Appendix

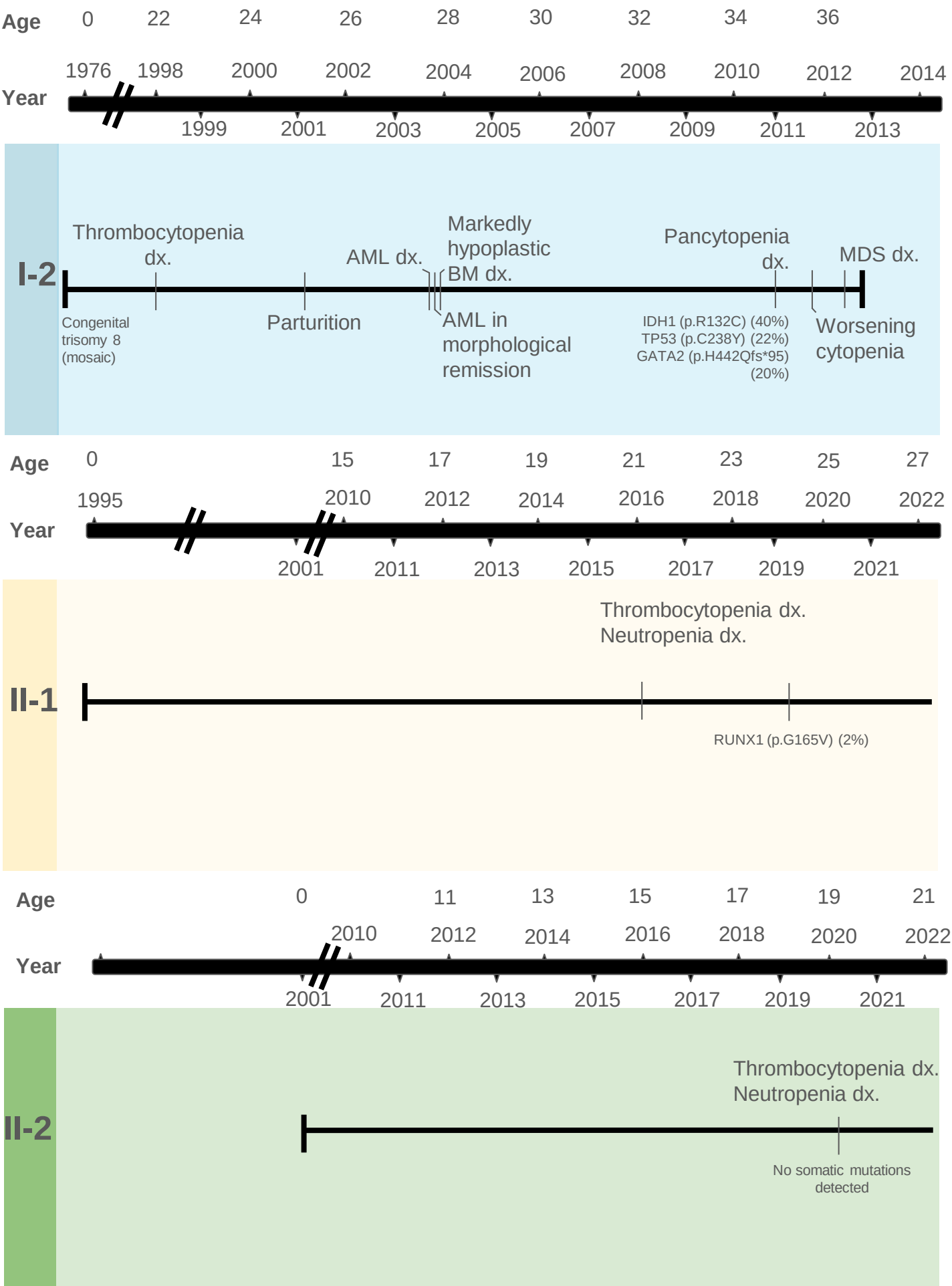
ERG Variants Research Network:

Panagiotis Baliakas[1,2], Piers Blombery[3], Lucy C. Fox[3], Jean Donadieu[4], Ana Rio-Machin[5], Orna Steinberg-Shemer[6,7], Neil V. Morgan[8], Joanne Y.Y. Ngeow[9], Julia Skokowa[10], Mark Pinese[11,12], Mark J. Cowley[11,12]

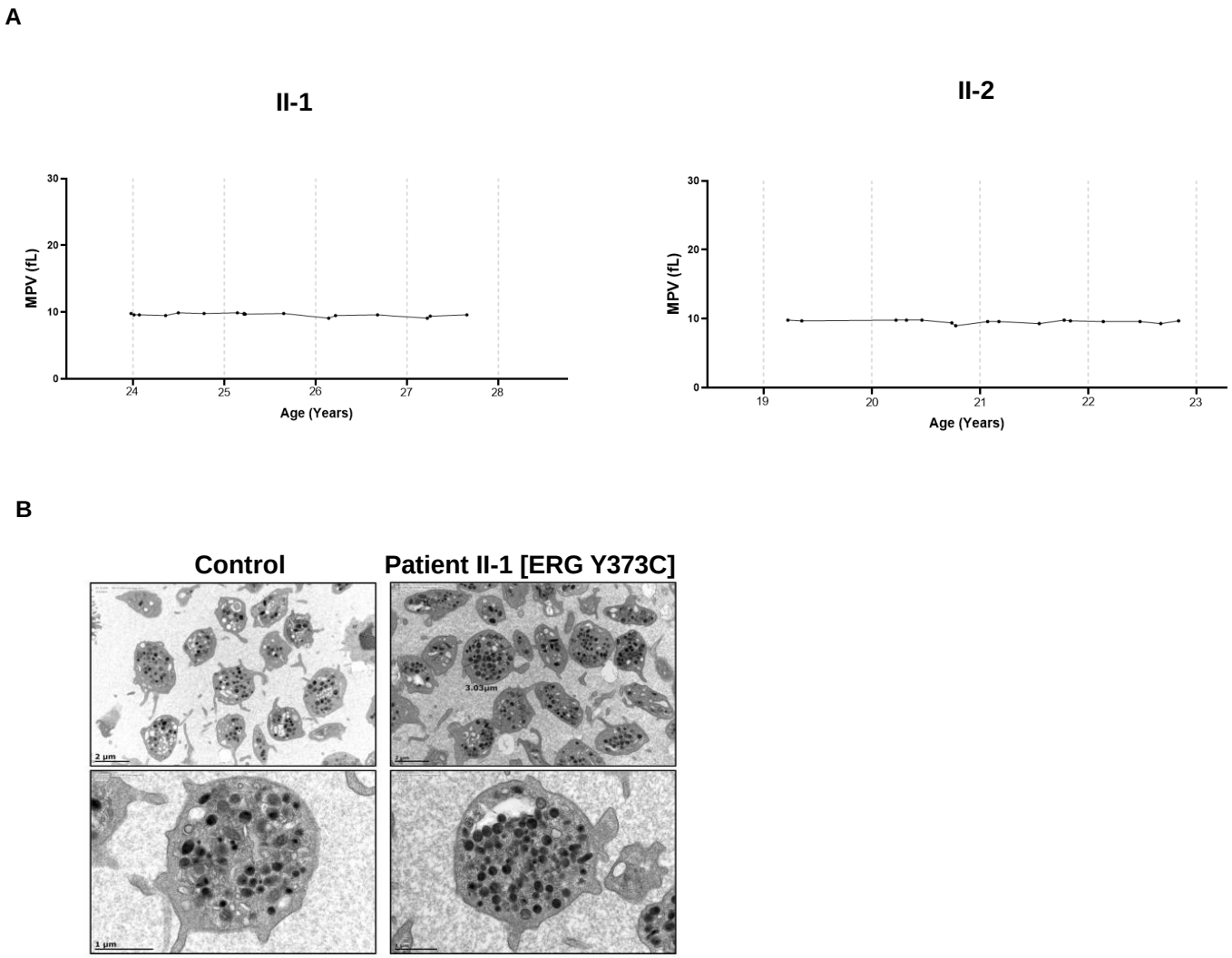
Affiliations:

1. Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden.
2. Department of Clinical Genetics, Uppsala University Hospital, Uppsala, Sweden
3. Peter MacCallum Cancer Centre, Melbourne, VIC, Australia.
4. Centre de Référence des Neutropénies Chroniques, Registre National des Neutropénies Congénitales, Service d'Hémato-oncologie Pédiatrique, Hôpital Armand-Trousseau, Assistance Publique Hôpitaux de Paris, Paris, France.
5. Centre for Haemato-Oncology, Barts Cancer Institute, Queen Mary University of London, London, UK.
6. Pediatric Hematology Research Laboratory, Felsenstein Medical Research Center, Tel-Aviv University, Tel-Aviv, Israel.
7. The Rina Zaizov Pediatric Hematology and Oncology Division Schneider Children's Medical Center of Israel, Petach Tikva, Israel.
8. Institute of Cardiovascular Sciences, College of Medical and Dental Sciences, University of Birmingham, Birmingham, United Kingdom.
9. Division of Medical Oncology, National Cancer Centre Singapore, Singapore, Singapore; Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore; Cancer Genetics Service, National Cancer Centre Singapore, Singapore, Singapore.
10. Department of Hematology, Oncology, Clinical Immunology, and Rheumatology, University Hospital Tübingen, Tübingen, Germany.
11. School of Clinical Medicine, UNSW Medicine & Health, UNSW Sydney, Sydney NSW, Australia.
12. Children's Cancer Institute Australia, Lowy Cancer Research Centre, Sydney, NSW, Australia.

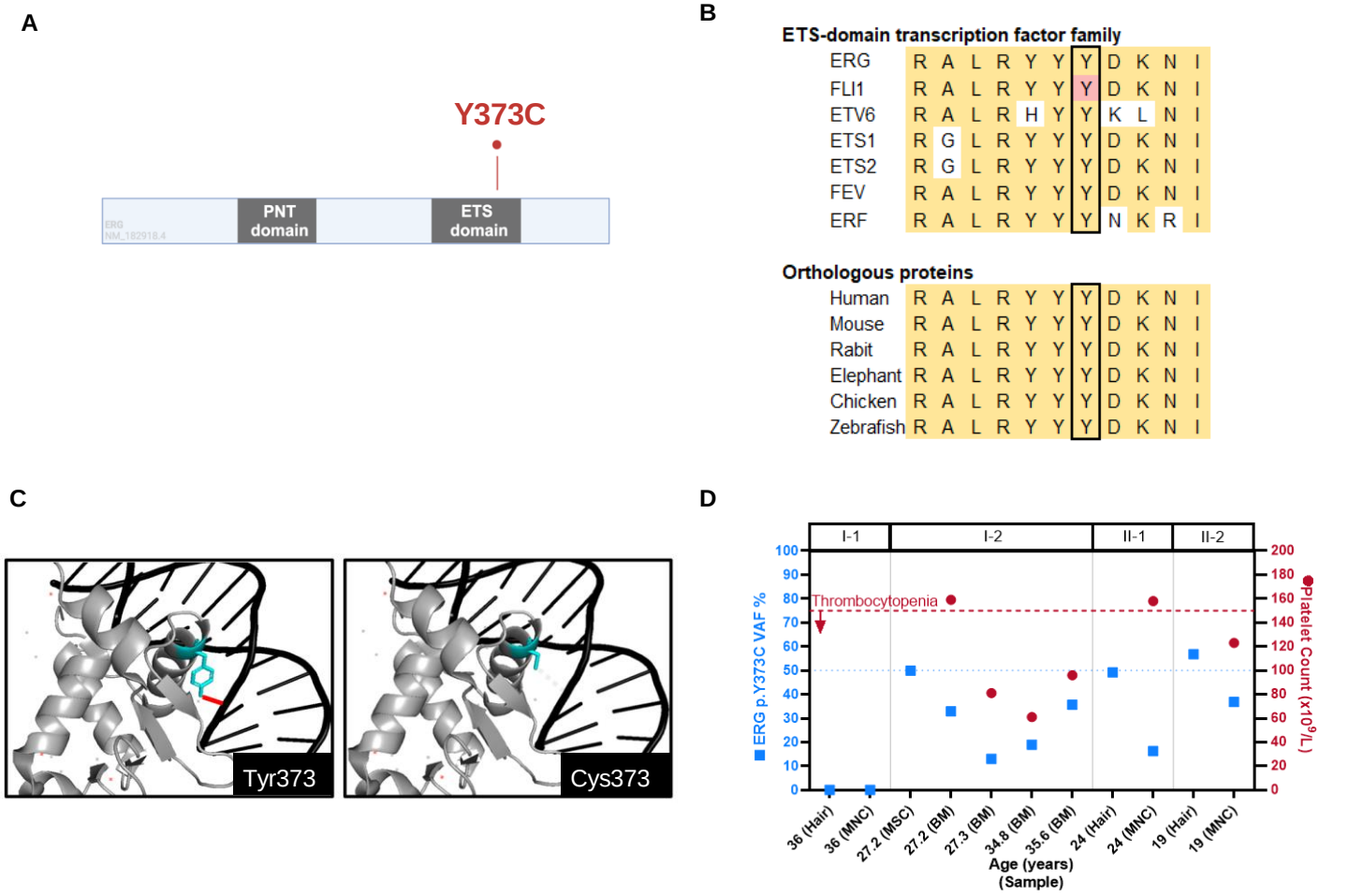
Supplemental Figures



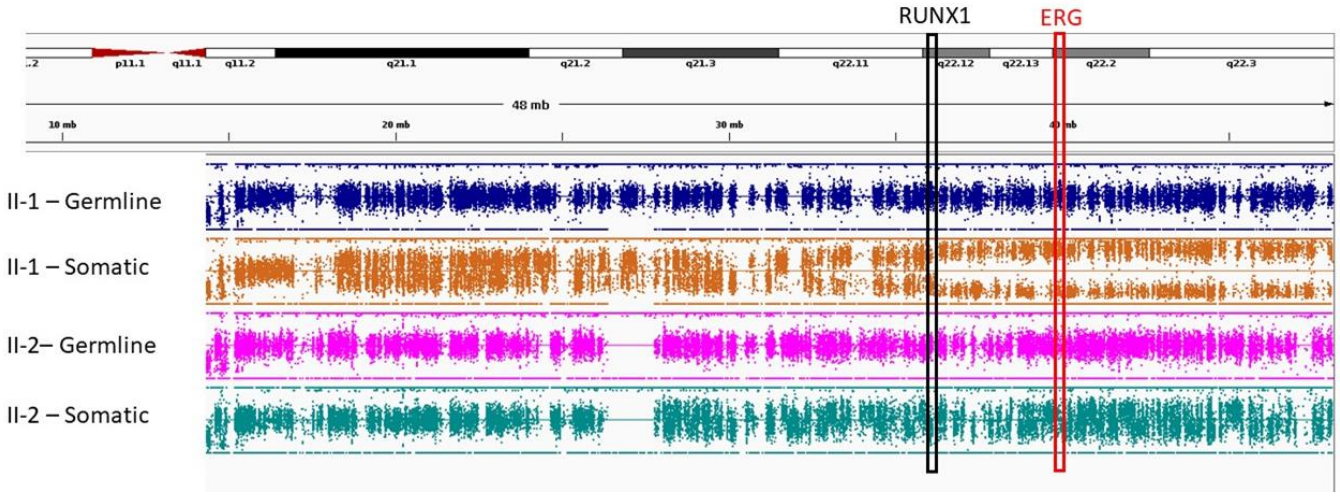
Supplemental Figure 1. Diagnosis timeline for members of Family 1. Patient 15 (blue), 16 (orange), 17 (green).



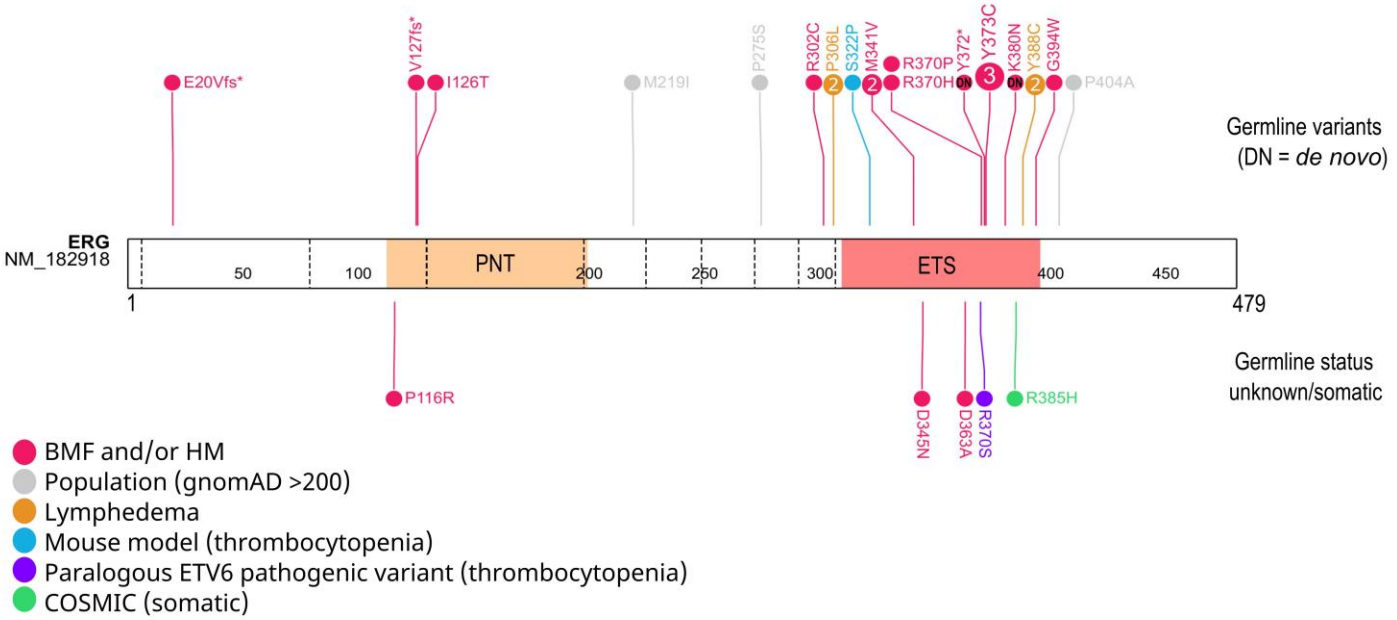
Supplemental Figure 2. Platelet analysis of Family 1. (A) Mean platelet volume of family 1. History of MPV of affected family members (Patient, 16, 17) (II-1, II-2). Mean platelet volume counts in femtoliters complete blood examinations (CBE) plotted with age (years). (B) Transmission electron micrographs of platelets from ERG (Y373C) carrier (II-1) and control. Pictures are indicative of ~200 platelets analyzed.



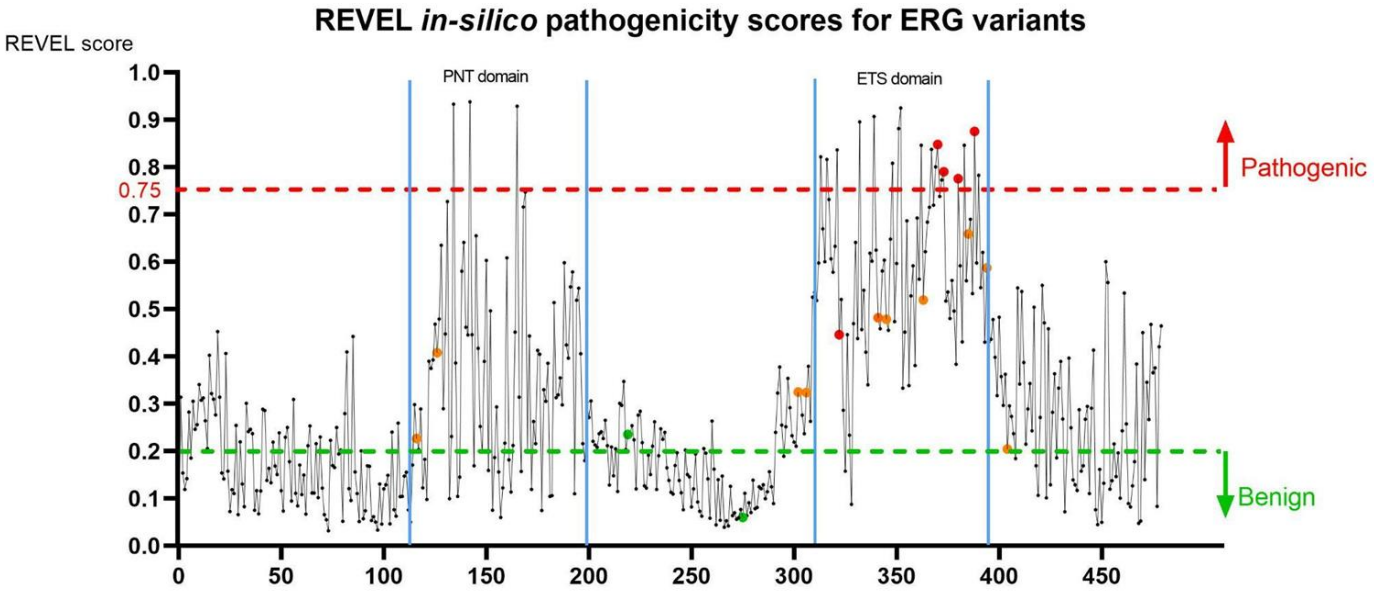
Supplemental Figure 3. Conservation and effect of germline ERG (p.Y373C) variant. (A) p.Y373C variant mapped on ERG protein (NM_182914). (B) Conservation of amino acids surrounding p.Y373C variant of paralogous human ETS transcription factors and orthologous ERG proteins across species. (C) Predicted effect of ERG p.Y373C variant on DNA binding. Predicted polar contacts (hydrogen bonds with solvent excluded) (red line) and predicted polar contact disruption due to mutagenesis (grey dotted line). 3D protein modelling performed on an ERG-DNA X-ray crystallography model (PDB ID: 6VGE A chain) obtained from UNIPROT online database. PyMOL was used to visualise the predicted structural impact of p.Y373C. (D) Comparison of p.Y373C VAF with platelet counts. Droplet digital PCR was used to determine VAFs from Family 1 members' (I-1, I-2, II-1, II-2) (Table 1 - patients 15, 16, 17, respectively) samples at time points shown, and platelet counts for the corresponding time points were also plotted. mononuclear cells (MNC), bone marrow (BM), mesenchymal stromal cells (MSC).



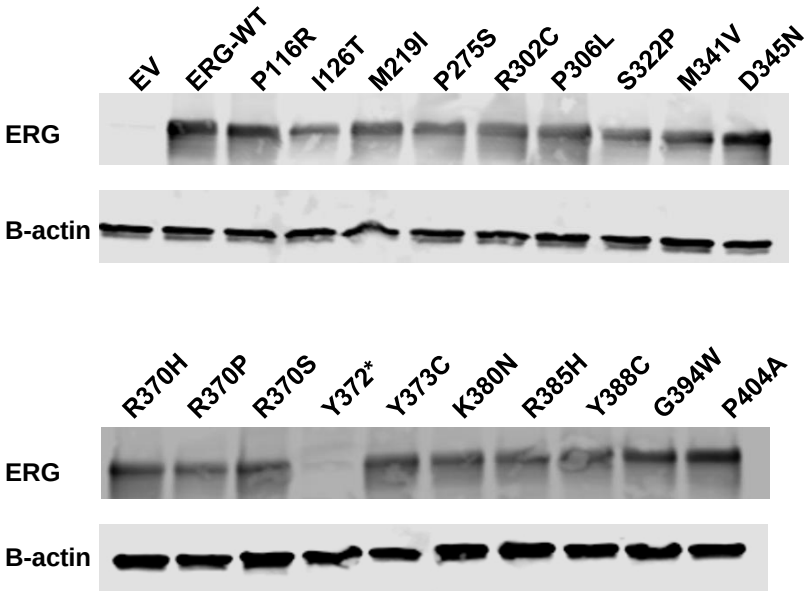
Supplemental Figure 4. B-allele frequencies of Family 1. B-allele plots generated with WGS across chr21 of II-1 and II-2 (Patients 16 and 17). Somatic, peripheral blood mononuclear cells; Germline, hair.



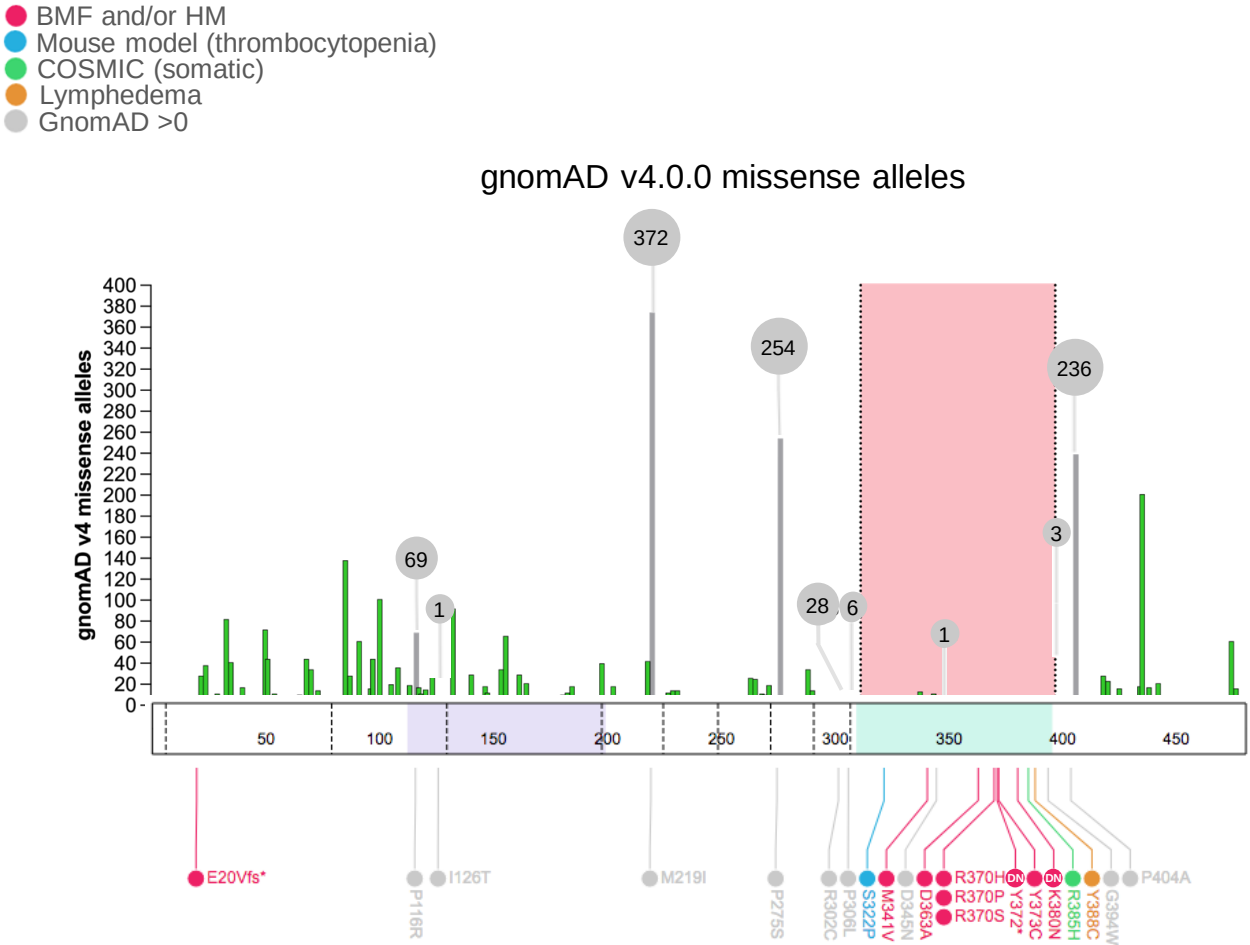
Supplemental Figure 5. *ERG* variants identified in patients with hematological conditions and lymphedema. Rare germline (above) and somatic or unconfirmed germline (below) *ERG* variants from patients carrying hematological-related phenotypes are mapped onto the ERG protein (isoform, NP_891548.1; transcript, NM_182918.4).[38] ETS DNA binding domain (ETS); Pointed domain (PNT); phenotypes and population controls (indicated by color); exon junctions (black dotted lines).



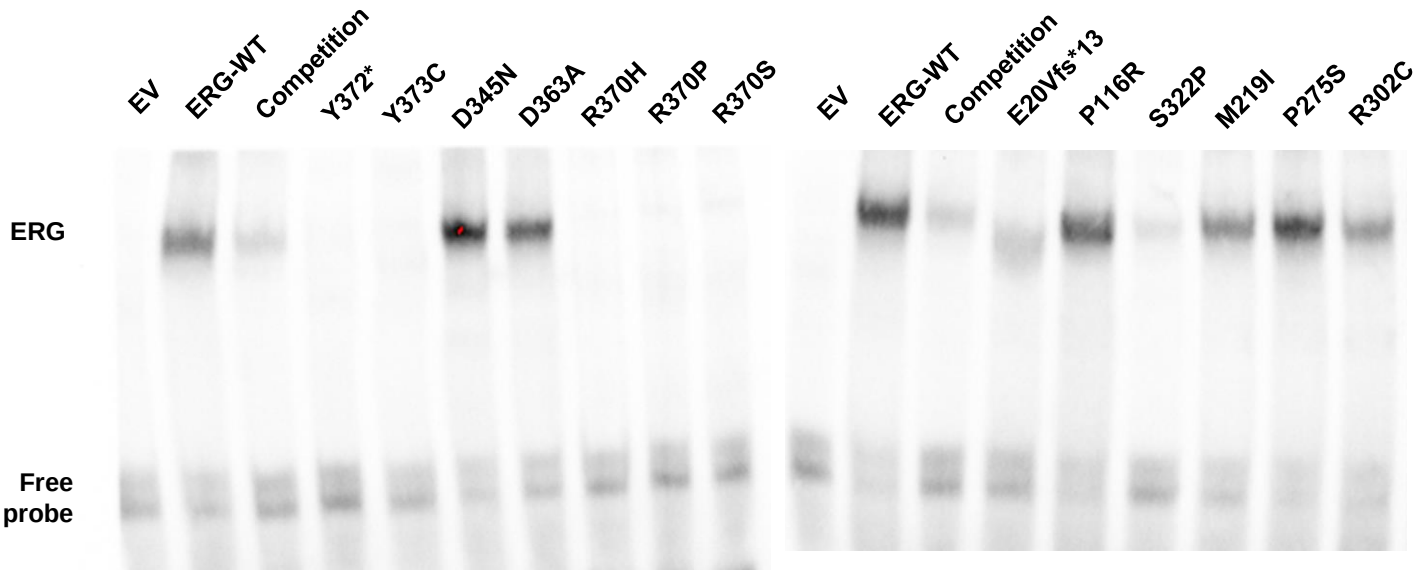
Supplemental Figure 6. ***In silico*** **REVEL pathogenicity prediction scores for** ***ERG*** **variants.** REVEL scores[5] of every possible variation at each amino acid position in NP_891548.1 were averaged and plotted. Average REVEL scores indicating pathogenicity (above red dotted line), average REVEL scores indicating benign residues (below green dotted line), amino acid position of Likely Pathogenic/Pathogenic missense variants from study (red), VUS missense variants from study (orange), Benign missense variants from study (green).



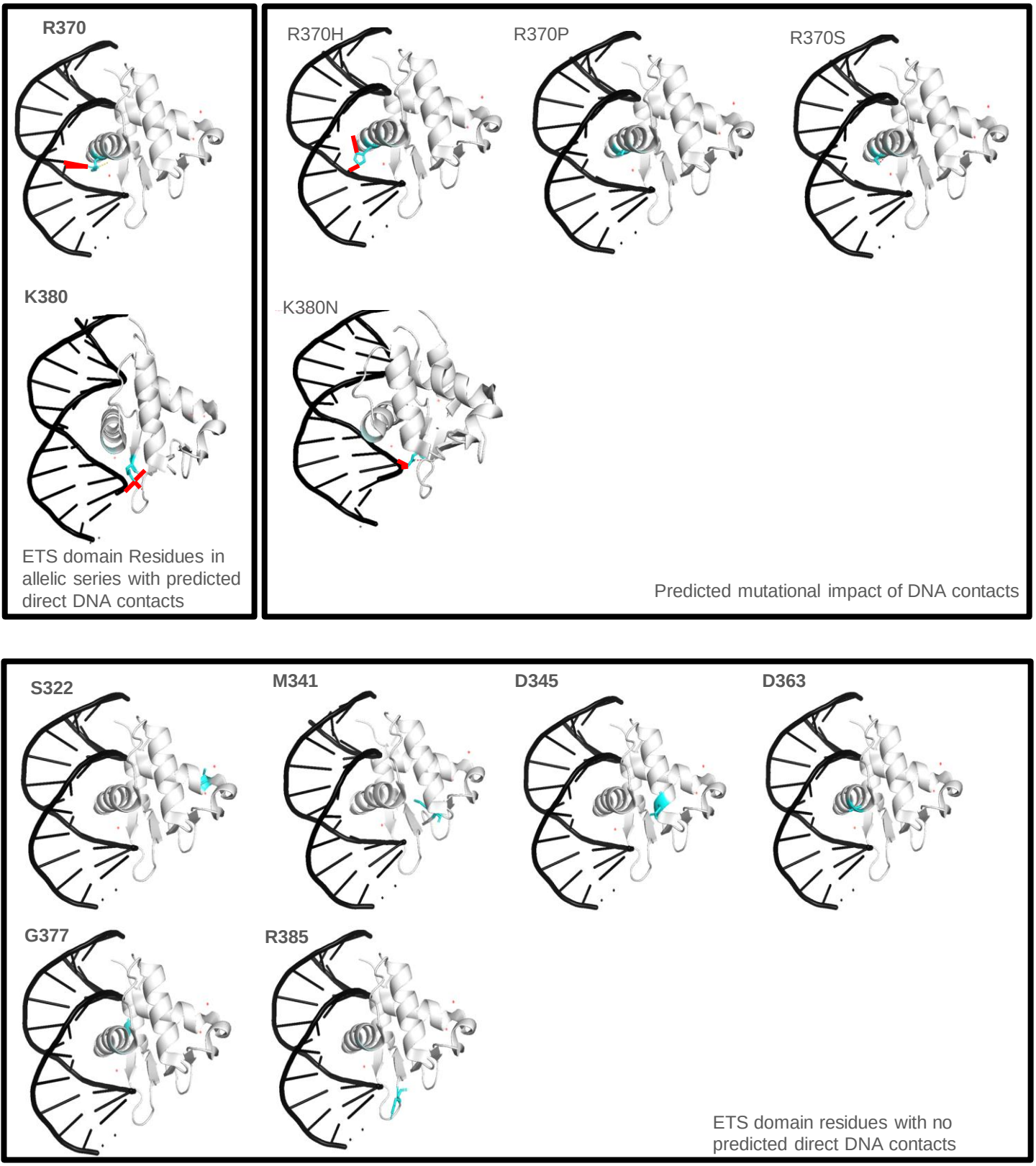
Supplemental Figure 7. Western blot of ERG mutant lysates. Equal amounts of protein were pooled from HEK293 lysates and electrophoresed on SDS-PAGE gels before being transferred to polyvinylidene difluoride membrane. Western blots were performed according to standard protocols probing with anti-ERG (ab92513, Abcam 1:1000) and a fluorescent secondary anti-rabbit antibody (680RD, Immunogen 1:10,000). Signal was detected using the Odyssey Clx infrared imaging system.



Supplemental Figure 8. Frequency of gnomAD missense alleles across ERG. gnomAD missense allele across *ERG* locus (NM_182918). ERG variants tested plotted underneath protein. Variants with gnomAD allele frequency > 0 (grey).

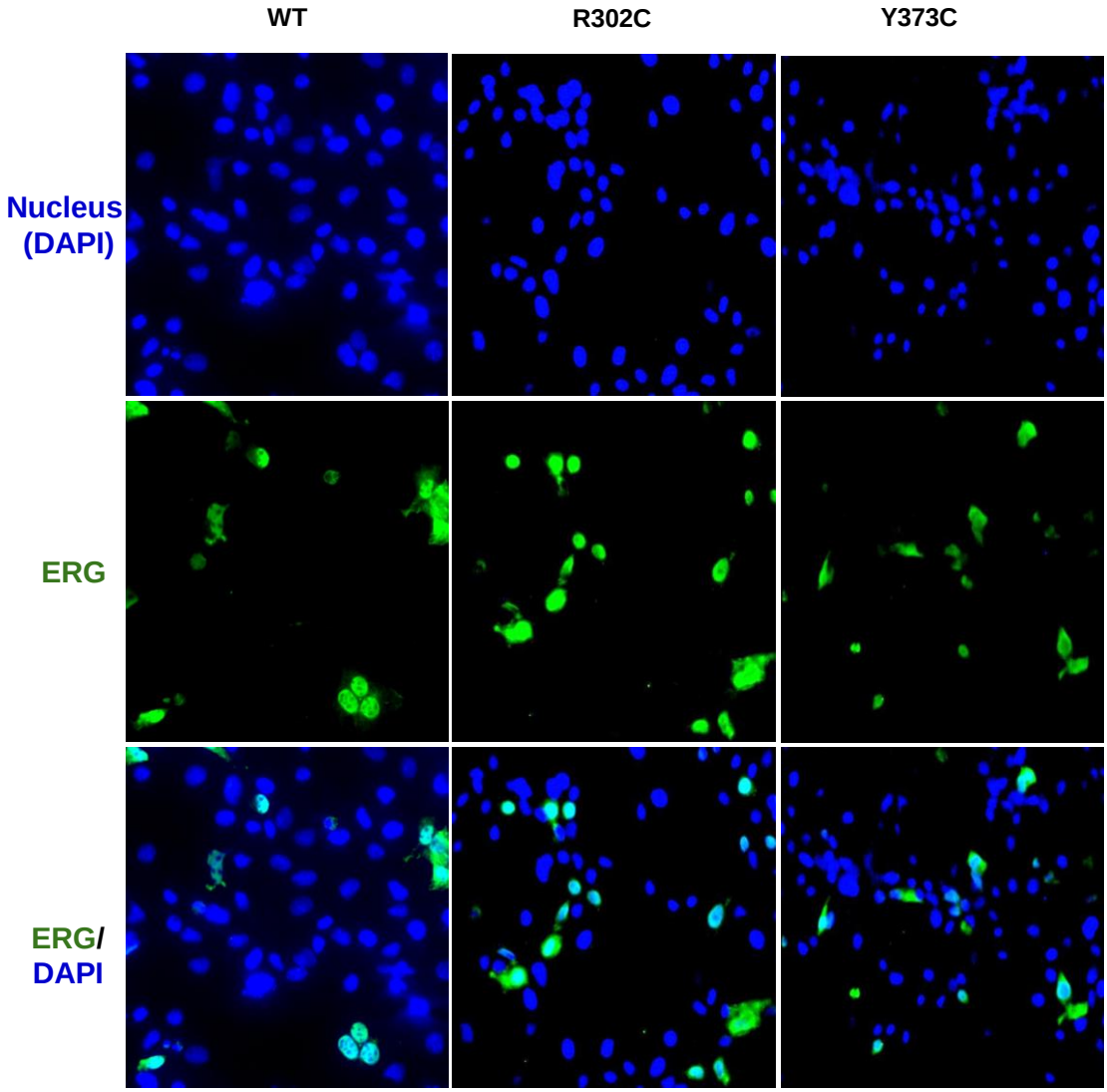


Supplemental Figure 9. ETS domain variants reduce ERG DNA binding affinity. Electrophoretic mobility shift assay (EMSA) of WT and mutant ERG. Whole lysates were prepared and bound to an oligonucleotide containing an ETS DNA consensus sequence. Probes were visualized using chemiluminescence. Competition, 3x unlabelled probe added to reaction. Blots are representative of 3x biological replicates.



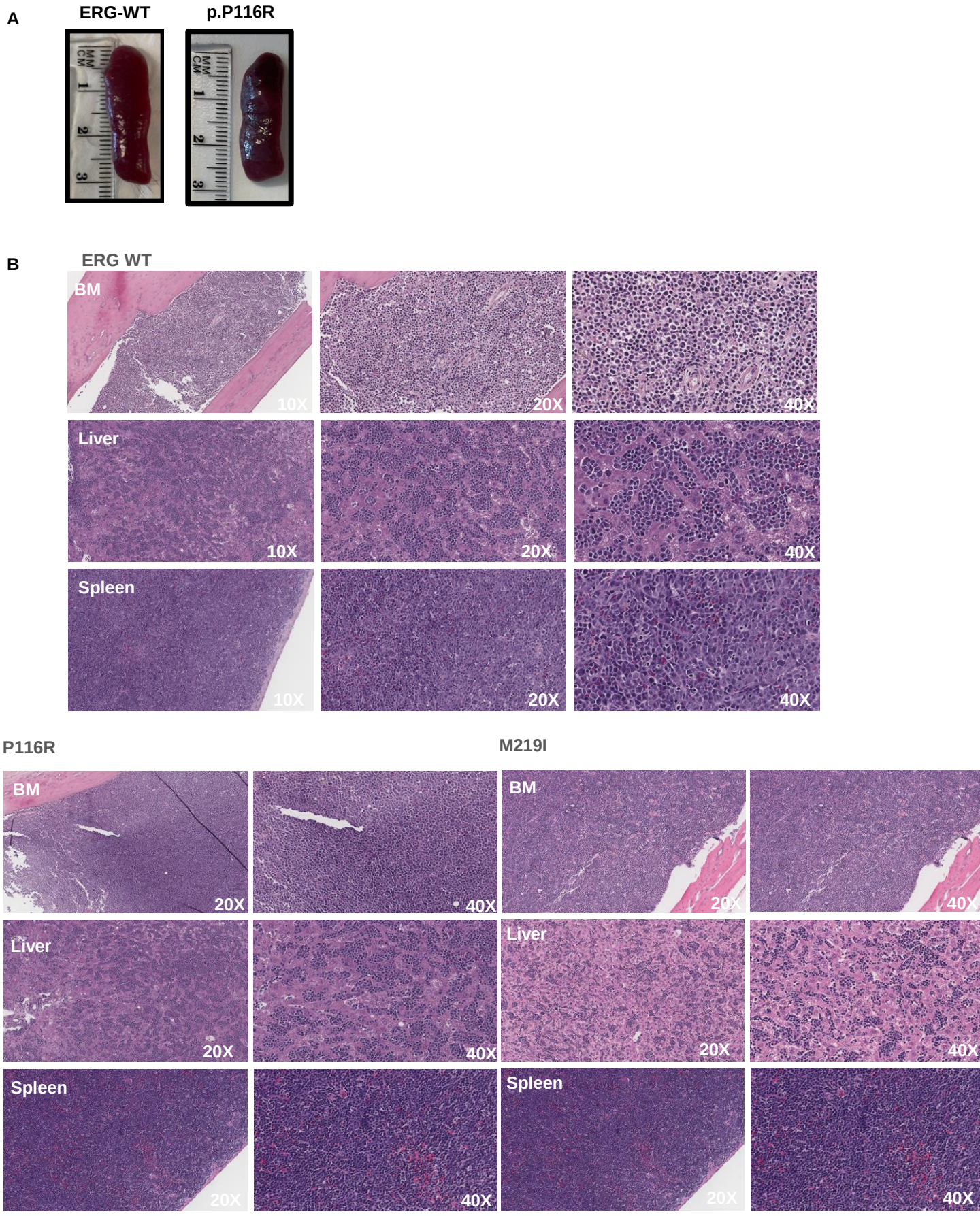
Supplemental Figure 10. Predicted modeling of ETS domain missense variants. Predicted effect of ERG ETS domain variants on DNA binding. Predicted polar contacts (hydrogen bonds with solvent excluded) (red line), DNA (black double helix), indicated amino acid (cyan). 3D protein modelling performed on an ERG-DNA X-ray crystallography model (PDB ID: 6VGE A chain) obtained from UNIPROT online database. PyMOL was used to visualise the predicted structural impact of each variant.

A

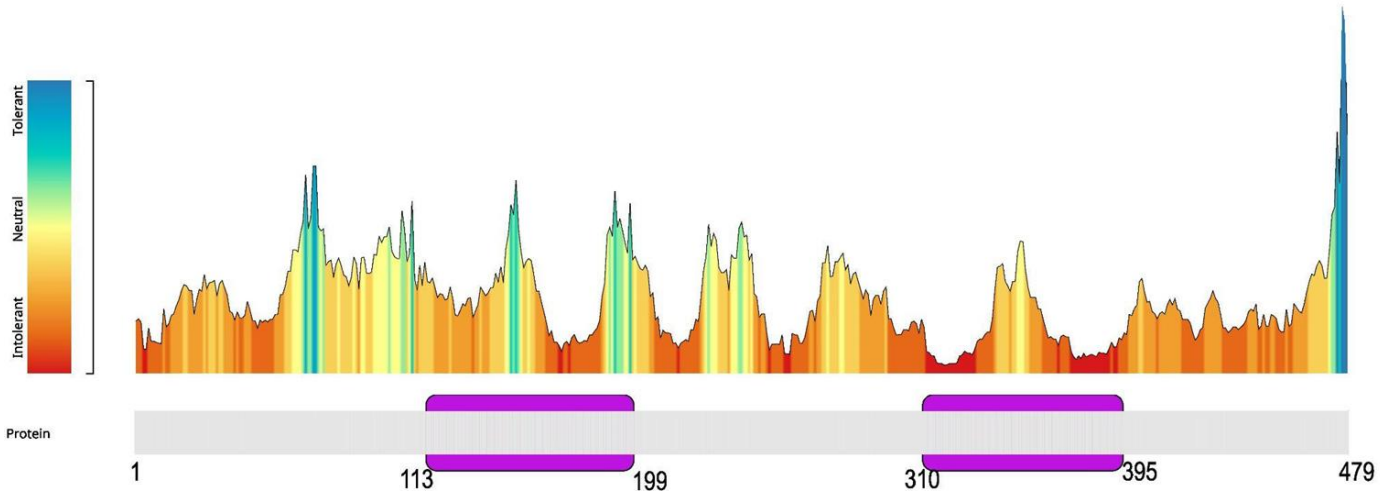


B

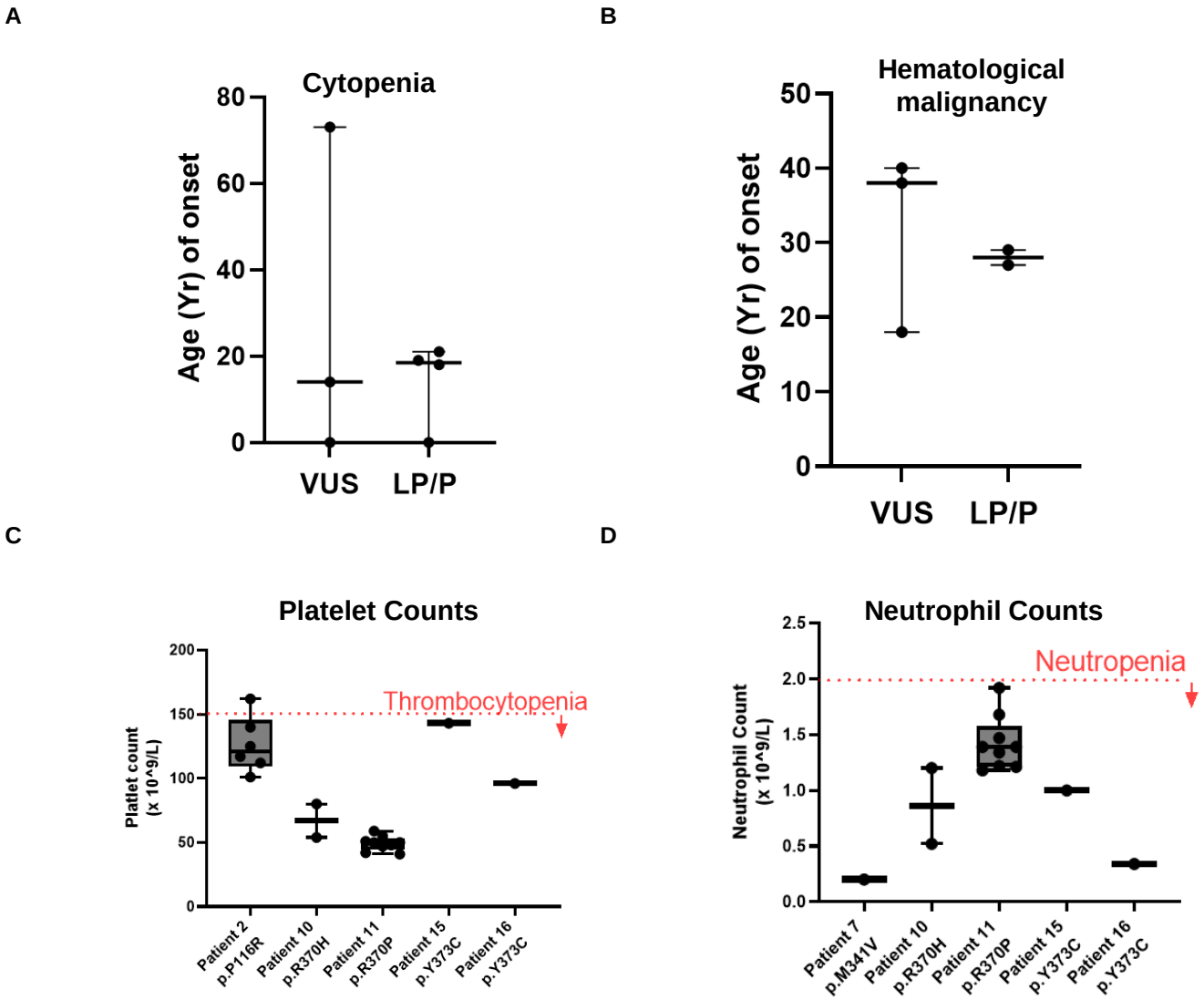
Supplemental Figure 11. Subcellular localization of ERG variants and conservation of critical nuclear localization signal (NLS) amino acids. (A) Immunofluorescence staining of ERG WT and mutants in COS-7 cells. WT and p.R302C are representative of mutants which localize predominantly in the nucleus (*e.g.* p.P116R, p.I126T, p.M219I, p.P275S, p.P306L, p.S322P, p.D345N, p.D363A, p.Y388C, p.G394W, p.P404A) while p.R370S and p.Y373C represent ETS domain variants that localize predominantly to the cytoplasm (*e.g.* p.R370H/P, p.Y373C, p.K380N, p.R385H). (B) Conservation of ETS domain and NLS of ETS transcription factors. Regions of complete homology to ERG (yellow); critical NLS amino acids (black boxes); ClinVar pathogenic variants with a hematological-related phenotype (red shading); residues that impact protein function[6] (pink shading). *ERG* (NM_182918.4), *FLI1* (NM_002017.4), *ETV6* (NM_001987.4), *ETS1* (NM_001143820.1), *ETS2* (NM_005239.5), *FEV* (NM_017521.2), *ERF* (NM_006494.2).



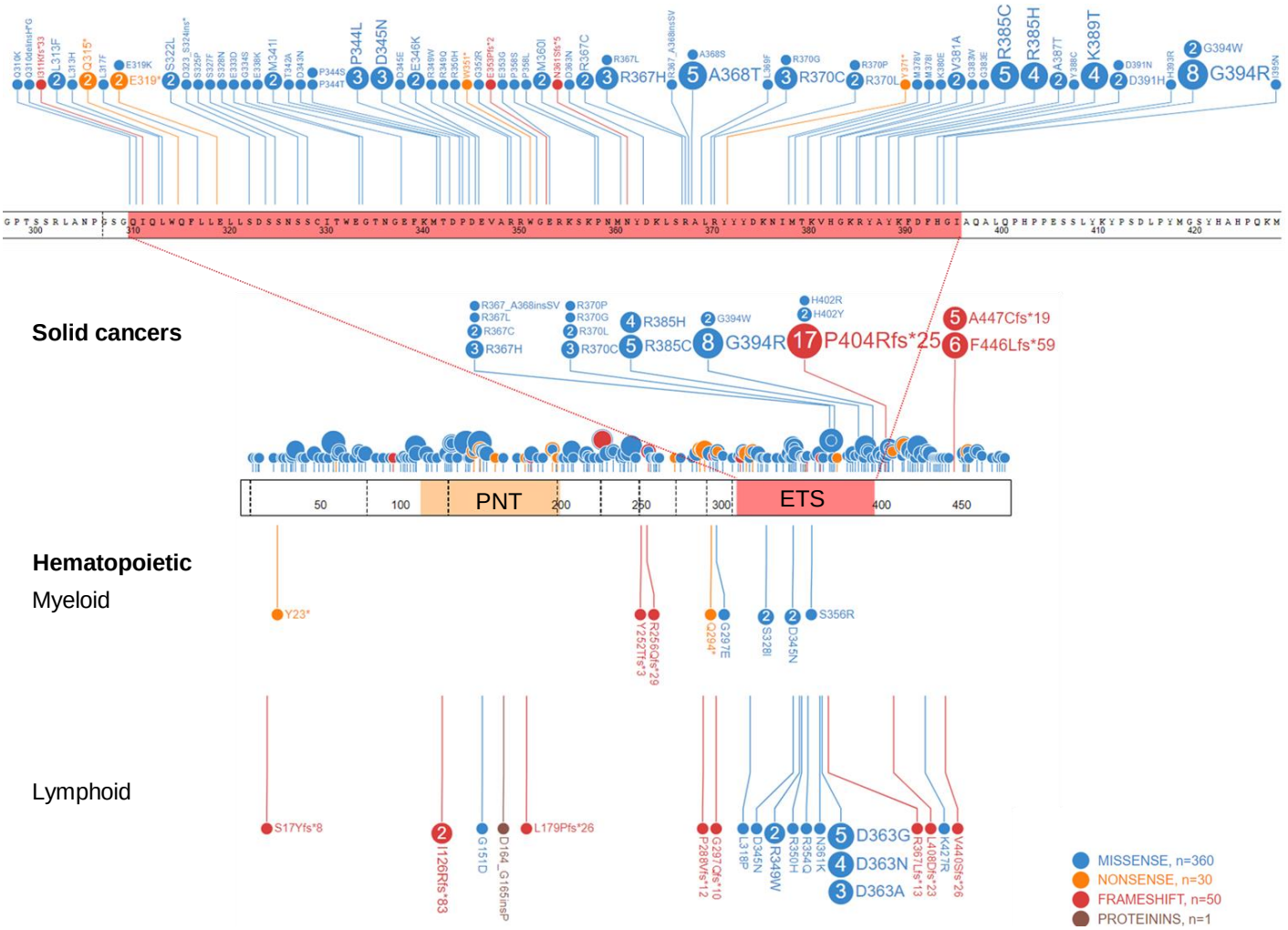
Supplemental Figure 12. Spleen and histology of erythroleukemia mice. (A) Enlarged spleen of ERG-WT and p.P116R (WT-like) mice. Spleen size of these mice is indicative of all erythroleukemic mice. (B) Histological (H&E) staining of bone marrow (BM), liver and spleen from ERG WT, p.P116R and p.M219I leukemic mice. Images are indicative of all erythroleukemic mice.



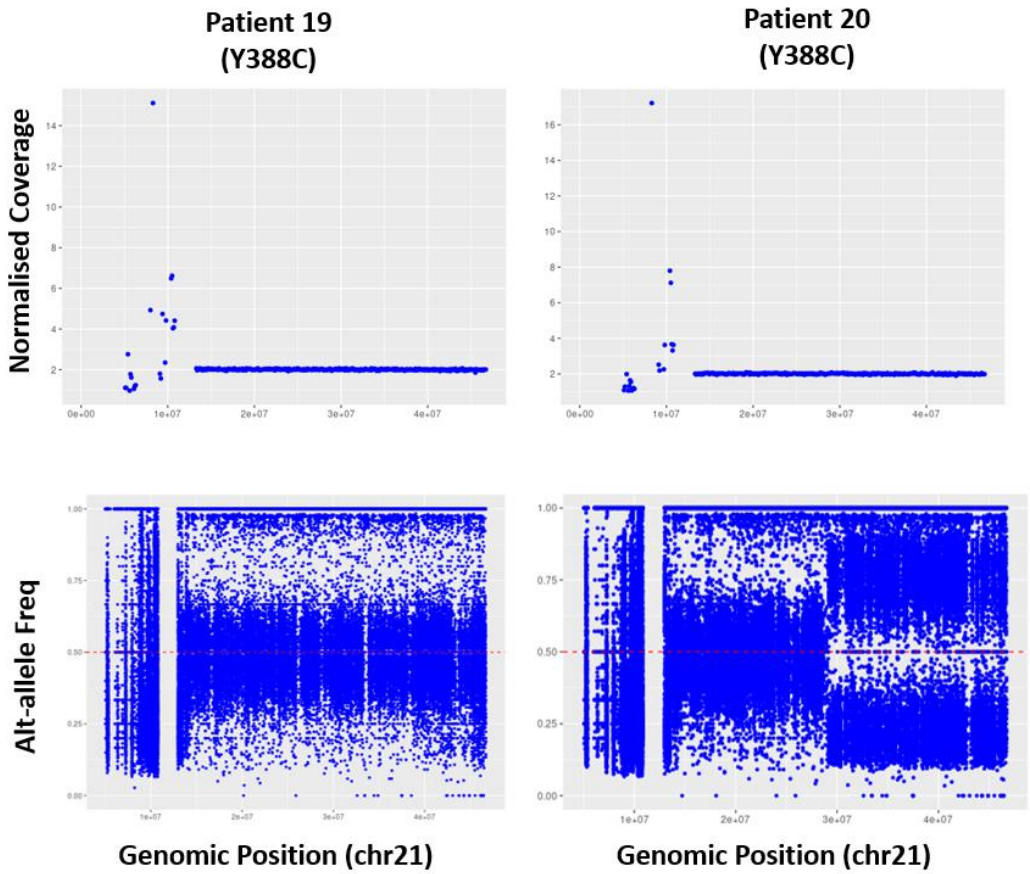
Supplemental Figure 13. Tolerance landscape of ERG protein (NM_182918). Tolerance is computed as a from missense over synonymous variant count ratio, which is calculated in a sliding window manner to provide a per-position indication of regional tolerance to missense variation. The variants are based on gnomAD and corrected for codon composition.[6]



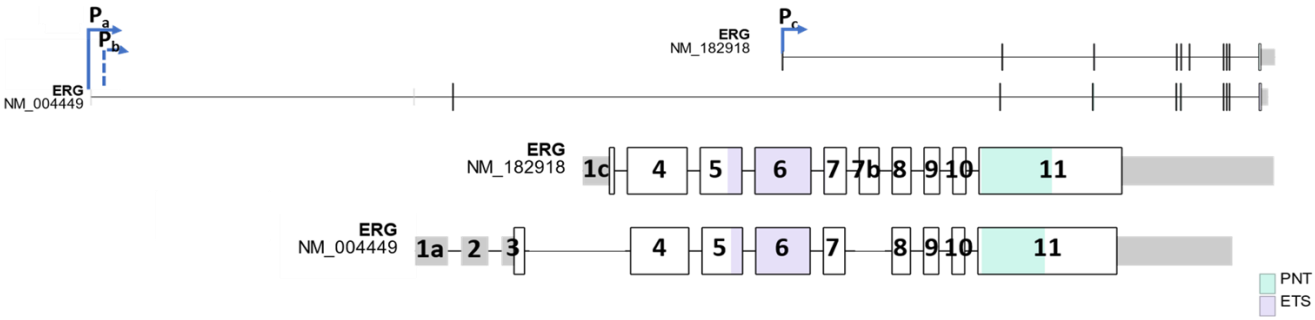
Supplemental Figure 14. Collated clinical data. (A) Age of onset for cytopenia diagnosis separated by ACMG classifications. (B) Age of onset for hematological malignancy diagnosis, separated by ACMG classifications. (C) Longitudinal platelet counts of patients with *ERG* variants. (D) Longitudinal neutrophil counts of patients with *ERG* variants.



Supplemental Figure 15. Analysis of somatic *ERG* mutations in COSMIC. *ERG* variants were extracted from 20,340 cancer samples in COSMIC (v99)[9] and filtered for missense, nonsense, frameshift and inframe indels (PROTEININS) occurring ≤ 3 times in the gnomAD (v4.0.0) population database. Solid cancers (above) and hematological malignancies (below) were separated and plotted in ProteinPaint.[10]



Supplemental Figure 16. Copy Neutral Loss of heterozygosity in *ERG* carriers. Normalized coverage (top panel) and Alt-allele frequency plots (bottom panel) of whole exome sequencing. Patient 20 shows a copy neutral loss of heterozygosity (cnLOH) event in chromosome 21 which encompasses the entire *ERG* gene.



Supplemental Figure 17. *ERG* alternative promoters, transcripts and isoforms. Most common ERG isoforms found in HSPCs are shown.[8]

Supplemental Methods

Whole exome sequencing and analysis. Whole exome sequencing data were aligned to the human reference genome (hg19: human_g1K_v37_decoy) using BWA mem (v.0.7.12). Variant calling was conducted with GATK (v.3) following established best practices. SNVs and small indels were jointly called using GATK HaplotypeCaller (v.3.8) and Joint genotyping was performed using GATK GenotypeGVCF. Structural variants were identified with Manta (v.1.3.1) using paired germline and somatic samples and annotated using AnnotSV[1].

A custom filtering strategy identified *ERG* variants through in-house developed variant analysis software, 'VariantGrid'[2] based on their low population frequency[3] predicted pathogenicity and oncogenicity. Integrative Genomics Viewer4 was used to visualize and validate the variants. Copy number variation (Log R Ratio) and zygosity (B-allele plot) measurements were performed using CytoSNP-850K BeadArray v1.2 (Illumina) and analyzed with BlueFuse multi v4.5 [Human hg19/GRCh37 assembly].

Somatic NGS. Platform: Targeted enrichment of selected coding exons and flanking intronic regions of 46 genes was performed using a custom-designed Haematological neoplasms capture panel manufactured by Integrated DNA Technologies (IDT; HaemV1) and sequenced using 2x150bp reads on the Illumina Sequencing System. This assay achieves >98% of bases with coverage >250x. Sample identification was confirmed by comparison of genotypes determined by NGS and an independent human ID panel (Agena iPLEX Pro Sample ID Panel). This process ensures sample integrity across the multi-step DNA extraction and NGS process.

Bioinformatics: Sequences were aligned to the human reference genome (GRCh38) using Burrows-Wheeler Aligner. Variant calling was performed using Vardict. This method detects single nucleotide variants and small insertion/deletion variants within exons, including splice site regions. The assay sensitivity is specific for each individual variant; however, at a target minimum unique depth of coverage of 250x this assay reliably detects SNV and indels present at 5-10% or greater. Variants detected with VAF <5% are reported where clinically significant. VAF estimates have a measurement of uncertainty of approximately 5%.

Variant classification: Variant annotation was performed using VariantGrid v3 and Alamut v2.15. The following population variation and cancer databases are commonly used to assist with variant interpretation: The Genome Aggregation Database (gnomAD; gnomad.broadinstitute.org), the Catalogue of Somatic Mutations in Cancer (COSMIC; cancer.sanger.ac.uk), cBioPortal (cbioportal.org) and PMKB (pmkb.weill.cornell.edu). Variants were prioritized based on clinical significance utilising data available at the time of reporting. Variants with a population frequency >1% in gnomAD are generally not reported. Variants with a VAF of approximately 50% or 100% and not in COSMIC are generally not reported. Splice predictions are performed in Alamut using several third-party tools (MaxEntScan, NNSPLICE, GeneSplicer, Splice

Cell culture. K562 cells were cultured in ATCC-formulated Iscove's Modified Dulbecco's Medium. HEK293 and COS-7 cells were cultured in Dulbecco's Modified Eagle's Medium. FLCs were cultured in StemSpan SFEM Hematopoietic Cell Culture Medium supplemented with mouse IL-3, IL-6, SCF, FLT-3, TPO cytokines. All mediums were supplemented with 10% FBS and 1% penicillin/streptomycin.

Generation of ERG variant vectors. *ERG* WT and variant expression vectors were derived from pcDNA3-ERG (1508 bp *ERG* cDNA; NM_004449) for luciferase assays and pcDNA3.1-ERG (C-terminal Myc-tag) (1437 bp *ERG* cDNA; NM_182918) for subcellular localization studies and Electrophoretic Mobility Shift Assays (Supplemental Figure 17). Primers were designed (PrimerX online platform) and mutations introduced by site-directed mutagenesis (QuikChange Site-Directed Mutagenesis Kit protocol, Agilent). Constructs were transformed (XL10-Gold Ultracompetent Cells, Agilent) and DNA extracted (Midiprep kit protocol, Qiagen).

Immunofluorescence staining. Wells were coated with Cell-Tak (354240, Corning). COS-7 cells were seeded, transiently transfected (Lipofectamine 2000) and fixed after ~20 h (4% paraformaldehyde). Cells were blocked (5% BSA in TBST) and probed with Anti-Myc antibody (9B11; New England Biolabs, 1:1000) and then Alexa Fluor 488 Rabbit anti-mouse (A27023) secondary antibody. Slides were mounted in Vectashield® mounting medium with DAPI (Vector Laboratories, Inc.) and visualized with an Olympus IX83 microscope. 100 cells were visually quantified and scored as 'nuclear', where protein was only present in the nucleus or 'cytoplasmic ' where protein was present in the cytoplasm and the nucleus.

Transduction. Selected *ERG* variants were generated in the MSCV-ERG-IRES-mCherry vector (NM_182918). HEK293T cells were transfected (Lipofectamine 3000) with ERG WT or variant MSCV-ERG-IRES-mCherry plasmids alongside pCAG-Eco and MSCV-Gag-Pol retroviral packaging vectors. Viral supernatant was harvested 48 hours post transfection and spun onto retronectin coated 12-well non-tissue culture treated plates (1.5 hrs @ 2500RPM). Freshly harvested fetal liver cells were Ter119-depleted using biotin-anti-ter119 antibody (BD Bioscience) followed by magnetic separation with MagniSort™ streptavidin negative selection beads (ThermoFisher Scientific). Ter119-depleted hematopoietic stem/progenitor cells were then plated onto the virus coated retronectin plates overnight before being transferred into culture media.

Droplet digital PCR (ddPCR). A custom ddPCR assay for Y373C was designed using the Bio-Rad algorithm. DNA samples were portioned into nl-sized DNA molecule droplets using a droplet generator (Bio-Rad) to create a water-oil emulsion. Using primers engineered with a fluorescent Taqman probe, each droplet was PCR amplified and analyzed using a 'mutation detection assay' on the QX200 Droplet Digital PCR system (Bio-Rad). QuantaSoft software was used to count PCR-positive and negative droplets which were differentiated by fluorescence colour to provide accurate quantification of ERG (Y373C) VAF.

Western blotting. Equal amounts of protein were pooled and electrophoresed on SDS-PAGE gels before being transferred to polyvinylidene difluoride membrane. Western blots were performed according to standard protocols probing with anti-ERG (ab92513, Abcam 1:1000) and a fluorescent secondary anti-rabbit antibody (680RD, Immunogen 1:10,000). Signal was detected using the Odyssey Clx infrared imaging system.

Supplemental Tables

Supplemental Table 1. Diagnostic panel of Hematology genes

ASXL1	CBL	ETV6	JAK2	NRAS	SF3B1	U2AF1
BCL2	CEBPA	EZH2	KIT	PHF6	SH2B3	WT1
BCOR	CSF3R	FLT3	KRAS	PPM1D	SRSF2	XPO1
BCORL1	CXCR4	GATA1	MPL	PSMB5	STAG2	ZRSR2
BIRC3	DDX41	GATA2	MYD88	PTPN11	STAT5B	
BRAF	DNMT3A	IDH1	NOTCH1	RUNX1	TET2	
CALR	ETNK1	IDH2	NPM1	SETBP1	TP53	

Supplemental Table 2. Rare heterozygous *ERG* variants and population controls. Rare ERG variants from similar phenotypic groups including BMF and/or HM and lymphedema, amongst population controls (gnomAD >200)s, a germline thrombocytopenic mouse mutation1, a paralogous ETV6 pathogenic variant (thrombocytopenia)37 and COSMIC mutation (somatic). Information unavailable (blank).

Class	g.HGVS (GRCh37/hg19) (Chr21)	c.HGVS (NM_004449.4) (ENST00000044248.5)	p.HGVS (NP_004440.1) (ENSP00000394694.1) (462 aa)	g.HGVS (GRCh38/hg38) (Chr21)	c.HGVS (NM_182918.4) (ENST00000288319.12)	p.HGVS (NP_891548.1) (ENSP00000288319.7) (479 aa)	Mouse (CL578L6) (NP_598420.1) (486 aa)	VAF (%)	Gender	Sample analyzed	gnomAD (v 4.0.0)	gnomAD (v2.1.1)	COSMIC (v99)	REVEL	CADD	Hematological-related Phenotype (BMF, HM) (age of onset/phenotype, yr)	Non-hematological phenotype (including lymphedema, other cancers)	Age onset of first phenotype (yr)	Patient outcomes (e.g. HSCT, treatment responses, survival, rate of progression)	Germline ERG (inherited/ *de novo*)	Cytogenetics (SNP array)	Cytogenetics (karyotype)	Blood or BM Flow cytometry	Platlet count (x 10^9/L)	Neutrophil count (x 10^9/L)	Somatic mutations (Myeloid somatic NGS Panel)
Frameshift	g.39817501_ 39817504del	c.80_83del	p.E27Vfs*13	g.38445578_ 38445581del	c.59_62del	p.E20Vfs*13		60	M	MSC	0	0	0	-	-	MDS		38		**Yes**	Normal Karyotype			n/a	n/a	
Missense	g.39795373G>C	c.368C>G	p.P123R	g.38423451G>C	c.347C>G	p.P116R	p.P123R	56	F	PB	69 0.004%	2 (2x P116G)	0 (1x P116Q)	0.36	31	Thrombocytopenia, Thrombocytopathy, platelet aggregation disorder	Hypertension, diabetes, cataract	73			-20q (mosaic) no obvious cnLOH			101 (2015), 117 (2016), 112 (2017), 162 (13.05.19), 140 (23.04.19), 125 (2021)		
Frameshift	g.39795340_ 39795341insCCCT	c.400_401insAGGG	p.V134Efs*82	g.38423418_ 38423419insCCCT	c.379_380insAGGG	p.V127Efs*82	p.V134Efs*	50			0	0	0	N/A	N/A	ALL		<18		**Yes**						
Missense	g.39795343A>G	c.398T>C	p.I133T	g.38423421A>G	c.377T>C	p.I126T	p.I133T	59	unknown	Skin fibroblasts	1 0.00016%	0	0	0.512	27.5	Chronic thrombocytopenia since childhood (<14 yr), AML (40 yr), relapse post HSCT (51 yr), died (52 yr). Familial history of dominant thrombocytopenia.	Not reported in clinical records	<14		**Yes** (evidenced in fibroblasts) / unknown transmission (no parental testing)	NA	NA	Not done	n/a	n/a	FLT3-ITD , IDH1 (Bone marrow at AML relapse)
Missense	g.39762932G>A	c.853C>T	p.R285C	g.38391010G>A	c.904C>T	p.R302C	p.R309C	40	M	PBMNC (Hair)	10 0.00068% (1xR302L) (17xR302H)	2	2 (5x R302H)	0.365	34	CLL	Not reported in clinical records	Unknown		**Yes**	NA					
Missense	g.39762919G>A	c.866C>T	p.P289L	g.38390997G>A	c.917C>T	p.P306L	p.P313L	40	F	PB	0 (6x P306W)	0	1	0.394	34	None	Lymphedema	<1		**Yes** (paternally inherited)	NA	NA	Reduced CD4/CD8 ratio	421	8.7	
Missense	g.39762919G>A	c.866C>T	p.P289L	g.38390997G>A	c.917C>T	p.P306L	p.P313L	57	M		0 (6x P306W)	0	1	0.394	34	None	Lymphedema			Yes						
Missense	g.39755744T>C	c.970A>G	p.M324V	g.38383822T>C	c.1021A>G	p.M341V	p.M348V	45	F	Skin fibroblasts	0	0	0 (2x M341I)	0.45	25.7	Severe congenital aplasia and abnormal B cells. Child received allo-HSCT.	Prematurity for acute fetal distress (33 weeks)	0	Allo-HSCT (~5 months), FU (8 yr), moderate attention disorder (prematurity?), type 2 renal failure (graft conditioning?)	**Yes** (paternally inherited)	del(7)(p21.3p12.3) 42.8 Mb	-7p	B-cell lymphopenia (increased T4/T8)	Transfused	<0.2	None (Myeloid panel-31 genes)
Missense	g.39755744T>C	c.970A>G	p.M324V	g.38383822T>C	c.1021A>G	p.M341V	p.M348V	45	F	Skin fibroblasts	0	0	0 (2x M341I)	0.45	25.7	MDS-RS-MLD, macrocytic anemia		56			NA					SF3B1 (K700E)
Missense	g.39755732C>T	c.982G>A	p.D328N	g.38383810C>T	c.1033G>A	p.D345N	p.D352N	24		PB	1 0.00012%	0	4 (1x D345E)	0.3149	33	MDS										
Missense	g.39755677T>G	c.1037A>C	p.D346A	g.38383755T>G	c.1088A>C	p.D363A	p.D370A	34		PB	0	0	3 (5x D363G) (5x D363N)	0.6209	28	MDS										
Missense	g.39755656C>T	c.1058G>A	p.R353H	g.38383734C>T	c.1109G>A	p.R370H	p.R377H	45	M	Skin fibroblasts and PB	0	0	0 (1x R370G, 1x R370P, 3x R370C, 2x R370L)	0.881	27.7	Neutropenia (at birth), pancytopenia (18 yr). Last follow up (19 yr). Family history of dominant cytopenia, aplastic anemia and AML.	Not reported in clinical records	0		**Yes** (evidenced in fibroblasts) / unknown transmission (no parental testing)	NA	NA	Not done	54 (diagnosis,18 yr), 80 (19 yr)	0.52 (at diagnosis at 18yo); 1.2 at 19yo	None (Blood at 18 yr)
Missense	g.39755656C>G	c.1058G>C	p.R353P	g.38383734C>G	c.1109G>C	p.R370P	p.R377P	44	M	PB Hair	0	0	1 (1x R370G, 3x R370C, 2x R370L)	0.888	28.2	MDS (asymptomatic thrombocytopenia and leukopenia)	Multiple health problems: an inguinal hernia on both sides, deformation (avascular necrosis) of the femoral head, heart problems (severe aortic valve insufficiency with secondary heart failure)	29		**Yes**	No obvious cnLOH	NA		48 (02.10.20), 42 (04.01.21), 50 (16.04.21), 41(14.07.21), 47 (13.10.21), 55 (20.01.22), 50 (20.04.22), 59 (18.07.22), 51 (17.10.22)	1.22 (02.10.20), 1.39 (04.01.21), 1.21 (16.04.21), 1.18 (14.07.21), 1.34 (13.10.21), 1.92 (20.01.22), 1.68 (20.04.22), 1.39 (18.07.22), 1.47 (17.10.22)	No
Truncation	g.39755649C>G	c.1065C>G	p.Y355*	g.38383727G>C	c.1116C>G	p.Y372*	p.Y379*	48	M	PB	0	0	0 (1x Y372=)	n/a	38	Congenital pancytopenia, bone marrow failure.		0		**Yes** (*de novo*)						
Missense	g.39755647T>C	c.1067A>G	p.Y356C	g.38383725A>G	c.1118A>G	p.Y373C	p.Y380C	17	M	PBMNC (Hair)	0	0	0	0.852	31	Thombocytopenia, neutropenia	Not reported in clinical records	21		**Yes** (maternally inherited)	2x 21q cnLOH events (mosaic)	NK		143	1	RUNX1 (G165V) (2%) Two 21q cnLOH events
Missense	g.39755647T>C	c.1067A>G	p.Y356C	g.38383725A>G	c.1118A>G	p.Y373C	p.Y380C	44	M	PBMNC (Hair)	0	0	0	0.852	31	Thombocytopenia, neutropenia	Not reported in clinical records	19		**Yes** (maternally inherited)	21q cnLOH (mosaic)					21q cnLOH event
Missense	g.39755647T>C	c.1067A>G	p.Y356C	g.38383725A>G	c.1118A>G	p.Y373C	p.Y380C	30	F	BM (Hair)	0	0	0	0.852	31	AML, Thrombocytopenia, t-MDS (RAEB2)	Not reported in clinical records	27		**Yes**	21q cnLOH (mosaic) Germline mosaic +8	Germline mosaic +8		96	0.34	IDH1 (R132C) (40%) TP53 (C238Y) (22%) GATA2 (H442Q6/*95) (20%) 21q cnLOH event
Missense	g.39755625C>G	c.1089G>C	p.K363N	g.38383703C>G	c.1140G>C	p.K380N	p.K387N	46		PB	0	0	0 (1x K380E)	0.642	22.9	Anemia, thrombocytopenia, pancytopenia, macrocytic anemia, abnormality of the spleen, eosinophilic infiltration of esophagus.	Frontal bossing, Hypoalbuminemia, Abnormal vitamin B12 level, Hemangioma, Hepatosplenomegaly, Vomiting, Diarrhea, Exocrine pancreatic insufficiency, Gastrointestinal inflammation, Erythema, Protein-losing enteropathy, Posteriorly rotated ears, Capillary malformation, Generalized-onset seizure, Autistic behavior, Delayed speech and language development, Weight loss.	>18		**Yes** (***de novo***)						
Missense	g.39755602T>C	c.1112A>G	p.Y371C	g.38383680T>C	c.1163A>G	p.Y388C	p.Y395C	50	F	PB	0	0	1	0.90	28.6	None	Lymphedema	8		**Yes** (maternally inherited)	ND		Reduced CD4/CD8 ratio	171	3.8	
Missense	g.39755602T>C	c.1112A>G	p.Y371C	g.38383680T>C	c.1163A>G	p.Y388C	p.Y395C	33	F	PB	0	0	1	0.90	28.6	None	lymphoscintigraphy showed slightly sluggish lymphatic drainage in both legs, but no significant impairment of lymphatic drainage. Venous duplex demonstrated varicose veins, often seen in lymphedema patients. No episodes of cellulitis or warts. No systemic lymphedema involvement. Treatment for supra-ventricular tachycardia.	50		**Yes**	21q cnLOH event					
Missense	g.39755585C>A	c.1129G>T	p.G377W	g.38383663C>A	c.1180G>T	p.G394W	p.G401W	15		PB	0 (3x G394R)	0	2 (8x G394R)	0.661	35	t-AML after treatment for DLBCL and prostate cancer	DLBCL and prostate cancer			**Yes**						
Missense	g.39774495C>T	c.678G>A	p.M226I	g.38402573C>T	c.657G>A	p.M219I				37M,69F		372 0.023% (1 hom)	106	1	0.05799	22.5	n/a				**Yes**						
Missense	g.39763629G>A	c.772C>T	p.P258S	g.38391707G>A	c.823C>T	p.P275S				81M,17F		254 0.016% (3 hom)	98	1	0.223	22.4	n/a				**Yes**						
Missense	g.39755801A>G	c.913T>C	p.S305P	g.38383879A>G	c.964T>C	p.S322P				n/a		0	0	0 (2x S322L)	0.574	31	ALL, thrombocytopenia				**Yes**						
Missense	g.39755657G>T	c.1057C>A	p.R353S	g.38383735G>T	c.1108C>A	p.R370S				n/a		0	0	0 (1x R370G/P, 3x R370C, 2x R370L)	0.803	33	ETV6 variant (ALL)				**No**						
Missense	g.39755611C>T	c.1103G>A	p.R368H	g.38383689C>T	c.1154G>A	p.R385H				unknown		0 (2x R385C)	0 (1xR385C)	4 (5x R385C)	0.674	27.6	n/a	2x BrCa, 1x Billiary, 1x Upper aerodigestive tract			**No**						
Missense	g.39755555G>C	c.1159C>G	p.P387A	g.38383633G>C	c.1210C>G	p.P404A				16M,10F		236 0.015%	26	0 (1x P404S, 3x P404L)	0.086	19.6	n/a		-		**Yes**						

Supplemental Table 3. ERG variants identified in international cohort samples. WGS from Cohort 8 (Genomics England); * 2 variants from this study and 3 identified in previous study.11

Cohort	Number of ERG variants identified	Number of ERG variant carriers	ERG variant	ACMG Classification	No. of patients screened in cohort	Sequencing Method (WGS, WES, Panel)	Phenotype(s) in cohort
1	2	4	p.Y373C	P	441 (215 families)	WES	Inherited BMF and/or HM
			p.R302C	VUS			
2	3	3	p.I126T	VUS	179	WES	Inherited BMF
			p.M341V	VUS			
			p.R370H	P			
3	5	5	p.E20Vfs*13	VUS	1655	WES	HM (all types)
			p.M341V	VUS			
			p.D345N	VUS			
			p.D363A	VUS			
			p.G394W	VUS			
4	2	2	p.P116R	VUS	700-800	WES (with panels for BMF and/or coagulation disorders)	Bleeding disorder and/or BMF
			p.R370P	P			
5	1	1	p.K380N	P	29	WGS	Rare Disease participants (unsolved patients) - many phenotypes, mostly neurological and immune related. (ERG variant was in the only thrombocytopenia case)
6	1	1	p.V127Efs*	VUS	6889	WES and/or WGS	Pediatric cancer types (>22) including: HMs - BALL (n=2037), TALL (n=451), AML (n=483), NHL (n=261), HL (n=367). Solid and brain cancers- ATRT, ACC, ACPG, EPD, EWS, GCT, HGG, HM, LGG, LIC, MB, MEL, NBL, OS, RB, RMS, WLM.
7	1	1	p.Y372*	P	>1000	WES	Several phenotypes including BMF
8	5*	2	p.P306L	VUS	95	WGS	Lymphedema
			p.Y388C	P			
9	0	N/A	N/A	N/A	55	WES	AA, MDS, cytopenia
10	0	N/A	N/A	N/A	~350	WES	Suspected predisposition for hematological malignancies, mainly myeloid neoplasms and thrombocytopenia.
11	0	N/A	N/A	N/A	37	Panel (Thrombocytopenia)	Suspected of having a platelet-based bleeding disorder
12	0	N/A	N/A	N/A	>100	unknown	unkknown
13	0	N/A	N/A	N/A	97	WGS	Neutropenia including congenital (n=14), Congenital anaemia (not DBA) (n=1), Diamond Blackfan Anaemia (n=12), Congenital thrombocytopenia (n=7), Telomere Biology Disorder (n=5), Hypocellular BMF (n=40), MDS/AML (n=18)
14	0	N/A	N/A	N/A	30	WES	Neutropenia
15	0	N/A	N/A	N/A	49	WES	Inherited MDS, AML, ALL, Lymphoma, CMML, AA, BMF
16	0	N/A	N/A	N/A	9	WES	Thrombocytopenia
17	0	N/A	N/A	N/A	47	WGS	Inherited BMF
18	0	N/A	N/A	N/A	496	WES	184 CNS 131 sarcomas 97 other solid 84 haem

Supplemental Table 4. Constraint scores of amino acid position changes.[7]

Domain	Amino Acid Position	Metadome derived constraint scores	Domain	Amino Acid Position	Metadome derived constraint scores
PNT	113	0.66	ETS	310	0.24
PNT	114	0.54	ETS	311	0.33
PNT	115	0.68	ETS	312	0.28
PNT	116	0.56	ETS	313	0.13
PNT	117	0.55	ETS	314	0.13
PNT	118	0.58	ETS	315	0.1
PNT	119	0.53	ETS	316	0.1
PNT	120	0.46	ETS	317	0.06
PNT	121	0.49	ETS	318	0.06
PNT	122	0.48	ETS	319	0.06
PNT	123	0.45	ETS	320	0.05
PNT	124	0.46	ETS	321	0.05
PNT	125	0.53	ETS	322	0.06
PNT	126	0.46	ETS	323	0.06
PNT	127	0.36	ETS	324	0.06
PNT	128	0.35	ETS	325	0.06
PNT	129	0.38	ETS	326	0.1
PNT	130	0.43	ETS	327	0.11
PNT	131	0.41	ETS	328	0.13
PNT	132	0.46	ETS	329	0.12
PNT	133	0.46	ETS	330	0.12
PNT	134	0.34	ETS	331	0.11
PNT	135	0.38	ETS	332	0.14
PNT	136	0.42	ETS	333	0.19
PNT	137	0.53	ETS	334	0.21
PNT	138	0.5	ETS	335	0.21
PNT	139	0.47	ETS	336	0.23
PNT	140	0.49	ETS	337	0.25
PNT	141	0.53	ETS	338	0.25
PNT	142	0.53	ETS	339	0.33
PNT	143	0.61	ETS	340	0.51
PNT	144	0.56	ETS	341	0.65
PNT	145	0.65	ETS	342	0.66
PNT	146	0.55	ETS	343	0.67
PNT	147	0.76	ETS	344	0.57
PNT	148	0.85	ETS	345	0.54
PNT	149	1.04	ETS	346	0.5
PNT	150	0.97	ETS	347	0.56
PNT	151	1.18	ETS	348	0.56
PNT	152	0.88	ETS	349	0.75
PNT	153	0.74	ETS	350	0.81
PNT	154	0.65	ETS	351	0.81
PNT	155	0.69	ETS	352	0.63
PNT	156	0.7	ETS	353	0.53
PNT	157	0.69	ETS	354	0.47

PNT	158	0.6	ETS	355	0.47
PNT	159	0.5	ETS	356	0.47
PNT	160	0.5	ETS	357	0.39
PNT	161	0.44	ETS	358	0.39
PNT	162	0.37	ETS	359	0.33
PNT	163	0.31	ETS	360	0.27
PNT	164	0.23	ETS	361	0.22
PNT	165	0.26	ETS	362	0.22
PNT	166	0.19	ETS	363	0.21
PNT	167	0.19	ETS	364	0.21
PNT	168	0.17	ETS	365	0.17
PNT	169	0.13	ETS	366	0.22
PNT	170	0.18	ETS	367	0.23
PNT	171	0.19	ETS	368	0.22
PNT	172	0.17	ETS	369	0.22
PNT	173	0.22	ETS	370	0.13
PNT	174	0.24	ETS	371	0.1
PNT	175	0.2	ETS	372	0.08
PNT	176	0.18	ETS	373	0.11
PNT	177	0.19	ETS	374	0.09
PNT	178	0.2	ETS	375	0.12
PNT	179	0.19	ETS	376	0.1
PNT	180	0.24	ETS	377	0.11
PNT	181	0.23	ETS	378	0.12
PNT	182	0.26	ETS	379	0.11
PNT	183	0.28	ETS	380	0.1
PNT	184	0.32	ETS	381	0.13
PNT	185	0.47	ETS	382	0.12
PNT	186	0.55	ETS	383	0.12
PNT	187	0.85	ETS	384	0.11
PNT	188	0.9	ETS	385	0.15
PNT	189	0.84	ETS	386	0.19
PNT	190	1.12	ETS	387	0.16
PNT	191	0.9	ETS	388	0.16
PNT	192	0.95	ETS	389	0.22
PNT	193	0.88	ETS	390	0.2
PNT	194	0.8	ETS	391	0.25
PNT	195	0.74	ETS	392	0.24
PNT	196	1.04	ETS	393	0.36
PNT	197	0.7	ETS	394	0.35
PNT	198	0.72	ETS	395	0.4
PNT	199	0.68			
		0.5208046			**0.257209302**

Supplemental table 5: ACMG classificaiton of ERG variants. ERG variants were classified using American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG-AMP) criteria to cytopenias, HMs and/or lymphedema (#). For functional assay criterion, for variants with complete loss-of-function in at least one functional assay (PS3_Strong), for hypomorphic activity (>50%) in one or more functional assays (PS3_Moderate) and for hypomorphic activity (<50%) in one or more assays (PS3_Supporting). PS3 criteria was not used where no functional data was ascertained and was not used to compare the difference of ACMF classficaiton without functional data.

Patient ID	**p.HGVS** (NP_891548.1) (ENSP00000288319.7) (479 aa)	**ACMG classification**	**ACMG criteria**	**ACMG classification** (without functional data)	**ACMG criteria** (with out functional data)
1	p.E20Vfs*13	Uncertain significance	PVS1_Moderate, PM2_Supporting	Uncertain significance	PVS1_Moderate, PM2_Supporting
2	p.P116R	Uncertain significance	BS3_Supporting	Uncertain significance	None
3	p.V127Efs*82	Uncertain significance	PVS1, PM2_Supporting	Uncertain significance	PVS1, PM2_Supporting
4	p.I126T	Uncertain significance	BS3_Supporting, PM2_Supporting	Uncertain significance	PM2_Supporting
5	p.R302C	Uncertain significance	PS3_Moderate	Uncertain significance	None
6	p.P306L	Uncertain significance	PS3_Supporting, PM2_Supporting	Uncertain significance	PM2_Supporting
7	p.M341V	Uncertain significance	PS3_Moderate, PM2_Supporting	Uncertain significance	PM2_Supporting
8	p.M341V	Uncertain significance	PS3_Moderate, PM2_Supporting	Uncertain significance	PM2_Supporting
9	p.D345N	Uncertain significance (somatic?)	PS3_Strong, PM2_Supporting	Uncertain significance	PM2_Supporting
10	p.D363A	Uncertain significance (somatic?)	PS3_Moderate, PM2_Supporting	Uncertain significance	PM2_Supporting
11	p.R370H	Pathogenic	PS3, PM1, PM5, PM2-Supporting, PP3	Likely Pathogenic	PM1, PM5, PM2-Supporting, PP3
12	p.R370P	Pathogenic	PS3, PM1, PM5, PM2-Supporting, PP3	Likely Pathogenic	PM1, PM5, PM2-Supporting, PP3
13	p.Y372*	Pathogenic	PVS1, PS2, PM2_Supporting	Pathogenic	PVS1, PS2, PM2_Supporting
14 (I-2)	p.Y373C	Likely Pathogenic	PS3, PM1, PM2_Supporting, PP1, PP3	Uncertain significance	PM1, PM2_Supporting, PP1, PP3
15 (II-1)	p.Y373C	Likely Pathogenic	PS3, PM1, PM2_Supporting, PP1, PP3	Uncertain significance	PM1, PM2_Supporting, PP1, PP3
16 (II-2)	p.Y373C	Likely Pathogenic	PS3, PM1, PM2_Supporting, PP1, PP3	Uncertain significance	PM1, PM2_Supporting, PP1, PP3
17	p.K380N	Pathogenic	PS2, PS3_Moderate, PM1, PM2_Supporting, PP3	Likely Pathogenic	PS2, PM1, PM2_Supporting, PP3
18	p.Y388C	Likely Pathogenic	PS3, PM2_Supporting, PP3	Uncertain significance	PM2_Supporting, PP3
19	p.G394W	Uncertain significance	PS3_Moderate, PM2_Supporting	Uncertain significance	PM2_Supporting
Population	p.M219I	Likely Benign	BS1, BS3_Supporting	Uncertain significance	BS1
Population	p.P275S	Likely Benign	BS1, BS3_Supporting	Uncertain significance	BS2
Mouse	p.S322P	Likely Pathogenic	PS3, PM2_Supporting, PP3	Uncertain significance	PM2_Supporting, PP3
ETV6	p.R370S	Pathogenic	PS3, PM1, PM5, PM2_Supporting, PP3	Likely Pathogenic	PM1, PM5, PM2_Supporting, PP3
COSMIC	p.R385H	Uncertain significance (somatic)	PS3_Moderate, PM2_Supporting, PP3	Uncertain significance	PM2_Supporting, PP3
Populatoin	p.P404A	Uncertain significance	BS1_Moderate	Uncertain significance	BS1_Moderate

Supplemental Table 6. Somatic ERG mutations in COSMIC. ERG variants (NM_182918.4) were extracted from 20,340 cancer samples in COSMIC (v99) and filtered for missense, nonsense, frameshift and inframe indels (PROTEININS) occurring ≤ 3 times in the gnomAD (v4.0.0) population database.

Gene	Chr position	Variant	gnomAD 4.0.0	Disease
ERG	38445621	E7K	0	Skin
ERG	38445615	L9V	0	Liver
ERG	38445611	S10L	0	Skin
ERG	38445609	V11I	3	Kidney
ERG	38445590	S17Yfs*8	0	Acute lymphoblastic B cell leukaemia
ERG	38445582	E20Q	0	Urinary tract
ERG	38445571	Y23*	0	Diffuse large B cell lymphoma
ERG	38445567	T25S	1	Lung
ERG	38445564	P26S	0	Endometrium
ERG	38445558	L28M	0	Endometrium
ERG	38445550	K30N	0	Ovary
ERG	38445549	T31A	0	Biliary tract
ERG	38445546	E32Q	0	Large intestine
ERG	38445546	E32Q	0	Lung
ERG	38445541	M33I	0	Breast
ERG	38445541	M33I	0	Breast
ERG	38445541	M33I	0	Breast
ERG	38445530	S37F	0	Skin
ERG	38445527	S38F	2	Skin
ERG	38445522	D40Y	0	Skin
ERG	38445506	S45F	0	NS
ERG	38445506	S45F	0	Skin
ERG	38445497	S48I	0	Lung
ERG	38445489	V51F	0	Stomach
ERG	38445479	Q54P	0	Skin
ERG	38445470	L57Q	0	Liver
ERG	38445470	L57Q	0	Liver
ERG	38445470	L57Q	0	Liver
ERG	38445470	L57Q	0	Liver
ERG	38445470	L57Q	0	Liver
ERG	38445465	Q59*	0	Cervix
ERG	38445464	Q59P	0	Skin
ERG	38445459	P61T	0	Liver
ERG	38445459	P61T	0	Liver
ERG	38445459	P61S	1	Skin
ERG	38445451	R63S	0	Skin
ERG	38445444	I66V	1	Pancreas
ERG	38445431	C70F	0	Biliary tract
ERG	38445422	S73N	0	Skin
ERG	38445421	S73R	2	Liver
ERG	38445421	S73R	2	Lung
ERG	38445417	V75M	0	NS
ERG	38445410	G77D	0	Thyroid
ERG	38445410	G77D	0	Thyroid
ERG	38423548	E84K	0	Skin
ERG	38423535	A88V	0	Skin
ERG	38423522	K92N	0	Large intestine
ERG	38423516	G95Afs*8	0	Lung
ERG	38423502	T99I	0	NS
ERG	38423497	G101R	0	Skin

ERG	38423497	G101W	0	Skin
ERG	38423494	M102L	3	Liver
ERG	38423490	N103S	2	Skin
ERG	38423484	G105V	1	Ovary
ERG	38423478	Y107F	0	Lung
ERG	38423474	M108I	1	Large intestine
ERG	38423474	M108I	1	Skin
ERG	38423474	M108I	1	Skin
ERG	38423474	M108I	1	Skin
ERG	38423473	E109K	0	Skin
ERG	38423473	E109K	0	Skin
ERG	38423454	P115L	0	Central nervous system
ERG	38423454	P115L	0	Central nervous system
ERG	38423451	P116Q	0	Upper aerodigestive tract
ERG	38423444	M118I	0	Lung
ERG	38423434	E122Q	1	Thyroid
ERG	38423421	I126Rfs*83	0	Acute lymphoblastic B cell leukaemia
ERG	38423421	I126Rfs*83	0	Acute lymphoblastic leukaemia
ERG	38423419	V127L	2	Lung
ERG	38423418	V127E	0	Breast
ERG	38423416	P128S	2	Breast
ERG	38423416	P128S	2	Breast
ERG	38423413	A129T	0	Large intestine
ERG	38423410	D130N	0	Small intestine
ERG	38403707	P131S	0	NS
ERG	38403707	P131S	0	NS
ERG	38403707	P131S	0	Skin
ERG	38403706	P131H	0	Endometrium
ERG	38403706	P131L	0	Lung
ERG	38403706	P131H	0	Stomach
ERG	38403696	W134C	0	Skin
ERG	38403680	R140W	3	Endometrium
ERG	38403680	R140W	3	Endometrium
ERG	38403680	R140W	3	Endometrium
ERG	38403680	R140W	3	Large intestine
ERG	38403680	R140W	3	Oesophagus
ERG	38403676	Q141P	0	Lung
ERG	38403673	W142L	0	Liver
ERG	38403665	E144_W145delinsG	0	Central nervous system
ERG	38403664	W145*	0	Skin
ERG	38403664	W145*	0	Skin
ERG	38403654	K148N	0	Breast
ERG	38403654	K148N	0	Oesophagus
ERG	38403654	K148N	0	Oesophagus
ERG	38403654	K148N	0	Oesophagus
ERG	38403654	K148N	0	Oesophagus
ERG	38403653	E149*	0	Large intestine
ERG	38403653	E149*	0	Large intestine
ERG	38403653	E149K	0	Skin
ERG	38403651	E149D	0	Endometrium
ERG	38403647	G151R	0	Large intestine
ERG	38403646	G151D	0	T cell large granular lymphocytic leukaemia
ERG	38403646	G151D	0	Stomach
ERG	38403646	G151D	0	Stomach
ERG	38403641	P153S	0	NS
ERG	38403641	P153S	0	Skin

ERG	38403625	L158*	0	Large intestine
ERG	38403610	I163T	0	Large intestine
ERG	38403606	D164_G165insP	0	Acute lymphoblastic B cell leukaemia
ERG	38403585	M171I	0	Upper aerodigestive tract
ERG	38403584	T172A	0	Kidney
ERG	38403569	Q177*	0	Skin
ERG	38403563	L179Pfs*26	0	Acute lymphoblastic B cell leukaemia
ERG	38403563	L179F	0	NS
ERG	38403559	S182Qfs*26	0	Large intestine
ERG	38403557	P181S	0	Skin
ERG	38403557	P181S	0	Skin
ERG	38403545	A185T	2	Large intestine
ERG	38403536	L188I	0	Ovary
ERG	38403518	Y194H	0	Breast
ERG	38403517	Y194C	0	Large intestine
ERG	38403517	Y194*	0	Ovary
ERG	38403516	Y194*	0	Kidney
ERG	38403511	R196K	1	Urinary tract
ERG	38403509	E197*	0	Skin
ERG	38402632	L200F	0	Central nervous system
ERG	38402629	P201S	0	Skin
ERG	38402621	L203F	0	Kidney
ERG	38402614	D206H	0	Urinary tract
ERG	38402614	D206H	0	Urinary tract
ERG	38402614	D206H	0	Urinary tract
ERG	38402611	D207Y	0	Endometrium
ERG	38402591	Q213H	0	Skin
ERG	38402586	S215F	0	Skin
ERG	38402586	S215F	0	Skin
ERG	38402584	P216S	0	Lung
ERG	38402584	P216T	0	Skin
ERG	38402581	R217W	1	Central nervous system
ERG	38402572	H220N	0	Breast
ERG	38402561	N223K	0	Lung
ERG	38400646	G226Vfs*65	0	Prostate
ERG	38400646	G226Vfs*65	0	Prostate
ERG	38400646	G226Vfs*65	0	Stomach
ERG	38400645	G225E	0	Skin
ERG	38400645	G225E	0	Skin
ERG	38400645	G225E	0	Skin
ERG	38400645	G225E	0	Skin
ERG	38400637	A228S	0	Endometrium
ERG	38400637	A228S	0	Kidney
ERG	38400625	P232T	0	Pancreas
ERG	38400624	P232L	0	Skin
ERG	38400624	P232L	0	Skin
ERG	38400612	V236A	0	Upper aerodigestive tract
ERG	38400606	P238H	0	Central nervous system
ERG	38400604	E239K	0	Breast
ERG	38400604	E239K	0	Breast
ERG	38400597	T241M	1	Breast
ERG	38400597	T241M	1	Prostate
ERG	38400591	R243I	0	Cervix
ERG	38400591	R243I	0	Large intestine
ERG	38400591	R243I	0	Large intestine
ERG	38400591	R243I	0	Large intestine

ERG	38400583	T246A	2	Stomach
ERG	38392441	Y252Tfs*3	0	Myelodysplastic syndrome
ERG	38392430	R256Qfs*29	2	Myelodysplastic syndrome
ERG	38392430	R256Gfs*35	0	Large intestine
ERG	38392430	R256Gfs*35	0	Lung
ERG	38392429	P254L	0	Skin
ERG	38392426	P255L	0	Upper aerodigestive tract
ERG	38392420	R257K	0	Urinary tract
ERG	38392417	S258L	0	Skin
ERG	38392415	A259S	1	Small intestine
ERG	38392410	W260C	0	Stomach
ERG	38392381	S270*	0	Stomach
ERG	38391710	Q274E	0	Urinary tract
ERG	38391688	P281H	0	Lung
ERG	38391680	E284K	2	Lung
ERG	38391680	E284K	2	NS
ERG	38391674	Q286*	0	Breast
ERG	38391674	Q286*	0	Breast
ERG	38391674	Q286E	0	Lung
ERG	38391673	Q286R	0	Breast
ERG	38391665	Q289*	0	Breast
ERG	38391665	Q289*	0	Breast
ERG	38391665	P288Vfs*12	0	Acute lymphoblastic leukaemia
ERG	38391665	Q289*	0	Skin
ERG	38391039	P292L	0	Central nervous system
ERG	38391039	P292L	0	Skin
ERG	38391034	Q294*	0	Myelodysplastic syndrome
ERG	38391033	Q294P	0	Endometrium
ERG	38391032	Q294Sfs*6	0	Lung
ERG	38391028	L296I	0	Endometrium
ERG	38391028	L296F	0	Urinary tract
ERG	38391025	G297Qfs*10	0	Acute lymphoblastic T cell leukaemia
ERG	38391025	G297*	0	Penis
ERG	38391024	G297E	0	Myelodysplastic syndrome
ERG	38391024	G297E	0	Skin
ERG	38391024	G297E	0	Stomach
ERG	38391015	S300I	0	Central nervous system
ERG	38391015	S300N	0	Large intestine
ERG	38390997	P306L	0	Skin
ERG	38383915	Q310K	0	Lung
ERG	38383913	Q310delinsH*G	0	Biliary tract
ERG	38383911	I311Kfs*33	0	Biliary tract
ERG	38383906	L313F	0	Central nervous system
ERG	38383906	L313F	0	NS
ERG	38383905	L313H	0	Breast
ERG	38383900	Q315*	0	Thyroid
ERG	38383900	Q315*	0	Thyroid
ERG	38383894	L317F	0	Skin
ERG	38383890	L318P	0	Peripheral T cell lymphoma unspecified
ERG	38383888	E319K	0	Breast
ERG	38383888	E319*	0	Thyroid
ERG	38383888	E319*	0	Thyroid
ERG	38383878	S322L	0	Ovary
ERG	38383878	S322L	0	Ovary
ERG	38383873	D323_S324ins*	0	Biliary tract
ERG	38383870	S325P	0	Stomach

ERG	38383863	S327F	0	Skin
ERG	38383860	S328N	1	Endometrium
ERG	38383860	S328I	1	Acute myeloid leukaemia
ERG	38383860	S328I	1	Acute myeloid leukaemia
ERG	38383844	E333D	0	Biliary tract
ERG	38383843	G334S	0	Skin
ERG	38383831	E338K	0	Skin
ERG	38383820	M341I	0	Large intestine
ERG	38383820	M341I	0	Large intestine
ERG	38383819	T342A	0	Stomach
ERG	38383816	D343N	0	Endometrium
ERG	38383813	P344T	0	Endometrium
ERG	38383813	P344S	0	Skin
ERG	38383812	P344L	0	Liver
ERG	38383812	P344L	0	Liver
ERG	38383812	P344L	0	Lung
ERG	38383810	D345N	1	Gastrointestinal tract (site indeterminate)
ERG	38383810	D345N	1	Acute myeloid leukaemia
ERG	38383810	D345N	1	Acute lymphoblastic B cell leukaemia
ERG	38383810	D345N	1	Acute myeloid leukaemia
ERG	38383810	D345N	1	Large intestine
ERG	38383810	D345N	1	Stomach
ERG	38383808	D345E	0	Kidney
ERG	38383807	E346K	0	Breast
ERG	38383807	E346K	0	Breast
ERG	38383798	R349W	3	Endometrium
ERG	38383798	R349W	3	Diffuse large B cell lymphoma
ERG	38383798	R349W	3	Acute lymphoblastic B cell leukaemia
ERG	38383797	R349Q	0	Large intestine
ERG	38383794	R350H	0	Breast
ERG	38383794	R350H	0	NK-T cell lymphoma
ERG	38383791	W351*	0	Urinary tract
ERG	38383789	G352R	0	Skin
ERG	38383786	E353Pfs*2	0	Biliary tract
ERG	38383785	E353G	0	Biliary tract
ERG	38383782	R354Q	1	Adult T cell lymphoma-leukaemia
ERG	38383775	S356R	0	Myelodysplastic syndrome
ERG	38383771	P358S	0	NS
ERG	38383770	P358L	0	Skin
ERG	38383763	M360I	1	Skin
ERG	38383763	M360I	1	Skin
ERG	38383761	N361Sfs*5	0	Biliary tract
ERG	38383760	N361K	0	Acute lymphoblastic leukaemia
ERG	38383756	D363N	0	Acute lymphoblastic T cell leukaemia
ERG	38383756	D363N	0	Acute lymphoblastic B cell leukaemia
ERG	38383756	D363N	0	Acute lymphoblastic B cell leukaemia
ERG	38383756	D363N	0	Acute lymphoblastic T cell leukaemia
ERG	38383756	D363N	0	Skin
ERG	38383755	D363A	0	Acute lymphoblastic B cell leukaemia
ERG	38383755	D363A	0	Acute lymphoblastic B cell leukaemia
ERG	38383755	D363A	0	Acute lymphoblastic B cell leukaemia
ERG	38383755	D363G	0	Acute lymphoblastic B cell leukaemia
ERG	38383755	D363G	0	Acute lymphoblastic B cell leukaemia
ERG	38383755	D363G	0	Acute lymphoblastic B cell leukaemia
ERG	38383755	D363G	0	Acute lymphoblastic B cell leukaemia
ERG	38383755	D363G	0	Acute lymphoblastic B cell leukaemia

ERG	38383744	R367C	0	Central nervous system
ERG	38383744	R367Lfs*13	0	Acute lymphoblastic B cell leukaemia
ERG	38383744	R367C	0	Prostate
ERG	38383743	R367L	0	Lung
ERG	38383743	R367H	0	Stomach
ERG	38383743	R367H	0	Stomach
ERG	38383743	R367H	0	Stomach
ERG	38383742	R367_A368insSV	0	Biliary tract
ERG	38383741	A368T	0	Cervix
ERG	38383741	A368T	0	Large intestine
ERG	38383741	A368T	0	Large intestine
ERG	38383741	A368T	0	Large intestine
ERG	38383741	A368T	0	Large intestine
ERG	38383741	A368S	0	Skin
ERG	38383738	L369F	0	Pancreas
ERG	38383735	R370G	0	Lung
ERG	38383735	R370C	0	Pancreas
ERG	38383735	R370C	0	Skin
ERG	38383735	R370C	0	Skin
ERG	38383734	R370L	0	Lung
ERG	38383734	R370L	0	Lung
ERG	38383734	R370P	0	Soft tissue
ERG	38383730	Y371*	0	Stomach
ERG	38383711	M378V	0	Urinary tract
ERG	38383709	M378I	0	Skin
ERG	38383705	K380E	0	Central nervous system
ERG	38383701	V381A	2	Central nervous system
ERG	38383701	V381A	2	Central nervous system
ERG	38383696	G383W	0	Skin
ERG	38383695	G383E	0	Skin
ERG	38383690	R385C	2	Bone
ERG	38383690	R385C	2	Liver
ERG	38383690	R385C	2	Liver
ERG	38383690	R385C	2	Liver
ERG	38383690	R385C	2	Skin
ERG	38383689	R385H	0	Biliary tract
ERG	38383689	R385H	0	Breast
ERG	38383689	R385H	0	Breast
ERG	38383689	R385H	0	Upper aerodigestive tract
ERG	38383684	A387T	3	Prostate
ERG	38383684	A387T	3	Prostate
ERG	38383680	Y388C	0	Stomach
ERG	38383677	K389T	0	Large intestine
ERG	38383677	K389T	0	Lung
ERG	38383677	K389T	0	Oesophagus
ERG	38383677	K389T	0	Stomach
ERG	38383672	D391N	1	Endometrium
ERG	38383672	D391H	0	Testis
ERG	38383672	D391H	0	Urinary tract
ERG	38383665	H393R	0	Urinary tract
ERG	38383663	G394W	0	Placenta
ERG	38383663	G394R	3	Skin
ERG	38383663	G394R	3	Skin
ERG	38383663	G394R	3	Skin
ERG	38383663	G394R	3	Skin
ERG	38383663	G394R	3	Skin

ERG	38383663	G394R	3	Skin
ERG	38383663	G394R	3	Stomach
ERG	38383663	G394R	3	Urinary tract
ERG	38383663	G394W	3	Urinary tract
ERG	38383659	I395N	1	Upper aerodigestive tract
ERG	38383652	Q397H	1	Large intestine
ERG	38383650	A398V	1	Skin
ERG	38383650	A398V	1	Skin
ERG	38383648	L399I	0	Large intestine
ERG	38383648	L399F	0	NS
ERG	38383641	P401H	0	Prostate
ERG	38383639	H402Y	0	Skin
ERG	38383639	H402Y	0	Skin
ERG	38383638	H402R	0	NS
ERG	38383637	P404Rfs*25	0	Biliary tract
ERG	38383637	P404Rfs*25	0	Bone
ERG	38383637	P404Rfs*25	0	Large intestine
ERG	38383637	P404Rfs*25	0	Large intestine
ERG	38383637	P404Rfs*25	0	Large intestine
ERG	38383637	P404Rfs*25	0	Large intestine
ERG	38383637	P404Rfs*25	0	Large intestine
ERG	38383637	P404Rfs*25	0	Large intestine
ERG	38383637	P404Rfs*25	0	Large intestine
ERG	38383637	P404Rfs*25	0	Large intestine
ERG	38383637	P404Rfs*25	0	Large intestine
ERG	38383637	P404Rfs*25	0	Large intestine
ERG	38383637	P404Rfs*25	0	Large intestine
ERG	38383637	P404Rfs*25	0	Large intestine
ERG	38383637	P404Rfs*25	0	Stomach
ERG	38383637	P404Rfs*25	0	Stomach
ERG	38383637	P404Rfs*25	0	Thymus
ERG	38383635	P403L	1	Skin
ERG	38383635	P403L	1	Upper aerodigestive tract
ERG	38383633	P404S	1	Skin
ERG	38383632	P404L	2	Biliary tract
ERG	38383632	P404L	2	Upper aerodigestive tract
ERG	38383632	P404L	2	Upper aerodigestive tract
ERG	38383626	S406*	0	Ovary
ERG	38383626	S406L	0	Skin
ERG	38383626	S406L	0	Skin
ERG	38383622	L408Dfs*23	0	Acute lymphoblastic B cell leukaemia
ERG	38383608	P412F	0	Skin
ERG	38383608	P412L	0	Skin
ERG	38383605	S413L	0	Breast
ERG	38383605	S413L	0	Breast
ERG	38383605	S413*	0	Urinary tract
ERG	38383605	S413*	0	Urinary tract
ERG	38383605	S413L	0	Urinary tract
ERG	38383597	P416T	0	Breast
ERG	38383596	P416R	0	Lung
ERG	38383596	P416R	0	Lung
ERG	38383593	Y417F	0	Breast
ERG	38383593	Y417F	0	Breast
ERG	38383593	Y417F	0	Breast
ERG	38383589	M418I	2	Liver
ERG	38383589	M418I	2	Skin

ERG	38383588	G419S	1	Skin
ERG	38383587	G419V	0	Skin
ERG	38383584	S420F	1	Skin
ERG	38383579	H422Y	0	Skin
ERG	38383579	H422Y	0	Skin
ERG	38383579	H422Y	0	Skin
ERG	38383579	H422Y	0	Skin
ERG	38383573	H424N	0	Skin
ERG	38383573	H424N	0	Thyroid
ERG	38383570	P425T	0	Bone
ERG	38383563	K427R	0	Acute lymphoblastic B cell leukaemia
ERG	38383561	M428V	0	Lung
ERG	38383559	M428I	1	Large intestine
ERG	38383559	M428I	1	NS
ERG	38383556	N429K	0	Salivary gland
ERG	38383556	N429K	0	Salivary gland
ERG	38383549	A432T	2	Skin
ERG	38383546	P433S	1	Urinary tract
ERG	38383543	H434Y	0	NS
ERG	38383540	P435T	2	Stomach
ERG	38383537	P436A	0	Breast
ERG	38383533	A437V	0	Skin
ERG	38383527	P439L	0	Skin
ERG	38383525	V440Sfs*26	0	Acute lymphoblastic T cell leukaemia
ERG	38383521	T441I	0	Lung
ERG	38383511	A447Cfs*19	0	Biliary tract
ERG	38383511	F446Lfs*59	0	Endometrium
ERG	38383511	A447Cfs*19	0	Large intestine
ERG	38383511	A447Cfs*19	0	Large intestine
ERG	38383511	F446Lfs*59	0	Large intestine
ERG	38383511	F446Lfs*59	0	Large intestine
ERG	38383511	F446Lfs*59	0	Large intestine
ERG	38383511	F446Lfs*59	0	Large intestine
ERG	38383511	F446Lfs*59	0	Small intestine
ERG	38383511	A447Cfs*19	0	Stomach
ERG	38383511	A447Cfs*19	0	Stomach
ERG	38383494	N450I	0	Lung
ERG	38383492	P451S	1	Skin
ERG	38383492	P451S	1	Skin
ERG	38383489	Y452H	1	Stomach
ERG	38383485	W453L	0	Skin
ERG	38383484	W453*	0	Skin
ERG	38383484	W453*	0	Skin
ERG	38383483	N454H	0	Large intestine
ERG	38383468	G459C	0	Central nervous system
ERG	38383467	G459D	0	Central nervous system
ERG	38383467	G459D	0	Large intestine
ERG	38383459	P462A	1	Biliary tract
ERG	38383459	P462T	0	Stomach
ERG	38383440	T468I	0	Pancreas
ERG	38383430	M471I	0	Lung
ERG	38383426	S473A	0	Skin
ERG	38383425	S473F	0	Central nervous system
ERG	38383421	H474Q	1	Stomach
ERG	38383416	G476V	1	Lung

Supplemental References

1. Geoffroy, V. et al. AnnotSV: an integrated tool for structural variations annotation. Bioinformatics 34, 3572–3574 (2018).

2. Tudini, E. et al. Shariant platform: Enabling evidence sharing across Australian clinical genetic-testing laboratories to support variant interpretation. Am. J. Hum. Genet. 109, 1960–1973 (2022).

3. Karczewski, K. J. et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 581, 434–443 (2020).

4. Robinson, J. T. et al. Integrative genomics viewer. Nat. Biotechnol. 29, 24–26 (2011).

5. Ioannidis, N. M. et al. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. Am. J. Hum. Genet. 99, 877–885 (2016).

6. Nishii, R. et al. Molecular basis of ETV6-mediated predisposition to childhood acute lymphoblastic leukemia. Blood 137, 364–373 (2021).

7. Wiel, L. et al. MetaDome: Pathogenicity analysis of genetic variants through aggregation of homologous human protein domains. Hum. Mutat. 40, 1030–1038 (2019).

8. Bohne A, Schlee C, Mossner M, Thibaut J, Heesch S, Thiel E, et al. Epigenetic control of differential expression of specific ERG isoforms in acute T-lymphoblastic leukemia. Leuk Res. 2009;33: 817–822.

9. Sondka, Z. et al. COSMIC: a curated database of somatic variants and clinical data for cancer. Nucleic Acids Res. 52, D1210–D1217 (2024).

10. Zhou, X. et al. Exploring genomic alteration in pediatric cancer using ProteinPaint. Nat. Genet. 48, 4–6 (2016).

11. Greene, D. et al. Genetic association analysis of 77,539 genomes reveals rare disease etiologies. Nat. Med. 29, 679–688 (2023).