

# UC Riverside

## UC Riverside Previously Published Works

### Title

TNRC18 engages H3K9me3 to mediate silencing of endogenous retrotransposons.

### Permalink

<https://escholarship.org/uc/item/5712r9nz>

### Journal

Nature: New biology, 623(7987)

### Authors

Zhao, Shuai

Pan, Bo

Fan, Huitao

et al.

### Publication Date

2023-11-01

### DOI

10.1038/s41586-023-06688-z

Peer reviewed



Published in final edited form as:

Nature. 2023 November ; 623(7987): 633–642. doi:10.1038/s41586-023-06688-z.

## TNRC18 engages H3K9me3 to mediate silencing of endogenous retrotransposons

Shuai Zhao<sup>1,2,3,4,12</sup>, Jiuwei Lu<sup>5,12</sup>, Bo Pan<sup>1,2</sup>, Huitao Fan<sup>4,11</sup>, Stephanie D. Byrum<sup>6</sup>, Chenxi Xu<sup>1,2,3,4</sup>, Arum Kim<sup>1,2,4</sup>, Yiran Guo<sup>1,2,7</sup>, Krishna L. Kanchi<sup>4</sup>, Weida Gong<sup>4</sup>, Tongyu Sun<sup>8</sup>, Aaron J. Storey<sup>6</sup>, Nathaniel T. Burkholder<sup>9</sup>, Samuel G. Mackintosh<sup>6</sup>, Peyton C. Kuhlers<sup>10</sup>, Ricky D. Edmondson<sup>6</sup>, Brian D. Strahl<sup>4,9</sup>, Yarui Diao<sup>2,8</sup>, Alan J. Tackett<sup>6</sup>, Jesse R. Raab<sup>4,10</sup>, Ling Cai<sup>2,3,10</sup>, Jikui Song<sup>5,13,✉</sup>, Gang Greg Wang<sup>1,2,3,4,9,13,✉</sup>

<sup>1</sup>Department of Pharmacology and Cancer Biology, Duke University School of Medicine, Durham, NC, USA.

<sup>2</sup>Duke Cancer Institute, Duke University School of Medicine, Durham, NC, USA.

<sup>3</sup>Department of Pathology, Duke University School of Medicine, Durham, NC, USA.

<sup>4</sup>Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC, USA.

<sup>5</sup>Department of Biochemistry, University of California, Riverside, CA, USA.

<sup>6</sup>Department of Biochemistry and Molecular Biology, University of Arkansas for Medical Sciences, Little Rock, AR, USA.

<sup>7</sup>Curriculum in Genetics and Molecular Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.

<sup>8</sup>Department of Cell Biology, Duke University School of Medicine, Durham, NC, USA.

<sup>9</sup>Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC, USA.

<sup>10</sup>Department of Genetics, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC, USA.

✉ Correspondence and requests for materials should be addressed to Jikui Song or Gang Greg Wang. jikui.song@ucr.edu; greg.wang@duke.edu.

**Author contributions** S.Z. and J.L. led the biological and structural aspects of this study, respectively, under the supervision of G.G.W. and J.S. S.Z., H.F., Y.G., A.K., L.C. and G.G.W. conducted cell-based characterizations and animal studies. S.Z. and C.X. conducted the protein-complex-related studies. S.D.B., A.J.S., S.G.M. and R.D.E. performed proteomics analysis under the supervision of A.J.T. S.Z. and T.S. conducted CasID under the guidance of G.G.W. and Y.D. S.Z., J.L. and N.T.B. performed biochemical and biophysical assays under the supervision of J.S., B.D.S. and G.G.W. S.Z., B.P., K.L.K., W.G., P.C.K., J.R.R., L.C. and G.G.W. analysed deep-sequencing datasets. G.G.W. and J.S. conceived the idea and designed and supervised the research. S.Z., J.L., J.S. and G.G.W. wrote the manuscript with input from all authors.

**Competing interests** B.D.S. is a co-founder of and is on the board of directors of EpiCypher.

Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41586-023-06688-z>.

**Peer review information** Nature thanks Jon Wilson, Déborah Bourc'his and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>

<sup>11</sup>Present address: The First Affiliated Hospital of Harbin Medical University, Harbin, P. R. China.

<sup>12</sup>These authors contributed equally: Shuai Zhao, Jiuwei Lu.

<sup>13</sup>These authors jointly supervised this work: Jikui Song and Gang Greg Wang.

## Abstract

Trimethylation of histone H3 lysine 9 (H3K9me3) is crucial for the regulation of gene repression and heterochromatin formation, cell-fate determination and organismal development<sup>1</sup>. H3K9me3 also provides an essential mechanism for silencing transposable elements<sup>1–4</sup>. However, previous studies have shown that canonical H3K9me3 readers (for example, HP1 (refs. 5–9) and MPP8 (refs. 10–12)) have limited roles in silencing endogenous retroviruses (ERVs), one of the main transposable element classes in the mammalian genome<sup>13</sup>. Here we report that trinucleotide-repeat-containing 18 (TNRC18), a poorly understood chromatin regulator, recognizes H3K9me3 to mediate the silencing of ERV class I (ERV1) elements such as LTR12 (ref. 14). Biochemical, biophysical and structural studies identified the carboxy-terminal bromo-adjacent homology (BAH) domain of TNRC18 (TNRC18(BAH)) as an H3K9me3-specific reader. Moreover, the amino-terminal segment of TNRC18 is a platform for the direct recruitment of co-repressors such as HDAC–Sin3–NCoR complexes, thus enforcing optimal repression of the H3K9me3-demarcated ERVs. Point mutagenesis that disrupts the TNRC18(BAH)-mediated H3K9me3 engagement caused neonatal death in mice and, in multiple mammalian cell models, led to derepressed expression of ERVs, which affected the landscape of cis-regulatory elements and, therefore, gene-expression programmes. Collectively, we describe a new H3K9me3-sensing and regulatory pathway that operates to epigenetically silence evolutionarily young ERVs and exert substantial effects on host genome integrity, transcriptomic regulation, immunity and development.

Almost half of the human genome is made up of transposable elements (TEs), many of which are retrotransposons such as long terminal repeats (LTRs) and non-LTRs<sup>15–17</sup>. Non-LTRs include long interspersed elements (LINEs) and short interspersed elements (SINEs), whereas LTRs consist of ERVs. To maintain host genome integrity, TEs are tightly regulated by various silencing mechanisms enforced by histone modification, DNA methylation, histone chaperone and/or noncoding RNAs<sup>15,18</sup>. Notably, H3K9me3, a histone mark characteristic of constitutive heterochromatin, is enriched at TEs<sup>1,15</sup>. H3K9me3 is deposited by its writer enzymes such as SETDB1 (also known as ESET)<sup>1,19–21</sup> and recognized by the H3K9me3-reading chromodomain harboured within heterochromatin protein 1 (HP1) family proteins<sup>5–9</sup>. SETDB1 loss causes global TE activation<sup>13</sup>; however, such a large TE-activation phenotype cannot be recapitulated by HP1 loss<sup>13</sup>, which indicates the existence of unknown H3K9me3-associated mechanisms for mediating TE repression. Chromodomain within M-phase phosphoprotein 8 (MPP8), a component of the human silencing hub (HUSH) complex<sup>2</sup>, was identified as an H3K9me3-specific reader<sup>10–12</sup>. However, depletion of the HUSH complex mainly caused LINE1 derepression, which demonstrates that MPP8–HUSH has a major role in silencing non-LTR (LINE1) TEs<sup>2–4</sup>. It remains elusive what H3K9me3 reader mediates silencing of LTRs (ERVs), a TE class that occupies approximately 8% of the human genome and potentially carries gene regulatory functions<sup>15,16,18</sup>.

## CasID identifies TNRC18 as an ERV binder

We sought to identify mammalian ERV regulators using CasID<sup>22</sup> (Extended Data Fig. 1a), in which a fusion protein of endonuclease-inactive Cas9 and a promiscuous biotin ligase (dCas9–BirA\*) was site-specifically targeted to ERVs by single guide RNAs (sgRNAs) targeting 20 representative ERV loci (including ERV1 and ERV classes II and III) in HEK293 cells. dCas9–BirA\* induced biotinylation of putative ERV-bound proteins, which were identified by mass spectrometry following NeutrAvidin pulldown (Fig. 1a and Supplementary Table 1). The most confident hits included well-established ERV co-repressors and repressors such as tripartite motif-containing protein 28 (TRIM28; also known as KAP1 and TIF1 $\beta$ ), MORC family CW-type zinc finger 2 (MORC2), H3K9me3-specific readers (CBX1 (also known as HP1 $\beta$ ) and CBX5 (also known as HP1 $\alpha$ )), DNA methyltransferase machinery (DNMT1 and UHRF1), small ubiquitin like modifier 2 (SUMO2), zinc finger proteins (ZFPs; about 50 ZFPs identified, such as the Kruppel C2H2-type ZFP family members ZNF384, ZNF362 and ZNF451), histone chaperones (ATRX, DAXX and CHAF1B) and N6-methyladenosine-related regulators (METTL3, YTHDC1 and YTHDF1)<sup>23–25</sup> (Fig. 1a and Supplementary Table 1). Previous studies have shown that TRIM28 interacts with KRAB-ZFPs, which bind specific DNA sequences at TE regions to mediate ERV silencing, and that SUMO2-based TRIM28 SUMOylation regulates genomic targeting of TRIM28 and promotes interactions between TRIM28 and SETDB1 (refs. 15,18). Most of the above-mentioned proteins were also previously identified from ERV reporter-based screening assays<sup>23,26–28</sup>, which supported the validity of our CasID-based approach. TNRC18, a protein with currently unknown biological functions, was identified as a highly ranked hit, showing 13-fold enrichment compared with control in the CasID screen (Fig. 1a, red). Immunofluorescence for TNRC18 revealed an exclusively nuclear pattern in cells using independent approaches (Extended Data Fig. 1b–e). We propose that TNRC18 acts as a chromatin-associated regulator of ERVs.

## TNRC18 mediates the silencing of ERV1

To determine putative TE regulatory roles of TNRC18, we performed TNRC18 knockdown (KD) in HEK293 cells and subsequent TNRC18 rescue by re-expressing exogenous TNRC18, which carries synonymous mutations designed to confer resistance to TNRC18-targeting short hairpin RNA (shRNA) (Extended Data Fig. 1f). Because traditional shRNA-based strategies only produced partial TNRC18 KD (Extended Data Fig. 1f), we also used a CRISPR–Cas9-based approach to achieve TNRC18 knockout (KO) in the same cells (Extended Data Figs. 2a,b). Subsequent RNA sequencing (RNA-seq)-based analyses of TEs revealed that among the various TE families, the ones most re-activated after TNRC18 loss (either KD or KO) compared with mock included LTR12 and LTR7, and that such reactivated ERVs became re-silenced following TNRC18 re-expression (Fig. 1b,c, Extended Data Fig. 1g and Supplementary Table 2). LTR12 is abundant, with more than 6,000 copies in the human genome, and evolutionarily young<sup>14</sup> (Extended Data Fig. 1h). Also, both full-length proviruses HERV9NC-int and solitary LTRs (for example, LTR12B, LTR12C and LTR12F, with their internal viral sequences deleted owing to homologous recombination events) of the LTR12 family were repressed by TNRC18. Notably, compared with endogenous TNRC18 expression, that of the exogenously introduced TNRC18 for

rescue was significantly higher (Extended Data Fig. 1f), which led to more substantial silencing of ERV1 subfamilies and other TE families such as ERV class II (Fig. 1b). RNA-seq analysis using both unique mapping reads and multi-mapping reads showed consistent results (Extended Data Fig. 2c). Reverse transcription followed by quantitative PCR (RT-qPCR) confirmed de-repression of the above ERVs following TNRC18 loss (KD or KO) and re-silencing of ERVs following TNRC18 re-expression (Fig. 1d,g). Thus, TNRC18 dose-dependently silences ERV1.

We next surveyed publicly available datasets and found that TNRC18 is widely expressed among various tissues (Extended Data Fig. 3a). Such ubiquitous TNRC18 expression suggests that this protein has a more general role that goes beyond HEK293 cells. To test this notion, we used independent designs of TNRC18-targeting sgRNAs and generated TNRC18 KO in four different human cancer cell lines of epithelial origin: 22Rv1, a human prostate carcinoma epithelial cell line; HT-29, a human colorectal adenocarcinoma line; SNU-1, a human gastric carcinoma line; and NCI-H23, a human lung adenocarcinoma line (Extended Data Figs. 3b,c). RT-qPCR showed that TNRC18 KO caused consistent and strong reactivation of LTR12 and LTR7 family TEs in all tested cell lines (Fig. 1g). The degree of TNRC18 KO-induced TE reactivation in these cells was either comparable with what was seen in HEK293 cells or more substantial, such as activation of LTR12, HERV9NC-int and/ or LTR7 by approximately 3–6-fold (Fig. 1d,g). RNA-seq of SNU-1 and NCI-H23 cells before and after TNRC18 KO also showed global activation of ERVs (Figs. 1h,i and Supplementary Table 2), as seen with TNRC18 KO in HEK293 cells.

## TNRC18 loss leads to immune activation

Retrotransposon activation generates double-stranded RNAs that stimulate interferon-responsive genes, a process known as viral mimicry<sup>29,30</sup>. Gene set enrichment analysis (GSEA) revealed that immunity-related pathways were upregulated following TNRC18 KD, and such activation of immunity-related genes was reversed after restoration of TNRC18 expression (Fig. 1e, Extended Data Figs. 2d,e and Supplementary Table 3). RT-qPCR verified that TNRC18 loss substantially activated the transcription of immunity-related genes such as IL11, IL18 and TLR4 (Fig. 1d, bottom). Analyses of RNA-seq profiles of three different human cells (HEK293, SNU-1 and HT-29) also showed that there was significant, consistent activation of immunity-related genes (Figs. 1d–f, Extended Data Figs. 2e,f and 3d and Supplementary Table 3). Depletion of the HUSH complex can cause activation of interferon-stimulated genes owing to derepressed LINE1s<sup>4</sup>. Here TNRC18 loss leads to preferential derepression of ERVs, which similarly induces viral mimicry.

## TNRC18 directly associates with ERVs

To determine genomic binding of TNRC18 in cells, we performed cleavage under targets and release using nuclease (CUT&RUN) for TNRC18 using independent approaches. We used either endogenous TNRC18 antibody in parental HEK293 cells or GFP antibody in cells stably expressing GFP-tagged TNRC18. CUT&RUN for endogenous TNRC18 and GFP-TNRC18 produced strong and highly correlated profiles, thereby confirming the validity of the mapped binding sites (Pearson  $r^2 = 0.82$ ; Extended Data Fig. 4a).

A significant proportion of TNRC18 peaks were called as TE-bound (for details, see Methods). Approximately 50% and 31% of these most confident TNRC18-bound TEs were annotated as SINEs and ERVs, respectively (Supplementary Fig. 4b). Consistently, subsequent enrichment analyses of TNRC18 binding also revealed that it was most enriched at ERVs, notably ERV1 (Fig. 2a and Extended Data Fig. 4c; columns in red and green), and at SINEs but not other TEs, including LINEs, satellite DNA and simple repeats (Fig. 2a and Extended Data Fig. 4c). CUT&RUN for TNRC18 in HeLa and K562 cells produced overall similar binding patterns (Extended Data Fig. 4d), as exemplified by what was detected at ERVs (LTR12C and MER52C) and promoter or intergenic target sites (Extended Data Fig. 4e). This result indicates that there is conservation in genomic targeting of TNRC18 across different cell types. CUT&RUN for H3K9me3 with independent antibodies showed that H3K9me3 overlapped with TE-associated TNRC18 peaks (Extended Data Fig. 4f), particularly those located at ERV regions (Extended Data Fig. 4g).

Next, we examined TNRC18 and H3K9me3 binding across all ERV subfamilies of TEs. Most sites that showed dual binding of TNRC18 and H3K9me3 belong to ERV1, such as LTR12C, LTR12D and LTR12E (Fig. 2b, blue). By contrast, certain subfamilies of ERV class II and III TEs exhibited high H3K9me3 positivity but had low enrichment for TNRC18 binding (Fig. 2b, orange and yellow). Note that our RNA-seq-based profiling identified that LTR12 family TEs were the most silenced by TNRC18, which is in line with the observation that they were the most strongly bound by TNRC18 and H3K9me3 based on CUT&RUN (Fig. 2b,c). Unbiased motif analysis of TNRC18 peaks revealed the most enrichments for NFY (CCAAT) and Sp (CCCCACC-like) motifs (Extended Data Fig. 4h and Fig. 2d), which are typical motifs in the U3 region of LTR12 (ref. 31). This result raises the possibility that certain DNA-sequence-specific binding factors are involved in directing preferential targeting of TNRC18 to ERVs. Pearson correlation revealed that TNRC18 binding was positively correlated with H3K9me3 at LTR12 regions (Fig. 2e), as exemplified by what was seen at a LTR12D copy in chromosome 13 (Fig. 2f). Together, these data demonstrate that TNRC18 binds H3K9me3-demarcated ERVs, particularly LTR12.

### TNRC18(BAH) specifically reads H3K9me3

Notable domains within TNRC18 include a C-terminal BAH domain (residues 2785–2968) that is conserved across species (Fig. 3a and Extended Data Fig. 5a). BAH domains of several chromatin-associated proteins such as ORC1 and BAHCC1 can recognize specific histone methylation marks<sup>32–36</sup>. Observed colocalization of TNRC18 and H3K9me3 at ERVs prompted us to assess the putative histone-binding activity of TNRC18(BAH). Isothermal titration calorimetry (ITC) analyses showed that purified TNRC18(BAH) protein bound H3K9me3 peptide, with a dissociation constant (K<sub>d</sub>) of 81 μM. By contrast, it bound H3K9me2 more weakly (K<sub>d</sub> = 252 μM) and showed no detectable interaction with H3K9me1 or H3K9me0 (Fig. 3b and Extended Data Fig. 5b–e). Unlike what was observed for H3K9me3, TNRC18(BAH) exhibited no binding to other tested histone trimethylation, including H3K4me3, H3K27me3, H3K36me3 or H4K20me3 (Extended Data Fig. 5c–e).

To determine the molecular basis underlying TNRC18(BAH)-mediated recognition of H3K9me3, we solved the crystal structure of TNRC18(BAH) in complex with H3(1–

22)K9me3 peptide at 1.84-Å resolution (Extended Data Fig. 6a,b and Extended Data Table 1). The structure of TNRC18(BAH) is dominated by an elongated  $\beta$ -fold comprising 12  $\beta$ -strands that are further packed against 5 intervening helices. Notably, residues K4–G12 of the H3(1–22)K9me3 peptide run in antiparallel with the  $\beta$ 9-strand (Fig. 3c), positioning the side chain of H3K9me3 into an aromatic cage formed by TNRC18(BAH) residues Y2837, W2858, Y2860 and E2864 for cation- $\pi$  and electrostatic interactions (Fig. 3d). The association of H3(1–22)K9me3 with TNRC18(BAH) is further maintained by a network of hydrogen bonds (Fig. 3e,f). For instance, the side chain of H3 T6 forms a water-mediated hydrogen bond with the backbone of TNRC18(BAH) A2830. The backbone of H3 A7 forms direct and water-mediated hydrogen bonds with the backbone of TNRC18(BAH) L2828 and A2830, respectively. The backbone of H3 S10 forms a hydrogen bond with the side chain of TNRC18(BAH) D2910, whereas the side chain of H3 S10 engages in a bifurcated hydrogen bond with the side chains of TNRC18(BAH) H2908 and D2910 (Fig. 3e,f). Additionally, the H3(1–22)K9me3–TNRC18(BAH) association is supported by van der Waals contacts involving H3K9me3-flanking residues such as the sidechain contacts between H3 T6 and TNRC18(BAH) L2828–A2830 (Fig. 3e,f).

The histone tail sequences flanking H3K9me3 or H3K27me3 share a so-called ARKS motif but differ at the –3 position (T6 compared with A24; Fig. 3g, bottom). The observation that T6 of H3(1–22)K9me3 engages in sidechain hydrogen-bonding and van der Waals contacts with residues L2828–A2830 indicates a possible mechanism for discriminating H3K9me3 peptide over H3K27me3 peptide (Fig. 3g, top). In support of this notion, replacement of H3(1–22)K9me3 T6 by alanine (T6A) or phosphorylated threonine (T6ph) abolished the binding between TNRC18(BAH) and the H3(1–22)K9me3 peptide. Conversely, phosphorylation of H3(1–22)K9me3 T11 did not appreciably affect binding (Fig. 3h and Extended Data Fig. 6c). Point mutagenesis of the caging residues of TNRC18(BAH) (Y2837A, W2858A, Y2860A or E2864A) or the residues that mediate histone backbone interactions (H2908A, D2910A or N2912A) all disrupted the TNRC18(BAH) binding to H3K9me3-containing peptides, as measured by ITC (Fig. 3i and Extended Data Fig. 6d). Phosphorylation of H3S10 abrogated the TNRC18(BAH) binding to H3(1–22)K9me3 peptide (Fig. 3h and Extended Data Fig. 6c), similar to previous reports for other H3K9me3 readers<sup>37</sup>.

Structural comparisons of the TNRC18(BAH)–H3K9me3 complex with previously reported BAH–histone complexes<sup>32,34,36</sup> revealed that BAH domains present a similar surface region for histone recognition, but with highly diverse interaction mechanisms (Extended Data Fig. 6e). Of note, BAHCC1(BAH) forms sequence-specific interactions with residues H3S28 and H3P30 of the H3K27me3-containing peptide (Extended Data Fig. 6e), in contrast to the TNRC18(BAH)–H3T6 interaction (Fig. 3g). Thus, BAH domains have evolved into a versatile family of histone readers.

It is worth noting that structural superposition of TNRC18(BAH) with the nucleosome core particle (NCP)-bound yeast Sir3 BAH domain (Sir3(BAH); Protein Data Bank (PDB) identifier 4KUD)<sup>38</sup> revealed that a basic patch on the surface of TNRC18(BAH) (Extended Data Fig. 6b, right) corresponds to a region in Sir3(BAH) involved in binding to the acidic patch of NCP (Extended Data Fig. 6f), which indicated the presence of additional

NCP-binding sites in TNRC18(BAH). Indeed, bilayer interferometry assays showed that TNRC18(BAH) binds to NCP containing a methyllysine analogue of histone H3K9 (H3Kc9me3) with a  $K_d$  of 18.8 nM, much tighter than the H3K9me3 peptide alone in vitro (Extended Data Fig. 6g,h). Together, these data point to a strong binding affinity between TNRC18 and NCP.

### H3K9me3-bound TNRC18 silences ERV1

To define the biological function of the H3K9me3-recognizing TNRC18(BAH), we used CRISPR–Cas9-based gene editing and generated HEK293 cells harbouring an engineered homozygous mutation in TNRC18 to disrupt its BAH-mediated H3K9me3 engagement (TNRC18W2858A; Extended Data Fig. 7a,b). RNA-seq-based profiling of TEs in TNRC18(BAH)-mutant cells, relative to wild-type (WT), showed that the majority of activated TEs belong to ERV1, especially the LTR12 and LTR7 family of TEs covering both full-length proviruses (such as HERV9-int and HERVH-int) and solitary LTRs (such as LTR12 and LTR7) (Fig. 4a and Supplementary Table 4). Consistent with TE activation, immunity-related pathways were activated in the TNRC18-(BAH)mutant cells (Fig. 4b, Extended Data Fig. 7c and Supplementary Table 4). TNRC18(BAH)-associated transcriptional alterations were further confirmed by RT–qPCR (Fig. 4c), reminiscent of those caused by TNRC18 depletion (Fig. 1d).

TEs can serve as gene regulatory elements following activation<sup>17,18</sup>. We examined expression of the genes in proximity to TNRC18-bound sites. The overall expression of genes associated with all TNRC18 peaks or those with TNRC18-bound transcription start site (TSS) regions did not show significant changes in TNRC18(BAH)-mutant cells compared with WT cells (Extended Data Fig. 7d). By contrast, those associated with the TNRC18-bound LTRs exhibited significant overall activation (Extended Data Fig. 7e). At genomic regions harbouring LTRs bound by both H3K9me3 and TNRC18, the genes upregulated in TNRC18(BAH)-mutant cells relative to WT cells showed closer proximity to the nearby TNRC18 peaks than those downregulated or without expression change (with an average of 4.81 kb, 10.26 kb and 26.32 kb for the upregulated, downregulated or stably expressed genes, respectively, from a nearby TNRC18 peak; Extended Data Fig. 7f). A similar gene regulatory mode was reported for ERVs activated after TRIM28 loss<sup>39</sup>. Collectively, these results show that TNRC18 silences ERV1, which influences the expression of nearby genes in a BAH-dependent and H3K9me3-dependent manner.

### TNRC18 silences transcription from TEs

Treatment with DNA methyltransferase and histone deacetylase (HDAC) inhibitors can induce genome-wide activation of treatment-induced non-annotated TSSs (TINATs), over 80% of which overlapped TEs, with the LTR12 family most enriched<sup>40</sup>. Our RT–qPCR-based results showed representative TINATs, including those in CRADD, DNAH12 and FMN1, were all activated in TNRC18(W2858A)-mutant cells compared with their expression in WT cells (Extended Data Fig. 7g). To confirm that TINAT activation occurs at a genome-wide level, we conducted cap analysis of gene expression followed by sequencing (CAGE–seq)<sup>41</sup> (Extended Data Fig. 7h). CAGE–seq accurately identifies



TSSs and distinguishes multiple transcription isoforms<sup>41</sup>. Consistent with the RNA-seq-based analyses, CAGE-seq demonstrated that transcripts from the LTR12 family TEs were globally activated in TNRC18(W2858A)-mutant cells compared with WT cells (Fig. 4d). For example, the DHRS2, ACSBG1, ETV7 and SYNGR1 transcripts that have LTR12 as TSSs were all upregulated in the TNRC18(W2858A)-mutant cells compared with control cells (Fig. 4e and Extended Data Fig. 7i). Another example is the TSSs of TNFRSF10B, which contain a canonical TSS and a cryptic one residing in a nearby LTR12C locus<sup>42</sup>, and HDAC inhibitor treatment activated the latter TSS<sup>42</sup>. In TNRC18(W2858A)-mutant cells, TNFRSF10B transcription was activated from the LTR12C-associated cryptic TSS, without having an obvious effect on that from canonical TSS (Fig. 4f). Additionally, the TNRC18W2858A mutation activated the transcription of LTR12C located at noncoding RNA TSSs such as lncRNA11-11N5.1 and lncRNA11-299H22.1 (Extended Data Fig. 7j). Collectively, we revealed a BAH-dependent repressive role of TNRC18 in the transcription of LTRs and LTR-regulated nearby genes.

### Tnrc18 mutation causes death in mice

Activation of ERVs, interferon-stimulated genes and cryptic transcripts in TNRC18(BAH)-mutant cells implied a wider functionality for TNRC18. To further evaluate the functional relevance of TNRC18(BAH) in development, we used a CRISPR-Cas9-based approach to produce heterozygous mice harbouring an H3K9me3-binding-defective mutation of Tnrc18BAH (mouse Tnrc18W2745A/Y2747A, equivalent to human TNRC18W2858A/Y2860A; Extended Data Fig. 8a). Breeding of Tnrc18W2745A/Y2747A heterozygous mice produced homozygous mutant pups at birth at a rate of 8% (13 out of a total of 161), which was significantly lower than the expected frequency (Fig. 4g). Additionally, the surviving homozygous mutant mice, both males and females, exhibited a dwarfism phenotype compared with their WT or heterozygous littermates (Extended Data Fig. 8b).

To further assess the function of TNRC18 in mouse tissues, we isolated five different primary tissues (lung, liver, heart, brain and stomach) from WT mice and paired littermates carrying the Tnrc18BAH homozygous mutation. RNA-seq demonstrated strong ERV activation in the lung and liver, less so in the heart and stomach, and no significant change in the brain of Tnrc18BAH mutant mice compared with that of WT mice (Fig. 4h, Extended Data Fig. 8c and Supplementary Table 4). Most reactivated TEs were known targets of SETDB1. RNA-seq also confirmed that the same immunity-related pathways were substantially activated in the lung and liver from Tnrc18-mutated compared with WT mice (Fig. 4i, Extended Data Fig. 8d,e and Supplementary Table 4). We next established primary mouse embryonic fibroblasts (MEFs) from WT and Tnrc18BAH homozygous mutant embryos (Extended Data Fig. 8f). RNA-seq analyses showed that ERVs were significantly activated in Tnrc18BAH mutant MEFs compared with WT MEFs (Extended Data Fig. 8g and Supplementary Table 4). Concurrent with ERV derepression was activation of interferon-stimulated and immunity-related genes (Extended Data Fig. 8h and Supplementary Table 4). These observations suggest that the loss-of-function mutation of Tnrc18BAH has widespread effects on deregulating TEs and genes in multiple tissues and organs during development, which may explain the embryonic death and dwarfism seen with Tnrc18-mutated mice.

## TNRC18 recruits HDACs and co-repressors

Next, we sought to explore the mechanism underlying TNRC18-mediated ERV silencing. BioID identified TNRC18-associated proteins in HEK293 cells, which included major components of the Sin3–HDAC complex, the NCoR–HDAC complex<sup>43</sup> and the HDAC-associated CCT complex<sup>44</sup> (Fig. 5a and Supplementary Table 5). TRIM28, a hub protein known to mediate ERV silencing<sup>26,45</sup>, and SUMOylation-related proteins, which regulate viral silencing activities of TRIM28 (ref. 26), were also identified in the BioID assay (Fig. 5a and Supplementary Table 5). Additional BioID for TNRC18 in HeLa cells produced similar results to those for HEK293 cells (Extended Data Fig. 9a and Supplementary Table 5). Co-immunofluorescence showed TNRC18 overlapped with TRIM28 (Pearson  $r = 0.681$ ) and HDAC2 (Pearson  $r = 0.756$ ) (Fig. 5b and Extended Data Fig. 9b). The Sin3–HDAC complex binds partner proteins through interactions between the paired amphipathic helix (PAH) domain of Sin3 and the Sin3-interacting motif of the partner (enriched for a consensus LXXLL-like sequence, with X indicating any amino acid)<sup>46</sup>. On the basis of sequence prediction, we synthesized peptides for each of the potential Sin3-interacting motifs within TNRC18 (Supplementary Table 7) and used ITC to assess their potential binding to the PAH domain of Sin3. Among these, amino acids 718–792 of TNRC18 (Fig. 5c left, and Extended Data Fig. 9c; named as a Sin3-interaction motif of TNRC18) bound to the PAH domain of Sin3 with a  $K_d$  of 104.5  $\mu\text{M}$ . By contrast, two independent mutations of this fragment largely abolished its binding to the PAH of the Sin3 domain (Fig. 5c, right). Next, we focused on TNRC18(L760), a residue located at the centre of the Sin3-interaction motif (Extended Data Fig. 9c) and crucial for efficient binding to the PAH domain of Sin3 (Fig. 5c, right). Here we used the CRISPR–Cas9 technique to introduce a TNRC18L760A mutation into HEK293 cells (Extended Data Fig. 9d). RNA-seq profiling of TNRC18(L760A)-mutant cells and TNRC18 WT cells showed reactivation of ERV family TEs in the mutant cells (Extended Data Fig. 9e and Supplementary Table 6). RT–qPCR analyses showed that there was approximately 1.5–2-fold activation of LTR12 and LTR7 family TEs (Extended Data Fig. 9f). Such TE reactivation resembles what was seen with TNRC18 KO or its BAH domain mutation, although the degree of TE reactivation in the Sin3-binding mutant (L760A) seemed milder. Nevertheless, the same immunity-related gene sets were activated in TNRC18(L760A)-mutant cells compared with TNRC18 WT cells, as seen with TNRC18 KO (Extended Data Fig. 9g,h and Supplementary Table 6). Collectively, these results establish a crucial requirement for TNRC18 and the associated Sin3–HDAC complex for mediating ERV repression.

Consistent with the co-immunofluorescence results (Fig. 5b and Extended Data Fig. 9b), CUT&RUN for HDAC, TRIM28 and SETDB1 showed that all of them colocalized well with TNRC18-bound genomic regions (Fig. 5d and Extended Data Fig. 9i), as exemplified by their co-binding at ITGB3BP, ZNF180 and ZNF260 (Supplementary Fig. 9j). Of note, motif analysis of TNRC18 CUT&RUN peaks also revealed the DNA-binding motifs of the TRIM28-associated KRAB–ZnF proteins such as ZFP809 (ref. 47) (Extended Data Fig. 9k). These observations not only point to functional cooperation between TRIM28 and TNRC18 but also indicate the potential involvement of DNA-binding ZNFs in mediating and/or stabilizing binding of TNRC18 to genomic targets. SETDB1 KO in HEK293 cells

(Extended Data Fig. 10a) led to a partial but significant decrease in TNRC18 levels in these cells compared with WT cells (Extended Data Fig. 10b, left; see anti-TNRC18). This effect is reminiscent of the observed decreased H3K9me3 and H3K9me2 regulator levels following depletion of the associated partner Rif1 in stem cells<sup>48</sup>. As expected, CUT&RUN showed a global decrease in the genomic binding of TNRC18 in SETDB1-depleted cells compared with WT cells (Extended Data Fig. 10c). Although the underlying mechanism of this effect remains unclear, the observed partial loss-of-function of TNRC18 after SETDB1 loss further supports a functional link between the two. Next, we sought to determine whether TNRC18 functions to recruit co-repressors to genomic sites such as ERVs for silencing. Towards this end, we conducted CUT&RUN for TNRC18, ERV co-repressors (HDAC2, TRIM28 and SETDB1) and histone modifications (H3K9me3 and H3 acetylation (H3ac)) in WT and TNRC18W2858A-mutated HEK293 cells. Genome-wide binding of TNRC18 was significantly decreased in the TNRC18W2858A-mutated cells relative to WT cells (Fig. 5e, red). There were concurrent decreases in binding by three examined co-repressors (HDAC2, TRIM28 and SETDB1) at TNRC18-bound genomic sites in the TNRC18-mutant cells compared with WT cells (Fig. 5e). The total SETDB1 protein level was unaffected following TNRC18 KO or mutation to its BAH domain (Extended Data Fig. 10b, right). Similarly, global H3K9me3 and H3ac showed no changes, as assessed by immunoblotting and CUT&RUN (Extended Data Fig. 10d–f). However, H3ac levels at LTR12 subclasses (LTR12C, LTR12D and LTR12E), but not other TEs (such as LINEs and SINEs), were increased in TNRC18W2858A-mutated cells compared with WT cells. By contrast, H3K9me3 levels at these LTR12s and not LINEs or SINEs were decreased (Fig. 5f), as exemplified in what was seen at the LTR12C locus on chromosome 14 (Fig. 5g). Whole-genome bisulfite sequencing showed that the TNRC18BAH mutation did not change DNA methylation levels, either globally (Extended Data Fig. 10g) or at LTR12 and LTR7 families (Extended Data Fig. 10h). Thus, the TNRC18BAH mutation seems to mainly affect the histone modification status (relative levels of H3K9me3 compared with H3ac), instead of DNA methylation, at ERVs.

To further validate the observed repressive function of TNRC18, we used the GAL4 DNA-binding domain (DBD)-fusion-based reporter assay (Extended Data Fig. 10i). Relative to mock, TNRC18 induced a significant transcriptional repression effect (Extended Data Fig. 10i). In agreement with the CUT&RUN results, chromatin immunoprecipitation with qPCR (ChIP–qPCR) revealed that concurrent with the expected increase in TNRC18 binding to the GAL4 target sites, the binding levels of HDAC2, TRIM28 and SETDB1 were all significantly increased, whereas those of H3Kac were decreased (Fig. 5h).

Together, these results support a model in which H3K9me3 recognition by TNRC18—potentially synergistic with the engagement of DNA-binding factors (for example, ZNFs)—help recruit and/or tether TNRC18 and its associated complex onto target sites (such as ERVs) to maintain a repressed chromatin state.

## Discussion

This study reports that TNRC18(BAH) is a mammalian H3K9me3 reader, and our biochemical and structural characterization demonstrated that such specificity to H3K9me3

engagement partially stems from various interactions of TNRC18(BAH) with histone backbone sequences adjacent to H3K9me3, such as H3T6 and H3S10. Genome-wide profiling of TNRC18 showed enriched binding to ERVs, notably the ERV1 family TEs such as LTR12 and LTR7. Both TNRC18 loss and an H3K9me3-binding-defective mutation of TNRC18(BAH) led to significantly activated expression of ERV1, inducing viral mimicry. Preferential silencing effects by TNRC18 on ERVs are in contrast to those induced by the HUSH complex on LINE1s<sup>3,4</sup>. This result indicates that there is a division of labour among various H3K9me3 reader complexes in ensuring repression of various TE families. Significant enrichment in CCAAT motifs at TNRC18-bound sites, as well as ZNFs identified through TNRC18 BioID, suggest that additional DNA sequence-specific factors (in addition to chromatin-associating factors) are likely to mediate genomic recruitment of TNRC18 (for details, see Supplementary Note 1). These observations are consistent with the notion of an ‘arms race’ in which various epigenetic machineries and associated DNA-binding ZNFs are deployed to silence and ‘tame’ viral insertions during evolution<sup>17,18,49,50</sup>. TNRC18 depletion, or mutagenesis of the H3K9me3-binding TNRC18(BAH) domain, in a range of different model systems resulted in activation of LTR12s, LTR12-associated cryptic TSSs and immune-related genes. Moreover, mice harbouring an H3K9me3-binding-defective mutation in TNRC18(BAH) exhibited severe phenotypes of neonatal death and dwarfism in adult surviving animals. Overall, this study demonstrates that a TNRC18(BAH)-based H3K9me3-sensing pathway operates to modulate the landscape of promoter–enhancer activities by controlling TEs and the embedded cis-elements, and that TNRC18 is a crucial regulator of immunity and development.

### Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-023-06688-z>.

## Methods

### Cell lines and tissue culture

Cell lines used in this work included HEK293 from the American Type Culture Collection (ATCC, CRL-1573), HEK293T (ATCC, CRL-3216), HeLa (ATCC, CCL-2), K562 (ATCC, CCL-243), 22Rv1 (ATCC, CRL-2505), SNU-1 (ATCC, CRL-5971), HT-29 (ATCC, HTB-38) and NCI-H23 (ATCC, CRL-5800). Cells were cultured and maintained following the vendor-recommended protocols. Authentication of identity of cells, including parental and derived lines, was ensured by using genetic signature profiling and fingerprinting analyses<sup>51</sup>. Generally, we used cells at a passage of fewer than ten times. Every month, we examined the cell lines in culture for possible mycoplasma contamination using a MycoAlert Plus Mycoplasma Detection kit (Lonza, LT27–286).

### Plasmids and antibodies

The sgRNAs targeting representative ERV loci were cloned into pLentiGuidePuro (Addgene, 52963). The pLKO.1-puro-based shRNA plasmids for KD of TNRC18 were

purchased from Sigma. Detailed information of sgRNA and shRNA is provided in Supplementary Table 7. Full-length TNRC18 cDNA was synthesized (Wuxi Qinglan Biotech), with synonymous mutations introduced to confer shRNA resistance or to create unique internal restriction enzyme sites for facilitating mutagenesis of TNRC18(BAH). The synthesized full-length TNRC18 cDNA was first ligated into a PiggyBac Transposase Expression vector, PB-EF1 $\alpha$ -MCS-IRES-Neo (System Biosciences, PB533A-2). In-frame fusion of a GFP tag (which was used for immunofluorescence (IF), immunoblotting and CUT&RUN), BirA\* (used for BioID) or the DBD of GAL4 (GAL4-DBD) with TNRC18 was conducted by ligating the corresponding PCR product into a unique restriction enzyme site flanking either the N or C terminus of TNRC18 in the above PiggyBac-EF1 $\alpha$  expression plasmid. The design of the TNRC18 mammalian expression vector is illustrated in Extended Data Fig. 1c. The pcDNA3.1-based mammalian expression plasmid for TNRC18 was created by subcloning. All of the used plasmids were verified by Sanger sequencing. The antibodies used for immunoblotting, IF, co-IP and CUT&RUN are listed in Supplementary Table 7.

### IF and colocalization analysis

Cells were grown on polylysine-coated coverslips (Corning, 354085) overnight in a 37 °C incubator. The cover slips were washed with PBS and then fixed in 4% formaldehyde for 10 min at room temperature. Fixed cell samples were washed with cold PBS three times and incubated in PBS plus 0.1% Triton X-100 for 10 min, followed by washing with PBS three times and incubation in blocking buffer (1% BSA in PBS plus 0.1% Tween-20) for 30 min. After removal of the blocking buffer, the fixed samples were then incubated with diluted primary antibody (anti-TNRC18 1:1,000, anti-Flag 1:1,000, anti-GFP 1:1,000; see the Nature Portfolio Reporting Summary for detailed information) in blocking buffer overnight at 4 °C in a humidified chamber, and washed with PBS plus 0.1% Tween-20 three times (3 min each time). Last, the samples were incubated with secondary antibody (diluted at 1:2,000) conjugated to appropriate fluorophore for 1 h at room temperature and washed three times with PBS before adding the mounting medium (Thermo Scientific, P36935). The slides were then dried overnight in the dark before imaging on an Olympus FV1000 confocal microscope with a  $\times 100/1.4$  NA plan apochromat oil-immersion objective.

Colocalization analysis was performed using the EzColocalization plugin in Fiji (v.1.53)<sup>52</sup>. Colocalization was measured using the Pearson correlation coefficient as previously described<sup>53,54</sup>. For analysis, nuclei were manually segmented by tracing with the polygon selection tool, and then converted into binary masks used in the EzColocalization plugin to restrict colocalization analysis to the nuclei.

### Immunoblotting

Samples of total protein were prepared using lysis buffer (50 mM Tris-HCl (pH 8.0), 150 mM NaCl and 1% IGEPAL CA-630 detergent), followed by brief sonication and centrifugation. The same amount of extracted sample was loaded onto a SDS-PAGE gel for immunoblot analysis. All antibodies for immunoblotting were diluted 1:1,000. Detailed information for antibodies is provided in the Nature Portfolio Reporting Summary. Uncropped scans of blots are shown in Supplementary Fig. 1.

## IP

Samples of total protein sample were prepared as described above, followed by brief sonication and high-speed centrifugation. Extracted samples were incubated with primary antibodies (2  $\mu$ g) on a rotator overnight at 4 °C. Dynabeads (Invitrogen) were added and incubated for an additional 2 h. Beads were washed three times in lysis buffer, resuspended in 50  $\mu$ l of 2 $\times$  sample loading buffer and boiled at 90 °C for 5 min before loading onto a SDS-PAGE gel. Western blotting was performed using standard protocols with a PVDF membrane, and signals were visualized using an ImageQuant LAS 4000 Luminescent Image Analyzer (GE Healthcare). Repeated immunoblotting using either freshly prepared total protein samples or those from anti-TNRC18 IP detected full-length TNRC18 and the two TNRC18 protein cleavage products (with sizes of about 220 and 80 kDa) that interact with each another (data not shown), reminiscent to what was reported for MLL (also known as KMT2A)<sup>55</sup>.

## RT-qPCR

Total RNA (1  $\mu$ g) was extracted from cells or tissues using a RNeasy Mini kit (Qiagen, 74106; with on-column digestion for 30 min by DNase set, Qiagen, 79254), and RT was carried out using a High Capacity cDNA Reverse Transcription kit (Applied Biosystems, 4368814) according to the provider's protocols. qPCR was run in triplicate using SYBR green master mix reagent (Bio-Rad, 1725125) on an ABI 7900HT fast real-time PCR system (ABI). Data from at least three independent experiments are presented as the mean  $\pm$  s.d. after normalization against internal control signals. Information on primers used for RT-qPCR is provided in Supplementary Table 7.

## CasID

Cells with stable expression of the dCas9-BirA\* fusion protein were transfected with pooled plasmids of pLentiGuidePuro that contain a set of sgRNAs targeting the representative ERV loci (for details, see Supplementary Table 7) or empty vector (which served as the control group). After treatment with 50  $\mu$ M biotin for 24 h, cells were collected from two 15-cm plates and then washed twice with cold PBS. Cell pellets were resuspended in 1 ml RIPA lysis buffer (10% glycerol, 25 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM EDTA, 0.1% SDS, 1% NP-40 and 0.2% sodium deoxycholate), and 1  $\mu$ l benzonase (Sigma-Aldrich, E1014) was added to the lysate and incubated on ice for 1 h. After centrifugation at maximum speed for 30 min at 4 °C, the supernatant was collected and incubated with Neutravidin beads (Thermo Fisher, 29204) overnight at 4 °C. The Neutravidin beads were washed twice with RIPA buffer and TAP lysis buffer (10% glycerol, 350 mM NaCl, 2 mM EDTA, 0.1% NP-40 and 50 mM HEPES, pH 8.0) sequentially. Finally, the beads were washed three times with ABC buffer (50 mM ammonium bicarbonate, pH 8.0) and subjected to mass-spectrometry (MS)-based analysis. We focused on the top 300 hits showing the most significant enrichment in CasID versus control, which contained many of the previously identified TE regulators based on reporter-based assays<sup>23,26–28</sup>, and suggested that our CasID approach works well. We scrutinized top hits by annotating their domain composition using the Simple Modular Architecture Research Tool (<http://smart.embl-heidelberg.de/>),

which showed that TNRC18, an under-studied protein, contains the putative chromatin-binding domain, namely, BAH studied in this work.

## BioID

A BirA\* cDNA fragment was inserted in-frame into the PB-EF1 $\alpha$ -based TNRC18 expression plasmid, followed by the establishment of stable expression lines in HEK293 and HeLa cells. BioID was performed as previously described<sup>56,57</sup>. In brief, cells, cultured in two 15-cm plates, treated with 50  $\mu$ M biotin for 24 h and then washed twice with cold PBS. Cell pellets were resuspended in 1 ml RIPA lysis buffer and 1  $\mu$ l benzonase (Sigma-Aldrich, E1014) was added to the lysate followed by incubation on ice for 1 h. After centrifugation at top speed for 30 min at 4 °C, the supernatant was collected and incubated with Neutravidin beads (Thermo Fisher, 29204) overnight at 4 °C. The Neutravidin beads were washed twice with RIPA buffer and TAP lysis buffer sequentially. Finally, the beads were washed three times with ABC buffer and subjected to MS-based analysis. The identified BioID hits were further analysed using STRING<sup>58</sup>.

## MS-based protein identification

Proteins were eluted from beads by adding 50  $\mu$ l 2 $\times$  Laemmli buffer and heating at 95 °C for 5 min. A total of 50  $\mu$ l of each sample was resolved by SDS-PAGE using a 4–20% Tris-glycine wedge well gel (Invitrogen) and visualized by Coomassie staining. Each SDS-PAGE gel lane was sectioned into 12 segments of equal volume. Each segment was subjected to in-gel trypsin digestion as follows. Gel slices were destained in 50% methanol, 50 mM ammonium bicarbonate (Sigma-Aldrich), followed by reduction in 10 mM Tris[2-carboxyethyl] phosphine and alkylation in 50 mM iodoacetamide. Gel slices were then dehydrated in acetonitrile, followed by the addition of 100 ng porcine sequencing-grade modified trypsin (Promega) in 50 mM ammonium bicarbonate and incubation at 37 °C for 12–16 h. Peptide products were then acidified in 0.1% formic acid. Tryptic peptides were separated by reverse-phase XSelect CSH C18 2.5  $\mu$ m resin (Waters) on an in-line 150  $\times$  0.075 mm column using a nanoAcquity UPLC system (Waters). Peptides were eluted using a 60 min gradient from 98:2 to 65:35 buffer A:B ratio (buffer A, 0.1% formic acid, 0.5% acetonitrile; buffer B, 0.1% formic acid, 99.9% acetonitrile). Eluted peptides were ionized by electrospray (2.2 kV) followed by MS/MS analysis using higher-energy collisional dissociation (HCD) on an Orbitrap Fusion Tribrid mass spectrometer (Thermo) in top-speed data-dependent mode. MS data were acquired using the FTMS analyser in profile mode at a resolution of 240,000 over a range of 375 to 1,500 m/z. Following HCD activation, MS/MS data were acquired using the ion trap analyser in centroid mode and normal mass range with precursor mass-dependent normalized collision energy between 28.0 and 31.0. Proteins were identified by searching the UniProtKB database restricted to Homo sapiens using Mascot (Matrix Science) with a parent ion tolerance of 3 ppm and a fragment ion tolerance of 0.5 Da, fixed modifications for carbamidomethyl of cysteine, and variable modifications for oxidation on methionine and acetyl on the N terminus. Scaffold (Proteome Software) was used to verify MS/MS-based peptide and protein identifications. Peptide identifications were accepted if they could be established with less than 1.0% false discovery by the Scaffold Local false discovery rate algorithm. Protein identifications were accepted if they could be established with less than 1.0% false discovery and contained at

least two identified peptides. Protein probabilities were assigned using the Protein Prophet algorithm<sup>59</sup>. Proteins were filtered out if they had a spectral count <8 in all sample groups and the counts were normalized to log2NSAF values.

### Protein expression and purification

DNA encoding the BAH domain of human TNRC18 (amino acids 2785–2968, TNRC18(BAH)) was inserted into a modified pRSF-Duet vector preceded by an N-terminal hexahistidine (His6)-SUMO tag and a ubiquitin-like protease 1 (ULP1) cleavage site. BL21(DE3) RIL cells harbouring the expression plasmids were induced through the addition of 0.4 mM isopropyl  $\beta$ -d-1-thiogalactopyranoside (IPTG) when the cell density reached an optical density of 0.8 at 600 nm and continued to grow at 16 °C overnight. Cells were collected and lysed in buffer containing 50 mM Tris-HCl (pH 8.0), 1 M NaCl, 25 mM imidazole, 10% glycerol and 1 mM PMSF. Subsequently, the fusion protein was purified through a nickel column, followed by removal of the His6-SUMO tag by ULP1 cleavage, ion-exchange chromatography on a heparin column (GE Healthcare) and size-exclusion chromatography on a HiLoad 16/600 Superdex 75 pg column (GE Health-care). The purified protein samples were concentrated in 20 mM Tris-HCl (pH 7.5), 100 mM NaCl, 5% glycerol and 5 mM DTT, and stored at –80 °C. The constructs for various TNRC18(BAH) domain mutants were produced by site-directed mutagenesis, and purification of these mutant proteins was conducted in the same way as described above.

### Crystallization and data collection

For crystallization, 1.5 mM TNRC18(BAH) was mixed with the H3K9me3-containing peptide, comprising residues 1–22 of histone H3 and a C-terminal tryptophan (for the purpose of peptide quantification), in a molar ratio of 1:3. Subsequently, the protein complex was incubated with 0.1 M sodium cacodylate (pH 6.5), 0.2 M magnesium acetate and 16–20% PEG8000 using the hanging drop vapour-diffusion method at 4 °C. The crystals were soaked in the well solution supplemented with 25% (v/v) glycerol before being flash-frozen in liquid nitrogen. X-ray diffraction data were collected on beamline 24-ID-E at Advanced Photon Source (APS), Argonne National Laboratory. The datasets were processed using the HKL3000 program (v.721.3)<sup>60</sup>. The structure of the TNRC18(BAH)–H3K9me3 complex was solved by molecular replacement with PHASER (v.2.7.16)<sup>61</sup> using the structure of mouse ORC1(BAH) (PDB identifier 4DOV) as a search model. Iterative cycles of model rebuilding and refinement were carried out using COOT (WinCoot 0.8.9 EL)<sup>62</sup> and PHENIX (V.1.20.1\_4487)<sup>63</sup>, respectively. Structural figures were generated using PyMol (v.2.5.2). Statistics for data processing and structure refinement are summarized in Extended Data Table 1.

### ITC

WT or mutant TNRC18(BAH) protein and peptides were respectively dialysed against buffer containing 25 mM Tris-HCl (pH 7.5), 100 mM NaCl and 1 mM DTT at 4 °C overnight. For ITC titration, the syringe and well contained 2 mM peptide and 0.2 mM TNRC18(BAH), respectively. A MicroCal iTC200 system (GE Healthcare) was used to conduct the ITC measurements. A total of 15–17 injections with a spacing of 180 s and a reference power of 5  $\mu$ cal s<sup>-1</sup> were performed at 7 °C. The ITC curves were processed



using the software ORIGIN (v.7.0) (MicroCal) with one-site fitting model as previously described<sup>36,64</sup>.

### Preparation of chemically modified nucleosome

The NCP carrying a mimic of H3K9me3 modification (H3Kc9me3) was prepared using previously reported protocols<sup>65,66</sup>. In essence, histones (H2A, H2B, K9C/C110S-mutated H3 and H4) from *Xenopus laevis* were expressed in *Escherichia coli* BL21 (DE3) RIL and purified through ion-exchange chromatography using sequential Q-XL and SP-HP (GE Healthcare) columns in a denaturing buffer (20 mM Tris-HCl pH 7.5, 7 M urea and 1 mM  $\beta$ -mercaptoethanol) and a salt gradient of 0 to 1 M KCl. Purified histone proteins were thoroughly dialysed against distilled water containing 2 mM  $\beta$ -mercaptoethanol, lyophilized and stored at  $-80^{\circ}\text{C}$ . The K9C-mutated H3 protein was alkylated as previously described<sup>67</sup>. Subsequently, the histone octamer was reconstituted in refolding buffer (20 mM Tris-HCl pH 7.5, 2 M NaCl, 1 mM EDTA and 5 mM 2-mercaptoethanol) and purified by size-exclusion chromatography on a HiLoad Superdex 200 column. 5'-biotin-labelled Widom 601 DNA (biotin-CTCTCTCCGTAAACATGCTGGAGAAATCGAGAATCCCGGTGCCGAGGCCGCTCAATTGGTCGTAGACAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTCCCCGCGTTTAAACCGCCAAGGGGATTACTCCCTAGTCTCCAGGCACGTGTCAGATATATACATCCGAT; Widom 601 DNA sequence is underlined) was PCR amplified. The NCP was assembled using the purified histone octamer mixed with the Wisdom 601 DNA in a 1:1 molar ratio, followed by step-wise dialysis against a buffer containing 10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 1 mM DTT and 0.25–2 M KCl. Finally, the assembled NCP was dialysed against a low-salt buffer containing 10 mM Tris-HCl (pH 7.5) and 1 mM DTT overnight. Biotinylated 601 DNA or biotinylated/H3Kc9me3-modified NCP were analysed in 5% Tris-borate-EDTA (TBE) native gel in 0.2 $\times$  TBE buffer (89 mM TBE, pH 8.3) at  $4^{\circ}\text{C}$  and ran at 100 V for about 30 min. The DNA band was visualized using SYBR Gold staining and scanned using a ChemiDoc imager (Bio-Rad).

### Biolayer interferometry

The binding affinity of TNRC18(BAH) for H3Kc9me3-modified NCP was measured at  $30^{\circ}\text{C}$  by biolayer interferometry on a gator prime instrument (Gatorbio) at a shaking speed of 1,000 r.p.m. using purified TNRC18(BAH) and H3Kc9me3-labelled NCP as described above. The binding buffer contained 20 mM Tris-HCl (pH 7.5), 100 mM NaCl, 1 mM DTT and 0.01% Tween-20. Before the experiments, all biosensors were pre-equilibrated in the binding buffer for 10 min. Assembled nucleosome was diluted to 10 nM to be immobilized on streptavidin biosensors (Gatorbio) to a shift of around 1.0. After a 120-s baseline step in binding buffer, the ligand-loaded biosensors were submerged in various concentrations (2-fold serial dilution from 50 nM to 3.1 nM) of TNRC18(BAH) protein solutions for 120 s and then transferred into the binding buffer for 600 s to measure TNRC18(BAH) association and dissociation kinetics. Data were aligned, inter-step corrected to the association step and further analysed using the Gatorbio data analysis software. The data were fitted using 1:1 mode and results were plotted in GraphPad Prism (v.9.1.0). Experiments were repeated two times to ensure consistency.

## CRISPR–Cas9-based genome editing for KO or site-specific mutation of TNRC18

For KO of TNRC18 in cells, we used the Alt-R CRISPR–Cas9 system in combination with a pair of targeting crRNAs. For introduction of a point mutation in TNRC18 in cells, we used the same Alt-R CRISPR–Cas9 system in combination with one crRNA and one single-stranded oligodeoxynucleotide donor (ssODN). The Alt-R CRISPR–Cas9 system includes a trans-activating CRISPR RNA (tracrRNA) with ATTO 550, a crRNA targeting the site to be mutated (the sequence information of crRNA is listed in Supplementary Table 7), S.p. HiFi Cas9 nuclease V3 and an electroporation enhancer (all purchased from IDT). The ssODN was custom designed (the sequence information of the ssODN is listed in Supplementary Table 7; <https://benchling.com/crispr/>) and then ordered from commercial sources (IDT). Preparation and delivery of a CRISPR–Cas9 ribonucleoprotein (RNP) complex (Alt-R CRISPR–Cas9 crRNA–tracrRNA and S.p. HiFi Cas9 nuclease), electroporation enhancer and ssODN into cells were performed according to the manufacturer's guidelines. In brief, crRNA and tracrRNA were mixed in equimolar concentrations, followed by heating at 95 °C for 5 min. The mixture was cooled to room temperature (15–25 °C) on the bench top, allowing the crRNA–tracrRNA duplex to form. The RNP complex was made by diluting the crRNA–tracrRNA duplex and Cas9 enzyme components in PBS, followed by incubation at room temperature for 10–20 min. The RNP complex mixed with the electroporation enhancer and ssODN (for KO, no ssODN is required) was delivered to cells using electroporation with an Amaxa Cell Line Nucleofector kit (Lonza). At 18 h after electroporation, ATTO-550-positive cells were sorted by FACS (UNC flow core) and split into 96-well plates to establish clonal lines for PCR and sequencing-based genotyping. After genotyping, lines with homozygous mutation/deletion were further validated at both DNA and RNA levels by Sanger sequencing of PCR products that covered the mutation/deletion site.

### RNA-seq and data analysis

Total RNA was extracted using a RNeasy Mini kit (Qiagen, 74106) (digested on column for 0.5 h by DNase set, Qiagen, 9254) and then prepared using a NEBNext rRNA Depletion kit (NEB, E6310L) and a NEB Ultra II DNA Library Prep kit (NEB, E7103) per the manufacturer's instructions<sup>57,68</sup>. The multiplexed RNA-seq libraries were subjected to deep sequencing using an Illumina NextSeq500 or NovaSeq 6000 sequencer platform. FastQC was used for quality control of high-throughput raw sequencing data.

The fastq files of human cell and mouse cell RNA-seq data were aligned to the GRCh38 human genome (GRCh38.d1.vd1.fa) and the mm10 mouse genome, respectively, using STAR (v.2.4.2)<sup>69</sup> with the following parameters: --outSAMtype BAM Unsorted --quantMode TranscriptomeSAM. The transcript abundance for each sample was estimated using salmon (v.1.4.0)<sup>70</sup> to quantify the transcriptome defined by Gencode v.22. Quality control was performed using Picard (v.2.22.4) CollectRnaSeqMetric and FastQC (v.0.11.9). Gene level counts were summed across isoforms, and genes with low counts (maximum expression < 10) were filtered out from downstream analyses. We tested genes for differential expression in DESeq2 (v.1.38.2)<sup>71</sup> in R, and GSEA on Molecular Signatures database (MSigDB) C2 curated gene sets<sup>72</sup> was conducted using GSVA (v.1.30.1)<sup>73</sup> in R.

To comprehensively quantify expression by all repetitive elements, including TEs in RNA-seq data, we used REdiscoverTE (v.1.0.1)<sup>74</sup>, which annotates the genes defined in Gencode and all RepeatMasker sequences in the human genome down to the TE subfamily level after the repetitive element expression quantification by salmon. TE count matrix was appended to gene expression data matrix for differential expression analysis using DESeq2 (v.1.38.2)<sup>71</sup>, and TEs showing significant differential expression were defined as having adjusted P values of less than 0.05 and a log<sub>2</sub> value of fold change greater than 1 in the experimental versus control group (with a base mean value greater than 10).

Because REdiscoverTE is specific to human, we used SalmonTE (v.0.4)<sup>75</sup> to calculate the TE abundance in the mouse samples. By performing read mapping against the TE sequences, SalmonTE reassigns multi-mapping reads using the expectation-maximization algorithm before determining the read count of each sequence. From the counts table, the differentially expressed TEs were identified and the statistical analysis function of SalmonTE generated the summary of each class and clade through the generalized linear model.

### CAGE-seq and data analysis

CAGE-seq library preparation, sequencing, mapping, gene expression, motif discovery analysis and gene ontology (GO) enrichment analysis were performed by DNAFORM (Kanagawa, Japan). Qualities of total RNA were assessed using Bioanalyzer (Agilent) to ensure that the RNA integrity number was greater than 7.0. The cDNAs were synthesized from total RNA using random primers. The ribose diols in the 5' cap structures of RNAs were oxidized and then biotinylated. The biotinylated RNA/cDNAs were selected using streptavidin beads (cap-trapping). After RNA digestion using RNaseONE/H and adaptor ligation to both ends of cDNA, double-stranded cDNA libraries (CAGE libraries) were constructed. CAGE libraries were sequenced using single-end reads of 75 nucleotides on a NextSeq 500 instrument (Illumina). Obtained reads (CAGE tags) were mapped to the human hg19 genome using BWA (v.0.7.17). Unmapped reads were then mapped using HISAT2 (v.2.0.5). CAGE tag clustering, detection of differentially expressed genes and motif discovery were performed by pipeline RECLU (v.1.0)<sup>76</sup>. Tag count data were clustered using the modified Paraclu program. Clusters with count per million < 0.1 were discarded. Regions that had 90% overlap between replicates were extracted using BEDtools (v.2.12.0). The clusters with an irreproducible discovery rate > 0.1 and clusters longer than 200 bp were discarded. Differentially expressed genes were detected using the edgeR package (v.3.22.5). TEs were annotated using the RepeatMasker bed file from the UCSC table browser. Read counts were normalized by library size, and enrichment scores for each TE family were computed as the normalized count ratio between mutant and WT samples.

### CUT&RUN and data analysis

CUT&RUN<sup>77</sup> was performed as previously described<sup>78,79</sup> with a commercially available kit according to the manufacturer's instruction (EpiCypher CUTANA pAG-MNase for ChIC/CUT&RUN, 15–1116). In brief, 1 million cells were first collected, washed in CUT&RUN wash buffer (20 mM HEPES, pH 7.5, 150 mM NaCl, 0.5 mM spermidine and 1× complete protease inhibitor cocktail), and then bound to activated ConA beads (Bangs Laboratories,

BP531). Next, the cell-bead sample was incubated with antibodies against the protein target (all antibodies were added at 1:100 dilution, also refer to the Nature Portfolio Reporting Summary for detailed information) and then permeabilized in digitonin-containing buffer (CUT&RUN wash buffer plus 0.02% digitonin), which was then followed by washing in the digitonin buffer, incubation with pAG-MNase and another wash in the digitonin buffer to remove unbound pAG-MNase. After the final wash, cells were subjected to digestion (after pAG-MNase activation) through the addition of pAG-MNase digestion buffer (digitonin buffer plus 2 mM CaCl<sub>2</sub>), followed by incubation on a nutator for 2 h at 4 °C. Solubilized chromatin was then released using CUT&RUN stop buffer (340 mM NaCl, 20 mM EDTA, 4 mM EGTA, 50 µg ml<sup>-1</sup> RNase A, 50 µg ml<sup>-1</sup> glycogen and 0.2 µg fly genomic DNA) and DNA purification was carried out using a PCR cleanup kit (NEB, Monarch PCR & DNA Cleanup kit, T1030). Next, 10 ng of the purified CUT&RUN DNA was used for preparation of multiplexed libraries with a NEB Ultra II DNA Library Prep kit per the manufacturer's instruction (NEB, E7103). Sequencing was conducted using an Illumina NextSeq 500 sequencing system.

Fastq files were mapped to the reference genome (hg19) using bow-tie (v.2.3.4.1)<sup>80</sup>. The non-primary alignment and PCR duplicates were removed from aligned data using Samtools (v.1.9) (--mapq 40 for unique mapping reads), the Picard 'MarkDuplicates' function (v.2.22.04), and bedtools (v.2.28.0), respectively. Peak calling was performed using MACS2 (v.2.2.7.1) (macs2 callpeak -f BAMPE -g hs/mm --keep-dup 1 -q 1e-5 -c IgG.control.bam)<sup>81</sup>.

Distribution of peaks was analysed using the R package ChIPpeakAnno (v.3.34.1)<sup>82</sup> (bindingType="fullRange", bindingRegion=c(-1,1), ignore.peak.strand=TRUE, select="bestOne"). By using the default parameters of ChIPpeakAnno, 75% of the TNRC18 peaks called by MACS2 (7,242 of a total of 9,656) overlapped TEs (data not shown). However, two parameters when using ChIPpeakAnno needed to be further considered for annotation of peaks as TEs: (1) the feature of the reference TE itself and (2) the distance between the TE to peaks (in other words, coverage of a called peak by TE). For the first parameter, certain very short simple repeat motifs existed in the reference TEs, such as (AT)<sub>n</sub>, with n = 3. And peaks with such short simple motifs may be incorrectly called as TEs. As this work is focused mainly on long TEs, we only retained those repetitive elements or TE features with a size of no less than 50 bp in the reference (for LINES, the size was no less than 15% of a full-length one, or 800 bp). For the second parameter, ChIPpeakAnno annotates peaks within 500 bp around TEs as TE peaks by default. In this study, we adopted a more stringent criterion by applying an overlap of no less than 50 bp for calling of TE peaks. In total, 32% of the TNRC18 peaks (3,090 of 9,656) were annotated as TEs by using the above stringent criterion (that is, the length of TE feature no less than 50 bp and peak TE overlap of no less than 50 bp). We conclude that a significant proportion of TNRC18 peaks are TEs, which can range from 32% (when using our custom-defined, stringent criteria) to 75% (by the default, relaxed parameters). To further annotate TE features, we chose to use the stringent criteria and focus on those TNRC18 peaks most confidently annotated as TEs by ChIPpeakAnno. Among them, around a half of them were further annotated as SINEs and 31% as ERVs (Extended Data Fig. 4b). The overall distribution of each TE subclass in the genome was considered and used for normalization in assessing its enrichment.

Motif analysis was performed using hypergeometric optimization of motif enrichment (HOMER) (v.4.8.2)<sup>83</sup>. Deeptools (v.3.2.0) was used to generate bigwig files, heatmaps and averaged plotting of CUT&RUN signals<sup>84</sup>. Genomic binding profiles were generated using the deepTools 'bamCoverage' and 'computeMatrix' functions, and blacklist regions were removed from coverage. The reads at repeated element regions, downloaded from the UCSC Genome Browser (repeatMasker Table), were fractionally assigned using featureCounts (v.2.0.0) with the following parameters: --ignoreDup -p -d 30 -D 1200 -M --fraction<sup>85</sup>. These were then normalized to the sequencing depth. The log<sub>2</sub> ratio of TNRC18 mutation relative to WT or specific IP relative to IgG was calculated to generate box plots using R.

Whole genome-wide bisulfite sequencing and data analysis Whole genome-wide bisulfate sequencing (WGBS) and data analyses were conducted as previously described<sup>86</sup>. In brief, total genomic DNA was purified and then used for the preparation of WGBS libraries and deep sequencing (performed by Admera Health). In brief, bisulfite treatment of genomic DNA was conducted using an EZ DNA Methylation-Gold kit (Zymo Research, D5005) per the vendor-provided instructions. Then, bisulfite-converted single-stranded DNA was recovered and used for library construction with an Accel-NGS@ Methyl-seq DNA Library kit (Swift BioSciences, 30024) based on the manufacturer's instructions. DNA enrichment was performed using PCR with primers compatible with Illumina sequencing. After quantity assessment and molecular size measurement, the WGBS libraries were pooled and sequenced using an Illumina NovaSeq S4 sequencer platform (150-bp read lengths in paired-end mode, with an output of approximately 750 million reads or 375 million in each direction per sample, and with 20% PhiX DNA added as spike-in controls). Read mapping and CpG calling were generated using the Bismark pipeline<sup>87</sup>. Then, the annotation of CpG sites were performed on the basis of UCSC Genome Browser (repeatMasker and CpG islands tracks). Data were filtered, normalized and calculated using the R packages 'methylKit' (v.1.26.0) (filterByCoverage(lo.count = 10, lo.perc = NULL, hi.count = NULL, hi.perc = 99.9), normalizeCoverage(method="median")). The violin plots of CpG methylation representing each site with the same annotation group were generated using the R package ggpubr.

### **CRISPR–Cas9-based Tnrc18 mutation KI in mice and phenotypic characterization**

Tnrc18 KI mice bearing the W2745A/Y2747A mutation were generated in the Animal Models Core (AMC) facility affiliated to UNC School of Medicine by pronuclear microinjection of CRISPR–Cas9 reagents in mouse embryos with a previously described protocol<sup>35</sup>. Cas9 gRNAs targeting the mouse Tnrc18W2745A/Y2747A-encoding codon were identified using Benchling software. Three gRNAs overlapping this region codon were selected for activity testing. Selected gRNAs were cloned into a T7 promoter vector followed by in vitro transcription and spin column purification. Functional testing was performed by transfecting a MEF cell line with gRNA and recombinant Cas9 protein (UNC Protein Expression and Purification Core Facility). Following transfection, the gRNA target site in Tnrc18 was PCR amplified from transfected cells and analysed using a T7endo1 assay to detect Cas9-mediated mutation. The guide RNA selected for genome editing in embryos was Tnrc18-sg85T (protospacer sequence 5'-GTGGTTCTACCATCCGG-3'). A donor oligonucleotide, Tnrc18-WY-AA-KI-B (sequence 5'-

TCACCTGGCCCTCATGGAAGTCTTGCCCGGGCTGGTCTCCTCCGGATGGGCGAA GGCCTTGACACGAACCACCATATTGTTGCCCCAGGACTCCCACATGCTCTGGA –3'), was designed for homologous recombination to insert the TGG to GCC and TAC to GCC codon mutations.

Microinjected embryos were implanted in recipient pseudopregnant females. Five resulting pups were screened by PCR and sequencing for the presence of the mutant allele. The correct mutant allele was detected in two females born from microinjection. One positive founder was from the 25 ng  $\mu\text{l}^{-1}$  guide RNA mix, and the other founder was from the 10 ng  $\mu\text{l}^{-1}$  mix. We routinely examined around 8–10 most highly predicted off-target sites to ensure that they were unaffected in the founders. Female founders were mated to WT C57BL/6J males, but only one transmitted the mutant allele through the germline to offspring. To establish the mutant colony for further studies, F1 animals heterozygous for the correct KI allele were back crossed with WT C57b/6 mice for more than six times to remove/segregate potential off-target mutation. Then, we used the back-crossed heterozygous mice for breeding with WT C57b/6 mice to maintain this line. Colony maintenance, daily care and monitoring of the lines were carried out in the collaboration with the Animal Studies Core, UNC.

### Generation of primary MEF cells

The mice were set up for breeding, and embryonic day 13.5 embryos collected. The isolation and establishment of primary MEF cells were conducted as previously described<sup>88</sup>. In brief, unwanted organs of embryos were removed, followed by tissue homogenization and enzymatic dissociation of cells. Genotypes of cells were determined by PCR and Sanger sequencing. Cells were then cultured in DMEM base medium supplemented with 15% FBS for 25 passages to establish stable MEF cell lines.

### Luciferase reporter assay

Cells seeded in 24-well plates were co-transfected with plasmids of GAL4-DBD-TNRC18, a luciferase reporter pGL2-5 $\times$ GAL4-SV40 (Pro- mega) and an internal control (pRL-CMV-Renilla). Forty-eight hours after transfection, luciferase activity was measured using a Dual Luciferase Reporter Assay system (Promega, E1910). Data were normalized to Renilla luciferase signals.

### Ethics oversight

All animal experiments were approved by and performed in accordance with the guidelines of Institutional Animal Care and Use Committee of the institutions. Experienced staff at the core facilities provide support of daily care and monitoring. Mice were housed in a specific pathogenfree facility with 12-h dark–12-h light cycle. Cages were well ventilated, softly lit and subjected to a light–dark cycle. The relative humidity was kept at 45–65%. Mouse rooms and cages were kept at a temperature range of 20–24 °C.

### Statistics and reproducibility

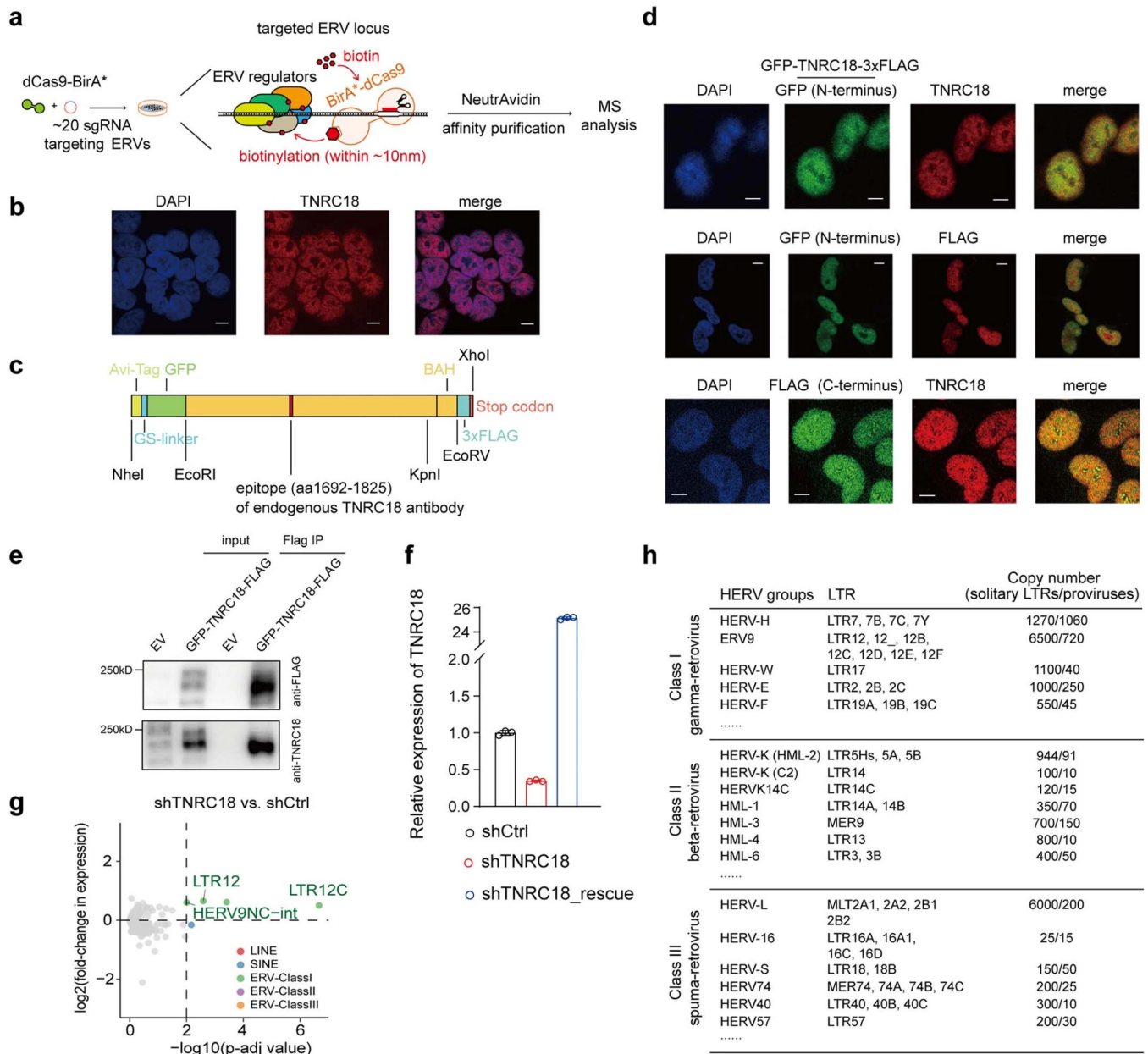
Experimental data are presented as the mean  $\pm$  s.d. of three independent experiments unless otherwise noted. The results for IF (Fig. 1b and Extended Data Fig. 1b,d), Coomassie

brilliant blue staining (Extended Data Fig. 5b), ethidium bromide staining (Extended Data Fig. 6g) and immunoblotting (Extended Data Figs. 1e, 2b, 3c, 7b, 8f, 9d and 10b,d) are representative of at least three biologically independent experiments.  $P < 0.05$  was considered significant. Significance levels are denoted as follows: \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ , \*\*\*\* $P < 0.0001$ . All exact  $P$  values in the main figures and Extended Data Figures are shown in Supplementary Table 8.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Extended Data



**Extended Data Fig. 1 | Unbiased CasID-based proteomic approach identified a previously less-studied nuclear protein TNRC18 as a putative binding protein at ERV regions.**

a. Workflow of CasID. A set of sgRNAs targeting representative ERV regions (sgRNA sequence information provided in Supplementary Table 7) were transduced into cells with stable expression of dCas9-BirA\*. The sgRNA-based targeting of dCas9-BirA\* to ERVs results in the BirA\*-mediated biotinylation of proteins in proximity (within approximately 10 nm) in the presence of biotin. Then, biotinylated proteins were enriched by NeutrAvidin affinity purification, followed by mass spectrometry-based protein identification.



b. Representative confocal immunofluorescence (IF) microscopy images showing the nuclear localization pattern of TNRC18 in HEK293 cells, probed with either DAPI (left) or anti-TNRC18 antibodies (middle). Scale bars, 10  $\mu\text{m}$ . Data shown represent 3 independent experiments.

c. The cDNA of TNRC18 used for mammalian cell expression, which contains an Avi-tag and a GFP tag at the N-terminus and 3 $\times$ FLAG tag at the C-terminus.

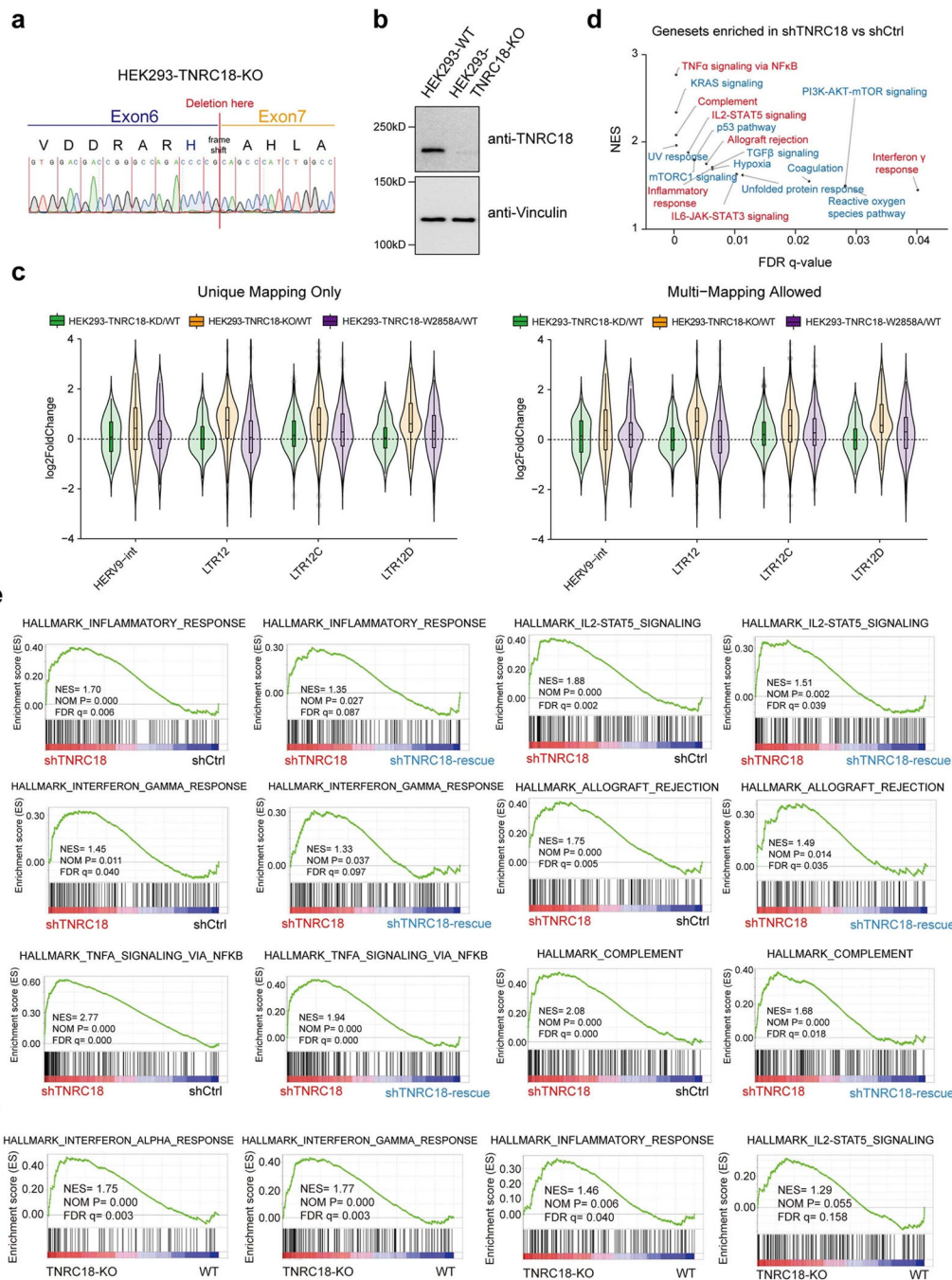
d. Representative confocal IF microscopy images for the above tagged TNRC18 in the HEK293 stable expression cells, co-stained with the mouse anti-GFP antibody and the rabbit antibody against endogenous TNRC18 (top panel), or with the rabbit anti-GFP antibody and the mouse anti-FLAG antibody (middle panel), or with the mouse anti-FLAG antibody and the rabbit antibody against endogenous TNRC18 (bottom panel), all of which exhibited a co-localization pattern in the nucleus. Data shown represent 3 independent experiments.

e. Immunoprecipitation (IP) of HEK293 cells expressing GFP-3 $\times$ FLAG-tagged TNRC18, compared to empty vector (EV)-transduced control cells, by using anti-FLAG beads. The IP sample was immunoblotted with anti-FLAG antibody (top panel) or that of endogenous TNRC18 (bottom panel; for TNRC18 protein size, refer to Methods of IP as well).

f. RT-qPCR of TNRC18 in the HEK293 cells with knockdown (KD) of endogenous TNRC18 (shTNRC18), compared to scramble controls (shCtrl), and in the cells with endogenous TNRC18 KD followed by the rescued re-expression of shTNRC18-resistant TNRC18 (i.e., shTNRC18\_rescue; n = 3 biologically independent experiments). Data were plotted as the mean  $\pm$  s.d. after normalization to the signals of an internal control (GAPDH) and to those of the control samples.

g. Scatter plot showing the indicated transposable element families exhibiting significant expression change, based on RNA-seq profiles of HEK293 cells with shTNRC18, compared to scramble controls (shCtrl) (n = 2 biologically independent experiments). The cut-off of statistical significance is log<sub>2</sub> value of fold-change in expression (y-axis) over 0.58 and adjusted P value (x-axis) less than 0.01 for transcripts with basemean read counts over 10. Adjusted P value is calculated by negative binomial model-based methods (DESeq2).

h. Classification of endogenous retroviruses in the human cells. Figure adapted from ref. 89, with permission from Elsevier and under a Creative Commons licence CC-BY 4.0, and ref. 14, BioMed Central.



**Extended Data Fig. 2 | TNRC18 knock-down (KD) or knockout (KO) results in activation of immunity-related genes in HEK293 cells.**

- Sanger sequencing to show frame-shifting mutation and knock-out of TNRC18 in HEK293 cells.
- Western blot to show the knock-out of TNRC18 in HEK293 cells. Vinculin is the sample processing control.
- RNA-seq analysis using unique mapping reads (left) and multi-mapping reads (right) of the indicated HEK293 cells. The boundaries of box plots indicate the 25th and 75th percentiles, the center line indicates the median, and the whiskers indicate 1.5× the

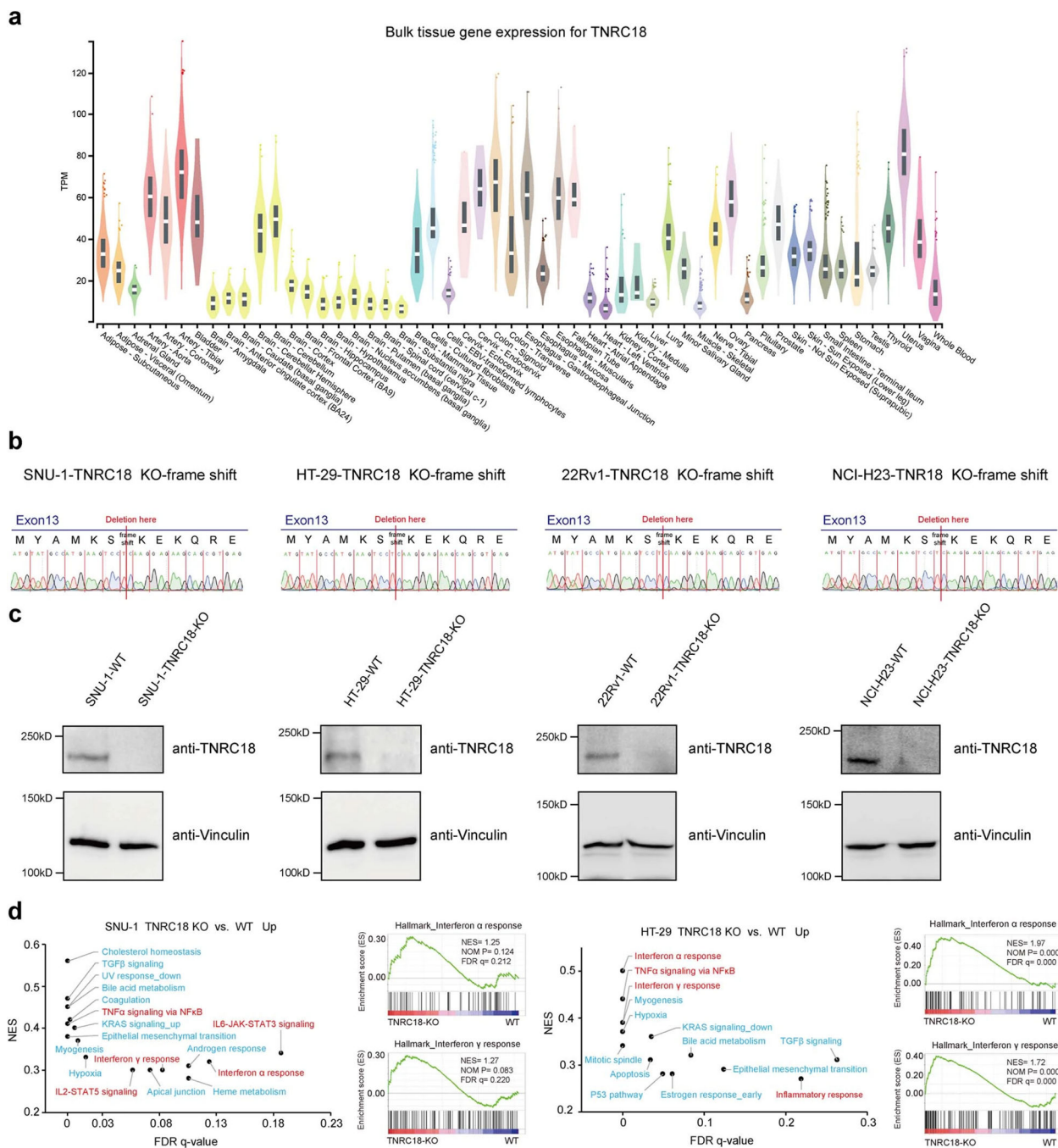
interquartile range. The number of analyzed copies per transposon type: group ‘unique mapping only’ (LTR12 (n = 122 copies), LTR12C (n = 612 copies), LTR12D (n = 54 copies), HERV9-int (n = 58 copies)); group ‘multi-mapping allowed’ (LTR12 (n = 124 copies), LTR12C (n = 678 copies), LTR12D (n = 55 copies), HERV9-int (n = 60 copies)). Sample size of each box plot is also listed in Supplementary Table 8.

d. Gene set enrichment analysis (GSEA) revealed enrichment for the indicated pathways in cells with TNRC18 KD (shTNRC18), compared to mock controls (shCtrl) (n = 2 biologically independent experiments). Immunity-related gene sets are labelled in red. The y-axis and x-axis showed the normalized enrichment score (NES) and false discovery rate (FDR) q-values of GSEA, respectively.

e. GSEA revealed the positive correlation between activation of genes related to immunity and TNRC18 KD (shTNRC18) in HEK293 cells, compared to those with the scramble controls (shCtrl) or with the rescued re-expression of TNRC18 (shTNRC18\_rescue).

NES, normalized enrichment score. The P value was calculated by a two-sided empirical phenotype-based permutation test; the false discovery rate q-value is adjusted for gene set size and several hypotheses testing whereas the P value is not.

f. GSEA revealing the positive correlation between activation of genes related to immunity and TNRC18 KO in HEK293 cells, compared to WT controls. The P value was calculated by a two-sided empirical phenotype-based permutation test; the false discovery rate q-value is adjusted for gene set size and several hypotheses testing whereas the P value is not..

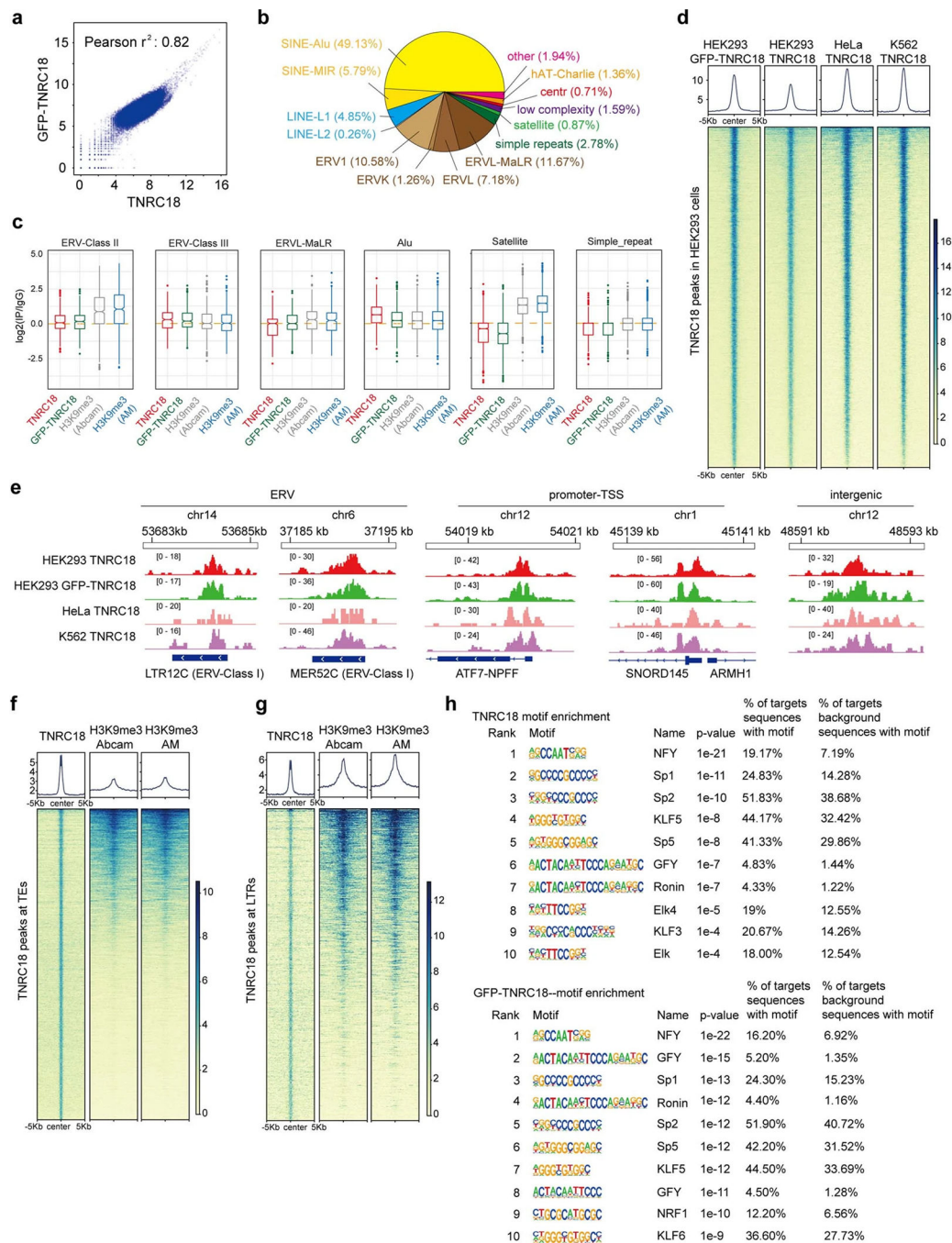


**Extended Data Fig. 3 |. TNRC18 KO using four different human cancer cell lines of epithelial origins.**

a. Bulk tissue gene expression for TNRC18 in the GTEx Analysis Release V8<sup>90</sup>. Expression values are shown in TPM (Transcripts Per Million), calculated from a gene model with isoforms collapsed to a single gene. The boundaries of box plots indicate the 25th and 75th percentiles, the center line indicates the median, and the whiskers indicate 1.5× the interquartile range. Sample size of each box plot is listed in Supplementary Table 8.

b-c. Sanger sequencing (b) and Western blot (c) to show frame-shifting mutation and KO of TNRC18 in the indicated edited cells, in comparison to WT. Vinculin is the sample processing control. For TNRC18 protein size, see Methods as well.

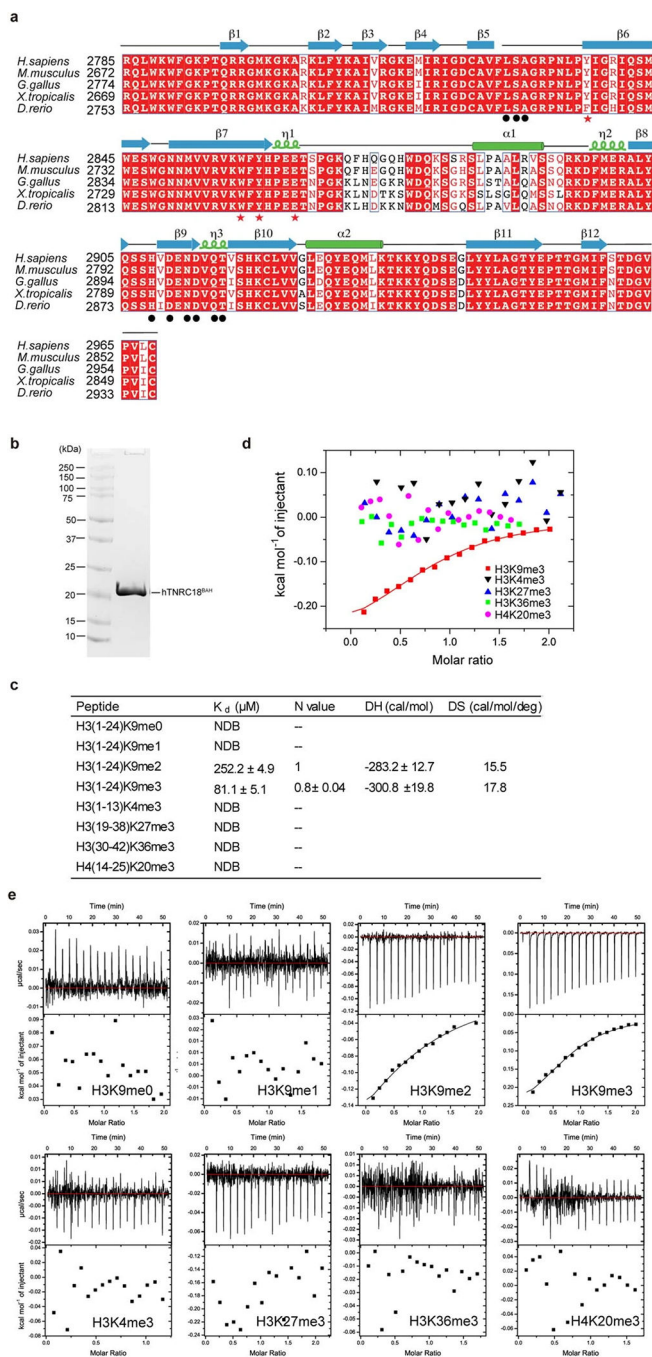
d. GSEA revealed enrichment for the indicated pathways in SNU-1 cells (left) and NCI-H23 cells (right) with TNRC18 KO, compared to WT controls. Immunity- related gene sets are labelled in red. NES, normalized enrichment score. The P value was calculated by a two-sided empirical phenotype-based permutation test; the false discovery rate q-value is adjusted for gene set size and several hypotheses testing whereas the P value is not.



**Extended Data Fig. 4 | TNRC18 and H3K9me3 co-localize at ERV regions.**

- Scatter plot showing correlation between signals of CUT&RUN for endogenous TNRC18 in parental HEK293 cells (x-axis; using anti-TNRC18 antibody) and those for GFP-TNRC18 following its stable expression into the HEK293 cells (y-axis; using anti-GFP antibody). Pearson correlation coefficient is shown.
- Pie chart showing distribution of the indicated TE annotation using the TNRC18 CUT&RUN peaks annotated as TE by ChIPpeakAnno.

- c. Box plots showing the log<sub>2</sub> values of fold-changes in signals of CUT&RUN for the indicated protein, relative to IgG controls, at different TE classes in HEK293 cells. CUT&RUN for TNRC18 was conducted with antibody of endogenous TNRC18 in parental HEK293 cells, or antibody of GFP in cells stably expressed with GFP-TNRC18. CUT&RUN for H3K9me3 was performed by using two independent antibodies, from either Abcam or Active Motif Inc. (AM). The boundaries of box plots indicate the 25th and 75th percentiles, the center line indicates the median, and the whiskers indicate 1.5× the interquartile range. Sample size of each box plot is listed in Supplementary Table 8.
- d. Heatmap of anti-GFP CUT&RUN signals in HEK293 cells stably expressed with GFP-TNRC18 (1st column), or those for TNRC18, probed by antibody of endogenous TNRC18, in either parental HEK293 (2nd column), HeLa (3rd column) or K562 cells (4th column), across ±5 kb from the most confident TNRC18 peaks defined in HEK293 cells (n = 7545; peaks common to GFP-TNRC18 [1st column) and TNRC18 [2nd column] in HEK293 cells were used).
- e. IGV tracks showing the CUT&RUN signal for TNRC18 in the indicated cells at the representative ERV (left), promoter-TSS (middle) or intergenic target site (right). TSS, transcription start site.
- f,g. Heatmap for the CUT&RUN signals of TNRC18 and H3K9me3 (probed with two independent antibodies), across ±5 kb from those TNRC18 peaks that were annotated as TEs (f) or LTRs (g).
- h. Motif search analysis revealing the most enriched motifs at the TNRC18 (top panel) and GFP-TNRC18 peaks (bottom panel) in HEK293 cells..



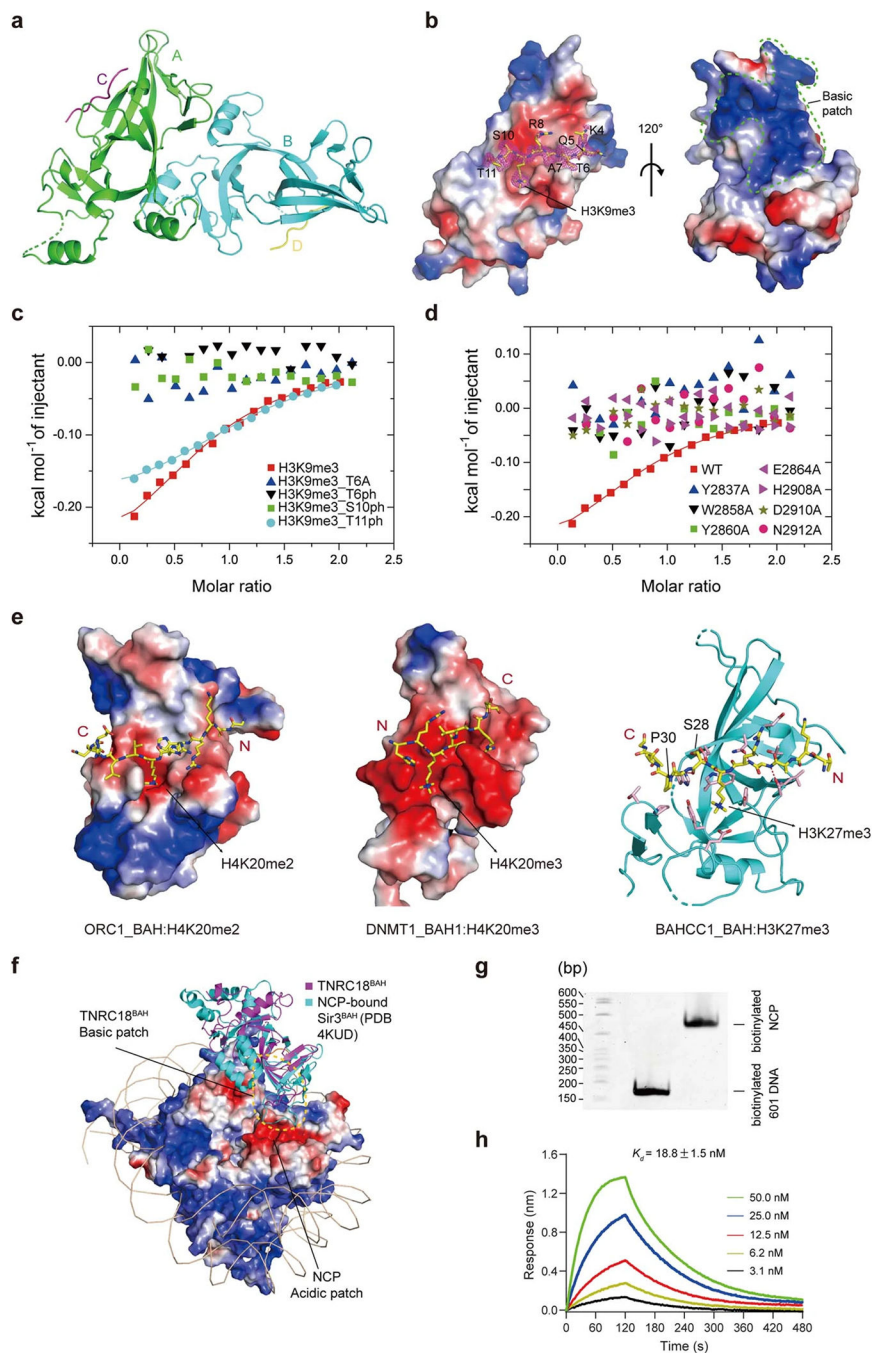
**Extended Data Fig. 5 | TNRC18BAH is a conserved domain binding specifically to H3K9me3.**

a. Sequence alignment of TNRC18BAH among different species. The secondary structures of human TNRC18BAH are indicated on top. The H3K9me3-binding pocket residues are indicated by red asterisks, and the rest of H3-binding sites are indicated by filled black circles at the bottom.

b. SDS-PAGE image of the purified recombinant protein of TNRC18BAH used for biochemical and structural studies.

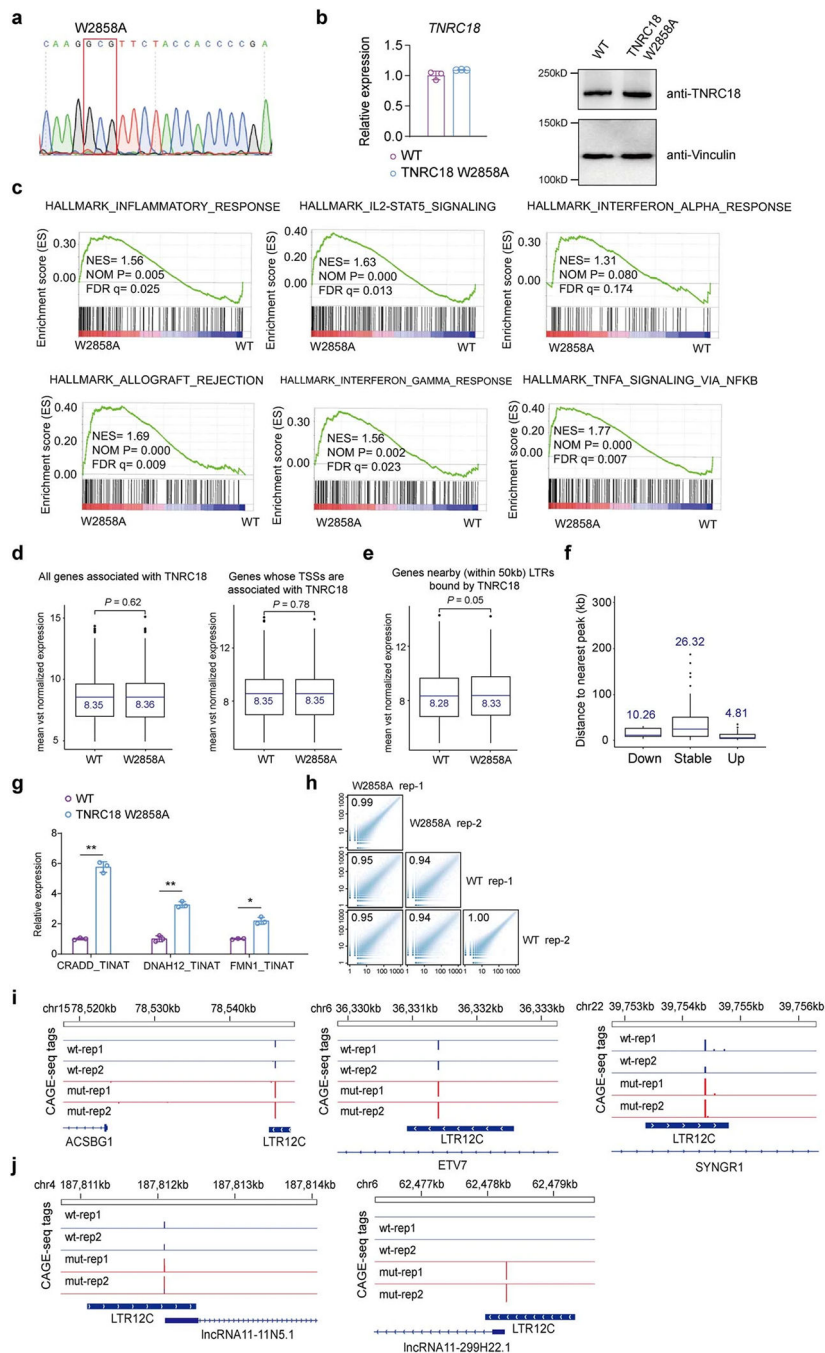


- c. ITC fitting parameters of TNRC18BAH binding to various histone peptides. NDB, no detectable binding.
- d. ITC fitting curves of TNRC18BAH against the indicated histone peptides trimethylated at different lysine sites.
- e. Original ITC binding curves of the recombinant TNRC18BAH protein against histone peptides with the indicated modification.



**Extended Data Fig. 6 | Structural basis of TNRC18BAH binding to H3K9me3 peptide and biochemical analysis of TNRC18BAH binding to H3K9me3- modified nucleosome.**

- a. Crystal structures of the two color-coded human TNRC18BAH molecules in one asymmetric unit, with the chain identifiers labeled (chain A and B). Each TNRC18BAH molecule is complexed with one H3K9me3 peptide (chain C or D). The TNRC18BAH-H3K9me3 complex with the best model- to-map fit (chain B and D) was selected for structural analysis.
- b. Electrostatic surface view of human TNRC18BAH bound to the H3K9me3 (yellow sticks). (left) The  $F_o - F_c$  omit map of the H3K9me3 peptide, contoured at  $1.5\sigma$  level, is shown as magenta mesh. (right) The surface patch enriched with basic residues is indicated as the potential binding site to nucleosome/DNA.
- c. ITC binding curves of TNRC18BAH against histone peptides with the indicated modification.
- d. ITC binding curves of the indicated TNRC18BAH mutants against the H3K9me3 peptide.
- e. Structural comparison of ORC1BAH-H4K20me2 (electrostatic surface and stick representation; PDB 4DOW), DNMT1BAH1-H4K20me3 (electrostatic surface and stick representation; PDB 7LMK) and BAHCC1BAH-H3K27me3 (ribbon and stick representation; PDB 6VIL).
- f. Structural superposition of TNRC18BAH with the nucleosome core particle (NCP)-bound yeast Sir3 BAH domain (Sir3BAH; PDB 4KUD) reveals that the basic patch of TNRC18BAH (Extended Data Fig. 6b, right panel) corresponds to a similar region of Sir3BAH involved in binding to the acidic patch of NCP.
- g. Assessment of purified H3K9me3-modified NCP on a native 5% TBE gel. The positions of reconstituted NCP and biotinylated 601 DNA are labeled on the right.
- h. The BLI kinetic curves for the TNRC18BAH-NCP binding. The concentrations of TNRC18BAH used for each kinetic measurement are indicated. The  $K_d$  value and s.d. were derived from two independent measurements..



**Extended Data Fig. 7 | A H3K9me3-binding-defective mutation of TNRC18BAH leads to the activated expression of the LTR12 family TEs, TINATs and interferon-stimulated genes.**

- Sanger sequencing verified homozygous knock-in (KI) mutation of TNRC18W2858A, introduced by the CRISPR-Cas9-based gene editing, in HEK293 cells.
- RT-qPCR and Western blot of TNRC18 in HEK293 cells, either wild type (WT) or with homozygous mutation of TNRC18W2858A. Data were plotted as the mean  $\pm$  s.d. after normalization to the signals of an internal control, GAPDH, and to those of WT cells (n = 3 independent experiments). Vinculin is the sample processing control.

c. GSEA revealed the positive correlation between activation of the indicated immunity-related gene sets and the H3K9me3-binding-defective mutation of TNRC18 (W2858A), relative to WT, in HEK293 cells (n = 2 independent experiments). The P value was calculated by a two sided empirical phenotype-based permutation test; the false discovery rate q-value is adjusted for gene set size and several hypotheses testing whereas the P value is not.

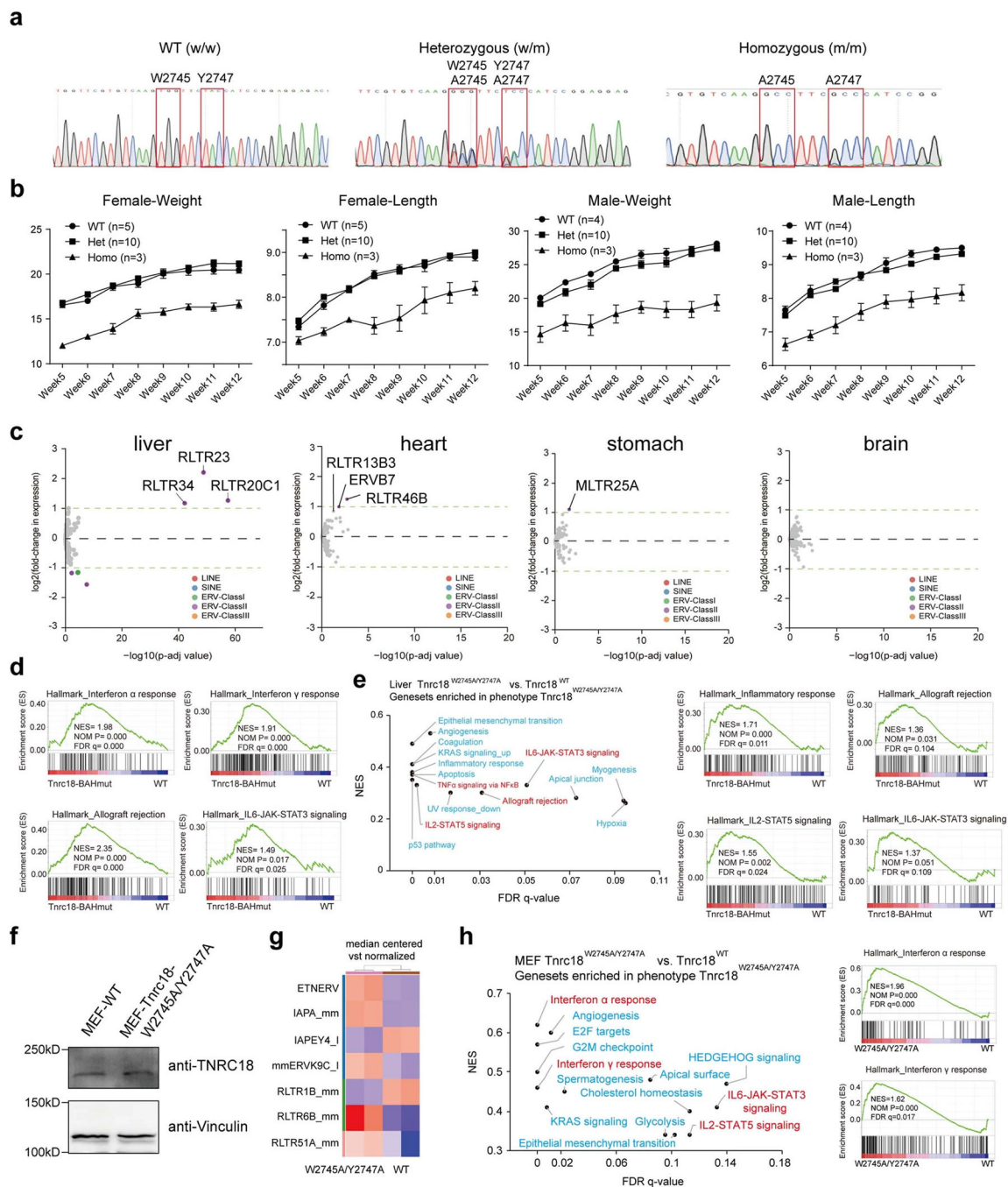
d-e. Overall expression levels of genes associated with all TNRC18 peaks (d, left), those with TNRC18-bound promoter/TSS regions (d, right), or genes close to (within 50 kb) the TNRC18-bound LTRs (e), based on RNA-seq profiling of HEK293 cells, either WT (left) or carrying the TNRC18W2858A homozygous mutation (right). VST, variance-stabilizing transformation. The boundaries of box plots indicate the 25th and 75th percentiles, the center line indicates the median, and the whiskers indicate 1.5× the interquartile range. Sample size of each box plot is listed in Supplementary Table 8. The P value was calculated by two-sided Wilcoxon test.

f. Averaged distance to the nearest TNRC18/H3K9me3-bound LTRs from genes exhibiting either down-regulation (left), no expression change (stable; middle) or up-regulation (right) in HEK293 cells carrying the TNRC18W2858A homozygous mutation, relative to WT, based on RNA-seq. The boundaries of box plots indicate the 25th and 75th percentiles, the center line indicates the median, and the whiskers indicate 1.5× the interquartile range. P value was calculated by two-sided Wilcoxon test. Sample size of each box plot is listed in Supplementary Table 8.

g. RT-qPCR of representative treatment-induced non-annotated TSSs (TINATs40) in HEK293 cells carrying the TNRC18W2858A homozygous mutation, relative to WT (n = 3 independent experiments; \*P < 0.05; \*\*P < 0.01, two-sided t-test). Data were plotted as the mean ±s.d. after normalization to signals of GAPDH and then to those of WT. The exact P value is shown in Supplementary Table 8.

h. Scatter plot showing Pearson correlation of the CAGE-seq profiles of HEK293 cells, WT or with the TNRC18W2858A homozygous mutation (n = 2 independent experiments). Rep-1/2, biological replicate 1 or 2.

i-j. IGV view of CAGE-seq profiles at the indicated gene (i) or long non-coding RNA (lncRNA; j) containing a nearby LTR12 in HEK293 cells, WT or with the TNRC18W2858A homozygous mutation.



**Extended Data Fig. 8 |** Compared to WT mice, those harboring the homozygous H3K9me3-binding-defective mutation of TNRC18BAH exhibited severe phenotypes of neonatal lethality and dwarfism in adult surviving animals, as well as ERV de-repression in their organs and MEF cells.

- Sanger sequencing results verified the genotype of mice, either WT (w/w) or with the Tncr18W2745A/Y2747A heterozygous (w/m) or homozygous (m/m) mutation.
- Measurement of body size and weight at the indicated time point post-birth showing that the Tncr18W2745A/Y2747A homozygous mice (female, n = 3; male, n = 3) were smaller, compared to their WT (female, n = 5; male, n = 4) and heterozygous (female, n = 10; male,

n = 10) littermates, both in the cohort of females (left) and males (right). Data are presented as mean value  $\pm$  SEM.

c. Scatter plot showing the indicated TE families that exhibit significant expression change, based on RNA-seq profiles of different primary tissue samples isolated from mice with the *Tnrc18W2745A/Y2747A* homozygous mutation (n = 6, two replicated experiments for three mice), compared to those of WT littermates (n = 4, two replicated experiments for two mice). Adjusted P value is calculated by negative binomial model-based methods (DESeq2).

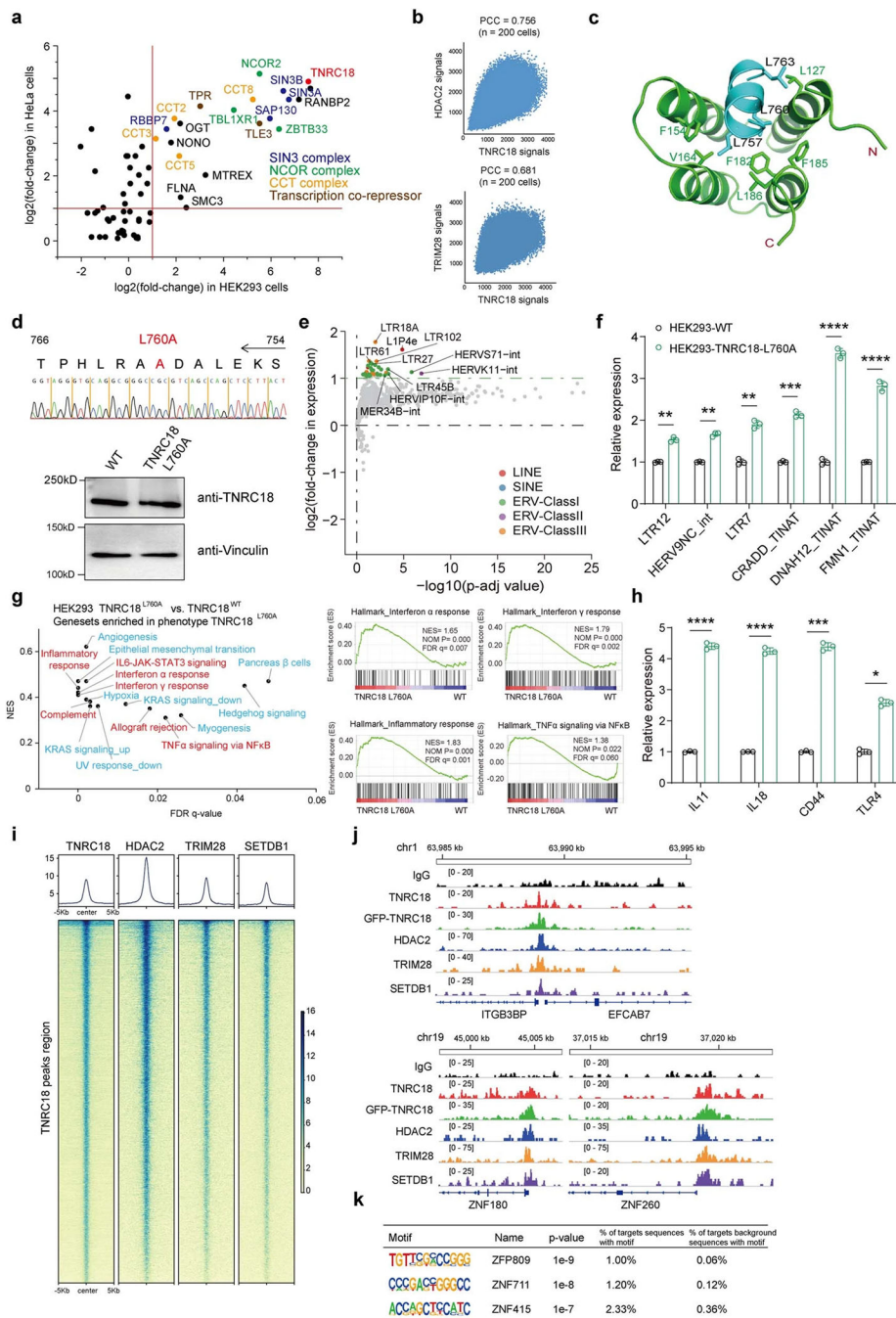
d. GSEA revealed the positive correlation between activation of the indicated immunity-related gene sets in the lung and the *Tnrc18W2745A/Y2747A* homozygous mutation, relative to WT controls. The P value was calculated by a two-sided empirical phenotype-based permutation test; the false discovery rate q-value is adjusted for gene set size and several hypotheses testing whereas the P value is not.

e. GSEA revealed enrichment for the indicated pathways in the liver of mice with the *Tnrc18W2745A/Y2747A* homozygous mutation, compared to WT controls. Immunity-related gene sets are labelled in red. The P value was calculated by a two-sided empirical phenotype-based permutation test; the false discovery rate q-value is adjusted for gene set size and several hypotheses testing whereas the P value is not.

f. Western blot of TNRC18 in WT MEF cells and those with the *Tnrc18W2745A/Y2747A* homozygous mutation. Vinculin is the sample processing control.

g. Heatmap showing TE families with significant expression change, based on RNA-seq profiles of MEFs harboring the *Tnrc18W2745A/Y2747A* homozygous mutation versus WT (n = 2 independent experiments). The cut-off of statistical significance is log<sub>2</sub> value of fold-change in expression over 0.58 and adjusted P value less than 0.01 for transcripts with basemean counts over 10.

h. GSEA revealed the positive correlation between activation of the indicated interferon response-related gene sets and homozygous *Tnrc18W2745A/Y2747A* mutation in MEF cells, relative to WT (n = 2 independent experiments). Immunity-related gene sets are labelled in red. The P value was calculated by a two sided empirical phenotype-based permutation test; the false discovery rate q-value is adjusted for gene set size and several hypotheses testing whereas the P value is not.



**Extended Data Fig. 9 | TNRC18 binds corepressors, mediating transcriptional repression.**

a. Scatter plots of the indicated TNRC18-interacting proteins identified by BioID in HEK293 (x-axis) and HeLa cells (y-axis). The mass spectrometry data following BioID in HeLa cells is shown in Supplementary Table 5.

b. Pearson correlation plot using the IF signals of the indicated protein in HEK293 cells.

c. Structural model of the Sin3A’s PAH domain (green) in complex with the TNRC18 peptide (cyan). The structure was predicted via PHYRE2 (<http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index>). The model of the complex was generated by the Coot

program, using the structure of mouse Sin3A PAH1-SAP25 SID complex (PDB 2RMS) as template.

d. Sanger sequencing (top) and Western blot (bottom) for the TNRC18 L760A KI mutation introduced into HEK293 cells. Vinculin is the sample processing control.

e. Scatter plot showing the indicated TE families that exhibit significant expression change, based on RNA-seq profiles of HEK293 cells with the TNRC18 L760A mutation versus WT controls. Adjusted P value is calculated by negative binomial model-based methods (DESeq2).

f. RT-qPCR of the indicated TEs in HEK293 cells with TNRC18 L760A mutation versus WT controls (n = 3 independent experiments; \*\*P < 0.01; \*\*\*P < 0.001; \*\*\*\*P < 0.0001, two-sided t-test). Data were plotted as the mean  $\pm$ s.d. after normalization to signals of GAPDH and then to those of WT. The exact P value is shown in Supplementary Table 8.

g. GSEA revealed enrichment for the indicated pathways in HEK293 cells with the TNRC18 L760A mutation versus WT controls. Immunity-related gene sets are labelled in red. The P value was calculated by a two sided empirical phenotype-based permutation test; the false discovery rate q-value is adjusted for geneset size and several hypotheses testing whereas the P value is not.

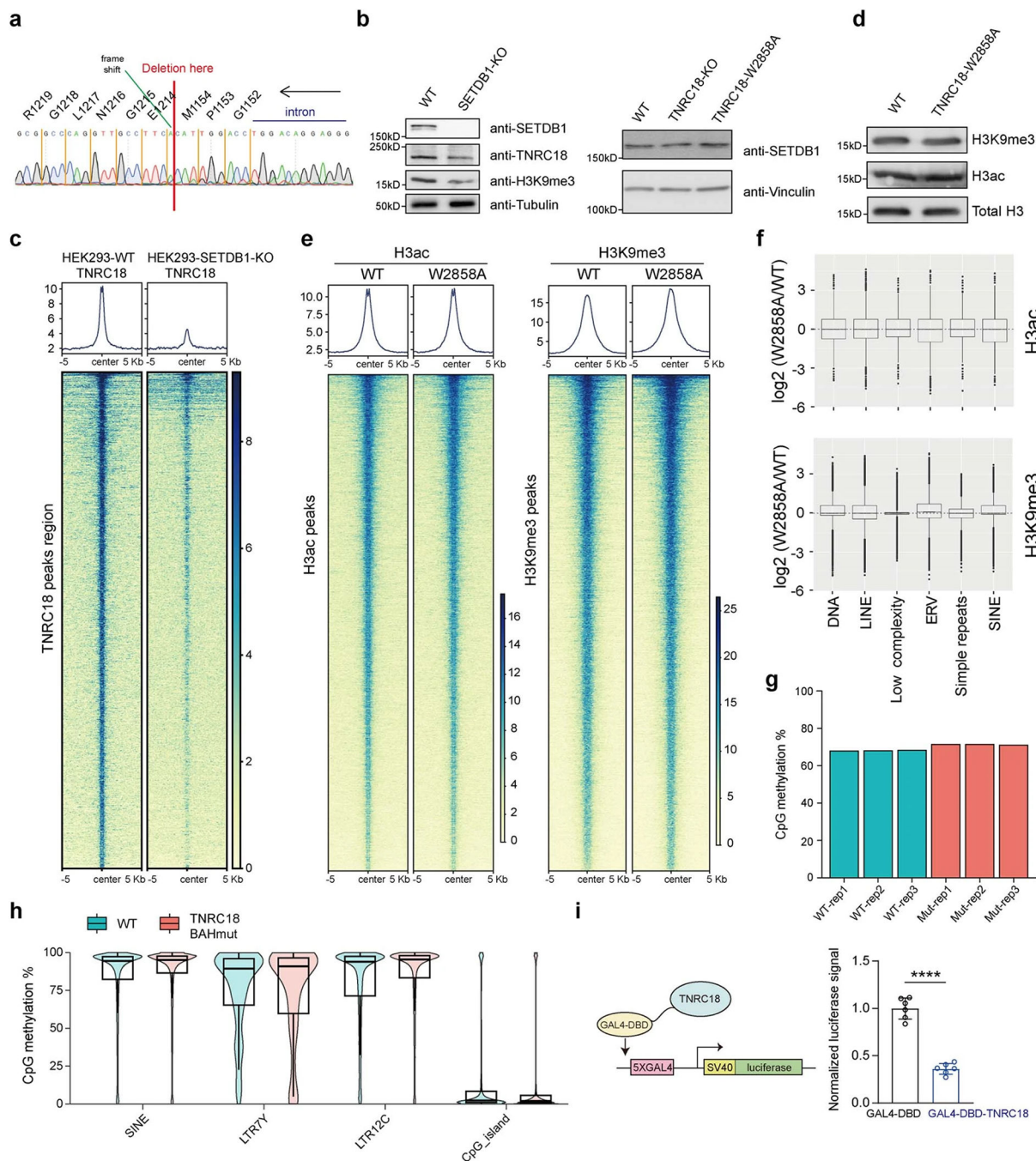
h. RT-qPCR of the indicated immunity-related genes in WT and TNRC18-L760A- mutated HEK293 cells (n = 3 independent experiments; \*P < 0.05; \*\*\*P < 0.001; \*\*\*\*P < 0.0001, two-sided t-test). Data were plotted as the mean  $\pm$ s.d. after normalization to signals of GAPDH and then to those of WT. The exact P value is shown in Supplementary Table 8.

i. Heatmap of CUT&RUN signals for TNRC18, HDAC2, TRIM28 and SETDB1 in HEK293 cells, across  $\pm$ 5 kb from the TNRC18 peaks (n = 7545; defined as the common peaks of TNRC18 and GFP-TNRC18, based on CUT&RUN in HEK293 cells).

j. IGV tracks showing the CUT&RUN signals of the indicated protein at the reported TRIM28 target gene in HEK293 cells.

k. The motif search analysis revealed the binding motifs of KRAB-ZnF proteins to be enriched at the TNRC18 peaks.





**Extended Data Fig. 10 | The TNRC18BAH-H3K9me3 engagement mediates optimal recruitment of TNRC18 and associated co-repressors, maintaining a silenced chromatin state at targets.**

a. Sanger sequencing to verify the CRISPR-Cas9-induced frameshift and KO of SETDB1 in HEK293 cells.

b. Left: Western blot of SETDB1, TNRC18 and H3K9me3 in WT and SETDB1-KO HEK293 cells. Right: Western blot of SETDB1 in HEK293 cells with WT TNRC18, TNRC18 KO and TNRC18 W2858A mutation. Vinculin and Tubulin are the sample processing control. The representative results from 3 independent experiments are shown.

- c. Heatmap of TNRC18 CUT&RUN signal across  $\pm 5$  kb at the TNRC18 peaks region in WT and SETDB1-KO HEK293 cells.
- d. Immunoblotting for H3K9me3, histone H3 acetylation (H3ac) and total H3 in HEK293 cells, either WT or with the homozygous mutation of TNRC18W2858A. Data shown are the representative results from 3 independent experiments.
- e. The heatmap of CUT&RUN signals of H3ac (left) and H3K9me3 (right) in HEK293 cells, either WT or with the homozygous mutation of TNRC18W2858A, across  $\pm 5$  kb from the called peaks.
- f. Box plot showing the log<sub>2</sub> values of normalized CUT&RUN sequencing counts in HEK293 cells with homozygous TNRC18W2858A mutation, compared to WT, at the indicated different TE classes. The boundaries of box plots indicate the 25th and 75th percentiles, the center line indicates the median, and the whiskers indicate 1.5 $\times$  the interquartile range. Sample size of each box plot is listed in Supplementary Table 8.
- g. Global methylation levels of WT and TNRC18W2858A-mutated HEK293 cells.
- h. DNA methylation levels of WT and TNRC18W2858A-mutated HEK293 cells at SINE, LTR7Y, LTR12C and CpG<sub>+</sub> island. The boundaries of box plots indicate the 25th and 75th percentiles, the center line indicates the median, and the whiskers indicate 1.5 $\times$  the interquartile range. Sample size of each box plot is listed in Supplementary Table 8.
- i. The dual-luciferase reporter-based system for assaying the activity of full-length TNRC18 fused to the GAL4's DNA binding domain (GAL4-DBD), relative to GAL4-DBD alone (n = 6 biologically independent samples; data presented as the mean  $\pm$  s.d.; \*\*\*\*P < 0.0001, two-sided t-test). The luciferase signals of cells with GAL4-DBD-TNRC18 fusion were first normalized to those of internal control (Renilla luciferase) and then normalized to those of cells with GAL4-DBD alone. The exact P value is shown in Supplementary Table 8..

#### Extended Data Table 1

Crystallographic data collection and refinement statistics of the TNRC18\_BAH:H3(1–22)K9me3 complex

hTNRC18 <sup>BAH</sup> -H3 <sub>1-22</sub> K9me3	
<b>Data collection</b>	
Space group	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	73.8, 120.8, 46.6
$\alpha$ , $\beta$ , $\gamma$ (°)	90, 90, 90
Resolution (Å)	37.48–1.84 (1.91–1.84) <sup>a</sup>
<i>R</i> <sub>merge</sub>	0.124 (1.2)
<i>I</i> / $\sigma$ <i>I</i>	8.5 (1.2)
CC <sub>(1/2)</sub>	0.996 (0.602)
Completeness (%)	98.8 (94.7)
Redundancy	5.2 (4.5)
<b>Refinement</b>	
No. reflections	36526
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub> (%)	20.7/23.5

hTNRC18 <sup>BAH</sup> -H3 <sub>1,22</sub> K9me3	
No. atoms	
Protein	2784
Water	204
<i>B</i> factors (Å <sup>2</sup> )	
Protein	38.8
Water	42.6
R.m.s. deviations	
Bond lengths (Å)	0.013
Bond angles (°)	1.3

<sup>a</sup>Values in parentheses are for highest-resolution shell. The dataset was collected from a single crystal.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

We thank the members of the Wang, Song and Cai laboratories for discussion and technical support; Z. Li and C. Pecot for sharing the reagents used in the work; staff members at the NE-CAT beamlines (GM124165), Advanced Photo Source (DE-AC02-06CH11357) and Argonne National Laboratory for access to the X-ray beamline; staff at institution-affiliated core facilities, including those for imaging, high-throughput deep sequencing, bioinformatics, flow cytometry and sorting, tissue culture and animal-related studies, for their professional assistance of this work. The cores affiliated to UNC Cancer Center are supported in part by the UNC Lineberger Comprehensive Cancer Center Core Support Grant P30-CA016086. This work was supported in part by NIH grants R01CA271603, R01CA268519 and R01CA268384 to G.G.W.; R35GM119721 to J.S.; R24GM137786, P20GM121293 and R01CA236209 to A.J.T., R35GM126900 to B.D.S.; R35GM147286 to J.R.R.; R01CA262903 to L.C.; and U01HL156064 and R35HG011328 to Y.D. J.R.R. is also supported in part by DoD Award W81XWH-19-1-0423. G.G.W. is an American Cancer Society Research Scholar and a Leukemia and Lymphoma Society Scholar.

## Data availability

Next-generation sequencing data have been deposited into the NCBI's Gene Expression Omnibus database under accession number GSE200839. Coordinates and structural factors for the TNRC18(BAH)-H3K9me3 complex have been deposited into the PDB (code 8DS8). PDB 4DOV was used as a search model for structure determination of the TNRC18(BAH)-H3K9me3 complex. Source data are provided with this paper.

## References

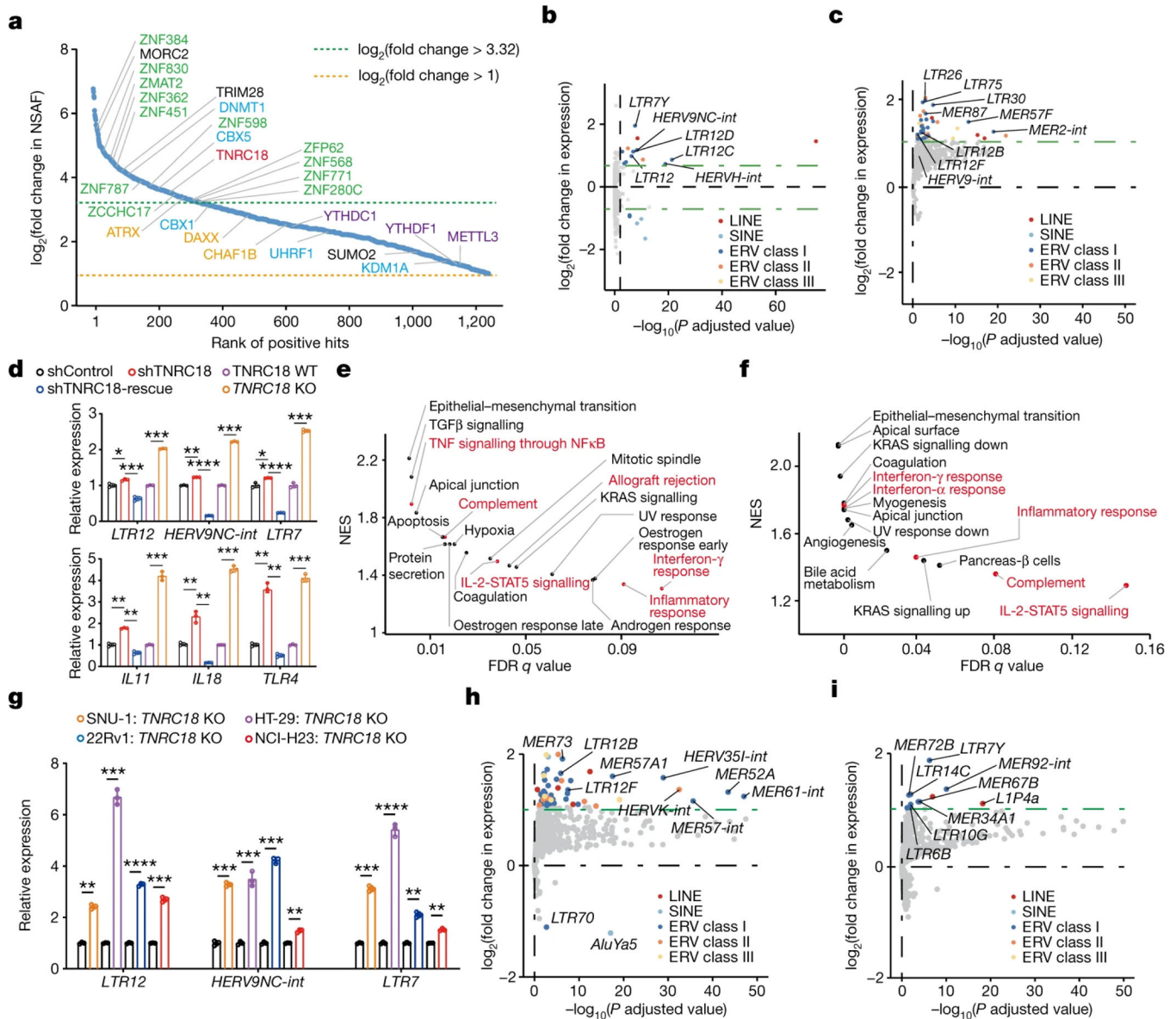
1. Padeken J, Methot SP & Gasser SM Establishment of H3K9-methylated heterochromatin and its functions in tissue differentiation and maintenance. *Nat. Rev. Mol. Cell Biol.* 23, 623–640 (2022). [PubMed: 35562425]
2. Tchasovnikarova IA et al. Epigenetic silencing by the HUSH complex mediates position-effect variegation in human cells. *Science* 348, 1481–1485 (2015). [PubMed: 26022416]
3. Liu N et al. Selective silencing of euchromatic L1s revealed by genome-wide screens for L1 regulators. *Nature* 553, 228–232 (2018). [PubMed: 29211708]
4. Tunbak H et al. The HUSH complex is a gatekeeper of type I interferon through epigenetic regulation of LINE-1s. *Nat. Commun.* 11, 5387 (2020). [PubMed: 33144593]

5. Lachner M, O'Carroll D, Rea S, Mechtler K & Jenuwein T Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* 410, 116–120 (2001). [PubMed: 11242053]
6. Jacobs SA et al. Specificity of the HP1 chromo domain for the methylated N-terminus of histone H3. *EMBO J.* 20, 5232–5241 (2001). [PubMed: 11566886]
7. Bannister AJ et al. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* 410, 120–124 (2001). [PubMed: 11242054]
8. Nakayama J, Rice JC, Strahl BD, Allis CD & Grewal SI Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science* 292, 110–113 (2001). [PubMed: 11283354]
9. Kaustov L et al. Recognition and specificity determinants of the human cbx chromodomains. *J. Biol. Chem.* 286, 521–529 (2011). [PubMed: 21047797]
10. Kokura K, Sun L, Bedford MT & Fang J Methyl-H3K9-binding protein MPP8 mediates E-cadherin gene silencing and promotes tumour cell motility and invasion. *EMBO J.* 29, 3673–3687 (2010). [PubMed: 20871592]
11. Chang Y, Horton JR, Bedford MT, Zhang X & Cheng X Structural insights for MPP8 chromodomain interaction with histone H3 lysine 9: potential effect of phosphorylation on methyl-lysine binding. *J. Mol. Biol.* 408, 807–814 (2011). [PubMed: 21419134]
12. Li J et al. Structural basis for specific binding of human MPP8 chromodomain to histone H3 methylated at lysine 9. *PLoS One* 6, e25104 (2011). [PubMed: 22022377]
13. Maksakova IA et al. H3K9me3-binding proteins are dispensable for SETDB1/H3K9me3-dependent retroviral silencing. *Epigenetics Chromatin* 4, 12 (2011). [PubMed: 21774827]
14. Babaian A & Mager DL Endogenous retroviral promoter exaptation in human cancer. *Mob. DNA* 7, 24 (2016). [PubMed: 27980689]
15. Geis FK & Goff SP Silencing and transcriptional regulation of endogenous retroviruses: an overview. *Viruses* 12, 884 (2020). [PubMed: 32823517]
16. Mao J, Zhang Q & Cong YS Human endogenous retroviruses in development and disease. *Comput. Struct. Biotechnol. J.* 19, 5978–5986 (2021). [PubMed: 34849202]
17. Fueyo R, Judd J, Feschotte C & Wysocka J Roles of transposable elements in the regulation of mammalian transcription. *Nat. Rev. Mol. Cell Biol.* 23, 481–497 (2022). [PubMed: 35228718]
18. Friedli M & Trono D The developmental control of transposable elements and the evolution of higher species. *Annu. Rev. Cell Dev. Biol.* 31, 429–451 (2015). [PubMed: 26393776]
19. Rea S et al. Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* 406, 593–599 (2000). [PubMed: 10949293]
20. Schultz DC, Ayyanathan K, Negorev D, Maul GG & Rauscher FJ 3rd 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). [PubMed: 11959841]
21. Wang H et al. mAM facilitates conversion by ESET of dimethyl to trimethyl lysine 9 of histone H3 to cause transcriptional repression. *Mol. Cell* 12, 475–487 (2003). [PubMed: 14536086]
22. Schmidtman E, Anton T, Rombaut P, Herzog F & Leonhardt H Determination of local chromatin composition by CasID. *Nucleus* 7, 476–484 (2016). [PubMed: 27676121]
23. Chelmicki T et al. m6A RNA methylation regulates the fate of endogenous retroviruses. *Nature* 591, 312–316 (2021). [PubMed: 33442060]
24. Xu W et al. METTL3 regulates heterochromatin in mouse embryonic stem cells. *Nature* 591, 317–321 (2021). [PubMed: 33505026]
25. Liu J et al. The RNA m6A reader YTHDC1 silences retrotransposons and guards ES cell identity. *Nature* 591, 322–326 (2021). [PubMed: 33658714]
26. Yang BX et al. Systematic identification of factors for provirus silencing in embryonic stem cells. *Cell* 163, 230–245 (2015). [PubMed: 26365490]
27. Fukuda K, Okuda A, Yusa K & Shinkai Y A CRISPR knockout screen identifies SETDB1-target retroelement silencing factors in embryonic stem cells. *Genome Res.* 28, 846–858 (2018). [PubMed: 29728365]

28. Groh S et al. Morc3 silences endogenous retroviruses by enabling Daxx-mediated histone H3.3 incorporation. *Nat. Commun.* 12, 5996 (2021). [PubMed: 34650047]
29. Roulois D et al. DNA-demethylating agents target colorectal cancer cells by inducing viral mimicry by endogenous transcripts. *Cell* 162, 961–973 (2015). [PubMed: 26317465]
30. Chiappinelli KB et al. Inhibiting DNA methylation causes an interferon response in cancer via dsRNA including endogenous retroviruses. *Cell* 162, 974–986 (2015). [PubMed: 26317466]
31. Yu X et al. The long terminal repeat (LTR) of ERV-9 human endogenous retrovirus binds to NF-Y in the assembly of an active LTR enhancer complex NF-Y/MZF1/GATA-2. *J. Biol. Chem.* 280, 35184–35194 (2005). [PubMed: 16105833]
32. Kuo AJ et al. The BAH domain of ORC1 links H4K20me2 to DNA replication licensing and Meier–Gorlin syndrome. *Nature* 484, 115–119 (2012). [PubMed: 22398447]
33. Zhao D et al. The BAH domain of BAHD1 is a histone H3K27me3 reader. *Protein Cell* 7, 222–226 (2016). [PubMed: 26850261]
34. Fan H et al. BAHCC1 binds H3K27me3 via a conserved BAH module to mediate gene silencing and oncogenesis. *Nat. Genet.* 52, 1384–1396 (2020). [PubMed: 33139953]
35. Fan H et al. A conserved BAH module within mammalian BAHD1 connects H3K27me3 to Polycomb gene silencing. *Nucleic Acids Res.* 49, 4441–4455 (2021). [PubMed: 33823544]
36. Ren W et al. DNMT1 reads heterochromatic H4K20me3 to reinforce LINE-1 DNA methylation. *Nat. Commun.* 12, 2490 (2021). [PubMed: 33941775]
37. Fischle W et al. Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. *Nature* 438, 1116–1122 (2005). [PubMed: 16222246]
38. Yang D et al. Na-acetylated Sir3 stabilizes the conformation of a nucleosome-binding loop in the BAH domain. *Nat. Struct. Mol. Biol.* 20, 1116–1118 (2013). [PubMed: 23934152]
39. Rowe HM et al. TRIM28 repression of retrotransposon-based enhancers is necessary to preserve transcriptional dynamics in embryonic stem cells. *Genome Res.* 23, 452–461 (2013). [PubMed: 23233547]
40. Brocks D et al. DNMT and HDAC inhibitors induce cryptic transcription start sites encoded in long terminal repeats. *Nat. Genet.* 49, 1052–1060 (2017). [PubMed: 28604729]
41. Takahashi H, Lassmann T, Murata M & Carninci P 5' end-centered expression profiling using cap-analysis gene expression and next-generation sequencing. *Nat. Protoc.* 7, 542–561 (2012). [PubMed: 22362160]
42. Beyer U, Kronung SK, Leha A, Walter L & Dobbelstein M Comprehensive identification of genes driven by ERV9-LTRs reveals TNFRSF10B as a re-activatable mediator of testicular cancer cell death. *Cell Death Differ.* 23, 64–75 (2016). [PubMed: 26024393]
43. Perissi V, Jepsen K, Glass CK & Rosenfeld MG Deconstructing repression: evolving models of co-repressor action. *Nat. Rev. Genet.* 11, 109–123 (2010). [PubMed: 20084085]
44. Banks CAS et al. Differential HDAC1/2 network analysis reveals a role for prefoldin/CCT in HDAC1/2 complex assembly. *Sci. Rep.* 8, 13712 (2018). [PubMed: 30209338]
45. Rowe HM et al. KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature* 463, 237–240 (2010). [PubMed: 20075919]
46. Le Guezennec X, Vermeulen M & Stunnenberg HG Molecular characterization of Sin3 PAH-domain interactor specificity and identification of PAH partners. *Nucleic Acids Res.* 34, 3929–3937 (2006). [PubMed: 16914451]
47. Wolf G et al. The KRAB zinc finger protein ZFP809 is required to initiate epigenetic silencing of endogenous retroviruses. *Genes Dev.* 29, 538–554 (2015). [PubMed: 25737282]
48. Dan J et al. Rif1 maintains telomere length homeostasis of ESCs by mediating heterochromatin silencing. *Dev. Cell* 29, 7–19 (2014). [PubMed: 24735877]
49. Almeida MV, Vernaz G, Putman ALK & Miska EA Taming transposable elements in vertebrates: from epigenetic silencing to domestication. *Trends Genet.* 38, 529–553 (2022). [PubMed: 35307201]
50. Bruno M, Mahgoub M & Macfarlan TS The arms race between KRAB-zinc finger proteins and endogenous retroelements and its impact on mammals. *Annu. Rev. Genet.* 53, 393–416 (2019). [PubMed: 31518518]

51. Yu M et al. A resource for cell line authentication, annotation and quality control. *Nature* 520, 307–311 (2015). [PubMed: 25877200]
52. Stauffer W, Sheng H & Lim HN EzColocalization: an ImageJ plugin for visualizing and measuring colocalization in cells and organisms. *Sci. Rep.* 8, 15764 (2018). [PubMed: 30361629]
53. Gao L et al. Structure of DNMT3B homo-oligomer reveals vulnerability to impairment by ICF mutations. *Nat. Commun.* 13, 4249 (2022). [PubMed: 35869095]
54. Ahn JH et al. Phase separation drives aberrant chromatin looping and cancer development. *Nature* 595, 591–595 (2021). [PubMed: 34163069]
55. Hsieh JJ, Cheng EH & Korsmeyer SJ Taspase1: a threonine aspartase required for cleavage of MLL and proper HOX gene expression. *Cell* 115, 293–303 (2003). [PubMed: 14636557]
56. Roux KJ, Kim DI, Burke B & May DG BioID: a screen for protein–protein interactions. *Curr. Protoc. Protein Sci.* 91, 19.23.1–19.23.15 (2018).
57. Li J et al. ZMYND11–MBTD1 induces leukemogenesis through hijacking NuA4/TIP60 acetyltransferase complex and a PWWP-mediated chromatin association mechanism. *Nat. Commun.* 12, 1045 (2021). [PubMed: 33594072]
58. Szklarczyk D et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 47, D607–D613 (2019). [PubMed: 30476243]
59. Nesvizhskii AI, Keller A, Kolker E & Aebersold R A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* 75, 4646–4658 (2003). [PubMed: 14632076]
60. Minor W, Cymborowski M, Otwinowski Z & Chruszcz M HKL-3000: the integration of data reduction and structure solution—from diffraction images to an initial model in minutes. *Acta Crystallogr. D Biol. Crystallogr.* 62, 859–866 (2006). [PubMed: 16855301]
61. McCoy AJ et al. Phaser crystallographic software. *J. Appl. Crystallogr.* 40, 658–674 (2007). [PubMed: 19461840]
62. Emsley P & Cowtan K Coot: model-building tools for molecular graphics. *Acta Crystallogr.* 60, 2126–2132 (2004).
63. Adams PD et al. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr.* 58, 1948–1954 (2002).
64. Ren W et al. Direct readout of heterochromatic H3K9me3 regulates DNMT1-mediated maintenance DNA methylation. *Proc. Natl Acad. Sci. USA* 117, 18439–18447 (2020). [PubMed: 32675241]
65. Dyer PN et al. Reconstitution of nucleosome core particles from recombinant histones and DNA. *Methods Enzymol.* 375, 23–44 (2004). [PubMed: 14870657]
66. Zhang W, Sankaran S, Gozani O & Song J A Meier–Gorlin syndrome mutation impairs the ORC1–nucleosome association. *ACS Chem. Biol.* 10, 1176–1180 (2015).
67. Simon MD et al. The site-specific installation of methyl-lysine analogs into recombinant histones. *Cell* 128, 1003–1012 (2007). [PubMed: 17350582]
68. Ren Z et al. PHF19 promotes multiple myeloma tumorigenicity through PRC2 activation and broad H3K27me3 domain formation. *Blood* 134, 1176–1189 (2019). [PubMed: 31383640]
69. Dobin A et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21 (2013). [PubMed: 23104886]
70. Patro R, Duggal G, Love MI, Irizarry RA & Kingsford C Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* 14, 417–419 (2017). [PubMed: 28263959]
71. Love MI, Huber W & Anders S Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550 (2014). [PubMed: 25516281]
72. Subramanian A et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* 102, 15545–15550 (2005). [PubMed: 16199517]
73. Hanzelmann S, Castelo R & Guinney J GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* 14, 7 (2013). [PubMed: 23323831]
74. Kong Y et al. Transposable element expression in tumors is associated with immune infiltration and increased antigenicity. *Nat. Commun.* 10, 5228 (2019). [PubMed: 31745090]

75. Jeong HH, Yalamanchili HK, Guo C, Shulman JM & Liu Z An ultra-fast and scalable quantification pipeline for transposable elements from next generation sequencing data. *Pac. Symp. Biocomput.* 23, 168–179 (2018). [PubMed: 29218879]
76. Ohmiya H et al. RECLU: a pipeline to discover reproducible transcriptional start sites and their alternative regulation using capped analysis of gene expression (CAGE). *BMC Genomics* 15, 269 (2014). [PubMed: 24779366]
77. Skene PJ, Henikoff JG & Henikoff S Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nat. Protoc.* 13, 1006–1019 (2018). [PubMed: 29651053]
78. Wang J et al. EZH2 noncanonically binds cMyc and p300 through a cryptic transactivation domain to mediate gene activation and promote oncogenesis. *Nat. Cell Biol.* 24, 384–399 (2022). [PubMed: 35210568]
79. Wang J et al. A cryptic transactivation domain of EZH2 binds AR and AR's splice variant, promoting oncogene activation and tumorous transformation. *Nucleic Acids Res.* 50, 10929–10946 (2022). [PubMed: 36300627]
80. Langmead B, Trapnell C, Pop M & Salzberg SL Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10, R25 (2009). [PubMed: 19261174]
81. Zhang Y et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9, R137 (2008). [PubMed: 18798982]
82. Zhu LJ et al. ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics* 11, 237 (2010). [PubMed: 20459804]
83. Heinz S et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* 38, 576–589 (2010). [PubMed: 20513432]
84. Ramirez F et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* 44, W160–W165 (2016). [PubMed: 27079975]
85. Liao Y, Smyth GK & Shi W featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930 (2014). [PubMed: 24227677]
86. Xu J et al. Subtype-specific 3D genome alteration in acute myeloid leukaemia. *Nature* 611, 387–398 (2022). [PubMed: 36289338]
87. Krueger F & Andrews SR Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* 27, 1571–1572 (2011). [PubMed: 21493656]
88. Garfield AS Derivation of primary mouse embryonic fibroblast (PMEF) cultures. *Methods Mol. Biol.* 633, 19–27 (2010). [PubMed: 20204617]
89. Jakobsson J & Vincendeau M SnapShot: human endogenous retroviruses. *Cell* 185, 400–400.e1 (2022). [PubMed: 35063077]
90. Consortium GTEx. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* 45, 580–585 (2013). [PubMed: 23715323]



**Fig. 1 | TNRC18 functions as an ERV repressor.**

a, CasID identified ERV-bound proteins in HEK293 cells, with hits ranked by the fold change of normalized spectral abundance factor (NSAF) relative to control. Label font colors: green, zinc finger proteins; blue, histone and DNA-methylation-related factors; yellow, histone chaperones; purple, N6-methyladenosine-related factors; red, TNRC18.

b,c, Scatter plots showing TE families exhibiting expression changes based on RNA-seq of HEK293 cells with endogenous TNRC18 KD (using TNRC18 shRNA (shTNRC18); b) relative to TNRC18 KD followed by TNRC18 re-expression (shTNRC18-rescue; n = 2 independent experiments), or cells with endogenous TNRC18 KO (TNRC18-KO; c) relative to WT (n = 3 independent experiments). The significance cut-off is the fold change in expression over 1.50 and adjusted P value less than 0.01 for transcripts with base mean read counts over 10. Adjusted P value was calculated using negative binomial model-based methods (DESeq2).

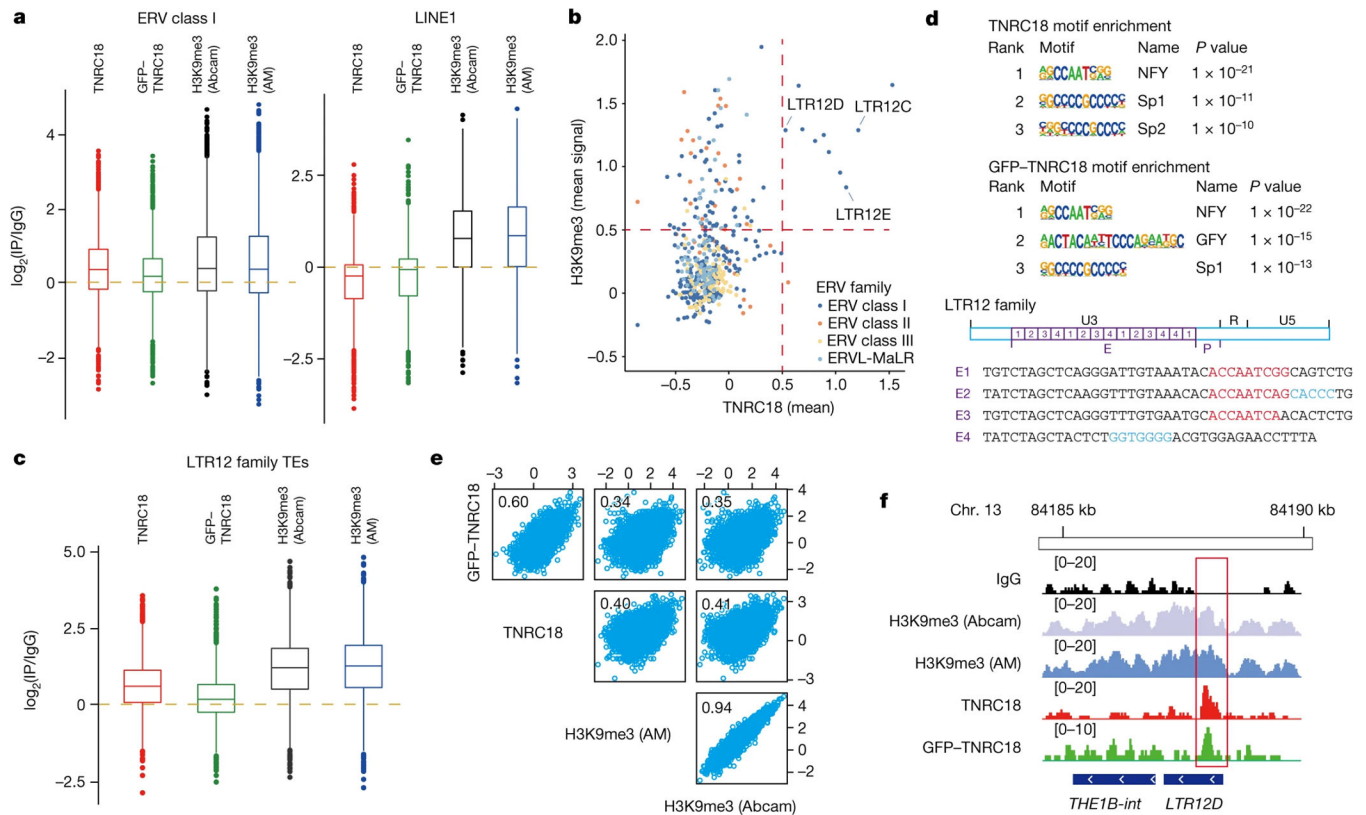


d, RT-qPCR for TEs (top) and immunity-related genes (bottom) in HEK293 cells with shTNRC18 versus control shRNA (shControl) or shTNRC18- rescue, or TNRC18 KO versus WT (n = 3 independent experiments; plotted as mean  $\pm$  s.d. after normalization to GAPDH and control samples). \*P < 0.05; \*\*P < 0.01; \*\*\*P < 0.001; \*\*\*\*P < 0.0001, two-sided t-test. Exact P values are shown in Supplementary Table 8.

e,f, GSEA revealed pathway enrichment in cells with shTNRC18 versus shControl (e), or TNRC18 KO versus WT (f). Immunity-related gene sets are labelled in red. The y axis and x axis show normalized enrichment score (NES) and false discovery rate (FDR) q values, respectively.

g, RT-qPCR for TEs in the indicated TNRC18 KO cells versus WT cells (n = 3 independent experiments; plotted as the mean  $\pm$  s.d. after normalization to GAPDH and to WT). \*P < 0.05; \*\*P < 0.01; \*\*\*P < 0.001; \*\*\*\*P < 0.0001, two-sided t-test. Exact P values are shown in Supplementary Table 8.

h,i, RNA-seq revealed TEs exhibiting expression changes in SNU-1 cells (h) and HT-29 cells (i) with TNRC18 KO versus WT (n = 3 independent experiments). Significance cut-off is the same as in b,c. Adjusted P values were calculated using negative binomial model-based methods (DESeq2).



**Fig. 2 | TNRC18 colocalizes with H3K9me3 at ERV regions.**

a,c, Box plots showing  $\log_2$  fold change values in CUT&RUN signals of the indicated factor, TNRC18 (probed with either endogenous TNRC18 antibodies or anti-GFP antibodies for GFP-tagged TNRC18) or H3K9me3 (probed with antibodies from Abcam or Active Motif (AM)), compared with those from IgG controls, at ERV1 (a, left) including the LTR12 subfamily of ERV1 (c), or LINE1 regions (a, right). The boundaries of box plots indicate the 25th and 75th percentiles, the centre line indicates the median, and the whiskers indicate  $1.5 \times$  the interquartile range. Sample sizes of each box plot in a–c are listed in Supplementary Table 8. IP, immunoprecipitation.

b, Scatter plots showing averaged CUT&RUN signals of TNRC18 and H3K9me3 at the indicated individual ERV subfamily.

d, Motif search analysis revealing the most enriched motifs at the called TNRC18 (top) or GFP-TNRC18 (middle) peaks. The bottom panel illustrates LTR12 at the  $\beta$ -globin gene locus. The LTR consists of U3 (unique 3'-end), R (repeated) and U5 (unique 5'-end) regions. The U3 region includes an enhancer (E), which contains 14 tandemly repeated elements of E1, E2, E3 and E4, and a promoter (P). Nucleotide colors (bottom): red, NFY motif in E1–E4; blue, Sp1/2 motif.

e, Scatter plots showing the correlation among CUT&RUN signals of TNRC18, GFP-TNRC18 and H3K9me3 probed with two independent antibodies at the LTR12 family TEs. Pearson correlation coefficient values are indicated at the upper left of each plot.

f, Integrated genome viewer (IGV) tracks of the indicated CUT&RUN signal at the genomic locus containing THE1B-int and LTR12D. Box highlights the enrichment of TNRC18 at

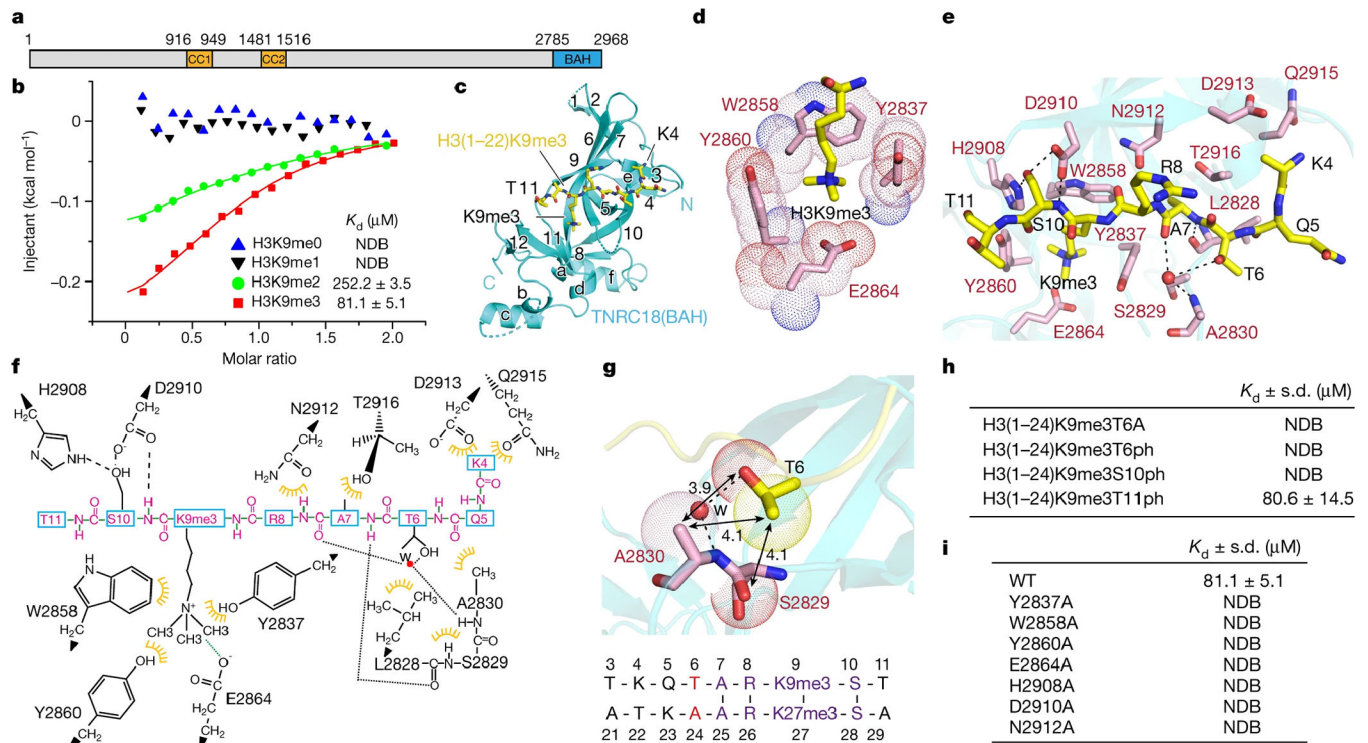
the U3 region of LTR12D. Bottom panel of d is adapted from ref. 31, under a Creative Commons CC-BY 4.0 license.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Fig. 3 |. The BAH module of TNRC18 specifically reads H3K9me3.**

a, Domain architecture of human TNRC18, with the boundaries of individual domains marked.

b, ITC binding curves for human TNRC18(BAH) and peptides with different degrees of H3K9 methylation. NDB, no detectable binding.

c, Overall structure of TNRC18(BAH) (cyan) bound to H3K9me3 peptide (yellow sticks). The  $\alpha$ -helices and  $\beta$ -strands are counted in alphabetic and numeric orders, respectively.

d, Positioning of the H3K9me3 side chain (yellow) within the aromatic cage of TNRC18(BAH). The cage residues (pink) of human TNRC18(BAH) are labelled in red.

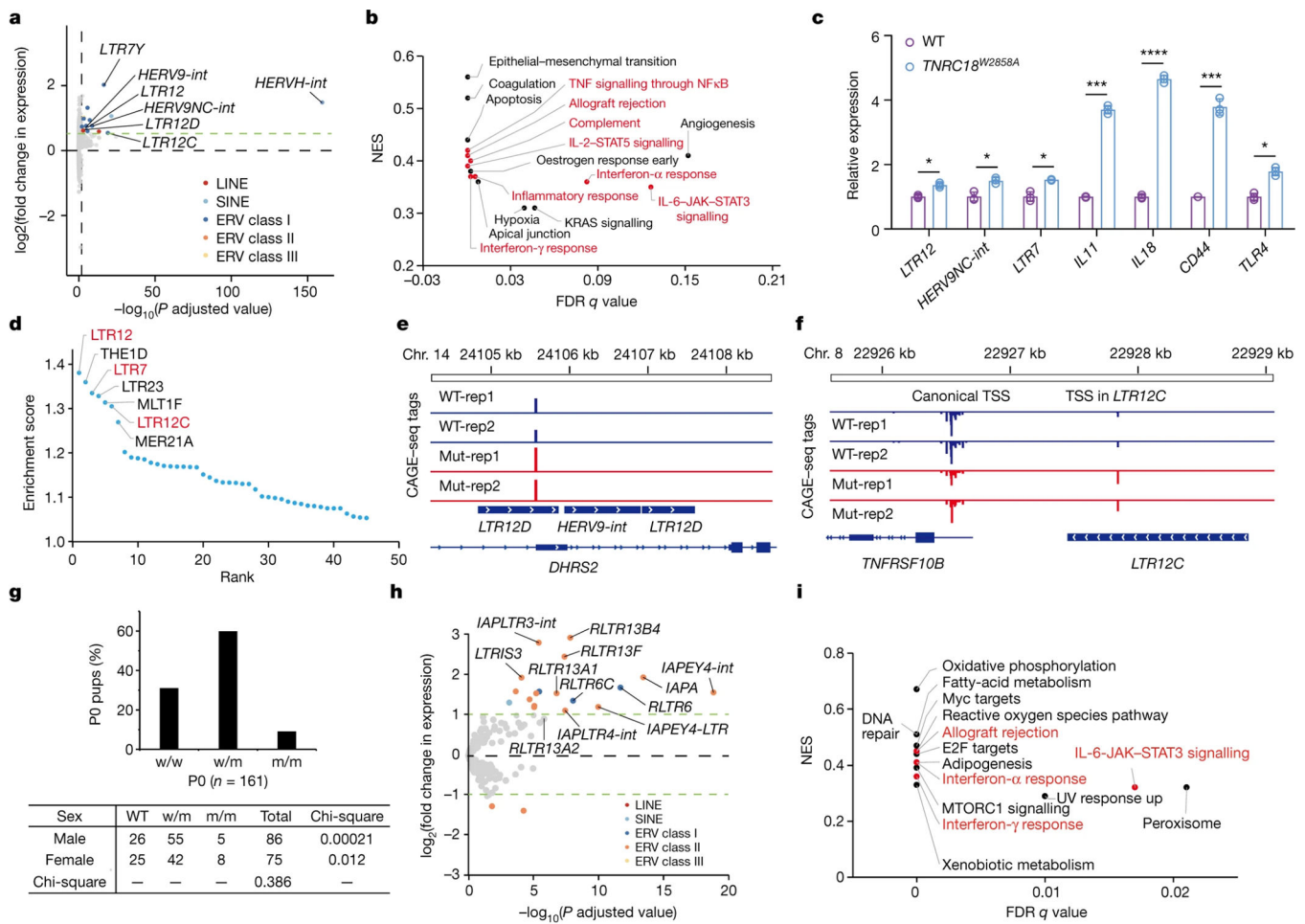
e, Close-up view of TNRC18(BAH)–H3K9me3 peptide interactions. The interacting residues of TNRC18(BAH) and peptide are shown as pink and yellow sticks, respectively. Dashed lines denote a hydrogen bond.

f, Schematic of TNRC18(BAH) (black) and H3K9me3 (magenta) interactions. Hydrogen bonds and electrostatic interactions are shown as black and green dashed lines, respectively. Hydrophobic and van der Waals interactions are in yellow.

g, Top, close-up view of the interaction between Thr6 of H3K9me3 peptide and TNRC18(BAH). Hydrogen bonds are shown as dashed lines. Van der Waals interactions are shown as solid lines, with distance labelled in angstrom units. Bottom, sequence alignment of the histone H3 sequences around H3K9me3 and H3K27me3.

h, Summary of ITC measurements for WT TNRC18(BAH) binding to H3 peptides with the indicated modification.

i, Summary of ITC measurement for binding by TNRC18(BAH), WT or the indicated mutant to H3K9me3 peptide.



**Fig. 4 | TNRC18(BAH)-H3K9me3 engagement mediates TE silencing.**

a, RNA-seq identified TE subfamilies exhibiting expression changes in HEK293 cells with a TNRC18(BAH) homozygous knock-in (KI) mutation (W2858A) versus WT (n = 2 independent experiments). Significance cut-off is the fold change in expression over 1.50 and adjusted P value less than 0.01 (calculated using negative binomial model-based methods and DESeq2) for transcripts with base mean read counts over 10.

b, GSEA revealed the gene set enrichment in HEK293 cells with a TNRC18W2858A homozygous mutation versus WT (n = 2 independent experiments).

c, RT-qPCR for the indicated TEs or genes in HEK293 cells with a TNRC18W2858A homozygous mutation versus WT (n = 3 independent experiments; plotted as the mean  $\pm$  s.d. after normalization to GAPDH and to WT). \*P < 0.05, \*\*P < 0.01, \*\*\*P < 0.001, \*\*\*\*P < 0.0001, two-sided t-test. Exact P values are shown in Supplementary Table 8.

d, CAGE-seq revealing enrichment for the indicated TEs activated in HEK293 cells with a TNRC18W2858A homozygous mutation versus WT (n = 2 independent experiments).

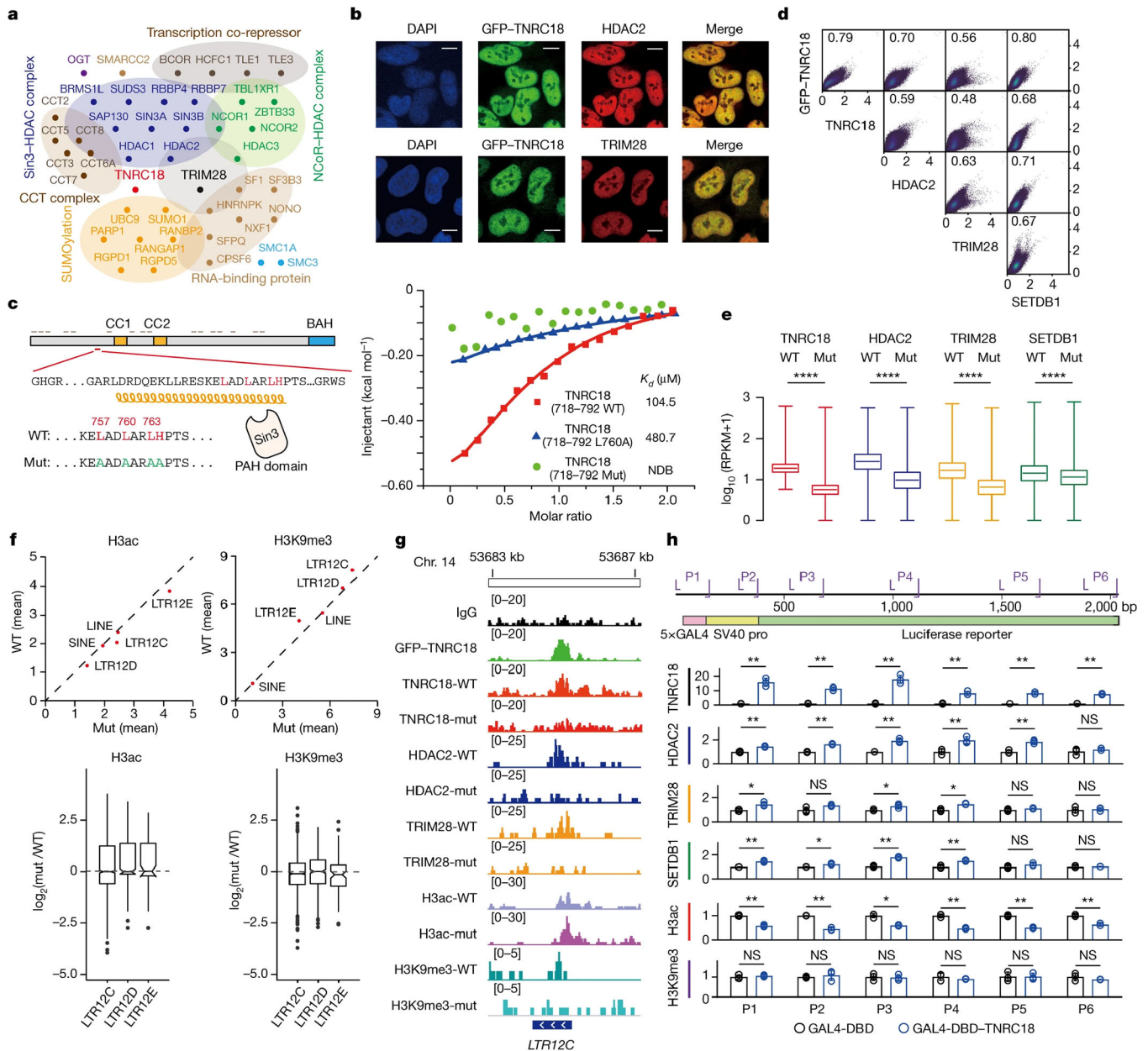
e,f, IGV view of CAGE-seq signals at LTR12 close to DHRS2 (e) and TNFRSF10B (f) in HEK293 cells, with WT TNRC18 or TNRC18W2858A homozygous mutation. Rep, replicate.

g, Bar chart (top) and Chi-square test (bottom) of the indicated genotype, WT (w/w; n = 51) or with heterozygous (w/m; n = 97) or homozygous (m/m; n = 13) Tnrc18W2745A/

Y2747A mutation, among newly born pups (P0) from breeding of Tnrc18W2745A/Y2747A heterozygous mice. n, cohort size.

h, RNA-seq revealed the indicated TEs that exhibited expression changes in primary lung tissues from mice with a Tnrc18W2745A/Y2747A mutation (n = 3 mice with two technical replicates each) compared with those of WT littermates (n = 2 mice with two technical replicates each). Significance cut-off is the same as in a, except that the expression fold change is set to 2. Adjusted P value was calculated using negative binomial model-based methods (DESeq2).

i, GSEA revealed pathway enrichment in lungs from mice with a Tnrc18W2745A/Y2747A mutation versus WT. For b and i, immunity-related gene sets are labelled in red.



**Fig. 5 | TNRC18 interacts with co-repressors to maintain repression.**

a, BioID of HEK293 cells identified TNRC18-associated proteins, grouped in different multi-subunit complexes using STRING.

b, Immunofluorescence of the indicated protein in HEK293 cells. Representative results from three independent experiments. Scale bar, 10  $\mu$ m.

c, Right, ITC fitting curves of WT or mutant (Mut) TNRC18 amino acids 718–792 titrated against the PAH domain of Sin3.

d, Scatter plots showing CUT&RUN signal correlations in HEK293 cells, with Pearson correlation coefficient values indicated at the upper left.

e, Plots for the indicated CUT&RUN signals at TNRC18 peaks (n = 7,545; common peaks based on both TNRC18 and GFP–TNRC18 CUT&RUN) in HEK293 cells, WT or

TNRC18(W2858A)-mutant. RPKM, reads per kilobase per million mapped reads. Boxes indicate 25th to 75th percentile, whiskers the full data range, and the centre line the median. P values were calculated using two-sided Wilcoxon matched-pairs signed-rank test. \*\*\*\*P < 0.0001. Exact P values are shown in Supplementary Table 8.

f, Top, averaged H3ac and H3K9me3 CUT&RUN signals at LTR12 families (LTR12C, LTR12D and LTR12E), LINEs and SINEs in WT (y axis) or TNRC18(W2858A)-mutant (x axis) HEK293 cells. Bottom, box plots of log<sub>2</sub> values of the ratios for CUT&RUN counts at TEs in TNRC18(W2858A)-mutant versus WT cells. Box boundaries indicate 25th and 75th percentiles, the centre line the median, and whiskers 1.5× the interquartile range. Sample sizes are provided in Supplementary Table 8.

g, CUT&RUN signals for the indicated factor at a LTR12C in chromosome 14 in WT or TNRC18(W2858A)-mutant HEK293 cells. h, ChIP-qPCR for the indicated factor at the integrated 5×GAL4-luciferase reporter regions (P1–P6; top) in cells stably transduced with GAL4-DBD–TNRC18 relative to GAL4-DBD (n = 3 independent experiments; plotted as the mean ± s.d.). The y axis shows relative signals after normalization to input and to GAL4-DBD-expressing cells. \*P < 0.05, \*\*P < 0.01, \*\*\*P < 0.001, \*\*\*\*P < 0.0001, two-sided t-test. NS, not significant. Exact P values are shown in Supplementary Table 8.