Human thymopoiesis is influenced by a common genetic variant within the TCRA-TCRD locus

Emmanuel Clave1,2*, Itauá Leston Araujo1,2*, Cécile Alanio3,4,5*, Etienne Patin6,7,8*, Jacob Bengstedt9*, Alejandra Urrutia3,4,5,10, Silvia Lopez-Lastra5,11, Yan Li5,11, Bruno Charbit3, Cameron Ross MacPherson3, Milena Hasan3, Breno Luiz Melo-Lima1,2, Corinne Douay1,2, Noémie Saut12,13, Marine Germain14,15, David-Alexandre Trégouët3,4,5, Pierre-Emmanuel Morange12,13, Magnus Fontes10,16, Darragh Duffy3,4,5, James P. Di Santo5,11, Lluís Quintana-Murci3,4,5,7,81, Matthew L. Albert3,4,5,10‡, Antoine Toubert1,2,17‡, The Milieu Intérieur Consortium§

The thymus is the primary lymphoid organ where naïve T cells are generated; however, with the exception of age, the parameters that govern its function in healthy humans remain unknown. We characterized the variability of thymic function among 1000 age- and sex-stratified healthy adults of the Milieu Intérieur cohort, using quantification of T cell receptor excision circles (TRECs) in peripheral blood T cells as a surrogate marker of thymopoiesis. Age and sex were the only nonheritable factors identified that affect thymic function. TREC amounts decreased with age and were higher in women compared to men. In addition, a genome-wide association study revealed a common variant (rs2204985) within the T cell receptor TCRA-TCRD locus, between the DD2 and DD3 gene segments, which associated with TREC amounts. Strikingly, transplantation of human hematopoietic stem cells with the rs2204985 GG genotype into immunodeficient mice led to thymopoiesis with higher TRECs, increased thymocyte counts, and a higher TCR repertoire diversity. Our population immunology approach revealed a genetic locus that influences thymopoiesis in healthy adults, with potentially broad implications in precision medicine.

INTRODUCTION

In healthy individuals, continuous production of naïve self-tolerant T cells by the thymus ensures potent immune responses toward newly encountered antigens and contributes to maintenance of the naïve T cell repertoire (1). Thymic function has been extensively studied for its capacity to shape the adaptive immune repertoire through positive and negative selection (2, 3). However, little is known about the environmental or genetic determinants of thymopoiesis in healthy individuals. Such insights would be relevant for optimizing regenerative strategies (4), especially in conditions where thymic function is altered, such as aging (5, 6), HIV infection (7), or alloge neic hematopoietic stem cell transplantation (allo-HSCT) (8).

A bilateral cross-talk between thymocytes and thymic stromal cells directs sequential intrathymic T cell development and helps maintain activity of thymic stromal niches (9, 10). Thymocyte progenitors receive signals from cortical thymic epithelial cells (TECs) for their commitment to the T cell lineage via the engagement of the NOTCH1 receptor with Delta-like 4 ligand, a major Forkhead box protein (FOX)N1 target in the thymic epithelium (11). The medulla, via medullary TECs (mTECs) and dendritic cells, has a critical role in establishing self-tolerance by negative selection and induction of regulatory T cells (Treg), especially but not exclusively via mTECs expressing the autoimmune regulator gene (AIRE) (9). Naïve T cells are heterogeneous and include so-called recent thymic emigrants (RTEs), a subset that undergoes further post-thymic maturation (12). Some phenotypic markers have been proposed to identify RTEs, such as CD31 (PECAM-1) in CD4+ T cells. However, CD31 expression can be maintained during cytokine-driven proliferation of CD4+ T cells, limiting its use as a specific marker of thymopoiesis. RTEs are enriched in T cell receptor excision circles (TRECs), which are produced during thymic TCR somatic recombination (13). TRECs persist within mature T cells as episomal DNA (14), cannot replicate, and are diluted out by peripheral cell divisions. Their quantification in peripheral blood provides a noninvasive surrogate marker of thymopoiesis, especially relevant in steady-state homeostatic conditions of the T cell compartment.

Signal joint TRECs (sjTRECs) are generated during the recombination of the TCRβ chain, in double-positive (DP) CD4+CD8+ thymocytes, before positive and negative selection and lineage commitment (14). Polymerase chain reaction (PCR)–based quantification of sjTRECs is used in clinical laboratories as a diagnostic test for the recovery of the naïve T cell repertoire during HIV treatment, after allo-HSCT, and in the screening of severe combined immunodeficiencies in newborns (15–17). Similarly, assays are available to...
measure βTRECs generated during the TCRβ chain recombination at the CD4+CD8+ double-negative (DN) 3 stage. Because the β chain recombines before the α chain, βTRECs are much less abundant than sjTRECs in the periphery and frequently fall below the detection threshold in quantitative PCR. Given the dilution of βTRECs at each cell division between βTREC and sjTREC generation, the log2 transformation of the sjTREC/βTREC ratio gives an estimate of the number of intrathymic divisions occurring between DN and DP stages (7).

Here, we quantified TRECs from the peripheral blood of 1000 healthy individuals of western European ancestry [the Milieu Intérieur (MI) cohort] at immunological steady state, stratified by sex and age across five decades of life from 20 to 69 years (18). This population immunology approach revealed determinants of heterogeneity in human thymic function and identified a common genetic variation within the TCRA-TCRD locus directly affecting thymopoiesis.

RESULTS

Validation of TRECs as surrogate markers of thymic function in the MI cohort

We first standardized and validated sjTREC and βTREC high-throughput assays (Fig. 1A and fig. S1) and then applied them to DNA from the 1000 MI donors. sjTREC counts normalized per 150,000 whole blood cells were used in subsequent analyses and correlated ($r^2 = 0.99, P < 10^{-16}$) with sjTRECs calculated as absolute numbers per microliter of blood (fig. S1C), the latter being not affected by T cell peripheral divisions. Log$_{10}$-transformed values of sjTRECs [log$_{10}$ sjTRECs] showed a normal distribution (kurtosis test, $P = 0.25$), with a mean of 2.4 ± 0.03 (minimum to maximum range, 0.2 to 4.1; fig. S1D). By contrast, log$_{10}$-transformed values of βTRECs [log$_{10}$ βTRECs] showed a bimodal distribution, with 368 donors having samples below the limit of assay detection. In donors with detectable βTRECs in whole blood, log$_{10}$ βTRECs followed a normal distribution (kurtosis test, $P = 0.70$), with a mean of 1.75 ± 0.06 (minimum to maximum range, 0 to 3.1). Finally, the number of intrathymic divisions was also normally distributed across the healthy donors with detectable βTRECs (kurtosis test, $P = 0.72$), with a mean of 3.0 ± 0.21 (minimum to maximum range, −4.3 to 10.6; fig. S1D).

We evaluated whether TRECs are associated with any of 173 immune cell variables, defined through 10 eight-color immunophenotyping flow cytometry panels (19). Using a generalized linear mixed model approach controlling for potential confounders and batch effects, sjTRECs were found to be strongly associated with naïve CD8+ and CD4+ T cell counts or other cell types that are known to develop within the thymus, including naïve T$_{reg}$ and invariant natural killer T (iNKT) cells (Fig. 1B and fig. S2). Naïve CD8+ T cell counts doubled with a 10-fold increase in sjTRECs [odds of having detectable amounts decreased by 3.36% per year (CI, 0.03 to 5.9%; K-R F test, adjusted $P = 6 	imes 10^{-10}$; fig. 2A) yet remained detectable in >95% of 60- to 69-year-old donors. Among donors with detectable βTRECs, we detected a 2% decrease per year (CI, 0.7 to 3.4%; adjusted $P = 8 	imes 10^{-3}$; fig. 2B). We also observed fewer donors with detectable amounts of βTRECs as a function of age [odds of having detectable amounts decreased by 3.36% per year (CI, 0.03 to 5.9%; K-R F test, adjusted $P = 9 	imes 10^{-3}$)]. Strikingly, sex also showed a strong effect on sjTRECs amounts with 67% (CI, 38 to 102%; adjusted $P = 2 	imes 10^{-15}$) higher sjTREC amounts in women of all ages, relative to men (Fig. 2D). In contrast, no associations were found between sex and either the probability of having detectable βTRECs, βTREC amounts for donors with detectable βTRECs, or the number of intrathymic divisions (adjusted $P > 0.05$; figs. S5 and S6).

Association of a genetic variation at the TCRA-TCRD locus with sjTRECs

We next conducted a genome-wide association study of log$_{10}$-transformed sjTREC numbers on 5,699,237 common single-nucleotide polymorphisms (SNPs) with a linear mixed model adjusted for age, sex, genetic relatedness, and other covariates selected using a data-driven variable selection scheme (24). No association was detected at genome-wide significance (LRT, $P < 5.0 	imes 10^{-8}$). Nevertheless, seven independent genomic regions on chromosomes 2, 4, 5, 10, 11, 14, and 17 showed suggestive evidence for association (LRT, $P < 1.0 	imes 10^{-5}$; Fig. 3A). To test for replication of these suggestive associations, we measured sjTRECs in an independent cohort, the Marseille Thrombosis Association (MARTHA) cohort, which includes 612 unrelated patients of European descent affected with venous thromboembolism (25). We validated in this cohort the association of decreased sjTRECs with increasing age (4.05% per year; CI, 3.55 to 4.56%; K-R F test, $P = 5 	imes 10^{-45}$; fig. S8A), and their higher abundance in women, relative to men (86%; CI, 60 to 116%; $P = 1.6 	imes 10^{-15}$; fig. S8B). Among 14 SNPs tagging the seven suggestively associated loci, only variants on chromosome 14 showed statistical evidence for replication in the MARTHA cohort (table S2).
Fig. 1. Thymic function associates with naïve T cell immune phenotypes. (A) ηTRECs (blue) are episomal DNA generated during the TCRβ recombination. sjTRECs (purple) derive from the deletion of the TCRD locus during TCRα locus recombination (shown in fig. S1B). DN, double negative; ISP, immature single positive; DP, double positive; SP, single positive. (B) Effect sizes of significant associations [adjusted P values (adj. P) < 0.05] between sjTRECs and immune cells and parameters measured by flow cytometry in 969 healthy individuals from the MI cohort. Effect sizes were estimated in a mixed model (see Supplementary Materials and Methods). MFI, mean fluorescence intensity. (C) Relationships between sjTRECs and the log10-transformed number of naïve CD4+ and CD8+ T cells, naïve Treg, and iNKT cells. Regression lines were fitted using linear regression. Adjusted P values were obtained using the mixed model and based on the Kenward-Rogers F test.
These variants all mapped within a 25-kb region included in the TCRA-TCRD locus (Fig. 3B).

To fine-map the signal, we genotyped the eight most informative imputed SNPs within this region in the MI cohort and combined these data with array-based or imputed genotype data from the MARTHA cohort. This led us to identify four SNPs in linkage disequilibrium (rs8013419, rs10873018, rs12147006, and rs2204985) with genome-wide statistical significance (DerSimonian and Laird meta-analysis, \( P < 2 \times 10^{-8} \); table S2) located in the intergenic DD2 and DD3 segments (Fig. 3C). Among them, rs2204985 (located 472 bases upstream of DD3) was considered the most likely candidate variant (effect allele frequency of 0.49; meta-analysis, \( P = 1.9 \times 10^{-8} \); Fig. 4C). We next studied thymocyte developmental stages on human CD45+ cells by flow cytometry (fig. S9). We observed that mice grafted with cells from rs2204985 genotype GG donors had larger thymocyte counts at all stages, starting as early as the CD3−CD4−CD8− DN population (Fig. 4D). These data support the hypothesis of a T cell–intrinsic effect of the identified genetic variant, which associates with thymocyte counts.

As shown in fig. S1, sjTRECs are produced by the \( \delta \text{Rec}-\Psi/\alpha(\text{Ja}61) \) recombination leading to the prominent TCRD locus deletion. However, there are alternative rearrangements including the one between \( \delta \text{Rec} \) and \( \alpha/\text{Ja}58 \) gene segments that represents 23% of total \( \delta \text{Rec} \) rearrangement (fig. S1B) (29). We found a similar effect of the rs2204985 genotype on the alternative \( \delta \text{Rec}-\alpha/\text{Ja}58 \) rearrangement as on sjTRECs (fig. S10), excluding an effect of the genetic variant on the \( \alpha/\text{Ja} \) segment usage during primary TCRA rearrangements. Evaluating the TCRA-TCRD repertoire diversity according to rs2204985 genotypes (table S3), we found that the numbers of total and productive rearrangements did not differ (Mann-Whitney \( U \) test, \( P > 0.05 \)). We found no specific overlap of TCRA-TCRD clonotypes, as calculated by the Morisita index, between mice grafted with the same fetal liver CD34+ cells or even with the same rs2204985 genotype (fig. S11). Conversely, repertoire diversity, as quantified by productive clonality or Shannon equilibrium indexes, was significantly greater in mice grafted with cells of the GG genotype (Mann-Whitney \( U \) test, \( P = 0.016 \) and \( P = 0.003 \), respectively; table S3). Whereas no differences in TCRAV and TCRAJ gene segment usage were observed among mice grafted with cells of the AA or GG genotypes (Mann-Whitney \( U \) test, \( P > 0.05 \); Fig. 5, A and B), large differences were found in TCRDV and TCRDJ usage, with a preferential usage of gene segments close to the variant region (DJ1, DV2, and DV3) in rs2204985 AA individuals (adjusted \( P < 0.05 \); Fig. 5B). Accordingly, the calculated frequency of T cells carrying a productive TCRD rearrangement was higher in AA individuals (Mann-Whitney \( U \) test, \( P = 0.012 \); Fig. 5C). A more detailed analysis of TCRDV and TCRDJ usage restricted to productive TCRD rearrangements showed that DV1, DD2, and DJ1 segments were used preferentially in GG, whereas DV2, DD3, and DJ3 were used preferentially in AA individuals (Fig. 5D), confirming that the rs2204985 variant locally affects TCRD rearrangements.

### Influence of the TCRA-TCRD genetic polymorphism on T cell development in immunodeficient mice

Immunodeficient mice engrafted with human hematopoietic stem cells (HSCs) are able to develop a diverse repertoire of thymus-dependent human T cells (27). To directly evaluate in vivo the impact of the rs2204985 polymorphism on thymopoiesis, we reconstituted immunodeficient Balb/c Rag2−/−Il2rg−/−Sirpa−/−NOD (BRGS) mice (28) with human CD34+ hematopoietic progenitors harvested from fetal livers having different genotypes for the rs2204985 variant (Fig. 4A). Controlling for mouse recipient sex in a linear model, we observed significantly higher sjTRECs (multiplicative effect size CI, 1.24 to 2.16; \( t \) test, \( P = 7 \times 10^{-4} \); Fig. 4B) and total CD3+ thymocyte numbers (multiplicative effect size CI, 1.63 to 3.92; \( P = 9 \times 10^{-5} \); Fig. 4C) in thymi of mice reconstituted with CD34+ progenitors of the rs2204985 GG genotype, as compared to mice reconstituted with AA or GA genotypes. Significant results were also obtained when controlling for the origin of the human fetal liver sample (1.6 times increase in sjTRECs: CI, 1.1 to 2.5; K-R \( F \) test, \( P = 0.047 \); 2.5 times increase in thymocytes: CI, 1.54 to 4.25; \( P = 6.6 \times 10^{-3} \)). We next studied thymocyte developmental stages on human CD45+ cells by flow cytometry (fig. S9). We observed that mice grafted with cells from rs2204985 genotype GG donors had larger thymocyte counts at all stages, starting as early as the CD3−CD4−CD8− DN population (Fig. 4D). These data support the hypothesis of a T cell–intrinsic effect of the identified genetic variant, which associates with thymocyte counts.

### Modeling the variance of thymic function in healthy adults

Finally, we developed a model that estimates TREC content in healthy adults as a function of the rs2204985 genotype, age, and sex. We combined data of the MI and MARTHA cohorts in a mixed model, controlling for population stratification and batch variables. We found a 43% increase of sjTRECs in rs2204985 GG homozygotes, relative to AA homozygotes in the MI cohort (marginal CI, 22 to 69%; Fig. 6A). Similarly, in the MARTHA cohort, we found a 44% increase of sjTRECs in rs2204985 GG homozygotes, relative to AA homozygotes (marginal CI, 21 to 71%; Fig. 6B). The relative contribution of age, sex, and the rs2204985 variant to the variance of log_{10} sjTRECs was estimated to be 37.8, 4.78, and 1.32% in the MI cohort and 25.6, 8.5, and 1.3% in the MARTHA cohort, respectively (Fig. 6C). There was no indication that the effect of age on sjTRECs
Fig. 3. Genome-wide association study reveals an impact of TCRA-TCRD genetic variation on thymic function. (A) Manhattan plot for genetic association with sjTRECs in the 969 donors of the MI cohort. Light and dark gray lines indicate the threshold for suggestive association ($P = 1.0 \times 10^{-5}$) and genome-wide significant association ($P = 5.0 \times 10^{-8}$), respectively. (B) Detailed view of the TCRA-TCRD locus. Primers (sjTREC-F/R) and probe (sjTREC-P) used to quantify sjTRECs are shown in red and cyan, respectively. (C) Fine mapping of the genetic association between the TCRA-TCRD locus and sjTRECs. Meta-analysis $P$ values were obtained by combining array-based, probe-based, and imputed genotypes of the MI and MARTHA cohorts (table S2). Variants that are significantly associated at the genome-wide level are indicated in red. (D) Physical position of the four most strongly associated variants, relative to active transcription activity (26). The position of Dδ3 is indicated.
was dependent on rs2204985 genotypes (CI: 0.94 to 0.95, 0.94 to 0.96, and 0.94 to 0.96 for AA, GA, and GG, respectively, in MI; CI: 0.95 to 0.97, 0.95 to 0.96, and 0.95 to 0.98 for AA, GA, and GG, respectively, in MARTHA). We next sought to express the effect of the \textit{TCRA-TCRD} genetic variation as a function of “thymic age,” defined as the age of a male carrying the AA genotype with sjTRECs equal to those predicted by a linear model fitted on age, sex, and the rs2204985 genotype, using combined data of the MI and MARTHA cohorts. We then estimated the difference between actual age and thymic age for women and men carrying the GG genotype of 18.5 years (CI, 15 to 22.2) and 7.3 years (CI, 4.57 to 10.1), respectively (Fig. 6D).

To support the application of rs2204985 genotyping in future clinical studies, we have developed the Shiny application allowing interactive visualization of the MI data (https://mithymus.pasteur.fr) (fig. S12).

**DISCUSSION**

The thymus is the primary lymphoid organ where T lymphocytes are generated in the adaptive immune system of all vertebrates, through spatiotemporal interactions between thymocytes and specialized microenvironments (9). The thymus is sensitive to insults received throughout life upon inflammation and infections, reflected in its functional decline with age (5, 6, 15). It is, however, an extremely plastic tissue endowed with endogenous regenerative capacities after an acute damage during chemotherapies or irradiation (4, 30, 31). However, the parameters that control the levels of thymic function in homeostatic conditions remain largely unknown, an unmet need to develop precision and regenerative medicine. Here, by combining TREC quantification and a population immunology approach, we report the assessment of nongenetic and genetic determinants of thymic function in healthy adults.

sjTRECs are produced by the thymus and diluted out during T cell divisions (15). Taking into account the dynamics of TREC and peripheral cell division in young healthy individuals, an average of 4% per year involution in thymic output was previously estimated on the basis of the dynamics of TREC and peripheral cell division in young healthy individuals (32), which is in line with our values of a 4.9 and 4% decrease per year in sjTREC amounts in the MI and MARTHA cohorts, respectively. In addition, our study in immunodeficient mice reconstituted with human CD34+ HSC allowed a direct investigation of the impact of \textit{TCRA-TCRD} genetic variation on the developing thymocytes independent of any peripheral dilution of TRECs.

**Fig. 4. Effect of TCRA-TCRD human genetic variation on thymic function in humanized immunodeficient mice.** (A) Immunodeficient Balb/c Rag2\(^{-/-}\)Il2rg\(^{-/-}\)Sirpa\(^{NOD}\) (BRGS) mice were reconstituted with human CD34\(^{+}\) hematopoietic progenitors harvested from fetal livers with rs2204985 genotype AA (orange), GA (brown), or GG (purple). (B) Effects of rs2204985 genotypes on sjTRECs in all mice (AA, \(n = 19\); GA, \(n = 58\); GG, \(n = 15\)). (C) Effects of rs2204985 genotypes in immunophenotyped mice (AA, \(n = 5\); GA, \(n = 31\); GG, \(n = 13\)) on number of CD3\(^{+}\) thymocytes and (D) on thymocyte subsets at different developmental stages. Indicated \(P\) values correspond to the genotype effect in a linear model including genotype and mouse recipient sex.
The only nonheritable factors that we found strongly affecting thymopoiesis in the healthy population were age and sex, with a higher thymic function in women relative to men. Previous studies reported higher thymic mass, as measured by computed tomography in young (20 to 30 years old) women relative to young men (33). We demonstrate that the impact of sex on sjTREC is observed during all of adulthood. In mice, androgens have a direct detrimental effect on stromal TECs (34), and male cortical TECs express low levels of genes implicated in thymocyte expansion and positive selection (35). We suggest that the sex differences observed in our study could similarly reflect sex differences in TEC function, resulting in a more efficient bilateral cross-talk between thymocytes and thymic stroma and higher thymopoiesis in women (9). Overall, the strong and replicated effect of sex on TREC content reinforces the need of stratifying immunological studies by sex (36).

Twin studies reported that RTE numbers are highly heritable, although no genetic associations have been found so far (37). In addition, naïve CD27+ CD4 T cell counts have a high estimated heritability in healthy twins (38) and in the MI cohort (19). Collectively, these studies estimated a higher heritability of naïve rather than

---

**Fig. 5. Effects of TCRA-TCRD human genetic variation on thymic TCR repertoire in humanized immunodeficient mice.** The human TCRA-TCRD locus was sequenced using genomic DNA from 8 (3 males and 5 females) and 12 (4 males and 8 females) immunodeficient mice thymi grafted with rs2204985 AA (orange) and GG (purple) human fetal livers, respectively (table S3). **(A)** Effects of the donor genotype on V (left) and J (right) gene usage, among TCRα and TCRδ productive rearrangements. **(B)** Ratio of median percentage of V (left) or J (right) gene usage in GG-grafted mice, over that in AA-grafted mice. Gene segments used specifically by TCRα genes are indicated in red, by TCRδ in gray, and shared by both in cyan. Whiskers indicate bias-corrected and accelerated bootstrap 95% CIs. **(C)** Effect of genotypes on the percentage of DJ genes/total J genes (%) in GG-grafted mice, over that in AA-grafted mice. **(D)** Effect of genotypes on the percentages of Delta genes among total J genes (%) in GG-grafted mice, over that in AA-grafted mice. Genes are ordered according to their genomic location (see Fig. 3B). Blue asterisks indicate P < 0.05 obtained using nonparametric Mann-Whitney U test, adjusted for multiple testing using the false discovery rate as error rate.
is predicted from a regression model, where AA men are assumed as the baseline.

...required further studies, the data collected in immunodeficient mice served association of sjTRECs with genetic polymorphism will require additional upstream elements to promote TCRD accessibility directly or through the binding of transcription factors and participate in the regulation of the TCRD recombination center (45). It will be interesting to investigate whether this polymorphism affects the generation of the different TCRγδ T cell subsets (41, 42). It remains also to be explained how the TCR genetic polymorphism could be linked to thymocyte survival or thymocyte proliferation at the DN stage. Notably, physiological DNA double-strand breaks generated in developing lymphocytes activate a broad transcriptional program (46), some of them promoting lymphocyte survival via, for instance, the activation of p38MAPK in DN thymocytes (47). In addition, transcription factors binding the rs2204985 genomic region might affect DN survival/proliferation, such as FOXM1, which is required for cellular proliferation in normal cells (48). It is intriguing to find evidence for genetic control of T cell generation in loci deleted in all mature peripheral T cells, TCRγδ through TCRD rearrangements and TCRαβ T cells through sjTREC generation. This suggests selective pressure at a critical step in T cell development, which might be otherwise unnecessary or possibly harmful if functional in the periphery. In support of the pathogenic potential of this genomic region is its proposed involvement during oncogene activation in T cell acute lymphoblastic leukemia (49).

By providing reference values of thymic function in a large healthy population via a key genetic control, our data provide a resource that may be useful in the context of precision medicine and regenerative strategies for diverse diseases. This study contributes to a better understanding of aging of the immune system, a major public health concern (50). We showed that the decrease in thymic output with age was different in men and women and was independent of several environmental factors, including latent CMV infection, previously shown to associate with exhaustion of differentiated T cells (19, 51). About 50% of the variance in sjTREC numbers remained unexplained, suggesting a role for still unknown environmental or genetic factors. Nonetheless, we showed differences in healthy thymic function depending on the TCRα-TCRD genetic variation in two independent cohorts of western European origin. It is important to estimate this impact in other ethnic groups, especially given the differences in frequency of the rs2204985 G allele across populations, ranging from 25% in East Asia to >80% in South America (52).
Considering the clinical implications of our findings, we anticipate that there may be settings where it would be beneficial to achieve a higher potential for T cell production. This would be the case, for instance, in an uncomplicated allo-HSCT setting or in the recovery of lymphopenic conditions in young patients. In contrast, it would be detrimental to fuel the system if the thymic environment is damaged as, for instance, in older individuals, in graft versus host disease in allo-HSCT (8, 16) or in autoimmune conditions where women are known to have an overall higher susceptibility (53). Such cases could result in the generation of T cells defective in their selection process with an autoreactive potential which could be pathogenic.

**MATERIALS AND METHODS**

**Study design**

**MI cohort**

The 1000 healthy donors of the MI cohort were recruited from September 2012 to August 2013 by Biotrial, stratified by sex (500 men and 500 women) and age (200 individuals from each decade between 20 and 69 years of age). Donors were selected on the basis of inclusion and exclusion criteria detailed elsewhere (18). To avoid the influence of hormonal fluctuations in women during the perimenopausal phase, only pre- or postmenopausal women were included. To avoid issues related to population stratification, the study was restricted to French citizens with Metropolitan French origin for three generations. The clinical study was approved by the Comité de Protection des Personnes–Ouest 6 on 13 June 2012 and by the French Agence Nationale de Sécurité du Médicament on 22 June 2012. The study is sponsored by the Institut Pasteur (Pasteur ID-RCB number: 2012-A00238-35) and was conducted as a single center study without any investigational product. The protocol is registered under ClinicalTrials.gov (study number NCT01699893). Primary data for the humanized mouse experiments are shown in table S5.

**Replication cohort**

Our replication cohort included 612 patients from the MARTHA cohort (25). Donors were all of European descent and were examined between January 1994 and October 2005 for having suffered a single venous thrombosis event, without detectable cause. The study was approved by the Institutional Ethics Committee (“Département Santé de la Direction Générale de la Recherche et de l’Innovation”; Projects DC; 2008–880 & 09.576), and written informed consent was obtained from each subject. MARTHA biobank is hosted by the HEMOVASC biosource center (CRB APHM). sjTRECs of all donors were quantified in DNA extracted from blood. Genotypes for candidate variants were obtained from the Illumina Human610-Quad SNP array (25) or probe-based genotyping.

**Statistical analysis**

We tested for association between TRECs and immunophenotypes, and TRECs and nonheritable factors by fitting linear mixed models, using the mmi R package (https://github.com/jacobbergstedt/mmi). The CIs were false coverage-adjusted intervals designed to keep the rate of false coverage at 5%. Hypothesis tests were done using K-R F tests with the false discovery rate as error rate. Impact of nonheritable factors on sjTREC detection status was analyzed using logistic regression and LRTs. Genome-wide association studies were conducted using linear mixed models controlling for nonheritable variables and using the GRM as one of the correlation matrices. A similar model was used to compute effect sizes and 95% CIs for the rs2204985 polymorphism, age, and sex, with respect to sjTRECs in the MI cohort. For the MARTHA cohort, the four principal components of the genotype matrix that explained most variance were used instead of the GRM, and the hypothesis test was conducted using the K-R F test. The DerSimonian and Laird method was used to compute the meta-analysis P values. Both linear regression models and linear mixed models were used to compute 95% CIs and P values for the effect of the rs2204985 polymorphism on sjTREC numbers and thymus T cell progenitors in humanized mice. For gene segment usage, nonparametric 95% CIs were estimated by a bootstrap procedure. Thymic age and proportion of variance was estimated from a linear regression model with log10-transformed sjTRECs as response and age, sex, and the rs2204985 polymorphism as predictors. Details on these analyses can be found in the Supplementary Materials.

**SUPPLEMENTARY MATERIALS**

www.sciencetranslationalmedicine.org/cgi/content/full/10/457/eaao2966/DC1

Composition of the MI Consortium

Materials and Methods

Fig. S1. Technical workflow of the study and validation of the TREC assay in the MI cohort.

Fig. S2. Association of sjTRECs with immune cell counts and parameters.

Fig. S3. Association of sjTRECs with immune cell counts and parameters.

Fig. S4. Association of sjTREC numbers with nonheritable factors.

Fig. S5. Association of sjTREC numbers with nonheritable factors.

Fig. S6. Association of intrathymic division number with nonheritable factors.

Fig. S7. Association of thymic function parameters with specific nonheritable factors.

Fig. S8. Age and sex impact on sjTRECs in the MARTHA cohort.

Fig. S9. Human immune system mice flow cytometry gating strategy.

Fig. S10. Effect of the SNP rs2204985 polymorphism on the thymic T-cell progenitors in humanized mice.

Fig. S11. Assessment of TCR repertoire overlap.

Table S1. Demographic, medical, and lifestyle variables included in the MI study.

Table S2. Statistics of association with sjTRECs for suggestive loci and replication at the TCRA-TCRD locus.

Table S3. TCRA-TCRD next-generation sequencing data.

Table S4. Primers and probes used for sex determination and TREC quantification.

Table S5. Primary human immune system mice data (Excel file).

References (54–60)

**REFERENCES AND NOTES**

Changes in thymic function with age and during the treatment of HIV infection. 


N. Lopes, H. Vachon, J. Marie, M. Irla, Administration of RANKL boosts thymic regeneration due to p38 MAP kinase by DNA double-strand breaks in VDJ recombination induces a G2/M cell cycle checkpoint. 


Acknowledgments: We thank J. Fellay, P. Scepanovic, and C. A. W. Thorball for support with genetic analysis and J.-M. Doisne, G. M. Ranson, and H. Strick-Marchand for support with humanized mouse experiments. We thank V. Asnafi, E. Macintyre, A. Cieslak, and E. Patin for helpful discussions. We also thank the Centre d’Immunologie Humaine (Institut Pasteur, Paris, France) for support. Funding: This work was supported by the French government’s Invest in the Future Program, managed by the Agence Nationale de la Recherche (ANR; 10-LABX-69-01). I.L.A. was a recipient of the Science without Borders PhD program from Brazil Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). J.B. is a member of the Lund Center for Control of Complex Engineering Systems (LCCC) Linnaeus Center and the Excellence Center at Linköping-Lund in Information Technology (ELLIT) Excellence Center at Lund University and is supported by the ELLIT Excellence Center. J.P.D.S., Y.L., and S.L.-L. received funding from Institut Pasteur, INSERM, the Laboratoire d’Excellence REVIVE. A.T. and E.C. received funding from ANR grant PriCelAge (ANR-14-CE14-0030-01). Genetic investigations in the MARTHA study were supported by the GENMED Laboratory of Excellence on Medical Genomics (ANR-10-LABX-0013). Author contributions: A.T. and M.L.A. conceived the project. E.C., I.L.A., and C.A. contributed to the design of experiments, supervised and conducted the experiments, generated the manuscript figures, and wrote the manuscript. E.P. and J.B. supervised and conducted statistical analyses, generated the manuscript figures, and wrote the manuscript. A.U., S.L.-L., Y.L., B.C., C.R.M., M.H., B.L.M.-L., and C.D. assisted in conducting experiments. N.S., M.G., D.-A.T., and P.-E.M. provided materials and data. M.F., D.D., and J.P.D.S. contributed to the supervision of the project and data analyses. A.T. and M.L.A. supervised the project, contributed to the design of the experiments and their analyses, and wrote the manuscript. M.L.A., L.Q.-M., and A.T. secured the funding. All authors reviewed and accepted the manuscript. Competing interests: E.C., I.L.A., C.A., L.Q.-M., M.L.A., and A.T. hold a patent on “Common genetic variations at the TCRA-TCRD locus control thymic function in humans” (PCT/EP2018/055873). All other authors declare that they have no competing interests. Data and materials availability: All data associated with this study are present in the paper or the Supplementary Materials. TCR repertoire data are accessible through ImmunACCESS database under the “Genetic of Thymic function in mMice” name project (https://clients.adaptivebiotech.com/immunaccess). The SNP array data have been deposited in the European Genome-Phenome Archive with the accession code EGAS00001002460. The flow cytometric data and the code implementing statistical analyses can be downloaded as an R package (http://github.com/JacobBergstedt/mmi) and explored with the online Shiny application (http://milieu-interieur.cytogwas.pasteur.fr/).

Composition of the Milieu Intérieur Consortium: The MI Consortium is composed of the following team leaders: Laurent Abel (Institut Imagine, Paris); Andres Alcover, Hugues Aschard, and Kalle Åström (Lund University, Lund); Philippe Bousso, Pierre Bruhns, Ana Cuman, Caroline Demangel, Ludovic Deriano, James Di Santo, Françoise Dromer, Darragh Duffy, Gérard Eberl, Jost Enninga, and Jacques Fellay (EPFL, Lausanne); Magnus Fontes, Antonio Freitas, Odile Gelpi, Ivo Gomperts-Boneca, Milena Hasan, and Serge Hercberg (Université Paris 13, Paris); Olivier Lantz (Institut Curie); Claude Leclerc, Hugo Mouquet, Etienne Patin, Sandra Pellegrini, and Stanislas Pol (Hôpital Cochin); Antonio Rausell (INSERM UMR 1163 – Institut Imagine); Lars Rogge, Anavaj Sakuntabhai, Olivier Schwartz, Benno Schwikowski, Spencer Shorte, and Vassili Soumelis (Institut Curie); Frédéric Tangy and Eric Tartour (Hôpital Européen George Pompidou, Paris); Antoine Toubert (Hôpital Saint-Louis, Paris); Mathilde Touvier (Université Paris 13); Marie-Noélle Ungeheuer; Lluis Quintana-Murci; and Matthew L. Albert. Additional information can be found at http://www.milieuinterieur.fr/en.

Submitted 5 July 2017
Resubmitted 11 April 2018
Accepted 25 July 2018
Published 5 September 2018
10.1126/scitranslmed.aao2966

Human thymopoiesis is influenced by a common genetic variant within the \textit{TCRA-TCRD} locus


\textit{Sci Transl Med} \textbf{10}, eaao2966.
DOI: 10.1126/scitranslmed.aao2966

\textbf{Mining the Milieu Intérieur}

Personalized medicine partly depends on understanding what causes variance even outside the context of overt disease. The Milieu Intérieur Consortium enrolled 1000 healthy adults to study how genetics and the environment influence the immune system. Clave et al. leveraged samples from this cohort to see how thymic output, known to decrease over time, is affected by other factors. In addition to seeing sex-dependent differences, a genome-wide association study revealed variants that were associated with thymic output, which was confirmed in an independent cohort and mouse models. The authors have also developed a Web application for other investigators to examine the Milieu Intérieur data.