Mechanisms Of Progression Of Myeloid Preleukemia To Transformed Myeloid Leukemia In Children With Down Syndrome

Maurice Labuhn[1^], Kelly Perkins[2^], Sören Matzk[4,5], Leila Varghese[6], Catherine Garnett[2], Elli Papaemmanuil[3], Marlen Metzner[2], Alison Kennedy[2], Vyacheslav Amstislavskiy[5], Thomas Risch[5], Raj Bhayadia[4], David Samulowski[4], David Cruz Hernandez[2], Bilyana Stoilova[2], Valentina Iotchkova[2]; Udo Oppermann[7] Carina Scheer[1], Kenichi Yoshida[8], Adrian Schwarzer[1], Jeffrey Taub[9], John D. Crispino[10], Mitchell J Weiss[11], Asuhide Hayashi[12], Takashi Taga[13], Etsuro Ito[14], Seishi Ogawa[7,15], Dirk Reinhardt[16], Marie-Laure Yaspo[5], Peter J Campbell[17], Irene Roberts[2,18], Stefan Constantinescu[6], Paresh Vyas[2,19*], Dirk Heckl[1,4*], Jan-Henning Klusmann[4*].

[1]Pediatric Hematology and Oncology, Hannover Medical School, 30625 Hannover Germany. [2]MRC MHU, BRC Hematology Theme, Oxford Biomedical Research Centre, Oxford Centre for Haematology, WIMM, Radcliffe Department of Medicine, University of Oxford, UK OX3 9DU. [3]Departments of Epidemiology and Biostatistics and Cancer Biology, MSKCC, New York NY 10065, USA. [4]Pediatric Hematology and Oncology, Martin-Luther-University Halle-Wittenberg, 06120, Halle, Germany. [5]Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany. [6]Ludwig Institute for Cancer Research Brussels Branch, Brussels BE1200 Belgium. [7]Botnar Research Centre, NDORMS, Oxford NIHR BRC and Structural Genomics Consortium, UK University of Oxford, Oxford, OX3 7LD UK [8]Department of Pathology and Tumor Biology, Graduate School of Medicine, Kyoto University, Sakyo-ku Kyoto, 606-8315 Japan. [9]Division of Pediatric Hematology/Oncology, Children's Hospital of Michigan, Wayne State University School of Medicine, Detroit MI 48201, USA. [10]Division of Hematology/Oncology, Northwestern University, Chicago IL 60611, USA. [11]Hematology Department, St. Jude Children's Research Hospital, Memphis, TN 38105, USA. [12]Institute of Physiology and Medicine, Jobu University, Takasaki-shi, Gunma 370-0033, Japan; [13]Department of Pediatrics, Shiga University of Medical Science, Shiga 520-2192, Japan; [14]Department of Pediatrics, Hirosaki University Graduate School of Medicine, Hirosaki 036-8562, Japan. [15]Center for Hematology and Regenerative Medicine, Karolinska Institute, SE-171 77, Stockholm, Sweden. [16]Pediatric Hematology and Oncology, Pediatrics III, University Hospital Essen, 45122 Essen Germany. [17]Cancer Genome Project, Wellcome Trust Sanger Institute, Hinxton CB10 1SA, UK. [18]Department of Paediatrics University of Oxford OX3 9DS. [19]Department of Haematology, Oxford University Hospitals NHS Trust OX3 7LE.

^ These authors contributed equally
* Senior authors

Corresponding authors: Paresh Vyas (paresh.vyas@imm.ox.ac.uk), Jan-Henning Klusmann (Jan-Henning.Klusmann@uk-halle.de), Dirk Heckl (Heckl.Dirk@MH-Hannover.de).

Lead contact: Paresh Vyas

Highlights

- Genetic and functional analyses of myeloid preleukemia & leukemia in Down Syndrome
- Non-GATA1 preleukemic mutations are often not required for preleukemia
- Previously undescribed transforming hotspot mutation in CSF2RB identified
- Loss of function of 18 genes validated in transformation of preleukemia to leukemia

Summary

Myeloid leukemia in Down syndrome (ML-DS) clonally evolves from transient abnormal myelopoiesis (TAM), a preleukemic condition in DS newborns. To define mechanisms of leukemic transformation, we combined exome and targeted re-sequencing of 111 TAM and 141 ML-DS samples with functional analyses. TAM requires trisomy 21 and truncating mutations in *GATA1*; additional TAM variants are usually not pathogenic. By contrast, in ML-DS, clonal and subclonal variants are functionally required. We identified a recurrent and oncogenic hotspot gain-of-function mutation in myeloid cytokine receptor *CSF2RB*. By a multiplex CRISPR-Cas9 screen in an *in vivo* murine TAM model we tested loss-of-function of 22 recurrently mutated ML-DS genes. Loss of 18 different genes produced leukemias that phenotypically, genetically and transcriptionally mirrored ML-DS.

Key words

Acute Myeloid Leukemia, Down syndrome, preleukemia, cancer transformation, CRISPR Screen, GATA1

Significance

Transformation from human preleukemia to leukemia is incompletely understood. Transient abnormal myelopoiesis (TAM) and myeloid leukemia in Down syndrome (ML-DS) are genetically simple models, where samples from two temporally separable, and ascertainable stages in leukemogenesis can be studied. Here, we provide a comprehensive exonic landscape of these two stages. Failure to ascribe function to most TAM variants, besides *GATA1* mutations,

demonstrates the importance of functional interrogation of variants, even in genes recurrently mutated in malignancy. Using a functional genomics approach, we defined genetic lesions acquired in ML-DS that induce transformation providing a basis to develop refined treatment approaches. Our work also exemplifies a general approach to study transformation from a precursor cancerous lesion.

Introduction

As the genetic landscape of precursor lesions and fully transformed cancer is revealed, it is now critically important to functionally annotate the importance of genetic variants during cancer initiation and progression. Childhood leukemia is an excellent tractable model to conduct these functional studies as temporally distinct preleukemic states exist (Hong, 2008; Roberts et al., 2013; Wechsler et al., 2002), allowing preleukemia and leukemia to be studied separately. Furthermore, childhood leukemia is genetically simple with only 1-2 exonic recurrent somatic mutations in the fully transformed state (Andersson et al., 2015; Ma et al., 2018; Network, 2013).

Myeloid leukemia of Down syndrome (ML-DS) is a megakaryoblastic/erythroid leukemia, that occurs with a 150-fold increased risk in children with trisomy 21, despite having a lower incidence of other cancers (Hasle et al., 2000; Lange et al., 1998). Remarkably, ML-DS is preceded by a clonally related, unique, preleukemic stage called transient abnormal myelopoiesis (TAM). Approximately 30% of neonates with Down syndrome acquire N-terminal truncating mutations in the hematopoietic transcription factor GATA1 in fetal hematopoietic stem/progenitor cells (HSPCs) (Roberts et al., 2013; Wechsler et al., 2002) and develop TAM. TAM has a variable presentation that ranges from clinically silent (about half of cases) to clinically overt; and in some cases, life-threatening (Flasinski et al., 2018; Gamis et al., 2011; Klusmann et al., 2008; Roberts et al., 2013). Further, ~20% of children with TAM progress to ML-DS within the first five years of life when persistent *GATA1* mutant cells acquire additional mutations, most frequently in genes encoding members of cohesin protein family, epigenetic regulators and signaling molecules (Nikolaev et al., 2013; Walters et al., 2006; Yoshida et al., 2013).

The goal of our study was to define the mutational landscape of TAM and ML-DS in a large cohort of patient samples and leverage this data to functionally test which genetic variants were functionally important in preleukemic and leukemic transformation.

Results

Genomic landscape of Down syndrome–related myeloid neoplasms

To characterize the exonic mutational landscape during progression from TAM to ML-DS, we studied 252 peripheral blood or bone marrow samples (141 ML-DS and 111 TAM) from 128 ML-DS patients and 103 TAM patients; including 12 paired TAM and ML-DS samples, and 3 paired ML-DS samples at diagnosis and at relapse from chemotherapy (**Table S1**). We used three sequencing methods (**Table S1** and **Figure S1A**): exome sequencing (7 TAM and 13 ML-DS samples, including 3 paired TAM and ML-DS samples), targeted re-sequencing using RNA baits covering 111 genes and chromosomal changes in myeloid malignancies (85 TAM and 107 ML-DS samples; **Table S1** and **Figure S1B**), and amplicon-based sequencing of 62 genes (53 TAM and 58 ML-DS samples; **Table S1** and **Figure S1C**). Four samples were sequenced by all three methods and 67 samples by combinations of two methods (**Table S1**), which provided validation of each method.

In addition to a truncating *GATA1* variant, which was present in all samples, 1.6 single-nucleotide variants (SNVs) and small nucleotide insertions or deletions (indels) were detected in ML-DS cases, which was significantly more than in TAM cases (0.4 variants per sample; $p=1.48x10^{-7}$; **Figure 1** and **Table S1**). In ML-DS the most frequently acquired variants were either in genes encoding the cohesin complex, including *STAG2*, *RAD21*, *SMC1A*, *SMC3, CTCF* and *NIPBL* (n=66; 47% of samples) or in JAK family kinases, *MPL* and *KIT* (n=68; 48% of samples). There was also a high frequency of variants in epigenetic regulators (n=51; 36% of samples), including *KANSL1, EZH2*, and *SUZ12* and a further 14% of samples (n=20) had variants in RAS family members. Though much of the landscape of acquired genetic changes is similar to that described previously (Walters et al., 2006; Yoshida et al., 2013), a hotspot mutation (A455D/T) in *CSF2RB*, encoding the common beta chain of the IL-3, IL-5 and GM-CSF cytokine receptors, was present in 6 ML-DS cases (4.7%; 5 cases with A455D and 1 case with A455T; **Table S1** and see below). Interestingly, we did not detect this mutation in a previously published RNA-sequencing dataset of juvenile myelomonocytic leukemia (Sakaguchi et al., 2013), which is characterized by constitutive GM-CSF signaling.

Though variants in addition to those in *GATA1* were detected in most ML-DS samples, in 25.5% (n=36) of ML-DS samples no additional SNVs/indels, chromosomal aberrations or fusion transcripts were identified. In two samples with sufficient RNA for sequencing, no fusion transcripts were identified (data not shown). Conversely, though most TAM samples did not contain variants besides those in *GATA1*, in 23% of samples (n=25), variants were detected in genes recurrently mutated in myeloid malignancies, raising the question of whether these were functional or not.

Analyzing the co-mutational pattern in ML-DS samples revealed significant co-occurrence of ten different combinations of copy number changes and variants in classes of genes (**Figure S2A** and **S2B**). By way of example, there was a significant co-occurrence of variants in genes encoding tyrosine kinases and RAS proteins with variants in epigenetic regulators ($p=5.28x10^{-8}$) or cohesin genes ($p=1.26x10^{-6}$).

Overall, our analysis depicts a mutational landscape of TAM to ML-DS transformation characterized by the potential cooperation between activated signaling and deregulation of epigenetic processes.

Clonal architecture, evolution and driver mutations of ML-DS

Taking advantage of paired and sequential samples, we next asked how clones evolved during disease progression. We monitored the variant allele frequencies (VAFs) of *GATA1* variants (hereafter *GATA1s*) and co-occurring variants in paired samples from six patients (**Figure 2A-2C**). In four cases (#55, #66, #4 and #6), two or more *GATA1* clones were detected, with minor clones dominating at transformation to ML-DS (#55, #4 and #6). Clonal diversification with multiple additional oncogenic lesions occurred in most patients (#55, #66, #3, #4 and #6) during progression or ML-DS transformation. These additional non-*GATA1* variants were not detected early in the TAM phase, even using highly sensitive droplet digital PCR (ddPCR) (data not shown). Only in one patient (#5) was a *SMC1A* mutation detectable in late TAM (day 31) and only with ddPCR (**Figure S2C**).

To gain insight into the putative function of additional non-*GATA1* recurrent variants in transformation, we asked which functional categories of genes contained variants that were

either clonal or subclonal, with respect to the *GATA1* mutation VAF, which marks the size of the ML-DS/TAM clone (**Figure 2D** and **2E**) in 68 ML-DS and 24 TAM samples. Variants were clonal in 72% of ML-DS cases (n=49), exemplified by cases #217 and #214 (**Figure 2D**). Subclonal variants were detected in 28% of ML-DS cases (n=19/68; 19.1% [n=17/87] at diagnosis (**Table S1**), exemplified by cases #211, #203, #205 and #215 (**Figure 2D**). In contrast, non-*GATA1* subclonal variants in TAM samples were present at a greater level in 83% of samples (n=20; 82% at diagnosis [n=18]; **Table S1**). Given the results, we wanted to clarify if variants in specific functional gene categories were more, or less, likely to be clonal in ML-DS compared to TAM (**Figure 2E** and **Table S2**). In ML-DS, variants leading to loss of function in genes encoding cohesin core subunits, *CTCF*, *NIPBL*, *JAK1-3*, *SRSF2* and *SF3B1,* were likely to be clonal, as were annotated gain of function variants within the tyrosine kinase class (*JAK1*, *JAK2*, *JAK3*). Of note, the A455D *CSF2RB* variant was clonal in all ML-DS cases. In one patient, the *CSF2RB* A455D variant co-occurred with activating mutations in *JAK2* (V617F). However, single cell DNA sequencing revealed that the *JAK2* mutation was not present in the dominant *CSF2RB* mutant clone (*CSF2RB* A455D in 122 out of 130 evaluable cells), but instead was present in a small subclone (*JAK2* V617F in 5 out of 92 evaluable cells) (**Table S2**).

Assigning functions to acquired mutations

Based on the observation of subclonal diversity and proportions, we asked if the predicted functional consequences of the variants in TAM and ML-DS differed between the different functional gene categories (**Figure 3A** and **Figure S3A**). In total 85/114 (75%) variants in ML-DS, in cohesin components (**Figure 3B**), *SH2B3* (**Figure 3C**), and epigenetic regulators (**Figure 3D**) were frameshift or nonsense variants or large deletions. In contrast, in TAM these likely loss-of-function of variants were only detected in 3/20 (15%) cases ($p=1.142x10^{-16}$). *SRSF2* variants were a mixture of missense variants and deletions (**Figure 3E**). *SRSF2* variants are common in myelodysplastic syndromes and secondary AML, where they have been shown to alter function. Mutations in *RUNX1*, *KIT*, *KMT2A/C* and *TET2* were missense mutations of unknown function.

In contrast, variants with documented gain-of-function (either in previous studies or in this study) were detected in JAK kinases (31/41 variants, 76%) (**Figure 3F**), *MPL* (9/10 variants,

90%) (**Figure 3G**), *CSF2RB* (6/7 variants, 90%) (**Figure 3G**), and RAS family members (14/16 variants, 88%) (**Figure 3H**). These include known gain-of-function variants described in myeloproliferative neoplasms including: *JAK2* V617F (Baxter et al., 2005; Campbell et al., 2005; James et al., 2005; Levine et al., 2005); *JAK2* M535I (Malinge et al., 2008) and *JAK3* A572V and A573V in acute myeloid leukemia (AML) (Kiyoi et al., 2007); as well as *JAK2* R683G as in DS acute lymphoblastic leukemia (Bercovich et al., 2008).

None of the *JAK1, JAK2, JAK3* or *MPL* variants identified at the TAM stage were activating in dual luciferase assays with a STAT5 reporter (**Figure S3B-S3I**). Three out of five novel mutations in *JAK2* (*JAK2* p.L153fs*13, p.W157* and K857SD) were found in TAM cases (**Figure 3F**) and resulted in a premature stop codon. No autonomous or cytokine-induced signaling was observed for two of these mutants (**Figure S3D**), as expected with lack of the kinase domain. No synergistic effects were noted with the co-expression of another novel truncated *JAK2* variant (p.R433*), and the well-described activating mutant *JAK3* p.M511I, which co-occurred in one ML-DS case (**Figure S3I).** Of the six novel *JAK3* variants, two (del561-3 and P906H) had demonstrable autonomous activity and were present in ML-DS (**Figure S3E-S3G)**. The most common *MPL* variant was S505N (in 7 out of 12 *MPL*-mutated cases), a well-described mutant previously shown to be auto-activating in the context of familial thrombocythemia (Ding et al., 2004). In contrast, previously un-reported *MPL* variants I492T and I492V, detected at the ML-DS and TAM stages, respectively, did not cause autonomous receptor activity (**Figure S3B**). In most ML-DS cases with non-activating *JAK* or *MPL* mutations, an additional mutation, known or presumed to be activating, was identified in the same patient (**Figure 1** and **Table S1**). In summary, we did not find other types of mutations unique to ML-DS in these genes. Failure to ascribe function to most TAM variants, besides *GATA1* mutations, demonstrates the importance of functional interrogation of variants, even in genes recurrently mutated in malignancy.

CSF2RB variant A455D is an oncogenic hotspot mutation in ML-DS

One striking finding from the genetic analyses was a recurrent, previously undocumented, variant A455D in *CSF2RB* in ML-DS. CSF2RB complexes with receptor-specific α chains and cytokines to trigger signaling through JAK kinases and downstream STAT, PI3K-AKT-mTOR and

MEK/ERK pathways to promote hematopoietic cell survival, proliferation and differentiation (reviewed in (Broughton et al., 2015; Hercus et al., 2013)). Three features suggested that *CSF2RB* hotspot A455D variant could be an important driver in transformation of TAM to ML-DS. First, the A455D variant was detected in 6/128 ML-DS cases (4.7%) (**Figure 3G**) and was clonal (**Figure 2E**). Second, we have previously reported elevated wild-type CSF2RB mRNA expression in ML-DS (Bourquin et al., 2006). Third, substituting the small amino acid alanine with the bulky negatively charged aspartic acid in A455D is predicted to promote pathological dimerization of the transmembrane domains either as a homomeric interaction with another CSF2RB chain or a heteromeric interaction with partner α chains of the IL3, IL5 and GM-CSF receptors (**Figure 4A**). This would result in an active conformation for downstream JAK signaling (Hansen et al., 2008), as demonstrated for the negatively charged bulky variant residue V449E (Jenkins et al., 1995), associated with myeloproliferative disorder in mice (D'Andrea et al., 1998), or for mutations found in the related cytokine receptors, G-CSFR and MPL (Bargmann and Weinberg, 1988; Ding et al., 2004; Maxson et al., 2016; Pikman et al., 2006; Plo et al., 2009; Staerk et al., 2006).

To test if CSF2RB A455D was functional, we first asked if this variant conferred cytokine independent growth in the GM-CSF-dependent (thus GM-CSFRα expressing) human erythroleukemia cell line TF1 (**Figure 4B**). While TF1 cells transduced with CSF2RB WT or empty vector ceased to proliferate, cells transduced with CSF2RB A455D proliferated in cytokine-free conditions accompanied by STAT5 phosphorylation (**Figure 4C** and **Figure S4A**), which was reversed by the JAK kinase inhibitor Ruxolitinib (**Figure 4B** and **4C).** We did not observe phosphorylated AKT (**Figure S4B** and **S4C**), confirming selectivity of CSF2RB A455D for ligand-independent activation of STAT5 via JAK kinases. Autonomous activation of CSF2RB A455D was further validated *in vivo* by transplantation of transduced TF1-cells into sub-lethally irradiated NSG mice (NOD.Cg-*Prkdc*^{scid} *Il2rg*^{tm1Wjl}/SzJ). Mouse cytokines, including GM-CSF, show limited reactivity with corresponding human cytokine receptors (Manz, 2007), thus allowing us to test growth of TF1 cells in effectively a cytokine-deprived *in vivo* environment, where they have previously been shown to grow poorly (Kiser et al., 2001). Mice injected with CSF2RB

A455D-transduced TF1 cells had significantly shorter survival than mice injected with empty vector- (median survival 23 vs 58 days; p=0.0039) or CSF2RB WT-transduced TF1 cells (74 days; p=0.0097; **Figure 4D**).

We next investigated the functional consequences of CSF2RB A455D on human hematopoietic proliferation and differentiation using lentivirally transduced $CD34^+$ HSPCs. Under conditions promoting megakaryocytic differentiation, CSF2RB A455D-transduced $CD34^+$ HSPCs were arrested in differentiation with a 7.5-fold reduction in mature $CD41^+/CD42b^+$ megakaryocytes compared to the empty vector control (p=0.0062) and 6.4-fold reduction compared to CSF2RB WT (p=0.005; **Figure 4E** and **4F**). The percentage of immature $CD41^+/CD42b^-$ megakaryocytic cells was similarly reduced by CSF2RB A455D (**Figure S4D**). The block of megakaryocytic differentiation was accompanied by a 30-fold expansion of mainly $CD117^+CD41^-CD33^+CD71^+CD235a^+$ immature erythroid progenitor cells after 18 days of culture over CSF2RB WT or the empty vector control (p=0.046; **Figure 4E** and **4G** and **Figure S4E**). In conditions supporting both erythroid and megakaryocytic differentiation, we confirmed that mutant CS2RB A455D promoted the unexpected preferential growth of erythroid cells (**Figure S4F-S4H**). Importantly, the expansion of immature erythroid progenitor cells induced by CSF2RB A455D during megakaryocytic differentiation could be reversed by Ruxolitinib (**Figure 4E**), whilst leaving empty vector-transduced cells mainly unperturbed (**Figure S4I**).

In contrast, under myeloid differentiation conditions there was a smaller enhancement of myeloid proliferation and no change in expression of myeloid markers CD15, CD66b, CD14, and CD11b (**Figure S4J-S4L**). Of note, we did not see cytokine-independent proliferation when murine cells (BA/F3 cell line and primary murine fetal liver cells) were transduced with CSF2RB A455D (data not shown). Though several reasons may account for this, one explanation could be species-specific differences in CSF2RB oncogenic signaling.

Taken together, our data provide compelling evidence that CSF2RB A455D has an oncogenic function in ML-DS by promoting megakaryocytic erythroid proliferation with an attendant block in terminal megakaryocyte maturation.

CRISPR-Cas9 loss-of-function screen identifies transforming mutations in ML-DS in a murine fetal liver model of TAM

To obtain a more comprehensive understanding of the transforming landscape of TAM to ML-DS transition, we performed a multiplex *in vivo* loss-of-function screening in a murine model closely resembling TAM (**Figure 5A**). We transduced a sgRNA targeting *Gata1* exon 2 (sgRNA G1.4; hereafter sgGata1s) – for which we verified optimal on-target editing efficacy and absence of indels/SNVs at the top five predicted off-target sites in the murine genome (**Figure S5A** and **Table S3**) – into embryonic day 13.5 (E13.5) disomic fetal liver cells (FLCs) of *Cas9*-knock-in mice (*Rosa26:Cas9−EGFP*$^{ki/WT}$) (Platt et al., 2014). In this way, we introduced the *Gata1* mutation (hereafter *Gata1s*) leading to exclusive GATA1s protein expression (**Figure S5B**) in the right developmental context, since TAM initiates in fetal cells. We used disomic murine background, as opposed to the three existing DS mouse models, *Tc1, Ts65Dn* and *Ts1Cje*, because we have previously shown that these mice do not develop TAM or ML-DS and they do not provide an appropriate background to test cooperation between *Gata1s* and ML-DS transforming genes (Alford et al., 2010; Klusmann et al., 2010a). sgGata1s induced a hyperproliferative phenotype in FLCs (**Figure S5B** and **S5C**) that was comparable to the previously reported *Gata1*$^{\Delta e2}$ model (Li et al., 2005) (**Figure S5C)** and the phenotype produced by an alternative sgGata1s (sgRNA G1.5; **Figure S5B** and **S5C)**. Importantly, upon transplantation into sub-lethally irradiated syngenic recipients (C57Bl/6j) *Gata1s*-FLCs were abundant in the peripheral blood at weeks 2 to 4 but disappeared over time (**Figure S5D**).

Thus, this model provided us with fetal GATA1s-expressing cell populations that produced a transient, developmental stage appropriate, *in vivo* myeloproliferative disorder; a platform to test the impact of predicted loss-of-function variants for transformation to ML-DS. 22 genes were selected for a multiplex CRISPR-Cas9 *in vivo* screen based on the frequency of indels/SNVs in ML-DS samples and taking their function, class, family and location into consideration to avoid over- or underrepresentation (**Table S3**). The selected genes comprised members of the cohesin complex (*Rad21*, *Stag2*, *Nipbl*, *Ctcf*, *Smc1a*), epigenetic regulators (*Kdm6a*, *Ezh2*, *Asxl1*, *Bcor*, *Setbp1*, *Kmt2a/b/e*, *Kansl1* and *Phf8*) and transcription factors (*Trp53*, *Runx1*

[located on chromosome 21]). We also selected negative regulators of the RAS (*Nf1*) and JAK-STAT signaling pathways (*Cbl*, *Sh2b3 or Rnf41*) to mimic autonomous signaling activation, which was detected in 48% of ML-DS samples.

A pre-tested pooled set of sgRNAs (**Table S3)**, with uniform high cutting efficiency (>70%) in HEL cells (data not shown), was introduced into *Gata1s*-FLCs to create multiplex loss of function of the 22 genes and we monitored their potential to induce leukemia *in vivo*. Based on patterns of mutation association, we also compiled 5 pools, containing only 8 or 15 sgRNAs (**Table S3**). Mutations or deletions in the tumor suppressor *TP53* are found in ~40% of human tumors that tend to be more aggressive (Bailey et al., 2018) and *Trp53*$^{-/-}$ mice show spontaneous development of cancer (Donehower et al., 1992). Therefore, we decided to additionally test a pool of 21 sgRNAs against 21 genes, but excluding *Trp53*, as well as individually targeting *Trp53*.

Recipient mice (n=25 for the entire pool of 22 genes) developed leukemia with full penetrance and short latency (median survival 36 days; **Figure 5B** and **Figure S5E**). 12/16 primary leukemias tested engrafted into secondary recipients, confirming a fully transformed phenotype (**Figure S5F**). Mice transplanted with pools lacking *sgTrp53* developed leukemias with a latency of 47.5 days (median survival; $p<0.0001$ compared to the entire pool) and a penetrance of 91% (n=20/22; **Figure 5B**). Conversely, targeting of *Trp53* alone was also sufficient to induce leukemia with 100% penetrance and a latency of 41.5 days (median survival; $p=0.005$ compared to entire pool; **Figure 5B**). This underlines the strong oncogenic potential of loss of *Trp53,* but also of the other recurrently mutated genes in ML-DS, in cooperation with *Gata1s*.

Importantly, in mice transduced with pools of sgRNAs omitting sgRNA directed against *Trp53*, we neither detected changed levels of *Trp53* mRNA expression compared to normal murine stem/progenitor cells (**Figure S5G**), nor alterations in *Trp53* cDNA in 6/6 samples tested by Sanger sequencing (data not shown), consistent with intact p53 function in leukemias generated from pooled approaches lacking the sgRNA directed against *Trp53*.

Two out of 14 mice transplanted with *Gata1s*-FLCs transduced with a control non-targeting control sgRNA developed leukemia (**Figure 5B**). As a further control, we transplanted 5

recipients with wild-type *Gata1*-FLCs transduced with the pool of sgRNAs (omitting *sgTrp53*) (**Figure S5H**). This resulted in a stable engraftment (**Figure S5I**). Notably, only 1/5 mice harboring sgRNAs against *Bcor*, *Cbl* and *Nipbl* (data not shown) developed leukemia but with a much longer latency of 88 days. Taken together, this suggests that genetic lesions engineered by the sgRNA pool require *Gata1s* in FLCs for initiation of high penetrance and short latency murine leukemias.

Detailed characterization of all leukemias highlighted an immature erythroid phenotype (dual $CD117^{+}Ter119^{+}$) with a minority of megakaryoblastic ($CD41a^{+}CD117^{+}$) and megakaryocytic ($CD41a^{+}$) cells with virtual absence of myeloid (CD11b) and lymphoid (CD3 and CD19) cell surface marker expression (**Figure 5C** and **5D**). ML-DS is typically a megakaryocyte/erythroid leukemia and the erythroid bias here may reflect the preponderance of erythroid cells in the fetal liver. Furthermore, leukemic mice presented with hepatomegaly and splenomegaly due to infiltration of leukemic cells in the liver and spleen as well as bone marrow (**Figure S5J** and **S5K**).

To determine more globally the lineage progenitor affiliation of the experimental murine leukemias, we compared RNA-seq profiles from six representative murine leukemias with RNA-seq profiles of FACS purified murine fetal liver stem/early progenitor populations. Unsupervised hierarchical clustering of the 1,613 most variable genes demonstrated that the mouse leukemias clustered most closely with megakaryocyte-erythroid progenitors (MEP; $Lin^{-}Sca1^{-}cKit^{+}CD34^{-}FcγR^{-}$) rather than early stem/early progenitors ($Lin^{-}Sca1^{+}cKit^{+}$ [LSK]), common myeloid progenitors (CMP; $Lin^{-}Sca1^{-}cKit^{+}CD34^{+}FcγR^{low}$) and granulocyte-monocyte progenitors (GMP; $Lin^{-}Sca1^{-}cKit^{+}CD34^{+}FcγR^{+}$) (**Figure 5E**).

Furthermore, we asked which human leukemias shared the closest similarities with the murine leukemias we had experimentally generated. Gene set enrichment analyses (GSEA) showed that the murine leukemias were positively enriched for gene signatures upregulated in human ML-DS and negatively enriched for genes expression signatures downregulated in human ML-DS (**Figure 5F** and **Table S3**) (Bourquin et al., 2006; Schwarzer et al., 2017). Finally, gene expression signatures of other pediatric AMLs – for example, those with t(8;21), *MLL*-

rearrangement or non-DS-AMKL – were negatively enriched in the gene expression profiles from the murine experimental leukemias, underlining the specificity of the human ML-DS-like expression profile in the murine leukemias (**Figure 5F**).

In summary, our multiplex CRISPR-Cas9 loss of function screening approach transformed murine *Gata1s*-FLCs to erythroid/megakaryoblastic leukemic blasts with an RNA expression profile similar to the one observed in human ML-DS.

Murine ML-DS-like leukemias share mutational spectrum of ML-DS samples

Having established ML-DS-like leukemias in 50 mice, we performed amplicon re-sequencing to identify the genomic loci targeted by sgRNAs in 38 mice, where DNA was available (**Figure 6A** and **Table S4**). On average, 2.7 mutations were detected per leukemia in addition to *Gata1s*, ranging from only one mutation to up to five per sample (**Figure S6A** and **Table S4**). Indels and SNVs constituted 96.2% and 3.8% of variants introduced by sgRNAs/Cas9, respectively, which were present in 18/22 genes targeted. Importantly, mutations in 7 genes accounted for 76% of all variants detected (79/104; **Figure 6A**). 61% (n=11/18) of the sequenced murine ML-DS-like leukemias that were originally transduced with the sgTrp53-containing pools (n=18) harbored mutations in *Trp53*. Incomplete penetrance of *Trp53* mutations in the sgTrp53-containing groups and significantly prolonged survival of leukemias only harboring *Trp53* mutations (**Figure S6B**) compared to *Trp53* co-mutated samples (median survival 41.5 days vs 35 days, p=0.01), underlined the importance of combinations of driver mutations mediating ML-DS transformation. Detailed mutation analysis revealed frequent targeting of RAS and JAK-STAT signaling cascade and epigenetic modifiers (**Figure 6A**), a mutational landscape that has similarities to human ML-DS (**Figure 6B**), but interestingly with reduced frequency of cohesin and *Ctcf* mutations ($p<0.0001$). In 74% of the leukemic mice, mutations in *Nf1*, *Cbl* or *Sh2b3* were detected (**Figure 6A**). *Nf1* was the most frequently altered single gene in the samples. Loss-of-function mutations in epigenetic regulators, *Ezh2* (member of the Polycomb Repressive Complex 2, PRC2), *Asxl1* (polycomb-associated protein), *Kdm6a* (histone lysine demethylase), *Bcor* (Polycomb

Repressive Complex 1, PRC1) and other epigenetic modifiers were found in 87% of leukemia samples. Consistent with this, gene set enrichment analysis of the RNA profiles of murine ML-DS-like leukemias revealed activation of genes normally repressed by the PRC2 and positively regulated by the histone demethylase LSD1 (**Figure S6C**), suggesting that de-repression of PRC2-target genes could be an important step for transition from TAM to ML-DS. Supporting this hypothesis, *in vitro* cell growth of leukemic blasts from two patients with ML-DS was retarded by the LSD1 inhibitor T-3775440 (Ishikawa et al., 2017) with median lethal concentration (LC_{50}) values of 9 nM and 20 nM, respectively, while leukemic blasts from one AML patient with MLL-rearrangement (MLLr) showed no response, and inhibition of normal $CD34^+$ HSPCs growth had LC_{50} values greater than 10 μM (**Figure S6D**).

Probing mutational cooperation, we found co-occurrence of variants in *Cbl* and *Ezh2* (p=0.059; **Figure S6E**). Conversely, mutations of *Nf1* with mutations in *Cbl*, *Ezh2* and *Ctcf* showed a negative correlation (p=0.156, p=0.024 and p=0.099, respectively; **Figure S6E**). Mutations in *Sh2b3* and *Nf1* were detected in four leukemias. Single cell sequencing of one of the leukemias verified that these mutations were in the same cells, implying functional cooperativity between RAS and JAK-STAT signaling activation (**Table S4**), but it is worth noting that no patient acquired both *SH2B3* and *RAS* gene variants. Taken together, ML-DS-like mouse leukemias share aspects of the mutational spectrum and mutation co-occurrence seen in ML-DS samples. Overall, these functional data validate 82% (18/22) of genes with putative loss-of-function variants in human ML-DS samples cooperate with *Gata1s* to transform fetal hematopoietic cells.

Discussion

Here, we present a study of genomic variants in clinical samples taken from a total 252 TAM and ML-DS cases that guided functional analysis. We delineated cooperativity between trisomy 21, *GATA1s* and secondary transforming events. One previously undescribed transforming event is the activating hotspot mutation in *CSFR2B.* More globally, we functionally validated loss of

function of 18 genes cooperates with GATA1 mutants in oncogenic transformation in developmentally appropriate, primary hematopoietic cells.

In most cases *GATA1s* with trisomy 21 is sufficient to result in TAM. Exonic variants in TAM were infrequent, mainly subclonal and we failed to assign function to most JAK and *MPL* TAM variants. In contrast, in most cases, variants in ML-DS had either known or predicted functional consequence or we could experimentally validate their oncogenic function. Although we cannot formally rule out rare germline mutations in at least in some patients, such as in *KANSL1* (Zhou et al., 2017), most transforming events were somatically acquired.

Examination of paired sequential TAM and ML-DS samples was clinically informative. First, transforming variants were usually not detectable at birth; only in one out of six cases, was a transforming variant detectable. Thus, surveying neonates with TAM at birth for additional transforming mutations may not be a good way to identify infants at risk of ML-DS. Persistence of a *GATA1* mutation beyond 3 months is a better predictor of progression to ML-DS ((Flasinski et al., 2018) and Roberts and Vyas unpublished observations). Moreover, our data imply that monitoring the dominant *GATA1s* TAM clone alone is not a clinically appropriate strategy for residual disease monitoring.

There are six previous reports of *CSF2RB* variants in hematopoietic malignancies in lymphoma, (https://cancer.sanger.ac.uk/cosmic/), and one germline variant (R461C) in pediatric T cell acute lymphoblastic leukemia (Watanabe-Smith et al., 2016). These variants have not been functionally characterized. The *CS2RB* A455D variant causes ligand-independent STAT5 phosphorylation, promoting cytokine independent growth abrogated by the JAK kinase inhibitor Ruxolitinib, providing a potential therapeutic option to ML-DS patients. Interestingly, onset of *CSF2RB* expression marks a prospectively purifiable primary human CD131^{+} (the CD marker for CSF2RB) *GATA1*-expressing cell population within the CMP compartment with megakaryocyte, erythroid, mast cell and basophil potential at a clonal level (Drissen et al., 2019). This raises the hypothesis that *GATA1*s expressing CD131^{+} cells within the CMP fraction could be a cell of origin of transformation to ML-DS in those patients that acquire the *CSF2RB A455D* mutation. Concordantly, *CSF2RB* is one of the most highly expressed RNAs in ML-DS (Bourquin et al.,

2006; Schwarzer et al., 2017) and ML-DS cells can have eosinophil (Maroz et al., 2014) as well as megakaryocyte and erythroid gene expression programs (Bourquin et al., 2006).

Our murine *in vivo* transplantation model reconstitutes key aspects of TAM: fetal developmental stage, *Gata1s*-expression and a transient *in vivo* myeloproliferative phenotype. With this model, loss-of-function of 18 of 22 selected genes led to a rapid ML-DS-like transformation providing compelling support for cooperation with *Gata1s*. One notable feature of our model is the absence of a trisomic background for genes on human chromosome 21. We used the disomic murine background, as opposed to three existing DS mouse models, *Tc1, Ts65Dn* and *Ts1Cje*, because we have previously shown that these do not provide an appropriate background to test cooperation between *Gata1s* and ML-DS transforming genes (Alford et al., 2010; Klusmann et al., 2010a). However, the lack of a trisomic background could be one explanation for the reduced frequency of loss-of-function mutations of cohesin family members in the murine leukemias. If true, intact dosage of cohesin function may be especially important in preventing leukemia in trisomic cells. Increased transcription factor binding site accessibility and expression of trisomic genes *ERG* and *RUNX1* in cohesin mutant cells supports this hypothesis (Mazumdar et al., 2015). Alternatively, cohesin mutations may just not read out in murine *Gata1s*-cells, and/or in a syngeneic, pre-leukemic cell transplantation model; though this not the case for all hematologic cancers (Heckl et al. unpublished data). Thus, analysis of loss of cohesin function in ML-DS still requires further work in addition to that published (Mullenders et al., 2015; Viny et al., 2015) but in GATA1s trisomic human cells with appropriate haploinsufficiency of cohesin genes.

Loss-of-function mutations in PRC regulators in patient samples and in murine leukemias point towards an important role of the PRC function in ML-DS transformation. Notably, mutations in epigenetic factors are rarely found in other pediatric AML samples (Bolouri et al., 2018); and even though they are more common in adult AML, they usually are early events rather than secondary events associated with transformation (Corces-Zimmerman et al., 2014). Hence, the order of events is reversed in ML-DS compared to adult AML underlining the distinct pathogenesis of ML-DS. Although the observation that leukemic blasts from two patients with

ML-DS were sensitive against LSD1-inhibitor T-3775440 (Ishikawa et al., 2017) requires validation, it potentially opens therapeutic inhibition of LSD1 as a potentially less toxic therapy in children with ML-DS who often have more toxicity with conventional chemotherapy (Uffmann et al., 2017).

In summary, by genetically characterizing the mutational landscape of TAM and ML-DS and functionally dissecting variants required for transformation, the field is now well-placed to perform detailed mechanistic studies of oncogenic cooperativity that will necessary for novel, rational therapeutic approaches for this leukemia.

Acknowledgements

We thank D. Trono of EPFL, Switzerland for pMD2.G (Addgene plasmid 12259) psPAX2 (Addgene plasmid 12260). The following funding is acknowledged: J.H.K by ERC via European Union Horizon 2020 (No 714226) & St. Baldrick's Robert Arceci Innovations Award; D.H by German Cancer Aid (111743); J.H.K. and D.H. by DFG (KL-2374/1-3, HE-7482/1-1); M.L. by Hannover Biomedical Research School. P.V. and I.R. by Bloodwise Specialist Programme Grant 13001, NIHR Oxford Biomedical Research Fund. P.V. by MRC MHU (MC_UU_12009/11); C.G. by Wellcome Trust Clinical Training Fellowship; L.V. by Mouve-in Louvain and Haas-Teichen Postdoctoral Fellowships; S.N.C. by WelBio F 44/8/5-MCF/UIG. 10955, Action de Recherche Concertée 16/21-073, Salus Sanguinis, Foundation "Les avions de Sébastien" and Fondation contre le cancer, Belgium.

Author Contributions

ML, KP, ND, CG, RB, DS performed experiments, analyzed data, contributed to editing. EP, SM, VA, TR, VI and AS analyzed data contributed to editing. MLY and PC supervised analyses, edited the manuscript. KY, JT, JDC, MJW, AH, TT EI, SO and DR provided patient material, data and edited the manuscript. LV and CS performed experiments, analyzed data and edited the manuscript. IR, SC, PV, DH, JHK designed the study, analyzed data, wrote the manuscript and directed the project.

Declaration of Interests

D.R. has consulting/advisory roles for Roche, Celgene, Hexal, Pfizer, Novartis, Boehringer and receives Celgene research funding. D.R. received travel, accommodation, expenses from Jazz Pharmaceuticals and Griffols. J.D.C. receives research funding from Scholar Rock and Forma Therapeutics. All other others have nothing to declare.

Figure Legends

Figure 1. Acquired genetic variant landscape in ML-DS and TAM.

Sequencing data from 141 ML-DS samples and 111 TAM samples. Samples are arranged in columns with genes labeled and grouped in rows. Deletion of chromosome 5q (del5q) is indicated. Samples subject to exome sequencing are shown at the bottom. Total number of variants per gene is indicated on the right. Where more than one variant was detected in a sample, the colored boxes are stacked rectangles. The predicted nature and function of variant is color-coded. The code is shown at the bottom of the figure. Tr. factor; transcription factor. The p value is calculated by 2-tailed paired Student's t-test relative to non *GATA1* variants present in ML-DS versus TAM cases.

See also **Figures S1 and S2** and **Table S1.**

Figure 2. Clonal and subclonal evolution in ML-DS and TAM.

(A-C) Plots of variant allele frequency (VAF) adjusted to copy number in sequential samples (day or month of life from birth) from patients with TAM that progressed (prog) within 3 months of birth. **(A)**; with TAM that transformed to ML-DS **(B)**; and a patient with TAM who progressed to ML-DS, relapsed after chemotherapy and died **(C)**. Circles, mean VAF; error lines represent SD. Left of (**A**) color-coded key of classes of mutations.

(D) VAF for six representative ML-DS cases. Percentage of blasts in the sample is shown on top. Circles; mean VAF; vertical bars; 95% confidence intervals (CI). P values are from 2-tailed paired Student's t-test relative to *GATA1*.

(E) Bean-plot of variant distribution relative to *GATA1* VAF. Grey horizontal lines indicate mean; black or white smaller horizontal lines represent individual data points; polygons represent

estimated data density. P values ($p<0.05$) were obtained from 2-tailed paired Student's t-test. Abbreviations: TK; tyrosine kinase, Tr.factor; transcription factor.

See also **Figure S2** and **Tables S1** and **S2**.

Figure 3. Predicted functional nature and distribution of variants within more commonly mutated genes in TAM and ML-DS.

(A) Column graph of frequency and type of non-*GATA1* variants in ML-DS and TAM samples for the indicated genes. Predicted functional consequence of variant is color-coded; legend on the right. Values within columns show number of variants annotated within each functional category. Large deletion refers to entire gene loss.

(B-H) Schematic of amino acid position of variants, predicted amino acid changes of common variants in TAM and ML-DS divided by functional categories: cohesin family members **(B)**, SH2B3 adaptor protein **(C)**, epigenetic regulators **(D)**, SRSF2 splicing factor protein **(E),** JAK kinase family members **(F)**, cytokine receptors including CSFR2B **(G)**, and RAS family members **(H)**. Protein binding recognition regions in black, conserved domain motifs in grey. Position of TAM (red arrowhead) and ML-DS (black arrowhead) variants are shown. Previously undocumented variants blue text; previously documented variants black text. Positions of sgRNAs used in CRISPR-Cas9 experiments are in pink font. * = stop codon, fs = frameshift, SA = splice acceptor variant, SD = splice donor variant. Cytogenetic abnormalities leading to loss of heterozygosity are denoted. Functionally validated gain of function variants in this or other studies denoted by green circle. ATPase: ATP-binding Walker A (N-terminal) and B (C-terminal); SCD: STAG (Stromal Antigen) Conserved Domain; PDS5B/WAPAL: Sister chromatid cohesion protein PDS5 homolog B/Wings Apart-Like binding sites; SA1/2: STAG1/2 binding sites; FERM: Four.1-Ezrin-Radixin-Moesin plasma membrane localization motif; ps.kinase JH1; pseudo-kinase JH1; SH2: Src Homology 2 domain; KAT8; lysine acetyltransferase 8 domain; HAT; histone acetyltransferase domain; trans; transmembrane domain.

See also **Figure S3** and **Table S1**.

Figure 4. CSF2RB A455D confers cytokine-independent JAK-STAT signaling and impairs megakaryocytic differentiation.

(A) Model of common receptor beta chain encoded by *CSF2RB* together with possible alpha chains. The A455D variant is in transmembrane domain.

(B) Ratio of transduced (CSF2RB A445D, CSF2RB WT and empty vector) to un-transduced TF1-cells cultured in the presence (upper panel) or absence (lower panel) of GM-CSF normalized to day 0 (n=3). Additionally, CSF2RB A455D expressing cells were treated with the JAK2-inhibitor Ruxolitinib (1 μM). Data are presented as mean ± SEM. *p ≤ 0.05; **p ≤ 0.01 (unpaired t-test)

(C) Representative flow cytometry (FACS) histogram (left) and statistics (right) of phosphorylated STAT5 (pSTAT5) in transduced TF1 cells without GM-CSF stimulus (n=3). Horizontal bars represent means; **p ≤ 0.01 (unpaired t-test)

(D) Kaplan-Meier survival curve of NSG-mice transplanted with CSF2RB A445D, CSF2RB WT or empty vector transduced TF1 cells. **p ≤ 0.01 (log-rank-test compared to CSF2RB A445D group)

(E) Representative FACS plots of transduced human cord blood-derived $CD34^+$ HSPCs grown in megakaryocytic differentiation medium with and without the JAK2 inhibitor Ruxolitinib after 12 days of culture.

(F) Percentage of mature megakaryocytic cells ($CD41^+CD42b^+$) as depicted in **(E,** first row**)** normalized to empty vector transduced control cells. Horizontal bars represent means; **p ≤ 0.01 (unpaired t-test and one-sample t-test for normalized data)

(G) Cell count of human HSPCs grown in megakaryocytic differentiation medium normalized to day 0. Data are presented as mean ± SEM. *p ≤ 0.05 (unpaired t-test)

See also **Figure S4**.

Figure 5. CRISPR-Cas9 loss-of-function *Gata1s* mutations in murine FLCs generate ML-DS-like leukemias.

(A) Schematic of *in vivo* multiplexed CRISPR-Cas9 approach to model and progression of TAM to ML-DS.

(B) Kaplan-Meier survival curve of recipients transplanted with *GATA1s*-expressing FLCs transduced with sgRNA libraries. **p ≤ 0.01 (log-rank-test).

(C) Representative FACS plots of bone marrow (BM) derived leukemic cells.

(D) Percentage of BM leukemic cells in *Gata1s*-FLCs. Monocytes: $CD11b^+/Gr1$, granulocytes: $CD11b^+/Gr1^+$. Data are presented as means ± SD.

(E) Heat map and unsupervised hierarchical clustering of RNA-seq data, using the 1613 most variable genes (standard deviation >1) across six murine bone marrow samples from mice with leukemia and FACS-sorted normal murine fetal liver LSKs, CMPs, GMPs, and MEPs.

(F) Bar graph showing normalized enrichment scores (NES) from GSEA of significantly up- or down-regulated gene sets from previously derived ML-DS and other signatures (Bourquin et al., 2006; Schwarzer et al., 2017) in the murine leukemia samples compared to normal progenitor populations. *FDR q value < 0.25, **FDR q value < 0.05.

See also **Figure S5** and **Table S3**

Figure 6. CRISPR-Cas9 *in vivo* screening reveals driving mutations in ML-DS.

(A) 104 nonsense/missense, frameshift and in frame mutations or insertions/deletions (indels) detected in 38 mice with leukemia where sufficient material was available (out of 50 leukemic mice). Mice are arranged in columns, genes labeled and grouped along rows.

(B) Pie charts depicting relative frequency of mutations in functional categories of genes in TAM/ML-DS patient cohort in this paper (left) and mouse cohort (right). Percentages and absolute numbers of cases are shown.

See also **Figure S6** and **Table S4**.

STAR Methods

KEY RESOURCES TABLE – provided as a separate document

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources should be directed to and will be fulfilled by Paresh Vyas (paresh.vyas@imm.ox.ac.uk).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Patients, human samples, karyotype data

Parents or custodians of all patients gave written informed consent in accordance with the Declaration of Helsinki, and study approved by the Thames Valley Research (06MRE12-10; NIHR portfolio no. 6362), Hannover Medical School (#2456-2014) and Martin-Luther-University Halle-Wittenberg (#2018-03) Ethics Committees. Neonates with Transient Abnormal myelopoiesis (TAM) were (i) sourced within 3 months from birth (majority under 14 days of age), (ii) positive for a *GATA1* variant by deep sequencing/conventional Sanger sequencing and (iii) evidence of subsequent clinical and/or *GATA1* mutation resolution. ML-DS samples were allocated if ≥3 months from birth, positive for *GATA1* variant (by criteria above), and ML-DS relapse if post treatment and *GATA1* mutation positive. For sequential samples, “Progressive TAM” refers to samples taken ≤3 months from birth where neonates still had a clinical diagnosis of TAM, irrespective of treatment status. **Table S1** lists karyotype, gender, diagnosis, blast percentage, source tissue and use of paired non-tumor controls for all samples. For the latter, metrics are as follows: (i) tumor and matched non-tumor sample re-sequenced: TAM=55, ML-DS=42, ML-DS relapse=3; n=100 (ii) paired non-tumor sample unavailable: TAM=56, ML-DS=94, ML-DS relapse=2; n=152. Karyotype data was obtained from referring clinicians, the National Down Syndrome Cytogenetic Register, Queen Mary University of London (England) and regional cytogenetic centers (West Scotland and South-east Scotland). This was supplemented by analysis of bait re-sequencing and exome sequencing (Papaemmanuil et al., 2013; Van Loo et al., 2010).

Animal studies

All experiments using mice were performed in accordance with relevant institutional and national guidelines and regulations and were approved by the local authorities (Niedersächsisches Landesamt für Verbraucherschutz und Lebensmittelsicherheit). NOD.Cg-*Prkdc*^{scid} *Il2rg*^{tm1Wjl}/SzJ (NSG mice), B6J.129(Cg)-*Gt(ROSA)26Sor*$^{tm1.1(CAG-cas9*,-EGFP)Fezh}$/J (Cas9 knock-in mice on C57BL/6J background), C57BL/6J mice (Jackson Laboratory, Bar Harbor), and Gata1$^{\Delta e2}$ mice (described earlier (Li et al., 2005)) were maintained in a specific pathogen free environment in individual ventilated cages and fed with autoclaved food and water at Medical School Hannover.

Cell lines and cell culture

All human cell lines were purchased from the German National Resource Center for Biological Material (DSMZ) and cultured in recommended media, typically RPMI medium with 1% streptomycin/penicillin (Sigma-Aldrich), 10% FCS (Capricorn Scientific) and 1% L-Glutamine (Millipore) (HEL and K-562). Medium for the TF1 cell line was additionally supplemented with 5 ng/mL GM-CSF (Peprotech). The adherent cell line 293T was grown in DMEM medium (Gibco) with 10% FCS, 1% L-Glutamine (Millipore) and 1% streptomycin/penicillin. Murine fetal liver cells were cultured and transduced in DMEM medium with 10% FCS, 1% L-Glutamine (Millipore), 1% streptomycin/penicillin, 30 ng/ml Scf and 30 ng/ml Thpo (both Peprotech). *Gata1s*-FLCs were cultures under low cytokine conditions in DMEM medium with 10% FCS, 1% L-Glutamine (Millipore), 1% streptomycin/penicillin, 2 ng/ml Scf and 20 ng/ml Thpo. Cord blood was provided with the parents' consent by the Department of Gynecology and Obstetrics, Hannover Medical School, and experiments were approved by the local ethics committee. Human CD34^{+} HSPCs were cultured in StemSpan SFEM (Stemcell Technologies) with 1% streptomycin/penicillin, 20 ng/ml THPO, 10 ng/ml IL3, 10 ng/ml IL6, 50 ng/ml SCF, and 50 ng/ml FLT3L (all Peprotech).

METHODS DETAILS

DNA extraction and *GATA1* analysis

Genomic DNA and RNA was extracted from patient blood and bone marrow samples (DNeasy Blood and Tissue Kit, Qiagen) and murine leukemias (QIAmp DNA Blood Mini Kit, Qiagen; Quick-RNA Microprep Kit, Zymo Research). Patient samples were tested for *GATA1* mutations in exon2 / partial exon3 by Sanger sequencing and denaturing high-performance liquid chromatography as previously described (Alford et al., 2011; Roberts et al., 2013).

Exome, bait, Amplicon and sgRNA amplicon re-sequencing

Samples subject to each sequencing approach is set out in **Table S1**. A list of genes with accompanying coverage depth is provided in **Table S1** and **Figure S1A**. 13 ML-DS and 7 TAM samples were subject to whole-exome sequencing using Illumina methodology. 107 ML-DS and 86 TAM samples were re-sequenced using RNA bait analysis (SureSelect, Agilent) as described (Papaemmanuil et al., 2013). Genes (n=82) were categorized as follows: “myeloid” - mutation by recurrent somatic mutation (Papaemmanuil et al., 2013) (n=54), “TAM/ML-DS” - mutations from in-house data (n=28) (**Table S1**). 52 TAM, 54 ML-DS, 5 ML-DS relapse samples were re-sequenced using n=1070 150-210 bp amplicons spanning n=62 genes (48.48 Fluidigm Access Array, Fluidigm) (**Table S1** and **Figure S1B**) using manufacturers conditions and sequenced using Illumina methodology.

To quantify the sgRNA frequency in murine leukemias, the sgRNA sequences were amplified from genomic DNA and prepared for sequencing as previously described(Joung et al., 2017). For RNA sequencing, paired-end libraries with 2 × 75 bp reads were prepared from total RNA using the TruSeq Stranded total RNA LT Sample Prep (RiboZero Gold, Illumina) using Illumina methodolgy.

Droplet digital PCR (ddPCR)

ddPCR was performed on a QX200 platform using conditions outlined by the manufacturer (Bio-Rad) and analyzed using QuantaSoft software, version 1.7. Primers and probes were designed as per guidelines (Bio-Rad) (**Table S5**; obtained from Thermo Fisher Scientific and Integrated DNA Techonologies (IDT). Gradient PCR was employed prior to use to determine optimal (1)

primer / probe annealing (2) separation of mutant and wild-type droplets. Samples were quantified using Qubit HS-dsDNA kit (Thermo Fisher Scientific). 40 ng DNA was used per ddPCR reaction (2 ng/µl: ~12,080 Genome Equivalent Units/20,000 droplets). Number of individual reactions for each variant is as follows: TAM (n=16), ML-DS (n=3, with the exception of individual 1; n=16); relapse (n=3), non-template control (n=2), post-treatment (n=8) and wild-type controls (n=8).

Lentivirus production and transduction

Lentiviral particles were generated by co-transfecting 293T cells with the viral constructs, pMD2.G and psPAX2 (Addgene #12259 and #12260) using the polyethylenimin-transfection method as previously described (Bhayadia et al., 2018; Emmrich et al., 2014). A list of sgRNAs is in **Table S3** (designed using CCTop and CRISPRater online tools (Labuhn et al., 2018; Stemmer et al., 2015)) that have been cloned into the SGL40C.EFS.E2Crimson or SGL40C.EFS.dTomato vectors (Addgene #100894 and #89395) (Reimer et al., 2017). Codon optimized CSF2RB WT and A455D cDNAs used for expression in cells were cloned into LeGO-iG2 vectors (Addgene, #27341) (Weber et al., 2011). TF1 cells were directly transduced with viral supernatant in presence of 5 µg/ml hexadimethrine bromide (Polybrene; Life Technologies). Primary human and mouse HSPCs were transduced with concentrated lentiviral particles with 2 µg/ml Polybrene in Retronectin-coated plates (ClonTech) according to manufacturer's instructions.

TF1 cell line transplantation in NSG mice

TF1 cells were transduced with CSF2RB expression vectors and 48 hr after transduction cells were selected with 1.5 µg/ml puromycin (Life Technologies) for 72 hr. The selection was verified by flow cytometry. Cells were then transplanted *i.v*. into sublethally irradiated (2.5 Gy) NSG mice in groups of five mice per condition ($1x10^7$ cells per mouse) as previously described (Bhayadia et al., 2018).

FLC transplantation in C57BL/6J mice

Ter119 depleted FLCs of E13.5 Cas9 knock-in mouse embryos were transduced with the Gata1-sgRNA expression vector and cultured for 3 weeks under low cytokine conditions for selection of

transduced *Gata1s*-cells. After the second transduction with pools of sgRNA expression vectors, cells were transplanted *i.v*. into sublethally irradiated (7.5 Gy) C57BL/6J mice in groups of five to ten mice per condition ($2x10^6$ cells per mouse).

Hematopoietic differentiation assays

Cord blood $CD34^+$ HSPCs were enriched from anonymous healthy donors using the EasySep™ Human CD34 Positive Selection Kit (Stemcell Technologies) and transduced with *CSF2RB*^{A455D}, wild-type or empty control vector concentrated virus as described before and sorted for GFP^+ cells 48 hr after transduction. Transduced cells were cultured in myeloid (week 1: RPMI (Gibco, Life Technologies) with 1% streptomycin/penicillin (Sigma-Aldrich), 10% FCS (Capricorn Scientific), 1% L-Glutamine (Millipore), 5 ng/mL SCF, 5 ng/mL GM-CSF, 10 ng/mL G-CSF, 5 ng/ml IL3 (all Peprotech); week 2: RPMI with 1% streptomycin/penicillin, 10% fetal calf serum, 1% L-Glutamine, 5 ng/mL GM-CSF, 10 ng/mL G-CSF (all Peprotech)), megakaryocytic (StemSpan SFEM (Stemcell Technologies) with 1% streptomycin/penicillin, 1x CD-Lipid Concentrate (Gibco, Life Technologies), 100 ng/mL THPO and 10 ng/mL SCF (all Peprotech)) or megakaryocytic/erythroid (week 1: StemSpan SFEM with 1% streptomycin/penicillin, 1% L-Glutamine, 1x CD-Lipid Concentrate, 100 ng/mL SCF, 50 ng/mL THPO, 10 ng/mL IL3 and 10 ng/mL IL6; week 2: StemSpan SFEM with 1% streptomycin/penicillin, 1% L-Glutamine, 1x CD-Lipid Concentrate, 100 ng/mL SCF, 10 ng/mL THPO, 10 ng/mL IL3, 10 ng/mL IL6 and 0.2U/mL EPO (GoldBio)) conditions as described before (Emmrich et al., 2014; Klusmann et al., 2010b; Maroz et al., 2014). Cell numbers were determined every second day starting from day 0 of differentiation. Differentiation markers were analysed by flow cytometry after 14 days (myeloid and megakaryocytic/erythroid differentiation) or 12 days (megakaryocytic differentiation) of culture.

Leukemic blast assays

To test LSD1-inhibitor T-3775440 (MedChemExpress), leukemic blasts were cultured in StemSpan SFEM (Stemcell Technologies) with 1% streptomycin/penicillin, 1x CD-Lipid Concentrate, 50 ng/mL SCF, 50 ng/mL FLT3, 50 ng/mL THPO, 10 ng/mL IL6, 2.5 ng/mL IL3 and 0.75 µM Stemregenin1 (SR1; StemCell Technologies) with increasing concentrations of T-3775440.

Signal transduction Dual luciferase assays and cell lines used

TF1 cells were transduced with CSF2RB A455D, CSF2RB WT or empty control vector as described before with a transduction rate of 5-10%. After 48 hr of culture (5 ng/ml GM-CSF, Peprotech), cells were transferred to different GM-CSF concentrations and cultured for 8 days. The percentage of GFP^+ cells was assessed by daily flow cytometry analysis (Ahmed et al., 2004).

For signal transduction assays, transduced GFP^+ TF1 cells were FACS-purified using a BD FACSARIA FUSION (BD Biosciences). $7x10^5$ purified cells were starved overnight at 37°C in GM-CSF- and FCS-free medium containing 0.5% bovine serum albumin (Sigma-Aldrich). GM-CSF was added at various concentrations for 15 minutes at 37°C. Cells were then fixed in 2% paraformaldehyde and permeabilized using ice-cold methanol. Phosphorylated STAT5 and AKT proteins were detected by flow cytometry (Anti-Akt (pS473), clone M89-61 and Anti-Stat5 (pY694), clone 47/Stat5(pY694); both BD Biosciences).

The previously described pMEGIX JAK2 (Dusa et al., 2008) was used while HA-MPL, Jak1 and JAK3 were subcloned into pMX-IRES-GFP. Mutated MPL, Jak1, JAK2 and JAK3 cDNAs were generated using the Quikchange method (Stratagene). Luciferase assays for STAT5 transcriptional activity in MPL, JAK2 and JAK3 were carried out in γ2A cells, a *JAK2*-deficient human fibrosarcoma cell line (Kohlhuber et al., 1997), while the *JAK1*-deficient U4C cell line was used for Jak1 (Guschin et al., 1995). Both γ2A and U4C cell lines were cultured in Dulbecco's Modified Eagles Medium (DMEM), supplemented with 10% FCS (Capricorn Scientific). STAT5 transcriptional activity in g2A cells with overexpression of MPL, JAK2 and JAK3 mutants, or in U4C cells with Jak1 mutants, was measured using a dual luciferase assay kit (Promega) on a GlowMax Discover luminometer (Promega). We used the STAT5 reporter pSpi-Luc (Wood et al., 1997), together with the pRL-TK plasmid, containing the renilla luciferase gene, as an internal control. Cells were transiently transfected using lipofectamine 2000 (Thermo Fisher) with pMX-IRES-GFP Stat5a, pRL-TK and pSpi-Luc, as well as MPL and JAK2 (for MPL and JAK2 mutants), Jak1 and JAK3 (for Jak1 mutants), or JAK2 and JAK3 (for JAK3 mutants). IL9Rα and γc were used instead of TMPL to examine transcription via JAK1 and JAK3. Dual luciferase

assays were performed as previously described (Leroy et al., 2016) with cells stimulated or not with 10 ng/mL of recombinant human THPO or 16 ng/mL of recombinant human IL9 (Miltenyi Biotec) 4 hr post transfection.

Analysis of CRISPR-Cas9 screening

Gata1 sequencing primers and amplicon sequencing primers are in **Table S5**. The 317 bases sequencing reads were assigned to 38 different samples using the unique barcode encoded in the first 6 bases of the reads. The reads were quality and adapter-trimmed using the bbduk tool from BBTools package (https://jgi.doe.gov/data-and-tools/bbtools/) and mapped on the mouse reference version GRCm38 using the bwa mem (Li and Durbin, 2009). All mutations (SNVs or short indels) occurred in any primary mapped read inside of the genes target positions were collected using the pysam module (Li et al., 2009) and subsequently annotated using the annovar tool (Wang et al., 2010). For each sgRNA a number of reads containing its spacer sequence was counted. The 22 bp spacer sequence was locked up in sequencing reads at the positions 23-49.

Murine RNA sequencing

RNA reads were aligned to GRCm38 using STAR (v2.6.0) and used for integrated data analysis. Mapped reads were annotated using Ensembl v.91 (Aken et al., 2017; Dobin et al., 2013). Gene expression levels were quantified in fragments per kb of exon model per million mapped reads (FPKM). FPKM calculation was performed using the R package edgeR including TMM normalization (Mortazavi et al., 2008; Robinson and Oshlack, 2010). The calculation of functional enrichments in ML-DS leukemia mice models in comparison to sorted murine fetal liver early stem/progenitor cells was done using gene set enrichment analysis (GSEA; v3.0) on $\log_2$-transformed FPKM values (Subramanian et al., 2005). The analysis included previously described and curated gene sets (Schwarzer et al., 2017) and the ML-DS signatures (Bourquin et al., 2006; Schwarzer et al., 2017). Human gene symbols of the published leukemia gene sets were mapped to murine gene symbols using orthologue annotations provided by Ensembl (Aken et al., 2017) considering only one-to-one orthologue relationships. Standard deviation of gene expression, across murine fetal liver early stem/progenitor cells and murine leukemias, was

calculated using log_2-transformed FPKM values. 1613 gene that showed a standard deviation >1 were selected as most variable genes and utilized for unsupervised hierarchical clustering.

Flow cytometry and sorting

Flow cytometry was performed on a BD FACSCanto™ (BD Biosciences) and Kaluza 1.3/1.5 (Beckman Coulter) and FlowJo 10.0.8r1 (FlowJo LLC). Antibodies are in the **Key Resources Table**. Sorting of murine HSPCs (LSKs: Lin$^-$Sca1$^+$cKit$^+$; MEPs: Lin$^-$Sca1$^-$cKit$^+$CD34$^-$FcγR$^-$; CMPs: Lin$^-$Sca1$^-$cKit$^+$CD34$^+$FcγR^{low}; GMPs: Lin$^-$Sca1$^-$cKit$^+$CD34$^+$FcγR$^+$) directly into in RNA lysis buffer (Zymo Research, Quick-RNA Microprep R1050) was performed on a BD FACSAria FUSION (BD Biosciences).

RNA Sequencing

ML-DS samples 187 (bone marrow) and 008 (peripheral blood), where no additional somatic variants were identified, were further assessed for oncogenic fusion transcripts. Thawed samples were stained with anti-human CD19-PECy5, CD3-PECy5, CD45-APC-H7, CD117-PE and Hoechst 33258. The antibodies used are listed in **Key Resources Table** Live CD19$^-$ CD3$^-$ CD45^{mid}CD117$^+$ cells were FACS-sorted purified using unstained, single stained and Fluorescence Minus One (FMO) controls were used to set the compensation for each channel and determine background staining. 100 cells were sorted directly into lysis buffer containing RNAse inhibitor (Clontech). cDNA synthesis was performed using SmartSeq-V4 Ultra low input RNA kit (Clontech). The Nextera XT DNA sample preparation kit and Index Kit was used to generate Illumina libraries (Illumina UK). Library size and quality were checked using the Agilent Bioanalyser (Agilent Technologies UK). Libraries were sequenced on the Illumina NextSeq 150 bp paired end reads.

Sequencing reads were mapped to the GRCh37 human genome reference using STAR (Dobin et al., 2013). Fusions were detected using Cicero (Roberts et al., 2014) and FusionCatcher (Nicorici et al., 2014). Visualisation of putative fusion transcripts was carried out using the St Jude Cloud Platform (Rapid RNA-Seq Fusion Detection - St. Jude Cloud Documentation. Available at: https://stjude.github.io/sjcloud-docs/guides/tools/rapid-rnaseq/. (Accessed: 3rd April 2019)). To assess gene expression profile, read counts for each annotated gene were

quantitated using featureCounts (Liao et al., 2014) and gene expression level normalization calculated as Transcripts Per Million (TPM).

Single Cell Sequencing of human and mouse cells

Human Single Cell Sorting: Single cells from ML-DS sample 186-2 (bone marrow) were analyzed to determine whether mutations in *JAK2* and *CSF2RB* occurred in the same or different cells in this patient. Single live, $CD19^{-}CD3\text{-}CD45^{mid}$ $CD117^{+}$ cells were sorted into 96-well plates on the BD FACSAria FUSION with antibodies as described above. Cord blood single cells were sorted similarly to serve as a control for the single cell sequencing. Primers are listed in **Table S5**.

Mouse single cell sorting: Single murine leukemia cells from mouse 322 that harboured frameshift mutations in both *Sh2b3* and *Nf1* were analysed to determine whether these mutations occurred in the same cells. Live, GFP^{+} cells from frozen murine bone marrow were single cell sorted into 96 well-plates on the BD Fusion. Control mouse cells that were not genome-edited were single cell sorted as a control for the single cell sequencing. Primers are listed in **Table S5**.

Single human or mouse cells were lysed and whole genome amplified using a REPLI-g Single Cell Kit (Qiagen). Primers for targeted re-sequencing of patient mutations were designed to include patient-specific heterozygous SNPs within the same amplicon (listed in **Table S5**). For the murine leukemia the primers from the CRISPR-Cas9 screening were used (listed in **Table S5**).Targeted amplification was performed using KAPA2G Fast Multiplex PCR Kit (Sigma-Aldrich), followed by addition of sample barcodes and sequencing adapters using FastStart High Fidelity PCR System (Roche) and Access Array™ Barcode Library for Illumina Sequencers (Fluidigm), as previously described (Quek et al., 2018). A total of 240 human patient and control cells (x3 amplicons) and 96 mouse leukemia and control cells (x2 amplicons) were sequenced on an Illumina MiSeq using a Nano Reagent Kit (Illumina) and 300-cycle paired-end sequencing.

Bioinformatic analysis of human single cells: Variant allele frequencies (VAFs) of the 3 mutations of interest were obtained using the UNIX command ‘grep’ (global regular expression print), counting the number of reference and mutant reads in the fastq files. Control cord blood

cells were used to determine VAF thresholds for each mutation (1). Minimum coverage was set at 100, 50 or 10 reads for *GATA1*, *CSF2RB* and *JAK2*, respectively. Co-amplification of proximal heterozygous SNPs in the patient was used to determine allelic dropout and helped exclude cells where the mutated allele was not amplified. All cells were scored according to the presence, absence or allelic dropout of the 3 mutations GATA1 p.V74_splice, CSF2RB p.A455D and JAK2 p.V617F.

Bioinformatic analysis of mouse single cells: Raw fastq files from mouse single cells with amplified *Sh2b3* and *Nf1* regions were analyzed using the CRISPRessoPooled software pipeline (Canver et al., 2018). Fastq files containing pooled amplicon sequences were demultiplexed by local alignment to given amplicon sequences. CRISPResso then quantifies the percentage of non-homologous end-joining sequences, including types and locations of insertions and deletions in the regions of interest.

QUANTIFICATION AND STATISTICAL ANALYSIS

Bioinformatic pipelines and variant calling

Variant analysis was performed on 111 TAM, 141 ML-DS samples with confirmed *GATA1* mutation(s) identified using ≥1 of the following next generation sequencing approaches. *GATA1* mutational data from custom RNA bait/exome re-sequencing was processing using previously described approaches (Roberts et al., 2013) and an in-house coupled mapping and variant pipeline (Quek et al., 2016). Adapter sequences were trimmed and aligned to hg19 human reference genome with Stampy (Lunter and Goodson, 2011) and SortSAM and SAMtools v0.1.9 (r783) (Li et al., 2009) were used to output pileup and sorted BAM files. The variant detection pipeline generated VarScan somatic data (Koboldt et al., 2009), with a secondary pipeline containing VarScan normal (reference sequence), Pindel (Ye et al., 2009) and GATK (McKenna et al., 2010) output. VarScan somatic (Koboldt et al., 2009) was primarily used for documenting allele frequencies present in a tumour sample and were compared to a *GATA1* variant negative sample from the same individual (post-treatment or resolution of TAM) or two unmatched control normal neonate samples (Roberts et al., 2013). VarScan somatic parameters (i) Access Array data [min-coverage 100, min-avg-qual 35, min-var-freq 0.05, min-freq-for-hom 0.9, p value 0.05],

(ii) Bait and Exome re-sequencing data: [min-coverage 6, min-avg-qual 35, min-var-freq 0.05, min-freq-for-hom 0.9, p value 0.05]. Annovar (Wang et al., 2010) provided annotation data, including Refseq, prediction of functional impact (Polyphen), and population frequency sourced from 1000 Genomes (1000g2012apr_all), COSMIC (Forbes et al., 2015), snp137, and exome server data (esp6500_all).

Exome bait, amplicon and variant pipeline output used initial filtering criteria as follows:

(i) Removal of variants resulting in a synonymous amino acid change, intergenic (excepting ± 15 bp/exon resulting in perturbation of splice consensus) or with ambiguous (unknown) annotations.

(ii) Removal if identified as a constitutional SNP/ in a public database (including dbSNP146, and 1000 Genomes) or absent as a confirmed somatic event in COSMIC (Forbes et al., 2015)/PubMed.

(iii) Removal of variants with allele frequencies of 0.45–0.55 if not identified in steps (i-ii), present at commensurate VAF in a matched non-tumor sample.

(iv) Removal of highly recurrent calls present in both tumor samples and controls (<10%) indicative of sequencing artifacts.

(v) Individual manual inspection of variants passing (i-iii) using the IGV: Integrated Genome Viewer.

(vi) Subsequent removal of variants subject to (v) localizing duplicated regions (SeqDups of the UCSC Genome browser), or as a result of low mapping / sequencing quality (incorporating proximity to homopolymers, and inspection of Phred score ± 10 bp).

Variants passing the above parameters were subjected to secondary criteria:

(i) Removal of missense variants with <10% VAF if:

(A) Previously undocumented and detected using a single methodology, present in a single sample (i.e, novel, but not recurrent).

(B) Not considered “oncogenic” if not present in prior literature or database, or a truncating variant in a gene implicated in myeloid malignancies by means of loss of function, according to filtering parameters applied by (Papaemmanuil et al., 2013).

(ii) If criteria in (i) met, variants were passed if 5-fold above a threshold value calculated specifically for each variant, by setting a maximum background value based on population analysis of n=96 samples by Grep (on processed reads with Phred score <35; average threshold =0.1%; range: 0-1%), then confirmed using the modified Thompson Tau outlier test.

(iii) Previously unannotated missense SNVs ≥10% VAF were included, subject to confirmation using population and fold difference criteria outlined in (ii).

For multiple samples from a single individual (e.g., TAM and ML-DS; sequential TAM; multiple pre-treatment ML-DS), a variant sequence present in ≥ 1 sample, above the VarScan somatic threshold and passing the initial filtering process, was subsequently analyzed in associated samples for VAF <5% by sequence-specific Grep. Results were only accepted if Secondary criteria (ii) above were met. A full list of annotated variants for each sample, including allele frequency, is provided in **Table S1**.

Statistical analysis

Modified Thompson Tau test calculation was: $\tau = \bullet (n-1)/(\sqrt{n}\sqrt{n-2+t^2})$; t = student t value based on $\alpha = 0.05$, degrees of freedom = 2. Confidence limits (95%) were calculated using sample mean ($\overline{x}$) and standard deviation (σ) of population, with ($\overline{x} \pm 0.95_{a/2} * \sigma/\sqrt{n}$) used for lower and upper bounds. P values were obtained using a paired 2-tailed Student's *t* test (Prism 7; GraphPad Software). Bean plots used calculations from R software (BoxPlotR; http://boxplot.bio.ed.ac.uk) to calculate density of data and mean values, with full statistics listed in **Table S2**. All statistical tests and sample numbers are disclosed in respective Figure Legends/ Supplementary Tables.

DATA AND SOFTWARE AVAILABILITY

DNA-seq and RNA-seq data of murine leukemias were deposited in the European Nucleotide Archive (ENA) under the accession number PRJEB30376; in the European Genome-phenome

Archive (EGA) under the accession numbers EGAD00001000783, EGAD00001000879, and EGAD00001000070; and in ArrayExpress under the accession number E-MTAB-7729.

Supplemental video Excel table title and legends

Table S1, related to Figure 1: Patient samples, sequencing approaches and metrics.

Table S2, related to Figure 2: Analysis of clonal structure in TAM and ML-DS samples

Table S3, related to Figure 5: sgRNA cutting efficacies, sgRNA-sub-pools and GSEA analysis

Table S4, related to Figure 6: Sequencing of murine leukemias

Table S5, related to STAR Methods: Primer sequences

References

Ahmed, M., Sternberg, A., Hall, G., Thomas, A., Smith, O., O'Marcaigh, A., Wynn, R., Stevens, R., Addison, M., King, D.*, et al.* (2004). Natural history of GATA1 mutations in Down syndrome. Blood *103*, 2480-2489.

Aken, B. L., Achuthan, P., Akanni, W., Amode, M. R., Bernsdorff, F., Bhai, J., Billis, K., Carvalho-Silva, D., Cummins, C., Clapham, P.*, et al.* (2017). Ensembl 2017. Nucleic Acids Res *45*, D635-D642.

Alford, K. A., Reinhardt, K., Garnett, C., Norton, A., Bohmer, K., von Neuhoff, C., Kolenova, A., Marchi, E., Klusmann, J. H., Roberts, I.*, et al.* (2011). Analysis of GATA1 mutations in Down syndrome transient myeloproliferative disorder and myeloid leukemia. Blood *118*, 2222-2238.

Alford, K. A., Slender, A., Vanes, L., Li, Z., Fisher, E. M., Nizetic, D., Orkin, S. H., Roberts, I., and Tybulewicz, V. L. (2010). Perturbed hematopoiesis in the Tc1 mouse model of Down syndrome. Blood *115*, 2928-2937.

Andersson, A. K., Ma, J., Wang, J., Chen, X., Gedman, A. L., Dang, J., Nakitandwe, J., Holmfeldt, L., Parker, M., Easton, J.*, et al.* (2015). The landscape of somatic mutations in infant MLL-rearranged acute lymphoblastic leukemias. Nat Genet *47*, 330-337.

Bailey, M. H., Tokheim, C., Porta-Pardo, E., Sengupta, S., Bertrand, D., Weerasinghe, A., Colaprico, A., Wendl, M. C., Kim, J., Reardon, B.*, et al.* (2018). Comprehensive Characterization of Cancer Driver Genes and Mutations. Cell *173*, 371-385 e318.

Bargmann, C. I., and Weinberg, R. A. (1988). Oncogenic activation of the neu-encoded receptor protein by point mutation and deletion. EMBO J *7*, 2043-2052.

Baxter, E. J., Scott, L. M., Campbell, P. J., East, C., Fourouclas, N., Swanton, S., Vassiliou, G. S., Bench, A. J., Boyd, E. M., Curtin, N.*, et al.* (2005). Acquired mutation of the tyrosine kinase JAK2 in human myeloproliferative disorders. Lancet *365*, 1054-1061.

Bercovich, D., Ganmore, I., Scott, L. M., Wainreb, G., Birger, Y., Elimelech, A., Shochat, C., Cazzaniga, G., Biondi, A., Basso, G.*, et al.* (2008). Mutations of JAK2 in acute lymphoblastic leukaemias associated with Down's syndrome. Lancet *372*, 1484-1492.

Bhayadia, R., Krowiorz, K., Haetscher, N., Jammal, R., Emmrich, S., Obulkasim, A., Fiedler, J., Schwarzer, A., Rouhi, A., Heuser, M.*, et al.* (2018). Endogenous Tumor Suppressor microRNA-193b: Therapeutic and Prognostic Value in Acute Myeloid Leukemia. J Clin Oncol *36*, 1007-1016.

Bolouri, H., Farrar, J. E., Triche, T., Jr., Ries, R. E., Lim, E. L., Alonzo, T. A., Ma, Y., Moore, R., Mungall, A. J., Marra, M. A.*, et al.* (2018). The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions. Nat Med *24*, 103-112.

Bourquin, J. P., Subramanian, A., Langebrake, C., Reinhardt, D., Bernard, O., Ballerini, P., Baruchel, A., Cave, H., Dastugue, N., Hasle, H.*, et al.* (2006). Identification of distinct molecular phenotypes in acute megakaryoblastic leukemia by gene expression profiling. Proc Natl Acad Sci U S A *103*, 3339-3344.

Broughton, S. E., Nero, T. L., Dhagat, U., Kan, W. L., Hercus, T. R., Tvorogov, D., Lopez, A. F., and Parker, M. W. (2015). The betac receptor family - Structural insights and their functional implications. Cytokine *74*, 247-258.

Campbell, P. J., Scott, L. M., Buck, G., Wheatley, K., East, C. L., Marsden, J. T., Duffy, A., Boyd, E. M., Bench, A. J., Scott, M. A.*, et al.* (2005). Definition of subtypes of essential thrombocythaemia and relation to polycythaemia vera based on JAK2 V617F mutation status: a prospective study. Lancet *366*, 1945-1953.

Canver, M. C., Haeussler, M., Bauer, D. E., Orkin, S. H., Sanjana, N. E., Shalem, O., Yuan, G. C., Zhang, F., Concordet, J. P., and Pinello, L. (2018). Integrated design, execution, and analysis of arrayed and pooled CRISPR genome-editing experiments. Nat Protoc *13*, 946-986.

Corces-Zimmerman, M. R., Hong, W. J., Weissman, I. L., Medeiros, B. C., and Majeti, R. (2014). Preleukemic mutations in human acute myeloid leukemia affect epigenetic regulators and persist in remission. Proc Natl Acad Sci U S A *111*, 2548-2553.

D'Andrea, R. J., Harrison-Findik, D., Butcher, C. M., Finnie, J., Blumbergs, P., Bartley, P., McCormack, M., Jones, K., Rowland, R., Gonda, T. J., and Vadas, M. A. (1998). Dysregulated

hematopoiesis and a progressive neurological disorder induced by expression of an activated form of the human common beta chain in transgenic mice. J Clin Invest *102*, 1951-1960.
Ding, J., Komatsu, H., Wakita, A., Kato-Uranishi, M., Ito, M., Satoh, A., Tsuboi, K., Nitta, M., Miyazaki, H., Iida, S., and Ueda, R. (2004). Familial essential thrombocythemia associated with a dominant-positive activating mutation of the c-MPL gene, which encodes for the receptor for thrombopoietin. Blood *103*, 4198-4200.
Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15-21.
Donehower, L. A., Harvey, M., Slagle, B. L., McArthur, M. J., Montgomery, C. A., Jr., Butel, J. S., and Bradley, A. (1992). Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours. Nature *356*, 215-221.
Drissen, R., Thongjuea, S., Theilgaard-Monch, K., and Nerlov, C. (2019). Identification of two distinct pathways of human myelopoiesis. Science Immunology *Invited resubmssion resubmitted for review*.
Dusa, A., Staerk, J., Elliott, J., Pecquet, C., Poirel, H. A., Johnston, J. A., and Constantinescu, S. N. (2008). Substitution of pseudokinase domain residue Val-617 by large non-polar amino acids causes activation of JAK2. J Biol Chem *283*, 12941-12948.
Emmrich, S., Rasche, M., Schoning, J., Reimer, C., Keihani, S., Maroz, A., Xie, Y., Li, Z., Schambach, A., Reinhardt, D., and Klusmann, J. H. (2014). miR-99a/100~125b tricistrons regulate hematopoietic stem and progenitor cell homeostasis by shifting the balance between TGFbeta and Wnt signaling. Genes Dev *28*, 858-874.
Flasinski, M., Scheibke, K., Zimmermann, M., Creutzig, U., Reinhardt, K., Verwer, F., de Haas, V., van der Velden, V. H. J., von Neuhoff, C., Zwaan, C. M.*, et al.* (2018). Low-dose cytarabine to prevent myeloid leukemia in children with Down syndrome: TMD Prevention 2007 study. Blood Adv *2*, 1532-1540.
Forbes, S. A., Beare, D., Gunasekaran, P., Leung, K., Bindal, N., Boutselakis, H., Ding, M., Bamford, S., Cole, C., Ward, S.*, et al.* (2015). COSMIC: exploring the world's knowledge of somatic mutations in human cancer. Nucleic Acids Res *43*, D805-811.
Gamis, A. S., Alonzo, T. A., Gerbing, R. B., Hilden, J. M., Sorrell, A. D., Sharma, M., Loew, T. W., Arceci, R. J., Barnard, D., Doyle, J.*, et al.* (2011). Natural history of transient myeloproliferative disorder clinically diagnosed in Down syndrome neonates: a report from the Children's Oncology Group Study A2971. Blood *118*, 6752-6759; quiz 6996.
Guschin, D., Rogers, N., Briscoe, J., Witthuhn, B., Watling, D., Horn, F., Pellegrini, S., Yasukawa, K., Heinrich, P., Stark, G. R., and et al. (1995). A major role for the protein tyrosine kinase JAK1 in the JAK/STAT signal transduction pathway in response to interleukin-6. EMBO J *14*, 1421-1429.
Hansen, G., Hercus, T. R., McClure, B. J., Stomski, F. C., Dottore, M., Powell, J., Ramshaw, H., Woodcock, J. M., Xu, Y., Guthridge, M.*, et al.* (2008). The structure of the GM-CSF receptor complex reveals a distinct mode of cytokine receptor activation. Cell *134*, 496-507.
Hasle, H., Clemmensen, I. H., and Mikkelsen, M. (2000). Risks of leukaemia and solid tumours in individuals with Down's syndrome. Lancet *355*, 165-169.
Hercus, T. R., Dhagat, U., Kan, W. L., Broughton, S. E., Nero, T. L., Perugini, M., Sandow, J. J., D'Andrea, R. J., Ekert, P. G., Hughes, T.*, et al.* (2013). Signalling by the betac family of cytokines. Cytokine Growth Factor Rev *24*, 189-201.
Hong, S. H., Gupta, R and Enver, T (2008). Initiating and cancer propagating cells in TEL-AML1 associated childhood leukaemia. Science *319*, 336-339.
Ishikawa, Y., Gamo, K., Yabuki, M., Takagi, S., Toyoshima, K., Nakayama, K., Nakayama, A., Morimoto, M., Miyashita, H., Dairiki, R.*, et al.* (2017). A Novel LSD1 Inhibitor T-3775440 Disrupts GFI1B-Containing Complex Leading to Transdifferentiation and Impaired Growth of AML Cells. Mol Cancer Ther *16*, 273-284.
James, C., Ugo, V., Le Couedic, J. P., Staerk, J., Delhommeau, F., Lacout, C., Garcon, L., Raslova, H., Berger, R., Bennaceur-Griscelli, A.*, et al.* (2005). A unique clonal JAK2 mutation leading to constitutive signalling causes polycythaemia vera. Nature *434*, 1144-1148.

Jenkins, B. J., D'Andrea, R., and Gonda, T. J. (1995). Activating point mutations in the common beta subunit of the human GM-CSF, IL-3 and IL-5 receptors suggest the involvement of beta subunit dimerization and cell type-specific molecules in signalling. EMBO J *14*, 4276-4287.
Joung, J., Konermann, S., Gootenberg, J. S., Abudayyeh, O. O., Platt, R. J., Brigham, M. D., Sanjana, N. E., and Zhang, F. (2017). Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. Nat Protoc *12*, 828-863.
Kiser, M., McCubrey, J. A., Steelman, L. S., Shelton, J. G., Ramage, J., Alexander, R. L., Kucera, G. L., Pettenati, M., Willingham, M. C., Miller, M. S., and Frankel, A. E. (2001). Oncogene-dependent engraftment of human myeloid leukemia cells in immunosuppressed mice. Leukemia *15*, 814-818.
Kiyoi, H., Yamaji, S., Kojima, S., and Naoe, T. (2007). JAK3 mutations occur in acute megakaryoblastic leukemia both in Down syndrome children and non-Down syndrome adults. Leukemia *21*, 574-576.
Klusmann, J. H., Creutzig, U., Zimmermann, M., Dworzak, M., Jorch, N., Langebrake, C., Pekrun, A., Macakova-Reinhardt, K., and Reinhardt, D. (2008). Treatment and prognostic impact of transient leukemia in neonates with Down syndrome. Blood *111*, 2991-2998.
Klusmann, J. H., Godinho, F. J., Heitmann, K., Maroz, A., Koch, M. L., Reinhardt, D., Orkin, S. H., and Li, Z. (2010a). Developmental stage-specific interplay of GATA1 and IGF signaling in fetal megakaryopoiesis and leukemogenesis. Genes Dev *24*, 1659-1672.
Klusmann, J. H., Li, Z., Bohmer, K., Maroz, A., Koch, M. L., Emmrich, S., Godinho, F. J., Orkin, S. H., and Reinhardt, D. (2010b). miR-125b-2 is a potential oncomiR on human chromosome 21 in megakaryoblastic leukemia. Genes Dev *24*, 478-490.
Koboldt, D. C., Chen, K., Wylie, T., Larson, D. E., McLellan, M. D., Mardis, E. R., Weinstock, G. M., Wilson, R. K., and Ding, L. (2009). VarScan: variant detection in massively parallel sequencing of individual and pooled samples. Bioinformatics *25*, 2283-2285.
Kohlhuber, F., Rogers, N. C., Watling, D., Feng, J., Guschin, D., Briscoe, J., Witthuhn, B. A., Kotenko, S. V., Pestka, S., Stark, G. R*., et al.* (1997). A JAK1/JAK2 chimera can sustain alpha and gamma interferon responses. Mol Cell Biol *17*, 695-706.
Labuhn, M., Adams, F. F., Ng, M., Knoess, S., Schambach, A., Charpentier, E. M., Schwarzer, A., Mateo, J. L., Klusmann, J. H., and Heckl, D. (2018). Refined sgRNA efficacy prediction improves large- and small-scale CRISPR-Cas9 applications. Nucleic Acids Res *46*, 1375-1385.
Lange, B. J., Kobrinsky, N., Barnard, D. R., Arthur, D. C., Buckley, J. D., Howells, W. B., Gold, S., Sanders, J., Neudorf, S., Smith, F. O., and Woods, W. G. (1998). Distinctive demography, biology, and outcome of acute myeloid leukemia and myelodysplastic syndrome in children with Down syndrome: Children's Cancer Group Studies 2861 and 2891. Blood *91*, 608-615.
Leroy, E., Dusa, A., Colau, D., Motamedi, A., Cahu, X., Mouton, C., Huang, L. J., Shiau, A. K., and Constantinescu, S. N. (2016). Uncoupling JAK2 V617F activation from cytokine-induced signalling by modulation of JH2 alphaC helix. The Biochemical journal *473*, 1579-1591.
Levine, R. L., Wadleigh, M., Cools, J., Ebert, B. L., Wernig, G., Huntly, B. J., Boggon, T. J., Wlodarska, I., Clark, J. J., Moore, S*., et al.* (2005). Activating mutation in the tyrosine kinase JAK2 in polycythemia vera, essential thrombocythemia, and myeloid metaplasia with myelofibrosis. Cancer Cell *7*, 387-397.
Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754-1760.
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.
Li, Z., Godinho, F. J., Klusmann, J. H., Garriga-Canut, M., Yu, C., and Orkin, S. H. (2005). Developmental stage-selective effect of somatically mutated leukemogenic transcription factor GATA1. Nat Genet *37*, 613-619.
Liao, Y., Smyth, G. K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923-930.
Lunter, G., and Goodson, M. (2011). Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. Genome Res *21*, 936-939.

Ma, X., Liu, Y., Liu, Y., Alexandrov, L. B., Edmonson, M. N., Gawad, C., Zhou, X., Li, Y., Rusch, M. C., Easton, J.*, et al.* (2018). Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. Nature *555*, 371-376.
Malinge, S., Ragu, C., Della-Valle, V., Pisani, D., Constantinescu, S. N., Perez, C., Villeval, J. L., Reinhardt, D., Landman-Parker, J., Michaux, L.*, et al.* (2008). Activating mutations in human acute megakaryoblastic leukemia. Blood *112*, 4220-4226.
Manz, M. G. (2007). Human-hemato-lymphoid-system mice: opportunities and challenges. Immunity *26*, 537-541.
Maroz, A., Stachorski, L., Emmrich, S., Reinhardt, K., Xu, J., Shao, Z., Kabler, S., Dertmann, T., Hitzler, J., Roberts, I.*, et al.* (2014). GATA1s induces hyperproliferation of eosinophil precursors in Down syndrome transient leukemia. Leukemia *28*, 1259-1270.
Maxson, J. E., Luty, S. B., MacManiman, J. D., Paik, J. C., Gotlib, J., Greenberg, P., Bahamadi, S., Savage, S. L., Abel, M. L., Eide, C. A.*, et al.* (2016). The Colony-Stimulating Factor 3 Receptor T640N Mutation Is Oncogenic, Sensitive to JAK Inhibition, and Mimics T618I. Clin Cancer Res *22*, 757-764.
Mazumdar, C., Shen, Y., Xavy, S., Zhao, F., Reinisch, A., Li, R., Corces, M. R., Flynn, R. A., Buenrostro, J. D., Chan, S. M.*, et al.* (2015). Leukemia-Associated Cohesin Mutants Dominantly Enforce Stem Cell Programs and Impair Human Hematopoietic Progenitor Differentiation. Cell Stem Cell *17*, 675-688.
McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res *20*, 1297-1303.
Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods *5*, 621-628.
Mullenders, J., Aranda-Orgilles, B., Lhoumaud, P., Keller, M., Pae, J., Wang, K., Kayembe, C., Rocha, P. P., Raviram, R., Gong, Y.*, et al.* (2015). Cohesin loss alters adult hematopoietic stem cell homeostasis, leading to myeloproliferative neoplasms. J Exp Med *212*, 1833-1850.
Network, C. G. A. R. (2013). Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. N Engl J Med *368*, 2059-2074.
Nicorici, D., Satalan, M., Edgren, H., Kangaspeska, S., Murumagi, A., Kallioniemi, O., Viratanen, S., and Kilkku, O. (2014). FusionCatcher - a tool for finding somatic fusion genes in paired-end RNA-sequencing data. bioRxiv *https:doi.org/10.1101/011650*.
Nikolaev, S. I., Santoni, F., Vannier, A., Falconnet, E., Giarin, E., Basso, G., Hoischen, A., Veltman, J. A., Groet, J., Nizetic, D., and Antonarakis, S. E. (2013). Exome sequencing identifies putative drivers of progression of transient myeloproliferative disorder to AMKL in infants with Down syndrome. Blood *122*, 554-561.
Papaemmanuil, E., Gerstung, M., Malcovati, L., Tauro, S., Gundem, G., Van Loo, P., Yoon, C. J., Ellis, P., Wedge, D. C., Pellagatti, A.*, et al.* (2013). Clinical and biological implications of driver mutations in myelodysplastic syndromes. Blood *122*, 3616-3627; quiz 3699.
Pikman, Y., Lee, B. H., Mercher, T., McDowell, E., Ebert, B. L., Gozo, M., Cuker, A., Wernig, G., Moore, S., Galinsky, I.*, et al.* (2006). MPLW515L is a novel somatic activating mutation in myelofibrosis with myeloid metaplasia. PLoS Med *3*, e270.
Platt, R. J., Chen, S., Zhou, Y., Yim, M. J., Swiech, L., Kempton, H. R., Dahlman, J. E., Parnas, O., Eisenhaure, T. M., Jovanovic, M.*, et al.* (2014). CRISPR-Cas9 knockin mice for genome editing and cancer modeling. Cell *159*, 440-455.
Plo, I., Zhang, Y., Le Couedic, J. P., Nakatake, M., Boulet, J. M., Itaya, M., Smith, S. O., Debili, N., Constantinescu, S. N., Vainchenker, W.*, et al.* (2009). An activating mutation in the CSF3R gene induces a hereditary chronic neutrophilia. J Exp Med *206*, 1701-1707.
Quek, L., David, M. D., Kennedy, A., Metzner, M., Amatangelo, M., Shih, A., Stoilova, B., Quivoron, C., Heiblig, M., Willekens, C.*, et al.* (2018). Clonal heterogeneity of acute myeloid leukemia treated with the IDH2 inhibitor enasidenib. Nat Med *24*, 1167-1177.
Quek, L., Otto, G. W., Garnett, C., Lhermitte, L., Karamitros, D., Stoilova, B., Lau, I. J., Doondeea, J., Usukhbayar, B., Kennedy, A.*, et al.* (2016). Genetically distinct leukemic stem

cells in human CD34- acute myeloid leukemia are arrested at a hemopoietic precursor-like stage. J Exp Med *213*, 1513-1535.
Reimer, J., Knoss, S., Labuhn, M., Charpentier, E. M., Gohring, G., Schlegelberger, B., Klusmann, J. H., and Heckl, D. (2017). CRISPR-Cas9-induced t(11;19)/MLL-ENL translocations initiate leukemia in human hematopoietic progenitor cells in vivo. Haematologica *102*, 1558-1566.
Roberts, I., Alford, K., Hall, G., Juban, G., Richmond, H., Norton, A., Vallance, G., Perkins, K., Marchi, E., McGowan, S.*, et al.* (2013). GATA1-mutant clones are frequent and often unsuspected in babies with Down syndrome: identification of a population at risk of leukemia. Blood *122*, 3908-3917.
Roberts, K. G., Li, Y., Payne-Turner, D., Harvey, R. C., Yang, Y. L., Pei, D., McCastlain, K., Ding, L., Lu, C., Song, G.*, et al.* (2014). Targetable kinase-activating lesions in Ph-like acute lymphoblastic leukemia. N Engl J Med *371*, 1005-1015.
Robinson, M. D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol *11*, R25.
Sakaguchi, H., Okuno, Y., Muramatsu, H., Yoshida, K., Shiraishi, Y., Takahashi, M., Kon, A., Sanada, M., Chiba, K., Tanaka, H.*, et al.* (2013). Exome sequencing identifies secondary mutations of SETBP1 and JAK3 in juvenile myelomonocytic leukemia. Nat Genet *45*, 937-941.
Schwarzer, A., Emmrich, S., Schmidt, F., Beck, D., Ng, M., Reimer, C., Adams, F. F., Grasedieck, S., Witte, D., Kabler, S.*, et al.* (2017). The non-coding RNA landscape of human hematopoiesis and leukemia. Nat Commun *8*, 218.
Staerk, J., Lacout, C., Sato, T., Smith, S. O., Vainchenker, W., and Constantinescu, S. N. (2006). An amphipathic motif at the transmembrane-cytoplasmic junction prevents autonomous activation of the thrombopoietin receptor. Blood *107*, 1864-1871.
Stemmer, M., Thumberger, T., Del Sol Keyer, M., Wittbrodt, J., and Mateo, J. L. (2015). CCTop: An Intuitive, Flexible and Reliable CRISPR/Cas9 Target Prediction Tool. PLoS One *10*, e0124633.
Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., and Mesirov, J. P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A *102*, 15545-15550.
Uffmann, M., Rasche, M., Zimmermann, M., von Neuhoff, C., Creutzig, U., Dworzak, M., Scheffers, L., Hasle, H., Zwaan, C. M., Reinhardt, D., and Klusmann, J. H. (2017). Therapy reduction in patients with Down syndrome and myeloid leukemia: the international ML-DS 2006 trial. Blood *129*, 3314-3321.
Van Loo, P., Nordgard, S. H., Lingjaerde, O. C., Russnes, H. G., Rye, I. H., Sun, W., Weigman, V. J., Marynen, P., Zetterberg, A., Naume, B.*, et al.* (2010). Allele-specific copy number analysis of tumors. Proc Natl Acad Sci U S A *107*, 16910-16915.
Viny, A. D., Ott, C. J., Spitzer, B., Rivas, M., Meydan, C., Papalexi, E., Yelin, D., Shank, K., Reyes, J., Chiu, A.*, et al.* (2015). Dose-dependent role of the cohesin complex in normal and malignant hematopoiesis. J Exp Med *212*, 1819-1832.
Walters, D. K., Mercher, T., Gu, T. L., O'Hare, T., Tyner, J. W., Loriaux, M., Goss, V. L., Lee, K. A., Eide, C. A., Wong, M. J.*, et al.* (2006). Activating alleles of JAK3 in acute megakaryoblastic leukemia. Cancer Cell *10*, 65-75.
Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res *38*, e164.
Watanabe-Smith, K., Tognon, C., Tyner, J. W., Meijerink, J. P., Druker, B. J., and Agarwal, A. (2016). Discovery and functional characterization of a germline, CSF2RB-activating mutation in leukemia. Leukemia *30*, 1950-1953.
Weber, K., Thomaschewski, M., Warlich, M., Volz, T., Cornils, K., Niebuhr, B., Tager, M., Lutgehetmann, M., Pollok, J. M., Stocking, C.*, et al.* (2011). RGB marking facilitates multicolor clonal cell tracking. Nat Med *17*, 504-509.
Wechsler, J., Greene, M., McDevitt, M. A., Anastasi, J., Karp, J. E., Le Beau, M. M., and Crispino, J. D. (2002). Acquired mutations in GATA1 in the megakaryoblastic leukemia of Down syndrome. Nat Genet *32*, 148-152.

Wood, T. J., Sliva, D., Lobie, P. E., Goullieux, F., Mui, A. L., Groner, B., Norstedt, G., and Haldosen, L. A. (1997). Specificity of transcription enhancement via the STAT responsive element in the serine protease inhibitor 2.1 promoter. Mol Cell Endocrinol *130*, 69-81.
Ye, K., Schulz, M. H., Long, Q., Apweiler, R., and Ning, Z. (2009). Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics *25*, 2865-2871.
Yoshida, K., Toki, T., Okuno, Y., Kanezaki, R., Shiraishi, Y., Sato-Otsubo, A., Sanada, M., Park, M. J., Terui, K., Suzuki, H.*, et al.* (2013). The landscape of somatic mutations in Down syndrome-related myeloid disorders. Nat Genet *45*, 1293-1299.
Zhou, J. X., Yang, X., Ning, S., Wang, L., Wang, K., Zhang, Y., Yuan, F., Li, F., Zhuo, D. D., Tang, L., and Zhuo, D. (2017). Identification of KANSARL as the first cancer predisposition fusion gene specific to the population of European ancestry origin. Oncotarget *8*, 50594-50607.