

ARTICLE

Constrained chromatin accessibility in PU.1-mutated agammaglobulinemia patients

Carole Le Coz¹, David N. Nguyen^{2,3,4,5,6}, Chun Su^{7,8}, Brian E. Nolan^{9,10}, Amanda V. Albrecht¹¹, Suela Khani¹¹, Di Sun¹, Benjamin Demaree^{12,13}, Piyush Pillarisetti¹, Caroline Khanna¹, Francis Wright^{2,3}, Peixin Amy Chen^{3,4,5,6}, Samuel Yoon¹, Amy L. Stiegler¹⁴, Kelly Maurer¹, James P. Garifallou¹⁵, Amy Rymaszewski¹⁶, Steven H. Kroff¹⁷, Timothy S. Olson^{18,19}, Alix E. Seif^{18,19}, Gerald Wertheim^{20,21}, Struan F.A. Grant^{7,19,22,23}, Linda T. Vo^{4,5}, Jennifer M. Puck^{24,25,26}, Kathleen E. Sullivan^{19,27}, John M. Routes¹⁶, Viktoria Zakharova²⁸, Anna Shcherbina²⁹, Anna Mukhina²⁹, Natasha L. Rudy³⁰, Anna C.E. Hurst^{30,31}, T. Prescott Atkinson³¹, Titus J. Boggon^{14,32}, Hakon Hakonarson^{15,19}, Adam R. Abate^{12,13,33}, Joud Hajjar^{34,35}, Sarah K. Nicholas^{34,35}, James R. Lupski^{36,37,38}, James Verbsky¹⁶, Ivan K. Chinn^{34,35}, Michael V. Gonzalez¹⁵, Andrew D. Wells^{8,20,27}, Alex Marson^{2,3,4,5,6,33,39}, Gregory M.K. Poon¹¹, and Neil Romberg^{1,19,27}

The pioneer transcription factor (TF) PU.1 controls hematopoietic cell fate by decompacting stem cell heterochromatin and allowing nonpioneer TFs to enter otherwise inaccessible genomic sites. PU.1 deficiency fatally arrests lymphopoiesis and myelopoiesis in mice, but human congenital PU.1 disorders have not previously been described. We studied six unrelated agammaglobulinemic patients, each harboring a heterozygous mutation (four de novo, two unphased) of *SPI1*, the gene encoding PU.1. Affected patients lacked circulating B cells and possessed few conventional dendritic cells. Introducing disease-similar *SPI1* mutations into human hematopoietic stem and progenitor cells impaired early in vitro B cell and myeloid cell differentiation. Patient *SPI1* mutations encoded destabilized PU.1 proteins unable to nuclear localize or bind target DNA. In PU.1-haploinsufficient pro-B cell lines, euchromatin was less accessible to nonpioneer TFs critical for B cell development, and gene expression patterns associated with the pro- to pre-B cell transition were undermined. Our findings molecularly describe a novel form of agammaglobulinemia and underscore PU.1’s critical, dose-dependent role as a hematopoietic euchromatin gatekeeper.

¹Division of Immunology and Allergy, Children’s Hospital of Philadelphia, Philadelphia, PA; ²Division of Infectious Diseases, Department of Medicine, University of California San Francisco, San Francisco, CA; ³Department of Microbiology and Immunology, University of California San Francisco, San Francisco, CA; ⁴Diabetes Center, University of California San Francisco, San Francisco, CA; ⁵Innovative Genomics Institute, University of California Berkeley, Berkeley, CA; ⁶Gladstone–University of California San Francisco Institute of Genomic Immunology, San Francisco, CA; ⁷Division of Human Genetics, Children’s Hospital of Philadelphia, Philadelphia, PA; ⁸Center for Spatial and Functional Genomics, Children’s Hospital of Philadelphia, Philadelphia, PA; ⁹Division of Rheumatology, Ann & Robert H. Lurie Children’s Hospital of Chicago, Chicago, IL; ¹⁰Department of Pediatrics, Feinberg School of Medicine, Northwestern University, Chicago, IL; ¹¹Department of Chemistry and Center for Diagnostics and Therapeutics, Georgia State University, Atlanta, GA; ¹²Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, San Francisco, CA; ¹³University of California Berkeley–University of California San Francisco Graduate Program in Bioengineering, University of California, San Francisco, CA; ¹⁴Departments of Pharmacology, Yale University, New Haven, CT; ¹⁵Center for Applied Genomics, Children’s Hospital of Philadelphia, Philadelphia, PA; ¹⁶Division of Allergy and Immunology, Department of Pediatrics, Medical College of Wisconsin, Milwaukee, WI; ¹⁷Department of Pathology, Medical College of Wisconsin, Milwaukee, WI; ¹⁸Division of Oncology, Children’s Hospital of Philadelphia, Philadelphia, PA; ¹⁹Department of Pediatrics, Perelman School of Medicine, Philadelphia, PA; ²⁰Department of Pathology and Laboratory Medicine, Children’s Hospital of Philadelphia, Philadelphia, PA; ²¹Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; ²²Division of Diabetes and Endocrinology, Children’s Hospital of Philadelphia, Philadelphia, PA; ²³Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; ²⁴Division of Allergy, Immunology, and Bone Marrow Transplantation, Department of Pediatrics, University of California, San Francisco, CA; ²⁵University of California San Francisco Institute for Human Genetics and Smith Cardiovascular Research Institute, University of California, San Francisco, CA; ²⁶UCSF Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, CA; ²⁷Institute for Immunology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; ²⁸Laboratory of Molecular Biology, Dmitry Rogachev National Medical Research Center of Pediatric Hematology, Oncology and Immunology, Moscow, Russia; ²⁹Department of Immunology, Dmitry Rogachev National Medical Research Center of Pediatric Hematology, Oncology and Immunology, Moscow, Russia; ³⁰Department of Genetics, University of Alabama at Birmingham, Birmingham, AL; ³¹Department of Pediatrics, University of Alabama at Birmingham, Birmingham, AL; ³²Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT; ³³Chan Zuckerberg Biohub, San Francisco, CA; ³⁴William T. Shearer Center for Human Immunobiology, Texas Children’s Hospital, Houston, TX; ³⁵Department of Immunology, Allergy and Rheumatology, Baylor College of Medicine, Houston, TX; ³⁶Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; ³⁷Texas Children’s Hospital, Houston, TX; ³⁸Baylor-Hopkins Center for Mendelian Genomics, Houston, TX; ³⁹Parker Institute for Cancer Immunotherapy, San Francisco, CA.

Correspondence to Neil Romberg: romberg@chop.edu.

© 2021 Le Coz et al. This article is distributed under the terms of an Attribution–Noncommercial–Share Alike–No Mirror Sites license for the first six months after the publication date (see <http://www.rupress.org/terms/>). After six months it is available under a Creative Commons License (Attribution–Noncommercial–Share Alike 4.0 International license, as described at <https://creativecommons.org/licenses/by-nc-sa/4.0/>).

Introduction

Pioneer transcription factors (TFs) are early and important epigenetic determinants of cell fate (Mayran and Drouin, 2018). Their outsized developmental role is well illustrated by the hematopoietic pioneer PU.1. Unlike nonpioneer TFs, PU.1 is capable of decompacting dense stem cell heterochromatin to create stably accessible euchromatic regions (Heinz et al., 2010; Pham et al., 2013; Sherwood et al., 2014). Once in open chromatin, PU.1 can directly control gene expression by binding genetic regulatory elements and can also more broadly influence transcription by recruiting nonpioneers, such as interferon regulatory factors (IRFs), to otherwise inaccessible genomic regions (Ciau-Uitz et al., 2013; Heinz et al., 2013; Pongubala and Atchison, 1997; Sherwood et al., 2014). Homozygous PU.1 mutations are embryonically lethal to mice, arresting lymphoid and myeloid development while erythropoiesis and thrombopoiesis remain intact (DeKoter et al., 2007a; Scott et al., 1994). Depending on the mouse model, subtle decreases in PU.1 dose either discourage B cell development or favor it over myelopoiesis (Houston et al., 2007; DeKoter et al., 2007a, 2007b; DeKoter and Singh, 2000). Additionally, PU.1 low-expressing mice develop acute myeloid leukemia (AML; Rosenbauer et al., 2004; Will et al., 2015), a potential parallel to human AML, where PU.1 can be somatically mutated and is often epigenetically repressed (Mueller et al., 2002; Steidl et al., 2006). No inborn diseases of PU.1 have been previously described.

Congenital agammaglobulinemia encompasses a group of primary disorders caused by early-stage B cell developmental arrest. Most congenital agammaglobulinemia cases are caused by X-linked BTK deficiency, but rarer autosomal recessive (IGHM, CD79A, CD79B, IGLL1, BLINK, and PIK3R1) and autosomal dominant (TCF3) forms of disease have also been described (Bruton, 1952; Conley et al., 2009). We identified six unrelated agammaglobulinemic patients with heterozygous mutations of *SPI1*, the autosomal gene encoding PU.1. PU.1-mutated agammaglobulinemia (PU.MA) patients demonstrated early B cell developmental arrest and deficiencies of PU.1^{hi} conventional dendritic cells (cDCs). Herein, we molecularly describe why mutated PU.1 proteins fail to properly interact with target DNA, demonstrate how these defects alter chromatin accessibility, and determine which hematopoietic lineages are affected.

Results

Description of patients

We performed whole-exome sequencing on a multi-institution cohort of 30 undiagnosed agammaglobulinemia patients. From these, we identified six unrelated patients (ages 15 mo to 37 yr; Fig. 1 A and Table 1) with novel, germline, heterozygous mutations in *SPI1*, the gene encoding PU.1 (Fig. 1 B). At study enrollment, all patients lacked circulating B cells (Fig. 1 C and Table 2). For five patients, B cell loss appeared congenital, as serum Igs were undetectable at diagnosis. The age of disease onset was less clear for patient C.II.1, who was first clinically assessed in his 23rd year. At that time point, he lacked circulating B cells but possessed detectable serum IgG (333 mg/dl) and a protective tetanus titer. Over the subsequent decade, he developed frank agammaglobulinemia and progressive lymphopenia.

All PU.MA patients experienced sinopulmonary infections with encapsulated bacteria before initiation of Ig replacement

therapy (Table 1). Patient B.II.2 was additionally diagnosed with meningococcal and enterococcal meningitis at age 9 mo. Systemic enteroviral infections were documented in two patients. Patient A.II.2 cleared prolonged echovirus-25 viremia, which was associated with transient neutropenia and reactive arthritis, after a course of oral plicapavir. He subsequently underwent hematopoietic stem cell transplantation from a matched sibling donor (A.II.1) but remained dependent on antibody replacement therapy after the procedure. Patient D.II.1 was the only patient to receive live-attenuated oral polio vaccine, and afterward he developed acute flaccid paralysis. Polio virions were isolated from his cerebral spinal fluid. Two patients (B.II.2 and F.II.1) received live Bacille Calmette-Guérin (BCG) vaccines, and neither experienced infectious complications. Noninfectious diagnoses reported by patients included idiopathic myositis (E.I.1), idiopathic neonatal lipodystrophy (B.II.2), and childhood-onset insulin-dependent diabetes (B.II.2). At the time of manuscript submission, all patients were alive, and none had experienced a malignant disease.

Each PU.MA patient harbored a different, novel *SPI1* mutation (Fig. 1 B). We predicted that four mutations (G109Sfs*78, Q111X, Y122X, and L233Afs*53; three de novo and one unphased) would be highly damaging to PU.1 and would convey a loss of function. Of these, one mutation (G109Sfs*78) frameshifted and two inserted premature termination codons into the central PU.1 PEST domain (Q111X, Y122X; Fig. 1 B). The fourth mutation (L233Afs*53) was predicted to generate a slightly longer protein than WT PU.1 by replacing the DNA-binding ETS domain with 53 nonsense amino acids. The two remaining *SPI1* mutations were missense variants (H212P and V242G, de novo and unphased) that substituted key ETS residues invariant among orthologues (Chang et al., 2018). Although common in our multi-institution agammaglobulinemia cohort, protein-altering *SPI1* variants were rare in large, publicly available human genetic databases. For instance, no loss-of-function *SPI1* variants were listed in the 141,456 exomes/genomes comprising the Genome Aggregation Database v2.1.1 (gnomAD; predicted loss intolerance score = 0.98; Lek et al., 2016). Further, *SPI1*'s nonsynonymous mutation load among the general population was similar to that of autosomal dominant disease-causing genes listed in Online Mendelian Inheritance in Man (Gene Damage Index Phred-Scaled Score = 1.335; Itan et al., 2015). The additional slim odds of four de novo mutations being chance identified in a *SPI1*-sized gene (1.1×10^{-20} to 1; Kessler et al., 2020), the tight segregation of mutations with disease in patient pedigrees (Fig. 1 A), and the lack of alternative genetic explanations (Tangye et al., 2020) indicated PU.1 mutations were likely contributors to our patients' immunodeficient phenotype.

Circulating B cells and cDCs are scarce in PU.MA patients

To determine the potential effects of mutant PU.1 on circulating blood cells, we performed in-depth cellular and molecular analyses of patient blood samples. Using an N-terminal detection antibody, mean PU.1 protein intensities were far greater in B cell-depleted healthy donor (HD) peripheral blood mononuclear cell (PBMC) lysates than in patient PBMC lysates (0.75 versus 0.05; $P < 0.001$; Fig. S1 A). To investigate such large

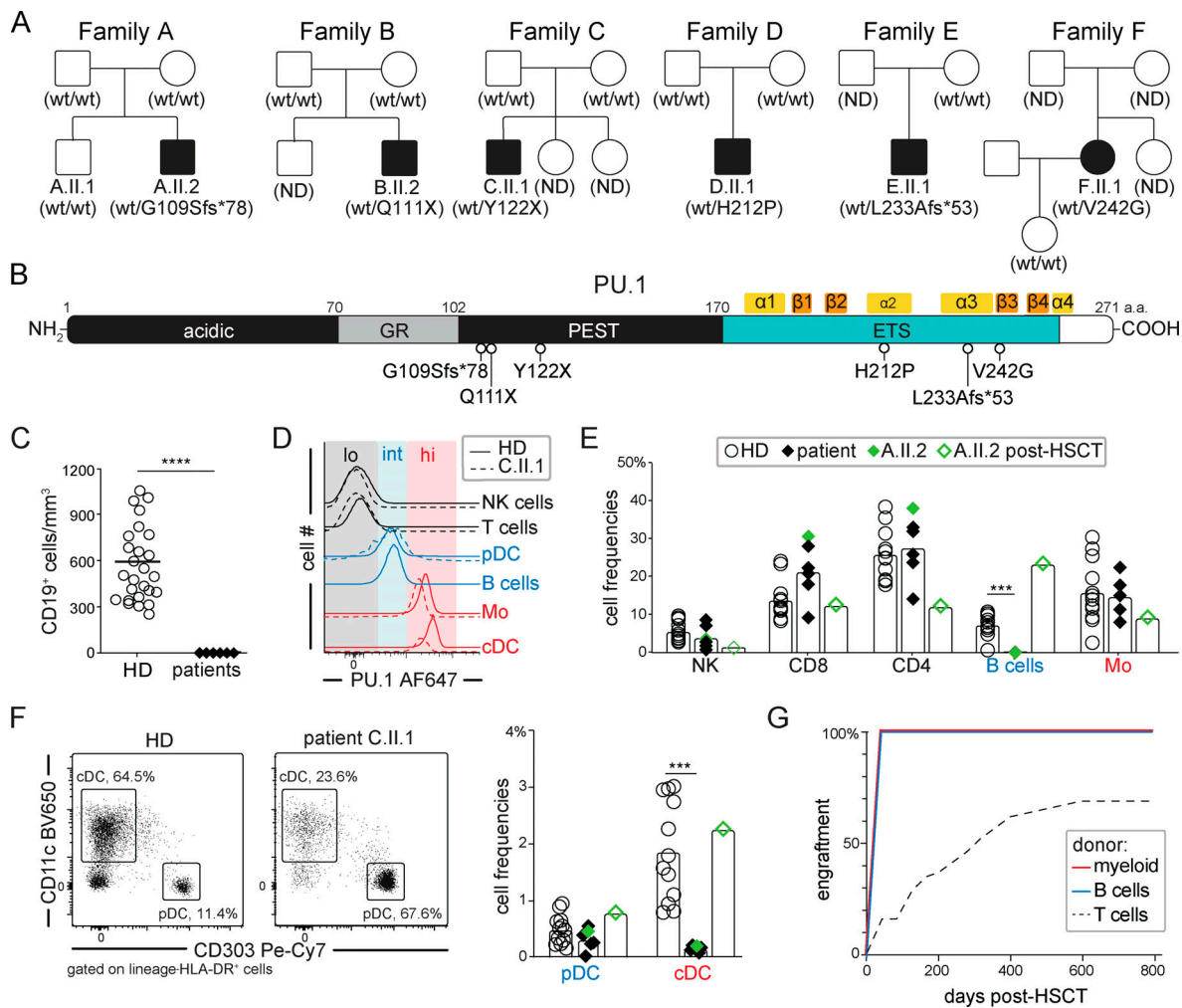


Figure 1. PU.MA patients are B cell and cDC deficient. (A and B) The distribution of PU.1 mutations across patient pedigrees (A) and along the protein’s functional domains (B) are illustrated. Tertiary structures comprising the DNA-binding ETS domain are labeled. **(C)** Peripheral CD19⁺ B cell concentrations for HD (*n* = 26) and PU.MA patients (*n* = 6) are shown with horizontal lines indicating mean values. **(D)** Histograms show PU.1 expression of indicated PBMC subsets from a representative HD (unbroken line) and PU.MA patient C.II.1 (dashed line). **(E)** Bars indicate average subset frequencies relative to total PBMCs from HDs (*n* = 12) and patients (*n* = 6). Patient A.II.2 subset frequencies before (filled green diamond) and after hematopoietic stem cell transplant (HSCT; unfilled green diamonds) are shown. **(F)** Flow cytometry plots show peripheral pDCs and cDCs from a representative HD and patient C.II.1. pDC and cDC frequencies relative to total PBMCs from HD (*n* = 12) and patients (*n* = 6) are displayed. **(G)** The proportion of donor-specific peripheral myeloid cells, B cells, and T cells circulating in patient A.II.2 are shown at specified time points after hematopoietic stem cell transplantation (HSCT). ***, *P* < 0.001; ****, *P* < 0.0001 (two-sided Mann–Whitney *U* tests). Mo, monocyte; ND, not done.

differences, we surveyed HD leukocyte subsets for PU.1 expression using flow cytometry and divided leukocytes into three groups: PU.1^{lo} (T cell, natural killer [NK] cell, and neutrophil), PU.1^{int} (B cell and plasmacytoid dendritic cell [pDC]), or PU.1^{hi}-expressing cells (monocyte and cDC; Fig. 1 D; and Fig. S1, B and C). PU.1^{lo} cells, especially CD8⁺ T cells, were overrepresented in patient PBMCs compared with HD counterparts (Fig. 1 E), while PU.1^{int} B cell and PU.1^{hi} cDC frequencies were significantly diminished (7.5% versus 0.06% and 1.8% versus 0.14% of PBMCs, respectively; *P* < 0.001 for both comparisons; Fig. 1 F). In addition to being less numerous (Fig. S2 A), PU.1, IL-12, and IL-6 expression lagged considerably in patient cDCs compared with HD counterparts (Fig. 1 D, Fig. S1 D, and Fig. 2 B, top row). Despite cDC-specific defects, patients’ unfractionated myeloid cells secreted IL-12 and IL-6 at frequencies comparable to that of HD counterparts (Fig. S2 B, bottom row). Other than patient C.II.1,

who was cytopenic at study enrollment, PU.MA patient pDC, NK cell, platelet, erythrocyte, T helper cell subsets, monocyte subsets, innate lymphoid cell (ILC) subsets, and granulocyte subsets were not appreciably different from those of HD controls (Fig. S2, C–G and Table 2).

After adoptive transfer of hematopoietic stem and progenitor cells (HSPCs) from his unaffected sibling (A.II.1), patient A.II.2’s PU.1 expression normalized and his B cell/cDC deficiencies resolved (Fig. 1, E and F). Serial measurement of peripheral blood chimerism after transplantation showed rapid engraftment of donor B cells and myeloid cells by post-transplant day 37 (Fig. 1 G). In contrast, significant mixed chimerism (~40% recipient) persisted in the T lymphocyte pool past post-transplant day 800, suggesting that normalized PU.1 expression affords a competitive advantage to developing B cells and myeloid cells but not to developing T cells. Thus, PU.MA is a hematopoietic

Table 1. Patient demographic, genotypic, and clinical information

Patient	A.II.2	B.II.2	C.II.1	D.II.1	E.II.1	F.II.1
Age	15 mo	7 yr	37 yr	32 yr	8 yr	16 yr
Sex	Male	Male	Male	Male	Male	Female
Heterozygous <i>SPI1</i> variant	g.18716_18718delinsAG c.325_327delGGCinsAG p.G109Sfs*78	g.18722C>T c.331C>T p.Q111X	g.19603C>A c.366C>A p.Y122X	g.23169A>C c.635A>C p.H212P	g.23230_23231delGC c.696_697delGC p.L233Afs*53	g.23259T>G c.725T>G p.V242G
Variant CADD score	—	42	35	26.9	—	28.9
Variant origin	De novo	De novo	De novo	De novo	Unphased	Unphased
Infections	Sinopulmonary, echovirus 25 viremia	Sinopulmonary, meningitis	Sinopulmonary	Sinopulmonary, polioencephalitis	Sinopulmonary	Sinopulmonary
Complications	Transient neutropenia, reactive arthritis	Neonatal lipodystrophy, T1D	Progressive lymphopenia	Paralytic poliomyelitis	Myositis	—
Therapies	IgG replacement, pocapavir, HSCT	IgG replacement	IgG replacement	IgG replacement	IgG replacement	IgG replacement

CADD, combined annotation-dependent depletion; HSCT, hematopoietic stem cell transplant; T1D, type 1 diabetes mellitus.

cell-intrinsic disease remarkable for scarcity of PU.1-expressing cells such as PU.1^{hi} cDCs and particularly PU.1^{int} B cells.

PU.1 loss undermines B cell development

To understand the role of PU.1 loss on human hematopoiesis, we used cellular indexing of transcriptomes and epitopes by sequencing (CITE-seq) to capture the expression of 132 hematopoietic cell-surface proteins on mononuclear bone marrow cells from patient A.II.2 and an age-, gender-matched HD control (Stoeckius et al., 2017). We dimensionally reduced protein-associated reads with uniform manifold approximation and projection (UMAP; Fig. 2 A) and labeled cell clusters based upon expression of key proteins/transcripts (Fig. S3). In HD marrow, *SPI1* transcripts were concentrated in hematopoietic stem cell, lymphoid progenitor, myeloid lineage (granulocyte macrophage progenitor, cDC, monocyte), and B cell clusters (Fig. 2 A, left). Among HD B cells, pre-B1 cells expressed the most *SPI1*, a finding confirmed by measuring intracellular PU.1 with flow cytometry (Fig. 2 B). As a contrast to the full B cell development arc observed in HD marrow, A.II.2 marrow contained only pro-B cells, indicating lineage arrest before the pre-B1 PU.1 pulse. In addition to B cell alteration, A.II.2 marrow contained ~3.5 times more $\alpha\beta$ T cells than the HD sample did. This increase, which was driven primarily by an enlarged activated CD8⁺ T cell pool, may have been exaggerated by active, systemic echoviremia at the time of A.II.2 marrow aspiration. Other quantitative differences between HD and patient marrow cell clusters appeared more minor.

To determine transcriptional changes associated with B cell developmental arrest in PU.MA patients, we identified differentially expressed genes (DEGs) between single CD19⁺CD20⁻IgD⁻ pro-B cells from HD ($n = 279$) and patient A.II.2 ($n = 35$; Fig. 2 C). As expected, *SPI1* transcripts were reduced by approximately half (\log_2 fold change [FC] = -0.9; $P < 0.01$) in A.II.2 pro-B cells (Fig. 2 C). This relatively modest decrease was accompanied by significantly larger drops (\log_2 FC < -1) in the expression of a myriad of genes involved in B cell

development (*IGLL5*, *BTGI*, *IGHM*), migration (*RGS2*), activation (*TCLIA*, *FCRLA*), and regulation (*CD37*; Hystad et al., 2007; Facchetti et al., 2002; van Spruiel et al., 2012). Also differentially down-regulated in A.II.2 pro-B cells were genes known to cause other forms of agammaglobulinemia such as *BTK*, *IGHM*, and *CD79B* (Tangye et al., 2020). Notably, *BTK*, *CD79B*, *TCLIA*, *CD37*, and *BTGI* are all regulatory targets of PU.1 in human B cells (ENCODE Project Consortium, 2012). Prominent among up-regulated patient pro-B cell transcripts was *DNTT* (\log_2 FC = 0.922; $P < 0.01$), the gene encoding terminal deoxynucleotidyl transferase (TdT). As with other forms of agammaglobulinemia, most PU.MA B cell precursors strongly stained TdT positive, indicating loss of pre-B cell receptor-mediated feedback (Fig. 2, D and E; Conley et al., 2009; Wasserman et al., 1997). Similarly, immunohistochemical (IHC) staining of an A.II.2 marrow biopsy yielded easily identifiable CD79a⁺ pro-B cells but few CD20⁺ pre-B cells and zero CD138⁺ terminally differentiated plasma cells (Fig. 2 D). Hence, B cell lymphopoiesis arrested between the pro- and pre-B cell stages for patient A.II.2, the precise time point when demand increased for PU.1 and PU.1-regulated gene products such as *BTK*, *CD79B*, *TCLIA*, *CD37*, and *BTGI*.

As animal models of PU.1 downmodulation have generated conflicting results (DeKoter and Singh, 2000; DeKoter et al., 2007a; Houston et al., 2007), we modeled PU.MA patient hematopoiesis in human cord blood HSPCs by editing *SPI1* with CRISPR/Cas9. We designed guide RNAs (gRNAs) targeting sites near PU.MA patient mutations in either exon four, which encodes the PEST domain, or *SPI1* exon five, which encodes the ETS domain, or as a control at the *AAVS1* intronic safe harbor site (Fig. 3 A). To determine the site-specific efficiencies of each *SPI1* edit, we single-cell genotyped HSPCs after electroporation with CAS9 and gRNAs (Demaree et al., 2021). The majority (76%) of *SPI1* exon four-targeted cells possessed either monoallelic or biallelic edits, and the frequency of cells with exon five edits was even higher, exceeding 97% (Fig. S4 A). In a separate set of experiments, we electroporated HSPCs from two cord blood donors

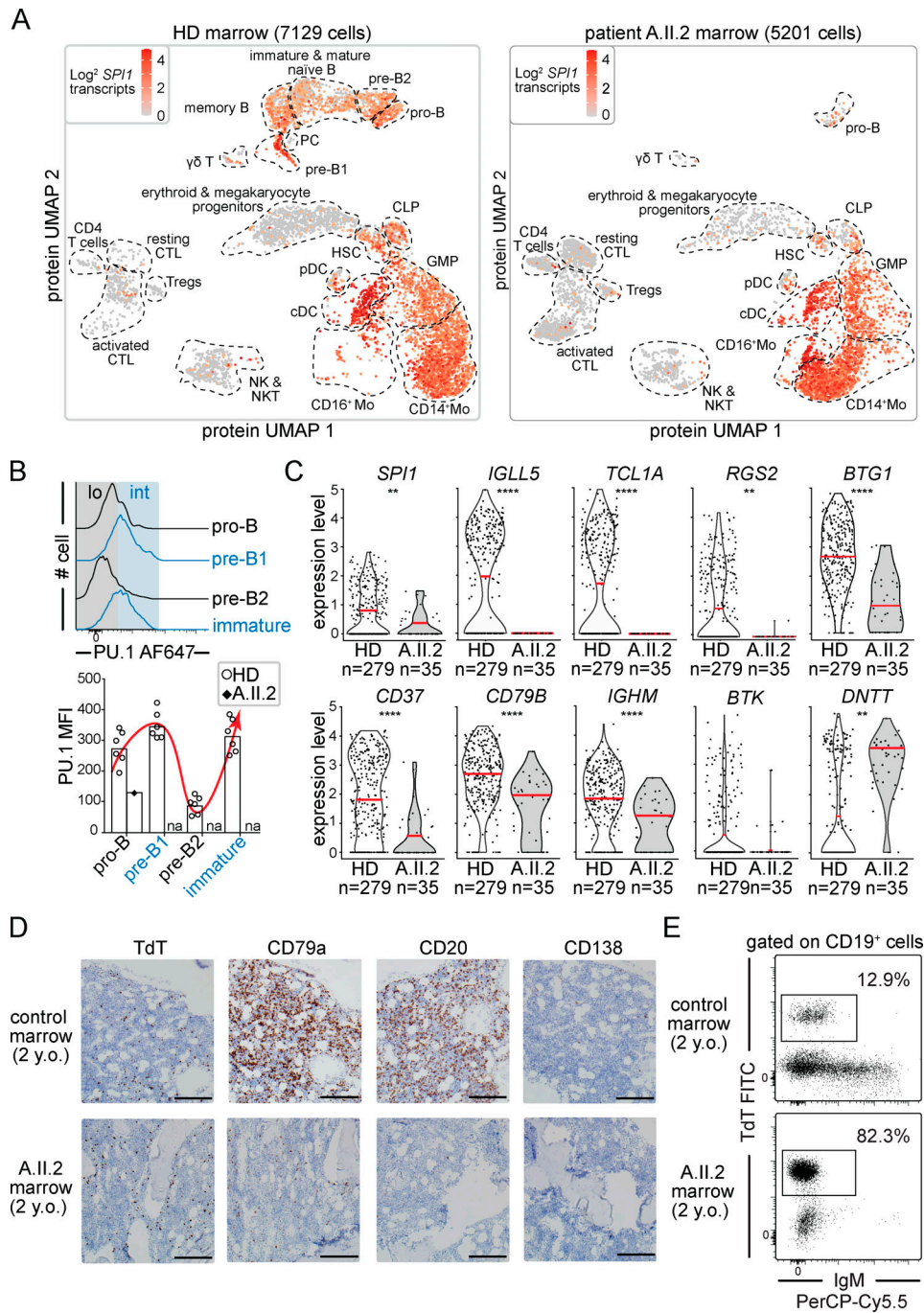


Figure 2. Patient A.II.2: B cell development arrests before the pre-B1 stage. (A) Heat maps of \log_2 *SPI1* transcripts are overlaid upon pediatric HD and patient A.II.2 mononuclear marrow cells. The cells were clustered based on expression of 132 cell-surface proteins using UMAP dimensional reduction. Single-cell protein expression was simultaneously captured with transcript expression using CITE-seq. Labels for cell clusters were anchored on key identifying proteins and transcripts (Fig. S3). **(B)** Histograms show intracellular PU.1 expression in B cell precursors from a representative HD measured via flow cytometry with PU.1 mean fluorescence intensities (MFIs) of six pediatric HDs and patient A.II.2 B cell precursors. A red line marks the PU.1 expression trend through HD B cell development. **(C)** Differential expression of indicated transcripts between HD and patient A.II.2. $CD19^+CD20^-IgD^-$ pro-B cells are depicted with violin plots. Red bars indicate mean expression values. **(D)** Bone marrow biopsies from patient A.II.2 and an unrelated, age- and sex-matched control are displayed with IHC staining of TdT (pro-B cell marker), CD79a (pan-B cell marker), CD20 (expressed from pre-B to plasmablast stages), and CD138 (plasma cell marker). Scale bars indicate 200 μ m. **(E)** Flow cytometry plots show TdT expressing B cell precursors in marrow aspirates from an age-, sex-matched HD and patient A.II.2. TdT expression frequencies are reported as a percentage of total $CD19^+$ marrow cells. **, $P < 0.01$; ****, $P < 0.0001$ (two-sided Mann-Whitney *U* tests). $\gamma\delta$ T, $\gamma\delta$ T cell; CLP, common lymphoid progenitor; CTL, Cytotoxic T lymphocyte; GMP, granulocyte-monocyte progenitor; HSC, hematopoietic stem cell; Mo, monocyte; na, not applicable; NKT, natural killer T cell; PC, plasma cell; Tregs, regulatory T cell; y.o., years old.

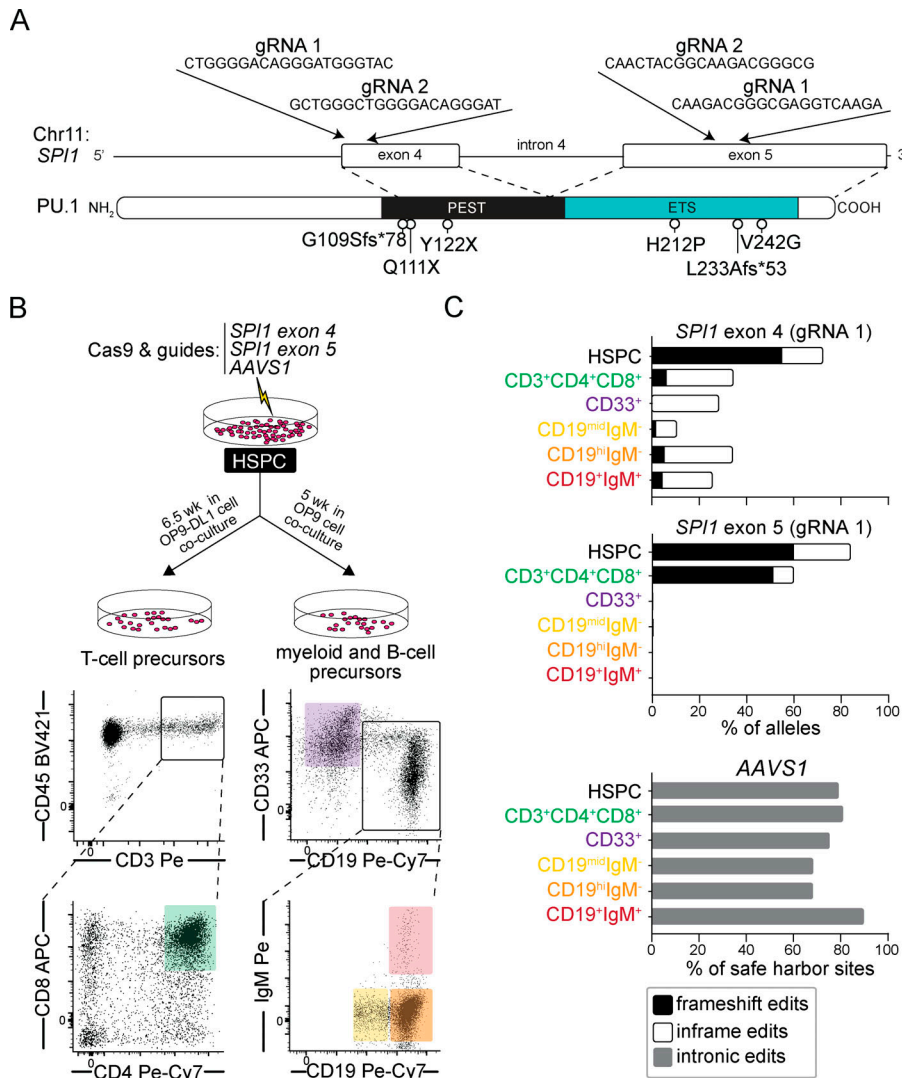


Figure 3. PU.1 ETS domain loss alters human myelopoiesis and B cell lymphopoiesis. (A) gRNA sequences and genomic target sites within *SPI1* exons four and five are illustrated at positions relative to patient PU.1 mutations. **(B)** Strategies to in vitro differentiate CRISPR/Cas9-edited human hematopoietic stem cells and to sort T cell, B cell, and myeloid precursors are depicted. **(C)** Frequencies of in-frame and frame-shifting edits introduced into *SPI1* exons four and five or intronic edits introduced into a separate AAVS1 safe harbor site are shown for hematopoietic stem cells before differentiation and for cells surviving differentiation. Data represent one of two independent cord blood donor experiments with one of two *SPI1* gRNA sets.

with CAS9 and gRNAs before in vitro differentiating them into T cell, B cell, and myeloid precursors (Fig. 3 B). Consistent with single-cell genotyping results, most edited HSPCs contained either frame-shifting or in-frame (*SPI1*) or intronic (*AAVS1*) mutations before differentiation (Fig. 3 C; and Fig. S4, B and C). After differentiation, approximately half of *SPI1* alleles in T cell precursors (CD3⁺CD4⁺CD8⁺) retained exon five-frameshifting indels. In contrast, almost no myeloid (CD33⁺) or B cell precursors (CD19^{int}IgM⁻, CD19^{hi}IgM⁻, or CD19^{hi}IgM⁺) possessed *SPI1* exon five mutations of any kind despite very high initial editing efficiency at that genomic site. Similarly, in-frame exon four edits that preserved the ETS were tolerated in B cell precursors, whereas ETS-altering frameshift edits of exon four were not. Thus, biallelic PU.1 ETS expression is important for human B and myeloid cell development in vitro and in vivo, whereas developing T cells appear more tolerant of partial PU.1 ETS loss.

Patient *SPI1* mutations impair PU.1 expression, nuclear localization, and DNA binding

We predicted that *SPI1* mutations identified in patients A.II.2, B.II.2, and C.II.1 (G109Sfs*78, Q111X, and Y122X) would generate

truncated proteins (28 kD and 26 kD, respectively) lacking full PEST and ETS domains (Fig. 1 B), but we were unable to detect these products in patient PBMC lysates (Fig. S1 A) or overexpress them in HEK293 cells (Fig. 4 A). In cDNA, mutant *SPI1* transcripts were significantly underrepresented for patient A.II.2 but not patient B.II.2 or C.II.1 (Fig. S1 E), suggesting susceptibility of the G109Sfs*78 encoding mRNA to decay and the Q111X or Y122X proteins to post-transcriptional degradation.

Proteins encoded by patient D.II.1's, E.II.1's, and F.II.1's mutant *SPI1* alleles (H212P, L233Afs*53, and V242G) could be overexpressed in HEK293 cells, albeit at lower concentrations than WT PU.1 (Fig. 4 A). To determine if mutated PU.1 could promote transcription, we used a HEK293 PU.1-reporter line containing an enhanced GFP (*EGFP*) gene under the transcriptional control of a PU.1-dependent, λB-based promoter/enhancer nucleotide sequence (Fig. S5 A; Munde et al., 2014). Compared with reporter cells expressing unmutated PU.1 or PU.1 altered to contain one of three common, nonsynonymous gnomAD variants (Lek et al., 2016), reporter cells expressing H212P, L233Afs*53, and V242G mutated PU.1 proteins emitted EGFP light less frequently (25.4–29.4% versus 5.4–10%; P < 0.0001; Fig.

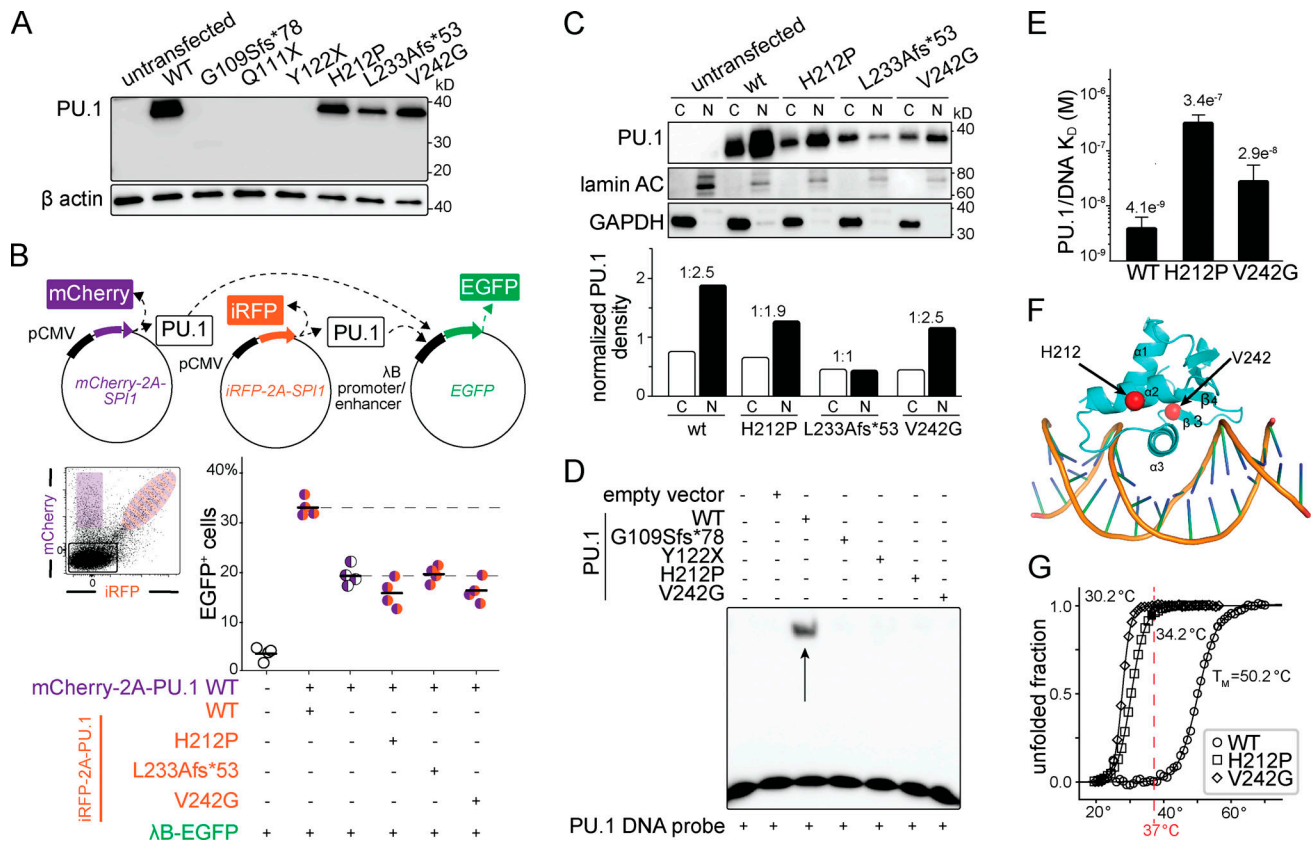


Figure 4. ETS-mutated PU.1 mutants encode destabilized proteins that cannot bind target DNA. (A) HEK293-overexpressed PU.1 proteins are detected by an N-terminal anti-PU.1 antibody. **(B)** Top: A diagram depicts the cotransfection strategy used to determine if PU.1 mutants affect the ability of WT PU.1 to drive EGFP expression by binding an upstream λB promoter. Bottom: iRFP (orange) and/or mCherry (purple) expression signifies cells successfully transfected with *SPI1*. Dashed horizontal lines indicate the fraction of cells expressing EGFP after transfection with full (100 ng) or half doses (50 ng) of unmutated *SPI1* plasmids in four independent experiments. The frequency of EGFP⁺ cells coexpressing WT and mutant PU.1 proteins is displayed. **(C)** HEK293-overexpressed PU.1 proteins are detected in cytoplasmic (C) or nuclear (N) lysates, normalized to GAPDH or lamin AC, respectively, and ratioed. Data are from one of two representative experiments. **(D)** PU.1 DNA probe shifting (arrow) of WT PU.1-expressing HEK293 cell lysates from one of two independent experiments is shown. **(E)** Dissociation constants (K_D s) of 1:1 of PU.1 ETS/DNA complexes was determined by three independent experiments. **(F)** The DNA-bound murine PU.1 ETS crystal structure is depicted with the human H212 and V242 residue positions indicated (red spheres). K_D , dissociation constant. **(G)** Melting curves represent normalized unfolded fractions of unmutated and mutated PU.1 ETS segments from three independent experiments. Average melting temperatures (T_M) are listed, and a dashed red line marks physiological temperature.

S5, B and C). To determine if mutant PU.1 transcriptionally interfered with unmutated PU.1, we transfected reporter cells with 50:50 mixtures of unmutated and mutated *SPI1*. In successfully cotransfected cells, the ability of unmutated PU.1 to drive EGFP expression was not influenced by PU.1 mutant proteins (Fig. 4 B). In contrast, coexpression of an ETS protein fragment significantly reduced unmutated PU.1's ability to drive EGFP expression by presumably competing for λB binding sites (Fig. S5 D; Albrecht et al., 2018). Hence, PU.1-associated *SPI1* mutations encode functionally damaged proteins that do not interfere with unmutated PU.1 but may instead convey deleterious effects via haploinsufficiency.

The PU.1 PEST domain mediates interactions with transcriptional partners (Pongubala and Atchison, 1997); the ETS domain governs nuclear localization and DNA binding (Klemsz et al., 1990; Zhong et al., 2005; Kwok et al., 2007). To determine why H212P, L233Afs*53, and V242G mutated PU.1 proteins failed to drive transcription in reporter cells, we assessed if they could successfully interact with transactivation partners. Consistent

with intact PEST domains, all three proteins coimmunoprecipitated with IRF4 and IRF8 (Fig. S5 E). The PU.1 nuclear localization signal (NLS) lies in the ETS domain and consists of a series of positively charged lysine and arginine residues (Kwok et al., 2007; Zhong et al., 2005). While preserved in H212P and V242G PU.1 mutant proteins, the NLS is obliterated by the L233Afs*53 frameshift. To determine if mutant PU.1 proteins could localize to HEK293 nuclei, we stained them with an N-terminal antibody and determined their cellular distributions using fluorescent microscopy and fractionated immunoblots. As anticipated, H212P and V242G PU.1 proteins were primarily nuclear concentrated, whereas L233Afs*53 mutant was more cytoplasmically relegated (Fig. 4 C and Fig. S5 F).

The PU.1 ETS crystal structure is a winged helix-turn-helix motif containing three α-helices and a four-stranded β-sheet (Fig. 1 B; Kodandapani et al., 1996). In the structure, α3, the “wing” between β3 and β4 and the loop between α2 and α3 directly contact DNA. Although all three points of contact are disturbed/destroyed in L233Afs*53 PU.1, each is retained in

H212P and V242G mutants. Despite this, both H212P and V242G proteins failed to bind PU.1 motif-containing DNA probes in electrophoretic mobility shift assays (Fig. 4 D). To more precisely quantify altered DNA binding, we calculated DNA dissociation constants (K_d) of mutated and unmutated PU.1 ETS segments that comprised amino acids 165–270 (Δ N165). K_d values for H212P and V242G Δ N165 segments were 100-fold and 10-fold higher, respectively, than unmutated ones (Fig. 4 E). We speculate that mutation of His-212 to Pro might break the α 2 helix and perturb the neighboring α 2- α 3 loop and that mutation of Val-242 to Gly might weaken the β 3 hydrophobic core and the α 3 helix stacking against it (Fig. 4 F). To determine if H212P and V242G variants destabilize ETS tertiary structure, we generated melting curves of mutated and unmutated PU.1 Δ N165 segments. Even above 37°C, unmutated segments were stable, whereas H212P and V242G mutated segments became completely unfolded below physiological temperatures (Fig. 4 G). Structural instability may explain why full-length H212P and V242G proteins fail to bind cognate DNA and why these proteins are quantitatively diminished in transfected HEK293 cells relative to unmutated PU.1 (Fig. 4 A). Hence, *SPII* mutations that decay mRNA transcripts, destabilize PU.1 proteins, interfere with PU.1 nuclear localization, and/or alter PU.1 ETS domain-DNA binding cause PU.MA.

PU.1 loss decreases chromatin accessibility and undermines pro-B cell gene expression patterns

To better understand the epigenetic effects of PU.1 haploinsufficiency on pro-B cells, we employed assay for transposase-accessible chromatin using sequencing (ATAC-seq) to generate genome-wide chromatin accessibility maps of human RS4;11 pro-B cell lines either unedited (*SPII*^{+/+}) or CRISPR/Cas9 edited with monoallelic (*SPII*^{-/-}) or biallelic (*SPII*^{-/-}) *SPII* exon one indels that abolished PU.1 expression. Among lines, we identified 189,774 open chromatin regions (OCRs), and of these, 15,878 contained one or more PU.1-binding motifs (Kodandapani et al., 1996). PU.1 OCRs were significantly more accessible in *SPII*^{+/+} than *SPII*^{+/-} or *SPII*^{-/-} cell lines (2.9 versus 2.82 or versus 2.56 log₂ fragments per kilobase of transcript per million mapped reads (FPKM); false discovery rate (FDR) adjusted $P < 1.5e^{-102}$ and $P = \text{zero}$; Fig. 5 A). OCRs containing the motifs of nonpioneer TFs, such as IRF4, IRF8, and TCF3, were also most accessible in *SPII*^{+/+} cell lines. As expected by PU.1's pioneer status, accessibility differences were accentuated in IRF4, IRF8, and TCF3 OCRs that also contained a PU.1-binding motif and were blunted in OCRs that did not (Fig. 5 A). Among OCRs, we identified 13,171 differentially accessible regions less open (\downarrow DARs) in *SPII*^{+/-} cells than in unedited lines. Epigenetic changes between lines appeared concentrated at PU.1-binding sites since \downarrow DARs overlapped with ENCODE human lymphoblast PU.1 chromatin immunoprecipitation followed by sequencing (ChIP-seq) peaks with significantly greater frequency than would be expected by chance (15.6% versus 9.9%; $P = 1.43 e^{-111}$). Further, we found the PU.1 binding site in 1,931 \downarrow DARs, the most significant enrichment of any TF motif (FDR adjusted $P = \text{zero}$; Fig. 5 B). Motifs of other TFs, including many critical to B cell development (ETS1, RUNX1, RUNX2, EBF1, SPIB, and GABPA) and/or already implicated in

human primary immunodeficiencies (TCF3, IRF3, IRF4, and IRF8; Hambleton et al., 2011; Boisson et al., 2013; Andersen et al., 2015; Bravo García-Morato et al., 2018; Tangye et al., 2020), were also significantly overrepresented. Hence, PU.1 haploinsufficiency specifically limits the access of nonpioneer TFs, including those relevant to B cell development and/or associated with primary immunodeficiencies, to pro-B cell open chromatin.

To determine how chromatin accessibility changes associated with PU.1 loss affected gene expression, we transcriptionally profiled *SPII*^{+/+} and *SPII*^{+/-} RS4;11 lines with RNA sequencing (RNA-seq). The DEG set between lines was significantly enriched for genes differentially expressed between human pro- and pre-B cells (Hystad et al., 2007) and also genes differentially expressed between HD and patient A.II.2 single pro-B cells ($P = 0.009$ and 0.001 ; Fig. 5 C). Leading edge transcripts shared by all three gene sets included key down-regulated (*IGLL5*, *TCLIA*, *BTGI*, *CD37*, and *CD79B*) and up-regulated (*DNMT1*) transcripts of biological interest. To determine how changes in chromatin access affected transcription, we assigned each RS4;11-expressed gene to its nearest OCR. Changes in gene expression between *SPII*^{+/+} and *SPII*^{+/-} lines generally correlated with OCR accessibility changes ($R = 0.3$; $P = \text{zero}$; Fig. 5 D). The gene-OCR pairs that correlated perfectly most often (Spearman's correlation = 1) were ones with OCR peaks within 2,000 upstream bases of a transcription start site (TSS; 25% versus 21%; $P = 2e^{-32}$; Fig. 5 E). This region typically contains the gene promoter. Although the promoter regions of several RS4;11-line DEGs, including *TCLIA*, *CD79B*, *CD37*, and *BTGI*, contained either a PU.1-binding motif and/or overlapped an ENCODE PU.1 ChIP-seq peak, 68% of these gene promoter regions did not. Hence, the effects of PU.1 loss on human pro-B cell lines include reduced gene promoter access and cascading transcriptional changes that undermine the pro-B to pre-B cell transition.

Discussion

Our findings describe a previously unreported, monogenic immunodeficiency disease caused by PU.1 haploinsufficiency. The clearest pathological feature of PU.MA, B cell developmental arrest, clarifies conflicting mouse data generated with different PU.1 dose-reduction techniques. With retroviral gene transfer, low PU.1 expression favored B cell, not myeloid, development (DeKoter and Singh, 2000; Houston et al., 2007; DeKoter et al., 2007a), whereas B cell development arrested in mice carrying low-expressing *SPII* alleles (Heinz et al., 2013). Here, we describe six patients with heterozygous *SPII* alleles encoding destabilized, cytoplasmically relegated, and/or transcriptionally inert PU.1 proteins. The patients' immunophenotype provides compelling evidence that the human developmental time point most sensitive to reduced PU.1 dose is the pro- to pre-B cell transition when there is rising demand for PU.1-dependent chromatin accessibility and for PU.1-regulated gene products such as BTK, CD79B, *TCLIA*, *CD37*, and *BTGI*. Although five patients in our cohort presented with agammaglobulinemia in their first year of life, the serum IgG and tetanus titer detected in patient C.II.1 as a young adult indicates delayed disease onset is also possible. Why B cell development would arrest early in life versus later is currently unclear but could conceivably be explained by variability in the complex, cooperative interactions between PU.1 and other TFs.

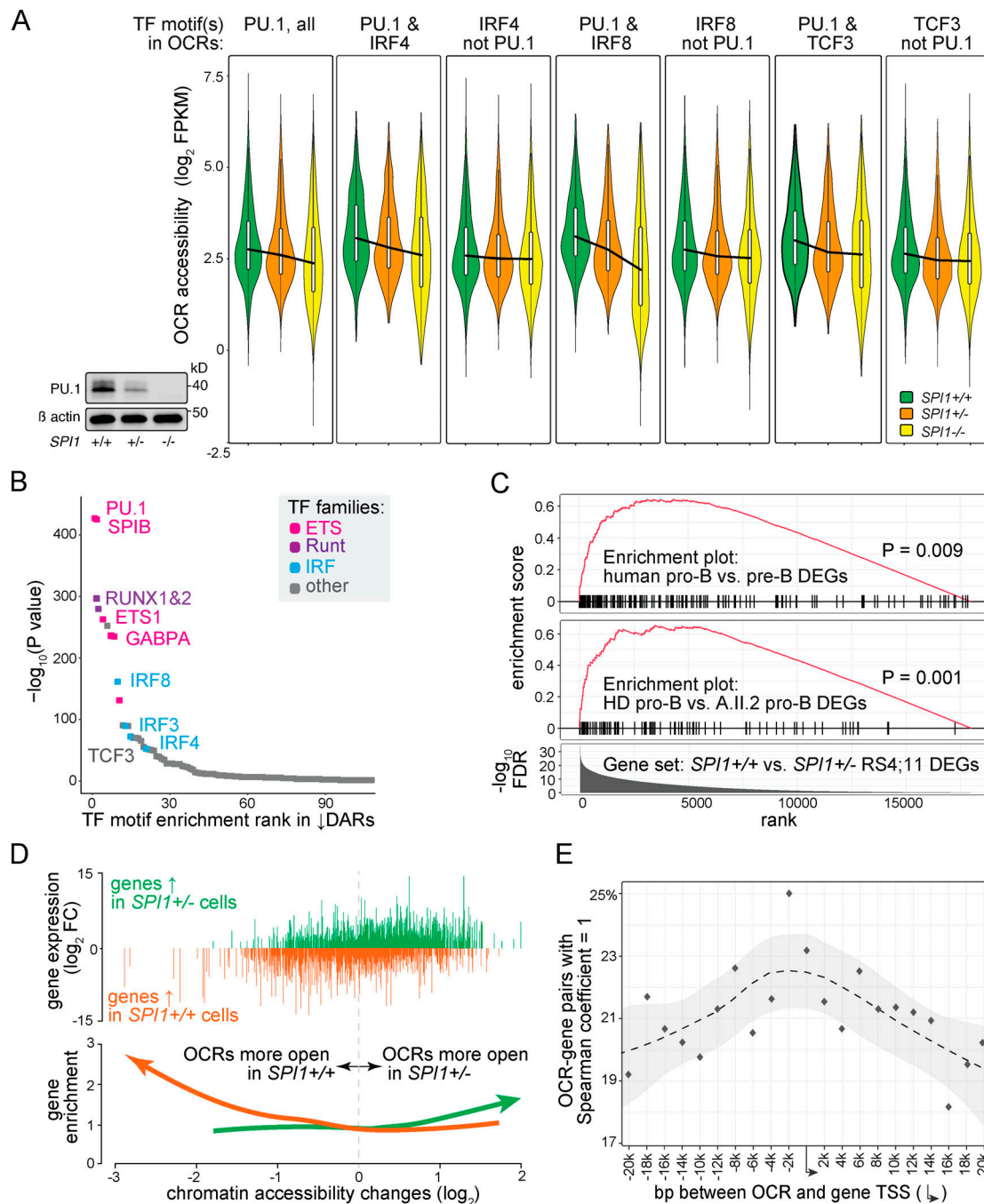


Figure 5. PU.1 haploinsufficiency constrains chromatin accessibility for nonpioneer TFs. (A) Accessibility, measured in FPKM, of OCRs containing the indicated TF motifs or TF motif pairs, is represented for *SPI1* $+/+$ (dark green), *SPI1* $+/-$ (orange), and *SPI1* $-/-$ (yellow) RS4;11 pro-B cells. Thick black lines show average OCR accessibility trends across lines. An immunoblot shows PU.1 expression in each line. **(B)** Binding motifs of TFs and TF families are ranked by decreasing enrichment in chromatin regions differentially less accessible (\downarrow DARs) in *SPI1* $+/-$ RS4;11 cells compared with unedited *SPI1* $+/+$ cells. **(C)** Enrichment plots show published sets of DEGs between human pro-B and pre-B cells (top) or DEGs between HD and PU.MA patient A.II.2 pro-B cells (middle) that overlap with DEGs between *SPI1* $+/+$ and *SPI1* $+/-$ cells that were ranked by decreasing $-\log_{10}$ FDR (bottom). **(D)** Top: \log_2 FCs of DEGs between *SPI1* $+/+$ and *SPI1* $+/-$ cells are ranked by decreasing accessibility of their nearest, paired OCR. A dashed zero-line marks OCRs equally accessible in either cell line. Bottom: A smoothed track indicates FC gene expression differences between *SPI1* $+/+$ (orange) and *SPI1* $+/-$ (green) cells divided by the global-FC average. **(E)** The proportion of OCR-gene pairs with perfect correlations between differential accessibility and expression (Spearman's coefficient = 1) are displayed relative to the bp distances between each pair's OCR peak and gene TSS. Gray shading indicates the 95% confidence interval. Data generated from three independent experiments. GSEA P values were calculated using permutation tests.

The sinopulmonary, invasive bacterial and systemic enteroviral infections observed in PU.MA patients are similar to the infectious illnesses experienced by patients with other forms of congenital agammaglobulinemia. Unlike PU.1 low-expressing mice (Rosenbauer et al., 2004; Will et al., 2015), no PU.MA patients have so far developed AML or any other malignancies. We hope this encouraging trend persists but must acknowledge that PU.1 haploinsufficiency does appear to alter the myeloid lineage, specifically circulating cDCs. Like PU.MA patients, PU.1-haploinsufficient mice also selectively lack cDCs due to decreased expression of FLT3L and DC-SCRIPT (Carotta et al., 2010; Chopin et al., 2019). Interestingly, IRF8-deficient patients also lack cDCs but normally generate B cells, whereas agammaglobulinemic patients with dominant negative *TCF3* mutations exhibit the opposite cellular profile (Hambleton et al., 2011; Boisson et al., 2013). The PU.MA immunophenotype, which straddles features of both disorders, illustrates how loss of a pioneer TF may phenotypically mimic primary deficiencies of nonpioneer TFs that rely upon pioneers for chromatin access.

Despite numerical and functional cDC deficiencies, the patients described here did not experience mycobacterial infections despite BCG vaccine exposures. Their apparent ability to clear mycobacteria may differentiate PU.MA patients from cDC-deficient patients with *IRF8*, *SPPL2A*, and *GATA2* mutations, who have been described as exquisitely susceptible to mycobacterial infections (Tangye et al., 2020). As IRF8-deficient patients also lack monocytes (Hambleton et al., 2011), *SPPL2A*-deficient mycobacteria-specific CD4⁺ T cells fail to secrete IFN- γ (Kong et al., 2018), and *GATA2*-deficient patients experience multilineage cytopenias (cDC, pDC, monocyte, NK cells, and B cells; Bigley et al., 2011; Hsu et al., 2011), it may be that cDCs are dispensable for mycobacterial clearance if non-cDC myeloid and T cell populations can be preserved. Nonetheless, until safety can be more conclusively studied, we recommend PU.MA patients do not receive live immunizations such as BCG and should absolutely avoid oral polio vaccination.

PU.MA is not the first inborn error of a pioneer/pioneer-like TF. Although the neuronal (HASH-1; de Pontual et al., 2003), cardiac (*GATA4*, *GATA6*; Garg et al., 2003; Kodo et al., 2009), pancreatic islet (*NEUROD1*; Malecki et al., 1999), eye (*SOX2*; Kelberman et al., 2006), muscle (*PAX7*; Feichtinger et al., 2019), inner ear (*ESRRB*; Collin et al., 2008), erythroid/megakaryocytic (*GATA1*; Nichols et al., 2000), and now immune (PU.1) pathologies of pioneer TF-associated diseases appear superficially unrelated, they may share constrained promoter region accessibility as a common underlying epigenetic disease mechanism. Notably, most disease-associated pioneer TFs, including PU.1, exhibit tissue-specific haploinsufficiency. Such exquisite sensitivity to reduced pioneer TF dose highlights the delicate regulation and critical role of euchromatin access during organogenesis.

Materials and methods

Human samples

Between 2012 and 2020, 30 agammaglobulinemic patients without molecular diagnoses and 30 age-matched healthy controls (range, 5–42 yr) were selected from a multi-institution

cohort for genomic analysis. 19 of 30 agammaglobulinemic patients were male (5 were PU.MA patients). 16 were white non-Hispanic (4 were PU.MA patients), 8 were white Hispanic (1 was a PU.MA patient), 3 were African American non-Hispanic (1 was a PU.MA patient), and 3 were Asian. Patient clinical information was abstracted from medical records. Blood samples were obtained by standard phlebotomy. Bone marrow aspirates were collected as surgical waste remaining after clinically indicated procedures. A malignancy-free, age-matched bone marrow biopsy obtained for cancer relapse monitoring was used as a control for IHC staining. Blood and bone marrow donors provided written informed consent in accordance with research protocols approved by institutional review boards of Children's Hospital of Philadelphia, Baylor College of Medicine, Medical College of Wisconsin, University of Alabama Birmingham, or Dmitry Rogachev National Medical Research Center.

Three de-identified human cord blood samples were obtained from StemExpress.

SPI1 variant sequencing

SPI1 variants were identified by whole-exome sequencing data performed on either a clinical or research basis. For research samples, whole-exome sequencing was performed by the Human Genome Sequencing Center (HGSC) at Baylor College of Medicine through the Baylor-Hopkins Center for Mendelian Genomics initiative. With 1 μ g of genomic DNA, an Illumina paired-end precapture library was constructed according to the manufacturer's protocol with modifications as described in the Baylor College of Medicine-HGSC Illumina Barcoded Paired-End Capture Library Preparation protocol. Precapture libraries were pooled into four-plex library pools and then hybridized in solution to the HGSC-designed Core capture reagent (52 Mb; NimbleGen), or six-plex library pools used the custom VCRome 2.1 capture reagent (42 Mb; NimbleGen) according to the manufacturer's protocol (NimbleGen SeqCap EZ Exome Library SR User's Guide) with minor revisions. The sequencing run was performed in paired-end mode using the Illumina HiSeq 2000 platform, with sequencing-by-synthesis reactions extended for 101 cycles from each end and an additional 7 cycles for the index read. With a sequencing yield of >9 Gb, samples achieved >90% of the targeted exome bases covered to a depth of 20 \times or greater. Illumina sequence analysis was performed using the HGSC Mercury analysis pipeline, which moves data through various analysis tools from the initial sequence generation on the instrument to annotated variant calls (single-nucleotide polymorphisms and intra-read in/dels). Annotated variants were then prioritized according to the American College of Medical Genetics and Genomics guidelines. Evolutionary conservation at mutated sites was determined by Aminode query (Chang et al., 2018).

All identified *SPI1* variants were annotated to cDNA transcript GenBank accession no. NM_001080547.2 and PU.1 protein isoform one (GenBank accession no. NP_003111.2) after Sanger sequencing confirmation of peripheral blood and/or buccal genomic DNA. Copy number variation was assessed and excluded with clinical chromosomal microarrays. When available, *SPI1* variant phasing was performed on parental buccal DNA samples.

PBMC immunophenotyping

Agammaglobulinemic and age-/sex-matched healthy patient mononuclear cells were isolated from PBMC and bone marrow aspirates using Ficoll-Paque PLUS density gradient centrifugation (GE Healthcare Life Sciences). For neutrophil studies, whole blood was incubated at room temperature for 10 min in red blood cell lysing solution (BD Bioscience). Cells were immunophenotyped using an LSRI Fortessa flow cytometer (BD Bioscience) with antibodies against CD3 (OKT3), CD4 (OKT4), CD8 (HIT8a), CD11b (ICRF44), CD11c (3.9), CD14 (HDC14), CD16 (B73.1), CD19 (HIB19), CD20 (2H7), CD25(BC96), CD34 (561), CD45 (2D1), CD45RO (UCHL1), CD56 (HCD56), CD117 (104D2), CD127 (A019D5), CD303 (201A), CCR6 (11A9), CRTH2 (BM16), CXCR3 (G025H7), CXCR5 (J252D4), FOXP3 (150D), IL-6 (MQ2-13A5), IL-12 (C11.5), PD1 (EH12.2H7), and PU.1 (7C6B05) all from BioLegend; IgM (G20-127) and HLA-DR (G46-6) from BD Bioscience; and TdT (TdT6; Invitrogen). FOXP3, IL-6, IL12, PU.1, and TdT intracellular staining was performed after fixation and permeabilization with the Foxp3/Transcription Factor Staining Buffer Set (eBioscience) in accordance with the manufacturer's instructions. Dead cells were excluded using LIVE/DEAD stain (Invitrogen). To analyze B cell precursors, we identified pro- (CD34⁺CD20⁻), pre-B1 (CD34⁻CD20⁻), and pre-B2 cells (CD34⁻CD20⁺) among CD19⁺IgM⁻ marrow mononuclear cells. Flow cytometric analyses were visualized with FlowJo software (TreeStar).

In vitro identification of IL-6 and IL-12 secreting myeloid cells

To assess IL-6 and IL-12 secretion by myeloid cells, PBMCs were rested overnight at 37°C and then stimulated for 6 h with either gardiquimod (10 µg/ml; Invivogen) or LPS (10 ng/ml; Sigma) in the presence of Brefeldin A (5 µg/ml; BD Bioscience). After activation, cells were incubated with a surface antibody cocktail for 30 min at 4°C, LIVE/DEAD stained, fixed, permeabilized, and intracellularly stained with anti-IL-6 and anti-IL-12 antibodies as described above.

CITE-seq

Viable marrow mononuclear cells were resuspended at 20 × 10⁶ cells/ml in CITE-seq staining buffer (BioLegend) and incubated with Human TruStain FcX Fc Blocking reagent for 10 min at 4°C to block nonspecific antibody binding. Following Fc blocking, cells were incubated with a pool of 140 antibodies conjugated to an antibody-derived tag (ADT), inclusive of eight isotype controls (TotalSeq-C Human Universal Cocktail; BioLegend) for 30 min at 4°C. After incubation, cells were washed three times with 3.5 ml of staining buffer to remove antibody excess. Cells were passed through a 40-µm filter to remove any cell clumps and resuspended in 10% FBS RPMI media at 1 × 10⁶ cells/ml for 10× Genomics 5' single-cell RNA-seq. Next-generation sequencing libraries were prepared using the 10× Genomics Chromium Single Cell 5' Library and Gel Bead kit v1 with Feature Barcoding Technology for Cell Surface Protein per manufacturer's "Chromium Single Cell V(D)J Reagent Kits" protocol. Libraries were uniquely indexed using the Single Index Kit N set A and sequenced on the Illumina NovaSeq 6000 sequencer (v1.5 chemistry) in a paired-end, single indexing run. Sequencing for each library targeted either 20,000 mean reads per cell for the 5'

chemistry gene expression libraries or 5,000 read pairs per cell for the cell surface/feature barcoding libraries. Data were processed using the cellranger count pipeline (10× Genomics, v.3.1) for demultiplexing, alignment of sequencing reads to the transcriptome (10× Genomics, human GRCh38) or antibody barcodes for the antibody capture feature barcode libraries, and creation of feature-barcode matrices. Secondary analysis on the feature barcode matrices was performed using the Seurat package (v.4.0) within the R computing environment. Count matrices for CITE-seq expression were adjusted based on isotype-specific control expression on a per cell basis. Briefly, raw unique molecular identifier counts of the isotype control surface marker were subtracted from the unique molecular identifier count of each ADT surface-expressed marker within a given cell. These altered count matrices were then read into the R computing environment and normalized via the CLR method in Seurat. Variable features, normalization, data scaling, and principal components were determined for each assay separately (ADT and RNA) followed by hierarchical clustering and visualization using UMAP. The likely identities of UMAP clustered cells were determined using the expression of key cell-type-specific proteins and transcripts (Fig. S3). DEGs between groups were determined using a Mann-Whitney *U* test. To determine transcriptional differences in genes known to be associated with other forms of congenital agammaglobulinemia, *P* value adjustment was performed using a Bonferroni correction based on the total number of such genes as determined by International Union of Immunological Societies (Tangye et al., 2020).

IHC

Bone marrow biopsies were IHC stained with mouse anti-human CD79a (1:200 dilution; Dako), CD20 (1:1,000; Dako), CD138 (1:200; Dako), and TdT (1:50; Leica). IHC images were captured on a BX53 (Olympus).

HSPC cultures

HSPC growth medium was prepared with X-Vivo15 (Lonza) with gentamicin and additional 1% human serum albumin (Grifols) and additional cytokines (Peprotech): stem cell factor (SCF) at 100 ng/ml, thrombopoietin at 100 ng/ml, Flt3 at 100 ng/ml, and IL-3 at 20 ng/ml. Human primary cord blood CD34⁺ hematopoietic stem cells were grown at an initial concentration of 250,000 cells/ml for 48 h before editing.

Cas9 ribonucleoprotein (RNP) formulation

Cas9 and RNPs were prepared immediately before electroporation. Synthetic CRISPR RNA and trans-activating CRISPR RNA were synthesized (Dharmacon Horizon), resuspended in duplex buffer (IDT) at a concentration of 160 µM, and stored at -80°C. gRNA was formulated mixing CRISPR RNA and trans-activating CRISPR RNA 1:1 by volume and annealing at 37°C for 30 min. 15–50 kD poly-L-glutamic acid (Sigma) was resuspended in water to a concentration of 100 mg/ml, then mixed with the gRNA at a volume ratio of 0.8:1 before complexing. Cas9-NLS (University of California Berkeley; QB3 Macrolab) at 40 µM was added to gRNA/poly-L-glutamic acid at 1:1.8 volume ratio for the

desired molar ratio of gRNA:Cas9-NLS of 2:1 and a final RNP concentration of 20 μ M.

Electroporation

Immediately before electroporation in a Lonza 4D 96-well format Nucleofector (Lonza), HSPCs were centrifuged for 10 min at 90 \times g, the media was aspirated, and cells were resuspended at a concentration of 100,000 cells per 18.5 μ l of electroporation buffer P3 (Lonza). For each condition, 50 pmol of RNP was added to 100,000 cells, mixed briefly, and then electroporated with pulse code DS-137. Immediately following electroporation, cells were rescued by adding HSPC growth media directly into each electroporation well and then incubated at 37°C for 15 min. After rescue, cells were diluted to 250,000 cells/ml in HSPC growth media and incubated at 37°C for an additional 48 h. A 10,000-cell aliquot was reserved for predifferentiation sequencing, and the remaining cells were equally divided for co-culture on OP9 or OP9-DL1 mouse stromal cells.

Single-cell *SPI1* genotyping

To perform single-cell genomic DNA analysis of CRISPR-edited alleles, 100,000 edited primary HSPCs from a cord blood donor were encapsulated on a single-cell DNA genotyping platform (Mission Bio; Tapestry; Pellegrino et al., 2018) according to the manufacturer's protocol to perform lysis, single-cell barcoding, and targeted amplicon generation for a predesigned panel of genes (Demaree et al., 2021) modified to include primers specific to *SPI1* exon 4 and exon 5. Recovery and cleanup of single-cell libraries proceeded according to the Mission Bio V2 protocol. All DNA libraries were run on a Bioanalyzer High Sensitivity DNA electrophoresis chip (Agilent Technologies) to verify complete removal of primer-dimer products. Libraries were quantified by fluorometer (Qubit 3.0; Invitrogen) and sequenced on an Illumina NextSeq 300 cycle Mid-Output Kit with a 20% spike-in of PhiX control DNA (Illumina) to generate paired-end 150-bp reads.

Sequencing data were processed using a custom pipeline (github/AbateLab/scCRISPR; Demaree et al., 2021). Reads were demultiplexed by cell barcode to create single-cell FASTQ files for further processing. Amplicon sequence reads were aligned and processed with the CRISPResso2 algorithm (Clement et al., 2019) to quantify modifications (insertions and deletions) of the target region. Cells with a minimum of 15 aligned reads at the edited locus were selected for further analysis, and analysis only included cells with 75% or more of the reads contributed by one or two alleles, thereby excluding barcodes containing mixed alleles arising from microfluidic cell multiplets. For each cell passing filtering, the top two alleles were classified by coding sequence length as WT (including possible substitutions), in-frame, or frameshift mutations.

OP9 and OP9-DL1 co-culture with edited HSPCs

OP9/DL1 medium was prepared by reconstituting α -Minimum Essential Media (Gibco; α -MEM) in diH₂O with 2.2 g/liter of sodium bicarbonate, 20% FBS (Sigma), 50 μ g/ml penicillin-streptomycin (Thermo Fisher Scientific), and 2 mM L-glutamine

(Thermo Fisher Scientific). OP9 and OP9-DL1 (La Motte-Mohs et al., 2005) were cultured in 1.5 ml of media in six-well plates and maintained at max ~80% confluency by passaging every 2–3 d. For co-culture, 30,000 OP9 or 50,000 OP9-DL1 per well was seeded in 1.5-ml media in six-well plates with additional IL-7 (5 ng/ μ l), Flt3 (5 ng/ μ l), and SCF (30 ng/ μ l) for 24 h; then edited HSPCs were directly added. Differentiating HSPCs were replated on freshly maintained OP9/DL1 cells every 7 d. To plate HSPCs on fresh stroma, medium was gently pipetted to loosen cells, then filtered through a 40- μ m mesh into round-bottom tubes. For the first 2 wk of differentiation, stroma was trypsinized and filtered into the round-bottom tubes. After week two, trypsin was no longer used. Cells were spun at 1,500 rpm for 5 min and resuspended in fresh OP9/DL1 media containing IL-7, Flt3, and SCF. All cells were then plated with freshly seeded stroma on six-well plates, with additional wells at a 1:2 split as needed to prevent confluency.

Fluorescence-activated cell sorting of differentiated cells

Developing B lymphocytes and myeloid lineage cells were maintained for 5 wk on OP9 co-culture, then harvested and stained for B cell and myeloid markers with labeled antibodies against CD19 (HIB19), CD20 (2H7), CD33 (WM53), and IgM (MHM-38), all from BioLegend, and CD45 (HI30; BD-Bioscience). Developing T lymphocytes were maintained on OP9-DL1 co-culture for 6.5 wk, then harvested and stained for T cell markers with labeled antibodies against CD3 (UCHT1), CD4 (OKT4), and CD8 (HIT8a), all from BioLegend. Cell populations were sorted on a BD FACS Aria (BD Bioscience) with a 100- μ m nozzle. Sorted cells were washed in 50% FCS in PBS, pelleted, and frozen for later processing.

Amplicon sequencing to determine editing efficiency

Genomic DNA was extracted from cell pellets using the Zymo Quick-Extract DNA MiniPrep kit (Zymo Research). Extracted DNA was purified by SPRI bead cleanup and resuspended in 20 μ l water. A two-step PCR amplicon sequencing approach was used to amplify and index samples before next-generation sequencing. For PCR, 1.8 μ l of clean genomic DNA (50 μ l total reaction) was amplified with touchdown-PCR conditions for 30 total cycles with KapaHiFi polymerase (Kapa Biosystems) using validated primers targeting an ~100–150-bp region around the predicted cut sites. After an 0.8 \times SPRI bead cleanup, a second PCR was then performed with NEB Q5 2 \times Master Mix Hot Start Polymerase for 10 cycles to append P5 and P7 Illumina Nextera sequencing adaptors and sample-specific indices, followed again by a 1.0 \times SPRI bead cleanup. Concentrations were normalized across samples and pooled, and the amplicon library was sequenced with a 15% spike-in of PhiX control DNA on an Illumina MiniSeq to obtain paired-end 150 base reads. For samples with <50,000 aligned amplicon reads, PCR 1 and PCR 2 were redone, and sequencing was performed on a HiSeq with single-end 100 base reads. Amplicon sequence reads were aligned and processed with the CRISPResso2 algorithm in batch mode (Clement et al., 2019) to analyze editing efficiency and quantify modifications (insertions and deletions) of the target region in each sample.

Plasmid construction

Plasmids used to construct PU.1 reporter cells included (1) an EGFP-based reporter controlled by a minimal promoter downstream of a synthetic λ B-based enhancer and (2) a pcDNA3.1(+) construction containing, under the promoter CMV (pCMV), the sequence *SPII* fused to either infrared RFP (iRFP) or mCherry genes via self-cleaving 2A peptide sequence (Lek et al., 2016). *SPII* mutations were introduced into the pCMV-iRFP-2A-*SPII* (corresponding to isoform one) plasmid by site-directed mutagenesis using the Q5 Site-Directed Mutagenesis kit (NEB). Each *SPII* plasmid sequence was validated by Sanger sequencing before transfection.

PU.1 reporter assay

For single *SPII* plasmid transfections, mutated or unmutated pCMV-iRFP-2A-*SPII* (100 ng) were Lipofectamine (Invitrogen) co-transfected with λ B-promoter/enhancer-EGFP (Munde et al., 2014; 10 ng) into 2×10^5 HEK293 cells, a line lacking native PU.1 expression. After 48 h, cells were stained for live/dead cell discrimination using a LIVE/DEAD kit (Invitrogen). The frequencies of live iRFP⁺ cells were acquired by a flow cytometer. For transfections using two *SPII* plasmids, mutated or unmutated pCMV-iRFP-2A-*SPII* (50 ng) and unmutated pCMV-mCherry-2A-*SPII* (50 ng) were cotransfected with λ B-promoter/enhancer-EGFP (10 ng) into 2×10^5 HEK293 cells. A pCMV-*SPII*_ETS plasmid (Albrecht et al., 2018) encoding a PU.1 ETS fragment, was used as a dominant negative control. The frequencies of live mCherry⁺ and/or mCherry⁺iRFP⁺ cells were measured after 48 h.

Immunoblots

pCMV-iRFP-2A-*SPII* (500 ng) was transfected into 5×10^5 HEK293 cells with Lipofectamine. After 48 h, live iRFP⁺ cells were sorted with a MoFlo Astrios EQ (Beckman Coulter) and then lysed for 10 min with ice-cold RIPA buffer (150 mM NaCl, 5 mM EDTA, 50 mM Tris, pH 8, 1% IGEPAL, 0.5% sodium deoxycholate, and 0.1% SDS) supplemented with protease inhibitor. Cell lysates were centrifuged at $8,000 \times g$ for 5 min. After quantification by Bradford testing, 7 μ g of proteins were incubated with 4 \times Laemmli sample buffer containing 2- β -mercaptoethanol at 95°C for 5 min. Proteins were separated by 4–15% SDS-PAGE gel and transferred to PVDF membranes using the trans-blot turbo transfer system (all from Bio-Rad) and probed with a rabbit anti-PU.1 antibody recognizing an epitope within the first 100 N-terminal amino acids (EPR3158Y; Abcam) or a rabbit anti- β actin antibody (13E5; Cell Signaling). Proteins were detected by chemiluminescence using species-specific HRP-linked antibodies (Cell Signaling). Immunoblots were visualized with a ChemidocMP imaging system (Bio-Rad) and quantified with ImageJ software (National Institutes of Health freeware).

To determine PU.1 localization, live iRFP⁺ HEK293 cells were sorted as described above, then incubated for 15 min on ice-cold hypotonic buffer (20 mM Tris-HCl, 10 mM NaCl, 3 mM MgCl₂, and 10% IGEPAL) before centrifuging at $8,000 \times g$ for 10 min. Cytoplasmic and nuclear protein fractions were extracted after a 30-min incubation in ice-cold RIPA lysis buffer. Gel separation and transfer of proteins were performed as described above. Rabbit anti-GAPDH (D16H11; cell signaling) and mouse anti-

Lamin A/C (E-1; Santa Cruz) antibodies were respectively used as cytoplasmic and nuclear controls.

PU.1 expression was also semi-quantified on PU.MA PBMCs and compared with HD PBMCs in which B cells were depleted via CD20 microbead depletion (Miltenyi). Undepleted cell lysates were normalized and immunoblotted as described.

Co-immunoprecipitation

HEK293 cells were transfected with pCMV-iRFP-2A-*SPII* and either pCMV-IRF4-HA or pCMV-IRF8-FLAG (GeneCopia). Following 48 h, cells were lysed with Reliablots lysis buffer (250 mM NaCl, 50 mM Tris, pH 8, 5 mM EDTA, and 0.5% IGEPAL; Bethyl Labs). 200 μ g of lysate proteins were incubated with either mouse anti-hemagglutinin (HA) beads or rabbit anti-Flag beads (Sigma) overnight at 4°C and then incubated with RIPA lysis buffer before gel separation and transfer as described above. Rabbit anti-Flag (D6W5B; Cell Signaling) and rabbit anti-HA (C29F4; Cell Signaling) antibodies were used to confirm sample loading.

Confocal microscopy

HEK293 cells (0.4×10^5) seeded onto Lab-Tek (Nunc) were transfected with mutated or unmutated pCMV-iRFP-2A-*SPII*. After 48 h, cells were washed with PBS, fixed for 15 min in 4% paraformaldehyde, and permeabilized for 15 min in 0.1% Triton X-100 in PBS at room temperature. After 30 min in blocking buffer (PBS 0.1% Triton X-100 with 2% BSA), cells were incubated for 2 h with a rabbit anti-PU.1 monoclonal antibody (EPR3158Y; Abcam), washed again, and incubated for 1 h in blocking buffer containing goat anti-rabbit Alexa Fluor 488 antibody (Invitrogen). After a final wash, cells were DAPI stained and mounted on slides using Fluorescence Mounting Medium (Dako). Images were obtained on a Leica TCS SP8 inverted confocal microscope with a 63 \times oil immersion objective (Leica Microsystems). Images were analyzed with ImageJ software.

Electrophoretic mobility shift assay

PU.1/probe DNA interactions were assessed using gelshift chemiluminescence (ActiveMotif) in accordance with manufacturer's instructions. Nuclear extracts from live, iRFP⁺ HEK293 cells transfected with WT or mutated pCMV-iRFP-2A-*SPII* plasmids were generated as described. WT, H212P, and V242G PU.1 proteins were semi-quantified by immunoblot to enable normalized PU.1 loading. As G109Sfs*78 and Y122X PU.1 proteins were undetectable in nuclear lysates, we included them in blots as additional negative controls. Nuclear extracts were incubated for 10 min at room temperature with 20 fmol biotin-PU.1 probe DNA in the presence of poly(dI-dC) (50 ng/ μ l). DNA-protein complexes were separated on a 6% retardation gel (Thermo Fisher Scientific) with 0.5 \times Tris/Borate/EDTA solution. The nucleotide sequences of PU.1 probes were 5'bio-AAAAAGAGGAAGTGAAC-3' and 5'-GTTTCACTTCTCTTTT-3' (IDT).

Production and purification of ETS protein fragments

Codon-optimized DNA fragments encoding various point mutants of the ETS domains (human number 165–270, termed Δ N165; Xhani et al., 2020) were synthesized by IDT DNA Technologies. The fragments were subcloned into NcoI/HindIII sites

of pCDF-1b vector (Novagen) for overexpression in *Escherichia coli*. All constructs were verified by Sanger sequencing.

Heterologous overexpression in BL21(DE3)pLysS *E. coli* was performed as previously described (Esaki et al., 2017). In brief, expression cultures in lysogeny broth media were induced at an OD₆₀₀ of 0.6 with 0.5 mM isopropyl β-D-1-thiogalactopyranoside for 4 h at 25°C. Harvested cells were lysed by sonication in 0.1 M TrisHCl, pH 7.4, containing 0.5 M NaCl and 1 mM PMSF. After centrifugation, cleared lysate was extracted with Co-NTA and eluted in up to 15 ml of elution buffer containing 150 mM imidazole. The eluate was dialyzed overnight in the presence of 10 U of thrombin (MPBio) against 10 mM NaH₂PO₄/Na₂HPO₄, pH 7.4, containing 0.5 M NaCl and polished on Sepharose SP (HiTrap; GE Healthcare). After extensive washing in this buffer, the protein was eluted in an NaCl gradient at ~1 M. Purified protein was dialyzed extensively and diluted as needed with dialysate. Protein concentrations were determined by UV absorption at 280 nm using the extinction coefficient 22,460 M⁻¹cm⁻¹.

Quantitative fluorescence polarization titrations

DNA binding by protein was measured by steady-state anisotropy of a Cy3-labeled DNA probe encoding the optimal PU.1-binding sequence 5'-AGCGGAAGTG-3'. In brief, 0.5 nM of DNA probe was titrated with protein in a 10 mM TrisHCl buffer, pH 7.4, containing 0.1% wt/vol BSA and 0.15 M NaCl. Steady-state anisotropies $\langle r \rangle$ were measured at 595 nm in 384-well black plates (Corning) in a Molecular Dynamics Paradigm plate-reader with 530-nm excitation. The signal represented the fractional bound DNA probe (F_b), scaled by the limiting anisotropies of the i -th bound $\langle r_i \rangle$ and unbound states $\langle r_0 \rangle$ as follows:

$$\langle r \rangle = F_b \left(\sum_{i=1}^n \langle r_i \rangle - \langle r_0 \rangle \right) + \langle r_0 \rangle = F_b \sum_{i=1}^n \Delta r_i + \langle r_0 \rangle, \quad (1)$$

where F_b is a function of total protein concentration as taken. As PU.1 is known to form 1:1 and 2:1 complexes with a single ETS binding site (Esaki et al., 2017), two stepwise K_{d1} s were used to describe the direct binding of the labeled DNA probe by PU.1:

$$K_{d1} = \frac{[P][D]}{[PD]} \quad (2)$$

$$K_{d2} = \frac{[PD][P]}{[P_2D]},$$

where P and D denote PU.1 and DNA. The observed anisotropy change represented the summed contributions of the two complexes as expressed by Eq. 1. The equations are solved numerically in terms of the unbound protein, P. The solution, which is cubic in [P], is

$$0 = \varphi_0 + \varphi_1[P] + \varphi_2[P]^2 + \varphi_3[P]^3, \quad (3)$$

$$\begin{cases} \varphi_0 = K_{d1}K_{d2}[P]_t \\ \varphi_1 = K_{d1}K_{d2} - K_{d2}[P]_t + K_{d2}[D]_t \\ \varphi_2 = K_{d2} - [P]_t + 2[D]_t \\ \varphi_3 = 1 \end{cases}$$

where the subscript “t” represents the total concentration of the referred species. With [P] in hand, [D], [PD], and [P₂D] were computed from Eq. 2 and the corresponding equations of state. Nonlinear least-square regression was performed using Origin (v2019b).

Circular dichroism spectroscopy

Proteins were diluted with dialysate to 10 μM and monitored at 222 nm in 10 mM NaH₂PO₄/Na₂HPO₄, pH 7.4, and 0.15 M NaCl with a Jasco J-810 instrument. The slit width was set at 2 nm. Samples were heated at 0.75°C/min followed by cooling at the same rate to determine reversibility. The circular dichroism signal at each temperature was integrated for 30 s. Under these conditions, ΔN165 exists predominantly as a monomer (Xhani et al., 2020). Melting curves were therefore analyzed according to denaturation of a two-state monomer as previously described to estimate their melting temperatures (Poon et al., 2002).

Three-dimensional PU.1 modeling

To three-dimensionally model patient PU.1 variants, the crystal structure of murine PU.1 bound to duplex DNA was used. This structure has been assigned Protein Data Bank accession no. 1PUE (Kodandapani et al., 1996). Ribbon representation of the interaction was generated with PyMOL v2.2.3 (Schrödinger, LLC).

Generation of SPI1 edited RS4;11 lines

RS4;11 cells were electroporated alone (unedited, *SPI1*^{+/+}) or with Cas9 enzyme and a gRNA sequence (5'-GGGGUGGAAGUCCCAGUAA-3') targeting exon one of *SPI1* (Synthego). Editing efficiency approached 50%. Clonal populations of mono (*SPI1*^{+/-}) and biallelically edited cells (*SPI1*^{-/-}) were obtained via limiting dilution and clonal expansion. Genomic edits were confirmed via Sanger sequencing (data not shown) and with PU.1 immunoblots (Fig. 5 A).

ATAC-seq library preparation and sequencing

100,000-cell aliquots of *SPI1*^{+/+}, *SPI1*^{+/-}, and *SPI1*^{-/-} RS4;11 cells were harvested in triplicate, washed twice, resuspended in 50 μl cold lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl₂, and 0.1% IGEPAL CA-630), and then spun down immediately at 550 ×g for 10 min, 4°C. The supernatant was discarded, and nuclei were immediately resuspended in cold transposition reaction mix from Illumina Tagment DNA Enzyme kit. The transposition reaction was incubated at 37°C for 45 min. Immediately following transposition, 10 μl of 3M Sodium Acetate was added to the DNA, which was then purified using a Qiagen MinElute Kit. Transposed DNA fragments were PCR amplified for 12 cycles using NEBNext High-Fidelity 2× PCR Master Mix with Nextera primers (Illumina). The PCR reaction was cleaned up using AMPureXP beads (Agencourt) and then paired-end sequenced on the Illumina NovaSeq platform (51-bp read length) at the Center for Spatial and Functional Genomics.

RNA-seq library preparation and sequencing

1,000,000-cell aliquots of *SPI1*^{+/+}, *SPI1*^{+/-}, and *SPI1*^{-/-} RS4;11 cells were harvested in triplicate and stored in Trizol. RNA was extracted and purified using Direct-zol RNA Microprep (Zymo

Research). Approximately 600 ng of high-quality RNA was used to prepare each library by using the TruSeq Stranded mRNA Library Prep Kit and the IDT for Illumina-TruSeq RNA UD Indexes (Illumina). Samples were sequenced on the Illumina NovaSeq platform (51-bp read length) at the Center for Spatial and Functional Genomics.

ATAC-seq peak calling and accessibility estimate

ATAC-seq peaks (OCRs) were called independently for each sample using the ENCODE ATAC-seq pipeline (<https://www.encodeproject.org/atac-seq/>). Briefly, pair-end reads from all replicates for each cell type were aligned to hg19 genome using bowtie2, and duplicate reads were removed from the alignment. Aligned tags were generated by offsetting +4 bp for all the reads aligned to the forward strand and -5 bp for all the reads aligned to the reverse strand. Narrow peaks were called independently for pooled replicates for each cell type using MACS2 (Zhang et al., 2008; -p 0.01-nomodel-shift -75-extsize 150 -B-SPMR-keep-dup all-call-summits), and ENCODE blacklist regions (wgEncodeDacMapabilityConsensusExcludable.bed.gz) were removed from called peaks. To improve the reproducibility on peak calling, the OCR sets for each condition were consolidated into one list by union operations on the OCRs present in at least two replicates. This generates a total of 139,424, 98,732, and 122,539 OCRs for *SPII*^{+/+}, *SPII*^{+/-}, and *SPII*^{-/-} samples, respectively. A final consensus of 189,774 OCRs was generated by combining the OCR sets across all the conditions using bedtools intersect (v2.25.0).

To estimate the accessibility in OCRs, the de-duplicated read counts for consensus OCRs were calculated for each replicate. Accessibility was estimated by FPKM for each called peak and each library and averaged across all replicates for each condition.

Differential analysis of chromatin accessibility

To determine whether an OCR is differentially accessible between unedited (*SPII*^{+/+}) and CRISPR/Cas9 edited with monoallelic (*SPII*^{+/-}) or biallelic indels (*SPII*^{-/-}), the de-duplicated read counts for consensus OCRs were calculated for each replicate and normalized against background (10-K bins of genome) using the R package C_{saw} (v1.8.1; Lun and Smyth, 2016). OCRs with median value of <1.5 counts per million (CPM; 5–15 reads per OCR) across all replicates were removed from further differential analysis. Similar to gene differential analysis, accessibility differential analysis was performed pairwise (*SPII*^{+/-} versus *SPII*^{+/+}, *SPII*^{-/-} versus *SPII*^{+/+}, and *SPII*^{-/-} versus *SPII*^{+/-}) using glmQLFit approach fitting model ~condition in edgeR (v3.16.5) but with the normalization scaling factors calculated from c_{saw}. Differential OCRs between conditions were identified if FDR < 0.05 and absolute log₂ FC > 1.

Motif enrichment and TF binding site prediction

The HOMER motif database contains 332 motif matrices and is mostly based on the analysis of public ChIP-seq datasets (<http://homer.ucsd.edu/homer/motif/motifDatabase.html>). We used this motif collection for both TF enrichment and TF-binding site prediction. The 13,171 OCRs that were less accessible in monoallelic indels compared with unedited were applied to motif enrichment analysis using findMotifsGenome.pl from HOMER

(v4.10) with parameter setting size given. The cumulative binomial distribution was used for motif ranking. Any enriched TFs (Benjamini-Hochberg [BH] <0.05) were removed from ranking if they were not expressed or were minimally expressed in unedited samples (mean transcript per million < 1).

Protein interaction quantification (Sherwood et al., 2014) was used to predict TF binding sites from the assembly gap masked genome sequence as described in <https://github.com/orzechoj/piq-single>. Briefly, HOMER motifs were first converted to jaspar format using R package universalmotif (<http://bioconductor.org/packages/release/bioc/html/universalmotif.html>) and were used for generating the position weight matrix (PWM) hits across masked genome. The protein interaction quantification was run separately for unedited, monoallelic, and biallelic samples after merging the BAM files. A binding site candidate was defined by using the purity score cutoff 0.7 in at least one condition and overlapping with precalled OCRs.

RNA-seq preprocessing, gene differential expression analysis, and gene set enrichment analysis (GSEA)

The pair-end FASTQ file for each sample was mapped to genome assembly hg19 by STAR (v2.6.0c) independently. GencodeV19 annotation was used for gene feature annotation, and the raw read count for gene feature was calculated by htseq-count (v0.6.1; Anders et al., 2015) with parameter settings -f bam -r pos -s reverse -t exon -m union. The gene features localized on chrM or annotated as ribosomal RNAs were removed from the final sample-by-gene read count matrix.

The differential analysis was performed in R (v3.3.2) using R package edgeR (v3.16.5; Robinson et al., 2010). Briefly, the raw reads on gene features were applied to CPM (read CPM total reads). The gene features with median value of <0.25 CPM (10~14 reads per gene feature) across all samples were removed from differential analysis. The trimmed mean of M-values method was used to calculate normalization scaling factors, and the quasi-likelihood negative binomial generalized log-linear (glmQLFit) approach was applied to the count data with model fitting ~condition for each pairwise comparison (*SPII*^{+/-} versus *SPII*^{+/+}, *SPII*^{-/-} versus *SPII*^{+/+}, and *SPII*^{-/-} versus *SPII*^{+/-}). The DEGs were identified with the cutoff FDR < 0.05 and absolute logFC > 1.

To determine whether the DEGs detected in RS4;11 lines were enriched for transcripts vital to the pro- to pre-B cell transition, we performed preranked GSEA using R package fgsea (Korotkevich et al., 2019 Preprint; Subramanian et al., 2005). The DEGs between *SPII*^{+/-} and *SPII*^{-/-} were ranked on significance level (-log₁₀FDR) and compared with gene sets differentially expressed between pro-B and pre-B in healthy human (Hystad et al., 2007) and DEG sets between patient and control pro-B cells through CITE-seq. Enrichment score was calculated to reflect how often members of gene set occur at the top or bottom of the ranked dataset. Statistical significance of the enrichment score was determined by 100 permutations on RS4;11 DEG significance level rank.

Correlating gene expression with OCR accessibility

To explore the association between gene expression and OCR accessibility, we first focused on the expression FCs of DEGs between *SPII*^{+/-} and *SPII*^{-/-} RS4;11 lines and compared them with

accessibility FC of promoter OCRs that are located within gene promoter (−1,500 bp approximately +500 bp around TSS). A 50% LOESS regression on the normalized expression FCs was used to smooth gene expression over chromatin openness changes (Moskowitz et al., 2017). To predict the effect of distance on the gene expression–OCR accessibility correlation, each OCR was assigned to its closest gene, and the Spearman coefficient was calculated for each OCR–gene pair. The proportion of OCR–gene pairs in which expression perfectly coordinated with accessibility (Spearman coefficient = 1) was determined for each bin of distance between OCR and gene TSS. Compared with distal regions, gene expression–OCR accessibility correlation was peaked within 2,000 bp upstream of gene TSS, and statistical significance was determined by one-tailed proportion test.

Predicting interactions between PU.1 and gene promoters

To predict which gene promoter regions PU.1 interacts with, we defined promoter regions as the −2,000 bp around TSSs. We used bedtools to determine whether the promoter regions overlapped with optimal irreproducible discovery rate peaks from PU.1 ChIP-seq in human B lymphocyte cell line GM12878 from ENCODE (<https://www.encodeproject.org/experiments/ENCSTR000BGQ>) or contained a PU.1-binding motif (as described above in the Motif enrichment section).

Statistical analyses

Prism v9 (GraphPad) was used to perform nonparametric Mann-Whitney and one-way ANOVA tests for analysis of immunophenotypic data and PU.1 reporter experiments. Paired nonparametric Mann-Whitney *U* tests were used to compare the rank order of OCR accessibilities (\log_2 FPKM) among experimental conditions (*SPII*^{+/+} versus *SPII*^{+/-} versus *SPII*^{-/-}). χ^2 tests were used to evaluate the overlapping probability between less open DARs in edited *SPII* lines and PU.1 ChIP-seq peaks. Correlation between promoter region accessibilities and gene expression was evaluated by GSEA preranked module (Subramanian et al., 2005). Briefly, the software calculated enrichment score, which is a running sum statistic incremented by the absolute value of the ranking metric when a gene belongs to the set. The statistical significance of the enrichment score was assessed by permutation test ($n = 1,000$). Binomial tests were used to calculate motif enrichment, and P value was adjusted with BH multiple-comparisons procedure. Negative binomial generalized linear models were implemented using the “edgeR” to perform the differential analysis of replicated count data from RNA-seq and ATAC-seq. P value was adjusted with BH multiple-comparisons procedure. Hypergeometric test was used for pathway enrichment. P value was adjusted with BH multiple-comparisons procedure. Pro-B cell gene set enrichment was evaluated by GSEA preranked module (Subramanian et al., 2005), in which enrichment score was incremented by the ranking metrics and P value was estimated by permutation test ($n = 100$). One-tailed one-sample proportion test was used to determine statistically significant higher proportion of coordinate OCR accessibility–gene expression when OCR was located within 2,000 bp upstream of gene TSS.

Data availability statement

RNA-seq data and ATAC-seq data were deposited and are publicly available through EMBL-EBI ArrayExpress under accession nos. E-MTAB-8677 and E-MTAB-8676. CITE-seq data were deposited and are publicly available through Gene Expression Omnibus under accession no. GSE165645. *SPII* mutations described here have been deposited in the National Center for Biotechnology Information ClinVar database under accession nos. SCV001134082, SCV001134083, SCV001134083, SCV001134085, SCV001451930, and SCV001451931.

Online supplemental material

Fig. S1 shows PU.1 expression in PBMCs, PBMC subsets, and neutrophils from PU.MA patients and healthy controls. Chromatograms show relative proportions of mutated and unmutated PU.MA patient *SPII* transcripts. Fig. S2 shows the number of dendritic cell subsets and the frequency of cytokine-secreting myeloid cells from PU.MA patients and HD controls. Relative frequencies of T helper cell subsets, monocyte subsets, and innate lymphoid cells are also displayed. Fig. S3 shows expression of key proteins and transcripts that define mononuclear bone marrow populations. Fig. S4 shows the distribution of *SPII* and/or AAVS mutations in single HSPCs after editing but before differentiation and in batch-sorted cells after in vitro differentiation. Fig. S5 shows EGFP expression by PU.1 reporter cells that express either unmutated PU.1 or patient PU.1 variants or common gnomAD variants or coexpressing unmutated PU.1 and an ETS fragment. Immunoblots of coimmunoprecipitations show how unmutated PU.1 and patient PU.1 variants interact with IRF4 and IRF8. Confocal images of HEK293 cells show intracellular distribution of unmutated and mutated PU.1.

Acknowledgments

We thank the patients and their families. We thank Children’s Hospital of Philadelphia’s Centers for Childhood Cancer Research, Spatial and Functional Genomics, and Robert’s Individualized Medical Genomics for pediatric marrow control samples, next-generation sequencing, and clinical sequencing support, respectively. We appreciate Donna M. Muzny, Shalini N. Jhangiani, Richard A. Gibbs, and Zeynep H. Coban-Akdemir from the Baylor College of Medicine HGSC and Baylor-Hopkins Center for Mendelian Genomics for exome sequencing and bioinformatic support. We thank Drs. Elena Raikina, Vera Tseitlina, and Dmitri Pershin for clinical analyses and treating physicians from the Immunology Department of the D. Rogachev Center. We thank Tiancheng Wang from Children’s Hospital of Philadelphia’s Center for Applied Genomics for assistance with genotyping. We thank GeneDx bioinformatics for identifying our index patient’s *SPII* variant and identifying the p.L233Afs*53 variant. We thank Richard P. Lifton and Michael Silverman for manuscript review and advice.

This work was supported by a Translational Research Program grant award (N. Romberg and A. Marson) from the Jeffrey Model Foundation, grants from the National Institutes of Health, National Institute of Allergy and Infectious Diseases, (AI146026 to N. Romberg and A.D. Wells), National Science Foundation

(MCB 1545160 to G.M.K. Poon), and a pilot grant from Children's Hospital of Philadelphia's Center for Spatial and Functional Genomics to N. Romberg. N. Romberg receives salary support from the Jeffrey Model Foundation. A. Marson and A.R. Abate are an investigators at the Chan Zuckerberg Biohub. A. Marson holds a Career Award for Medical Scientists from the Burroughs Wellcome Fund and the Cancer Research Institute Lloyd J. Old STAR award, has received funding from the Innovative Genomics Institute, and is a member of the Parker Institute for Cancer Immunotherapy. D.N. Nguyen received funding from the Alpha Stem Cell Clinic Fellowship, the CIS CSL Behring Fellowship, and National Institutes of Health, National Institute of Allergy and Infectious Diseases grant K08AI153767. D. Sun received grant funding from the National Institute of Child Health and Human Development (T32HD043021). F. Wright received a GERM grant from the Infectious Diseases Society of America Foundation. L.T. Vo is supported by a fellowship from the Damon Runyon Cancer Research Foundation. The University of California San Francisco Flow Cytometry Core is supported by the National Institutes of Health Diabetes Research Center Grant (P30 DK063720). Effort by the Baylor-Hopkins Center for Mendelian Genomics was funded by National Institutes of Health–National Human Genome Research Institute grant UMI HG006542.

Author contributions: C. Le Coz, B.E. Nolan, A.V. Albrecht, S. Khani, P. Pillarisetti, C. Khanna, F. Wright, P.A. Chen, S. Yoon, K. Maurer, and D. Sun performed experiments and analyzed data. D.N. Nguyen and B. Demaree performed experiments and bioinformatics for hematopoietic stem cell editing experiments. J.M. Puck and L.T. Vo designed, supervised, and interpreted in vitro hematopoietic stem cell differentiation assays. C. Su performed bioinformatic ATAC-seq and RNA-seq analyses on RS4;11 cell lines. M.V. Gonzalez and J.P. Garifallou performed bioinformatic analyses of CITE-seq experiments. A.L. Stiegler performed three-dimensional prediction modeling. A. Rymaszewski and V. Zakharova performed bioinformatics analysis of whole exome sequencing. S.H. Kroft evaluated pathology samples. G. Wertheim performed histological analyses of patient samples. T.S. Olson, A.E. Seif, J. Hajjar, K.E. Sullivan, J.M. Routes, S.K. Nicholas, J. Verbsky, A. Shcherbina, A. Mukhina, N.L. Rudy, and T.P. Atkinson evaluated patients and collected data. J.M. Routes, J.R. Lupski, I.K. Chinn, A.C.E. Hurst, H. Hakonarson, and S.F.A. Grant supervised bioinformatics analysis including whole-exome sequencing. T.J. Boggon supervised predictive modeling and provided advice. A.D. Wells, A. Marson, G.M.K. Poon, and A.R. Abate planned and supervised research and data analysis and provided advice. N. Romberg conceptualized the study, coordinated research, supervised experimental work, and analyzed data. C. Le Coz, N. Romberg, and G.M.K. Poon prepared the manuscript.

Disclosures: D. Nguyen reported a patent to PCT/US2019/066079 with royalties paid. T. Olson reported personal fees from Bluebird Bio outside of the submitted work. J. Puck reported other from Invitae (spouse's employer) and other from UpToDate (royalties) outside the submitted work. J. Hajjar reported grants from Immune Deficiency Foundation, the US immunodeficiency network, Chao-physician Scientist award, the Texas Medical Center Digestive Diseases Center, and the Jeffrey Modell

Foundation, other from Horizon, Pharming, Baxalta, CSL Behring, and the National Guard and Al-Faisal University Hospital outside the submitted work. J. Lupski reported grants from NIH/NINDS (R35 NS105078), and NIH/NIGMS (R01 GM106373), personal fees from Regeneron Genetics Center and Novartis, and other from 23andMe outside the submitted work. A. Marson reported personal fees from Arsenal Biosciences, Spotlight Therapeutics, PACT Pharma, Merck, Vertex, AlphaSights, ALDA, Amgen, Trizell, Juno Therapeutics, Health Advances, Lonza, Bernstein, Abbvie, Genentech, Illumina, Arcus, Jackson Laboratories, NanoString Technologies, GLG, and Rupert Case Management, grants from Anthem, Gilead, GlaxoSmithKline, Juno Therapeutics, Epinomics, Sanofi, and Parker Institute for Cancer Immunotherapy (PICI), non-financial support from Illumina, other from Parker Institute for Cancer Immunotherapy (PICI), ThermoFisher, and Third Rock Ventures outside the submitted work. In addition, A. Marson had a patent to WO 2016/123578 licensed (The identity of the licensee has not been made public) and a patent to PCT/US19/66079 licensed (The identity of the licensee has not been made public); and is an investor in and informal advisor to Offline Ventures. No other disclosures were reported.

Submitted: 13 August 2020

Revised: 9 February 2021

Accepted: 16 March 2021

References

- Albrecht, A.V., H.M. Kim, and G.M.K. Poon. 2018. Mapping interfacial hydration in ETS-family transcription factor complexes with DNA: a chimeric approach. *Nucleic Acids Res.* 46:10577–10588. <https://doi.org/10.1093/nar/gky894>
- Anders, S., P.T. Pyl, and W. Huber. 2015. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics.* 31:166–169. <https://doi.org/10.1093/bioinformatics/btu638>
- Andersen, L.L., N. Mork, L.S. Reinert, E. Kofod-Olsen, R. Narita, S.E. Jorgensen, K.A. Skipper, K. Honing, H.H. Gad, L. Ostergaard, et al. 2015. Functional IRF3 deficiency in a patient with herpes simplex encephalitis. *J. Exp. Med.* 212:1371–1379. <https://doi.org/10.1084/jem.20142274>
- Bigley, V., M. Haniffa, S. Doulatov, X.-N. Wang, R. Dickinson, N. McGovern, L. Jardine, S. Pagan, I. Dimmick, I. Chua, et al. 2011. The human syndrome of dendritic cell, monocyte, B and NK lymphoid deficiency. *J. Exp. Med.* 208:227–234. <https://doi.org/10.1084/jem.20101459>
- Boisson, B., Y.-D. Wang, A. Bosompem, C.S. Ma, A. Lim, T. Kochetkov, S.G. Tangye, J.-L. Casanova, and M.E. Conley. 2013. A recurrent dominant negative E47 mutation causes agammaglobulinemia and BCR(-) B cells. *J. Clin. Invest.* 123:4781–4785. <https://doi.org/10.1172/JCI71927>
- Bravo García-Morato, M., F.J. Aracil Santos, A.C. Briones, A. Blázquez Moreno, Á. Del Pozo Maté, Á. Domínguez-Soto, M.J. Beato Merino, L. Del Pino Molina, J. Torres Canizales, A.V. Marin, et al. 2018. New human combined immunodeficiency caused by interferon regulatory factor 4 (IRF4) deficiency inherited by uniparental isodisomy. *J. Allergy Clin. Immunol.* 141:1924–1927.e18. <https://doi.org/10.1016/j.jaci.2017.12.995>
- Bruton, O.C. 1952. Agammaglobulinemia. *Pediatrics.* 9:722–728.
- Carotta, S., A. Dakic, A. D'Amico, S.H.M. Pang, K.T. Greig, S.L. Nutt, and L. Wu. 2010. The transcription factor PU.1 controls dendritic cell development and Flt3 cytokine receptor expression in a dose-dependent manner. *Immunity.* 32:628–641. <https://doi.org/10.1016/j.immuni.2010.05.005>
- Chang, K.T., J. Guo, A. di Ronza, and M. Sardiello. 2018. Aminode: Identification of evolutionary constraints in the human proteome. *Sci. Rep.* 8: 1357. <https://doi.org/10.1038/s41598-018-19744-w>
- Chopin, M., A.T. Lun, Y. Zhan, J. Schreuder, H. Coughlan, A. D'Amico, L.A. Mielke, F.F. Almeida, A.J. Kueh, R.A. Dickins, et al. 2019. Transcription factor PU.1 promotes conventional dendritic cell identity and function

- via induction of transcriptional regulator DC-SCRIPT. *Immunity*. 50: 77–90.e5. <https://doi.org/10.1016/j.immuni.2018.11.010>
- Ciau-Uitz, A., L. Wang, R. Patient, and F. Liu. 2013. ETS transcription factors in hematopoietic stem cell development. *Blood Cells Mol. Dis.* 51:248–255. <https://doi.org/10.1016/j.bcmd.2013.07.010>
- Clement, K., H. Rees, M.C. Canver, J.M. Gehrke, R. Farouni, J.Y. Hsu, M.A. Cole, D.R. Liu, J.K. Joung, D.E. Bauer, and L. Pinello. 2019. CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.* 37:224–226. <https://doi.org/10.1038/s41587-019-0032-3>
- Collin, R.W.J., E. Kalay, M. Tariq, T. Peters, B. van der Zwaag, H. Venselaar, J. Oostrik, K. Lee, Z.M. Ahmed, R. Çaylan, et al. 2008. Mutations of ESRRB encoding estrogen-related receptor beta cause autosomal-recessive nonsyndromic hearing impairment DFNB35. *Am. J. Hum. Genet.* 82: 125–138. <https://doi.org/10.1016/j.ajhg.2007.09.008>
- Conley, M.E., A.K. Dobbs, D.M. Farmer, S. Kilic, K. Paris, S. Grigoriadou, E. Coustan-Smith, V. Howard, and D. Campana. 2009. Primary B cell immunodeficiencies: comparisons and contrasts. *Annu. Rev. Immunol.* 27:199–227. <https://doi.org/10.1146/annurev.immunol.021908.132649>
- de Pontual, L., V. Népote, T. Attié-Bitach, H. Al Halabiah, H. Trang, V. Elghouzzi, B. Levacher, K. Benihoud, J. Augé, C. Faure, et al. 2003. Noradrenergic neuronal development is impaired by mutation of the proneural HASH-1 gene in congenital central hypoventilation syndrome (Ondine's curse). *Hum. Mol. Genet.* 12:3173–3180. <https://doi.org/10.1093/hmg/ddg339>
- DeKoter, R.P., and H. Singh. 2000. Regulation of B lymphocyte and macrophage development by graded expression of PU.1. *Science*. 288:1439–1441. <https://doi.org/10.1126/science.288.5470.1439>
- DeKoter, R.P., M.B. Kamath, and I.B. Houston. 2007a. Analysis of concentration-dependent functions of PU.1 in hematopoiesis using mouse models. *Blood Cells Mol. Dis.* 39:316–320. <https://doi.org/10.1016/j.bcmd.2007.06.004>
- DeKoter, R.P., B.L. Schweitzer, M.B. Kamath, D. Jones, H. Tagoh, C. Bonifer, D.A. Hildeman, and K.J. Huang. 2007b. Regulation of the interleukin-7 receptor α promoter by the Ets transcription factors PU.1 and GABP1 protein in developing B cells. *J. Biol. Chem.* 282:14194–14204. <https://doi.org/10.1074/jbc.M700377200>
- Demaree, B., C.L. Delley, H.N. Vasudevan, C.A.C. Peretz, D. Ruff, C.C. Smith, and A.R. Abate. 2021. Joint profiling of DNA and proteins in single cells to dissect genotype-phenotype associations in leukemia. *Nat. Commun.* 12:1583. <https://doi.org/10.1038/s41467-021-21810-3>
- ENCODE Project Consortium. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 489:57–74. <https://doi.org/10.1038/nature11247>
- Esaki, S., M.G. Evich, N. Erlitzki, M.W. Germann, and G.M.K. Poon. 2017. Multiple DNA-binding modes for the ETS family transcription factor PU.1. *J. Biol. Chem.* 292:16044–16054. <https://doi.org/10.1074/jbc.M117.798207>
- Facchetti, F., M. Cella, S. Festa, D.H. Fremont, and M. Colonna. 2002. An unusual Fc receptor-related protein expressed in human centroblasts. *Proc. Natl. Acad. Sci. USA*. 99:3776–3781. <https://doi.org/10.1073/pnas.022042699>
- Feichtinger, R.G., B.E. Mucha, H. Hengel, Z. Orfi, C. Makowski, J. Dort, G. D'Anjou, T.T.M. Nguyen, R. Buchert, H. Juenger, et al. 2019. Biallelic variants in the transcription factor PAX7 are a new genetic cause of myopathy. *Genet. Med.* 21:2521–2531. <https://doi.org/10.1038/s41436-019-0532-z>
- Garg, V., I.S. Kathiriyai, R. Barnes, M.K. Schluterman, I.N. King, C.A. Butler, C.R. Rothrock, R.S. Eapen, K. Hirayama-Yamada, K. Joo, et al. 2003. GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5. *Nature*. 424:443–447. <https://doi.org/10.1038/nature01827>
- Hambleton, S., S. Salem, J. Bustamante, V. Bigley, S. Boisson-Dupuis, J. Azevedo, A. Fortin, M. Haniffa, L. Ceron-Gutierrez, C.M. Bacon, et al. 2011. IRF8 mutations and human dendritic-cell immunodeficiency. *N. Engl. J. Med.* 365:127–138. <https://doi.org/10.1056/NEJMoa1100066>
- Heinz, S., C. Benner, N. Spann, E. Bertolino, Y.C. Lin, P. Laslo, J.X. Cheng, C. Murre, H. Singh, and C.K. Glass. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell.* 38:576–589. <https://doi.org/10.1016/j.molcel.2010.05.004>
- Heinz, S., C.E. Romanoski, C. Benner, K.A. Allison, M.U. Kaikkonen, L.D. Orozco, and C.K. Glass. 2013. Effect of natural genetic variation on enhancer selection and function. *Nature*. 503:487–492. <https://doi.org/10.1038/nature12615>
- Houston, I.B., M.B. Kamath, B.L. Schweitzer, T.M. Chlon, and R.P. DeKoter. 2007. Reduction in PU.1 activity results in a block to B-cell development, abnormal myeloid proliferation, and neonatal lethality. *Exp. Hematol.* 35: 1056–1068. <https://doi.org/10.1016/j.exphem.2007.04.005>
- Hsu, A.P., E.P. Sampaio, J. Khan, K.R. Calvo, J.E. Lemieux, S.Y. Patel, D.M. Frucht, D.C. Vinh, R.D. Auth, A.F. Freeman, et al. 2011. Mutations in GATA2 are associated with the autosomal dominant and sporadic monocytopenia and mycobacterial infection (MonoMAC) syndrome. *Blood*. 118:2653–2655. <https://doi.org/10.1182/blood-2011-05-356352>
- Hystad, M.E., J.H. Myklebust, T.H. Bø, E.A. Sivertsen, E. Rian, L. Forfang, E. Munthe, A. Rosenwald, M. Chiorazzi, I. Jonassen, et al. 2007. Characterization of early stages of human B cell development by gene expression profiling. *J. Immunol.* 179:3662–3671. <https://doi.org/10.4049/jimmunol.179.6.3662>
- Itan, Y., L. Shang, B. Boisson, E. Patin, A. Bolze, M. Moncada-Vélez, E. Scott, M.J. Ciancanelli, F.G. Lafaille, J.G. Markle, et al. 2015. The human gene damage index as a gene-level approach to prioritizing exome variants. *Proc. Natl. Acad. Sci. USA*. 112:13615–13620. <https://doi.org/10.1073/pnas.1518646112>
- Kelberman, D., K. Rizzoti, A. Avilion, M. Bitner-Glindzicz, S. Cianfarani, J. Collins, W.K. Chong, J.M. Kirk, J.C. Achermann, R. Ross, et al. 2006. Mutations within Sox2/SOX2 are associated with abnormalities in the hypothalamo-pituitary-gonadal axis in mice and humans. *J. Clin. Invest.* 116:2442–2455. <https://doi.org/10.1172/JCI28658>
- Kessler, M.D., D.P. Loesch, J.A. Perry, N.L. Heard-Costa, D. Taliun, B.E. Cade, H. Wang, M. Daya, J. Ziniti, S. Datta, et al. TOPMed Population Genetics Working Group. 2020. De novo mutations across 1,465 diverse genomes reveal mutational insights and reductions in the Amish founder population. *Proc. Natl. Acad. Sci. USA*. 117:2560–2569. <https://doi.org/10.1073/pnas.1902766117>
- Klemsz, M.J., S.R. McKercher, A. Celada, C. Van Beveren, and R.A. Maki. 1990. The macrophage and B cell-specific transcription factor PU.1 is related to the ets oncogene. *Cell*. 61:113–124. [https://doi.org/10.1016/0092-8674\(90\)90219-5](https://doi.org/10.1016/0092-8674(90)90219-5)
- Kodandapani, R., F. Pio, C.-Z. Ni, G. Piccialli, M. Klemsz, S. McKercher, R.A. Maki, and K.R. Ely. 1996. A new pattern for helix-turn-helix recognition revealed by the PU.1 ETS-domain-DNA complex. *Nature*. 380:456–460. <https://doi.org/10.1038/380456a0>
- Kodo, K., T. Nishizawa, M. Furutani, S. Arai, E. Yamamura, K. Joo, T. Takahashi, R. Matsuoka, and H. Yamagishi. 2009. GATA6 mutations cause human cardiac outflow tract defects by disrupting semaphorin-plexin signaling. *Proc. Natl. Acad. Sci. USA*. 106:13933–13938. <https://doi.org/10.1073/pnas.0904744106>
- Kong, X.-F., R. Martinez-Barricarte, J. Kennedy, F. Mele, T. Lazarov, E.K. Deenick, C.S. Ma, G. Breton, K.B. Lucero, D. Langlais, et al. 2018. Disruption of an antimicrobial circuit between dendritic and helper T cells in human SPPL2a deficiency. *Nat. Immunol.* 19:973–985. <https://doi.org/10.1038/s41590-018-0178-z>
- Korotkevich, G., V. Sukhov, and A. Sergushichev. 2019. Fast gene set enrichment analysis. *bioRxiv*. doi: (Preprint posted October 22, 2019). <https://doi.org/10.1101/060012>
- Kwok, J.C., J. Perdomo, and B.H. Chong. 2007. Identification of a monopartite sequence in PU.1 essential for nuclear import, DNA-binding and transcription of myeloid-specific genes. *J. Cell. Biochem.* 101:1456–1474. <https://doi.org/10.1002/jcb.21264>
- La Motte-Mohs, R.N., E. Herer, and J.C. Zúñiga-Pflücker. 2005. Induction of T-cell development from human cord blood hematopoietic stem cells by Delta-like 1 in vitro. *Blood*. 105:1431–1439. <https://doi.org/10.1182/blood-2004-04-1293>
- Lek, M., K.J. Karczewski, E.V. Minikel, K.E. Samocha, E. Banks, T. Fennell, A.H. O'Donnell-Luria, J.S. Ware, A.J. Hill, B.B. Cummings, et al. Exome Aggregation Consortium. 2016. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 536:285–291. <https://doi.org/10.1038/nature19057>
- Lun, A.T.L., and G.K. Smyth. 2016. csaw: a Bioconductor package for differential binding analysis of ChIP-seq data using sliding windows. *Nucleic Acids Res.* 44:e45. <https://doi.org/10.1093/nar/gkv1191>
- Malecki, M.T., U.S. Jhala, A. Antonellis, L. Fields, A. Doria, T. Orban, M. Saad, J.H. Warram, M. Montminy, and A.S. Krolewski. 1999. Mutations in NEUROD1 are associated with the development of type 2 diabetes mellitus. *Nat. Genet.* 23:323–328. <https://doi.org/10.1038/15500>
- Mayran, A., and J. Drouin. 2018. Pioneer transcription factors shape the epigenetic landscape. *J. Biol. Chem.* 293:13795–13804. <https://doi.org/10.1074/jbc.R117.001232>
- Moskowitz, D.M., D.W. Zhang, B. Hu, S. Le Saux, R.E. Yanes, Z. Ye, J.D. Buenrosto, C.M. Weyand, W.J. Greenleaf, and J.J. Goronzy. 2017. Epigenomics of human CD8 T cell differentiation and aging. *Sci. Immunol.* 2: eaag0192. <https://doi.org/10.1126/sciimmunol.aag0192>

- Mueller, B.U., T. Pabst, M. Osato, N. Asou, L.M. Johansen, M.D. Minden, G. Behre, W. Hiddemann, Y. Ito, and D.G. Tenen. 2002. Heterozygous PU.1 mutations are associated with acute myeloid leukemia. *Blood*. 100: 998–1007. <https://doi.org/10.1182/blood.V100.3.998>
- Munde, M., S. Wang, A. Kumar, C.E. Stephens, A.A. Farahat, D.W. Boykin, W.D. Wilson, and G.M.K. Poon. 2014. Structure-dependent inhibition of the ETS-family transcription factor PU.1 by novel heterocyclic diamidines. *Nucleic Acids Res.* 42:1379–1390. <https://doi.org/10.1093/nar/gkt955>
- Nichols, K.E., J.D. Crispino, M. Poncz, J.G. White, S.H. Orkin, J.M. Maris, and M.J. Weiss. 2000. Familial dyserythropoietic anaemia and thrombocytopenia due to an inherited mutation in GATA1. *Nat. Genet.* 24:266–270. <https://doi.org/10.1038/73480>
- Pellegrino, M., A. Sciambi, S. Treusch, R. Durruthy-Durruthy, K. Gokhale, J. Jacob, T.X. Chen, J.A. Geis, W. Oldham, J. Matthews, et al. 2018. High-throughput single-cell DNA sequencing of acute myeloid leukemia tumors with droplet microfluidics. *Genome Res.* 28:1345–1352. <https://doi.org/10.1101/gr.232272.117>
- Pham, T.-H., J. Minderjahn, C. Schmidl, H. Hoffmeister, S. Schmidhofer, W. Chen, G. Längst, C. Benner, and M. Rehli. 2013. Mechanisms of in vivo binding site selection of the hematopoietic master transcription factor PU.1. *Nucleic Acids Res.* 41:6391–6402. <https://doi.org/10.1093/nar/gkt355>
- Pongubala, J.M.R., and M.L. Atchison. 1997. PU.1 can participate in an active enhancer complex without its transcriptional activation domain. *Proc. Natl. Acad. Sci. USA.* 94:127–132. <https://doi.org/10.1073/pnas.94.1.127>
- Poon, G.M.K., P. Gross, and R.B. Macgregor Jr. 2002. The sequence-specific association of the ETS domain of murine PU.1 with DNA exhibits unusual energetics. *Biochemistry.* 41:2361–2371. <https://doi.org/10.1021/bi015572q>
- Robinson, M.D., D.J. McCarthy, and G.K. Smyth. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 26:139–140. <https://doi.org/10.1093/bioinformatics/btp616>
- Rosenbauer, F., K. Wagner, J.L. Kutok, H. Iwasaki, M.M. Le Beau, Y. Okuno, K. Akashi, S. Fiering, and D.G. Tenen. 2004. Acute myeloid leukemia induced by graded reduction of a lineage-specific transcription factor, PU.1. *Nat. Genet.* 36:624–630. <https://doi.org/10.1038/ng1361>
- Scott, E.W., M.C. Simon, J. Anastasi, and H. Singh. 1994. Requirement of transcription factor PU.1 in the development of multiple hematopoietic lineages. *Science.* 265:1573–1577. <https://doi.org/10.1126/science.8079170>
- Sherwood, R.I., T. Hashimoto, C.W. O'Donnell, S. Lewis, A.A. Barkal, J.P. van Hoff, V. Karun, T. Jaakkola, and D.K. Gifford. 2014. Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape. *Nat. Biotechnol.* 32:171–178. <https://doi.org/10.1038/nbt.2798>
- Steidl, U., F. Rosenbauer, R.G.W. Verhaak, X. Gu, A. Ebraldidze, H.H. Otu, S. Klippel, C. Steidl, I. Bruns, D.B. Costa, et al. 2006. Essential role of Jun family transcription factors in PU.1 knockdown-induced leukemic stem cells. *Nat. Genet.* 38:1269–1277. <https://doi.org/10.1038/ng1898>
- Stoeckius, M., C. Hafemeister, W. Stephenson, B. Houck-Loomis, P.K. Chattopadhyay, H. Swerdlow, R. Satija, and P. Smibert. 2017. Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods.* 14:865–868. <https://doi.org/10.1038/nmeth.4380>
- Subramanian, A., P. Tamayo, V.K. Mootha, S. Mukherjee, B.L. Ebert, M.A. Gillette, A. Paulovich, S.L. Pomeroy, T.R. Golub, E.S. Lander, and J.P. Mesirov. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA.* 102:15545–15550. <https://doi.org/10.1073/pnas.0506580102>
- Tangye, S.G., W. Al-Herz, A. Bousfiha, T. Chatila, C. Cunningham-Rundles, A. Etzioni, J.L. Franco, S.M. Holland, C. Klein, T. Morio, et al. 2020. Human inborn errors of immunity: 2019 update on the classification from the International Union of Immunological Societies Expert Committee. *J. Clin. Immunol.* 40:24–64. <https://doi.org/10.1007/s10875-019-00737-x>
- van Spriël, A.B., S. de Keijzer, A. van der Schaaf, K.H. Gartlan, M. Sofi, A. Light, P.C. Linssen, J.B. Boezeman, M. Zuidschermwoude, I. Reinieren-Beeren, et al. 2012. The tetraspanin CD37 orchestrates the $\alpha(4)\beta(1)$ integrin-Akt signaling axis and supports long-lived plasma cell survival. *Sci. Signal.* 5:ra82. <https://doi.org/10.1126/scisignal.2003113>
- Wasserman, R., Y.S. Li, and R.R. Hardy. 1997. Down-regulation of terminal deoxynucleotidyl transferase by Ig heavy chain in B lineage cells. *J. Immunol.* 158:1133–1138.
- Will, B., T.O. Vogler, S. Narayanagari, B. Bartholdy, T.I. Todorova, M. da Silva Ferreira, J. Chen, Y. Yu, J. Mayer, L. Barreyro, et al. 2015. Minimal PU.1 reduction induces a preleukemic state and promotes development of acute myeloid leukemia. *Nat. Med.* 21:1172–1181. <https://doi.org/10.1038/nm.3936>
- Khani, S., S. Lee, H.M. Kim, S. Wang, S. Esaki, V.L.T. Ha, M. Khanezarrin, G.L. Fernandez, A.V. Albrecht, J.M. Aramini, et al. 2020. Intrinsic disorder controls two functionally distinct dimers of the master transcription factor PU.1. *Sci. Adv.* 6:eaay3178. <https://doi.org/10.1126/sciadv.aay3178>
- Zhang, Y., T. Liu, C.A. Meyer, J. Eeckhoute, D.S. Johnson, B.E. Bernstein, C. Nusbaum, R.M. Myers, M. Brown, W. Li, and X.S. Liu. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9:R137. <https://doi.org/10.1186/gb-2008-9-9-r137>
- Zhong, H., A. Takeda, R. Nazari, H. Shio, G. Blobel, and N.R. Yaseen. 2005. Carrier-independent nuclear import of the transcription factor PU.1 via RanGTP-stimulated binding to Nup153. *J. Biol. Chem.* 280:10675–10682. <https://doi.org/10.1074/jbc.M412878200>

Supplemental material

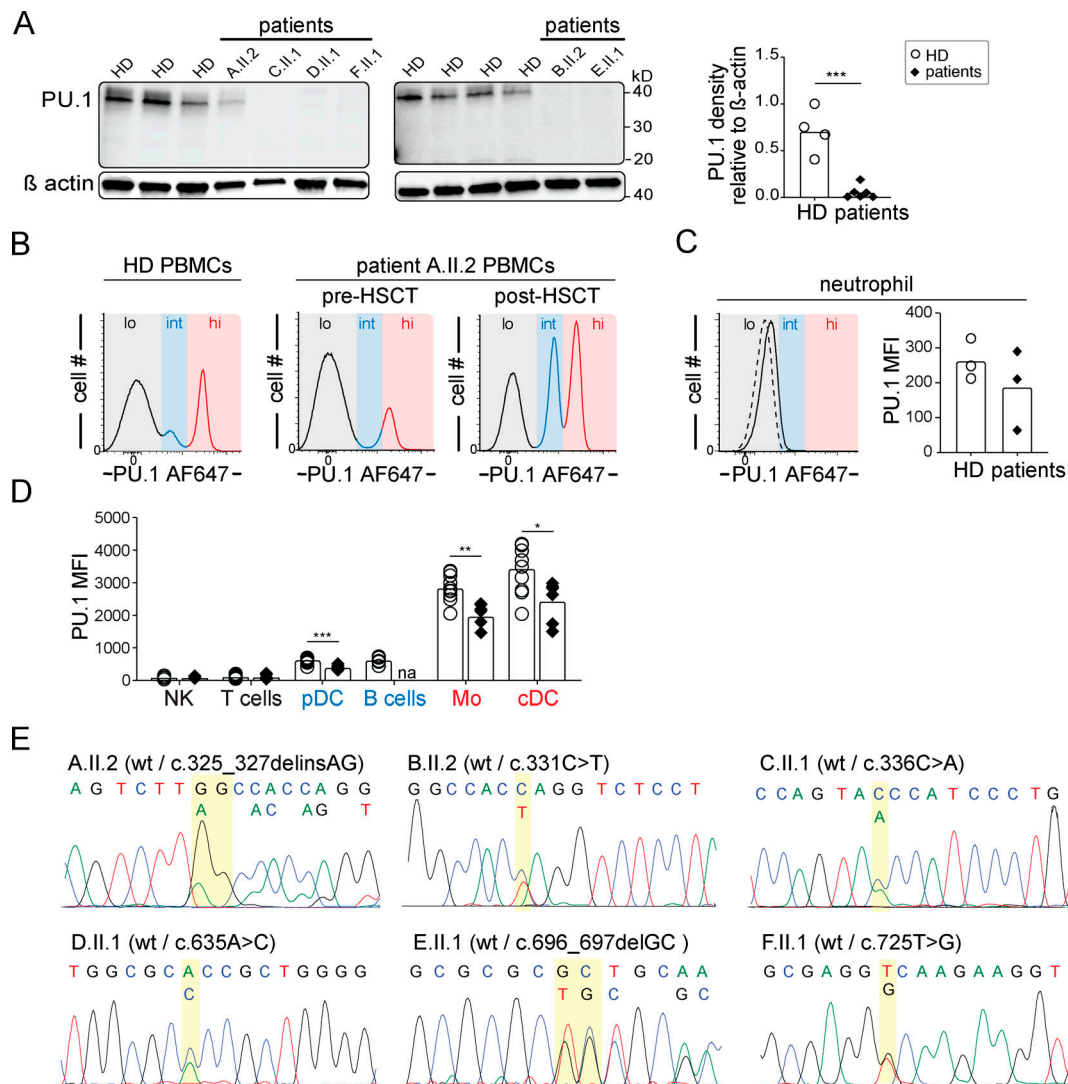


Figure S1. **PU.1 expression is reduced in PU.MA patients.** (A) Left: An anti-PU.1 N-terminal antibody detects PU.1 in B cell-depleted HD and PU.MA patient PBMC lysates. 10 μ g of total protein was loaded per lane. Right: PU.1 to β -actin staining intensity ratios for four HD controls and six patients. Bars represent mean values. (B) Histograms show PU.1 staining of PBMCs from a representative HD and from patient A.II.2 before and after hematopoietic stem cell transplant (HSCT). (C) Histograms show PU.1 expression by HD and patient C.II.1 circulating CD11b⁺ neutrophils. Bars show average neutrophil PU.1 mean fluorescence intensities (MFIs) for HDs (n = 3) and patients (n = 3). (D) Bars show average PU.1 MFIs for HD (n = 10) and patient (n = 6) NK cells (CD3⁻CD20⁻CD56⁺CD16⁺), T cells, B cells, pDCs (CD3⁻CD20⁻CD14⁻HLADR⁺CD11c⁻CD303⁺), monocytes (Mo), and cDCs (CD3⁻CD20⁻CD14⁻HLADR⁺CD11c⁺CD303⁻). (E) Chromatograms depict *SPI1* transcripts from patient PBMC cDNA libraries. Positions where nucleotides are replaced, inserted, or deleted are highlighted (yellow). *, P < 0.05; **, P < 0.01; and ***, P < 0.001 (two-sided Mann-Whitney U tests). na, not applicable due to the absence of circulating B cells.

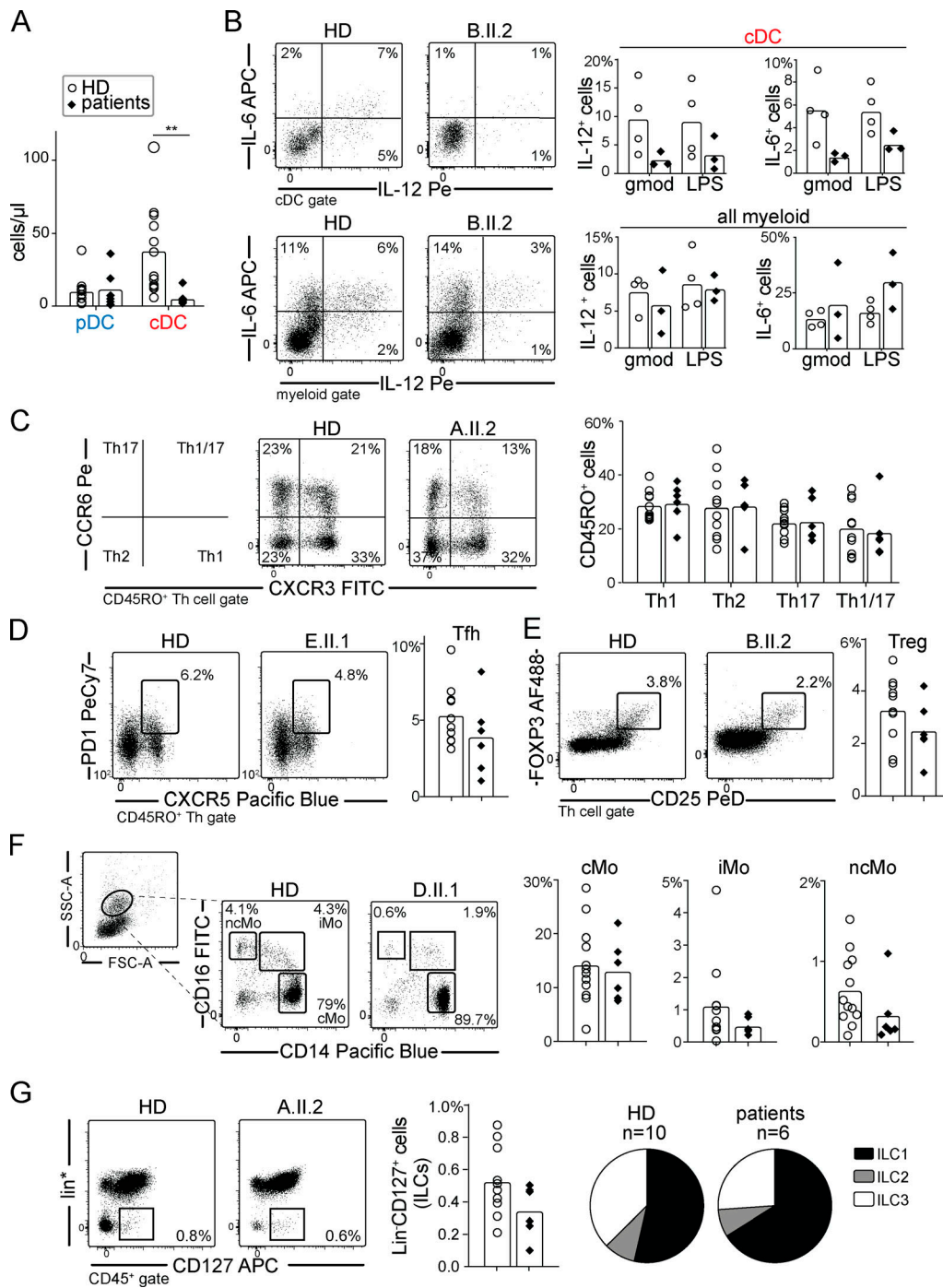


Figure S2. IL-12- and IL-6-secreting cDCs are scarce in PU.MA patient blood samples. (A) Absolute pDC and cDC concentrations (cell/ μ l) in HD ($n = 12$) and PU.MA patient ($n = 6$) blood samples are shown. (B) The frequency of IL-12- and/or IL-6- expressing cDCs (top) or total myeloid cells (inclusive of cDCs; bottom) from HDs ($n = 3$) and patients ($n = 3$) is identified by flow cytometry after 6 h of gardiquimod (gmod) or LPS activation. (C and D) The gating strategy and frequencies of HD ($n = 12$) and patient ($n = 6$) memory T helper (Th) cell subsets including Th1 (CXCR3⁺CCR6⁻), Th2 (CCR6⁻CXCR3⁻), Th17 (CXCR3⁺CCR6⁺), Th1-17 (CXCR3⁺CCR6⁺; C), and cTfh (CXCR5⁺PD1⁺; D) as percentages of CD45RO⁺ Th cells. (E) Regulatory T cell (T reg cell; CD25⁺FOXP3⁺) frequencies are displayed as percentages of total Th cells. (F) Classical (cMo; CD14⁺CD16^{lo}), intermediate (iMo; CD14⁺CD16⁺), and nonclassical (ncMo; CD14^{lo}CD16²⁺) monocyte frequencies are shown as a percentage of total HD ($n = 12$) and patient ($n = 6$) PBMCs. (G) Flow cytometry plots illustrate the strategy used to identify circulating ILCs. Lineage (lin*)-positive staining is defined as cells expressing any of the following proteins: CD3, CD14, CD19, or CD11c. ILC frequencies are shown as a percentage of total PBMCs for HDs ($n = 10$) and patients ($n = 6$). Pie charts depict averaged ratios of ILC subsets from HD and patients (ILC1, CD117⁻CRTH2⁺; ILC2, CRTH2⁺; ILC3, CD117⁺CRTH2⁻). Bars represent mean values throughout. **, $P < 0.01$ (two-sided Mann-Whitney U tests). cTfh, circulating T follicular helper cell; FSC-A, forward scatter-A; SSC-A, side scatter-A.

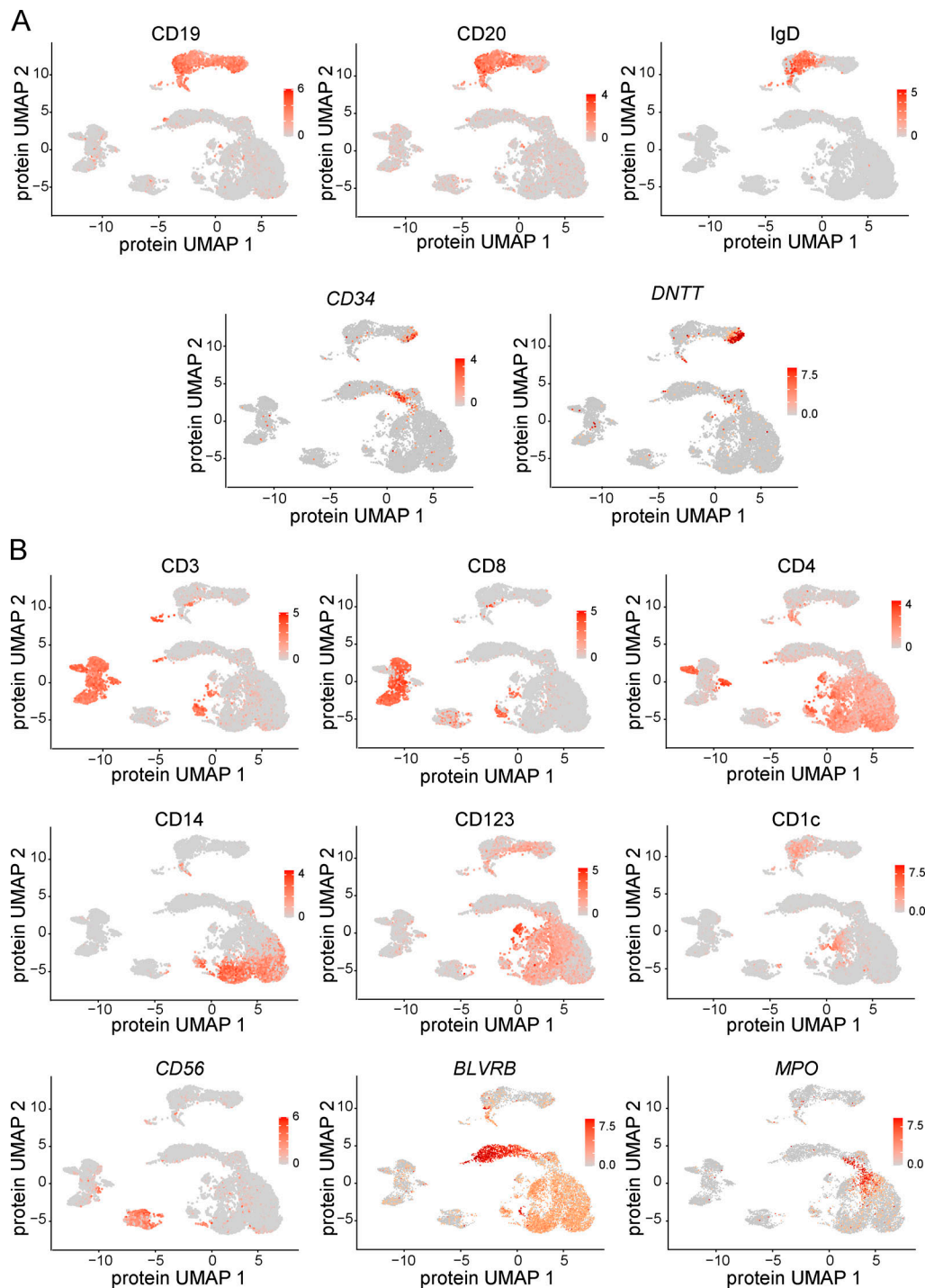


Figure S3. **Protein and transcript heat maps of bone marrow cells.** (A) Heat maps show \log_2 expression of B cell-relevant proteins (CD19, CD20, IgD) and transcripts (*CD34* and *DNNT*) overlaid upon dimensionally reduced UMAP plots of merged HD and patient A.II.2 bone marrow mononuclear cells. (B) Heat maps show \log_2 expression of key proteins (CD3, CD8, CD4, CD14, CD123, CD1c, and CD56 proteins) and transcripts (*BLVRB* and *MPO* transcripts) used to label UMAP clustered cells. For labeling cells, types were defined as follows: pro-B: CD19⁺CD20⁻IgD⁻, pre-B1: CD19⁺CD34⁻CD20⁻, pre-B2: CD19⁺CD20⁺IgD⁻, immature and mature naive B: CD19⁺CD20⁺IgD⁺, memory B: CD19⁺CD27⁺, plasma cell (PC): *SDCL1*, $\gamma\delta$ T cell: TCRV δ , erythroid progenitor: *BLVRB*, common lymphoid progenitor (CLP): *MME*, hematopoietic stem cells (HSCs): *CD34*, granulocyte macrophage progenitor (GMP): *MPO*, cDC: CD1c, pDC: CD123, CD14⁺ monocytes: CD14. CD16⁺ monocytes: CD16. NK and NKT: CD56. Activated CTL: CD8 and *CCL5*, resting CTL: CD8 and *CCR7*, CD4 T cell: CD4, T reg cell: *FOXP3*. NKT, natural killer T cell.

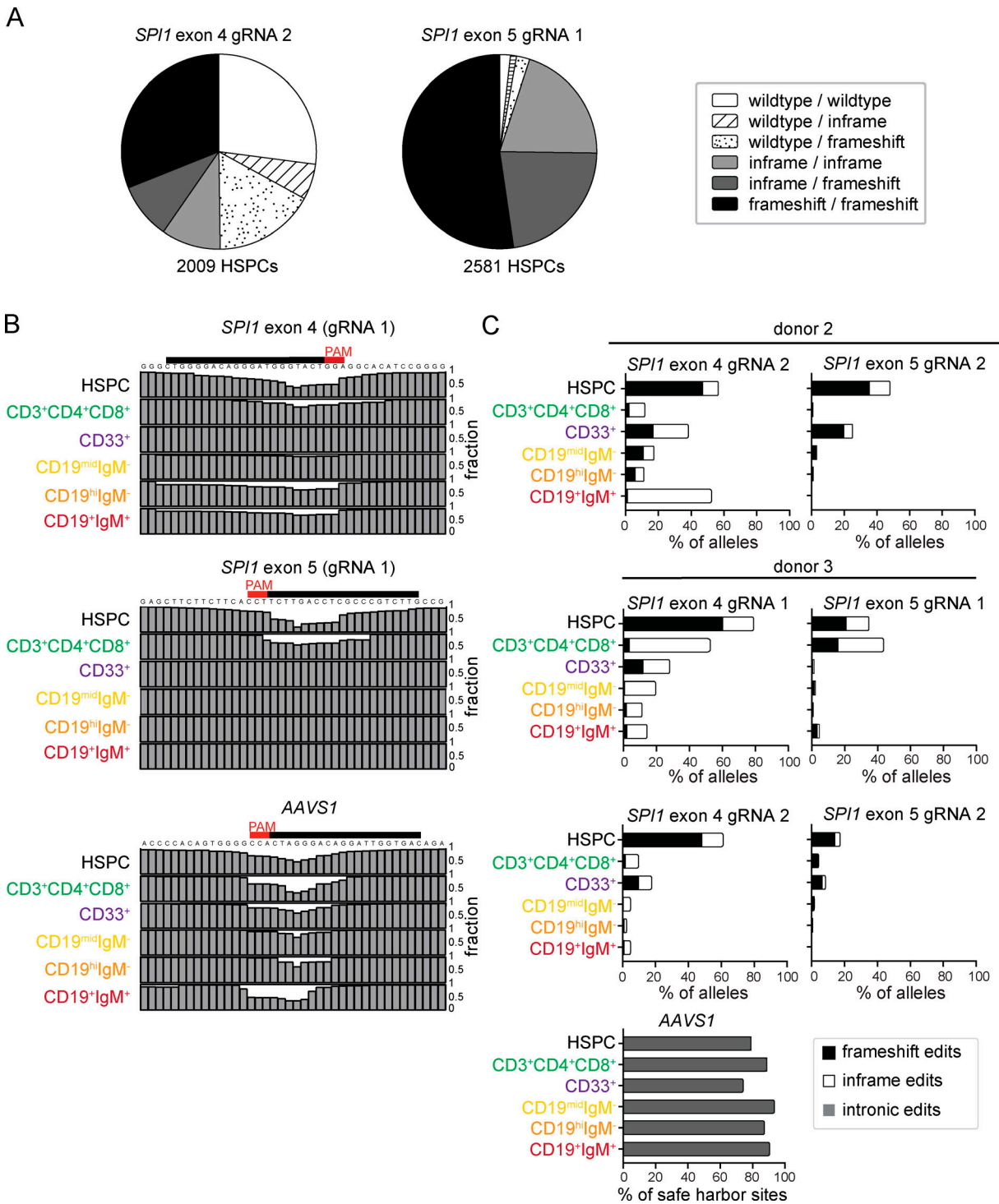


Figure S4. *SPI1* edits before and after in vitro differentiation. **(A)** *SPI1* allele distributions (based on coding sequence length) determined by single-cell genomic DNA amplicon sequencing. 2,009 HSPCs from cord blood donor one were analyzed at Exon 4. 2,581 HSPCs were analyzed at Exon 5. **(B)** The fractions of amplicon reads, in indicated donors and cell types, correctly aligned to reference are indicated with single-nucleotide resolution (gray rectangles) for each genomic target region. gRNAs used to introduce edits are shown aligned with the protospacer adjacent motifs (PAMs) at each cut site. Figure depicts cells from cord blood donors two and three. **(C)** Frequencies of in-frame (white), frame-shifting (black), and intronic (gray) genomic edits introduced into *SPI1* exons and AAVS1 using indicated guides are displayed.

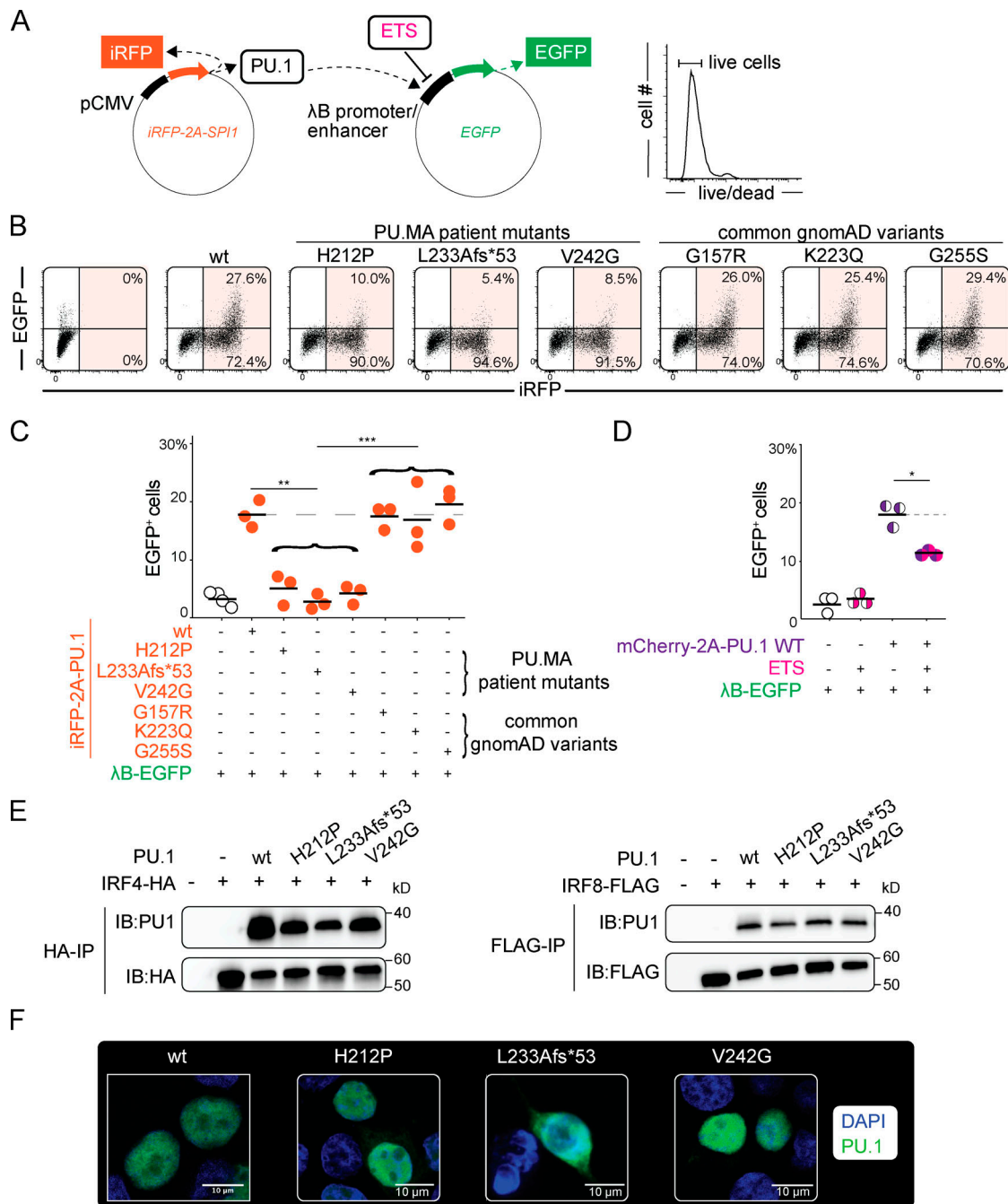


Figure S5. Functional assessment of mutant PU.1 proteins in HEK293 cells. (A) Ectopic PU.1 activity was measured in a HEK293 reporter cell line containing EGFP under the transcriptional control of a PU.1-dependent, λ B-based promoter/enhancer nucleotide sequence. An ETS fragment competes with PU.1 for promoter binding sites. (B) Representative flow cytometry dot plots from a single experiment displays frequencies of live, EGFP-positive cells successfully transfected with WT or mutated SPI1. Transfection status determined by iRFP expression (orange shading). (C and D) The frequencies of EGFP⁺ HEK293 reporter cells expressing either WT PU.1, PU.MA-associated PU.1 mutants, or common PU.1 variants listed in gnomAD (C) or coexpressing WT PU.1 and an ETS fragment (D). The mean allelic frequencies of selected gnomAD variants were all $>1.5 \times 10^{-4}$. Data from one of three representative experiments. (E) HA-tagged IRF4 and FLAG-tagged IRF8 similarly coimmunoprecipitated with WT and patient-associated H212P, L233Afs*53, and V242G PU.1. Data are representative of two independent experiments. (F) Representative confocal images of anti-PU.1 and DAPI-stained HEK293 cells overexpressing the indicated PU.1 proteins are shown. Images from one of two representative experiments. *, $P < 0.05$; **, $P < 0.01$; and ***, $P < 0.001$ (paired Student *t* tests). IP, immunoprecipitation.