Germline ERG haploinsufficiency defines a new syndrome with cytopenia and hematological malignancy predisposition

Jiarna R Zerella[1,2], Claire C Homan[1,2,3#], Peer Arts[1,2,3#], Xuzhu Lin[4#], Sam J Spinelli[1], Parvathy Venugopal[1,3], Milena Babic[1,3], Peter J Brautigan[1,3], Lynda Truong[4], Luis Arriola-Martinez[1], Sarah Moore[3], Rachel Hollins[3], Wendy T Parker[1,3], Hung Nguyen[3], Karin S Kassahn[2,3], Susan Branford[2,3], Simone Feurstein[5], Lise Larcher[6], Flore Sicre De Fontbrune[7], Serwet Demirdas[8], Sonja de Munnik[9], Helene A Poirel[10], Benedicte Brichard[11], Sahar Mansour[12,13], Kristiana Gordon[12,14], ERG Variants Research Network, Marcin W. Wlodarski[15,16], Ashwin Koppayi[17], Sara Dobbins[12], Pim G N J Mutsaers[18], Kim E. Nichols[19], Ninad Oak[19], Desiree DeMille[20], Rong Mao[20,21], Ali Crawford[22], Julie McCarrier[23], Donald Basel[23], Josue Flores-Daboub[24], Michael W Drazer[25], Kerry Phillips[26], Nicola K Poplawski[26], Graeme M Birdsey[27], Daniela Pirri[27], Pia Ostergaard[12], Annet Simons[9], Lucy A Godley[17], David M Ross[1,2,28,29], Devendra K Hiwase[2,28,29], Jean Soulier[6], Anna L Brown[1,2,3], Catherine L Carmichael[4,30], Hamish S Scott[1,2,3,31]*, Christopher N Hahn[1,2,3]*.

equal second author
* equal senior authorship

Affiliations:

1.) Centre for Cancer Biology, SA Pathology and University of South Australia, Adelaide, SA, Australia.
2.) Adelaide Medical School, University of Adelaide, Adelaide, SA, Australia.
3.) Department of Genetics and Molecular Pathology, SA Pathology, Adelaide, SA, Australia.
4.) Centre for Cancer Research, Hudson Institute of Medical Research, Clayton, VIC, Australia.
5.) Department of Internal Medicine, Section of Hematology, Oncology & Rheumatology, University Hospital Heidelberg, Heidelberg, Germany.
6.) Université Paris Cité, Inserm and Hôpital Saint-Louis, APHP, Paris, France.
7.) Service d'hématologie-greffe, Hôpital Saint-Louis, Université Paris Cité, Paris, France.
8.) Department of Clinical Genetics Erasmus Medical Center Rotterdam The Netherlands.
9.) Department of Human Genetics, Radboud University Medical Center, Nijmegen, the Netherlands.
10.) Sciensano, Rue Juliette Wytsman 14, 1050 Brussels, Belgium.
11.) Department of Pediatric hematology and oncology, Cliniques universitaires Saint-Luc, Brussels, Belgium
12.) Cardiovascular and Genomics Research Institute, St George's University of London, London, United Kingdom.
13.) South West Thames Regional Centre for Genomics, St. George's Universities Hospitals NHS Foundation Trust, London, United Kingdom
14.) Dermatology & Lymphovascular Medicine, St. George's Universities NHS Foundation Trust, London, United Kingdom
15.) Department of Pediatrics and Adolescent Medicine, Division of Pediatric Hematology and Oncology, Medical Center, Faculty of Medicine, University of Freiburg, Freiburg, Germany
16.) Department of Hematology, St. Jude Children's Research Hospital, Memphis, TN
17.) Robert H. Lurie Comprehensive Cancer Center and Northwestern University.

18.) Department of Hematology, Erasmus Medical Centre Rotterdam, Rotterdam, The Netherlands.
19.) Department of Oncology, St. Jude Children's Research Hospital, Memphis, TN, USA.
20.) ARUP Institute for Clinical and Experimental Pathology, ARUP Laboratories, Salt Lake City, UT, USA.
21.) Department of Pathology, University of Utah, Salt Lake City, UT, USA.
22.) Illumina, Inc., San Diego, CA, USA.
23.) Department of Pediatrics, Medical College of Wisconsin, Milwaukee WI, USA.
24.) Pediatrics, University of Utah School of Medicine, Salt Lake City, UT, USA.
25.) Section of Hematology/Oncology, Department of Medicine, The University of Chicago, IL, USA.
26.) Adult Genetics Unit, Royal Adelaide Hospital, Adelaide, SA, Australia.
27.) National Heart and Lung Institute, Imperial College London, London, United Kingdom.
28.) South Australian Health & Medical Research Institute (SAHMRI), Adelaide, SA, Australia.
29.) Royal Adelaide Hospital, Central Adelaide Local Health Network, Adelaide, SA, Australia.
30.) Department of Molecular and Translational Medicine, Faculty of Medicine, Nursing and Health Sciences, Monash University, Clayton, VIC, 3168
31.) ACRF Genomics Facility, Centre for Cancer Biology, Adelaide, SA, Australia.

Christopher N. Hahn (chris.hahn@sa.gov.au)
Hamish S. Scott (hamish.scott@sa.gov.au)

Data Sharing Statement

For original data, please contact chris.hahn@sa.gov.au or hamish.scott@sa.gov.au

Shortened title

ERG predisposes to cytopenias and blood cancers

Key Points

Germline ERG loss-of-function variants predispose to cytopenias and hematological malignancies.

Somatic genetic rescue of ERG pathogenic variants in hematopoietic tissues impacts diagnosis, disease severity and potential for correction.

Abstract

The genomics era has facilitated discovery of new genes predisposing to bone marrow failure (BMF) and hematological malignancy (HM). We report the discovery of *ERG* as a novel autosomal dominant BMF/HM predisposition gene. ERG is a highly constrained transcription factor critical for definitive hematopoiesis, stem cell function and platelet maintenance. ERG colocalizes with other transcription factors including RUNX1 and GATA2 on promoters/enhancers of genes orchestrating hematopoiesis. We identified a rare heterozygous *ERG* missense variant in 3 thrombocytopenic individuals from one family and 14 additional *ERG* variants in unrelated individuals with BMF/HM including 2 *de novo* cases and 3 truncating variants. Phenotypes associated with pathogenic germline *ERG* variants included cytopenias (thrombocytopenia, neutropenia, pancytopenia) and HMs (acute myeloid leukemia, myelodysplastic syndrome, acute lymphoblastic leukemia) with onset before 40 years. Twenty *ERG* variants (19 missense, 1 truncating) including 3 missense population variants were functionally characterized. Thirteen potentially pathogenic ETS domain missense variants displayed loss-of-function characteristics disrupting transcriptional transactivation, DNA-binding and/or nuclear localization. Selected variants overexpressed in mouse fetal liver cells failed to drive myeloid differentiation and cytokine-independent growth in culture, and to promote acute erythroleukemia when transplanted into mice, concordant with these variants being loss-of-function. Four individuals displayed somatic genetic rescue by copy neutral loss of heterozygosity. Identification of predisposing germline *ERG* variants has clinical implications for patient/family diagnosis, counselling, surveillance, and treatment strategies including selection of bone marrow donors or cell/gene therapy.

Manuscript
Introduction

Tightly controlled regulation of hematopoiesis is essential to ensure adequate supply of healthy blood cells, and ability to respond to increased demand. Transcription factors (TFs), including ERG, TAL1, LYL1, LMO2, GATA2, RUNX1, MEIS1, PU.1, FLI1 and GFI1B form part of a network that, in varying combinations, control regulation of hematopoietic stem/progenitor cells.[1,2] This TF network is important for regulating cellular self-renewal, lineage specification, differentiation, and migration. Dysregulation can lead to cellular malfunction, impaired differentiation programs, and aberrant stem cell self-renewal, all of which have implications for human disease. Highlighting this, recurrent somatic variants within several of these TFs, including chromosomal rearrangements, point mutations, insertions and deletions, are detected in hematological malignancies (HM). Further, germline pathogenic variants in *RUNX1* (Familial platelet disorder with predisposition to myeloid malignancies, RUNX1-FPD; MONDO: 0011071), *GATA2* (GATA2 deficiency with susceptibility to MDS/AML; MONDO:0042982) and *FLI1* (Bleeding disorder, platelet-type, 21; OMIM 617443) are associated with development of hematologic disorders typified by increased risk for cytopenias and/or hematological malignancies (HMs). Roles for other members of this key hematopoietic TF network in predisposing to HM remain to be identified.[3]

The ETS-related gene (ERG), a member of the ETS TF family, was first reported as critical for normal hematopoiesis in 2008.[4] Like most ETS TFs, ERG demonstrates various homeostatic functions, binding to specific GGA(A/T) motifs to regulate genes in hematopoietic as well as

non-hematopoietic contexts. This includes binding at gene regulatory sites with other key hematopoietic transcriptional regulators including HM predisposition genes *GATA2* and *RUNX1*.[5] In normal hematopoiesis, ERG is essential for maintaining quiescence and preventing differentiation of hematopoietic stem cells (HSC)[6], promoting HSC self-renewal after hematopoietic stress (*e.g.* BM transplantation),[7] and supporting definitive hematopoiesis, adult HSC function and maintenance of peripheral blood platelet numbers.[4]

The perturbation of hematopoiesis by aberrant ERG expression and the contribution to HM via *ERG* overexpression has been well documented[8–18], particularly in acute myeloid leukemia (AML)[19–22]. Consequences of dysregulated *ERG* are also evident in other diseases including cardiovascular disease,[23–26] prostate cancers,[27,28], Ewing's sarcoma[29] and B-cell acute lymphoid leukemia (ALL)[30], the latter three via chromosomal translocations and genomic rearrangements. The first report of disease due to a germline *ERG* variant was heterozygous *Erg* (S322P) causing thrombocytopenia in a mouse model.[4] More recently, germline loss-of-function *ERG* variants have been associated with predisposition to primary lymphedema.[31]

We report for the first time, germline *ERG* variants in patients with a range of malignant and non-malignant hematological phenotypes. This discovery marks the identification of the third ETS TF with autosomal dominant pathogenic germline variants and HM and/or BMF (in addition to *ETV6*[32] and *FLI1*[33,34]) and adds to the growing list of master hematopoietic TFs already included on germline targeted sequencing panels for HM predisposition (*RUNX1, CEBPA, GATA2, ETV6, MECOM, PAX5, IKZF1*).[35] Our findings suggest a pathogenic role of *ERG* haploinsufficiency/hypomorphic actions, in contrast to the prototypical oncogenic nature of *ERG,* defining “ERG Deficiency Syndrome” as a new disease entity.[9] This paradox implies that the consequence of dysregulated *ERG* may not be consistent across all spatial and temporal cellular contexts, a phenomenon that is not uncommon in hereditary HM predisposition (*e.g. GATA2*).[36] It also implies a strict expression threshold, where dysregulation of *ERG* may upset the stoichiometry within a TF complex, leading to predisposition and/or initiation of HM-related disease. We systematically examined the functional implications of *ERG* variants on DNA binding, subcellular localization and transactivation of gene expression, and focused on specific variants to demonstrate their effect on *ERG*-mediated myeloid differentiation and cytokine independence *ex vivo,* and impact on *ERG*-driven leukemia in an *in vivo* murine model.

Methods

Gene discovery. Samples were obtained from Australian Familial Haematological Conditions Study (AFHCS), which was approved by the Women's and Children's Health Network Human Research Ethics Committee (Adelaide; HREC approval 2020/HRE00981)).

Genomic DNA was extracted (QiaAmp DNA mini kit, Qiagen), exonic sequences captured using xGEN (Integrated DNA Technologies), and libraries sequenced using NextSeq 550 (Illumina) to an average depth of 50x (hair) and 100x (blood). In addition, PCR-free short-read whole genome sequencing (WGS) was performed at the Australian Genome Research Facility (AGRF, Melbourne, Australia).

Identification of additional ERG variants in hematological cohorts. To identify additional kindreds with blood phenotypes and rare *ERG* variants, we contacted collaborators with existing genomic data from routine clinical testing and research studies. Genematcher[37] was used to further expand our patient cohort. All patient samples and data were covered by local institutional Human Research Ethics Committees in accordance with the Declaration of Helsinki.

Luciferase assay. An Integrin, Alpha 2b (*ITGA2B*) in pGL4.10-Luc was kindly donated by Dr Marie-Christine Kopp (University of Sydney).[33] K562 cells were seeded, transfected (Lipofectamine 2000) and then lyzed after 20 h (Dual-luciferase reporter kit, Promega). pUC18 was used to normalize the amount of DNA transfected. Luciferase levels were measured using the Explorer Multimode Microplate Reader (Promega).

Immunofluorescence staining. COS-7 cells were transiently transfected with pcDNA3.1-ERG(Myc-tag) WT or variant expression vector, fixed after 20 h (4% paraformaldehyde) and probed with anti-Myc antibody (9B11; New England Biolabs) and Alexa Fluor 488 Rabbit anti-mouse (A27023) secondary antibody. Fluorescent cells (100) were scored as 'nuclear' (protein only in nucleus) or 'cytoplasmic' (protein in nucleus and cytoplasm).

Electromobility shift assay (EMSA). HEK293 cells were seeded, transiently transfected (Lipofectamine 3000) and lysates prepared (RIPA 9806S, CST). Both biotin labeled and unlabeled double stranded DNA oligonucleotides containing an ERG binding site were synthesized (5'Biotin-GGCACTCACTTCCGGCTTGGCCGTCGA-3'). EMSAs were performed using Lightshift Chemiluminescent EMSA kit (Thermofisher) with 20 pmol probe.

ERG overexpression in fetal liver cells (FLCs). FLCs harvested from E14.5 WT C57BL/6 mouse embryos were transduced (Supplemental Methods) with MSCV-IRES-mCherry based retrovirus containing either ERG WT or variants (P116R, M219I, D345N R370P, Y372*, Y373C).[8] Transduced cells were cultured in StemSpan™ (Stem Cell Technologies) supplemented with Il3 (10ng/ml), Il6 (10ng/ml), Flt3L (50 ng/ml), Scf (50 ng/ml) and Tpo (50 ng/ml). FACS analysis was performed weekly for 4-6 weeks. For cytokine independence assays, all cytokines were removed when cell populations reached 80-90% mCherry^{+}. Cell viability and survival were measured every 3 days for 12 days.

In vivo leukemia model driven by ERG overexpression. MSCV-ERG-IRES-mCherry retrovirus transduced FLCs (above) were cultured for 3 days before intravenous injection into

sub-lethally irradiated 8-week old C57BL/6 mice. Mice received neomycin water for 3-weeks post-irradiation and their blood was monitored every 2 weeks for mCherry expression.

Results

Identification of BMF and/or HM families and individuals with rare ERG variants

Family 1 presented with a range of hematological abnormalities which included thrombocytopenia, neutropenia and AML (Figure 1A,B, Supplemental Figure 1). Patient I-2 developed AML at age 27. The patient entered morphological remission, later developing therapy-related MDS and died at age 36 (Supplemental Figure 1). A gene panel (Supplemental Table 1) on all 3 affected family members showed no germline pathogenic variants in known BMF/HM predisposition genes. Analysis of whole exome sequencing on the unaffected father confirmed the absence of any germline variants that may explain the phenotypes seen in both children. Cytogenetic analysis identified a constitutional mosaic trisomy 8 in patient I-2 (Supplemental Table 2) which was not present in either of her children, and may have contributed to myeloid malignancy progression. A pathogenic somatic RUNX1 (G165V) variant was identified in individual II-1 at a low VAF (2%) that may be an early indicator/marker of clonal progression to malignancy. Platelet morphology studies in Individual II-1 identified minor platelet abnormalities including a slight increase in alpha granule numbers, slight dilation of the open canalicular system and mildly enlarged platelets, despite no noticeable mean platelet volume (MPV) abnormalities (Supplemental Figure 2A,B).

Whole exome sequencing on all 4 individuals revealed heterozygosity of a novel *ERG* variant (chr21g.38383725A>G (hg38); c.1118A>G; p.Y373C) in the 3 affected individuals altering a highly conserved amino acid and segregated with thrombocytopenia (Table 1, Supplemental Figure 3A,B). This variant was absent in the general population (gnomAD)[38] and predicted to disrupt DNA binding (Supplemental Figure 3C). Interestingly, SNP-arrays identified copy neutral loss-of-heterozygosity (cnLOH) events on chromosome 21q favoring the *ERG* wildtype (WT) copy in all three carriers, strongly suggesting somatic genetic rescue (SGR) of the germline *ERG* deleterious variant (c.1118A>G; VAF 30% (I-2), 17% (II-1) and 44% (II-2)) (Figure 1C, Supplemental Figure 3D). Intriguingly, cnLOH encompassed the entire *RUNX1* gene in two patients (I-2 and II-2), and the recombination breakpoint in the most telomeric event in Individual II-1 was within the *RUNX1* gene (Figure 1C). With the addition of whole genome sequencing, we confirmed the presence of the cnLOH events, and the absence of deleterious variants in *RUNX1* including non-coding and structural variations that might explain the phenotype (Supplemental Figure 4). Note, cnLOH towards WT is not a commonly described mechanism of somatic reversion in RUNX1-FPD.[39]

Through international collaborations and GeneMatcher[37] (genotype matching resource), we identified an additional 13 rare *ERG* heterozygous variants associated with BMF/HM-related disease and 2 associated with lymphedema identified in a primary lymphoedema cohort in the 100,000 Genomes Project.[40] This included 15 probands; 12 were confirmed germline including 2 *de novo* (Table 1, Supplemental Table 2, Supplemental Figure 5), and germline samples were not available for the remaining three variants. Three variants were predicted to cause premature protein termination and 10 were missense variants that clustered within the ETS domain (310-395 aa). For 4 missense variants, REVEL scores indicated pathogenicity (score >0.85; Table 1, Supplemental Figure 6). To determine the prevalence of rare *ERG* variants in different study cohorts, we tabulated the number of patients screened, phenotypes encompassed, and number of *ERG* variants found, including multiple cohorts without rare *ERG* variants (Supplemental

Table 3). Because of the wide range of phenotypes covered by different research cohorts, associated ascertainment biases and inconsistent variant filtering strategies, determining the prevalence or penetrance of germline pathogenic *ERG* variants to different phenotypes remains to be determined with incorporation of additional well defined cohort studies and systematic analyzes.

In vitro functional characterization of ERG variants

The impact of *ERG* variants on transactivation, DNA binding and nuclear localization was examined (Figure 2). Western blot analysis showed all ERG variants produced protein, except for Y372*, which was unstable (Supplemental Figure 7). To assess transactivation ability, assays were performed using a platelet-specific *ITGA2B* promoter-luciferase reporter in K562 myeloid cells. Most ETS domain variants showed either complete LOF (S322P, R370H/P/S, Y372*, Y373C, Y388C) or were hypomorphic (M341V, D345N, D363A) compared to WT (Figure 2A). In contrast, a somatic variant (R385H), reported in multiple cancers in Catalogue Of Somatic Mutations In Cancer (COSMIC),[41] and R302C increased transactivation. *ERG* variants observed more than once in the population (*i.e.* >0.00012% gnomAD)[27] (Supplemental Figure 8) showed no impact on transactivation ability in this assay (Figure 2A, Table 1).

Electromobility shift assays were performed to measure the effect of ERG variants on DNA binding (Figure 2B, Supplemental Figure 9). Several ETS domain variants (S322P, R370H/P/S, Y372*, Y373C) entirely ablated DNA binding, consistent with R370 and Y373 being critical contacts for DNA binding (Supplemental Figure 3C, 10).[42]

Immunofluorescence was used to quantify the effect of ERG variants on subcellular localization (Figure 2C). R302C and P306L displayed an increased nuclear to cytoplasmic ratio, but their effect was not statistically significant. Conversely, ETS domain variants R370H/P/S, Y372*, Y373C, K380 and, R385H reduced nuclear protein localization (*$p<0.05$, Figure 2C, Supplemental Figure 11A). Equivalent amino acid residues to ERG (R370, Y373) in this highly conserved ETS domain in ETV6 (R399)[43] and FLI (Y343),[44] respectively, are similarly critical for proper nuclear localization of these TFs (Supplemental Figure 11B).

ERG LOF variants fail to drive myeloid differentiation and expansion of cytokine-independent stem/progenitors in fetal liver cell cultures

Retroviral-driven expression of ERG WT in fetal liver cells (FLCs) drives expansion of an immature stem/progenitor cell population with megakaryocytic features, as demonstrated by intermediate $cKIT^+$ and high $CD41^+$ cell surface expression[8] (Figure 3A,B). Several representative variants including 3 complete LOF, 1 hypomorphic variant and 2 WT-like variants from *in vitro* assays were chosen to study their effect on *ex vivo* ERG-driven megakaryocytic expansion and cytokine independent cell growth. Like WT, P116R, a population (*i.e.* benign) variant (M219I) and a hypomorphic variant (D345N) also drove this expansion. Strikingly, ETS domain complete LOF variants (R370P, Y372*, Y373C), after 4-5 weeks in culture, did not show expansion of this immature stem/progenitor cell population, replicating non-transduced FLCs or FLCs transduced with an empty MSCV-IRES-mCherry control retrovirus (EV).

Furthermore, we observed that ERG WT, P116R, M219I and D345N were each able to induce cytokine independence within FLCs after 6 weeks in culture, driving cell survival in the absence of essential cytokines. Conversely, ETS domain complete LOF variants (R370P,

Y372*, Y373C) were not able to drive cytokine independent growth, with all cells dying within 9 days of cytokine removal (Figure 3C) replicating non-transduced FLCs.

Complete LOF ERG variants fail to drive leukemia in an ERG overexpression mouse model
Overexpression of ERG in murine FLCs followed by transplantation into sub-lethally irradiated C57BL/6 mice leads to development of a well-described erythro-megakaryocytic leukemia characterized by an accumulation of immature erythroblasts that infiltrate the bone marrow, spleen, liver, and lung.[8,10,45] Enforced expression of ERG WT, a WT-like variant (P116R) and a population variant (M219I) in murine FLCs similarly drove erythro-megakaryocytic leukemia development (Figure 4A), by mCherry engraftment (Figure 4B), within 220 days following transplantation in 10/15 animals. Moribund mice displayed a large liver and spleen (Supplemental Figure 12A) and flow cytometric analysis of bone marrow and spleen revealed a $cKit^+CD71^+mCherry^+$ leukemic cell population consistent with ERG-driven erythro-megakaryocytic leukemia (Figure 4C,D), as previously described.[8] Histology further confirmed leukemic cell infiltration in bone marrow, spleen and liver (Supplemental Figure 12B). In contrast, all mice (20/20) receiving ERG ETS domain complete (R370P, Y372*, Y373C) or hypomorphic (D345N) variants showed no signs of disease or $mCherry^+$ peripheral blood cells in the first 220 days demonstrating LOF in this *in vivo* assay. Notably, one variant (D345N) is hypomorphic in transactivation assays (Figure 2A), acts WT-like in cytokine independence assays (Figure 3A,C) and is complete LOF in this murine erythroleukemia assay (Figure 4A). Hence, choice of assay(s) is critical in ERG functional studies, especially when applying to variant classification for diagnostic and clinical applications.

Classification of germline ERG variants and their phenotypes.
Based on clinical, *in silico* predictive and functional data, all *ERG* variants were classified using American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG-AMP) criteria with heterozygous LOF variants (including missense) and resulting haploinsufficiency as the disease mechanism[46] (Supplemental Table 2). ERG ETS domain is highly conserved (average constraint score 0.26; Supplemental Figure 13, Supplemental Table 4),[47] and LOF variants across the protein are rare in the normal population (LOEUF=0.23 and pLI=1, gnomADv4.0.0; Supplemental Figure 8)[38] suggesting intolerance of haploinsufficiency.

Five variants associated with BMF/HM (R370H, R370P, Y372*, Y373C, K380N) demonstrated complete LOF in one or more functional assays and were classified as Pathogenic, while Y388C (lymphedema) reached Likely Pathogenic (Figure 5, Supplemental table 5). Consistent with other BMF/HM predisposition genes, we observed variability in clinical presentation associated with these variants. Thrombocytopenia, neutropenia or pancytopenia was diagnosed in all patients with a likely pathogenic/pathogenic *ERG* variant (Table 1, Supplemental Figure 14A, Supplemental Table 2). Of this cohort, 28% developed a HM (AML and/or MDS) with a median age of onset 29 years (Supplemental Figure 14B). Longitudinal monitoring of Individuals 10, 11, 15 (II-1) and 16 (II-2) consistently showed thrombocytopenia (platelets $<150x10^9/L$) and neutropenia ($<2.0x10^9/L$) (Supplemental Figure 14C,D).

Ten variants were classified as variants of uncertain significance (VUS), including 4 (M341V, D345N, D363A, G394W) that demonstrated complete LOF or were hypomorphic in transactivation, DNA binding and/or erythroleukemia assays, and 1 variant (V127Efs*82)

which is likely to be pathogenic based on complete LOF of other premature termination variants (*i.e.* Y372* and lymphedema variants[31]) (Figure 5). Seven VUSs were associated with HM (MDS, AML, ALL, CLL, DLBCL), 3 with BMF and 1 with lymphedema. Note, one thrombocytopenic patient carrying a germline I126T variant also harbored a germline pathogenic *RUNX1* whole gene deletion (Supplemental Table 2) likely explaining the phenotype; whether the *ERG* variant also contributes to the phenotype is unclear. Two variants (D345N and D363A) could not be verified as germline due to unavailable samples, and have been reported multiple times as somatic, with D363 as a "hotspot" of somatic variant in multiple cancers (Supplemental Figure 15, Supplemental Table 6). Somatic analysis of bone marrow or blood identified variants in Patients 2, 3, 7, 14, 15 and 16, two of whom (3, 16) acquired *IDH1* variants; no variants were found in Patients 11 and 12 (Table 1).

Somatic ERG variants in HM and other cancers.
COSMIC was filtered for *ERG* variants occurring ≤ 3 times (*i.e.* very rare) in gnomAD v4.0.0. There were 64 unique ETS domain missense variants. Screening for somatic missense variants at the same amino acids affected by germline ETS domain variants in this study, revealed all except one (Y373C) to be seen somatically (Supplemental Figure 15), adding weight to these amino acid substitutions being drivers of malignancy.

In a separate pediatric familial cancer cohort (St Jude Children's Research Hospital), a single germline *ERG* (V127Efs*82) variant was identified (Table 1) as well as several predicted pathogenic somatic *ERG* variants, predominantly in children with B-ALL, but also AML, NHL and some solid cancers (Supplemental Table 6). Interestingly, ERG (V127Efs*82) was identified both as a germline and somatic variant in different patients with B-ALL.

Discussion

We report discovery of *ERG* as a new autosomal dominant BMF and HM predisposition gene, expanding its disease phenotypes beyond recently reported primary lymphedema.[31] *ERG* clinical disease resembles that of *RUNX1*,[48] *ETV6*[49] and *ANKRD26*[50] (thrombocytopenia, myeloid and lymphoid HM), *GATA2*[36,51] (BMF, myeloid HM, lymphedema), and *ERG's* most closely related ETS family member, *FLI1* (thrombocytopenia, but to date without reports of HM).[33,34,52,53] These overlapping, but non-identical, phenotypes are consistent with many of these TFs being part of a complex homeostatic network that is crucial for normal hematopoiesis.[5] Further, consistent with a role for *ERG* and *GATA2* in the lymphatics predisposing to lymphedema, in single cell studies, there is high expression of *ERG* and *GATA2* in bone marrow endothelial cells, in contrast to *RUNX1* where no primary lymphatic phenotype has been described.[54]

Intriguingly, cnLOH favoring the WT allele was identified in all three *ERG* (Y373C) carriers in index Family 1. SGR events such as these can complicate the detection of germline variants from hematopoietic tissues by mimicking the lower VAFs of somatic variants. Consistent with this, individuals with both Y388C and S182Afs*22 (recently published familial germline *ERG* variant in a patient with lymphedema[31]) had cnLOH, but no hematological disease (Supplementary Figure 16). Notably, the patient with S182Afs*22 presented with 4% VAF in the peripheral blood,[31] which upon further investigation was due to a cnLOH event correcting 92% of cells. SGR may lead to asymptomatic carriers, milder symptoms, later age of onset or missed diagnosis ("hidden" predisposition), and reveal potential for disease prevention via cell

and/or gene therapy.[55,56] Hence, identified *ERG* variants should be tested in "germline" non-hematopoietic tissues such as fibroblasts, mesenchymal stromal cells or hair to confirm germline or somatic status.

Functional studies demonstrated that all rare *ERG* variants within the ETS domain resulted in complete/hypomorphic LOF in transactivation, DNA binding and/or subcellular localization assays, and selected variants similarly exhibited LOF in fetal liver hematopoietic cell growth assays *ex vivo* and leukemogenic assays *in vivo*. This contrasts the gain-of-function nature of common *ERG* fusions[19–21,27,28] and hence shows that *ERG* can act as both a tumor suppressor gene and an oncogene. Notably, disease manifestations stemming from germline LOF mechanisms extend to other members of the ETS family (*ETV6* and *FLI1*)[33,34,52,53]. *ETV6* and *FLI1* variants often cluster in the ETS domain[32], and several are analogous to pathogenic *ERG* variants in our study (Supplemental Figure 11B). For example, a single consanguineous family with 2 siblings affected with moderate thrombocytopenia and a lifelong bleeding history was described with a rare FLI1 ETS domain missense variant in homozygosity (R324W, NM_002017.4).[33] This variant (analogous to ERG R354) was hypomorphic in i*n vitro* analyzes[33] similar to 3 variants in our study in the same region (M341V, D345N, D363A) (Figure 2A). To date, we have not seen patients with homozygous *ERG* germline variants or evidence of autosomal recessive disease although this might be possible for hypomorphic variants.

Structural modelling shows that several complete LOF variants (in DNA binding and transactivation assays) affect residues that contact DNA (Supplemental Figure 9).[42] Additionally, the nuclear localization (NLS) 2 region of the FLI1 protein is identical to this ETS domain region in ERG[44,57]. Unsurprisingly, 6 variants within this paralogous region were unable to appropriately localize to the nucleus (Figure 2C, Supplemental Figure 10). Therefore, because of the high conservation of the ETS domain in paralogous TFs, reported "function-disrupting" variants in one ETS TF are likely to affect protein function at the corresponding amino acid position in others. Indeed, paralogous amino acids cause similar loss of DNA binding and nuclear localization, and several have been classified in ClinVar as pathogenic for their relevant disease phenotypes. Pathogenic variants and their functional analyzes in highly conserved paralogous genes and domains, such as ETS TFs and ETS domains, is topical for variant curation expert panel discussion in generating gene-specific guidelines.

In addition to germline variants, *ERG* somatic variants are found in a proportion of sporadic cancers including HM (B-ALL, MDS, DLBCL, AML, T-ALL) and solid cancers (predominantly skin, gastrointestinal, breast, lung). Interestingly, the majority of the missense variants in HMs are located within the ETS domain with several variant "hotspots" including ones we tested (*i.e.* D345N, D363A), which were hypomorphic in our system; this raises the likelihood of a strict threshold of ERG activity in certain cellular contexts including concurrent with other somatic variants. For germline *ERG* HM cases, as for *GATA2*, we have not identified somatic variants on the other allele (*i.e.* biallelic) as occurs for *RUNX1*,[58] *CEBPA*[58,59] and *ETV6*[43]).

Our data establishes a gene-disease association for *ERG* in the pathogenicity of BMF and HM in addition to lymphedema. Six ETS domain variants in this study were classified as pathogenic or likely pathogenic, while 10 variants were VUS (Figure 5). Addition of functional data criteria changed ACMG classification of 5 *ERG* variants (3 VUS to likely pathogenic, 2 VUS

to benign) (Supplemental Table 5), highlighting the importance of generating faithful functional assays. For the VUSs, despite their seemingly WT-like behavior in functional assays, it remains plausible that some may have critical functional consequences that were not detected in our overexpression systems. The functional consequences of these variants may impact via temporal and spatial mechanisms, and therefore caution is required when interpreting "WT-like" VUS because protein-protein interactions or target binding may vary in different cellular contexts. It is probable though, that some variants classified as VUS (*i.e.* I126T and P116R) are not monogenic fully penetrant variants, given their lack of functional consequence in overexpression assays used in this study.

Our data (Table 2, Figure 6) defines *ERG* as a new predisposition gene to be added to germline screening panels for BMF and HM syndromes. Demonstration of *de novo* germline variants emphasizes testing for inheritance for family planning and counselling. We stress the importance of screening true germline samples of patients and family members as we posit SGR and use of hematopoietic samples may mask asymptomatic or mildly symptomatic carriers, potentially missing more severe clinical presentation in other family members including perinatal and neonatal lethality as has been observed in *MECOM*[60,61]. Enticingly, SGR highlights potential for pre-emptive cell therapy and/or gene editing strategies to prevent or alleviate *ERG*-related disease in the blood stem cell compartment. We only observed several inherited cases (4 small families), which may indicate low penetrance, mild phenotypes and/or incomplete family data. However, of 3 families in the Genomics England "Rareservoir" study with primary lymphedema carrying *ERG* premature termination variants[31] together with 2 missense *ERG* variants reported here, 4 showed familial inheritance and one was *de novo*. Whether lymphedema is a highly penetrant phenotype and hence seen in family units remains unanswered. While *ERG* variants in lymphedema and HM phenotypes have not yet coincided, we anticipate with a larger, more defined cohort study, they will overlap, mirroring the history of *GATA2* variants and Emberger syndrome (lymphedema and MDS).[51]

Our current disease model (Figure 7) proposes that, in *ERG* carriers, interplay of physiological and environmental stressors impact critical threshold-sensitive gene expression/biological pathways required for normal hematopoiesis resulting in BMF, and adaptive (somatic genetic rescue) and/or maladaptive (hematological malignancy) selective processes. It is possible for both adaptive and maladaptive clones to be present in the same individual, wherein context dependent competitive fitness of the clone determines the physiological outcome - a "game of clones". Notably, *ERG* expression is high in primary human HSC and drops in transition to HPC[62] and during differentiation. Analysis of *ERG* downstream effectors and pathways impacted in primary bone marrow and blood cells in *ERG* carriers and BMF/HM patients is warranted to better understand disease initiation and progression. Clearly, description of disease mechanisms, phenotypic and mutational spectrum, as well as natural history of diseases caused by *ERG* germline variants, has only just commenced for what may well be another pleiotropic and protean transcriptopathy. As population scale genomic studies such as the UK Biobank and "All of Us" become popular, our study demonstrates that they should not become de rigueur. Careful clinical and laboratory observations with professional networking will remain important in describing new *ERG*-associated disease and other disease entities.

Acknowledgements
We would like to thank the families and individuals for their participation in these studies. We would also like to acknowledge the RUNX1 Research Program for their consumer voice and advocacy in germline thrombocytopenia and leukemia predisposition and the Monash University Histology Platform, Monash University Animal Research Platform, and Monash Health Translation Precinct (MHTP) Animal Facility and Flowcore.

This research was supported by funding from the National Health and Medical Research Council of Australia (APP1024215, APP1164601, APP1023059, APP1182318); Catalyst Grant for Model Organism/System Study through the Australian Functional Genomics Network (AFGN) Medical Research Future Fund (MRFF) and Genomic Futures Health Mission grant (MRF2007498); Cancer Council Project Research Project Grant (APP1125849); The Hospital Research Foundation (PA); Maddie Riewoldt's Vision (MRV0017 - fellowship PV); University of South Australia, Centre for Cancer Biology funding (PV, CCH); MRFF Early to Mid-Career Researchers (APP2023357); The Royal Adelaide Hospital Research Fund (fellowship - PA); The University of Adelaide Australian Government Research Training Program (RTP) Scholarship (JRZ); Leukaemia Foundation of Australia Strategic Ecosystem Research Partnerships (SERP) (ALB, LAG, CCH); Peter Nelson Leukemia Research Fellowship, Cancer Council of South Australia (CCH); the American Lebanese Syrian Associated Charities (ALSAC; KEN, NO); Swiss Federal National Fund for Scientific Research (CRSII5_177191/1); UKRI Medical Research Council (MR/P011543/1); British Heart Foundation (RG/17/7/33217, PG/20/16/35047). Identification of ERG-associated lymphedema samples was made possible through access to data in the National Genomic Research Library, managed by Genomics England Limited (a wholly owned company of the Department of Health and Social Care). The National Genomic Research Library holds data provided by patients and collected by the NHS as part of their care and data collected as part of their participation in research. The National Genomic Research Library is funded by the National Institute for Health Research and NHS England. The Wellcome Trust, Cancer Research UK and the Medical Research Council have also funded research infrastructure.

Authorship Contributions
JRZ wrote the manuscript and was involved in all aspects of the project including research design, manuscript preparation, and collecting/analyzing experimental and clinical data; ALB, CLC, HSS, CNH were involved in research design, data analysis, manuscript writing and preparation, American College of Medical Genetics and Genomics (ACMG)-variant classification; and providing scientific insight. CCH, PA, PV contributed to different aspects of design and analysis of clinical and experimental data and manuscript writing and preparation. XL, SJS, LT helped collect experimental data. MB, PJB helped with sample preparation, LAM, KSK performed bioinformatic analyzes and SM, RH, WT, MRH and HN performed cytogenetic, somatic NGS and/or diagnostic arrays for Family 1. SB and MWW were involved in figure generation. SF, LL, FSF, SD, SM, HAP, BB, SM, KG, AK, SD, PGM, KEN, NO, DD, RM, AC, JM, DB, JFD, MWD, KP, NKP, GMB, DP, PO, AS, LAG, DMR, DKH and JS provided clinical and variant information from patients. MWW and The ERG Variant Research Network; PB, PB, LF, JD, ARM, OSS, NVM, JN, and JS, confirmed the absence of rare *ERG* variants in their cohorts. All authors critically reviewed and approved the manuscript.

Disclosure of Conflicts of Interest
The authors declare no competing financial interests.

References

1. Wilson, N. K. *et al.* Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators. *Cell Stem Cell* **7**, 532–544 (2010).
2. Thoms, J. A. I. *et al.* Disruption of a GATA2-TAL1-ERG regulatory circuit promotes erythroid transition in healthy and leukemic stem cells. *Blood* **138**, 1441–1455 (2021).
3. Zerella, J. R. *et al.* Transcription factor genetics and biology in predisposition to bone marrow failure and hematological malignancy. *Front. Oncol.* **13**, (2023).
4. Loughran, S. J. *et al.* The transcription factor Erg is essential for definitive hematopoiesis and the function of adult hematopoietic stem cells. *Nat. Immunol.* **9**, 810–819 (2008).
5. Diffner, E. *et al.* Activity of a heptad of transcription factors is associated with stem cell programs and clinical outcome in acute myeloid leukemia. *Blood* **121**, 2289–2300 (2013).
6. Knudsen, K. J. *et al.* ERG promotes the maintenance of hematopoietic stem cells by restricting their differentiation. *Genes Dev.* **29**, 1915–1929 (2015).
7. Ng, A. P. *et al.* Erg is required for self-renewal of hematopoietic stem cells during stress hematopoiesis in mice. *Blood* **118**, 2454–2461 (2011).
8. Carmichael, C. L. *et al.* Hematopoietic overexpression of the transcription factor Erg induces lymphoid and erythro-megakaryocytic leukemia. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 15437–15442 (2012).
9. Salek-Ardakani, S. *et al.* ERG is a megakaryocytic oncogene. *Cancer Res.* **69**, 4665–4673 (2009).
10. Tsuzuki, S., Taguchi, O. & Seto, M. Promotion and maintenance of leukemia by ERG.

Blood **117**, 3858–3868 (2011).

11. Xie, Y. *et al.* Reduced Erg Dosage Impairs Survival of Hematopoietic Stem and Progenitor Cells. *Stem Cells* **35**, 1773–1785 (2017).
12. Nikolova-Krstevski, V. *et al.* ERG is required for the differentiation of embryonic stem cells along the endothelial lineage. *BMC Dev. Biol.* **9**, 72 (2009).
13. Liu, F. & Patient, R. Genome-wide analysis of the zebrafish ETS family identifies three genes required for hemangioblast differentiation or angiogenesis. *Circ. Res.* **103**, 1147–1154 (2008).
14. Marcucci, G. *et al.* Overexpression of the ETS-related gene, ERG, predicts a worse outcome in acute myeloid leukemia with normal karyotype: a Cancer and Leukemia Group B study. *J. Clin. Oncol.* **23**, 9234–9242 (2005).
15. Metzeler, K. H. *et al.* ERG expression is an independent prognostic factor and allows refined risk stratification in cytogenetically normal acute myeloid leukemia: a comprehensive analysis of ERG, MN1, and BAALC transcript levels using oligonucleotide microarrays. *J. Clin. Oncol.* **27**, 5031–5038 (2009).
16. Baldus, C. D. *et al.* Acute myeloid leukemia with complex karyotypes and abnormal chromosome 21: Amplification discloses overexpression of APP, ETS2, and ERG genes. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 3915–3920 (2004).
17. Zaliova, M. *et al.* ERG deletion is associated with CD2 and attenuates the negative impact of IKZF1 deletion in childhood acute lymphoblastic leukemia. *Leukemia* **28**, 182–185 (2014).
18. Clappier, E. *et al.* An intragenic ERG deletion is a marker of an oncogenic subtype of B-cell precursor acute lymphoblastic leukemia with a favorable outcome despite frequent IKZF1 deletions. *Leukemia* **28**, 70–77 (2014).

19. Martens, J. H. A. Acute myeloid leukemia: a central role for the ETS factor ERG. *Int. J. Biochem. Cell Biol.* **43**, 1413–1416 (2011).

20. Chen, S. *et al.* Ewing sarcoma with ERG gene rearrangements: A molecular study focusing on the prevalence of FUS-ERG and common pitfalls in detecting EWSR1-ERG fusions by FISH. *Genes Chromosomes Cancer* **55**, 340–349 (2016).

21. Kim, J. *et al.* Detection of FUS-ERG chimeric transcript in two cases of acute myeloid leukemia with t(16;21)(p11.2;q22) with unusual characteristics. *Cancer Genet. Cytogenet.* **194**, 111–118 (2009).

22. Weber, S. *et al.* Gain of chromosome 21 or amplification of chromosome arm 21q is one mechanism for increased ERG expression in acute myeloid leukemia. *Genes Chromosomes Cancer* **55**, 148–157 (2016).

23. Kalna, V. *et al.* The Transcription Factor ERG Regulates Super-Enhancers Associated With an Endothelial-Specific Gene Expression Program. *Circ. Res.* **124**, 1337–1349 (2019).

24. Shah, A. V., Birdsey, G. M. & Randi, A. M. Regulation of endothelial homeostasis, vascular development and angiogenesis by the transcription factor ERG. *Vascul. Pharmacol.* **86**, 3–13 (2016).

25. Sperone, A. *et al.* The transcription factor Erg inhibits vascular inflammation by repressing NF-kappaB activation and proinflammatory gene expression in endothelial cells. *Arterioscler. Thromb. Vasc. Biol.* **31**, 142–150 (2011).

26. Dufton, N. P. *et al.* Dynamic regulation of canonical TGFβ signalling by endothelial transcription factor ERG protects from liver fibrogenesis. *Nat. Commun.* **8**, 895 (2017).

27. Kissick, H. T. *et al.* The transcription factor ERG increases expression of neurotransmitter receptors on prostate cancer cells. *BMC Cancer* **15**, 604 (2015).

28. García-Perdomo, H. A., Chaves, M. J., Osorio, J. C. & Sanchez, A. Association between TMPRSS2:ERG fusion gene and the prostate cancer: systematic review and meta-analysis. *Cent European J Urol* **71**, 410–419 (2018).

29. Yin, T. *et al.* Gastrointestinal Ewing Sarcoma: A Clinicopathological and Molecular Genetic Analysis of 25 Cases. *Am. J. Surg. Pathol.* **48**, 275–283 (2024).

30. Zhang, J. *et al.* Deregulation of DUX4 and ERG in acute lymphoblastic leukemia. *Nat. Genet.* **48**, 1481–1489 (2016).

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

32. Feurstein, S. & Godley, L. A. Germline ETV6 mutations and predisposition to hematological malignancies. *Int. J. Hematol.* **106**, 189–195 (2017).

33. Stevenson, W. S. *et al.* Paris-Trousseau thrombocytopenia is phenocopied by the autosomal recessive inheritance of a DNA-binding domain mutation in FLI1. *Blood* **126**, 2027–2030 (2015).

34. Saultier, P. *et al.* Macrothrombocytopenia and dense granule deficiency associated with FLI1 variants: ultrastructural and pathogenic features. *Haematologica* **102**, 1006–1016 (2017).

35. Martin, A. R. *et al.* PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. *Nat. Genet.* **51**, 1560–1565 (2019).

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

37. Sobreira, N., Schiettecatte, F., Valle, D. & Hamosh, A. GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. *Hum. Mutat.* **36**, 928–930 (2015).

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

39. Glembotsky, A. C. *et al.* First description of revertant mosaicism in familial platelet disorder with predisposition to acute myelogenous leukemia: correlation with the clinical phenotype. *Haematologica* **105**, e535 (2020).

40. Mark Caulfield, Jim Davies, Martin Dennys, Leila Elbahy, Tom Fowler, Sue Hill, Tim Hubbard, Luke Jostins, Nick Maltby, Jeanna Mahon-Pearson, Gil McVean, Katrina Nevin-Ridley, Matthew Parker, Vivienne Parry, Augusto Rendon, Laura Riley, Clare Turnbull, Kerrie Woods. National Genomic Research Library. *National Genomic Research Library Protocol v5.1.* https://doi.org/10.6084/m9.figshare.4530893.v7 (2017).

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

42. Adamo, P. & Ladomery, M. R. The oncogene ERG: a key factor in prostate cancer. *Oncogene* **35**, 403–414 (2016).

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

44. Hu, W., Philips, A. S., Kwok, J. C., Eisbacher, M. & Chong, B. H. Identification of nuclear import and export signals within Fli-1: roles of the nuclear import signals in Fli-1-dependent activation of megakaryocyte-specific promoters. *Mol. Cell. Biol.* **25**, 3087–3108 (2005).

45. Thoms, J. A. I. *et al.* ERG promotes T-acute lymphoblastic leukemia and is transcriptionally regulated in leukemic cells by a stem cell enhancer. *Blood* **117**, 7079–7089 (2011).

46. Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: a

joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* **17**, 405–424 (2015).

47. Wiel, L. *et al.* MetaDome: Pathogenicity analysis of genetic variants through aggregation of homologous human protein domains. *Hum. Mutat.* **40**, 1030–1038 (2019).
48. Hayashi, Y., Harada, Y. & Harada, H. Myeloid neoplasms and clonal hematopoiesis from the RUNX1 perspective. *Leukemia* **36**, 1203–1214 (2022).
49. Zhang, M. Y. *et al.* Germline ETV6 mutations in familial thrombocytopenia and hematologic malignancy. *Nat. Genet.* **47**, 180–185 (2015).
50. 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).
51. 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).
52. Stockley, J. *et al.* Enrichment of FLI1 and RUNX1 mutations in families with excessive bleeding and platelet dense granule secretion defects. *Blood* **122**, 4090–4093 (2013).
53. Di Paola, J. & Porter, C. C. ETV6-related thrombocytopenia and leukemia predisposition. *Blood* **134**, 663–667 (2019).
54. McKinney-Freeman, S. & Hall, T. Evaluation of Ranzoni et al.: Integrative Single-Cell RNA-Seq and ATAC-Seq Analysis of Human Developmental Hematopoiesis. *Cell stem cell* vol. 28 357–358 (2021).
55. Miyazawa, H. & Wada, T. Reversion Mosaicism in Primary Immunodeficiency

Diseases. *Front. Immunol.* **12**, 783022 (2021).

56. Gutierrez-Rodrigues, F., Sahoo, S. S., Wlodarski, M. W. & Young, N. S. Somatic mosaicism in inherited bone marrow failure syndromes. *Best Pract. Res. Clin. Haematol.* **34**, 101279 (2021).
57. Lu, J. *et al.* Types of nuclear localization signals and mechanisms of protein import into the nucleus. *Cell Commun. Signal.* **19**, 60 (2021).
58. Brown, A. L., Hahn, C. N. & Scott, H. S. Secondary leukemia in patients with germline transcription factor mutations (RUNX1, GATA2, CEBPA). *Blood* **136**, 24–35 (2020).
59. Dufour, A. *et al.* Acute myeloid leukemia with biallelic CEBPA gene mutations and normal karyotype represents a distinct genetic entity associated with a favorable clinical outcome. *J. Clin. Oncol.* **28**, 570–577 (2010).
60. Niihori, T. *et al.* Phenotypic heterogeneity in individuals with MECOM variants in 2 families. *Blood Adv* **6**, 5257–5261 (2022).
61. Venugopal, P. *et al.* Unravelling Facets of MECOM-Associated Syndrome: Somatic Genetic Rescue, Clonal Hematopoiesis and Phenotype Expansion. *Blood Adv* (2024) doi:10.1182/bloodadvances.2023012331.
62. Subramanian, S. *et al.* Genome-wide transcription factor-binding maps reveal cell-specific changes in the regulatory architecture of human HSPCs. *Blood* **142**, 1448–1462 (2023).
63. Homan, C. C. *et al.* Somatic mutational landscape of hereditary hematopoietic malignancies caused by germline variants in RUNX1, GATA2, and DDX41. *Blood Adv* **7**, 6092–6107 (2023).
64. Antony-Debré, I. *et al.* Somatic mutations associated with leukemic progression of familial platelet disorder with predisposition to acute myeloid leukemia. *Leukemia* **30**,

999–1002 (2016).

65. Sahoo, S. S. *et al.* Clinical evolution, genetic landscape and trajectories of clonal hematopoiesis in SAMD9/SAMD9L syndromes. *Nat. Med.* **27**, 1806–1817 (2021).

Tables

Table 1. Rare heterozygous *ERG* variants and population variants. Rare *ERG* variants (NP_891548.1) from similar phenotypic groups including BMF and/or HM and lymphedema, amongst population variants (gnomAD >200)[38], a germline thrombocytopenic mouse mutation[1], a paralogous *ETV6* pathogenic variant (thrombocytopenia)[49] and COSMIC mutation (somatic)[41]. Information unavailable (blank); gnomAD v4.0.0, alternative variants at the same amino acid position (brackets). Not available (NA), Not determined (ND), Mesenchymal stem cells (MSC), Peripheral blood (PB), Skin fibroblasts (SF).

Patient ID	Gender	ERG variant (479 aa)	VAF (%)	gnomAD v4.0.0	REVEL	Hematological-related Phenotype	Non-hematological Phenotype	Age onset of first phenotype (yr)	Germline ERG (inherited/ *de novo*) (Sample)	Somatic variants	Chr21q cnLOH
1	M	E20Vfs*13	60	0	-	MDS		38	Yes (MSC)		No
2	F	P116R	56	69 0.004%	0.36	Thrombocytopenia, Thrombocytopathy, platelet aggregation disorder	Hypertension, diabetes, cataract	73			No
3		V127Efs*82	50	0	N/A	ALL		<18	Yes (NA)		
4		I126T	59	1 0.00016%	0.512	Chronic thrombocytopenia AML	Not reported in clinical records	<14 40	Yes (SF)	FLT3-ITD, IDH1	NA
5	M	R302C	40	10 0.00068% (1x R302L) (17x R302H)	0.365	CLL	Not reported in clinical records	Unknown	Yes (Hair)		NA
6	F	P306L	44	0 (6x P306W)	0.394	None	Lymphedema	<1	Yes (inherited) (PB)		No
7	M	P306L	57	0 (6x P306W)	0.394	None	Lymphedema		Yes		ND

8	F	M341V	45	0	0.45	Severe congenital aplasia and abnormal B cells.	Prematurity for acute fetal distress (33 weeks)	0	Yes (inherited) (SF)	None	
9	F	M341V	45	0	0.45	MDS		56			No
10		D345N	24	1 0.00012%	0.3149	MDS					No
11		D363A	34	0	0.6209	MDS					No
12	M	R370H	45	0	0.881	Neutropenia		0	Yes (SF)	None (Blood at 18 yr)	NA
						Pancytopenia		18			
13	M	R370P	44	0	0.888	MDS (asymptomatic thrombocytopenia and leukopenia)	Deformation (avascular necrosis) of femoral head, severe aortic valve insufficiency with secondary heart failure	29	Yes (Hair)	None	No
14	M	Y372*	48	0	n/a	Congenital pancytopenia, bone marrow failure.		0	Yes (*de novo*) (PB)		
15	F	Y373C	30	0	0.852	AML, Thrombocytopenia, t-MDS (RAEB2)	Not reported in clinical records	27	Yes (Hair)	IDH1 (R132C) (40%) TP53 (C238Y) (22%) GATA2 (H442Qfs*95) (20%)	Yes
16	M	Y373C	17	0	0.852	Thrombocytopenia, neutropenia	Not reported in clinical records	21	Yes (inherited) (Hair)	RUNX1 (G165V) (2%)	Yes (2 events)
17	M	Y373C	44	0	0.852	Thrombocytopenia, neutropenia	Not reported in clinical records	19	Yes (inherited) (Hair)	None	Yes
18		K380N	46	0	0.642	Anemia, Thrombocytopenia, Pancytopenia, Macrocytic anemia, Abnormality of the spleen, Eosinophilic infiltration of the	Hemangioma, Hepatosplenomegaly, Gastrointestinal inflammation, Erythema, Capillary	>18	Yes (*de novo*) (PB)		No

						esophagus.	malformation.				
19	F	Y388C	50	0	0.90	None	Lymphedema	8	Yes (inherited) (PB)		No
20	F	Y388C	33	0	0.90	None	.Lymphedema (Bilateral intermittent lower limb swelling)	50	Yes (PB)		Yes
21		G394W	15	0 (3x G394R)	0.661	t-AML after treatment for DLBCL and prostate cancer	DLBCL and prostate cancer		Yes (PB)		No
Population		M219I		372 0.023% (1 hom)	0.058	n/a			Yes		
Population		P275S		254 0.016% (3 hom)	0.223	n/a			Yes		
Mouse		S322P		0	0.574	ALL, thrombocytopenia			Yes		
ETV6		R370S		0	0.803	ALL			No		
COSMIC		R385H		0 (2x R385C)	0.674	n/a	2x BrCa, 1x Biliary, 1x Upper aerodigestive tract		No		
Population		P404A		236 0.015%	0.086	n/a			Yes		

Table 2. Summary of functional assays for phenotypic *ERG* variants in this study. *ERG* variants are ordered in descending order of transactivation ability. Ticks indicate degree of WT-like activity in transactivation, DNA binding, subcellular localization, FLC myeloid differentiation, FLC cytokine independence and leukemogenesis assays. X indicates complete LOF in these assays. Not done, ND.

Variant	Phenotypes	Transactivation	DNA binding	Subcellular localization	FLC myeloid differentiation	FLC cytokine independence	Leukemogenesis assay	ACMG classification
WT	N/A	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	N/A
I126T	AML	✓✓✓✓	✓✓✓✓	✓✓✓✓	ND	ND	ND	VUS
P116R	Cytopenia	✓✓✓✓	✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	VUS
R302C	CLL	✓✓✓✓	ND	✓✓✓✓	ND	ND	ND	VUS
K380N	Cytopenia	✓✓✓✓	✓✓✓	✓	ND	ND	ND	P
P306L	Lymphedema	✓✓✓	ND	✓✓✓✓	ND	ND	ND	VUS
G394W	AML	✓✓✓	✓✓✓✓	✓✓✓✓	ND	ND	ND	VUS
M341V	Cytopenia	✓✓	✓✓✓✓	✓✓✓✓	ND	ND	ND	VUS
D345N	MDS	✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✕	VUS
D363A	MDS	✓✓	✕	✓✓✓✓	ND	ND	ND	VUS
Y388C	Lymphedema	✕	ND	✓✓✓✓	ND	ND	ND	LP
R370H	Cytopenia	✕	✕	✓✓	ND	ND	ND	P
R370P	Cytopenia + MDS	✕	✕	✓✓	✕	✕	✕	P
Y373C	Cytopenia + AML	✕	✕	✓	✕	✕	✕	LP
Y372*	Cytopenia	✕	✕	✓	✕	✕	✕	P
E20Vfs*13	MDS	ND	ND	ND	ND	ND	ND	VUS
V127Efs*82	ALL	ND	ND	ND	ND	ND	ND	VUS

Figure legends

Figure 1. Germline *ERG* variant identified in a family with hematological conditions. (A) Pedigree of Family 1 containing the ERG (Y373C) variant segregating with thrombocytopenia and AML. Diagnosed (dx); Died (d); Age of onset (years). (B) History of platelet and neutrophil counts of affected family members (Patient 15, 16, 17) (I-1, II-1, II-2). Absolute neutrophil/platelet counts per microlitre of blood from complete blood examinations (CBE) plotted with age (years). Normal lower limit for neutrophils/platelets is marked with a dotted red line. (C) Log R Ratio and B-allele plots of SNP-array analysis. All individuals show a copy neutral loss of heterozygosity (cnLOH) event which encompasses the entire *ERG* gene (yellow highlight). Individual II-1 (Patient 16) had a second cnLOH event (Blue highlight). Log R Ratios (top panels) show no loss of copy number, and B-allele frequencies (bottom panels) show regions of LOH, together demonstrating cnLOH. Samples used: I-2 - bone marrow cytogenetic pellet; II-1 and II-2 - peripheral blood (PB) mononuclear cells.

Figure 2. Functional characterization of *ERG* variants. (A) ERG variants residing in the 'ETS domain' reduce transactivation. K562 cells were transfected with pcDNA3 empty vector (EV) or pcDNA3-*ERG* (WT or variants). All constructs were co-transfected with a luciferase reporter plasmid driven by an *ITGA2B* promoter (quadruplicate replicates, repeated 3 times). Fold change (mean ± S.E.M.) compared to the WT is plotted. Pairwise comparisons are shown (* $P < 0.05$ compared to WT). (B) ETS domain variants reduce ERG DNA binding affinity. Electromobility shift assay (EMSA) of ERG WT and variants. Transfected HEK293 whole cell lysates were prepared and bound to an oligonucleotide containing an ETS DNA consensus sequence. Probes were visualized using chemiluminescence. Pairwise comparisons are shown (* $P < 0.05$ compared to WT). (C) ERG ETS domain variants alter subcellular localization. A Myc-tag was added to ERG WT and variants. COS-7 cells were transfected with pcDNA3 empty vector (EV), pcDNA3-*ERG*-Myc (WT) and pcDNA3-*ERG*-Myc variants. Cells were stained for a Myc-tag and DAPI. The nuclear:cytoplasmic ratio of each variant was quantified and the fold-change (mean ± S.E.M.) compared to the WT is plotted. Pairwise comparisons are shown (* $P < 0.05$ compared to WT). Nuclear:cytoplasmic ratio is obtained from 3 independent experiments. In all comparisons, a Student's *t-test* was used.

Figure 3. ERG loss-of-function variants fail to drive megakaryocytic differentiation and cytokine independence. (A) *In vitro* differentiation of fetal liver cells (FLCs) overexpressing ERG WT and variants. Immature stem/progenitor cell population with megakaryocytic features were stained using antibodies against $cKIT^+$ and $CD41^+$, with the percentage of double positive cells measured by flow cytometry (6 independent experiments). Pairwise comparisons are shown (* $P < 0.05$ compared to WT). In all comparisons, a Student's *t-test* was used. (B) Example of gating strategy for $cKIT^+$ and $CD41^+$ expression on ERG WT or variant transduced FLCs. Cells were first gated for viability and mCherry expression. (C) Cell viability of FLC culture after cytokine removal measured over time.

Figure 4. *ERG* variants are loss-of-function in an *in vivo* leukemia model driven by *ERG* overexpression. (A) Enforced expression of ERG WT and WT-like variants (P116R, M219I) in mice led to development of erythro-megakaryocytic leukemia within 220 days. (B) mCherry engraftment over time in peripheral blood of mice. Numbering refers to three ERG WT mice succumbing to disease (see 5A). (C) Infiltration of ERG WT/$mCherry^+$ transplanted FLCs into recipient mouse bone marrow (BM), spleen and peripheral blood as measured by FACS

analysis. Recipient mouse #3 shown is representative of all ERG WT and WT-like (P116R, M219I) mice. (D) ERG WT and WT-like variant mice developed mCherry$^+$ erythro-megakaryocytic leukemia with cKit$^+$ and CD71$^+$ expression. FACS plots of a representative ERG WT leukemia are shown. Non-leukemic (*i.e.* mCherry$^-$) BM and spleen cells are shown as comparison.

Figure 5. ACMG classification of germline *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 and BS3 criteria were not applied where variants showed no change in functional assays or were not tested.

Figure 6. Functional consequences of rare ERG variants. Rare *ERG* variants from similar phenotypic groups including BMF and/or HM and lymphedema, amongst population variants (gnomAD >200)[38], a germline thrombocytopenic mouse variant[1], a paralogous *ETV6* pathogenic variant (thrombocytopenia)[49] and COSMIC mutation (somatic)[41] are mapped onto the ERG protein (isoform, NP_891548.1; transcript, NM_182918.4). Functional characterization of each variant via transactivation, DNA binding, subcellular localization, FLC myeloid differentiation, FLC cytokine independence and leukemogenesis assays are displayed.

Figure 7. Proposed model for ERG deficiency syndrome-associated phenotypes. Adaptive/maladaptive events in ERG heterozygous carriers. (top panel) It is proposed that *ERG* carriers harbor a threshold of activity that is insufficient under certain physiological and environmental stressors leading to a range of phenotypic outcomes. Malignant cells (yellow cells). Maladaptive and adaptive events observed in our study (arrow). Other potential (maladaptive/adaptive) somatic events seen in other BMF and/or HM predisposition genes[63–65] (dotted arrow) or hypothetical events (question mark preceding dotted arrow). Somatic genetic reversion (SGR); bone marrow failure (BMF), copy neutral loss of heterozygosity (cnLOH). **Game of clones.** (bottom panel) The development of different physiological outcomes in patients is the result of a "game of clones", where context-dependent competitive fitness of the clone(s) determines progression and observed outcome. Notably, adaptive and maladaptive events, and subsequent clonal selection, may occur simultaneously. The clonal output shown is based on mature blood cells, reflecting collectively the impact on mature, and stem and progenitor cells.