A mutant ASXL1-EHMT complex contributes to heterochromatin dysfunction in clonal hematopoiesis and chronic monomyelocytic leukemia

Zhen Dong[1,2,3,4], Hugo Sepulveda[1,2,3,4], Leo Josue Arteaga[1], Chad Blouin[1], Jenna Fernandez[5], Moritz Binder[5], Wen-Chien Chou[8], Hwei-Fang Tien[8], Geoffrey J Faulkner[6,7], Samuel A Myers[1,3] Mrinal Patnaik[5], Anjana Rao [1,2,3,4]

[1] Department of Signaling and Gene Expression, La Jolla Institute for Allergy and Immunology, 9420 Athena Circle, La Jolla, California 92037, USA

[2] Sanford Consortium for Regenerative Medicine, La Jolla, California 92037, USA.

[3] Department of Pharmacology, University of California, San Diego, California 92093, USA.

[4] Moores Cancer Center, University of California, San Diego, California 92093, USA

[5] Division of Hematology, Department of Internal Medicine, Mayo Clinic, MN

[6] Mater Research Institute, University of Queensland, TRI Building, Woolloongabba, QLD, 4102, Australia

[7] Queensland Brain Institute, University of Queensland, Brisbane QLD 4072, Australia

[8] Division of Hematology, Department of Internal Medicine, National Taiwan University Hospital, No. 7, Chung-Shan S Rd, Taipei 10002, Taiwan

Abstract

ASXL1 is one of the three most frequently mutated genes in age-related clonal hematopoiesis (CH), with the others being *DNMT3A* and *TET2* [1,2]. CH can progress to myeloid malignancies including chronic monomyelocytic leukemia (CMML), and is also strongly associated with inflammatory cardiovascular disease and all-cause mortality in humans [3,4,5]. DNMT3A and TET2 regulate DNA methylation and demethylation pathways respectively [6,7], and *DNMT3A* and *TET2* loss-of-function mutations in CH reduce DNA methylation in heterochromatin, allowing de-repression of silenced elements in heterochromatin [8,9,10]. In contrast, the mechanisms that connect mutant ASXL1 and CH are not yet fully understood. CH/CMML-associated *ASXL1* mutations encode C-terminally truncated proteins that enhance the deubiquitinase activity of the ASXL-BAP1 “PR-DUB” deubiquitinase complex, which removes mono-ubiquitin from H2AK119Ub [11,12,13]. Here we show that ASXL1 mutant proteins interact with the EHMT1-EHMT2 methyltransferase complex, which generates H3K9me1 and me2, the latter a repressive modification in constitutive heterochromatin. Compared to cells from age-matched wildtype mice, we found that expanded myeloid cells from old (≥18-month-old) *Asxl1tm/+* mice [14], a heterozygous knock-in mouse model of CH, display genome-wide decreases of H3K9me2, H3K9me3 and H2AK119Ub as well as an associated increase in expression of transposable elements (TEs) and satellite repeats. Increased TE expression was also observed in monocytes from *ASXL1*-mutant CMML patients compared to monocytes from healthy control individuals. Our data suggest that mutant ASXL1 proteins compromise the integrity of both constitutive and facultative heterochromatin in an age-dependent manner, by reducing the levels of H3K9me2/3 and H2AK119Ub respectively. The resulting increase in TE expression can alter the expression of nearby genes and promote the expression of inflammation-associated and interferon-inducible genes (ISGs).

Introduction

ASXL1, ASXL2 and ASXL3, the human homologues of the *Drosophila Asx* (*additional sex combs*) gene [11], are obligate regulatory components of the PR-DUB BAP1 deubiquitinase complex, which removes ubiquitin from H2AK119Ub [12]. Somatic mutations in *ASXL1* are frequent in clonal hematopoiesis (CH), an age-associated condition that can progress to myeloid malignancies including myelodysplastic syndrome (MDS), chronic monomyelocytic leukemia (CMML) and acute myeloid leukemia (AML), and germline mutations of *ASXL1* and *ASXL3* have been implicated in the neurodevelopmental disorders Bohring-Opitz and Bainbridge-Ropers syndromes respectively [15,16,17].

We previously showed that CH/CMML-associated ASXL1-mutant proteins enhance BAP1 activity [13]. Here we show that C-terminally truncated ASXL1 mutant proteins also interact with the EHMT1-EHMT2 (GLP-G9a) H3K9 methyltransferase complex, which is responsible for generation of the H3K9me1 and me2 marks [18,19]. Using a knock-in mouse model of CH (*Asxl1tm/*+ mice [14]), we show that expanded myeloid cells from old mice (≥18-month-old) display genome-wide decreases of H3K9me2, H3K9me3 and H2AK119Ub and substantial changes in gene expression compared to cells from age-matched wildtype mice; cells from young (4-month-old) mice showed no appreciable differences. Compared to control LK (lineage-negative, c-Kit+) cell populations enriched for hematopoietic stem/precursor cells (HSPC), LK cells from aged *Asxl1tm/*+ mice showed increased expression of transposable elements (TEs) and satellite repeats, indicating perturbed heterochromatin integrity. Increased TE and gene expression in heterochromatin were also observed in bone marrow samples from *ASXL1*-mutant CMML patients compared to healthy controls. In a small number of cases in which individual TEs could be mapped uniquely to the genome, increased TE expression was shown to alter the expression of nearby genes in *cis*, and was associated with altered expression of inflammation-associated and interferon-inducible genes (ISGs).

In summary, we find that mutant ASXL1 compromises heterochromatin integrity in an age-dependent manner by decreasing H3K9me2/3 levels in constitutive heterochromatin and H2AK119Ub levels in facultative heterochromatin. We propose that the top three mutations in clonal hematopoiesis share the common feature of promoting heterochromatin dysfunction, through reduction of heterochromatic DNA methylation in *TET2*- and *DNMT3A*-mutant cells [10,20] and decreased levels of H3K9me2/3 and H2AK119Ub histone modification in *ASXL1*-mutant cells.

Results

Cancer-associated ASXL1 mutant proteins bind heterochromatin-associated proteins

The heterozygous *ASXL1* and *ASXL3* mutations observed in CH, CMML, Bohring-Opitz and Bainbridge-Ropers syndrome invariably encode C-terminally truncated proteins [15,16,17,21,22]. Essentially all frameshift and nonsense mutations occur in the large penultimate exons of *ASXL* genes, thus all ASXL-mutant proteins retain the DEUBAD deubiquitinase adapter domain (aa 240-390 of human ASXL1) (shown for *ASXL1* in **Fig. 1a**, *top* and **Suppl. Figure S1**). The frameshift mutation c.1934dupG; p.G646WfsX12 (here termed ASXL1G646fs) is one of the most common, both in pre-neoplastic CH samples and in myeloid neoplasms (MPN) including the myelodysplastic/ myeloproliferative overlap neoplasm CMML [21,22] (**Suppl. Figure S1; Fig. 1a**, *top*).

To understand the association of mutant ASXL1 with CH, we identified proteins interacting with mutant ASXL1 by mass spectrometry. Full-length FLAG-tagged ASXL1 ectopically expressed in HEK 293T cells was both cytoplasmic and nuclear (**Fig. 1a**, *top panel*) as previously reported [23], whereas two distinct C-terminal truncation mutants of human ASXL1 (ASXL1G646fs and ASXL1Y591*) were predominantly present in the nucleus (**Fig. 1a**, *middle and bottom panels*). Because our *ASXL1*-mutant mouse model (*Asxl1tm/*+) bears the ASXL1G646fs mutation [14], we used the ASXL1G646fs mutant protein to identify interacting proteins by mass spectrometry.

Mutant ASXL1G646fs is more stable than full-length ASXL1 and the expression of both mutant and full-length ASXL1 is further stabilized by co-expression of BAP1 [23]. Previous mass spectrometry experiments were performed after immunoprecipitation of ASXL1 that was ectopically expressed in the absence of BAP1, leading to results that were variable and in some cases contradictory between different experiments [23,24,25,26,27,28,29]. We reasoned that this might be due to low or variable nuclear expression of ASXL1 proteins that were not stably folded in the absence of BAP1. To avoid this problem and ensure high levels of a properly configured ASXL1G646fs-BAP1 complex in the nucleus, we expressed both BAP1 and FLAG-tagged ASXL1G646fs in HEK293T cells, then co-immunoprecipitated (co-IP'd) interacting proteins from nuclear extracts (**Suppl. Fig. S2a-c; Suppl. Table 1**).

Mass spectrometric analysis (**Fig. 1b**) revealed several known ASXL1 interactors (BAP1, HCF1, OGT, FOXK1, FOXK2, BRD4) [23,24,25,26,27,28,29], as well as several proteins whose functions are related to heterochromatin. Among these were the H3K9 methyltransferase EHMT1 (GLP), which forms a tight dimer with EHMT2 (G9a) to generate

H3K9me1 and me2 [18,19]; HP1γ (heterochromatin-associated protein-1 gamma), one of three conserved HP1 proteins that bind H3K9me3 [8,30]; and NPM1 (nucleophosmin), a nucleolar protein that is recurrently mutated in AML such that it loses its nucleolar localization signal and becomes localized to the cytoplasm [31]. We focused on EHMT1 and EHMT2 which possess enzymatic activity and contribute to heterochromatin integrity by generating H3K9me1 and me2.

Co-IP data showed that ASXL1-FOXK1 and ASXL1-FOXK2 interactions were strengthened if BAP1 was co-expressed (**Fig. 1c, 1d**, *top panel*, *compare lanes 2, 4 to lanes 3, 5*). In contrast, the interaction of ASXL1G646fs with EHMT1 and EHMT2 was relatively insensitive to co-expression of BAP1 (**Fig. 1d,** *top panels, compare lanes 4, 5 in*); this was also true for its interactions with NPM1 and HP1γ (**Suppl Fig S2d, S2e**). The data suggest that BAP1 is not a stable component of the ASXL1G646fs-EHMT1-EHMT2 complex, but stabilizes and increases the nuclear expression level of mutant ASXL1. Indeed, EHMT1 and EHMT2 did not co-IP effectively with full-length ASXL1, potentially because of its low expression compared to mutant ASXL1G646fs (**Fig. 1d**). In a reciprocal experiment, however, V5-tagged EHMT2 effectively pulled down endogenous EHMT1, endogenous BAP1 and FLAG-tagged mutant as well as full-length ASXL1 from HEK293T cell nuclear lysates (**Fig. 1e**).

K562 cells express both full-length ASXL1 and the ASXL1Y591X mutant protein. Because of poor sensitivity of the commercially available antibodies to endogenous EHMT1, we were unable to co-IP the endogenous ASXL1Y591X with endogenous EHMT1/2 in K562 cell nuclear lysates. However, antibodies to endogenous EHMT2 did pull down both FLAG-tagged full-length ASXL1 and mutant ASXL1Y591X ectopically and stably expressed in K562 cells (**Suppl Fig S2f**).

These findings suggest that ASXL1 can serve as a component of two distinct protein complexes, one containing BAP1, HCF1, OGT and FOXK1/2 and the other containing EHMT1 and EHMT2. The ASXL1-EHMT1- EHMT2 complex may form preferentially with mutant ASXL1, suggesting that loss of the ASXL1 C-terminal region facilitates the efficient interaction of ASXL1 with the EHMT1/2 dimer.

Age-dependent changes in hematopoietic differentiation and gene expression in ASXL1 mutant mice

We next showed that the heterozygous *Asxl1tm/+* mouse strain is a reliable model of age-related clonal hematopoiesis. The *Asxl1tm* mutation is the mouse correlate of the human G646WfsX12 frame-shift mutation "knocked-in" to the *Asxl1* locus (**Fig. 2a**). The G646WfsX12 mutant protein strongly potentiates the ability of BAP1 to deubiquitinate H2AK119Ub[13]. To mirror CH in humans in which *DNMT3A*, *TET2* and *ASXL1* mutations are initially always heterozygous, we examined heterozygous *Asxl1tm/+* knock-in mice. Consistent with the previous report [14], we did not observe any change in the frequencies of hematopoietic cell subsets in bone marrow of young (4-month-old) heterozygous *Asxl1tm1/+* knock-in mice compared to young WT mice; however, we did observe significantly increased frequencies of $CD11b^+$ $Gr\text{-}1^+$ myeloid cells (**Figs. 2b, 2c**) and long-term hematopoietic stem cells (LT-HSC, **Figs. 2d, 2e**), in bone marrow cells from older *Asxl1tm1/+* mice (18- to 20-month-old) compared to age-matched WT controls. There was also a significant decrease in the frequencies of multipotent myeloid progenitor cells (MPP, **Figs. 2d, 2e**) and $B220^+$ B cells, pro-B cells ($B220^+CD43^+$), pre-B cells ($B220^+CD43^-$) and $Ter119^+$ erythroid cells (**Suppl. Fig S3a-S3c**), in old *Asxl1tm/+* mice. The frequencies of common myeloid progenitors (CMP), megakaryocyte-erythrocyte progenitors (MEP) and granulocyte-monocyte progenitors (GMP) were not appreciably altered in either young or old mice (**Suppl. Fig S2d**).

Consistent with these observations, there were almost no changes in gene expression in $CD11b^+$ myeloid cells from 4-month-old *Asxl1tm/+* vs WT mice) (2 upregulated and 2 downregulated genes based on criteria of log2 fold-change ($\leq$1 or >1and p-value <0.05; **Suppl. Fig S4a**). Even in older 18-month-old mice, only 40 genes were upregulated and 20 genes were downregulated in *Asxl1tm/+* compared to WT $CD11b^+$ cells (**Suppl. Fig S4b**), and fewer than 200 genes were downregulated or upregulated in LK cells (which are enriched in HSPC) from the same mice (*not shown*). However, there were clear changes in the expression of interferon-stimulated genes (ISGs) [32] ([32], **Suppl Table 2**) in old *ASXL1*-mutant $CD11b^+$ cells, including *Ptgs2* (encoding prostaglandin-endoperoxide synthase 2, also known as COX2, the inducible isoform of cyclooxygenase), and *Aldh1a1* and *Aldh1a7* (encoding aldehyde dehydrogenases), and *Hsp1a* and *Hsp1b* (encoding heat shock proteins) (**Suppl. Fig S4d, S4e, S4g**). These findings suggest dysregulation of interferon signaling in old compared to young mutant *Asxl1tm/+* $CD11b^+$ cells.

We reanalyzed RNA-seq data from long-term hematopoietic stem cells (LT-HSCs) from mice in which a different *ASXL1* mutation (ASXL1 p.E635RfsX15), was inducibly expressed from the *Rosa26* locus [33]. Again, the gene expression differences from age-matched controls were more striking in old compared to young mice, and there were marked changes in ISG expression (**Suppl. Fig S4d, S4e**). Notably, however, there was little or no overlap in either study (ours and [33]) between genes differentially expressed in young versus old cells, and the few genes whose

expression was altered in common in young and old cells did not fall into any obvious functional categories (**Suppl. Fig S4c, 4f**).

Together these data indicate that unlike conventional transcription factors, mutant ASXL1 proteins do not reproducibly regulate a common set of genes, and moreover that *Asxl1* mutations lead in an age-dependent manner to activation of inflammatory pathways.

Global decrease in H3K9me3 and H2AK119Ub in CD11b+ myeloid cells of old Asxl1tm1/+ mice

The EHMT1/EHMT2 dimer coordinately generates H3K9me1 and me2, and the H3K9me2 modification may serve as the substrate for deposition of H3K9me3 in heterochromatin by SUVARH1/2, SETDB1, and other H3K9me3 methyltransferases [8,34,35,36,37,38]. The interaction of mutant ASXL1G646fs with the heterochromatin-associated proteins EHMT1 and EHMT2 prompted us to ask whether *Asxl1tm/+* knock-in mice displayed age-related changes in the genome-wide distribution of the histone modifications H3K9me2 and me3. We also examined the distribution of H2AK119Ub in *Asxl1tm/+* myeloid cells. H2AK119Ub, which is deposited by the RING1A and RING1B subunits of PRC1 [8,12,39], and independently by BRCA1 [40,41], showed a greater decrease in cells expressing mutant ASXL1-BAP1 PR-DUB complexes compared to wildtype[13]. Lin$^-$c-Kit$^+$ (LK) cells from mice lacking ASXL1 were reported to show decreased levels of H3K27me3 [42]; similarly, LK cells from transgenic mice expressing mutant ASXL1 proteins were reported to have decreased levels of H3K4me3 [26]. We therefore examined the global levels of all these histone modifications by CUT&RUN for H3K9me2/3, H2AK119Ub, H3K27me3 and H3K4me3 in CD11b$^+$ cells from old (18-to 20-month-old) *Asxl1tm/+* compared to age-matched WT mice. We used *Drosophila* chromatin as a "spike-in" control, and divided the genome into euchromatic (Hi-C A) and heterochromatic (Hi-C B) regions by drawing on our previous Hi-C data from CD11b$^+$ cells [43].

As expected, H3K9me2 and H3K9me3 adorned broad regions of constitutive heterochromatin (**Fig. 3a, 3b; Suppl. Fig. S5a**), while H3K4me3 formed sharp peaks at transcribed gene promoters in euchromatin (**Fig. 3c; Suppl. Fig. S5b**). H2AK119Ub and H3K27me3, histone modifications deposited by PRC1 and PRC2 respectively, are present in facultative heterochromatin, a sub-compartment of euchromatin (Hi-C A) containing genes that are not expressed but are poised to be expressed during cell differentiation or cell activation in response to external signals (**Fig. 3a, 3b; Suppl. Fig. S5a, S5b**). Both modifications are broadly present across the gene body as well as 5' and 3' of genes, with the difference that H2AK119Ub marks both expressed and non-expressed genes whereas H3K27me3 primarily marks non-expressed genes [13].

In CD11b$^+$ cells of 18- to 20-month-old *Asxl1tm/+* mice compared to age-matched WT mice, we observed significant genome-wide reductions of H2AK119Ub, H3K9me2 and H3K9me3 levels (**Figs. 3a, 3b, Suppl. Fig. S5a**). H3K4me3 levels were moderately reduced while H3K27me3 levels were not appreciably changed (**Fig. 3c**, **Suppl. Fig. S5b**). Western blotting of acid-extracted histones (a less sensitive technique) confirmed a perceptible reduction of the global levels of H2AK119Ub and H3K9me3 in bone marrow cells and CD11b$^+$ cells of old *Asxl1tm/+* compared to age-matched WT mice (**Figs. 3d, 3e**). In contrast, western blotting of acid-extracted histones from bone marrow cells of young (4-month-old) *Asxl1tm/+* mice showed no differences in H2AK119Ub, H3K9me2 and H3K9me3 levels compared to age-matched WT mice (**Suppl. Fig. S5c**). In the genome browser views and western blots shown, there is less reduction of H3K9me2 than H3K9me3 levels (**Fig. 3a, 3d, 3e**, **Suppl. Figs. S5a, S5c**); however, analysis of 10 kb windows, assigned to genome compartments using Hi-C data, showed a decrease in the average levels as well as a tightening of the distribution of both H3K9me2 and H3K9me3 in heterochromatin (**Fig. 3b,** violin plots).

We also tested OCI-AML2 cells transduced with full-length wildtype or mutant (MT) ASXL1 Y591X, and with GFP as a control. OCI-AML2 cells have no baseline mutations in *ASXL1* or *TET2*, but western blotting of acid-extracted histones from the transduced cells showed a substantial reduction of both H3K9me3 and H2AK119Ub, with no obvious reduction in H3K9me2; both effects were intensified when BAP1 was co-expressed (**Suppl. Fig. S5d**).

Taken together, these data show that the global reduction in H2AK119Ub, H3K9me2 and H3K9me3 levels occurs progressively with age, in parallel with the observed increases in myeloid cell frequencies, the dysregulation of other hematopoietic lineages and the alterations in differential gene expression with age (**Fig. 2, Suppl. Figs. S3, S4**).

Old *Asxl1tm/+* mice display increased expression of transposable and repetitive elements, increased levels of R-loops and DNA damage

Decreased H2AK119Ub and H3K9me2/3 levels in heterochromatin are associated with increased expression of transposable elements (TEs) and satellite repeats that are normally highly suppressed in healthy cells [8, 34,35,36,37, 39,40]. We compared TE expression in CD11b$^+$ cells of young (4-month) and older (20-month) *Asxl1tm1/+* vs WT mice by

qRT-PCR as previously described [40], with DNase I treatment to eliminate DNA contamination. We observed a marked increase in the expression of LINE-1 (L1) and satellite repeats between 4 and 20 months (**Fig. 4a**). RNA-seq on total RNA from CD11b$^+$ cells of 18-month-old *Asxl1tm1/+* compared to WT mice revealed increased expression of an L1 subfamily as well as an ERVK family of endogenous retroviruses (RMER17C) (**Fig. 4b**). Increased TE expression was even more pronounced in LK cells (which are enriched for HSPC) of *Asxl1tm/+* mice, compared to LK cells from aged-matched wildtype control mice (**Fig. 4c, 4d**).

Cells with decreased H2K119Ub or decreased DNA or H3K9 methylation often display increased levels of R-loops and DNA damage [8,10,35,36,37,44]. In *C. elegans* lacking the SETDB1 and SUV39H homologs MET-2 and SET-25, R-loops were substantially enriched over repeat elements that were derepressed in the absence of H3K9me3 [36], while in humans and other species, a variety of genomic repeats have been reported to be enriched for R-loops [45,46]. Using recombinant V5-tagged catalytically inactive (D210N) RNASE H1 protein (dRNase H1) in flow cytometry [47] , we observed that R-loops levels increased significantly in CD11b$^+$ myeloid cells from young (4-month-old) and more strikingly in old (20-month-old) *Asxl1tm/+* compared to WT mice (**Suppl. Fig. 6a**). Increased R-loops are highly correlated with DNA damage [48,49,50]; specifically, excess R-loops are excised by the xeroderma pigmentosa proteins XPG and XPF, which can generate both single-strand and double-strand DNA breaks [49,50]. However, a significant increase in DNA damage (double-stranded DNA breaks), assessed by flow cytometry for γH2AX, was observed only in old *Asxl1tm/+* compared to WT mice (**Suppl. Fig. S6b**). While the correlation of increased TE and satellite repeat expression with increased R-loops and DNA damage is intriguing, new long-reads sequencing methods will be needed to ask if the genomic location of DNA damage associated with newly emerging R-loops can be connected directly to increased expression of specific uniquely mapped TEs and satellite repeats.

CMML patients show increased expression of interferon-stimulated genes (ISGs) and TEs located in heterochromatin

ASXL1 mutations that result in C-terminally truncated proteins are associated with age-related CH, as well as with CMML, MDS and AML [21,22]. In CMML, approximately 40% of patients have truncating *ASXL1* mutations and these mutations are associated with a proliferative phenotype, resistance to epigenetic therapies and shortened overall survival (<2 years) [51,52]. We reanalysed published RNA-seq data [51] from bone marrow mononuclear cells (primarily malignant monocytes) of *ASXL1*-mutant CMML patients; the variant allele frequencies of the *ASXL1* mutations ranged from 31-48% (**Fig. 5a; Suppl Figure S1**). A PCA plot showed that the gene expression patterns of healthy control patients clustered tightly together but were separated in the PC1 dimension from the gene expression patterns of CMML patients harboring ASXL1 mutations (**Fig. 5b**).

Since CMML is a highly inflammatory disease, with 30% of patients showing concomitant systemic inflammatory manifestations [21,53], we assessed inflammatory phenotypes by examining interferon-stimulated genes (ISGs), which include genes encoding IRF family members, pro-inflammatory cytokines (IL6, IL1β) and cytokine and chemokine receptors ([32], **Suppl Table 2**). Indeed, as in older myeloid cells from *Asxl1tm/+* mice (**Suppl. Fig. S4**), a substantial proportion of differentially expressed genes in CMML patients were ISGs (**Fig. 5a**). Notable among these were *PTGS2* (encoding COX2, the inducible cyclooxygenase isoform) and genes belonging to the (*ALDH*) aldehyde dehydrogenase family, including *ALDH1A3* and *ALDH16A1* (**Fig. 5a**; **Suppl. Fig. S8a, S8b**). As noted above, *Ptgs2* and related *Aldh*-family genes were also found upregulated in myeloid cells from old *Asxl1tm/+* mice (**Suppl. Fig S4b, S4g**). All three genes are located in heterochromatin (Hi-C B compartment), and low ALDH levels are associated with more favorable outcomes in AML [54].

As noted (**Suppl. Fig S4b, S4g**), *Ptgs2* and related *Aldh*-family genes were also found upregulated in myeloid cells from old *Asxl1tm/+* mice. All three genes are located in heterochromatin (i.e. are poorly expressed) in normal CD11b$^+$ cells, and low ALDH levels are associated with more favorable outcomes in AML [54].

We asked whether the bulk of gene expression changes in CMML patients were in euchromatin or heterochromatin. When we plotted the changes in expression of individual genes using available Hi-C data from normal human monocytes [55], most differential gene expression occurred in the gene-rich euchromatic compartment as expected (**Suppl Fig. S7a**). To examine expression of genes as well as TEs, we plotted differential RNA expression in 10 kb genomic windows (**Fig. 5c**). This analysis confirmed that most changes occurred in euchromatin and involved decreased RNA expression; in contrast in heterochromatin, more windows showed increased RNA expression, suggesting that CMML monocytes showed de-repression of normally silenced transcriptional units in this genomic compartment (**Fig. 5c, Suppl Fig. S7b, S7c**).

We confirmed this point by analyzing changes in TE expression in *ASXL1*-mutant CMML patients (**Fig. 5d**). Compared to healthy individuals (controls), *ASXL1*-mutant CMML patients showed a striking increase in the expression of many

TE subfamilies including L1s, LTRs (human endogenous retroviruses, HERVs) and DNA transposons as well as alpha (ALR-ASAT) and beta (LSAU) satellite repeats (**Fig. 5d**). Notably, these TE subfamilies and repeat classes are located primarily in heterochromatin (**Fig. 5e**). The data suggest that with more extensive analyses involving a broader spectrum of patients, it might be possible to use RNA-seq and TE expression to subclassify CMML for disease stratification or for prognostic or therapeutic purposes. HERVK LTRs were prominent among the upregulated TEs and their relevance to cancer immunotherapy has been noted previously: specifically, HERVK-int expression is potentially immunogenic [56], and expression of HERVK loci has been used to stratify lung cancer patients [57].

Since TEs can act as gene promoters or enhancers [9,58,59,60], we asked if the increased TE expression observed in CMML patients was accompanied by changes in the expression of nearby genes. Only a very few individual upregulated TEs – all of which were located in heterochromatin in healthy human monocytes – could be mapped uniquely using the available RNA-seq data (50x50 paired-end reads) [51] (**Suppl. Fig. S7d-f**). One of these uniquely-mapped elements – an evolutionarily young L1PA6 element [61] – was located 5' of the *PTH2R* gene, and its increased expression correlated with increased *cis*-expression of the *PTH2R* gene in *ASXL1*-mutant CMML (**Fig. 5f**). The clinical significance of increased *PTH2R* expression is not known, but the data are reminiscent of the limited heterochromatin-to-euchromatin switch we previously documented in expanded myeloid cells of inducible *Tet1/2/3*-deficient mice [43].

Together these findings suggest that ASXL1 is not a conventional transcriptional regulator that directly controls gene expression by binding to DNA or acting as a coactivator/corepressor for DNA-binding transcription factors. Rather, the decrease in histone modifications in facultative and constitutive heterochromatin, and the consequent increase of TE expression in *ASXL1*-mutant CMML, may drive stochastic increases in expression of nearby genes in *cis*. This hypothesis may explain why reproducible changes in gene expression are rarely observed in *ASXL1*-mutant CMML [51] or in different mouse models of ASXL1 mutations (**Suppl. Fig. S4**).

Discussion

Here, we show that CH/CMML-associated ASXL1 mutant proteins co-immunoprecipitate (co-IP) not only the known interactors BAP1, HCF1, OGT and FOXK1/2, but also a previously undescribed set of heterochromatin-related proteins (**Suppl. Fig. S9a**). Notable among these are EHMT1 and EHMT2, a dimer of H3K9 methyltransferases that coordinately generate H3K9me2 [18,19]. Pairwise co-IP experiments (**Fig. 1d**) suggest, however, that BAP1 is not necessarily required for the ASXL1-EHMT1-EHMT2 interaction. Consistent with EHMT1/EHMT2 interaction, myeloid cells of *Asxl1tm/+* mice – a model of age-related CH (**Fig. 2**) – showed a genome-wide decrease in H3K9me2 and me3 that is pronounced in old (18-20-month-old, **Fig. 3**) but not young (4-month-old) mice. Further studies will be needed to define the structures of the two postulated complexes containing mutant ASXL1 (**Suppl. Fig. S9a**) and to determine how mutant ASXL1 in the ASXL1-EHMT1-EHMT2 complex might inhibit EHMT1/EHMT2 activity. It will also be important to identify the H3K9me3 methyltransferases (e.g. SUVARH1/2, SETDB1/2), whose reduced activity is responsible for the observed reduction in H3K9me3 levels in *Asxl1tm/+* cells.

Mutant ASXL1G646fs also interacted with three other well-established heterochromatin-associated proteins, NPM1 (nucleophosmin), HP1γ and CENPV. The nucleolar protein NPM1 is localized to heterochromatin and recurrently mutated in AML; mutant NPM1 loses its C-terminal nucleolar localization signal and becomes localized to the cytoplasm [31,62]. Notably, ASXL1 and NPM1 mutations are mutually exclusive in AML [63], suggesting that they act in the same molecular pathway. Perinucleolar heterochromatin contains inactive rDNA arrays that can be pressed into service during periods of rapid cell growth [64], thus one function of ASXL1 may be to cooperate with NPM1 in this specialized heterochromatic compartment. HP1γ, also a heterochromatin-associated protein that binds H3K9me3, was listed among the interaction partners of ASXL1 E635RfsX15 and ASXL1 Y588X in previous studies but was not further pursued [26,27]. The role of the ASXL1-CENPV (centromere protein V) interaction is less clear, primarily because the functions of CENPV have not been extensively investigated. In HeLa cells, overexpression of CENPV leads to hypercondensation of pericentromeric heterochromatin and altered H3K9me3 distribution, whereas CENPV depletion correlates with a number of chromosomal abnormalities in metaphase [65]. Similarly, in mouse oocytes, CENPV is reported to be crucial for spindle formation, chromosome segregation and faithful meiosis [66]. The functions of mutant ASXL1 in perinucleolar and pericentromeric heterochromatin remain to be investigated.

Myeloid cells from older *Asxl1tm/+* mice also showed a genome-wide decrease in mono-ubiquitylated H2AK119Ub (**Fig. 3**), consistent with the well-established ability of C-terminally truncated ASXL1/2/3 proteins to potentiate BAP1 deubiquitinase activity [13] through the interaction of ASXL DEUBAD domains with the UCH-CC1 domains of BAP1 [67,68,69,70] (**Suppl. Fig. S9a**, *left*). H2AK119Ub is deposited not only by RING1/2 in the PRC1 complex [39] but also by the BRCA1-BARD1 RING heterodimer complex [28,29,40,41], although other lysine residues of H2A may also be modified

by this complex [41]. BAP1 (originally BRCA1-associated protein 1) counters ubiquitylation of cellular substrates by BRCA1-BARD1; BAP1 also removes ubiquitin from HCF1, the OGT partner that is part of the MLL H3K4 methyltransferase complex [71]. This may happen in the context of a presumed second complex that includes OGT-HCF1 and FOXK1/2 in addition to ASXL1-BAP1 (**Suppl. Fig. S9a**, *left*).

Our data prompt the unifying hypothesis that all three of the most frequent mutations in CH and haematopoietic malignancies – in *DNMT3A*, *TET2* and *ASXL1* – involve a progressive loss of heterochromatin integrity [8] (**Suppl. Fig. S9b**). TET deficiency in a variety of different cell types is associated with decreased DNA methylation in heterochromatin; likewise, DNMT3A deficiency results in decreased DNA methylation genome-wide and synergizes with TET2 deficiency in mice to induce more rapid dysregulation of myeloid phenotypes and reduced survival [10]. Moreover, reduction of DNA methylation or H3K9me2/me3 modifications in constitutive heterochromatin is associated with a variety of cellular and genomic aberrations – increased expression of TEs and other repetitive elements, increased R-loops, increased mutations and genome instability, and increased expression of genes related to inflammation [8,10,34,35,36,37,72]. Double *Tet2, Tet3*-deficient (*Tet2/3 DKO*) iNKT cells [10], *Tet2/3 DKO* B cells [47], triple Tet1/2/3-deficient (*Tet iTKO*) myeloid cells (López-Moyado, *unpubl.)* and ASXL1-mutant myeloid cells (*this study*) display increased expression of TEs and satellite repeats, increased R-loops and increased DNA damage. *Tet iTKO* mESC display chromosome missegregation and aneuploidies [73]; and *Tet2-/-* peritoneal macrophages [74], *Tet2-/-* bone-marrow-derived macrophages (BMDM) [75], *Tet iTKO* BMDM [76] and *DNMT3A*- or *TET2*-deficient human monocyte-derived macrophages [77] all show increased expression of pro-inflammatory cytokines, cytokine receptors and interferon-stimulated genes (ISGs). These changes are all manifestations of heterochromatin dysfunction [8,10,72].

ALDH-family and *PTGS2* genes were prominent among the ISGs upregulated both in our mouse model of CH in older *Asxl1tm/+* mice and in monocytes from CMML patients (**Suppl. Fig. S4b**; **Fig. 5a**). Several ALDH-family genes show high expression in AML, and low *ALDH1A1* expression correlates with improved survival in AML[54,78,]. Mouse LSK cells express high levels of *Aldh1a1* mRNA and ALDH1A1 protein compared to total bone marrow cells, and *Aldh1a1* and *Aldh3a1* together regulate hematopoietic (particularly B cell) lineage specification in mice [79]. However, following HSPC differentiation into mouse CD11b$^+$ cells or differentiated human monocytes, these genes become localized to heterochromatin, which contains the majority of L1s and LTRs (**Suppl. Fig. 8c**). Our data suggest that increased expression of these genes during age-dependent CH evolution in mice and CH to CMML progression in humans reflects their slow release from heterochromatin-mediated silencing, followed by positive selection for clones with increased expression of protein-coding genes and TEs in heterochromatin [80]. Indeed, we previously documented the occurrence of a limited heterochromatin-to-euchromatin switch in expanded *Tet iTKO* CD11b$^+$ myeloid cells, followed by expansion of new myeloid populations expressing higher levels of clustered, originally heterochromatin-localised, *stefin/cystatin* genes [43].

DNA methyltransferase inhibitors (DNMTi) such as azacitidine, decitabine and decitabine with the cytidine deaminase inhibitor cedazuridine are US FDA approved therapies for myeloid neoplasms such as MDS, CMML and AML [81]. The clinical improvement observed after treatment of responder patients with DNMTi does not appear to be due to preferential elimination of mutant HSPC clones (e.g. by the immune system), alteration in mutational allele burdens, or a decrease in the differentiation capability of mutant HSPC [82,83]; rather, it is thought that these agents likely act in the short term by increasing inflammation through “viral mimicry”, thereby potentiating anti-tumor responses [84,85,86]. Our data suggest that in the longer-term, therapy with DNMTI would be not improve outcomes for patients with MDS, AML or CMML: the slow, genome-wide, replication-dependent reduction of DNA methylation that accompanies such therapies might worsen heterochromatin dysfunction and disease progression. Rather, it would be ideal, if possible, to devise therapies that delay the slow loss of DNA and H3K9 methylation during CH to CMML progression.

Acknowledgements

We thank Steve Henikoff and Kami Ahmad (Fred Hutchinson) for reagents, protocols, suggestions and support regarding CUT&RUN experiments, and the Department of Laboratory Animal Care (DLAC) and the animal facility for excellent support. We also thank C. Kim, S Alarcon and Z. Mikulski and colleagues of the La Jolla Institute (LJI) Flow Cytometry, Next-Generation Sequencing and Microscopy Core Facilities for help with cell sorting, next-generation sequencing, and microscopy respectively. The FACSAria II Cell Sorter (S10 RR027366), the CyTOF Mass Cytometer (S10 RR027366), the NovaSeq 6000 (S10OD025052) and the HiSeq 2500 (S10 OD018499) were acquired through the Shared Instrumentation Grant (SIG) Program. This research was supported by NIH grants R35 CA210043 to A.R., R01 CA 272496 to M.P., NIH NIGMS R35GM147554 to S.A.M., NHMRC Investigator Grant (GNT1173711) and the Mater Foundation to G.J.F., NSC 112-2314-B-075A-004- and 112-2314-B-002-095-MY3 to W.-C.-C, and MOHW109-

TDU-B-211-134009 to W-C.C. and H.-F. T. H.S. was supported by the Pew Latin-American Fellows Program from The Pew Charitable Trusts, and by a Fellowship from the California Institute for Regenerative Medicine.

Author contributions

Z.D. designed and performed experiments, analysed data and prepared the sequencing libraries; H.S. and L.J.A. performed computational analyses of the RNA-seq and CUT&RUN data; C.B. assisted with cell isolation and flow cytometry; W-C.C. and H.F.T. provided the ASXL1 mutant knock-in mouse strain; G.J.F. advised on and interpreted the data on transposable element expression together with H.S., S.A.M. performed the mass spectrometry experiments and analyzed the data, J.F., M.B. and M.P. provided RNA-sequencing data for CMML samples and healthy controls. Z.D., H.S. and A.R. wrote the manuscript, with all authors contributing to writing, editing and providing feedback on the manuscript.

Competing interests

AR is a member of the Scientific Advisory Board of biomodal, formerly Cambridge Epigenetix. All other authors declare no competing interests.

References

1. Genovese, G. *et al.* Clonal Hematopoiesis and Blood-Cancer Risk Inferred from Blood DNA Sequence. *New England Journal of Medicine* **371**, 2477–2487 (2014).
2. Jaiswal, S. *et al.* Clonal Hematopoiesis and Risk of Atherosclerotic Cardiovascular Disease. *N Engl J Med* **377**, 111–121 (2017).
3. Evans, M. A., Sano, S. & Walsh, K. Cardiovascular Disease, Aging, and Clonal Hematopoiesis. *Annual Review of Pathology: Mechanisms of Disease* **15**, 419–438 (2020).
4. Jaiswal, S. & Libby, P. Clonal haematopoiesis: connecting ageing and inflammation in cardiovascular disease. *Nat Rev Cardiol* **17**, 137–144 (2020).
5. Bick, A. G. *et al.* Inherited causes of clonal haematopoiesis in 97,691 whole genomes. *Nature* **586**, 763–768 (2020).
6. Lio, C.-W. J., Yuita, H. & Rao, A. Dysregulation of the TET family of epigenetic regulators in lymphoid and myeloid malignancies. *Blood* **134**, 1487–1497 (2019).
7. Mattei, A. L., Bailly, N. & Meissner, A. DNA methylation: a historical perspective. *Trends in Genetics* **38**, 676–707 (2022).
8. Janssen, A., Colmenares, S. U. & Karpen, G. H. Heterochromatin: Guardian of the Genome. *Annual Review of Cell and Developmental Biology* **34**, 265–288 (2018).
9. Chuong, E. B., Rumi, M. A. K., Soares, M. J. & Baker, J. C. Endogenous retroviruses function as species-specific enhancer elements in the placenta. *Nat Genet* **45**, 325–329 (2013).
10. López-Moyado, I. F. *et al.* Paradoxical association of TET loss of function with genome-wide DNA hypomethylation. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 16933–16942 (2019).
11. Fisher, C. L., Randazzo, F., Humphries, R. K. & Brock, H. W. Characterization of Asxl1, a murine homolog of Additional sex combs, and analysis of the Asx-like gene family. *Gene* **369**, 109–118 (2006).
12. Scheuermann, J. C. *et al.* Histone H2A deubiquitinase activity of the Polycomb repressive complex PR-DUB. *Nature* **465**, 243–247 (2010).
13. Balasubramani, A. *et al.* Cancer-associated ASXL1 mutations may act as gain-of-function mutations of the ASXL1-BAP1 complex. *Nat Commun* **6**, 7307 (2015).
14. Hsu, Y.-C. *et al.* The distinct biological implications of Asxl1 mutation and its roles in leukemogenesis revealed by a knock-in mouse model. *Journal of Hematology & Oncology* **10**, 139 (2017).
15. Hoischen, A. *et al.* De novo nonsense mutations in ASXL1 cause Bohring-Opitz syndrome. *Nat Genet* **43**, 729–731 (2011).
16. Bainbridge, M. N. *et al.* De novo truncating mutations in ASXL3 are associated with a novel clinical phenotype with similarities to Bohring-Opitz syndrome. *Genome Medicine* **5**, 11 (2013).
17. Kuechler, A. *et al.* Bainbridge–Ropers syndrome caused by loss-of-function variants in ASXL3: a recognizable condition. *Eur J Hum Genet* **25**, 183–191 (2017).
18. Sanchez, N. A., Kallweit, L. M., Trnka, M. J., Clemmer, C. L. & Al-Sady, B. Heterodimerization of H3K9 histone methyltransferases G9a and GLP activates methyl reading and writing capabilities. *Journal of Biological Chemistry* **297**, 101276 (2021).
19. Tachibana, M. *et al.* Histone methyltransferases G9a and GLP form heteromeric complexes and are both crucial for methylation of euchromatin at H3-K9. *Genes Dev.* **19**, 815–826 (2005).
20. Zhang, X. *et al.* DNMT3A and TET2 compete and cooperate to repress lineage-specific transcription factors in hematopoietic stem cells. *Nat Genet* **48**, 1014–1023 (2016).
21. Patnaik, M. M. & Tefferi, A. Chronic Myelomonocytic leukemia: 2020 update on diagnosis, risk stratification and management. *Am J Hematol* **95**, 97–115 (2020).
22. Gelsi-Boyer, V. *et al.* Mutations in ASXL1 are associated with poor prognosis across the spectrum of malignant myeloid diseases. *Journal of Hematology & Oncology* **5**, 12 (2012).

23. Xia, Y.-K. *et al.* Tumor-derived neomorphic mutations in ASXL1 impairs the BAP1-ASXL1-FOXK1/K2 transcription network. *Protein Cell* **12**, 557–577 (2021).

24. Dey, A. *et al.* Loss of the Tumor Suppressor BAP1 Causes Myeloid Transformation. *Science* **337**, 1541–1546 (2012).

25. Li, Z. *et al.* ASXL1 interacts with the cohesin complex to maintain chromatid separation and gene expression for normal hematopoiesis. *Sci. Adv.* **3**, e1601602 (2017).

26. Inoue, D. *et al.* A novel ASXL1–OGT axis plays roles in H3K4 methylation and tumor suppression in myeloid malignancies. *Leukemia* **32**, 1327–1337 (2018).

27. Yang, H. *et al.* Gain of function of ASXL1 truncating protein in the pathogenesis of myeloid malignancies. *Blood* **131**, 328–341 (2018).

28. Wu, X. *et al.* Tumor suppressor ASXL1 is essential for the activation of INK4B expression in response to oncogene activity and anti-proliferative signals. *Cell Res* **25**, 1205–1218 (2015).

29. Kolovos, P. *et al.* PR-DUB maintains the expression of critical genes through FOXK1/2- and ASXL1/2/3-dependent recruitment to chromatin and H2AK119ub1 deubiquitination. *Genome Res.* **30**, 1119–1130 (2020).

30. Lomberk, G., Wallrath, L. & Urrutia, R. The Heterochromatin Protein 1 family. *Genome Biology* **7**, 228 (2006).

31. Falini, B., Brunetti, L., Sportoletti, P. & Martelli, M. P. NPM1-mutated acute myeloid leukemia: from bench to bedside. *Blood* **136**, 1707–1721 (2020).

32. Li, Q. *et al.* RNA editing underlies genetic risk of common inflammatory diseases. *Nature* **608**, 569–577 (2022).

33. Fujino, T. *et al.* Mutant ASXL1 induces age-related expansion of phenotypic hematopoietic stem cells through activation of Akt/mTOR pathway. *Nat Commun* **12**, 1826 (2021).

34. Bulut-Karslioglu, A. *et al.* Suv39h-Dependent H3K9me3 Marks Intact Retrotransposons and Silences LINE Elements in Mouse Embryonic Stem Cells. *Molecular Cell* **55**, 277–290 (2014).

35. Velazquez Camacho, O. *et al.* Major satellite repeat RNA stabilize heterochromatin retention of Suv39h enzymes by RNA-nucleosome association and RNA:DNA hybrid formation. *eLife* **6**, e25293 (2017).

36. Zeller, P. *et al.* Histone H3K9 methylation is dispensable for Caenorhabditis elegans development but suppresses RNA:DNA hybrid-associated repeat instability. *Nat Genet* **48**, 1385–1395 (2016).

37. Padeken, J., Methot, S. P. & Gasser, S. M. Establishment of H3K9-methylated heterochromatin and its functions in tissue differentiation and maintenance. *Nat Rev Mol Cell Biol* **23**, 623–640 (2022).

38. Schultz, D. C., Ayyanathan, K., Negorev, D., Maul, G. G. & Rauscher, F. J. SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes Dev.* **16**, 919–932 (2002).

39. Blackledge, N. P. & Klose, R. J. The molecular principles of gene regulation by Polycomb repressive complexes. *Nat Rev Mol Cell Biol* **22**, 815–833 (2021).

40. Zhu, Q. *et al.* BRCA1 tumour suppression occurs via heterochromatin-mediated silencing. *Nature* **477**, 179–184 (2011).

41. Kalb, R., Mallery, D. L., Larkin, C., Huang, J. T. J. & Hiom, K. BRCA1 Is a Histone-H2A-Specific Ubiquitin Ligase. *Cell Reports* **8**, 999–1005 (2014).

42. Abdel-Wahab, O. *et al.* Deletion of Asxl1 results in myelodysplasia and severe developmental defects in vivo. *Journal of Experimental Medicine* **210**, 2641–2659 (2013).

43. Yuita, H. *et al.* Inducible disruption of *Tet* genes results in myeloid malignancy, readthrough transcription, and a heterochromatin-to-euchromatin switch. *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2214824120 (2023).

44. Zhu, Q. *et al.* Heterochromatin-Encoded Satellite RNAs Induce Breast Cancer. *Molecular Cell* **70**, 842-853.e7 (2018).

45. Zeng, C., Onoguchi, M. & Hamada, M. Association analysis of repetitive elements and R-loop formation across species. *Mobile DNA* **12**, 3 (2021).

46. Nadel, J. *et al.* RNA:DNA hybrids in the human genome have distinctive nucleotide characteristics, chromatin composition, and transcriptional relationships. *Epigenetics & Chromatin* **8**, 46 (2015).

47. Shukla, V. *et al.* TET deficiency perturbs mature B cell homeostasis and promotes oncogenesis associated with accumulation of G-quadruplex and R-loop structures. *Nat Immunol* **23**, 99–108 (2022).

48. Santos-Pereira, J. M. & Aguilera, A. R loops: new modulators of genome dynamics and function. *Nat Rev Genet* **16**, 583–597 (2015).

49. Sollier, J. & Cimprich, K. A. Breaking bad: R-loops and genome integrity. *Trends in Cell Biology* **25**, 514–522 (2015).

50. Sollier, J. *et al.* Transcription-Coupled Nucleotide Excision Repair Factors Promote R-Loop-Induced Genome Instability. *Molecular Cell* **56**, 777–785 (2014).

51. Binder, M. *et al.* Oncogenic gene expression and epigenetic remodeling of cis-regulatory elements in ASXL1-mutant chronic myelomonocytic leukemia. *Nat Commun* **13**, 1434 (2022).

52. Carr, R. M. *et al.* RAS mutations drive proliferative chronic myelomonocytic leukemia via a KMT2A-PLK1 axis. *Nat Commun* **12**, 2901 (2021).

53. Zahid, M. F. *et al.* Spectrum of autoimmune diseases and systemic inflammatory syndromes in patients with chronic myelomonocytic leukemia. *Leukemia & Lymphoma* **58**, 1488–1493 (2017).

54. Dancik, G. M., Voutsas, I. F. & Vlahopoulos, S. Lower RNA expression of ALDH1A1 distinguishes the favorable risk group in acute myeloid leukemia. *Mol Biol Rep* **49**, 3321–3331 (2022).

55. Xia, Y. *et al.* Capturing 3D Chromatin Maps of Human Primary Monocytes: Insights From High-Resolution Hi-C. *Frontiers in Immunology* **13**, (2022).

56. Jansz, N. & Faulkner, G. J. Endogenous retroviruses in the origins and treatment of cancer. *Genome Biology* **22**, 147 (2021).

57. Ng, K. W. *et al.* Antibodies against endogenous retroviruses promote lung cancer immunotherapy. *Nature* **616**, 563–573 (2023).

58. Friedli, M. & Trono, D. The Developmental Control of Transposable Elements and the Evolution of Higher Species. *Annual Review of Cell and Developmental Biology* **31**, 429–451 (2015).

59. Faulkner, G. J. *et al.* The regulated retrotransposon transcriptome of mammalian cells. *Nat Genet* **41**, 563–571 (2009).

60. Frost, J. M. *et al.* Regulation of human trophoblast gene expression by endogenous retroviruses. *Nat Struct Mol Biol* **30**, 527–538 (2023).

61. Khan, H., Smit, A. & Boissinot, S. Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates. *Genome Res.* **16**, 78–87 (2006).

62. Zarka, J., Short, N. J., Kanagal-Shamanna, R. & Issa, G. C. Nucleophosmin 1 Mutations in Acute Myeloid Leukemia. *Genes* **11**, 649 (2020).

63. Carbuccia, N. *et al.* Mutual exclusion of ASXL1 and NPM1 mutations in a series of acute myeloid leukemias. *Leukemia* **24**, 469–473 (2010).

64. McStay, B. Nucleolar organizer regions: genomic ‘dark matter’ requiring illumination. *Genes Dev.* **30**, 1598–1610 (2016).

65. Tadeu, A. M. B. *et al.* CENP-V is required for centromere organization, chromosome alignment and cytokinesis. *The EMBO Journal* **27**, 2510–2522 (2008).

66. Nabi, D. *et al.* CENP-V is required for proper chromosome segregation through interaction with spindle microtubules in mouse oocytes. *Nat Commun* **12**, 6547 (2021).

67. Sahtoe, D. D., van Dijk, W. J., Ekkebus, R., Ovaa, H. & Sixma, T. K. BAP1/ASXL1 recruitment and activation for H2A deubiquitination. *Nat Commun* **7**, 10292 (2016).

68. De, I. *et al.* Structural Basis for the Activation of the Deubiquitinase Calypso by the Polycomb Protein ASX. *Structure* **27**, 528-536.e4 (2019).

69. Ge, W. *et al.* Basis of the H2AK119 specificity of the Polycomb repressive deubiquitinase. *Nature* **616**, 176–182 (2023).

70. Thomas, J. F. *et al.* Structural basis of histone H2A lysine 119 deubiquitination by Polycomb repressive deubiquitinase BAP1/ASXL1. *SCIENCE ADVANCES* (2023).

71. Eletr, Z. M. & Wilkinson, K. D. An Emerging Model for BAP1's Role in Regulating Cell Cycle Progression. *Cell Biochem Biophys* **60**, 3–11 (2011).

72. López-Moyado, I. F. et al. TET Enzymes in the Immune System: From DNA Demethylation to Immunotherapy, Inflammation, and Cancer. *Annual Review of Immunology, in press* (2024)

73. Georges, R. O. *et al.* Acute deletion of TET enzymes results in aneuploidy in mouse embryonic stem cells through decreased expression of Khdc3. *Nat Commun* **13**, 6230 (2022).

74. Cull, A. H., Snetsinger, B., Buckstein, R., Wells, R. A. & Rauh, M. J. Tet2 restrains inflammatory gene expression in macrophages. *Experimental Hematology* **55**, 56-70.e13 (2017).

75. Zhang, Q. *et al.* Tet2 is required to resolve inflammation by recruiting Hdac2 to specifically repress IL-6. *Nature* **525**, 389–393 (2015).

76. Onodera, A. *et al.* Roles of TET and TDG in DNA demethylation in proliferating and non-proliferating immune cells. *Genome Biology* **22**, 186 (2021).

77. Cobo, I. *et al.* DNA methyltransferase 3 alpha and TET methylcytosine dioxygenase 2 restrain mitochondrial DNA-mediated interferon signaling in macrophages. *Immunity* **55**, 1386-1401.e10 (2022).

78. Gasparetto, M. *et al.* Targeted therapy for a subset of acute myeloid leukemias that lack expression of aldehyde dehydrogenase 1A1. *Haematologica* **102**, 1054–1065 (2017).

79. Gasparetto, M. *et al.* Aldehyde dehydrogenases are regulators of hematopoietic stem cell numbers and B-cell development. *Experimental Hematology* **40**, 318-329.e2 (2012).

80. Steensma, D. P. *et al.* Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. *Blood* **126**, 9–16 (2015).

81. Stomper, J., Rotondo, J. C., Greve, G. & Lübbert, M. Hypomethylating agents (HMA) for the treatment of acute myeloid leukemia and myelodysplastic syndromes: mechanisms of resistance and novel HMA-based therapies. *Leukemia* **35**, 1873–1889 (2021).

82. Thoms, J. A. I. *et al.* Clinical Response to Azacytidine (AZA) Is Associated with Increased Contribution from Mutated Blood Progenitors: Insights from Single Cell Genotyping of Matched Stem/Progenitor and Mature Blood Cells from MDS/CMML Patients Pre- and Post-AZA Treatment. *Blood* **140**, 9751–9752 (2022).

83. Merlevede, J. *et al.* Mutation allele burden remains unchanged in chronic myelomonocytic leukaemia responding to hypomethylating agents. *Nat Commun* **7**, 10767 (2016).

84. Roulois, D. *et al.* DNA-Demethylating Agents Target Colorectal Cancer Cells by Inducing Viral Mimicry by Endogenous Transcripts. *Cell* **162**, 961–973 (2015).

85. Chiappinelli, K. B. *et al.* Inhibiting DNA Methylation Causes an Interferon Response in Cancer via dsRNA Including Endogenous Retroviruses. *Cell* **162**, 974–986 (2015).

86. Chen, R., Ishak, C. A. & De Carvalho, D. D. Endogenous Retroelements and the Viral Mimicry Response in Cancer Therapy and Cellular Homeostasis. *Cancer Discovery* **11**, 2707–2725 (2021).

Figure legends

Figure 1. C-terminally truncated ASXL1 proteins reside in the nucleus and recruit novel epigenetic regulators

- **a.** *Top*, Diagram of full-length (FL) ASXL1 and two C-terminally truncated ASXL1 mutant proteins encoded by ASXL1 mutations associated with CH/CMML (also see **Suppl. Fig. S1**). Full-length ASXL1 contains a predicted DNA binding (HTH) domain at its N-terminus, followed by the DEUBAD deubiquitinase adapter domain and a C-terminal atypical PHD Zinc-finger. A 3xFLAG epitope tag was engineered at the C terminus. *Bottom*, Subcellular localisation of FLAG-tagged full length and mutant ASXL1G646fs and Y591X proteins in HEK293T cells. Full-length ASXL1 is cytoplasmic in most cells, but three cells with nuclear ASXL1 are present at the right, whereas mutant ASXL1G646fs is nuclear in all cells.
- **b.** Proteins interacting with mutant ASXL1G646fs. HEK293T cells were co-transfected with 3xFLAG-tagged ASXL1G646fs plus BAP1 or with GFP as control, and mutant ASXL1-interacting proteins were co-immunoprecipitated (IP'd) with anti-FLAG antibody and identified by mass spectrometry. The proteins labelled in green in the volcano plot are proteins already reported to interact with ASXL1, while the proteins labelled in red are novel heterochromatin-related regulators identified in this study.
- **c.** FOXK1 and FOXK2 bind both full-length (fl) ASXL1 and mutant (mt) ASXL1G646fs and the binding is potentiated by BAP1. HEK293T cells were transfected with FLAG-tagged ASXL1 and ASXL1G646fs with or without BAP1. Nuclear fractions were isolated and immunoprecipitated with anti-FLAG. ASXL1-bound FOXK1, FOXK2 and BAP1 were detected with antibodies to the endogenous proteins.
- **d.** The H3K9 methyltransferases EHMT1 and EHMT2 co-IP more strongly with mutant (mt) ASXL1G646fs compared to full-length (FL, fl) ASXL1. HEK293T cells were transfected with FLAG-tagged ASXL1 or mutant ASXL1G646fs with or without BAP1. Nuclear fractions were isolated and immunoprecipitated with anti-FLAG. ASXL1-bound FOXK1, EHMT1 and EHMT2 were detected by immunoblotting (IB) using antibodies to the endogenous proteins.
- **e.** Reciprocal immunoprecipitation of V5-tagged EHMT2 with ASXL1. HEK293T cells were co-transfected with V5-EHMT2 and FLAG-tagged full-length (fl) or mutant (mt) ASXL1G646fs-FLAG (the FLAG tag is at the C-terminus), and EHMT2 bound proteins were detected with anti-FLAG, anti-EHMT1 and anti-BAP1.

Figure 2. Heterozygous expression of C-terminally truncated ASXL1G646fs at physiological levels in *Asxl1tm/+* knock-in mice results in myeloid expansion with age.

- **a.** Schematic illustration of the mouse *Asxl1tm* knock-in mutation corresponding to the human CH/CMML-associated ASXL1G646WfsX12 mutation (ASXL1G646fs). The "G" inserted into the mutated allele is indicated in red. The protein sequence is in capitals and the asterisk represents the stop codon generated as a result of the frameshift.
- **b.** Flow cytometry of myeloid (Gr-1$^+$/CD11b$^+$) cell populations in the bone marrow of young mice (4 months old) and old mice (18 months old).
- **c.** Quantification of flow cytometry in young and old mice (n=4 per genotype).
- **d.** The frequency of MPP (multipotent progenitor cells), short-term (ST) and long-term (LT) hematopoietic stem cells (HSC) in bone marrow cells of young and old mice. Lineage$^-$/c-kit$^+$/Scal$^+$ (LSK) cells were gated for this panel.
- **e.** Summary of the frequency of MPP, ST-HSCs (short-term HSCs), and LT-HSCs (long-term HSCs) in bone marrow from young and old mice.

Figure 3. Epigenomic alterations in myeloid cells from *ASXL1tm/+* mutant mice

- **a.** Genome browser views of CUT&RUN data for the indicated histone modifications in CD11b$^+$ myeloid cells from 18-month-old *Asxl1tm/+* mice. *Drosophila* S2 cells were used as spike-ins for normalization. A ~26 Mb region of chr. 5 is shown. Hi-C PC1 values were used to divide the genome into euchromatic and heterochromatic compartments (positive and negative PC1 values respectively). The data show global reduction of H3K9me2/3 in heterochromatin (*negative Hi-C PC1 values, dark blue*), as well as H2AK119Ub in euchromatic regions (*positive Hi-C PC1 values, green*). H2AK119Ub is present as broad regions in

facultative heterochromatin (annotated as being in the euchromatic (Hi-C A) compartment). H3K9me2 and me3 are characteristic heterochromatic marks that cover broad regions of the genome.

b. Violin plots showing the quantification of H2AK119ub, H3K9me2 and H3K9me3 enrichment at euchromatic and heterochromatic regions in aged *Asxl1tm/+* and wildtype mice. Signals from replicate experiments in each group were normalized to signals from *Drosophila* S2 cells that had been spiked in to the CUT&RUN reactions; IgG signals were then subtracted and the results were averaged.

c. Distribution of H3K4me3 and H3K27me3 signals in CD11b$^+$ cells from aged mice (20 months old). CUT&RUN was performed to capture the H3K4me3 and H3K27me3 signals in CD11b$^+$ cells and *Drosophila* S2 cells as the spike-in for normalization. IgG signals were subtracted.

d. Western blots of acid-extracted histones from bone marrow (BM) cells from aged WT and *Asxl1tm/+* mice (18 months old) for H2AK119Ub, H3K9me2 and H3K9me3 modifications.

e. Western blots of acid-extracted histones for H3K9me2 and H3K9me3 in CD11b$^+$ cells from old WT and *Asxl1tm/+* mice (20 months old).

Figure 4. Increased expression of transposable elements and satellite repeats in aged *ASXL1tm/+* mutant cells

a. Total RNA isolated from CD11b$^+$ cells from young (4 months old) and old (18 months old) *Asxl1 tm/+* and WT mice was treated with DNase I and expression of satellite repeats and TEs was assessed by quantitative RT-PCR (qRT-PCR). Note the increased expression of major and minor satellite repeats and LINE-1s in cells old but not young mice.

b. Volcano plot of RNA-seq data showing limited changes of transposable and repetitive elements in CD11b$^+$ cells from old mice. P<0.05, Log2FC+/-0.5.

c. Quantitative PCR of repetitive elements in LK (Lineage$^-$c-Kit$^+$) cells from 20 months-old mice. Note the more highly increased expression of TEs and major and minor satellite repeats in LK cells compared to CD11b+ cells (**a**) from old *Asxl1 tm/+* mice.

d. Volcano plot of RNA-seq data showing markedly increased expression of L1s in LK cells compared to CD11b+ cells (**b**) from old *Asxl1 tm/+* mice (18 months old). P<0.05, Log2FC+/-0.5.

Figure 5. Increased expression of Interferon-Stimulated Genes (ISGs) and transposable elements in monocytes from ASXL1-mutant CMML patients

a. Differential gene expression in cells from ASXL1-mutant CMML patients. The MA-plot highlights many differentially expressed genes (DEGs), with differentially expressed ISGs indicated as black dots. The numbers of DEGs and differentially expressed ISGs are indicated. Red dots represent genes expressed at high levels in ASXL1-mutant CMML samples while blue dots represent downregulated genes.

b. Principal Component Analysis (PCA) plot of RNA-seq in monocytes from ASXL1-mutant CMML patients bearing different mutations in addition to ASXL1.

c. Dot plot of gene expression (RNA-seq) in ASXL1-mutant CMML and healthy control samples in euchromatin (positive Hi-C PC1 values) and heterochromatin (negative Hi-C PC1 values). Each dot represents RNA transcripts arising within a 10 kb window of the genome.

d. Volcano plot of differentially expressed transposable elements (TEs) in monocytes from ASXL1-mutant CMML patients compared to healthy controls. The bulk expression of TE families was analyzed from Poly (A) RNA-seq data using the TE transcripts package. Each dot represents a specific TE-subfamily. Color codes show the different classes of TEs. Differentially expressed TEs were considered as log2 fold change (log2FC) of +/- 1 and a padj value <0.05 (calculated by adjusting the p-value with the false discovery rate [FDR]).

e. Distribution of TE subfamilies in euchromatin or heterochromatin in normal human monocytes (Hi-C data from published paper [55]). For each subfamily of TEs (represented as a single dot), we counted the number of individual TE copies located in euchromatin (positive Hi-C PC1 values) or heterochromatin (negative Hi-C PC1 values). The log2 ratio of the number of TE copies is plotted on the y-axis, while the total number of copies found in the genome (T2T) is shown on the x-axis. Enriched TEs were defined as log2 ratio > or < +/- 0.5.

f. Genome browser view showing that a uniquely mapped *L1PA6* element (red rectangle) is upregulated in monocytes from *ASXL1*-mutant CMML patients; a downstream gene, *PTH2R,* located in heterochromatin in normal human monocytes (dark blue), also displayed higher expression in *ASXL1*-mutant cells.

Supplementary figure legends

Supplementary Figure S1. Spectrum of ASXL1 mutations observed in CH and CMML. The frameshift mutation c.1934dupG; p.G646WfsX12 (here termed ASXL1G646fs) is by far the most common, both in pre-neoplastic CH (pre-myeloid samples, *top*) and in myeloproliferative neoplasms (MPN, *bottom*). This mutation yields a mutant ASXL1 protein containing the first 646 amino acids of ASXL1 plus 12 additional amino acids not found in wildtype ASXL1. Also depicted is the Y591X mutation, which is represented only in MPN but encodes a biochemically better-behaved protein, all of whose amino acids are present in full-length ASXL1. No CH/CMML *ASXL1* mutations occur in the DEUBAD deubiquitinase adapter domain (aa 240-390). Most mutations are encoded in the large penultimate exon (exon 12) of *ASXL1*.

Supplementary Figure S2. Mass spectrometry identifies novel heterochromatin-related proteins interacting with mutant ASXL1 G646fs.

- **a.** Work flow of immunoprecipitation followed by mass spectrometry (IP-MS) in HEK293T cells co-expressing FLAG-ASXL1G646fs and BAP1. Cells expressing GFP were used as control.
- **b.** Silver staining of nuclear proteins co-immunoprecipitated with ASXL1G646fs-FLAG using anti-FLAG M2 beads. ASXL1G646fs, BAP1, HCF1 and OGT are present in the immunoprecipitates. FLAG-tagged GFP is the control for IP.
- **c.** Western blot for OGT, HCF1 and BAP1 in immunoprecipitates of FLAG-tagged full-length ASXL1 or ASXL1G646Wfs from HEK293T nuclear extracts. FLAG-tagged GFP is the control for IP.
- **d.** Verification of the interaction between mutant ASXL1G646Wfs and NPM1 or HP1γ in 293T cells.
- **e.** Immunoprecipitation in K562 cells constitutively expressing FLAG-tagged full length ASXL1 and mutant ASXL1Y591X shows that HP1γ preferentially interacts with mutant ASXL1.
- **f.** Endogenous EHMT2 pulls down FLAG-tagged mutant ASXL1Y591X from K562 cells stably expressing this protein.

Supplementary Figure S3. Hematopoietic development in young and old mice.

- **a.** Flow cytometry to assess T cell ($CD3^+$) and B cell ($B220^+$) lymphoid lineages in bone marrow of young (4 months) and old (18-20 months) WT and *Asxl1tm/*+ mice.
- **b.** Flow cytometry and frequency of B cell lymphoid (B220/CD43) populations in the bone marrow of young (4 months) and old (18-20 months) WT and *Asxl1tm/*+ mice.
- **c.** Flow cytometry and frequency of erythroid (CD71/Ter-119) populations in bone marrow of young (4 months) and old (18-20 months) WT and *Asxl1tm/*+ mice.
- **d.** Flow cytometry of hematopoietic stem/progenitor cells (HSPC) in young (4 months) and old (18-20 months) WT and *Asxl1tm/*+ mice. Cell populations were gated on $Lineage^-$c-Kit^+ $ScaI^-$ (LK) cells. CMP, common multipotent progenitors; GMP, granulocyte-monocyte progenitors; MEP, megakaryocyte-erythrocyte progenitors.

Supplementary Figure S4. RNA-seq analysis of differentially expressed genes in $CD11b^+$ cells and LK cells from young and old *Asxl1tm/*+ mice.

- **a.** MA plot of RNA-seq of $CD11b^+$ cells from young *Asxl1tm/*+ mice (4 months old). There were only 2 upregulated genes and 2 down regulated genes.
- **b.** Differentially expressed genes in $CD11b^+$ cells from old *Asxl1tm/*+ mice (18 months). Red dots represent genes expressed at high levels in *Asxl1tm/*+ cells while blue dots represent downregulated genes. Some ISGs are upregulated, including *Aldh1a1*, *Aldh1a7*, *Hspa1a/b* and *Ptgs2*.
- **c.** Venn diagram of differentially expressed genes in $CD11b^+$ cells from young and old *Asxl1tm/*+ mice. No genes were upregulated in common in young and old *Asxl1tm/*+ $CD11b^+$ cells, while only one gene, the

Rps4y2 gene encoding the ribosomal protein S4 Y-linked 2, is expressed at a low level in both young and old *Asxl1tm/*+ CD11b$^+$ cells.

d. RNA-seq of genes in long-term hematopoietic stem cells (LT-HSCs) in young mice bearing the ASXL1 p.E635RfsX15 mutation. Reanalysis of RNA-seq data [33].

e. There were more differentially expressed genes in LT-HSC of old mice bearing the ASXL1 p.E635RfsX15 mutation. Reanalysis of RNA-seq data [33].

f. Overlap of differentially expressed genes in HSCs from young and old Asxl1 mutant mice (reanalysis of RNA-seq data [33]). Only 16 genes showed high expression in both young and old *Asxl1tm/*+ HSCs, while 4 genes were downregulated.

g. Normalized expression of ALDH-family genes upregulated in old (18 months) *Asxl1tm/*+ CD11b$^+$ cells (*this study*). Data are shown as Log_2CPM (counts per million reads) after normalization from RNA-seq data.

Supplementary Figure S5. Altered histone modifications in *ASXL1tm/*+ cells.

a. Genome browser view of CUT&RUN data for H2AK119ub, H3K9me2 and H3K9me3 enrichment at the *Arap2* and neighboring genes as an example in CD11b$^+$ cells from 18-month-old *Asxl1tm/*+ mice. *Drosophila* S2 cells were used as spike-ins for normalization. Hi-C PC1 values were used to divide the genome into euchromatic (Hi-C A, positive PC1 values) and heterochromatic (Hi-C B) compartments. The data show global reduction of H3K9me2/3 in heterochromatin (*dark blue*), as well as H2AK119Ub in euchromatic regions (*green*).

b. Genome browser view of H3K27me3 and H3K4me3 enrichment at euchromatin (*green*) and heterochromatin (*dark blue*) in CD11b$^+$ cells from 18-month-old mice. H3K27me3 and H3K4me3 are present as discrete peaks in euchromatin.

c. Western blots of acid-extracted histones from bone marrow cells from young WT and *Asxl1tm/*+ mice (4 months old) for H2K119ub, H3K9me2 and H3K9me3 modifications.

d. Western blots of of acid-extracted histones from OCI-AML2 cells ectopically expressing FLAG-ASXL1G646fs with or without BAP1 co-expression. Levels of the H2K119ub, H3K9me2 and H3K9me3 modifications were assessed.

Supplementary Figure S6. Increased R-loops and γH2Ax in CD11b$^+$ cells from old *Asxl1tm/*+ mice.

a. Flow cytometry and frequency of R-loops, assessed by staining using V5-tagged catalytically-dead RNase H1, in CD11b$^+$ cells from 4 months-old and 20 months-old WT and *Asxl1tm/*+ mice.

b. **b.** Flow cytometry and frequency of γH2Ax in CD11b$^+$ cells from 4 months-old and 20 months-old WT and *Asxl1tm/*+ mice.

Supplementary Figure S7. Increased expression of genes and TEs normally contained in heterochromatin in ASXL1-mutant CMML patients compared to healthy controls.

a. Dot plot showing the distribution of differentially expressed genes (RNA-seq) located in euchromatin (positive Hi-C PC1 values) and heterochromatin (negative Hi-C PC1 values) in monocytes from *ASXL1*-mutant CMML and healthy control samples Each dot represents a gene.

b-e. Genome browser views showing genes (**b, c**) and TEs (**d-g**) from healthy (*top*, blue) or *Asxl1*-mutant CMML (*bottom*, purple) samples. **d-e**, unique L1HS (**d**) and HERK-int (**e**) elements that are highly expressed in all *Asxl1*-mutant CMML samples.

f. Clusters of TEs in heterochromatin that displayed higher expression in *ASXL1*-mutant CMML samples compared to healthy controls. Hi-C data from normal monocytes was used for chromatin compartmentalization [55]. PC1 positive values (*green*) indicate euchromatin while negative values (*blue*) indicate heterochromatin.

Supplementary Figure S8. Relative expression of *ALDH*-family genes in *ASXL1*-mutant CMML and genome browser view of *Aldh1a1/a7* from RNA-seq in old CD11b$^+$ cells

a. Normalized expression of *ALDH*-family genes in *ASXL1*-mutant CMML. Data were presented as Log_2CPM (counts per million reads) after normalization from RNA-seq.

b. Normalized expression of *PTGS1* and *PTGS2* in *ASXL1*-mutant CMML.

c. Genome browser view shows that *Aldh1a1* and *Aldh1a7* localize at heterochromatin (defined by Hi-C PC1 values) in *WT* CD11b$^+$ cells (*this study*), and TEs are enriched at those regions.

Supplementary Figure S9. Working model for the role of mutant ASXL1.

a. Two distinct ASXL1 protein complexes act in facultative and constitutive heterochromatin to cause global reduction of H2AK119Ub and H3K9me3 respectively. In *Asxl1*-mutant cells, C-terminally truncated ASXL1 forms a canonical complex with BAP1, OGT, HCF1 and FOXK1/2, which reduces H2AK119ub modification at both euchromatin and facultative heterochromatin. Our study reveals the existence of a new predicted complex, in which ASXL1 mutant proteins interact with EHMT1/2 (histone methyltransferase), resulting in reduced H3K9me2/me3 at constitutive heterochromatin. Both changes are observed in clonal hematopoiesis caused by *Asxl1* mutations in our *Asxl1tm/+* mouse model.

b. The three most frequent mutations in clonal hematopoiesis (in *TET2*, *DNMT3A* and *ASXL1*) compromise heterochromatin integrity. Our previous studies showed that loss-of-function mutations of *TET2* and *DNMT3A* both resulted in decreased DNA methylation in heterochromatin. We show here that C-terminally truncated ASXL1 proteins decrease H3K9me2/me3 levels in heterochromatin. Thus, the top three mutations in clonal haematopoiesis contribute to heterochromatin dysfunction and consequent genome instability.

Methods

Mice

The *Asxl1tm* strain of mutant knock-in mice were generated in the laboratory of Dr. Hwei-Fang Tien. The cognate mouse mutation is predicted to be c.1925dupG; p.G643WfsX12, encoding 654 amino acids mimicking the most common form of human mutant ASXL1 protein p.G646WfsX12. Mice were housed in a pathogen-free animal facility at the La Jolla Institute. They were used according to protocols approved by the Institutional Animal Care and Use Committee (IACUC).

Plasmids

Mammalian cytomegalovirus (CMV) promoter (pCMV)-driven plasmids that encode full-length human ASXL1 and BAP1, C-terminally FLAG-tagged ASXL1 p.G646Wfs*12 and p.Y591X were obtained from Addgene. The MSCV-T2A-Puromycin vectors encoding 3xFLAG-tagged full-length ASXL1 and truncated ASXL1 were generated by PCR amplification followed by sub-cloning using Gibson assembly.

Flow cytometry

Antibody staining and flow cytometric analysis was performed as previously described [1]. Briefly, bone marrow cells were flushed out of femurs and tibiae in Hanks balanced salt solution (HBSS) containing 2% FBS and 10M HEPES. Red blood cells were depleted prior to flow cytometry. Cells were stained with monoclonal antibodies in FACS buffer (PBS containing 3% heat-inactivated FBS, 2 mM EDTA and 0.1% (w/v) sodium azide. All antibodies were purchased from Biolegend and BD Bioscience. The following monoclonal antibodies were used: PerCP/Cy5.5 anti-mouse Cd11b, BV650 anti mouse Gr1, APC anti-mouse CD3ε, BV421 anti-mouse B220.

As a lineage cocktail, biotinylated antibodies against CD3ε (145-2C11), CD45R/B220 (RA3-6B2), Mac-1/CD11b (M1/70), Gr-1 (RB6-8C5), Ter-119 (Ter-119) (all from Mouse hematopoietic Lineage Biotin Panel, Biolegend) and IL-7Rα (A7R34, Biolegend) were used in combination with a Streptavidin-PerCP-Cy5.5 conjugate (Biolegend). Live/Dead eFluor 780 (Invitrogen) was used to exclude dead cells. Flow cytometric analyses were performed using FACS Celesta (BD Biosciences) and data were analyzed using FlowJo software. $Cd11b^+$ ($Mac\text{-}1^+$) cells were isolated from unfractionated bone marrow using Cd11b MicroBeads (Miltenyi Biotec) following the manufacturers' instructions.

Immunoprecipitation and mass spectrometry

For sample generation, nuclear extracts were prepared from approximately 70 million HEK293T cells transiently? transfected with untagged BAP1 and mutant ASXL1G646fs bearing a C-terminal 3xFLAG-tag. The cells were washed twice with cold PBS, and then resuspended in hypotonic buffer (10 mM HEPES pH 7.9 [don't need brackets for pH], 1.5 mM $MgCl_2$, 10 mM KCl). After 15 min incubation on ice, cells were pelleted (650*g*, 5 min), and resuspended in isotonic buffer (10 mM Tris pH 7.5, 2 mM $MgCl_2$, 3 mM $CaCl_2$, 0.3 M sucrose). Cells were disrupted using a Dounce homogenizer, followed by centrifugation (10,000*g*, 20 min) to collect pellets containing nuclei. Pellets were resuspended in nuclear lysis buffer (10 mM Hepes pH 7.9, 400 mM NaCl, and 1 mM EDTA) supplemented with benzonase (500 U/ml), and incubated at 4ºC for 30 min on a rotator. Debris was removed by centrifugation at 13,000 rpm for 15 min, and the supernatant (nuclear fraction) was recovered. An equal volume of 2x conversion buffer (10 mM Tris-HCl pH 7.5, 280 mM NaCl, 1 mM EDTA, 1 mM EGTA, 0.2% Triton X-100, 10% glycerol) was added to the nuclear proteins. Anti-FLAG M2 magnetic beads (Sigma M8823) were washed with RIPA buffer before adding 1 µl beads/ mg protein to protein samples with added benzonase and inverting continuously at 4°C overnight. Reactions were carried out overnight at 4ºC and washed three times with wash buffer I (50 mM Tris HCl, pH 7.4, with 150 mM NaCl, 1 mM EDTA, and 1% TRITON- X-100) and wash buffer II (50 mM Tris-HCl pH 7.4, 150 mM NaCl). 5% of beads suspension was aliquoted and heated at 90ºC in 1× LDS sample buffers with 10% 2-mercaptoethanol to elute proteins. Immunoprecipitated proteins were analyzed using immunoblotting and silver staining. The remaining protein-beads complexes were used for LC-MS/MS as previously described [2]. Protein quantification was done using Pierce™ BCA Protein Assay Kits (Thermo Fisher Scientific 23225). Proteins were processed and analyzed as previously described with minor modifications [2]. Samples were analyzed on an Orbitrap Eclipse mass spectrometer (Thermo Fisher Scientific).

Mass Spectrometry data analysis

Raw data were processed using MaxQuant software [3] using the default settings and searched against the human UniProt database (12/2017) with common contaminant entries. The settings used for MaxQuant analysis were: enzymes set as LysC/P and Trypsin/P, with maximum of 2 missed cleavages; fixed modification was Carbamidomethyl

(Cys); variable modifications were Acetyl (Protein N-term) and Oxidation (Met); label-free quantification was used with a minimum ratio count of 2 and classic normalization; False Discovery Rate for both protein and peptide identification was 0.01. The 'Match between runs" feature was enabled. The proteinGroups file from the MaxQuant analysis was used as the input for all downstream analyses. Gene names that were not automatically assigned by MaxQuant were manually added from their Entrez protein IDs. Of the 2,617 entries in the proteinGroups file, we filtered out 33 entries identified as contaminants. Each of the three replicates within a condition were annotated. LFQ intensities were log2 transformed and proteins with less than 2 LFQ intensity values across both conditions were filtered out. Missing values were imputed with the minimum detected intensity. A two-tailed t-test with a p-value cutoff of 0.05 and FDR of 0.01 was conducted alongside the calculation of log2 fold changes for construction of the volcano plot.

Immunostaining

HEK293T cells were seeded on round coverslips treated with L-poly-lysine in 24-well plate and transfected with plasmids. Cells were fixed with 4% formaldehyde for 15 min at room temperature and blocked with 3% normal goat serum and in PBST (0.2% Triton-X 100 in 1X PBS) for 20 min at room temperature. Staining was carried out by incubating overnight with primary antibodies (1: 500) in PBST supplemented with 3% normal goat serum. The cells were rinsed three times with 1X PBS and incubated with fluorochrome-conjugated secondary antibodies (1:1,000) in PBST for 1 h at room temperature. The cells were rinsed three times with 1X PBS and coated with ProLong Gold antifade mounting medium with DAPI (4',6-diamidino-2-phenylindole, 100 ng/ml) (Life Technologies, P-36931) on slides. Images were obtained on an LSM 880 laser-scanning confocal microscope.

RNA extraction and qRT-PCR

Total RNA was isolated with RNeasy Plus Mini Kit or with Trizol following the manufacturers' instructions. To prevent contamination from genomic DNA, the purified RNA was subjected to DNase I treatment (1 U/10 ng RNA) for 30 minutes at 37°C, then the DNase I was inactivated by adding 5 mM EDTA and heating at 75°C for 10 minutes. RNA was concentrated using 5 μg linear acrylamide and 0.5 M ammonium acetate, followed by 100% ETOH to precipitate RNA by centrifugation at 14,000 rpm for 15 min at 4°C. cDNA was synthesized using SuperScript III reverse transcriptase, and qRT-PCR was performed using SYBR Green Master Mix on a StepOnePlus Real-time PCR system (Thermo Fisher Scientific). Gene expression was normalized to 18s rRNA. Primers for qRT-PCR are listed in Table S3.

RNA-seq

Total RNA was isolated from CD11b$^+$ cells and LK cells using RNeasy plus mini kit (Qiagen 74134) [source?]. RNA libraries were prepared using NEB UltraII RNA library Prep kit for Illumina (NEB) according to the manufacturer's instructions, and sequenced using Illumina NovaSeq 6000 to yield 30 million paired-end reads (50x50 bases) of per each RNA-seq library. For mice or human samples, reads were aligned against the mouse (mm10) or human (T2T) genomes respectively, using STAR [4][104], with the parameters --outFilterMultimapNmax 1, --outSAMtype BAM SortedByCoordinate –sjdbOverhang 100 to analyze gene expression. HT-Seq [5][105] was used to quantify the gene expression levels using the options htseq-count -s yes, -r pos, -a 10. Normalization and differential expression analyses were performed using DESEq2 [6][106], with the parameters fitType parametric, alpha 0.05 and using the Benjamini & Hochberg method. Differentially expressed genes (DEGs) between WT and mutant cells were defined as those with adjusted P value <0.05 and a fold change (log2FC) +/- 1. For visualization of the data, we generated tracks using Deeptools [7][107], with the option bamCoverage. All related plots were made using R-Studio [8] [108] and Integrative Genome Viewer (IGV) [9][109]. To measure the expression of Transposable Elements (TEs), we used the TE-transcript package [110] with the option -multi to quantify ambiguously mapped reads and then perform the differential expression (DE) analysis with DESEQ2 (FDR cutoff of p<0.05). Genes or TEs with less than 10 reads total were pre-filtered in all comparisons as an initial step. For TE analysis on mouse samples, we specifically used the mm10 repeatMasker version from TE-transcripts [10] [110].

Cleavage under target and release using nuclease (CUT&RUN)

CUT&RUN was performed as previously described[11]. All antibodies were used at 1:100 dilution. The spike-in Cut&Run was performed with a 1:20 ratio of *Drosophila* S2 cells to mouse cells in each reaction[12]. Briefly, cells were harvested as above and washed twice with PBS followed by two washes with wash buffer (20 mM Hepes-NaOH pH 7.5, 150 mM NaCl, 0.5 mM spermidine, and 1× protease phosphatase inhibitor). Cells and spike-in *Drosophila* S2 cells were then incubated with activated Con A beads at room temperature for 20 min, followed by incubation with the primary

antibody or IgG control overnight. Unbound antibody was removed by two washes with wash buffer, after which protein A/G–micrococcal nuclease (pA/G-MNase) was added at a final concentration of 700 ng/ml, and incubated for 1 hour at 4°C. Following incubation, the bound complexes were washed twice with wash buffer and placed in a cold metal block for 5 min to prechill the tubes. $CaCl_2$ was added to activate MNase at a final concentration of 2 mM and incubated for 30 min in a cold metal block on ice. The digestion reaction was neutralized by the addition of an equal volume of 2× STOP buffer (200 mM NaCl, 20 mM EDTA, 4 mM EGTA, 50 µg/ml ribonuclease A, and 40 µg/ml glycogen) followed by incubation at 37°C for 10 min to release cleaved DNA fragments. The protein-DNA complex was collected by centrifugation at 16,000*g* for 5 min. Digested DNA was purified via spin column according to the manufacturer's instructions. CUT&RUN DNA libraries for next-generation sequencing were generated using the NEBNext Ultra II DNA library Prep Kit following the manufacturer's instructions. Libraries were pooled in equal quantities and sequenced using Illumina HiSeq 2500 (Illumina) to produce ~30 million paired-end reads (50x50 bases) per library.

Paired-end reads were aligned against genome reference mm10 using Bowtie2 (v2.4.5) [13] using default parameters. Duplicated reads and unassembled contigs were removed using Picard. The number of reads aligning to mouse or Drosophila chromosomes were calculated using Samtools (v1.6) view -c -F 260. The ratio between the total number of reads mapping against mouse or Drosophila were used to estimate the normalization factors. bamCoverage was used to generate the bedGraph maps, applying the normalization factors. bedGraphToBigWig (UCSC) was used to generate the bigwig tracks. To remove the technical noise, IgG signals were removed from the from the CUT&RUN signals using bigwigCompare --operation subtract [7]. Average tracks per condition were estimated by using BigwigMerge and considering the number of samples. To measure genome-wide changes, we used a 10kb approach, using bedtools map. CUT&RUN 10Kb windows data was intersected with HiC data to determine the presence of each region in A or B-compartments. HiC dataset was reanalyzed from REF as described before REF. All related plots were made using R-Studio [8] and Integrative Genome Viewer (IGV).

Data Analysis and Statistical Methods

For quantitative real-time PCR, mRNA level was determined using the Ct method with normalization to the housekeeping gene 18S ribosomal RNA. For all statistical tests, the exact "n" values can be found in figure legends. All quantitative data are presented as mean ± standard deviation from at least three samples or experiments, unless otherwise stated. The significance was determined by two-tailed un-paired Student's t test, one-way ANOVA with Tukey's multiple comparison, two-way ANOVA with Tukey's multiple comparison. Otherwise, methods are stated in figure legends. Statistical significance is shown as: *, $P < 0.5$; **, $P < 0.01$; and ***, $P < 0.001$. Statistical analyses were performed using Graphpad PRISM software.

References

1. Balasubramani, A. *et al.* Cancer-associated ASXL1 mutations may act as gain-of-function mutations of the ASXL1-BAP1 complex. *Nat Commun* **6**, 7307 (2015).
2. Myers, S. A. *et al.* SOX2 O-GlcNAcylation alters its protein-protein interactions and genomic occupancy to modulate gene expression in pluripotent cells. *eLife* **5**, e10647 (2016).
3. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology* **26**, 1367–1372 (2008).
4. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
5. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
6. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, 550 (2014).
7. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research* **44**, W160–W165 (2016).
8. R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
9. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics* **14**, 178–192 (2013).

10. Jin, Y. & Hammell, M. Analysis of RNA-Seq Data Using TEtranscripts. in *Transcriptome Data Analysis: Methods and Protocols* (eds. Wang, Y. & Sun, M.) 153–167 (Springer New York, New York, NY, 2018). doi:10.1007/978-1-4939-7710-9_11.
11. Skene, P. J., Henikoff, J. G. & Henikoff, S. Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nature Protocols* **13**, 1006–1019 (2018).
12. Sarthy, J. F. *et al.* Histone deposition pathways determine the chromatin landscapes of H3.1 and H3.3 K27M oncohistones. *eLife* **9**, e61090 (2020).
13. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**, 357–359 (2012).

Figure 1

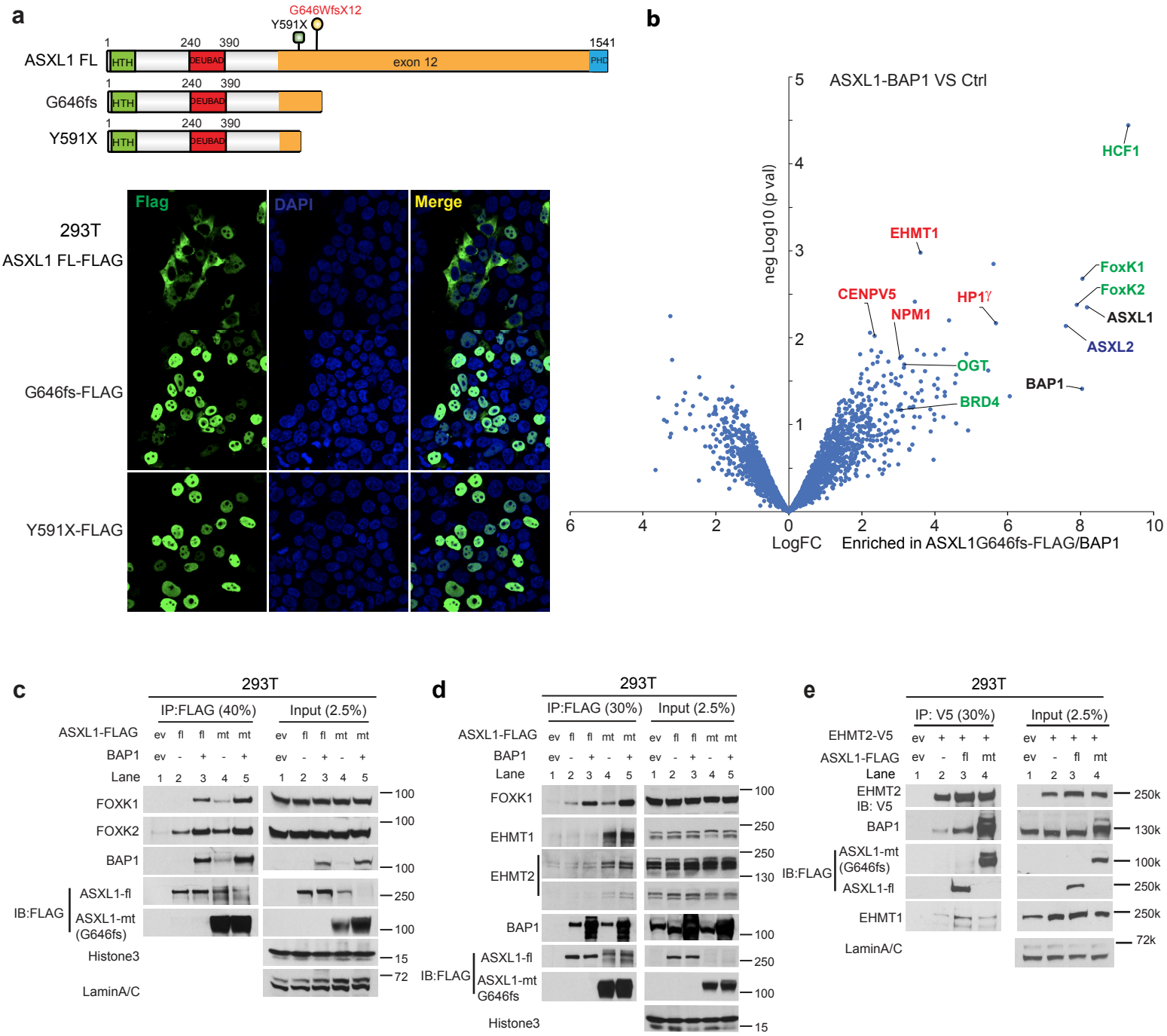

a
G646WfsX12
Y591X
ASXL1 FL
G646fs
Y591X
HTH
DEUBAD
exon 12
PHD
1
240
390
1541
293T
ASXL1 FL-FLAG
G646fs-FLAG
Y591X-FLAG
Flag
DAPI
Merge
b
ASXL1-BAP1 VS Ctrl
neg Log10 (p val)
LogFC
Enriched in ASXL1G646fs-FLAG/BAP1
HCF1
EHMT1
FoxK1
FoxK2
CENPV5
NPM1
HP1γ
ASXL1
ASXL2
OGT
BAP1
BRD4
c
293T
IP:FLAG (40%)
Input (2.5%)
ASXL1-FLAG
BAP1
Lane
FOXK1
FOXK2
BAP1
IB:FLAG
ASXL1-fl
ASXL1-mt (G646fs)
Histone3
LaminA/C
d
293T
IP:FLAG (30%)
Input (2.5%)
ASXL1-FLAG
BAP1
Lane
FOXK1
EHMT1
EHMT2
BAP1
IB:FLAG
ASXL1-fl
ASXL1-mt G646fs
Histone3
e
293T
IP: V5 (30%)
Input (2.5%)
EHMT2-V5
ASXL1-FLAG
Lane
EHMT2 IB: V5
BAP1
IB:FLAG
ASXL1-mt (G646fs)
ASXL1-fl
EHMT1
LaminA/C

Figure 2

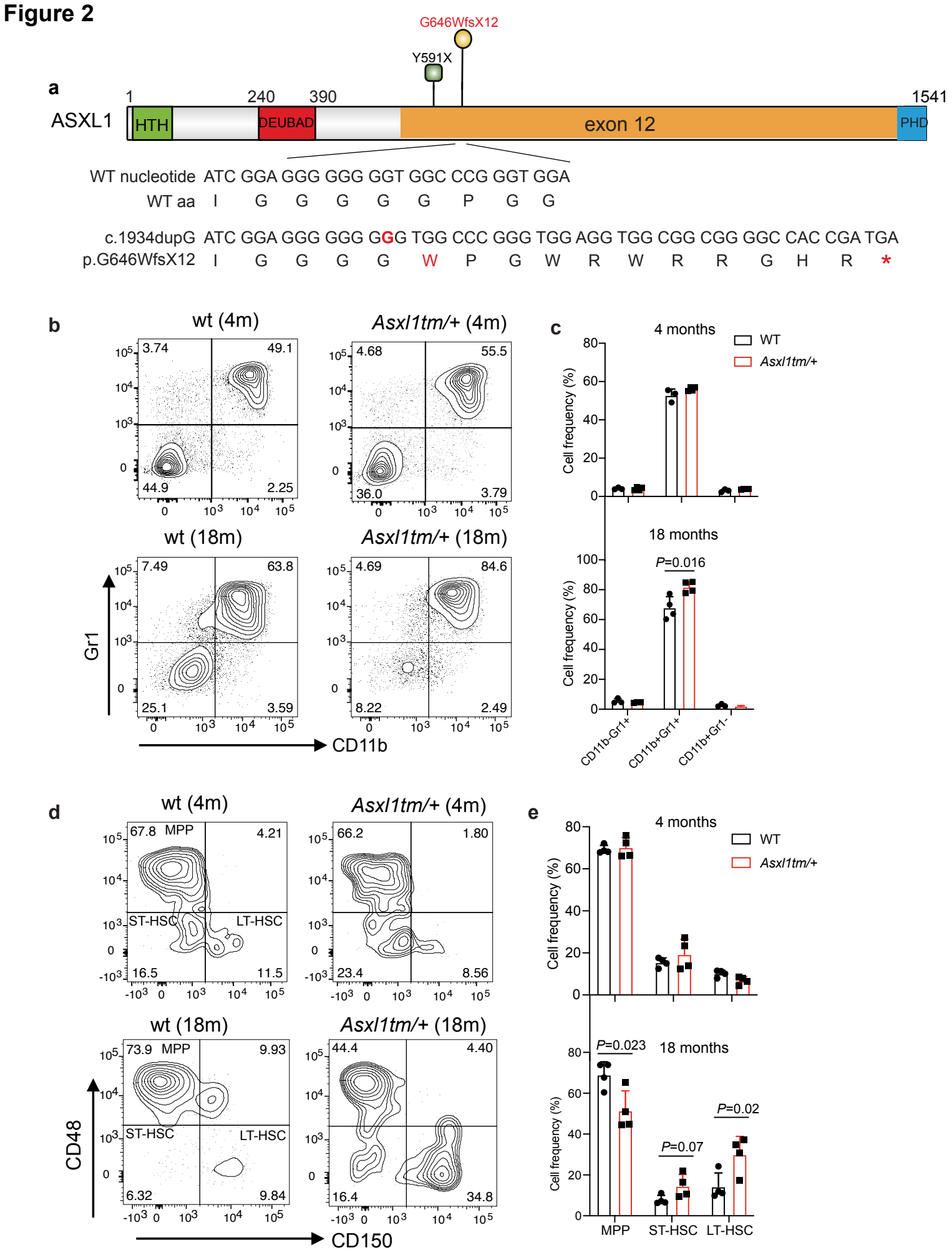

a
ASXL1
1
240
390
1541
HTH
DEUBAD
exon 12
PHD
Y591X
G646WfsX12
WT nucleotide ATC GGA GGG GGG GGT GGC CCG GGT GGA
WT aa I G G G G G P G G
c.1934dupG ATC GGA GGG GGG GGG TGG CCC GGG TGG AGG TGG CGG CGG GGC CAC CGA TGA
p.G646WfsX12 I G G G G W P G W R W R R G H R *
b
wt (4m)
Asxl1tm/+ (4m)
wt (18m)
Asxl1tm/+ (18m)
Gr1
CD11b
c
4 months
18 months
WT
Asxl1tm/+
Cell frequency (%)
P=0.016
CD11b-Gr1+
CD11b+Gr1+
CD11b+Gr1-
d
wt (4m)
Asxl1tm/+ (4m)
wt (18m)
Asxl1tm/+ (18m)
MPP
ST-HSC
LT-HSC
CD48
CD150
e
4 months
18 months
WT
Asxl1tm/+
Cell frequency (%)
P=0.023
P=0.07
P=0.02
MPP
ST-HSC
LT-HSC

Figure 3

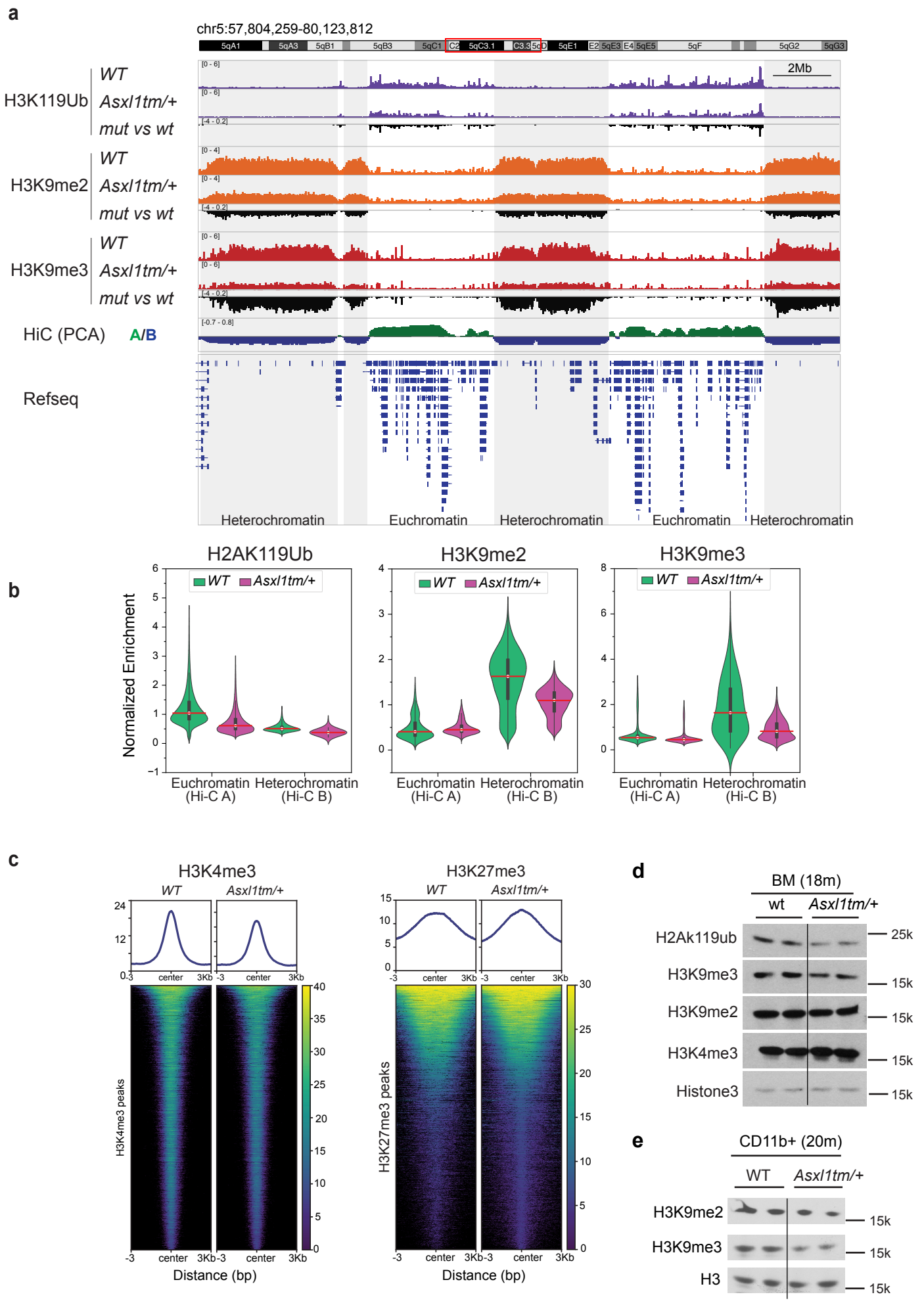

a
chr5:57,804,259-80,123,812
2Mb
H3K119Ub
WT
Asxl1tm/+
mut vs wt
H3K9me2
H3K9me3
HiC (PCA)
A/B
Refseq
Heterochromatin
Euchromatin
b
H2AK119Ub
H3K9me2
H3K9me3
WT
Asxl1tm/+
Normalized Enrichment
Euchromatin (Hi-C A)
Heterochromatin (Hi-C B)
c
H3K4me3
H3K27me3
H3K4me3 peaks
H3K27me3 peaks
Distance (bp)
d
BM (18m)
wt
H2Ak119ub
H3K9me3
H3K9me2
H3K4me3
Histone3
25k
15k
e
CD11b+ (20m)
WT
H3K9me2
H3K9me3
H3

Figure 4

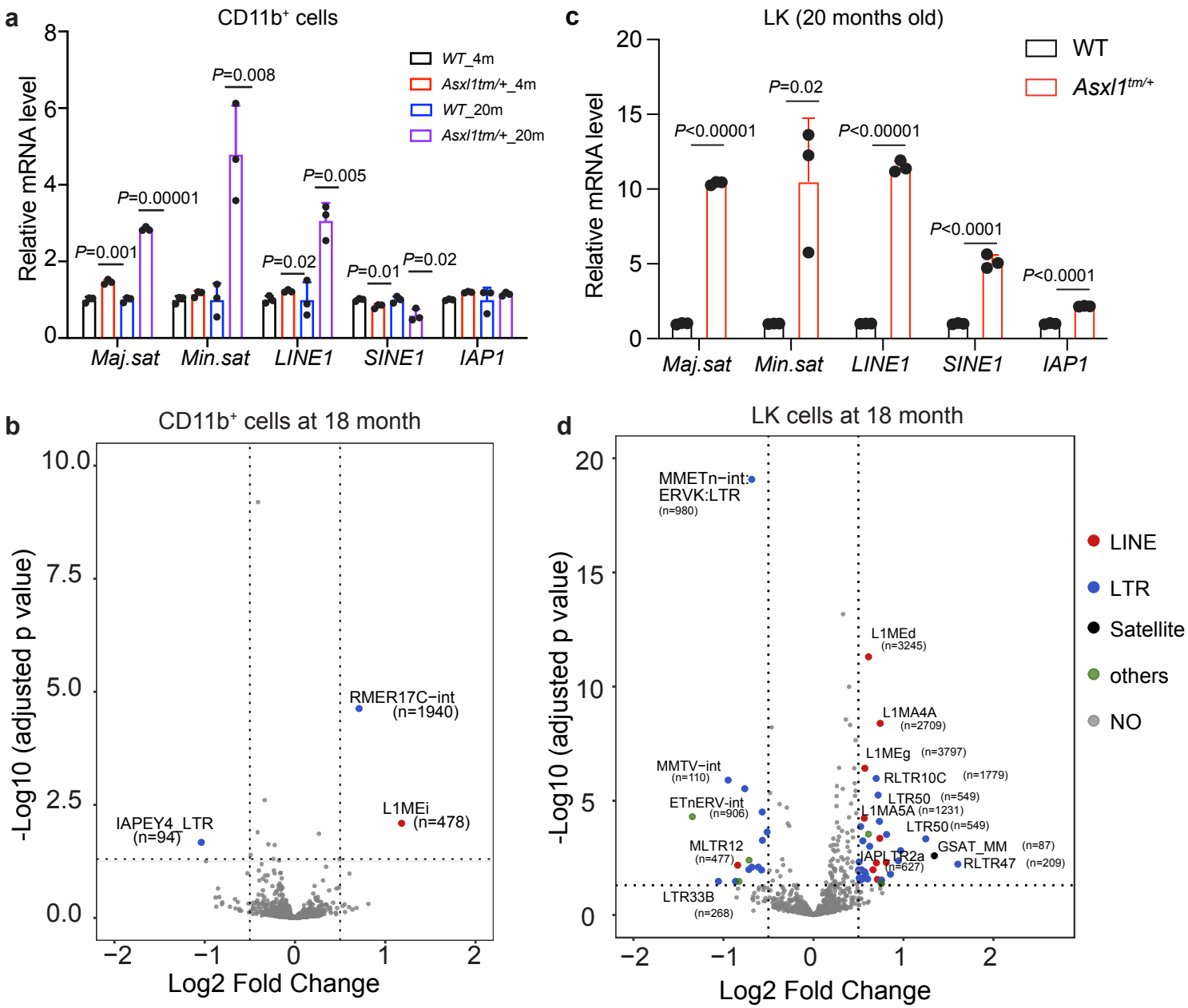
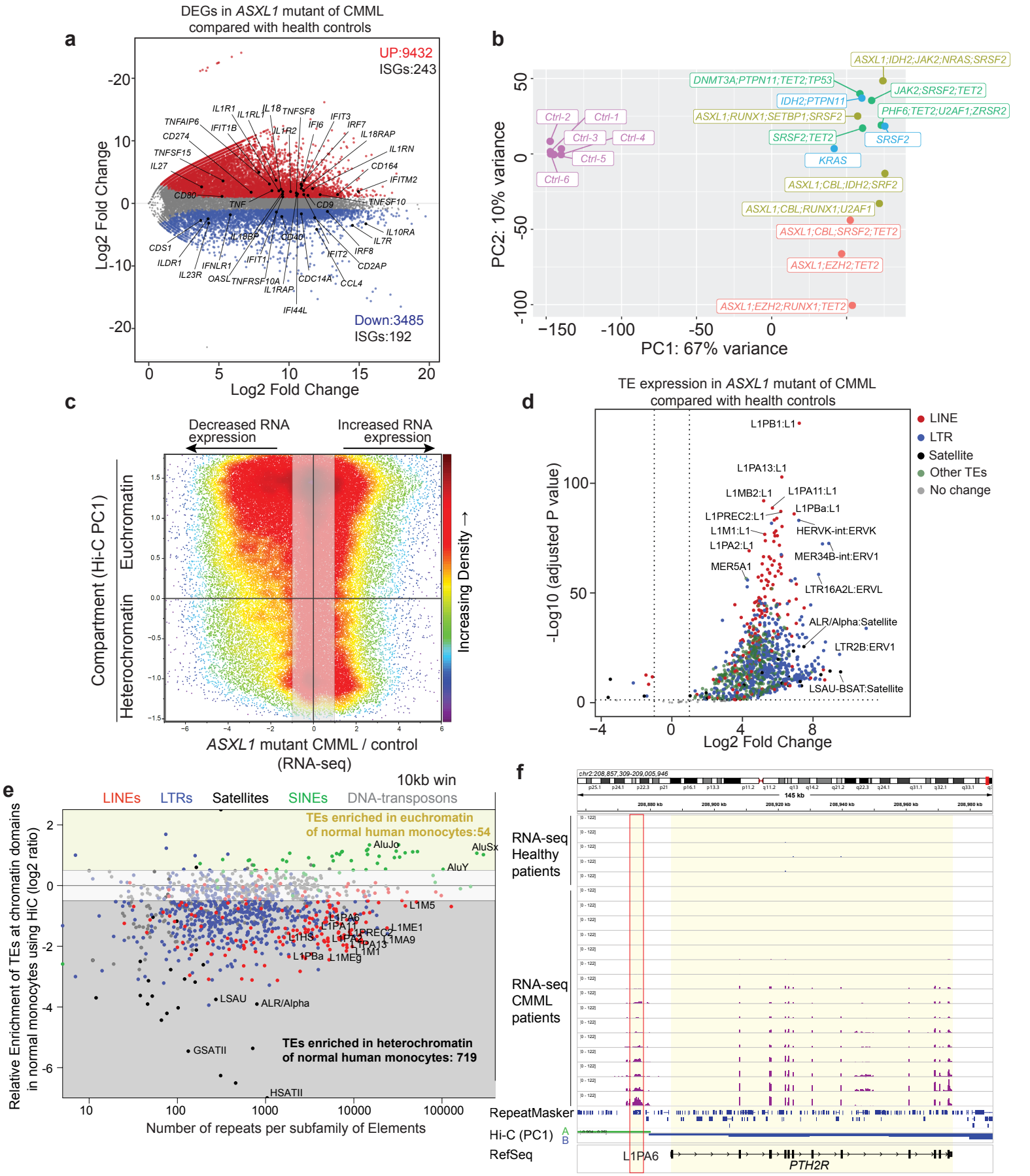
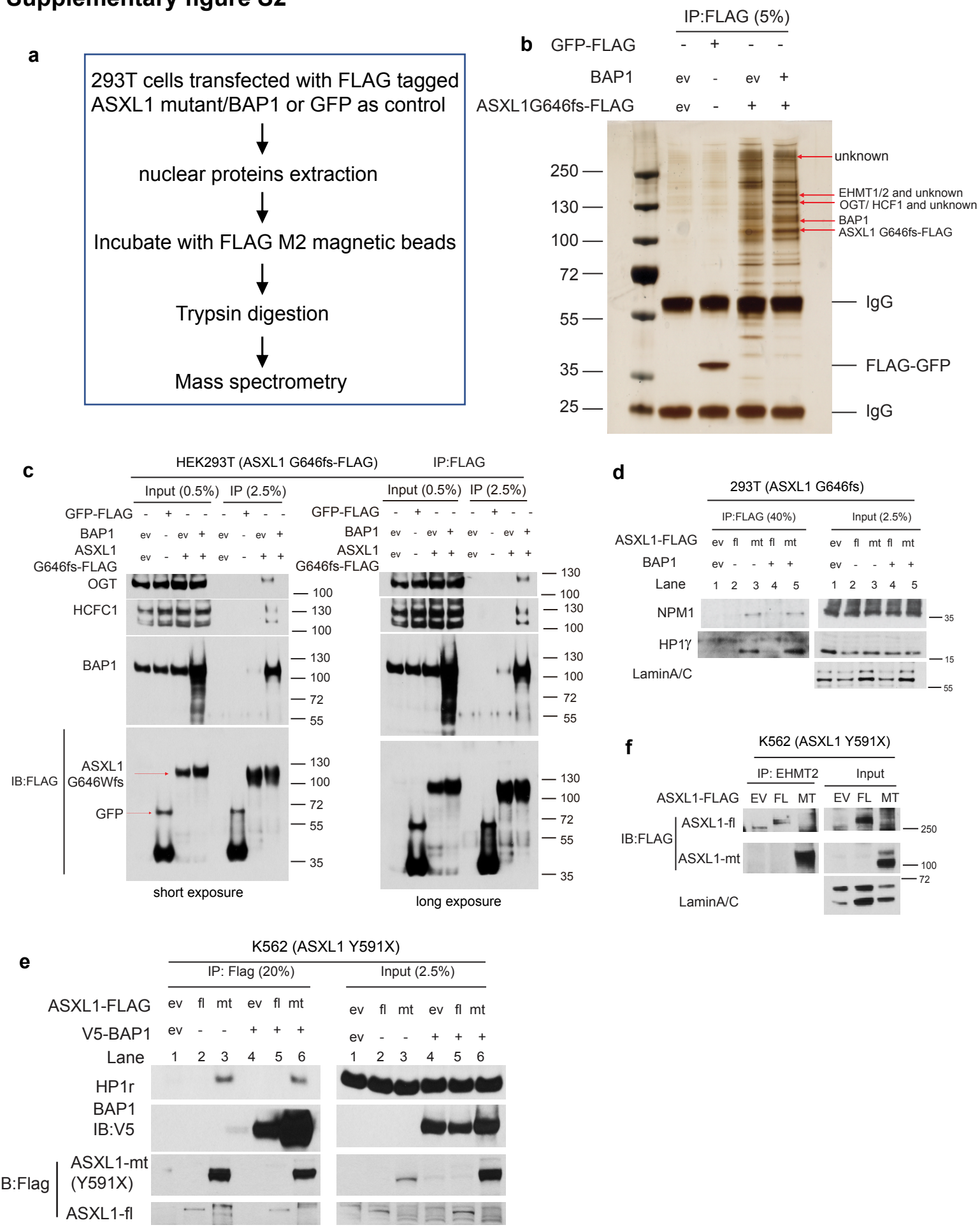

a
CD11b⁺ cells
Relative mRNA level
WT_4m
Asxl1tm/+_4m
WT_20m
Asxl1tm/+_20m
P=0.001
P=0.00001
P=0.008
P=0.02
P=0.005
P=0.01
P=0.02
Maj.sat
Min.sat
LINE1
SINE1
IAP1
c
LK (20 months old)
WT
Asxl1tm/+
P<0.00001
P=0.02
P<0.00001
P<0.0001
P<0.0001
b
CD11b⁺ cells at 18 month
-Log10 (adjusted p value)
Log2 Fold Change
RMER17C-int (n=1940)
L1MEi (n=478)
IAPEY4_LTR (n=94)
d
LK cells at 18 month
MMETn-int:ERVK:LTR (n=980)
L1MEd (n=3245)
L1MA4A (n=2709)
L1MEg (n=3797)
RLTR10C (n=1779)
LTR50 (n=549)
L1MA5A (n=1231)
LTR50 (n=549)
GSAT_MM (n=87)
RLTR47 (n=209)
IAPLTR2a (n=627)
MMTV-int (n=110)
ETnERV-int (n=906)
MLTR12 (n=477)
LTR33B (n=268)
LINE
LTR
Satellite
others
NO

Figure 5

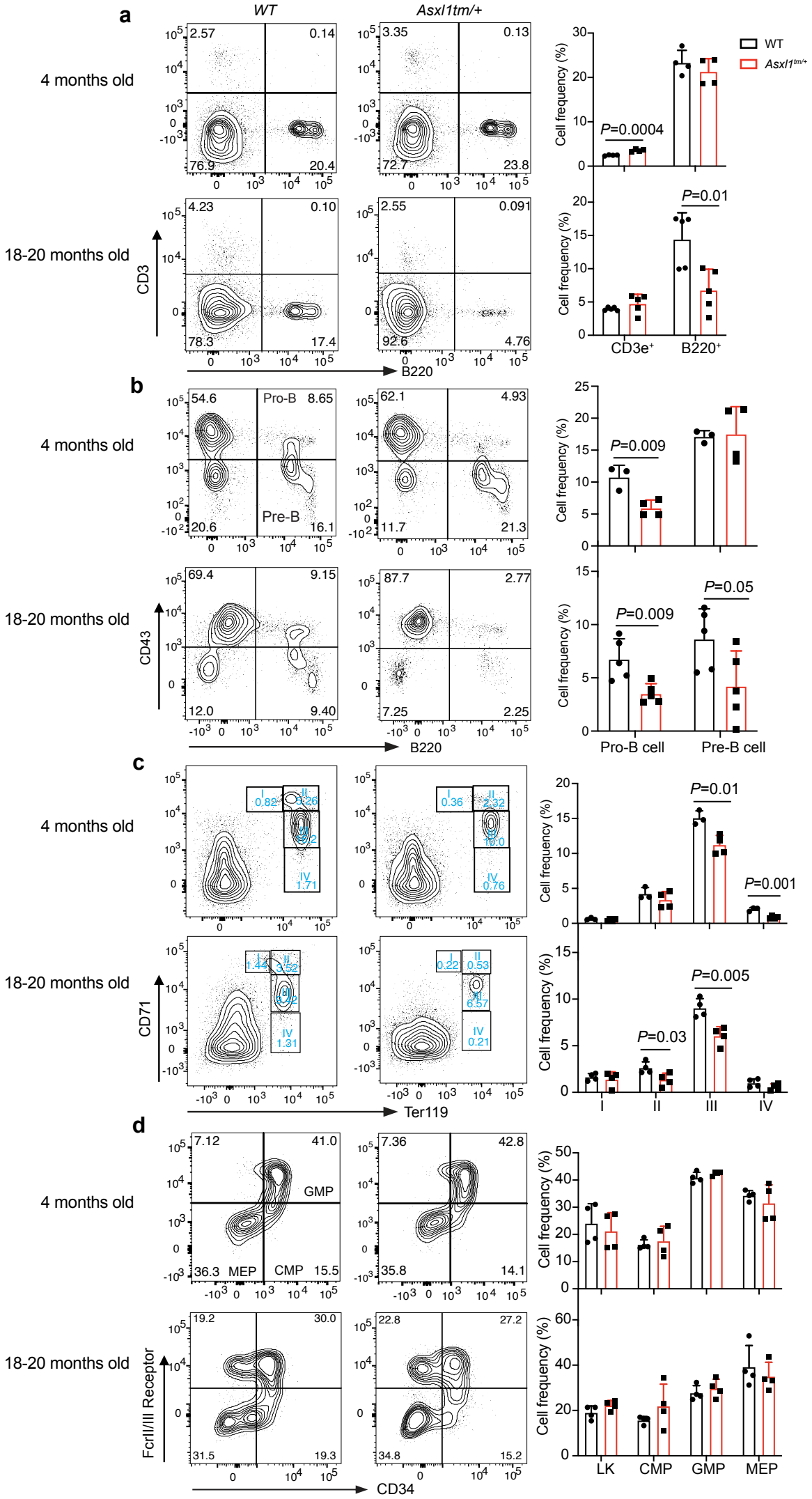

a
DEGs in ASXL1 mutant of CMML compared with health controls
UP:9432
ISGs:243
Down:3485
ISGs:192
Log2 Fold Change
b
PC2: 10% variance
PC1: 67% variance
ASXL1;IDH2;JAK2;NRAS;SRSF2
DNMT3A;PTPN11;TET2;TP53
JAK2;SRSF2;TET2
IDH2;PTPN11
PHF6;TET2;U2AF1;ZRSR2
ASXL1;RUNX1;SETBP1;SRSF2
SRSF2;TET2
SRSF2
KRAS
ASXL1;CBL;IDH2;SRF2
ASXL1;CBL;RUNX1;U2AF1
ASXL1;CBL;SRSF2;TET2
ASXL1;EZH2;TET2
ASXL1;EZH2;RUNX1;TET2
Ctrl-1
Ctrl-2
Ctrl-3
Ctrl-4
Ctrl-5
Ctrl-6
c
Decreased RNA expression
Increased RNA expression
Compartment (Hi-C PC1)
Euchromatin
Heterochromatin
Increasing Density
ASXL1 mutant CMML / control (RNA-seq)
d
TE expression in ASXL1 mutant of CMML compared with health controls
LINE
LTR
Satellite
Other TEs
No change
L1PB1:L1
L1PA13:L1
L1MB2:L1
L1PA11:L1
L1PBa:L1
L1PREC2:L1
L1M1:L1
HERVK-int:ERVK
L1PA2:L1
MER34B-int:ERV1
MER5A1
LTR16A2L:ERVL
ALR/Alpha:Satellite
LTR2B:ERV1
LSAU-BSAT:Satellite
-Log10 (adjusted P value)
Log2 Fold Change
e
10kb win
LINEs
LTRs
Satellites
SINEs
DNA-transposons
TEs enriched in euchromatin of normal human monocytes:54
TEs enriched in heterochromatin of normal human monocytes: 719
AluJo
AluSx
AluY
L1M5
L1PA6
L1PA11
L1ME1
L1PREC2
L1MA9
L1HS
L1PA2
L1PA13
L1M1
L1PBa
L1ME9
LSAU
ALR/Alpha
GSATII
HSATII
Relative Enrichment of TEs at chromatin domains in normal monocytes using HiC (log2 ratio)
Number of repeats per subfamily of Elements
f
RNA-seq Healthy patients
RNA-seq CMML patients
RepeatMasker
Hi-C (PC1)
RefSeq
L1PA6
PTH2R

Supplementary figure S1

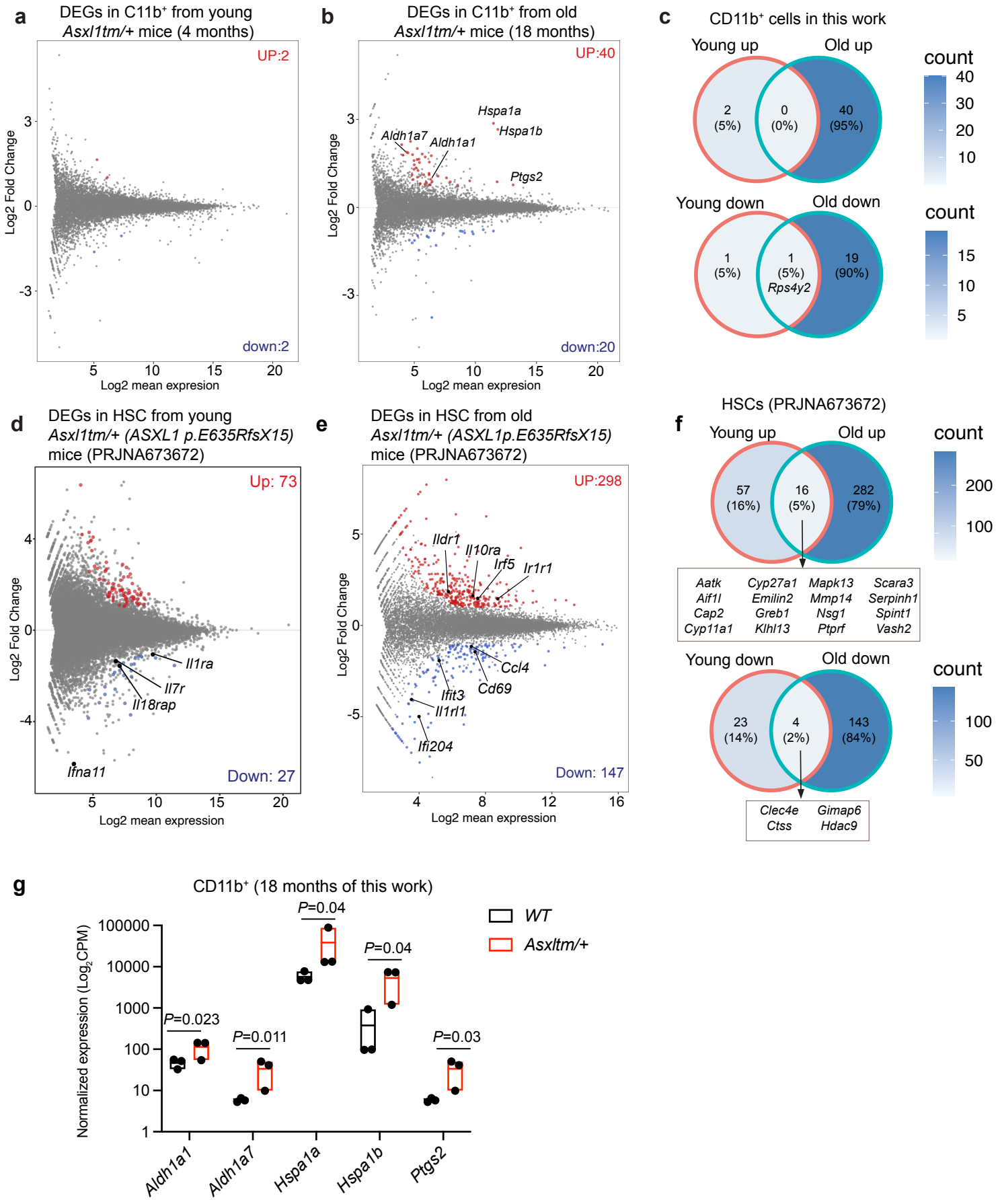
Supplementary figure S2



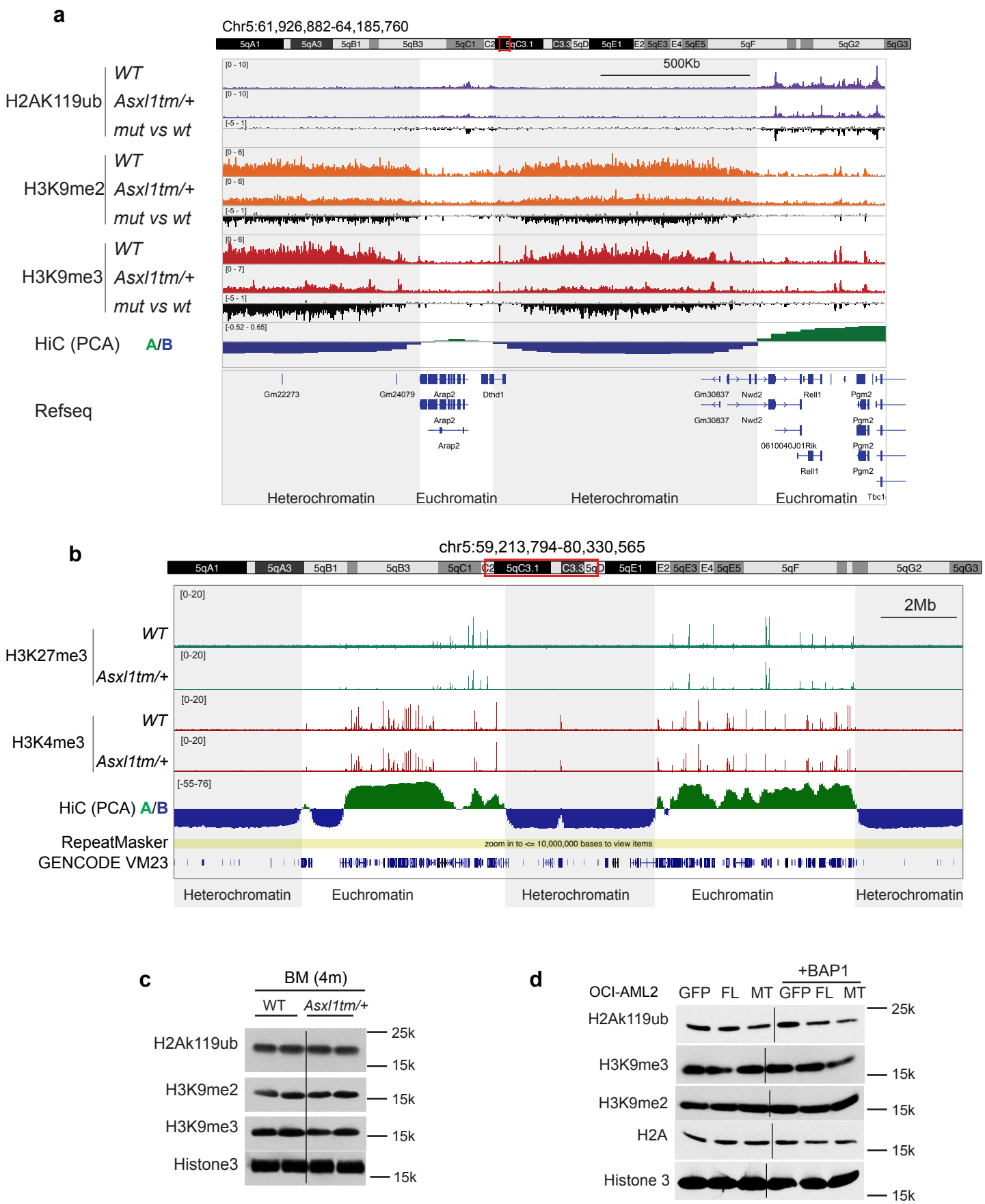
Supplementary figure S3



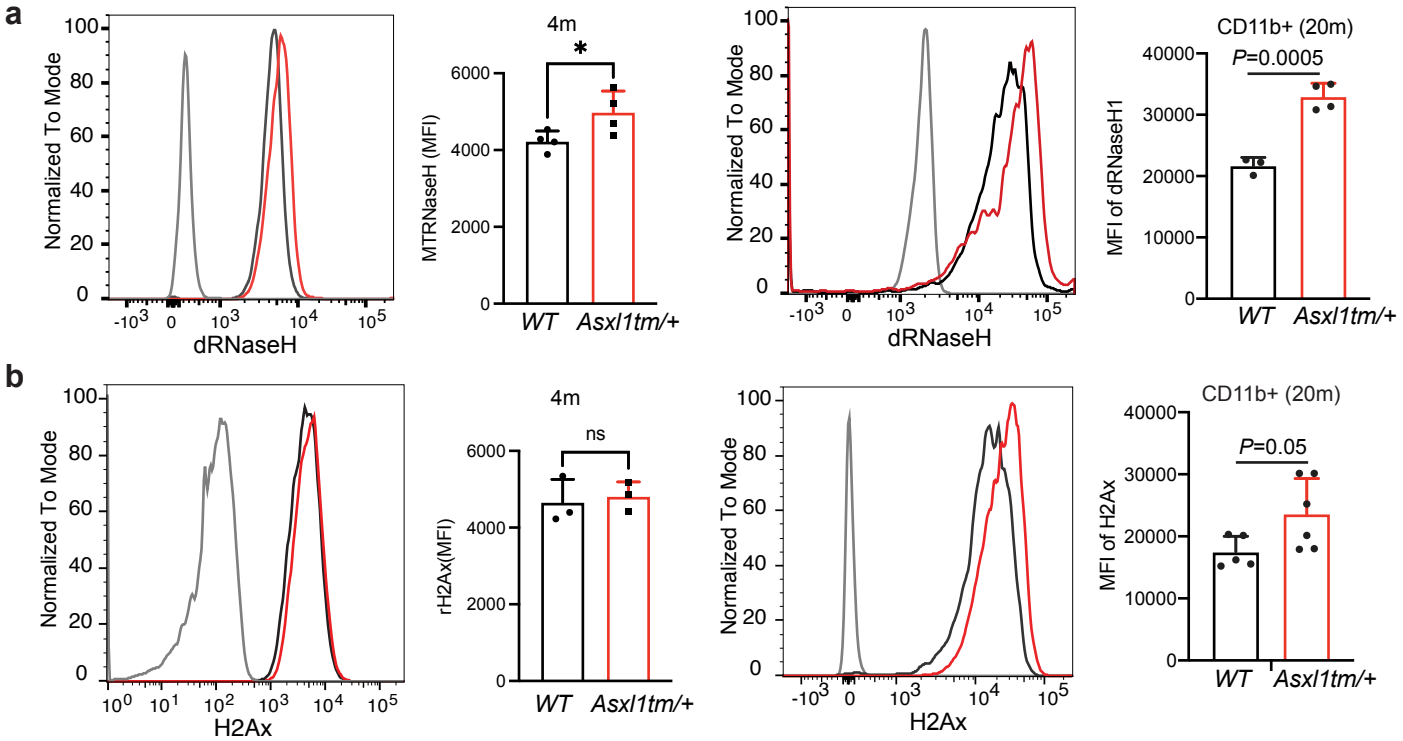
Supplementary figure S4



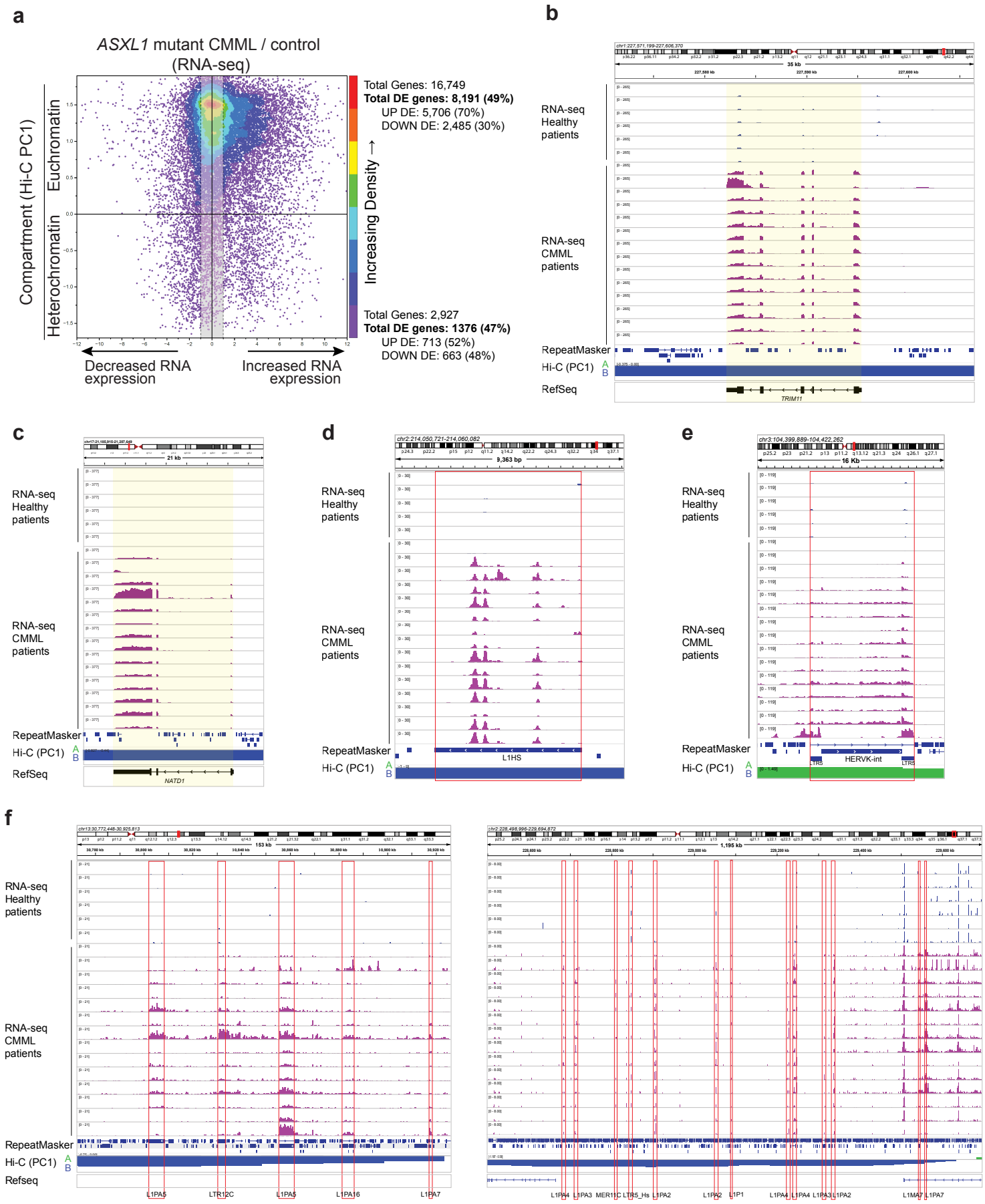
Supplementary figure S5



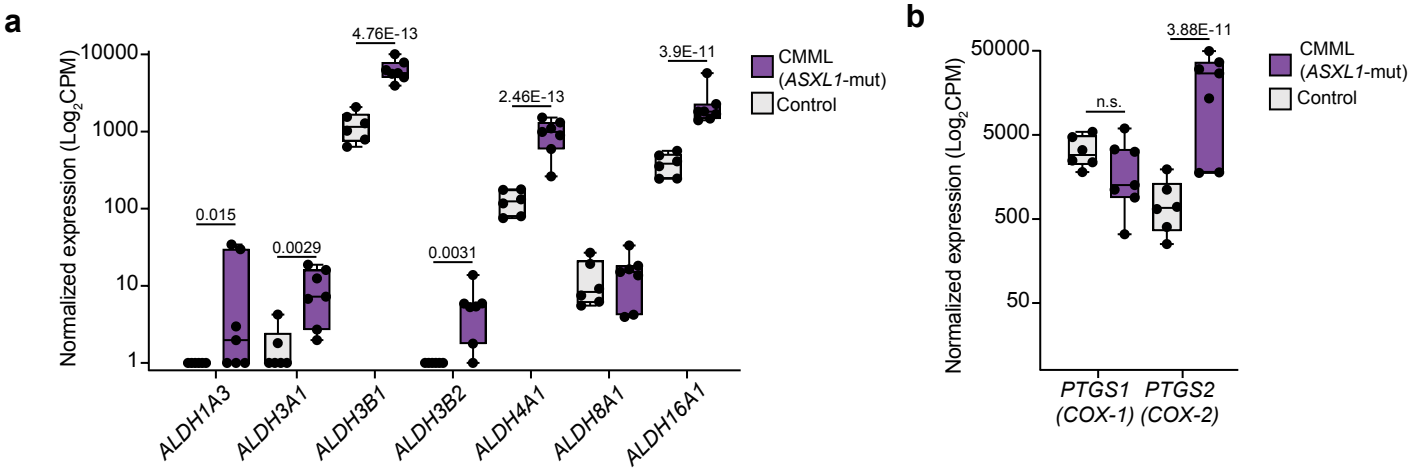
Supplementary figure S6

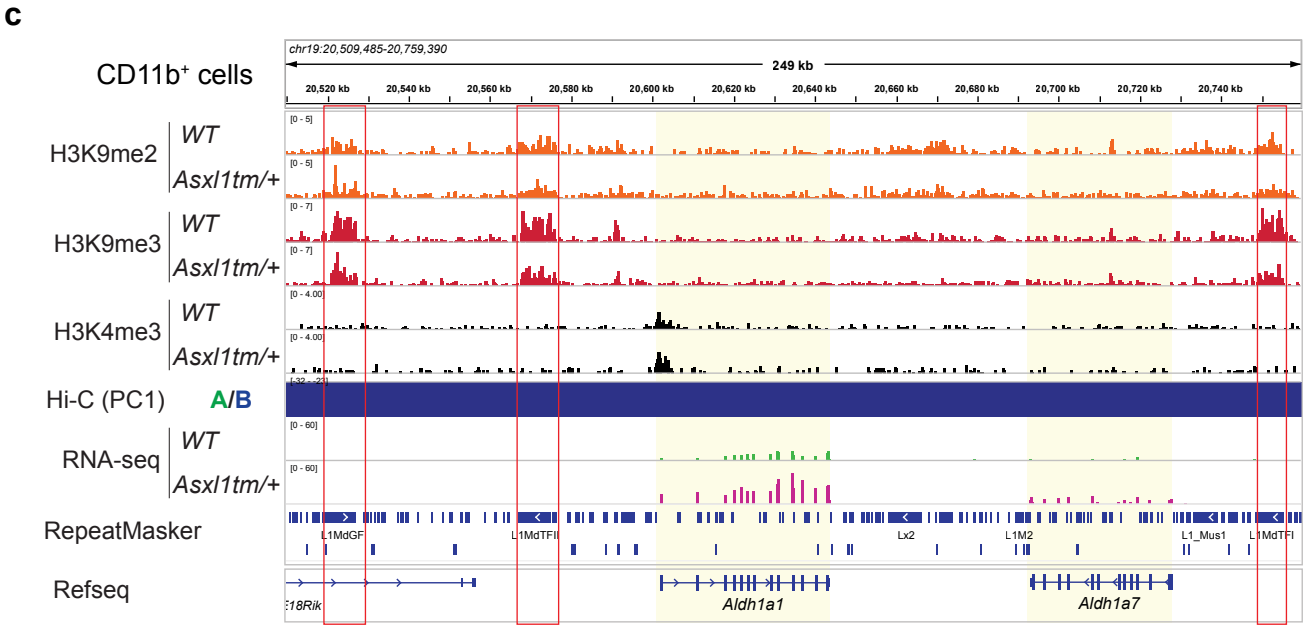


Supplementary figure S7



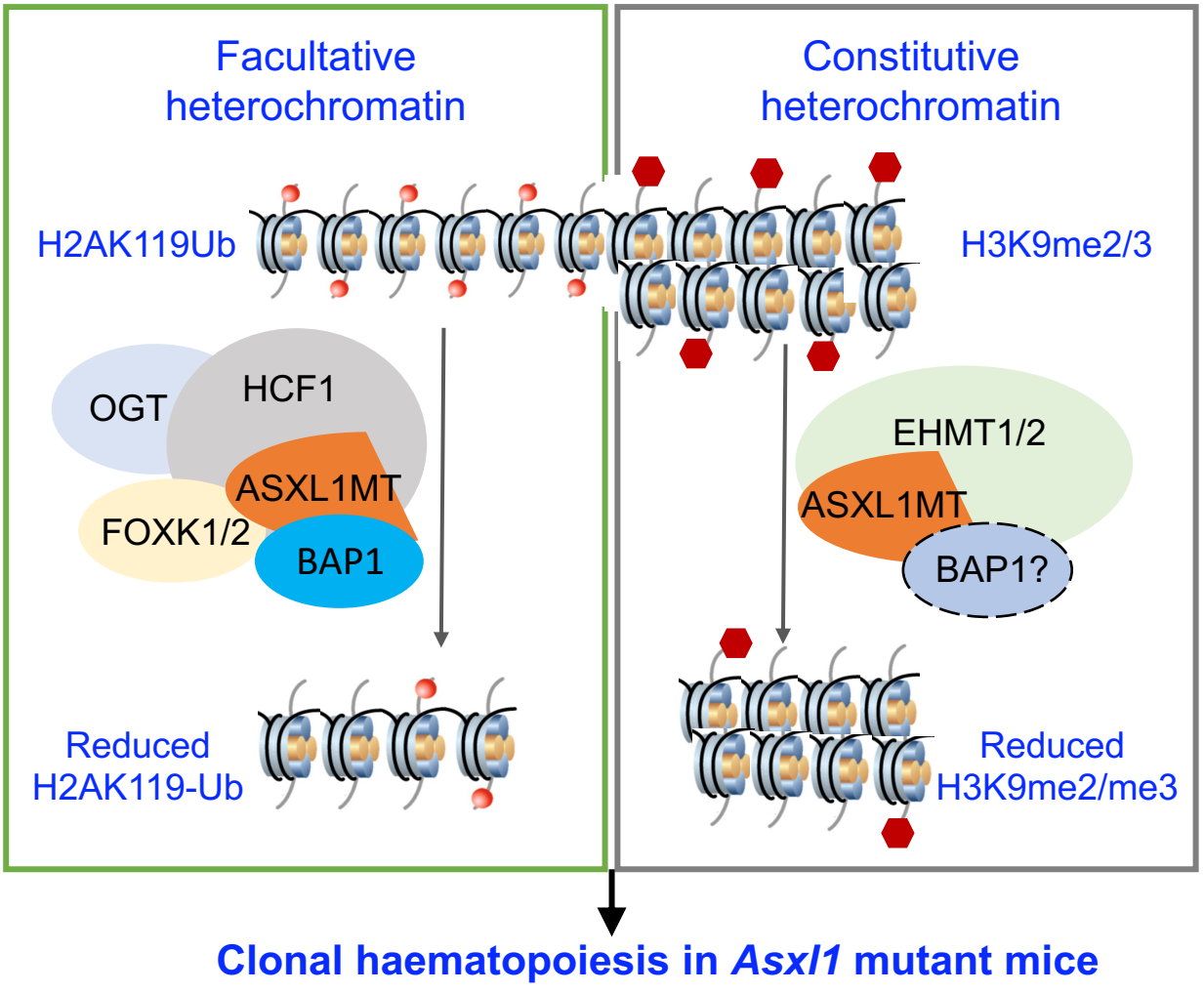
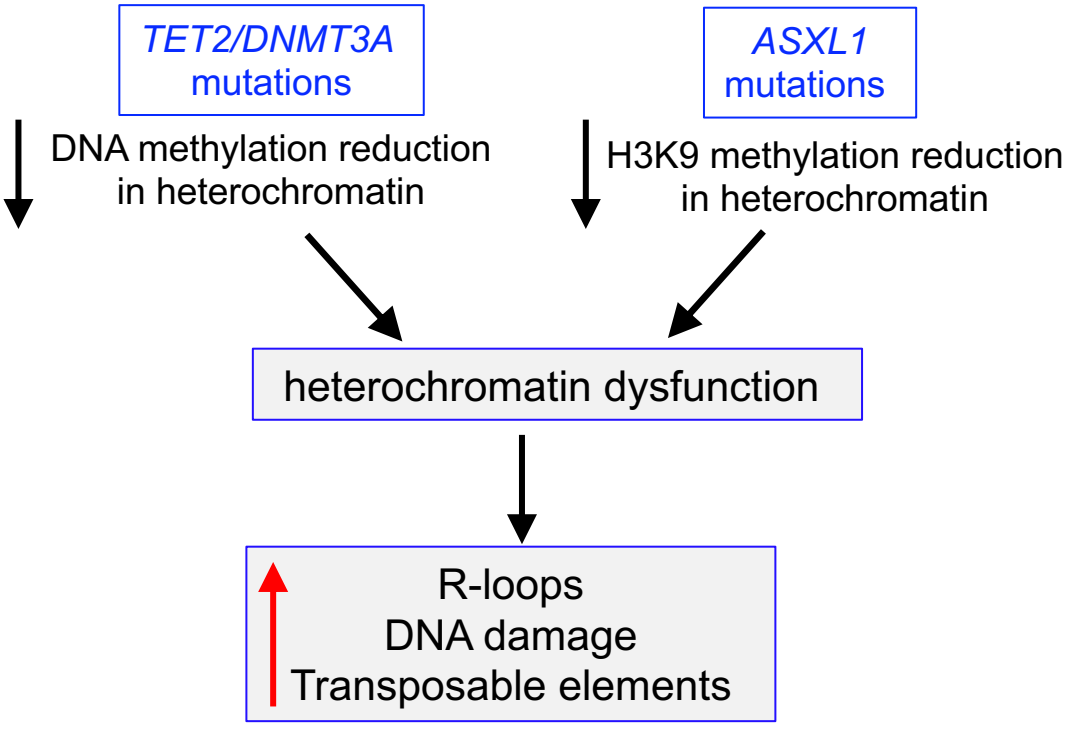
Supplementary figure S8





Supplementary figure S9

a Global histone modification changes in haematopoietic cells with CH/CMML-associated *Asxl1* mutations



b