

RESEARCH ARTICLE

# Pathogenic implications for autoimmune mechanisms derived by comparative eQTL analysis of CD4<sup>+</sup> versus CD8<sup>+</sup> T cells

Silva Kasela<sup>1,2</sup>, Kai Kisand<sup>3</sup>, Liina Tserel<sup>3</sup>, Epp Kaleviste<sup>2,3</sup>, Anu Remm<sup>3</sup>, Krista Fischer<sup>1</sup>, Tõnu Esko<sup>1</sup>, Harm-Jan Westra<sup>4,5,6</sup>, Benjamin P. Fairfax<sup>7</sup>, Seiko Makino<sup>7</sup>, Julian C. Knight<sup>7</sup>, Lude Franke<sup>8</sup>, Andres Metspalu<sup>1,2</sup>, Pärt Peterson<sup>3\*</sup>, Lili Milani<sup>1\*</sup>

**1** Estonian Genome Center, University of Tartu, Tartu, Estonia, **2** Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia, **3** Institute of Biomedicine and Translational Medicine, University of Tartu, Tartu, Estonia, **4** Divisions of Genetics and Rheumatology, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, United States of America, **5** Partners Center for Personalized Genetic Medicine, Boston, MA, United States of America, **6** Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, United States of America, **7** Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom, **8** Department of Genetics, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

\* [part.peterson@ut.ee](mailto:part.peterson@ut.ee) (PP); [lili.milani@ut.ee](mailto:lili.milani@ut.ee) (LM)



**OPEN ACCESS**

**Citation:** Kasela S, Kisand K, Tserel L, Kaleviste E, Remm A, Fischer K, et al. (2017) Pathogenic implications for autoimmune mechanisms derived by comparative eQTL analysis of CD4<sup>+</sup> versus CD8<sup>+</sup> T cells. *PLoS Genet* 13(3): e1006643. doi:10.1371/journal.pgen.1006643

**Editor:** Tuuli Lappalainen, New York Genome Center & Columbia University, UNITED STATES

**Received:** July 5, 2016

**Accepted:** February 18, 2017

**Published:** March 1, 2017

**Copyright:** © 2017 Kasela et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** Gene expression data from this study have been submitted to the NCBI Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo/>) under accession number GSE78840. Individual level genotype data of the subjects from the Estonian biobank are handled in accordance with the regulations of the Human Genes Research Act, where data can be accessed upon ethical approval by submitting a data release request to the Estonian Genome Center, University of Tartu (<http://www.geenivaramu.ee/en/access-biopank/data-access>). We have made a browser available

## Abstract

Inappropriate activation or inadequate regulation of CD4<sup>+</sup> and CD8<sup>+</sup> T cells may contribute to the initiation and progression of multiple autoimmune and inflammatory diseases. Studies on disease-associated genetic polymorphisms have highlighted the importance of biological context for many regulatory variants, which is particularly relevant in understanding the genetic regulation of the immune system and its cellular phenotypes. Here we show cell type-specific regulation of transcript levels of genes associated with several autoimmune diseases in CD4<sup>+</sup> and CD8<sup>+</sup> T cells including a *trans*-acting regulatory locus at chr12q13.2 containing the rs1131017 SNP in the *RPS26* gene. Most remarkably, we identify a common missense variant in *IL27*, associated with type 1 diabetes that results in decreased functional activity of the protein and reduced expression levels of downstream *IRF1* and *STAT1* in CD4<sup>+</sup> T cells only. Altogether, our results indicate that eQTL mapping in purified T cells provides novel functional insights into polymorphisms and pathways associated with autoimmune diseases.

## Author summary

Variation in regulatory regions as well as coding regions of the genome can affect the expression of genes. Many of these variants have been associated with different diseases and other traits, but the underlying biological pathways are often left unknown. Analysing the effect of single nucleotide polymorphisms (SNPs) on gene expression levels, referred to as expression quantitative trait loci (eQTL), in specific cell types can be used to gain insight into specific mechanisms of disease. By analyzing eQTLs in CD4<sup>+</sup> and CD8<sup>+</sup> T

for all significant *cis*- and *trans*-eQTLs in CD4<sup>+</sup> and CD8<sup>+</sup> T cells detected at probe-level false discovery rate of 0.5 at <http://genenetwork.nl/cd4cd8eqtlbrowser>. This browser enables downloading the results files and performing queries based on a SNP or a gene for in depth comparisons.

**Funding:** This work was supported by the University of Tartu for the Center of Translational Genomics (SP1GVARENG) [AM], Estonian Research Council grants IUT20-60 [AM] and IUT2-2 [PP], the European Union through the European Regional Development Fund (Project No. 2014-2020.4.01.15-0012, <http://archimedes.ee/str/en/toetuse-edenemine/>) [AM], EU H2020 grant ePerMed (grant no. 692145, <https://ec.europa.eu/>) [AM], EU FP7 grant BBMRI-LPC (grant no 313010, <https://ec.europa.eu/>) [AM], ERA-Net.Rus grant EGIDA [PP], Wellcome Trust (Grants 074318 [JCK], and 090532/Z/09/Z [core facilities WTCHG]), the European Research Council (FP7/2007-2013; ERC Grant agreement number 281824 [JCK], Medical Research Council (98082 [JCK]), and National Institute for Health Research Oxford Biomedical Research Centre. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

cells, essential elements of adaptive immune response, we identified both *cis*- and *trans*-acting SNPs in genes associated with several autoimmune diseases.

## Introduction

T cells are essential elements of the adaptive immune response [1]. CD4<sup>+</sup> T cells, together with an appropriate cytokine environment, are required for the activation and differentiation of CD8<sup>+</sup> T cells that mediate defense and pathogen clearance during various infections [2]. The role of CD4<sup>+</sup> T cells is also necessary for B cells and macrophages to execute their protective functions.

T cells are activated and start to differentiate in response to complex stimuli involving cytokines, membrane receptors and transcription factors. Signaling through T cell receptor and costimulatory molecules induces their rapid proliferation and clonal expansion, and differentiation into effector and memory T cell subtypes [3]. Faulty activation or inadequate regulation of CD4<sup>+</sup> and CD8<sup>+</sup> T cells may contribute to the initiation and progression of multiple autoimmune diseases, including type 1 diabetes (T1D), rheumatoid arthritis (RA), autoimmune thyroiditis, systemic lupus erythematosus (SLE), multiple sclerosis, psoriasis, inflammatory bowel disease, as well as allergy and asthma [4,5]. In addition, a distinct lineage of CD4<sup>+</sup> T cells, regulatory T cells, have a central role in the control of the pathogenesis of autoimmune and inflammatory diseases [6].

Genome-wide association studies (GWAS) have identified thousands of single nucleotide polymorphisms (SNPs) associated with various immune related diseases [7–12]. Studies of the consequences of the risk alleles in biologically relevant contexts for different diseases have stressed the importance of the tissue and cell type-specificity of the regulatory variants [13–15]. We have recently shown that by mapping eQTLs in a sufficiently large sample set, it is possible to identify cell type-specific effects in whole blood, but the challenge to distinguish the cells responsible for the associations remains [16,17]. The power of studying expression-associated genetic variants in purified cells types has now been illustrated for B cells, monocytes, neutrophils, T regulatory cells and CD4<sup>+</sup> T cells [18–21], which allowed the identification of functional roles for several polymorphisms at autoimmune and even neurodegenerative disease loci. Meanwhile, some particularly interesting *cis*- and *trans*-eQTLs identified in whole blood could not be attributed to these cell types, and require an expanded survey of cells involved in immune response.

To this end, we purified CD4<sup>+</sup> and CD8<sup>+</sup> T cells from the peripheral blood of 313 healthy individuals for genome-wide mapping of genetic variation affecting the expression of genes involved in immune response. Our analysis characterizes both the extent of genetic control of gene expression in T cells and its cellular specificity, as well as variation in isoform levels of genes and their association with epigenetic changes. We show that integrating knowledge from GWAS with eQTLs enables us to clarify the functional consequences of disease-associated variants and assess the enrichment of autoimmune response. Finally, we highlight a common T1D-associated missense variant in *IL27* affecting the STAT1 and IRF1 pathway in CD4<sup>+</sup> T cells. Our analysis provides insights into the basic processes of the regulation of gene expression in T cells and advances our understanding about the pathways involved with disease susceptibility in the adaptive immune system.

## Results

### Genetic control of gene expression in T cells

We purified CD4<sup>+</sup> and CD8<sup>+</sup> T cells from peripheral blood mononuclear cells (PBMCs) of 313 healthy European individuals from the Estonian Biobank [22]. The purified cells were subjected

to genome-wide gene expression analysis, genotyping and imputation using the 1000 Genomes reference panel. After stringent quality control and filtering, close to 6 million SNPs, and expression data from 38,839 probes within 23,704 genes were included in the analysis. DNA methylation of 450,000 CpG sites was determined in the same purified cells of 100 of the individuals [23] to add explanations to observed differential expression between cell types.

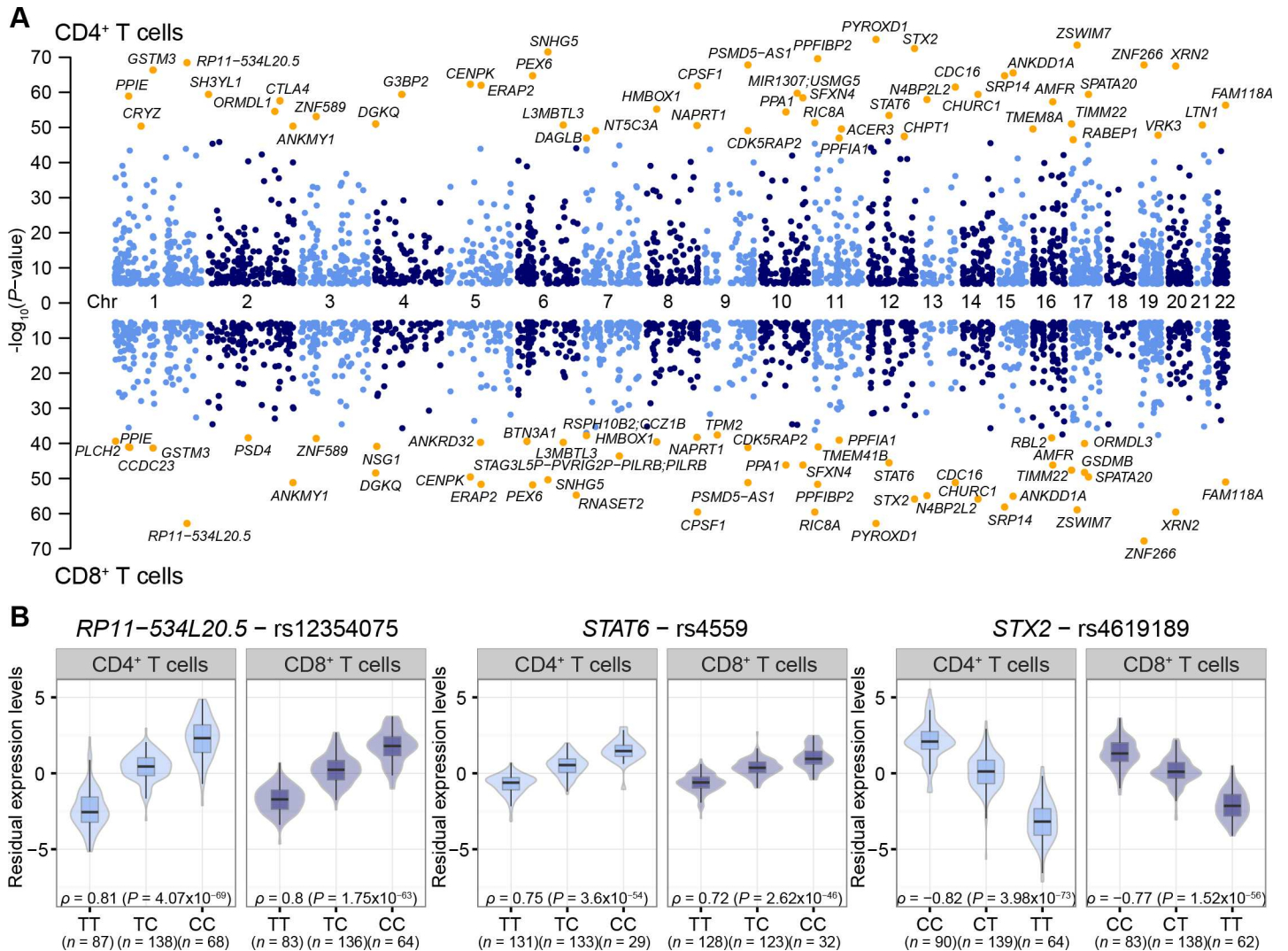
To characterize the extent of genetic control of gene expression in T cells and its cellular specificity, we first tested the association between SNPs and gene expression within 1Mb intervals, referred to as *cis*-eQTLs. In total, we identified *cis* regulatory SNPs for 2,605 genes in CD4<sup>+</sup> T cells and 2,056 genes in CD8<sup>+</sup> T cells at probe-level false discovery rate (FDR) < 0.05 (Fig 1A, S1 Table), with an overlap of 1,637 genes and estimated replication rate  $\pi_1 = 0.99$ . The high replication rate reflects the similarity of the two cell types compared, and also highlights the limitations of arbitrary cut-off levels for significance. Moreover, about half of the significant eQTLs detected in a large meta-analysis of peripheral blood from 5000 individuals [16] could be replicated in the CD4<sup>+</sup> and CD8<sup>+</sup> T cells ( $\pi_1 = 0.51$  and 0.45, respectively, S1 Fig), indicating a high level of T cell-specific eQTL effects [24,25]. As examples of *cis*-eQTLs in CD4<sup>+</sup> and CD8<sup>+</sup> T cells, we observed evidence of genetic regulation of long coding RNA *RP11-534L20.5* located 7 kb downstream of the *IKBKE* gene which has an essential role in regulating inflammatory responses to viral infections; *STAT6*, a transcriptional activator involved in T cell differentiation, and *STX2* (syntaxin 2), involved in intracellular transport of vesicles (Fig 1B).

## Cell type-specificity of eQTLs

We next used a Bayesian model averaging and hierarchical modelling [26] that combines information across genes to jointly call eQTLs in one framework to assess the specificity and proportion of shared eQTLs in CD4<sup>+</sup> and CD8<sup>+</sup> T cells. Of 3871 genes associated with eQTLs, all showed very strong posterior probability to be shared between CD4<sup>+</sup> and CD8<sup>+</sup> T cells (S2 Table).

One of the strengths of the multi-tissue joint *cis*-eQTL analysis is its capability to account for incomplete power [26,27]. As our sample size for both of the cell types was similar, we noted the advantage of Bayesian methods in particular for genes with eQTLs that have modest effects. Moreover, for eQTLs where the absolute effect size was higher in one cell type, there was also a tendency for higher expression levels in that gene in the given cell type (chi-square test  $P$ -value <  $2 \times 10^{-16}$ , S2 Fig). For a standard tissue-by-tissue analysis, higher effect sizes and small standard deviations lead to more statistical evidence against the null, thus resulting in lower  $P$ -values.

The effect of the magnitude of the expression levels is exemplified by the effect sizes and  $P$ -values of the probes in the inhibitory immune checkpoint gene, *CTLA4*, known to be associated with T1D and other endocrine autoimmune diseases [28,29]. *CTLA4* is covered by three probes all showing consistent eQTL effects shared by both cell types in the multi-tissue analysis (Fig 2A). In the tissue-by-tissue analysis, expression levels of *CTLA4* were associated with 122 SNPs (S3A Fig), including T1D and RA-associated variants rs231775 (+49A/G) [28], rs3087243 (CT60A/G) [28,30], rs231806 (MH30C/G) [28], and rs231735 [8] (pair-wise  $r^2$  between the SNPs ranging from 0.51 to 0.90, and strong linkage disequilibrium (LD,  $r^2 > 0.8$ ) between MH30 and CT60, S3C Fig). The CT60 polymorphism is also in strong LD with length variation of the (AT)<sub>n</sub> repeat in the 3'UTR of *CTLA4* with long variants of the repeat associating with lower *CTLA4* mRNA expression in autoimmune T cell clones [31]. Our results extend this finding to the polyclonal population of CD4<sup>+</sup> and CD8<sup>+</sup> T cells, where the disease risk alleles were associated with lower expression levels of *CTLA4* (S3B Fig). Genetic control of the expression



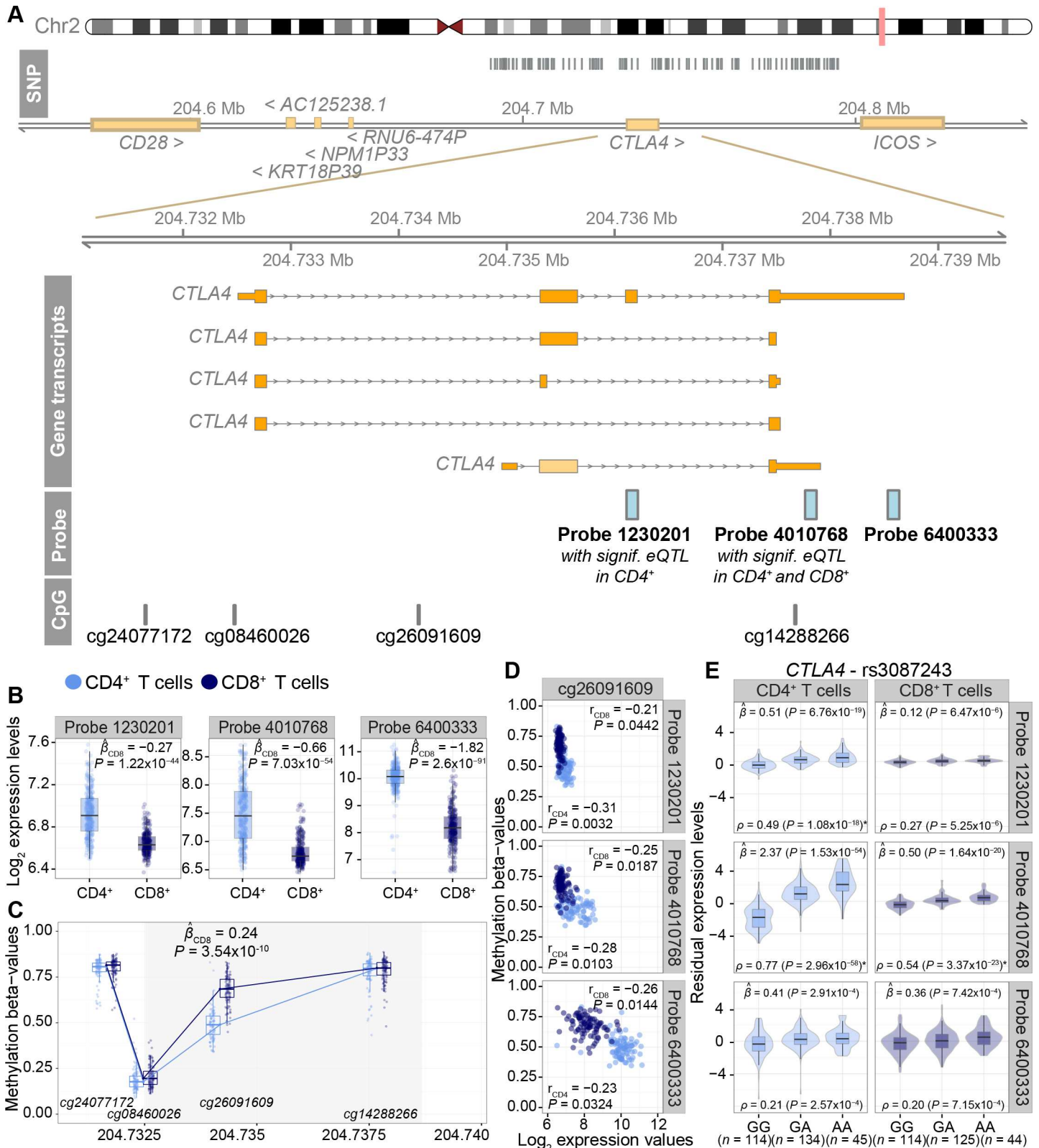
**Fig 1. Summary of cis-eQTL analysis in CD4<sup>+</sup> and CD8<sup>+</sup> T cells.** (A) Manhattan plots of the significant cis-eQTLs per gene in CD4<sup>+</sup> (top) and CD8<sup>+</sup> T cells (bottom). The names of the top 50 genes per cell type are shown. (B) Examples of genes with cis-eQTLs: *RP11-534L20.5* is located downstream of the *IKBKE* gene which has an essential role in regulating inflammatory responses to viral infections; *STAT6* is involved in T cell differentiation; and *STX2* is involved in intracellular transport of vesicles. The allelic effect of the SNPs on gene expression levels are shown by boxplots within violin plots. Violin plot shows the density plot of the data on each side, the lower and upper border of the box correspond to the first and third quartiles, respectively, the central line depicts the median, and whiskers extends from the borders to  $\pm 1.5 \times \text{IQR}$ , where IQR stands for inter-quantile range, the distance between the first and third quartiles.

doi:10.1371/journal.pgen.1006643.g001

of the alternatively spliced third exon, that is omitted from the secretory variant of *CTLA4* [28], was significantly detectable in CD4<sup>+</sup> T cells only, whereas genetic association with the expression of the fourth exon was significant in both cell types (Fig 2B and 2E). Further investigation of the *CTLA4* gene locus revealed differential methylation between CD4<sup>+</sup> and CD8<sup>+</sup> T cells (Fig 2C). Specifically, methylation of a CpG site near the second exon of *CTLA4* (cg26091609) was inversely correlated with the expression levels of the three different probes in the gene (Fig 2D), possibly explaining the expression differences in CD4<sup>+</sup> and CD8<sup>+</sup> T cells, and thus failing to detect significant eQTL effects in the tissue-by-tissue analysis.

Interestingly, divergent DNA methylation between CD4<sup>+</sup> and CD8<sup>+</sup> T cells was similarly detectable for an additional 73 out of 187 genes covered by more than one probe, and with





**Fig 2. Probe-level differences between CD4<sup>+</sup> and CD8<sup>+</sup> T cells in the *CTLA4* gene.** (A) Overview of the *CTLA4* gene locus. On the top panel the SNPs associating with different isoform levels between the two cell types in the *CTLA4* gene are depicted as grey bars, the gene transcripts are drawn, and the positions of the three gene expression probes and four CpG sites in the region are shown at the bottom. (B) The expression levels (y-axis, quantile

normalized and log<sub>2</sub> transformed) for each of the probes are shown by boxplots for CD4<sup>+</sup> and CD8<sup>+</sup> T cells. The mean of the differences in expression levels between CD4<sup>+</sup> and CD8<sup>+</sup> T cells are estimated by linear mixed models. (C) Methylation levels (*y*-axis) for each of the CpG sites along the genomic position (*x*-axis) are shown by boxplots for CD4<sup>+</sup> and CD8<sup>+</sup> T cells. Medians of the methylation values per CpG sites are connected by a line in both cell types. The mean of the differences in methylation levels between CD4<sup>+</sup> and CD8<sup>+</sup> T cells are estimated by linear mixed models. Only estimates with *P*-value < 0.05 are marked on the figure. (D) Correlation between methylation and gene expression levels are shown on the scatter plot between the differentially methylated CpG site cg26091609 for CD4<sup>+</sup> and CD8<sup>+</sup> T cells (*y*-axis) and three expression probes (*x*-axis). (E) The effect of rs3087243 on the expression levels (*y*-axis, residual gene expression levels) of different probes in the *CTLA4* gene are shown by boxplots within violin plots depicted as in Fig 1. Effect sizes and corresponding *P*-values are reported for the multi-tissue and tissue-by-tissue analysis above and below of the violin plots, respectively. Significant eQTLs in the tissue-by-tissue analysis are indicated with a star (\*).

doi:10.1371/journal.pgen.1006643.g002

seemingly isoform-specific eQTL effects (S3 Table). Among the 74 genes, multi-tissue analysis resulted in the identification of eQTL effects for two or more probes in 66 genes which were shared between the cell types. These observations suggest that the differential splicing effects, plausibly regulated by epigenetic DNA methylation, influence the gene expression levels, and thus the power to significantly detect eQTLs in the tissue-by-tissue analysis.

### T cell specific trans-acting regulatory locus

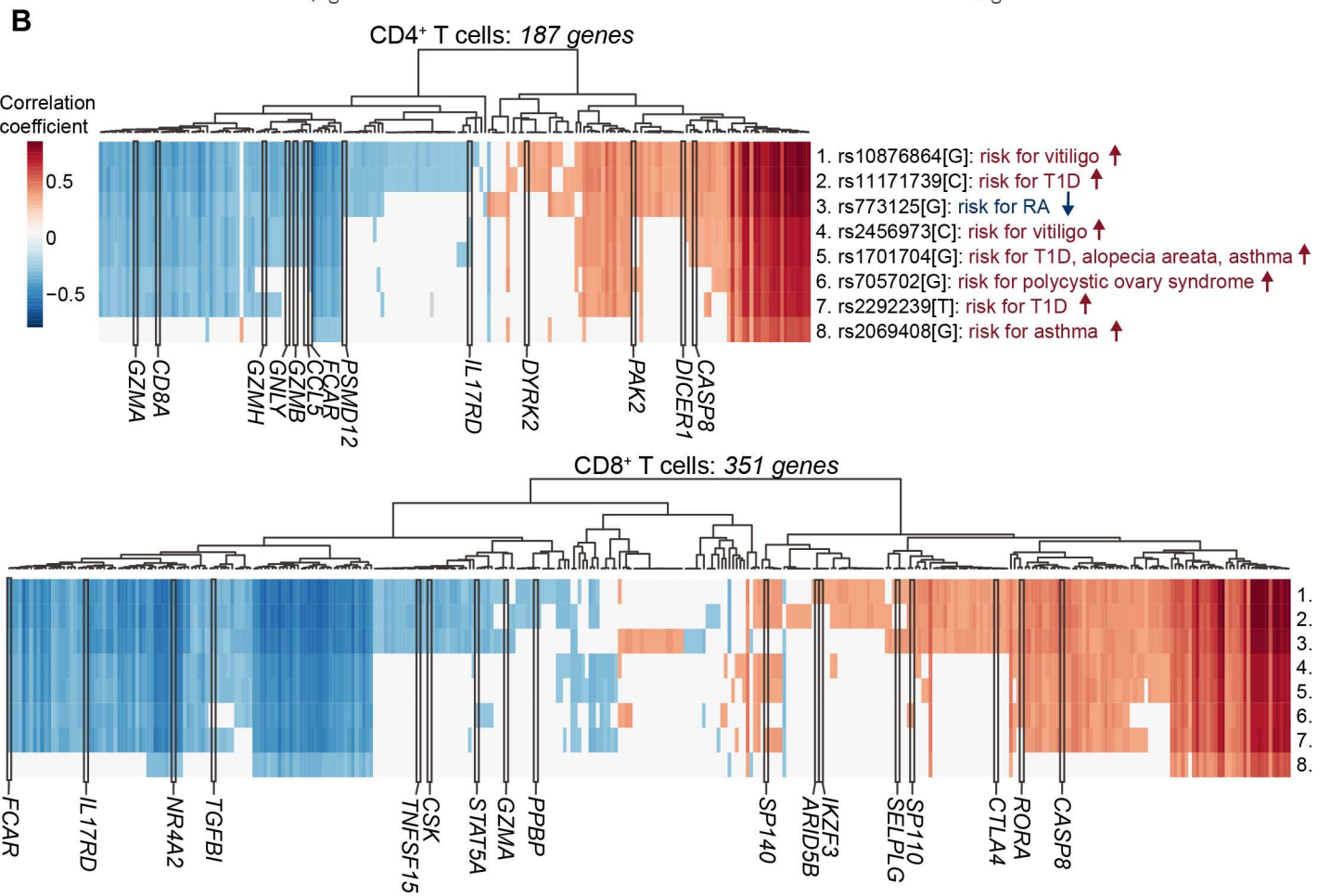
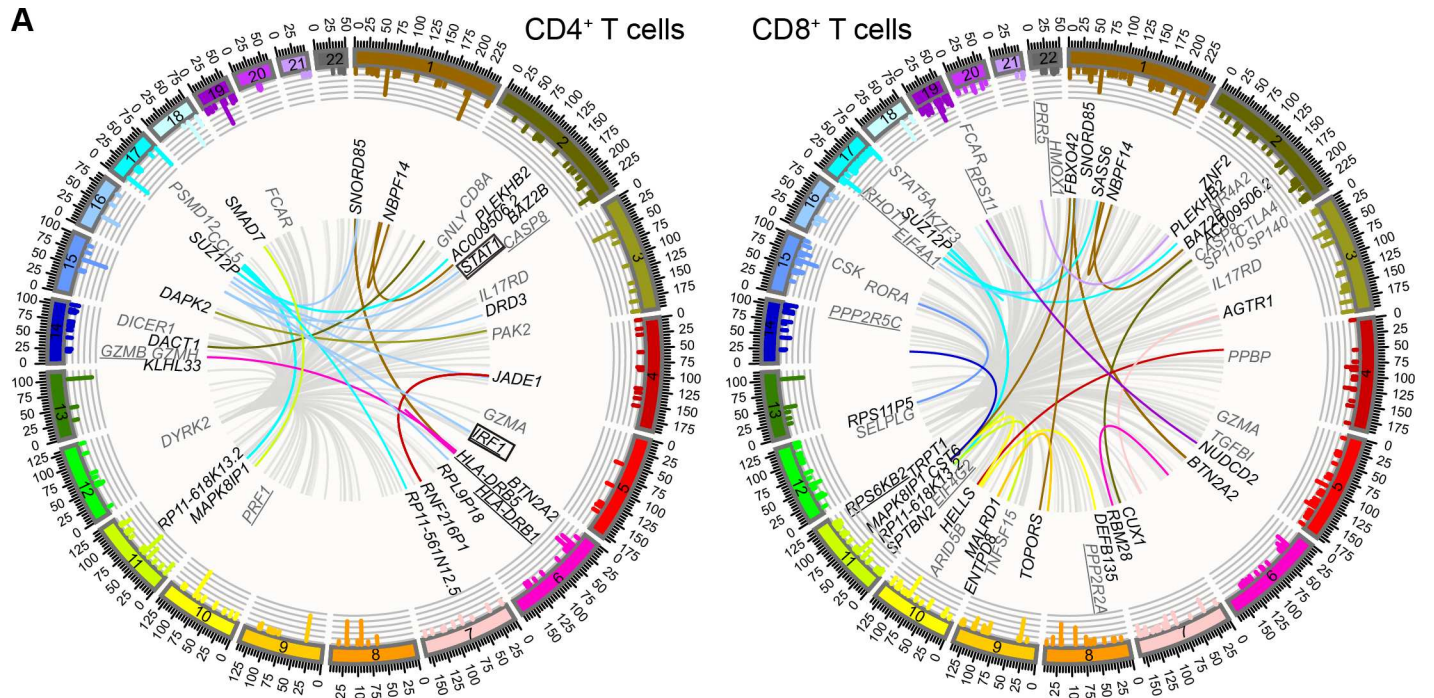
For the analysis of SNPs affecting the expression of distal genes (>5 Mb apart), referred to as *trans*-eQTLs, we selected all 4,638 genome-wide significant (*P*-value < 5×10<sup>-8</sup>) SNPs from the GWAS catalog [32] (accessed 24/03/2015). After correcting gene expression levels for *cis*-eQTL effects, we identified 36 and 40 GWAS SNPs associated with the expression levels of 209 and 378 distant genes in CD4<sup>+</sup> and CD8<sup>+</sup> T cells, respectively (overlap of 21 SNPs and 133 genes; Fig 3A, S4 Table). The functions of the genes associated with the *trans*-acting GWAS SNPs highlighted their role in T1D (Ingenuity pathway analysis, *P* = 4.39×10<sup>-5</sup>, Fig 3A) and mTOR signaling (*P* = 3.84×10<sup>-3</sup>, Fig 3A) in CD4<sup>+</sup> and CD8<sup>+</sup> T cells, respectively.

Strikingly, we observed a *trans*-acting regulatory locus at chr12q13.2 in CD4<sup>+</sup> and CD8<sup>+</sup> T cells with broad-range impact on hundreds of genes. Five SNPs in that region were previously implicated in B cell-specific trans-associations by Fairfax *et al.* [18]. In T cells, we identified the lead SNP in this region as rs1131017 located in the 5'UTR of the ribosomal small subunit protein *RPS26* gene (S5 Table). The rs1131017 SNP is in LD with eight GWAS SNPs implicated in T1D [34,36,40], vitiligo [33,35] and other autoimmune and inflammatory diseases [30,37–39]. The pair-wise *r*<sup>2</sup> values of the nine SNPs are shown in S4B Fig and their eQTL effects in S4C Fig, illustrating the strong LD (*r*<sup>2</sup> > 0.8) between the top *trans*-eQTL SNP rs1131017 and several of the GWAS SNPs. In addition to *RPS26*, the *trans*-acting region contains the *CDK2*, *RAB5B*, *SUOX*, *IKZF4* and *ERBB3* genes (S4A Fig) and associates with the expression levels of 187 and 351 genes in CD4<sup>+</sup> and CD8<sup>+</sup> T cells, respectively, with an overlap of 124 genes (Fig 3B). Many of these genes are highly expressed and have specific roles in T cells such as *CTLA4*, *GZMA*, *GZMB*, *GZMH*, *GZML*, *GZMY*, *GZNL* and *CD8A* (Fig 3B).

### Missense variant in *IL27* as a candidate disease variant for T1D revealing significant *trans*-eQTL effects in CD4<sup>+</sup> T cells

We also identified a *trans*-eQTL (GWAS SNP rs4788084[T]) on chr16p11.2 close to the *IL27* gene associated with lower expression of *IRF1* (*P* = 1.84×10<sup>-9</sup>) and *STAT1* (*P* = 2.91×10<sup>-8</sup>) in CD4<sup>+</sup> T cells only. The rs4788084[T] is associated with lower expression of *STAT1* in peripheral blood [16] and reduced risk for T1D [7,11]. To further explore the effects of genetic variants in the *IL27* region, we mapped *trans*-eQTLs for all SNPs in the chromosome 16:28,2–29,1 Mb region (829 SNPs in total, S6 Table, Fig 4A).

This analysis revealed an even stronger *trans*-eQTL signal for a missense SNP rs181206[G] within the *IL27* gene (Fig 4B) and decreased expression levels of *IRF1*, *STAT1* and *REC8* (Fig 4C).





**Fig 3. Trans-associations in CD4<sup>+</sup> and CD8<sup>+</sup> T cells.** (A) The outermost rim of the circos plot shows the histogram of the  $-\log_{10}(P\text{-value})$  of the associations between known genes and SNPs. Per every significant known gene only the highest  $-\log_{10}(P\text{-value})$  is depicted. The innermost network represents the *trans*-associations between the SNPs and the most significant gene expression probes per gene. The lines are colored by the chromosome of the given SNP, except for the locus on chr12q13.2 with over 100 *trans*-associations that are colored in grey. An arbitrary selection of genes is depicted. All genes not affected by chr12q13.2 SNPs are colored in black and a set of genes affected by chr12q13.2 *trans*-acting regulatory SNPs are selected based on their known importance in immune system related processes and are colored in grey. The underlined genes are involved in T1D and mTOR signaling in CD4<sup>+</sup> T cells and CD8<sup>+</sup> T cells, respectively. Two boxed genes among the genes with *trans*-eQTLs in CD4<sup>+</sup> T cells are associated with the T1D susceptibility variant rs4788084 close to the *IL27* gene. (B) Heatmap of the correlation coefficients between chr12q13.2 *trans*-acting regulatory SNP allele dosages and gene expression levels in CD4<sup>+</sup> and CD8<sup>+</sup> T cells are shown. The most significant gene expression probe per gene is chosen. An arbitrary selection of genes based on their importance in immune system related processes are marked with black borders and gene symbol. The depicted SNPs are linked with their role in disease susceptibility based on GWAS studies [30,33–40] and the numbers indicated in the lower panel refer to the same SNPs listed in the top panel.

doi:10.1371/journal.pgen.1006643.g003

The rs181206[G] allele, with a minor allele frequency of 29% in the 1000 Genomes phase 3 European population, causes the amino acid Leu119Pro change in the alpha-helical domain of IL-27 (S5 Fig). Strong LD ( $r^2$  ranges from 0.62 in the Estonian population to 0.84 in the other European populations, 1000 Genomes phase 3) between rs181206 and the GWAS SNP rs4788084 supports the protective role of the rs181206[G] allele for T1D among individuals of European ancestry. Of note, the effect of the missense SNP on *IRF1* and *STAT1* remains after regressing out the effect of the GWAS SNP, but not the other way around. Moreover, Bayesian test for colocalisation [41] adds strength to the hypothesis that T1D susceptibility and changes of expression in *IRF1* and *STAT1* are associated with the region, and share a single causal variant with the most posterior support for rs181206[G] (S7 Table).

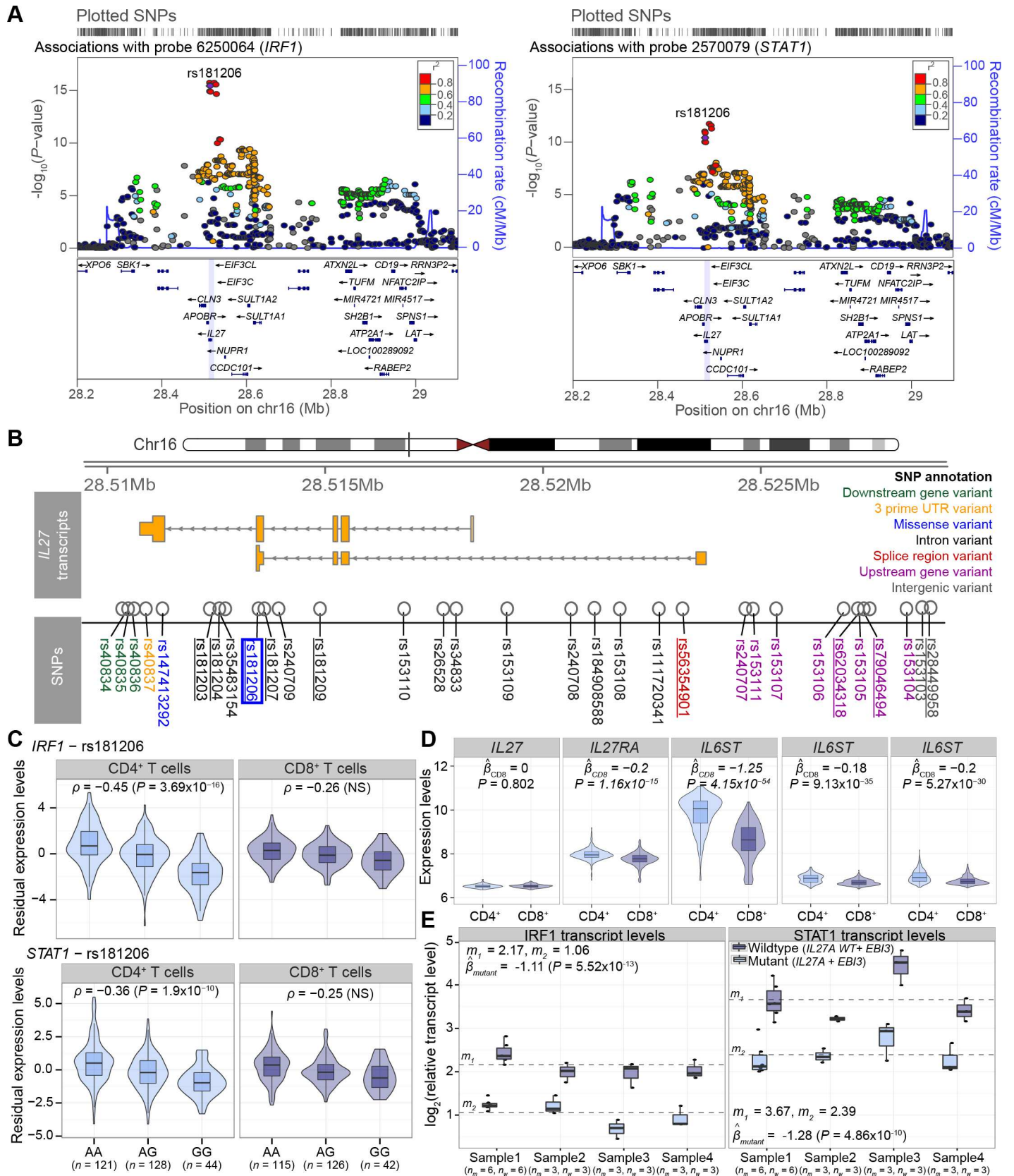
Importantly, despite the relatively small sample size, our focus on T cell subtypes enabled us to identify this effect as specific to CD4<sup>+</sup> T cells, as we did not detect significant *trans*-eQTL effects in this region in CD8<sup>+</sup> T cells (S6 Fig). We also confirmed the absence of the signal in B cells and monocytes by re-analyzing the data from Fairfax *et al.* [18] (S6 Fig). Notably, the *trans*-eQTL locus had no effects on the expression level of *IL27* in *cis* (S7 Fig). In agreement with the cell type-specific effect, we found higher expression levels of the *IL27RA* and *IL6ST* (gp130) genes, which together act as a receptor for the IL-27 cytokine, in CD4<sup>+</sup> cells in comparison to CD8<sup>+</sup> T cells (Fig 4D).

Given the strong positive correlation between the expression levels of *IRF1* and *STAT1*, we used structural equation modeling to determine whether the *IL27* SNP rs181206 affects both genes independently or via each other. Overall, the best-fitting scenario suggested *IRF1* to mediate the SNP and *STAT1* relationship (model 1, S8 Table). The finding was supported by a simulation experiment (S8 Fig) and suggested a mechanism for the effect of IL-27 on *IRF1* and *STAT1* expression (S9 Fig).

### Functional effects of the *IL27* missense SNP suggest causal role in decreased expression of *IRF1* and *STAT1*

In order to confirm the functional effect of the *IL27* rs181206[G] allele we investigated its effect on *IRF1* and *STAT1* expression in human PBMCs by additional experiments. IL-27 is produced by innate immune cells, and after forming a heterodimer with EB13, it interacts with its receptor *IL27RA* and activates the *STAT1/STAT3* pathway in T cells [42]. After binding to interferon stimulated response elements (ISRE), the *STAT1/STAT3* pathway induces transcription of several interferon-induced genes, including *IRF1* and *STAT1* itself. We cloned cDNA variants of the IL-27 wild-type (Leu119) and missense (Pro119), as well as EB13. After transfection into HEK293 cells, we combined the cell supernatants containing either IL-27 Leu119 or Pro119 protein with an equal amount of EB13 protein and studied their effect on *IRF1* and *STAT1* expression by real-time PCR in human PBMCs from four healthy individuals.





**Fig 4. Trans associations with *IRF1* and *STAT1* gene expression levels in CD4<sup>+</sup> T cells.** (A) Regional association plots of the *IL27* region SNPs with association *P*-values for expression levels of *IRF1* (chr5) and for *STAT1* (chr2) in CD4<sup>+</sup> T cells. The missense SNP rs181206 is used as the index SNP for showing LD between the SNPs. (B) Overview of the *IL27* transcripts and SNPs located in the *IL27* gene region. In addition to the missense SNP rs181206, other SNPs with the lowest association *P*-values in the upper LD cloud in panel (A) are intronic variants rs181203, rs181204, rs181207, rs181209, splice region variant rs56354901, upstream gene variants rs62034318, rs79046494, and intergenic variant rs28449958. (C) The effect of the missense SNP rs181206 on *IRF1* and *STAT1* gene expression levels are shown. *P*-values greater than  $9.2 \times 10^{-8}$  in CD8<sup>+</sup> T cells resulted in FDR > 0.05 noted as NS (not significant). (D) Expression levels (*y*-axis, quantile normalized and log<sub>2</sub>-transformed) grouped by cell type (*x*-axis) are shown for the three genes: *IL27* (probe 6520523 in the last exon), *IL27RA* gene (probe 4250735 in 3' downstream sequence), and *IL6ST* gene (probe 4010100 at the end of the last exon, 3830048 at the beginning of the last exon, 4260333 in the middle of the gene, from left to right). Linear mixed effects model is used to estimate the mean of the differences in expression levels between CD4<sup>+</sup> and CD8<sup>+</sup> T cells. (E) The effect of the mutant and wild-type alleles of rs181206 on the expression levels of *IRF1* and *STAT1*. The log<sub>2</sub> relative transcript levels (*y*-axis) are shown as a boxplot per allele and sample, with four samples in total. Every sample was run in multiple parallel reactions, indicated by  $n_m$  (number of mutant reactions) and  $n_w$  (number of wild-type reactions), Sample1 was used in two parallel sets. The mean expression in each class is shown by grey dashed lines, where  $m_1$  and  $m_2$  are indicating the mean among wild-type and mutant samples, respectively. The effect of the mutant SNP on transcript levels is evaluated by linear mixed effects models. Boxplots within violin plots are depicted as in Fig 1.

doi:10.1371/journal.pgen.1006643.g004

As shown in Fig 4E, the missense SNP resulting in Pro119 in IL-27 induced lower *STAT1* and *IRF1* transcript levels compared to IL-27 Leu119. The comparison of the fixed effects of the rs181206[G] allele resulted in highly significant estimates:  $\hat{\beta}_{mutant} = -1.11$  ( $P = 5.52 \times 10^{-13}$ ) for *IRF1* and  $\hat{\beta}_{mutant} = -1.28$  ( $P = 4.86 \times 10^{-10}$ ) for *STAT1*. This result supports our *trans*-eQTL analysis, suggesting that the Pro119 in the *IL27* gene is the causal SNP of these associations.

## Discussion

We here report eQTL mapping in purified CD4<sup>+</sup> and CD8<sup>+</sup> T cells and reveal multiple effects on regulation of genes associated with autoimmune diseases. The eQTL studies in cell subtypes have both demonstrated a high level of specificity as their effects vary across cell types, as well as remarkable sensitivity despite their several fold smaller sample sizes [24,25]. Indeed, we also observed many *cis*-eQTLs identified in CD4<sup>+</sup> and CD8<sup>+</sup> T cells that were shared according to the joint analysis, but were seemingly cell type-specific by the tissue-by-tissue analysis. The genes with such notable patterns include *CTLA4*, known for its inhibitory role in T cell mediated immune responses and association with many organ-specific autoimmune diseases [29]. The autoimmunity-susceptible SNPs in this region (rs231775, rs3087243, and rs231806) have been associated with various functional effects including *CTLA4* transcript levels, splicing, production of the soluble form of CTLA-4 and posttranslational modifications [28,43–45]. In addition, our results show differential DNA methylation in the *CTLA4* region to be associated with the potential isoform-specific differences in CD4<sup>+</sup> and CD8<sup>+</sup> T cells. This is likely due to negative correlation between DNA methylation and gene expression, which eventually leads to lower expression levels and moderate effect sizes. To confirm whether the observed effects are due to true differences in isoforms instead of the quality of the designed probes at capturing variation in gene expression levels, a follow-up study with RNA-seq data is needed.

The observed *trans*-acting regulatory region on chr12q13.2 in CD4<sup>+</sup> and CD8<sup>+</sup> T cells indicates the important role of genetic variants in that region affecting the expression levels of over a hundred genes across the genome. Interestingly, the region includes the *RPS26* gene, which may constitute a mechanism for the detected *trans*-eQTL effects. Our lead SNP rs1131017 is located in an oligopyrimidine tract of the 5' UTR of the *RPS26* mRNA and its T1D risk allele rs1131017[C] correlates positively with *RPS26* expression levels [46,47]. The oligopyrimidine tract controls the translation of many mammalian ribosomal protein genes [48], and the effect of the SNP on *RPS26* ribosomal distribution has been reported [49]. *RPS26* is a main component of the ribosomal region involved in the recruitment of cellular mRNA during translational initiation and in maintenance of the path of mRNA molecules to the ribosomal exit site [50]. Hence, it is conceivable that changed *RPS26* protein levels may affect the stability or

translational efficiency of a large number of cytosolic mRNAs. Nevertheless, the exact functional role of rs1131017[C] in T1D remains to be identified, as well as the effect of other candidate genes in this region such as *IKZF4*, *ERBB3* [51] and *CDK2* [52].

Strikingly, we identified a common missense variant in cytokine *IL27* as a significant *trans*-eQTL for *IRF1* and *STAT1* in CD4<sup>+</sup> T cells. The effect of IL-27 in T cells is regarded as anti-inflammatory but it has also been shown as a growth and survival factor for T cells [42]. As a heterodimer with EBI3, IL-27 activates the STAT1/STAT3 pathway that induces the transcription of several interferon-induced genes, including *IRF1* and *STAT1* itself. Our model suggests that *IRF1* mediates the effect of the *IL27* SNP on expression levels of *STAT1*. Moreover, our functional studies with the mutated form of IL-27 confirmed its decreased capacity to activate the STAT1 pathway, and we showed that a potential causal variant (the missense variant rs181206) for T1D susceptibility and changes in *IRF1* and *STAT1* expression is shared. Furthermore, our findings are supported by studies of a T1D mouse model with high levels of IL-27 and delayed T1D onset after treatment with an IL-27 blocking antibody [53]. Altogether these results suggest that the rs181206[G] variant of the *IL27* gene confers protection against T1D through the inhibited expression of *IRF1* and *STAT1* in CD4<sup>+</sup> T cells. Our results also suggest that IL-27 may promote autoimmunity toward pancreatic islets via the upregulation of the STAT1 pathway.

In conclusion, the analysis of genetic modulators of gene expression profiles as intermediate phenotypes between human traits and underlying genetic variation offers new instruments to refine our understanding of disease susceptibility. Moreover, eQTL studies in purified cell types instead of whole blood enable us to establish the specific mechanisms and pathways involved in diseases progression and create the basis for future explorations and drug interventions.

## Materials and methods

### Study design

The aims of this study were to perform *cis*- and *trans*-eQTL mapping in purified CD4<sup>+</sup> and CD8<sup>+</sup> T cells and evaluate their pathogenic implications for autoimmune mechanisms. Based on similar studies [18], we considered the sample size of 300 individuals to be sufficient to ensure statistical power to detect eQTL effects in purified cells. This study participants were healthy donors of the Estonian Genome Center of the University of Tartu [22]. In total, 313 subjects were selected for the study, with median age 54 (standard deviation 17.8), 154 females and 159 males. The study was approved by the Ethics Review Committee of Human Research of the University of Tartu, Estonia (permission no 206/T-4, date of issue 25.08.2011) and it was carried out in compliance with the Helsinki Declaration. A written informed consent to participate in the study was obtained from each individual prior to recruitment. All methods were carried out in accordance with approved guidelines.

DNA from the samples were genotyped using HumanOmniExpress BeadChips (Illumina), according to the manufacturer's instructions. We imputed both datasets using the 1000 Genomes project reference by using IMPUTE v2 [54], resulting in 5,879,386 autosomal SNPs with a minor allele frequency (MAF) of > 0.05 for downstream analyses. CD4<sup>+</sup> and CD8<sup>+</sup> T cells were extracted from the peripheral blood mononuclear cells (PBMC) by consecutive positive separation using microbeads (CD4<sup>+</sup> #130-045-101; CD8<sup>+</sup> #130-045-201) and AutoMACS technology (Miltenyi Biotec) according to the manufacturer's protocol. Gene expression data was generated using HumanHT-12v4 BeadChips (Illumina), according to the standard protocol. Preprocessing and quality control of the data was done using R [55] and the Bioconductor packages *lumi* [56] and *arrayQualityMetrics* [57]. The number of samples retained for further analysis was 293 for CD4<sup>+</sup> and 283 for CD8<sup>+</sup> T cells, 303 unique individuals. The effect of the



mutant and wild-type *IL27* was experimentally tested in HEK 293 cells. The procedures are described in details in the [S1 File](#).

### Cis- and trans-eQTL mapping

The effects of SNPs on local (*cis*-eQTL) and distant (*trans*-eQTL) genes were determined via eQTL mapping as described in eQTL mapping analysis cookbook developed by the University Medical Center Groningen at the Genetics Department and the Genomics Coordination Center (<https://github.com/molgenis/systemsgenetics/wiki/eQTL-mapping-analysis-cookbook>) and previously in [16]. For *cis*-analysis the distance between the probe midpoint and SNP genomic location was up to 1 Mb and for *trans*-analysis the distance was more than 5 Mb or the probe and SNP were on different chromosomes. Only SNPs with a MAF > 0.05, call rate > 0.95 and a Hardy-Weinberg equilibrium *P*-value > 0.001 were included in the analyses. We used the Spearman correlation coefficient (Spearman's rho) to detect associations between the coding allele of the SNP (directly genotyped or imputed allele dosages) and the variations in gene expression levels (residual gene expression levels obtained after corrections as described in "Gene expression quality control and normalization" in Supplementary Methods). To control for multiple testing, we applied more conservative probe-level false discovery rate (FDR) procedure. The eQTL mapping procedures are described in details in the [S1 File](#).

### Proportion of true positives

To estimate the sharing of *cis*-eQTLs between the two cell types and peripheral blood we used the  $\pi_1$  statistic (Storey and Tibshirani *q*-value approach [58] implemented in the R *qvalue* package, default settings used). Based on the list of *P*-values, the overall proportion of true null hypotheses ( $\pi_0$ ) is estimated, i.e. those following the Uniform(0,1) distribution. An estimate of true alternative tests is  $\pi_1 = 1 - \pi_0$ , i.e.  $\pi_1$  is the proportion of true positives. The reported replication rate (proportion of sharing) between two groups (CD4<sup>+</sup> or CD8<sup>+</sup> T cells or peripheral blood) is the  $\pi_1$  statistic estimated by taking the significant SNP-probe list from one group and using the corresponding *P*-value distribution in the other group.

### Multi-tissue joint discovery of eQTLs

We used Bayesian framework for multi-tissue joint eQTL analysis by Flutre *et al.* [26] in CD4<sup>+</sup> and CD8<sup>+</sup> T cells implemented in the eQTLBma program. The *cis* candidate region was defined as +/- 1 Mb from probe midpoint. As an input to the program, residual gene expression levels obtained after corrections as described above and the corresponding genotypes were used. Firstly, the raw Bayes factors for each configuration (eQTL active in CD4<sup>+</sup>, eQTL active in CD8<sup>+</sup>, eQTL active in both cell types) were computed assessing the support in the data for each probe-SNP pair being an eQTL. Secondly, information across all genes were combined by the hierarchical model with an EM algorithm to get maximum-likelihood estimates of the configuration probabilities. Thirdly, by Bayesian model averaging using the raw Bayes factors weighted by the estimated configuration probabilities, the posterior probabilities for the eQTL to be active in a given configuration were obtained. As a final result, only the best SNP per probe is picked, based on the posterior probability for a SNP to be "the" eQTL for a probe. To obtain the posterior probabilities, an estimate of the probability for a probe to have no eQTL in any cell type ( $\pi_0$ ) is needed. This was estimated using Storey and Tibshirani *q*-value approach [58] via probe-level *P*-values obtained by 10,000 permutations.

## Pathway analysis

The Ingenuity Pathway Analysis (IPA) tool was used to establish associations with processes and diseases. The standard IPA enrichment analysis was performed with the set of all human genes on the Illumina array as background.

## Bayesian test for colocalisation

The R package *coloc* [41] was used to assess whether two association signals are consistent with a shared causal variant. Assuming a single causal SNP for both of the traits in a region, posterior support for following hypothesis is estimated:

- $H_0$ : neither trait has a genetic association in the region;
- $H_1$ : only  $trait_1$  has a genetic association in the region;
- $H_2$ : only  $trait_2$  has a genetic association in the region;
- $H_3$ : both traits are associated, but with different causal variants;
- $H_4$ : both traits are associated and share a single causal variant.

For  $trait_1$ , we chose T1D susceptibility and obtained necessary summary statistics from Onengut-Gumuscu *et al.* [59] from [www.t1dbase.org](http://www.t1dbase.org) (accessed 13/09/2016) webpage. For  $trait_2$ , we chose the association with *IRF1* and *STAT1* expression in CD4<sup>+</sup> T cells. Following Giambartolomei *et al.* [41], we set prior probability ( $p_1$ ) that a variant is associated with  $trait_1$  to  $10^{-4}$ , prior probability ( $p_2$ ) for *trans*-eQTL effect to  $10^{-5}$ , and prior probability ( $p_{12}$ ) that a variant is associated with both traits to  $10^{-6}$ . To show the consistency of the results, we varied  $p_2$  and  $p_{12}$  values from  $10^{-4}$  to  $10^{-5}$  and  $10^{-5}$  to  $10^{-6}$ , respectively.

## Structural equation modeling (SEM) and mediation analysis

We compared the plausibility of potential models for the effects of the *IL27* SNP on gene expression levels of *STAT1* and *IRF1* using structural equation modeling (SEM). First, to test the model's concordance to the true population we used the chi-square test. Next, to assess the goodness of fit of the resulting models, we used the following measures: Comparative Fit Index (CFI), Tucker-Lewis Index (TLI) and Root Mean Square Error of Approximation (RMSEA). The CFI and TLI goodness of fit measures indicate good models for values higher than 0.9. RMSEA values less or equal to 0.05 indicate reasonable fit between the model and the data. To compare the models, we used the Akaike information criterion (AIC). Lower values of that theoretic measure indicate better models.

We applied the Sobel test to test whether a mediator carries the influence of the causal variable to the outcome, i.e. testing whether the indirect effect of the causal variable on the outcome via the mediator is significantly different from zero. The test statistic was calculated using the Sobel formula [60]:  $ab/\sqrt{b^2s_a^2 + a^2s_b^2}$ , where  $a$  is the effect of the causal variable on the mediator and  $b$  is the effect of the mediator on the outcome and  $s_a$  and  $s_b$  are the corresponding standard errors of  $a$  and  $b$ .

## *IL27-IRF1-STAT1* simulation study

We conducted a simulation study to investigate the relationship between the coding allele of the SNP and expression levels of two genes, *IRF1* and *STAT1*. Therefore, we performed 1000 simulations using a sample size of  $n = 1000$ . We generated SNP genotypes with the minor allele frequency being the sum of two random binary traits from a binomial distribution  $B(1, 0.37)$ ,

*IRF1* expression levels depending only on the SNP ( $irf1 = 0.903 - 1.237 \times snp + N(0,1)$ ) and *STAT1* expression levels depending only on *IRF1* expression levels ( $stat1 = 0.031 - 0.534 \times irf1 + N(0,1)$ ). Then using simulated data, we estimated the significance of the SNP allelic effect and *IRF1* expression levels on *STAT1* expression levels and the SNP allelic effect and *STAT1* expression levels on *IRF1* expression levels by linear models.

## Graphics packages

Graphs were generated using R [55] base packages and *ggplot2* [61]. The LD plots were generated using *LDheatmap* [62], the regional association plots were generated with Locus Zoom [63] using hg19/1000 Genomes Nov 2014 EUR for showing LD between the variants, and the circos plots were drawn using the Circos visualization tool [64].

## Statistical analysis

We used linear mixed effects models to estimate the mean of the differences in methylation and expression levels between CD4<sup>+</sup> and CD8<sup>+</sup> T cells assessing random intercepts for each of the individuals adjusted for sex, age and batch effects (chip and position on chip). The effect of the mutant SNP on *IRF1* and *STAT1* transcript levels was also evaluated by linear mixed effects models.

## Supporting information

### S1 Fig. Estimated replication rate of peripheral blood *cis*-eQTLs in CD4<sup>+</sup> and CD8<sup>+</sup> T cells.

The histograms show the distribution of *P*-values of significant SNP-probe pairs (probe-level FDR < 0.05) discovered in the meta-analysis by Westra *et al.* [16] in (A) CD4<sup>+</sup> and (B) CD8<sup>+</sup> T cells. The  $\pi_1$  statistic [58] estimates the proportion of true positives from the *P*-value distribution, interpreted as the proportion of replicated *cis*-eQTL effects.

(TIF)

**S2 Fig. Association between the eQTL effect size and gene expression levels.** The scatterplot shows the effect of the “best” SNP allele dosages on gene expression for 4385 significant probes in CD4<sup>+</sup> T cells (*x*-axis) and in CD8<sup>+</sup> T cells (*y*-axis). For eQTLs where the absolute effect size was higher in one cell type, there was also a tendency for higher expression levels in the given cell type (chi-square test *P*-value <  $2 \times 10^{-16}$ ).

(TIF)

### S3 Fig. Associations of the *CTLA4* region SNPs with different expression probes in *CTLA4* gene.

(A) Regional association plots of the *CTLA4* region (chr2:203,736–205,738 Mb) SNPs with association *P*-values for gene expression levels of the three probes in the *CTLA4* gene (1230201, 4010768 and 6400333) in CD4<sup>+</sup> and CD8<sup>+</sup> T cells. The SNP rs3087243 is the lead eQTL SNP and is used as the index SNP for showing linkage disequilibrium between the SNPs. (B) Heatmap of the correlation coefficients between the T1D and/or RA-associated variants rs231735 [8], rs231806 [28] (MH30C/G), rs3087243 [28,30] (CT60A/G), rs231775 [28] (+49A/G) and gene expression levels of the three different probes in the *CTLA4* gene in CD4<sup>+</sup> and CD8<sup>+</sup> T cells. Allele in brackets indicates the assessed allele which is also the MAF allele. (C) Linkage disequilibrium (LD) plot for the four T1D and/or RA-associated variants is shown. LD between the SNPs is measured by pairwise  $r^2$  calculated using the genotypes of 99 individuals from 1000 Genomes project phase 3 CEU population. UCSC genes (based on RefSeq) and their location with respect to SNPs are shown on the top of the LD plot. In the Estonian population, the LD between rs231735 and rs231806, rs3087243 is stronger ( $r^2$  of 0.85



and 0.76, respectively), the other  $r^2$  values are similar.  
(TIF)

**S4 Fig. Overview of the chr12q13.2 *trans*-acting locus.** (A) Expression levels ( $y$ -axis, quantile normalized and  $\log_2$ -transformed) of the six genes in the region by cell type ( $x$ -axis) are shown by box plots incorporated into violin plots. Violin plot shows the density plot of the data on each side, the lower and upper border of the box correspond to the first and third quartiles, respectively, the central line depicts the median, and whiskers extends from the borders to  $\pm 1.5 \times \text{IQR}$ , where IQR stands for inter-quantile range, the distance between the first and third quartiles. (B) Linkage disequilibrium (LD) plot for the lead eQTL SNP and eight GWAS SNPs on chromosome 12 *trans*-acting region chr12q13.2 is shown. The GWAS SNPs are linked with their role in disease susceptibility in Fig 3. LD between the SNPs is measured by pairwise  $r^2$  calculated from 99 individuals from 1000 Genomes project phase 3 CEU population. UCSC genes (based on RefSeq) and their location with respect to SNPs are shown on the top of the LD plot. There is strong LD ( $r^2 > 0.8$ ) between the lead eQTL SNP rs1131017 and three GWAS SNPs rs10876864, rs11171739 and rs773125 ( $r^2$  of 1.00, 0.92 and 0.88, respectively, in CEU population and  $r^2$  of 0.98, 0.98 and 0.84, respectively, in the Estonian population). (C) Regional association plots including all SNPs at chr12q13.2. The smallest association  $P$ -values for the SNPs are shown in CD4<sup>+</sup> and CD8<sup>+</sup> T cells. The SNP rs1131017 is the lead SNP in the region with the smallest  $P$ -value, and the eight GWAS SNPs are highlighted.  
(TIF)

**S5 Fig. Structure model of the human IL-27 (Q8NEV9; residues 61 to 134, based on <http://www.proteinmodelportal.org/>) and location of the amino acid lysine (L119) that is changed to proline (P119) by the rs181206 missense mutation.** We identified a common missense variant rs181206[A/G] in cytokine *IL27* as a *trans*-eQTL for *IRF1* and *STAT1* in CD4<sup>+</sup> T cells. The G allele of the variant alters an amino acid in the alpha-helical domain from leucine to proline at position 119.  
(TIF)

**S6 Fig. Associations of the *IL27* region SNPs with *IRF1* and *STAT1* gene expression levels in CD8<sup>+</sup> T cells, B cells and in monocytes.** Regional association plots of the *IL27* region (chr16:28,2–29,1 Mb) SNPs with association  $P$ -values for (A) *IRF1* (chr5) and (B) *STAT1* (chr2) gene expression levels in CD8<sup>+</sup> T cells, B cells, and monocytes. The SNP rs181206 is used as the index SNP for showing linkage disequilibrium between the SNPs.  
(TIF)

**S7 Fig. Associations of the *IL27* region SNPs with *IL27* gene expression levels in CD4<sup>+</sup> T cells, CD8<sup>+</sup> T cells, B cells and in monocytes.** Regional association plots of the *IL27* region (chr16:28,2–29,1 Mb) SNPs with association  $P$ -values for the *IL27* gene expression levels in CD4<sup>+</sup> T cells, CD8<sup>+</sup> T cells, B cells, and monocytes. The SNP rs181206 is used as the index SNP for showing linkage disequilibrium between the SNPs.  
(TIF)

**S8 Fig. Comparison of two models  $IRF \sim STAT1 + SNP$  and  $STAT1 \sim IRF1 + SNP$  on observed and simulated data.** We performed 1000 simulations using a sample size of  $n = 1000$ . We generated SNP genotypes, *IRF1* and *STAT1* expression levels according to the plausible causal model 1)  $SNP \rightarrow IRF1 \rightarrow STAT1$  as follows: we generated SNP genotypes with minor allele frequency as the sum of a two random binary traits from binomial distribution  $B(1, 0.37)$ , *IRF1* expression levels depending only on the SNP ( $irf1 = 0.903 - 1.237 \times snp + N(0,1)$ ) and *STAT1* expression levels depending only on *IRF1* expression levels ( $stat1 = 0.031$

+ 0.534 x *irf1* + N(0,1)). Then using simulated and observed data, we compared the estimates obtained from two linear models IRF ~ STAT1 + SNP (upper panel) and STAT1 ~ IRF1 + SNP (lower panel). Parameter estimates with 95% confidence intervals are shown for every explanatory variable for observed and simulation data. A similar pattern of the parameter estimates supports the validity of causal model 1).

(TIF)

**S9 Fig. A scheme on IL-27 role in activation of *IRF1* and *STAT1* gene expression.** IL27 (as a heterodimer with EBI3), upon binding to its receptor, activates the STAT1/STAT3 pathway. After binding to interferon stimulated response elements (ISRE), STAT1/STAT3 pathway induces transcription of several interferon-induced genes, including *IRF1* and *STAT1* itself. IRF1 is a transcription factor that enhances the expression of *STAT1* gene. We identified a common missense variant in cytokine *IL27* as a *trans*-eQTL for *IRF1* and *STAT1* in CD4<sup>+</sup> T cells. Our model suggests that *IRF1* mediates the SNP and *STAT1* relationship. Moreover, our functional studies with the mutated form of IL-27 (that is associated with protection against T1D via linkage disequilibrium with GWAS SNP rs4788084) confirmed its decreased capacity to activate the STAT1 pathway.

(TIF)

**S1 Table. Significant *cis*-eQTL effects in CD4<sup>+</sup> and CD8<sup>+</sup> T cells (probe-level false discovery rate < 0.05).**

(XLSX)

**S2 Table. Significant results of multi-tissue joint analysis (FDR < 0.05).** Bayesian model averaging and hierarchical modelling framework by Flutre *et al.* was used implemented in the eQTLBma program. Only the best SNP per probe is picked, based on the posterior probability for a SNP to be “the” eQTL for that probe. Posteriors of all configurations together with summary statistics in CD4<sup>+</sup> and CD8<sup>+</sup> T cells are reported. There are 4385 probes with eQTLs within 3871 genes (FDR < 0.05).

(XLSX)

**S3 Table. Significant differences in DNA methylation levels between CD4<sup>+</sup> and CD8<sup>+</sup> T cells among genes with seemingly isoform-specific effects between the two cell types.** There were 408 out of 3,024 genes affected by a SNP which were covered by more than one expression probe with *cis*-eQTL effects in CD4<sup>+</sup> and/or CD8<sup>+</sup> T cells. We could find CpG sites in the proximity (+/- 1000 basepairs) of the gene for 405 genes covered by 8665 CpG sites. We tested those CpG sites for differential methylation between CD4<sup>+</sup> and CD8<sup>+</sup> T cells. The effect on the mean of the differences in CD8<sup>+</sup> compared to CD4<sup>+</sup> T cells was evaluated using linear mixed models. False discovery rate based multiple testing was used based on all tested CpG sites. CpG sites with the most significant differences in methylation are listed per 74 gene out of 187 with seemingly isoform-specific effects in the tissue-by-tissue analysis. Multi-tissue analysis could find eQTL effect for two or more probes in 66 genes (of the 74 genes) with high posterior confidence to be shared between the two cell types.

(XLS)

**S4 Table. Significant *trans*-eQTL effects with SNPs associated with human diseases or traits in CD4<sup>+</sup> and CD8<sup>+</sup> T cells (probe-level false discovery rate < 0.05).**

(XLS)

**S5 Table. Significant *trans*-eQTL effects with SNPs at chr12q13.2 in CD4<sup>+</sup> and CD8<sup>+</sup> T cells (probe-level false discovery rate < 0.05).**

(XLS)

**S6 Table. *IL27* region *trans*-eQTL mapping results for the *IRF1* and *STAT1* genes in CD4<sup>+</sup> T cells, CD8<sup>+</sup> T cells, monocytes, and B cells.**

(XLS)

**S7 Table. T1D/eQTL colocalisation.** Results of the colocalisation analysis between the *trans*-eQTLs for *IRF1* (A) and *STAT1* (B) in CD4<sup>+</sup> T cells and the type 1 diabetes (T1D) susceptibility using different prior probabilities. The columns “*T1D pval*” and “*eQTL pval*” note the lowest *P*-value found for the association with T1D from Onengut-Gumuscu *et al.* study using T1DBase database ([www.t1dbase.org](http://www.t1dbase.org)), and for the expression association in CD4<sup>+</sup> T cells respectively, with the corresponding SNP name (“*T1D SNP*” and “*eQTL SNP*”), “*Best Causal*” reports the SNP with the highest posterior probability to be the true causal variant among the two. Different prior probabilities for observing *trans*-eQTL effect ( $p_2$ ) and different prior probabilities for the variant being associated with both traits ( $p_{12}$ ) are used to estimate the poster probability for different signal (different causal variant for associated traits, *PP3*) and common signal (shared single causal variant for associated traits, *PP4*).

(DOCX)

**S8 Table. SEM fit statistics for three alternative causal models.** To test the model’s concordance to true population we used chi-square test. To assess the goodness of fit of the resulting models, we used the following measures: Comparative Fit Index (CFI), Tucker-Lewis Index (TLI) and Root Mean Square Error of Approximation (RMSEA). The CFI and TLI goodness of fit measures indicate good models for values higher than 0.9. RMSEA values less or equal to 0.05 indicate reasonable fit between the model and the data. To compare models we used the Akaike information criterion (AIC). Lower values of the theoretic measures indicate better models.

(DOCX)

**S1 File. Supplementary material and methods.**

(PDF)

## Acknowledgments

We thank Mr Kaur Alasoo for valuable discussion, Ms Maire Pihlap and Mr Viljo Soo for their assistance with laboratory work, Mr Marc Jan Bonder for setting up the web browser, and Dr Tomi M. Pastinen, Mr Bing Ge, Prof Stephen Sawcer for their for assistance in evaluating our results. This work was carried out in part in the High Performance Computing Center of University of Tartu.

## Author Contributions

**Conceptualization:** LM PP.

**Data curation:** SK TE HJW LF BPF SM.

**Formal analysis:** SK KK KF.

**Funding acquisition:** JCK LF AM PP.

**Investigation:** SK KK LT EK AR BPF.

**Methodology:** SK KF HJW LF.

**Project administration:** LM PP AM.

**Resources:** HJW BPF JCK LF AM PP LM.



**Supervision:** AM PP LM.

**Visualization:** SK.

**Writing – original draft:** SK KK PP LM.

**Writing – review & editing:** BPF JCK HJW LF.

## References

- Swain SL, McKinstry KK, Strutt TM. Expanding roles for CD4<sup>+</sup> T cells in immunity to viruses. *Nat Rev Immunol.* 2012; 12: 136–48. doi: [10.1038/nri3152](https://doi.org/10.1038/nri3152) PMID: [22266691](https://pubmed.ncbi.nlm.nih.gov/22266691/)
- Tscharke DC, Croft NP, Doherty PC, La Gruta NL. Sizing up the key determinants of the CD8<sup>+</sup> T cell response. *Nat Rev Immunol.* 2015; 15: 705–16. doi: [10.1038/nri3905](https://doi.org/10.1038/nri3905) PMID: [26449178](https://pubmed.ncbi.nlm.nih.gov/26449178/)
- Kaech SM, Cui W. Transcriptional control of effector and memory CD8<sup>+</sup> T cell differentiation. *Nat Rev Immunol.* 2012; 12: 749–761. doi: [10.1038/nri3307](https://doi.org/10.1038/nri3307) PMID: [23080391](https://pubmed.ncbi.nlm.nih.gov/23080391/)
- Liblau RS, Wong FS, Mars LT, Santamaria P. Autoreactive CD8 T cells in organ-specific autoimmunity: emerging targets for therapeutic intervention. *Immunity.* 2002; 17: 1–6. PMID: [12150886](https://pubmed.ncbi.nlm.nih.gov/12150886/)
- Walter U, Santamaria P. CD8<sup>+</sup> T cells in autoimmunity. *Curr Opin Immunol.* 2005; 17: 624–31. doi: [10.1016/j.coi.2005.09.014](https://doi.org/10.1016/j.coi.2005.09.014) PMID: [16226438](https://pubmed.ncbi.nlm.nih.gov/16226438/)
- Noack M, Miossec P. Th17 and regulatory T cell balance in autoimmune and inflammatory diseases. *Autoimmun Rev.* 2014; 13: 668–77. doi: [10.1016/j.autrev.2013.12.004](https://doi.org/10.1016/j.autrev.2013.12.004) PMID: [24418308](https://pubmed.ncbi.nlm.nih.gov/24418308/)
- Barrett JC, Clayton DG, Concannon P, Akolkar B, Cooper JD, Erlich HA, et al. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat Genet.* 2009; 41: 703–7. doi: [10.1038/ng.381](https://doi.org/10.1038/ng.381) PMID: [19430480](https://pubmed.ncbi.nlm.nih.gov/19430480/)
- Gregersen PK, Amos CI, Lee AT, Lu Y, Remmers EF, Kastner DL, et al. REL, encoding a member of the NF-kappaB family of transcription factors, is a newly defined risk locus for rheumatoid arthritis. *Nat Genet.* 2009; 41: 820–3. doi: [10.1038/ng.395](https://doi.org/10.1038/ng.395) PMID: [19503088](https://pubmed.ncbi.nlm.nih.gov/19503088/)
- Han J-W, Zheng H-F, Cui Y, Sun L-D, Ye D-Q, Hu Z, et al. Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. *Nat Genet.* 2009; 41: 1234–7. doi: [10.1038/ng.472](https://doi.org/10.1038/ng.472) PMID: [19838193](https://pubmed.ncbi.nlm.nih.gov/19838193/)
- Dubois PCA, Trynka G, Franke L, Hunt KA, Romanos J, Curtotti A, et al. Multiple common variants for celiac disease influencing immune gene expression. *Nat Genet.* 2010; 42: 295–302. doi: [10.1038/ng.543](https://doi.org/10.1038/ng.543) PMID: [20190752](https://pubmed.ncbi.nlm.nih.gov/20190752/)
- Plagnol V, Howson JMM, Smyth DJ, Walker N, Hafler JP, Wallace C, et al. Genome-wide association analysis of autoantibody positivity in type 1 diabetes cases. *PLoS Genet.* 2011; 7: e1002216. doi: [10.1371/journal.pgen.1002216](https://doi.org/10.1371/journal.pgen.1002216) PMID: [21829393](https://pubmed.ncbi.nlm.nih.gov/21829393/)
- Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature.* 2012; 491: 119–24. doi: [10.1038/nature11582](https://doi.org/10.1038/nature11582) PMID: [23128233](https://pubmed.ncbi.nlm.nih.gov/23128233/)
- Fu J, Wolfs MGM, Deelen P, Westra H-J, Fehrmann RSN, Te Meerman GJ, et al. Unraveling the regulatory mechanisms underlying tissue-dependent genetic variation of gene expression. *PLoS Genet.* 2012; 8: e1002431. doi: [10.1371/journal.pgen.1002431](https://doi.org/10.1371/journal.pgen.1002431) PMID: [22275870](https://pubmed.ncbi.nlm.nih.gov/22275870/)
- Grundberg E, Small KS, Hedman ÅK, Nica AC, Buil A, Keildson S, et al. Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat Genet.* 2012; 44: 1084–1089. doi: [10.1038/ng.2394](https://doi.org/10.1038/ng.2394) PMID: [22941192](https://pubmed.ncbi.nlm.nih.gov/22941192/)
- Trynka G, Sandor C, Han B, Xu H, Stranger BE, Liu XS, et al. Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nat Genet.* 2013; 45: 124–30. doi: [10.1038/ng.2504](https://doi.org/10.1038/ng.2504) PMID: [23263488](https://pubmed.ncbi.nlm.nih.gov/23263488/)
- Westra H-J, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet.* 2013; 45: 1238–1243. doi: [10.1038/ng.2756](https://doi.org/10.1038/ng.2756) PMID: [24013639](https://pubmed.ncbi.nlm.nih.gov/24013639/)
- Westra H-J, Arends D, Esko T, Peters MJ, Schurmann C, Schramm K, et al. Cell Specific eQTL Analysis without Sorting Cells. *PLoS Genet.* 2015; 11: e1005223. doi: [10.1371/journal.pgen.1005223](https://doi.org/10.1371/journal.pgen.1005223) PMID: [25955312](https://pubmed.ncbi.nlm.nih.gov/25955312/)
- Fairfax BP, Makino S, Radhakrishnan J, Plant K, Leslie S, Dilthey A, et al. Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. *Nat Genet.* 2012; 44: 502–10. doi: [10.1038/ng.2205](https://doi.org/10.1038/ng.2205) PMID: [22446964](https://pubmed.ncbi.nlm.nih.gov/22446964/)

19. Raj T, Rothamel K, Mostafavi S, Ye C, Lee MN, Replogle JM, et al. Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes. *Science*. 2014; 344: 519–23. doi: [10.1126/science.1249547](https://doi.org/10.1126/science.1249547) PMID: [24786080](https://pubmed.ncbi.nlm.nih.gov/24786080/)
20. Ferraro A, D'Alise AM, Raj T, Asinovski N, Phillips R, Ergun A, et al. Interindividual variation in human T regulatory cells. *Proc Natl Acad Sci U S A*. 2014; 111: E1111–20. doi: [10.1073/pnas.1401343111](https://doi.org/10.1073/pnas.1401343111) PMID: [24610777](https://pubmed.ncbi.nlm.nih.gov/24610777/)
21. Naranbhai V, Fairfax BP, Makino S, Humburg P, Wong D, Ng E, et al. Genomic modulators of gene expression in human neutrophils. *Nat Commun*. 2015; 6: 7545. doi: [10.1038/ncomms8545](https://doi.org/10.1038/ncomms8545) PMID: [26151758](https://pubmed.ncbi.nlm.nih.gov/26151758/)
22. Leitsalu L, Haller T, Esko T, Tammesoo M-L, Alavere H, Snieder H, et al. Cohort Profile: Estonian Biobank of the Estonian Genome Center, University of Tartu. *Int J Epidemiol*. 2015; 44: 1137–1147. doi: [10.1093/ije/dyt268](https://doi.org/10.1093/ije/dyt268) PMID: [24518929](https://pubmed.ncbi.nlm.nih.gov/24518929/)
23. Tserel L, Kolde R, Limbach M, Tretyakov K, Kasela S, Kisand K, et al. Age-related profiling of DNA methylation in CD8<sup>+</sup> T cells reveals changes in immune response and transcriptional regulator genes. *Sci Rep*. 2015; 5: 13107. doi: [10.1038/srep13107](https://doi.org/10.1038/srep13107) PMID: [26286994](https://pubmed.ncbi.nlm.nih.gov/26286994/)
24. Knight JC. Genomic modulators of the immune response. *Trends Genet*. 2013; 29: 74–83. doi: [10.1016/j.tig.2012.10.006](https://doi.org/10.1016/j.tig.2012.10.006) PMID: [23122694](https://pubmed.ncbi.nlm.nih.gov/23122694/)
25. Nica AC, Dermitzakis ET. Expression quantitative trait loci: present and future. *Philos Trans R Soc Lond B Biol Sci*. 2013; 368: 20120362. doi: [10.1098/rstb.2012.0362](https://doi.org/10.1098/rstb.2012.0362) PMID: [23650636](https://pubmed.ncbi.nlm.nih.gov/23650636/)
26. Flutre T, Wen X, Pritchard J, Stephens M, Frazer K, Murray S, et al. A Statistical Framework for Joint eQTL Analysis in Multiple Tissues. Gibson G, editor. *PLoS Genet*. Public Library of Science; 2013; 9: e1003486. doi: [10.1371/journal.pgen.1003486](https://doi.org/10.1371/journal.pgen.1003486) PMID: [23671422](https://pubmed.ncbi.nlm.nih.gov/23671422/)
27. GTEx Consortium TGte, Welter D, MacArthur J, Morales J, Burdett T, Hall P, et al. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*. American Association for the Advancement of Science; 2015; 348: 648–60. doi: [10.1126/science.1262110](https://doi.org/10.1126/science.1262110) PMID: [25954001](https://pubmed.ncbi.nlm.nih.gov/25954001/)
28. Ueda H, Howson JMM, Esposito L, Heward J, Snook H, Chamberlain G, et al. Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. *Nature*. 2003; 423: 506–11. doi: [10.1038/nature01621](https://doi.org/10.1038/nature01621) PMID: [12724780](https://pubmed.ncbi.nlm.nih.gov/12724780/)
29. Scalapino KJ, Daikh DI. CTLA-4: a key regulatory point in the control of autoimmune disease. *Immunol Rev*. 2008; 223: 143–55. doi: [10.1111/j.1600-065X.2008.00639.x](https://doi.org/10.1111/j.1600-065X.2008.00639.x) PMID: [18613834](https://pubmed.ncbi.nlm.nih.gov/18613834/)
30. Okada Y, Wu D, Trynka G, Raj T, Terao C, Ikari K, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature*. 2014; 506: 376–81. doi: [10.1038/nature12873](https://doi.org/10.1038/nature12873) PMID: [24390342](https://pubmed.ncbi.nlm.nih.gov/24390342/)
31. de Jong VM, Zaldumbide A, van der Slik AR, Laban S, Koeleman BPC, Roep BO. Variation in the CTLA4 3'UTR has phenotypic consequences for autoreactive T cells and associates with genetic risk for type 1 diabetes. *Genes Immun*. Macmillan Publishers Limited; 2016; 17: 75–8. doi: [10.1038/gene.2015.51](https://doi.org/10.1038/gene.2015.51) PMID: [26656450](https://pubmed.ncbi.nlm.nih.gov/26656450/)
32. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res*. 2014; 42: D1001–6. doi: [10.1093/nar/gkt1229](https://doi.org/10.1093/nar/gkt1229) PMID: [24316577](https://pubmed.ncbi.nlm.nih.gov/24316577/)
33. Tang X-F, Zhang Z, Hu D-Y, Xu A-E, Zhou H-S, Sun L-D, et al. Association analyses identify three susceptibility Loci for vitiligo in the Chinese Han population. *J Invest Dermatol*. 2013; 133: 403–10. doi: [10.1038/jid.2012.320](https://doi.org/10.1038/jid.2012.320) PMID: [22951725](https://pubmed.ncbi.nlm.nih.gov/22951725/)
34. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007; 447: 661–78. doi: [10.1038/nature05911](https://doi.org/10.1038/nature05911) PMID: [17554300](https://pubmed.ncbi.nlm.nih.gov/17554300/)
35. Jin Y, Birlea SA, Fain PR, Ferrara TM, Ben S, Riccardi SL, et al. Genome-wide association analyses identify 13 new susceptibility loci for generalized vitiligo. *Nat Genet*. 2012; 44: 676–80. doi: [10.1038/ng.2272](https://doi.org/10.1038/ng.2272) PMID: [22561518](https://pubmed.ncbi.nlm.nih.gov/22561518/)
36. Hakonarson H, Qu H-Q, Bradfield JP, Marchand L, Kim CE, Glessner JT, et al. A novel susceptibility locus for type 1 diabetes on Chr12q13 identified by a genome-wide association study. *Diabetes*. 2008; 57: 1143–6. doi: [10.2337/db07-1305](https://doi.org/10.2337/db07-1305) PMID: [18198356](https://pubmed.ncbi.nlm.nih.gov/18198356/)
37. Petukhova L, Duvic M, Hordinsky M, Norris D, Price V, Shimomura Y, et al. Genome-wide association study in alopecia areata implicates both innate and adaptive immunity. *Nature*. 2010; 466: 113–7. doi: [10.1038/nature09114](https://doi.org/10.1038/nature09114) PMID: [20596022](https://pubmed.ncbi.nlm.nih.gov/20596022/)
38. Hirota T, Takahashi A, Kubo M, Tsunoda T, Tomita K, Doi S, et al. Genome-wide association study identifies three new susceptibility loci for adult asthma in the Japanese population. *Nat Genet*. 2011; 43: 893–6. doi: [10.1038/ng.887](https://doi.org/10.1038/ng.887) PMID: [21804548](https://pubmed.ncbi.nlm.nih.gov/21804548/)

39. Shi Y, Zhao H, Shi Y, Cao Y, Yang D, Li Z, et al. Genome-wide association study identifies eight new risk loci for polycystic ovary syndrome. *Nat Genet.* 2012; 44: 1020–5. doi: [10.1038/ng.2384](https://doi.org/10.1038/ng.2384) PMID: [22885925](https://pubmed.ncbi.nlm.nih.gov/22885925/)
40. Todd JA, Walker NM, Cooper JD, Smyth DJ, Downes K, Plagnol V, et al. Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nat Genet.* 2007; 39: 857–64. doi: [10.1038/ng2068](https://doi.org/10.1038/ng2068) PMID: [17554260](https://pubmed.ncbi.nlm.nih.gov/17554260/)
41. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. *PLoS Genet.* 2014; 10: e1004383. doi: [10.1371/journal.pgen.1004383](https://doi.org/10.1371/journal.pgen.1004383) PMID: [24830394](https://pubmed.ncbi.nlm.nih.gov/24830394/)
42. Yoshida H, Hunter CA. The Immunobiology of Interleukin-27. *Annu Rev Immunol.* 2015; 33: 417–443. doi: [10.1146/annurev-immunol-032414-112134](https://doi.org/10.1146/annurev-immunol-032414-112134) PMID: [25861977](https://pubmed.ncbi.nlm.nih.gov/25861977/)
43. Anjos S, Nguyen A, Ounissi-Benkhalha H, Tessier M-C, Polychronakos C. A common autoimmunity predisposing signal peptide variant of the cytotoxic T-lymphocyte antigen 4 results in inefficient glycosylation of the susceptibility allele. *J Biol Chem.* 2002; 277: 46478–86. doi: [10.1074/jbc.M206894200](https://doi.org/10.1074/jbc.M206894200) PMID: [12244107](https://pubmed.ncbi.nlm.nih.gov/12244107/)
44. Atabani SF, Thio CL, Divanovic S, Trompette A, Belkaid Y, Thomas DL, et al. Association of CTLA4 polymorphism with regulatory T cell frequency. *Eur J Immunol.* 2005; 35: 2157–62. doi: [10.1002/eji.200526168](https://doi.org/10.1002/eji.200526168) PMID: [15940668](https://pubmed.ncbi.nlm.nih.gov/15940668/)
45. Pruul K, Kisand K, Alnek K, Metsküla K, Heilman K, Peet A, et al. Expression of B7 and CD28 family genes in newly diagnosed type 1 diabetes. *Hum Immunol.* 2013; 74: 1251–7. doi: [10.1016/j.humimm.2013.07.007](https://doi.org/10.1016/j.humimm.2013.07.007) PMID: [23911738](https://pubmed.ncbi.nlm.nih.gov/23911738/)
46. Dixon AL, Liang L, Moffatt MF, Chen W, Heath S, Wong KCC, et al. A genome-wide association study of global gene expression. *Nat Genet.* 2007; 39: 1202–7. doi: [10.1038/ng2109](https://doi.org/10.1038/ng2109) PMID: [17873877](https://pubmed.ncbi.nlm.nih.gov/17873877/)
47. Schadt EE, Molony C, Chudin E, Hao K, Yang X, Lum PY, et al. Mapping the genetic architecture of gene expression in human liver. *PLoS Biol.* 2008; 6: e107. doi: [10.1371/journal.pbio.0060107](https://doi.org/10.1371/journal.pbio.0060107) PMID: [18462017](https://pubmed.ncbi.nlm.nih.gov/18462017/)
48. Levy S, Avni D, Hariharan N, Perry RP, Meyuhos O. Oligopyrimidine tract at the 5' end of mammalian ribosomal protein mRNAs is required for their translational control. *Proc Natl Acad Sci U S A.* 1991; 88: 3319–23. PMID: [2014251](https://pubmed.ncbi.nlm.nih.gov/2014251/)
49. Li Q, Makri A, Lu Y, Marchand L, Grabs R, Rousseau M, et al. Genome-wide search for exonic variants affecting translational efficiency. *Nat Commun.* 2013; 4: 2260. doi: [10.1038/ncomms3260](https://doi.org/10.1038/ncomms3260) PMID: [23900168](https://pubmed.ncbi.nlm.nih.gov/23900168/)
50. Sharifulin D, Khairulina Y, Ivanov A, Meschaninova M, Ven'yaminova A, Graifer D, et al. A central fragment of ribosomal protein S26 containing the eukaryote-specific motif YxxPKxYxK is a key component of the ribosomal binding site of mRNA region 5' of the E site codon. *Nucleic Acids Res.* 2012; 40: 3056–65. doi: [10.1093/nar/gkr1212](https://doi.org/10.1093/nar/gkr1212) PMID: [22167470](https://pubmed.ncbi.nlm.nih.gov/22167470/)
51. Lempainen J, Laine A-P, Hammias A, Toppari J, Simell O, Veijola R, et al. Non-HLA gene effects on the disease process of type 1 diabetes: From HLA susceptibility to overt disease. *J Autoimmun.* 2015; 61: 45–53. doi: [10.1016/j.jaut.2015.05.005](https://doi.org/10.1016/j.jaut.2015.05.005) PMID: [26074154](https://pubmed.ncbi.nlm.nih.gov/26074154/)
52. Li L, Iwamoto Y, Berezovskaya A, Boussiotis VA. A pathway regulated by cell cycle inhibitor p27Kip1 and checkpoint inhibitor Smad3 is involved in the induction of T cell tolerance. *Nat Immunol.* 2006; 7: 1157–65. doi: [10.1038/ni1398](https://doi.org/10.1038/ni1398) PMID: [17013388](https://pubmed.ncbi.nlm.nih.gov/17013388/)
53. Wang R, Han G, Wang J, Chen G, Xu R, Wang L, et al. The pathogenic role of interleukin-27 in autoimmune diabetes. *Cell Mol Life Sci.* 2008; 65: 3851–60. doi: [10.1007/s00018-008-8540-1](https://doi.org/10.1007/s00018-008-8540-1) PMID: [18931971](https://pubmed.ncbi.nlm.nih.gov/18931971/)
54. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* 2009; 5: e1000529. doi: [10.1371/journal.pgen.1000529](https://doi.org/10.1371/journal.pgen.1000529) PMID: [19543373](https://pubmed.ncbi.nlm.nih.gov/19543373/)
55. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2015.
56. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. *Bioinformatics.* 2008; 24: 1547–8. doi: [10.1093/bioinformatics/btn224](https://doi.org/10.1093/bioinformatics/btn224) PMID: [18467348](https://pubmed.ncbi.nlm.nih.gov/18467348/)
57. Kauffmann A, Gentleman R, Huber W. arrayQualityMetrics—a bioconductor package for quality assessment of microarray data. *Bioinformatics.* 2009; 25: 415–6. doi: [10.1093/bioinformatics/btn647](https://doi.org/10.1093/bioinformatics/btn647) PMID: [19106121](https://pubmed.ncbi.nlm.nih.gov/19106121/)
58. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A. National Academy of Sciences;* 2003; 100: 9440–5. doi: [10.1073/pnas.1530509100](https://doi.org/10.1073/pnas.1530509100) PMID: [12883005](https://pubmed.ncbi.nlm.nih.gov/12883005/)

59. Onengut-Gumuscu S, Chen W-M, Burren O, Cooper NJ, Quinlan AR, Mychaleckyj JC, et al. Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. *Nat Genet.* 2015; 47: 381–386. doi: [10.1038/ng.3245](https://doi.org/10.1038/ng.3245) PMID: [25751624](https://pubmed.ncbi.nlm.nih.gov/25751624/)
60. Sobel ME. Asymptotic intervals for indirect effects in structural equations models. *Sociol Methodol.* 1982; 13: 290–312.
61. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. 1st ed. Springer-Verlag New York; 2009.
62. Shin J-H, Blay S, McNeney B, Graham J. LDheatmap: An R Function for Graphical Display of Pairwise Linkage Disequilibria Between Single Nucleotide Polymorphisms. *J Stat Soft.* 2006; 16: Code Snippet 3.
63. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics.* 2010; 26: 2336–7. doi: [10.1093/bioinformatics/btq419](https://doi.org/10.1093/bioinformatics/btq419) PMID: [20634204](https://pubmed.ncbi.nlm.nih.gov/20634204/)
64. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. *Genome Res.* 2009; 19: 1639–45. doi: [10.1101/gr.092759.109](https://doi.org/10.1101/gr.092759.109) PMID: [19541911](https://pubmed.ncbi.nlm.nih.gov/19541911/)