Genomics of inherited bone marrow failure and myelodysplasia

Michael Yu Zhang

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2015

Reading Committee:

Mary-Claire King, Chair

Akiko Shimamura

Marshall Horwitz

Program Authorized to Offer Degree:

Molecular and Cellular Biology

University of Washington

ABSTRACT

Genomics of inherited bone marrow failure and myelodysplasia

Michael Yu Zhang

Chair of the Supervisory Committee:

Professor Mary-Claire King

Department of Medicine (Medical Genetics) and Genome Sciences

Bone marrow failure and myelodysplastic syndromes (BMF/MDS) are disorders of impaired blood cell production with increased leukemia risk. BMF/MDS may be acquired or inherited, a distinction critical for treatment selection. Currently, diagnosis of these inherited syndromes is based on clinical history, family history, and laboratory studies, which directs the ordering of genetic tests on a gene-by-gene basis. However, despite extensive clinical workup and serial genetic testing, many cases remain unexplained. We sought to define the genetic etiology and pathophysiology of unclassified bone marrow failure and myelodysplastic syndromes. First, to determine the extent to which patients remained undiagnosed due to atypical or cryptic presentations of known inherited BMF/MDS, we developed a massively-parallel, next-generation DNA sequencing assay to simultaneously screen for mutations in 85 BMF/MDS genes. Querying 71 pediatric and adult patients with unclassified BMF/MDS using this assay revealed 8 (11%) patients with constitutional, pathogenic mutations in *GATA2*, *RUNX1*, *DKC1*, or *LIG4*. All eight patients lacked classic features or laboratory findings for their syndromes. These data suggest that the current clinical practice of directed testing of individual genes failed

to diagnose patients with cryptic disease presentations and inform the integration of broad genetic screening into the diagnostic workup of BMF/MDS. In cases where broad screening of known genes failed to identify an inherited cause, we sought to discover new genes responsible for inherited BMF/MDS. We identified kindreds with unclassified disease presenting in multiple family members and sequenced the complete coding regions of their genomes (exome). In a family with low blood counts and malignancy, we discovered a novel germline heterozygous mutation in the transcription factor *ETV6* segregating with the disease phenotype. The mutant protein is defective in transcriptional regulation, inhibits wild-type ETV6 in a dominant negative manner, and impairs hematopoiesis when expressed in human hematopoietic stem cells. Together, our studies provide new methodologies and genetic associations to enhance the diagnosis of inherited BMF/MDS.

Acknowledgements

I would like to thank Akiko Shimamura for imparting upon me intellectual curiosity, critical thinking, creativity, and an unquenchable enthusiasm for science. I would like to thank Mary-Claire King for sharing with me her boundless energy, scientific rigor, and bold thinking. They always put my needs as a trainee before all else. I could not imagine mentors who are better models as scientists and human beings.

I would like to thank the rest of my thesis committee, Marshall Horwitz, Stephen Tapscott, and Barry Stoddard, as well as former members Linda Wordeman and Chip Asbury, for insightful suggestions, constructive criticisms, and guidance.

I would like to thank the members of the Shimamura lab, including Scott Coats, Katie Lombardo, Tomoka Nakamura, Nick Burwick, Marilyn Sanchez-Bonilla, Take Furuyama, Melisa Ruiz-Gutierrez, Aaron Seo, Phoenix Ho, and Ang Scott, for their patience in showing me the ropes and making the lab a warm and intellectually stimulating community.

I would like to thank members of the King lab, Tom Walsh, Ming Lee, Suleyman Gulsuner, and Amanda Watts, for being such brilliant and hard-working collaborators. It's been a true honor to work with them.

I would like to thank all the patients and their families, for entrusting us with the most sensitive part of their being: their blueprint. They have made these studies possible and are the ultimate motivation behind all of our work.

I would like to thank my parents, Peng and Ning, and brother Mark for their unwavering love, encouragement, and every possible form of support.

Finally, I thank my wife Solina and son Benji. Solina is the most positive, generous, resilient, and hard-working person I know. She has made this work possible, and I share it with her. Benji is my night and day; everything here is for him.

Table of Contents

Chapter 1 – Introduction

1.1 Inherited bone marrow failure and myelodysplastic syndromes

Bone marrow failure (BMF) is a deficiency in blood cell production in the bone marrow which leads to low peripheral blood counts, or cytopenias, in one or more hematopoietic lineages. Myelodysplastic syndrome (MDS) is a heterogeneous group of disorders characterized by ineffective hematopoiesis, cytopenias, and increased risk of progression to acute myeloid leukemia (AML)[1]. BMF/MDS may be sporadically acquired or genetically inherited. Examples of inherited BMF/MDS include dyskeratosis congenita, caused by mutations in the teleomerase complex[2], Fanconi anemia, caused by biallelic mutations in DNA repair genes[3], Diamond-Blackfan anemia[4], caused by mutations in ribosomal protein genes, and Shwachman-Diamond syndrome, caused by mutations in a ribosomal biogenesis factor gene, *SBDS*[5]. In addition to hematologic abnormalities, these syndromes are also associated with increased risk of solid tumors as well as physical congenital anomalies[6].

1.2 Traditional diagnostic paradigm of inherited BMF/MDS

Accurate and timely diagnosis of inherited bone marrow failure (BMF) and inherited myelodysplastic syndromes (MDS) is essential to ensure appropriate medical management and treatment. Standard treatments for acquired BMF/MDS, such as cyclosporine and anti-thymocyte globulin, have poor efficacy in patients with inherited syndromes[7]. Diagnosis of an underlying inherited BMF/MDS allows clinical monitoring for early signs of leukemia development. The recognition of an inherited disorder also informs hematopoietic stem cell donor selection as it allows unambiguous identification of affected siblings and guides appropriate genetic counseling of family members.

Though critical for guiding clinical care, diagnosis of these inherited disorders remains challenging. Variable presentation within each syndrome and overlapping clinical features between different syndromes complicates the decision of which genetic tests to order based on clinical and laboratory findings. A substantial fraction of patients presenting with familial or syndromic marrow failure do not meet the diagnostic criteria for an inherited BMF/MDS[8]. Our genomic studies of these patients may reveal causal mutations in known BMF/MDS genes which were clinically unexpected due to atypical presentations of their disorder and thus expand our understanding of the phenotypic scope of these syndromes. In addition, discovery of mutations in genes not previously associated with inherited BMF/MDS will define new syndromes that provide insights into the role of these genes in hematopoiesis and malignancy.

Evidence-based guidelines for genetic screening for inherited BMF/MDS are scarce. A family history of MDS or leukemia is a clue to an underlying inherited cause. Physicians also rely on clinical stigmata to suggest inherited BMF/MDS[6]. Substantial phenotypic overlap between different syndromes, as well as absent, subtle, or previously unreported clinical findings of inherited BMF/MDS, render diagnosis challenging[9].

1.3 Next-generation massively-parallel sequencing technology

Traditional clinical genetic testing is based on PCR amplification and Sanger sequencing. The high accuracy but low-throughput of this technology renders it suitable for sequencing of genes on an individual basis. Next generation sequencing technologies have reduced the base-pair costs of DNA sequencing by four to five orders of magnitude compared to Sanger sequencing[10]. In these technologies, sheared DNA fragments (100-300bp) are amplified in clusters at high-density on the surface of a flow cell. The base-pairs are then determined by CCD-camera capture of fluorescently labeled nucleotides as they are incorporated base by base in the

synthesis of DNA fragments complementary to each cluster sequence. Several million clusters are amplified and sequenced simultaneously in each flow cell.

Despite this reduction in sequencing cost enabled by the dramatic increase in parallelization and throughput, sequencing the entire genome at an informative depth of coverage using next-generation technology remains cost-prohibitive for most applications. To make high-throughput sequencing more broadly applicable, methods to target specific portions of the genome were developed. Targeted portions included either all of the protein-coding regions of the genome, known as the exome, or specific gene subsets, known as panels. One technology that enables such enrichment of targeted regions is in-solution hybrid capture, in which RNA probes complementary to the regions of interest (cRNA) are conjugated to magnetic beads and pooled in a single reaction for multiplexed enrichment of target molecules. In this way, a subset of the genome may be sequenced at high depth at a fraction of the cost of whole genome sequencing. Furthermore, unique barcodes may be added to the ends of captured DNA for each patient, and multiple patient samples may be pooled for simultaneous sequencing followed by bioinformatic identification of the patient of origin for each read.

1.4 Outline of thesis

We hypothesize that the traditional clinical genetic work-up of BMF/MDS patients fails to identify underlying inherited disease due to 1) the patient manifesting a known inherited BMF/MDS disorder in an atypical or otherwise cryptic fashion such that diagnosis was elusive or 2) the patient having an as-of-yet uncharacterized genetic disorder caused by mutation of a gene not previously associated with inherited BMF/MDS. The subject of this thesis is to advance the diagnostic paradigm and expand the genetic repertoire of inherited BMF/MDS disorders using next-generation sequencing. In Chapter 2, I describe the development of a targeted capture

assay which enables the simultaneous querying of mutations in all known BMF/MDS genes using high-throughput sequencing. This assay was then used to determine the genetic cause of inherited disease in a group of patients with unclassified BMF/MDS after traditional clinical genetic workup. The assay showed that a subset of these patients indeed had mutations in known BMF/MDS genes but had manifested the disease in a cryptic or atypical manner. In Chapter 3, I describe the whole exome sequencing of BMF/MDS kindreds to identify germline mutations in *ETV6* as the genetic cause of a new autosomal dominant genetic disorder characterized by thrombocytopenia and predisposition to hematologic malignancy. Functional analysis of the patient mutations show that the mutations cause disease via a dominant negative mechanism. Appendices A and B contain supplemental information for chapters 2 and 3, respectively.

Chapter 2 - Genomic analysis of bone marrow failure and myelodysplastic syndromes

Note: This chapter is based on the following published paper:

Michael Y. Zhang, Siobán B. Keel, Tom Walsh, Ming K. Lee, Suleyman Gulsuner, Amanda C. Watts, Colin C. Pritchard, Stephen Salipante, Michael R. Jeng, Inga Hofmann, David A. Williams, Mark D. Fleming, Janis L. Abkowitz, Mary-Claire King, and Akiko Shimamura. Genomic analysis of bone marrow failure and myelodysplastic syndromes reveals phenotypic and diagnostic complexity. *Haematologica*, 100, 42-48 (2015).

M.Y.Z. and S.B.K. contributed equally to this work.

2.1 Abstract

Accurate and timely diagnosis of inherited bone marrow failure (BMF) and inherited myelodysplastic syndromes (MDS) is essential to guide clinical management. Distinguishing inherited from acquired BMF/MDS poses a significant clinical challenge. At present, diagnostic genetic testing for inherited BMF/MDS is performed gene-by-gene, guided by clinical and laboratory evaluation. We hypothesized that standard clinically-directed genetic testing misses patients with cryptic or atypical presentations of inherited BMF/MDS. In order to screen simultaneously for mutations of all classes in all known BMF/MDS genes, we developed and validated a panel of 85 genes for targeted capture and multiplexed massively-parallel sequencing. In patients with clinical diagnoses of Fanconi anemia, genomic analysis resolved subtype assignment, including those of patients with inconclusive complementation test results. Eight patients with idiopathic BMF/MDS were found to harbor damaging germline mutations in *GATA2*, *RUNX1*, *DKC1*, or *LIG4*. All eight of these patients lacked classical clinical stigmata or laboratory findings of these syndromes and only four had a family history suggestive of inherited disease. These results reflect the extensive genetic heterogeneity and phenotypic complexity of BMF/MDS phenotypes. This study supports the integration of broad unbiased genetic screening into the diagnostic workup of children and young adults with BMF/MDS.

2.2 Introduction

Timely and accurate diagnosis of inherited bone marrow failure (BMF) and inherited myelodysplastic syndromes (MDS) is essential to ensure appropriate medical management and treatment[11,12]. Early diagnosis of an underlying inherited BMF/MDS allows clinical monitoring for signs of clonal evolution in order to initiate hematopoietic stem cell transplantation prior to leukemia development. Secondary leukemias arising in these clinical contexts carry particularly poor prognoses. Many of these syndromes require reduced intensity transplantation regimens to

avoid excessive toxicities. Furthermore, the recognition of an inherited disorder informs hematopoietic stem cell donor selection as it allows unambiguous identification of affected siblings and guides appropriate genetic counseling of family members.

Evidence-based guidelines for genetic screening for inherited BMF/MDS are scarce. A family history of MDS or leukemia is a clue to an underlying inherited cause. Physicians also rely on clinical stigmata of inherited BMF/MDS to guide diagnosis. Substantial phenotypic overlap between different syndromes, as well as absent, subtle, or previously unreported clinical findings of inherited BMF/MDS, render diagnosis challenging[9]. The clinical benefit of applying next generation sequencing approaches for comprehensive genetic screening of seemingly idiopathic BMF/MDS remains a critical question.

To investigate the clinical utility of broad genomic analysis for cryptic or atypical presentations of inherited BMF/MDS, we developed a targeted capture gene panel coupled with high-throughput, multiplexed, massively parallel sequencing. This panel, referred to here as MarrowSeq, includes 85 genes responsible for inherited and acquired marrow failure syndromes and MDS. Using this assay, we queried patients deemed to have idiopathic disease for an underlying genetic cause of their BMF and MDS.

2.3 Methods

Additional details on the design and methods of this study are provided in Appendix A.

Subjects. DNA samples were obtained from pediatric and adult patients with idiopathic BMF or MDS at Seattle Children's Hospital, Seattle Cancer Care Alliance, the University of Washington Medical Center, and Boston Children's Hospital. Patients were treated between 2000 and 2013.

Patients had been previously screened for various individual genes causing inherited marrow failure syndromes based on clinical history and physical findings but remained unclassified after genetic workup. Inclusion criteria for pediatric patients included presentation to a pediatric hematology clinic with idiopathic marrow failure (hypoproductive cytopenias including any of the following: absolute neutrophil count less than 1500/μL, hemoglobin low for age, platelet count less than 150,000/μL, hypocellular marrow for age) or MDS as determined by the treating physician according to WHO criteria[13]. Inclusion criteria for patients presenting to an adult hematology clinic included: 1) MDS or hypoproductive cytopenia (hypocellular marrow for age and any of the following: ANC < 1500/uL, hemoglobin low for age, or platelet count less than 150,000/uL), and 2) presentation at age less than 40 years or a family history suggestive of inherited BMF/MDS. Patients or their legal guardians provided written informed consent in accordance with protocols approved by the institutional review board at each institution. Genomic DNA was isolated from peripheral blood, bone marrow mononuclear cells, and/or fibroblasts using the All-prep DNA/RNA kit (Qiagen) or as previously described[14].

Genomics. Oligonucleotide probes were designed using the eArray website (Agilent Technologies). Probes were 120 basepairs (bp) in length and designed to tile each region with 3x depth. Probes covered coding regions and 20 bp of intronic sequence flanking each exon of 85 genes, as well as the promoter of *DKC1*[15] and the 5'UTR of *ANKRD26*[16], which contain known pathogenic variants. Pathogenic variants in an intronic region of *GATA2* were described recently[17], so all samples were screened for mutations in this small intronic region by Sanger sequencing. Mutations in *RTEL1* and *ERCC4* leading to dyskeratosis congenita[18] and Fanconi anemia[19], respectively, were also reported after completion of the initial screens for this study. These regions have been incorporated into subsequent versions of the gene capture pools.

Targeted gene capture and sequencing were performed as previously described[20,21]. Reads were aligned to the human reference genome (hg19) using the Burrows-Wheeler aligner[22]. Single nucleotide and small insertion-deletion (indel) variants were called by three independent bioinformatics pipelines, as previously described[21,23,24] and the results were jointly analyzed. Alignment to the whole genome facilitated exclusion of variants that fell in pseudogenes. Copy number variants (CNVs) were identified as previously described[25].

Each variant in each of the 85 genes was classified by predicted effect on protein function, as previously described. Any variants previously reported to be either germline or somatic pathogenic alleles for BMF/MDS were specifically noted. Loss-of-function or likely damaging mutations in genes not previously reported in inherited BMF/MDS were considered variants of unknown clinical significance.

All potentially damaging variants were validated by Sanger sequencing. In addition, to rule out somatically acquired mutations, variants were confirmed as germline by sequencing fibroblast DNA.

2.4 Results

2.4.1 Validation of Gene Capture Assay

Genes known to contribute to inherited bone marrow failure and myelodysplastic syndromes were selected for capture and sequencing (Table 2.1). For all samples evaluated, median coverage across the 383kb targeted region was 549X, with 97.8% of bases having >50X coverage and 98.2% of bases having >10X coverage. This depth of coverage enabled identification of all classes of mutations, including point mutations, small indels, copy number

variants, and genomic rearrangements. The assay was validated by blinded analysis of genomic DNA from 14 patients with known mutations in nine genes representing a variety of mutation classes (Appendix Table A.1). All mutations, including copy number variants (Figure 2.1A), were correctly identified.

2.4.2 Genetic Determination of Fanconi Anemia Subtype

Simultaneous targeted capture and sequencing of all Fanconi anemia (FA) genes was applied to six patients who had been clinically diagnosed with Fanconi anemia. By prior complementation studies, four patients had been reported as subtype A (FA-A), and two patients remained unclassified after clinical complementation testing for A, C, and G subtypes. The critical mutations were unknown for all six patients. We identified biallelic deleterious mutations in *FANCA* in FH-9, FH-73, FH-124, and FH-241 (Figure 2.2A), confirming the results of their complementation tests. Three of the eight *FANCA* mutations were copy number variants (Figure 2.1B-D).

Patient FH-42, without a previous subtype assignment, carried two damaging mutations in *FANCD2*: c.2715+1G>A (p.Glu906Ilefs*4) and c.2048T>C (p.Leu683Pro) (Appendix Table A.2). Sequencing of subcloned cDNA from this patient indicated that the two mutations were in *trans* and indicated the insertion of 27 base pairs of intron 28 in the *FANCD2* transcript, consistent with aberrant splicing. Immunoblotting of protein from patient-derived fibroblast lysates showed nearly absent levels of FANCD2 protein (Figure 2.2B), confirming subtype D2 for this patient.

Clinical complementation testing of Patient FH-3 was negative for FA subtypes A, C, and G. However, genetic analysis revealed two mutations in *FANCA* for FH-3 (Figure 2.2A). No other pathogenic mutations were found in any other FA genes. By immunoblot analysis, fibroblasts of

FH-3 were deficient in FANCD2 monoubiquitination, which was restored by the introduction of wild-type *FANCA* (Figure 2.2C), confirming the FA-A subtype for this patient.

2.4.3 Broad Genetic Screening Identifies Cryptic Presentations of Inherited BMF/MDS

We next tested for BMF/MDS gene mutations in 71 patients deemed to have idiopathic disease after clinical, laboratory, and clinically-directed genetic evaluation (Table 2.2 and Appendix Table A.3). Marrow failure in these patients remained unclassified after directed testing of candidate genes, chromosomal breakage testing for possible Fanconi anemia[26], telomere length analysis for possible dyskeratosis congenita[27], and tests of pancreatic enzyme levels for possible Shwachman-Diamond syndrome[28]. Of the 71 subjects, 58 patients were age 18 years or younger and drawn from the pediatric clinic, and 13 patients were older than 18 years and drawn from the adult clinic. Thirty-two patients, including 6 of the 13 adults, had a positive family history.

Eight patients carried damaging germline mutations in one of the BMF/MDS genes (Table 2.3). Five patients were heterozygous for damaging mutations in the hematopoietic transcription factor *GATA2*, despite the absence of clinical features of MonoMac[29–32] or Emberger syndrome[33]. *GATA2* contains two Cys4 zinc finger domains, the first of which is responsible for interaction with co-regulator FOG1 (Friend of GATA1) and the second mediating DNA binding. Two patients with idiopathic MDS carried nonsense mutations that truncate the protein in the first zinc finger domain (*GATA2* c.988C>T [p.Arg330*]) or in the second zinc finger domain (*GATA2* c.1084C>T [p.Arg362*]). GATA2 p.Arg330* was previously reported in an individual with MonoMac characteristics[32]; our patient presented with familial leukemia/myelodysplasia. GATA2 p.Arg362* has not been previously reported. Two other patients carried *GATA2* missense mutations altering highly-conserved residues in the second zinc finger domain:

GATA2 c.1078T>A [p.Trp360Arg] and *GATA2* c.1082G>A [p.Arg361His]. These missense mutations have not been previously reported, but GATA2 p.Arg361Leu and p.Arg361Cys have been reported in Emberger[33] and MonoMac[34] syndromes, respectively, and GATA2 p.Arg361His has been reported as an acquired mutation in acute myelogenous leukemia (AML)[35]. Trp360 resides in a LWRR motif that is conserved across N- and C-terminal zinc fingers in all human GATA proteins (Appendix Figure A.1A). Mutation of the buried hydrophobic Trp360 to a basic arginine residue would be predicted to disrupt the zinc-finger domain folding (Appendix Figure A.1B). These data support recent reports of cryptic presentations of GATA2 mutations without features of MonoMac syndrome, Emberger syndrome, or familial MDS in patients with BMF/MDS[34,36].

One patient (FH-178) presented with MDS. His antecedent medical history was significant for multiple congenital anomalies, neutropenia, and thrombocytopenia (Table 2.3). A heterozygous mutation (c.567C>G [p.Tyr189*]) in *RUNX1* was identified. Heterozygous *RUNX1* mutations cause a familial platelet disorder with propensity to develop myeloid malignancy[37,38]. RUNX1 p.Tyr189* truncates the protein C-terminal to the RUNT domain, with loss of the transactivation domain, and is predicted to function in a dominant negative fashion[39].

One patient (FH-70) was originally diagnosed at an outside hospital with Shwachman-Diamond syndrome based on his clinical presentation of failure to thrive, neutropenia, enzyme-responsive steatorrhea, and low fecal elastase. He lacked mutations in *SBDS* and expressed normal levels of SBDS protein. Telomere lengths fell between the 1st and 10th percentiles for age in all 6 leukocyte subsets tested for FH-70, and hence did not raise prior clinical suspicion for dyskeratosis congenita using published criteria[27]. We identified a hemizygous mutation in *DKC1* (c.-141C>G) located in the promoter of the *DKC1* gene. This promoter mutation was not found in normal control databases dbSNP138, the Exome Variant Server, nor 1000Genomes Project.

Whole exome sequencing analysis of this family did not identify alternative candidate genes. This promoter mutation had been previously reported in two unrelated kindreds with X-linked dyskeratosis congenita and segregated with the dyskeratosis congenita (DC) phenotype[15,40]. This mutation is located in a GC-rich element required for Sp1 transcription factor binding[41] and leads to reduced transcript levels of *DKC1*[15]. In *DKC1* (c.-141C>G) patient-derived CD34+ cells, expression of $DKC1^{WT}$, but not catalytically-inactive $DKC1^{Asp125Ala}$, rescued hematopoietic colony formation[15]. The absence of telomere lengths shorter than the first percentile was consistent with prior reports of other DC patients carrying this promoter mutation in *DKC1*[15]. Thus, broad genetic screening identified a clinically unsuspected genetic syndrome presenting in an atypical fashion.

Patient CH-103 was originally diagnosed with Seckel syndrome based on his presentation with short stature, microcephaly, bird-like facies, and mental retardation (Table 2.3). He subsequently developed marrow failure evolving to MDS and AML. Seckel syndrome with marrow failure is associated with mutations in the *ATR* gene, but no deleterious *ATR* mutations were identified for CH-103. Instead, CH-103 harbored biallelic truncating mutations *LIG4* (c.2440C>T [p.Arg814*])[42] and *LIG4* (c.1751_1755delTAAGA [p.Ile584Argfs*2]). *LIG4* encodes DNA ligase IV. *LIG4* mutations are associated with an autosomal recessive syndrome that, like Seckel syndrome, is characterized by microcephaly, facial dysmorphism, growth retardation, developmental delay, and pancytopenia, but MDS and AML have not been previously reported for LIG4 syndrome[42]. DNA ligase IV functions in DNA double-strand break repair via non-homologous end-joining and V(D)J recombination[43]. Patients with LIG4 syndrome exhibit sensitivity to ionizing radiation, so the correct diagnosis informs choice of treatments and imaging modalities. Patients with LIG4 syndrome develop pancytopenia[42,44] and lymphoid malignancies[45–47]. However, MDS and AML were not previously associated with LIG4 syndrome and thus directed testing for *LIG4* mutations had not been pursued for this patient. Thus, broad

multi-gene sequencing approach overcomes limitations posed by incomplete published knowledge of the range of clinical phenotypes for these syndromes.

2.5 Discussion

Currently, the selection of specific genetic tests for a given patient is driven by clinical suspicion based on history, physical exam, and laboratory evaluation. Our results demonstrate that multi-gene screening identifies patients harboring mutations in known BMF/MDS genes which were clinically unsuspected and therefore were not diagnostically pursued. Our study subjects were selected after a diagnostic workup including prior screening for cryptic presentations of inherited marrow failure syndromes such as chromosomal breakage testing for Fanconi anemia[26], telomere length testing for dyskeratosis congenita[48,49], and pancreatic enzyme testing for Shwachman-Diamond syndrome[50]. Our results therefore underestimate the frequency of genetic causes of BMF/MDS in children, young adults, and adults with suggestive family histories.

Of the patients for whom no mutations were found in any of the 85 BMF/MDS genes, 47 patients had family history and/or syndromic features suggestive of inherited, or at least constitutional, disease. Our results therefore suggest that additional as yet unidentified genes remain to be found for BMF and MDS. Conversely, of the eight patients for whom constitutional damaging mutations were identified, only four had a suggestive family history. Of the genetic diagnoses in the absence of a family history, two most likely involve *de novo* mutation (of *GATA2* in FH-82 and FH-154), one involved recessive inheritance (of *LIG4* in CH-103), and one occurred in an adopted child with no available family history (*RUNX1* in FH-178).

Substantial phenotypic overlap among inherited BMF/MDS syndromes as well as pleiotropy and variable expressivity within syndromes further complicate diagnosis based purely on clinical

presentation. Identification of causal genetic lesions thus plays a critical role in the diagnostic workup. Our multiplexed genetic approach resolved the diagnosis of two patients, FH-70 and CH-103, by simultaneously revealing the absence of mutations in genes associated with their initial clinical diagnoses (*SBDS* and *ATR*, respectively) and identifying pathogenic mutations in clinically unsuspected genes (*DKC1* and *LIG4,* respectively).

For the purposes of clinical diagnostic testing, custom targeted gene panels offer important advantages over whole exome sequencing. Clinical genetic testing requires sensitive and accurate mutation detection across a defined set of clinically actionable genes with reporting of results in a timely manner. Whole exome sequencing, while well-suited for gene discovery, does not ensure deep coverage across all exons of clinical interest. In contrast, high depth of coverage across all genes of interest is achieved through targeted gene capture[51]. The deep coverage afforded by this targeted sequencing approach detected CNVs with a single assay, in contrast to alternative approaches combining targeted capture, array comparative genomic hybridization (aCGH), and RNA sequencing[52]. This is especially important for inherited BMF/MDS which are frequently caused by deletions in ribosomal protein genes[53,54], *RUNX1*[37,55], and *FANCA*[56,57]. Finally, whole exome sequencing queries thousands of genes unrelated to BMF/MDS and thus may reveal incidental mutations not immediately relevant to the diagnosis that prompted the sequencing. Patients for whom a germline cause for his or her disease is not revealed by MarrowSeq may benefit from whole exome sequencing for gene discovery in a research setting. New BMF/MDS genes can be incorporated into probe sets for MarrowSeq as they are reported.

These results demonstrate the clinical utility of broad screening of apparently idiopathic cases of BMF and MDS to identify cryptic presentations of inherited disease. Genomic analysis revealed previously unreported phenotypes due to mutations in known genes for syndromic disease. The

integration of genomic analysis with clinical and laboratory evaluation in the diagnosis of patients with BMF/MDS can guide clinical management. Multiplexed, clinically unbiased genetic screening provides a powerful approach to elucidate the genetic heterogeneity and phenotypic complexity of inherited syndromes.

2.6 Figures

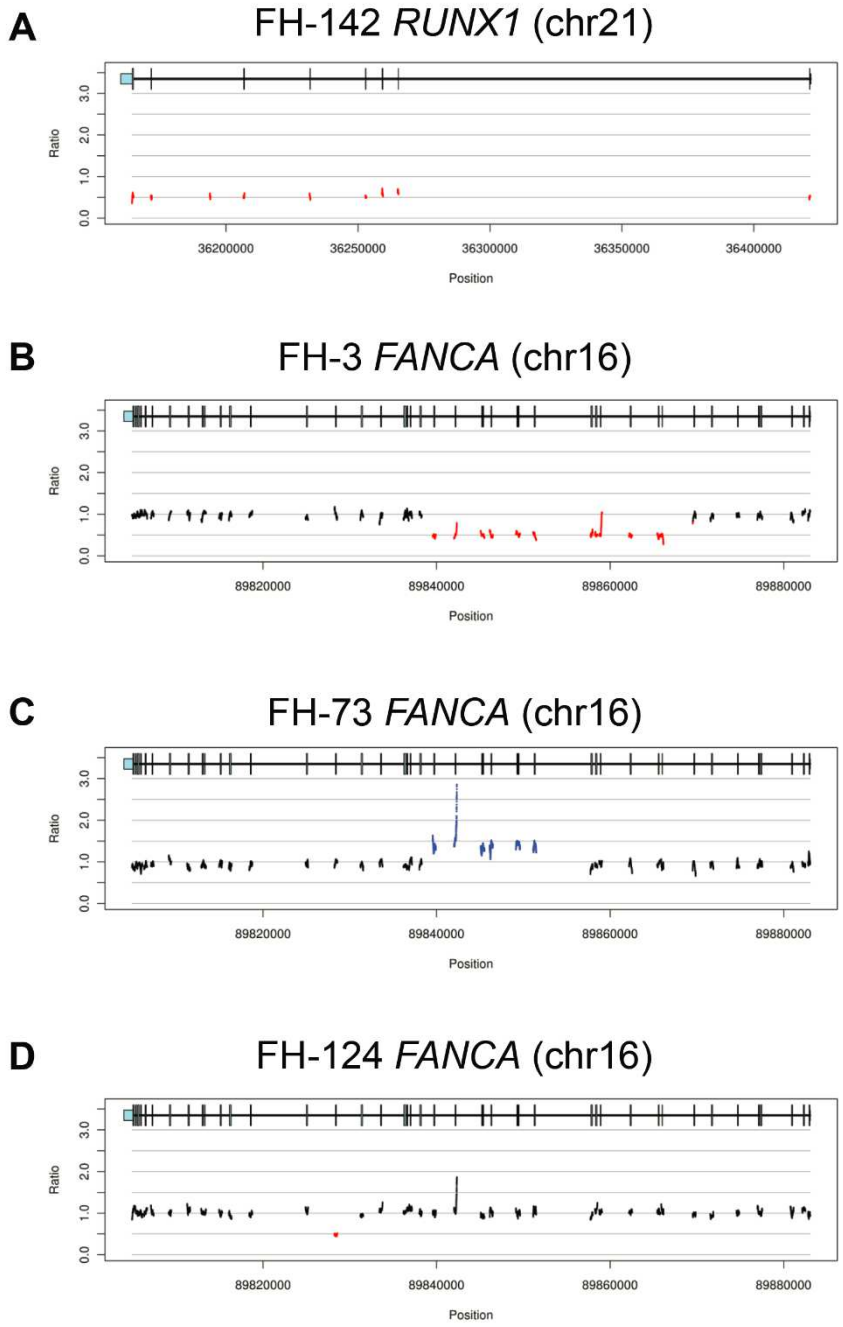


Figure 2.1. Detection of Genomic Copy Number Variants. Ratios of sample to median corrected depth of coverage within a flow cell lane are plotted across targeted genomic regions of the indicated gene. Diploid bases are shown in black. Deletions and duplications are shown in red and blue, respectively. Genomic positions of exons (vertical bars) and untranslated regions (light blue rectangles) are shown above ratio plots. (A) Whole gene deletion of *RUNX1*. No diploid bases were present in this region. (B) Deletion of *FANCA* exons 9-22. (C) Amplification of *FANCA* exons 15-22. (D) Deletion of *FANCA* exon 29.

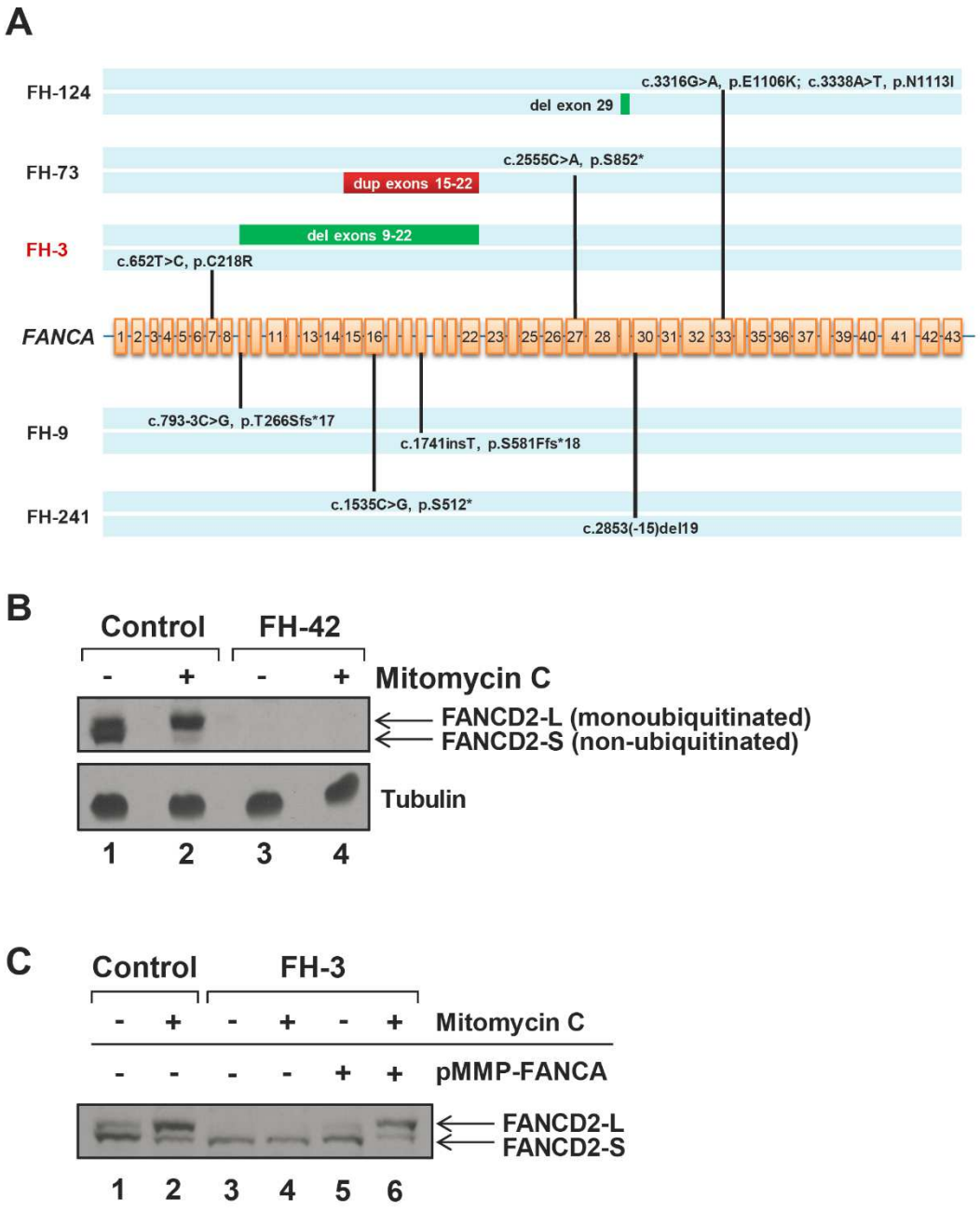


Figure 2.2. Targeted Gene Capture Correction of Fanconi anemia subtype assignment. (A) Biallelic *FANCA* mutations identified by MarrowSeq in five patients. FA patient FH-3 (highlighted in red) was non-ACG subtype by clinical complementation testing. (B) Protein extracts of bone marrow fibroblasts isolated from healthy controls or Fanconi anemia patient FH-42 (*FANCD2*, p.[Leu683Pro];[Glu906Ilefs*4]) were immunoblotted for FANCD2 with or without 24-hour mitomycin C treatment. Fibroblasts from FH-42 exhibit low FANCD2 protein expression (lanes 3 and 4) in comparison to cells from controls (lanes 1 and 2). α-tubulin was used to ascertain equivalent protein loading. (C) Functional validation of Fanconi anemia subtype A in FA patient FH-3 (FANCA p.[Cys218Arg];[Val265Leufs*8]). Protein extracts of bone marrow fibroblasts isolated from healthy control or FH-3 were immunoblotted for FANCD2 with

or without 24-hour mitomycin C treatment. Fibroblasts from healthy control show both non-ubiquitinated (FANCD2-S) and monoubiquitinated (FANCD2-L) FANCD2 forms (lane 1), with an increased ratio of monoubiquitinated FANCD2 relative to non-ubiquitinated FANCD2 upon mitomycin C treatment (lane 2). Fibroblasts from FH-3 show only the non-ubiquitinated FANCD2-S form with and without mitomycin C (lanes 3 and 4). FANCD2 monoubiquitination is restored upon infection with a pMMP retroviral vector encoding the wild-type *FANCA* cDNA (lanes 5 and 6).

2.7 Tables

Table 2.1. Inherited bone marrow failure and myelodysplastic syndrome genes

Dyskeratosis congenita	Fanconi anemia	Diamond-Blackfan anemia	Congenital neutropenia	Other inherited BMF/MDS	Familial MDS/Leukemia	AML and MDS	
CTC1	*FANCA*	*GATA1*	*ELANE*	*ABCB7*	*CBL*	*ABL1*	*NPM1*
DKC1	*FANCB*	*RPS7*	*G6PC3*	*AK2*	*CEBPA*	*ASXL1*	*NRAS*
NHP2	*FANCC*	*RPS10*	*GFI1*	*ANKRD26*	*GATA2*	*BCL2L11*	*PML*
NOP10	*BRCA2 (FANCD1)*	*RPS17*	*HAX1*	*ATR*	*PAX5*	*BCOR*	*PRPF40B*
RTEL1[a]	*FANCD2*	*RPS19*	*WAS*	*LIG4*	*RUNX1*	*BCR*	*RARA*
TERC	*FANCE*	*RPS24*		*MPL*		*BRAF*	*RPS14*
TERT	*FANCF*	*RPS26*		*NBN*		*DNMT3A*	*SF1*
TINF2	*FANCG*	*RPL5*		*RMRP*		*ETV6*	*SF3A1*
WRAP53	*FANCI*	*RPL11*		*SBDS*		*FLT3*	*SF3B1*
	BRIP1 (FANCJ)	*RPL35A*		*SRP72*		*IDH1*	*SRSF2*
	FANCL					*IDH2*	*TET2*
	FANCM					*JAK2*	*TP53*
	PALB2 (FANCN)					*KIT*	*U2AF1*
	RAD51C (FANCO)					*KRAS*	*U2AF2*
	SLX4 (FANCP)					*MET*	*WT1*
	ERCC4 (FANCQ) [a]					*MLL*	*ZRSR2*

[a]Genes added after initial screens

Table 3.2. Characteristics of patients with idiopathic BMF/MDS

Characteristic	Pediatric patients (N = 58)	Adult patients (N = 13)
Sex		
Male	32 (57%)	5 (38%)
Female	26 (43%)	8 (62%)
Median age in years (range)	8.5 (1-18)	24 (20-67)
BMF and MDS		
BMF only	51 (88%)	12 (92%)
MDS only	4 (7%)	1 (8%)
BMF and MDS	3 (5%)	0 (0)
Family history of related phenotypes	26/56 (46%)	6/11 (55%)
Congenital physical or structural anomalies	28 (48%)	2 (15%)
Short telomeres	8/38 (21%)	3/10 (30%)

BMF, bone marrow failure; MDS, myelodysplastic syndrome

Table 3. Clinical features and genetic diagnoses of patients with previously unclassified BMF/MDS

ID	Sex	Age (yr)	Marrow pathology	Hematologic	Immunologic	Transplant	Family history	Congenital physical anomalies	Other clinical features	Initial diagnosis	Gene	Mutation
FH-181	F	22	Hypocellular, transient trisomy 8, no morphologic dysplasia	Neutrophils 900-1200/µl, HCT 34% (normocytic), monocytes 0-20/µl (reference range 0 – 800)	Normal quantitative immunoglobulins	-	MDS/AML (maternal uncle), hematologic malignancies with HSCT (maternal niece and maternal grandmother)	None	2 episodes of pneumonia, 1 episode of severe skin infection of thumb requiring debridement	Marrow failure vs. hypocellular MDS	*GATA2*	c.988C>T (p.R330*)
CH-119	M	12	Hypocellular, transient trisomy 8, mild erythroid and myeloid morphologic dysplasia	Pancytopenia, monocytes 0-32/µl (reference range 0-900)	-	Mismatched unrelated donor transplant complicated by engraftment failure, died during second transplant with matched unrelated donor	Leukopenia (mother, maternal aunt, maternal uncle), neutropenia (maternal cousins)	None	ICU admission for febrile illness with hepatic and pulmonary complications (pulmonary hemorrhage and bronchiolitis obliterans)	Marrow failure	*GATA2*	c.1078T>A (p.W360R)
FH-202	F	12	Hypocellular, mild tri-lineage morphologic dysplasia	Transfusion-dependent anemia, platelets 70k/µl, neutrophils 500/µl, monocytes 10-40/µl (reference range 0-900)	Decreased B-cell precursors with normal mature B-cell numbers	-	Pre-B ALL (maternal cousin, age 3y), breast cancer (maternal aunt, age 45y)	None	Hypothyroidism, psoriasis, erythema multiforme, abdominal migraine, dysplastic nevus	Marrow failure	*GATA2*	c.1082G>A (p.R361H)
FH-154	F	17	Hypocellular, mild multi-lineage morphologic dysplasia	Pancytopenia, red cell macrocytosis, monocytes 90/µl (reference range 0-900)	-	-	None	Strabismus, hearing loss	Hematoma after tooth extraction	Marrow failure	*GATA2*	c.1084C>T (p.R362*)
FH-82	F	16	Hypocellular, mild morphologic tri-lineage dysplasia	Neutrophils 740/µl, monocytes 0-171/µl (reference range 0-900)	Lymphocytes 860/µl, CD2+ 834/µl, CD19+ 17/µl, abnormal lymphocyte stimulation with tetanus and candida	-	None	Sensorineural hearing loss, pseudotumor cerebri, onycholysis	Developmental delay, anxiety, recurrent abdominal pain and vomiting, hepatic transaminitis, abdominal nevus	Marrow failure	*GATA2*	c.1084C>T (p.R362*)
FH-178	M	12	Hypocellular, del(5)(q15q33) clone	Platelets 60-70k/µl, Neutrophils 900-1200/µl.	-	-	Unknown (adopted)	Chiari I malformation, scoliosis, umbilical hernia	Myopathy, chronic obstructive pulmonary disease	MDS	*RUNX1*	c.567C>G (p.Y189*)
FH-70	M	15	Hypocellular, no morphologic dysplasia	Neutrophils 1000-1500/µl	-	-	Cytopenias, marrow hypocellularity, diarrhea (two siblings)	Short stature, laryngomalacia, Narrow rib cage	eczema, steatorrhea, clubbed digits, chronic cough, encephalitis, angioedema, chronic joint pain, dry eyes	Shwachman-Diamond syndrome	*DKC1*	c.-141C>G (hemizygous)
CH-103	M	10	Hypocellular, tri-lineage morphologic dysplasia. Progressed to MDS and AML with del(7)q21q31-32)[9],idem +21[10]_46,XY[1]	Transfusion-dependent anemia and thrombocytopenia, Neutrophils <500 µl, macrocytosis (MCV 106fL)	Low IgG/IgM, low B- and T-cell counts, normal lymphocyte stimulation with mitogens	Matched unrelated donor, death from transplant complications	None	Primordial dwarfism, bird-like facies, undescended testes	Intrauterine growth retardation, failure to thrive, otitis media, developmental delay, renal dysplasia	Atypical Seckel syndrome	*LIG4*	c.[2440C>T] (p.R814*); [1751_1755del] (p.I584Rfs*2)

ALL, acute lymphoblastic leukemia; AML, acute myeloid leukemia; F, female; HCT, hematocrit; HSCT, hematopoietic stem cell transplant; ICU, intensive care unit; M, male; MCV, mean corpuscular volume, MDS, myelodysplastic syndrome

Chapter 3 - Germline *ETV6* Mutations in Familial Thrombocytopenia and Hematologic Malignancy

Note: This chapter is based on the following published paper:

Michael Y. Zhang, Jane E. Churpek, Siobán B. Keel, Tom Walsh, Ming K. Lee, Keith R. Loeb, Suleyman Gulsuner, Colin C. Pritchard, Marilyn Sanchez-Bonilla, Jeffrey J. Delrow, Ryan S. Basom, Melissa Forouhar, Boglarka Gyurkocza, Bradford S. Schwartz, Barbara Neistadt, Rafael Marquez, Christopher J. Mariani, Scott A. Coats, Inga Hofmann, R. Coleman Lindsley, David A. Williams, Janis L. Abkowitz, Marshall S. Horwitz, Mary-Claire King, Lucy A. Godley, and Akiko Shimamura. Germline *ETV6* mutations in familial thrombocytopenia and hematologic malignancy. *Nature Genetics*, 47(2), 180-5 (2015).

3.1 Abstract

We report germline missense mutations in *ETV6* segregating with the dominant transmission of thrombocytopenia and hematologic malignancy in three unrelated kindreds, defining a new hereditary syndrome featuring thrombocytopenia with susceptibility to diverse hematologic neoplasms. Two mutations, p.Arg369Gln and p.Arg399Cys, reside in the highly-conserved ETS DNA binding domain. The third mutation, p.Pro214Leu, lies within the internal linker domain which regulates DNA binding. These three amino acid sites are hot-spots for recurrent somatic mutation in malignancies. Functional studies show that the mutations abrogate DNA binding, alter subcellular localization, decrease transcriptional repression in a dominant negative fashion, and impair hematopoiesis. These familial genetic studies reveal a central role for *ETV6* in hematopoiesis and malignant transformation. The identification of germline predisposition to cytopenias and cancer informs diagnosis and medical management of at-risk individuals.

3.2 Introduction and Results

Few genes predisposing to familial myelodysplastic syndrome (MDS) and acute leukemia have been identified to date. The genes currently known are *RUNX1*[37], *CEBPA*[58], *GATA2*[31,59], *ANKRD26*[16,60], and *SRP72*[61] for MDS and acute myelogenous leukemia (AML), and *PAX5*[62,63] and *TP53*[64,65] for acute lymphoblastic leukemia (ALL). However, most cases of familial MDS-leukemia remain unexplained.

We studied a family of German and Native American ancestry (Family A) with genetically undefined familial thrombocytopenia and malignancy (Figure 3.1 and Appendix B Supplementary Note). Exome sequencing of II-4, II-5, III-1, III-2, and III-3 revealed five protein-

altering variants, in *ETV6*, *TOP3B*, *GPR144*, *ITGA8*, and *PLEC*, affecting evolutionarily-conserved amino acids and segregating with thrombocytopenia and malignancy under the assumption of an autosomal dominant mode of inheritance (Appendix Table B.1). Sanger sequencing of these five mutations in II-1 and II-3 revealed only one variant to be absent in both unaffected individuals: a heterozygous germline *ETV6* variant, c.1195C>T (NM_001987.4), encoding p.Arg399Cys (NP_001978.1) (Appendix Figure B.1a). The proband (III-2) of Family A demonstrated easy bruising in infancy and menorrhagia in her teenage years. Affected members of Family A also developed diverse hematologic malignancies including myelodysplastic syndrome in III-2 at age 17, pre-B cell acute lymphoblastic leukemia (ALL) in III-1 at age 7.5, and multiple myeloma in II-5 at age 51 (Table 3.1). Additionally, subject II-5 developed stage III colorectal adenocarcinoma at age 45.

Targeted sequencing of *ETV6* and 84 additional genes associated with marrow failure and MDS/AML (Appendix Table B.2)[66] for an additional 55 individuals with idiopathic familial leukemia or MDS (all lacking germline *GATA2*, *RUNX1*, *CEBPA*, and *PAX5* mutations), and 153 individuals with idiopathic cytopenias and/or bone marrow failure identified two additional families with thrombocytopenia and hematologic malignancy harboring germline *ETV6* mutations. Family B, of Scottish ancestry, harbored the heterozygous *ETV6* variant c.1106G>A (p.Arg369Gln) (Figure 3.1). Affected individuals in Family B had thrombocytopenia with petechiae, and epistaxis. I-1 developed chronic myelomonocytic leukemia (CMML) at age 82. III-8 was diagnosed with stage IV colon cancer at age 43. DNA sequencing of skin fibroblasts from the proband (II-1) of Family C (Figure 3.1), of African-American ancestry, revealed a heterozygous *ETV6* variant c.641C>T (p.Pro214Leu). II-1 had a long history of nosebleeds and menorrhagia. She was found to have thrombocytopenia unresponsive to standard therapies for immune thrombocytopenia. At age 50, she developed T-/myeloid mixed phenotype acute leukemia (MPAL). Following standard induction chemotherapy, she had delayed recovery of

both platelets and red blood cells and remained transfusion-dependent for over 5 months until undergoing allogeneic hematopoietic stem cell transplantation. During this interval, she had two bone marrow biopsies without evidence of residual leukemia or signs of early regeneration of normal hematopoiesis.

The *ETV6* variant segregation pattern was consistent with a dominant transmission pattern of thrombocytopenia and elevated cancer risk. All individuals who carried an *ETV6* variant had thrombocytopenia, and all individuals tested who developed a hematologic malignancy and/or thrombocytopenia carried an *ETV6* variant (Supplementary Table 3).

The three *ETV6* variants were absent from public databases dbSNP139, the Exome Variant Server, and the 1000 Genomes Project (see URLs). No germline copy number changes in *ETV6* were found in the affected members. No damaging germline mutations were found in *RUNX1, CEBPA, GATA2*, *SRP72*, *ANKRD26*, *TP53*, or *PAX5* or additional marrow failure genes (Supplementary Table 2) in any of the affected individuals.

ETV6 encodes the ETS family transcriptional repressor, Ets variant 6. ETV6 harbors a highly-conserved ETS DNA binding domain shared by all ETS family proteins. Arg369 and Arg399 reside in the second β-sheet and third α-helix of the ETV6 ETS domain, respectively (Figure 3.2a). Arg399 directly contacts DNA at the first guanine of the ETS binding element GGA(A/T) via bidentate hydrogen bonds (Figure 3.2b)[67]. Molecular modeling of the p.Arg399Cys substitution predicts a weakened interaction with DNA[68] (Figure 3.2c). Arg369 is involved in a hydrogen bond with the backbone carbonyl oxygen of Arg414. Arg414 is involved in electrostatic interactions with DNA upstream (5') of the GGA(A/T) motif[67]. Thus the p.Arg369Gln mutation may reduce ETV6 DNA binding via destabilization of the ETS domain and/or by altering the Arg414-DNA interaction (Appendix Figure B.2). DNA binding of murine Etv6 at the

ETS domain is auto-inhibited via a C-terminal inhibitory domain (CID, aa 426-436)[67,69,70]. The p.Pro214Leu mutation resides in a linker inhibitory domain (aa 127-331) that indirectly promotes DNA binding by attenuating the inhibitory effects of the CID[69]. Thus, all three mutations fall within ETV6 domains affecting DNA binding. The linker domain is additionally important for the interaction of ETV6 with transcriptional corepressor complexes[71].

The effect of ETS domain p.Arg369Gln and p.Arg399Cys mutations on DNA binding was tested by electrophoretic mobility shift assay using a DNA probe containing a consensus ETS binding site. A shift in mobility of the DNA probe was detected with 50nM of the purified recombinant wild-type ETV6 ETS domain, whereas no shift was observed following the addition of up to 500 nM of mutant ETV6 p.Arg369Gln or p.Arg399Cys ETS domains (Figure 3.2d,e), demonstrating that the p.Arg369Gln and p.Arg399Cys mutations abrogate DNA binding of ETV6.

Fluorescence microscopy of EGFP-tagged ETV6 in HeLa cells showed that wild-type ETV6 concentrated in cell nuclei (Figure 3.3a,b). In contrast, ETV6 p.Pro214Leu exhibited predominantly cytoplasmic localization and ETV6 p.Arg369Gln and ETV6 p.Arg399Cys showed reduced nuclear localization (Figure 3.3a,b). Concordant with these fluorescence microscopy data, fractionation of HeLa cells transiently expressing wild-type or p.Arg399Cys *ETV6* cDNAs showed increased ETV6 p.Arg399Cys protein levels in the cytoplasmic fraction and decreased levels in the nuclear fraction compared to wild-type (Appendix Figure B.3). Thus, the p.Pro214Leu, p.Arg369Gln, and p.Arg399Cys mutations alter ETV6 localization. These results concur with prior reports that the ETV6 C-terminal residues 332-452, which includes the ETS DNA binding domain, affect ETV6 nuclear localization[72]. The p.Pro214Leu mutation might affect intracellular localization through indirect effects of the linker region on ETV6 DNA binding[69]. Mutations resulting in predominantly cytoplasmic localization may contribute to a dominant negative effect via oligomerization with WT ETV6 resulting in its cytoplasmic sequestration.

While protein levels were comparable between exogenous WT and mutant ETV6 proteins (Figure 3.4b), potential effects of ETV6 overexpression cannot be ruled out. Additional studies of the molecular mechanisms regulating the intracellular localization of endogenous ETV6 are warranted.

Since ETV6 functions as a transcriptional repressor of promoters harboring ETS binding sites (EBS)[71,73–76], we tested the effects of *ETV6* mutations on transcriptional repression of firefly luciferase reporter constructs containing the *MMP3* or *PF4* promoters, which harbor EBS (Figure 3.4a). While wild-type ETV6 repressed expression of both reporter genes, no repression was seen with the ETV6 mutants (Figure 3.4c,d). Expression of an ETV6 ETS-domain deletion mutant has been previously shown to inhibit wild-type ETV6 transrepression in a dominant negative manner[74]. To test whether the patient-derived missense mutations are dominant negative, the effect of increasing levels of mutant *ETV6* cDNA, co-transfected together with a set quantity of wild-type *ETV6* cDNA, was measured using the *PF4*-firefly luciferase reporter construct. All three patient-derived mutants antagonized wild-type ETV6-mediated repression in a dose-dependent manner (Figure 3.4e, compare 3rd through the 11th bars versus the 2nd bar).

The PNT domain of ETV6 mediates homo-oligomerization, a property required for stable ETV6 binding to DNA harboring tandem EBS[69]. We hypothesized that the transrepression-defective ETV6 missense mutants inhibit transrepression by forming dysfunctional PNT domain-mediated heteromeric complexes with WT ETV6. We introduced into the p.Arg399Cys mutant additional PNT domain missense mutations p.Ala93Asp and p.Val112Glu previously demonstrated to disrupt PNT domain oligomerization[77]. In contrast to oligomerization-competent ETV6 p.Arg399Cys, monomeric ETV6 p.Arg399Cys failed to inhibit wild-type ETV6-mediated repression (Figure 3.4e, compare 12th through the 14th bars versus the 2nd bar). These results

suggest that dominant negative ETV6 mutants inhibit wild-type ETV6 transrepression in an oligomerization-dependent manner.

In murine models, *Etv6* is required for hematopoietic stem cell maintenance[78], but hematopoiesis is unperturbed by heterozygous loss of one *Etv6* allele[79]. To test the effect of dominant negative ETV6 mutants in hematopoietic stem cells, we measured the proliferation of human CD34+ hematopoietic stem/progenitor cells (HSPC) transduced with lentiviral vectors expressing WT or mutant ETV6. Proliferation of CD34+ cells expressing WT ETV6 was similar to that of empty vector (Figure 3.5a). In contrast, proliferation of CD34+ cells expressing any of the three ETV6 mutants was markedly reduced (Figure 3.5a). No increase in apoptosis was noted.

To further compare the functional consequences of the three *ETV6* mutations, we performed RNA sequence profiling of the K562 myeloid cell line expressing wild-type, p.Pro214Leu, p.Arg369Glu, or p.Arg399Cys ETV6. Principal component analysis and K-means clustering revealed similar gene signature patterns for cells expressing any of the 3 missense mutations, which distinctly differed from the expression profiles of cells expressing WT ETV6 (Figure 3.5b). There were 311 genes whose expression was reduced by all three ETV6 mutants compared to WT ETV6 (Supplementary Table 4) and 349 genes whose expression was increased by all three ETV6 mutants compared with WT ETV6 (Supplementary Table 5). GO-seq analysis identified platelet-associated gene sets robustly expressed with WT ETV6 but reduced by all three missense mutants (Supplementary Tables 6 and 7). These data are consistent with the notion that all three ETV6 mutations result in similar impairment of ETV6 function.

To identify mutations acquired during malignant progression in the context of germline *ETV6* mutation, we examined paired tumor and fibroblast samples from Family A for mutations in 194

cancer-related genes utilizing a targeted gene capture panel[23]. No deletion or mutation of the remaining wild-type *ETV6* allele was observed in any of the neoplasms (Supplementary Table 8 and Appendix Figure B.1b). In individual II-5, different sets of somatic mutations were identified in the colon adenocarcinoma (*BRAF, CTNNB1*, *GNAS*, *PTEN*, and *TP53)* versus in the multiple myeloma (*CDK8* and *KMT2A*) (Supplementary Table 8). In the colon cancer, the *BRAF* V600E and *CTNNB1* mutations were early events, followed by mutations in *GNAS* and *PTEN*. The acquisition of multiple distinct mutations of different variant allele fractions within both *GNAS* and *PTEN* was suggestive of convergent subclonal evolution. In individual III-2, sequencing of an MDS sample upon progression to refractory anemia with excess blasts 1 (RAEB-1) identified acquired truncating mutations in *BCOR* and *RUNX1* and an activating mutation in *KRAS,* all present in the same dominant clone (Supplementary Table 9). Sequencing of these mutations in an earlier sample taken prior to the development of excess blasts revealed the *BCOR* and *RUNX1* mutations, but the *KRAS* mutation was absent (Appendix Figure B.4). This indicated that the *KRAS* mutation arose during the progression to high-grade MDS.

3.3 Discussion

Acquired mutations in *ETV6* frequently arise in human cancers. We report the first germline *ETV6* mutations segregating with cytopenias and hematological malignancy. The *ETV6* mutations impair DNA-binding, alter intracellular localization, and reduce transrepressive activity. In addition, the mutants inhibit wild-type ETV6 function via hetero-oligomerization, and expression of these dominant negative mutants in human hematopoietic stem and progenitor cells impair hematopoiesis. Since the publication of this work, our conclusions have been confirmed by two studies which together describe five additional unrelated families with similar phenotypes harboring germline *ETV6* mutations[118,119].

Transcription factors with separable domains responsible for DNA-binding and protein-protein interactions are susceptible to dominant negative alleles. The DNA-binding domain is a frequent hot-spot for substitution mutations which confer dominant negative activity. Indeed, in familial MDS/AML disorders caused by germline *GATA2* mutation, dominant negative missense mutations (e.g. p.Thr354Met) are also observed[31,59]. In ETV6, ETS domain missense mutations at Arg369 and Arg399 have both been previously reported as acquired mutations in mature T-cell ALL[91], colorectal carcinoma[92,93], pre-B cell ALL[94], and CMML[95] (Figure 3.2a). Moreover, the residue orthologous to Arg399 in the ETS domain of PU.1 has been observed to undergo recurrent mutation to cysteine, serine, and histidine in a mouse model of radiation-induced AML[96]. Alternatively, truncation mutations eliminating all or part of the ETS domain but retaining the PNT domain can also act in a dominant negative manner[74,97,98] and are also recurrently acquired in cancer (Figure 3.2a). Thus, heritable and acquired dominant negative *ETV6* mutations affecting the ETS domain are a recurrent pathogenic mechanism.

ETV6 harbors phosphosites at serines 22, 213, and 257[99,100]. Phosphorylation at these sites by mitogen activated protein (MAP) kinases p38 and ERK inhibit ETV6 function[99,100]. These phosphorylated serine residues each have an adjacent downstream proline residue; the serine and proline residues together comprise the minimal MAP kinase phosphorylation motif[101]. At Pro214, acquired mutations to leucine and serine have been observed in colorectal carcinoma[93] and melanoma[102], respectively. Additional somatic mutations within ETV6 phosphorylation motifs observed in cancer include p.Ser22Cys in renal carcinoma[103] and p.Pro258Ser in melanoma[102]. Further investigation of the effect of these mutations on ETV6 phosphorylation, how they affect in ETV6 function, and their role in malignant transformation is warranted.

In addition to thrombocytopenia, multiple affected individuals across Family A and B developed gastrointestinal diseases, including gastroesophageal reflux disease, esophageal strictures, and

colon cancer at early ages (42 and 45) (Table 3.1). Conversely, individuals in Family A and B who did not harbor *ETV6* mutations did not develop these gastrointestinal disorders (Table 3.1). The role of ETV6 in the gastrointestinal tract is unknown. *ETV6* is transcriptionally expressed in the mouse intestine[78,124] and zebrafish gut[125]. In addition, acquired ETV6 translocations and point mutations have been recurrently observed in colorectal carcinoma[92,93], and *ETV6* expression is decreased in colorectal carcinoma[124]. Although none of the subsequently described families with ETV6 mutation had known gastrointestinal disorders[118,119], we anticipate further refinement of the association between ETV6 and non-hematologic manifestations, such as in the gastrointestinal tract, as more families with germline *ETV6* mutations are described.

Autosomal dominant transmission of thrombocytopenia and predisposition to MDS/acute leukemia caused by germline *ETV6* mutations is reminiscent of phenotypes associated with mutations in *RUNX1*[37] and *ANKRD26*[60,104], respectively. *ETV6*, *RUNX1*, and *ANKRD26* are all highly expressed in hematopoietic stem cells and megakaryocyte-erythroid progenitors[105]. Recent evidence suggests that *ANKRD26* is transcriptionally regulated by RUNX1 and the ETS family transcription factor FLI1, and that autosomal dominant thrombocytopenia (THC2)-associated mutations in the 5' UTR of *ANKRD26* alter RUNX1- and FLI1-mediated regulation of *ANKRD26*[106]. The potential intersection of pathways regulated by *ANKRD26*, *RUNX1*, and ETS family transcription factors in megakaryopoiesis and hematopoietic transformation warrants further study.

Somatic point mutations in *ETV6* have been recurrently observed by recent large-scale cancer genome sequencing efforts (Figure 3.2a)[92,93,107–110], but the role of *ETV6* mutations in malignant transformation remained unclear. We identified germline missense *ETV6* mutations affecting amino acids recurrently altered across diverse malignancies (Figure 3.2a). The association of these mutations with cancer predisposition supports a role for *ETV6* point mutations as initiating

events in the early steps of malignant transformation. The study of familial cancer syndromes thus complements cancer genome sequencing approaches to identify driver mutations in malignancy.

3.4 Methods

Subjects and samples. Subjects provided written informed consent in accordance with protocols approved by the institutional review boards of the Fred Hutchinson Cancer Research Center and Seattle Children's Hospital for Family A and The University of Chicago for Family B and C.

Exome sequencing. Family A II-4, II-5, III-1, III-2, and III-3 were subjected to exome sequencing, as previously described[24,80]. Briefly, paired end libraries with 250bp inserts were hybridized to the SeqCap EZ Human Exome Library v2.0 (NimbleGen). Sequencing was performed with 2x101bp reads using SBS v3 on a HiSeq2000 (Illumina). Rare and private variants were classified by predicted function to include all missense, nonsense, frameshift, or splice-site alleles. Variants were filtered based on an autosomal dominant mode of inheritance.

Targeted gene panel sequencing. For *ETV6* mutational screening, capture probes were designed to target all coding exons and 20 basepairs (bp) of flanking sequence for *ETV6* and 84 other genes involved in inherited bone marrow failure and MDS/AML (Supplementary Table 2). Targeted capture, sequencing, and bioinformatic analysis was performed as previously described[21]. Identification of somatic alterations in a panel of 194 cancer genes was performed on paired tumor and fibroblasts samples as previously described[23]. *Cis* or *trans* relationships between variants were determined using Integrated Genome Viewer.

Plasmids. Human *ETV6* cDNA (NM_001987.4) was cloned into pHAGE-CMV-MCS-IRES-ZsGreen (pHAGE)[81] and the resultant plasmid was used for generation of ETV6 mutants (p.Pro214Leu, p.Arg369Gln, p.Arg399Cys, p.[Ala93Asp; p.Val112Glu; p.Arg399Cys]) by QuikChange site-directed mutagenesis (Agilent). Wild-type and mutant human *ETV6* cDNA were cloned into pEGFP-N3 (Clontech). Promoters of human *MMP3* and *PF4* were cloned into pGL3-Basic (Agilent). The ETS domain of human ETV6 (amino acids 335-430) was cloned into pHAT10 (Clontech) for bacterial expression (Supplementary Table 10).

Cell culture. HeLa cells were cultured in DMEM supplemented with 10% fetal bovine serum, 1% glutamine, and 1% penicillin-streptomycin. CD34+ cells were isolated from anonymous discarded full-term human umbilical cord blood using the CD34 Microbead kit (Miltenyi Biotec) as previously described[82]. Cells were cultured in StemSpan SFEM II (StemCell Technologies) supplemented with penicillin/streptomycin and 100 ng/ml each of human stem cell factor, thrombopoietin, IL-6, and Flt-3 ligand (PeproTech). K562 cells were grown in RPMI-1640 with 10% FBS, 1% glutamine, 1% penicillin-streptomycin, and 1 mM sodium pyruvate.

RNA-seq expression analysis. K562 cells were electroporated with the pHAGE ETV6 constructs using the Cell Line Nucleofector Kit V (Lonza) according to manufacturer instructions, grown in growth media for 48 hours and flow sorted for ZsGreen positivity. Positive cells were grown for another 24 hours before lysis in TRIzol (Invitrogen). RNA was extracted with the RNeasy Total RNA cleanup kit (Qiagen). RNA integrity was measured using an Agilent 2200 TapeStation (Agilent Technologies). RNA-seq libraries were prepared from total RNA using the TruSeq RNA Sample Prep Kit (Illumina) and a Sciclone NGSx Workstation (PerkinElmer). Sequencing was performed using an Illumina HiSeq 2500 in rapid-output mode employing a paired-end, 50 base read length (PE50) sequencing strategy. Image analysis and base calling were performed using Illumina's Real Time Analysis software.

RNA-seq data analysis. Reads of low quality were filtered prior to alignment to the reference genome (UCSC hg19 assembly) using TopHat v2.0.12[83]. Counts were generated from TopHat alignments for each gene using the Python package HTSeq v0.6.1[84]. Genes with low counts in greater than 3 samples were removed prior to identification of differentially expressed genes using the Bioconductor package edgeR v3.4.2[85]. A false discovery rate (FDR) method was employed to correct for multiple testing[86]. Differential expression was defined as $|\log_2 (\text{ratio})| \geq 0.585$ (± 1.5-fold) with the FDR set to 5%. K-means cluster analysis was performed for those genes found to be differentially expressed in one or more comparison. The normalized $\log_2$ signal intensities were mean-centered at the gene-level and replicate samples were averaged prior to clustering. The number of clusters was selected using the figure of merit (FOM) method[87]. K-means clustering and cluster number estimation were performed using the TM4 microarray software suite MultiExperimental Viewer (MeV)[88]. Overrepresented GO Biological Process terms comprised of genes found in K-means clusters were identified using the Bioconductor package GOseq[89]. Principal component analysis (PCA) plots were generated using R.

Lentiviral transduction. Lentiviruses were produced by transient transfection of 293T cells using polyethylenimine and concentrated by low-speed centrifugation. CD34+ cells were transduced with pHAGE lentiviral bicistronic vectors expressing wild-type or mutant *ETV6* cDNA and a ZsGreen marker at MOI 10 in the presence of 8 μg/ml hexadimethrine bromide (Sigma-Aldrich)[90]. ZsGreen-positive cells were selected by flow cytometry.

Western blotting. Whole cell extracts were obtained by lysing cells in RIPA buffer (1% NP-40, 0.5% NaDOC, and 0.2% SDS in PBS) with 1 mg/ml Pefabloc (Sigma-Aldrich), 1 μg/ml Pepstatin (Sigma-Aldrich), and 1x Complete EDTA-free protease inhibitor cocktail (Roche). Cell

fractionation was performed using NE-PER Nuclear and Cytoplasmic Extraction Reagents (Pierce). Samples were run on a 10% acrylamide SDS-PAGE gel, transferred onto nitrocellulose, and probed with antibodies against ETV6 (N-19X or H-214, Santa Cruz Biotechnology), α-tubulin (DM1A, Sigma-Aldrich), NPM1 (FC82291, Abcam), or GAPDH (ab9485, Abcam). Western Lightning Plus ECL (PerkinElmer) was used for detection.

Immunofluorescence. HeLa cells were plated on poly-L lysine-coated coverslips and transfected with pEGFP constructs using Attractene Transfection Reagent (Qiagen). After 48 hours, cells were fixed in 4% paraformaldehyde in phosphate-buffered saline, mounted with VECTASHIELD Mounting Media with DAPI (Vector Laboratories), and visualized using a Nikon ECLIPSE E800 microscope.

Recombinant protein expression and purification. Recombinant proteins were expressed and purified as previously described[90]. Purified proteins were dialyzed overnight into 20 mM sodium citrate pH 5.3, 500 mM KCl, 1 mM EDTA, 1 mM DTT, 0.2 mM phenylmethylsulfonyl fluoride, and 10% glycerol.

EMSA probes. DNA probes were modified from Green et al.[69] (Supplementary Table 10). Probes were labeled by 3' biotinylation of the forward strand. Probes were annealed by incubation at 95 °C for 1 minute and slow cooling to room temperature for 2 hours.

Gel shift assay. Protein and probes were incubated in EMSA buffer (25 mM Tris pH 8.0, 50 mM KCl, 1 mM DTT, 10% glycerol, 6 mM $MgCl_2$, 1 mM EDTA, 50 ng/µl poly (dI-dC), 0.1 mg/ml bovine serum albumin) for 20 minutes at room temperature. Samples were run in a 6% acrylamide, 0.5% TBE native gel for 70 minutes at 100 volts at 4 °C. Protein-nucleic acid complexes were transferred to a nylon membrane for 35 mins at 380 mA. Nucleic acids were

cross-linked to the nylon membrane by UV at 120 mJ/cm^2. Biotin-labeled probes were detected on the membrane using the Chemiluminescent Nucleic Acid Detection Module (Pierce).

Luciferase assay. HeLa cells were co-transfected with pGL3 reporter construct, pHAGE expression construct, and pCS2 Renilla luciferase construct using Attractene Transfection Reagent (Qiagen). Empty pHAGE plasmid was added to maintain a constant DNA concentration per transfection. Cells were harvested 48 hours after transfection using Passive Lysis Buffer (Promega). Renilla and firefly luciferase levels were assayed with the Dual Luciferase Report Assay System (Promega) using a GloMax Microplate Luminometer with Dual Injectors (Promega). The pCS2 Renilla luciferase was used to normalize for transfection efficiency.

CD34+ proliferation assay. Cord blood-derived CD34+ cells were purified, transduced, and cultured as described above. Cells were cultured in triplicate. Viable cells, as determined by trypan blue staining, were counted every 2 days.

3.5 Figures

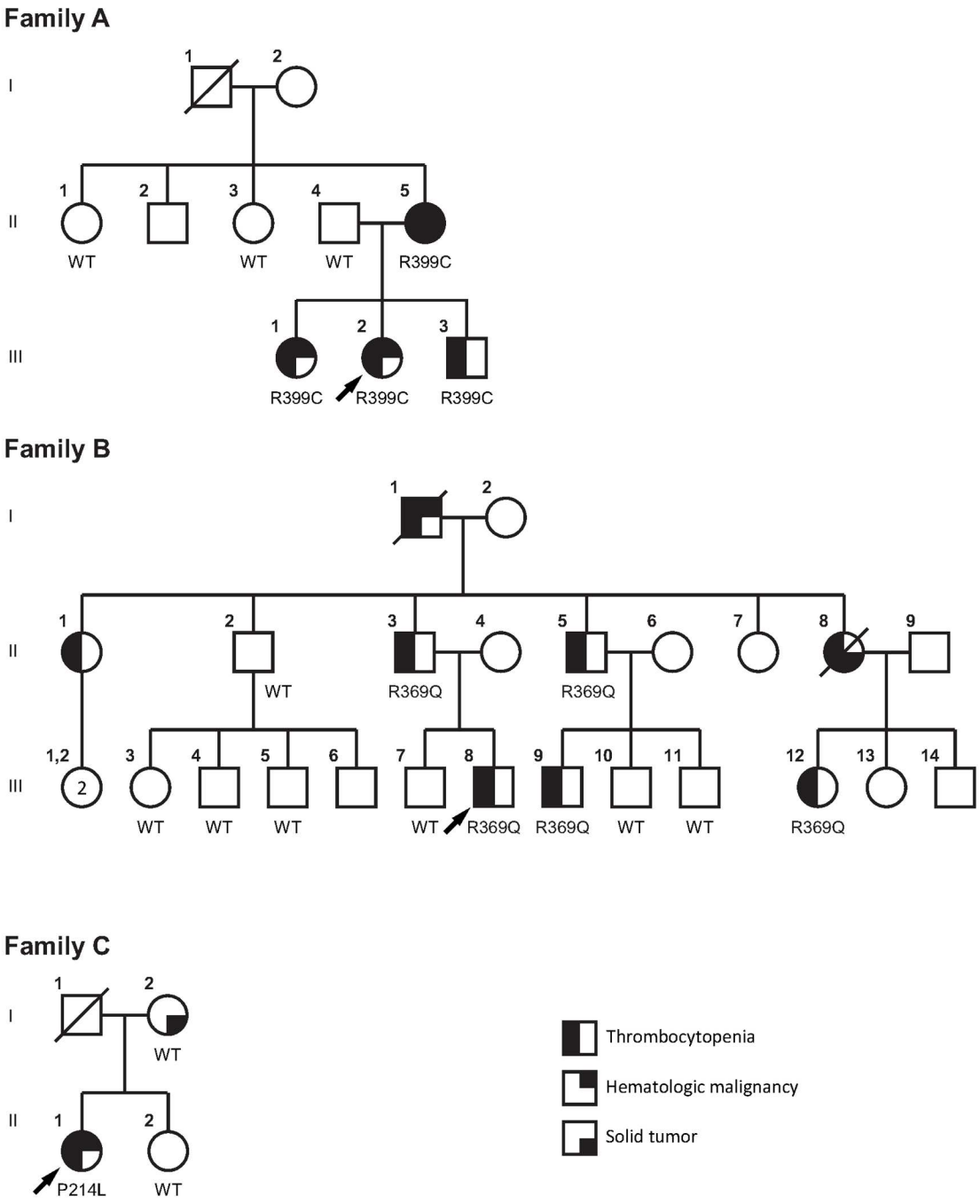
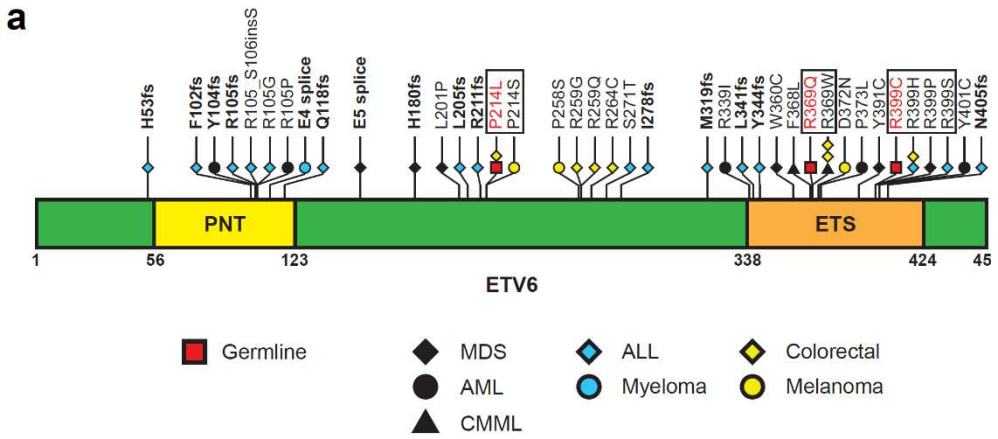
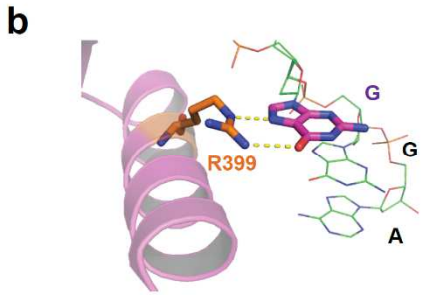
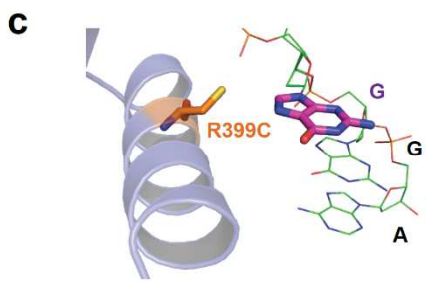
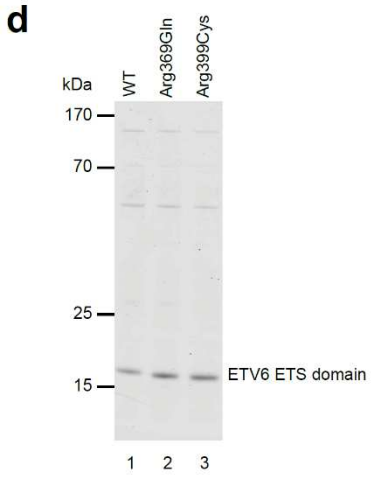


Figure 3.1. Novel ETV6 germline variants p.Pro214Leu, p.Arg369Gln, and p.Arg399Cys in association with thrombocytopenia and hematologic malignancy. Families A, B, and C with ETV6 p.Arg399Cys, p.Arg369Gln, and p.Pro214Leu, respectively, segregating with thrombocytopenia and hematologic malignancy in each family. WT indicates genotyped subjects with only wild-type *ETV6* alleles; R399C, R369Q, and P214L indicate subjects heterozygous for the variant allele. Arrows indicate the proband in each family.









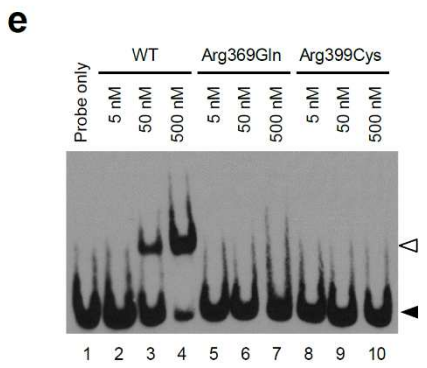


Figure 3.2. Missense mutations in the ETS domain abrogate ETV6 DNA binding.
(a) Positions of germline and somatic mutations in ETV6 relative to the Pointed (PNT) oligomerization and ETS DNA-binding domains. The germline mutations reported in this study are highlighted in red. Somatic mutations affecting the same amino acids as the germline mutations are boxed. Somatic mutations reported in the literature include those associated with

MDS[108,110,111], AML[107,108,112–114], CMML[95], immature T-cell ALL[109], mature T-cell ALL[91], B-cell precursor ALL[94,115], hypodiploid ALL[64], multiple myeloma[116], colorectal adenocarcinoma[92,93], and melanoma[102]. Truncating mutations including nonsense, frameshift, and splice mutations are in bold. (b) Hydrogen bonding (dotted lines) of Arg399 (orange) with guanine (magenta) in the ETS binding element. Protein structure of the murine Etv6 ETS domain (PDB ID: 4MHG)[67] is shown. The ETS domains of murine and human ETV6 have 100% protein sequence identity. (c) Molecular modeling of the Arg399Cys (orange) variant using SWISS-MODEL predicts loss of hydrogen bonding to DNA. (d) Coomassie-stained SDS-PAGE gel with recombinant wild-type (WT), p.Arg369Gln, or p.Arg399Cys HAT-tagged ETV6 ETS domains. (e) Electrophoretic mobility shift assay of ETV6 wild-type (WT), p.Arg369Gln, or p.Arg399Cys ETS domains. Biotinylated EBS DNA probe was incubated with the indicated concentrations of purified, recombinant wild-type or mutant HAT-tagged ETV6 ETS domain protein. Open and closed triangles indicate positions of the protein-bound and unbound probes, respectively.

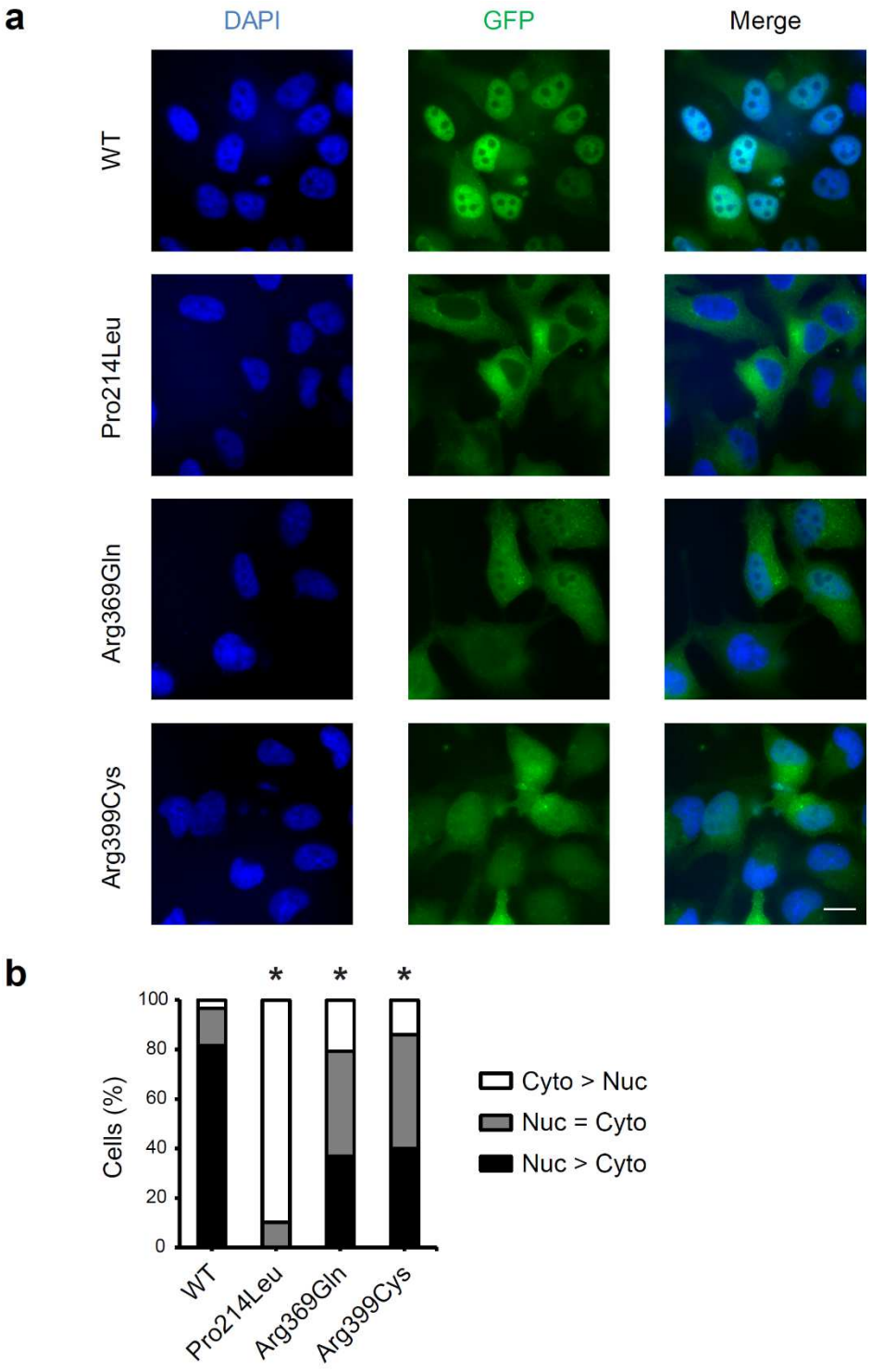


Figure 3.3. ***ETV6* mutation reduces nuclear localization.** (**a**) Fluorescence images of HeLa cells transiently expressing EGFP-tagged ETV6 wild-type, p.Pro214Leu, p.Arg369Gln, or p.Arg399Cys. Scale bar represents 25 μm. (**b**) Percentage of cells exhibiting predominantly nuclear, predominantly cytoplasmic, or equivalent nuclear and cytoplasmic EGFP signal. Three individual experiments were performed, and at least 300 total cells were counted for each condition. Pairwise comparisons between wild-type and each mutant were performed using the Chi-square test (*$P < 10^{-48}$).

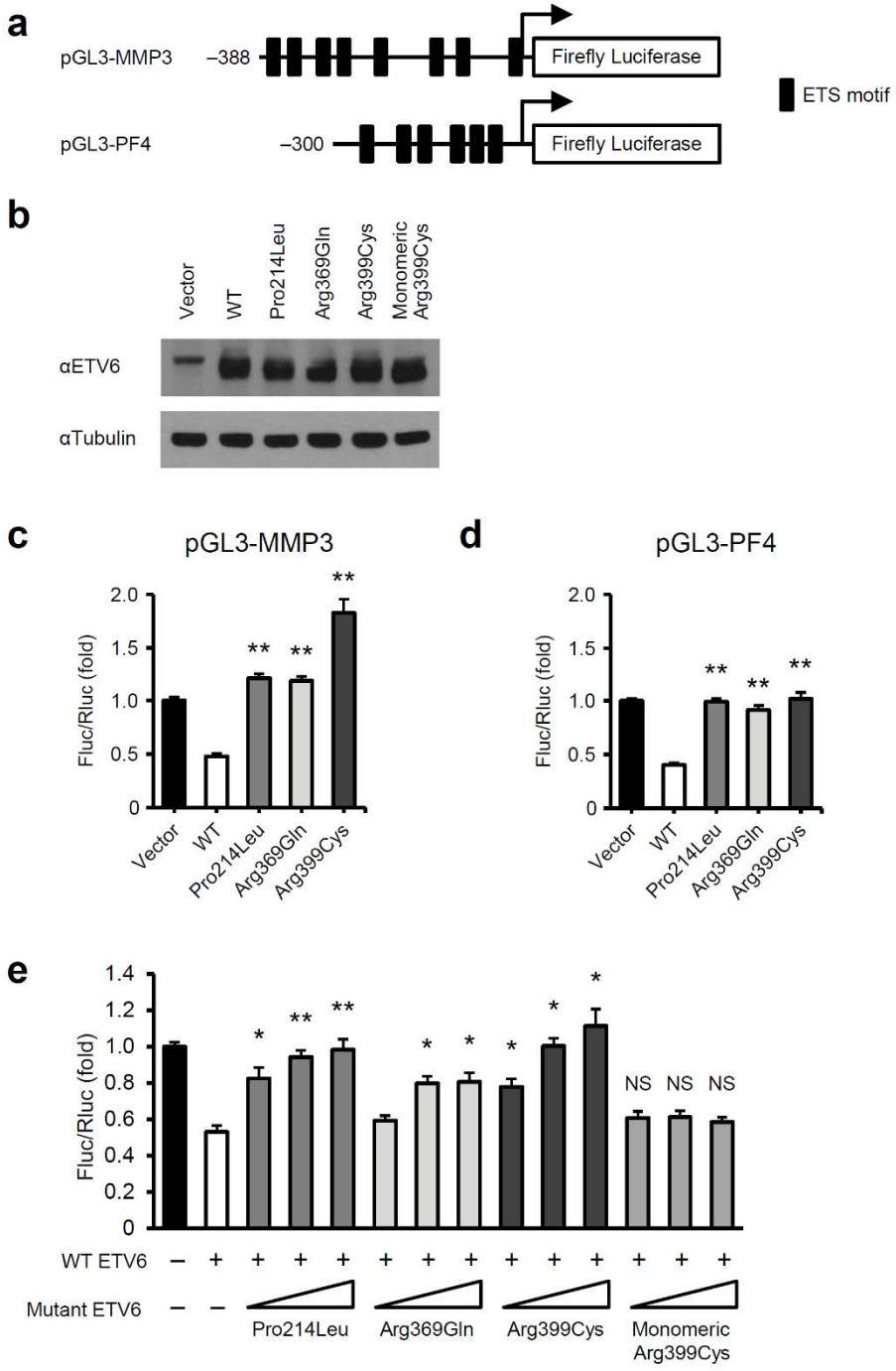


Figure 3.4. ETV6 mutants are deficient in transcriptional repression and act in a dominant negative manner. (**a**) Schematic of pGL3 reporter constructs harboring *MMP3* and *PF4* promoters upstream of firefly luciferase. Black rectangles represent core ETS DNA-binding motifs. (**b**) Western blot analysis of ETV6 expression in HeLa whole cell lysates. (**c**) HeLa cells were co-transfected with the pGL3-MMP3 reporter construct, a pHAGE expression vector (empty vector, wild-type *ETV6*, or mutant *ETV6*), and pCS2 Renilla luciferase. Firefly to Renilla luciferase ratios (Fluc/Rluc) were calculated to control for transfection efficiency. Bars show mean (+ s.e.m.) fold-change of Fluc/Rluc relative to empty vector. Data represents at least 2 individual experiments for each condition with duplicate measurements. Pairwise Student's *t*-tests were performed comparing each condition to WT (**$P < 0.0005$) (**d**) Same as in (**c**) except

the pGL3-PF4 reporter construct was used. Data represents at least 3 individual experiments for each condition with duplicate measurements. Pairwise Student's *t*-tests were performed comparing each condition to WT (**$P < 0.0005$) (**e**) HeLa cells were co-transfected with 50 ng of wild-type *ETV6* expression vector with increasing (50, 150, and 250 ng) amounts of *ETV6* p.Pro214Leu, p.Arg369Gln, p.Arg399Cys, or monomeric p.Arg399Cys expression vectors, pGL3-PF4 reporter construct, and pCS2 Renilla luciferase. Bars show mean (+ s.e.m.) fold-change of Fluc/Rluc relative to empty vector (lane 1). Data represents at least 3 individual experiments for each condition with duplicate measurements. Pairwise Student's *t*-tests were performed comparing each condition to WT alone (lane 2) (*$P < 0.005$ and **$P < 0.0005$; NS, not significant).

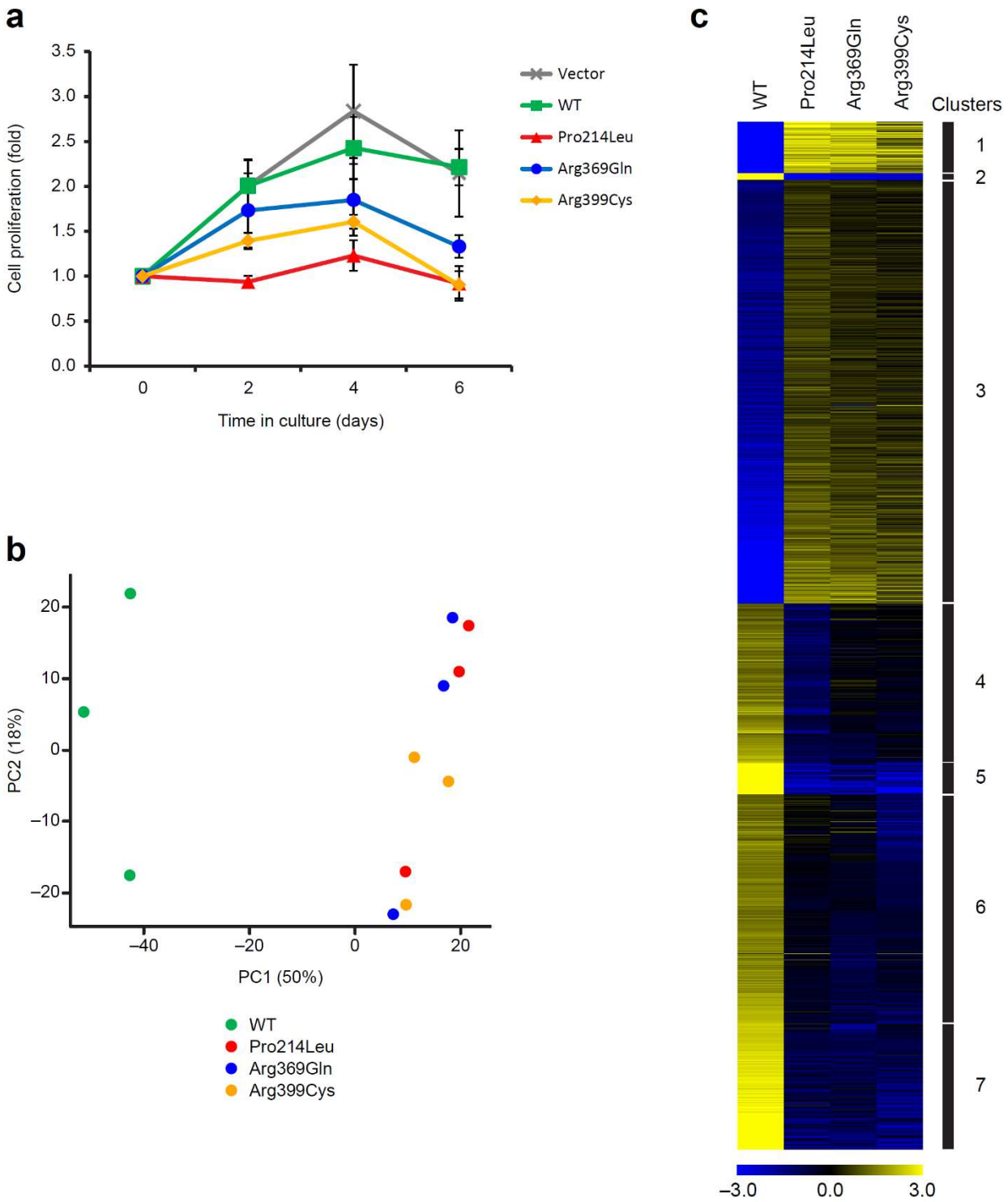


Figure 3.5. ETV6 mutants impair hematopoietic stem cell proliferation and alter the *ETV6* transcriptome. **(a)** Proliferation of human CD34+ cells expressing wild-type (WT), p.Pro214Leu, p.Arg369Gln, or p.Arg399Cys ETV6 cultured under non-differentiating conditions. Viable cells were counted in triplicate every 2 days. Plotted points represent mean ± s.d.

Pairwise Student's *t*-tests were performed comparing each mutant vs WT ETV6 on day 6 (*$P <$ 0.01). (**b**) Genome-wide mRNA expression profiling with WT or mutant ETV6. Principle component analysis (PCA) plot of the first two principle components representing 68% of the total variance in the transcriptome dataset from K562 cells expressing wild-type or the indicated mutant ETV6 species. The data from three independent experiments are shown. **(c)** Heat map showing log2-transformed and mean-centered transcript levels for differentially expressed genes in K562 cells expressing wild-type or the indicated mutant ETV6 species. Differentially expressed genes were partitioned into 7 distinct clusters by k-means clustering using Euclidian distance. Yellow and blue indicate higher and lower expression, respectively.

3.6 Tables

Table 3.1. Clinical features of individuals in the study families

Family	ETV6 mutation status	Individual	Cytopenias	Malignancies[a]	Additional features
A	p.Arg399Cys	II-5	Thrombocytopenia, neutropenia	Stage III colorectal carcinoma (45), multiple myeloma (51)	
A	p.Arg399Cys	III-1	Thrombocytopenia	Pre-B cell acute lymphoblastic leukemia (7)	
A	p.Arg399Cys	III-2 (proband)	Thrombocytopenia, neutropenia, anemia	Refractory anemia (9), refractory anemia with excess blasts type I (21)	Myopathy, GI dysmotility, GERD, developmental delay, seizures, degenerative dental disease, delayed puberty
A	p.Arg399Cys	III-3	Thrombocytopenia, anemia		Myopathy, undefined GI symptoms
B	DNA unavailable	I-1	Thrombocytopenia	Skin cancer, chronic myelomonocytic leukemia (82)	
B	DNA unavailable	II-1	Thrombocytopenia (initially diagnosed ITP)		
B	WT	II-2		Skin cancer	
B	p.Arg369Gln	II-3 (proband)	Thrombocytopenia		Reading disability, GERD
B	p.Arg369Gln	II-5	Thrombocytopenia		Esophageal stricture, GERD
B	p.Arg369Gln (obligate carrier)	II-8	Thrombocytopenia	Stage IV colon cancer (42)	GERD
B	WT	III-3		Skin cancer (35)	
B	WT	III-7			Reading disability
B	p.Arg369Gln	III-8	Thrombocytopenia		Reading disability, GERD
B	p.Arg369Gln	III-9	Thrombocytopenia	Skin cancer (34)	Esophageal stricture, GERD
B	p.Arg369Gln	III-12	Thrombocytopenia		
C	WT	I-2		Colon cancer (68)	
C	p.Pro214Leu	II-1 (proband)	Thrombocytopenia (initially diagnosed ITP)	Mixed phenotype acute leukemia (50)	

ITP, immune thrombocytopenia; GERD, gastrointestinal esophageal reflux disease; GI, gastrointestinal

[a]Parentheses indicate age of diagnosis in years.

Chapter 4 – Discussion

4.1 Summary

We began with the hypothesis that the traditional clinical genetic work-up of BMF/MDS patients in which directed genetic tests are ordered in succession based on clinical suspicion fails to identify underlying inherited disease. We submitted two primary possibilities causing this failure. The first possibility is that the patient may have been manifesting a known inherited BMF/MDS disorder in an atypical or otherwise cryptic fashion such that diagnosis was elusive. The second possibility is that the patient may have an as-of-yet uncharacterized genetic disorder caused by mutation of a gene not previously associated inherited BMF/MDS. We utilized genomic technologies and high-throughput sequencing to address both of these possibilities.

First, we designed the MarrowSeq targeted capture panel to simultaneously sequence 85 BMF/MDS genes with high accuracy and sequencing depth. After validating this platform on patients with previously known mutations, we applied it to interrogate germline mutations in 71 patients with BMF/MDS which have remained unclassified after traditional clinical work-up. Broad sequencing identified the causal genetic lesion in 11% (8/71) of these previously unclassified patients, demonstrating the utility of high-throughput, multi-gene sequencing in the diagnosis of known inherited BMF/MDS disorders in patients with cryptic or atypical presentations of disease. In addition, 47 of the 63 patients for whom no mutations were found in any of the 85 BMF/MDS genes had family history and/or syndromic features suggestive of inherited or constitutional disease, suggesting that additional unidentified genes remain to be discovered for inherited hematopoietic disorders.

Second, we selected patients with unclassified BMF/MDS who had family histories of hematologic dysfunction or malignancy and queried patient and family member samples for

mutations in all protein coding genes using whole exome sequencing. Exome sequencing of a family with thrombocytopenia and diverse hematologic malignancies revealed a segregating germline mutation in the hematopoietic transcription factor *ETV6*. Further screening of additional unclassified families with thrombocytopenia revealed two additional kindreds harboring distinct deleterious mutations in *ETV6* which segregated with disease. We showed that these mutations disrupt localization, DNA binding, hematopoiesis, and targeted transcription. The *ETV6* gene was known to be necessary for hematopoiesis in mice[78,79] and to recurrently undergo somatic mutation, via translocation or point mutation, in a range of hematologic malignancies[115,117]; however, germline mutations in the gene have never previously been associated with an inherited disorder. Our results define a new genetic syndrome characterized by thrombocytopenia and malignant predisposition, and thus expands our knowledge of genes responsible for inherited BMF/MDS. Since the publication of our study detailing these findings, our conclusions have been further confirmed by two studies which together describe five additional unrelated families with similar phenotypes harboring germline *ETV6* mutations[118,119].

4.2 Significance and Future Directions

Our results demonstrate the advantages of incorporating broad unbiased sequencing into the diagnostic work up of pediatric and young adult patients with BMF/MDS. The study demonstrates that the current paradigm of gene-by-gene sequencing misses the diagnosis of inherited disease in a subset of patients. Identification of the correct inherited syndrome is critical for tailoring treatment and appropriate surveillance of malignancies. In addition, the study shows that broad unbiased sequencing using the MarrowSeq platform is able to identify this subset of patients and determine their pathogenic genetic mutation. Knowledge of the precise pathogenic mutations are necessary to guide transplant donor selection, genetic testing of family members, and pre-implantation genetic diagnosis. On the horizon, disease-associated

mutations may be corrected within the patient's hematopoietic cells using gene editing, serving as the first step towards transplantation of autologous genetically-edited, disease-free bone marrow cells[120].

In addition to the eight patients with damaging mutations, 25 variants of unknown significance (VUS) were also identified in the sequencing of our unclassified patient group (Appendix Table A.4 and A.5). VUSs represent variants whose pathogenicity is uncertain due to a lack of supporting functional data. Thus, they fail to provide a diagnosis and must be interpreted with caution in clinical decision-making. As high-throughput sequencing is increasingly applied in the clinical arena, the number of VUSs discovered will rapidly accelerate. Advancement in the interpretation of these VUSs may benefit from pre-generated compendiums of possible variants, available for retrieval during the analysis of patient mutations. The generation of these databases necessitates methodologies to biologically assay variants in a rapid, high-throughput, and cost-effective manner. One such methodologic paradigm is saturation mutagenesis, in which all possible variants in a genetic region (e.g. open reading frame or 5' UTR) are genetically engineered and assayed in parallel for a specific biological function[121]. Recently, this methodology has been applied in human cells to systematically assess the effect of exonic single nucleotide variants (SNVs) introduced by CRISPR/Cas9 genome editing technology on gene transcript abundance and splicing efficiency[122]. A profound technical challenge remains in the extension of these functional genetic methods to assay hematopoietic function. The development of techniques to efficiently create specific genetic alterations in functional hematopoietic stem and progenitor cells represent a large step towards this goal[123]. The systematic functional characterization of the effect of genetic variants on hematopoiesis is critical for the advancement of genetic diagnosis in inherited blood disorders.

The study of genes causing inherited cancer predisposition syndromes sheds light on the functional relevance of acquired mutations in sporadic neoplasms in the general population. While the driving role of *ETV6* translocations in cancer has been widely accepted due to their high recurrence rate and an understanding of their transformative mechanism[126], the significance of *ETV6* point mutations has been less clear. Somatic point mutations in *ETV6* have been recurrently identified by recent large-scale cancer genome sequencing efforts in both liquid and solid neoplasms, albeit at low frequency[92,107,108,127,128]. A recent study reported only a 1.5% overall incidence of *ETV6* point mutations in 970 hematologic malignancies of varying lineages[115]. Our observation of germline missense *ETV6* mutations associated with predisposition to a multitude of hematopoietic malignancies including MDS (RAEB-1), pre-B ALL, multiple myeloma, CMML, and MPAL supports a role for *ETV6* point mutations as initiating events in the early steps of malignant transformation in different hematopoietic lineages. Further study of the prognostic potential of *ETV6* point mutations across liquid and solid malignancies is warranted. The study of familial cancer predisposition syndromes thus complements large-scale cancer genome sequencing in the identification of driver mutations in malignancy.

Appendix A – Supplemental Materials for Chapter 2

A.1 Supplementary Methods

Genomics. Libraries were prepared in 96-well format with a Bravo liquid handling robot (Agilent Technologies). One to two micrograms of genomic DNA were sheared to a peak size of 150 bp using a Covaris E series instrument. DNA was end-repaired, A-tailed, and ligated to Illumina adapters. Libraries were amplified for 5 cycles with flanking primers (PE 1.0 and PE 2.0), and 500 nanograms of purified amplified library were then hybridized to the custom biotinylated cRNA oligonucleotides (Agilent SureSelect Target Enrichment system) for 24 hours at 65^0C. Following a series of washes with increasing stringency, the biotin cRNA-DNA hybrids were purified with streptavidin conjugated magnetic beads. Post-capture PCR was performed with 96 different 6 bp barcoded primers for 14 cycles, and the final concentration of captured library was determined with a TapeStation (Agilent Technologies).

Forty-eight barcoded libraries were pooled in equimolar amount to carry out cluster amplification. Sequencing was performed on an Illumina HiSeq 2500 with a standard paired end 2x101 bp recipe in rapid mode. Yield of Q30 base pairs ranged from 31 to 33 GB.

An average per patient of 291 ± 26 (mean ± SD) variants with $\geq$8X coverage and confirmatory reads from both minus and plus strands were detected. Variants were excluded if present in <5% of reads at the site, if the minor allele frequency was >1% on dbSNP138, the Exome Variant Server or 1000Genomes, or if present in >50% of samples; 23 ± 5 variants per sample remained.

Custom bioinformatic analysis was designed for *SBDS* given the complexities due to gene conversion from its pseudogene, *SBDSP*.[129] Reads were aligned to the *SBDS* gene in isolation,

and variant allele fractions of pseudogene-derived variants were analyzed to detect the presence of gene conversion.

Consequences of splice site mutations for transcription were evaluated by subcloning and Sanger sequencing cDNA generated from patients' cells. Compound heterozygous variants were confirmed to be in *trans* by genotyping DNA from parents, or by subcloning and Sanger sequencing genomic DNA or cDNA.

Fanconi anemia pathway analysis. Bone marrow fibroblasts were infected with a pMMP retrovirus expressing a wild-type FANCA cDNA or empty vector control. Immunoblotting for FANCD2 monoubiquitination was performed as previously described.[130]

A.2 Supplementary Figure

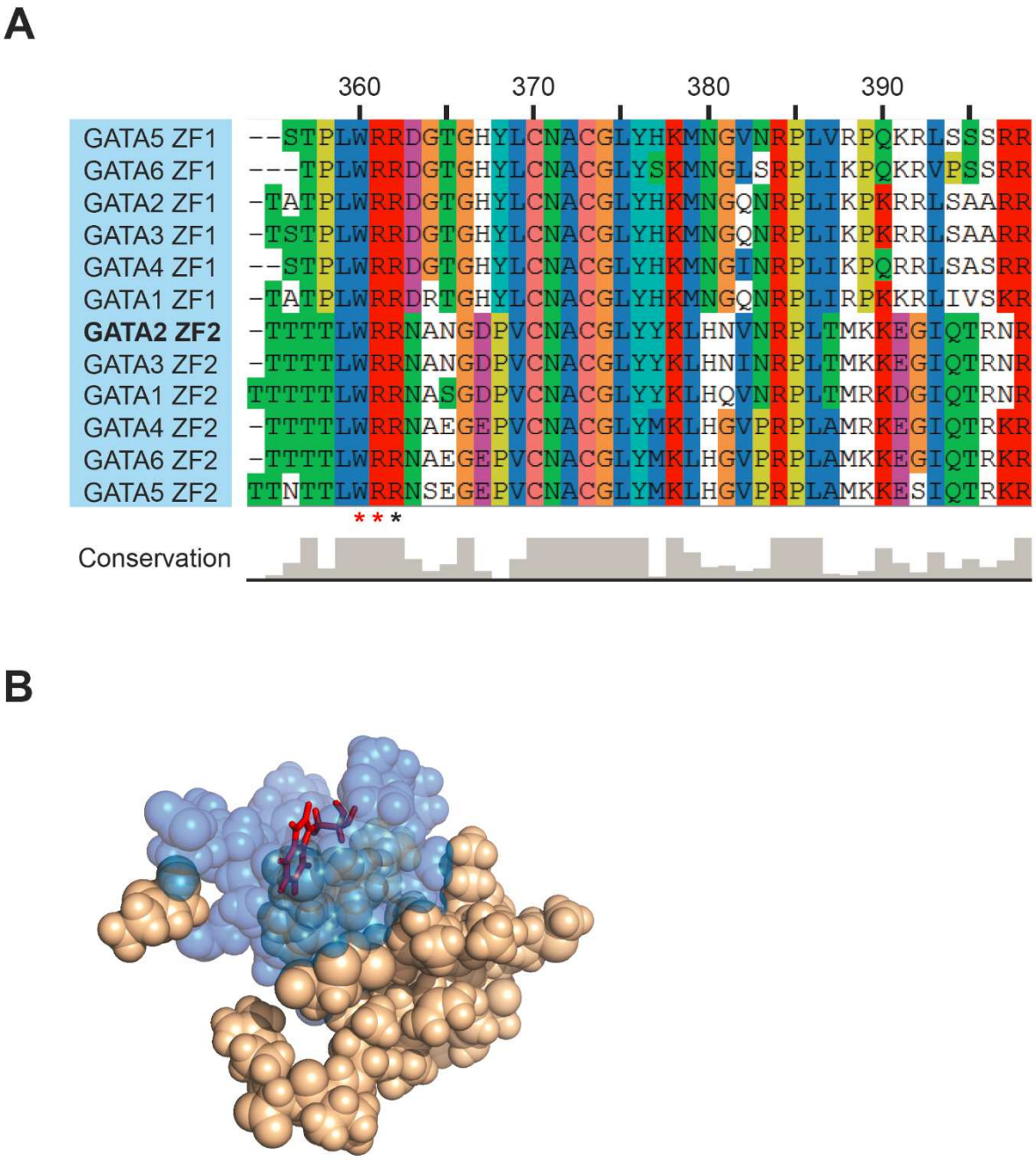


Figure A.1. Evolutionary and structural analysis of mutations in *GATA2*. (A) *GATA2* mutations identified in this study disrupt residues within an LWRR motif conserved across both zinc finger domains of the human GATA family. Missense and nonsense mutations reported in this study are indicated by red and black asterisks, respectively. Amino acid numbers correspond to the second zinc finger of GATA2. (B) Solution structure of murine GATA1 ZF1 (PDB ID: 1GNF).[131] The residue orthologous to human GATA2 Trp360 (highlighted in red with surrounding residues rendered semi-transparent) is largely buried within the GATA zinc finger.

A.3 Supplementary Tables

Table A.1. Known mutations used for blinded validation of MarrowSeq

Mutation class	Study Code	Diagnosis/Phenotype	Gene	Zygosity	Mutation[a]	Effect
Copy number variant	FH-142	Thrombocytopenia	*RUNX1*	Heterozygous	Whole gene deletion	Haploinsufficiency
	FH-16	Diamond-Blackfan anemia	*RPS17*	Heterozygous	Whole gene deletion	Haploinsufficiency
Insertion-deletion	FH-6	Dyskeratosis congenita	*DKC1*	Heterozygous	c.1512_1514dupGAA	p.K505dup
	FH-11	Diamond-Blackfan anemia	*RPL5*	Heterozygous	c.169_172delAACA	p.N57Qfs*12
Single nucleotide variant – exonic	FH-28	Diamond-Blackfan anemia	*RPS19*	Heterozygous	c.301C>T	p.R101C
	FH-43	Dyskeratosis congenita (AD)	*TERT*	Heterozygous	c.2266C>T	p.R756C
	FH-97	Aplastic anemia	*GATA2*	Heterozygous	c.1061C>T	p.T354M
	FH-130	Dyskeratosis congenita	*TERT*	Compound heterozygous	c.1760T>C	p.I587T
					c.2110C>T	p.P704S
	FH-138	Familial platelet disorder with predisposition to myeloid malignancy	*RUNX1*	Heterozygous	c.861C>A	p.Y287*
	FH-240	Pre-B cell ALL predisposition	*PAX5*	Heterozygous	c.547G>A	p.G183S
	FH-158	Shwachman-Diamond syndrome	*SBDS*	Compound heterozygous	c.183_184TA>CT	p.K62*
					c.258+2T>C	p.C84Yfs*4
Single-nucleotide variant – intronic splice site	FH-30	Shwachman-Diamond syndrome	*SBDS*	Homozygous	c.258+2T>C	p.C84Yfs*4
	CH-130	Shwachman-Diamond syndrome	*SBDS*	Compound heterozygous	c.258+2T>C	p.C84Yfs*4
					c.653G>A	p.R218Q
Single nucleotide variant – promoter	FH-31	Dyskeratosis congenita	*DKC1*	Hemizygous	c.-141C>G	Promoter mutation

[a]RefSeq IDs: *RUNX1* NM_001754.4; *RPS17* NM_001021.3; *DKC1* NM_001363.4; *RPL5* NM_000969.3; *RPS19* NM_001022.3; *TERT* NM_198253.2; *GATA2* NM_032638.4; *PAX5* NM_016734.2; *SBDS* NM_016038.2

Asterisk denotes a protein truncation resulting from the gene mutation.

Table A.2. Damaging mutations identified in patients with Fanconi Anemia

Study code	Clinical complementation subtype	Gene	Mutation[a]	Effect
FH-9	FA-A	*FANCA*	c.793_3C>G	p.T266Sfs*17
			c.1741dupT	p.S581Ffs*18
FH-73	FA-A	*FANCA*	c.1360_2014dup	p.V672Gfs*31
			c.2555C>A	p.S852*
FH-124	FA-A	*FANCA*	c.1462_1535del	p.L927Afs*10
			c.3338A>T, c.3316G>A	p.N1113I, p.E1106K
FH-241	FA-A	*FANCA*	c.1535C>G	p.S512*
			c.2853-15_2856del	Splice site mutation
FH-42	Non-A,C,G	*FANCD2*	c.2048T>C	p.L683P
			c.2715+1G>A	p.E906Ifs*4
FH-3	Non-A,C,G	*FANCA*	c.652T>C	p.C218R
			c.793_2014del	p.V265Lfs*8

[a]RefSeq IDs: *FANCA* NM_000135.2; *FANCD2* NM_033084.3
Asterisk denotes a protein truncation resulting from the gene mutation.

Table A.3. Clinical features of patients with idiopathic BMF/MDS

Study code	Gender	Age (years)	Diagnosis	Family history	Congenital anomalies	Short telomeres*	Causal gene identified by MarrowSeq
CH-103	M	10	Marrow Failure, MDS	-	+	NA	*LIG4*
CH-110	M	10	Marrow Failure	-	-	NA	-
CH-119	M	12	Marrow Failure, MDS	+	-	NA	*GATA2*
CH-140	M	10	Marrow Failure	-	-	NA	-
CH-144	M	1	Marrow Failure	-	-	NA	-
FH-8	M	7	Marrow Failure	-	-	-	-
FH-10	M	8	Marrow Failure	-	+	-	-
FH-13	F	1	Marrow Failure, MDS	-	+	+	-
FH-14	F	21	Marrow Failure	+	+	+	-
FH-15	F	9	Marrow Failure	-	+	NA	-
FH-20	M	1	Marrow Failure	+	-	-	-
FH-21	F	2	Marrow Failure	-	+	+	-
FH-22	M	3	Marrow failure	-	-	NA	-
FH-23	F	17	Marrow Failure	-	-	NA	-
FH-24	F	1	Marrow Failure	+	+	NA	-
FH-27	M	14	Marrow Failure	+	+	-	-
FH-45	M	2	Marrow Failure	-	-	-	-
FH-50	M	12	Marrow Failure	+	-	+	-
FH-53	F	24	Marrow Failure	Unknown (Adopted)	-	-	-
FH-54	M	10	Marrow Failure	-	+	NA	-
FH-55	F	7	Marrow Failure	+	-	-	-
FH-58	M	18	Marrow Failure	+	+	-	-
FH-60	F	1	Marrow Failure	+	-	+	-
FH-64	M	4	Marrow Failure	+	+	NA	-
FH-65	M	3	Marrow Failure	-	-	-	-
FH-66	M	4	Marrow Failure	+	-	-	-
FH-67	F	7	MDS	+	-	-	-
FH-68	M	9	Marrow Failure	Unknown (adopted)	-	+	-
FH-69	F	16	Marrow Failure	-	-	NA	-
FH-70	M	15	Marrow Failure	+	+	-	*DKC1*
FH-75	F	9	Marrow Failure	-	-	-	-
FH-78	F	13	MDS, RAEB	+	-	-	-
FH-79	F	57	Marrow Failure	+	+	-	-
FH-82	F	16	Marrow Failure	-	+	-	*GATA2*
FH-93	F	16	Marrow Failure	-	-	+	-

FH-98	M	23	Marrow Failure with trisomy 8	-	-	+	-
FH-105	F	31	Marrow Failure	-	-	-	-
FH-106	M	5	Marrow Failure	+	-	NA	-
FH-109	M	23	Marrow Failure	-	-	NA	-
FH-113	F	64	Marrow Failure	+	-	NA	-
FH-120	F	14	Marrow Failure	+	-	-	-
FH-121	F	17	Marrow Failure	-	+	-	-
FH-125	M	8	Marrow Failure	-	+	-	-
FH-126	M	11	Marrow Failure	+	+	-	-
FH-129	M	14	Marrow Failure	-	+	+	-
FH-134	M	4	Marrow Failure	-	+	NA	-
FH-141	M	10	Marrow Failure	-	-	NA	-
FH-145	F	1	Marrow Failure	+	+	-	-
FH-147	M	37	Marrow Failure	-	-	-	-
FH-149	F	3	Marrow Failure	+	+	-	-
FH-153	M	17	Marrow Failure	+	+	NA	-
FH-154	F	17	Marrow Failure	-	+	-	*GATA2*
FH-156	F	2	MDS	-	-	NA	-
FH-159	F	5	Marrow Failure	+	+	-	-
FH-164	F	11	Marrow Failure	-	-	-	-
FH-165	F	7	Marrow Failure	-	-	-	-
FH-166	M	2	Marrow Failure	+	+	-	-
FH-169	M	1	Marrow Failure	-	-	-	-
FH-178	M	12	MDS, 5q-	Unknown (adopted)	+	-	*RUNX1*
FH-179	F	18	Marrow Failure	+	-	-	-
FH-180	F	2	Marrow Failure	-	+	NA	-
FH-181	F	22	Marrow Failure	+	-	NA	*GATA2*
FH-182	M	67	Marrow Failure	+	-	-	-
FH-184	M	6	Marrow Failure	+	+	NA	-
FH-185	M	8	Marrow Failure	+	+	-	-
FH-186	F	6	Marrow Failure	-	+	NA	-
FH-187	F	20	MDS	Unknown	-	-	-
FH-189	F	32	Marrow Failure	-	-	-	-
FH-190	M	24	Marrow Failure	+	-	+	-
FH-202	F	12	Marrow Failure	+	-	-	*GATA2*
UW-8	M	18	Marrow Failure	+	-	+	-

F, female; M; male; MDS, myelodysplastic syndrome; NA, not available

Table A.4. Damaging mutations and variants of unknown clinical significance (VUS) at highly conserved sites in patients with previously unclassified BMF/MDS

Variant type[a]	Study code	Sex[b]	DNA source[c]	Gene	RefSeq Transcript	Mutation	Effect	Total reads	Variant reads
Damaging	CH-103	M	MF	*LIG4*	NM_002312.3	c.2440C>T[d]	p.R814*	523	247
Damaging	CH-103	M	MF	*LIG4*	NM_002312.3	c.1751_1755delTAAGA[d]	p.I584Rfs*2	541	244
VUS	CH-119	M	MF	*KIT*	NM_000222.2	c.1889A>G	p.H630R	394	176
VUS	CH-119	M	MF	*SRP72*	NM_006947.3	c.1448C>A	p.T483N	342	177
VUS	CH-119	M	MF	*BCR*	NM_004327.3	c.1889C>G	p.A630G	491	222
Damaging	CH-119	M	MF	*GATA2*	NM_032638.4	c.1078T>A	p.W360R	317	153
VUS	FH-68	M	PB	*TINF2*	NM_012461.2	c.734C>A[e]	p.S245Y	498	498
Damaging	FH-70	M	MF	*DKC1*	NM_001363.4	c.-141C>G	Promoter mutation	92	92
Damaging	FH-82	F	LCL, MF	*GATA2*	NM_032638.4	c.1084C>T	p.R362*	354	182
VUS	FH-105	F	PB	*ETV6*	NM_001987.4	c.380G>A	p.R127Q	722	307
VUS	FH-134	M	LCL	*DNMT3A*	NM_022552.4	c.347C>G	p.A116G	299	129
VUS	FH-147	M	PB	*KMT2A*	NM_001197104.1	c.11770G>T	p.D3924Y	647	291
VUS	FH-147	M	PB	*U2AF1*	NM_001025203.1	c.470A>G	p.Q157R	607	261
VUS	FH-147	M	PB	*U2AF1*	NM_001025203.1	c.101C>T	p.S34F	648	283
VUS	FH-149	F	PB	*GFI1*	NM_005263.3	c.49C>G	p.Q17E	532	266
VUS	FH-154	F	BM	*ANKRD26*	NM_014915.2	c.3899A>G	p.K1300R	195	81
VUS	FH-154	F	BM	*CBL*	NM_005188.3	c.522T>G	p.F174L	542	269
Damaging	FH-154	F	BM, MF	*GATA2*	NM_032638.4	c.1084C>T	p.R362*	361	178
VUS	FH-165	F	PB	*CBL*	NM_005188.3	c.12C>A	p.N4K	219	96
VUS	FH-166	M	PB	*PRPF40B*	NM_001031698.2	c.1649G>A	p.R550H	407	189
VUS	FH-169	M	BM	*BCOR*	NM_001123385.1	c.3308A>C	p.E1103A	167	167
Damaging	FH-178	M	PB, MF	*RUNX1*	NM_001754.4	c.567C>G	p.Y189*	547	286
VUS	FH-181	F	PB	*RARA*	NM_000964.3	c.743G>A	p.G248D	380	190
Damaging	FH-181	F	PB	*GATA2*	NM_032638.4	c.988C>T	p.R330*	195	79
VUS	FH-186	M	PB	*KMT2A*	NM_001197104.1	c.9400C>T	p.L3134F	659	306
Damaging	FH-202	F	PB, MF	*GATA2*	NM_032638.4	c.1082G>A	p.R361H	130	62

[a]VUS, variant of unknown significance
[b]F, female; M, male;
[c]BM, bone marrow; LCL, lymphoblast cell line; MF, marrow fibroblasts; PB, peripheral blood
[d]These two mutations were experimentally determined to be in *trans*.
[e]A patient with aplastic anemia has previously been reported to harbor this variant as a heterozygote.[132] This variant is present with an allele frequency of 0.05% in 1000Genomes and 0.0326% in the Exome Variant Server. The pathogenicity of this variant as a homozygote is unknown.
Asterisk denotes a protein truncation resulting from the gene mutation.

Table A.5. Damaging carrier mutations and carrier variants of unknown clinical significance at highly conserved sites in patients with previously unclassified BMF/MDS

Variant type[a]	Study code	Sex[b]	DNA source[c]	Gene	RefSeq Transcript	Mutation	Effect	Total reads	Variant reads
Damaging	FH-22	M	LCL	*FANCL*	NM_018062.3	c.1096_1099dupATTA	p.T367Nfs*13	574	241
VUS	FH-50	M	MF	*FANCL*	NM_018062.3	c.1007_1009delTAT	p.I336_C337delinsS	611	286
VUS	FH-55	F	LCL	*FANCM*	NM_020937.2	c.4931G>A	p.R1644Q	463	232
VUS	FH-65	M	MF	*FANCI*	NM_001113378.1	c.1813C>T	p.L605F	466	238
Damaging	FH-66	M	MF	*SBDS*	NM_016038.2	c.258+2T>C	p.C84fs*3	742	304
Damaging	FH-69	F	MF	*AK2*	NM_001625.3	c.219+5G>A	Splice site mutation	456	195
Damaging	FH-98	M	LCL	*MPL*	NM_005373.2	c.744_747dupTGGC	p.N250Wfs*13	382	151
Damaging	FH-113	F	BM	*PALB2*	NM_024675.3	c.1163dupC	p.L389Sfs*12	489	236
Damaging	FH-121	F	BM	*CTC1*	NM_025099.5	c.2452C>T	p.R818*	396	193
VUS	FH-154	F	BM	*FANCI*	NM_001113378.1	c.1813C>T	p.L605F	410	187
VUS	FH-156	F	BM	*FANCI*	NM_001113378.1	c.1264G>A	p.G422R	718	340
VUS	FH-180	F	PB	*AK2*	NM_001625.3	c.224G>T	p.S75I	459	222
VUS	FH-186	M	PB	*AK2*	NM_001625.3	c.49C>G	p.R17G	122	63
VUS	UW-8	M	LCL	*NBN*	NM_002485.4	c.706A>G	p.K236E	223	101

[a]VUS, variant of unknown significance
[b]M, male; F, female
[c]BM, bone marrow; LCL, lymphoblast cell line; MF, marrow fibroblasts; PB, peripheral blood
Asterisk denotes a protein truncation resulting from the gene mutation.

Appendix B – Supplemental Materials for Chapter 3

B.1 Supplementary Note

Clinical summaries

Family A

III-2 was born at term with a birth weight of 3.2 kilograms. Pregnancy was complicated by maternal thrombocytopenia and hyperemesis gravidarum. There was a maternal history of five miscarriages. The proband developed bruising in infancy and was diagnosed with chronic thrombocytopenia with platelet counts around 50,000. Her thrombocytopenia worsened in her late teenage years. At age 19, she developed severe menorrhagia requiring menstrual suppression with oral contraceptives. She developed progressive anemia and intermittent mild neutropenia. Evaluation for hemoglobinopathies and iron deficiency were negative. Her cytopenias persisted, with poor response to IVIG therapy. She intermittently required platelet transfusions prior to surgeries.

A bone marrow exam at age 9 showed mild marrow hypocellularity of 60% and multilineage dysplasia with many small hypolobated megakaryocytes (micromegakaryocytes) and hypogranular neutrophils with abnormal nuclear segmentation. The erythroid lineage was expanded and left shifted with occasional megaloblastic features without dyserythropoiesis. Cytogenetic studies demonstrated a normal female karyotype. Serial marrow exams were stable until age 15, when a translocation t(X;5)(p11.2;q22) was noted in 2 of 20 metaphases (10%). This clonal population progressively increased to 5 of 20 metaphases (25%) at age 16, and 8 of 20 metaphases (40%) by age 17, together with progressive multilineage dysplasia and 1-2% blasts. This clone became undetectable on subsequent bone marrow exams. FISH studies for -5/5q-, -7/7q-, +8, 20q-, *AML1/ETO*, *PML/RARA*, and *CBFβ/MYH11* were negative. Bone marrow exam at age 22 showed multilineage dysplasia with an expanded population of myeloid

blasts representing 7% of the marrow cellularity by morphology consistent with a diagnosis of refractory anemia with excess blasts-1 (RAEB-1). She had adequate numbers of erythroid precursors with normal maturation. Myeloid precursors were decreased in number with left-shifted maturation, and a subset of myeloid cells showed hypogranulation and hypolobation (Supplementary Fig. 5a). Megakaryocyte dysplasia was present with numerous atypical forms including many micromegakaryocytes and a subset with splayed nuclei (Supplementary Fig. 5c). The iron stain demonstrated no ring sideroblasts.

The patient underwent hematopoietic stem cell transplantation from an HLA-matched unrelated donor at age 21 years and 11 months. The conditioning regimen consisted of intravenous (IV) treosulfan 14g/m^2/day on days -6 to -4, fludarabine 30 mg/m^2/day IV on days -6 to -2, and 2 Gy TBI on day 0, followed by infusion of peripheral blood stem cells from an unrelated donor. Graft-versus-host disease (GVHD) prophylaxis consisted of tacrolimus and methotrexate, as described previously[1]. Neutrophil engraftment occurred on day +26. She developed engraftment syndrome/hyperacute GVHD with grade 3 skin involvement on day +21, and was treated with high-dose steroids and psoralen plus ultraviolet A. She was tapered off all immunosuppression by day +240 and remains in remission, without evidence of GVHD at last follow up two years post-transplant.

The mother (II-5), sister (III-1), and brother (III-3) of the proband also exhibited thrombocytopenia (Supplementary Table 3). The mother's platelet counts ranged between 40,000-80,000/μL. Mother developed stage III colorectal adenocarcinoma at age 45 and multiple myeloma at age 51. Details of the mother's treatment are not available. The proband's sister was diagnosed with pre-B cell acute lymphoblastic leukemia (ALL) at age 7.5. She was treated with chemotherapy under the Pediatric Oncology Group (POG) protocol 9201 but medical reports are not available. She remains in remission 20 years after her diagnosis, but

she has persistent mild thrombocytopenia with platelet counts around 120,000/μL. The proband's brother has platelet counts in the range of 60,000-120,000/μL and anemia. The brother also has an undefined myopathy and gastrointestinal symptoms attributed to inflammatory bowel disease.

Sanger sequencing of *GATA2*, *CEBPα*, *PTEN*, and *TP53* yielded no damaging germline mutations. Peripheral blood DNA from the proband at age 20 years showed heterozygosity for *RUNX1* c.424dupG, predicted to result in premature protein termination (p.Ala142Glyfs*2). This mutation was not present in skin fibroblasts from the proband and was also absent from her parents and siblings, indicating an acquired event.

Family B

III-8 was first evaluated at age 17, due to recurrent nosebleeds about twice monthly lasting up to 40 minutes in duration. Laboratory evaluation at that time revealed isolated thrombocytopenia with a platelet count of 135,000/μL. His platelets were normal in size. Platelet aggregation studies showed decreased aggregation with arachidonic acid, collagen, and ADP as well as decreased thrombin and collagen release, but normal arachidonic acid and ADP release and normal ristocetin aggregation. Platelet electron microscopy showed normal platelet morphology. A strong family history of thrombocytopenia and bleeding in other individuals in the family was noted, including his grandfather (I-1), father (II-3), two aunts (II-1 and II-8), uncle (II-5), and cousins (III-9 and III-12). He was given a diagnosis of a likely congenital platelet disorder of unknown etiology.

Malignancies have been diagnosed in two individuals in this family. At age 82, I-1 was noted to have an elevated monocyte count (5700/μL) as well as peripheral blood blasts. Bone marrow biopsy was hypercellular with 85% cellularity with 16% blasts expressing CD13, CD33, CD34,

CD117, HLA-DR, and CD4 (subset), consistent with a diagnosis of chronic myelomonocytic leukemia-2 (CMML-2). FISH for *BCR/ABL1* and *FIP1L1-PDGFRA* fusion was negative. Dysplasia was noted in both erythroid and granulocytic precursors. Megakaryocytes were decreased. He declined therapy. II-8 was diagnosed with stage IV colon cancer at age 43.

Family C

II-1 presented with thrombocytopenia at the age of 43. She was treated for the provisional diagnosis of ITP with steroids, IVIG, and the thrombopoietin receptor agonist, eltrombopag, without a sustained response. She did well until the age of 50, when she developed exertional dyspnea and was found to have anemia (hemoglobin 7.7 g/dL) and thrombocytopenia (90,000/uL). A bone marrow biopsy revealed a hypercellular bone marrow with 80% cellularity with 60% blasts, which were positive for CD2 (subset), cytoplasmic CD3, CD25, CD33, CD117, HLA-DR, MPO, and TdT, consistent with a diagnosis of T-/myeloid mixed phenotype acute leukemia. Cytogenetic analysis showed a normal karyotype (46, XX[22]). FISH for recurrent genetic abnormalities in MDS and the common chromosomal fusions in AML was negative. She underwent induction chemotherapy with infusional cytarabine and daunorubicin. Induction was complicated by prolonged cytopenias. She remained transfusion dependent for both platelets and red blood cells at day 70 post-induction. A repeat bone marrow around day 60 following consolidation was hypocellular, with 10-15% cellularity, but with evidence of early regeneration of all three lineages and no residual leukemia. She underwent one cycle of high dose cytarabine consolidation three months after her initial induction, which she tolerated well, but she remained transfusion dependent for both platelets and red blood cells throughout this time. A repeat marrow again showed hypocellularity with 10% cellularity, with normal appearance of all three lineages and no residual leukemia. She then underwent an alternative donor allogeneic hematopoietic stem cell transplantation using a haploidentical sibling (II-2) and an umbilical cord blood unit. She received fludarabine 30 mg/m^2 IV on Day -7 through Day -3, melphalan 70

mg/m2 IV on Day -3 and Day -2, and rabbit anti-thymocyte globulin 1.5 mg/kg IV on Day -5, Day -3, and Day -1. She received tacrolimus and mycophenolate mofetil for GVHD prophylaxis. Neutrophil engraftment occurred on Day +11. Her post-transplant course was complicated by tacrolimus related microangiopathy and persistent anemia and thrombocytopenia, requiring occasional transfusions. She remains well two months post-transplant but awaits full recovery of her red blood cell and platelet counts.

The proband's sister (II-2), who served as her haploidentical stem cell donor, is healthy without a history of bleeding or thrombocytopenia. Blood counts prior to stem cell collection were normal, except for mild leukopenia (white blood cell count 2.3k/μL), consistent with her African-American descent. Her mother (I-2) was well until age 64 when she developed rectal bleeding. Colonoscopy revealed a rectal mass. She underwent resection which showed Stage II moderately differentiated adenocarcinoma. Peripheral blood counts at the time of diagnosis were normal. The proband's father died in his 20s due to an accident. Limited health information is known for the proband's father and the rest of his side of the family.

B.2 Supplementary Figures

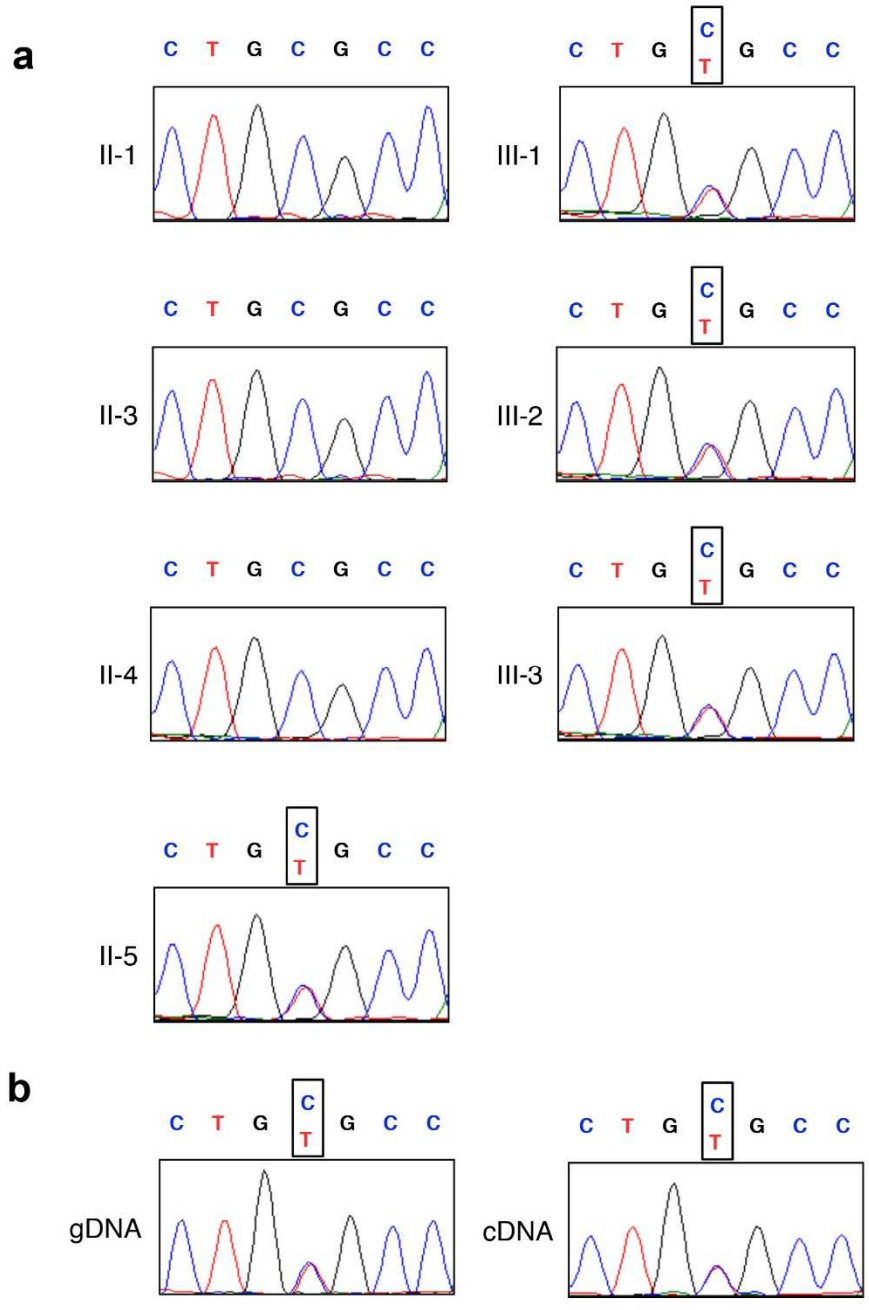


Figure B.1. Sanger sequencing of *ETV6* c.1195C>T mutation in Family A. (**a**) Electropherograms from Sanger sequencing of *ETV6* c.1195C>T in peripheral blood genomic DNA of Family A. (**b**) Sanger sequencing of *ETV6* c.1195C>T in genomic and complementary DNA derived from RAEB-1 MDS bone marrow sample of Family A III-2 shows no loss of heterozygosity.

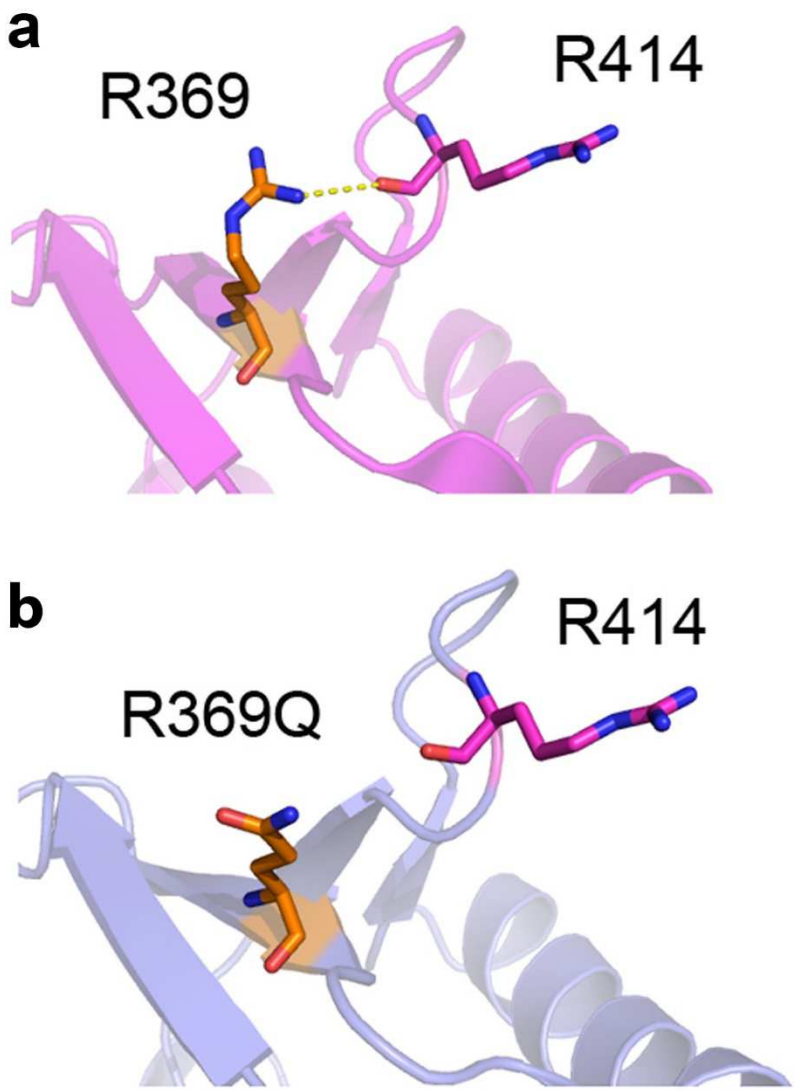


Figure B.2. ETV6 p.Arg369Gln mutation disrupts internal hydrogen bonding. (**a**) Hydrogen bonding (dotted lines) between the guanidinium nitrogen of Arg369 (orange) in β-sheet 2 with the backbone carbonyl oxygen of Arg414 (magenta) in the wing of the ETS domain. Protein structure of the murine Etv6 ETS domain (PDB ID: 4MHG)[2] is shown. The ETS domains of murine and human ETV6 have 100% amino acid sequence identity. (**b**) Molecular modeling of the Arg369Gln (orange) variant using SWISS-MODEL predicts loss of this hydrogen bonding interaction.

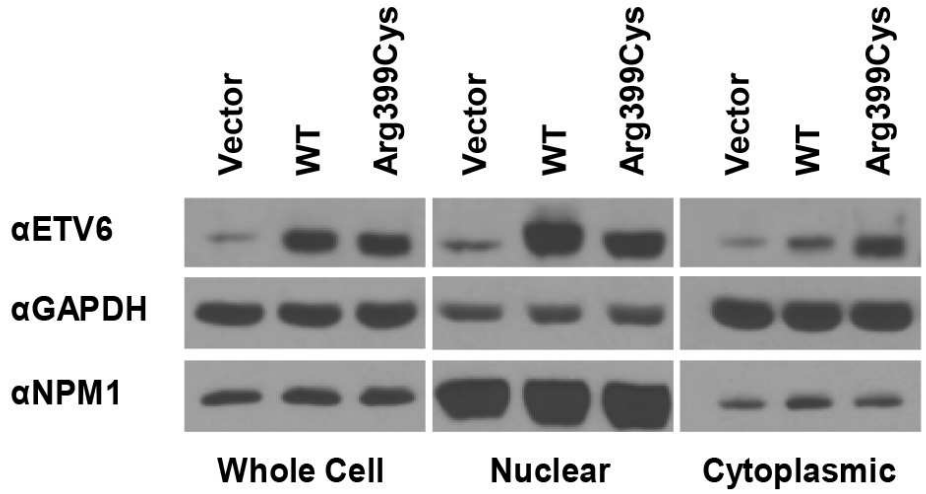


Figure B.3. Cell fractionation shows decreased nuclear localization of mutant ETV6. HeLa cells were transfected with empty vector, *ETV6* wild-type cDNA, or *ETV6* p.Arg399Cys cDNA. Lysates of nuclear versus cell fractions as well as whole cell lysates were analyzed by western blot for ETV6, GAPDH (cytoplasmic marker), and NPM1 (nuclear marker).

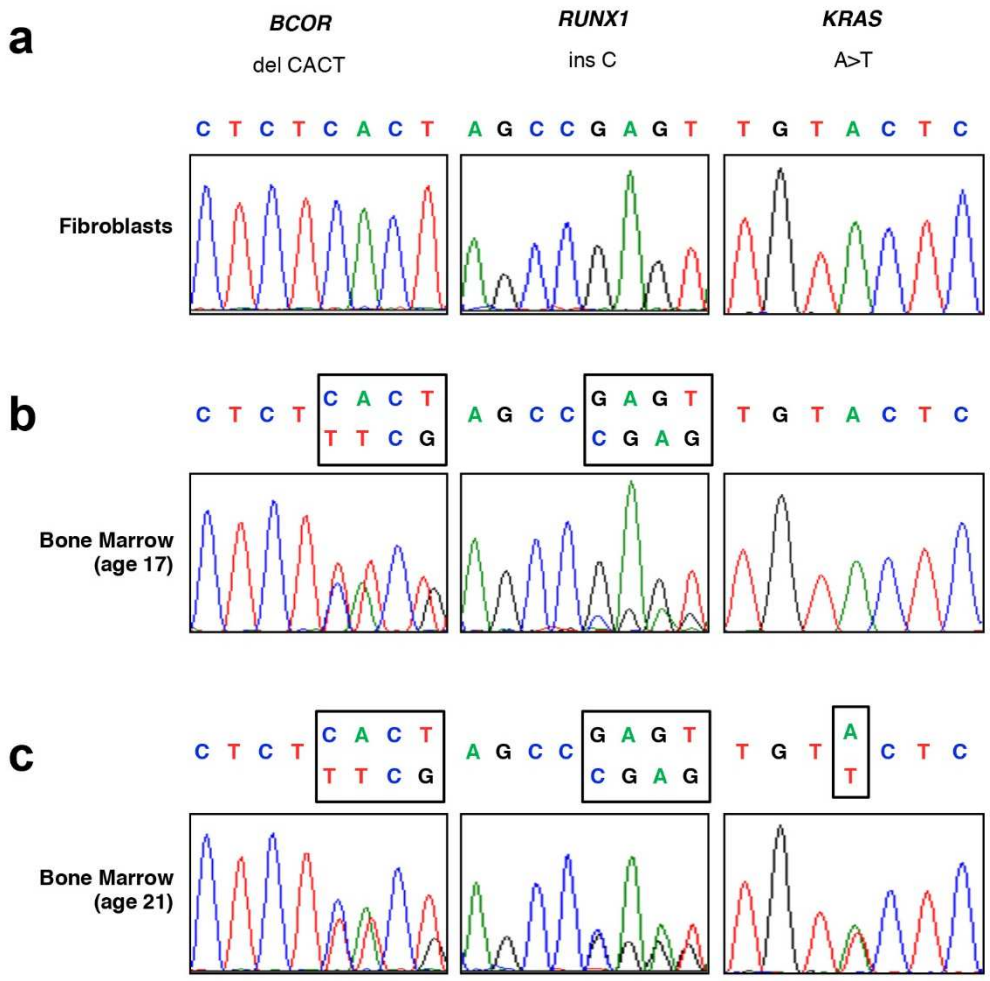


Figure B.4. Clonal evolution of MDS in the context of germline ETV6 mutation. Analysis of germline and serial bone marrow samples in Family A III-2. Electropherograms from Sanger sequencing of *BCOR*, *RUNX1*, and *KRAS* mutations in genomic DNA derived from (**a**) marrow fibroblasts, (**b**) bone marrow mononuclear cells at age 17 when the patient had refractory cytopenias with multilineage dysplasia (RCMD), and (**c**) bone marrow mononuclear cells at age 21 when the patient's disease progressed to refractory anemia with excess blasts type 1 (RAEB-1).

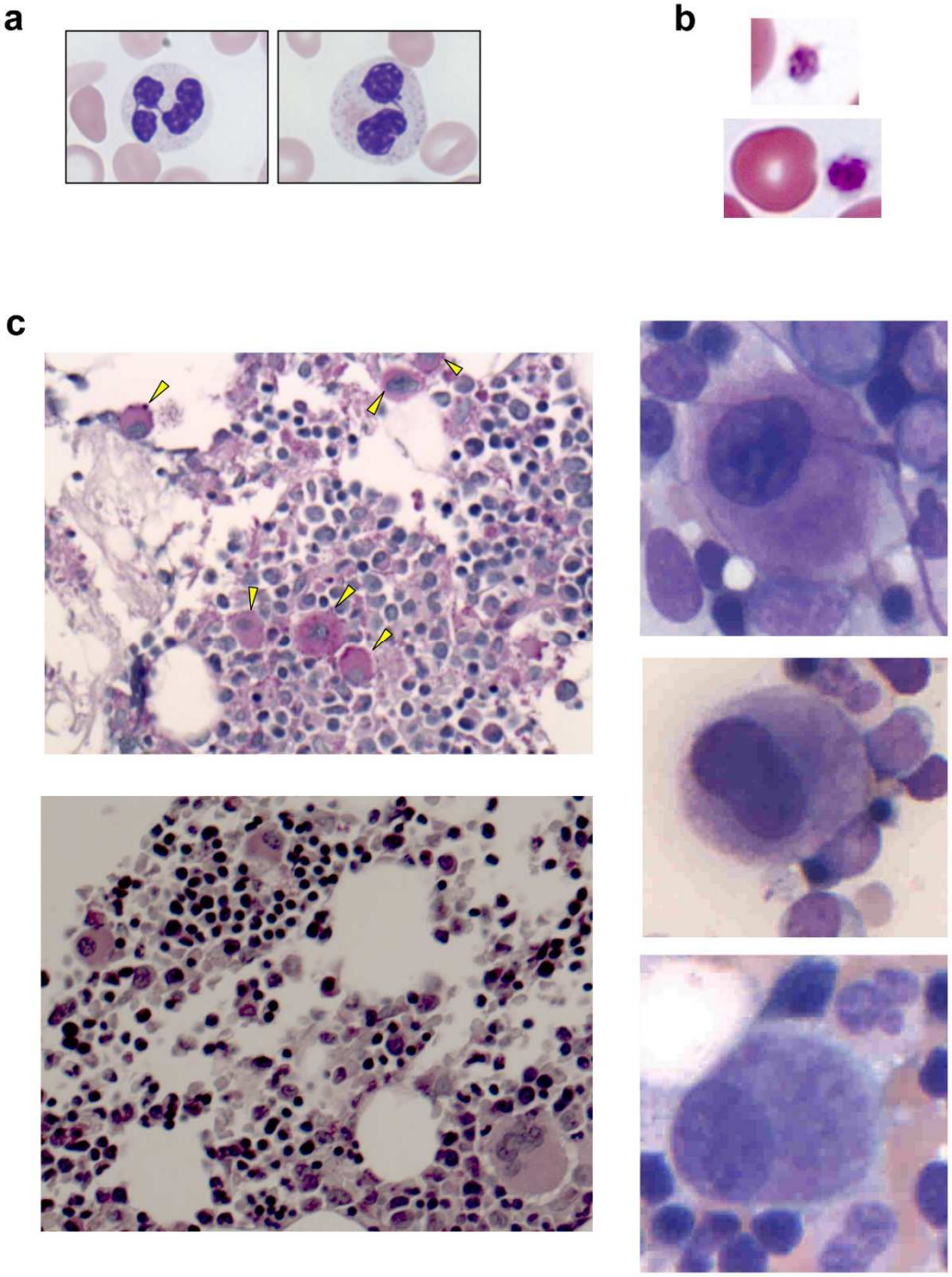


Figure B.5. Bone marrow morphology in III-2 of Family A. (**a**) Pre-transplant peripheral blood smears from patient III-2 illustrating hypogranulated neutrophils (left panel) and hypolobated neutrophils (pseudo Pelger-Huët cell) (right panel). (**b**) Platelets from a Wright-Giemsa stained peripheral blood smear. (**c**) Upper left panel: Wright-Giemsa stained particle preparation depicting small, hypolobated megakaryocytes (triangles). Lower left panel: Bone marrow biopsy depicting hypolobated micro-megakaryocytes. Right panels: Megakaryocyte dysplasia.

B.3 Supplementary Tables

Table B.1. Candidate protein-altering germline variants segregating with thrombocytopenia and malignancy in Family A.

Germline variants were derived from exome sequencing of II-4, II-5, III-1, III-2, and III-3 of Family A. Candidate variants were selected based on segregation with thrombocytopenia and malignancy under the assumption of an autosomal dominant mode of inheritance, alteration of protein sequence (nonsense, missense, splice site, indels), and evolutionary conservation of the altered site across species.

Gene	RefSeq	Mutation	Effect
ETV6	NM_001987.4	c.1195C>T	p.R399C
GPR144	NM_001161808.1	c.2365G>T	p.A789S
ITGA8	NM_003638.1	c.746T>C	p.L249P
PLEC	NM_201383.1	c.2248C>T	p.R750C
TOP3B	NM_003935.3	c.1492G>A	p.E498K

Table B.2. Gene panel for targeted capture followed by multiplexed high throughput sequencing.

ABCB7	*FANCE*	*MET*	*RPS26*
ABL1	*FANCF*	*MLL*	*RUNX1*
AK2	*FANCG*	*MPL*	*SBDS*
ANKRD26	*FANCI*	*NBN*	*SF1*
ASXL1	*FANCJ/BRIP1*	*NHP2*	*SF3A1*
ATR	*FANCL/PHF9*	*NOP10*	*SF3B1*
BCL2L11	*FANCM*	*NPM1*	*SRP72*
BCOR	*FANCN/PALB2*	*NRAS*	*SRSF2*
BCR	*FANCO/RAD51C*	*PAX5*	*TERC*
BRAF	*FANCP/SLX4*	*PML*	*TERT*
CBL	*FLT3*	*PRPF40B*	*TET2*
CEBPA	*G6PC3*	*RARA*	*TINF2*
CTC1	*GATA1*	*RMRP*	*TP53*
DKC1	*GATA2*	*RPL5*	*U2AF1*
DNMT3A	*GFI1*	*RPL10*	*U2AF2*
ELANE	*HAX1*	*RPL11*	*WAS*
ETV6	*IDH1*	*RPL35A*	*WRAP53*
FANCA	*IDH2*	*RPS7*	*WT1*
FANCB	*JAK2*	*RPS14*	*ZRSR2*
FANCC	*KIT*	*RPS17*	
FANCD1/BRCA2	*KRAS*	*RPS19*	
FANCD2	*LIG4*	*RPS24*	

Table B.3. Hematologic parameters and genotypes of individuals in the study families.

Individual	ETV6 Genotype*	Age	wbc ($x10^9$/l)	Rbc ($x10^{12}$/l)	Hgb (g/dl)	HCT (%)	PLT ($x10^9$/l)	Retic. (%)	MCV (fl)	MCH (pg)	MCHC (g/dl)	RDW (%)	MPV (fl)	ANC ($x10^9$/l)
A: II-4	WT	50 yr, 0 mo	7.1	4.77	14.9	42.8	213		89.7	31.3	34.9	12.8	9.9	5.9
A: II-5	R399C	35 yr, 6 mo	4.3	4.39	14.3	41.2	111		93.8	32.6	34.7	12.2	7.1	2.6
		43 yr, 10 mo	2.1	4.45	14.5	41.7	60		93.7	32.7	34.9	14.8	6.4	1.0
		45 yr, 8 mo	2.5	4.15	14.5	41.6	96	0.99	100.2	35.0	34.9	11.3	6.5	1.2
		46 yr, 3 mo	3.9	3.54	12.1	34.5	59		97.4	34.2	35.2	16.3	7.4	1.3
		47 yr, 6 mo	2.3	2.93	10.6	30.2	71		103.1	36.1	35	13.2	6.7	1.0
		49 yr, 1 mo	3.5	3.60	12.7	36.4	80		101.1	35.3	34.9	15.2	7.0	1.6
		51 yr, 2 mo	2.4	3.14	11.1	31.1	66		99.1	35.4	35.7	16.0	8.6	0.8
		51 yr, 5 mo	3.3	3.06	10.2	29.6	56	2.67	96.7	33.3	34.5	19.7	7.3	1.7
A: III-1	R399C	8 yr, 8 mo	1.3	2.27	7.6	21.4	72	2.90	94.3	34.3	35.5	22.0	6.7	
		9 yr, 5 mo	0.9	2.04	7.7	21.9	119		107.4	37.7	35.1	13.9	6.7	0.4
		10 yr, 2 mo	2.3	3.42	12.7	36.2	98		105.8	37.1	35.1	14.2		
		12 yr, 0 mo	3.9	4.36	14.3	41.4	103		95.1	32.7	34.4	12.8	6.2	2.0
		15 yr, 9 mo	4.3	4.32	13.2	37.6	117	2.60	87.2	30.5	35.0	15.5	7.3	2.3
A: III-2	R399C	6 yr, 0 mo	4.0	3.82	12.6	36.0	34	1.40	94.2	32.9	34.9	13.5	7.8	1.6
		7 yr, 9 mo	4.6	4.40	12.8	37.8	95		85.8	29.1	33.9	15.8		
		9 yr, 11 mo	4.3	3.61	12.2	35.3	57		97.7	33.7	34.5	12.6	8.8	1.6
		12 yr, 7 mo	4.7	3.75	11.5	34.4	47		91.7	30.7	33.4	16.3	8.2	1.8
		14 yr, 9 mo	2.7	3.76	10.7	32.4	39		86.0	28.4	33.1	19.0	9.7	1.0
		16 yr, 10 mo	2.2	3.08	9.1	27.2	43	2.04	88.4	29.6	33.4	20.4	9.3	0.6
		17 yr, 5 mo	4.8	3.07	9.1	27.8	10		90.6	29.6	32.7	18.9	10.8	
		21 yr, 8 mo	5.5	4.51	9.9	31.0	21		68.7	22.1	32.1	24.3	12.6	
		21 yr, 11 mo	2.6	3.59	8.0	26.2	17		73.0	22.3	30.5	25.0	12.6	2.4
		22 yr, 6 mo‡	3.4	3.40	12.1	35.3	141		103.8	35.6	34.3	13.4	7.0	2.3
		23 yr, 3 mo‡	5.1	4.09	13.1	39.8	201		97.4	32.1	32.9	13.4	7.4	3.5
A: III-3	R399C	3 yr, 11 mo	9.9	4.23	13.3	38.0	127	1.20	89.9	31.4	35.0	12.6		
		9 yr, 10 mo	4.2	4.25	11.3	33.1	141		77.9	26.6	34.1	21.0	8.4	1.9
		11 yr, 3 mo	3.3	4.05	10.5	31.3	112		77.2	26.0	33.7	20.0	8.7	1.6
		14 yr, 11 mo	3.2	4.74	12.6	37.7	128		79.5	26.6	33.5	19.2	8.5	1.5
		15 yr, 6 mo	2.4	4.65	12.0	35.4	136		76.1	25.8	33.9	23.4	9.5	1.1
		16 yr, 3 mo	1.8	4.23	12.0	35.5	102	1.85	84.0	28.4	33.8	16.8	7.3	0.3
B: I-1	No DNA available	82 yr, 1 mo	15.1	2.95	9.6	28.1	42	0.60	95.3	32.5	34.2	16.7	9.1	7.1
B: II-2	WT	51 yr, 1 mo	7.2	5.19	15.5	46.1	228		88.9	29.9	33.6	13.1	6.9	5.7
		59 yr, 4 mo	7.0	5.34	15.3	46.1	197		86.3	28.7	33.2	13.7	10.0	4.6
B: II-3	R369Q	48 yr, 0 mo	5.6	4.77	15.1	44.6	140		93.5	31.7	33.9	12.8		
		57 yr, 9 mo	8.1	5.27	16.7	48.4	125		91.8	31.7	34.5	14.8	8.4	6.1
B: II-5	R369Q	57 yr, 0 mo	6.1	5.36	17.1	49.4	76		92.2	31.9	34.6	13.3	6.3	4.2
B: II-8	Obligate carrier	48 yr, 0 mo¶	15.4	2.95	10.6	31.6	100		107.1	35.9	33.5	25.7	10.2	10.6
B: III-3	WT	30 yr, 3 mo	7.9	4.59	13.5	40.2	211		87.6	29.4	33.6	13.0	10.4	6.2
		32 yr, 3 mo	3.2	4.56	13.2	39.6	181		86.8	28.9	33.3	12.2	9.2	1.1
		32 yr, 6 mo	5.3	4.69	13.7	41.3	197		88.1	29.2	33.2	12.3	9.0	3.6

		33 yr, 2 mo	5.2	4.87	13.9	42.4	194	87.1	28.5	32.8	13.0	9.3	3.4
B: III-4	WT	31 yr, 6 mo	6.8	5.54	15.9	47.4	249	85.5	28.7	33.6	13.2	7.8	3.9
B: III-5	WT	24 yr, 5 mo	7.1	5.05	13.9	42.8	358	84.8	27.5	32.5	13.0	10.5	4.5
B: III-7	WT	38 yr, 0 mo	9.4	5.21	14.7	44.3	231	85.0	28.2	33.2	13.7	10.1	5.4
B: III-8	R369Q	17 yr, 3 mo	5.0	4.62	14.3	42.7	135	92.5	31.0	33.5	11.6	8.0	
		17 yr, 5 mo	5.0	5.00	15.9	46.0	93	91.0		35	12.0		
		32 yr, 11 mo	7.3	5.27	17.0	47.7	106	90.5	32.3	35.6	13.6	9.2	
		34 yr, 0 mo	5.6	5.09	15.9	46.7	108	91.7	31.2	34	13.9	8.1	
B: III-10	WT	29 yr, 5 mo§	16.2	4.80	15.6	43.0	174	90.0	32.5	36	12.0	9.1	15.1
B: III-11	WT	28 yr, 11 mo	6.8	5.40	16.0	47.0	223	87.0			12.1		
B: III-12	R369Q	20 yr, 8 mo¥	13.2	3.51	11.2	32.1	138	92.0	32.0	34.9	11.3	6.8	10.4
		20 yr, 8 mo¥	7.6	3.51	11.3	32.0	131	91.0	32.1	35.2	12.6	6.7	4.6
		20 yr, 9 mo©	12.5	3.26	10.2	29.9	133	92.0	31.5	34.3	12.2	6.0	10.4
		20 yr, 9 mo©	18.7	3.23	10.5	29.8	150	92.0	32.3	35	12.2	6.3	
C: I-2	WT	62 yr, 4 mo	5.5	4.34	11.9	35.9	353	82.7	27.4	33.1	13.7	9.4	2.6
C: II-1	P214L	45 yr, 11 mo	4.3	4.38	14.0	39.4	42	90.0	32.0	35.5	14.8		2.3
		51 yr, 0 mo	6.0		7.7	22.3	90	94.9	32.8		19.5		3.9
		51 yr, 3 mo	3.1	2.40	6.9	19.0	26	79.2	28.8	36.3	12.6	8.4	1.5
		51 yr, 6 mo	4.0	2.86	7.9	22.3	12	78.0	27.6	35.4	12.5	8.9	2.8
C: II-2	WT	50 yr, 2 mo	3.2	4.73	14.3	42.5	229	89.6	30.2	33.7	13.0	11.3	1.9

ANC, absolute neutrophil count; HCT, hematocrit; Hgb, hemoglobin; MCH, mean corpuscular hemoglobin; MCHC, mean corpuscular hemoglobin concentration; MCV, mean corpuscular volume; MPV, mean platelet volume; Plt, platelet count; Rbc, red blood cell count; RDW, red cell distribution width; Retic., reticulocyte count; wbc, white blood cell count.
*WT, wild-type; P214L, R369Q, and R399C represent heterozygosity for the variant allele.
‡Post-hematopoietic cell transplant
¶Sample taken during metastatic colon cancer
§Sample taken during viral illness
¥Sample taken during splenic rupture
©Sample taken during rebleed of spleen

Table B.4. Genes downregulated in mutant *ETV6* compared to WT *ETV6*

Symbol	Description	Chr	Map location	Type of gene
EPHA1-AS1	EPHA1 antisense RNA 1	7	7q35	miscRNA
ESPN	espin	1	1p36.31	protein-coding
PDGFA	platelet-derived growth factor alpha polypeptide	7	7p22	protein-coding
PPP2R2C	protein phosphatase 2, regulatory subunit B, gamma	4	4p16.1	protein-coding
TRAF1	TNF receptor-associated factor 1	9	9q33-q34	protein-coding
ABTB1	ankyrin repeat and BTB (POZ) domain containing 1	3	3q21	protein-coding
AIM2	absent in melanoma 2	1	1q22	protein-coding
AXL	AXL receptor tyrosine kinase	19	19q13.1	protein-coding
BEST1	bestrophin 1	11	11q13	protein-coding
CD226	CD226 molecule	18	18q22.3	protein-coding
CD44	CD44 molecule (Indian blood group)	11	11p13	protein-coding
COL1A2	collagen, type I, alpha 2	7	7q22.1	protein-coding
CTSE	cathepsin E	1	1q31	protein-coding
DNAH12	dynein, axonemal, heavy chain 12	3	3p14.3	protein-coding
DNALI1	dynein, axonemal, light intermediate chain 1	1	1p35.1	protein-coding
EGR1	early growth response 1	5	5q31.1	protein-coding
EGR3	early growth response 3	8	8p23-p21	protein-coding
EPHB6	EPH receptor B6	7	7q33-q35	protein-coding
GPR183	G protein-coupled receptor 183	13	13q32.3	protein-coding
IL20RB	interleukin 20 receptor beta	3	3q22.3	protein-coding
LOC100128593	uncharacterized LOC100128593	9	9q34.3	miscRNA
MAGEA4	melanoma antigen family A, 4	X	Xq28	protein-coding
MDGA1	MAM domain containing glycosylphosphatidylinositol anchor 1	6	6p21	protein-coding
MLC1	megalencephalic leukoencephalopathy with subcortical cysts 1	22	22q13.33	protein-coding
NFASC	neurofascin	1	1q32.1	protein-coding
NOX5	NADPH oxidase, EF-hand calcium binding domain 5	15	15q23	protein-coding
OR2AT4	olfactory receptor, family 2, subfamily AT, member 4	11	11q13.4	protein-coding
OSBPL10	oxysterol binding protein-like 10	3	3p22.3	protein-coding
PAEP	progestagen-associated endometrial	9	9q34	protein-coding

	protein			
PAH	phenylalanine hydroxylase	12	12q22-q24.2	protein-coding
PDCD1	programmed cell death 1	2	2q37.3	protein-coding
PDGFB	platelet-derived growth factor beta polypeptide	22	22q13.1	protein-coding
PPL	periplakin	16	16p13.3	protein-coding
RASSF6	Ras association (RalGDS/AF-6) domain family member 6	4	4q13.3	protein-coding
S100A16	S100 calcium binding protein A16	1	1q21	protein-coding
SDC2	syndecan 2	8	8q22-q23	protein-coding
SLC4A1	solute carrier family 4 (anion exchanger), member 1	17	17q21.31	protein-coding
STC1	stanniocalcin 1	8	8p21-p11.2	protein-coding
SYTL5	synaptotagmin-like 5	X	Xp21.1	protein-coding
TANC1	tetratricopeptide repeat, ankyrin repeat and coiled-coil containing 1	2	2q24.2	protein-coding
THBS1	thrombospondin 1	15	15q15	protein-coding
TRPV6	transient receptor potential cation channel, subfamily V, member 6	7	7q34	protein-coding
VEPH1	ventricular zone expressed PH domain-containing 1	3	3q24-q25	protein-coding
YPEL4	yippee-like 4 (Drosophila)	11	11q12.1	protein-coding
SNX21	sorting nexin family member 21	20	20q13.12	protein-coding
SPAG4	sperm associated antigen 4	20	20q11.21	protein-coding
SPRED3	sprouty-related, EVH1 domain containing 3	19	19q13	protein-coding
SSR4P1	signal sequence receptor, delta pseudogene 1	21	21q22.3	pseudo
TAL2	T-cell acute lymphocytic leukemia 2	9	9q32	protein-coding
TENM1	teneurin transmembrane protein 1	X	Xq25	protein-coding
TMSB10	thymosin beta 10	2	2p11.2	protein-coding
TNFRSF9	tumor necrosis factor receptor superfamily, member 9	1	1p36	protein-coding
TRPV3	transient receptor potential cation channel, subfamily V, member 3	17	17p13.2	protein-coding
TSPAN5	tetraspanin 5	4	4q23	protein-coding
WDR54	WD repeat domain 54	2	2p13.1	protein-coding
YPEL5	yippee-like 5 (Drosophila)	2	2p23.1	protein-coding
ZC2HC1C	zinc finger, C2HC-type containing 1C	14	14q24.3	protein-coding
ZNF778	zinc finger protein 778	16	16q24.3	protein-coding

ZNF862	zinc finger protein 862	7	7q36.1	protein-coding
ABCB1	ATP-binding cassette, sub-family B (MDR/TAP), member 1	7	7q21.12	protein-coding
ABCG2	ATP-binding cassette, sub-family G (WHITE), member 2	4	4q22	protein-coding
AKR1C1	aldo-keto reductase family 1, member C1	10	10p15-p14	protein-coding
ALDH3B1	aldehyde dehydrogenase 3 family, member B1	11	11q13	protein-coding
APLF	aprataxin and PNKP like factor	2	2p13.3	protein-coding
ARHGAP27	Rho GTPase activating protein 27	17	17q21.31	protein-coding
ASNS	asparagine synthetase (glutamine-hydrolyzing)	7	7q21.3	protein-coding
BMP8B	bone morphogenetic protein 8b	1	1p35-p32	protein-coding
C22orf23	chromosome 22 open reading frame 23	22	22q13.1	protein-coding
CD3D	CD3d molecule, delta (CD3-TCR complex)	11	11q23	protein-coding
CD80	CD80 molecule	3	3q13.3-q21	protein-coding
CDKN1A	cyclin-dependent kinase inhibitor 1A (p21, Cip1)	6	6p21.2	protein-coding
CDNF	cerebral dopamine neurotrophic factor	10	10p13	protein-coding
CHRD	chordin	3	3q27	protein-coding
CRYL1	crystallin, lambda 1	13	13q12.11	protein-coding
CSRNP1	cysteine-serine-rich nuclear protein 1	3	3p22	protein-coding
CTH	cystathionase (cystathionine gamma-lyase)	1	1p31.1	protein-coding
DGKG	diacylglycerol kinase, gamma 90kDa	3	3q27.2-q27.3	protein-coding
DUSP18	dual specificity phosphatase 18	22	22q12.2	protein-coding
ERRFI1	ERBB receptor feedback inhibitor 1	1	1p36	protein-coding
FAM114A1	family with sequence similarity 114, member A1	4	4p14	protein-coding
FAM214A	family with sequence similarity 214, member A	15	15q21.2-q21.3	protein-coding
FAM69A	family with sequence similarity 69, member A	1	1p22.1	protein-coding
FBXL16	F-box and leucine-rich repeat protein 16	16	16p13.3	protein-coding
FCGRT	Fc fragment of IgG, receptor, transporter, alpha	19	19q13.3	protein-coding
FRMD6	FERM domain containing 6	14	14q22.1	protein-coding
GATS	GATS, stromal antigen 3 opposite strand	7	7q22.1	protein-coding
GGCX	gamma-glutamyl carboxylase	2	2p12	protein-coding
GNB5	guanine nucleotide binding protein (G	15	15q21.2	protein-coding

	protein), beta 5			
GRK5	G protein-coupled receptor kinase 5	10	10q26.11	protein-coding
GSDMB	gasdermin B	17	17q12	protein-coding
H6PD	hexose-6-phosphate dehydrogenase (glucose 1-dehydrogenase)	1	1p36	protein-coding
HBA2	hemoglobin, alpha 2	16	16p13.3	protein-coding
HBZ	hemoglobin, zeta	16	16p13.3	protein-coding
IL6ST	interleukin 6 signal transducer (gp130, oncostatin M receptor)	5	5q11.2	protein-coding
ITGB5	integrin, beta 5	3	3q21.2	protein-coding
JHDM1D	jumonji C domain containing histone demethylase 1 homolog D (S. cerevisiae)	7	7q34	protein-coding
KIAA1107	KIAA1107	1	1p22.1	protein-coding
KIAA1324L	KIAA1324-like	7	7q21.12	protein-coding
KIAA1549	KIAA1549	7	7q34	protein-coding
KLF6	Kruppel-like factor 6	10	10p15	protein-coding
KLHL3	kelch-like family member 3	5	5q31	protein-coding
L3MBTL1	l(3)mbt-like 1 (Drosophila)	20	20q13.12	protein-coding
LGALS1	lectin, galactoside-binding, soluble, 1	22	22q13.1	protein-coding
LOC100133123	uncharacterized LOC100133123	X	Xp22.2	miscRNA
LOC339862	uncharacterized LOC339862	3	3p24.3	protein-coding
LPAR5	lysophosphatidic acid receptor 5	12	12p13.31	protein-coding
LYRM9	LYR motif containing 9	17	17q11.2	protein-coding
MAPK13	mitogen-activated protein kinase 13	6	6p21.31	protein-coding
MME	membrane metallo-endopeptidase	3	3q25.2	protein-coding
MRC2	mannose receptor, C type 2	17	17q23.2	protein-coding
MUC15	mucin 15, cell surface associated	11	11p14.3	protein-coding
MXD4	MAX dimerization protein 4	4	4p16.3	protein-coding
MYBPHL	myosin binding protein H-like	1	1p13.3	protein-coding
MYH14	myosin, heavy chain 14, non-muscle	19	19q13.33	protein-coding
MYL9	myosin, light chain 9, regulatory	20	20q11.23	protein-coding
NBPF14	neuroblastoma breakpoint family, member 14	1	1q21.1	protein-coding
OR51B2	olfactory receptor, family 51, subfamily B, member 2	11	11p15	protein-coding
OSBPL5	oxysterol binding protein-like 5	11	11p15.4	protein-coding
PAM	peptidylglycine alpha-amidating monooxygenase	5	5q14-q21	protein-coding

PARD6A	par-6 family cell polarity regulator alpha	16	16q22.1	protein-coding
PERP	PERP, TP53 apoptosis effector	6	6q24	protein-coding
PLEKHH1	pleckstrin homology domain containing, family H (with MyTH4 domain) member 1	14	14q24.1	protein-coding
PLEKHH2	pleckstrin homology domain containing, family H (with MyTH4 domain) member 2	2	2p21	protein-coding
PPFIBP2	PTPRF interacting protein, binding protein 2 (liprin beta 2)	11	11p15.4	protein-coding
PTPLA	protein tyrosine phosphatase-like (proline instead of catalytic arginine), member A	10	10p14-p13	protein-coding
RAB6B	RAB6B, member RAS oncogene family	3	3q22.1	protein-coding
RABL2B	RAB, member of RAS oncogene family-like 2B	22	22q13.33	protein-coding
REEP1	receptor accessory protein 1	2	2p11.2	protein-coding
RHBDF1	rhomboid 5 homolog 1 (Drosophila)	16	16p13.3	protein-coding
RHOB	ras homolog family member B	2	2p24	protein-coding
RNF181	ring finger protein 181	2	2p11.2	protein-coding
RSAD2	radical S-adenosyl methionine domain containing 2	2	2p25.2	protein-coding
SCRN1	secernin 1	7	7p14.3-p14.1	protein-coding
SEC14L6	SEC14-like 6 (S. cerevisiae)	22	22q12.2	protein-coding
SHISA4	shisa family member 4	1	1q32.1	protein-coding
SLC2A1-AS1	SLC2A1 antisense RNA 1	1	1p34.2	miscRNA
SNPH	syntaphilin	20	20p13	protein-coding
TAC3	tachykinin 3	12	12q13-q21	protein-coding
TBC1D19	TBC1 domain family, member 19	4	4p15.2	protein-coding
TCTN1	tectonic family member 1	12	12q24.11	protein-coding
TES	testis derived transcript (3 LIM domains)	7	7q31.2	protein-coding
TMEM150A	transmembrane protein 150A	2	2p11.2	protein-coding
TMEM233	transmembrane protein 233	12	12q24.23	protein-coding
TMEM45A	transmembrane protein 45A	3	3q12.2	protein-coding
TNFRSF12A	tumor necrosis factor receptor superfamily, member 12A	16	16p13.3	protein-coding
TPD52	tumor protein D52	8	8q21	protein-coding
ULK4	unc-51 like kinase 4	3	3p22.1	protein-coding
VENTX	VENT homeobox	10	10q26.3	protein-coding
VTN	vitronectin	17	17q11	protein-coding
YPEL3	yippee-like 3 (Drosophila)	16	16p11.2	protein-coding
ZNF396	zinc finger protein 396	18	18q12	protein-coding

ZSWIM5	zinc finger, SWIM-type containing 5	1	1p34.1	protein-coding
ZYX	zyxin	7	7q32	protein-coding
ADARB1	adenosine deaminase, RNA-specific, B1	21	21q22.3	protein-coding
AIM1	absent in melanoma 1	6	6q21	protein-coding
AKR1C2	aldo-keto reductase family 1, member C2	10	10p15-p14	protein-coding
AKR1C3	aldo-keto reductase family 1, member C3	10	10p15-p14	protein-coding
ALAS2	aminolevulinate, delta-, synthase 2	X	Xp11.21	protein-coding
ANKEF1	ankyrin repeat and EF-hand domain containing 1	20	20p12.2	protein-coding
ASIC1	acid-sensing (proton-gated) ion channel 1	12	12q12	protein-coding
C11orf21	chromosome 11 open reading frame 21	11	11p15.5	protein-coding
C5	complement component 5	9	9q33-q34	protein-coding
C7orf41	chromosome 7 open reading frame 41	7	7p14.3	protein-coding
CALCOCO1	calcium binding and coiled-coil domain 1	12	12q13.13	protein-coding
CBLB	Cbl proto-oncogene B, E3 ubiquitin protein ligase	3	3q13.11	protein-coding
CCT6B	chaperonin containing TCP1, subunit 6B (zeta 2)	17	17q12	protein-coding
CD9	CD9 molecule	12	12p13.3	protein-coding
CLIP2	CAP-GLY domain containing linker protein 2	7	7q11.23	protein-coding
COL6A1	collagen, type VI, alpha 1	21	21q22.3	protein-coding
COL6A3	collagen, type VI, alpha 3	2	2q37	protein-coding
CR2	complement component (3d/Epstein Barr virus) receptor 2	1	1q32	protein-coding
DDIT4	DNA-damage-inducible transcript 4	10	10q22.1	protein-coding
DENND2C	DENN/MADD domain containing 2C	1	1p13.2	protein-coding
DOPEY2	dopey family member 2	21	21q22.2	protein-coding
DUSP8	dual specificity phosphatase 8	11	11p15.5	protein-coding
ENC1	ectodermal-neural cortex 1 (with BTB domain)	5	5q13	protein-coding
EPB42	erythrocyte membrane protein band 4.2	15	15q15-q21	protein-coding
ETV4	ets variant 4	17	17q21	protein-coding
EVI5L	ecotropic viral integration site 5-like	19	19p13.2	protein-coding
FILIP1L	filamin A interacting protein 1-like	3	3q12.1	protein-coding
FMNL2	formin-like 2	2	2q23.3	protein-coding
FNDC4	fibronectin type III domain containing 4	2	2p23.3	protein-coding
GOLGA8S	golgin A8 family, member S	15	15q11.2	pseudo

GPC1	glypican 1	2	2q35-q37	protein-coding
GPR55	G protein-coupled receptor 55	2	2q37	protein-coding
HOMER3	homer homolog 3 (Drosophila)	19	19p13.11	protein-coding
IL4R	interleukin 4 receptor	16	16p12.1-p11.2	protein-coding
ITGB4	integrin, beta 4	17	17q25	protein-coding
KISS1R	KISS1 receptor	19	19p13.3	protein-coding
LGALS3	lectin, galactoside-binding, soluble, 3	14	14q22.3	protein-coding
LMOD1	leiomodin 1 (smooth muscle)	1	1q32	protein-coding
MAPK8IP1	mitogen-activated protein kinase 8 interacting protein 1	11	11p11.2	protein-coding
NAB2	NGFI-A binding protein 2 (EGR1 binding protein 2)	12	12q13.3	protein-coding
NBPF16	neuroblastoma breakpoint family, member 16	1	1q21.2	protein-coding
NNMT	nicotinamide N-methyltransferase	11	11q23.1	protein-coding
NR2F2	nuclear receptor subfamily 2, group F, member 2	15	15q26	protein-coding
PABPC1L	poly(A) binding protein, cytoplasmic 1-like	20	-	protein-coding
PAQR9	progestin and adipoQ receptor family member IX	3	3q23	protein-coding
PCYT1B	phosphate cytidylyltransferase 1, choline, beta	X	Xp22.11	protein-coding
PHGDH	phosphoglycerate dehydrogenase	1	1p12	protein-coding
PHYHIP	phytanoyl-CoA 2-hydroxylase interacting protein	8	8p21.3	protein-coding
PRSS8	protease, serine, 8	16	16p11.2	protein-coding
RABL2A	RAB, member of RAS oncogene family-like 2A	2	2q13	protein-coding
RDH10	retinol dehydrogenase 10 (all-trans)	8	8q21.11	protein-coding
S100P	S100 calcium binding protein P	4	4p16	protein-coding
SELM	selenoprotein M	22	22q12.2	protein-coding
SEMA4F	sema domain, immunoglobulin domain (Ig), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 4F	2	2p13.1	protein-coding
SERPINB9	serpin peptidase inhibitor, clade B (ovalbumin), member 9	6	6p25	protein-coding
SH3BP2	SH3-domain binding protein 2	4	4p16.3	protein-coding
SYTL1	synaptotagmin-like 1	1	1p36.11	protein-coding
TF	transferrin	3	3q22.1	protein-coding
TMEM133	transmembrane protein 133	11	11q22.1	protein-coding

TMEM86A	transmembrane protein 86A	11	11p15.1	protein-coding
ZC3H6	zinc finger CCCH-type containing 6	2	2q13	protein-coding
AIF1L	allograft inflammatory factor 1-like	9	9q34.13-q34.3	protein-coding
ARHGAP30	Rho GTPase activating protein 30	1	1q23.3	protein-coding
ARRDC4	arrestin domain containing 4	15	15q26.3	protein-coding
BACH2	BTB and CNC homology 1, basic leucine zipper transcription factor 2	6	6q15	protein-coding
BOK	BCL2-related ovarian killer	2	2q37.3	protein-coding
C7orf63	chromosome 7 open reading frame 63	7	7q21.13	protein-coding
CALB1	calbindin 1, 28kDa	8	8q21.3	protein-coding
CCNG2	cyclin G2	4	4q21.1	protein-coding
CD69	CD69 molecule	12	12p13	protein-coding
CLU	clusterin	8	8p21-p12	protein-coding
CRYAB	crystallin, alpha B	11	11q22.3-q23.1	protein-coding
DIRAS3	DIRAS family, GTP-binding RAS-like 3	1	1p31	protein-coding
FN1	fibronectin 1	2	2q34	protein-coding
GPNMB	glycoprotein (transmembrane) nmb	7	7p15	protein-coding
IGF1	insulin-like growth factor 1 (somatomedin C)	12	12q23.2	protein-coding
IL6	interleukin 6 (interferon, beta 2)	7	7p21	protein-coding
INSR	insulin receptor	19	19p13.3-p13.2	protein-coding
LINC00173	long intergenic non-protein coding RNA 173	12	12q24.22	miscRNA
LINC00520	long intergenic non-protein coding RNA 520	14	14q22.3	miscRNA
MAFF	v-maf avian musculoaponeurotic fibrosarcoma oncogene homolog F	22	22q13.1	protein-coding
MEIS3	Meis homeobox 3	19	19q13.32	protein-coding
NBPF10	neuroblastoma breakpoint family, member 10	1	1q21.1	protein-coding
NEBL	nebulette	10	10p12	protein-coding
SYN1	synapsin I	X	Xp11.23	protein-coding
TMEM158	transmembrane protein 158 (gene/pseudogene)	3	3p21.3	protein-coding
TP53INP1	tumor protein p53 inducible nuclear protein 1	8	8q22	protein-coding
TTLL7	tubulin tyrosine ligase-like family, member 7	1	1p31.1	protein-coding
TUBA1A	tubulin, alpha 1a	12	12q13.12	protein-coding
UBTD2	ubiquitin domain containing 2	5	5q35.1	protein-coding

ULBP1	UL16 binding protein 1	6	6q25	protein-coding
UNC5B	unc-5 homolog B (C. elegans)	10	10q22.1	protein-coding
VGF	VGF nerve growth factor inducible	7	7q22.1	protein-coding
YPEL2	yippee-like 2 (Drosophila)	17	17q22	protein-coding
ZFHX2	zinc finger homeobox 2	14	14q11.2	protein-coding
ZFP90	ZFP90 zinc finger protein	16	16q22.1	protein-coding
ADM2	adrenomedullin 2	22	22q13.33	protein-coding
ALDH1L2	aldehyde dehydrogenase 1 family, member L2	12	12q23.3	protein-coding
ANPEP	alanyl (membrane) aminopeptidase	15	15q25-q26	protein-coding
ARHGAP42	Rho GTPase activating protein 42	11	11q22.1	protein-coding
CHAC1	ChaC, cation transport regulator homolog 1 (E. coli)	15	15q15.1	protein-coding
ETV5	ets variant 5	3	3q28	protein-coding
FGD4	FYVE, RhoGEF and PH domain containing 4	12	12p11.21	protein-coding
HBE1	hemoglobin, epsilon 1	11	11p15.5	protein-coding
HEY1	hairy/enhancer-of-split related with YRPW motif 1	8	8q21	protein-coding
IGDCC4	immunoglobulin superfamily, DCC subclass, member 4	15	15q22.31	protein-coding
INHBE	inhibin, beta E	12	12q13.3	protein-coding
LAMC1	laminin, gamma 1 (formerly LAMB2)	1	1q31	protein-coding
LRRK2	leucine-rich repeat kinase 2	12	12q12	protein-coding
MAGEA8	melanoma antigen family A, 8	X	Xq28	protein-coding
MORC4	MORC family CW-type zinc finger 4	X	Xq22.3	protein-coding
NR4A1	nuclear receptor subfamily 4, group A, member 1	12	12q13	protein-coding
PLEKHN1	pleckstrin homology domain containing, family N member 1	1	1p36.33	protein-coding
PVRIG	poliovirus receptor related immunoglobulin domain containing	7	7q22.1	protein-coding
SEMA3F	sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3F	3	3p21.3	protein-coding
SLC30A4	solute carrier family 30 (zinc transporter), member 4	15	15q21.1|15q21.1	protein-coding
SMIM14	small integral membrane protein 14	4	4p14	protein-coding
TEX9	testis expressed 9	15	15q21.3	protein-coding
TGM1	transglutaminase 1	14	14q11.2	protein-coding
TIMP1	TIMP metallopeptidase inhibitor 1	X	Xp11.3-p11.23	protein-coding

TXNDC2	thioredoxin domain containing 2 (spermatozoa)	18	-	protein-coding
ULBP2	UL16 binding protein 2	6	6q25	protein-coding
ZNF662	zinc finger protein 662	3	3p22.1	protein-coding
ACE	angiotensin I converting enzyme	17	17q23.3	protein-coding
ARMC9	armadillo repeat containing 9	2	2q37.1	protein-coding
BTG1	B-cell translocation gene 1, anti-proliferative	12	12q22	protein-coding
C9orf135	chromosome 9 open reading frame 135	9	9q21.12	protein-coding
EPAS1	endothelial PAS domain protein 1	2	2p21-p16	protein-coding
EPS8L2	EPS8-like 2	11	11p15.5	protein-coding
MYO5B	myosin VB	18	18q21	protein-coding
RIN1	Ras and Rab interactor 1	11	11q13.2	protein-coding
SERPINE1	serpin peptidase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 1	7	7q22.1	protein-coding
SLC2A14	solute carrier family 2 (facilitated glucose transporter), member 14	12	12p13.31	protein-coding
SLC41A2	solute carrier family 41 (magnesium transporter), member 2	12	12q23.3	protein-coding
TINAGL1	tubulointerstitial nephritis antigen-like 1	1	1p35.2	protein-coding
UNC13A	unc-13 homolog A (C. elegans)	19	19p13.11	protein-coding
ASS1	argininosuccinate synthase 1	9	9q34.1	protein-coding
DYSF	dysferlin, limb girdle muscular dystrophy 2B (autosomal recessive)	2	2p13.3	protein-coding
FOLR1	folate receptor 1 (adult)	11	11q13.3-q14.1	protein-coding
KCP	kielin/chordin-like protein	7	7q32.1	protein-coding
LAMA4	laminin, alpha 4	6	6q21	protein-coding
LONRF2	LON peptidase N-terminal domain and ring finger 2	2	2q11.2	protein-coding
MGAT5B	mannosyl (alpha-1,6-)-glycoprotein beta-1,6-N-acetyl-glucosaminyltransferase, isozyme B	17	17q25.2	protein-coding
PTPRN	protein tyrosine phosphatase, receptor type, N	2	2q35-q36.1	protein-coding
S100A10	S100 calcium binding protein A10	1	1q21	protein-coding
SCN4A	sodium channel, voltage-gated, type IV, alpha subunit	17	17q23.3	protein-coding
SLCO2B1	solute carrier organic anion transporter family, member 2B1	11	11q13	protein-coding
SPATA31A6	SPATA31 subfamily A, member 6	9	9p11.2	protein-coding
FOS	FBJ murine osteosarcoma viral oncogene	14	14q24.3	protein-coding

	homolog			
HES2	hairy and enhancer of split 2 (Drosophila)	1	1p36.31	protein-coding
RASGRP3	RAS guanyl releasing protein 3 (calcium and DAG-regulated)	2	2p25.1-p24.1	protein-coding
S1PR3	sphingosine-1-phosphate receptor 3	9	9q22.1-q22.2	protein-coding
TCP11L2	t-complex 11, testis-specific-like 2	12	12q23.3	protein-coding
EPSTI1	epithelial stromal interaction 1 (breast)	13	13q13.3	protein-coding
GSN	gelsolin	9	9q33	protein-coding
BCL6	B-cell CLL/lymphoma 6	3	3q27	protein-coding
COL6A4P2	collagen, type VI, alpha 4 pseudogene 2	3	3q22.1	pseudo
ISG20	interferon stimulated exonuclease gene 20kDa	15	15q26	protein-coding

Table B.5. Genes upregulated in mutant ETV6 compared to WT ETV6.

Symbol	Description	Chr	Map location	Type of gene
AHSP	alpha hemoglobin stabilizing protein	16	16p11.2	protein-coding
AIF1	allograft inflammatory factor 1	6	6p21.3	protein-coding
ARHGDIB	Rho GDP dissociation inhibitor (GDI) beta	12	12p12.3	protein-coding
C15orf26	chromosome 15 open reading frame 26	15	15q25.1	protein-coding
CD24	CD24 molecule	6	6q21	protein-coding
GPR85	G protein-coupled receptor 85	7	7q31	protein-coding
ITGB2	integrin, beta 2 (complement component 3 receptor 3 and 4 subunit)	21	21q22.3	protein-coding
ACPP	acid phosphatase, prostate	3	3q22.1	protein-coding
ACSL5	acyl-CoA synthetase long-chain family member 5	10	10q25.1-q25.2	protein-coding
ACVRL1	activin A receptor type II-like 1	12	12q13.13	protein-coding
APOBR	apolipoprotein B receptor	16	16p11	protein-coding
CD33	CD33 molecule	19	19q13.3	protein-coding
CLC	Charcot-Leyden crystal galectin	19	19q13.1	protein-coding
COL9A2	collagen, type IX, alpha 2	1	1p33-p32	protein-coding
CSF2RB	colony stimulating factor 2 receptor, beta, low-affinity (granulocyte-macrophage)	22	22q13.1	protein-coding
EPX	eosinophil peroxidase	17	17q23.1	protein-coding
IFI35	interferon-induced protein 35	17	17q21	protein-coding
IFIT3	interferon-induced protein with tetratricopeptide repeats 3	10	10q24	protein-coding
IFITM2	interferon induced transmembrane protein 2	11	11p15.5	protein-coding
IL18RAP	interleukin 18 receptor accessory protein	2	2q12	protein-coding
INPP5D	inositol polyphosphate-5-phosphatase, 145kDa	2	2q37.1	protein-coding
ITGAM	integrin, alpha M (complement component 3 receptor 3 subunit)	16	16p11.2	protein-coding
KRT12	keratin 12	17	17q12	protein-coding
LINC00482	long intergenic non-protein coding RNA 482	17	17q25.3	miscRNA
LRG1	leucine-rich alpha-2-glycoprotein 1	19	19p13.3	protein-coding
NCF2	neutrophil cytosolic factor 2	1	1q25	protein-coding
PIEZO2	piezo-type mechanosensitive ion channel component 2	18	18p11.22	protein-coding
PIK3CG	phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit gamma	7	7q22.3	protein-coding

PLXNA2	plexin A2	1	1q32.2	protein-coding
PRG2	proteoglycan 2, bone marrow (natural killer cell activator, eosinophil granule major basic protein)	11	11q12	protein-coding
PRG3	proteoglycan 3	11	11q12	protein-coding
PRSS1	protease, serine, 1 (trypsin 1)	7	7q34	protein-coding
PSMB8	proteasome (prosome, macropain) subunit, beta type, 8	6	6p21.3	protein-coding
PSMB9	proteasome (prosome, macropain) subunit, beta type, 9	6	6p21.3	protein-coding
PYCARD	PYD and CARD domain containing	16	16p11.2	protein-coding
RASGRP4	RAS guanyl releasing protein 4	19	19q13.1	protein-coding
RINL	Ras and Rab interactor-like	19	19q13.2	protein-coding
AMHR2	anti-Mullerian hormone receptor, type II	12	12q13	protein-coding
CTSS	cathepsin S	1	1q21	protein-coding
DOCK2	dedicator of cytokinesis 2	5	5q35.1	protein-coding
ELOVL6	ELOVL fatty acid elongase 6	4	4q25	protein-coding
FAM115C	family with sequence similarity 115, member C	7	7q35	protein-coding
GCH1	GTP cyclohydrolase 1	14	14q22.1-q22.2	protein-coding
HYAL1	hyaluronoglucosaminidase 1	3	3p21.3-p21.2	protein-coding
ID3	inhibitor of DNA binding 3, dominant negative helix-loop-helix protein	1	1p36.13-p36.12	protein-coding
IGSF1	immunoglobulin superfamily, member 1	X	Xq25	protein-coding
IKBKE	inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase epsilon	1	1q32.1	protein-coding
IL1RL1	interleukin 1 receptor-like 1	2	2q12	protein-coding
LCMT2	leucine carboxyl methyltransferase 2	15	15q15.3	protein-coding
LOC731424	uncharacterized LOC731424	4	-	miscRNA
MAGEH1	melanoma antigen family H, 1	X	Xp11.21	protein-coding
MUC1	mucin 1, cell surface associated	1	1q21	protein-coding
P2RX6	purinergic receptor P2X, ligand-gated ion channel, 6	22	22q11.21	protein-coding
PAQR8	progestin and adipoQ receptor family member VIII	6	6p12.1	protein-coding
PEX6	peroxisomal biogenesis factor 6	6	6p21.1	protein-coding
RAC2	ras-related C3 botulinum toxin substrate 2 (rho family, small GTP binding protein Rac2)	22	22q13.1	protein-coding
RASSF5	Ras association (RalGDS/AF-6) domain family member 5	1	1q32.1	protein-coding

RILP	Rab interacting lysosomal protein	17	17p13.3	protein-coding
RNASEL	ribonuclease L (2',5'-oligoisoadenylate synthetase-dependent)	1	1q25	protein-coding
SCARF1	scavenger receptor class F, member 1	17	17p13.3	protein-coding
SIGLEC6	sialic acid binding Ig-like lectin 6	19	19q13.3	protein-coding
SLFN14	schlafen family member 14	17	17q12	protein-coding
SYK	spleen tyrosine kinase	9	9q22	protein-coding
TAF1B	TATA box binding protein (TBP)-associated factor, RNA polymerase I, B, 63kDa	2	2p25	protein-coding
TESPA1	thymocyte expressed, positive selection associated 1	12	12q13.2	protein-coding
TMEM173	transmembrane protein 173	5	5q31.2	protein-coding
TMEM74B	transmembrane protein 74B	20	20p13	protein-coding
TRIM67	tripartite motif containing 67	1	1q42.2	protein-coding
UPP1	uridine phosphorylase 1	7	7p12.3	protein-coding
VWA1	von Willebrand factor A domain containing 1	1	1p36.33	protein-coding
APOL1	apolipoprotein L, 1	22	22q13.1	protein-coding
AQP1	aquaporin 1	7	7p14	protein-coding
C10orf54	chromosome 10 open reading frame 54	10	10q22.1	protein-coding
CCDC173	coiled-coil domain containing 173	2	2q31.1	protein-coding
CD19	CD19 molecule	16	16p11.2	protein-coding
COL15A1	collagen, type XV, alpha 1	9	9q21-q22	protein-coding
DES	desmin	2	2q35	protein-coding
ESR2	estrogen receptor 2 (ER beta)	14	14q23.2	protein-coding
GPSM3	G-protein signaling modulator 3	6	6p21.3	protein-coding
IGFBP4	insulin-like growth factor binding protein 4	17	17q12-q21.1	protein-coding
LRMP	lymphoid-restricted membrane protein	12	12p12.1	protein-coding
LST1	leukocyte specific transcript 1	6	6p21.3	protein-coding
LYZ	lysozyme	12	12q15	protein-coding
METTL7A	methyltransferase like 7A	12	12q13.12	protein-coding
NAV1	neuron navigator 1	1	1q32.3	protein-coding
PLAU	plasminogen activator, urokinase	10	10q22.2	protein-coding
PROS1	protein S (alpha)	3	3q11.2	protein-coding
PTGDR2	prostaglandin D2 receptor 2	11	11q12-q13.3	protein-coding
SLAMF6	SLAM family member 6	1	1q23.2	protein-coding

TDRG1	testis development related 1 (non-protein coding)	6	6p21.2	miscRNA
TIMP3	TIMP metallopeptidase inhibitor 3	22	22q12.3	protein-coding
TOR4A	torsin family 4, member A	9	9q34.3	protein-coding
TREML2	triggering receptor expressed on myeloid cells-like 2	6	6p21.1	protein-coding
WNT5B	wingless-type MMTV integration site family, member 5B	12	12p13.3	protein-coding
ZNF467	zinc finger protein 467	7	7q36.1	protein-coding
CEBPE	CCAAT/enhancer binding protein (C/EBP), epsilon	14	14q11.2	protein-coding
COL25A1	collagen, type XXV, alpha 1	4	4q25	protein-coding
CORO1A	coronin, actin binding protein, 1A	16	16p11.2	protein-coding
EVI2B	ecotropic viral integration site 2B	17	17q11.2	protein-coding
LINC00689	long intergenic non-protein coding RNA 689	7	7q36.3	miscRNA
SERPING1	serpin peptidase inhibitor, clade G (C1 inhibitor), member 1	11	11q12.1	protein-coding
CD68	CD68 molecule	17	17p13	protein-coding
HNMT	histamine N-methyltransferase	2	2q22.1	protein-coding
HSPA7	heat shock 70kDa protein 7 (HSP70B)	1	1q23.3	pseudo
MCF2	MCF.2 cell line derived transforming sequence	X	Xq27	protein-coding
SPATA22	spermatogenesis associated 22	17	17p13.3	protein-coding
STAP1	signal transducing adaptor family member 1	4	4q13.2	protein-coding
WAS	Wiskott-Aldrich syndrome	X	Xp11.4-p11.21	protein-coding
ACP5	acid phosphatase 5, tartrate resistant	19	19p13.2	protein-coding
CD53	CD53 molecule	1	1p13	protein-coding
CKMT1B	creatine kinase, mitochondrial 1B	15	15q15	protein-coding
DOK2	docking protein 2, 56kDa	8	8p21.3	protein-coding
FCER1A	Fc fragment of IgE, high affinity I, receptor for; alpha polypeptide	1	1q23	protein-coding
PDE4DIP	phosphodiesterase 4D interacting protein	1	1q12	protein-coding
SLAMF7	SLAM family member 7	1	1q23.1-q24.1	protein-coding
ANTXR1	anthrax toxin receptor 1	2	2p13.1	protein-coding
FCGR2C	Fc fragment of IgG, low affinity IIc, receptor for (CD32) (gene/pseudogene)	1	1q23.3	protein-coding
IL16	interleukin 16	15	15q26.3	protein-coding
MYO1F	myosin IF	19	19p13.3-p13.2	protein-coding

PTPN6	protein tyrosine phosphatase, non-receptor type 6	12	12p13	protein-coding
VIPR2	vasoactive intestinal peptide receptor 2	7	7q36.3	protein-coding
ARL11	ADP-ribosylation factor-like 11	13	13q14.2	protein-coding
GBGT1	globoside alpha-1,3-N-acetylgalactosaminyltransferase 1	9	9q34.13-q34.3	protein-coding
GNG11	guanine nucleotide binding protein (G protein), gamma 11	7	7q21	protein-coding
STAC3	SH3 and cysteine rich domain 3	12	12q13.3	protein-coding
TMEM154	transmembrane protein 154	4	4q31.3	protein-coding
C10orf128	chromosome 10 open reading frame 128	10	10q11.23	protein-coding
ICAM1	intercellular adhesion molecule 1	19	19p13.3-p13.2	protein-coding
PLAT	plasminogen activator, tissue	8	8p12	protein-coding
FAM178B	family with sequence similarity 178, member B	2	2q11.2	protein-coding
FCGR2B	Fc fragment of IgG, low affinity IIb, receptor (CD32)	1	1q23	protein-coding
CKMT1A	creatine kinase, mitochondrial 1A	15	15q15	protein-coding
ANXA2R	annexin A2 receptor	5	5p12	protein-coding
CLDN7	claudin 7	17	17p13.1	protein-coding
ID2	inhibitor of DNA binding 2, dominant negative helix-loop-helix protein	2	2p25	protein-coding
KCNN1	potassium intermediate/small conductance calcium-activated channel, subfamily N, member 1	19	19p13.1	protein-coding
PKI55	DKFZp434H1419	2	2q35	miscRNA
TCN1	transcobalamin I (vitamin B12 binding protein, R binder family)	11	11q11-q12	protein-coding
ZNF208	zinc finger protein 208	19	19p12	protein-coding
ABHD14A	abhydrolase domain containing 14A	3	3p21.1	protein-coding
C7orf71	chromosome 7 open reading frame 71	7	7p15.2	protein-coding
CD84	CD84 molecule	1	1q24	protein-coding
ENO3	enolase 3 (beta, muscle)	17	17p13.2	protein-coding
G0S2	G0/G1switch 2	1	1q32.2	protein-coding
HEPACAM2	HEPACAM family member 2	7	7q21.3	protein-coding
METTL7B	methyltransferase like 7B	12	12q13.2	protein-coding
PLEK	pleckstrin	2	2p13.3	protein-coding
TAX1BP3	Tax1 (human T-cell leukemia virus type I) binding protein 3	17	17p13	protein-coding
TRIM22	tripartite motif containing 22	11	11p15	protein-coding

CD82	CD82 molecule	11	11p11.2	protein-coding
CDKN1C	cyclin-dependent kinase inhibitor 1C (p57, Kip2)	11	11p15.5	protein-coding
FIZ1	FLT3-interacting zinc finger 1	19	19q13.42	protein-coding
GM2A	GM2 ganglioside activator	5	5q33.1	protein-coding
LGALS12	lectin, galactoside-binding, soluble, 12	11	11q13	protein-coding
LRCH2	leucine-rich repeats and calponin homology (CH) domain containing 2	X	Xq23	protein-coding
MGAT3	mannosyl (beta-1,4-)-glycoprotein beta-1,4-N-acetylglucosaminyltransferase	22	22q13.1	protein-coding
MPP1	membrane protein, palmitoylated 1, 55kDa	X	Xq28	protein-coding
NEK9	NIMA-related kinase 9	14	14q24.3	protein-coding
RABEP2	rabaptin, RAB GTPase binding effector protein 2	16	16p11.2	protein-coding
RHCE	Rh blood group, CcEe antigens	1	1p36.11	protein-coding
RNASE1	ribonuclease, RNase A family, 1 (pancreatic)	14	14q11.2	protein-coding
SDC1	syndecan 1	2	2p24.1	protein-coding
SOSTDC1	sclerostin domain containing 1	7	7p21.1	protein-coding
SPI1	spleen focus forming virus (SFFV) proviral integration oncogene	11	11p11.2	protein-coding
STAT6	signal transducer and activator of transcription 6, interleukin-4 induced	12	12q13	protein-coding
STK36	serine/threonine kinase 36	2	2q35	protein-coding
TNFAIP2	tumor necrosis factor, alpha-induced protein 2	14	14q32	protein-coding
VAV1	vav 1 guanine nucleotide exchange factor	19	19p13.2	protein-coding
ACSS1	acyl-CoA synthetase short-chain family member 1	20	20p11.23-p11.21	protein-coding
ALKBH7	alkB, alkylation repair homolog 7 (E. coli)	19	19p13.3	protein-coding
ARHGAP25	Rho GTPase activating protein 25	2	2p13.3	protein-coding
ARHGAP4	Rho GTPase activating protein 4	X	Xq28	protein-coding
CCBP2	NA	NA	NA	NA
CREB5	cAMP responsive element binding protein 5	7	7p15.1	protein-coding
F8	coagulation factor VIII, procoagulant component	X	Xq28	protein-coding
GDF15	growth differentiation factor 15	19	19p13.11	protein-coding
HCLS1	hematopoietic cell-specific Lyn substrate 1	3	3q13	protein-coding
HDAC7	histone deacetylase 7	12	12q13.1	protein-coding
IL18R1	interleukin 18 receptor 1	2	2q12	protein-coding

ITGA2B	integrin, alpha 2b (platelet glycoprotein IIb of IIb/IIIa complex, antigen CD41)	17	17q21.32	protein-coding
KCNN4	potassium intermediate/small conductance calcium-activated channel, subfamily N, member 4	19	19q13.2	protein-coding
KIAA0040	KIAA0040	1	1q24-q25	protein-coding
LOC100505702	uncharacterized LOC100505702	4	-	miscRNA
LPCAT2	lysophosphatidylcholine acyltransferase 2	16	16q12.2	protein-coding
LPPR3	lipid phosphate phosphatase-related protein type 3	19	19p13.3	protein-coding
NBEAL2	neurobeachin-like 2	3	3p21.31	protein-coding
NYNRIN	NYN domain and retroviral integrase containing	14	14q12	protein-coding
ORAI1	ORAI calcium release-activated calcium modulator 1	12	12q24.31	protein-coding
POU3F2	POU class 3 homeobox 2	6	6q16	protein-coding
PPM1H	protein phosphatase, Mg2+/Mn2+ dependent, 1H	12	12q14.1	protein-coding
PTPN22	protein tyrosine phosphatase, non-receptor type 22 (lymphoid)	1	1p13.2	protein-coding
RMDN1	regulator of microtubule dynamics 1	8	8q21.3	protein-coding
RMND1	required for meiotic nuclear division 1 homolog (S. cerevisiae)	6	6q25.1	protein-coding
RNF25	ring finger protein 25	2	2q35	protein-coding
RNF39	ring finger protein 39	6	6p21.3	protein-coding
8-Sep	septin 8	5	5q31	protein-coding
SLC40A1	solute carrier family 40 (iron-regulated transporter), member 1	2	2q32	protein-coding
SMOX	spermine oxidase	20	20p13	protein-coding
SORBS3	sorbin and SH3 domain containing 3	8	8p21.3	protein-coding
SPTBN2	spectrin, beta, non-erythrocytic 2	11	11q13	protein-coding
TLR8-AS1	TLR8 antisense RNA 1	X	Xp22.31	miscRNA
TMEM253	transmembrane protein 253	14	14q11.2	protein-coding
TNIP2	TNFAIP3 interacting protein 2	4	4p16.3	protein-coding
TTC14	tetratricopeptide repeat domain 14	3	3q26.33	protein-coding
UNC13D	unc-13 homolog D (C. elegans)	17	17q25.1	protein-coding
USP11	ubiquitin specific peptidase 11	X	Xp11.23	protein-coding
VILL	villin-like	3	3p21.3	protein-coding
VPS25	vacuolar protein sorting 25 homolog (S. cerevisiae)	17	17q21.31	protein-coding
WDR55	WD repeat domain 55	5	5q31.3	protein-coding

YIF1B	Yip1 interacting factor homolog B (S. cerevisiae)	19	19q13.2	protein-coding
ZNF792	zinc finger protein 792	19	19q13.11	protein-coding
ANKLE1	ankyrin repeat and LEM domain containing 1	19	19p13.11	protein-coding
ARHGEF1	Rho guanine nucleotide exchange factor (GEF) 1	19	19q13.13	protein-coding
BAI3	brain-specific angiogenesis inhibitor 3	6	6q12	protein-coding
C19orf54	chromosome 19 open reading frame 54	19	19q13.2	protein-coding
CACNB3	calcium channel, voltage-dependent, beta 3 subunit	12	12q13	protein-coding
CARD11	caspase recruitment domain family, member 11	7	7p22	protein-coding
CHCHD5	coiled-coil-helix-coiled-coil-helix domain containing 5	2	2q13	protein-coding
CORO1B	coronin, actin binding protein, 1B	11	11q13.2	protein-coding
CSPG4	chondroitin sulfate proteoglycan 4	15	15q24.2	protein-coding
DNAH8	dynein, axonemal, heavy chain 8	6	6p21.2	protein-coding
DND1	DND microRNA-mediated repression inhibitor 1	5	5q31.3	protein-coding
DRAM1	DNA-damage regulated autophagy modulator 1	12	12q23.2	protein-coding
EDARADD	EDAR-associated death domain	1	1q42.3	protein-coding
FASTKD1	FAST kinase domains 1	2	2q31	protein-coding
FJX1	four jointed box 1 (Drosophila)	11	11p13	protein-coding
FRAT1	frequently rearranged in advanced T-cell lymphomas	10	10q24.1	protein-coding
GALM	galactose mutarotase (aldose 1-epimerase)	2	2p22.1	protein-coding
GLRX	glutaredoxin (thioltransferase)	5	5q14	protein-coding
HBBP1	hemoglobin, beta pseudogene 1	11	11p15.5	pseudo
HEATR6	HEAT repeat containing 6	17	17q23.1	protein-coding
HES1	hairy and enhancer of split 1, (Drosophila)	3	3q28-q29	protein-coding
KAZALD1	Kazal-type serine peptidase inhibitor domain 1	10	10q24.31	protein-coding
KRT10	keratin 10	17	17q21	protein-coding
LOC100506746	uncharacterized LOC100506746	4	4q21	miscRNA
LOC153684	uncharacterized LOC153684	5	5p12	miscRNA
LOC388553	uncharacterized LOC388553	19	19q13.32	miscRNA
METTL6	methyltransferase like 6	3	3p25.1	protein-coding
MITD1	MIT, microtubule interacting and transport,	2	2q11.2	protein-coding

	domain containing 1			
MORF4L2-AS1	MORF4L2 antisense RNA 1	X	Xq22.2	miscRNA
NFIB	nuclear factor I/B	9	9p24.1	protein-coding
NTRK1	neurotrophic tyrosine kinase, receptor, type 1	1	1q21-q22	protein-coding
PHLDA1	pleckstrin homology-like domain, family A, member 1	12	12q15	protein-coding
PICALM	phosphatidylinositol binding clathrin assembly protein	11	11q14	protein-coding
PPP1R14A	protein phosphatase 1, regulatory (inhibitor) subunit 14A	19	19q13.1	protein-coding
PTPRC	protein tyrosine phosphatase, receptor type, C	1	1q31-q32	protein-coding
RCSD1	RCSD domain containing 1	1	1q24.2	protein-coding
ROBO2	roundabout, axon guidance receptor, homolog 2 (Drosophila)	3	3p12.3	protein-coding
SELPLG	selectin P ligand	12	12q24	protein-coding
SLC17A7	solute carrier family 17 (vesicular glutamate transporter), member 7	19	19q13	protein-coding
SLC2A4	solute carrier family 2 (facilitated glucose transporter), member 4	17	17p13	protein-coding
SMIM10	small integral membrane protein 10	X	Xq26.3	protein-coding
SNRNP200	small nuclear ribonucleoprotein 200kDa (U5)	2	2q11.2	protein-coding
SNTA1	syntrophin, alpha 1	20	20q11.2	protein-coding
SORT1	sortilin 1	1	1p13.3|1p21.3-p13.1	protein-coding
SP110	SP110 nuclear body protein	2	2q37.1	protein-coding
SPRED2	sprouty-related, EVH1 domain containing 2	2	2p14	protein-coding
SYNE2	spectrin repeat containing, nuclear envelope 2	14	14q23.2	protein-coding
TBRG1	transforming growth factor beta regulator 1	11	11q24.2	protein-coding
TEX19	testis expressed 19	17	17q25.3	protein-coding
TGM2	transglutaminase 2	20	20q12	protein-coding
TNPO3	transportin 3	7	7q32.1	protein-coding
TSPAN14	tetraspanin 14	10	10q23.1	protein-coding
WIPF1	WAS/WASL interacting protein family, member 1	2	2q31.1	protein-coding
ZBTB37	zinc finger and BTB domain containing 37	1	1q25.1	protein-coding
ANKRD55	ankyrin repeat domain 55	5	5q11.2	protein-coding
APLNR	apelin receptor	11	11q12	protein-coding

C10orf71	chromosome 10 open reading frame 71	10	10q11.23	protein-coding
C21orf59	chromosome 21 open reading frame 59	21	21q22.1	protein-coding
C5orf63	chromosome 5 open reading frame 63	5	5q23.2	protein-coding
DDN	dendrin	12	12q13.12	protein-coding
DERA	deoxyribose-phosphate aldolase (putative)	12	12p12.3	protein-coding
EGLN3	egl-9 family hypoxia-inducible factor 3	14	14q13.1	protein-coding
FADS1	fatty acid desaturase 1	11	11q12.2-q13.1	protein-coding
FAM110A	family with sequence similarity 110, member A	20	20p13	protein-coding
FLJ46284	uncharacterized LOC441369	8	8q22.1	miscRNA
INO80C	INO80 complex subunit C	18	18q12.2	protein-coding
LCP2	lymphocyte cytosolic protein 2 (SH2 domain containing leukocyte protein of 76kDa)	5	5q35.1	protein-coding
LOC100507173	NA	NA	NA	NA
LOC257358	uncharacterized LOC257358	5	5q35.1	miscRNA
LOC728613	programmed cell death 6 pseudogene	5	5p15.33	pseudo
MILR1	mast cell immunoglobulin-like receptor 1	17	17q23.3	protein-coding
MTMR9	myotubularin related protein 9	8	8p23-p22	protein-coding
PANX1	pannexin 1	11	11q21	protein-coding
PLEKHA7	pleckstrin homology domain containing, family A member 7	11	11p15.1	protein-coding
RAB20	RAB20, member RAS oncogene family	13	13q34	protein-coding
RAB3D	RAB3D, member RAS oncogene family	19	19p13.2	protein-coding
RHOG	ras homolog family member G	11	11p15.5-p15.4	protein-coding
RPF1	ribosome production factor 1 homolog (S. cerevisiae)	1	1p22.3	protein-coding
SLC24A2	solute carrier family 24 (sodium/potassium/calcium exchanger), member 2	9	9p22.1	protein-coding
SLC25A41	solute carrier family 25, member 41	19	19p13.3	protein-coding
SLIT2	slit homolog 2 (Drosophila)	4	4p15.2	protein-coding
SQRDL	sulfide quinone reductase-like (yeast)	15	15q15	protein-coding
TMEM156	transmembrane protein 156	4	4p14	protein-coding
TMEM99	transmembrane protein 99	17	17q21.2	protein-coding
TNFRSF1A	tumor necrosis factor receptor superfamily, member 1A	12	12p13.2	protein-coding
CCDC39	coiled-coil domain containing 39	3	3q26.33	protein-coding
FYB	FYN binding protein	5	5p13.1	protein-coding

GPRC5B	G protein-coupled receptor, family C, group 5, member B	16	16p12	protein-coding
IFITM3	interferon induced transmembrane protein 3	11	11p15.5	protein-coding
LAPTM5	lysosomal protein transmembrane 5	1	1p34	protein-coding
LHFPL2	lipoma HMGIC fusion partner-like 2	5	5q14.1	protein-coding
LINC00534	long intergenic non-protein coding RNA 534	8	-	miscRNA
MAML3	mastermind-like 3 (Drosophila)	4	4q28	protein-coding
NCKAP1L	NCK-associated protein 1-like	12	12q13.1	protein-coding
PEX12	peroxisomal biogenesis factor 12	17	17q12	protein-coding
SERPINB1	serpin peptidase inhibitor, clade B (ovalbumin), member 1	6	6p25	protein-coding
TRANK1	tetratricopeptide repeat and ankyrin repeat containing 1	3	3p22.2	protein-coding
ADAMTS14	ADAM metallopeptidase with thrombospondin type 1 motif, 14	10	10q21	protein-coding
APOM	apolipoprotein M	6	6p21.33	protein-coding
ASIC4	acid-sensing (proton-gated) ion channel family member 4	2	2q35	protein-coding
BMX	BMX non-receptor tyrosine kinase	X	Xp22.2	protein-coding
DBNDD2	dysbindin (dystrobrevin binding protein 1) domain containing 2	20	20q13.12	protein-coding
GPR158	G protein-coupled receptor 158	10	10p12.1	protein-coding
HTR1F	5-hydroxytryptamine (serotonin) receptor 1F, G protein-coupled	3	3p12	protein-coding
IFI27L2	interferon, alpha-inducible protein 27-like 2	14	14q32.12	protein-coding
MLKL	mixed lineage kinase domain-like	16	16q23.1	protein-coding
MPZ	myelin protein zero	1	1q23.3	protein-coding
ARHGAP9	Rho GTPase activating protein 9	12	12q13.3	protein-coding
CCDC19	coiled-coil domain containing 19	1	1q22	protein-coding
CCRL2	chemokine (C-C motif) receptor-like 2	3	3p21	protein-coding
FOXH1	forkhead box H1	8	8q24.3	protein-coding
NSUN5	NOP2/Sun domain family, member 5	7	7q11.23	protein-coding
PCSK9	proprotein convertase subtilisin/kexin type 9	1	1p32.3	protein-coding
RNPEP	arginyl aminopeptidase (aminopeptidase B)	1	1q32	protein-coding
TMTC2	transmembrane and tetratricopeptide repeat containing 2	12	12q21.31	protein-coding
ZNF311	zinc finger protein 311	6	6p22.1	protein-coding

C19orf66	chromosome 19 open reading frame 66	19	19p13.2	protein-coding
CNRIP1	cannabinoid receptor interacting protein 1	2	2p14	protein-coding
LGALS3BP	lectin, galactoside-binding, soluble, 3 binding protein	17	17q25	protein-coding
MESDC1	mesoderm development candidate 1	15	15q13	protein-coding
NDRG2	NDRG family member 2	14	14q11.2	protein-coding
PLCB2	phospholipase C, beta 2	15	15q15	protein-coding
ARAP3	ArfGAP with RhoGAP domain, ankyrin repeat and PH domain 3	5	5q31.3	protein-coding
C17orf62	chromosome 17 open reading frame 62	17	17q25.3	protein-coding
C3AR1	complement component 3a receptor 1	12	12p13.31	protein-coding
EFNA1	ephrin-A1	1	1q21-q22	protein-coding
ID1	inhibitor of DNA binding 1, dominant negative helix-loop-helix protein	20	20q11	protein-coding
SCG3	secretogranin III	15	15q21	protein-coding
APOL4	apolipoprotein L, 4	22	22q11.2-q13.2	protein-coding
C9orf64	chromosome 9 open reading frame 64	9	9q21.32	protein-coding
UBE2D1	ubiquitin-conjugating enzyme E2D 1	10	10q21.1	protein-coding
C21orf67	chromosome 21 open reading frame 67	21	21q22.3	miscRNA
HVCN1	hydrogen voltage-gated channel 1	12	12q24.11	protein-coding
PTGS1	prostaglandin-endoperoxide synthase 1 (prostaglandin G/H synthase and cyclooxygenase)	9	9q32-q33.3	protein-coding
UBXN8	UBX domain protein 8	8	8p12-p11.2	protein-coding
LOC100505875	NA	NA	NA	NA
OAS3	2'-5'-oligoadenylate synthetase 3, 100kDa	12	12q24.2	protein-coding
LOC154761	family with sequence similarity 115, member C pseudogene	7	7q35	pseudo

Table B.6. GO-Seq categories for genes downregulated in mutant ETV6 compared to WT ETV6.

Category	Term	Over-represented p-value	Over-represented FDR	# D.E. in cat.	# genes in cat
GO:0002376	immune system process	6.51E-19	7.14E-15	172	1256
GO:0006955	immune response	8.13E-17	4.46E-13	113	716
GO:0006952	defense response	1.10E-15	3.20E-12	112	729
GO:0009611	response to wounding	1.17E-15	3.20E-12	98	610
GO:0045321	leukocyte activation	2.85E-15	6.25E-12	68	346
GO:0001775	cell activation	6.38E-15	1.17E-11	82	474
GO:0009605	response to external stimulus	1.60E-13	2.50E-10	127	948
GO:0050896	response to stimulus	4.87E-13	6.67E-10	384	4129
GO:0002684	positive regulation of immune system process	1.89E-12	2.07E-09	68	389
GO:0040011	locomotion	1.84E-12	2.07E-09	99	733
GO:0044707	single-multicellular organism process	2.64E-12	2.63E-09	292	3009
GO:0002682	regulation of immune system process	4.16E-12	3.51E-09	94	644
GO:0032943	mononuclear cell proliferation	4.10E-12	3.51E-09	32	111
GO:0070661	leukocyte proliferation	8.35E-12	6.54E-09	32	114
GO:0046649	lymphocyte activation	1.06E-11	7.74E-09	55	296
GO:0043207	response to external biotic stimulus	1.25E-11	8.30E-09	61	334
GO:0046651	lymphocyte proliferation	1.29E-11	8.30E-09	31	109
GO:0048583	regulation of response to stimulus	2.11E-11	1.28E-08	186	1722
GO:0050865	regulation of cell activation	2.29E-11	1.32E-08	45	213
GO:0002694	regulation of leukocyte activation	2.73E-11	1.50E-08	43	198
GO:0050776	regulation of immune response	2.87E-11	1.50E-08	71	444
GO:0032501	multicellular organismal process	3.03E-11	1.51E-08	295	3107
GO:0009607	response to biotic stimulus	3.96E-11	1.89E-08	62	352
GO:0048870	cell motility	4.48E-11	2.05E-08	79	562
GO:0016477	cell migration	6.30E-11	2.66E-08	74	516
GO:0051707	response to other organism	6.26E-11	2.66E-08	59	329
GO:0002253	activation of immune response	1.01E-10	4.09E-08	48	251
GO:0023052	signaling	1.14E-10	4.30E-08	276	2887
GO:0044700	single organism signaling	1.14E-10	4.30E-08	276	2887
GO:0006954	inflammatory response	1.43E-10	5.23E-08	46	235

GO:0007154	cell communication	1.60E-10	5.57E-08	279	2932
GO:0048584	positive regulation of response to stimulus	1.64E-10	5.57E-08	106	826
GO:0051716	cellular response to stimulus	1.68E-10	5.57E-08	314	3365
GO:0007165	signal transduction	2.51E-10	8.09E-08	258	2669
GO:0007166	cell surface receptor signaling pathway	5.04E-10	1.58E-07	155	1425
GO:0002757	immune response-activating signal transduction	5.44E-10	1.66E-07	44	230
GO:0050778	positive regulation of immune response	6.20E-10	1.84E-07	51	288
GO:0098542	defense response to other organism	8.67E-10	2.50E-07	37	166
GO:0001816	cytokine production	1.00E-09	2.81E-07	51	291
GO:0002274	myeloid leukocyte activation	1.36E-09	3.73E-07	23	74
GO:0001817	regulation of cytokine production	1.44E-09	3.86E-07	47	259
GO:0006928	cellular component movement	2.84E-09	7.42E-07	101	856
GO:0050900	leukocyte migration	3.56E-09	9.09E-07	30	133
GO:0050867	positive regulation of cell activation	3.80E-09	9.47E-07	32	141
GO:0032944	regulation of mononuclear cell proliferation	4.61E-09	1.12E-06	24	86
GO:0002252	immune effector process	4.73E-09	1.13E-06	54	333
GO:0051249	regulation of lymphocyte activation	6.47E-09	1.51E-06	36	175
GO:0045087	innate immune response	7.59E-09	1.71E-06	70	496
GO:0070663	regulation of leukocyte proliferation	7.63E-09	1.71E-06	24	88
GO:0051239	regulation of multicellular organismal process	8.00E-09	1.75E-06	120	1063
GO:0051674	localization of cell	9.00E-09	1.94E-06	53	357
GO:0042110	T cell activation	1.52E-08	3.20E-06	38	203
GO:0050670	regulation of lymphocyte proliferation	1.79E-08	3.70E-06	23	85
GO:0042098	T cell proliferation	2.32E-08	4.71E-06	22	79
GO:0002696	positive regulation of leukocyte activation	2.49E-08	4.91E-06	30	136
GO:0042221	response to chemical	2.51E-08	4.91E-06	184	1817
GO:0007155	cell adhesion	3.04E-08	5.86E-06	61	453
GO:0050851	antigen receptor-mediated signaling pathway	3.26E-08	6.17E-06	23	89
GO:0022610	biological adhesion	3.39E-08	6.31E-06	61	454
GO:0006950	response to stress	3.78E-08	6.85E-06	205	2064
GO:0050864	regulation of B cell activation	3.81E-08	6.85E-06	17	51

GO:0042060	wound healing	4.96E-08	8.77E-06	55	383
GO:0048534	hematopoietic or lymphoid organ development	6.51E-08	1.13E-05	46	296
GO:0070887	cellular response to chemical stimulus	7.82E-08	1.34E-05	134	1245
GO:0051251	positive regulation of lymphocyte activation	8.25E-08	1.39E-05	28	128
GO:0065008	regulation of biological quality	8.70E-08	1.45E-05	168	1672
GO:0065007	biological regulation	9.08E-08	1.49E-05	475	5822
GO:0002429	immune response-activating cell surface receptor signaling pathway	9.32E-08	1.50E-05	29	139
GO:0034097	response to cytokine	1.24E-07	1.98E-05	53	351
GO:0045577	regulation of B cell differentiation	1.45E-07	2.27E-05	9	15
GO:0040012	regulation of locomotion	2.18E-07	3.37E-05	46	314
GO:0050878	regulation of body fluid levels	2.22E-07	3.38E-05	52	366
GO:0002764	immune response-regulating signaling pathway	3.36E-07	5.05E-05	46	304
GO:0044763	single-organism cellular process	4.19E-07	6.21E-05	535	6655
GO:0030097	hemopoiesis	5.61E-07	8.20E-05	42	275
GO:0002520	immune system development	8.94E-07	0.000129	46	318
GO:0007599	hemostasis	9.63E-07	0.000137	46	317
GO:0001818	negative regulation of cytokine production	1.08E-06	0.000152	20	81
GO:0031349	positive regulation of defense response	1.50E-06	0.000209	30	163
GO:0050853	B cell receptor signaling pathway	1.61E-06	0.000221	10	24
GO:0051241	negative regulation of multicellular organismal process	1.66E-06	0.000224	32	183
GO:0007596	blood coagulation	1.75E-06	0.000234	45	314
GO:0010033	response to organic substance	1.78E-06	0.000235	139	1366
GO:1902531	regulation of intracellular signal transduction	1.93E-06	0.000252	92	842
GO:0050817	coagulation	2.07E-06	0.000268	45	316
GO:0006935	chemotaxis	2.16E-06	0.000272	43	302
GO:0042330	taxis	2.16E-06	0.000272	43	302
GO:0050870	positive regulation of T cell activation	2.37E-06	0.000295	21	94
GO:0009966	regulation of signal transduction	2.40E-06	0.000296	135	1359
GO:0051270	regulation of cellular component movement	2.43E-06	0.000297	44	321
GO:0050871	positive regulation of B cell activation	2.69E-06	0.000324	12	35

GO:0002521	leukocyte differentiation	2.85E-06	0.00034	35	221
GO:0040013	negative regulation of locomotion	3.06E-06	0.00036	21	104
GO:0032101	regulation of response to external stimulus	3.56E-06	0.000415	40	264
GO:0051607	defense response to virus	3.82E-06	0.000441	25	126
GO:0070665	positive regulation of leukocyte proliferation	4.19E-06	0.000479	16	60
GO:0044699	single-organism process	4.94E-06	0.000559	577	7326
GO:0030888	regulation of B cell proliferation	5.29E-06	0.000592	10	27
GO:0009617	response to bacterium	5.42E-06	0.0006	29	163
GO:0030334	regulation of cell migration	5.50E-06	0.000603	38	268
GO:0050789	regulation of biological process	5.84E-06	0.000634	447	5549
GO:0030154	cell differentiation	5.95E-06	0.00064	158	1673
GO:0031347	regulation of defense response	6.42E-06	0.000683	42	284
GO:0071310	cellular response to organic substance	6.93E-06	0.000731	108	1034
GO:0002250	adaptive immune response	7.07E-06	0.000739	24	126
GO:2000145	regulation of cell motility	7.18E-06	0.000743	40	291
GO:0042113	B cell activation	8.01E-06	0.000813	22	114
GO:0048869	cellular developmental process	7.99E-06	0.000813	169	1821
GO:0002683	negative regulation of immune system process	8.95E-06	0.000893	21	103
GO:0023051	regulation of signaling	8.96E-06	0.000893	143	1488
GO:0010646	regulation of cell communication	1.03E-05	0.001016	143	1493
GO:0050863	regulation of T cell activation	1.07E-05	0.001044	24	128
GO:0060326	cell chemotaxis	1.08E-05	0.001052	17	73
GO:0009615	response to virus	1.18E-05	0.00114	30	174
GO:0030155	regulation of cell adhesion	1.23E-05	0.00117	28	176
GO:0003013	circulatory system process	1.25E-05	0.001186	28	167
GO:0035456	response to interferon-beta	1.35E-05	0.001265	7	12
GO:0097530	granulocyte migration	1.44E-05	0.001335	10	29
GO:0006909	phagocytosis	1.52E-05	0.001406	21	109
GO:0032946	positive regulation of mononuclear cell proliferation	1.58E-05	0.001441	15	59
GO:0071345	cellular response to cytokine stimulus	1.67E-05	0.00151	41	285
GO:0048731	system development	1.69E-05	0.001517	175	1925
GO:0003008	system process	3.03E-05	0.002704	64	556

GO:0032103	positive regulation of response to external stimulus	3.27E-05	0.002877	17	79
GO:0050852	T cell receptor signaling pathway	3.28E-05	0.002877	16	72
GO:0033993	response to lipid	3.31E-05	0.00288	44	328
GO:0042100	B cell proliferation	3.41E-05	0.002949	10	32
GO:0002237	response to molecule of bacterial origin	3.45E-05	0.002956	23	126
GO:0097529	myeloid leukocyte migration	3.61E-05	0.00307	12	44
GO:0032496	response to lipopolysaccharide	3.67E-05	0.003096	22	118
GO:0030593	neutrophil chemotaxis	3.76E-05	0.003126	8	20
GO:1990266	neutrophil migration	3.76E-05	0.003126	8	20
GO:0002275	myeloid cell activation involved in immune response	4.07E-05	0.003332	10	33
GO:0016337	cell-cell adhesion	4.05E-05	0.003332	26	170
GO:0048514	blood vessel morphogenesis	4.28E-05	0.00348	29	199
GO:0002263	cell activation involved in immune response	5.11E-05	0.004021	18	91
GO:0002366	leukocyte activation involved in immune response	5.11E-05	0.004021	18	91
GO:0009653	anatomical structure morphogenesis	5.13E-05	0.004021	118	1249
GO:0030029	actin filament-based process	5.05E-05	0.004021	42	334
GO:0071621	granulocyte chemotaxis	5.08E-05	0.004021	9	27
GO:0050671	positive regulation of lymphocyte proliferation	5.51E-05	0.004285	14	58
GO:0006897	endocytosis	6.01E-05	0.004642	39	301
GO:0006812	cation transport	6.31E-05	0.004839	50	398
GO:0006691	leukotriene metabolic process	6.51E-05	0.004955	7	15
GO:0045071	negative regulation of viral genome replication	6.60E-05	0.004991	9	26
GO:0030098	lymphocyte differentiation	6.70E-05	0.005032	24	148
GO:0048513	organ development	6.80E-05	0.005073	125	1324
GO:0002697	regulation of immune effector process	7.04E-05	0.00522	24	141
GO:0018108	peptidyl-tyrosine phosphorylation	7.11E-05	0.005237	23	140
GO:0009967	positive regulation of signal transduction	7.25E-05	0.0053	66	593
GO:0045619	regulation of lymphocyte differentiation	7.70E-05	0.005592	14	60
GO:0018212	peptidyl-tyrosine modification	8.21E-05	0.005921	23	141
GO:0050777	negative regulation of immune response	8.41E-05	0.006029	10	34

GO:0002819	regulation of adaptive immune response	8.78E-05	0.006213	15	67
GO:0019370	leukotriene biosynthetic process	8.75E-05	0.006213	6	11
GO:0042742	defense response to bacterium	8.91E-05	0.006266	12	45
GO:0030595	leukocyte chemotaxis	9.20E-05	0.006429	13	55
GO:0045576	mast cell activation	9.38E-05	0.006508	9	28
GO:0023056	positive regulation of signaling	9.54E-05	0.00658	67	611
GO:0033627	cell adhesion mediated by integrin	9.69E-05	0.006604	9	28
GO:0048518	positive regulation of biological process	9.69E-05	0.006604	217	2481
GO:0050854	regulation of antigen receptor-mediated signaling pathway	0.000105	0.00706	8	23
GO:0060412	ventricular septum morphogenesis	0.000105	0.00706	7	18
GO:0042129	regulation of T cell proliferation	0.000107	0.007184	14	62
GO:0050729	positive regulation of inflammatory response	0.00011	0.007339	9	28
GO:0003143	embryonic heart tube morphogenesis	0.000115	0.0076	9	32
GO:0002704	negative regulation of leukocyte mediated immunity	0.000117	0.007688	5	8
GO:0010647	positive regulation of cell communication	0.00012	0.007859	67	616
GO:0044351	macropinocytosis	0.000124	0.008049	4	5
GO:0034340	response to type I interferon	0.000128	0.008287	12	46
GO:0007265	Ras protein signal transduction	0.000139	0.008882	26	180
GO:0030336	negative regulation of cell migration	0.00014	0.008882	16	88
GO:0042107	cytokine metabolic process	0.00014	0.008882	12	50
GO:0033628	regulation of cell adhesion mediated by integrin	0.000142	0.008926	8	23
GO:0030183	B cell differentiation	0.000144	0.009046	13	61
GO:0001944	vasculature development	0.000147	0.009188	37	301
GO:0002768	immune response-regulating cell surface receptor signaling pathway	0.000154	0.009531	31	224
GO:1902533	positive regulation of intracellular signal transduction	0.000155	0.00957	47	388
GO:0050793	regulation of developmental process	0.000159	0.009731	92	933
GO:0002703	regulation of leukocyte mediated immunity	0.00016	0.009777	14	64
GO:0050795	regulation of behavior	0.000168	0.010158	13	60
GO:0002820	negative regulation of adaptive immune response	0.000171	0.010318	5	9

GO:0042102	positive regulation of T cell proliferation	0.000181	0.010867	11	42
GO:0048872	homeostasis of number of cells	0.000188	0.011208	20	119
GO:0030036	actin cytoskeleton organization	0.000192	0.011408	38	310
GO:0007275	multicellular organismal development	0.0002	0.01181	203	2379
GO:0046596	regulation of viral entry into host cell	0.000208	0.012189	5	9
GO:0002695	negative regulation of leukocyte activation	0.000211	0.012217	12	53
GO:0035455	response to interferon-alpha	0.00021	0.012217	6	13
GO:0050727	regulation of inflammatory response	0.000212	0.01225	18	99
GO:0048562	embryonic organ morphogenesis	0.000222	0.01272	20	125
GO:2000146	negative regulation of cell motility	0.000223	0.012762	16	91
GO:0045579	positive regulation of B cell differentiation	0.000235	0.013265	5	10
GO:0071300	cellular response to retinoic acid	0.000234	0.013265	8	26
GO:0001947	heart looping	0.000236	0.013287	8	28
GO:0030198	extracellular matrix organization	0.000239	0.013298	24	168
GO:0043062	extracellular structure organization	0.000239	0.013298	24	168
GO:0001568	blood vessel development	0.000251	0.01392	31	243
GO:0019221	cytokine-mediated signaling pathway	0.000255	0.014075	31	220
GO:0035556	intracellular signal transduction	0.000258	0.014167	128	1380
GO:0006690	icosanoid metabolic process	0.000268	0.014463	10	36
GO:0060586	multicellular organismal iron ion homeostasis	0.000265	0.014463	4	6
GO:1901568	fatty acid derivative metabolic process	0.000268	0.014463	10	36
GO:0002685	regulation of leukocyte migration	0.000274	0.014671	11	44
GO:0032970	regulation of actin filament-based process	0.000274	0.014671	24	163
GO:0043303	mast cell degranulation	0.000276	0.014671	7	20
GO:0032632	interleukin-3 production	0.000294	0.015588	3	3
GO:0002443	leukocyte mediated immunity	0.000296	0.01562	21	129
GO:0050794	regulation of cellular process	0.000304	0.015974	416	5292
GO:0001525	angiogenesis	0.000312	0.016281	27	200
GO:0002218	activation of innate immune response	0.000315	0.016381	20	121
GO:0051271	negative regulation of cellular component movement	0.000331	0.017126	16	94
GO:0009887	organ morphogenesis	0.000335	0.017261	44	391

GO:0030225	macrophage differentiation	0.000345	0.017699	7	20
GO:0008015	blood circulation	0.000374	0.019061	19	114
GO:0007266	Rho protein signal transduction	0.000386	0.019433	11	49
GO:0035295	tube development	0.000386	0.019433	31	251
GO:0065009	regulation of molecular function	0.000384	0.019433	131	1420
GO:0038003	opioid receptor signaling pathway	0.000392	0.019621	3	3
GO:0002279	mast cell activation involved in immune response	0.000395	0.019637	7	21
GO:0030574	collagen catabolic process	0.000396	0.019637	8	31
GO:1902105	regulation of leukocyte differentiation	0.000399	0.019718	18	105
GO:0048732	gland development	0.000401	0.019727	20	136
GO:0043299	leukocyte degranulation	0.000408	0.019804	8	28
GO:0046718	viral entry into host cell	0.000408	0.019804	7	21
GO:0051094	positive regulation of developmental process	0.000408	0.019804	48	422
GO:0061371	determination of heart left/right asymmetry	0.000416	0.020112	8	30
GO:0031294	lymphocyte costimulation	0.000427	0.020448	9	34
GO:0031295	T cell costimulation	0.000427	0.020448	9	34
GO:0048568	embryonic organ development	0.000438	0.020882	28	211
GO:0000165	MAPK cascade	0.000466	0.021747	40	337
GO:0008154	actin polymerization or depolymerization	0.000461	0.021747	17	99
GO:0035050	embryonic heart tube development	0.000466	0.021747	9	38
GO:0060337	type I interferon signaling pathway	0.000466	0.021747	11	45
GO:0071357	cellular response to type I interferon	0.000466	0.021747	11	45
GO:0042089	cytokine biosynthetic process	0.000482	0.022415	11	49
GO:0008283	cell proliferation	0.000489	0.022639	96	981
GO:0071396	cellular response to lipid	0.000492	0.022691	21	136
GO:0030041	actin filament polymerization	0.000497	0.02279	15	81
GO:0023014	signal transduction by phosphorylation	0.000505	0.023058	41	351
GO:0044243	multicellular organismal catabolic process	0.00051	0.023204	8	32
GO:0046456	icosanoid biosynthetic process	0.000519	0.023413	8	26
GO:1901570	fatty acid derivative biosynthetic process	0.000519	0.023413	8	26
GO:0072511	divalent inorganic cation transport	0.000529	0.023763	21	138
GO:0050866	negative regulation of cell activation	0.000553	0.024769	12	58

GO:0046597	negative regulation of viral entry into host cell	0.000558	0.024861	4	6
GO:0045088	regulation of innate immune response	0.000563	0.025012	25	173
GO:0033003	regulation of mast cell activation	0.000573	0.02536	6	16
GO:0045595	regulation of cell differentiation	0.000584	0.025729	68	673
GO:0045069	regulation of viral genome replication	0.000588	0.025801	10	41
GO:0002823	negative regulation of adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains	0.000617	0.026847	4	7
GO:0007264	small GTPase mediated signal transduction	0.000616	0.026847	42	369
GO:0051250	negative regulation of lymphocyte activation	0.000636	0.027581	10	44
GO:0030182	neuron differentiation	0.000652	0.028152	60	594
GO:0050857	positive regulation of antigen receptor-mediated signaling pathway	0.000658	0.02832	4	7
GO:0048521	negative regulation of behavior	0.000685	0.029335	5	12
GO:0042127	regulation of cell proliferation	0.000691	0.029494	74	724
GO:0030001	metal ion transport	0.000723	0.03075	37	308
GO:0042116	macrophage activation	0.000737	0.031231	6	17
GO:0045089	positive regulation of innate immune response	0.000764	0.032238	21	138
GO:0050730	regulation of peptidyl-tyrosine phosphorylation	0.000773	0.032502	16	93
GO:0050856	regulation of T cell receptor signaling pathway	0.000795	0.033282	6	17
GO:0002460	adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains	0.000801	0.033422	18	110
GO:0061448	connective tissue development	0.000811	0.033683	17	110
GO:0030260	entry into host cell	0.000844	0.034042	7	23
GO:0032649	regulation of interferon-gamma production	0.00084	0.034042	8	28
GO:0044409	entry into host	0.000844	0.034042	7	23
GO:0051806	entry into cell of other organism involved in symbiotic interaction	0.000844	0.034042	7	23
GO:0051828	entry into other organism involved in symbiotic interaction	0.000844	0.034042	7	23
GO:0052126	movement in host environment	0.000844	0.034042	7	23
GO:0052192	movement in environment of other	0.000844	0.034042	7	23

	organism involved in symbiotic interaction				
GO:0060411	cardiac septum morphogenesis	0.000834	0.034042	7	24
GO:0002707	negative regulation of lymphocyte mediated immunity	0.000866	0.034795	4	7
GO:0048646	anatomical structure formation involved in morphogenesis	0.000874	0.034969	52	500
GO:0002224	toll-like receptor signaling pathway	0.000889	0.035367	17	102
GO:0032963	collagen metabolic process	0.00089	0.035367	10	49
GO:0071222	cellular response to lipopolysaccharide	0.000902	0.035587	12	58
GO:2000026	regulation of multicellular organismal development	0.000902	0.035587	67	673
GO:0048699	generation of neurons	0.000932	0.036591	64	651
GO:0070838	divalent metal ion transport	0.000934	0.036591	20	135
GO:0003281	ventricular septum development	0.000947	0.036955	7	25
GO:0048856	anatomical structure development	0.000951	0.036969	196	2345
GO:0003231	cardiac ventricle development	0.000983	0.037975	10	46
GO:0050818	regulation of coagulation	0.000982	0.037975	9	36
GO:0043408	regulation of MAPK cascade	0.000991	0.038054	35	295
GO:0080134	regulation of response to stress	0.000992	0.038054	58	542
GO:0002221	pattern recognition receptor signaling pathway	0.001018	0.0389	18	114
GO:0002244	hematopoietic progenitor cell differentiation	0.001031	0.039188	11	54
GO:0032956	regulation of actin cytoskeleton organization	0.001033	0.039188	21	146
GO:0002822	regulation of adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains	0.001047	0.039595	12	59
GO:0050731	positive regulation of peptidyl-tyrosine phosphorylation	0.001056	0.039794	12	62
GO:0044767	single-organism developmental process	0.001072	0.040278	226	2752
GO:0035587	purinergic receptor signaling pathway	0.001118	0.041851	4	7
GO:0002758	innate immune response-activating signal transduction	0.001128	0.042099	18	115
GO:0001776	leukocyte homeostasis	0.001137	0.042256	9	38
GO:0002444	myeloid leukocyte mediated immunity	0.001149	0.042573	8	33
GO:0022008	neurogenesis	0.00117	0.043212	67	694

GO:0032502	developmental process	0.001204	0.044328	228	2784
GO:0045730	respiratory burst	0.001224	0.044914	6	17
GO:0045621	positive regulation of lymphocyte differentiation	0.00127	0.046425	9	38
GO:0048525	negative regulation of viral process	0.001321	0.04815	10	46
GO:0071219	cellular response to molecule of bacterial origin	0.001348	0.048967	12	61
GO:0030890	positive regulation of B cell proliferation	0.001368	0.049528	6	19

Table B.7. GO-Seq categories for genes upregulated in mutant ETV6 compared to WT ETV6.

Category	Term	Over-represented p-value	Over-represented FDR	# D.E. in cat.	# genes in cat.
GO:0032501	multicellular organismal process	2.84E-12	3.12E-08	134	3107
GO:0044707	single-multicellular organism process	7.58E-12	4.16E-08	130	3009
GO:0048731	system development	3.47E-11	1.27E-07	95	1925
GO:0007275	multicellular organismal development	2.73E-10	7.49E-07	107	2379
GO:0032502	developmental process	2.89E-09	5.53E-06	116	2784
GO:0044767	single-organism developmental process	3.03E-09	5.53E-06	115	2752
GO:0048513	organ development	4.80E-09	7.53E-06	69	1324
GO:0030154	cell differentiation	9.14E-09	1.25E-05	80	1673
GO:0048856	anatomical structure development	1.15E-08	1.40E-05	101	2345
GO:0048869	cellular developmental process	3.75E-08	4.12E-05	83	1821
GO:0023052	signaling	7.76E-08	7.10E-05	114	2887
GO:0044700	single organism signaling	7.76E-08	7.10E-05	114	2887
GO:0050896	response to stimulus	1.31E-07	0.000110356	147	4129
GO:0001775	cell activation	1.62E-07	0.000126572	33	474
GO:0007154	cell communication	1.76E-07	0.000128742	114	2932
GO:0048545	response to steroid hormone	3.28E-07	0.000224832	17	153
GO:0009653	anatomical structure morphogenesis	5.62E-07	0.000362655	61	1249
GO:0007165	signal transduction	7.63E-07	0.000464661	104	2669
GO:0032940	secretion by cell	9.49E-07	0.000547558	29	413
GO:0046903	secretion	1.09E-06	0.000596155	31	462
GO:0051897	positive regulation of protein kinase B signaling	1.27E-06	0.000665693	9	43
GO:0002576	platelet degranulation	1.53E-06	0.000720918	10	57
GO:0042574	retinal metabolic process	1.58E-06	0.000720918	4	5
GO:0097305	response to alcohol	1.49E-06	0.000720918	15	133
GO:0001944	vasculature development	1.64E-06	0.000721367	24	301
GO:0014070	response to organic cyclic compound	2.26E-06	0.00095406	25	344
GO:0006928	cellular component movement	2.65E-06	0.001075034	46	856
GO:0042127	regulation of cell proliferation	3.07E-06	0.001203603	40	724

GO:0008283	cell proliferation	3.61E-06	0.001366164	49	981
GO:2000026	regulation of multicellular organismal development	5.46E-06	0.001996541	38	673
GO:0009611	response to wounding	6.65E-06	0.002351815	35	610
GO:0030168	platelet activation	7.21E-06	0.002470296	15	150
GO:0001101	response to acid	1.07E-05	0.00332156	15	156
GO:0009753	response to jasmonic acid	1.09E-05	0.00332156	3	3
GO:0051239	regulation of multicellular organismal process	1.07E-05	0.00332156	51	1063
GO:0071395	cellular response to jasmonic acid stimulus	1.09E-05	0.00332156	3	3
GO:0018149	peptide cross-linking	1.43E-05	0.004245427	4	7
GO:0009605	response to external stimulus	1.71E-05	0.004931176	46	948
GO:0042221	response to chemical	1.76E-05	0.00494602	74	1817
GO:0016477	cell migration	1.85E-05	0.00508363	31	516
GO:0040011	locomotion	2.29E-05	0.006131259	39	733
GO:0001568	blood vessel development	2.75E-05	0.007002243	19	243
GO:0009584	detection of visible light	2.71E-05	0.007002243	7	35
GO:0050865	regulation of cell activation	2.93E-05	0.007176344	17	213
GO:0051896	regulation of protein kinase B signaling	2.94E-05	0.007176344	9	62
GO:0001523	retinoid metabolic process	3.19E-05	0.007613257	6	25
GO:0010942	positive regulation of cell death	3.68E-05	0.008589634	20	281
GO:0048870	cell motility	3.81E-05	0.008712243	32	562
GO:0002696	positive regulation of leukocyte activation	3.94E-05	0.008816507	13	136
GO:0006887	exocytosis	4.52E-05	0.00991043	16	195
GO:0016101	diterpenoid metabolic process	4.89E-05	0.010520851	6	27
GO:0002138	retinoic acid biosynthetic process	5.48E-05	0.011342129	3	4
GO:0016102	diterpenoid biosynthetic process	5.48E-05	0.011342129	3	4
GO:0065008	regulation of biological quality	5.68E-05	0.011527489	68	1672
GO:0008285	negative regulation of cell proliferation	5.93E-05	0.01162137	22	336
GO:0050867	positive regulation of cell activation	5.83E-05	0.01162137	13	141
GO:0014910	regulation of smooth muscle cell migration	7.33E-05	0.014111751	5	18
GO:0001525	angiogenesis	8.24E-05	0.014663477	16	200
GO:0009583	detection of light stimulus	8.27E-05	0.014663477	7	41
GO:0045321	leukocyte activation	8.11E-05	0.014663477	22	346

GO:0046649	lymphocyte activation	7.90E-05	0.014663477	20	296
GO:1901342	regulation of vasculature development	8.29E-05	0.014663477	10	86
GO:0033993	response to lipid	9.69E-05	0.01687607	21	328
GO:0006721	terpenoid metabolic process	0.0001	0.017823102	6	31
GO:0006954	inflammatory response	0.00011	0.01817888	17	235
GO:0002828	regulation of type 2 immune response	0.00012	0.018773448	4	11
GO:0032943	mononuclear cell proliferation	0.00011	0.018773448	11	111
GO:2000379	positive regulation of reactive oxygen species metabolic process	0.00012	0.018773448	5	21
GO:0043627	response to estrogen	0.00012	0.019598182	9	76
GO:0014074	response to purine-containing compound	0.00013	0.019947695	9	76
GO:0042573	retinoic acid metabolic process	0.00013	0.019947695	3	5
GO:0010951	negative regulation of endopeptidase activity	0.00013	0.020019747	9	77
GO:0002548	monocyte chemotaxis	0.00015	0.020595389	4	12
GO:0002694	regulation of leukocyte activation	0.00015	0.020595389	15	198
GO:0006955	immune response	0.00015	0.020595389	35	716
GO:0007603	phototransduction, visible light	0.00015	0.020595389	6	32
GO:0009888	tissue development	0.00014	0.020595389	37	748
GO:0010466	negative regulation of peptidase activity	0.00014	0.020595389	9	78
GO:0030198	extracellular matrix organization	0.00016	0.020595389	14	168
GO:0043062	extracellular structure organization	0.00016	0.020595389	14	168
GO:0043491	protein kinase B signaling	0.00014	0.020595389	9	75
GO:0070661	leukocyte proliferation	0.00015	0.020595389	11	114
GO:0071383	cellular response to steroid hormone stimulus	0.00015	0.020595389	6	32
GO:0045622	regulation of T-helper cell differentiation	0.00016	0.020819061	4	12
GO:0048583	regulation of response to stimulus	0.00016	0.020944634	68	1722
GO:0050793	regulation of developmental process	0.00017	0.021046152	43	933
GO:0051716	cellular response to stimulus	0.00017	0.021376522	115	3365
GO:0007586	digestion	0.00018	0.022244249	5	22
GO:0009628	response to abiotic stimulus	0.00018	0.022244249	29	552
GO:0043068	positive regulation of programmed cell death	0.00019	0.023620464	18	268

GO:0006952	defense response	0.0002	0.024418821	35	729
GO:0014909	smooth muscle cell migration	0.00021	0.024783135	5	22
GO:0042063	gliogenesis	0.00021	0.024783135	10	96
GO:0030638	polyketide metabolic process	0.00023	0.025249832	3	6
GO:0030647	aminoglycoside antibiotic metabolic process	0.00023	0.025249832	3	6
GO:0042092	type 2 immune response	0.00022	0.025249832	4	13
GO:0044597	daunorubicin metabolic process	0.00023	0.025249832	3	6
GO:0044598	doxorubicin metabolic process	0.00023	0.025249832	3	6
GO:0051674	localization of cell	0.00023	0.025249832	22	357
GO:0046683	response to organophosphorus	0.00024	0.026293389	8	66
GO:0016114	terpenoid biosynthetic process	0.00025	0.026814132	3	6
GO:0023056	positive regulation of signaling	0.00026	0.026814132	31	611
GO:0043154	negative regulation of cysteine-type endopeptidase activity involved in apoptotic process	0.00026	0.026814132	7	50
GO:0051591	response to cAMP	0.00025	0.026814132	7	50
GO:0070887	cellular response to chemical stimulus	0.00025	0.026814132	52	1245
GO:0010001	glial cell differentiation	0.00026	0.027051122	9	81
GO:0071229	cellular response to acid	0.00027	0.027203008	9	82
GO:0048514	blood vessel morphogenesis	0.00027	0.027558134	15	199
GO:0009966	regulation of signal transduction	0.00028	0.028219333	56	1359
GO:0007166	cell surface receptor signaling pathway	0.00029	0.02840361	58	1425
GO:0009725	response to hormone	0.0003	0.02840361	24	427
GO:0010647	positive regulation of cell communication	0.0003	0.02840361	31	616
GO:0045628	regulation of T-helper 2 cell differentiation	0.00029	0.02840361	3	6
GO:0050670	regulation of lymphocyte proliferation	0.0003	0.02840361	9	85
GO:0090280	positive regulation of calcium ion import	0.00029	0.02840361	3	6
GO:0043367	CD4-positive, alpha-beta T cell differentiation	0.0003	0.028517361	5	24
GO:0043370	regulation of CD4-positive, alpha-beta T cell differentiation	0.00031	0.029420847	4	14
GO:0032944	regulation of mononuclear cell proliferation	0.00032	0.029648673	9	86
GO:2000117	negative regulation of cysteine-type endopeptidase activity	0.00032	0.029648673	7	52

GO:1901700	response to oxygen-containing compound	0.00035	0.032369269	32	656
GO:0022008	neurogenesis	0.00036	0.032708151	34	694
GO:0002250	adaptive immune response	0.00037	0.033142547	11	126
GO:0002684	positive regulation of immune system process	0.00037	0.033142547	22	389
GO:0070663	regulation of leukocyte proliferation	0.00038	0.033683803	9	88
GO:0051251	positive regulation of lymphocyte activation	0.0004	0.034770911	11	128
GO:0071384	cellular response to corticosteroid stimulus	0.0004	0.034770911	4	15
GO:0070374	positive regulation of ERK1 and ERK2 cascade	0.0004	0.034851361	6	39
GO:0007602	phototransduction	0.00041	0.034939258	6	38
GO:0046651	lymphocyte proliferation	0.00044	0.036732007	10	109
GO:1902533	positive regulation of intracellular signal transduction	0.00043	0.036732007	22	388
GO:0002443	leukocyte mediated immunity	0.00044	0.037040567	11	129
GO:0042060	wound healing	0.00045	0.037060989	22	383
GO:0040013	negative regulation of locomotion	0.00046	0.03737374	10	104
GO:2000377	regulation of reactive oxygen species metabolic process	0.00045	0.03737374	6	41
GO:0002460	adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains	0.00047	0.037587278	10	110
GO:0030098	lymphocyte differentiation	0.00047	0.037587278	12	148
GO:0031589	cell-substrate adhesion	0.00047	0.037587278	12	142
GO:0045064	T-helper 2 cell differentiation	0.00047	0.037587278	3	7
GO:0002376	immune system process	0.00049	0.038019135	51	1256
GO:0050673	epithelial cell proliferation	0.00049	0.038019135	12	146
GO:0071407	cellular response to organic cyclic compound	0.00049	0.038019135	11	128
GO:0000165	MAPK cascade	0.00051	0.038436899	20	337
GO:0023051	regulation of signaling	0.00051	0.038436899	59	1488
GO:0072358	cardiovascular system development	0.0005	0.038436899	24	428
GO:0072359	circulatory system development	0.0005	0.038436899	24	428
GO:0001822	kidney development	0.00053	0.038518193	10	107
GO:0001935	endothelial cell proliferation	0.00053	0.038518193	7	55
GO:0007596	blood coagulation	0.00054	0.038518193	19	314

GO:0010512	negative regulation of phosphatidylinositol biosynthetic process	0.00057	0.038518193	2	2
GO:0010646	regulation of cell communication	0.00056	0.038518193	59	1493
GO:0035710	CD4-positive, alpha-beta T cell activation	0.00052	0.038518193	5	27
GO:0035788	cell migration involved in metanephros development	0.00057	0.038518193	2	2
GO:0035789	metanephric mesenchymal cell migration	0.00057	0.038518193	2	2
GO:0035793	positive regulation of metanephric mesenchymal cell migration by platelet-derived growth factor receptor-beta signaling pathway	0.00057	0.038518193	2	2
GO:0051249	regulation of lymphocyte activation	0.00051	0.038518193	13	175
GO:0060326	cell chemotaxis	0.00053	0.038518193	8	73
GO:0060664	epithelial cell proliferation involved in salivary gland morphogenesis	0.00053	0.038518193	2	2
GO:0061061	muscle structure development	0.00057	0.038518193	17	260
GO:1900238	regulation of metanephric mesenchymal cell migration by platelet-derived growth factor receptor-beta signaling pathway	0.00057	0.038518193	2	2
GO:2000514	regulation of CD4-positive, alpha-beta T cell activation	0.00055	0.038518193	4	16
GO:2000589	regulation of metanephric mesenchymal cell migration	0.00057	0.038518193	2	2
GO:2000591	positive regulation of metanephric mesenchymal cell migration	0.00057	0.038518193	2	2
GO:0007599	hemostasis	0.0006	0.038700461	19	317
GO:0014812	muscle cell migration	0.0006	0.038700461	5	27
GO:0030195	negative regulation of blood coagulation	0.00061	0.038700461	4	17
GO:0030336	negative regulation of cell migration	0.00059	0.038700461	9	88
GO:0050671	positive regulation of lymphocyte proliferation	0.0006	0.038700461	7	58
GO:0050817	coagulation	0.00059	0.038700461	19	316
GO:0071505	response to mycophenolic acid	0.0006	0.038700461	2	2
GO:0071506	cellular response to mycophenolic acid	0.0006	0.038700461	2	2
GO:0072126	positive regulation of glomerular mesangial cell proliferation	0.0006	0.038700461	2	2
GO:1900047	negative regulation of hemostasis	0.00061	0.038700461	4	17
GO:1902531	regulation of intracellular signal transduction	0.00062	0.039114879	38	842

GO:0002262	myeloid cell homeostasis	0.00063	0.039572179	8	74
GO:0035556	intracellular signal transduction	0.00064	0.039984551	55	1380
GO:0046636	negative regulation of alpha-beta T cell activation	0.00065	0.040229671	3	7
GO:0032946	positive regulation of mononuclear cell proliferation	0.00066	0.040683857	7	59
GO:0022603	regulation of anatomical structure morphogenesis	0.00068	0.04155294	22	392
GO:0032355	response to estradiol	0.00068	0.04155294	6	43
GO:0050866	negative regulation of cell activation	0.00068	0.04155294	7	58
GO:0008284	positive regulation of cell proliferation	0.00069	0.041806687	21	376
GO:0042493	response to drug	0.0007	0.042056058	14	203
GO:0045766	positive regulation of angiogenesis	0.00071	0.042749828	6	42
GO:0009967	positive regulation of signal transduction	0.00073	0.042882607	29	593
GO:0021700	developmental maturation	0.00074	0.042882607	10	114
GO:0031960	response to corticosteroid	0.00073	0.042882607	7	60
GO:0044699	single-organism process	0.00074	0.042882607	211	7326
GO:0045624	positive regulation of T-helper cell differentiation	0.00074	0.042882607	3	8
GO:0070665	positive regulation of leukocyte proliferation	0.00073	0.042882607	7	60
GO:2000146	negative regulation of cell motility	0.00074	0.042882607	9	91
GO:0006959	humoral immune response	0.00075	0.042935285	6	44
GO:0010757	negative regulation of plasminogen activation	0.00076	0.042935285	2	2
GO:0072215	regulation of metanephros development	0.00075	0.042935285	3	8
GO:0009719	response to endogenous stimulus	0.00078	0.044009849	34	741
GO:0046634	regulation of alpha-beta T cell activation	0.00079	0.044303116	5	29
GO:0072593	reactive oxygen species metabolic process	0.00083	0.046351277	8	82
GO:0002526	acute inflammatory response	0.00085	0.04710004	6	44
GO:0022414	reproductive process	0.00085	0.04710004	26	518
GO:0097529	myeloid leukocyte migration	0.00085	0.04710004	6	44
GO:0023014	signal transduction by phosphorylation	0.00086	0.047402822	20	351
GO:0002252	immune effector process	0.00088	0.04757797	19	333
GO:0009581	detection of external stimulus	0.00087	0.04757797	7	59

GO:0070372	regulation of ERK1 and ERK2 cascade	0.00089	0.047989911	7	62
GO:0030097	hemopoiesis	0.00089	0.048100756	17	275
GO:0007517	muscle organ development	0.00093	0.049283931	13	179
GO:0050819	negative regulation of coagulation	0.00093	0.049283931	4	19
GO:0051271	negative regulation of cellular component movement	0.00093	0.049283931	9	94

Table B.8. *ETV6* mutation variant allele fractions in tumors of affected individuals in Family A.

Individual	Cancer type	Mutation	Effect	Variant reads	Total reads	VAF
II-5	Multiple myeloma	c.1195C>T	p.R399C	466	989	0.47
II-5	Colon cancer	c.1195C>T	p.R399C	333	738	0.45
III-2	MDS (RAEB-1)	c.1195C>T	p.R399C	313	701	0.45

RAEB-1, refractory anemia with excess blasts 1; VAF, variant allele fraction

Table B.9. Somatic mutations in cancers from affected members of Family A.

Paired tumor and fibroblast samples were subjected to targeted capture and high-throughput sequencing of 194 cancer-associated genes[3]. Mutations were verified to be somatic by their absence in the matched marrow fibroblast DNA.

Individual	Cancer type	Gene	RefSeq	Mutation	Effect	Variant reads	Total reads	VAF
II-5	Colon cancer	*BRAF*	NM_004333.4	c.1799T>A	p.V600E	261	627	0.42
II-5	Colon cancer	*CTNNB1*	NM_001904.3	c.1004A>T	p.K335I	233	608	0.38
II-5	Colon cancer	*CTNNB1*	NM_001904.3	c.1178C>A	p.T393N	247	674	0.37
II-5	Colon cancer	*GNAS*	NM_000516.4	c.601C>A[#]	p.R201S	35	574	0.06
II-5	Colon cancer	*GNAS*	NM_000516.4	c.601C>T[#]	p.R201C	113	574	0.20
II-5	Colon cancer	*GNAS*	NM_000516.4	c.602G>A[#]	p.R201H	22	577	0.04
II-5	Colon cancer	*PTEN*	NM_000314.4	c.70G>A[#]	p.D24N	14	310	0.05
II-5	Colon cancer	*PTEN*	NM_000314.4	c.71A>G[#]	p.D24G	14	309	0.05
II-5	Colon cancer	*PTEN*	NM_000314.4	c.202_203dup	p.N69fs	7	90	0.08
II-5	Colon cancer	*TP53*	NM_000546.5	c.659A>G	p.Y220C	21	268	0.08
II-5	Multiple myeloma	*CDK8*	NM_001260.1	c.324C>G	p.I108M	45	670	0.07
II-5	Multiple myeloma	*KMT2A*	NM_005933.3	c.9932C>G	p.A3311G	103	1203	0.09
III-2	MDS (RAEB-1)	*BCOR*	NM_001123385.1	c.3231_3234del	p.1077_1078del	181	518	0.35
III-2	MDS (RAEB-1)	*KRAS*	NM_004985.3	c.190T>A	p.Y64N	224	571	0.39
III-2	MDS (RAEB-1)	*RUNX1*	NM_001754.4	c.424dupG	p.A142fs	351	869	0.40

RAEB-1, refractory anemia with excess blasts-1; VAF, variant allele fraction
[#]Mutations in *trans*

Table B.10. Primer and probe sequences.

pGL3 MMP3 Promoter Acc65I Forward	GGTACCCCAAGTCAACCTACTTCTTCTC
pGL3 MMP3 Promoter XhoI Reverse	CTCGAGTTCCACTGGCTTTACTTAGC
pGL3 PF4 Promoter Acc65I Forward	GGTACCTGCTTGTATGTAAGTTGAGGTTGGAGGT
pGL3 PF4 Promoter XhoI Reverse	CTCGAGGAAACTCGGGCTGGGTCTCT
pHAT ETV6 ETS KpnI Forward	GGTACCATAGCAGACTGTAGACTGCTTTGG
pHAT ETV6 ETS EcoRI-stop Reverse	GAATTCTCATCGGCCACTCATGATTTCATCTG
ETV6 genomic exon 5 Forward	TTCTGGGGAGAAAGGTCC
ETV6 genomic exon 5 Reverse	AAGCCTCCAATTGCCCTC
ETV6 genomic exon 6 Forward	TGGTTAGTGCCTCAACAAGAAA
ETV6 genomic exon 6 Reverse	TTGCCCAGCAAGGAATAAAC
ETV6 genomic exon 7 Forward	GCTAAAGGGTGGGGACAGAT
ETV6 genomic exon 7 Reverse	TTTGGCTTTATTACAGGACAGACA
p.Ala93Asp QuikChange Forward	TGAAATGAATGGCAAAGATCTCCTGCTGCTGACCA
p.Ala93Asp QuikChange Reverse	TGGTCAGCAGCAGGAGATCTTTGCCATTCATTTCA
p.Val112Glu QuikChange Forward	TCCTCATTCAGGTGATGAGCTCTATGAACTCCTTC
p.Val112Glu QuikChange Reverse	GAAGGAGTTCATAGAGCTCATCACCTGAATGAGGA
p.Pro214Leu QuikChange Forward	GCCGCCTCTCCCTGGCTGAGAGAGC
p.Pro214Leu QuikChange Reverse	GCTCTCTCAGCCAGGGAGAGGCGGC
p.Arg369Gln QuikChange Forward	CAAAGAATCCAAAATATTCCAGATAGTGGATCCCAACGGAC
p.Arg369Gln QuikChange Reverse	GTCCGTTGGGATCCACTATCTGGAATATTTTGGATTCTTTG
p.Arg399Cys QuikChange Forward	AAATGTCCAGAGCCCTGTGCCACTACTACAAACTA
p.Arg399Cys QuikChange Reverse	TAGTTTGTAGTAGTGGCACAGGGCTCTGGACATTT
ETS consensus EMSA probe Forward	CGGCCAAGCCGGAAGTGAGTGCCTCGA
ETS consensus EMSA probe Reverse	GGCACTCACTTCCGGCTTGGCCGTCGA

Bibliography

1. Greenberg, P. *et al.* International Scoring System for Evaluating Prognosis in Myelodysplastic Syndromes. *Blood* **89,** 2079–2088 (1997).
2. Mason, P. J. Stem cells, telomerase and dyskeratosis congenita. *BioEssays News Rev. Mol. Cell. Dev. Biol.* **25,** 126–133 (2003).
3. Kee, Y. & D'Andrea, A. D. Molecular pathogenesis and clinical management of Fanconi anemia. *J. Clin. Invest.* **122,** 3799–3806 (2012).
4. Ball, S. Diamond Blackfan Anemia. *ASH Educ. Program Book* **2011,** 487–491 (2011).
5. Burroughs, L., Woolfrey, A. & Shimamura, A. Shwachman-Diamond syndrome: a review of the clinical presentation, molecular pathogenesis, diagnosis, and treatment. *Hematol. Oncol. Clin. North Am.* **23,** 233–248 (2009).
6. Shimamura, A. Clinical approach to marrow failure. *Hematol. Educ. Program Am. Soc. Hematol. Am. Soc. Hematol. Educ. Program* 329–337 (2009). doi:10.1182/asheducation-2009.1.329
7. Marsh, J. C. W. *et al.* Guidelines for the diagnosis and management of aplastic anaemia. *Br. J. Haematol.* **147,** 43–70 (2009).
8. Teo, J. T. *et al.* Clinical and Genetic Analysis of Unclassifiable Inherited Bone Marrow Failure Syndromes. *Pediatrics* **122,** e139–e148 (2008).
9. Shimamura, A. & Alter, B. P. Pathophysiology and management of inherited bone marrow failure syndromes. *Blood Rev.* **24,** 101–122 (2010).
10. Liu, L. *et al.* Comparison of Next-Generation Sequencing Systems. *J. Biomed. Biotechnol.* **2012,** (2012).

11. Shimamura, A. Clinical approach to marrow failure. *ASH Educ. Program Book* **2009,** 329–337 (2009).

12. Churpek, J. E. *et al.* Proposal for the clinical detection and management of patients and their family members with familial myelodysplastic syndrome/acute leukemia predisposition syndromes. *Leuk. Lymphoma* **54,** 28–35 (2013).

13. Vardiman, J. W. *et al.* The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. *Blood* **114,** 937–951 (2009).

14. Miller, S. A., Dykes, D. D. & Polesky, H. F. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* **16,** 1215 (1988).

15. Bellodi, C. *et al.* H/ACA small RNA dysfunctions in disease reveal key roles for noncoding RNA modifications in hematopoietic stem cell differentiation. *Cell Rep.* **3,** 1493–1502 (2013).

16. Pippucci, T. *et al.* Mutations in the 5′ UTR of *ANKRD26*, the ankirin repeat domain 26 gene, cause an autosomal-dominant form of inherited thrombocytopenia, THC2. *Am. J. Hum. Genet.* **88,** 115–120 (2011).

17. Hsu, A. P. *et al.* GATA2 haploinsufficiency caused by mutations in a conserved intronic element leads to MonoMAC syndrome. *Blood* **121,** 3830–3837 (2013).

18. Walne, A. J., Vulliamy, T., Kirwan, M., Plagnol, V. & Dokal, I. Constitutional mutations in *RTEL1* cause severe dyskeratosis congenita. *Am. J. Hum. Genet.* **92,** 448–453 (2013).

19. Bogliolo, M. *et al.* Mutations in *ERCC4*, encoding the DNA-repair endonuclease XPF, cause Fanconi anemia. *Am. J. Hum. Genet.* **92,** 800–806 (2013).

20. Walsh, T. *et al.* Detection of inherited mutations for breast and ovarian cancer using genomic capture and massively parallel sequencing. *Proc. Natl. Acad. Sci.* **107,** 12629–12633 (2010).
21. Walsh, T. *et al.* Mutations in 12 genes for inherited ovarian, fallopian tube, and peritoneal carcinoma identified by massively parallel sequencing. *Proc. Natl. Acad. Sci.* **108,** 18032–18037 (2011).
22. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinforma. Oxf. Engl.* **25,** 1754–1760 (2009).
23. Pritchard, C. C. *et al.* Validation and implementation of targeted capture and sequencing for the detection of actionable mutation, copy number variation, and gene rearrangement in clinical cancer specimens. *J. Mol. Diagn.* **16,** 56–67 (2014).
24. Gulsuner, S. *et al.* Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network. *Cell* **154,** 518–529 (2013).
25. Nord, A. S., Lee, M., King, M.-C. & Walsh, T. Accurate and exact CNV identification from targeted high-throughput sequence data. *BMC Genomics* **12,** 184 (2011).
26. Pinto, F. O. *et al.* Diagnosis of Fanconi anemia in patients with bone marrow failure. *Haematologica* **94,** 487–495 (2009).
27. Alter, B. P. *et al.* Very short telomere length by flow fluorescence in situ hybridization identifies patients with dyskeratosis congenita. *Blood* **110,** 1439–1447 (2007).
28. Ip, W. F. *et al.* Serum pancreatic enzymes define the pancreatic phenotype in patients with Shwachman-Diamond syndrome. *J. Pediatr.* **141,** 259–265 (2002).

29. Hsu, A. P. *et al.* Mutations in *GATA2* are associated with the autosomal dominant and sporadic monocytopenia and mycobacterial infection (MonoMAC) syndrome. *Blood* **118,** 2653–2655 (2011).

30. Dickinson, R. E. *et al.* Exome sequencing identifies *GATA-2* mutation as the cause of dendritic cell, monocyte, B and NK lymphoid deficiency. *Blood* **118,** 2656–2658 (2011).

31. Kazenwadel, J. *et al.* Loss-of-function germline *GATA2* mutations in patients with MDS/AML or MonoMAC syndrome and primary lymphedema reveal a key role for GATA2 in the lymphatic vasculature. *Blood* **119,** 1283–1291 (2012).

32. Pasquet, M. *et al.* High frequency of *GATA2* mutations in patients with mild chronic neutropenia evolving to MonoMac syndrome, myelodysplasia, and acute myeloid leukemia. *Blood* **121,** 822–829 (2013).

33. Ostergaard, P. *et al.* Mutations in *GATA2* cause primary lymphedema associated with a predisposition to acute myeloid leukemia (Emberger syndrome). *Nat. Genet.* **43,** 929–931 (2011).

34. Spinner, M. A. *et al.* GATA2 deficiency: a protean disorder of hematopoiesis, lymphatics, and immunity. *Blood* **123,** 809–821 (2014).

35. Fasan, A. *et al. GATA2* mutations are frequent in intermediate-risk karyotype AML with biallelic *CEBPA* mutations and are associated with favorable prognosis. *Leukemia* **27,** 482–485 (2013).

36. Dickinson, R. E. *et al.* The evolution of cellular deficiency in *GATA2* mutation. *Blood* **123,** 863–874 (2014).

37. Song, W.-J. *et al.* Haploinsufficiency of *CBFA2* causes familial thrombocytopenia with propensity to develop acute myelogenous leukaemia. *Nat. Genet.* **23,** 166–175 (1999).

38. Liew, E. & Owen, C. Familial myelodysplastic syndromes: a review of the literature. *Haematologica* **96,** 1536–1542 (2011).

39. Mangan, J. K. & Speck, N. A. *RUNX1* mutations in clonal myeloid disorders: from conventional cytogenetics to next generation sequencing, a story 40 years in the making. *Crit. Rev. Oncog.* **16,** 77–91 (2011).

40. Knight, S. W. *et al.* Identification of novel *DKC1* mutations in patients with dyskeratosis congenita: implications for pathophysiology and diagnosis. *Hum. Genet.* **108,** 299–303 (2001).

41. Salowsky, R., Heiss, N. S., Benner, A., Wittig, R. & Poustka, A. Basal transcription activity of the dyskeratosis congenita gene is mediated by Sp1 and Sp3 and a patient mutation in a Sp1 binding site is associated with decreased promoter activity. *Gene* **293,** 9–19 (2002).

42. O'Driscoll, M. *et al.* DNA ligase IV mutations identified in patients exhibiting developmental delay and immunodeficiency. *Mol. Cell* **8,** 1175–1185 (2001).

43. Frank, K. M. *et al.* Late embryonic lethality and impaired V (D)J recombination in mice lacking DNA ligase IV. *Nature* **396,** 173–177 (1998).

44. Gruhn, B. *et al.* Successful bone marrow transplantation in a patient with DNA ligase IV deficiency and bone marrow failure. *Orphanet J. Rare Dis.* **2,** 5 (2007).

45. Toita, N. *et al.* Epstein–Barr virus-associated B-cell lymphoma in a patient with DNA ligase IV (LIG4) syndrome. *Am. J. Med. Genet. A.* **143A,** 742–745 (2007).

46. Bacon, C. M. *et al.* Epstein-Barr virus–independent diffuse large B-cell lymphoma in DNA ligase 4 deficiency. *J. Allergy Clin. Immunol.* **131,** 1237–1239.e1 (2013).

47. Riballo, E. *et al.* Identification of a defect in DNA ligase IV in a radiosensitive leukaemia patient. *Curr. Biol.* **9,** 699–S2 (1999).

48. Yamaguchi, H. *et al.* Mutations of the human telomerase RNA gene (*TERC*) in aplastic anemia and myelodysplastic syndrome. *Blood* **102,** 916–918 (2003).

49. Yamaguchi, H. *et al.* Mutations in *TERT*, the gene for telomerase reverse transcriptase, in aplastic anemia. *N. Engl. J. Med.* **352,** 1413–1424 (2005).

50. Myers, K. C. *et al.* Variable clinical presentation of Shwachman-Diamond syndrome: update from the North American Shwachman-Diamond Syndrome Registry. *J. Pediatr.* **164,** 866–870 (2014).

51. Puchalka, J., Kohistani, N. & Klein, C. Novel NGS-based platforms for molecular diagnosis of severe congenital neutropenia. *Blood* **122,** 1034–1034 (2013).

52. Chandrasekharappa, S. C. *et al.* Massively parallel sequencing, aCGH, and RNA-Seq technologies provide a comprehensive molecular diagnosis of Fanconi anemia. *Blood* **121,** e138–e148 (2013).

53. Farrar, J. E. *et al.* Ribosomal protein gene deletions in Diamond-Blackfan anemia. *Blood* **118,** 6943–6951 (2011).

54. Quarello, P. *et al.* High frequency of ribosomal protein gene deletions in Italian Diamond-Blackfan anemia patients detected by multiplex ligation-dependent probe amplification assay. *Haematologica* **97,** 1813–1817 (2012).

55. Béri-Dexheimer, M. *et al.* Clinical phenotype of germline *RUNX1* haploinsufficiency: from point mutations to large genomic deletions. *Eur. J. Hum. Genet.* **16,** 1014–1018 (2008).

56. Morgan, N. V., Tipping, A. J., Joenje, H. & Mathew, C. G. High frequency of large intragenic deletions in the Fanconi anemia group A gene. *Am. J. Hum. Genet.* **65,** 1330–1341 (1999).

57. Castella, M. *et al.* Origin, functional role, and clinical impact of Fanconi anemia *FANCA* mutations. *Blood* **117,** 3759–3769 (2011).

58. Smith, M. L., Cavenagh, J. D., Lister, T. A. & Fitzgibbon, J. Mutation of *CEBPA* in familial acute myeloid leukemia. *N. Engl. J. Med.* **351,** 2403–2407 (2004).

59. Hahn, C. N. *et al.* Heritable *GATA2* mutations associated with familial myelodysplastic syndrome and acute myeloid leukemia. *Nat. Genet.* **43,** 1012–1017 (2011).

60. Noris, P. *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).

61. Kirwan, M. *et al.* Exome sequencing identifies autosomal-dominant *SRP72* mutations associated with familial aplasia and myelodysplasia. *Am. J. Hum. Genet.* **90,** 888–892 (2012).

62. Shah, S. *et al.* A recurrent germline *PAX5* mutation confers susceptibility to pre-B cell acute lymphoblastic leukemia. *Nat. Genet.* **45,** 1226–1231 (2013).

63. Auer, F. *et al.* Inherited susceptibility to pre B-ALL caused by germline transmission of *PAX5* c.547G>A. *Leukemia* **28,** 1136–1138 (2014).

64. Holmfeldt, L. *et al.* The genomic landscape of hypodiploid acute lymphoblastic leukemia. *Nat. Genet.* **45,** 242–252 (2013).

65. Powell, B. C. *et al.* Identification of *TP53* as an acute lymphocytic leukemia susceptibility gene through exome sequencing. *Pediatr. Blood Cancer* **60,** E1–E3 (2013).

66. Zhang, M. Y. *et al.* Genomic analysis of bone marrow failure and myelodysplastic syndromes reveals phenotypic and diagnostic complexity. *Haematologica* (2014). doi:10.3324/haematol.2014.113456

67. De, S. *et al.* Steric mechanism of auto-inhibitory regulation of specific and non-specific DNA binding by the ETS transcriptional repressor ETV6. *J. Mol. Biol.* doi:10.1016/j.jmb.2013.11.031

68. Biasini, M. *et al.* SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* **42,** W252–258 (2014).

69. Green, S. M., Coyne, H. J., 3rd, 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).

70. Coyne, H. J. *et al.* Autoinhibition of ETV6 (TEL) DNA binding: appended helices sterically block the ETS domain. *J. Mol. Biol.* **421,** 67–84 (2012).

71. Chakrabarti, S. R. & Nucifora, G. The leukemia-associated gene *TEL* encodes a transcription repressor which associates with SMRT and mSin3A. *Biochem. Biophys. Res. Commun.* **264,** 871–877 (1999).

72. Park, H., Seo, Y., Kim, J. I., Kim, W. & Choe, S. Y. Identification of the nuclear localization motif in the ETV6 (TEL) protein. *Cancer Genet. Cytogenet.* **167,** 117–121 (2006).

73. Fenrick, R. *et al.* Both TEL and AML-1 contribute repression domains to the t(12;21) fusion protein. *Mol. Cell. Biol.* **19,** 6566–6574 (1999).

74. 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).

75. Lopez, R. G. *et al.* TEL is a sequence-specific transcriptional repressor. *J. Biol. Chem.* **274,** 30132–30138 (1999).

76. 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).

77. Kim, C. A. *et al.* Polymerization of the SAM domain of TEL in leukemogenesis and transcriptional repression. *EMBO J.* **20,** 4173–4182 (2001).

78. Wang, L. C. *et al.* The *TEL/ETV6* gene is required specifically for hematopoiesis in the bone marrow. *Genes Dev.* **12,** 2392–2402 (1998).

79. Hock, H. *et al.* Tel/Etv6 is an essential and selective regulator of adult hematopoietic stem cell survival. *Genes Dev.* **18,** 2336–2341 (2004).

80. Walsh, T. *et al.* Whole exome sequencing and homozygosity mapping identify mutation in the cell polarity protein GPSM2 as the cause of nonsyndromic hearing loss DFNB82. *Am. J. Hum. Genet.* **87,** 90–94 (2010).

81. Mostoslavsky, G., Fabian, A. J., Rooney, S., Alt, F. W. & Mulligan, R. C. Complete correction of murine Artemis immunodeficiency by lentiviral vector-mediated gene transfer. *Proc. Natl. Acad. Sci.* **103,** 16406–16411 (2006).

82. Delaney, C., Varnum-Finney, B., Aoyama, K., Brashem-Stein, C. & Bernstein, I. D. Dose-dependent effects of the Notch ligand Delta1 on ex vivo differentiation and in vivo marrow repopulating ability of cord blood cells. *Blood* **106,** 2693–2699 (2005).

83. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinforma. Oxf. Engl.* **25,** 1105–1111 (2009).

84. Anders, S., Pyl, P. T. & Huber, W. HTSeq - A Python framework to work with high-throughput sequencing data. *Bioinforma. Oxf. Engl.* (2014). doi:10.1093/bioinformatics/btu638

85. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinforma. Oxf. Engl.* **26,** 139–140 (2010).

86. Reiner, A., Yekutieli, D. & Benjamini, Y. Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinforma. Oxf. Engl.* **19,** 368–375 (2003).

87. Yeung, K. Y., Haynor, D. R. & Ruzzo, W. L. Validating clustering for gene expression data. *Bioinforma. Oxf. Engl.* **17,** 309–318 (2001).

88. Saeed, A. I. *et al.* TM4: a free, open-source system for microarray data management and analysis. *BioTechniques* **34,** 374–378 (2003).

89. Young, M. D., Wakefield, M. J., Smyth, G. K. & Oshlack, A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* **11,** R14 (2010).

90. Burwick, N., Coats, S. A., Nakamura, T. & Shimamura, A. Impaired ribosomal subunit association in Shwachman-Diamond syndrome. *Blood* **120,** 5143–5152 (2012).

91. Griesinger, F., Janke, A., Podleschny, M. & Bohlander, S. K. Identification of an *ETV6-ABL2* fusion transcript in combination with an *ETV6* point mutation in a T-cell acute lymphoblastic leukaemia cell line. *Br. J. Haematol.* **119,** 454–458 (2002).

92. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487,** 330–337 (2012).

93. Seshagiri, S. *et al.* Recurrent R-spondin fusions in colon cancer. *Nature* **488,** 660–664 (2012).

94. Zhang, J. *et al.* Key pathways are frequently mutated in high-risk childhood acute lymphoblastic leukemia: a report from the Children's Oncology Group. *Blood* **118,** 3080–3087 (2011).

95. Padron, E. *et al. ETV6* and signaling gene mutations are associated with secondary transformation of myelodysplastic syndromes to chronic myelomonocytic leukemia. *Blood* **123,** 3675–3677 (2014).

96. Suraweera, N. *et al.* Mutations of the PU.1 Ets domain are specifically associated with murine radiation-induced, but not human therapy-related, acute myeloid leukaemia. *Oncogene* **24,** 3678–3683 (2005).

97. Gunji, H. *et al.* TEL/AML1 shows dominant-negative effects over TEL as well as AML1. *Biochem. Biophys. Res. Commun.* **322,** 623–630 (2004).

98. Vlierberghe, P. V. *et al. ETV6* mutations in early immature human T cell leukemias. *J. Exp. Med.* **208,** 2571–2579 (2011).

99. Maki, K. *et al.* Leukemia-Related Transcription Factor TEL Is Negatively Regulated through Extracellular Signal-Regulated Kinase-Induced Phosphorylation. *Mol. Cell. Biol.* **24,** 3227–3237 (2004).

100. Arai, H. *et al.* Functional regulation of TEL by p38-induced phosphorylation. *Biochem. Biophys. Res. Commun.* **299,** 116–125 (2002).

101. Lu, K. P., Liou, Y. C. & Zhou, X. Z. Pinning down proline-directed phosphorylation signaling. *Trends Cell Biol.* **12,** 164–172 (2002).

102. Hodis, E. *et al.* A Landscape of Driver Mutations in Melanoma. *Cell* **150,** 251–263 (2012).

103. Forbes, S. A. *et al.* COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.* **39,** D945–D950 (2011).

104. Marquez, R. *et al.* A new family with a germline *ANKRD26* mutation and predisposition to myeloid malignancies. *Leuk. Lymphoma* (2014). doi:10.3109/10428194.2014.903476

105. Novershtern, N. *et al.* Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell* **144,** 296–309 (2011).

106. Bluteau, D. *et al.* Thrombocytopenia-associated mutations in the *ANKRD26* regulatory region induce MAPK hyperactivation. *J. Clin. Invest.* **124,** 580–591 (2014).

107. Welch, J. S. *et al.* The origin and evolution of mutations in acute myeloid leukemia. *Cell* **150,** 264–278 (2012).

108. Walter, M. J. *et al.* Clonal diversity of recurrently mutated genes in myelodysplastic syndromes. *Leukemia* **27,** 1275–1282 (2013).

109. Zhang, J. *et al.* The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. *Nature* **481,** 157–163 (2012).

110. Bejar, R. *et al.* Clinical effect of point mutations in myelodysplastic syndromes. *N. Engl. J. Med.* **364,** 2496–2506 (2011).

111. Xu, L. *et al.* Genomic landscape of CD34+ hematopoietic cells in myelodysplastic syndrome and gene mutation profiles as prognostic markers. *Proc. Natl. Acad. Sci.* 201407688 (2014). doi:10.1073/pnas.1407688111

112. Ding, L. *et al.* Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* **481,** 506–510 (2012).

113. Yoshida, K. *et al.* Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* **478,** 64–69 (2011).

114. Dolnik, A. *et al.* Commonly altered genomic regions in acute myeloid leukemia are enriched for somatic mutations involved in chromatin remodeling and splicing. *Blood* **120,** e83–92 (2012).

115. Wang, Q. *et al. ETV6* mutation in a cohort of 970 patients with hematologic malignancies. *Haematologica* haematol.2014.104406 (2014). doi:10.3324/haematol.2014.104406

116. Lohr, J. G. *et al.* Widespread Genetic Heterogeneity in Multiple Myeloma: Implications for Targeted Therapy. *Cancer Cell* **25,** 91–101 (2014).

117. Golub, T. R. *et al.* Fusion of the *TEL* gene on 12p13 to the *AML1* gene on 21q22 in acute lymphoblastic leukemia. *Proc. Natl. Acad. Sci.* **92,** 4917–4921 (1995).

118. Noetzli, L. *et al.* Germline mutations in ETV6 are associated with thrombocytopenia, red cell macrocytosis and predisposition to lymphoblastic leukemia. *Nat. Genet.* **advance online publication,** (2015).

119. Topka, S. *et al.* Germline ETV6 Mutations Confer Susceptibility to Acute Lymphoblastic Leukemia and Thrombocytopenia. *PLoS Genet.* **11,** e1005262 (2015).

120. Genovese, P. *et al.* Targeted genome editing in human repopulating haematopoietic stem cells. *Nature* **advance online publication,** (2014).

121. Chronopoulou, E. G. & Labrou, N. E. Site-saturation mutagenesis: a powerful tool for structure-based design of combinatorial mutation libraries. *Curr. Protoc. Protein Sci. Editor. Board John E Coligan Al* **Chapter 26,** Unit 26.6 (2011).

122. Findlay, G. M., Boyle, E. A., Hause, R. J., Klein, J. C. & Shendure, J. Saturation editing of genomic regions by multiplex homology-directed repair. *Nature* **513,** 120–123 (2014).

123. Mandal, P. K. *et al.* Efficient Ablation of Genes in Human Hematopoietic Stem and Effector Cells using CRISPR/Cas9. *Cell Stem Cell* **15,** 643–652 (2014).

124. Jedlicka, P. & Gutierrez-Hartmann, A. Ets transcription factors in intestinal morphogenesis, homeostasis and disease. *Histol. Histopathol.* **23,** 1417–1424 (2008).

125. Rasighaemi, P., Onnebo, S. M. N., Liongue, C. & Ward, A. C. ETV6 (TEL1) regulates embryonic hematopoiesis in zebrafish. *Haematologica* **100,** 23–31 (2015).

126. De Braekeleer, E. *et al. ETV6* fusion genes in hematological malignancies: a review. *Leuk. Res.* **36,** 945–961 (2012).

127. Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. *Nature* **474,** 609–615 (2011).

128. Biankin, A. V. *et al.* Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. *Nature* **491,** 399–405 (2012).

129. Boocock, G. R. B. *et al.* Mutations in *SBDS* are associated with Shwachman-Diamond syndrome. *Nat. Genet.* **33,** 97–101 (2003).

130. Shimamura, A. *et al.* A novel diagnostic screen for defects in the Fanconi anemia pathway. *Blood* **100,** 4649–4654 (2002).

131. Kowalski, K., Czolij, R., King, G. F., Crossley, M. & Mackay, J. P. The solution structure of the N-terminal zinc finger of GATA-1 reveals a specific binding face for the transcriptional co-factor FOG. *J. Biomol. NMR* **13,** 249–262 (1999).

132. Walne, A. J., Vulliamy, T., Beswick, R., Kirwan, M. & Dokal, I. *TINF2* mutations result in very short telomeres: analysis of a large cohort of patients with dyskeratosis congenita and related bone marrow failure syndromes. *Blood* **112,** 3594–3600 (2008).