**Impact of genetic polymorphisms on human immune cell gene expression**

Benjamin J. Schmiedel1, Divya Singh1, Ariel Madrigal1, Alan G. Valdovino-Gonzalez1, Brandie M. White1, Jose Zapardiel-Gonzalo1, Brendan Ha1, Gokmen Altay1, Jason A. Greenbaum1, Graham McVicker2, Grégory Seumois1, Anjana Rao1,3,4, Mitchell Kronenberg1,5, Bjoern Peters1,6, Pandurangan Vijayanand1,6,7,8.

1La Jolla Institute for Allergy and Immunology, La Jolla, CA 92037, USA.

2Salk Institute for Biological Studies, La Jolla, CA 92037, USA.

3Sanford Consortium for Regenerative Medicine, La Jolla, CA 92037, USA.

4Department of Pharmacology and Moores Cancer Centre, University of California San Diego (UCSD), La Jolla, CA 92037, USA.

5Division of Biological Sciences, University of California San Diego, La Jolla, CA 92037, USA.

6Department of Medicine, University of California San Diego, [La Jolla, CA 92093](https://maps.google.com/?q=9500+Gilman+Dr+%230656La+Jolla,+CA+92093&entry=gmail&source=g), USA.

7University of Southampton, Faculty of Medicine, Southampton SO16 6YD, UK.

8Lead contact.

Correspondence to: Pandurangan Vijayanand (vijay@lji.org).

**SUMMARY**

While many genetic variants have been associated with risk for human diseases, how these variants affect gene expression in various cell types remains largely unknown. To address this gap, the DICE (Database of Immune Cell Expression, Expression quantitative trait loci (eQTLs) and Epigenomics) project was established. Considering all human immune cell types and conditions studied, we identified *cis*-eQTLs for a total of 12,254 unique genes, which represent 61% of all protein-coding genes expressed in these cell types. Strikingly, a large fraction (41%) of these genes showed a strong *cis*-association with genotype only in a single cell type. We also found that biological sex is associated with major differences in immune cell gene expression in a highly cell-specific manner. These datasets will help reveal the effects of disease risk-associated genetic polymorphisms on specific immune cell types, providing mechanistic insights into how they might influence pathogenesis (http://dice-database.org).

**INTRODUCTION**

Genomic surveys of individuals from multiple populations have revealed significant genetic heterogeneity, with over 80 million autosomal single nucleotide polymorphisms (SNPs), including ~8 million common variants (1000 Genomes Project Consortium, 2015). Disease risk associations for several thousand of these SNPs have been identified by genome-wide association studies (GWAS) (Visscher et al., 2012), however, the overwhelming majority of GWAS hits lie in non-coding regions of the genome, and thus very little is known about exactly which genes they effect, in which cell types they act and what their functional relevance is to disease. In parallel, large-scale epigenomic initiatives analyzing multiple human cell types have attributed *cis*-regulatory function to nearly 20% of the non-coding DNA sequences, and showed that disease risk variants are enriched in cell-specific *cis*-regulatory DNA sequences (Roadmap Epigenomics et al., 2015). This indicates that disease risk variants are likely to affect gene expression in a context-dependent manner, in a specific subset of cell types in the human body. Defining the cell types in which disease risk variants modulate gene expression will aid mechanistic and functional studies that investigate the genetic basis of specific diseases.

The Genotype-Tissue Expression (GTEx) consortium assessed the association of genetic variants with gene expression levels (expression quantitative trait loci or eQTLs) across 44 human tissues (GTEx Consortium, 2015, 2017). Its results provide a valuable resource for narrowing down tissue-specific *cis*-regulatory effects of disease risk variants, but do not provide information on eQTLs in immune cell types. Previous large-scale eQTL studies of cells in the human immune system were performed in heterogeneous cell types and cell lines (Grundberg et al., 2012; Wright et al., 2014; Zhernakova et al., 2017) or in a relatively small number of homogeneous cell types from the same study subjects, thus limiting cell-specific eQTL analyses (Chen et al., 2016; Fairfax et al., 2014; Lee et al., 2014; Quach et al., 2016; Raj et al., 2014). The DICE (Database of Immune Cell Expression, Expression quantitative trait loci (eQTLs) and Epigenomics) project was established to define the transcriptional and epigenomic landscape of many human immune cell types in relation to genetic variation, thereby identifying both the cell-type restriction and potential function of human disease-associated SNPs.

In this first report from the DICE project, we identified common genetic variants that are associated with the expression of several thousand protein-coding and non-coding RNAs in 13 types of immune cells and 2 activation conditions, providing a map as to how human genetic variations affect gene expression in immune cell types. We have built a central resource (http://dice-database.org) to integrate and disseminate our datasets and to allow scientists to predict the effects of disease-associated genetic variants in immune cell types, thus connecting the genetic risk of disease to cell types and molecular pathways.

**RESULTS**

**Transcriptome of common human immune cell types**

In the first phase of the DICE project, we present transcriptomic data generated from 13 immune cell types isolated from 106 leukapheresis samples provided by 91 healthy subjects, including 15 who provided longitudinal samples. The cell types presented here comprise over 60% of all circulating mononuclear cells and include three innate immune cell types (CD14high CD16— classical monocytes, CD14— CD16+ non-classical monocytes, CD56dim CD16+ NK cells), four adaptive immune cell types that have not encountered cognate antigen in the periphery (naïve B cells, naïve CD4+ T cells, naïve CD8+ T cells and naïve regulatory T cells (TREG)), six CD4+ memory or more differentiated T cell subsets (TH1, TH1/17, TH17, TH2, follicular helper T cell (TFH) and memory TREG) and two activated cell types (naïve CD4+ and CD8+ T cells that were stimulated *ex vivo*) (**Fig. 1A,B**, **Fig. S1A-F** and **Table S1**).

RNA-Seq analysis showed that similar numbers of transcripts (ranging from 14,500 to 17,000, ~39% of all annotated transcripts) were detected (>1.0 transcripts per million, TPM) across all cell types and activation conditions (**Fig. 1C**). About 18% of these transcripts were long-noncoding RNAs (lncRNAs) (**Fig. 1D**), which as a class were expressed at lower levels compared to protein-coding transcripts (**Fig. S1G**). Though a majority of transcripts were detected in multiple cell types (**Fig. S1H,I**), their relative abundance varied substantially, resulting in distinct clustering patterns of different cell types (**Fig. 1E**).

**Cell-specific transcriptional signatures**

We classified the differentially expressed transcripts into groups that had significantly higher expression in one cell type than in the others (**Fig. 2** and **Table S2**). 4,602 transcripts showed highly cell-specific expression patterns (**Fig. 2A**), and were classified as cell-specific protein-coding transcripts, including several whose function is well-known (*CD163*, *SIGLEC10*, *GZMB*, *FCRL2*, *CD40LG*, *CRTAM*) and others where it is still poorly defined in the respective cell type (*PADI4*, *CLEC4F*, *C1orf21*, *TSPAN13*, *FHIT*) (**Fig. 2A,B**). Interestingly, several lncRNAs showed high abundance in specific immune cell types. Though very little is known about their function, the cell specificity of their expression raises the speculation that these lncRNAs may have an important role in the development and/or function of the respective immune cell type.

Differential gene expression analysis among the CD4+ T cell subsets revealed over 700 genes that were enriched for expression in one specific CD4+ T cell subset (**Fig. 2C**). Besides known CD4+ T cell subset-specific genes (*e.g.*, *FOXP3* in TREG cells, *IL17A* in TH17 cells, *CXCR5* in TFH cells), several novel protein-coding transcripts and lncRNAs (*e.g.*, *RP11-265P11.2* in naïve TREG cells, *AP001189.4* in memory TREG cells, *GPR56* in TH1 cells, *TNFRSF11A* in TH2 cells), whose function has not been fully characterized, with cell-specific expression patterns in CD4+ T cell subsets were identified (**Fig. 2C,D**). Overall, these findings emphasize the value of this resource in defining a reference for cell-specific transcripts, several of which may have important functional roles in the respective cell types.

**Expression QTLs are highly cell-specific**

We identified common single-nucleotide polymorphisms (SNPs) associated with variation in transcript abundance of a nearby gene, which we refer to as *cis*-eQTLs for convenience. Considering all immune cell types and including differences related to activation of naïve CD4+ and CD8+ T cells, we identified *cis*-eQTLs for a total of 12,254 unique genes, representing 61% of all expressed protein-coding genes and 49% of all lncRNAs, respectively, and majority of these were replicated in other eQTL studies (**Fig. 3A** and **Table S3**). For simplicity, genes with at least one significant *cis*-eQTL in any cell type are hereafter referred to as eGenes. The *cis*-eQTLs were enriched within intronic regions of the genome (**Fig. 3B**) and 12% of all *cis*-eQTLs were located within ±10 kb of the TSS (**Fig. 3C**). However, 37% of all *cis*-eQTLs were located >100 kb from the TSS of the target gene (**Fig. 3C**), suggesting that genetic variants may correlate with expression of more distant target genes.

We identified >2,000 eGenes (ranging from 2,421 to 3,590 eGenes/cell type) for each of the indicated six common cell types (**Fig. 3D**), and the majority (51%) of these genes showed a significant *cis*-association with genotype only in one specific cell type (**Fig. 3E-G**). The cell specificity of SNPs’ association was not entirely due to cell-specific expression of eGenes, as these associations were observed even for genes that were expressed at similar levels across multiple cell types, with some cell types highly susceptible to the effect of the SNP and others not (**Fig. 3F,G** and **Fig. S2A**). For example, *IRAK3* showed a peak *cis*-eQTL (rs1732886, adj. association *P* = 6.82 x 10-17) only in naïve B cells even though it was expressed at even higher levels in classical and non-classical monocytes (**Fig. 3G**). IRAK3 is a member of the interleukin-1 receptor-associated kinase family and known to function as a negative regulator of Toll-like receptor (TLR) signaling regulating immune homeostasis and tolerance (Hubbard and Moore, 2010).

Additional examples of cell-specific eGenes that we identified include two genes in the *CD96* locus that showed strong *cis-*eQTLs exclusively in naïve B cells (rs149612031, adj. *P* = 3.01 x 10-16 for *CD96* and 1.81 x 10-16 for *ZBED2*, respectively; **Fig. 3G** and **Fig. S2A**). Homozygosity of the GA allele of rs149612031 (64% of our subjects) was associated with lack of expression of the *CD96* and *ZBED2* transcripts in B cells but not in other cell types such as NK cells and T cells, suggesting that these genetic variants, if causally linked to gene expression, profoundly affect only susceptible cell types. CD96 belongs to the immunoglobulin family of receptors that bind to nectin-like proteins and is expressed by NK cells and T cells where it functions as an important negative regulator of cellular activation and cytokine responses (Chan et al., 2014); ZBED2 is a transcription factor. An interesting eQTL (rs2511162) associated with *GAB2* showed inverse effects on gene expression in B cells versus T cells; rs2511162A/A was associated with lower *GAB2* expression in B cells, whereas in T cells it was associated with increased expression of *GAB2*. Knocking down *GAB2* expression in naïve CD4+ T cells resulted in a significant reduction in the expression of activation markers (CD69, CD25 and HLA-DR) and cytokine transcripts (*IFNG* and *IL2*) following anti-CD3 and anti-CD28 stimulation (**Fig. S2B-D**). GAB2 has previously been shown to be an important adaptor protein involved in B and T cell signaling pathways (Osinalde et al., 2016); thus, if this genetic variant modulates *GAB2* protein expression, there could be divergent functional consequences in T cells and in B cells.

A ~240 kb dense haploblock of NK cell- and CD8+ T cell-specific *cis-*eQTLs at the 12p13.2 locus was associated with the expression of three lncRNAs and six C-type lectin-like receptors, including 5 activating NK receptors (*KLRK1*, *KLRC4-KLRK1*, *KLRC4*, *KLRC3*, *KLRC2*) and an inhibitory receptor (*KLRC1*) (**Fig. 3H**, top panel). Non-coding SNPs in general have previously been shown to perturb the function of cell-specific enhancers and thus affect gene expression in specific cell types (Schmiedel et al., 2016). As expected, several of the *cis*-eQTLs in the 12p13.2 locus directly overlapped many cell-specific DNase sites present in NK cells and naïve CD8+ T cells (**Fig. 3H**, middle panel). The G allele at rs2927561 was associated with increased expression of all five activating NK receptors and two lncRNAs in the locus, which suggested that this allele, if causally linked to gene expression, is likely to confer NK cells with an enhanced activation potential (**Fig. 3G** and **3H**, bottom panel). Given the importance of NK cells in anti-viral and tumor immunity, it would be interesting to test whether the 12p13.2 genetic variants that correlate with expression of an important group of NK receptors could have donor-specific consequences.

Previous studies have reported that several genetic variants are associated with gene expression only following perturbation of resting immune cells by relevant environmental (extrinsic) stimuli (Fairfax et al., 2014; Lee et al., 2014). To uncover such environment-responsive *cis*-eQTLs, we activated naïve CD4+ and CD8+ T cells *ex vivo* and analyzed the association of SNPs with gene expression in an activation-induced manner. We identified additional *cis*-eQTLs for >1,300 genes exclusively in activated naïve CD4+ and/or CD8+ T cells (**Fig. 3I** and **Table S3A**). These activation-induced eQTLs were associated with several genes encoding products known to influence T cell function, such as, the immune checkpoint molecule *PDCD1* (PD-1) (Keir et al., 2008), the co-stimulatory molecule *TNFRSF9* (4-1BB) (Watts, 2005), the co-inhibitory molecule CD59 (Sivasankar et al., 2009) and SGK1 (Wu et al., 2013), a kinase that regulates differentiation of TH17 and TH2 cells (**Fig. 3J** and **Fig. S2E**). In a time course experiment, we confirmed that following activation, a greater proportion of CD4+ T cells expressed cell surface 4-1BB protein in subjects with the *TNFRSF9* (4-1BB) expression-increasing allele (T/T genotype at rs9657975) compared to those with C/C genotype (**Fig. 3K**). This effect was not observed for other markers of activation such as CD69, CD25 or HLA-DR (**Fig. 3K** and **Fig. S2F**), which suggested that despite similar levels of activation, 4-1BB protein expression differed based on genotype. Since 4-1BB is currently tested as a co-stimulatory cancer immunotherapy target (Chester et al., 2018), our description of genotype-dependent effects on expression of 4-1BB in T cells may aid in the development of genetic biomarkers that stratify potential responders to 4-1BB targeted therapies.

**CD4**+ **T cell subset-specific eQTLs**

Naïve CD4+ T cells when activated *in vivo* following pathogen exposure, differentiate into several lineage-specific memory subsets such as TH1, TH1/17, TH17, TH2, and TFH subsets (Zhou et al., 2009), which have distinct gene expression programs and functional roles in host defense (**Fig. 2C**). Naïve TREG cells are another important CD4+ T cell lineage that develops in the thymus and differentiates into memory TREG cells in the periphery where they play an immune regulatory role (Sakaguchi et al., 2010). Our eQTL analysis of CD4+ memory and TREG subsets yielded over 7,800eGenes, of which >2,500 were unique to these subsets (**Fig. 4A**). Further, we identified >2,000 eGenes (ranging from 2,381 to 3,560 eGenes/cell type, **Fig. 4B**) for each of the eight CD4+ T cell subsets and a large proportion (44%) of these genes showed a significant *cis*-association with genotype only in one cell type (**Fig. 4C,D**).

A noteworthy example of an eGene specific to the CD4+ memory and TREG subsets was *ENTPD1* that encodes for CD39, a cell-surface ectonucleotidase, which has been widely considered to play a role in the generation of immunosuppressive microenvironments through adenosine production from ATP (Takenaka et al., 2016). Several other eGenes showed highly cell-type-restricted effects (**Fig. 4E**), for example, *CCNG2*, encoding for a cell cycle protein, cyclin G2, showed a peak eQTL (rs4859751) specifically in TH1/17 and TH17 cells although it was also highly expressed in other TH subsets and immune cell types. Whereas, *NCF4,* which encodes neutrophilic cytosolic factor 4, a component of the NADPH oxidase complex, showed prominent cis-QTLs specifically in CD4+ T subsets (**Fig. 4E**).

We also identified 2 genes in the Fc receptor like (*FCRL*) gene locus that showed strong *cis*-eQTLs primarily in TREG subsets (rs2065883, adj. association *P* = 5.6 x 10-18 for *FCRL1* and 3.8 x 10-18 for *FCRL2* and 1.81 x 10-16 in naïve TREG cells, respectively; **Fig. 4E**). The function of FCRL1 and FCRL2in TREG cells is not known, but given its role in modulating B cell receptor signaling pathway in B cell subsets (Li et al., 2014; Rostamzadeh et al., 2018), where it is highly expressed, it is conceivable that these proteins may module TREG activation and function in a genotype dependent manner. Our data emphasize the importance of studying purified, homogeneous cell types to identify cells and genes most susceptible to the effects of genetic variants. It will also be important to confirm directly that the SNPs indeed affect the biological functions of the corresponding cell types, using appropriate experimental systems of development or activation.

**Cell types most susceptible to GWAS SNPs**

To shed light on the disease relevance of the eQTLs identified in human immune cells, we next focused on the subset of eGenes linked to eQTLs that directly overlapped haploblocks of significant GWAS SNPs (*PGWAS* < 5 x 10-8) that were associated with one or more of 846 unique human diseases and traits (**Table S4** and **STAR Methods**). Most importantly, about 19% of all the eGenes identified in each human immune cell type (2,297 unique GWAS eGenes; 599 to 987 eGenes/cell type) were also associated with GWAS SNPs (**Fig. 5A,B**), and for about 36% of the eGenes associations were observed in only one or two cell types (**Fig. S3A**). Together, these results suggest that disease risk alleles may have a profound impact on immune cell gene expression and that their effects may be highly restricted to specific cell types.

We highlight in **Fig. 5C**, **Fig. S4** and **Table S4** the cell types most susceptible to the potential effects of genetic variants associated with common immune-mediated diseases. rs12936231, a variant associated with several autoimmune diseases and asthma, whose functional effects have been previously described (Schmiedel et al., 2016), is associated with the modulation of expression of *ORMDL3* and *GSDMB* in lymphocyte subsets and not in cells of myeloid origin. In contrast, the *ERAP2* peak eQTL (rs2548540), which was also a lead GWAS SNP for multiple autoimmune diseases, caused genotype-dependent expression differences across all the immune cell types studied (**Fig. S3B**).

Coding and non-coding polymorphisms in the *LACC1* gene*,* encoding for the protein FAMIN (fatty acid metabolism-immunity nexus) that drives fatty acid oxidation and ATP regeneration, were previously shown to be associated with increased risk of Crohn’s disease (CD). Due to its strong expression in macrophages, studies on the function of FAMIN and its loss-of-function variants have so far focused on this cell type where it was shown to control inflammasome activation (Cader et al., 2016) and the release of pro-inflammatory cytokines in response to TLR signaling (Lahiri et al., 2017). However, we observed that *LACC1* was upregulated more than 10-fold following activation of T cells and most importantly, the CD-risk variant (rs9567293C/C) was associated with reduced expression of *LACC1* in CD4+ memory T cell subsets, as well as in resting and activated naïve T cells but not in monocytes (**Fig. 5D** and **Fig. S3C**). We therefore tested if the function of T cells was affected by FAMIN. As FAMIN modulated TLR signaling responses in monocytes, we specifically tested the role of FAMIN in T cell responses to TLR co-stimulation. We found that *LACC1* knockdown in CD4+ T cells when compared to control siRNA treatment resulted in significantly reduced activation (CD25 and HLA-DR surface expression) and expression of cytokine transcripts (*IFNG* and *IL2*) following TLR1/2 co-stimulation and CD3/CD28 stimulation (**Fig. 5E** and **Fig. S3D-F**). This result suggested that the reduced *LACC1* levels observed in subjects carrying the CD-risk variant (rs9567293C/C) is likely to have important functional consequences in their T cells, and thus raised the hypothesis that T cells may be dominant players in driving the *LACC1*-related genetic risk of Crohn’s disease.

We identified several eQTLs that influenced the expression of genes associated by GWAS studies with risk for the rare autoimmune disease primary biliary cholangitis (PBC), including *NAB1* and *SYNGR1*. Interestingly, we found that the expression of *NAB1* displayed prominent genotype-dependent association (rs13028201, adj. *P* = 1.41 x 10-14) only in NK cells (**Fig. 5F** and **Fig. S3C**), though it was widely expressed across all the other immune cell types studied. NAB1 is a corepressor for early growth response (EGR) transcription factors and its mouse homologue NAB2 has been shown to be required for preventing activation-induced cell death in CD8+ T cells that received CD4+ T cell help (Wolkers et al., 2012). The PBC-risk variant rs74804869 influenced *SYNGR1* expression inNK and also CD8+ T cells (**Fig. 5F** and **Fig. S3C**). To date, no studies have examined the functional role of *NAB1* or *SYNGR1* in NK cells or linked NK cells to PBC pathogenesis, our results thus are likely to generate hypothesis that motivate further investigation. As a first step in that process, we knocked down these genes in human NK cells and tested their impact on NK cell activity (Fauriat et al., 2010). We found that reducing *NAB1* expression in NK cells significantly reduced the release of pro-inflammatory cytokines/chemokines such as IFN-γ, MIP-1β, RANTES and TNF by human NK cells activated by IgG crosslinking (**Fig. 5G** and **Fig. S3G,H**). *NAB1* knockdown also reduced IFN-γ release when NK cells were activated by co-culture with MHC devoid K-562 cancer cells (**Fig. 5H**). Knocking down *SYNGR1* also impacted cytokine production by activated NK cells (**Fig. 5G,H** and **Fig. S3G,H**). Together, these results suggested that PBC-risk variants that modulate *NAB1* and *SYNGR1* expressionare likely to impact NK cell function and influence PBC disease susceptibility.

The expression of several lncRNAs was influenced by disease-risk SNPs in a cell-restricted manner. For example, NK cellsdisplayed highly significant eQTLs associated with the lncRNAs *RP11-973H7.1* and *RP11-305L7.1*; T cell subsets an association between *RP11-77C3.3* and a SNPlinked to Type 1 diabetes (**Fig. 5C**, **Fig. S4** and **Fig. S5A**). These associations suggest that these lncRNAs may play an important role in disease pathogenesis, although this requires further study.

Interesting examples of disease-risk variants that display cell type-restricted effects in specific CD4+ T memory subsets (**Fig. 5C**, **Fig. S4** and **Fig. S5B**) include an RA-risk variant (rs3093026) and celiac disease-risk variant (rs7374671) that were associated with expression of chemokine receptor genes *CCR6* and *CCR2*, respectively, in TH1/17 and TH17 subsets. A CD-risk variant (rs528101238) was associated with expression of *EDN3*, encoding for endothelin 3, mainly in TH2 cells (**Fig. 5C** and **Fig. S5B**), and notably, endothelin signaling has been shown modulate T cell function (Tanaka et al., 2014). TFH cells that play an important role in ‘helping’ B cell responses showed a prominent association for an eGene (*TSPAN32*) linked to a B cell malignancy (chronic lymphocytic leukemia, CLL). Interestingly, Tspan32-deficent murine T cells have been shown to exhibit increased proliferative responses (Tarrant et al., 2002), thus raising speculation about the role of this protein in modulating TFH cell–B cell ‘help’ function to drive the genetic risk linked to CLL.

TREGcells were most susceptible to the effects of several autoimmune disease-risk variants. Notable eGenes associated with these variants include *IKZF4*, *PDE4A* and *PTGIR* (**Fig. 5C,I** and **Fig. S5B**). EOS, encoded by *IKZF4*, is a transcription factor that is required for the suppression of several FOXP3 target genes (Pan et al., 2009). More recently, it has been shown that downregulation of EOS leads to TREG plasticity, which can result in pathogenic conversion of TREG to effector cells in autoimmune diseases (Sharma et al., 2013). Therefore, we speculate that TREG cells from carriers with the *IKZF4* expression-reducing autoimmune disease risk-allele (rs1701704C/C (Hakonarson et al., 2008) in tight linkage with rs2640562T/T) may be more susceptible to conversion to pathogenic effector cells in autoimmune diseases. Phosphodiesterase 4A, encoded by *PDE4A*, hydrolyzes the second messenger cAMP that is known to play an important role in driving immunosuppressive signals in T cells, and TREG cells may also transfer cAMP through gap junctions into effector cells to dampen their functional responsiveness (Wehbi and Tasken, 2016). Whereas, prostaglandin I2 (PGI2) signaling through the PGI2 receptor (IP), encoded by *PTGIR*, increases intracellular cAMP levels, and recent studies in Ptgir-deficient T cells have shown an important role of PGI2-IP signaling in TREG development and function (Bloodworth et al., 2016; Liu et al., 2013). Thus, the products of autoimmune risk-eGenes *PDE4A* and *PTGIR* are likely to modulate cAMP-mediated immunoregulation and influence disease pathogenesis. Our results suggested that genotype-dependent perturbations in the expression patterns of several important genes involved in human TREG function is likely to have an impact on autoimmune disease pathogenesis.

Overall, these examples show that the effects of disease risk alleles may not necessarily be most prominent in the cell type in which the associated gene is most highly expressed. Studies of the functional impact of these alleles should therefore focus on the cell types most susceptible to their expression-modulating effects.

**Sex has major effects on immune cell gene expression**

It is well known that immune responses in humans and model organisms are different between sexes. For example, females mount more robust innate and antibody responses to infections and vaccination, but also display increased susceptibility to many autoimmune diseases (Klein and Flanagan, 2016). To determine if sex as a biological factor significantly influences the transcriptome of immune cells, we performed differential gene expression (DE) analysis between sexes considering ancestry and age as covariates. 1,875 unique transcripts, including 1,553 protein-coding transcripts and 196 lncRNAs, were differentially expressed between sexes (‘sex-biased transcripts’) across all the cell types investigated in this study, of which 93% were transcribed from autosomes (**Fig. 6A,B** and **Table S5A**). The majority (>72%) of these transcripts showed pronounced sex-biased expression pattern in only a single immune cell type, suggesting that sex-related transcriptional differences are highly cell-specific (**Fig. 6C** and **Fig. S6A**).

Independent pathway analysis of female-biased transcripts for each cell type revealed enrichment of genes encoding products involved in the canonical interferon and pattern recognition receptor pathways (**Table S5B**). For example, in classical monocytes alone, over 50 transcripts linked to the interferon pathway, including inflammatory cytokines and chemokines like *IL6*, *TNF*, *CXCL10,* displayed sex-biased expression, and enhanced expression was very prominent in some female subjects (**Fig. 6D,E**). Interferons (type 1 and type 2), interferon regulatory factors (IRFs), toll-like receptors (TLRs), and TLR ligands were also predicted by Ingenuity pathway analysis (IPA) to be upstream regulators of sex-biased transcripts (**Fig. 6E** and **Table S5C**).

Several female-biased transcripts we identified are worthy of specific mention. *FAM13A* (family with sequence similarity 13 gene) was expressed at higher levels in the naïve CD4+ T cells of female subjects. We also found a *cis*-eQTL (rs1246642) whose impact on expression was stronger in female subjects (**Fig. 6F,G** and **Fig. S6B,C**); homozygous females carrying the CC versus TT genotype had 3.5-fold greater levels of *FAM13A* transcripts compared to a <1.5-fold increase in males. This effect was also observed in TH1/17, TH17 and TFH CD4+ T cell subsets (**Fig. S6B,C**). GWAS studies have strongly implicated *FAM13A* as a susceptibility gene for chronic obstructive pulmonary disease (COPD) (Cho et al., 2014), and subjects with the risk alleles express higher levels of *FAM13A* in lung tissue (Kim et al., 2014). The higher expression of *FAM13A* among females that we found may help to explain their greater risk for developing severe early onset COPD with less cigarette-smoking exposure compared to males (Aryal et al., 2014). Further, our studies point to CD4+ T cells, especially from females, as highly susceptible to the effects of COPD-risk alleles in the *FAM13A* locus. To investigate the function of FAM13A in T cells, we performed a screening assay that is based on transcriptomic analysis of antigen-specific T cells following *FAM13A* knockdown (**Fig. 6H** and **Fig. S6D-F**). We found that *IL22* transcript level was most significantly reduced in *FAM13A* knock down conditions compared to control siRNA treatment (**Fig. 6H**). We then independently confirmed at the protein level a significant reduction in the proportion of IL-22 producing memory CD4+ T cells following *FAM13A* knockdown (**Fig. 6I** and **Fig. S6G-I**). Our results suggested that FAM13A is a positive regulator of IL-22 production by memory CD4+ T cells and thus females with higher expression of *FAM13A* are likely to have more pronounced IL-22 responses *in vivo*. IL-22 acts on epithelial cells to promote the release of anti-microbial peptides and chemokines as well as wound healing and regenerative gene programs, and thus plays an important role in barrier immunity as well as in promoting excessive inflammatory responses in skin diseases, where IL-22 blocking therapies have shown success (Guttman-Yassky et al., 2018). Recent studies have found that serum and sputum IL22 levels were elevated in patients with COPD (Zhang et al., 2013), implying IL-22 may also have a pathogenic role in COPD. Further studies in model organisms are required to confirm the link between FAM13A-induced IL-22 production by T cells and COPD pathogenesis.

We also found female-biased expression of *NLRP2,* which encodes an intracellular pattern recognition receptor from the NOD-like receptor (NLR) family NLRP2, which is involved in inflammasome assembly and secretion of caspase-1-dependendent inflammatory cytokines such as IL-1β (Van Gorp et al., 2014) (**Fig. 6J** and **Fig. S6J**). *PTGER2,* which encodes the prostaglandin receptor E2 (EP2), displayed one of the most significant female-biased increased expression patterns, specifically in naïve and memory T cell subsets (**Fig. 6J** and **Fig. S6J**). Prostaglandin E2 signaling through EP2 has been shown to promote the differentiation and pro-inflammatory functions of human and murine TH17 and TH1 cells (Boniface et al., 2009). Furthermore, blocking EP2 receptors in T cells has been shown to be effective in preventing the development of arthritis in mouse models (Sreeramkumar et al., 2016). Given the increased propensity of women to develop autoimmune arthritis and the pathogenic role of TH17 cells in driving its disease pathogenesis, we speculate that the increased expression of *PTGER2* transcripts in females may play a role in amplifying pathogenic TH17 responses that lead to the development of arthritis. Further functional studies are required to formally test these hypotheses.

Overall, we report on transcripts with sex-biased expression in specific immune cell types, and hypothesize that their products may play an important role in driving the sex differences in immune cell function.

**DISCUSSION**

Despite robust associations between genetic variants and several human diseases, the molecular mechanisms of how non-coding genetic variants perturb gene expression remains unclear in most cases. To address this gap several large-scale studies have described the association of common genetic variants with gene expression in a wide-range of tissues (GTEx Consortium, 2015, 2017), and uncovered the scale of eGenes (nearly 80% of all protein-coding transcripts) and the tissue types where their effects were observed. However, previously published eQTL studies in immune cell types were restricted to either heterogeneous populations or a small number of homogenous cell types, thus making it difficult to tease out cell-specificity of eQTLs. Therefore, the simultaneous study of multiple and diverse cell types is essential to uncover all potential cell-specific effects of genetic variants.

In this first report from the DICE project, we describe results of transcriptomic and eQTL analyses of 13 immune cell types *ex vivo* as well as activated naïve CD4+ and CD8+ T cells. We discovered a total of 12,254 genes with *cis*-eQTLs and a large fraction (41%) of these genes showed a strong *cis*-association with genotype only in a single cell type. The novelty and strength of our study is the simultaneous analysis of many purified cell types and conditions (*n* = 15) from the same subjects (*n* = 91), compared to previous reports where just a few purified immune cell types were compared. This approach allowed us to map cell-specific effects of genetic variants and importantly, link GWAS variants to cell types where their effects are most pronounced. We found that biological sex had a profound effect on the gene expression profiles of immune cells, with a total of over 1,800 genes showing sex-biased expression patterns. We report on several sex-biased genes that may contribute to the risk of developing autoimmune diseases in female subjects as well as explain the previously reported differences in response to infections and vaccination.

Overall, the DICE project provides a reference of transcriptomic and eQTL data for the human immune system, which is likely to facilitate mechanistic and functional investigations into the role of disease risk variants in relevant cell types. Additionally, the sex-related differences evident in various immune cell types will spur interest in understanding the biological consequences of such changes. The next phases of the DICE project will further enrich this dataset with descriptions from additional rare immune cell types and the inclusion of cell-specific epigenomic profiles that can allow pinpointing functional disease-risk variants.

**ACKNOWLEDGMENTS**

We thank the La Jolla Institute (LJI) Flow Cytometry, Clinical Studies and Sequencing Core. This work was funded by the William K. Bowes Jr Foundation (P.V.) and the NIH grants R24 AI108564 (P.V., B.P., A.R., M.K.), S10RR027366 (BD FACSAria II) and S10OD016262 (Illumina HiSeq 2500).

**AUTHOR CONTRIBUTIONS**

B.J.S., A.R., M.K., B.P., and P.V. conceived the work, designed and analyzed the experiments, B.J.S. and P.V. wrote the paper; B.J.S., D.S., A.G.V.-G. and B.M.W. performed the preparation of PBMC, FACS sorting and/or RNA isolation under the supervision of B.J.S. and P.V.; B.J.S. performed the functional experiments; D.S. performed RNA-Seq under the supervision of B.J.S., G.S. and P.V.; A.M., J.Z.-G., B.H. and G.A. performed processing and analysis of sequencing data under the supervision of B.J.S., J.A.G., G.M., B.P. and P.V.

**DECLARATION OF INTERESTS**

The authors declare no competing interests.

**REFERENCES**

1000 Genomes Project Consortium (2015). A global reference for human genetic variation. Nature *526*, 68-74.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166-169.

Aryal, S., Diaz-Guzman, E., and Mannino, D.M. (2014). Influence of sex on chronic obstructive pulmonary disease risk and treatment outcomes. Int J Chron Obstruct Pulmon Dis *9*, 1145-1154.

Bacher, P., Heinrich, F., Stervbo, U., Nienen, M., Vahldieck, M., Iwert, C., Vogt, K., Kollet, J., Babel, N., Sawitzki, B.*, et al.* (2016). Regulatory T Cell Specificity Directs Tolerance versus Allergy against Aeroantigens in Humans. Cell *167*, 1067-1078 e1016.

Ballouz, S., Verleyen, W., and Gillis, J. (2015). Guidance for RNA-seq co-expression network construction and analysis: safety in numbers. Bioinformatics *31*, 2123-2130.

Bardou, P., Mariette, J., Escudie, F., Djemiel, C., and Klopp, C. (2014). jvenn: an interactive Venn diagram viewer. BMC Bioinformatics *15*, 293.

Bloodworth, M.H., Goleniewska, K., and Peebles, R.S. (2016). Prostaglindin I2 Receptor (IP) Signaling Increases Regulatory T (Treg) Cell Induction and Function and Renders T Effector (Teff) Cells More Susceptible to Treg-Mediated Suppression. J Allergy Clin Immun *137*, Ab270-Ab270.

Boniface, K., Bak-Jensen, K.S., Li, Y., Blumenschein, W.M., McGeachy, M.J., McClanahan, T.K., McKenzie, B.S., Kastelein, R.A., Cua, D.J., and de Waal Malefyt, R. (2009). Prostaglandin E2 regulates Th17 cell differentiation and function through cyclic AMP and EP2/EP4 receptor signaling. J Exp Med *206*, 535-548.

Cader, M.Z., Boroviak, K., Zhang, Q., Assadi, G., Kempster, S.L., Sewell, G.W., Saveljeva, S., Ashcroft, J.W., Clare, S., Mukhopadhyay, S.*, et al.* (2016). C13orf31 (FAMIN) is a central regulator of immunometabolic function. Nat Immunol *17*, 1046-1056.

Chan, C.J., Martinet, L., Gilfillan, S., Souza-Fonseca-Guimaraes, F., Chow, M.T., Town, L., Ritchie, D.S., Colonna, M., Andrews, D.M., and Smyth, M.J. (2014). The receptors CD96 and CD226 oppose each other in the regulation of natural killer cell functions. Nat Immunol *15*, 431-438.

Chen, L., Ge, B., Casale, F.P., Vasquez, L., Kwan, T., Garrido-Martin, D., Watt, S., Yan, Y., Kundu, K., Ecker, S.*, et al.* (2016). Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune Cells. Cell *167*, 1398-1414 e1324.

Chester, C., Sanmamed, M.F., Wang, J., and Melero, I. (2018). Immunotherapy targeting 4-1BB: mechanistic rationale, clinical results, and future strategies. Blood *131*, 49-57.

Cho, M.H., McDonald, M.L., Zhou, X., Mattheisen, M., Castaldi, P.J., Hersh, C.P., Demeo, D.L., Sylvia, J.S., Ziniti, J., Laird, N.M.*, et al.* (2014). Risk loci for chronic obstructive pulmonary disease: a genome-wide association study and meta-analysis. Lancet Respir Med *2*, 214-225.

Deng, J.E., Sham, P.C., and Li, M.X. (2015). SNPTracker: A Swift Tool for Comprehensive Tracking and Unifying dbSNP rs IDs and Genomic Coordinates of Massive Sequence Variants. G3 (Bethesda) *6*, 205-207.

Fairfax, B.P., Humburg, P., Makino, S., Naranbhai, V., Wong, D., Lau, E., Jostins, L., Plant, K., Andrews, R., McGee, C.*, et al.* (2014). Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. Science *343*, 1246949.

Fauriat, C., Long, E.O., Ljunggren, H.G., and Bryceson, Y.T. (2010). Regulation of human NK-cell cytokine and chemokine production by target cell recognition. Blood *115*, 2167-2176.

Gondro, C., Porto-Neto, L.R., and Lee, S.H. (2014). SNPQC--an R pipeline for quality control of Illumina SNP genotyping array data. Anim Genet *45*, 758-761.

Grundberg, E., Small, K.S., Hedman, A.K., Nica, A.C., Buil, A., Keildson, S., Bell, J.T., Yang, T.P., Meduri, E., Barrett, A.*, et al.* (2012). Mapping cis- and trans-regulatory effects across multiple tissues in twins. Nature genetics *44*, 1084-1089.

GTEx Consortium (2015). Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science *348*, 648-660.

GTEx Consortium (2017). Genetic effects on gene expression across human tissues. Nature *550*, 204-213.

Guttman-Yassky, E., Brunner, P.M., Neumann, A.U., Khattri, S., Pavel, A.B., Malik, K., Singer, G.K., Baum, D., Gilleaudeau, P., Sullivan-Whalen, M.*, et al.* (2018). Efficacy and safety of fezakinumab (an IL-22 monoclonal antibody) in adults with moderate-to-severe atopic dermatitis inadequately controlled by conventional treatments: A randomized, double-blind, phase 2a trial. J Am Acad Dermatol *78*, 872-881 e876.

Hakonarson, H., Qu, H.Q., Bradfield, J.P., Marchand, L., Kim, C.E., Glessner, J.T., Grabs, R., Casalunovo, T., Taback, S.P., Frackelton, E.C.*, et al.* (2008). A novel susceptibility locus for type 1 diabetes on Chr12q13 identified by a genome-wide association study. Diabetes *57*, 1143-1146.

Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S.*, et al.* (2012). GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res *22*, 1760-1774.

Howie, B.N., Donnelly, P., and Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet *5*, e1000529.

Hubbard, L.L., and Moore, B.B. (2010). IRAK-M regulation and function in host defense and immune homeostasis. Infect Dis Rep *2*.

Keir, M.E., Butte, M.J., Freeman, G.J., and Sharpe, A.H. (2008). PD-1 and its ligands in tolerance and immunity. Annual review of immunology *26*, 677-704.

Kim, W.J., Lim, M.N., Hong, Y., Silverman, E.K., Lee, J.H., Jung, B.H., Ra, S.W., Choi, H.S., Jung, Y.J., Park, Y.B.*, et al.* (2014). Association of lung function genes with chronic obstructive pulmonary disease. Lung *192*, 473-480.

Klein, S.L., and Flanagan, K.L. (2016). Sex differences in immune responses. Nature reviews Immunology *16*, 626-638.

Lahiri, A., Hedl, M., Yan, J., and Abraham, C. (2017). Human LACC1 increases innate receptor-induced responses and a LACC1 disease-risk variant modulates these outcomes. Nat Commun *8*, 15614.

Lee, M.N., Ye, C., Villani, A.C., Raj, T., Li, W., Eisenhaure, T.M., Imboywa, S.H., Chipendo, P.I., Ran, F.A., Slowikowski, K.*, et al.* (2014). Common genetic variants modulate pathogen-sensing responses in human dendritic cells. Science *343*, 1246980.

Li, F.J., Won, W.J., Becker, E.J., Jr., Easlick, J.L., Tabengwa, E.M., Li, R., Shakhmatov, M., Honjo, K., Burrows, P.D., and Davis, R.S. (2014). Emerging roles for the FCRL family members in lymphocyte biology and disease. Curr Top Microbiol Immunol *382*, 29-50.

Liu, W., Li, H., Zhang, X., Wen, D., Yu, F., Yang, S., Jia, X., Cong, B., and Ma, C. (2013). Prostaglandin I2-IP signalling regulates human Th17 and Treg cell differentiation. Prostaglandins Leukot Essent Fatty Acids *89*, 335-344.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550.

Osinalde, N., Sanchez-Quiles, V., Blagoev, B., and Kratchmarova, I. (2016). Changes in Gab2 phosphorylation and interaction partners in response to interleukin (IL)-2 stimulation in T-lymphocytes. Sci Rep *6*, 23530.

Pan, F., Yu, H., Dang, E.V., Barbi, J., Pan, X., Grosso, J.F., Jinasena, D., Sharma, S.M., McCadden, E.M., Getnet, D.*, et al.* (2009). Eos mediates Foxp3-dependent gene silencing in CD4+ regulatory T cells. Science *325*, 1142-1146.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J.*, et al.* (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. American journal of human genetics *81*, 559-575.

Quach, H., Rotival, M., Pothlichet, J., Loh, Y.E., Dannemann, M., Zidane, N., Laval, G., Patin, E., Harmant, C., Lopez, M.*, et al.* (2016). Genetic Adaptation and Neandertal Admixture Shaped the Immune System of Human Populations. Cell *167*, 643-656 e617.

Raj, T., Rothamel, K., Mostafavi, S., Ye, C., Lee, M.N., Replogle, J.M., Feng, T., Lee, M., Asinovski, N., Frohlich, I.*, et al.* (2014). Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes. Science *344*, 519-523.

Ramos, E.M., Hoffman, D., Junkins, H.A., Maglott, D., Phan, L., Sherry, S.T., Feolo, M., and Hindorff, L.A. (2014). Phenotype-Genotype Integrator (PheGenI): synthesizing genome-wide association study (GWAS) data with existing genomic resources. Eur J Hum Genet *22*, 144-147.

Roadmap Epigenomics, C., Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J.*, et al.* (2015). Integrative analysis of 111 reference human epigenomes. Nature *518*, 317-330.

Rosales, S.L., Liang, S., Engel, I., Schmiedel, B.J., Kronenberg, M., Vijayanand, P., and Seumois, G. (2018). A Sensitive and Integrated Approach to Profile Messenger RNA from Samples with Low Cell Numbers. Methods Mol Biol *1799*, 275-301.

Rostamzadeh, D., Kazemi, T., Amirghofran, Z., and Shabani, M. (2018). Update on Fc receptor-like (FCRL) family: new immunoregulatory players in health and diseases. Expert Opin Ther Targets *22*, 487-502.

Sakaguchi, S., Miyara, M., Costantino, C.M., and Hafler, D.A. (2010). FOXP3+ regulatory T cells in the human immune system. Nature reviews Immunology *10*, 490-500.

Schmiedel, B.J., Seumois, G., Samaniego-Castruita, D., Cayford, J., Schulten, V., Chavez, L., Ay, F., Sette, A., Peters, B., and Vijayanand, P. (2016). 17q21 asthma-risk variants switch CTCF binding and regulate IL-2 production by T cells. Nat Commun *7*, 13426.

SEQC Consortium (2014). A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. Nature biotechnology *32*, 903-914.

Seumois, G., Zapardiel-Gonzalo, J., White, B., Singh, D., Schulten, V., Dillon, M., Hinz, D., Broide, D.H., Sette, A., Peters, B.*, et al.* (2016). Transcriptional Profiling of Th2 Cells Identifies Pathogenic Features Associated with Asthma. J Immunol *197*, 655-664.

Shabalin, A.A. (2012). Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics *28*, 1353-1358.

Sharma, M.D., Huang, L., Choi, J.H., Lee, E.J., Wilson, J.M., Lemos, H., Pan, F., Blazar, B.R., Pardoll, D.M., Mellor, A.L.*, et al.* (2013). An inherently bifunctional subset of Foxp3+ T helper cells is controlled by the transcription factor eos. Immunity *38*, 998-1012.

Sivasankar, B., Longhi, M.P., Gallagher, K.M., Betts, G.J., Morgan, B.P., Godkin, A.J., and Gallimore, A.M. (2009). CD59 blockade enhances antigen-specific CD4+ T cell responses in humans: a new target for cancer immunotherapy? J Immunol *182*, 5203-5207.

Sreeramkumar, V., Hons, M., Punzon, C., Stein, J.V., Sancho, D., Fresno, M., and Cuesta, N. (2016). Efficient T-cell priming and activation requires signaling through prostaglandin E2 (EP) receptors. Immunol Cell Biol *94*, 39-51.

Storey, J.D., Bass, A.J., Dabney, A., and Robinson, D. (2015). qvalue: Q-value estimation for false discovery rate control. R package version 2.6.0.

Takenaka, M.C., Robson, S., and Quintana, F.J. (2016). Regulation of the T Cell Response by CD39. Trends Immunol *37*, 427-439.

Tanaka, K., Yoshioka, K., Tatsumi, K., Kimura, S., and Kasuya, Y. (2014). Endothelin regulates function of IL-17-producing T cell subset. Life Sci *118*, 244-247.

Tarrant, J.M., Groom, J., Metcalf, D., Li, R., Borobokas, B., Wright, M.D., Tarlinton, D., and Robb, L. (2002). The absence of Tssc6, a member of the tetraspanin superfamily, does not affect lymphoid development but enhances in vitro T-cell proliferative responses. Mol Cell Biol *22*, 5006-5018.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105-1111.

Van Gorp, H., Kuchmiy, A., Van Hauwermeiren, F., and Lamkanfi, M. (2014). NOD-like receptors interfacing the immune and reproductive systems. FEBS J *281*, 4568-4582.

Visscher, P.M., Brown, M.A., McCarthy, M.I., and Yang, J. (2012). Five years of GWAS discovery. American journal of human genetics *90*, 7-24.

Wagner, G.P., Kin, K., and Lynch, V.J. (2012). Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. Theory Biosci *131*, 281-285.

Watts, T.H. (2005). TNF/TNFR family members in costimulation of T cell responses. Annual review of immunology *23*, 23-68.

Wehbi, V.L., and Tasken, K. (2016). Molecular Mechanisms for cAMP-Mediated Immunoregulation in T cells - Role of Anchored Protein Kinase A Signaling Units. Front Immunol *7*, 222.

Wolkers, M.C., Gerlach, C., Arens, R., Janssen, E.M., Fitzgerald, P., Schumacher, T.N., Medema, J.P., Green, D.R., and Schoenberger, S.P. (2012). Nab2 regulates secondary CD8+ T-cell responses through control of TRAIL expression. Blood *119*, 798-804.

Wright, F.A., Sullivan, P.F., Brooks, A.I., Zou, F., Sun, W., Xia, K., Madar, V., Jansen, R., Chung, W., Zhou, Y.H.*, et al.* (2014). Heritability and genomics of gene expression in peripheral blood. Nature genetics *46*, 430-437.

Wu, C., Yosef, N., Thalhamer, T., Zhu, C., Xiao, S., Kishi, Y., Regev, A., and Kuchroo, V.K. (2013). Induction of pathogenic TH17 cells by inducible salt-sensing kinase SGK1. Nature *496*, 513-517.

Zhang, L., Cheng, Z., Liu, W., and Wu, K. (2013). Expression of interleukin (IL)-10, IL-17A and IL-22 in serum and sputum of stable chronic obstructive pulmonary disease patients. COPD *10*, 459-465.

Zhernakova, D.V., Deelen, P., Vermaat, M., van Iterson, M., van Galen, M., Arindrarto, W., van 't Hof, P., Mei, H., van Dijk, F., Westra, H.J.*, et al.* (2017). Identification of context-dependent expression quantitative trait loci in whole blood. Nature genetics *49*, 139-145.

Zhou, L., Chong, M.M., and Littman, D.R. (2009). Plasticity of CD4+ T cell lineage differentiation. Immunity *30*, 646-655.

**MAIN FIGURE LEGENDS**

**Figure 1.** Transcriptome of common human immune cell types. (**A**) Study overview. (**B**) The pie chart shows relative proportions of the 13 primary cell types (details provided in **Fig. S1** and **Table S1A**). (**C**) RNA-Seq analysis of 1,544 samples (details provided in **Table S1B**) showing the total number of detected transcripts categorized based on their mean expression levels (TPM, transcript per million) in the indicated cell types and activation conditions. (**D**) The fraction of expressed transcripts belonging to each biotype as defined in the Ensembl database. For comparison, the GRCh37.p13 bar shows the corresponding fractions for all 57,820 annotated transcripts in the reference genome. (**E**) Unsupervised hierarchical clustering of the transcriptomes of the indicated cell types based on the 1,000 most variable transcripts expressed; each line represents an independent sample.

**Figure 2.** Cell-specific transcriptional signatures. (**A** and **C**) RNA-Seq analysis showing differentially expressed genes (**STAR Methods**), presented as row-wise *z*-scores of normalized TPM counts in the indicated cell types. (**B** and **D**) t-SNE plots of transcriptomes (symbol) for each of the indicated cell types for all subjects, and the expression of indicated cell-specific transcripts in each subject is shown in corresponding plots. Each symbol represents an independent sample; color scale represents the *z*-scores of normalized TPM counts in the indicated cell types.

**Figure 3.** Expression QTLs are highly cell-specific. (**A**) The fraction of eGenes belonging to each biotype as defined in the Ensembl database. (**B**) Genomic location of all analyzed SNPs and the *cis-*eQTLs identified in this study. (**C**) Distance of *cis-*eQTLs from the transcription start site (TSS) of the associated transcript (eGene); left panel shows the adj. association *P* value for all peak *cis-*eQTLs (**STAR Methods**) and the dotted line indicates adj. *P* < 0.05; right panel shows the fraction of *cis-*eQTLs identified across all cell types and activation conditions (*n* = 15) in the indicated regions up and downstream of the TSS, each dot represents a specific cell type. Error bars are mean ± SEM; \*\*\**P* < 0.001 by Student’s paired two-tailed t-test. (**D**) The total number of eGenes identified across the indicated cell types and activation conditions. (**E**) Overlap of eGenes identified in the indicated cell types (*n* = 6). The pie chart shows the fraction of eGenes identified in varying numbers of cell types. (**F**) Cell-specific eGene analysis, showing row-wise z-scores of the adj. association *P* values (left panel) and average expression levels (right panel) for cell-specific eGenes (one per row) in the indicated cell types. (**G**) Mean expression levels (TPM) of selected eGenes in the indicated cell types from subjects categorized based on the genotype at the indicated peak *cis-*eQTL; each symbol represents an individual subject; \* adj. association *P* value < 0.05. (**H**) Top panel, University of California Santa Cruz (UCSC) tracks showing chromosomal location and genes present in the 12p13.2 locus (containing a large haplotype block of *cis-*eQTLs (black lines); red line indicates location of peak *cis*-eQTL of *KLRK1* (rs2927561)); UCSC tracks of DNase hypersensitivity sites (provided by NIH Roadmap Epigenomics Mapping Consortium). Regions with active DHS in NK cells and CD8+ T cells, but not in the other cell types, are highlighted in blue. Bottom panel shows row-wise z-scores of TPM counts for the indicated eGenes in NK cells, where each column represents an individual subject. (**I**) Overlap of eGenes identified in resting and activated naïve CD4+ and CD8+ T cells. (**J**) Mean expression levels (TPM) of eGenes in resting and activated naïve CD4+ and CD8+ T cells from subjects categorized based on the genotype; each symbol represents an individual subject; \* adj. association *P* value < 0.05. (**K**) FACS analysis of surface expression of 4-1BB and CD25 in naïve CD4+ T cells following stimulation at the indicated time points. The subjects (*n* = 13) were categorized based on the genotype of the cis-eQTL associated with *TNFRSF9* (4-1BB) expression (C/C or T/T genotype at rs9657975); results of two-way ANOVA testing for significant differences between both groups are indicated for each surface marker.

**Figure 4.** Expression QTLs in CD4+ T cell subsets. (**A**) Venn diagram indicates overlap of eGenes from cell types and activation conditions discussed in **Fig. 3D** (*n* = 8) with CD4+ memory and TREG subsets (*n* = 7). (**B**) The total number of eGenes identified across the indicated cell types and activation conditions. (**C**) Overlap of eGenes identified in the indicated cell types (*n* = 8). The pie chart shows the relative cell specificity of the eGenes of the indicated cell types, the bar graph shows the fraction of eGenes identified in each cell type. (**D**) Cell-specific eGene analysis, showing row-wise z-scores of the adj. association *P* values (left panel) and average expression levels (right panel) for cell-specific eGenes (one per row) in the indicated cell types (*n* = 91 subjects); each column represents a cell type. (**E**) Mean expression levels (TPM) of selected eGenes in the indicated cell types from subjects categorized based on the genotype at the indicated peak *cis-*eQTL; each symbol represents an individual subject; \* adj. association *P* value < 0.05.

**Figure 5.** Cell types susceptible to GWAS SNPs. (**A**) Fraction of eGenes overlapping with significant SNPs (*PGWAS* < 5 x 10-8) that emerged from the catalogue of GWAS studies (GWAS SNPs) comprising 540 unique human diseases and traits (**Table S4** and **STAR Methods**). (**B**) The total number of eGenes identified across the indicated cell types and activation conditions. (**C**) For each disease, the adj. association *P* value for the peak *cis-*eQTL with the indicated GWAS eGenes, excluding HLA genes, in each cell type and activation condition is shown. Example eGenes discussed in the results are printed in bold. (**D**) Mean expression levels (TPM) of *LACC1* in the indicated cell types and activation conditions from subjects categorized based on the genotype at the *cis-*eQTL rs9567293; each dot represents an individual subject; \* adj. association *P* value < 0.05. (**E**) Real-time PCR quantification of the effects of *LACC1* knockdown on the induction of *IFNG* and *IL2* expression in naive CD4+ T cells activated for 6 hours with antibodies to CD3 and CD28 in absence or presence of the TLR1/2 ligand Pam3CSK4 (*n* = 5); \**P* < 0.05 by Student’s paired two-tailed t-test; NS, not significant. (**F**) Mean expression levels (TPM) of *NAB1* and *SYNGR1* in the indicated cell types and activation conditions from subjects categorized based on the genotype at the indicated peak *cis-*eQTL; each dot represents an individual subject; \* adj. association *P* value < 0.05. (**G**) Effects of *NAB1* and *SYNGR1* knockdown on the release of IFN-γ by NK cells activated for 24 hours with coated human IgG (*n* = 8); \**P* < 0.05 by Wilcoxon matched-pairs signed-rank test. (**H**) Effects of *NAB1* and *SYNGR1* knockdown on the release of IFN­-γ by NK cells co-cultured for 24 hours with MHC devoid K-562 target cells (*n* = 8); \**P* < 0.05 by Wilcoxon matched-pairs signed-rank test. (**I**) Mean expression levels (TPM) of *IKZF4* and *PDEA4* in the indicated cell types from subjects categorized based on the genotype at the indicated peak *cis-*eQTL; each symbol represents an individual subject; \* adj. association *P* value < 0.05.

**Figure 6**. Sex has major effects on gene expression. (**A**) RNA-Seq analysis of differentially expressed transcripts (sex-biased transcripts, each dot) in each immune cell type from female (*n* = 37) versus male (*n* = 54) subjects (**STAR Methods**). Right, fractions of sex-biased transcripts in autosomes, sex chromosomes and mitochondrial DNA. (**B**) Total number of sex-biased transcripts identified across all cell types and activation conditions. (**C**) Overlap of sex-biased transcripts in the indicated cell types (*n* = 15). The pie chart shows the relative cell specificity of the transcripts in the indicated cell types, the bar graph shows the fraction of sex-biased transcripts identified per cell type. (**D**) Ingenuity pathway analysis of sex-biased transcripts identified in classical monocytes that are regulated by interferon-γ and encode products with various functions (key indicates product type). (**E**) Column-wise z-scores of TPM counts for transcripts in the interferon (IFN, left panel), interferon regulatory factor (IRF, middle panel) and toll-like receptor (TLR, right panel) response pathways expressed in classical monocytes; each row represents an individual subject. (**F**) Mean expression levels (TPM) of sex-biased transcript *FAM13A* in the indicated cell types from female and male subjects; each symbol represents an individual subject; \* adj. *P* value < 0.05. (**G**) Mean expression levels (TPM) of *FAM13A* in naïve CD4+ T cells and classical monocytes from female and male subjects categorized based on genotype at the indicated peak *cis-*eQTL; each symbol represents an individual subject; \* adj. association *P* value < 0.05. (**H**) RNA-Seq analysis of genes (one per row, row-wise *z*-scores of normalized TPM counts shown) expressed differentially by *ex vivo* activated hCMV-specific CD4+ T cells following *FAM13A* siRNA knockdown versus those treated with siControl (adj. *P* value < 0.1 (DESeq2 analysis; Benjamini-Hochberg test)). Each column represents an individual biological replicate sample (siControl, *n* = 12; siFAM13A, *n* = 11); the color-code on top refers to the specific donors (*n* = 4). (**I**) Representative FACS plots showing intracellular staining of IL-2 and IL-22 in memory CD4+ T cells activated for 6 hours with PMA and Ionomycin (after knockdown with siRNA pool for control siRNA or *FAM13A*); frequencies of cytokine producing cells for IL-22, IFN-γ and IL-2 in each donor are shown below (*n* = 10); \**P* < 0.05 by Wilcoxon matched-pairs signed-rank test; NS, not significant. (**J**) Mean expression levels (TPM) of sex-biased transcripts in the indicated cell types from female and male subjects; each symbol represents an individual subject; \* adj. *P* value < 0.05.

**SUPPLEMENTAL FIGURE LEGENDS**

**Figure S1.** Sorting strategies, immune cell types isolated and additional data for RNA-Seq analysis. Related to Figures 1 and 2. (**A-D**) Representative FACS plots describing the sorting strategies used for the isolation of the indicated immune cell types using four different FACS panels from peripheral blood mononuclear cells (PBMC). (**E**) Surface markers used for isolation of the indicated cell types (details in **Table S6A**). (**F**) Relative proportions of the indicated immune cell populations in PBMC samples (*n* = 91 each); each symbol represents an individual subject. (**G**) RNA-Seq analysis of 1,544 samples (details in **Table S1B**) showing the total number of detected protein-coding transcripts and lncRNAs categorized based on their mean expression levels (TPM, transcript per million) in the indicated cell types and activation conditions. (**H**) The pie charts show the relative cell specificity of the expressed transcripts, globally (left) and according to the biotype as defined in the Ensembl database. (**I**) Overlap of all expressed transcripts identified across cell types and conditions (*n* = 15). The bar graph shows the fraction of transcripts identified per cell type.

**Figure S2.** eQTLs and activation-induced QTLs in innate and naïve cells types. Related to Figure 3. (**A**) Extended data set of **Fig. 3G**; mean expression levels (TPM) of selected eGenes in the indicated cell types from subjects categorized based on the genotype at the indicated peak *cis-*eQTL; each symbol represents an individual subject; \* adj. association *P* value < 0.05. (**B**) Real-time PCR quantification of *GAB2* transcript levels (relative to the housekeeping gene *YWHAZ*) in naïve CD4+ T cells 48 hours after knockdown with control siRNA or *GAB2* siRNA pools (*n* = 4); \**P* < 0.05 by Student’s paired two-tailed t-test. (**C**) FACS analysis of the effects of *GAB2* knockdown of the surface expression of the activation markers CD69, CD25 and HLA-DR in naive CD4+ T cells activated for 48 hours with antibodies to CD3 and CD28; \**P* < 0.05 by Student’s paired two-tailed t-test. (**D**) Real-time PCR quantification of the effects of *GAB2* knockdown on the expression of *IFNG*, *IL2* and *TNF* transcripts in naive CD4+ T cells activated for 6 hours with antibodies to CD3 and CD28; \**P* < 0.05 by Student’s paired two-tailed t-test; NS, not significant. (**E**) Mean expression levels (TPM) of selected eGenes in the indicated cell types from subjects categorized based on the genotype at the indicated peak *cis-*eQTL; each symbol represents an individual subject; \* adj. association *P* value < 0.05. (**F**) FACS analysis of the surface expression of CD69 and HLA-DR in naïve CD4+ T cells following anti-CD3 and anti-CD28 stimulation at the indicated time points. The subjects (*n* = 13) were classified based on the genotype of the cis-eQTL associated with modulated *TNFRSF9* (4-1BB) expression (C/C or T/T genotype at rs9657975); results of two-way ANOVA testing for significant differences between both groups are indicated for each surface marker.

**Figure S3.** Cell types most susceptible to GWAS SNPs and role of GWAS eGenes in cellular functions. Related to Figure 5. (**A**) Overlap of GWAS eGenes across all cell types and activation conditions (*n* = 15). The pie chart shows the relative cell specificity of the GWAS eGenes, the bar graph shows the fraction of GWAS eGenes identified per cell type. (**B**) Mean expression levels (TPM) of selected eGenes in the indicated cell types from subjects categorized based on the genotype at the indicated peak *cis-*eQTL; each symbol represents an individual subject; \* adj. association *P* value < 0.05. (**C**) Extended data set of **Fig. 5D** and **Fig. 5F**; mean expression levels (TPM) of selected eGenes in the indicated cell types from subjects categorized based on the genotype at the indicated peak *cis-*eQTL; each symbol represents an individual subject; \* adj. association *P* value < 0.05. (**D**) Real-time PCR quantification of *LACC1* transcript levels (relative to the housekeeping gene *YWHAZ*) in naïve CD4+ T cells 48 hours after knockdown with control siRNA or *LACC1* siRNA pools (*n* = 5); \**P* < 0.05 by Student’s paired two-tailed t-test. (**E**) FACS analysis of surface expression of activation markers CD69, CD25 and HLA-DR in naive CD4+ T cells activated for 48 hours with antibodies to CD3 and CD28 in absence or presence of the TLR1/2 ligand Pam3CSK4 (*n* = 4) after knockdown with control siRNA or *LACC1* siRNA pools. \**P* < 0.05 by Student’s paired two-tailed t-test; NS, not significant. (**F**) Real-time PCR quantification of the effects of *LACC1* knockdown on the expression of *TNF* transcript levels in naive CD4+ T cells activated for 6 hours with antibodies to CD3 and CD28 in absence or presence of the TLR1/2 ligand Pam3CSK4 (*n* = 5); NS, not significant by Student’s paired two-tailed t-test. (**G**) Real-time PCR quantification of *NAB1* and *SYNGR1* transcript levels (relative to the housekeeping gene *YWHAZ*) in NK cells 48 hours after knockdown with control siRNA, *NAB1* or *SYNGR1* siRNA pools (*n* = 8); \**P* < 0.05 by Student’s paired two-tailed t-test. (**H**) Effects of *NAB1* and *SYNGR1* knockdown on the release of MIP-1, RANTES and TNF by NK cells activated for 24 hours with coated human IgG (*n* = 8); \**P* < 0.05 by Wilcoxon matched-pairs signed-rank test; NS, not significant.

**Figure S4.** Cell types most susceptible to GWAS SNPs. Related to Figure 5. For each disease, the adj. association *P* value for the peak *cis-*eQTL with the indicated GWAS eGenes, excluding HLA genes, in each cell type and activation condition is shown. Example eGenes discussed in the results are printed in bold.

**Figure S5.** Cell types most susceptible to GWAS SNPs. Related to Figure 5. (**A** and **B**) Mean expression levels (TPM) of selected GWAS eGenes in the indicated cell types from subjects categorized based on the genotype at the indicated peak *cis-*eQTL; each symbol represents an individual subject, \* adj. association *P* value < 0.05.

**Figure S6.** Determination of sex-biased gene expression in immune cells. Related to Figure 6. (**A**) Cell-specific analysis of sex-biased transcripts, showing row-wise z-scores of the adj. *P* values (left panels), fold change in expression (middle panels) and average expression levels (right panels) for transcripts (one per row) in the indicated cell types (*n* = 91 subjects); each column represents a cell type. (**B**) Extended data set of **Fig. 6F**; mean expression levels (TPM) of *FAM13A* expression in the indicated cell types from female and male subjects; each symbol represents an individual subject. (**C**) Extended data set of **Fig. 6G**; mean expression levels (TPM) of *FAM13A* in the indicated cell types and activation conditions from female and male subjects categorized based on genotype of the peak cis-eQTL (rs1246642); each symbol represents an individual subject; \* adj. *P* value < 0.05. (**D**) Experimental design to assess the effects of knocking down *FAM13A* in CD4+ T cells followed by stimulation with a hCMV peptide pool and FACS sorting for analysis by RNA-Seq. (**E**) Real-time PCR quantification of *FAM13A* transcript levels (relative to the housekeeping gene *YWHAZ*) in CD4+ T cells 48 hours after knockdown with control siRNA or FAM13A siRNA pools (*n* = 4); \**P* < 0.05 by Student’s paired two-tailed t-test. (**F**) FACS plots describing the sorting strategy used for enrichment of CD154+ hCMV antigen-responding, non-naïve CD4+ T cells activated for 6 hours with a hCMV peptide pool (after knockdown with control siRNA or FAM13A siRNA pools) in presence of autologous PBMC to allow optimal antigen presentation (PBMC pre-stained with CellTrace Violet for exclusion). (**G**) Real-time PCR quantification of *FAM13A* transcript levels (relative to the housekeeping gene *YWHAZ*) in memory CD4+ T cells 48 hours after knockdown with control siRNA or *FAM13A* siRNA pools (*n* = 8); \**P* < 0.05 by Student’s paired two-tailed t-test. (**H**) Representative FACS plots showing intracellular staining of IFN-γ, IL-17A, IL-2, and TNF in memory CD4+ T cells activated for 6 hours with PMA and Ionomycin (after knockdown with siRNA pool for control siRNA or *FAM13A*). (**I**) Percentage of cytokine producing cells for IL-17A and TNF in each donor is shown (*n* = 10); NS, not significant by Wilcoxon matched-pairs signed-rank test. (**J**) Extended data set of **Fig. 6J**; mean expression levels (TPM) of the indicated genes in the indicated cell types and activation conditions from female and male subjects categorized based on genotype at the indicated peak cis-eQTL; each symbol represents an individual subject; \* adj. *P* value < 0.05.

**STAR METHODS**

**CONTACT FOR REAGENT AND RESOURCE SHARING**

Further information and requests for reagents may be directed to the corresponding author/lead contact, Pandurangan Vijayanand (vijay@lji.org).

**EXPERIMENTAL MODEL AND SUBJECT DETAILS**

The Institutional Review Board (IRB) of the La Jolla Institute for Allergy and Immunology (LJI; IRB protocol no. SGE-121-0714) approved the study. A total of 91 healthy volunteers were recruited in the San Diego area, who provided leukapheresis samples at the San Diego Blood Bank (SDBB) after written informed consent. The cohort consisted of 54 males and 37 female subjects with a median age of 27 years (range 18 to 61 years). All study subjects self-reported ethnicity and race details, and were tested negative for hepatitis B, hepatitis C and human immunodeficiency virus (HIV). Details of the study subjects are provided in **Table S1A**.

**METHOD DETAILS**

**Sample processing**

Peripheral blood mononuclear cells (PBMC) were obtained from leukapheresis samples by density gradient centrifugation according to the manufacturer’s instructions and cryopreserved in liquid nitrogen. For the isolation of immune cell types of interest, cryopreserved PBMCs were thawed, washed, stained directly with cocktails of fluorescently conjugated antibodies or pre-enriched for total B cells or memory CD4+ T cells using the ‘Human B Cell Isolation Kit II’ (Miltenyi Biotec; Cat.No. 130-091-151) or the ‘Memory CD4+ T Cell Isolation Kit’ (Miltenyi Biotec; Cat.No. 130-091-893), respectively, following the manufacturer’s instructions before staining with antibodies in four different FACS panels (**Fig. S1** and **Table S6A**) and sorted on a BD FACSAria II (Becton Dickinson) using the gating strategies shown in **Fig. S1A-D**; flow cytometry data were analyzed using FlowJo software (version 10). The FACS-sorted cells were washed and lysed in TRIzol solution (Invitrogen) for subsequent isolation of total RNA.

For activation conditions, FACS-sorted naïve CD4+ and CD8+ T cells were resuspended at a concentration of 1x106/mL in 1 mL of pre-warmed IMDM medium, supplemented with 5% (vol/vol) heat-inactivated fetal bovine serum (FBS) and 2% (vol/vol) human AB serum (CellGro), and activated *ex vivo* with GibcoTM DynabeadsTM Human T-Activator CD3/CD28 (ThermoFisher Scientific) at a bead-to-cell ratio of 1:1 for 4 hours at 37°C.

**RNA sequencing**

Total RNA was purified using the miRNeasy Micro Kit (Qiagen), quantified as described previously (Seumois et al., 2016), and quality assessed using the High Sensitivity RNA ScreenTape Kit (Agilent Life Sciences). Uniquely indexed Illumina sequencing libraries were generated from 180-600 ng of purified total RNA per sample using the TruSeq Stranded mRNA HT Sample Prep Kit (Illumina). Sequencing library preparation steps were performed in a 96-well format, utilizing the automated platform Biomek FXP (Beckman Coulter), to reduce assay-to-assay variability. Samples failing any quality control step (RNA quality and quantity, PCR amplification, library fragment size) were eliminated from further downstream steps. Libraries were sequenced on the HiSeq 2500 Illumina platform to obtain 50-bp single-end reads (HiSeq SBS Kit v4; Illumina), generating a total of over 33.1 billion uniquely mapped reads (median of 20.1 million uniquely mapped reads per sample) (**Table S1B**).

**RNA-Seq analysis**

RNA-Seq data were mapped against the hg19 reference genome using TopHat v1.4.1 (Trapnell et al., 2009) and GENCODE annotation v19 (GRCh37.p13) (Harrow et al., 2012) as gene reference model (including a total of 57,820 annotated transcripts, sub-classified into 20,776 protein-coding genes, 13,870 long non-coding RNAs (lncRNA), 9,013 short non-coding RNAs (sncRNA) and 14,161 pseudogenes). Sequencing read coverage per gene was counted using HTSeq (https://htseq.readthedocs.io) (Anders et al., 2015) and transcript abundance expressed as transcripts per million (TPM) units (Wagner et al., 2012). Transcripts with mean TPM < 1.0 across all samples of a given cell type were excluded from the analysis. The identity of all individual RNA-Seq samples was confirmed by cross-comparison with genotyping data for a given subject by genome-wide analysis of SNPs overlapping exons. We observed no batch effects by standard principal component analysis and multidimensional scaling analysis. To diminish the effects of possible outliers, the regularized log transformation of the DESeq2 package (Love et al., 2014) was applied and the homoscedastic data were then quantile-normalized. Further, samples from each specific cell types were processed for RNA-Seq in one batch to avoid any experimental batch effects.

Differentially expressed (DE) genes across cell types (and activation conditions) were identified by pairwise comparisons using DESeq2 (Love et al., 2014); we considered a transcript to be expressed differentially by any comparison when the DESeq2 analysis resulted in a Benjamini-Hochberg adj. *P* value of < 0.05 and fold change (FC) ≥ 2. Sex-biased transcripts were identified by performing pairwise comparison (female (*n* = 37) *versus* male (*n* = 54) subjects) using DESeq2 and considering ancestry and age as covariates; we considered a transcript expressed differentially when the DESeq2 analysis resulted in a Benjamini-Hochberg adj. *P* value of < 0.05. Shared expression of transcripts across cell types was determined using the jvenn diagram viewer (http://jvenn.toulouse.inra.fr) (Bardou et al., 2014).

**Clustering, statistical analysis and data display**

Unbiased hierarchical clustering (HC) was performed to determine gene expression similarities of single samples across cell types using the TPM values in a log2-transformed scale (a value of 1 was added prior to transformation to account for values = 0). Distance metrics were obtained based on Spearman correlation as D = 1 - abs(spearman\_corr); the average method was used for hierarchical clustering using the 1,000 most variable transcripts expressed. Spearman correlation coefficient (r value) was calculated to assess significance of correlations between the levels of any two components of interest. Principal component analysis (PCA) was performed on the entire matrix of expressed transcripts and a set of first principal components were used to perform t-SNE dimensional reduction. GraphPad Prism 7.0d software was used for generating graphs and performing statistical significance tests. The R package with custom scripts was utilized for generating dendrograms, t-SNE plots, heat maps and volcano plots.

**Genotyping, imputation and determination of eQTL**

Genomic DNA was isolated from PBMC using the DNeasy Blood & Tissue Kit (Qiagen) and utilized for genotyping using the Infinium Multi-Ethnic Global-8 Kit (Illumina) following the manufacturer’s instructions. Raw data from the genotyping analysis was exported using GenomeStudio v2011.1 (Illumina). The data quality was assessed using the snpQC package (Gondro et al., 2014) for R v3.2.3 (http://www.R-project.org) and low-quality SNPs were removed for downstream analysis: SNPs failing in more than 5% of the samples and SNPs with Illumina’s GC scores less than 0.2 in more than 10% of the samples were filtered out. Sex, ethnicity and relatedness of the subjects was inferred from the genotype data using PLINK v1.90b3w (Purcell et al., 2007). IMPUTE v2.3.2 (Howie et al., 2009) was used to phase observed genotypes and impute SNPs using the 1,000 Genomes Project phase 3 (1000 Genomes Project Consortium, 2015) as the reference panel, resulting in >82 million imputed SNPs (~84.5 million SNPs in total; about 5.8 million SNPs were utilized in eQTL analysis). Notably, the eQTL analysis did not cover the pseudoautosomal regions of chr X or chr Y, thereby slightly lowering the annotated transcript count from 57,820 to 57,215 (605 missing genes).

To identify eQTLs, the full set of SNPs (measured and imputed) was tested for association with the expression of sequenced genes using the Matrix eQTL package v2.1.1 (Shabalin, 2012). Briefly, a linear model was fitted for the expression of each transcript for the underlying genotype of all SNPs located within a window around the annotated transcription start site (TSS) of a given gene (±1 Mb of TSS). The first two principal components of the genotyping data were used as covariates to minimize possible differences introduced by the different ancestry in the samples. SNPs missing for more than 5% of the samples or with a minimum allele frequency (MAF) lower than 5% were removed from the analysis. The gene expression data were processed utilizing an approximation (Ballouz et al., 2015) of the criteria established in the Sequencing Quality Control (SEQC) project by the SEQC/MAQC-III Consortium (SEQC Consortium, 2014). Briefly, transcripts with low expression values – the bottom one-third of all transcripts for at least 80% of the samples – were filtered out. The regularized log transformation of the DESeq2 package (Love et al., 2014) was applied and the homoscedastic data were then quantile-normalized to diminish the effects of possible outliers. For each transcript, the associated SNP with the lowest *P* value was selected and utilized to correct for multiple comparisons. The permutation-based *P* values were calculated by random shuffling of the sample labels and test statistics were then recalculated as described above. A null distribution was built and corrected *P* values, Q values and false discovery rate (FDR; adj. association *P* value) were calculated using the qvalue package (http://github.com/jdstorey/qvalue) (Storey et al., 2015). For the determination of *cis-*acting eQTL and affected transcripts and their comparison across cell types, the following thresholds were applied: FDR < 0.05, raw *P* value < 0.0001 and TPM > 1.0. If not stated otherwise, the associated SNP (called as peak SNP) with the lowest FDR and *P* value closest to the TSS was chosen for a given transcript.

**GWAS SNP database construction and analysis**

To define the collection of SNPs from GWAS studies, we downloaded the dataset of the Phenotype-Genotype Integrator (PheGenI) (Ramos et al., 2014), which merges NHGRI GWAS Catalog data with several databases of the National Center for Biotechnology Information (NCBI), including Gene, dbGaP, OMIM, eQTL and dbSNP (September 2017; https://www.ncbi.nlm.nih.gov/gap/phegeni). SNP positions were back-translated from PheGenI’s GRCh38 annotation to GRCh37.p13 using SNPTracker (Deng et al., 2015), and the resulting catalog of SNP-trait associations included 14,321 significant associations (lead GWAS SNPs; defined as *P* < 5 x 10-8) and 846 distinct phenotypes/traits.

Linkage disequilibrium (LD) was calculated using PLINK v1.90b3w (Purcell et al., 2007) for continental ‘super populations’ (AFR, AMR, EAS, EUR, SAS) based on data from the phase 3 of the 1,000 Genomes Project (1000 Genomes Project Consortium, 2015). SNPs in tight genetic linkage with GWAS lead SNPs (LD threshold r2 > 0.8) in any of the five super-populations were retrieved along with the SNP information (e.g. genomic location, allelic variant, allele frequencies). Utilizing this data set, GWAS SNPs (lead SNPs and SNPs in LD) were analyzed for overlap with *cis-*eQTLs (FDR < 0.05, raw *P* value < 0.0001 and TPM > 1.0) separately for each cell type to identify transcripts regulated by SNPs in haploblocks containing GWAS lead SNP(s).

**Knowledge-based network generation and pathway analysis**

The biological relevance of differentially expressed transcripts identified by DESeq2 analysis was investigated using the Ingenuity Pathways Analysis (IPA) platform. The enrichment of canonical pathways (pre-defined, well-described metabolic and signaling pathways curated from literature reviews) among these transcripts was assessed, with significance determined by right-tailed Fisher's exact test (*P* < 0.05). For network analysis, differentially expressed transcripts were progressively linked together on the basis of a measure of their interconnection, derived from previously characterized functional interactions.

**Validation of gene function in primary immune cells**

4-1BB surface expression studies: Naïve CD4+ T cells were isolated from PBMC using the ‘Naive CD4+ T Cell Isolation Kit II’ (Miltenyi Biotec; Cat.No.: 130-094-131). Subsequently, the naïve CD4+ T cells were stimulated with pre-coated anti-CD3 antibodies (10 µg/mL) and soluble anti-CD28 antibodies (1 µg/mL) in presence of IL-7 (5 ng/mL; Miltenyi Biotec). The expression of surface markers (4-1BB, CD25, CD69 and HLA-DR) was analyzed by FACS at various time points); see **Table S6B** for details.

All siRNA knockdown studies were performed as previously described (Schmiedel et al., 2016). Primary immune cells were negatively pre-enriched for the cell populations of interest by MACS-based isolation (Miltenyi Biotec) following the manufacturer’s instructions, and cultured in Iscove's Modified Dulbecco's Medium (IMDM; Invitrogen) supplemented with 5% (vol/vol) heat-inactivated fetal bovine serum (FBS) and 2% (vol/vol) human AB serum (CellGro).

*GAB2* and *LACC1* knockdown studies: Naive CD4+ T cells were transfected with 0.5 nmoL (per 1 x 106 cells) of siRNA pools specific for *GAB2* and *LACC1* or non-targeting siRNA (ON-TARGETplus SMARTpools from Dharmacon) using the Neon Transfection System (Invitrogen) according to the manufacturer's protocol (settings: 2,200 V, 10 ms, 3 pulses). Knockdown efficiency was analyzed 48 hours after transfection by real-time PCR for transcript levels and the viability of cells was analyzed by FACS-based Annexin-V and DAPI staining (data not shown). Subsequently, the siRNA-transfected naïve CD4+ T cells were stimulated with pre-coated anti-CD3 antibodies (3 µg/mL), soluble anti-CD28 antibodies (1 µg/mL) and IL-7 (5 ng/mL; Miltenyi Biotec), in presence or absence of the TLR1/2 ligand Pam3CSK4 (100 ng/mL; Invivogen). Transcript levels of cytokines were analyzed by real-time PCR after 6 hours of stimulation, the surface expression of activation markers was measured by FACS after 48 hours of stimulation. Total RNA was extracted using the miRNeasy Micro Kit (Qiagen); cDNA was reverse-transcribed with the SuperScript III First-Strand Synthesis System (Life Technologies). Real-time PCR was employed the Fast Start Universal SYBR Green Master Mix (Roche); see **Table S6C** for primer sequences. Data were acquired on the QuantStudio 6 Flex (Applied Biosystems); all results are presented in arbitrary units relative to expression of the housekeeping gene *YWHAZ*.

*NAB1* and *SYNGR1* knockdown studies: NK cells were isolated from PBMC using the ‘NK Cell Isolation Kit’ (Miltenyi Biotec; Cat.No.: 130-092-657) and transfected with 0.5 nmoL (per 1 x 106 cells) of siRNA pools specific for *NAB1* and *SYNGR1* or non-targeting siRNA (ON-TARGETplus SMARTpools from Dharmacon) using the Neon Transfection System (Invitrogen) according to the manufacturer's protocol (settings: 1,900 V, 10 ms, 3 pulses). Knockdown efficiency was analyzed as described above. Subsequently, the siRNA-transfected NK cells were stimulated with pre-coated human IgG (3 µg/mL) or K-562 target cells (cell:cell ratio 1:1). The cytokine secretion was measured after 24 hours of stimulation with U-PLEX assay kits (Meso Scale Discovery) following the manufacturer’s instructions. The degranulation of lytic granules by NK cells after culture with K-562 target cells for 3 hours was assessed by surface staining for CD107a.

*FAM13A* knockdown studies: Memory CD4+ T cells were isolated from PBMC using the ‘Memory CD4+ T Cell Isolation Kit’ (Miltenyi Biotec; Cat.No. 130-091-893) and transfected with 0.5 nmoL (per 1 x 106 cells) of siRNA pools specific for *FAM13A* or non-targeting siRNA (ON-TARGETplus SMARTpools from Dharmacon) using the Neon Transfection System (Invitrogen) according to the manufacturer's protocol (settings: 2,200 V, 10 ms, 3 pulses). Knockdown efficiency was analyzed as above. Subsequently, the siRNA-transfected memory CD4+ T were intracellularly stained (ICS) for cytokine detection after stimulation with PMA and Ionomycin for 6 hours; Brefeldin A (5 µg/mL; Sigma-Aldrich) was added for the final two hours of culture. ICS assay was performed following the instructions of BioLegend’s ‘Intracellular flow cytometry staining protocol’.

Screening for effects of *FAM13A* knockdown in antigen-specific T cells: The activation and FACS-based enrichment of antigen-specific CD4+ T cells from PBMC was previously described (Bacher et al., 2016), modifications to the protocol are outlined in **Fig. S6D**. In brief, CD4+ T cells were isolated from PBMC using the ‘CD4+ T Cell Isolation Kit’ (Miltenyi Biotec; Cat.No. 130-096-533) and transiently transfected with 0.5 nmoL (per 1 x 106 cells) of siRNA pool specific for *FAM13A* or non-targeting siRNA (ON-TARGETplus SMARTpools from Dharmacon) using the Neon Transfection System (Invitrogen) according to the manufacturer's protocol (settings: 2,200 V, 10 ms, 3 pulses). Knockdown efficiency was analyzed as described above. 48 hours after transfection, siRNA-transfected CD4+ T cells were pooled with autologous PBMC (stained with CellTrace Violet (ThermoFisher Scientific) to discriminate from siRNA-transfected CD4+ T cells) and rested for 2 hours. Subsequently, PepTivator hCMV pp65 (Miltenyi Biotec) was added and the cells were cultured for 6 hours in presence of a blocking anti-CD40 antibody (Miltenti Biotec). For FACS-based enrichment of activated hCMV-specific CD4+ T cells, cells were stained for the activation markers CD154 and 4-1BB. For each replicate, 400 CD154+ (marker of antigen-specific activation) non-naïve CD4+ T cells were FACS-sorted directly into lysis buffer using a BD FACSAria Fusion (Becton Dickinson); the gating strategy is shown in **Fig. S6F**. RNA isolation and preparation of RNA-Seq libraries was done following the low-input protocol described previously (Rosales et al., 2018). Differentially expressed (DE) genes across samples from treatment with siRNA pool for *FAM13A* or non-targeting siRNA were identified by pairwise comparisons of all samples including replicates, using DESeq2 (Love et al., 2014); we considered a transcript to be expressed differentially in a comparison when the DESeq2 analysis resulted in a Benjamini-Hochberg adj. *P* value of < 0.1.

**QUANTIFICATION AND STATISTICAL ANALYSIS**

Processing of data, applied methods and codes are described in the respective section in the **STAR Methods**. The number of subjects, samples and replicates analyzed, and the statistical test performed are indicated in the figure legends.

Any samples failing established quality controls (QC) were eliminated from downstream analysis; assay details and QCs on RNA-Seq samples included are provided in **Table S1B** (1,544 samples in total). Differentially expressed (DE) genes across cell types (and activation conditions) were identified by pairwise comparisons using DESeq2; we considered a transcript to be expressed differentially by any comparison when the DESeq2 analysis resulted in a Benjamini-Hochberg adj. *P* value of < 0.05 and fold change (FC) ≥ 2. Sex-biased transcripts were identified by performing pairwise comparison (female (*n* = 37) *versus* male (*n* = 54) subjects) using DESeq2 and considering ancestry and age as covariates; we considered a transcript expressed differentially when the DESeq2 analysis resulted in a Benjamini-Hochberg adj. *P* value of < 0.05.

To identify eQTLs, the full set of SNPs (measured and imputed) was tested for association with the expression of sequenced genes using the Matrix eQTL package v2.1.1. For the determination of *cis-*acting eQTL and affected transcripts and their comparison across cell types, the following thresholds were applied: FDR < 0.05, raw *P* value < 0.0001 and TPM > 1.0. If not stated otherwise, the associated SNP (called as peak SNP) with the lowest FDR and *P* value closest to the TSS was chosen for a given transcript.

Statistical analysis for comparison between two groups was assessed with Student’s paired two-tailed t-test or Wilcoxon matched-pairs signed-rank test using GraphPad Prism 7.0d. The two-way ANOVA test was utilized for the time course experiment.

**DATA AND SOFTWARE AVAILABILITY**

The DICE project provides anonymized data for public access at http://dice-database.org. Individual–specific sequencing and genotype data was submitted to the database of Genotypes and Phenotypes (dbGaP). The accession number for data reported in this paper is: (To be provided).

**SUPPLEMENTAL TABLES**

**Table S1.** Subject characteristics & RNA-Seq metrics. Related to Figure 1.

**Table S2.** Cell-specific gene lists. Related to Figure 2.

**Table S3.** eQTLs identified in all cell types & overlap of eGenes with previous eQTL studies. Related to Figures 3 and 4.

**Table S4.** GWAS eGenes identified in all cell types. Related to Figure 5.

**Table S5.** Sex-biased transcripts & IPA analysis of sex-biased transcripts. Related to Figure 6.

**Table S6.** FACS antibodies and reagents. Related to STAR Methods.