6.26×10^{-14}	chr10:97932006/rs117592294	same SNP	1.35×10^{-15}
7	*DNTT(p), OPALIN(p), BLNK(o), ENTPD1(b)*	chr10:98088623	A/G	0.978	CD39+ CD4+ AC	−0.777 (0.094)	6.05	1.87×10^{-16}	chr10:98088623	same SNP	1.809×10^{-20}
8	*NCAM1(b)*	chr11:112706386/rs76771478	G/T	0.890	lymphosum %P	0.455 (0.064)	4.51	1.62×10^{-12}	chr11:112707378/rs1992842	0.96	7.18×10^{-17}
9	*CD4(p,e,o)*	chr12:6899181/rs2855537	G/T	0.606	naive (CD4+CD8+) AC	0.315 (0.048)	4.70	5.94×10^{-11}	chr12:6898460/rs7956804	1	4.77×10^{-13}
10	*TNFSF13B(p), LIG4(o)*	chr13:108957063/rs9520836	A/G	0.513	B cell %GP	−0.239 (0.035)	2.95	1.45×10^{-11}	chr13:108957063/rs9520836	same SNP	1.39×10^{-14}
11	*CIITA(p,o)*	chr16:10974355/rs9924520	A/G	0.778	HLA DR+T lymphocyte %P	−0.435 (0.039)	8.15	2.20×10^{-28}	chr16:10975311/rs4781011	0.994	9.29×10^{-50}

Locus	Candidate Genes	topSNP (chr:position/rsID)	A1/A2	Freq A1	Trait	Effect (SE)	Var. Expl.	p Value (n = 1,629)	SNP for Validation (chr:position/rsID)	r2 with topSNP	Validation p Value (n = 2,870)
12	*ATXN2(p), SH2B3(p,o)*	chr12:111973358/rs597808	G/A	0.539	T lymphocyte AC	−0.195 (0.035)	2.01	3.84×10^{-08}	chr12:111973358/rs597808	same SNP	1.87×10^{-09}
13	*SLFN13(p), SLFN12L(p,c), CCL1(e)*	chr17:33797371/rs9916257	T/G	0.568	NK %GP	−0.212 (0.035)	2.54	9.78×10^{-10}	chr17:33797371/rs9916257	same SNP	4.72×10^{-20}

The independently associated variants for each locus are tabulated, along with the association parameters. Indicated are, from left to right: the locus number; the candidate genes potentially regulated by the variant (for each candidate gene, a letter indicates the reason for inclusion: p, position; e, eQTL; c, coding; o, OMIM; b, biological candidate); the chromosomal position on hg19/GRCh37 genomic build of the lead variant and the corresponding SNP identification number (rs ID), when available; the major and minor alleles (A1 and A2) and the frequency of the major allele; the corresponding associated trait (CD8+ corresponds to the summation of CD8 bright and CD8 dim cells); the effect size in standard deviation units per each copy of allele A1; the standard error; the variance explained as percentage; and the p value. The last three columns report parameters for the SNP used in the validation step; the chromosome position with the corresponding identification number, the correlation with the lead SNP, and the p value of the validation data set are listed, respectively. See also Tables S3 and S4A.

[a]Trans eQTLs.

Table 2

ing Associations with Complex Diseases

ion	Immune Trait	SNP	Effect Allele/Other	Effect (SE)	p Value	Disease	SNP Disease	Best Reported p Value	Risk Allele/Other	r^2	Risk Allele/ Corresponding Trait Allele (Effect)	Source
s II (chr6p21.1)	CD45RA− CD28− CD8+ AC	rs6923504	G/C	−0.249 (0.037)	2.81×10^{-11}	Hodgkin's lymphoma	rs6903608	2.84×10^{-5}◦	G/A	0.99	G/G (decrease)	1
						Systemic sclerosis	rs3129882	1.89×10^{-27}	G/A	0.803	G/G (decrease)	1
						Ulcerative colitis	rs9268877	3.90×10^{-23}	T/C	0.83	G/G (decrease)	1
						Parkinson's disease	rs3129882	1.90×10^{-1}◦	G/A	0.803	G/G (decrease)	1
r10p15.1)	D25hi CD4+ %P	rs61839660	C/T	−0.484 (0.072)	2.38×10^{-11}	Type 1 diabetes	rs61839660	5.10×10^{-9}	C/T	1	C/C (decrease)	1
	D45RA− D25hi CD4+ not Treg AC	rs61839660	C/T	−0.484 (0.072)	1.05×10^{-10}	Type 1 diabetes	rs61839660	5.10×10^{-9}			C/C (decrease)	1
	D45RA− D25hi CD4+ not Treg %P	rs61839660	C/T	−0.484 (0.072)	1.85×10^{-11}	Type 1 diabetes	rs61839660	5.10×10^{-9}			C/C (decrease)	1
TXN2 (chr12q24.12)	T lymphocyte AC	rs597808	G/A	−0.195(0.035)	3.84×10^{-8}	Type 1 diabetes	rs3184504	2.80×10^{-27}	T/C	0.95	T/A (increase)	1,2
						Celiac disease	rs3184504	5.40×10^{-21}	T/C		T/A (increase)	2
						Primary hypothyroidism	rs3184504	2.60×10^{-12}	T/C		T/A (increase)	1
						Primary sclerosing cholangitis	rs3184504	5.91×10^{-11}	T/C		T/A (increase)	3
						Juvenile rheumatoid arthritis	rs3184504	2.60×10^{-9}	T/C		T/A (increase)	2
						Rheumatoid arthritis[a]	rs3184504	6.00×10^{-6}	T/C		T/A (increase)	1
						Coronary heart disease[a]	rs3184504	6.35×10^{-6}	T/C		T/A (increase)	1
						Multiple sclerosis[a]	rs3184504	6.70×10^{-5}	T/C		T/A (increase)	2
	CD4+ AC	rs597808	G/A	−0.195(0.036)	4.66×10^{-8}	Type 1 diabetes	rs3184504	2.80×10^{-27}			T/A (increase)	1,2
						Celiac disease	rs3184504	5.40×10^{-21}			T/A (increase)	2
						Primary hypothyroidism	rs3184504	2.60×10^{-12}			T/A (increase)	1
						Primary sclerosing cholangitis	rs3184504	5.91×10^{-11}			T/A (increase)	3
						Juvenile rheumatoid arthritis	rs3184504	2.60×10^{-9}			T/A (increase)	2
						Rheumatoid arthritis[a]	rs3184504	6.00×10^{-6}			T/A (increase)	1

ion	Immune Trait	SNP	Effect Allele/Other	Effect (SE)	p Value	Disease	SNP Disease	Best Reported p Value	Risk Allele/Other	r^2	Risk Allele/ Corresponding Trait Allele (Effect)	Source
						Coronary heart disease[a]	rs3184504	6.35×10^{-6}			T/A (increase)	1
						Multiple sclerosis[a]	rs3184504	6.70×10^{-5}			T/A (increase)	2
	CD4+ not Treg AC	rs597808	G/A	−0.195(0.036)	4.80×10^{-8}	Type 1 diabetes	rs3184504	2.80×10^{-27}			T/A (increase)	1,2
						Celiac disease	rs3184504	5.40×10^{-21}			T/A (increase)	2
						Primary hypothyroidism	rs3184504	2.60×10^{-12}			T/A (increase)	1
						Primary sclerosing cholangitis	rs3184504	5.91×10^{-11}			T/A (increase)	3
						Juvenile rheumatoid arthritis	rs3184504	2.60×10^{-9}			T/A (increase)	2
						Rheumatoid arthritis[a]	rs3184504	6.00×10^{-6}			T/A (increase)	1
						Coronary heart disease[a]	rs3184504	6.35×10^{-6}			T/A (increase)	1
						Multiple sclerosis[a]	rs3184504	6.70×10^{-5}			T/A (increase)	2
	T lymphocyte AC	rs597808	G/A	−0.195(0.035)	3.84×10^{-8}	Celiac disease	rs653178	7.15×10^{-21}	C/T	0.96	C/A (increase)	1,2
						Chronic kidney disease	rs653178	3.50×10^{-11}	C/T		C/A (increase)	1
						Rheumatoid arthritis[a]	rs653178	1.50×10^{-5}	C/T		C/A (increase)	1,2
	CD4+ AC	rs597808	G/A	−0.195(0.036)	4.66×10^{-8}	Celiac disease	rs653178	7.15×10^{-21}			C/A (increase)	1,2
						Chronic kidney disease	rs653178	3.50×10^{-11}			C/A (increase)	1
						Rheumatoid arthritis[a]	rs653178	1.50×10^{-5}			C/A (increase)	1,2
	CD4+ not Treg AC	rs597808	G/A	−0.195(0.036)	4.80×10^{-8}	Celiac disease	rs653178	7.15×10^{-21}			C/A (increase)	1,2
						Chronic kidney disease	rs653178	3.50×10^{-11}			C/A (increase)	1
						Rheumatoid arthritis[a]	rs653178	1.50×10^{-5}			C/A (increase)	1,2
	T lymphocyte AC	rs597808	G/A	−0.195(0.035)	3.84×10^{-8}	Primary biliary cirrhosis	rs11065979	2.87×10^{-9}	T/C	0.92	T/A (increase)	2
	CD4+ AC	rs597808	G/A	−0.195(0.036)	4.66×10^{-8}	Primary biliary cirrhosis	rs11065979	2.87×10^{-9}				2
	CD4+ not Treg AC	rs597808	G/A	−0.195(0.036)	4.80×10^{-8}	Primary biliary cirrhosis	rs11065979	2.87×10^{-9}				2
	T lymphocyte AC	rs597808	G/A	−0.195(0.035)	3.84×10^{-8}	Vitiligo	rs4766578	3.54×10^{-18}	T/A	0.96	T/A (increase)	1
	CD4+ AC	rs597808	G/A	−0.195(0.036)	4.66×10^{-8}	Vitiligo	rs4766578	3.54×10^{-18}				1

ion	Immune Trait	SNP	Effect Allele/Other	Effect (SE)	p Value	Disease	SNP Disease	Best Reported p Value	Risk Allele/Other	r^2	Risk Allele/ Corresponding Trait Allele (Effect)	Source
	CD4+ not Treg AC	rs597808	G/A	−0.195(0.036)	4.80×10^{-8}	Vitiligo	rs4766578	3.54×10^{-18}				1
r16p13.13)	HLA DR+ T lymphocyte AC	rs9924520	A/G	−0.425 (0.042)	1.46×10^{-23}	Ulcerative colitis[a]	rs4781011	3.23×10^{-6}	T/G	0.99	T/G (increase)	1
	HLA DR+ T lymphocyte %P	rs9924520	A/G	−0.435 (0.039)	2.20×10^{-28}	Ulcerative colitis[a]	rs4781011	3.23×10^{-6}	T/G			1
	HLA DR+ T lymphocyte %GP	rs9924520	A/G	−0.449(0.041)	2.35×10^{-28}	Ulcerative colitis[a]	rs4781011	3.23×10^{-6}	T/G			1

statistics from the immune trait analyses are reported in the first six columns. The pathology, the disease-associated variant, its best-reported p value in public repositories, and the risk allele are the seventh, eighth, ninth, and tenth columns, respectively. The LD (r^2) between immune trait variant and the disease-associated variant are shown in column 11, whereas column 12 lists the risk ed with the corresponding immune trait allele (and its effect). The last column indicates whether the disease was reported in GWAS Catalog (1), ImmunoBase (2), or PMID:23603763 (3). See S3 and Table S6.

e-associated variants in these rows do not reach the standard genome-wide association threshold ($p < 5 \times 10^{-8}$) in public databases.