Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Steri M, Orrù V, Idda ML, et al. Overexpression of the cytokine BAFF and autoimmunity risk. N Engl J Med 2017;376:1615-26. DOI: 10.1056/NEJMoa1610528

Supplementary Appendix for

"Overexpression of the cytokine BAFF and autoimmunity risk"

Table of contents

Authors

Maristella Steri, Ph.D.*, Valeria Orrù, Ph.D.*, M. Laura Idda, Ph.D.*, Maristella Pitzalis, Ph.D., Mauro Pala, Ph.D., Ilenia Zara, B.Eng., Carlo Sidore, Ph.D., Valeria Faà, M.S., Matteo Floris, Ph.D., Manila Deiana, Isadora Asunis, Eleonora Porcu, Ph.D., Antonella Mulas, Ph.D., Maria G. Piras, M.S., Monia Lobina, Ph.D., Sandra Lai, M.S., Mara Marongiu, Ph.D., Valentina Serra, Ph.D., Michele Marongiu, B.S., Gabriella Sole, M.S., Fabio Busonero, Ph.D., Andrea Maschio, M.S., Roberto Cusano, M.S., Gianmauro Cuccuru, M.S., Francesca Deidda, Ph.D., Fausto Poddie, M.S., Gabriele Farina, M.D., Mariano Dei, M.S., Francesca Virdis, Ph.D., Stefania Olla, Ph.D., Maria A. Satta, M.D., Mario Pani, M.D., Alessandro Delitala, M.D., Eleonora Cocco, M.D., Jessica Frau, M.D., Giancarlo Coghe, M.D., Lorena Lorefice, M. D., Giuseppe Fenu, M.D., Paola Ferrigno, M.D., Maria Ban, Ph.D., Nadia Barizzone, Ph.D., Maurizio Leone, M.D., Franca R. Guerini, M.D., Matteo Piga, M.D., Davide Firinu, M.D., Ingrid Kockum, Ph.D., Izaura Lima Bomfim, Ph.D., Tomas Olsson, Ph.D., Lars Alfredsson, Ph.D., Ana Suarez, M.D., Patricia E. Carreira, M.D., Maria J. Castillo-Palma, M.D., Joseph H. Marcus, B.S., Mauro Congia, M.D., Andrea Angius, Ph.D., Maurizio Melis, M.D., Antonio Gonzalez, Ph.D., Marta E. Alarcón Riquelme, Ph.D., Berta M. da Silva, Ph.D., Maurizio Marchini, Ph.D., Maria G. Danieli, M.D., Stefano Del Giacco, M.D., Alessandro Mathieu, M.D., Antonello Pani, M.D., Stephen B. Montgomery, Ph.D., Giulio Rosati, M.D. †, Jan Hillert, M.D., Ph.D., Stephen Sawcer, Ph.D., Sandra D'Alfonso, Ph.D., John A. Todd, Ph.D., John Novembre, Ph.D., Gonçalo R. Abecasis, Ph.D., Michael B. Whalen, Ph.D., Maria G. Marrosu, M.D., Alessandra Meloni, Serena

Sanna, Ph.D., Myriam Gorospe, Ph.D., David Schlessinger, Ph.D., Edoardo Fiorillo, Ph.D., Magdalena Zoledziewska, Ph.D., and Francesco Cucca, M.D.

* These authors contributed equally to this work.

Correspondence to: Francesco Cucca (fcucca@uniss.it)

† Prof. Giulio Rosati passed away on 15 August 2016. This work is dedicated to his memory.

Methods

Sardinian sample sets

The primary Sardinian sample sets comprise autoimmune case-control samples and a general population cohort to dissect quantitative immune traits. Only individuals who had at least 3 Sardinian grandparents were included in the study.

The case-control samples are unrelated Sardinians collected from the main clinical and blood transfusion centers on the island. They consist of 2,934 MS and 411 SLE patients and a shared control group including 3,110 blood donors as well as 282 Affected Family BAsed Controls (AFBAC) pseudo-controls derived as those with alleles not transmitted from unaffected parents to affected children in MS disease family trios[39].

MS patients (female-to-male ratio 2.2:1) were diagnosed according to the McDonald criteria: 92.6% of them had a bout-onset disease course (mean age at onset of 31.3 ±10.51, range 5-88).

SLE patients (female-to-male ratio 6.9:1, mean age of onset 32.4 ± 13.02 – range 7-86) were recruited to fulfill at least 4 of the American College of Rheumatology (ACR) 1982 criteria for the classification of SLE.

The SardiNIA general population cohort (6,921 volunteers, 3,985 of them female, aged 18–102), has been previously described[6,13].

MS replication cohorts

To replicate the association of SNP rs12874404 with MS initially observed in Sardinia, cohorts from several populations were assembled. All MS cases were diagnosed using the McDonald criteria. The MS cohorts comprised:

i. 2,292 cases and 2,563 healthy individuals from mainland Italy. MS patients (female-to-male ratio 1.9:1, mean age of onset 31.39 ± 10.26, 89% with a relapsing remitting disease course at onset) were recruited through the Italian multicenter PROGEMUS (PROgnosticGEnetic factors in MUltiple Sclerosis) consortium.

   Controls consisted of 2,563 healthy individuals (female-to-male ratio 0.76:1) including blood donors and medical staff from the same areas where the patients were collected.

   All individuals studied were from the Continental Italian population. Individuals of Sardinian origin, or those related to enrolled volunteers, were excluded.

ii. 4,548 cases and 3,481 controls from Sweden, collected by the Karolinska Institute. Samples from Sweden were obtained from both the "Genes and Environment in Multiple Sclerosis"[40] (GEMS) and the "Epidemiological Investigation of Multiple Sclerosis"[41] (EIMS) studies. GEMS is a population-based case-control study with cases identified through the national Swedish MS Registry in 2008. EIMS is an ongoing multi-center population-based case-control study with newly diagnosed MS patients between 16 and 70 years of age recruited through 42 neurology units. Only samples collected from April 1st 2005 to May 31th 2012 were used for this study. Controls for both EIMS and GEMS were chosen from the Swedish Total Population register and matched by age, gender and area of residence. Genotypes were successfully obtained for

90.1% (4,548 cases and 3,481 controls) among 8,911 individuals. In total, 5,449 individuals were enrolled in the GEMS study, while 2,580 participated in the EIMS study. The female-to-male ratio was 2.56 among cases and 3.08 among controls. The mean age at onset was 36.59 ± 11.12 (range 8–75) and 91% of the cases had a relapsing remitting disease course at onset.

iii. 3,176 cases and 2,958 controls from the United Kingdom[2,26].

The UK cases were collected through a national recruitment project ("The genetic analysis of multiple sclerosis") coordinated by the Department of Clinical Neurosciences at the University of Cambridge. The demographics of these cases are typical, with an average age at recruitment of 48, an average age at disease onset of 33, a female-to-male ratio of 3:1, a mean EDSS of 4.9 and a primary progressive MS proportion of 14%. Controls were recruited from the NIHR Cambridge BioResource access which was approved by the Cambridge BioResource Scientific Advisory Board (http://www.cambridgebioresource.org.uk).

SLE replication cohorts

To test and extend the association with BAFF-var observed in Sardinia, SLE cohorts were assembled. All patients in the cohorts fulfilled at least 4 of the American College of Rheumatology 1982 criteria for the classification of SLE. The SLE cohorts comprised:

i. 503 cases and 2,563 controls from mainland Italy.

The SLE patients (female-to-male ratio 8.5:1, mean age of onset 30.40 ± 12.64) were recruited at four Italian centers. 2,653 healthy individuals (female: male

ratio: 0.76:1), including blood donors and medical staff, were analyzed as controls. All individuals belonged to the Continental Italian population and individuals of Sardinian origin or those related to enrolled volunteers were excluded.

ii. 1,120 cases and 1,300 controls from the Iberian Peninsula were collected by GENYO and the Hospital Clinico Universitario de Santiago.
646 SLE patients and 664 controls from GENYO, comprising individuals from Portugal and Spain, were recruited in Granada, Spain.

SLE patients and control from Hospital Clinico Universitario de Santiago were obtained at 4 Spanish and 1 Portuguese hospitals. Recruiters asked participants for their ancestry, and only those reporting ancestry completely from their respective country were included. In addition, six top ancestry informative markers for European population differentiation were checked to exclude differences between cases and controls from each recruitment center, as previously described[42]. This procedure resulted in the inclusion of 474 patients with SLE and 636 healthy controls.

The SLE controls from mainland Italy were the same ones used in the MS cohort analyses (see above).

Each study was approved by the local Ethical Review Boards and all enrolled subjects provided written informed consent.
Relevant features of these cohorts are given in **Table S1.**

Clinical data on MS and SLE patients

Information about the response to β-interferon therapy was available for 501 MS Sardinian samples. More specifically, to avoid possible bias, only MS patients treated with β-interferon as first therapy during the first 24 months after disease onset were considered. Patients were then categorized as responders, if they satisfied the Rio criteria[43], as super responders, if they were treated for more then 5 years with β-interferon without disability progression or worsening, or as non responders.

Clinical and serological manifestations were assessed in 319 Sardinian and 368 Italian SLE patients, as described in **Table S14**. In particular, anti-dsDNA autoantibodies were measured at the disease onset (at the time of diagnosis, before therapeutical treatment) in the 319 Sardinian SLE patients by RIA (Farr Assay).

Genotyping, imputation and BAFF-var characterization

To perform a GWAS for MS in Sardinia, a sample set of 2,273 cases and 2,148 controls (1,917 blood donors and 231 AFBAC from MS disease family trios) were used. Cases and controls were genotyped with the Affymetrix Genome-Wide Human SNP Array 6.0. For genotype calling, Birdseed v2 software (http://archive.broadinstitute.org/mpg/birdsuite/birdseed.html) was used. Stringent quality controls were applied as previously described[14]. A baseline scaffold of 574,519 autosomal SNPs that passed quality control filters was used for imputation.

Likewise, association analyses for quantitative variables in the *TNFSF13B* region were performed in the SardiNIA population cohort. Briefly, 6,602 samples were genotyped with four Illumina arrays (OmniExpress, ImmunoChip, Cardio-MetaboChip and ExomeChip) as previously described[6].

Imputation was performed, on a genome wide scale, on the case-control cohort using a Sardinian sequence-based reference panel of 2,120 individuals, and the software *minimac* (http://genome.sph.umich.edu/wiki/Minimac) on pre-phased genotypes, as previously carried out for the SardiNIA cohort[6]. After imputation, only the markers with imputation quality (RSQR)>0.3 for estimated MAF≥1% or >0.6 if the MAF was <1% were retained for association analyses[44] leading to ~12.2 million variants useful for analyses. This is in line with the number of variants, ~13.6 millions, we previously obtained for the SardiNIA cohort[6].

To evaluate the presence of indels in the *TNFSF13B* region, variant calling on the sequenced samples was performed with the GATK option "HaplotypeCaller" (https://www.broadinstitute.org/gatk) in the 1.5Mb window including *TNFSF13B* (chr13:108,200,000-109,700,000). Low quality variants were filtered out using GATK VQSR (https://www.broadinstitute/guide/article?id=39) and genotype refinement and haplotype phasing were performed using Beagle (faculty.washington.edu/browning/beagle/beagle.html). We generated a high quality reference panel -- which included the "BAFF-var" composite polymorphism [GCTGT/A] (hg19 chr13:g.108960380_108960384delinsA)[45] -- and used this panel for local re-imputation in both the SardiNIA and case-control cohorts. Sanger sequencing of 96 healthy carriers confirmed the presence of BAFF-var and its allelic configuration. We note that BAFF-var is called as two features in the 1000 Genomes Project variant set: an indel ([GCTG/-], rs200748895) and a SNP ([T/A], rs374039502).

Custom TaqMan assays (Thermo Fisher Scientific) were designed to genotype: SNP rs12874404 in the Sardinian, Italian, Sweden and UK MS case-control samples and BAFF-var in the Sardinian and Italian MS case-control datasets as well as in the

Sardinian, Italian and Iberian SLE case-control samples. Genotyped variants in the genotyped cohorts satisfied standard quality controls parameters: call rate ≥95%, Hardy-Weinberg P>0.05. The concordance rates between imputed and TaqMan derived genotypes in Sardinian MS case-control dataset (from GWAS) were 76% and 84% for SNP rs12874404 and BAFF-var, respectively.

Association analyses

GWAS analysis in the Sardinian MS case-controls was based on a likelihood ratio test implemented in *mach2dat* (http://csg.sph.umich.edu/abecasis/MACH/). The association model was adjusted for the first ten principal component axes (PCAs) estimated by SmartPCA software[46] using a randomly chosen set of 100,000 of the quality controlled autosomal markers outside the HLA region, in order to correct for any possible population stratification; however, no large scale substructure was observed in the samples. Also, the genomic control parameter was 1.072, indicating little if any overall inflation in the association statistics.

To prioritize MS GWAS candidate loci, we selected SNPs reaching a *P* value of $5x10^{-6}$. Association plots of the corresponding genomic regions were then inspected and discarded if the best signal was not supported by nearby SNPs consistent with association (significance at least $P<10^{-4}$). Coincident associations with signals for quantitative immune traits in the SardiNIA study[4] were specifically assessed, and SNP rs12874404 in the *TNFSF13B* gene was selected for follow-up.

Following BAFF-var imputation in the MS GWAS dataset, conditional analyses were performed in the *TNFSF13B* region by adding BAFF-var or SNP rs12874404 as covariates to the model adjusted for PCAs.

Association analyses for the extended MS Sardinia data and for the non-Sardinian replication were performed using the same statistical test employed in the GWAS analyses, implemented in PLINK v.1.9 (http://pngu.mgh.harvard.edu/purcell/plink/). No adjustment for PCA components was possible in the replication samples. In Sardinian SLE samples, the same strength of association was confirmed after adjusting for age and sex. Combined *P*-values and effect sizes for SLE were obtained by an inverse-variance meta-analysis approach, implemented in Metal (http://genome.sph.umich.edu/wiki/METAL_Program) [47].

Targeted association analyses of 10,000 variants across the *TNFSF13B* region with quantitative immune traits were carried out in the SardiNIA cohort with EPACTS [http://genome.sph.umich.edu/wiki/EPACTS]. All assessed traits were normalized with inverse-normal transformation and adjusted for sex, age and age2 as covariates. To adjust for multiple testing, Bonferroni correction was applied to the standard significance threshold (P=0.05). For each specific hypothesis, the number of independent tests was calculated as the product of the number of variants in the analyses (10,000) and the number of independent traits (corresponding to the number of absolute cell counts and Median Fluorescence Intensities – MFI, when measured).

Conditional analyses were performed for each trait that reached genome-wide significance ($P<6.09\text{x}10^{-09}$)[6] by adding either BAFF-var or SNP rs12874404 as covariates to the model adjusted for age, age squared and sex.

BAFF-var association analyses of quantitative traits (B cells and soluble BAFF [sBAFF]) in MS and SLE cases and relative controls were performed with a linear model in R software, after inverse-normal transformation for each trait and sex and age adjustment.

Quantitative trait comparisons between cases and controls were assessed with Wilcoxon-Mann Whitney test, as implemented in R.

BAFF-var frequency heterogeneity in different clinical subgroups was assessed by univariate logistic and multivariate linear regression models implemented in R; the statistical significance was determined by adjusting the standard significance threshold (P=0.05) for the number of sample sets analysed.

Population Attributable Risk

The Population Attributable Risk (PAR), a measure of disease risk attributable to an exposure or a genetic variant in the population as a whole, was calculated as:

$$PAR = f(OR - 1)/[f(OR - 1) + 1],$$

where f is BAFF-var frequency among the controls and OR is the odds ratio[48]. OR was estimated using logistic regression analysis under multiplicative, additive, recessive and dominant models of transmission of alleles, using PLINK v.1.9. Similar OR values were achieved under the different models; only results from the multiplicative model are reported for all the populations tested.

Flow cytometric measurements

Peripheral blood was drawn into heparinized tubes, then antibody-stained and analyzed by flow cytometry (FACSCanto II, BD Biosciences). To avoid time-dependent artifacts, cells were processed and stained in the recruitment center within two hours after blood collection, as previously described[4].

Circulating immune cell types -- including T cell and dendritic cell subtypes along with broad monocytes, natural killer and B cells -- encompassing 188 immune traits

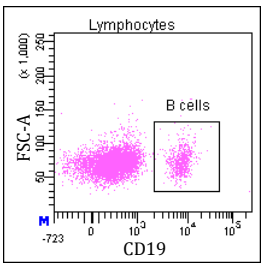
(including 66 actual counts and 122 percentages with respect to hierarchically higher cell populations), were initially characterized by FACS analysis as previously described[4] in 2,040 individuals added to the previously reported 1,629 samples, providing overall data for 3,669 immuno-phenotyped and genotyped volunteers for the analyses.

In addition, to identify B cell and monocyte subtypes associated with BAFF-var, two new antibody panels were designed, as described below, and assessed by flow cytometry in a subset of 1,902 individuals, providing estimates of 6 B cell and 3 monocyte subsets. Furthermore, for each subtype, the percentages with respect to its parental and (if available) grandparental cells were also assessed, along with the MFI of specific markers. This resulted in measures of a total of 34 B cell and 31 monocyte-related traits (see **Table S8** and **S10**).

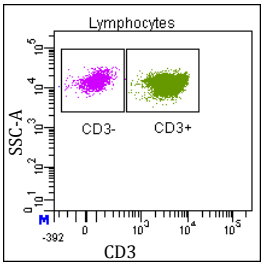
B cell classification. B cells were characterized using a 7-multicolor antibody panel (**Figure S10a-h**). Cellular aggregates were eliminated using morphology parameters (FSC-A and FSC-H).

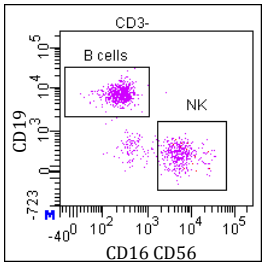
The gating strategy to identify B cells was based on preliminary tests indicating that CD19 positivity alone provides highly pure B cells. In calibration phenotyping experiments on 100 volunteers, the values for B cells obtained as CD19+ lymphocytes (plot 1) and B cells obtained with a more stringent gating strategy as CD3- CD16- CD56- CD19+ lymphocytes (plot 2-3) were 99.98% concordant.

1)



2) 3)





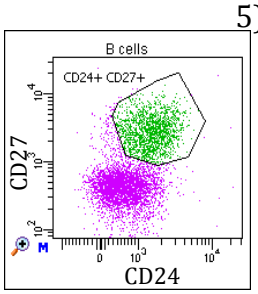
Hence, B cells were identified as CD19 positive lymphocytes.

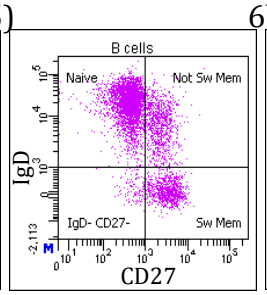
B cells were subdivided using the following classification criteria:

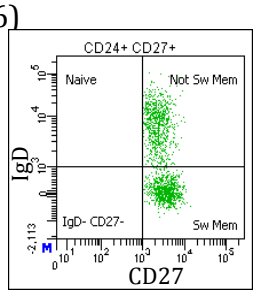
- CD24 vs CD38 classification to identify transitional B cells as CD24+ CD38hi.
- CD27 vs IgD classification to discriminate switched memory (CD27+ IgD-) un-switched memory (CD27+ IgD+), naïve (CD27- IgD+) and CD27- IgD- B cells.
- CD24 vs CD27 classification to identify CD24+ CD27+ memory cells.

This population is heterogeneous, including both IgD+ and IgD- cells. The intersection of CD24+ CD27+ cells (plot 4) with the four populations identified by CD27 vs IgD (plot 5) from 22 randomly selected SardiNIA samples shows that 50% of CD24+ CD27+ are unswitched memory (IgD+) and the remaining 50% are switched memory (IgD-) cells (plot 6).

4) 5) 6)







The MFI of CD25 and BAFF receptor (BAFF-R) were also assessed for each B cell subset.

The markers, fluorochromes and clones used are listed in **Table S2**.

Monocyte classification. Monocytes were identified by morphologic parameters and by their positivity for HLA-DR. Monocytes were then divided into classical (CD14+ CD16-), non-classical (CD14- CD16+), and intermediate (CD14+ CD16+). Each subset was assessed for CD40, CD64, CCR2, CX3CR1, and PD-L1 expression level (**Figure S10i-q**).

The markers, fluorochromes and clones used are listed in **Table S2**.

Absolute count panel. Absolute cell counts were obtained with a Lyse-No-Wash protocol using BD TruCount™ (Becton Dickinson) absolute counting tubes (see manufacturer's instructions). This 3-color panel identified B cells (as CD19+ lymphocytes) and CD14 positive monocytes. CD45 leukocyte positivity was used to discriminate cells from erythrocyte debris remaining after the Lyse-No-Wash protocol (**Figure S10r-u)**. Cells were morphologically identified as singlets or doublets in order to consider the latter as two cells in the absolute count. This counting approach improved the correlation between FACS and Cell counter enumeration for 200 samples processed the same day and compared with and without taking into account cell doublets.

The markers, fluorochromes and clones used are listed in **Table S2**.

The sums of the measures thus provided 34 B cell traits (including total B cells, 6 cell subtypes, 13 percentages with respect to parental and grandparental cell populations and 14 fluorescence intensities) and 31 monocyte traits (including total monocytes, 3 cell subtypes, 3 percentages with respect to total monocytes, and 24 fluorescence intensities) (see **Tables S8** and **S10**).

Flow Cytometry Instrument Setting and Reproducibility of Measurements were previously described[4].

MFI panel normalization. Fresh blood was sampled from volunteers over a period of about 2 years. To control for batch effects in MFIs due to variability in antibody lots and any seasonal shifts, the distribution was normalized for overall and daily changes. In detail, values for each trait were normalized by calculating the cohort mean (CM) of all the samples and the daily means (DM) of the samples analyzed on the same day. Each MFI value was then multiplied by the ratio between CM and DM to compensate for any daily fluctuation. The normalization was calculated independently for each MFI trait.

Serum quantifications for quantitative trait analyses

sBAFF was measured in serum by ELISA (R&D Systems) in:

a) 2,733 SardiNIA individuals, including the 41 volunteers with MS;
b) 76 SLE patients; and
c) 79 blood donors used as controls for SLE.

Immunoglobulin levels of IgA, IgG1, IgG2, IgG3, IgG4 and IgM were measured in 2,898 SardiNIA sera using a Bio-Plex Pro™ Human Isotyping Panel, 6-plex (Bio-Rad). IgG was calculated as the sum of the 4 IgG subclasses.

Hemocytometer parameters were quantified in up to 5,937 genotyped volunteers as previously described[13], using a COULTER LH 700 Cell Counter (Beckman Coulter, Inc).

Transcriptional studies

To monitor BAFF transcription, reads mapping to the *TNFSF13B* gene were extracted from RNA-sequencing data of leukocyte samples from 606 SardiNIA individuals. *TNFSF13B* gene-level quantifications were generated and variance stabilized with DESeq[49]; confounding factors were corrected using PEER[50].

TNFSF13B 3′ UTR length variation was measured as:

$$UTRLengthScore = \frac{C_1}{C_2}\frac{L_2}{L_1}$$

where C1 is the number of *read pairs* mapping downstream of BAFF-var. Because gene level variation among individuals must be taken into account, C1 was divided by C2, which is the number of *read pairs* mapping to the region spanning exon 6 and the full 3`UTR, with L1=1659 nt and L2=448 nt. *Read pairs* mapping to multiple genomic positions were not considered.

The 3`UTR length score might be biased because BAFF-var includes both an indel and a deletion. If so, however, the coverage profile would shift similarly upstream and downstream of the genomic position of BAFF-var, but no such shift was observed (see **Figure 2**).

Both PEER residuals and 3`UTR Length Score were inverse-normally transformed before eQTL mapping (normality was assessed by the Shapiro test, resulting in a *P* value=1 for both traits). Association analyses were performed using *Merlin* (http://csg.sph.umich.edu//abecasis/Merlin/) and *lmekin*() function from the R Software.

To establish whether the BAFF-var effect on sBAFF levels is fully accounted for by effects on transcription, a conditional linear regression model was run on 522

individuals for whom both sBAFF protein levels and RNAseq data were available. We fitted two linear models using *lmekin*() to adjust for familial clustering: one uses only BAFF-var as predictor of sBAFF levels, and the other also includes the *TNFSF13B* 3`UTR levels. The effect of BAFF-var on sBAFF levels, with and without *TNFSF13B* 3`UTR transcription levels, was assessed with a Likelihood Ratio Test.

3'UTR/microRNA effects on BAFF expression

Cell isolation. Expression studies were performed with primary monocytes, HeLa, HEK293T and THP-1 cell lines cultured at 37°C, 5% CO2. Peripheral Blood Mononuclear Cells (PBMCs) were purified from whole blood using Histopaque-1077 (Sigma-Aldrich) density centrifugation. Primary monocytes were isolated by negative selection, using an EasySep Human Monocyte Enrichment Kit (Stemcell Technologies) following manufacturer's instructions. The protocol provides ~ 90% purity, confirmed by flow cytometry.

RNA isolation and qPCR. Total RNA, including miRNA, was isolated from cells using TRIzol-RNA Isolation Reagent and the PureLink RNA mini Kit (Life Technologies), according to manufacturer's protocol. Total RNA was reverse-transcribed into cDNA using Superscript III First-Strand Synthesis System (Invitrogen) with poly-dT primer. *TNFSF13B* mRNA expression in THP1 cell lines transfected with miRNA precursor was analyzed by quantitative (qPCR) using SyBR Green mix (Thermo Fisher Scientific) and an ABI 7900HT instrument (Applied Biosystems). *GAPDH* mRNA was used as internal control and relative mRNA expression calculated by the $2^{-\Delta Ct}$ method using primers listed in **Table S3**.

miRNA expression was assessed by TaqMan qPCR following the manufacturer's protocol (Thermo Fisher Scientific). Briefly, RNA obtained from THP1, HeLa, HEK293T or primary monocytes was reverse-transcribed using TaqMan MicroRNA Reverse Transcription kit and miRNA-specific stem-loop primers. qPCR was performed using a TaqMan microRNA assay (hsa-miR-15a-5p, Thermo Fisher Scientific) and TaqMan Universal PCR Master Mix (Thermo Fisher Scientific). Relative expression of miR-15a was determined as $2^{-\Delta Ct}$ using U6 RNA levels as an endogenous control.

The efficiency of the polyadenylation signal created by BAFF-var was measured by qPCR. RNA was isolated from primary monocytes of individuals with extreme genotypes (6 homozygous WT and 7 BAFF-var). As a proxy of the efficiency of the polyadenylation signal, the proportion of the long UTR transcript (*3'UTR long*) per unit of *TNFSF13B* mRNA was used. In particular, the expression of the *3'UTR long* was quantified using the amplicon generated by primers A and B (**Figure S4**). To account for variation in absolute *TNFSF13B* RNA levels across individuals, ΔΔCt values were calculated using the signal of the amplicon spanning exons 3-4 (primers C and D); $\Delta\Delta Ct = Ct_{AB}-Ct_{CD.}$ Significance was established using a *t*-test on ΔΔCt values and the effect size was determined as effect = $1- [(2^{-\Delta\Delta Ct})_{BAFF\text{-}var}/(2^{-\Delta\Delta Ct})_{WT}]$). Primers used to amplify 3′UTR fragments are listed in **Table S3.**

"In silico" prediction of microRNA targeting of TNFSF13B 3' UTR. To identify candidate miRNAs differentially targeting BAFF 3′UTR-WT and BAFF 3′UTR-var, the results of three prediction programs were integrated: TargetScan [http://www.targetscan.org/], miRDB [http://mirdb.org/miRDB/] and RNA 22 [https://cm.jefferson.edu/rna22/].

Candidate miRNAs were sorted based on predicted possible interactions with the fragment lost in the UTR in the presence of BAFF-var (nt 2130-2582, **Figure S5a**). The miRNAs and predicted binding sites (**Table S4)** were then considered for further validation.

Pull-down assay to assess miRNA binding. To synthesize biotinylated RNA, forward PCR primers contained the T7 RNA polymerase promoter sequence. The list of PCR primers is in **Table S3**.

After purification of the PCR products, the following biotin-RNAs (bi-RNA) were synthetized using MaxiScript T7 kit (Ambion): bi-3′ WT corresponding to the WT 3′UTR of *TNFSF13B*; bi-3′ var corresponding to the 3′ UTR generated by BAFF-var; and antisense RNA fragment (bi-as) that was used to calibrate enrichment levels (**Figure S5a**). Bi-RNA pull-down assay was performed as previously described[51] using 200 μg of whole-cell lysates from THP1. The biRNA:miRNA complexes were then isolated by TRIzol-RNA (Thermo Fisher) reagent following the manufacturer's protocol. For reverse transcription, MicroRNA First-Strand Synthesis (Clontech) was used, and miRNA enrichment was then assessed by qPCR analysis with primers listed in **Table S3**. Fold enrichment for the biRNA:miRNA complexes was calculated as relative enrichment, $2^{-\Delta\Delta Ct}$, using RNA U6 as an endogenous control. Pull-down sample levels were normalized and plotted relative to an antisense bi-RNA to calibrate the results.

microRNA and LNA–anti-miR oligonucleotide transfection, protein analysis by Western Blot and ELISA. Relevant pre-miR-miRNA precursors (Ambion) and LNA–anti-miR-15a (miRCURY LNA™ Power microRNA inhibitor fluorescein labeled,

Exicon) along with relative controls, were transfected into THP1 cells with Lipofectamine 2000 (Invitrogen) following the manufacturer's protocol.

Forty-eight hours later, cells were washed and lysed and proteins concentration determined with a Bradford protein assay kit (Biorad). Proteins were separated and western blot analysis performed using primary antibody, anti-BAFF (C-term)-rabbit polyclonal antibody (Millipore), followed by incubation with horseradish peroxidase (HRP)-conjugated secondary antibody (Sigma). Blots were then probed with mouse anti-HSP90 α/β (F-8) monoclonal antibody (SantaCruz) and levels of expression corrected for loading differences. Exposed films were scanned and protein bands quantified using ImageJ Software (NIH, USA); all values were plotted relative to pre-miR-negative control sample.

Primary monocytes were seeded in 12-well plates for 48 h before transfection. LNA–anti-miR-15a and the LNA–anti-miR-control oligonucleotide were transfected with Lipofectamine 2000 (Invitrogen) following the manufacturer's protocol. Seventy-two hours later for THP1 cells and 96 h later for primary monocytes, supernatants were harvested and quantification of sBAFF levels performed using ELISA kits according to the manufacturer's instructions (human soluble BAFF, ELISA Kit (hypersensitive) AdipoGen). The optical density of known standards was used to construct a calibration curve and the cytokine values were calculated for each sample. Values were plotted relative to LNA–anti-miR control samples.

*Luciferase Constructs and reporter assay***.** A modified pmirGLO Dual-Luciferase miRNA target expression vector (Promega) was used to generate the 3′UTR constructs to perform luciferase assays. Briefly, the 3′UTR of the *Firefly* Luciferase gene was removed up to the AATAAA sequence by site-specific mutagenesis and

EcoRI restriction sites were inserted at both ends. The following constructs were produced **(Figure 4a**):

- PmirGLO-BAFF 3′UTR-WT: The full-length fragment of BAFF 3´UTR-WT, containing two putative miRNA-15a binding sequences, amplified using the primers listed in **Table S3**, incorporating *EcoRI* restriction sites. Fragments were cloned downstream of luciferase in the modified pmirGLO vector.

- PmirGLO-BAFF 3′ UTR-Var: The short BAFF-3′UTR was generated by PCR and cloned downstream of the *Firefly* luciferase, using primers listed in **Table S3**, bearing *EcoRI* restriction sites.

- PmirGLO-BAFF 3′UTR-WT with mutated miRNA15a binding sites: to generate BAFF 3′UTR-WT carrying mutated sequence in the seeding regions of miRNA-15a (TGCTGCTA), site directed mutagenesis was performed using the QuickChange II site-directed mutagenesis kit (Agilent Technologies) according to the manufacturer's instructions. The resulting fragments were cloned into the modified pmirGLO vector yielding the following constructs:

    1. pmirGLO-BAFF 3′UTR-WT – Mut. 1° binding site
    2. pmirGLO-BAFF 3′UTR-WT – Mut. 2° binding site
    3. pmirGLO-BAFF 3′UTR-WT – Mut. 1° and 2° binding sites

  All construct sequences were verified by Sanger sequencing.

Reporter assays were carried out in HeLa cells on 12-multiwell plates. Cells were co-transfected with 10 nM of either miRNA-15a precursor or miRNA-Control (Ambion) and 500 ng reporters comprising the entire generated construct. 48 h after transfection,

the relative luciferase activity was measured by Dual-Luciferase Reporter Assay System (Promega) following the manufacturer's instructions, using a Synergy 2 Plate Reader (Biotek). *Firefly* luciferase activity was normalized to Renilla luciferase activity for each transfected well and all values plotted relative to WT construct. Four independent experiments were performed in triplicate.

Statistical significance of all results was calculated by *t*-test, with $P<0.05$ considered as significant.

Statistical tests to assess BAFF-var differentiation and positive selection

All analyses were conducted on a subset of unrelated samples extracted from the 2,120 Sardinian whole-genome sequenced samples[6]. Relatedness was estimated by computing the genome-wide proportion of pairwise IBD (π) on a random set of 1 million SNPs with a minor allele frequency (MAF) >0.05 in the 1000 Genomes Project populations (Phase 3 v5)[19]. For each pair of individuals with $\pi>0.05$, offspring were preferentially removed if they occurred in a trio; otherwise, the individual with the larger summed value of π across all other relationships with $\pi>0.05$ was removed, leading to a total of 1,081 samples for the final analyses. For both the frequency-based and the haplotype-based analyses, variants with MAF<0.01 and Hardy-Weinberg proportion test *P* value$<10^{-6}$ in Sardinian samples were first removed; this conservative filter allowed a better reconstruction of haplotypes by reducing errors in genotype calls.

To assess whether BAFF-var was a target of positive selection, five standard frequency-based and haplotype-based statistical tests were used: i) $\mathbf{F_{ST}}$, which evaluates allelic differentiation between populations[15]; ii) Population-Branch Statistics (**PBS)**, which we used to estimate the magnitude of the Sardinian-specific allele frequency change[16]; then iii) **iHS**[17] and iv) **nSL**[18], both of which evaluate haplotype diversity

among allele carriers in a single population, and v) **xp-EHH**, a statistic that compares the extent of haplotype diversity in different populations[52]. Additionally, a modified version of xp-EHH was also developed applied (**as-xp-iHH**[10]) to test differences in the length of the haplotype carrying BAFF-var in Sardinians compared to other populations.

The **F_{ST} statistic** is a measure of population differentiation: an F_{ST} value close to 1 means that the maximal possible differentiation between two populations is observed at a segregating site. To calculate the F_{ST} statistic, we used the Weir-Cockerham formula implemented in *vcftools* v.0.1.12b (http://vcftools.sourceforge.net/[53]) to compare Sardinians with populations from the 1000 Genomes Project[19].

To further investigate the strength and the direction of this difference in allele frequency, we used the **PBS**, which estimates the allele frequency difference on a specific population branch of a 3-taxa population tree. PBS was calculated comparing Sardinians with Tuscans from Italy (TSI) and with British from England and Scotland (GBR) to estimate the magnitude of allele frequency change since the divergence of Sardinian and Tuscan populations. PBS values were calculated combining the F_{ST} among the three tested populations.

Among the haplotype-based tests, **iHS** is based on the ratio of integrated haplotype homozygosity for the haplotypes carrying the derived allele (iHH_D) and the ancestral allele iHH_A at a candidate site.

An alternative approach, **nSL,** has been proposed as a haplotype-based statistic for detecting, in a single population, both hard and soft sweeps. A hard sweep denotes an instance when a new advantageous mutation arises and spreads quickly to fixation due to natural selection[54]. A soft sweep indicates an instance when a neutral allele

becomes favored due to a driving force of positive selection and increases in frequency, or when multiple independent mutations at a single locus are all favored and all increase in frequency simultaneously until the sum of the frequencies is 1 but no single favored allele will reach fixation during the selective event[54]. The main difference between iHS and nSL statistics is the method used to calculate the length of a segment of haplotype homozygosity: iHS is based on the genetic distance, while nSL relies on the number of mutations in the region. Therefore, no genetic map is required to calculate the nSL statistic.

Another haplotype-based score, the **xp-EHH** (cross population-Extended Haplotype Homozygosity), compares the integrated EHH profiles between two populations at the same SNP; the xp-EHH only gains power for nearly fixed alleles, because it does not consider the specific allele under selection: this means that sampling error from haplotypes of the unselected allele is minimized when the selected allele is near fixation. Indeed, the xp-EHH test has the most elevated power for selective sweeps in which the selected allele has risen to high frequency or fixation in one population, but remains polymorphic in the human population as a whole.

The **as-xp-iHH** (allele specific - cross population - iHH), is defined as:

$$as\text{-}xp\text{-}iHH = log(iHH_{d1}/\, iHH_{d2})$$

where $d1$ and $d2$ refer to the same derived allele in populations 1 and 2. Whereas the iHS statistic compares the integrated EHH profiles between two alleles at a given SNP in the same population, the as-xp-iHH statistic compares the integrated EHH profiles of the specific allele putatively under selection between two populations. With as-xp-

iHH, we assessed evidence for the specific hypothesis that BAFF-var resides on a longer haplotype in Sardinians with respect to other populations.

The extended haplotype homozygosity (EHH) and the unstandardized iHS, nSL and xp-EHH values were calculated using *selscan* (release 1.1.0 - 07MAY2015)[55].

Empirical percentile calculation for positive selection tests. A genomic background distribution was constructed by calculating the statistics on a set of 3,042 randomly selected variants matching BAFF-var in three genetic features in Sardinians: minor allele count (MAC±10), measure of background selection (B score[56], ±50 units), and recombination rate in a 50kb region around the variant (±0.5 cM/Mb).

To estimate the local recombination rate (in cM/Mb), an interpolation of the genetic map from Delaneau et al[57] (where physical positions are in NCBI b37 coordinates) was performed. To exclude variants in LD, variant pairs with $r^2 \geq 0.1$ were filtered using the "pruning" procedure implemented in *Plink* (https://www.cog-genomics.org/plink2/).

A high-resolution genetic map based on LD patterns was then estimated by linearly interpolating the genetic map files in IMPUTE format (physical positions in NCBI b37/hg19 coordinates) from https://mathgen.stats.ox.ac.uk/impute/1000GP%20Phase%203%20haplotypes%206%20October%202014.html. The estimated genetic distances were then used when calculating iHS and xp-EHH statistics.

Results

The MS GWAS associated loci

Table S5 lists the 19 MS top associated variants. In sum, 6 loci (*TNFRSF14*, *CBLB*, *PTGER4*, *IL2RA*, *CLEC16A*) were previously reported[1,14]; 3 loci

(chr2:112464359, *FAM134B*, *MPRIP*) were already tested in the context of our previous GWAS on MS[14] (and unpublished data); and the *INPP5A* locus was reported in Burton et al.[58], with r^2=0.07 as our top variant. When coincident associations with signals for quantitative immune traits in the SardiNIA study[4] were specifically assessed, only SNP rs12874404 in the *TNFSF13B* gene was found to be associated with B cells.

MS disease-specific association analyses

After BAFF-var was identified and characterized by sequencing, imputation of the *TNFSF13B* region in the MS GWAS case-control dataset was performed using the high-quality reference panel that included the BAFF-var polymorphism. The imputation quality for BAFF-var and SNP rs12874404 was 0.68 and 0.78, respectively. The association results adjusted for the first ten principal components (OR=1.29, SE=0.06, P=5.21x10^{-06} for BAFF-var; and OR=1.26, SE=0.05, P=2.49x10^{-06} for rs12874404) were closely comparable to those obtained without such correction (OR=1.29, SE=0.06, P=2.53x10^{-06}, and OR=1.28, SE=0.05, P=1.10x10^{-06} for BAFF-var and rs12874404, respectively), indicating no impact of population stratification on the signal.

Reciprocal conditional analyses for BAFF-var and rs12874404 showed that either variant could statistically account for the entire association at the locus (BAFF-var P value after conditioning for rs12874404=0.49; rs12874404 P value after conditioning for BAFF-var=0.17; best P value after conditional analyses=3.37x10^{-03}).

Given the concordance rate between imputed and Taqman derived genotypes (84% and 76% for BAFF-var and SNP rs12874404, respectively), the association analyses were repeated replacing imputed with genotyped data, confirming the previous

obtained values (OR=1.24, SE=0.05, P=4.62x10^{-06} and OR=1.24, SE=0.05, P=3.26x10^{-06} for BAFF-var and rs12874404, respectively).

Joint analysis of "extended" MS Sardinian sample showed a strong association (OR=1.27, SE=0.04, P=1.23x10^{-09} for BAFF-var; OR=1.27, SE=0.04, P=1.52x10^{-09} for rs12874404, **Table S6**).

Lack of replication of SNP rs12874404 and statistical power estimation

With an average allele frequency of 26.5%, BAFF-var is especially common across Sardinia, but is also common throughout Southern Europe (5.7% in Italy and 4.9% in Spain), and appreciable in Northern Europe (1.8% in UK and Sweden), but is scarcely seen in South Asia and even rarer in East Asia and Africa, as shown in **Figure S8a**. Corresponding to allele frequencies, 45.8% of Sardinians, 13.5% of individuals from mainland Italy, and 3.6% in Northern Europe (UK and Sweden) carry at least one BAFF-var allele. Furthermore, the patterns of LD with surrounding SNPs differ in Sardinia vs other European populations (**Figures S1a-b**). For instance, in the tested MS sample set from Northern Europe, BAFF-var has a frequency of 1.8% and its LD with rs12874404 is weaker (r^2=0.44) than in Sardinia (r^2=0.76). These features rationalize the initial lack of replication of MS association with SNP rs12874404 in non-Sardinians. Typing this SNP in those populations would have required a sample set of 17,000 cases and a similar number of controls to have an 80% chance to detect association at a nominal level of significance (P=0.05) assuming an OR=1.2. Even directly typing BAFF-var would require 12,000 MS cases and a similar number of controls to have an 80% chance to detect association at a nominal level of significance, and 61,000 MS cases and matched controls to see a genome-wide significant association with the effect size observed in Sardinia.

Population Attributable Risk estimation

To assess the contribution of BAFF-var to the prevalence of MS and SLE in Sardinia we then estimated the population-attributable risk, taking into account its effect size and frequency[48]. Even with the moderate effect of the variant, given its high frequency in Sardinia, the calculated risk predicts a substantial contribution: if this variant were absent in Sardinia, disease prevalence would be 6.7% and 10% lower for MS and SLE, respectively (**Table S15)**.

Results of sBAFF evaluation in MS and SLE patients

To compare sBAFF levels in MS cases vs controls and avoid biases, we performed a nested case-control analysis within our SardiNIA cohort study, in which protocols and conditions of blood withdrawal, serum preparation, storage and sample processing are reproducible and are the same for cases and controls. We thus evaluated the sBAFF levels in 41 SardiNIA general population volunteers who had MS and 88 volunteers from the same study without MS or other autoimmune disease. We found sBAFF significantly increased (P=1.07x10^{-5}) in MS cases vs controls. Furthermore, when we stratified the data according to BAFF-var, the levels of sBAFF were increased in a genotype-dependent way but showed an increase in cases vs controls even in individuals who are homozygous for the wild type genotype (**Table S12**, **Figure S7a**). The longitudinal nature of our SardiNIA study also offered us the unusual opportunity to measure sBAFF levels in available sera taken during their first visit (from 2001 to 2004) of 37 of the aforementioned 41 MS patients. These included 20 volunteers whose sera was collected before the diagnosis of the disease (from less than 1 to 12 years before diagnosis, median 4.9 years). This retrospective analysis shows that sBAFF

levels in these pre-clinical individuals (median= 960.9 pg/ml) were similar to those observed in the 17 individuals who had already developed overt disease (median= 925.4 pg/ml, *P*= 0.20). Increased levels of sBAFF were also observed in a subset of 6 individuals who developed MS ≥ 10 years later (median= 1096.2 pg/ml). Taken together these findings are consistent with a primary role of increased sBAFF levels in MS pathogenesis.

Further comparisons to assess the impact of sBAFF in SLE could not rely on SLE cases in the SardiNIA longitudinal study, which were too few (N=7) to perform meaningful analyses. We therefore measured, using the same batch of ELISA kit, available sera from 76 Sardinian SLE patients (collected at the time of disease diagnosis and before therapy) and 79 controls, both from the case-control genetic analyses. We found that sBAFF was significantly increased (P=1.03x10^{-9}) in SLE cases vs controls (**Table S12**, **Figure S7b**).

Notably, as seen for MS, when we stratified the data according to BAFF-var, the levels of sBAFF increased in a genotype dependent way, and at the baseline level even in cases vs controls homozygous for the wild type genotype (i.e., lacking BAFF-var).

Results of B cell count evaluation in MS and SLE patients

B cell measures were available for 36 MS cases belonging to the SardiNIA general population cohort volunteers and who agreed to participate in systematic FACS-based immune cell profiling. In this sample set, BAFF-var showed a non-significant trend toward an increase of B cells (P=0.28 corrected for sex and age).

The number of SLE cases in our SardiNIA study was too small (N=7) to allow reliable genetic analysis of effects of BAFF-var on B cell count. However, considering the

effects of BAFF-var on the 36 MS and 7 SLE cases (43 patients) jointly, we observed a more significant trend with increased B cell count (*P*=0.097 corrected for sex and age, **Table S16**).

Dissection of B cell outliers in the SardiNIA volunteers

In 3,653 general population volunteers whose cells were FACS-analyzed, 123 individuals had B cell count at least + 3 standard deviation above the mean. In these volunteers, no shared underlying/concurrent condition was observed, but lymphoproliferative disorders, although still rare, showed an increased prevalence (3 in the 123 B outliers vs 5 in the remaining 3,530 immuno-profiled volunteers; *P*=0.002). Interestingly, in the same outliers, BAFF-var was enriched in frequency (43% vs 31% in the non-outliers, $P=2.2\times10^{-4}$), and even after removing these outlier samples from the analyses, BAFF-var remained strongly statistically associated with increased B cell count ($P=1.99\times10^{-10}$).

Stratification of MS and SLE patients based on clinical data

Considering that treatment with β-interferon could raise sBAFF blood levels[59], we analyzed the impact of BAFF-var in the response to this therapy stratifying 501 Sardinian MS patients into three cathegories: responders (N=195), super responders (N=186) and non responders (N=129), and assessing the frequency of BAFF-var in these sets of patients. No significant heterogeneity of BAFF-var frequency was observed among the three cathegories (*P*=0.13 and table below).

Cathegory (A vs B)	N samples (A / B)	BAFF-var freq in A	BAFF-var freq in B	*P* value*	OR	SE	95% CI
(responders+super responders) vs non responders	510 (381/129)	0.33	0.35	0.68	0.94	0.15	0.70-1.26

responders vs non responders	324 (195/129)	0.36	0.35	0.62	1.08	0.17	0.78-1.51
super responders vs non responders	315 (186/129)	0.30	0.35	0.20	0.80	0.17	0.57-1.12
super responders vs responders	381 (186/195)	0.30	0.36	0.05	0.74	0.15	0.55-1.00

*Significance threshold *P*=0.01

BAFF-var could also impact on SLE outcomes by increasing sBAFF levels; we thus stratified SLE patients according to different classification criteria (as reported in **Table S14**) to assess if clinical and serological manifestations preferentially correlated with BAFF-var genotypes. In particular, we analyzed up to 13 different clinical manifestations in 319 Sardinian and 368 Italian SLE patients. We found no significant correlation with any of the tested conditions (**Table S17**). Similarly, no BAFF-var association had been found with the quantitative levels of anti-dsDNA autoantibodies in the same 319 Sardinian SLE patients (*P*=0.66, effect=+0.04 s.d.).

Nevertheless, as all these results are obtained in small sample sets, they are only indicative and need to be futher confirmed in larger, and consequently more powerful, datasets.

Relative contribution of 3`UTR length and mRNA levels to sBAFF association with BAFF-var

In order to establish whether the BAFF-var effect on protein variation (sBAFF) levels is fully accounted for by effects on transcription, a conditional model was run. The analysis was restricted to the 522 individuals for whom both sBAFF levels and RNAseq data were available. Conditional analyses were carried out with the *lmekin()* function. Before the association analysis, all the traits were inverse normally transformed and sex, age and age^2 effects were removed with a linear model.

The effect of BAFF-var on the protein levels with and without the *TNFSF13B* 3`UTR (or mRNA) levels was compared with a Likelihood Ratio Test.

In the subset of 522 individuals, sBAFF was strongly associated with BAFF-var ($P<10^{-16}$, $r^2_{tot}=0.25$). Conditioning for *TNFSF13B* mRNA, the association at sBAFF was slightly decreased ($P<10^{-16}$, $r^2_{resRNA}=0.19$), but when conditioning for the 3`UTR Length Score, the amount of variance explained was drastically reduced ($P=10^{-11}$, $r^2_{resUTR}=0.06$). The reported r^2 was adjusted for the number of covariates.

This finding indicates that the proportion of sBAFF variation due to BAFF-var was mostly correlated with the variation in 3`UTR length, $1-(r^2_{resUTR}/r^2_{tot})=1-(0.06/0.25)\sim0.76$, and much less with the change in RNA level, $1-(r^2_{resRNA}/r^2_{tot})=1-(0.19/0.25)\sim0.24$. In agreement, when the same analysis was performed with *Merlin* software, the contribution of the RNA levels to sBAFF levels was 0.27.

TNFSF13B gene-level expression (PEER residuals) and UTR length scores were inversely correlated (Pearson correlation coefficient r=-0.36, $r^2=0.13$, $P=3.42\text{x}10^{-20}$).

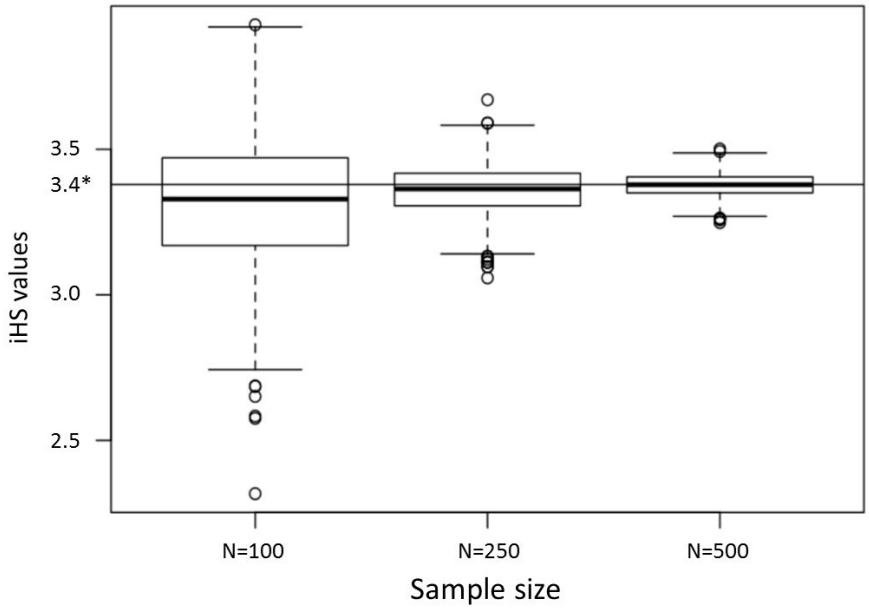
Further results of differentiation and positive selection analyses

We examined the extent of haplotype homozygosity to see if it reinforced the strong evidence for BAFF-var differentiation between Sardinians and the 1000G populations, inferred from two frequency-based tests - F_{ST} (**Table S13**) and Population Branch Statistic (**PBS**) (**Figure S8b**). We applied three haplotype-based statistics: iHS, nSL, xp-EHH, and a modified version of xp-EHH, as-xp-EHH.

Compared to haplotypes possessing similar genetic features in Sardinians, haplotypes carrying BAFF-var are significantly longer than those carrying the ancestral allele (**Table S18** and **Figure S9a**). Less significant results were found for the other

European populations (**Table S18**), although the estimate of the extent of haplotype homozygosity could be affected by smaller sample size of those population samples.

To assess the impact on EHH (and consequently on iHS) of the slight differences in sample size between Sardinians and 1000 Genomes Project populations, 1,000 simulations were performed at the *TNFSF13B* locus by randomly sampling 100, 250 and 500 Sardinian individuals (equally distributed among the SardiNIA and the MS case-control studies) to reproduce the average 1,000 Genomes Project population size (see figure below, where the asterisk in the Y axis indicates the iHS value obtained with the full set of Sardinians).



This simulation showed that the higher the sample size, the more robust the estimate of the haplotype homozygosity. Thus, we can state that the observed selection signals in the Sardinian population are robust; however, we cannot completely exclude that a stronger signal might be observed in the 1000 Genomes population if the sample size for European populations were larger.

Moreover, xp-EHH and as-xp-EHH results showed no significant differences in the length and structure of the core haplotype around BAFF-var in Sardinians compared to other European populations in which the variant is detected at appreciable frequencies **(Figure S9b** and **Table S19**).

Overall, these findings point to high frequencies of BAFF-var arising substantially as an adaptation to a selective pressure that has been relatively common in Southern Europe and extremely prevalent in Sardinia.

Discussion

BAFF and autoimmunity

Soon after the discovery of BAFF/BlyS, its effects were studied in animal models[24,25] with a strong indication for its role in autoimmunity from transgenic mice overexpressing BAFF and developing symptoms resembling human SLE and Sjögren's syndrome[11]. Later case-control studies of human autoimmune disease comparing sBAFF levels in serum and other biological fluids reported some indication of increased sBAFF in patients with rheumatic autoimmune diseases such as Sjögren's disease, SLE and Rheumatoid Arthritis, as previously described[60-62], but found no significant differences in plasma and in cerebrospinal fluid in patients versus non-MS controls[63]. Overall, the results were inconclusive and difficult to interpret in terms of causal relationships.

Part of the difficulty in comparing immune variables in patients versus controls stems from the fact that they can be secondarily confounded by the disease process and its therapy. For instance, both interferon-β and alemtuzumab have been found to augment sBAFF levels in MS patients[59]. This led to the proposal of a protective/stabilizing role of increased sBAFF in MS[64]. Likewise, it has also been

suggested that BAFF/APRIL signaling may represent a negative feedback by which the brain controls neuroinflammation[63]. In standard case-control designs for quantitative variables, problems extend to difficulties in controlling the experimental conditions, logistical and experimental, which can create biases and are difficult to detect.

More robust genetic studies comparing intrinsic DNA variation in patients versus controls are not prone to second order effects but have been similarly inconclusive. In particular, previous well-powered GWAS have provided no evidence for the role of this molecule in autoimmune disease risk[2,26]. One recent published small, under-powered study of 270 Indian cases and 555 controls described a genetic association of the BAFF region with MS susceptibility driven by the common allele at rs7318477[27], but which, as the authors pointed out, "doesn't survive Bonferroni correction". The LD between rs7318477 and BAFF-var is very low, $r2<0.04$ both in Sardinians and Europeans, suggesting that the variants are essentially unrelated. It is currently unknown whether BAFF-var is present in the Indian population. However, rs7318477 is common also in Sardinia but showed no significant association with disease or with any of the multiple related endophenotypes reported in our study. Moreover, given the low power of the study in India, a significant association with a strong and unusual effect (OR=2.5; standard error not reported) is unlikely to be accounted by an indirect effect (i.e., LD between the common allele at rs7318477 and a rare untyped causal variant).

In conclusion, the standard epidemiological approach in cases and controls has difficulties inferring causality, and the small size of many previous cohorts and the rarity of alleles of large effect size (such as BAFF-var) in most populations limited the power of genetic studies. Associations linking intrinsic genetic variation with both

disease risk and disease-related endophenotypes, with power considerably augmented by the high frequency of the large effect size BAFF-var allele in a Sardinian cohort of substantial size, provide stronger and convergent evidence about both a primary role and direction of effect of BAFF in autoimmunity. Furthermore, our genetic findings are corroborated by the unusual availability of longitudinal preclinical data, available through the SardiNIA cohort study, for MS individuals who, as reported above in "Results of sBAFF evaluation in MS and SLE patients", allowed us to show elevated sBAFF levels in individuals up to 11 years before they were diagnosed with disease.

Therapeutic implications

Researchers early on[12,65] considered that inhibition of BAFF might have therapeutic potential in some autoimmune diseases. Probably the most immediate suggestion for a role of anti-BAFF therapies for autoimmune disorders emerged from the use of BAFF receptor-Ig fusion proteins as BAFF inhibitors in lupus-prone mice, reducing disease symptoms[66]. Still, despite the approval of BAFF-targeting therapies for some forms of SLE, its use has been only partially successful, and is even now the subject of considerable controversy[67]. Based on our findings we can now anticipate that patients stratified for BAFF-var status could show differential benefit from anti-BAFF therapies. Several additional factors, however, might affect the efficacy of anti-BAFF therapies. One is their specificity, given the failure of compounds such as Atacicept, which binds both sBAFF and the related molecule sAPRIL[68] and worsens the clinical status of MS patients. It has been suggested that this failure results from interference of Atacicept with the equilibrium between natural soluble TACI and BAFF, resulting in an imbalance between effector B cells and regulatory B cells[69]. Whatever the mechanism of harmful effects of Atacicept, the current data suggest that selective

inhibition of sBAFF might be more effective, at least in the treatment of MS. A further factor could be the ability of anti-BAFF drugs to pass the blood brain barrier to reach a site where sBAFF is also produced by brain-resident cells[25] and up-regulated in lesions in MS patients[70]. These considerations provide a rationale for developing small molecules for MS and other autoimmune diseases that specifically and selectively antagonize sBAFF.

Open mechanistic issues

An interesting mechanistic issue related to anti-BAFF therapy regards the fact that we found transitional cells only marginally increased in BAFF-var subjects, whereas they were reported as a prime target of BAFF-mediated survival and were shown to promptly decrease in SLE when BAFF is blocked. How can one reconcile these apparently contradictory observations?

In line with our findings, the fact that the level of transitional cells is not the primary indicator of BAFF action is also supported by the BAFF transgenic model, which shows that the impact of the overexpression of this cytokine is more evident on mature than on immature (and hence transitional) B cells[25,71]. Moreover, studies performed on human peripheral B cells cultured with BAFF showed no significant increase of transitional cells (with the main effects again on mature B cells)[72,73]. Still, it remains puzzling that transitional B cells were not significantly augmented in peripheral blood when sBAFF was increased, despite their rapid decrease when BAFF is blocked in SLE patients[74].

Currently, we have no experimental data to explain these asymmetric effects, but we suggest a "kinetic" model in which the net effect of sBAFF *in vivo* may be the acceleration of formation of mature (naïve and memory) B cells, so that increased flow

through the pool of transitional B cells would balance their activation-increased survival, leaving the net transitional pool size roughly the same, at least in peripheral blood. Indeed, all these measures pertain to peripheral blood and may not reflect B cell levels in tissues, lymph nodes, gut or other critical sites.

Another interesting finding is the correlation of SLE risk-BAFF-var with B cell lymphocytosis, whereas B cell lymphocytopenia is also observed with SLE. How can the B lymphocytopenia observed in SLE be harmonized with the increases that we see? Our inference is that B cells (and especially mature B cells) are primarily increased in SLE patients, consistent with findings in mice that overexpress BAFF and develop symptoms resembling human SLE. One possibility is therefore that B cell lymphocytopenia in SLE patients results from secondary effects of the disease process and/or its therapy. This inference is also consistent with the efficacy of B cell depleting therapies in SLE (and with the correlation of B lymphocytopenia with better outcome)[75]. B lymphocytopenia could also result from autoantibodies binding specific antigens on the surface of human B lymphocytes[76]. Thus it seems likely that B cell lymphocytopenia could result from reverse causation.

Conclusions and prospects

Taking advantage of a natural human overexpressing model we directly assessed the broad range of effects of sBAFF in the general population and, through coincident genetic associations, linked them to MS and SLE risk. However, the effects of the overexpression of sBAFF on the downstream endophenotypes, such as specific B-cell and immunoglobulin subtypes, are broad and their involvement in MS and SLE

pathogenesis remains unclear. The next phase of work will focus on further refinement of such broad effects in autoimmune disease risk. For example, analyses to see if additional variants across the genome increase risk of autoimmunity and show coincident associations converging on specific B cell and immunoglobulin subtypes will help to narrow the range of causal relationships.

The findings may also impact future studies aimed at a better understanding of the mechanisms of naturally acquired immunity to malaria, which would be germane to reducing malaria burden by developing effective vaccines and treatments.

Author contributions

F.C. conceived and supervised the study. F.C. and D.S. drafted and revised the manuscript, M.S., V.O., M.L.I., M.Pi., M.Pala., M.F., C.S., S.Saw, J.A.T., M.B.W., J.N, G.R.A, S.San., M.G., M.Z. and E.F. contributed to write specific sections and revised the manuscript; M.Z., F.D., M. Pi., Mar.D., S.L prepared DNA samples; Ma.Maro. prepared RNA samples; M.Z. supervised array genotyping that performed along with A.Mu., F.B., and An.Ma; I.Z, G.Cu. and M.Pi. generated and administrated the Sardinian autoimmune disease database; Mi.Maro. upgraded and administrated the SardiNIA project database; F.B., An.Ma, R.C. and A.A. performed DNA and RNA sequencing; C.S. and S.San. analyzed DNA sequencing data and G.R.A supervised the analyses; E.F., V.O., M.C. and F.C. designed flow cytometric panels; V.O., E.F., F.V., Mar.D., S.L., V.S., M.Lo. and M.G.P. measured immunophenotypes; E.F. and V.O. supervised immunophenotypes measurements; M.S., I.Z., E.P., C.S., G.S. and S.San. carried out genetic association; M.Pi. and M.Z. selected TNFSF13B as a prime candidate from initial disease GWAS analysis; M.Z. and M.Pi. designed and carried out single variant Taqman genotyping and analyzed the data; M.Pala performed RNA

sequencing data analyses, and M.Pala and S.B.M designed the analyses; M.F., M.S. and J.H.M. carried out population genetic tests, and M.S. and J.N. supervised the analyses; M.F. and S.O. carried out bioinformatics analyses; M.L.I., V.F., I.A., Man.D. and A.Me. performed functional experiments; A.Me., M.L.I. and M.G. designed microRNA experiments; A.D. clinically evaluated SardiNIA immunophenotypes; F.P., G.F., M.A.S., M.Pani, E.C., J.F., G.Co., L.L., G.F., P.F., M.B., N.B., M.Le., F.R.G., I.K., I.L.B., T.O., L.A., M.Mel., M.R., S.D.A., J.H., J.A.T., S.Saw., M.G.M. provided samples and data for Multiple Sclerosis association testing and F.P., M.A.S., M.Pani, M.Piga, D.F., A.S., P.E.C., M.J.C.-P., A.G., M.A.R., B.M.D.S., M.March., M.G.D, S.D.G., Al.Ma., A.P. and S.D.A. provided samples and data for Systemic Lupus Erythematosus association testing; Al.Ma. and M.Piga supervised Systemic Lupus Erythematosus clinical stratification analyses, G.R.A., M.G., D.S., and F.C. provided funds and reagents. All authors read the paper and contributed to its final form.

Additional acknowledgements and funding

MS cohorts

Italy - University of Eastern Piedmont, Novara

The authors thank **the PROGEMUS Consortium**: Paola Naldi, Cristoforo Comi, Roberto Cantello (MS Centre, AOU "Maggiore della Carità", Novara); Marco Salvetti (Neurology and Centre for Experimental Neurologica Therapy (CENTERS), Sapienza University of Rome); Paola Cavalla (Department of Neurology, University of Turin, A.O.U. San Giovanni Battista, Turin); Maria Donata Benedetti (Policlinico G. B. Rossi, Verona); Alessia Di Sapio (II Neurological Department and Multiple Sclerosis Center, San Luigi Hospital, Orbassano, Turin); Domenico Caputo (Multiple Sclerosis Unit, Don C Gnocchi Foundation, IRCCS, S Maria Nascente, Milan); Roberto Bergamaschi

(Fondazione "Istituto Neurologico C. Mondino" IRCCS, Pavia, Italy); Daniela Galimberti (Department of Neurological Sciences, Centro Dino Ferrari, University of Milan, Fondazione Cà Granda, Ospedale Maggiore Policlinico, Milan); Elio Scarpini (Department of Neurological Sciences, Centro Dino Ferrari, University of Milan, Fondazione Cà Granda, Ospedale Maggiore Policlinico, Milan); Gabriella Rosso (Centro Sclerosi Multipla – Ospedale S. Croce e Carle, Cuneo); Susanna Cordera (SC di Neurologia Ospedale Parini di Aosta); Roberto Cavallo (San Giovanni Bosco Hospital, Turin), Claudio Solaro (Centro Sclerosi Mutipla ASL3 Genova Sestri Ponente).

UK- University of Cambridge

The work carried out by the JDRF/Wellcome Trust Diabetes and Inflammatory Laboratory was funded by the JDRF (9-2011-253), the Wellcome Trust (091157) and the National Institute of Health Research (NIHR) Cambridge Biomedical Research Centre. The Cambridge Institute of Medical Research (CIMR) is in receipt of a Wellcome Trust Strategic Award (100140).

Sweden-Karolinska Institute

The collection of samples, phenotypic data and DNA extraction of samples in the EIMS and GEMS cohorts was funded with grants from Swedish Research council, Knut and Alice Wallenberg Foundation, the AFA foundation, The Swedish Brain foundation and Swedish Research Council for Health, Working Life and Welfare.

SLE cohorts

Italy - University of Eastern Piedmont, Novara

The collection of samples was supported by the Italian Foundation for Multiple Sclerosis [FISM grants, 2011/R/14 and 2015/R/10] and Gruppo Italiano LES to Sandra D'Alfonso.

Iberia - GENYO

Samples from patients with SLE and controls were recruited with the help of Unit of Autoimmunity: José Luis Callejas Rubio and Norberto Ortego-Centeno, and Department of Rheumatology: Enrique Raya (Hospital Clínico Universitario de Granada), Unit of Autoimmunity: Maria del Mar Ayala Gutiérrez and Enrique De Ramón (Hospital Carlos Haya, Málaga), Carlos Vasconcelos (Centro Hospitalar do Porto /Hospital Santo Antonio and UMIB/ICBAS, Porto), Maria de los Angeles Aguirre, Rafaela Castro, and Eduardo Collantes (Department of Rheumatology, Hospital Reina Sofía, Córdoba), coordinated by Marta E. Alarcón-Riquelme with the help of Nieves Varela. The collection of samples was supported by FEDER funds of the European Union through the Instituto de Salud Carlos III (grant # IP/2012-02558) to Marta E. Alarcon Riquelme.

Iberia – Hospital Clinic of the University of Santiago

Samples from patients with SLE and controls were also provided by Francisco J Blanco (INIBIC-CH Universitario A Coruña, Spain), and Maria José Santos (Hospital Garcia de Orta and Rheumatology Reseach Unit, Instituto de Medicina Molecular, Lisbon, Portugal). The collection of samples was partially supported by: Fondo de Investigacion Sanitaria of the Instituto de Salud Carlos III (Spain) grants 14/01651 and RETICS Program RD12/0009/0008 that are partially financed by the European

Regional Development Fund of the European Union, to Antonio Gonzalez, Instituto de Salud Carlos III.

The authors thank Carmen Pena for taking care of the samples.

Figures

Figure S1. LD extension around rs12874404 and BAFF-var in Sardinians and Europeans.

Linkage Disequilibrium around (*a*) the initial GWAS-leading SNP rs12874404 and (*b*) the BAFF-var are shown. Each panel reports the LD calculated as r^2 (in ordinate) between the variant of interest and the other variants ± 300 kb around it (x-axis) in Sardinians (blue diamonds) and 1000 Genomes-Europeans (red squares).

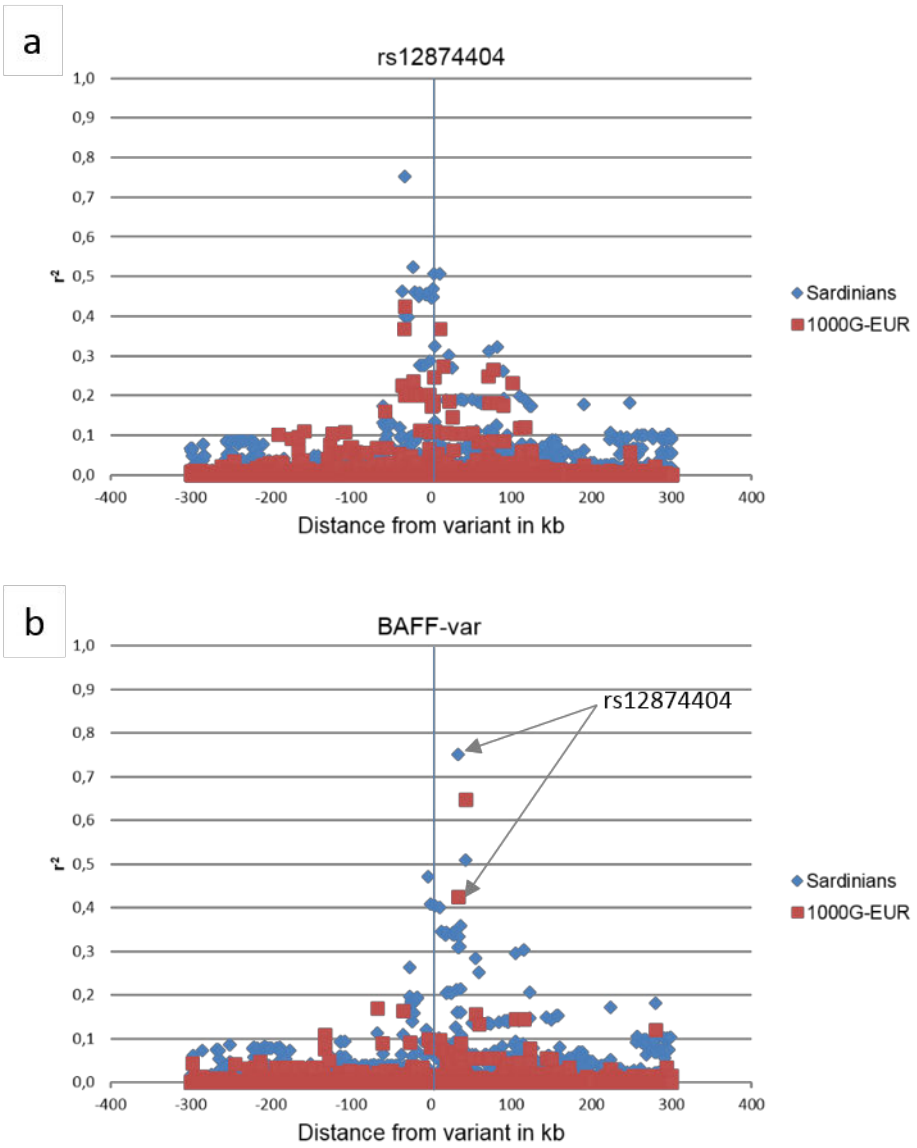


Figure S2. Immunoglobulin association plots for the *TNFSF13B* region.

Each panel represents the association strength (–log10 p-value is reported on the ordinate) versus the genomic positions around BAFF-var (in the hg19/GRCh37 genomic build, X axis). Other variants in the region are color-coded to reflect their LD with BAFF-var, as in the inset (taken from pairwise r^2 values calculated on Sardinian haplotypes). Genes and the position of exons, as well as the direction of transcription, are noted in the box below the data plots. Plots are drawn using the standalone version of the LocusZoom package[77].

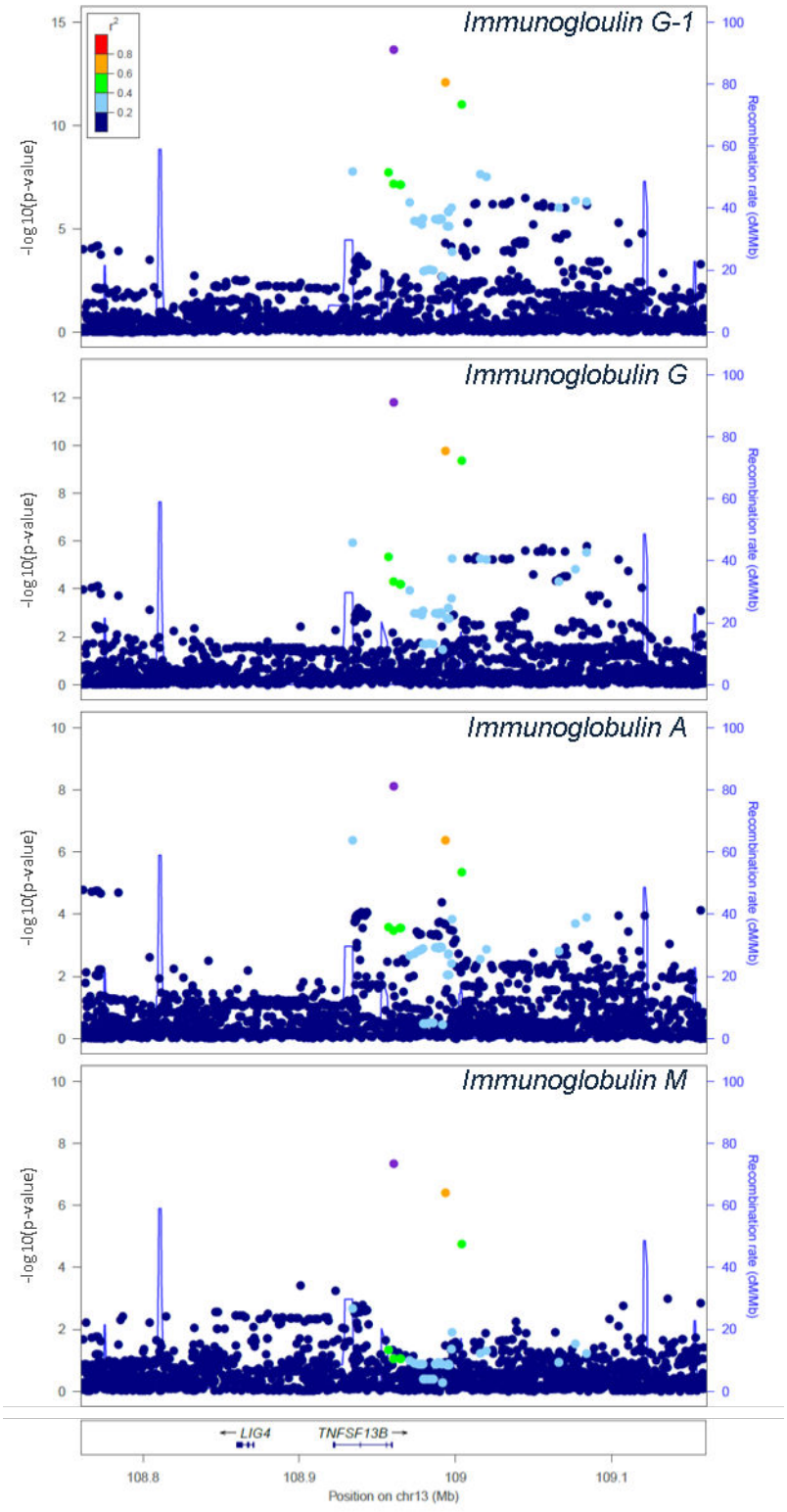

Immunogloulin G-1
Immunoglobulin G
Immunoglobulin A
Immunoglobulin M
-log10(p-value)
Recombination rate (cM/Mb)
r²
0.8
0.6
0.4
0.2
LIG4
TNFSF13B
108.8
108.9
109
109.1
Position on chr13 (Mb)

Figure S3. *TNFSF13B* mRNA and 3′ UTR length associated with BAFF-var.

Boxplots of (*a*) *TNFSF13B* RNAseq PEER corrected data; (*b*) 3′UTR length; (*c*) *P* value profile for all the variants tested *in cis* with *TNFSF13B* mRNA (PEER corrected data); (*d*) *P* value profile for all the variants tested *in cis* with *TNFSF13B* 3′ UTR length.

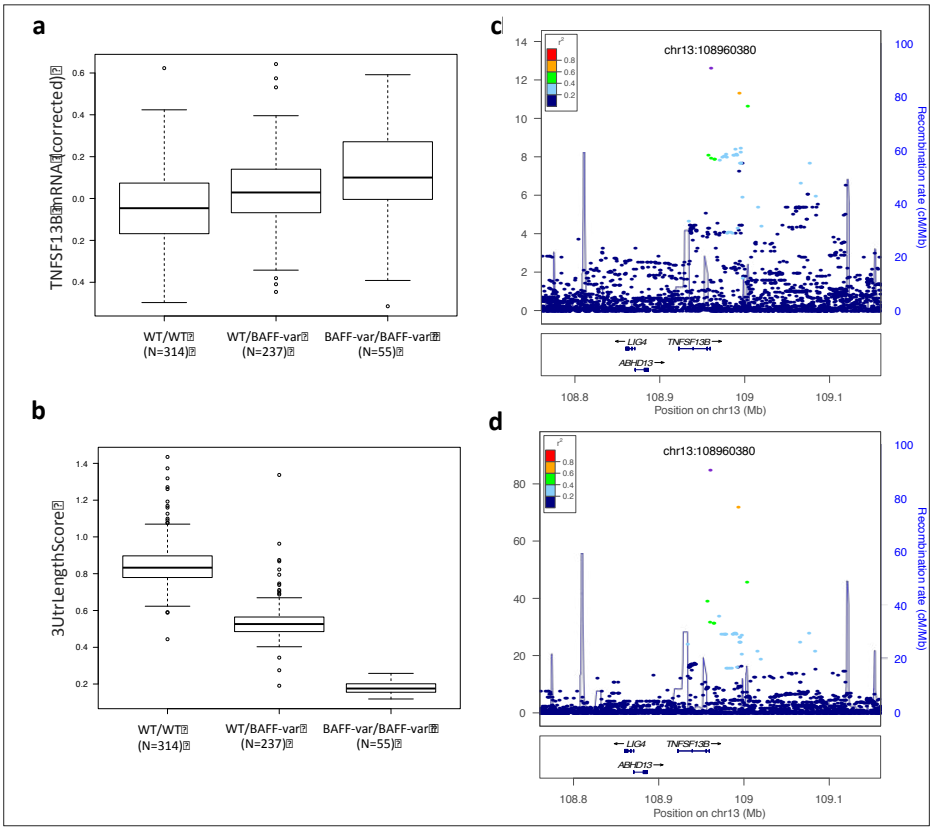


Figure S4. Quantitative analysis of the 3′UTR of *TNFSF13B* conditioned on the genotypes of interest.

(*a*) Schematic representation of the *TNFSF13B* cDNA, depicting the coding region and the 3′UTR. In WT/WT individuals (above), only the canonical polyA signal is used and one isoform is produced. In BAFF-var/BAFF-var individuals (below), the alternative polyadenylation (APA) signal is preferentially used and the short isoform predominates. Gray lines indicate the long isoform; red lines indicate the short isoform. The red arrow indicates the position of the APA. Black arrows indicate the position of the primers used for the TaqMan assay**.** (*b*) Results of TaqMan assay performed on primary monocytes from individuals carrying the extreme genotypes of interest. Level of significance is indicated with asterisks (** indicates a *P* value<0.01).

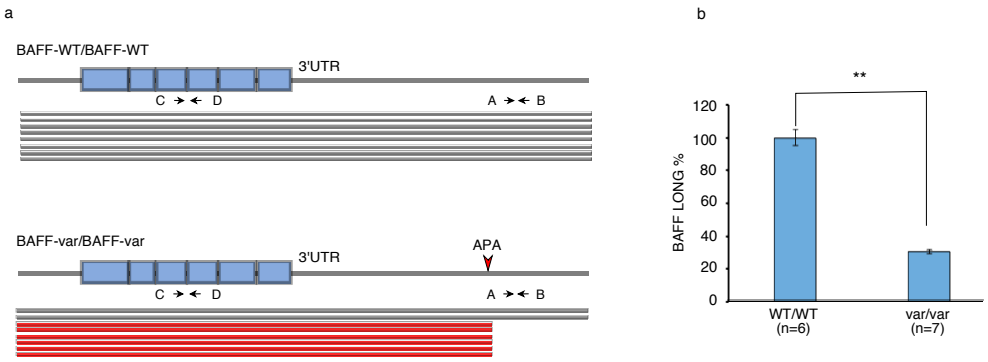


Figure S5. qRT-PCR analysis of miRNA-BAFF 3′UTR pull-downs using biotinylated RNA (bi-RNA).

(*a*) Schematic representation of *TNFSF13B* mRNA and bi-RNA used for pull-down experiments. The arrow indicates the location of the APA. CR denotes the coding region. (*b*) qPCR analysis of the pull-down complex. To calculate the fold enrichment for the miRNA:bi-RNA complexes the delta-delta Ct calculation was used. The relative enrichment with respect to the endogenous control, U6, was first calculated and the pull-down samples were then normalized to a bi-RNA antisense (as) used as calibrator.

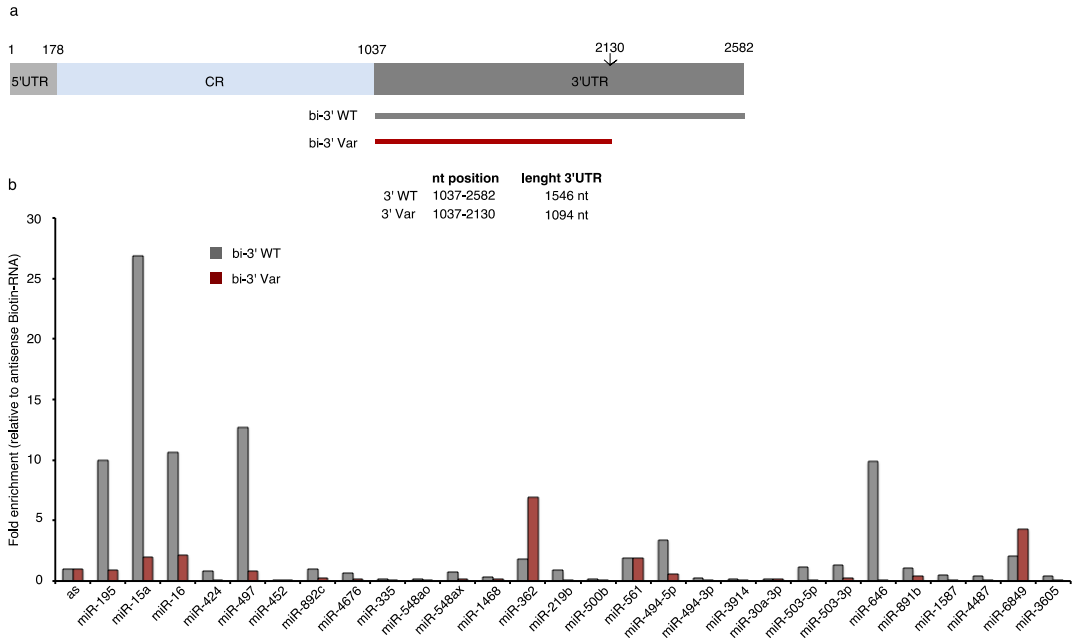


Figure S6. miR15a expression in different cell lines.

Endogenous miRNA-15a expression was measured by qPCR in HEK293T, THP1, and HeLa cells. Relative expression was determined as $2^{-\Delta Ct}$ using U6 as endogenous control. Error bars represent standard deviation of three independent experiments.

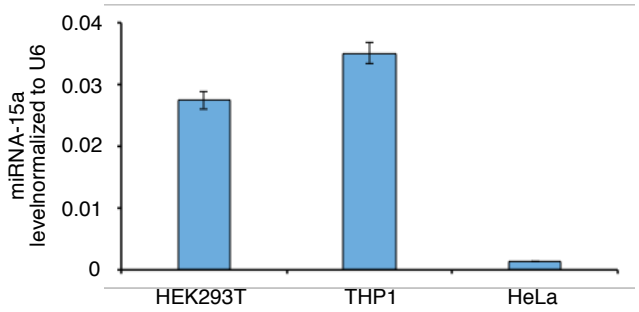


Figure S7. Comparison of sBAFF levels in MS and SLE cases versus controls.

X-axis label indicates the number of samples per class, separating the full set (ALL) from the homozygotes wild type (1/1), heterozygotes (1/2) and homozygotes for BAFF-var (2/2). *a*) Comparison between MS cases (MS) and controls (CT) from the SardiNIA cohort. *b*) Comparison between SLE cases (SLE) and controls (CT) from the case-control set. Median values and association *P* values are reported in **Table S12**.

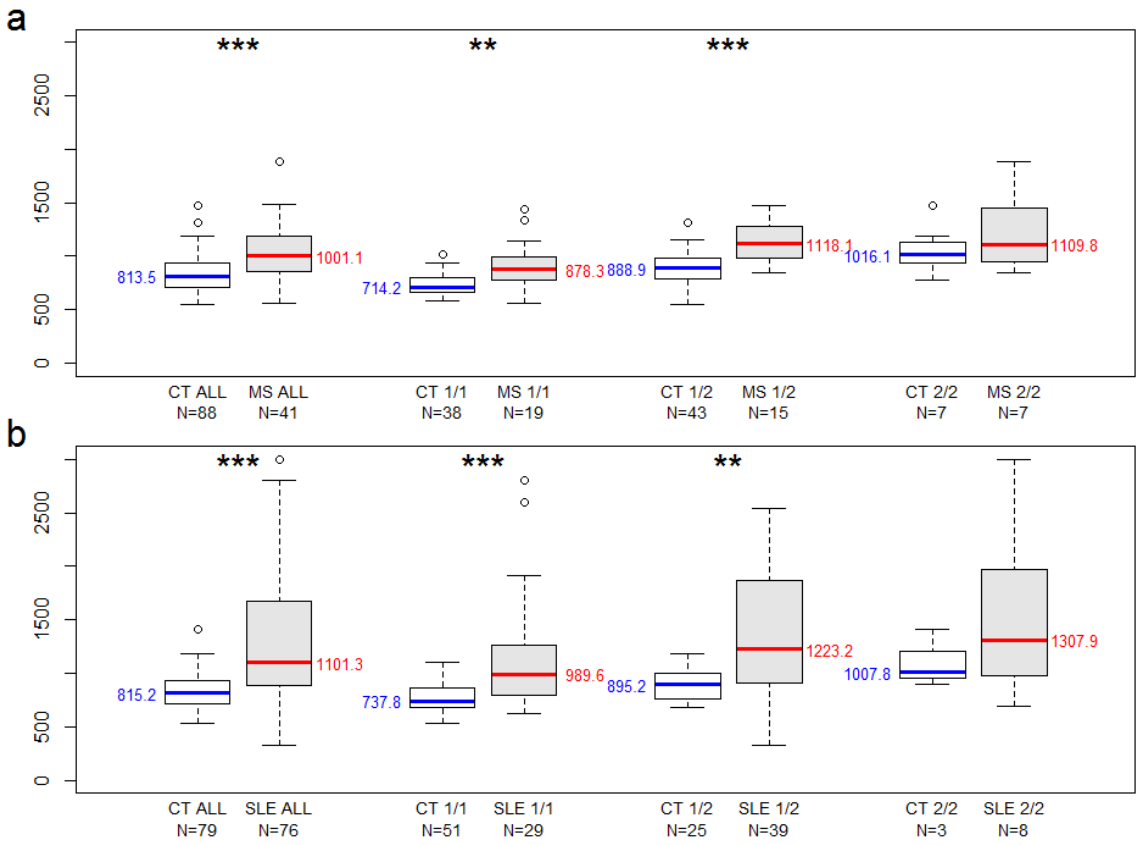


Figure S8. Frequency and haplotype-based tests for BAFF-var differentiation and selection.

(*a*) BAFF-var frequency in Sardinians (red triangle) and in 1000 Genomes Project populations (black dots). (*b*) Genomic distribution for Sardinian PBS relative to Tuscans-TSI and British-GBR from 1000 Genomes Project. (*c*) Genomic distribution of the absolute unstandardized iHS in Sardinians. In (*b,c*), the distribution for ~3,000 variants matched with BAFF-var by allele frequency in Sardinians, local recombination rate and B score is shown; the vertical dotted lines represent BAFF-var PBS and iHS values, respectively. 1000 Genomes population acronyms are specified in http://www.1000genomes.org/category/population/.

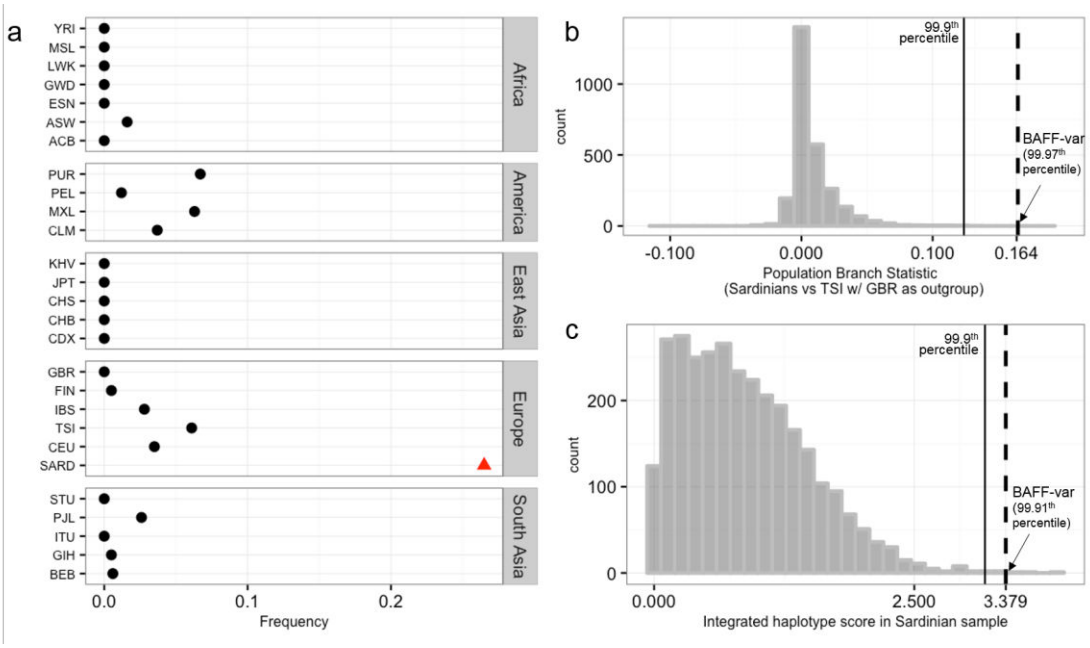


Figure S9. Haplotype-based tests for BAFF-var selection.

Distribution of: (*a*) unstandardized nSL results in Sardinians, and (*b*) unstandardized xp-EHH results (Sardinians versus CEU), each (*a,b*) based on ~3,000 BAFF-var matched variants. The vertical dotted lines represent BAFF-var values.

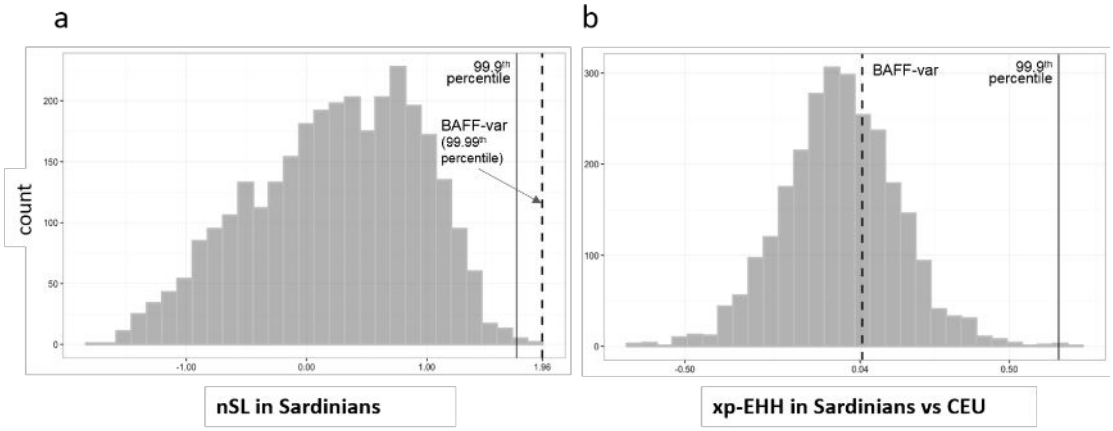
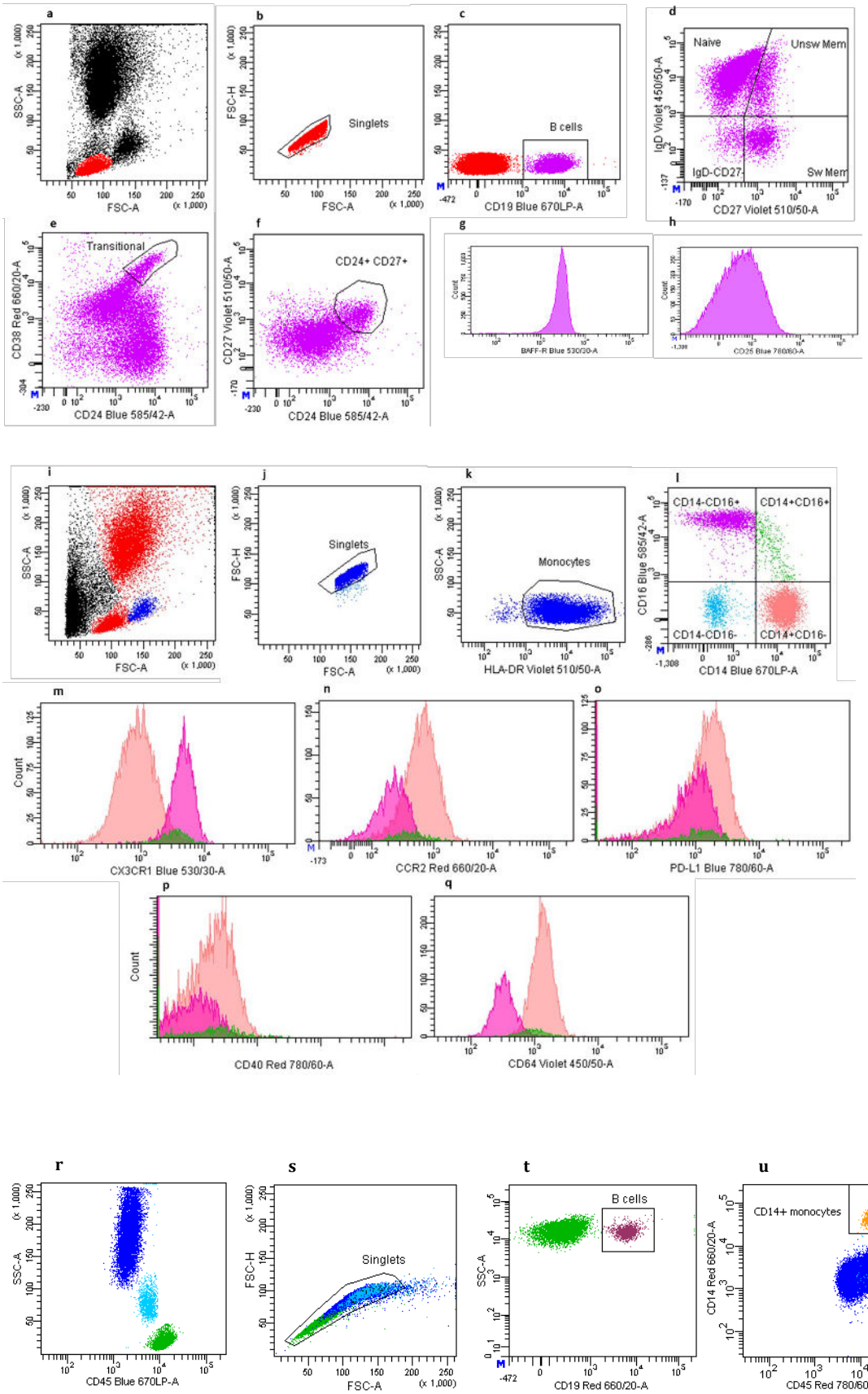


Figure S10. Gating strategy of the B cell, monocyte and absolute count flow cytometry panels.

B cell panel. (*a, b*) Lymphocyte (red) single cells were identified by morphological parameters; (*c*) B lymphocytes (violet) were isolated by positivity for CD19 antigen; (*d*) CD27 vs IgD distinguishes naïve (IgD+CD27-), unswitched memory (IgD+CD27+), switched memory (IgD-CD27+) and IgD-CD27- subsets; (*e*) CD24 vs CD38 identifies transitional B cells as CD24++CD38++; (*f*) CD24 vs CD27 discriminates CD24+CD27+ memory B cells; (*g, h*) BAFF-R and CD25 expression levels were measured on total B cells as well as on their subsets (not shown).

Monocyte panel. (*i-k*) Monocyte single cells (blue) were identified by morphological parameters and HLA DR positivity; (*l*) Monocytes were divided into CD14+CD16- (classic, light pink), CD14-CD16+ (non-classic, violet) and CD14+CD16+ (intermediate, green); (*m-q*) the three monocyte subsets were assessed for the expression levels of the chemokine receptors CX3CR1 and CCR2, the co-stimulatory molecules PD-L1 and CD40 and the pro-inflammatory protein CD64.

Absolute count panel. (*r*) Lymphocytes (green) and monocytes (light blue) were identified by morphological properties and CD45 expression level; (*s*) cell singlets were discriminated using morphologic parameters (FSC-A vs FSC-H) (*t*) B cells (purple) were identified based on their positivity for CD19 antigen; (*u*) within monocytes identified in (*r*), monocytes expressing CD14 marker (orange) were discriminated.


a
SSC-A
FSC-A
b
FSC-H
Singlets
FSC-A
c
B cells
CD19 Blue 670LP-A
d
Naive
Unsw Mem
IgD-CD27-
Sw Mem
IgD Violet 450/50-A
CD27 Violet 510/50-A
e
Transitional
CD38 Red 660/20-A
CD24 Blue 585/42-A
f
CD24+ CD27+
CD27 Violet 510/50-A
CD24 Blue 585/42-A
g
BAFF-R Blue 530/30-A
h
CD26 Blue 780/60-A
i
SSC-A
FSC-A
j
FSC-H
Singlets
FSC-A
k
SSC-A
Monocytes
HLA-DR Violet 510/50-A
l
CD14-CD16+
CD14+CD16+
CD14-CD16-
CD14+CD16-
CD16 Blue 585/42-A
CD14 Blue 670LP-A
m
Count
CX3CR1 Blue 530/30-A
n
CCR2 Red 660/20-A
o
PD-L1 Blue 780/60-A
p
Count
CD40 Red 780/60-A
q
CD64 Violet 450/50-A
r
SSC-A
CD45 Blue 670LP-A
s
FSC-H
Singlets
FSC-A
t
B cells
SSC-A
CD19 Red 660/20-A
u
CD14++ monocytes
CD14 Red 660/20-A
CD45 Red 780/60-A

Tables

Table S1. Features of the case-control cohorts.

Listed are: (1st column) the origin of the MS and SLE cohorts; (2nd, 7th) the number of patients; (3rd, 8th) the female vs male ratio in cases; (4th, 9th) the mean age at onset with the year range specified; (5th, 10th) the number of control individuals; (6th, 11th) the female vs male ratio in controls (N=3,110, excluding AFBAC). The Iberian cohort consisted of two groups, one from the University of Granada-Junta and one from the Hospital Clinico Universitario de Santiago.

	MS					SLE				
Population	***N cases***	***F:M ratio***	***Mean age at onset yr ± s.d.***	***N controls***	***F:M ratio***	***N cases***	***F:M ratio***	***Mean age at onset yr ±s.d.***	***N controls***	***F:M ratio***
Sardinians	2,934	2.2:1.0	31.3 ± 10.51	3,392	0.3:1.0*	411	6.9:1.0	32.4 ± 13.02	3,392	0.3:1.0*
Italians	2,292	1.9:1.0	31.4 ± 10.26	2,563	0.8:1.0	503	8.5:1.0	30.4 ± 12.64	2,563	0.8:1.0
Iberians	-	-	-	-	-	1,120	16.2:1.0	29.9 ± 12.89	1,300	2.4:1.0
Swedish	4,548	2.6:1.0	36.6 ± 11.12	3,481	3.1:1.0	-	-	-	-	-
UK	3,176	3.0:1.0	33.0 ± 10.00	2,958	1.0:1.0	-	-	-	-	-

*The F:M ratio is calculated on 3,110 blood donor samples

Table S2. List of markers, fluorochromes and clones used in flow cytometric measurements.

Listed, from left to right, are: the antibody; the fluorochrome; and the clone used in each flow cytometric panel.

Antibody	*Fluorochrome*	*Clone*
B cell panel		
IgD	BV421	IA6-2
CD27	V500	M-T271
BAFF-R	FITC	11C1
CD24	PE	ML5
CD19	PerCPCy5.5	SJ25C1
CD25	PECy7	2A3
CD38	APC	AB-7
Monocyte panel		
CD64	V450	10.1
HLA-DR	V500	G46-6
CD16	PE	Leu 11c
CD14	PerCPCy5.5	MφP9
PDL-1	PECy7	MIH1
CCR2	Alexa647	48607
CD40	APCH7	5C3
Cx3CR1	FITC	2A9-1
Absolute count panel		
CD19	PerCPCy5	SJ25C1
CD14	APC	MφP9
CD45	APCH7	2D1

Table S3. List of primers used for in vitro studies.

From left to right the table lists: the primer names, their sequences and the assay for which they were used.

Primer	***Sequence (5'-3')***	***Use***
bi-3´ WT-F	CCTACTTACACCATGTCTGTAG	PCR to generate biotinylated RNA fragments
bi-3´ WT-R	ATGAGGCAAAAGGCAATGAA	PCR to generate biotinylated RNA fragments
bi-3´ Var-F	CCTACTTACACCATGTCTGTAG	PCR to generate biotinylated RNA fragments
bi-3´ Var-R	ATTCTTAAGTATGGTACTTTAT	PCR to generate biotinylated RNA fragments
bi-3´ anti-sense-F	AATACTTATCTTTTATAACATGT	PCR to generate biotinylated RNA fragments
bi-3´anti-sense-R	ATGAGGCAAAAGGCAATGAA	PCR to generate biotinylated RNA fragments
A	AAACAGTAGGTGGAAAAATAGATGC	TaqMan Assay
B	CCTTGAAAAGTATTTTGATATAGATGG	TaqMan Assay
C	AAAAAGGATCTTACACATTTGTTCC	TaqMan Assay
D	TCTTTTTCTTCTAGGGCACTTCC	TaqMan Assay
BAFF 3´ UTR WT-F	TAGAATTCCTACTTACACCATGTCTGTAGCTATT	Cloning
BAFF 3´ UTR WT-R	ATGAATTCATGAGGCAAAAGGCAATGAATGTTT	Cloning
BAFF 3´ UTR Var-F	TAGAATTCCTACTTACACCATGTCTGTAGCTATT	Cloning
BAFF 3´ UTR Var-R	TAGAATTCGACAAGCAAAGCGAAGGAAACATGTTA	Cloning
BAFF-F	CACAATTCAAAGGGGCAGTAA	qPCR
BAFF-R	ACTGAAAAGGAGGGAGTGCAT	qPCR
GAPDH-F	TGCACCACCAACT GCTTAGC	qPCR
GAPDH-R	GGCATGGACTGTGGTCATGAG	qPCR

Table S4. List of predicted miRNA validated with miRNA-RNA pull down.

To identify miRNAs differentially targeting BAFF 3′ UTR-WT and BAFF 3′ UTR-Var, the results of three prediction programs: TargetScan [http://www.targetscan.org/], miRDB [http://mirdb.org/miRDB/] and RNA 22 [https://cm.jefferson.edu/rna22/] has been integrated. Wherever the programs report a miRNA- BAFF 3′ UTR interaction, the candidate miRNAs were sorted based on the interaction with the fragment lost with BAFF-var (nt 2130-2582). From left to right the table lists: the identified miRNAs, their sequences and the primers used to perform the quantitative PCR.

miRNA	*Mature sequence*	*Upstream primer sequence used for qPCR*
hsa-miR-195-5p	uagcagcacagaaauauuggc	tagcagcacagaaatattggc
hsa-miR-15a-5p	uagcagcacauaaugguuugug	tagcagcacataatggtttgtg
hsa-miR-16-5p	uagcagcacguaaauauuggcg	tagcagcacgtaaatattggcg
hsa-miR-424-5p	cagcagcaauucauguuuugaa	cagcagcaattcatgttttgaa
hsa-miR-497-5p	cagcagcacacugugguuugu	cagcagcacactgtggtttgt
hsa-miR-452-5p	aacuguuugcagaggaaacuga	aactgtttgcagaggaaactga
hsa-miR-892c-3p	cacuguuuccuuucugagugga	cactgtttcctttctgagtgga
hsa-miR-4676-3p	cacuguuucaccacuggcucuu	cactgtttcaccactggctctt
hsa-miR-335-3p	uuuuucauuauugcuccugacc	tttttcattattgctcctgacc
hsa-miR-548ao-5p	agaaguaacuacgguuuuugca	agaagtaactacggtttttgca
hsa-miR-548ax	agaaguaauugcgguuuugcca	agaagtaattgcggttttgcca
hsa-miR-1468-3p	agcaaaauaagcaaauggaaaa	agcaaaataagcaaatggaaaa
has-miR-362-5p	aauccuuggaaccuaggugugagu	aatccttggaacctaggtgtgagt

hsa-miR-219b-3p	agaauugcguuuggacaaucagu	agaattgcgtttggacaatcagt
hsa-miR-500b-5p	aauccuugcuaccugggu	aatccttgctacctgggt
hsa-miR-561-3p	caaaguuuaagauccuugaagu	caaagtttaagatccttgaagt
hsa-miR-494-5p	agguuguccguguugucuucucu	aggttgtccgtgttgtcttctct
hsa-miR-494-3p	ugaaacauacacgggaaaccuc	tgaaacatacacgggaaacctc
hsa-miR-3914	aaggaaccagaaaaugagaagu	aaggaaccagaaaatgagaagt
hsa-miR-30a-3p	cuuucagucggauguuugcagc	ctttcagtcggatgtttgcagc
hsa-miR-503-5p	uagcagcgggaacaguucugcag	tagcagcgggaacagttctgcag
hsa-miR-503-3p	gggguauuguuuccgcugccagg	ggggtattgtttccgctgccagg
hsa-miR-646	aagcagcugccucugaggc	aagcagctgcctctgaggc
hsa-miR-891b	ugcaacuuaccugagucauuga	tgcaacttacctgagtcattga
hsa-miR-1587	uuggggcugggcuggguuggg	ttgggctgggctgggttggg
has-miR-4487	agagcuggcugaagggcag	agagctggctgaagggcag
hsa-miR-6849-5p	gaguggauaggggaguguugga	gagtggataggggagtgtgtgga
has-mir 3605-5p	ugaggauggauagcaaggaagcc	tgaggatggatagcaaggaagcc

Table S5. Multiple Sclerosis GWAS results for non-HLA variants with significance level of P <5x10^{-6}.

Listed, from left to right, are: (1st column) the chromosome in which the SNP is localized; (2nd) the SNP identification number; (3rd) the chromosome position based on hg19 and dbSNP Build 137; (4th) the effect (1) and alternate (2) alleles; (5th) the effect allele frequency in the overall case-control dataset and, in brackets, in cases and controls; (6th) the imputation quality (RSQR); (7th) the effect size of the tested allele expressed in standard deviation units; (8th) the odds ratio (OR); (9th) the statistical significance (*P* value); (10th) the gene in which the SNP is localized and, in case of intergenic SNP, the nearest genes within 50kb; (11th) the SNP localization with respect to the genes; (12th) markers previously reported as associated with multiple sclerosis; (13th) correlation between associated SNPs and previously reported SNPs.

Chr	*rsID*	*Position*	*Allele1, Allele2*	*Allele1 freq (Cases, Controls)*	*Rsqr*	*Effect Allele1*	*OR*	*P value*	*Gene*	*SNP localization*	*Previously reported marker*	*r2 with rsID*
1	rs60733400	2516781	G,A	0.726 (0.748, 0.702)	0.989	0.24	1.27	1.38E-06	*TNFRSF14*	intergenic	rs4648356	0.86
2	rs7606128	101288820	A,C	0.695 (0.672, 0.719)	0.998	-0.22	0.80	2.51E-06		intergenic		
2	rs12623828	112464359	T,C	0.586 (0.561, 0.613)	1.000	-0.21	0.81	2.12E-06		intergenic		
2	rs10185078	135212421	A,G	0.939 (0.951, 0.927)	0.987	0.45	1.56	1.12E-06	*MGAT5, TMEM163*	intergenic		
2	rs4665042	160073895	C,G	0.639 (0.616, 0.663)	0.989	-0.22	0.81	1.80E-06	*TANC1*	intronic		
3	rs162735	295255	C,T	0.792 (0.812, 0.772)	0.992	0.25	1.28	3.24E-06	*CHL1*	intronic		
3	rs9657904	105586714	T,C	0.853 (0.873, 0.832)	0.996	0.31	1.37	2.09E-07	*CBLB*	intergenic	rs9657904	1.00
5	rs6887031	16586718	G,A	0.499 (0.475, 0.524)	0.999	-0.21	0.81	2.15E-06	*FAM134B*	intronic		

5	rs28886404	18509962	G,A	0.874 (0.889, 0.857)	0.915	0.31	1.36	3.85E-06		intergenic		
5	rs78077440	40323767	G,A	0.738 (0.760, 0.716)	0.996	0.23	1.26	2.48E-06	*PTGER4*	intergenic	rs6880778	0.43
6	rs62427099	144261379	G,A	0.874 (0.858, 0.891)	0.941	-0.32	0.73	2.37E-06	*PLAGL1, ZC2HC1B*	intergenic		
8	rs56309766	41031690	A,C	0.789 (0.772, 0.808)	0.900	-0.26	0.78	4.66E-06		intergenic		
10	rs12253981	6092346	G,T	0.550 (0.580, 0.519)	0.975	0.27	1.31	1.12E-09	*IL2RA*	intronic	rs7090512	0.60
10	rs2819716	134406650	T,C	0.746 (0.768, 0.722)	0.998	0.23	1.26	3.28E-06	*INPP5A*	intronic		
13	rs12874404*	108993494	A,G	0.687 (0.667, 0.710)	0.799	-0.23	0.79	4.83E-06	*TNFSF13B*	intergenic		
16	rs6498160	11199447	C,T	0.612 (0.585, 0.640)	0.999	-0.24	0.79	6.77E-08	*CLEC16A*	intronic	rs12708716	0.88
17	rs12603417	17001291	T,C	0.939 (0.928, 0.951)	0.971	-0.43	0.65	2.19E-06	*MPRIP*	intronic		
19	rs36247	8321946	T,C	0.742 (0.760, 0.722)	0.816	0.27	1.31	9.75E-07	*CERS4*	synonymous		
19	rs36075281	17924918	G,A	0.744 (0.757, 0.731)	0.362	0.38	1.46	3.91E-06	*B3GNT3, INSL3*	intergenic		

Table S6. BAFF-var association with Multiple Sclerosis (MS) and Systemic Lupus Erythematosus (SLE) in different populations.

Listed, from left to right, are: (1st column) the origin of the cohorts; (2nd) the total number of analyzed samples with the number of cases and controls specified in brackets; (3rd) the BAFF-var frequency in each cohort and (in brackets) in cases and controls; (4th) the statistical significance (*P* value); (5th) the standard error (SE); (6th) the odds ratio (OR); (7th) the 95% confidence interval (95% CI).

Cohort	*N samples (Cases, Controls)*	*BAFF-var freq (Cases, Controls)*	*P value*	*SE*	*OR*	*95% CI*
		MS				
Sardinians	6326 (2934, 3392)	0.288 (0.314, 0.265)	1.23E-09	0.0398	1.273	1.178-1.377
Italians	4855 (2292, 2563)	0.063 (0.070, 0.057)	9.84E-03	0.0868	1.251	1.055-1.483
		SLE*				
Sardinians	3803 (411, 3392)	0.275 (0.335, 0.267)	4.09E-05	0.079	1.384	1.377-1.391
Italians	3066 (503, 2563)	0.061 (0.082, 0.056)	2.25E-03	0.132	1.493	1.481-1.506
Iberians	2420 (1120, 1300)	0.061 (0.075, 0.050)	3.05E-04	0.123	1.552	1.540-1.564

* SLE combined dataset: OR = 1.44 (SE = 0.0595); *P* value = 6.74E-10; Heterogenity *P* value = 0.7064.

Table S7. Circulating Dendritic cells (cDCs), Maturation stages of T cell (MT), T-B-NK cells and Regulatory T (Treg) cell panel immunophenotypes: statistics and associate on results.

Listed, from left to right, are: (1st column) the panel name in which the trait has been analyzed; (2nd) the trait name (AC stands for Absolute Count expressed as 10^6 cells/litre); (3rd) the number of assessed samples; (4th, 5th, 6th, 7th) the minimum, mean, median and maximum trait values, respectively; (8th) BAFF-var frequency; (9th) the statistical significance (*P* value); (10th) BAFF-var effect size expressed in standard deviation units; (11th) the standard error (SE); (12th) the heritability (H2) explained. The *P* value threshold is 7.57e-08 (rows in bold), after Bonferroni correction of the nominal *P* value for 10K assessed variants and 66 absolute cell counts.

Panel	*Traits*	*N samples*	*Min*	*Mean*	*Median*	*Max*	*BAFF-var freq*	*P value*	*Effect*	*SE*	*H2 explained*
TBNK	**B cell %lymphocyte**	**3669**	**0.72**	**12.88**	**12.32**	**69.49**	**0.3147**	**9.36E-23**	**0.2627**	**0.0266**	**0.025970**
TBNK	**T/B cell**	**3653**	**0.32**	**6.58**	**5.76**	**113.10**	**0.3150**	**3.42E-21**	**-0.2546**	**0.0268**	**0.024180**
TBNK	**B cell AC**	**3653**	**16.78**	**322.50**	**286.60**	**5241.00**	**0.3150**	**4.23E-12**	**0.1832**	**0.0264**	**0.013080**
TBNK	**NK %CD3- lymphocyte**	**3669**	**6.10**	**51.03**	**50.94**	**96.03**	**0.3147**	**5.06E-11**	**-0.1713**	**0.0260**	**0.011710**
TBNK	**B cell %CD3- lymphocyte**	**3669**	**3.66**	**44.72**	**44.50**	**92.46**	**0.3147**	**6.20E-11**	**0.1699**	**0.0259**	**0.011600**
TBNK	CD3- lymphocyte %leukocyte	3669	1.42	10.84	10.37	47.54	0.3147	7.91E-07	0.1353	0.0274	0.006633
TBNK	T lymphocyte %lymphocyte	3669	22.09	70.28	71.19	93.64	0.3147	7.06E-06	-0.1181	0.0263	0.005492
TBNK	CD3- lymphocyte %lymphocyte	3669	6.36	29.72	28.81	77.91	0.3147	7.06E-06	0.1181	0.0263	0.005492
TBNK	HLA DR+ NK %CD3- lymphocyte	3596	0.00	3.55	2.80	26.34	0.3130	1.22E-05	-0.1186	0.0271	0.005316
MT	TD CD4+ AC	3395	1.42	39.41	24.22	2696.00	0.3158	3.42E-04	-0.1005	0.0280	0.003776
MT	TD CD4+ %T lymphocyte	3427	0.14	2.24	1.48	49.07	0.3158	4.18E-04	-0.0945	0.0268	0.003632
TBNK	CD14+ monocyte AC	3651	37.98	346.20	324.50	1740.00	0.3149	4.81E-04	-0.0904	0.0259	0.003338

MT	TD CD4+ %CD4+	3427	0.22	3.63	2.38	66.68	0.3158	5.68E-04	-0.0930	0.0270	0.003466
TBNK	CD4+ CD8dim %lymphocyte	3668	0.00	0.86	0.42	58.24	0.3146	8.42E-04	-0.0895	0.0268	0.003039
TBNK	CD4+ CD8dim AC	3652	0.00	22.31	10.11	4251.00	0.3149	9.07E-04	-0.0909	0.0274	0.003014
cDC	cDC AC	3387	0.36	28.28	24.68	212.60	0.3154	1.74E-03	-0.0839	0.0268	0.002897
TBNK	NKT %lymphocyte	3669	0.08	7.16	5.46	78.04	0.3147	1.92E-03	-0.0853	0.0275	0.002624
TBNK	NKT AC	3653	1.72	179.50	126.70	5697.00	0.3150	2.41E-03	-0.0828	0.0273	0.002522
TBNK	CD4+ CD8dim %T lymphocyte	3668	0.00	1.22	0.60	62.19	0.3146	4.44E-03	-0.0760	0.0267	0.002208
TBNK	HLA DR+ NK AC	3580	0.00	25.06	18.84	324.80	0.3134	4.54E-03	-0.0769	0.0271	0.002251
cDC	myeloid cDC AC	3387	0.34	17.50	13.95	190.40	0.3154	6.12E-03	-0.0763	0.0278	0.002220
TBNK	leukocyte AC	3653	1260.00	6805.00	6545.00	24800.00	0.3150	6.13E-03	-0.0731	0.0267	0.002057
cDC	plasmacytoid cDC AC	3387	0.02	10.79	9.60	50.73	0.3154	6.15E-03	-0.0716	0.0261	0.002217
TBNK	granulocyte AC	3653	588.80	3819.00	3593.00	16350.00	0.3150	6.33E-03	-0.0747	0.0273	0.002042
TBNK	CD3- lymphocyte AC	3653	67.60	730.10	679.00	5876.00	0.3150	9.10E-03	0.0707	0.0271	0.001863
TBNK	NKT %T lymphocyte	3669	0.12	10.04	7.84	83.34	0.3147	1.10E-02	-0.0695	0.0273	0.001763
TBNK	CD8+ %lymphocyte	3668	3.55	21.63	20.89	69.91	0.3146	1.28E-02	-0.0667	0.0268	0.001690
TBNK	CD8+ AC	3652	56.73	541.00	488.00	4013.00	0.3149	1.56E-02	-0.0633	0.0261	0.001604
MT	CD45RA+ CD4+ AC	3395	4.92	463.30	409.10	2732.00	0.3158	1.77E-02	-0.0617	0.0260	0.001659
TBNK	CD8br %lymphocyte	3668	3.11	19.52	18.73	68.83	0.3146	1.81E-02	-0.0633	0.0268	0.001525
TBNK	T lymphocyte AC	3653	306.10	1743.00	1661.00	6836.00	0.3150	1.85E-02	-0.0627	0.0266	0.001520
MT	CM CD4+ %CD4+	3427	0.00	28.35	27.79	61.33	0.3158	1.93E-02	0.0650	0.0278	0.001599
TBNK	CD8br AC	3652	51.94	489.30	435.40	3864.00	0.3149	2.01E-02	-0.0608	0.0261	0.001481
Treg	CD4+ Treg AC	3407	7.33	60.31	55.15	220.20	0.3159	2.58E-02	-0.0622	0.0279	0.001459
MT	CM CD4+ %T lymphocyte	3427	0.00	17.76	17.20	47.38	0.3158	3.28E-02	0.0603	0.0282	0.001331
TBNK	HLA DR+ NK %NK	3596	0.00	7.20	5.66	61.14	0.3130	3.61E-02	-0.0580	0.0277	0.001223
MT	CD45RA+ (CD4- CD8-) %T lymphocyte	3427	0.02	1.04	0.76	14.81	0.3158	3.62E-02	0.0565	0.0270	0.001282
Treg	CD45RA+ CD4+ not Treg AC	3408	5.48	431.70	378.70	2443.00	0.3160	3.82E-02	-0.0534	0.0257	0.001262

Treg	CD28- CD8br AC	3408	5.26	190.60	132.50	3594.00	0.3160	3.86E-02	-0.0570	0.0276	0.001257
TBNK	CD4+ %lymphocyte	3668	7.59	43.82	43.93	73.00	0.3146	4.05E-02	-0.0546	0.0266	0.001145
Treg	secreting CD4+ Treg AC	3407	3.92	33.06	30.23	139.30	0.3159	4.11E-02	-0.0580	0.0284	0.001226
TBNK	NK AC	3653	19.24	378.60	333.00	3766.00	0.3150	4.19E-02	-0.0535	0.0263	0.001134
MT	naive CD4+ AC	3395	0.00	423.90	369.20	2303.00	0.3158	4.37E-02	-0.0522	0.0259	0.001199
TBNK	CD14+ monocyte %leukocyte	3667	0.68	5.12	4.98	15.49	0.3146	4.41E-02	-0.0516	0.0256	0.001106
MT	CD45RA- CD4+ %T lymphocyte	3427	9.35	36.49	35.53	73.25	0.3158	4.81E-02	0.0519	0.0262	0.001141
TBNK	NK %lymphocyte	3669	1.25	15.61	14.20	65.36	0.3147	5.35E-02	-0.0491	0.0254	0.001017
MT	TD CD8br AC	3395	0.47	75.65	48.61	3252.00	0.3158	5.53E-02	-0.0540	0.0281	0.001083
Treg	activated CD4+ Treg AC	3407	2.24	15.45	13.83	71.88	0.3159	5.56E-02	-0.0541	0.0283	0.001076
MT	TD (CD4- CD8-) %T lymphocyte	3427	0.02	0.91	0.62	14.20	0.3158	5.88E-02	0.0525	0.0278	0.001043
MT	CD45RA- CD4+ %CD4+	3427	19.79	58.92	58.57	99.18	0.3158	6.96E-02	0.0469	0.0258	0.000962
MT	CD45RA+ CD4+ %CD4+	3427	0.82	41.08	41.43	80.21	0.3158	6.96E-02	-0.0469	0.0258	0.000962
TBNK	HLA DR+ T lymphocyte AC	3580	9.45	85.81	66.37	1826.00	0.3134	7.24E-02	-0.0477	0.0266	0.000903
MT	CD45RA+ CD8br AC	3395	5.91	203.00	172.90	3260.00	0.3158	7.78E-02	-0.0419	0.0237	0.000917
Treg	CD45RA- CD4+ not Treg %T lymphocyte	3440	9.33	34.37	33.67	72.17	0.3159	8.11E-02	0.0453	0.0259	0.000885
MT	CD45RA- (CD4- CD8-) %CD4- CD8-	3427	4.83	74.38	78.41	98.71	0.3158	8.15E-02	-0.0482	0.0277	0.000886
MT	CD45RA+ (CD4- CD8-) %CD4- CD8-	3427	1.29	25.62	21.59	95.17	0.3158	8.15E-02	0.0482	0.0277	0.000886
TBNK	CD4+ AC	3652	161.60	1081.00	1030.00	3567.00	0.3149	8.35E-02	-0.0466	0.0269	0.000821
TBNK	lymphocyte %leukocyte	3669	9.15	36.80	36.81	69.52	0.3147	8.45E-02	0.0467	0.0271	0.000812
Treg	CD45RA- CD28+ CD8br %CD8br	3440	2.26	33.41	31.65	83.01	0.3159	8.65E-02	0.0469	0.0274	0.000855
TBNK	CD4+ CD8br %T lymphocyte	3668	0.00	0.49	0.34	36.68	0.3146	8.92E-02	0.0455	0.0268	0.000788
TBNK	HLA DR++ CD14+ monocyte %CD14+ monocyte	3597	0.18	12.05	11.50	41.78	0.3132	9.18E-02	0.0462	0.0274	0.000791

Treg	CD45RA+ CD28+ CD8dim %T lymphocyte	3440	0.00	0.27	0.22	4.79	0.3159	9.18E-02	0.0430	0.0255	0.000827
MT	EM (CD4- CD8-) %CD4- CD8-	3427	4.83	72.89	76.66	98.55	0.3158	9.32E-02	-0.0466	0.0277	0.000823
Treg	CD4+ not Treg AC	3408	150.60	1011.00	961.70	3268.00	0.3160	9.44E-02	-0.0464	0.0277	0.000822
TBNK	HLA DR+ T lymphocyte %lymphocyte	3596	0.53	3.49	2.85	35.47	0.3130	9.52E-02	-0.0428	0.0256	0.000775
Treg	resting CD4+ Treg AC	3407	0.03	13.19	10.66	104.00	0.3159	9.93E-02	-0.0415	0.0252	0.000799
cDC	CD62L- plasmacytoid cDC %cDC	3307	0.00	0.79	0.56	14.02	0.3129	1.04E-01	0.0451	0.0277	0.000802
Treg	CD45RA- CD28+ CD8dim AC	3408	0.67	26.09	20.70	275.70	0.3160	1.06E-01	-0.0401	0.0248	0.000770
MT	TD (CD4- CD8-) %CD4- CD8-	3427	0.70	22.14	17.91	95.17	0.3158	1.14E-01	0.0440	0.0278	0.000731
Treg	CD28- CD8br %CD8br	3440	1.56	35.72	33.72	97.00	0.3159	1.23E-01	-0.0381	0.0247	0.000693
Treg	CD45RA+ CD28+ CD8dim %CD8dim	3440	0.00	8.96	7.06	73.73	0.3159	1.26E-01	0.0413	0.0270	0.000680
cDC	CD62L- myeloid cDC AC	3387	0.10	5.44	1.93	167.20	0.3154	1.28E-01	-0.0434	0.0285	0.000684
MT	CM CD8br %CD8br	3427	0.00	4.82	4.21	57.73	0.3158	1.31E-01	0.0410	0.0271	0.000666
Treg	CD28- CD8br %T lymphocyte	3440	0.44	10.64	8.39	75.07	0.3159	1.31E-01	-0.0404	0.0268	0.000663
TBNK	CD4+/CD8+	3652	0.13	2.50	2.25	11.54	0.3149	1.41E-01	0.0403	0.0273	0.000596
MT	naive (CD4- CD8-) %CD4- CD8-	3427	0.00	3.48	2.24	39.76	0.3158	1.43E-01	0.0363	0.0247	0.000628
Treg	CD25hi CD4+ AC	3407	14.32	173.20	157.80	750.30	0.3159	1.45E-01	-0.0406	0.0278	0.000624
TBNK	CD4- CD8dim AC	3652	2.36	52.63	40.48	696.30	0.3149	1.59E-01	-0.0371	0.0263	0.000543
MT	TD CD8br %T lymphocyte	3427	0.08	4.27	2.95	69.19	0.3158	1.62E-01	-0.0385	0.0275	0.000571
MT	CD45RA+ CD4+ %T lymphocyte	3427	0.38	25.99	25.54	64.37	0.3158	1.65E-01	-0.0376	0.0271	0.000562
MT	naive (CD4- CD8-) %T lymphocyte	3427	0.00	0.13	0.08	1.48	0.3158	1.67E-01	0.0311	0.0225	0.000558
MT	TD CD8br %CD8br	3427	0.77	14.40	11.34	87.32	0.3158	1.70E-01	-0.0371	0.0270	0.000551
Treg	CD45RA- CD4+ not Treg %CD4+	3440	17.92	55.84	55.44	92.51	0.3159	1.72E-01	0.0348	0.0255	0.000543
MT	CM CD8br %T lymphocyte	3427	0.00	1.24	1.08	14.77	0.3158	1.73E-01	0.0364	0.0267	0.000543
TBNK	Granulocyte %leukocyte	3669	23.31	55.57	55.58	88.14	0.3147	1.77E-01	-0.0369	0.0273	0.000498

Treg	CD45RA+ CD4+ not Treg %CD4+	3440	1.02	38.45	38.76	77.74	0.3159	1.77E-01	-0.0344	0.0254	0.000531
TBNK	HLA DR++ CD14+ monocyte AC	3581	0.47	41.51	37.42	245.60	0.3135	1.84E-01	-0.0351	0.0264	0.000494
MT	naive CD4+ %CD4+	3427	0.00	37.45	37.82	77.88	0.3158	1.87E-01	-0.0333	0.0252	0.000509
MT	EM CD8br AC	3395	7.78	264.00	220.90	3221.00	0.3158	2.02E-01	-0.0362	0.0283	0.000481
MT	naive CD8br AC	3395	0.00	127.30	99.97	917.30	0.3158	2.06E-01	-0.0235	0.0186	0.000473
TBNK	CD4- CD8dim %lymphocyte	3668	0.06	2.14	1.71	22.52	0.3146	2.09E-01	-0.0341	0.0271	0.000431
MT	CD45RA- CD8br AC	3395	17.75	285.60	241.10	3236.00	0.3158	2.23E-01	-0.0346	0.0284	0.000438
TBNK	CD4+/CD8br	3652	0.13	2.62	2.35	12.58	0.3149	2.41E-01	0.0320	0.0272	0.000377
TBNK	CD4+ %T lymphocyte	3668	10.39	62.58	63.16	88.24	0.3146	2.52E-01	0.0308	0.0269	0.000359
Treg	CD45RA- CD28+ CD8br %T lymphocyte	3440	0.76	8.75	8.12	34.33	0.3159	2.61E-01	0.0312	0.0277	0.000368
TBNK	CD8+ %T lymphocyte	3668	4.19	30.60	29.87	83.74	0.3146	2.71E-01	-0.0295	0.0268	0.000331
Treg	CD39+ resting CD4+ Treg %resting CD4+ Treg	3439	0.00	4.99	3.73	96.97	0.3158	2.72E-01	0.0298	0.0271	0.000352
Treg	CD45RA+ CD28+ CD8br AC	3408	0.00	145.40	117.10	953.00	0.3160	2.78E-01	-0.0209	0.0193	0.000345
MT	naive CD4+ %T lymphocyte	3427	0.00	23.75	23.28	63.39	0.3158	2.96E-01	-0.0276	0.0264	0.000319
TBNK	CD8br %T lymphocyte	3668	4.13	27.63	26.72	79.76	0.3146	3.08E-01	-0.0274	0.0268	0.000284
cDC	CD62L- cDC %cDC	3410	1.98	19.29	14.73	98.31	0.3155	3.15E-01	0.0282	0.0280	0.000297
TBNK	CD4+ CD8br %lymphocyte	3668	0.00	0.34	0.24	27.41	0.3146	3.15E-01	0.0269	0.0268	0.000276
TBNK	CD4- CD8- AC	3652	4.58	88.56	67.45	1338.00	0.3149	3.15E-01	-0.0257	0.0256	0.000277
Treg	CD39+ CD8br %CD8br	3440	0.00	2.36	1.78	36.57	0.3159	3.27E-01	0.0272	0.0278	0.000280
MT	EM (CD4- CD8-) AC	3395	0.97	63.22	44.10	1044.00	0.3158	3.32E-01	-0.0255	0.0263	0.000277
MT	CD45RA- (CD4- CD8-) AC	3395	0.97	64.16	45.21	1051.00	0.3158	3.36E-01	-0.0252	0.0262	0.000273
Treg	CD39+ secreting CD4+ Treg AC	3407	0.22	10.25	9.14	68.55	0.3159	3.56E-01	-0.0262	0.0284	0.000251
Treg	CD45RA+ CD4+ not Treg %T lymphocyte	3440	0.35	24.17	23.67	61.53	0.3159	3.66E-01	-0.0242	0.0267	0.000238
MT	CD45RA+ (CD4- CD8-) AC	3395	0.39	18.17	12.60	375.50	0.3158	3.66E-01	0.0243	0.0269	0.000241

MT	EM CD4+ AC	3395	50.17	314.00	288.50	1642.00	0.3158	3.68E-01	-0.0254	0.0282	0.000239
cDC	CD86+ cDC %cDC	3409	0.36	10.91	6.72	81.91	0.3154	3.85E-01	0.0246	0.0283	0.000222
Treg	CD45RA- CD25hi CD4+ not Treg %T lymphocyte	3439	0.77	5.95	5.37	28.95	0.3158	3.88E-01	0.0244	0.0282	0.000217
cDC	CD62L- cDC AC	3387	0.50	7.38	3.84	169.30	0.3154	3.89E-01	-0.0243	0.0282	0.000220
Treg	CD28- CD8dim AC	3408	0.41	35.39	20.78	729.50	0.3160	3.98E-01	-0.0237	0.0280	0.000210
Treg	CD45RA- CD25hi CD4+ not Treg %CD4+	3439	1.63	9.56	8.78	37.21	0.3158	4.03E-01	0.0236	0.0282	0.000204
TBNK	lymphocyte AC	3653	472.40	2473.00	2366.00	8229.00	0.3150	4.17E-01	-0.0220	0.0271	0.000181
Treg	CD45RA- CD4+ not Treg AC	3408	90.54	579.00	547.20	2249.00	0.3160	4.18E-01	-0.0231	0.0285	0.000193
MT	TD (CD4- CD8-) AC	3395	0.39	15.87	10.38	375.50	0.3158	4.21E-01	0.0222	0.0276	0.000191
Treg	CD39+ CD4+ Treg AC	3407	0.44	19.56	17.99	103.80	0.3159	4.28E-01	-0.0223	0.0281	0.000184
TBNK	CD4+ CD8br AC	3652	0.00	8.29	5.60	428.90	0.3149	4.37E-01	0.0209	0.0270	0.000165
TBNK	CD4- CD8- %lymphocyte	3668	0.23	3.56	2.84	30.38	0.3146	4.43E-01	-0.0200	0.0260	0.000160
cDC	CD86+ cDC AC	3386	0.07	4.54	1.72	134.40	0.3153	4.48E-01	-0.0217	0.0286	0.000170
Treg	secreting CD4+ Treg %CD4+	3439	0.69	3.15	3.03	8.91	0.3158	4.51E-01	-0.0207	0.0274	0.000166
Treg	CD4+ not Treg %T lymphocyte	3440	9.68	58.54	59.17	85.01	0.3159	4.54E-01	0.0208	0.0279	0.000163
MT	EM CD4+ %T lymphocyte	3427	4.34	18.72	17.69	55.67	0.3158	4.57E-01	0.0191	0.0256	0.000162
Treg	resting CD4+ Treg %CD4+	3439	0.02	1.19	1.06	6.13	0.3158	4.70E-01	-0.0180	0.0249	0.000152
Treg	CD39+ CD4+ AC	3408	0.04	55.10	52.95	312.70	0.3160	4.71E-01	-0.0202	0.0280	0.000153
MT	CM (CD4- CD8-) AC	3395	0.00	0.94	0.61	13.18	0.3158	4.74E-01	-0.0167	0.0234	0.000151
MT	CD45RA+ CD8br %T lymphocyte	3427	0.59	11.26	10.58	69.27	0.3158	4.86E-01	-0.0170	0.0243	0.000142
Treg	CD39+ activated CD4+ Treg %activated CD4+ Treg	3439	3.68	60.34	73.38	98.63	0.3158	4.96E-01	-0.0188	0.0275	0.000135
Treg	CD45RA- CD25hi CD4+ not Treg AC	3406	6.98	99.75	88.58	564.80	0.3159	5.07E-01	-0.0186	0.0280	0.000130
Treg	CD4+ Treg %CD4+	3439	1.64	5.71	5.54	13.31	0.3158	5.07E-01	-0.0184	0.0277	0.000129

Treg	CD4+ not Treg %CD4+	3440	86.69	94.29	94.46	100.00	0.3159	5.11E-01	0.0184	0.0280	0.000126
MT	CM (CD4- CD8-) %CD4-CD8-	3427	0.00	1.49	0.98	20.37	0.3158	5.16E-01	0.0165	0.0253	0.000123
Treg	CD45RA- CD28+ CD8br AC	3408	8.79	148.50	133.70	1169.00	0.3160	5.23E-01	-0.0179	0.0279	0.000120
TBNK	T lymphocyte %leukocyte	3669	5.67	25.96	25.63	55.53	0.3147	5.26E-01	-0.0167	0.0263	0.000110
Treg	CD45RA+ CD25hi CD4+ not Treg AC	3407	0.02	13.22	9.88	160.70	0.3159	5.37E-01	-0.0174	0.0281	0.000112
Treg	CD39+ activated CD4+ Treg AC	3407	0.22	9.27	7.97	54.09	0.3159	5.40E-01	-0.0174	0.0285	0.000110
MT	CD45RA- CD4+ AC	3395	96.61	614.30	586.10	2467.00	0.3158	5.59E-01	-0.0167	0.0286	0.000101
cDC	myeloid cDC %cDC	3410	7.99	54.56	53.88	96.19	0.3155	5.62E-01	-0.0163	0.0281	0.000099
Treg	CD39+ resting CD4+ Treg %CD4+ Treg	3439	0.00	0.99	0.64	12.47	0.3158	5.81E-01	0.0151	0.0273	0.000089
Treg	CD45RA+ CD28+ CD8br %CD8br	3440	0.00	30.88	29.00	88.25	0.3159	5.89E-01	0.0112	0.0207	0.000085
Treg	CD39+ CD8br %T lymphocyte	3440	0.00	0.60	0.47	14.42	0.3159	5.95E-01	0.0147	0.0278	0.000082
MT	EM CD4+ %CD4+	3427	6.48	30.57	28.65	91.01	0.3158	6.06E-01	0.0135	0.0262	0.000078
Treg	CD39+ secreting CD4+ Treg %CD4+ Treg	3439	0.78	17.68	17.79	58.74	0.3158	6.38E-01	-0.0130	0.0275	0.000065
TBNK	HLA DR+ T lymphocyte %T lymphocyte	3596	0.66	5.00	4.13	44.29	0.3130	6.38E-01	-0.0117	0.0250	0.000062
Treg	activated CD4+ Treg %CD4+	3439	0.29	1.50	1.38	5.42	0.3158	6.43E-01	-0.0125	0.0270	0.000063
Treg	CD45RA+ CD28+ CD8dim AC	3408	0.00	4.65	3.59	77.17	0.3160	6.43E-01	0.0112	0.0243	0.000063
Treg	CD39+ CD4+ Treg %CD4+ Treg	3439	1.56	33.33	36.54	83.70	0.3158	6.64E-01	-0.0117	0.0270	0.000055
Treg	CD39+ activated CD4+ Treg %CD4+ Treg	3439	0.66	15.62	16.18	51.45	0.3158	6.80E-01	0.0113	0.0273	0.000050
MT	CD45RA- CD8br %CD8br	3427	11.63	58.84	59.39	97.04	0.3158	6.89E-01	0.0094	0.0235	0.000047
MT	CD45RA+ CD8br %CD8br	3427	2.96	41.16	40.61	88.37	0.3158	6.89E-01	-0.0094	0.0235	0.000047
Treg	CD45RA+ CD28+ CD8br %T lymphocyte	3440	0.00	8.04	7.19	39.63	0.3159	6.92E-01	0.0076	0.0192	0.000046
MT	naive (CD4- CD8-) AC	3395	0.00	2.30	1.38	32.19	0.3158	6.92E-01	0.0087	0.0221	0.000046

MT	CM (CD4- CD8-) %T lymphocyte	3427	0.00	0.05	0.04	0.85	0.3158	6.94E-01	0.0094	0.0238	0.000045
MT	naive CD8br %CD8br	3427	0.00	26.75	24.53	86.96	0.3158	7.03E-01	0.0077	0.0203	0.000042
Treg	resting CD4+ Treg %CD4+ Treg	3439	0.43	20.90	19.52	78.56	0.3158	7.04E-01	-0.0092	0.0242	0.000042
Treg	CD39+ CD4+ %T lymphocyte	3440	0.00	3.33	3.27	16.36	0.3159	7.16E-01	0.0097	0.0267	0.000038
cDC	CD62L- myeloid cDC %cDC	3410	0.48	12.50	7.48	93.58	0.3155	7.32E-01	-0.0096	0.0282	0.000035
cDC	plasmacytoid cDC %cDC	3410	0.44	37.01	37.36	80.99	0.3155	7.33E-01	-0.0095	0.0277	0.000034
Treg	CD45RA- CD28+ CD8dim %T lymphocyte	3440	0.05	1.53	1.25	16.54	0.3159	7.43E-01	-0.0085	0.0259	0.000031
Treg	CD25hi CD4+ %T lymphocyte	3439	1.29	10.25	9.57	42.01	0.3158	7.54E-01	0.0087	0.0279	0.000029
Treg	CD45RA+ CD25hi CD4+ not Treg %T lymphocyte	3439	0.00	0.77	0.59	8.73	0.3158	7.59E-01	0.0087	0.0284	0.000027
Treg	CD39+ secreting CD4+ Treg %secreting CD4+ Treg	3439	1.07	31.79	33.94	79.09	0.3158	7.64E-01	-0.0083	0.0277	0.000026
Treg	CD4+ Treg %T lymphocyte	3439	0.45	3.54	3.41	8.71	0.3158	7.93E-01	-0.0074	0.0284	0.000020
Treg	CD45RA+ CD25hi CD4+ not Treg %CD4+	3439	0.01	1.20	0.96	12.99	0.3158	8.07E-01	0.0069	0.0283	0.000017
TBNK	CD4- CD8dim %T lymphocyte	3668	0.06	3.02	2.44	28.01	0.3146	8.18E-01	-0.0062	0.0269	0.000014
Treg	CD28- CD8dim %CD8dim	3440	2.17	46.44	44.87	98.29	0.3159	8.31E-01	-0.0058	0.0270	0.000013
MT	naive CD8br %T lymphocyte	3427	0.00	7.00	6.10	37.55	0.3158	8.34E-01	0.0039	0.0185	0.000013
cDC	CD62L- plasmacytoid cDC AC	3284	0.00	0.23	0.15	5.22	0.3128	8.43E-01	0.0054	0.0272	0.000012
Treg	CD25hi CD4+ %CD4+	3439	4.45	16.47	15.60	51.86	0.3158	8.50E-01	0.0053	0.0278	0.000010
MT	EM CD8br %T lymphocyte	3427	0.38	15.17	13.58	70.09	0.3158	8.50E-01	-0.0051	0.0268	0.000010
Treg	CD45RA- CD28+ CD8dim %CD8dim	3440	1.59	44.61	44.70	95.49	0.3159	8.66E-01	-0.0045	0.0269	0.000008
TBNK	HLA DR++ CD14+ monocyte %leukocyte	3597	0.01	0.62	0.56	3.29	0.3132	8.79E-01	0.0041	0.0267	0.000006
Treg	CD39+ CD8br AC	3408	0.00	10.02	7.66	156.70	0.3160	8.80E-01	-0.0043	0.0283	0.000007
MT	CM CD4+ AC	3395	0.00	300.30	281.40	1071.00	0.3158	8.83E-01	-0.0041	0.0277	0.000006

Treg	CD28- CD8dim %T lymphocyte	3440	0.05	2.02	1.23	28.56	0.3159	9.02E-01	-0.0034	0.0278	0.000004
Treg	CD39+ CD4+ Treg %CD4+	3439	0.07	1.90	1.85	9.22	0.3158	9.06E-01	-0.0032	0.0274	0.000004
Treg	activated CD4+ Treg %CD4+ Treg	3439	5.73	25.89	25.36	60.48	0.3158	9.13E-01	0.0029	0.0263	0.000003
MT	EM CD8br %CD8br	3427	1.47	54.03	54.14	95.72	0.3158	9.21E-01	0.0023	0.0234	0.000003
Treg	CD39+ resting CD4+ Treg AC	3407	0.00	0.61	0.35	7.69	0.3159	9.29E-01	0.0024	0.0275	0.000002
TBNK	CD4- CD8- %T lymphocyte	3668	0.24	5.02	4.07	38.61	0.3146	9.47E-01	0.0017	0.0264	1.20E-06
MT	CM CD8br AC	3395	0.00	21.59	17.53	305.90	0.3158	9.48E-01	-0.0017	0.0259	1.26E-06
MT	CD45RA- CD8br %T lymphocyte	3427	2.32	16.41	14.94	70.54	0.3158	9.51E-01	-0.0017	0.0270	1.12E-06
MT	CD45RA- (CD4- CD8-) %T lymphocyte	3427	0.09	3.64	2.73	31.40	0.3158	9.58E-01	-0.0014	0.0269	7.98E-07
Treg	secreting CD4+ Treg %CD4+ Treg	3439	10.46	55.54	55.68	87.37	0.3158	9.65E-01	0.0012	0.0264	5.77E-07
Treg	CD39+ CD4+ %CD4+	3440	0.01	5.39	5.42	22.67	0.3159	9.69E-01	-0.0010	0.0267	4.39E-07
MT	EM (CD4- CD8-) %T lymphocyte	3427	0.08	3.59	2.68	31.28	0.3158	9.87E-01	-0.0004	0.0265	7.61E-08

Table S8. B cell panel immunophenotypes: statistics and association results.

Listed, from left to right, are: (1st column) the trait name (AC stands for Absolute Count expressed as 10^6 cells/litre); (2nd) the number of assessed samples; (3rd, 4th, 5th, 6th) the minimum, mean, median and maximum trait values, respectively; (7th) BAFF-var frequency; (8th) the statistical significance (*P* value); (9th) BAFF-var effect size expressed in standard deviation units; (10th) the standard error (SE); (11th) the heritability (H2) explained. The *P* value threshold is 2.38×10^{-7} (rows in bold), after Bonferroni correction of the nominal P value for 10K assessed variants and 21 independent traits (1 B cell count + 6 B cell subtype absolute counts + 14 median fluorescence intensities).

Traits	*N samples*	*Min*	*Mean*	*Median*	*Max*	*BAFF-var freq*	*P value*	*Effect*	*SE*	*H2 explained*
Absolute counts and percentages										
B cell %lymphocyte	**1901**	**0.74**	**13.03**	**12.39**	**74.59**	**0.3422**	**1.45E-19**	**0.3150**	**0.0344**	**0.042270**
CD24+ CD27+ %lymphocyte	**1899**	**0.18**	**3.89**	**3.27**	**40.60**	**0.3423**	**1.00E-14**	**0.2678**	**0.0343**	**0.031140**
naive %lymphocyte	**1899**	**0.20**	**8.12**	**7.49**	**54.58**	**0.3423**	**4.44E-12**	**0.2408**	**0.0346**	**0.024990**
sw mem %lymphocyte	**1899**	**0.10**	**2.18**	**1.91**	**36.82**	**0.3423**	**2.48E-10**	**0.2207**	**0.0347**	**0.020930**
B cell AC	**1902**	**5.39**	**343.30**	**300.60**	**15600.00**	**0.3423**	**2.94E-10**	**0.2170**	**0.0342**	**0.020720**
CD24+ CD27+ AC	**1899**	**2.69**	**102.20**	**78.72**	**2063.00**	**0.3423**	**1.45E-09**	**0.2126**	**0.0350**	**0.019150**
unsw mem %lymphocyte	**1899**	**0.07**	**1.96**	**1.60**	**38.41**	**0.3423**	**5.71E-09**	**0.2011**	**0.0344**	**0.017760**
naive AC	1899	8.08	207.10	182.30	2751.00	0.3423	2.61E-07	0.1792	0.0347	0.013910
sw mem AC	1899	2.23	56.91	45.91	2096.00	0.3423	2.77E-06	0.1659	0.0353	0.011530
unsw mem AC	1899	1.46	51.09	39.18	1098.00	0.3423	4.40E-06	0.1591	0.0346	0.011070

IgD- CD27- AC	1899	1.08	19.97	14.62	2340.00	0.3423	5.68E-02	0.0672	0.0353	0.001914
transitional AC	1899	0.02	10.48	8.31	64.69	0.3423	8.48E-02	0.0590	0.0342	0.001568
IgD- CD27- %B cell	1899	0.22	5.74	4.96	73.64	0.3423	1.55E-03	-0.1105	0.0349	0.005275
IgD- CD27- %lymphocyte	1899	0.04	0.74	0.59	42.81	0.3423	2.69E-03	0.1052	0.0350	0.004745
transitional %B cell	1899	0.05	3.18	2.88	21.82	0.3423	8.57E-03	-0.0920	0.0350	0.003643
CD24+ CD27+ %B cell	1899	2.51	30.29	27.62	89.00	0.3423	9.43E-03	0.0921	0.0354	0.003553
transitional %lymphocyte	1899	0.00	0.42	0.34	2.71	0.3423	2.67E-02	0.0762	0.0343	0.002591
unsw mem %B cell	1899	1.18	15.03	13.41	69.59	0.3423	4.84E-01	0.0246	0.0352	0.000258
naive %B cell	1899	2.03	61.88	63.15	95.92	0.3423	6.91E-01	-0.0140	0.0352	0.000084
sw mem %B cell	1899	0.84	17.35	15.95	93.22	0.3423	9.89E-01	0.0005	0.0356	9.29E-08
Normalized Median Fluorescent Intensities										
CD25 on naive	1673	141	336	321.7	930.4	0.34116	1.95E-03	0.1141	0.0368	0.00573
CD25 on B cell	1675	185.8	443.2	427.7	1215	0.34105	4.33E-03	0.1071	0.0375	0.00485
CD25 on unsw mem	1673	273.8	745	736.2	1769	0.34116	7.10E-03	0.0987	0.0366	0.00433
CD25 on CD24+ CD27+	1673	328.6	739.8	731.4	1624	0.34116	9.31E-02	0.0610	0.0363	0.00169
BAFF-R on unsw mem	1673	11.38	1694	1738	3062	0.34116	1.66E-01	-0.0521	0.0376	0.00115
CD25 on sw mem	1673	285.3	637.8	633.1	1226	0.34116	2.06E-01	0.0464	0.0367	0.00096
BAFF-R on CD24+ CD27+	1673	14.08	1698	1746	3155	0.34116	3.04E-01	-0.0391	0.0380	0.00063
BAFF-R on B cell	1675	25.92	1756	1797	3185	0.34105	4.13E-01	-0.0307	0.0375	0.00040
BAFF-R on naive	1673	26.31	1793	1835	3149	0.34116	4.24E-01	-0.0300	0.0375	0.00038
BAFF-R on IgD- CD27-	1673	44.75	1572	1595	2937	0.34116	4.27E-01	-0.0298	0.0375	0.00038
CD25 on transitional	1673	143.9	354.9	348.5	1225	0.34116	5.44E-01	0.0221	0.0364	0.00022
BAFF-R on sw mem	1673	16.1	1699	1736	3170	0.34116	5.87E-01	-0.0207	0.0381	0.00018
BAFF-R on transitional	1673	33.38	1764	1787	3465	0.34116	8.81E-01	-0.0056	0.0376	0.00001
CD25 on IgD- CD27-	1673	94.77	565.4	555.7	1175	0.34116	9.78E-01	-0.0010	0.0367	4.40E-07

Table S9. Serum proteins and hemocytometer-assessed parameters: statistics and results.

Listed, from left to right, are: (1st column) the trait name (counts are expressed as 10^6 cells/litre); (2nd) the number of assessed samples; (3rd, 4th, 5th, 6th) the minimum, mean, median and maximum trait values, respectively; (7th) BAFF-var frequency; (8th) the statistical significance (*P* value); (9th) BAFF-var effect size expressed in standard deviation units; (10th) the standard error (SE); (11th) the heritability (H2) explained. The *P* value thresholds are 5.0e-06, 1.0e-06 and 8.33e-7 (rows in bold) for soluble BAFF, immunoglobulins and hemocytometer parameters. Bonferroni correction has been applied to the nominal *P* value considering 10K variants and the number of traits measured (6 immunoglobulins and 5 absolute counted hemocytometer-assessed parameters).

Traits	*N samples*	*Min*	*Mean*	*Median*	*Max*	*BAFF-var freq*	*P value*	*Effect*	*SE*	*H2 explained*
Cytokine										
soluble BAFF (*pg/ml*)	**2733**	**353.9**	**825**	**795**	**3279**	**0.3148**	**8.47E-150**	**0.7752**	**0.0279**	**0.22030**
Immunoglobulins										
IgG-1 (mg/dL)	**2898**	**1.43**	**520.54**	**470.01**	**13029.00**	**0.3123**	**2.24E-14**	**0.2299**	**0.0300**	**0.01996**
IgG (mg/dL)	**2886**	**9.03**	**989.56**	**934.58**	**9977.16**	**0.3122**	**1.68E-12**	**0.2160**	**0.0305**	**0.01715**
IgA (mg/dL)	**2885**	**0.73**	**132.08**	**118.26**	**1869.90**	**0.3118**	**7.64E-09**	**0.1679**	**0.0290**	**0.01152**
IgM (mg/dL)	**2898**	**0.12**	**151.41**	**126.30**	**4038.70**	**0.3123**	**4.70E-08**	**0.1567**	**0.0286**	**0.01026**
IgG-3 (mg/dL)	2897	0.17	53.27	43.86	4324.10	0.3121	1.02E-05	0.1362	0.0308	0.00671
IgG-4 (mg/dL)	2889	0.06	54.16	40.87	588.46	0.3122	1.69E-04	0.1132	0.0301	0.00489

IgG-2 (mg/dL)	2896	7.35	379.03	342.78	13182.00	0.3125	1.11E-03	0.0980	0.0300	0.00367
Hemocytometer traits										
Monocyte count	**5894**	**4.80**	**389.31**	**365.60**	**3504.60**	**0.3161**	**9.07E-13**	**-0.1595**	**0.0223**	**0.00863**
Monocyte %WBC	**5894**	**0.10**	**5.93**	**5.80**	**53.10**	**0.3161**	**5.48E-07**	**-0.1110**	**0.0221**	**0.00425**
WBC	5937	800.00	6674.35	6500.00	21800.00	0.3159	2.66E-05	-0.0929	0.0221	0.00297
Neutrophil count	5899	303.20	3817.24	3613.40	14862.50	0.3163	2.96E-03	-0.0648	0.0218	0.00150
Eosinophil count	5827	2.10	173.63	141.00	3433.80	0.3172	6.44E-03	-0.0609	0.0223	0.00127
Lymphocyte count	5899	150.40	2273.08	2188.80	18268.40	0.3163	9.57E-03	-0.0576	0.0222	0.00114
Basophil count	5899	0.00	23.48	20.00	862.50	0.3163	1.06E-01	-0.0351	0.0217	4.44E-04
Lymphocyte %WBC	5899	3.20	34.55	34.50	83.80	0.3163	1.51E-01	0.0313	0.0218	3.50E-04
Eosinophil % WBC	5827	0.10	2.61	2.20	29.10	0.3172	2.07E-01	-0.0277	0.0219	2.74E-04
Neutrophil %WBC	5899	5.80	56.59	56.50	90.80	0.3163	8.12E-01	0.0052	0.0218	9.56E-06
Basophil %WBC	5899	0.00	0.36	0.30	12.50	0.3163	8.76E-01	0.0033	0.0213	4.15E-06

Table S10. Monocyte panel immunophenotypes: statistics and association results.

Listed, from left to right, are: (1st column) the trait name (AC stands for Absolute Count expressed as 10^6 cells/litre); (2nd) the number of assessed samples; (3rd, 4th, 5th, 6th) the minimum, mean, median and maximum trait values, respectively; (7th) BAFF-var frequency; (8th) the statistical significance (*P* value); (9th) BAFF-var effect size expressed in standard deviation units; (10th) the standard error (SE); (11th) the heritability (H2) explained. The *P* value threshold is 1.78e-07, after Bonferroni correction of the nominal *P* value for 10K assessed variants and 28 traits (1 Monocyte count + 3 Monocyte subtype absolute counts + 24 median fluorescence intensities).

Traits	***N samples***	***Min***	***Mean***	***Median***	***Max***	***BAFF-var freq***	***P value***	***Effect***	***SE***	***H2 explained***
Absolute counts and percentages										
CD14+ CD16- AC	1856	125.00	408.20	390.30	1062.00	0.34168	5.92E-04	-0.1166	0.0339	0.00636
monocyte AC	1856	153.70	484.90	463.40	1206.00	0.34168	2.56E-03	-0.1008	0.0334	0.00490
CD14+ CD16- %monocyte	1856	58.13	84.27	84.61	99.98	0.34168	5.68E-02	-0.0662	0.0347	0.00196
CD14- CD16+ %monocyte	1856	0.01	11.42	10.92	33.61	0.34168	6.25E-02	0.0641	0.0344	0.00187
CD14+ CD16+ %monocyte	1856	0.01	4.31	4.04	25.52	0.34168	2.80E-01	0.0389	0.0359	0.00063
CD14+ CD16+ AC	1856	0.02	21.37	18.42	127.80	0.34168	6.95E-01	-0.0137	0.0350	0.00008
CD14- CD16+ AC	1856	0.07	55.30	50.45	271.20	0.34168	8.79E-01	-0.0051	0.0336	0.00001
Normalized Median Fluorescent Intensities										
HLA DR on CD14- CD16+	1632	2840	7471	7236	17700	0.34027	4.02E-02	-0.0769	0.0375	0.00258
CX3CR1 on CD14- CD16+	1594	1582	3214	3163	5957	0.34323	7.85E-02	-0.0656	0.0373	0.00194
PD-L1 on CD14- CD16+	1632	153.4	406	398.2	1521	0.34027	2.20E-01	-0.0472	0.0385	0.00092

CCR2 on CD14+ CD16+	1632	136	374.3	363.4	970.7	0.34027	2.71E-01	-0.0414	0.0376	0.00074
PD-L1 on CD14+ CD16+	1632	273.8	516	506.2	1252	0.34027	3.17E-01	-0.0375	0.0375	0.00061
HLA DR on CD14+ CD16+	1632	1277	18510	18560	34100	0.34027	3.72E-01	-0.0340	0.0381	0.00049
CD64 on CD14+ CD16-	1625	189.2	1503	1467	3645	0.34021	4.13E-01	0.0309	0.0376	0.00041
CD64 on Monocyte	1625	206.8	1384	1349	3401	0.34021	4.42E-01	0.0291	0.0378	0.00037
CD40 on CD14- CD16+	1632	25.63	279.4	261.3	1793	0.34027	5.87E-01	0.0207	0.0380	0.00018
CX3CR1 on CD14+ CD16-	1594	503.8	1249	1226	2876	0.34323	6.65E-01	-0.0164	0.0378	0.00012
PD-L1 on Monocyte	1632	219	373.3	368.9	657.4	0.34027	6.86E-01	-0.0152	0.0377	0.00010
HLA DR on Monocyte	1632	1186	3608	3533	8032	0.34027	7.01E-01	-0.0147	0.0382	9.06E-05
CCR2 on CD14+ CD16-	1632	243.8	400.7	391.5	698.5	0.34027	7.12E-01	-0.0136	0.0368	8.37E-05
CD40 on CD14+ CD16+	1632	110.7	485.9	472.6	1419	0.34027	7.48E-01	0.0123	0.0383	6.31E-05
CCR2 on Monocyte	1632	232.4	368.4	360.8	656.4	0.34027	7.59E-01	-0.0115	0.0374	5.80E-05
HLA DR on CD14+ CD16-	1632	1096	3046	2979	6482	0.34027	7.62E-01	-0.0115	0.0382	5.61E-05
CD64 on CD14+ CD16+	1625	342.3	1271	1237	3107	0.34021	7.75E-01	-0.0107	0.0376	5.02E-05
PD-L1 on CD14+ CD16-	1632	224.9	391.2	385.7	670.1	0.34027	8.45E-01	-0.0075	0.0381	2.36E-05
CD40 on Monocyte	1632	107.5	255.8	247.7	640.7	0.34027	8.77E-01	0.0059	0.0383	1.47E-05
CX3CR1 on Monocyte	1594	538.8	1364	1334	3075	0.34323	8.86E-01	-0.0054	0.0379	1.29E-05
CD64 on CD14- CD16+	1625	215	432	417.7	1054	0.34021	8.91E-01	-0.0052	0.0375	1.17E-05
CCR2 on CD14- CD16+	1632	127.1	227.1	222.5	688	0.34027	9.05E-01	0.0046	0.0382	8.84E-06
CX3CR1 on CD14+ CD16+	1594	968.7	2321	2278	6127	0.34323	9.29E-01	-0.0034	0.0381	4.95E-06
CD40 on CD14+ CD16-	1632	108.7	259.4	252.5	602.4	0.34027	9.82E-01	-0.0008	0.0379	2.98E-07

Table S11. Reciprocal conditional analyses for BAFF-var and rs12874404 with the 27 genome-wide significantly associated immune traits.

Listed, from left to right, are: (1st column) the trait name (AC stands for Absolute Count expressed as 10^6 cells/litre); (2nd) the number of assessed samples; (3rd) BAFF-var frequency; BAFF-var association results, in particular (4th) the statistical significance (*P* value), and (5th) the effect size and its standard error (SE); BAFF-var association results after conditioning for rs12874404, with (6th) the *P* value and (7th) the effect size with standard error (SE) in brackets; (8th) rs12874404 frequency; (9th, 10th) rs12874404 association results, in particular *P* value, effect size with SE in brackets, respectively; (11th, 12th) rs12874404 association results after conditioning for BAFF-var, with *P* value, effect size with SE in brackets, respectively. The effect sizes are expressed in standard deviation units. rs12874404-A>G association results are referred to the G allele, in phase with BAFF-var.

Traits	***N samples***	***BAFF-var association***			***BAFF-var association after conditioning for rs12874404***		***rs12874404 association****			***rs12874404 association* after conditioning for BAFF-var***	
		Freq	***P value***	***Effect (SE)***	***P value***	***Effect (SE)***	***Freq***	***P value***	***Effect (SE)***	***P value***	***Effect (SE)***
TBNK panel											
B cell %lymphocyte	3669	0.315	9.36E-23	0.263 (0.03)	1.24E-05	0.243 (0.06)	0.317	2.20E-18	0.233 (0.03)	7.31E-01	0.019 (0.06)
T/B cell	3653	0.315	3.42E-21	-0.255 (0.03)	1.55E-05	-0.240 (0.06)	0.317	4.31E-17	-0.226 (0.03)	7.63E-01	-0.017 (0.06)
B cell AC	3653	0.315	4.23E-12	0.183 (0.03)	1.62E-03	0.174 (0.06)	0.317	6.32E-10	0.163 (0.03)	8.64E-01	0.009 (0.05)

NK %CD3- lymphocyte	3669	0.315	5.06E-11	-0.171 (0.03)	1.02E-03	-0.174 (0.05)	0.317	7.66E-09	-0.150 (0.03)	9.62E-01	-0.003 (0.05)
B cell %CD3- lymphocyte	3669	0.315	6.20E-11	0.170 (0.03)	8.29E-04	0.177 (0.05)	0.317	1.18E-08	0.148 (0.03)	9.72E-01	-0.002 (0.05)
B cell panel											
B cell %lymphocyte	1901	0.342	1.45E-19	0.315 (0.03)	1.28E-03	0.233 (0.07)	0.344	5.90E-18	0.304 (0.03)	1.70E-01	0.101 (0.07)
IgD+ %lymphocyte	1899	0.342	3.93E-17	0.291 (0.03)	7.43E-03	0.193 (0.07)	0.344	3.55E-16	0.285 (0.03)	1.18E-01	0.114 (0.07)
IgD+ CD38- %lymphocyte	1899	0.342	1.77E-15	0.277 (0.03)	1.76E-04	0.273 (0.07)	0.344	1.69E-12	0.249 (0.03)	9.23E-01	0.007 (0.07)
CD24+ CD27+ %lymphocyte	1899	0.342	1.00E-14	0.268 (0.03)	2.22E-03	0.226 (0.07)	0.344	1.30E-12	0.248 (0.03)	5.81E-01	0.041 (0.07)
IgD+ CD24+ %lymphocyte	1899	0.342	1.75E-13	0.252 (0.03)	3.97E-03	0.210 (0.07)	0.344	1.21E-11	0.234 (0.03)	5.51E-01	0.044 (0.07)
memory %lymphocyte	1899	0.342	2.01E-12	0.244 (0.03)	1.75E-03	0.231 (0.07)	0.344	4.09E-10	0.219 (0.03)	9.09E-01	0.009 (0.07)
IgD+CD38-CD27- %lymphocyte	1885	0.341	2.01E-12	0.246 (0.03)	4.42E-05	0.300 (0.07)	0.343	5.31E-09	0.207 (0.04)	4.50E-01	-0.056 (0.07)
naive %lymphocyte	1899	0.342	4.44E-12	0.241 (0.03)	1.93E-02	0.172 (0.07)	0.344	4.04E-11	0.232 (0.03)	2.85E-01	0.080 (0.07)
IgD-CD25- %lymphocyte	1900	0.342	1.53E-11	0.239 (0.03)	5.79E-03	0.132 (0.05)	0.344	1.36E-05	0.098 (0.04)	7.08E-01	-0.018 (0.05)
IgD+ CD38- AC	1899	0.342	5.27E-11	0.229 (0.03)	1.81E-01	0.098 (0.07)	0.344	3.35E-10	0.218 (0.03)	8.45E-02	0.128 (0.07)
IgD+CD38-CD27+ %lymphocyte	1885	0.341	5.90E-11	0.230 (0.03)	6.96E-03	0.198 (0.07)	0.343	1.84E-09	0.214 (0.04)	5.90E-01	0.040 (0.07)
sw mem %lymphocyte	1899	0.342	2.48E-10	0.221 (0.03)	1.16E-02	0.189 (0.07)	0.344	7.77E-09	0.204 (0.04)	6.45E-01	0.035 (0.08)
B cell AC	1902	0.342	2.94E-10	0.217 (0.03)	1.45E-01	0.107 (0.07)	0.344	2.45E-10	0.220 (0.03)	1.01E-01	0.122 (0.07)
naive-mature %lymphocyte	1899	0.342	3.79E-10	0.220 (0.03)	2.04E-01	0.094 (0.07)	0.344	1.23E-10	0.228 (0.04)	5.18E-02	0.145 (0.07)

IgD+ AC	1899	0.342	5.68E-10	0.213 (0.03)	5.70E-03	0.202 (0.07)	0.344	2.08E-09	0.211 (0.04)	6.42E-01	0.034 (0.07)
CD24+ CD27+ AC	1899	0.342	1.45E-09	0.213 (0.03)	3.67E-02	0.154 (0.07)	0.344	7.33E-09	0.205 (0.04)	3.47E-01	0.070 (0.07)
IgD+ CD24- %lymphocyte	1899	0.342	4.81E-09	0.203 (0.03)	2.20E-01	0.090 (0.07)	0.344	2.21E-09	0.209 (0.03)	8.01E-02	0.130 (0.07)
unsw mem %lymphocyte	1899	0.342	5.71E-09	0.201 (0.03)	1.65E-01	0.102 (0.07)	0.344	6.16E-09	0.203 (0.03)	1.55E-01	0.106 (0.07)
Cytokine											
soluble BAFF (*pg/ml*)	2733	0.315	8.47E-150	0.775 (0.03)	1.05E-05	0.138 (0.03)	0.314	5.64E-124	0.711 (0.03)	8.74E-01	0.005 (0.03)
Immunoglobulins											
IgG-1 (mg/dL)	2898	0.312	2.24E-14	0.230 (0.03)	4.03E-03	0.176 (0.06)	0.314	8.55E-13	0.214 (0.03)	2.98E-01	0.064 (0.06)
IgG (mg/dL)	2886	0.312	1.68E-12	0.216 (0.03)	2.60E-03	0.188 (0.06)	0.314	1.67E-10	0.195 (0.03)	6.38E-01	0.029 (0.06)
Hemocytometer traits											
Monocyte count	5894	0.316	9.07E-13	-0.160 (0.02)	1.42E-01	-0.033 (0.02)	0.318	4.92E-10	-0.139 (0.02)	7.12E-01	0.008 (0.02)

* rs12874404 (A>G) association results are referred to G allele, in phase with BAFF-var

Table S12. sBAFF levels in case-control analyses.

Listed, from left to right, are: (1st, 2nd column) the datasets assessed (MS or SLE cases and relative controls); (3rd) the number of assessed samples; (4th) BAFF-var frequency; (5th, 6th, 7th, 8th) the minimum, mean, median and maximum of sBAFF levels, expressed in pg/ml, respectively; (9th) case-control comparison statistical significance (CS-CT P value); (10th) BAFF-var association statistical significance (*P* value); (11th) BAFF-var effect size expressed in standard deviation units (Beta); (12th) the standard error (SE).

				BAFF-var	sBAFF levels					Association with BAFF-var		
	Dataset		N	Freq (%)	Min	Mean	Median	Max	CS-CT P value	P value	Beta	SE
MS		CT	88	32.3	550.7	834.0	813.5	1470.5	1.07×10^{-5}	2.43×10^{-8}	0.70	0.11
		MS	41	35.3	562.2	1040.8	10011.2	1882.5		1.93×10^{-4}	0.53	0.14
SLE		CT	79	19.6	527.2	829.6	815.2	1410.1	1.03×10^{-9}	7.79×10^{-8}	0.96	0.16
		SLE	76	36.1	328.6	1530.9	1101.3	9327.1		0.001	0.49	0.15

Table S13. F_{ST} results among Sardinians and 1000 Genomes populations.

Listed, from left to right, are: (1st column) the population code; (2nd) the population description; (3rd) the number of chromosomes analyzed; (4th) BAFF-var frequency; (5th) the F_{ST} value; (6th) the genomic percentile.

Population code	***Population description***	***# Chromosomes***	***BAFF-var freq***	***F_{ST}***	***Genomic percentile***
EUR*	*Europeans*	1006	0.027	0.207	99.97
EUR* excluding TSI	*Europeans excluding Tuscans*	792	0.018	0.196	99.97
CEU	*Utah Residents (CEPH) with Northern and Western European Ancestry*	198	0.035	0.162	99.90
TSI	*Tuscans in Italy*	214	0.061	0.163	99.97
IBS	*Iberian Population in Spain*	214	0.028	0.162	99.93
ASW	*Americans of African Ancestry in SW USA*	122	0.016	0.156	85.98
AMR**	*Admixed Americans*	694	0.045	0.190	99.13
CLM	*Colombians from Medellin, Colombia*	188	0.037	0.161	99.53
MXL	*Mexican Ancestry from Los Angeles USA*	128	0.063	0.156	94.90

PEL	*Peruvians from Lima, Peru*	170	0.012	0.160	87.57
PUR	*Puerto Ricans from Puerto Rico*	208	0.067	0.161	99.83
PJL	*Punjabi from Lahore, Pakistan*	192	0.026	0.162	97.35

Table S14. SLE clinical feature description.

The table lists the clinical features assessed and their description. Specification of criteria used for the classification are reported as: * definition based on 2012 SLICC classification criteria for SLE; ** definition based on SLEDAI-2K disease activity system; *** classification according to the 2006 Sidney Criteria.

Clinical Feature	Description
Acute cutaneous lupus*	Lupus malar rash (do not count if malar discoid), bullous lupus toxic epidermal necrolysis variant of SLE, maculopapular lupus rash, photosensitive lupus rash; in ABSENCE OF dermatomyositis OR subacute cutaneous lupus (no indurated psoriasiform and/or annular polycyclic lesions that resolve without scarring, although occasionally with post inflammatory depigmentation or telangiectasias)
Chronic cutaneous lupus*	Classic discoid rash localized (above the neck), generalized (above and below the neck), hypertrophic (verrucous) lupus, lupus panniculitis (profundus), mucosal lupus, lupus erythematosus tumidus, Chilblains lupus, discoid lupus/lichen planus overlap
Oral ulcers*	Palate, buccal, Tongue OR nasal ulcers, in the ABSENCE OF other causes, such as vasculitis, Bechet's disease, infection (herpesvirus), inflammatory bowel disease, reactive arthritis, and acidic food.
Non scarring alopecia*	Diffuse thinning or hair fragility with visible broken hairs, in the ABSENCE OF other causes such as alopecia areata, drugs, iron deficiency, and androgenic alopecia

Synovitis*	Synovitis involving 2 or more joints, characterized by swelling or effusion OR tenderness in 2 or more joints and at least 30 minutes of morning stiffness
Serositis*	Typical pleurisy for more than 1 day OR pleural effusions OR pleural rub. Typical pericardial pain (pain with recumbency improved by sitting forward) for more than 1 day OR pericardial effusion OR pericardial rub OR pericarditis by electrocardiography in the ABSENCE OF other causes, such as infection, uremia, and Dressler's pericarditis.
Renal disorders*	Urine protein–to-creatinine ratio (or 24-hour urine protein) representing 500 mg protein/24 hours OR red blood cell casts
Neurologic disorders*	Seizures, Psychosis, Mononeuritis multiplex in the ABSENCE OF other known causes such as primary vasculitis. Myelitis, Peripheral or cranial neuropathy in the ABSENCE OF other known causes such as primary vasculitis, infection, and diabetes mellitus. Acute confusional state in the ABSENCE OF other causes, including toxic/metabolic, uremia, drugs.
Leukopenia*	Leukopenia (<4,000/mm3 at least once) in ABSENCE OF other known causes such as Felty's syndrome, drugs, and portal hypertension OR Lymphopenia (<1,000/mm3 at least once) in the ABSENCE OF other known causes such as corticosteroids, drugs, and infection
Thrombocytopenia*	Thrombocytopenia (<100,000/mm3) at least once in the ABSENCE OF other known causes such as drugs, portal hypertension, and thrombotic thrombocytopenic purpura
Hemolytic anemia*	ANA level above laboratory reference range
Vasculitis**	Ulceration, gangrene, tender finger nodules, periungual, infarction, splinter hemorrhages, or biopsy or angiogram proof of vasculitis
APL syndrome***	according to Sidney 2006 criteria

Table S15. Odds ratio (OR) and population attributable risk (PAR) results from the multiplicative model.

Listed, from left to right, are: (1st column) the assessed population; the OR and the percentage of PAR in MS (2nd, 3rd) and SLE (4th, 5th) cohorts.

	MS		*SLE*	
Population	***OR***	***PAR (%)***	***OR***	***PAR (%)***
Sardinians	1.27	6.67	1.38	10
Italians	1.27	1.52	1.49	2.7
Iberians	-	-	1.55	2.68

Table S16. B cell count evaluation in MS and SLE patients.

Listed, from left to right, are: (1st column) the datasets assessed (SLE or MS cases); (2nd) the number of analysed samples; (3rd) BAFF-var frequency; (4th, 5th, 6th, 7th) the minimum, mean, median and maximum of B cell counts, respectively; (8th) BAFF-var association statistical significance (*P* value); (9th) BAFF-var effect size expressed in standard deviation units (Beta); (10th) the standard error (SE).

		BAFF-var	B cell count				Association with BAFF-var		
Dataset	N	Freq (%)	Min	Mean	Median	Max	P value	Beta	SE
MS	36	33.3	54.25	310.2	294.8	711	0.28	0.30	0.28
SLE	7	-	41.57	282.2	281.2	516.2	-	-	-
MS+SLE	43	32.5	41.57	305.6	288.4	711	0.097	0.45	0.26

Table S17. BAFF-var association in Sardinian SLE samples stratified for clinical manifestations.

Listed, from left to right, are: (1st column) the clinical manifestation assessed; (2nd) the total number of analyzed samples with, specified in brackets, the number of patients having (A) and not having (B) the clinical manifestation; (3rd) BAFF-var frequency in A; (4th) BAFF-var frequency in B; (5th) the BAFF-var association statistical significance (*P* value); (6th) BAFF-var effect size expressed in standard deviation units; (7th) the odds ratio (OR); (8th) the standard error (SE); (8th) the 95% confidence interval (95% CI). The *P* value threshold is 0.002, after Bonferroni correction of the nominal *P* value for 24 groups analyzed.

Clinical manifestation type	N samples (A / B)	BAFF-var freq in A	BAFF-var freq in B	*P* value	BAFF- var effect	OR	SE	95% CI
Sardinian patients								
Acute_cutaneous_lupus	297 (157/140)	0.356	0.362	0.89	-0.02	0.98	0.17	0.69 - 1.37
Chronic_cutaneous_lupus	291 (48/243)	0.362	0.359	0.96	0.01	1.01	0.24	0.64 - 1.60
Oral_ulcers	292 (92/200)	0.318	0.371	0.22	-0.24	0.79	0.19	0.54 - 1.15
Non_scarring_alopecia	186 (60/126)	0.425	0.353	0.18	0.32	1.36	0.23	0.87 - 2.12
Synovitis	298 (266/32)	0.351	0.350	0.99	0.00	1.00	0.29	0.57 - 1.76
Serositis	295 (99/196)	0.345	0.354	0.84	-0.04	0.96	0.19	0.67 - 1.39
Renal_disorders	302 (76/226)	0.320	0.354	0.44	-0.15	0.86	0.20	0.58 - 1.27
Neurologic_disorders	298 (53/245)	0.382	0.352	0.57	0.13	1.14	0.23	0.73 - 1.77
Leukopenia	270 (139/131)	0.368	0.357	0.78	0.05	1.05	0.18	0.74 - 1.50
Thrombocytopenia	275 (59/216)	0.379	0.354	0.62	0.11	1.12	0.22	0.73 - 1.71
Hemolytic_anemia	274 (23/251)	0.428	0.356	0.35	0.31	1.36	0.33	0.72 - 2.57

Vasculitis	282 (50/232)	0.390	0.351	0.47	0.17	1.18	0.23	0.76 - 1.84
APL_syndrome	298 (41/257)	0.375	0.349	0.65	0.11	1.12	0.25	0.69 - 1.82
Italian patients								
Oral_ulcers	275 (23, 252)	0.09	0.09	1.00	0.00	1.00	0.53	0.92 - 1.20
Non_scarring_alopecia	162 (22, 140)	0.14	0.09	0.36	0.42	1.52	0.46	0.62 - 2.43
Synovitis	364 (272, 92)	0.08	0.09	0.88	-0.04	0.96	0.30	0.37 - 1.54
Serositis	363 (128, 235)	0.07	0.09	0.39	-0.25	0.78	0.29	0.22 - 1.34
Renal_disorders	362 (153, 209)	0.11	0.07	0.05	0.53	1.70	0.27	1.17 - 2.22
Neurologic_disorders	221 (45, 176)	0.10	0.09	0.72	0.14	1.15	0.39	0.39 - 1.91
Leukopenia	362 (176, 186)	0.11	0.06	0.02	0.62	1.87	0.28	1.33 - 2.41
Thrombocytopenia	299 (81, 218)	0.09	0.09	0.96	0.02	1.02	0.32	0.39 - 1.64
Hemolytic_anemia	242 (32, 210)	0.03	0.09	0.16	-1.10	0.33	0.77	-1.18 - 1.85
APL_syndrome	75 (15, 60)	0.10	0.13	0.72	-0.22	0.80	0.61	-0.40 - 2.00
Vasculitis	73 (10, 63)	0.20	0.11	0.33	0.55	1.74	0.57	0.62 - 2.85

Table S18. iHS results in Sardinians and in 1000 Genomes populations.

Listed, from left to right, are: (1st column) the population code; (2nd) the integrated Haplotype Score (iHS) values; (3rd) the number of matching variants with BAFF-var; (4th) the number of matching variants with BAFF-var with an absolute iHS value better then BAFF-var iHS; (5th) the genomic percentile.

Population code	*iHS*	*# matching variants*	*# matching variants with abs(iHS)> BAFF-var iHS*	*Genomic percentile*
Sardinians	3.3790	3042	5	99.91
ASW	1.8039	27713	23718	19.50
CEU	3.2868	11896	1220	94.27
CLM	3.4215	9487	855	91.38
IBS	3.4305	11112	821	95.80
MXL	3.1348	4108	284	97.03
PEL	1.8941	10150	3888	60.05
PJL	4.0280	14597	430	85.70
PUR	4.0654	6859	62	99.19
TSI	2.9306	7985	531	96.48
EUR	3.2013	76602	11463	99.93

Table S19. Cross-population results when comparing Sardinian vs 1000 Genomes populations.

Listed, from left to right, are: (1st column) the population code; (2nd) the BAFF-var frequency, (3rd) the integrated Haplotype Homozygosity of the derived allele (BAFF-var, iHHd) in Sardinians; (4th) the iHHd in the 1000 Genomes populations; (5th) the cross population-Extended Haplotype Homozygosity (xp-EHH) value, and (6th) its genomic percentile; (7th) the allele specific – cross population - iHH (as-xp-IHH) value and its genomic percentile with respect to (8th) the ancestral and (9th) the derived allele.

Population code	*BAFF-var freq*	*iHHd Sardinians*	*iHHd 1000G population*	*xp-EHH*	*xp-EHH genomic percentile*	*as-xp-iHH*	*as-xp-iHH - ancestral allele - genomic percentile*	*as-xp-iHH - derived allele - genomic percentile*
EUR	0.026	0.008	0.007	0.109	20.3	0.047	36.0	24.1
CEU	0.035	0.010	0.008	0.177	65.9	0.077	76.5	80.2
TSI	0.060	0.010	0.008	0.190	93.9	0.082	49.0	88.5

References

39. Thomson G. Mapping disease genes: family-based association studies. Am J Hum Genet 1995;57(2):487–98.

40. Hedström AK, Hillert J, Olsson T, Alfredsson L. Alcohol as a modifiable lifestyle factor affecting multiple sclerosis risk. JAMA Neurol 2014;71(3):300–5.

41. Hedström AK, Bäärnhielm M, Olsson T, Alfredsson L. Tobacco smoking, but not Swedish snuff use, increases the risk of multiple sclerosis. Neurology 2009;73(9):696–701.

42. Alonso-Perez E, Suarez-Gestal M, Calaza M, et al. Further Evidence of Subphenotype Association with Systemic Lupus Erythematosus Susceptibility Loci: A European Cases Only Study. PLoS ONE 2012;7(9):e45356.

43. Río J, Nos C, Tintoré M, et al. Assessment of different treatment failure criteria in a cohort of relapsing-remitting multiple sclerosis patients treated with interferon beta: implications for clinical trials. Ann Neurol 2002;52(4):400–6.

44. Pistis G, Porcu E, Vrieze SI, et al. Rare variant genotype imputation with thousands of study-specific whole-genome sequences: implications for cost-effective study designs. Eur J Hum Genet 2015;23(7):975–83.

45. Dunnen den J, Antonarakis S. Nomenclature for the description of human sequence variations. Human Genetics 2001;109(1):121–4.

46. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006;38(8):904–9.

47. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 2010;26(17):2190–1.

48. Rothman KJ, Greenland S. Modern Epidemiology. Lippincott Williams & Wilkins; 1998.

49. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol 2010;11(10):R106.

50. Stegle O, Parts L, Piipari M, Winn J, Durbin R. Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. Nat Protoc 2012;7(3):500–7.

51. Lee EK, Kim W, Tominaga K, et al. RNA-binding protein HuD controls insulin translation. Mol Cell 2012;45(6):826–35.

52. Sabeti PC, Fry B, Lohmueller J, et al. Genome-wide detection and characterization of positive selection in human populations. Nature 2007;449(7164):913–8.

53. Danecek P, Auton A, Abecasis G, et al. The variant call format and VCFtools. Bioinformatics 2011;27(15):2156–8.

54. Pritchard JK, Pickrell JK, Coop G. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. Curr Biol 2010;20(4):R208–15.

55. Szpiech ZA, Hernandez RD. selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. Molecular Biology and Evolution 2014;31(10):2824–7.

56. McVicker G, Gordon D, Davis C, Green P. Widespread Genomic Signatures of Natural Selection in Hominid Evolution. PLoS Genetics 2009;5(5):e1000471.

57. Delaneau O, Marchini J, McVean GA, et al. Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. Nature Communications 2014;5:3934.

58. Duncanson A, McCarthy MI, Ouwehand WH, et al. Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. Nat Genet 2007;39(11):1329–37.

59. Thompson SAJ, Jones JL, Cox AL, Compston DAS, Coles AJ. B-Cell Reconstitution and BAFF After Alemtuzumab (Campath-1H) Treatment of Multiple Sclerosis. Journal of Clinical Immunology 2009;30(1):99–105.

60. Cheema GS, Roschke V, Hilbert DM, Stohl W. Elevated serum B lymphocyte stimulator levels in patients with systemic immune-based rheumatic diseases. Arthritis Rheum 2001;44(6):1313–9.

61. Zhang J, Roschke V, Baker KP, et al. Cutting edge: a role for B lymphocyte stimulator in systemic lupus erythematosus. The Journal of Immunology 2001;166(1):6–10.

62. Groom J, Kalled SL, Cutler AH, et al. Association of BAFF/BLyS overexpression and altered B cell differentiation with Sjögren's syndrome. Journal of Clinical Investigation 2002;109(1):59–68.

63. Piazza F, DiFrancesco JC, Fusco ML, et al. Cerebrospinal fluid levels of BAFF and APRIL in untreated multiple sclerosis. Journal of Neuroimmunology 2010;220(1-2):104–7.

64. Mameli G, Cocco E, Frau J, et al. Serum BAFF levels, Methypredsinolone therapy, Epstein-Barr Virus and Mycobacterium avium subsp. paratuberculosis infection in Multiple Sclerosis patients. Sci Rep 2016;6:29268.

65. Furie R, Zamani O, Wallace D, Tegzová D, Petri M. Belimumab, a BLyS-specific inhibitor, reduced disease activity and severe flares in seropositive SLE patients: BLISS-76 study results through wk 76. Arthritis Rheum; 2010.

66. Gross JA, Johnston J, Mudri S, et al. TACI and BCMA are receptors for a TNF homologue implicated in B-cell autoimmune disease. Nature 2000;404(6781):995–9.

67. Ehrenstein MR, Wing C. The BAFFling effects of rituximab in lupus: danger ahead? Nat Rev Rheumatol 2016;12(6):367–72.

68. Mackay F, Schneider P. Cracking the BAFF code. Nat Rev Immunol 2009;9(7):491–502.

69. Hoffmann FS, Kuhn P-H, Laurent SA, et al. The immunoregulator soluble TACI is released by ADAM10 and reflects B cell activation in autoimmunity. J Immunol 2015;194(2):542–52.

70. Krumbholz M, Theil D, Derfuss T, et al. BAFF is produced by astrocytes and up-regulated in multiple sclerosis lesions and primary central nervous system lymphoma. J Exp Med 2005;201(2):195–200.

71. Rauch M, Tussiwand R, Bosco N, Rolink AG. Crucial role for BAFF-BAFF-R signaling in the survival and maintenance of mature B cells. PLoS ONE 2009;4(5):e5456.

72. Palanichamy A, Barnard J, Zheng B, et al. Novel human transitional B cell populations revealed by B cell depletion therapy. J Immunol 2009;182(10):5982–93.

73. Malaspina A, Moir S, Ho J, et al. Appearance of immature/transitional B cells in HIV-infected individuals with advanced disease: correlation with increased IL-7. Proc Natl Acad Sci USA 2006;103(7):2262–7.

74. Jacobi AM, Huang W, Wang T, et al. Effect of long-term belimumab treatment on B cells in systemic lupus erythematosus: extension of a phase II, double-blind, placebo-controlled, dose-ranging study. Arthritis Rheum 2010;62(1):201–10.

75. Dias SS, Rodriguez-Garcia V, Nguyen H, Pericleous C, Isenberg D. Longer duration of B cell depletion is associated with better outcome. Rheumatology (Oxford) 2015;54(10):1876–81.

76. Cappione AJ, Pugh-Bernard AE, Anolik JH, Sanz I. Lupus IgG VH4.34 antibodies bind to a 220-kDa glycoform of CD45/B220 on the surface of human B lymphocytes. The Journal of Immunology 2004;172(7):4298–307.

77. Pruim RJ, Welch RP, Sanna S, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 2010;26(18):2336–7.