Constitutional *SAMD9L* mutations cause familial myelodysplastic syndrome and transient monosomy 7

Victor B. Pastor,[1,2]* Sushree S. 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, 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, Germany; [10]Clinical Centre South West, Department of Pathology, Böblingen Clinics, Germany; [11]Genetics & Genomic Medicine, UCL Great Ormond Street Institute of Child Health, University College 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, 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, Denmark and [17]Lübeck Institute of Experimental Dermatology, Germany

*VBP and SS contributed equally to this manuscript.

Figure S1. Shared variants found by WES in P1 and P2
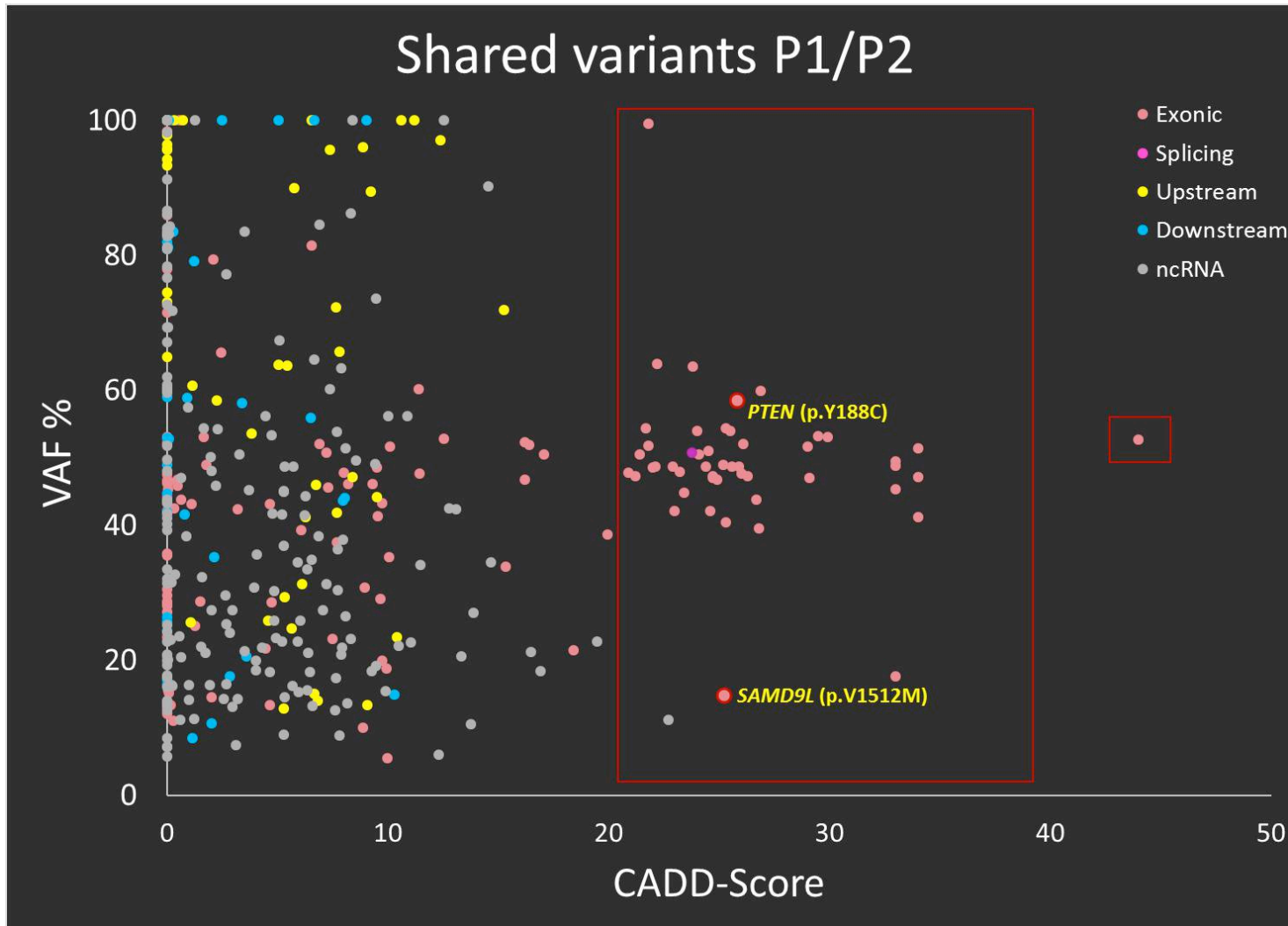
Legend: Y-axis, average VAF % of all shared mutations in patient's 1 and 2; X-axis, CADD-score. The vast majority of mutations depict a low degree of deleteriousness with CADD-scores lower than 20. Framed in red, 50 most deleterious variants of which 48 were of exonic type. Among them, germline mutations in *SAMD9L* (p.V1512M) and *PTEN* (p.Y188C) were observed.

Figure S2. Individual mutations found in P1 by WES
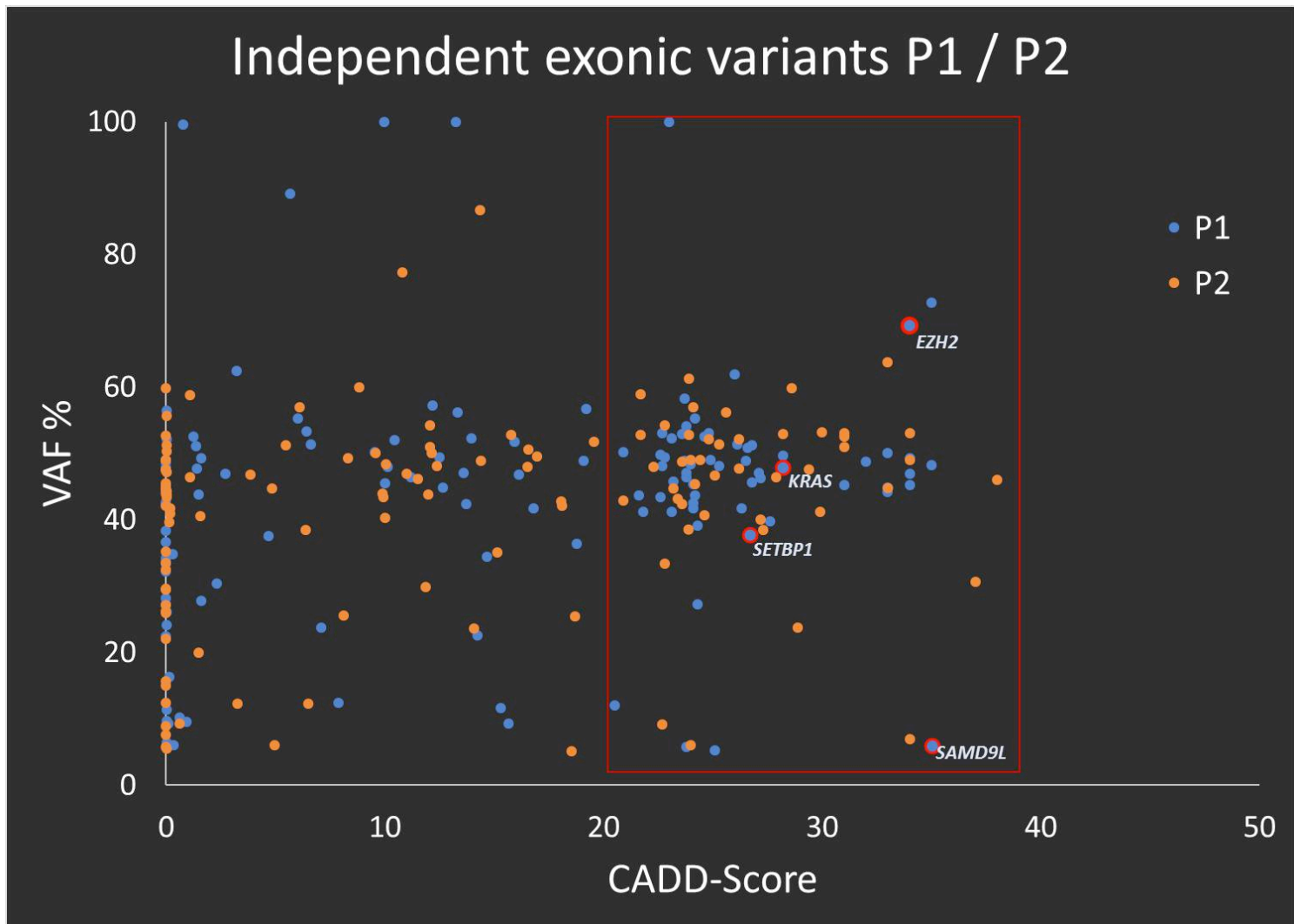
Legend: Individual exonic variants in patient's 1 and 2. Highly deleterious (CADD≥ 20) variants were 58 and 45 in P1 and P2 respectively. Somatic mutations in known leukemic driver genes were detected in P1 in addition to a novel truncating variant in the *SAMD9L* gene. In contrast, no significant mutations involved in blood disease or cancer were found in P2.

Table S1. Allelic frequencies of informative chromosome 7q SNPs in patient 7

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) was 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 hemato-oncologic 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, SNPs 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 an "86-gene 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(3).

Bioinformatics

The variant annotation was compiled using ANNOVAR(4) 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)(5), ExAC Browser (Version 0.3)(6), 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(7). In addition, we used the standard *in-silico* predictors: PolyPhen2(8), SIFT(9), Mutation Taster(10) and PredictSNP(11). The evolutionary conservation of nucleotides/amino acids across species was evaluated with PhyloP and PhastCons, and the physicochemical difference between amino acids was estimated with the Grantham score(12). Sanger sequencing was performed using a Sanger ABI sequencer. DNA sequences were visualized in Sequence Pilot (SeqPilot v 3.5.2) and CodonCode Aligner v6.0.2.

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.

3.   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.

4.   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.

5.   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.

6.   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.

7.   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-+.

8.   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.

9.   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.

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

11.   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.

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