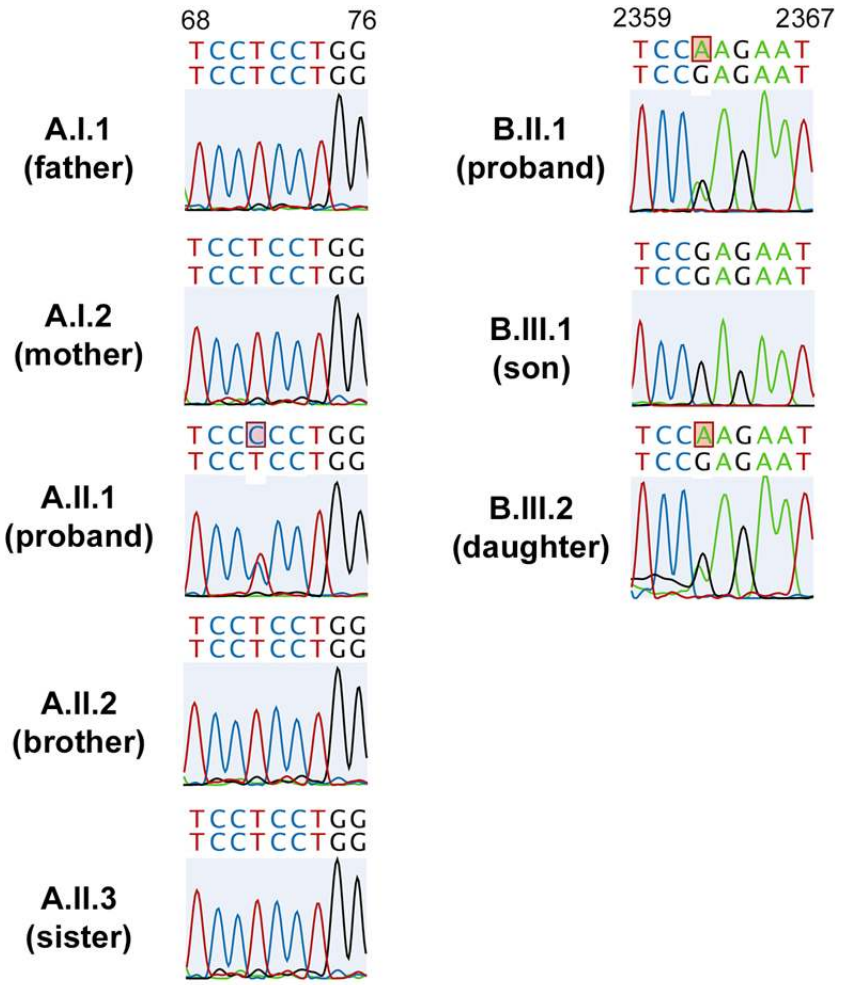
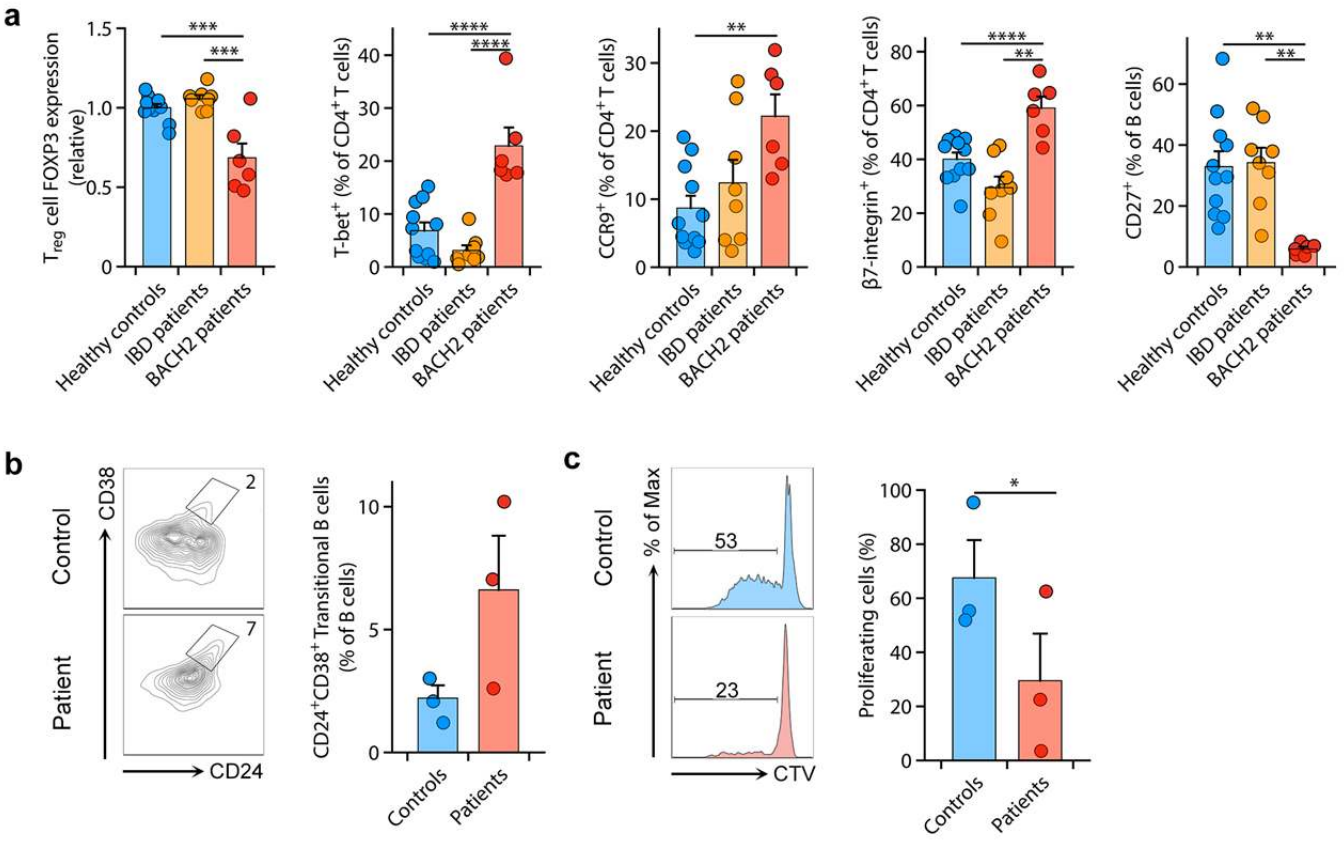


Supplementary Figure 1

Sanger sequencing chromatograms for the two families.

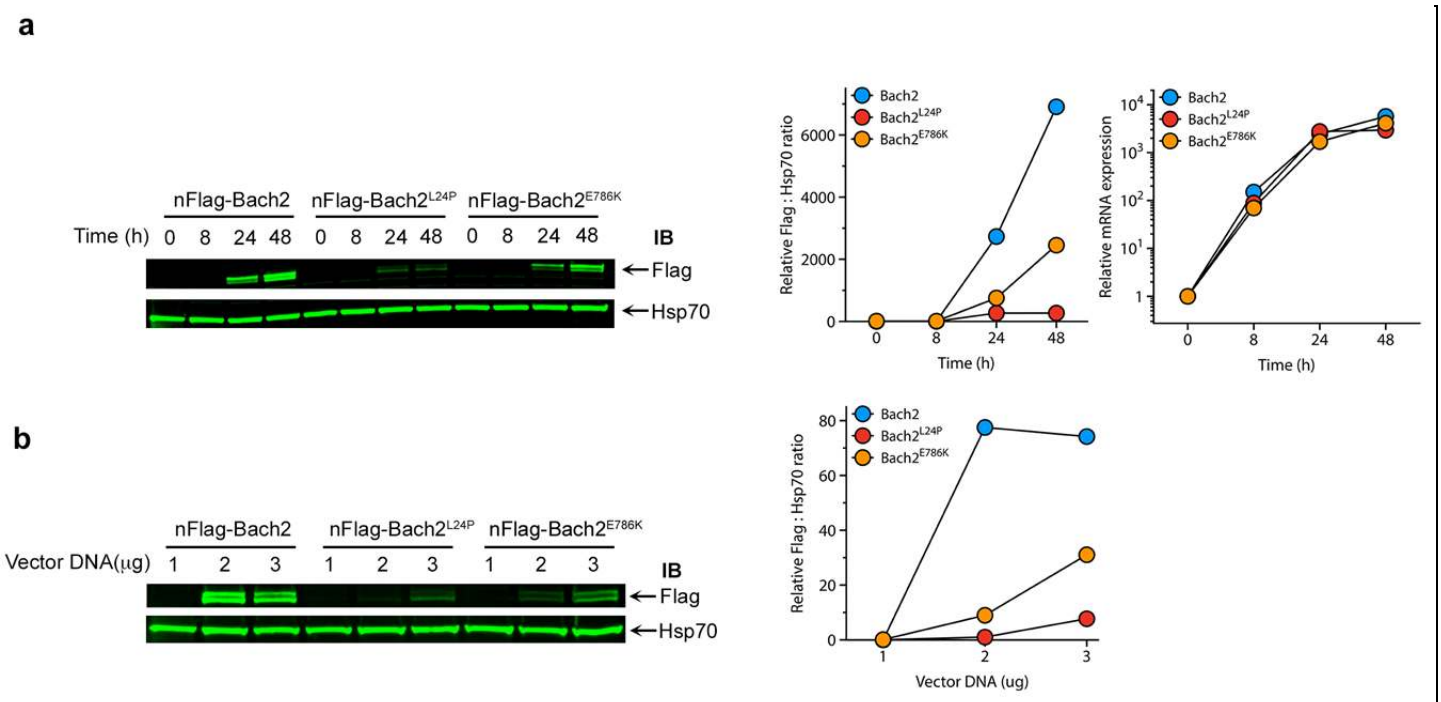
Shown are the Sanger sequencing chromatograms for the two *BACH2* mutations (A.II.1 – left; B.II.1 and B.III.2 – right) and unaffected family members (Family A on the left, family B on the right). For each individual the two alleles of the sequenced region and base positions are shown above the sequencing chromatograms. Subject A.II.1 had a heterozygous T to C mutation at coding position 71 whereas patients B.II.1 and B.III.2 were heterozygous for G to A base substitutions at position 2362.



Supplementary Figure 2

Additional phenotyping of patient cells.

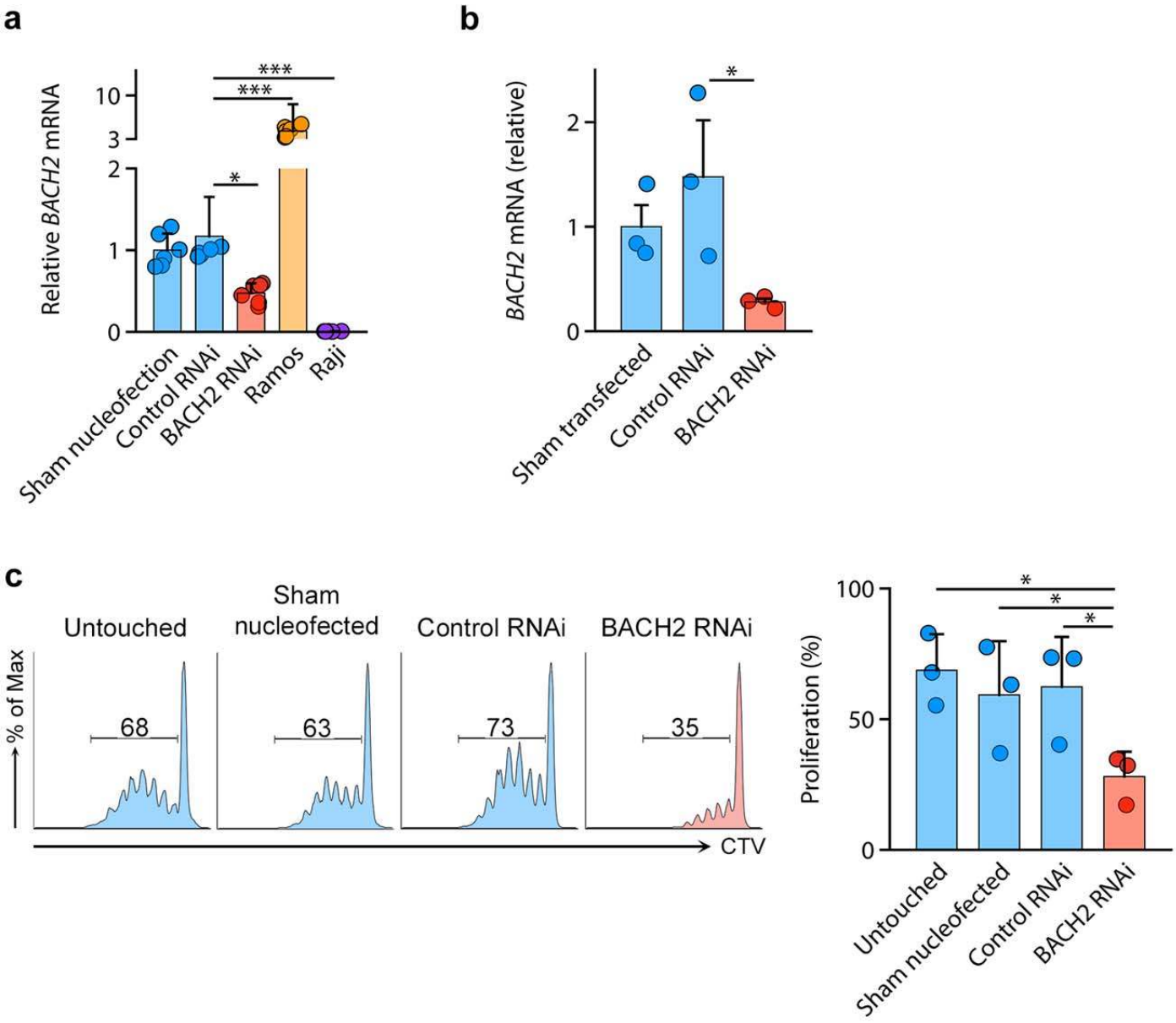
(**a**) Expression of FOXP3, T-bet, CCR9 and β7-integrin in CD4$^+$ T cells (left four panels, respectively) and frequency of memory B cells (right panel) in peripheral blood of healthy (n=11) and disease controls (n=8), compared to the patients with BACH2 mutations. (**b**) CD38$^+$CD24$^+$ transitional B cells[28] in patients compared to controls. Shown are representative flow cytometry plots (left) and cumulative data (right) from all patients and matched controls. Patients had elevated proportions of CD38$^+$CD24$^+$ transitional B cells compared to controls but the observed difference does not reach statistical significance. (**c**) Proliferation (Cell Trace Violet (CTV) dilution) of primary patient CD4$^+$ T cells in response to anti-CD3 and anti-CD28. Shown in **b-c** are representative FACS plots and cumulative data. Bars show mean ± sem throughout. *p<0.05 **p<0.01 ***p<0.001 ****p<0.0001 by ANOVA (**a**) and t-test (**c**).



Supplementary Figure 3

Expression of BACH2 over time and titration of vector DNA

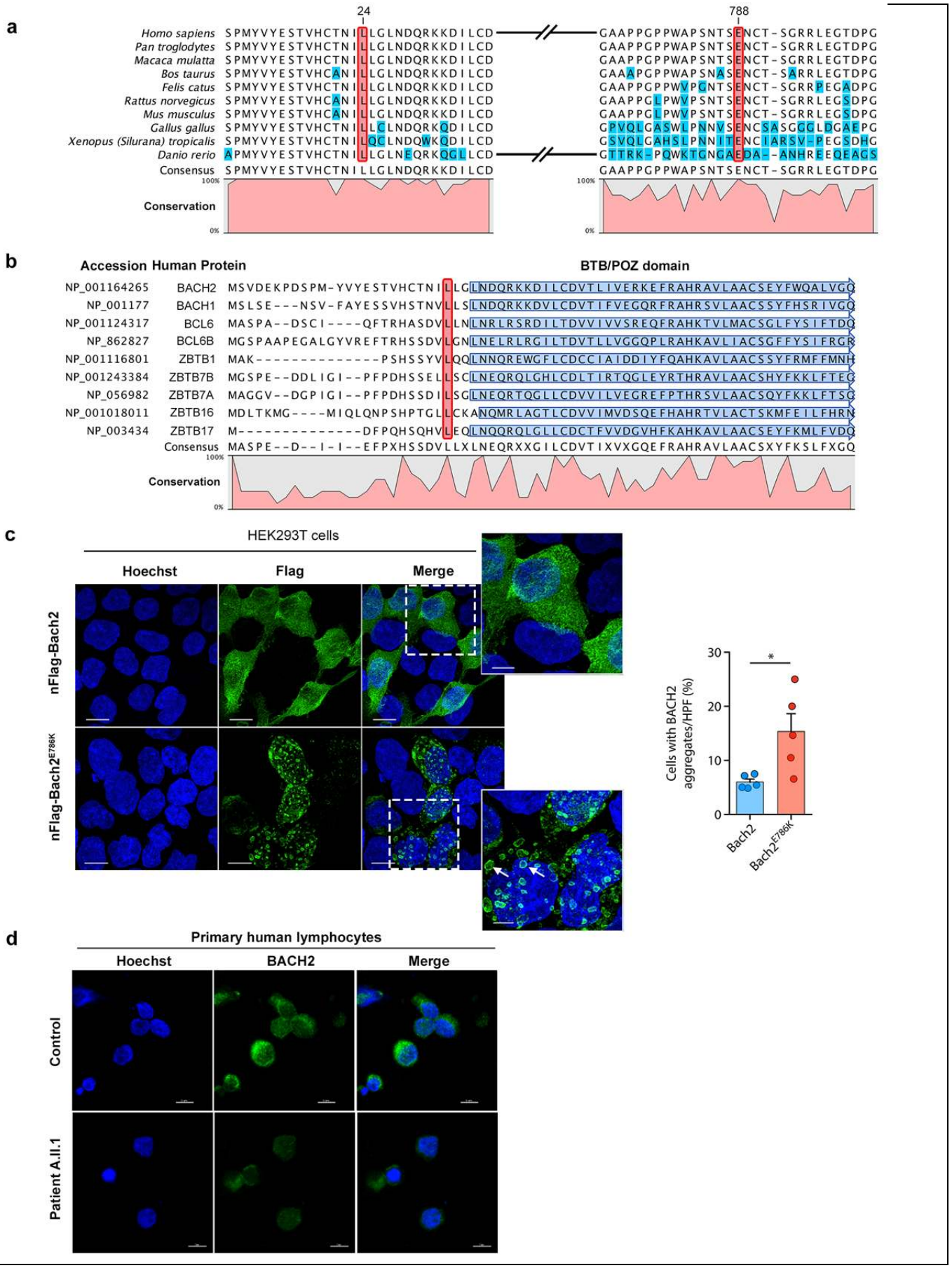
(**a**) Immunoblot showing expression of Flag over time in HEK293T cells transfected with Flag-tagged WT and mutant forms of Bach2 (left panel), with Flag quantification (middle panel) and paired *Bach2* mRNA expression (right panel) over the same time course. (**b**) Immunoblot showing expression of Flag 48h after transfection of HEK293T cells with titrated doses of Flag-tagged WT and mutant *Bach2*-expressing vector (left panel). Flag quantification is shown in the right panel. (**a** and **b**) show representative examples of n=2 experiments.



Supplementary Figure 4

Silencing of *BACH2* in primary healthy donor cells recapitulates T and B cell phenotype of patients

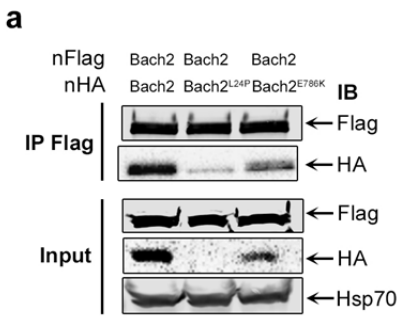
(**a** and **b**) BACH2 mRNA expression in primary healthy donor $CD4^+$ T cells (**a**) and naïve B cells (**b**) transfected with control RNAi or RNAi specific for BACH2. Shown are mean ± sem from n=6 (**a**) and n=3 (**b**) experiments; high (Ramos cell line) and low (Raji cell line) controls for BACH2 expression are additionally shown in **a**. (**c**) Proliferation (Cell Trace Violet (CTV) dilution) of primary healthy control T cells transfected with control RNAi or RNAi specific for BACH2 (n=3 from two independent experiments). Shown are representative flow cytometry examples and cumulative mean ± sem from multiple experiments. *p<0.05 **p<0.01 ***p<0.001 by ANOVA.


a
24
788
Homo sapiens
Pan troglodytes
Macaca mulatta
Bos taurus
Felis catus
Rattus norvegicus
Mus musculus
Gallus gallus
Xenopus (Silurana) tropicalis
Danio rerio
Consensus
Conservation
100%
0%
b
Accession
Human Protein
BTB/POZ domain
NP_001164265 BACH2
NP_001177 BACH1
NP_001124317 BCL6
NP_862827 BCL6B
NP_001116801 ZBTB1
NP_001243384 ZBTB7B
NP_056982 ZBTB7A
NP_001018011 ZBTB16
NP_003434 ZBTB17
Consensus
Conservation
c
HEK293T cells
Hoechst
Flag
Merge
nFlag-Bach2
nFlag-Bach2E786K
Cells with BACH2 aggregates/HPF (%)
30
20
10
0
*
Bach2
Bach2E786K
d
Primary human lymphocytes
Hoechst
BACH2
Merge
Control
Patient A.II.1

Supplementary Figure 5

***BACH2* coding mutations affect conserved amino acids.**

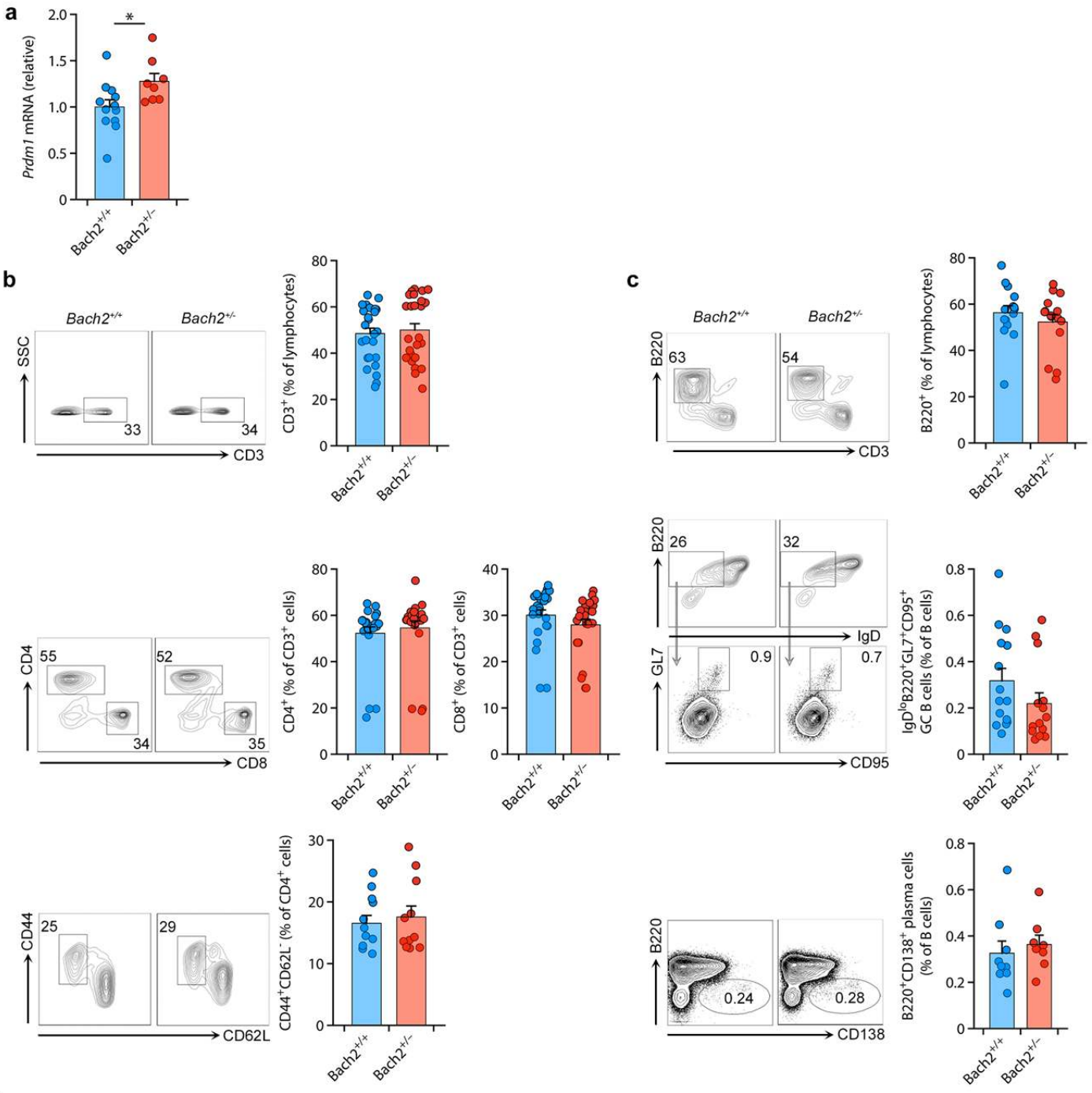
(**a**) Conservation of both mutated residues across species. (**b**) Conservation of N-terminus mutation across other BTB/POZ family members. Highlighted in red are the residues mutated in patients, light blue represents non-conserved residues and purple the location of the BTB/POZ domains. Note that the BTB/POZ domain is from residues 27-133 in human BACH2 but that the first α-helix extends from residues 18-34 of the protein. (**c**) Confocal imaging of HEK293T cells transfected with Flag-tagged murine WT or Bach2^{E786K} and stained for Flag (green) and Hoechst (blue). Insets show enlarged images of single cells and localization of cytoplasmic Flag aggregates (white arrows). Images are representative from 5 independent experiments. The bar graph (right) shows quantification (mean ± sem) of the number of cells containing aggregates per high power field (HPF) from n=5 experiments. White scale bar represents 10 μm in sections and 5 μm in insets. (**d**) Confocal images of primary lymphocytes from healthy control and patient A.II.1 stained for BACH2 (green) and Hoechst (blue). Scale bars: 5 μm. *$p<0.05$ by t-test.



Supplementary Figure 6

Mutant forms of *Bach2* do not exert dominant negative effects

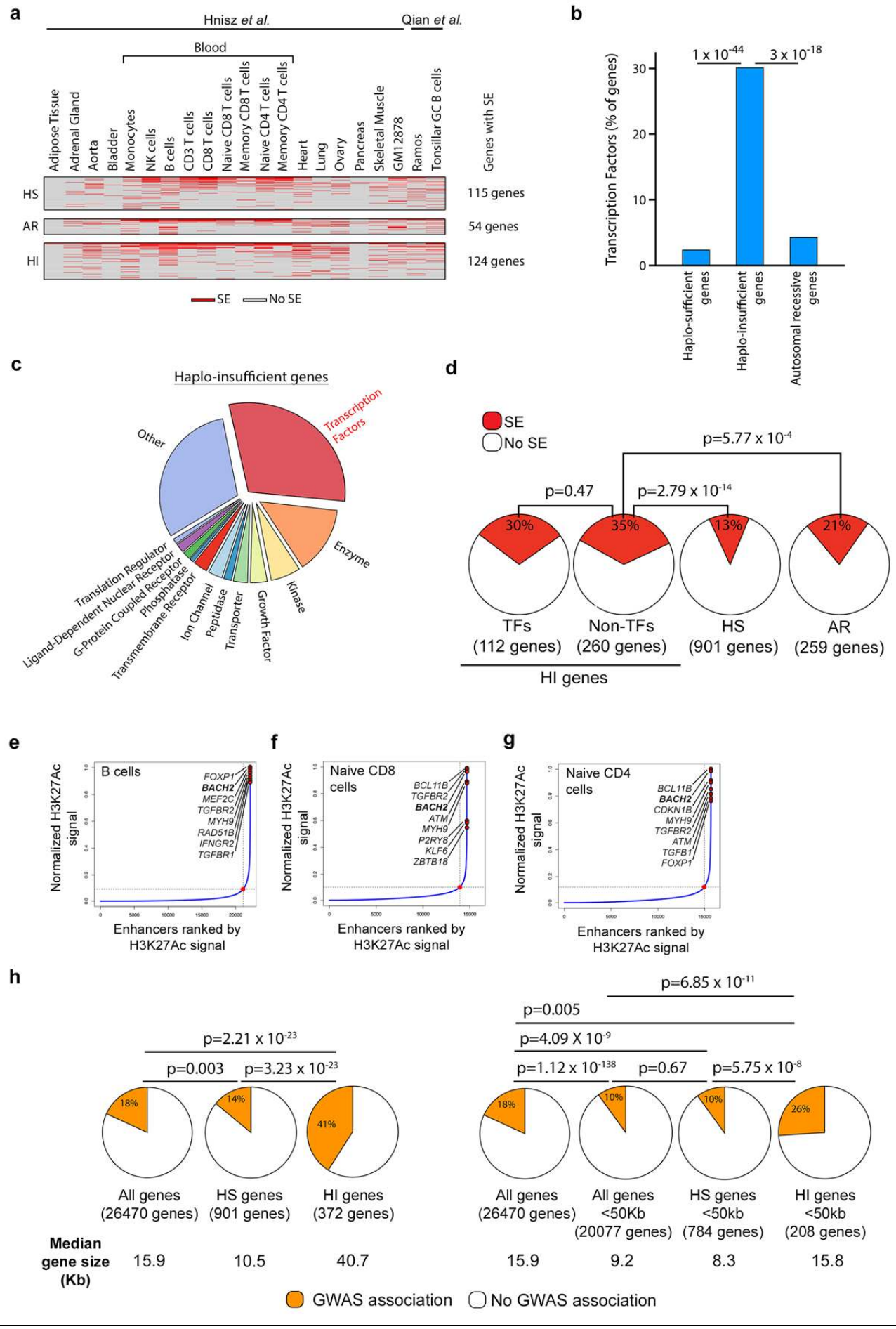
(**a**) Co-immunoprecipitation of Flag-tagged WT with HA-tagged mutant forms of murine *Bach2* transfected into HEK293T cells at 1:1 ratio. Shown is a representative example from n=3 independent experiments.



Supplementary Figure 7

Additional phenotyping of T and B cells of *Bach2* heterozygous mice

(**a**) Expression of *Prdm1* mRNA in B cells of *Bach2*$^{+/+}$ and *Bach2*$^{+/-}$ mice. (**b**) Proportions of CD3$^+$ T cells (top panels), CD4$^+$ and CD8$^+$ T cells (middle panels) and effector CD4$^+$ T cells (CD44$^+$CD62L$^-$; lower panels) in *Bach2*$^{+/-}$ mice compared to *Bach2*$^{+/+}$ littermates. (**c**) B cell subsets in unimmunized heterozygous mice, showing total B cells (top panels), germinal center (GC) B cells (middle panels) and plasma cells (bottom panels). Data in **b** and **c** show representative flow cytometry plots with bar graphs depicting mean ± sem values from a minimum of n=8 mice per group.


a
Hnisz et al.
Qian et al.
Blood
Adipose Tissue
Adrenal Gland
Aorta
Bladder
Monocytes
NK cells
B cells
CD3 T cells
CD8 T cells
Naive CD8 T cells
Memory CD8 T cells
Naive CD4 T cells
Memory CD4 T cells
Heart
Lung
Ovary
Pancreas
Skeletal Muscle
GM12878
Ramos
Tonsillar GC B cells
Genes with SE
HS
AR
HI
115 genes
54 genes
124 genes
SE
No SE
b
Transcription Factors (% of genes)
1 x 10-44
3 x 10-18
Haplo-sufficient genes
Haplo-insufficient genes
Autosomal recessive genes
c
Haplo-insufficient genes
Transcription Factors
Enzyme
Kinase
Growth Factor
Transporter
Peptidase
Ion Channel
Transmembrane Receptor
Phosphatase
G-Protein Coupled Receptor
Ligand-Dependent Nuclear Receptor
Translation Regulator
Other
d
SE
No SE
p=5.77 x 10-4
p=0.47
p=2.79 x 10-14
30%
35%
13%
21%
TFs (112 genes)
Non-TFs (260 genes)
HS (901 genes)
AR (259 genes)
HI genes
e
B cells
Normalized H3K27Ac signal
FOXP1
BACH2
MEF2C
TGFBR2
MYH9
RAD51B
IFNGR2
TGFBR1
Enhancers ranked by H3K27Ac signal
f
Naive CD8 cells
BCL11B
TGFBR2
BACH2
ATM
MYH9
P2RY8
KLF6
ZBTB18
g
Naive CD4 cells
BCL11B
BACH2
CDKN1B
MYH9
TGFBR2
ATM
TGFB1
FOXP1
h
p=2.21 x 10-23
p=0.003
p=3.23 x 10-23
18%
14%
41%
All genes (26470 genes)
HS genes (901 genes)
HI genes (372 genes)
p=6.85 x 10-11
p=0.005
p=4.09 X 10-9
p=1.12 x 10-138
p=0.67
p=5.75 x 10-8
18%
10%
10%
26%
All genes (26470 genes)
All genes <50Kb (20077 genes)
HS genes <50kb (784 genes)
HI genes <50kb (208 genes)
Median gene size (Kb)
15.9
10.5
40.7
15.9
9.2
8.3
15.8
GWAS association
No GWAS association

Supplementary Figure 8

Superenhancer (SE) and GWS association of haploinsufficient genes

(**a**) Tissue distribution of genes with SE structure. Shown are only the genes with SE structure in the haplosufficient (HS, top panel), autosomal recessive (AR, middle panel) and haploinsufficient (HI, bottom panel) lists. The presence (red fill) or absence (grey fill) of an SE structure in each tissue is indicated. Each row represents a single gene. Tissue types are shown in columns. Indicated in the "overall" column, is the number of genes with SE architecture in at least one tissue type (see also **supplementary Table 3**). Source data are indicated. (**b-c**) Percentage of genes among HS, HI and AR genes that are transcription factors (**b**) and pie-chart showing function of genes causing HI diseases (**c**). Source data for **b** and **c** are from Qiagen Ingenuity Pathway Analysis. (**d**) Percentage of SE-regulated genes among transcription factor and non-transcription factor HI genes in comparison to HS and AR genes. (**e-g**) Ranked order of H3K27Ac-loaded enhancers in human B cells (**e**), naïve $CD8^+$ (**f**) and naïve $CD4^+$ (**g**) T cells. Indicated are the relative positions, ranked according to signal intensity (higher = greater signal intensity), of the top 8 genes from the HI gene list in those cells. Source data from[29]. (**h**) Percentage of genome-wide association study (GWAS)-associated genes in HS and HI genes compared to all genes, with (right) and without (left) correction for size. Shown are median gene sizes for each gene category. All p-values are from exact Fisher tests.