Natural variation in the parameters of innate immune cells is preferentially driven by genetic factors

Etienne Patin, Milena Hasan, Jacob Bergstedt, Vincent Rouilly, Valentina Libri, Alejandra Urrutia, Cécile Alanio, Petar Scepanovic, Christian Hammer, Friederike Jönsson, et al.

Natural variation in innate immune cell parameters

is preferentially driven by genetic factors

Etienne Patin[1-3,26,*], Milena Hasan[4,26], Jacob Bergstedt[5,6,26], Vincent Rouilly[3,4], Valentina Libri[4], Alejandra Urrutia[4,7-9], Cécile Alanio[4,7,8], Petar Scepanovic[10,11], Christian Hammer[10,11], Friederike Jönsson[12,13], Benoît Beitz[4], Hélène Quach[1-3], Yoong Wearn Lim[9], Julie Hunkapiller[14], Magge Zepeda[15], Cherie Green[16], Barbara Piasecka[1-4], Claire Leloup[14], Lars Rogge[4,17], François Huetz[18,19], Isabelle Peguillet[20-22], Olivier Lantz[20-23], Magnus Fontes[6,24], James P. Di Santo[4,8,25], Stéphanie Thomas[4,7,8], Jacques Fellay[9,10], Darragh Duffy[4,7,8], Lluís Quintana-Murci[1-3,27], Matthew L. Albert[4,7-9,27,*], for The Milieu Intérieur Consortium

[1]Unit of Human Evolutionary Genetics, Department of Genomes & Genetics, Institut Pasteur, Paris 75015, France. [2]CNRS URA3012, Paris 75015, France. [3]Center of Bioinformatics, Biostatistics and Integrative Biology, Institut Pasteur, Paris 75015, France. [4]Center for Translational Science, Institut Pasteur, Paris 75015, France. [5]Department of Automatic Control, Lund University, Lund SE-221, Sweden. [6]International Group for Data Analysis, Institut Pasteur, Paris 75015, France. [7]Laboratory of Dendritic Cell Immunobiology, Department of Immunology, Institut Pasteur, Paris 75015, France. [8]INSERM U1223, France. [9]Department of Cancer Immunology, Genentech, South San Francisco, California 94080, USA. [10]School of Life Sciences, École Polytechnique Fédérale de Lausanne, Lausanne 1015, Switzerland. [11]Swiss Institute of Bioinformatics, Lausanne 1015, Switzerland. [12]Antibodies in Therapy and Pathology, Department of Immunology, Institut Pasteur, Paris 75015, France. [13]INSERM U1222, France. [14]Department of Human Genetics, Genentech, South San Francisco, California 94080, USA. [15]Employee Donation Program, Genentech, South San Francisco, California 94080, USA. [16]Department of Development Sciences, Genentech, South San Francisco, California 94080, USA. [17]Immunoregulation Unit, Department of Immunology, Institut Pasteur, Paris 75015, France. [18]INSERM U783, Faculté de Médecine, Site Necker-Enfants Malades, Université Paris Descartes, Paris 75015, France. [19]Lymphocyte Population Biology, CNRS URA 1961, Institut Pasteur, Paris 75015, France. [20]Center of Clinical Investigations CIC-BT1428 IGR/Curie, Paris 75005,

France. [21]Equipe Labellisée de la Ligue de Lutte Contre le Cancer, Institut Curie, Paris 75005, France. [22]Department of Biopathology, Institut Curie, Paris 75005, France. [23]INSERM/Institut Curie U932, France. [24]Centre for Mathematical Sciences, Lund University, Lund SE-221, Sweden. [25]Innate Immunity Unit, Institut Pasteur, Paris 75015. [26]These authors contributed equally to this work. [27]These authors jointly directed this work.

*Correspondence should be addressed to: M.L.A. (albertm7@gene.com), E.P. (epatin@pasteur.fr)

Abstract

The enumeration and characterization of circulating immune cells provide key indicators of human health and disease. To identify the relative impact that environmental and genetic factors have on variation of innate and adaptive immune cell parameters in homeostatic conditions, we combined standardized flow cytometric analysis of blood leukocytes and genome-wide DNA genotyping in 1,000 healthy, unrelated individuals of western European ancestry. We show that smoking, together with age, sex and latent cytomegalovirus infection, are the main non-genetic factors affecting human variation in immune cell parameters. Genome-wide association studies of 166 immunophenotypes identified 15 loci that are enriched in disease-associated variants. Finally, we demonstrate that innate cell parameters are more strongly controlled by genetic variation than adaptive cell parameters, which are primarily driven by environmental exposures. Our data establish a resource that generates new hypotheses in immunology and highlight the role of innate immunity in the susceptibility to common autoimmune diseases.

Introduction

The immune system plays an essential role in maintaining homeostasis in individuals challenged by microbial infections, a physiological mechanism conceptualized by the French physician Claude Bernard in 1865, when he defined the notion of "*milieu intérieur*"[1]. Host-pathogen interactions trigger immune responses through the activation of specialized immune cell populations, which may eventually result in pathogen clearance. The study of immune cell populations circulating in the blood provides a view into innate cells that are transiting between the bone marrow and tissues, and adaptive cells that are recirculating through lymphoid organs. Clinical studies of patients with past or chronic latent infections have reported profound perturbations of subsets of circulating immune cells due to altered trafficking, selective expansion or attrition[2,3]. However, several studies have suggested that extensive differences in white blood cell composition also exist among healthy individuals[4,5]. The evaluation of the naturally occurring variation of immune cell parameters, together with its environmental and genetic determinants, could accelerate hypothesis generation in basic immunology, and ultimately improve the characterization of pathological states.

Population immunology approaches, which compare the immune status across a large number of healthy individuals, have highlighted the predominant effect of intrinsic factors such as age and sex on human blood cell composition[6]. Several activated and memory T cell subpopulations increase with age[7], which may partially result from diminished thymic activity[8] and explain reduced vaccination efficacy in the elderly[9]. Seasonal fluctuations in B cells, regulatory T (T_{reg}) cells and monocytes[10] and a strong effect of cohabitation on human immune profiles[11] have been observed, suggesting that environmental exposures also drive immune variation. For instance, latent cytomegalovirus (CMV) infection, detected in 40% to >90% of the general population[12], has been associated with an increased number of effector memory T cells[13], which could in turn alter immune responses to heterologous infection[14]. However, the respective impact of age, sex and CMV infection on both innate and adaptive cells, as well as the precise nature of the environmental factors affecting immune variation, are largely unknown.

Recent technological advances in flow cytometry, combined with genome-wide DNA genotyping, now allow the dissection of the genetic basis of interindividual variation in immune cell parameters. A

seminal genome-wide association study identified 13 genetic loci strongly associated with the proportion of different leukocyte subpopulations, in a cohort of 249 Sardinian families[15]. Another study reported the deep immunophenotyping of ~1,800 independent traits in 245 healthy twin pairs, identifying 11 independent genetic loci that accounted for up to 36% of the variation of 19 different traits[16]. A third study estimated the genetic heritability of 95 different immune cell frequencies in 105 healthy twin pairs, and suggested that variation in immune cells is largely explained by non-heritable factors[17]. Finally, four novel loci were associated to B and T cell traits in a cohort of 442 healthy donors, in a study that dissected both non-genetic and genetic factors affecting immune cell traits mediating adaptive immunity[10]. Together, these studies provided valuable insights into the contribution of genetic factors to inter-individual differences in adaptive immune cell populations, but largely neglected several major innate cell types in circulation. An integrated evaluation of the nature and respective impact of intrinsic, environmental and genetic factors driving human variation in both innate and adaptive immunity is thus lacking.

Here, we report the use of standardized flow cytometry to comprehensively establish the white blood cell composition of 1,000 healthy, unrelated individuals of western European ancestry, which compose the Milieu Intérieur cohort. We confirm with this broad resource that age, sex, CMV seropositivity and smoking have major, independent effects on innate and adaptive immune cell parameters. We identified, through a genome-wide association study, 15 loci associated with parameters of circulating leukocyte subpopulations, 12 of which are novel. Finally, we show that cellular mediators of innate and adaptive immunity are differentially affected by non-genetic and genetic factors under homeostatic conditions.

Results

Variation of immune cell parameters in the general population

The Milieu Intérieur cohort includes 500 men and 500 women, stratified across five decades of age from 20 to 69 years. Subjects were surveyed for a number of demographic variables, including past infections, vaccination and surgical histories and health-related habits (**Supplementary Table 1**). Detailed inclusion and exclusion criteria used to define "healthy" subjects recruited into the cohort have been previously reported[18].

To describe natural variation of both innate and adaptive immune cells in the 1,000 subjects, we used ten 8-color immunophenotyping flow cytometry panels (**Supplementary Figs. 1-10** and **Supplementary Table 2**; **Online Methods**), which allowed us to report a total of 166 distinct immunophenotypes (**Supplementary Table 3**). Our resource includes 75 (46%) and 91 (54%) immunophenotypes obtained in innate and adaptive immune cells, respectively. Innate cells were defined as those lacking somatic recombination of the genome[19], and included granulocytes (neutrophils, basophils and eosinophils), monocytes, natural killer (NK) cells, dendritic cells and innate lymphoid cells (ILCs) (**Fig. 1**). Adaptive cells were defined by their dependence on RAG1/2 activity and included T cells (γδ T, MAIT, NKT, T_{reg} and T_H cells) and B cells. The immunophenotypes in both innate and adaptive immune cells included 76 absolute counts of circulating cells, 87 expression levels of cell-surface protein markers (quantified by the mean fluorescence intensity, or MFI), and 3 ratios of cell counts or MFI (**Supplementary Fig. 11** and **Supplementary Table 3**).

To reduce technical variation introduced by sample temperature fluctuations and pre-analytical procedures, we strictly followed a standardized protocol for tracking and processing samples[20]. We verified that measured immunophenotypes were highly reproducible using technical replicates (**Supplementary Figs. S12** and **S13** and **Supplementary Table 3**), demonstrating the high precision of the data. We nevertheless identified two technical batch effects that impacted flow cytometric analyses. One effect corresponded to the hour at which the blood sample was drawn from fasting subjects (**Supplementary Fig. 14a**), which may possibly be explained by the spike in cortisol at the time of waking[21]. The second effect corresponded to temporal variation of immunophenotypes over

the one-year sampling period, which did not follow the periodic distribution observed for cellular traits under seasonal fluctuations[11], and primarily affected MFI measures (**Supplementary Fig. 14b**). We corrected for these batch effects in all subsequent analyses (**Supplementary Fig. 15**; **Online Methods**), and provide the distribution, ranges and statistics of all batch-corrected immune cell counts (**Supplementary Table 3**), thereby facilitating comparisons with cytometry data collected as part of routine clinical practice. This resource can be accessed through a user-friendly web application (http://104.236.137.56:3838/LabExMICytometryBrowser_ShinyApp/), which can be queried based on personal characteristics, such as age or sex.

Owing to the hierarchical structure of immune cell differentiation (i.e., cellular lineages emerge from common progenitor cells), a substantial portion of the immune cell counts measured in this study were highly correlated (**Supplementary Fig. 16**). These correlations were not directly attributable to the influence of factors such as age or sex, which were regressed out in this analysis. We observed correlations between circulating levels of ILC and NK populations, reflecting their common developmental pathway and dependence on γ_c cytokines[22]. Likewise, MAIT cells and $CCR6^+$ $CD8^+$ T cells were also correlated, owing to the former being the major subset of $CCR6^+$ T cells in circulation[23]. Finally, we identified a strong correlation between the number of T_{reg} and conventional $CD4^+$ T cells, validating previous experimental work that defined an IL-2-driven self-regulatory circuit that integrates the homeostasis of these cell populations[24].

Impact of age, sex and CMV infection on innate and adaptive cell parameters

Prior studies have shown that two intrinsic factors, age and sex, are responsible for inter-individual variation in white blood cell composition[6,7,10,14,25–27]. We used linear mixed models to quantify the respective impact of each of these intrinsic factors on variation in innate and adaptive cell composition. We observed a significant effect of age on 35% of immune cell parameters (adjusted $P<0.01$; **Fig. 2a** and **Supplementary Fig. 17a**), among which only 29% were measured in innate cells. We detected a general decline in the number of ILC and plasmacytoid dendritic cells (pDCs) and an increase in the number of $CD16^{hi}$ monocytes with increasing age (**Fig. 2a**), which might contribute to the altered immune response to viral infections in elderly persons and age-associated

inflammation[14,28,29]. We found a modest increase in the number of memory T cells with age, supporting the view that the observed expansion of these cell populations in elderly subjects is not due to aging per se, but to CMV seropositivity[13], which we accounted for in the model. Our analyses also revealed that naive $CD8^+$ T cells decrease more than twice as rapidly with age as compared to naive $CD4^+$ T cells, at a rate of 3.6 % (99% FCR-adjusted Confidence Interval (99%CI): [3.0%, 4.1%]) and 1.6 % (99%CI: [1.1%, 2.1%]) per year, respectively (**Fig. 2a-c**), supporting the view that $CD8^+$ T cells are more susceptible to concentrations of homeostatic cytokines and/or that the production of $CD4^+$ T cells is preferentially enhanced in the human thymus[30].

Although sex differences have been previously reported for various immune responses and diseases[25], studies examining circulating cellular parameters have reported inconsistent results, owing to both differences between flow cytometry procedures and relatively small, underpowered or poorly-stratified study cohorts. We report a significant impact of sex on 16% of measured immunophenotypes (adjusted $P<0.01$, **Fig. 2d** and **Supplementary Fig. 17b**), of which 38% were measured in innate cells. We found a higher number of activated NK cells in men, as compared to women. By contrast, MAIT cells were systematically increased in women, across all age decades (**Fig. 2e-f**), collectively suggesting a lasting effect of early hormonal differences on immune cell development and biology.

Environmental exposures are also known to drive immune variation, among which persistent CMV infection is one of the strongest candidates[6,13,14,17]. We observed a significant effect of latent CMV infection on 13% of immune cell parameters (**Fig. 2g** and **Supplementary Fig. 17c**), of which more than 75% were measured in adaptive cells. We confirm that CMV triggers a major change in the number of memory T cells, which is independent from age effects[13,17]. In particular, CMV seropositivity associated with a 12.5-fold (99%CI: [8.8, 17.6]) higher number of $CD4^+$ effector memory RA T cells (T_{EMRA}), and a 4.6-fold (99%CI: [3.5, 6.0]) higher number of $CD8^+$ T_{EMRA} cells (**Fig. 2g-i**). However, we did not find evidence that CMV infection impacts the number of naive or central memory (T_{CM}) T cell compartments. Supporting this observation, the total number of $CD8^+$ and $CD4^+$ T cells increased in parallel with the expanded number of memory T cells, thus suggesting independent regulation of the naive and T_{EM} and/or T_{EMRA} cell pools. CMV seropositive donors also

149 presented lower numbers of circulating NKT and MAIT cells (**Fig. 2g**). Together, our broad resource
150 provides a comprehensive quantification of the respective impact that age, sex and CMV infection
151 have on immune cell parameters. In doing so, our results suggest a stronger impact of these factors on
152 adaptive cells, relative to innate cells.
153
154 **Tobacco smoking extensively alters innate and adaptive cell numbers**
155 Capitalizing on the detailed lifestyle and demographic data obtained for the Milieu Intérieur cohort,
156 we evaluated the influence of additional environmental factors on immune cell parameters, controlling
157 for the defined effects of age, sex and CMV serological status. A total of 39 variables were chosen for
158 analysis and tested for association with each immunophenotype. These include socio-economic
159 characteristics, past infections, health-related habits and surgery and vaccination history
160 (**Supplementary Fig. 18** and **Supplementary Table 1**). We identified a unique environmental factor
161 that significantly alters circulating numbers of immune cells: active tobacco cigarette smoking, which
162 affects 36% of measured immunophenotypes (**Fig. 3a** and **Supplementary Fig. 19**), of which 36%
163 were measured in innate cells.
164 We observed a 23% (99%CI: [11%, 37%]) increase in the number of circulating $CD45^+$ cells, and
165 a 26% (99%CI: [10%, 45%]) increase in the number of conventional lymphocytes in smokers as
166 compared to non-smokers (**Fig. 3b**). Previous studies suggested that smokers have alterations in
167 circulating cell populations due to diminished adherence of leukocytes to blood vessel walls, possibly
168 as a result of lower antioxidant concentrations[31]. Furthermore, we found in active smokers a
169 significant increase of 43% (99%CI: [17%, 76%]) and 41% (99%CI: [15%, 71%]) of activated and
170 memory T_{reg} cells, respectively, a pattern that was also observed to a lesser extent in past smokers
171 (**Fig. 3b-d**). Active smokers also showed decreased numbers of NK cells, ILCs, γδ T cells and
172 different subsets of MAIT cells (**Fig. 3b**). These findings are consistent with a study showing that
173 smoking triggers local release of IL-33 by the lung epithelium[32], in turn engaging the IL-33 receptor,
174 ST2, on both innate and non-classical lymphocytes[33]. Collectively, these findings reveal that active
175 smoking has a profound impact on immune cell parameters, which is similar in magnitude to that of
176 age, and affects both innate and adaptive cells.

Genome-wide association study of 166 immune cell parameters

To identify common genetic variants affecting inter-individual variation in immune cell parameters, the Milieu Intérieur cohort was genotyped at 945,213 SNPs, enriched in exonic SNPs (**Online Methods**). After quality control (**Supplementary Fig. 20**), genotype imputation was performed and yielded a total of 5,699,237 highly accurate SNPs, which were tested for association with the 166 immunophenotypes using linear mixed models. The models were adjusted for the genetic relatedness among subjects and any non-genetic variable identified as predictive of each specific immunophenotype by stability selection based on elastic net regression (**Supplementary Table 3**; **Online Methods**). We confirmed our power to identify medium-effect genotype-phenotype associations by simulations, and by empirically replicating well-known genetic associations with non-immune traits, such as eye and hair color or uric acid and cholesterol levels (**Online Methods**).

With respect to immune traits, we found 14 independent genetic loci associated with 42 out of 166 immunophenotypes (25%), at a conservative genome-wide significant threshold of $P<1.0\text{x}10^{-10}$ (**Fig. 4a**, **Table 1**, **Supplementary Fig. 21**, **Supplementary Tables 4** and **5**). We then conducted conditional GWAS, by adjusting these 42 immunophenotypes on the 14 leading associated variants (**Table 1**), and found an additional independent locus reaching genome-wide significance (**Supplementary Fig. 22** and **Supplementary Table 6**). Genome-wide significant associations were replicated in an independent cohort of 75 European-descent donors, for all immune traits measured in this replication cohort ($P<0.05$; **Table 1**; **Online Methods**). Also, we confirmed that our immune cell measurements were stable, as all genome-wide significant associations were confirmed for immunophenotypes measured in a new blood draw taken in 500 of the 1,000 subjects of the Milieu Intérieur cohort, sampled 7 to 44 days after the initial visit ($P<10^{-3}$; **Table 1**). We also provide a list of 26 suggestive association signals ($P<5.0\text{x}10^{-8}$), including a number of biologically relevant candidate genes (**Supplementary Table 6**). The associated genetic loci were enriched in SNPs associated by GWAS with diseases (31% observed *vs.* 5% expected, resampling $P=0.0032$), most of which were autoimmune diseases, including rheumatoid arthritis, Vogt-Koyanagi-Harada syndrome and atopic

dermatitis (**Supplementary Table 4)**. These findings highlight the importance of loci altering immune cell populations in the context of ultimate organismal traits affecting human health.

Genetic associations primarily identify immune cell-specific protein QTLs

Of the 42 immunophenotypes for which a significant genetic association was detected, 36 (86%) were MFI, which measures the cell-specific expression of protein markers conventionally used to determine the differentiation or activation state of leukocytes. For 28 of these 36 MFI measurements (78%), the genetic association was observed between the protein MFI and SNPs located in the vicinity of the gene encoding the corresponding protein (**Table 1** and **Supplementary Fig. 21**), i.e., local protein QTLs (local-pQTLs). For instance, genetic variation close to the *ENPP3* gene was associated with CD203c MFI in basophils (rs2270089, P=2.1x10^{-28}), *CD24* with CD24 MFI in marginal zone B cells (rs12529793, P=3.8x10^{-21}) and *CD8A* with CD8a MFI in CD69^{+} CD16^{hi} NK cells (rs71411868, P=5.9x10^{-58}).

We identified two independent local-pQTLs in the *FCGR* gene cluster (**Table 1**), which encodes the most important Fc receptors for inducing phagocytosis of opsonized microbes. Genetic variation close to *FCGR3A* was associated here with CD16 MFI in CD16^{hi} NK cells (rs3845548, P=3.0x10^{-87}). The same variants were also shown to affect the number of CD62L- myeloid cDCs in a previous study[15]. The second signal associated *FCGR2B* variation with CD32 MFI in basophils (rs61804205, P=1.7x10^{-36}), but not in eosinophils and neutrophils. Consistently, it is known that basophils express both CD32a and CD32b proteins, while eosinophils and neutrophils predominantly express CD32a[34]. Conversely, a local-pQTL was identified at the *SELL* gene, which was associated with CD62L MFI in eosinophils and neutrophils (rs2223286, P=1.6x10^{-35} and 8.8x10^{-13}, respectively), but not in basophils (**Fig. 4b, c**).

A number of other local-pQTLs were found to be cell-specific; three different association signals were found in the *HLA-DR* gene region, with HLA-DR MFI in pDCs and CD14^{hi} monocytes (rs114973966, P=2.2x10^{-56}), in cDC1 (rs2760994, P=6.1x10^{-38}) and in cDC3 cells (rs143655145, P=2.6x10^{-11}). To verify if these signals were independent from each other, we conducted omnibus association tests on imputed HLA alleles (**Online Methods**). We found that the association signals in

CD14^{hi} monocytes, pDCs and cDC1 actually resulted from different amino acid-altering variants at the same multi-allelic position 13 of the HLA-DRβ1 protein (P=2.0x10^{-47}, 7.0x10^{-90} and 5.3x10^{-41} in CD14^{hi} monocytes, pDC and cDC1, respectively; **Supplementary Tables 7** and **8**), recently shown to explain a large part of the association signal in the *HLA* locus for type 1 diabetes[35]. A different amino-acid variant, at position 67 of HLA-DRβ1, was identified in cDC3s (P=3.9x10^{-13}). Conditional analyses also revealed independent associations of HLA-DR cell-surface expression with two residues in class I *HLA-B* gene (position 97 and 194; P=3.8x10^{-17} and 1.3x10^{-18}; **Supplementary Tables 7** and **8**). Collectively, these results show that the protein expression of markers of immune cell differentiation and activation can be affected by common genetic variants, of which some are known to be implicated in human pathogenesis.

Immune cell local protein QTLs control mRNA levels of nearby genes

Although four of the 9 local-pQTLs identified by our analyses are likely explained by amino acid-altering variants in surrounding genes (**Supplementary Tables 4** and **7**), the remaining signals do not present obvious candidate causal variants. To dissect the functional basis of these associations, we tested if the corresponding SNPs were also associated with mRNA levels of nearby genes (i.e., expression QTL, eQTL) using gene expression data obtained from the same donors[36] and results from the Genotype-Tissue Expression (GTEx) Project[37]. Five of the local-pQTLs were strongly associated with the transcript levels of a surrounding gene (P<1.0x10^{-5}; **Fig. 4d**). The SNPs controlling the MFI of CD16 in CD16^{hi} NK cells and CD32 in basophils, CD62L in eosinophils, CD8a in CD69^{+} CD16^{hi} NK cells and CD203c in basophils were associated with mRNA levels of *FCGR2B*, *SELL*, *CD8A*, and *ENPP3*, respectively (**Supplementary Table 4**). These analyses indicate that genetic variants associated with immunophenotypes can directly affect gene expression of markers of immune cells in whole blood. This suggests that eQTL mapping in different immune cell compartments can greatly improve our knowledge of the genetic factors controlling human inter-individual variation in flow cytometric parameters.

Novel trans-acting genetic associations with immune cell parameters

We detected six loci that do not exclusively act as local-pQTLs on immunophenotypes (**Table 1** and **Supplementary Fig. 21**). These included variants that are associated with immune cell counts, or that are genetically independent from the genes encoding immune cell markers with which they are associated (i.e., trans-pQTLs). A variant in the vicinity of the *S1PR1* gene was associated with CD69 MFI in $CD16^{hi}$ NK cells (rs6693121, $P=4.8x10^{-37}$). CD69 is known to downregulate cell-surface expression of the sphingosine-1-phosphate receptor-1 (S1P1) on lymphocytes, a mechanism that elicits egress from the thymus and secondary lymphoid organs[38]. Genetic variation in an intron of the *ACOXL* gene, close to *BCL2L11*, was associated with the absolute count of $CD8a^{+}$ $CD56^{hi}$ NK cells (rs12986962, $P=9.1x10^{-19}$). BCL2L11 (also known as BIM) is an important regulator of lymphocyte apoptosis[39], and is associated with chronic lymphocytic leukemia and total blood cell number[40]. A third association involved genetic variants close to the *ACTL9* gene and the ratio of CD16 MFI in $CD16^{hi}$ and $CD56^{hi}$ NK cells (rs114412914, $P=4.3x10^{-30}$). The same variants have been also found to be associated with $CD56^{++}$ $CD16^{-}$ NK cells in another study[10].

Although identified here for their trans effects on markers of immune cell differentiation or activation, three trans-acting genetic associations were also local-eQTLs for nearby immune-related genes[37] (**Supplementary Tables 4** and **6**). The MFI of CCR7 in $CD4^{+}$ and $CD8b^{+}$ naive T cells was associated with a variant in the *TMEM8A* gene (rs11648403, $P=3.0x10^{-19}$), which also controls *TMEM8A* mRNA levels ($P=2.5x10^{-27}$). TMEM8A is expressed on the surface of resting T cells and is down-regulated after cell activation[41], suggesting a possible functional association and/or co-regulation with CCR7. Variants in the vicinity of the *ALOX15* gene were associated with increased protein levels of the high-affinity IgE receptor in eosinophils (rs56170457, $P=9.2x10^{-14}$) and increased *ALOX15* mRNA levels ($P=2.7x10^{-13}$). These results, together with the high expression of the ALOX15 protein and its pro-inflammatory effect in circulating eosinophils[42], suggest that this lipoxygenase plays an important role in IgE-dependent allergic reactions. Finally, conditional GWAS identified an additional trans-acting association, between a variant close to the *CD83* gene and HLA-DR MFI in cDC1 (rs72836542, $P=2.8x10^{-12}$, **Supplementary Fig. 22**), the same variant being also identified as a local-eQTL of *CD83* gene expression ($P=5.4x10^{-21}$). These results suggest that CD83,

an early activation marker of human DCs, upregulates HLA-DR expression in activated dendritic cells.

Natural variation of innate immune cell parameters is preferentially driven by genetic factors

A large proportion of both MFI and cell number immunophenotypes that presented a genome-wide association were detected in innate immune cells (35/44, 80%), including granulocytes, monocytes, NK and dendritic cells (**Table 1**), while 47% of all immunophenotypes were measured in innate cells (**Supplementary Table 3**). Furthermore, of the adaptive cell immunophenotypes showing genetic associations, 3 of the 9 measurements (33%) were related to naive T or B cells, while naive adaptive cell parameters represented <10% of all adaptive cell measurements. These observations suggest a stronger effect of genetic variants on innate and naive adaptive cell subpopulations, relative to differentiated or experienced adaptive immune cells.

In support of this hypothesis, the presence of HLA-DR molecules, which was assessed at the surface of both innate and adaptive immune cells, was strongly associated with *HLA-DR* genetic variation in monocytes, NK and dendritic cells (**Table 1**), but not in memory $CD4^+$ or $CD8^+$ T_{CM}, T_{EM} and T_{EMRA} cells ($P>1.0x10^{-6}$; **Supplementary Table 5**). Because we observed substantial correlations among $HLA\text{-}DR^+$ memory T cell numbers ($R^2\approx0.3$, $P<0.05$; **Supplementary Fig. 16**), we hypothesized that they were at least partly controlled by the same genetic factors, which were further examined using a multivariate GWAS (**Online Methods**). This refined approach detected a suggestive genetic association close the *HLA-DRB1* gene with a variant (rs35743245, $P=1.0x10^{-8}$) in strong linkage disequilibrium with that detected in pDCs, monocytes and NK cells ($r^2=0.92$; **Supplementary Fig. 23**). This finding provides proof-of-concept that immunophenotypes in both innate and adaptive cells can be controlled by the same genetic factors, but their effects are stronger in innate cells, relative to experienced adaptive cells.

We next systematically quantified the impact of genetic and non-genetic factors on innate and adaptive cells. We established, for each immunophenotype, a linear regression model that included the four most impactful non-genetic variables (**Figs. 2** and **3**) and all genome-wide significant and suggestive variants (**Table 1** and **Supplementary Table 6**), and estimated their respective

contribution to the total variance (**Online Methods**). We found that a larger proportion of the variance of innate cell immunophenotypes was explained by genetic factors (**Fig. 5b** and **5d**), relative to adaptive cell immunophenotypes (**Fig. 5a** and **5c**). Inversely, the variance in adaptive cell numbers was dominated by non-genetic factors such as age and CMV serostatus (**Fig. 5a**). To test if these differences were significant, we used a mixed model that accounted for correlations among immunophenotypes (**Online Methods**). Conclusively, we estimated that the variance explained by genetics was 66% larger for innate cell measurements, relative to adaptive cells (95%CI: [13%-143%]; bootstrap P=0.012; Mann-Whitney U test: P=0.032), while the variance explained by non-genetic factors was 46% smaller for innate cell measurements (95%CI: [22%-63%]; bootstrap P=1.8x10^{-3}; Mann-Whitney U test: P=8.1x10^{-3}). When considering non-genetic factors separately, the ratio of explained variance between innate and adaptive cell measurements was the smallest for smoking (0.46, 95%CI: [0.17-1.25]), followed by age (0.63, 95%CI: [0.42-0.95]), CMV infection (0.71, 95%CI: [0.51-0.99]), and sex (0.95, 95%CI: [0.60-1.51]). Taken together, our results indicate that genetic factors account for a substantial fraction of human variation in immune cell parameters, with their influence being stronger in innate immune cells, relative to adaptive immune cell phenotypes.

Discussion

Over the last two decades, research in human immunology has employed multi-parametric cytometry to enumerate and assess the activation state of immune cells in healthy and disease conditions. Although immune cell parameters do vary in the general population, the extent to which intrinsic, environmental and genetic factors explain this variability remained elusive. To tackle these questions, we generated a broad resource by combining standardized flow cytometry with genome-wide DNA genotyping in a demographically well-defined cohort of 1,000 healthy individuals. We confirm the strong and independent impacts of age and CMV infection on naive and memory T cell populations, respectively, and provide robust evidence for sex differences in innate and adaptive cell numbers. We show that immune homeostasis is altered upon chronic cigarette smoke exposure, which elicits both a decline of MAIT cells, possibly due to their increased migration to sites of inflammation, and an increase in the numbers of activated and memory T_{reg} cells, suggesting a role for these immunosuppressive populations in the increased susceptibility of smokers to infection[43]. Furthermore, we found that human genetic variation substantially impacts immune cell parameters, particularly the cell-surface expression of markers conventionally used to identify leukocyte differentiation or activation. These results highlight the need to consider non-genetic and genetic features when interpreting parameters such as circulating white blood cells of patients, a critical aspect in clinical monitoring. For instance, HLA-DR expression on monocytes is routinely measured by flow cytometry to predict the clinical course of septic shock and identify patients who should benefit from immunoadjuvant therapies[44]. We identified a strong effect of HLA-DRβ1 coding variation on HLA-DR expression in $CD14^{hi}$ monocytes, suggesting that prognostic tools of fatal outcome in sepsis should be tailored to patient's genetic makeup.

The most prominent result of our study is the lower number of genetic associations detected in memory T and B cells, relative to innate cells, an observation that could be explained by their strong dependence on the varying individual history of past infections. Adaptive immune cells are known to possess a much longer half-life as compared to myeloid innate cells, in mice and humans[45,46]. Stimulus-induced differentiation and expansion may also result in the possible masking of genetic associations for adaptive cell types. Consistently, genetic associations in adaptive immune cells were

primarily observed for immunophenotypes of naive adaptive cells. Our observations are further supported by a GWAS of 36 blood traits in 173,480 individuals, which found that the genetic heritability of monocyte and eosinophil counts was larger than that of lymphocyte counts[27]. This is however at odds with another recent study, which concluded that adaptive immune traits are more affected by genetics, whereas innate immune traits are more affected by environment, based on the estimated genetic heritability of 23,394 immune phenotypes in 497 adult female twins[47]. We suggest that such deep immunophenotyping in large-scale cohorts, combined with statistical tests for differences in heritability that account for inherent correlations among phenotypes, may reveal a more balanced contribution of genetics on the natural variation of innate and adaptive immune cell traits.

Our findings that genetic factors preferentially controls variation in innate immune cells have other important consequences. A previous study of 105 healthy twin pairs concluded that variation in cell population frequencies is largely driven by non-heritable influences[17]. We find instead that genetic variation explains a large part of the variance of immune cell parameters, particularly MFIs (i.e., cell-surface expression of protein markers) measured in innate cells. This discrepancy may stem from the fact that this previous study considered only a fraction of innate myeloid and lymphoid populations[48], and possibly because of its limited power due to a moderate sample size. Also, our results suggest that the genetic control of cell-surface expression of immune cell markers is stronger than that of cell counts, and the former were not assessed in most previous population immunology studies[10,15,17].

Finally, the mapping of genetic loci that control immune cell parameters identified cell-specific pQTLs that are enriched in genetic variants associated with human diseases and traits. For example, we identified the position 13 of the HLA-DRβ1 protein as a predictor of HLA-DR expression at the cell-surface of pDCs and monocytes, which in turn is strongly associated with type 1 diabetes[35], suggesting the implication of innate immunity in the disease[27]. Furthermore, the expression of CD56 and CD16 in NK cells is controlled by genetic variants close to the *ACTL9* gene, which were shown to be associated with atopic dermatitis[49], suggesting a possible involvement of NK cells in this pathology[50]. More generally, genetic variants found to modulate innate immune cell parameters, in this and previous studies[10,15,16], have been directly implicated in the aetiology of several autoimmune

disorders, such as inflammatory bowel disease, ulcerative colitis and atopic dermatitis. Together, these findings illustrate the value of our approach, which mapped novel genetic associations to specific cell populations and cellular states, providing new insights into the mechanisms underlying disease pathogenesis. Further evaluations of the natural variability in cellular mediators of immunity, together with the elucidation of their environmental and genetic determinants, will facilitate a detailed dissection of the immune system in human health and disease.

References:

1. Bernard, C. *Introduction à l'étude de la médecine expérimentale*. (Libraires de l'Académie Impériale de Médecine, 1865).
2. Altfeld, M. & Gale, M. Innate immunity against HIV-1 infection. *Nat. Immunol.* **16,** 554–562 (2015).
3. Orme, I. M., Robinson, R. T. & Cooper, A. M. The balance between protective and pathogenic immune responses in the TB-infected lung. *Nat. Immunol.* **16,** 57–63 (2015).
4. Tollerud, D. J. *et al.* The Influence of Age, Race, and Gender on Peripheral Blood Mononuclear-Cell Subsets in Healthy Nonsmokers. *J. Clin. Immunol.* **9,** 214–222 (1989).
5. Reichert, T. *et al.* Lymphocyte Subset Reference Ranges in Adult Caucasians. *Clin. Immunol. Immunopathol.* **60,** 190–208 (1991).
6. Liston, A., Carr, E. J. & Linterman, M. A. Shaping Variation in the Human Immune System. *Trends Immunol.* **37,** 637–646 (2016).
7. Goronzy, J. J. & Weyand, C. M. Successful and maladaptive T cell aging. *Immunity* **46,** 364–378 (2017).
8. Sauce, D. & Appay, V. Altered thymic activity in early life : how does it affect the immune system in young adults ? *Curr. Opin. Immunol.* **23,** 543–548 (2011).
9. Furman, D. *et al.* Apoptosis and other immune biomarkers predict influenza vaccine responsiveness. *Mol. Syst. Biol.* **9,** 1–14 (2013).
10. Aguirre-Gamboa, R. *et al.* Differential Effects of Environmental and Genetic Factors on T and B Cell Immune Traits. *Cell Rep.* **17,** 1–14 (2016).
11. Carr, E. J. *et al.* The cellular composition of the human immune system is shaped by age and cohabitation. *Nat. Immunol.* **17,** 461–468 (2016).
12. Boeckh, M. & Geballe, A. P. Cytomegalovirus: pathogen, paradigm, and puzzle. *J Clin Invest* **121,** 1673–1680 (2011).
13. Wertheimer, A. M. *et al.* Aging and Cytomegalovirus Infection Differentially and Jointly Affect Distinct Circulating T Cell Subsets in Humans. *J. Immunol.* **192,** 2143–2155 (2014).
14. Furman, D. *et al.* Cytomegalovirus infection enhances the immune response to influenza. *Sci. Transl. Med.* **7,** 281ra43 (2015).
15. Orrù, V. *et al.* Genetic variants regulating immune cell levels in health and disease. *Cell* **155,** 242–56 (2013).

16. Roederer, M. *et al.* The Genetic Architecture of the Human Immune System: A Bioresource for Autoimmunity and Disease Pathogenesis. *Cell* **161,** 387–403 (2015).
17. Brodin, P. *et al.* Variation in the Human Immune System Is Largely Driven by Non-Heritable Influences. *Cell* **160,** 37–47 (2015).
18. Thomas, S. *et al.* The Milieu Intérieur study — An integrative approach for study of human immunological variance. *Clin. Immunol.* **157,** 277–293 (2015).
19. Vivier, E. *et al.* Innate or Adaptive Immunity? The Example of Natural Killer Cells. *Science (80-.).* **331,** 44–49 (2011).
20. Hasan, M. *et al.* Semi-automated and standardized cytometric procedures for multi-panel and multi-parametric whole blood immunophenotyping. *Clin. Immunol.* **157,** 261–276 (2015).
21. Patterson, S. *et al.* Cortisol Patterns Are Associated with T Cell Activation in HIV. *PLoS One* **8,** e63429 (2013).
22. Serafini, N., Vosshenrich, C. A. J. & Di Santo, J. P. Transcriptional regulation of innate lymphoid cell fate. *Nat. Rev. Immunol.* **15,** 415–428 (2015).
23. Dusseaux, M. *et al.* Human MAIT cells are xenobiotic-resistant, tissue-targeted, CD161hi IL-17–secreting T cells. *Blood* **117,** 1250–1260 (2011).
24. Amado, I. F. *et al.* IL-2 coordinates IL-2-producing and regulatory T cell interplay. *J. Exp. Med.* **210,** 2707–2720 (2013).
25. Pennell, L. M., Galligan, C. L. & Fish, E. N. Sex affects immunity. *J. Autoimmun.* **38,** J282–J291 (2012).
26. Furman, D. *et al.* Systems analysis of sex differences reveals an immunosuppressive role for testosterone in the response to influenza vaccination. *Proc. Natl. Acad. Sci.* **111,** 869–874 (2014).
27. Astle, W. J. *et al.* The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell* **167,** 1415–1429 (2016).
28. Della Bella, S. *et al.* Peripheral blood dendritic cells and monocytes are differently regulated in the elderly. *Clin. Immunol.* **122,** 220–228 (2007).
29. Puchta, A. *et al.* TNF Drives Monocyte Dysfunction with Age and Results in Impaired Anti-pneumococcal Immunity. *PLoS Pathog.* **12,** e1005368 (2016).
30. Vrisekoop, N. *et al.* Sparse production but preferential incorporation of recently produced naïve T cells in the human peripheral pool. *Proc. Natl. Acad. Sci.* **105,** 6115–6120 (2008).

31. Tsuchiya, M. *et al.* Smoking a Single Cigarette Rapidly Reduces Combined Concentrations of Nitrate and Nitrite and Concentrations of Antioxidants in Plasma. *Circulation* **105,** 1155–1157 (2002).
32. Kearley, J. *et al.* Cigarette Smoke Silences Innate Lymphoid Cell Function and Facilitates an Exacerbated Type I Interleukin-33-Dependent Response to Infection. *Immunity* **42,** 566–579 (2015).
33. Monticelli, L. A. *et al.* Articles Innate lymphoid cells promote lung-tissue homeostasis after infection with influenza virus. *Nat. Immunol.* **12,** 1045–1055 (2011).
34. Cassard, L., Jönsson, F., Arnaud, S. & Daëron, M. Fcγ Receptors Inhibit Mouse and Human Basophil Activation. *J. Immunol.* **189,** 1–13 (2012).
35. Hu, X. *et al.* Additive and interaction effects at three amino acid positions in HLA-DQ and HLA-DR molecules drive type 1 diabetes risk. *Nat. Genet.* **47,** 898–905 (2015).
36. Piasecka, B. *et al.* Distinctive Roles of Age, Sex and Genetics in Shaping Transcriptional Variation of Human Immune Responses to Microbial Challenges. *Proc. Natl. Acad. Sci.* **TBD,** 1–55 (2017).
37. GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science (80-.).* **348,** 648–660 (2015).
38. Garris, C. S., Blaho, V. A., Hla, T. & Han, M. H. Sphingosine-1-phosphate receptor 1 signalling in T cells : trafficking and beyond. *Immunology* **142,** 347–353 (2014).
39. Pellegrini, M. *et al.* Loss of Bim Increases T Cell Production and Function in Interleukin 7 Receptor–deficient Mice. *J. Exp. Med.* **200,** 1189–1195 (2004).
40. van der Harst, P. *et al.* Seventy-five genetic loci influencing the human red blood cell. *Nature* **492,** 369–375 (2012).
41. Motohashi, T. *et al.* Molecular Cloning and Chromosomal Mapping of a Novel Protein Gene, M83. *Biochem Biophys Res Commun* **250,** 244–250 (2000).
42. Feltenmark, S. *et al.* Eoxins are proinflammatory arachidonic acid metabolites produced via the 15-lipoxygenase-1 pathway in human eosinophils and mast cells. *Proc. Natl. Acad. Sci.* **105,** 680–685 (2008).
43. Stämpfli, M. R. & Anderson, G. P. How cigarette smoke skews immune responses to promote infection, lung disease and cancer. *Nat Rev Immunol* **9,** 377–384 (2009).
44. Venet, F., Lukaszewicz, A.-C., Payen, D., Hotchkiss, R. & Monneret, G. Monitoring the immune response in sepsis: a rational approach to administration of immunoadjuvant therapies. *Curr. Opin. Immunol.* **25,** 477–483 (2013).

45. Kolaczkowska, E. & Kubes, P. Neutrophil recruitment and function in health and inflammation. *Nat. Rev. Immunol.* **13,** 159–175 (2013).

46. Farber, D. L., Yudanin, N. A. & Restifo, N. P. Human memory T cells: Generation, compartmentalization and homeostasis. *Nat. Rev. Immunol.* **14,** 24–35 (2014).

47. Mangino, M., Roederer, M., Beddall, M. H., Nestle, F. O. & Spector, T. D. Innate and adaptive immune traits are differentially affected by genetic and environmental factors. *Nat. Commun.* **8,** 13850 (2017).

48. Casanova, J. & Abel, L. Disentangling Inborn and Acquired Immunity in Human Twins. *Cell* **160,** 13–15 (2015).

49. Paternoster, L. *et al.* Multi-ancestry genome-wide association study of 21,000 cases and 95,000 controls identifies new risk loci for atopic dermatitis. *Nat. Genet.* **47,** 1449–1456 (2015).

50. von Bubnoff, D. *et al.* Natural killer cells in atopic and autoimmune diseases of the skin. *J Allergy Clin Immunol* **125,** 60–68 (2010).

Data Availability:

The SNP array data that support the findings of this study have been deposited in the European Genome-Phenome Archive (EGA) with the accession code EGAS00001002460. The flow cytometric data can be downloaded as an R package (XX) and explored with the Shiny web application available in http://milieu_interieur_cytoGWAS.pasteur.fr.

Code Availability:

The code developed to identify non-genetic factors that impact immunophenotypes and quantify their effects has been made available in http://github.com/JacobBergstedt/mmi.

Acknowledgments:

This work benefited from support of the French government's Program *Investissement d'Avenir*, managed by the Agence Nationale de la Recherche (ANR, reference 10-LABX-69-01). We thank the Center for Translational Science, Institut Pasteur; and the OMNI Biomarker Development-Flow Cytometry Biomarker group, Genentech for their expert support. J.B. is a member of the LCCC Linnaeus Center and the ELLIIT Excellence Center at Lund University and is supported by the ELLIIT Excellence Center.

Author contributions:

Author contributions were as follows: Conceptualization, E.P., L.Q.-M., M.L.A.; Methodology, M.H., V.L., A.U., F.J., B.B., C.L., F.H., L.R., I.P., O.L., J.P.D.; Software, J.B., E.P., V.R., P.S., C.H., B.P., J.F.; Validation, M.H., V.R., F.J., Y.W.L., M.L.A.; Formal analysis, E.P., J.B., V.R., P.S., C.G., C.H., B.P., J.F.; Investigation, E.P., M.H., J.B., V.L., A.U., C.A., F.J., H.Q., M.Z., B.B., C.L., F.H., L.R., O.L., J.P.D., M.L.A.; Data curation, E.P., M.H., J.B., V.R., V.L., A.U., B.P., C.L., L.R., I.P., O.L., J.P.D.; Writing – Original Draft, E.P., J.B., C.A., D.D., M.L.A.; Writing – Review & Editing, E.P., J.B., M.H., C.A., L.R., O.L., M.F., J.F., J.P.D., L.Q.-M., M.L.A.; Supervision, E.P., M.H., M.F., J.F.,

D.D., M.L.A.; Project Administration, S.T., D.D., J.H.; Funding Acquisition, M.F., J.F., L.Q.-M., M.L.A.

Competing financial interests:

A.U., C.H., Y.W.L., J.H., M.Z., C.G. and M.L.A. are employees of Genentech Inc., a member of The Roche Group.

Figure legends

Figure 1 Immune cell counts and cell-surface markers measured in the Milieu Intérieur cohort. Panel numbers refer to the cytometric analyses performed, grouped based on cellular lineage (**Supplementary Figs. 1-10** and **Supplementary Tables 2** and **3**). The expression of phenotypic markers of differentiation or activation was quantified based on their mean fluorescent intensity (MFI), indicated per panel. Interconnecting lines illustrate cellular lineages or differentiation states. Red and blue squares indicate immunophenotypes significantly associated in this study with non-genetic or genetic factors, respectively.

Figure 2 Respective effects of age, sex and CMV infection on innate and adaptive cell counts in 1,000 healthy individuals. Significant multiplicative effects (adjusted $P<0.01$) of (**a-c**) increasing age, (**d-f**) female sex and (**g-i**) CMV seropositivity on circulating levels of immune cells. (**a, d, g**) Effect sizes were estimated in a linear mixed model with a log-transformed immunophenotype as response, controlling for batch effects and genome-wide significant SNPs, and then transformed to the original scale. Adaptive and innate immune cells are represented in grey and black, respectively. The 99% confidence intervals (99%CIs) were false coverage-adjusted. (**b, e, h**) Regression lines were fitted using local polynomial regression. (**b**) Impact of age on naive $CD8b^+$ (in dark green) and $CD4^+$ (in light green) T cells. (**e**) Impact of age and sex on the absolute count of MAIT cells. Females are represented in pink and men in blue. (**h**) Impact of age and CMV serostatus on $CD4^+$ EMRA T cells. CMV+ individuals are represented in red and CMV- in orange. (**c**) Flow cytometry plots of naive $CD8b^+$ and $CD4^+$ T cells for representative persons in their 20s and their 60s. (**f**) Flow cytometry plots of EMRA $CD4^+$ T cells in representative CMV- and CMV+ subjects. (**i**) Flow cytometry plots of MAIT cells in representative woman and man. The significant effects of age, sex and CMV seropositivity on MFI can be found in **Supplementary Fig. 17**.

Figure 3 Effects of smoking on innate and adaptive immune cell counts in 1,000 healthy individuals. (**a**) Levels of association (i.e., $-\log_{10}$(q-values)) between 39 non-genetic factors and adaptive and innate cell counts, at a false discovery rate (FDR) < 1%. Except when their effects were specifically measured, immunophenotypes were regressed on age, sex, CMV status, batch effects and genome-wide significant SNPs (**Table 1**). (**b**) Significant multiplicative effects (adjusted P<0.01) of active and past smoking on circulating levels of immune cells. The multiplicative effect sizes were estimated in a linear mixed model with a log-transformed immunophenotype as response, controlling for age, sex, CMV serostatus, batch effects and genome-wide significant SNPs, and then transformed to the original data scale. 99%CIs were false coverage-adjusted. Adaptive and innate immune cells are represented in grey and black, respectively. (**c**) Impact of age and smoking on the number of circulating T_{reg} cells. Brown indicates active smokers, orange indicates past smokers and yellow indicates non-smokers. Regression lines were fitted using local polynomial regression. (**d)** Flow cytometry plots of HLA-DR expression in T_{reg} cells of representative non-smoker and active smoker. The effect of smoking on MFI can be found in **Supplementary Fig. 19**.

Figure 4 Genome-wide significant associations with 166 immunophenotypes measured in 1,000 healthy individuals. (**a**) Manhattan plots of genome-wide significant associations with variants acting locally (local-pQTLs, in blue) or not (cell count QTLs or *trans*-pQTLs, in yellow) on immunophenotypes. The gray line indicates the genome-wide significance threshold (P<10^{-10}). Zoomed Manhattan plots for all hits are shown in **Supplementary Figure 21**. (**b**) Differential expression of the CD62L protein marker in granulocytes of representative individuals homozygous for the major (T/T, in dark colors) and minor (C/C, in light colors) rs2223286 alleles. (**c**) Cell-specific CD62L expression is shown for age-matched individuals homozygous for the major (open distribution with solid line) or minor (shaded distribution with dotted line) rs2223286 alleles. (**d**) Zoomed Manhattan plots of genetic associations between SNP rs2223286 in the *SELL* gene and cell-surface expression on CD62L in eosinophils or *SELL* mRNA levels in whole blood. Each point is a SNP,

whose color represents its level of linkage disequilibrium (r^2) with the best hit (in purple). Blue lines indicate local recombination rates.

Figure 5 Proportion of variance of innate and adaptive cell parameters explained by non-genetic and genetic factors. Flow cytometric measurements were separated into (**a**, **b**) 76 absolute counts and 2 count ratios of circulating immune cells and (**c**, **d**) 87 MFIs and a ratio of MFIs. The total variance R^2 of the 91 adaptive (**a**, **c**) and 75 innate (**b**, **d**) cell parameters was decomposed into proportions explained by intrinsic factors (age and sex; **Fig. 2**), environmental exposures (CMV infection and smoking; **Figs. 2** and **3**) and genetic factors (independent significant and suggestive GWAS hits, **Table 1** and **Supplementary Table 6**).

Locus	FACS panel	Immunophenotype	Other immunophenotypes[a]	*P*-value	Replication *P*-value[b]	*P*-value for biological replicates[c]	Identified by a previous study	Effect size (SE)	Chr	Position	Candidate variant	Effect allele[d]	Other allele	EAF[d]	Candidate gene	Distance to TSS (kb)
1	4	CD69 in $CD16^{hi}$ NK cells	$CD69^{+}$ $CD16^{hi}$ NK cells; CD69 in $CD8a^{+}$ and $CD69^{+}$ $CD16^{+}$ NK cells	4.8×10^{-37}	6.3×10^{-4}	2.0×10^{-16}	-	0.14 (0.01)	1	101744633	rs6693121	A	C	0.40	*S1PR1*	41.0
2	4	CD16 in $CD16^{hi}$ NK cells	CD16 in $CD56^{hi}$ NK cells; HLA-DR in $CD16^{hi}$, $CD8a^{+}$ $CD16^{+}$ and $CD69^{+}$ $CD16^{+}$ NK cells	3.0×10^{-87}	7.1×10^{-7}	2.6×10^{-41}	Orrù et al., *Cell* 2013	22.77 (1.04)	1	161507448	rs3845548	C	T	0.87	*FCGR3A*	12.4
3	7	CD32 in basophils	-	1.7×10^{-36}	3.6×10^{-7}	1.6×10^{-18}	-	11.23 (0.86)	1	161653737	rs61804205	C	T	0.10	*FCGR2B*	20.8
4	7	CD62L in eosinophils	CD62L in neutrophils	1.6×10^{-35}	3.7×10^{-2}	1.4×10^{-8}	-	542.78 (42.08)	1	169665632	rs2223286	C	T	0.33	*SELL*	0.0
5	4	CD8a in $CD69^{+}$ $CD16^{hi}$ NK cells	CD8a in CD16hi, $CD56^{hi}$, $CD69^{+}$ $CD56^{hi}$, $CD8^{+}$ $CD56^{hi}$, $CD8a^{+}$ $CD16^{hi}$ and HLA-DR^{+} $CD16^{hi}$ NK cells	5.9×10^{-58}	5.9×10^{-2}	3.4×10^{-24}	Orrù et al., *Cell* 2013	0.44 (0.03)	2	87026807	rs71411868	A	G	0.76	*CD8A*	0.0
6	4	Number of $CD8a^{+}$ $CD56^{hi}$ NK cells	$CD56^{hi}$ NK cells; CD69+ $CD56^{hi}$ NK cells; $CD56^{+}$ ILC	9.1×10^{-19}	2.7×10^{-2}	2.5×10^{-9}	-	1.57 (0.18)	2	111808558	rs12986962	A	G	0.62	*ACOXL / BCL2L11*	0.0
7	8	HLA-DR in cDC3	-	2.6×10^{-11}	-	3.1×10^{-10}	-	0.11 (0.02)	6	32340176	rs143655145	T	C	0.19	*HLA-DRA*	67.4

8	8	HLA-DR in cDC1	-	6.1×10^{-38}	-	1.3×10^{-17}	-	0.12 (0.01)	6	32574308	rs2760994	T	C	0.63	*HLA-DRB1*	16.7
9	8	HLA-DR in pDC	CD86 in pDC; HLA-DR^{+} CD56^{hi} NK cells; HLA-DR in CD14^{hi} monocytes	2.2×10^{-56}	-	2.7×10^{-26}	-	9.06 (0.54)	6	32599163	rs114973966	T	C	0.18	*HLA-DRB1*	41.5
10	6	CD24 in IgM^{+} marginal zone B cells	CD24 in B cells, and in naive, memory, double negative memory, IgM^{-} marginal zone and marginal zone B cells	3.8×10^{-21}	-	5.5×10^{-10}	-	0.20 (0.02)	6	107168676	rs12529793	C	T	0.92	*CD24*	254.7
11	7	CD203c in basophils	-	2.1×10^{-28}	3.2×10^{-2}	3.9×10^{-14}	-	8.83 (0.77)	6	132043056	rs2270089	G	A	0.09	*ENPP3*	0.0
12	1	CCR7 in CD4^{+} naive T cells	CCR7 in CD8b^{+} naive T cells	3.0×10^{-19}	-	2.0×10^{-7}	-	0.07 (0.01)	16	429129	rs11648403	C	T	0.57	*TMEM8A*	0.0
13	7	FCεRI in eosinophils	-	9.2×10^{-14}	5.1×10^{-5}	1.9×10^{-7}	-	0.96 (0.13)	17	4560141	rs56170457	G	T	0.75	*ALOX15*	25.9
14	4	Ratio of CD16 MFI in CD16^{hi} and CD56^{hi} NK cells	CD16 in CD56^{hi} NK cells	4.3×10^{-30}	2.4×10^{-2}	8.9×10^{-13}	Aguirre-Gamboa et al., *Cell Reports* 2016	0.39 (0.03)	19	8788184	rs114412914	G	A	0.85	*ACTL9*	21.0

Table 1 Genome-wide signals of association with immunophenotypes in the Milieu Intérieur cohort.

[a]Other immunophenotypes correspond to any measured immunophenotype in the Milieu Intérieur cohort that was also significantly associated with the candidate variant, but to a lesser extent than the main immunophenotype.

[b]Replication was performed in an independent cohort of 75 European-descent Americans. Only panels 4 and 7 could be used, due to sample limitations. Effects were in the same direction as in the primary cohort.

[c]*P*-values for biological replicates were estimated based on immunophenotypes measured from a new blood draw taken ~17 days after the initial visit, in 500 subjects of the Milieu Intérieur cohort.

[d]EAF is the frequency of the effect allele, which was defined as the allele with a positive effect on the immunophenotype.

Online Methods

A summary of the Online Methods can be found in the Life Sciences Reporting Summary.

The Milieu Intérieur cohort

The 1,000 healthy donors of the Milieu Intérieur cohort were recruited by BioTrial (Rennes, France), and included 500 women and 500 men, and 200 individuals from each decade of life, between 20 and 69 years of age. Donors were selected based on stringent inclusion and exclusion criteria, detailed elsewhere[18]. The clinical study was approved by the Comité de Protection des Personnes — Ouest 6 (Committee for the protection of persons) on June 13th, 2012 and by the French Agence Nationale de Sécurité du Médicament (ANSM) on June 22nd, 2012. The study is sponsored by the Institut Pasteur (Pasteur ID-RCB Number: 2012-A00238-35), and was conducted as a single center study without any investigational product. The protocol is registered under ClinicalTrials.gov (study# NCT01699893).

Human material and staining protocol

Whole blood samples were collected from the 1,000 healthy, fasting donors on Li-heparin, every working day from 8 to 11AM, from September 2012 to August 2013, in Rennes, France. Tracking procedures were established in order to ensure delivery to Institut Pasteur, Paris, within 6 hours of blood draw, at a temperature between 18°C and 25°C. To check the stability of our flow cytometry measures through time, a second blood sample was drawn for half of the cohort during a second visit, ~17 days on average after the first visit, ranging from 7 to 44 days. After receipt, samples were kept at room temperature prior to sample staining. Details on staining protocols can be found elsewhere[20].

Reproducibility testing and assay development

For optimization studies and panel development, whole blood samples were collected from healthy volunteers enrolled at the Institut Pasteur Platform for Clinical Investigation and Access to Research Bioresources (ICAReB) within the Diagmicoll cohort. The biobank activity of ICAReB platform is NF S96-900 certified. The Diagmicoll protocol was approved by the French Ethical Committee (CPP)

Ile-de-France I, and the related biospecimen collection was declared to the Research Ministry under the code N° DC 2008-68. The reproducibility tests were performed as detailed elsewhere[20].

Cytometric analyses

Ten 8-color flow cytometry panels were developed. Details on staining antibodies can be found in **Supplementary Table 2**. A unique lot of each antibody was used for the entire study. Each antibody was selected and titrated as described earlier[20]. Gating strategies are described in **Supplementary Figures 1-10**. The acquisition of cells was performed using two MACSQuant analyzers (Serial numbers 2420 & 2416), each fit with identical three lasers and ten detector optical racks (FSC, SSC and eight fluorochrome channels). Calibration of instruments was performed using MacsQuant calibration beads (Miltenyi, ref. 130-093-607). Flow cytometry data were generated using MACSQuantify™ software version 2.4.1229.1 and saved as .mqd files (Miltenyi). The files were converted to FCS compatible format and analyzed by FlowJo software version 9.5.3. A total of 313 immunophenotypes were exported from FlowJo. These included 110 cell proportions, 106 cell counts, 89 MFI and 8 ratios. We excluded from subsequent analyses all cell proportions, 35 immunophenotypes that were measured several times on different panels and were exported for quality controls, and two MFI that were measured with a problematic clone (**Supplementary Table 3**). A total of 166 flow cytometry measurements were thus analysed, including 76 cell counts, 87 MFI and 3 ratios (**Supplementary Table 3**). Problems in flow cytometry processing, such as abnormal lysis or staining, were systematically flagged by trained experimenters, which resulted in 8.70% missing data among the 166,000 measured values.

Outlier removal

Despite the exclusion of flagged problematic values, a limited number of outlier values were observed. As the goal of this study was to identify common non-genetic and genetic factors controlling immune cell levels, we removed these outlier values. Outliers were detected using a distance-based algorithm instead of a parametric method (e.g., removal based on a number of standard deviations from the mean), because of the substantial and highly variable skewness of the

distributions of flow cytometry measurements. A value in the higher tail was considered an outlier if the distance to the closest point in the direction of the mean of the distribution was more than 60% of the total range of the sample, while a value in the lower tail was considered an outlier if that distance was more than 15% of the total range of the sample. To choose these threshold values, we simulated 10,000 log-normal distributions with a skewness similar to that of the flow cytometry measurements. We then searched for threshold values so that simulated values outside of these ranges were observed in less than 5% of the distributions. Outliers were only looked for in the 50 highest and lowest values. This threshold was chosen to make sure that we do not miss any effect on immunophenotypes of common genetic variants (minor allele frequency>5%), or that of one of 39 continuous or common categorical non-genetic factors studied here. All values more extreme than the points labelled as outliers were also labelled outliers. A total of 24 values was removed at this stage.

Batch effects on flow cytometry measurements

Two batch effects on flow cytometry measurements were considered: the hour at which blood samples were drawn (from 8h to 11h in the morning) and the day at which samples were processed (8 to 12 samples per day, from September 2012 to august 2013). The hour of blood draw effect was evaluated with linear regression on all immunophenotypes. We observed that hour of blood draw impacts a limited number of cell counts, mainly $CD16^{hi}$ NK cells (**Supplementary Fig. 14a**). The sampling day effect was evaluated by estimating its variance component on all immunophenotypes. Visual inspection was used to determine whether temporal fluctuations – observed for those immunophenotypes with a large variance explained – were seasonal or not. We observed that sample processing day has a substantial impact on MFI. Fluctuations in MFI across time were strongly discontinuous, suggesting technical issues possibly related to the compensation matrix, rather than seasonal effects (**Supplementary Fig. 14b**).

Inclusion and imputation of candidate non-genetic factors

A large number of demographic variables were available in the Milieu Intérieur cohort[18]. These included infection and vaccination history, childhood diseases, health-related life habits, and socio-

demographic variables. Of these, 39 variables were chosen for subsequent analyses (**Supplementary Table 1**), based on the fact that they are intrinsic factors (i.e., age, sex), or measure the exposure of individuals to exogenous factors, and thus may not be affected by the immunophenotypes themselves. These variables were filtered based on their distribution (i.e., categorical variables with only rare levels, such as infrequent vaccines, were excluded) and on their levels of dependency with other variables (e.g., height and BMI). The dependency matrix among the 39 non-genetic variables, together with batch variables, was obtained based on the generalized R^2 measures for pairwise fitted generalized linear models. If the response was a continuous variable, we used a Gaussian linear model. If the response was binary, we used logistic regression. Categorical variables were used only as predictors. Missing values were imputed using the random forest-based R package *missForest*[57].

Impact of candidate non-genetic factors on immunophenotypes

To analyse the impact of non-genetic factors on immunophenotypes, we fitted a linear mixed model for each of the 166 immunophenotypes and each of the 39 non-genetic treatment variables. A total of 6,474 models were therefore fitted using the *lme4* R package[51]. All models were fitted to complete cases. Due to lack of a priori knowledge on how the non-genetic variables impact the immunophenotypes, we did not attempt to make a full causal structural equation model for all variables. Instead, we chose to keep the amount of controls in the models small to increase interpretability of the results, and to make the study easier to reproduce. We included age, sex and CMV seropositivity as fixed-effect controls for all models (**Fig. 3** and **Supplementary Fig. 19**), except when they were the treatment variable to be tested (**Fig. 2** and **Supplementary Figs. 17**). The intrinsic factors, i.e., age and sex, were included as covariates because they are known to have an impact on immunophenotypes[6,7,10,14,25–27], as well as on many of the other environmental exposures, and are therefore possible confounders. CMV seropositivity was included because it has been shown to strongly affect some immunophenotypes[6,13,14,17]. We also controlled for genome-wide significant SNPs for corresponding immunophenotypes (**Table 1**). Genetic variants were included to reduce the residual variance of the models and to make the inferences more robust. To correct for the batch effect related to the day of sample processing, we included it as a random effect for all models: we included

713 a constant for each day and assumed that all constants were drawn from the same normal distribution.
714 This procedure models correlation among subjects processed during the same day. We also included
715 the hour of blood draw as a fixed-effect control for all models.
716 The distributions of the immunophenotypes have variable skewness. We considered normal,
717 lognormal and negative binomial response distributions, and chose to model all immunophenotypes as
718 lognormal based on diagnostic plots, AIC measures and our aim to have comparable results across
719 immunophenotypes and facilitate the interpretation of effect sizes. A total of 46 immunophenotypes
720 had zero values. A unit value was added to those before log-transformation.
721 For each model, we tested the hypothesis that the regression parameter for the treatment variable
722 was zero by an F-test with the Kenward-Roger approximation. This test has better small- and
723 medium-sample properties than the traditional chi-square-based likelihood ratio test for mixed
724 models[52] and can readily be applied using the *pbkrtest* R package[53]. We assumed that our sample size
725 was large enough for this test to be appropriate and chose therefore not to do parametric
726 bootstrapping. We considered all 6,474 tests as one multiple testing family and we used the false
727 discovery rate (FDR) as error rate. An effect was considered significant if the adjusted *P*-value was
728 smaller than 0.01. If a test was significant, confidence intervals were constructed using the profile
729 likelihood method in such a way that the false coverage rate was controlled at a level of 0.01. The
730 false coverage rate measures the rate of confidence intervals that do not cover the true parameter and
731 is needed if confidence intervals are selected based on a criterion that makes these intervals especially
732 interesting, for instance significant hypothesis tests[54]. FCR-adjusted confidence intervals are always
733 wider than regular intervals. All these analyses were done, and can be reproduced, with the *mmi* R
734 package (http://github.com/JacobBergstedt/mmi).
735
736 **Genome-wide DNA genotyping**
737 The 1,000 subjects of the Milieu Intérieur cohort were genotyped at 719,665 SNPs by the
738 HumanOmniExpress-24 BeadChip (Illumina, California). SNP call rate was higher than 97% in all
739 donors. To increase coverage of rare and potentially functional variation, 966 of the 1,000 donors
740 were also genotyped at 245,766 exonic SNPs by the HumanExome-12 BeadChip (Illumina,

California). HumanExome SNP call rate was lower than 97% in 11 donors, which were thus removed from this dataset. We filtered out from both datasets SNPs that: (i) were unmapped on dbSNP138, (ii) were duplicated, (iii) had a low genotype clustering quality (GenTrain score < 0.35), (iv) had a call rate < 99%, (v) were monomorphic, (vi) were on sex chromosomes and (vii) were in Hardy-Weinberg disequilibrium (HWE $P < 10^{-7}$). These SNP quality-control filters yielded a total of 661,332 and 87,960 SNPs for the HumanOmniExpress and HumanExome BeadChips, respectively. The two datasets were then merged, after excluding triallelic SNPs, SNPs with discordant alleles between arrays (even after allele flipping), SNPs with discordant chromosomal position, and SNPs shared between arrays that presented a genotype concordance rate < 99%. Average concordance rate for the 16,753 SNPs shared between the two genotyping platforms was 99.9925%, and individual concordance rates ranged from 99.80% to 100%, validating that no problem occurred during DNA sample processing. The final dataset included 732,341 QC-filtered genotyped SNPs.

Genetic relatedness and structure

Possible pairs of genetically related subjects were detected using an estimate of the kinship coefficient and the proportion of SNPs that are not identical-by-state between all possible pairs of subjects, obtained with KING[55]. Genetic structure was visualized with the Principal Component Analysis (PCA) implemented in EIGENSTRAT[56]. For comparison purposes, the analysis was performed on 261,827 independent SNPs and 1,723 individuals, which include the 1,000 Milieu Intérieur subjects together with a selection of 723 individuals from 36 populations of North Africa, the Near East, western and northern Europe[57].

Genotype imputation

Prior to imputation, we phased the final SNP dataset with SHAPEIT2[58] using 500 conditioning haplotypes, 50 MCMC iterations, 10 burn-in and 10 pruning iterations. SNPs and allelic states were then aligned to the 1,000 Genomes Project imputation reference panel (Phase1 v3.2010/11/23). We removed SNPs that have the same position in our data and in the reference panel but incompatible alleles, even after allele flipping, and ambiguous SNPs that have C/G or A/T alleles. Genotype

imputation was performed by IMPUTE v.2[59], considering 1-Mb windows and a buffer region of 1 Mb. Out of the 37,895,612 SNPs obtained after imputation, 37,164,442 were imputed. We removed 26,005,463 imputed SNPs with information ≤ 0.8, 43,737 duplicated SNPs, 955 monomorphic SNPs, and 449,903 SNPs with missingness >5% (individual genotype probabilities < 0.8 were considered as missing data). After quality-control filters, a total of 11,395,554 high-quality SNPs were further filtered for minor allele frequencies >5%, yielding a final set of 5,699,237 SNPs for association analyses.

Genome-wide association analyses

Prior to the genome-wide association study, we transformed immunophenotypes using a different procedure than that used for the analysis of non-genetic factors. This is because we tested for association between immunophenotypes and millions of genetic variants, among which some have an unbalanced genotypic distribution (i.e., SNPs with a low minor allele frequency), which makes this analysis more sensitive to deviations from distributional assumptions. Our primary aim was therefore to use transformations that make the GWAS as robust as possible against such deviations. Also, we map loci associated with immunophenotypes based on *P*-values, so it was less important to keep effect sizes on the same scale, in contrast with the analysis of non-genetic factors, for which we favoured the interpretability of effect sizes. A unit value was first added to all phenotypes with zero values. The transformations were then chosen based on an AIC measure using the Jacobian-adjusted Gaussian likelihood, among three possible choices of increasing skewness: identity transformation, squareroot-tranformation and log-transformation. We kept the amount of possible transformations low to minimize the amount of added unmodelled stochasticity. The added unit value was kept only for immunophenotypes for which the log-transformation was chosen.

After transformation, a second round of outlier removal was done, to remove extreme values on the new scale. The thresholds for the lower and higher tail were 20%, obtained as for the first step of outlier removal (see description of the distance-based outlier removal algorithm above), but on the Gaussian scale. The immunophenotypes were then imputed using the *missForest* R package[57], as missing data is not allowed by the subsequent analyses. We finally adjusted all immunophenotypes

for the batch effect of processing days. We used the ComBat non-parametric empirical-Bayes framework[60], instead of the mixed model described above (see section "Impact of candidate non-genetic factors on immunophenotypes" above), because the GEMMA mixed model used to conduct GWAS (see below) includes only the random effect capturing genetic relatedness. ComBat adjusts for batch effects by leveraging multivariate correlations among response variables. We did not include variables of interest in the ComBat model (none of the non-genetic variables were significantly different across sample processing days, with the exception of smoking (regression P=0.002)).

To reduce the residual variance of GWAS models and make the inferences more robust[61], we sought to adjust models for covariates selected among 42 variables. These included the 39 non-genetic variables (**Supplementary Table 1**), the hour of blood draw variable, and the two first principal components of a PCA based on genetic data (**Supplementary Fig. 20b**). Covariates were selected by stability selection[62,63], with elastic net regression as the selection algorithm. A selection algorithm uses a cost function that drives regression parameters of non-predictive variables to zero, unlike least-square regressions. The elastic net method was used in particular because it has lower variance than stepwise methods and overcomes limitations of the LASSO method related to correlated variables[64]. To perform stability selection, we estimated, for each of the $i \in \{1, \ldots, 42\}$ variables, the probability $p_i = P(\beta_i = 0)$ that the elastic net regression parameter β_i of variable i equals zero. Specifically, we first took 50 subsamples of half of the data, performed variable selection on each subsample, and estimated p_i as the number of subsamples in which $\beta_i > 0$, divided by the total number of subsamples. The variables were then chosen to be controls in the GWAS models by thresholding the probability $\hat{p}_i$. It has been shown that this procedure, with the right threshold and under certain assumptions, controls the false discovery rate of selected variables[63]. The procedure is more stable than selecting variables by, for instance, stepwise regression or elastic net without stability selection, and thus adds less unmodelled variability to the estimates. Still, because this approach does select predictive variables for each individual response variable, it adds more variance to the model selection, relative to models in which only age, sex, CMV infection and smoking would be systematically included. However, controlling for the selected variables is expected to generate more parsimonious models (i.e., the inclusion of unnecessary covariates could reduce power[65]), and to decrease the risk of type 1 errors

(e.g., some of the many rare genetic variants that are tested could associate, by chance, with an immunophenotype when the model does not fulfil inference assumptions due to a specific, unmodelled covariate).

The univariate genome-wide association study was conducted for each imputed, transformed and batch-effect corrected immunophenotype using the linear mixed model implemented in GEMMA[66], adjusting on selected covariates. GEMMA is an efficient mixed model that controls for genetic relatedness among donors and allows for multivariate analyses. Genetic relatedness matrices (GRM) were estimated for each chromosome separately, using the 21 other chromosomes, to exclude from the GRM estimation potentially associated SNPs (*i.e.*, "leave-one-chromosome" approach; see [67]). A conditional GWA analysis was also carried out for each of the 14 immunophenotypes that showed the strongest genome-wide significant signals ("main immunophenotypes" in **Table 1**), by including as a covariate in GEMMA the genotypes of the most strongly associated variant. A multivariate GWAS was conducted on a set of 6 candidate immunophenotypes (*i.e.*, number of HLA-DR+ memory T cells), using GEMMA linear mixed model adjusted on covariates that were selected for at least one of the six traits. For all genome-wide association analyses, a conservative genome-wide significant threshold of $P<10^{-10}$ was used, to account for testing multiple SNPs and immunophenotypes.

Power estimation

We used simulations to estimate the minimum effect of a variant that we could detect with 95% power by our GWAS. Namely, we sampled 100,000 times a SNP in our data, and simulated an immunophenotype by adding to a randomly sampled immunophenotype the effect k of that SNP, k being drawn from a uniform distribution of bounds 0 and 1 (k is expressed in unit of phenotype standard deviations, as in 'scheme 1' of ref[68]). We then ran the GEMMA mixed model on the simulated data, and estimated the probability that the variant was detected, assuming our genome-wide significant threshold of $P<10^{-10}$. We found that we have 95% power to detect a SNP with a medium effect of 0.6 phenotype standard deviation. We also confirmed empirically the power to identify medium-effect genotype-phenotype associations in the *Milieu Intérieur* cohort by replicating well-known genetic associations with non-immune traits, including *OCA2-HERC2* genes with eye and

hair color (rs12913832, P=6.7x10^{-138} and 8.5x10^{-18}, respectively), *SLC45A2* with hair color (rs16891982, P=3.2x10^{-9}), *UGT1A* gene cluster with bilirubin levels (rs6742078, P=2.6x10^{-75}), *SLC2A9* with uric acid levels (rs6832439, P=4.3x10^{-14}), and *CETP* with HDL levels (rs711752, P=4.5x10^{-8}).

Enrichment in variants associated with diseases

We explored the implication of our 15 genome-wide significant variants in human diseases and traits using previously published hits of genome-wide association studies (GWAS), obtained from the 31/08/2017 version of the EBI-NHGRI GWAS Catalog. A candidate variant was considered as implicated in a disease/trait if it was previously associated with such a disease/trait with a P<5x10^{-8}, or if it was in linkage disequilibrium (LD) with a variant associated with such a disease/trait (r^2>0.6). We tested if our 15 genome-wide significant variants were enriched in known associations with diseases/traits by resampling. Namely, we sampled 100,000 times 15 random SNPs with minor allele frequencies matched to those observed, and we calculated for each resampled set the proportion of variants known to be, or in LD with a variant known to be, associated with a disease. The enrichment P-value was estimated as the proportion of resamples for which this proportion was larger than that observed in our set. LD was precomputed for all 5,699,237 SNPs with PLINK 1.9 (options '--show-tags all --tag-kb 500 --tag-r2 0.6')[69].

HLA typing and association tests

Four-digit classical alleles and variable amino acid positions in the HLA class I and II proteins were imputed with SNP2HLA v 1.03[70]. 104 HLA alleles and 738 amino acid residues (at 315 positions) with MAF >1% were included in the analysis. Conditional haplotype-based association tests were performed using PLINK v. 1.07[71], as well as multivariate omnibus tests used to test for association at multi-allelic amino acid positions.

Replication cohort

We recruited 75 donors through the Genentech Genotype and Phenotype (gGAP) Registry. This sample size provides 95% power to replicate SNPs with an effect > 0.9 phenotype standard deviation. Ethical agreement was obtained for all gGAP donors. Samples were received at room temperature and processed 1 h after blood draw. Prior to staining, the blood was washed with PBS 1X. Except for the CD32 antibodies, the antibodies for population identification were titrated using the same clones and providers as in the primary study (**Supplementary Table 2**). Cell labelling were performed manually in deep-well plates. Data acquisition was performed within one hour using a calibrated FacsCantoII (Becton Dickinson). We selected panels 4 and 7 for the replication study, because 10 of the 16 GWAS hits were identified with these panels, and because of sample limitations. Immunophenotypes were transformed based on models chosen in the primary cohort. The GEMMA linear mixed model was used to test for replication, with age and sex as covariates and a GRM estimated from 1,960,432 autosomal SNPs obtained by the Illumina HumanOmni1-Quad v1.0 array.

Gene expression assays

NanoString nCounter®, a hybridization-based multiplex assay, was used to measure gene expression in non-stimulated whole blood of the 1,000 Milieu Intérieur subjects, with the Human Immunology v2 Gene Expression CodeSet. This data is described in detail in a separate work[36]. Expression probes that bind to cDNAs in which at least 3 known common SNPs segregate in humans were removed from the analyses (*i.e.*, *HLA-DQB1*, *HLA-DQA1*, *HLA-DRB1*, *HLA-B* and *C8G*). After quality control filters, mRNA levels were available for 986 individuals at 90 candidate genes, i.e., immunity-related genes in a 1-Mb window around the genome-wide significant and suggestive associations identified in this study. For each sample, probe counts were $\log_2$ transformed, normalized and adjusted for batch effects. eQTL mapping was performed in a 1-Mb window around corresponding association signals, using the linear mixed model implemented in GenABEL[72]. All models were adjusted on the proportion of eight major cell populations, including neutrophils, $CD19^+$ B cells, $CD4^+$ T cells, $CD8^+$ T cells, $CD4^+CD8^+$ T cells, $CD4^-$ $CD8^-$ T cells, NK cells, and $CD14^+$ monocytes, to account for the effect of heterogeneous blood cell composition on gene expression.

Decomposition of the proportion of variance explained

We analysed each of the 166 batch-corrected and transformed immunophenotypes (see section "Genome-wide association analyses" of Online Methods) with a linear regression model including the four most impactful non-genetic factors (**Fig. 2**), i.e., age, sex, CMV seropositivity status and smoking, and both genome-wide significant ($P<10^{-10}$) and suggestive ($P<5x10^{-8}$) genetic factors. The contribution of each of these variables to the variance of each immunophenotype was calculated by averaging over the sums of squares in all orderings of the variables in the linear model, using the *lmg* metric in the *relaimpo* R package[73].The averaging over orderings was done to avoid bias due to correlations among predictors.

The difference in contribution to explained variance between innate and adaptive immunophenotypes was tested using linear mixed models, where we used the log-transformed proportions of variance of each immunophenotype explained by age, sex, CMV serostatus, smoking or genetics as different response variables, and indicator variables for the immunophenotype being innate or adaptive, and being a count or an MFI. The sum of the individual contributions of associated genetic variants was used to estimate the overall contribution of genetics. Since some of the immunophenotypes are correlated, their proportion of variance explained are also correlated. To account for this, we included a random effect term whose covariance matrix was modelled as a variance component multiplied by the sample correlation matrix among the immunophenotypes. Due to the small sample size, hypothesis testing was done by building a null distribution of likelihood ratios using the parametric bootstrap. The models were fitted using the R package *lme4qtl* (http://github.com/variani/lme4qtl). Because the distribution of variance explained by genetics was zero-inflated, we also tested for differences in the proportion of variance explained by non-genetic and genetic factors between innate and adaptive cell measurements with a non-parametric Mann-Whitney U test. Because the Mann-Whitney U test cannot account for correlations among immune cell measurements, we conducted this test on a subset of immunophenotypes that were selected to be uncorrelated ($h<0.6$ with the *protoclust* R package). Fifty immunophenotypes were kept, including 19 adaptive and 31 innate cell measures, among which the median Pearson's r was 0.039.

References

51. Bates, D., Maechler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67,** 1–48 (2015).

52. Kenward, M. G. & Roger, J. H. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* **53,** 983–997 (1997).

53. Halekoh, U. & Højsgaard, S. A Kenward-Roger Approximation and Parametric Bootstrap Methods for Tests in Linear Mixed Models - The R Package pbkrtest. *J. Stat. Softw.* **59,** 1–30 (2014).

54. Benjamini, Y. & Yekutieli, D. False discovery rate–adjusted multiple confidence intervals for selected parameters. *J. Am. Stat. Assoc.* **1000,** 71–93 (2005).

55. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics* **26,** 2867–2873 (2010).

56. Patterson, N., Price, A. L. & Reich, D. Population Structure and Eigenanalysis. *PLoS Genet.* **2,** e190 (2006).

57. Behar, D. M. *et al.* The genome-wide structure of the Jewish people. *Nature* **466,** 238–242 (2010).

58. Delaneau, O., Zagury, J.-F. & Marchini, J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat. Methods* **10,** 5–6 (2013).

59. Howie, B. N., Donnelly, P. & Marchini, J. A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies. *PLoS Genet.* **5,** e1000529 (2009).

60. Johnson, W. E. & Li, C. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8,** 118–127 (2007).

61. Mefford, J. & Witte, J. S. The Covariate's Dilemma. *PLoS Genet.* **8,** 1–2 (2012).

62. Meinshausen, N. & Bühlmann, P. Stability selection. *J. R. Stat. Soc. Ser. B* **72,** 417–473 (2010).

63. Shah, R. D. & Samworth, R. J. Variable selection with error control : another look at stability selection. *J. R. Stat. Soc. Ser. B* **75,** 55–80 (2013).

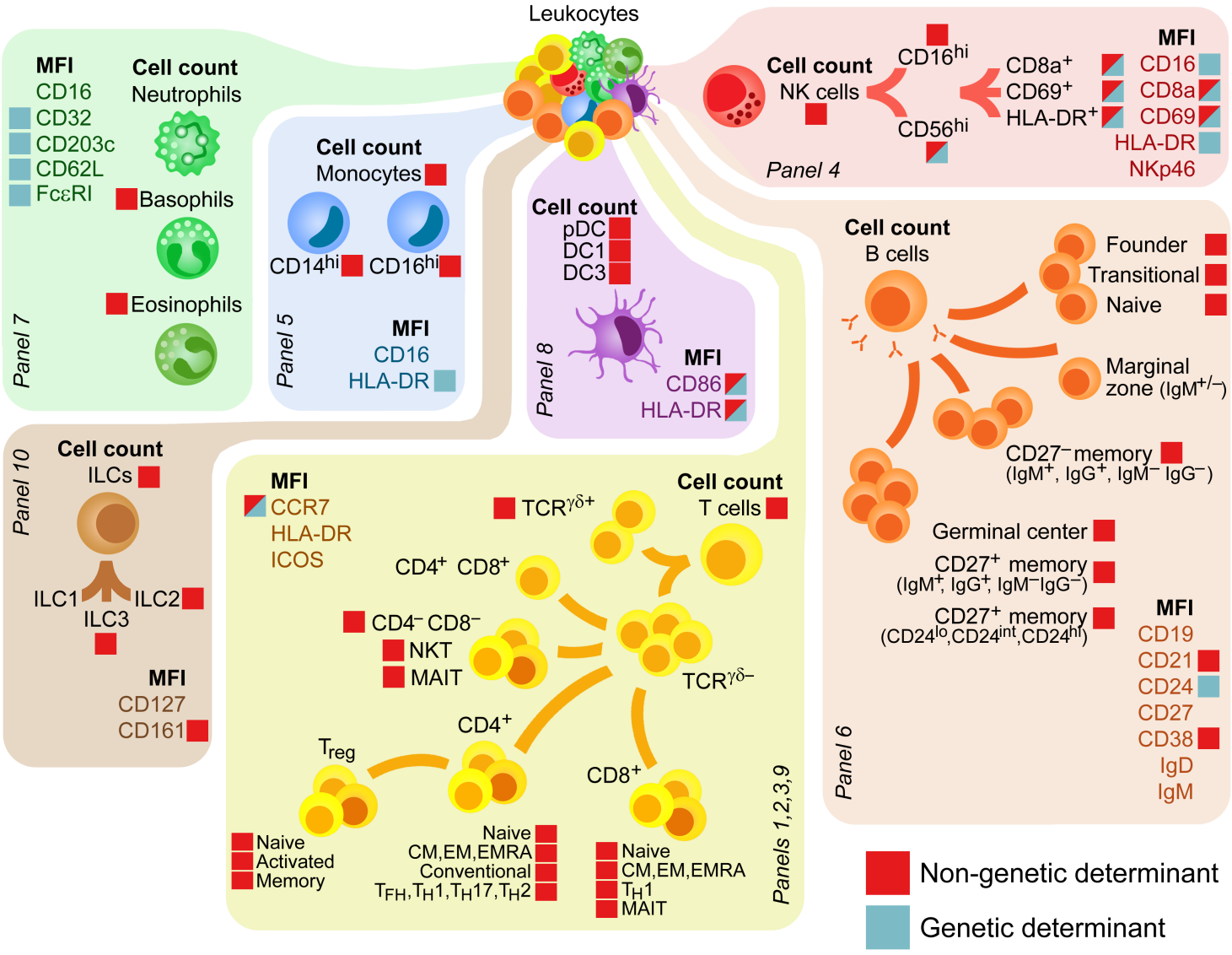
64. Hastie, T., Tibshirani, R. & Friedman, J. *Elements of statistical learning*. (Springer, 2009).

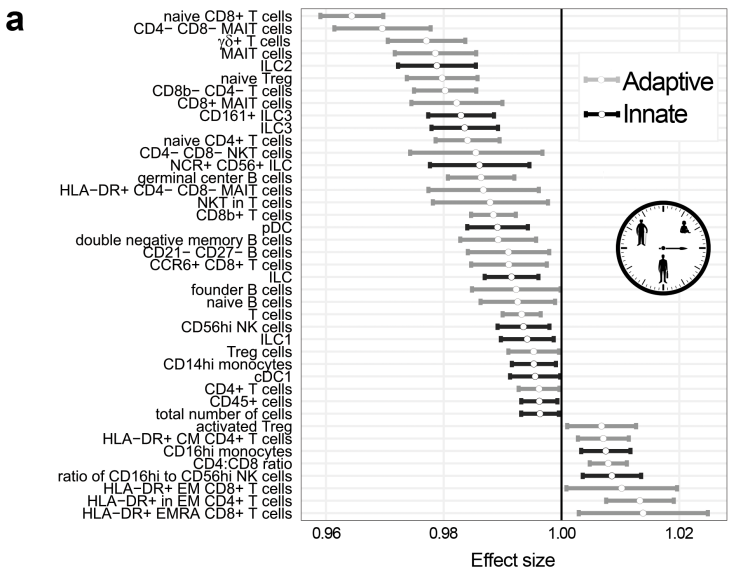
65. Wakefield, J. *Bayesian and Frequentist Regression Methods*. (Springer, 2013).

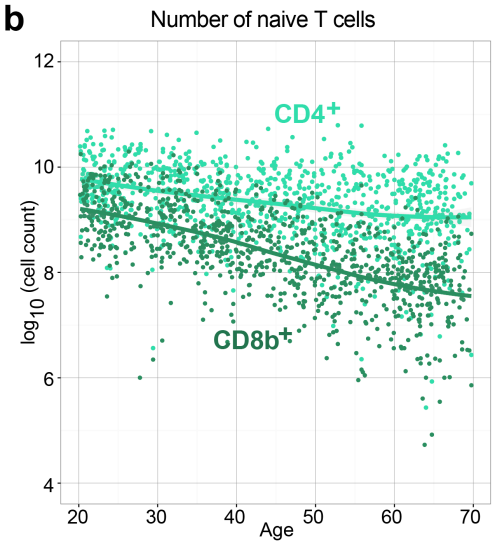
66. Zhou, X. & Stephens, M. Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nat. Methods* **11,** 407–409 (2014).

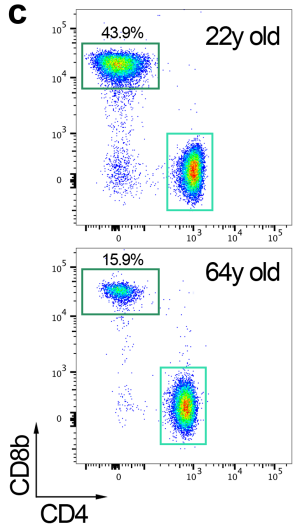
67. Yang, J., Zaitlen, N. A., Goddard, M. E., Visscher, P. M. & Price, A. L. Advantages and pitfalls in the application of mixed-model association methods. *Nat. Genet.* **46,** 100–6 (2014).

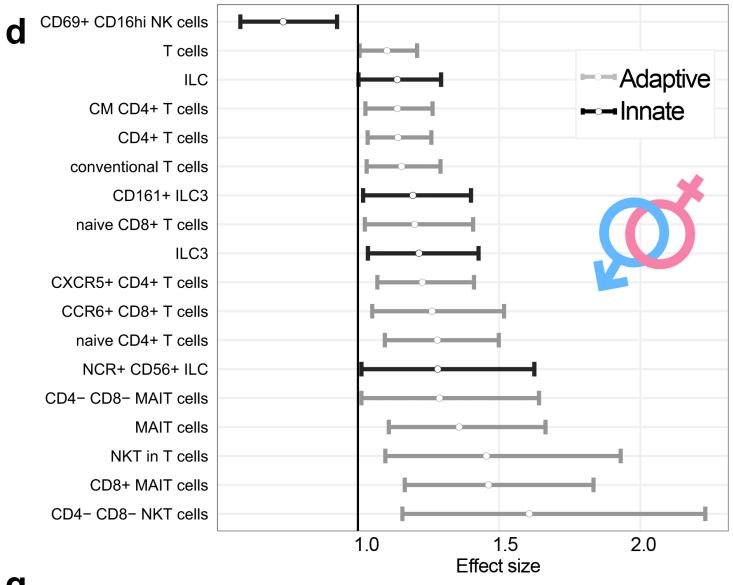
966 68. Zhang, Z. *et al.* Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.*
967 **42,** 355–360 (2010).
968 69. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets.
969 *Gigascience* **4,** 1–16 (2015).
970 70. Jia, X. *et al.* Imputing Amino Acid Polymorphisms in Human Leukocyte Antigens. *PLoS One* **8,** e64683
971 (2013).
972 71. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage
973 Analyses. *Am. J. Hum. Genet.* **81,** 559–575 (2007).
974 72. Aulchenko, Y. S., Ripke, S., Isaacs, A. & van Duijn, C. M. GenABEL : an R library for genome-wide
975 association analysis. *Bioinformatics* **23,** 1294–1296 (2007).
976 73. Grömping, U. Relative Importance for Linear Regression in R: The Package relaimpo. *J. Stat. Softw.* **17,**
977 1–27 (2006).
978

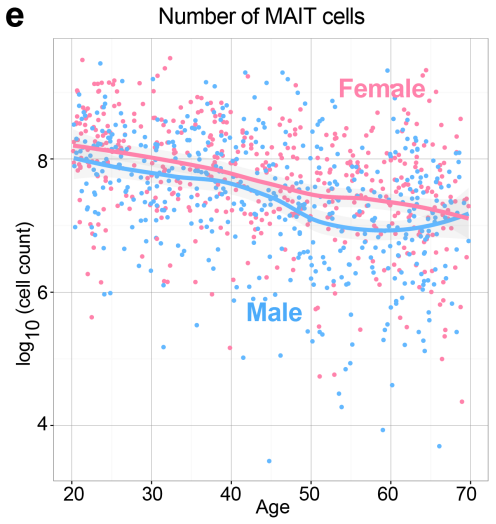

Leukocytes
Panel 7
MFI
CD16
CD32
CD203c
CD62L
FcεRI
Cell count
Neutrophils
Basophils
Eosinophils
Panel 5
Cell count
Monocytes
CD14hi
CD16hi
MFI
CD16
HLA-DR
Panel 8
Cell count
pDC
DC1
DC3
MFI
CD86
HLA-DR
Panel 4
Cell count
NK cells
CD16hi
CD56hi
CD8a+
CD69+
HLA-DR+
MFI
CD16
CD8a
CD69
HLA-DR
NKp46
Panel 6
Cell count
B cells
Founder
Transitional
Naive
Marginal zone (IgM+/−)
CD27− memory (IgM+, IgG+, IgM− IgG−)
Germinal center
CD27+ memory (IgM+, IgG+, IgM−IgG−)
CD27+ memory (CD24lo,CD24int,CD24hi)
MFI
CD19
CD21
CD24
CD27
CD38
IgD
IgM
Panel 10
Cell count
ILCs
ILC1
ILC2
ILC3
MFI
CD127
CD161
Panels 1,2,3,9
MFI
CCR7
HLA-DR
ICOS
TCRγδ+
Cell count
T cells
CD4+ CD8+
CD4− CD8−
NKT
MAIT
TCRγδ−
CD4+
Treg
CD8+
Naive
Activated
Memory
Naive
CM,EM,EMRA
Conventional
TFH,TH1,TH17,TH2
Naive
CM,EM,EMRA
TH1
MAIT
Non-genetic determinant
Genetic determinant

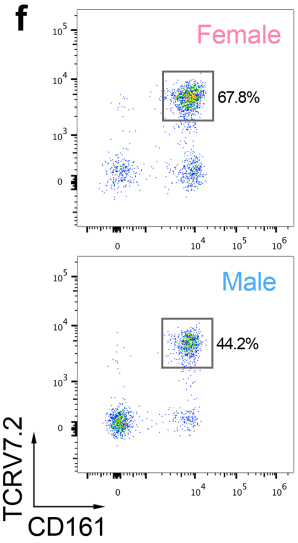

a
naive CD8+ T cells
CD4− CD8− MAIT cells
γδ+ T cells
MAIT cells
ILC2
naive Treg
CD8b− CD4− T cells
CD8+ MAIT cells
CD161+ ILC3
ILC3
naive CD4+ T cells
CD4− CD8− NKT cells
NCR+ CD56+ ILC
germinal center B cells
HLA−DR+ CD4− CD8− MAIT cells
NKT in T cells
CD8b+ T cells
pDC
double negative memory B cells
CD21− CD27− B cells
CCR6+ CD8+ T cells
ILC
founder B cells
naive B cells
T cells
CD56hi NK cells
ILC1
Treg cells
CD14hi monocytes
cDC1
CD4+ T cells
CD45+ cells
total number of cells
activated Treg
HLA−DR+ CM CD4+ T cells
CD16hi monocytes
CD4:CD8 ratio
ratio of CD16hi to CD56hi NK cells
HLA−DR+ EM CD8+ T cells
HLA−DR+ in EM CD4+ T cells
HLA−DR+ EMRA CD8+ T cells
Adaptive
Innate
0.96
0.98
1.00
1.02
Effect size

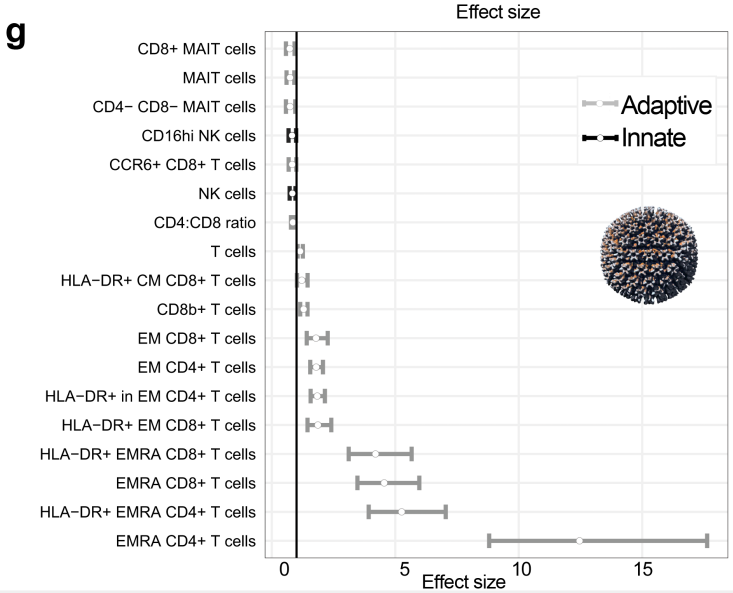

b
Number of naive T cells
CD4+
CD8b+
log10 (cell count)
12
10
8
6
4
20
30
40
50
60
70
Age

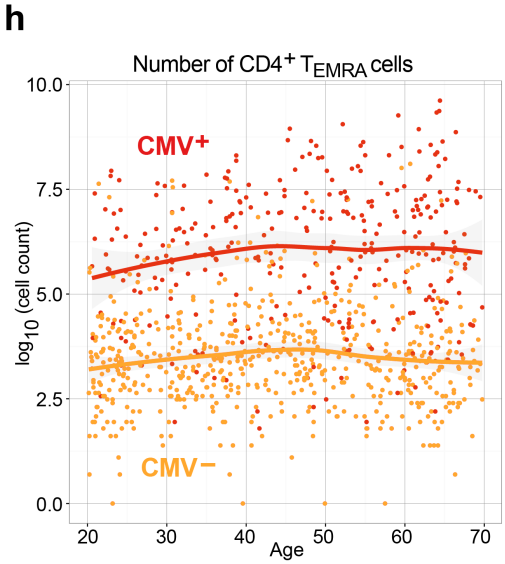

c
43.9%
22y old
15.9%
64y old
CD8b
CD4

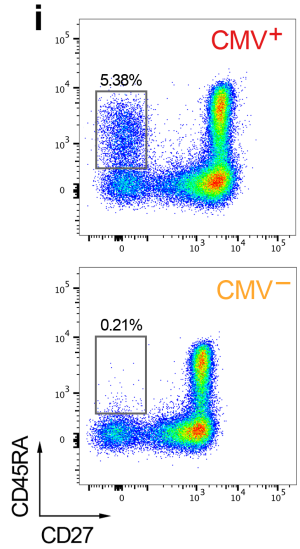

d
CD69+ CD16hi NK cells
T cells
ILC
CM CD4+ T cells
CD4+ T cells
conventional T cells
CD161+ ILC3
naive CD8+ T cells
ILC3
CXCR5+ CD4+ T cells
CCR6+ CD8+ T cells
naive CD4+ T cells
NCR+ CD56+ ILC
CD4− CD8− MAIT cells
MAIT cells
NKT in T cells
CD8+ MAIT cells
CD4− CD8− NKT cells
Adaptive
Innate
1.0
1.5
2.0
Effect size


e
Number of MAIT cells
Female
Male
log10 (cell count)
8
6
4
20
30
40
50
60
70
Age

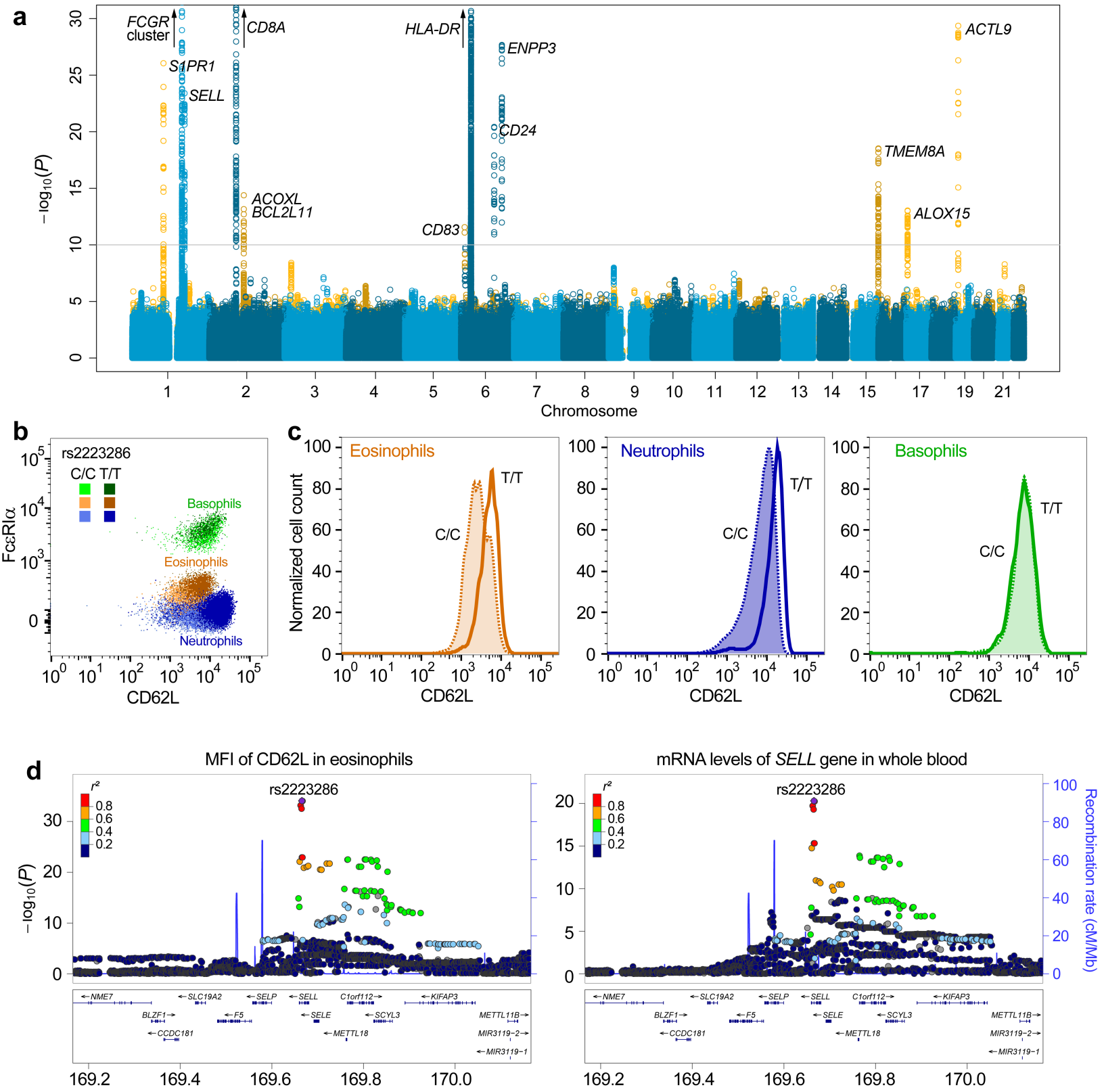

f
Female
67.8%
Male
44.2%
TCRV7.2
CD161


g
CD8+ MAIT cells
MAIT cells
CD4− CD8− MAIT cells
CD16hi NK cells
CCR6+ CD8+ T cells
NK cells
CD4:CD8 ratio
T cells
HLA−DR+ CM CD8+ T cells
CD8b+ T cells
EM CD8+ T cells
EM CD4+ T cells
HLA−DR+ in EM CD4+ T cells
HLA−DR+ EM CD8+ T cells
HLA−DR+ EMRA CD8+ T cells
EMRA CD8+ T cells
HLA−DR+ EMRA CD4+ T cells
EMRA CD4+ T cells
Adaptive
Innate
0
5
10
15
Effect size


h
Number of CD4+ TEMRA cells
CMV+
CMV−
log10 (cell count)
10.0
7.5
5.0
2.5
0.0
20
30
40
50
60
70
Age


i
CMV+
5.38%
CMV−
0.21%
CD45RA
CD27


a
Adaptive cell counts
Innate cell counts
FDR
(0,2]
(2,4]
(4,7]
(7,12]
(12,20]
(20,36]
(36,63]
(63,112]
CD8+ MAIT cells
T cells
MAIT cells
CD4− CD8− MAIT cells
CCR6+ CD8+ T cells
germinal center B cells
CD4+ T cells
Treg cells
NKT in T cells
naive CD8+ T cells
naive CD4+ T cells
naive B cells
HLA−DR+ in EM CD4+ T cells
HLA−DR+ EMRA CD8+ T cells
HLA−DR+ EM CD8+ T cells
HLA−DR+ CM CD4+ T cells
double negative memory B cells
CXCR5+ CD4+ T cells
conventional T cells
CM CD4+ T cells
CD8b+ T cells
CD4:CD8 ratio
CD4− CD8− NKT cells
CCR6+ CD4+ T cells
activated Treg
naive Treg
memory Treg
memory B cells
HLA−DR+ EMRA CD4+ T cells
HLA−DR+ CM CD8+ T cells
HLA−DR+ CD4− CD8− MAIT cells
γδ+ T cells
founder B cells
EMRA CD8+ T cells
EMRA CD4+ T cells
EM CD8+ T cells
EM CD4+ T cells
CRTh2+ CD4+ cells
CD8b− CD4− T cells
CD24low memory B cells
CD24hi memory B cells
CD21− CD27+ B cells
CD21− CD27− B cells
B cells
total number of cells
NCR+ CD56+ ILC
ILC3
ILC
CD45+ cells
pDC
CD56hi NK cells
CD161+ ILC3
CD14hi monocytes
ratio of CD16hi to CD56hi NK cells
NK cells
NCR− CD56+ ILC
monocytes
ILC2
ILC1
HLA−DR+ CD56hi NK cells
cDC1
CD69+ CD16hi NK cells
CD56+ ILC
CD16hi NK cells
CD16hi monocytes
basophils
Non-genetic variables
Age
Smoking
Sex
CMV+
Heart rate
Concentration
Owns a house
Metabolic score
BMI
Yellow fever vaccine
Whooping cough vaccine
Typhoid vaccine
MMR vaccine
HBV vaccine
HAV vaccine
Flu vaccine
Cannabis
Sleep problems
Temperature
Recent crisis
Physical activity
Low appetite
Lives with partner
Lives with kids
Listless
Income
Hours of sleep
Tonsillectomy
Rubella
Had mumps
Measles
Chicken pox
Appendicectomy
Flu IgG+
Employed
Education
Exposure to dust
Depression score
Born in a city
b
Past smoker
Active smoker
Adaptive
Innate
CD8+ MAIT cells
CD4− CD8− MAIT cells
MAIT cells
NCR− CD56+ ILC
CCR6+ CD8+ T cells
CD56+ ILC
HLA−DR+ CD56hi NK cells
ILC
CD56hi NK cells
CRTh2+ CD4+ cells
monocytes
basophils
CD14hi monocytes
T cells
pDC
B cells
CD45+ cells
total number of cells
CD4+ T cells
CXCR5+ CD4+ T cells
CCR6+ CD4+ T cells
double negative memory B cells
conventional T cells
CM CD4+ T cells
NCR+ CD56+ ILC
HLA−DR+ CM CD4+ T cells
CD24hi memory B cells
Treg cells
germinal center B cells
activated Treg
memory Treg
memory B cells
CD21− CD27+ B cells
CD24low memory B cells
0.5
1.0
1.5
2.0
2.5
Effect size
c
Number of Treg cells
log10 (cell count)
9
8
7
6
20
30
40
50
60
70
Age
Active smoker
Past smoker
Non smoker
d
HLA-DR in Treg cells
Active smoker
25.1%
Non smoker
10.5%


a
FCGR cluster
S1PR1
SELL
CD8A
ACOXL
BCL2L11
HLA-DR
CD83
ENPP3
CD24
TMEM8A
ALOX15
ACTL9
−log10(P)
Chromosome
b
rs2223286
C/C T/T
Basophils
Eosinophils
Neutrophils
FcεRIα
CD62L
c
Eosinophils
Neutrophils
Basophils
C/C
T/T
Normalized cell count
CD62L
d
MFI of CD62L in eosinophils
mRNA levels of SELL gene in whole blood
rs2223286
r²
0.8
0.6
0.4
0.2
−log10(P)
Recombination rate (cM/Mb)
NME7
SLC19A2
SELP
SELL
C1orf112
KIFAP3
BLZF1
F5
SELE
SCYL3
METTL11B
CCDC181
METTL18
MIR3119-2
MIR3119-1
169.2
169.4
169.6
169.8
170.0


a
Adaptive cells
b
Innate cells
Absolute cell counts
Genetics
CMV+
Age
Smoking
Sex
CD4:CD8 ratio
HLA–DR+ CM CD8+ T cells
HLA–DR+ CM CD4+ T cells
naive Treg
activated Treg
naive CD4+ T cells
EM CD4+ T cells
germinal center B cells
CD8b– CD4– T cells
HLA–DR+ EMRA CD8+ T cells
HLA–DR+ EMRA CD4+ T cells
HLA–DR+ in EM CD4+ T cells
EMRA CD4+ T cells
EMRA CD8+ T cells
HLA–DR+ EM CD8+ T cells
EM CD8+ T cells
CD8b+ T cells
T cells
MAIT cells
CD4– CD8– MAIT cells
CD8+ MAIT cells
naive CD8+ T cells
γδ+ T cells
double negative memory B cells
HLA–DR+ CD4– CD8– MAIT cells
CD21– CD27– B cells
CCR6+ CD8+ T cells
naive B cells
CD24low memory B cells
Treg cells
memory Treg
CD21– CD27+ B cells
memory B cells
CM CD4+ T cells
conventional T cells
CD4+ T cells
B cells
CXCR5+ CD4+ T cells
CD24hi memory B cells
CCR6+ CD4+ T cells
CRTh2+ CD4+ cells
NKT in T cells
CD4– CD8– NKT cells
transitional B cells
IgM+ marginal zone B cells
IgM– marginal zone B cells
marginal zone B cells
founder B cells
HLA–DR+ CD4– CD8– NKT cells
CM CD8+ T cells
CD69+ CD16hi NK cells
CD8a+ CD56hi NK cells
cDC3
CD56hi NK cells
HLA–DR+ CD56hi NK cells
CD69+ CD56hi NK cells
CD56+ ILC
NCR– CD56+ ILC
CD8a+ CD16hi NK cells
pDC
NK cells
ILC
ratio of CD16hi to CD56hi NK cells
CD16hi NK cells
CD161+ ILC3
ILC2
ILC3
CD16hi monocytes
NCR+ CD56+ ILC
ILC1
CD45+ cells
total number of cells
CD14hi monocytes
neutrophils
monocytes
basophils
cDC1
eosinophils
10%
20%
30%
40%
c
d
Mean fluorescence intensities
CCR7 in naive CD4+ T cells
CD24 in IgM+ marginal zone B cells
CD24 in double negative memory B cells
CD24 in memory B cells
CD24 in IgM– marginal zone B cells
CD24 in marginal zone B cells
CD21 in transitional B cells
CD24 in naive B Cells
CD24 in B cells
CCR7 in naive CD8+ T cells
ICOS in activated Treg
CD21 in B cells
CD21 in memory B cells
CD21 in CD24int memory B cells
CD21 in IgM– marginal zone B cells
CCR7 in CM CD8+ T cells
CD21 in marginal zone B cells
CCR7 in CM CD4+ T cells
CD21 in CD24hi memory B cells
CD27 in B cells
CD21 in double negative memory B cells
CD21 in naive B cells
CD38 in memory B cells
CD38 in B cells
IgM in B cells
CD38 in double negative memory B cells
CD21 in germinal center B cells
CD21 in CD24low memory B cells
CD24 in germinal center B cells
ICOS in memory Treg
CD38 in germinal center B cells
CCR7 in EMRA CD4+ T cells
IgD in B cells
CD21 in IgM+ marginal zone B cells
CD19 in B cells
ICOS in naive Treg
CCR7 in EMRA CD8+ T cells
CCR7 in EM CD8+ T cells
CCR7 in EM CD4+ T cells
CD16 in CD16hi NK cells
HLA–DR in CD8a+ CD16+ NK cells
HLA–DR in CD16hi NK cells
CD8a in CD16hi NK cells
HLA–DR in pDC
CD16 in CD56hi NK cells
CD8a in CD69+ CD16+ NK cells
HLA–DR in CD69+ CD16+ NK cells
HLA–DR in cDC1
CD69 in CD8a+ CD16+ NK cells
CD8a in CD8a+ CD16+ NK cells
CD69 in CD16hi NK cells
HLA–DR in CD14hi monocytes
CD16 MFI in CD16hi to CD56hi NK cells
CD32 in basophils
CD62L in eosinophils
CD203c in basophils
CD8a in CD56hi NK cells
CD69 in CD69+ CD16+ NK cells
CD8a in HLA–DR+ CD16+ NK cells
CD8a in CD69+ CD56+ NK cells
CD8a in CD8a+ CD56+ NK cells
CD86 in CD14hi monocytes
HLA–DR in CD8a+ CD56+ NK cells
HLA–DR in CD56hi NK cells
CD86 in pDC
CD62L in neutrophils
HLA–DR in cDC3
FcεRI in eosinophils
HLA–DR in HLA–DR+ CD16+ NK cells
CD16 in neutrophils
CD16 in CD16hi monocytes
CD69 in CD8a+ CD56+ NK cells
CD69 in CD69+ CD56+ NK cells
CD161 in ILC1
CD16 in CD14hi monocytes
CD69 in CD56hi NK cells
CD161 in ILC3
CD127 in ILC1
CD86 in cDC3
CD86 in cDC1
FcεRI in neutrophils
CD16 in eosinophils
FcεRI in basophils
CD16 in basophils
HLA–DR in HLA–DR+ CD56+ NK cells
HLA–DR in CD69+ CD56+ NK cells
CD69 in HLA–DR+ CD16+ NK cells
NKp46 in NK cells
10%
20%
30%
40%
Proportion of variance explained
Proportion of variance explained