Genetics, transcriptomics and murine models reveal alterations of erythroid master regulator activities in human erythroleukemia

Alexandre Fagnan[1], Frederik Otzen Bagger[2#], Maria Riera Piqué Borràs[2#], Cathy Ignacimouttou[1#], Alexis Caulier[3], Cécile K Lopez[1], Elie Robert[1], Robert Ivanek[2], Véronique Gelsi-Boyer[4], Zakia Aid[1], Cécile Thirant[1], Ute Moll[5], Amina Kurtovic[6], Jaroslaw Maciejewski[7], Christine Dierks[8], Alessandro Rambaldi[9], Thomas Pabst[10], Kazuya Shimoda[11], Hélène Lapillonne[12], Sébastien Malinge[1], Stéphane DeBotton[1], Jean-Baptiste Micoll[1], Benjamin Kile[13], Catherine Carmichael[13], Ilaria Iacobucci[14], Charles Mullighan[14], Martin Caroll[15], Peter Valent[16], Olivier A. Bernard[1], Eric Delabesse[17], Paresh Vyas[18], Daniel Birnbaum[4], Eduardo Anguita[19], Loïc Garcon[3], Eric Soler[20], Juerg Schwaller[2,&], Thomas Mercher[1,21,&]

[1]Institut Gustave Roussy, INSERM U1170, Université Paris Diderot, 94800 Villejuif, France,
[2]University Children's Hospital Beider Basel (UKBB) and Department of Biomedicine, University of Basel, 4031 Basel, Switzerland
[3]EA 4666 HEMATIM UPJV and Service Hématologie Biologique, CHU Amiens, Amiens, France,
[4]Centre de Recherche en Cancérologie de Marseille, Marseille, France
[5]Institute of Molecular Oncology, University Medical Center Göttingen, Göttingen 37077, Germany; Department of Pathology, Stony Brook University, Stony Brook, NY 11794, USA
[6]Research Institute of Molecular Pathology, Vienna, Austria
[7]Department of Translational Hematology and Oncologic Research, Cleveland Clinic Taussig Cancer Institute, Cleveland, United States
[8]Hämatologie, Onkologie und Stammzelltransplantation, Klinik für Innere Medizin I, Freiburg, Germany
[9]Università degli Studi di Milano and Ospedale Papa Giovanni XXIII, University of Milan, Bergamo, Italy
[10]Department of Hematology, Bern University Hospital, Bern, Switzerland,
[11]University Of Miyazaki Hospital, Miyazaki, Japan
[12]Sorbonne Université, CRSA – Unité INSERM, AP-HP, Hôpital Trousseau, 75012 Paris, France
[13]Monash University Biomedicine Discovery Institute, Clayton, Australia
[14]Department of Pathology, St. Jude Children's Research Hospital, Memphis, TN, USA
[15]Division of Hematology and Oncology, University of Pennsylvania, Pennsylvania, United States
[16]Department of Internal Medicine I, Division of Hematology and Hemostaseology and Ludwig Boltzmann Institute for Hematology and Oncology, Medical University of Vienna, Vienna, Austria
[17]INSERM U1037, Team 16, Center of Research of Cancerology of Toulouse, Hematology laboratory, IUCT-Oncopole, France
[18]Radcliffe Department of Medicine, Oxford university, United Kingdom
[19]Servicio de Hematología y Hemoterapia/Haematology Department, Hospital Clínico San Carlos, Madrid, Spain
[20]Institut de Génétique Moléculaire de Montpellier, Montpellier, France
[21]Equipe labellisée Ligue Nationale Contre le Cancer, 75013 Paris, France

[#]Equal contribution as 2nd author
[&]Equal contribution as senior author

To whom correspondence should be addressed:
Thomas Mercher
Institut Gustave Roussy, INSERM U1170
39 rue Camille Desmoulins, 94800 Villejuif
Phone: +33 1 42 11 44 83
E-mail: thomas.mercher@inserm.fr
or
Juerg Schwaller
University Children's Hospital Basel (UKBB)
Department of Biomedicine (DBM), University of Basel
ZLF-Lab 202
Hebelstrasse 20
CH-4031 Basel, Switzerland
Phone: +41 61 265 3504
E-mail: j.schwaller@unibas.ch

Running title: Altered transcription activities in human erythroleukemia
Keywords: AEL, MDS, AML, GATA1, Transcription, Models
Number of figures: 7

ABSTRACT

Acute erythroleukemia (AML-M6 or AEL) is a rare but aggressive hematologic malignancy. Previous studies showed that AEL often carry complex karyotypes and mutations in known AML-associated genes. However, the underlying molecular mechanisms driving the erythroid phenotype remain poorly understood. Here, we report the genetic and transcriptional characterization in a series of 57 AEL patients. Our data suggests at least three genetic AEL sub-groups including patients carrying *TP53* mutations (~30%), epigenetic mutations such as in *DNMT3A, TET2* or *IDH2* (~30%), and others with fewer known mutations. To distinguish the myeloid and erythroid features of AEL, we established a transcriptomics-based space in which, independently of the genetic subgroup, a significant fraction of AEL patients clustered closest to the erythroid lineage, apart from MDS and other AML types. Interestingly, >25% of AEL patients showed aberrant expression of transcriptional regulators, including *SKI*, *ERG*, and *ETO2,* that are related to the activity of the erythroid master regulator GATA1. Expression of these factors immortalized murine erythroid progenitors *in vitro* and led to erythroid or mixed erythroid/myeloid proliferations *in vivo*, phenocopying several aspects of human AEL. Collectively, our data indicates that AEL is a genetically heterogeneous disease with an erythroid identity that results in part from the aberrant activity of key erythroid transcription factors.

INTRODUCTION

Acute myeloid leukemia (AML) of the erythroid lineage (Erythroleukemia, AML-M6 or AEL) accounts for 3-5% of the patients and is inherently associated with poor outcome, however the disease-driving molecular mechanisms remain poorly understood [1–3]. Although AEL can occur at any age, the majority of the patients is >65 years, and the disease often occurs secondary to other neoplasms, including *JAK2*$^{V617F+}$ myeloproliferative neoplasms (MPN), or after genotoxic cancer treatment. Two major morphological subtypes have been proposed: pure erythroleukemia (PEL, AML-M6b, a.k.a. “Di Giuglielmo disease”) with >80% of blasts committed to the erythroid lineage; and AML-M6a characterized by the presence of erythroid precursors and myeloid blasts [1–3]. Notably, although the 2016 WHO classification integrated AML-M6a into myelodysplastic syndromes (MDS) or not otherwise specified AML (AML-NOS), the revised classification remains a matter of debates [4–6].

In AML, functional studies suggested that 2-5 genetic driver lesions on a ground of pre-existing alterations in HSPC might be sufficient to induce disease [7,8]. For AEL, earlier work showed that leukemic cells often have complex karyotypes with alterations and targeted DNA sequencing revealed the presence of several known AML-associated mutations [9–12] but erythroleukemia-specific mutations have rarely been functionally validated. Strikingly, single or multiple TP53 mutations have been shown to be a molecular hallmark of PEL [13].

Normal erythroid differentiation is controlled by the activity of both extrinsic factors, including erythropoietin (EPO) mediating its effects through the EPOR signaling pathways, and the intrinsic multimeric transcription factors complexes, including hematopoietic master regulators like GATA-binding protein 1 (GATA1), T-cell acute lymphocytic leukemia protein 1 (TAL1), LIM domain-only 2 (LMO2) and CBFA2/RUNX1 partner transcriptional co-repressor 3 2 (CBFA2T3, a.k.a. ETO2) and LIM-domain-binding protein 1 (LDB1), thereafter broadly named "GATA1-complexes", that can activate or repress transcription of target genes [14,15]. The

LDB1/GATA1-complex has been shown to contribute to terminal erythroid differentiation through binding to gene loci and induction of essential erythroid regulators (e.g. hemoglobin). This process is also regulated by Kruppel-like Factor 1 (KLF1), which binds DNA next to the LDB1/GATA1 complex to coregulate erythroid genes [16,17]. Functional synergism between these transcriptional regulators and the EPO/EPOR signaling to establish the erythroid differentiation program could be mediated by the presence of phosphorylated STAT5 binding in the neighborhood of GATA1 and KLF1 [18,19]. Accordingly, mutations in these factors have been associated with human altered erythropoiesis [20,21]. For example, GATA-1 mutations are associated with congenital erythroid hypoplasia (Diamond-Blackfan anemia; DBA) or X-linked dyserythropoietic anemia [22]. Alterations in regulators of GATA1-complexes activity, like ETO2 involved in a NFIA-ETO2 fusion in pediatric PEL [23] suggest that altered activity of these complexes may contribute to human erythroid leukemogenesis.

To better understand the molecular mechanisms that control the erythroid identity, we characterized the genetic and transcriptional landscape in a series of 57 patients diagnosed with AEL. We found genetic subgroups composed of patients carrying 1-*TP53* mutations (~1/3 of patients), 2-various combinations of mutations previously found in AML and MDS such as *DNMT3A, TET2* or *IDH2* (~1/3 of patients), and 3- those with none of these alterations and an average lower number of mutations. Comparative transcriptomics allowed to establish an erythro-myeloid differentiation expression signature space that distinguished most AEL from MDS or AML patient samples. Notably, leukemic cells from several AEL patients expressed aberrantly high levels of multiple transcription factors involved in the GATA1 complex as SKI, ERG or ETO2. Depending on the cellular hierarchy, experimental expression in hematopoietic stem and progenitor cells induced lethal erythroid or mixed erythroid/myeloid diseases in mice phenocopying several aspects of human AEL.

MATERIALS AND METHODS

Patients samples

Human patient samples were obtained with the informed consent of the patient in accordance with national ethics rules. AEL patient diagnostics were established according to the WHO 2008 classification. Patient blood or bone marrow (BM) samples and clinical data were collected with the informed consent of the patient accordingly with the national ethic rules. Mononuclear cell fractions were obtained from patient samples by using Ficoll gradient, and frozen in FBS supplemented with 10% DMSO. DNA extraction, RNA extraction and immune phenotype were done on fresh or frozen samples. DNA was extracted using bulk or sorted-cells ($CD36^+$ for blast cell population and $CD3^+$ or $CD19^+$ for non-tumor cell population). RNA was extracted from patient samples (n=17) and xenografted samples (n=4) from bulk or sorted-cells ($CD36^+$ or $CD45^+$ cells) respectively.

Murine models

C57BL/6JOlaHsd mice were purchased from Envigo and *NOD.Cg-Prkdc^{scid}Il2rg^{tm1Wjl}/SzJ* (NSG) mice from the Jackson Laboratory (005557). *TP53^{R248Q}* knock-in mice were described previously [24]. To generate double transgenic TET2$^{-/-}$/GATA1s mice, we inter-crossed *Tet2$^{-/-}$* and *Gata1$^{\Delta e2}$* (here named *Gata1s)* mice [25,26]. Mice were maintained at the Gustave Roussy preclinical facility and all experiments were approved by the French national animal care and use committee (CEEA#26: projects 2017-082-12726 and 2017-084-12799).

Flow cytometry and cell sorting

Antibodies used for flow cytometry are listed in a **Supplementary Table 1**. Cells were stained in 1X phosphate buffered saline (PBS) supplemented with 2% fetal bovine serum (FBS, Gibco) at 4°C for 30min and washed prior analysis. Whole BM or spleen cells were analysed without

red blood cell lysis. For cell sorting, total BM cells underwent red blood cells lysis. To obtain hematopoietic stem and progenitor cells (HSPC), total BM was depleted from the major haematopoietic cell lineage (Lin^-) using the Mouse Haematopoietic Progenitor (Stem) Cell enrichment Set [Becton Dickinson (BD), 558451]. Progenitors populations were further purified by FACS according to the following phenotypes: Hematopoietic stem cells (HSC) were defined as $Lin^-/Sca1^+/KIT^+/CD34^-/CD48^-$, megakaryocytic-erythroid progenitors (MEP) were defined as Lin^- $Sca1^-$ $KIT^+/CD34^-/CD16/32^-$ and granulocyte-macrophage progenitors (GMP) were defined as $Lin^-/Sca1^-/KIT^+/CD34^+/CD16/32^+$. To obtain mouse erythroid progenitors, BM cells were first depleted using biotin-conjugated antibodies against CD3, B220, Gr-1, and CD11b (BD) followed by fluorescence-activated cell sorting (FACS) according to the population described as $CD71^+/Ter119^+/KIT^+$. Flow cytometric analysis was performed by using ARIAII, CANTO-II or CANTO-X instruments (BD), and data were analysed using the FlowJo software (Flowjo 9.3.2).

Cell Culture

Erythroid progenitor cells were expanded in StemSpan™ serum-free expansion medium (SFEM, Stem Cell Technologies) supplemented with Penicillin (100U/mL)-Streptomycin (100µg/mL), mSCF (10ng/mL), mIL3 (10ng/mL), mIL6 (10ng/mL), hEPO (2U/mL), 0.4% Cholesterol and dexamethasone (10^{-6} M). Mouse erythroleukemia (MEL) cells were maintained in RPMI-1640 (Gibco) supplemented with 10% FBS, Penicillin (100U/mL)-Streptomycin (100µg/mL) and 2mM L-Glutamine (Gibco). The mouse cell line Ba/F3 was maintained in RPMI 1640 (Gibco) supplemented with 10% FBS, Penicillin (100U/mL)-Streptomycin (100µg/mL), mIL3 (10ng/mL) and 2mM L-Glutamine (Gibco). Human embryonic kidney (HEK-293T) cells were grown in DMEM (Gibco) supplemented with 10% FBS, Penicillin (100U/mL)-Streptomycin (100µg/mL) and 2mM L-Glutamine (Gibco).

Retroviral constructs, particles production and cell transduction

cDNA constructs were cloned into the retroviral *pMSCV-IRES-EGFP* or *-mCherry* backbone. The *SKI* cDNA is a kind gift from Dr. Suzana Atansoski (Basel, Switzerland). The other cDNA were synthesized. For retroviral particles production, 2 million HEK-293T cells were plated one day before co-transfection with the expression constructs co-expressing EGFP or mCherry and mouse *SKI*, human *ERG*, *ETO2*, *GATA1s* ORFs using the X-tremeGENE™-9 DNA Transfection Reagent (Roche), accordingly to manufacturer recommendations. Culture media was changed 24h post-transfection and supernatants containing viral particles were harvested 48h and 72h post-transfection. Murine cells were transduced by spinoculation (90 minutes at 2500 rpm, 33°C) with supernatants containing viral particles supplemented with 5µg/mL polybrene in 7.5mM HEPES buffer.

Bone marrow transplantation

Total BM ($0.4x10^6$ cells) and/or transduced progenitor cells were transplanted through intravenous injection in lethally (9.5Gy) or sub-lethally (5Gy) irradiated 8-10-week-old *C57BL/6J* recipient mice.

Single cell differentiation assay

HSC, GMP and MEP progenitors were sorted and incubated for 2h at 37°C before spinoculation with retrovirus expressing SKI or an empty control. Single GFP^+ cells were sorted 24h after transduction into 96-well plates and maintained in RPMI-1640 supplemented with mSCF, mIL3, mIL6 and hEPO. Clones were counted and differentiation was evaluated by flow cytometry analysis after 5 to 7 days.

RNA extraction and RT-qPCR

RNA was extracted by using RNeasy Mini Kit (Qiagen) or AllPrep DNA/RNA Mini Kit (Qiagen), accordingly to manufacturer recommendations and quantified using NanoDrop (ThermoScientific). Reverse transcription was done using SuperScript II (Invitrogen). Quantitative PCR was performed using SYBR Select Master mix or Taqman Gene expression Master mix (Applied Biosystems) on a 7500HT Fast Real-Time PCR System (Applied Biosystems) following manufacturer's recommendations. Primers sequences are listed in **Supplementary Table 2**.

Whole-exome sequencing

Whole-exome sequencing was conducted on paired-samples from 11 patients. DNA from sorted CD3+ or CD19+ non-tumor cells were used for exome capture using SureSelect All Exon V4, V5 kit (Agilent Technologies). We performed paired-end sequencing (100bp) using HiSeq2000 sequencing instruments at Gustave Roussy genomic platform. Reads were mapped to the reference genome hg19 using the Burrows–Wheeler Aligner (BWA) alignment tool version 0.7.10. PCR duplicates were removed using Picard tools – Mark Duplicates (version 1.119). Local realignment around indels and base quality score recalibration were performed using GATK 3.3 (Genome Analysis Tool Kit). Reads with a mapping quality score < 30 < 20 were removed. Somatic single-nucleotide variations (SNV) and indels were called using Varscan (v2.3.7). For candidate somatic mutations, the variants were adopted as candidate mutations when P value was <0.001 and allele frequency was <0.1 in the reference sample. Variants were annotated with Annovar (v141112). We excluded synonymous SNVs, variants located in intergenic, intronic, untranslated regions and non-coding RNA regions. The mean coverage in the targeted regions was respectively 85,4X and 91,2X for tumoral and control samples. Functional variants filtering was done by using the open platform Cancer Genome

Interpreter (CGI). Only known-variants or variants predicted as driver were kept and visualized in IGV (2.3.88).

RNA sequencing

RNA-seq was performed as described [27]. Sequences were aligned to the reference genome with TopHat2 version 2.0.9 using the following parameters --bowtie1 --fusion-search --library-type fr-firststrand --read-realign-edit-dist 0 -p 8 -r 50 (or 2.0.14 for mice datasets) and Bowtie1 version 1.0.0. The number of reads per genes (RefSeq database) was counted with HTSeq-count version 0.5.4p5 using the "union" mode. The counts were then normalized with the DESeq2 method, which takes-into-account the library size of each sample. DESeq2 was also used to detect differential expression signatures between two conditions. Molecular pathway and gene set enrichment analyses (GSEA) of differentially expressed genes were performed using GSEA (2.2.3), the Molecular Signatures Database (6.2, Broad Institute)

Transcription factor activity inference

Gene Regulatory Network inference: ARACNe-AP software was used to infer a Gene Regulatory Network using scRNAseq data from normal human progenitors to predict a list of target genes for each TF [28,29]. ARACNe was ran over the log2 normalized counts in bootstrap mode (100 iterations), with a p-value threshold of 1e-8 and a custom curated list of 2171 TFs. Therefore, the activity of each TF in a normal context was computed in a network. For each AEL sample, TF activities were inferred by interrogating this network with AEL transcriptome data and expressed as Normalized Enrichment Score (NES) using the R library viper, as described in the bioconductor package manual [30]. NES were used to test differential activity by t-test and p-value correction by Benjamini-Hochenberg (FDR cut-off at 0.05).

Datasets

Sequencing data has been submitted to EBI (ega-box-1242).

Statistical analysis

Statistical significance was calculated using Prism (version 6.0a) and is indicated as p values (Student's t test except when otherwise specified). *: $p<0.05$, **: $p<0.01$, ***: $p<0.001$.

RESULTS

Genetic alterations in AEL patients

We collected samples from 57 AEL patients, including 30 adults >60 years, 14 between 40-59 years, 6 young adults (21-39 years), and 7 patients not characterized by age, diagnosed according to the 2008 WHO classification. 29 patients were diagnosed with *de novo* AEL, including 26 AML-M6a and 3 AML-M6b, 15 patients were diagnosed secondary to MDS and a more precise diagnosis was lacking for 13 patients (**Figure 1A**, **Supplementary Table 3**). Thereafter, the term "AEL" was used for all patients. For several AEL samples lacking sufficient number of viable cells, we attempted to expand the tumor cells by xenografting them in *NOD.Cg-Prkdc*^{scid}*Il2rg*^{tm1Wjl}*/SzJ* (NSG) mice. This approach provided additional tumor material to isolate RNA (7 patients) and DNA (4 patients). Together, we performed exome sequencing on 11 paired leukemic and non-tumoral (either $CD3^+$ or $CD19^+$ cells from the same patient) samples and RNA-sequencing of 29 tumor samples. Combining exome and RNA-seq data, we identified sequence variants (as compared to the paired non-tumoral samples for exomes, and as compared to the reference genome for RNA-seq), and filtered them using the Cancer Genome Interpreter platform to retain only known variants or variants with predicted functional consequences in 62 genes **(Figure 1B**).

This genetic data including high variant allele frequency (**Figure 1C**) allowed us to assign patients into three molecular sub-groups (**Figure 1B**). First, subgroup-1 presenting with *TP53* mutations (n=12, 36.3% of patients) and an average number of 4.41 mutations per sample. Subgroup-2 (n=11, 33.3%) mostly presenting *TET2* nonsense mutations (n=8) and *DNMT3A* mutations (n=5), including 2 patients with both *TET2* and *DNMT3A* mutations, and an average number of 5.72 mutations per sample. Several patients with *TET2* and/or *DNMT3A* mutations also carried *SRSF2*$^{P95H/R}$ or *IDH2*^{R140Q} mutations. Of note, in the only sample presenting both a *TET2* and an *IDH2* mutation, the variant allele frequencies were 60% and 13% respectively,

indicating that we cannot exclude the possibility of having two independent clones in this sample (data not shown). Interestingly, one case (#17) of subgroup-2 harbored a TET2 loss-of-function mutation and a GATA1 mutation, predicted to encode for a GATA1s. Additional mutations affected transcription factors (e.g. *WT1, RUNX1*), epigenetic regulators (e.g. *ASXL1, EP300, BCOR*), signaling mediators (e.g. *NOTCH2, IL7R*) and other genes in this group of patients. Finally, subgroup-3 (n=10, 30.4%) was constituted by samples without *TP53*, *TET2/DNMT3A* mutations or recurrent known variants and an average number of 1.60 mutations per sample. Overall, our results confirmed that AEL is a heterogeneous disease and that the recurrent mutations present comparable incidence in several cohorts [11,27].

AEL transcriptional signatures correlate with erythroid differentiation but not genetics

As the genetic alterations did not directly explain the erythroid phenotype of these leukemia, we investigated the erythroid identity by comparing gene expression signatures (GES). We observed no significant correlation between the GES and the genetic subgroups (**Figure 2A**). Similarly, the immunophenotypes (determined prior to RNA extraction) also appeared poorly reflected by GES (**Figure 2B**).

As the AEL classification is based on amount of erythroid and myeloid BM blasts, we used the xCell program, a digital cellular deconvolution method, to compute an enrichment in erythroid, myeloid and other hematopoietic cell types (**Supplementary Figure 1A)** [28]. The majority of samples had the highest signal for the "erythrocyte" signature (n=20) while some AEL samples presented more signal for immature (MPP, CMP, GMP) or myeloid (monocyte, neutrophil) signature (n=9). To further explore the link between AEL patient GES and different stages of human erythroid maturation, we then compared our GES with those obtained experimentally after *in vitro* differentiation of peripheral blood mononuclear cells into colony forming unit erythroid ("Day2", $CD71^{+}/CD235^{-}$), pro-erythroblasts ("Day 4", $CD71^{+}/CD235^{low}$),

intermediate ("Day 7", CD71^{+}/CD235^{high}) and late erythroblasts ("Day 10", CD71^{low}/CD235^{high})[29] and observed that our AEL samples clustered according to different maturation stages (**Figure 2C, D**). Importantly, AEL transcriptomes from an independent patient cohort also clustered similarly (**Figure 2E, F**). Random forest classification was used to define top differentially expressed genes between these clusters including several well-known regulators of erythroid maturation (**Figure 2G**). Notably, GES similar to early stages of differentiation (CFU-E/Pro-E) expressed *EPOR*, *glycophorin-A* (*GYPA)*, *KLF1*, or *carbonic anhydrases* (*CA-1/2*) well known markers of early erythroid differentiation that were not seen in cases that were closer to later differentiation stages (**Figure 2G, Supplementary Table 4**).

Collectively, our genetic analysis indicates that AEL patients are characterized by a high prevalence of *TP53* mutations, followed by mutations affecting various genes involved in chromatin and splicing biology (e.g. *TET2*, *DNMT3A*, *IDH2*, *SRSF2*) that are commonly found in MDS, AML and other myeloid malignancies. However, the AEL gene expression program seems more influenced by the stage of erythroid differentiation than by the presence of particular genetic lesions, suggesting that the erythroid identity in human AEL might depend on the cellular origin and the activity of regulators driving cellular differentiation.

Transcriptome-based space to map AEL, MDS and other AML to erythroid and myeloid lineage trajectories and compute erythroid transcription factor activities

As the 2016 WHO classification assigns most cases previously diagnosed as AEL to MDS or AML [7], we aimed at designing a transcriptome-based space that is able to compare AEL, MDS and other AML forms. To this end, we retrieved cellular signatures from the ENCODE database [32] and computed erythroid and myeloid differentiation expression trajectories (**Figure 3A**). As expected, our AEL samples clustered between the erythroid and myeloid trajectories. However, although the distribution appeared more or less as a continuum,

we were able to define, at least, two distinct groups (cases above and below the 0 value of PC2) of which one group (17 cases) mapped closer to the erythroid axis and the second group mapping closer to the myeloid axis (n=6) (**Figure 3B**). The latter group appeared very similar to MDS-derived signatures and surprisingly even closer to the myeloid trajectory than the MDS cases retrieved from public expression databases. However, the majority of the AEL cases clustered in this space apart from MDS samples (**Figure 3B**). A similar distribution was observed after including GES from a recently published AEL patient cohort (**Figure 3C**). We then also included published GES from AML and APL patients that clustered apart from most AEL and closer to the myeloid trajectory (**Figure 3D**). Interestingly, among our AEL samples that mapped closer to the myeloid axis and other AML samples, sample #24 showed a high expression of *SPI1* (**Supplementary Figure 1B**) generally associated with other AML subtypes. These data support that, although some overlap exists, the transcriptional programs of the majority of AEL cases differs from those of MDS and other AML subtypes and also support the WHO-2016 reclassification of some AEL cases as AML-NOS.

Myelo-erythroid differentiation is controlled by expression and activity of a relatively small group of transcription factors. Using the ARACNe and VIPER packages [30,31] and a large dataset from human normal progenitor cells transcriptomes [32], we inferred lists of targets genes and computed the activity of transcription factors (**Supplementary Figure 2A, B**). Interestingly, in both, our, and the *Iacobucci et al*. AEL patient cohorts (respectively M6 cohort 1 and 2) we observed a gradual decrease in expression of erythroid transcription factors (e.g. KLF1, GATA1, NFE2, TAL1, NFIA) and their predicted activity when going from the erythroid to the myeloid trajectories and an inverse correlation with myeloid factors (e.g. CEPBA and SPI1) factors (**Figure 3E and Supplementary Figure 2C**). This confirmed that AEL can be defined based on their proximity to the normal erythroid lineage trajectory and the relative activity of master transcription factors that control erythroid differentiation.

Based on these data, we hypothesized that some AEL cases could be driven by aberrant expression and activity of erythroid transcription factors. Therefore, we focused on factors known to be predominantly expressed during erythroid differentiation and/or to control the activity of the GATA1 erythroid master regulator[14]. Using a threshold of 4-fold higher expression level than the average, we observed that several AEL patients aberrantly expressed high levels of *ERG* (n=2), *GFI1* (n=1), *RUNX1T1* (=1) and *CBFA2T3* (n=1) **(Figure 3F**). *GATA3*, which enforced expression resulted in erythroid bias[33], was also highly expressed in 3 samples. Notably, we also found aberrantly high expression of the transcriptional co-repressor *SKI* (*v-Ski avian sarcoma viral oncogene homolog*) in 2 patients from genetic subgroup-3 (**Figure 3F**). Interestingly, *v-Ski* was previously reported to transform chicken erythroid directly with GATA1 to repress erythroid differentiation and to enhance self-renewal of mouse HSC [34–37]. Therefore, our finding suggest that SKI not only influences experimental erythroid differentiation but also contributes to human AEL pathogenesis.

The detection of fusion transcripts using RNAseq data revealed additional alterations, including one in-frame *BCR-ABL1* fusion gene and two novel out-of-frame fusion transcripts, notably all of them in a *TP53*-mutated context (**Supplementary Figure 3A-C**). Sample #37 harbored an out-of-frame fusion of *YWHAE* (Tyrosine 3-Monooxygenase/Tryptophan 5-Monooxygenase Activation Protein Epsilon) with *EPO* and sample #ES3 showed an out-of-frame fusion of *HSD17B11* (Hydroxysteroid 17-Beta Dehydrogenase 11) with *B4GALNT3* (Beta-1,4-N-Acetyl-Galactosaminyltransferase 3). They were associated with ectopic expression of *EPO* and *B4GALNT3*, respectively. Although the role of B4GALNT3 in human erythropoiesis remains unclear, an out-of-frame fusion leading to overexpression of B4GALNT3 has been previously reported in thyroid carcinoma [38]. Sample #ES1 from subgroup-3 presented with an out-of-frame fusion targeting the middle of the *DNMT3B* locus and associated with lower *DNMT3B* expression as compared to other samples, supporting

DNMT3B inactivation. Notably, *Dnmt3a* and *Dnmt3b* expression were previously reported to be regulated during erythroid maturation in mice [39].

Overall, genetic and transcriptional alterations in factors erythroid regulators, including physical or functional interactors of the GATA1 transcriptional complexes were found in 9/33 patients (27%) (**Figure 3G**). Notably, these cases had a trend toward poorer overall survival (**Figure 3H**), which was confirmed after performing a similar classification in another AEL samples dataset (**Figure 3I**) [27]. Together, transcriptome analysis revealed that the majority of AEL samples are significantly different from MDS and other AML subtypes and that AEL frequently presents with epigenomic alterations that converge on aberrant GATA1-complexes.

Overexpression of AEL-associated erythroid transcription factors transforms mouse erythroid progenitors

To functionally test whether aberrant expression and activities of erythroid transcription factors found in AEL may contribute to the transformation of the erythroid lineage, we first explored the consequences of ectopic expression of *SKI*, *ERG*, *ETO2*, *GATA1s*, *EPO, SPI1* and *B4GALNT3* on proliferation of murine erythroid progenitors (**Figure 4A**). FACS-purified $KIT^+CD71^+Ter119^+$ progenitors were transduced with retroviruses encoding these genes and grown *in vitro* (**Figure 4B**). In contrast to vector-transduced controls that proliferated for only ~7 days, ectopic expression of *ERG*, *SPI1*, *ETO2*, *SKI, and B4GALNT3* significantly maintained proliferation of erythroid cells presenting with a $CD71^+KIT^+Ter119^-$ phenotype and a pro-erythroblast morphology for >30 days (**Figure 4B and 4C**). Notably, ectopic expression of *EPO* or *GATA1s* alone were not sufficient to expand erythroblasts longer than 10 days (**Figure 4B**). Quantitative analysis confirmed overexpression of each gene and the erythroid identity of the cells, by a higher expression of other erythroid regulators (*GATA1*, *ALAS2* and *NFE2*) compared to the murine erythroleukemia (MEL) cell line or to non-erythroid Ba/F3 cells serving as positive and

negative controls, respectively. [39] (**Figure 4D and Supplementary Figure 4A**). Notably, transplantation of ERG, ETO2 and SKI-transformed erythroblasts into irradiated syngeneic recipients rapidly induced a fully penetrant fatal AEL-like phenotype characterized mostly by accumulation of $CD71^{+}Ter119^{-}$ and few $CD71^{+}Ter119^{+}$ blasts both lacking expression of myeloid or lymphoid markers (**Figure 4E and 4F**). Histopathological analysis of symptomatic mice showed infiltration of BM, spleen and livers with typical erythroblast expressing nuclear GATA1 (**Figure 4G and Supplemental Figure 4B**). Collectively, these data demonstrate that ectopic expression of *ERG*, *ETO2*, *SKI* and *B4GALNT3* can efficiently immortalize murine erythroblasts *in vitro*.

TP53^{R248Q} mutation cooperate with *ERG* to transform erythroid progenitors *in vivo*.

To gain insights into the *in vivo* transformation potential of alterations found in the 3 major AEL molecular subgroups, we developed additional bone marrow transplantation (BMT) and transgenic models. In TP53-mutated AEL, aberrant *ERG* expression was observed in 2/12 samples (**Figure 3G**). Previous murine BMT approaches using adult donor cells indicated that *ERG* overexpression alone does not efficiently lead to AEL but rather to T-cell leukemia *in vivo* [40,41]. As the majority of AEL-associated *TP53* alterations are mostly gain-of-function (GOF) mutations affecting the DNA binding domain of the protein [13], we investigated the requirements for a cooperation between *TP53* GOF mutations (*TP53*^{R248Q}) and high *ERG* expression in AEL development. We transduced adult lineage-negative (Lin^{-}) HPSC from wild-type (WT) or *TP53*^{R248Q} knock-in mice with either empty (Ctrl) and *ERG*-expressing retroviruses and transplanted them into lethally irradiated syngeneic recipients. In both WT and *TP53*^{R248Q} contexts, *ERG* expression led to an abnormal accumulation of erythroid progenitors in the BM of recipient mice starting 4 weeks post-transplant. However, all mice later developed fatal T-cell leukemia (**Supplementary Figure 5A, B)**. To assess the long-term consequences of high

ERG expression in erythroid progenitors, we transplanted flow-sorted GFP$^+$ erythroblasts taken from primary recipients (28 days post-transplant) into sub-lethally irradiated secondary recipients (**Figure 5A**). In this setting, recipients of *TP53*^{R248Q} erythroblasts overexpressing *ERG* developed a fatal leukemia with a median survival of 60 days, while recipients of *ERG*-expressing wildtype erythroblasts developed disease after over 4 months (**Figure 5B**). The *TP53*^{R248Q} + *ERG*-induced disease was characterized by anemia, thrombocytopenia (**Figure 5C**) and the accumulation of CD71$^+$Ter119$^+$ erythroid and to a lesser extent CD11b$^+$Gr1$^+$ myeloid progenitors in the BM (**Figure 5D**), with infiltration in non-hematopoietic organs (**Figure 5E**). Together, these data indicate that a *TP53* GOF mutation can cooperate with other AEL-associated alterations, including aberrantly high *ERG* expression, to enhance the proliferative capacity of erythroid progenitors leading to leukemia with several features of the human disease.

Cooperation of *Tet2* loss-of-function mutations with *Gata1s*

Concomitant mutations in *GATA1* and epigenetic regulators was found in an AEL patient from subgroup-2 (*GATA1* and *TET2*, **Figure 1B**) and was also reported recently (*GATA1* and *IDH2*) [27]. Mutations leading to expression of *GATA1s* impair erythroid differentiation but do not lead to the development of a leukemic phenotype *in vivo* [26,42]. Similarly, *Tet2* loss-of-function mutations have pleiotropic effects on hematopoiesis, including alterations of the erythroid differentiation in zebrafish and mice but are generally also not sufficient to induce a *bona fide* leukemia in mice [25,43–46]. To address whether cooperation between *Gata1s* and *Tet2* inactivating mutations *in vivo* may be sufficient to induce erythroleukemia, we crossed *Tet2*-deficient [25] (thereafter named *Tet2*$^{-/-}$) mice with *Gata1*$^{\Delta e2}$ knock-in [26] (thereafter named *Gata1s*) mice **[25,26]**. We flow-purified erythroid progenitor cells from wild-type, *Tet2*$^{-/-}$ only, *Gata1s* only and *Tet2*$^{-/-}$+*Gata1s* mice and compared *in vitro* proliferation (**Figure 6A**). Only *Tet2*$^{-/-}$+*Gata1s*

erythroblasts proliferated over two months whereas *Gata1s* or *Tet2*$^{-/-}$ erythroblasts could not be expanded >10-15 days (**Figure 6B**). Expanded cells exhibited high CD71 expression and an erythroid morphology (**Figure 6B**). Injection of these CD45.2^{+} *Tet2*$^{-/-}$+*Gata1s* erythroid cells into sublethally-irradiated CD45.1^{+} recipients rapidly induced a fatal leukemia associated with infiltration of mostly CD71^{+}Ter119$^{+/-}$ CD45.2^{+} cells (**Figure 6C**). Histopathological analysis confirmed the infiltration with GATA1^{+} erythroblasts in spleens and livers (**Figure 6D**).

As primary *Tet2*$^{-/-}$+*Gata1s* mice did not develop hematological malignancies or alterations of myeloid and erythroid lineage differentiation during a follow-up period of 14 months (**Supplementary Figure 5C, D**), we investigated their competitive fitness by transplanting 0.5x10^{6} Lin^{-} HSPC into lethally irradiated recipients (**Figure 6E**). As opposed to recipients of *Tet2*$^{-/-}$-only cells, recipients of *Tet2*$^{-/-}$+*Gata1s* cells developed a rapid and fully penetrant lethal disease associated with anemia, cytopenia and splenomegaly (**Figure 6F, G**). Flow cytometry analysis indicated that leukemic blasts were primarily CD11b^{+}Gr1^{+} myeloid cells (**Figure 6H**). Histopathological analysis confirmed that BM, spleen and liver were highly infiltrated by blasts with myeloid features (**Supplementary Figure 5E**). Of note, we also observed emperipolesis that was previously described in murine GATA1s models. Together, these data demonstrate that *Tet2* loss-of-function and *Gata1s* mutations cooperate to transform murine erythroid progenitors *in vitro*, and promote AML *in vivo*.

High expression of SKI induces an MDS-like phenotype with erythroid features from HSC and MEP but not GMP

Recurrent overexpression of *SKI* in subgroup-3 (**Figure 3G**), led us to investigate its role on *in vivo* disease development. First, we confirmed previous results[36] showing that transplantation of Lin^{-} HSPC retrovirally overexpressing *SKI* induced a lethal disease (**Supplementary Figure 6A**) characterized by anemia, thrombocytopenia, and increased myeloid cells in the periphery

(**Supplementary Figure 6B**) associated with hypercellular BM and spleens showing high percentage of mostly myeloid or erythroid GFP^+ cells (**Supplementary Figure 6C-E**). Notably, the presence of virus in myeloid, erythroid cells and platelets (**Supplementary Figure 6F**) suggested that *SKI* overexpression may affect early multipotent stem or progenitor cells.

To investigate whether the transforming activity of SKI depends on the hematopoietic target cell, we purified, transduced and transplanted long-term multipotent HSC, erythroid-enriched (MEP) or myeloid-committed granulocyte-macrophage (GMP) progenitors (**Figure 7A**). Three weeks post-transplant, transduced cells were detectable in the blood for all three conditions (**Supplementary Figure 6G**). However, transduced cells were only found in the BM of recipients transplanted with *SKI*-transduced HSC and MEP, but not GMP (**Supplementary Figure 6G**). Later, recipients of *SKI*-transduced HSC and MEP, but not of GMP, developed symptomatic diseases (**Figure 7B**) presenting with anemia, thrombocytopenia and splenomegaly (**Supplementary Figure 6H**) and presenting with both $CD11b^+Gr1^+$ myeloid and $CD71^+Ter119^+$ erythroid features (**Figure 7C-E**). Symptomatic recipients of HSC or MEP showed an increase in basophilic, polychromatophilic and orthochromatic erythroblasts and in reticulocytes associated with a relative decrease of mature red cells (**Figure 7F**), suggesting that *SKI* delays but does not fully block erythroid differentiation *in vivo*. Histopathological analyses confirmed the BM hypercellularity and revealed infiltration of erythroid cells in the spleen and liver (**Figure 7G**).

Taken together, these data indicate that aberrantly high *SKI* expression transforms various cells of the hematopoietic hierarchy including HSC and MEP, but not myeloid-restricted progenitors like GMP. While aberrant *SKI* expression in erythroid-restricted progenitors leads to pure erythroid proliferation, expression in more immature HSPC resulted in increased self-renewal capacity with aberrant differentiation toward both myeloid and erythroid lineages indicative of an AEL/MDS-like disease.

Discussion

AEL is an aggressive human cancer, often difficult to diagnose due to the close resemblance with other forms of hematopoietic malignancies presenting with variable compositions of cells with erythroid features, like MDS or certain AML subtypes. We found several novel features of this disease. Firstly, genetic and transcriptional data indicate that the majority of human AEL cluster apart from MDS but closer to the erythroid differentiation trajectory. Secondly, aberrant expression of various transcriptional regulators known to modulate GATA1 activity is frequently found in AEL and may represent a common molecular module that controls erythroid identity. Thirdly, *in vivo* models demonstrate that the relative composition of the erythroid and myeloid features is strongly dependent on the type of hematopoietic cell in which a driving oncogene is expressed, providing a basis for a better understanding the highly heterogeneous clinical appearance of the disease.

The genetic data described here are in line with previous reports and the most recent study of the genetic landscape of pediatric and adult erythroleukemia [9–12,27]. Collectively, all these studies support the idea that AEL patients can be classified into different genetic subgroups, including patients with *TP53* mutations (~1/3 of our cases), patients with genetic alterations in genes previously associated with clonal hematopoiesis of indeterminate potential and MDS (e.g. *DNMT3A*, *TET2* and *IDH1/2* mutations) (~1/3 of our cases) and another group of patients presenting with none of these recurrent alterations. Consistent with the lack of pediatric patients in our cohort, we did not observe any cases of the other recently described subgroups including those with *NUP98*, *KMT2A* and other in-frame fusions [27]. In older adult AEL, we rather observed out-of-frame fusion transcripts associated with altered expression of one of the partner genes not reported before. For example, the fusion between *YWHAE* and *EPO* in a TP53-mutated patient, is associated with ectopic expression of *EPO*, and the concomitant high expression of *EPOR* suggesting an autocrine EPO/EPOR signaling

mechanism [15]. Interestingly, alterations of multiple signaling intermediates, including downstream of EPO/EPOR, were recently found in up to 48% of human AEL samples[27] and acquired activating KIT mutations were also essential to induce a bona fide erythroleukemia in a SPI1/PU.1 transgenic murine model [47], indicating their importance oncogenic transformation in the erythroid lineage.

As the vast majority of these AEL-associated mutations are also found in a wide spectrum of human myeloid malignancies, it is essential to gain insights into their functional role in the definition of the erythroid phenotype that leads to a diagnosis of AEL. Here, we identified genetic alterations targeting erythroid master regulators, including a *YWHAE-EPO* out-of-frame fusion and a mutation in the erythroid transcription factor GATA1, that represent molecular candidates underlying the erythroid phenotype. Importantly, the recently reported *APLP2-EPOR* and *MYB1-GATA1* fusion genes [27] are predicted to result in similar functional consequences underlining their importance for the biology of AEL.

The mutational landscape and the gene expression profiles revealed that many AEL-associated alterations may ultimately interfere with GATA1 activity either through direct interaction within the GATA1 transcriptional complexes (e.g. aberrantly expressed *ETO2*, *ERG*, *SKI*) or through functional interaction (e.g. *SPI1*, *KLF1*) [14,48,49]. As shown previously for *SPI1* [50], and described here in a functional *in vitro* transformation assay using murine erythroblasts, erythroid progenitors are susceptible to transformation upon altered expression of factors that regulate GATA1 activity, including *CBFA2T3*, *ERG* and *SKI*. Interestingly, several epigenomic alterations that functionally impair erythroid master may ultimately represent a common theme that defines lineage identity. Indeed, a novel signaling pathway based on JAK2-mediated phosphorylation of TET2 leading to interaction with KLF1 was recently reported [51]. Combined TET2 and DNMT3A inactivation was also reported to upregulate expression of *KLF1* and *EPOR* in HSC [44]. Therefore, the concomitant *TET2* and *DNMT3A* mutations observed in two AEL

patients and the presence of *TET2* and *GATA1s* mutations in another AEL sample may result in a functional synergism between alterations of KLF1 and GATA1 transcriptional programs leading to erythroid differentiation blockage. Based on these observations, we hypothesize that the erythroid phenotype in AEL is the consequence of cooperation of genetic and transcriptional alterations. As proposed for other subtypes of leukemia, interference with the activity of altered erythroid master regulators, for example through targeting of critical protein-protein interactions may therefore represent promising therapeutic strategies for AEL [52,53].

Our observations may also have implications for the classification of AEL patients that is currently based on morphology indicating the predominantly affected lineage. Comparative analysis of AEL expression signatures with normal erythroid and myeloid differentiation indicated that AEL is heterogeneously spread along the differentiation with some patient samples being clustering next to progenitors retaining myeloid features and other patient samples being closer to the erythroid trajectory. Also, while several oncogenes (e.g. *SKI*) can transform restricted erythroid lineages, they led to mixed erythroid/myeloid hematopoietic malignancies upon expression in multipotent murine progenitors. These data indicate that the relative composition of myeloid vs. erythroid elements at time of diagnosis is not solely based on the type of genetic alterations but likely also reflects the type of progenitor targeted by these mutations. Notably, the relation of the gene expression signatures to normal differentiation trajectories was not clearly visible when comparing the reported immunophenotypes of the blasts and no correlation was found with the different genetic subgroups. These data strongly suggest that, in some AEL patients, the erythroid phenotype maybe initiated either by strong genetic alterations that interfere with erythroid differentiation, or by mutations that provide advantages to erythroid-restricted progenitors, while, in others, the erythroid phenotype may originate from genetic alterations in multipotent progenitors with a subsequent epigenetic drift toward the erythroid lineage.

The relevance of these observations likely extends beyond AEL as several of the genetic alterations are also found in other AML subtypes. Indeed, the *GATA1s* mutation is also strongly associated with Down's syndrome acute megakaryoblastic leukemia (AMKL) [54]. Interestingly, development of normal and malignant megakaryocytes shares dependencies with the erythroid lineages on GATA1 function [22,53,55–57]. These observations suggest that epigenomic alterations that functionally affect GATA1 activity can have cell-context depend consequences and may also underlie the confounding effects on lineage allocation at diagnosis in several AML cases. Although, the rarity of fresh AEL samples available to us did not allow to prospectively purify the various hematopoietic progenitors and assess clonal architecture, the better definition of the molecular mechanisms of erythroid identity in AEL is a first step toward the development of improved therapies.

ACKNOWLEDGEMENTS

We are grateful to Olivier Hermine and Michaela Fontenay for expert advices and useful discussions. We also thank Suzana Atanasoski (Basel, Switzerland) for providing the human *SKI* cDNA. Handling, conditioning and storing of one patient samples was performed by the FILOtheque (N° BB-0033-00073), tumor bank of the FILO group (Cochin hospital, Paris).

This work was supported by Institut National Du Cancer (PLBIO-2014-176 and PLBIO-2018-169 to TM), Ligue contre le cancer (AF: PhD grant, TM: équipe labelisée), INCa-PlanCancer ‘‘Soutien pour la formation’’ 2009–2013 (to C.I.), Cancéropôle Ile-de-France (to CKL 2014-2017), SIRIC-SOCRATE (INCa-DGOS-INSERM 6043 to TM), Fondation pour la Recherche Médicale (to CI, ZA: FRM-ING20150532273, CKL), Fondation de France (FdF-00057925 to CT and TM), Gustave Roussy Genomic Core Facility (Taxe d'apprentissage TA2018-ALFA (to AF). The work was also supported by grants from the Swiss Cancer League (KFS-3487-08-2014), the Gertrude von Meissner Foundation (Basel, Switzerland), the San Salvatore Foundation (Lugano, Switzerland) and the Wilhelm Sander Foundation (Munich, Germany) to JS.

AUTHORSHIP CONTRIBUTIONS

Performed and analyzed experiments:

Alexandre Fagnan, Maria Riera Piqué Borràs, Cathy Ignacimouttou, Alexis Caulier, Cécile K Lopez, Zakia Aid, Cécile Thirant, Loïc Garcon, Eric Soler, Jürg Schwaller, Thomas Mercher

Performed bioinformatics analyses:

Alexandre Fagnan, Maria Riera Piqué Borràs, Frederik Otzen Bagger, Robert Ivanek, Elie Robert

Provided patient samples and clinical information:

Véronique Gelsi-Boyer, Amina Kurtovic, Jaroslaw Maciejewski, Christine Dierks, Alessandro Rambaldi, Thomas Pabst, Kazuya Shimoda, Hélène Lapillonne, Stéphane DeBotton, Jean-Baptiste Micoll, Ilaria Iacobucci, Charles Mullighan, Benjamin Kile, Catherine Carmichael, Martin Caroll, Peter Valent, Eric Delabesse, Paresh Vyas, Daniel Birnbaum, Eduardo Anguita,

Provided major intellectual inputs and reagents:

Alexandre Fagnan, Ute Moll, Olivier A. Bernard, Daniel Birnbaum, Eduardo Anguita, Loïc Garcon, Eric Soler, Jürg Schwaller, Thomas Mercher

Conceived, supervised the project and drafted the manuscript: Thomas Mercher, Jürg Schwaller.

All authors revised and approved the final version of the manuscript.

REFERENCES

1. Hasserjian RP, Zuo Z, Garcia C, et al. Acute erythroid leukemia: a reassessment using criteria refined in the 2008 WHO classification. *Blood*. 2010;115(10):1985–1992.
2. Wang SA, Hasserjian RP. Acute Erythroleukemias, Acute Megakaryoblastic Leukemias, and Reactive Mimics: A Guide to a Number of Perplexing Entities. *Am. J. Clin. Pathol.* 2015;144(1):44–60.
3. Boddu P, Benton CB, Wang W, et al. Erythroleukemia-historical perspectives and recent advances in diagnosis and management. *Blood Rev.* 2018;32(2):96–105.
4. Arber DA, Orazi A, Hasserjian R, et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood*. 2016;127(20):2391–2405.
5. Arber DA. Revisiting erythroleukemia. *Curr. Opin. Hematol.* 2017;24(2):146–151.
6. Qiu S, Jiang E, Wei H, et al. An analysis of 97 previously diagnosed de novo adult acute erythroid leukemia patients following the 2016 revision to World Health Organization classification. *BMC Cancer*. 2017;17(1):534.
7. Cancer Genome Atlas Research Network, Ley TJ, Miller C, et al. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N. Engl. J. Med.* 2013;368(22):2059–2074.
8. Papaemmanuil E, Gerstung M, Bullinger L, et al. Genomic Classification and Prognosis in Acute Myeloid Leukemia. *N. Engl. J. Med.* 2016;374(23):2209–2221.
9. Grossmann V, Bacher U, Haferlach C, et al. Acute erythroid leukemia (AEL) can be separated into distinct prognostic subsets based on cytogenetic and molecular genetic characteristics. *Leukemia*. 2013;27(9):1940–1943.
10. Cervera N, Carbuccia N, Garnier S, et al. Molecular characterization of acute erythroid leukemia (M6-AML) using targeted next-generation sequencing. *Leukemia*. 2016;30(4):966–970.
11. Cervera N, Carbuccia N, Mozziconacci M-J, et al. Revisiting gene mutations and prognosis of ex-M6a-acute erythroid leukemia with regard to the new WHO classification. *Blood Cancer J*. 2017;7(8):e594.
12. Ping N, Sun A, Song Y, et al. Exome sequencing identifies highly recurrent somatic GATA2 and CEBPA mutations in acute erythroid leukemia. *Leukemia*. 2017;31(1):195–202.
13. Montalban-Bravo G, Benton CB, Wang SA, et al. More than 1 TP53 abnormality is a dominant characteristic of pure erythroid leukemia. *Blood*. 2017;129(18):2584–2587.
14. Kerenyi MA, Orkin SH. Networking erythropoiesis. *J. Exp. Med.* 2010;207(12):2537–2541.
15. Kuhrt D, Wojchowski DM. Emerging EPO and EPO receptor regulators and signal transducers. *Blood*. 2015;125(23):3536–3541.
16. Soler E, Andrieu-Soler C, de Boer E, et al. The genome-wide dynamics of the binding of Ldb1 complexes during erythroid differentiation. *Genes Dev.* 2010;24(3):277–289.
17. Li L, Freudenberg J, Cui K, et al. Ldb1-nucleated transcription complexes function as primary mediators of global erythroid gene activation. *Blood*. 2013;121(22):4575–4585.
18. Gillinder KR, Tuckey H, Bell CC, et al. Direct targets of pSTAT5 signalling in erythropoiesis. *PLoS ONE*. 2017;12(7):e0180922.
19. Perreault AA, Benton ML, Koury MJ, Brandt SJ, Venters BJ. Epo reprograms the epigenome of erythroid cells. *Exp. Hematol.* 2017;51:47–62.
20. Singleton BK, Frayne J, Anstee DJ. Blood group phenotypes resulting from mutations in erythroid transcription factors. *Curr. Opin. Hematol.* 2012;19(6):486–493.
21. Doshi BS, Abramowsky C, Briones M, Bunting ST. Concomitant a novel ALAS2 mutation and GATA1 mutation in a newborn: a case report and review of the literature. *Am J Blood Res*. 2014;4(1):41–45.

22. Crispino JD, Horwitz MS. GATA factor mutations in hematologic disease. *Blood*. 2017;129(15):2103–2110.
23. Micci F, Thorsen J, Panagopoulos I, et al. High-throughput sequencing identifies an NFIA/CBFA2T3 fusion gene in acute erythroid leukemia with t(1;16)(p31;q24). *Leukemia*. 2013;27(4):980–982.
24. Alexandrova EM, Yallowitz AR, Li D, et al. Improving survival by exploiting tumour dependence on stabilized mutant p53 for treatment. *Nature*. 2015;523(7560):352–356.
25. Quivoron C, Couronné L, Della Valle V, et al. TET2 inactivation results in pleiotropic hematopoietic abnormalities in mouse and is a recurrent event during human lymphomagenesis. *Cancer Cell*. 2011;20(1):25–38.
26. Li Z, Godinho FJ, Klusmann J-H, et al. Developmental stage-selective effect of somatically mutated leukemogenic transcription factor GATA1. *Nat. Genet.* 2005;37(6):613–619.
27. Iacobucci I, Wen J, Meggendorfer M, et al. Genomic subtyping and therapeutic targeting of acute erythroleukemia. *Nat. Genet.* 2019;51(4):694–704.
28. Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* 2017;18(1):220.
29. Merryweather-Clarke AT, Atzberger A, Soneji S, et al. Global gene expression analysis of human erythroid progenitors. *Blood*. 2011;117(13):e96-108.
30. Lachmann A, Giorgi FM, Lopez G, Califano A. ARACNe-AP: gene network reverse engineering through adaptive partitioning inference of mutual information. *Bioinformatics*. 2016;32(14):2233–2235.
31. Alvarez MJ, Shen Y, Giorgi FM, et al. Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nat. Genet.* 2016;48(8):838–847.
32. Velten L, Haas SF, Raffel S, et al. Human haematopoietic stem cell lineage commitment is a continuous process. *Nat. Cell Biol.* 2017;19(4):271–281.
33. Chen D, Zhang G. Enforced expression of the GATA-3 transcription factor affects cell fate decisions in hematopoiesis. *Exp. Hematol.* 2001;29(8):971–980.
34. Larsen J, Meyer S, Steinlein P, Beug H, Hayman MJ. Transformation of chicken bone marrow cells by the v-ski oncogene. *Oncogene*. 1993;8(12):3221–3228.
35. Ueki N, Zhang L, Hayman MJ. Ski negatively regulates erythroid differentiation through its interaction with GATA1. *Mol. Cell. Biol.* 2004;24(23):10118–10125.
36. Singbrant S, Wall M, Moody J, et al. The SKI proto-oncogene enhances the in vivo repopulation of hematopoietic stem cells and causes myeloproliferative disease. *Haematologica*. 2014;99(4):647–655.
37. Muench DE, Ferchen K, Velu CS, et al. SKI controls MDS-associated chronic TGF-β signaling, aberrant splicing, and stem cell fitness. *Blood*. 2018;132(21):e24–e34.
38. Costa V, Esposito R, Ziviello C, et al. New somatic mutations and WNK1-B4GALNT3 gene fusion in papillary thyroid carcinoma. *Oncotarget*. 2015;6(13):11242–11251.
39. Shearstone JR, Pop R, Bock C, et al. Global DNA demethylation during mouse erythropoiesis in vivo. *Science*. 2011;334(6057):799–802.
40. Thoms JAI, Birger Y, Foster S, et al. ERG promotes T-acute lymphoblastic leukemia and is transcriptionally regulated in leukemic cells by a stem cell enhancer. *Blood*. 2011;117(26):7079–7089.
41. Carmichael CL, Metcalf D, Henley KJ, et al. Hematopoietic overexpression of the transcription factor Erg induces lymphoid and erythro-megakaryocytic leukemia. *Proc. Natl. Acad. Sci. U.S.A.* 2012;109(38):15437–15442.
42. Hollanda LM, Lima CSP, Cunha AF, et al. An inherited mutation leading to production of only the short isoform of GATA-1 is associated with impaired erythropoiesis. *Nat. Genet.* 2006;38(7):807–812.

43. Ge L, Zhang R, Wan F, et al. TET2 plays an essential role in erythropoiesis by regulating lineage-specific genes via DNA oxidative demethylation in a zebrafish model. *Mol. Cell. Biol.* 2014;34(6):989–1002.
44. Zhang X, Su J, Jeong M, et al. DNMT3A and TET2 compete and cooperate to repress lineage-specific transcription factors in hematopoietic stem cells. *Nat. Genet.* 2016;48(9):1014–1023.
45. Yan H, Wang Y, Qu X, et al. Distinct roles for TET family proteins in regulating human erythropoiesis. *Blood*. 2017;129(14):2002–2012.
46. Moran-Crusio K, Reavie L, Shih A, et al. Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. *Cancer Cell*. 2011;20(1):11–24.
47. Kosmider O, Denis N, Lacout C, et al. Kit-activating mutations cooperate with Spi-1/PU.1 overexpression to promote tumorigenic progression during erythroleukemia in mice. *Cancer Cell*. 2005;8(6):467–478.
48. Rodriguez P, Bonte E, Krijgsveld J, et al. GATA-1 forms distinct activating and repressive complexes in erythroid cells. *EMBO J.* 2005;24(13):2354–2366.
49. Tallack MR, Whitington T, Yuen WS, et al. A global role for KLF1 in erythropoiesis revealed by ChIP-seq in primary erythroid cells. *Genome Res.* 2010;20(8):1052–1063.
50. Rimmelé P, Kosmider O, Mayeux P, Moreau-Gachelin F, Guillouf C. Spi-1/PU.1 participates in erythroleukemogenesis by inhibiting apoptosis in cooperation with Epo signaling and by blocking erythroid differentiation. *Blood*. 2007;109(7):3007–3014.
51. Jeong JJ, Gu X, Nie J, et al. Cytokine-Regulated Phosphorylation and Activation of TET2 by JAK2 in Hematopoiesis. *Cancer Discov*. 2019;9(6):778–795.
52. Wichmann C, Becker Y, Chen-Wichmann L, et al. Dimer-tetramer transition controls RUNX1/ETO leukemogenic activity. *Blood*. 2010;116(4):603–613.
53. Thirant C, Ignacimouttou C, Lopez CK, et al. ETO2-GLIS2 Hijacks Transcriptional Complexes to Drive Cellular Identity and Self-Renewal in Pediatric Acute Megakaryoblastic Leukemia. *Cancer Cell*. 2017;31(3):452–465.
54. Wechsler J, Greene M, McDevitt MA, et al. Acquired mutations in GATA1 in the megakaryoblastic leukemia of Down syndrome. *Nat. Genet.* 2002;32(1):148–152.
55. Thiollier C, Lopez CK, Gerby B, et al. Characterization of novel genomic alterations and therapeutic approaches using acute megakaryoblastic leukemia xenograft models. *J. Exp. Med.* 2012;209(11):2017–2031.
56. Gruber TA, Larson Gedman A, Zhang J, et al. An Inv(16)(p13.3q24.3)-encoded CBFA2T3-GLIS2 fusion protein defines an aggressive subtype of pediatric acute megakaryoblastic leukemia. *Cancer Cell*. 2012;22(5):683–697.
57. Masetti R, Pigazzi M, Togni M, et al. CBFA2T3-GLIS2 fusion transcript is a novel common feature in pediatric, cytogenetically normal AML, not restricted to FAB M7 subtype. *Blood*. 2013;121(17):3469–3472.
58. Novershtern N, Subramanian A, Lawton LN, et al. Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell*. 2011;144(2):296–309.
59. Han GC, Vinayachandran V, Bataille AR, et al. Genome-Wide Organization of GATA1 and TAL1 Determined at High Resolution. *Mol. Cell. Biol.* 2016;36(1):157–172.
60. Wadman IA, Osada H, Grütz GG, et al. The LIM-only protein Lmo2 is a bridging molecule assembling an erythroid, DNA-binding complex which includes the TAL1, E47, GATA-1 and Ldb1/NLI proteins. *EMBO J.* 1997;16(11):3145–3157.
61. Tsang AP, Visvader JE, Turner CA, et al. FOG, a multitype zinc finger protein, acts as a cofactor for transcription factor GATA-1 in erythroid and megakaryocytic differentiation. *Cell*. 1997;90(1):109–119.

62. Schuh AH, Tipping AJ, Clark AJ, et al. ETO-2 associates with SCL in erythroid cells and megakaryocytes and provides repressor functions in erythropoiesis. *Mol. Cell. Biol.* 2005;25(23):10235–10250.

FIGURE LEGENDS

Figure 1. Genomic and transcriptional landscape of human AEL

(A) Patient distribution according to age: older adult (>60 years), adult (40-59 years) and young adult (21-39 years) or according to diagnosis: AML-M6a *de novo*, AML-M6a secondary to MDS and AML-M6b *de novo*.

(B) Heatmap representation of genetic mutations in the 3 AEL genomic subgroups: TP53-mutated, epigenetic modifier and other. Each column represents a single patient sample and colors represents type of mutation: missense (blue), nonsense (red), frameshift (yellow), splice mutation (green), copy number (CN) gain (purple) and CN loss-of-heterozygotity (LOH, brown)

(C) Variants allele frequency of recurrent mutated genes in AEL patient.

Figure 2. Transcriptomic landscape of AEL

(A) Principal component analysis (PCA) of AEL patient samples based on gene expression. Each point represents one sample, colored according to their genomic sub-group.

(B) PCA of AEL patient samples based on gene expression. Samples are colored according to immunophenotyping group profile: Immature ($CD34^+$ KIT+ $CD36^-$), BFU/CFU-E-like ($CD34^+$ $KIT^{+/-}$ $CD36^+$) and erythroid ($CD34^-$ KIT^- $CD36^+$)

(C) Heatmap representation of correlation between AEL patient samples and the Merryweather-Clarke et al. erythroid signature [29]. Merryweather-Clarke et al. describes the expression profile of enriched human colony-formation unit-erythroid (CFU-E), pro-erythroblasts (PRO-E), intermediate (Int-E), and late (Late-E) erythroblast based on surface cell marker expression. Patient-derived M6 cell expression signatures clustered in an unsupervised manner into four groups according to differentiation stage: CFU-E, Pro-E, CFU-&Pro-E and Int-E&Late-E.

(D) PCA of AEL patient samples, colored according to differentiation stage groups found derived from clustering in Merryweather-Clarke et al. dataset.

(E) Heatmap of correlation between AEL samples from *Iacobucci et al*. and the Merryweather-Clarke et al. dataset, clustered in an unsupervised manner into four groups according to differentiation stage: CFU-E, Pro-E, CFU-&Pro-E and Int-E&Late-E.

(F) PCA of AEL samples from *Iacobucci et al*. [27] according to differentiation stage groups found when compared with Merryweather-Clarke et al. dataset.

(G) Heatmap of mRNA expression of 50 most important genes for classification. Important genes are derived from a random forest model for prediction of clusters derived unsupervised from correlation with Merryweather-Clarke et al. data (C-D and E-F).

Figure 3. Transcription factors activity of AEL

(A) Principal component analysis (PCA) of data from Differentiation Map (DMAP) [58] with regression to cell types in erythroid and myeloid compartment. Regression line fit to erythroid (green) and myeloid (gray) cells in the PCA space of genes significantly (FDR>05, LogFC>2) segregating each hematopoietic population: basophiles (BASO), common myeloid progenitor cells (CMP), eosinophil (EOS), erythrocytes (ERY), granulocyte-monocyte progenitor cells (GMP), granulocytes (GRAN), hematopoietic stem cells (HSC), megakaryocytes (MEGA), megakaryocytes-erythroid progenitor cells (MEP) and monocytes (MONO).

(B) PCA with regression lines from previous plot (A) with projection of AEL patient samples and MDS samples from 5 different studies.

(C) PCA with regression lines from plot (A) with projection of AEL patient samples from *Iacobucci et al*. [27] and MDS samples.

(D) PCA with regression lines from plot (A) with projection of AEL patient samples, AML and APL samples from Blueprint consortium.

(E) PCA with regression lines from plot (A) with projection of AEL patient samples and AEL patient samples from *Iacobucci et al* [27], colored with KLF1, GATA1 and CEBPA expression and activity.

(F) Histogram representation of *ETO2, ERG, GFI1, GATA3* and *SKI* gene expression in AEL patients and normal human BFU-E and CFU-E. Positive patient samples (red bars) were defined as presenting an expression above the threshold set as 4-fold the average of AEL samples. Dotted bars represent the average expression of AEL samples.

(G) Table indicate patient samples presenting with genetic alteration (green) or transcriptional alteration (blue) of GATA1 associated genes [59–62] in the 3 genomic subgroups of AEL: TP53-mutated, epigenetics and others.

(H) Kaplan-Meier survival plot of AEL patients grouped according to the presence (or absence) of genetic or transcriptional alterations affecting GATA1 associated genes. p-value using Log rank Mantel-Cox test is indicated.

(I) Kaplan-Meier survival plot of AEL patients from *Iacobucci et al.* [27] and grouped according to the presence (or not) of genetic or transcriptional alterations affecting GATA1 associated genes. p-value using Log rank Mantel-Cox test is indicated.

Figure 4. Modeling AEL *in vitro* by expressing GATA complex partners

(A) Experimental Design: mouse erythroid progenitors ($CD71^{+}Ter119^{+}KIT^{+/low}$) were sorting using BM depleted of CD3, B220, CD11b and Gr-1 positive cells, transduced by retroviral expressing *SKI*, *ERG*, *ETO2*, *GATA1s* or an empty vector (Ctrl) and maintained in StemSpan SFEM with cytokines (mSCF, mIL3, mIL6, hEPO, cholesterol and dexamethasone).

(B) Representation of cell number and percentage of GFP positive cells of mouse erythroid progenitors transduced with *SKI, ERG, ETO2, GATA1s, EPO, SPI1, B4GALNT3* or empty vector (Ctrl), maintained *in vitro* for 15 days.

(C) Flow cytometry analysis of CD71 and Ter119 expression and cytospots, analyzed with May Grunwald Giemsa staining, of mouse erythroblast maintained *in vitro* with either *SKI, ERG, ETO2, SPI1* or *B4GALNT3* overexpression.

(D) Histogram representation of *GATA1, ALAS2*, and *NFE2* gene expression detected using RT-qPCR in mouse erythroblast (Ctrl), erythroblast maintained with expression of either *SKI*, *ERG*, *ETO2*, *SPI1* or *B4GALNT3* overexpression and the mouse erythroid MEL cell line.

(E) Kaplan-Meier survival plot of mice engrafted with mouse erythroblast transduced with *SKI*, *ERG, ETO2, SPI1or B4GALNT3* and maintained over 2 months *in vitro*.

(F) Flow cytometry analysis of CD11b and Gr1 or CD71 and Ter119 gene expression, gated into viable GFP positive cells, in mice injected with mouse erythroblast expressing *SKI, ERG or ETO2* maintained in culture over 2 months before injection.

(G) Histopathological analyses of BM, spleen and liver sections from mice engrafted with mouse erythroblast expressing *SKI* or from control (Ctrl) animals, stained with hematoxylin-and-eosin (HE) or GATA1.

Statistical significance is indicated as p values (Student's t test except when otherwise specified). *: $p<0.05$, **: $p<0.01$, ***: $p<0.001$.

Figure 5. *Tp53* mutation functionally cooperates with *ERG* overexpression in erythroleukemia

(A) Experimental design: GFP positive erythroid progenitors ($CD71^+Ter119^+KIT^+$) from primary recipient mice, engrafted with HSPC cells transduced either with *ERG* or empty vector, were sorted and engrafted into sub-lethally secondary recipient mice.

(B) Kaplan-Meier survival plot of secondary recipient mice engrafted with sorted GFP^+ erythroid progenitors ($CD71^+Ter119^+KIT^+$) from primary mice engrafted with *WT* or *TP53* HSPC retrovirally transduced with *ERG* or empty vector (Ctrl).

(C) Peripheral blood counts of white blood cells (WBC), red blood cells (RBC) and platelets (PLT) in mice engrafted with *WT* or *TP53*^{R248Q} HSPC retrovirally transduced with *ERG* or empty vector (Ctrl).

(D) Flow cytometry analysis of myeloid cells (CD11b$^+$Gr-1$^+$), erythroid progenitors (CD71$^+$ Ter119$^+$), B cells (B220$^+$) and T cells (CD4$^+$CD8$^+$) gated in GFP positive or negative cells in secondary recipient mice engrafted with sorted GFP$^+$ erythroid progenitors (CD71$^+$ Ter119$^+$ KIT$^+$) from primary mice engrafted with *WT* or *TP53*^{R248Q} HSPC retrovirally transduced with *ERG* or empty vector (Ctrl).

(E) Histopathology analysis of spleen and liver of secondary recipient mice engrafted with sorted GFP$^+$ erythroid progenitors (CD71$^+$Ter119$^+$KIT$^+$) from primary mice engrafted with *WT* or *TP53*^{R248Q} HSPC retrovirally transduced with *ER*G or empty vector (Ctrl).

Figure 6. *Tet2* inactivation functionally cooperates with *Gata1s* mutation for leukemogenesis

(A) Experimental design: mouse erythroid progenitors (CD71$^+$ Ter119$^+$ KIT$^+$) were sorted using BM depleted cells for CD3, B220, CD11b and Gr-1 and maintained in StemSpan SFEM with cytokines to count viable cells. 10^5 cells were then injected into recipient mice for disease development.

(B) Left panel: numbers of erythroblasts maintained in liquid culture from WT, GATA1s, TET2$^{-/-}$ and GATA1s TET2$^{-/-}$ conditions: Right panel: cytospot of erythroblasts from GATA1s TET2$^{-/-}$ mice, maintained in liquid culture over 2 months.

(C) Flow cytometry analysis of myeloid cells (CD11b$^+$Gr-1$^+$), erythroid progenitors (CD71$^+$ Ter119$^+$), B cells (B220$^+$) and T cells (CD4$^+$CD8$^+$) gated in CD45.2 positive cells in

immunodeficient mice engrafted with CD45.2 positive erythroid progenitors from GATA1s $TET2^{-/-}$ mice maintained in liquid culture for 2 months.

(D) Histopathology analysis of spleen and liver of immunodeficient mice engrafted with erythroblast from GATA1s $TET2^{-/-}$ mice, maintained in liquid culture for 2 months before injection.

(E) Experimental design: mouse $CD45.2^{+}$ HSPC were sorted and injected into $CD45.1^{+}$ irradiated recipients followed for disease development.

(F) Kaplan-Meier survival plot and spleen weights of recipient mice engrafted with HSPC cells from $TET2^{-/-}$ or GATA1s+$TET2^{-/-}$ mice.

(G) Peripheral blood counts (WBC, RBC and PLT) of immunodeficient mice engrafted with HSPC cells from $TET2^{-/-}$ or $TET2^{-/-}$+GATA1s transgenic mice. WBC: White blood count, RBC: Red blood count, PLT: Platelet count.

(H) Flow cytometry analysis of myeloid cells ($Gr1^{+}CD11b^{+}$), erythroid progenitors ($CD71^{+}$ $Ter119^{+}$), B cells ($B220^{+}$) and T cells ($CD4^{+}CD8^{+}$) gated into $CD45.2^{+}$ donor cells from $TET2^{-/-}$ or GATA1s+$TET2^{-/-}$ mice.

Figure 7. SKI corrupts HSC and MEP to promote *in vivo* MDS-like disease with erythroid component

(A) Experimental design of hematopoietic stem cells (HSC), granulocyte-monocyte progenitors (GMP) and megakaryocyte-erythroid progenitors (MEP) sorting before retroviral transduction by *SKI* or empty vector (Ctrl) and injected into lethally irradiated recipient mice. Cells were sorting as follow: HSC was defined as $Kit^{+}Sca1^{+}CD34^{-}CD48^{-}$, MEP was defined as $Kit^{+}Sca1^{-}CD34^{-}CD16/32^{-}$ and GMP was defined as $Kit^{+}Sca1^{-}CD34^{+}CD16/32^{+}$.

(B) Left: percentage of GFP positive cells in BM. Right: Kaplan-Meier survival plot of recipient mice.

(C) Flow cytometry analysis of myeloid cells ($CD11b^+Gr\text{-}1^+$), erythroid progenitors ($CD71^+$ $Ter119^+$), B cells ($B220^+$) and T cells ($CD4^+CD8^+$) gated in viable GFP positive cells, in spleen of HSC (Control and SKI) and MEP-SKI primary engrafted mice.

(D) Percentage of myeloid cells ($CD11b^+Gr\text{-}1^+$), erythroid progenitors ($CD71^+Ter119^+$), B cells ($B220^+$) and T cells ($CD4^+CD8^+$) gated into viable GFP positive cells in HSC-Control (Ctrl, n=6), HSC-SKI (n=6) and MEP-SKI (n=6) in primary mice BM and spleen.

(E) Number of GFP positive and negative total bone marrow cells, myeloid cells ($CD11b^+Gr\text{-}1^+$) and erythroid progenitors ($CD71^+$ $Ter119^+$) in primary mice BM.

(F) Flow cytometry analysis of terminal erythroid differentiation in BM of HSC (Control and SKI) and MEP-SKI engrafted mice, using Forward scatter (FSC-A) and CD44 gene expression gated into viable Ter119 positive cells. Lower panel represents mean+/-SD (n=6) of the percentage of each population.

(G) Histopathology analysis of BM spleen and liver from HSC-Ctrl and HSC-SKI recipient mice with May Grunwald Giemsa and GATA1 staining.

SUPPLEMENTAL FIGURES

Supplemental Figure1. Hematopoietic lineage enrichment profile of AEL

(A) Heatmap representation of hematopoietic lineage enrichment score profile obtained using xCell software [28] on AEL RNAseq data.

(B) *SPI1* expression in AEL patients (normalized number of counts are represented). The red bar highlights sample that expressed more than 3-fold the highest expression observed in either normal human BFU-E or CFU-E.

Supplemental Figure 2. Transcription factor activity inference

(A) Schematic representation of the generation of a transcription factor (TF) network using the ARACNe algorithm to compute a human HSPC network and the VIPER algorithm to compute the activity of TF based on our AEL patient transcriptome signatures.

(B) Heatmap representation of the most differentially transcription factors activated in AEL patient samples.

Supplemental Figure 3. Fusion transcripts in human AEL

(A) Heatmap representation of patient carried out-of-frame (green) or in-frame (red) fusion transcript according to their molecular subgroup.

(B) Sequences of *YWHAE-EPO* and *HSD17B11-B4GALNT3* reads spanning the fusion points. The number of reads supporting the fusion are indicated.

(C) Paired histogram representation of *YWHAE* and *EPO, HSD17B11* and *B4GALNT3, LINC01* and *DNMT3B* gene expression in AEL patients, normal human BFU-E and CFU-E. Patients with red bars indicate those presenting the fusion transcripts.

Supplemental Figure 4. Modeling GATA1-associated genes alterations in mouse pro-erythroblast.

(A) Histogram representation of *SKI, ERG, ETO2, B4GALNT3, SPI1, GATA1, ALAS2*, and *NFE2* gene expression detected using RT-qPCR in WT erythroblast (Ctrl) or in erythroblast transformed with either *SKI, ERG, ETO2, SPI1* or *B4GALNT3* overexpression as compared to the erythroleukemia MEL cell line.

(B) Histopathological analysis of BM, spleen and Liver of mice engrafted with mouse erythroblasts expressing *ETO2* or *ERG*, maintained *in vitro* over 2 months.

Supplemental Figure 5. Modeling AEL oncogenic cooperation in mice

(A) Kaplan-Meier survival plot of lethally irradiated mice engrafted with HSPC cells from either WT or $TP53^{R248Q}$ mice transduced with ERG or empty vector (Ctrl).

(B) Flow cytometry analysis of myeloid cells ($Gr1^+CD11b^+$), erythroid progenitors ($CD71^+$ $Ter119^+$), B cells ($B220^+$) and T cells ($CD4^+CD8^+$) gated into GFP positive cells in irradiated mice engrafted with HSPC cells from either WT or $TP53^{R248Q}$ mice transduced with ERG or empty vector (Ctrl) at pre-leukemic or leukemic state of disease development.

(C) Flow cytometry analysis of myeloid cells ($CD11b^+Gr\text{-}1^1$) and erythroid progenitors ($CD71^+$ $Ter119^+$) in BM of WT, $TET2^{-/-}$, GATA1s or $TET2^{-/-}$+GATA1s transgenic mice at 8 months.

(D) Flow cytometry analysis of CD44 gene expression and forward scatter gated in $Ter119^+$ cells of total BM of WT, $TET2^{-/-}$, GATA1s or $TET2^{-/-}$+GATA1s transgenic mice at 8 months.

(E) Histopathology of BM, spleen and liver sections from recipient mice engrafted with HSPC from TET2$^{-/-}$+GATA1s mice.

Supplemental Figure 6. SKI overexpression in HSPC leads to a myeloid/erythroid disease in mice.

(A) Percentage of GFP positive cells in BM and Kaplan-Meier survival plot of lethally irradiated mice transplanted with 0.5x10^6 lineage negative stem cells transduced with either murine *SKI* cDNA (n=5) or empty vector (Ctrl; n=5).

(B) Peripheral blood counts of recipient mice: white blood cells (WBC), red blood cells (RBC), platelets (PLT), lymphocytes (Lym), monocytes (Mono), neutrophils and granulocytes (N/Gr).

(C) Flow cytometry analysis of myeloid cells (CD11b$^+$Gr-1$^+$), erythroid progenitors (CD71$^+$ Ter119$^+$), B cells (B220$^+$CD19$^+$) and T cells (CD4$^+$CD8$^+$) gated in GFP positive cells, in BM of recipient mice engrafted with SKI or empty vector (Ctrl) cells.

(D) Percentage of myeloid cells (CD11b$^+$Gr-1$^+$), erythroid progenitors (CD71$^+$Ter119$^+$), B cells (B220$^+$CD19$^+$) and T cells (CD4$^+$CD8$^+$) gated into viable GFP positive cells in BM or spleen of *SKI* (n=5) and control (Ctrl; n=5) recipient mice.

(E) Number of GFP positive and negative myeloid cells (CD11b$^+$Gr-1$^+$), erythroid progenitors (CD71$^+$Ter119$^+$), B cells (B220$^+$CD19$^+$) and T cells (CD4$^+$CD8$^+$) in BM or spleen of primary engrafted mice.

(F) Flow cytometry analysis of GFP positive cells in peripheral blood: platelets (PLT), red blood cells (RBC) and white blood cells (WBC) of primary engrafted mice

(G) Percentage of GFP positive cells in peripheral blood (WBC, RBC, PLT) and BM in recipient mice engrafted with HSC, MEP and GMP expressing *SKI* or empty vector (Ctrl), 3 weeks post-engraftment.

(H) Peripheral blood counts of WBC, RBC and PLT in recipient mice engrafted with HSC and MEP expressing *SKI* or empty vector (Ctrl), at day of sacrifice.

(I) Histopathology of BM and spleen of MEP expressing *SKI* engrafted mice analyzed with GATA1 or HE stainings.

Supplemental Table 1. Antibodies used in this study

Supplemental Table 2. Oligonucleotides used in this study

Supplemental Table 3. Patients' information

Supplemental Table 4. Differentially expressed genes used to cluster AEL patients based on *Merryweather-Clark et al*. normal erythroid expression data.