nature medicine

Minimal PU.1 reduction induces a preleukemic state and promotes development of acute myeloid leukemia

Britta Will[1,8], Thomas O Vogler[1,8], Swathi Narayanagari[1], Boris Bartholdy[1], Tihomira I Todorova[1], Mariana da Silva Ferreira[1], Jiahao Chen[1], Yiting Yu[2], Jillian Mayer[1], Laura Barreyro[1], Luis Carvajal[1], Daniela Ben Neriah[1], Michael Roth[1], Johanna van Oers[1], Sonja Schaetzlein[1], Christine McMahon[3], Winfried Edelmann[1,4,5], Amit Verma[2,4,6,7] & Ulrich Steidl[1,2,4,7]

Modest transcriptional changes caused by genetic or epigenetic mechanisms are frequent in human cancer. Although loss or near-complete loss of the hematopoietic transcription factor PU.1 induces acute myeloid leukemia (AML) in mice, a similar degree of PU.1 impairment is exceedingly rare in human AML; yet, moderate PU.1 inhibition is common in AML patients. We assessed functional consequences of modest reductions in PU.1 expression on leukemia development in mice harboring DNA lesions resembling those acquired during human stem cell aging. Heterozygous deletion of an enhancer of PU.1, which resulted in a 35% reduction of PU.1 expression, was sufficient to induce myeloid-biased preleukemic stem cells and their subsequent transformation to AML in a DNA mismatch repair–deficient background. AML progression was mediated by inhibition of expression of a PU.1-cooperating transcription factor, Irf8. Notably, we found marked molecular similarities between the disease in these mice and human myelodysplastic syndrome and AML. This study demonstrates that minimal reduction of a key lineage-specific transcription factor, which commonly occurs in human disease, is sufficient to initiate cancer development, and it provides mechanistic insight into the formation and progression of preleukemic stem cells in AML.

Genomic studies have shown that, in human cancer, somatic DNA alterations often occur within the noncoding part of the genome, are enriched in gene-regulatory regions and cause only moderate transcriptional changes. It is currently not well understood if and how such moderate gene expression changes contribute to malignant transformation.

The progression from a hematopoietic stem cell (HSC) to a fully differentiated cell is a multistep process[1]. A set of key transcriptional regulators establishes stable, lineage- and cell type–specific gene expression and thereby controls cell fate and differentiation outcomes[2]. One such master regulator is the Ets-family transcription factor PU.1, which is indispensable for HSC function and the differentiation of cells within the myeloid and lymphoid lineages[3–5].

AML is the most frequent acute leukemia in adults, and the median age at diagnosis is 67 years (ref. 6). It develops through a multistep transformation process that originates in HSCs. Initial genetic or epigenetic aberrations lead to the formation of preleukemic stem cells with altered function and an increased propensity for subsequent progression to AML[7]. AML consists of transplantable 'leukemia-initiating cells' and a tumor bulk of myeloid cells incapable of terminal differentiation ('leukemic blasts') that accumulate in peripheral blood and bone marrow[8]. Genes encoding transcription factors are frequently mutated, rearranged or otherwise deregulated in human AML, and mouse models of leukemia have demonstrated roles for several deregulated lineage-determining transcriptional master regulators, including PU.1, in the initiation of AML[9–12]. Reduction of PU.1 expression by 80–100% induces AML in mice, whereas PU.1 halpoinsufficiency causes subtle changes in hematopoietic differentiation but is not sufficient to induce leukemia[3,9,13,14]. The greatly diminished PU.1 levels required to induce AML in mice do not resemble the relatively moderate reduction in PU.1 levels frequently observed in human AML. Several molecular mechanisms through which PU.1 expression or its activity is impaired in human AML cells have been described, but, while common, their effects on PU.1 are relatively modest[15–20]. Homozygous mutations or deletions of the *SPI1* gene (which encodes PU.1) have not been observed in human AML; only some rare cases with heterozygous mutations or heterozygous deletions have been reported[21,22]. We hypothesized that minimal reduction in PU.1 expression can be a founding event for myeloid transformation, specifically in the context of acquired mutations accumulating during aging.

The exact mechanisms of how HSCs and preleukemic stem cells in AML acquire disease-relevant mutations is currently not well resolved, but several lines of evidence support a role of impaired

[1]Department of Cell Biology, Albert Einstein College of Medicine, Bronx, USA. [2]Department of Medicine (Oncology), Division of Hemato-Oncology, Albert Einstein College of Medicine–Montefiore Medical Center, Bronx, USA. [3]Department of Pathology, Albert Einstein College of Medicine, Bronx, USA. [4]Albert Einstein Cancer Center, Albert Einstein College of Medicine, Bronx, USA. [5]Department of Genetics, Albert Einstein College of Medicine, Bronx, USA. [6]Department of Developmental and Molecular Biology, Albert Einstein College of Medicine, Bronx, USA. [7]Institute for Stem Cell and Regenerative Medicine Research, Albert Einstein College of Medicine, Bronx, USA. [8]These authors contributed equally to this work. Correspondence should be addressed to B.W. (britta.will@einstein.yu.edu) or U.S. (ulrich.steidl@einstein.yu.edu).

Received 24 June; accepted 5 August; published online 7 September 2015; doi:10.1038/nm.3936

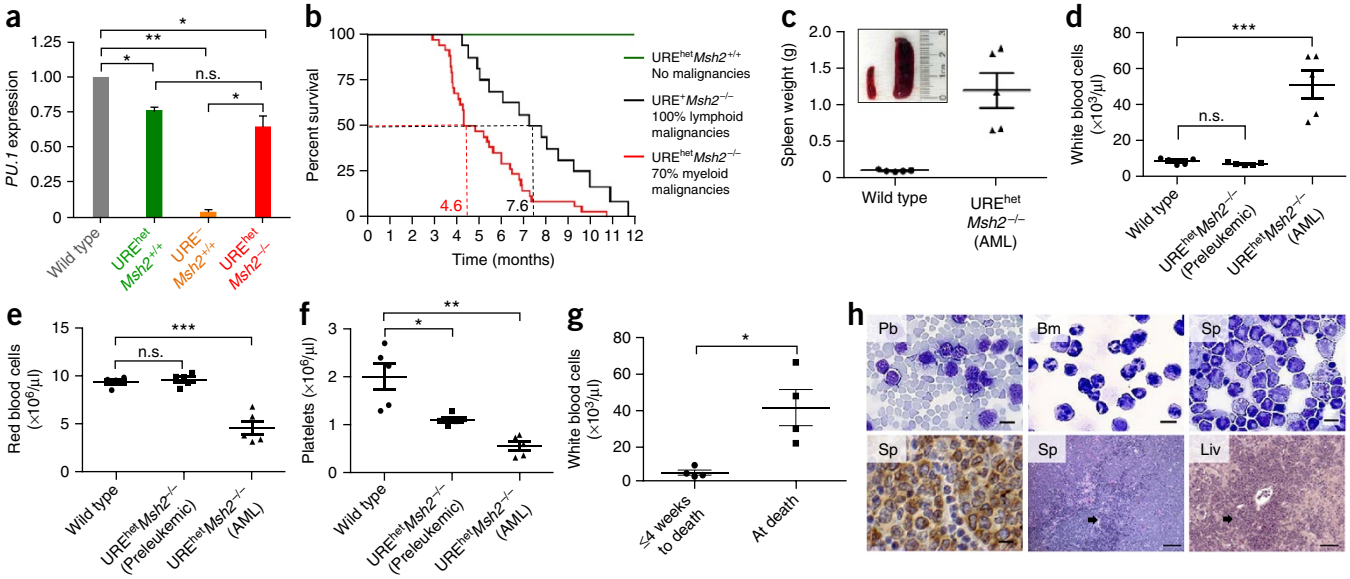
DNA mismatch repair (MMR) in leukemogenesis[23–25]. Mice lacking *Msh2*, whose product is the key component of both the MutSα and MutSβ complexes that mediate DNA MMR, display a genetic phenotype that closely mimics the spectrum of mutations found in aging human HSCs and in patients with myeloid leukemia (transition mutations and small insertions and deletions)[26,27]. We therefore generated a mouse model carrying a heterozygous deletion of an upstream enhancer of *Spi1* (also called *PU.1* here) and a homozygous deletion of *Msh2* to evaluate the role of minimal PU.1 reduction in the context of acquired mutations.

RESULTS

Minimal reduction of PU.1 expression leads to AML

To assess the effects of minimal PU.1 inhibition in the context of an elevated number of point mutations, in particular C to T or G to A transitions and small insertions or deletions resembling the mutations acquired in aging human individuals and in patients with AML, we crossed mice with a heterozygous deletion of a regulatory element 14 kb upstream of the transcriptional start site of *PU.1* (referred to as 'URE^{het}' here)[9] with *Msh2*$^{-/-}$ mice[28]. URE^{het}*Msh2*$^{-/-}$ mice were born at Mendelian frequencies. PU.1 expression in hematopoietic multipotent stem and progenitor cells sorted from URE^{het} mice exhibited a significant ($P < 0.05$) but very modest reduction of *PU.1* expression as compared to wild-type (WT) littermates (37% ± 8% in Lin$^-$Sca-1$^+$cKit$^+$ (LSK) cells, 33% ± 4% in common myeloid progenitor (CMP) cells, and 26% ± 20% in granulocytic-monocytic progenitor (GMP) cells) (**Fig. 1a** and **Supplementary Fig. 1a**,**b**). Western blotting confirmed minimal impairment of PU.1 expression at the protein level (36% reduction in myeloid progenitor cells and 21% reduction in mature neutrophils; **Supplementary Fig. 1c**). As previously reported[9], URE$^-$ mice showed a much greater reduction in *PU.1* levels (97% ± 2% reduction in LSK cells, 92% ± 3% reduction in CMP cells, and 76% ± 5% reduction in GMP cells) (**Fig. 1a** and **Supplementary Fig. 1a**,**b**).

In agreement with previous reports, we observed that *Msh2*$^{-/-}$ mice developed T cell lymphomas with 100% penetrance and a late onset[29] (**Fig. 1b**). Minimal reduction of PU.1 in combination with the *Msh2* deficiency had a two-pronged effect: (i) the median survival time was reduced from 7.6 months for URE$^+$*Msh2*$^{-/-}$ mice to 4.6 months for URE^{het}*Msh2*$^{-/-}$ mice ($P = 0.0026$; log-rank test), and (ii) the disease phenotype was altered to AML in more than two



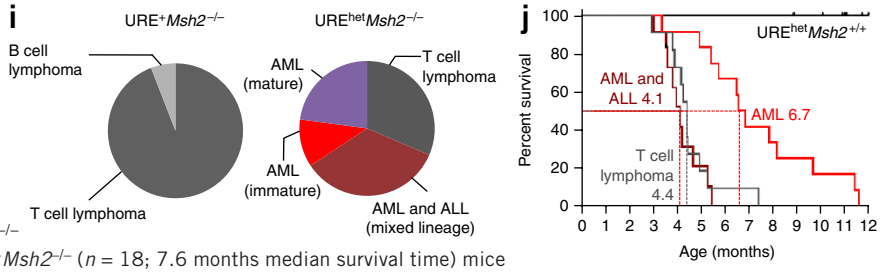


Figure 1 Minimal reduction of PU.1 expression is sufficient to trigger AML development in mice. (**a**) Quantitative RT-PCR analysis of *PU.1* expression in isolated Lin$^-$cKit$^+$Sca-1$^+$ cells (LSK) from 4- to 6-week-old URE^{het}*Msh2*$^{+/+}$, URE$^-$*Msh2*$^{+/+}$ and URE^{het}*Msh2*$^{-/-}$ mice ($n = 3$ per group). Shown are averages and s.d., expressed as fold changes as compared to wild-type controls. (**b**) Kaplan-Meier analysis of the survival of URE^{het}*Msh2*$^{+/+}$ ($n = 16$), URE^{het}*Msh2*$^{-/-}$ ($n = 34$; 4.6 months median survival time) and URE$^+$*Msh2*$^{-/-}$ ($n = 18$; 7.6 months median survival time) mice ($P = 0.0026$, log-rank (Mantel-Cox) test; compared to URE$^+$*Msh2*$^{-/-}$ mice). (**c**) Weights of spleens from moribund URE^{het}*Msh2*$^{-/-}$ mice with AML and from age-matched wild-type controls. Shown are averages and s.e.m. of $n = 5$ mice per genotype. (**d**–**g**) Peripheral blood counts. White blood cell counts (×10^3/µl) (**d**), red blood cell counts (×10^6/µl) (**e**) and platelet counts (×10^6/µl) (**f**) of preleukemic URE^{het}*Msh2*$^{-/-}$ mice, age-matched URE$^+$*Msh2*$^{+/+}$ mice and URE^{het}*Msh2*$^{-/-}$ animals with AML ($n = 5$ per group; averages and s.e.m.). (**g**) White blood cell counts (×10^3/µl) of URE^{het}*Msh2*$^{-/-}$ with AML at 2–4 weeks before death and at time of death ($n = 4$ per group; averages and s.e.m.). (**h**) May-Giemsa (Diff-quick)-, H&E- or myeloperoxidase (MPO)-stained cells in peripheral blood (Pb), bone marrow (Bm) spleen (Sp) and liver (Liv) from URE^{het}*Msh2*$^{-/-}$ mice with AML. Cytospins of Diff-quick–stained peripheral blood, bone marrow and spleen cells (upper row; scale bars, 8 µm). Tissue sections (lower row) stained with MPO (left; scale bar, 8 µm) or H&E (center and right; scale bars, 80 µm) showing infiltration with blast cells (arrows). (**i**) Pie charts depicting the different disease phenotypes in URE^{het}*Msh2*$^{-/-}$ (right, $n = 35$) and URE$^+$*Msh2*$^{-/-}$ mice (left, $n = 17$). (**j**) Kaplan-Meier survival analysis of URE^{het}*Msh2*$^{+/+}$ mice ($n = 18$) and URE^{het}*Msh2*$^{-/-}$ mice with either AML ($n = 18$), mixed lineage AML and ALL ($n = 10$) or T cell lymphoma ($n = 11$). If not specified otherwise, significance values are indicated as $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$ (Student's *t*-test); n.s., not significant.

thirds of URE^{het}*Msh2*$^{-/-}$ mice (**Fig. 1c–j**). AML was never observed in URE^{+}*Msh2*$^{-/-}$ mice. To exclude the possibility that the observed myeloid phenotype was caused by further inhibition of PU.1 expression or function, we examined expression of known PU.1 target genes in different stem and progenitor cells and found only minimal reduction in their expression levels in URE^{het}*Msh2*$^{+/+}$ mice and no further reduction in their expression levels in URE^{het}*Msh2*$^{-/-}$ mice (**Supplementary Fig. 1d**). In mice with AML, we also ensured that the PU.1-encoding gene or its regulatory regions had not acquired mutations that further reduced PU.1 activity (the one mouse out of the 20 mice examined in which this was the case was excluded from further analyses) (**Supplementary Fig. 1e**).

Macroscopically, URE^{het}*Msh2*$^{-/-}$ leukemic mice presented with severe splenomegaly (**Fig. 1c**) and hepatomegaly (data not shown). Moribund URE^{het}*Msh2*$^{-/-}$ AML mice showed elevated white blood cell counts (WBCs) and reduced red blood cell and platelet counts as compared to 4- to 12-week-old preleukemic mice of the same genotype or to age-matched WT animals (**Fig. 1d–f**). We assessed WBCs over time in URE^{het}*Msh2*$^{-/-}$ mice and found that the WBCs rapidly increased shortly before death (**Fig. 1g**), indicating the presence of an acute disease. Aberrant myeloid blasts were present in the peripheral blood, bone marrow and spleen of mice, stained positive for the myeloid marker myeloperoxidase and disrupted the tissue architecture of the spleen, liver and bone marrow (**Fig. 1h** and **Supplementary Fig. 1f–h**). Cytogenetic examination of AML cells showed evidence for clonality (**Supplementary Table 1**).

The bone marrow and spleens of URE^{het}*Msh2*$^{-/-}$ mice with AML showed an expanded cKit^{+}CD8a^{-}CD4^{-}B220^{-} cell population that also expressed high levels of CD44 (**Supplementary Fig. 1i–k**). Further analysis revealed variations in disease phenotype (**Fig. 1i**). In 17% of mice with AML (termed 'AML (immature)'), the aberrant populations appeared morphologically immature (**Supplementary Fig. 1l**), and the majority of cKit^{+} cells lacked expression of the mature myeloid marker CD11b (**Supplementary Fig. 1j**). 33% of the URE^{het}*Msh2*$^{-/-}$ mice with AML (termed 'AML (mature)') displayed aberrant cKit^{+}CD44^{+}CD8a^{-}CD4^{-}B220^{-} cells coexpressing CD11b (**Supplementary Fig. 1j**, third panel from the top), with blasts in these mice resembling more mature metamyelocyte-like neutrophils with ring-shaped nuclei (**Supplementary Fig. 1l**). 50% of the URE^{het}*Msh2*$^{-/-}$ mice with AML (termed 'AML and ALL (mixed lineage)') harbored an aberrant myeloid cKit^{+}CD44^{+}CD8a^{-}CD4^{-}B220^{-} population and a separate, expanded CD8a^{+} and/or CD4^{+} morphologically immature cell population within the hematopoietic organs (**Supplementary Fig. 1i,j**, bottom panels, and **Supplementary Fig. 1m–o**). Kaplan-Meier survival analysis of URE^{het}*Msh2*$^{-/-}$ mice further revealed that mice succumbing to mixed lineage leukemia

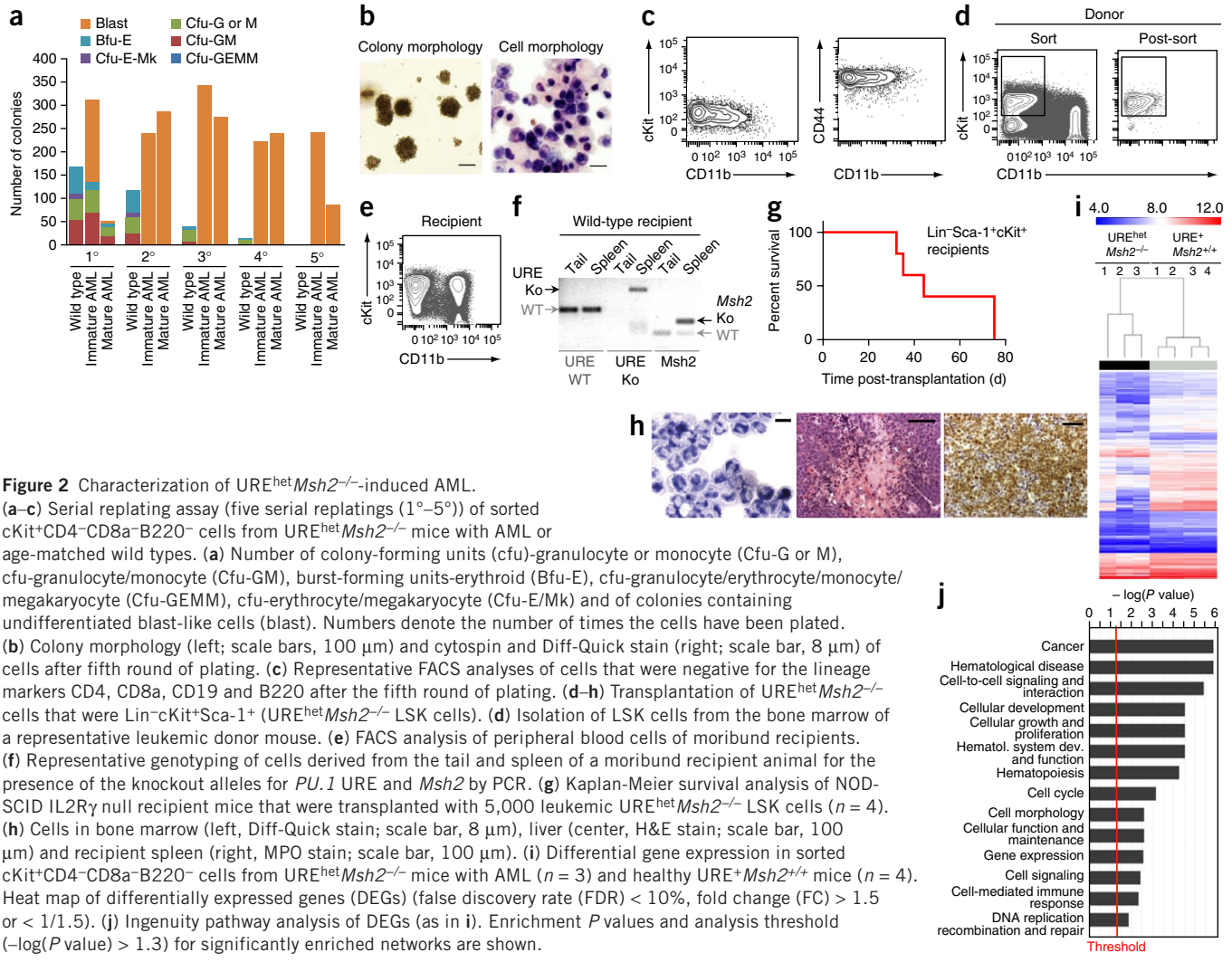


Figure 2 Characterization of URE^{het}*Msh2*$^{-/-}$-induced AML. (**a–c**) Serial replating assay (five serial replatings (1°–5°)) of sorted cKit^{+}CD4^{-}CD8a^{-}B220^{-} cells from URE^{het}*Msh2*$^{-/-}$ mice with AML or age-matched wild types. (**a**) Number of colony-forming units (cfu)-granulocyte or monocyte (Cfu-G or M), cfu-granulocyte/monocyte (Cfu-GM), burst-forming units-erythroid (Bfu-E), cfu-granulocyte/erythrocyte/monocyte/megakaryocyte (Cfu-GEMM), cfu-erythrocyte/megakaryocyte (Cfu-E/Mk) and of colonies containing undifferentiated blast-like cells (blast). Numbers denote the number of times the cells have been plated. (**b**) Colony morphology (left; scale bars, 100 μm) and cytospin and Diff-Quick stain (right; scale bar, 8 μm) of cells after fifth round of plating. (**c**) Representative FACS analyses of cells that were negative for the lineage markers CD4, CD8a, CD19 and B220 after the fifth round of plating. (**d–h**) Transplantation of URE^{het}*Msh2*$^{-/-}$ cells that were Lin^{-}cKit^{+}Sca-1^{+} (URE^{het}*Msh2*$^{-/-}$ LSK cells). (**d**) Isolation of LSK cells from the bone marrow of a representative leukemic donor mouse. (**e**) FACS analysis of peripheral blood cells of moribund recipients. (**f**) Representative genotyping of cells derived from the tail and spleen of a moribund recipient animal for the presence of the knockout alleles for *PU.1* URE and *Msh2* by PCR. (**g**) Kaplan-Meier survival analysis of NOD-SCID IL2Rγ null recipient mice that were transplanted with 5,000 leukemic URE^{het}*Msh2*$^{-/-}$ LSK cells ($n = 4$). (**h**) Cells in bone marrow (left, Diff-Quick stain; scale bar, 8 μm), liver (center, H&E stain; scale bar, 100 μm) and recipient spleen (right, MPO stain; scale bar, 100 μm). (**i**) Differential gene expression in sorted cKit^{+}CD4^{-}CD8a^{-}B220^{-} cells from URE^{het}*Msh2*$^{-/-}$ mice with AML ($n = 3$) and healthy URE^{+}*Msh2*$^{+/+}$ mice ($n = 4$). Heat map of differentially expressed genes (DEGs) (false discovery rate (FDR) < 10%, fold change (FC) > 1.5 or < 1/1.5). (**j**) Ingenuity pathway analysis of DEGs (as in **i**). Enrichment *P* values and analysis threshold (–log(*P* value) > 1.3) for significantly enriched networks are shown.

or T cell lymphoma had a significantly faster rate of disease progression as compared to mice that developed AML alone ($P < 0.001$, log-rank test) (**Fig. 1j**). These data show that minimal reduction of PU.1 expression leads to development of AML in 70% of URE^{het}*Msh2*$^{-/-}$ animals, which show phenotype variations resembling those in human disease.

Leukemia-initiating cells in URE^{het}*Msh2*$^{-/-}$-derived AML

We characterized self-renewal and differentiation capacity of the aberrant CD8a$^-$CD4$^-$B220$^-$ (Lymph$^-$) cKit$^+$ population from URE^{het}*Msh2*$^{-/-}$ mice with AML. Compared to WT cKit$^+$Lymph$^-$ cells, cKit$^+$Lymph$^-$ cells from URE^{het}*Msh2*$^{-/-}$ mice with immature AML formed 94% more total colonies, whereas cells from URE^{het}*Msh2*$^{-/-}$ mice with mature AML gave rise to 70% fewer colonies in the initial plating of the colony-forming assays (**Fig. 2a**). However, URE^{het}*Msh2*$^{-/-}$ cells of either AML subtype showed greater clonogenic capacity in the second to fifth plating and continued to give rise to aberrant blast-like colonies after the WT colony-initiating cells were exhausted. The number of myeloid colonies with normal morphology derived from immature or mature AML cells was found to be reduced in the first plating (**Fig. 2a**). URE^{het}*Msh2*$^{-/-}$-derived aberrant colonies were comprised of differentiation-impaired myeloid cells with blast morphology and became the sole colony type after the initial cell plating (**Fig. 2a–c**). Cells isolated from the blast colonies maintained the same immunophenotype as the primary AML bulk cell population (**Fig. 2d** and **Supplementary Fig. 1i–k**). Next, we transplanted purified cKit$^+$Lymph$^-$ cells and stem cell–enriched LSK populations into immunocompromised NOD-SCID IL2Rγ-null (NOG) mice. Recipient animals displayed a massive expansion of a donor-derived bulk tumor population with the same phenotype as that observed in the primary tumors (**Fig. 2e–h** and **Supplementary Fig. 2a–d**). All recipients died of AML within 3–11 weeks, demonstrating that the disease is transplantable from the stem cell–containing cell compartment (**Fig. 2g** and **Supplementary Fig. 2d**).

Gene expression analysis of leukemia-initiating cKit$^+$Lymph$^-$ cells from URE^{het}*Msh2*$^{-/-}$ mice revealed alterations that substantially affected several cellular networks. Comparison with cKit$^+$Lymph$^-$ cells from age-matched healthy URE$^+$*Msh2*$^{+/+}$ mice identified 587 annotated, differentially expressed genes (**Supplementary Table 2**), which separated URE^{het}*Msh2*$^{-/-}$ AML cells from WT cells in a hierarchical cluster analysis (**Fig. 2i**). Pathway analysis showed that differentially expressed genes were enriched for molecules involved in several key cellular processes (**Fig. 2j** and **Supplementary Fig. 2e**).

Role for *Irf8* downregulation in AML induction

Among the differentially expressed genes found in cKit$^+$Lymph$^-$ cells from URE^{het}*Msh2*$^{-/-}$ AML, we identified *Irf8* (also known as *ICSBP*) to be significantly reduced compared to cells from age-matched URE$^+$*Msh2*$^{+/+}$ (WT) controls ($P = 0.025$, Student's *t*-test) (**Fig. 3a**). As Irf8 cooperatively regulates the expression of target genes together with PU.1 (refs. 30–32), we examined whether a subset of the deregulated genes in URE^{het}*Msh2*$^{-/-}$ AML are common targets of PU.1 and Irf8. Using published chromatin immunoprecipitation and sequencing (ChIP-seq) data sets, we identified genes with PU.1 or with combined PU.1 and Irf8 occupancy within their promoter regions that were also differentially expressed in URE^{het}*Msh2*$^{-/-}$ AML (**Supplementary Tables 3** and **4**). Both subsets of genes showed a significant (–log (*P* value) > 1.3) enrichment of genes involved in AML-relevant pathways (**Fig. 3b** and **Supplementary Fig. 3a**). Motif enrichment analysis designated a PU.1-IRF composite DNA-binding motif containing a PU-box and an adjacent IRF consensus site as the top enriched motif (**Fig. 3c**). Intersection of the genes occupied by PU.1 and Irf8 in their promoter regions with the differentially expressed genes in URE^{het}*Msh2*$^{-/-}$ AML cells (as compared to those in URE$^{+/+}$*Msh2*$^{+/+}$ cells) showed reduced expression of several co-occupied genes (**Supplementary Fig. 3b**). We next tested whether these could be rescued by Irf8 restoration using a retroviral vector system containing a GFP reporter. Compared to the empty vector control, transduction with a retrovirus encoding Irf8 led to restoration of Irf8 protein levels (**Fig. 3d**) and, concomitantly, a significant increase ($P < 0.05$, Student's *t*-test) in the expression of several putative PU.1-Irf8 target genes, including *Fam132a*, *Rnf13*, *Osbpl3* and previously established PU.1-Irf8 co-regulated targets *Etv3*, *Nf1* and *Emr1* in URE^{het}*Msh2*$^{-/-}$ AML cells (**Fig. 3e** and **Supplementary Fig. 3c,d**). Irf8 expression restoration also induced myeloid differentiation of URE^{het}*Msh2*$^{-/-}$ AML cells as evidenced by morphological analysis and immunophenotyping (**Fig. 3f,g**). Furthermore, we observed induction of apoptosis in leukemic cells upon rescue of Irf8 expression ($P < 0.05$, Student's *t*-test) (**Fig. 3h**). Restoration of Irf8 expression also led to significantly impaired self-renewal of URE^{het}*Msh2*$^{-/-}$ AML cells ($P < 0.05$, Student's *t*-test) and a progressive loss of colony-initiating cKit-expressing cells in serial replating assays (**Fig. 3i** and **Supplementary Fig. 3e**). Mice injected with Irf8-rescued (GFP$^+$) URE^{het}*Msh2*$^{-/-}$ AML cells survived significantly longer than mice transplanted with GFP$^+$ URE^{het}*Msh2*$^{-/-}$ AML cells infected with an empty vector control ($P < 0.001$, log-rank test; **Fig. 3j**). Competitive transplantation of a mix of Irf8-restored (GFP$^+$) and nonrestored (GFP$^-$) URE^{het}*Msh2*$^{-/-}$ AML cells demonstrated that low Irf8 levels were critical to conferring growth advantage *in vivo*, as URE^{het}*Msh2*$^{-/-}$ AML cells with restored Irf8 levels were outcompeted by GFP$^-$ AML cells in the recipients' bone marrow (BM) and spleen (SP); this observation was not made with the empty vector control (**Fig. 3j**). The rescue of PU.1 expression led to similar differentiation and apoptosis-inducing effects as those observed with the rescue of Irf8 in the mouse AML cells (**Supplementary Fig. 3f–h**). Together, these data show that reduced expression of Irf8 contributes to the myeloid differentiation block, impaired apoptosis, *in vitro* self-renewal and the competitive growth advantage of leukemic cells in URE^{het}*Msh2*$^{-/-}$ AML.

Minimal *PU.1* expression reduction induces a preleukemic state

Irf8 was not decreased at the preleukemic stage (**Fig. 4a**). In contrast, *PU.1* expression, although minimally reduced, did not change upon progression to AML (**Fig. 4a**). FACS analysis of phenotypically defined HSC and lymphoid-myeloid multipotent progenitor cells (LMPP) (**Supplementary Fig. 4a**) revealed a significant two-fold increase of myeloid-biased HSCs (Lin$^-$IL7Rα$^-$cKit$^+$Sca-1$^+$CD150$^+$Flt3$^-$CD41$^+$ (ref. 33) and CD150^{high}Lin$^-$IL7Rα$^-$Flt3$^-$cKit$^+$Sca-1$^+$CD48$^-$ (ref. 34)) in URE^{het} mice (versus age-matched URE$^+$ mice; $P < 0.01$) and a significant decrease of phenotypic lineage-unbiased HSCs (Lin$^-$IL7Rα$^-$cKit$^+$Sca-1$^+$CD150$^+$Flt3$^-$CD41$^-$ and CD150^{medium}Lin$^-$IL7Rα$^-$Flt3$^-$cKit$^+$Sca-1$^+$CD48$^-$; $P < 0.05$; **Fig. 4 b,c** and **Supplementary Fig. 4b**). We next assessed multilineage reconstitution by highly purified, phenotypic lineage-unbiased HSCs or myeloid-biased HSCs from URE^{het}*Msh2*$^{+/+}$ animals; reconstitution with either population resulted in a marked increase in donor-derived myeloid cells in the peripheral blood of recipient animals versus animals transplanted with WT HSCs (**Fig. 4d** and **Supplementary Fig. 4c**). *In vivo* BrdU labeling revealed an increase in cycling cells within the phenotypic myeloid-biased HSC population in URE^{het} mice (**Supplementary Fig. 4d**). These results demonstrate that minimal PU.1 expression reduction

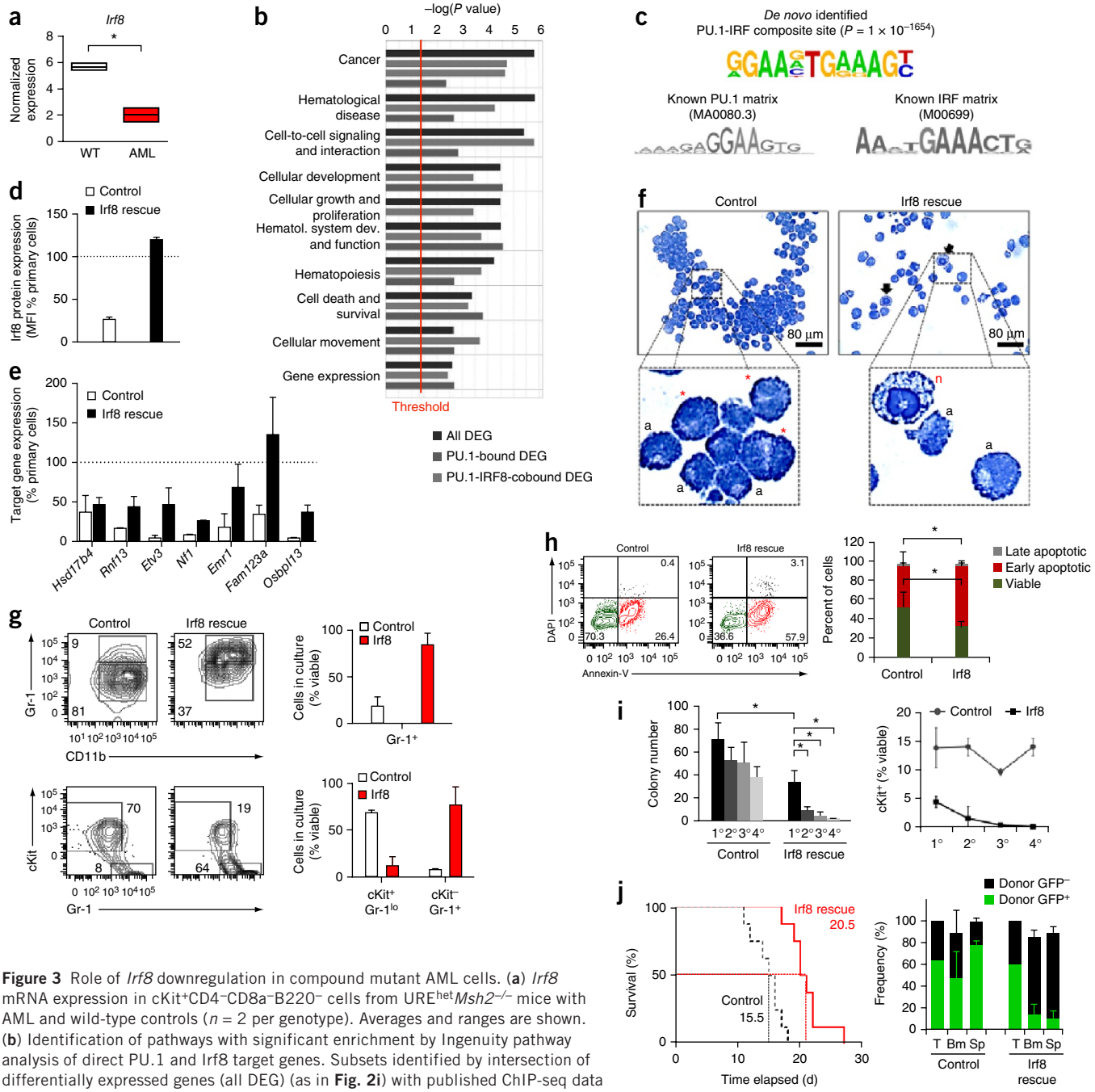


Figure 3 Role of *Irf8* downregulation in compound mutant AML cells. (**a**) *Irf8* mRNA expression in cKit+CD4−CD8a−B220− cells from URE^{het}*Msh2*$^{-/-}$ mice with AML and wild-type controls ($n = 2$ per genotype). Averages and ranges are shown. (**b**) Identification of pathways with significant enrichment by Ingenuity pathway analysis of direct PU.1 and Irf8 target genes. Subsets identified by intersection of differentially expressed genes (all DEG) (as in **Fig. 2i**) with published ChIP-seq data revealed subsets of DEG bound either by PU.1 alone (PU.1-bound DEG) or by PU.1 and IRF8 (PU.1-IRF8–cobound DEG) in regions spanning the transcriptional start site (TSS) and up to 20 kb upstream. (**c**) Top composite motif (ETS-IRF) using the HOMER software for motif discovery. Analyis showed that PU.1 and IRF8 ChIP-seq peaks co-occurred in regions 0–5 kb upstream of the TSS with high similarity to known consensus motifs. (**d–j**) Rescue of Irf8 expression in URE^{het}*Msh2*$^{-/-}$ AML cells by transduction with a retroviral construct expressing either Irf8-eGFP or eGFP as a control. (**d**) Irf8 protein levels (background-subtracted mean fluorescence intensities (MFI)) in rescued and control AML cells as compared to primary WT cKit+FcγR16/32+Lymph− cells (expressed as percentage). (**e**) mRNA expression changes of PU.1 and Irf8 co-regulated target genes. Shown are averages and s.d. of technical triplicates expressed as 'percentage of primary myeloid progenitor cell expression'. (**f**) Diff-Quick stain of AML cells transduced with an empty vector control (left) or an Irf8-expressing (right) vector. Dividing cells (red asterisks), morphologically more mature (arrow) neutrophil-like cells (n) and myelocyte and metamyelocyte (a) stages are indicated. Scale bar, 80 µm. (**g**) Representative FACS plots (left) and quantification of cell surface marker positive, eGFP+ cells in three independent experiments (right; $n = 3$; averages and s.d. are shown). (**h**) Representative FACS density plots of annexin V and DAPI stains. Quantification of GFP+ viable (DAPI−annexin V−), early apoptotic (DAPI−annexin V+) and late apoptotic (DAPI+annexin V+) cells 48 h after transduction (three independent experiments). (**i**) Left, serial replating assay (four serial replatings (1°–4°)) of URE^{het}*Msh2*$^{-/-}$ AML cells. Right, FACS analysis of cKit+ URE^{het}*Msh2*$^{-/-}$ AML cells in culture during four serial replatings (1°–4°) ($n = 4$). (**j**) Left, Kaplan-Meier survival analysis of NOG mice that received 40,000 eGFP+ control ($n = 8$; 15.5 d median survival time) or Irf8-restored ($n = 8$; 20.5 d median survival time) URE^{het}*Msh2*$^{-/-}$ AML cells ($P = 0.0002$, log-rank test). Right, FACS analysis of competitive transplantation of unfractionated AML cells after transduction (T) and quantification of eGFP+ and eGFP− donor (CD45.2+) cells in recipients' spleen (Sp) and bone marrow (Bm) ($n = 4$ per group). $^*P < 0.05$ (Student's *t*-test). Error bars show s.d.

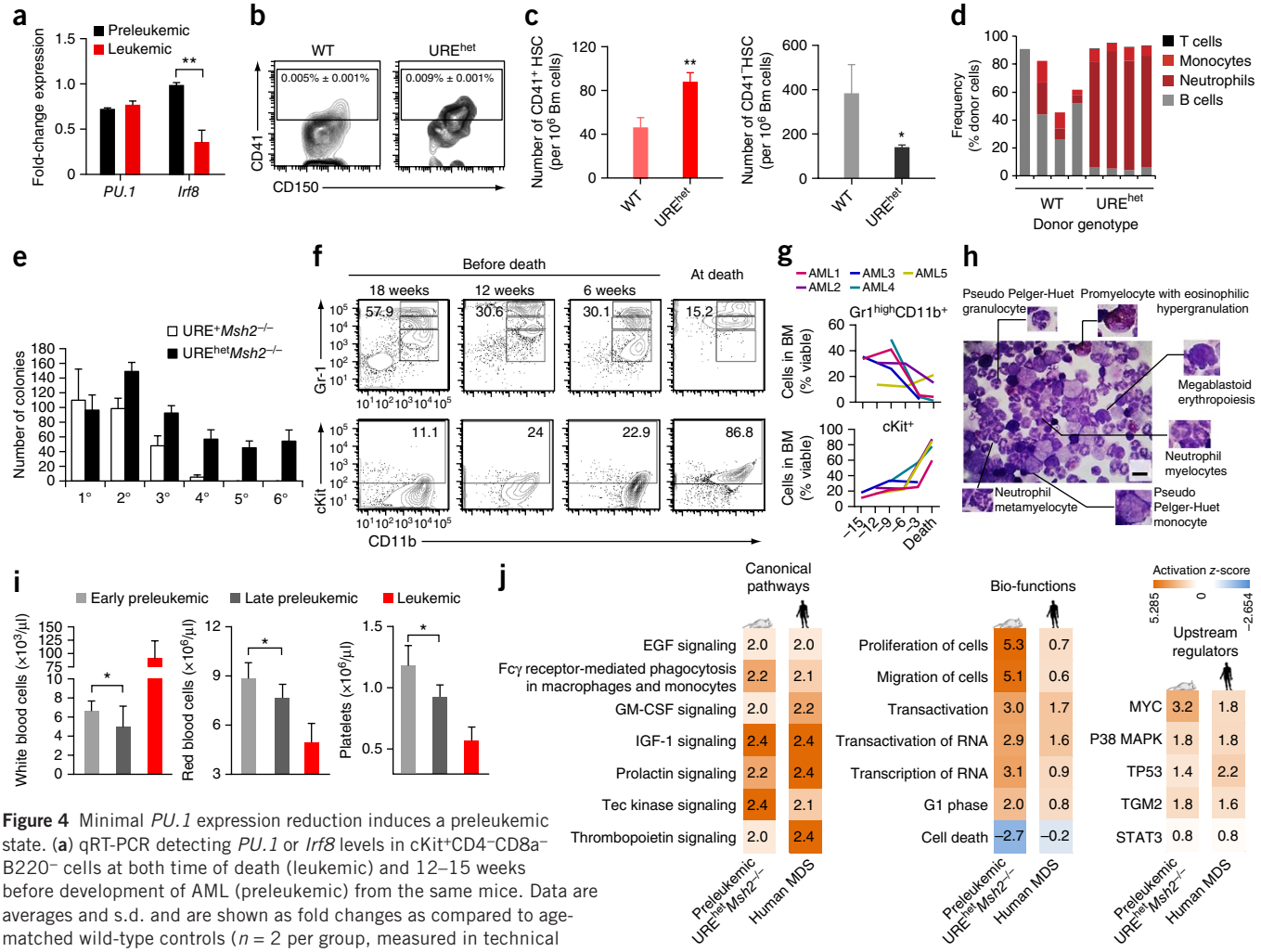


Figure 4 Minimal *PU.1* expression reduction induces a preleukemic state. (**a**) qRT-PCR detecting *PU.1* or *Irf8* levels in cKit$^+$CD4$^-$CD8a$^-$B220$^-$ cells at both time of death (leukemic) and 12–15 weeks before development of AML (preleukemic) from the same mice. Data are averages and s.d. and are shown as fold changes as compared to age-matched wild-type controls ($n = 2$ per group, measured in technical triplicates). (**b**) Representative FACS plot of CD41 expression on CD150$^+$Flt3$^-$IL7Rα$^-$KSL HSC in 6–8-week-old URE^{het} (right) or WT (left) animals. (**c**) Quantification of phenotypical CD41$^+$ (left) and CD41$^-$ (right) HSC ($n = 3$ mice per genotype). (**d**) Lineage output of 100 transplanted CD41$^-$ HSC 90 d after competitive transplantation. Donor-derived (CD45.2$^+$) T cells (CD4$^+$Gr-1$^-$B220$^-$CD11b$^-$), monocytes (CD4$^-$B220$^-$Gr-1^{medium}CD11b$^+$), neutrophils (CD4$^-$B220$^-$Gr-1$^+$CD11b$^+$) and B cells (CD4$^-$Gr-1$^-$CD11b$^-$B220$^+$) were assessed. Shown are values from individual recipients ($n = 4$ per group). (**e**) Serial replating assay (six serial replatings (1°–6°)) of bone marrow cells derived from 4–6-week-old URE^{het}*Msh2*$^{-/-}$ and URE$^+$*Msh2*$^{-/-}$ mice ($n = 4$). (**f**,**g**) Longitudinal analysis of bone marrow from URE^{het}*Msh2*$^{-/-}$ mice that ultimately develop AML. (**f**) FACS plots of viable CD4$^-$CD8a$^-$B220$^-$ (Lymph$^-$) cells at 18, 12 and 6 weeks before death and at time of death. (**g**) Quantification of Gr-1$^+$CD11b$^+$ neutrophils (top) and cKit$^+$ cells (bottom) from five individual URE^{het}*Msh2*$^{-/-}$ mice at various time points (15, 12, 9, 6 and 3 weeks before death and at death). (**h**) Bone marrow cells from preleukemic URE^{het}*Msh2*$^{-/-}$ mouse showing cell morphological myelodysplastic abnormalities. Scale bar, 8 μm. (**i**) Differential peripheral blood cell counts at the early preleukemic (12–30 weeks before overt AML), late preleukemic (4–6 weeks before overt AML) and leukemic stages. Data are averages and s.d. of consecutive measurements from three animals per time point. (**j**) Comparative pathway analysis of DEGs in cKit$^+$Lymph$^-$ cells from preleukemic URE^{het}*Msh2*$^{-/-}$ mice (versus age-matched wild-type controls; FDR < 10%, FC > 1.5 or FC < 1/1.5, $n = 4$) and human patients with MDS (GSE19429, $n = 174$) versus healthy controls ($n = 17$; FDR < 5%, FC > 1.5 or FC < 1/1.5) (colored tiles indicate activation *z*-score). $^*P < 0.05$, $^{**}P < 0.01$ (Student's *t*-test). Error bars show s.d.

is sufficient to induce functional alterations in the immature hematopoietic cell compartment.

HSC-enriched cell populations derived from preleukemic URE^{het}*Msh2*$^{-/-}$ mice (using URE$^+$*Msh2*$^{-/-}$ mice as controls) had an extended self-renewal capacity with colony initiation beyond the fourth plating (**Fig. 4e**). Using serial bone marrow aspirations and peripheral blood sampling of preleukemic URE^{het}*Msh2*$^{-/-}$ mice, we observed a progressive increase in the number of immature myeloid cells (cKit$^+$CD11b$^{-/lo}$Lymph$^-$) along with a gradual decrease in the number of mature myeloid cells (Gr-1^{high}CD11b$^+$Lymph$^-$) (**Fig. 4f**,**g** and **Supplementary Fig. 4e**,**f**) as well as expansion of the phenotypic HSC-MPP compartments (**Supplementary Fig. 4g**). Furthermore, in preleukemic mice, we detected cells with dysplastic morphological features, including pseudo-Pelger-Huet cells, megablastoid erythropoiesis, eosinophilic hypergranulation and a mild increase in numbers of promyelocytes and neutrophilic myelocytes (**Fig. 4h**). Longitudinal complete blood cell count (CBC) analysis showed normal numbers of total white blood cells and red blood cells and slightly reduced platelet numbers (**Fig. 1f**) in the early preleukemic phase (12–30 weeks before overt AML), a slight but significant reduction of cells in all three lineages during the late preleukemic phase (4–6 weeks before overt AML) and an up to 25-fold increase in white blood cells accompanied by a progressive reduction of red blood cells and platelets in the leukemic stage (**Fig. 4i**). Comparative pathway analysis of differential

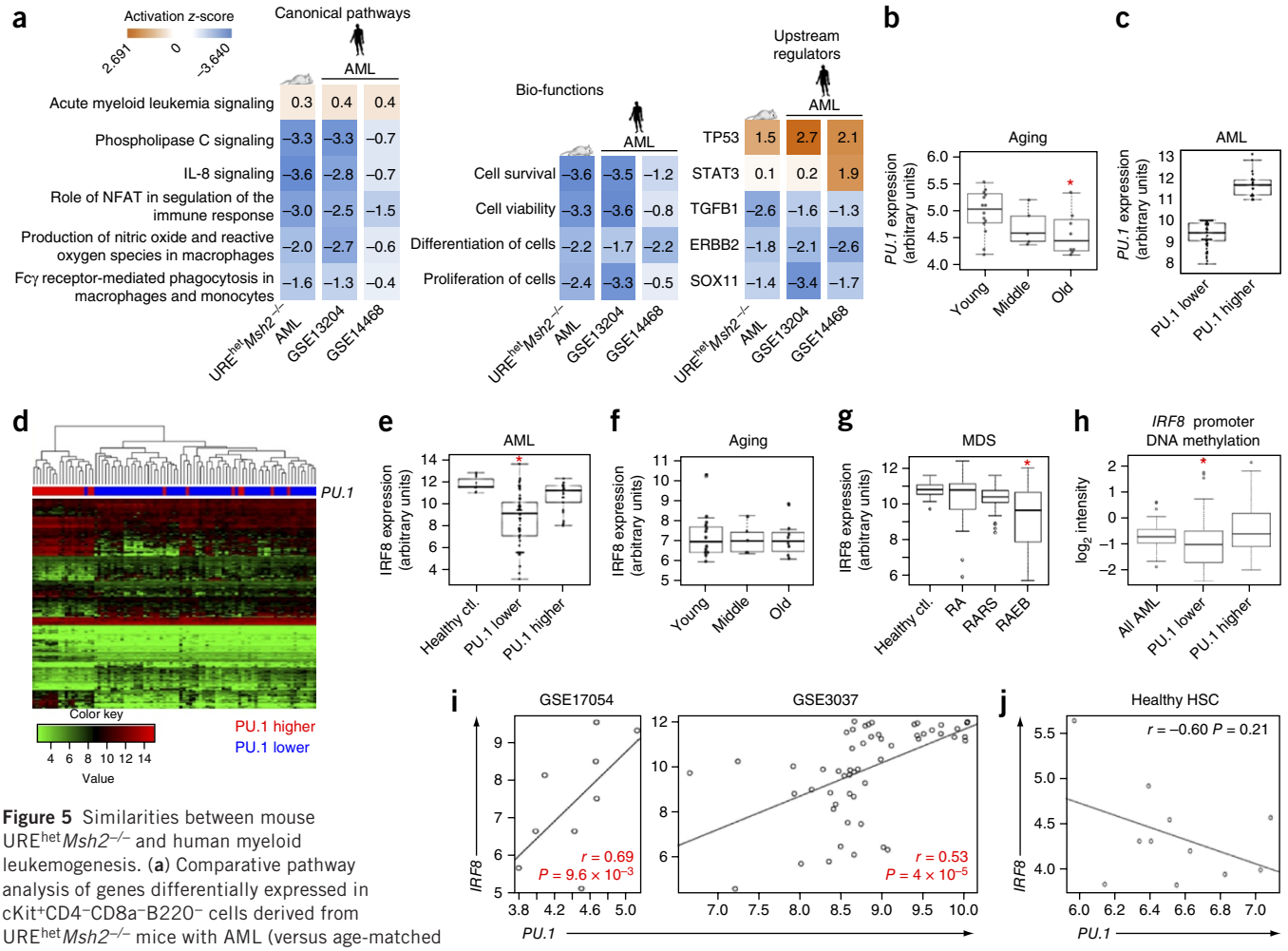


Figure 5 Similarities between mouse URE^{het}*Msh2*$^{-/-}$ and human myeloid leukemogenesis. (**a**) Comparative pathway analysis of genes differentially expressed in cKit^{+}CD4^{-}CD8a^{-}B220^{-} cells derived from URE^{het}*Msh2*$^{-/-}$ mice with AML (versus age-matched wild-type controls (FDR < 10%, FC > 1.5 or FC < 1/1.5; $n = 4$)) or human patients with AML (GSE13204 ($n = 351$); FDR < 1%, log(FC) >1.1 or log(FC)< −1.1; GSE14468 ($n = 212$) versus healthy controls ($n = 11$); FDR < 0.2%, log(FC) >1.5 or log(FC) < −1.5) (colored tiles indicate activation *z*-score). (**b**) *PU.1* expression in human hematopoietic stem cells during aging (GSE32719; $n_{(Young)} = 14$, $n_{(Middle)} = 5$, $n_{(Old)} = 8$). (**c**) AML patients (GSE14468) were dichotomized into two groups on the basis of their *PU.1* expression: 'PU. 1 higher' (>80th percentile, $n = 25$) and 'PU.1 lower' (<50th percentile, $n = 62$). (**d**) Hierarchical cluster analysis performed on the basis of 325 probe IDs of DEGs between 'PU.1 higher' and 'PU.1 lower' groups of patients that also showed differential expression in URE^{het}*Msh2*$^{-/-}$ mice with AML ($q < 0.01$, $\log_2$(FC) is ±1.2). (**e**–**g**) *IRF8* gene expression analysis in (**e**) AML patient– and age-matched healthy individual–derived (Healthy ctl) CD34^{+} cells (GSE14468) (*$P < 0.05$, Wilcox sum rank test), (**f**) young and aged HSCs (GSE32719) and (**g**) patients with MDS (grouped by WHO classification into patients with refractory anemia (RA; $n = 51$), refractory anemia with ringed sideroblasts (RARS; $n = 44$), refractory anemia with excess of blasts (RAEB; $n = 79$)) and age-matched healthy controls ($n = 17$). (**h**) DNA cytosine methylation analysis (GSE18700, $n = 125$) (*$P < 0.001$, Student's *t*-test). (**i**,**j**) Correlation plots of *PU.1* and *IRF8* expression in leukemia stem cells from two independent studies (**i**) and in healthy individuals (GSE35008) (**j**). Axes show arbitrary units of $\log_2$-transformed array values. *$P < 0.05$, **$P < 0.01$; ***$P < 0.001$ (Student's *t*-test, if not specified otherwise).

gene expression profiles of cKit^{+}Lymph^{-} cells derived from preleukemic mice (as compared to those of age-matched WT mice) revealed a set of differentially expressed genes (**Supplementary Table 5**) affecting pathways that were also found significantly deregulated in human myelodysplastic syndrome (MDS) ($P > 1.3 \times 10^{-4}$; **Fig. 4j** and **Supplementary Table 6**). These observations demonstrate that minimally impaired PU.1 induces myeloid-biased mouse HSPCs and, in the context of *Msh2* deficiency, increases self-renewal and leads to the development of a preleukemic state that resembles the cellular and molecular features of human myelodysplasia, a preleukemic disease.

Similarities between mouse URE^{het}*Msh2*$^{-/-}$-derived AML and human AML

Comparative pathway analysis of differentially expressed genes in leukemic URE^{het}*Msh2*$^{-/-}$ cKit^{+}Lymph^{-} cells (relative to age-matched WT controls) identified significantly dysregulated pathways ($P > 1.3 \times 10^{-4}$) that were shared with human AML (derived by comparing persons with AML versus healthy individuals) (**Fig. 5a** and **Supplementary Table 7**). We next evaluated whether molecular alterations seen in mouse URE^{het}*Msh2*$^{-/-}$-induced leukemogenesis were observed during the process of myeloid transformation in humans. As human AML development is highly age dependent[35], we first tested whether changes in *PU.1* expression occur during human aging. Our analyses revealed slightly but significantly decreased expression levels of PU.1 in highly purified HSC from healthy >65-year-old donors when compared to levels in HSCs from young, 20- to 35-year-old donors ($P < 0.05$; **Fig. 5b**). We next assessed whether the reduction in *PU.1* expression correlates with the expression of its transcriptional target genes in human AML. We divided individuals from two published gene expression studies into two subgroups: 'PU.1 lower'

and 'PU.1 higher' individuals (**Fig. 5c** and **Supplementary Fig. 5a**). We determined the differential gene expression profiles between the two groups and identified 325 probe IDs, which corresponded to 219 unique annotated genes, with expression changes that were also found to be differentially expressed in mouse AML (versus age-matched WT controls; **Fig. 5d** and **Supplementary Fig. 5b**). Among the differentially expressed genes we found a significant enrichment of genes harboring the PU.1-binding motif in their promoters ($P < 0.05$, Fisher's exact *t*-test; **Supplementary Fig. 5c**), suggesting that the reduction in *PU.1* expression has an impact on the expression of its target genes. To test the hypothesis that the molecular alterations seen in our mouse AML model show more similarities to human AML with lower PU.1 expression than to human AML with higher PU.1 expression, we used two different approaches. We performed comparative gene set enrichment analysis (GSEA) and found that dysregulated gene expression programs in the 'PU.1 lower' group of individuals with AML enriched for 1,223 gene sets that were also concordantly altered in mice with AML, whereas GSEA between the 'PU.1 higher' group of people with AML and mice with AML revealed only 464 shared gene sets (**Supplementary Fig. 5d**). Using a second analysis strategy we directly tested whether genes that were differentially regulated in mouse AML are dysregulated in human AML that is associated with lower or higher PU.1 expression. The set of genes aberrantly expressed in mouse AML showed enrichment in the group of aberrantly expressed genes found in the 'PU.1 lower' group of people with AML, whereas they were not enriched in the 'PU.1 higher' group (**Supplementary Fig. 5e**). Interestingly, among the genes correlated with 'PU.1 higher' or 'PU.1 lower' expression status in AML patients, we also found *IRF8* (**Fig. 5e** and **Supplementary Fig. 5f**). We found no significant *IRF8* expression change during aging (**Fig. 5f**) or in patients with lower risk MDS (**Fig. 5g**); yet, we detected lower *IRF8* expression in advanced stage MDS patients (with refractory anemia and an excess of blasts (RAEB)), whose disease has a higher propensity for progressing to AML (**Fig. 5g**). Consistent with this observation, MDS patients with lower *IRF8* expression had a significantly worse overall probability of survival ($P = 0.0144$) than patients with higher IRF8 expression (**Supplementary Fig. 5g**). AML patients with lower *PU.1* expression showed a significant ($P < 0.001$) hypermethylation of the *IRF8* promoter (**Fig. 5h**). Lastly, we compared expression changes found in URE^{het}*Msh2*$^{-/-}$ leukemia-initiating cells with differential gene expression in human leukemic stem cells (LSCs) versus healthy HSCs. Notably, although we detected a significantly positive correlation between *PU.1* and *IRF8* expression in LSCs from two independent studies (**Fig. 5i**), *PU.1* and *IRF8* expression patterns in healthy HSCs did not show a similar correlation (**Fig. 5j**). Together, these data show a marked molecular resemblance between mouse URE^{het}*Msh2*$^{-/-}$-derived AML and human acute myeloid leukemia pathogenesis. When we restored IRF8 expression in human AML cell lines showing IRF8 'lower' (such as Kasumi-3 or ML-2 cells) or IRF8 'higher' (such as Nomo-1 or Molm14 cells) expression patterns (**Supplementary Fig. 5h**), we found that increased IRF8 levels led to induction of differentiation in all four of the IRF8 'lower' cell lines and apoptosis in two of these AML cell lines, whereas ectopic expression of IRF8 in the IRF8 'higher' AML cell lines neither induced differentiation nor apoptosis (**Supplementary Fig. 5i**).

DISCUSSION

Modest impairment of PU.1 activity or expression is common in human AML pathogenesis, yet mouse models have failed to demonstrate a functional relevance for minimal dosage alterations of this key hematopoietic regulator in AML. Here we provide genetic evidence that minimal inhibition of PU.1 can be a founding event in leukemogenesis.

Our data reveal that a 35% reduction in PU.1 expression introduces a myeloid bias in multipotent mouse HSPCs. Our phenotypic marker analysis suggested there was an expanded myeloid-biased HSC population in mice with reduced PU.1. However, cell cycle studies showed that these cells were less quiescent than their wild-type counterparts, in line with a previous report[36]. Adoptive cell transfer experiments showed that, although myeloid biased, these cells were still capable of multilineage reconstitution. Thus, even though cell surface markers define this population as myeloid-biased HSCs, the phenotypic characteristics are more consistent with an early multipotent progenitor. Notably, these myeloid lineage–biased multipotent cells did not give rise to overt myeloid leukemia, which is in line with previous observations from PU.1-haploinsufficient mice[37,38]. Furthermore, we found minimally reduced PU.1 expression levels in aged, but otherwise healthy, human HSCs, which have also been shown to harbor a myeloid lineage bias[39,40].

Molecularly, PU.1 exerts many of its functions by binding highly specific DNA sequence motifs, often in concert with other transcription factors. Among the genes whose expression was significantly reduced in immature myeloid cells from URE^{het}*Msh2*$^{-/-}$ mice with AML, we identified *Irf8*, a PU.1 cofactor[30–32] that is frequently lost or impaired in human myeloid leukemia[30,41]. *Irf8*-deficient mice show dysfunctional granulocytic-monocytic lineage determination and develop a myeloproliferative neoplasm[42]. In human hematopoietic stem and progenitor cells, monocytes and macrophages, PU.1-IRF8 consensus DNA-binding motifs discriminate PU.1 consensus sites with actual PU.1 occupancy from those that are not bound by the transcription factor[43]. We found that restoration of Irf8 expression in leukemic URE^{het}*Msh2*$^{-/-}$ cells rescued the impaired expression of genes harboring PU.1-IRF consensus-binding sites, led to the loss of aberrant self-renewal, promoted myeloid differentiation and induced apoptosis, which demonstrates that Irf8 impairment functionally cooperates with lower PU.1 expression in our model. IRF8 downregulation has also been shown to cause resistance to apoptosis and to be associated with disease progression in human leukemia[44]. Our results provide evidence that minimal PU.1 reduction cooperates with IRF8 impairment in human leukemogenesis: (i) patients with myelodysplastic syndrome (MDS) with a higher risk for progression to AML had lower IRF8 levels, suggesting that inactivation of IRF8 promotes leukemogenesis, (ii) lower IRF8 expression was detected specifically in AML patients with lower PU.1 levels, (iii) restoration of IRF8 expression induced differentiation in 'IRF8 low' AML cells, and (iv) a positive correlation of PU.1 and IRF8 expression was found in human leukemia stem cells, but not in healthy HSCs.

Aging hematopoietic stem cells progressively accumulate mutations and molecular alterations which can impair cellular function, give rise to preleukemic stem cells and initiate age-associated myeloid malignancies[45,46]. By modeling the accumulation of small DNA lesions during the aging process by using an *Msh2*-deficient mouse background, we found that a minimal reduction in PU.1 expression triggered significant cellular and molecular alterations that were consistent with MDS, a preleukemic disorder that frequently progresses to AML[47]. Compound mutant URE^{het}*Msh2*$^{-/-}$ mice also frequently developed AML that, as in human AML, showed considerable phenotype heterogeneity. This variability is most likely due to differences in acquired genetic and epigenetic alterations in *Msh2*-deficient cells and further supports a role for PU.1-dependent stem cell fate

dysregulation as an early preleukemic event during leukemogenesis. Recent studies in patients with AML have confirmed the functional importance and clinical significance of preleukemic stem cell populations as the cellular origin of leukemia-initiating cells in human disease[7,27,47,48]. However, our functional and mechanistic knowledge of preleukemic stem cells and their progression is still very limited. Our study reveals a mechanism of leukemic transformation that is mediated by the minimal inhibition of a key transcriptional regulator of hematopoiesis, PU.1, which alters hematopoietic stem cell fate and sensitizes them to further malignant transformation.

Genetic models have demonstrated that alteration of enhancer function can cause hematologic diseases[49,50]. Very recently, AML has also emerged as such an 'enhanceropathy'[51]. Our study provides further proof that even the heterozygous disruption of critical *cis*-regulatory regions can be potent drivers of malignant transformation. It is possible that other key transcriptional regulators that exhibit small changes in expression levels play critical roles in malignant transformation, particularly in the early stages of tumorigenesis, and thus deserve our attention in the future.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Gene Expression Omnibus: gene expression data generated from preleukemic and leukemic myeloid progenitor cells derived from URE^{het}*Msh2*$^{-/-}$ mice as well as age-matched wild-type control animals are provided under accession code GSE65671.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We would like to thank K. Ozato for kindly providing the Irf8 expression vector. We thank A. Skoultchi, K. Gritsman and members of the Steidl laboratory for very helpful discussions and suggestions. We also thank G. Simkin and D. Sun of the Einstein Stem Cell Isolation and Xenotransplantation Facility (funded through New York Stem Cell Science (NYSTEM) grant no. C029154), C. Montagna and Z.X. Yang from the Einstein Genome Imaging Facility, D. Reynolds and W. Tran from the Einstein Genomics Core Facility and P. Schultes from the Einstein Department of Cell Biology for expert technical assistance. This work was supported by US National Institutes of Health (NIH) grant R00CA131503 (U.S.), Albert Einstein Cancer Center Core Support grant P30CA013330 and The Gabrielle's Angel Foundation for Cancer Research (U.S.). U.S. is a Research Scholar of the Leukemia and Lymphoma Society and the Diane and Arthur B. Belfer Faculty Scholar in Cancer Research of the Albert Einstein College of Medicine.

AUTHOR CONTRIBUTIONS

B.W., U.S., A.V. and W.E. designed the study and experiments. B.W., T.O.V., S.N., T.I.T., J.M., M.d.S.F., L.C., D.B.N., M.R., J.v.O. and S.S. conducted experiments. B.B., J.C., Y.Y., L.B. and B.W. performed gene expression, large data set and pathway analyses. C.M. and A.V. performed cell pathological analyses. B.W., T.O.V. and U.S. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Orkin, S.H. Diversification of hematopoietic stem cells to specific lineages. *Nat. Rev. Genet.* **1**, 57–64 (2000).
2. Iwasaki, H. *et al.* The order of expression of transcription factors directs hierarchical specification of hematopoietic lineages. *Genes Dev.* **20**, 3010–3021 (2006).
3. Iwasaki, H. *et al.* Distinctive and indispensable roles of PU.1 in maintenance of hematopoietic stem cells and their differentiation. *Blood* **106**, 1590–1600 (2005).
4. Dakic, A. *et al.* PU.1 regulates the commitment of adult hematopoietic progenitors and restricts granulopoiesis. *J. Exp. Med.* **201**, 1487–1502 (2005).
5. Scott, E.W., Simon, M.C., Anastasi, J. & Singh, H. Requirement of transcription factor PU.1 in the development of multiple hematopoietic lineages. *Science* **265**, 1573–1577 (1994).
6. Juliusson, G. *et al.* Age and acute myeloid leukemia: real world data on decision to treat and outcomes from the Swedish Acute Leukemia Registry. *Blood* **113**, 4179–4187 (2009).
7. Jan, M. *et al.* Clonal evolution of preleukemic hematopoietic stem cells precedes human acute myeloid leukemia. *Sci. Transl. Med.* **4**, 149ra118 (2012).
8. Hope, K.J., Jin, L. & Dick, J.E. Acute myeloid leukemia originates from a hierarchy of leukemic stem cell classes that differ in self-renewal capacity. *Nat. Immunol.* **5**, 738–743 (2004).
9. Rosenbauer, F. *et al.* Acute myeloid leukemia induced by graded reduction of a lineage-specific transcription factor, PU.1. *Nat. Genet.* **36**, 624–630 (2004).
10. Kuo, Y.H. *et al.* Cbfβ-SMMHC induces distinct abnormal myeloid progenitors able to develop acute myeloid leukemia. *Cancer Cell* **9**, 57–68 (2006).
11. Krivtsov, A.V. *et al.* Transformation from committed progenitor to leukaemia stem cell initiated by MLL-AF9. *Nature* **442**, 818–822 (2006).
12. Shimizu, R. *et al.* Leukemogenesis caused by incapacitated GATA-1 function. *Mol. Cell. Biol.* **24**, 10814–10825 (2004).
13. Metcalf, D. *et al.* Inactivation of PU.1 in adult mice leads to the development of myeloid leukemia. *Proc. Natl. Acad. Sci. USA* **103**, 1486–1491 (2006).
14. Cook, W.D. *et al.* PU.1 is a suppressor of myeloid leukemia, inactivated in mice by gene deletion and mutation of its DNA-binding domain. *Blood* **104**, 3437–3444 (2004).
15. Steidl, U. *et al.* Essential role of Jun family transcription factors in PU.1 knockdown–induced leukemic stem cells. *Nat. Genet.* **38**, 1269–1277 (2006).
16. Steidl, U. *et al.* A distal single-nucleotide polymorphism alters long-range regulation of the *PU.1* gene in acute myeloid leukemia. *J. Clin. Invest.* **117**, 2611–2620 (2007).
17. Yoshida, H. *et al.* PML–retinoic acid receptor α inhibits PML IV enhancement of PU.1-induced C/EBPε expression in myeloid differentiation. *Mol. Cell. Biol.* **27**, 5819–5834 (2007).
18. Mueller, B.U. *et al.* ATRA resolves the differentiation block in t(15;17) acute myeloid leukemia by restoring PU.1 expression. *Blood* **107**, 3330–3338 (2006).
19. Mizuki, M. *et al.* Suppression of myeloid transcription factors and induction of STAT response genes by AML-specific *Flt3* mutations. *Blood* **101**, 3164–3173 (2003).
20. Vangala, R.K. *et al.* The myeloid master regulator transcription factor PU.1 is inactivated by AML1-ETO in t(8;21) myeloid leukemia. *Blood* **101**, 270–277 (2003).
21. Bonadies, N., Pabst, T. & Mueller, B.U. Heterozygous deletion of the PU.1 locus in human AML. *Blood* **115**, 331–334 (2010).
22. Mueller, B.U. *et al.* Heterozygous PU.1 mutations are associated with acute myeloid leukemia. *Blood* **100**, 998–1007 (2002).
23. Mao, G. *et al.* Preferential loss of mismatch repair function in refractory and relapsed acute myeloid leukemia: potential contribution to AML progression. *Cell Res.* **18**, 281–289 (2008).
24. Zhu, Y.M., Das-Gupta, E.P. & Russell, N.H. Microsatellite instability and p53 mutations are associated with abnormal expression of the *MSH2* gene in adult acute leukemia. *Blood* **94**, 733–740 (1999).
25. Diouf, B. *et al.* Somatic deletions of genes regulating MSH2 protein stability cause DNA mismatch repair deficiency and drug resistance in human leukemia cells. *Nat. Med.* **17**, 1298–1303 (2011).
26. Zhang, S., Lloyd, R., Bowden, G., Glickman, B.W. & de Boer, J.G. Thymic lymphomas arising in Msh2-deficient mice display a large increase in mutation frequency and an altered mutational spectrum. *Mutat. Res.* **500**, 67–74 (2002).
27. Welch, J.S. *et al.* The origin and evolution of mutations in acute myeloid leukemia. *Cell* **150**, 264–278 (2012).
28. Smits, R. *et al.* Somatic *Apc* mutations are selected upon their capacity to inactivate the β-catenin downregulating activity. *Genes Chromosom. Cancer* **29**, 229–239 (2000).
29. Reitmair, A.H. *et al.* *MSH2*-deficient mice are viable and susceptible to lymphoid tumours. *Nat. Genet.* **11**, 64–70 (1995).
30. Schmidt, M., Bies, J., Tamura, T., Ozato, K. & Wolff, L. The interferon regulatory factor ICSBP/IRF-8 in combination with PU.1 upregulates expression of tumor suppressor p15^{Ink4b} in murine myeloid cells. *Blood* **103**, 4142–4149 (2004).
31. Pourcet, B. *et al.* LXRα regulates macrophage arginase 1 through PU.1 and interferon regulatory factor 8. *Circ. Res.* **109**, 492–501 (2011).
32. Meraro, D., Gleit-Kielmanowicz, M., Hauser, H. & Levi, B.Z. IFN-stimulated gene 15 is synergistically activated through interactions between the myelocyte-lymphocyte–specific transcription factors, PU.1, IFN regulatory factor 8–IFN consensus sequence–binding protein and IFN regulatory factor 4: characterization of a new subtype of IFN-stimulated response element. *J. Immunol.* **168**, 6224–6231 (2002).
33. Gekas, C. & Graf, T. CD41 expression marks myeloid-biased adult hematopoietic stem cells and increases with age. *Blood* **121**, 4463–4472 (2013).
34. Morita, Y., Ema, H. & Nakauchi, H. Heterogeneity and hierarchy within the most primitive hematopoietic stem cell compartment. *J. Exp. Med.* **207**, 1173–1182 (2010).
35. Appelbaum, F.R. *et al.* Age and acute myeloid leukemia. *Blood* **107**, 3481–3485 (2006).

36. Staber, P.B. *et al.* Sustained PU.1 levels balance cell-cycle regulators to prevent exhaustion of adult hematopoietic stem cells. *Mol. Cell* **49**, 934–946 (2013).
37. Dahl, R. *et al.* Regulation of macrophage and neutrophil cell fates by the PU.1: C/EBPα ratio and granulocyte colony-stimulating factor. *Nat. Immunol.* **4**, 1029–1036 (2003).
38. Genik, P.C. *et al.* Leukemogenesis in heterozygous *PU.1* knockout mice. *Radiat. Res.* **182**, 310–315 (2014).
39. Dykstra, B., Olthof, S., Schreuder, J., Ritsema, M. & de Haan, G. Clonal analysis reveals multiple functional defects of aged murine hematopoietic stem cells. *J. Exp. Med.* **208**, 2691–2703 (2011).
40. Pang, W.W. *et al.* Human bone marrow hematopoietic stem cells are increased in frequency and myeloid-biased with age. *Proc. Natl. Acad. Sci. USA* **108**, 20012–20017 (2011).
41. Schmidt, M., Hochhaus, A., Nitsche, A., Hehlmann, R. & Neubauer, A. Expression of nuclear transcription factor interferon consensus sequence–binding protein in chronic myeloid leukemia correlates with pretreatment risk features and cytogenetic response to interferon-α. *Blood* **97**, 3648–3650 (2001).
42. Holtschke, T. *et al.* Immunodeficiency and chronic myelogenous leukemia–like syndrome in mice with a targeted mutation of the *ICSBP* gene. *Cell* **87**, 307–317 (1996).
43. Pham, T.H. *et al.* Mechanisms of *in vivo* binding site selection of the hematopoietic master transcription factor PU.1. *Nucleic Acids Res.* **41**, 6391–6402 (2013).
44. Hu, X. *et al.* IRF8 regulates acid ceramidase expression to mediate apoptosis and suppresses myelogeneous leukemia. *Cancer Res.* **71**, 2882–2891 (2011).
45. Holstege, H. *et al.* Somatic mutations found in the healthy blood compartment of a 115-yr-old woman demonstrate oligoclonal hematopoiesis. *Genome Res.* **24**, 733–742 (2014).
46. Xie, M. *et al.* Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nat. Med.* **20**, 1472–1478 (2014).
47. Corces-Zimmerman, M.R., Hong, W.J., Weissman, I.L.,Medeiros, B.C. & Majeti, R. Preleukemic mutations in human acute myeloid leukemia affect epigenetic regulators and persist in remission. *Proc. Natl. Acad. Sci. USA* **111**, 2548–2553 (2014).
48. Shlush, L.I. *et al.* Identification of preleukaemic hematopoietic stem cells in acute leukaemia. *Nature* **506**, 328–333 (2014).
49. Mercola, M., Wang, X.F., Olsen, J. & Calame, K. Transcriptional enhancer elements in the mouse immunoglobulin heavy chain locus. *Science* **221**, 663–665 (1983).
50. Herranz, D. *et al.* A NOTCH1-driven *MYC* enhancer promotes T cell development, transformation and acute lymphoblastic leukemia. *Nat. Med.* **20**, 1130–1137 (2014).
51. Gröschel, S. *et al.* A single oncogenic enhancer rearrangement causes concomitant *EVI1* and *GATA2* deregulation in leukemia. *Cell* **157**, 369–381 (2014).

ONLINE METHODS

Mice. URE^{het}*Msh2*$^{-/-}$ were generated by mating *Msh2*-deficient mice (Msh2$^{-/-}$) (C57Bl/6) to *PU.1* URE^{het} mice. NOD-Scid IL2Rγ-null (NOG) and B6.SJL-Ptprca Pepcb/BoyJ mice were purchased from Jackson Labs (Bar Harbor, ME). All mice were housed in a special pathogen-free (SPF) barrier facility. All experimental procedures conducted on mice were approved by the Institutional Animal Care and Use Committee (IACUC; protocol #2013-1202, Albert Einstein College of Medicine). Mice with the indicated genotypes were included in the study without any further preselection or formal randomization and comprised balanced numbers from both genders; we used age- and gender-matched mice. Investigators were not blinded to genotype group allocations.

Histology. Femoral bones, spleens and livers were fixed for >24 h in neutrally buffered formalin at room temperature, subjected to paraffin embedding, cut into sections using a microtome and stained with hematoxylin and eosin (H&E) according to standard protocols. Cytospins of single-cell suspensions from bone marrow and spleen samples were prepared after erythrocyte lysis using ACK buffer pH 7.4 (0.15 M NH_4Cl, 10 mM $KHCO_3$, 1.0 mM EDTA). Cytospun cells were stained using a modified Giemsa stain (ThermoScientific Shandon Kwik-Diff Stains) according to the manufacturer's recommendation. Cell and tissue morphology was evaluated using an Axiovert 200M microscope (Zeiss, Maple Grove, MN) or an EVOS FL Auto microscope (Life Technologies, Grand Island, NY).

Complete blood counts. Peripheral blood was obtained from the mouse facial vein using standard techniques and analyzed using the Forcyte Hematology Analyzer (Oxford Science Inc., Oxford CT) according to the manufacturer's instructions.

Analysis and purification of hematopoietic stem and progenitor cells. Total bone marrow cells were isolated from the tibiae, femurs and pelvic bones of mice as previously described[52]. Isolated cells were treated with ACK buffer pH 7.4 to lyse red blood cells. After two washes with phosphate-buffered saline (without $MgCl_2$) containing 2% FBS (PBS-FBS), immature hematopoietic cells were enriched using magnetic bead negative enrichment with Dynabeads (Invitrogen, Carlsbad, CA). Cells were incubated with the following antibodies (all labeled with PE-Cy5–Tricolor): CD4 (GK1.5), CD8a (53-6.7), CD19 (1D3) (1:100 in PBS-FBS; all from eBioscience, San Diego, CA) and B220 (RA3-6B2), Gr-1 (RB6-8C5) (1:50 in PBS-FBS, both from Invitrogen, Carlsbad, CA) (referred to as "Lin"). After immunomagnetic separation, unbound cells were washed once and stained with the following antibody cocktail for isolation of hematopoietic stem (HSC) and progenitor cells (all 1:30 in PBS-FBS) for 30 min on ice: APC Alexa 750 CD117 (eBioscience, 2B8), Pacific Blue Sca-1 (BioLegend, D7), PE-Cy7 CD127 (eBioscience, A7R34), FITC CD41 (eBioscience, MWReg30), PE CD150 (BioLegend, TC15-12F12.2), APC Flt3 (eBioscience, A2F10). Cells were washed once and immediately sorted using an Aria II Special Order flow cytometer (Beckton Dickinson, San Jose, CA). Analysis of FACS data was done using the BD FACSDiva software (Beckton Dickinson).

Metaphase preparation of primary hematopoietic cells. cKit$^+$Lymph$^-$ cells were isolated from ACK-treated single-cell suspensions of the bone marrow or spleens of mice by FACS. To promote cell proliferation, single-cell suspensions were precultured for 12 h in M5300 containing 50 ng/ml rmSCF, 10 ng/ml rmIL-3, 10 ng/ml rmIL-6 (all from Peprotech) and 10 IU/ml rhEpo (Epogen, Amgen). The cells were then exposed to colcemid (1 h at 37 °C, KaryoMAX 10 μg/ml, Invitrogen), centrifuged for 10 min at 300*g* and incubated with hypotonic KCl solution (0.075 M KCl prewarmed at 37 °C, ThermoFisher). Afterwards, the cells were fixed and washed four times with a methanol:acetic acid solution (3:1). 40 μl of the cell suspension was dropped onto a clean slide, in 48% humidity and at 24 °C, de-identified for blinded analysis and then stored at 37 °C until analyzed. Chromosomes were counted in 10 metaphases for each condition.

Spectral karyoptyping (SKY). SKY was performed as described before[53]. Briefly, the slides with metaphases were denatured with 50% formamide in 2× SSC (Sigma) at 80 °C for 1.5 min and then dehydrated with serial ethanol washing steps (ice-cold 70%, 90% and 100% ethanol for 3 min each). Mouse SKY paint probes (Applied Spectral Imaging) were denatured in the hybridization solution (50% dextran sulfate in 2× SSC) at 85 °C for 5 min and then applied to the metaphase slides. After 72 h, the slides were then washed three times for 5 min each with 50% formamide in 2× SSC, 1× SSC and 4× SSC with 0.1% Tween. Slides were dehydrated with serial ethanol washing steps and mounted with ProLong Gold antifade reagent with DAPI (Invitrogen) for imaging. Images were acquired using an Axiovert 200 microscope (Zeiss, Germany) connected to an imaging interferometer (SD200, Applied Spectral Imaging, Migdal HaEmek, Israel). Defined rearrangements with nomenclature rules from the International Committee on Standard Genetic Nomenclature for Mice were used for classification.

Transplantation of bone marrow–derived cells. Total bone marrow cells were isolated from the tibiae, femurs and pelvic bones. Single-cell suspensions were generated and red blood cells were lysed using ACK-buffer treatment for 30 s on ice. Cells were washed once with PBS and stained with antibodies. Using an ARIA II Special Order System flow cytometer (Beckton Dickinson), Lin$^-$cKit$^+$Sca-1$^+$CD150$^+$IL7Rα$^-$Flt3$^-$CD48$^-$CD41$^-$ (lineage-unbiased HSC), CD150^{high}cKit$^+$Sca-1$^+$Lin$^-$IL7Rα$^-$Flt3$^-$CD48$^-$ (myeloid-biased HSC), Lin$^-$Sca-1$^+$cKit$^+$ (LSK), cKit$^+$B220$^-$CD4$^-$CD8a$^-$ (cKit$^+$Lymph$^-$) or cKit$^-$B220$^-$CD4$^-$CD8a$^-$ (cKit$^-$Lymph$^-$) cells were isolated. Sorted LSK, cKit$^+$Lymph$^-$ or cKit$^-$Lymph$^-$ cells were resuspended in HBSS (Thermo Fisher) and then transplanted into sublethally irradiated 4- to 6-week-old NOG recipient animals via retro-orbital injection 4 h after irradiation. Total body irradiation was delivered in a single dose of 250 cGy using a Shepherd 6810 sealed-source ^{137}Cs irradiator. Lineage-unbiased HSCs and myeloid-biased HSCs were transplanted into congenic recipient mice (B6.SJL-Ptprca Pepcb/BoyJ) 4 h after lethal irradiation (950 cGy total body irradiation) along with 2×10^5 CD45.1 and CD45.2 double-positive nucleated bone marrow cells for support. Engraftment of donor cells was monitored by analysis of CD45.2 and CD45.1 expression on peripheral blood cells as indicated in the figures. We stained ACK-treated peripheral blood cells with antibodies (all from eBioscience) against CD45.1 (A20), CD45.2 (104), CD11b (M1/70), Gr-1, CD4, B220 and Ter119 (TER-119) (1:100 in PBS, 2% FBS) and analyzed their binding by flow cytometry on an ARIA II Special order System (Beckton Dickinson). Animal numbers were chosen on the basis of previous experiences with transplanting hematopoietic cells from the parental mouse strain[9,16].

***In vivo* HSC proliferation assay.** To characterize *in vivo* cell cycle activity of HSCs, we used a previously described bromodeoxyuridine (BrdU) incorporation protocol[36]. Briefly, URE^{het} and wild-type mice were injected once with BrdU (100 mg per kg of body weight (intraperitoneally)) and were then given 0.8 mg/ml BrdU in the drinking water for 16 h. Bone marrow from mice were harvested, stained with antibodies against cell surface markers and fixed. BrdU incorporation was detected by using the BrdU Flow Kit (BD Biosciences) according to the manufacturer's recommendation. BrdU incorporation was measured by flow cytometry on gated CD150^{high}Sca1$^+$ckit$^+$CD48$^-$Lin$^-$Flt3$^-$ cells.

***In vitro* colony formation assay and serial replating assay.** To characterize the clonogenic capacity of cells, we plated 2,000 cKit$^+$CD4$^-$CD8a$^-$B220$^-$ or 5,000 GFP$^+$URE^{het}*Msh2*$^{-/-}$ AML cells (after transduction) in MethoCult M3434 GF+ (Stem Cell Technologies, Vancouver, BC) containing IL-3, IL-6, SCF and EPO as previously described[52]. Colonies were scored 8–10 d after plating using an Axiovert 200M microscope (Zeiss, Maple Grove, MN) or an EVOS FL Auto microscope (Life Technologies). After the first plating, we proceeded with serial replating assays until colony formation ceased. Cells (5,000–10,000 cells/ml) were replated in M3434 MethoCult GF+ and colonies were scored again after 8–10 d.

Restoration of Irf8 or PU.1 expression. Total bone marrow was isolated from a URE^{het}*Msh2*$^{-/-}$ mouse with AML, a single-cell suspension was prepared, and cells were cultured in M5300 (Stem Cell Technologies) containing 50 ng/ml recombinant mouse (rm) SCF, 10 ng/ml rmIL-3, 10 ng/ml rmIL-6 and 10 IU/ml human Epo for five weeks with weekly passage. After the initial five passages, cells were grown in cytokine-free M5300 for 10 more passages. Characterization of cell surface marker expression by FACS revealed high-intermediate cKit

doi:10.1038/nm.3936

expression, intermediate CD11b and Gr-1 expression, high CD44 (IM7) expression and the absence of B220, CD19, CD4 and CD8a expression. Exponentially growing URE^{het}*Msh2*$^{-/-}$ AML cells were transduced at a cell density of 5×10^5/ml with either a retroviral expression construct allowing for the ectopic expression of Irf8-eGFP or empty vector–eGFP alone using multiplicities of infection (MOI) ranging between 3 and 10 as previously described[54]. A lentiviral expression construct was used to ectopically express PU.1-IRES-GFP or GFP alone as described before[15]. 48 h after transduction, cultures were analyzed or sorted for eGFP$^+$ cells. Overexpression was verified by qRT-PCR and intracellular FACS. Cell death was assessed using the Annexin-V PE Apoptosis Detection kit according to the manufacturer's recommendations (eBioscience). For IRF8 expression rescue experiments, human AML cell lines were transduced with Irf8-eGFP or eGFP vectors[54]. Briefly, exponentially growing Kasumi-3, ML-2 and Molm14 cells were cultured in RPMI-1640 medium with 20% FBS, and HL-60, Thp-1 and Nomo-1 cells cultured in RPMI-1640 medium with 10% FBS were transduced in the presence of polybrene by spin infection. 72 h after transduction, cells were analyzed for GFP expression by FACS.

Assessment of protein expression. PU.1 protein levels were measured in purified myeloid progenitor cells (~22,000) and neutrophils (~300,000) by western blotting using a rabbit polyclonal antibody to PU.1 (Santa Cruz, clone T-21) as previously described[9]. IRF8 protein abundance was detected by intracellular FACS analysis in mouse and human AML cells using a monoclonal antibody (conjugated with APC) that detects both human and mouse IRF8 (eBioscience, clone V3GYWCH) following the manufacturer's recommendations. A mouse IgG1 K APC-conjugated isotype control (eBioscience, clone P3.6.2.8.1) was used to assess nonspecific binding. IRF8 protein abundance was expressed as isotype-subtracted mean fluorescence intensity (MFI).

RNA purification, real-time PCR and gene expression analysis by microarray. RNA was extracted from FACS-purified hematopoietic stem and progenitor cells using the RNeasy Micro kit (Qiagen). RNA quantity and quality was assessed using a 2100 Bioanalyzer (Agilent, Santa Clara, CA) device. For real time PCR, RNA was reverse transcribed using Superscript II (Invitrogen, Carlsbad, CA). Amplification of target genes was performed using the Universal PCR Power SYBR Green mix or TaqMan Universal PCR Master Mix (both from Applied Biosystems, Carlsbad, CA). cDNA was amplified in a final volume of 15 µl in 96-well, or 8 µl in 384-well, microtiter plates according to the manufacturer's recommendation. Primers and probes used for real-time PCR can be found in **Supplementary Table 8**. We performed real-time PCR using an ViiA7 instrument (Life Technologies) with one cycle of 50 °C (for 2 min) and 95 °C (for 10 min) followed by 40 cycles of amplification, with each cycle comprising the steps: 95 °C (for 15 sec) and 60 °C (for 1 min). Specific amplification for the target gene products was validated by melting curve analysis and Sanger sequencing. Target gene expression quantification was calculated using the Pfaffl model and normalized to *GAPDH* expression levels. For global gene expression analysis using microarrays, high-quality RNA (RNA integrity number ≥ 8) was amplified using the WT Ovation Pico RNA amplification system (Nugen, San Carlos, CA). After labeling with the GeneChip WT terminal labeling kit (Affymetrix, Santa Clara, CA), labeled cRNA of each individual sample was hybridized to an Affymetrix Mouse Gene 1.0ST microarray (Affymetrix), stained and scanned by GeneChip Scanner 3000 7G system (Affymetrix) according to standard protocols. Data reported in the manuscript are tabulated in the supplementary information and are available at GEO (GSE65671).

Analysis of microarray data. Raw data was normalized with the RMA algorithm (oligo package v. 1.3.0) under R (v. 3.1.2)/Bioconductor (v. 3.0). Differentially expressed genes (DEGs) were determined after SAM analysis (EMA package, v. 1.4.4); FDR and fold change (FC) cutoffs were as indicated. After filtering out unannotated and duplicate genes, the remaining genes were clustered by hierarchical clustering (using Euclidean distance and complete linkage) using R (plots generated using *gplots* (v.2.16.0)). Gene set enrichment analysis was performed using gene set enrichment analysis (GSEA; Broad Institute). Pathway analyses were performed using DAVID[55] and Ingenuity Pathway Analysis (IPA, Qiagen, Redwood City, www.qiagen.com/ingenuity).

Integrative analysis with published data sets. For direct target gene identification, we intersected binding peaks from published data sets of whole-genome chromatin immunoprecipitation (ChIP-seq) for PU.1 and Irf8 (GSE38824). We used the published bedgraph tracks GSM950325_201104_s_5_chipseq.ucsc.bedGraph (for Irf8) and GSM1031977_Tot2-IRF8_PU.1_Tags.bedGraph (for PU.1) and retained peaks with cutoffs of scores of 15 and 30, respectively. This was sufficiently stringent to eliminate background signal (from GSM950323_201104_s_4_chipseq.ucsc.bedGraph). We next identified genes with PU.1 or PU.1 and IRF8 co-occupancy within promoter regions (using several different region cutoffs, as indicated in the figure legends) that were also differentially expressed in URE^{het}*Msh2*$^{-/-}$ AML. HOMER was used to retrieve information about nearest TSS and for identification of known DNA binding motifs according to TRANSFAC. Bedtools (v.2.20.1) was used to find geometric overlaps. PWMScan was used to identify HOMER motifs.

Comparative pathway analyses between differentially expressed genes in mouse and human AML were performed using IPA and GSEA. Briefly, to identify common pathways with differential activation in mouse and human MDS or AML, we used the comparative analysis module under IPA and compared GSEA results on differentially expressed genes between mouse preleukemic or leukemic cells (versus age-matched wild-type cells) and human MDS (GSE19429, excluding −7/(del)7 patients) or AML (GSE13204, normal karyotype AML; GSE14468, excluding patients with −7/(del)7) versus healthy controls. To test whether mouse AML was molecularly similar to human AML with lower PU.1 expression, we generated gene sets from differentially expressed genes found in human 'PU.1 lower' AML samples as compared to those in healthy controls and 'PU.1 higher' AML versus healthy controls and performed GSEA analysis using genes found to have significantly altered expression in mouse AML (compared to wild type control). Significant enrichment was identified using FDR *q*-values of 5% (GSE13204) and 1% (GSE14468) as cutoff.

Statistical analysis. Statistical analysis of group comparisons was performed using Student's *t*-test, Wilcox rank sum test and log-rank test in Excel and GraphPad Prism or R, as indicated. Statistical significance was set at $P < 5\%$. Statistical evaluation of microarray expression data was performed using the built-in functions of MeV, Ingenuity Pathway analysis, GSEA or the respective R/Bioconductor packages. Sample sizes chosen are indicated in the individual figure legends and were not based on formal power calculations to detect prespecified effect sizes.

52. Will, B. *et al.* Satb1 regulates the self-renewal of hematopoietic stem cells by promoting quiescence and repressing differentiation commitment. *Nat. Immunol.* **14**, 437–445 (2013).
53. Montagna, C. *et al.* The *Septin* 9 (*MSF*) gene is amplified and overexpressed in mouse mammary gland adenocarcinomas and human breast cancer cell lines. *Cancer Res.* **63**, 2179–2187 (2003).
54. Tsujimura, H. *et al.* ICSBP–IRF-8 retrovirus transduction rescues dendritic cell development *in vitro*. *Blood* **101**, 961–969 (2003).
55. Huang, D., Sherman, B.T. & Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).