



A frameshift variant in specificity protein 1 triggers superactivation of Sp1-mediated transcription in familial bone marrow failure

Hemanth Tummala^{a,1}, Amanda J. Walne^a, Findlay Bewicke-Copley^{b,c}, Alicia Ellison^a, Nikolas Pontikos^a, Maria G. Bridger^a, Ana Rio-Machin^b, Jasmin K. Sidhu^a, Jun Wang^c, Henrik Hasle^d, Jude Fitzgibbon^b, Tom Vulliamy^{a,2}, and Inderjeet Dokal^{a,2}

^aCentre for Genomics and Child Health, Blizard Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, E1 2AT London, United Kingdom; ^bCentre for Cancer Genomics and Computational Biology, Barts Cancer Institute, Queen Mary University of London, EC1M 6BQ London, United Kingdom; ^cCentre for Molecular Oncology, Barts Cancer Institute, Queen Mary University of London, EC1M 6BQ London, United Kingdom; and ^dDepartment of Pediatrics, Aarhus University Hospital, 8200 Aarhus, Denmark

Edited by Robert Tjian, University of California, Berkeley, CA, and approved June 9, 2020 (received for review February 17, 2020)

Inherited bone marrow failure (BMF) syndromes are a heterogeneous group of diseases characterized by defective hematopoiesis and often predisposing to myelodysplastic syndrome (MDS) and acute myelogenous leukemia. We have studied a large family consisting of several affected individuals with hematologic abnormalities, including one family member who died of acute leukemia. By whole-exome sequencing, we identified a novel frameshift variant in the ubiquitously expressed transcription factor specificity protein 1 (SP1). This heterozygous variant (c.1995delA) truncates the canonical Sp1 molecule in the highly conserved C-terminal DNA-binding zinc finger domains. Transcriptomic analysis and gene promoter characterization in patients' blood revealed a hypermorphic effect of this Sp1 variant, triggering superactivation of Sp1-mediated transcription and driving significant up-regulation of Sp1 target genes. This familial genetic study indicates a central role for Sp1 in causing autosomal dominant transmission of BMF, thereby confirming its critical role in hematopoiesis in humans.

bone marrow failure | transcription | Sp1

Over the last three decades, the genetic basis of many of the classic bone marrow failure (BMF) syndromes, such as Fanconi anemia, Diamond Blackfan anemia, Shwachman Diamond syndrome, and dyskeratosis congenita, have been elucidated (1). However, there are still families in which two or more first-degree relatives have hematologic abnormalities and whose biological basis remains unknown. Here we report a large family with several cases of cytopenias and one case of acute leukemia (Fig. 1A). The index case was a 16-y-old boy (Fig. 1A, III-3) who presented with recurrent nosebleeds and several nonsevere infectious episodes. He had leukopenia, variable neutropenia, and thrombocytopenia (Fig. 1A and *SI Appendix, Table S1*). His bone marrow showed reduced cellularity but normal trilineage morphology and karyotype. His mother, age 47 y, also had leukopenia, variable neutropenia, thrombocytopenia, and macrocytosis (Fig. 1A and *SI Appendix, Table S1*). The hematologic abnormalities were first documented in his mother (II-3) around age 20 y, and her bone marrow examination was considered morphologically normal. His maternal aunt (II-1) had died from acute leukemia at age 19 y.

Results

A Frameshift Variant in SP1 Segregates with Hematologic Disease. To investigate a possible genetic basis for the hematologic disease in this family, we performed whole-exome sequencing and variant calling (2) on the index case and his mother. No variants were detected in any of the known BMF genes or those known to cause familial thrombocytopenia/MDS/leukemia. However, we did identify a previously unreported heterozygous deletion (c.1995delA) in the specificity protein 1 gene (*SP1*) that was shared by the index case and his mother. There are several remarkable things about this

variant, notably the fact that there is just one other loss-of-function (LoF) *SP1* variant reported in the Genome Aggregation Database (gnomAD). Several other novel heterozygous variants in other genes were identified to be shared between mother and index case (*SI Appendix, Table S2*), but all of these variants have lower combined annotation-dependent depletion (CADD) scores compared with the c.1995delA variant in *SP1* (3). Sanger sequencing confirmed the presence of the *SP1* variant in the DNA and RNA (Fig. 1B) from whole blood and Epstein-Barr virus (EBV)-transformed lymphoblastoid cell lines (LCLs) of patients (II-3 and III-3). The variant allele frequency observed in the cDNA was 0.46, indicating no loss of expression of the frameshift allele by nonsense-mediated RNA decay (NMD). Segregation analysis showed that his two half-brothers (III-1 and III-2), both of whom had thrombocytopenia, as well as his maternal grandmother (I-1), who had macrocytosis, were also heterozygous for c.1995delA (*SI Appendix, Table S1*). Although there remains a possibility that the disease results from a different mutation, we find it compelling that this

Significance

Bone marrow failure (BMF) syndromes are inherited life-threatening conditions characterized by low blood cell production and predisposition to cancer. In this study we report a germ line frameshift variant in the Sp1 transcription factor in a family of patients with BMF and acute leukemia. Sp1 is ubiquitously expressed in human tissues and regulates transcription for blood cell lineage specification. Dissecting the molecular function of this SP1 variant revealed a hypermorphic effect, triggering superactivation of Sp1-mediated transcription in the patients' blood. To our knowledge, this is the first report of a naturally occurring germ line variant in SP1 that alters transcriptional networks and disrupts hematopoiesis in humans.

Author contributions: H.T., J.F., T.V., and I.D. designed research; H.T., A.J.W., A.E., M.G.B., A.R.-M., and J.K.S. performed research; H.H. and I.D. contributed new reagents/analytic tools; H.T., A.J.W., F.B.-C., N.P., M.G.B., A.R.-M., J.W., H.H., J.F., T.V., and I.D. analyzed data; and H.T., T.V., and I.D. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Data deposition: Original RNA sequencing files are deposited in National Center for Biotechnology Information Gene Expression Omnibus database (<https://www.ncbi.nlm.nih.gov/geo/info/seq.html>) and available under accession no. [GSE152262](https://www.ncbi.nlm.nih.gov/geo/info/seq.html).

¹To whom correspondence may be addressed. Email: h.tummala@qmul.ac.uk.

²T.V. and I.D. contributed equally to this work.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2002857117/-DCSupplemental>.

First published July 7, 2020.

highly unusual *SP1* variant is associated with dominant transmission of variable hematologic abnormalities in several members of the family.

Truncated Sp1 Variant Alters WT Sp1 Regulation in Patient Cells. Sp1 belongs to the specificity protein (Sp)/Krüppel-like family of transcription factors and binds the G-rich (5'(G/T)GGGCGG(G/A)(G/A)(C/T)-3') DNA motif to regulate gene expression (4). It is ubiquitously expressed (5) and regulates generic processes, such as cell cycle progression, metabolism, and apoptosis (6). Interestingly, homozygous genetic ablation of *Sp1* in mice is embryonic lethal (7), and heterozygous *Sp1*^{+/-} mice are normal (8). However, combined conditional knockout of *Sp1* and *Sp3* in adult mice revealed specific defects in hematopoietic terminal differentiation, especially in megakaryopoiesis (9). The C-terminal region of human Sp1 consists of three Cys-2-His-2 zinc finger motifs that are involved in DNA binding (10, 11), two of which would be lost in the truncated Sp1 protein predicted to result from the c.1995delA variant (p.K665Nfs*32) (Fig. 1C). Previous studies have shown that abolition of Sp1 DNA-binding activity in mice causes progressive impairment of hematopoiesis (12). Immunoblotting of patient cell lysates clearly indicated the presence of both wild-type (WT) and truncated Sp1 proteins at reduced levels compared with an age-matched control using an antibody targeting

the N-terminal region of Sp1 protein (Fig. 1D). The truncated Sp1 variant appears to be reduced in intensity compared to WT Sp1, indicating its less stable in the patient cell lysates (Fig. 1D). C-terminal targeted antibodies showed reduced levels of WT Sp1 protein in patient cell lysates compared with controls (Fig. 1D). The specificity of both antibodies was verified by immunoblotting of chromatin extracts prepared from HEK 293 cells treated by *SP1* transcript-specific siRNA (Fig. 1E).

We also noted a differential pattern of bands for the WT Sp1 molecules in control cell lysates compared with patients (Fig. 1D). In mammalian cells, Sp1 turnover is maintained by various post-translational modifications, including phosphorylation and subsequent degradation by sumoylation (13–15). Therefore, the observed pattern of bands for WT Sp1 in control cells may reflect its posttranslational modifications and subsequent degradation via the ubiquitin proteasomal pathway. An absence of this pattern in patient cells suggests that the presence of the truncated Sp1 variant has altered WT Sp1 regulation.

Using an enzyme-linked immunosorbent assay (ELISA) that targets an Sp1 epitope between residues 520 and 534, which is present in both WT and truncated forms of Sp1, DNA-binding ability appeared to be reduced in the nuclear extracts of patient cells compared with control (Fig. 1F). Collectively, these studies indicate that the presence of a less stable, truncated Sp1 variant

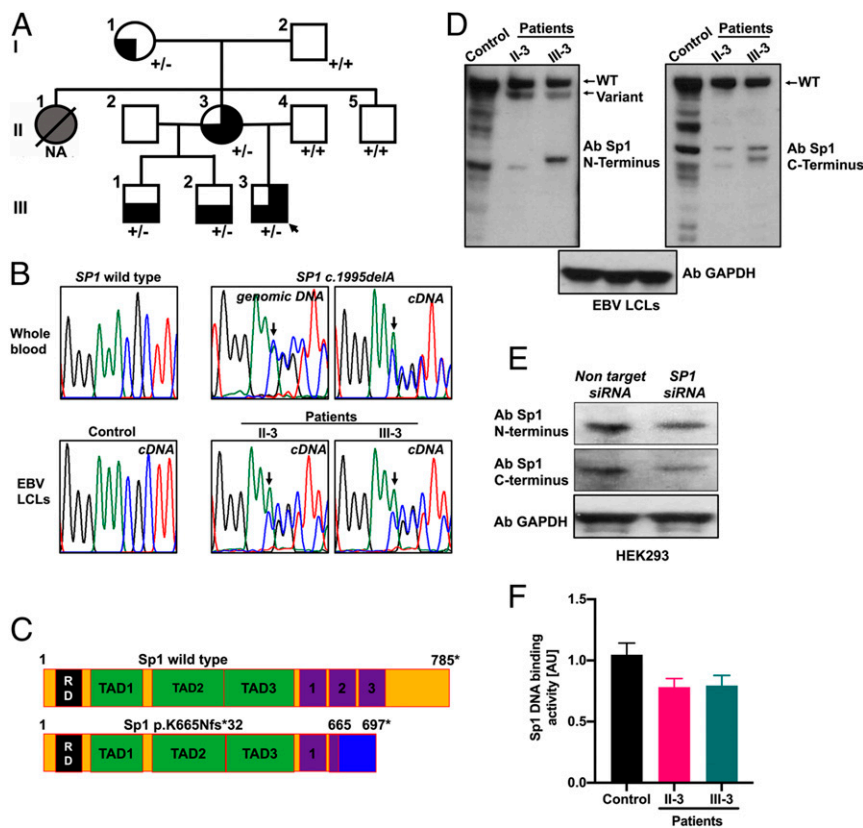


Fig. 1. A frameshift variant in *SP1* alters WT Sp1 regulation. (A) The family tree shows the presence of heterozygous variant in *SP1* c.1995delA by a filled left lower half quadrant in the index case (arrowhead), his half-brothers, mother, and maternal grandmother. The presence of thrombocytopenia is indicated by a filled right lower quadrant. The presence of leukopenia with variable neutropenia is indicated by a filled top right quadrant. The presence of acute leukemia is indicated by gray shading. Genotypes for the *SP1* variant are indicated as +/- for heterozygous and +/+ for WT. NA refers to an unknown genotype due to sample unavailability. (B) Sanger sequencing traces showing the WT trace from normal control and a heterozygous deletion of residue A at the position marked by the arrow in both genomic DNA and cDNA samples of patient blood and patient-specific EBV-transformed LCLs. (C) The predicted consequence of the *SP1* variant on protein structure (p.K665Nfs*32). The three zinc finger motifs are shown as purple boxes. The blue shaded box shows the extent of the frameshifted amino acids. RD, repressor domain; TAD, transactivation domain. (D) Immunoblotting of EBV-transformed lymphoblastoid cell extracts showing WT Sp1 expression levels in controls compared with both WT and the reduced truncated allele expression only in the patient cells. (E) Antibodies targeting N-terminal and C-terminal regions verified by immunoblotting of nuclear extracts from HEK293 cells treated with *SP1* specific siRNA. (F) ELISA analysis of Sp1 DNA-binding affinity in control and patient cell nuclear extracts.

alters WT Sp1 regulation, compromising Sp1 DNA-binding activity in these patient cells.

Superactivation of Sp1-Mediated Transcription in Patient Cells. To assess the impact of the truncated Sp1 variant on Sp1-mediated transcription, we performed RNA sequencing (RNA-seq) of white blood cell RNA in affected cases (I-1, III-2, and III-3) and WT individuals (I-2 in duplicate and II-5) from the family. Principal component analysis of this RNA-seq data revealed a clear separation of gene expression signatures of the three cases from their WT relatives (Fig. 2A). A total of 1,247 genes were differentially expressed with a false discovery rate (FDR) <0.01, of which 1,071 genes were up-regulated in the cases compared with their WT relatives, giving rise to a markedly skewed volcano plot (Fig. 2B and Dataset S1). A subset of these up-regulated genes was validated by qRT-PCR (Fig. 2C). Gene set enrichment analysis (GSEA) showed significant up-regulation of genes involved in signaling pathways, notably for platelet activation signaling and aggregation (Fig. 2D and Dataset S2). The enrichment seen for up-regulated genes increased as the FDR decreased and was closely mirrored when the search was restricted to a list of Sp1 target genes derived from ENCODE chromatin immunoprecipitation sequencing datasets (SI Appendix, Table S3 and Dataset S3). In addition, transcription factor-binding analysis (HOMER; ref. 16) identified the Sp1 consensus binding site as the most significantly enriched motif in the promoters of the 1,071 up-regulated genes (Fig. 2E).

To cross-examine this finding of enhanced Sp1-mediated transcription in patients' blood, we interrogated control and patient LCLs by transduction with lentiviral particles that report on Sp1-mediated transcription. In cells cultured under puromycin selection to generate stable reporter lines, we analyzed Sp1-mediated transcription by monitoring the amount of green fluorescent protein signal that is driven by Sp1 response elements (Fig. 2F). This analysis revealed increased Sp1-mediated transcription in both patient LCLs compared with control, as determined by the GFP expression level using fluorescence microscopy and Western blot analysis (Fig. 2G and H). This finding is reminiscent of previous studies in vitro that showed the superactivation of WT Sp1 in the presence of zinc fingerless Sp1 molecules (16, 17).

To ascertain whether the up-regulation of Sp1 target genes in the patients' blood was due to the presence of the germ line *SP1* variant, we conducted a parallel analysis on RNA-seq datasets acquired from patients with BMF without an *SP1* genetic lesion (SI Appendix, Table S4). Based on the blood counts, we classified these samples into two distinct categories, termed "full-BMF," in which all three lineages are affected, and "partial-BMF," which have variable degrees of BMF and were specifically selected to match the blood counts seen in the *SP1* heterozygotes of the family (SI Appendix, Table S4). A pairwise comparison of these datasets with control samples revealed a completely different pattern of gene dysregulation (FDR <0.01) compared with that seen in the *SP1* family (SI Appendix, Fig. S1A and B). Neither the full-BMF nor the partial-BMF datasets showed any skewing toward up-regulation and although they overlap one another to some extent, the number of genes overlapping the *SP1* family data were minimal (SI Appendix, Fig. S1A and B). Furthermore, the HOMER analysis showed no significant enrichment for the SP1-binding motif among the up-regulated genes in these samples (SI Appendix, Fig. S1C). We conclude that the increase in Sp1-mediated transcription in the affected individuals of the family appears to be specific to the presence of the germ line *SP1* variant and not a secondary or a reactive effect to the cytopenias. Collectively, these studies demonstrate that the truncated Sp1 protein that we have identified triggers superactivation of Sp1-mediated transcription in the affected individuals, leading to dysregulated hematopoiesis and ultimately BMF.

Discussion

We report on a family with autosomal dominant transmission of BMF in affected individuals caused by a germ line frameshift variant in the ubiquitously expressed transcription factor Sp1 (Fig. 1). The identified *SP1* variant does not undergo premature NMD, as the truncated variant was readily detectable in both RNA and protein of patient cells (Fig. 1B and D). This is likely due to the presence of c.1995delA variant in the penultimate exon of the *SP1* gene, giving rise to a premature termination codon in the final exon. This is not predicted to trigger NMD (18). In contrast, the only other LoF variant reported on gnomAD (p.Val518GlufsTer15) is predicted to undergo NMD based on the same rules, as the variant is located upstream in exon 3 of *SP1*, resulting in expression of only the WT allele. In this context, we note that genetic ablation of *SP1* allele in the heterozygous state in mice and mouse embryonic stem cells had no profound effect on hematopoiesis (6–8, 12).

Sp1 undergoes various posttranslational modifications, including phosphorylation at the threonine 739 residue to prevent its degradation (13–15, 19). The truncated Sp1 variant that we have identified lacks this phosphorylation site and may be targeted for early degradation by binding to SUMO (13–15). The presence of residual truncated Sp1 protein also appeared to reduce Sp1 DNA-binding affinity in patient cells, presumably due to loss of the zinc finger domains. The residual truncated Sp1 protein that we observe in the patient cells is highly reminiscent of a superactivating Sp1 protein previously studied in vitro (16, 17). In these studies, it was shown that Sp1 can form multimeric complexes and that WT Sp1 could be transcriptionally enhanced through interaction with a fingerless Sp1. More recently, it has been demonstrated that the low-complexity domains of Sp1 allow the formation of regulatory hubs which can recruit the transcription elongation machinery without binding to DNA (20).

We view these in vitro experiments as highly relevant to the stark increase in Sp1-mediated transcription that we have observed in the patients' blood. We propose that an enhanced multimerization of the mutant Sp1 is the likely mechanism driving the Sp1 superactivation. The extensive further work needed to establish this point is beyond the scope of the present study, in which we report on the impact of the first human disease-associated Sp1 variant.

It has been established elsewhere that transcription activation by Sp1 exerts a hierarchical control on blood cell development (5, 12, 21, 22). In mice, Sp1-mediated transcriptional dosage is required for hematopoietic cell lineage specification of megakaryocyte differentiation and platelet production (12). Recent transcriptomic studies on differentiated mouse embryonic stem cells expressing Sp1 mutant that lack DNA binding (Sp1^{ΔDBD/ΔDBD}) showed severely disturbed hematopoietic cell differentiation trajectories (22). The affected cases of this family presented with variable cytopenias, including thrombocytopenia and leukopenia, indicating a variable effect exerted by the truncated Sp1 variant in hematopoietic specification and differentiation. Irrespective of the precise mechanism, it is clear that from mice to human, *SP1* lesions alter transcriptional networks that regulate bona fide hematopoiesis.

We find it intriguing that the aunt of this family's index case died of leukemia at 19 y of age. Although we were unable to genotype this individual, the occurrence of hematologic malignancy in this family is reminiscent of thrombocytopenia/leukemia phenotypes associated with mutations in RUNX1 (23), CEBPA (24), ANKRD26 (25), and ETV6 (26). Transcription factor mutations are known to perturb gene networks that regulate hematopoiesis and drive clonal neoplasms in blood (27). Sp1, ETV6, RUNX1, and ANKRD26 are all expressed in hematopoietic stem cells and megakaryocyte-erythroid progenitors (28–30). Sp1, along with MYC, RUNX1, and GATA2, have been

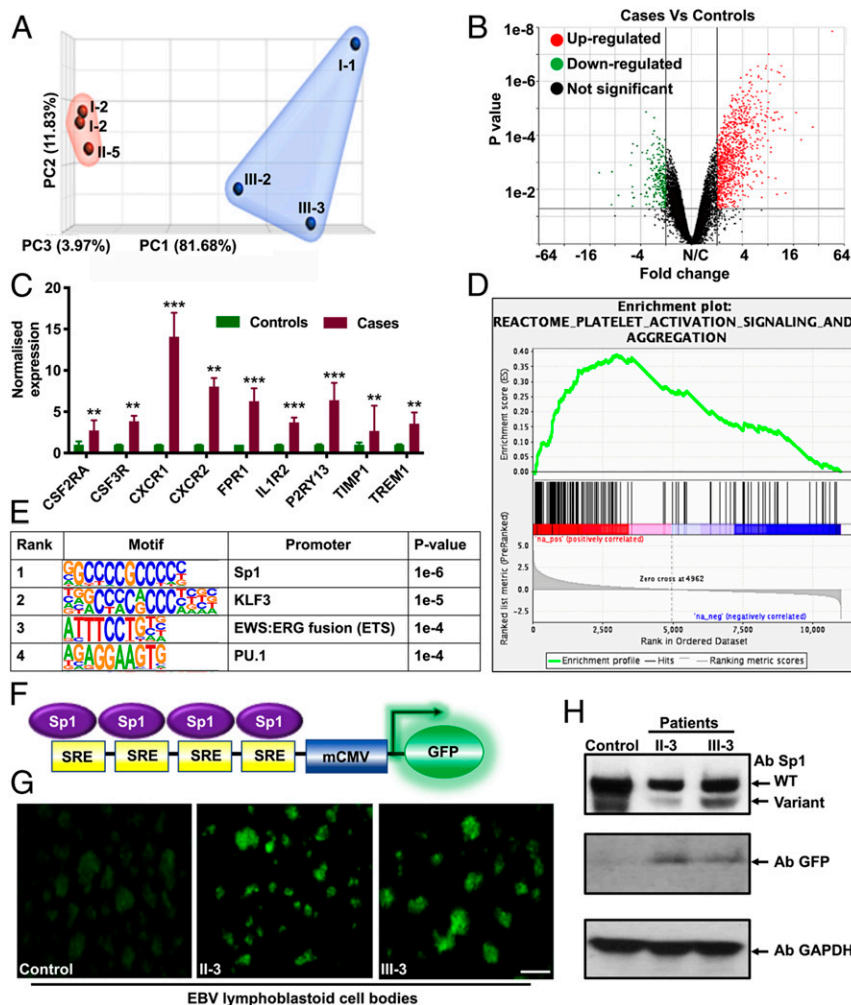


Fig. 2. Up-regulation of Sp1-mediated gene expression signature in patients' white blood cells. (A) Principal component analysis of RNA-seq expression data showing separation of three cases (I-1, III-2, and III-3) from the family controls [I-2 (dup), II-5]. (B) Volcano plot display of differentially expressed transcripts in cases ($n = 3$) carrying heterozygous *SP1* variant (c.1995delA) versus WT ($n = 3$) family controls. (C) qRT-PCR indicating that cases carrying the heterozygous *SP1* variant (c.1995delA) show up-regulation of the selected transcripts compared with WT in blood. The expression of each gene after normalization to three control genes (*USB1*, *CSNK2B*, and *GAPDH*) is shown. Bars represent the median expression ratio. * $P < 0.05$; ** $P < 0.01$, Mann-Whitney U test. (D) GSEA showing significant up-regulation of platelet activation signaling and aggregation. (E) Transcription factor binding site enrichment detected using HOMER, in the promoters of 1,071 up-regulated genes in patients' white blood cells with a threshold FDR < 0.01 . (F) Schematic representation of Sp1 transcriptional activation using a lentiviral-based reporter system (pGreenFire) that constitutes plasmid-containing Sp1 response elements (SRE) under the control of the minimal cytomegalovirus (mCMV) promoter that drives GFP expression on Sp1 binding. (G and H) GFP expression levels in EBV-transformed lymphoblastoid cell clusters from controls and the two patients stably expressing pGreenFire lentivirus as assessed by fluorescence microscopy and Western blot analysis. GAPDH levels were determined for the loading control. (Scale bar: 500 μm .)

identified as key transcriptional regulators driving oncogene expression in acute myelogenous leukemia (31). Taken together, these observations indicate an association of Sp1 function in cancer predisposition, but to date its role in malignant transformation remains unclear.

In summary, we report a novel frameshift variant in the ubiquitous transcription factor *SP1* segregating with hematologic disease. This is a naturally occurring germ line *SP1* variant associated with a human Mendelian disorder. Functional studies on patient cells revealed a significant up-regulation of gene expression that is driven to a large extent by Sp1, suggesting a molecular basis for the hematologic disease in this family.

Methods

Patient Samples, Sequencing, and Bioinformatics. All patient samples were obtained with written consent under the approval of our local Research Ethics Committee (London-City and East). Exome sequencing was performed on the Illumina HiSeq 2000 platform using the Agilent SureSelect XT All Exon

V5 enrichment kit. Exome data were processed and analyzed using Phenopolis (<https://phenopolis.org>) as described previously (32). All relevant variants identified were validated by Sanger sequencing. For white blood cell transcriptomic analysis, blood was collected in EDTA tubes following venipuncture. RNA was extracted using the QIAmp RNA Blood Mini Kit (Qiagen) following the manufacturer's instructions. The quality and quantity of RNA were assessed using the Agilent 2100 Bioanalyzer system. Library preparation was performed using the NEBNext Ultra II RNA Library Prep Kit for Illumina with optional mRNA enrichment and NEBNext multiplexing oligos (New England BioLabs) following the manufacturer's protocol. RNA sequencing was done on the Illumina HiSeq 4000 platform. The reads were aligned to hg38 using the HISAT aligner. Raw gene counts were obtained using HTSeq (33), and differential expression was carried out using edgeR (34) according to the documentation. Significantly deregulated genes were identified as those that had an FDR < 0.01 . The whole gene list was ranked in descending order of \log_2 fold change in the *SP1* heterozygotes, and GSEA was carried out using the GSEA preranked module on the GenePattern (35) platform using the KEGG and Reactome gene sets.

Selected genes were analyzed by qRT-PCR using a relative standard curve method by serially diluting a sample of known input concentration. The relative amplification efficiencies of each gene were compared and found to be between 90% and 110%. The geometric mean for three control genes (*USB1*, *CSNK2B*, and *GAPDH*) was used to normalize the levels of each gene of interest (GOI). These control genes were selected based on the fact that they were expressed at similar levels in all samples. The relative amount of each GOI was expressed as a ratio to the combined levels for the normalizer genes for its own sample. The final fold change was determined by the ratio of the means for the WT and case samples. HOMER analysis was carried out on the list of significantly deregulated genes using the findMotifs.pl script. The genes were passed to HOMER as a file of gene names, and the human promoter set was selected.

Cell Culture, Immunoblotting, and Sp1 DNA-Binding Activity Analysis. EBV-transformed control and patient LCLs were cultured in standard RPMI 1640, 15% FCS, and 10% antibiotics. Control and patient LCL lysates were separated on sodium dodecyl sulfate polyacrylamide gel electrophoresis and probed using antibodies against Sp1 (Merck; 07-645) and the C terminus of Sp1 (Abcam; 13370). GAPDH (Abcam 13370) and TATA binding protein antibody (Abcam 818) served as loading controls. EBV-transformed lymphoblastoid cells of controls and patients were lysed in nuclear extraction lysis buffer supplied in the Sp1 transcription factor assay ELISA kit (Abcam 207226). The lysates were centrifuged at 14,000 rpm at 4 °C and subsequently used following the instructions in the kit.

Sp1 Transcription Reporter Assay. The lentiviral-based Sp1 transcription reporter system pGreenFire was obtained from System Biosciences (catalog no. TRO36PA-P). In brief, equal amounts of infection units encoding lentiviral particles of Sp1 response elements were transduced into control and patient lymphoblastoid cells via spinfection at 2,500 rpm for 90 min. After transduction, cells were cultured in RPMI medium supplemented with 10% FCS and antibiotics for 96 h. Stable cell lines were established under puromycin selection (2 µg/mL) for 2 wk and subsequently analyzed by fluorescence microscopy and Western blot analysis.

Statistical Analysis. Quantitative analysis results are presented as mean ± SD (or SEM) from repeated experiments as indicated in the figure legends. The pairwise Student's *t* test was used to analyze statistical significance, except for the qRT-PCR results, which were analyzed using the Mann–Whitney *U* test.

Data Sharing Statement. Original RNA sequencing files are deposited in National Center for Biotechnology Information Gene Expression Omnibus database (<https://www.ncbi.nlm.nih.gov/geo/info/seq.html>) and available under accession no. GSE152262. Differentially regulated gene set data are provided in online [Datasets S1 to S3](#).

ACKNOWLEDGMENTS. We thank the families and clinicians who contributed to this research. This work was supported by grants from the Medical Research Council (MR/P018440/1) and Blood Cancer UK (14032).

1. J. Collins, I. Dokal, Inherited bone marrow failure syndromes. *Hematology* **20**, 433–434 (2015).
2. H. Tummala *et al.*, Genome instability is a consequence of transcription deficiency in patients with bone marrow failure harboring biallelic *ERCC6L2* variants. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 7777–7782 (2018).
3. P. Rentzsch, D. Witten, G. M. Cooper, J. Shendure, M. Kircher, CADD: Predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* **47**, D886–D894 (2019).
4. W. S. Dynan, R. Tjian, Isolation of transcription factors that discriminate between different promoters recognized by RNA polymerase II. *Cell* **32**, 669–680 (1983).
5. S. H. Orkin, Transcription factors and hematopoietic development. *J. Biol. Chem.* **270**, 4955–4958 (1995).
6. G. Suske, E. Bruford, S. Philipsen, Mammalian SP/KLF transcription factors: Bring in the family. *Genomics* **85**, 551–556 (2005).
7. M. Marin, A. Karis, P. Visser, F. Grosveld, S. Philipsen, Transcription factor Sp1 is essential for early embryonic development but dispensable for cell growth and differentiation. *Cell* **89**, 619–628 (1997).
8. I. Krüger *et al.*, Sp1/Sp3 compound heterozygous mice are not viable: Impaired erythropoiesis and severe placental defects. *Dev. Dyn.* **236**, 2235–2244 (2007). Erratum in: *Dev. Dyn.* **236**, 2970 (2007).
9. M. Meinders *et al.*, Sp1/Sp3 transcription factors regulate hallmarks of megakaryocyte maturation and platelet formation and function. *Blood* **125**, 1957–1967 (2015).
10. J. T. Kadonaga, K. R. Carner, F. R. Masiarz, R. Tjian, Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. *Cell* **51**, 1079–1090 (1987).
11. J. Letovsky, W. S. Dynan, Measurement of the binding of transcription factor Sp1 to a single GC box recognition sequence. *Nucleic Acids Res.* **17**, 2639–2653 (1989).
12. J. Gilmour *et al.*, A crucial role for the ubiquitously expressed transcription factor Sp1 at early stages of hematopoietic specification. *Development* **141**, 2391–2401 (2014).
13. J. Li, W. X. Zou, K. S. Chang, Inhibition of Sp1 functions by its sequestration into PML nuclear bodies. *PLoS One* **9**, e94450 (2014).
14. L. Gong *et al.*, Sumoylation differentially regulates Sp1 to control cell differentiation. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 5574–5579 (2014).
15. Y. T. Wang, W. B. Yang, W. C. Chang, J. J. Hung, Interplay of posttranslational modifications in Sp1 mediates Sp1 stability during cell cycle progression. *J. Mol. Biol.* **414**, 1–14 (2011).
16. A. J. Courey, D. A. Holtzman, S. P. Jackson, R. Tjian, Synergistic activation by the glutamine-rich domains of human transcription factor Sp1. *Cell* **59**, 827–836 (1989).
17. E. Pascal, R. Tjian, Different activation domains of Sp1 govern formation of multimers and mediate transcriptional synergism. *Genes Dev.* **5**, 1646–1656 (1991).
18. R. G. H. Lindeboom, M. Vermeulen, B. Lehner, F. Supek, The impact of nonsense-mediated mRNA decay on genetic disease, gene editing and cancer immunotherapy. *Nat. Genet.* **51**, 1645–1651 (2019).
19. H. D. Ulrich, Ubiquitin, SUMO, and phosphate: How a trio of posttranslational modifiers governs protein fate. *Mol. Cell* **47**, 335–337 (2012).
20. S. Chong *et al.*, Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science* **361**, eaar2555 (2018).
21. S. Heinz *et al.*, Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
22. J. Gilmour *et al.*, Robust hematopoietic specification requires the ubiquitous Sp1 and Sp3 transcription factors. *Epigenetics Chromatin* **12**, 33 (2019).
23. W. J. Song *et al.*, Haploinsufficiency of *CBFA2* causes familial thrombocytopenia with propensity to develop acute myelogenous leukaemia. *Nat. Genet.* **23**, 166–175 (1999).
24. M. L. Smith, J. D. Cavenagh, T. A. Lister, J. Fitzgibbon, Mutation of *CEBPA* in familial acute myeloid leukemia. *N. Engl. J. Med.* **351**, 2403–2407 (2004).
25. P. Noris *et al.*, Mutations in *ANKRD26* are responsible for a frequent form of inherited thrombocytopenia: Analysis of 78 patients from 21 families. *Blood* **117**, 6673–6680 (2011).
26. M. Y. Zhang *et al.*, Germline *ETV6* mutations in familial thrombocytopenia and hematologic malignancy. *Nat. Genet.* **47**, 180–185 (2015).
27. J. E. Churpek, E. H. Bresnick, Transcription factor mutations as a cause of familial myeloid neoplasms. *J. Clin. Invest.* **129**, 476–488 (2019).
28. A. Heydemann, G. Juang, K. Hennessy, M. S. Parmacek, M. C. Simon, The myeloid-cell-specific *c-fes* promoter is regulated by Sp1, PU.1, and a novel transcription factor. *Mol. Cell Biol.* **16**, 1676–1686 (1996).
29. K. D. Fischer, A. Haese, J. Nowock, Cooperation of GATA-1 and Sp1 can result in synergistic transcriptional activation or interference. *J. Biol. Chem.* **268**, 23915–23923 (1993).
30. M. Merika, S. H. Orkin, Functional synergy and physical interactions of the erythroid transcription factor GATA-1 with the Krüppel family proteins Sp1 and EKLf. *Mol. Cell Biol.* **15**, 2437–2447 (1995).
31. R. Pippa *et al.*, MYC-dependent recruitment of RUNX1 and GATA2 on the SET oncogene promoter enhances PP2A inactivation in acute myeloid leukemia. *Oncotarget* **8**, 53989–54003 (2016).
32. N. Pontikos *et al.*, UKIRDC, Phenopolis: An open platform for harmonization and analysis of genetic and phenotypic data. *Bioinformatics* **33**, 2421–2423 (2017).
33. S. Anders, P. T. Pyl, W. Huber, HTSeq—A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
34. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
35. M. Reich *et al.*, GenePattern 2.0. *Nat. Genet.* **38**, 500–501 (2006).