Chromatin remodelling factor SMARCD2 regulates transcriptional networks controlling differentiation of neutrophil granulocytes

DOI:
10.1038/ng.3833

Document Version
Accepted author manuscript

Chromatin remodelling factor SMARCD2 regulates transcriptional networks controlling differentiation of neutrophil granulocytes

Maximilian Witzel[1,2], Daniel Petersheim[1], Yanxin Fan[1], Ehsan Bahrami[1], Tomas Racek[1], Meino Rohlfs[1], Jacek Puchałka[1§], Christian Mertes[2], Julien Gagneur[2,3], Christoph Ziegenhain[4], Wolfgang Enard[4], Asbjørg Stray-Pedersen[5], Peter D. Arkwright[6], Miguel R. Abboud[7], Vahid Pazhakh[8], Graham J. Lieschke[8], Peter M. Krawitz[9], Maik Dahlhoff[10], Marlon R. Schneider[10], Eckhard Wolf[10], Hans-Peter Horny[11], Heinrich Schmidt[1], Alejandro A. Schäffer[12] and Christoph Klein[1,2]

[1]Department of Pediatrics, Dr. von Hauner Children's Hospital, Ludwig-Maximilians-Universität München, Munich, Germany

[2]Gene Center Munich, Ludwig-Maximilians-Universität München, Munich, Germany

[3]Department of Informatics, Technical University of Munich, Munich, Germany

[4]Anthropology and Human Genomics, Department Biology II, Faculty of Biology, Ludwig-Maximilians-Universität München, Munich, Germany

[5]Norwegian National Unit for Newborn Screening, Oslo University Hospital, Oslo, Norway

[6] University of Manchester, Department of Paediatric Allergy & Immunology, Royal Manchester Children's Hospital, Manchester, UK

[7]Department of Pediatrics and Adolescent Medicine, American University of Beirut Medical Center, Beirut, Lebanon

[8]Australian Regenerative Medicine Institute, Monash University, Clayton, Australia

[9]Medical Genetics and Human Genetic, Charite University Hospital, Berlin, Germany

[10]Molecular Animal Breeding and Biotechnology, Gene Center Ludwig-Maximilians-Universität München, Munich, Germany

[11]Pathology Institute, Faculty of Medicine, Ludwig-Maximilians-Universität München, Munich, Germany

[12]National Center for Biotechnology Information, National Institutes of Health, Department of Health and Human Services, Bethesda, MD USA

[§]Deceased (in September 2015)

Correspondence:

Christoph Klein MD PhD

Department of Pediatrics

Dr. von Hauner Children's Hospital

Ludwig-Maximilians-Universität München

Lindwurmstrasse 4

D-80337 Munich

Tel +49-89-4400 5 7700

Fax +49-89-4400 5 7703

Email christoph.klein@med.uni-muenchen.de

Key words:

Hematopoiesis, primary immunodeficiency disorder, SWI/SNF complex, specific granule deficiency, leukemogenesis, congenital neutropenia

Abstract

Differentiation of hematopoietic stem cells follows a hierarchical program of transcription factor regulated events[1-3]. Early myeloid cell differentiation is dependent on PU.1 and CEBPA (CCAAT/enhancer binding protein alpha), late myeloid differentiation is orchestrated by CEBPE (CCAAT/enhancer binding protein epsilon)[4]. The influence of SWI/SNF (SWItch/Sucrose Non-Fermentable) chromatin remodelling factors as novel master regulators of hematopoietic differentiation is only beginning to be explored[3,5,6]. Here, we identify SMARCD2 *(SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily d, member 2)* as a critical regulator of myeloid differentiation in humans, mice, and zebrafish. Studying patients from three unrelated pedigrees characterized by neutropenia, specific granule deficiency, myelodysplasia with excess of blast cells and various developmental aberrations, we identified three loss-of-function mutations in *SMARCD2*. Using mice and zebrafish as model systems, we showed that SMARCD2 controls early steps in the differentiation of myeloid-erythroid progenitor cells. *In vitro*, SMARCD2 interacts with the transcription factor CEBPE and controls expression of neutrophil proteins stored in specific granules. Defective expression of SMARCD2 leads to transcriptional and chromatin changes in acute myeloid leukemia (AML) human promyelocytic cells. In summary, SMARCD2 is a key factor controlling myelopoiesis and a potential tumor suppressor in leukemia.

Article

Differentiation of hematopoietic cells is controlled by transcription-factor mediated instructive events and less well-defined permissive events orchestrated by a variety of epigenetic modulators[7,8]. Dynamic chromatin remodelling adds another level of complexity.

Embedding of promoter DNA into nucleosome landscapes restricts accessibility of cognate binding sites to transcription factors and restricts gene expression [9-11]. The SWI/SNF complex is composed of multimeric units that use energy derived from ATP hydrolysis to unwrap or restructure nucleosomes[12]. SMARCD2 is a component of the SWI/SNF complex in hematopoietic stem cells (HSC) and other hematopoietic cells[6,13,14]. The two paralogous proteins, SMARCD1 (BAF60A) and SMARCD3 (BAF60C), control embryonic stem cell[15] and heart muscle cell differentiation[16], respectively.

Here, we investigated three independent pedigrees with four patients who presented as neonates with delayed separation of umbilical cord and subsequently developed severe bacterial infections associated with neutropenia, parasitosis or chronic diarrhea (Table S1). Extrahematopoietic findings included mild to moderate developmental delay and dysmorphic features (Figure S1, Table S1). The bone marrow of patients showed hypercellularity, paucity of neutrophil granulocytes, dysplastic features (Figure 1), and progressive development of myelodysplasia (Figure 1, S2). Neutrophil granulocytes were characterized by absence of granules (Figure S3).

In search of the underlying genetic defect, we performed homozygosity mapping and whole exome sequencing (WES), followed by Sanger sequencing of patients and family members (see Supplementary Materials & Methods for details). Homozygosity mapping identified an especially large perfect marker interval of over 50Mbp in family A on chromosome 17; within this interval, family B had two non-adjacent perfect intervals spanning 1.8Mbp and 0.5Mbp. The asymptotic LOD scores for these intervals are +4.2 (+1.8 for A and +2.4 for B) and peak observed LOD scores with a more realistic disease haplotype frequency of 0.05, are 3.0 (+1.2, +1.8). There were approximately 36 genes located in the two shared intervals, including *SMARCD2*.

We identified distinct segregating homozygous mutations in *SMARCD2* in all three pedigrees (Figure 2a, b, c). Mutations are described by their putative effect on transcript *SMARCD2-001* (ENST00000448276). Effects on hypothetical transcripts are shown in Table S2. At the DNA level, the mutations in pedigrees A and C affect splice sites, while the mutation in pedigree B is a duplication of 25bp, leading to a frameshift and premature termination (Table S2). Western blot analyses showed an absence of SMARCD2 protein in patient cells (Figure 2d). To confirm that the *SMARCD2* mutations lead to a loss of function, we sequenced reverse transcribed mRNA from patient cells (Figure 2e) and determined their putatively encoded proteins. We then cloned 2 isoforms of patient AII.1 (AII.1a: p.Ile362Cysfs*3 and AII1b: p.Ser394Argfs*1), one isoform of patient BII.1. (BII.1: p.Gln147Glufs*5) and one isoform of patient CII.1 (CII.1: p.Arg73Valfs*8). FLAG-tagged expression vectors carrying mutated *SMARCD2* versions and a red fluorescence protein gene separated by an Internal Ribosomal Entry Sequence (IRES.RFP) were transfected into 293T cells and were investigated for co-immunoprecipitation with native SWI/SNF core members. As shown in Figure 2f, only the wild-type version of SMARCD2 was able to co-precipitate with SMARCA4 (BRG1), SMARCC2 (BAF170), SMARCC1 (BAF155), and SMARCB1 (BAF47); none of the mutant versions were able to co-precipitate with any of these proteins, suggesting that the mutations constitute loss-of-function alleles.

Since all SMARCD2-deficient patients had either been subjected to allogeneic HSCT or had died due to their disease, primary SMARCD2-deficient hematopoietic stem cells were not available for further experiments. To further study the role of SMARCD2 in neutrophil differentiation, we established several *in vivo* and *in vitro* models.

As a first model, we used zebrafish (*Danio rerio*), in which *smarcd2* (XP_692749.2) is the ortholog of human *SMARCD2*. Using antisense morpholino-oligonucleotides (MOs), we created Smarcd2-deficient zebrafish in two reporter strains with fluorescent neutrophil

granulocytes: Tg (mpx:EGFP)^{i114} (Figure S4a,b,c) and Tg (lyz:dsRed)^{nz50} (Figure 3c)[17-19]. *Smarcd2* MOs were designed to block either translation initiation (label ATG) or splicing (labels SB1 and SB2, for MOs targeting splice donor and acceptor sites respectively) of *smarcd2*. In both fish strains, there was a significant reduction in the number of neutrophil granulocytes compared to controls at 72 hours post fertilization (hpf) for the ATG and SB1 MOs (Figure S4c, 3c). MO SB2, which failed to disrupt *smarcd2* splicing (Figure S4a), provided an additional negative control indicating specificity of the on-target smarcd2-MO effect to reduce neutrophil abundance. Using CRISPR/Cas9 genome editing in zebrafish, we created a frameshift mutant *smarcd2* allele Smarcd2$^{1/1}$ (Figure S4d), which also showed reduced granulocyte abundance at 72 hpf compared to wild-type controls (Figure 3a, b). Collectively, these zebrafish models provide concordant evidence that a requirement for SMARCD2 in neutrophil granulocyte differentiation is evolutionarily conserved.

A second *in vivo* model was generated by injection of *Smarcd2*$^{+/-}$ murine ES cells (KOMP repository) into blastocysts and transferring them into pseudo-pregnant mice. Chimeric offspring were mated with wild-type mice, resulting in *Smarcd2*$^{+/-}$ mice, which were intercrossed (Figure S5a, b, f). We found that *Smarcd2*$^{-/-}$ embryos died late during fetal development (Figure 5c, d) and are characterized by reduced size, pallor, and decreased temporal vascularization (Figure 3d), suggestive of a compromised hematopoietic system. However, we did find Mendelian ratios of *Smarcd2*$^{-/-}$ embryos at 14.5dpc (Figure S5d, e). Flow cytometry analysis of fetal liver single cell suspensions showed comparable numbers of hematopoietic stem cells (Figure S5e), yet a striking reduction in CD11b^{+}Gr1^{+} neutrophil granulocytes and CD11b^{+}Ly6c^{+} monocytes in *Smarcd2*$^{-/-}$ embryos (Figure 3h,k).

To assess the differentiation capacity of hematopoietic stem cells, we next purified CD45.2^{+}Lin^{-}Mac$^{+/low}$ Sca1^{+}cKit^{+} (LSK) cells from wild-type, heterozygous and homozygous fetal livers and performed colony-forming unit (CFU) assays, *in vitro*. In comparison to CFU

colonies derived from wild-type or heterozygous mice, *Smarcd2*$^{-/-}$ CFU colonies showed a marked reduction in size and numbers (data not shown and Figure S6a) and maturation arrest (Figure 3f). *Smarcd2*$^{-/-}$ myeloid CFU colonies, generated in the presence of myeloid cytokine cocktail, were deficient in cell surface expression of CD11b, Gr1 and Ly6c (Figure S6b). A block in myeloid differentiation was also seen when LSK cells (native) were exposed to either GM-CSF, M-CSF, or G-CSF, suggesting that none of the corresponding cytokine-receptors were able to induce myeloid cell growth (Figure 3i).

Aberrant hematopoiesis was not restricted to the myeloid compartment in *Smarcd2*$^{-/-}$ embryos, but also affected erythroid differentiation. Fetal/umbilical cord blood cytology at 14.5dpc showed marked dysplastic changes in *Smarcd2*$^{-/-}$ erythropoiesis: In contrast to wild-type embryos, characterized by normochromic, orthochromatic erythrocytes and presence of few nucleated erythrocytes, *Smarcd2*$^{-/-}$ embryos showed extensive anisocytosis of erythrocytes, multinucleated cells, perturbed mitosis and increased apoptosis (Figure 3e). Furthermore, *in vitro* erythroid differentiation of LSK cells in the presence of rm SCF, rm IL-3, rh IL-6, rh EPO hints towards a partial differentiation block or delay at the immature S1 stage, as determined by CD71/Ter119 expression[20] in *Smarcd2*$^{-/-}$ GEMM colonies (Figure 3j,l). Taken together, murine SMARCD2-deficient hematopoietic cell differentiation is characterized by a maturation arrest in myeloid and erythroid cells *in vitro* and *in vivo*, reminiscent of the hematological phenotype in *SMARCD2*$^{-/-}$ patients.

Various previous studies identified that SWI/SNF complex members increase or decrease primitive or definite hematopoiesis[6]. Here, we hypothesize that A) the functional effects of SMARCD2 deficiency on granulopoiesis are due to its absence from SWI/SNF complexes, B) SWI/SNF complexes that contain SMARCD2 have a specific role in granulopoiesis, and C) mechanistically, SMARCD2 governs granulopoiesis via chromatin accessibility and interaction with CEBPE.

To identify alterations in transcriptional networks controlling differentiation of fetal liver hematopoietic stem cells, we isolated LSK cells from 5 *Smarcd2*$^{+/+}$ and 9 *Smarcd2*$^{-/-}$ fetal livers and profiled their transcriptome by RNA-sequencing. Among a total of 12,362 detected genes, we found 4,290 to be differentially expressed at a False Discovery Rate (FDR, see Material and Methods) lower than 10%; Smarcd2 showed the largest expression ratio among all genes, as expected (Figure 3g, Table S4). Interestingly, the majority (79%) of the 605 genes with a relatively large difference (fold-change > 1.4, FDR<1%) were upregulated and not downregulated. This had also been reported for embryonic fibroblasts deficient for SMARCB1 (Snf5) and SMARCA4 (Brg1), two other members of the SWI/SNF complex[11]. The upregulated genes were most enriched in categories related to membrane proteins, including MHC complexes, immunoglobulin domains and G-protein coupled receptors that included signalling pathways related to immunodeficiency and host defence (Table S5, Figure S7a). A subset of CEBPE-dependent genes is also deregulated in *Smarcd2*$^{-/-}$ murine LSK cells (Figure S7 b,c). Consistent with the finding that CpG island (CGI) promoters can facilitate promiscuous induction without a requirement for SWI/SNF[21], we found that genes containing CGI promoters are significantly under-represented within the group of differentially expressed genes (Fisher's exact test, p = 0.00441004, odds ratio=0.71).
Thus, a considerable fraction of the genes that are found to be differentially expressed are directly dependent on SWI/SNF and/or transcription factors.

Even though these experiments suggest that SMARCD2 orchestrates transcriptional networks in early hematopoietic stem cells, they cannot directly explain the striking absence of neutrophil granules and perturbed differentiation of mature neutrophils seen in SMARCD2-deficient patients. To shed light on the mechanisms of SMARCD2 in late neutrophil maturation, we set out to establish a human *in vitro* system to further study the function of SMARCD2. We chose the promyelocytic cell line NB4 that is responsive to retinoic acid

156 signalling and can be differentiated toward mature neutrophil granulocytes *in vitro*. Since our
157 attempts to generate SMARCD2-deficient NB4 cells using CRISPR/Cas9 tools was
158 unsuccessful, we decided to make use of RNA interference to establish cell lines
159 characterized by lower SMARCD2 protein expression. We designed lentiviral shRNA
160 constructs expressing a *SMARCD2*-specific shRNA and the marker gene GFP, transduced and
161 flow-sorted NB4 cells for further analysis.
162
163 NB4 cells express *SMARCD1*, *SMARCD2*, *SMARCD3* and *CEBPE* RNA/cDNA at detectable
164 levels (Figure 4a and [22]). RNA expression of *SMARCD2*, but not of the family members
165 *SMARCD1* and *SMARCD3* was significantly reduced upon lentiviral expression of shRNA
166 directed against *SMARCD2* (Figure 4a). The expression of *CEBPE* was not affected by
167 *SMARCD2* knock down and increased after differentiation with all trans retinoic acid (ATRA)
168 (data not shown) as previously described (e.g.[23]). Next, we systematically analysed RNA
169 expression of genes encoding proteins that are expressed and stored in primary and specific
170 granules in neutrophil granulocytes. Interestingly, during differentiation with ATRA,
171 transcript levels of primary granule proteins cathelicidin (CAMP) and alpha-1-antitrypsin
172 (AAT) as well as specific granule proteins matrix metalloproteinase-8 (MMP8),
173 transcobalamin (TCN1) and lactoferrin (LTF), were all significantly reduced (Figure 4a) in
174 SMARCD2-deficient cells.
175 Mice with targeted mutations in *Cebpe*[24] and human patients with rare mutations in *CEBPE*[25]
176 are characterized by specific granule deficiency and susceptibility to bacterial infections. In
177 view of these phenotypic similarities, we asked whether SMARCD2 controls the effects of
178 CEBPE. RNA-expression of *CEBPE* was not directly affected in SMARCD2-deficient cells.
179 As an alternative, we hypothesized that SMARCD2 may be relevant for recruiting CEBPE to
180 open chromatin and thus facilitating expression of CEBPE-dependent genes. Indeed, co-
181 expression and immune precipitation of HA-tagged CEBPE and Flag-tagged SMARCD2

suggested a direct protein-protein interaction of both proteins in mammalian cells (Figure 4c). A functional link between SMARCD2 and CEBPE is further supported by our finding that documented CEPBE-dependent genes (Table S3) are deregulated in the absence of SMARCD2 in human (Figure 4g, S8c, d) and murine hematopoietic cells (Figure S7b, c).

The consequences of defective nucleosome positioning in dysfunctional SWI/SNF molecules may be complex. We attempted to interrogate effects of SMARCD2 deficiency on global chromatin accessibility using ATAC sequencing. We compared all genes that showed differential chromatin accessibility in *SMARCD2*-knockdown cells with differentially expressed genes determined by RNA-sequencing studies in undifferentiated and ATRA differentiated promyelocytic leukemia cell line NB4. A specific subset of genes was found deregulated in both assays, ATAC-Seq and RNA-Seq (Figure 4c-f), affecting vesicular trafficking, migration and signalling. Differentially expressed genes in both, murine transcriptome (Table S4) and human transcriptome (Table S6,S7), cluster significantly in signalling pathways relevant to immune system functions (Figure S7a and Figure S8a, b, respectively). Taken together, DNA accessibility studies, transcriptome studies and protein-protein interaction studies suggest that SMARCD2 has a direct role to remodel the chromatin and to mediate downstream effects partly by interaction with the myeloid transcription factor CEBPE. In contrast to CEBPE deficiency, SMARCD2 deficiency causes not only absence of specific granule expression, but also defects in early hematopoietic cells associated with AML/myelodysplasia (Figure S2) as well as non-hematopoietic syndromic features (Figure S1, Table S1).

In summary, our clinical and molecular characterization of a previously unrecognized human disease reveals SMARCD2 as a key factor controlling transcriptional networks governing

hematopoietic stem cell differentiation and highlights the relevance for chromatin remodelling in lineage specification in the hematopoietic system.

Patients, Materials, and Methods

Patients

Patients were referred by AS-P, PDA, and MA for genetic assessment of congenital neutrophil deficiencies. The study was approved by the ethics committee of the University Medical School of Hannover and the Faculty of Medicine at LMU, Munich. Patient recruitment, genetic analysis, and data handling were done in accordance with the tenets of the Declaration of Helsinki. Patients or their parents gave informed consent for the genetic and functional studies and for publication of their pictures.

Hematology, biochemistry, and pathological bone marrow studies

Clinical laboratory-based assays, such as blood cell counting, were done by referring centers according to good clinical practices. Bone marrow histological studies were performed on paraffin-embedded samples provided by the referring clinical immunology centers. Following standard histopathological procedures, specimens were cut by microtome (Leica) and stained by SAKURA Tissue-Tek Prisma & Film Automated Slide Stainer (hematoxylin-eosin) or BenchMark XT fully automated IHC/ISH staining instrument (immune histochemistry). In addition to anti-lactoferrin antibody ab15811 (Abcam), antibodies against myeloperoxidase #A0398 (Dako) , CD15 #PNIM1921 (Beckman Coulter), glycophorin C #M0820 (Dako) and CD61 #760-4249 (Ventana/Roche) were used according to the manufacturers' instructions.

Homozygosity mapping and next generation sequencing

Patient AII.1 served as the index case. Patient BII.1, previously described as clinical case report[26] and patient BII.2 (not described) served as reference case for homozygosity mapping using the Affymetrix 6.0 chip as in[27].We searched for perfectly segregating intervals in the SNP data using the software findhomoz[28]. To compute LOD scores, we assumed that the

parents of the affected individuals are second cousins as in[29,30] because they are known not to be first cousins and if they are more distantly related than second cousins, then the LOD scores would be higher. Indeed, in the initial case report, family B was described as "non-consanguineous"[26]. Asymptotic LOD scores, assuming the frequency of the disease-associated marker haplotype decreases in the limit towards 0, were computed by hand using the principle that each meiosis after the first contributes $\log_{10}2$ to the score. We used FASTLINK v. 4.1P[31-33] to check the asymptotic scores and to compute scores with more realistic marker allele frequencies. For LOD score computations, we assumed full penetrance and a very rare disease associated allele. The scores shown here are computed for the two families separately and summed.

Genomic DNA of the two parents and two affected children in family B was enriched for all coding exons using Agilent's SureSelect Human All Exon kit V3-50MB (Agilent Technologies) according to the manufacturer's protocol and subjected to sequencing on an Illumina Genome Analyzer II. Short sequence reads were mapped to the human reference genome GRCh37 with Novoalign and variants were detected as previously described[34-36]. For each possible mutation found in family B, we designed a sequencing assay to test the affected individual in family A (our index patient) for that mutation. Since this failed, we performed high-throughput sequencing in family A, and identified a likely pathological variant in *SMARCD2*: c.1181+1G>A (NM_001098426, ENST00000448276) confirmed by Sanger sequencing. Sanger sequencing of *SMARCD2* in family B revealed a large homozygous insertion in patients BII.1 and BII.2 (c.414_438dup), segregating in family B.

Within our cohort of SCN patients, in patient CII.1, a homozygous mutation in *SMARCD2* (c.401+2T>C) was identified by whole exome sequencing with SureSelect XT Human All Exon V3 + UTRs kit according to the manufacturer's instructions (Agilent Technologies) using SOLiD 5500 next generation sequencing platform (LifeTechnologies) to an average coverage depth of 100x (75 bp forward and 35bp reverse pair-end). Segregation of this

variant in family C was confirmed by Sanger sequencing. In all three families (A, B, C), *CEBPE* and several other candidate genes were excluded (i.e., shown not to contain germline biallelic mutations) by Sanger sequencing or whole exome sequencing ([26], and new data, not shown).

Sanger sequencing of *SMARCD2*

Human *SMARCD2* isoform SMARCD2-001 (ENST00000448276) is consistently annotated (CCDS45756) and was used as the reference sequence for specific sequence-based experiments. Targeted sequencing included all 13 exons of ENST00000448276 as well as one potential alternative Exon 1 derived from isoform SMARCD2-003 (ENST00000323347). Throughout the text, mutations are described by their putative effect on transcript SMARCD2-001 only. Effects on other transcripts are shown in Supplementary Table S2.

DNA was extracted from adherent cells or suspension cells with QIAamp DNA Blood Mini Kit (Qiagen #51106) and used for further application. RNA was extracted with Qiagen RNeasy Micro Kit #74004, RNase inhibitor ribolock (Thermo) was added, RNA was stored at -80°C or used for cDNA transcription using High-Capacity cDNA Reverse Transcription Kit # 4368813 (Life Technologies).

Sanger sequencing was performed on both gDNA and cDNA. The exonic regions of gDNA were amplified by PCR reaction. Per reaction, 2.5µl HiFi buffer, 2.5µl dNTP 2mmol/l, 0.125µl HiFi polymerase, DMSO 1.25µl or betaine 5µl, 1.5µl primer forward/ reverse 10pmol/l, ≥20ng DNA, up to 25µl nuclease free water. The PCR reaction conditions were 95°C melting for 5-10', followed by 35-40 cycles of loops consisting of 90°C melting for 30'', 56°C annealing for 30'', 72°C elongation for 30'' (primer list for exons upon request) or 1'30'' for amplification of full length cDNA exon 1 to exon 13 of ENST00000448276 (SMARCD2-FW GAGCGATGTCGGGCCGAG; SMARCD2-REV ATCCCTGAGCAGTTAGGTCAGGCGAAT). The full length SMARCD2 transcript

amplification was performed with the aim to enrich all potential transcript variants conserved in the N and C termini of SMARCD2-001 ENST00000448276. It includes mutated transcripts of SMARCD2-001 (ENST00000448276) altered by exon skipping, intron retention or insertion/ duplications. Bands were visualized on 1% agarose gel. Clean up of PCR products was performed with ExoSAP-IT (Affimetrix AF 78202) or with QIAquick Gel Extraction (Qiagen #28706). Sanger sequencing was performed in house on a ABI 3130xl Genetic Analyzer or outsourced to Eurofins Company, Munich, Germany. Results were analyzed by Seqman (DNASTAR) or ApE (M. Wayne Davis, Utah[37]) software.

Cell lines

Standard cell lines (NB4, 293T) were purchased from the German Collection of Microorganisms and Cell Cultures (DSMZ). Patient cell lines were cultivated from skin biopsies (fibroblasts) or peripheral blood after infection with Epstein-Barr virus (EBV) (the term we use for the transformed cells lines is EBV-transformed lymphoblastoid cell lines). Adherent cell lines 293T and fibroblasts from healthy donors and patients AII.1 and BII.1 were cultured in DMEM, supplemented with 10%FCS, 50U/ml penicillin, 50 µg/ml streptomycin, 2mM L-Glu. Suspension cell lines NB4 and EBV-LCL were cultured in RPMI, supplemented with 10% FCS, 50 U/ml penicillin, 50 µg/ml streptomycin, 2mM L-Glu, 10mM HEPES buffer.

Plasmids and molecular cloning

SMARCD2 was amplified from a human healthy donor sample or from patients' cDNA (SMARCD2-FW GACGGGACGGAGCGATGT; SMARCD2-REV GAGCAGTTAGGTCAGGCGAATT). Analysis on agarose gels revealed differences in fragment size and/or number of fragments in patients versus healthy donor. Fragments were

gel extracted and cloned into an in-house, CMV driven plasmid modified from the pCHGFPW plasmid[38] using 5' prime XhoI primer (with Kozak and FLAG tag: SMARCD2-Xho1-FLAG-FW: AAACTCGAGGCCACCATGGACTACAAAGACGATGACGACAAGTCGGGCCGAGGCGCG) and 3' prime SpeI primer (SMARCD2-SpeI-REV: TTTACTAGTTTAGGTCAGGCGAATTCCC). Due to individual truncations at the 3´-terminus of mutated proteins, cloning was restricted to 5´-Flag tags for mutated proteins. *SMARCD2*-specific pGIPZ. shRNA constructs and pGIPZ non-silencing control were purchased from Thermo Scientific (shRNA 1-3: clone ID V3LHS_300463; V3LHS_300461; V3LHS_400374 and non-silencing control # RHS4531). The shRNA sequences were cloned via Mlu1 and Xho1 restriction sites into pGIPZ.SF.GFP.2 plasmid (kindly provided by Axel Schambach, MHH Hannover). Viral particles were produced in 293T cells with gag-pol, VSVG and rev helper plasmids. NB4 cell lines were stably transduced, GFP sorted and expanded. Knockdown efficiency was determined by expression of protein (Western blot) and mRNA/cDNA (qPCR method, see section: Expression of neutrophil specific granule genes in NB4 cells).

Immunoprecipitation experiments and western blotting:

Transfection of wild-type and mutant SMARCD2 Flag-tagged proteins or HA-tagged CEBPE with calcium phosphate into semi-confluent 293T with 10µg plasmid/10cm dish was performed. Cells were harvested on day 3, pelleted and lysed in freshly prepared RIPA Buffer (450mM NaCl, 25mM TrisHCl pH7.5, 1mM EDTA, 1%NP40, 5% Glycerol, 25mM Na-Pyrophosphate, 50mM Na-Fluoride, EDTA-free protease inhibitor (Roche)). For each immune-precipitation one confluent 10cm dish with 293T cells has been used. Lysates were cleared by centrifugation (21,000g x for 15 min at 4°C). IP with FLAG affinity gel (Sigma) or HA affinity gel (Thermos Scientific) was performed overnight at 4°C on a rotating laboratory

wheel. Samples were washed 5 times in RIPA buffer. Elution was performed with FLAG peptide 37.5µg (Sigma F3290-4MG) or boiling in 2x Laemmli sample buffer. Lysates were analyzed by Western blot.

Similarly, SMARCD2 expression in fibroblasts and EBV-LCLs of healthy donor and patients was analyzed. Cells were lysed with freshly prepared RIPA buffer. Lysates were cleared by centrifugation (21,000 x g, 10 min, 4 °C). Protein quantification was performed with Bradford reagent (Bio-Rad Laboratories) using ELISA plate readers (Synergy H1 Hybrid Reader, BioTek; infinite M200, Tecan). Equal amounts of protein - achieved by Bradford (SMARCD2 expression in healthy donor and patient cells) or counting of input cells (Immune precipitation) were separated by sodium dodecyl sulphate polyacrylamide (SDS-PAGE) gel electrophoresis and blotted onto polyvinyl difluoride (PVDF) membranes. The membranes were blocked in PBS containing 0.1 % Tween-20 (PBS-T) supplemented with 5 % BSA or 5% non-fat dry milk for 2 hours, followed by incubation with primary antibodies overnight at 4 °C or for 1 h at room temperature in (PBS-T) with 5 % BSA or 5% non-fat dry milk. Antibodies used included anti-Flag, mouse, clone M2, F1804-200ug (Sigma) or anti FLAG-HRP, A8592-.2MG (Sigma), SMARCD2, mouse monoclonal antibody clone F-34, SC-101162, (SCBT), GAPDH mouse monoclonal antibody clone 6C5, SC-32233 (SCBT), anti-HA, rabbit, ab9110 (Abcam), anti BRG1, rabbit, clone EPNCIR111A, ab110641, (Abcam), anti SMARCC2/BAF170, rabbit, #8829S (CST), anti SMARCC1/ BAF 155, rabbit, clone D7F8S, #11956S (CST), anti SNF5/BAF47, rabbit, clone D9C2, #8745S (CST). After washing the PVDF membranes in PBS-T, secondary horse-radish peroxidase conjugated antibodies anti-mouse (BD Pharmingen) and anti-rabbit (CST) were added for 1h at room temperature. After development with chemiluminescent substrate (Pierce ECL western blotting substrate), digital images were acquired on a Chemidoc XRS Imaging System (Bio-Rad Laboratories). Blots were stripped between detection of different antibody probes using

Restore Western Blot Stripping Buffer (Thermo Scientific). Data analysis was performed using Image LabTM software (Bio-Rad Laboratories).

Mouse model

Generation of SMARCD2-deficient murine model

The C57BL/6 embryonic stem cell clone 11930A-F4 carrying a mutant *Smarcd2* allele was generated by Regeneron Pharmaceuticals and obtained from the KOMP repository (www.komp.org). To generate Smarcd2-deficient mice, clonal embryonic stem cells were injected into C57BL/6BrdCrHsd-Tyrc (albino) blastocysts and transferred to pseudo-pregnant NMRI foster mothers. The resulting chimeras were crossed to C57BL/6 albino mice to identify germ line transmission of the targeted allele and to produce mice heterozygous for the mutation. F1 intercrosses of heterozygous mice resulted in *Smarcd2*$^{+/+}$, *Smarcd2*$^{+/-}$, and *Smarcd2*$^{-/-}$ embryos/ mice, which were genotyped using standard PCR reaction conditions and the primers for the wild-type allele: FW: CCATCTGTAACGAAATCCGATGCCC; REV: TTATCCCTCAGGTTCCTGACAAGGC, amplicon size 264bp and for the knock-out allele: FW: GAGTCTAGGGCCTTCTCTTCCTTGC; REV: GCAGCCTCTGTTCCACATACACTTCA, amplicon size 569bp (see Fig S5).
Animals were maintained under specific pathogen-free conditions at 23°C, 65% humidity and with 12h light/dark cycle and had free access to a standard rodent diet (V1534, Ssniff, Soest, Germany) and water. All animal experiments were carried out in accordance with the German Animal Welfare Act with permission from the responsible veterinary authority.

Flow cytometry (FACS)

For FACS analysis of fetal liver hematopoietic cells, single cell suspensions by homogenization of fetal liver tissue with a 1ml Eppendorf pipette and Hank´s buffered salt solution (HBSS) with 3% fetal calf serum (FCS) were prepared. Fetal liver cells were kept on

ice until the genotyping. For FACS analysis of CFU derived hematopoietic cells, CFU colonies were picked after light microscopic evaluation and washed once in HBSS with 3% FCS.

Fetal liver hematopoietic cells and CFU derived cells were stained with the following fluorochrome- or biotin-conjugated monoclonal antibodies for 20 min on ice: anti-B220-AlexaFluor®780 (eBioscience), anti-CD3ε-FITC(eBioscience), anti-CD19-PeCy7(eBioscience), anti-Ter119-PE (BD Pharmigen), anti-Gr1-FITC (BD Pharmigen), anti-Ly6c-PerCP-Cy5.5 (eBioscience), anti-Mac2/CD11b-biotin/- eFluor 450 (eBioscience) and anti-CD71-FITC (BD Pharmigen). Cells stained with biotinylated monoclonal antibodies were washed and incubated with Streptavidin-APC (eBioscience). Samples were acquired on either FACSCanto or LSR II flow cytometer (BD), and data were analyzed using FlowJo software (Tree Star). Fluorescence intensity plots are shown in $\log_{10}$ scales. Relative abundances (percentage of parental gate) were analyzed by Prism software (GraphPad); statistical center value: Mean, standard error: SEM, p-values, and two-tailed unpaired t-tests were used.

Flow cytometry sorting

Murine fetal liver $Lin^{-}Sca1^{+}ckit^{+}$ (LSK) early progenitor cells, were isolated to perform CFU assays and RNA Sequencing (RNA-Seq). Individual embryos were genotyped, fetal liver tissue was suspended in HBSS with 3% FCS, and LSK cells were defined as follows: $CD45.2^{+}$, $lineage^{-}$, $Mac1^{low/+}$, $ckit^{+}$ and $Sca1^{+}$ cells. Cells were then stained with anti-CD45.2-FITC (BD Pharmingen), biotinylated lineage antibodies (anti-B220, -CD3, Gr-1, and -Ter-119; all BD), anti-Mac1/CD11b-eFluor 450 (eBioscience), anti-CD117/c-kit-Alexa Fluor®780 (eBioscience), and anti-Sca-1-PeCy7 (eBioscience). Biotinylated monoclonal antibodies were labeled by incubation with Streptavidin-PerCP/PerCP-Cy5.5 (eBioscience).

LSK cells were sorted into Iscove's Modified Dulbecco's Medium (IMDM) with 3% FCS for CFU assays or directly into 1%TritonX supplemented with RNAse Inhibitor (Promega) for RNA-Seq. Cell sorting was performed using a FACS Aria III cell sorter and FACS Diva software.

Colony Forming Unit Assays

Flow-sorted fetal liver LSK cells were washed and resuspended in 50µl IMDM without FCS. Between 500-1500 LSK cells were plated per 35mm Petri dishes containing 1.3ml MethoCult® (M3231 or M3434, Stem Cell Technologies). M3434 (rm SCF, rm IL-3, rh IL-6, rh EPO) was used to examine erythro-myeloid maturation. M3231 (with addition of murine G-CSF [50 ng/ml], murine GM-CSF [50ng/ml] or murine M-CSF [50ng/ml] (cytokines from Peprotech)) was used to assess myeloid maturation to specific cytokines. CFU colonies were assessed daily from day 3 onwards. Colony forming units (>20 cells) and lineage differentiation potential were assessed using an inverted microscope (Axiovert-II, Zeiss) at day 7 - 12. CFUs were photographed (data not shown), counted and analyzed by FACS. Cytological assessment was performed by May-Grünwald-Giemsa stain after cytospin centrifugation (Shandon Cytofunnel Thermo). CFU counts were normalized to LSK cell input and analyzed by Prism software (GraphPad); statistical center value: Mean, standard error: SEM, p-values, and two-tailed unpaired t-tests were used.

Mouse fetal blood cytology

Fetal blood was recovered from sacrificed embryos and washed in HBSS with 3% FCS. Cytological assessment of equal numbers or nucleated cells was performed by cytospins (Shandon Cytofunnel Thermo) and May-Grünwald-Giemsa stain. Blood cells were morphologically assessed using an inverted microscope (Axiovert-II, Zeiss) and photographed.

443 **Zebrafish experiments**

444 **Zebrafish**

445 Tg(*mpx*:EGFP)^{i114} [19] and Tg(*lyz*:dsRed)^{nz50} [17] strains were used. Fish were held in the
446 FishCore (Monash University) using standard practices. Embryos were held in egg water
447 (0.06 g/L salt (Red Sea, Sydney, Australia)) or E3 medium (5mM NaCl, 0.17 mM KCl, 0.33
448 mM $CaCl_2$, 0.33 mM $MgSO_4$, equilibrated to pH 7.0); from 12 hpf, 0.003% 1-phenyl-2-
449 thiourea (Sigma-Aldrich) was added to inhibit pigmentation. Embryos were held at 28°C in
450 an incubator (Thermoline Scientific) upon collection. Animal experiments followed NHMRC
451 guidelines ("Australian code of the care and use of animals for scientific purposes" 8th edition,
452 NHMRC, 2013) and were approved by the Monash University Animal Ethics Committees.
453

454 **Morpholino knockdown experiments**

455 Microinjection of morpholino oligonucleotides was performed as follows: Antisense
456 morpholino oligonucleotides (Gene Tools, LLC (Eugene, OR)) (Table S3) were resuspended
457 as stock in milli-Q water at 1 mM and microinjected at highest non-toxic concentration for
458 each morpholino (700 μM for all MO-smarcd2). 10-20 min post fertilization embryos were
459 collected in egg water and placed on a 4% agarose gel block aligned to grooves on the gel
460 surface. Microinjection of 1-cell embryos was performed using a standard microinjection
461 apparatus and large-bore needle, positioned at the border of cell and yolk sac.
462 Knockdowns of *smarcd2* by the splice-blocking morpholino oligonucleotides were examined
463 by RT-PCR as follows: Whole embryo RNA was extracted using TRIzol® Reagent (Life
464 Technologies) and cDNA synthesis was performed using SuperScript™ III Reverse
465 Transcriptase (Invitrogen). Phusion High Fidelity DNA Polymerase (Thermo scientific) was
466 used for cDNA amplification. 50 μl PCR reaction was consisting of 1 μl Phusion DNA
467 Polymerase, 10 μl 5X Phusion HF Buffer, 1 μl dNTP (10 mM), 1 μl forward primer (10 μM),
468 1 μl reverse primer (10 μM), 2 μl RT reaction product (cDNA) and 34 μl of nuclease free

water. Biorad T100 thermal cycler with following program was used for amplification: 90 seconds at 95°C as initial denaturation, followed by 30 cycles of 30 sec at 95°C for denaturation, 30 sec at 56 °C for annealing, 30 sec at 72 °C for extension, and final extension at 72 °C for 5 min. Primer sequences in Table S3.

Total numbers of fluorescent neutrophils in digital images of control and morphant embryos were manually counted at 72 hours post fertilization (hpf) using an Olympus MVX10 microscope fitted with an Olympus DP72 camera.

CRISPR/Cas9 mutagenesis in zebrafish model

Single guide RNA (sgRNA) synthesis for CRISPR mutagenesis

The zebrafish smarcd2 gene was mutated by CRISPR/Cas9 technology using the method of Gagnon et al.[39] . Briefly, the web tool "CHOPCHOP" (https://chopchop.rc.fas.harvard.edu/)[39] was used to design gene-specific spacer sequences to contribute to two single guide RNAs (sgRNAs) for smarcd2 targeting (named S1 and S2 in Table S3a). All CHOPCHOP results were checked on zebrafish genome database by Ensembl genome browser. DNA templates for sgRNA synthesis resulted from annealing two single-stranded DNA oligonucleotides (Sigma Aldrich) followed by T4 DNA polymerase (NEB) fill-in, to make a full double-stranded DNA oligonucleotide. Each for each sgRNA DNA template, one oligonucleotide provides the site specific sequence (incorporating either S1 or S2) and the second "constant" oligonucleotide one supplied the binding site for Cas9 enzyme. sgRNAs were generated by in vitro transcription (mMESSAGE mMACHINE® SP6 or T7 Transcription Kit, Thermo Fisher Scientific).Transcribed sgRNA was cleaned (Sephadex G-50 spin columns, Roche Diagnostics) and its integrity was checked on 1% agarose (Bioline, BIO-41025) gel made in 0.5% TBE.

sgRNA microinjection

Individual sgRNAs (50-200 ng/µl) mixed with Cas9 Nuclease 20 µM (NEB) at a 1:1 ratio were microinjected (500–1000 pg) into the cytoplasm of 1-cell stage Tg(mpx:EGFP) embryos.

Genotyping of zebrafish

Smard2 locus genotyping was performed by DNA sequencing. DNA samples were extracted from single embryos or fin clips of adult fish using the HotSHOT protocol[40] and amplified by PCR (primers and PCR conditions, see Table S3a) . Following gel electrophoresis, excised bands (AccuPrep® Gel purification kit, BIONEER) were sequenced in the Micromon sequencing facility (Monash University) using an "Applied Biosystems 3730s Genetic Analyzer". F0 genotyping documented sgRNA activity. F1 genotyping was used to identify founders carrying mutated alleles. F2 genotyping assisted colony management and confirmed the genotypes of all embryos contributing to the phenotype comparison.

Sequencing analysis of zebrafish results

Sequencing traces were analyzed in DNASTAR navigator (Version 2.2.1.1) and ApE (A Plasmid Editor v.2.0.47, similar to[37]). Analysis of complex compound CRISPR/Cas9 genotypes required manual curation and interpretation of sequence chromatograms.

Phenotype analysis of zebrafish

EGFP-positive neutrophils in digital images of control and F2 Tg(mpx:EGFP) embryos of various smarcd2 allelotypes were manually counted at 72 hours post fertilization (hpf) in the tail distal to the tip of the yolk extension, which includes the leukocyte-rich caudal hematopoietic tissue (CHT) using Olympus MVX10 microscope fitted with an Olympus DP72 camera.

Descriptive and analytical statistics were prepared in Prism 5.0c (GraphPad Software Inc). p-values are from two-tailed unpaired t-tests, statistical center value: Mean, error bar: standard deviation.

Murine LSK transcriptome – RNA-Seq

Murine LSK cell populations were sorted into lysis buffer composed of 0.2 % Triton X-100 (Sigma) and 2 U/µl of RNase Inhibitor (Promega). ERCC spike- in controls (Life Technologies) were added to the cell lysis mix at 1:1,000 dilution. RNA was cleaned up from the crude lysate with Agencourt RNAclean XP SPRI beads (Beckman-Coulter). cDNA was synthesized and pre-amplified from 5 µl of lysate as described elsewhere[41]. 0.7 ng of pre-amplified cDNA was used as input for tagmentation by the Nextera XT Sample Preparation Kit (Illumina), where a second amplification round was performed for 12 cycles. For each sample, 5 ng of final library was pooled.10 pmol of the library pool was sequenced 1 x 50 bases on an Illumina HiSeq1500.

RNA-Seq data analysis

We chose a minimum sample size of n=5 according to recommendations of power in RNA-Seq[42]. All sorted murine LSK samples were processed, none was excluded. The murine fetal LSK samples/cell lysates were randomized for RNA-Seq library preparation by assigning a random sample number. During analysis, samples had to be unblinded. Sequencing reads were demultiplexed from the Nextera (i5 and i7) indices. Demultiplexed reads were aligned to the mouse genome (mm10) and ERCC reference using NextGenMap[43]. Count data were generated from mapped reads using feature Counts[44] on Ensembl gene models (GRCm38.74). To remove noise from genes with low expression levels, count data sets were subjected to data-driven gene filtering using the HTSFilter R package[45].

Differential expression (DE) analysis was done in the DESeq2 R package[46]. The full set of the 12362 detected genes, their estimated $\log_2$ fold-change and the adjusted p-value of the Wald test is given in Table S4. For Figure 3g, we used the 50 genes showing the largest difference between the two groups and applied hierarchical clustering gene-wise and sample-wise with complete linkage based on Euclidian distances of variance stabilized counts of DE genes. We displayed the two-dimensional hierarchical cluster results as a heatmap. The reference expression value is the expression average of wild-type LSK cells. For testing enrichment of functional categories we used upregulated ($\log_2$ fold change>0.5) and downregulated genes (<-0.5) as the input list and all 12,362 detected genes as background list for functional annotation clustering using DAVID[47] . Results obtained using default parameters and a cutoff of a 2-fold enrichment are shown in Table S5. Data deposition: GSE84703

Expression of neutrophil specific granule genes in NB4 cells

NB4 AML cells transduced with specific shRNA against SMARCD2 clone 1 V3LHS_300463 and clone 2 V3LHS_400374 or non-silencing control # RHS4531 were cultivate in complete RPMI medium with ATRA 1µM (dissolved in DMSO), after 3 days RPMI medium supplemented with ATRA was exchanged. Cells were analyzed on day 3 and 6. RNA was extracted, cDNA was transcribed and expression levels of *SMARCD1* (FW: GTCAGATGCCGAGGATGGGGA; REV: GTGGCATCATATTTGGACAAGGCTG), SMARCD2 (FW: ATCTCTTGGCTTTTGAGCGGAAGCTG; REV: CTTGCTGGGACTGAACGTATTGGA), *SMARCD3* (FW: GGAGCCGCAGTGCCAAGA; REV: TAAGCCTGGGACTCGGGGAC), as well as granule genes *LFT* (FW: GCTGGAGACGTGGCTTTTATCAGA; REV GTAACTCATACTCGTCCCTTTCAGC), *MMP8* (FW: CCGAAGAAACATGGACCAACACCTC; REV: TGAGCGAGCCCCAAAGAATG), *TCN1* (FW: CACATTTAGCACAGGAGAAGC; REV: TGTTGGCAATTCCAGTCAT), *CALM* (FW: AGAAATCACCCAGCAGGGCAAA; REV:

GTATGGGGACAGTGACCCTCAACC), *AAT* (FW: GAACTCACCCACGATATCATCACC; REV: TGGACAGTTTGGGTAAATGTAAGC) normalized to *GAPDH* (FW: TGATGACATCAAGAAGGTGGTGAAG; REV: TCCTTGGAGGCCATGTGGGCCAT) were detected by SYBR green based qPCR on an ABI Step one plus cycler. Differential expression of genes was calculated by the ΔΔ Ct method. Data points represent the relative fold change of shRNA clone 1 or 2 vs non-silencing control and individual repeat differentiation experiments. Descriptive and analytical statistics were prepared in Prism 5.0 (GraphPad Software Inc) and p-values are from two-tailed unpaired t-tests, statistical center value: Mean, error bar: standard deviation.

RNA-Seq in differentiated NB4 AML cells

NB4 AML cells transduced with specific shRNA against SMARCD2 clone 1 V3LHS_300463 and clone 2 V3LHS_400374 or non-silencing control # RHS4531 were cultivate in complete RPMI medium with ATRA 1µM (dissolved in DMSO) or DMSO-only control or 3 days. RNA was extracted from 1,000,000 NB4 cells (shRNA treated cells (clone 1 V3LHS 300463 and clone 2 V3LHS 400374) and the control cells (RHS4531)) with or without ATRA differentiation. Cells were extracted by GeneJET RNA Purification Kit (Thermo Fisher Scientific) and RNA-Seq library preparation was performed with NEBNext® Ultra RNA Library Prep Kit from Illumina® (#E7530 S) according to the manufacturer´s instructions and sequenced on an Illumina NextSeq 500 at the Dr. von Hauner Children´s Hospital NGS facility. 6 libraries were sequenced together using a Mid output cartridge (FC-404-2001, 150 cycles, paired-end sequencing) reaching approximately 2x 5 Gb per sample.

ATAC-Seq in differentiated NB4 AML cells

Assay for Transposase-Accessible Chromatin with high throughput sequencing was performed as described previously[48] .

NB4 cells (ACC207) were cultured as described above. DMSO (Sigma-Aldrich) was used as a carrier for ATRA. Cells were kept in logarithmic growth and stimulated with 1μM ATRA or DMSO as a control. After 72 hours, 50,000 cells per condition were harvested and nuclei preparation was done as described[48] . Isolated nuclei were treated with Tn5 transposase from the Nextera DNA Library Preparation Kit (Illumina, Catalog # FC-140-1089) for 30 minutes, 300 rpm in a Thermomixer. Transposed DNA was purified with the Qiagen MinElute Reaction Cleanup Kit (Qiagen, Catalog # 28204) and amplified with Illumina Tn5 compatible barcoding primers (NEBNext Multiplex Oligos for Illumina,NEB). We ran a qPCR side-reaction with 5μl of the previously amplified library to determine the minimum number of additional PCR cycles needed (Primers: FW 5′ AATGATACGGCGACCACCGAGAT 3′ and REV 5′ CAAGCAGAAGACGGCATACGA 3′). Minimally PCR-amplified libraries were again purified with the Qiagen MinElute Reaction Cleanup Kit. Libraries were analyzed on an Agilent Bioanalyzer 2100 (High Sensitivity DNA Chip) and size selection for the fragments was performed using AMPure beads.16 ATAC libraries were pooled and sequenced using a Mid output cartridge (FC-404-2001, 150 cycles, paired-end sequencing) on a Illumina NextSeq 500 reaching approximately 2x1.8 Gb per sample.

Human RNA-Seq and ATAC-Seq data analysis

Demultiplexed FASTQ files were generated using bcl2fastq v2.17 (BCL2FASTQ Conversion Software 2.17, Illumina). The ATAC-Seq reads were mapped with BWA-MEM[49] with default parameters to the human genome (GRCh37.p13). The RNA-Seq reads were mapped with STAR (v2.5.0a) to the same genome in combination with the gene model annotation of GENCODE 19[50].

620 The R/Bioconductor[51] package GenomicAlignments was utilized to generate count data for
621 the RNA-Seq data based on the gene level. The same gene model as in the alignment step was
622 used. Only reads mapping uniquely to one feature were counted by setting the
623 *summarizeOverlaps* function option to mode='Union'. Furthermore all other options were set
624 to FALSE (ignore.strand=FALSE, inter.feature=FALSE, fragments=FALSE). To reduce
625 noise from lowly expressed genes, only genes were kept, if the 95 % quantile of the coverage
626 across all samples was below 10 reads as suggested by DESeq2[46] package. DESeq2 was
627 utilized to carry out a differential expression analysis between shRNA treated cells (clone 1
628 V3LHS 300463 and clone 2 V3LHS 400374) and the control cells (RHS4531) of the
629 remaining 13,244 genes. Furthermore, a differential expression analysis was performed
630 between ATRA treated cells and DMSO treated cells as control. A significance cutoff of FDR
631 < 0.1 was applied to the results before further downstream analysis. The DESeq2 results of
632 the significantly differentially expressed genes are listed in Table S6 and S7 with their raw
633 counts and their normalized counts.
634 For the ATAC-Seq analysis we followed the proposed workflow by Lun and Smyth[52]. Hence
635 the reads were first pulled together from all samples. On all reads MACS2[53] was used with
636 default parameters to call peaks. The resulting peaks were used as genomic features to
637 generate count data for each sample with the R/Bioconductor package GenomicAlignment[53].
638 The count data was again subjected to filtering steps to reduce the noise. To call significantly
639 differentially occupied genomic regions the R/ Bioconductor package EdgeR[54] was utilized.
640 A significance cutoff of p value < 0.005 was applied to the results before further downstream
641 analysis. The EdgeR results of the significantly differentially occupied peaks are listed in
642 Table S6 and S7 with their raw counts and their normalized counts.
643 Heatmaps were plotted based on the $\log_2$-fold-changes. The dendrogram was obtained by
644 gene-wise and sample-wise hierarchical clustering with complete linkage. Fold changes and

p-values per gene are given in Table S6 (DMSO/undifferentiated) and Table S7 (ATRA/differentiated).

Pathway analysis

Pathway analysis of human and murine transcriptomes was carried out by Cytoscape 3.3.0[55] together with Reactome Functional Interaction plugin[56,57]. In brief, for pathway analysis, differentially expressed murine genes (log_2 FC -0.5<x>0.5, padj<0.1) or differentially expressed human genes (padj<0.1) or intersections thereof with CEBPE target genes were loaded into the Gene set / Mutation Analysis interface. Networks were generated with or without linker genes as indicated. Spectral partition based network clustering according to[58] was performed and individual spectral clusters where analyzed by Reactome Pathway Enrichment[5959]. Abstraction of spectral clusters (Figure S7, S8) and lists of significantly enriched pathways are shown (Table S3). For intersection of gene lists the Venny Venn online tool[60] was used.

Acknowledgements

We thank all medical and laboratory staff members involved in taking care of patients and performing scientific experiments, in particular Raffaele Conca (FACS sorting) and Jens Hinke (genomic facility), Peter Robinson and Stefan Mundlos for NGS expertise and contribution to the identification of homozygous variants in family A and B. We thank Bernward Zeller for clinical care of patient AII.1 and Anne Tierens for intial FACS and histological workup of patient AII.1. We acknowlegde Julie Lessard for providing antibodies and critical discussion.

The study has been supported by the European Research Council (ERC Advanced Grant „Explore“), the Else-Kröner Fresenius Research School, the German Center for Infection

Research DZIF, the Deutsche Forschungsgemeinschaft (Gottfried Wilhelm Leibniz Program), the German PID-NET (BMBF) and the Care-for-Rare Foundation.
VP was supported by a Monash International Postgraduate Research Scholarship (MIPRS) and Monash Graduate Scholarship (MGS). GL was supported by NHMRC (1069284, 1044754). The Australian Regenerative Medicine Institute (ARMI) is supported by grants from the State Government of Victoria and the Australian Government. This research was supported in part by the Intramural Research Program of the National Institutes of Health, NLM. WE and CZ were supported by the Deutsche Forschungsgemeinschaft (DFG) through LMUexcellent and the SFB1243 (Subproject A14)

Contributions:

MW designed, performed and interpreted experiments, performed writing and editing of the paper. DP performed ATAC-Seq and RNA-Seq, YF, EB TR, MR were involved in genomic and biochemical analyses, JP led the computational biology efforts, CM, JG analysed ATAC-Seq, RNA Seq, CZ and WE performed mouse LSK RNA-Seq and DGE analysis. AS-P, PDA, MRA provided clinical care of patients, VP and GJL generated and analysed zebrafish models, PMK analysed WES in intial patients. MD, MRS, EW generated mice. HPH performed ICH analysis of bone marrow biopsies, HS provided expert clinical genetic consulting, AAS guided bioinformatics studies and helped with writing, editing. CK designed and guided the study, supervised MW, provided laboratory resources and wrote the paper.

Author information:

The authors declare no competing financial interests. Murine LSK RNA-Seq data has been deposited here: GSE84703. Correspondence and requests for material should be addressed to CK (christoph.klein@med.uni-muenchen.de)

References

1 Krumsiek, J., Marr, C., Schroeder, T. & Theis, F. J. Hierarchical differentiation of myeloid progenitors is encoded in the transcription factor network. *PloS one* **6**, e22649, doi:10.1371/journal.pone.0022649 (2011).

2 Orkin, S. H. & Zon, L. I. Hematopoiesis: an evolving paradigm for stem cell biology. *Cell* **132**, 631-644, doi:10.1016/j.cell.2008.01.025 (2008).

3 Krosl, J. *et al.* A mutant allele of the Swi/Snf member BAF250a determines the pool size of fetal liver hemopoietic stem cell populations. *Blood* **116**, 1678-1684, doi:10.1182/blood-2010-03-273862 (2010).

4 Friedman, A. D. Transcriptional control of granulocyte and monocyte development. *Oncogene* **26**, 6816-6828, doi:10.1038/sj.onc.1210764 (2007).

5 Griffin, C. T., Brennan, J. & Magnuson, T. The chromatin-remodeling enzyme BRG1 plays an essential role in primitive erythropoiesis and vascular development. *Development* **135**, 493-500, doi:10.1242/dev.010090 (2008).

6 Huang, H. T. *et al.* A network of epigenetic regulators guides developmental haematopoiesis in vivo. *Nature cell biology* **15**, 1516-1525, doi:10.1038/ncb2870 (2013).

7 Alvarez-Errico, D., Vento-Tormo, R., Sieweke, M. & Ballestar, E. Epigenetic control of myeloid cell differentiation, identity and function. *Nature reviews. Immunology* **15**, 7-17, doi:10.1038/nri3777 (2015).

8 Cedar, H. & Bergman, Y. Epigenetics of haematopoietic cell development. *Nature reviews. Immunology* **11**, 478-488, doi:10.1038/nri2991 (2011).

9 Cairns, B. R. The logic of chromatin architecture and remodelling at promoters. *Nature* **461**, 193-198, doi:10.1038/nature08450 (2009).

10 de la Serna, I. L., Ohkawa, Y. & Imbalzano, A. N. Chromatin remodelling in mammalian differentiation: lessons from ATP-dependent remodellers. *Nature reviews. Genetics* **7**, 461-473, doi:10.1038/nrg1882 (2006).

11 Tolstorukov, M. Y. *et al.* Swi/Snf chromatin remodeling/tumor suppressor complex establishes nucleosome occupancy at target promoters. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 10165-10170, doi:10.1073/pnas.1302209110 (2013).

12 Wilson, B. G. & Roberts, C. W. SWI/SNF nucleosome remodellers and cancer. *Nature reviews. Cancer* **11**, 481-492, doi:10.1038/nrc3068 (2011).

13 Buscarlet, M. *et al.* Essential role of BRG, the ATPase subunit of BAF chromatin remodeling complexes, in leukemia maintenance. *Blood* **123**, 1720-1728, doi:10.1182/blood-2013-02-483495 (2014).

14 Jojic, V. *et al.* Identification of transcriptional regulators in the mouse immune system. *Nature immunology* **14**, 633-643, doi:10.1038/ni.2587 (2013).

15 Alajem, A. *et al.* Differential association of chromatin proteins identifies BAF60a/SMARCD1 as a regulator of embryonic stem cell differentiation. *Cell reports* **10**, 2019-2031, doi:10.1016/j.celrep.2015.02.064 (2015).

16 Lickert, H. *et al.* Baf60c is essential for function of BAF chromatin remodelling complexes in heart development. *Nature* **432**, 107-112, doi:10.1038/nature03071 (2004).

17 Hall, C., Flores, M. V., Storm, T., Crosier, K. & Crosier, P. The zebrafish lysozyme C promoter drives myeloid-specific expression in transgenic fish. *BMC developmental biology* **7**, 42, doi:10.1186/1471-213X-7-42 (2007).

18 Liongue, C., Hall, C. J., O'Connell, B. A., Crosier, P. & Ward, A. C. Zebrafish granulocyte colony-stimulating factor receptor signaling promotes myelopoiesis

and myeloid cell migration. *Blood* **113**, 2535-2546, doi:10.1182/blood-2008-07-171967 (2009).

19 Renshaw, S. A. *et al.* A transgenic zebrafish model of neutrophilic inflammation. *Blood* **108**, 3976-3978, doi:10.1182/blood-2006-05-024075 (2006).

20 Koulnis, M. *et al.* Identification and analysis of mouse erythroid progenitors using the CD71/TER119 flow-cytometric assay. *Journal of visualized experiments : JoVE*, doi:10.3791/2809 (2011).

21 Ramirez-Carrozzi, V. R. *et al.* A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling. *Cell* **138**, 114-128, doi:10.1016/j.cell.2009.04.020 (2009).

22 Iriyama, N. *et al.* Enhancement of differentiation induction and upregulation of CCAAT/enhancer-binding proteins and PU.1 in NB4 cells treated with combination of ATRA and valproic acid. *International journal of oncology* **44**, 865-873, doi:10.3892/ijo.2013.2236 (2014).

23 Tanaka, M., Gombart, A. F., Koeffler, H. P. & Shiohara, M. Expression of bactericidal/permeability-increasing protein requires C/EBP epsilon. *International journal of hematology* **85**, 304-311, doi:10.1532/IJH97.05162 (2007).

24 Lekstrom-Himes, J. & Xanthopoulos, K. G. CCAAT/enhancer binding protein epsilon is critical for effective neutrophil-mediated response to inflammatory challenge. *Blood* **93**, 3096-3105 (1999).

25 Lekstrom-Himes, J. A., Dorman, S. E., Kopar, P., Holland, S. M. & Gallin, J. I. Neutrophil-specific granule deficiency results from a novel mutation with loss of function of the transcription factor CCAAT/enhancer binding protein epsilon. *The Journal of experimental medicine* **189**, 1847-1852 (1999).

26 Wynn, R. F. *et al.* Intractable diarrhoea of infancy caused by neutrophil specific granule deficiency and cured by stem cell transplantation. *Gut* **55**, 292-293, doi:10.1136/gut.2005.081927 (2006).

27 Kotlarz, D. *et al.* Loss-of-function mutations in the IL-21 receptor gene cause a primary immunodeficiency syndrome. *The Journal of experimental medicine* **210**, 433-443, doi:10.1084/jem.20111229 (2013).

28 Glocker, E. O. *et al.* Inflammatory bowel disease and mutations affecting the interleukin-10 receptor. *The New England journal of medicine* **361**, 2033-2045, doi:10.1056/NEJMoa0907206 (2009).

29 Bohn, G. *et al.* A novel human primary immunodeficiency syndrome caused by deficiency of the endosomal adaptor protein p14. *Nature medicine* **13**, 38-45, doi:10.1038/nm1528 (2007).

30 Hamada, T. *et al.* Lipoid proteinosis maps to 1q21 and is caused by mutations in the extracellular matrix protein 1 gene (ECM1). *Human molecular genetics* **11**, 833-840 (2002).

31 Cottingham, R. W., Jr., Idury, R. M. & Schaffer, A. A. Faster sequential genetic linkage computations. *American journal of human genetics* **53**, 252-263 (1993).

32 Lathrop, G. M., Lalouel, J. M., Julier, C. & Ott, J. Strategies for multilocus linkage analysis in humans. *Proceedings of the National Academy of Sciences of the United States of America* **81**, 3443-3446 (1984).

33 Schaffer, A. A., Gupta, S. K., Shriram, K. & Cottingham, R. W., Jr. Avoiding recomputation in linkage analysis. *Human heredity* **44**, 225-237 (1994).

34 Krawitz, P. *et al.* Microindel detection in short-read sequence data. *Bioinformatics* **26**, 722-729, doi:10.1093/bioinformatics/btq027 (2010).

35 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

36 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033 (2010).

37 Zhu, L. J., Holmes, B. R., Aronin, N. & Brodsky, M. H. CRISPRseek: a bioconductor package to identify target-specific guide RNAs for CRISPR-Cas9 genome-editing systems. *PloS one* **9**, e108424, doi:10.1371/journal.pone.0108424 (2014).

38 Van Maele, B., De Rijck, J., De Clercq, E. & Debyser, Z. Impact of the central polypurine tract on the kinetics of human immunodeficiency virus type 1 vector transduction. *Journal of virology* **77**, 4685-4694 (2003).

39 Gagnon, J. A. *et al.* Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and large-scale assessment of single-guide RNAs. *PloS one* **9**, e98186, doi:10.1371/journal.pone.0098186 (2014).

40 Meeker, N. D., Hutchinson, S. A., Ho, L. & Trede, N. S. Method for isolation of PCR-ready genomic DNA from zebrafish tissues. *BioTechniques* **43**, 610, 612, 614 (2007).

41 Picelli, S. *et al.* Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nature methods* **10**, 1096-1098, doi:10.1038/nmeth.2639 (2013).

42 Liu, Y., Zhou, J. & White, K. P. RNA-seq differential expression studies: more sequence or more replication? *Bioinformatics* **30**, 301-304, doi:10.1093/bioinformatics/btt688 (2014).

43 Sedlazeck, F. J., Rescheneder, P. & von Haeseler, A. NextGenMap: fast and accurate read mapping in highly polymorphic genomes. *Bioinformatics* **29**, 2790-2791, doi:10.1093/bioinformatics/btt468 (2013).

44 Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930, doi:10.1093/bioinformatics/btt656 (2014).

45 Rau, A., Gallopin, M., Celeux, G. & Jaffrezic, F. Data-based filtering for replicated high-throughput transcriptome sequencing experiments. *Bioinformatics* **29**, 2146-2152, doi:10.1093/bioinformatics/btt350 (2013).

46 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).

47 Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* **4**, 44-57, doi:10.1038/nprot.2008.211 (2009).

48 Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]* **109**, 21 29 21-29, doi:10.1002/0471142727.mb2129s109 (2015).

49 Li, H. Aligning reads, clone sequences and assembly contigs with BWAMEM. http://arxiv.org/abs/1303.3997v2 (2013).

50 Harrow, J. *et al.* GENCODE: the reference human genome annotation for The ENCODE Project. *Genome research* **22**, 1760-1774, doi:10.1101/gr.135350.111 (2012).

51 Lawrence, M. *et al.* Software for computing and annotating genomic ranges. *PLoS computational biology* **9**, e1003118, doi:10.1371/journal.pcbi.1003118 (2013).

52 Lun, A. T. & Smyth, G. K. De novo detection of differentially bound regions for ChIP-seq data using peaks and windows: controlling error rates correctly. *Nucleic acids research* **42**, e95, doi:10.1093/nar/gku351 (2014).

53 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome biology* **9**, R137, doi:10.1186/gb-2008-9-9-r137 (2008).

54 Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140, doi:10.1093/bioinformatics/btp616 (2010).

55 Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* **13**, 2498-2504, doi:10.1101/gr.1239303 (2003).

56 Croft, D. *et al.* The Reactome pathway knowledgebase. *Nucleic acids research* **42**, D472-477, doi:10.1093/nar/gkt1102 (2014).

57 Milacic, M. *et al.* Annotating cancer variants and anti-cancer therapeutics in reactome. *Cancers* **4**, 1180-1211, doi:10.3390/cancers4041180 (2012).

58 Newman, M. E. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 8577-8582, doi:10.1073/pnas.0601602103 (2006).

59 ReactomeWiki. ReactomeFIViz. http://wiki.reactome.org/index.php?title=ReactomeFIViz&oldid=7168 (2016).

60 Oliveros, J. C. Venny. An interactive tool for comparing lists with Venn's diagrams. http://bioinfogp.cnb.csic.es/tools/venny/index.html (2007-2015).

Figure Legends

Figure 1: Bone marrow and peripheral blood cell analysis

(a-c) Bone marrow and peripheral blood cells from healthy donor. (a) Hematoxylin eosin (HE) stained normal cellular bone marrow biopsy showing regular maturation of all hematopoietic cell lineages and no infiltration by blast cells. Insert shows magnification of normal hematopoiesis. (b) Giemsa stained healthy donor peripheral blood cells showing mature, segmented neutrophil granulocytes (c) Giemsa stained healthy donor bone marrow cells showing maturation of red and white blood cell lineages.

(d-f) Bone marrow and peripheral blood cells from patient AII.1. (d) HE stained bone marrow biopsy with diffuse and compact infiltration by blast cells, absence of megakaryocytes and erythroid islands. Insert shows immature neutrophilic cells. (e) Giemsa stained bone marrow cells showing atypical mature neutrophilic cells with hypogranulated cytoplasm, hyposegmented nuclei and Pseudo Pelger-Huët anomaly (PPHA) (black arrow head). (f) Giemsa stained bone marrow cells showing left shifted neutrophilic granulopoiesis, blast cells and PPHA (black arrow head)

(g-i) Bone marrow and peripheral blood cells from family B (g) HE stained bone marrow biopsy from patient BII.2 (boy) showing a marked hypercellularity with subtotal depletion of adipocytes and normal erythroid precursors. Diffuse infiltration by blast cells and starry sky pattern with disseminated activated macrophages (empty arrow heads). Insert shows immature neutrophilic cells. (h) Giemsa stained peripheral blood cells from patient BII.1 (girl) showing blood smear with circulating atypical neutrophil cells and PPHA (black arrow head).

(i) Giemsa stained bone marrow cells from patient BII.1 (girl) showing left shifted atypical neutrophilic granulocytopoiesis with increase of blast cells. PPHA (black arrow head) and atypical mature neutrophils (empty arrow head).

(j-l) Bone marrow cells from patient CII.1. (j) HE stained bone marrow biopsy from patient CII.1 showing marked hypercellularity with subtotal depletion of adipocytes and depletion of normal erythrocytes. Diffuse and compact infiltration by blast cells, scattered activated macrophages (empty arrow heads). Insert showing pleomorphic blast cells with round nuclei and small but distinguishable nucleoli (black arrow heads). (k) Bone marrow anti GlycoC staining demonstrates superseding of erythropoiesis (empty arrow heads) by blast invasion (asterisks). (l) immunohistochemical anti CD-61 staining shows loosely scattered, small and immature megakaryocytes (micro-megakaryocytes).

Figure 2: Identification of biallelic loss-of-function mutations in *SMARCD2*

(a, b, c.) Pedigrees and Sanger sequencing chromatogram of patient (Pat) compared to reference sequence (Ref), and specification of homozygous mutations (Mut).

(d) Western blot revealing absence of SMARCD2 protein expression (molecular weight 58.9kDa, filled arrow head) in fibroblasts (healthy donors 1 (HD1) and healthy donor 2 (HD2), patients AII.1, patients BII.1) and in EBV-transformed B cell lines (Healthy donor (HD), patient CII.1)

(e) *SMARCD2* mRNA transcripts detected in patients cells, open reading frames are shown in black. Healthy donor (HD) transcript ENST00000448276, CCDS45756, in comparison to patients AII.1 (a: p.I362CfsX2, b: p.S394RfsX1, c: p.I362VfsX85), BII.1 (p.Q147EfsX4) and CII.1 (p.R73VfsX8). See also Table S2.

(f) Immunoprecipitation showing defective binding of patient-specific SMARCD2-mutated proteins to SWI/SNF core complex members BRG1, BAF170, BAF155, and BAF47. FLAG-tagged SMARCD2 proteins (wildtype and mutants), expressed in 293T cells, were immunoprecipitation using anti-FLAG tag. Native co-immunoprecipitation of SWI/SNF complex members was visualized using Western blotting of input and immunoprecipitated samples. Short exposure of FLAG stained membrane shows similar amounts of FLAG-precipitated SMARCD2-WT and SMARCD2-AII1.a and SMARCD2-AII.1b proteins (no shown); long exposure of FLAG stained membrane reveals presence of FLAG-IP precipitated SMARCD2-WT, SMARCD2-AII1.a, SMARCD2-AII.1b, SMARCD2-BII.1 and SMARCD2-CII.1 proteins.

Figure 3: Smarcd2 deficiency in genetic model organisms

(a-c) Defective neutrophil differentiation in zebrafish models

(a) Fluorescence image of *Danio rerio* strain Tg(mpx:EGFP)i114, *smarcd2*$^{wt/wt}$ and *smarcd2*$^{1/1}$ (knock out). Reduced numbers of GFP expressing neutrophils are observed in the *smarcd2*$^{1/1}$ mutant fish embryo. Numbers of fluorescence labled neutrophils were evaluated in the caudal hematopoietic tissue per individual fish embryos.

(b) Statistics of neutrophil numbers in wildtype *smarcd2*$^{wt/wt}$ vs. knock out *smarcd2*$^{1/1}$ zebrafish. Numbers of fluorescence labled neutrophils were evaluated in the caudal hematopoietic tissue per individual fish embryos. (Center value: mean and error bar: SD, p-values: two-tailed unpaired t-tests.)

(c) Statistics of neutrophil numbers in Tg(lyz:dsRed)^{nz50} zebrafish at 72 hours post fertilization after injection of morpholino oligonucleotides (unspecific control versus translation start site blocker (ATG) and splice site blocker (SB1 and SB2) against *smarcd2*. Data points represent numbers of fluorescence labeled neutrophils per individual fish embryos. (Center value: mean and error bar: SD, p-values: two-tailed unpaired t-tests).

(d-l) Defective hematopoiesis in murine Smarcd2-deficient embryos

(d) Gross morphology of murine litter mate embryos *Smarcd2*$^{+/+}$, *Smarcd2* $^{+/-}$ and *Smarcd2*$^{-/-}$ at 14.5dpc, (representative pictures chose out of 4 litters (Wt n=4, Ht n=10, Ko n=9)), (e) *Smarcd2*$^{+/+}$ and *Smarcd2*$^{-/-}$ blood cytology at 14.5dpc, May-Grünwald/Eosin 20x and 63x showing an increase of dysplastic red cell precursors in *Smarcd2*$^{-/-}$ blood e.g. anisocytosis (empty arrow head, 23x), increased mitosis (black arrow heads, 63x) and multinucleated cells (empty arrow heads, 63x), 2 independent experiments in 2 litter, (f) May-Grünwald/Eosin stained CFU cells derived from *Smarcd2*$^{+/+}$, *Smarcd2*$^{+/-}$ and *Smarcd2*$^{-/-}$ hematopoietic stem cells upon differentiation with SCF, Il3, Il6 show promyelocytic arrest. Murine neutrophils

(anular shaped nucleus) are absent in *Smarcd2*$^{-/-}$ colonies. (g) RNA-Seq analysis of *Smarcd2*$^{+/+}$ (n=5) and *Smarcd*$^{-/-}$ (n=9) 14-15dpc fetal liver LSK cell samples. Shown is a heatmap of 50 significantly (padj < 0.1) differentially expressed genes with highest absolute fold change. Each column represents a LSK sample of one embryo. *Smarcd2*$^{+/+}$ is depleted in all *Smarcd2*$^{-/-}$ samples (black arrow). Color key below heat map indicates range of log2 fold changes. For detailed statistic methodology please refer to material/ method section.

(h, k) Representative FACS scatter blots of fetal liver cells stained for CD11b^{+}Gr1^{+} neutrophil granulocytes and CD11b^{+}Ly6c^{+} monocytes and statistical analysis. (h) Here, *Smarcd2*$^{+/+}$ fetal liver blood cells express CD11b^{+} (21.3%) vs CD11b^{+}Gr1^{+} (12.1%) and CD11b^{+}Ly6c^{+} (10%), *Smarcd2*$^{+/-}$ fetal liver blood cells express CD11b^{+} (33.1%) vs CD11b^{+}Gr1^{+} (8.27%) and CD11b^{+}Ly6c^{+} (7.51%) and *Smarcd2*$^{-/-}$ fetal liver blood cells express CD11b^{+} (13.9%) vs CD11b^{+}Gr1^{+} (0.47%) and CD11b^{+}Ly6c^{+} (1.1%). (k) statistical analysis of (h), CD11b^{+}Gr1^{+} and CD11b^{+}Ly6c^{+} cells are significantly reduced in *Smarcd2*$^{-/-}$ vs *Smarcd2*$^{+/+}$. Representation of 2 experiments with 6 litters (Wt n=10, Ht n=14, Ko n=9) with center value: Mean, Error bar: SEM p-values: two-tailed unpaired t-tests. Experiment was repeated 3 times with a total of 8 litters.

(i) Count of myeloid colonies derived from *Smarcd2*$^{+/+}$, *Smarcd2*$^{+/-}$ and *Smarcd2*$^{-/-}$ LSK cells upon differentiation with myeloid cocktail M3434, GM-CSF, M-CSF or G-CSF, respectively. CFU from LSK of Wt n=4, Ht n=5, Ko n= 5, derived from 5 litter, 3 independent experiments, center value: Mean, Error bar: SD, p-values: two-tailed unpaired t-tests).

(j,l) FACS analysis of erythropoietic progenitors derived from *Smarcd2*$^{+/+}$, *Smarcd2*$^{+/-}$ and *Smarcd2*$^{-/-}$ CFU GEMM colonies after differentiation with M3434 (myeloid cocktail). (j) FACS scatter blots show representative CD71/Ter119 distribution of erythropoietic cells from GEMM colonies (l) Display shows distribution of of erythroid stages S0-S5 in 8 GEMM colonies derived per each of 3 WT (i.e. 3 x 8 GEMM colonies), 2 HT (i.e. 2x 8 GEMM

colonies) and 3 KO (i.e. 3x8 GEMM colonies) from fetal liver LSKs (i.e. 64 data points). Statistical analysis by 2-way ANOVA, center value: Mean, Error bar: SEM)

Figure 4: SMARCD2, granule formation and transcriptional regulation

(a) Relative mRNA expression of *SMARCD1* (dots), *SMARCD2* (dots), *SMARCD3* (dots) and primary granule genes (empty squares) *LL37*, *AAT* and secondary granule genes (filled squares) *MMP8*, *TCN1*, *LTF* is shown. The human AML-NB4 cell line was lentivirally transduced with either an unspecific control (CTRL) or 1 of 2 specific shRNAs against *SMARCD2*. Data points show the relative expression of shRNA vs CTRL in 3 independent experiments with 2 shRNAs (i.e. 6 data points) for *SMARCD1 , SMARCD3, LL37, AAT, MMP8*, or in 4 independent experiments with 2 shRNAs (i.e. 8 data points) for *SMARCD2* and *LTF*. *SMARCD1*, *SMARCD2* and *SMARCD3* expression levels were determined in undifferentiated cells; granule gene expression was measured after differentiation with ATRA 1µM for 6 days. To describe the effects of *SMARCD2* knock down, the relative expression levels in all samples were compared to the relative expression of *SMARCD1*. Statistics: center value: Mean, p-values: two-tailed unpaired t-tests. Relative expression of granule genes after knock down and differentiation with ATRA 1µM for 3 days show similar results (data not shown).

(b) Co- overexpression of CEBPE-N-HA and SMARCD2-N-FLAG shows protein-protein interaction *in vitro* in 293T cells. SMARCD2-N-FLAG co-precipitates with HA-immune precipitated CEBPE-N-HA and vice versa.

(c-f) The intersection of differentially enriched genes (ATAC-Seq and RNA-Seq) in NB4 knock down SMARCD2 vs. control is shown.

(c) In undifferentiated NB4 cells (UD) a distinct subset of genes shows both, changes of chromatin compaction measured by ATAC-Seq and gene expression measured by RNA-Seq.

(d) Fold change of transcription (FC RNA) and chromatin accessibility (FC ATAC) are indicated for genes, affected in both assays in undifferentiated NB4 cells. Color key (same for

(d) and (f)) below heat map indicates range of log2 fold changes. For detailed statistic methodology please refer to material/ method section.

(e) In differentiated NB4 cells (ATRA 1μM for 3 days), a distinct subset of genes shows both, changes of chromatin compaction measured by ATAC-Seq and gene expression measured by RNA-Seq.

(f) Fold change of transcription (FC RNA) and chromatin accessibility (FC ATAC) are indicated for genes, affected in both assays in differentiated NB4 cells. Color key below heat map (d) indicates range of log2 fold changes (for (d) and (f)). For detailed statistic methodology please refer to material/ method section.

(g) SMARCD2 regulates expression of CEBPE dependent genes. Intersection of differentially expressed genes in undifferentiated vs differentiated SMARCD2 knock down cells vs CEBPE targets is shown, for intersections see Table S3.

For detailed statistical methodology of ATAC-Seq and RNA-Seq data analysis please refer to Materials and Methods section.

Fig. 1

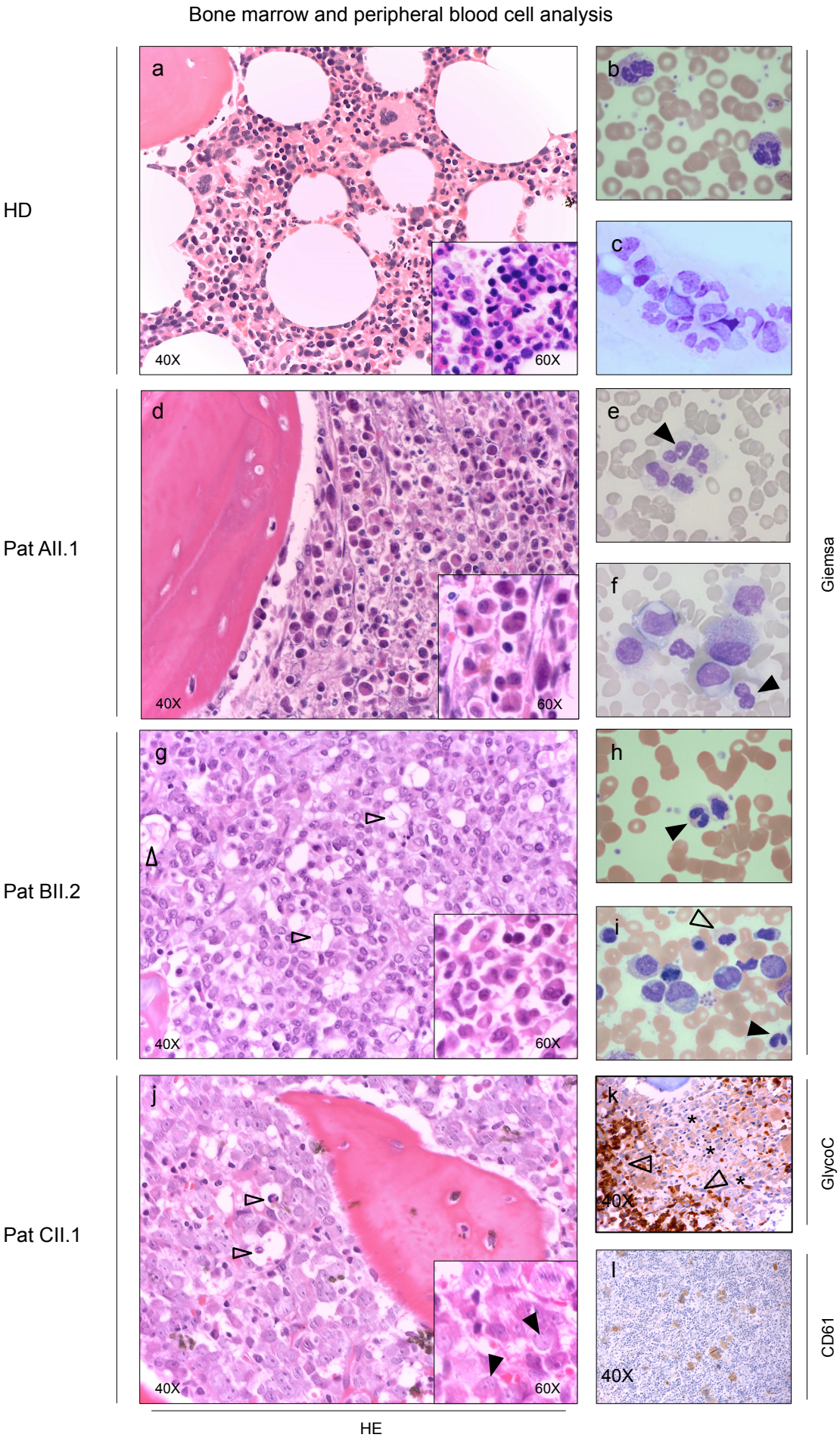
Bone marrow and peripheral blood cell analysis
HD
Pat AII.1
Pat BII.2
Pat CII.1
a
b
c
d
e
f
g
h
i
j
k
l
40X
60X
Giemsa
GlycoC
CD61
HE

Fig. 2

Identification of biallelic loss-of-function mutations in *SMARCD2*

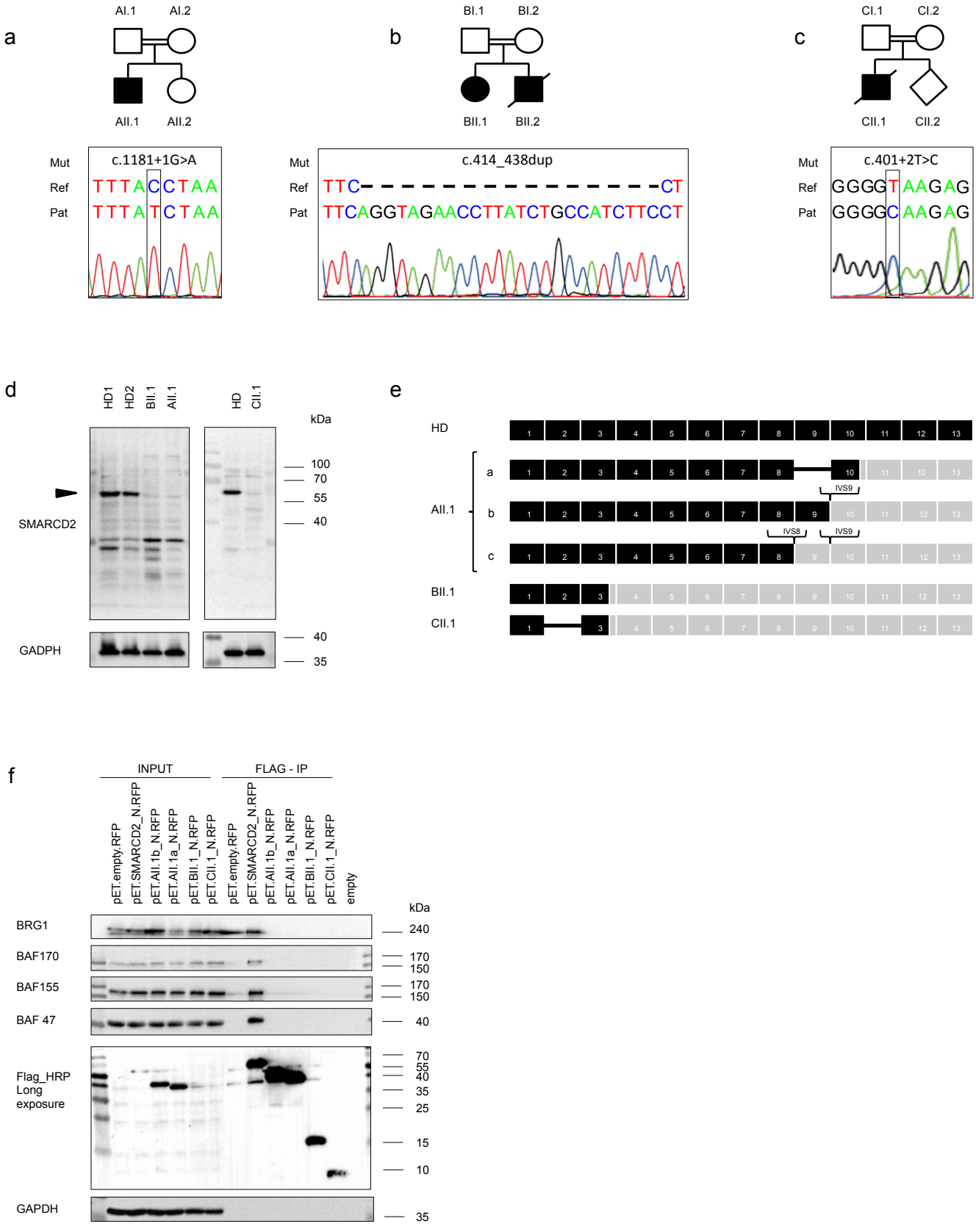

Fig. 3

Smarcd2 deficiency in genetic model organisms

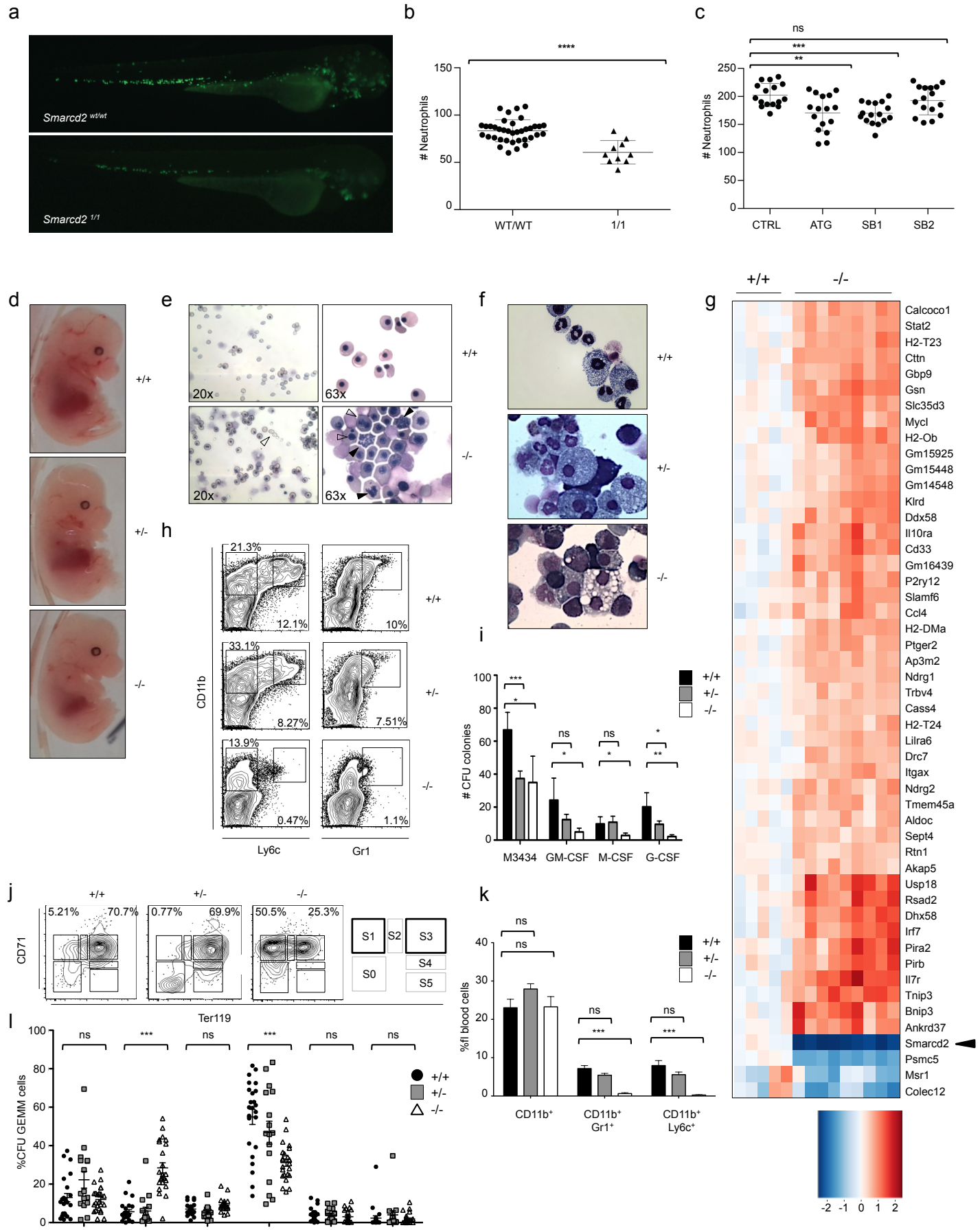

Fig. 4

SMARCD2, granule formation and transcriptional regulation