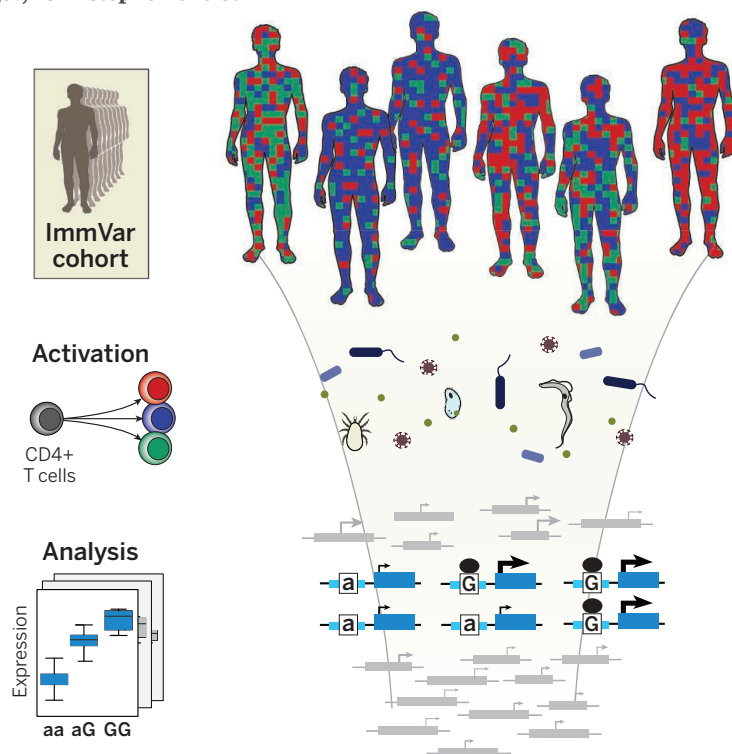IMMUNOGENETICS

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

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

Characterizing the variation in T cell response. A total of 348 individuals from the ImmVar cohort donated blood for CD4+ T cell isolation and genotyping. The T cells were subsequently activated to mimic the recognition of the cognate antigen. The variability in adaptive immune response, which may have genetic and nongenetic components, was characterized by profiling the response of activated T cells. A computational analysis uncovered genetic drivers for this variation and pinpointed molecular mechanisms by which they may act.

INTRODUCTION: The activation of CD4+ T lymphocytes by antigen initiates adaptive immune responses, amplifying rare antigen-specific clones and leading to functional specialization of effector T cells, in particular through the spectrum of cytokines they produce. Pathogens have exerted selective pressure during recent human evolution and migrations, but selection for a type of response that is optimal against one microbe class could carry a price in less-effective responses to other microbes or in impaired self-tolerance and autoimmune disease. Little is known about inter-

individual variation in the responsiveness of human CD4+ T cells and how genetic variation affects the tone and intensity of T cell responses in human individuals.

RATIONALE: As the third arm of the ImmVar project, we performed a rigorously controlled analysis of the responses of human blood CD4+ T cells to activation in unbiased conditions or in a culture regimen that promotes differentiation to the T helper 17 cell (T_H17) phenotype. To permit an appreciation of the breadth of variation across humans, we investigated these responses in

a cohort of 348 healthy human volunteers representing three different ancestries (African, Asian, and European). Responses to activation were evaluated by gene expression profiling, focused on the transcripts that best represent the response, its variability, and its functional consequences. Relating these data to dense single-nucleotide polymorphism genotypes from the participants, we identified the genetic contributions to this variation by using heritability analysis and fine-mapped previously unknown expression quantitative trait loci (eQTLs) that control these responses by using trans-ethnic meta-analysis.

RESULTS: We observed a high degree of interindividual variability, much of which was reproducible for a given subject. This variability followed complex patterns and did not reduce simply to dominant T_H1, 2, or 17 types. We identified 39 loci associated in cis with gene activation in T cells. These explained on average 25% of the repeatable variation, but a major element could not be ascribed to simple genetic effects, instead reflecting environmental influences, immunologic history, or complex integration of network regulation. Of activation-induced genes, cytokines showed the most variability, but with little or no cis genetic control, in contrast to cytokine receptors, which were less variable but for which several eQTLs were detected. Ancestry of the donors markedly influenced T cell responses,

with stronger T_H17 being associated with African descent. We fine-mapped and validated experimentally a single-base variant that modulates binding of the transcription factor YY1 and hence the activity of an enhancer element controlling the autoimmune-associated *IL2RA* gene, affecting its activity in activated but not regulatory T cells.

ON OUR WEB SITE

Read the full article at http://dx.doi .org/10.1126/ science.1254665

CONCLUSION: Echoing the linked ImmVar studies, we find that the relevant cell type and context are essential for discovering the genetic drivers of cell-specific responses and of connected autoinflammatory diseases. Our study lays the groundwork for further explorations into the relative contributions of genes and their environment on immunological processes, which should aid in our understanding of autoimmune disease and its genetic underpinnings. ∎

IMMUNOGENETICS

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 T. Wilson,[2] Natasha Asinovski,[2] Cristin McCabe,[1,3] 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]* Christophe Benoist[1,2]*

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 T helper 17 (T$_H$17) conditions from 348 healthy participants representing European, Asian, and African ancestries. We observed interindividual variability, most marked for cytokine transcripts, with clear biases on the basis of ancestry, and following patterns more complex than simple T$_H$1/2/17 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, interindividual variability affects the fundamental immunologic process of T helper activation, with important connections to autoimmune disease.

A ctivation 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): T helper 1 (T$_H$1) cells producing interferon-γ (IFN-γ) and tumor necrosis factor–α (TNF-α) are most effective against viruses and intracellular microbes; T$_H$2 cells producing interleukin-5 (IL-5) and IL-4/13 are most effective in controlling parasites; T$_H$17 cells and the IL-17 family cytokines partake in antifungal and antibacterial defenses at mucosal interfaces; and follicular T$_H$ 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 T$_H$17 identity through IL-23. These pathways also drive major immunoinflammatory diseases. Pathogenic T$_H$1 or T$_H$17 cells have been implicated in rheumatoid arthritis, multiple sclerosis (MS), and inflammatory bowel disease (IBD), and T$_H$2-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. However, selection for a T$_H$ phenotype, advantageous in one environment, might lead to subpar responses to dominant pathogens in other geographic locales, or carry a penalty in terms of self/nonself 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 interindividual variation in the responsiveness of human CD4+

T cells. Are some individuals inherently more responsive to T cell activation or more prone to mount T$_H$1- or T$_H$17-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). Although 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 interindividual variation

We followed a stepwise approach to dissect the variation in the responses of primary T cells in 348 healthy participants of African (91), Asian (74), or European (183) ancestry from the Boston PhenoGenetic Project, which were also analyzed in other arms of the ImmVar study (17, 18) (Fig. 1A and table S1). By design, we limited the age range (18 to 56 years) to constrain the effects of age on immune function and sampled the participants in a manner that minimizes 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 T$_H$17 phenotype (supplementation with TGF-β, IL-6, IL-23, IL-1B, anti–IL-4, and anti-IFN-γ—hereafter "α3+28/T$_H$17") or with addition of IFN-β. These conditions allowed us to map interindividual variation in T cell transcriptional responses that mimic activation through the antigen-specific receptor in an unbiased setting or when coerced toward the T$_H$17 phenotype. Genome-wide expression profiling established the time points to sample (with cells from a pool of donors) and the extent of interindividual 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 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 Technologies, Seattle, WA) 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 interindividual variation from a subset of eight donors who were sampled twice and distinguished components attributable to cis genetic or to nongenetic 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

[1]Broad Institute of Massachusetts Institute of Technology (MIT) and Harvard, Cambridge, MA 02142, USA. [2]Division of Immunology, Department of Microbiology and Immunobiology, Harvard Medical School, Boston, MA 02115, USA. [3]Program in Translational NeuroPsychiatric Genomics, Institute for the Neurosciences, Departments of Neurology and Psychiatry, Brigham and Women's Hospital, Boston, MA 02115, USA. [4]Department of Medicine Lung Biology Center, University of California, San Francisco, San Francisco, CA 94158, USA. [5]Section of Genetic Medicine, University of Chicago, Chicago, IL 60637, USA. [6]Howard Hughes Medical Institute, Department of Biology, MIT, Cambridge, MA 02139, USA.
*Corresponding author. E-mail: aregev@broad.mit.edu (A.R.); cbdm@hms.harvard.edu (C.B.)
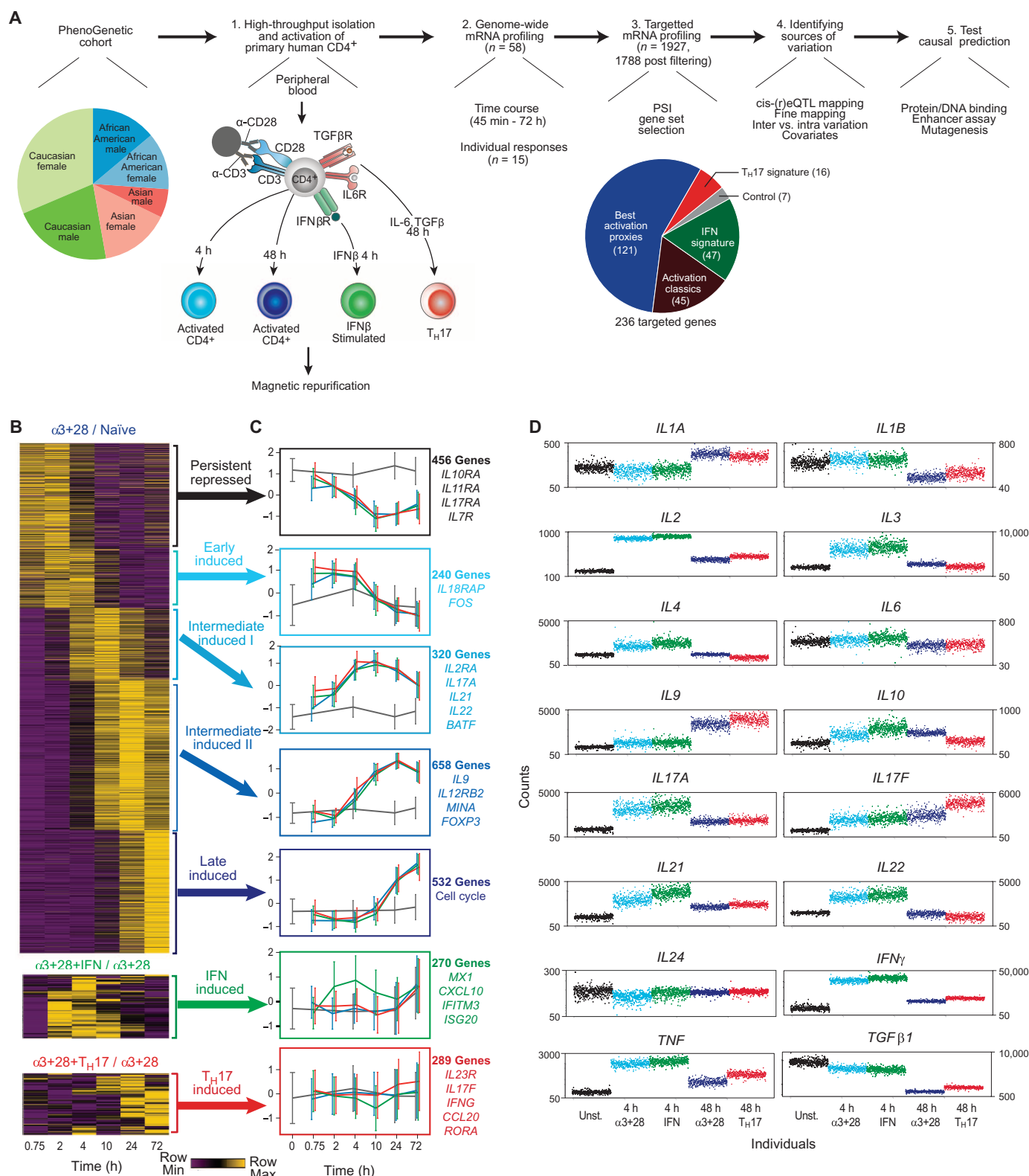
Fig. 1. A systematic approach to characterize the variation in T cell activation and response. (**A**) Overview of study, from (1) cell purification and activation in four conditions, (2) genome-wide or (3) signature expression profiling, (4) computational analysis to decompose expression variance and identify eQTLs, and (5) functional validation of potential causal variants. (**B**) Time course profiles of gene expression in unbiased α3+28 cells (fold change relative to 0-hour baseline; top), and response to IFN-β (fold change relative to unbiased α3+28; middle) or to T_H17 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)].

a single-nucleotide polymorphism (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. 1, B and 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. 1, B and C), affecting 1750 induced and 456 repressed genes (fold change >1.68; table S3), including proliferation and effector genes (*23*, *24*). In T$_H$17-biased conditions, few transcripts (289) were affected that were not already induced in unbiased α3+28 conditions (we did observe the expected overinduction of the *IL17* family and, more surprisingly, of *IFNG*, the prototypical T$_H$1 cytokine). The 270 genes induced in response to IFN-β peaked between 2 and 4 hours, including canonical IFN-responsive targets. From these data, we selected representative time points at 4 and 48 hours for α3+28,

4 hours for α3+28+IFN-β, and 48 hours for α3+28/T$_H$17 activations, and unstimulated controls harvested at 4 hours.

We determined the extent of interindividual 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 (for example, *IL4*, *IL17A*, *IL22*, and *IFNG*) and the temporal disconnect between *IL17A* and *IL17F*]. There was, however, a spectrum in the extent of interindividual variability (Fig. 1D): large for *IL3* and *IL17A* (9.3- and 5.3-fold, respectively) and 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, $P < 1 \times 10^{-7}$), 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 popu-

lation, despite a slight positive correlation between *IL17F* and *IFNG* responses at 48 hours 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, that is, 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 (fold change >1.68, table S4) at 48 hours 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 participant gender or age (fig. S1, B to 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 [principal component 2 (PC2): Pearson $r = -0.52$, $P < 1 \times 10^{-16}$; PC3: Pearson $r = 0.36$, $P < 1 \times 10^{-12}$] with principal components computed over the entire expression matrix (fig. S1, E and F).
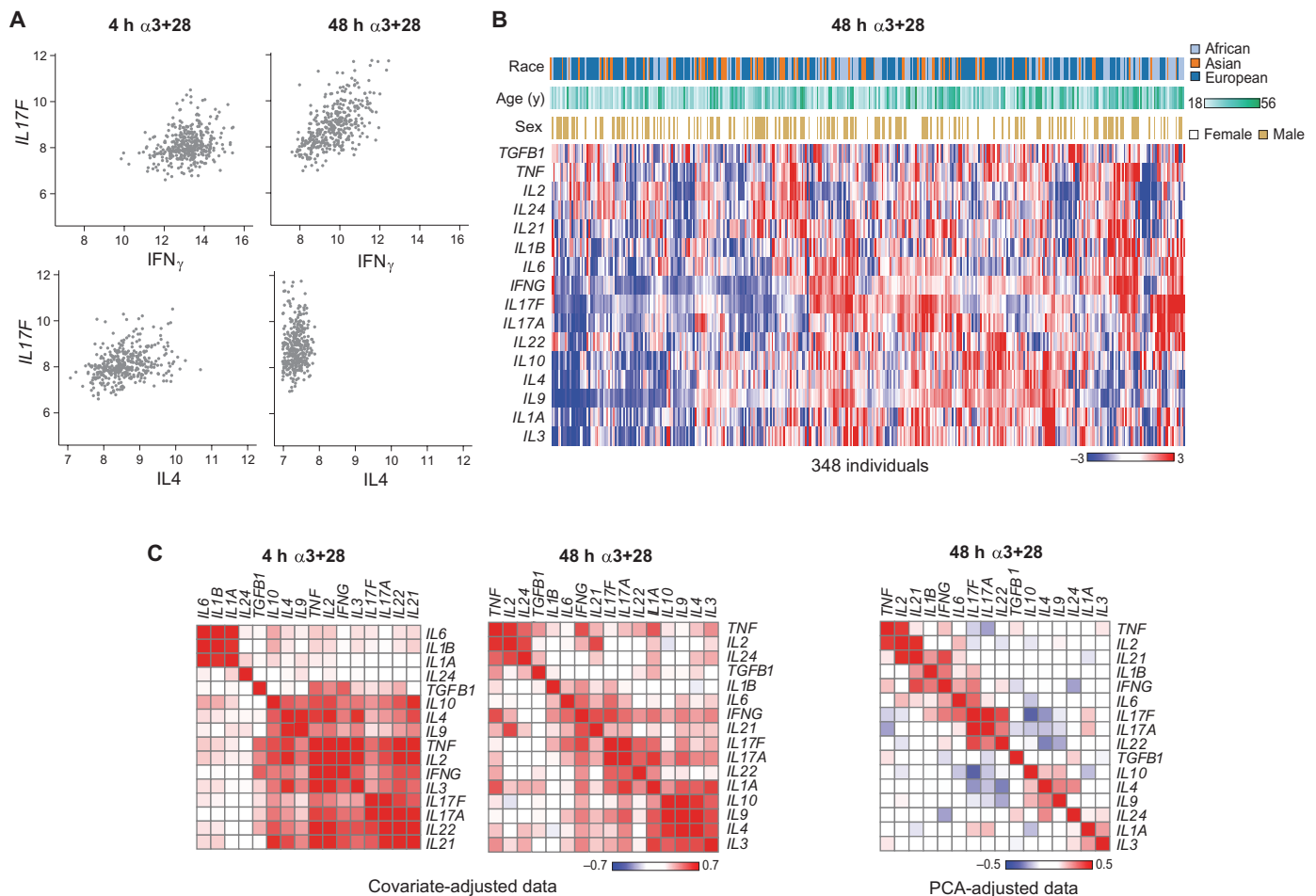
Fig. 2. Interindividual variation in T cell response. (**A**) Relation between expression values (log$_2$ scale) for pairs of T$_H$ defining cytokines (IL-4, IL-17F, and IFN-γ) in unbiased 4-hour α3+28 and 48-hour α3+28 conditions; each point is an individual sample. (**B**) Heatmap of 16 cytokines clustered by expression across 348 individuals in 48-hour 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 principal components analysis (PCA)–adjusted data to account for overall responsiveness (right).

We asked whether some individuals could be distinguished by "response types," skewed toward the cytokine signature of a particular T_H subtype. No simple "T_HX-type" individuals emerged from the expression of signature cytokines: The levels of *IL17F* and *IL4* were largely independent (Pearson $r = 0.07$, $P = 0.16$), whereas the levels of *IL17F* and *IFNG* were positively correlated at 48 hours (Pearson $r = 0.6$, $P < 1 \times 10^{-15}$) (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 T_H1/T_H2 opposition, but this was certainly not the norm. 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 covariation 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 T_H17 cytokines and a well-individualized cluster of inflammatory cytokines (*IL1A/B* and *IL6*; Fig. 2C, left). Clusters became more refined after 48 hours (Fig. 2C, middle), some of which were expected (*IL17A/F* and *IL22*), and others more surprising (for example, *IFNG* with *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 eight 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 hours and *IL17F* expression at 48 hours and between early *IFNG* and later *IFNG* production, indicating that interindividual variation sets the type of response. This early bias was not an amplification of a predetermined state, because 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*, and *STAT2*) predicting later *IL17F* induction, suggesting an interesting connection between tonic IFN-I signals and the ability to mount IL-17 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 intraindividual repeatability with data from eight individuals whose cells were sampled on two different dates and stimulated in the same conditions (Fig. 3A). The average repeatability (*27*) (linear mixed model bootstrap $\bar{R} \pm SE_{\bar{R}}$, table S5, medium fill in Fig. 3A) was highest in late response (45 ± 2% in either 48-hour α3+28 or 48-hour T_H17) and lower in early response (20 ± 3% in 4-hour α3+28). Although there was no general trend between repeatability and the coefficient of variation, 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 (for example, age, gender, height, weight, and blood pressure) was generally low (2.4 to 3.6%, adjusted $\bar{r}^2_{fixed} \pm SE_{\bar{r}^2_{fixed}}$, light fill in Fig. 3, A and B), with sex [10 genes, *t*-test false discovery rate (FDR) < 0.05] and age (53 genes, *F*-test FDR < 0.05) the only covariates significantly correlated with gene expression in 48-hour α3+28 (table S6).

The average proportion of variance explained by cis heritability (*28*, *29*) (permutation $\bar{h}^2_{cis} \pm SE_{\bar{h}^2_{cis}}$, dark fill in Fig. 3A) was higher in 48-hour activated conditions (9.2 ± 0.9%) than in unstimulated T cells (5.8%) and 4-hour activated conditions (~6.6%). Whereas the latter are consistent with previous estimates in whole tissue with identity by descent (*28*), the higher 48-hour estimates may reflect increased power to detect activation-specific eQTLs.

Repeatable variance attributable to common cis genetic and nongenetic factors varied widely between genes (Fig. 3, A and B). Cytokine transcripts were repeatable in their expression upon activation for the eight individuals tested, but common genetic variation did not contribute substantially to their variability across the full data set 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 may be 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 of 229 (41%) genes in the two 48-hour 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 stronger trend of overexpression for donors of African ancestry, lower for European ancestry, and a mixed pattern for Asian ancestry (Fig. 3C). Differentially responsive genes include key indicators of T_H phenotypes, *IL17* family cytokines (overinduced in

individuals of African ancestry), and *IFNG*, which showed an opposite pattern. We are cognizant that such differences can be artifactually generated by systematic bias in sampling, batch, or lifestyle confounders, but the differences observed likely have some genetic underpinning because 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 (ancestry explaining a median of 2.4 and 2.7% of expression variation in the 48-hour conditions; Fig. 3D and fig. 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-hour stimulation, indicating that the population differentiation required activation to be manifest.

In addition to genes with known population differentiation resulting from population-specific genetic polymorphisms [*UTS2* and *IFITM3* (*18*, *34*)], we detected population differentiation in transcripts encoding cytokines, chemokines, or their receptors (table S6 and Fig. 3C). Notably, *IL2RA* was one of the most statistically differentially expressed genes (*t* test, $P < 1 \times 10^{-16}$ in both 48-hour 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 T_H17 polarization

To identify the genetic variants underpinning expression differences, we associated ~10 million imputed genotypes [1000 Genomes Project (*35*), minor allele frequency (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 and table S7) had a significant cis-eQTL within 1 Mb 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 T_H17 polarizing or IFN-β conditions beyond those detected in unbiased activation alone, although a few eQTLs (eight in T_H17, five in IFN-β) had larger effect sizes (table S7). Some of the condition-specific cis-eQTLs were related to transcript levels (Fig. 4, C and D), but this was not the case for many (Fig. 4, E and F)—*IL2RA*

was expressed before stimulation, but it only had a significant cis-eQTL in the late activation conditions (rs12251836, meta $P < 1 \times 10^{-16}$). 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 and rs2863204) in the α3+28/T_H17 condition with the change in expression of IL21 between 48-hour α3+28/T_H17 and unbiased stimulation (table S10), explaining 11 ± 8.7% of the variance (permutation FDR <0.2; mediation
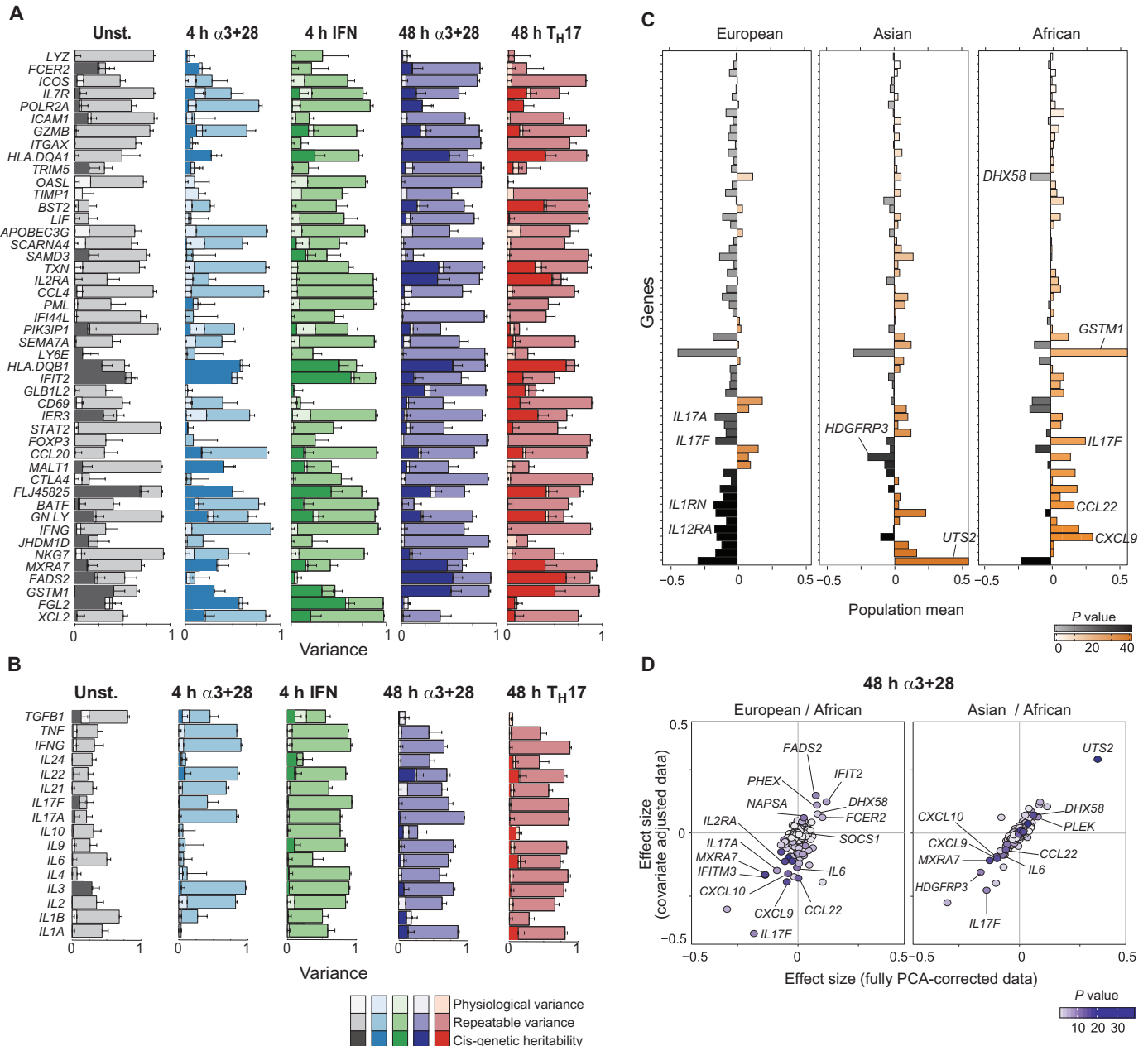

Fig. 3. Sources of variation across T cell activation conditions. (**A** and **B**) Top 56 non-cytokine genes (A) or top 16 key cytokines (B) ranked by maximum repeatability across all conditions estimated from eight recalled individuals. For each gene, repeatability (medium shading) and contributions of cis heritability (dark shading) and of physiological covariates [age, gender, body mass index (BMI), 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 1 Mb 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, and blood pressure estimated from multiple regression. Error bars indicate bootstrap SEs around each estimate. (**C**) Percent difference of average population expression (median) from overall average (median) that shows 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-hour 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.
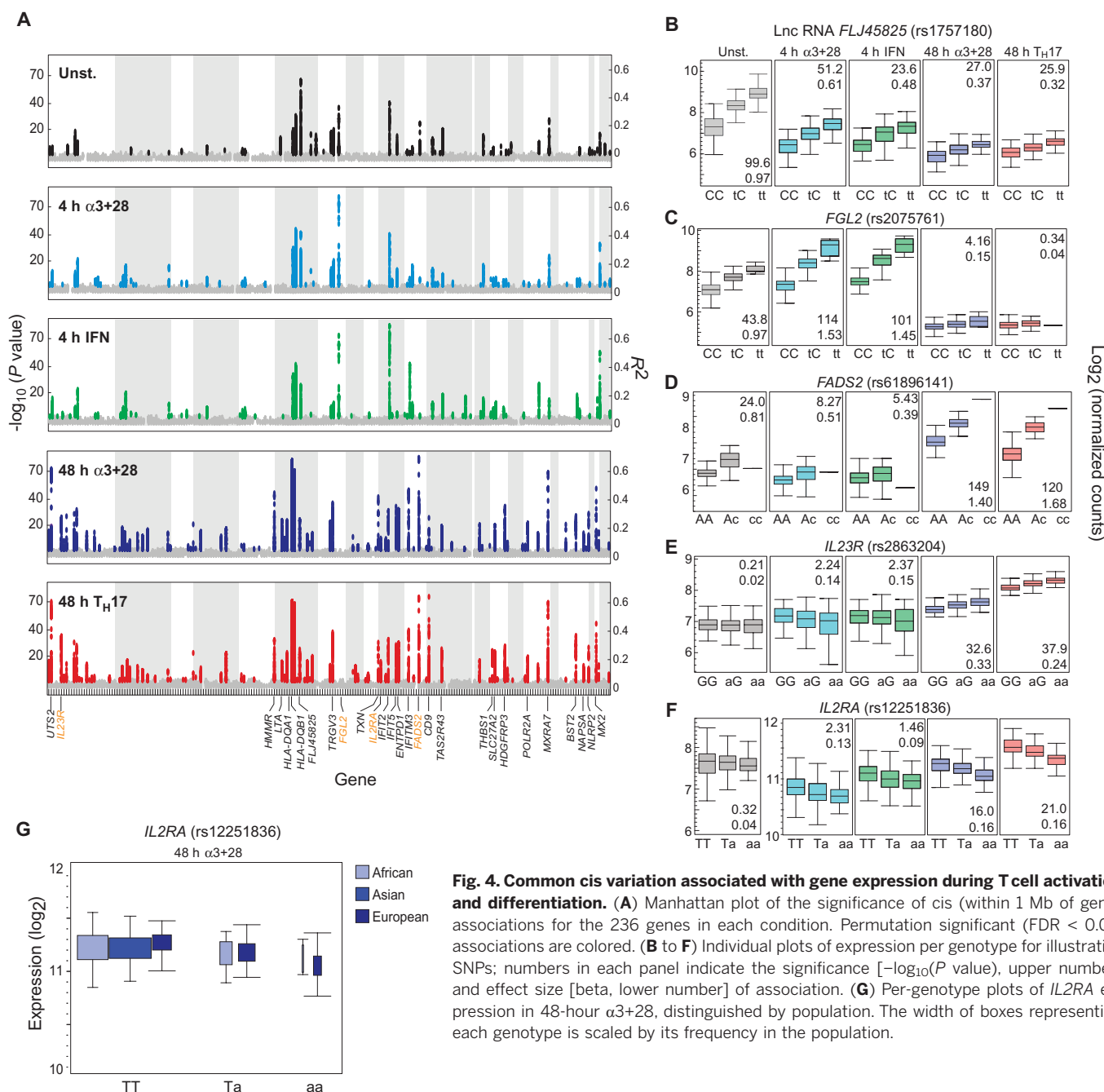
Fig. 4. Common cis variation associated with gene expression during T cell activation and differentiation. (**A**) Manhattan plot of the significance of cis (within 1 Mb of gene) associations for the 236 genes in each condition. Permutation significant (FDR < 0.01) associations are colored. (**B** to **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-hour α3+28, distinguished by population. The width of boxes representing each genotype is scaled by its frequency in the population.

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 IL-21 production by IL-23 signals during T$_H$17 differentiation (*36, 37*).

Overlap of genetically variable T cell responses with genome-wide association study loci identifies potential disease-relevant variants

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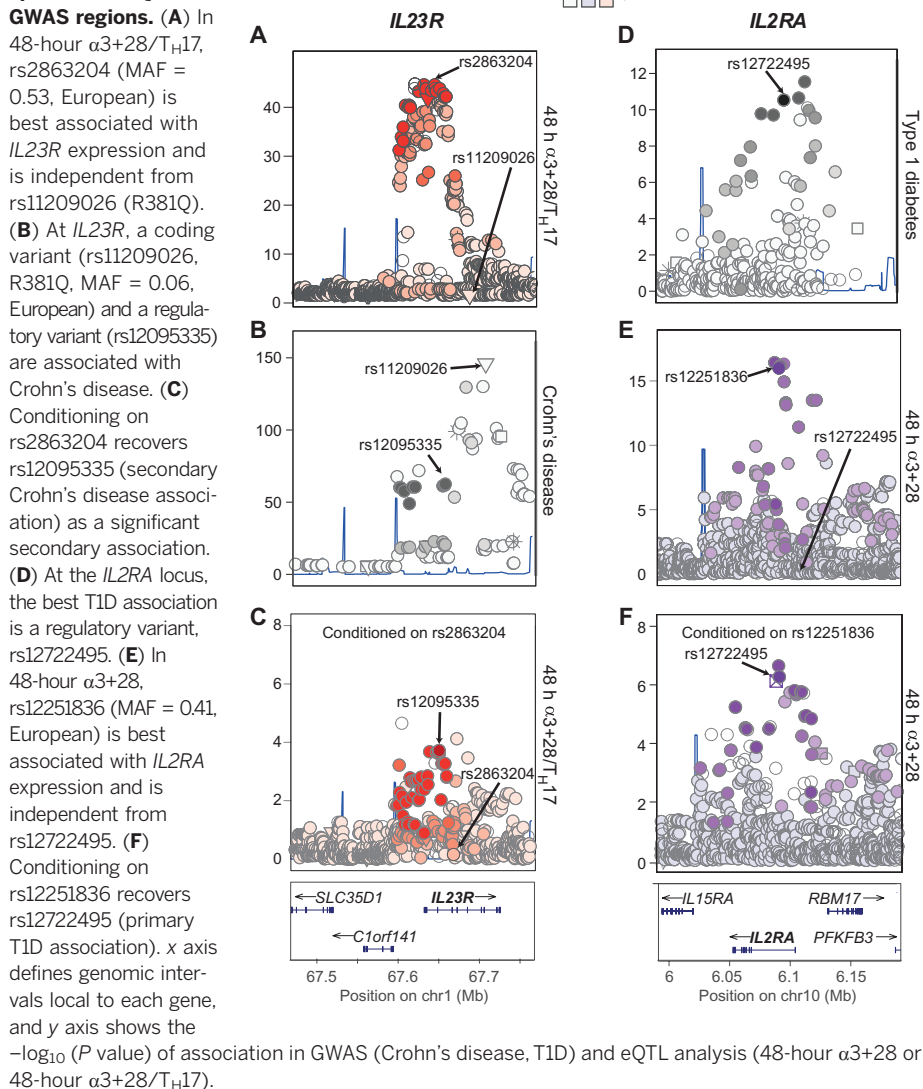ly near $IL23R$ and $IL2RA$ (Fig. 5 and table S11), two key cytokine receptors implicated in a number of autoimmune diseases.

$IL23R$ encodes the receptor for IL-23, partakes in establishing stable T$_H$17 differentiation (*38*), and is strongly associated with IBD (*39, 40*). Here, rs2863204 was the variant most significantly associated with $IL23R$ expression in 48-hour α3+28/ T$_H$17 (Fig. 5A). It is not on the same haplotype as the IBD-associated nonsynonymous variant (rs11209026, R381Q) (*40*). However, meta-analysis of ulcerative colitis and Crohn's disease separately (*40*) revealed allelic heterogeneity in $IL23R$ where an additional variant, tagged by rs12095335, is associated with Crohn's disease but not ulcerative colitis (Fig. 5B). When conditioned on rs2863204,

rs12095335 appeared as a secondary association to $IL23R$ mRNA levels (meta P = 2 × 10⁻⁴; Fig. 5C), suggesting that there may be at least two independent effects on $IL23R$ expression in the region. The trans-association between these $IL23R$ variants and $IL21$ expression in the T$_H$17 condition further implies functional consequences for the region with regard to T$_H$17 differentiation.

Perhaps most intriguing was the discovery of SNPs associated with $IL2RA$ expression. $IL2RA$ encodes CD25, the high-affinity IL-2 receptor, and has intricately balanced functions, reflecting the divergent roles of IL-2 as a trophic factor for inhibitory T regulatory (T$_{reg}$) cells and in aiding the expansion of proinflammatory effector T cells

Fig. 5. Trans-ancestry fine mapping identifies activation-specific cis-eQTLs in GWAS regions. (**A**) In 48-hour α3+28/T$_H$17, rs2863204 (MAF = 0.53, European) is best associated with *IL23R* expression and is independent from rs11209026 (R381Q). (**B**) At *IL23R*, a coding variant (rs11209026, R381Q, MAF = 0.06, European) and a regulatory variant (rs12095335) are associated with Crohn's disease. (**C**) Conditioning on rs2863204 recovers rs12095335 (secondary Crohn's disease association) as a significant secondary association. (**D**) At the *IL2RA* locus, the best T1D association is a regulatory variant, rs12722495. (**E**) In 48-hour α3+28, rs12251836 (MAF = 0.41, European) is best associated with *IL2RA* expression and is independent from rs12722495. (**F**) Conditioning on rs12251836 recovers rs12722495 (primary T1D association). *x* axis defines genomic intervals local to each gene, and *y* axis shows the −log$_{10}$ (*P* value) of association in GWAS (Crohn's disease, T1D) and eQTL analysis (48-hour α3+28 or 48-hour α3+28/T$_H$17).

or natural killer 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 with titers of soluble IL-2 receptor, expression of CD25 at the cell surface, or IL-2 signaling efficacy (*12, 43, 44*). Three independent haplotypes have been reported to be associated with T1D. The strongest, a protective haplotype tagged by rs12722495 (Fig. 5D), leads to higher expression of *IL2RA* in steady-state CD45RA$^-$ memory and T$_{reg}$ cells (*43–45*). Here, rs12251836, the most significant variant associated with *IL2RA* expression in 48-hour 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* = 5.3 × 10^{-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 T$_{reg}$ cells, because *IL2RA* mRNA in unstimulated CD4$^+$ T cells would predominantly come from CD25^{hi} T$_{reg}$ 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 T$_{reg}$ cells. On balance, these two variants would then control the effectiveness of IL-2 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 inves-

tigated the functional effects of the top five candidate SNPs located in the 5′ flanking and first intron of *IL2RA* (Fig. 6A). We first tested for enhancer activity by introducing ~200–base pair (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 [+/− phorbol 12-myristate 13-acetate (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 than 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* < 1 × 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 electrophoretic mobility shift assay (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 T$_H$ 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
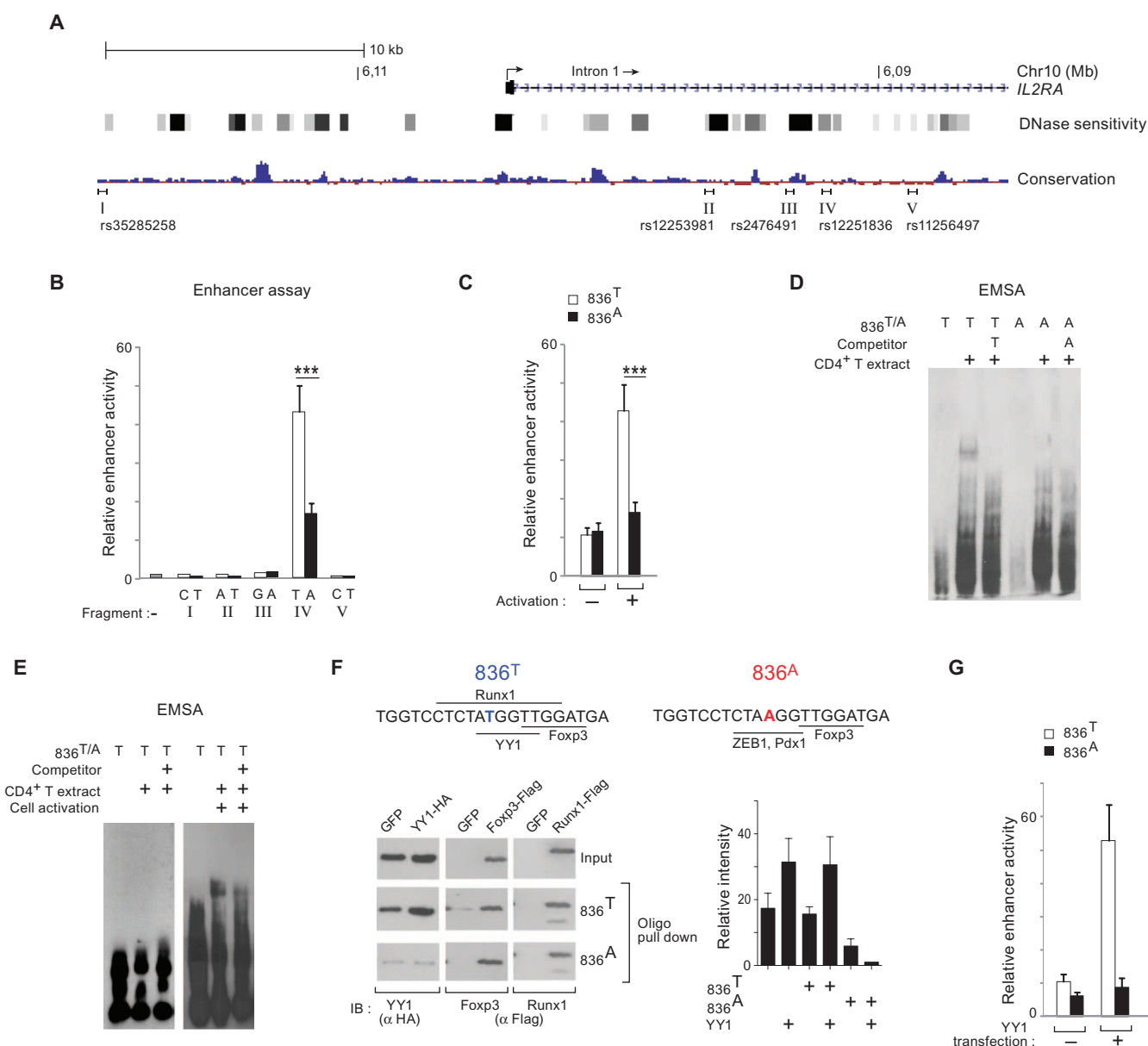
Fig. 6. Validation of rs12251836, a causal variant in the *IL2RA* locus. (**A**) Genomic locations of five top associated SNPs in the *IL2RA* locus, together with mammalian sequence conservation and deoxyribonuclease (DNase) hypersensitivity sites (in T_H1 cells). The ~200-bp fragments centered around each of the variants used in enhancer assays are marked (I to V). (**B**) Enhancer activity in activated Jurkat cells for both alleles (as marked) of the five selected fragments cloned into a minimal promoter vector (firefly luciferase, all values normalized to cotransfected *Renilla*). (**C**) Enhancer activity of fragment IV, as in (B), with or without PMA + ionomycin 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-hour 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 human embryonic kidney 293 cells, hemagglutinin (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 overexpressed 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 indicate SD.

were profiled with respect to fresh immunocytes at baseline (*17*) and to activation responses in dendritic cells (*18*). Sixty-nine 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* and *IL2RA*). Of the eQTLs for IFN-responsive genes, 11

of 15 overlapped with those detected in IFN-treated dendritic cells (*18*), reflecting the sharing of IFN-β 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 T_H–biased responses? Of the changes elicited by T cell

triggering, cytokine transcripts showed the strongest change in interindividual variance. This diversification did not generally track along simple T_H1/2/17 models but adopted more-complex patterns. This observation is of importance 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. 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, whereas the cytokines themselves may be under more uniform genetic control but vary with environmental cues or the individual's immunological history. In addition, the interindividual variance exhibited a general "responsiveness": a widespread correlation between all cytokines, fine distinctions between clusters of co-regulated cytokines only becoming 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 toward 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 IL-17 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 of 73 in 48-hour α3+28/T_H17) 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 but rare in African and Asian descendants. The lower responses to activation and the overrepresentation of the corresponding alleles in European-derived populations suggest that tolerance mechanisms to avoid immunopathology resulting from exuberant T cell responses may also have led to the evolution of protection from autoimmunity.

REFERENCES AND NOTES

1. J. Zhu, H. Yamane, W. E. Paul, Differentiation of effector CD4+ T cell populations. *Annu. Rev. Immunol.* **28**, 445–489 (2010). doi: 10.1146/annurev-immunol-030409-101212; pmid: 20192806

2. M. Fumagalli *et al.*, Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. *PLOS Genet.* **7**, e1002355 (2011). doi: 10.1371/journal.pgen.1002355; pmid: 22072984

3. R. M. Zinkernagel, H. Hengartner, T-cell–mediated immunopathology versus direct cytolysis by virus: Implications for HIV and AIDS. *Immunol. Today* **15**, 262–268 (1994). doi: 10.1016/0167-5699(94)90005-1; pmid: 7915115

4. R. Medzhitov, D. S. Schneider, M. P. Soares, Disease tolerance as a defense strategy. *Science* **335**, 936–941 (2012). doi: 10.1126/science.1214935; pmid: 22363001

5. K. A. Siminovitch, *PTPN22* and autoimmune disease. *Nat. Genet.* **36**, 1248–1249 (2004). doi: 10.1038/ng1204-1248; pmid: 15565104

6. D. M. Evans *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.* **43**, 761–767 (2011). doi: 10.1038/ng.873; pmid: 21743469

7. Genetic Analysis of Psoriasis Consortium & the Wellcome Trust Case Control Consortium 2 *et al.*, A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between *HLA-C* and *ERAP1*. *Nat. Genet.* **42**, 985–990 (2010). doi: 10.1038/ng.694; pmid: 20953190

8. K. Wang, M. Li, H. Hakonarson, Analysing biological pathways in genome-wide association studies. *Nat. Rev. Genet.* **11**, 843–854 (2010). doi: 10.1007/s10709-010-9467-7; pmid: 21085203

9. C. E. Lowe *et al.*, Large-scale genetic fine mapping and genotype-phenotype associations implicate polymorphism in the *IL2RA* region in type 1 diabetes. *Nat. Genet.* **39**, 1074–1082 (2007). doi: 10.1038/ng2102; pmid: 17676041

10. J. C. Barrett *et al.*, Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat. Genet.* **41**, 703–707 (2009). doi: 10.1038/ng.381; pmid: 19430480

11. J. D. Cooper *et al.*, Meta-analysis of genome-wide association study data identifies additional type 1 diabetes risk loci. *Nat. Genet.* **40**, 1399–1401 (2008). doi: 10.1038/ng.249; pmid: 18978792

12. L. M. Maier *et al.*, *IL2RA* genetic heterogeneity in multiple sclerosis and type 1 diabetes susceptibility and soluble interleukin-2 receptor production. *PLOS Genet.* **5**, e1000322 (2009). doi: 10.1371/journal.pgen.1000322; pmid: 19119414

13. International Multiple Sclerosis Genetics Consortium (IMSGC) *et al.*, Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nat. Genet.* **45**, 1353–1360 (2013). doi: 10.1371/journal.pgen.1000322; pmid: 24076602

14. International Multiple Sclerosis Genetics Consortium *et al.*, Risk alleles for multiple sclerosis identified by a genomewide study. *N. Engl. J. Med.* **357**, 851–862 (2007). pmid: 17660530

15. S. B. Montgomery, E. T. Dermitzakis, From expression QTLs to personalized transcriptomics. *Nat. Rev. Genet.* **12**, 277–282 (2011). doi: 10.1038/nrg2969; pmid: 21386863

16. Y. Gilad, S. A. Rifkin, J. K. Pritchard, Revealing the architecture of gene regulation: The promise of eQTL studies. *Trends Genet.* **24**, 408–415 (2008). doi: 10.1016/j.tig.2008.06.001; pmid: 18597885

17. T. Raj *et al.*, Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes. *Science* **344**, 519–523 (2014). doi: 10.1126/science.1249547; pmid: 24786080

18. M. N. Lee *et al.*, Common genetic variants modulate pathogen-sensing responses in human dendritic cells. *Science* **343**, 1246980 (2014). doi: 10.1126/science.1246980; pmid: 24604203

19. Y. Donner, T. Feng, C. Benoist, D. Koller, Imputing gene expression from selectively reduced probe sets. *Nat. Methods* **9**, 1120–1125 (2012). doi: 10.1038/nmeth.2207; pmid: 23064520

20. N. Zaitlen, B. Paşaniuc, T. Gur, E. Ziv, E. Halperin, Leveraging genetic variability across populations for the identification of causal variants. *Am. J. Hum. Genet.* **86**, 23–33 (2010). doi: 10.1016/j.ajhg.2009.11.016; pmid: 20085711

21. L. Li, J. Nishio, A. van Maurik, D. Mathis, C. Benoist, Differential response of regulatory and conventional CD4+ lymphocytes to CD3 engagement: Clues to a possible mechanism of anti-CD3 action? *J. Immunol.* **191**, 3694–3704 (2013). doi: 10.4049/jimmunol.1300408; pmid: 23986534

22. J. A. Best *et al.*, Transcriptional insights into the CD8+ T cell response to infection and memory T cell formation. *Nat. Immunol.* **14**, 404–412 (2013). doi: 10.1038/ni.2536; pmid: 23396170

23. S. Feske, J. Giltnane, R. Dolmetsch, L. M. Staudt, A. Rao, Gene regulation mediated by calcium signals in T lymphocytes. *Nat. Immunol.* **2**, 316–324 (2001). doi: 10.1038/86318; pmid: 11276202

24. K. Hess *et al.*, Kinetic assessment of general gene expression changes during human naive CD4+ T cell activation. *Int. Immunol.* **16**, 1711–1721 (2004). doi: 10.1093/intimm/dxh172; pmid: 15492022

25. M. K. Jenkins, J. J. Moon, The role of naive T cell precursor frequency and recruitment in dictating immune response magnitude. *J. Immunol.* **188**, 4135–4140 (2012). doi: 10.4049/jimmunol.1102661; pmid: 22517866

26. M. M. Delmastro *et al.*, Modulation of redox balance leaves murine diabetogenic TH1 T cells "LAG-3-ing" behind. *Diabetes* **61**, 1760–1768 (2012). doi: 10.2337/db11-1591; pmid: 22586584

27. D. S. Falconer, T. F. C. Mackay, *Introduction to Quantitative Genetics* (Longman, Essex, England, ed. 4, 1996).

28. A. L. Price *et al.*, Single-tissue and cross-tissue heritability of gene expression via identity-by-descent in related or unrelated individuals. *PLOS Genet.* **7**, e1001317 (2011). doi: 10.1371/journal.pgen.1001317; pmid: 21383966

29. J. Yang, S. H. Lee, M. E. Goddard, P. M. Visscher, GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011). doi: 10.1016/j.ajhg.2010.11.01; pmid: 21167468

30. J. D. Storey *et al.*, Gene-expression variation within and among human populations. *Am. J. Hum. Genet.* **80**, 502–509 (2007). doi: 10.1086/512017; pmid: 17273971

31. B. E. Stranger *et al.*, Patterns of *cis* regulatory variation in diverse human populations. *PLOS Genet.* **8**, e1002639 (2012). doi: 10.1371/journal.pgen.1002639; pmid: 22532805

32. N. A. Rosenberg *et al.*, Genetic structure of human populations. *Science* **298**, 2381–2385 (2002). doi: 10.1126/science.1078311; pmid: 12493913

33. T. Lappalainen *et al.*, Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* **501**, 506–511 (2013). doi: 10.1038/nature12531; pmid: 24037378

34. B. E. Stranger *et al.*, Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* **315**, 848–853 (2007). doi: 10.1126/science.1136678; pmid: 17289997

35. 1000 Genomes Project Consortium *et al.*, An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012). doi: 10.1038/nature11632; pmid: 23128226

36. S. Aggarwal, N. Ghilardi, M. H. Xie, F. J. de Sauvage, A. L. Gurney, Interleukin-23 promotes a distinct CD4 T cell activation state characterized by the production of interleukin-17. *J. Biol. Chem.* **278**, 1910–1914 (2003). doi: 10.1074/jbc.M207577200; pmid: 12417590

37. L. Zhou *et al.*, IL-6 programs T_H-17 cell differentiation by promoting sequential engagement of the IL-21 and IL-23 pathways. *Nat. Immunol.* **8**, 967–974 (2007). doi: 10.1038/ni1488; pmid: 17581537

38. T. Korn, E. Bettelli, M. Oukka, V. K. Kuchroo, IL-17 and Th17 cells. *Annu. Rev. Immunol.* **27**, 485–517 (2009). doi: 10.1146/annurev.immunol.021908.132710; pmid: 19132915

39. R. H. Duerr *et al.*, A genome-wide association study identifies *IL23R* as an inflammatory bowel disease gene. *Science* **314**, 1461–1463 (2006). doi: 10.1126/science.1135245; pmid: 17068223

40. L. Jostins *et al.*, Host–microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491**, 119–124 (2012). doi: 10.1038/nature11582; pmid: 23128233

41. O. Boyman, J. Sprent, The role of interleukin-2 during homeostasis and activation of the immune system. *Nat. Rev. Immunol.* **12**, 180–190 (2012). doi: 10.1038/nri3156; pmid: 22343569

42. D. J. Smyth *et al.*, Shared and distinct genetic variants in type 1 diabetes and celiac disease. *N. Engl. J. Med.* **359**, 2767–2777 (2008). doi: 10.1056/NEJMoa0807917; pmid: 19073967

43. C. A. Dendrou *et al.*, Cell-specific protein phenotypes for the autoimmune locus *IL2RA* using a genotype-selectable human bioresource. *Nat. Genet.* **41**, 1011–1015 (2009). doi: 10.1038/ng.434; pmid: 19701192

44. K. Cerosaletti *et al.*, Multiple autoimmune-associated variants confer decreased IL-2R signaling in CD4+CD25hi T cells of type 1 diabetic and multiple sclerosis patients. *PLOS One* **8**, e83811 (2013). doi: 10.1371/journal.pone.0083811; pmid: 24376757

45. V. Orrù *et al.*, Genetic variants regulating immune cell levels in health and disease. *Cell* **155**, 242–256 (2013). doi: 10.1016/j.cell.2013.08.041; pmid: 24074872

46. M. Ono *et al.*, Foxp3 controls regulatory T-cell function by interacting with AML1/Runx1. *Nature* **446**, 685–689 (2007). doi: 10.1038/nature05673; pmid: 17377532

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/345/6202/1254665/suppl/DC1
Materials and Methods
Figs. S1 to S8
Tables S1 to S11
References (*47–55*)