Published Ahead of Print on December 7, 2017, as doi:10.3324/haematol.2017.180778.

Constitutional SAMD9L mutations cause familial myelodysplastic syndrome and transient monosomy 7

by Victor B. Pastor, Sushree Sahoo, Jessica Boklan, Georg C. Schwabe, Ebru Saribeyoglu, Brigitte Strahm, Dirk Lebrecht, Matthias Voss, Yenan T. Bryceson, Miriam Erlacher, Gerhard Ehninger, Marena Niewisch, Brigitte Schlegelberger, Irith Baumann, John C. Achermann, Akiko Shimamura, Jochen Hochrein, Ulf Tedgård, Lars Nilsson, Henrik Hasle, Melanie Boerries, Hauke Busch, Charlotte M. Niemeyer, and Marcin W. Wlodarski

Haematologica 2017 [Epub ahead of print]

Citation: Victor B. Pastor, Sushree Sahoo, Jessica Boklan, Georg C. Schwabe, Ebru Saribeyoglu, Brigitte Strahm, Dirk Lebrecht, Matthias Voss, Yenan T. Bryceson, Miriam Erlacher, Gerhard Ehninger, Marena Niewisch, Brigitte Schlegelberger, Irith Baumann, John C. Achermann, Akiko Shimamura, Jochen Hochrein, Ulf Tedgård, Lars Nilsson, Henrik Hasle, Melanie Boerries, Hauke Busch, Charlotte M. Niemeyer, and Marcin W. Wlodarski . Constitutional SAMD9L mutations cause familial myelodysplastic syndrome and transient monosomy 7. Haematologica. 2017; 102:xxx doi:10.3324/haematol.2017.180778

Constitutional SAMD9L mutations cause familial myelodysplastic syndrome and transient monosomy 7

Victor B. Pastor[1,2*], Sushree Sahoo[1,2,3*], Jessica Boklan[4], Georg C Schwabe[5], Ebru Saribeyoglu[5], Brigitte Strahm[1], Dirk Lebrecht[1], Matthias Voss[6], Yenan T. Bryceson[6], Miriam Erlacher[1,7], Gerhard Ehninger[8], Marena Niewisch[1], Brigitte Schlegelberger[9], Irith Baumann[10], John C. Achermann[11], Akiko Shimamura[12], Jochen Hochrein[13], Ulf Tedgård[14], Lars Nilsson[15], Henrik Hasle[16], Melanie Boerries[7,13], Hauke Busch[13,17], Charlotte M. Niemeyer[1,7], and Marcin W. Wlodarski[1,7]

1) Department of Pediatrics and Adolescent Medicine, Division of Pediatric Hematology and Oncology, Medical Center, Faculty of Medicine, University of Freiburg, Germany
2) Faculty of Biology, University of Freiburg, Germany
3) Spemann Graduate School of Biology and Medicine, University of Freiburg, Germany
4) Center for Cancer and Blood Disorders, Phoenix Children's Hospital, Phoenix, AZ, USA
5) Children's Hospital, Carl-Thiem-Klinikum Cottbus, Germany
6) Department of Medicine, Huddinge, Hematology and Regenerative Medicine, Karolinska Institute, Stockholm, Sweden
7) German Cancer Consortium (DKTK), Freiburg, Germany and German Cancer Research Center (DKFZ), Heidelberg, Germany
8) Internal Medicine of Hematology/Medical Oncology, University Hospital, Dresden, Germany
9) Institute of Human Genetics, Hannover Medical School, Hannover, Germany
10) Clinical Centre South West, Department of Pathology, Böblingen Clinics, Böblingen, Germany
11) Genetics & Genomic Medicine, UCL Great Ormond Street Institute of Child Health, University College London, London, UK
12) Boston Children's Hospital, Dana Farber Cancer Institute, and Harvard Medical School, MA, USA
13) Institute of Molecular Medicine and Cell Research, University of Freiburg, Freiburg, Germany
14) Department of Pediatric Oncology and Hematology, Skåne University Hospital, Lund, Sweden
15) Department of Hematology, Oncology and Radiation Physics, Skåne University Hospital, Lund, Sweden
16) Department of Pediatrics, Aarhus University Hospital, Aarhus, Denmark
17) Lübeck Institute of Experimental Dermatology, Lübeck, Germany

* Equal contribution

The authors declare no conflict of interest

Corresponding author:
Marcin W. Wlodarski, MD.[1, 7-]
Mathildenstr. 1, 79106 Freiburg, Germany
Phone: + 49 761 270 -45020, Fax: -46160
Email: marcin.wlodarski@uniklinik-freiburg.de

Article type: Original article
Running title: SAMD9L-related familial MDS
Key words: monosomy 7, familial myelodysplastic syndromes, germline predisposition
Abstract / manuscript word count: 202 / 3750 (introduction, methods, results, discussion)
Figures: 6, Tables: 2, References: 40

ABSTRACT

Familial myelodysplastic syndromes arise from haploinsufficiency of genes involved in hematopoiesis and are primarily associated with early-onset disease. Here we describe a familial syndrome in 7 patients from 4 unrelated pedigrees presenting with myelodisplastic syndrome and loss of chromosome 7/7q. Median age at diagnosis was 2.1 (1-42) years, all patients presented with thrombocytopenia with or without additional cytopenias and a hypocellular marrow without increase of blasts. Genomic studies identified constitutional mutations (p.H880Q, p.R986H, p.R986C and p.V1512M) in *SAMD9L* gene on 7q21, with decreased allele frequency in hematopoiesis. The nonrandom loss of mutated *SAMD9L* alleles was attained via monosomy 7, deletion 7q, UPD7q, or acquired truncating *SAMD9L* variants p.R1188X and p.S1317RfsX21. Incomplete penetrance was noted in 30% (3/10) of mutation carriers. Long-term observation revealed divergent outcomes with either leukemic progression and/or accumulation of driver mutations (n=2), persistent monosomy 7 (n=4), and transient monosomy 7 followed by spontaneous recovery with *SAMD9L*-wildtype UPD7q (n=2). Dysmorphic features or neurological symptoms were absent in our patients, pointing to the notion that myelodisplasia with monosomy 7 can be a sole manifestation of SAMD9L disease. Collectively, our results define a new subtype of familial myelodisplastic syndrome and provide an explanation for the phenomenon of transient monosomy 7. www.clinicaltrials.gov; #NCT00047268

INTRODUCTION

Germline predisposition has been recognized as an underlying cause for the development of myelodysplastic syndromes (MDS) in children. Recently, it has also been gaining importance in the etiology of adult myeloid neoplasia (MN), particularly in cases with a positive family history. In addition to inherited bone marrow failure syndromes (IBMFS), a number of genes are known to be associated with heritable forms of MDS and acute myeloid leukemia (AML)(1, 2), i.e. *GATA2*(3), *CEBPA*(4), *RUNX1*(5), *ANKRD26*(6), *ETV6*(7) and *DDX41*(8). Germline mutations in *DDX41* can result in adult onset MN, while aberrations in *RUNX1* and *GATA2* are associated with MN in younger individuals. We recently reported that GATA2 deficiency is the most common genetic cause of primary childhood MDS, accounting for 15% of all advanced MDS cases, and 37% of MDS with monosomy 7 (MDS/-7)(9). However, in the majority of pediatric MDS, and also in a considerable number of familial MN cases, the presumed germline cause has not yet been discovered(10) (11).

Monosomy 7 is the most frequent cytogenetic lesion in children with MDS and, unlike in adults, it often occurs as the sole cytogenetic abnormality(12). Due to the rapid and progressive disease course it is considered an urgent indication for hematopoietic stem cell transplantation (HSCT)(13). However, transient monosomy 7 has been occasionally documented in childhood MDS(14-16). Considering that complete (-7) and partial (del(7q)) deletion of chromosome 7 are common aberrations in MDS, extensive efforts have been undertaken to discern causative tumor suppressor genes located on chromosome 7. Asou and colleagues identified *SAMD9* (Sterile Alpha Motif Domain-containing 9), its paralogue *SAMD9L* (SAMD9-like), and *Miki/HEPACAM2* as commonly deleted genes within a 7q21 cluster in patients with MN(17).

Notably *Samd9l*-haploinsufficient mice were shown to develop myeloid malignancies characterized by different cytopenias and mimicking human disease with monosomy 7(18).

In line with these findings, germline heterozygous gain-of-function (GoF) *SAMD9L* mutations p.H880Q, p.I891T, p.R986C, and p.C1196S were recently discovered in 4 pedigrees with variable degrees of neurological (ataxia, balance impairment, nystagmus, hyperreflexia, dysmetria, dysarthria) and hematological (single to tri-lineage cytopenias, MDS/-7) symptoms. For most carriers, the clinical presentation was compatible with the diagnosis of ataxia-pancytopenia (AT) syndrome(19, 20). Similarly, in two recent studies, we and others reported *de novo* GoF mutations in *SAMD9* in a total of 18 patients with MIRAGE syndrome (myelodysplasia, infection, restriction of growth, adrenal hypoplasia, genital phenotypes, and enteropathy), of whom 4 notably also developed MDS/-7(21, 22), however not all patients develop the full MIRAGE disease spectrum, as documented in one family with SAMD9-related MDS(23). *SAMD9* and *SAMD9L* genes share 62% sequence identity and apart from their putative role as myeloid tumor suppressors, the general function and their specific effect pertaining to hematopoiesis are not well-understood(18).

In this study we aimed to identify the genetic cause in pedigrees with nonsyndromic familial MDS. We discovered constitutional *SAMD9L* mutations associated with nonrandom patterns of clonal escape leading to loss of the mutant allele. We further demonstrate in two cases that *SAMD9L*–related disease can be associated with transient -7, occurring as a one-time clonal event followed by somatic correction of hematopoiesis achieved by UPD7q with double wild type *SAMD9L*.

METHODS

Patients

Diagnosis of MDS was established according to the WHO criteria(24, 25). Patients 1 (P1), 2 (P2), 5 (P5), and 7 (P7) were enrolled in the prospective study 98 of the European Working Group of MDS in Childhood (EWOG-MDS) (www.clinicaltrials.gov; #NCT00047268). Patient 6 (P6) was the father of P5 (family III). Family II (P3 and P4) was referred for evaluation of familial MDS from the Phoenix Children´s Hospital, USA. The study had been approved by the institutional ethics committee (University of Freiburg, CPMP/ICH/135/95 and 430/16). Written informed consent had been obtained from patients and parents.

Genomic studies and bioinformatics

Exploratory whole exome sequencing (WES) was performed in bone marrow granulocytes (BM-GR) in P1 and P2, as outlined in the supplement. Targeted deep sequencing (DS) for *SAMD9*/*SAMD9L* and MDS/IBMFS-related genes was performed in other patients, except for P3 and P6 due to unavailability of material. All relevant variants were validated using Sanger sequencing. For germline confirmation, DNA was extracted from skin fibroblast and/or hair follicles, and targets were amplified and sequenced as previously described(9). The degree of deleteriousness was calculated using the combined annotation-dependent depletion scoring system (CADD-score)(26). The variants with CADD-scores higher than 20 were further evaluated towards their role in hematological disease or cancer, thereby focusing on the top 1% most deleterious variants in the human genome. In addition, the pathogenicity calculations were achieved using standard prediction tools. The evolutionary conservation across species and the

physicochemical difference between amino acids was estimated by PhyloP, PhastCons and Grantham score respectively(27). Mutant clonal size was inferred from allelic frequencies and the total number of NGS reads normalized to the ploidy level. Further details are provided in supplementary methods.

Evaluation of variant allelic configuration

Genomic DNA of P1 collected at the time point of progress to chronic myelomonocytic leukemia (CMML) was amplified to obtain a 1333 bp region encompassing both *SAMD9L* mutations p.V1512M (germline) and p.R1188X (acquired). PCR products were TA-cloned and sequenced as previously reported(28). Sequences of 170 colonies were evaluated for the presence of *SAMD9L* mutations.

Cellular and functional studies

Metaphase karyotyping and interphase fluorescence in situ hybridization (FISH) were performed using bone marrow (BM) specimens according to standard procedures(12). Human colony forming cell (CFC) assays were performed in P1 (at CMML disease stage) and in P7 (at diagnosis) as previously described(29). Further, to evaluate the effect of the patient-derived *SAMD9L* p.V1512M and p.R986C mutations on cellular proliferation, 293FT cells were dye-labeled and consequently transfected with wild type or mutant teal fluorescent protein (TFP)-SAMD9L as previously described(20). The transfected cells were tracked by flow cytometry for TFP-SAMD9L expression and dye dilution as an indicator of cell division. *SAMD9L* variant p.T233N recently reported as “disease-protective” (20) was used as control.

RESULTS

Clinical phenotype of patients

Patients P1-P6 (4 males and 2 females) belong to 3 unrelated families of German descent and were diagnosed with *bona fide* familial MDS after the exclusion of known IBMFS achieved by targeted sequencing and functional tests (Figure 1, Table 1). Index patient P7 is the only child of a non-consanguineous Swedish family. Detailed clinical description of all patients is outlined in the supplement. All affected individuals had normal measurements without dysmorphic stigmata at birth and at last follow-up (Table 1). Psychomotor development and neurocognitive function were normal, in particular no ataxia or movement disorder were diagnosed in the 10 mutation carriers (7 patients and 3 silent carriers with *SAMD9L* mutations). Previous family histories were unremarkable for cytopenias, neurological disease, malignancies, or stillbirths, with the exception of father of P7 who at the last follow-up presented with unclear ataxia. Prior non-invasive recurrent respiratory tract infections were noted in 3/7 patients (P1, P2, and P4) and endogenous eczema in 2 (P1 and P3). Moreover, P1 developed transient pancytopenia during infancy and suffered from self-limiting seizures during infancy with no structural brain abnormalities or neurological deficits identified. Peripheral blood findings at diagnosis included isolated thrombocytopenia in one (P2), thrombocytopenia with neutropenia in two (P1, P3) and pancytopenia in 4 patients (P4-7). MCV was elevated in 5/7 patients at diagnosis, and HbF in 2 out of 3 tested patients. MDS manifested at the median age of 2.1 (1.0-42) years as hypocellular RCC or in the adult (P6) refractory cytopenia with multilineage dysplasia (RCMD, Table 1). Severe dysplasia with vacuolization was observed in 2 patients (P1-2). A common cytogenetic feature in all patients was the complete or partial loss of chromosome 7 (Table 1).

The clinical course was remarkable for several patients. P1 developed 3 years after the initial diagnosis of RCC severe infections and hepatosplenomegaly, and for the first time required platelet transfusions, while his blood smear depicted 21% blasts compatible with the diagnosis of CMML (with no *in vitro* GM-CSF hypersensitivity), (Figure 2). Following HSCT, he developed late-onset acute GVHD and died from cerebral hemorrhage. At the same time, his younger sister (P2) was diagnosed with RCC/-7, however the parents decided against follow-up in hematology clinic. Unexpectedly, her complete blood count (CBC) normalized 3.7 years later and remained stable until last follow-up 20 years after the initial diagnosis. She achieved a spontaneous remission as supported by normalization of BM morphology/cellularity and cytogenetics (Figure 2, Table 1). A similar clinical picture was seen in P7 who was diagnosed with RCC with a -7 clone at the age of 2.1 years, and experienced rapid cytogenetic remission with normal marrow and blood counts until the last follow-up, 16 years after diagnosis (Table 1). In P4, shortly after initial manifestation, hematology normalized, however FISH revealed chromosome 7 loss in BM, which gradually increased and after 3.5 months culminated in the emergence of two independent clones with -7 and del(7q) (Table 1).

Due to high-risk cytogenetics and disease progression, 5/7 patients (P1, P3-6) underwent HSCT after myeloablative conditioning. At last follow-up, 6 out of 7 patients were alive, 4 after HSCT, and 2 without therapy.

Constitutional and acquired *SAMD9L* mutations

Exploratory WES performed in family I identified two shared candidate variants in P1 and P2 evaluated as highly conserved and deleterious by *in silico* prediction: *SAMD9L* (p.V1512M) and

PTEN (p.Y188C), (Table 2, Figure S1). Sequencing of DNA from hair follicles confirmed the constitutional nature of both novel mutations. *SAMD9L* p.V1512M variant was inherited from the mother (Figure 1A) whereas *PTEN* p.Y188C was of paternal origin; both parents were asymptomatic and had normal CBC at the time of testing. Finally, truncating acquired *SAMD9L* mutation p.R1188X (VAF 5.9%) was identified in P1 in hematopoiesis (Table 2).

In pedigree II, targeted NGS in P4 revealed *SAMD9L* p.R986H as the most plausible candidate constitutional mutation predicted to be highly conserved and deleterious (Table 2, Figure 1A-B). This mutation was found in 4 individuals in ExAC (out of 120976 alleles). Additional missense variants in *JAK3* p.A877V (ExAC: 11 individuals, 121372 alleles), and *FANCM* p.L57F (ExAC: 195 individuals, 121190 alleles) had lower and moderate pathogenicity scores, respectively (Table 2). Chromosomal breakage studies on P4 were negative thus arguing against a pathogenic role of the heterozygous FANCM variant. Germline analysis revealed *SAMD9L* and *JAK3* variants in P3, P4, and their father, while the *FANCM* variant was transmitted from the mother only to P4. Both parents were asymptomatic.

In pedigree III, the *SAMD9L* p.R986C mutation was identified in P5 and the affected father, P6 (Figure 1A). This mutation has been reported in a family with ataxia-pancytopenia phenotype, with 1 of 3 carriers developing MDS/-7 at the age of 18 months(20). The HLA-identical brother of P5 was thoroughly evaluated as a potential BM donor. He was clinically healthy and had normal CBC, but he did not qualify as a donor because of hypocellular BM with mild dysplastic features. He was also carrier of the p.R986C mutation.

In P7 of pedigree IV, targeted NGS identified two *SAMD9L* mutations (Table 2): missense p.H880Q with VAF of 27% out of 8139 reads (likely constitutional; this mutation was reported in multiple individuals within a family with ataxia pancytopenia but no MDS phenotype) and nonsense p.S1317RfsX21 likely acquired in a subclone as inferred from the much lower VAF of 10% (5934 reads). In summary, inherited *SAMD9L* mutations p.V1512M, p.R986H, p.R986C were identified in 3 families (each with 2 individuals diagnosed with MDS/-7 and 1 healthy carrier, Figure 1A-B), and p.H880Q in P7 who presented with transient monosomy 7.

Acquired mutations in known oncogenes

All patients with exception of P3 and P6 were evaluated for the presence of somatic mutations in leukemia-associated genes using WES or targeted NGS. In P1 previously reported leukemia driver mutations *SETBP1* p.D868N, *EZH2* p.V582M, *KRAS* p.Q61P were identified as somatic (Table 2, Figure S2). Similarly, in P5 a somatic *RUNX1* mutation c.413_427+5dup20bp was detected (Table 2). This is a novel splice-donor site mutation with no occurrence in databases. No additional mutations in leukemia driver genes were observed in other affected cases.

Clonal escape mechanisms from *SAMD9L* mutations are not random

In comparison with other constitutional variants, *SAMD9L* missense mutations showed significantly lower median allelic frequencies across all patients (53% vs. 20%, $p<0.05$, Table 2). This finding was corroborated by the consistent partial or complete loss of chromosome 7 (Figure 3, Table 1). In P1, -7 progressively expanded from 60% at time of RCC diagnosis to 95% at CMML progression. In addition, P1 and P7 harbored subclones with acquired stop-gain

SAMD9L mutations located upstream of the constitutional missense substitutions (Figure 3). TA cloning confirmed *cis* orientation of both mutations on the same chromosome in P1 (Figure 1C). Western blot of 293T cells transiently transfected with double-mutated (p.V1512M/p.R1188X) SAMD9L plasmid revealed stable expression of truncated protein (not shown). In P2 and P7, the initial -7 disappeared and was replaced by a large somatic clone with uniparental disomy (UPD) 7q that duplicated the paternal wildtype *SAMD9L* allele (Figure 3-4).

Natural history of SAMD9L-related MDS reveals divergent clinical outcomes

We next studied in detail the disease course in patients who were left untreated for longer periods of time. In P4, the evolution from normal karyotype to independent clones with -7 and del(7q) was rapid and occurred within a few months (Table 1). Cytogenetic progress was associated with partial recovery of CBC and normalization of BM cellularity, pointing to the possibility of HSC niche repopulation by -7/del(7q) retaining only the wildtype *SAMD9L* allele. P1 developed CMML 3.6 years after initial diagnosis of MDS/-7. This patient carried somatic driver mutations representing major subclones co-existing within -7 background (Table 2). Additional somatic *SAMD9L* mutation p.R1188X (co-occuring in *cis*- with p.V1512M) was found in a minor clonal fraction of ~6%.

In contrast, P2 and P7 showed an unexpected clinical course with spontaneous hematologic recovery, disappearing monosomy 7 and the presence of large double wildtype UPD7q clone in BM (Figure 4-5). Both patients remained healthy, had normal follow-up BM examinations with no signs of dysplasia (Table 1), and normal CBC until last FUP, 20 (P1) and 16 (P7) years after initial diagnosis.

SAMD9L mutations inhibit cell proliferation

Inhibitory effect on cell proliferation reported for *SAMD9L* mutants overexpressed in 293FT cells *in vitro* was termed as gain-of-function (GoF). In contrary, ectopic expression of p.T233N was shown to mitigate cell proliferation to a smaller extent in comparison to WT *SAMD9L*, and was categorized as a disease-protective or loss-of-function (LoF) variant(20). *SAMD9L* p.R986H was previously functionally studied and shown to be GoF (20), while p.H880Q was shown to induce LOH by -7/del(7q) or UPD7q in an EBV-transformed cell line *in vitro*(19). To determine the effect of the *SAMD9L* MT identified in families I and III, we transiently transfected 293FT cells with vectors containing disease-associated mutations p.V1512M and p.R986C respectively, along with the LoF variant p.T233N. Cell proliferation was assessed in dye dilution assays. Both *SAMD9L* p.V1512M and p.R986C decreased dye dilution in comparison to *SAMD9L* WT and p.T233N (Figure 6A-B), pointing to an amplified growth restrictive effect of the disease-associated variants (Figure 6B).

DISCUSSION

In this study, we describe a familial MDS syndrome caused by heterozygous missense mutations in *SAMD9L* gene located on chromosome 7q21. We present 7 individuals from 4 unrelated pedigrees who developed MDS/-7 from the age of 1 to 42 years without any neurological involvement. Compared to the reported SAMD9L mutation carriers (19, 20), hematology was the sole clinical manifestation in our cases. Other novel findings outlined here are the description of the somatic mutational landscape likely contributing to MDS progression, the observation of transient monosomy 7, and finally the occurrence of nonrandom revertant mosaicism leading to complete hematological recovery.

The *SAMD9L* mutations p.R986H, p.R986C, p.V1512M identified in this cohort affect evolutionary highly conserved amino acid residues, are assessed as pathogenic by *in silico* prediction. The mutation p.H880Q (P7) shows a weak conservation score, however, this mutation had already been reported as causative for the ataxia pancytopenia phenotype (19). We were not able to test *SAMD9L* genetics in P7, however the unclear ataxia this patient had been evaluated for at last visit points to a carrier status and indicates that there must be an overlap between sole hematological and ataxia phenotype in *SAMD9L* disease. Summarizing all *SAMD9L* mutations recently reported or identified in our cohort, a total of 6 germline mutations can be discerned (p.H880Q, p.I981T, p.R986H, p.R986C, p.C1196S, and p.V1512M). Of note, all these mutations cluster exclusively to the C-terminal half of the protein. Further, upon comparing reported mutations in the paralogue gene *SAMD9* (p.R982H/C)(22) with that of the present study in *SAMD9L* (p.R986H/C), we identified a potential mutational hotspot affecting

highly conserved regions in both SAMD9L (p.984-989: GVRIIH) and SAMD9 (p980-985: GVRIIH) proteins.

The reported constitutional variants in both *SAMD9*/*SAMD9L* were classified as GoF based on the observation of decreased cell proliferation in a 293T cell line(20-22). Similarly, we observed growth deficiency in 293T cells harboring *SAMD9L* p.V1512M and p.R986C. Based on these findings, one cautious speculation hints at a GoF effect that is cell toxic. This can be supported by the discovery of an acquired stopgain *SAMD9L* mutations in P1 and P7 that likely "eliminate" germline missense mutations. In the cases studied here, complete or partial deletion of chromosome 7 and also UPD7q was nonrandom and each time resulted in loss of the germline-mutated *SAMD9L* gene copy. Additional studies are essential to further define the effect of *SAMD9L* variants, might be challenging due to the growth inhibitory effect of the alterations. It also remains to be answered if *SAMD9L* missense mutations lead to an increased protein stability, alter the protein structure, enhance an unknown functional domain, or exert a completely neomorphic effect.

We describe 3 silent mutation carriers from separate families demonstrating no previous medical history. Despite normal CBC and MCV, the brother of P5 exhibited a hypocellular marrow with mild dysplasia, evidently attributed to the identical pathogenic *SAMD9L* mutation. This finding emphasizes the necessity of thorough hematological workup including marrow studies in potential sibling donors especially when they lack a genetic marker for familial disease. The intrafamiliar heterogeneity regarding the hematologic presentation remains

elusive; one can speculate that other yet unknown genetic or epigenetic mechanisms might act as modifiers.

Thus far only limited knowledge exists about the regulation and cellular functions of *SAMD9L*. It has been postulated that both *SAMD9L* and the adjacent paralogous *SAMD9* gene share functional redundancy, shaped by a long term, possibly virus-induced selective pressure(30). Both genes can be upregulated by type I and II interferons(20, 31-33) and by this might suppress inflammatory pathways and exert anti-viral properties(31, 34). Although there was no evidence for a defined immunodeficiency in our cohort, 3/7 patients experienced recurrent respiratory infections with or without cytopenias before they developed RCC/-7.

Further evidence supports that *SAMD9*/*SAMD9L* genes act as tumor suppressors as their inactivation is associated with increased cellular proliferation, i.e. in normophosphatemic familiar tumoral calcinosis (*SAMD9*), in hepatitis B virus-associated hepatocellular carcinoma (*SAMD9L*), in MDS/AML with microdeletion in 7q21 (both genes), and in SAMD9L-haploinsufficient mice. Based on the observations in this murine model, the cytokine-receptor complexes cannot be properly degraded due to impaired endosomal function in SAMD9L-haploinsufficient cells, which results in constitutive intracellular signaling with prolonged cell survival(18). Moreover, *SAMD9L*$^{+/-}$ and *SAMD9L*$^{-/-}$ mice develop MDS with normo/hypercellular BM, drawing an a parallel to human SAMD9L-related MDS where the initial marrow hypocellularity associated with “toxic” mutation is restored upon the loss of mutant *SAMD9L* allele. This loss is accomplished through -7, del(7q) or UPD7q, and leads to proliferative advantage with clonal expansion. In our cohort, although initially all patients presented with

hypocellular marrow, longer observation periods in P1, P2, P4, and P7 revealed successive normalization of marrow cell content associated with increasing -7/del(7q) clone or appearance of UPD7q. In P1, the leukemic progression was further aggravated with the accumulation of typical MDS driver mutations in *SETBP1, KRAS* and *EZH2* within the -7 clone. P5 also demonstrated an acquired splice site mutation in *RUNX1,* a known leukemic driver gene. Notably, typical adult MDS driver mutations (i.e. *TET2, DNMT3A, IDH1/2*) were not encountered in our cohort. This is in line with previously published findings discussing rather *SETBP1,* RAS pathway mutations and *RUNX1* (identified in our SAMD9L-mutated patients) as recurrent drivers of pediatric MDS (10). Building on our observations we propose the following mechanism of MDS evolution in *SAMD9L* disease: BM attempts to circumvent the toxicity of constitutional *SAMD9L* mutation and selects for fitter, yet premalignant -7/del(7q) clones (with only one wildtype *SAMD9L* copy), or benign clones with truncated *SAMD9L* (Figure 3). Over time the resulting haploinsufficiency of tumor supressor genes on 7q (e.g. *EZH2* or *CUX1*) in all patients likely provide the first step towards progression. Finally, additional driver somatic mutations might be encountered in some but not all patients.

Somatic revertant mosaicism has been reported in IBMFS with hypocellular BM, including telomeropathy with germline mutations in *TERT*(35), and Fanconi anemia where mosaicism in blood occurs at rates of up to 140x higher than in general population(36, 37). However, in general, revertant mosaicism is a rare facet to clonal hematopoiesis because spontaneous correction of the pathogenic allele is a random event. In our study, we report two patients (P2 and P7) who presented with RCC/-7 at young age and demonstrated complete hematopoietic remission with normal cytogenetics throughout an observation period ranging from 16 to 20

years. The clinical picture of these patients fits the previously described transient monosomy 7 syndrome; to our knowledge 8 patients with primary MDS with transient -7 or del(7q) were reported in the literature (14-16, 38-40). Their age at diagnosis ranged from 8 months to 3 years, and spontaneous remission was achieved within 1-20 months from diagnosis. It seems that the -7 clones in our patients either spontaneously vanished or were outcompeted by "fitter" UPD7q-corrected clones with diploid copy of wild type *SAMD9L* allele. Based on these observations, a watch-and-wait strategy might be proposed for younger patients with RCC/-7 who have no additional somatic driver mutations and are clinically stable. However, prolonged "watchful waiting" poses the risk of progression as witnessed in P1, who advanced to CMML and acquired oncogenic mutations 3.6 years after he was diagnosed of RCC.

In conclusion, our observations establish the molecular basis of a distinct subtype of familial MDS and point to the notion that MDS with chromosome 7 loss can be the sole and common manifestation of *SAMD9L*-related disease. The negative mutational effect leads to escape and outgrowth of clones carrying -7/del(7q) with only wildtype *SAMD9L* allele, which might spontaneously dissapear or persist and provide the first step towards disease progression. Finally, this is the first description of long-term revertant mosaicism due to nonrandom UPD7q in SAMD9L disease, and a plausible explanation for transient monosomy 7 syndrome.

ACKNOWLEDGMENTS

Provided in the supplement.

AUTHOR CONTRIBUTION

MWW, VBP, SS designed the study and wrote the manuscript. VBP, SS, DL, MV, YTB, MB, JH, HB, and MWW performed experiments and analyzed data. CMN, JCA, BS, and ME contributed to study conception and data interpretation. JB, GCS, ES, BS, GE, MN, BSch, IB, AS, UT, LN, HH, CMN were involved in patient care, testing procedures and sample acquisition. All authors contributed to writing and approved the manuscript.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

1. Churpek JE, Godley LA. How I diagnose and manage individuals at risk for inherited myeloid malignancies. Blood. 2016;128(14):1800-1813.
2. Feurstein S, Drazer MW, Godley LA. Genetic predisposition to leukemia and other hematologic malignancies. Semin Oncol. 2016;43(5):598-608.
3. Hahn CN, Chong CE, Carmichael CL, et al. Heritable GATA2 mutations associated with familial myelodysplastic syndrome and acute myeloid leukemia. Nat Genet. 2011;43(10):1012-1017.
4. Smith ML, Cavenagh JD, Lister TA, Fitzgibbon J. Mutation of CEBPA in familial acute myeloid leukemia. N Engl J Med. 2004;351(23):2403-2407.
5. Song WJ, Sullivan MG, Legare RD, et al. Haploinsufficiency of CBFA2 causes familial thrombocytopenia with propensity to develop acute myelogenous leukaemia. Nat Genet. 1999;23(2):166-175.
6. Pippucci T, Savoia A, Perrotta S, et al. Mutations in the 5' UTR of ANKRD26, the ankirin repeat domain 26 gene, cause an autosomal-dominant form of inherited thrombocytopenia, THC2. Am J Hum Genet. 2011;88(1):115-120.
7. Zhang MY, Churpek JE, Keel SB, et al. Germline ETV6 mutations in familial thrombocytopenia and hematologic malignancy. Nat Genet. 2015;47(2):180-185.
8. Polprasert C, Schulze I, Sekeres MA, et al. Inherited and Somatic Defects in DDX41 in Myeloid Neoplasms. Cancer Cell. 2015;27(5):658-670.
9. Wlodarski MW, Hirabayashi S, Pastor V, et al. Prevalence, clinical characteristics, and prognosis of GATA2-related myelodysplastic syndromes in children and adolescents. Blood. 2016;127(11):1387-1397.
10. Pastor V, Hirabayashi S, Karow A, et al. Mutational landscape in children with myelodysplastic syndromes is distinct from adults: specific somatic drivers and novel germline variants. Leukemia. 2017;31(3):759-762.
11. Churpek JE, Pyrtel K, Kanchi KL, et al. Genomic analysis of germ line and somatic variants in familial myelodysplasia/acute myeloid leukemia. Blood. 2015;126(22):2484-2490.
12. Gohring G, Michalova K, Beverloo HB, et al. Complex karyotype newly defined: the strongest prognostic factor in advanced childhood myelodysplastic syndrome. Blood. 2010;116(19):3766-3769.
13. Kardos G, Baumann I, Passmore SJ, et al. Refractory anemia in childhood: a retrospective analysis of 67 patients with particular reference to monosomy 7. Blood. 2003;102(6):1997-2003.
14. Mantadakis E, Shannon KM, Singer DA, et al. Transient monosomy 7 - A case series in children and review of the literature. Cancer. 1999;85(12):2655-2661.
15. Parker TM, Klaassen RJ, Johnston DL. Spontaneous remission of myelodysplastic syndrome with monosomy 7 in a young boy. Cancer Genet Cytogenet. 2008;182(2):122-125.
16. Leung EW, Woodman RC, Roland B, Abdelhaleem M, Freedman MH, Dror Y. Transient myelodysplastic syndrome associated with isochromosome 7q abnormality. Pediatr Hematol Oncol. 2003;20(7):539-545.
17. Asou H, Matsui H, Ozaki Y, et al. Identification of a common microdeletion cluster in 7q21.3 subband among patients with myeloid leukemia and myelodysplastic syndrome. Biochem Biophys Res Commun. 2009;383(2):245-251.
18. Nagamachi A, Matsui H, Asou H, et al. Haploinsufficiency of SAMD9L, an endosome fusion facilitator, causes myeloid malignancies in mice mimicking human diseases with monosomy 7. Cancer Cell. 2013;24(3):305-317.
19. Chen DH, Below JE, Shimamura A, et al. Ataxia-Pancytopenia Syndrome Is Caused by Missense Mutations in SAMD9L. Am J Hum Genet. 2016;98(6):1146-1158.

20. Tesi B, Davidsson J, Voss M, et al. Gain-of-function SAMD9L mutations cause a syndrome of cytopenia, immunodeficiency, MDS and neurological symptoms. Blood. 2017;129(16):2266-2279.
21. Narumi S, Amano N, Ishii T, et al. SAMD9 mutations cause a novel multisystem disorder, MIRAGE syndrome, and are associated with loss of chromosome 7. Nat Genet. 2016;48(7):792-797.
22. Buonocore F, Kuhnen P, Suntharalingham JP, et al. Somatic mutations and progressive monosomy modify SAMD9-related phenotypes in humans. J Clin Invest. 2017;127(5):1700-1713.
23. Schwartz JR, Wang S, Ma J, et al. Germline SAMD9 mutation in siblings with monosomy 7 and myelodysplastic syndrome. Leukemia. 2017;31(8):1827-1830.
24. Baumann I, Niemeyer CM BJ, Shannon K. WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues. Lyon: IARC Press; 2008, 2008:104-107.
25. Vardiman JW, Harris NL, Brunning RD. The World Health Organization (WHO) classification of the myeloid neoplasms. Blood. 2002;100(7):2292-2302.
26. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. 2014;46(3):310-315.
27. Grantham R. Amino acid difference formula to help explain protein evolution. Science. 1974;185(4154):862-864.
28. Wlodarski MW, O'Keefe C, Howe EC, et al. Pathologic clonal cytotoxic T-cell responses: nonrandom nature of the T-cell-receptor restriction in large granular lymphocyte leukemia. Blood. 2005;106(8):2769-2780.
29. Vraetz T, Emanuel PD, Niemeyer CM. In vitro regulation of colony stimulating factor-mediated hematopoiesis in healthy individuals and patients with different types of myeloproliferative disease. Methods Mol Biol. 2003;215:293-309.
30. Lemos de Matos A, Liu J, McFadden G, Esteves PJ. Evolution and divergence of the mammalian SAMD9/SAMD9L gene family. BMC Evol Biol. 2013;13:121.
31. Chefetz I, Ben Amitai D, Browning S, et al. Normophosphatemic familial tumoral calcinosis is caused by deleterious mutations in SAMD9, encoding a TNF-alpha responsive protein. J Invest Dermatol. 2008;128(6):1423-1429.
32. Tanaka M, Shimbo T, Kikuchi Y, Matsuda M, Kaneda Y. Sterile alpha motif containing domain 9 is involved in death signaling of malignant glioma treated with inactivated Sendai virus particle (HVJ-E) or type I interferon. Int J Cancer. 2010;126(8):1982-1991.
33. Hershkovitz D, Gross Y, Nahum S, et al. Functional characterization of SAMD9, a protein deficient in normophosphatemic familial tumoral calcinosis. J Invest Dermatol. 2011;131(3):662-669.
34. 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. J Virol. 2011;85(7):3270-3282.
35. Jongmans MCJ, Verwiel ETP, Heijdra Y, et al. Revertant somatic mosaicism by mitotic recombination in dyskeratosis congenita. Am J Hum Genet. 2012;90(3):426-433.
36. Waisfisz Q, Morgan NV, Savino M, et al. Spontaneous functional correction of homozygous fanconi anaemia alleles reveals novel mechanistic basis for reverse mosaicism. Nat Genet. 1999;22(4):379-383.
37. Reina-Castillon J, Pujol R, Lopez-Sanchez M, et al. Detectable clonal mosaicism in blood as a biomarker of cancer risk in Fanconi anemia. Blood Advances. 2017;1(5):319-329.

38. Nagasawa M, Tomizawa D, Tsuji Y, et al. Pancytopenia presenting with monosomy 7 which disappeared after immunosuppressive therapy. Leuk Res. 2004;28(3):315-319.

39. Benaim E, Hvizdala EV, Papenhausen P, Moscinski LC. Spontaneous remission in monosomy 7 myelodysplastic syndrome. Br J Haematol. 1995;89(4):947-948.

40. Scheurlen W, Borkhardt A, Ritterbach J, Huppertz HI. Spontaneous hematological remission in a boy with myelodysplastic syndrome and monosomy 7. Leukemia. 1994;8(8):1435-1438.

Table 1. Clinical data of SAMD9L-mutated patients.

Patient # (UPN); sex	Gestational age; measurements	Dysmorphic feat., neurol. symptoms	Prior medical problems	Timepoint; MDS subtype	PB findings (Plt, WBC, ANC: x10^9/L; Hb: g/dL)	BM findings			Cytogenetics	
						Cellularity	Dysplasia	Blasts (%)	Metaphases	FISH chr. 7
P1 (D084); male	39w: 3550g (P>50-75), 50cm (P20)	none	recurrent RTI and endogenous eczema since infancy, transient pancytopenia at 6 mo, Plt↓, ANC↓ at 2 yrs; self-limiting seizures at 3 yrs	3.4yrs; RCC	Plt↓ (41), WBC↓/ANC↓ (4.3/0.34), MCV↑	↓	+++ vacuolization in E+M	<5	45,XY,-7 [6] / 46,XY [10]	60%
				7yrs; progress: CMML	Plt↓, WBC↑ (mono: 21%, blasts: 10%, erythro-blasts: 26%)	N	+++	5	45,XY,-7 [5]	95%
P2 (D154); female	34w: 2670g (P75-90), 49cm (P90)	none	recurrent RTI since age 1.5 yrs	2.0yrs; RCC	Plt↓ (96), MCV↑	↓	+++ vacuolization in M	<5	45,XX,-7‡	77%
				5.7yrs	normal, MCV↑	n.p.	n.p.	n.p.	n.p	n.p
				12yrs	normal, MCV↑	N	+	<5	46,XX	n.p
				13yrs	normal, MCV↑	N	N	<5	n.p.	normal
				14, 15, 17, 18yrs	normal, MCV↑	N	N	<5	46,XX	normal
				22yrs	normal, MCV↑	n.p.	n.p.	n.p.	n.p.	n.p.
P3 (US1); female	40w: 4080g (P90-97), 53cm (P50)	none	endogenous eczema	20mo; RCC	Plt↓ (88), ANC↓ (0.54), MCV↑	↓	+	<5	45,XX,-7 [3] / 46,XX [18]	16%
P4 (US2); male	41w: 3540g (P25-50), 52cm (P40)	none	recurrent RTI and endogenous eczema since infancy	12mo; RCC	Plt↓ (5), ANC↓ (0.43), Hb↓ (8.3), HbF↑ (5.2%)	↓	++	<5	46,XY [20]	normal
				13mo	ANC↓ (0.88), Hb↓ (9.8)	N	N	7	46,XY [20]	normal
				15mo	normal	N	N	<5	46,XY [20]	5.5%
				17mo	normal, HbF↑ (11.3%)	N	N	<5	n.p.	15%
				18.5mo	ANC↓ (0.6)	N	N	<5	45,XY,-7 [6] / 46,XY,del(7)(q11.2q36) [4] / 46,XY [10]	19%
P5 (D637); male	40w: 3875g (P75), 54cm (P75)	none	none	7.7yrs; RCC	Plt↓ (64), WBC↓/ANC↓ (1.7/0.08), Hb↓ (9.8), MCV↑, HbF↑ (5.7%)	↓	++	<5	45,XY,-7, der(18;21)(q10;q10),+21 [20/20]	n.p.
P6 (D637f); male	term: normal	none	none	42yrs; RCMD	Plt↓ (72), WBC↓/ANC↓ (2.0/1.4), Hb↓ (5.8), MCV↑	↓	++	<5	45,XY,der(1;7)(q10;p10)[11]/ 46,XY[5]	n.p.
P7 (SC054); female	term: normal	none	pancytopenia and hypocellular marrow at 1.7yrs	2.1yrs; RCC	Plt↓ (74), WBC↓/ANC↓ (4.0/0.8), Hb↓ (10.2)	N	+	<5	45,XX,-7 [4] / 46,XX [17]	-7 confirmed
				2.3yrs	Plt↓ (90), Hb↓ (10.5)	↑	+	<5	46,XX [25]	normal
				5, 6, 7.5, 11, 12, 18yrs	normal, HbF↑ (1.2-2.8%)	N	N	<5	normal	normal

UPN, unique patient number; syndr., syndromic; w, gestational week; RTI, respiratory tract infection (including otitis, bronchitis, pneumonia); yrs.; years of age; mo, months of age; Dx, diagnosis; RCC, refractory cytopenia of childhood; CMML, chronic myelomonocytic leukemia; BM, bone marrow; E, erythropoiesis, M, myelopoiesis; n.p., not performed; PB, peripheral blood; Plt, platelets; MCV, mean corpuscular volume (according to age); WBC, white blood count; ANC, absolute neutrophil count; Dysplasia: +, mild; ++, moderate; +++, severe. N, normal;

*with occasional small dysplastic megakaryocytes. ‡ 51% of metaphases with monosomy 7, of those 17% additionally showed hypoploid metaphases with involvement of chromosome 9, 14, 19, 21.

Table 2. Overview of germline and somatic mutations.

Mut type		Patient # (UPN)	Time point, age	-7% (FISH / Metaphases)	Genotype			Myeloid Sample	VAF% (depth)	Germline source	Conservation / PhysChem diff.	Effect (SIFT / MutTaster / Polyphen 2 / PredictSNP)	CADD-Score	ExAC browser n°./ % in population
GERMLINE	Family I	P1 (D084)	7yrs (CMML)	60% / [6/16]	***SAMD9L***^{m}	c.4534G>A	p.V1512M	BM-GR	^{WES}8.3% (277), ^{DS}13.3%(2755)	-	High / Small	D/P/D/D(65%)	25.2	none
					PTEN^{p}	c.563A>G	p.Y188C	BM-GR	^{WES}57.5% (80)	-	High / Large	D/-/D/D(61%)	25.8	none
		P2 (D154)	17yrs	0% / -	***SAMD9L***^{m}	c.4534G>A	p.V1512M	BM-GR	^{WES}21.5% (424), ^{DS}19.3% (1261)	HR (Sanger)	High / Small	D/P/D/D(65%)	25.2	none
					PTEN^{p}	c.563A>G	p.Y188C	BM-GR	^{WES}59.7% (62)	HR (Sanger)	High / Large	D/-/D/D(61%)	25.8	none
		Father		-	*PTEN*	c.563A>G	p.Y188C	PB	Sanger	HR (Sanger)	High / Large	D/-/D/D(61%)	25.8	none
		Mother		-	***SAMD9L***	c.4534G>A	p.V1512M	PB	^{DS}49.0% (13116)	HR (Sanger)	High / Small	D/P/D/D(65%)	25.2	none
	Family II	P3 (US1)	20mo	16% / [3/21]	***SAMD9L***^{p}	c.2957G>A	p.R986H	PB	Sanger	HR (Sanger)	High / Small	D/D/D/D(87%)	26.5	4/0.003%
					JAK3^{p}	c.2630C>T	p.A877V	PB	Sanger	HR (Sanger)	Weak / Small	T/N/B/N(74%)	23.3	11/0.01%
		P4 (US2)	15mo	5.5% / [0/20]	***SAMD9L***^{p}	c.2957G>A	p.R986H	PB	^{DS}43.0% (252)	HR (Sanger)	High / Small	D/D/D/D(87%)	26.5	4/0.003%
					JAK3^{p}	c.2630C>T	p.A877V	PB	^{DS}48.1% (n.a.)	HR (Sanger)	Weak / Small	T/N/B/N(74%)	23.3	11/0.01%
					FANCM^{m}	c.171G>C	p.L57F	PB	^{DS}49.0% (401)	HR (Sanger)	Weak / Small	D/D/B/D(52%)	17.9	195/0.32%
		Father		-	***SAMD9L***	c.2957G>A	p.R986H	PB	Sanger	HR (Sanger)	High / Small	D/D/D/D(87%)	26.5	4/0.003%
					JAK3	c.2630C>T	p.A877V	PB	Sanger	HR (Sanger)	Weak / Small	T/N/B/N(74%)	23.3	11/0.01%
		Mother		-	*FANCM*	c.171G>C	p.L57F	PB	Sanger	HR (Sanger)	Weak / Small	D/D/B/D(52%)	17.9	195/0.32%
	Family III	P5 (D637)	7.7yrs	- / [20/20]	***SAMD9L***^{p}	c.2956C>T	p.R986C	BM-GR	^{DS}7.5% (3422)	FB (Sanger)	High / Large	D/N/B/D(87%)	21.7	none
		P6 (father)	42yrs	- / [11/16]	***SAMD9L***	c.2956C>T	p.R986C	PB	Sanger	-	High / Large	D/N/B/D(87%)	21.7	none
		Brother		-	***SAMD9L***^{p}	c.2956C>T	p.R986C	PB	Sanger	-	High / Large	D/N/B/D(87%)	21.7	none
	Family IV	P7 (SC054)	2.3yrs	- / [0/25]	***SAMD9L***	c.2640C>A	p.H880Q	BM-GR	^{DS}27% (8139)	-	Weak / Small	T/-/D/-(-)	23.7	none
ACQUIRED		P1 (D084)	7yrs (CMML)	60% / [6/16]	***SAMD9L***	c.3562C>T	p.R1188X	BM-GR	^{WES}5.9% (202) ^{DS}8.0% (884)	-		novel, stop-gain	35	-
					KRAS	c.182A>C	p.Q61P	BM-GR	^{WES}37.7% (106)	-		known driver mut.	28.2	-
					SETBP1	c.2602G>A	p.D868N	BM-GR	^{WES}47.8% (355)	-		known driver mut.	26.7	-
					EZH2	c.1744G>A	p.V582M	BM-GR	^{WES}69.2% (130)	-		known driver mut.	34	-
		P5 (D637)	7.7yrs	- / [20/20]	*RUNX1*	c.413_427+5dup20bp		BM-GR	Sanger	FB (Sanger)		novel, splice donor	n.a.	-
		P7 (SC054)	2.3yrs	- / [0/25]	***SAMD9L***	c.3951_3955delTAAAG*	p.S1317RfsX21	BM-GR	^{DS}10% (5934)	-		novel, stop-gain	28.8	-

Abbreviations: Mut, mutation; UPN, unique patient number; BM, bone marrow; PB, peripheral blood; GR, granulocytes; HR, hair follicles; FB, skin fibroblast; m., maternal origin; p., paternal origin; VAF, variant allelic frequency; WES, whole exome sequencing; DS, targeted deep sequencing; Sanger, identified by Sanger sequencing; n.a., not available; +yrs/mo, years/months after diagnosis. Evolutionary conservation scores, Phylop and PhastCons; PhysChem diff., physicochemical difference between amino acids. In-silico prediction: SIFT: T-tolerated, D-deleterious; Mutation Taster: D-disease causing, N-polymorphism, P-polymorphism automatic; PolyPhen2: D-probably damaging, B- benign; PredictSNP consensus classifier: D-deleterious, N-neutral (% accuracy). Combined annotation-dependent depletion (CADD-score) of 20 means that a variant is amongst the top 1% of deleterious variants in the human genome; CADD-20=1%, CADD-30=0.1%, CADD-40=0.01%, CADD-50=0.001%. * mutation classified as acquired based on low allelic frequency

Gene annotations: SAMD9L (NM_001303500.1), EHZ2 (NM_152998), SETBP1 (NM_015559), KRAS (NM_004985.4), FANCM (NM_001308134), JAK3 (NM_000215), PTEN (NM_000314.4), RUNX1 (001001890).

FIGURE LEGENDS

Figure 1. Germline *SAMD9L* mutations in pedigrees with familial MDS.

(A) Identification of 4 pedigrees with MDS and monosomy 7 harboring germline heterozygous *SAMD9L* mutations: p.V1512M (pedigree I), p.R986H (pedigree II) and p.R986C (pedigree III), p. p.H880Q (pedigree IV), and somatic mutations: p.R1188X (P1) and p.S1317RfsX21 (P7). Dotted symbols indicate healthy mutation carriers. Sanger sequencing on DNA extracted from hair follicles (HR) confirmed germline status of mutations as visualized in electropherograms. Sequencing in P1 was performed on peripheral blood granulocytes (GR) revealing a minor mutational peak, corresponding to a variant allelic frequency of 8.3% by WES. In pedigree III, the mutation in P5 was confirmed in fibroblast (FB) DNA, while for P6 and remaining family members whole blood (WB) was analyzed. In pedigree IV other family members were not tested (n.t.). (B) *SAMD9L* and *SAMD9* gene orientation on 7q22 in reverse strand direction (3'-5'). SAMD9L protein is coded by one exon and contains two known functional sites: N-terminal sterile alpha motif (SAM) and nuclear localization sequence (NLS). Four germline and two somatic (*) mutations were identified in *SAMD9L*. Germline missense mutations are evolutionarily highly conserved. (C) TA cloning of double mutated *SAMD9L* region of P1 revealed *cis*-configuration of mutations p.V1512M (germline) and p.R1188X (somatic) in 10 out of 172 clones tested.

Figure 2. Bone marrow findings in P1, P2 at different timepoint of disease course.

H&E staining of bone marrow (BM) at diagnosis of RCC in P1 showing dysplastic granulopoiesis with hypergranulation and a pseudo Pelger cell (top left), myeloblast and dysplastic eosinophil (top right). BM at diagnosis in P2 (synchronous with monosomy 7) showing hypergranulation and vacuolization in myelocytes, and dysplastic erythropoiesis with double nuclei (bottom left). Normal BM morphology in P2, 15 years after initial BM confirming spontaneous phenotype reversion (bottom right).

Figure 3. Mechanisms of clonal escape from *SAMD9L* germline mutations.

Multiple mechanisms of clonal escape from damaging germline missense *SAMD9L* mutations are observed and lead to complete (monosomy 7) or partial (deletion 7q) loss of chromosome 7 with decreasing mutant *SAMD9L* allele (red circles), both situations can lead to MDS development; UPD7q and truncating somatic *SAMD9L* mutations (green circles), which show benign outcome and contribute to normal hematopoiesis. Multiple clonal outcomes can occur in single patient.

Figure 4. Loss of mutated *SAMD9L* allele due to genomic deletion or mitotic recombination.

Variant allelic frequency (VAF) scores for chromosome 7 in P1 and P2. SNPs and Indels detected using whole exome sequencing (~4000 variants with a VAF score > 5% and < 95%), depict a complete loss of chromosome 7 in P1, as the VAF scores are either low or high. P2, unlike P1 demonstrates a partial loss of the Chr 7 after position 7q11.22 towards the q terminal site. The read depth of the SNPs for P2 was maintained throughout for Chr7 with no loss thus confirming the LOH is due to UPD and not -7q. Whole exome sequencing VAF values are marked by a star within the graph. Abbreviations: VAF, variant allelic frequency; UPD, uniparental isodisomy. Blue line: centromere; red line: *SAMD9L* gene position; yellow dotted line: start of UPD.
For P7, targeted NGS identified 14 informative (heterozygous) polymorphisms located on chromosome 7q with an average depth of 1036 reads (supplementary table 1). SNPs are represented in a VAF graph depicting the skewing of heterozytosity towards one allelle ocurring after position g.66098482 (rs3764903).

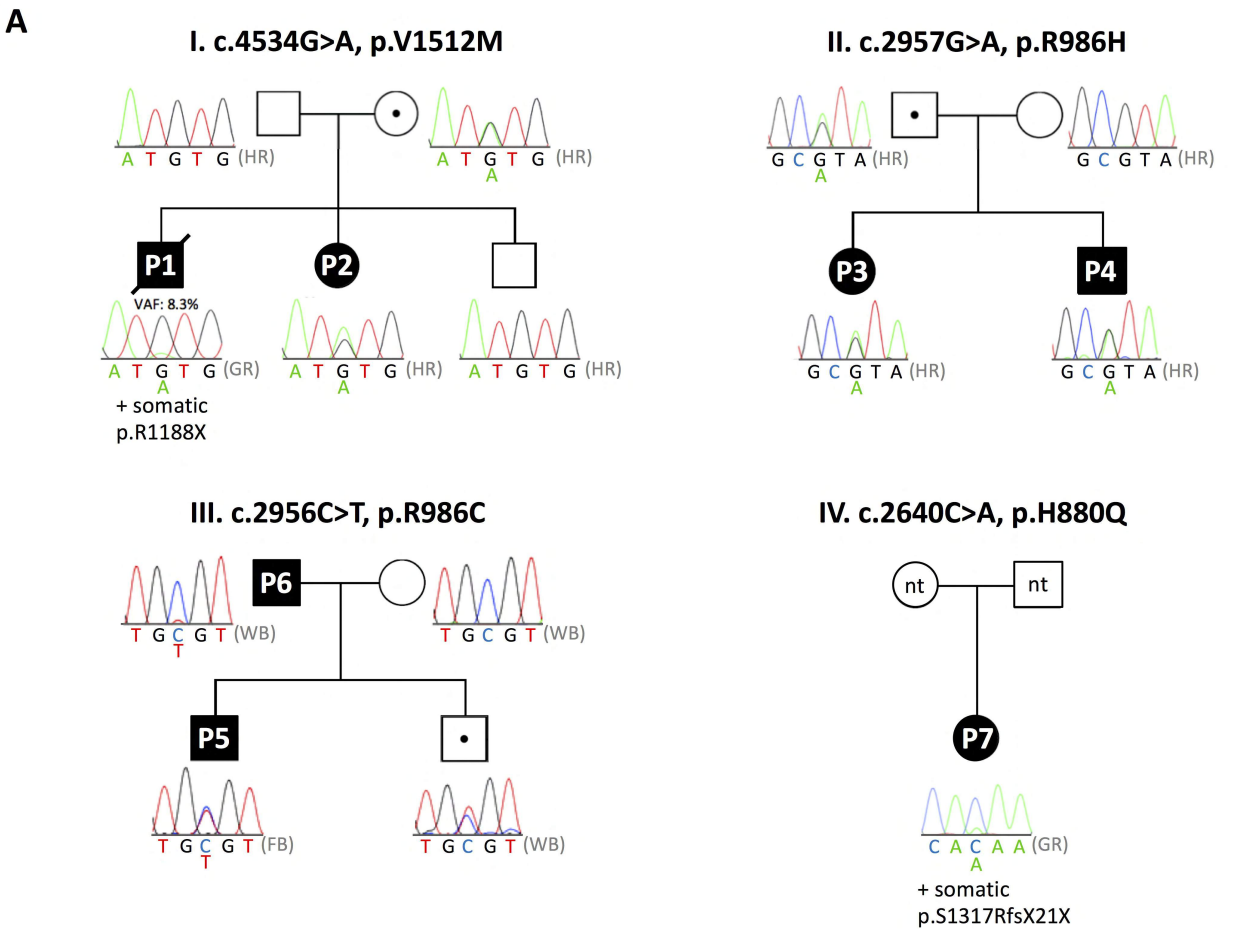
Figure 5. Clonal evolution and spontaneous reversion due to UPD7q.

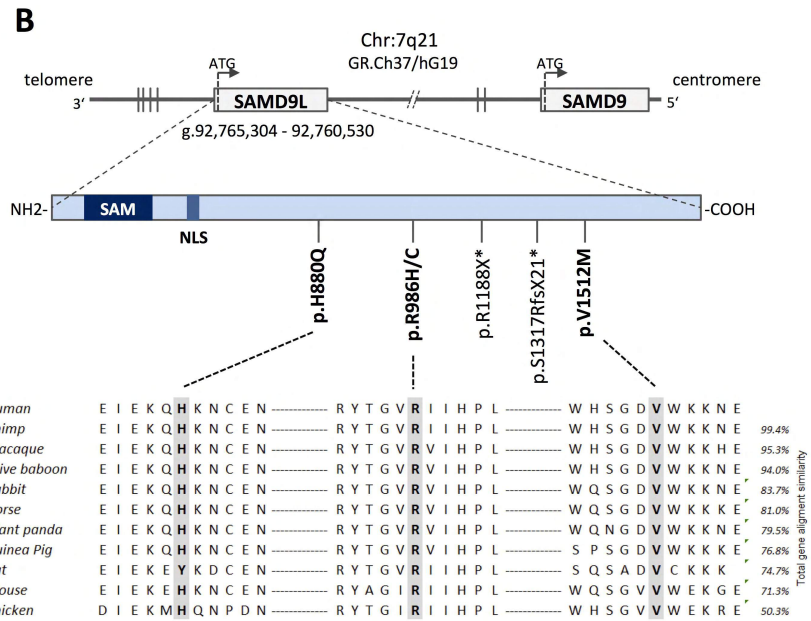
Clonal evolution model in P2 (D154) depicting disease history during an observation period of 20 years. At diagnosis, initial bone marrow harbored monosomy 7 (77% by FISH and 51% by metaphase karyotyping). Blood counts normalized 3.7 years later and since then P2 maintained normal CBC until last follow-up at age of 22 years. From age of 12 years, repetitive yearly BM examinations revealed normocellular hematopoiesis with no dysplasia and normal cytogenetics. BM collected at age of 17 years (*) was subjected to whole exome and targeted deep sequencing. Germline heterozygous *SAMD9L* mutation p.V1512M was detected at a variant allelic frequency (VAF) of ~20%, corresponding to a clonal size of ~40% in a diploid chromosome 7 background. Concurrently, a spontaneous genetic correction of the *SAMD9L* locus occurred resulting from uniparental isodisomy (UPD)7q of paternal origin. This self-corrected clone occurred either initially (dotted line) or after termination of monosomy 7 and contributed to normal hematopoiesis. Abbreviations: Dx, diagnosis; pat, paternal origin; mat, maternal origin; UPD; uniparental isodisomy, LFU; last follow-up.

Figure 6. Functional evaluation of *SAMD9L* mutations.

(A-B) The effect of *SAMD9L* mutations on cell proliferation was assessed by dye dilution assays. 293FT cells were transiently transfected with TFP-SAMD9L wild type (WT), the disease-associated mutations p.R986C and p.V1512M, and the protective variant p.T233N previously reported by Tesi et al(20). (A) Histograms depict the dye levels in transfected cells. Dye levels were monitored in TFP-transfected cells (filled grey histograms) and compared to cells expressing uniformly intermediate levels of TFP-SAMD9L wild type (blue histograms) or variants (red/orange lines), as indicated. A single representative experiment is shown. (B) Cumulative summary of three independent experiments on inhibition of cell proliferation associated with indicated TFP-*SAMD9L* mutations. Values (mean +/- SD) are calculated based on a scale defined by 0 (dye levels in TFP-transfected cells) and -1 (dye levels in cells transfected with TFP-SAMD9L wild type). Unpaired t-test, two tailed: * $p<0.05$; ** $p<0.005$; *** $p<0.001$.

Figure 1. Germline SAMD9L mutations in pedigrees with familial MDS





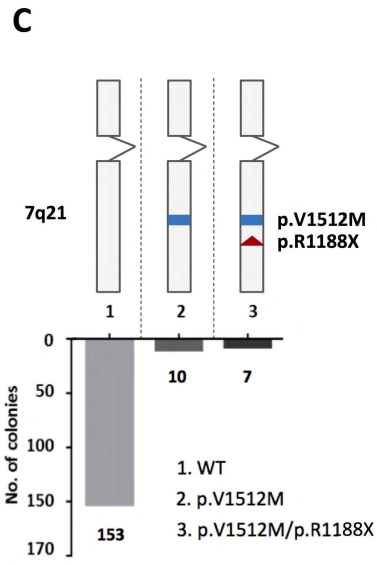


Figure 2. Bone marrow morphology in P1 and P2

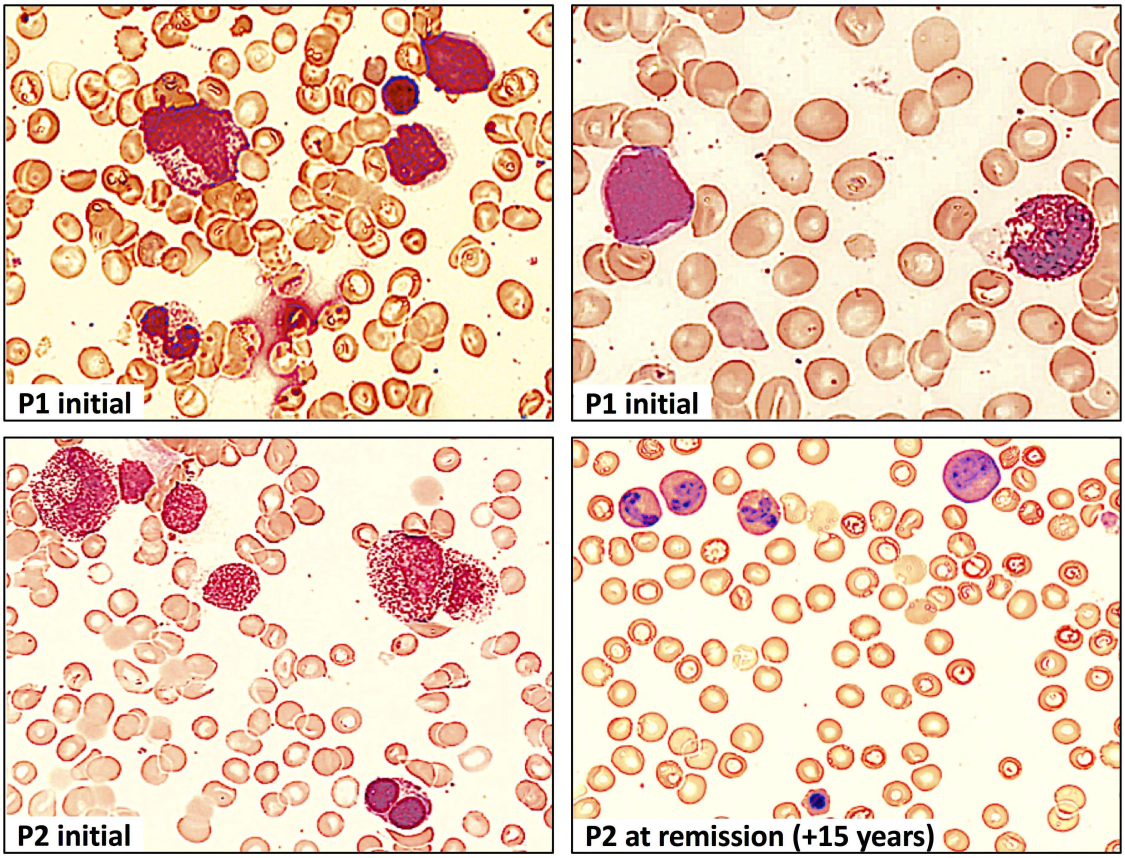


Figure 3. Mechanisms of clonal escape from SAMD9L germline mutations

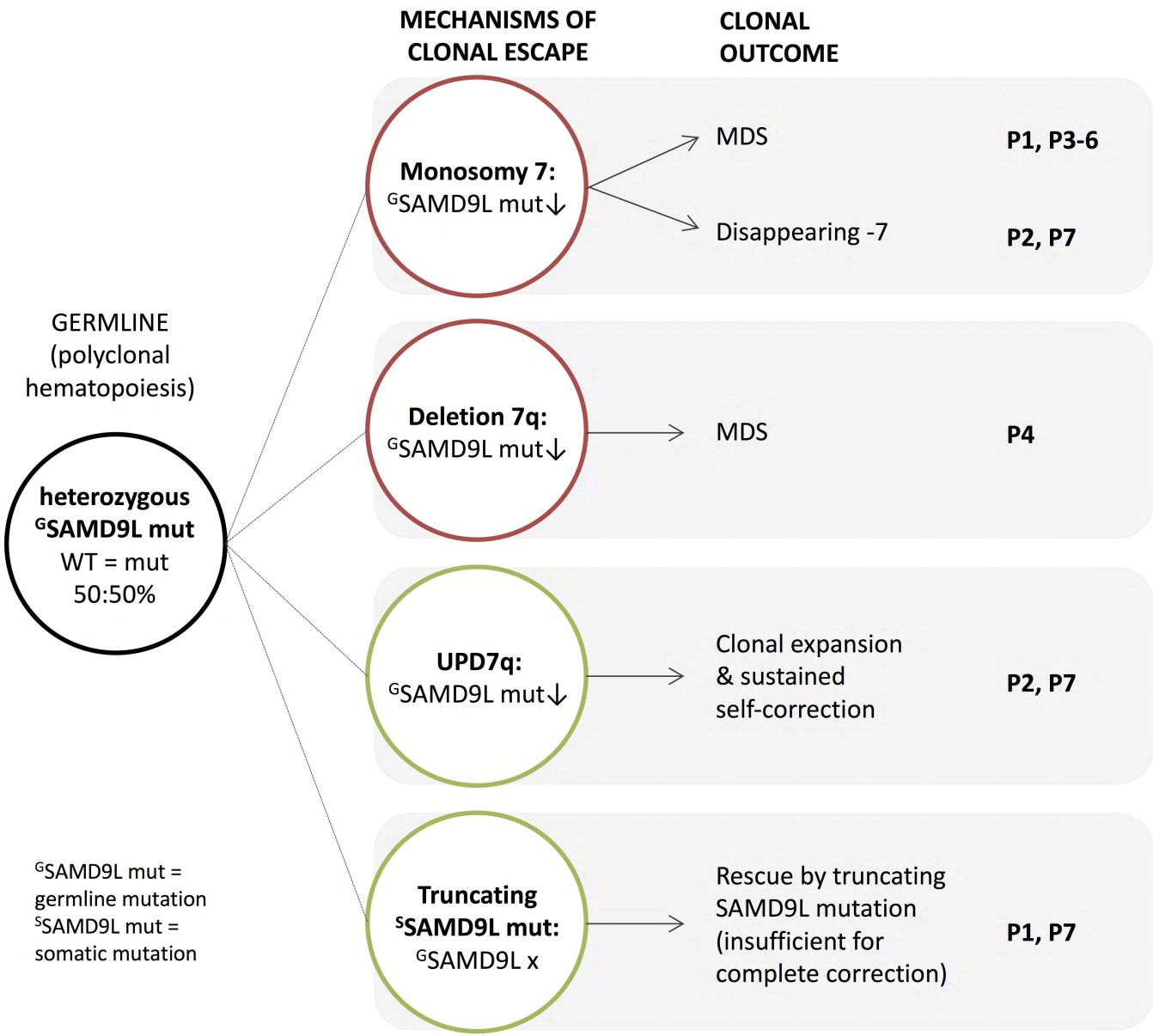


Figure 4. Loss of mutated SAMD9L allele due to genomic deletion or copy-number neutral recombination

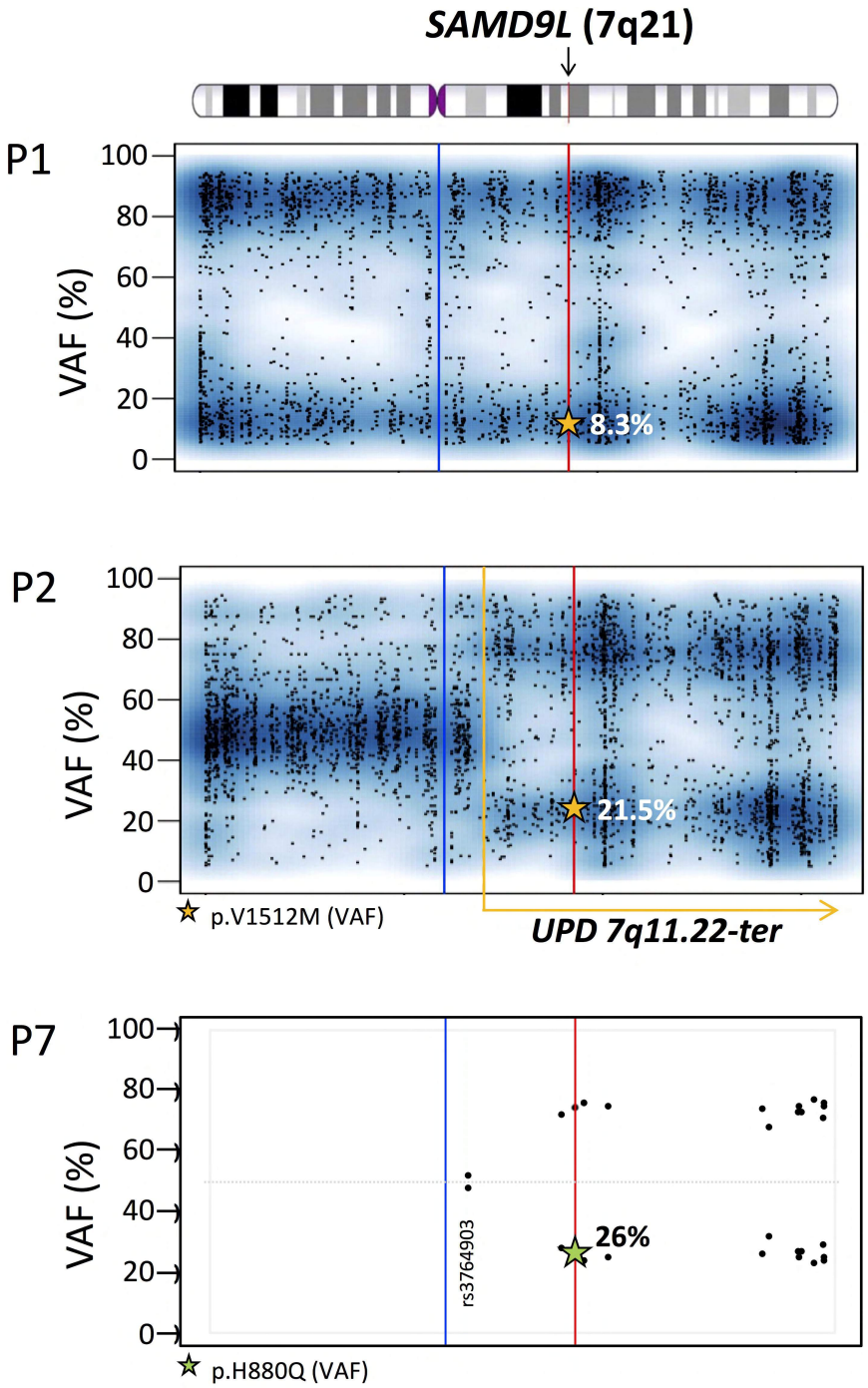


Figure 5. Clonal evolution and spontaneous reversion due to UPD7q

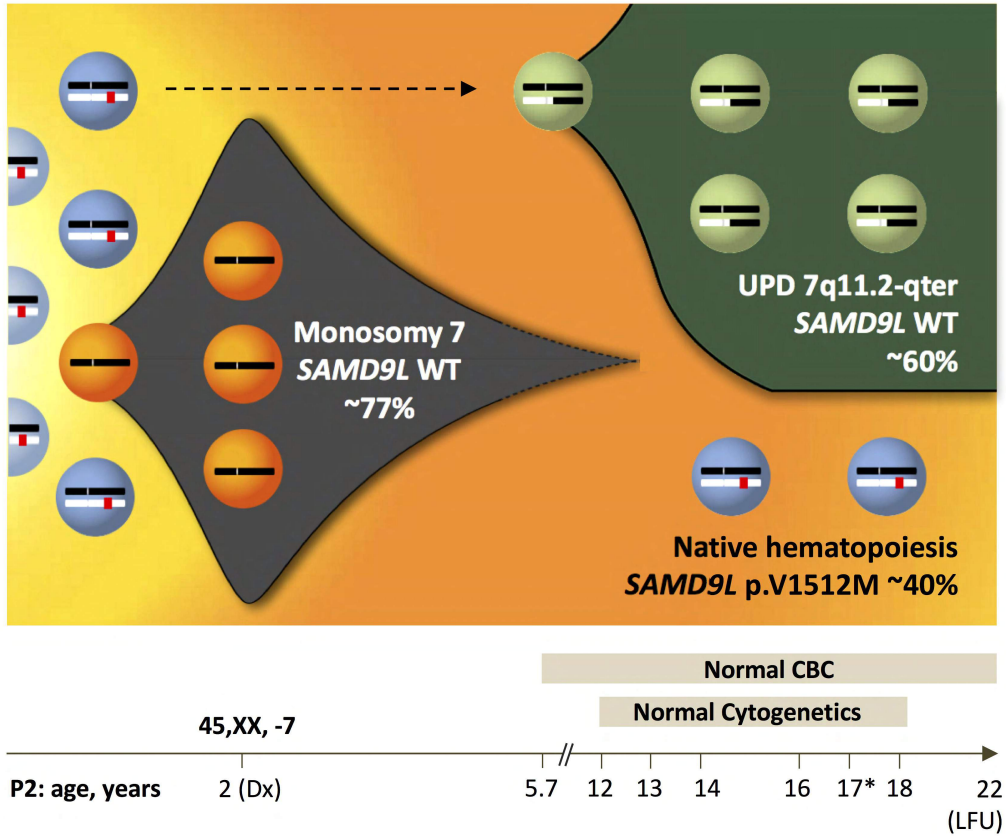
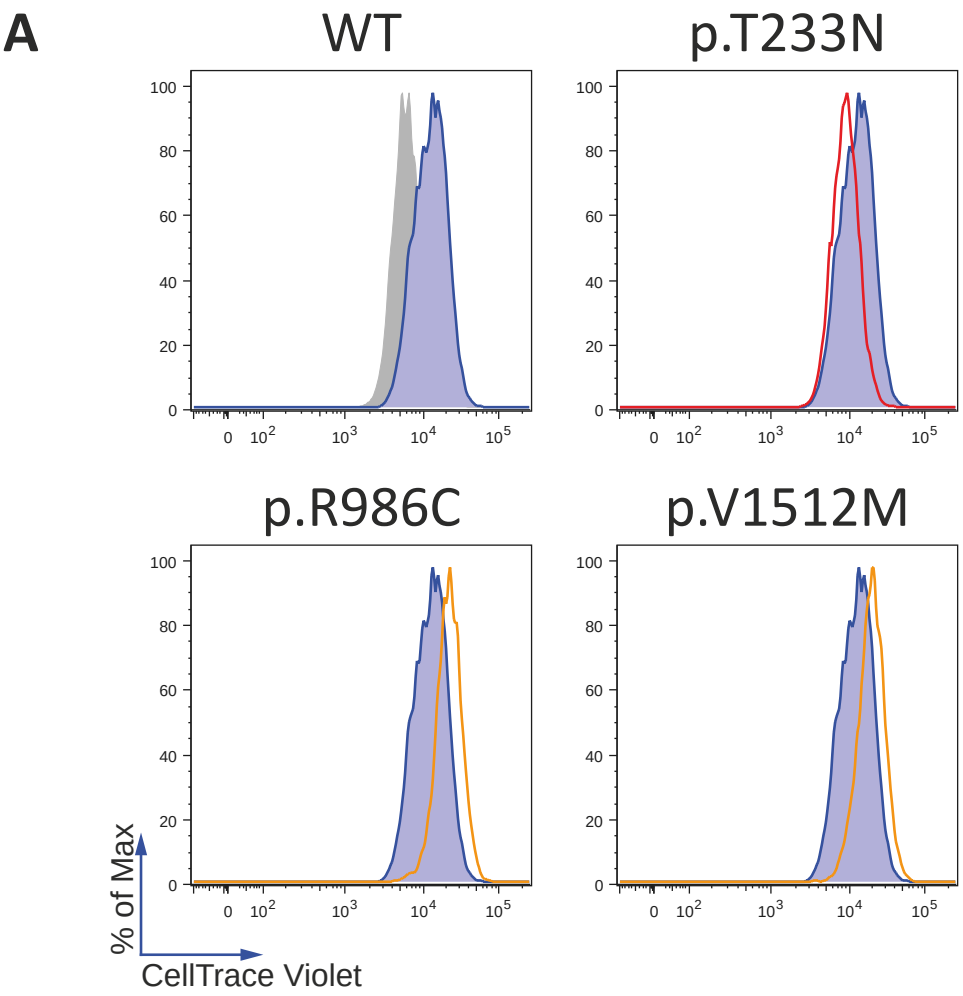


Figure 6. Functional evalution of SAMD9L mutations.



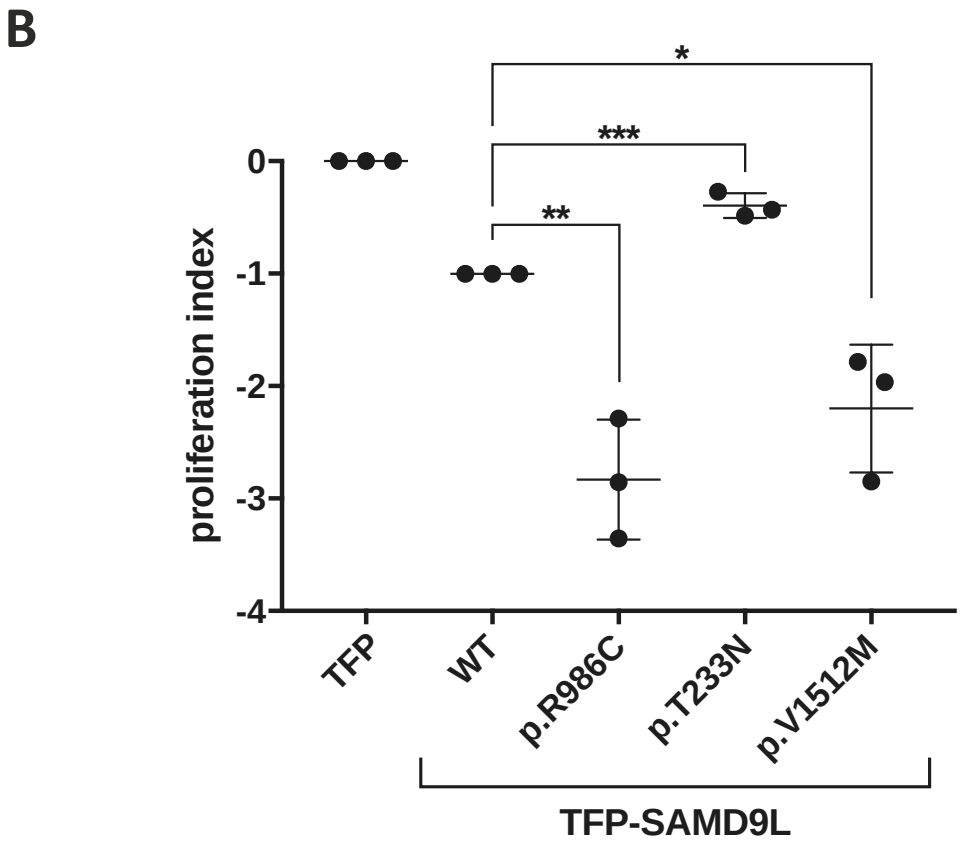
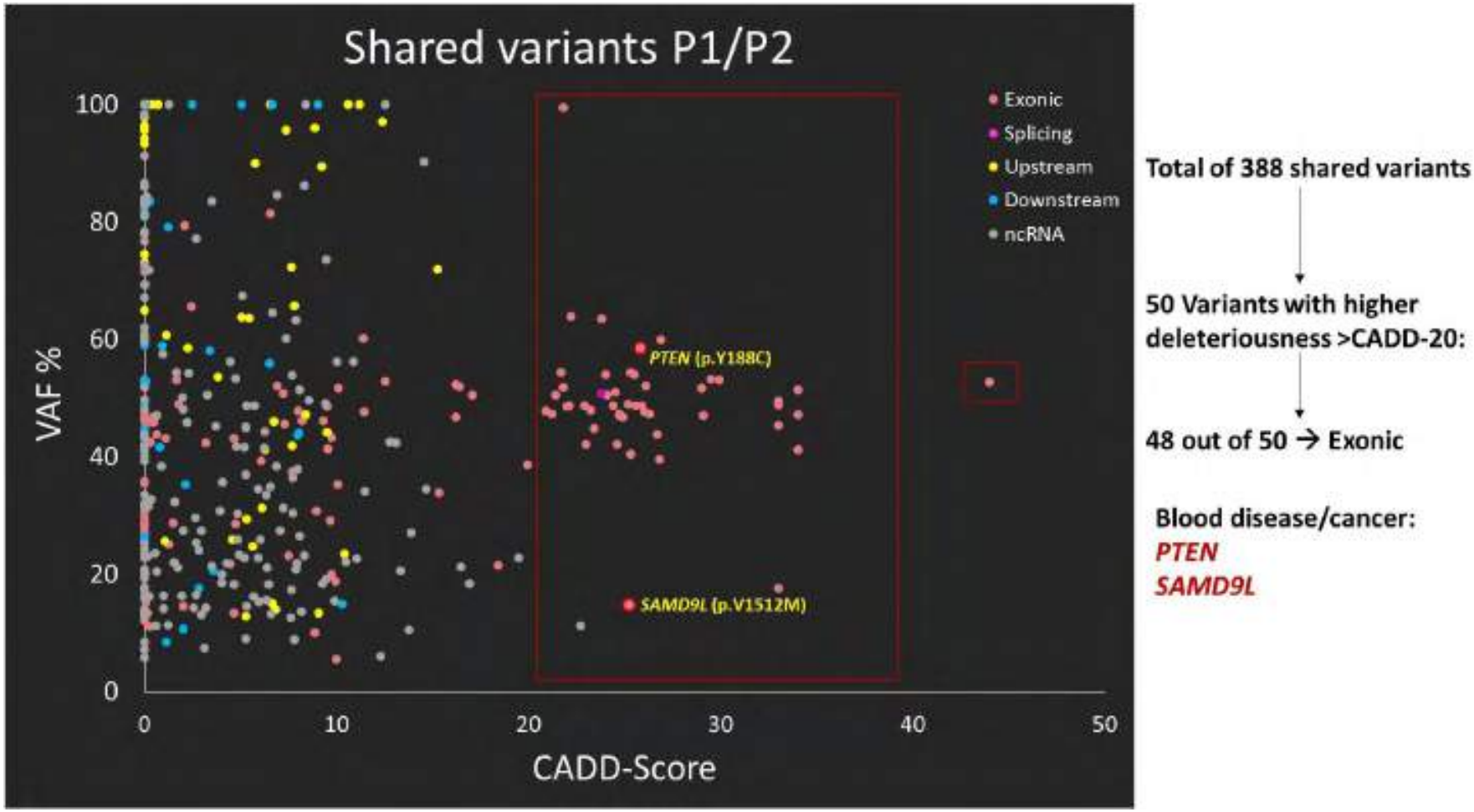
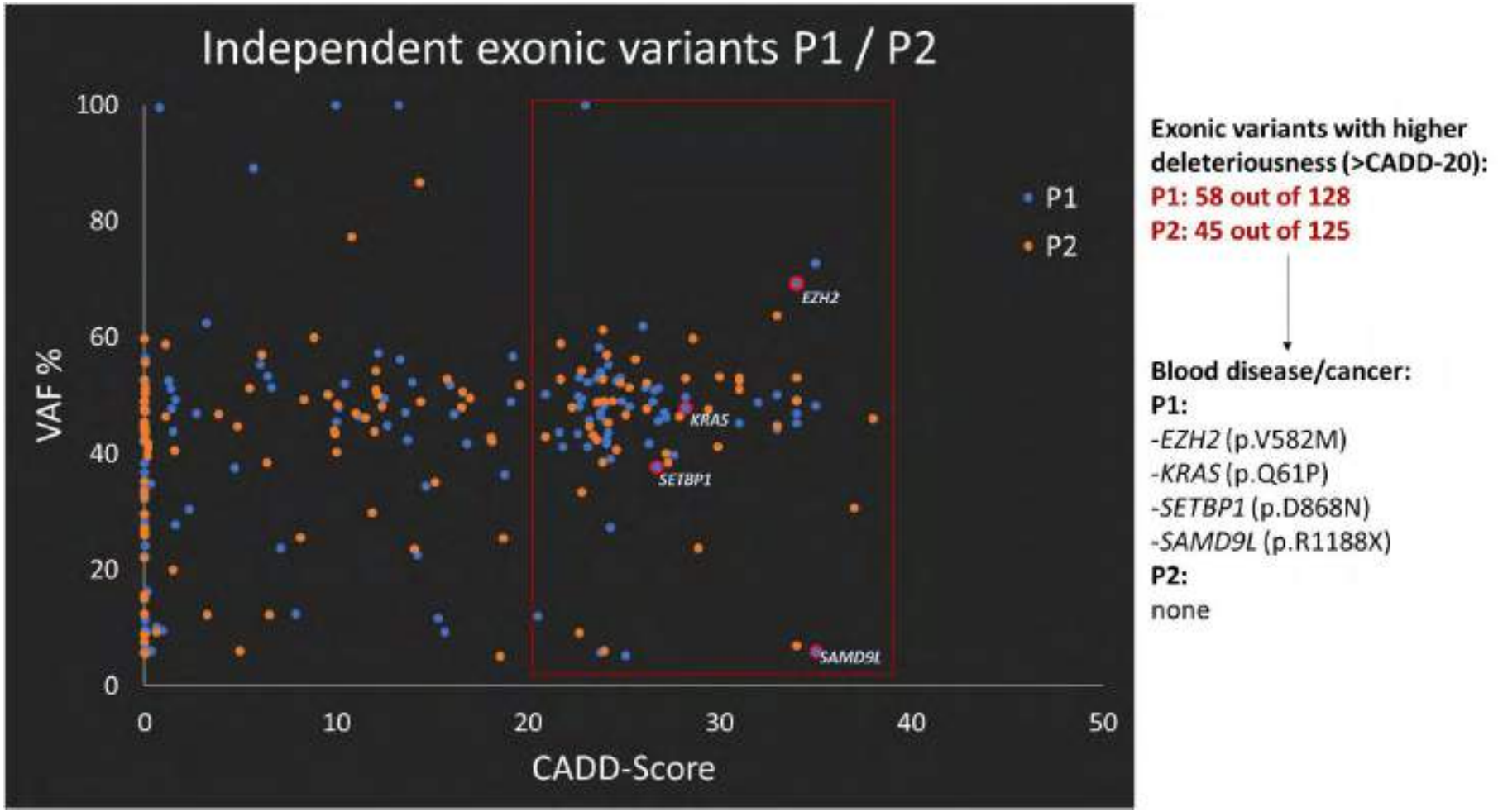


Figure S1. Shared variants found by WES in P1 and P2



Legend: The Y axis corresponds to the average VAF % of all detected shared mutations in both siblings and the X axis the calculated CADD-scores. The vast majority of mutations show a low degree of deleteriousness with a CADD-score lower than 20. Framed in red, 50 highly deleterios variants of which 48 were exonic. Among them, germline mutations in *SAMD9L* (p.V1512M) and *PTEN* (p.Y188C) were observed.

Figure S2. Individual mutations found in P1 by WES



Legend: Indivudial exonic variants in P1 and P2. Highly deleterios (CADD>20) variants were 58 and 45 variant in P1 and P2 respectively. Known leukemic driver genes were affected uniquely in P1 in addtion to a novel non-sense variant in the *SAMD9L* gene. In contrast, no significant mutations involved in blood disease or cancer were found in P2.

Table S1. Allelic frequencies of informative chromosome 7q polymorphisms in P7

Gene	Isoform	Genomic Position	SNP	cDNA		VAF (allele 1)		VAF (allele 2)		dbSNP:MAF
KCTD7	NM_001167961	66098482	rs3764903	c.314+51G>A	Het.	G	52%	A	48%	46%
GTPBP10	NM_033107	89984409	rs42664	c.329A>G	Het.	A	**28%**	G	72%	49%
DYNC1I1	NM_004411	95709602	rs42083	c.1702-73C>T	Het.	C	**24%**	T	76%	44%
DYNC1I1	NM_004411	95709666	rs42082	c.1702-9C>A	Het.	C	**24%**	A	76%	44%
CUX1	NM_001202544	102000000	rs11540899	c.1089G>A	Het.	G	**75%**	A	**25%**	46%
TAS2R4	NM_016944	141478574	rs2234001	c.286G>C	Het.	G	74%	C	**26%**	47%
CLCN1	NM_000083	143042837	rs2272251	c.2154C>T	Het.	C	68%	T	**32%**	48%
TMEM176B	NM_014020	150491084	rs3173833	c.280A>C	Het.	A	73%	C	**27%**	46%
SLC4A2	NM_001199694	150768786	rs2303937	c.2160G>A	Het.	G	**25%**	A	75%	45%
PRKAG2	NM_016203	151504499	rs10257529	c.115-20872C>T	Het.	C	73%	T	**27%**	42%
DPP6	NM_001936	154667643	rs2230064	c.1725G>A	Het.	G	77%	A	**23%**	18%
UBE3C	NM_014671	156971581	rs870745	c.616+40T>C	Het.	T	71%	C	**29%**	37%
DNAJB6	NM_005494	157162068	rs9692250	c.346+1891T>C	Het.	T	**24%**	C	76%	34%
DNAJB6	NM_005494	157162103	rs12672981	c.346+1926A>G	Het.	A	**25%**	G	75%	45%

Average read depth obtained for all 14 polymorphisms: 1036x.

CASE DESCRIPTIONS

In **family I**, the older brother P1 suffered from recurrent respiratory tract infections (RTI) and endogenous eczema manifesting in infancy. The first documented complete blood count (CBC) during an infectious episode at the age of 6 months showed pancytopenia (platelets $39x10^9$/L neutrophils $0.67x10^9$/L Hb 7.7g/dL). Anemia spontaneously improved but platelet and neutrophil counts remained low normal. The first bone marrow (BM) aspirate performed at the age of 2 years showed a reduced cell content and dysplasia with vacuolization in the myeloid and erythroid lineages and BM eosinophilia. Hemoglobin F, Vitamin B12, folate and immunoglobulin levels as well as lymphocyte subpopulations were within normal range and there were no signs of autoimmunity. The patient remained stable without transfusions. A second BM aspirate performed at the age of 3.4 years confirmed myelodysplasia (Figure 2); chromosomal studies showed monosomy 7 in 6 of 16 metaphases. P1 was diagnosed with refractory cytopenia of childhood (RCC). He suffered transient cerebral seizures without the need of anticonvulsant therapy, while a cranial CT scan was normal. At the age of 6.5 years, he presented with hemorrhagic varicella with secondary pneumonia and required platelet transfusions for a first time. At the age of 7 years, he had developed hepatosplenomegaly and showed 21% blasts on peripheral blood smear. There was no spontaneous growth in *in vitro* GM-CSF hypersensitivity assay AML induction chemotherapy was complicated by bacterial sepsis and respiratory failure. Following recovery, thioguanine maintenance therapy was administered. He received an HSCT from a matched unrelated donor (MUD). Hematological reconstitution was slightly delayed (WBC day +19, platelets day +63, RBC day +49) with stable complete chimerism. At day +138 late, grade III acute GvHD (gut stage 3, liver stage 2, liver stage 2) manifested, and he died 10 days later from acute CNS hemorrhage.

P2 is the younger sister of P1 and presented with recurrent respiratory tract infections at the age of 18 months. The first documented CBC at the age of 2 years revealed thrombocytopenia (Table 1). BM was hypocellular with severe dysplasia, vacuolization similar to what had been observed in P1 (Figure 2). Monosomy 7 was detected in 77% FISH-interphases and 51% metaphases with hypoploid metaphases in one subclone (Table 1). After the death of her brother, P2 did not visit the hematology clinic until two years later for tonsillectomy due to recurrent RTI. Her CBC revealed only mild thrombocytopenia ($85x10^9$/L) and leukopenia ($4.4x10^9$/L) without neutropenia. When she was 5.7 years old her CBC normalized and remained within the normal range since. A second BM examination at the age of 12 years revealed normal cellularity with slight dysplastic changes and normal cytogenetics. All subsequent marrow analyses in nearly yearly intervals were normal without any signs of MDS or monosomy 7 (Table 1, Figure 2). P2 did not suffer from infectious episodes and at last follow-up (FUP) at the age of 22 years she was in good general condition.

In **family II** the older sister (P3) had unremarkable postnatal development with the exception of chronic endogenous eczema, and petechial rash manifesting at the age of 1 year. No recurrent infections were observed. Her CBC performed during routine 18 months evaluation revealed thrombocytopenia, neutropenia and elevated MCV. Initial BM was hypocellular with mild erythroid dysplasia, and without increase of blasts. Cytogenetics uncovered monosomy 7 in 3/21 metaphases and 16% FISH-interphases. Six months later she was successfully transplanted from a MUD after myeloablative conditioning. At last FUP she was 6 years 4 months old and in good general condition.

P4 is the younger brother of P3. He suffered from recurrent RTI since infancy. His initial hematologic presentation during an episode of acute otitis media at the age of 12 months was pancytopenia with severely hypocellular BM and dysgranulopoiesis (Table 1). Evolving MDS-RCC or toxic stromal damage secondary to a viral insult was suspected. Viral workup was unremarkable. His CBC recovered and subsequent BM examinations demonstrated normocellular BM without significant dysplasia. However, 3 months after initial presentation, a small monosomy 7 clone was detected by FISH in 11/200 interphase-cells but was absent in metaphase cytogenetics. Finally, 3 months later, two independent clones emerged, with -7 and del7q in 6/20 and 4/20 metaphases, respectively (Table 1). At last FUP, P4 was 21 months old and undergoing MUD-HSCT.

In **family III**, the affected son (P5) and his father (P6) did not have any infectious or hematologic problems until they presented with pancytopenia and were diagnosed with hypocellular MDS at the age of 7.7 and 42 years, respectively. Complete or partial loss of chromosome 7 in addition to other concomitant lesions (Table 1) were found in BM. Both P5 and P6 were transplanted from an unrelated donor, or HLA-identical brother at the age of 8 and 43 years, respectively. They were alive at last FUP when they were 17.4 and 54 years old.

In **family IV**, P7 was the affected child of non-consanguineous parents with negative family history for hematooncologic or neurologic diseases. At the last FUP, P7´father had been evaluated by neurologists for unclear ataxia symptoms (his SAMD9L status is unknown). The girl presented at the age of 1 year and 8 months with pancytopenia and petechial rash on the trunk. Initial bone marrow showed a cellularity of 15%. Marrow analysis performed 5 months later revealed normal cellularity and dysplasias compatible with the diagnosis of MDS-RCC, while cytogenetics showed complete monosomy 7 in 4 out of 21 metaphases as confirmed by interphase FISH. Unexpectedly, cytogenetics normalized 3 months later, cellularity gradually increased, and there were only slight dysplastic traits observed in the myeloid lineage. At that time point NGS analysis in BM specimen identified a large UPD7q clone (double wildtype for SAMD9L). Subsequent marrow examinations and CBC analyses were normal and the patient remained healthy until last follow-up at the age of 18 years.

SUPPLEMENTARY METHODS

Genomic studies

Genomic DNA from P1 and P2 was subjected to exploratory whole exome sequencing (WES). After exon region capture using Agilent SureSelect v5.0 (Agilent, Santa Clara, CA) and library construction by Illumina TruSeq (Illumina, San Diego, CA), 100bp paired-end sequencing was performed on Illumina HiSeq2000 instrument. Following dynamic trimming, the resulting reads were aligned to the hg19 reference genome with BWA20 followed by base quality score recalibration, indel realignment, duplicate removal, SNP and INDEL discovery using standard filtering parameters, and variant quality score recalibration according to GATK[1, 2]. Confirmatory deep sequencing was performed for *SAMD9L* in P1 and P2 and P4 for respective mutations (Table 2). P1, P2, P5 and P7 were subjected to an in house 28-gene pediatric MDS panel (University of Freiburg, Germany). P4 was analyzed using a "57-genes myeloid malignancies" NGS panel (ARUP laboratories, Salt Lake City, USA) and a "86-genes bone marrow failure" NGS panel (Claritas Genomics, Cambridge, USA).

All mutations identified by NGS were validated using Sanger sequencing. For germline confirmation, DNA from skin fibroblast and/or hair follicles was extracted using automated extraction system (AS 1290 kit, MaxWell 16 System, Promega). The targets were amplified using Amplitaq Gold Polymerase (Lifetechnologies, USA) and sequenced as previously described[4].

Bioinformatics

The variant annotation was compiled using ANNOVAR[5] from the 1000 Genomes Project requiring a minimum of 8 reads and a variant allele frequency > 0.05%. SNPs reported at a minor allele frequency <0.1% in the ExAc Browser were considered rare. We applied a minimum quality score of 30 (Q30, 99.9% base call accuracy) to minimize the incorrect base calls (Q30=1/1000 incorrect base calls). A minimum coverage of 150X was employed for genetic variants with a variant allelic frequency (VAF) ≥ 5%. Those quality thresholds were recommended by the Illumina Technical Notes: 1) Quality Scores for Next-Generation Sequencing, 2011, 2) Somatic Variant Caller, 2012.

The analysis of identified variants was performed using Alamut Visual 2.9 (Interactive Biosoftware, Rouen, France) and evaluated with open source genomic databases i.e. catalogue of somatic mutations in cancer (COSMIC v77)[6], ExAC Browser (Version 0.3)[7], UniProtKB/Swiss-Prot database, Ensembl genome browser 85 and dbSNP - NCBI - National Institutes of Health. The degree of deleteriousness was calculated using a PHRED-like scoring system "Combined Annotation Dependent Depletion" (CADD) - score. Using this meta-annotation tool, CADD-scores are assigned to genetic variants using an algorithm that creates a consensus based on 60 functional prediction tools. A CADD score of 10 refers to the top 10% of deleterious variants detected in the human genome. CADD score above 20 represents the top 1% of deleterious variants.[8] Finally, the pathogenicity was evaluated using standard *in-silico* predictors: PolyPhen2[9], SIFT[10], Mutation Taster[11] and PredictSNP[12]. The evolutionary conservation of nucleotides/amino acids across species and the physicochemical difference between amino acids was estimated by PhyloP, PhastCons and Grantham score respectively[13]. Sanger sequencing was performed using a Sanger ABI sequencer. Sanger sequences were visualized in Sequence Pilot (SeqPilot v 3.5.2) and CodonCode Aligner v6.0.2.

ACKNOWLEDGMENTS

We are thankful to Sophia Hollander, Christina Jäger, Yahaira Pastor, Alexandra Fischer, Wilfried Truckenmüller (Freiburg), Tamara Szattler (Stockholm) and Bart Przychodzien (USA) for excellent laboratory assistance and data management, and to Dr. Anett Schmidt, Dr. Dagmar Möbius, and Dr. Elisabeth Holfeld (Cottbus) for clinical care.

Funding: Deutsche Krebshilfe (Max Eder grant #109005) to MWW, BMBF (DKTK German cancer consortium, topic molecular diagnostics of pediatric malignancies) to CMN and MWW, BMBF (e:Med FKZ 01ZX1409B) to MB, the German Science Foundation (DFG, SFB 850, to MB and EXE306 to HB), the Research Council, Swedish Cancer Society, and Foundation for Strategic Research to YTB. MWW and ME are past trainees of EHA-ASH translational research training in hematology. MV is an EMBO Long-term Fellow (ALTF 206-2015 co-funded by the European Commission (LTFCOFUND2013, GA-2013-609409). JCA is a Wellcome Trust Senior Research Fellow in Clinical Science (grant 098513/Z/12/Z) with support from the National Institute for Health Research Biomedical Research Centre at Great Ormond Street Hospital for Children, NHS Foundation Trust and UCL, and Great Ormond Street Children's Charity. This study was supported in part by the Excellence Initiative of the German Research Foundation (GSC-4, Spemann Graduate School). The authors are grateful to the Genomics Core Facility at the German Cancer Research Center/DKFZ, Heidelberg, Germany for performing whole exome sequencing and Claritas Genomics (Cambridge, MA, USA) for targeted sequencing in P4.

SUPPLEMETARY REFERENCES

1. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A*, et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010 Sep; **20**(9)**:** 1297-1303.

2. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A*, et al.* From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* 2013; **43:** 11 10 11-33.

4. Wlodarski MW, Hirabayashi S, Pastor V, Stary J, Hasle H, Masetti R*, et al.* Prevalence, clinical characteristics, and prognosis of GATA2-related myelodysplastic syndromes in children and adolescents. *Blood* 2016 Mar 17; **127**(11)**:** 1387-1397.

5. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010 Sep; **38**(16)**:** e164.

6. Forbes SA, Beare D, Gunasekaran P, Leung K, Bindal N, Boutselakis H*, et al.* COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res* 2015 Jan; **43**(Database issue)**:** D805-811.

7. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T*, et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016 Aug 18; **536**(7616)**:** 285-291.

8. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 2014 Mar; **46**(3)**:** 310-+.

9. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P*, et al.* A method and server for predicting damaging missense mutations. *Nat Methods* 2010 Apr; **7**(4)**:** 248-249.

10. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 2009; **4**(7)**:** 1073-1081.

11. Schwarz JM, Rodelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods* 2010 Aug; **7**(8)**:** 575-576.

12. Bendl J, Stourac J, Salanda O, Pavelka A, Wieben ED, Zendulka J*, et al.* PredictSNP: robust and accurate consensus classifier for prediction of disease-related mutations. *PLoS Comput Biol* 2014 Jan; **10**(1)**:** e1003440.

13. Grantham R. Amino acid difference formula to help explain protein evolution. *Science* 1974 Sep 06; **185**(4154)**:** 862-864.