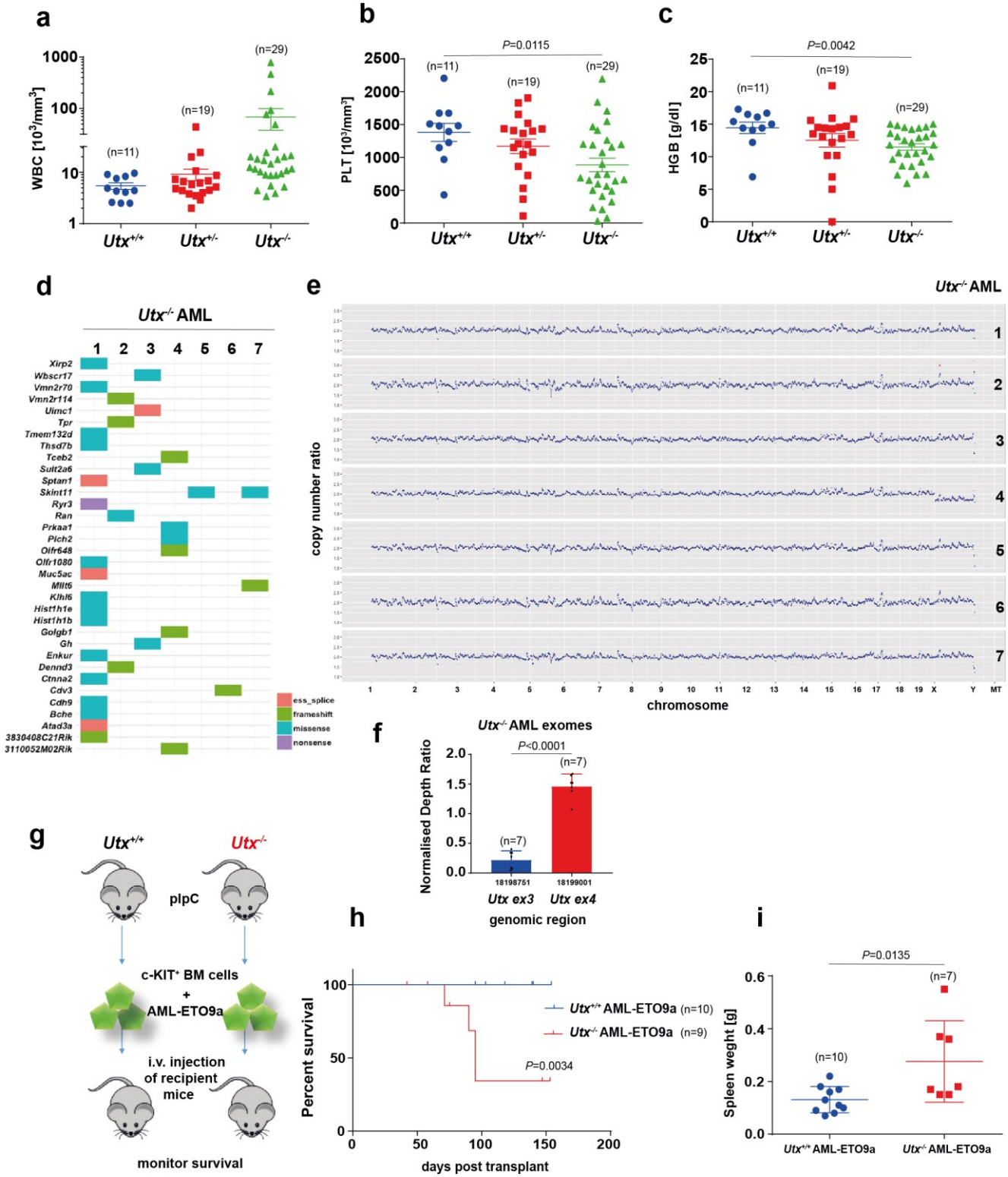
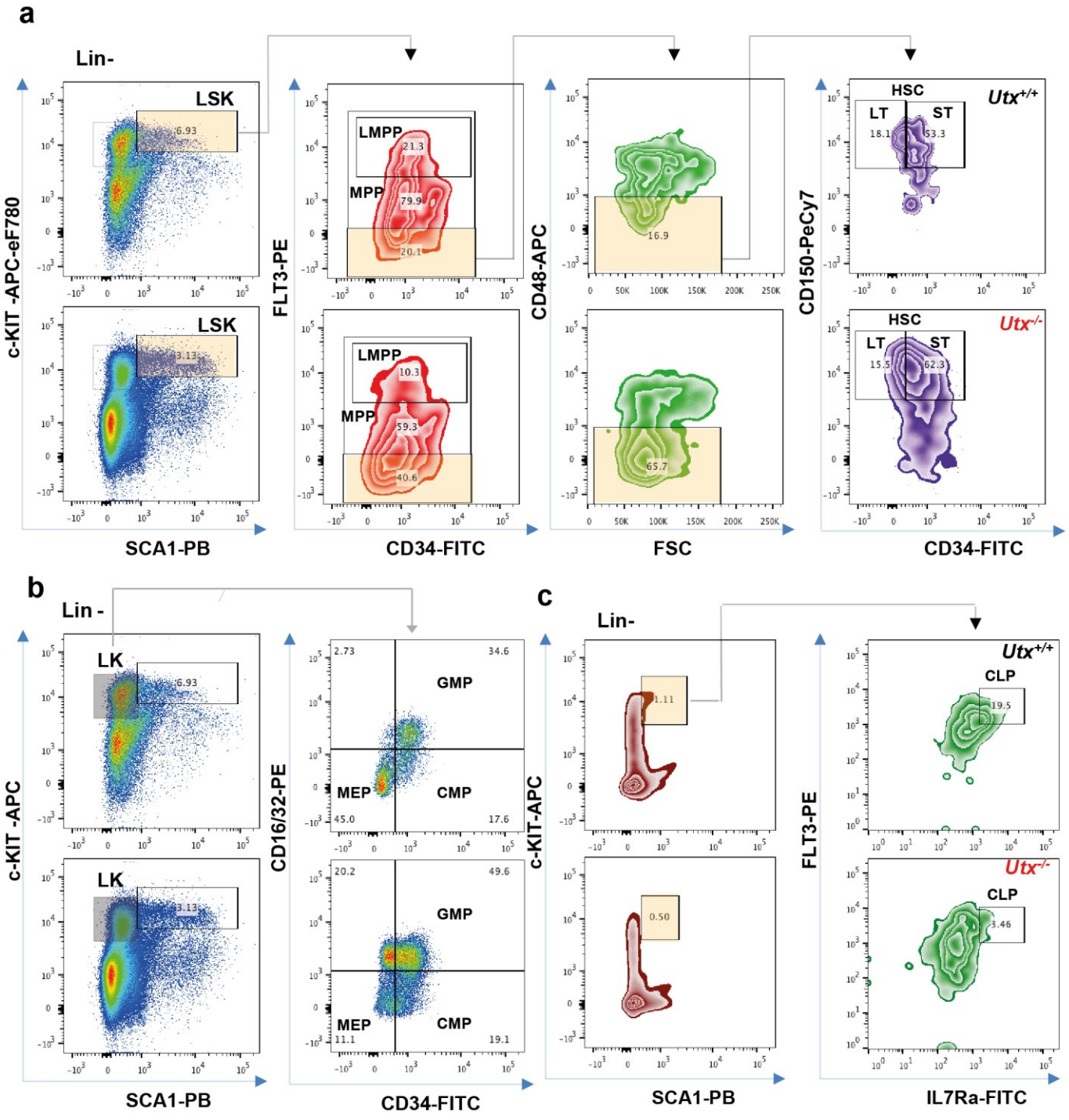
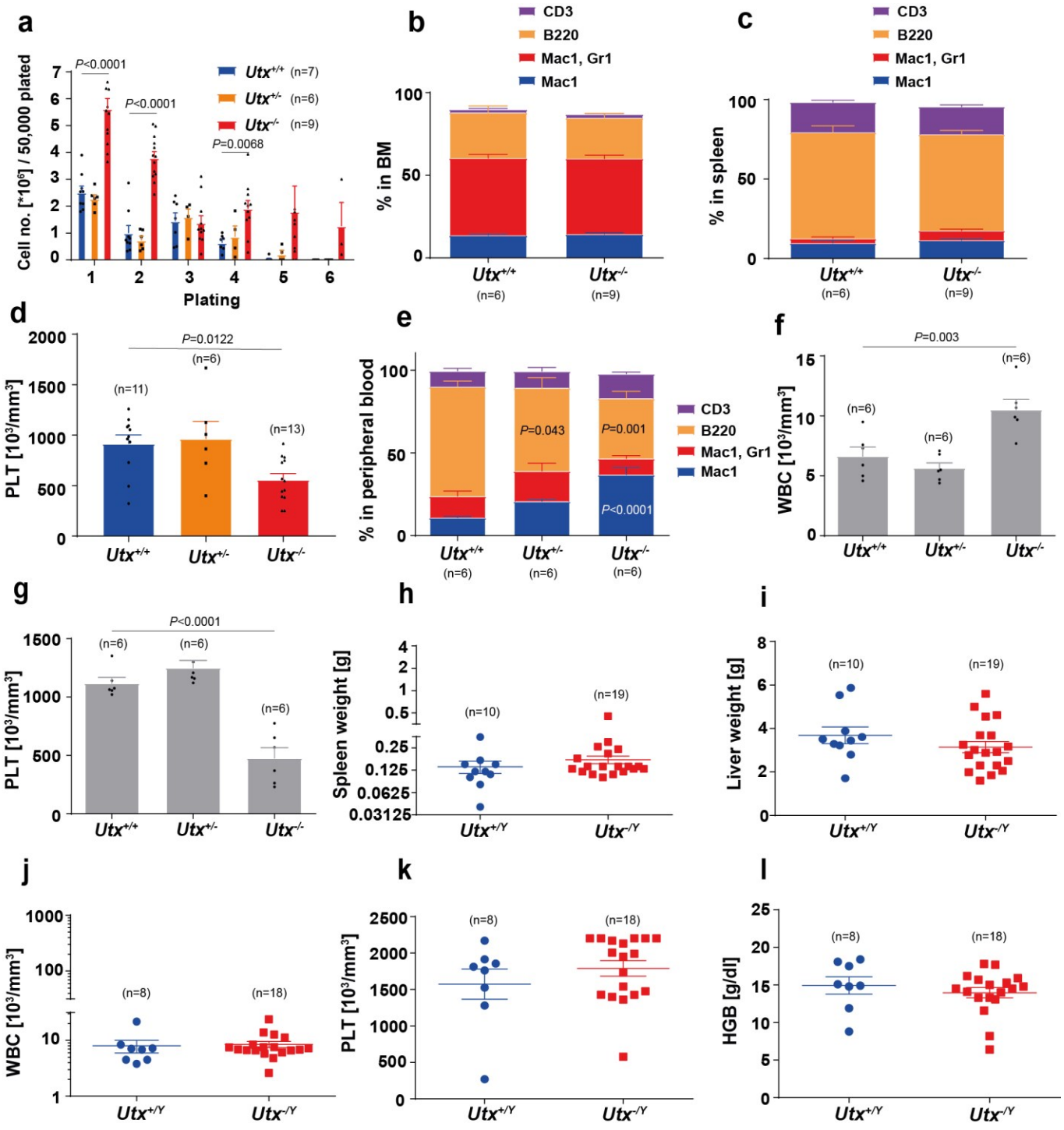
Supplementary Figure 1



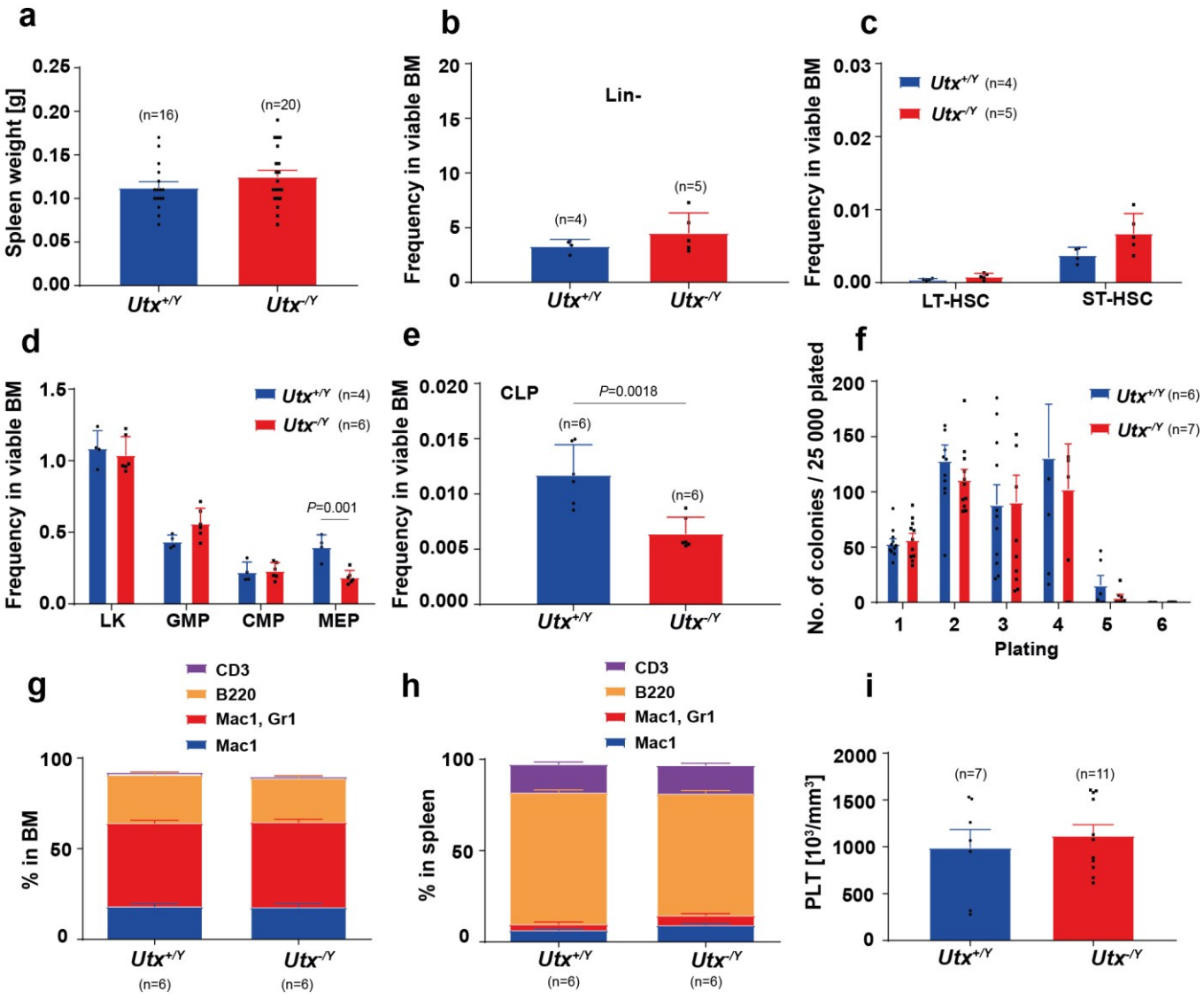
Supplementary Figure 2



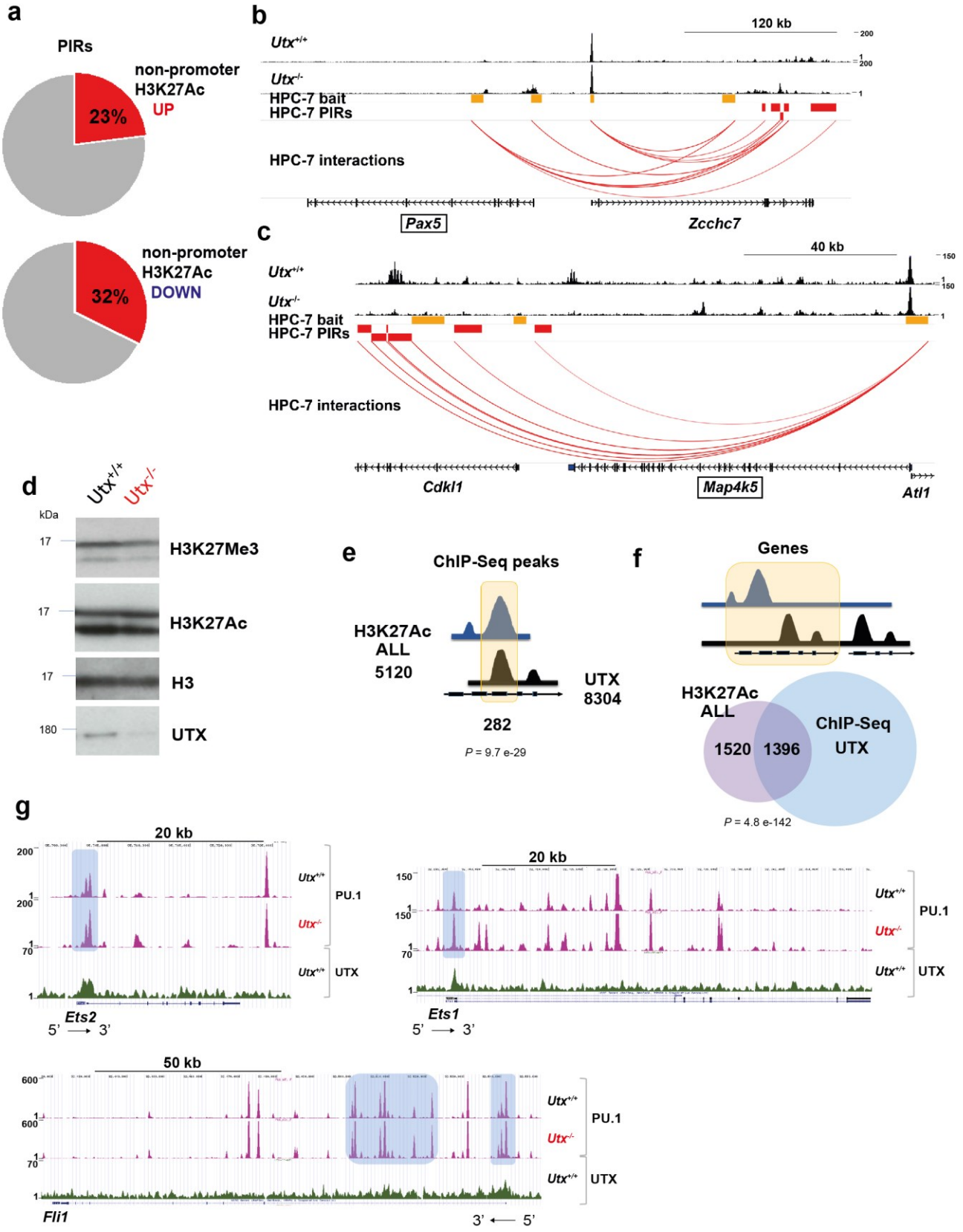
Supplementary Figure 3



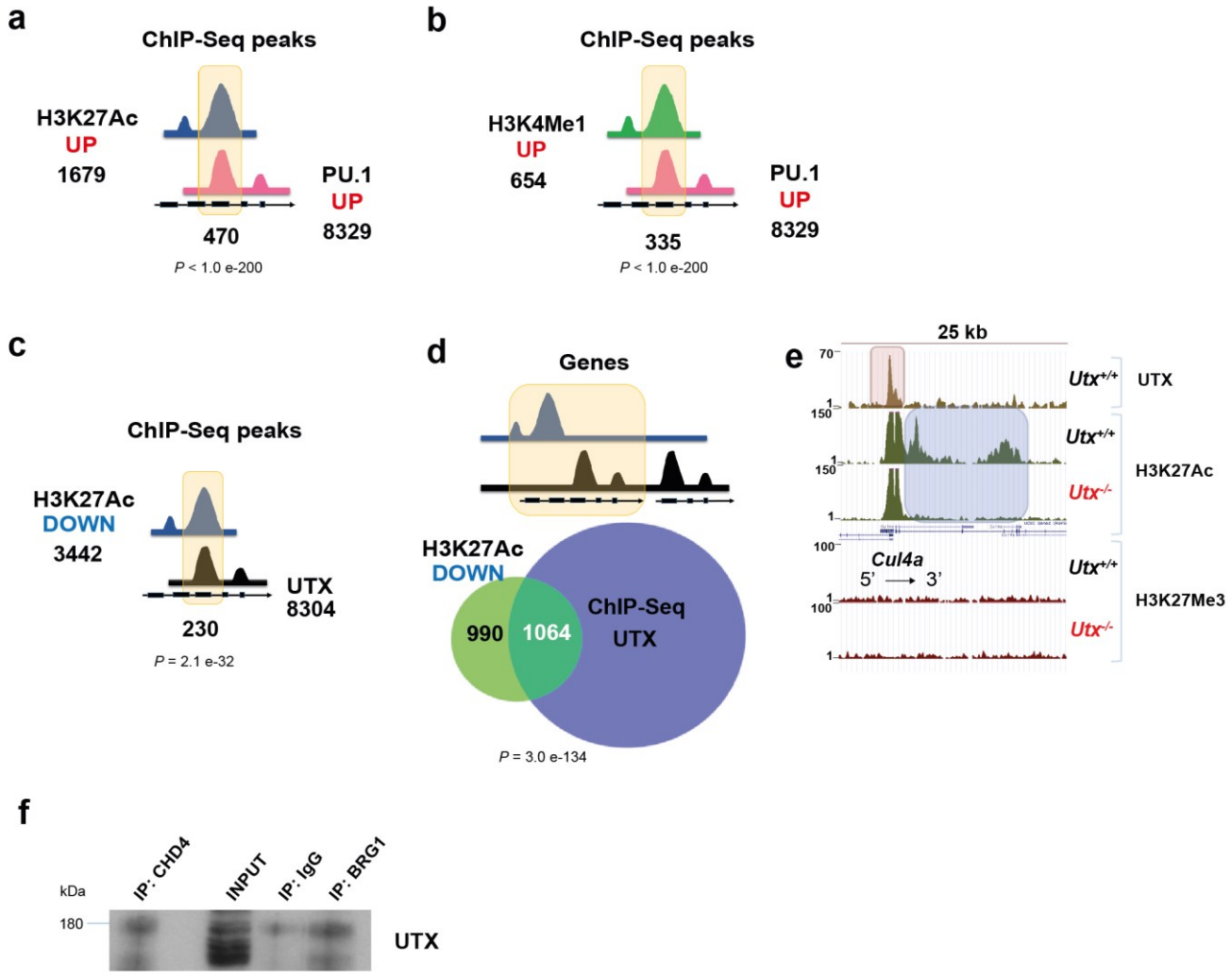
Supplementary Figure 4



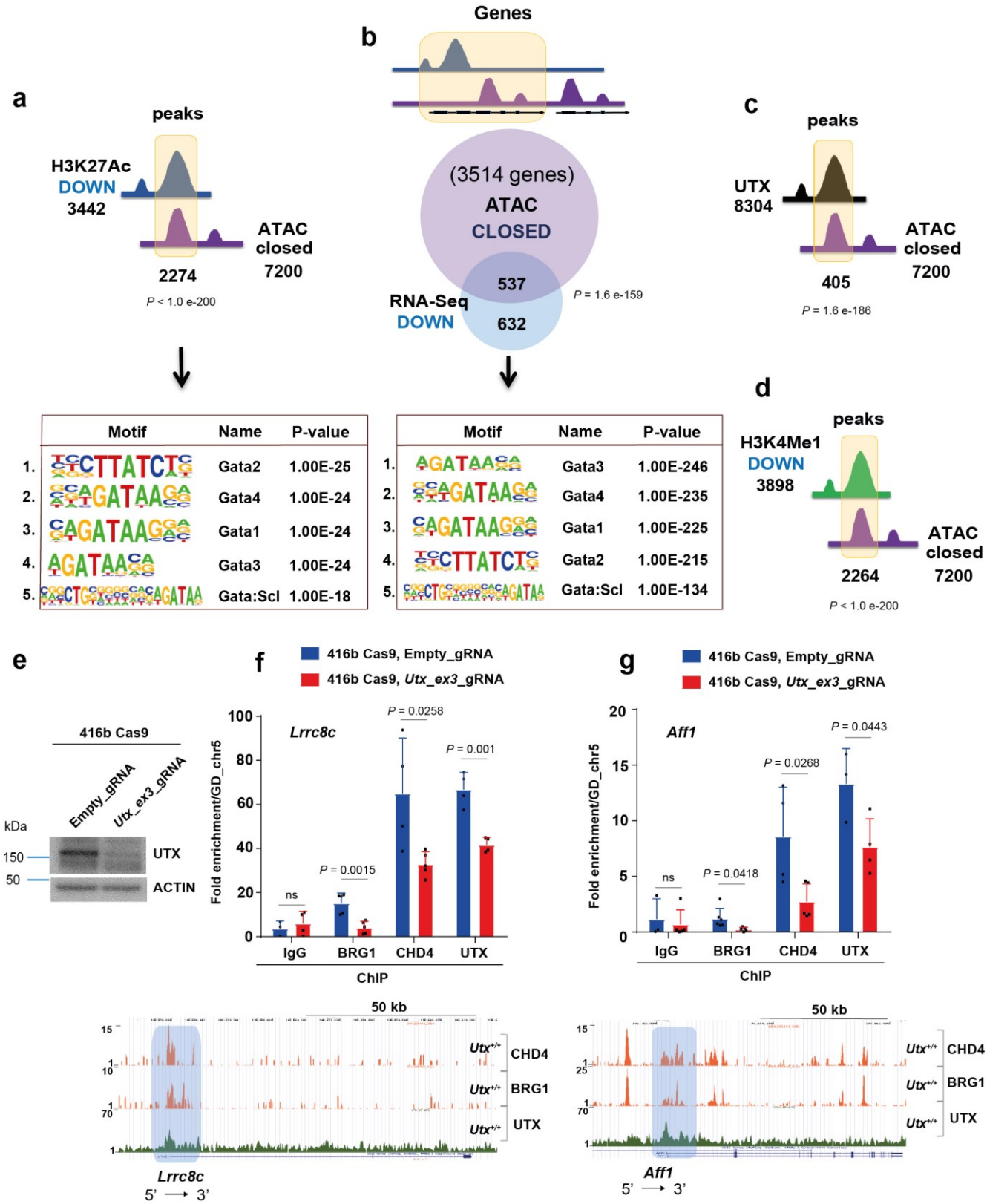
Supplementary Figure 5



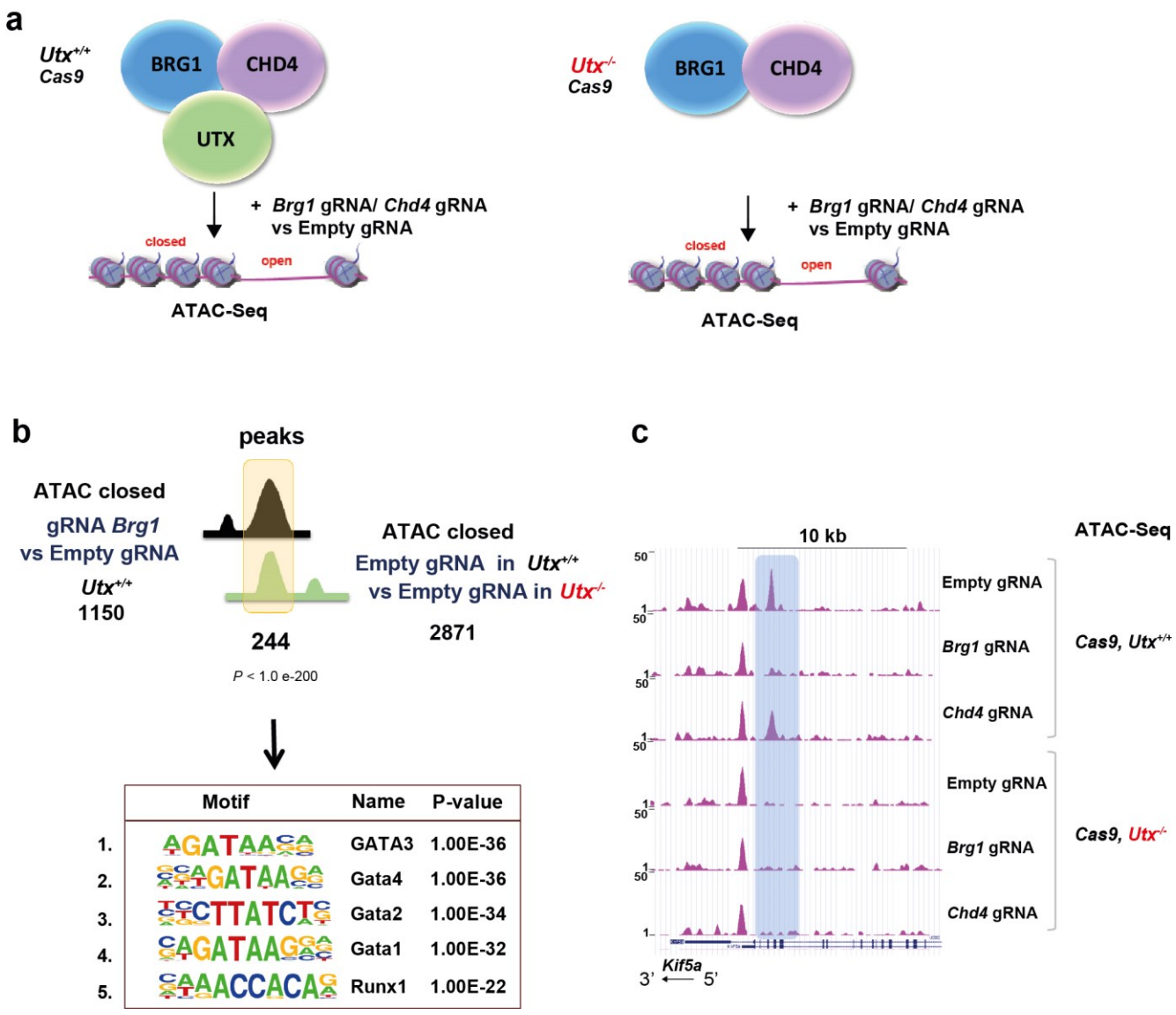
Supplementary Figure 6



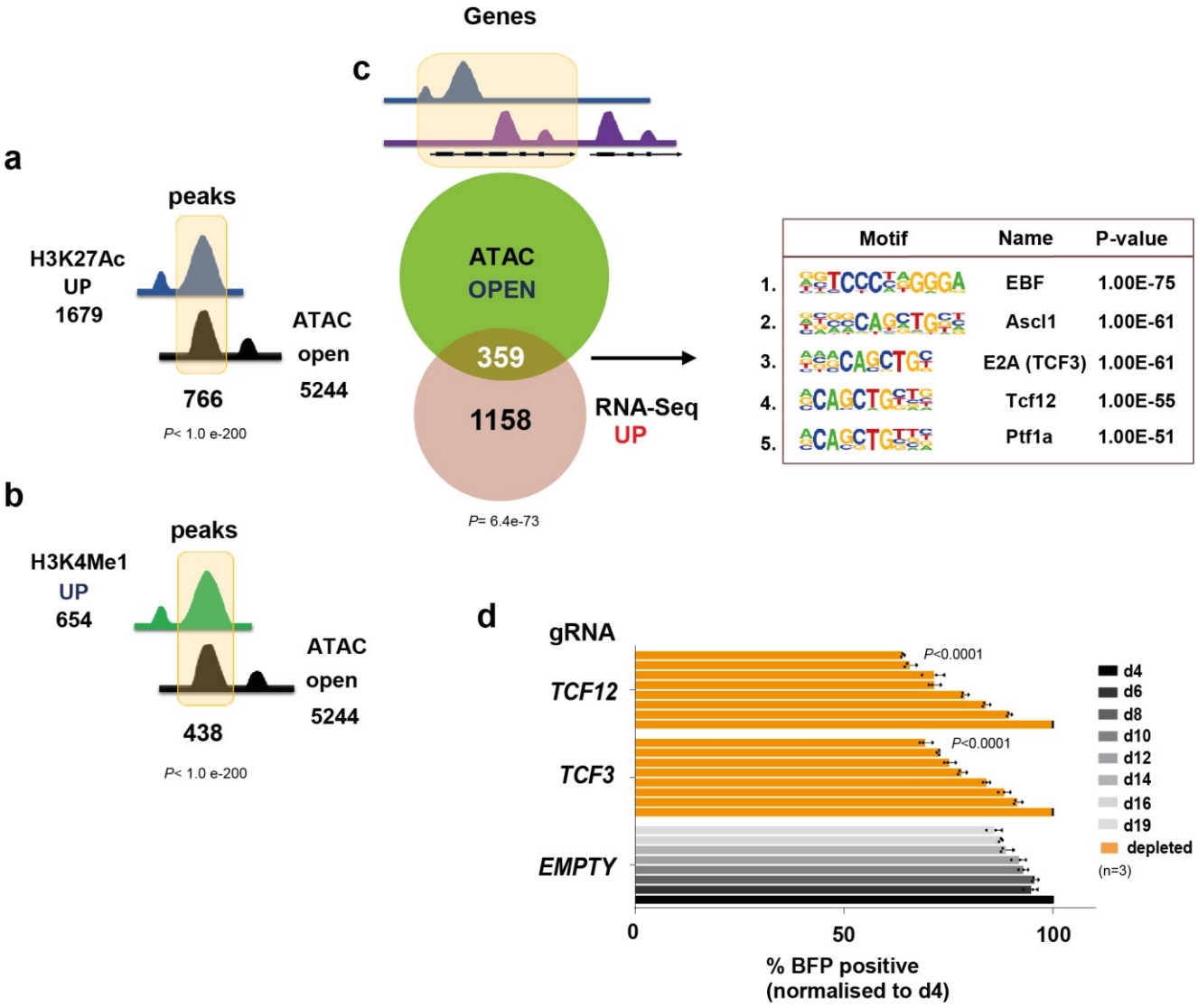
Supplementary Figure 7



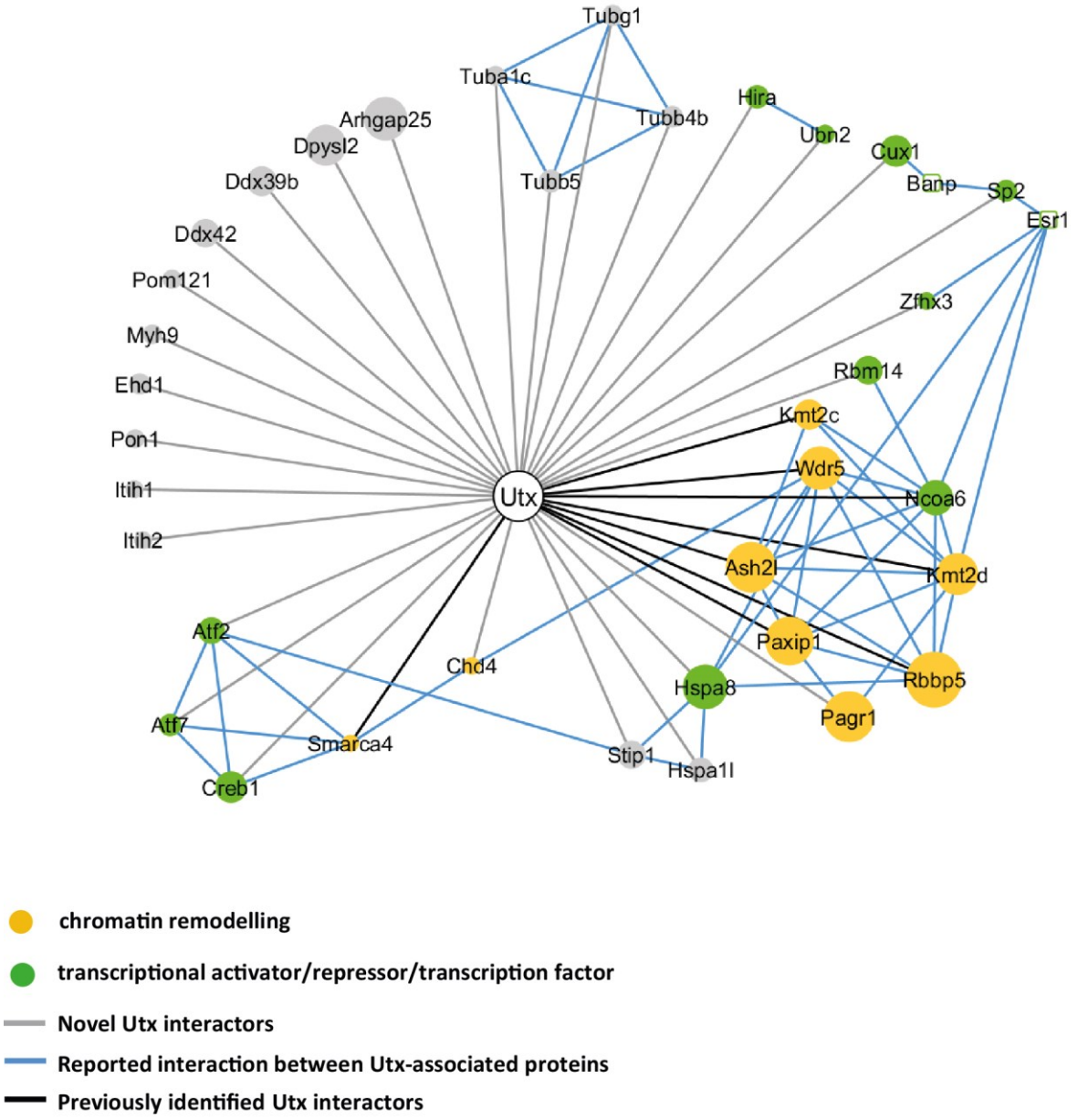
Supplementary Figure 8



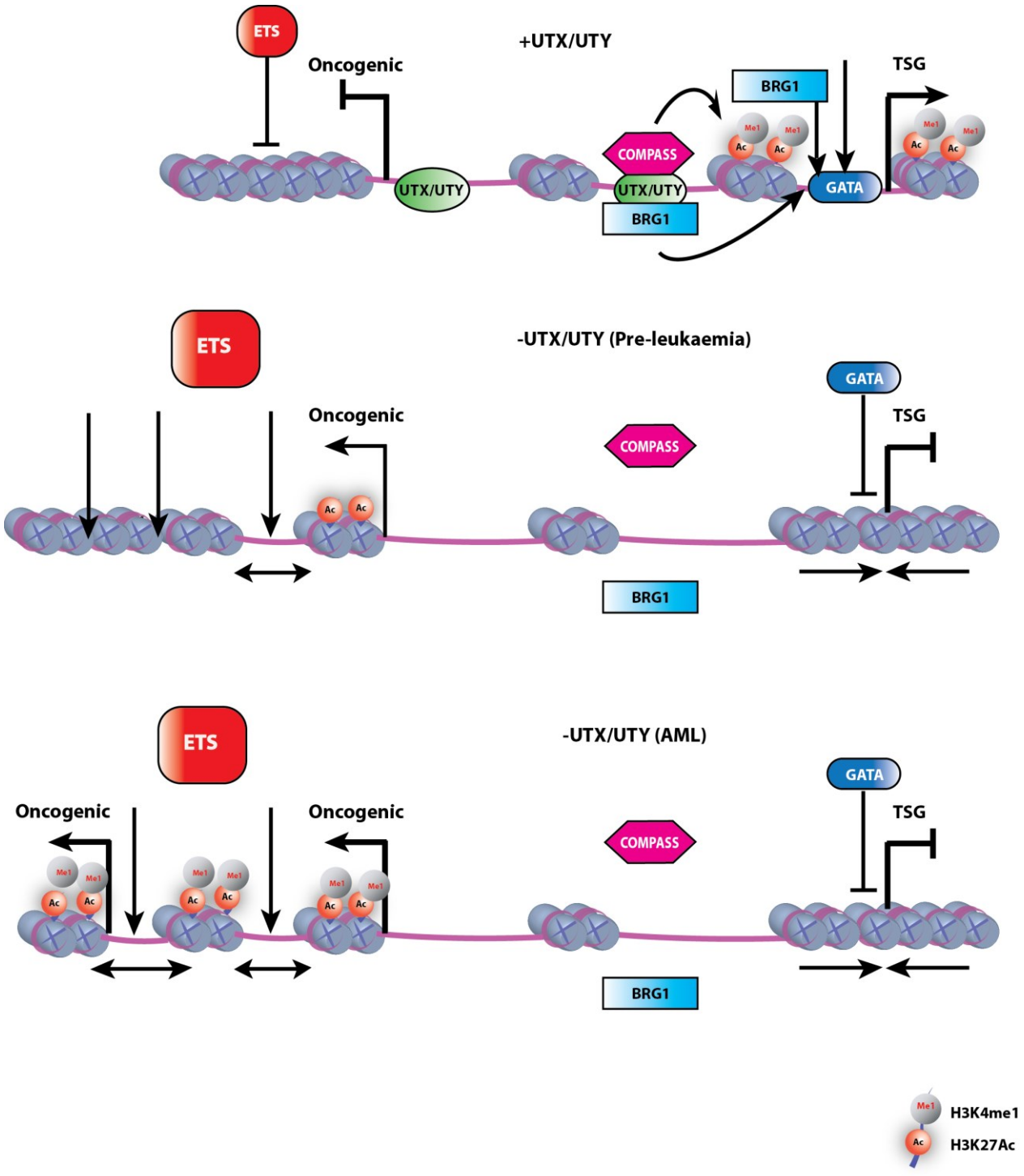
Supplementary Figure 9



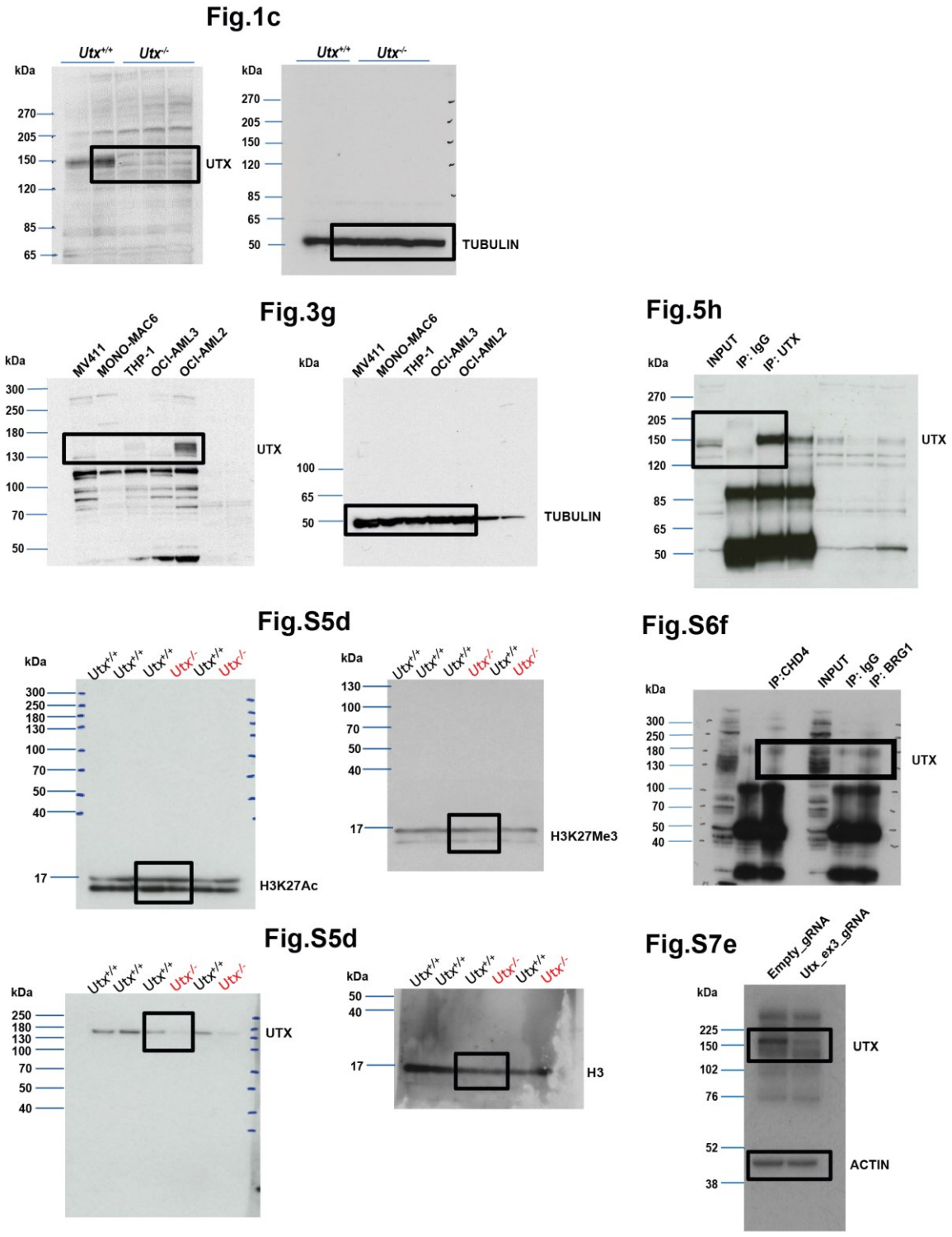
Supplementary Figure 10



Supplementary Figure 11



Supplementary Figure 12



Supplementary Figure Legends

Supplementary Figure 1. Additional blood count results from sick Utx$^{-/-}$ vs Utx$^{+/+}$ mice. Exome sequencing.
(**a**) WBC (**b**) PLT and (**c**) HGB counts of diseased *Utx*$^{+/+}$, *Utx*$^{+/-}$ and *Utx*$^{-/-}$ mice; the mean ± s.e.m is shown; n = number of mice per genotype. P value was determined by one-way ANOVA with Bonferroni correction; for PLT: t = 2.733, df=56; for HGB: t = 2.43; df = 56. (**d**) Mutated genes and (**e**) copy-number changes (from exome sequencing) in 7 individual *Utx*$^{-/-}$ AMLs. (**f**) Deletion of *Utx* exon3 in comparison to exon4 was detected in all *Utx*$^{-/-}$ AMLs. The mean ± s.e.m is shown; n = number of mice; *P* by two-sided t-test (t = 13, df = 12). (**g**) c-KIT^{+} BM cells were isolated form *Utx*$^{+/+}$ and *Utx*$^{-/-}$mice (n = 3 mice per genotype) and transduced with lentivral vectors expressing AML-ETO9a. Cells were transplanted into lethally irradiated syngeneic recipient mice and mouse survival was monitored. (**h**) Kaplan-Meier survival curves of mice transplanted with *Utx*$^{-/-}$; AML-ETO9a and control *Utx*$^{+/+}$; AML-ETO9a cells; n = number of mice; *P* by Log-rank (Mantel-Cox) test, df = 1. (**i**) Spleen sizes of mice in h, the mean ± s.e.m is shown; n = number of mice; *P* by two-sided t-test (t = 2.798, df = 15).

Supplementary Figure 2. Gating strategy for progenitor analysis by FACS.
(**a**) Gating strategy and fluorophores used for separation of LT-HSC, ST-HSC, LSK, MPP, LMPP as well as (**b**) GMP, CMP, MEP and (**c**) CLP populations in *Utx*$^{+/+}$ and *Utx*$^{-/-}$mice.

Supplementary Figure 3. Mature blood cell numbers in *Utx*$^{-/-}$, *Utx*$^{+/-}$ and *Utx*$^{-/Y}$ mice.
(**a**) Cell numbers from pre-leukemic *Utx*$^{+/+}$, *Utx*$^{+/-}$, and *Utx*$^{-/-}$ in a serial re-plating assay (for *Utx*$^{+/+}$ vs *Utx*$^{-/-}$ in plating: 1, t = 7.018; df = 25; plating 2, t = 8.668, df = 25; plating 4 t = 3.342, df = 19). Mature cell frequencies in pre-leukemic (**b**) BM and (c) spleen as well as (c) PLT counts in *Utx*$^{+/+}$, *Utx*$^{+/-}$,and *Utx*$^{-/-}$ mice (t = 2.977, df = 27). (**e**) Relative mature cell frequencies in blood of 36-week old mice; *P* calculated vs *Utx*$^{+/+}$ control; only significant *P* values are shown. For *Utx*$^{+/+}$ vs *Utx*$^{-/-}$ comparison, MAC1 (t = 6.718, df = 15), B220 (4.442, df = 15); for *Utx*$^{+/+}$ vs *Utx*$^{+/-}$ comparison, MAC1 (t = 2.571; df = 15) (**f**) WBC (t = 3.879, df = 15) and (**g**) PLT counts (t = 6.487, df = 15) from aged (36-week-old) *Utx*$^{+/+}$, *Utx*$^{+/-}$ and *Utx*$^{-/-}$. (**h**) Spleen (**i**) liver weights, (**j**) WBC, (**k**) PLT and (**l**) HGB in aged (for 22 months) *Utx*$^{+/Y}$, and *Utx*$^{-/Y}$. The mean ± s.e.m is shown; n = number of mice per genotype. In **a, d, e, f,** and **g**, *P* value was determined by one-way ANOVA with Bonferroni correction. In **b, c, h-i** *P* was determined by two-sided t-test as ns.

Supplementary Figure 4. Extended hemopoietic phenotyping of pre-leukemic *Utx*$^{+/Y}$ and *Utx*$^{-/Y}$ mice.
(**a**) Spleen weights of pre-leukemic *Utx*$^{+/Y}$, and *Utx*$^{-/Y}$. (**b**) HSPC/Lin- and (**c**) LT-HSC/ST-HSC frequency in BM cells from *Utx*$^{+/Y}$, and *Utx*$^{-/Y}$. (**d**) LK, CMP, GMP, MEP and (**e**) CLP frequency in BM from pre-leukemic *Utx*$^{+/Y}$, and *Utx*$^{-/Y}$. (**f**) BM colonies from *Utx*$^{+/Y}$, and *Utx*$^{-/Y}$ in a serial re-plating assay. (**g**) Mature cell frequencies in pre-leukemic BM and (**h**) spleen as well as (**i**) PLT counts in *Utx*$^{+/Y}$, and *Utx*$^{-/Y}$ mice. In **a-i** the mean ± s.e.m is shown; n = number of mice per genotype, *P* by two-sided t-test. For CLP, t = 4.206, df = 10; for MEP, t = 5.028, df = 8. Only significant *P* values are shown.

Supplementary Figure 5. Interaction of promoter and putative enhancer regions in *Utx*$^{-/-}$ mice.
(**a**) Proportion of differential H3K27ac peaks at non-promoter regions overlapped with promoter-interacting regions (PIRs). Based on promoter capture Hi-C data in HPC-7 cell line, a total number of 54,339 PIRs were defined as HindIII digested fragments that form significant interactions (CHiCAGO score ≥5) with promoter baits. (**b**) Regions marked by increased H3K27ac in *Utx*$^{-/-}$ vs *Utx*$^{+/+}$ HSPCs interact with the Pax5 promoter in HPC-7 cells. (**c**) Regions marked by decreased H3K27ac in *Utx*$^{-/-}$ vs *Utx*$^{+/+}$ HSPCs interact with the Map4k5 promoter in HPC-7 cells.(**d**) Immunoblots showing global levels of H3K27me3, H3K27ac and UTX protein in the *Utx*$^{-/-}$ and *Utx*$^{+/+}$ pre-leukemic mouse BM cells. Histone 3 (H3) was used as protein loading control. Results of one representative experiment are shown (n = 3 experiments). Uncropped images are shown in Supplementary Fig. 12 (**e**) Overlap of peaks and (**f**) peak-associated genes between UTX-bound sites (6734 loci) and all differential H3K27ac modifications (2916 loci). Differential H3K27ac peaks (FDR<1%, -1.5>FC>1.5) were defined by DiffBind tool; n = 2 mice per genotype. (**g**) Genomic snapshot of UTX and PU.1 demonstrates enhanced PU.1 binding in the promoter region of *Ets1*, *Ets2* and *Fli1* in the absence of *Utx*. UTX also binds

to the promoter of these genes in *Utx*$^{+/+}$ mice. In **e**, *P* by Fisher's exact test for peak comparisons; in **f**, *P* by hypergeometric test for gene comparisons.

Supplementary Figure 6. Overlap of UTX and PU.1 binding with differential H3K27ac and H3K4me1.
(**a**) Overlap of peaks between PU.1-enriched sites and increased H3K27ac or (**b**) gained H3K4me1 sites. (**c**) Overlap of peaks and (**d**) peak-associated genes between UTX-bound regions (2916 genes) and regions with loss of H3K27ac sites (2054 genes) in the *Utx*$^{-/-}$. (**e**) Genomic snapshot of UTX, H3K27ac, H3K27me3 ChIP-seq in *Utx*$^{+/+}$ and *Utx*$^{-/-}$ Lin− at the Cul4a locus showing no direct co-localisation of UTX binding with changes in H3K27ac. Instead UTX binds in proximity to these changes. (**f**) Co-IP showing association between UTX and both SMARCA4 and CHD4. Uncropped images are shown in Supplementary Fig. 12 *P* by Fisher's exact test for peak comparisons; *P* by hypergeometric test for gene comparisons. Differential peaks were defined by DiffBind tool; n = 2 mice per genotype for H3K27ac, H3K4me1 and UTX; n = 3 mice per genotype for PU.1 ChIP-seq.

Supplementary Figure 7. Correlation of closed chromatin regions with UTX binding and altered H3K27ac.
(**a**) Intersection of peaks between differentially closed chromatin and loss of H3K27ac followed by motif analysis of overlapping sites; number indicates motif rank. (**b**) Comparison of downregulated genes to closed chromatin regions in *Utx*$^{-/-}$ followed by motif analysis of the overlap; number indicates motif rank. Motif analysis in lower panels of **a** and **b** - by HOMER software. (**c**) Intersection of peaks between UTX-bound and open chromatin sites. (**d**) Intersection of closed chromatin sites and loss of H3K4me1. (**e**) Immunoblot showing UTX protein level in 416B-Cas9 cells edited with Utx gRNA or empty gRNA control. Actin was used as loading control (repeated in n = 3 experiments). Uncropped images are shown in Supplementary Fig.12 (**f**) ChIP for SMARCA4, CHD4, UTX followed by qPCR for Aff1 and (**g**) Lrrc8c loci shown as fold enrichment over control loci (GD_chr5, gene desert on chromosome 5) in 416B-Cas9 cells with edited *Utx* vs Empty-gRNA control. Genomic snapshots show chromatin co-occupacy of UTX, SMARCA4 and CHD4 at studied regions. The mean ± s.e.m is shown; n = 3-5 independent cell cultures; *P* by two-sided t-test. In **f** (Utx_ex3 gRNA vs Empty-gRNA) for SMARCA4 t = 2.371; df = 9; for CHD4 t = 2.793, df = 7; for UTX t = 2.671, df = 5. In g (Utx_ex3 gRNA vs Empty-gRNA) for SMARCA4 t = 4.723; df = 8; for CHD4 t = 2.818, df = 7; for UTX t = 6.009, df = 6. In **a, c**, and **d**, *P* by Fisher's exact test; in **b**, *P* by hypergeometric test. Differential peaks were defined by DiffBind tool; n = 2 mice per genotype for H3K27ac and UTX; n = 3 mice per genotype for ATAC-seq.

Supplementary Figure 8. Functional redundancy between UTX and SMARCA4 loss for chromatin accessibility on GATA sites.
(**a**) Schematic representation of experimental strategy: *Cas9* expressing *Utx*$^{-/-}$ and *Utx*$^{+/+}$ HSPCs were transduced with lentiviral vector expressing gRNA for Smarca4 and Chd4 and empty gRNA as a control. Six days post transduction we analysed chromatin accessibility by ATAC-seq. (**b**) Differential peaks between empty gRNA and Smarca4 gRNA in *Utx*$^{+/+}$ cells (FDR<0.05; FC< −1.25) were compared with differential peaks lost upon *Utx* deletion (empty gRNA in *Utx*$^{+/+}$ was compared to empty gRNA in *Utx*$^{-/-}$; FDR<0.01; FC< −1.5) followed by motif analysis of the overlap. Differential peaks were defined by DiffBind tool; n = 2 mice per genotype, *P* for peak comparison by Fisher's exact test. Motif and statistical analysis in lower panel of b was determined by HOMER software. (**c**) Genomic snapshot of ATAC-seq at *Kif5a* locus. Note that *Smarca4* editing but not *Chd4* knockout phenocopy *Utx* loss at indicated in blue region.

Supplementary Figure 9. Correlation of open chromatin sites with gain of H3K27ac, H3K4me1 and gene expression.
(**a**) Intersection of peaks between differentially opened chromatin and gained H3K27ac and (**b**) gained H3K4me1 sites. (**c**) Comparison of upregulated genes to opened chromatin regions (2175 genes) in *Utx*$^{-/-}$ followed by motif analysis of the overlap, number indicates motif rank. *P* in **a** and **b**, by Fisher's exact test; *P* in c by hypergeometric test. Motif and statistical analysis was determined by HOMER software. Differential peaks were defined by DiffBind tool; n = 2 mice per genotype for H3K27ac and H3K4me1; n = 3 mice per genotype for ATAC-seq. (d) MONO-MAC6 proliferation upon editing of TCF3 and TCF12. 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 edited gene on day 19 vs control gRNA (EMPTY) d19; for TCF12, t = 18.81, df = 6; for TCF3, t = 14.22, df=6. d – day in culture.

Supplementary Figure 10. UTX interaction network based on the mass spectrometry data.
Previously reported interactions were obtained from STRING database, Uniprot annotations and literature searches. The graph was drawn with Cytoscape using the organic layout, which was modified manually to allow better visualization of the connections. Node size represents abundance of the protein. Grey edges represent novel interactions identified by UTX IP-MS, light blue edges represent interactions reported in the literature, and black edges represent previously reported interactions. Note that Esr1 and Banp were not identified in the UTX IP-MS experiment. The protein were included in the network to illustrate their interactions with the newly identified UTX-binding proteins.

Supplementary Figure 11. Model.
(**a**) In the presence of UTX or UTY, the oncogenic ETS transcriptional program is suppressed by tight regulation of ETS TF expression levels and through coordinated recruitment of chromatin remodelers such as SMARCA4 and the COMPASS complex to open chromatin and lay enhancer marks for important tumour suppressor TF such as GATA2. (**b**) However, when UTX/UTY are initially lost (in the pre-leukemic phase of disease) an increased expression and binding of ETS TFs occur across the genome, leading to novel enhancer generation with immediate effects upon gene expression at some loci (full arrow). For a larger number of regions ETS TFs bind to closed and compacted chromatin (dashed arrows). In addition, loss of UTX/UTY leads to a lack of coordinated binding and activity of SMARCA4 and the COMPASS complex closing off chromatin accessibility and leading to a loss of enhancer function at GATA binding sites, switching off the GATA program. In addition, the loss of H3K4me1 further leads to a failure to recruit and activate SMARCA4. (**c**) Subsequently and potentially related to the effects of cooperating mutations, the pioneering effects of ETS TFs lead to gene activation and this evolution of the leukemic transcriptional programs leads to the development of overt AML.

Supplementary Figure 12.
Uncropped western blot images presented in the main and supplementary figures.