nature cell biology

Article

https://doi.org/10.1038/s41556-022-01008-5

KCNQ1OT1 promotes genome-wide transposon repression by guiding RNA–DNA triplexes and HP1 binding

Received: 18 December 2021

Accepted: 7 September 2022

Published online: 20 October 2022

Check for updates

Xiaoli Zhang[1,7], Quanlong Jiang[2,3,7], Jiyang Li[2,3,4,7], Shiqiang Zhang[1,2,3,7], Yaqiang Cao[2,3], Xian Xia[5], Donghong Cai[2,3], Jiaqi Tan[1], Jiekai Chen[6] and Jing-Dong J. Han[1]

Transposon (de)repression and heterochromatin reorganization are dynamically regulated during cell fate determination and are hallmarks of cellular senescence. However, whether they are sequence specifically regulated remains unknown. Here we uncover that the KCNQ1OT1 lncRNA, by sequence-specific Hoogsteen base pairing with double-stranded genomic DNA via its repeat-rich region and binding to the heterochromatin protein HP1α, guides, induces and maintains epigenetic silencing at specific repetitive DNA elements. Repressing KCNQ1OT1 or deleting its repeat-rich region reduces DNA methylation and H3K9me3 on KCNQ1OT1-targeted transposons. Engineering a fusion KCNQ1OT1 with an ectopically targeting guiding triplex sequence induces de novo DNA methylation at the target site. Phenotypically, repressing KCNQ1OT1 induces senescence-associated heterochromatin foci, transposon activation and retrotransposition as well as cellular senescence, demonstrating an essential role of KCNQ1OT1 to safeguard against genome instability and senescence.

Transposon activation and heterochromatin reorganization are critical events in cell fate determination and senescence[1–4]. A most prominent and common early event during cellular senescence is the formation of H3K9 trimethylation (H3K9me3) foci in the nucleus, which are called senescence-associated heterochromatin foci (SAHF). SAHF often contain DNA sequence from both coding and noncoding regions, and are characterized by several heterochromatin marks, including H3K9me3 and HP1 proteins[4–6]. Recent studies also revealed several long noncoding RNAs (lncRNAs) involved in heterochromatin formation and stability[7–10], while many other lncRNAs are related to senescent phenotypes[11–14]. For example, we recently found many lncRNAs to regulate NFkB signalling and hence regulate senescence-associated secretory phenotypes (SASPs)[14], and another study demonstrates that the lncRNA GUARDIN is necessary for maintaining the stability of telomeres[10]. Transposon repression and activation are often sequence specific—for example, only the evolutionarily young AluY subfamily, but not older Alu subfamilies, are repressed in immune cells[15]. Why only a select set of transposons are activated or repressed during development and ageing, and what makes the broad domain epigenetic (de)repression sequence specific remain elusive. It is also unknown whether any lncRNAs regulate senescence through shaping transposon activity and heterochromatin landscape. KCNQ1OT1 is a paternally expressed lncRNA and correlates with transcriptional silencing of eight to ten protein-coding genes in *cis*[16–18]. The more than 91 kb-long KCNQ1OT1

[1]Peking-Tsinghua Center for Life Sciences, Academy for Advanced Interdisciplinary Studies, Center for Quantitative Biology (CQB), Peking University, Beijing, China. [2]CAS Key Laboratory of Computational Biology, Shanghai Institute of Nutrition and Health, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China. [3]University of Chinese Academy of Sciences, Beijing, China. [4]School of Life Science and Technology, ShanghaiTech University, Shanghai, China. [5]Department of Pharmacology, Nanjing University of Chinese Medicine, Nanjing, China. [6]CAS Key Laboratory of Regenerative Biology, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou, China. [7]These authors contributed equally: Xiaoli Zhang, Quanlong Jiang, Jiyang Li, Shiqiang Zhang. e-mail: jackie.han@pku.edu.cn

https://doi.org/10.1038/s41556-022-01008-5

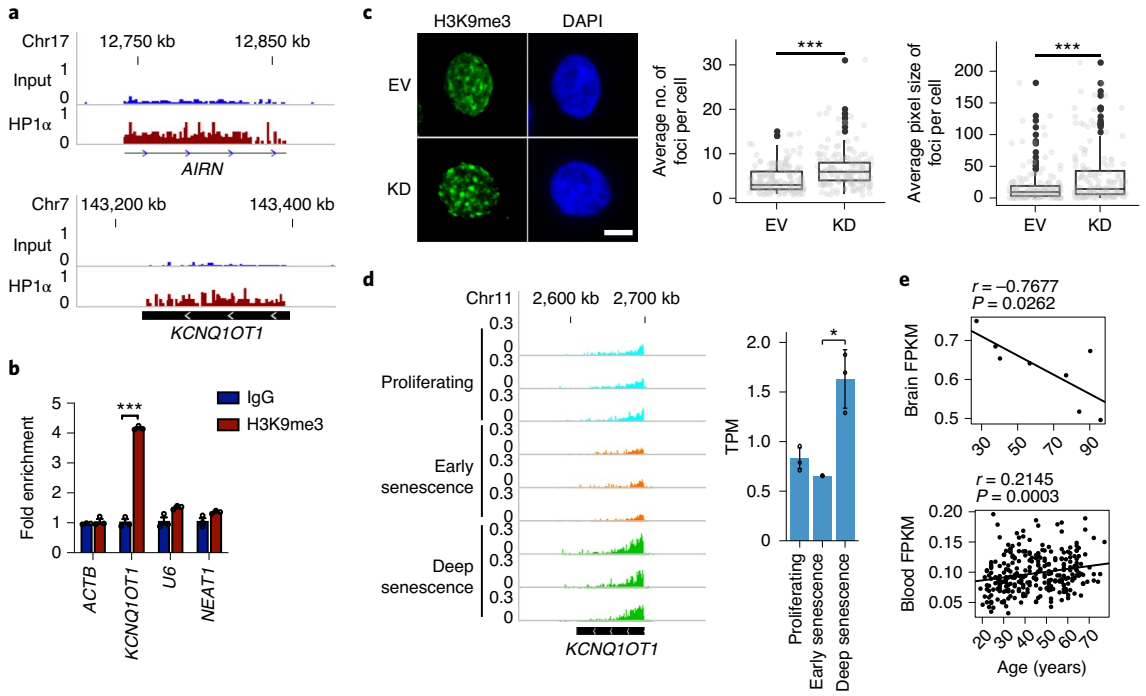


Fig. 1 | The lncRNA KCNQ1OT1 binds HP1 and suppresses H3K9me3 foci. a, Representative normalized HP1α eCLIP–seq reads on the lncRNAs AIRN and KCNQ1OT1 in NIH3T3 cells. **b**, RIP using antibodies to control immunoglobulin G (IgG) or H3K9me3 in human HEK293T cells. Data are the mean ± s.e.m. relative to the input ($n = 3$ independent experiments). *ACTB*, *U6* small nuclear RNA and *NEAT1* were used as negative controls. **c**, Increase in H3K9me3 foci in *KCNQ1OT1*-KD HEK293T cells as shown by immunofluorescence staining of H3K9me3 (left). Nuclei were counterstained with 4,6-diamidino-2-phenylindole (DAPI). Scale bar, 5 µm. Average number of H3K9me3 foci per cell (middle) and average size of H3K9me3 foci per cell (right; $n = 183$ (EV) and 159 *KCNQ1OT1*-KD (KD) cells pooled across three independent experiments). Box-and-whisker plots: the bound of a box shows the 25% (lower quartile) and 75% (upper quartile) of all values in the group; the line in the box shows the median value; and values outside the middle 50% values (interquartile range, IQR) but within 1.5×IQR are shown as whiskers, otherwise as black dots. All values are shown as grey spots. **d**, RNA-seq tracks showing the expression of KCNQ1OT1 in proliferating, early senescence and deep senescence human lung fibroblast cells (left). Levels of KCNQ1OT1 expression in proliferating, early senescence and deep senescence human lung fibroblast cells (right). TPM, transcripts per kilobase million. **e**, Expression pattern of KCNQ1OT1 in human brain tissue (top; National Omics Data Encyclopedia, OEP001041) and blood (bottom; GSE106670) derived from different ages. Pearson correlation coefficient (r) and P values are shown. **b**–**d**, Statistical significance was determined using a two-tailed Student's t-test; $^{*}P < 0.05$ and $^{***}P < 0.001$. Source numerical data are provided.

contains several repeat motifs at the 5′ end and also harbours a large approximately 50-kb repeat-rich (predominately long interspersed nuclear element) region at the 3′ half of KCNQ1OT1 transcripts[19,20]. The repeats located at the 5′ end right after the transcription start site of *KCNQ1OT1* are conserved in mammals[21] and play an important role in silencing the flanking genes, whereas the function of the repeat-rich region in *KCNQ1OT1* is unknown[18,22,23]. Furthermore, KCNQ1OT1 was previously assumed to only function in *cis* to repress nearby coding genes[18,19,24]. However, its *cis*-only action cannot explain the much higher expression levels of KCNQ1OT1 (fragments per kilobase of exon per million mapped fragments (FPKM) of 2.6 in HEK293T cells according to our RNA-sequencing (RNA-seq) data) than most lncRNAs (Extended Data Fig. 1a), such as HOTTIP (FPKM of 0.21), a well-known *trans*-acting lncRNA[25]. The difference between *cis*- and *trans*-acting is that sequence specificity is needed for *trans* targeting but not for *cis* targeting. Here we identified through HP1α enhanced crosslinking and immunoprecipitation (eCLIP) with sequencing (eCLIP–seq), computational prediction and detailed molecular characterization that KCNQ1OT1 lncRNA guides DNA methylation, H3K9me3 and transposon repression in *trans* at specific repetitive DNA elements genome wide. We find that, guided by forming an RNA–double-stranded DNA (dsDNA) triplex between the repeat-rich region of KCNQ1OT1 and the target DNA as well as simultaneous binding to HP1α, KCNQ1OT1 is required for maintaining H3K9me3 and DNA methylation on its targeting sequences to represses their transcription. Deletion of the repeat-rich region of KCNQ1OT1 alone leads to activation of KCNQ1OT1-targeted Alu, L1 and satellite repeat elements, and knockdown (KD) of KCNQ1OT1 results in spontaneous cellular senescence. In addition, engineering a new guiding sequence is sufficient to ectopically target KCNQ1OT1 for DNA methylation.

Results

KCNQ1OT1 is associated with heterochromatin

Several studies showed that HP1 protein not only contains chromodomain that binds to H3K9me3 but also possess an RNA-binding domain[26–28]. Using HP1α eCLIP[29] in NIH3T3 cells, we found two antisense lncRNAs, AIRN and KCNQ1OT1, were more than twofold enriched by HP1α (Fig. 1a). These two lncRNAs are known as imprinting genes, playing a role in silencing the parental genes in *cis* and regulating histone H3 lysine 9 methylation through interaction with G9a during early development[19,30]. We confirmed the interactions of KCNQ1OT1 with heterochromatin through H3K9me3 RNA immunoprecipitation (RIP), followed by complementary DNA synthesis and quantitative real-time PCR (qPCR) in both NIH3T3 and HEK293T cells. The RIP results showed that KCNQ1OT1 was specifically enriched by antibodies to H3K9me3 (Fig. 1b and Extended Data Fig. 1b). Given that KCNQ1OT1 is highly conserved in the genomic sequence, position and structure in mammals, and more ubiquitously and highly expressed among different tissues compared with AIRN, which is consistent with our RNA-seq data (Extended Data Fig. 1a), we focused on KCNQ1OT1 lncRNA in this study.

To further validate the interaction between KCNQ1OT1 and HP1α, we performed chromatin oligonucleotide affinity precipitation (ChOP)[31] using biotinylated antisense DNA oligonucleotides

(Supplementary Table 1) to capture proteins bound by KCNQ1OT1, followed by western blotting in NIH3T3 cells. The DNA probes were designed according to a method used for chromatin isolation by RNA purification with sequencing (ChIRP–seq)[32]. A similar set of *lacZ* probes served as a negative control, which has no specific targets in human cells. We found that KCNQ1OT1 probes enriched HP1α proteins more efficiently and specifically compared with the control probes (Extended Data Fig. 1c). Given these results and that previous studies showed that KCNQ1OT1 RNA locates at nucleolar periphery and nuclear periphery regions[21,33], which are the well-known HP1 and heterochromatin attachment sites[34,35], we investigated whether KCNQ1OT1 regulates the stability of heterochromatin. To investigate this, we knocked down or silenced the expression of KCNQ1OT1 using clustered regularly interspaced palindromic repeats (CRISPR)–nuclease-inactive Cas9 (dCas9) fused to the Krüppel-associated box (KRAB) repressor[36], with single guide RNA (sgRNA) targeting the promoter region of *KCNQ1OT1* (Supplementary Table 2). As a result, *KCNQ1OT1* KD in both HEK293T and NIH3T3 cells are highly effective, close to 100% in HEK293T and 80% in NIH3T3 (Extended Data Fig. 1d). In addition, we found that H3K9me3 foci were clearly more visible in the *KCNQ1OT1*-KD cells compared with cells transfected with empty vector (EV; Fig. 1c and Extended Data Fig. 1e (left)), and the number and size of the foci increased significantly following *KCNQ1OT1* KD (Fig. 1c and Extended Data Fig. 1e (right)). To further verify the association of KCNQ1OT1 with cellular senescence, we reanalysed a published RNA-seq dataset derived from proliferating, early senescent and late senescent cells[37]. We found that KCNQ1OT1 expression decreased in early senescent cells but then increased in late senescent cells (Fig. 1d). The KCNQ1OT1 expression changes were also observed during human blood and brain tissue ageing but with opposite trends—that is, an age-related decrease in the brain and an increase in blood (Fig. 1e). Given that brain cells have a much longer lifespan than blood cells, this may imply that there are more brain cells in early senescence and more blood cells in late senescence, providing further support for KCNQ1OT1 being a dynamic senescence biomarker. These data show that KCNQ1OT1 regulates the stability of heterochromatin and is dynamically expressed during cellular senescence and ageing.

Repetitive elements are the main genomic targets of KCNQ1OT1

To reveal the DNA targets of KCNQ1OT1, we performed chromatin isolation by ChIRP–seq[32,38] in HEK293T cells using the same antisense DNA oligonucleotides that were used for the ChOP described earlier. KCNQ1OT1 RNA can be specifically enriched by *KCNQ1OT1* probes but cannot be enriched by *lacZ* probes (Fig. 2a). Two independent probe sets were designed and numbered together according to their positions along the KCNQ1OT1 RNA sequence, with an odd to even number offset. Only ChIRP–seq signals obtained by both even and odd probes were considered as true KCNQ1OT1 binding sites and used for further analysis (Methods). A total of 4,889 KCNQ1OT1 binding sites were identified by unique mapping of our ChIRP–seq. Among these, 77% of the DNA sequences (in total base pairs) targeted by KCNQ1OT1 lncRNA were repetitive elements and the most abundant repeats were Alu and L1 ($P = 1 \times 10^{-100}$ and 3.5×10^{-7} for Alu and L1, respectively; Fisher's exact test against all repeat elements), which accounted for 36% and 18% of all KCNQ1OT1-targeted repetitive elements, respectively (Fig. 2b). The DNA sequences bound by KCNQ1OT1 were primarily smaller than 4 kb (Extended Data Fig. 2a). Among the various Alu subfamilies, the evolutionarily youngest AluY elements were predominantly targeted, accounting for 36% of all targeted Alu elements (Supplementary Table 3), and ranked at the top among all transposable element subfamilies enriched by KCNQ1OT1 ChIRP–seq (Fig. 2c). Overall, across subfamilies of different evolutionary ages, the percentage of both L1 and Alu elements targeted by KCNQ1OT1 increased with a decrease in evolutionary age and plateaued from L1PA4 to L1Hs for L1 (Fig. 2d; Spearman's rank correlation coefficient to evolutionary age = −0.82 and −0.61, and $P = 9.8 \times 10^{-7}$ and 0.044 for L1 and Alu, respectively). L1HS (previously known as L1PA1) is the evolutionarily youngest and the only active L1 subfamily in the human genome that is responsible for L1 and Alu reposition in human cells. Consistent with this, L1HS is the second-most-targeted L1 subfamily by KCNQ1OT1, right after L1PA2 (Fig. 2d), suggesting that KCNQ1OT1 might be required for suppressing L1 and Alu reposition in human cells. In addition, the interactions between KCNQ1OT1 transcript and genomic DNA occur on all chromosomes rather than being restricted to neighbouring regions of the *KCNQ1OT1* gene (Extended Data Fig. 2b). Consistent with the patterns observed on total genomic sequences bound by KCNQ1OT1, over 84% of the KCNQ1OT1 ChIRP–seq peaks on each chromosome contain an overlap of at least 1 bp with a repetitive element (Extended Data Fig. 2b), making the total overlap of KCNQ1OT1 ChIRP–seq peak length to repeat elements 90.8%, which is significantly higher than random expectation (binomial distribution test, $P < 2.2 \times 10^{-16}$; empirical, $P < 0.001$; Extended Data Fig. 2c). Furthermore, the overlaps of KCNQ1OT1 chromatin immunoprecipitation with sequencing (ChIP–seq) peaks with evolutionarily young L1 and Alu subfamilies are especially enriched or highly significantly above background, whereas the overlaps with old L1 and Alu subfamilies are depleted or below background expectation (Fig. 2e), further confirming the target specificity of KCNQ1OT1 for these young transposable elements. These data indicate that KCNQ1OT1 functions not solely through a *cis*-acting mechanism, as previous studies demonstrated[18,19,24], but mainly *trans*-targets evolutionarily young repeat elements in the genome. To validate the specificity and reliability of the genomic sequences identified by ChIRP–seq, we performed ChIRP–qPCR using DNA isolated by *lacZ* or *KCNQ1OT1* probes as templates. Ten ChIRP peak regions distributed on ten chromosomes were randomly selected and qPCR primers were designed within these regions. The result showed that DNA sequences derived from ChIRP–seq data were simultaneously enriched by *KCNQ1OT1*

Fig. 2 | Genomic targets of KCNQ1OT1 lncRNA and overlap with heterochromatin marks in HEK293T cells. a, KCNQ1OT1 ssDNA probes specifically enrich KCNQ1OT1 RNA but no detectable GAPDH or U6 RNA. *LacZ* ssDNA probes do not enrich detectable KCNQ1OT1, GAPDH or U6 RNA. **b**, Percentage of non-repetitive and repetitive DNA elements targeted by KCNQ1OT1 by length (left), and the percentage of the most abundant repeats targeted by KCNQ1OT1 by count among the total 9,997 KCNQ1OT1-targeted repeats (with an overlap of at least 1 bp; right). **c**, Transposable element subfamilies enriched for KCNQ1OT1 binding regions among all transposable element subfamilies in the human genome. The bubble size indicates the Bonferroni-corrected hypergeometric test enrichment *P* value and the enrichment score are indicated by colour. **d**, Percentage of KCNQ1OT1-targeted L1 (left) and Alu (right) elements in each subfamily according to evolutionary age, from old to young. **e**, Heatmap of the percentage of KCNQ1OT1 ChIRP–seq peaks and background fragments that contain an overlap of at least 1 bp with L1 (top) and Alu (bottom) subfamilies. Background fragments were generated by 1,000 times random sampling of genomic fragments with the same length distribution as the ChIRP–seq peaks. Evolutionary age (million years) of each subfamily, $\log_2$-transformed fold change of KCNQ1OT1 percentage relative to the background and significance determined by binomial distribution test after Bonferroni correction (−log(*q*-value)) are shown on the left of the heatmaps. **f**, Validation of ten randomly selected KCNQ1OT1-targeted genomic regions on ten chromosomes by ChIRP–qPCR. Even and odd indicate two independent probe sets that are designed and numbered according to their positions along the KCNQ1OT1 or LacZ RNA sequence. **g**, Overlap of KCNQ1OT1 ChIRP–seq, H3K9me3 ChIP–seq and MeDIP–seq peaks (≥1 bp overlap). The Venn diagram was plotted using $\log_2$-transformed values of the peak numbers. **h**, H3K9me3 ChIP–seq (top) and MeDIP–seq (bottom) signals on KCNQ1OT1 target regions, and L1 and Alu elements in EV and *KCNQ1OT1*-KD HEK293T cells. DNAme, DNA methylation. **a**,**f**, Data are the mean ± s.e.m. ($n = 3$ independent experiments). Two-tailed Student's *t*-test; $^{**}P < 0.01$ and $^{***}P < 0.001$. Source numerical data are provided. RPM, reads per million.

https://doi.org/10.1038/s41556-022-01008-5

odd and even probes but they were not enriched by *lacZ* odd or even probes (Fig. 2f).

Next, we examined whether the KCNQ1OT1 lncRNA binding sequences are characterized by heterochromatin marks H3K9me3 and DNA methylation. H3K9me3 ChIP–qPCR in HEK293T cells showed that at two of the three randomly selected ChIRP peak sites (Extended Data Fig. 2d), H3K9me3 was specifically enriched relative to the negative control sites (Extended Data Fig. 2e). To improve the resolution of H3K9me3 immunoprecipitation on KCNQ1OT1 target regions, we developed a sequential ChIRP–ChIP method based on widely used ChIRP and ChIP technologies (Extended Data Fig. 2f). This analysis revealed that H3K9me3 was specifically enriched at three randomly selected KCNQ1OT1 binding sites that were isolated by *KCNQ1OT1* probes but it was not enriched at negative control sites (Extended Data

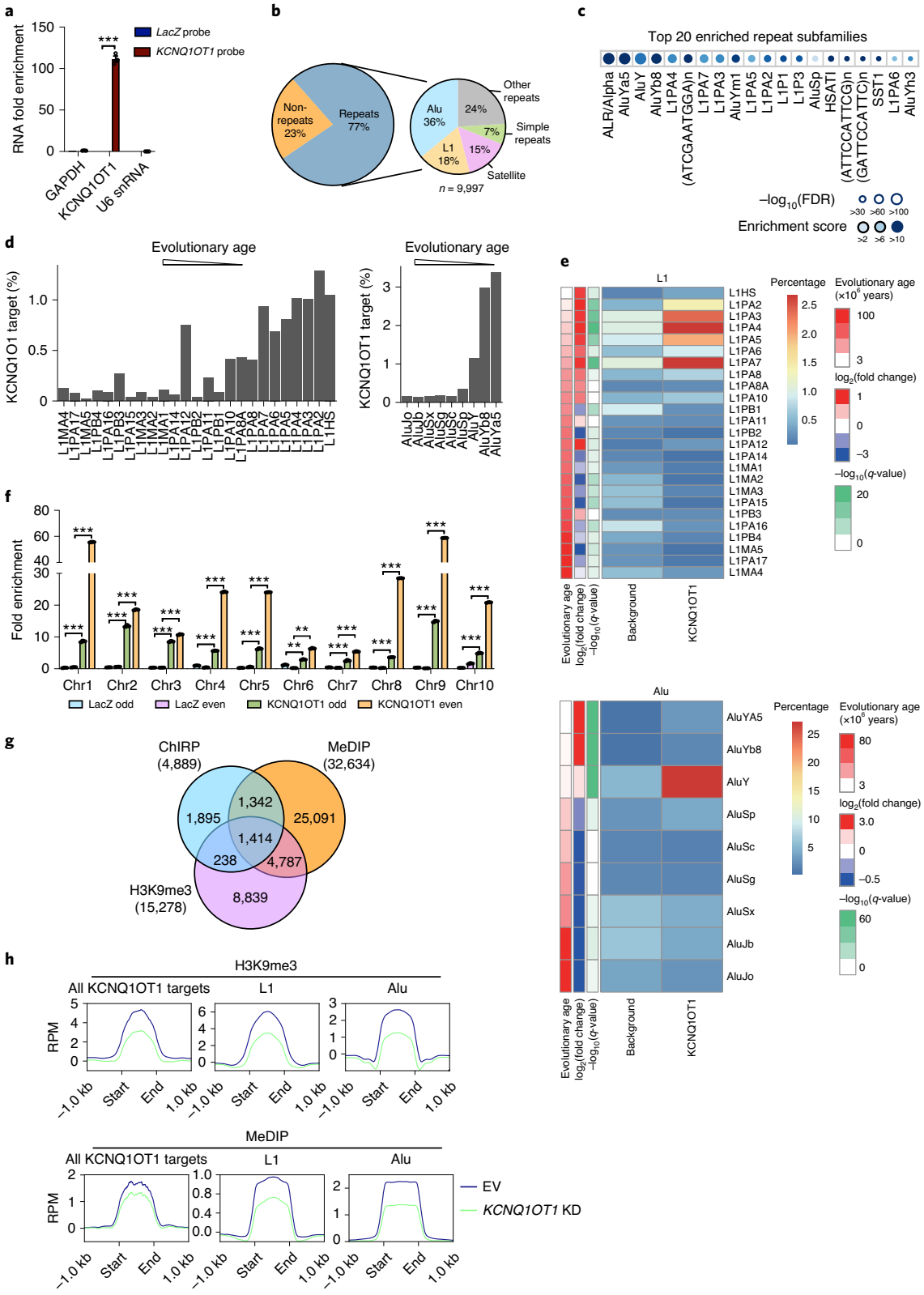


Fig. 2g). Moreover, genome-wide ChIP–seq of H3K9me3 revealed that around 34% of the KCNQ1OT1 binding sites are marked by H3K9me3 (Fisher's exact test, $P = 0.03$ using the total uniquely mapped ChIRP–seq, 5mC antibody DNA immunoprecipitation, followed by sequencing (MeDIP–seq) and H3K9me3 peaks as background; Fig. 2g). In addition, we performed MeDIP–seq in HEK293T cells and found that more than 50% of the ChIRP–seq peaks were marked by DNA methylation (Fig. 2g). All these data demonstrate that the genomic sequences targeted by KCNQ1OT1 lncRNA possess canonical heterochromatin marks.

We then examined whether KCNQ1OT1 lncRNA mediates epigenetic modifications of its target sequences. H3K9me3 ChIP–seq showed *KCNQ1OT1* KD decreased the average H3K9me3 levels on all KCNQ1OT1-targeted regions, and KCNQ1OT1-targeted L1 and Alu regions by an average of 43, 46 and 58%, respectively (Fig. 2h and Extended Data Fig. 2h). Our MeDIP–seq data revealed decreased average DNA methylation levels on all KCNQ1OT1-targeted regions, and KCNQ1OT1-targeted L1 and Alu regions by an average of 26, 41 and 39%, respectively (Fig. 2h and Extended Data Fig. 2h). These results suggest that KCNQ1OT1 lncRNA plays a critical role in maintaining the H3K9me3 modification and DNA methylation status of repetitive elements. To ensure correct targeting of the KD, we overexpressed KCNQ1OT1 through the activating CRISPRa approach[39] (Extended Data Fig. 2i) and then performed *KCNQ1OT1* KD using antisense oligonucleotides (ASOs). The results showed that the decrease in DNA methylation of L1 element (chr1:65797529–65797674) caused by *KCNQ1OT1* KD can be largely prevented by close to twofold overexpression of KCNQ1OT1 (Extended Data Fig. 2j).

Repeat-rich region of KCNQ1OT1 represses repetitive elements

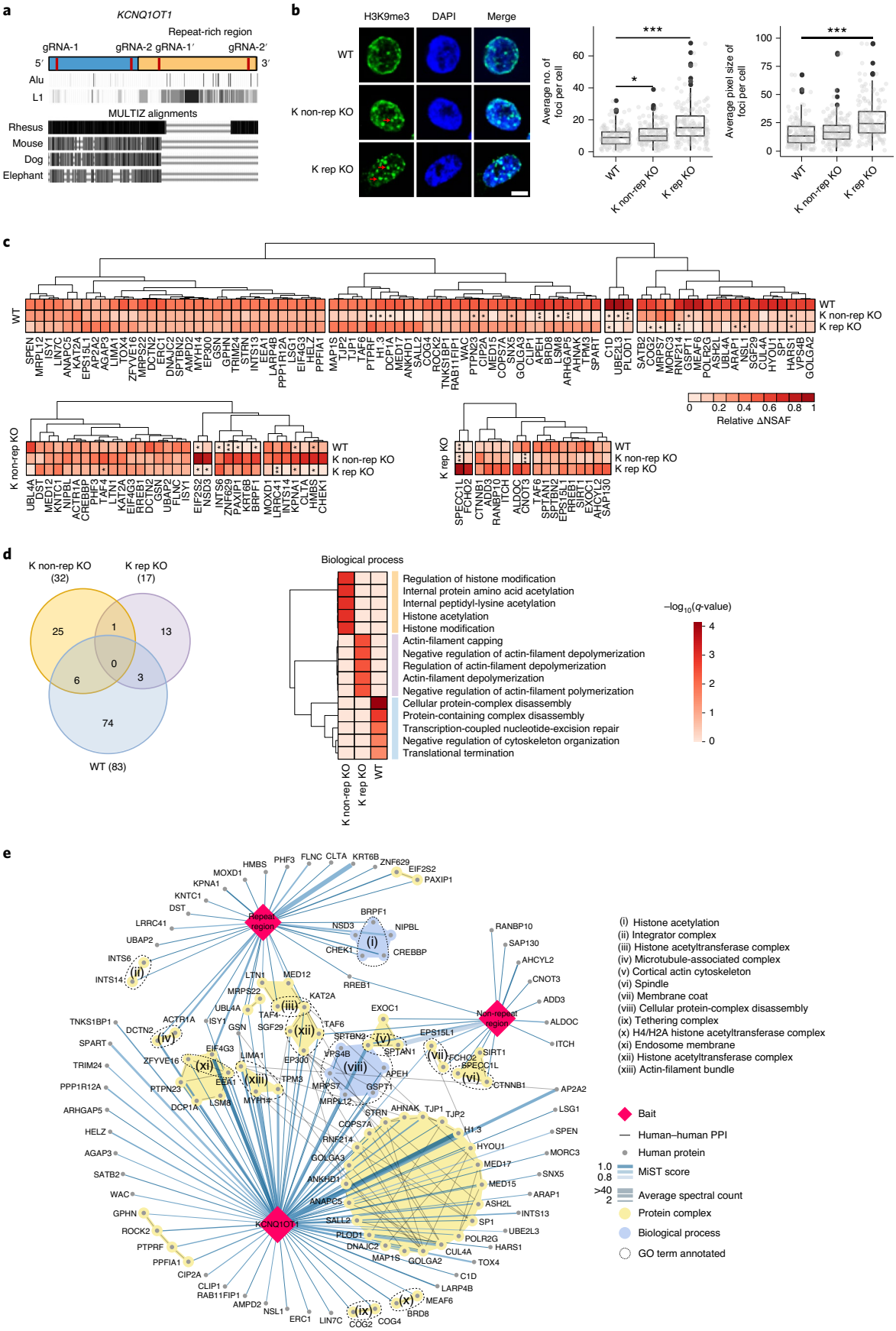
The 91 kb-long KCNQ1OT1 is characterized by a large fraction of repetitive sequences (around 50 kb in length), primarily L1 subfamilies, clustered in the 3′ half of the KCNQ1OT1 transcript (Fig. 3a). Unlike the 5′ half non-repeat-rich region, the 3′ half repeat-rich region of KCNQ1OT1 shows much lower sequence conservation among different species, which indicates its species-specific functions (Fig. 3a). To investigate the previously unexplored function of this repeat-rich region and compare with the non-repeat-rich region of KCNQ1OT1, we performed large-fragment deletions through CRISPR–Cas9 using paired gRNAs that target these regions separately (Fig. 3a and Extended Data Fig. 3a). For better comparison, we deleted comparable lengths of the repeat-rich region (approximately 37 kb, between the transcription start site at +50 kb to about 87 kb, named K rep knockout (KO)) and non-repeat-rich region (approximately 32 kb, between the transcription start site at +5 kb to about 37 kb, named K non-rep KO). Note that although KCNQ1OT1 is antisense to the coding gene *KCNQ1*, this gene is not expressed in HEK293T cells; furthermore, no change in KCNQ1 expression was observed following KO of either region or *KCNQ1OT1* KD (Extended Data Fig. 3b) and the effect of KO is therefore unlikely to be due to KCNQ1. In addition, KCNQ1 overlaps with both the repeat and non-repeat regions; the distinct or differential consequences of the KO of the two regions cannot thus be attributed to the *KCNQ1* gene. We first investigated whether heterochromatin organization was differentially changed in the two KO cell lines through immunostaining of H3K9me3. We found that heterochromatin foci were induced by K non-rep KO as well as K rep KO, with significantly more and larger foci formed in K rep KO cells compared with K non-rep KO cells (Fig. 3b). We then examined whether the K non-rep and K rep sequences bind to different sets of proteins by performing ChOP-mass spectrometry (ChOP-MS) in wild-type (WT), K non-rep KO and K rep KO HEK293T cells using *lacZ* probes as controls. High correlations were observed between replicate experiments of each sample (Extended Data Fig. 3c). We found that when the binding and non-binding are determined in a binary manner using the SAINTexpress and MiST scoring algorithms[40,41], the vast majority of WT bound proteins were absent in either KOs, with K rep KO losing more binding proteins than K non-rep KO (Fig. 3c). This led to the identification of 83 high-confidence interacting proteins for KCNQ1OT1, and 32 and 17 high-confidence interacting proteins for the repetitive region (sans the non-repeat region) in K non-rep KO cells and non-repetitive region of KCNQ1OT1 (sans the non-repeat region) in K rep KO cells, respectively (Fig. 3d and Supplementary Tables 4–6). Our analysis revealed that intact, repetitive and non-repetitive regions of KCNQ1OT1-associated proteins are enriched for different biological processes and distinct protein complex (Fig. 3d,e). Gene ontology (GO) enrichment analysis revealed that the proteins bound by the non-repetitive region alone are related to the organization of actin filament, proteins bound by the repetitive region alone are involved in histone modification and proteins bound by intact KCNQ1OT1 are involved in both nucleotide-excision repair and cytoskeleton organization (Fig. 3d). These data imply that the non-repeat-rich region is more likely to associate with the cytoskeleton/nucleoskeleton and thus attachment of heterochromatin to nucleoskeleton, whereas the repeat-rich region of KCNQ1TO1 is preferentially involved in histone modification and DNA-damage repair.

As repetitive elements are the major genomic sequences targeted by KCNQ1OT1, we compared the repression of repetitive element expression by the repetitive and non-repetitive regions of KCNQ1OT1. First, we randomly selected several Alu and L1 elements that are targeted by KCNQ1OT1 to quantify by qPCR with reverse transcription (RT–qPCR) and found that two thirds of the tested Alu and L1 elements showed significant upregulation in K rep KO cells but not in K non-rep KO cells (Extended Data Fig. 3d,e). Confirming that the increased expression of L1 elements and DNA methylation decrease in K rep KO cells were not caused by stochastic clonal epigenetic variation[42], we found similar derepression of L1 expression and DNA methylation in another K rep KO cell line (Extended Data Fig. 3f,g). Consistent with these data, using MeDIP–seq we found that the average DNA

Fig. 3 | Repeat-rich and non-repeat-rich regions of KCNQ1OT1 have different protein interactomes. a, Physical map of *KCNQ1OT1* RNA and KO regions targeted by paired gRNAs. The Alu and L1 tracks denote the locations of Alu and L1 elements on KCNQ1OT1. The tracks of MULTIZ alignments show the conservation of *KCNQ1OT1* across different species. **b**, Immunostaining of H3K9me3 in WT, K non-rep KO and K rep KO HEK293T cells (left). Nuclei were counterstained with DAPI. Scale bar, 5 µm. Average number of foci (middle) and average area of the foci in each cell (right) in the images on the left (exemplary foci are indicated with red arrows). Box-and-whisker plots: the bound of a box shows the 25% (lower quartile) and 75% (upper quartile) of all values in the group; the line in the box shows the median value; and values outside the middle 50% values (IQR) but within 1.5×IQR are shown as whiskers, otherwise as black dots. All values are shown as grey spots; $n = 155$ cells in each condition pooled across three independent experiments. **c**, The relative ΔNSAF of high-confidence binding proteins to full length, repeat region and non-repeat region of KCNQ1OT1 identified in WT, K non-rep KO and K rep KO cells (Methods). The preys of each sample are grouped into four clusters by hierarchical clustering. **d**, Venn diagram showing high-confidence KCNQ1OT1-associated proteins detected by ChOP-MS in WT, K non-rep KO and K rep KO HEK293T cells (left). The size of the circles is proportional to the $\log_2$-transformed values of the corresponding protein numbers. GO biological processes enriched among the high-confidence binding proteins to full-length, repeat region and non-repeat region of KCNQ1OT1 identified in WT, K non-rep KO and K rep KO cells (right). The top five most significant terms for each KCNQ1OT1 variant are shown. **e**, High-confidence interactions between the bait KCNQ1OT1 variants (red diamonds) and prey proteins (circles). Yellow shades indicate protein complexes and biological process are shaded in blue. Thin black lines denote physical interactions among proteins. Blue edges link prey proteins to their baits. Dashed circles show the GO annotation. The edge colour is proportional to the MiST score and the edge thickness is proportional to the spectral counts (Methods); $n = 3$ independent biological replicates. **b**,**c**, Statistical significance was determined using a two-tailed Student's *t*-test; $^{*}P < 0.05$, $^{**}P < 0.01$ and $^{***}P < 0.001$. Source numerical data are provided.

https://doi.org/10.1038/s41556-022-01008-5

methylation levels on all KCNQ1OT1 target sites decreased in both K rep and non-rep KO cells, with a larger decrease in the K rep KO cells (Fig. 4a), which was similarly observed at L1 and Alu sequences targeted by KCNQ1OT1 (Fig. 4b–d). This phenomenon might be partially attributed to the preferential DNMT binding to K-rep sequences compared with K non-rep sequences (Extended Data Fig. 3h).



https://doi.org/10.1038/s41556-022-01008-5

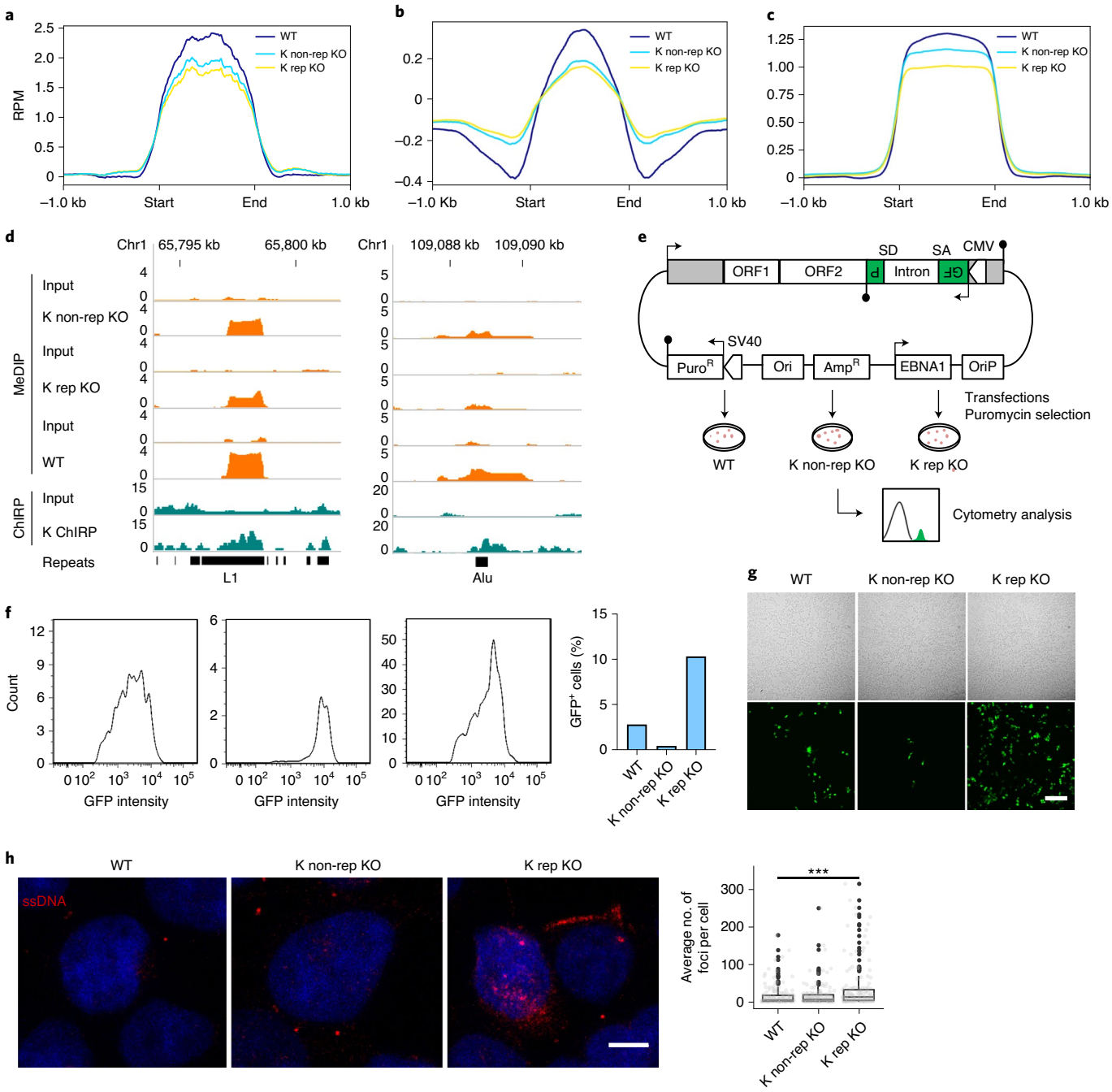


Fig. 4 | Decreased DNA methylation and increased L1 transposition in K rep KO HEK293T cells. a, MeDIP–seq signal at KCNQ1OT1 ChIRP–seq target regions in WT, K non-rep KO and K rep KO HEK293T cells. **b**, MeDIP–seq signal at KCNQ1OT1 target L1 elements in WT, K non-rep KO and K rep KO HEK293T cells. **c**, MeDIP–seq signal at KCNQ1OT1 target Alu elements in WT, K non-rep KO and K rep KO HEK293T cells. **d**, Genome-browser tracks of MeDIP–seq and ChIRP–seq alignments at randomly selected L1 and Alu repeat regions. **e**, Schematic map of the retrotransposition reporter construct and the assay for retrotransposition. Puromycin-resistant cells containing the reporter constructs were analysed by flow cytometry for retrotransposition-activated GFP signal. Puro^R, puromycin resistance; Amp^R, ampicillin resistance. **f**, Flow cytometry counts of GFP-tagged retrotransposition reporter-positive cells normalized to untransfected HEK293T cells (left); 50,000 cells were sorted per sample. Percentage of GFP$^+$ HEK293T cells in the WT, K non-rep KO and K rep KO groups (right). The three graphs are the density plot of flowcytometry count of cells with indicated GPF intensity (*x* axis) for WT, K non-rep KO and K-rep KO from left to right. **g**, Cells transfected with the GFP-tagged retrotransposition reporter construct viewed under a bright-field (top) and fluorescence (bottom) microscope. Scale bar, 100 µm. **f**,**g**, Cells were selected with puromycin for two weeks. Transfection was performed three times and similar results were obtained, as shown by the number and sizes of GFP foci per field in three biological replicates. **h**, Immunostaining of ssDNA in WT, K non-rep KO and K rep KO cells (left). Nuclei were counterstained with DAPI. Scale bar, 5 µm. Number of ssDNA puncta observed in the immunofluorescence images (right); $n = 213$ (WT), 289 (K non-rep KO) and 191 (K rep KO) cells pooled from three independent experiments. Box-and-whisker plots: the bound of a box shows the 25% (lower quartile) and 75% (upper quartile) of all values in the group; the line in the box shows the median value; and values outside the middle 50% values (IQR) but within 1.5×IQR are shown as whiskers, otherwise as black dots. All values are shown as grey spots. Statistical significance was determined using a two-tailed Student's *t*-test; ***$P < 0.001$. Source numerical data are provided.

https://doi.org/10.1038/s41556-022-01008-5

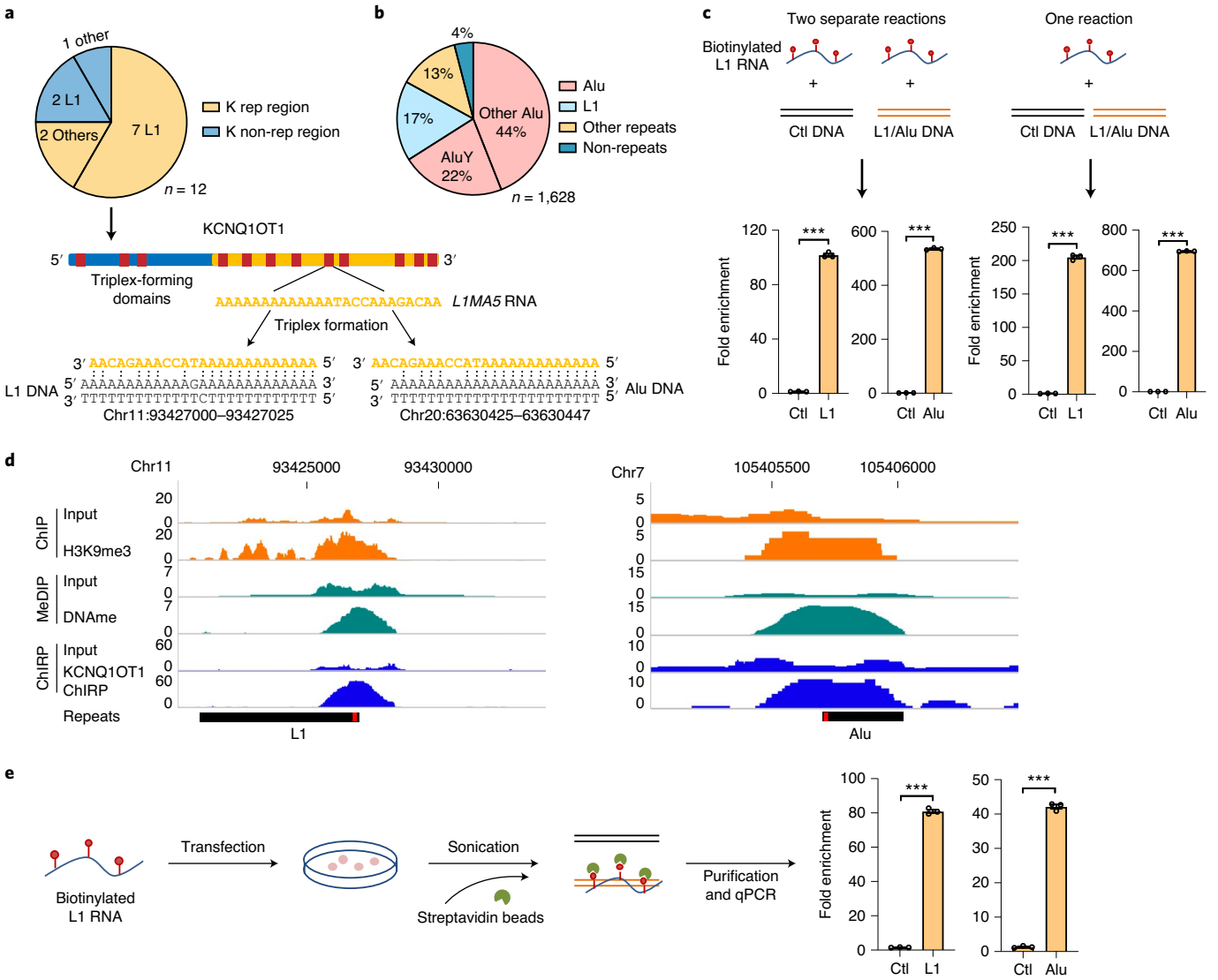


Fig. 5 | KCNQ1OT1 target genomic sequences are guided by RNA–dsDNA triplex. a, Distribution and repeat types of predicted triplex-forming sites on KCNQ1OT1 (top). Locations of predicted triplex-forming sites on KCNQ1OT1 (red blocks; middle) and an example of a predicted triplex formed between L1MA5 RNA derived from KCNQ1OT1 and L1/Alu DNA via Hoogsteen base pairing (denoted by two dots; bottom). **b**, Percentage of predicted triplex-forming KCNQ1OT1 ChIRP–seq repeat and non-repeat targets in HEK293T cells. **c**, In vitro triplex pulldown performed by incubating biotinylated L1 RNA with control (Ctl) DNA and Alu DNA separately (left) or together (right). The RNA-associated DNA was analysed by qPCR and normalized to the input DNA. **d**, Genome browser views showing the H3K9me3 and DNA methylation status on the KCNQ1OT1-targeted L1 and Alu elements and the predicted KCNQ1OT1 triplex (indicated with a red line) within these elements. DNAme, DNA methylation. **e**, Schematic of the in vivo triplex pulldown by transfecting biotinylated L1 and Alu RNA into HEK293T cells at the sites indicated in **d** (left). The RNA-associated DNA was analysed by qPCR and normalized to the input DNA (right). The qPCR primers used amplify unique sequences within the repetitive elements. **c**,**e**, Data are the mean ± s.e.m. ($n = 3$ independent experiments). Statistical significance was determined using a two-tailed Student's t-test; ***$P < 0.001$. Source numerical data are provided.

The repeat-rich region of KCNQ1OT1 represses retrotransposon transposition

Previous studies reported that increased activity of Alu and L1 is associated with cellular senescence[37,43] and the successful retrotransposition of Alu elements depends on the L1-provided open reading frame 2 (ORF2) protein[44]. We thus examined whether transposon derepression in K rep and K rep KO cells induces transposition activity and genome instability using an L1 transposition reporter system where transposition events are marked as GFP foci (Fig. 4e). After selection with puromycin, the cells were subjected to flow cytometry analysis. The proportion of GFP$^+$ K rep KO cells was around 10%, much higher than GFP$^+$ WT cells (approximately 3%) but there were significantly fewer GFP$^+$ K non-rep KO cells (approximately 0.5%; Fig. 4f), which is consistent with the microscopy observations (Fig. 4g). These suggest that the low increase in L1 expression in the K non-rep KO cells is associated with L1 transposition suppression and high increase in L1 expression in the K rep KO cells is associated with L1 transposition enhancement. As L1 transposition is known to lead to ssDNA accumulation in the cytoplasm[37], we also stained cells with an antibody specifically targeting ssDNA, as described previously[45], and found that only K rep KO cells (and not K non-rep KO cells) displayed a significantly higher level of ssDNA than WT cells (Fig. 4h). These data revealed that the repetitive region of KCNQ1OT1 is essential for repressing L1 transposition activity, which when activated is known to cause DNA damage, genome instability, chromatin-structure dysfunction and ultimately cellular senescence[46].

KCNQ1OT1 targets genomic sequences by forming an RNA–dsDNA triplex

These results led us to investigate whether the stronger effects of evolutionarily non-conserved repeat-rich region compared with the conserved non-repeat-rich region, in addition to its larger protein interactome, could also be attributed to an upstream prerequisite function of the K rep region to guide KCNQ1OT1 to its target sites, whereas K non-rep is a downstream partial effector moiety. To explore how KCNQ1OT1 recognizes its target sites, we examined whether KCNQ1OT1 can form a triplex with target double-stranded genomic DNA via Hoogsteen base pairing[47], which is a mechanism found in gene regulation mediated by a few lncRNAs[48–50]. We applied a computational method to predict RNA–dsDNA triplexes[47] and identified 12 triplex-forming domains on KCNQ1OT1—nine were L1 sequences and seven of these were located at the repeat-rich region of KCNQ1OT1 (Fig. 5a)—and 1,628 KCNQ1OT1 ChIRP–seq binding sites were predicted to form triplexes with KCNQ1OT1 (Fig. 5b). Notably, AluY and L1 are the top two repeat subfamilies possessing triplex-forming potential with *KCNQ1OT1* (Fig. 5b). The direct RNA–DNA interactions mainly occur between A-rich sequences through Hoogsteen base pairing (Fig. 5a (bottom) for examples). Based on this analysis, we then validated the triplex-forming ability of KCNQ1OT1 with target sites using in vitro and in vivo triplex pulldown assays. We observed that the L1 RNA sequence derived from KCNQ1OT1 can specifically form a triplex with L1 or Alu DNA both in vitro and in vivo (Fig. 5c–e). The specific interaction between L1 RNA and L1 DNA formed between poly(A) and poly(dA) was abolished by replacing the poly(A) in L1 RNA with poly(G) in a mutant L1 RNA (Extended Data Fig. 4a and Supplementary Table 8). Consistent with KCNQ1OT1's preferential targeting of the evolutionarily young L1 and Alu subfamilies (Fig. 2d), we found that L1 and Alu subfamilies that are evolutionary younger contained significantly more predicted triplexes, or triplex target sites, with KCNQ1OT1 than evolutionarily older subfamilies (Fig. 6a (top); Spearman's rank correlation coefficient to evolutionary age = −0.81 and −0.71, and $P = 1.9 \times 10^{-6}$ and 0.031 for L1 and Alu, respectively). With a decrease in evolutionary age, the primary distribution peak of predicted number of triplexes on either L1 or Alu gradually shifted from zero to one triplex; for L1, a secondary peak also gradually appeared at around five and shifted towards nine and ten triplexes. Consistent with the highly enhanced L1 retrotransposition we observed in K rep KO cells (Fig. 4e–h), L1HS elements—the only active L1 subfamily in human cells—had the highest number of KCNQ1OT1 triplexes, with the most prominent secondary distribution peak sharply centred at ten *KCNQ1OT1* triplexes, providing the suggestion of the highest affinity of KCNQ1OT1 to target this active L1 subfamily (Fig. 6a (bottom)). Furthermore, we found the number of triplexes predicted on L1 and Alu elements follows a cyclic pattern (Fig. 6b), perhaps due to the structural and sequencing constraints of Hoogsteen base pairing[51]. Also in line with the induction of L1 transposition by K rep KO but not K non-rep KO (Fig. 4e–h), the repeat-rich and non-repeat-rich regions preferentially formed triplexes with different repeat subfamilies (Extended Data Fig. 4b,c). In particular, among all of the L1 subfamilies, L1HS was the most preferentially targeted by triplexes in the K rep region; similarly, the youngest Alu subfamily was the most preferentially targeted by triplexes in the K rep region (Fig. 6c).

Next, we investigated whether the strength of the interactions between KCNQ1OT1 and its target, reflected by the predicted number of triplexes, has any effect on the ability of repetitive elements being targeted by KCNQ1OT1 and epigenetically silenced. To address this question, we first analysed how many of the L1 and Alu elements identified by ChIRP–seq are characterized by triplex-forming sequences, DNA methylation and H3K9me3. The result showed that without a triplex-forming sequence, H3K9me3 or DNA methylation, only 0.05% L1 and 0.04% Alu were bound by KCNQ1OT1; however, triplex formation remarkably increased the odds of KCNQ1OT1 binding when the L1 and Alu were concurrently marked by DNA methylation and H3K9me3—that is, 36% for L1 and 17% for Alu (Fig. 6d,e). KCNQ1OT1 binding seems to be more associated with DNA methylation than H3K9me3. Most importantly, KCNQ1OT1 binding progressively increased with the number of triplexes formed, peaking at four triplexes, with 100% L1 and 50% Alu elements bound by KCNQ1OT1. Interestingly, the triplex-forming region on Alu elements seem to evict or mobilize one nucleosome in the H3K9me3 profile (Extended Data Fig. 4d).

To further infer potential causal relationships among the two chromatin marks and triplex formation as well as downstream expression changes, we computed a Bayesian network among the four events based on binary yes or no signals of L1 and Alu separately using an approach we developed earlier[52] (Methods). As KCNQ1OT1 ChIRP–seq peaks are far from saturation (Extended Data Fig. 4e), perhaps due to limited probes, we did not include them in this analysis. Consistent with forming an 'AND' gate requiring both triplex-guided targeting and H3K9me3 modification for tight repression of the transposons by DNA methylation, on both L1 and Alu, triplex formation and H3K9me3 are the most upstream events, together leading to DNA methylation, which in turn leads to transcription repression (Fig. 6f). However, due to the limitations of the acyclic structure of the Bayesian network, there could be feedback mutual stabilizing interactions missed in the network; thus, the conflicting causal directions between triplex formation and H3K9me3 on L1 and Alu could be due to a mutual feedback stabilization.

To investigate whether triplex formation is sufficient in guiding KCNQ1OT1 to a specific target site to enable subsequent DNA methylation, we designed an artificial RNA sequence that is predicted to form a triplex with a genomic sequence marked with H3K9me3 but without DNA methylation and a scrambled RNA sequence that is unable to form a triplex with any genomic sequence (Fig. 6g), and investigated whether this engineered guiding sequence can lead to de novo DNA methylation at the ectopic site. Using CRISPR knock-in, we inserted the DNA corresponding to the triplex-forming or scrambled RNA sequence into the non-repeat-rich region (to avoid disrupting the potential DNMT binding site in the repeat-rich region; Extended Data Fig. 3h) of *KCNQ1OT1* (Fig. 6g). The MeDIP–qPCR data demonstrated that the DNA methylation levels at the region (qPCR region 2) recognized by our engineered RNA sequence and a nearby region (qPCR region 1) are increased in cells with the engineered triplex-forming sequence on *KCNQ1OT1* but not in cells containing the scrambled sequence on *KCNQ1OT1* (Fig. 6g,h). These data demonstrate that triplex-forming sequences contained in KCNQ1OT1 are sufficient to guide KCNQ1OT1 and necessary for ensuring maximal DNA methylation, and also make the KCNQ1OT1 a mammalian endogenous Cas9-like epigenetic repression system amenable to new guiding RNA design.

KCNQ1OT1 deficiency results in spontaneous cellular senescence

To investigate the impact of KCNQ1OT1 deficiency on untransformed cells, we induced premature senescence in WI38 cells (a cell line commonly used for cellular senescence studies) with H_2O_2 and found that the expression of KCNQ1OT1 was dynamic with 48 h of H_2O_2 treatment—it decreased at 24 h and then increased at 48 h (Fig. 7a,b). However, genes associated with cell-cycle arrest—*CDKN2A* (also known as *P16*) and *CDKN1A* (also known as *P21*)—SASP (*IL6*) and L1 elements were consistently upregulated (Fig. 7a,b). In addition, we performed *KCNQ1OT1* KD in WI38 cells using ASOs. Gene expression analysis by RT–qPCR demonstrated that KCNQ1OT1 decreased by 60% at 24 h after transfection, genes including *CDKN1A*, SASP (*IL6*) and type-I interferon response (*IFNB1* and *IFNK*) were significantly increased, whereas *CDKN2A* expression remained unchanged and L1 elements were not consistently activated (Fig. 7c). This gene expression pattern changed at 48 h after transfection, with upregulation of *KCNQ1OT1* and *CDKN2A*, decrease in *CDKN1A*, and more activated L1 elements (Fig. 7d). These expression changes observed in senescent WI38 cells resemble those observed in IMR90 cells undergoing natural early to late senescence[37]. In addition, we found that the activity of the late senescence marker

https://doi.org/10.1038/s41556-022-01008-5

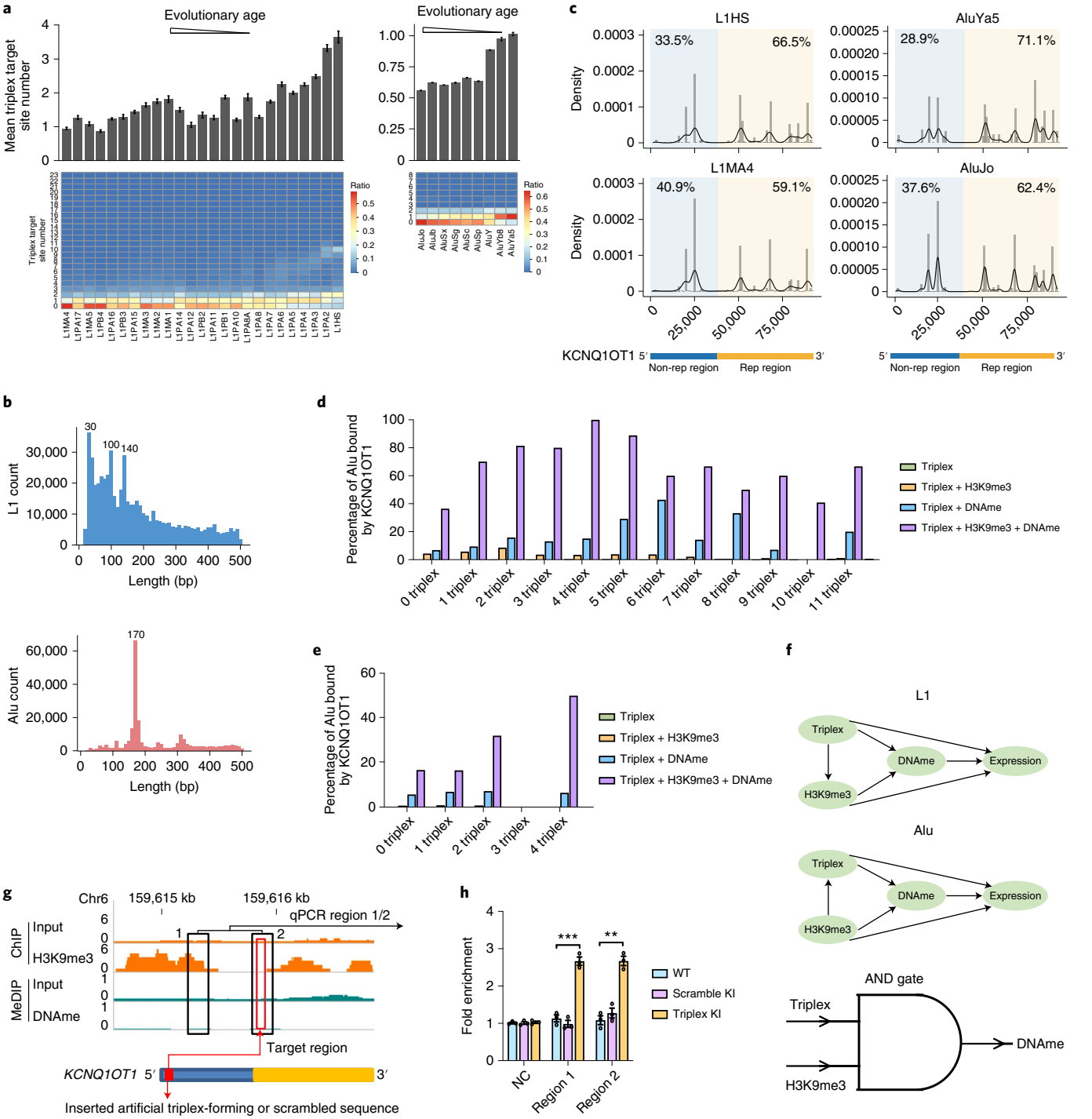


Fig. 6 | The RNA–dsDNA triplex contributes to KCNQ1OT1 binding to genomic DNA. a, Average number of predicted KCNQ1OT1 triplex target sites on L1 (left) and Alu (right) elements in different subfamilies according to evolutionary age (from old to young; top).The heatmaps show the proportion of elements predicted to form *n* number of triplex target sites, with *n* ranging from 0 to 23 for L1, and 0 to 8 for Alu (bottom). **b**, Distribution of the distance between predicted triplex target sites on L1 (top) and Alu (bottom) elements; zero marks the centre of a predicted triplex target site. Each triplex site is counted once as the centre, the count of all other triplex sites are shown. **c**, Probability density and histogram of the number of triplexes with the youngest and oldest L1 and Alu subfamilies in each 100-bp (density) and 1-kb (histogram) interval along the *KCNQ1OT1* genes. **d**,**e**, Percentage of L1 (**d**) and Alu (**e**) that contain different numbers of predicted triplex-forming sequences with or without DNA methylation and/or H3K9me3 detected as KCNQ1OT1 targets, determined through ChIRP–seq in HEK293T cells. **f**, Bayesian network inferred causal relationships among triplex, H3K9me3, DNA methylation and gene expression on L1 (top) and Alu (middle). Schematic model showing the AND gate formed between triplex formation and H3K9me3 to consequent DNA methylation (bottom). **g**, Genome browser view showing the H3K9me3 modification and DNA methylation states of the region targeted by the inserted sequences in KCNQ1OT1 (top). The red box denotes the triplex target region; the two black boxes denote two regions detected by MeDIP–qPCR in **h**. Schematic diagram indicating the knock-in site of artificial scrambled or triplex-forming sequence on *KCNQ1OT1* (bottom). **h**, MeDIP–qPCR performed in WT, scramble knock-in and triplex knock-in cells, followed by qPCR analysis. Data are the fold enrichment relative to the input. NC, negative control region that is not marked by DNA methylation in WT cells; KI, knock-in. **a**,**h**, Data are the mean ± s.e.m. ($n = 3$ independent experiments). Statistical significance was determined using a two-tailed Student's *t*-test; $^{**}P < 0.01$ and $^{***}P < 0.001$. DNAme, DNA methylation. The total number of repeat elements, mean ± s.e.m of the predicted KCNQ1OT1 triplex target sites in each L1 and Alu, and other source numerical data are provided.

https://doi.org/10.1038/s41556-022-01008-5

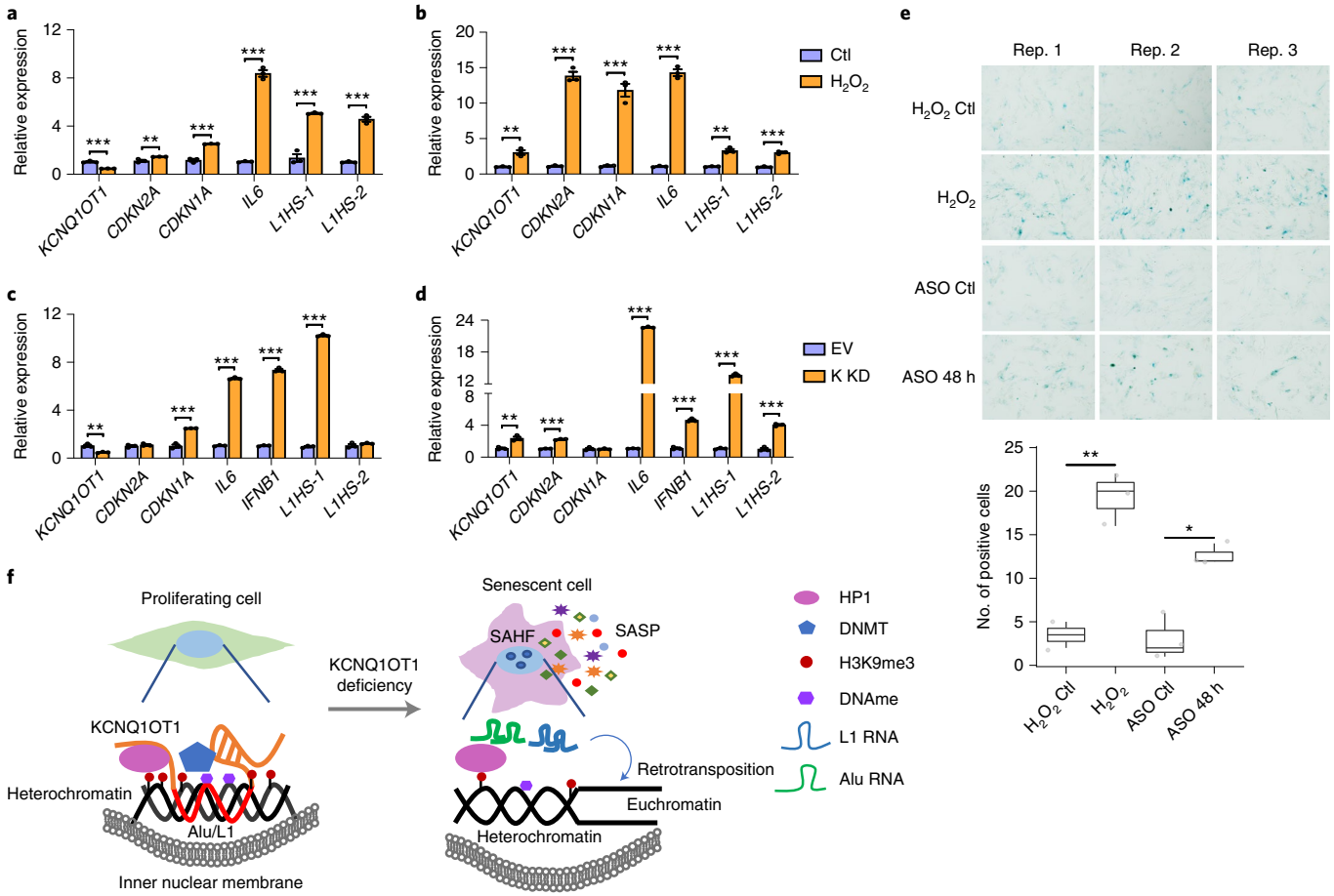


Fig. 7 | KCNQ1OT1 repression induces cellular senescence. a,**b**, Senescence-related gene expression detected in WI38 cells treated with 400 μM H_2O_2 for 24 (**a**) or 48 h (**b**). This expression pattern was consistently observed in three repeated experiments. **c**,**d**, Gene expression, determined by RT–qPCR, in WI38 cells transfected with ASO at 24 (**c**) and 48 h (**d**) after transfection. **e**, SA-β-gal staining 48 h after WI38 cells transfected with or without ASOs to KCNQ1OT, or after induction with or without 400 μM H_2O_2 (top). Representative images from three replicates (Rep. 1–3) are shown. Number of cells staining positive for SA-β-gal (bottom). Box-and-whisker plots: the bound of a box shows the 25% (lower quartile) and 75% (upper quartile) of all values in the group; the line in the box shows the median value; and values outside the middle 50% values (IQR) but within 1.5×IQR are shown as whiskers, otherwise as black dots. All values are shown as grey spots. **f**, Schematic model showing cellular senescence-associated events caused by KCNQ1OT deficiency, including heterochromatin decompaction and detachment from the nuclear membrane, loss of H3K9me3 and DNA methylation (DNAme), activation of retrotransposable elements and formation of SAHF and SASP. **a**–**d**, Data are the mean ± s.e.m. ($n = 3$ independent experiments) relative expression normalized to *β-actin* messenger RNA. **a**–**d**,**e**, Statistical significance was determined using a two-tailed Student's *t*-test; $^{*}P < 0.05$, $^{**}P < 0.01$ and $^{***}P < 0.001$. Ctl, control. Source numerical data are provided.

senescence-associated β-galactosidase (SA-β-gal) was also significantly increased in WI38 cells at 48 h after transfection with KCNQ1OT1 ASOs to half the level of H_2O_2-induced senescence (Fig. 7e). These data indicate that a reduction in KCNQ1OT1 not only induced early senescence but also deep senescence and that the upregulation of KCNQ1OT1 in cells in late senescence is insufficient to reverse the course of cellular senescence. Together, our data suggest that a deficiency in KCNQ1OT1 lncRNA induces spontaneous cellular senescence accompanied by disorganization of heterochromatin, inefficiency of histone modification and DNA methylation, activation of repetitive elements and SASPs (Fig. 7f).

Discussion

Together, we uncovered (1) a causal and consequential link between the 'weak' heterochromatin mark H3K9me3 and the 'strong' heterochromatin mark DNA methylation mediated by KCNQ1OT1 lncRNA, (2) a mammalian guiding RNA mechanism to achieve sequence specificity in transposon silencing and heterochromatin stabilization, (3) an AND gate between H3K9me3/HP1 mark and target DNA sequence specificity for deep silencing of transposons by DNA methylation and (4) a critical function of lncRNA guarding against genome instability, heterochromatin foci formation and cellular senescence.

In contrast to the previously established *cis* targeting nearby coding gene silencing function of KCNQ1OT1, in this study we found that the repeat-rich region of KCNQ1OT1 contains the guiding RNA sequence by forming RNA–dsDNA triplex base pairing to *trans*-target genomic sequences of transposons for deep repression. Deletion of the species-specific repeat-rich region of KCNQ1OT1 significantly enhanced L1 transposition events, whereas deletion of only the evolutionarily conserved non-repeat-rich region of KCNQ1OT1 resulted in lower-than-WT L1 transposition events, suggesting a regulatory function of the non-repetitive region that can be modulated to fine-tune transposition frequency.

Interestingly, not only the guiding sequences are encoded in repeat elements within KCNQ1OT1, the target affinity seems to also be a function of the number of RNA–dsDNA triplex formed, with the four triplexes having the highest targeting affinity for both L1 and Alu elements. We further demonstrated the necessity and sufficiency

of the RNA–dsDNA triplex in target guiding by ectopically targeting and DNA methylating an H3K9me3 region using an engineered triplex sequence; this makes it the first dCas9-like RNA originated from a mammalian source, including human cells, and hence a valuable endogenous epigenome-modifying arsenal.

Intriguingly, our data and reanalysis of previously published data[37] reveal that the expression pattern of KCNQ1OT1 is dynamic during the process of cellular senescence, decreases at early senescence and increases at late senescence in both human and mouse cells, and that a decrease in or lack of KCNQ1OT1 alone is sufficient to induce early cellular senescence, making it a common senescence marker and an essential and evolutionarily conserved gatekeeper against genome instability, as evidenced by frequent L1 retrotransposition and cellular senescence across many cell types (Fig. 6h). Together, our study revealed a new paradigm and mechanism of lncRNA functions to guard against genome instability and senescence through epigenetic modification-sensitive and sequence-specific tight DNA methylation repression of transposon elements. The epigenetic modification-sensitivity and sequence specificity are achieved through HP1 binding and RNA–dsDNA triplex pairing, respectively, forming a molecular logical AND gate on the RNA together with the effector moieties introducing DNA methylation forming an autonomous RNA-based regulation circuitry. The structure of KCNQ1OT1 represents a novel endogenous dCas9-like repression system in mammals, more permanent than dCas9, which induces more transient H3K9me3. In terms of technology, the modularized target specificity, which intrinsically formed by the guiding sequences in the clustered repetitive sequences in KCNQ1OT1, can be engineered for ectopic re-targeting, making it a useful epigenome-editing tool. In a biological context, the highly conserved and essential role of KCNQ1OT1 in maintaining genome stability and cell vitality manifested by the spontaneous senescence following its loss have foreseeable strong implications in ageing and senescence interventions and rejuvenation.

Online content

Any methods, additional references, Nature Research 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/s41556-022-01008-5.

References

1. De Cecco, M. et al. Genomes of replicatively senescent cells undergo global epigenetic changes leading to gene silencing and activation of transposable elements. *Aging Cell* **12**, 247–256 (2013).
2. Muotri, A. R. et al. Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature* **435**, 903–910 (2005).
3. Becker, J. S., Nicetto, D. & Zaret, K. S. H3K9me3-dependent heterochromatin: barrier to cell fate changes. *Trends Genet.* **32**, 29–41 (2016).
4. Zhang, R. & Adams, P. D. Heterochromatin and its relationship to cell senescence and cancer therapy. *Cell Cycle* **6**, 784–789 (2007).
5. Chandra, T. et al. Global reorganization of the nuclear landscape in senescent cells. *Cell Rep.* **10**, 471–483 (2015).
6. Kane, A. E. & Sinclair, D. A. Epigenetic changes during aging and their reprogramming potential. *Crit. Rev. Biochem. Mol. Biol.* **54**, 61–83 (2019).
7. Johnson, W. L. & Straight, A. F. RNA-mediated regulation of heterochromatin. *Curr. Opin. Cell Biol.* **46**, 102–109 (2017).
8. Chen, C. K. et al. Xist recruits the X chromosome to the nuclear lamina to enable chromosome-wide silencing. *Science* **354**, 468–472 (2016).
9. Deng, Z., Norseen, J., Wiedmer, A., Riethman, H. & Lieberman, P. M. TERRA RNA binding to TRF2 facilitates heterochromatin formation and ORC recruitment at telomeres. *Mol. Cell* **35**, 403–413 (2009).
10. Hu, W. L. et al. GUARDIN is a p53-responsive long non-coding RNA that is essential for genomic stability. *Nat. Cell Biol.* **20**, 492–502 (2018).
11. Abdelmohsen, K. & Gorospe, M. Noncoding RNA control of cellular senescence. *Wiley Interdiscip. Rev. RNA* **6**, 615–629 (2015).
12. Ghanam, A. R. et al. Shining the light on senescence associated lncRNAs. *Aging Dis.* **8**, 149–161 (2017).
13. Grammatikakis, I., Panda, A. C., Abdelmohsen, K. & Gorospe, M. Long noncoding RNAs (lncRNAs) and the molecular hallmarks of aging. *Aging* **6**, 992–1009 (2014).
14. Cai, D. & Han, J.-D.J. Aging-associated lncRNAs are evolutionarily conserved and participate in NFκB signaling. *Nat. Aging* **1**, 438–453 (2021).
15. Su, M., Han, D., Boyd-Kirkup, J., Yu, X. & Han, J. J. Evolution of Alu elements toward enhancers. *Cell Rep.* **7**, 376–385 (2014).
16. Smilinich, N. J. et al. A maternally methylated CpG island in KvLQT1 is associated with an antisense paternal transcript and loss of imprinting in Beckwith–Wiedemann syndrome. *Proc. Natl Acad. Sci. USA* **96**, 8064–8069 (1999).
17. Mancini-Dinardo, D., Steele, S. J., Levorse, J. M., Ingram, R. S. & Tilghman, S. M. Elongation of the Kcnq1ot1 transcript is required for genomic imprinting of neighboring genes. *Genes Dev.* **20**, 1268–1282 (2006).
18. Kanduri, C. Kcnq1ot1: a chromatin regulatory RNA. *Semin. Cell Dev. Biol.* **22**, 343–350 (2011).
19. Pandey, R. R. et al. Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol. Cell* **32**, 232–246 (2008).
20. Paulsen, M., Khare, T., Burgard, C., Tierling, S. & Walter, J. Evolution of the Beckwith–Wiedemann syndrome region in vertebrates. *Genome Res.* **15**, 146–153 (2005).
21. Mohammad, F. et al. Kcnq1ot1/Lit1 noncoding RNA mediates transcriptional silencing by targeting to the perinucleolar region. *Mol. Cell. Biol.* **28**, 3713–3728 (2008).
22. Fitzpatrick, G. V., Soloway, P. D. & Higgins, M. J. Regional loss of imprinting and growth deficiency in mice with a targeted deletion of KvDMR1. *Nat. Genet.* **32**, 426–431 (2002).
23. Horike, S. et al. Targeted disruption of the human LIT1 locus defines a putative imprinting control element playing an essential role in Beckwith–Wiedemann syndrome. *Hum. Mol. Genet.* **9**, 2075–2083 (2000).
24. Mohammad, F., Mondal, T., Guseva, N., Pandey, G. K. & Kanduri, C. Kcnq1ot1 noncoding RNA mediates transcriptional gene silencing by interacting with Dnmt1. *Development* **137**, 2493–2499 (2010).
25. Elcheva, I. A. & Spiegelman, V. S. The role of *cis*- and *trans*-acting RNA regulatory elements in leukemia. *Cancers* **12**, 3854 (2020).
26. Alekseyenko, A. A. et al. Heterochromatin-associated interactions of *Drosophila* HP1a with dADD1, HIPP1, and repetitive RNAs. *Genes Dev.* **28**, 1445–1460 (2014).
27. Muchardt, C. et al. Coordinated methyl and RNA binding is required for heterochromatin localization of mammalian HP1α. *EMBO Rep.* **3**, 975–981 (2002).
28. Bannister, A. J. et al. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410**, 120–124 (2001).
29. Van Nostrand, E. L. et al. Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat. Methods* **13**, 508–514 (2016).
30. Nagano, T. et al. The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* **322**, 1717–1720 (2008).
31. Frank, S. et al. yylncT defines a class of divergently transcribed lncRNAs and safeguards the T-mediated mesodermal commitment of human PSCs. *Cell Stem Cell* **24**, 318–327 (2019).

32. Chu, C., Qu, K., Zhong, F. L., Artandi, S. E. & Chang, H. Y. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Mol. Cell* **44**, 667–678 (2011).
33. Sachani, S. S. et al. Nucleoporin 107, 62 and 153 mediate Kcnq1ot1 imprinted domain regulation in extraembryonic endoderm stem cells. *Nat. Commun.* **9**, 2795 (2018).
34. Bizhanova, A. & Kaufman, P. D. Close to the edge: heterochromatin at the nucleolar and nuclear peripheries. *Biochim. Biophys. Acta Gene Regul. Mech.* **1864**, 194666 (2021).
35. Padeken, J. & Heun, P. Nucleolus and nuclear periphery: velcro for heterochromatin. *Curr. Opin. Cell Biol.* **28**, 54–60 (2014).
36. Thakore, P. I. et al. Highly specific epigenome editing by CRISPR–Cas9 repressors for silencing of distal regulatory elements. *Nat. Methods* **12**, 1143–1149 (2015).
37. De Cecco, M. et al. L1 drives IFN in senescent cells and promotes age-associated inflammation. *Nature* **566**, 73–78 (2019).
38. Chu, C. & Chang, H. Y. Understanding RNA–chromatin interactions using chromatin isolation by RNA purification (ChIRP). *Methods Mol. Biol.* **1480**, 115–123 (2016).
39. Konermann, S. et al. Genome-scale transcriptional activation by an engineered CRISPR–Cas9 complex. *Nature* **517**, 583–588 (2015).
40. Teo, G. et al. SAINTexpress: improvements and additional features in significance analysis of INTeractome software. *J. Proteom.* **100**, 37–43 (2014).
41. Jager, S. et al. Global landscape of HIV–human protein complexes. *Nature* **481**, 365–370 (2011).
42. Streva, V. A., Faber, Z. J. & Deininger, P. L. LINE-1 and Alu retrotransposition exhibit clonal variation. *Mob. DNA* **4**, 16 (2013).
43. Ohsawa, R., Seol, J. H. & Tyler, J. K. At the intersection of non-coding transcription, DNA repair, chromatin structure, and cellular senescence. *Front Genet.* **4**, 136 (2013).
44. Dewannieux, M., Esnault, C. & Heidmann, T. LINE-mediated retrotransposition of marked Alu sequences. *Nat. Genet.* **35**, 41–48 (2003).
45. Thomas, C. A. et al. Modeling of TREX1-dependent autoimmune disease using human stem cells highlights L1 accumulation as a source of neuroinflammation. *Cell Stem Cell* **21**, 319–331 (2017).
46. Andrenacci, D., Cavaliere, V. & Lattanzi, G. The role of transposable elements activity in aging and their possible involvement in laminopathic diseases. *Ageing Res. Rev.* **57**, 100995 (2020).
47. Kuo, C. C. et al. Detection of RNA–DNA binding sites in long noncoding RNAs. *Nucleic Acids Res.* **47**, e32 (2019).
48. O'Leary, V. B. et al. PARTICLE, a triplex-forming long ncRNA, regulates locus-specific methylation in response to low-dose irradiation. *Cell Rep.* **11**, 474–485 (2015).
49. Postepska-Igielska, A. et al. LncRNA Khps1 regulates expression of the proto-oncogene SPHK1 via triplex-mediated changes in chromatin structure. *Mol. Cell* **60**, 626–636 (2015).
50. Mondal, T. et al. MEG3 long noncoding RNA regulates the TGF-β pathway genes through formation of RNA–DNA triplex structures. *Nat. Commun.* **6**, 7743 (2015).
51. Kunkler, C. N. et al. Stability of an RNA*DNA–DNA triple helix depends on base triplet composition and length of the RNA third strand. *Nucleic Acids Res.* **47**, 7213–7222 (2019).
52. Yu, H., Zhu, S., Zhou, B., Xue, H. & Han, J. D. Inferring causal relationships among different histone modifications and gene expression. *Genome Res.* **18**, 1314–1324 (2008).

Methods

Cell culture

NIH3T3 (SCSP-515), HEK293T (SCSP-502) and WI38 (GNHu42) cells were obtained from the Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences. The NIH3T3 and HEK293T cells were cultured in DMEM medium supplemented with 10% fetal bovine serum and 1% penicillin–streptomycin. The WI38 cells were cultured in MEM medium containing 10% fetal bovine serum and 1% penicillin–streptomycin. All cells were maintained at 37 °C with 5% CO_2. For drug treatment, the cells were exposed to H_2O_2 (400 μM) 24 h after plating. At the end of treatment, the cells were collected using TRIzol reagent (Invitrogen) and stored at −80 °C for subsequent RNA extraction and RT–qPCR analysis.

Lentiviral transduction

Plasmid expressing sgRNA and dCas9–KRAB for *KCNQ1OT1* KD; plasmids encoding dCas9-VP64, MS2-P65-HSF1 activator helper complex and sgRNA (MS2) for *KCNQ1OT1* activation; the second-generation packaging plasmid psPAX2 and the envelope plasmid pMD2.G are available from Addgene (plasmid nos 71236, 61425, 61426, 61427, 12260 and 12259, respectively). The production of lentivirus and lentiviral transduction were performed as previously described[36]. Briefly, HEK293T cells were seeded at a density of 5.0×10^3 cells cm^{-2} one day before being cotransfected with dCas9–KRAB lentiviral expression plasmid for *KCNQ1OT1* KD or plasmids for *KCNQ1OT1* activation (5 μg), psPAX2 (4 μg) and pMD2.G (1.5 μg) with Lipofectamine 3000 reagent (Invitrogen). After 24 h, the medium containing lentivirus was collected and the cells were supplemented with fresh medium. The second batch of medium containing lentivirus was collected after a further 24 h and combined with the first-batch medium. Residual producer cells were removed by filtering the lentiviral supernatant through 0.45-μm filters. The filtered viral supernatant was aliquoted, snap-frozen in liquid nitrogen and stored at −80 °C for future use. For transduction, the viral supernatant was diluted 1:3 with DMEM medium. Cationic polymer polybrene was added to the viral medium at a final concentration of 4 μg ml^{-1}. One day after transduction, the medium was exchanged for fresh medium. Four days after transduction, puromycin, blasticidin, hygromycin and zeocin were added to the medium at a final concentration of 1, 20, 250 and 300 μg ml^{-1}, respectively, to select transduced cells.

eCLIP

The eCLIP was performed according to a published protocol[29]. Briefly, the cells were crosslinked with ultraviolet light (254 nm; 400 mJ cm^{-2}), followed by lysis in eCLIP lysis buffer (50 mM Tris–HCl, pH 7.4, 100 mM NaCl, 1% NP-40, 0.1% SDS, 0.5% sodium deoxycholate and freshly added protease inhibitor) and digestion with RNase I (Ambion). Immunoprecipitation was performed using anti-HP1α (Abcam, ab109028) and M-280 sheep anti-rabbit IgG beads (Invitrogen). After 3′ RNA adaptor ligation, the samples were run on protein gels and transferred to membranes (polyvinylidene fluoride membrane for western blotting and nitrocellulose membrane for library preparation). A region 75 kDa above the HP1α band on the nitrocellulose membrane was cut and treated with proteinase K to isolate RNA. The RNA was converted to cDNA, followed by 3′ DNA adaptor ligation. The sample was then PCR amplified and fragment size-selected by agarose gel electrophoresis and gel extraction.

Immunofluorescence staining

Cells were fixed in 4% paraformaldehyde for 15 min at room temperature, permeabilized with 0.5% Triton X-100 for 10 min, blocked with 2% BSA and incubated overnight at 4 °C with primary antibodies diluted in 0.5% BSA as follows: anti-H3K9me3 (1:500; Abcam, ab8898) and anti-DNA, single-stranded-specific (5 μg ml^{-1}; Millipore, MAB3299). The next day the cells were incubated with secondary antibodies at 4 °C for 1 h. The secondary antibodies, conjugated to Alexa Fluor 488 (Abcam, ab150073) and 647 (Abcam, ab150115), were diluted 1:200 in 0.5% BSA. Before mounting, the DNA was stained with DAPI. For ssDNA staining, the cells were treated according to a method described previously[45]. Briefly, the cells were fixed with 4% paraformaldehyde, followed by overnight incubation at −20 °C in methanol. Before blocking with 3% BSA, the cells were treated with 200 μg ml^{-1} RNase at 37 °C for 4 h. The cells were stained according to the procedure described above. Heterochromatin foci and ssDNA puncta were quantified using the Fiji software. No statistical methods were used to pre-determine sample sizes but our sample sizes are similar to those reported in previous publications[37,53].

RT–qPCR

Total RNA was isolated using TRIzol reagent, followed by treatment with TURBO DNase (Invitrogen) according to the manufacturer's instructions. Purified total RNA (1 μg) was reverse-transcribed into single-stranded cDNA using SuperScript III reverse transcriptase (Invitrogen) in 20-μl reactions. The reverse transcription product (2 μl) was used for subsequent real-time qPCR reactions using the SYBR Green system (Takara); *β-actin* mRNA was used as the normalization control in all experiments. The results are expressed as fold changes, determined using the $\Delta\Delta C_t$ method. Primers for each gene and repetitive elements are provided in Supplementary Table 7.

MeDIP

Genomic DNA was isolated using a QIAamp DNA mini kit (Qiagen). For each MeDIP experiment, 2 μg DNA was fragmented into 100–500 bp by sonication. The DNA fragments were subjected to end repair and adaptor ligation using End repair/dA-tailing (NEB, E7442) and Ligation (NEB, E7445) modules, respectively. Immunoprecipitation was performed with 2 μg anti-5mC (Abcam, ab214727) in IP buffer (100 mM sodium phosphate, pH 7.0, 1.4 M NaCl and 0.5% Triton X-100). After an overnight incubation at 4 °C, 20 μl pre-washed protein A/G magnetic beads (Thermo Scientific) were added to the IP buffer, followed by an additional incubation at 4 °C for 2 h. The beads were washed with IP buffer and DNA was purified with phenol:chlorophorm:isoamyl alcohol after proteinase K digestion. The isolated DNA was amplified by PCR and size-selected for deep sequencing.

ChOP

The ChOP assay was conducted according to a previous method[31] with minor modifications. Briefly, 2×10^7 HEK293T cells were crosslinked with 1% formaldehyde for 10 min at room temperature and quenched with 125 mM glycine, followed by three washes with ice-cold PBS. The cell pellets were resuspended in 2 ml buffer A (3 mM $MgCl_2$, 10 mM Tris–HCl, pH 7.4, 10 mM NaCl, 0.5% NP-40, 0.5 mM PMSF and 100 units ml^{-1} RNase inhibitor) and incubated on ice for 20 min. After centrifugation at 13,000*g* for 10 min at 4 °C, the nuclear pellets were resuspended in 1.2 ml buffer B (50 mM Tris–HCl, pH 7.4, 10 mM EDTA, 0.5% Triton X-100, 0.1% SDS, 0.5 mM PMSF and 100 units ml^{-1} RNase inhibitor), incubated on ice for 40 min. Next, an equal volume of buffer C (15 mM Tris–HCl, pH 7.4, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.5 mM PMSF and 100 units ml^{-1} RNase inhibitor) was added to the samples, followed by incubation on ice for 15 min. The samples were sonicated for 30 cycles using a Qsonica sonicator at 4 °C (30% amplitude; 30 s on and 30 s off). The cell debris were removed by centrifugation and the supernatant was transferred to new tubes. Biotinylated DNA probes, complimentary to KCNQ1OT1 RNA or non-specific probes, were added to the supernatant together with 100 μg ml^{-1} yeast transfer RNA and 100 μg ml^{-1} salmon sperm DNA. After an overnight incubation at 4 °C, streptavidin beads were added to samples, followed by incubation at 4 °C for an additional 3 h. The beads were washed and proteins were eluted from the beads. The eluted proteins were subjected to western blotting to detect interactions between KCNQ1OT1 RNA and specific

proteins. To identify all proteins associated with KCNQ1OT1 RNA, the eluted proteins were analysed using mass spectrometry.

Western blotting

Protein samples obtained from ChOP were mixed with 2×Laemmli buffer at a 1:1 ratio and boiled for 10 min at 100 °C. The proteins were separated through SDS–PAGE and transferred onto polyvinylidene fluoride membranes. Non-specific binding was blocked with 5% milk in PBS-T buffer by incubation at room temperature for 1 h. The membranes were incubated overnight at 4 °C with primary antibodies (anti-HP1α (Abcam, ab109028) and anti-β-actin (Cell Signaling Technology, 4967)) diluted in blocking buffer. The membranes were then incubated with secondary antibodies diluted in blocking buffer at room temperature for 1 h. Protein bands were detected using West pico PLUS chemiluminescent substrate (Thermo Scientific) and autoradiography.

Sequential ChIRP–ChIP

The procedure of sequential ChIRP–ChIP was performed by carrying out ChIRP, followed by ChIP. The ChIRP and ChIP procedures are described in the following sections. In the sequential experiment, the streptavidin C1 beads used in ChIRP were replaced with streptavidin agarose beads (Thermo Scientific). After the last wash with wash buffer in ChIRP, the chromatin–RNA complex attached to the agarose beads was subjected to subsequent incubation with antibody and protein A/G magnetic beads and the ChIP steps described later.

ChIRP

ChIRP was performed as previously described[32]. Briefly, cells were crosslinked with 3% formaldehyde for 30 min at room temperature and then quenched with glycine. After three washes with ice-cold PBS, the cells were snap-frozen in liquid nitrogen and stored at −80 °C. The cell pellets (2×10^7) were resuspended in cell lysis buffer (50 mM Tris–HCl, pH 7.0, 10 mM EDTA, 1% SDS, freshly added 1 mM PMSF (Sigma-Aldrich), protease inhibitors (Roche) and RNase inhibitors (NEB)) and subsequently subjected to sonication to shear chromatin to ensure the bulk was smaller than 500 bp. The lysate was centrifuged to remove insoluble material and the supernatant was transferred to new tubes. Two inputs were saved for RNA-enrichment analysis by RT–qPCR and DNA samples. For hybridization, a 2×volume of freshly prepared hybridization buffer (750 mM NaCl, 1% SDS, 50 mM Tris–HCl, pH 7.0, 1 mM EDTA, 15% formamide, 1 mM PMSF, protease inhibitors and RNase inhibitors) was added along with probe at a concentration of 100 pM. Incubation was performed at 37 °C with end-to-end rotation for 16 h. Pre-washed streptavidin C1 beads (Invitrogen; 50 µl) were then added, followed by an additional incubation of 30 min. The beads were washed five times with wash buffer (2×SSC, 0.5% SDS and 1 mM PMSF) at 37 °C for 5 min. RNA and DNA were isolated separately and subjected to qPCR analysis or deep sequencing.

ChIP

Cells were crosslinked with 1% formaldehyde for 10 min at room temperature, followed by quenching with 125 mM glycine and subsequent washing with ice-cold PBS. For each sample, 5×10^6 nuclei were resuspended in 300 µl ChIP lysis buffer (50 mM HEPES–KOH, pH 7.5, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100 and 0.1% SDS) and sonicated using a Qsonica sonicator at 4 °C to shear chromatin to fragments of 200–500 bp. The samples were centrifuged for 15 min at 13,000*g* to remove the insoluble debris. The supernatants were transferred to new tubes and 1% supernatant was saved to be used as input. After incubating protein A/G magnetic beads (Thermo Scientific) with 5 µg of antibodies, anti-H3K9me3 (Abcam, ab8898), or non-specific control IgG (Cell Signaling Technology, 2729), the beads were washed twice with lysis buffer, resuspended with the sheared chromatin and incubated on a rotator overnight at 4 °C. The next morning the beads were washed three times with ice-cold lysis buffer, three times with ice-cold high-salt buffer (50 mM HEPES–KOH, pH 7.5, 350 mM NaCl, 1 mM EDTA, 1% Triton X-100 and 0.1% SDS), three times with ice-cold LiCl buffer (0.25 M LiCl, 1% NP-40, 1 mM EDTA and 10 mM Tris–HCl, pH 8.0) and once with ice-cold TE buffer (10 mM Tris–HCl, pH 7.4 and 1 mM EDTA) at 4 °C. Chromatin was eluted from the beads using elution buffer (50 mM Tris–HCl, pH 7.4, 10 mM EDTA and 1% SDS) with incubation at 65 °C for 1 h, followed by formaldehyde crosslinking reversal through overnight incubation at 65 °C. The DNA was purified using phenol:chloroform:isoamyl alcohol (Sigma-Aldrich).

RIP

The RIP experiment was performed according to a protocol from Abcam, with minor modifications. Briefly, 1×10^7 cells were resuspended in 10 ml nuclear isolation buffer (256 mM sucrose, 8 mM Tris–HCl, pH 7.5, 4 mM $MgCl_2$, 0.8% Triton X-100 and 0.2×PBS) and incubated on ice for 20 min. After resuspending nuclei in RIP buffer (150 mM KCl, 25 mM Tris, pH 7.4, 5 mM EDTA, 0.5 mM dithiothreitol, 0.5% NP-40, protease and RNase inhibitors), the chromatin was sheared using a dounce homogenizer. The insoluble debris were removed by centrifugation at 13,000*g* for 10 min. An aliquot of the lysate was saved as the input. Immunoprecipitation was conducted using 10 µg antibody and 40 µl protein A/G beads. RNA was extracted with TRIzol and treated with TURBO DNase. The purified RNA was used for reverse transcription and subsequent RT–qPCR analysis.

L1 retrotransposition assay

The L1 retrotransposition construct contains a full-length and retrotransposition-competent L1 tagged with a GFP retrotransposition indicator cassette and contains a puromycin-resistance marker. Plasmid transfection was performed using Lipofectamine 3000 reagent (Invitrogen) according to the manufacturer's instructions. Two days after transfection, the medium was supplemented with puromycin at 1 µg ml^{-1} to select transfected cells. After two weeks of puromycin selection, the surviving cells were analysed by FACS to determine the percentage of GFP^+ cells.

CRISPR–Cas9-mediated knockout and knock-in

GuideScan (www.guidescan.com) was used to design sgRNAs (Supplementary Table 2), which were cloned into the px330-mCherry vector (Addgene, 98750). For fragment deletion, plasmids containing up sgRNA and plasmids containing down sgRNA were co-transfected into cells using Lipofectamine 3000 reagent. For fragment insertion, ssDNA donors, containing 60-bases-long upstream and downstream homology arms and 70-bases-long engineered triplex-forming sequence (5′-GGAAAGAAGATAGGTGGGTAGATGGGTAGGAGGAAGGAGTAAGAATAAAGGGGTTATAAAAAGGGGGAGA-3′) or scrambled sequence (5′-GAGTATGGAAGAGTGTAAGAGAGGATGGAGGAGGGATAGAGGATAGTAAGAAGAATGGTAAGGAAGAGAG-3′), were prepared and co-transfected with px330 plasmid into cells according to a published method[54]. Single $mCherry^+$ cells were sorted into 96-well plates by FACS 24 h after transfection. The desired single-cell clones were identified by PCR and Sanger sequencing.

Triplex pulldown assay

Both in vitro and in vivo triplex pulldown assays were performed according to a previous method[49], with minor modifications. Briefly, Alu and L1 DNA were synthesized through PCR using genomic DNA as the template and biotinylated L1 RNA was generated by in vitro transcription using L1 DNA derived from the *KCNQ1OT1* gene as the template (Supplementary Table 8). For in vitro pulldown, 100 fmol DNA was incubated with 1 pmol biotin-labelled RNA in triplex-forming buffer (10 mM Tris–HCl, pH 7.5, 20 mM KCl, 10 mM $MgCl_2$, 0.05% Tween 20 and 100 U RNase inhibitor (NEB)) for 20 min at room temperature. The RNA–DNA complexes were then isolated using streptavidin-coated Dynabeads (Invitrogen), followed by three washes with wash buffer 1 (150 mM KCl, 10 mM Tris–HCl, pH 7.5, 5 mM $MgCl_2$, 0.5% NP-40 and

100 U RNase inhibitor (NEB)) and once with wash buffer 2 (15 mM KCl, 10 mM Tris–HCl, pH 7.5 and 5 mM $MgCl_2$). The RNA-associated DNA was eluted from the beads by incubating in elution buffer (1% SDS, 50 mM Tris–HCl, pH 8.0 and 10 mM EDTA) for 5 min at 65 °C. The DNA was purified with phenol:chloroform:isoamyl alcohol (Sigma-Aldrich) following digestion with RNase A (50 ng μl^{-1}; 30 min at 37 °C) and proteinase K (200 ng μl^{-1}; 30 min at 55 °C). Recovered DNA was applied to qPCR analysis and normalized to the input DNA.

For in vivo pulldown, biotinylated RNA was transfected into HEK293T cells. The cells were collected 24 h after transfection and nuclei were isolated. The nuclei were then sonicated in triplex buffer (10 mM Tris–HCl, pH 7.5, 20 mM KCl, 10 mM $MgCl_2$ and 100 U RNase inhibitor) containing 100 ng proteinase K. The RNA–DNA complexes were enriched with streptavidin-coated Dynabeads (Invitrogen). After washing with buffer 1 and buffer 2, the RNA-associated DNA was eluted and purified as per the treatment for in vitro pulldown. The recovered DNA was analysed by qPCR and normalized to the input DNA.

SA-β-gal cell staining

Cells were fixed and stained with a senescence-β-galactosidase staining kit (Cell Signaling Technology, 9860) according to the manufacturer's protocol. Staining was performed overnight at 37 °C. The cells were observed under a microscope the following day and positive-staining cells were quantified.

Mass spectrometry data analysis

The SEQUEST HT engine[55] in the Proteome Discoverer 2.2 software (Thermo Fisher Scientific) was used to search against the human proteome (UniProt-reviewed sequences; accessed 10 March 2020)[56]. The peptides of the same protein were merged, controlling the protein false-discovery rate (FDR) confidence at 1%, and proteins with <2 unique peptides were removed to obtain the prey protein data for subsequent analysis. The normalized spectral abundance factor (NSAF)[57] was used to calculate the relative abundance of prey proteins in each sample. The MiST (https://github.com/kroganlab/mist)[41,58] and SAINTexpress (v3.6.3)[40] software were then used with default parameters to identify high-confidence protein–lncRNA interactions. For SAINTexpress, annotated protein interactions in the CORUM database (accessed 24 December 2020)[59] were used. Only protein–lncRNA interactions with MiST score of ≥0.7, SAINTexpress Bayesian FDR ≤ 0.1 and average spectral count ≥ 2 were kept for further analyses. Cytoscape (v3.8.0)[60] was used to visualize high-confidence protein–lncRNA interactions. In addition to the high-confidence protein–lncRNA interactions obtained by mass spectrometry, the physical interactions between prey proteins in the BIOGRID database (Release 4.1.190)[61] were also included in the network. The ClusterProfiler package[62] in the R software (v3.6.1) was used to annotate the enriched protein complexes and biological processes on the network and among high-confidence proteins of each sample. For each network module, if there were redundant enriched GO terms, a representative was manually selected as the annotation. For functional annotation, the top five most enriched results of each sample were retained for visualization and clustering. The threshold for enrichment is q-value ≤ 0.05 after Benjamini–Hochberg multiple testing correction. For any protein i in cell j, we calculated the relative ΔNSAF as follows:

Step 1: Calculate the ΔNSAF of protein i in cell j:

$$\Delta\mathrm{NSAF}_{ij} = \begin{cases} \sum_{n=1}^{KCNQ1OT1\ \text{probes}} \mathrm{NSAF}_{ij} - \sum_{n=1}^{lacZ\ \text{probes}} \mathrm{NSAF}_{ij} \left(\Delta\mathrm{NSAF}_{ij} > 0\right) \\ 0 \left(\Delta\mathrm{NSAF}_{ij} \le 0\right) \end{cases}$$

Step 2: Calculate the relative ΔNSAF a_{ij} of protein i in cell j:

$$a_{ij} = \frac{\Delta\mathrm{NSAF}_{ij}}{\sum_{j=1}^{\mathrm{cell}} \Delta\mathrm{NSAF}_{ij}}$$

Where j cell types included WT, K non-rep KO and K rep KO HEK293T cells, and n, the number of replicates of *KCNQ1OT1* and *lacZ* probes.

High-throughput sequencing data analysis

RNA-seq reads derived from NIH3T3 and HEK293T cells were mapped to GRCm38 and GRCh38, respectively, using STAR-2.6.0c[63]. Uniquely mapped mouse reads were quantified by featureCount and normalized to FPKM. The reads per million value of repeats was quantified using iteres (version 0.3.2-r121) and $\log_2$-transformed. Mapped reads of human were quantified using STAR in strand-specific manner and normalized to the ERCC RNA spike-in (Invitrogen). Annotation of putative active L1 was obtained from L1Base 2. For eCLIP–seq data, reads were mapped using STAR-2.6.0c, and mapped reads were quantified and normalized using featureCount and FPKM, respectively. For the ChIRP, ChIP and MeDIP data, uniquely mapped reads were aligned using bowtie2 (ref. [64]) and peaks were identified using MACS2 (ref. [65]) with the default parameters. For ChIRP peaks, only those simultaneously obtained by both odd and even probes were considered as true signals. The mean profiles of the MeDIP peaks were generated using deepTools[66] by unique sequence mapping using the default parameters.

Transposable element subfamily enrichment analysis

The enrichment of a transposable element subfamily for KCNQ1OT1 binding regions (overlap of at least 1 bp) was tested using a hypergeometric test and the P value was corrected using Bonferroni's method as done previously[67]. An FDR < 1×10^{-100} was assigned 1×10^{-100} and an enrichment score >10 was assigned 10. The enrichment score over the background was calculated as

$$\text{Enrichment score} = \frac{k}{m} / \frac{n}{N}$$

where k is the number of specific transposable element subfamily in the input list, m is the number of transposable elements in the input list, n is the number of the specific subfamily in the genome and N is the total number of transposable elements in the genome.

Statistical analysis

General statistical analyses, such as mean and t-tests, were performed using Excel. The details of individual tests are reported in the respective figure legends. For heterochromatin foci and ssDNA quantification, the immunofluorescence images were randomly selected and scored.

Bayesian network inference

H3K9me3 and DNA methylation reads within ±1 kb of the Alu/L1 centre were counted in FPKM. DNA methylation, H3K9me3, Alu/L1 expression and predicted triplexes were discretized to one and zero according to the following standards: FPKM > 0.1 for MeDIP–seq, H3K9me3 and Alu/L1 expression; and normalized counts > 4 for predicted triplex-forming sequences. The WinMine package (https://www.microsoft.com/en-us/research/project/winmine-toolkit/) was used to calculate joint conditional probability and build the preliminary potential Bayesian networks as described[52].

Reporting summary

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

Data availability

Sequencing data that support the findings of this study have been deposited in the Gene Expression Omnibus (GEO) under the accession code GSE208017. The protein mass spectrometry data have been deposited to ProteomeXchange under the accession code PXD035996. Previously published data that were reanalysed here are available:

KCNQ1OT1 expression in the human brain (National Omics Data Encyclopedia, OEP001041) and blood (GEO, GSE106670), H3K9me2 ChIP–seq data in HEK293T cells (GEO, GSE67317), H3K27me3 ChIP–seq data in HEK293T cells (GEO, GSE97326), and RNA-seq data in proliferating, early senescent and deep senescent human lung fibroblast cells (GEO, GSE109700). Source data are provided with this paper. All other data supporting the findings of this study are available from the corresponding author on reasonable request.

Code availability

The WinMine package was used for Bayesian network inference. Triplex Domain Finder characterizes the triplex-forming potential between RNA and DNA regions[47]. For high-confidence protein–lncRNA interactions analyses, MiST (https://github.com/kroganlab/mist), SAINTexpress (v3.6.3), BioGRID database (release 4.1.190) and the CORUM database (accessed 24 December 2020) were used[40,41,47,58,59].

References

53. Emam, A. et al. Stalled replication fork protection limits cGAS–STING and P-body-dependent innate immune signalling. *Nat. Cell Biol.* **24**, 1154–1164 (2022).
54. Miura, H., Quadros, R. M., Gurumurthy, C. B. & Ohtsuka, M. Easi-CRISPR for creating knock-in and conditional knockout mouse models using long ssDNA donors. *Nat. Protoc.* **13**, 195–215 (2018).
55. Eng, J. K., McCormack Al Fau, -, Yates, J. R. & Yates, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass. Spectrom.* **5**, 976–989 (1994).
56. UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480–D489 (2021).
57. Paoletti, A. C. et al. Quantitative proteomic analysis of distinct mammalian Mediator complexes using normalized spectral abundance factors. *Proc. Natl Acad. Sci. USA* **103**, 18928–18933 (2006).
58. Verschueren, E. et al. Scoring large-scale affinity purification mass spectrometry datasets with MiST. *Curr. Protoc. Bioinform.* **49**, 8.19.1–8.19.16 (2015).
59. Giurgiu, M. et al. CORUM: the comprehensive resource of mammalian protein complexes-2019. *Nucleic Acids Res.* **47**, D559–D563 (2019).
60. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
61. Oughtred, R. et al. The BioGRID database: a comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Sci.* **30**, 187–200 (2021).
62. Yu, G. et al. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
63. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
64. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
65. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
66. Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, W187–W191 (2014).
67. Cao, Y. et al. Widespread roles of enhancer-like transposable elements in cell identity and long-range genomic interactions. *Genome Res.* **29**, 40–52 (2019).

Acknowledgements

We thank J. Wang (Sun Yat-Sen University) for providing L1 retrotansposation vector and B. Zhu (Institute of Biophysics) for critical reading of the manuscript and invaluable suggestions, the technology platform of the National Center for Protein Sciences, Peking University for mass spectrometry and flow cytometry analysis, and the Center for Quantitative Biology, Peking University for microscopy imaging. This work was supported by grants from the National Natural Science Foundation of China (grant nos 92049302 and 32088101), China Ministry of Science and Technology (grant no. 2020YFA0804000) and Shanghai Municipal Science and Technology Major Project (grant no. 2017SHZDZX01) to J.-D.J.H.

Author contributions

X.Z. and J.-D.J.H. designed experiments. J.-D.J.H., X.Z. and J.C. designed analyses. X.Z. and J.T. performed wet experiments. Q.J., J.L., S.Z., Y.C., X.X. and D.C. performed computational analyses, and X.Z., J.-D.J.H., Q.J., J.L. and S.Z. wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

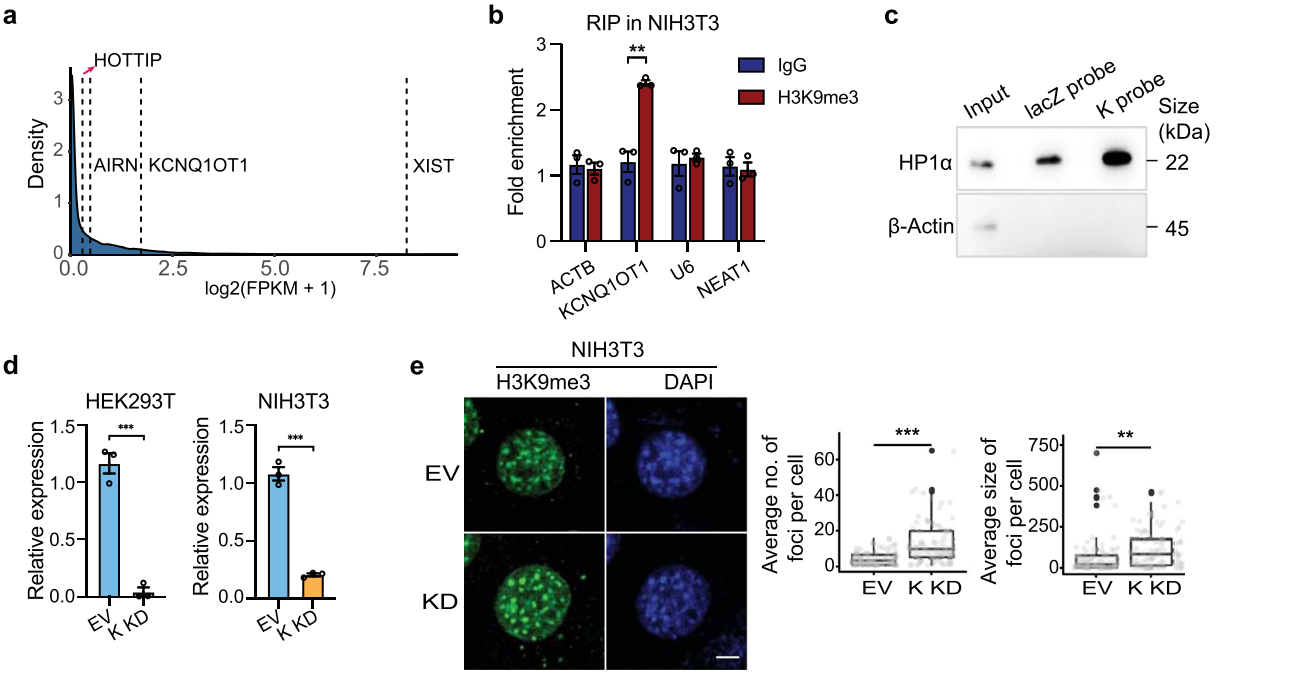
Extended data is available for this paper at https://doi.org/10.1038/s41556-022-01008-5.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41556-022-01008-5.

Correspondence and requests for materials should be addressed to Jing-Dong J. Han.

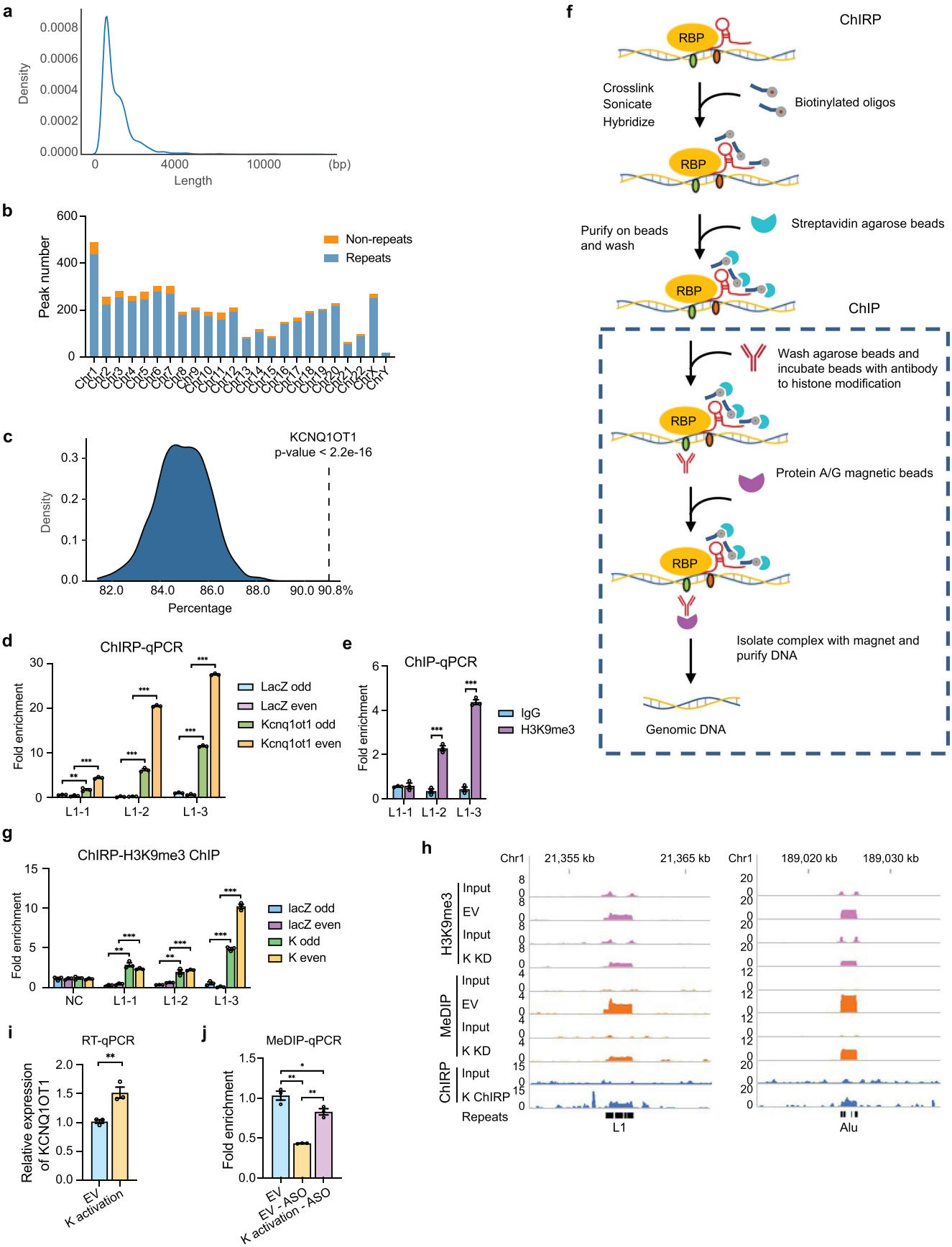
Peer review information *Nature Cell Biology* thanks Edward Chuong, John Sedivy and Rory Johnson for their contribution to the peer review of this work. Peer reviewer reports are available.

https://doi.org/10.1038/s41556-022-01008-5



Extended Data Fig. 1 | LncRNA KCNQ1OT1 binds HP1 and suppresses H3K9me3 foci in mouse NIH3T3 cells. a, Expression rank of XIST, KCNQ1OT1, AIRN, and HOTTIP among all lncRNAs derived from HEK293T RNA-seq data. **b**, RIP performed using antibodies against control immunoglobulin G (IgG) or H3K9me3 in mouse NIH3T3 cells. **c**, *In vitro* RNA pull down with biotinylated DNA probes complementary to KCNQ1OT1 (K probe) or control probes (*lacZ* probe) followed by western blotting. This experiment was repeated three times with similar results. **d**, Expression of KCNQ1OT1 in control (transfected with empty vector, EV) and K KD cells (left, HEK293T; right, NIH3T3). **e**, Left, increase of H3K9me3 foci in KCNQ1OT1 KD 3T3 cells shown by immunofluoresence against H3K9me3. Nuclei were counterstained with DAPI. Scale bar, 5 µm. Right, quantification of heterochromatin foci. n = 94 (EV) and 110 (KD) cells pooled across three independent experiments. The bound of a box shows 25% (lower quartile) and 75% (upper quartile) of all values in the group. The centre of the box shows the median value. Values outside middle 50% value (Interquartile Range, or IQR) but within 1.5xIQR are shown as whiskers, otherwise as black dots. All values are shown as grey spots. Data in **a** and **d** are shown as mean ± s.e.m. (n = 3 independent experiments), and as fold enrichment relative to input. ACTB, U6 snRNA, and NEAT1 are used as negative controls for RIP-qPCR. Statistical significance was determined by two-tailed Student's *t*-test, *p < 0.05, ** p < 0.01, *** p < 0.001. Source numerical data and unprocessed blots are available in Source Data.

https://doi.org/10.1038/s41556-022-01008-5

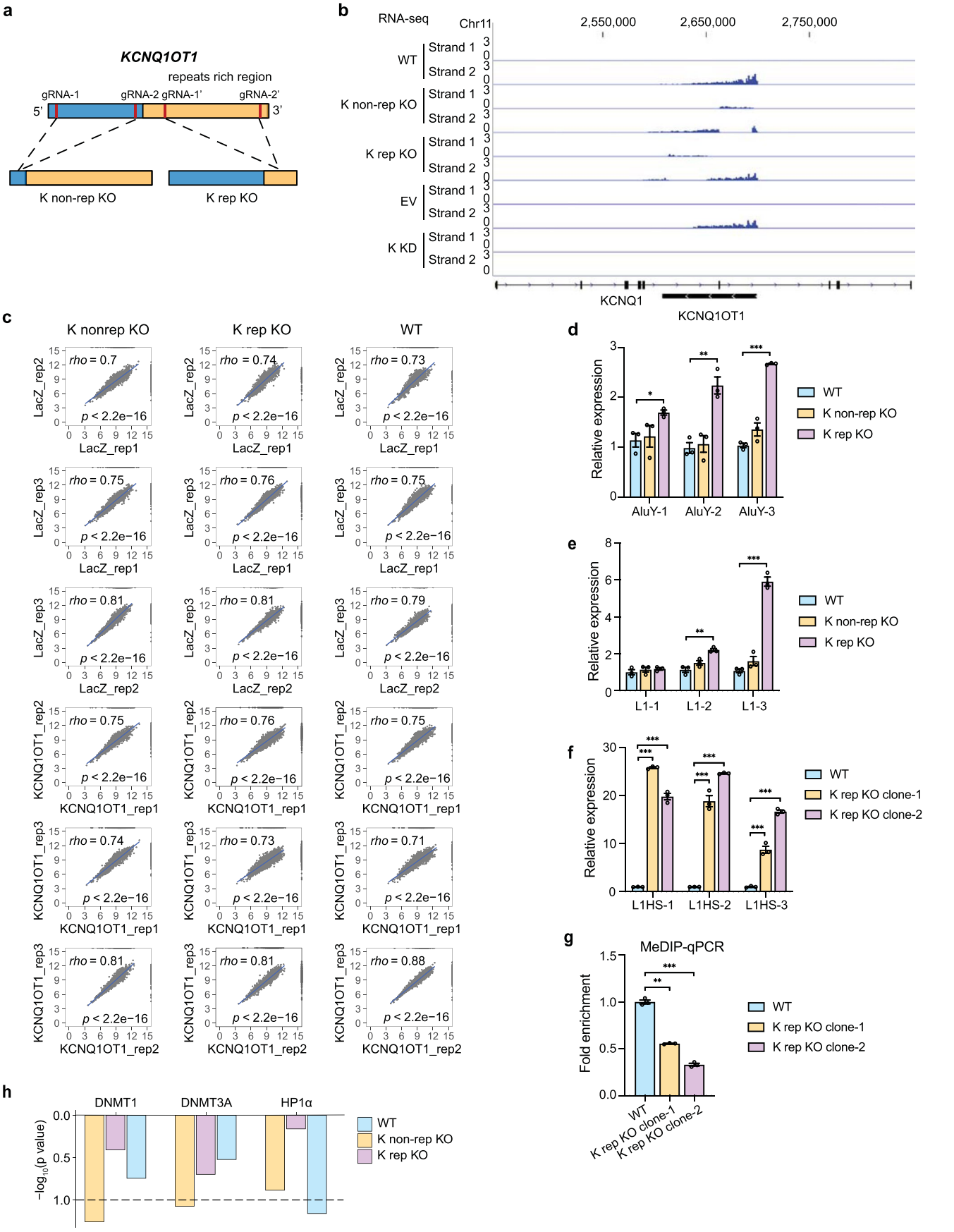


Extended Data Fig. 2 | See next page for caption.

https://doi.org/10.1038/s41556-022-01008-5

Extended Data Fig. 2 | KCNQ1OT1 lncRNA binding regions overlap heterochromatin marks in HEK293T cells. a, Length distribution of DNA sequences bound by KCNQ1OT1. **b**, Number of KCNQ1OT1 ChIRP–seq peaks ≥ 1 bp overlapped with at least one repeat or no repeats on each chromosome. **c**, Compared to 1000 times random sampling of the same number of genomic fragments of the exact same length distribution as KCNQ1OT1 ChIRP–seq peaks (panel 2a), the total 90.8% overlap of KCNQ1OT1 ChIRP–seq peak length to repeat elements is significantly larger than those generated by random sampling with binomial distribution test $p < 2.2e\text{-}16$, empirical $p < 0.001$. **d**, Validation of 3 randomly selected KCNQ1OT1 targeted L1 repeats by ChIRP–qPCR. **e**, ChIP-qPCR of H3K9me3 in HEK293T cells at the same 3 L1 repeats shown in **d** and at one negative control region. **f**, Flow-chart of sequential ChIRP–ChIP. **g**, Sequential ChIRP–ChIP-qPCR showing H3K9me3 at 3 randomly selected KCNQ1OT1 lncRNA associated L1 repeats regions and one negative control (NC) region. **h**, Genome-browser tracks of H3K9me3 ChIP–seq, MeDIP-seq, and ChIRP–seq alignments at two representative repeats regions, including L1 and Alu. **i**, Expression of KCNQ1OT1 in control (transfected with empty vector, EV) and KCNQ1OT1 activation (transfected with gRNA expression vector) cells. **j**, DNA methylation status on a representative KCNQ1OT1 binding L1 element (qPCR region: chr1:65,797,529-65,797,674) in EV, EV-ASO, and K activation-ASO cells detected by MeDIP-qPCR. Data in **d**, **e**, **g**, **i**, and **j** are mean ± s.e.m. (n = 3 independent experiments). Statistical significance was determined by two-tailed Student's *t*-test, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. Source numerical data are available in Source Data.

https://doi.org/10.1038/s41556-022-01008-5

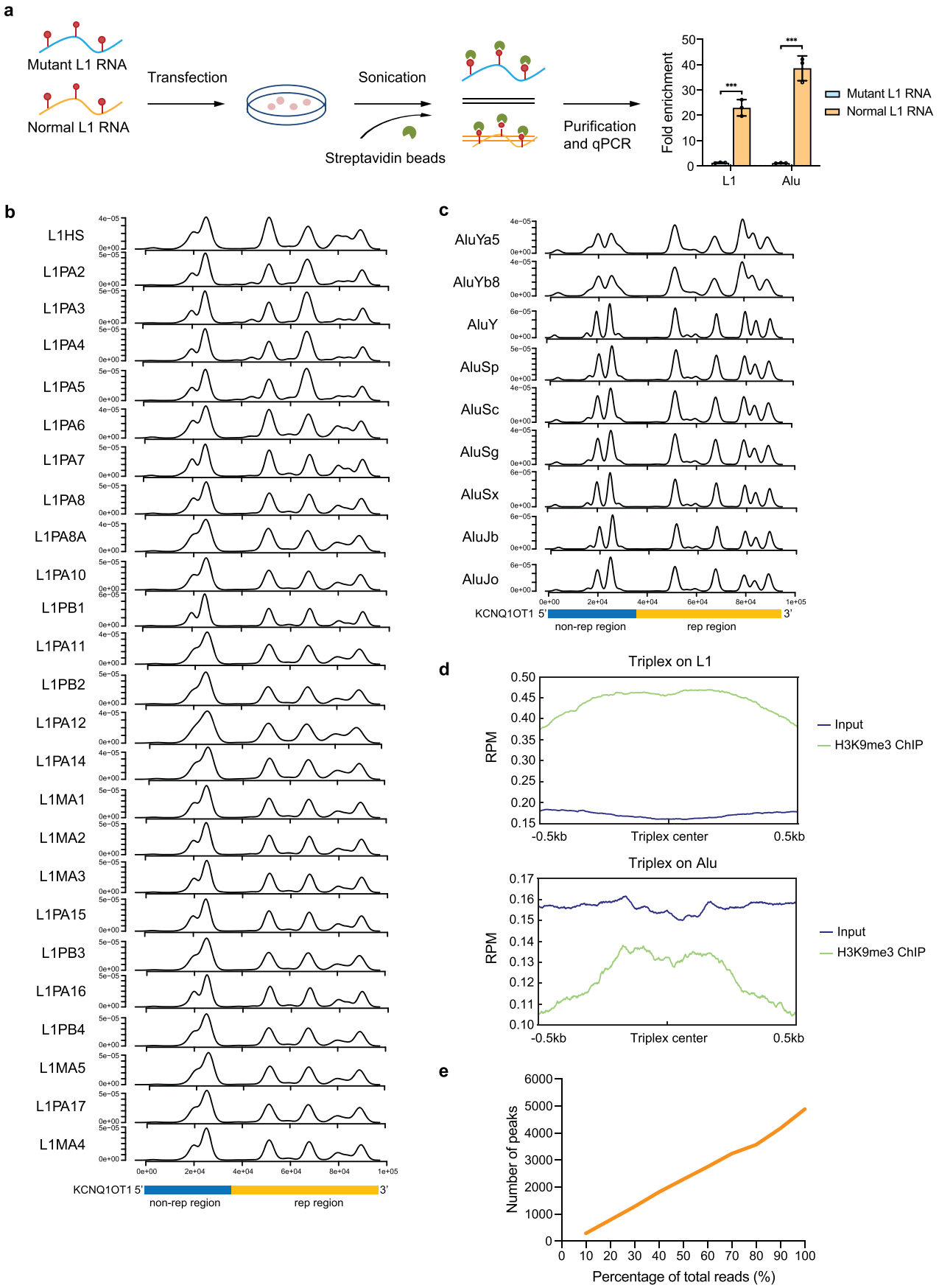


Extended Data Fig. 3 | See next page for caption.

https://doi.org/10.1038/s41556-022-01008-5

Extended Data Fig. 3 | Binding proteins and repetitive elements expression changes in K non-rep KO and K rep KO versus WT HEK293T cells. a, Diagram showing regions of KCNQ1OT1 deleted in K non-rep KO and K rep KO cells. **b**, Strand-specific RNA-seq tracks show the expression of *KCNQ1* and *KCNQ1OT1* in WT, K non-rep KO, K rep KO, EV, and K KD HEK293T cells. **c**, Correlation analysis of the ChOP-MS dataset shows good correlation between biological replicates. Spearman's rank correlation coefficient (rho) and p-value between two replicate samples are shown. KCNQ1OT1 denotes KCNQ1OT1 probe, LacZ denotes LacZ probe. **d**, Expression of randomly selected AluY elements that are targeted by KCNQ1OT1 assessed by RT-qPCR. **e**, Expression of randomly selected L1 elements that are targeted by KCNQ1OT1 assessed by RT-qPCR. **f**, Expression of L1HS in two different K rep KO clonal cell lines detected by RT-qPCR. **g**, DNA methylation status on a representative KCNQ1OT1 binding L1 element in WT, K rep KO clone-1, and K rep KO clone-2 cells detected by MeDIP-qPCR. The qPCR primers amplify unique sequence within the repetitive element (chr1:65,797,529-65,797,674). **h**, Significance of DNMT1, DNMT3A and HP1α in WT, K non-rep KO and K rep KO cells (use NSAF for one-tailed paired Student's *t*-test). Data in **d**, **e**, **f**, and **g** are mean ± s.e.m. (n = 3 independent experiments). Statistical significance was determined by two-tailed Student's *t*-test, *p < 0.05, **p < 0.01, ***p < 0.001. Source numerical data are available in Source Data.

https://doi.org/10.1038/s41556-022-01008-5



Extended Data Fig. 4 | See next page for caption.

https://doi.org/10.1038/s41556-022-01008-5

Extended Data Fig. 4 | KCNQ1OT1 binding at L1 and Alu elements that contain triplex-forming sequences. a, In vivo triplex pulldown by transfecting biotinylated normal L1 (containing poly(A)) or mutant L1 RNA (containing poly(G)) targeting the site indicated in Fig. 5d (left panel) into HEK293T cells. RNA-associated DNA was analysed by qPCR and normalized to input DNA. The qPCR primers amplify unique sequences within the repetitive elements (L1, chr11:93,426,919-93,427,079; Alu, chr7:105,405,644-105,405,823). **b** and **c**, Probability density of number of triplexes on KCNQ1OT1 with each L1 (**b**) and Alu (**c**) subfamilies in each 100 bp interval along the KCNQ1OT1 genes. **d**, Average input and H3K9me3 ChIP–seq signal at the triplexes predicted on L1 and Alu elements. **e**, Saturation analysis for ChIRP–seq data. Peak numbers detected at the indicated percentage of reads randomly sampled. Data in **a** are shown as mean ± s.e.m. (n = 3 independent experiments). Statistical significance was determined by two-tailed Student's *t*-test, ** $p < 0.01$, *** $p < 0.001$. Source numerical data are available in Source Data.

nature portfolio

Corresponding author(s): Jing-Dong J. Han

Last updated by author(s): Sep 2, 2022

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Collection of qPCR data: Agilent AriaMx software version 1.7 Microscope imaging: Leica Application Suite X Immunoblot imaging: Tanon 4800 Multi Imaging System Mass spectrometry: Proteome Discoverer (version 2.2), Thermo Fisher Scientific
Data analysis	For RNA-seq data analyses, the following software were used: STAR-2.6.0c for mapping, featureCounts for counting reads, iteres-0.3.2 for repeats quantification, GSEApy-v0.9.4 for GSEA analysis. For eCLIP-seq data analyses, STAR-2.6.0c and featureCounts were used. For ChIRP-seq, ChIP-seq, and MeDIP-seq data, bowtie2 was used for mapping, MACS2 was used for calling peaks and deepTools was used for generating mean profiles. WinMine package was used for bayesian network inference. For high-confidence protein-lncRNA interactions analyses, MiST (https://github.com/kroganlab/mist), SAINTexpress (version 3.6.3), BioGRID database (Release 4.1.190) and CORUM database (December 24, 2020) were used. For network visualization, Cytoscape (version 3.8.0) was used. For functional annotation of enriched proteins, ClusterProfiler package (version 3.12.0) was used.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Sequencing data that support the findings of this study have been deposited in the Gene Expression Omnibus (GEO) under accession code GSE208017. The protein mass spectrometry data have been deposited to the ProteomeXchange under accession code PXD035996. Previously published data that were re-analysed here are available, KCNQ1OT1 expression in human brain NODE (OEP001041) and blood (GSE106670); GSE67317: H3K9me2 ChIP-seq data in HEK293T cells; GSE97326: H3K27me3 ChIP-seq data in HEK293T cells; GSE109700: RNA-seq data in proliferating, early senescent, and deep senescent human lung fibroblast cells. Source data are provided with this study. All other data supporting the findings of this study are available from the corresponding author on reasonable request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample sizes used in all experiments were explicitly described.
Data exclusions	No data exclusion was used.
Replication	Number of biological replicates were described in figure legends. All attempts at replication were successful. No findings were not replicated or could not be reproduced.
Randomization	Genomic sites used for ChIRP-qPCR and ChIP-qPCR validation were randomly selected based on ChIRP-seq data. Immunofluorescence images used for quantification were randomly selected.
Blinding	The investigators were blinded when quantifying immunofluorescence images.

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	*Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).*
Research sample	*State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.*
Sampling strategy	*Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.*
Data collection	*Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.*
Timing	*Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Non-participation	*State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.*

Randomization

If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description

Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates.

Research sample

Describe the research sample (e.g. a group of tagged Passer domesticus, all Stenocereus thurberi within Organ Pipe Cactus National Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source.

Sampling strategy

Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.

Data collection

Describe the data collection procedure, including who recorded the data and how.

Timing and spatial scale

Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken

Data exclusions

If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.

Reproducibility

Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.

Randomization

Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.

Blinding

Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.

Did the study involve field work? ☐ Yes ☐ No

Field work, collection and transport

Field conditions

Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).

Location

State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).

Access & import/export

Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).

Disturbance

Describe any disturbance caused by the study and how it was minimized.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☐	☒ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

For eCLIP, anti-HP1 alpha antibody (Abcam, ab109028) was used;
For immunofluorescence, anti-H3K9me3 antibody (Abcam, ab8898), anti-DNA antibody, single stranded specific (Millipore, MAB3299), and the secondary antibody conjugated to Alexa Fluor 488 (Abcam, ab150073) and 647 (Abcam, ab150115) were used;
For Western Blotting, anti-HP1 alpha antibody (Abcam, ab109028), and anti-β-actin antibody (Cell Signaling Technology, 4967) were used;
For MeDIP, anti-5mC antibody (Abcam, ab214727) was used;
For ChIP and RIP, anti-H3K9me3 antibody (Abcam, ab8898) and anti-IgG antibody (Cell Signaling Technology, 2729) were used.

Validation

Anti-HP1 alpha antibody, Abcam, ab109028
Host species: Rabbit
Application: Flow Cyt (Intra), ICC, WB, IP, IHC-P
Reacts with: Mouse, Rat, Human
Knockout validated

Anti-H3K9me3 antibody, Abcam, ab8898
Host species: Rabbit
Application: WB, IHC-P, ICC, ChIP
Reacts with: Mouse, Cow, Human

Anti-DNA antibody, single stranded specific, Millipore, MAB3299
Host species: Mouse
Application: Flow Cyt, ICC, IHC, IH(P)
Reacts with: All species

Anti-β-actin antibody, Cell Signaling Technology, 4967
Host species: Rabbit
Application: WB, IP, IHC, ChIP, IF, Flow Cyt, ELISA
Reacts with: All species

Anti-5mC antibody, Abcam, ab214727
Host species: Rabbit
Application: MeDIP, ELISA, ICC, IHC-P, Flow Cyt, Dot blot
Reacts with: 5-methylcytosine in both single-stranded and double-stranded DNA

Anti-IgG antibody, Cell Signaling Technology, 2729
Host species: Rabbit
Application: WB, IP, IHC, ChIP, IF, Flow Cyt, ELISA
Reacts with: All species

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

NIH3T3, HEK293T, and WI38 cells are obtained from the cell bank affiliated to Shanghai Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences.

Authentication

All cell lines were authenticated by the vendor listed above.

Mycoplasma contamination

All cell lines were tested negative for mycoplasma infection by the vendor.

Commonly misidentified lines
(See ICLAC register)

No commonly misidentified cell lines were used.

Palaeontology and Archaeology

Specimen provenance

Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information). Permits should encompass collection and, where applicable, export.

Specimen deposition

Indicate where the specimens have been deposited to permit free access by other researchers.

Dating methods

If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Ethics oversight

Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	*For laboratory animals, report species, strain, sex and age OR state that the study did not involve laboratory animals.*
Wild animals	*Provide details on animals observed in or captured in the field; report species, sex and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.*
Field-collected samples	*For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.*
Ethics oversight	*Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about studies involving human research participants

Population characteristics	*Describe the covariate-relevant population characteristics of the human research participants (e.g. age, gender, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."*
Recruitment	*Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.*
Ethics oversight	*Identify the organization(s) that approved the study protocol.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	*Provide the trial registration number from ClinicalTrials.gov or an equivalent agency.*
Study protocol	*Note where the full trial protocol can be accessed OR if not available, explain why.*
Data collection	*Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.*
Outcomes	*Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.*

Dual use research of concern

Policy information about dual use research of concern

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No	Yes	
☐	☐	Public health
☐	☐	National security
☐	☐	Crops and/or livestock
☐	☐	Ecosystems
☐	☐	Any other significant area

Experiments of concern

Does the work involve any of these experiments of concern:

No	Yes	
☐	☐	Demonstrate how to render a vaccine ineffective
☐	☐	Confer resistance to therapeutically useful antibiotics or antiviral agents
☐	☐	Enhance the virulence of a pathogen or render a nonpathogen virulent
☐	☐	Increase transmissibility of a pathogen
☐	☐	Alter the host range of a pathogen
☐	☐	Enable evasion of diagnostic/detection modalities
☐	☐	Enable the weaponization of a biological agent or toxin
☐	☐	Any other potentially harmful combination of experiments and agents

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links
May remain private before publication.

https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE208017
http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD035996

Files in database submission

WT_rep1_1.mzid, WT_rep1_2.mzid, WT_rep1_3.mzid, WT_rep1_4.mzid, WT_rep2_1.mzid, WT_rep2_2.mzid, WT_rep2_3.mzid, WT_rep2_4.mzid, WT_rep3_1.mzid, WT_rep3_2.mzid, WT_rep3_3.mzid, WT_rep3_4.mzid, WT_CTL_rep1_1.mzid, WT_CTL_rep1_2.mzid, WT_CTL_rep1_3.mzid, WT_CTL_rep1_4.mzid, WT_CTL_rep2_1.mzid, WT_CTL_rep2_2.mzid, WT_CTL_rep2_3.mzid, WT_CTL_rep2_4.mzid, WT_CTL_rep3_1.mzid, WT_CTL_rep3_2.mzid, WT_CTL_rep3_3.mzid, WT_CTL_rep3_4.mzid, K non-rep KO_rep1_1.mzid, K non-rep KO_rep1_2.mzid, K non-rep KO_rep1_3.mzid, K non-rep KO_rep1_4.mzid, K non-rep KO_rep2_1.mzid, K non-rep KO_rep2_2.mzid, K non-rep KO_rep2_3.mzid, K non-rep KO_rep2_4.mzid, K non-rep KO_rep3_1.mzid, K non-rep KO_rep3_2.mzid, K non-rep KO_rep3_3.mzid, K non-rep KO_rep3_4.mzid, K non-rep KO_CTL_rep1_1.mzid, K non-rep KO_CTL_rep1_2.mzid, K non-rep KO_CTL_rep1_3.mzid, K non-rep KO_CTL_rep1_4.mzid, K non-rep KO_CTL_rep2_1.mzid, K non-rep KO_CTL_rep2_2.mzid, K non-rep KO_CTL_rep2_3.mzid, K non-rep KO_CTL_rep2_4.mzid, K non-rep KO_CTL_rep3_1.mzid, K non-rep KO_CTL_rep3_2.mzid, K non-rep KO_CTL_rep3_3.mzid, K non-rep KO_CTL_rep3_4.mzid, K rep KO_rep1_1.mzid, K rep KO_rep1_2.mzid, K rep KO_rep1_3.mzid, K rep KO_rep1_4.mzid, K rep KO_rep2_1.mzid, K rep KO_rep2_2.mzid, K rep KO_rep2_3.mzid, K rep KO_rep2_4.mzid, K rep KO_rep3_1.mzid, K rep KO_rep3_2.mzid, K rep KO_rep3_3.mzid, K rep KO_rep3_4.mzid, K rep KO_CTL_rep1_1.mzid, K rep KO_CTL_rep1_2.mzid, K rep KO_CTL_rep1_3.mzid, K rep KO_CTL_rep1_4.mzid, K rep KO_CTL_rep2_1.mzid, K rep KO_CTL_rep2_2.mzid, K rep KO_CTL_rep2_3.mzid, K rep KO_CTL_rep2_4.mzid, K rep KO_CTL_rep3_1.mzid, K rep KO_CTL_rep3_2.mzid, K rep KO_CTL_rep3_3.mzid, K rep KO_CTL_rep3_4.mzid, ./Human_ChIRP/B0dd_L4_P703502.R1.fastq.gz, ./Human_ChIRP/B0dd_L4_P703502.R2.fastq.gz, ./Human_ChIRP/Even_L3_P703503.R1.fastq.gz, ./Human_ChIRP/Even_L3_P703503.R2.fastq.gz, ./Human_ChIRP/Input_L4_320320.R1.fastq.gz, ./Human_ChIRP/Input_L4_320320.R2.fastq.gz, ./Human_H3K9me3/H3K9me3-EV.R1.fastq.gz, ./Human_H3K9me3/H3K9me3-EV.R2.fastq.gz, ./Human_H3K9me3/H3K9me3-KD.R1.fastq.gz, ./Human_H3K9me3/H3K9me3-KD.R2.fastq.gz, ./Human_H3K9me3/H3K9me3.R1.fastq.gz, ./Human_H3K9me3/H3K9me3.R2.fastq.gz, ./Human_H3K9me3/Input-EV.R1.fastq.gz, ./Human_H3K9me3/Input-EV.R2.fastq.gz, ./Human_H3K9me3/Input-KD.R1.fastq.gz, ./Human_H3K9me3/Input-KD.R2.fastq.gz, ./Human_H3K9me3/Input.R1.fastq.gz, ./Human_H3K9me3/Input.R2.fastq.gz, ./Human_MeDIP/InputC-KO_L4_A004.R1.fastq.gz, ./Human_MeDIP/InputC-KO_L4_A004.R2.fastq.gz, ./Human_MeDIP/Input-EV_L2_A001.R1.fastq.gz, ./Human_MeDIP/Input-EV_L2_A001.R2.fastq.gz, ./Human_MeDIP/InputK-KD_L2_A002.R1.fastq.gz, ./Human_MeDIP/InputK-KD_L2_A002.R2.fastq.gz, ./Human_MeDIP/InputL-KO_L2_A005.R1.fastq.gz, ./Human_MeDIP/InputL-KO_L2_A005.R2.fastq.gz, ./Human_MeDIP/InputWT_L4_A003.R1.fastq.gz, ./Human_MeDIP/InputWT_L4_A003.R2.fastq.gz, ./Human_MeDIP/MeDIP-K non-rep-KO_L2_A004.R1.fastq.gz, ./Human_MeDIP/MeDIP-K non-rep-KO_L2_A004.R2.fastq.gz, ./Human_MeDIP/MeDIP-EV_.R1.fastq.gz, ./Human_MeDIP/MeDIP-EV_.R2.fastq.gz, ./Human_MeDIP/MeDIP-K-KD_L4_A002.R1.fastq.gz, ./Human_MeDIP/MeDIP-K-KD_L4_A002.R2.fastq.gz, ./Human_MeDIP/MeDIP-K rep-KO_.R1.fastq.gz, ./Human_MeDIP/MeDIP-K rep-KO_.R2.fastq.gz, ./Human_MeDIP/MeDIP-WT_.R1.fastq.gz, ./Human_MeDIP/MeDIP-WT_.R2.fastq.gz, ./Mouse_HP1_eClip/HP1_L3_A002.R1.clean.fastq.gz, ./Mouse_HP1_eClip/Input_L3_A001.R1.clean.fastq.gz

Genome browser session
(e.g. UCSC)

Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.

Methodology

Replicates

One replicate.

Sequencing depth

EV: 40594060 total reads, 9166138 uniquely mapped reads, 150bp, paired-end; KD: 60261242 total reads, 19102813 uniquely mapped reads, 150bp, paired-end; WT: 36025271 total reads, 22803996 uniquely mapped reads, 150bp, paired-end; EV-input: 38502124 total reads, 6318198 uniquely mapped reads, 150bp, paired-end; KD-input: 16968667 total reads, 3010241 uniquely

	mapped reads, 150bp, paired-end, WT-input: 37619124 total reads, 28101485 uniquely mapped reads, 150bp, paired-end.
Antibodies	Anti-H3K9me3 antibody, Abcam, ab8890
Peak calling parameters	macs2 callpeak --broad --SPMR --nomodel --fix-bimodal -f BAMPE -g hs
Data quality	WT 15215 peaks; EV 26364 peaks; K KD 13455 peaks
Software	macs2

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Before experiment, cells were collected from culture dishes through trypsin digestion and resuspended in DMEM culture medium containing less than 2% FBS.
Instrument	BD Biosciences Aira SORP was used.
Software	Data were analysed using FlowJo 7.6.
Cell population abundance	Through microscopic detection, the abundance of the relevant cell populations within post-sort fractions was over 95%.
Gating strategy	For WT, K non-rep KO, and K rep KO HEK293T cell samples, living cells were selected by forward scatter and side scatter. GFP high living cells were sorted.

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

Design type	*Indicate task or resting state; event-related or block design.*
Design specifications	*Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.*
Behavioral performance measures	*State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).*

Acquisition

Imaging type(s)	*Specify: functional, structural, diffusion, perfusion.*
Field strength	*Specify in Tesla*
Sequence & imaging parameters	*Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.*
Area of acquisition	*State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.*

Diffusion MRI ☐ Used ☐ Not used

Preprocessing

Preprocessing software	*Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).*

Normalization

If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.

Normalization template

Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.

Noise and artifact removal

Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).

Volume censoring

Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.

Statistical modeling & inference

Model type and settings

Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation).

Effect(s) tested

Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference
(See Eklund et al. 2016)

Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.

Correction

Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).

Models & analysis

n/a	Involved in the study
☐	☐ Functional and/or effective connectivity
☐	☐ Graph analysis
☐	☐ Multivariate modeling or predictive analysis

Functional and/or effective connectivity

Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).

Graph analysis

Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).

Multivariate modeling and predictive analysis

Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.