nature genetics

Germline mutations in *ETV6* are associated with thrombocytopenia, red cell macrocytosis and predisposition to lymphoblastic leukemia

Leila Noetzli[1,2,18], Richard W Lo[3,4,18], Alisa B Lee-Sherick[1], Michael Callaghan[5], Patrizia Noris[6], Anna Savoia[7,8], Madhvi Rajpurkar[5], Kenneth Jones[9], Katherine Gowan[9], Carlo L Balduini[6], Alessandro Pecci[6], Chiara Gnan[7,8], Daniela De Rocco[7,8], Michael Doubek[10], Ling Li[3], Lily Lu[3], Richard Leung[3], Carolina Landolt-Marticorena[11], Stephen Hunger[1], Paula Heller[12], Arthur Gutierrez-Hartmann[9,13], Liang Xiayuan[14], Fred G Pluthero[3], Jesse W Rowley[15,16], Andrew S Weyrich[15,16], Walter H A Kahr[3,4,17–19], Christopher C Porter[1,18,19] & Jorge Di Paola[1,2,18,19]

Some familial platelet disorders are associated with predisposition to leukemia, myelodysplastic syndrome (MDS) or dyserythropoietic anemia[1,2]. We identified a family with autosomal dominant thrombocytopenia, high erythrocyte mean corpuscular volume (MCV) and two occurrences of B cell–precursor acute lymphoblastic leukemia (ALL). Whole-exome sequencing identified a heterozygous single-nucleotide change in *ETV6* (*ets* variant 6), c.641C>T, encoding a p.Pro214Leu substitution in the central domain, segregating with thrombocytopenia and elevated MCV. A screen of 23 families with similar phenotypes identified 2 with *ETV6* mutations. One family also had a mutation encoding p.Pro214Leu and one individual with ALL. The other family had a c.1252A>G transition producing a p.Arg418Gly substitution in the DNA-binding domain, with alternative splicing and exon skipping. Functional characterization of these mutations showed aberrant cellular localization of mutant and endogenous ETV6, decreased transcriptional repression and altered megakaryocyte maturation. Our findings underscore a key role for ETV6 in platelet formation and leukemia predisposition.

ETV6 (also known as *TEL*) encodes a transcriptional repressor in the ETS family[3]. *ETV6* was initially identified as a tumor suppressor by its involvement in somatic translocations in childhood leukemia, including in *ETV6-RUNX1* fusion. These translocations often occur together with a somatic mutation in the intact *ETV6* allele, suggesting that loss of *ETV6* function contributes to the development of leukemia[4–6]. Somatic *ETV6* mutations have also been described in patients with MDS and T cell leukemias[7,8], but germline mutations had not been described until recently[9]. *ETV6* also has a key role in hematopoiesis, as demonstrated in animal models[10,11]. Mice with conditional *Etv6* knockout in megakaryocytic-erythroid cells are thrombocytopenic, and megakaryocyte colony formation is absent in homozygous knockout hematopoietic cells and decreased in heterozygous knockout cells, indicating the involvement of *Etv6* in thrombopoiesis[12,13]. ETV6 is also one of four transcription factors shown to be sufficient to differentiate mouse fibroblasts into hematopoietic-lineage cells[14].

We present here three families with germline mutations in *ETV6* and defects in hematopoiesis. Affected members (n = 5) in the original family from the United States (family 1; **Fig. 1a**) had variable thrombocytopenia (67,000–132,000 platelets/μl, normal range 150,000–450,000 platelets/μl) and elevated red blood cell MCV (92.5–101.5 fl, normal range 80–95 fl), suggesting a defect affecting megakaryocytic-erythroid precursors. Hematocrit and other hematologic indexes were within normal ranges (**Supplementary Table 1**). Platelets had normal mean volume and ultrastructure, with some elongated α-granules (**Supplementary Fig. 1**). Patients exhibited

[1]Department of Pediatrics, University of Colorado Anschutz Medical Campus (AMC), Aurora, Colorado, USA. [2]Human Medical Genetics and Genomics Program, University of Colorado AMC, Aurora, Colorado, USA. [3]Program in Cell Biology, Research Institute, The Hospital for Sick Children, Toronto, Ontario, Canada. [4]Department of Biochemistry, University of Toronto, Toronto, Ontario, Canada. [5]Children's Hospital of Michigan, Department of Pediatrics, Wayne State University, Detroit, Michigan, USA. [6]Department of Internal Medicine, Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS) Policlinico San Matteo Foundation, University of Pavia, Pavia, Italy. [7]Department of Medical Sciences, University of Trieste, Trieste, Italy. [8]Institute for Maternal and Child Health, IRCCS Burlo Garofolo, Trieste, Italy. [9]Department of Biochemistry and Molecular Genetics, University of Colorado AMC, Aurora, Colorado, USA. [10]Department of Internal Medicine, Haematology/Oncology, University Hospital Brno, Brno, Czech Republic. [11]Department of Medicine, Division of Rheumatology, University of Toronto, University Health Network, Toronto, Ontario, Canada. [12]Instituto de Investigaciones Médicas Alfredo Lanari, Universidad de Buenos Aires, Buenos Aires, Argentina. [13]Department of Medicine, University of Colorado AMC, Aurora, Colorado, USA. [14]Department of Pathology, University of Colorado AMC, Aurora, Colorado, USA. [15]Department of Internal Medicine, University of Utah, Salt Lake City, Utah, USA. [16]Molecular Medicine Program, University of Utah, Salt Lake City, Utah, USA. [17]Department of Paediatrics, Division of Haematology/Oncology, University of Toronto and The Hospital for Sick Children, Toronto, Ontario, Canada. [18]These authors contributed equally to this work. [19]These authors jointly supervised this work. Correspondence should be addressed to J.D.P. (jorge.dipaola@ucdenver.edu), C.C.P. (chris.porter@ucdenver.edu) or W.H.A.K. (walter.kahr@sickkids.ca).

Received 28 October 2014; accepted 25 February 2015; published online 25 March 2015; doi:10.1038/ng.3253

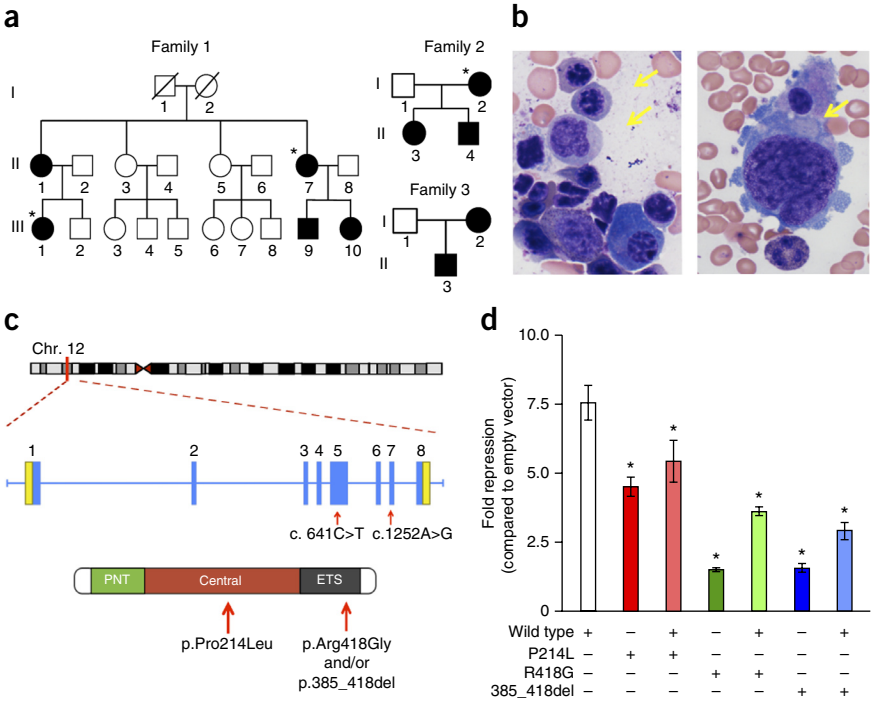


Figure 1 Mutation analysis of *ETV6*. (**a**) Pedigrees for three affected families. Filled symbols represent affected individuals with thrombocytopenia and high red blood cell MCV. Asterisks indicate individuals who developed B cell leukemia. (See **Supplementary Table 1** for complete blood count (CBC) values). (**b**) Representative images of a bone marrow aspirate for one of the affected individuals without leukemia, showing mild dyserythropoiesis (left; yellow arrows) and an immature, hypolobulated megakaryocyte (right; yellow arrow) (magnification 600×). (**c**) Top, schematic of *ETV6*, which is composed of eight exons encoding UTRs (yellow) and protein-coding sequences (blue). The two mutations identified, c.641C>T in exon 5 and c.1252A>G in exon 7, are depicted. Bottom, schematic of the ETV6 protein showing its different domains, pointed (PNT), central and ETS (the DNA-binding domain), and the location of the amino acid changes. (**d**) Effects of the *ETV6* mutant alleles (encoding p.Pro214Leu, p.Arg418Gly and p.385_418del) on the activity of an ETV6-responsive reporter plasmid (pGL2-754TR) when expressed alone or in combination with the wild-type allele. Luciferase activity, represented by fold repression, is shown relative to that observed with empty vector and normalized using an internal control plasmid expressing *Renilla* luciferase. Wild-type ETV6 repressed luciferase expression by approximately 7.5-fold in comparison to empty vector, whereas Pro214Leu, Arg418Gly and 385_418del ETV6 repressed transcription 4.5-, 1.5- and 1.5-fold, respectively (*$P < 0.0001$, one-way ANOVA). Coexpression of wild-type and mutant proteins caused intermediate repression of the *stromelysin-1* (*MMP3*) promoter. Experiments were performed in triplicate and repeated four times (error bars, s.d.).

mild to moderate bleeding, and two developed B cell–precursor ALL at ages 3 (III-1) and 37 (II-7) years. Histopathological analysis of bone marrow from affected individuals without leukemia showed small, hypolobulated megakaryocytes and abnormal red blood cell precursors (**Fig. 1b**).

Whole-exome sequencing of family 1 identified a heterozygous single-nucleotide change in *ETV6*, which segregated completely with thrombocytopenia and high red blood cell MCV. This missense variant, c.641C>T, encodes a proline-to-leucine amino acid change (p.Pro214Leu) in the central domain of ETV6 (NM_001987, P41212) (**Fig. 1c**); the variant was not found in the 1000 Genomes Project or dbSNP databases and was predicted to be possibly damaging by PolyPhen-2. We screened 23 additional European families with autosomal dominant thrombocytopenia, high red blood cell MCV and, in some cases, increased incidence of leukemia via Sanger sequencing of the exons and exon-intron boundaries of *ETV6*. This approach led to the identification of two additional families with germline *ETV6* mutations. One (family 2; **Fig. 1a**) had affected members with platelet counts of 44,000–115,000 platelets/μl, red blood cell MCV of 88–97 fl and ALL in individual I-2 at age 14 years, with all affected members exhibiting a c.641C>T mutation identical to the one identified in family 1. In the other family (family 3; **Fig. 1a**), having individuals with platelet counts of 99,000–101,000 platelets/μl and red blood cell MCV of 93–98 fl but no malignancies, we detected a mutation affecting the DNA-binding domain (c.1252A>G; p.Arg418Gly) that was not observed in the 1000 Genomes Project database and was predicted to be highly damaging by PolyPhen-2. Sequence alignment demonstrated that both mutations affect amino acids that are highly conserved across multiple species (**Supplementary Note**). Both mutations were present in the Catalogue of Somatic Mutations in Cancer (COSMIC) database, indicating that somatic acquisition of these mutations may be oncogenic.

The c.1252A>G mutation is located in the last codon of exon 7, which is split with exon 8, raising the possibility of this variant disrupting a splice site. To test this possibility, we isolated RNA from the peripheral blood cells of two individuals with the c.1252A>G mutation. RT-PCR detected two different transcripts, one of the expected size (386 bp) and another of 285 bp, indicating an alternatively spliced product (**Supplementary Fig. 2**). Sequencing of the 285-bp product showed skipping of exon 7. This c.1153_1253del mutation is predicted to cause a partial deletion of the putative DNA-binding domain (p.385_418del) and a subsequent p.Asn385Valfs*7 frameshift alteration resulting in the introduction of a premature stop codon, raising the possibility of a truncated protein. Although this truncated protein was expressed in transfection assays in HEK293T cells (**Supplementary Fig. 3**), it was not detected in platelets obtained from patients (**Supplementary Fig. 4**), suggesting that it is not functional in megakaryocytes. The alternative splicing event did not affect all mutant RNA, as sequencing of the 386-bp RT-PCR product, as well as plasmids into which this product was cloned, showed the presence of the G nucleotide in addition to the wild-type A nucleotide, indicating that the Arg418Gly form of ETV6 is likely to be expressed in patients from this family. This type of genetic aberration, in which an exonic mutation creates an alternative splicing site as well as an amino acid change, has been observed in other hematological disorders such as hemoglobin E disease[15].

ETV6 is a 57-kDa protein with 452 amino acids and 3 functional domains: the N-terminal pointed (PNT), central regulatory and C-terminal DNA-binding (ETS) domains (**Fig. 1c**). The nuclear localization and transcriptional repression activity of ETV6 require homodimerization via the pointed domain[16]. ETV6 modulates the activity of other ETS transcription factors such as FLI1—for which hemizygous deletions of the corresponding gene cause Paris-Trousseau–related thrombocytopenia[17,18]—and several other known ETV6 interaction partners that are present in platelets[19,20] and

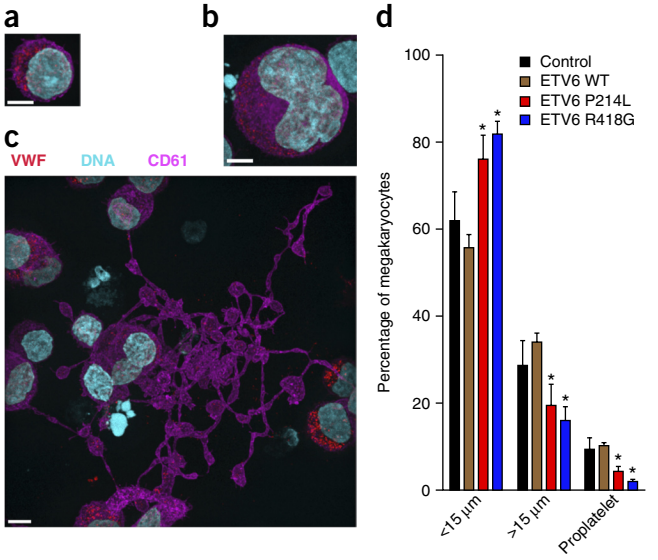
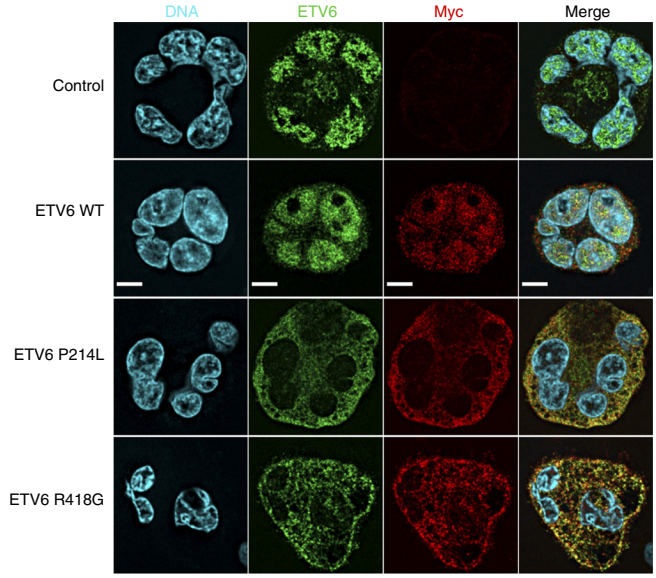


Figure 2 Abnormal development of cultured megakaryocytes expressing mutant ETV6 at day 12. Control cells and those transduced with lentivirus to express Myc-tagged forms of ETV6 (wild type, Pro214Leu and Arg418Gly) were assessed via immunofluorescence microscopy imaging; scale bars, 5 µm. (**a**–**c**) Megakaryocytes (control cells shown) were identified via expression of CD61 (magenta) and VWF (red) and staged by diameter, <15 µm (**a**) or >15 µm (**b**), or by the presence of proplatelets (**c**). (**d**) Population distributions showed no significant differences between control cells and those transduced with lentivirus expressing wild-type ETV6, whereas megakaryocytes transduced with lentivirus for Pro214Leu ETV6 showed a significantly higher proportion (*$P < 0.05$, two-tailed *t* test) of cells in the earlier developmental stage (<15 µm) and fewer cells in the mature, proplatelet-forming stage (control, $n = 7$; wild type (WT), $n = 3$; Pro214Leu, $n = 3$; Arg418Gly, $n = 4$ cultures, with >300 cells for each; error bars, s.d.). Images for control cells and those transduced with lentivirus expressing wild-type, Pro214Leu and Arg418Gly forms of ETV6 are shown in **Supplementary Figure 6**.

presumably in megakaryocytes. The function of the central domain is not well understood, but this domain has been shown to undergo post-translational modifications and to be essential for the repressive function of ETV6 in *in vitro* reporter assays[21].

To investigate the effect of the p.Pro214Leu, p.Arg418Gly and p.385_418del alterations on ETV6 transcriptional repression, we used an *in vitro* reporter assay to measure luciferase expression induced by the promoter of the known ETV6 target *stromelysin-1* (*MMP3*)[8,22]. As expected, wild-type ETV6 repressed luciferase expression in comparison to empty vector. Expression of each of the three mutant forms of ETV6 resulted in less transcriptional repression than detected with wild-type ETV6 when the mutant was expressed alone or together with wild-type ETV6 (**Fig. 1d**). We determined that the protein expression of wild-type ETV6 and all the mutants in transfected cells was equivalent by immunoblot, suggesting that the mutations do not influence protein stability (**Supplementary Fig. 3**).

Thrombopoiesis involves a complex sequence of cellular events in bone marrow megakaryocytes culminating in the elaboration of proplatelet extensions that release platelets into the circulation. Approximately 1×10^{11} platelets must be produced daily to maintain the normal concentrations of $150–400 \times 10^9$ platelets/l in human blood[23]. To determine the effect of the p.Pro214Leu and p.Arg418Gly ETV6 substitutions on megakaryocyte differentiation, we transduced human CD34$^+$ cells with lentivirus containing *ETV6* alleles encoding wild-type, Pro214Leu or Arg418Gly protein and cultured them with thrombopoietin to support megakaryocyte development. Cells transduced with lentivirus expressing Pro214Leu or Arg418Gly ETV6 showed delayed/decreased maturation when compared to control cells and those transduced with lentivirus expressing wild-type ETV6, as indicated by increased numbers of small, immature megakaryocytes (similar to the abnormalities observed in the bone marrow of affected individuals; **Fig. 1b**) and decreased generation of proplatelets (**Fig. 2a–d**). These findings indicate that the identified mutations affect megakaryocyte development and likely platelet production.



As ETV6 requires dimerization to exert its transcriptional repression, it is possible that the p.Pro214Leu, p.Arg418Gly and p.385_418del alterations might affect this interaction in a dominant-negative manner. We found that all mutants dimerized with wild-type ETV6, as demonstrated by concomitant pulldown of both differentially His-tagged wild-type ETV6 and Myc-tagged ETV6 mutants (**Supplementary Fig. 5**). We examined the intracellular distribution of protein produced by transduced constructs for the Pro214Leu, Arg418Gly and wild-type forms of ETV6 in megakaryocytes by immunofluorescence staining for their Myc tag, which showed the expected nuclear localization for wild-type ETV6, whereas Pro214Leu and Arg418Gly ETV6 were concentrated in the cytoplasm (**Supplementary Fig. 6**). Staining for ETV6 (**Fig. 3**) showed that this protein was concentrated in the nuclei of untransduced controls and cells transduced with wild-type ETV6 and that it was largely in the cytoplasm of cells expressing mutant ETV6 (see also **Supplementary Videos 1–4**). These results indicate that the mutant forms of ETV6

Figure 3 Aberrant cytoplasmic localization of ETV6 in cultured megakaryocytes transduced with lentivirus expressing ETV6 mutants. Comparison via immunofluorescence microscopy of ETV6 localization in cultured control megakaryocytes and those transduced with lentivirus for wild-type or mutant (Pro214Leu or Arg418Gly) Myc-tagged forms of ETV6 at day 12. Confocal *z* sections of representative mature cells (>15 µm) are shown stained for DNA (blue), ETV6 (both endogenous and Myc-tagged forms; green) and Myc (red). In control cells, endogenous ETV6 is concentrated in the nucleus, as is endogenous and Myc-tagged wild-type ETV6 in cells transduced with lentivirus for wild-type ETV6. In contrast, cells transduced with mutant Myc-tagged ETV6 (Pro214Leu or Arg418Gly) showed this protein concentrated in the cytoplasm with little visible in the nucleus. Scale bars, 5 µm. See also **Supplementary Videos 1–4**.

mislocalize inside cells and, by dimerizing with endogenous wild-type protein, may also prevent nuclear localization, further compromising the repression activity of ETV6.

Several cases of leukemia involving somatic *ETV6* alterations have been associated with loss of ETV6 function[4–6]. To identify additional genetic aberrations contributing to leukemogenesis, we performed whole-exome and RNA sequencing (RNA-seq) on leukemia blast cells collected at the time of diagnosis and on bone marrow cells collected after remission in the same patient (III-1, family 1). RNA-seq showed that the wild-type *ETV6* allele was structurally normal and that wild-type transcripts and those encoding p.Pro214Leu were expressed at equal levels in the bone marrow samples. We found several genetic variants and a new gene fusion between *PAX5* and *SHB* in the leukemia sample that were not present in the remission sample (**Supplementary Table 2**). These variants may represent candidate mutations cooperative with the germline *ETV6* mutation in the development of leukemia[24,25].

In summary, we present new germline mutations in *ETV6*, a gene known to be involved in human leukemogenesis and in mouse and zebrafish hematopoiesis, in three families with thrombocytopenia and high red blood cell MCV, along with leukemia predisposition in two of the three families (with the mutation encoding p.Pro214Leu). These mutations partially disrupt ETV6 transcriptional repression *in vitro* and cause aberrant cytoplasmic localization of both mutant and endogenous ETV6, suggesting a dominant-negative effect. Recently, while this manuscript was under review, a manuscript by Zhang *et al.* showed similar findings in three unrelated families[9]. These two studies represent effective independent replications and lend further support to the findings. The identified mutations in *ETV6* also impair megakaryocyte development and proplatelet formation in culture. Furthermore, platelet RNA-seq of patients with the p.Pro214Leu alteration showed decreased expression of platelet-specific transcripts associated with the mutation, including considerable reduction in the levels of several cytoskeletal transcripts (**Supplementary Figs. 7** and **8**, **Supplementary Note** and **Supplementary Data Set**). The mutations in these families show the implications of viable germline mutations in *ETV6* and provide insights into the importance of ETV6 for platelet biology and tumor-suppressor activity.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. BioProject, PRJNA276621; Sequence Read Archive (SRA), SRP056053; database of Genotypes and Phenotypes (dbGaP), phs000873.v1.p1.

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

ACKNOWLEDGMENTS

We are grateful to the families studied for their contribution to this project. We are also grateful to T. Shaikh, R. Spritz and J. Murray for their insightful comments. This work was supported by the Postle Family Chair in Pediatric Cancer and Blood Disorders (J.D.P.) and by US National Institutes of Health grants HL112311 (A.S.W.) and GM103806 (J.W.R.). W.H.A.K. was supported by operating grants from the Canadian Institutes of Health Research (CIHR; MOP-81208 and MOP-259952). P.N. and A.S. were supported by grant GGP13082 from the Telethon Foundation.

AUTHOR CONTRIBUTIONS

L.N., R.W.L., A.B.L.-S., A.S.W., W.H.A.K., C.C.P. and J.D.P. conceived and designed the experiments. L.N., R.W.L., A.B.L.-S., R.L., F.G.P., L. Li, L. Lu, A.S., C.G. and D.D.R. performed experiments and provided critical data. M.C., M.R., P.N., C.L.B., A.P., M.D., A.G.-H., L.X. and C.L.-M. provided patient samples and study materials, and collected and assembled data. S.H., P.H. and A.G.-H. analyzed data. K.J., K.G. and J.W.R. analyzed genomic and transcriptome data. L.N., R.W.L., A.B.L.-S., F.G.P., A.S.W., W.H.A.K., C.C.P. and J.D.P. wrote the manuscript. All authors reviewed and contributed to the final version of the manuscript. A.S.W., W.H.A.K., C.C.P. and J.D.P. jointly supervised the research.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Song, W.J. *et al.* Haploinsufficiency of CBFA2 causes familial thrombocytopenia with propensity to develop acute myelogenous leukaemia. *Nat. Genet.* **23**, 166–175 (1999).
2. Nichols, K.E. *et al.* Familial dyserythropoietic anaemia and thrombocytopenia due to an inherited mutation in *GATA1*. *Nat. Genet.* **24**, 266–270 (2000).
3. Kar, A. & Gutierrez-Hartmann, A. Molecular mechanisms of ETS transcription factor–mediated tumorigenesis. *Crit. Rev. Biochem. Mol. Biol.* **48**, 522–543 (2013).
4. Romana, S.P. *et al.* Deletion of the short arm of chromosome 12 is a secondary event in acute lymphoblastic leukemia with t(12;21). *Leukemia* **10**, 167–170 (1996).
5. Patel, N. *et al.* Expression profile of wild-type ETV6 in childhood acute leukaemia. *Br. J. Haematol.* **122**, 94–98 (2003).
6. Barjesteh van Waalwijk van Doorn-Khosrovani, S. *et al.* Somatic heterozygous mutations in ETV6 (TEL) and frequent absence of ETV6 protein in acute myeloid leukemia. *Oncogene* **24**, 4129–4137 (2005).
7. Bejar, R. *et al.* Clinical effect of point mutations in myelodysplastic syndromes. *N. Engl. J. Med.* **364**, 2496–2506 (2011).
8. Van Vlierberghe, P. *et al.* *ETV6* mutations in early immature human T cell leukemias. *J. Exp. Med.* **208**, 2571–2579 (2011).
9. Zhang, M.Y. *et al.* Germline *ETV6* mutations in familial thrombocytopenia and hematologic malignancy. *Nat. Genet.* **47**, 180–185 (2015).
10. Rasighaemi, P., Onnebo, S.M., Liongue, C. & Ward, A.C. ETV6 (TEL1) regulates embryonic hematopoiesis in zebrafish. *Haematologica* **100**, 23–31 (2015).
11. Wang, L.C. *et al.* Yolk sac angiogenic defect and intra-embryonic apoptosis in mice lacking the Ets-related factor TEL. *EMBO J.* **16**, 4374–4383 (1997).
12. Wang, L.C. *et al.* The *TEL*/*ETV6* gene is required specifically for hematopoiesis in the bone marrow. *Genes Dev.* **12**, 2392–2402 (1998).
13. Hock, H. *et al.* Tel/Etv6 is an essential and selective regulator of adult hematopoietic stem cell survival. *Genes Dev.* **18**, 2336–2341 (2004).
14. Pereira, C.F. *et al.* Induction of a hemogenic program in mouse fibroblasts. *Cell Stem Cell* **13**, 205–218 (2013).
15. Orkin, S.H. *et al.* Abnormal RNA processing due to the exon mutation of β E-globin gene. *Nature* **300**, 768–769 (1982).
16. Green, S.M., Coyne, H.J. III, McIntosh, L.P. & Graves, B.J. DNA binding by the ETS protein TEL (ETV6) is regulated by autoinhibition and self-association. *J. Biol. Chem.* **285**, 18496–18504 (2010).
17. Kwiatkowski, B.A. *et al.* The ets family member Tel binds to the Fli-1 oncoprotein and inhibits its transcriptional activity. *J. Biol. Chem.* **273**, 17525–17530 (1998).
18. Raslova, H. *et al.* *FLI1* monoallelic expression combined with its hemizygous loss underlies Paris-Trousseau/Jacobsen thrombopenia. *J. Clin. Invest.* **114**, 77–84 (2004).
19. Million, R.P., Harakawa, N., Roumiantsev, S., Varticovski, L. & Van Etten, R.A. A direct binding site for Grb2 contributes to transformation and leukemogenesis by the Tel-Abl (ETV6-Abl) tyrosine kinase. *Mol. Cell. Biol.* **24**, 4685–4695 (2004).
20. Roukens, M.G., Alloul-Ramdhani, M., Moghadasi, S., Op den Brouw, M. & Baker, D.A. Downregulation of vertebrate Tel (ETV6) and *Drosophila* Yan is facilitated by an evolutionarily conserved mechanism of F-box-mediated ubiquitination. *Mol. Cell. Biol.* **28**, 4394–4406 (2008).
21. Lopez, R.G. *et al.* TEL is a sequence-specific transcriptional repressor. *J. Biol. Chem.* **274**, 30132–30138 (1999).
22. Fenrick, R. *et al.* TEL, a putative tumor suppressor, modulates cell growth and cell morphology of ras-transformed cells while repressing the transcription of stromelysin-1. *Mol. Cell. Biol.* **20**, 5828–5839 (2000).
23. Machlus, K.R. & Italiano, J.E. Jr. The incredible journey: from megakaryocyte development to platelet formation. *J. Cell Biol.* **201**, 785–796 (2013).
24. Mullighan, C.G. *et al.* Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature* **446**, 758–764 (2007).
25. Strehl, S., Konig, M., Dworzak, M.N., Kalwak, K. & Haas, O.A. *PAX5*/*ETV6* fusion defines cytogenetic entity dic(9;12)(p13;p13). *Leukemia* **17**, 1121–1123 (2003).

ONLINE METHODS

Patient tissue acquisition. All patients in the original cohort (family 1) were recruited at the Hematology Clinic at the Children's Hospital of Michigan. The study received institutional review board approval from the University of Colorado Anschutz Medical Campus, and informed consent was obtained for all participants. Patients in families 2 and 3 were recruited in clinical centers in the Czech Republic and Italy as part of a European consortium focused on inherited platelet disorders. The institutional review board of the IRCCS Policlinico San Matteo Foundation of Pavia, Italy, approved the study protocol. Informed consent was obtained from all patients in the study. Studies were performed in accordance with the Declaration of Helsinki. Genomic DNA was extracted from whole blood using the Gentra PureGene DNA Extraction kit (Qiagen). Platelets were purified from whole blood and negatively selected using $CD45^+$ MACS MicroBeads (Miltenyi Biotec). Bone marrow aspirates were obtained at diagnosis, and permission was obtained from families to use images.

Exome sequencing. Genomic DNA (3–5 μg) was sheared, size selected (~400–600 bp), ligated to sequencing adaptors and PCR amplified following standard library preparation procedures. The library obtained after PCR amplification was used for exome capture with the Agilent SureSelect Human All Exon v4 Capture kit. Eight exome-enriched samples were sequenced using two lanes (2 × 100 bp) on an Illumina HiSeq 2000. Exome sequencing was performed in eight individuals from family 1: II1, II3, II5, II7, III1, III2, III9 and III10. We screened for *ETV6* mutations in the European families by Sanger sequencing of the promoter region, exons and exon-flanking regions of *ETV6*.

Bioinformatics of exome sequencing. Reads passing the Illumina chastity filter were subjected to a quality filter step that removed low-quality bases from the 3′ end and retained pairs of reads if the trimmed reads for both members of the pair were 50 bp or longer. Paired reads that passed the quality filter step were mapped to the reference human genome sequence (hg19) with GSNAP (Genomic Short-read Nucleotide Alignment Program, version 2012-07-20)[26]. Sequence calls for SNPs and insertions and deletions (indels) were performed using GATK (the Broad Institute's Genome Analysis Toolkit, v2.1-8-g5efb575)[27].

The program ANNOVAR (Annotate Variation, version 2012-03-08) was used to classify variants and to cross-reference all variants across various genetic variation databases. Included in ANNOVAR are databases to identify nonsynonymous and splice-site variants (refGene.txt), variants in conserved genomic regions (phastConsElements46way.txt), variants in segmental duplications (genomicSuperDups.txt) and variant frequencies from the 1000 Genomes Project database (hg19_ALL.sites.2012_02.txt). Variants located outside of conserved regions or with frequencies >1% were excluded from further analysis[28]. Only nonsynonymous changes (SNPs and indels), splice-site variants and/or changes resulting in an aberrant stop codon were considered for further analysis. All insertion and deletion variants were considered to be damaging, whereas SNP variants were cross-referenced to dbNSFP (database of functional predictions for nonsynonymous SNPs, version 2.0) to determine whether the changes would be considered tolerable or damaging using four algorithms (Sorting Intolerant From Tolerant (SIFT), PolyPhen-2 (Polymorphism Phenotyping v2), likelihood-ratio test (LRT) and MutationTaster)[29].

The final filtered list of variants for each affected family member was then intersected to find putative causal variants.

Platelet RNA sequencing. RNA was isolated from leukoreduced platelet preparations stored in TRIzol as previously described[30]. RNA-seq libraries were prepared and barcoded using Illumina TruSeq V2 with oligo(dT) selection. We generated 50 cycles of single-end reads on a single lane of the HiSeq 2000 and aligned the reads using Novoalign (Novocraft Technologies) to UCSC genome version hg19 with known and shuffled splice junctions included. Normalization of read counts and differential expression analysis were performed with DESeq2 (ref. 31). Sample-to-sample variability in the levels of leukocyte transcripts, which can significantly alter read counts (J.W.R., unpublished observations), was corrected for by including the ratio of *PTPRC* (encoding the leukocyte marker CD45) to *ITGA2B* (encoding a platelet marker) as a factor in the model for significance testing. Relationship to the affected individual was also included in the model. Euclidean distance computation, clustering (complete linkage) and heat map analysis of the regularized log-transformed read counts was performed in R as described in the DESeq2 vignette. A list of 351 expressed transcripts involved in platelet biogenesis or function was curated from Reactome and from transcripts enriched in platelets in comparison to all other tissues in Illumina's Human Body Map 2.0. Gene Set Enrichment Analysis (GSEA)[32] of the 177 differentially expressed targets included testing against GO biological process, Biocarata, KEGG and Reactome gene sets.

Bone marrow RNA sequencing. Total RNA was isolated from samples taken from bone marrow at diagnosis of leukemia and again during remission. cDNA libraries were constructed for each sample using the TruSeq mRNA sample preparation kit (Illumina). The poly(A) mRNA fraction was isolated from total RNA using selection via poly(T) oligonucleotide attached to magnetic beads. RNA was chemically fragmented and randomly primed before reverse transcription and cDNA generation. cDNA fragments then underwent end repair, the addition of a single A base to the 3′ end and ligation to adaptors. Finally, the products were purified and PCR enriched to create the double-stranded cDNA library. The uniquely indexed libraries were pooled and sequenced on a single lane of an Illumina HiSeq 2000 flow cell, yielding 2 × 100-bp reads.

Bioinformatics of RNA sequencing. Reads passing the Illumina chastity filter were subjected to a quality filter step as described above. Paired reads that passed the quality filter step were mapped to the reference human genome sequence (hg19) with GSNAP (version 2012-07-20)[26]. The aligned reads were then searched for gene fusions using two separate algorithms: TopHat-Fusion, v2.0.9 (ref. 33), and FusionMap, version 7.0.1.25 (ref. 34). Intersection of the resulting gene fusion predictions from the two programs resulted in a single high-confidence candidate.

Plasmids and mutagenesis. A pCMV-Entry vector containing cDNA for human *ETV6* (*ets* variant 6) and a sequence encoding a C-terminal Myc/DDK tag was obtained from Origene. Site-directed mutagenesis was carried out using the QuikChange Site-Directed Mutagenesis kit (Agilent Technologies) and primers designed to contain the mutations of interest (c.641C>T and c.1252A>G) with flanking complementary sequence. Wild-type *ETV6* was also cloned into a pCMV-Entry vector encoding a C-terminal His tag (Origene). A construct designed to encode the alternatively spliced and truncated *ETV6* (p.385_418del, p.Asn385Valfs*7) was generated by GenScript and cloned into the pCMV-Entry vector encoding a C-terminal Myc/DDK tag. The luciferase reporter construct (pGL2-754TR) was originally published by S. Hiebert and was obtained from A. Ferrando (Columbia University)[8]. The pCMV-Renilla vector was also obtained from A. Ferrando. All constructs were transformed into Mix & Go Competent *Escherichia coli* cells (Zymogen Research) and grown up with 100 μg/ml kanamycin or ampicillin.

Luciferase reporter assays. HEK293T cells (a gift from D. Ginsburg, University of Michigan) were cultured in MEM (Invitrogen) supplemented with 10% FBS. Cells were seeded at 1.5×10^6 cells/60-mm dish and transfected using Trans-IT 293 (Mirus) 24 h after seeding with either 2 μg of empty pCMV-Entry vector or vector encoding wild-type, Pro214Leu, Arg418Gly or 385_418del ETV6 or with 1 μg of wild-type *ETV6* and 1 μg of mutant *ETV6* cDNA for cotransfection experiments. Cells were also cotransfected with 2 μg of pGL2-754TR luciferase reporter, along with 15 ng of pCMV-Renilla as a transfection control. Cells were collected 48 h after transfection, lysed with Passive Lysis Buffer (Promega) and analyzed using the Dual-Luciferase Reporter Assay kit (Promega). Before experiments, cells were tested for mycoplasma contamination, and contamination was ruled out.

Pulldown assays. HEK293T cells were transfected with construct encoding Myc/DDK- or His-tagged wild-type ETV6 alone or cotransfected with construct encoding His-tagged wild-type ETV6 and Myc/DDK-tagged wild-type, Pro214Leu, Arg418Gly or 385_418del ETV6, and cells were collected as previously described. Protein complexes containing the Myc/DDK tag were purified

from cell lysate using the c-Myc Protein Mild Purification kit (Medical and Biological Laboratories). Protein was eluted using c-Myc peptide, and eluate was run on SDS-PAGE and probed using a primary antibody to DDK (TA50011-100, clone 4C5, Origene) or His (polyclonal, 2365, Cell Signaling Technology).

Lentiviral CD34$^+$ cell transduction. Lentivirus generation and CD34$^+$ cell transduction were performed following methods previously described[35]. Specifically, *ETV6* sequences encoding c-Myc–tagged wild-type, Pro214Leu and Arg418Gly protein were cloned into the lentiviral vector PLJM1, and the vectors were cotransfected into HEK293T cells with PMD2.G, PMDLg/pRRE and pRSV-Rev. Supernatants were collected every 24 h for up to 72 h, and virus was concentrated via ultracentrifugation. CD34$^+$ cells were affinity purified from G-CSF–mobilized blood progenitor cells from bone marrow transplant donors using the CD34 MicroBeads kit (Miltenyi Biotec) and cultured with IMDM (Gibco) supplemented with 10% BIT 9500 (Stemcell Technologies), 2 mM L-glutamine (Wisent), 1% penicillin-streptomycin (Wisent) and 50 ng/ml thrombopoietin (a gift from the Kirin Brewery Company). On day 2, cells were pelleted and resuspended in culture medium containing viruses encoding wild-type, Pro214Leu or Arg418Gly ETV6. Cell-virus mixtures were placed onto plates coated with Retronectin (Clontech) for transduction, a process that was repeated 12 h later. Cells were transferred to a new culture dish after 12 h. On day 8 of culture, cells were seeded onto coverslips coated with a 1:6 dilution of Matrigel (Corning) in DMEM (Winsent) as previously described[36].

High-resolution laser immunofluorescence confocal microscopy. On day 12 of culture, cells growing in Matrigel on coverslips were fixed with paraformaldehyde, permeabilized and immunostained with human-specific antibodies: megakaryocyte-specific proteins were detected with rabbit antibody to CD61 (ab75872, Abcam) and sheep antibody to VWF (AHP062, AbD Serotec), Myc-tagged ETV6 was detected using mouse antibody to c-Myc (AHP062, Covance), and endogenous and tagged ETV6 was detected via rabbit antibody to ETV6 (Prestige Antibodies, HPA000264, Sigma-Aldrich). Antibodies were visualized using donkey secondary antibodies: anti-rabbit Alexa Fluor 647 or 488, anti-sheep Alexa Fluor 555, and anti-mouse Alexa Fluor 488 or 546 (Life Technologies). DNA was stained with DAPI, and samples were prepared with fluorescent mounting medium (Dako).

Images were obtained with an oil immersion objective (60×, 1.35 N.A.) using a Quorum Olympus spinning-disc confocal inverted epifluorescence microscope equipped with four solid-state lasers (Spectral Applied Research)—405 nm, 491 nm, 561 nm and 642 nm—an Improvision Piezo Focus Drive, a 1.5× magnification lens (Spectral Applied Research), a Hamamatsu C9100-13 back-thinned EM-CCD camera and a Yokogawa CSU X1 spinning-disc confocal scan head with the Spectral Aurora Borealis upgrade. Acquisition, image processing and cell counting were performed using PerkinElmer Volocity software (versions 5.5–6.1). Confocal images were deconvolved and exported as *z* sections or extended-focus images to Adobe Photoshop and/or Illustrator for final presentation. Images were rendered as three-dimensional volumes (maximum intensity) and animated using Imaris 7.6 to create videos.

Transmission electron microscopy and immunoblots. Transmission electron microscopy of platelets and preparation of platelet lysates for immunoblot analysis were carried out as previously described[37]. The antibodies used for the immunoblots presented throughout were polyclonal rabbit antibody to ETV6 (PA5-35371, Thermo Scientific), mouse monoclonal antibody to GAPDH (MAB374, Millipore), mouse monoclonal 4C5 antibody to DDK (Origene) and rabbit polyclonal antibody to His (Cell Signaling Technology).

26. Wu, T.D. & Nacu, S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* **26**, 873–881 (2010).
27. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
28. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
29. Liu, X., Jian, X. & Boerwinkle, E. dbNSFP: a lightweight database of human nonsynonymous SNPs and their functional predictions. *Hum. Mutat.* **32**, 894–899 (2011).
30. Rowley, J.W. *et al.* Genome-wide RNA-seq analysis of human and mouse platelet transcriptomes. *Blood* **118**, e101–e111 (2011).
31. Love, M., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *bioRxiv.* http://dx.doi.org/10.1101/002832.
32. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**, 15545–15550 (2005).
33. Kim, D. & Salzberg, S. TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. *Genome Biol.* **12**, R72 (2011).
34. Ge, H. *et al.* FusionMap: detecting fusion genes from next-generation sequencing data at base-pair resolution. *Bioinformatics* **27**, 1922–1928 (2011).
35. Wilcox, D.A. *et al.* Induction of megakaryocytes to synthesize and store a releasable pool of human factor VIII. *J. Thromb. Haemost.* **1**, 2477–2489 (2003).
36. Kahr, W.H. *et al.* Abnormal megakaryocyte development and platelet function in *Nbeal2*$^{-/-}$ mice. *Blood* **122**, 3349–3358 (2013).
37. Urban, D. *et al.* The VPS33B-binding protein VPS16B is required in megakaryocyte and platelet α-granule biogenesis. *Blood* **120**, 5032–5040 (2012).