Published in final edited form as:

Nat Med. 2021 October ; 27(10): 1806–1817. doi:10.1038/s41591-021-01511-6.

Corresponding author: Marcin W Wlodarski, MD, PhD, 262 Danny Thomas Place, Memphis, 38103 TN, USA, Phone: +901-595-2484, Fax: +901-595-2176, marcin.wlodarski@stjude.org.

*A list of authors and their affiliations appears at the end of the paper.

AUTHOR CONTRIBUTIONS

MWW, SSS and CMN designed the research; VBP, SSS, CG, AS, DL, PKP, JW and RDD performed genomic studies and analyzed data; SSS, EK, AS and CG accomplished functional studies; MWW, RKV, PN, ME, BS and CMN analyzed and interpreted clinical data; PN performed clinical statistical analysis; MD, JS, FL, RM, MS, BDM, AC, KK, DT, HH, JB, KJ, MU, SP, OPS, OF, SB, VH, IB, SSF, MRN, MGS, BB, PL, PB, RB, IM, MHA, RM, AS, GC, SH, GG, AYN, MC, ME, CF, BS, CMN and MWW were involved in patient care, testing and data presentation; SSS, MWW and CMN wrote the manuscript. All authors contributed to the manuscript and approved its final version.

COMPETING INTERESTS

Jochen Buechner declares personal fees, advisory board/ steering committee honorarium and/or nonfinancial support from Novartis, Pfizer, Kite and Janssen Pharma corporations. Peter Bader declares research grants from Neovii, Reimser and Medac to his affiliated institution; is on the advisory board for Novartis, Cellgene, Amgen, Medac and Servier both for personal and on behalf of his affiliated institution, and is part of the speaker's bureau for Miltenyi, Jazz, Reimser, Novartis and Amgen on behalf of his affiliated institution and holds Patent and receives Royalties from Medac. Robert Durruthy-Durruthy is employed by Mission Bio and owns equity in Mission Bio.

*EWOG-MDS (European Working Group of Childhood MDS) Consortium Members

Michael Dworzak[4], Jan Starý[5], Riccardo Masetti[7], Markus Schmugge[8], Barbara De Moerloose[9], Albert Catala[10], Krisztián Kallay[11], Dominik Turkiewicz[12], Henrik Hasle[13], Jochen Buechner[14], Kirsi Jahnukainen[15], Marek Ussowicz[16], Sophia Polychronopoulou[17], Owen Smith[18], Shlomit Barzilai[20], Valérie De Haas[21], Irith Baumann[22], Stephan Schwarz-Furlan[22,23], Gudrun Gohring[37], Charlotte Niemeyer[2,38], Karin Nebral[39], Ingrid Simonitsch-Kluppp[40], Pascale De Paepe[41], Nadine Van Roy[42], Vit Campr[43], Zuzana Zemanova[44], Erik Clasen-Linde[45], Tine Plesner[45], Brigitte Schlegelberger[38], Martina Rudelius[46], Kalliopi Manola[47], Kalliopi Stefanaki[48], Judit Csomor[49], Hajnalka Andrikovics[50], David Betts[51], Maureen O'Sullivan[52], Yaniv Zohar[53], Marta Jeison[54], Rita De Vito[55], Francesco Pasquali[56], Jadwiga Maldyk[57], Olga Haus[58], Helena Alaiz[59], Paula Kjollerstrom[60], Luis Mascarenhas de Lemos[61], Ivana Bodova[62], Martin Čermák[63], Lukas Plank[64], Barbara Gazic[65], Marko Kavcic[66], Helena Podgornik[66], Margarita Llavador Ros[67], Jose Cervera[68], Carole Gengler[69], Joelle Tchinda[70], Berna Beverloo[71,72], Roos Leguit[73].

[39]Labdia Labordiagnostik GmbH, Clinical Genetics, Vienna, Austria. [40]Department of Dermatology Clinical Institute of Pathology, Medical University of Vienna, Vienna, Austria. [41]Department of Pathology, AZ Sint-Jan Brugge-Oostende, Brugge, Belgium. [42]Center for Medical Genetics, Ghent University Hospital, Gent, Belgium. [43]Department of Pathology and Molecular Medicine, Charles University and University Hospital Motol, Prague, Czech Republic. [44]Center of Oncocytogenomics, Institute of Medical Biochemistry and Laboratory Diagnostics, General University Hospital and First Faculty of Medicine, Charles University, Prague, Czech Republic. [45]Department of Pathology, Copenhagen University Hospital, Copenhagen, Denmark.[46]Department of Pathology, University Hospital, LMU Munich, Munich, Germany. [47]Department of Biodiagnostic Sciences and Technologies, INRASTES, National Centre for Research 'Demokritos', Athens, Greece. [48]Department of Pathology, Aghia Sophia Children's Hospital, Athens, Greece. [49]First Department of Pathology and Experimental Cancer Research, Semmelweis University, Budapest, Hungary. [50]Laboratory of Molecular Diagnostics, Central Hospital of Southern Pest, Budapest, Hungary. [51]National Children's Cancer Service, Children's Health Ireland at Crumlin, Dublin, Ireland. [52]Department of Histology, Our Lady's Children's Hospital, Dublin, Ireland. Department of Pathology, Rambam Medical Center, Haifa, Israel. [53]Department of Pathology, Rambam Medical Center, Haifa, Israel. [54]Cancer Cytogenetic and Molecular Cytogenetic Laboratory, Schneider Children's Medical Center of Israel, Petach Tikva, Israel.[55]Cancer Cytogenetic and Molecular Cytogenetic Laboratory, Schneider Children's Medical Center of Israel, Petach Tikva, Israel.[56]Medical Genetics, Department of Medicine and Surgery, University of Insubria, Varese, Italy. [57]Department of Pediatrics, Hematology and Oncology, Medical University of Warsaw, Warsaw, Poland.[58]Department of Clinical Genetics, Nicolaus Copernicus University, Bydgoszcz, Poland.[59]Laboratory of Hemato Oncology, Portuguese Oncology Institute Francisco Gentil, Lisbon, Portugal.[60]Pediatric Hematology Unit, Hospital Dona Estefânia, Centro Hospitalar Universitário de Lisboa Central, Lisbon, Portugal.**[61]**Department of Surgery and Human Morphology, Universidade NOVA de Lisboa, Lisbon, Portugal. [62]Bone marrow transplantation unit, Detská fakultná nemocnica spoliklinikou Bratislave, Bratislava, Slovakia.[63]Department of Genetics, The National Institute of Oncology, Bratislava, Slovakia.[64]Department of Pathological Anatomy, Comenius University in Bratislava, Bratislava, Slovakia.[65]Department of Pathology, Institute of Oncology Ljubljana, Ljubljana, Slovenia.[66]Unit of Oncology and Haematology, University Children's Hospital, Ljubljana University Medical Centre, Ljubljana, Slovenia.[67]Department of Pathology, Hospital Universitari i Politècnic la Fe, Valencia, Spain.[68]Department of Hematology, Genetics Unit and CIBERONC, Hospital Universitario y Politécnico La Fe, Valencia, Spain.[69]Institut universitaire de pathologie, Lausanne University Hospital (CHUV), Lausanne, Switzerland.[70]Laboratory for Oncology, University Children's Hospital Zürich, Zürich, Switzerland.[71]Department of Clinical Genetics, Erasmus MC, University Medical Center Rotterdam, Rotterdam, the Netherland.[72]Dutch Childhood Oncology Group, Princess Máxima Centre, Utrecht, Netherlands.[73]Department of Pathology, University Medical Center Utrecht, Utrecht, Netherlands.

Clinical Evolution, Genetic Landscape, and Trajectories of Clonal Hematopoiesis in SAMD9/SAMD9L Syndromes

A full list of authors and affiliations appears at the end of the article.

Abstract

Germline *SAMD9* and *SAMD9L* mutations (*SAMD9/9L*^{mut}) predispose to myelodysplastic syndromes (MDS) with propensity for somatic rescue. Here, we investigated a clinically annotated pediatric MDS cohort (n=669) to define the prevalence, genetic landscape, phenotype, therapy outcome and clonal architecture of SAMD9/9L syndromes. In consecutively diagnosed MDS, germline *SAMD9/9L*^{mut} accounted for 8% and were mutually exclusive with *GATA2* mutations present in 7%. Among *SAMD9/9L*^{mut} cases, refractory cytopenia was the most prevalent MDS subtype (90%), acquired monosomy 7 was present in 38%, constitutional abnormalities in 57%, and immune dysfunction in 28%. The clinical outcome was independent of germline mutations. In total, 67 patients had 58 distinct germline *SAMD9/9L*^{mut} clustering to protein middle regions. Despite inconclusive *in silico* prediction, 94% of *SAMD9/9L*^{mut} suppressed HEK293 cell growth, and mutations expressed in CD34+ cells induced overt cell death. Furthermore, we found that 61% of *SAMD9/9L*^{mut} patients underwent somatic genetic rescue (SGR) resulting in clonal hematopoiesis, of which 95% was maladaptive (monosomy 7 +/− cancer mutations) and 51% had adaptive nature (revertant UPD7q, somatic *SAMD9/9L*^{mut}). Finally, bone marrow single-cell DNA sequencing revealed multiple competing SGR events in individual patients. Our findings demonstrate that SGR is common in *SAMD9/9L*^{mut} MDS and exemplify the exceptional plasticity of hematopoiesis in children.

Keywords

MDS predisposition; *SAMD9* ; *SAMD9L* ; somatic genetic rescue

INTRODUCTION

Sterile alpha motif domain protein 9 (SAMD9) and its paralogue SAMD9-like (SAMD9L) are cytoplasmic proteins encoded by 2 juxtaposed single-exon genes on chromosome 7q21. They share a 60% amino acid sequence identity and likely originated from a duplication of a common ancestral gene[1]. Their function remains enigmatic; they have been linked to tumor suppression[2], inflammation[3], stress response[4], development[4], endosomal fusion[5,6], and protein translation[7,8]. Both proteins were also shown to function as restriction factors forming a cross-species barrier for poxvirus infection[9–12]. Structural analysis of these large proteins has been limited to homology modeling, which predicted identical domains in either protein (SAM, ALBA2, SIR2, P-loop/NTPase, and OB-fold)[13]. Moreover, these genes exhibit tight regulation during embryonic development and transition to ubiquitous expression levels in adult tissues[14,15].

Notably, Samd9l-haploinsufficient mice develop myeloid neoplasia mimicking human myelodysplastic syndrome (MDS) with monosomy 7[5]. Several groups reported germline *SAMD9* or *SAMD9L* mutations (*SAMD9/9L*^{mut}) underlying new human syndromes with a propensity for cytopenia, bone marrow failure (BMF), and MDS with non-random monosomy 7 or deletion of 7q[6,16–28]. *SAMD9* mutations (*SAMD9*^{mut}) were initially linked to a fatal, early-onset condition with Myelodysplasia, Infections, Restriction of growth, Adrenal hypoplasia, Genital phenotypes, and Enteropathy (MIRAGE)[6,29]. In contrast, *SAMD9L* mutations (*SAMD9L*^{mut}) were originally described in families with a progressive neurological phenotype, multi-lineage cytopenia, and bone marrow hypoplasia (ataxia-pancytopenia syndrome)[16,17]. Recent reports broadened this spectrum and found missense *SAMD9/9L*^{mut} in non-syndromic familial MDS[30–33], truncating *SAMD9L*^{mut} in children with autoinflammatory panniculitis resembling Chronic Atypical Neutrophilic Dermatosis with Lipodystrophy and Elevated Temperatures (CANDLE) syndrome[34] or B-cell aplasia[8], and missense *SAMD9*^{mut} in steroid-resistant nephrotic syndrome[35]. Thus far, 64 mostly missense germline mutations (38 *SAMD9*; 26 *SAMD9L*), have been reported[6,7,16–32,34–37]. However, none of the studies were performed in population-based registries but rather using institutional cohorts driven by specific phenotypes. Here, we studied an international cohort of children consecutively diagnosed with primary MDS to assess the prevalence, hematological features, and clinical outcome of *SAMD9/9L*^{mut}. We also describe genotype and phenotype spectrum and depict clonal architecture of somatic genetic rescue.

RESULTS

Clinical characteristics and prevalence of germline mutations

Our first aim was to assess the prevalence, hematological features, and outcome of SAMD9/9L disorders, compared to GATA2 deficiency, a common germline driver in pediatric MDS. We first studied 548 children and adolescents consecutively diagnosed with primary MDS over a period of 18 years in Germany (cohort A, Fig. 1a). In this population-based cohort, 8% (42/548) of patients carried germline *SAMD9/9L*^{mut}, 7% (38/548) were *GATA2*-mutated (*GATA2*^{mut}), and 85% (468/548) were categorized as wildtype (Table 1). Among *SAMD9/9L*^{mut} patients, 90% (38/42) presented with refractory cytopenia of childhood (RCC) and 10% (4/42) with MDS with excess blasts (MDS-EB). In contrast, in *GATA2*^{mut} group, MDS-EB was more common (34%, 13/38). Next, we assessed the proportion of germline mutations within MDS subtypes (Fig. 1b, Supplementary Table 1). *SAMD9/9L*^{mut} had a similar prevalence across subtypes, while *GATA2*^{mut} was more frequent in MDS-EB compared to RCC (16% vs. 5%, $p<0.001$) (Fig. 1b). There was no difference in age at diagnosis, sex, blood counts, or bone marrow (BM) findings between *SAMD9*^{mut} vs. *SAMD9L*^{mut} or comparing *SAMD9/9L*^{mut} and wildtype cohorts (Table 1, Supplementary Table 2). Monosomy 7 and del(7q) (referred to as −7) was the predominant cytogenetic lesion, significantly enriched in *SAMD9/9L*^{mut} (38%) and *GATA2*^{mut} (57%) compared to wildtype (8%) (Table 1). Assessing the prevalence of germline mutations in MDS with −7, we found that half of the −7 cohort carried mutations in either *SAMD9/9L* (21%) or *GATA2* (30%) (Fig. 1b, Supplementary Table 1). The prevalence of *GATA2*^{mut} increased with age, while *SAMD9/9L*^{mut} were found at similar frequencies across all age groups (Fig. 1c).

Clinical outcome between 26 *SAMD9*^{mut} and 16 *SAMD9L*^{mut} cases appeared similar, with a 5-years overall survival (OS) of 84% and 93%, respectively (Fig. 1d). Further outcome analysis focused on the largest group; patients diagnosed with RCC. Normal karyotype predicted a 5-years OS of ≥90% (Fig. 1e). In contrast, the presence of −7 implied a less favorable outcome with a 5-years OS of approximately 70–80%, which appeared independent of the underlying germline predisposition (Fig. 1e).

Non-hematopoietic features

For extended studies shown henceforth, we included 25 additional *SAMD9L*^{mut} cases identified among 121 MDS patients from non-consecutive cohort B (Fig. 1a), increasing the number of *SAMD9/9L*^{mut} patients to 67 (Supplementary Table 3). Constitutional abnormalities were documented in 57% (38/67), with similar frequencies between *SAMD9* and *SAMD9L* genotypes (Extended Data Fig. 1). We found that clinical phenotypes previously considered gene-idiosyncratic, like endocrine features associated with *SAMD9* or neurologic problems with *SAMD9L*, can manifest with both genotypes. For example, central nervous system abnormalities, including cerebral atrophy, spina bifida, and meningocele were found in 3 *SAMD9*^{mut} patients. Vice versa, urogenital malformations were ascertained in 5 *SAMD9L*^{mut} patients of whom 2 (D1285, CH026) had features of MIRAGE syndrome. Immune abnormalities were recorded in 26% (10/38) of *SAMD9*^{mut} and 31% (9/29) of *SAMD9L*^{mut} patients (Supplementary Table 4).

Germline and somatic spectrum of *SAMD9/9L* mutations

Unless specified by expert panels i.e. ClinGen, variants in dominant genes fall within the pathogenic range if their gnomAD population minor allele frequency (MAF) is below 0.1%[38]. Using this threshold, we identified 67 individuals with *SAMD9/9L*^{mut} in cohorts A and B. To rule out other etiologies, whole exome sequencing (WES) was performed and an alternative genetic cause due to non-synonymous changes was found in only 1 patient (germline *PTPN11* in GR010) (Supplementary Table 5). Of the 67 patients, 49 (73%) had novel germline *SAMD9/9L* mutations, while 18 (27%) had variants present in gnomAD (Supplementary Table 6–7). We next defined a set of clinical features specific to SAMD9/9L syndromes (Fig. 2a). Remarkably, these features were associated only with novel or ultra-rare variants (<0.005% MAF), found in 49 and 8 subjects, respectively. Since the full SAMD9/9L disease spectrum is unknown, we retained all variants with population MAF <0.1%, hereby establishing a cohort of 67 (38 *SAMD9*^{mut} and 29 *SAMD9L*^{mut}) patients with 58 distinct germline variants. According to ACMG criteria, 72% (42/58) were categorized as variants of uncertain significance (VUS), 19% (11/58) as likely pathogenic, and 9% (5/58) as pathogenic (Supplementary Table 6). The majority (85%, 49/58) of variants were singletons (Fig. 2b). Nearly all patients (96%, 64/67) carried a single variant, 1 patient (D991) had 2 germline *SAMD9L*^{mut} and 2 (I363, D1248) had variants in both genes. Of note, 86% (50/58) of variants had not been reported in human disease. Clustering analysis revealed significant over-representation of germline variants (48%, 28/58) between amino acid 635 and 998, encompassing predicted P-loop/NTPase domains specified here as middle region (Fig. 2c, Supplementary Table 8). Multiple smaller mutational clusters were found in both genes. In one-quarter (16/67) of the cohort we also found second-site somatic *SAMD9/9L*^{mut} with random distribution. In 9 analyzed cases, these somatic mutations were

found to be *in cis* with the germline *SAMD9L*^{mut} (Fig. 2b; Supplementary Table 6). Somatic mutations split evenly into truncating and missense alterations (Fig. 2d). In comparison, 91% of the germline *SAMD9/9L*^{mut} were missense.

SAMD9/9L mutational effect

Baseline *SAMD9/9L* mRNA and protein expression were comparable in patient and control fibroblasts (Fig. 3a). To determine if truncating variants alter gene expression, we studied fibroblasts with germline heterozygous *SAMD9L* R406X variant present at 0.2% MAF in gnomAD. RNA expression was unaffected and western blot revealed robust expression of wildtype and truncated proteins (Fig. 3a). Informed by these results that *SAMD9/9L*^{mut} do not affect expression, we next employed *in silico* prediction to evaluate putative variant effect. CADD algorithm predicted 66% (38/58) of *SAMD9/9L*^{mut} as deleterious and 34% (20/58) as benign (Supplementary Table 6). To understand if the high rate of benign prediction was *SAMD9/9L*^{mut} specific, we also studied *GATA2*^{mut}. Compared to gnomAD variants, patient *GATA2*^{mut} had significantly higher (pathogenic) rank scores across multiple algorithms (Fig. 3b, Extended Data Fig. 2a–b). In contrast, no such discrimination was achieved for *SAMD9/9L* variants, rendering prediction with applied tools ineffective.

Knowing that *SAMD9/9L*^{mut} can inhibit proliferation, we expressed mutant proteins in human embryonic kidney (HEK) 293 cell line. Wildtype and mutant proteins yielded similar expression levels (Extended Data Fig. 3a). Of the 48 germline *SAMD9/9L*^{mut} tested, 45 showed a significant growth-suppressive effect compared to wildtype, and 3 were neutral (Fig. 3c). Second-site somatic missense *SAMD9/9L*^{mut} co-expressed *in cis* with the respective patient's germline variant ameliorated its growth-suppressive effect (Extended Data Fig. 3b). Unexpectedly, when investigated alone, somatic missense *SAMD9/9L*^{mut} also inhibited growth. As controls, we used 15 recurrent gnomAD variants showing no significant effect (Fig. 3c, Supplementary Table 9). We next revisited the initial *in silico* scoring and could reclassify 82% of the patient variants predicted benign by CADD to likely pathogenic based on a significant *in vitro* readout (Extended Data Fig. 3c). To answer if *SAMD9/9L*^{mut} affect primary hematopoietic cells, we transduced CD34+ cells with SAMD9 E974K and SAMD9L V1512M lentiviruses. Both mutants caused significant growth arrest and apoptosis compared to wildtype (Fig. 3d, Extended Data Fig. 4a–b). Of note, overexpression of SAMD9L wildtype protein was also detrimental to cell survival.

Somatic rescue and cancer mutations in SAMD9/9L-mutated MDS

Somatic genetic rescue (SGR) is a post-meiotic mechanism by which the pathogenic germline allele is modified or lost. A prototypic rescue event in SAMD9/9L syndromes is the nonrandom loss of chromosome 7 (containing *SAMD9/9L*^{mut} allele), causing maladaptive SGR. All our *SAMD9/9L*^{mut} patients with −7 had decreased *SAMD9/9L*^{mut} allele frequency (median 33%) in hematopoietic cells (Extended Data Fig. 5). In addition, we found benign, adaptive SGR events including somatic *SAMD9/9L*^{mut} and uniparental isodisomy 7q (UPD7q) in one-third (21/64 evaluable) of the cohort (Extended Data Fig. 6a–c). Altogether, SGR events accounted for 61% (41/67) of *SAMD9/9L*^{mut} patients, of which 51% (21/41) had benign SGR and 95% (39/41) carried −7 (Fig. 4a–c). All benign SGR was linked to germline *SAMD9/9L*^{mut} localized within middle region and C-terminal

(Supplementary Table 10), and the *SAMD9L* R986 residue was affected in one-fourth (5/21) of these patients (Supplementary Table 6). While somatic *SAMD9/9L*^{mut} occurred in both genotypes, UPD7q was exclusive to *SAMD9L*^{mut} cases with −7 and correlated with younger age at diagnosis (Fig. 4a, Extended Data Fig. 7, Supplementary Table 11).

Somatic cancer (leukemia driver) gene mutations were detected in 30% (19/64 evaluable) of *SAMD9/9L*^{mut} patients, and the majority (84%, 16/19) of those had −7 karyotype (Fig. 4a). In total, we detected 39 mutations in 12 cancer genes, with recurrent hits in *SETBP1*, *ASXL1*, *RUNX1*, *EZH2*, *PTPN11*, *CBL*, and *ETV6* (Fig. 4a, Supplementary Table 12). We also looked at cancer mutations in other pediatric MDS subgroups and found similar mutational landscape and overlap with −7 across *SAMD9/9L*^{mut}, *GATA2*^{mut}, and wildtype MDS (Fig. 4d, Supplementary Table 11). Assessing for interconnections, almost all secondary changes arose in context of −7 (Fig. 4b, Extended Data Fig. 6c). To understand how cancer mutations evolve, we analyzed serial specimens in 14 *SAMD9/9L*^{mut} patients and found that 2 acquired new mutations, 3 retained existing mutations, while 9 remained negative (Fig. 4a, Supplementary Table 12).

Single-cell clonal architecture

Aiming to characterize clonal evolution in an exemplary *SAMD9*^{mut} patient, we sequenced single colony forming cells (CFC) obtained from serial BM specimens (Fig. 4e). All colonies carried −7; early subclonal *SETBP1* mutation was dominant and expanded over time, while late *ASXL1* mutation was short-lived. Analysis of −7 clonal size in 4 additional patients demonstrated rapid expansion of −7 clones in culture (Fig. 4f), precluding accurate clonal quantification using CFC assay. This prompted us to develop a single-cell DNA sequencing (scDNAseq) method allowing to detect and quantify mutations in *SAMD9/9L* and cancer genes, and simultaneously measure loss-of-heterozygosity (LOH) of 7q (Supplementary Table 13). Patient D1248 with 2 germline variants (*SAMD9L* and *SAMD9*) had 5 independent clones, including native state hematopoiesis, 2 clones with benign SGR (UPD7q, *cis* somatic *SAMD9L*^{mut}), and 2 clones with malignant SGR potential (−7 alone, −7 with additional *EZH2* mutation) (Fig. 4g). Clonal zygosity and ploidy were determined from genotype data (Fig. 4h, Extended Data Fig. 8a–c). Using this approach, we discovered chromosome 7 LOH due to −7 or copy neutral UPD7q (Fig. 4h). Multiclonal patterns were also evident in D769 and D991, each carrying 2 independent rescue clones with somatic *SAMD9*^{mut} or *SAMD9L*^{mut} (Extended Data Fig. 8a–c). Longitudinal scDNAseq in D769 revealed expansion of −7 clone from 6.3% to 16% at the expense of clone with somatic *SAMD9* R1529H mutation. Finally, clonal architecture in D1248, D778 (scDNAseq) and D130 (CFC) was consistent with a linear model of clonal evolution where ancestral −7 clone gives rise to subclones carrying additional mutations (Fig. 4e, Extended Data Fig. 8b). Notably, in 3 of 4 tested samples, scDNAseq uncovered lesions (UPD7q, *SAMD9* P1431L, and *MYB*) missed by bulk sequencing.

Association of genetic lesions with clinical outcomes

Integrating clinical phenotyping, karyotyping, bulk sequencing, and proof-of-concept scDNAseq, we inferred multiple clonal outcomes (Fig. 5). SGR with −7 showed 3 distinct routes: i) acquisition of somatic cancer mutations, ii) no mutations, and iii) disappearance

of the clone. In contrast, UPD7q and somatic *SAMD9/9L*^{mut}, both considered benign (adaptive) clonal hematopoiesis, had no cancer gene mutations. Between the 2 benign lesions, only UPD7q demonstrated complete rescue potential, as indicated by long-lasting clinical remission. The remaining 26 patients had no detectable SGR at chromosome 7 and/or within *SAMD9/9L* loci. Five of them developed clonal hematopoiesis: 2 with *ASXL1* mutations and 3 with karyotype progression (Fig. 5).

Finally, to understand how SGR might modify the disease outcome, we reviewed patients' clinical course and established 3 categories with distinctive features: stable disease (5/67), remission (5/67), and high-risk/progressed (21/67) (Extended Data Fig. 9, Supplementary Table 14). Stable disease group was defined by prolonged chronic cytopenia (>1 year) without the need for therapies. These patients were older (median 11.8 and 21.5 years at diagnosis and last follow up, respectively), had normal karyotype, and only 2 of them carried somatic *ASXL1* mutations. Remission, defined as resolution of hematologic symptoms, was seen in 5 patients, 2 (*SAMD9L*^{mut}) with transient −7 succeeded by UPD7q, and 3 (*SAMD9*^{mut}) with normal karyotype. These patients showed stable blood counts over a long period (median 11.6 years) and had no additional acquired changes (Extended Data Fig. 9). In the high-risk/progressed group (advanced MDS, somatic cancer mutations, disease progression) all patients required hematopoietic stem cell transplantation (HSCT) and 17/21 were alive at last follow up, despite a high prevalence of −7 (76%, 16/21) and somatic cancer mutations (81%, 17/21) (Fig. 4a, Supplementary Table 3). Comparing the age at diagnosis for these 3 categories, patients with remission were diagnosed at a younger age (1.6–4.3 years) than stable disease (9.7–12.7 years) or high-risk/progressed group (0.4–17.6 years) (Extended Data Fig. 10).

DISCUSSION

One of the hallmarks of primary MDS in childhood is an expected high rate of germline predisposition. GATA2 deficiency has been previously implicated as the major germline driver[39]; however, SAMD9/9L syndromes recently emerged as a new hereditary cause of early-onset cytopenia and MDS. A major gap in understanding these syndromes is the incomplete knowledge of the prevalence, genetic spectrum, and clinical consequences in the context of population-based cohorts. Investigating 548 children consecutively diagnosed with MDS over 18 years, we found that *SAMD9/9L* and *GATA2* mutations show a comparable prevalence of 8% and 7%, respectively. We also demonstrated that half of the MDS/−7 cases arise from germline *SAMD9/9L* or *GATA2* mutations. These findings confirm −7 as the strongest predictor of germline predisposition in pediatric MDS. Unlike *GATA2*^{mut} with their inherent high rate of advanced MDS, *SAMD9/9L*^{mut} cases presented predominantly with RCC and hypocellular marrow. In comparison, in a French national cohort of patients with suspected BMF syndrome[32], 16 of 86 patients carried germline *SAMD9/9L*^{mut}, while in a single-institution study, 8 of 77 children with MDS had SAMD9/9L disease[30]. These observations point to the dichotomy of these disorders that can mimic a BMF-like condition or manifest as advanced MDS.

Somatic genetic rescue (SGR) is rare in BMF syndromes and generally leads to phenotypic correction[40] (Supplementary Table 15). We demonstrated that SGR in SAMD9/9L

syndromes is highly prevalent (61%) and very diverse, pointing to the extraordinary plasticity of childhood hematopoiesis. The observed repertoire includes benign SGR with remission potential (UPD7q or somatic *SAMD9/9L*^{mut}) and maladaptive SGR (−7) with propensity for acquisition of somatic cancer mutations and progression. Among all Mendelian disorders, post-zygotic mosaicism has been estimated in up to 20–23% of cases[41,42]. This designates SAMD9/9L syndromes as diseases with the highest rate of somatic mosaicism known in humans. Although somatic rescue events were previously described in *SAMD9/9L*^{mut} patients, profiling of hematopoiesis to reflect mosaicism on single-cell level was unattainable. Using scDNAseq, we provide for the first time a snapshot into the complex clonal architecture, demonstrating the presence of SGR with multiple layers of functional redundancy (indicated by coexistence of independent populations with somatic *SAMD9/9L*^{mut}, UPD7q, −7 or −7 with additional cancer mutations in single individuals). At this point, we assume that not all these clones are sustained; some could be evolutionary dead-ends with insufficient propagating potential, while others might outgrow the dysfunctional *SAMD9/9L*^{mut} hematopoiesis.

In a first attempt to identify a genotype-phenotype correlation, we found that benign SGR occurs only in patients with germline *SAMD9/9L*^{mut} localized within middle region and C-terminal. Another striking finding is the exclusive correlation of UPD7q with *SAMD9L* genotype found in 7 patients; similarly, majority (11/14) of reported UPD7q cases had *SAMD9L*^{mut}[17,26,31–33]. Arguably, the stronger growth-suppressive effect of *SAMD9L* (observed *in vitro*) might drive the development of UPD7q. Conversely, *SAMD9*^{mut} potentially exert a less pronounced negative selective pressure in hematopoiesis. This goes along with our observation that remission unrelated to UPD7q and stable disease were associated with *SAMD9*^{mut} genotype. Identification of prognostic factors for remission is essential for therapy stratification. In our cohort, patients with UPD7q had earlier disease manifestation compared to UPD7q-negative patients, and 2/7 experienced long-lasting remission with the disappearance of −7 and no evidence for malignant transformation. While this suggests that UPD7q supports normal hematopoiesis without clonal exhaustion, lifetime surveillance will better define such patients' risks. Intriguingly, 4 patients with UPD7q were found to carry −7 with cancer mutations, and exemplary scDNAseq pointed to independent evolution of these lesions in separate clones. While we are still enthusiastic about UPD7q as a natural gene therapy, one must consider the substantial risk of leukemia coming from co-evolving −7 clones.

A novelty of our study contrasting the current *status quo* that most somatic *SAMD9/9L*^{mut} are truncating, is the equal frequency of missense and truncating somatic mutations. Unpredictably, somatic missense *SAMD9/9L*^{mut} when studied alone exerted growth-inhibitory effect *in vitro*. When co-expressed with the respective patient germline mutation, they showed compensatory ("neutralizing") effect, which is in line with the concept of *cis* suppression of disease mutations[43]. This observation was affirmed by scDNAseq studies suggesting that hematopoietic cells with somatic missense or truncating *SAMD9/9L*^{mut} clones outgrow their germline *SAMD9/9L*^{mut} ancestry. However, thus far no evidence exists for these somatic *SAMD9/9L*^{mut} to facilitate clinical remission.

Acquisition of somatic cancer mutations is common in MDS, and we found 30% of *SAMD9/9L*^{mut} patients, predominantly with −7 karyotype, to harbor such alterations. The mutational spectrum and association with −7 were comparable across the 3 MDS genotypes (*SAMD9/9L*^{mut}, *GATA2*^{mut}, and wildtype). This exemplifies that the cancer mutation landscape in pediatric MDS is independent of germline predisposition and rather dictated by −7. It is obvious that −7 represents a maladaptive rescue attempt to escape from the hematotoxic nature of *SAMD9/9L*^{mut}. While these cells are not malignant *per se*, they have a propensity to acquire cancer driving mutations, as observed in almost half of −7 *SAMD9/9L*^{mut} cohort. In a broader context, this opens the question of how −7 aberration transitions to a leukemic state and if epigenetic regulators on chromosome 7 play a role.

Predicting the pathogenicity of *SAMD9/9L* variants is challenging because i) these syndromes are novel and their phenotypes continuously expand, ii) most of the patient mutations are private, iii) functional knowledge is lacking, iv) *SAMD9/9L* genes have low mutational constraint with tolerance for missense variation in humans, and v) variable expressivity and reduced penetrance are common. Unlike *GATA2,* where disease-causing variants are absent in healthy controls, we found that 21% of germline *SAMD9/9L*^{mut} occur in the general population, albeit at a very low frequency. One approach for establishing a population frequency threshold that extracts pathogenic variants is to consider disease-specific phenotypes. While most of the patients with novel or ultra-rare variants (<0.005%) exhibited SAMD9/9L-specific findings, 10 patients carrying population variants that are more common (0.005–0.1%) lacked such features, suggesting these variants either have low penetrance or are not the primary disease drivers, or are not pathogenic. One of these patients (GR010) had germline *PTPN11* mutation associated with Noonan syndrome. However, he carried a *SAMD9*^{mut} with decreased allelic burden in hematopoiesis due to −7 and growth-suppressive effect *in vitro,* suggesting its deleterious nature. The recurrent pathogenic *SAMD9L* mutation R986H is found in 4/112952 non-Finnish European alleles, implying that 1 in ~14 thousand Europeans is a carrier. This observation could be attributed to the variable expressivity of *SAMD9/9L* alleles and points to the need of testing larger patient cohorts across all ages. The exact disease frequency in adults remains unknown and thus far, only 10 adult patients (diagnosed at age 21–58 years) had been reported[17,32,33]. It will be important to screen adults with SAMD9/9L-phenotype spectrum, i.e., MDS/−7 or ataxia.

We also identified a mutational cluster containing half of all germline *SAMD9/9L*^{mut} that we termed as the middle region. P-loop/NTPase encompasses most of this region, supporting its predicted functional relevance. P-loop/NTPases are present in 5–10% of all proteins and play a role in various cellular processes including apoptosis[44]. No function has yet been assigned to P-loop/NTPases in SAMD9/9L, and merely homology modelling established a link to STAND superfamily of NTPases[13]. A striking resemblance exists between SAMD9/9L and autosomal dominant inflammasome syndromes caused by gain-of-function (GOF) mutations located within P-loop/NTPase domains of *NLRP3*, *NLRP1,* and *NLRC4* genes[45].

It is noteworthy that while previously reported germline *SAMD9/9L*^{mut} were missense, a recent study found germline truncating *SAMD9L*^{mut} in 6 children with CANDLE

syndrome [34] (all developed cytopenia, 3 had marrow hypocellularity, 1 acquired −7, personal communication Dr. Goldbach-Mansky). We found germline truncating mutations in 4 *SAMD9*^{mut} and 1 *SAMD9L*^{mut} patients, thus expanding this new category. The effect of germline missense mutations was proposed as GOF. While this terminology is provisional, we infer from HEK293 assays that both missense and protein-truncating mutations show a GOF effect with respect to suppression of cell growth. When exemplary missense *SAMD9/9L*^{mut} were expressed in primary hematopoietic CD34+ cells, we observed severe growth suppression and cell death precluding further work on hematopoietic differentiation. Recent studies evaluating *in vitro* effect of 1 truncating[8] and 4 missense[7] patient variants confirmed our observations of increased cell death and also showed suppressed translational capacity of affected cells. Another study depicted reduced cellular proliferation and enhanced endocytosis in mouse hematopoietic cells with *Samd9L*_D764N mutation[46]. Taken together, the mechanisms by which *SAMD9/9L*^{mut} regulate cell growth remain elusive.

Specific constitutional abnormalities have been associated with *SAMD9*^{mut} (MIRAGE-endocrine features, growth retardation) and *SAMD9L*^{mut} (neurological anomalies). Hence, our observation of endocrine, urogenital, and growth anomalies in *SAMD9L*^{mut} patients was unexpected. This is the first study to describe MIRAGE phenotype in 2 *SAMD9L*^{mut} and recurrent neuropathology in several *SAMD9*^{mut} patients. Therefore, we propose to refer to these disorders as “SAMD9/9L syndromes”, rather than separate entities. Previously, a poor outcome from HSCT was observed in several *SAMD9*^{mut} children with MIRAGE syndrome[36,47]. In our study, there was no difference in outcome between *SAMD9*^{mut} and *SAMD9L*^{mut} groups, however due to the small size of our *SAMD9/9L*^{mut} MDS cohort, larger studies might be required to reach definitive conclusions. Further, we found that survival from HSCT was comparable between *SAMD9/9L*^{mut} and other MDS subgroups.

Our findings should be viewed in light of some limitations. Obvious caveats are the referral bias, possibly restricting the phenotypic spectrum and the lack of parental screening required for estimating penetrance and anticipation. Due to lack of information on race and ethnicity we were not able to factor these variables into the analysis. Although we employed deep sequencing panels supplemented by WES, we missed several small clones later picked up by scDNAseq. Accordingly, we are under detecting somatic events and the true prevalence of SGR in *SAMD9/9L*^{mut} patients is likely higher than the 61% reported here. Lastly, the HEK293 growth assay does not model a hematopoietic system and relies on artificial overexpression that cannot grade the variants based on their clinical severity. Modification of endogenous *SAMD9/9L* loci might provide insights into phenotypic effects of these variants, however, this is beyond the scope of the present study and warrants future explorations.

In conclusion, navigating through the complex genetic and clinical landscape of a large cohort of pediatric MDS we were able to advance the existing knowledge about SAMD9/9L syndromes by uncovering new genotypes and mutational clusters, cataloguing the phenotype spectrum, and demonstrating that these unique syndromes show the highest rates and diversity of clonal rescue hematopoiesis known in human disease.

ONLINE METHODS

Study cohort

The diagnosis of refractory cytopenia of childhood (RCC) or primary MDS with excess blasts (MDS-EB) was established according to the WHO classification for pediatric MDS[1]. All patients were enrolled in the prospective registry studies 1998 (NCT00047268) or 2006 (NCT00662090) of the European Working Group for MDS in Childhood (EWOG-MDS). University of Freiburg institutional ethics committee approved the research (CPMP/ICH/135/95 and 430/16). Written informed consent was obtained from patients and/or guardians. The consecutive cohort (cohort A) consisted of 548 children and adolescents with primary MDS diagnosed in Germany between July 1st, 1998, and December 31st, 2016 (Fig. 1a). To increase the number of *SAMD9L*^{mut} cases, we screened additional 121 pediatric MDS cases (cohort B) with clinical presentation suggestive of SAMD9/9L-related phenotypes. Of the 67 patients with germline *SAMD9/9L* mutations (cohort A and B), 7 had MDS-EB at the time of diagnosis, including 1 patient (PL082) with an initial BM blast count of 22% and a subsequent clinical course consistent with MDS-EB. Two patients (D154 and SE007) experienced spontaneous normalization of their karyotype and complete blood count (CBC), like what had been previously described by us and others[2–4]. We performed a detailed history and physical examination to exclude patients with known constitutional bone marrow failure (BMF) syndromes. All studied patients had a normal chromosomal breakage test to rule out Fanconi anemia (FA).

Patient samples and genomic studies

We used genomic DNA from bone marrow (BM) and/or peripheral blood (PB) granulocytes obtained at diagnosis. Genomic analysis was accomplished using a custom panel targeting *SAMD9*, *SAMD9L,* and 22 single nucleotide variants (SNV) on chromosome 7q (allele frequency >35% in all ethnic sub-populations in gnomAD) (Ampliseq #IAD104171)[5] in 666/669 cases (Cohort A: 547; Cohort B: 119/121), while in 544 cases of Cohort A additional sequencing of 28 MDS-associated genes (*GATA2, RUNX1, HOXA9, CEBPA, GATA1, KRAS, NRAS, CBL, PTPN11, ASXL1, EZH2, SETBP1, FLT3, KIT, JAK2, JAK3, CSF3R, MPL, SH2, BCOR, BCORL1; RAD21, STAG2, CTCF, TP53, PTEN, CALR, VPS45*) was performed using targeted NGS panel (Ampliseq #IAD51150) (Supplementary Table 16). Libraries for targeted NGS were prepared using NEBNext Ultra II DNA library prep kit (New England BioLabs, cat#E7645S/L) per manufacturer's instruction and samples were sequenced on an Illumina Miseq 2000 sequencer with 2 × 150 bp reads. In addition, n=125 cases (*SAMD9/9L*^{mut}: 63/67; *GATA2*^{mut} 24/35, MDS wildtype 38/471) were subjected to whole exome sequencing (WES) either using SureSelect Human All Exon V6 enrichment (n=117; Agilent, cat# 5190–8863) and Human Comprehensive Exome (n=8; TWIST Biosciences). The generated libraries were sequenced on the Illumina Hiseq 2500 with 150bp paired-end reads. FASTQ files were mapped and aligned using SeqNext platform (JSI medical system, Germany) to hg19 reference sequences of a virtual panel of 300 genes (including 28 MDS-associated genes, *SAMD9*, and *SAMD9L*) relevant to BMF, MDS predisposition, and hematological cancers as per the Pan-Cancer studies with cohorts of >10,000 cancers (Supplementary Table 17)[6–12]. Initial variant calling was restricted to ≥Q15 (40% threshold) with a minimum 20x per direction coverage, 30x absolute minimum

coverage, ≥ variant allelic frequency (VAF) of 5%, and tagging of variants in homopolymer regions with >5bp. Intronic regions adjacent to exon boundaries (± 30bp) were also included. Then the generated BAM and VCF files were exported to SeqNext and Alamut Visual v2.11 (Sophia Genetics, USA) for visualization and subsequent variant annotation with ANNOVAR[13] and dbNSFP3.5[14] Longitudinal sequencing analysis was performed in 14 *SAMD9/9L*^{mut} patients. Germline status was confirmed in non-myeloid specimens, including skin fibroblasts (11 patients), hair follicles (18), purified T-lymphocytes (22), nails (1), buccal swabs (2), or by family inheritance (5). For characterization of somatic mutational landscape across all the 3 germline MDS (*GATA2*^{mut}, *SAMD9/9L*^{mut}, wildtype) as shown in Fig. 4d, a combinatorial analysis of targeted gene panels in 559 patients (Cohort A: 548, Cohort B: 11/21 *SAMD9/9L*^{mut}) and WES in 125 patients (63 *SAMD9/9L*^{mut}, 24 *GATA2*^{mut}, 38 wildtype) was accomplished.

For variant analysis, we employed American College of Medical Genetics and Genomics - Association for Molecular Pathology (ACMG-AMP) 2015 guidelines[5] and Sherloc[15] parameters and developed a toolbox for *SAMD9/9L* variant interpretation (Online methods: Variant selection, Supplementary Table 18). The variant effect was predicted using dbNSFP-adapted rank scoring[16] for REVEL[16], CADD[17], DANN[18], M_CAP[16], FATHMM[16], SIFT[19], Polyphen2HDIV[20], Polyphen2HVAR, MutationTaster[21] and MutationAssessor[22]. The size of clones harboring LOH of chromosome 7 was calculated from VAF of heterozygous SNVs.

Variant selection

We used an in-house classification framework for data analysis and variant classification that included procedures from the ACMG-AMP 2015 guidelines[5] and Sherloc[15]. The following steps were applied for variant classification and selection of variants with clinical relevance:

1. Population allele frequency. The Genome Aggregation Database (gnomAD)[23] v2.1.1 was used as a population control cohort. Variants with population minor allele frequency (MAF) <0.1% were considered to fall within a potentially “pathogenic range” and further evaluated, whereas common variants with population MAF≥0.1% were filtered out. Ethnic sub-populations were individually sorted to identify ethnicity-enriched polymorphisms. Variants that were enriched in gnomAD ethnic populations (definition: population MAF >0.1%) were excluded from further analysis.

2. Quality assessment. All variants with ≥ 20x reads were considered for further evaluation. At least 8 altered reads in the forward and reverse strands (4/4) were required. Visual verification of BAM files with Alamut Visual v2.11 (Alamut Visual (Sophia Genetics, USA)) and SeqNext (JSI Medical Systems, Germany) was performed to facilitate the recognition of low-quality reads and homopolymer regions surrounding variants of interest. All variants below 1% VAF and a quality score of 30 (Q30, 99.9% base call accuracy) were automatically excluded.

3. Mutation type. We considered non-synonymous changes causing missense or null/protein-truncating effect for further evaluation. Synonymous variants were filtered out from our study and not evaluated.

4. Computational prediction. CADD v1.3 scoring model was used for variant prediction as it is not only widely used, but can also classify variants other than missense[24]. To further increase the robustness of CADD scores, we applied gene-specific thresholds provided by the Gene-Aware Variant Interpretation (GAVIN_calibrations_r0.5) for medical use[25]. For selection of candidate germline variants from WES, we applied an initial CADD threshold of >15, followed by excluding variants that fell below GAVIN 0.5 – gene-specific threshold (Category C1 and C2)[25].

   For comparative analysis of patient-derived and gnomAD missense variants in *SAMD9/9L* and *GATA2*, we also used DANN[18] and REVEL[14] algorithms because of their high sensitivity and specificity in calling missense variants[25]. Canonical scoring values were replaced by rank scores provided by dbNSFP[26] that range from 0 to 1, where higher values indicate pathogenicity[26].

5. Clinical observations. Clinical evidence was based on previously reported cases linking known disease phenotype to identified genotype.

ACMG/AMP variant classification toolbox for *SAMD9/9L* genes

After the initial selection based on the above criteria, variants were assigned to 5 categories according to ACMG/AMP criteria, including pathogenic, likely pathogenic, variant of unknown significance (VUS), likely benign, and benign (Supplementary Table 6). None of the *SAMD9/9L* variants scored as benign or likely benign. In our study, we have not used the PS3 criterion (functional studies) for *SAMD9/9L* variant scoring because the functional studies have not been standardized. Specifically, the following evidence codes were applied to *SAMD9/9L* variants:

- PP1 for cases with familial disease: D084, D154, D637, SE007.
- PS1 for the same amino acid change described as pathogenic in other studies. This does not apply to patients D154, D084, D870 that were previously published by us.
- PS2 for confirmed *de novo* variants, subdivided into PS2_strong: in patients with MIRAGE syndrome phenotype (D769 and D1285), and PS2_moderate in patients with a phenotype consistent with SAMD9/9L syndromes but specific to these syndromes (D1127, D1300, E084, and H005).
- PS4 (prevalence of the variant in affected individuals is significantly increased compared with controls). The significance was calculated using Fisher's Exact test comparing the prevalence of variants in our screened patient cohort A and B (n=669) versus total analyzed cases (~140,000) from the gnomAD population database (Supplementary Table 18). PS4 was subdivided into PS4_supporting (if 1 allele present in patients), PS4_moderate (2–3 alleles), and PS4_strong

(≥4 alleles). Alleles refer to all known patients, studied here and previously published.

- PM4 for in-frame deletion variants in D130 and D1266.
- PM5 for missense changes at an amino acid residue where a different change was reported as pathogenic.
- PM6_moderate (for variants assumed *de novo* in patients with MIRAGE phenotype because all published MIRAGE cases thus far had *de novo* mutations).
- PP3 or BP4 (based on CADD score threshold of 20).
- BP5 for patient GR010 with alternate genetic cause (*PTPN11* mutation).

The following categories were deliberately not utilized in variant classification:

- PS3 (well-established *in vitro* or *in vivo* functional studies support the damaging effect of the variant), because the HEK293/T/FT cell line growth assays for *SAMD9/9L* variants have different readouts and are not standardized across laboratories.
- PP2 (missense variant in a gene that has a low rate of benign missense variation and in which missense variants are a common mechanism of disease): even though the exceeding majority of *SAMD9/9L* germline mutations are missense, we have not applied this category because of the low constraint of *SAMD9/9L* genes for benign missense variation (not reaching Z-score 3.09).
- PM1 (variant located in a mutational hot spot and/or critical and well-established functional domain without benign variation) was not used due to not enough functional evidence.
- PVS1 (null variant in a gene where loss-of-function is a common disease cause) for germline null (frameshift, nonsense) mutations.

Germline *SAMD9L* mutations K770E and M892R were interpreted as VUS in patients D1215 and D1297, respectively. However, the same variants were confirmed through parental testing as *de novo* in patients D1285 (K770E) and H005 (M892R), elevating the category to likely pathogenic in these 2 patients.

For the interpretation of variants in other highly conserved genes such as *KRAS*, *HRAS*, *MAP2K2*, *RAF1*, *SHOC2*, and *SOS1*, PP2 has been utilized as PP2_supporting, despite of Z-score <3.09, according to ClinGen Rasopathy panel specifications.[27]

Furthermore, to exclude other potential underlying syndromes, we performed WES on 63 *SAMD9/9L*-mutated (*SAMD9/9L*^{mut}) patients with available material (WES analysis pipeline described below). We identified 84 additional variants with a median VAF of 48% in 47 *SAMD9/9L*^{mut} patients (Supplementary Table 5). Based on ACMG/AMP[5] criteria, only 1 variant found in patient GR010 scored as pathogenic (hotspot *PTPN11* E139D mutation previously reported as pathogenic in Noonan syndrome). This *PTPN11* mutation was confirmed as germline in hair follicles. This patient carried a novel germline *SAMD9*

mutation H244R with a low VAF due to monosomy 7. Among the remaining 83 variants found by WES, 8 were classified as likely-pathogenic (6 truncating, 1 in-frame deletion, and 1 cryptic-splicing acceptor) while 75 were categorized as VUS.

Four patients within the *SAMD9/9L*^{mut} cohort carried variants in telomere-related genes (VUS in *ACD*, *NHP2*, *WRAP53,* and a likely-pathogenic variant in *PARN*). Telomere length analysis was normal in 3 patients (D430, D662, D1267) with heterozygous VUS, and below 1st percentile in D1215 with the homozygous *PARN* variant c.1262+2_1262+3dupTA. Patient D1215 had a *SAMD9L* mutation (K770E) that was also detected as a *de novo* in another patient with MIRAGE phenotype, monosomy 7, and additional second-site rescue *SAMD9L* mutation (D1285). D1215 presented with SAMD9L specific disease phenotype, including ataxia, psychomotor retardation, and cryptorchidism (Supplementary Table 4). Nineteen heterozygous variants in FA genes were detected in 16 patients: *FANCA* (5), *FANCD2* (4), *FANCE* (1), *FANCD1* (1), *FANCS* (2), *FANCQ* (3), *FANCP* (2), *RAD51C* (1). As mentioned above, all patients had normal chromosomal breakage tests and a phenotype not consistent with FA. Additional VUS in myeloid neoplasia predisposing genes were excluded as potential candidates based on mutation effect, localization, and phenotype: for *TET2* only frameshift truncating mutations were found in cases with familial myeloid neoplasia (our patients carried only missense VUS)[28,29]; for *HLTF* only one mutation in DEAD helicase domain was found in one family with adult-onset MDS (our patients carried missense VUS outside this domain)[30]. Other VUS in genes associated with bone marrow failure syndromes (*ERCCL6*, *MPL*, *SBDS,* and *EFL1*) were heterozygous and not consistent with the bi-allelic status described in these syndromes.

We have previously reported 5 patients: 1 with *SAMD9* mutation (D870)[31] and 4 with *SAMD9L* mutations (D084, D154, D637, SE007)[4].

Determination of clonal size of −7 from bulk sequencing

The clonal size was inferred from allelic frequencies, and the total number of deep sequencing reads normalized to the ploidy level, not considering clonal heterogeneity. Bi-allelic clones with normal karyotype (A|B), mono-allelic clones with −7 (A|-), and bi-allelic (A|A) for known uniparental disomy of chromosome 7q (UPD7q) were assumed. Per patient, an average of 1400 common variants on whole chromosome 7 were selected (900 of those were on chromosome 7q). Heterozygous variants were assumed as bi-allelic clones (A|B). The median clonal size of −7 was calculated by subtracting the allelic count of each heterozygous call (multiplied by 2) from the allelic number (total depth). The remaining alleles were counted as single cells with monoallelic genotype (A|-), corresponding to a −7 clone.

Determination of loss of heterozygosity (LOH) of chromosome 7 from WES data

LOH of chromosome 7 was inferred from *SAMD9/9L*^{mut} patients' WES data by targeting all informative variants on chromosome 7 and plotting the allelic frequency of variants that are present in dbSNP and gnomAD databases. Variant calling was restricted variants with ≥ Q30 (40% threshold), minimum 20x per direction coverage, 30x absolute minimum coverage, ≥ VAF of 5%, and tagging of variants in homopolymer regions with >5bp.

VCF files were then exported and annotated with ANNOVAR. Finally, a text file with the genomic position and VAF ratio of the alternative allele was visualized using RStudio v3.6 (rstudio.com).

Karyotyping

Metaphase karyotyping and interphase fluorescence *in situ* hybridization was performed using patient BM specimens according to standard protocol[32]. Complex karyotypes were defined by the presence of 3 chromosomal abnormalities, including at least one structural aberration[32].

Colony forming cell (CFC) assay

Based on material availability, we procured BM mononuclear cells (MNC) from 4 patients (D534, D769, D743, D870) from diagnosis and 1 patient (D130) from 3 time points throughout the disease course. Cryopreserved BM MNCs were thawed, washed (2x), and plated at a density of 30,000 cells per 35mm petri dish in Methocult (StemSpan H4434). The cells were cultured for 12–14 days until recognizable myeloid and erythroid CFC units were observed. The colonies were hand-picked under an inverted light microscope using 100ul pipette and suspended in 45ul of Chelex solution (0.9ug of Chelex, 18ul of Triton X-100, 18ml of nuclease-free water). Further, the Chelex-CFC mixture was briefly centrifuged and incubated at 56°C for 15 min and 99°C for 8 min. After that, the supernatant was transferred to a 96 well Multiscreen Durapore membrane (Millipore #MSHVN4510) plate with a collection plate underneath and centrifuged at 1000g for 15 minutes.

The purified native DNA was subjected to PCR amplification of the targeted regions followed by bi-directional sequencing for the following mutational targets: *SAMD9*: V1276I, Y1192X, R1188P, R1529H, L642del, F920GfsX4, I983S, N765TfsX13; *SAMD9L*: R986C, R589X; *SETBP1*: D868N; *ASXL1*: G656WfsX12; *PTPN11*: E76Q, Q61P. We targeted common polymorphisms on chromosome 7q (found to be heterozygous in germline specimens of the 5 investigated patients): rs876118, rs2537193, rs2303937, rs3173833, rs2657332, rs2234001, rs2235013, rs9692250, rs34311813, rs2293353, rs9692250, rs870745, rs4407778. These variants were used to assess the loss of heterozygosity of 7q in single CFC colonies compared to germline bulk specimens.

Analysis of allelic configuration of *SAMD9/9L* mutations

Genomic DNA from bone marrow of n=5 patients harboring both germline and somatic *SAMD9/9L* mutations in D769 (R1188P/R1529H), D851 (R986H/W695C), D927 (G454R/I876T), D1248 (R986H/R1524H), and D084 (V1512M/R1188X) was PCR-amplified to obtain haplotypes encompassing analyzed mutational spots. The obtained PCR products were TA-cloned, and Sanger sequenced. The sequences were analyzed in Codoncode Aligner v7.1 to assess if the germline and somatic *SAMD9/9L* mutations are travelling together (*cis*) or occurring in separate sequences (*trans*). In additional n=5 patients, we indirectly assessed the haplotype and allelic configuration of germline and somatic *SAMD9/9L*^{mut} mutations using single CFC sequencing (n=3) in D534 (V1276/Y1192X), D743 (R986C/R589X), and D870 (I983S/N769KfsX5); and scDNAseq (n=2) in (R1281S/

L635F, R1281S/N769KfsX5) and D769 (R1188P/P1431L, P1341L was previously missed in bulk sequencing). In all the 9 patients described above, both variants were found to be in *cis*.

Functional variant analysis

HEK293 growth assays

The functional impact of *SAMD9/9L* mutations was assessed by their effect on the proliferation of the human embryonic kidney cell line (HEK293). However, this research assay's results were not considered when classifying *SAMD9/9L* variants according to ACMG/AMP criteria. Here, we introduced patient-specific single (germline n=48, somatic n=7) and double (*cis* germline/somatic n=6) *SAMD9* or *SAMD9L* mutations into the vector with wildtype *SAMD9* or *SAMD9L* cDNA by site-directed mutagenesis (QuickChange multisite-directed mutagenesis kit, Agilent Technologies). In addition, 4 rare (gnomAD population MAF <0.1%) and 11 common (MAF ≥0.1%) missense variants (total n=15) were introduced (Supplementary Table 9). All constructs were sequenced to confirm mutations. For assessment of cellular proliferation, HEK293 cells (ATCC) were transfected at ~50% confluency separately with *SAMD9/9L* variants, wildtype (WT), and empty vector using FuGENE (Promega). Pathogenic mutations (*SAMD9* I983S, *SAMD9L* R986C) previously shown to inhibit proliferation served as controls. Transfected cells were subjected to G418 (700ug/ml) (Gibco) selection for 14 days and collected once all untransfected control cells were dead. Cells were trypsinized, washed, trypan-blue stained, and counted in Countess II analyzer (Thermofisher). This experiment was performed in triplicates and on 3 independent occasions for every construct. Results were represented as mean ± SEM for all 3 studies.

CD34+ cell apoptosis assay

GCSF mobilized peripheral blood CD34+ cells from healthy donors (Fred Hutch Cancer Research center, IRB protocol 985.03) were pre-stimulated for 24 hours in X-Vivo 10 (#BEBP01–055Q, Lonza) supplemented with 100ng/ml stem cell factor (SCF, #255-SC/CF R&D Systems), 100ng/ml thrombopoietin (TPO, cat#288-TP/CF R&D Systems) and 100ng/ml recombinant Flt3 ligand (#308-FK/CF R&D Systems). After stimulation the cells were centrifuged at 300g for 10 min and then re-suspended in fresh X-Vivo 10 with cytokines (SCF, TPO, Flt3, 100ng/ml each), 1% HSA, 2mM L-ala-L-glu and transduction enhancers (protamine sulphate 8ug/ml, 1X PGE2, 1X Lentiboost). Then 200,000 CD34+ cells were mixed with CL20c-MSCV-ires-GFP empty or SAMD9 or SAMD9L wildtype or mutant lentiviral particles (MOI400) and seeded in a total of 100ul volume/well in a flat bottom non-tissue culture 96-well plate (Fisher Scientific). The plate was incubated at 37°C for 24 hours in a tissue culture incubator (5% CO2 and 21% O2). Following incubation, the transduced cells were collected, washed 2X with 1X sterile PBS and stained for Annexin V-APC following the manufacturer's protocol (#640941 BioLegend). Dead cell exclusion was achieved with DAPI stain. For each sample at least 10,000 total events were recorded using BD LSRFortessa flow cytometer and the data were analyzed in FlowJo_v10.7.1 (BD Biosciences). This experiment was performed in triplicates and on 3 independent occasions for every construct. Results were represented as mean ± SEM for all 3 studies.

RT-PCR

RNA was extracted using Qiagen RNeasy Mini kit (Qiagen, cat#74104) and ZymoResearch Direct-zol RNA (ZymoResearch, cat#R2061) from patients' fibroblasts and transduced CD34+ cells, respectively. The cDNA conversion was performed using SuperScript IV VILO Master Mix (#11756050 ThermoFisher Scientific) as per the manufacturer's instruction. Lastly, ThermoScientific PowerUp (#A25742, ThermoFisher Scientific) and PowerTrack (#A46012, ThermoFisher Scientific) SYBR green master mix was used for expression analysis under Fast cycling mode. PowerUP PCR conditions: 1 cycle 50°C 2 minutes, 1 cycle 95°C 2 minutes, 40 cycles (95°C 3 seconds, 60°C 30 seconds). PowerTrack PCR conditions: 1 cycle 95°C 2 minutes, 40 cycles (95°C 5 seconds, 60°C 30 seconds).

Primer sequences for RT-PCR:

SAMD9 Fwd GCATCACACATGGACCAGCTA

SAMD9 Rev GTTGTACTCATTGCAGACGGATT

SAMD9L Fwd GAAACAGGAGCACTCAATCTCA

SAMD9L Rev CAGCCTTACTGGTGATTTTCACA

GAPDH Fwd GGAGCGAGATCCCTCCAAAAT

GAPDH Rev GGCTGTTGTCATACTTCTCATGG

Western blot

Whole-cell lysates were prepared using RIPA buffer (#89900, ThermoFisher Scientific) containing a complete phosphatase-protease inhibitor cocktail (#78440, ThermoFisher Scientific). The protein concentration was determined by using the Pierce BCA protein assay kit (#23225, ThermoFisher Scientific) according to the manufacturer's protocol. To reach a final volume of 30ul, an appropriate amount of cell lysate containing 5–10ug of protein was mixed with 7.5ul of Laemmli buffer and the missing amount of RIPA. The samples were then incubated at 95°C for 10 minutes, loaded on Bolt 4%−12% Bris-Tris gel, subjected to electrophoresis at 100V for ~2 hours, and transferred to PVDF membrane using Criterion Blotter (BIORAD) at 100V for 45mins. Afterwards, the membrane was blocked in blocking buffer (5% powdered milk in Tris-buffered saline Tween20 (TBST)) for 1 hour at room temperature. Then, the membrane was incubated with anti-SAMD9 (#ab180575; Abcam, dilution 1:5000 in blocking buffer) or anti-SAMD9L (#25173–1-AP, Proteintech, dilution 1:5000 in blocking buffer) with anti-GAPDH (#2118L, Cell Signaling, dilution 1:1000) as loading control for overnight at 4°C, followed by 3 washes with TBST. The secondary anti-rabbit IgG HRP (#7074s, Cell Signaling) at dilution 1:10000 was applied for 1 hour at room temperature. The protein detection was performed using SignalFire Plus ECL Reagent (#12630S, Cell Signaling) according to the manufacturer's instructions.

Single-cell DNA sequencing (scDNAseq) and data processing

Single-cell DNA sequencing with antibody-oligonucleotide staining was performed using the Mission Bio Tapestri single-cell DNA sequencing platform, per the manufacturer's instructions. In short, custom scDNAseq panels targeting 250 heterozygous gnomAD population polymorphisms on 7q arm and 69 amplicons in *SAMD9/9L* and cancer genes were designed and manufactured by Mission Bio (Supplementary Table 13). Cryopreserved patient samples (n=5 bone marrow from n=4 patients) were thawed, washed with RPMI (GIBCO), quantified using a Cellometer Auto T4 (Nexcelom Bioscience), and diluted to a concentration of 4,000,000 cells/mL in Cell Buffer (Mission Bio). Next, 50uL of cell suspension was loaded onto a microfluidics cartridge and cells were encapsulated on the Tapestri instrument followed by the cell lysis and protease digestion on a thermal cycler within the individual droplet. The cell lysate was reintroduced onto the cartridge and barcoded such that each cell had a unique molecular identifier. Amplification of the targeted DNA regions was performed by incubating the barcoded DNA emulsions in a thermocycler as follows: 98°C for 6 min (4°C/sec); 10 cycles of 95°C for 30 sec, 72°C for 10 sec, 61°C for 3 min, 72°C for 20 sec (1°C/sec); 10 cycles of 95°C for 30 sec, 72°C for 10 sec, 48°C for 3 min, 72°C for 20 sec (1°C/sec); and 72°C for 6 min (4°C/sec). Emulsions were broken and DNA digested and purified with 0.7X Ampure XP beads (Beckman Coulter). The beads were pelleted, washed with 80% ethanol, and the DNA targets were eluted in nuclease-free water. Indexed Illumina libraries were generated by amplifying DNA libraries with Mission Bio V2 Index Primers in the thermocycler using the following program: 95°C for 3 min; 12 cycles of 98°C for 20 sec, 62°C for 20 sec, 72°C for 45 sec; 72°C for 2 min. Final libraries were purified with 0.69X Ampure XP beads. All libraries were sized and quantified using an Agilent Bioanalyzer and pooled for sequencing on an Illumina NovaSeq6000 with 150 base-paired ending multiplexed runs.

FASTQ files generated by sequencers were processed using the Tapestri Pipeline V2 and included adapter trimming, sequence alignment (BWA), barcode correction, cell finding, and variant calling (GATK v4/Haplotype caller). Loom and h5 files were then processed with Tapestri Insights v2.2 (Mission Bio) and the Python-based Mosaic package (GitHub). Tapestri Insights analysis used default filter criteria (e.g., genotype quality ≥30, reads/cell/target ≥10) and annotation-based information (e.g., ClinVar, DANN). Only cells with complete genotype information for all variants (previously detected in bulk sample) were included for downstream processing. Also, samples with low cell yield (<100 total cells) were removed from the analysis.

Statistical analysis

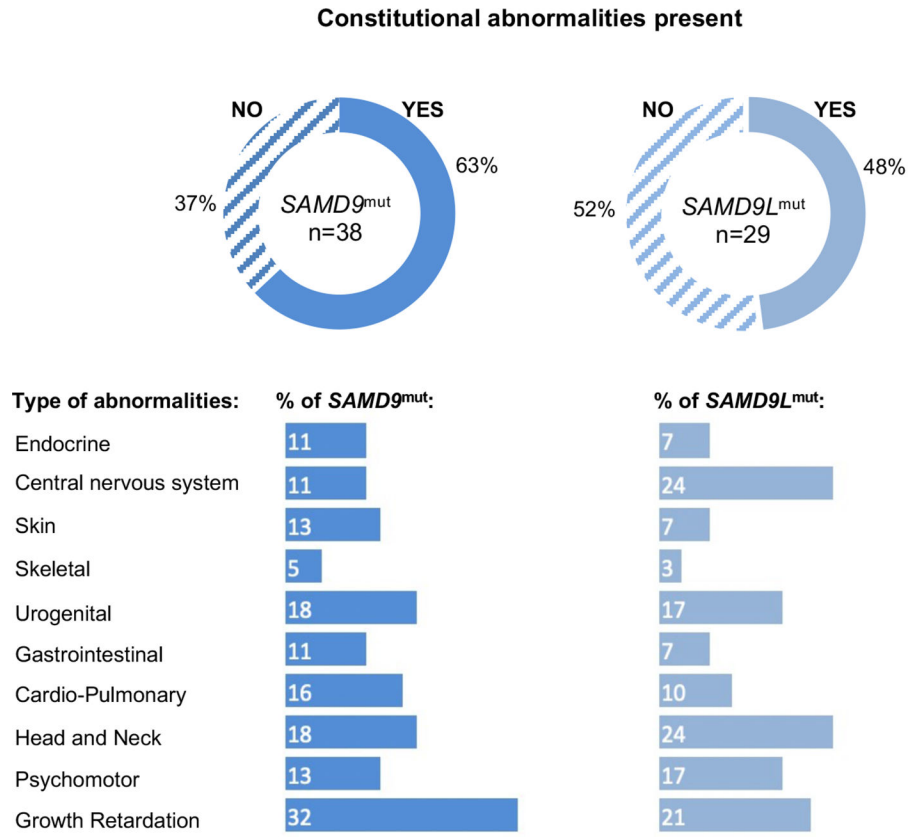
Analysis of clinical variables and outcome was performed with data locked on May 24th, 2019. Overall survival (OS) was defined as the time from diagnosis to death or last follow-up, while event-free survival (EFS) depicts time from HSCT to death, disease recurrence, or last follow-up. The Kaplan-Meier method was used to estimate survival rates, and the two-sided log-rank test was employed to evaluate the equality of the survivorship functions in different subgroups. The Mann-Whitney test with Bonferroni correction was used for comparison of differences in medians of continuous variable. Chi-Square-test was used to examine the statistical significance of the association between mutational status and

categorized factors. Unpaired two-tailed *t*-test was used to compare the association between germline *SAMD9/9L*^{mut} allelic frequency and karyotype. Welch's *t*-test was employed for comparing age at diagnosis between specific phenotype (UPD7q vs non-UPD7q) or disease outcome (remission vs stable disease vs high-risk/progressed) groups with different sample sizes and unequal variances. Fisher's exact test was calculated for 2×2 contingency analyses. Kruskal-Wallis one-way analysis of variance was employed for comparing *in silico* pathogenic prediction scores for gnomAD population vs. *GATA2*^{mut} or *SAMD9/9L*^{mut} MDS patients' variants. Two-tailed paired *t*-tests were implemented to represent the mean ±SEM from n=3 independent HEK293 cell proliferation and n=3 independent CD34+ apoptosis assay. Ordinary one-way ANOVA was employed to show the mean with ±SEM for n=2 independent SAMD9/9L RT-PCR from patients' and healthy controls' primary fibroblasts. All p-values were two-sided, and values <0.05 were statistically significant. Software packages SPSS for Windows 24.0.0 (IBM Corp, NY), NCSS 2004 (NCSS, Kaysville, UT), and GraphPad Prism 9 were employed.

DATA AVAILABILITY

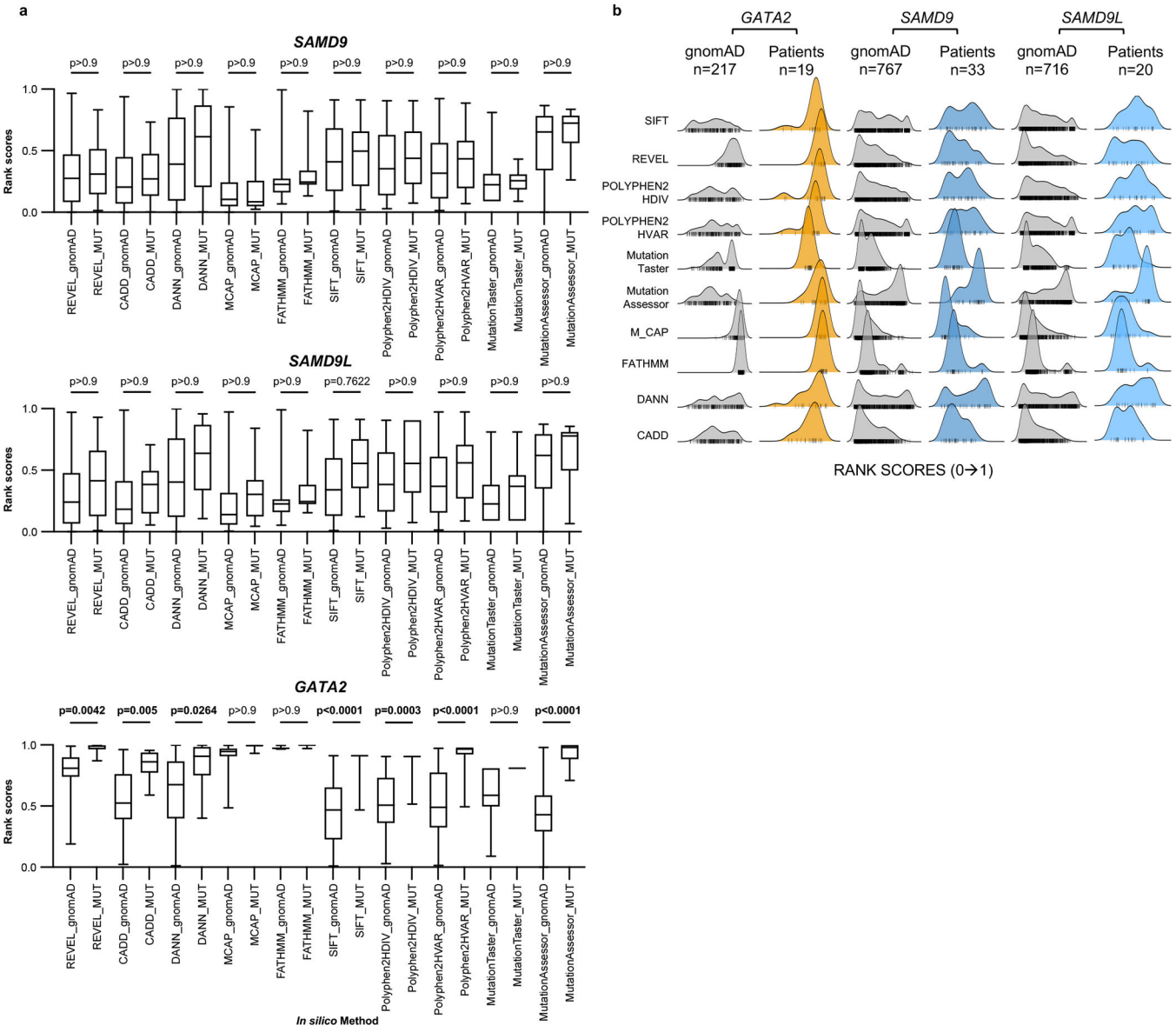
All data are available in the manuscript or supplementary information. Population minor allele frequency of studied variants are from gnomAD v2.1.1 aggregation database (https://gnomad.broadinstitute.org/). Python-based Mosaic package for analysis of single-cell DNA sequencing is available at GitHub (https://github.com/MissionBio/mosaic). Raw sequence datasets have been deposited at the European Genome-Phenome Archive (EGA, http://www.ebi.ac.uk/ega/) hosted by the European Bioinformatics Institute under accession numbers EGAS00001005431 (targeted panel sequencing), EGAS00001005432 (whole exome sequencing), and EGAS00001005433 (single-cell DNA sequencing). The detailed clinical annotation of *SAMD9/9L*^{mut} cohort is provided in Supplementary Tables 1–4.

Extended Data



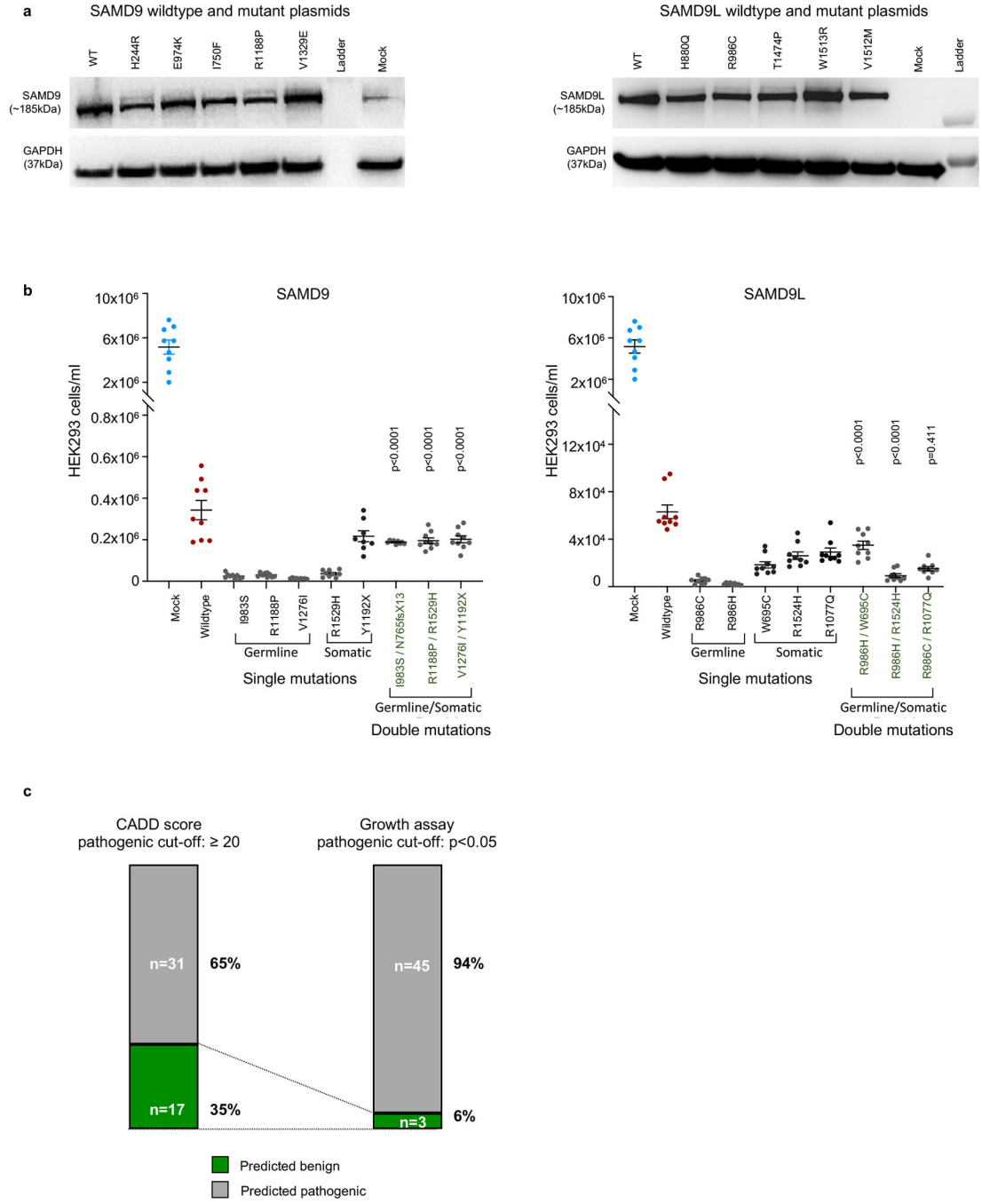
Extended Data Fig. 1. Constitutional abnormalities in patients with germline *SAMD9/9L* mutations.

Frequency of all (pie charts) and individual (bar graphs) non-hematopoietic abnormalities in n=67 *SAMD9/9L*^{mut} patients (n=39 *SAMD9*^{mut} and n=29 *SAMD9L*^{mut}). Detailed phenotypic findings are outlined in Supplemental Table 4.



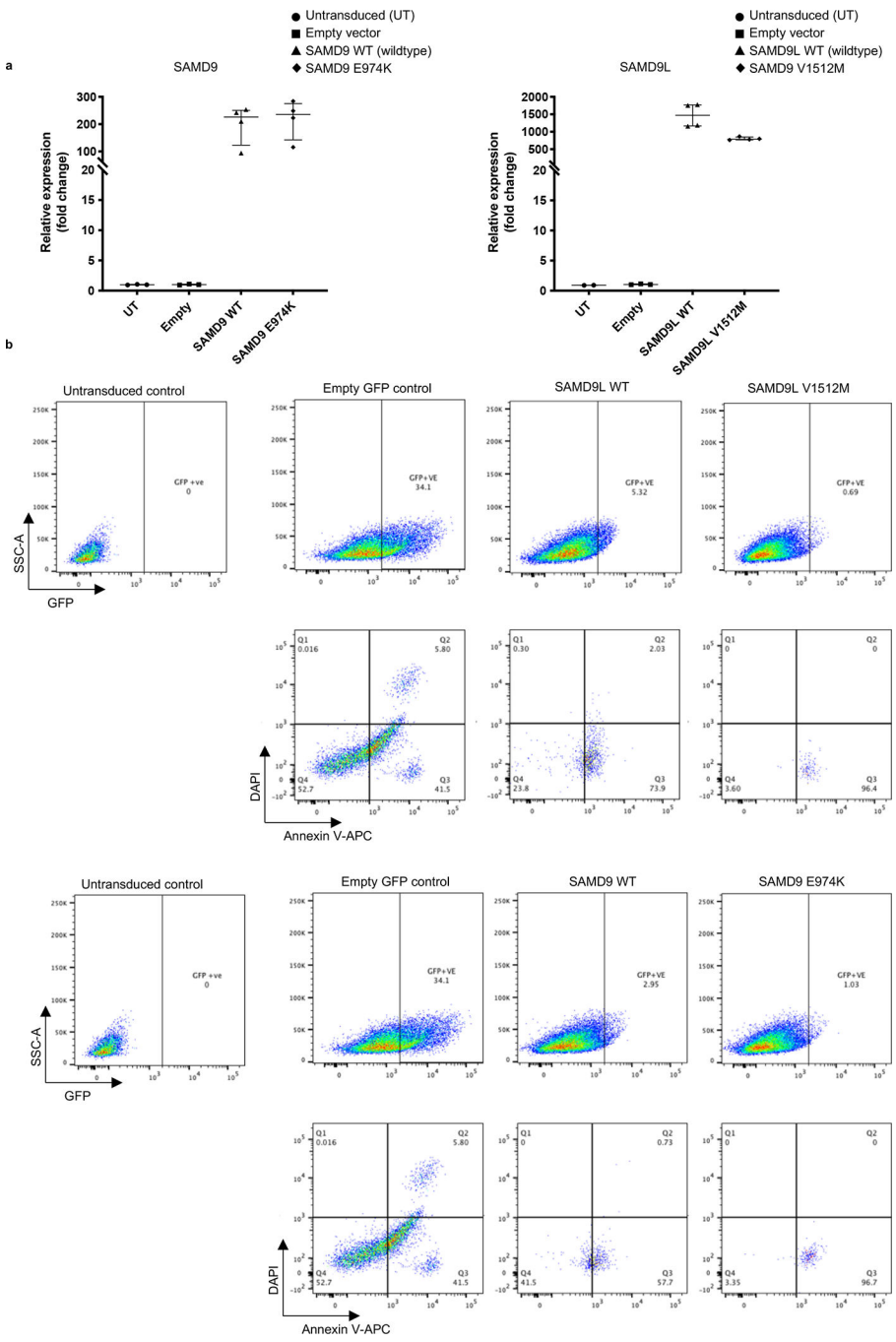
Extended Data Fig. 2. Computational assessment of germline *SAMD9/9L* and *GATA2* mutations.

Rank scores of *GATA2*, *SAMD9*, and *SAMD9L* population variants (gnomAD) or patient-specific germline mutations (MUT) was assessed using 10 *in silico* algorithms: REVEL, CADD, DANN, M_CAP, FATHMM, SIFT, Polyphen2 HDIV, Polyphen2 HVAR, Mutation Taster and Mutation Assessor. The Y-axis depicts the rank scores (0, benign; 1, pathogenic). Analyzed number of gnomAD variants: *GATA2* n=217, *SAMD9* n=767, *SAMD9L* n=716 and patient germline mutations from this study: *GATA2* n=19, *SAMD9* n=33, *SAMD9L* n=20. **(a)** Box plots depicting the minimum, lower quartile, median, upper quartile and maximum of each analyzed data set, p-values are calculated by comparing the medians of each data set using Kruskal-Wallis one-way analysis of variance. The significant p-values (<0.05) are bolded. **(b)** Density plot showing the distribution of rank scores from gnomAD and patient missense variants across *GATA2*, *SAMD9* and *SAMD9L* genes.



Extended Data Fig. 3. Functional consequence and *in silico* discordance of *SAMD9/9L* mutations.
(a) Representative Western blot showing the protein levels 24 hours after transfection with plasmid expressing SAMD9 wildtype (WT), SAMD*9* n=5 patient mutations (left), and SAMD9L wildtype (WT) with SAMD9L n=5 patient mutations (right). The mock control was HEK293 transfected with empty vector. **(b)** Effect of co-occurring germline and second-site (somatic) *SAMD9/9L*^{mut} on HEK293 cell proliferation when expressed either alone or in *cis* from single constructs. Mean ± SEM from n=3 independent experiments performed in triplicate are presented. P-values using paired *t*-test are shown for each double mutant

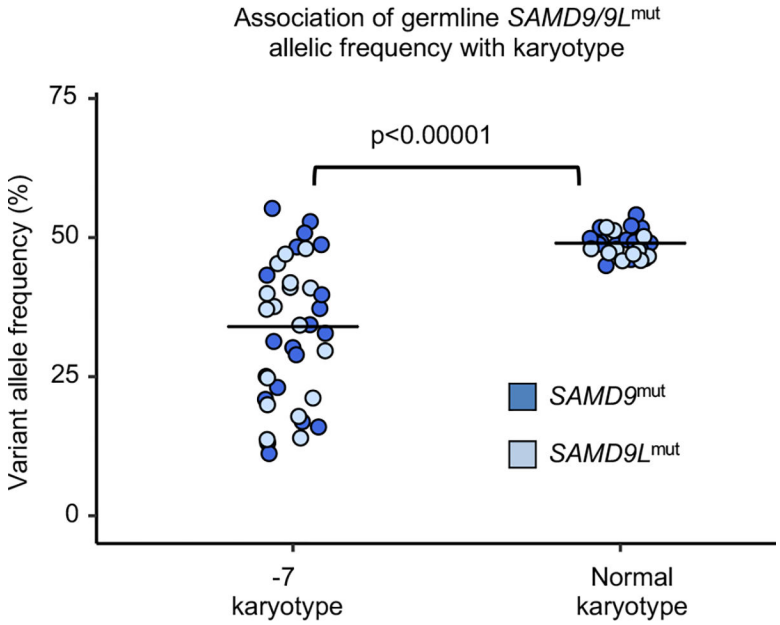
in comparison to the respective germline *SAMD9/9L*^{mut}. **(c)** Comparative evaluation of pathogenicity of 49 germline *SAMD9/9L* mutations by CADD prediction and HEK293 growth assay.



Extended Data Fig. 4. Transduction of CD34+ cells with *SAMD9/9L* mutant lentiviruses.

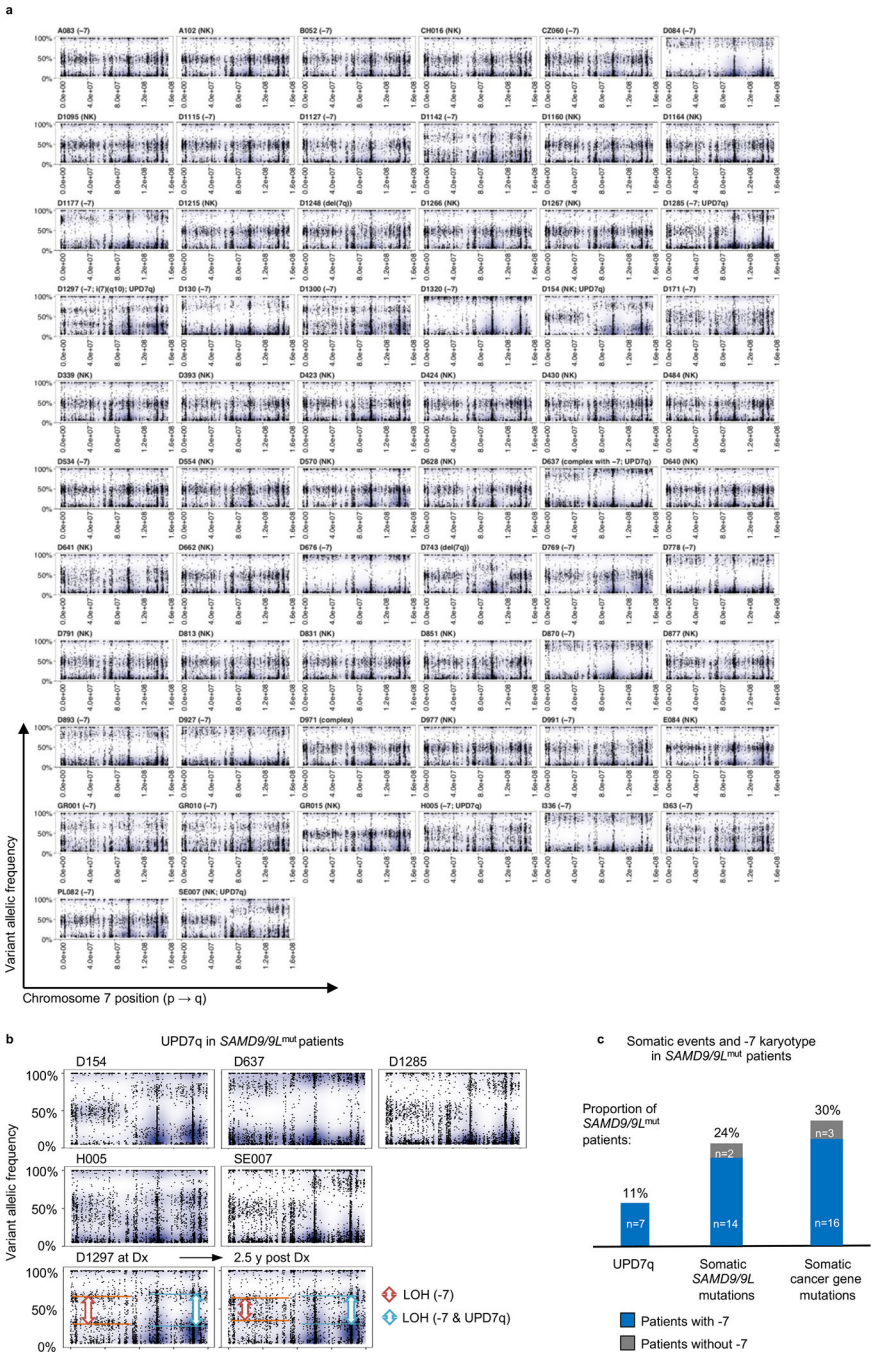
(a) RT-PCR confirming the overexpression of SAMD9 wildtype, SAMD9 E974K, SAMD9L wildtype and SAMD9L V1512M lentiviruses compared to empty control in sorted GFP positive CD34+ cells 24 hours post transduction. Mean ± SEM from n=2 independent

experiments are presented. SAMD9 and SAMD9L expression data is normalized to GAPDH, and untransduced CD34+ sample is set to 1. **(b)** Flow cytometry plots outlining staining strategy with Annexin V-APC and DAPI in GFP positive CD34+ cells transduced with lentiviral constructs (top: SAMD9L, bottom SAMD9). Corresponding data are shown in main manuscript Fig. 3d.



Extended Data Fig. 5. Association of germline *SAMD9/9L* mutation allelic frequency with karyotype.

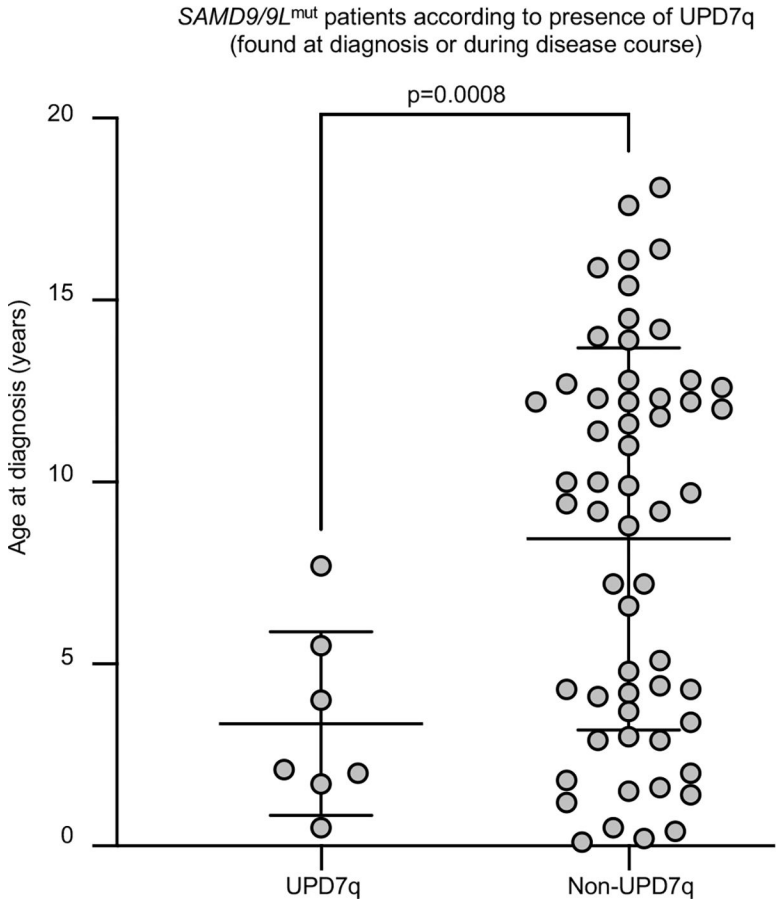
Variant allelic frequencies of germline *SAMD9/9L* mutations in n=65 patients either with monosomy 7 (−7), n=37 or normal, karyotype, n=28. The median allelic frequency of each group is marked, and the p-value is calculated using unpaired *t*-test.



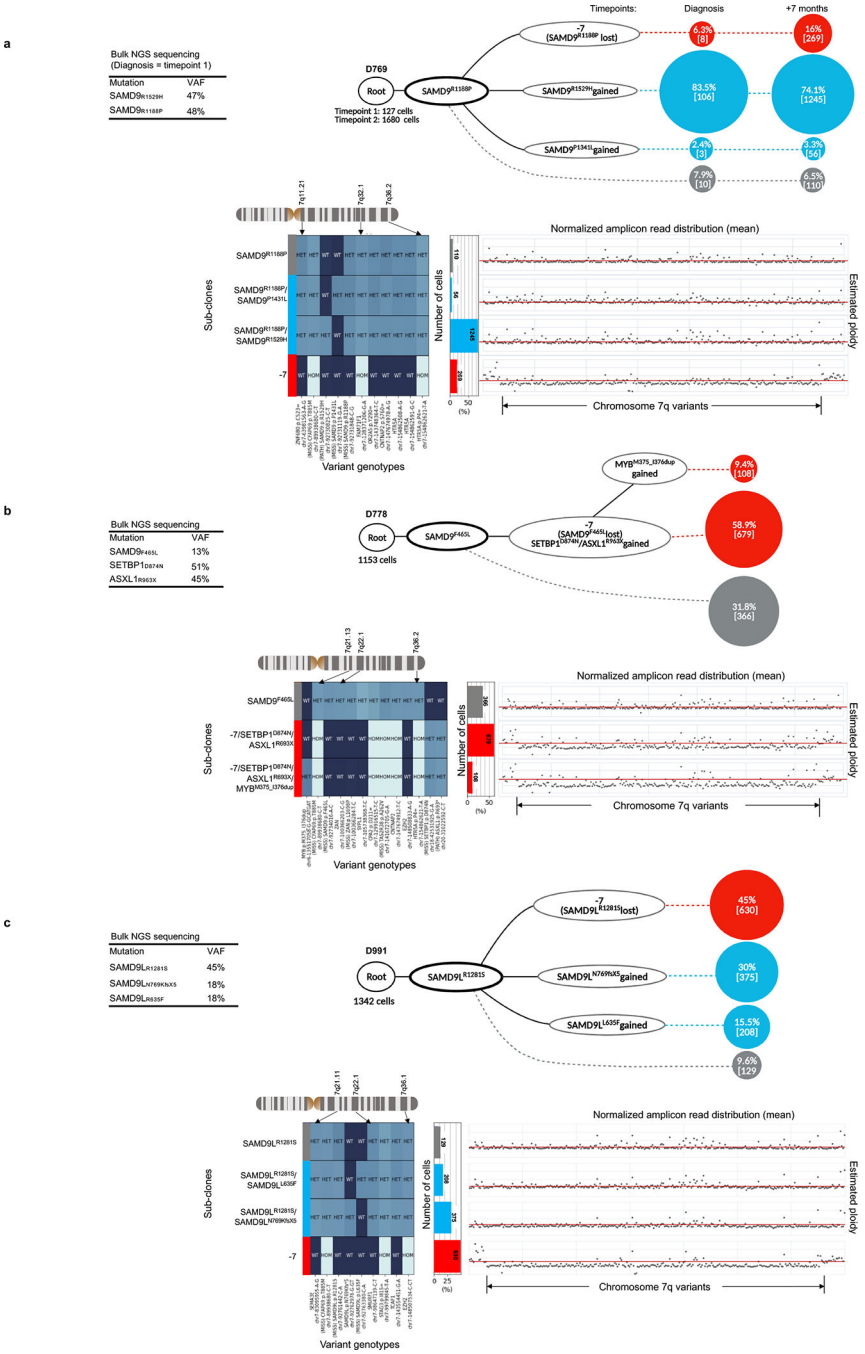
Extended Data Fig. 6. Loss of heterozygosity (LOH) plots for chromosome 7 and the association of −7 with somatic events.

Allelic frequency of all informative chromosome 7 variants detected by bone marrow whole exome sequencing (variant allele frequency (VAF) between >5% and <95%) plotted for all sequenced *SAMD9/9L*^{mut} cases. The karyotypes at time of sequencing are shown, including monosomy 7 (−7), normal karyotype (NK), and uniparental isodisomy of 7q (UPD7q). **(b)** Chromosome 7 plots for patients with UPD7q. Two types of loss of heterozygosity (LOH) detected in patient D1297 are indicated by blue (−7) and orange (−7 & UPD7q) arrows.

(c) Proportion of somatic events in *SAMD9/9L*^{mut} patients. Stacked bars demonstrate the percentage of *SAMD9/9L*^{mut} patients with different somatic events (UPD7q, somatic second-site *SAMD9/9L*^{mut} and somatic cancer gene mutations) within the −7 (blue) and other than −7 karyotype (grey). Within each bar the absolute number of patients for each group is shown.



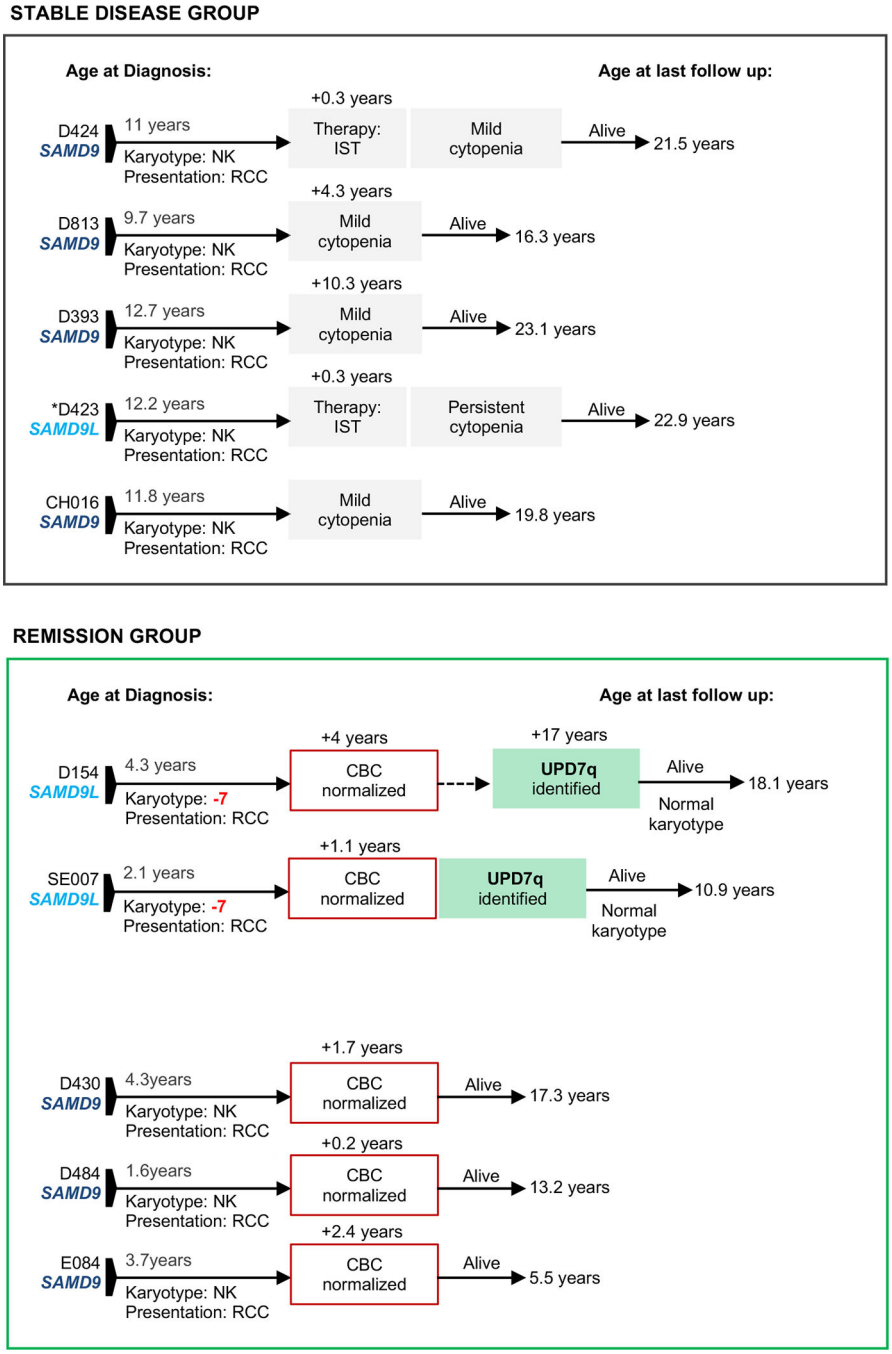
Extended Data Fig. 7. Age at diagnosis in *SAMD9/9L*^{mut} patients with UPD7q.
Patients with detected uniparental isodisomy 7q (UPD7q), n=7, had a diagnosis (Dx) at a younger age of median 2.1, range 0.5 – 7.7 years) in comparison to n=59 patients without detectable UPD7q (median 9.4, range 0.1 – 18.1 years). The p-value is calculated using unpaired *t*-test with Welch's correction (due to unequal cohort size). Dot plots represent the mean ±SD.



Extended Data Fig. 8. Clonal architecture and chromosome 7 copy number inferred from single-cell DNA sequencing (scDNAseq).

Pattern of clonal evolution (created with BioRender.com) and visualization of genotypes of individual clones constructed from high quality single cells (allelic dropout rate <0.9) detected in 3 patients with *SAMD9* **(a-b)** or *SAMD9L* **(c)** germline mutations. Mutational phylogeny was inferred from scDNAseq data using Tapestri Insights v2.2 and Mosaic packages. Root denotes the total number of cells analyzed for each sample and bolded circle symbolizes ancestral clone with germline *SAMD9/9L*^{mut}. Percentage and number

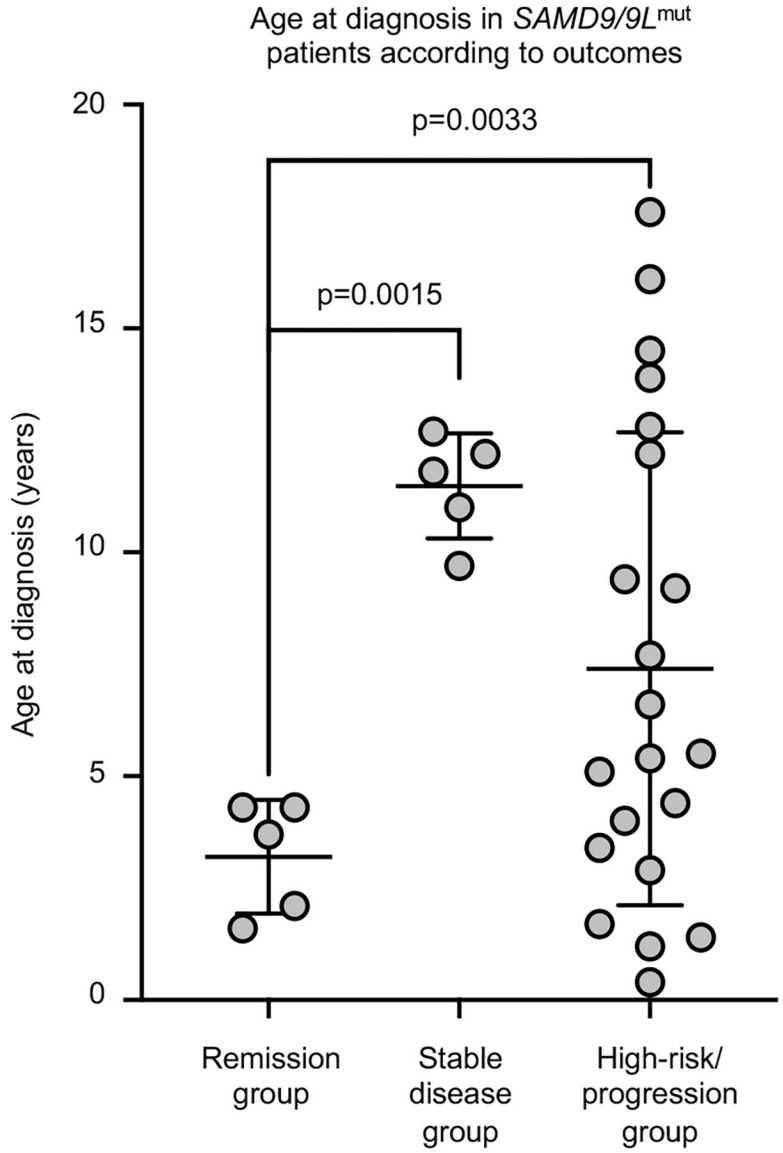
of single cells appear within colored circles; native state hematopoiesis (grey), second-site *SAMD9L* mutation (blue), UPD7q (green), −7 and −7 with somatic cancer mutations (both red). Variant allele frequency (VAF) from bulk sequencing is shown for reference. Panels **(a)** and **(c)** exemplify patients with branching evolution of independent benign and malignant SGR events arising from germline *SAMD9/9L*^{mut} hematopoiesis. Panel **(b)** depicts the linear evolution of malignant −7 clone with *SETBP1*/*ASXL1* mutation to acquire an additional *MYB* M375_I376dup mutation. Lower part of panels **a-c** depict the genotype annotation of the observed individual clones (shown above each genotype plot). Selected variants flanking *SAMD9* or *SAMD9L* have either wildtype (WT), heterozygous (HET) or homozygous (HOM) genotype states. Right of lower panels show normalized amplicon read distribution of informative variants from scDNAseq, with red line marking the diploid state referenced from diploid cells.



Extended Data Fig. 9. *SAMD9/9L*^{mut} patients with stable disease or remission.

Timeline depicting the clinical course in n=10 *SAMD9/9L*^{mut} patients with a follow up longer than 1 year who had stable disease (defined as no need for transfusions or therapeutic interventions, and no infections) or natural hematopoietic improvement. All patients presented with refractory cytopenia of childhood (RCC). The upper panel (black box) shows n=5 patients with stable disease course characterized by chronic cytopenia, normal karyotype (NK); n=4 had *SAMD9*^{mut}. Two patients underwent immunosuppressive therapy (IST). Lower panel (green box) shows cases with spontaneous hematological

improvement; 2 *SAMD9L*^{mut} patients with previous monosomy 7 (−7) had normalization of karyotype and complete blood counts (CBC).



Extended Data Fig. 10. Age at diagnosis in *SAMD9/9L*^{mut} patients according to outcomes.
Patients in remission group (n=5) were diagnosed at a younger age in comparison to stable disease (n=6) or high-risk/progression (n=21) groups. The p-value was calculated using unpaired *t*-test with Welch's correction.. The dot plots represent the mean ± SD. Group were defined as follows: Remission - long-lasting resolution of hematologic symptoms; Stable disease - prolonged chronic cytopenia (>1 year) without the need for therapies; High-risk/progressed - advanced MDS, somatic cancer mutations, and disease progression (progression of karyotype from normal to abnormal or MDS subtype from refractory cytopenia of childhood (RCC) to MDS with excess blast (MDS-EB))

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Sushree S Sahoo[1,2], Victor B Pastor[2], Charnise Goodings[1], Rebecca K Voss[2], Emilia J Kozyra[2,3], Amina Szvetnik[2], Peter Noellke[2], Michael Dworzak[4], Jan Stary[5], Franco Locatelli[6], Riccardo Masetti[7], Markus Schmugge[8], Barbara De Moerloose[9], Albert Catala[10], Krisztián Kállay[11], Dominik Turkiewicz[12], Henrik Hasle[13], Jochen Buechner[14], Kirsi Jahnukainen[15], Marek Ussowicz[16], Sophia Polychronopoulou[17], Owen P Smith[18], Oksana Fabri[19], Shlomit Barzilai[20], Valerie de Haas[21], Irith Baumann[22], Stephan Schwarz-Furlan[22,23],
European Working Group of MDS in Children (EWOG-MDS)[*],
Marena R Niewisch[2], Martin G Sauer[24], Birgit Burkhardt[25], Peter Lang[26], Peter Bader[27], Rita Beier[28], Ingo Müller[29], Michael H Albert[30], Roland Meisel[31], Ansgar Schulz[32], Gunnar Cario[33], Pritam K Panda[2], Julius Wehrle[34,35], Shinsuke Hirabayashi[2], Marta Derecka[1], Robert Durruthy-Durruthy[36], Gudrun Göhring[37], Ayami Yoshimi-Noellke[2], Manching Ku[2], Dirk Lebrecht[2], Miriam Erlacher[2,38], Christian Flotho[2,38], Brigitte Strahm[2], Charlotte M Niemeyer[2,38], Marcin W Wlodarski[1,2]

Affiliations

[1)] Department of Hematology, St. Jude Children's Research Hospital, Memphis, TN, USA

[2)] Department of Pediatrics and Adolescent Medicine, Division of Pediatric Hematology and Oncology, Medical Center, Faculty of Medicine, University of Freiburg, Freiburg, Germany

[3)] Faculty of Biology, University of Freiburg, Freiburg, Germany

[4)] Department of Pediatrics, St. Anna Children's Hospital and Children's Cancer Research Institute, Medical University of Vienna, Vienna, Austria

[5)] Department of Paediatric Haematology and Oncology, Second Faculty of Medicine, Charles University and University Hospital Motol, Prague, Czech Republic

[6)] Department of Pediatric Hematology and Oncology, IRCCS Ospedale Pediatrico Bambino Gesù; Sapienza University of Rome, Italy

[7)] Paediatric Oncology and Haematology, IRCCS Azienda Ospedaliero-Universitaria di Bologna, Italy

[8)] Department of Hematology and Oncology, University Children's Hospital, Zurich, Switzerland

[9)] Department of Paediatric Haematology-Oncology, Ghent University Hospital Ghent, Belgium

[10] Department of Hematology and Oncology, Hospital Sant Joan de Deu, Barcelona, Spain

[11] Department of Pediatric Hematology and Stem Cell Transplantation, Central Hospital of Southern Pest - National Institute of Hematology and Infectious Diseases, Budapest, Hungary

[12] Department of Pediatric Oncology/Hematology, Skåne University Hospital, Lund, Sweden

[13] Department of Pediatrics, Aarhus University Hospital, Aarhus, Denmark

[14] Department of Pediatric Hematology and Oncology, Oslo University Hospital, Oslo, Norway

[15] Division of Hematology-Oncology and SCT Children´s Hospital, University of Helsinki and Helsinki University Hospital, Hus, Finland

[16] Department of Paediatric Bone Marrow Transplantation, Oncology and Hematology, BMT Unit CIC 817, Wroclaw Medical University, Wroclaw, Poland

[17] Department of Pediatric Hematology/Oncology, Aghia Sophia Children's Hospital, Athens, Greece

[18] Department of Pediatric Haematology/Oncology, Children's Health Ireland at Crumlin, Dublin, Ireland

[19] Department. of Haematology and Transfusiology, National Institute of Children's Diseases Faculty of Medicine, Comenius University, Bratislava, Slovakia

[20] Pediatric Hematology Oncology, Schneider Children's Medical Center of Israel, Petah Tikva, and Sackler Faculty of Medicine, Tel Aviv University, Israel

[21] Dutch Childhood Oncology Group, Princess Máxima Center for Pediatric Oncology, Utrecht, The Netherlands

[22] Institute of Pathology, Klinikum Kaufbeuren-Ravensburg, Kaufbeuren, Germany

[23] Institute of Pathology, University Hospital Erlangen, Erlangen, Germany

[24] Department of Pediatric Hematology and Oncology, Hannover Medical School, Hannover, Germany

[25] Pediatric Hematology and Oncology, University Hospital Muenster, Muenster, Germany

[26] Department of Hematology/Oncology and General Pediatrics, Children's University Hospital, University of Tübingen, Tübingen, Germany

[27] Division for Stem Cell Transplantation and Immunology, Department for Children and Adolescents, University Hospital Frankfurt, Frankfurt am Main, Germany

[28] University Hospital Essen, Pediatric Haematology and Oncology, Essen, Germany

[29] Division of Pediatric Hematology and Oncology, Clinic of Pedatric Hematology and Oncology, University Medical Center of Hamburg-Eppendorf, Hamburg, Germany

[30] Department of Pediatrics, Dr. von Hauner Children´s Hospital, University Hospital, LMU Munich, Munich, Germany

[31] Department of Pediatric Oncology, Hematology and Clinical Immunology, Division of Pediatric Stem Cell Therapy, Medical Faculty, Heinrich-Heine-University, Duesseldorf, Germany

[32] Department of Pediatrics, University Medical Center Ulm, Ulm, Germany

[33] Department of Pediatrics, University Hospital Schleswig-Holstein, Campus Kiel, Kiel, Germany

[34] Department of Medicine I, Medical Center - University of Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany

[35] Institute of Digitalization in Medicine, Faculty of Medicine, University of Freiburg, Freiburg, Germany

[36] Mission Bio Inc., South San Francisco, CA, USA

[37] Department of Human Genetics, Hannover Medical School, Hannover, Germany

[38] German Cancer Consortium (DKTK), Heidelberg and Freiburg, Germany

ACKNOWLEDGEMENTS

This work was supported by grants from the Deutsche Krebshilfe Max-Eder-Nachwuchsgruppenprogramm 70109005 (to MWW), ERA PerMED GATA2-HuMo German Federal Ministry of Education and Research (BMBF) 2018–123/01KU1904 (to MWW), German Cancer Consortium DKTK (to MWW, CMN), Fritz-Thyssen Foundation 10.17.1.026MN (to MWW), Deutsche Kinderkrebsstifung 2017.03 (to MWW); BMBF MyPred 01GM1911A (to MWW, ME, CMN, GG, BS, CF), José Carreras Leukämie-Stiftung (to VBP), DFG SFB1160 (to MK), AIRC (Associazione Italiana Ricerca sul Cancro Special Program Metastatic disease: the key unmet need in oncology 5 per mille 2018 Project Code 21147 (to FL), Cancer Center Core Grant (CA021765, to St. Jude) and Cooperative Centers of Excellence in Hematology NIDDK U54 Grant (DK106829, to Fred Hutch). SSS is a previous recipient of Spemann Graduate School of Biology and Medicine (SGBM) scholarship.

We thank Sophie Krueger, Christina Jaeger, Sandra Zolles, Sophia Hollander, Marco Teller, Ali-Riza Kaya for excellent laboratory assistance; Annemarie Breier, Alexandra Fischer, Wilfried Truckenmueller, Maria Siskou-Zwecker, Axel Gebert, Melanie Boerries and Hauke Busch for data management (all Freiburg); David Cullins (St. Jude) for FACS sorting services and Parag Mitra for technical support. We also acknowledge the Hilda Biobank Freiburg and Genomics Core Facility at the German Cancer Research Center/DKFZ, Heidelberg, Germany for specimen processing. Patient care within the EWOG-MDS consortium would not have been possible without the continuous effort of the National Reference Pathologists, National Reference Cytogeneticists, physicians, nurses and other staff of pediatric oncology units and transplant centers in all 17 participating countries (www.ewog-mds-saa.org).

REFERENCES

1. Lemos de Matos A, Liu J, McFadden G & Esteves PJ Evolution and divergence of the mammalian SAMD9/SAMD9L gene family. BMC Evol Biol 13, 121 (2013). [PubMed: 23758988]
2. Wu L, et al. miR-96 induces cisplatin chemoresistance in non-small cell lung cancer cells by downregulating SAMD9. Oncol Lett 11, 945–952 (2016). [PubMed: 26893673]
3. Topaz O, et al. A deleterious mutation in SAMD9 causes normophosphatemic familial tumoral calcinosis. Am J Hum Genet 79, 759–764 (2006). [PubMed: 16960814]

4. Liu J & McFadden G SAMD9 is an innate antiviral host factor with stress response properties that can be antagonized by poxviruses. J Virol 89, 1925–1931 (2015). [PubMed: 25428864]
5. Nagamachi A, et al. Haploinsufficiency of SAMD9L, an endosome fusion facilitator, causes myeloid malignancies in mice mimicking human diseases with monosomy 7. Cancer Cell 24, 305–317 (2013). [PubMed: 24029230]
6. Narumi S, et al. SAMD9 mutations cause a novel multisystem disorder, MIRAGE syndrome, and are associated with loss of chromosome 7. Nat Genet 48, 792–797 (2016). [PubMed: 27182967]
7. Thomas ME 3rd, et al. Pediatric MDS and bone marrow failure-associated germline mutations in SAMD9 and SAMD9L impair multiple pathways in primary hematopoietic cells. Leukemia (2021).
8. Allenspach EJ, et al. Germline SAMD9L truncation variants trigger global translational repression. J Exp Med 218(2021).
9. Liu J, Wennier S, Zhang L & McFadden G M062 is a host range factor essential for myxoma virus pathogenesis and functions as an antagonist of host SAMD9 in human cells. Journal of virology 85, 3270–3282 (2011). [PubMed: 21248034]
10. Meng X, Krumm B, Li Y, Deng J & Xiang Y Structural basis for antagonizing a host restriction factor by C7 family of poxvirus host-range proteins. Proc Natl Acad Sci U S A 112, 14858–14863 (2015). [PubMed: 26578811]
11. Nounamo B, et al. An interaction domain in human SAMD9 is essential for myxoma virus host-range determinant M062 antagonism of host anti-viral function. Virology 503, 94–102 (2017). [PubMed: 28157624]
12. Meng X, et al. A paralogous pair of mammalian host restriction factors form a critical host barrier against poxvirus infection. PLoS Pathog 14, e1006884 (2018). [PubMed: 29447249]
13. Mekhedov SL, Makarova KS & Koonin EV The complex domain architecture of SAMD9 family proteins, predicted STAND-like NTPases, suggests new links to inflammation and apoptosis. Biol Direct 12, 13 (2017). [PubMed: 28545555]
14. Jiang Q, et al. The Samd9L gene: transcriptional regulation and tissue-specific expression in mouse development. J Invest Dermatol 131, 1428–1434 (2011). [PubMed: 21412262]
15. Li CF, et al. Human sterile alpha motif domain 9, a novel gene identified as down-regulated in aggressive fibromatosis, is absent in the mouse. BMC Genomics 8, 92 (2007). [PubMed: 17407603]
16. Chen DH, et al. Ataxia-Pancytopenia Syndrome Is Caused by Missense Mutations in SAMD9L. Am J Hum Genet 98, 1146–1158 (2016). [PubMed: 27259050]
17. Tesi B, et al. Gain-of-function SAMD9L mutations cause a syndrome of cytopenia, immunodeficiency, MDS and neurological symptoms. Blood (2017).
18. Schwartz JR, et al. Germline SAMD9 mutation in siblings with monosomy 7 and myelodysplastic syndrome. Leukemia 31, 1827–1830 (2017). [PubMed: 28487541]
19. Amano N, et al. Genetic defects in pediatric-onset adrenal insufficiency in Japan. Eur J Endocrinol 177, 187–194 (2017). [PubMed: 28546232]
20. Jeffries L, et al. A novel SAMD9 mutation causing MIRAGE syndrome: An expansion and review of phenotype, dysmorphology, and natural history. Am J Med Genet A 176, 415–420 (2018). [PubMed: 29266745]
21. Kim YM, et al. A case of an infant suspected as IMAGE syndrome who were finally diagnosed with MIRAGE syndrome by targeted Mendelian exome sequencing. BMC Med Genet 19, 35 (2018). [PubMed: 29506479]
22. Shima H, et al. Two patients with MIRAGE syndrome lacking haematological features: role of somatic second-site reversion SAMD9 mutations. J Med Genet 55, 81–85 (2018). [PubMed: 29175836]
23. Shima H, et al. MIRAGE syndrome is a rare cause of 46,XY DSD born SGA without adrenal insufficiency. PLoS One 13, e0206184 (2018). [PubMed: 30403727]
24. Csillag B, et al. Somatic mosaic monosomy 7 and UPD7q in a child with MIRAGE syndrome caused by a novel SAMD9 mutation. Pediatr Blood Cancer 66, e27589 (2019). [PubMed: 30565860]

25. Thunstrom S & Axelsson M Leukoencephalopathia, demyelinating peripheral neuropathy and dural ectasia explained by a not formerly described de novo mutation in the SAMD9L gene, ends 27 years of investigations - a case report. BMC Neurol 19, 89 (2019). [PubMed: 31053103]
26. Cheah JJC, et al. A novel germline SAMD9L mutation in a family with ataxia-pancytopenia syndrome and pediatric acute lymphoblastic leukemia. Haematologica 104, e318–e321 (2019). [PubMed: 30923096]
27. Mengen E & Kucukcongar Yavas A A Rare Etiology of 46, XY Disorder of Sex Development and Adrenal Insufficiency: A case of MIRAGE syndrome caused by mutations in SAMD9 gene. J Clin Res Pediatr Endocrinol (2019).
28. Perisa MP, et al. A novel SAMD9 variant identified in patient with MIRAGE syndrome: Further defining syndromic phenotype and review of previous cases. Pediatr Blood Cancer 66, e27726 (2019). [PubMed: 30900330]
29. Buonocore F, et al. Somatic mutations and progressive monosomy modify SAMD9-related phenotypes in humans. J Clin Invest (2017).
30. Schwartz JR, et al. The genomic landscape of pediatric myelodysplastic syndromes. Nat Commun 8, 1557 (2017). [PubMed: 29146900]
31. Wong JC, et al. Germline SAMD9 and SAMD9L mutations are associated with extensive genetic evolution and diverse hematologic outcomes. JCI Insight 3(2018).
32. Bluteau O, et al. A landscape of germ line mutations in a cohort of inherited bone marrow failure patients. Blood 131, 717–732 (2018). [PubMed: 29146883]
33. Pastor VB, et al. Constitutional SAMD9L mutations cause familial myelodysplastic syndrome and transient monosomy 7. Haematologica 103, 427–437 (2018). [PubMed: 29217778]
34. de Jesus AA, et al. Distinct interferon signatures and cytokine patterns define additional systemic autoinflammatory diseases. J Clin Invest (2020).
35. Ishiwa S, et al. A girl with MIRAGE syndrome who developed steroid-resistant nephrotic syndrome: a case report. BMC Nephrol 21, 340 (2020). [PubMed: 32787808]
36. Ahmed IA, et al. Outcomes of Hematopoietic Cell Transplantation in Patients with Germline SAMD9/SAMD9L Mutations. Biol Blood Marrow Transplant 25, 2186–2196 (2019). [PubMed: 31306780]
37. Yoshida M, et al. Prevalence of germline GATA2 and SAMD9/9L variants in paediatric haematological disorders with monosomy 7. Br J Haematol (2020).
38. Bamshad MJ, et al. Exome sequencing as a tool for Mendelian disease gene discovery. Nat Rev Genet 12, 745–755 (2011). [PubMed: 21946919]
39. Wlodarski MW, Collin M & Horwitz MS GATA2 deficiency and related myeloid neoplasms. Semin Hematol 54, 81–86 (2017). [PubMed: 28637621]
40. Revy P, Kannengiesser C & Fischer A Somatic genetic rescue in Mendelian haematopoietic diseases. Nat Rev Genet 20, 582–598 (2019). [PubMed: 31186537]
41. Huisman SA, Redeker EJ, Maas SM, Mannens MM & Hennekam RC High rate of mosaicism in individuals with Cornelia de Lange syndrome. J Med Genet 50, 339–344 (2013). [PubMed: 23505322]
42. Erickson RP Somatic gene mutation and human disease other than cancer: an update. Mutat Res 705, 96–106 (2010). [PubMed: 20399892]
43. Jordan DM, et al. Identification of cis-suppression of human disease mutations by comparative genomics. Nature 524, 225–229 (2015). [PubMed: 26123021]
44. Leipe DD, Koonin EV & Aravind L STAND, a class of P-loop NTPases including animal and plant regulators of programmed cell death: multiple, complex domain architectures, unusual phyletic patterns, and evolution by horizontal gene transfer. J Mol Biol 343, 1–28 (2004). [PubMed: 15381417]
45. Cordero MD, Alcocer-Gomez E & Ryffel B Gain of function mutation and inflammasome driven diseases in human and mouse models. J Autoimmun 91, 13–22 (2018). [PubMed: 29610014]
46. Nagamachi A, et al. Multiorgan failure with abnormal receptor metabolism in mice mimicking Samd9/9L syndromes. J Clin Invest 131(2021).

47. Sarthy J, et al. Poor outcome with hematopoietic stem cell transplantation for bone marrow failure and MDS with severe MIRAGE syndrome phenotype. Blood Adv 2, 120–125 (2018). [PubMed: 29365320]

REFERENCES

1. Baumann I, Niemeyer CM, Bennett JM. Childhood myelodysplastic syndrome, in: WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues. in The 2016 Revision to the World Health Organization Classificatio of Myelodysplastic Syndromes (ed. al., S.S.e.) (International Agency for Research on Cancer, Lyon 2017).
2. Tesi B, et al. Gain-of-function SAMD9L mutations cause a syndrome of cytopenia, immunodeficiency, MDS and neurological symptoms. Blood (2017).
3. Bluteau O, et al. A landscape of germ line mutations in a cohort of inherited bone marrow failure patients. Blood 131, 717–732 (2018). [PubMed: 29146883]
4. Pastor VB, et al. Constitutional SAMD9L mutations cause familial myelodysplastic syndrome and transient monosomy 7. Haematologica 103, 427–437 (2018). [PubMed: 29217778]
5. 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). [PubMed: 25741868]
6. Bailey MH, et al. Comprehensive Characterization of Cancer Driver Genes and Mutations. Cell 173, 371–385 e318 (2018). [PubMed: 29625053]
7. Ma X, et al. Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. Nature 555, 371–376 (2018). [PubMed: 29489755]
8. Grobner SN, et al. The landscape of genomic alterations across childhood cancers. Nature 555, 321–327 (2018). [PubMed: 29489754]
9. Zhang MY, et al. Germline ETV6 mutations in familial thrombocytopenia and hematologic malignancy. Nat Genet 47, 180–185 (2015). [PubMed: 25581430]
10. Lindsley RC, et al. Prognostic Mutations in Myelodysplastic Syndrome after Stem-Cell Transplantation. N Engl J Med 376, 536–547 (2017). [PubMed: 28177873]
11. Papaemmanuil E, et al. Clinical and biological implications of driver mutations in myelodysplastic syndromes. Blood 122, 3616–3627; quiz 3699 (2013). [PubMed: 24030381]
12. Pastor V, et al. Mutational landscape in children with myelodysplastic syndromes is distinct from adults: specific somatic drivers and novel germline variants. Leukemia 31, 759–762 (2017). [PubMed: 27876779]
13. Wang K, Li M & Hakonarson H ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 38, e164 (2010). [PubMed: 20601685]
14. Ioannidis NM, et al. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. Am J Hum Genet 99, 877–885 (2016). [PubMed: 27666373]
15. Nykamp K, et al. Sherloc: a comprehensive refinement of the ACMG-AMP variant classification criteria. Genet Med 19, 1105–1117 (2017). [PubMed: 28492532]
16. Shihab HA, et al. An integrative approach to predicting the functional effects of non-coding and coding sequence variation. Bioinformatics 31, 1536–1543 (2015). [PubMed: 25583119]
17. Mather CA, et al. CADD score has limited clinical validity for the identification of pathogenic variants in noncoding regions in a hereditary cancer panel. Genet Med 18, 1269–1275 (2016). [PubMed: 27148939]
18. Quang D, Chen Y & Xie X DANN: a deep learning approach for annotating the pathogenicity of genetic variants. Bioinformatics 31, 761–763 (2015). [PubMed: 25338716]
19. Ng PC & Henikoff S SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res 31, 3812–3814 (2003). [PubMed: 12824425]
20. Adzhubei IA, et al. A method and server for predicting damaging missense mutations. Nat Methods 7, 248–249 (2010). [PubMed: 20354512]
21. Schwarz JM, Rodelsperger C, Schuelke M & Seelow D MutationTaster evaluates disease-causing potential of sequence alterations. Nat Methods 7, 575–576 (2010). [PubMed: 20676075]

22. Reva B, Antipin Y & Sander C Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic Acids Res 39, e118 (2011). [PubMed: 21727090]
23. Lek M, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature 536, 285–291 (2016). [PubMed: 27535533]
24. Kircher M, et al. A general framework for estimating the relative pathogenicity of human genetic variants. Nature Genetics 46, 310–+ (2014). [PubMed: 24487276]
25. van der Velde KJ, et al. GAVIN: Gene-Aware Variant INterpretation for medical sequencing. Genome Biol 18, 6 (2017). [PubMed: 28093075]
26. Liu X, Wu C, Li C & Boerwinkle E dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs. Hum Mutat 37, 235–241 (2016). [PubMed: 26555599]
27. Gelb BD, et al. ClinGen's RASopathy Expert Panel consensus methods for variant interpretation. Genet Med 20, 1334–1345 (2018). [PubMed: 29493581]
28. Duployez N, et al. Familial myeloid malignancies with germline TET2 mutation. Leukemia 34, 1450–1453 (2020). [PubMed: 31827242]
29. Kaasinen E, et al. Impact of constitutional TET2 haploinsufficiency on molecular and clinical phenotype in humans. Nat Commun 10, 1252 (2019). [PubMed: 30890702]
30. Takaoka K, et al. A germline HLTF mutation in familial MDS induces DNA damage accumulation through impaired PCNA polyubiquitination. Leukemia 33, 1773–1782 (2019). [PubMed: 30696947]
31. Buonocore F, et al. Somatic mutations and progressive monosomy modify SAMD9-related phenotypes in humans. J Clin Invest (2017).
32. Gohring G, et al. Complex karyotype newly defined: the strongest prognostic factor in advanced childhood myelodysplastic syndrome. Blood 116, 3766–3769 (2010). [PubMed: 20802024]

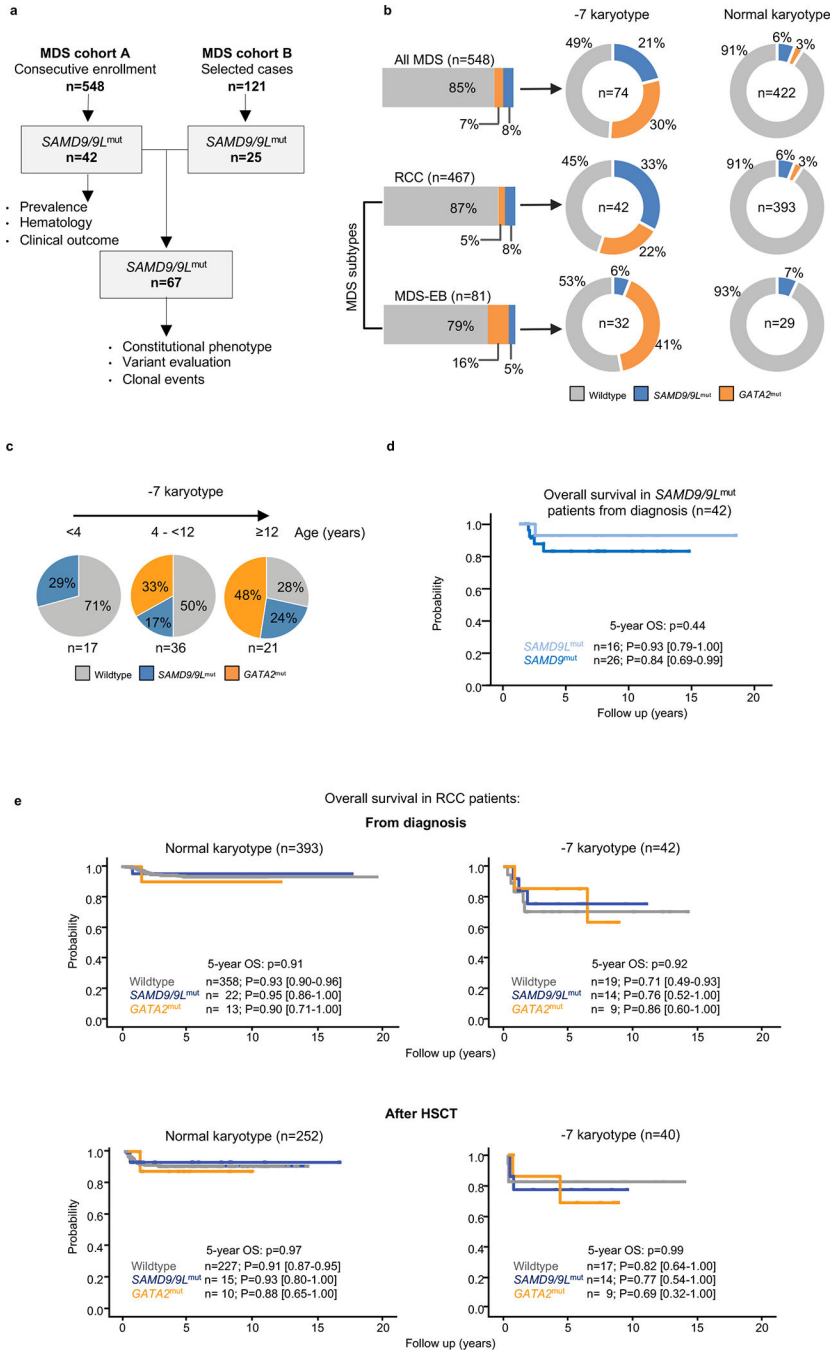


Figure 1. Study cohorts, prevalence of germline *SAMD9*, *SAMD9L* and *GATA2* mutations and overall survival.

(**a**) Study design: cohort A included 548 patients with myelodysplastic syndromes (MDS) consecutively registered in Germany between 1998 and 2016 and cohort B comprised 121 MDS patients selected for possible SAMD9 and SAMD9L (SAMD9/9L) disease-related phenotypes. (**b**) Prevalence of germline mutations (*GATA2, SAMD9/9L*) in cohort A in all MDS and the MDS disease subtypes refractory cytopenia of childhood (RCC) and MDS with excess blasts (MDS-EB). Prevalence within karyotype groups is depicted by donut

plots. (**c**) Prevalence of *SAMD9/9L* or *GATA2* germline mutations in MDS with monosomy 7 (−7) across the age spectrum in cohort A. (d-e) Kaplan-Meier survival curves depicting the probability of overall survival (OS) from diagnosis or after hematopoietic stem cell transplantation (HSCT). Log-rank tests are employed to compare the 5-year OS. P-values are indicated within the graph.

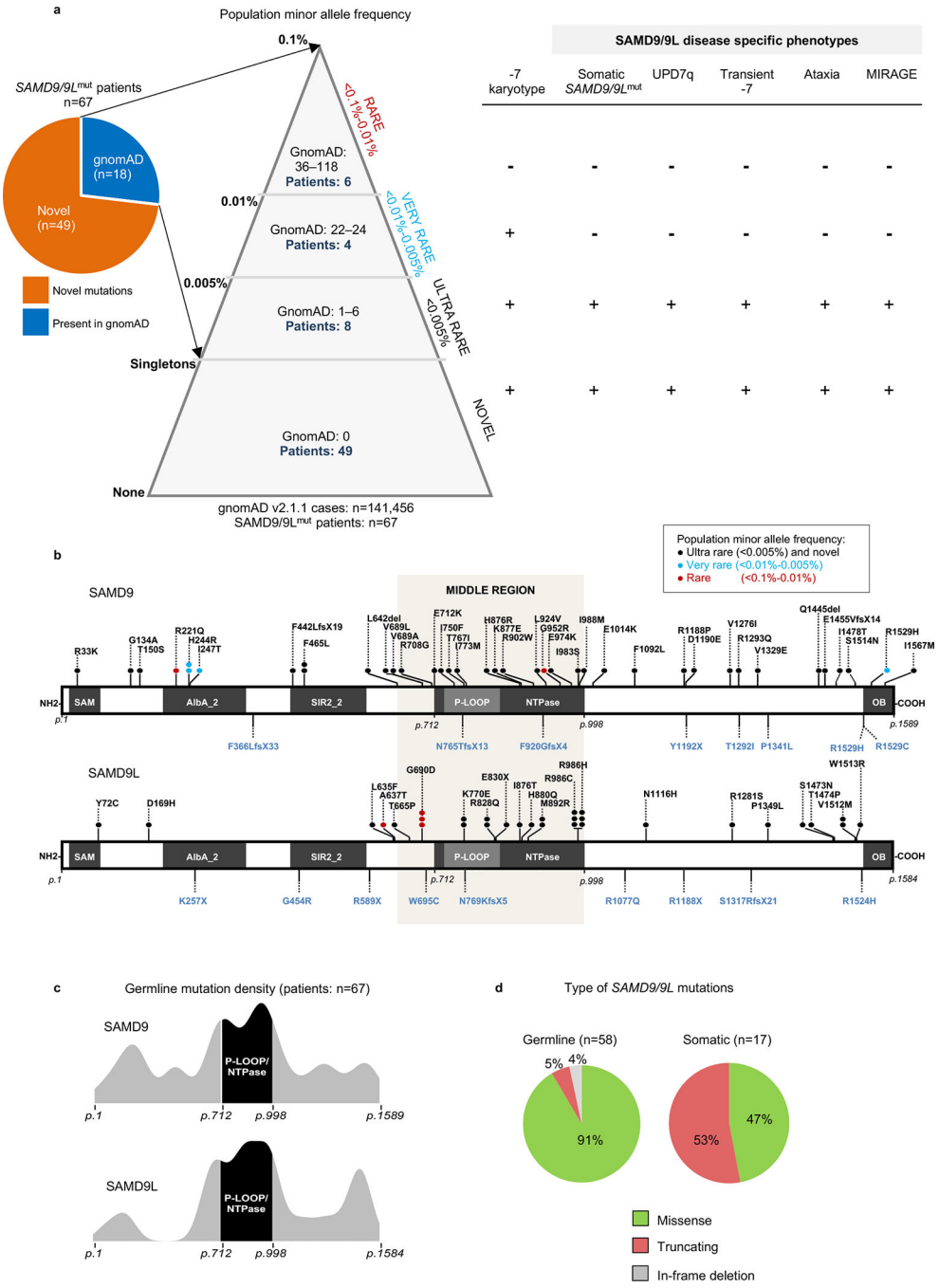


Figure 2. Mutational landscape in *SAMD9* and *SAMD9L* genes.

(**a**) Ranking of germline *SAMD9* and *SAMD9L* mutations (*SAMD9/9L*^{mut}) based on gnomAD population allele frequency. The numbers inside the pyramid show the number of gnomAD individuals or patients (n=18) from our study within the defined population minor allele frequency categories. Grey horizontal lines indicate frequency thresholds. Presence (+) or absence (−) of reported SAMD9/9L disease-specific phenotypes found in patients is marked on the right. These phenotypes include second-site somatic *SAMD9/9L*^{mut}, uniparental isodisomy 7q (UPD7q), transient monosomy 7 (−7), ataxia and syndrome

complex of myelodysplasia, infections, growth restriction, adrenal hypoplasia, genital phenotypes, and enteropathy (MIRAGE). Monosomy 7 (−7) karyotype was not considered SAMD9/9L disease-specific and is shown for comparison. (**b**) Distribution of germline (black font) and somatic (blue font) mutations across SAMD9 or SAMD9L proteins. Predicted protein domains are indicated inside each bar, dots represent single patients, and colors refer to gnomAD population allele frequency. (**c**) Density plot illustrating the *SAMD9/9L* germline mutational distribution. (**d**) Proportion of mutational subtypes of germline or somatic *SAMD9/9L* mutations.

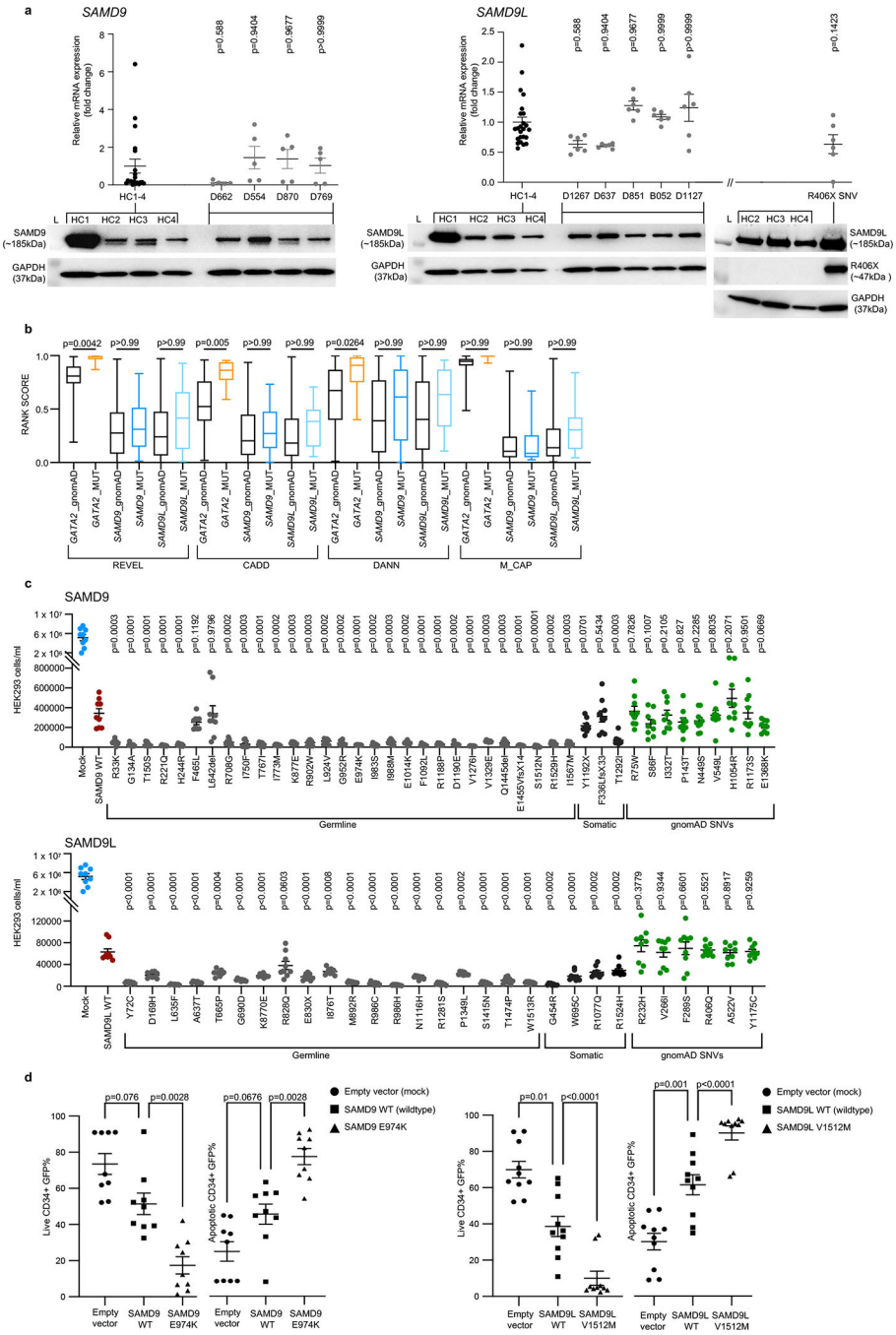


Figure 3. Assessment of *SAMD9* and *SAMD9L* mutations.

(a) *SAMD9* and *SAMD9L* (*SAMD9/9L*) mRNA and protein expression was measured in primary fibroblasts from n=9 *SAMD9/9L*^{mut} patients, n=4 healthy controls (HC), and n=1 individual with germline truncating *SAMD9L* R406X single nucleotide variant (SNV), present at a frequency of 612/280244 alleles in gnomAD. All data points from n=2 independent RT-PCR experiments with mean ± SEM were plotted, and p-values were calculated using ordinary one-way ANOVA. One representative western blot from n=2 independent analysis is shown. (**b**) Rank scores of *GATA2*, *SAMD9*, and *SAMD9L*

population variants (gnomAD) or patient-specific germline mutations (MUT) assessed using 4 variant prediction algorithms. The Y-axis depicts rank scores (0, benign; 1, pathogenic). Number of gnomAD missense variants: *GATA2* n=217, *SAMD9* n=767, *SAMD9L* n=716; number of patient germline missense mutations from this study: *GATA2* n=19, *SAMD9* n=33, and *SAMD9L* n=20 (Extended Data Fig. 2a–b). The minimum, lower quartile, median, upper quartile, and maximum of each analyzed data set is depicted by boxplots and the p-values are calculated using Kruskal-Wallis one-way analysis of variance. (**c**) Bar graphs depicting growth inhibition of HEK293 cells upon overexpression of SAMD9 (n=42) or SAMD9L (n=30) mutant constructs. The p-values (mean ± SD from at least n=3 independent studies in triplicate, paired two-tailed *t*-test) are shown for each mutant in comparison to the respective wildtype constructs. *SAMD9L* R1524H mutation was confirmed as either germline (D791) or somatic (D769) and is depicted here under somatic category. (**d**) Effect of SAMD9/9L wildtype and mutant (SAMD9 E974K, SAMD9L V1512M) lentiviral overexpression on proliferation and apoptosis in healthy donor CD34+ cells. 200,000 CD34+ cells were transduced and assessed after 24 hours by flow cytometry for GFP expression and Annexin V/DAPI (live and apoptotic cells). All data points are from n=3 independent experiments with mean ± SEM and p-values were calculated using paired two-tailed *t*-test.

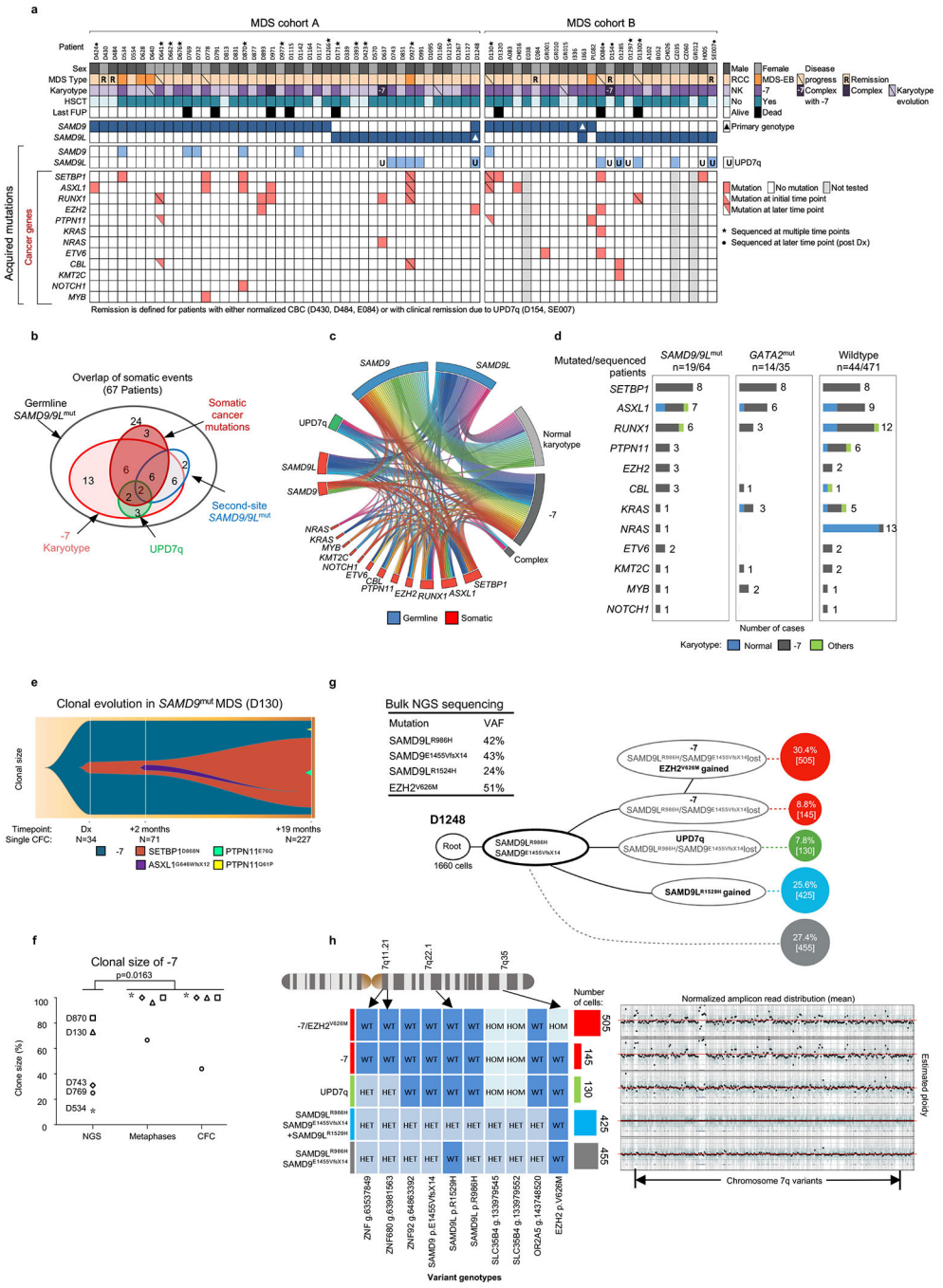


Figure 4. Clonal events in MDS patients with germline *SAMD9* and *SAMD9L* mutations.

(**a**) Clinical status and genetic findings in 67 patients with germline *SAMD9* or *SAMD9L* mutations (*SAMD9/9L*^{mut}). Germline *SAMD9/9L*^{mut} are depicted using dark blue boxes, all other mutations below are somatic. Abbreviations: MDS, myelodysplastic syndrome; HSCT, hematopoietic stem cell transplantation; FUP, follow up; RCC, refractory cytopenia of childhood; MDS-EB, MDS with excess blasts; −7, monosomy 7; NK, normal karyotype; Dx, diagnosis; CBC, complete blood counts; UPD, uniparental isodisomy. (**b**) Venn diagram with overlaps between benign (*SAMD9/9L*^{mut} or UPD7q) and malignant (−7 karyotype

and somatic cancer mutations) somatic genetic rescue events. (**c**) Circos plot delineating interconnections between germline and somatic *SAMD9/9L*^{mut}, somatic cancer mutations, and karyotypes. Lines represent individual cases. (**d**) Somatic cancer mutations identified in *SAMD9/9L*^{mut} patients (from cohort A and B) compared to *GATA2*-mutated (*GATA2*^{mut}) and wildtype MDS from cohort A. Details on genomic studies appear in online methods. (**e**) Clonal evolution depicted from sequencing single colony forming units (CFC) derived from serial bone marrow samples for one representative *SAMD9*^{mut} patient (D130). (**f**) Comparison of monosomy 7 (−7) clone size in n=5 patients using next generation sequencing (NGS), metaphase karyotyping, and CFC assay. P-value was calculated using paired two-tailed *t*-test. (**g**) Clonal architecture in bone marrow of patient D1248 inferred from single-cell DNA sequencing with 5 distinct clonal populations identified: native hematopoiesis (grey), somatic second-site *SAMD9L* mutation (blue), UPD7q (green), −7 and −7 with somatic *EZH2* mutation (both red). Root denotes analyzed single-cell numbers and bolded circle symbolizes ancestral clone with germline *SAMD9/9L*^{mut}. Percentage and number of single cells appear within colored circles. Variant allele frequency (VAF) from bulk sequencing is shown for reference. (**h**) Genotype annotation of individual clones in D1248 (same color coding as in g). Selected variants flanking *SAMD9L* locus have either wildtype (WT), heterozygous (HET) or homozygous (HOM) genotype states. Right panel depicts normalized amplicon read distribution from informative variants, with red line marking the diploid state referenced from diploid clone. Results from additional patients are shown in Extended Data Fig. 8a–c. Panel g was created with BioRender.com.

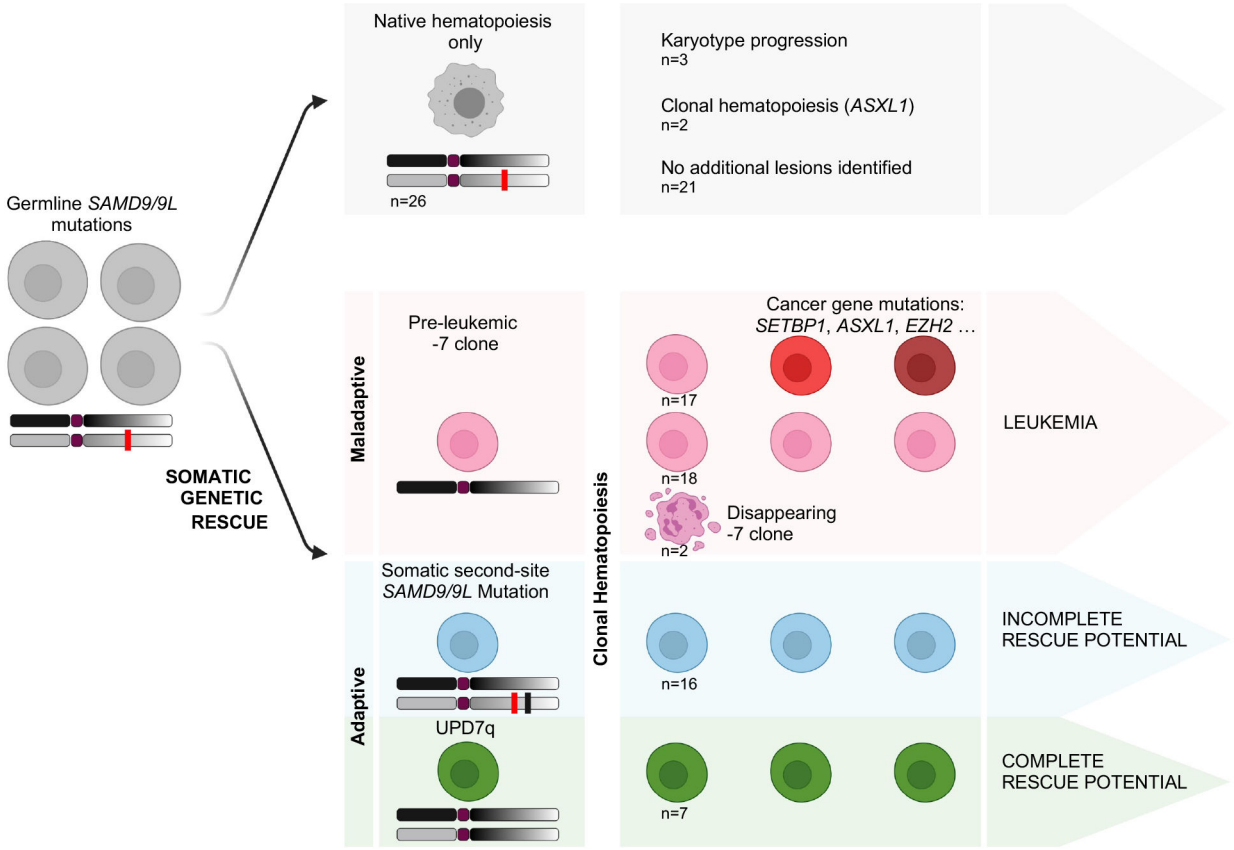


Figure 5. Trajectories of clonal hematopoiesis arising from germline *SAMD9* and *SAMD9L* mutations.

Four major states of hematopoiesis were inferred from comprehensive molecular and cytogenetic characterization of 67 patients with germline *SAMD9/9L* mutations. Upper panel represents patients with native state hematopoiesis only (no rescue events identified at chromosome 7 or within *SAMD9/9L* gene locus) and low rates of clonal progression. Lower panel depicts cases with somatic genetic rescue events resulting in adaptive (benign) clonal hematopoiesis including UPD7q with complete rescue potential and somatic *SAMD9/9L* mutations associated with incomplete rescue, or maladaptive clones with monosomy 7 (−7) that can gain additional cancer mutations. Twenty-six patients had overlapping somatic genetic rescue events as depicted in Fig. 4a. Created with BioRender.com.

Table 1.

Baseline characteristics of 548 children with consecutively diagnosed MDS (cohort A)

	Cohort A: n=548				
	***SAMD9/9L*^{mut} 8% (n=42)**	***GATA2*^{mut} 7% (n=38)**	**Wildtype 85% (n=468)**	**P-value *SAMD9/9L*^{mut} vs. Wildtype**	**P-value *GATA2*^{mut} vs. Wildtype**
Sex: male, % (# pts)	57% (24)	47% (18)	58% (272)	0.90	0.20
Age at diagnosis in years, median (range)	10.0 (0.2–17.6)	12.5 (5.2–16.8)	9.9 (0.8–18.7)	0.84	**0.007**
Diagnosis, % (# pts)				0.45	**0.001**
Refractory cytopenia of childhood	90% (38)	66% (25)	86% (404)		
MDS with excess blasts	10% (4)	34% (13)	14% (64)		
Hypocellular bone marrow, % (# pts)	76% (32)	61% (23)	77% (362)	0.86	**0.02**
Karyotype*, % (# pts)				**<0.001**	**<0.001**
Normal	57% (24)	35% (13)	85.5% (385)		
−7	38% (16)	57% (21)	8% (36)		
Structural complex**	5% (2)	3% (1)	1% (4)		
Other	0%	5% (2)	5.5% (25)		
Familial disease	2% (1)	18% (7)	1.5% (7)	0.66	**<0.001**
Alive at last follow up, % (# pts)	88% (37)	76% (29)	90% (423)	0.63	**0.007**

*Karyotype available in 529/548 patients; missing karyotype in n=1 patient with *GATA2*^{mut} and n=18 with wildtype genotype. The Mann-Whitney test with Bonferroni correction was used for comparison of differences in medians of continuous variables. Chi-Square-test was used to compare differences in percentages for groups. Bolded p-values are significant.

**Defined by the presence of 3 chromosomal abnormalities, including at least 1 structural aberration.

Abbreviations: # pts, number of patients; MDS, myelodysplastic syndromes; −7, monosomy 7 or deletion 7q.