nature
immunology

BACH2 immunodeficiency illustrates an association between super-enhancers and haploinsufficiency

Behdad Afzali[1,2,17], Juha Grönholm[3,17], Jana Vandrovcova[4,5,17], Charlotte O'Brien[5], Hong-Wei Sun[1], Ine Vanderleyden[6], Fred P Davis[1], Ahmad Khoder[5], Yu Zhang[3] , Ahmed N Hegazy[7,8], Alejandro V Villarino[1] , Ira W Palmer[1], Joshua Kaufman[1], Norman R Watts[1], Majid Kazemian[9], Olena Kamenyeva[3], Julia Keith[7], Anwar Sayed[5], Dalia Kasperaviciute[10], Michael Mueller[10], Jason D Hughes[11], Ivan J Fuss[3], Mohammed F Sadiyah[6], Kim Montgomery-Recht[12], Joshua McElwee[11], Nicholas P Restifo[13], Warren Strober[3], Michelle A Linterman[6], Paul T Wingfield[1], Holm H Uhlig[7,14], Rahul Roychoudhuri[6] , Timothy J Aitman[5,15], Peter Kelleher[5], Michael J Lenardo[3], John J O'Shea[1], Nichola Cooper[5,18] & Arian D J Laurence[7,16,18]

The transcriptional programs that guide lymphocyte differentiation depend on the precise expression and timing of transcription factors (TFs). The TF BACH2 is essential for T and B lymphocytes and is associated with an archetypal super-enhancer (SE). Single-nucleotide variants in the *BACH2* locus are associated with several autoimmune diseases, but *BACH2* mutations that cause Mendelian monogenic primary immunodeficiency have not previously been identified. Here we describe a syndrome of BACH2-related immunodeficiency and autoimmunity (BRIDA) that results from *BACH2* haploinsufficiency. Affected subjects had lymphocyte-maturation defects that caused immunoglobulin deficiency and intestinal inflammation. The mutations disrupted protein stability by interfering with homodimerization or by causing aggregation. We observed analogous lymphocyte defects in *Bach2*-heterozygous mice. More generally, we observed that genes that cause monogenic haploinsufficient diseases were substantially enriched for TFs and SE architecture. These findings reveal a previously unrecognized feature of SE architecture in Mendelian diseases of immunity: heterozygous mutations in SE-regulated genes identified by whole-exome/genome sequencing may have greater significance than previously recognized.

The inheritance patterns of genetic diseases fall along a spectrum that ranges from the vast majority, representing polygenic susceptibility variants (usually identified in genome-wide associate studies (GWAS)), to the minority, which are monogenic and manifest in either a recessive or a dominant manner. It is now appreciated that mutations in more than 300 different genes can cause primary immunodeficiencies (PIDs), many of which affect the function of T and B lymphocytes[1–4]. PIDs are often paradoxically associated with autoimmunity[3–7]. Common variable immunodeficiency (CVID), a major form of PID with antibody deficiency, is typically associated with recurrent infections and autoimmunity[8]. Recently developed gene-sequencing technologies now allow for the rapid identification of PIDs, but they have also raised the important question of how to interpret the many heterozygous mutations seen in both affected individuals and healthy controls. Relatively few PID syndromes are caused by haploinsufficiency, an autosomal dominant pattern of disease inheritance in which one allele is damaged and only a single functional allele remains[9]. Genes such as *CTLA4* are particularly susceptible to haploinsufficiency, although the reasons for this are unknown[10]. Given that many healthy people harbor heterozygous loss-of-function or hypomorphic variants, why should partial changes in gene expression have significant consequences for health?

[1]Lymphocyte Cell Biology Section (Molecular Immunology and Inflammation Branch), Biodata Mining and Discovery Section and Protein Expression Laboratory, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, Maryland, USA. [2]MRC Centre for Transplantation, King's College London, London, UK. [3]Molecular Development of the Immune System Section, NIAID Clinical Genomics Program, Biological Imaging Section (Research Technologies Branch) and Mucosal Immunity Section, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland, USA. [4]Molecular Neuroscience, Institute of Neurology, Faculty of Brain Sciences, University College London, London, UK. [5]Department of Medicine, Imperial College London, London, UK. [6]Laboratory of Lymphocyte Signaling and Development, Babraham Institute, Cambridge, UK. [7]Translational Gastroenterology Unit, Nuffield Department of Medicine, John Radcliffe Hospital, Oxford, UK. [8]Kennedy Institute of Rheumatology, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK. [9]Department of Biochemistry and Department of Computer Science, Purdue University, West Lafayette, Indiana, USA. [10]Imperial BRC Genomics Facility, Hammersmith Hospital, London, UK. [11]Merck Research Laboratories, Merck & Co. Inc., Boston, Massachusetts, USA. [12]Clinical Research Directorate/CMRP, Leidos Biomedical Research Inc., NCI at Frederick, Frederick, Maryland, USA. [13]National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA. [14]Department of Paediatrics, University of Oxford, Oxford, UK. [15]Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK. [16]Department of Haematology, Northern Centre for Cancer Care, Newcastle upon Tyne, UK. [17]These authors contributed equally to this work. [18]These authors jointly supervised this work. Correspondence should be addressed to B.A. (behdad.afzali@nih.gov) or N.C. (n.cooper@imperial.ac.uk).

Promoters and enhancer elements govern gene expression. Most genes, such as the housekeeping gene *ACTB*, are regulated by a limited number of associated enhancers, known as 'typical enhancers'[11]. By contrast, 5–10% of genes have a complex enhancer structure consisting of multiple enhancers that collectively are described as SEs[12,13]. Genes with associated SEs have a highly regulated pattern of gene expression; single-nucleotide polymorphisms shown to be associated with autoimmune diseases in GWAS are preferentially enriched in SE regions[14]. These findings suggest that minor changes in regulatory function at SE regions could have significant consequences for genes regulated by SEs, and thus potentially for the immune system.

BACH2 is a typical example of an SE-regulated gene associated with autoimmune disease. The protein it encodes is a highly conserved member of the basic leucine zipper (bZIP)-domain superfamily of TFs, and a critical regulator of differentiation and maturation in both T and B lymphocytes[15,16]. Polymorphisms in the human gene locus are associated with several autoimmune diseases, including asthma[17], insulin-dependent diabetes mellitus[18], Crohn's and celiac diseases[19,20], vitiligo[21] and multiple sclerosis[16,22]. The *Bach2* gene locus has the largest SE structure seen in mouse lymphocytes[14]. Homozygous deletion of *Bach2* in mice results in spontaneous fatal autoimmunity between 3 and 9 months of age[15]. Functionally, BACH2 acts as a repressive 'guardian' TF that regulates the balance among a network of other TFs critical to the specification and maturation of T and B cells. In B cells, BACH2 controls the balance between PAX5 and BLIMP1 by repressing the latter[23,24] to decelerate plasma cell differentiation and permit antibody class-switch recombination (CSR) (thereby allowing the expression of IgA, IgG and IgE isotypes)[25]. Consequently, mice that lack BACH2 have B cells with impaired CSR that rapidly differentiate into IgM-restricted plasma cells. In T cells, BACH2 regulates networks of genes that control effector T cell lineages[14] and cellular senescence[26], thus limiting differentiation into effector T cells[15] and promoting the development of FoxP3$^+$ regulatory T cells (T$_{reg}$ cells). T$_{reg}$ cells are a nonredundant suppressive lineage of T cells that prevent the development of autoimmune diseases by controlling overactivation of the immune system[27]. Thus, mice deficient in BACH2 have both a paucity of T$_{reg}$ cells and an excess of memory/effector T cells that age and die prematurely, resulting in autoimmunity.

Structurally, BACH2 contains a BTB/POZ domain that mediates homo- and heterodimerization at its N terminus, and a bZIP domain at the C terminus that is required for DNA binding. The dimerization domain is an α-helical structure that contains a cysteine residue capable of forming a disulfide bond with its opposite partner[28]. Thus homodimerization is likely to be stabilized by a covalent modification that occurs soon after protein folding. BACH2 dimers translocate to the nucleus, where they interact with target DNA loci at palindromic Maf recognition elements, either alone or in collaboration with other members of the bZIP family, such as the small Maf proteins (MafF, MafG and MafK)[16]. This interaction—for example, at the mouse *Prdm1* locus that encodes BLIMP1—represses gene expression.

Here we describe a novel PID caused by haploinsufficiency of *BACH2*, and propose a shared genetic mechanism to explain why some genes are particularly likely to be associated with disease caused by haploinsufficiency. We conclude that the interpretation of heterozygote variants in these genes should be regarded as significant and should be prioritized in any investigation of novel genetic disease by whole-exome sequencing.

RESULTS

### *BACH2* mutations are associated with CVID and colitis

We investigated a woman (**Fig. 1a,b**; family A) with infancy-onset colitis who became ill at 19 years old with noninfectious fever, splenomegaly

(21.7 cm, compared with 10–12 cm in healthy adults) (**Fig. 1c**) and pancytopenia. Fever and cytopenia improved with corticosteroids, but lymphopenia, immunoglobulin deficiency (IgM, IgG, IgA and IgE), ongoing colitis, lung infiltrates and recurrent upper respiratory tract infections persisted (**Fig. 1c**, **Table 1** and **Supplementary Table 1**). A colonic biopsy showed inflammatory changes with crypt branching and prominent lymphocytic infiltrates around the crypts (**Fig. 1d**), with significantly reduced numbers of FoxP3$^+$ T$_{reg}$ cells compared with those in healthy controls or subjects with classical inflammatory bowel disease (IBD) (**Fig. 1e**). The early disease onset and unusual symptoms combined with the absence of family history prompted us to carry out whole-exome sequencing on this subject and her healthy parents as a trio. After we had excluded all variants with minor allele frequencies of >0.01, no candidate variants remained to support a hypothesis of recessive inheritance. We found a novel heterozygous *de novo*, nonsynonymous mutation in *BACH2*, c.T71C, that was predicted to be deleterious (**Supplementary Table 2**), substituting a highly conserved leucine by proline (L24P), and was not present in healthy family members (**Fig. 1b** and **Supplementary Fig. 1**). A second family (family B) that had been previously investigated by exome sequencing (**Fig. 1a**) was found to have a heterozygous point mutation in *BACH2*, c.G2362A (**Fig. 1b**), in which glutamic acid was substituted by lysine (E788K), in a father and daughter, both of whom presented with inflammation of both small and large bowel, as well as pulmonary disease including recurrent sino-pulmonary infections, bronchiectasis and fibrosis (**Fig. 1c** and **Supplementary Table 1**). The *BACH2* mutation was not seen in healthy family members (**Supplementary Fig. 1**). The father (proband) was deficient in all immunoglobulin subtypes; the daughter had undetectable levels of IgA (**Supplementary Table 1**). Detailed clinical features are described in **Supplementary Note 1**, **Table 1** and **Supplementary Figure 1**. We found no low-minor-allele-frequency variants or causative mutations in genes that cause monogenic IBD or other recognized PIDs[29–31].

In the lymphocytes of affected individuals, we found decreased expression of FoxP3 in CD4$^+$CD25^{hi}CD127^{lo} T$_{reg}$ cells (**Fig. 2a**), and increased expression of the type 1 helper T cell (T$_H$1 cell) TF T-bet and two gut-homing receptors, CCR9 and β$_7$ integrin, on CD4$^+$ T cells[32,33] (**Fig. 2b**). In B cells from these subjects, we found a marked reduction in CD19$^+$CD27$^+$ memory and IgG class-switched CD27$^+$IgG$^+$ B cells (**Fig. 2c**). These features were not present in healthy controls or in subjects with IBD (**Supplementary Fig. 2a**). Furthermore, numbers of CD24$^+$CD38$^+$ transitional B cells were increased in affected subjects (**Supplementary Fig. 2b**). *In vitro* activation of naive B cells from affected individuals resulted in significantly impaired plasmablast generation, CSR and class-switched antibody secretion in the presence of IL-21 (**Fig. 2d,e**), which suggested a defect in B cell maturation toward memory and plasma cells, similar to that seen in *Bach2*-knockout mice[34]. Polyclonal activation of T cells resulted in reduced CD4$^+$ T cell proliferation compared with that in healthy controls (**Supplementary Fig. 2c**). In summary, the immunophenotype of individuals with mutations in *BACH2* consisted of compromised T$_{reg}$ cells, enhanced T$_H$1 cell differentiation, impaired proliferation and defective maturation of B cells, and immunoglobulin class-switching.

BACH2 silencing mimics immunodeficient cell phenotypes

We next measured BACH2 protein expression by flow cytometry and found that it was reduced in CD4$^+$, CD8$^+$ and B lymphocytes in affected subjects, despite normal mRNA expression in these people compared with that in healthy controls (**Fig. 3a,b**). We measured the protein expression of Flag-tagged vectors encoding wild-type or mutant forms of BACH2 in transfected HEK293T cells and found
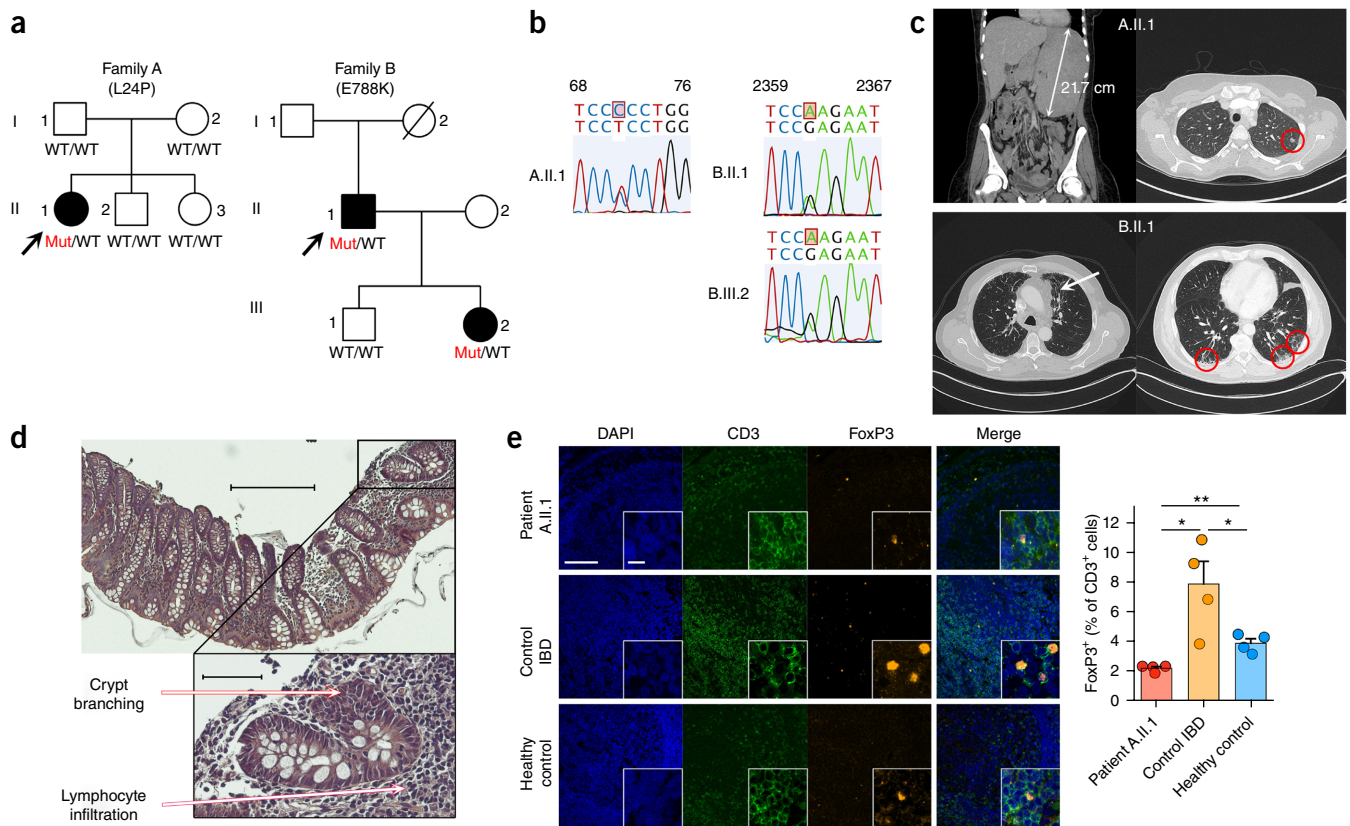
Figure 1 Pedigrees and phenotypes of subjects with mutations in *BACH2*. (**a**) The pedigrees of two families with heterozygous missense coding mutations in *BACH2* that result in L24P or E788K amino acid substitutions. Shown are affected heterozygotes (black symbols) and unaffected family members (white symbols). Arrows indicate probands. WT, wild-type allele; Mut, mutant allele. (**b**) Sanger-sequencing chromatograms for the affected individuals from both families. For each individual, the two alleles of the sequenced region of *BACH2* and base positions are shown above the chromatograms. Subject A.II.1 had a heterozygous T-to-C mutation at coding position 71; patients B.II.1 and B.III.2 were heterozygous for G-to-A base substitutions at position 2362. (**c**) Computerized tomography scans showing splenomegaly (arrow in the upper left image) and pulmonary nodules (red circle in the upper right image) in patient A.II.1, and bronchiectasis (dilated airways; arrow in the lower left image) and fibrosis ('honeycombing' within red circles in the lower right image) in patient B.II.1. (**d**) Photomicrographs of a hematoxylin-and-eosin-stained section from a colonic biopsy from patient A.II.1, showing crypt branching and lymphocytic inflammatory infiltrates around the crypts. Scale bars, 200 μm (main image) or 50 μm (inset). (**e**) Colonic biopsies from patient A.II.1, a control subject with IBD, and a healthy control, showing immunofluorescent staining for nuclear DNA (DAPI; blue), CD3 (green) and FoxP3 (orange). Shown are representative sections (left) and cumulative (mean and s.e.m.) quantification (right) from four low-power fields per patient (500–3,000 CD3$^+$ cells counted per low-power field). Scale bars, 100 μm (main image) or 2 μm (inset). *$P < 0.05$, **$P < 0.01$, *t*-test.

that mutant forms of the protein accumulated less than the wild-type protein (**Fig. 3c**) at all time points measured and concentrations of vector used (**Supplementary Fig. 3a,b**). *PRDM1*, which encodes the protein BLIMP1, is a target of BACH2-mediated transcriptional repression[24]. We found that naive B cells and CD4$^+$ T cells from affected subjects expressed significantly higher levels of *PRDM1* mRNA than observed in healthy controls, which suggested a release from BACH2 repression (**Figs. 3d,e**). Furthermore, this difference could be reversed by forced expression of wild-type BACH2 in CD4$^+$ lymphocytes of affected subjects (**Fig. 3e**). These observations suggested a causal relationship between reduced BACH2 expression and cellular phenotype. To confirm this relationship, we silenced BACH2 expression in T and B cells from healthy controls by ~50% via RNA interference (RNAi) and carried out functional phenotyping (**Supplementary Fig. 4a,b**). BACH2 silencing in control CD4$^+$ T cells led to a significant increase in *PRDM1* mRNA expression (**Fig. 3f**) and resulted in reduced proliferation of CD4$^+$ T cells (**Supplementary Fig. 4c**) compared with that in healthy controls, similar to what we observed in primary CD4$^+$ T cells from affected subjects (**Supplementary Fig. 2c**). In addition, BACH2 silencing in B cells from healthy controls significantly suppressed *in vitro* CSR

toward the IgG and IgA isotypes (**Fig. 3g**). Thus, experimental silencing of BACH2 in healthy T and B cells recapitulated the phenotype seen in primary cells of affected subjects.

BACH2 mutations impair protein stability
Both of the mutations that we identified affect highly conserved amino acid residues in BACH2 (**Fig. 4a**). Mouse and human BACH2 share 90% sequence identity, and Leu24 is conserved across species and across other members of the BTB/POZ domain family (**Supplementary Fig. 5a,b** and **Supplementary Table 2**). Leu24 resides within α-helix 1 (residues 18–34) of the BTB/POZ domain, a key part of the BACH2 homodimerization interface (**Fig. 4b,c**). The mutant proline residue probably perturbs α-helix 1 of the BTB/POZ domain and places a polar residue into the hydrophobic face of that helix, which we predicted would decrease dimer stability (**Supplementary Table 2**). We expressed and purified the BTB domains from both wild-type and L24P mutant proteins. The wild-type protein was soluble and formed dimers (**Fig. 4d**), whereas the L24P mutant was insoluble in solution, was probably misfolded, and formed multiple aggregated species (**Fig. 4e**). Glu788, the site of the C-terminus mutation, was again highly conserved (**Supplementary**

Table 1 Summary clinical characteristics of subjects with missense mutations in BACH2

| Demographic and clinical characteristics	Patient		
	A.II.1	B.II.1	B.III.2
Age, sex	19, F	63, M	40, F
Lymphadenopathy	Yes	Yes	Yes
Splenomegaly	Yes	No	No
Intestinal manifestations	Yes	Yes	Yes
Chronic diarrhea	Yes	Yes	Yes
IBD	Colitis	Not biopsied	UC at age 10; Crohn's disease at age 32
Pulmonary manifestations	Yes	Yes	Yes
Recurrent sino-pulmonary infections	Yes	Yes	Yes
Radiographic changes on chest CT	Yes	Yes	Not imaged
Immunoglobulins[a]			
IgM	Low	Low	High
IgG	Low	Low	High[b]
IgA	Low	Low	Low
IgE	Low	Low	Normal
On IvIg treatment	Yes	Yes	No
EBV antibodies	N/A (DNA negative)	N/A	High
RhF	N/A	N/A	N/A
dsDNA antibodies	Negative	N/A	N/A
ANCA	Positive (p-ANCA)[c]	N/A	N/A
ANA	Negative	N/A	Negative

IvIg, intravenous immunoglobulin; EBV, Epstein–Barr virus; RhF, rheumatoid factor; dsDNA, double-stranded DNA; ANCA, anti-neutrophil cytoplasmic antibody; p-ANCA, perinuclear ANCA; ANA, antinuclear antibody; UC, ulcerative colitis; N/A, not assessed.
[a]Absolute values are given in **Supplementary Table 1**. [b]Progressive decline in IgG levels. [c]Positive by immunofluorescence but negative for myeloperoxidase and proteinase III antibodies by ELISA.

Fig. 5a). Though it has not been characterized by structural studies, it is in proximity to a nuclear export signal (**Fig. 4a**). We found that wild-type BACH2 protein was evenly distributed in both cytoplasm and nucleus, whereas the E788K mutant protein aggregated in the cytoplasm, with relatively little found in the nucleus (**Fig. 4f** and **Supplementary Movies 1** and **2**). We observed similar protein
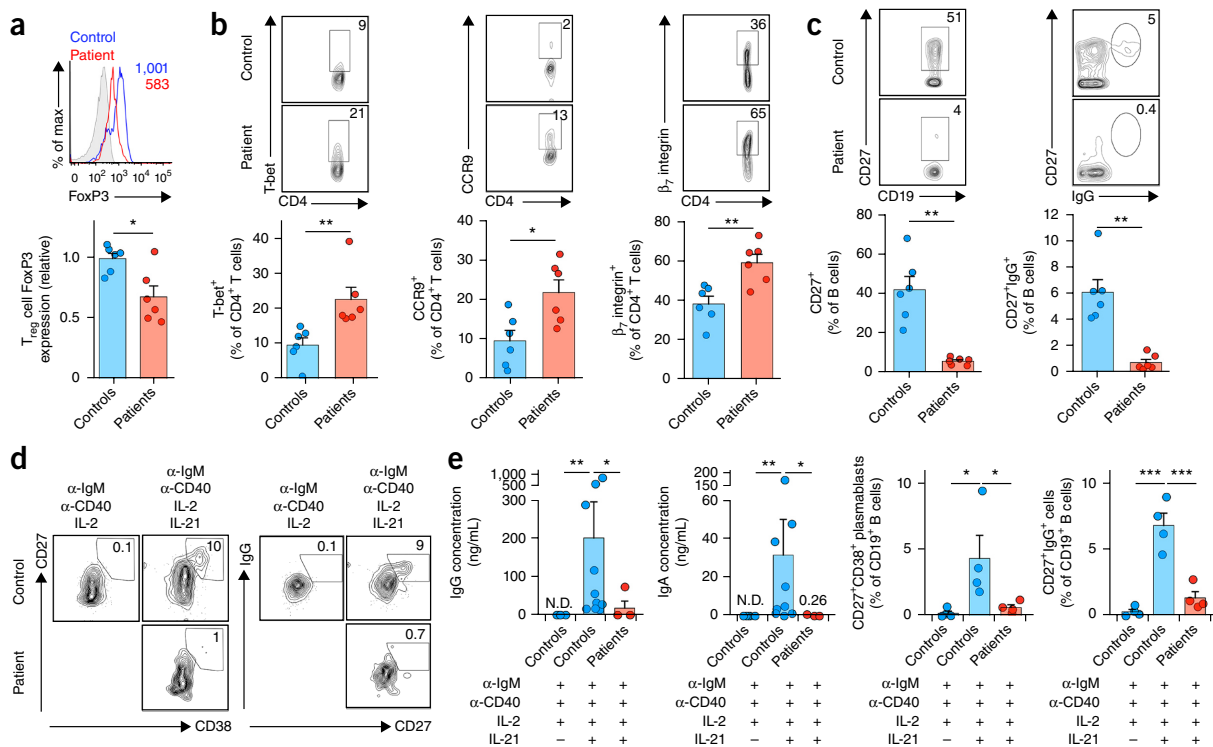
Figure 2 Immunophenotypes of patients with mutations in *BACH2*. (a–c) T$_{reg}$ cell (a), T cell (b) and B cell (c) immunophenotypes of peripheral blood cells from affected subjects and healthy controls. Shown are total FoxP3 expression (mean fluorescent intensity) in CD4$^+$CD25^{hi}CD127^{lo} cells (a), expression of the TF T-bet and of gut-homing receptors (CCR9 and β$_7$ integrin) in bulk CD4$^+$ T cells (b), and total memory (c, left) and class-switched memory B cells (c, right) in bulk B cells. (d,e) Plasmablast formation (d, left), IgG CSR (d, right) and immunoglobulin secretion (e) in naive B cells from affected subjects and healthy controls, activated *in vitro* as indicated. Shown are representative flow cytometry plots (d) and cumulative data (e). N.D., not detected; very low values are shown above the respective bar for clarity. In a–d, representative flow cytometry plots are shown with cumulative data from all patients and matched controls. In all flow cytometry plots, numbers in corners indicate the percentage of cells in the gate. Bar graphs show mean and s.e.m. throughout; data points represent individual subjects. Data are from $n = 3$ (a–c) or 4 (d,e) independent experiments. Note that IgG secretion in e does not include that for patient B.III.2, who had normal IgG secretion. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, determined by *t*-test (a–c), one-way analysis of variance (ANOVA) (d) or Kruskal–Wallis test (e).
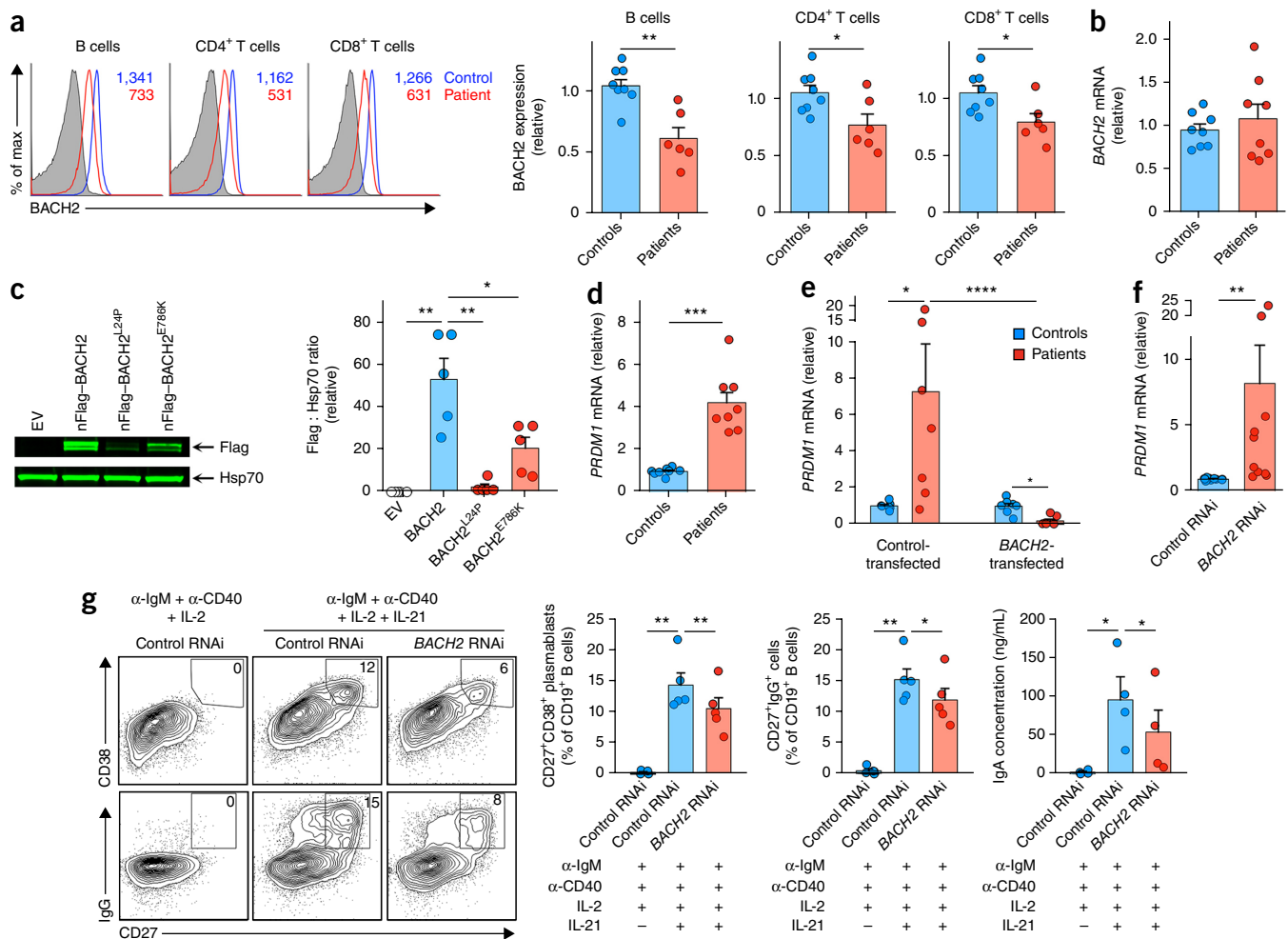
Figure 3 The cellular phenotype is attributable to reduced BACH2 protein expression. (**a**) BACH2 protein expression in primary immune cells from affected subjects and healthy controls. Shown are representative flow cytometry plots (left; numbers indicate the mean fluorescence intensity) and cumulative BACH2 protein expression (right) in affected subjects relative to controls ($n = 3$ independent experiments). (**b**) Cumulative *BACH2* mRNA expression from naive B cells of affected subjects and healthy controls ($n = 3$ experiments). (**c**) Results of immunoblotting for Flag and Hsp70 from lysates of HEK293T cells transfected with empty vector (EV) and Flag-tagged wild-type or mutant mouse BACH2 (L24P or E786K, the mouse equivalent of E788K). Shown are a representative blot (left) and cumulative quantifications from $n = 5$ experiments (right). (**d**) *PRDM1* mRNA expression in naive B cells from affected subjects and healthy controls ($n = 3$ experiments). (**e**,**f**) *PRDM1* mRNA expression in CD4+ T lymphocytes from healthy controls and affected subjects transfected with either control (eGFP) or *BACH2* mRNA ($n = 4$ experiments) (**e**) and in healthy donor CD4+ T lymphocytes subjected to *BACH2* RNAi or control RNAi ($n = 4$ experiments) (**f**). (**g**) Plasmablast formation, IgG CSR and IgA secretion in naive B cells from healthy controls subjected to control RNAi or *BACH2* RNAi and activated *in vitro* as shown. The panel shows representative flow cytometry plots and cumulative data ($n = 5$, 5 and 4 experiments, respectively, from left to right). Numbers in corners indicate the percentage of cells in the gate. Bar graphs show mean and s.e.m. throughout; data points represent either individual subjects (**a**,**b**, **d**–**g**) or individual experiments (**c**). *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$ by t-test (**a**,**d**), Wilcoxon test (**f**) or ANOVA (**c**,**e**,**g**).

aggregates in HEK293T cells transfected with this C-terminal variant (**Supplementary Fig. 5c** and **Supplementary Movies 3** and **4**). By contrast, aggregates were not detected in lymphocytes from patient A.II.1 that expressed the L24P mutant, although, as noted, expression levels of L24P-mutant BACH2 protein were lower than BACH2 expression levels in controls (**Supplementary Fig. 5d**). Thus, both gene mutations impair BACH2 protein stability.

BACH2 mutations are not dominant negative

In both families studied, the *BACH2* gene mutations had the potential to act in a dominant negative manner. To investigate this, we cotransfected HEK293T cells with Flag-tagged wild-type protein together with constructs expressing untagged wild-type or mutant BACH2. Neither mutant altered wild-type protein expression (**Fig. 5a**).

We repeated the experiment with HEK293T cells cotransfected with vectors encoding two tagged (hemagglutinin (HA)–BACH2 and Flag–BACH2) forms of the wild-type protein together with constructs encoding untagged wild-type or mutant BACH2 protein (**Fig. 5b**). Coimmunoprecipitation studies showed that untagged wild-type BACH2, but not mutant forms of the protein, interfered with dimerization between HA- and Flag-tagged wild-type BACH2. Furthermore, when we cotransfected Flag-tagged wild-type BACH2 with HA-tagged wild-type, L24P-mutant or E786K-mutant BACH2, we detected reduced amounts of mutant HA–BACH2 bound to Flag-tagged wild-type BACH2 after immunoprecipitation, in proportion to the reduction in protein accumulation, which implied limited, if any, effects on wild-type BACH2 (**Supplementary Fig. 6a**). All these results were consistent with our earlier findings of loss of stability of
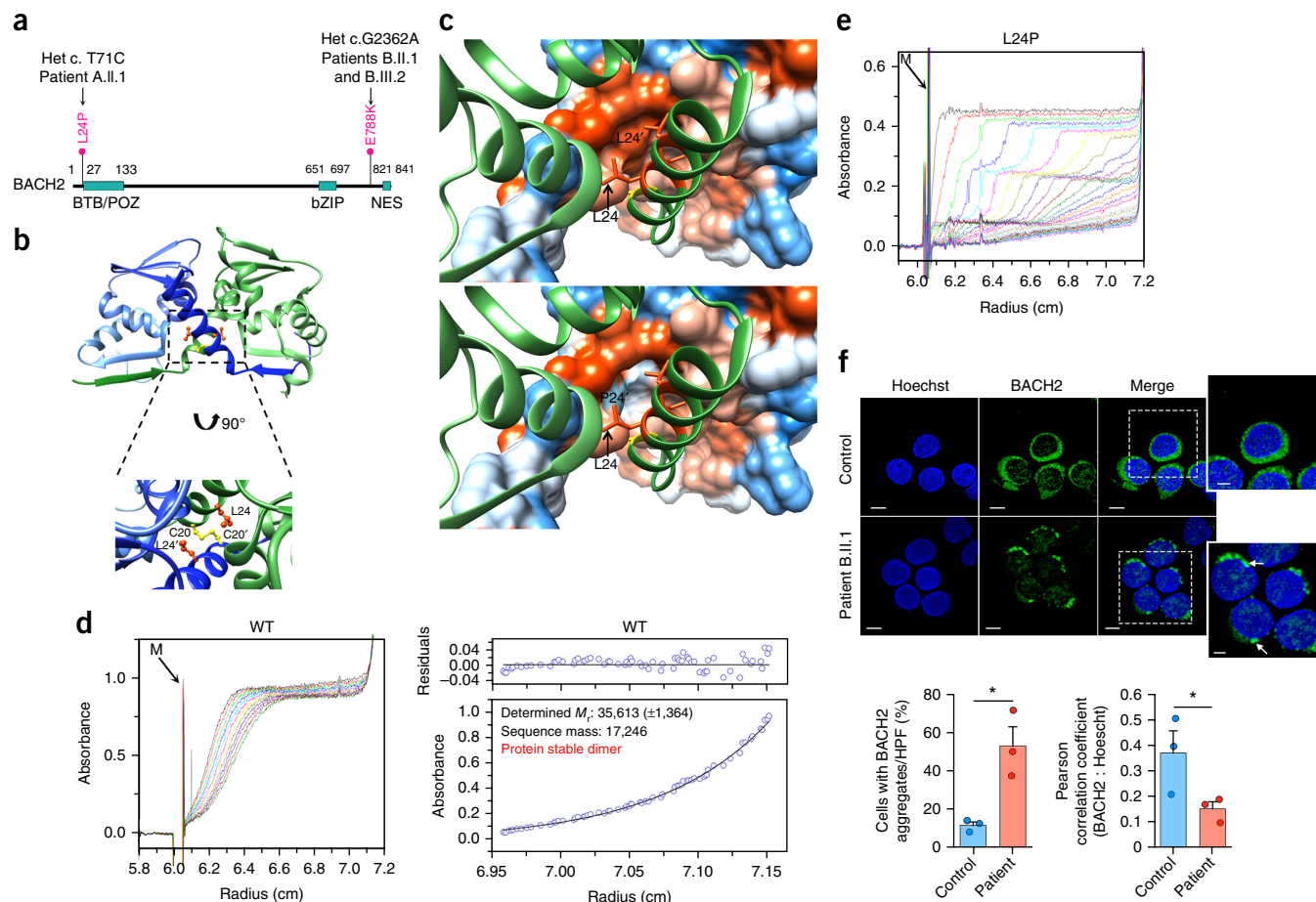
Figure 4 *BACH2* mutations produce unstable proteins. (**a**) Domain schematic of BACH2 protein and point substitutions in different subjects. NES, nuclear export signal. (**b**) Ribbon representations of the BACH2 POZ domain (crystal structure form II; PDB 3OHV) in the wild-type protein (top), with an expanded and rotated interface view (bottom). Yellow, intermolecular disulfide at position 20; orange, leucine residues at position 24. (**c**) Top, wild-type POZ domain dimer interface (PDB 3OHV); bottom, homology model of BACH2^{L24P}–wild-type POZ heterodimer, illustrating local changes. In each, one monomer is rendered as a partially transparent hydrophobicity surface (orange, hydrophobic; white, intermediate; blue, hydrophilic), and the other as a ribbon (green); selected side chains are shown as sticks. Cys20 (yellow) and Ile23, Leu24 and Leu27 (all orange) form a hydrophobic patch on α-helix 1; two of these patches are in close contact at the wild-type dimer interface. Note that the lower diagram is not meant to represent the structure accurately, but is shown merely to indicate regional changes. (**d,e**) Analytical ultracentrifugation of purified wild-type (WT) p.BACH2 (**d**) and mutant p.BACH2^{L24P} (**e**) BTB/POZ domain; the sedimentation direction is left to right. M, sample meniscus. The wild-type protein is dimeric (35 kDa), as determined by sedimentation equilibrium measurements (shown on the right in **d**), and migrates with a single boundary with a sedimentation coefficient (S) of 2.6. The mutant showed several boundaries (S values from 4 to 18), indicating heterogeneous large protein aggregates (**e**). (**f**) Top, representative confocal microscopy images of primary lymphocytes from a healthy control and from patient B.II.1, stained for BACH2 (green) and DNA (Hoechst; blue); arrows indicate cytoplasmic aggregates. Dashed outlines indicate the regions magnified in the respective insets. Scale bars, 5 μm (main image) or 2 μm (inset). Bottom, quantification (mean and s.e.m.; $n = 3$ experiments) of the number of cells containing aggregates per high-power field (HPF) and of BACH2 nuclear localization. *$P < 0.05$, t-test.

the mutant proteins compared with the wild-type protein (**Fig. 4**). Finally, we used retroviral constructs encoding wild-type or mutant mouse BACH2 to transduce *Prdm1*–YFP transgenic CD4$^+$ T cells. Forced expression of wild-type BACH2 alone led to a significant reduction in the expression of *Prdm1*–YFP, but cotransduction with either mutant form of BACH2 did not interfere with repression of the *Prdm1* reporter in primary mouse lymphocytes (**Fig. 5c**). Collectively, these data indicate that neither BACH2 mutation exerted a dominant negative effect.

Bach2$^{+/−}$ mice have impaired lymphocyte development

In the absence of a dominant negative effect, we turned to haploinsufficiency as an explanation. A complete absence of *Bach2* in mice results in B cell immunodeficiency and fatal autoimmunity later in life[15,16]. If haploinsufficiency were responsible for the defects in

lymphocyte development observed in our affected subjects, we would expect to see a similar effect in mice heterozygous for wild-type and null alleles (*Bach2*$^{+/−}$). We found that *Bach2*$^{+/−}$ mice had reduced *Bach2* mRNA expression (**Fig. 6a**) and protein expression (**Fig. 6b**), as well as elevated *Prdm1* mRNA expression (**Supplementary Fig. 7a**). There was no difference in the numbers of CD4$^+$ and CD8$^+$ T cells, B cells or plasma cells between heterozygous and wild-type unchallenged mice (**Supplementary Fig. 7b,c**), but *Bach2*$^{+/−}$ mice did show a small but significant reduction in the number of FoxP3$^+$ cells, as well as significant increases in numbers of CCR9$^+$ and β$_7$ integrin$^+$ cells among CD4$^+$ T cells (**Fig. 6c–e**). We next immunized wild-type and *Bach2*$^{+/−}$ mice with 4-hydroxy-3-nitrophenylacetyl hapten-conjugated chicken γ-globulin (NP-CGG) in alum and analyzed the splenic B cell response. Immunized *Bach2*$^{+/−}$ mice showed minimal induction of both IgG1 class-switched B220^{hi}CD138$^−$ B cells and B220^{lo}CD138$^+$
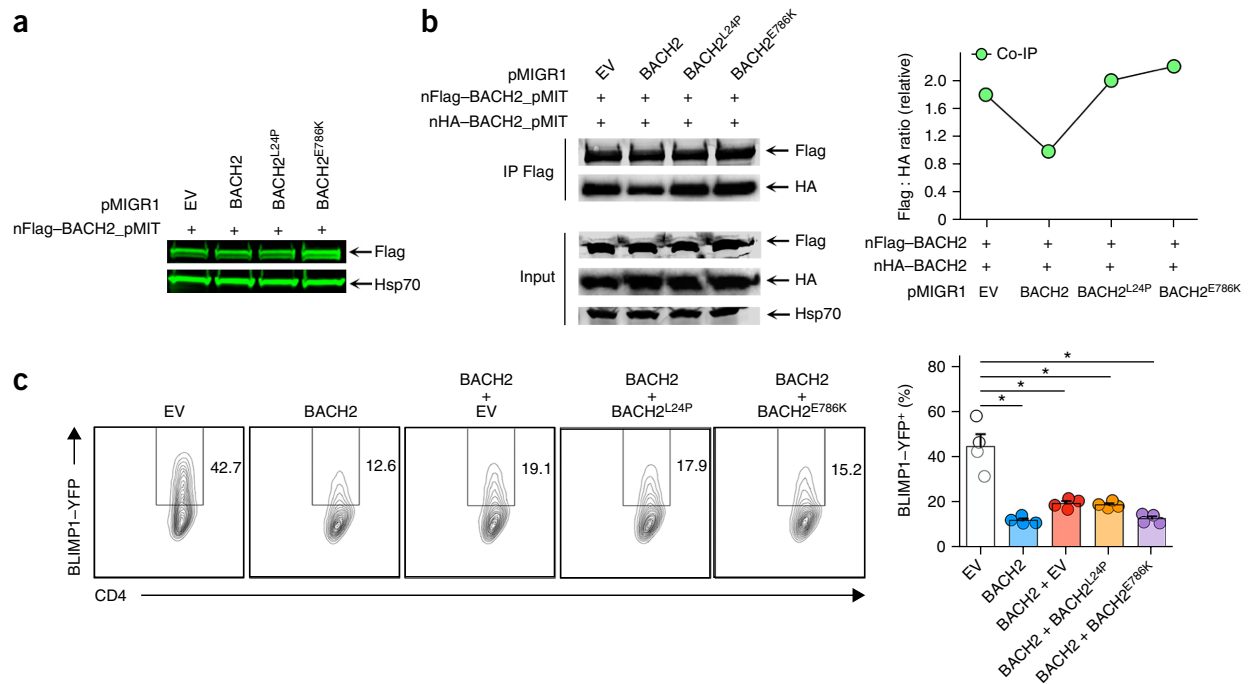
Figure 5 Mutant forms of *Bach2* do not exert dominant negative effects. (**a**) Immunoblot for Flag and Hsp70 in HEK293T cells cotransfected at a 1:1 ratio with Flag-tagged wild-type mouse BACH2 and untagged wild-type and mutant forms of mouse BACH2. The blot shown is representative of three independent experiments. (**b**) Coimmunoprecipitation of Flag- and HA-tagged wild-type BACH2 transfected into HEK293T cells together with untagged wild-type and mutant forms of mouse BACH2 at a 1:1:1 vector ratio. Shown are a representative blot from $n = 3$ independent experiments (left) and quantification of the coimmunoprecipitated (Co-IP) Flag and HA signals (right). (**c**) BLIMP1–YFP signal in BLIMP1–YFP transgenic mouse CD4$^+$ T cells cotransduced at a 1:1 ratio with retrovirus supernatants encoding wild-type and mutant forms of mouse BACH2. Shown are representative flow cytometry plots (left) and cumulative data (mean and s.e.m.) from $n = 4$ independent experiments (right). Numbers adjacent to outlines indicate the percentage of cells in the gate. EV, empty vector. *$P < 0.0001$, ANOVA.

plasma cells compared with that in wild-type mice (**Fig. 6f**). The proportion of germinal-center B220$^+$Ki67$^+$Bcl-6$^+$ B cells was also reduced in *Bach2$^{+/-}$* mice compared with that in wild-type mice (**Fig. 6g**), thus supporting a haploinsufficiency model.

SE-regulated genes are associated with haploinsufficiency

Taken together, our data argue that the maintenance of a threshold concentration of BACH2 is crucial for proper immunoregulation. Mutations in other TF-encoding genes have been reported to cause haploinsufficiency disorders[35]. *BACH2* expression is regulated in a complex manner, and the *BACH2* locus contains an archetypal SE[12–14,16,36,37] (**Fig. 7a**). We therefore hypothesized that SE structure may be enriched among genes associated with haploinsufficiency diseases. To investigate this, we compared genetic disorders mediated by haploinsufficiency (372 genes) or autosomal recessive (AR) inheritance (259 genes) to haplosufficient (HS) gene states (i.e., those for which single-allele deletions are inconsequential; 901 genes)[38] (Online Methods). To validate these three groups, we evaluated the probability of loss-of-function intolerance (pLI) (as estimated by ExAc[39]), for which a score of 0 predicts that the loss of a single copy of the gene is well tolerated, and a score of 1 predicts that the loss of a single copy is poorly tolerated and likely to result in a disease. As expected, the median pLI score for our haploinsufficiency list was considerably higher than for the others (median values of 0.86, 0.0005 and 0.004 for haploinsufficiency, HS and AR genes, respectively) (**Fig. 7b**). Moreover, haploinsufficiency genes were substantially more likely to have SE architecture, as reflected by especially high acetylation of histone H3 on lysine 27 (H3K27ac), a hallmark of active enhancers[40] (**Fig. 7c,d**, **Supplementary Fig. 8a**

and **Supplementary Tables 3** and **4**). In contrast, there were no differences in the frequency of typical enhancers among the three groups (**Fig. 7d**). We next compared the functions of genes among the three groups and found that haploinsufficiency genes were more likely to encode TFs compared with genes associated with AR inheritance or HS genes (**Fig. 7e** and **Supplementary Fig. 8b,c**). To address any potential confounding abundance of TF-encoding genes in SEs, we also divided our list of haploinsufficiency genes into those that code for TFs and those that code for all other proteins, and compared the frequency of SEs (**Supplementary Fig. 8c**). We found that even after TF genes were discounted, haploinsufficiency-disease-causing genes were heavily enriched for SE architecture compared with HS and AR genes (**Supplementary Fig. 8d**). We next asked whether SE-bearing genes have less tolerance to loss-of-function mutations and whether the 'size' of the SE correlates with this effect. We expanded our analysis to a collection of genes regulated by SEs from across more than 100 tissues (dbSuper database[41]) and observed both a striking increase in the pLI score with increasing SE signal size and a concomitant increase in the proportion of TF genes (**Fig. 7f**). Thus, not only the presence of an SE but also its 'size' correlates with the likelihood of disease caused by haploinsufficiency.

SE architecture was previously shown to be associated with human disease loci in GWAS[14,36,42]. This is the case for the *BACH2* gene[16,22], which was consistently in the top 1% of human SE genes as ranked by H3K27ac SE signal intensity in naive CD4$^+$ T cells and naive CD8$^+$ T and B cells (**Supplementary Fig. 8e–g**). On the basis of the SE enrichment among haploinsufficiency genes, we next asked whether there would be general enrichment of GWAS 'hits' in genes associated with haploinsufficiency disease. In agreement with this hypothesis,
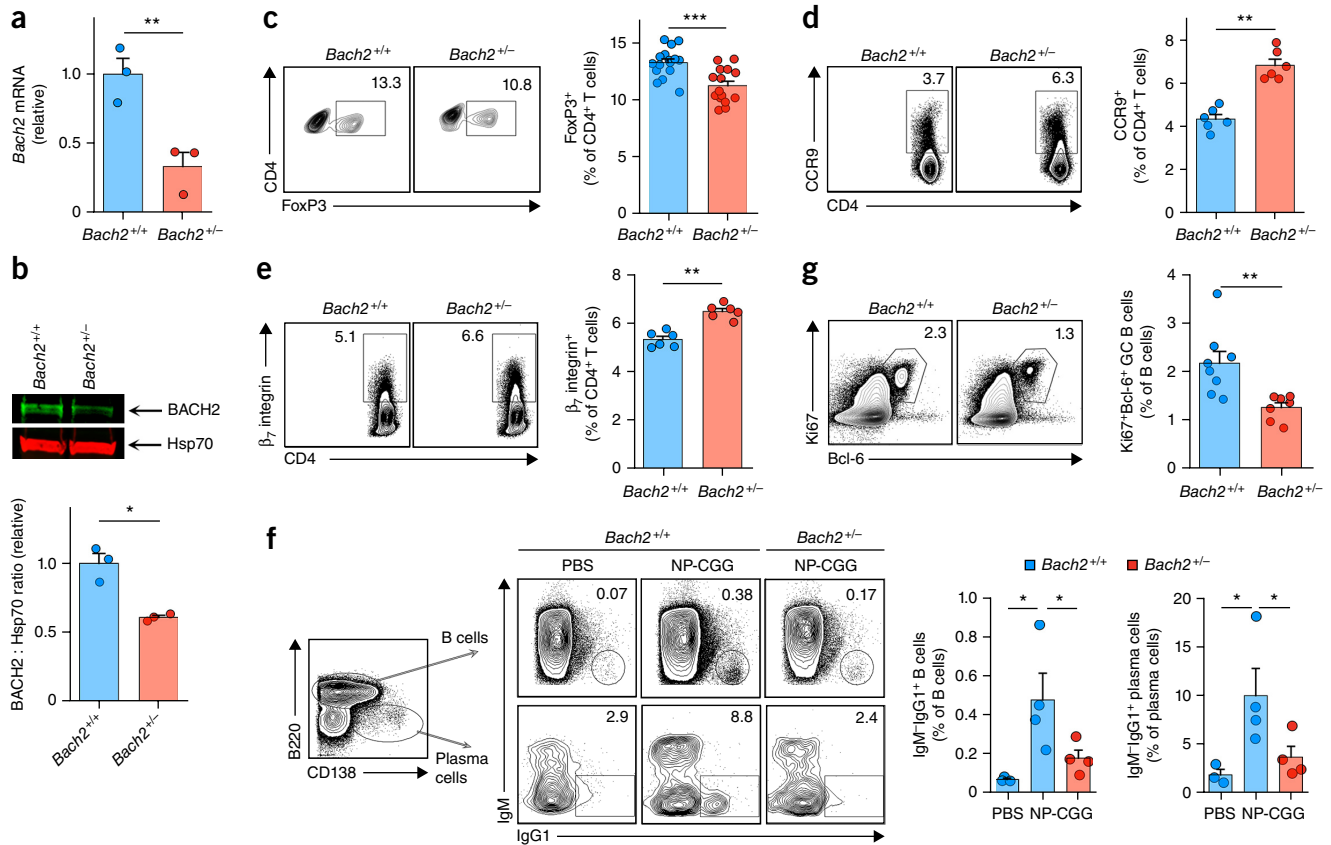
Figure 6 *Bach2*-haploinsufficient mice have abnormal B cell differentiation and T_{reg} cell numbers. (**a**) *Bach2* mRNA expression in B cells of *Bach2*[+/+] and *Bach2*[+/−] mice (*n* = 3 experiments). (**b**) BACH2 protein expression in splenic naive B cells of *Bach2*[+/+] and *Bach2*[+/−] mice. Shown are a representative blot (top) and cumulative quantification (bottom) from *n* = 3 independent experiments. (**c**–**e**) Flow cytometry analysis of CD4[+] splenocytes from *Bach2*[+/+] and *Bach2*[+/−] mice, showing the percentages of FoxP3[+] (**c**), CCR9[+] (**d**) and β_7 integrin[+] (**e**) cells. *n* = 3 (**c**) or 2 (**d**,**e**) experiments. (**f**) IgM and IgG1 expression in B cells and plasma cells among splenocytes of *Bach2*[+/+] and *Bach2*[+/−] mice 8 d after immunization with NP-CGG in alum or PBS treatment. (**g**) B220[+]Ki67[+]Bcl-6[+] germinal center (GC) B cells among splenocytes of *Bach2*[+/+] and *Bach2*[+/−] mice 8 d after immunization with NP-CGG in alum. Panels **c**–**f** show representative flow cytometry plots and cumulative quantification. In flow cytometry plots, numbers in corners indicate the percentage of cells in the gate. Bar charts show mean and s.e.m. throughout. *In vivo* experiments were carried out twice. Data points represent individual mice. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$ by *t*-test (**a**,**b**), one-way ANOVA (**f**) or Mann–Whitney *U*-test (**c**–**e**,**g**).

we found highly significant enrichment of disease-associated single-nucleotide variants in this gene set (**Fig. 7g** and **Supplementary Table 5**). To exclude gene size as a potential confounding factor, we repeated the analysis on subsets of genes of less than 50 kb, and again we found more GWAS associations in genes associated with haploinsufficiency syndromes than in HS genes (**Supplementary Fig. 8h**). Thus, haploinsufficiency genes are enriched for both SEs and GWAS hits.

DISCUSSION

Adaptive immunity is critically dependent on the appropriate differentiation and maturation of lymphocytes. Several complex differentiation steps are required for the formation of mature cells that occupy specific niches and carry out defined roles within the immune system. Key to the regulation of lymphocyte differentiation is precise control over the expression of many TFs that form complex regulatory networks. The identification of both mice and humans with dramatic early-onset stereotypical autoimmune disease associated with a homozygous loss of gene expression has led to the identification of many key regulatory TFs, most notably FoxP3, the master TF of T_{reg} cells[27].

Reductions in the cost and time requirements of whole-exome sequencing have allowed subjects with no family history of disease to be analyzed for genetic mutations. Comparing a person's DNA sequence with those of his or her healthy parents can identify the appearance of *de novo* mutations that would otherwise be missed if a positive family history were required before any investigation could take place. Through this strategy, a number of heterozygous mutations associated with autoimmune diseases have recently been discovered.

BACH2 has a major role in the regulation of the adaptive immune system. Its own expression is tightly regulated by the presence of a large SE region within the *Bach2* locus[14]. The role of BACH2 has been elucidated through the investigation of BACH2-deficient mice that have defects in B cell CSR and T_{reg} cell differentiation. In mice, this combination results in chronic variable immunodeficiency with a late-onset, progressively fatal autoimmune syndrome that includes inflammatory enteropathy and respiratory infiltrates[15]. In humans, in line with *BACH2*'s designation as an SE-associated gene, there is a link between single-nucleotide polymorphisms within the *BACH2* locus and a number of autoimmune and inflammatory diseases.

Herein we describe three subjects from two families that have heterozygous mutations in *BACH2*. Two of the three presented with early-onset autoimmune gastrointestinal disease, and the third presented with symptoms later in life. All three developed a chronic variable immunodeficiency characterized by recurrent respiratory tract infections
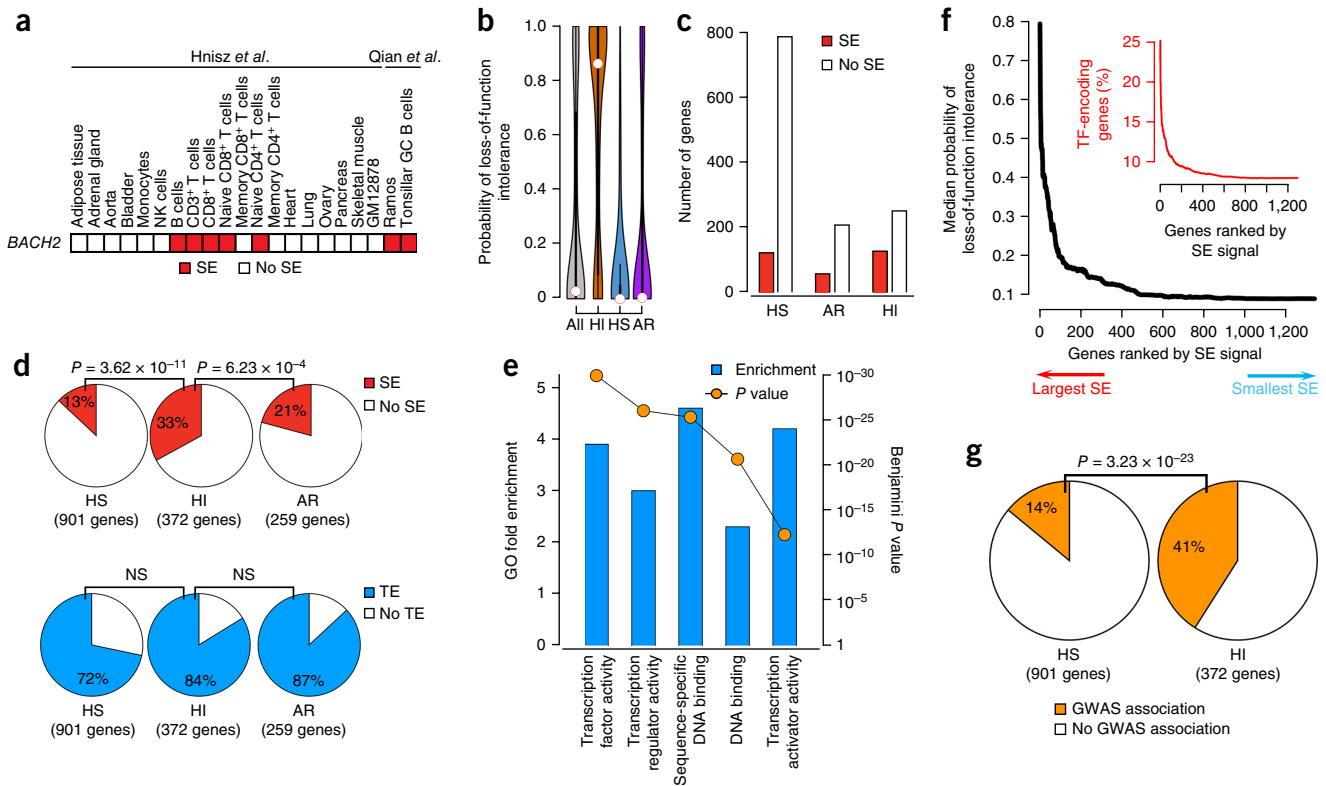
Figure 7 SE-regulated genes are associated with haploinsufficiency. (**a**) The *BACH2* locus has SE structures in multiple human immune cell types demarcated by H3K27ac loading. Red shading denotes the presence of an SE in the *BACH2* locus in the specified tissue. Data are from Hnisz *et al.*[36] or Qian *et al.*[37] as indicated. GC, germinal center. (**b**) pLI scores in haplosufficient (HS), autosomal recessive (AR) and haploinsufficient (HI) gene sets. The white circles indicate median values. Source data were taken from the ExAc database[39]. (**c**) Numbers of HS, AR and HI genes with and without associated SE architecture in humans (see also **Supplementary Fig. 8a** and **Supplementary Table 3**). (**d**) The frequency of SE and typical enhancer (TE) structures in HS, HI and AR genes. (**e**) Gene Ontology (GO) functional annotation enrichment in HI genes. Shown are enrichment scores and Benjamini *P* values for the top five most significantly enriched terms. (**f**) Median pLI (black line) versus SE signal size; the inset shows the percentage of genes that encode TFs (red line) versus the SE signal size. For reference, the red line is asymptotic to the expected level (the mean percentage of genes in the human genome that encode TFs is 7.5%). Source data were taken from the ExAc[39] and dbSuper[41] databases. (**g**) The percentage of HS or HI genes with GWAS disease associations. *P* values in **d** and **g** were determined by Fisher exact test. NS, not significant.

associated with an inability to generate appropriate antibody responses to vaccination. Our findings support a role for human BACH2 as a key regulator of the human adaptive immune system critical to the maintenance of T_{reg} cell function and B cell maturation. BACH2-deficient mice show accelerated T cell senescence[26,43], and, in keeping with this, T cells from our subjects showed defective cell proliferation associated with progressive T cell lymphopenia. Many of the autoimmune phenomena in our subject with the L24P mutation have been successfully treated with corticosteroids, although this has not reduced her chronic variable immunodeficiency or her pneumonitis, which is of some concern, as this is a key cause of early mortality in BACH2-deficient mice. The father with the E788K mutation developed bronchiectasis later in life. It remains to be seen whether the pneumonitis will be progressive in our L24P subject and result in chronic lung damage.

In the first family, the *BACH2*[T71C] gene mutation resulted in a protein that is predicted to be unable to dimerize and is unstable. In the second family, the mutant protein encoded by *BACH2*[E788K] again showed some evidence of a defect in stability, albeit less dramatic, and we saw more evidence of a defect in the localization of the protein, with reduced nuclear localization. We found little evidence that either mutant protein acted in a dominant negative manner. Thus we attribute the clinical phenotype to *BACH2* haploinsufficiency, and this conclusion is consistent with our findings and previous reports[44] that *Bach2*[+/−] heterozygotes have defects in CSR antibody responses.

Mammalian cells contain tens of thousands of gene enhancer sites that cluster in large numbers around a select subset of genes that make up some 5–10% of the total human genome. These clusters are collectively known as SEs. Mutations identified in GWAS tend to associate with these gene loci, but the significance of this remains unclear. Previous work suggests that SE genes encode proteins with functions that are highly dependent on transcription, small changes in which would lead to significant changes in cell development. On this basis, we hypothesize that SE genes would be susceptible to gene dosage effects in patients with heterozygous mutations. We conclude that the relationship between GWAS and SE-regulated genes exists not simply because these genes transcribe proteins that are important per se, but because small changes in the expression of SE genes result in large functional changes in the affected cells.

In summary, we describe a newly recognized disorder, BACH2-related immunodeficiency and autoimmunity (BRIDA), that is caused by heterozygous mutations in *BACH2*. We found that the mechanism of disease is *BACH2* haploinsufficiency, and that *BACH2* is a prototype haploinsufficiency gene with SE architecture. Given the prevalence of heterozygous variants in nonconsanguineous human genomes[45], it is difficult to predict which ones cause disease. We have demonstrated that haploinsufficiency diseases are associated with heterozygous mutations in SE-regulated genes. As SEs allow complex regulation of gene transcription, we conclude that haploinsufficiency

genes are carefully regulated owing to their SE association, and that small changes in their expression level can potentially lead to amplified changes in their associated network, especially for TF-encoding genes, thereby resulting in significant pathology. Thus, SE-regulated genes should be a higher priority in assessments of heterozygous variants discovered by whole-exome/genome sequencing.

METHODS
Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

AUTHOR CONTRIBUTIONS
B.A., J.G. and J.V. designed and performed experiments, analyzed data and wrote the manuscript. C.O'B., I.V., F.P.D., A.K., A.N.H., J. Keith, M.F.S., A.S., R.R., M.A.L., O.K., H.-W.S. and Y.Z. performed experiments and/or analyzed data. I.J.F., W.S., T.J.A., P.K. and N.C. provided patient samples and clinical and scientific input. K.M.-R. coordinated patient samples. Patient sequencing and sequence analysis were carried out by J.V., N.C., T.J.A., D.K., M.M., J.D.H., J.M. and Y.Z. A.V.V., N.R.W., H.H.U. and M.K. provided scientific input. P.T.W., I.W.P. and J. Kaufman provided scientific input, performed protein chemistry experiments and analyzed data. N.P.R. provided murine reagents for these experiments. M.J.L., J.J.O'S., N.C. and A.D.J.L. provided scientific input, supervised the project and wrote the manuscript.

1. Bousfiha, A. *et al.* The 2015 IUIS Phenotypic Classification for Primary Immunodeficiencies. *J. Clin. Immunol.* **35**, 727–738 (2015).
2. Picard, C. *et al.* Primary immunodeficiency diseases: an update on the classification from the International Union of Immunological Societies Expert Committee for Primary Immunodeficiency 2015. *J. Clin. Immunol.* **35**, 696–726 (2015).
3. Arason, G.J., Jorgensen, G.H. & Ludviksson, B.R. Primary immunodeficiency and autoimmunity: lessons from human diseases. *Scand. J. Immunol.* **71**, 317–328 (2010).
4. Notarangelo, L.D. Primary immunodeficiencies. *J. Allergy Clin. Immunol.* **125**, S182–S194 (2010).
5. Conley, M.E. & Casanova, J.-L. Discovery of single-gene inborn errors of immunity by next generation sequencing. *Curr. Opin. Immunol.* **30**, 17–23 (2014).
6. Deau, M.-C. *et al.* A human immunodeficiency caused by mutations in the *PIK3R1* gene. *J. Clin. Invest.* **125**, 1764–1765 (2015).
7. Lo, B. *et al.* Patients with LRBA deficiency show CTLA4 loss and immune dysregulation responsive to abatacept therapy. *Science* **349**, 436–440 (2015).
8. Cunningham-Rundles, C. The many faces of common variable immunodeficiency. *Hematology (Am. Soc. Hematol. Educ. Program)* **2012**, 301–305 (2012).
9. Rieux-Laucat, F. & Casanova, J.-L. Autoimmunity by haploinsufficiency. *Science* **345**, 1560–1561 (2014).
10. Lo, B. *et al.* CHAI and LATAIE: new genetic diseases of CTLA-4 checkpoint insufficiency. *Blood* **128**, 1037–1042 (2016).
11. Vahedi, G. *et al.* STATs shape the active enhancer landscape of T cell populations. *Cell* **151**, 981–993 (2012).
12. Whyte, W.A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).
13. Lovén, J. *et al.* Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153**, 320–334 (2013).
14. Vahedi, G. *et al.* Super-enhancers delineate disease-associated regulatory nodes in T cells. *Nature* **520**, 558–562 (2015).
15. Roychoudhuri, R. *et al.* BACH2 represses effector programs to stabilize T$_{reg}$-mediated immune homeostasis. *Nature* **498**, 506–510 (2013).
16. Igarashi, K., Ochiai, K., Itoh-Nakadai, A. & Muto, A. Orchestration of plasma cell differentiation by Bach2 and its gene regulatory network. *Immunol. Rev.* **261**, 116–125 (2014).
17. Ferreira, M.A.R. *et al.* Identification of *IL6R* and chromosome 11q13.5 as risk loci for asthma. *Lancet* **378**, 1006–1014 (2011).
18. Cooper, J.D. *et al.* Meta-analysis of genome-wide association study data identifies additional type 1 diabetes risk loci. *Nat. Genet.* **40**, 1399–1401 (2008).
19. Franke, A. *et al.* Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat. Genet.* **42**, 1118–1125 (2010).
20. Dubois, P.C.A. *et al.* Multiple common variants for celiac disease influencing immune gene expression. *Nat. Genet.* **42**, 295–302 (2010).
21. Jin, Y. *et al.* Genome-wide association analyses identify 13 new susceptibility loci for generalized vitiligo. *Nat. Genet.* **44**, 676–680 (2012).
22. International Multiple Sclerosis Genetics Consortium. *et al.* Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* **476**, 214–219 (2011).
23. Nakayama, Y. *et al.* A limited number of genes are involved in the differentiation of germinal center B cells. *J. Cell. Biochem.* **99**, 1308–1325 (2006).
24. Ochiai, K. *et al.* Plasmacytic transcription factor Blimp-1 is repressed by Bach2 in B cells. *J. Biol. Chem.* **281**, 38226–38234 (2006).
25. Muto, A. *et al.* The transcriptional programme of antibody class switching involves the repressor Bach2. *Nature* **429**, 566–571 (2004).
26. Kuwahara, M. *et al.* The Menin-Bach2 axis is critical for regulating CD4 T-cell senescence and cytokine homeostasis. *Nat. Commun.* **5**, 3555 (2014).
27. Povoleri, G.A.M. *et al.* Thymic versus induced regulatory T cells—who regulates the regulators? *Front. Immunol.* **4**, 169 (2013).
28. Rosbrook, G.O., Stead, M.A., Carr, S.B. & Wright, S.C. The structure of the Bach2 POZ-domain dimer reveals an intersubunit disulfide bond. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 26–34 (2012).
29. Uhlig, H.H. *et al.* The diagnostic approach to monogenic very early onset inflammatory bowel disease. *Gastroenterology* **147**, 990–1007.e3 (2014).
30. Deane, S., Selmi, C., Naguwa, S.M., Teuber, S.S. & Gershwin, M.E. Common variable immunodeficiency: etiological and treatment issues. *Int. Arch. Allergy Immunol.* **150**, 311–324 (2009).
31. Salzer, U. & Grimbacher, B. Monogenetic defects in common variable immunodeficiency: what can we learn about terminal B cell differentiation? *Curr. Opin. Rheumatol.* **18**, 377–382 (2006).
32. Iwata, M. *et al.* Retinoic acid imprints gut-homing specificity on T cells. *Immunity* **21**, 527–538 (2004).
33. Cassani, B. *et al.* Gut-tropic T cells that express integrin $\alpha_4\beta_7$ and CCR9 are required for induction of oral immune tolerance in mice. *Gastroenterology* **141**, 2109–2118 (2011).
34. Igarashi, K., Ochiai, K. & Muto, A. Architecture and dynamics of the transcription factor network that regulates B-to-plasma cell differentiation. *J. Biochem.* **141**, 783–789 (2007).
35. Seidman, J.G. & Seidman, C. Transcription factor haploinsufficiency: when half a loaf is not enough. *J. Clin. Invest.* **109**, 451–455 (2002).
36. Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013).
37. Qian, J. *et al.* B cell super-enhancers and regulatory clusters recruit AID tumorigenic activity. *Cell* **159**, 1524–1537 (2014).

38. Huang, N., Lee, I., Marcotte, E.M. & Hurles, M.E. Characterising and predicting haploinsufficiency in the human genome. *PLoS Genet.* **6**, e1001154 (2010).
39. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
40. Creyghton, M.P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. USA* **107**, 21931–21936 (2010).
41. Khan, A. & Zhang, X. dbSUPER: a database of super-enhancers in mouse and human genome. *Nucleic Acids Res.* **44**, D164–D171 (2016).
42. Parker, S.C.J. *et al.* Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc. Natl. Acad. Sci. USA* **110**, 17921–17926 (2013).
43. Roychoudhuri, R. *et al.* BACH2 regulates CD8[+] T cell differentiation by controlling access of AP-1 factors to enhancers. *Nat. Immunol.* **17**, 851–860 (2016).
44. Shinnakasu, R. *et al.* Regulated selection of germinal-center cells into the memory B cell compartment. *Nat. Immunol.* **17**, 861–869 (2016).
45. 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).

ONLINE METHODS

Ethics approval. Subjects and their participating relatives provided written informed consent and were investigated under National Institute of Allergy and Infectious Diseases (NIAID) Institutional Review Board–approved research protocols 89-I-0158 and 06-I-0015, with approval from the West London Research Ethics Committee (Ethics Protocol Reference Number 11/LO/0883) and the Oxford IBD cohort study (monogenic IBD subproject). All animal studies were carried out according to National Institutes of Health guidelines for the use and care of live mice and were approved by the Institutional Animal Care and Use Committee of the National Institute of Arthritis, Musculoskeletal and Skin Diseases (protocol number A014-03-02).

Histology and immunohistochemistry. A colonic biopsy was obtained from patient A.II.1 at the time of her clinical presentation, when she was 19 years old. The biopsy was stained with hematoxylin and eosin and was reviewed by pathologists at the Hammersmith Hospital (London, UK). Immunohistochemical staining was carried out on formalin-fixed paraffin-embedded (FFPE) sections from affected subjects and on tissue-matched FFPE sections from healthy control donors and from age-matched donors diagnosed with classical Crohn's disease (provided by the Oxford Centre of Histopathology Research and the Oxford Gastrointestinal Illness Biobank). For this we used antibodies to FoxP3 (Abcam; 236A/E7; validation information is available at http://www.abcam.com/foxp3-antibody-236ae7-ab20034-references.html) followed by tyramide signal amplification (PerkinElmer), and anti-CD3 (Dako; F7.2.38; validation information is available at http://www.agilent.com/en/products/immunohistochemistry/antibodies-controls/primary-antibodies/cd3-f7-2-38) followed by Alexa Fluor 488–conjugated goat anti-mouse IgG (Life Technologies). Nuclei were stained with Vectashield antifade mounting medium with DAPI (Vector Laboratories), and slides were examined with a Zeiss LSM510 inverted confocal microscope. ImageJ (ImageJ) and Photoshop (Adobe) were used for processing and presentation of the images.

Antibodies, cell lines and media. The following antibodies and reagents were used in the study: anti-human BACH2 (ab83364; Abcam); anti-human CD19 (HIB19), anti-human CD24 (ML5), anti-mouse CD3 (145-2C11), anti-mouse CD8 (53-6.7), anti-mouse CCR9 (9B1), anti-human CCR9 (LO53E8) and anti-human/mouse β_7 integrin (FIB504) (all BioLegend); anti-human CD4 (OKT4), anti-human CD25 (2A3), anti-human CD27 (M-T271), anti-human CD38 (HB-7), anti-human IgG (GI8-145), human BD Fc Block, anti-mouse CD4 (RM4-5), anti-mouse CD25 (7D4), anti-mouse CD44 (IM7), anti-mouse CD62L (MEL-14), anti-mouse CD138 (281-2), anti-mouse B220 (RA3-6B2), anti-mouse CXCR5 (2G8), anti-mouse IgG1 (A85-1), anti-mouse IgM (R6-60.2), streptavidin–allophycocyanin and streptavidin–fluorescein isothiocyanate (all from BD Biosciences); anti-human CD3 (OKT3), anti-human CD8 (RPA-T8), anti-human CD38 (HB7), anti-human CD127 (eBioRDR5), anti-human T-bet (eBio4B10), anti-human FoxP3 (PCH101), anti-mouse CD25 (BC61.5), anti-mouse CD127 (A7R34), anti-mouse GL7 (GL-7), anti-mouse Fas (15A7), anti-mouse NKp46 (29AI.4), anti-mouse IgD (11-26), anti-mouse IgM (11/41), anti-mouse PD1 (J43), anti-mouse GITR (DTA-1), anti-mouse FoxP3 (FJK-16s) and anti-Thy1.1 (HIS51) (all eBioscience); mouse anti-Flag (clone M2; Sigma); and goat anti-rabbit IgG–Alexa Fluor 488 (A-11034; Life Technologies). Live-Dead Flixable Aqua Dead Cell stain was from Thermo Fisher (Boston, MA, USA). Raji, Ramos and HEK293T cell lines were from ATCC. Unless specified, human cells and cell lines were maintained in RPMI 1640 media supplemented with 2 mM L-glutamine (Life Technologies), penicillin and streptomycin (100 IU/mL and 100 µg/mL, respectively; Life Technologies), and 10% FBS (Atlanta Biologicals). Mouse cells were cultured in identical medium supplemented with 2 mM β-mercaptoethanol (Sigma Aldrich). HEK293T cells were maintained in DMEM (Life Technologies) supplemented as done for the human cell culture medium. Cell lines were not tested for mycoplasma contamination.

Mice. C57BL/6J mice were from The Jackson Laboratory. *Bach2*$^{-/-}$ and *Bach2*$^{+/-}$ mice were generated and housed as previously described[15]. BLIMP1–YFP BAC transgenic mice have been previously described[46]. No statistical methods were used to predetermine sample size.

Cell isolation and culture. We isolated human peripheral blood mononuclear cells (PBMCs) from the blood of affected subjects and healthy donors by density-gradient centrifugation using Ficoll (GE Healthcare), and we lysed red blood cells with RBC lysis buffer (eBioscience). CD4$^+$ T cells, naive CD4$^+$ T cells and naive B cells were purified from PBMCs by negative selection with the human CD4$^+$ T cell isolation kit, human naive CD4$^+$ T cell isolation kit II and human naive B cell isolation kit II, respectively (all Miltenyi Biotec) according to the manufacturer's instructions. B cell subsets were sort-purified by FACSAria (BD Immunocytometry Systems, San Jose, CA, USA) with allophycocyanin–conjugated anti-CD19 (BioLegend, San Diego, CA, USA), phycoerythrin (PE)-conjugated anti-CD27 (BD Biosciences, San Jose, CA, USA), and PerCP–Cy5.5-conjugated anti-IgM (BD Biosciences). Naive B cells were defined as CD19$^+$CD27$^-$IgM$^+$ B cells with >98% purity[47].

CD4$^+$ T cells from the spleens and lymph nodes of 6–8-week-old mice were purified by negative selection and magnetic separation (Miltenyi), and then the naive CD4$^+$CD25$^-$CD62L$^+$CD44$^-$ population was sorted with a FACSAria II. Naive BLIMP1–YFP CD4$^+$ T cells were activated for 3 d by plate-bound anti-CD3 (2C11; BioXCell) and anti-CD28 (37.51; BioXCell), each at a concentration of 10 µg/ml in medium. Cells were stimulated in the presence of mouse IL-12 (20 ng/ml) and anti-mouse IL-4 (10 µg/ml) (T$_H$1 cell conditions) (both from R&D Systems) for 3 d, then split into fresh uncoated plates and supplemented with fresh medium and 100 IU/mL human IL-2 (NIH/NCI BRB Preclinical Repository).

B cell cultures and induction of class-switch recombination. Purified naive B cells were cultured in RPMI 1640 media containing L-glutamine (Sigma Aldrich, St. Louis, MO, USA), 10% FBS (Sigma Aldrich), 10 mM HEPES, pH 7.4 (Sigma-Aldrich), 0.1 mM nonessential amino acid solution (Sigma-Aldrich), 1 mM sodium pyruvate and 40 µg/ml apo-transferrin (Sigma-Aldrich), and supplemented with 60 µg/ml penicillin and 100 µg/ml streptomycin. To induce CSR, we added recombinant human CD40L (1 µg/ml; R&D Systems, Minneapolis, MN, USA), Fab fragment anti-human IgM (Jackson Immunoresearch, West Grove, PA, USA), IL-2 (100 IU/ml; PeproTech, Rocky Hill, NJ, USA) and IL-21 (50 ng/ml; PeproTech) at the beginning of the culture. Cells were cultured in 96-well round-bottom plates (Nunc, Roskilde, Denmark) for 5 d. Culture supernatants were collected for ELISA at the end of the culture period.

IgG and IgA ELISA. IgG and IgA secretion was assessed with the Ready-set-go total IgG and IgA kits (Thermo Fisher) according to the manufacturer's protocols. Absorbance was read at 450 nm within 30 min of the end of the reaction. The sensitivities and linear ranges were obtained with the standard immunoglobulin provided with the kit.

Whole-exome sequencing. DNA was extracted from EDTA-treated blood with the Maxwell 16 blood DNA purification kit (Promega) or from PBMCs with the DNeasy Blood & Tissue kit (Qiagen). A total of 3 µg of DNA was sheared with an E220 focused sonicator (Covaris), and exome libraries were generated with the SureSelect Human All Exon kits (Agilent) according to the manufacturer's protocol. The quality of generated libraries was inspected with the Agilent High Sensitivity DNA kit (Agilent) and quantified with a qPCR kit (Agilent). Samples were sequenced on an Illumina HiSeq 2000 (Illumina) to generate 100-bp paired-end reads. Sequences were aligned to human reference genome GRCh37 with bwa v.0.6.1 with the default parameters[48]. Variant calling (single-nucleotide variants and insertions/deletions) was done with the GATK v.2 toolkit[49], and variants were annotated with Annovar[50]. An in-house custom analysis pipeline was used to filter and prioritize variants on the basis of the likely genetic models and clinical pedigree for patients.

Sanger sequencing. DNA samples were extracted from blood or saliva with the Maxwell 16 blood DNA purification kit (Promega) or Oragene DNA (OG500) (Oragene), respectively. The candidate mutations in affected and unaffected individuals from both families in the study were validated with the BigDye Terminator sequencing kit (Life Technologies) and sequenced on an ABI3730xl genetic analyzer (Applied Biosystems). PCR primer sequences are available from the corresponding author(s) on request.

Flow cytometry. All flow cytometry was carried out in a final staining volume of 100–200 µL, with data acquisition on an LSR II, LSRFortessa or FACSVerse

(all BD Biosciences) within 24 h. Appropriate internal controls, isotype controls and FMO (fluorescence minus one) controls were used to assign gates. Rat anti-mouse CD16/CD32 (clone 2.4G2; BD Biosciences) was used for Fc blockade in mouse flow cytometry experiments. FACS data were analyzed with FlowJo (Tree Star Inc.). For intracellular staining, a BD Cytofix/Cytoperm plus Fixation/Permeabilization Solution Kit was used according to the manufacturer's instructions. For cytokine staining, 4 h of re-stimulation with the phorbol ester PMA (50 ng/mL) and ionomycin (1 mM) (both Sigma) in the presence of brefeldin A (GolgiPlug (BD Biosciences)) was carried out before fixation and permeabilization. FoxP3 staining was carried out with the kit from eBiosciences as per the manufacturer's instructions. We calculated relative FoxP3 and BACH2 levels by dividing the geometric mean fluorescence intensity of patient cells by that of the matched healthy control cells in each run. For assessment of cell proliferation by flow cytometry, T cells were stained with CellTrace Violet as per the manufacturer's instructions and then cultured in the presence of anti-CD3 and anti-CD28 (1 µg/mL of each; clones HIT3α and CD28.2, respectively, both from Biolegend) for 5 d before live/dead staining and data acquisition.

***In vivo* class-switch assay.** We injected 8–10-week-old *Bach2*[+/−] heterozygous and *Bach2*[+/+] wild-type mice intraperitoneally with 50 µg of NP-CGG (Biosearch Technologies) 1:1 in alum (Thermo Scientific) (vol/vol). We collected spleens after 8 d and made single-cell suspensions by passing the cells through a 40-µm strainer. We then carried out surface staining and flow cytometry as described above.

Quantitative RT-PCR. Total RNA was extracted with TRIzol reagent (Invitrogen) and treated with DNase I (Qiagen). RNA was reverse-transcribed to cDNA with the iScript cDNA synthesis kit (Bio-Rad) according to the manufacturer's instructions. Quantitative real-time PCR (qRT-PCR) experiments were carried out in triplicate with Taqman Universal PCR Master Mix (Applied Biosystems) in total reaction volumes of 20 µL and thermocycled in a CFX284 Touch real-time PCR detection system (Bio-Rad). The following Taqman gene-specific primer probes were purchased from Applied Biosystems: human *BACH2* (Hs00222364_m1), *PRDM1* (Hs00153357_m1), *ACTB* (Hs99999903_m1) and 18S ribosomal RNA (Hs99999901_s1); and mouse *Bach2* (Mm00464379_m1), *Prdm1* (Mm00476128_m1), *Bcl6* (Mm00477633_m1) and *Actb* (Mm00607939_s1). Cycle threshold (C_t) values were exported and normalized against the control probe via the $2^{-\Delta Ct}$ method and are reported as expression relative to a control condition.

Silencing of *BACH2* and BACH2 overexpression. 5×10^6 PBMCs per sample were nucleofected with 300 nM Dicer-substrate short interfering RNA (DsiRNA) negative control or predesigned *BACH2* DsiRNA (both TriFECTa; Integrated DNA Technologies) with the Amaxa human T cell nucleofector kit (Program-U014; Lonza) according to the manufacturer's instructions. 24 h after nucleofection, cells were labeled with the CellTrace Violet cell-proliferation kit (Thermo) and rested for 6 h in culture before activation of 1×10^5 cells per 96-well plate with plate-bound anti-CD3 (1 µg/ml; clone HIT3α) and anti-CD28 (1 µg/ml; clone CD28.2) (both from BioLegend). Cells were surface-stained and proliferation was analyzed by flow cytometry after 5 d.

Naive B cells or CD4[+] T cells were nucleofected with 2 µM MISSION universal negative control siRNA (Sigma) or *BACH2* siRNA (Hs01_00214431; Sigma) with the P3 primary cell 96-well Nucleofector kit (Lonza) according to the manufacturer's instructions. Cells were cultured for 24 h at 37 °C in the presence of 100 ng/ml human IL-7 before activation for CSR as described above.

5×10^6 blasting human CD4[+] T cells were mixed with 2–5 µg of either *BACH2* or eGFP mRNA (TriLink) in 50 µl of HyClone MaxCyte buffer and electroporated in an OC-100 PA electroporation chamber with a MaxCyte GT instrument (program T-02). After electroporation, cells were incubated for 20 min at 37 °C in electroporation buffer in 96-well plates and then transferred to 12-well plates in complete RPMI media containing 100 IU/ml human IL-2. *PRDM1* expression was analyzed after 24–48 h by qPCR.

Plasmid DNA and point mutagenesis. Wild-type *Bach2* cDNA expression vectors pMSCV-IRES-GFP (pMIGR1-*Bach2*) and pMSCV-IRES-Thy1.1 DEST (pMIT-*Bach2*) have been described previously[15]. We carried out gene

synthesis to achieve an N-terminal fusion of Flag and HA sequences preceded by a methionine translation-initiation codon (MDYKDDDDK and MYPYDVPDYA, respectively) to the wild-type BACH2 open reading frame. Synthesized DNA was subcloned into pMIT to generate pMIT-Flag-BACH2 and pMIT-HA-BACH2. Point mutagenesis to introduce the *Bach2*[T71C] (Bach2[L24P]) and *Bach2*[G2356A] (Bach2[E786K]) mutations was carried out with the Agilent QuickChange II XL site-directed mutagenesis kit (Agilent Technologies) according to the manufacturer's instructions, with the following primer pairs: *Bach2*[T71C]: forward, 5′-CATTGAGGCCCAGGGGGATGTTGGCACAG-3′ and reverse, 5′-CTGTGCCAACATCCCCCTGGGCCTCAATG-3′; *Bach2*[G2356A]: forward, 5′-AGAGGTACAATTCTTAGAGGTGTTGCTGGGCACC-3′ and reverse, 5′-GGTGCCCAGCAACACCTCTAAGAATTGTACCTCT-3′.

Transfection and production of retrovirus. Transfection was carried out in antibiotic-free medium with Lipofectamine LTX and Plus reagent (Invitrogen). Medium was replaced 7 h later. For retrovirus production, payload retroviral plasmid was cotransfected with pCL-Eco helper virus plasmid as described[51]. Transfected cells were harvested and viral supernatant was collected 48 h after transfection.

Retrovirus transduction. *Prdm1*–YFP BAC transgenic CD4[+] T cells were activated for 24 h with plate-bound anti-mouse CD3 (145-2C11; Biolegend) and anti-mouse CD28 (37.51; Biolegend). Activated cells were transduced with supernatants that contained retrovirus encoding Thy1.1 alone (EV) or together with mouse *Bach2* or mutant mouse *Bach2* conforming to the L24P or E786K mutation, in the presence of polybrene (4 µg/ml), by centrifugation at 2,200 r.p.m. for 50 min at 22 °C. Medium was replaced afterward with fresh culture medium, and cells were harvested 48 h after transduction.

Western blotting and Flag immunoprecipitation (IP). Clarified protein extracts were prepared by lysis of cell pellets in Pierce IP lysis buffer (Thermo Scientific) containing 1× cOmplete Protease Inhibitor cocktail (Roche). Protein concentrations were quantified (Micro BCA protein assay kit (Thermo Scientific)) to ensure equal loading. Proteins were resolved by SDS–PAGE on Any kD Criterion TGX gels (Bio-Rad) and electrotransferred onto nitrocellulose membranes (Bio-Rad). Immunoblotting was done with rabbit anti-BACH2 (Abcam), mouse anti-Flag M2 (Sigma), mouse anti-Hsp70 (B-6; Santa Cruz Biotechnology) and goat anti-mouse IRDye 800CW (926-32210; Li-Cor), and blots were scanned on an Odyssey imaging system (Li-Cor Biotechnology). Alternatively, anti-HA–HRP was used for development with SuperSignal West Pico chemiluminescent substrate (Thermo Scientific) and a ChemiDoc MP Imaging system (Bio-Rad). Flag IP was carried out with EZview Red Anti-Flag M2 affinity gel (Sigma) according to the manufacturer's instructions, after which proteins were eluted with 3× Flag peptide (Sigma).

Confocal microscopy. HEK293T cells (ATCC) were cultured and transfected on poly-L-lysine (Sigma)–coated round cover slips. Primary PBMCs were spun onto poly-L-lysine-coated cover slides with a Cytospin3 centrifuge (Shandon). Cells were fixed with 4% paraformaldehyde, permeabilized with 0.1% Triton X-100 in TBS, blocked with TBS containing 5% horse serum and 0.01% NaN₃, and stained with primary antibodies for 1–2 h at room temperature. Staining with secondary antibodies was carried out for 40 min at room temperature in the dark with 1:10,000 Hoechst. Cells were mounted with ProLong Diamond antifade mountant (Life Technologies). The following antibodies and dilutions were used for confocal microscopy: 1:100 mouse anti-Flag M2 (Sigma), 1:25 rabbit anti-human BACH2 (Abcam), 1:500 goat anti-mouse IgG–Alexa Fluor 488 (A-21121; Life Technologies), and 1:500 goat anti-rabbit IgG–Alexa Fluor 568 (A-11036; Life Technologies). Confocal microscopy of immunostained cells was done with a Leica SP8 inverted five-channel confocal microscope equipped with a motorized stage and ultra-sensitive hybrid detectors (Leica Microsystems). The following laser lines were used: diode for 405 nm, argon for 488 nm, and diode-pumped solid-state for 561 nm excitation wavelengths. The microscope configuration was set up for 3D (x, y, z) sequential scanning with a 63× objective, and z stacks of 0.3-µm optical slices (total of 10–15 µm) were collected. For statistical analysis of BACH2 localization, tiled images of transfected cell layers at a total cell number of 200 cells per field were collected. Images were processed with Imaris (Bitplane, Switzerland)

and Huygens (Scientific Volume Imaging, the Netherlands) software. The number of cells containing protein aggregates was determined from at least three tiled images. Pearson's correlation coefficients were calculated with Imaris.

Recombinant protein expression and purification of BACH2 and variants. Synthetic genes with codons optimized for *Escherichia coli* expression were from Genscript. BL21(DE3) cells with pET 28 vectors were grown in a fermenter, and cells were broken and initially processed as described[52]. Full-length human p.BACH2[1–841] and its p.L24P-variant, and mouse p.Bach2[1–133] and p.Bach2[1–133] L24P all contained an N-terminal His tag to facilitate purification (the sequence difference between human p.BACH2[1–133] and mouse p.Bach2[1–133] is at one position, amino acid 8, which is Asp in humans and Ala in mice). Human wild-type p.BACH2[1–841] was extracted from cell lysate with 100 mM sodium bicarbonate, pH 9.5, containing 2 M urea, and the L24P variant was extracted with 8 M guanidine–HCl. Wild-type proteins were expressed as soluble protein, but L24P variants were insoluble and were extracted with 8 M guanidine–HCl. Proteins were purified by a combination of Ni-chelate and size-exclusion chromatography with Ni-chelate Sepharose and Sephadex S200 (both from GE Healthcare). The L24P variants were folded by dialysis against 4 M urea and then stepped through lower concentrations until the urea was removed. DTT was present in all buffers to keep proteins reduced.

Analytical ultracentrifugation. We used a Beckman Optima XL-I analytical ultracentrifuge, absorption optics, an An-60 Ti rotor and standard double-sector centerpiece cells. Equilibrium measurements were taken at 20 °C, and concentration profiles were recorded after 16 h at 20,000 r.p.m. (BACH2[133]) or 10,000 r.p.m. (BACH2[841]). We established baselines by over-speeding at 45,000 r.p.m. for 3 h. Data (the average of eight scans collected with a radial step size of 0.001 cm) were analyzed with the standard Optima XL-I data-analysis software. Sedimentation velocity experiments were performed at 40,000 r.p.m., with scans recorded every 6 min for 3 h. Protein partial specific volumes, calculated from the amino acid compositions, and solvent densities were estimated with the program SEDNTERP (http://rasmb.org/sednterp/).

Protein concentrations. Estimated from amino compositions: absorbencies at 280 nm of 1 mg/ml of mBach2[133] and hBACH2[841] of 0.69 and 0.41, respectively, were used.

Analysis of mutations. Conservation scores for mutated sites (PhyloP, PhastCons and GERP) were obtained from the UCSC genome browser (GRCh37/hg19). Polyphen2, SIFT, LRT, MutationAssessor Functional Impact, MutationTaster and CADD scaled scores were derived with dbNSFP as described[53,54]. The CADD-based mutation significance cutoff at the 99% confidence interval was calculated as described[55].

Curation of haploinsufficient and autosomal recessive disease genes and of haplosufficient genes. We retrieved haploinsufficient genes from PubMed and Online Mendelian Inheritance in Man (OMIM), using the semi-automated method of Dang *et al.*[56]. Searches were restricted to the period from 12 November 2007 to 25 October 2015, and search results were merged with the data set that existed before 12 November 2007 (ref. 56). All retrieved items were manually curated by two independent physicians to ensure that only true positives (i.e., genes that cause haploinsufficient disease) were retained for further analysis. We identified autosomal recessive genes by downloading the OMIM database and extracting all entries inherited in an autosomal recessive fashion. We obtained haplosufficient genes from a list of high-confidence predictions ($P_{HI} < 0.05$) reported by Huang *et al.*[38]. We screened these predictions further by removing those that matched haploinsufficient genes (three genes in total). We carried out functional annotation analysis for genes by Gene Ontology enrichment analysis via DAVID[57,58] and Ingenuity Pathway Analysis (Qiagen).

Super-enhancer structures. Sequencing data were downloaded from GEO. URLs for data used in this article are listed in **Supplementary Table 4**. Reads were mapped to hg19 with bowtie0.12.8 (ref. 59). The HOMER suite of programs[60] was used to call SEs and typical enhancers according to the

guidelines presented by Whyte *et al.*[12]. Enhancers were assigned to the closest genes with PAPST[61]. K27ac signal graphs were created from data generated with HOMER.

We obtained estimated probabilities of human gene intolerance to loss-of-function mutations from the EXAC database[39] (*n* = 18,225 genes; release 0.3.1: ftp://ftp.broadinstitute.org/pub/ExAC_release/release0.3.1/functional_gene_constraint/fordist_cleaned_exac_r03_march16_z_pli_rec_null_data.txt; accessed 18 August 2016). We obtained SE calls (*n* = 65,950 SEs from 99 tissues/cells) from dbSuper[41] (http://bioinfo.au.tsinghua.edu.cn/dbsuper/data/bed/hg19/all_hg19_bed.bed; accessed 17 August 2016). These SEs were ranked according to signal intensity within each cell/tissue. We assigned each SE to the closest protein-coding gene promoter within a 50-kb radius[62] (ENSEMBL GRCh37.75; http://ftp.ensembl.org/pub/release-75/gtf/homo_sapiens/Homo_sapiens.GRCh37.75.gtf.gz; accessed 18 August 2016) with BEDTOOLS[63]. If a gene was near multiple SEs, we assigned it the highest observed SE rank. Finally, to explore the relationship between pLI score and enhancer architecture, we combined this gene-centric table of SE ranks with the EXAC pLI table. Specifically, we determined the median pLI score observed with varying thresholds of SE rank. To explore the specific roles of transcription factors, we obtained a comprehensive list of human transcription factors from AnimalTFDB[64] (http://www.bioguo.org/AnimalTFDB/download/Homo_sapiens_TF_EnsemblID.txt; accessed 14 September 2016). We then determined the fraction of transcription factors with varying thresholds of SE rank. We created the plots in R.

GWAS data (gwas_catalog_v1.0) were downloaded from the NHGRI–EBI GWAS catalog (http://www.ebi.ac.uk/gwas/downloads). We converted the hg38 single-nucleotide polymorphism coordinates to hg19 coordinates with liftOver from the UCSC Genome Browser (http://hgdownload.cse.ucsc.edu/downloads.html#source_downloads). Genomic-region overlapping analyses were conducted with BEDTools[63]. A single-nucleotide polymorphism was assigned to a gene if its coordinate was within the gene body (transcription start to transcription end, as defined by RefSeq hg19). Haplosufficient and haploinsufficient genes with GWAS associations are listed in **Supplementary Table 5**. Fisher exact tests were carried out with R3.2.0. Data extraction, data reformatting and data preparation for analysis were all facilitated with customized scripts of Bash, Python and R.

Data analysis and visualization. Data were analyzed with Microsoft Excel and GraphPad Prism (Graph Pad Software) and visualized with CLC Main Workbench 7 (CLCbio, Qiagen) and DataGraph 3.2 (Visual Data Tools, Inc.). Molecular graphics and analyses were generated with the UCSF Chimera package. Chimera is developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIGMS P41-GM103311). Statistical analyses were done with appropriate parametric and nonparametric tests. Multiple data sets were compared by repeated-measures ANOVA. Statistical analysis of data in contingency tables was done by Fisher exact test. Two-tailed *P* values of <0.05 were considered statistically significant.

Data availability. The data that support the findings of this study are available from the corresponding author upon request.

46. Rutishauser, R.L. *et al.* Transcriptional repressor Blimp-1 promotes CD8⁺ T cell terminal differentiation and represses the acquisition of central memory T cell properties. *Immunity* **31**, 296–308 (2009).
47. Khoder, A. *et al.* Regulatory B cells are enriched within the IgM memory and transitional subsets in healthy donors but are deficient in chronic GVHD. *Blood* **124**, 2034–2045 (2014).
48. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
49. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
50. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
51. Naviaux, R.K., Costanzi, E., Haas, M. & Verma, I.M. The pCL vector system: rapid production of helper-free, high-titer, recombinant retroviruses. *J. Virol.* **70**, 5701–5705 (1996).

52. Wingfield, P.T. *et al.* Biophysical and functional characterization of full-length, recombinant human tissue inhibitor of metalloproteinases-2 (TIMP-2) produced in *Escherichia coli*. Comparison of wild type and amino-terminal alanine appended variant with implications for the mechanism of TIMP functions. *J. Biol. Chem.* **274**, 21362–21368 (1999).

53. Liu, X., Jian, X. & Boerwinkle, E. dbNSFP: a lightweight database of human nonsynonymous SNPs and their functional predictions. *Hum. Mutat.* **32**, 894–899 (2011).

54. Liu, X., Jian, X. & Boerwinkle, E. dbNSFP v2.0: a database of human non-synonymous SNVs and their functional predictions and annotations. *Hum. Mutat.* **34**, E2393–E2402 (2013).

55. Itan, Y. *et al.* The mutation significance cutoff: gene-level thresholds for variant predictions. *Nat. Methods* **13**, 109–110 (2016).

56. Dang, V.T., Kassahn, K.S., Marcos, A.E. & Ragan, M.A. Identification of human haploinsufficient genes and their genomic proximity to segmental duplications. *Eur. J. Hum. Genet.* **16**, 1350–1357 (2008).

57. Huang, W., Sherman, B.T. & Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).

58. Huang, W., Sherman, B.T. & Lempicki, R.A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**, 1–13 (2009).

59. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).

60. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).

61. Bible, P.W. *et al.* PAPST, a user friendly and powerful Java platform for ChIP-seq peak co-localization analysis and beyond. *PLoS One* **10**, e0127285 (2015).

62. Aken, B.L. *et al.* The Ensembl gene annotation system. *Database (Oxford)* **2016**, baw093 (2016).

63. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

64. Zhang, H.-M. *et al.* AnimalTFDB: a comprehensive animal transcription factor database. *Nucleic Acids Res.* **40**, D144–D149 (2012).