UTX-mediated enhancer and chromatin remodeling suppresses myeloid leukemogenesis through non-catalytic inverse regulation of ETS and GATA programs

Malgorzata Gozdecka[1, 2], Eshwar Meduri[2], Milena Mazan[1, 2], Konstantinos Tzelepis[1], Monika Dudek[1], Andrew J. Knights[3], Mercedes Pardo[4], Lu Yu[4], Jyoti S. Choudhary[4], Emmanouil Metzakopian[5], Vivek Iyer[6], Haiyang Yun[2], Naomi Park[7], Ignacio Varela[8], Ruben Bautista[9], Grace Collord[1], Oliver Dovey[1], Dimitrios A. Garyfallos[1], Etienne De Braekeleer[1], Saki Kondo[10], Jonathan Cooper[1], Bertie Göttgens[11] Lars Bullinger[12], Paul A. Northcott[13,14], David Adams[6], George S. Vassiliou[1,2,15,16], Brian J.P. Huntly[2,11,15,16]

1. Haematological Cancer Genetics, Wellcome Trust Sanger Institute, Hinxton, Cambridge, CB10 1SA, UK.

2. Wellcome Trust–MRC Stem Cell Institute, Cambridge Biomedical Campus, University of Cambridge, Cambridge CB2 0XY, UK.

3. Genomics of gene regulation, Wellcome Trust Sanger Institute, Hinxton, Cambridge, CB10 1SA, UK.

4. Proteomic Mass Spectrometry, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton CB10 1SA, UK

5. Mouse Genomics, Wellcome Trust Sanger Institute, Hinxton, Cambridge, CB10 1SA, UK.

6. Experimental Cancer Genetics, Wellcome Trust Sanger Institute, Hinxton, Cambridge, CB10 1SA, UK.

7. Sequencing Research Group, Wellcome Trust Sanger Institute, Cambridge CB10 1SA, UK.

8. Instituto de Biomedicina y Biotecnología de Cantabria (CSIC-UC-Sodercan), Departamento de Biología Molecular, Universidad de Cantabria, 39011 Santander, Spain.

9. New Pipeline Group, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton CB10 1SA, UK

10. Laboratory of Molecular Genetics, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan.

11. Cambridge Institute for Medical Research and Wellcome Trust/Medical Research Council, Stem Cell Institute and Department of Haematology, University of Cambridge, Cambridge, UK

12. Department of Internal Medicine III, Ulm University Medical Centre, Ulm, Germany; Medical Department, Division of Hematology, Oncology and Tumour Immunology, Charité Universitätsmedizin Berlin, Germany

13. Division of Pediatric Neurooncology, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany.

14. Developmental Neurobiology, St Jude Children's Research Hospital, Memphis, Tennessee 38105, USA.

15. Department of Haematology, Cambridge University Hospitals NHS Trust, Cambridge CB2 0QQ, UK.

16. These authors contributed equally to this work

Correspondence: George S. Vassiliou gsv20@sanger.ac.uk

Brian J.P. Huntly bjph2@cam.ac.uk

Summary

The H3K27 lysine-specific demethylase UTX is targeted by loss-of-function mutations in multiple cancers. Here, we demonstrate that UTX suppresses myeloid leukemogenesis through non-catalytic functions, a property shared with its catalytically inactive Y-chromosome paralogue, UTY. In keeping with this, we demonstrate concomitant loss/mutation of *UTX* and *UTY* in multiple human cancers. Mechanistically, global genomic profiling revealed only minor changes in H3K27Me3, but significant and bidirectional alterations of H3K27Ac and chromatin accessibility, a predominant loss of H3K4Me1 modifications, alterations in ETS and GATA factor binding and altered gene expression upon *Utx* loss. By integrating proteomic and genomic analyses, we link these changes to UTX regulation of ATP-dependent chromatin remodeling, coordination of the COMPASS complex and enhanced pioneering activity of ETS factors during evolution to AML. Collectively, our findings reveal a dual role for UTX in suppressing acute myeloid leukaemia via repression of oncogenic ETS and upregulation of tumor-suppressive GATA programs.

Introduction

Enzymatic modifications of histones play a central role in the control of gene expression to orchestrate diverse biological processes[1]. The JmjC-domain-containing protein ubiquitously transcribed tetratricopeptide repeat, X-linked (UTX or KDM6A), which demethylates di- and tri-methylated lysine-27 of histone H3 (H3K27Me3)[2], is a frequent target of somatic loss-of-function mutations in multiple cancer types[3,4,5,6,7,8] including leukemia[9-12]. Re-introduction of intact UTX into mutation-bearing cancer cells leads to significant transcriptional changes and a reduction in proliferation[7] in keeping with its role as a tumor suppressor, however the mechanism whereby UTX suppresses malignancy are poorly studied. Mechanistic insights into the tumor suppressive function of UTX have come from studies of T-cell acute lymphoblastic leukaemia (T-ALL) where an absence of UTX-catalytic function is pivotal for T-ALL initiation and maintenance[9]. Importantly, *UTX* mutations in T-ALL are almost exclusively found in males, reflecting the fact that the gene is X-linked and escapes X-inactivation[13], such that females (but not males) with single allele loss-of-function *UTX* mutations retain *UTX* expression[14]. Interestingly, in T-ALL mutations are concentrated in the catalytic JmjC domain, whereas this bias is not seen in other cancers where the mutations are spread throughout the gene[15], raising the possibility of different tumor suppressive mechanisms. Of potential relevance, UTY (KDM6C), the Y-chromosome homologue of UTX, has markedly reduced demethylase activity due to point substitutions affecting substrate-binding[16]. By contrast, in common with UTX, UTY contains an intact tetratricopeptide repeat region involved in protein-protein interactions that mediate demethylase-independent functions[17]. Tantalizingly, deletion of *UTY* was seen more

frequently than expected in cancer cell lines with mutations in *UTX*, than in those without[7], raising the possibility of a functional redundancy between the two paralogues.

Using myeloid malignancies as an exemplar, we investigate the role of *UTX* loss in oncogenesis and its interaction with *UTY* in mice following haematopoietic-specific loss of *Utx*. Our findings reveal that UTX prevents leukemogenesis by coordinate repression of pro-oncogenic ETS (E-twenty-six) and maintenance of tumor-suppressive GATA transcriptional programs. These effects are mediated by differential effects on genome-wide H3K27 acetylation, H3K4 monomethylation and chromatin accessibility, and their functional consequences were rescued by both UTY and enzymatically dead UTX, confirming their independence of demethylase activity.

Results

Homozygous loss of *Utx* induces spontaneous leukemia in a murine model.

To investigate the function of UTX in the hematopoietic system, we generated conditional *Utx* knock-out mice, where exon 3 of *Utx* is flanked with loxP sites ($Utx^{f/f}$)[18] (**Fig.1a**). $Utx^{f/f}$ mice were crossed into the inducible *Mx1-Cre* line, enabling efficient *Utx* recombination in hematopoietic stem and progenitor cells (HSPC) following polyinosinic-polycytidylic acid (pIpC) administration that activates the *Mx1*-promoter. $Utx^{f/f}$;*Mx1-Cre* mice treated with pIpC (hereafter $Utx^{-/-}$) showed loss of *Utx* mRNA and protein (**Fig.1b-c)** compared to pIpC-treated $Utx^{f/f}$ mice (hereafter $Utx^{+/+}$/WT). Female $Utx^{-/-}$, $Utx^{+/-}$ and $Utx^{+/+}$ mice were then aged and monitored for leukaemia development for up to 22 months. $Utx^{-/-}$ mice demonstrated significantly decreased survival in comparison to $Utx^{+/-}$ and $Utx^{+/+}$ female mice (**Fig.1d**). At necropsy, $Utx^{-/-}$ mice had a significantly increased spleen weight (**Fig.1e**) and a predominance of myeloid cells in the spleen and bone marrow (BM) (**Fig.1f**). Blood leukocyte counts (WBC) showed a variable rise and both platelet (PLT) and hemoglobin (HGB) levels were decreased (**Fig.S1 a-c**). Histological examination revealed that the majority (63%) developed acute myeloid leukaemia (AML) (**Fig.1g-h**). In contrast, no $Utx^{+/-}$ or $Utx^{+/+}$ mice developed AML. Splenocytes from leukaemic $Utx^{-/-}$ mice propagated the disease in secondary recipients (**Fig.1i**) verifying their full leukemogenic potential. Exome sequencing of seven $Utx^{-/-}$ AMLs showed no recurrently mutated genes, with the exception of *Skint11* (2/7 samples), and only occasional copy number alterations (**Fig.S1 d-f**). To more accurately mimic human disease we expressed the *AML1-ETO* fusion gene that commonly co-occurs with *UTX* mutations[10,19,20] in $Utx^{-/-}$ HSPC, and observed significantly reduced survival of recipient mice (**Fig.S1 g-i**).

Deregulation of HSPC number, function and differentiation following *Utx* loss

Our findings suggested that *Utx* loss confers a pre-leukaemic state on HSPCs, with transformation reliant on additional mutations. To characterize this pre-leukemic phase, we analyzed mice early after *Utx* deletion (4-5 weeks post-pIpC). As reported previously[21], $Utx^{-/-}$ mice demonstrated splenic enlargement in comparison to $Utx^{+/+}$ and $Utx^{+/-}$ mice (**Fig.2a-b**). We next examined the effect of UTX on hematopoietic differentiation and composition of the HSPC compartment, likely to harbor the initial target cell for transformation. $Utx^{-/-}$ mice demonstrated a significant expansion of HSPC progenitors (lineage negative, Lin^-) (**Fig.2c**), long-term and short-term hematopoietic stem cell (LT-HSC; ST-HSC) frequency (**Fig.2d, Fig.S2a**), an increase in the granulocyte–monocyte progenitor (GMP) and common myeloid progenitor (CMP) and a decrease in the megakaryocyte-erythroid progenitor (MEP) compartments (**Fig.2e-f, Fig.S2b**), as well as a significant reduction in common lymphoid

progenitors (CLP)(**Fig.2g, Fig.S2c)**. To assess HSPC function, we performed serial re-plating assays, observing enhanced self-renewal and proliferative potential of *Utx*$^{-/-}$ progenitors (**Fig.2h and Fig.S3a**). As regards mature cell numbers in BM, spleen and blood, only peripheral blood (PB) thrombocytopenia was noted at 5 weeks post-pIpC (**Fig.S3b-d**). However, at later time points (36 weeks post pIpC) in otherwise well animals, there was also an increase in PB MAC1^{+} myeloid cells and total WBC, and a reduction in B-cells in *Utx*$^{-/-}$ compared to *Utx*$^{+/+}$ mice (**Fig.S3e-g**). Collectively, these results demonstrate that biallelic loss of *Utx* leads to dramatic and progressive alterations in the composition, function and differentiation of HSPCs and their progeny, including enhanced self-renewal, myeloid expansion and a block in lymphoid and erythroid/megakaryocytic differentiation (**Fig.2i**).

UTY also suppresses leukemia induction and rescues the UTX-deficient pre-leukemic phenotypes

To define any role for UTY in suppressing leukemogenesis we also monitored hemizygous (*Utx*$^{-/Y}$, lacking *Utx* but expressing *Uty*) and control (*Utx*$^{+/Y}$) male mice over the same time period. Remarkably, and in stark contrast to *Utx*$^{-/-}$ females, *Utx*$^{-/Y}$ males showed no difference in survival or hematological phenotype compared to *Utx*$^{+/Y}$ mice (**Fig.2j)**. In particular, we observed no differences in spleen and liver weights, WBC counts, platelet or hemoglobin levels (**Fig.S3h-l)**. Also, no *Utx*$^{-/Y}$ mice developed AML, indicating that UTY also suppresses myeloid leukemogenesis (**Fig.2k**). Similarly, the presence of UTY in hemizygous males was sufficient to abrogate the abnormalities in pre-leukemic HSPCs, apart from decreases in the MEP and CLP compartments (**Fig.2l** and **Fig.S4a-i**). Importantly, CRISPR/Cas9-mediated knockout of *Uty* in *Utx*$^{-/Y}$ mice increased HSPC self-renewal, phenocopying *Utx*$^{-/-}$ female mice (**Fig.2m-n**).

As the only significant difference between UTY and UTX proteins is the lack of catalytic activity in the former, these findings suggested that catalytic activity is dispensable for their tumor suppressor functions. To further test this hypothesis, we identified an AML cell line, MONO-MAC6, with deletion of both *UTX* and *UTY*. Lentiviral expression of *UTX*, *UTY* or a catalytically-dead *UTX* mutant (*UTX-MT2*)[22] in MONO-MAC6 confirmed this hypothesis, with all constructs significantly suppressing proliferation *in vitro* (**Fig.3a-c**). In xenotransplant assays, MONO-MAC6 expressing *UTX*, *UTX-MT2* or *UTY* demonstrated slower growth and a significant survival advantage over *FLAG*-expressing control cells (**Fig.3d-f**). Taken together these studies demonstrate that the tumor suppressor functions of UTX do not require its catalytic activity and are shared with its catalytically-inactive paralogue UTY.

Concomitant loss of both *UTX* and *UTY* tumor suppression occurs in multiple human cancer types

Our findings indicate that UTY can suppress myeloid leukemogenesis, however unlike T-ALL[14], *UTX*-mutated AML cases show no gender bias. We therefore analysed the status of *UTY* in human male AML cell lines carrying *UTX* mutations, identified through the COSMIC database (**Fig.3g**). Of 4 male AML lines with *UTX* mutations/deletions, we confirmed loss of *UTY* expression in all 4 (**Fig.3h**). Analysis of exome sequencing data from COSMIC, revealed that all harbored a *UTY* microdeletion. Systematically extending our analysis, a further 7 male hematopoietic cell line with *UTX* mutations were identified, of which 4 had *UTY* microdeletions (**TableS1**). Strikingly, in informative solid organ cancers, 20/25 (80%) *UTX*-mutant male cell lines also demonstrated *UTY* microdeletion/mutation (**TableS1**), and we confirmed loss of *UTY* expression in a subset (10/13**, Fig.3i**).

Integrated genome-scale analysis identifies altered enhancer function as a mediator of leukemogenesis following *Utx* loss

To determine the molecular basis of UTX-mediated leukemia suppression, we performed integrated genome-scale analyses (RNA-Seq, ChIP-Seq and ATAC-Seq) in HSPCs from age-matched pre-leukemic $Utx^{-/-}$, $Utx^{-/Y}$ and $Utx^{+/+}$ mice. As anticipated, in comparison to controls, the number of differentially expressed genes was significantly greater for $Utx^{-/-}$ than $Utx^{-/Y}$ (4497 vs 673 genes, $P < 0.05$; **TableS2-S4**). As loss of their single *Utx* allele did not lead to AML in hemizygous $Utx^{-/Y}$ males, we removed these 673 genes from subsequent analysis. Focusing on mRNAs differentially expressed with a $\log_2$ fold change > ±0.5, we identified 2686 genes (**Fig. 4a-b**). Interestingly, and somewhat counterintuitive to the perception of UTX as solely a transcriptional activator, similar numbers of genes were upregulated (1517, 57%) upon *Utx* loss, as were downregulated (1169, 43%). Importantly, although additional genes were also differentially expressed upon evolution to frank AML in $Utx^{-/-}$ mice, significant components of the pre-leukemic transcriptional programs were retained (**Fig. 4c, TableS5**). Using ChIP-Seq in WT mice, we documented 8304 UTX binding sites, corresponding to 6734 genes (**TableS6**), with the majority found at the promoter or within the gene body (**Fig. 4d**). Correlation with gene expression demonstrated that 581/1169 (50%) of downregulated and 614/1517 (40%) of upregulated gene loci were bound by UTX (**TableS7**), suggesting that around half of deregulated genes are direct UTX targets (**Fig. 4e**) and confirming UTX as both a transcriptional activator and repressor.

In keeping with our finding that H3K27-demethylase activity is redundant for tumor suppression, only 302 differentially modified H3K27Me3 peaks were observed between $Utx^{-/-}$ and $Utx^{+/+}$ HSPC . Further corroborating this notion, the majority (200/302, 67%) also showed decreased rather than increased modification (**Fig. 4f, TableS8**). In marked contrast, the coordinated acetyl modification at the same lysine residue, H3K27Ac, was markedly altered in its distribution in $Utx^{-/-}$ in comparison to

Utx$^{+/+}$ mice. We observed 5121 regions with differential H3K27Ac in either direction (corresponding to 2916 gene loci), including 3442 peaks (2054 gene loci) significantly decreased and 1679 (953 gene loci) increased, following UTX loss (**Fig.4g and Tables S9-S10**). Comparing these putative enhancer regions to a recently published promoter-based capture-HiC dataset in the hematopoietic stem and progenitor cell line HPC7[23], we observed that 23% of upregulated and 32% of downregulated regions interacted with promoters (**Fig.S5a-c**), suggesting significant enhancer remodeling on UTX loss. The observed changes were locus-specific, as global levels of H3K27Ac and H3K27Me3 were similar between *Utx*$^{+/+}$ and *Utx*$^{-/-}$ BM (**Fig.S5d**). In order to define direct co-occurrence on chromatin we overlapped UTX peaks and differential H3K27Ac regions. However, we found only a modest co-occurrence, with just 282/5121 regions (6%) in common (**Fig.S5e)**. UTX-peaks and differential H3K27Ac regions were then annotated to their associated/adjacent genes, defining larger genomic areas for comparison. In contrast to the limited peak-to-peak co-occurrence, when we compared whole gene loci showing differential H3K27Ac and UTX binding, we identified a highly significant overlap of 1396/2916 regions (48%, **Fig.S5f, TableS11**), suggesting that UTX indirectly regulates acetylation of regions adjacent to its binding. To further address the impact of UTX on enhancer function, we performed ChIP-Seq for the canonical enhancer defining mark H3K4Me1 in pre-leukaemic *Utx*$^{-/-}$ and *Utx*$^{+/+}$ HSPC. Of particular interest, H3K4Me1 is deposited by KMT2C/D, components of the COMPASS complex that are known UTX interaction partners[24]. We observed 4552 differentially modified H3K4Me1 peaks (**Fig. 4h**; **TableS12**); of which the majority, 3898, were downregulated in *Utx*$^{-/-}$ mice. Differentially downregulated H3K4Me1 regions highly correlated with peaks that also lost H3K27Ac (1589 common peaks - 46% overlap). UTX did not bind directly at these sites, proposing an indirect role of UTX and the COMPASS complex in early enhancer specification. Taken together, these data confirm that H3K27 demethylase activity is dispensable for tumor suppression, identify UTX as both a transcriptional activator and repressor and suggest that UTX/COMPASS-mediated indirect regulation of enhancers is important for transformation.

Loss of *Utx* activates an oncogenic ETS transcriptional program during leukemia development.

As our genomic data demonstrated both loss and gain of activating chromatin marks and up- and downregulation of gene expression upon *Utx* loss, we speculated that both processes were necessary for leukaemia development, but may be mediated by different mechanisms. We therefore analysed up and down regulated gene programs separately. Notably, several ETS transcription factors, including *Elf4*, *Etv6*, *Erg*, *Fli1*, *Ets2* and *Elk* were upregulated immediately following *Utx* loss in the pre-leukaemic phase (**Fig.5a**). Importantly, overexpression of these ETS factors persisted in *Utx*$^{-/-}$ AML (**Fig.5b)**. GSEA analysis also demonstrated a significant correlation between upregulated

pre-leukaemic transcriptional programs and genes repressed upon knockdown of the oncogenic ETS-factor fusion EWSR1-FLI1 (**Fig.5c**). In addition, motif analysis of 614 UTX-bound and upregulated genes demonstrated a significant enrichment in ETS binding motifs (**Fig.5d**, **TableS30**). To directly link an ETS program to leukemogenesis we mapped global chromatin occupancy of the ETS factor PU.1 in *Utx*$^{-/-}$ and *Utx*$^{+/+}$ HSPCs. We observed 8329 enriched and 5869 depleted PU.1 peaks in *Utx*$^{-/-}$ (**TableS13**). Many ETS-factors overexpressed in *Utx*$^{-/-}$ and directly bound by UTX in wild type cells, including *Ets2* and *Fli1* showed higher PU.1 promoter occupancy in the absence of UTX (**Fig.S5g**). These data suggest enhanced ETS factor binding and transcriptional auto-regulation of specific ETS genes in the absence of *Utx*. On a global scale however, UTX binding did not co-localize with PU.1 chromatin occupancy (only 115 peaks/14198, 0.8%), suggesting that the profound general redistribution of PU.1 binding may relate more to its overexpression. ETS factors are known to recruit histone acetyltransferases[25], and thus increase H3K27Ac deposition, therefore we next asked if the changes in differential H3K27Ac and H3K4Me1 observed in *Utx*$^{-/-}$ correlate with UTX-dependent redistribution of PU.1 binding. Indeed, we observed that 51% of gained H3K4Me1 (335/654) and 30% of gained H3K27Ac (470/1679) peaks also demonstrated increased PU.1 binding (**Fig**.**S6a-b**). These direct binding data provide a mechanistic explanation for the gain of enhancer marks upon *Utx* loss.

To investigate the functional significance of deregulated ETS factors we utilized CRISPR/Cas9 genome editing to ablate a number of these factors in *Cas9*-expressing, *UTX*/*UTY* mutant, MONO-MAC6 cells. Importantly, we observed significant growth suppression upon editing of the same *ETS* factors immediately overexpressed upon *Utx* loss: *FLI1*, *ERG*, *PU.1*, *ETS1*, and *ELF4* (**Fig.5e**). These data conclusively indicate that ETS factors drive leukemia induction and maintenance in the context of *UTX*/*UTY* loss.

Utx loss affects BRG1-dependent chromatin remodeling to repress a tumor suppressive GATA program during leukemia development.

Analysis of the downregulated gene-expression program using GSEA, demonstrated enrichment for GATA2 targets (**Fig.5f**). Furthermore, motif analysis at the 3442 regions with significantly decreased H3K27Ac upon *Utx* loss demonstrated a striking and almost exclusive enrichment in GATA binding motifs (**Fig.5g, TableS30**). Similarly to the ETS motifs, loss of H3K27Ac did not directly overlap with UTX binding (**Fig.S6c-d**) but occurred nearby and likely affected the same genes (**Fig.S6e**, **TableS14**). To better understand the indirect effect of UTX on acetylation changes at GATA and ETS sites, we performed pulldown of endogenous UTX in the murine myeloid cell line 416B, followed by mass spectrometric analysis. We did not observe UTX interaction with ETS-factors or GATA2 but we did

identify multiple known UTX interactors, including KMT2C/D/COMPASS members (**Fig.5h, Tables S15-S16**). Interestingly, our proteomic analysis also demonstrated lower level interactions between UTX and the ATP-dependent chromatin remodeling complex members BRG1 (SMARCA4) and CHD4, which we further verified (**Fig.S6f**). Speculating that changes in acetylation and GATA binding occurred through alterations in chromatin accessibility, we performed ATAC-Seq analysis in pre-leukaemic *Utx*$^{-/-}$ HSPC and *Utx*$^{+/+}$ controls. Strikingly, we observed significant and bidirectional changes in chromatin accessibility, including 7200 sites where chromatin accessibility decreased and 5244 sites that became accessible upon *Utx* loss (**TablesS17-18**). Loss of chromatin accessibility correlated strongly with decreased H3K27Ac (2274/3442 peaks, 73%) and H3K4Me1 deposition (2264/3898, 58%) (**Fig.S7a,d**) and decreased gene expression (**Fig.S7b, TableS19**). Furthermore, analysis of closed chromatin sites showed a striking enrichment in GATA motifs (**Fig.6a, Fig.S7a-b).** Comparisons of GATA2 peaks in HPC-7[26] with differential chromatin accessibility in *Utx*$^{-/-}$ HSPC demonstrated direct correlation only with closed chromatin regions (15%, 409/2796 peaks), as only 0.7% (19/2796 peaks) overlapped with open chromatin. As before, only limited overlap was found between closed chromatin and UTX binding (**Fig.S7c**). We next asked if UTX binds with BRG1 and CHD4 on chromatin. Importantly, we observed a highly significant overlap, 91% (7541/8304 peaks) of UTX sites with binding of both BRG1 and CHD4 from published ChIP-Seq[27] (**Fig.6c-e**). To validate co-occupancy we employed CRISPR/Cas9 genome editing of *Utx* in 416B cells targeting *Utx* exon-3, recapitulating our mouse model. We observed significantly lower chromatin binding for BRG1 and CHD4 at exemplar loci (*Aff1* and *Lrrc8c*) in the absence of UTX, in keeping with a role for UTX in the recruitment of complexes containing these proteins (**Fig.S7e-g**).

To further address molecular mechanism of UTX-dependent chromatin remodeling, we analysed the activity of BRG1 and CHD4 in *Utx*$^{-/-}$ and *Utx*$^{+/+}$ HSPCs. We crossed *Utx*$^{f/f}$;*Mx1-Cre* mice with *Cas9*-expressing mice[28] and induced *Utx* deletion. Five weeks post gene deletion we isolated HSPCs and used CRISPR/Cas9 genome editing to target *Brg1* and *Chd4*, using an empty gRNA construct as control. We then analyzed chromatin accessibility by ATAC-Seq, hypothesizing that loss of *Brg1* and/or *Chd4* would at least partially phenocopy the effect of UTX loss. We found 1150 sites with significantly decreased accessibility upon *Brg1* editing and only 16 sites gained, suggesting that BRG1 is mainly involved in the opening or maintenance of specific open chromatin loci (**TableS20**). We then overlapped these sites with regions that were differentially lost between *Utx*$^{+/+}$ versus *Utx*$^{-/-}$ cells using the same culture conditions (2871 peaks, **TableS21**). We observed that 21% (244/1150) of regions closed in *Brg1*-edited cells overlapped with sites that closed upon *Utx* deletion (**Fig.S8a-b**). Of note, there was no further alteration of accessibility at these 244 sites when *Brg1* was edited in *Utx*$^{-/-}$ cells (**TableS22**). This suggests a degree of functional redundancy between UTX and BRG1 loss

for chromatin accessibility. Performing motif analysis of these 244 sites, we again observed high enrichment for GATA motifs (**Fig.S8a-b**). Similar analysis for CHD4 demonstrated no significant overlap upon *Utx* loss (**Fig.S8c, TableS23**). These findings, along with our proteomic data suggest that UTX interacts with BRG1 to maintain chromatin accessibility at GATA bound regions.

Utx loss allows chromatin accessibility to other transcription factors and facilitates the pioneering function of ETS factors during AML evolution.

For sites newly accessible by ATAC-Seq following *Utx* loss, the converse was seen, with a correlation between these sites and increased H3K27Ac (766/1679 peaks, 45%)(**Fig.S9a**), H3K4Me1 (438/654 peaks, 67%)(**Fig.S9b**) and gene expression close to the accessible sites (389/1517, 25%)(**Fig.S9c, TableS24)**. Furthermore, motif analysis of newly accessible sites revealed enrichment for a number of transcription factors including ASCL1, E2A, EBF, PTF1a, TCF12, in addition to ETS factors (**Fig.6b**). Three transcription factors represented in the top five motifs (ASCL1, EBF, PTF1A) were not expressed in MONO-MAC6. Assessing the functional relevance of the two remaining expressed TFs, TCF3/E2A and TCF12/HEB, we utilized CRISPR/Cas9 editing in MONO-MAC6 and observed a significant reduction of cell growth upon knockout of either *TCF3* or *TCF12* (**Fig.S9d**) These data suggest that the transcriptional activity of TCF3 and TCF12 maintain AML growth in the absence of UTX. In assessing ETS sites, PU.1 binding sites gained in the absence of UTX only minimally overlapped with open chromatin at the pre-leukemic stage (758/8329 peaks, 9%). However, of significant interest, although the remaining 91% (7592 peaks, linked to 3450 genes) occurred in ATAC inaccessible chromatin in the PL stage and had no effect on gene expression following *Utx* loss, we documented that the expression of 691 of these linked genes (20%) was upregulated in the later transition to AML **(Fig.6f-h, TableS25-S26).** These data suggest that pioneering function of redistributed ETS TFs "primes" a later leukemogenic transcriptional program for upregulation during AML evolution.

Discussion

The mechanism of tumor suppression by UTX has been previously thought to be dependent on its demethylase catalytic function, a notion supported by data from T-ALL[9,14]. However, during the evolution of AML, we categorically demonstrate that the demethylase function of UTX is redundant for tumor suppression. Non-catalytic functions of UTX have previously been demonstrated in embryonic development[22,29,30] and in mammary luminal lineage development[31]. UTX-catalytic activity was reported to upregulate expression of the master transcription factor PLZF and to modulate

super-enhancer accessibility in invariant natural killer T cells[32]. Additionally, UTX function was recently linked to enhancer activity and gene activation via coordination of COMPASS mediated H3K4monomethylation and CREBBP/p300 H3K27Ac[33], as we also demonstrate here. However, no role for these functions in tumor suppression has been demonstrated until now. Furthermore, our study confirms and defines the molecular basis for the frequent co-occurrence of *UTX* and *UTY* mutation/loss[6,7], identifying UTY as a tumor suppressor in its own right and highlighting non-catalytic functions of UTX/UTY as the dominant mediators of tumor suppression.

Contrary to the prevailing perception that UTX is solely a transcriptional activator[17], our study identifies that it also functions as a repressor. We demonstrate that its loss leads to significant alterations in chromatin accessibility, bidirectional alterations in H3K27Ac, a predominant loss of H3K4Me1 and coordinated changes in gene expression that impart pro-leukemic properties on HSPCs, alterations that are maintained during evolution to AML (**Fig.S11**). In particular, UTX loss upregulates a transcriptional program driven by the ETS family of pioneer transcription factors[34,35,36]. ETS factors are known to be oncogenic[37,38], with overexpression of a single ETS factor, ERG, able to generate AML in mice[39] and *ERG* expression levels being one of the strongest prognostic factors in human AML[40]. Novel binding events of the exemplar ETS factor PU.1, lead to an increase in chromatin accessibility, activation of enhancer modifications and, via its pioneering activity, gene activation occurring at later timepoints during leukemia evolution. Furthermore, loss of UTX also downregulates a program of GATA-driven genes through loss of chromatin accessibility and local H3K27 acetylation. GATA factors are also critical regulators of hematopoiesis and leukemia[41], with germline and somatic loss-of-function mutations of *GATA2* described in AML[42,43]. Taken together, our data demonstrate that UTX loss coordinates a “double-hit” mechanism reminiscent of genomic inversion that removes a critical enhancer to downregulate GATA2 expression and relocates it to drive oncogenic expression of the transcription factor EVI1[44].

Our proteomic data also suggest that non-catalytic tumor suppressive functions operate through protein-protein interactions with tetratricopeptide repeats, the other major protein domain of UTX/UTY[17]. We could demonstrate protein interactions between endogenous UTX and the KMT2C/D-containing COMPASS complex and the ATP-dependent chromatin remodeling factors, BRG1 and CHD4 in myeloid cells. The interaction between UTX and BRG1 was previously demonstrated in T lymphocytes[45] and during cardiac development[46]. Loss of *Utx* led to a marked decrease in the deposition of the canonical early enhancer mark H3K4Me1. Notably, mutations of *KMT2D* are responsible for the majority (60-80%) of cases of Kabuki syndrome, with *UTX* mutations causing ~ 10% of cases, thus highlighting the mechanistic links between the two proteins. Furthermore, we

could demonstrate co-occupancy of UTX, BRG1 and CHD4[27] at specific genomic loci associated with alterations in chromatin accessibility upon UTX loss (**Fig. 6c-e**). Furthermore, we could functionally demonstrate that BRG1 loss at least in part phenocopied the loss of chromatin accessibility seen upon *Utx* deletion. Recent studies demonstrate that H3K4Me1 is required for binding of the BRG1-containing BAF complex to chromatin and enhances BAF chromatin-remodeling activity[47]. Our own data corroborate and extend this model further defining a role for UTX in linking H3K4Me1 deposition with chromatin remodeling via BRG1. Taken all together, these mechanistic data demonstrate that loss of UTX leads to an upregulated activity of ETS transcription factors with both immediate and later pioneering effect to facilitate chromatin accessibility, and loss of coordination of COMPASS-mediated H3K4Me1 enhancer specification and BRG1-mediated chromatin accessibility. Together, these lead to alter patterns of gene expression to induce and maintain leukemia (**Fig.S11**).

Our findings identify UTX as a complex transcriptional regulator capable of both activating and repressing transcription, through effects on pioneering transcription factors, enhancer function and chromatin accessibility, with obvious implications for its role in tumor suppression and other critical cellular processes. Finally, our framing of *UTY* as a *bona fide* tumor suppressor gene firmly establishes a pathogenic role for Y chromosome-specific genes in carcinogenesis, and throws new light on the role of Y chromosome loss in diverse cancer types and on the significance of age-related clonal hematopoiesis associated with "loss-of-Y" in otherwise healthy men[48-50].

Acknowledgements

This study was primarily funded by a joint Bloodwise Program Grant (17006) to B.H. and G.S.V. Work in the Huntly lab is also funded by an ERC consolidator award (grant 647685 COMAL), a CRUK program award, the Medical Research Council, (MRC) the Welcome Trust (WT) and the Cambridge NIHR BRC. We acknowledge the WT/MRC center grant (097922/Z/11/Z) and support from WT strategic award 100140. G.S.V. is funded by a Cancer Research UK Senior Cancer Research Fellowship (C22324/A23015). The Vassiliou laboratory is also supported by the Kay Kendall Leukemia Fund and core funding from the Sanger Institute (WT098051).

Author Contributions

M.G., G.S.V. and B.J.P.H conceived the study, designed the experiments and prepared the manuscript. M.G. conducted the majority of the experiments. E.M. performed ChIP-Seq, ATAC-Seq and Motif analysis. A.J.K performed ATAC-Seq experiment. M.P and M.G. prepared samples for mass spectrometry, M.P, L.Y. and J.C. conducted mass spectrometry and related data analysis. E.M.

designed and generated pKLV-puro vectors. V.I. and D.A. performed exome analysis. H.Y performed promoter-enhancer interaction analysis. N.P. and I.V. performed experimental and computational analysis. G.C., M.M., M.D., O.D., K.T., E.B and J.C. performed cell culture and mouse experiments. R.B. performed analysis of RNA-Seq data. P.A.N., B.G., and L.B. provided genomic data and expertise. S.K. helped with vector generation. All authors reviewed and agreed with the final submission.

Competing Financial Interests

Authors states no competing financial interests.

References

1 Jenuwein, T. & Allis, C. D. Translating the histone code. *Science* **293**, 1074-1080, doi:10.1126/science.1063127 (2001).

2 Agger, K. *et al.* UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. *Nature* **449**, 731-734, doi:10.1038/nature06145 (2007).

3 Dalgliesh, G. L. *et al.* Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. *Nature* **463**, 360-363, doi:10.1038/nature08672 (2010).

4 Bailey, P. *et al.* Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature* **531**, 47-52, doi:10.1038/nature16965 (2016).

5 Gui, Y. *et al.* Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder. *Nature genetics* **43**, 875-878, doi:10.1038/ng.907 (2011).

6 Robinson, G. *et al.* Novel mutations target distinct subgroups of medulloblastoma. *Nature* **488**, 43-48, doi:10.1038/nature11213 (2012).

7 van Haaften, G. *et al.* Somatic mutations of the histone H3K27 demethylase gene UTX in human cancer. *Nature genetics* **41**, 521-523, doi:10.1038/ng.349 (2009).

8 Huether, R. *et al.* The landscape of somatic mutations in epigenetic regulators across 1,000 paediatric cancer genomes. *Nat Commun* **5**, 3630, doi:10.1038/ncomms4630 (2014).

9 Ntziachristos, P. *et al.* Contrasting roles of histone 3 lysine 27 demethylases in acute lymphoblastic leukaemia. *Nature* **514**, 513-517, doi:10.1038/nature13605 (2014).

10 Cancer Genome Atlas Research, N. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *The New England journal of medicine* **368**, 2059-2074, doi:10.1056/NEJMoa1301689 (2013).

11 Jankowska, A. M. *et al.* Mutational spectrum analysis of chronic myelomonocytic leukemia includes genes associated with epigenetic regulation: UTX, EZH2, and DNMT3A. *Blood* **118**, 3932-3941, doi:10.1182/blood-2010-10-311019 (2011).

12 Wouters, B. J. & Delwel, R. Epigenetics and approaches to targeted epigenetic therapy in acute myeloid leukemia. *Blood* **127**, 42-52, doi:10.1182/blood-2015-07-604512 (2016).

13 Greenfield, A. *et al.* The UTX gene escapes X inactivation in mice and humans. *Hum Mol Genet* **7**, 737-742 (1998).

14 Van der Meulen, J. *et al.* The H3K27me3 demethylase UTX is a gender-specific tumor suppressor in T-cell acute lymphoblastic leukemia. *Blood* **125**, 13-21, doi:10.1182/blood-2014-05-577270 (2015).

15 Arcipowski, K. M., Martinez, C. A. & Ntziachristos, P. Histone demethylases in physiology and cancer: a tale of two enzymes, JMJD3 and UTX. *Curr Opin Genet Dev* **36**, 59-67, doi:10.1016/j.gde.2016.03.010 (2016).

16 Walport, L. J. *et al.* Human UTY(KDM6C) is a male-specific N-methyl lysyl demethylase. *J Biol Chem* **289**, 18302-18313, doi:10.1074/jbc.M114.555052 (2014).

17 Van der Meulen, J., Speleman, F. & Van Vlierberghe, P. The H3K27me3 demethylase UTX in normal development and disease. *Epigenetics* **9**, 658-668, doi:10.4161/epi.28298 (2014).

18 Skarnes, W. C. *et al.* A conditional knockout resource for the genome-wide study of mouse gene function. *Nature* **474**, 337-342, doi:10.1038/nature10163 (2011).

19 Faber, Z. J. *et al.* The genomic landscape of core-binding factor acute myeloid leukemias. *Nature genetics* **48**, 1551-1556, doi:10.1038/ng.3709 (2016).

20 Papaemmanuil, E. *et al.* Genomic Classification and Prognosis in Acute Myeloid Leukemia. *The New England journal of medicine* **374**, 2209-2221, doi:10.1056/NEJMoa1516192 (2016).

21 Thieme, S. *et al.* The histone demethylase UTX regulates stem cell migration and hematopoiesis. *Blood* **121**, 2462-2473, doi:10.1182/blood-2012-08-452003 (2013).

22 Wang, C. *et al.* UTX regulates mesoderm differentiation of embryonic stem cells independent of H3K27 demethylase activity. *Proc Natl Acad Sci U S A* **109**, 15324-15329, doi:10.1073/pnas.1204166109 (2012).

23 Wilson, N. K. *et al.* Integrated genome-scale analysis of the transcriptional regulatory landscape in a blood stem/progenitor cell model. *Blood* **127**, e12-23, doi:10.1182/blood-2015-10-677393 (2016).

24 Cho, Y. W. *et al.* PTIP associates with MLL3- and MLL4-containing histone H3 lysine 4 methyltransferase complex. *The Journal of biological chemistry* **282**, 20395-20406, doi:10.1074/jbc.M701574200 (2007).

25 Foulds, C. E., Nelson, M. L., Blaszczak, A. G. & Graves, B. J. Ras/mitogen-activated protein kinase signaling activates Ets-1 and Ets-2 by CBP/p300 recruitment. *Molecular and cellular biology* **24**, 10954-10964, doi:10.1128/MCB.24.24.10954-10964.2004 (2004).

26 Wilson, N. K. *et al.* Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators. *Cell stem cell* **7**, 532-544, doi:10.1016/j.stem.2010.07.016 (2010).

27 Morris, S. A. *et al.* Overlapping chromatin-remodeling systems collaborate genome wide at dynamic chromatin transitions. *Nat Struct Mol Biol* **21**, 73-81, doi:10.1038/nsmb.2718 (2014).

28 Tzelepis, K. *et al.* A CRISPR Dropout Screen Identifies Genetic Vulnerabilities and Therapeutic Targets in Acute Myeloid Leukemia. *Cell reports* **17**, 1193-1205, doi:10.1016/j.celrep.2016.09.079 (2016).

29 Shpargel, K. B., Starmer, J., Yee, D., Pohlers, M. & Magnuson, T. KDM6 demethylase independent loss of histone H3 lysine 27 trimethylation during early embryonic development. *PLoS Genet* **10**, e1004507, doi:10.1371/journal.pgen.1004507 (2014).

30 Morales Torres, C., Laugesen, A. & Helin, K. Utx is required for proper induction of ectoderm and mesoderm during differentiation of embryonic stem cells. *PloS one* **8**, e60020, doi:10.1371/journal.pone.0060020 (2013).

31 Yoo, K. H. *et al.* Histone Demethylase KDM6A Controls the Mammary Luminal Lineage through Enzyme-Independent Mechanisms. *Molecular and cellular biology* **36**, 2108-2120, doi:10.1128/MCB.00089-16 (2016).

32 Beyaz, S. *et al.* The histone demethylase UTX regulates the lineage-specific epigenetic program of invariant natural killer T cells. *Nature immunology* **18**, 184-195, doi:10.1038/ni.3644 (2017).

33 Wang, S. P. *et al.* A UTX-MLL4-p300 Transcriptional Regulatory Network Coordinately Shapes Active Enhancer Landscapes for Eliciting Transcription. *Molecular cell* **67**, 308-321 e306, doi:10.1016/j.molcel.2017.06.028 (2017).

34 Zaret, K. S. & Mango, S. E. Pioneer transcription factors, chromatin dynamics, and cell fate control. *Curr Opin Genet Dev* **37**, 76-81, doi:10.1016/j.gde.2015.12.003 (2016).

35 Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell* **38**, 576-589, doi:10.1016/j.molcel.2010.05.004 (2010).

36 Ciau-Uitz, A., Wang, L., Patient, R. & Liu, F. ETS transcription factors in hematopoietic stem cell development. *Blood Cells Mol Dis* **51**, 248-255, doi:10.1016/j.bcmd.2013.07.010 (2013).

37 May, W. A. *et al.* Ewing sarcoma 11;22 translocation produces a chimeric transcription factor that requires the DNA-binding domain encoded by FLI1 for transformation. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 5752-5756 (1993).

38 Tomlins, S. A. *et al.* Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* **310**, 644-648, doi:10.1126/science.1117679 (2005).

39 Goldberg, L. *et al.* Genome-scale expression and transcription factor binding profiles reveal therapeutic targets in transgenic ERG myeloid leukemia. *Blood* **122**, 2694-2703, doi:10.1182/blood-2013-01-477133 (2013).

40 Marcucci, G. *et al.* Overexpression of the ETS-related gene, ERG, predicts a worse outcome in acute myeloid leukemia with normal karyotype: a Cancer and Leukemia Group B study. *J Clin Oncol* **23**, 9234-9242, doi:10.1200/JCO.2005.03.6137 (2005).

41 Gao, J., Chen, Y. H. & Peterson, L. C. GATA family transcriptional factors: emerging suspects in hematologic disorders. *Exp Hematol Oncol* **4**, 28, doi:10.1186/s40164-015-0024-z (2015).
42 Hahn, C. N. *et al.* Heritable GATA2 mutations associated with familial myelodysplastic syndrome and acute myeloid leukemia. *Nature genetics* **43**, 1012-1017, doi:10.1038/ng.913 (2011).
43 Greif, P. A. *et al.* GATA2 zinc finger 1 mutations associated with biallelic CEBPA mutations define a unique genetic entity of acute myeloid leukemia. *Blood* **120**, 395-403, doi:10.1182/blood-2012-01-403220 (2012).
44 Groschel, S. *et al.* A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia. *Cell* **157**, 369-381, doi:10.1016/j.cell.2014.02.019 (2014).
45 Miller, S. A., Mohn, S. E. & Weinmann, A. S. Jmjd3 and UTX play a demethylase-independent role in chromatin remodeling to regulate T-box family member-dependent gene expression. *Molecular cell* **40**, 594-605, doi:10.1016/j.molcel.2010.10.028 (2010).
46 Lee, S., Lee, J. W. & Lee, S. K. UTX, a histone H3-lysine 27 demethylase, acts as a critical switch to activate the cardiac developmental program. *Developmental cell* **22**, 25-37, doi:10.1016/j.devcel.2011.11.009 (2012).
47 Local, A. *et al.* Identification of H3K4me1-associated proteins at mammalian enhancers. *Nature genetics* **50**, 73-82, doi:10.1038/s41588-017-0015-6 (2018).
48 Forsberg, L. A. *et al.* Mosaic loss of chromosome Y in peripheral blood is associated with shorter survival and higher risk of cancer. *Nat Genet* **46**, 624-628, doi:10.1038/ng.2966 (2014).
49 Dumanski, J. P. *et al.* Mutagenesis. Smoking is associated with mosaic loss of chromosome Y. *Science* **347**, 81-83, doi:10.1126/science.1262092 (2015).
50 Zhou, W. *et al.* Mosaic loss of chromosome Y is associated with common variation near TCL1A. *Nat Genet* **48**, 563-568, doi:10.1038/ng.3545 (2016).
51 Kogan, S. C. *et al.* Bethesda proposals for classification of nonlymphoid hematopoietic neoplasms in mice. *Blood* **100**, 238-245 (2002).
52 Morse, H. C., 3rd *et al.* Bethesda proposals for classification of lymphoid neoplasms in mice. *Blood* **100**, 246-258 (2002).
53 Koike-Yusa, H., Li, Y., Tan, E. P., Velasco-Herrera Mdel, C. & Yusa, K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nature biotechnology* **32**, 267-273, doi:10.1038/nbt.2800 (2014).

Figure Legends

Figure 1. *Utx*$^{-/-}$ mice develop acute myeloid leukaemia

(**a**) Structure of the *Utx* conditional allele. (**b**) qRT-PCR for exons 2-3 of *Utx* confirms *Utx* loss in *Utx*$^{-/-}$ HSPCs. The mean ± s.e.m is shown; n= number of mice per genotype; *P* by two-sided *t*-test, t=10.93, df=63. (**c**) Immunoblot showing loss of UTX protein in *Utx*$^{-/-}$ BM. Results of one representative experiment are shown (n=3 experiments). Uncropped images are shown in Fig.S12 (**d**) Kaplan-Meier survival curves for female *Utx*$^{-/-}$ (median 483 days), *Utx*$^{+/-}$ (median 661 days) and *Utx*$^{+/+}$(median survival not reached) mice; n= number of mice per genotype; *P* by Log-rank (Mantel-Cox) test, df=2. (**e**) Spleen weights of *Utx*$^{-/-}$, *Utx*$^{+/-}$ and *Utx*$^{+/+}$mice; mean ± s.e.m. is shown; n= number of mice per genotype; *P* value was determined by one-way ANOVA with Bonferroni correction, t=2.554, df=55. (**f**) Characteristic Mac1/Gr1 fluorescence activated cell sorting (FACS) analysis of BM and spleen cells from one diseased *Utx*$^{-/-}$ mice (similar result was observed in n=12). (**g**) Histopathological diagnoses of moribund mice of indicated genotypes. The number of mice with a cancer diagnosis and the total analysed is indicated for each genotype. B-ALL: B-cell ALL; MPN: myeloproliferative neoplasm; other: non-hematological tumors; other (n/s): non-specified. (**h**) Characteristic histology from one mouse with AML is shown (similar results were observed in n=20 mice). Sp=spleen, Li=liver. (**i**) Kaplan-Meier survival curves of mice transplanted with *Utx*$^{-/-}$ leukemias: AL: acute leukemia unclassified (n=5), T-ALL: T cell acute lymphoblastic leukemia (n=5), and two AML (n=9).

Figure 2. *Utx* loss expands hematopoietic stem/progenitor cells and imparts a myeloid bias, characteristics rescued by *Uty*

(**a**) Pre-leukaemic spleens, (**b**) spleen weights (t=10.93, df=63) (**c**) Lin- cell (HSPC)(t=6.908, df=9) and (**d**) LT-HSC (t=3.712, df=9) and ST-HSC frequency in BM from *Utx*$^{+/+}$ and *Utx*$^{-/-}$ (t=4.049, df=9) (**e**) Representative FACS-profiles of CMP, GMP and MEP (similar to other 5 mice). (**f**) Quantitation of LK ($Lin^-Sca1^-c\text{-}Kit^+$)(t=2.927, df=14), CMP(t=3.518, df=14), GMP (t=3.608, df=14), and MEP (t=6.181, df=15) from *Utx*$^{+/+}$ and *Utx*$^{-/-}$ mice. (**g**) CLP frequency in BM from *Utx*$^{+/+}$ and *Utx*$^{-/-}$ mice (t=3.534, df=5). (**h**) Serial re-plating of BM-derived colonies from *Utx*$^{+/+}$, *Utx*$^{+/-}$, *Utx*$^{-/-}$ (for *Utx*$^{+/+}$ vs *Utx*$^{-/-}$ in plating: 1, t=7.164; df=19; 2, t=3.991, df=19; 4, t=5.489, df=11; 5, t=3.292, df=11) (**i**) Schematic summary of progenitor differentiation in *Utx*$^{-/-}$. Green arrows= preferential differentiation; red lines= differentiation block. (**j**) Kaplan-Meier survival curves of *Utx*$^{-/Y}$ and *Utx*$^{+/Y}$ compared to *Utx*$^{-/-}$; *P*=ns between *Utx*$^{-/Y}$ and *Utx*$^{+/Y}$ by Log-rank (Mantel-Cox) test, df=1 (**k**) Histopathological diagnoses of moribund mice. Numbers of mice with a cancer diagnosis and total analysed indicated for each genotype. (**l**) Schematic summary of progenitor differentiation in *Utx*$^{-/Y}$. (**m**) serial re-plating and (**n**) proliferation of *Cas9*-expressing HSPCs upon *Uty* editing; for m and n, cells were isolated from n=3

mice per genotype; the mean ± s.e.m is shown; *P* by one-way ANOVA with Bonferroni correction (*P* shown for *Cas9*, *Utx*$^{-/Y}$ gRNA-*Uty* versus *Cas9*, *Utx*$^{+/Y}$ gRNA-*Uty*, in m for plating: 3 (t=4.313, df=8), 4(t=5.522,df=8); in n for culture days: 4 (t=3.176, df=8); 6, (t=3.994, df=8); 7, (t=4.537, df=7); . In c, d, f, g, the mean ± s.e.m is shown; n= number of mice; *P* by two-sided t-test. In b and h, mean ± s.e.m is shown; *P* by one-way ANOVA with Bonferroni correction.

Figure 3. H3K27 demethylase activity is not required for tumor suppression by UTX

(**a**) Experimental approach for study of *UTX*-mutant-MONO-MAC6 after expression of *FLAG*, *UTX*, *UTY* and *UTX-MT2*. (**b**) Compared to FLAG; UTX, UTY and UTX-MT2 reduced MONO-MAC6 proliferation; the mean ± s.e.m is shown; n=independent cultures; *P* by one-way ANOVA with Bonferroni correction. Day3 (compared to FLAG): UTX (t=5,633; df=8), UTY (t=3,95; df=8); day4: UTX (t=6.866; df=8), UTY(t=5.444; df=8); day5: UTX(t=4.976; df=8), UTY(t=4.216; df=8) (**c**) Colony formation in semisolid media, upper panel - similar result observed in n=3 cultures; lower panel-colony quantification; the mean ± s.e.m is shown; n=3 independent cultures; *P* by one-way ANOVA with Bonferroni correction; compared to FLAG: for UTX t=10.19, df=8; UTX-MT2 t=10.36, df=8; UTY t=7.955; df=8 (**d-e**) *in vivo* growth after transplantation into immunocompromised mice. Cells were also transduced with luciferase-expressing vector and mice imaged on days 6, 20 and 27 after transplant. The mean ± s.e.m is shown; *P* by one-way ANOVA with Bonferroni correction: ** *P* (FLAG vs UTX)=0.0131, t=3.317, df=16; *P* (FLAG vs UTY)= 0.0242, t=3.025, df=16; *P*(FLAG vs UTX-MT2)=0.0095, t=3.469, df=16; * *P*(FLAG vs UTX) =0.0013, t=4.408, df=16; *P*(FLAG vs UTY) = 0.1284, t=2.2, df=16; for *P*(FLAG vs UTX-MT2)=0.0010, t=4.524; df=16. (**f**) Kaplan-Meier survival curves of transplanted mice; n=number of mice, *P* by Log-rank (Mantel-Cox) test reported versus FLAG, df=1 (**g**) Immunoblot analysis of UTX in AML lines; similar result observed in n=3 experiments. Uncropped images are shown in Fig.S12 (**h**) qRT-PCR for *UTY* in AML and (**i**) in non-hematopoietic cancer cell lines with *UTX* mutations, for h and i, the mean ± s.e.m is shown; n=3 independent cell cultures.

Figure 4. *Utx* loss drives both up and down regulation of gene expression primarily through effects on H3K27 acetylation

(**a**) Pre-leukaemic gene expression changes in HSPCs from *Utx*$^{-/-}$ female (n=2 mice) and *Utx*$^{-/Y}$ male (n=2 mice) compared to sex-matched wild-type controls (n=2 mice); genes with adj. *P*<0.05 are shown. Subtraction of genes differentially expressed in males from those in females defines a differential transcriptional program of interest; $\log_2$FC (-0.5>$\log_2$FC>0.5)(**b**) Volcano plot of fold-change (-0.5>$\log_2$FC>0.5) and adj. *P* <0.05 (only transcripts with *P* values between 0.05 and 1e^{-38} are shown) for genes differentially expressed in the pre-leukaemic and AML setting. In a and b the *P* was

generated using a negative binominal generalized linear model (DESeq2). (**c**) Overlap between differentially expressed genes in *Utx*$^{-/-}$ pre-leukaemic HSPCs (n=2 mice) and AMLs (n=3 mice), each compared to *Utx*$^{+/+}$ HSPC (n=2 mice); *P* by hypergeometric test. (**d**) Distribution of UTX ChIP-Seq peaks in annotated regions of the genome. (**e**) Highly significant enrichment of UTX-bound genes amongst those differentially expressed in pre-leukaemic *Utx*$^{-/-}$ HSPCs; *P* by hypergeometric test. (**f**) H3K27Me3 (**g**) H3K27Ac and (**h**) H3K4Me1 density plots (left) and average read counts (right) across all (global changes) or differentially modified regions (local changes). H3K27Me3 signal density show that only 302 genomic regions were differential modified, in contrast to similar plots for H3K27Ac reveals 5120 and for H3K4Me1 4552 differential modifications in *Utx*$^{-/-}$ vs *Utx*$^{+/+}$ HSPC. Arrows show each replicate (mouse) per genotype. FDR was calculated using DiffBind tool; n=2 mice per genotype. Plots are peak centered, scaled and ± 1kb for each locus. Shaded region in the line graphs in f-h indicate the standard errors.

Figure 5. *Utx* loss activates an oncogenic ETS transcriptional program while suppressing a GATA program

Volcano plots of differentially expressed genes in (**a**) pre-leukaemic (n=2 mice) and (**b**) AML *Utx*$^{-/-}$ (n=3 mice) compared to wild-type controls (n=2 mice) reveal overexpression of multiple ETS factors (red dots). Fold-change (-0.5>$\log_2$FC>0.5) and adj. *P* <0.05 (only transcripts with *P* values between 0.05 and 1e^{-38} are shown in the graph); *P* was generated using DESeq2. (**c**) GSEA plot shows significant overlap with a known ETS oncogenic program driven by the EWSR1-FLI1 fusion. (**d**) Motif analysis of UTX ChIP-Seq peaks that overlap with overexpressed genes; number indicates motif rank (**e**) MONO-MAC6 proliferation upon editing of indicated gene. BFP-positive fraction was compared with the non-transduced population and normalized to day 4 (d4) for each gRNA. The mean ± s.d. is shown; n=independent cell cultures; *P* by one-way ANOVA with Bonferroni correction; *P* shown for day19 compared to control gRNA (EMPTY). Compare to EMPTY for: *ELF4* t=32.32; ETV6 t=10.03; FLI1 t=41.1; ETS1 t=15.85; PU.1 t=33.56; ELF1 t=3.967; ERG=12.35 and df=16. (**f**) GSEA plot showing enrichment of genes differentially expressed in *Utx*$^{-/-}$ HSPCs with a published dataset of GATA2 targets. (**g**) Motif analysis of 3442 downregulated H3K27Ac peaks (FDR<1%, FC<-1.5) identified in *Utx*$^{-/-}$ HSPCs, number indicates motif rank. (**h**) Selected proteins identified by mass spectrometry after immunoprecipitation of endogenous UTX from murine myeloid cells (416B) (n=2 independent cell cultures). Motif and statistical analysis in d and g was determined by HOMER software (see Methods and TableS30).

Figure 6. UTX interacts with chromatin modifiers to maintain chromatin accessibility

(**a**) Motif analysis of ATAC-Seq closed and (**b**) open peaks reveals dramatic enrichment for GATA motifs in the former and ETS, amongst other motifs, in the latter; number indicates motif rank. Motif and statistical analysis was determined by HOMER software (TableS30) (**c**) Genomic snapshot of GATA2, UTX, ATAC-Seq and H3K27Ac ChIP-Seq in *Utx*$^{+/+}$ and *Utx*$^{-/-}$ HSPCs at the *Ets2* and (**d**) *Steap3* loci. Note co-localization of GATA2 binding with dynamically closed chromatin and loss of H3K27Ac following UTX loss, without evidence for GATA2-UTX co-binding. By contrast, binding of the chromatin remodelers BRG1 and CHD4 directly co-localize with UTX binding (lower two tracks). At the *Ets2* locus, newly accessible chromatin is also seen following UTX loss, again at regions not directly bound by UTX or chromatin remodelers. (**e**) Density plots of UTX, BRG1 and CHD4 ChIP-Seq on UTX-bound genomic loci; Venn diagram shows overlap between all UTX, BRG1 and CHD4 ChIP-Seq peaks; *P* by Fisher's exact test for ChIP-Seq: UTX vs BRG1/CHD4 (**f**) Schematic representation of PU.1 occupancy occurring mostly on closed chromatin. (**g**) Overlap of genes associated with enhanced PU.1 binding (in *Utx*$^{-/-}$) on closed chromatin with gene expression changes from pre-leukemia (PL) to AML. For PU.1 ChIP-Seq n=3 mice; ATAC-Seq n=3 mice; PL RNA-Seq n=2 mice; AML RNA-Seq n=3 mice. *P* by hypergeometric test. (**h**) Genomic snapshot demonstrating enhanced PU.1 occupancy in *Utx*$^{-/-}$ HSPCs that occurs on closed chromatin at the *Rab11a* locus and correlation with PL and AML RNA-Seq. Note *Rab11a* expression increases only upon progression to AML.

Methods

Mice

The *In vivo* experiments were performed under the project licence PPL 80/2564 issued by the United Kingdom Home Office, in accordance with the Animal Scientific Procedures Act 1986. The *Utx* mouse model, C57Bl6, was developed at the Sanger Institute. $Utx^{f/f}$ mice were crossed with Flpe mice and then with *Mx1-Cre* mice. Cre expression was induced by intraperitoneal injection of 5-6 -week -old mice with pIpC (Sigma #P1530, 400 µg/mouse; 5 doses over a period of 10 days). All pre-leukaemic experiments were performed 4-6 weeks post pIpC injection. *Cas9*-expressing mice were reported previously[28].

Cell lines

293FT (Invitrogen) were cultured in DMEM (Invitrogen), supplemented with 10% FBS. 416B cells were cultured in RPMI1640 (Invitrogen), 10% FBS. SN12C, KU-19-19, KYSE-180 and HCC2998 were cultured in RPMI (Invitrogen) supplemented with 10% FBS (Invitrogen); J82, UM-UC-3 and FADU were cultured in EMEM (Invitrogen), 10% FBS; CAL-27 and VM-CUB-1 were cultured in DMEM (Invitrogen), 10% FBS. SW684 was cultured in L15, 10% FBS. LB996-RCC was cultured in IMDM (Invitrogen), 10% FBS (Invitrogen). D-423MG was cultured in Gibco Zinc Option (Invitrogen), 10% FBS; KYSE-270 in RPMI & Ham's F12, 2% FBS. Each of the media was supplemented with 1% penicillin/streptomycin/glutamine (PSG, Invitrogen). AML cell lines: MV4-11, MONO-MAC6 and THP1 were cultured in RPMI1640 (Invitrogen) supplemented with 10% FBS and 1% PSG. OCI-AML2 and OCI-AML3 were cultured in alpha-MEM (Lonza), 20% FBS and 1% PSG. All cancer cell lines were obtained from the Sanger Institute Cancer Cell Collection.

cDNA synthesis, PCR and qRT-PCR

cDNA synthesis was performed using a qScript cDNA SuperMix (Quanta Biosciences) according to the manufacturer's instructions. PCR was performed with REDTaq ReadyMix PCR Reaction Mix (Sigma) according to the manufacturer's instructions. qRT-PCR was performed using TaqMan Fast Universal PCR Master Mix (Thermo Fisher Scientific) and the Universal Probe Library system (Roche). The *TBP* housekeeping gene was used for data normalization. qRT-PCR primer sequences are presented in (**TableS27**).

Protein extraction, Immunoblot, co-immunoprecipitation

The cells were lysed in whole cell lysis buffer (50 mM Tris-HCl pH=8, 150 mM NaCl, 0.1% NP-40, 1 mM EDTA), supplemented with 1 mM DTT, protease inhibitors (Sigma), and phosphatase inhibitors

(Sigma). Protein concentrations were assessed by Bradford assay (Bio-Rad) and an equal amount of protein was loaded per track. Prior to loading, the samples were supplemented with SDS-PAGE sample buffer and DTT was added to each sample. 10-40 µg of protein was separated on a 10% SDS-PAGE gel, and blotted onto polyvinylidene difluoride membranes (Millipore). The following primary antibodies were used: anti-UTX (Bethyl, A302-374A), anti-UTX (GeneTex, GTX121246) and anti-IgG (Santa Cruz Biotechnology, sc-2027) and anti-α-tubulin (Sigma, T6074) or anti-ACTIN (Santa Cruz Biotechnology, sc-1616) a loading control. The secondary antibodies used in the study were as follows: HRP-linked donkey anti-rabbit (GE Healthcare UK); ECL HRP-linked anti-mouse (Santa Cruz Biotechnology, sc-2005). Visualization was performed via LumiGLO Chemiluminescent Substrate (KPL, 54-61-00). Co-immunoprecipitation (co-IP) was performed in the extracts isolated with cell lysis buffer (as above). 2-6 ug of antibody were bound to 20ul of Dynabeads Protein G (Thermo Fisher Scientific) beads and incubated with 500-1000 µg of freshly extracted protein for 1,5h at 4°C with rotation. IP was washed 4 times with IP wash buffer (10 mM Tris-HCl pH=8, 150 mM NaCl, 0.1% NP-40, 1 mM EDTA), supplemented with protease inhibitor (Sigma). IP samples with beads were then resuspended in 1x NuPAGE LDS Sample Buffer (Thermo Fisher Scientific), supplemented with NuPAGE Sample Reducing Agent (Thermo Fisher Scientific). The following antibodies were used for IP; anti-CHD4 (Abcam, ab72418), anti-BRG1 (Santa Cruz Biotechnology, sc-10768) anti-UTX (Bethyl, A302-374A) and IgG (Santa Cruz Biotechnology, sc-2027).

Histological analysis of mouse tissue

The tissues were fixed in 10% formaldehyde and subsequently paraffin embedded. Bones were decalcified using 0.38 M EDTA pH=7. Tissue sections (4 µm) were stained with Hematoxylin and Eosin (Thermo Fisher Scientific). Histology assessment was performed using the Bethesda criteria for mouse hematological tumors[51,52].

Blood count analysis

Blood count measurement was performed on a VetabC analyzer (Horiba ABX).

Isolation of mouse hematopoietic progenitors

Freshly isolated bone marrow was suspended in erythrocyte lysis solution (BD PharmLyse, BD Bioscience), followed by magnetic bead selection of Lin- cells, using the Lineage Cell Depletion Kit (Miltenyi Biotec, cat. no. 130-090-858) according to the manufacturer's instructions. c-KIT$^+$ progenitors were selected with mouse CD117 MicroBeads, (Miltenyi Biotec, #130-091-224) according to the manufacturer's instructions.

Culture of mouse hematopoietic progenitors

Primary mouse cells were cultured in X-VIVO 20 media (Lonza) supplemented with 5% serum (Stem Cell Technologies), 10ng/ml IL3 (Peprotech), 10ng/ml IL6 (Peprotech) 50ng ml/ml of SCF (Peprotech) and 1% penicillin-streptomycin-glutamine (Gibco).

FACS analysis

Bone marrow cells were incubated in erythrocyte lysis buffer: 0.85% NH_4Cl (Sigma) and blocked with anti-mouse CD16/32 (BD Pharmigen, #553142) and 10% mouse serum (Sigma M5905) for LSK and CLP staining or 10% mouse serum alone for LK staining. LSK, MPP, LMPP, LT/ST-HSC FACS staining was performed using the following antibody against: CD4 (Biolegend, #100514), CD5 (Biolegend, #100610), CD8a (Biolegend, # 100710), CD11b (Biolegend, #101210), B220 (Biolegend, #103210), TER-119 (Biolegend, #116210), GR-1 (Biolegend, #108410) as well as SCA-1 (Biolegend, #122520), CD117 (eBioscience, # 47-1171); CD48 (Biolegend, #103411), CD150 (Biolegend, #115913), CD34 (BD Pharmigen, #553733), FLT3 (eBioscienc, #12-1351). GMP, MEP, CMP staining was performed with the following biotin-conjugated lineage markers: MAC1, GR1, CD3, B220, TER119 (BD Pharmigen, #559971), and IL7Ra (Biolegend, #121103), Streptavidin (Biolegend, #405206) alongside with CD34 (BD Pharmigen, #553733), CD16/32 (BD Pharmigen, #553145), c-KIT (BioLegend, #105812), SCA-1 (Biolegend, #122520). For the detection of the CLP population cells were stained with the Flt3 (eBioscienc, #12-1351), IL7Ra (BioLegend, #135008), lineage-biotin conjugated markers: MAC1, GR1, CD3, B220, TER119 (BD Pharmigen, #559971), as well as NK (LSBio, LS-C62548) c-KIT (BioLegend, #105812) and SCA-1 (Biolegend, #122520). Differentiated bone marrow, spleen and peripheral blood cells were stained with CD45 (BD Pharmigen, #563891), CD11b (BD Pharmigen, #557657), B220 (Biolegend, #103210), GR1 (BD Pharmigen, #560603), c-KIT (BioLegend, #105812), TER119 (BD Pharmigen, #557915) and CD3e (eBioscience, #12-0031-82). LT-HSC: long term hematopoietic stem cells were defined as: Lin^-, c-KIT^+, $SCA1^+$, $FLT3^-$, $CD48^-$, $CD150^+$, $CD34^-$; ST-HSC: short term HSC as Lin^-, c-KIT^+, $SCA1^+$, $FLT3^-$, $CD48^-$, $CD150^+$, $CD34^+$; MPP: multipotent progenitors (Lin^-, c-KIT^+, $SCA1^+$, $FLT3^+$); LMPP: lymphoid primed multipotent progenitors (Lin^-, c-KIT^+, $SCA1^+$, $FLT3^{hi}$), CLP: common lymphoid progenitors (Lin^-, $FLT3^{hi}$, $IL7Ra^+$, c-KIT^{lo}, $SCA\text{-}1^{lo}$), GMP: granulocyte-monocyte progenitors (Lin^-, $IL7Ra^-$, c-KIT^+, $SCA1^-$, $CD34^+$, $CD16/32^+$); CMP: common myeloid progenitors (Lin^-, $IL7Ra^-$, c-KIT^+, $SCA1^-$, $CD34^+$, $CD16/32^-$), MEP: megakaryocyte-erythroid progenitors (Lin^-, $IL7Ra^-$, c-KIT^+, $SCA1^-$, $CD34^-$, $CD16/32^-$). Flow cytometry analysis was performed using the LSRFortessa instrument (BD) and analysed using FlowJo software.

Serial re-plating assay

For re-plating assays 50,000 bone marrow cells were plated in two wells of 6-well-plate of M3434 methylcellulose (Stem Cell Technologies). The colonies were counted 7 days later and further 30,000 cells re-seeded and re-counted after a week until no colonies were observed.

Proliferation assay

10^4 cells /well were plated onto 96-well plates and assayed daily for growth using CellTiter 96 AQueous Non-Radioactive Cell Proliferation Assay (Promega) according to manufactures' instructions.

Plasmids, cloning

FLAG-tagged versions of UTX, UTY, UTX-MT2 plasmids were purchased from Addgene (pCS2-UTX-F #24168, pCS2-UTY-F #17439 and pCS2-UTX-F-MT2 #40619). The lentiviral UTX, UTY, UTX-MT2 and FLAG expression vectors were constructed in pKLV-puro as follows. Firstly, the lentiviral backbone vector, pKLV-U6(Flip)gRNA(BbsI)-PGKpuro2ABFP[53] was digested with BbsI and KpnI to remove the U6gRNA(BbsI)-PGKpuro2ABFP cassette. Gibson cloning (NEB) was performed to clone the PCR products for EF1α promoter, UTX/UTY/UTX-MT2/FLAG cDNA and puromycin (Puro) resistance gene with the primers in (**TableS28)**. EF1α promoter was PCR amplified from pLVX-EF1α-IRES-ZsGreen1 construct (Clontech, 631982). Puro resistance gene was amplified from pKLV-U6(Flip)gRNA(BbsI)-PGKpuro2ABFP construct. Gibson cloning was performed according to manufacturer's specification. Firefly luciferase expressing plasmid (EF1α-GFP-T2A-Luciferase) was obtained from System Biosciences (BLIV503-MC-1-SBI). AML-ETO9a plasmid was published before[54]. For CRISPR/Cas9 experiments gRNA were cloned into BbsI digested pKLV2-U6gRNA(BbsI)PGKpuro2ABFP backbone[28]. Sequences of gRNA used in the study are provided in **TableS.29**.

Lentiviral vector production and transduction

Lentiviruses were produced in HEK293 cells using ViraPower Lentiviral Expression System (Invitrogen) according to manufacturer's instructions. Viral supernatant was concentrated by centrifugation at 6000g, 16h, at 4 °C. The cells were transduced by spinoculation (60 min, 800 g, 32 °C) in culture medium supplemented with 4μg/ml of polybrene (Millipore) and further incubated overnight at 37 °C. The following day, the transduced cells underwent selection on puromycin, (1.5ug/ml, Sigma) for three days.

Transplant; i*n vivo* imaging and quantification

MONO-MAC6 cell line was transduced with lentiviral vectors expressing UTX, UTY, UTX-MT2 or FLAG control, puromycin-selected for 4 days. $0.8x10^6$ cells were transplanted via i.v. tail injection into

immunocompromised recipient mouse (Il2rg$^{-/-}$; Rag2$^{-/-}$). Five mice were injected per group. For bioluminescence examination mice were injected with D-luciferin (BioVision; 3mg/20g, i.p.) and subsequently anesthetized with isoflurane. Bioluminescence was quantified with an In Vivo Imaging System IVIS Lumina II (Caliper), with Living Image version 4.3.1 software (PerkinElmer) according to manufactures' instructions.

AML-ETO9a transplants: c-KIT positive cells from the BM of *Utx*$^{+/+}$ and *Utx*$^{-/-}$ mice and transduced them with lentiviral vectors expressing the AML-ETO9a fusion. 1x10^6 cells were subsequently transplanted into lethally irradiated syngeneic recipient mice (n=9-10/group).

Secondary transplant of mouse *Utx*$^{-/-}$ leukemias: 1x10^6 splenocytes were injected into sublethally irradiated recipient mice: acute leukaemia unclassified (n=5 mice), T-ALL (n=5 mice), and two AML (n=9 mice).

RNA extraction, RNA-Seq analysis

RNA was extracted from HSPC/Lin$^-$ BM cells using Arcturus Picopure RNA Isolation Kit (Thermo Fisher Scientific) according to the manufacturer's instructions. RNA from murine primary AML samples was extracted with Trizol reagent (Thermo Fisher Scientific). RNA from human cell lines was extracted with RNeasy kit (Qiagen) according to the manufacturer's instructions. RNA-Seq library was generated using TruSeq Stranded mRNA Sample Prep Kit (Illumina) and sequenced on Illumina HiSeq2000 v4 chemistry, 75-bp paired-end sequencing. RNA-Seq reads for each library were mapped using TopHat version 2.0.13 against the mouse genome build GRCm38.68, downloaded via Ensembl BioMart data mining tool. HiSeq libraries were aligned with the following options: "--keep-fasta-order --no-sort-bam -r 100 -p 12 --library-type fr-firststrand --no-coverage-search --microexon-search --transcriptome-index=GRCm38.known" where GRCm38.known is a transcriptome index file in GTF format prepared beforehand by an initial single TopHat run without input reads. Raw counts for each gene in the Genome Reference Consortium genome assembly (GRCm38.68) were obtained with Bioconductor package GenomicAlignments version 1.2.2 with mode="Union". Differential expression analysis was carried out with these counts using Bioconductor package DESeq2 version 1.6.3 with BH independent filtering method with an FDR of 1%; both packages were used according to their vignettes. Fold changes in expression were also calculated by DESeq2 after correcting for differences in library sizes.

ChIP-Seq, ChIP-qPCR

ChIP-Seq experiments were performed on primary HSPC cells. For histone ChIP-Seq cells were fixed in 1% formaldehyde (FA, Thermo Fisher Scientific, #28906) for 5 mins at room temperature. The

reaction was stopped by the addition of glycine (0.125M, Sigma) and the cells were washed in ice-cold PBS. The cells were then processed with iDeal ChIP-Seq kit for Histones (Diagenode) with following antibody anti: H3K4Me1 (Diagenode, #pAb-194-050), H3K27Me3 (Abcam, #Ab6002), H3K27Ac (Diagenode, #pAb-196-050). For PU.1 ChIP-Seq cell were fixed with 1% FA for 10 mins at room temperature and processed with iDeal ChIP-Seq kit for Transcription Factors (Diagenode) according to manufacturer's instructions, with anti-PU.1 antibody (Santa Cruz Biotechnology, sc-352x). ChIP-Seq for UTX was performed in HSPCs. The cells were crosslinked with 2 mM Disuccinimidyl Glutarate (DSG, Sigma) for 30 mins at RT followed by the second crosslinking with 1% FA for 30 mins at 4°C. The cells were then processed with iDeal ChIP-seq kit for Transcription Factors, with anti-UTX antibody (Bethyl, A302-374A). Primer sequences used for ChIP-qPCR are listed in (**TableS27**).

For BRG1, CHD4, UTX ChIP-qPCR 416B cells were crosslinked with 2 mM Disuccinimidyl Glutarate (DSG, Sigma) for 30 mins at RT followed by the second crosslinking with 1% FA for 10 mins at RT. Crosslinking was stopped by addition of 125 mM glycine. Cells were resuspended in ChIP Lysis Buffer (1%SDS, 10mM EDTA, 50mM Tris-HCl pH=8, protease inhibitors) and sonicated in Bioruptor Pico (Diagenode) for 10 cycles. Sonicated chromatin was diluted 1:10 in modified RIPA buffer (1% Triton; 0.1% deoxycholate; 90mM NaCl; 10mM Tris-HCl pH8; EDTA free protease inhibitors) and incubated overnight with 3ug of anti-BRG1 (Santa Cruz Biotechnology, sc-10768X) or anti-CHD4 (Abcam, ab72418) antibody. Next protein A/G (50% A 50% G) Dynabeads (Invitrogen) were added to the chromatin and incubated 2h at 4°C followed by magnetic separation. Beads were subsequently washed twice with mixed micelle buffer (150 mM NaCl, 0.2% SDS, 20 mM Tris-Cl pH 8.0, 5 mM EDTA, 5.2% sucrose, 1% Triton X-100); high salt buffer (250 mM NaCl, 5 mM Tris-Cl pH 8.0, 0.5 mM EDTA, 0.05% sodium deoxycholate, 25 mM HEPES pH 8.0, 0.5% Triton X-100) and LiCl buffer (250 mM LiCl, 10 mM Tris-Cl pH 8.0, 10 mM EDTA, 0.5% NP40-Nonidet, 0.5% Sodium Deoxycholate) and once with elution buffer (1%SDS; 100mM NaHCO3). Beads were then resuspended in elution buffer supplemented with DNAse free RNAse (Roche, #11119915001). Cross-linking was reverted by the incubation at 37˚C for 30min followed by the incubation at 65˚C overnight. DNA was purified using the Chip-DNA purification kit (Zymo).

Chip-Seq, motif analysis and GSEA, data visualization

Adapter sequences were trimmed for all paired end reads and mapped against mm10 reference genome using Bowtie2[55]. All the samples were processed independently and uniquely mapped reads were retained. Peaks were called using SICER[56] with W200 and G600 for broad peaks and W200 and G200 for narrow peaks. Peak calling were perform for each of the replicate individually. Motif

analysis and peaks were annotated using HOMER[35]. Detail output of HOMER Motif analysis is included in **TableS30**. Peaks in intergenic regions were assigned to genes if they were within the 100kb window from the TSS. Differential binding analysis was performed using DiffBind[57] by groping replicates together. Overlapping peak analysis was performed using intersected from bedtools[58]. The statistical analysis of the overlapping peaks was performed using fisher's exact test from bedtools. Each of the ChIP-Seq experiments, with exception of the PU.1 ChIP-Seq, was performed in biological duplicates. PU.1 ChIP-Seq was performed in biological triplicates. Gene set enrichment analysis tools were obtained from The Broad Institute[59]. ChIP-Seq; RNA-Seq and ATAC-Seq data were visualized in UCSC Genome Browser[60]. Venn diagrams were performed using BioVenn web application[61]. All graphs were perform in GraphPad Prism unless specify otherwise.

Promoter-enhancer interaction analysis

Promoter-associated interaction matrix of multipotent hematopoietic progenitor cell line 7 (HPC-7) was generated in a previous study using Promoter Capture Hi-C (PCHi-C) method and data were analysed with CHiCAGO package[23]. Genome coordinates from the interaction matrix were converted from mm9 to mm10. In total, 54,339 regions that form significant interactions (CHiCAGO score ≥ 5) with promoter baits were defined as promoter-interacting regions (PIRs). Differential H3K27Ac peaks in *Utx*$^{-/-}$ versus *Utx*$^{+/+}$ were grouped as increased (UP) or decreased (DOWN) peaks and intersected with PIRs using bedtools. Interactions at specific gene loci were visualized in WashU Epigenome Browser.

ATAC

The ATAC-Seq method was used based on the established protocol[62] with modifications[63]. Briefly, 200,000 cells were washed in 0.3 mL of ice-cold Dulbecco's phosphate buffered saline without calcium and magnesium. This was followed by centrifugation at 300g for 3 minutes before resuspending in 400 µL of freshly-made ice-cold sucrose buffer (10 mM Tris-Cl pH 7.5, 3 mM $CaCl_2$, 2 mM $MgCl_2$ and 0.32 M sucrose) and incubated on ice for 12 minutes. 10% Triton X-100 was added to a final concentration of 0.5 % and the cells were vortexed briefly before incubating on ice for a further 6 minutes to access nuclei. The nuclei were briefly vortexed again before another centrifugation at 300g for 3 minutes at 4 °C. The sucrose/triton lysis buffer was removed before immediately resuspending the nuclei pellet in 50 µL of Nextera tagmentation master mix, comprising 25 µL 2x Tagment DNA buffer, 20 µL nuclease-free water and 5 µL Tagment DNA Enzyme 1 (Illumina FC-121-1030). The tagmentation reaction mixture was immediately transferred to a 1.5 mL low-bind microfuge tube and incubated at 37 °C for 30 minutes. The tagmentation reaction was stopped by

the addition of 500 µL Buffer PB (Qiagen). The tagmented chromatin was purified using the MinElute PCR purification kit (Qiagen 28004), according to the manufacturer's instructions, eluting in 10 µL of buffer EB (Qiagen). 10 µL of the tagmented chromatin was mixed with 2.5 µL Nextera PCR primer cocktail and 7.5 µL Nextera PCR mastermix (Illumina FC-121-1030) in a 0.2 mL low-bind PCR tube. 2.5 µL of an i5 primer and 2.5 µL of an i7 primer (Illumina FC-121-1011) were added per PCR, totaling 25 µL. PCR amplification was performed as follows: 72 °C for 3 minutes and 98 °C for 30 seconds, followed by 12 cycles of 98 °C for 10 seconds, 63 °C for 30 seconds and 72 °C for 3 minutes. Libraries were size-selected on a 1 % agarose TAE gel, collecting library fragments from 120 bp to 1 kb. Gel slices were extracted with the MinElute Gel Extraction kit (Qiagen 28604), eluting in 20 µL of Elution Buffer. Samples were further purified using Agencourt AMPure XP magnetic beads (Beckman Coulter A63880) at a ratio of 1.2 AMPure beads :1 PCR sample (v/v), according the manufacturer's instructions, eluting in 20 µL of Buffer EB (Qiagen). Before sequencing, each ATAC-seq library was assessed on an Agilent 2100 Bioanalyzer using a High Sensitivity DNA chip (Agilent Technologies 5067-4626).

ATAC Sequencing analysis

Similar to ChIP-Seq analyses all the adapter sequenced of ATAC-Seq paired end reads were trimmed and mapped against mm10 reference genome using Bowtie2. All the samples were processed independently and uniquely mapped reads were retained. Peaks were called with MACS2[64] with –nomodel and –nolambda parameters. Peak calling was performed individually for each of the replicate. Differential binding analysis was performed using DiffBind[57] by groping replicates together. Overlapping peak analysis were performed using intersected from bedtools[58]. The statistical analysis of the overlapping peaks were performed using fisher's exact test from bedtools. ATAC-Seq in HSPC was performed in biological triplicates, ATAC-Seq experiments for remodelers were performed in biological duplicates.

GEO accession codes for publicly available data sets

GATA2, GSM552234; CHD4, GSM1296403 and GSM1296404; BRG1, GSM1296402.

Preparation of IP samples for MS

10^7 416B cells were lysed in the whole cell lysis buffer (50 mM Tris-HCl pH=8, 150 mM NaCl, 0.1% NP-40, 1 mM EDTA), supplemented with 1 mM DTT, protease inhibitors (Sigma), and phosphatase inhibitors (Sigma). Cell were homogenized and lysate cleared by centrifugation. UTX immunoprecipitation was performed in the whole cell lysis buffer with 16 ug of antibody bound to

100ul of Dynabeads Protein G (Thermo Fisher Scientific) and incubated for 1,5h at 4°C with rotation. IP was washed five times with IP wash buffer (10 mM Tris-HCl pH=8, 150 mM NaCl, 0.1% NP-40, 1 mM EDTA), supplemented with protease inhibitor (Sigma). UTX immunoprecipitates were eluted by boiling in 1x LDS loading buffer, reduced with 5 mM TCEP, alkylated with 10 mM iodoacetamide and electrophoresed in Novex NuPAGE Bis-Tris 4%–12% gels (Life Technologies). Gels were stained with colloidal Coomassie (Sigma). Whole lanes were cut in slices and samples processed for MS analysis as described previously[65].

Sample preparation and LC-MS/MS analysis

The affinity purified material was electrophoresed in 4-12% Bis-Tris NuPAGE gels (Life Technologies). Gels were fixed and stained with Coomassie as previously described[65]. The whole gel lanes were excised into five sections and IgG bands were discarded. Gel pieces were digested with trypsin and peptides extracted as previously described[65]. The peptides were re-dissolved in 0.5% formic acid and analysed with on-line nano liquid chromatography (Ultimate 3000 RSLCnano System) tandem mass spectrometry on an LTQ Orbitrap Velos mass spectrometer. The sample was desalted on a PepMap C18 nano-trap (100 µm i.d. x 20 mm, 100Å, 5µm), then separated on a PepMap RSLC C18 column (75 µm i.d. x 250 mm, 100 Å, 2 µm) in a linear gradient of 4-32% CH_3CN/0.1% formic acid in 90 mins. The HPLC, columns and mass spectrometer were all from Thermo Fisher Scientific. The Orbitrap mass spectrometer was operated in the standard "top 15" data-dependent acquisition mode while the preview mode was disabled. The MS full scan was set at m/z 380 – 1600 with the resolution at 30,000 at m/z 400 and AGC at $1x10^6$ with a maximum injection time at 200 msec. The siloxane ion at 445.120030 was used as lock mass. The 15 most abundant multiply-charged precursor ions ($z \geq 2$), with a minimal signal above 3000 counts, were dynamically selected for CID (Collision Induced Dissociation) fragmentation in the ion trap, which had the AGC set at 5000 with the maximum injection time at 100 msec. The precursor isolation width was set at 2 Da. The normalized collision energy for CID MS/MS was set at 35%. The dynamic exclusion duration time for the selected ions for MS/MS was set for 60 sec with ±10 ppm exclusion mass width.

MS data analysis

Raw MS files were processed with Proteome Discoverer v 1.4 (ThermoFisher Scientific). Database searches were performed using Mascot (v 2.5, Matrix Science) with the mouse Swiss-Prot database (December 2015, 16942 sequences) supplemented with an in-house contaminant database. The search parameters were as follow: trypsin digestion, 2 missed cleavages, 10 ppm mass tolerance for precursor ions, 0.5 Da mass tolerance for fragment ions, variable modifications of carbamidomethyl

(C), N-acetylation (protein), formyl (N-term), oxidation (M), deamidated (NQ), and pyro-glu (N-term Q). Database search results were further processed with Percolator[66-68] within Proteome Discoverer. Protein identification required at least one high-confidence peptide (FDR < 1% based on *q*-value) and a minimum Mascot protein score of 20. Protein lists for bait and control experiments (2 and 3 replicates respectively) were compared using SAINTexpress with default settings[69]. External contaminants were removed for further analysis. Preys with SAINT probability score ≥ 0.99 are reported in the final high confidence interactors list. The complement protein C1qc was manually removed from the list because, despite having a SAINT probability of 1, it was detected with a very similar number of peptides in both bait and control samples (**TableS15, TableS16**).

Exome sequencing

DNA from 7 *Utx*$^{-/-}$ AML cases and matched normal DNA extracted from tail tips before pIpC-mediated *Utx* deletion, was extracted using DNeasy blood & tissue kit (Qiagen) according to the manufacturer's instructions. The extracted DNA was quantified (using Invitrogen's dsDNA Quant-IT PicoGreen), followed by normalizing each sample to 4.17ng/ul in 120ul in preparation for library creation. The first step of library preparation involved shearing the DNA into fragments of 150bp (using the Covaris LC220 and Agilent Bravo automated workstation for liquid handling) followed by library creation and PCR using unique index tags and adaptors (Agilent's SureSelectXT Automated Library Prep and Capture Kits and MJ Tetrad). The amplified libraries were then purified (using Agencourt AMPure XP and a Beckman Coulter Biomek NX96 for liquid handling) and eluted in nuclease-free water followed by another round of quantification (using the Caliper GX). The quantified, size-selected libraries were then diluted down to an appropriate concentration for introduction into the exome capture stage. Exome pulldown (or hybridisation) was performed using Mouse-All Exon RNA-baits (designed by Agilent, supplier ID: S0276129) for 23 hours at 65ºC. Eight uniquely indexed samples were baited and captured in a single pool as part of the agreed multiplexing strategy. The pulldown was then purified and eluted using streptavidin-coated Dynal beads ready to be amplified using PCR (MJ Tetrad). The PCR was then further purified using Agencourt AMPure XP (and Beckman Coulter Biomek NX96 for liquid handling), followed by quantification of the amplified pulldown product using the Agilent Bioanalyser. Samples were exome sequenced as paired-end 75bp inserts using Illumina HiSeq v4 flow-cell chemistry.

Exome data analysis

Somatic variants (point mutations and indels) were called using the Caveman[70] and Pindel[71,72] pipelines respectively, and filtered for artefacts (including sufficient tumor fraction, strand bias,

presence of tumor allele in the normal, presence of tumor allele in panel of normal samples also sequenced at Sanger). Genome-wide copy number variation was called using the Control Freek[73] software package. This packages accepts paired tumor/ normal sequence files. We noticed that in a few of our samples that the sequenced paired normal displayed an artifactual noisiness in sequencing depth and opted to use a single 'quiescent' normal (MD5280a) as a constant comparator for all samples. Control Freek was run using these parameters: step=1000000, window=5000000, breakpointtype=4, breakpointthreshold=1.2, readcountthreshold=50. The copy-number plots show the normalised bam depth-ratio as produced by ControlFreek (black points), as well as a mark (red points) for regions marked with CopyNumber = 2. Copy-number variation looking for a focal change at exon 3 of *Utx* was run (as before) with an unmatched normal - MD5280a - as well as the following parameters: step=250, window=500, breakpointthreshold=0.6, breakpointtype=4, readcountthreshold=50.

Statistical analysis

All statistical analyses were performed using two-sided Student's T test or one-way ANOVA as specified in figure legends. Error bars represent the standard error of the mean (s.e.m.) or the standard deviation (s.d.). P values ≤ 0.05 were considered statistically significant. Representative data/images were replicated in at least three independent experiments as specified in the relevant figure legend. Hypergeometric distribution was calculated with the online tool, GeneProf. The number of independent experiments used to generate statistically significant data is defined in the relevant figure legends.

URLs

TopHat version 2.0.13 (http://ccb.jhu.edu/software/tophat/)

BioMart data mining tool (http://www.ensembl.org/info/data/biomart/).

GenomicAlignments version 1.2.2
(http://bioconductor.org/packages/3.0/bioc/html/GenomicAlignments.html) with mode="Union".
DESeq2 version 1.6.3 (http://bioconductor.org/packages/3.0/bioc/html/DESeq2.html)

UCSC Genome Browser (http://genome.ucsc.edu/)

BioVenn tool (http://www.biovenn.nl/)

Hypergeometric test, (https://www.geneprof.org/GeneProf/tools/hypergeometric.jsp)

Data Accessibility Statement

All sequencing and proteomic raw data have been deposited in public repositories. There are no restrictions to data access. ChIP-Seq and ATAC-Seq data are deposited under accession numbers: GSE86490 and GSE101307. The exome sequencing data was deposited in ENA under accession number ERP017908. MS data are available via ProteomeXchange with identifier PXD005011. For RNA-Seq data raw files are deposited in ENA: http://www.ebi.ac.uk/ena. Processed RNA-Seq data are accessible from: ftp://ngs.sanger.ac.uk/production/casm/2018/Gozdecka_et_al_NatGenet/UTX_FPKM_RNAseq/. Description of raw and processed RNA-Seq file is provided in the TableS31.

Supplementary references

54 Yan, M. *et al.* A previously unidentified alternatively spliced isoform of t(8;21) transcript promotes leukemogenesis. *Nat Med* **12**, 945-949, doi:10.1038/nm1443 (2006).

55 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).

56 Xu, S., Grullon, S., Ge, K. & Peng, W. Spatial clustering for identification of ChIP-enriched regions (SICER) to map regions of histone methylation patterns in embryonic stem cells. *Methods in molecular biology* **1150**, 97-111, doi:10.1007/978-1-4939-0512-6_5 (2014).

57 Ross-Innes, C. S. *et al.* Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481**, 389-393, doi:10.1038/nature10730 (2012).

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

59 Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 15545-15550, doi:10.1073/pnas.0506580102 (2005).

60 Kent, W. J. *et al.* The human genome browser at UCSC. *Genome research* **12**, 996-1006, doi:10.1101/gr.229102 (2002).

61 Hulsen, T., de Vlieg, J. & Alkema, W. BioVenn - a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC genomics* **9**, 488, doi:10.1186/1471-2164-9-488 (2008).

62 Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature methods* **10**, 1213-1218, doi:10.1038/nmeth.2688 (2013).

63 Kumasaka, N., Knights, A. J. & Gaffney, D. J. Fine-mapping cellular QTLs with RASQUAL and ATAC-seq. *Nature genetics* **48**, 206-213, doi:10.1038/ng.3467 (2016).

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

65 Pardo, M. *et al.* An expanded Oct4 interaction network: implications for stem cell biology, development, and disease. *Cell stem cell* **6**, 382-395, doi:10.1016/j.stem.2010.03.004 (2010).

66 Kall, L., Canterbury, J. D., Weston, J., Noble, W. S. & MacCoss, M. J. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nature methods* **4**, 923-925, doi:10.1038/nmeth1113 (2007).

67 Brosch, M., Yu, L., Hubbard, T. & Choudhary, J. Accurate and sensitive peptide identification with Mascot Percolator. *Journal of proteome research* **8**, 3176-3181, doi:10.1021/pr800982s (2009).
68 Spivak, M., Weston, J., Bottou, L., Kall, L. & Noble, W. S. Improvements to the percolator algorithm for Peptide identification from shotgun proteomics data sets. *Journal of proteome research* **8**, 3737-3745, doi:10.1021/pr801109k (2009).
69 Teo, G. *et al.* SAINTexpress: improvements and additional features in Significance Analysis of INTeractome software. *Journal of proteomics* **100**, 37-43, doi:10.1016/j.jprot.2013.10.023 (2014).
70 Jones, D. *et al.* cgpCaVEManWrapper: Simple Execution of CaVEMan in Order to Detect Somatic Single Nucleotide Variants in NGS Data. *Current protocols in bioinformatics* **56**, 15 10 11-15 10 18, doi:10.1002/cpbi.20 (2016).
71 Ye, K., Schulz, M. H., Long, Q., Apweiler, R. & Ning, Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* **25**, 2865-2871, doi:10.1093/bioinformatics/btp394 (2009).
72 Raine, K. M. *et al.* cgpPindel: Identifying Somatically Acquired Insertion and Deletion Events from Paired End Sequencing. *Current protocols in bioinformatics* **52**, 15 17 11-12, doi:10.1002/0471250953.bi1507s52 (2015).
73 Boeva, V. *et al.* Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics* **28**, 423-425, doi:10.1093/bioinformatics/btr670 (2012).

Figure 1

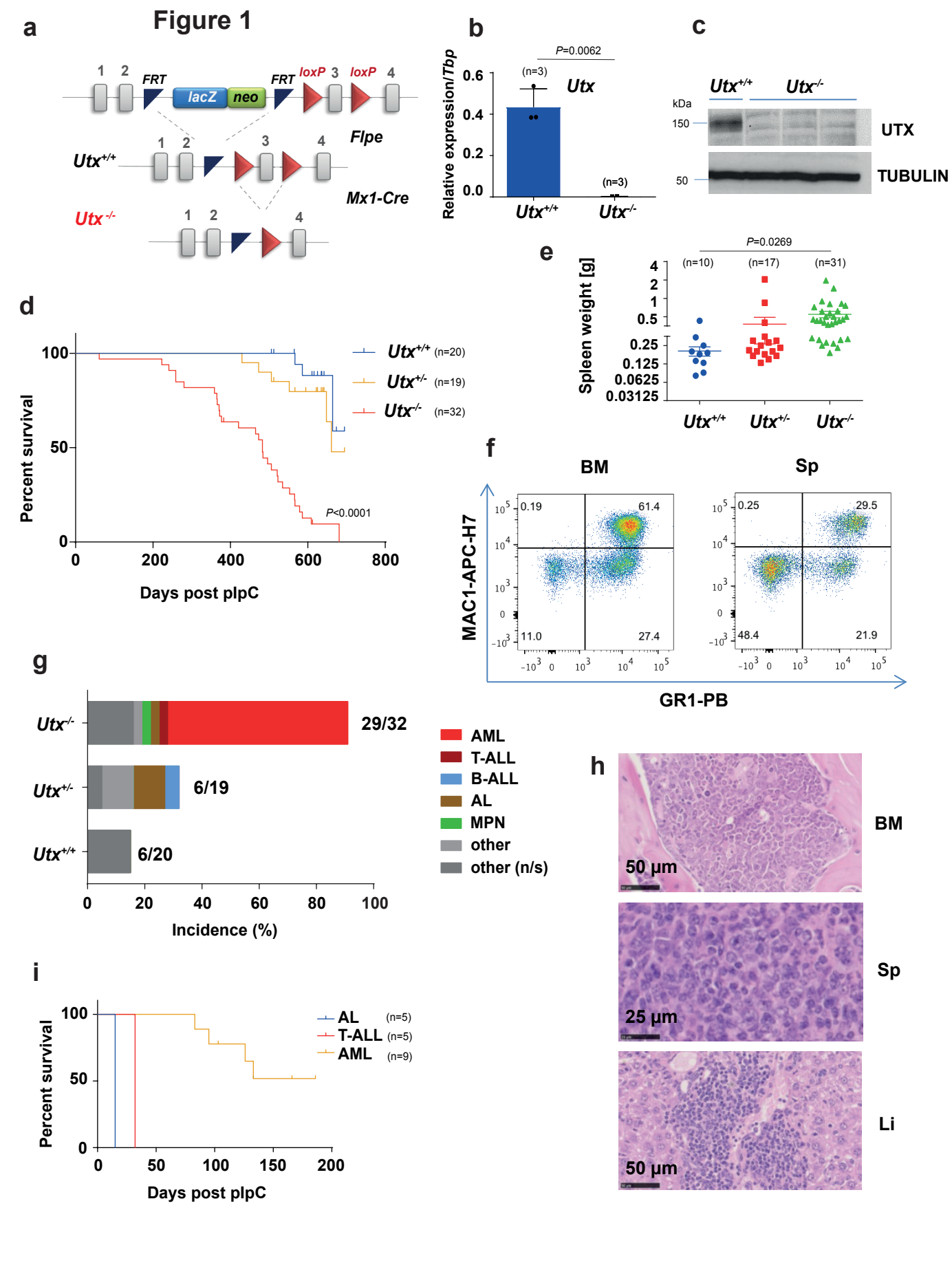

a
1 2 FRT lacZ neo FRT loxP 3 loxP 4
Flpe
Utx+/+ 1 2 3 4
Mx1-Cre
Utx -/- 1 2 4
b
P=0.0062
(n=3)
Utx
(n=3)
Relative expression/Tbp
0.6
0.4
0.2
0.0
Utx+/+
Utx-/-
c
Utx+/+ Utx-/-
kDa
150
UTX
50
TUBULIN
d
Percent survival
100
50
0
0 200 400 600 800
Days post plpC
Utx+/+ (n=20)
Utx+/- (n=19)
Utx-/- (n=32)
P<0.0001
e
P=0.0269
(n=10) (n=17) (n=31)
Spleen weight [g]
4
2
1
0.5
0.25
0.125
0.0625
0.03125
Utx+/+ Utx+/- Utx-/-
f
BM
Sp
MAC1-APC-H7
GR1-PB
0.19
61.4
11.0
27.4
0.25
29.5
48.4
21.9
g
Utx-/-
Utx+/-
Utx+/+
29/32
6/19
6/20
0 20 40 60 80 100
Incidence (%)
AML
T-ALL
B-ALL
AL
MPN
other
other (n/s)
h
BM
50 µm
Sp
25 µm
Li
50 µm
i
Percent survival
100
50
0
0 50 100 150 200
Days post plpC
AL (n=5)
T-ALL (n=5)
AML (n=9)

Figure 2

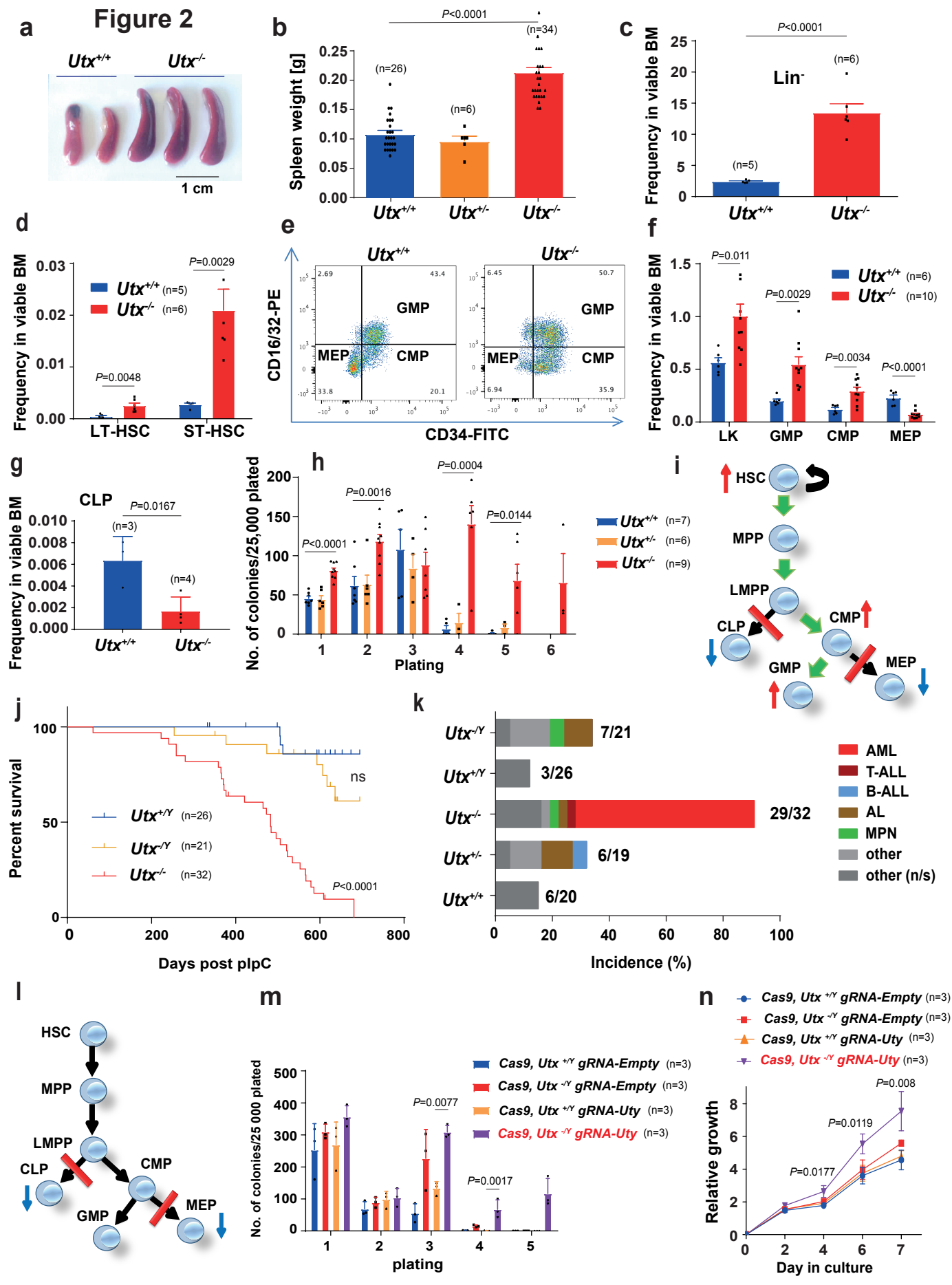

a
Utx+/+
Utx-/-
1 cm
b
P<0.0001
(n=34)
(n=26)
(n=6)
Spleen weight [g]
Utx+/+
Utx+/-
Utx-/-
c
P<0.0001
(n=6)
Lin-
(n=5)
Frequency in viable BM
Utx+/+
Utx-/-
d
P=0.0029
Utx+/+ (n=5)
Utx-/- (n=6)
P=0.0048
LT-HSC
ST-HSC
Frequency in viable BM
e
Utx+/+
Utx-/-
GMP
MEP
CMP
CD16/32-PE
CD34-FITC
f
P=0.011
P=0.0029
P=0.0034
P<0.0001
Utx+/+ (n=6)
Utx-/- (n=10)
LK
GMP
CMP
MEP
Frequency in viable BM
g
CLP
P=0.0167
(n=3)
(n=4)
Frequency in viable BM
Utx+/+
Utx-/-
h
P=0.0004
P=0.0016
P=0.0144
P<0.0001
Utx+/+ (n=7)
Utx+/- (n=6)
Utx-/- (n=9)
No. of colonies/25,000 plated
Plating
i
HSC
MPP
LMPP
CLP
CMP
GMP
MEP
j
ns
P<0.0001
Utx+/Y (n=26)
Utx-/Y (n=21)
Utx-/- (n=32)
Percent survival
Days post plpC
k
Utx-/Y
Utx+/Y
Utx-/-
Utx+/-
Utx+/+
7/21
3/26
29/32
6/19
6/20
AML
T-ALL
B-ALL
AL
MPN
other
other (n/s)
Incidence (%)
l
HSC
MPP
LMPP
CLP
CMP
GMP
MEP
m
Cas9, Utx +/Y gRNA-Empty (n=3)
Cas9, Utx -/Y gRNA-Empty (n=3)
Cas9, Utx +/Y gRNA-Uty (n=3)
Cas9, Utx -/Y gRNA-Uty (n=3)
P=0.0077
P=0.0017
No. of colonies/25 000 plated
plating
n
Cas9, Utx +/Y gRNA-Empty (n=3)
Cas9, Utx -/Y gRNA-Empty (n=3)
Cas9, Utx +/Y gRNA-Uty (n=3)
Cas9, Utx -/Y gRNA-Uty (n=3)
P=0.008
P=0.0119
P=0.0177
Relative growth
Day in culture

Figure 3

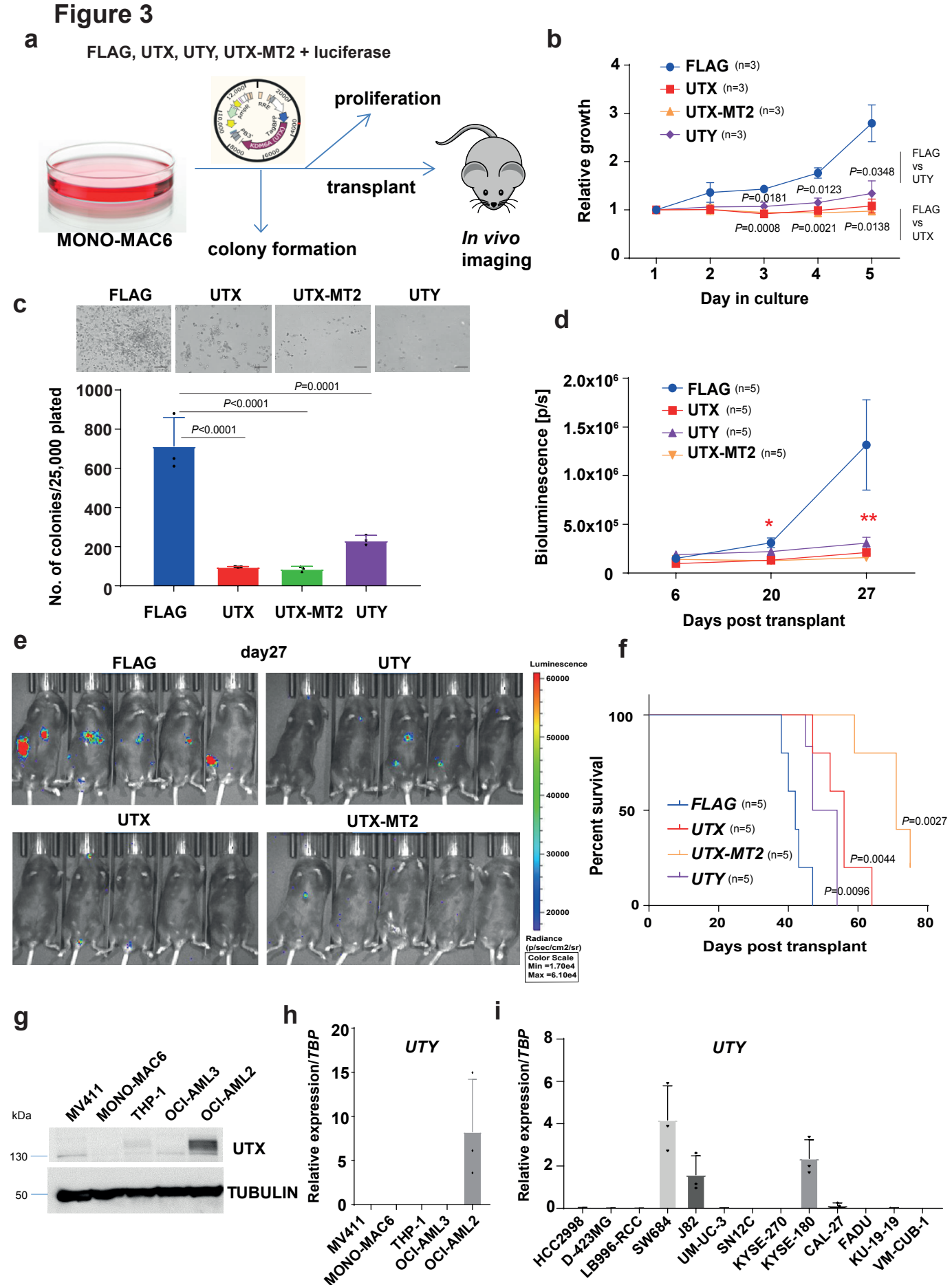

a
FLAG, UTX, UTY, UTX-MT2 + luciferase
proliferation
transplant
colony formation
MONO-MAC6
In vivo imaging
b
FLAG (n=3)
UTX (n=3)
UTX-MT2 (n=3)
UTY (n=3)
Relative growth
Day in culture
P=0.0348
P=0.0181
P=0.0123
P=0.0008
P=0.0021
P=0.0138
FLAG vs UTY
FLAG vs UTX
c
FLAG
UTX
UTX-MT2
UTY
No. of colonies/25,000 plated
P=0.0001
P<0.0001
P<0.0001
d
FLAG (n=5)
UTX (n=5)
UTY (n=5)
UTX-MT2 (n=5)
Bioluminescence [p/s]
Days post transplant
e
day27
FLAG
UTY
UTX
UTX-MT2
Luminescence
Radiance (p/sec/cm2/sr)
Color Scale Min =1.70e4 Max =6.10e4
f
FLAG (n=5)
UTX (n=5)
UTX-MT2 (n=5)
UTY (n=5)
Percent survival
Days post transplant
P=0.0027
P=0.0044
P=0.0096
g
kDa
MV411
MONO-MAC6
THP-1
OCI-AML3
OCI-AML2
UTX
TUBULIN
130
50
h
UTY
Relative expression/TBP
MV411
MONO-MAC6
THP-1
OCI-AML3
OCI-AML2
i
UTY
Relative expression/TBP
HCC2998
D-423MG
LB996-RCC
SW684
J82
UM-UC-3
SN12C
KYSE-270
KYSE-180
CAL-27
FADU
KU-19-19
VM-CUB-1

Figure 4

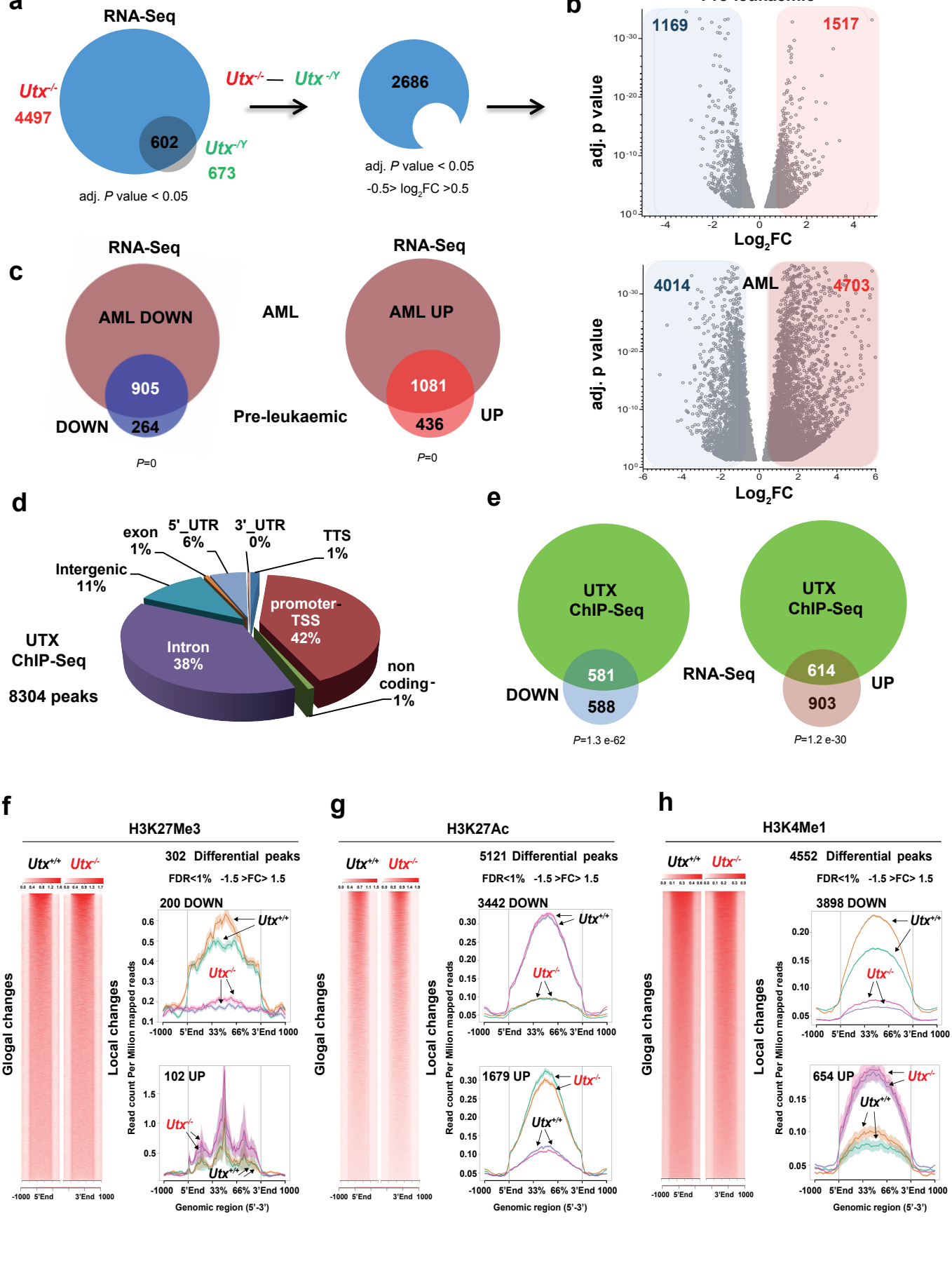
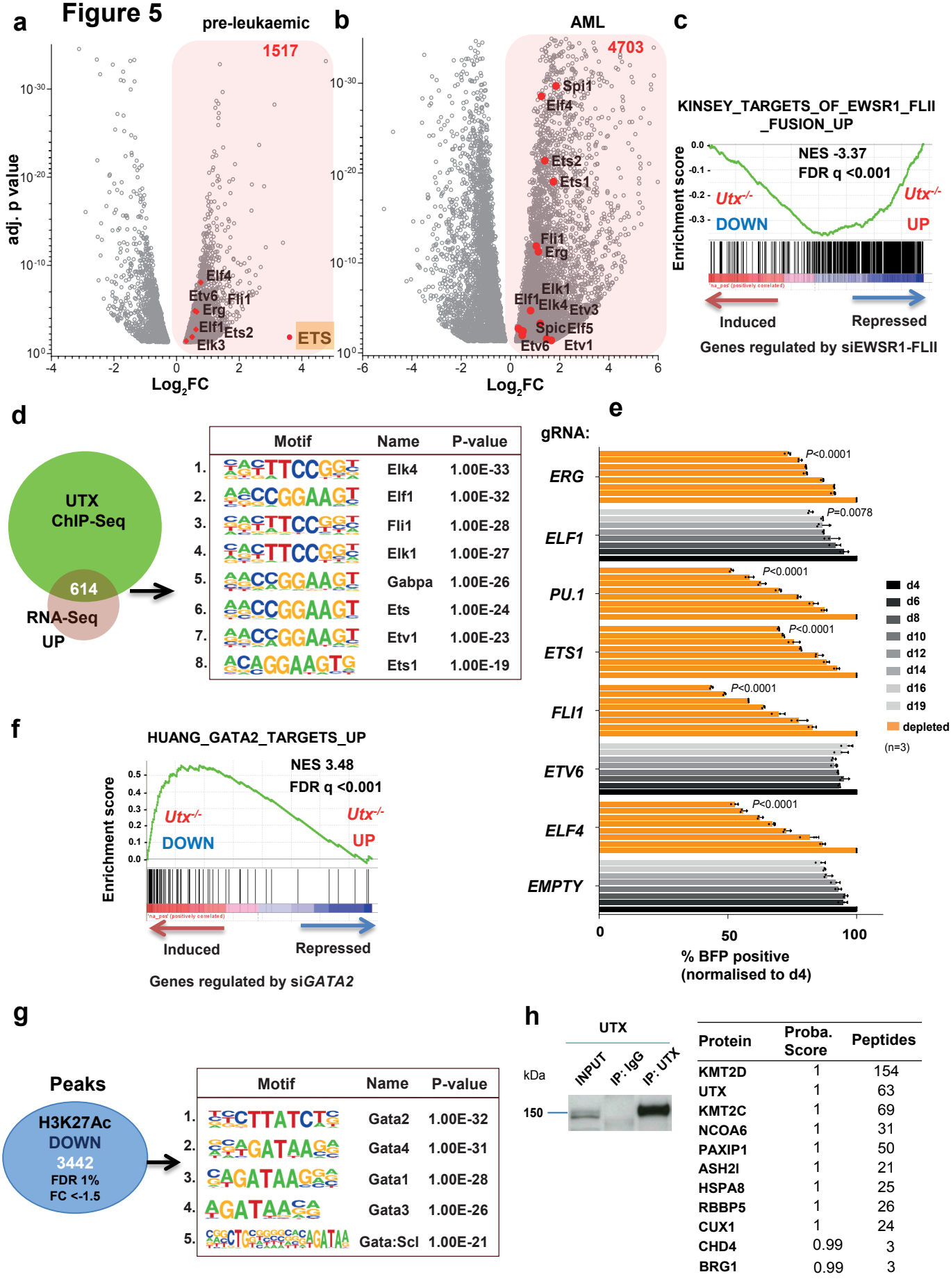
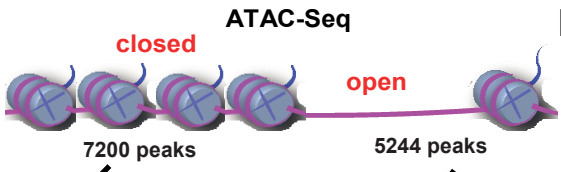

a
RNA-Seq
Utx-/-
4497
602
Utx-/Y
673
adj. P value < 0.05
Utx-/- — Utx-/Y
2686
adj. P value < 0.05
-0.5> log2FC >0.5
b
Pre-leukaemic
1169
1517
adj. p value
Log2FC
4014
AML
4703
c
RNA-Seq
AML DOWN
905
DOWN
264
P=0
AML
Pre-leukaemic
RNA-Seq
AML UP
1081
436
UP
P=0
d
exon 1%
5'_UTR 6%
3'_UTR 0%
TTS 1%
Intergenic 11%
promoter-TSS 42%
Intron 38%
non coding- 1%
UTX ChIP-Seq
8304 peaks
e
UTX ChIP-Seq
581
DOWN
588
P=1.3 e-62
RNA-Seq
UTX ChIP-Seq
614
903
UP
P=1.2 e-30
f
H3K27Me3
302 Differential peaks
FDR<1% -1.5 >FC> 1.5
200 DOWN
102 UP
Gloal changes
Local changes
Read count Per Million mapped reads
Genomic region (5'-3')
g
H3K27Ac
5121 Differential peaks
FDR<1% -1.5 >FC> 1.5
3442 DOWN
1679 UP
h
H3K4Me1
4552 Differential peaks
FDR<1% -1.5 >FC> 1.5
3898 DOWN
654 UP

Figure 5



Protein	Proba. Score	Peptides
KMT2D	1	154
UTX	1	63
KMT2C	1	69
NCOA6	1	31
PAXIP1	1	50
ASH2l	1	21
HSPA8	1	25
RBBP5	1	26
CUX1	1	24
CHD4	0.99	3
BRG1	0.99	3

Figure 6



a

	Motif	Name	P-value
1.		Gata3	1.00E-1920
2.		Gata4	1.00E-1919
3.		Gata2	1.00E-1882
4.		Gata1	1.00E-1847
5.		Gata:Scl	1.00E-1041

b

	Motif	Name	P-value
1.		Ascl1	1.00E-469
2.		E2A	1.00E-453
3.		EBF	1.00E-433
4.		Ptf1a	1.00E-425
5.		Tcf12	1.00E-376
15.		ETS1	1.00E-122
16.		SpiB	1.00E-114
17.		PAX5	1.00E-109
18.		Fli1	1.00E-105
19.		PU.1	1.00E-103