Intersection of population variation and autoimmunity genetics in human T cell activation

Citation: Ye, C. J. et al. "Intersection of Population Variation and Autoimmunity Genetics in Human T Cell Activation." Science 345.6202 (2014): 1254665–1254665.

As Published: http://dx.doi.org/10.1126/science.1254665

Publisher: American Association for the Advancement of Science (AAAS)

Persistent URL: http://hdl.handle.net/1721.1/105882

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

HHS Public Access

Author manuscript

Science. Author manuscript; available in PMC 2016 February 11.

Published in final edited form as:

Science. 2014 September 12; 345(6202): 1254665. doi:10.1126/science.1254665.

Intersection of population variation and autoimmunity genetics in human T cell activation

Chun Jimmie Ye[1], **Ting Feng**[2], **Ho-Keun Kwon**[2], **Towfique Raj**[1,3], **Michael Wilson**[2], **Natasha Asinovski**[2], **Cristin McCabe**[1,2], **Michelle H. Lee**[3], **Irene Frohlich**[3], **Hyun-il Paik**[2], **Noah Zaitlen**[4], **Nir Hacohen**[2], **Barbara Stranger**[5], **Philip De Jager**[1,3], **Diane Mathis**[1,2], **Aviv Regev**[1,6,*], and **Christophe Benoist**[1,2,*]

[1]Broad Institute of MIT and Harvard, Cambridge, MA

[2]Division of Immunology, Department of Microbiology and Immunobiology, Harvard Medical School, Boston, MA

[3]Program in Translational NeuroPsychiatric Genomics, Institute for the Neurosciences, Departments of Neurology and Psychiatry, Brigham and Women's Hospital, Boston, MA

[4]Department of Medicine Lung Biology Center, University of California, San Francisco, CA

[5]Section of Genetic Medicine, The University of Chicago, Chicago, IL 60637, USA

[6]Howard Hughes Medical Institute, Department of Biology, MIT, Cambridge, MA 02139

Abstract

T lymphocyte activation by antigen conditions adaptive immune responses and immunopathologies, but we know little about its variation in humans, and its genetic or environmental roots. We analyzed gene expression in $CD4^+$ T cells during unbiased activation or in Th17 conditions from 348 healthy subjects representing European, Asian and African ancestries. We observed inter-individual variability, most marked for cytokine transcripts, with clear biases on the basis of ancestry, and following patterns more complex than simple Th1/2/17 partitions. We identified 39 genetic loci specifically associated in cis with activated gene expression. We further fine-mapped and validated a single-base variant that modulates YY1 binding and the activity of an enhancer element controlling the autoimmune-associated *IL2RA* gene, affecting its activity in activated but not regulatory T cells. Thus, inter-individual variability affects the fundamental immunologic process of T helper activation, with important connections to autoimmune disease.

*Correspondence to: aregev@broad.mit.edu, cbdm@hms.harvard.edu.

Supplementary Materials:
Materials and Methods
Figures S1–S8
Tables S1–S11

Additional author notes should be indicated with symbols (for example, for current addresses).

INTRODUCTION

Activation of T lymphocytes occurs through the recognition of the cognate antigen presented by molecules of the Major Histocompatibility Complex (MHC) and is arguably the linchpin of the adaptive immune system. This process initiates the amplification of rare T cells specific for a given antigen within the naïve repertoire, which rapidly builds the cell cohorts required for an effective response. T cell amplification also coincides with phenotypic differentiation into various "flavors" of effector $CD4^+$ T cells, each of which is beneficial in distinct contexts of microbial challenge (1): Th1 cells producing IFNγ and TNFα are most effective against viruses and intracellular microbes; Th2 cells producing IL-5 and IL-4/13 are most effective in controlling parasites; Th17 cells and the IL-17 family cytokines partake in anti-fungal and anti-bacterial defenses at mucosal interfaces; and Tfh cells help B cells to produce high-affinity immunoglobulins. Choices between effector phenotypes are themselves modulated by the cytokine network, such as the reinforcement of the Th17 identity through IL23. These pathways also drive major immuno-inflammatory diseases. Pathogenic Th1 or Th17 cells have been implicated in rheumatoid arthritis, multiple sclerosis (MS) and inflammatory bowel disease (IBD), and Th2-type responses in asthma and other atopic diseases (1).

Pathogens exert strong selective pressures during human evolution, including the migrations of the last tens of millennia (2). Optimal fitness, however, must balance pathogen clearance with the cost to the organism that may result from collateral tissue damage or autoimmune deviation (3, 4). Akin to the diversification of MHC alleles, it is plausible that selection for responses to different types of pathogens has diversified the overall efficacy and/or functional bias of T cell activation in humans. But selection for a Th phenotype, advantageous in one environment, might lead to sub-par responses to dominant pathogens in other geographic locales, or carry a penalty in terms of self/non-self discrimination and autoimmune diseases. Indeed, there is evidence that genetic variants that influence T cell activation can enhance susceptibility to immunologic diseases (5–8), such as the variants in the *IL2RA* locus, which encodes a key trophic cytokine receptor, associated with MS and type 1 diabetes (T1D) (9–14).

Little is known, however, about inter-individual variation in the responsiveness of human $CD4^+$ T cells. Are some individuals inherently more responsive to T cell activation, or more prone to mount Th1- or Th17-biased responses? Genetic variants in regulatory elements that affect transcript abundance can be mapped as expression quantitative trait loci (eQTLs) that associate individual genetic variation with resulting variation in gene expression (15, 16). While these can in principle facilitate the interpretation of disease-associated variants, most eQTL studies have been performed in cell-types or states that are not directly relevant to pathogenic processes involving T lymphocytes.

An experimental and analytic strategy to dissect inter-individual variation

We followed a stepwise approach to dissect the variation in the responses of primary T cells in 348 healthy subjects of African (91), Asian (74) and European (183) ancestry from the Boston PhenoGenetic Project, which were also analyzed in other arms of the ImmVar study (17, 18) (Fig. 1A, table S1). By design, we limited the age range (18 – 56 years), to

constrain the effects of age on immune function, and sampled subjects in a manner to minimize variation related to circadian or hormonal secretion rhythms, or food consumption (fig. S1). First, we established a rigorously standardized protocol to purify and activate, in a highly parallel mode, fresh $CD4^+$ T cells from human blood. Cells were activated via the T cell receptor (anti-CD3+CD28 beads – hereafter "α3+28") alone or in biasing conditions that favor differentiation to a Th17 phenotype (supplementation with TGFβ, IL6, IL23, IL1B, anti-IL4 and anti-IFNγ – hereafter "α3+28/Th17"), or with addition of IFNβ. These conditions allowed us to map inter-individual variation in T cell transcriptional responses that mimic activation through the antigen-specific receptor in an unbiased setting, or when coerced towards the Th17 phenotype. Genome-wide expression profiling established the time points to sample (with cells from a pool of donors) and the extent of inter-individual variability in these responses (with cells from a set of 15 donors). From these data we selected the five most informative conditions (stimulus time-point pairs) and a gene set of 236 transcripts (table S2) that best capture the responses and their variance across donors and activation states, as previously described (19), complemented with transcripts of known importance, including 16 key defining cytokines. We then used direct molecule counting (Nanostring) to measure the 236-transcript gene set in $CD4^+$ T cells from all 348 donors, in each of these five conditions. We estimated the stable inter-individual variation from a subset of 8 donors who were sampled twice, and distinguished components attributable to cis genetic or to non-genetic effects (demographic, physical and environmental) from the 183 donors of European ancestry. Leveraging the enhanced resolution offered by meta-analysis across population groups (20), we fine-mapped eQTLs that control these responses. On the basis of these results, we validated and dissected molecular interactions around a SNP that controls *IL2RA* expression upon T cell activation.

Interindividual variation in T cell responses, co-regulated cytokine clusters

Genome-wide expression profiling determined the signatures of dynamic T cell activation and differentiation responses to different stimuli in a pool of donors of different genders and ancestries (Fig. 1B,C). After α3+28 activation, we observed successive waves of induction consistent with patterns of responses to T cell activation *in vivo* (21, 22) (Fig. 1B,C), affecting 1,750 induced and 456 repressed genes (fold change (FC) > 1.68; table S3), including proliferation and effector genes (23, 24). In Th17-biased conditions, few transcripts (289) were affected that were not already induced in unbiased α3+28 conditions (we did observe the expected over-induction of the *IL17* family, and more surprisingly of *IFNG*, the prototypical Th1 cytokine). The 270 genes induced in response to IFNβ peaked between 2–4 hrs, including canonical IFN responsive targets. From these data, we selected representative time-points at 4 and 48 hrs for α3+28, 4 hrs for α3+28+IFNβ, and 48 hrs for α3+28/Th17, and unstimulated controls harvested at 4 hrs.

We determined the extent of inter-individual variation in T cell responses across 348 individuals, with Nanostring profiling of the 236-transcript set. For cytokine transcripts, there was a generally consistent pattern of responses across individuals (Fig. 1D; note the surprisingly early response of many cytokine genes [e.g. *IL4*, *IL17A*, *IL22*, *IFNG*], and the temporal disconnect between *IL17A* and *IL17F*). There was, however, a spectrum in the extent of inter-individual variability (Fig. 1D): large for *IL3* and *IL17A* (9.3- and 5.3-fold,

respectively), tight for *IL2* and *TNF* (2.4-fold; for comparison, the range for *ACTB* was 1.07). Indeed, a set of 14 transcripts, enriched for cytokines and chemokines (13 of 14, $p < 10^{-7}$) showed an increase in variance relative to unstimulated baseline, contrary to the bulk of responsive genes whose variance tended to shrink upon activation (fig. S2). The range of variability in cytokine responses was largely independent of the proportion of $CD62L^{lo}$ "memory" cells in the starting population, despite a slight positive correlation between *IL17F* and *IFNG* responses at 48 hrs and the proportion of $CD62L^{lo}$ cells (Pearson r = 0.07 and 0.12, fig. S3).

The intensity of initial T cell activation is an important determinant of the outcome of a response to viral pathogens (25). To assess differences in overall responsiveness, i.e. whether individuals differ in the general intensity of T cell responses, we calculated a "response index" by averaging the centered expression values of 51 transcripts induced (FC > 1.68, table S4) at 48 hrs of α3+28 activation. This index spanned a somewhat narrow range (1.69 fold across the 348 donors; fig. S1A) and was not or only minimally affected by subject gender or age (fig. S1B–D). The response index correlated strongly with a similar index computed from cytokine transcripts, which ranged over 3.4-fold (fig. S1A). This variability in the activation response is significantly correlated (PC2: Pearson $r = -0.52$, $p < 10^{-16}$; PC3: Pearson r = 0.36, $p < 10^{-12}$) with principal components computed over the entire expression matrix (fig. S1E,F).

We asked whether some individuals could be distinguished by "response-types", skewed towards the cytokine signature of a particular Th subtype. No simple "ThX-type" individuals emerged from the expression of signature cytokines: the levels of *IL17F* and *IL4* were largely independent (Pearson r = 0.07, p = 0.16), while the levels of *IL17F* and *IFNG* were positively correlated at 48 hrs (Pearson r = 0.6, $p < 10^{-15}$) (Fig. 2A). This complexity in individual responses is exemplified by a cluster of donors characterized by higher *IL4* and *IL9* expression, but which included some high *IL17F* or *IFNG* responders as well (Fig. 2B). Some of the high *IFNG* responders had low *IL4* responses, as would be predicted from classic Th1/Th2 opposition, but this was certainly not the norm. Importantly, these response-defining cytokines are highly repeatable between replicate draws from the same donors (fig. S4).

Some groups of cytokines are co-regulated in mouse T cells in response to particular pathogen challenges. We asked whether similar co-variation patterns of cytokine transcripts exist across healthy humans, as affected by genetic and environmental variation. We calculated, for each condition, the pairwise correlation between 16 cytokine transcripts. Early in α3+28 activation, many classic T cell cytokines tended to be correlated, with only limited differentiation of Th17 cytokines, and a well-individualized cluster of inflammatory cytokines (*IL1A/B, IL6;* Fig. 2C, **left panel**). Clusters became more refined after 48 hrs (Fig. 2C, **middle panel**), some of which were expected (*IL17A/F, IL22),* others more surprising (e.g. *IFNG* and *IL21*, or *IL4* with *IL9, IL10, and IL3*). Expression of most cytokines still exhibited some degree of positive correlation, perhaps indicative of general responsiveness. Indeed, the distinctions between cytokines became more apparent when the data were adjusted for 8 principal components (Fig. 2C, *right*).

To test whether variation at early times may be predictive of later outcomes, we computed pairwise correlations between genes across time-points (fig. S5) and found a positive correlation between the expression of the *IL17* family at 4 hrs with *IL17F* expression at 48 hrs; and of early *IFNG* with later *IFNG* production, indicating that inter-individual variation sets the type of response. This early bias was not an amplification of a pre-determined state, as there was no correlation with the same in unstimulated samples. Other baseline transcripts were correlated to responsiveness (26) including the abundance of IFN-I responsive transcripts (*DDX58*, *IFIH1*, *STAT2*) predicting later *IL17F* induction, suggesting an interesting connection between tonic IFN-I signals and the ability to mount IL17 responses (26).

Contributions of genetics and environment to repeatable variation in T cell response

To estimate the relative contributions of genetic and environmental factors, we first estimated for each gene its intra-individual repeatability with data from 8 individuals whose cells were sampled on two different dates and stimulated in the same conditions (Fig. 3A). The average repeatability (27) (LMM bootstrap $\bar{R} \pm \mathrm{SE}_{\bar{R}}$, table S5, medium fill in Fig. 3A) was highest in late response (45%±2% in either 48 hrs α3+28 or Th17), and lower in early response (20%±3% in 4 hrs α3+28). Although there was no general trend between repeatability and the coefficient of variation (CV), some of the most variable genes, including *IL17A/F*, *IL4*, *IFNG*, and *IL9*, were also among the most repeatable (Fig. S6).

Next, we estimated the proportion of variance attributable to known covariates or genetic components in samples only from the 183 donors of European ancestry. The proportion of expression variance explained by physiological covariates together (e.g. age, gender, height, weight, blood pressure, etc.) was generally low (2.4 to 3.6%, adjusted $\bar{r}^2_{\mathrm{fixed}} \pm \mathrm{SE}_{\bar{r}^2_{\mathrm{fixed}}}$, light fill in Fig. 3A,B), with sex (10 genes, t-test FDR < 0.05) and age (53 genes, F-test FDR < 0.05) the only covariates significantly correlated with gene expression in 48 hrs α3+28 (table S6). The average proportion of variance explained by cis heritability (28, 29) (permutation $\bar{h}^2_{cis} \pm \mathrm{SE}_{\bar{h}^2_{cis}}$, dark fill in Fig. 3A) was higher in 48 hrs activated conditions (9.2%±0.9%) than in unstimulated T cells (5.8%) and 4 hrs activated conditions (~6.6%). While the latter are consistent with previous estimates in whole tissue with identity-by-descent (28), the higher 48 hr estimates may reflect increased power to detect activation-specific eQTLs.

Repeatable variance attributable to common cis genetic and non-genetic factors varied widely between genes (Fig. 3A,B). Cytokine transcripts were repeatable in their expression upon activation for the 8 individuals tested, but common genetic variation did not contribute substantially to their variability across the full dataset of 348 individuals (Fig. 3B). In contrast, repeatable variation in cytokine receptors, in particular *IL23R* and *IL2RA*, was significantly driven by cis genetic factors (~50%) in activated conditions. Overall, our results suggest that although the T cell response is generally reproducible, common cis genetic effects account for ~25% of repeatable variability, physiological covariates account for very little (~4%) and the rest maybe due to established environmental effects, immunological history or *trans*-genetic regulation.

Population differentiation in T cell response

When examining differences in T cell responses across the cohort of 348 donors with a linear model, ancestry differentiated the expression of 94/229 (41%) genes in the two 48 hrs conditions (linear regression *F* test, FDR < 0.01, table S6), explaining a median 7% of the expression variance. We observed ordering in the mean expression of response genes, with a strong trend of over-expression for donors of African ancestry, lower for European ancestry, and a mixed pattern for Asian anestry (Fig. 3C). Differentially responsive genes include key indicators of Th phenotypes, *IL17* family cytokines (over-induced in individuals of African ancestry) and *IFNG*, which showed an opposite pattern. We are cognizant that such differences can be artefactually generated by systematic bias in sampling, batch or lifestyle confounders, but the differences observed likely have some genetic underpinning as sampling was from the same geographical area, over the same time period, at the same location, under a controlled protocol, and samples from different ancestries were randomly distributed across experimental batches. After adjusting the data for the first eight principal components to account for unknown confounders the pattern was weakened but still present for many genes (median 2.4% and 2.7% in the 48 hr conditions; Figs. 3D, S7). Even the most stringent estimates of population differentiation are higher than previous estimates (17, 30–33) which may partially be due to enrichment of immune related genes in our gene set consistent with previous observations (17, 30–33). Further, although the trend was present in unstimulated samples, it was amplified after 48 hr stimulation, indicating that the population differentiation required activation to be manifest.

In addition to genes with known population differentiation (*UTS2*, *IFITM3*; (18, 34)), we detected population differentiation in transcripts encoding cytokines, chemokines or their receptors (table S6, Fig. 3C). Notably, *IL2RA* was one of the most statistically differentially expressed genes (t-test, $p < 10^{-16}$ in both 48 hrs activation conditions, p = 0.03 after PCA adjustment): on average, activated T cells from Europeans expressed ~15% less *IL2RA* mRNA than individuals of African ancestry. Together, these results suggest that there is strong population differentiation in T cell responses even after stringent removal of the general response and may include a genetic component.

Genetic variants associated with expression variation during T cell activation and Th17 polarization

To identify the genetic variants underpinning expression differences, we associated ~10M imputed genotypes (1000 Genomes Project (35), MAF > 0.05) from our cohort of 348 donors with gene expression levels in each condition. Almost half the genes in our gene set (112 of 222 - excluding X and Y-encoded genes; Fig. 4A, table S7) had a significant cis-eQTL within 1MB around the gene, in at least one state (permutation FDR < 0.01, table S8). Some cis-eQTLs were detected in all conditions (Fig. 4B), but 38 were state-specific (12 late, 2 early, 11 pan activation and 13 other patterns, table S7). We did not detect any cis-eQTLs specific to Th17 polarizing nor IFNβ conditions beyond those detected in unbiased activation alone, although a few eQTLs (8 in Th17, 5 in IFNβ) had larger effect sizes (table S7). Some of the condition-specific cis-eQTLs were related to transcript levels (Fig. 4C,D), but this was not the case for many (Fig. 4E,F) - *IL2RA* was expressed prior to stimulation, but only had a significant cis-eQTL in the late activation conditions (rs12251836, Meta p <

10^{-16}). The low-expression variant at rs12251836 was rare in African (MAF = 0.13) and Asian (MAF = 0.03), but common in European genomes (MAF = 0.41) (Fig. 4G).

We identified only six *trans*-eQTLs at genome-wide significance (permutation FDR < 0.2), including rs6498114, a variant in the CIITA locus, a gene known to control MHC class-II expression, which fittingly correlated with *HLA-DRA* expression (table S9). When restricting to test for *trans* impact only those SNPs identified as having local cis effects, we identified six other significant signals, such as the association of the *IL23R* cis-eQTLs (rs2064689, rs2863204) in the α3+28/Th17 condition with the change in expression of *IL21* between 48 hr α3+28/Th17 and unbiased stimulation (table S10), explaining 11%+/−8.7% of the variance (permutation FDR < 0.2; mediation analysis by conditioning on *IL23R* transcript levels excluded that the effect might be due to the nearby H3Q codon change); consistent with the amplification of IL21 production by IL23 signals during Th-17 differentiation (36, 37).

Overlap of genetically variable T cell responses with GWAS loci identifies potential disease relevant variants

To assess the relevance of activation eQTLs to disease, we intersected our eQTLs with the GWAS catalog and identified intriguing disease-associated variants, notably near *IL23R* and *IL2RA* (Fig. 5, table S11), two key cytokine receptors implicated in a number of autoimmune diseases.

IL23R encodes the receptor for IL23, partakes in establishing stable Th17 differentiation (38) and is strongly associated with IBD (39, 40). Here, rs2863204 was the variant most significantly associated with *IL23R* expression in 48 hr α3+28/Th17 (Fig. 5A). It is not on the same haplotype as the IBD-associated non-synonymous variant (rs11209026, R381Q) (40). However, meta-analysis of ulcerative colitis (UC) and Crohn's disease (CD) separately (40) revealed allelic heterogeneity in *IL23R* where an additional variant, tagged by rs12095335 is associated with CD, but not UC (Fig. 5B). When conditioned on rs2863204, rs12095335 appeared as a secondary association to *IL23R* mRNA levels (Meta $p = 2x10^{-4}$; Fig. 5C), suggesting that there may be two independent effects on *IL23R* expression in the region. The trans-association between these *IL23R* variants and *IL21* expression in the Th17 condition further implies functional consequences for the region with regards to Th17 differentiation.

Perhaps most intriguing was the discovery of SNPs associated with *IL2RA* expression. *IL2RA* encodes CD25, the high-affinity IL2 receptor, and has intricately balanced functions, reflecting the divergent roles of IL2 as trophic factor for inhibitory Treg cells and in aiding the expansion of pro-inflammatory effector T or NK cells (41). Complementary susceptible and protective haplotypes in MS and T1D have been demonstrated (9, 12, 42). These map to the promoter region or first intron of *IL2RA*, and have been diversely associated to titers of soluble IL2 receptor, expression of CD25 at the cell surface, or IL2 signaling efficacy (12, 43, 44). Three independent haplotypes have been reported to be associated to T1D. The strongest, a protective haplotype tagged by rs12722495 (Fig. 5D), leads to higher expression of *IL2RA* in steady-state $CD45RA^{-}$ memory and Treg cells (43–45). Here, rs12251836, the most significant variant associated with *IL2RA* expression in 48 hr α3+28 activated cells,

was unlinked to rs12722495 which showed little signal (Fig. 5E). However, by conditioning on rs12251836, we detected in our data a secondary association between rs12722495 and *IL2RA* expression in activated cells ($p = 5.3x10^{-7}$, Fig. 5F). Together, these results point to functional effects of at least two independent associations in the *IL2RA* locus, which may have balancing effects. We surmise that the associated variant rs12251836 is mainly active in acutely triggered T effector cells, but not in Treg cells, since *IL2RA* mRNA in unstimulated $CD4^+$ T cells would predominantly come from $CD25^{hi}$ Treg cells, and this association was not apparent in unstimulated cells. On the other hand the rs12722495 variant, weaker here but dominant in fresh $CD4^+$ blood T cells (43), might preferentially operate in chronically triggered cells such as Treg cells. On balance, these two variants would then control the effectiveness of IL2 signals in cells of opposite outcome in a disease context. The divergent GWAS results for T1D and MS may also result from this interplay.

A disease-related SNP affecting YY1 binding and activity of an enhancer motif in IL2RA

Given the overlap between the *IL2RA* eQTL and autoimmunity-associated GWAS SNPs, we investigated the functional effects of the top 5 candidate SNPs located in the 5′ flanking and first intron of *IL2RA* (Fig. 6A). We first tested for enhancer activity by introducing ~200 bp fragments centered around each SNP (Fig. 6A) into a luciferase reporter vector, driven by SV40 or *IL2RA* minimal promoters, and transfecting them into Jurkat T cells (+/− PMA +Ionomycin activation). Four of the segments were essentially inactive (or even had slight inhibitory effects, fig. S8) but fragment IV, which encompasses rs12251836, showed enhancer activity (“836” for short in Fig. 6B). The A allele, associated with lower *IL2RA* expression, had lower enhancer activity relative to the T allele (Fig. 6B). The higher enhancer function of the T allele was dependent on activation of the Jurkat cells (Fig. 6C), suggesting that enhancer activity is boosted upon cell activation by interactions affected by the single base change at rs12251836.

We examined whether the rs12251836 SNP affects the binding of transcription factors (TFs). After incubation of nuclear extracts from primary human CD4+ T cells, we detected protein binding to a 25 bp oligonucleotide encompassing $rs12251836^T$, but not to the $rs12251836^A$ counterpart (Fig. 6D). Furthermore, the formation of this DNA-protein complex in primary T cells depended on their stimulation (Fig. 6E), paralleling enhancer activity in the region. Searching for TF-binding motifs showed that the $rs12251836^T$ sequence has putative ($p < 10^{-5}$) binding sites for YY1 and RUNX1 straddling the SNP, with a FOX-binding site immediately downstream (Fig. 6F) (46). The 836^A allele maintains the FOX-binding motif, but the predicted RUNX and YY1 sites are lost. To test these predicted interactions, we expressed epitope-tagged versions of these three TFs and assayed for binding to oligonucleotides spanning this segment. FOXP3 and RUNX1 bound the rs12251836 oligonucleotides equally for both alleles, but binding of YY1 was restricted largely to $rs12251836^T$ allele. The allele-specific binding of YY1 was also confirmed by EMSA (fig. S8). Furthermore, transfection of a YY1 expression vector boosted the enhancer activity of fragment IV, but only with the plasmid carrying the $rs12251836^T$ variant (Fig. 6G). These results show that this variant is nested in an activation-augmented enhancer region, and mediates differential *IL2RA* expression by disrupting a functional YY1-binding site.

DISCUSSION

The exploration of how natural variation affects gene expression during human T lymphocyte activation and differentiation identified complex patterns of variability between individuals, more complex than simple Th paradigms. Ancestry had a surprisingly profound effect on these response patterns.

By focusing on primary cells, and profiling them in the activated cellular state that has the most likelihood to reveal regulatory variation related to autoimmune pathogenesis, we estimated higher cis-heritability and identified novel genetic associations. This work represents the third arm of the ImmVar project, in which the same donors were profiled with respect to fresh immunocytes at baseline (17) and to activation responses in dendritic cells (18). 69 of the 112 eQTLs discovered here were not detected in baseline $CD4^+$ T cells of Raj et al (17). These eQTLs were discovered for genes of direct importance in anti-infectious responses and autoimmunity (*IL23R*, *IL2RA*). Of the eQTLs for interferon responsive genes, 11/15 overlapped with those detected in interferon-treated DCs (18) reflecting the sharing of IFNβ response pathways. A comprehensive elucidation of the genetic control of immune responses will require similar analyses in an array of relevant cell-types under relevant conditions of stimulation.

Are some individuals more prone to mount Th-biased responses? Of the changes elicited by T cell triggering, cytokine transcripts showed the strongest change in inter-individual variance. This diversification did not generally track along simple Th1/2/17 models, but adopted more complex patterns. This observation is of significance for the relationship between T cell differentiation profiles and immune diseases and for the definition of biomarkers: rather than correlating disease or therapeutic outcome with levels of a single cytokine, it may be necessary to map individuals to a multidimensional outcome. Interestingly, very few cis-eQTLs were detected for cytokine transcripts but many were found for cytokine receptors. This contrast implies that genetic fluctuations are tolerated at cytokine receptor loci, while the cytokines themselves may be under more uniform genetic control but vary with environmental cues or the individual's immunological history. In addition, the inter-individual variance exhibited a general "responsiveness"; a widespread correlation between all cytokines, with fine distinctions between clusters of co-regulated cytokines that only were apparent over time in culture. Thus, individuals vary in the quantitative responsiveness of their T cells, as they do in the qualitative sense.

The study uncovered an unexpected but substantial population differentiation in overall immune response and in the expression of genes encoding key cytokines, with a trend towards higher activation in cells from donors of African ancestry. This differentiation was almost as pronounced as that of transcripts well recognized as being differentially expressed between human populations due to known genetic polymorphisms, such as *UTS2* or *GSTM1*. Suggestively, the IL17 cytokines were in that group, such that these differences may have direct consequences on immunopathology. It is possible that some differences in responsiveness reflect disparate nutritional or environmental influences in addition to the genetic component, but many (34/73 in 48 hr α3+28/Th17) involve cis genetic differences, for instance the rs12251836 variant that controls *IL2RA* through differential binding of YY1.

The low responder allele is a derived allele, present almost exclusively in donors of European descent, rare in African and Asian descendants. The lower responses to activation, and over-representation of the corresponding alleles, in European-derived population suggests that tolerance mechanisms to avoid immunopathology resulting from exuberant T cell responses may also have led to the evolution of protection from autoimmunity.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful to study participants in the Phenogenetic project for their contribution. We thank ImmVar participants, V. Kuchroo and D. Hafler for valuable discussions, K. Rothamel for processing of the microarray samples, C. Laplace for graphics. This work was supported by the US National Institute of General Medical Sciences grant RC2 GM093080 (C.B.). T.R. was supported by the NIH F32 Fellowship (F32 AG043267). P.L.D is a Harry Weaver Neuroscience Scholar Award Recipient of the National Multiple Sclerosis Society (JF2138A1). The gene expression data have been deposited at the Gene Expression Omnibus under accession no. 60236.

References and Notes

1. Zhu J, Yamane H, Paul WE. Differentiation of effector CD4 T cell populations (*). Annual review of immunology. 2010; 28:445–489.
2. Fumagalli M, et al. Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. PLoS Genet. 2011; 7:e1002355. [PubMed: 22072984]
3. Zinkernagel RM, Hengartner H. T-cell-mediated immunopathology versus direct cytolysis by virus: implications for HIV and AIDS. Immunology today. 1994; 15:262–268. [PubMed: 7915115]
4. Medzhitov R, Schneider DS, Soares MP. Disease tolerance as a defense strategy. Science. 2012; 335:936–941. [PubMed: 22363001]
5. Siminovitch KA. PTPN22 and autoimmune disease. Nat Genet. 2004; 36:1248–1249. [PubMed: 15565104]
6. Evans DM, et al. Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility. Nat Genet. 2011; 43:761–767. [PubMed: 21743469]
7. C Genetic Analysis of Psoriasis et al. A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between HLA-C and ERAP1. Nat Genet. 2010; 42:985–990. [PubMed: 20953190]
8. Wang K, Li M, Hakonarson H. Analysing biological pathways in genome-wide association studies. Nature reviews Genetics. 2010; 11:843–854.
9. Lowe CE, et al. Large-scale genetic fine mapping and genotype-phenotype associations implicate polymorphism in the IL2RA region in type 1 diabetes. Nat Genet. 2007; 39:1074–1082. [PubMed: 17676041]
10. Barrett JC, et al. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. Nat Genet. 2009; 41:703–707. [PubMed: 19430480]
11. Cooper JD, et al. Meta-analysis of genome-wide association study data identifies additional type 1 diabetes risk loci. Nat Genet. 2008; 40:1399–1401. [PubMed: 18978792]
12. Maier LM, et al. IL2RA genetic heterogeneity in multiple sclerosis and type 1 diabetes susceptibility and soluble interleukin-2 receptor production. PLoS Genet. 2009; 5:e1000322. [PubMed: 19119414]
13. C. International Multiple Sclerosis Genetics et al. Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. Nat Genet. 2013; 45:1353–1360. [PubMed: 24076602]

14. C. International Multiple Sclerosis Genetics et al. Risk alleles for multiple sclerosis identified by a genomewide study. The New England journal of medicine. 2007; 357:851–862. [PubMed: 17660530]
15. Montgomery SB, Dermitzakis ET. From expression QTLs to personalized transcriptomics. Nature reviews Genetics. 2011; 12:277–282.
16. Gilad Y, Rifkin SA, Pritchard JK. Revealing the architecture of gene regulation: the promise of eQTL studies. Trends in genetics: TIG. 2008; 24:408–415. [PubMed: 18597885]
17. Raj T, et al. Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes. 2014
18. Lee MN, et al. Common genetic variants modulate pathogen-sensing responses in human dendritic cells. Science. 2014; 343:1246980. [PubMed: 24604203]
19. Donner Y, Feng T, Benoist C, Koller D. Imputing gene expression from selectively reduced probe sets. Nature methods. 2012; 9:1120–1125. [PubMed: 23064520]
20. Zaitlen N, Pasaniuc B, Gur T, Ziv E, Halperin E. Leveraging genetic variability across populations for the identification of causal variants. American journal of human genetics. 2010; 86:23–33. [PubMed: 20085711]
21. Li L, Nishio J, van Maurik A, Mathis D, Benoist C. Differential response of regulatory and conventional CD4(+) lymphocytes to CD3 engagement: clues to a possible mechanism of anti-CD3 action? Journal of immunology. 2013; 191:3694–3704.
22. Best JA, et al. Transcriptional insights into the CD8(+) T cell response to infection and memory T cell formation. Nature immunology. 2013; 14:404–412. [PubMed: 23396170]
23. Feske S, Giltnane J, Dolmetsch R, Staudt LM, Rao A. Gene regulation mediated by calcium signals in T lymphocytes. Nature immunology. 2001; 2:316–324. [PubMed: 11276202]
24. Hess K, et al. Kinetic assessment of general gene expression changes during human naive CD4+ T cell activation. International immunology. 2004; 16:1711–1721. [PubMed: 15492022]
25. Jenkins MK, Moon JJ. The role of naive T cell precursor frequency and recruitment in dictating immune response magnitude. Journal of immunology. 2012; 188:4135–4140.
26. Delmastro MM, et al. Modulation of redox balance leaves murine diabetogenic TH1 T cells “LAG-3-ing” behind. Diabetes. 2012; 61:1760–1768. [PubMed: 22586584]
27. Falconer, DS.; Mackay, TFC. Introduction to quantitative genetics. 4. Longman; Essex, England: 1996. p. xiiip. 464
28. Price AL, et al. Single-tissue and cross-tissue heritability of gene expression via identity-by-descent in related or unrelated individuals. PLoS Genet. 2011; 7:e1001317. [PubMed: 21383966]
29. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. American journal of human genetics. 2011; 88:76–82. [PubMed: 21167468]
30. Storey JD, et al. Gene-expression variation within and among human populations. American journal of human genetics. 2007; 80:502–509. [PubMed: 17273971]
31. Stranger BE, et al. Patterns of cis regulatory variation in diverse human populations. PLoS Genet. 2012; 8:e1002639. [PubMed: 22532805]
32. Rosenberg NA, et al. Genetic structure of human populations. Science. 2002; 298:2381–2385. [PubMed: 12493913]
33. Lappalainen T, et al. Transcriptome and genome sequencing uncovers functional variation in humans. Nature. 2013; 501:506–511. [PubMed: 24037378]
34. Stranger BE, et al. Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science. 2007; 315:848–853. [PubMed: 17289997]
35. C. Genomes Project et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012; 491:56–65. [PubMed: 23128226]
36. Aggarwal S, Ghilardi N, Xie MH, de Sauvage FJ, Gurney AL. Interleukin-23 promotes a distinct CD4 T cell activation state characterized by the production of interleukin-17. The Journal of biological chemistry. 2003; 278:1910–1914. [PubMed: 12417590]
37. Zhou L, et al. IL-6 programs T(H)-17 cell differentiation by promoting sequential engagement of the IL-21 and IL-23 pathways. Nature immunology. 2007; 8:967–974. [PubMed: 17581537]

38. Korn T, Bettelli E, Oukka M, Kuchroo VK. IL-17 and Th17 Cells. Annual review of immunology. 2009; 27:485–517.

39. Duerr RH, et al. A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. Science. 2006; 314:1461–1463. [PubMed: 17068223]

40. Jostins L, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature. 2012; 491:119–124. [PubMed: 23128233]

41. Boyman O, Sprent J. The role of interleukin-2 during homeostasis and activation of the immune system. Nature reviews Immunology. 2012; 12:180–190.

42. Smyth DJ, et al. Shared and distinct genetic variants in type 1 diabetes and celiac disease. The New England journal of medicine. 2008; 359:2767–2777. [PubMed: 19073967]

43. Dendrou CA, et al. Cell-specific protein phenotypes for the autoimmune locus IL2RA using a genotype-selectable human bioresource. Nat Genet. 2009; 41:1011–1015. [PubMed: 19701192]

44. Cerosaletti K, et al. Multiple autoimmune-associated variants confer decreased IL-2R signaling in CD4+ CD25(hi) T cells of type 1 diabetic and multiple sclerosis patients. PloS one. 2013; 8:e83811. [PubMed: 24376757]

45. Orru V, et al. Genetic variants regulating immune cell levels in health and disease. Cell. 2013; 155:242–256. [PubMed: 24074872]

46. Ono M, et al. Foxp3 controls regulatory T-cell function by interacting with AML1/Runx1. Nature. 2007; 446:685–689. [PubMed: 17377532]

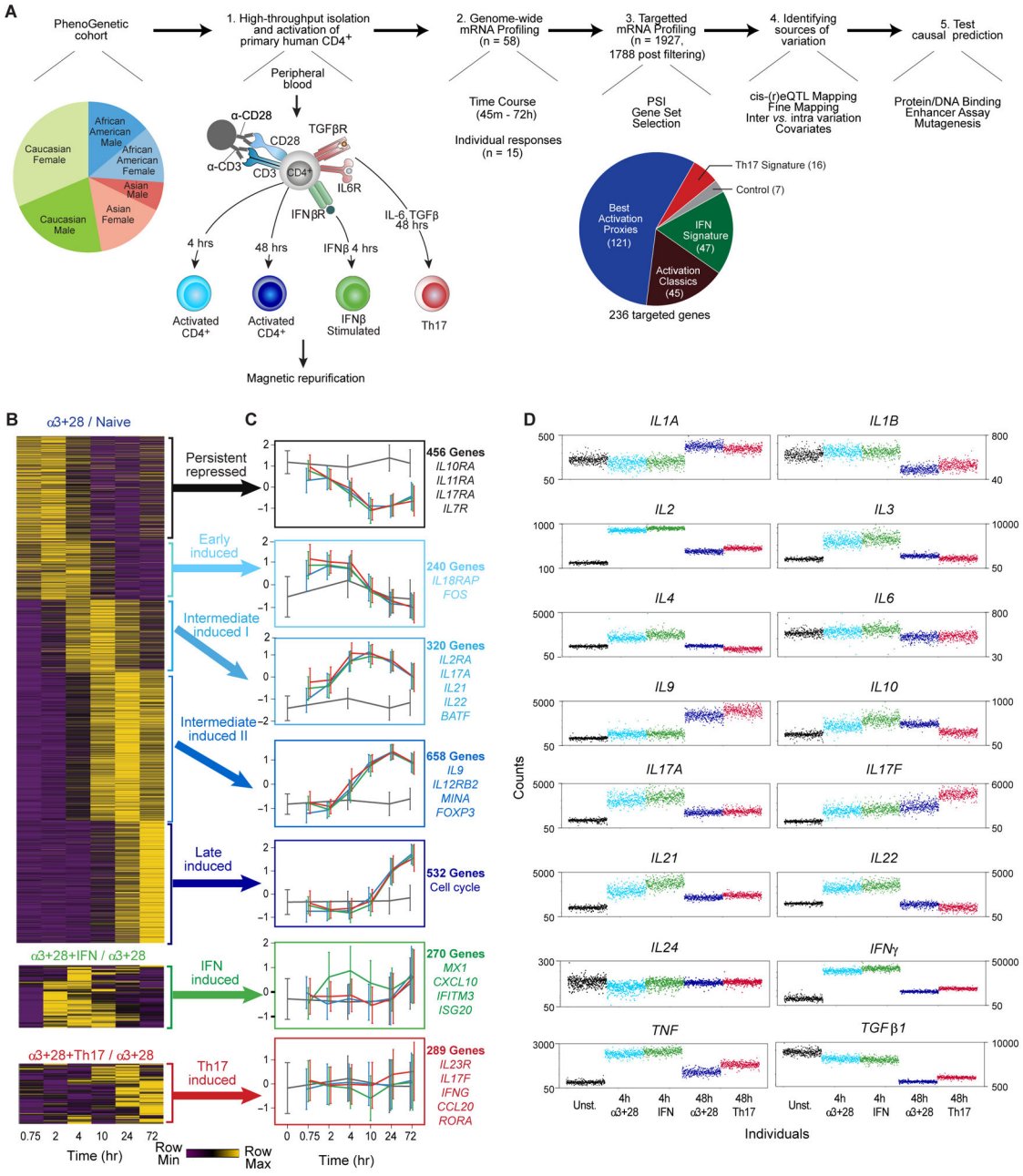

Fig. 1. A systematic approach to characterize the variation in T cell activation and response

(**A**) Overview of study, from (1) cell purification and activation in four conditions, (2) genomewide or (3) signature expression profiling, (4) computational analysis to decompose expression variance and identify eQTLs and (5) functional validation of potential causal variants. (**B**) Time course expression profiles of expression in unbiased α3+28 cells (fold change relative to 0 hr baseline, top), and response to IFNβ (fold change relative to unbiased α3+28, middle) or to Th17 polarization cytokines (fold change relative to unbiased α3+28, bottom). Each row (gene) is normalized to its mean fold change. (**C**) Profiles of gene expression for seven gene clusters, for conditions outlined and color-coded as in **A**. A

sample of representative genes in each cluster is listed. (**D**) Expression of 16 cytokines across 348 individuals and 5 conditions (black = unstimulated, other color codes as in **A**).

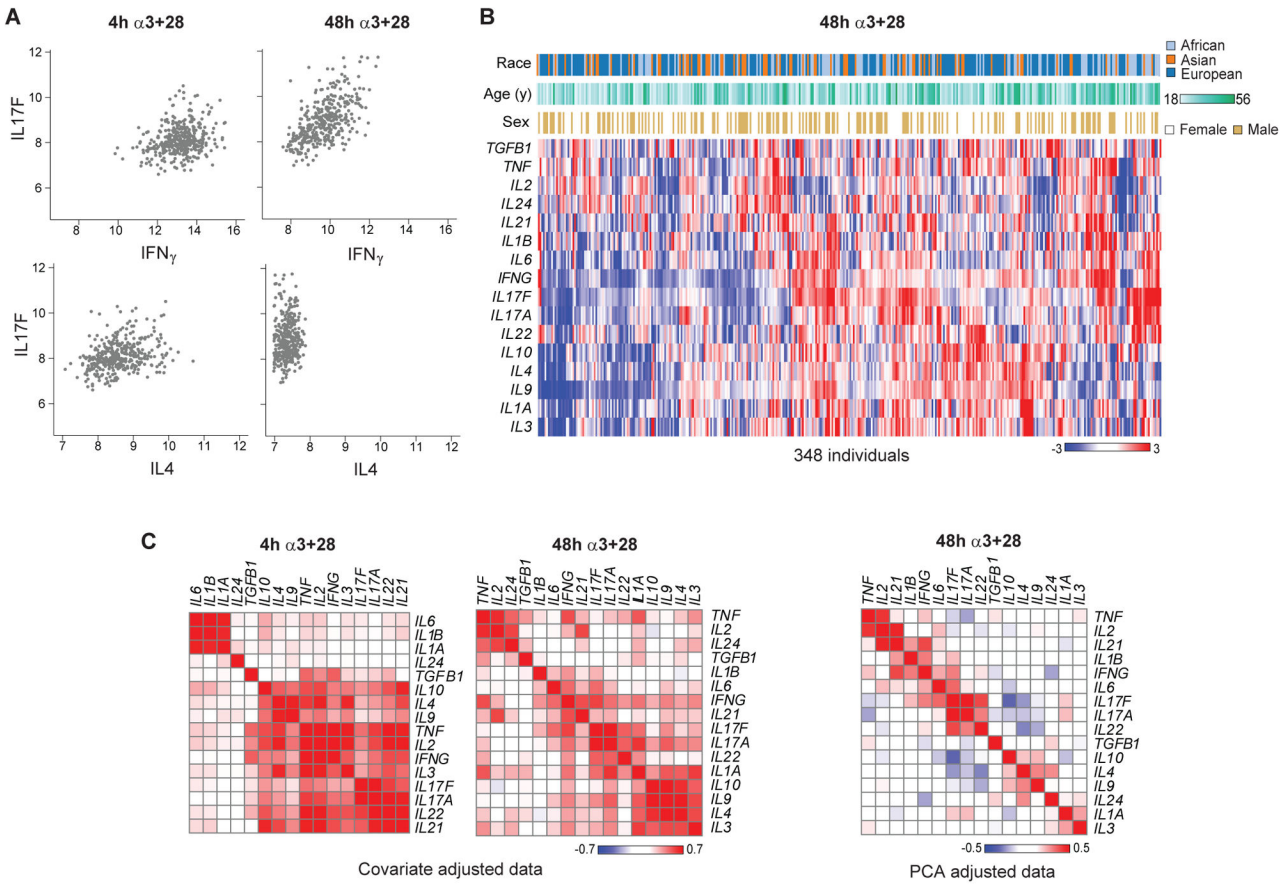

Fig. 2. Inter-individual variation in T cell response

(**A**) Relation between expression values (log2 scale) for pairs of Th defining cytokines (IL4, IL17F and IFNγ) in unbiased 4 hr α3+28 and 48 hr α3+28 conditions, each point is an individual sample. (**B**) Heatmap of 16 cytokines clustered by expression across 348 individuals in 48 hr α3+28 with demographic covariates race, age and sex. (**C**) Pairwise correlation between 16 cytokines in covariate-adjusted data during early activation (left), late activation (middle), and late activation with PCA-adjusted data to account for overall responsiveness (right).

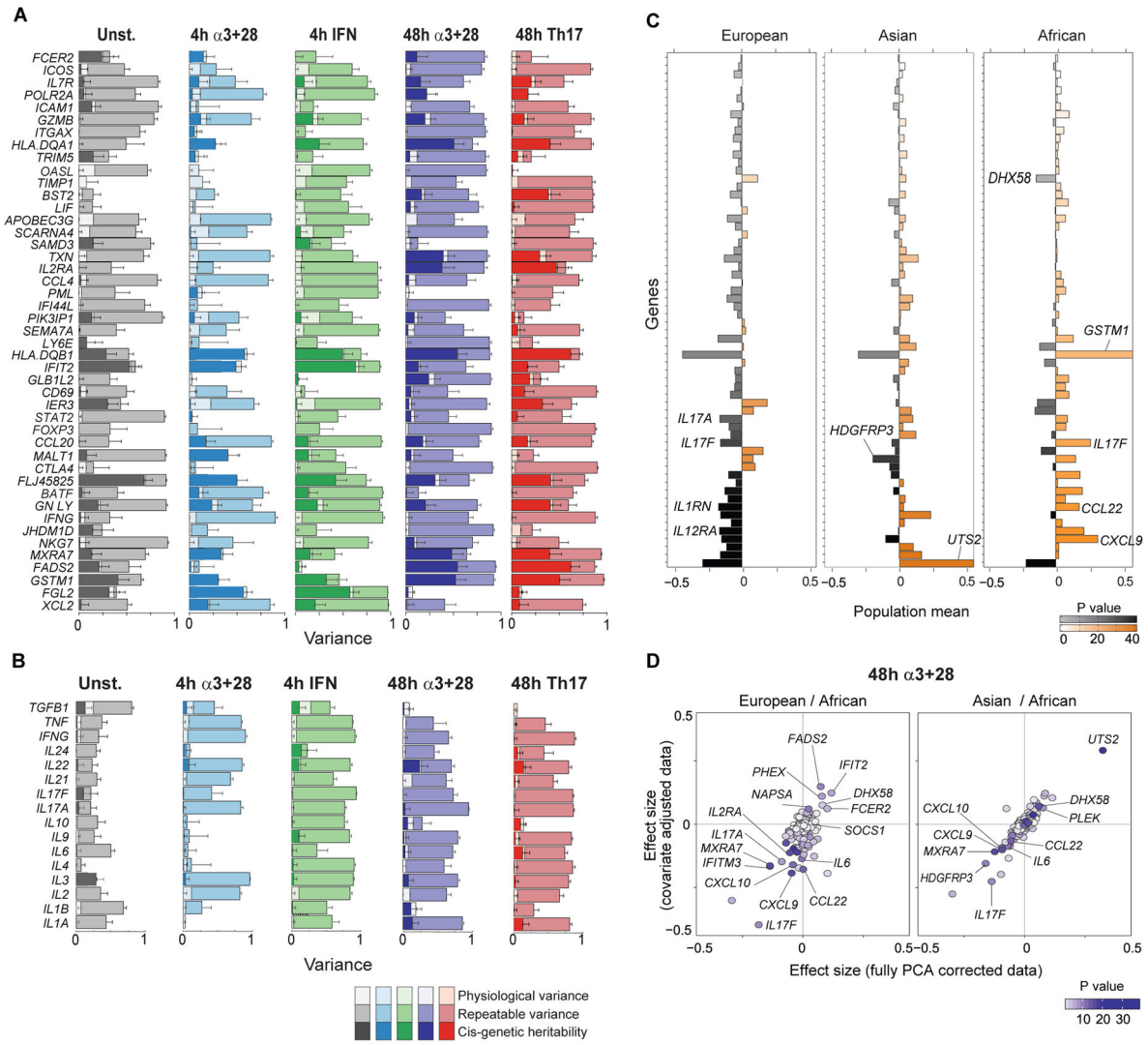

Fig. 3. Sources of variation across T cell activation conditions

(**A**) Top 56 non-cytokine genes or (**B**) 16 key cytokines ranked by maximum repeatability across all conditions estimated from 8 recalled individuals. For each gene, repeatability (medium shading), contributions of cis heritability (dark shading) and of physiological covariates (age, gender, body-mass index, weight, height, diastolic, systolic, season and month; light shading) to repeatable variation are estimated (European cohort only). Cis heritability is defined as the proportion of phenotypic variance explained by variants within 1MB of the gene estimated using linear mixed models (29). Physiological variance explained is the proportion of variance explained by all known covariates including gender, age, weight, height, BMI, blood pressure estimated from multiple regression. Error bars indicate bootstrap standard errors around each estimate. (**C**) Percent difference of average population expression (median) from overall average (median) that show population differentiation in expression. (**D**) Fold change (effect size) in normalized expression for donors of European (left, N = 183) and Asian (right, N = 74) ancestry relative to donors of African ancestry (N=91), comparing 48 hrs α3+28 data adjusted for all PCs (x-axis) or for

experimental and physiological covariates only (y-axis). Darker shading indicates more statistically significant population differentiated genes as determined by an omnibus *F* test.

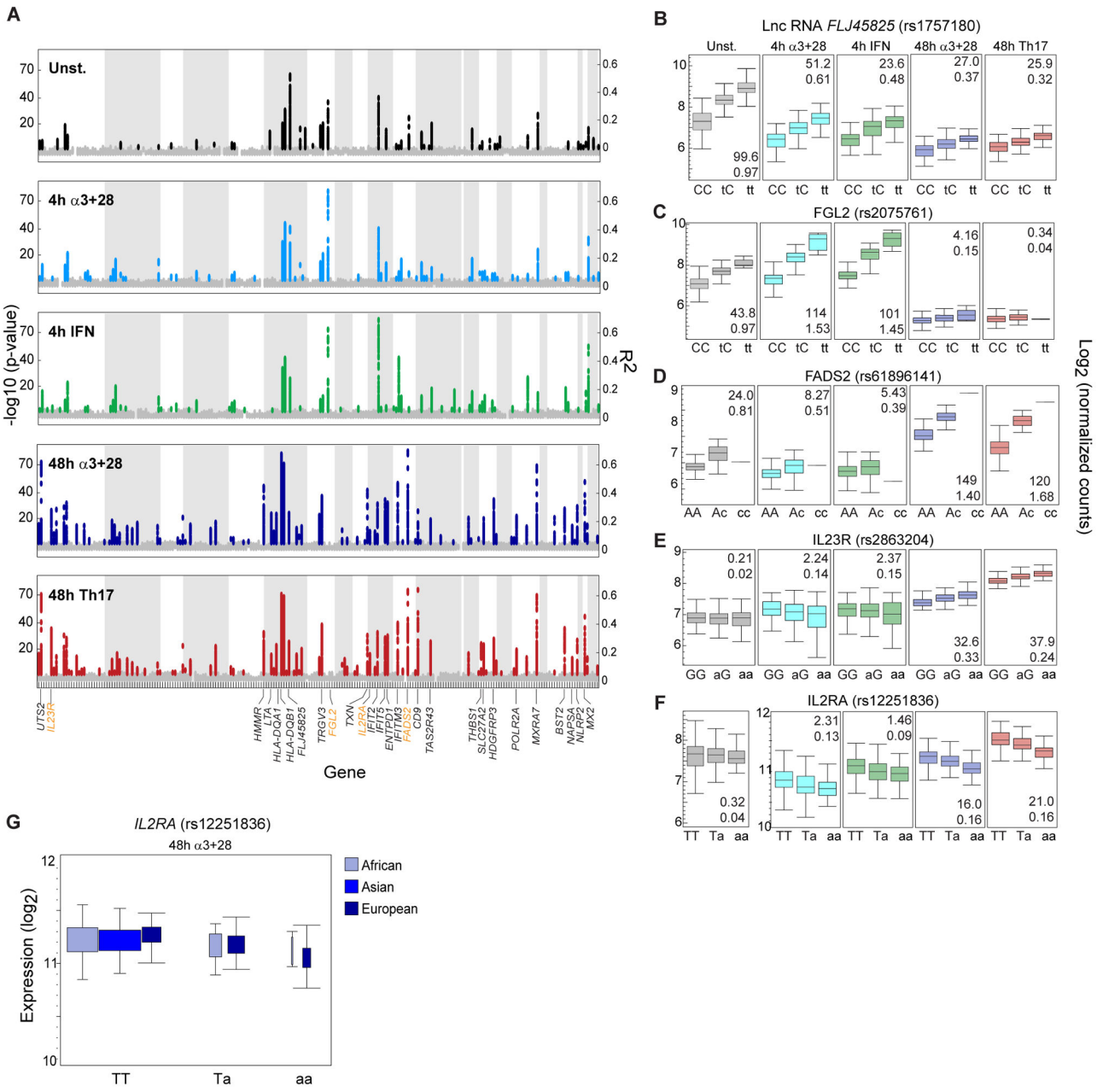

Fig. 4. Common cis variation associated with gene expression during T cell activation and differentiation

(**A**) Manhattan plot of the significance of cis (within 1M of gene) associations for the 236 genes in each condition. Permutation significant (FDR < 0.01) associations are colored. (**B–F**) Individual plots of expression per genotype for illustrative SNPs, numbers in each panel indicate the significance (−log10(p-value), upper number) and effect size (beta, lower number) of association. (**G**) Per genotype plots of *IL2RA* expression in 48 hr α3+28, distinguished by population. Width of boxes representing each genotype is scaled by its frequency in the population.

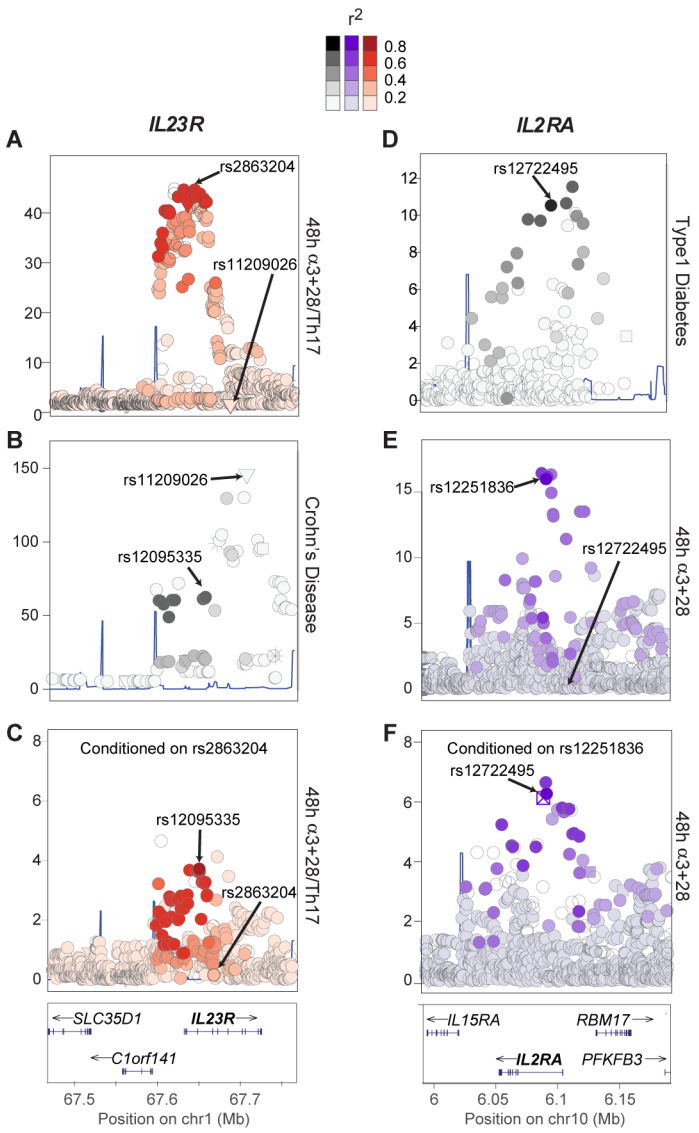

Fig. 5. Trans-ancestry fine mapping identifies activation specific cis-eQTLs in GWAS regions

(**A**) In 48 hr α3+28/Th17, rs2863204 (MAF = 0.53, European) is best associated with *IL23R* expression and is independent from rs11209026 (R381Q). (**B**) At *IL23R*, a coding variant (rs11209026, R381Q, MAF = 0.06, European) and a regulatory variant (rs12095335) are associated with Crohn's disease (CD). (**C**) Conditioning on rs2863204 recovers rs12095335 (secondary CD association) as a significant secondary association. (**D**) At the *IL2RA* locus, the best T1D association is a regulatory variant, rs12722495. (**E**) In 48 hr α3+28, rs12251836 (MAF = 0.41, European) is best associated with *IL2RA* expression and is independent from rs12722495. (**F**) Conditioning on rs12251836 recovers rs12722495 (primary T1D association). X-axis defines genomic intervals local to each gene and y-axis shows the −log10 p-value of association in GWAS (CD, T1D) and eQTL analysis (48 hr α3+28 or 48 hr α3+28/Th17).

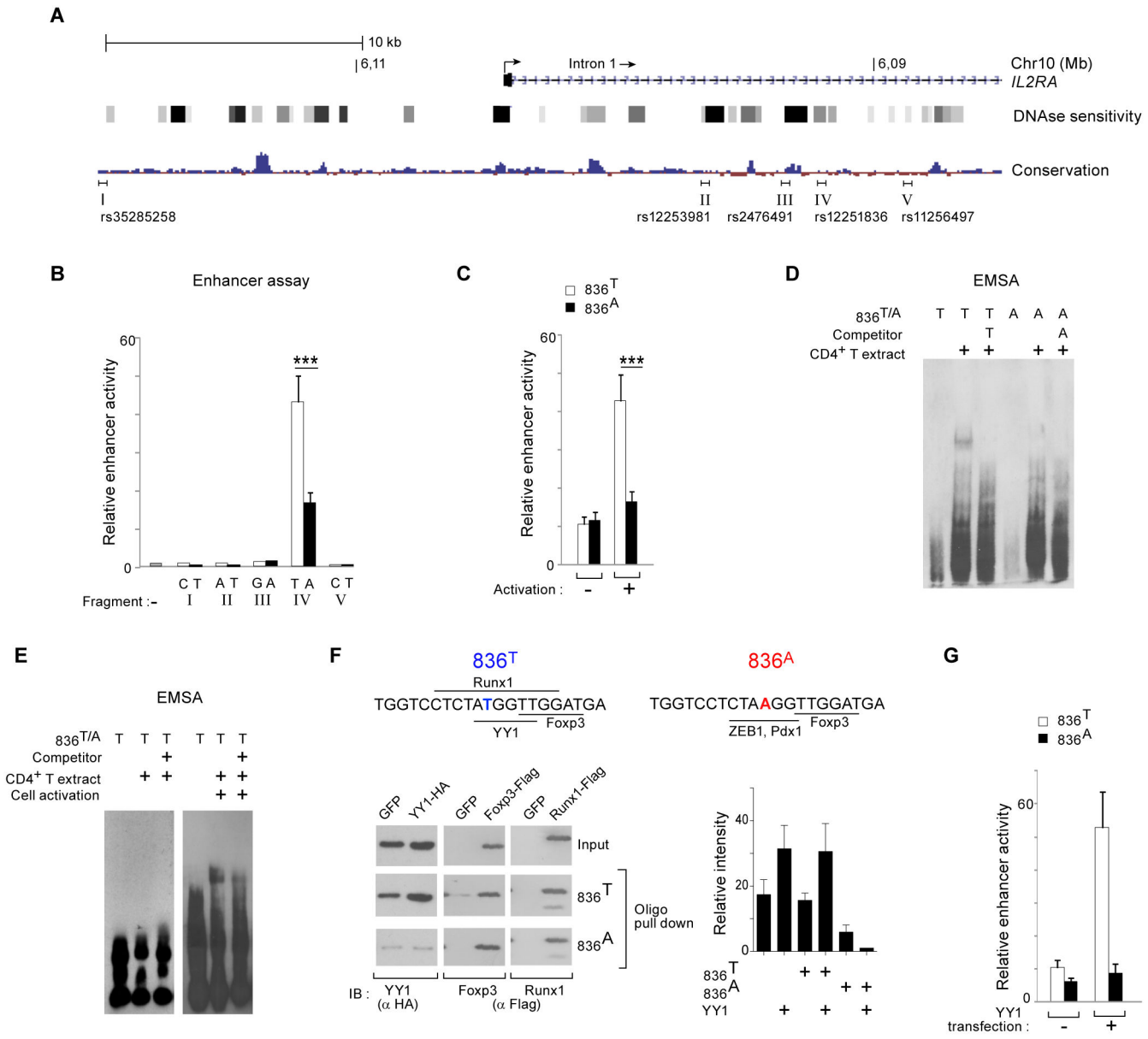

Fig. 6. Validation of rs12251836, a causal variant in the *IL2RA* locus

(**A**) Genomic locations of five top associated SNPs in the *IL2RA* locus, together with mammalian sequence conservation and DNAase hypersensitivity sites (in Th1 cells). The ~200bp fragments centered around each of the variants used in enhancer assays are marked (I–V). (**B**) Enhancer activity in activated Jurkat cells for both alleles (as marked) of the 5 selected fragments cloned into a minimal promoter vector (firefly luciferase, all values normalized to co-transfected Renilla). (**C**) Enhancer activity of fragment IV, as in B, with or without PMA+Ionomycin (P+I) stimulation. (**D**) Protein binding to the two alleles of rs12251836 tested by EMSA with nuclear extract from Jurkat or human primary $CD4^+$ T cells – these gels were run for different times and apparent sizes should not be compared. (**E**) Specific binding of YY1 to the $rs12251836^T$ allele in nuclear extract of primary $CD4^+$ T cells, with or without prior 6 hr activation with α3+28 (**F**) Top: predicted TF-binding sites

around the two alleles of rs12251836. Bottom: pull-down assay of the three main predicted TFs (transfected into HEK293 cells, HA or FLAG tagged) with oligonucleotides encompassing both rs12251836 alleles or an irrelevant control (EBNA), and detected by immunoblotting for the relevant tags. (**G**) Trans-activating activity of over-expressed YY1 in Jurkat cells also transfected with the enhancer reporter plasmids used in B (with either allele at rs12251836). Data are representative of three or more independent experiments; error bars = SD.