



**HAL**  
open science

## 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.

### ► To cite this version:

Etienne Patin, Milena Hasan, Jacob Bergstedt, Vincent Rouilly, Valentina Libri, et al.. Natural variation in the parameters of innate immune cells is preferentially driven by genetic factors. *Nature Immunology*, 2018, 19 (3), pp.302 - 314. 10.1038/s41590-018-0049-7 . pasteur-01768943

**HAL Id: pasteur-01768943**

**<https://pasteur.hal.science/pasteur-01768943>**

Submitted on 13 May 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Natural variation in innate immune cell parameters

### is preferentially driven by genetic factors

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

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

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

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

1 **Abstract**

2       The enumeration and characterization of circulating immune cells provide key indicators of human  
3 health and disease. To identify the relative impact that environmental and genetic factors have on  
4 variation of innate and adaptive immune cell parameters in homeostatic conditions, we combined  
5 standardized flow cytometric analysis of blood leukocytes and genome-wide DNA genotyping in  
6 1,000 healthy, unrelated individuals of western European ancestry. We show that smoking, together  
7 with age, sex and latent cytomegalovirus infection, are the main non-genetic factors affecting human  
8 variation in immune cell parameters. Genome-wide association studies of 166 immunophenotypes  
9 identified 15 loci that are enriched in disease-associated variants. Finally, we demonstrate that innate  
10 cell parameters are more strongly controlled by genetic variation than adaptive cell parameters, which  
11 are primarily driven by environmental exposures. Our data establish a resource that generates new  
12 hypotheses in immunology and highlight the role of innate immunity in the susceptibility to common  
13 autoimmune diseases.

14

## 15 **Introduction**

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

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

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

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

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

64

65 **Results**

66 **Variation of immune cell parameters in the general population**

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

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

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

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

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

111

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

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

121 inflammation<sup>14,28,29</sup>. We found a modest increase in the number of memory T cells with age,  
122 supporting the view that the observed expansion of these cell populations in elderly subjects is not due  
123 to aging per se, but to CMV seropositivity<sup>13</sup>, which we accounted for in the model. Our analyses also  
124 revealed that naive CD8<sup>+</sup> T cells decrease more than twice as rapidly with age as compared to naive  
125 CD4<sup>+</sup> T cells, at a rate of 3.6 % (99% FCR-adjusted Confidence Interval (99%CI): [3.0%, 4.1%]) and  
126 1.6 % (99%CI: [1.1%, 2.1%]) per year, respectively (**Fig. 2a-c**), supporting the view that CD8<sup>+</sup> T cells  
127 are more susceptible to concentrations of homeostatic cytokines and/or that the production of CD4<sup>+</sup> T  
128 cells is preferentially enhanced in the human thymus<sup>30</sup>.

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

138 Environmental exposures are also known to drive immune variation, among which persistent  
139 CMV infection is one of the strongest candidates<sup>6,13,14,17</sup>. We observed a significant effect of latent  
140 CMV infection on 13% of immune cell parameters (**Fig. 2g** and **Supplementary Fig. 17c**), of which  
141 more than 75% were measured in adaptive cells. We confirm that CMV triggers a major change in the  
142 number of memory T cells, which is independent from age effects<sup>13,17</sup>. In particular, CMV  
143 seropositivity associated with a 12.5-fold (99%CI: [8.8, 17.6]) higher number of CD4<sup>+</sup> effector  
144 memory RA T cells ( $T_{EMRA}$ ), and a 4.6-fold (99%CI: [3.5, 6.0]) higher number of CD8<sup>+</sup>  $T_{EMRA}$  cells  
145 (**Fig. 2g-i**). However, we did not find evidence that CMV infection impacts the number of naive or  
146 central memory ( $T_{CM}$ ) T cell compartments. Supporting this observation, the total number of CD8<sup>+</sup>  
147 and CD4<sup>+</sup> T cells increased in parallel with the expanded number of memory T cells, thus suggesting  
148 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<sup>+</sup> 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<sup>31</sup>. 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<sub>reg</sub> 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,  $\gamma\delta$  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<sup>32</sup>, in turn engaging the IL-33 receptor,  
174 ST2, on both innate and non-classical lymphocytes<sup>33</sup>. 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.

177 **Genome-wide association study of 166 immune cell parameters**

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

188 With respect to immune traits, we found 14 independent genetic loci associated with 42 out of  
189 166 immunophenotypes (25%), at a conservative genome-wide significant threshold of  $P < 1.0 \times 10^{-10}$   
190 (**Fig. 4a**, **Table 1**, **Supplementary Fig. 21**, **Supplementary Tables 4** and **5**). We then conducted  
191 conditional GWAS, by adjusting these 42 immunophenotypes on the 14 leading associated variants  
192 (**Table 1**), and found an additional independent locus reaching genome-wide significance  
193 (**Supplementary Fig. 22** and **Supplementary Table 6**). Genome-wide significant associations were  
194 replicated in an independent cohort of 75 European-descent donors, for all immune traits measured in  
195 this replication cohort ( $P < 0.05$ ; **Table 1**; **Online Methods**). Also, we confirmed that our immune cell  
196 measurements were stable, as all genome-wide significant associations were confirmed for  
197 immunophenotypes measured in a new blood draw taken in 500 of the 1,000 subjects of the Milieu  
198 Intérieur cohort, sampled 7 to 44 days after the initial visit ( $P < 10^{-3}$ ; **Table 1**). We also provide a list of  
199 26 suggestive association signals ( $P < 5.0 \times 10^{-8}$ ), including a number of biologically relevant candidate  
200 genes (**Supplementary Table 6**). The associated genetic loci were enriched in SNPs associated by  
201 GWAS with diseases (31% observed vs. 5% expected, resampling  $P = 0.0032$ ), most of which were  
202 autoimmune diseases, including rheumatoid arthritis, Vogt-Koyanagi-Harada syndrome and atopic

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

### 205 **Genetic associations primarily identify immune cell-specific protein QTLs**

206 Of the 42 immunophenotypes for which a significant genetic association was detected, 36 (86%) were  
207 MFI, which measures the cell-specific expression of protein markers conventionally used to determine  
208 the differentiation or activation state of leukocytes. For 28 of these 36 MFI measurements (78%), the  
209 genetic association was observed between the protein MFI and SNPs located in the vicinity of the  
210 gene encoding the corresponding protein (**Table 1** and **Supplementary Fig. 21**), i.e., local protein  
211 QTLs (local-pQTLs). For instance, genetic variation close to the *ENPP3* gene was associated with  
212 CD203c MFI in basophils (rs2270089,  $P=2.1 \times 10^{-28}$ ), *CD24* with CD24 MFI in marginal zone B cells  
213 (rs12529793,  $P=3.8 \times 10^{-21}$ ) and *CD8A* with CD8a MFI in CD69<sup>+</sup> CD16<sup>hi</sup> NK cells (rs71411868,  
214  $P=5.9 \times 10^{-58}$ ).

215 We identified two independent local-pQTLs in the *FCGR* gene cluster (**Table 1**), which encodes  
216 the most important Fc receptors for inducing phagocytosis of opsonized microbes. Genetic variation  
217 close to *FCGR3A* was associated here with CD16 MFI in CD16<sup>hi</sup> NK cells (rs3845548,  $P=3.0 \times 10^{-87}$ ).  
218 The same variants were also shown to affect the number of CD62L<sup>-</sup> myeloid cDCs in a previous  
219 study<sup>15</sup>. The second signal associated *FCGR2B* variation with CD32 MFI in basophils (rs61804205,  
220  $P=1.7 \times 10^{-36}$ ), but not in eosinophils and neutrophils. Consistently, it is known that basophils express  
221 both CD32a and CD32b proteins, while eosinophils and neutrophils predominantly express CD32a<sup>34</sup>.  
222 Conversely, a local-pQTL was identified at the *SELL* gene, which was associated with CD62L MFI in  
223 eosinophils and neutrophils (rs2223286,  $P=1.6 \times 10^{-35}$  and  $8.8 \times 10^{-13}$ , respectively), but not in basophils  
224 (**Fig. 4b, c**).

225 A number of other local-pQTLs were found to be cell-specific; three different association signals  
226 were found in the *HLA-DR* gene region, with HLA-DR MFI in pDCs and CD14<sup>hi</sup> monocytes  
227 (rs114973966,  $P=2.2 \times 10^{-56}$ ), in cDC1 (rs2760994,  $P=6.1 \times 10^{-38}$ ) and in cDC3 cells (rs143655145,  
228  $P=2.6 \times 10^{-11}$ ). To verify if these signals were independent from each other, we conducted omnibus  
229 association tests on imputed HLA alleles (**Online Methods**). We found that the association signals in

230 CD14<sup>hi</sup> monocytes, pDCs and cDC1 actually resulted from different amino acid-altering variants at  
231 the same multi-allelic position 13 of the HLA-DRβ1 protein ( $P=2.0 \times 10^{-47}$ ,  $7.0 \times 10^{-90}$  and  $5.3 \times 10^{-41}$  in  
232 CD14<sup>hi</sup> monocytes, pDC and cDC1, respectively; **Supplementary Tables 7 and 8**), recently shown to  
233 explain a large part of the association signal in the *HLA* locus for type 1 diabetes<sup>35</sup>. A different amino-  
234 acid variant, at position 67 of HLA-DRβ1, was identified in cDC3s ( $P=3.9 \times 10^{-13}$ ). Conditional  
235 analyses also revealed independent associations of HLA-DR cell-surface expression with two residues  
236 in class I *HLA-B* gene (position 97 and 194;  $P=3.8 \times 10^{-17}$  and  $1.3 \times 10^{-18}$ ; **Supplementary Tables 7 and**  
237 **8**). Collectively, these results show that the protein expression of markers of immune cell  
238 differentiation and activation can be affected by common genetic variants, of which some are known  
239 to be implicated in human pathogenesis.

240

#### 241 **Immune cell local protein QTLs control mRNA levels of nearby genes**

242 Although four of the 9 local-pQTLs identified by our analyses are likely explained by amino acid-  
243 altering variants in surrounding genes (**Supplementary Tables 4 and 7**), the remaining signals do not  
244 present obvious candidate causal variants. To dissect the functional basis of these associations, we  
245 tested if the corresponding SNPs were also associated with mRNA levels of nearby genes (i.e.,  
246 expression QTL, eQTL) using gene expression data obtained from the same donors<sup>36</sup> and results from  
247 the Genotype-Tissue Expression (GTEx) Project<sup>37</sup>. Five of the local-pQTLs were strongly associated  
248 with the transcript levels of a surrounding gene ( $P < 1.0 \times 10^{-5}$ ; **Fig. 4d**). The SNPs controlling the MFI  
249 of CD16 in CD16<sup>hi</sup> NK cells and CD32 in basophils, CD62L in eosinophils, CD8a in CD69<sup>+</sup> CD16<sup>hi</sup>  
250 NK cells and CD203c in basophils were associated with mRNA levels of *FCGR2B*, *SELL*, *CD8A*, and  
251 *ENPP3*, respectively (**Supplementary Table 4**). These analyses indicate that genetic variants  
252 associated with immunophenotypes can directly affect gene expression of markers of immune cells in  
253 whole blood. This suggests that eQTL mapping in different immune cell compartments can greatly  
254 improve our knowledge of the genetic factors controlling human inter-individual variation in flow  
255 cytometric parameters.

256

257

258 **Novel trans-acting genetic associations with immune cell parameters**

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

272 Although identified here for their trans effects on markers of immune cell differentiation or  
273 activation, three trans-acting genetic associations were also local-eQTLs for nearby immune-related  
274 genes<sup>37</sup> (**Supplementary Tables 4 and 6**). The MFI of CCR7 in CD4<sup>+</sup> and CD8b<sup>+</sup> naive T cells was  
275 associated with a variant in the *TMEM8A* gene (rs11648403,  $P=3.0 \times 10^{-19}$ ), which also controls  
276 *TMEM8A* mRNA levels ( $P=2.5 \times 10^{-27}$ ). *TMEM8A* is expressed on the surface of resting T cells and is  
277 down-regulated after cell activation<sup>41</sup>, suggesting a possible functional association and/or co-  
278 regulation with CCR7. Variants in the vicinity of the *ALOX15* gene were associated with increased  
279 protein levels of the high-affinity IgE receptor in eosinophils (rs56170457,  $P=9.2 \times 10^{-14}$ ) and increased  
280 *ALOX15* mRNA levels ( $P=2.7 \times 10^{-13}$ ). These results, together with the high expression of the  
281 *ALOX15* protein and its pro-inflammatory effect in circulating eosinophils<sup>42</sup>, suggest that this  
282 lipoygenase plays an important role in IgE-dependent allergic reactions. Finally, conditional GWAS  
283 identified an additional trans-acting association, between a variant close to the *CD83* gene and HLA-  
284 DR MFI in cDC1 (rs72836542,  $P=2.8 \times 10^{-12}$ , **Supplementary Fig. 22**), the same variant being also  
285 identified as a local-eQTL of *CD83* gene expression ( $P=5.4 \times 10^{-21}$ ). These results suggest that CD83,

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

288

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

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

298 In support of this hypothesis, the presence of HLA-DR molecules, which was assessed at the  
299 surface of both innate and adaptive immune cells, was strongly associated with *HLA-DR* genetic  
300 variation in monocytes, NK and dendritic cells (**Table 1**), but not in memory CD4<sup>+</sup> or CD8<sup>+</sup> T<sub>CM</sub>, T<sub>EM</sub>  
301 and T<sub>EMRA</sub> cells ( $P > 1.0 \times 10^{-6}$ ; **Supplementary Table 5**). Because we observed substantial correlations  
302 among HLA-DR<sup>+</sup> memory T cell numbers ( $R^2 \approx 0.3$ ,  $P < 0.05$ ; **Supplementary Fig. 16**), we  
303 hypothesized that they were at least partly controlled by the same genetic factors, which were further  
304 examined using a multivariate GWAS (**Online Methods**). This refined approach detected a  
305 suggestive genetic association close the *HLA-DRB1* gene with a variant (rs35743245,  $P = 1.0 \times 10^{-8}$ ) in  
306 strong linkage disequilibrium with that detected in pDCs, monocytes and NK cells ( $r^2 = 0.92$ ;  
307 **Supplementary Fig. 23**). This finding provides proof-of-concept that immunophenotypes in both  
308 innate and adaptive cells can be controlled by the same genetic factors, but their effects are stronger in  
309 innate cells, relative to experienced adaptive cells.

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

314 contribution to the total variance (**Online Methods**). We found that a larger proportion of the variance  
315 of innate cell immunophenotypes was explained by genetic factors (**Fig. 5b** and **5d**), relative to  
316 adaptive cell immunophenotypes (**Fig. 5a** and **5e**). Inversely, the variance in adaptive cell numbers  
317 was dominated by non-genetic factors such as age and CMV serostatus (**Fig. 5a**). To test if these  
318 differences were significant, we used a mixed model that accounted for correlations among  
319 immunophenotypes (**Online Methods**). Conclusively, we estimated that the variance explained by  
320 genetics was 66% larger for innate cell measurements, relative to adaptive cells (95%CI: [13%-  
321 143%]; bootstrap  $P=0.012$ ; Mann-Whitney U test:  $P=0.032$ ), while the variance explained by non-  
322 genetic factors was 46% smaller for innate cell measurements (95%CI: [22%-63%]; bootstrap  
323  $P=1.8 \times 10^{-3}$ ; Mann-Whitney U test:  $P=8.1 \times 10^{-3}$ ). When considering non-genetic factors separately, the  
324 ratio of explained variance between innate and adaptive cell measurements was the smallest for  
325 smoking (0.46, 95%CI: [0.17-1.25]), followed by age (0.63, 95%CI: [0.42-0.95]), CMV infection  
326 (0.71, 95%CI: [0.51-0.99]), and sex (0.95, 95%CI: [0.60-1.51]). Taken together, our results indicate  
327 that genetic factors account for a substantial fraction of human variation in immune cell parameters,  
328 with their influence being stronger in innate immune cells, relative to adaptive immune cell  
329 phenotypes.

330

331 **Discussion**

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

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

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

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

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

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

393 **References:**

- 394 1. Bernard, C. *Introduction à l'étude de la médecine expérimentale*. (Libraires de l'Académie Impériale de  
395 Médecine, 1865).
- 396 2. Altfeld, M. & Gale, M. Innate immunity against HIV-1 infection. *Nat. Immunol.* **16**, 554–562 (2015).
- 397 3. Orme, I. M., Robinson, R. T. & Cooper, A. M. The balance between protective and pathogenic immune  
398 responses in the TB-infected lung. *Nat. Immunol.* **16**, 57–63 (2015).
- 399 4. Tollerud, D. J. *et al.* The Influence of Age, Race, and Gender on Peripheral Blood Mononuclear-Cell  
400 Subsets in Healthy Nonsmokers. *J. Clin. Immunol.* **9**, 214–222 (1989).
- 401 5. Reichert, T. *et al.* Lymphocyte Subset Reference Ranges in Adult Caucasians. *Clin. Immunol.*  
402 *Immunopathol.* **60**, 190–208 (1991).
- 403 6. Liston, A., Carr, E. J. & Linterman, M. A. Shaping Variation in the Human Immune System. *Trends*  
404 *Immunol.* **37**, 637–646 (2016).
- 405 7. Goronzy, J. J. & Weyand, C. M. Successful and maladaptive T cell aging. *Immunity* **46**, 364–378  
406 (2017).
- 407 8. Sauce, D. & Appay, V. Altered thymic activity in early life : how does it affect the immune system in  
408 young adults ? *Curr. Opin. Immunol.* **23**, 543–548 (2011).
- 409 9. Furman, D. *et al.* Apoptosis and other immune biomarkers predict influenza vaccine responsiveness.  
410 *Mol. Syst. Biol.* **9**, 1–14 (2013).
- 411 10. Aguirre-Gamboa, R. *et al.* Differential Effects of Environmental and Genetic Factors on T and B Cell  
412 Immune Traits. *Cell Rep.* **17**, 1–14 (2016).
- 413 11. Carr, E. J. *et al.* The cellular composition of the human immune system is shaped by age and  
414 cohabitation. *Nat. Immunol.* **17**, 461–468 (2016).
- 415 12. Boeckh, M. & Geballe, A. P. Cytomegalovirus: pathogen, paradigm, and puzzle. *J Clin Invest* **121**,  
416 1673–1680 (2011).
- 417 13. Wertheimer, A. M. *et al.* Aging and Cytomegalovirus Infection Differentially and Jointly Affect Distinct  
418 Circulating T Cell Subsets in Humans. *J. Immunol.* **192**, 2143–2155 (2014).
- 419 14. Furman, D. *et al.* Cytomegalovirus infection enhances the immune response to influenza. *Sci. Transl.*  
420 *Med.* **7**, 281ra43 (2015).
- 421 15. Orrù, V. *et al.* Genetic variants regulating immune cell levels in health and disease. *Cell* **155**, 242–56  
422 (2013).

- 423 16. Roederer, M. *et al.* The Genetic Architecture of the Human Immune System: A Bioresource for  
424 Autoimmunity and Disease Pathogenesis. *Cell* **161**, 387–403 (2015).
- 425 17. Brodin, P. *et al.* Variation in the Human Immune System Is Largely Driven by Non-Heritable  
426 Influences. *Cell* **160**, 37–47 (2015).
- 427 18. Thomas, S. *et al.* The Milieu Intérieur study — An integrative approach for study of human  
428 immunological variance. *Clin. Immunol.* **157**, 277–293 (2015).
- 429 19. Vivier, E. *et al.* Innate or Adaptive Immunity? The Example of Natural Killer Cells. *Science* (80-. ).  
430 **331**, 44–49 (2011).
- 431 20. Hasan, M. *et al.* Semi-automated and standardized cytometric procedures for multi-panel and multi-  
432 parametric whole blood immunophenotyping. *Clin. Immunol.* **157**, 261–276 (2015).
- 433 21. Patterson, S. *et al.* Cortisol Patterns Are Associated with T Cell Activation in HIV. *PLoS One* **8**, e63429  
434 (2013).
- 435 22. Serafini, N., Vosshenrich, C. A. J. & Di Santo, J. P. Transcriptional regulation of innate lymphoid cell  
436 fate. *Nat. Rev. Immunol.* **15**, 415–428 (2015).
- 437 23. Dusseaux, M. *et al.* Human MAIT cells are xenobiotic-resistant, tissue-targeted, CD161hi IL-17–  
438 secreting T cells. *Blood* **117**, 1250–1260 (2011).
- 439 24. Amado, I. F. *et al.* IL-2 coordinates IL-2-producing and regulatory T cell interplay. *J. Exp. Med.* **210**,  
440 2707–2720 (2013).
- 441 25. Pennell, L. M., Galligan, C. L. & Fish, E. N. Sex affects immunity. *J. Autoimmun.* **38**, J282–J291  
442 (2012).
- 443 26. Furman, D. *et al.* Systems analysis of sex differences reveals an immunosuppressive role for  
444 testosterone in the response to influenza vaccination. *Proc. Natl. Acad. Sci.* **111**, 869–874 (2014).
- 445 27. Astle, W. J. *et al.* The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common  
446 Complex Disease. *Cell* **167**, 1415–1429 (2016).
- 447 28. Della Bella, S. *et al.* Peripheral blood dendritic cells and monocytes are differently regulated in the  
448 elderly. *Clin. Immunol.* **122**, 220–228 (2007).
- 449 29. Puchta, A. *et al.* TNF Drives Monocyte Dysfunction with Age and Results in Impaired Anti-  
450 pneumococcal Immunity. *PLoS Pathog.* **12**, e1005368 (2016).
- 451 30. Vrisekoop, N. *et al.* Sparse production but preferential incorporation of recently produced naïve T cells  
452 in the human peripheral pool. *Proc. Natl. Acad. Sci.* **105**, 6115–6120 (2008).

- 453 31. Tsuchiya, M. *et al.* Smoking a Single Cigarette Rapidly Reduces Combined Concentrations of Nitrate  
454 and Nitrite and Concentrations of Antioxidants in Plasma. *Circulation* **105**, 1155–1157 (2002).
- 455 32. Kearley, J. *et al.* Cigarette Smoke Silences Innate Lymphoid Cell Function and Facilitates an  
456 Exacerbated Type I Interleukin-33-Dependent Response to Infection. *Immunity* **42**, 566–579 (2015).
- 457 33. Monticelli, L. A. *et al.* Innate lymphoid cells promote lung-tissue homeostasis after infection  
458 with influenza virus. *Nat. Immunol.* **12**, 1045–1055 (2011).
- 459 34. Cassard, L., Jönsson, F., Arnaud, S. & Daëron, M. Fcγ Receptors Inhibit Mouse and Human Basophil  
460 Activation. *J. Immunol.* **189**, 1–13 (2012).
- 461 35. Hu, X. *et al.* Additive and interaction effects at three amino acid positions in HLA-DQ and HLA-DR  
462 molecules drive type 1 diabetes risk. *Nat. Genet.* **47**, 898–905 (2015).
- 463 36. Piasecka, B. *et al.* Distinctive Roles of Age, Sex and Genetics in Shaping Transcriptional Variation of  
464 Human Immune Responses to Microbial Challenges. *Proc. Natl. Acad. Sci.* **TBD**, 1–55 (2017).
- 465 37. GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation  
466 in humans. *Science (80-. )*. **348**, 648–660 (2015).
- 467 38. Garris, C. S., Blaho, V. A., Hla, T. & Han, M. H. Sphingosine-1-phosphate receptor 1 signalling in T  
468 cells : trafficking and beyond. *Immunology* **142**, 347–353 (2014).
- 469 39. Pellegrini, M. *et al.* Loss of Bim Increases T Cell Production and Function in Interleukin 7 Receptor–  
470 deficient Mice. *J. Exp. Med.* **200**, 1189–1195 (2004).
- 471 40. van der Harst, P. *et al.* Seventy-five genetic loci influencing the human red blood cell. *Nature* **492**, 369–  
472 375 (2012).
- 473 41. Motohashi, T. *et al.* Molecular Cloning and Chromosomal Mapping of a Novel Protein Gene, M83.  
474 *Biochem Biophys Res Commun* **250**, 244–250 (2000).
- 475 42. Feltenmark, S. *et al.* Eoxins are proinflammatory arachidonic acid metabolites produced via the 15-  
476 lipoxygenase-1 pathway in human eosinophils and mast cells. *Proc. Natl. Acad. Sci.* **105**, 680–685  
477 (2008).
- 478 43. Stämpfli, M. R. & Anderson, G. P. How cigarette smoke skews immune responses to promote infection,  
479 lung disease and cancer. *Nat Rev Immunol* **9**, 377–384 (2009).
- 480 44. Venet, F., Lukaszewicz, A.-C., Payen, D., Hotchkiss, R. & Monneret, G. Monitoring the immune  
481 response in sepsis: a rational approach to administration of immunoadjuvant therapies. *Curr. Opin.*  
482 *Immunol.* **25**, 477–483 (2013).

- 483 45. Kolaczowska, E. & Kubes, P. Neutrophil recruitment and function in health and inflammation. *Nat.*  
484 *Rev. Immunol.* **13**, 159–175 (2013).
- 485 46. Farber, D. L., Yudanin, N. A. & Restifo, N. P. Human memory T cells: Generation,  
486 compartmentalization and homeostasis. *Nat. Rev. Immunol.* **14**, 24–35 (2014).
- 487 47. Mangino, M., Roederer, M., Beddall, M. H., Nestle, F. O. & Spector, T. D. Innate and adaptive immune  
488 traits are differentially affected by genetic and environmental factors. *Nat. Commun.* **8**, 13850 (2017).
- 489 48. Casanova, J. & Abel, L. Disentangling Inborn and Acquired Immunity in Human Twins. *Cell* **160**, 13–  
490 15 (2015).
- 491 49. Paternoster, L. *et al.* Multi-ancestry genome-wide association study of 21,000 cases and 95,000 controls  
492 identifies new risk loci for atopic dermatitis. *Nat. Genet.* **47**, 1449–1456 (2015).
- 493 50. von Bubnoff, D. *et al.* Natural killer cells in atopic and autoimmune diseases of the skin. *J Allergy Clin*  
494 *Immunol* **125**, 60–68 (2010).
- 495

496 **Data Availability:**

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

501

502 **Code Availability:**

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

505

506 **Acknowledgments:**

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

513

514 **Author contributions:**

515 Author contributions were as follows: Conceptualization, E.P., L.Q.-M., M.L.A.; Methodology, M.H.,  
516 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.,  
517 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.,  
518 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.,  
519 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.,  
520 J.P.D.; Writing – Original Draft, E.P., J.B., C.A., D.D., M.L.A.; Writing – Review & Editing, E.P.,  
521 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.,

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

523 M.L.A.

524

525 **Competing financial interests:**

526 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

527 Roche Group.

528

529 **Figure legends**

530

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

538

539 **Figure 2** Respective effects of age, sex and CMV infection on innate and adaptive cell counts in  
540 1,000 healthy individuals. Significant multiplicative effects (adjusted  $P < 0.01$ ) of **(a-c)** increasing age,  
541 **(d-f)** female sex and **(g-i)** CMV seropositivity on circulating levels of immune cells. **(a, d, g)** Effect  
542 sizes were estimated in a linear mixed model with a log-transformed immunophenotype as response,  
543 controlling for batch effects and genome-wide significant SNPs, and then transformed to the original  
544 scale. Adaptive and innate immune cells are represented in grey and black, respectively. The 99%  
545 confidence intervals (99% CIs) were false coverage-adjusted. **(b, e, h)** Regression lines were fitted  
546 using local polynomial regression. **(b)** Impact of age on naive CD8b<sup>+</sup> (in dark green) and CD4<sup>+</sup> (in  
547 light green) T cells. **(e)** Impact of age and sex on the absolute count of MAIT cells. Females are  
548 represented in pink and men in blue. **(h)** Impact of age and CMV serostatus on CD4<sup>+</sup> EMRA T cells.  
549 CMV+ individuals are represented in red and CMV- in orange. **(c)** Flow cytometry plots of naive  
550 CD8b<sup>+</sup> and CD4<sup>+</sup> T cells for representative persons in their 20s and their 60s. **(f)** Flow cytometry plots  
551 of EMRA CD4<sup>+</sup> T cells in representative CMV- and CMV+ subjects. **(i)** Flow cytometry plots of  
552 MAIT cells in representative woman and man. The significant effects of age, sex and CMV  
553 seropositivity on MFI can be found in **Supplementary Fig. 17**.

554

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

569

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

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

583

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

591

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

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

592

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

594 <sup>a</sup>Other immunophenotypes correspond to any measured immunophenotype in the Milieu Intérieur cohort that was also significantly associated with the

595 candidate variant, but to a lesser extent than the main immunophenotype.

596 <sup>b</sup>Replication was performed in an independent cohort of 75 European-descent Americans. Only panels 4 and 7 could be used, due to sample limitations.

597 Effects were in the same direction as in the primary cohort.

598 <sup>c</sup>*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

599 subjects of the Milieu Intérieur cohort.

600 <sup>d</sup>EAF is the frequency of the effect allele, which was defined as the allele with a positive effect on the immunophenotype.

601

602 **Online Methods**

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

604

605 **The Milieu Intérieur cohort**

606 The 1,000 healthy donors of the Milieu Intérieur cohort were recruited by BioTrial (Rennes, France),  
607 and included 500 women and 500 men, and 200 individuals from each decade of life, between 20 and  
608 69 years of age. Donors were selected based on stringent inclusion and exclusion criteria, detailed  
609 elsewhere<sup>18</sup>. The clinical study was approved by the Comité de Protection des Personnes — Ouest 6  
610 (Committee for the protection of persons) on June 13th, 2012 and by the French Agence Nationale de  
611 Sécurité du Médicament (ANSM) on June 22nd, 2012. The study is sponsored by the Institut Pasteur  
612 (Pasteur ID-RCB Number: 2012-A00238-35), and was conducted as a single center study without any  
613 investigational product. The protocol is registered under ClinicalTrials.gov (study# NCT01699893).

614

615 **Human material and staining protocol**

616 Whole blood samples were collected from the 1,000 healthy, fasting donors on Li-heparin, every  
617 working day from 8 to 11AM, from September 2012 to August 2013, in Rennes, France. Tracking  
618 procedures were established in order to ensure delivery to Institut Pasteur, Paris, within 6 hours of  
619 blood draw, at a temperature between 18°C and 25°C. To check the stability of our flow cytometry  
620 measures through time, a second blood sample was drawn for half of the cohort during a second visit,  
621 ~17 days on average after the first visit, ranging from 7 to 44 days. After receipt, samples were kept at  
622 room temperature prior to sample staining. Details on staining protocols can be found elsewhere<sup>20</sup>.

623

624 **Reproducibility testing and assay development**

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

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

631

### 632 **Cytometric analyses**

633 Ten 8-color flow cytometry panels were developed. Details on staining antibodies can be found in  
634 **Supplementary Table 2**. A unique lot of each antibody was used for the entire study. Each antibody  
635 was selected and titrated as described earlier<sup>20</sup>. Gating strategies are described in **Supplementary**  
636 **Figures 1-10**. The acquisition of cells was performed using two MACSQuant analyzers (Serial  
637 numbers 2420 & 2416), each fit with identical three lasers and ten detector optical racks (FSC, SSC  
638 and eight fluorochrome channels). Calibration of instruments was performed using MacsQuant  
639 calibration beads (Miltenyi, ref. 130-093-607). Flow cytometry data were generated using  
640 MACSQuantify™ software version 2.4.1229.1 and saved as .mqd files (Miltenyi). The files were  
641 converted to FCS compatible format and analyzed by FlowJo software version 9.5.3. A total of 313  
642 immunophenotypes were exported from FlowJo. These included 110 cell proportions, 106 cell counts,  
643 89 MFI and 8 ratios. We excluded from subsequent analyses all cell proportions, 35  
644 immunophenotypes that were measured several times on different panels and were exported for  
645 quality controls, and two MFI that were measured with a problematic clone (**Supplementary Table**  
646 **3**). A total of 166 flow cytometry measurements were thus analysed, including 76 cell counts, 87 MFI  
647 and 3 ratios (**Supplementary Table 3**). Problems in flow cytometry processing, such as abnormal  
648 lysis or staining, were systematically flagged by trained experimenters, which resulted in 8.70%  
649 missing data among the 166,000 measured values.

650

### 651 **Outlier removal**

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

657 distributions of flow cytometry measurements. A value in the higher tail was considered an outlier if  
658 the distance to the closest point in the direction of the mean of the distribution was more than 60% of  
659 the total range of the sample, while a value in the lower tail was considered an outlier if that distance  
660 was more than 15% of the total range of the sample. To choose these threshold values, we simulated  
661 10,000 log-normal distributions with a skewness similar to that of the flow cytometry measurements.  
662 We then searched for threshold values so that simulated values outside of these ranges were observed  
663 in less than 5% of the distributions. Outliers were only looked for in the 50 highest and lowest values.  
664 This threshold was chosen to make sure that we do not miss any effect on immunophenotypes of  
665 common genetic variants (minor allele frequency > 5%), or that of one of 39 continuous or common  
666 categorical non-genetic factors studied here. All values more extreme than the points labelled as  
667 outliers were also labelled outliers. A total of 24 values was removed at this stage.

668

#### 669 **Batch effects on flow cytometry measurements**

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

681

#### 682 **Inclusion and imputation of candidate non-genetic factors**

683 A large number of demographic variables were available in the Milieu Intérieur cohort<sup>18</sup>. These  
684 included infection and vaccination history, childhood diseases, health-related life habits, and socio-

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

695

#### 696 **Impact of candidate non-genetic factors on immunophenotypes**

697 To analyse the impact of non-genetic factors on immunophenotypes, we fitted a linear mixed model  
698 for each of the 166 immunophenotypes and each of the 39 non-genetic treatment variables. A total of  
699 6,474 models were therefore fitted using the *lme4* R package<sup>51</sup>. All models were fitted to complete  
700 cases. Due to lack of a priori knowledge on how the non-genetic variables impact the  
701 immunophenotypes, we did not attempt to make a full causal structural equation model for all  
702 variables. Instead, we chose to keep the amount of controls in the models small to increase  
703 interpretability of the results, and to make the study easier to reproduce. We included age, sex and  
704 CMV seropositivity as fixed-effect controls for all models (**Fig. 3** and **Supplementary Fig. 19**),  
705 except when they were the treatment variable to be tested (**Fig. 2** and **Supplementary Figs. 17**). The  
706 intrinsic factors, i.e., age and sex, were included as covariates because they are known to have an  
707 impact on immunophenotypes<sup>6,7,10,14,25-27</sup>, as well as on many of the other environmental exposures,  
708 and are therefore possible confounders. CMV seropositivity was included because it has been shown  
709 to strongly affect some immunophenotypes<sup>6,13,14,17</sup>. We also controlled for genome-wide significant  
710 SNPs for corresponding immunophenotypes (**Table 1**). Genetic variants were included to reduce the  
711 residual variance of the models and to make the inferences more robust. To correct for the batch effect  
712 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<sup>52</sup> and can readily be applied using the *pbkrtest* R package<sup>53</sup>. 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<sup>54</sup>. 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,

741 California). HumanExome SNP call rate was lower than 97% in 11 donors, which were thus removed  
742 from this dataset. We filtered out from both datasets SNPs that: (i) were unmapped on dbSNP138, (ii)  
743 were duplicated, (iii) had a low genotype clustering quality (GenTrain score  $< 0.35$ ), (iv) had a call  
744 rate  $< 99\%$ , (v) were monomorphic, (vi) were on sex chromosomes and (vii) were in Hardy-Weinberg  
745 disequilibrium (HWE  $P < 10^{-7}$ ). These SNP quality-control filters yielded a total of 661,332 and  
746 87,960 SNPs for the HumanOmniExpress and HumanExome BeadChips, respectively. The two  
747 datasets were then merged, after excluding triallelic SNPs, SNPs with discordant alleles between  
748 arrays (even after allele flipping), SNPs with discordant chromosomal position, and SNPs shared  
749 between arrays that presented a genotype concordance rate  $< 99\%$ . Average concordance rate for the  
750 16,753 SNPs shared between the two genotyping platforms was 99.9925%, and individual  
751 concordance rates ranged from 99.80% to 100%, validating that no problem occurred during DNA  
752 sample processing. The final dataset included 732,341 QC-filtered genotyped SNPs.

753

#### 754 **Genetic relatedness and structure**

755 Possible pairs of genetically related subjects were detected using an estimate of the kinship coefficient  
756 and the proportion of SNPs that are not identical-by-state between all possible pairs of subjects,  
757 obtained with KING<sup>55</sup>. Genetic structure was visualized with the Principal Component Analysis  
758 (PCA) implemented in EIGENSTRAT<sup>56</sup>. For comparison purposes, the analysis was performed on  
759 261,827 independent SNPs and 1,723 individuals, which include the 1,000 Milieu Intérieur subjects  
760 together with a selection of 723 individuals from 36 populations of North Africa, the Near East,  
761 western and northern Europe<sup>57</sup>.

762

#### 763 **Genotype imputation**

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

769 imputation was performed by IMPUTE v.2<sup>59</sup>, considering 1-Mb windows and a buffer region of 1 Mb.  
770 Out of the 37,895,612 SNPs obtained after imputation, 37,164,442 were imputed. We removed  
771 26,005,463 imputed SNPs with information  $\leq 0.8$ , 43,737 duplicated SNPs, 955 monomorphic SNPs,  
772 and 449,903 SNPs with missingness  $>5\%$  (individual genotype probabilities  $< 0.8$  were considered as  
773 missing data). After quality-control filters, a total of 11,395,554 high-quality SNPs were further  
774 filtered for minor allele frequencies  $>5\%$ , yielding a final set of 5,699,237 SNPs for association  
775 analyses.

776

### 777 **Genome-wide association analyses**

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

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

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

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

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

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

841

#### 842 **Power estimation**

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

853 hair color (rs12913832,  $P=6.7 \times 10^{-138}$  and  $8.5 \times 10^{-18}$ , respectively), *SLC45A2* with hair color  
854 (rs16891982,  $P=3.2 \times 10^{-9}$ ), *UGT1A* gene cluster with bilirubin levels (rs6742078,  $P=2.6 \times 10^{-75}$ ),  
855 *SLC2A9* with uric acid levels (rs6832439,  $P=4.3 \times 10^{-14}$ ), and *CETP* with HDL levels (rs711752,  
856  $P=4.5 \times 10^{-8}$ ).

857

### 858 **Enrichment in variants associated with diseases**

859 We explored the implication of our 15 genome-wide significant variants in human diseases and traits  
860 using previously published hits of genome-wide association studies (GWAS), obtained from the  
861 31/08/2017 version of the EBI-NHGRI GWAS Catalog. A candidate variant was considered as  
862 implicated in a disease/trait if it was previously associated with such a disease/trait with a  $P < 5 \times 10^{-8}$ ,  
863 or if it was in linkage disequilibrium (LD) with a variant associated with such a disease/trait ( $r^2 > 0.6$ ).  
864 We tested if our 15 genome-wide significant variants were enriched in known associations with  
865 diseases/traits by resampling. Namely, we sampled 100,000 times 15 random SNPs with minor allele  
866 frequencies matched to those observed, and we calculated for each resampled set the proportion of  
867 variants known to be, or in LD with a variant known to be, associated with a disease. The enrichment  
868  $P$ -value was estimated as the proportion of resamples for which this proportion was larger than that  
869 observed in our set. LD was precomputed for all 5,699,237 SNPs with PLINK 1.9 (options ‘--show-  
870 tags all --tag-kb 500 --tag-r2 0.6’)<sup>69</sup>.

871

### 872 **HLA typing and association tests**

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

878

### 879 **Replication cohort**

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

892

### 893 **Gene expression assays**

894 NanoString nCounter®, a hybridization-based multiplex assay, was used to measure gene expression  
895 in non-stimulated whole blood of the 1,000 Milieu Intérieur subjects, with the Human Immunology v2  
896 Gene Expression CodeSet. This data is described in detail in a separate work<sup>36</sup>. Expression probes that  
897 bind to cDNAs in which at least 3 known common SNPs segregate in humans were removed from the  
898 analyses (*i.e.*, *HLA-DQB1*, *HLA-DQA1*, *HLA-DRB1*, *HLA-B* and *C8G*). After quality control filters,  
899 mRNA levels were available for 986 individuals at 90 candidate genes, *i.e.*, immunity-related genes in  
900 a 1-Mb window around the genome-wide significant and suggestive associations identified in this  
901 study. For each sample, probe counts were log<sub>2</sub> transformed, normalized and adjusted for batch  
902 effects. eQTL mapping was performed in a 1-Mb window around corresponding association signals,  
903 using the linear mixed model implemented in GenABEL<sup>72</sup>. All models were adjusted on the  
904 proportion of eight major cell populations, including neutrophils, CD19<sup>+</sup> B cells, CD4<sup>+</sup> T cells, CD8<sup>+</sup>  
905 T cells, CD4<sup>+</sup>CD8<sup>+</sup> T cells, CD4<sup>-</sup> CD8<sup>-</sup> T cells, NK cells, and CD14<sup>+</sup> monocytes, to account for the  
906 effect of heterogeneous blood cell composition on gene expression.

907

908 **Decomposition of the proportion of variance explained**

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

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

935

936 **References**

- 937 51. Bates, D., Maechler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *J.*  
938 *Stat. Softw.* **67**, 1–48 (2015).
- 939 52. Kenward, M. G. & Roger, J. H. Small sample inference for fixed effects from restricted maximum  
940 likelihood. *Biometrics* **53**, 983–997 (1997).
- 941 53. Halekoh, U. & Højsgaard, S. A Kenward-Roger Approximation and Parametric Bootstrap Methods for  
942 Tests in Linear Mixed Models - The R Package pbkrtest. *J. Stat. Softw.* **59**, 1–30 (2014).
- 943 54. Benjamini, Y. & Yekutieli, D. False discovery rate-adjusted multiple confidence intervals for selected  
944 parameters. *J. Am. Stat. Assoc.* **100**, 71–93 (2005).
- 945 55. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics*  
946 **26**, 2867–2873 (2010).
- 947 56. Patterson, N., Price, A. L. & Reich, D. Population Structure and Eigenanalysis. *PLoS Genet.* **2**, e190  
948 (2006).
- 949 57. Behar, D. M. *et al.* The genome-wide structure of the Jewish people. *Nature* **466**, 238–242 (2010).
- 950 58. Delaneau, O., Zagury, J.-F. & Marchini, J. Improved whole-chromosome phasing for disease and  
951 population genetic studies. *Nat. Methods* **10**, 5–6 (2013).
- 952 59. Howie, B. N., Donnelly, P. & Marchini, J. A Flexible and Accurate Genotype Imputation Method for  
953 the Next Generation of Genome-Wide Association Studies. *PLoS Genet.* **5**, e1000529 (2009).
- 954 60. Johnson, W. E. & Li, C. Adjusting batch effects in microarray expression data using empirical Bayes  
955 methods. *Biostatistics* **8**, 118–127 (2007).
- 956 61. Mefford, J. & Witte, J. S. The Covariate’s Dilemma. *PLoS Genet.* **8**, 1–2 (2012).
- 957 62. Meinshausen, N. & Bühlmann, P. Stability selection. *J. R. Stat. Soc. Ser. B* **72**, 417–473 (2010).
- 958 63. Shah, R. D. & Samworth, R. J. Variable selection with error control : another look at stability selection.  
959 *J. R. Stat. Soc. Ser. B* **75**, 55–80 (2013).
- 960 64. Hastie, T., Tibshirani, R. & Friedman, J. *Elements of statistical learning*. (Springer, 2009).
- 961 65. Wakefield, J. *Bayesian and Frequentist Regression Methods*. (Springer, 2013).
- 962 66. Zhou, X. & Stephens, M. Efficient multivariate linear mixed model algorithms for genome-wide  
963 association studies. *Nat. Methods* **11**, 407–409 (2014).
- 964 67. Yang, J., Zaitlen, N. A., Goddard, M. E., Visscher, P. M. & Price, A. L. Advantages and pitfalls in the  
965 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









