nature chemical biology

Article

https://doi.org/10.1038/s41589-024-01671-9

# Uncoupling histone modification crosstalk by engineering lysine demethylase LSD1

Received: 13 September 2023

Accepted: 7 June 2024

Published online: 04 July 2024

Check for updates

Kwangwoon Lee[1,2,8], Marco Barone[3,8], Amanda L. Waterbury[4,5], Hanjie Jiang[1,2], Eunju Nam[1,2], Sarah E. DuBois-Coyne[1,2], Samuel D. Whedon[1,2], Zhipeng A. Wang[1,2], Jonatan Caroli[3], Katherine Neal[1,2], Brian Ibeabuchi[1,2], Zuzer Dhoondia[1,6], Mitzi I. Kuroda[1,6], Brian B. Liau[4,5], Samuel Beck[7] ✉, Andrea Mattevi[3] ✉ & Philip A. Cole[1,2] ✉

Biochemical crosstalk between two or more histone modifications is often observed in epigenetic enzyme regulation, but its functional significance in cells has been difficult to discern. Previous enzymatic studies revealed that Lys14 acetylation of histone H3 can inhibit Lys4 demethylation by lysine-specific demethylase 1 (LSD1). In the present study, we engineered a mutant form of LSD1, Y391K, which renders the nucleosome demethylase activity of LSD1 insensitive to Lys14 acetylation. K562 cells with the Y391K LSD1 CRISPR knockin show decreased expression of a set of genes associated with cellular adhesion and myeloid leukocyte activation. Chromatin profiling revealed that the *cis*-regulatory regions of these silenced genes display a higher level of H3 Lys14 acetylation, and edited K562 cells show diminished H3 mono-methyl Lys4 near these silenced genes, consistent with a role for enhanced LSD1 demethylase activity. These findings illuminate the functional consequences of disconnecting histone modification crosstalk for a key epigenetic enzyme.

Nuclear DNA is intricately packaged in macromolecular protein–nucleic acid complexes in the form of chromatin. Chromatin comprises nucleosomes that consist of an octamer with pairs of the four core histones H2A, H2B, H3 and H4 wrapped by 147 base pairs (bp) of DNA[1,2]. The rich tapestry of histone modifications that adorn nucleosomes serves to regulate epigenetic states, gene expression, cell growth and differentiation. Among the myriad post-translational histone modifications, reversible methylation, acetylation and ubiquitination of specific Lys residues in histone tails are pre-eminent in impacting epigenetic states[3]. In some cases, there appears to be functional crosstalk between specific acetylation or methylation histone marks that have been referred to as the histone code. Well-known examples of such crosstalk are histone H2B Lys120 ubiquitination promoting histone H3 Lys79 methylation[4,5] as well as H3 Lys4 methylation (H3K4me)[6,7]. Positive interplay between H3K4me and H3 acetyl marks appears to reciprocally stimulate their respective writers[8–12], recruit selective chromatin regulators[13–15] or repel their erasers[16–19] to fine-tune gene expression.

Lysine-specific demethylase 1 (LSD1) is a histone demethylase that specifically removes methyl groups from mono-methylated or di-methylated Lys4 from histone H3 (H3K4me1 and H3K4me2). These histone marks are typically linked with transcriptionally active genes and functional enhancers[20]. LSD1 plays a role in gene silencing[21] that

[1]Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA. [2]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA, USA. [3]Department of Biology and Biotechnology Lazzaro Spallanzani, University of Pavia, Pavia, Italy. [4]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA, USA. [5]Broad Institute of Harvard and MIT, Cambridge, MA, USA. [6]Department of Genetics, Harvard Medical School, Boston, MA, USA. [7]Department of Dermatology, Boston University School of Medicine & Boston Medical Center, Boston, MA, USA. [8]These authors contributed equally: Kwangwoon Lee, Marco Barone. ✉e-mail: sambeck@bu.edu; andrea.mattevi@unipv.it; pacole@bwh.harvard.edu

impacts various cellular processes, such as epithelial-to-mesenchymal transition, proliferation and cell motility[22–24]. Notably, LSD1 has emerged as a promising therapeutic target for hematological malignancies, which has spurred the development of selective cyclopropylamine inhibitors, such as GSK2879552 (ref. 25). In 2018, it was reported that Lys14 acetylation of histone H3 (H3K14ac) inhibits demethylation of H3 Lys4 (H3K4me2) by LSD1 (ref. 19), revealing negative crosstalk between two histone marks. In cells, LSD1 is constitutively bound to the scaffolding protein CoREST, and this multi-protein complex sometimes also includes the histone deacetylase HDAC1 (refs. 26,27), another enzyme that can repress gene expression. The LSD1–HDAC1–CoREST (LHC) complex was shown to be efficient at deacetylating a wide range of acetyl-Lys on histone H3 and histone H2B modified nucleosomes except for H3K14ac, which was resistant to deacetylation[19,28,29]. Of note, other multi-protein HDAC1-containing complexes, such as MIDAC, SIN3 and NuRD, lack such sharp selectivity in rejecting H3K14ac versus other nucleosome acetyl-Lys sites[19,28,29]. Thus, there appears to be a selective reciprocal antagonism between the demethylase and deacetylase activities within LHC through both direct and indirect effects of the H3K4me–H3K14ac crosstalk.

It is conceivable that the co-occurrence of H3K4me and H3K14ac shields key genes from being inappropriately silenced by LSD1. Historically, dissecting the specific biological functions of histone modification relationships associated with reader or eraser enzymes has been difficult because of the many cellular isoforms of histone H3 in humans and the lack of precision tools to address the challenge.

Structural analysis of the LSD1–CoREST1 complex lacking HDAC1 (LC) bound to nucleosomes revealed that the SANT2 domain of CoREST1 contacts the histone octamer and core DNA, and a surface of LSD1 binds to extended linker DNA (Extended Data Fig. 2a)[30]. An additional dimension of complexity arose from a low-resolution structure of LHC trapped with nucleosomes using a propargyl warhead on H3, which showed a very distinct mode of nucleosome–LHC interaction[31]. LHC is often recruited to chromatin by SNAIL/Gfi-1 (SNAG) family transcription factors that appear to mimic H3 tail binding to the LSD1 substrate interaction site[23,24,32,33]. Because of this competitive mode of interaction, most cyclopropylamine LSD1 inhibitors, such as GSK-LSD1, dislodge SNAG family transcription factors from binding to LSD1 while impairing catalysis[33–36]. As such, LSD1 inhibitors have not been able to accurately discriminate between an enzymatic function and an adapter function for LSD1. It is, thus, possible that LHC is primarily important in deacetylating nucleosomes after recruitment to nucleosomes by SNAG family transcription factors, whereas LC could serve as the principal demethylase form of LSD1, but this has not been straightforward to address. Teasing apart these roles will greatly help clarify various newly discovered biological roles of LSD1 (refs. 37,38).

In the present study, we analyzed the demethylase activities of LC versus LHC toward H3K4me2-containing nucleosomes and obtained evidence that LC is a much more powerful demethylase complex than LHC. Based on this, we evaluated H3K14ac's impact on H3K4me2 nucleosome demethylation by LC and confirmed strong suppression of demethylase activity as observed previously with peptide substrates. We then screened for LSD1 mutants that render LC nucleosome demethylation activity insensitive to H3K14ac and identified one such mutant, Y391K LSD1. This Y391K mutant was knocked into K562 cells using CRISPR–Cas9, and the effects on gene expression and histone marks were evaluated and are described below.

## Results

### LC versus LHC in nucleosome demethylation

Previous studies suggested that LC was highly efficient at demethylating modified nucleosomes containing extended DNA linker segments[30,39]. However, these studies employed methyl-thiaLys-modified mimics at the H3K4 position rather than natural methyl-Lys, which could influence the demethylation rates[40]. Consequently, we prepared authentic H3K4me2-modified 185-bp nucleosomes using semi-synthetic histone H3 prepared with an engineered sortase to examine as substrates of purified LC and LHC complexes[19]. Remarkably, LHC was much slower at demethylating H3K4me2 nucleosomes compared with LC (Fig. 1a–c and Extended Data Fig. 1a–e). To gain more insight into this rate difference, we adapted microscale thermophoresis (MST) to measure the affinities of LC and LHC with unmodified 185-bp nucleosomes. To enable MST analysis, we prepared LC and LHC in fluorescent forms by N-terminally labeling the CoREST1 subunits via chemoselective ligation with fluorescein (Fig. 1d)[41,42]. These fluorescein-tagged LC and LHC showed similar enzymatic activities to the unlabeled versions, indicating that the fluorescein modification procedure did not perturb their structures (Extended Data Fig. 2c). MST with LC showed relatively tight binding to unmodified 185-bp nucleosomes with $K_D$ of 101 nM, similar to the previously reported affinity[39]. In contrast, the binding of LHC to unmodified 185-bp nucleosomes by MST was very weak ($K_D$ >12 μM) (Fig. 1e). This drastic contrast in affinity correlates with the difference in demethylase activity on nucleosomes containing H3K4me2. Moreover, previous studies showed that LHC and LSD1 alone show similar demethylation rates with H3K4me2 peptide substrates, supporting the notion that the rate differences with nucleosome substrates correlate with the weaker binding affinity of LHC[19].

We recently used a combination of the electron microscopy (EM) density map of LHC (EMD-10629)[31] and AlphaFold2 to model how HDAC1 interacts with CoREST1 (Extended Data Fig. 2b)[43]. This model shows key interactions between the CoREST1 SANT2 domain and HDAC1. Based on an X-ray structure, the SANT2 domain of CoREST1 in the LC complex also engages the nucleosome core region (Protein Data Bank (PDB): 6VYP)[39] (Extended Data Fig. 2a). It, thus, appears that LC's primary mode of interaction with nucleosomes is unavailable to LHC. It is important to note that different LC and LHC constructs were used in our measurements (LC: LSD1 amino acids (aa) 171–852 and CoREST1 aa 286–482 expressed in bacteria; LHC: full-length LSD1, full-length HDAC1 and CoREST1 aa 84–482 expressed in HEK293F cells). This choice was made to ensure optimal stability of the protein complexes. Although we speculate that the large difference in the nucleosome binding affinity between LC and LHC is attributed to HDAC1 binding, we cannot dismiss the possibility that the N-terminal portion of CoREST1 (aa 84–285; segment absent in our LC complex) hampers nucleosome recognition. Conversely, we contend that the N-terminal region of LSD1 (aa 1–170; segment absent in the LC complex in our studies) does not impact nucleosome recognition by the LC complex, as indicated by previous studies[44,45].

Considering LHC's weak nucleosome binding affinity, we hypothesize that LHC's demethylase mode might be activated only under specific conditions. The low-resolution EM density of the chemically linked LHC–nucleosome complex (EMD-10630) illustrates LHC's capability to recognize the nucleosome using the LSD1 active site, bypassing the need for the SANT2 domain of CoREST1 to interact with the nucleosome core[31]. However, based on the weak binding affinity and low demethylase activity, this conformation of LHC seems disfavored unless covalently attached. Therefore, it is plausible that LHC may necessitate nucleosome-binding interaction partners and/or post-translational modifications (PTMs) to activate its demethylase activity. Our over-expression system may have limited acquisition of these features. Recent proximity-dependent BioID studies identified numerous potential binding partners of LHC[46], suggesting the possibility of LHC's nucleosome-binding affinity enhancement with the help of binding partners. Alternatively, exchanging CoREST1 with CoREST2, which disfavors the binding of HDAC1/2, can also activate the demethylase function of the CoREST complex[47].

In our demethylase-focused study, we chose to use LC as a primary tool. This allows us to carefully examine the regulatory elements of the CoREST complex and to help explore potential engineering interventions to refine substrate sensitivity.

https://doi.org/10.1038/s41589-024-01671-9

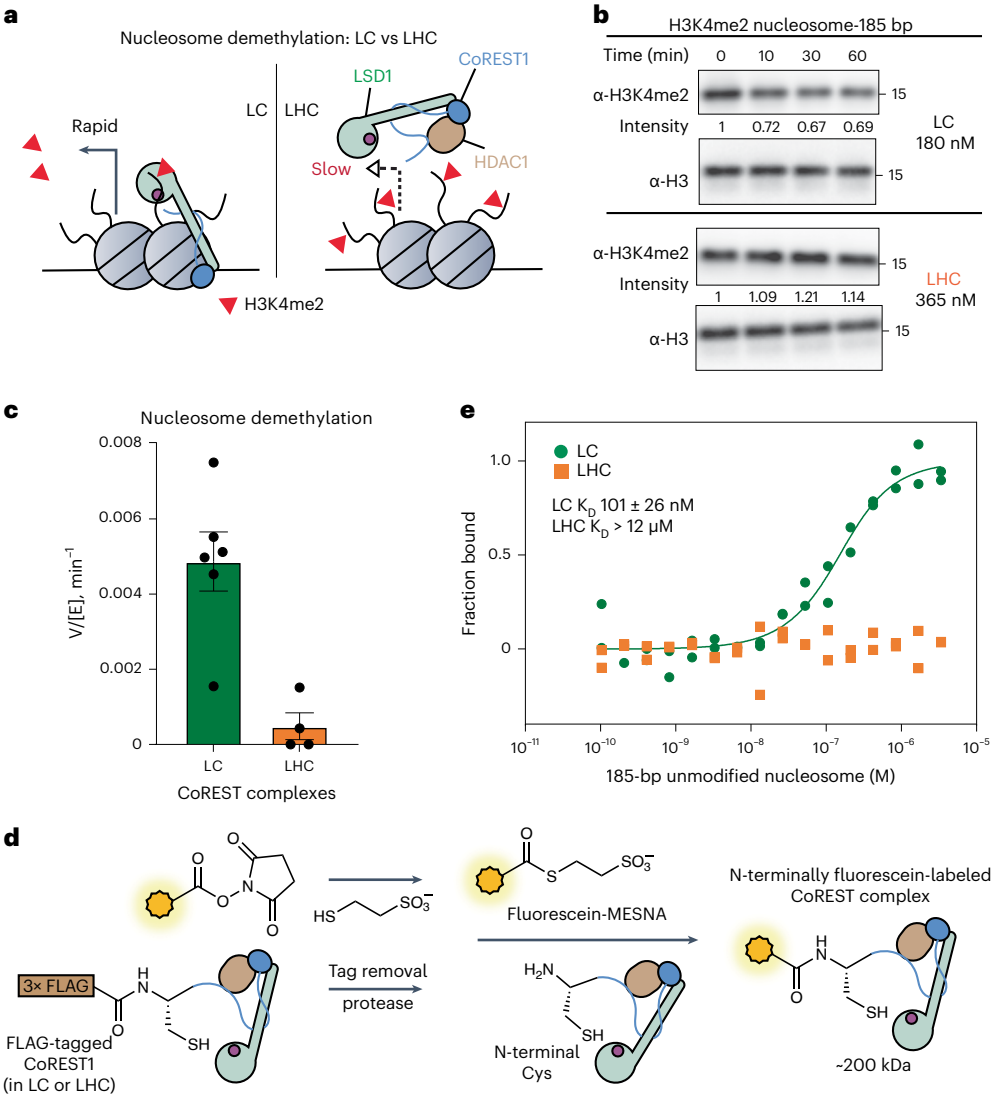



**Fig. 1 | LC and LHC complexes have different nucleosome demethylase activities. a**, The illustration highlights LC's ability to demethylate nucleosomes rapidly and LHC's slower nucleosome demethylation. Striped circles indicate nucleosomes, and the small purple circle on LSD1 highlights the active site. **b**, Western blot–based assays were employed to extrapolate the demethylase activity of LC (top) and LHC (bottom) on H3K4me2 nucleosomes. The anti-H3K4me2 signal at each timepoint was normalized by anti-H3. We consistently observed that, after $T_{30min}$, the demethylase activity gets very slow, possibly caused by product inhibition. **c**, The demethylase activities of LC and LHC are shown in the bar plot (mean ± s.e.m.; *n* for LC = 6 and *n* for LHC = 4). **d**, Site-specific fluorophore labeling of LC/LHC was achieved using chemoselective ligation of NHS-ester fluorescein to an N-terminal Cys of CoREST1. **e**, MST was used to measure the binding affinities of N-terminally fluorescein-labeled LC and LHC to the nucleosome (185 bp, unmodified histones). $K_D$ values and the standard errors are shown. LC exhibited strong engagement with the nucleosome in the nanomolar range, whereas LHC's $K_D$ toward the nucleosome was >12 μM. Technical duplicates of the MST measurements were performed (*n* = 2).

## Structural analysis of LC with H3K14ac peptide

The fact that LC may serve as the principal histone-demethylating form physiologically led to our focus on this complex to analyze the inhibitory role of $K14_{H3}$ acetylation. To understand the structural basis of why Lys14 acetylation in histone H3 results in diminished demethylation of H3K4me2, we obtained an X-ray crystal structure of LC bound to a synthetic H3 tail peptide (aa 1–21) containing K4M and K14ac. It was previously shown that a methionine replacing $K4_{H3}$ enhances binding affinity to LSD1, presumably by mimicking the neutral amine form of Lys, which would be stabilized by the active site to enhance catalysis[48]. A prior X-ray crystal structure of unacetylated K4M H3 tail peptide revealed an electrostatic interaction between the sidechain of $K14_{H3}$ and that of $E559_{LSD1}$ (ref. 48). In contrast, the crystal structure of K4M/K14ac H3 tail peptide revealed that the acetylated $K14_{H3}$ sidechain appears to pull away from $E559_{LSD1}$ (Fig. 2a). Interestingly, the sidechain of $K9_{H3}$ moves closer to $E559_{LSD1}$ in an apparent attempt to substitute for the $K14_{H3}$–$E559_{LSD1}$ salt bridge. This $K9_{H3}$ movement toward $E559_{LSD1}$ in the acetylated peptide causes $K9_{H3}$ to move away from $H564_{LSD1}$, which otherwise engages in polar contacts with both $T6_{H3}$ and $K9_{H3}$. We propose that disruption of these polar interactions negatively influences the enzymatic activity, as H564A mutation is deleterious to the demethylase activity (Fig. 2b). Supporting this idea, earlier structural studies of an alternative LC-H3 peptide complex, wherein propargylated $K4_{H3}$ and $FAD_{LSD1}$ are chemically linked to mimic a transition state analog (PDB: 2UXN)[49], showed a potential hydrogen bond between the imidazole ring of $H564_{LSD1}$ and the backbone carbonyl oxygen atom of $K4_{H3}$. This interaction may influence the catalytic step by controlling the sidechain orientation of H3K4me toward FAD. Thus, we speculate that the $H564_{LSD1}$–$K9_{H3}$ polar interaction may fine-tune the orientation and dynamics of the $H564_{LSD1}$ sidechain that can form hydrogen bonds

https://doi.org/10.1038/s41589-024-01671-9

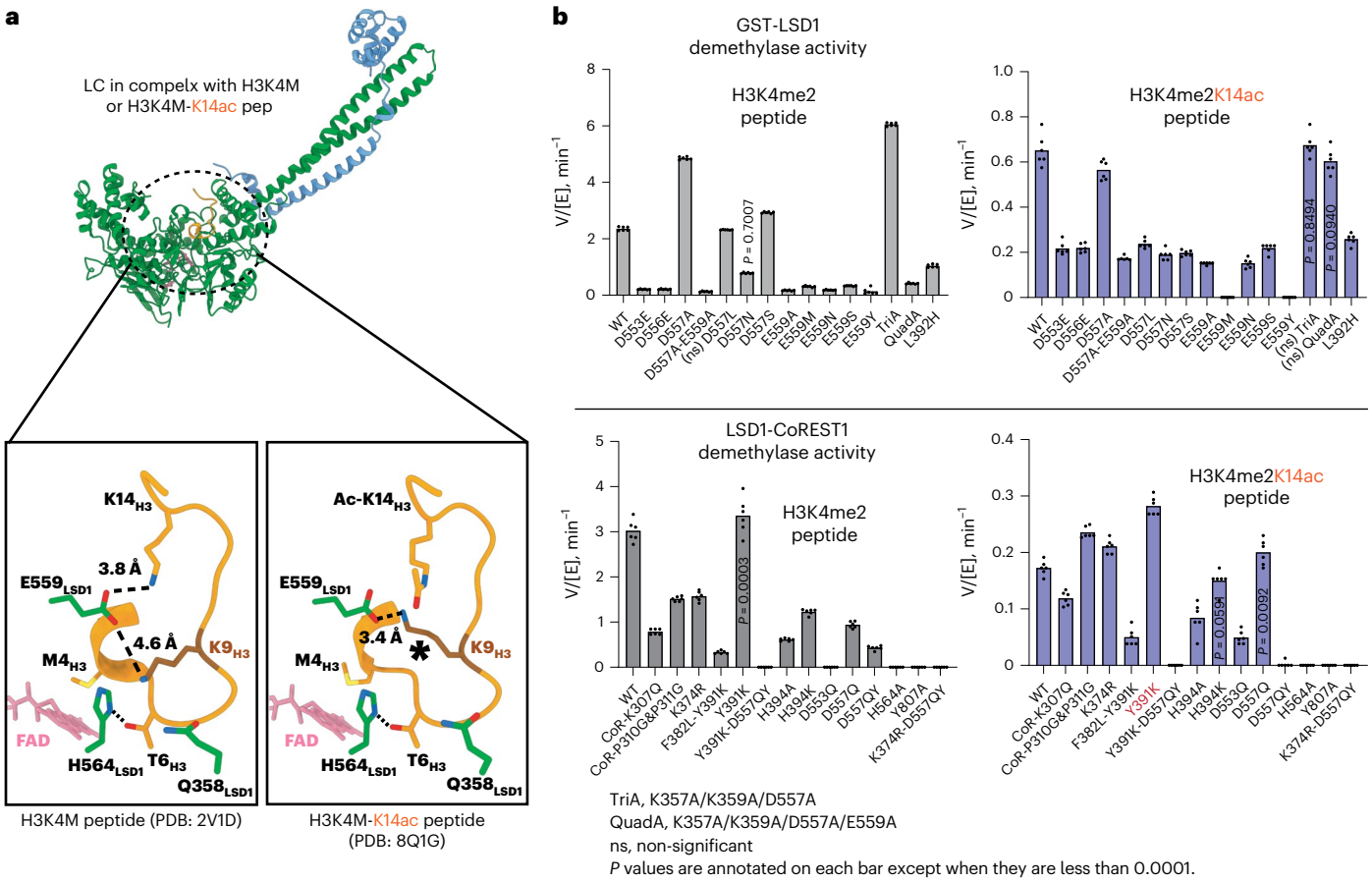



**Fig. 2 | Impact of H3K14ac on LSD1's demethylase activity. a**, Crystal structures of LC in complex with H3K4M (left) and H3K4M/K14ac (right) peptides are compared. Both peptides contain a K4M mutation to mimic methylated Lys. In H3K4M, K9 of H3 (dark brown) is situated close to H564 and Q358 of LSD1, and K14 of H3 forms a direct electrostatic contact with E559 of LSD1. However, in H3K4M/K14ac, K9 forms a compensatory salt bridge with E559 due to the charge neutralization of K14, resulting in H564 and Q358 being unoccupied. **b**, Mutant activities were evaluated using an HRP-coupled peptide demethylase assay with GST-LSD1 mutants (top) and H3K4me2/H3K4me2K14ac peptides (aa 1–21). Additional peptide assay screenings were conducted using the LC binary complex (bottom) to consider the CoREST1 binding effect in certain mutants. V/[E] ($min^{-1}$) values were derived from six data points obtained through duplicate continuous assays, and the bars in the graph depict the mean values. One-way ANOVA analysis with Dunnett's multiple comparisons test, with a single pooled variance, was employed to compare mean values across LSD1 mutants. *P* values are indicated on the bars, except for groups exhibiting $P < 0.0001$. 'ns' signifies groups with $P > 0.05$.

with nearby H3 resides, thereby facilitating H3K4me demethylation. This notion aligns with earlier findings that suggest a downregulatory role of $T6_{H3}$ phosphorylation (H3T6ph) and $K9_{H3}$ acetylation (H3K9ac) in LSD1 demethylase activity[16,19,50].

## Y391K LSD1 as a H3K4me/K14ac nucleosome demethylase

To identify LSD1 mutants that might show enhanced H3K4me demethylase activity in substrates containing H3K14ac modification, we visually inspected several regions of LSD1 using Pymol (version 1.2r3pre, Schrödinger) and ChimeraX[51] (Supplementary Fig. 1). Several areas emerged as potential targets for mutagenesis: Area 1 centered around $E559_{LSD1}$; Area 2 involving $K357_{LSD1}$ and $Lys359_{LSD1}$; Area 3 involving $K374_{LSD1}$; Area 4 in the αC helix; Area 5 somewhat distal to the substrate binding site; and Area 6 on CoREST1 at the LSD binding site. The following hypotheses guided the selections of these mutations: (1) mutants in Area 1 could more strongly interact with H3K14ac; (2) mutants in Areas 2, 3 and 4 might remodel or alter the substrate binding pocket, especially since the Lys residues in Areas 2 and 3 have been noted as key regulatory residues for the substrate interaction[52,53]; (3) mutants in Area 5 at a slightly distal region from the H3 binding site were projected to potentially promote LSD1 binding affinity to H3K14ac by gaining +1 charge lost by the Lys acetylation; and (4) mutants in Area 6 were predicted to locally disrupt the LSD1–CoREST1 binding interface to facilitate entry of the H3 substrate with a longer H3K14ac sidechain.

A total of 29 LSD1 mutants were screened for demethylase activity with H3K4me2 tail peptides ± $K14_{H3}$ acetylation. We opted to make mutations in both LSD1 alone and the LC complex to gain insights into how these mutations affect substrate and CoREST1 interactions to overcome H3K14ac. Most mutants showed diminished catalytic activity with one or both peptide substrates (Fig. 2b). Y391K LSD1 in the LC complex stood out, showing a modestly enhanced demethylase activity with H3K14ac relative to wild-type (WT) LC (Fig. 2b). This prompted us to test Y391K with H3K4me2 nucleosomes ± H3K14ac as well as with H3K4me1 nucleosomes ± H3K14ac. Y391K and WT LC showed similar and relatively strong demethylase activity with non-acetylated H3K4me2 and H3K4me1 nucleosomes (Fig. 3a–b and Extended Data Figs. 3a–i and 4a–k). As expected, whereas WT LC showed sharply reduced demethylase activity with H3K4me2/K14ac and H3K4me1/K14ac nucleosomes, Y391K LC demonstrated robust demethylase activity with H3K4me2/K14ac and H3K4me1/K14ac nucleosomes, close to the rate of WT and Y391K LC with non-acetylated nucleosomes (Fig. 3a,b and Extended Data Fig. 3a–i). This behavior in which Y391K LC's enzymatic activity appeared relatively insensitive to H3K14ac in

https://doi.org/10.1038/s41589-024-01671-9

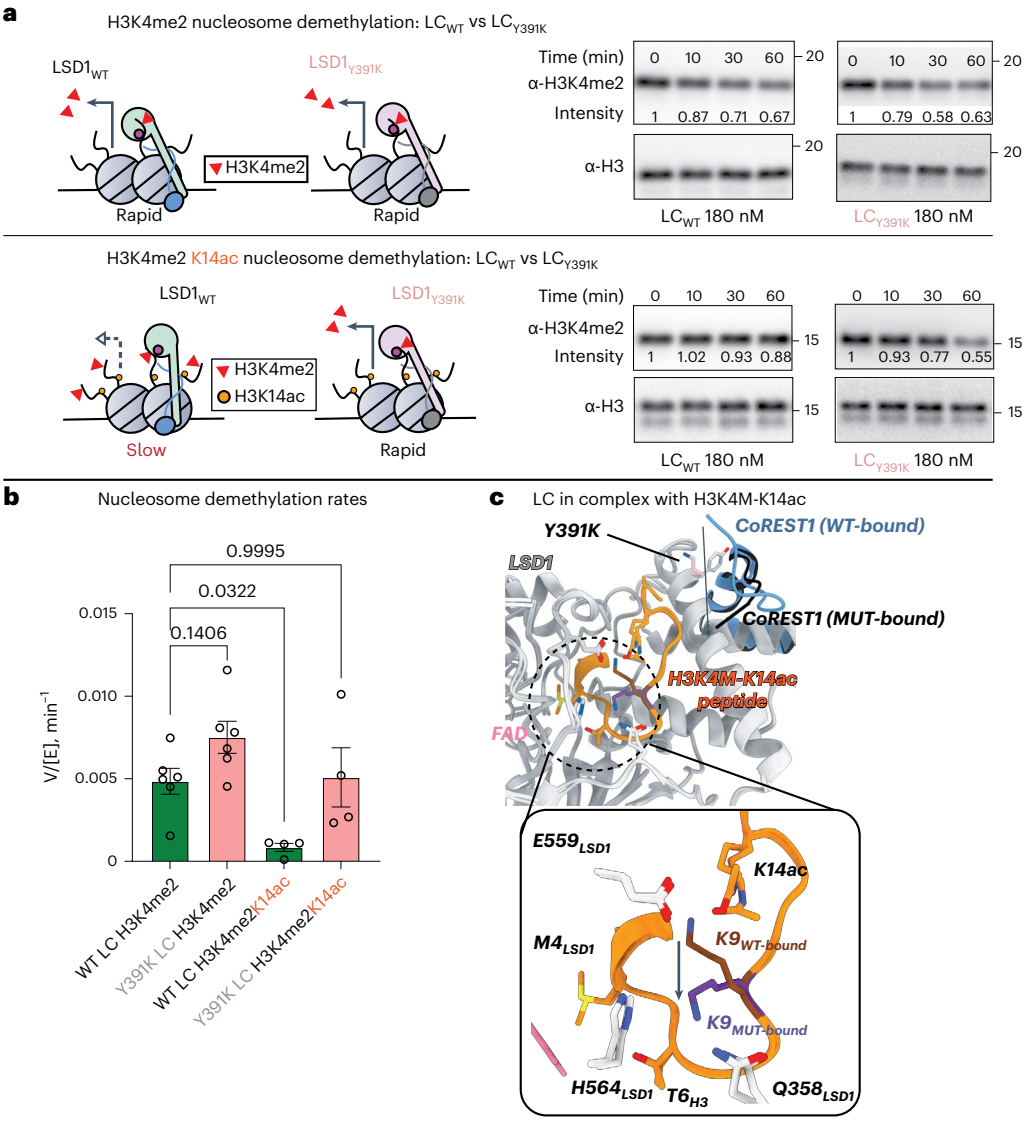



**Fig. 3 | Y391K LSD1 as an H3K14ac-agnostic demethylase. a**, Y391K LC complex's nucleosome demethylation activities are compared with those of the WT LC complex. Unlike the slow demethylation observed in the WT LC complex on H3K4me2/K14ac nucleosome, the Y391K LC complex can demethylate the H3K4me2/K14ac nucleosome at a similar rate to that of the H3K4me2 nucleosome. Representative images of the western blot–based demethylation of H3K4me2 (top right) and H3K4me2/K14ac nucleosomes (bottom right) by WT LC (left) and Y391K LC (right) are shown. Replicate 3 for WT on H3K4me2, replicate 1 for WT on H3K4me2K14ac, replicate 1 for Y391K on H3K4me2 and replicate 4 for Y391K on H3K4me2K14ac are selected. Images for the whole replicates are shown in Extended Data Figs. 1a and 3a–c. A total of six technical replicates on H3K4me2 nucleosomes and four technical replicates on H3K4me2K14ac nucleosomes were performed. **b**, The bar plot (bottom) demonstrates the demethylase activities of WT LC and Y391K LC toward H3K4me2 and H3K4me2/K14ac nucleosomes (two-way ANOVA with Dunnett's multiple comparisons test, with a single pooled variance; mean ± s.e.m.; $n = 6$ for H3K4me2 nucleosome and $n = 4$ for H3K4me2/K14ac nucleosome; $P$ values are indicated above each comparison group). Data points for WT LC toward H3K4me2 were referenced from Fig. 1c and Extended Data Figs. 1a and 3a–c. **c**, Crystal structure snapshot of WT and Y391K LC in complex with the H3K4M/K14ac peptide. Y391K LSD1-bound K9 ('K9$_{MUT-bound}$') restores its interaction with H564 of LSD1. Y391K point mutation also induces a conformational change in CoREST1 at the LSD1 binding interface, driven by the charge repulsion.

the setting of nucleosomes was what we were hoping to achieve by LSD1 engineering (Fig. 3a).

As Y391$_{LSD1}$ is rather remote from the active site of LSD1, we performed additional experiments to gain greater insight into how Y391K$_{LSD1}$ was able to overcome the effects of H3K14ac on LSD1 nucleosome demethylation. Using MST, we measured the dissociation constants of WT and Y391K LC with H3K14ac nucleosomes and found that the binding affinities were similar and strong ($K_D$ values 35 nM ± 17 nM and ~18 nM ± 3 nM for WT and Y391K LC, respectively; Extended Data Fig. 3h). These results indicate that Y391K LC's increased demethylase activity on H3K14ac nucleosomes relative to that of WT LC is not acquired by tighter substrate binding.

We then determined the X-ray crystal structure of Y391K LC in complex with K4M/K14ac H3 tail peptide. Of note, the orientation of the K9$_{H3}$ sidechain, which was perturbed in the presence of the H3K4M/K14ac peptide in the WT LC complex, has now been restored with the Y391K LC complex (Fig. 3c and Extended Data Fig. 5). Although the precise reason for this realignment in the Y391K LC complex is not completely clear, we observed that the sidechain of K391$_{LSD1}$ of LSD1 Y391K has rotated relative to Y391 and impinges on a CoREST1 basic patch, causing reconfiguration of CoREST1 basic residues R305$_{CoREST1}$, K307$_{CoREST1}$, R308$_{CoREST1}$, K309$_{CoREST1}$ and K312$_{CoREST1}$ at the LSD1-binding interface (Fig. 3c and Extended Data Fig. 5) that appears to trigger the reorientation of K9 of histone H3 and augment LSD1 demethylase activity in the

presence of $K14_{H3}$ acetylation. This impact may be more accentuated in the context of an H3K14ac nucleosome substrate, which displayed a more marked effect of Y391K stimulating demethylase activity. When we directly overlay the crystal structure of the Y391K LC and H3K14ac complex onto the pre-existing LC nucleosome structure (PDB: 6VYP)[30], it is evident that K312 of CoREST1 relocates approximately 11.5 Å closer to the linker DNA (Supplementary Fig. 3). This shift, in turn, could stabilize CoREST1's basic patch (R305, K307, R308 and K309) that promotes H3K9 to adopt a more catalytically favorable orientation.

Aside from the conformational differences in CoREST1 and the sidechain orientations of H3K9, the overall LSD1 protein structures are similar among all states (WT LC-H3K4M, WT LC-H3K4M/K14ac, Y391K LC-H3K4M and Y391K LC-H3K4M/K14ac) with the small Ca root mean square deviation (RMSD) values ranging between 0.27 Å and 0.36 Å. Consistent with the structural similarities, both WT and Y391K LC showed similar binding affinities toward the H3K4M and H3K4M/K14ac peptides as well as the SNAG domain peptide from SNAIL (Supplementary Fig. 4), suggesting that WT and Y391K share similar architecture and modes of recognition toward catalytic and non-catalytic targets.

Collectively, our comparative structural and enzymological studies with LSD1 mutants point to potential crosstalk between the CoREST1–LSD1 binding interface and the H3–LSD1 binding interface, influencing LSD1's sensitivity to H3K14ac. This crosstalk seems most pronounced when LC interacts with the nucleosome. In this scenario, when CoREST1's basic residues near the LSD1 binding interface directly engage the nucleosomal DNA, their conformational dynamics may undergo substantial changes, favoring $E559_{LSD1}$–$K9_{H3}$ interaction mediated by H3K14ac. This will strongly impede demethylation steps by disrupting the $H564_{LSD1}$–$K9_{H3}$ interaction (greater than 5-fold activity reduction; Extended Data Figs. 3 and 4), despite the tight binding between LC and H3K14ac nucleosome (Extended Data Fig. 3h). This potential downregulatory synergy in the nucleosome context appears to be less pronounced when LC recognizes peptide substrates. In this case, both the peptides and CoREST1 residues at the LSD1 binding interface may gain conformational flexibility. This increased flexibility may lead to a more frequent restoration of the $H564_{LSD1}$–$K9_{H3}$ interaction, allowing $H564_{LSD1}$ to adopt orientation and dynamics more optimal for demethylation than that of the nucleosome substrate.

We additionally evaluated the impact of Y391K on LHC's deacetylase activity. It was previously reported that LHC shows strong activity toward H3K9ac nucleosome substrate[19,28,29,54], and we employed H3K9ac nucleosome as substrate in our deacetylase assays here. As expected, WT and Y391K LHC showed similar deacetylase activities toward H3K9ac nucleosomes (Extended Data Fig. 6a–e), indicating that the Y391K LSD1 mutation does not appear to influence the HDAC1 subunit of LHC. This suggests that the potential cellular effects of Y391K LSD1 mutation are likely to be mediated by the demethylase functions of LC.

### Introducing Y391K LSD1 in K562 cells via CRISPR knockin

Given the desired nucleosome demethylase properties of Y391K LC in which Lys14 acetylation appeared no longer inhibitory, we proceeded to carry out genome editing to introduce this point mutant into a human cell line. We selected a K562 chronic myeloid leukemia cell line that was previously shown to tolerate LSD1 inhibition[33]. Green fluorescent protein (GFP) was inserted at the C-terminus of LSD1 to aid in cell selection during editing. Using designed single guide RNAs (sgRNAs) for the desired allele, CRISPR–Cas9 was employed to furnish the Y391K LSD1 allele knocked into two of three alleles of this triploid cancer cell line (Fig. 4a). As Y391K is anticipated to be a gain-of-function allele, we considered the heterozygous knockin with two edited alleles out of three in the triploid K562 cells sufficient for further analysis. These Y391K LSD1 knockin cells (subsequently referred to as 'edited K562' cells) had similar proliferation rates and LSD1 protein levels compared with 'parental K562' cells (Fig. 4a–c). To explore other phenotypic differences arising from the Y391K LSD1-mediated transcriptional reprogramming, we then analyzed the anti-proliferative pharmacological impacts on the 'edited K562' cells of the LSD1 inhibitor GSK2879552; the Abl tyrosine kinase inhibitor imatinib; the broad spectrum HDAC inhibitor SAHA; and the dual LSD1/HDAC1 inhibitor corin. Corin was previously found to show selectivity for HDAC1 inhibition of the LHC deacetylase complex[55]. Each of the inhibitors shared similar efficacy and potency on the proliferation of 'edited K562' cells versus 'parental K562' cells except for corin, which displayed a modestly greater potency (~2.5-fold) in slowing the proliferation of 'edited K562' cells (Fig. 4d–g). GSK2879552 LSD1 inhibitor did not appreciably slow the growth of either cell type, consistent with the tolerance of LSD1 inhibition in K562 cells. It appears that the presence of Y391K LSD1 may confer mildly enhanced cellular sensitivity to the dual inhibitor, but, in general, the cell proliferation properties of 'edited K562' cells are similar to those of 'parental K562' cells.

### Transcriptomic data analysis of Y391K LSD1 K562 cells

To understand how Y391K LSD1 knockin could affect gene expression, we performed RNA sequencing (RNA-seq) on the 'edited K562' cells and compared the RNA transcript levels to 'parental K562' cells. These results revealed across two clonal populations that there were 498 downregulated genes based on a two-fold mRNA reduction with $P < 0.05$. There were also 256 upregulated genes identified using the same criteria (Fig. 4h). These results are consistent with enhanced gene silencing associated with elevated gain-of-function demethylase activity of the Y391K LSD1 enzyme. It is worth noting that many of these differentially expressed genes in the 'edited K562' cells had a low baseline expression level (Supplementary Fig. 5). Analysis of the 498 genes with reduced expression using clusterProfiler enrichGO[56] indicated that the most prominent functional set are positive regulators of cell adhesion and myeloid leukocyte activation (Fig. 4i). Interestingly, 13 downregulated genes show an inversely correlated expression pattern with genes upregulated in K562 cells treated with the ORY-1001 LSD1 inhibitor (227 upregulated genes exhibiting >1.5-fold mRNA increase with $P < 0.05$)[57]. Among these genes are those involved in leukocyte differentiation, migration or adhesion (for example, *CD99* encoding a CD99 cell surface glycoprotein[58], *LGMN* encoding an LGMN cysteine protease[59] and *ARHGEF3* encoding an ARHGEF3 Rho-like GTPase[60]); leukemic proliferation (*STAP1* encoding a STAP1 TEC tyrosine protein kinase activator)[61], imatinib resistance in chronic myeloid leukemia (CML) (*PRG2* encoding a PRG2 eosinophil major basic protein)[62], leukemic stem cell maintenance (*TSPAN18* encoding a TSPAN18 membrane protein)[63], cell division (*TLE6* encoding a TLE6 transcriptional co-repressor)[64], Wnt signaling (*DLK1* encoding a DLK1 transmembrane protein)[65,66], apoptosis (*TMEM14A* encoding a TMEM14A mitochondrial membrane protein)[67], and pH regulation (*HVCN1* encoding an HVCN1 hydrogen voltage-gated channel)[68]. Collectively, based on prior work investigating LSD1's co-repressor activity and the transcriptomics analysis, it can be inferred that the Y391K LSD1 knockin results in 'on-target' enhanced gene repression.

### Chromatin analysis of Y391K LSD1 K562 cells

To investigate how Y391K LSD1 knockin may have influenced gene expression, we performed CUT&RUN[69] to analyze the positions of LSD1, H3K14ac, H3K4me1 and H3K4me2 in 'parental K562' and 'edited K562' cells. It is first worth noting that LSD1 from both parental and edited K562 cells had similar genomic distribution within or near the downregulated genes (Fig. 5a) and shared approximately 40% direct peak overlap with the LSD1 chromatin immunoprecipitation followed by sequencing (ChIP-seq) peaks from the ENCODE database (Extended Data Fig. 8d). About half of the LSD1 peaks in the 'parental K562' cells showed direct overlap with the LSD1 peaks in the 'edited K562' cells. Moreover, approximately 80% of the LSD1 peaks that did not overlap were found in non-promoter regions associated with LSD1-bound genes

https://doi.org/10.1038/s41589-024-01671-9

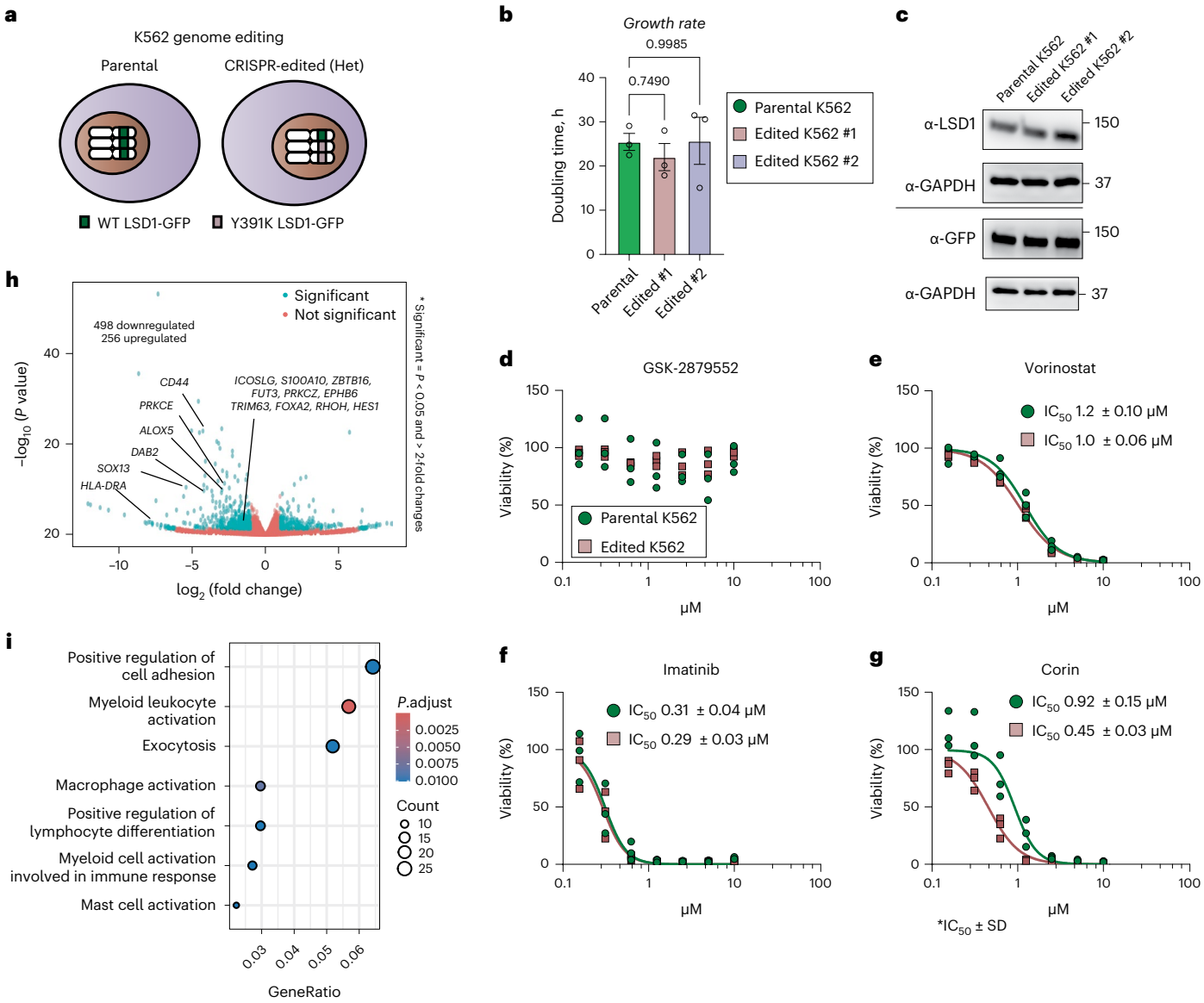



**Fig. 4 | Y391K LSD1 gene-edited K562 cells. a**, 'Parental K562' cells (left) and heterozygously Y391K LSD1-edited K562 cells ('edited K562', right) (C-terminal GFP included). 'Edited K562' cells contain Y391K LSD1 mutation in two out of three chromosomes. **b**, Doubling time comparison for parental and edited K562 cells ($n = 3$). The error bars represent s.e.m. *P* values shown are from one-way ANOVA analysis with Dunnett's multiple comparisons test. **c**, Western blots of LSD1 in parental and edited K562 cells. A total of 13 μg of cell lysate from one biological replicate ($n = 1$) for parental K562, edited K562 #1 and edited K562 #2 cells was loaded on two gels and analyzed using anti-LSD1, anti-GAPDH and anti-GFP antibodies. **d**–**g**, Drug sensitivity assays for parental and edited K562 cells, measured by Alamar blue fluorescence 4 d after initial treatment. $IC_{50}$ values (mean ± s.e.m.) are shown. Green circles indicate 'parental K562' cells, and pink squares indicate 'edited K562' cells. **d**, GSK-2879552 (LSD1 inhibitor) effects. **e**, Vorinostat (pan-HDAC1/2 inhibitor) effects. **f**, Imatinib (Abl kinase inhibitor) effects. **g**, Corin (LSD1–HDAC1 dual inhibitor) effects. **h**, Differential gene expression analysis between parental and edited K562 cells using RNA-seq. In total, 498 downregulated genes and 256 upregulated genes were identified in the edited K562 cells compared with the parental. Statistical analysis was conducted using DESeq2, applying a two-sided statistical test with adjustments made for multiple comparisons. **i**, Gene Ontology analysis of the downregulated genes from the transcriptomic analysis. Using clusterProfiler's enrichGO, over-representation analysis was performed with hypergeometric distribution to calculate the *P* values. The resulting *P* values were further adjusted for multiple comparisons.

(Extended Data Fig. 8d). Considering the large overlap in LSD1-bound genes between parental and edited K562 cells (Fig. 6a), it appears plausible that the introduction of Y391K LSD1 triggered alterations in the occupancy of these shared genes. However, because LSD1 in the 'edited K562' cells maintains a similar genomic distribution pattern across downregulated genes to that of 'parental K562' cells (Fig. 5a), we posit that the sole occupancy of LSD1 at specific loci is not a predominant factor driving stronger gene repression.

To further explore, we compared CUT&RUN signals involving LSD1 and histone marks between parental and edited K562 cells while focusing on (1) the 498 downregulated genes and their 5′ and 3′ flanking regions (±20 kb) compared with analogous regions of 498 randomly selected unaffected control genes and (2) LSD1 peaks across these regions and distal regulatory regions (±5 kb) identified by ChIPseeker[70]. In the 'parental K562' cells, average LSD1 levels were lower in the transcriptional start and end sites in the 498 downregulated genes compared with the 498 unaffected control genes (Extended Data Fig. 7a), whereas H3K14ac marks appeared to be moderately higher in the gene body regions of the 498 downregulated versus the unaffected control genes (Fig. 5b). Consistently, H3K14ac levels in 'parental K562' cells near

https://doi.org/10.1038/s41589-024-01671-9

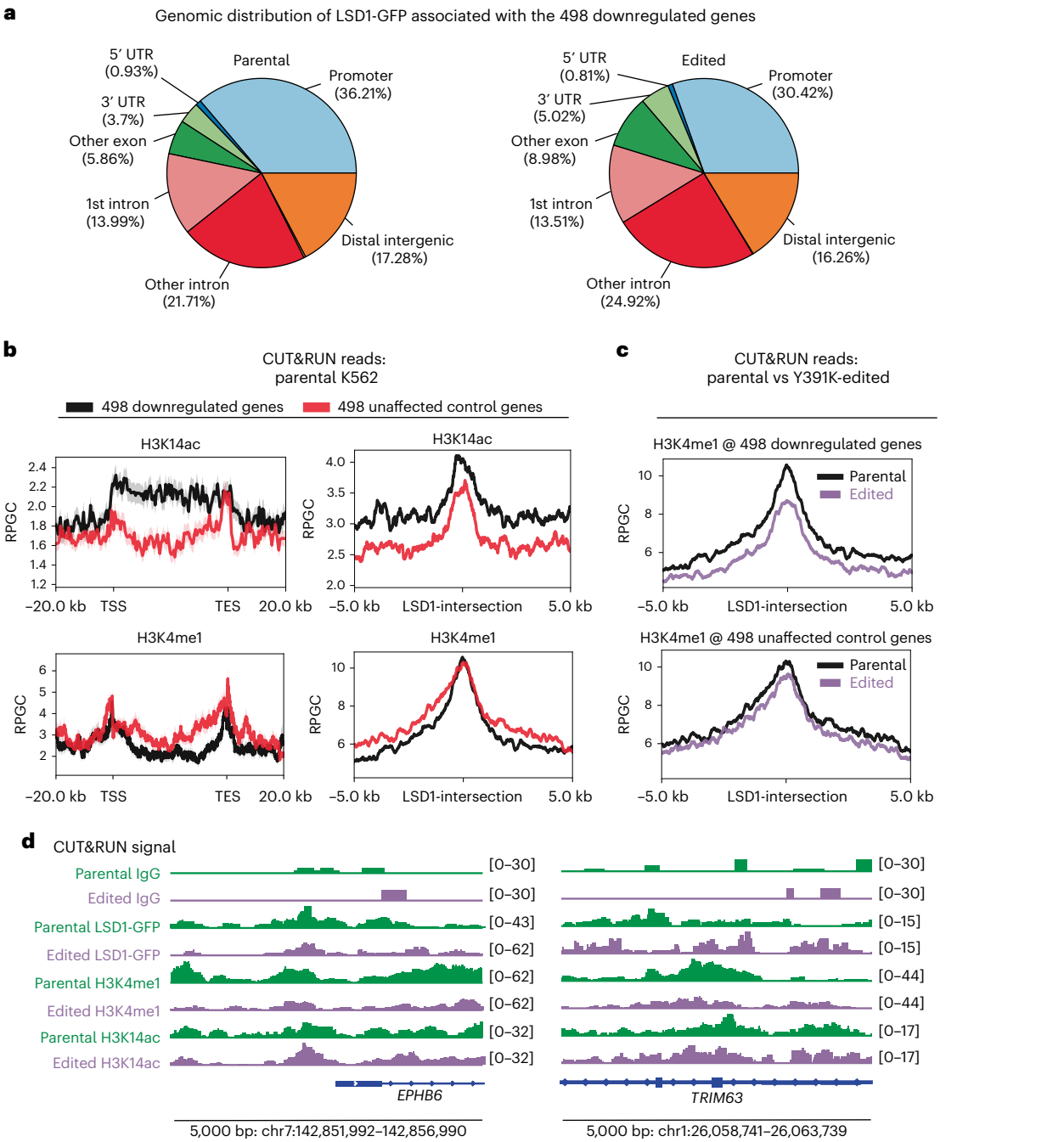



**Fig. 5 | CUT&RUN chromatin profiling analysis of parental and edited K562 cells, using biological duplicates for parental and edited K562 cell #1. a**, Pie chart displaying LSD1 occupancy at the regions associated with the 498 downregulated genes. ChIPseeker was used to annotate the LSD1 peaks across the 498 downregulated genes. Most LSD1 peaks were found in three major categories: promoter, introns and distal intergenic regions (up to ~80 kb away from the nearest transcription start site of the 498 downregulated genes). Downstream regions (≤300 bp) were also included in the analysis, with parental K562 cells showing 0.31% and edited K562 cells showing 0.08% LSD1 occupancy, but these were not separately labeled on the charts. **b**, Metagene plot (mean ± s.e.m.) representing H3K14ac (top) and H3K4me1 (bottom) from the scaled gene bodies ±20 kb of the 498 downregulated genes and 498 unaffected control genes (left) and at the LSD1 peaks (±5 kb) across these genes (right). ChIPseeker was used to identify these LSD1 peaks associated with the genes. The gene bodies of the downregulated genes show a stronger H3K14ac signal than the unaffected control genes, whereas H3K4me1 reads remain similar in both. Metagene plots were generated using deepTools. **c**, Metagene plot (mean) representing H3K4me1 at the LSD1 peaks (±5 kb) across the 498 downregulated genes (top) and the 498 unaffected control genes (bottom) between parental (black) and edited (pink) K562 cells. H3K4me1 at the LSD1 peaks in the downregulated genes show a reduction in the edited K562 cells, whereas those of the unaffected control genes remain similar. **d**, Genomic snapshot showing CUT&RUN signals for LSD1, H3K4me1 and H3K14ac in comparison with IgG for two representative genes: *EPHB6* and *TRIM63*. The H3K4me1 signal shows a reduction in the 'edited K562' cells.

LSD1 CUT&RUN peaks across the 498 downregulated genes were also elevated but not in the 498 unaffected control genes (Fig. 5b). Among the 498 downregulated genes, 341 were directly occupied by LSD1 in both parental and edited K562 cells. Interestingly, LSD1-bound genes displayed higher H3K14ac levels than the 157 LSD1-unbound genes (Fig. 6a). This observation provides additional insights, indicating that

https://doi.org/10.1038/s41589-024-01671-9

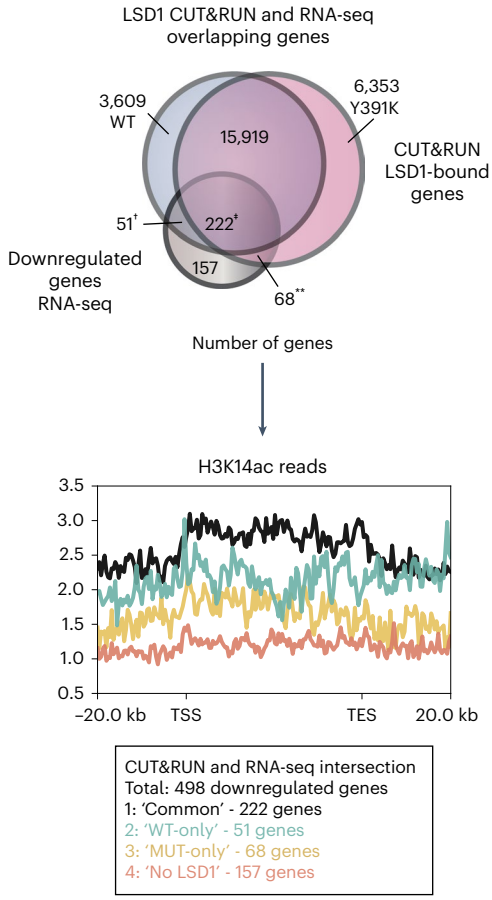



**Fig. 6 | Crosstalk between LSD1 and H3K14ac.** Venn diagram (top) showing a large overlap of LSD1-bound genes between parental and edited K562 cells. All genes were annotated using ChIPseeker and LSD1 peak files. Notably, most of the 498 downregulated genes share this overlap with LSD1-bound genes. Significance levels are denoted by symbols: ** indicates $P = 0.00583$; † indicates $P = 0.00004$; and ‡ indicates $P < 0.00001$. These $P$ values from hypergeometric analysis emphasize the statistical significance of the gene overlaps, indicating that they surpass what would be expected by random chance. Metagene plot (mean) (bottom) shows higher levels of H3K14ac (RPGC-normalized) in the LSD1-bound genes compared with LSD1-unbound downregulated genes. Among the 222 downregulated genes with a common overlap with LSD1-bound genes in both cells, the highest H3K14ac levels in their gene bodies signify the crosstalk between LSD1 and H3K14ac, showcasing the unique enzymatic gain of function of Y391K LSD1 toward H3K14ac.

direct occupancy of Y391K LSD1 could potentially lead to disruption of the H3K14ac–LSD1 crosstalk.

To gather additional evidence on whether Y391K LSD1 occupancy at H3K14ac-rich downregulated genes contributes to alleviating the impediment of demethylation posed by H3K14ac, we compared the H3K4me2 or H3K4me1 levels between parental and edited K562 cells across the downregulated genes. H3K4me1 levels were similar in 'parental K562' cells when comparing the 498 downregulated and unaffected control genes (Fig. 5b), whereas H3K4me2 displayed slightly lower levels in the 498 downregulated genes compared with the unaffected control genes (Extended Data Fig. 7b). It was especially interesting, however, that H3K4me1 levels were reduced near LSD1 in the downregulated genes in the 'edited K562' cells (Fig. 5c,d), whereas H3K4me2 showed only marginal differences (Extended Data Fig. 8a). Both H3K4me1 and H3K4me2 in the LSD1-bound non-promoter regions of these genes showed subtle but more pronounced differences than the LSD1-bound promoter regions (Extended Data Fig. 10a). These results suggest that Y391K LSD1 at the H3K14ac-rich non-promoter region had stronger demethylase activities. Interestingly, global levels of H3K4me1 and H3K4me2 remain similar between parental and edited K562 cells (Extended Data Fig. 9a–h), supporting Y391K LSD1's 'on-target' effect consistent with the transcriptomic changes. We speculate that one plausible scenario to account for this 'on-target' effect is the preferential presence of LC over LHC to attain stronger demethylase activity on chromatin. Alternatively, it is also plausible that the H3K4me–K14ac crosstalk is more selectively enriched in the gene bodies of the downregulated genes. As CUT&RUN analysis does not assess how much of the H3K4me and H3K14ac co-occur on the same H3 tail within a nucleosome, and it is predicted that the impact of H3K14ac on H3K4me demethylation by LSD1 is '*in cis*', our ability to link CUT&RUN data with enzymological results is limited.

Global levels of H3K14ac were also reduced in the 'edited K562' cells (Extended Data Figs. 8b and 9a–g). Given that LHC is a slow deacetylase for H3K14ac[19,28,29] and a poor nucleosome demethylase (Fig. 1b), we speculate that global H3K14ac reduction could stem from upregulation of the activity of the set of HDAC complexes like MiDAC that can efficiently cleave H3K14ac[28,29]. Alternatively, downregulated activities of H3K14ac-selective acetyltransferases, such as GCN5, MOZ/MORF and HBO1, could occur[12,71–73]. Regardless of the mechanism, the reduction of H3K14ac accompanied by a drop in H3K4me1 would likely lead to cooperative effects on gene repression. Beyond these findings, we also note from information in the ENCODE database on K562 cells that euchromatin-associated H3K27ac and H3K79me2 levels were lower while heterochromain-associated H3K27me3, SUZ12 and EZH2 levels were higher near the 498 downregulated genes than the unaffected control genes (Extended Data Fig. 7d–h), showing consistency with their low baseline expression levels (Supplementary Fig. 5) and suggesting that these silenced genes are less prevalent in the euchromatin state under normal conditions.

## Discussion

Despite the increasing number of examples of crosstalk between two or more histone modifications associated with a histone enzyme eraser or writer, the function of such interplay is typically inferred from indirect experiments. Knockout or overexpression of an enzyme often makes it challenging to distinguish between decreases or increases of a particular histone mark and how the separate histone marks can functionally communicate. Moreover, interpreting histone H3 point mutations is complicated because amino acid replacement mimics of PTMs are imprecise, and there are numerous isoforms of histone H3 that need to be considered. To address the intricate web of complexities within epigenetic mark crosstalk, protein engineering strategies, such as histone mark–DNA methyl mark disruptors or bivalent histone mark biosensors[74,75] have emerged as powerful tools offering selectivity and specificity. Our approach here describes an unconventional strategy to address the question of histone modification crosstalk by tailoring the enzyme's mode of substrate recognition involved in sensing and driving the histone mark interplay. This study identified an LSD1 mutant, Y391K, that allows for the negative catalytic influence of H3K14ac to be disabled by profoundly altering the demethylase activity toward its target site H3K4me on nucleosomes. The structural basis of Y391K in LSD1 conferring resistance to the inhibitory effect of H3K14ac appears to involve a cascade of conformational adjustments that position the histone H3 tail in a state resembling the unacetylated tail in the WT enzyme, orchestrated by the local conformational change of CoREST1 near Y391.

We were able to discern the functional impact of Y391K in cells through gene editing and showed that the Y391K LSD1 knockin alleles appear to silence a set of genes associated with cell adhesion and myeloid leukocyte activation. CUT&RUN analysis provides insights into these effects by revealing decreased H3K4me1 levels within the promoter, intragenic and distal regions of the silenced genes that are enriched in H3K14ac (Fig. 5b). Our findings support the biological

importance of LSD1 in tuning a delicate balance of gene expression by shaping and responding to histone mark crosstalk in the H3 tail.

Our study also reveals evidence that LC may serve as the primary nucleosome demethylase form of LSD1 in cells, whereas LHC's principal enzymatic role is as a histone deacetylase that is recruited by SNAG domain transcription factors. Although we cannot rule out that LHC could demethylate chromatin under special circumstances, the position of HDAC1 in LHC appears to obstruct the CoREST1 surface that has been shown to bind to the nucleosome core. As such, LSD1 would be impeded from demethylating the H3 tail in LHC. How the LSD1 pool is distributed among LC versus LHC complexes is not well understood but may involve partnering with particular CoREST isoforms because CoREST2 appears to have diminished HDAC affinity compared with CoREST1 (ref. 47). Alternatively, PTMs on the various LHC subunits or other protein interactors such as BHC80 could influence LC complex stabilities. Future studies can presumably shed light on the specific features of LC and LHC that drive distinct biological outputs.

## Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41589-024-01671-9.

## References


1. Kornberg, R. D. Structure of chromatin. *Annu. Rev. Biochem.* **46**, 931–954 (1977).
2. Richmond, T. J., Finch, J. T., Rushton, B., Rhodes, D. & Klug, A. Structure of the nucleosome core particle at 7 Å resolution. *Nature* **311**, 532–537 (1984).
3. Wang, Z. A. & Cole, P. A. The chemical biology of reversible lysine post-translational modifications. *Cell Chem. Biol.* **27**, 953–969 (2020).
4. Valencia-Sánchez, M. I. et al. Structural basis of Dot1L stimulation by histone H2B lysine 120 ubiquitination. *Mol. Cell* **74**, 1010–1019 (2019).
5. Worden, E. J., Hoffmann, N. A., Hicks, C. W. & Wolberger, C. Mechanism of cross-talk between H2B ubiquitination and H3 methylation by Dot1L. *Cell* **176**, 1490–1501 (2019).
6. Hsu, P. L. et al. Structural basis of H2B ubiquitination-dependent H3K4 methylation by COMPASS. *Mol. Cell* **76**, 712–723 (2019).
7. Kwon, M. et al. H2B ubiquitylation enhances H3K4 methylation activities of human KMT2 family complexes. *Nucleic Acids Res.* **48**, 5442–5456 (2020).
8. Dreveny, I. et al. The double PHD finger domain of MOZ/MYST3 induces α-helical structure of the histone H3 tail to facilitate acetylation and methylation sampling and modification. *Nucleic Acids Res.* **42**, 822–835 (2014).
9. Jain, K. et al. An acetylation-mediated chromatin switch governs H3K4 methylation read-write capability. *eLife* **12**, e82596 (2023).
10. Ringel, A. E., Cieniewicz, A. M., Taverna, S. D. & Wolberger, C. Nucleosome competition reveals processive acetylation by the SAGA HAT module. *Proc. Natl Acad. Sci. USA* **112**, E5461–E5470 (2015).
11. Takahashi, S. et al. HBO1-MLL interaction promotes AF4/ENL/P-TEFb-mediated leukemogenesis. *eLife* **10**, e65872 (2021).
12. Taverna, S. D. et al. Yng1 PHD finger binding to H3 trimethylated at K4 promotes NuA3 HAT activity at K14 of H3 and transcription at a subset of targeted ORFs. *Mol. Cell* **24**, 785–796 (2006).
13. Li, N. et al. ZMYND8 reads the dual histone mark H3K4me1-H3K14ac to antagonize the expression of metastasis-linked genes. *Mol. Cell* **63**, 470–484 (2016).
14. Local, A. et al. Identification of H3K4me1-associated proteins at mammalian enhancers. *Nat. Genet.* **50**, 73–82 (2018).
15. Morrison, E. A., Bowerman, S., Sylvers, K. L., Wereszczynski, J. & Musselman, C. A. The conformation of the histone H3 tail inhibits association of the BPTF PHD finger with the nucleosome. *eLife* **7**, e31481 (2018).
16. Forneris, F., Binda, C., Vanoni, M. A., Battaglioli, E. & Mattevi, A. Human histone demethylase LSD1 reads the histone code. *J. Biol. Chem.* **280**, 41360–41365 (2005).
17. Maltby, V. E. et al. Histone H3K4 demethylation is negatively regulated by histone H3 acetylation in *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* **109**, 18505–18510 (2012).
18. Shi, Y.-J. et al. Regulation of LSD1 histone demethylase activity by its associated factors. *Mol. Cell* **19**, 857–864 (2005).
19. Wu, M. et al. Lysine-14 acetylation of histone H3 in chromatin confers resistance to the deacetylase and demethylase activities of an epigenetic silencing complex. *eLife* **7**, e37231 (2018).
20. Barski, A. et al. High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823–837 (2007).
21. Shi, Y. et al. Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119**, 941–953 (2004).
22. Gu, F. et al. Biological roles of LSD1 beyond its demethylase activity. *Cell. Mol. Life Sci.* **77**, 3341–3350 (2020).
23. Lin, T., Ponn, A., Hu, X., Law, B. K. & Lu, J. Requirement of the histone demethylase LSD1 in Snai1-mediated transcriptional repression during epithelial-mesenchymal transition. *Oncogene* **29**, 4896–4904 (2010).
24. Lin, Y. et al. The SNAG domain of Snail1 functions as a molecular hook for recruiting lysine-specific demethylase 1. *EMBO J.* **29**, 1803–1816 (2010).
25. Mohammad, H. P. et al. A DNA hypomethylation signature predicts antitumor activity of LSD1 inhibitors in SCLC. *Cancer Cell* **28**, 57–69 (2015).
26. Humphrey, G. W. et al. Stable histone deacetylase complexes distinguished by the presence of SANT domain proteins CoREST/kiaa0071 and Mta-L1. *J. Biol. Chem.* **276**, 6817–6824 (2001).
27. You, A., Tong, J. K., Grozinger, C. M. & Schreiber, S. L. CoREST is an integral component of the CoREST-human histone deacetylase complex. *Proc. Natl Acad. Sci. USA* **98**, 1454–1458 (2001).
28. Wang, Z. A. et al. Diverse nucleosome site-selectivity among histone deacetylase complexes. *eLife* **9**, e57663 (2020).
29. Wang, Z. A. et al. Histone H2B deacylation selectivity: exploring chromatin's dark matter with an engineered sortase. *J. Am. Chem. Soc.* **144**, 3360–3364 (2022).
30. Kim, S.-A., Zhu, J., Yennawar, N., Eek, P. & Tan, S. Crystal structure of the LSD1/CoREST histone demethylase bound to its nucleosome substrate. *Mol. Cell* **78**, 903–914 (2020).
31. Song, Y. et al. Mechanism of crosstalk between the LSD1 demethylase and HDAC1 deacetylase in the CoREST complex. *Cell Rep.* **30**, 2699–2711 (2020).
32. Baron, R., Binda, C., Tortorici, M., McCammon, J. A. & Mattevi, A. Molecular mimicry and ligand recognition in binding and catalysis by the histone demethylase LSD1-CoREST complex. *Structure* **19**, 212–220 (2011).
33. Vinyard, M. E. et al. CRISPR-suppressor scanning reveals a nonenzymatic role of LSD1 in AML. *Nat. Chem. Biol.* **15**, 529–539 (2019).
34. Egolf, S. et al. LSD1 inhibition promotes epithelial differentiation through derepression of fate-determining transcription factors. *Cell Rep.* **28**, 1981–1992 (2019).
35. Ferrari-Amorotti, G. et al. Inhibiting interactions of lysine demethylase LSD1 with Snail/Slug blocks cancer cell invasion. *Cancer Res.* **73**, 235–245 (2013).
36. Ishikawa, Y. et al. A novel LSD1 inhibitor T-3775440 disrupts GFI1B-containing complex leading to transdifferentiation and impaired growth of AML cells. *Mol. Cancer Ther.* **16**, 273–284 (2017).
37. Rivera, C. et al. Unveiling RCOR1 as a rheostat at transcriptionally permissive chromatin. *Nat. Commun.* **13**, 1550 (2022).

38. Zeng, C. et al. Demethylase-independent roles of LSD1 in regulating enhancers and cell fate transition. *Nat. Commun.* **14**, 4944 (2023).
39. Kim, S.-A., Chatterjee, N., Jennings, M. J., Bartholomew, B. & Tan, S. Extranucleosomal DNA enhances the activity of the LSD1/CoREST histone demethylase complex. *Nucleic Acids Res.* **43**, 4868–4880 (2015).
40. Simon, M. D. et al. The site-specific installation of methyl-lysine analogs into recombinant histones. *Cell* **128**, 1003–1012 (2007).
41. Dempsey, D. R., Jiang, H., Kalin, J. H., Chen, Z. & Cole, P. A. Site-specific protein labeling with *N*-hydroxysuccinimide-esters and the analysis of ubiquitin ligase mechanisms. *J. Am. Chem. Soc.* **140**, 9374–9378 (2018).
42. Jiang, H. & Cole, P. A. N-terminal protein labeling with *N*-hydroxysuccinimide esters and microscale thermophoresis measurements of protein–protein interactions using labeled protein. *Curr. Protoc.* **1**, e14 (2021).
43. Lee, K., Whedon, S. D., Wang, Z. A. & Cole, P. A. Distinct biochemical properties of the class I histone deacetylase complexes. *Curr. Opin. Chem. Biol.* **70**, 102179 (2022).
44. Astro, V. et al. Fine-tuned KDM1A alternative splicing regulates human cardiomyogenesis through an enzymatic-independent mechanism. *iScience* **25**, 104665 (2022).
45. Dhall, A. et al. Nucleosome binding by the lysine specific demethylase 1 (LSD1) enzyme enables histone H3 demethylation. *Biochemistry* **59**, 2479–2483 (2020).
46. Barnes, C. E., English, D. M., Broderick, M., Collins, M. O. & Cowley, S. M. Proximity-dependent biotin identification (BioID) reveals a dynamic LSD1–CoREST interactome during embryonic stem cell differentiation. *Mol. Omics* **18**, 31–44 (2022).
47. Barrios, Á. P. et al. Differential properties of transcriptional complexes formed by the CoREST family. *Mol. Cell. Biol.* **34**, 2760–2770 (2014).
48. Forneris, F., Binda, C., Adamo, A., Battaglioli, E. & Mattevi, A. Structural basis of LSD1-CoREST selectivity in histone H3 recognition. *J. Biol. Chem.* **282**, 20070–20074 (2007).
49. Yang, M. et al. Structural basis of histone demethylation by LSD1 revealed by suicide inactivation. *Nat. Struct. Mol. Biol.* **14**, 535–9 (2007).
50. Metzger, E. et al. Phosphorylation of histone H3T6 by PKC$\beta_I$ controls demethylation at histone H3K4. *Nature* **464**, 792–796 (2010).
51. Pettersen, E. F. et al. UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).
52. Baron, R. & Vellore, N. A. LSD1/CoREST is an allosteric nanoscale clamp regulated by H3-histone-tail molecular recognition. *Proc. Natl Acad. Sci. USA* **109**, 12509–12514 (2012).
53. Nalawansha, D. A. & Pflum, M. K. H. LSD1 substrate binding and gene expression are affected by HDAC1-mediated deacetylation. *ACS Chem. Biol.* **12**, 254–264 (2017).
54. Wang, Z. A. et al. Structural basis of sirtuin 6-catalyzed nucleosome deacetylation. *J. Am. Chem. Soc.* **145**, 6811–6822 (2023).
55. Kalin, J. H. et al. Targeting the CoREST complex with dual histone deacetylase and demethylase inhibitors. *Nat. Commun.* **9**, 53 (2018).
56. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
57. Maes, T. et al. ORY-1001, a potent and selective covalent KDM1A inhibitor, for the treatment of acute leukemia. *Cancer Cell* **33**, 495–511 (2018).
58. Ali, A., Vaikari, V. P. & Alachkar, H. CD99 in malignant hematopoiesis. *Exp. Hematol.* **106**, 40–46 (2022).
59. Wang, H., Chen, B., Lin, Y., Zhou, Y. & Li, X. Legumain promotes gastric cancer progression through tumor-associated macrophages in vitro and in vivo. *Int. J. Biol. Sci.* **16**, 172–180 (2020).
60. D'Amato, L. et al. ARHGEF3 controls HDACi-induced differentiation via RhoA-dependent pathways in acute myeloid leukemias. *Epigenetics* **10**, 6–18 (2015).
61. Ishiura, M. et al. Positive interactions between STAP-1 and BCR-ABL influence chronic myeloid leukemia cell proliferation and survival. *Biochem. Biophys. Res. Commun.* **556**, 185–191 (2021).
62. Al-Jamal, H. A. N., Johan, M. F., Mat Jusoh, S. A., Ismail, I. & Wan Taib, W. R. Re-expression of bone marrow proteoglycan-2 by 5-azacytidine is associated with STAT3 inactivation and sensitivity response to imatinib in resistant CML cells. *Asian Pac. J. Cancer Prev.* **19**, 1585–1590 (2018).
63. Wesely, J. et al. Acute myeloid leukemia iPSCs reveal a role for RUNX1 in the maintenance of human leukemia stem cells. *Cell Rep.* **31**, 107688 (2020).
64. Akbari, M. et al. A novel variant in *TLE6* is associated with embryonic developmental arrest (EDA) in familial female infertility. *Sci. Rep.* **12**, 17664 (2022).
65. Park, E. C. & Rongo, C. RPM-1 and DLK-1 regulate pioneer axon outgrowth by controlling Wnt signaling. *Development* **145**, dev164897 (2018).
66. Vietor, I. et al. The negative adipogenesis regulator *Dlk1* is transcriptionally regulated by *Ifrd1* (TIS7) and translationally by its orthologue *Ifrd2* (SKMc15). *eLife* **12**, e88350 (2023).
67. Woo, I. S. et al. TMEM14A inhibits N-(4-hydroxyphenyl)retinamide-induced apoptosis through the stabilization of mitochondrial membrane potential. *Cancer Lett.* **309**, 190–8 (2011).
68. Hondares, E. et al. Enhanced activation of an amino-terminally truncated isoform of the voltage-gated proton channel HVCN1 enriched in malignant B cells. *Proc. Natl Acad. Sci. USA* **111**, 18078–83 (2014).
69. Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife* **6**, e21856 (2017).
70. Yu, G., Wang, L.-G. & He, Q.-Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382–2383 (2015).
71. Doyon, Y. et al. ING tumor suppressor proteins are critical regulators of chromatin acetylation required for genome expression and perpetuation. *Mol. Cell* **21**, 51–64 (2006).
72. Karmodiya, K., Krebs, A. R., Oulad-Abdelghani, M., Kimura, H. & Tora, L. H3K9 and H3K14 acetylation co-occur at many gene regulatory elements, while H3K14ac marks a subset of inactive inducible promoters in mouse embryonic stem cells. *BMC Genomics* **13**, 424 (2012).
73. Mishima, Y. et al. The Hbo1-Brd1/Brpf2 complex is responsible for global acetylation of H3K14 and required for fetal liver erythropoiesis. *Blood* **118**, 2443–2453 (2011).
74. Delachat, A. M. F. et al. Engineered multivalent sensors to detect coexisting histone modifications in living stem cells. *Cell Chem. Biol.* **25**, 51–56 (2018).
75. Noh, K.-M. et al. Engineering of a histone-recognition domain in Dnmt3a alters the epigenetic landscape and phenotypic features of mouse ESCs. *Mol. Cell* **69**, 533 (2018).

## Methods

### 185-bp DNA preparation

Two 185-bp dsDNA W601-related sequences were used to reconstitute nucleosomes:

W601 DNA #1: 5′-ATCGCTGTTCAATACATGCACAGGATGTATATATCTGACACGTGCCTGGAGACTAGGGAGTAATCCCCTTGGCGGTTAAAACGCGGGGGACAGCGCGTACGTGCGTTTAAGCGGTGCTAGAGCTGTCTACGACCAATTGAGCGGCCTCGGCACCGGGATTCTCCAGGGCGGCCGCGTATAGGGAT-3′

W601 DNA #2: 5′-ATCGCTGTTCAATACATGCACAGGATGTATATATCTGACACGTGCCTGGAGACTAGGGAGTAATCCCCTTGGCGGTTAAAACGCGGGGGACAGCGCGTACGTGCGTTTAAGCGGTGCTAGAGCTGTCTACGACCAATTGAGCGGCCTCGTTCACCTGCCTCTCCAGGGCGGCCGCGTATAGGGAT-3′

W601 DNA #1 was used to reconstitute nucleosomes for the demethylase assays, in accordance with the method as previously documented[19]. W601 DNA #2 was employed for nucleosome reconstitution to measure the binding affinities between nucleosomes and CoREST complexes using MST. To bacterially produce each DNA, 12 tandem repeats of W601 DNA #1 were inserted into a pUC57 vector, and four tandem repeats of W601 DNA #2 were inserted into a pUC18 vector (synthesized by GenScript). Each tandem repeat was separated by an EcoRV recognition site (5′-GAT/ATC-3′). Transformation of each plasmid into DH5α cells was performed on LB agar plates with 100 mg $L^{-1}$ ampicillin. A single colony was inoculated into 1 L of LB media in a baffled flask (Thomson, 931136-B) and maintained at 37 °C and 200 r.p.m. overnight. Cells were harvested via centrifugation at 2,702*g*. Harvested cells were resuspended in 11.3 ml of the resuspension buffer containing 10 mM Tris at pH 8.0 and 10 mM EDTA. Then, 50 mg of chicken egg lysozyme was added (Millipore Sigma), followed by thorough mixing. This mixture was left at room temperature for 5 min. The samples were divided into two portions, and 15 ml of lysis solution, composed of 0.2 M NaOH and 1% SDS, was added to each portion. After thorough mixing, the mixtures were left at room temperature for 5 min. To the samples, 12 ml of neutralization buffer was added, containing 1.5 M potassium acetate and 1.5 M acetic acid mixture at pH 4.8. The samples were vigorously mixed by inverting, but not vortexing, and incubated at room temperature for 5 min. After this, the samples were centrifuged at 4,000*g* for 10 min. The supernatants were combined and filtered using an Econo column (Bio-Rad). The obtained flowthrough was subsequently mixed with 0.6 eq of isopropanol, incubated at room temperature for 5 min and centrifuged at 4,000*g* for 10 min. The resulting pellets were resuspended in a total of 15 ml of the resuspension buffer (10 mM Tris at pH 8.0 and 10 mM EDTA). An equal volume of ice-cold 5 M lithium chloride (Millipore Sigma) was added to the suspension. The sample was incubated on ice for 5 min and centrifuged at 4,000*g* to precipitate RNA and proteins. The supernatant was subsequently mixed with 0.6 eq of isopropanol and incubated at room temperature for 5 min. After centrifugation at 4,000*g* for 10 min, the pellet was resuspended with 10 ml of 10 mM Tris at pH 8.0 and 1 mM EDTA. Then, 50 µl of 10 mg $ml^{-1}$ heat-treated RNase A (Bio Basic) was introduced. After thorough mixing, the sample was incubated at room temperature for 15 min. Then, 30 ml of isopropanol was added, mixed, incubated at room temperature for 5 min and centrifuged at 4,000*g* for 10 min. After discarding the supernatant, the pellet was dissolved in 10 ml of 1× rCutSmart buffer (New England Biolabs (NEB)), and 50 µl of EcoRV (NEB) was added followed by the incubation at 37 °C until the 185-bp DNA was fully digested from the vector (typically ~24 h). Upon completion of the cleavage, the sample was diluted 10-fold in a Q-resin buffer A containing 10 mM Tris at pH 7.5 and 1 mM EDTA. The diluted sample was then incubated with Q-sepharose FF resin (Cytiva), and the slurry was poured onto an Econo column. The resin was then washed with the Q-wash buffer (Q-resin buffer A with 300 mM NaCl), and the DNA was eluted with the Q-elution buffer (Q-resin buffer A with 800 mM NaCl). The eluant was concentrated using a 10-kDa molecular weight cutoff (MWCO) concentrator (Amicon, Millipore Sigma) and further purified by size-exclusion chromatography in which the column (Superdex 200 increase 10/300 GL) was pre-equilibrated with the size-exclusion buffer containing 10 mM Tris at pH 7.5, 1 M NaCl and 1 mM EDTA. The fractions pooled at the main peak around approximately 10 ml of the elution were re-injected into the size-exclusion column to attain a highly purified (>95%) sample.

### Peptide synthesis

Histone H3 peptides (aa 1–21 and aa 1–34). H3 peptides were synthesized and purified as previously reported[19,28,29,54,55]. Using the Fmoc-based solid-phase peptide synthesis strategy with Fmoc-Ala-Wang resin for aa 1–21 or with Rink Amide AM resin (Millipore Sigma), syntheses were performed using a Prelude peptide synthesizer (Gyros Protein Technologies) at a 0.1-mmol scale. The synthesis cycles typically followed deprotection and double-coupling procedures as below:

N-terminal Fmoc groups were deprotected using 5 ml of 20% piperidine in dimethyl formamide (DMF) mixed for 10 min. Each deprotection was repeated twice. Subsequently, 4 ml of DMF was added to the reaction vessel and mixed for 30 s. The vessel was drained, and five additional washes were performed with alternating solvent delivery from the top and bottom of the reaction vessel. For each coupling cycle, 0.4 mmol (4 eq) Fmoc-amino acid in 2 ml of DMF, along with 0.375 mmol (3.75 eq) HATU and 0.8 mmol (8 eq) *N*-methylmorpholine (NMM) in 4 ml of DMF were sequentially added. After mixing for 90 min, a wash with 4 ml of DMF was carried out, followed by a second identical coupling procedure. The resin was again washed five times with 4 ml of DMF before the next cycle of deprotection and coupling.

Upon completion of amino acid coupling, peptides were N-terminally deprotected, as described above. The resin was sequentially washed with DMF and dichloromethane (DCM) and dried under vacuum. Peptide cleavage from the resin and sidechain protecting group removal were achieved by adding Reagent B (5% water, 5% phenol, 2.5% triisopropyl silane (TIPS) and 87.5 % trifluoroacetic acid (TFA)) and gently mixing for 90 min. The supernatant was filtered through a disposable glass fritted column, and the peptide was precipitated by adding 10 vol eq of cold diethyl ether. The suspension was pelleted by centrifugation at 3,800*g*, and the supernatant was removed. The pellet was subsequently washed two times with 10 vol eq of cold diethyl ether, after which the pellet was dried under a stream of nitrogen. Washed peptide precipitates were dissolved in a solution containing 89.95% $H_2O$, 10% $CH_3CN$ and 0.05% TFA and purified by reversed-phase high-performance liquid chromatography (HPLC) with a C18 semi-preparative column (Varian Dynamax Microsorb 100, 250 × 21.4 mm) using a linear gradient from 7% $CH_3CN$/0.05% TFA to 30% $CH_3CN$/0.05% TFA over 30 min at a flow rate of 10 ml $min^{-1}$. Fractions containing the desired peptides were determined by MALDI-TOF MS (Dana Farber Cancer Institute Molecular Biology Core Facilities, 4800 MALDI TOF/TOF, Applied Biosystems/MDS SCIEX) or ESI-MS (Q Exactive, Thermo Fisher Scientific) and pooled. Pooled samples were lyophilized and stored at −80 °C as dry powders or concentrated stock solutions.

### Incorporating depsipeptide Fmoc-Thr(OtBu)-glycolic acid into H3 peptides (aa 1–34)

The synthesis of Thr(OtBu)-glycolic acid (TOG) was performed as previously described[19]. After the coupling of the first Gly, the resin (0.1 mmol) was treated with the desipeptide TOG (0.2 mmol, 2 eq), 0.18 mmol (1.8 eq) HATU and 0.4 mmol (4 eq) DIPEA in DMF. Subsequent amino acid couplings followed the general protocol as described above.

### Protein expression and purification

**LC for enzymology and MST.** $His_6$-tagged LSD1 (aa 171–852) and $His_6$-tagged (for enzymology) or $His_8$-SUMO-tagged (for fluorophore labeling) CoREST1 (Pro286-Ser482) were subcloned in pET15b and

pET28a vectors, respectively. Primers used for mutagenesis are listed in Supplementary Table 1. Then, 300 ng of each plasmid was co-transformed into BL21-CodonPlus (DE3)-RIPL competent cells (Agilent). A single colony was inoculated in a starter LB media containing 100 mg $L^{-1}$ ampicillin, 50 mg $L^{-1}$ kanamycin and 35 mg $L^{-1}$ chloramphenicol at 37 °C. The starter culture was further inoculated into a larger culture, and, when $A_{600}$ reached 0.6, 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) was added for induction at 18 °C for 20 h. The cells were harvested by centrifugation at 2,702*g*. Harvested cells were resuspended in lysis buffer containing 20 mM Tris pH 7.8, 200 mM NaCl and 0.5 mM tris-carboxy-ethyl-phosphine (TCEP), followed by cell lysis by French press. The soluble lysate was extracted via centrifugation at 20,853*g* and purified using nickel-nitriloacetic acid (NiNTA) resin (MCLAB). In brief, the resin was washed with the wash buffer (lysis buffer containing 30 mM imidazole), and LC was eluted with the elution buffer (lysis buffer containing 200 mM imidazole). The eluants were concentrated and buffer exchanged into the lysis buffer using a 10-kDa MWCO concentrator (Amicon, Millipore Sigma). Samples were then purified over a size-exclusion column pre-equilibrated with the lysis buffer to attain stoichiometric complexes (Superdex 200 increase 10/300 GL, Cytiva) that appeared more than 95% pure by SDS-PAGE stained with Coomassie. Fractions from the main peak (~12.5 ml elution volume) were concentrated down using a 10-kDa MWCO concentrator (Amicon, Millipore Sigma) for crystallographic studies or directly used for enzymological and binding studies without any further concentration. LC protein solutions were flash frozen and stored at −80 °C.

**GST-LSD1.** GST-tagged LSD1 (aa 171–852) was subcloned in a pGEX6P-1 vector and bacterially expressed and purified as previously reported[76]. Primers used for mutagenesis are listed in Supplementary Table 1. In brief, the plasmids encoding WT and mutant LSD1 were transformed into BL21-CodonPlus (DE3)-RIPL competent cells (Agilent), and single colonies were picked for inoculation in LB media containing 100 mg $L^{-1}$ ampicillin and 35 mg $L^{-1}$ chloramphenicol. When $A_{600}$ reached 0.6 at 37 °C, 0.5 mM IPTG was added to induce protein production at 16 °C for 20 h. Cells were harvested via centrifugation at 2,702*g* and resuspended in the lysis buffer (20 mM $Na_2HPO_4$, 3.6 mM $KH_2PO_4$, 5.4 mM KCl, 280 mM NaCl, 1 mM EDTA, 10 mM DTT and 10% glycerol at pH 7.4). Resuspended cells were lysed by French press and centrifuged at 20,853*g* for 30 min. The supernatant was then incubated with the glutathione agarose beads (MCLAB) and washed with the lysis buffer, and GST-LSD1 protein was eluted with lysis buffer containing 20 mM glutathione. The eluant was then concentrated and buffer exchanged against the lysis buffer containing 1 mM β-mercaptoethanol (BME) instead of 10 mM DTT, using a 50-kDa MWCO concentrator (Amicon, Millipore Sigma). GST-LSD1 appeared more than 70% pure by SDS-PAGE stained with Coomassie. GST-LSD1 protein solutions were flash frozen and stored at −80 °C.

**LHC.** Production of LHC was assessed as previously reported[19,55,77]. Full-length human LSD1, 3× flag-tagged CoREST1 (aa 84–482) and full-length human HDAC1 in pcDNA3.1 expression vectors were co-transfected into HEK293F cells with polyethylenimine (PEI) (Millipore Sigma, 408727-100ML). Cells were harvested via centrifugation at 2,702*g*, resuspended in the lysis buffer (50 mM HEPES at pH 7.5, 100 mM KCl, 0.3% Triton X-100, 5% glycerol and protease inhibitor cocktail (Thermo Fisher Scientific)), lysed by sonication and centrifuged at 20,853*g*. The supernatant was further purified over anti-flag affinity resin (Millipore Sigma), followed by TEV protease-assisted tag removal for CoREST1 and gel filtration using a Superose 6 10/300 GL column (Cytiva) pre-equilibrated with a buffer containing 50 mM HEPES pH 7.5, 50 mM KCl and 0.5 mM TCEP. The sample was concentrated to approximately 5 mM using a concentrator with a 30-kDa MWCO (Pall). LHC protein showed 1:1:1 stoichiometry and appeared more than 90% pure by SDS-PAGE stained with Coomassie. LHC protein solutions were flash frozen and stored at −80 °C.

**Histones.** *Xenopus laevis* histones H2A, H2B, H3 and H4 and globular H3 (gH3; aa 35–135) were bacterially expressed and purified as previously described[19,28,29,78,79]. In brief, core histone proteins were transformed into BL21-CodonPlus (DE3)-RIPL competent cells (Agilent). A single colony was inoculated in a starter LB media containing 100 mg $L^{-1}$ ampicillin and 35 mg $L^{-1}$ chloramphenicol, further inoculated into a larger culture, grown up to $A_{600}$ ~ 0.6 and induced with 1 mM IPTG for 3 h at 37 °C. The cells were harvested by centrifugation at 2,702*g* (4 °C for histones H2A, H2B, H3 and gH3; 25 °C for histone H4) and resuspended in the wash buffer containing 50 mM Tris-HCl at pH 7.5, 100 mM NaCl, 1 mM EDTA, 5 mM BME, 0.2 mM phenylmethylsulfonyl fluoride (PMSF) and 1% Triton X-100. After lysing cells by French press, the lysate was centrifuged at 20,853*g* for 30 min. The pellet was subsequently resuspended in the wash buffer once more and centrifuged again at 20,853*g* for 10 min. The washed pellet was further washed two more times at the same centrifugation cycle with the wash buffer without Triton X-100. Then, the pellet was resuspended in the denaturation buffer containing 20 mM Tris at pH 7.5, 7 M guanidinium hydrochloride and 10 mM DTT and agitated at room temperature for 10 min. The mixture was then centrifuged at 20,853*g* for 30 min. The supernatant was dialyzed against 4 L of ion exchange buffer (IEX) containing 10 mM Tris at pH 7.8, 7 M urea, 1 mM EDTA, 0.2 mM PMSF and 5 mM BME at room temperature overnight using a 3.5-kDa MWCO dialysis membrane (Spectra/Por 7). Samples were first purified over Q Sepharose FF resin (Cytiva) pre-equilibrated with the IEX buffer containing 100 mM NaCl. Then, the flowthrough was subsequently loaded onto SP Sepharose FF resin (Cytiva) pre-equilibrated with IEX buffer containing 100 mM NaCl. Each histone was then eluted from the SP resin with the IEX buffer containing 100 mM to 500 mM NaCl. Fractions containing histones were identified via SDS-PAGE, pooled and dialyzed against 2 mM BME using dialysis tubing (3.5-kDa MWCO, Spectra/Por). Dialyzed samples were then lyophilized and stored at −80 °C.

### F40 sortase

F40 sortase with the APXTG motif preference was used to generate semi-synthetic histone H3 with specific modifications at Lys4, Lys9 and Lys14. Expression and purification of the F40 sortase used for histone semi-synthesis was as previously described[19,28,29,54]. Generally, F40 subcloned in a pET21 vector was expressed from BL21(DE3) with 0.25 mM IPTG for 4 h of induction at 30 °C. Cells were harvested by centrifugation at 2,702*g* and then resuspended in lysis buffer containing 20 mM Tris at pH 8, 0.1% Triton X-100 and 1 mM PMSF. The cells were then lysed by the French press and centrifuged at 20,853*g* for 30 min. The lysate supernatant was incubated with the NiNTA resin (MCLAB), washed with the wash buffer containing 20 mM Tris at pH 8, 0.1% Triton X-100 and 500 mM NaCl and eluted with the elution buffer containing 20 mM Tris at pH 7.5, 500 mM NaCl and 400 mM imidazole. Finally, the eluant was dialyzed against the buffer containing 50 mM Tris at pH 7.5, 150 mM NaCl and 5 mM $CaCl_2$ (Thermo Fisher Scientific, 66110), followed by the concentration (Amicon 10-kDa MWCO, Millipore Sigma) to approximately 10 mg $ml^{-1}$. The purified protein was flash frozen and stored at −80 °C.

### Semi-synthesis of histones H3K4me1, H3K4me2, H3K9ac, H3K14ac, H3K4me1/K14ac and H3K4me2/K14ac using F40 sortase

Sortase-mediated ligation of the depsipeptide and gH3 was carried out as previously reported[19,28,29,54]. H3 depsipeptides (aa 1–34) were used in more than 15-fold excess relative to gH3. Peptide and protein were combined in a reaction buffer containing 50 mM PIPES at pH 7.0, 5 mM $CaCl_2$, 1 mM DTT and 300 μM F40 sortase overnight at 37 °C.

Precipitate was isolated by centrifugation and re-dissolved for ion exchange purification in 10 mM Tris at pH 7.8, 7 M urea, 1 mM EDTA, 0.2 mM PMSF and 5 mM BME. The re-dissolved pellet solution and reaction supernatant were combined, and ligated and unligated products were separated by chromatography with SP Sepharose FF resin using the IEX buffer with a 100–500 mM NaCl gradient, as described above in the histone purification step. Unligated gH3 eluted around 150 mM NaCl, and the ligated products eluted generally between 180 mM and 250 mM NaCl. Ligated products were then dialyzed against 2 mM BME at 4 °C and then against 0.05% TFA at 4 °C, concentrated (Amicon 3.5-kDa MWCO, Millipore Sigma) and lyophilized. Purity was confirmed by mass spectrometry (MALDI-TOF MS or ESI-MS), and proteins were lyophilized and stored in the same manner as other histone proteins described above.

### Octamer refolding and nucleosome reconstitution

Octamer refolding and nucleosome assembly were performed as previously reported[19,78,79]. In brief, the core histone proteins H2A, H2B, H3 and H4 were dissolved in denaturation buffer (7 M guanidinium hydrochloride, 20 mM Tris pH 7.5 and 10 mM DTT) and mixed at a molar ratio of 1.1:1.1:1:1, respectively. After dialyzing against the octamer high-salt buffer (20 mM Tris 7.5, 2.0 M NaCl, 1 mM EDTA and 5 mM BME) at 4 °C, the octamer was purified by gel filtration (Superdex 200 increase 10/300 GL). The 185-bp DNAs used for these nucleosomes were prepared by the aforementioned method. The histone octamer and DNA were mixed at a 1.1:1 molar ratio in the nucleosome high-salt buffer (10 mM Tris 7.5, 2.0 M KCl, 1 mM EDTA and 1 mM DTT), and the mixture was dialyzed to the nucleosome low-salt buffer (10 mM Tris 7.5, 0.25 M KCl, 1 mM EDTA and 1 mM DTT) in a linear gradient. The reconstituted nucleosome was HPLC purified (Waters) over the TEKgel DEAE-5PW (Tosoh Bioscience) using a linear gradient between buffer A (10 mM Tris at pH 7.5, 0.25 M KCl and 1 mM EDTA) and buffer B (10 mM Tris at pH 7.5, 0.6 M KCl and 1 mM EDTA) at a 1 ml $min^{-1}$ flow rate. The fractions were collected and dialyzed (Thermo Fisher Scientific, 66383) into storage buffer (20 mM Tris 7.5 and 1 mM DTT) and concentrated to 5–10 μM using a concentrator (Amicon 10-kDa MWCO, Millipore Sigma). Nucleosome concentration and purity were assessed by absorbance at A260 and A280 and native gel electrophoresis and shown to be more than 90%.

### Fluorescein labeling of LC and LHC

An N-terminal cysteine-selective chemical labeling strategy[41,42] was applied to generate site-specifically fluorescein-labeled LC and LHC complexes. First, CoREST1 underwent Cys point mutagenesis at the tag removal site. For example, CoREST1 in LC (Pro286Cys-Ser482) had REQIGG/C at the Ulp1-mediated SUMO tag removal site, and CoREST1 in LHC (Trp84-Ser482, where Cys-Ala-Met precedes Trp84) had ENLYFQ/C at the TEV protease-mediated FLAG tag removal site. This allowed for N-terminal cysteine exposure that is necessary for chemoselective ligation. Fluorescein-thioester was generated by incubating 1 mM NHS-fluorescein (Thermo Fisher Scientific, 46410) in the activation buffer containing 100 mM HEPES at pH 7.3, 1 mM TCEP, 500 mM sodium 2-mercaptoethane sulfonate (MESNA) and 50 mM KCl for 3 h at room temperature in the dark. Then, purified LC or LHC was incubated with 50 eq fluorescein-thioester in the reaction buffer containing 100 mM HEPES at pH 7.3, 1 mM TCEP, 80 mM MESNA and 50 mM KCl at 4 °C for 48 h. After checking the fluorescein-labeling efficiency by SDS-PAGE and fluorescence gel imaging (Amersham Typhoon FLA 9500, Cytiva), residual fluorophore was removed by desalting (Zeba spin desalting column, 7-kDa MWCO), and the LC/LHC complex was further purified by size exclusion (Superose 6 10/300 GL for LHC and Superdex 200 increase 10/300 GL for LC, Cytiva) pre-equilibrated with a buffer containing 50 mM HEPES at pH 7.5, 50 mM KCl and 0.5 mM TCEP for LHC and 20 mM Tris pH 7.5, 200 mM NaCl and 0.5 mM TCEP for LC.

### Crystallography

The human full-length WT LSD1 and Δ305-CoREST1 proteins were expressed in *Escherichia coli* and co-purified as described[44]. $His_6$-Y391K LSD1 (aa 171–852) and Δ305-CoREST1 complex was co-expressed and purified using the same procedure. Crystals were grown by vapor diffusion by mixing 10 mg $ml^{-1}$ LSD1/Δ305CoREST1 in 25 mM $KH_2PO_4$ pH 7.2, 5% (v/v) glycerol with a precipitant solution consisting of 1–1.3 M Na/K tartrate and 0.1 M 2,2′-[(2-amino-2-oxoethyl)azanediyl]diacetic acid (ADA) pH 6.5 (ref. 80). They were harvested in a stabilizing solution consisting of 1.3 M Na/K tartrate, 0.1 M ADA pH 6.5 and 1 mM H3K4M N-terminal peptide (residues 1–21) with or without acetylation on K14. Diffraction data were measured at the European Synchrotron Radiation Facility (Grenoble, France) beamlines and processed with XDS (BUILT = 20180126)[81] and CCP4 (version 7.0.044) programs[82]. Structure determination and refinement were performed with Phenix (version 1.20.1-4487)[83] and Coot (version 0.9.6)[84]. For validation, the WT LC in complex with the H3K4M peptide was used as a control, demonstrating a nearly identical structure to the deposited WT LC-H3K4M peptide structure (PDB: 2V1D). Atomic coordinates for WT LC in complex with H3K4M/K14ac peptide, Y391K LC in complex with H3K4M peptide and Y391K LC in complex with H3K4M/K14ac peptide were deposited in the PDB (https://www.rcsb.org/). Crystallographic statistics are shown in Supplementary Table 2.

### SNAIL peptide binding assays

SNAIL binding affinities were measured by fluorescence polarization following the protocols described in ref. 85. We used full-length WT LSD1/Δ305CoREST1 and $His_6$-Y391K LSD1 (aa 171–852)/Δ305CoREST1 LC complexes identical to those employed in crystallography. The experiments were performed using a peptide containing the first nine residues of human SNAIL plus an additional C-terminal lysine where the TAMRA fluorophore was covalently attached. The assay solution contained 15 mM $KH_2PO_4$ pH 7.2, 5% (v/v) glycerol, 0.01% Tween 20 and 10 nM SNAIL (1-9)-TAMRA. The LC concentrations ranged from 2 μM for a total of 16 points and $n = 2$ independent experiments. Data were fit to a binding curve as described[85].

### H3 peptide demethylase assays

The horseradish peroxidase (HRP)-coupled assay measurements of the demethylation of H3 peptide substrate by WT or mutant GST-LSD1 and LC were performed as previously described[55,86]. In brief, 150 μM H3K4me2 or H3K4me2/K14ac peptides (aa 1–21) were added to the demethylase assay buffer containing 50 mM HEPES at pH 7.5, 0.1 mM 4-aminoantipyrine, 1 mM 3,5-dichloro-2-hydroxybenzenesulfonic acid, 0.04 mg $ml^{-1}$ HRP (Worthington Biochemical) and GST-LSD1 or LC (from 50 nM to 200 nM) at 25 °C. Absorbance changes were measured at 515 nm with 15-s intervals over a 20–40-min time course, and the product formation was quantified using the extinction coefficient of 26,000 $M^{-1}$ $cm^{-1}$. For GST-LSD1 WT and mutants, the steady-state region between 2-min and 4-min timepoints was used for analysis. For LC WT and mutants, the steady-state region between 1-min and 3-min timepoints was used for analysis for the H3K4me2 peptides, and the region between 10-min and 12-min timepoints was used for analysis for the H3K4me2K14ac peptides. Statistical analysis of the data was conducted using GraphPad Prism 9 with one-way ANOVA with Dunnett's multiple comparisons test, with a single pooled variance. Two replicates of continuous assays were performed, and each replicate included three data points within the specific time range aforementioned, resulting in a total of six data points for analysis within each group.

### Nucleosome demethylase assays

H3K4me2, H3K4me2/K14ac, H3K4me1 and H3K4me1/K14ac 185-bp nucleosomes (100 nM) were treated with LC (180 nM) or LHC (365 nM) in a buffer containing 50 mM HEPES at pH 7.5, 7 mM KCl, 2.1% glycerol and 0.2 mg $ml^{-1}$ BSA at 25 °C. At each timepoint, 18 μl of samples

were taken and quenched with 12 µl of 4× SDS sample loading buffer. Each quenched aliquot was heated for 1 min at 95 °C and resolved by SDS-PAGE using 4–20% gradient pre-cast Tris-glycine gels (Thermo Fisher Scientific). Gels were cut and transferred to nitrocellulose membranes (iBlot, Thermo Fisher Scientific), blocked with 5% BSA/1× TBST solution (20 mM Tris at pH 7.4, 150 mM NaCl and 0.1% Tween 20), washed with 1× TBST and blotted with anti-H3K4me2 (1:2,000, Abcam, ab32356), anti-H3K4me1 (1:1,000, Thermo Fisher Scientific, 710795), anti-H3K14ac (1:2,000, Millipore Sigma, 07-353), anti-H3K9ac (1:2,000, Abcam, ab32129) and anti-H3 (1:2,000, Abcam, ab1791) antibodies. After thoroughly washing the membrane with 1× TBST, anti-rabbit IgG HRP-linked antibody (1:1,000, Cell Signaling Technology, 7074S) was added and incubated. After the incubation, the membrane was washed with 1× TBST. Each membrane was visualized by ECL (Clarity, Bio-Rad) using a chemiluminescence imager (G:Box mini, SynGene) and Genesys (version 1.8.5.0). The density of each lane was quantified by ImageJ (National Institutes of Health); the data were normalized by total H3 at each timepoint; and the relative intensities (relative to $T_0$) were fitted into an exponential decay function using GraphPad Prism 9 with the constraints of $Y_0$ at 1 and plateau at 0. The extrapolated rate constant was converted to V/[E], $min^{-1}$. All measurements were done in four or six replicates, and two-way ANOVA analysis with Dunnett's multiple comparisons test, with a single pooled variance, was employed to compare statistical differences in LC complexes.

### Nucleosome deacetylase assays

H3K9ac 185-bp nucleosomes (100 nM) were treated with fluorescein-labeled WT and Y391K LHC (90 nM and 120 nM) in a buffer containing 50 mM HEPES at pH 7.5, 100 mM KCl, 100 µM inositol hexaphosphate (IP6) and 0.2 mg $ml^{-1}$ BSA at 37 °C. At each timepoint, 6.5-µl aliquots were quenched with an equivalent amount of a quenching buffer containing 20 mM EDTA and 2× SDS loading dye. Each quenched aliquot was heated for 5 min at 95 °C and resolved by SDS-PAGE using 4–20% gradient pre-cast Tris-glycine gels (Thermo Fisher Scientific). Samples were transferred to nitrocellulose membranes (iBlot, Thermo Fisher Scientific), blocked with 5% BSA/1× TBST solution, washed with 1× TBST and blotted with anti-H3K9ac (1:2,000, Abcam, ab32129) and anti-H3 (1:2,000, Abcam, ab1791) antibodies. After thoroughly washing the membrane with 1× TBST, anti-rabbit IgG HRP-linked antibody (1:1,000, Cell Signaling Technology, 7074S) was added and incubated. After the incubation, the membrane was washed with 1× TBST. Each membrane was visualized by ECL (Clarity, Bio-Rad) using a chemiluminescence imager (G:Box mini, SynGene) and Genesys (version 1.8.5.0). The intensity of each lane was quantified using ImageJ. Subsequently, the relative intensities of H3K9ac, in relation to the initial timepoint ($T_0$), were subjected to exponential decay curve fitting using GraphPad Prism 9 with the constraints of $Y_0$ at 1 and plateau at 0. The extrapolated rate constant was converted to V/[E], $min^{-1}$. All experiments were done in four technical replicates.

### MST binding assay

Fluorescein-labeled WT and Y391K LC and LHC (100 nM) were titrated with unmodified or H3K14ac 185-bp nucleosomes to measure the binding affinity ($n = 2$ technical replicates). Each measurement was assessed in the binding buffer containing 10 mM HEPES at pH 7.5, 10 mM Tris at pH 7.5, 0.5 mM DTT, 0.5 mM TCEP, 1 mg $ml^{-1}$ BSA, 0.05% Tween 20 and 50 mM NaCl at 23 °C. Nucleosomes were prepared with a two-fold serial dilution for each titration point, with the starting concentration ranging between 2.38 µM and 3.38 µM, and fluorescein-labeled LC/LHC at a final concentration of 100 nM was added, mixed and incubated for 10 min at room temperature. MST signal was measured with Monolith NT.115 (NanoTemper) with Nano BLUE detector in technical duplicates. The output was fitted into the quadratic equation to extrapolate the $K_D$ values using MO.Affinity Analysis software (version 2.3) and GraphPad Prism 9, where AB is the concentration of the CoREST complex bound to the nucleosome, $A_T$ is the total concentration for the CoREST complex and $B_T$ is the total concentration for the nucleosome.

$$AB = \frac{(A_T + B_T + K_D) - \sqrt{(A_T + B_T + K_D)^2 - 4(A_T B_T)}}{2}$$

### Cellular experiments

**Cell culture.** HEK293F cells were purchased from Thermo Fisher Scientific (FreeStyle 293-F, Gibco), and K562 cells were obtained from the American Type Culture Collection. K562 cells were authenticated by short tandem repeat profiling (Genetica). All cells were routinely tested for mycoplasma (Lonza). HEK293F cells were cultured in FreeStyle 293 Expression media (Thermo Fisher Scientific) and maintained at 37 °C and 8% $CO_2$ on an orbital shaker platform rotating at 125 r.p.m. K562 cells were cultured in RPMI 1640 media (Cytiva) supplemented with 2 mM L-glutamine, 10% FBS (Millipore Sigma), 100 U $ml^{-1}$ penicillin and 100 mg $ml^{-1}$ streptomycin (Millipore Sigma) and maintained at 37 °C and 5% $CO_2$.

**CRISPR knockin of LSD1-GFP ('parental K562' cells).** Homozygously 'edited K562' cells expressing LSD1-GFP were generated according to published procedures[87]. mEGFP followed by a 'GGGSGGGS' linker was knocked into the C-terminus of LSD1 in K562 cells. sgRNA (sg92: TGTGAGACAGATGCATTCTA) targeting the C-terminus of LSD1 was cloned into a Cas9 plasmid, PX459 (Addgene plasmid 48139), and introduced into K562 cells by electroporation according to the manufacturer's protocol (Neon transfection system; 1,350 V, 10 ms, four pulses) along with a repair vector containing the mEGFP CDS and linker flanked by 750 bp of genomic homology sequence of KDM1A C-terminus on either side. Cells were FACS sorted on a MoFlo Astrios EQ cell sorter for $GFP^+$ into 96-well plates, expanded and screened by genomic PCR followed by validation with Sanger sequencing and western blot.

**CRISPR knockin of Y391K LSD1-GFP—'edited K562' cells #1 and #2.** Heterozygously 'edited K562' cells expressing Y391K LSD1-GFP cell lines were generated according to published procedures. Y391K LSD1 point mutation was knocked into LSD1-GFP in 'parental K562' cells. Using electroporation (Neon transfection system; 1,480 V, 10 ms, four pulses), sgRNA ('677 FW' - TGGTAGAGCAAGAGTTTAAC) targeting exon 11 of LSD1 was delivered to the 'parental K562' cells as an RNP complex with Cas9 and the repair template DNA. The repair template contained a synonymous mutation at the PAM site (R384; 22 bp away from Y391) that masks the native AgeI restriction site while introducing the Y391K mutation.

Repair template: 5′-AAGATGAAATGGTAGAGCAAGAGTTTAAC-*<u>**AGA**</u>TTGCTAGAAGCTACATCT**<u>**AAG**</u>CTTAGTCATCAACTAGACTTCAATGTCCTCAATAATAAGCCTGTGTCCCT-3′; PAM site mutant is indicated by *, and Y391K mutant is indicated by **.

Cells were FACS-sorted on a BD Fusion sorter for $GFP^+$ into 96-well plates and expanded. Genomic DNA was extracted from the sorted cells, and the mutation site was amplified using PCR (GoTaq Green Master Mix, Promega). For screening, the resulting amplicon was subjected to restriction digestion with AgeI (recognizing 5′-A/CCGGT-3′; NEB), Sanger sequencing (Quintara Biosciences), and next-generation sequencing (MGH CCIB DNA Core). The AgeI restriction site was removed upon editing, allowing for efficient identification of edited cells. As a result, two clonal populations of the Y391K-'edited K562' cells were obtained ('edited K562' #1 and #2). Both K562 cells (triploid) contained two out of three alleles edited with the Y391K mutation as well as the synonymous mutation at the PAM site (R384; CGG to AGA). The last chromosome in each cell included native Y391 along with the PAM site mutation, resulting in synonymous protein expression as the parental cells while masking the PAM site against Cas9 after the edit.

The third allele with Y391: 5′-AAGATGAAATGGTAGAGCAAGAGT-TTAAC*AGATTGCTAGAAGCTACATCT**TACCTTAGTCATCAACTAGA CTTCAATGTCCTCAATAATAAGCCTGTGTCCCT-3′; PAM site mutant is indicated by *, and Y391 is indicated by **.

**Compounds.** GSK-2879552 (LSD1 inhibitor, MedChemExpress), vorinostat (HDAC inhibitor, MedChemExpress), corin (LSD1-HDAC1 dual inhibitor, MedChemExpress) and imatinib (BCR-ABL inhibitor, Ambeed) were dissolved in DMSO to make 8 mM concentrated stock solutions before the drug sensitivity assay. All inhibitor compound stock solutions were stored at −80 °C.

**Drug sensitivity assay.** 'Parental K562' cells and 'edited K562' cells #1 were seeded in 96-well plates (non-treated/flat, NEST) at a 100,000 cells per milliliter density in 195 µl of RPMI 1640 media (Cytiva) supplemented with 2 mM L-glutamine, 10% FBS (Millipore Sigma), 100 U $ml^{-1}$ penicillin and 100 mg $ml^{-1}$ streptomycin (Millipore Sigma). Then, 5 µl of the serially diluted drugs in a buffer containing 1× Dulbecco's phosphate-buffered saline (dPBS, Millipore Sigma) and 5% DMSO was added to make the final concentrations of each drug at 0 µM, 0.156 µM, 0.313 µM, 0.625 µM, 1.25 µM, 2.5 µM, 5 µM and 10 µM with 0.125% DMSO per well. After allowing cells to grow at 37 °C and 5% $CO_2$ for 96 h, 90 µl of culture was transferred to black-bottom 96-well plates (Thermo Fisher Scientific) and mixed with 10 µl of Alamar blue reagent (A50100, Thermo Fisher Scientific), mixed and incubated at 37 °C for 1 h. Then, each well's fluorescence was measured (excitation/emission at 560/590 nm) using a plate reader (BioTek Cytation 5, Agilent). Three biological replicates were performed for parental and edited K562 cells. The percentage of cell viability was determined by multiplying the fluorescence intensity ratio of the treated cells to the DMSO control by 100. These percentages were fitted into the [Inhibitor] vs. response – Variable slope (four parameters) function of GraphPad Prism 9 with the constraints of the Y-intercept at 100 and plateau at 0 to extrapolate half-maximal inhibitory concentration ($IC_{50}$) ± s.e.m. values.

**Growth rate measurement.** 'Parental K562' cells and 'edited K562' cells #1 and #2 were seeded in 12-well plates (non-treated/flat, NEST) at a 200,000 cells per milliliter density. Every 24 h, 10 µl of culture was mixed with 10 µl of dPBS and 20 µl of Trypan blue, and the cell density was measured using hemocytometer-based cell counting under the microscope. After the initial 24 h of the adaptation period, cells entered the log phase between 24-h and 72-h timepoints. After reaching confluency at the 96th hour, cell densities in the log phase were used to extrapolate the doubling rate using the exponential (Malthusian) growth function of GraphPad Prism 9. Three replicates of the independently split cells were assessed for parental and edited K562 cells #1 and #2 ($n = 3$ for each) to calculate the doubling time (mean ± s.e.m.), and the $P$ values were calculated using ordinary one-way ANOVA with Dunnett's multiple comparisons test, with a single pooled variance.

**Cell lysate western blotting.** In total, 1.5 million 'parental K562' cells and 'edited K562' cells #1 and #2 were harvested via centrifugation (500$g$ for 5 min), and the cell pellets were resuspended in the RIPA lysis buffer containing 50 mM Tris at pH 7.4, 150 mM NaCl, 1% NP-40, 0.5% deoxycholate and 0.1% SDS. Protein contents were estimated by the BCA protein assay kit (Thermo Fisher Scientific) using the absorbance measurement at 562 nm (BioTek Cytation 5, Agilent). Then, 13 µg of total protein from each group was resolved by SDS-PAGE (4–20% gradient Tris-Glycine pre-cast gel, Thermo Fisher Scientific). Gels were excised and transferred to nitrocellulose membranes (iBlot, Thermo Fisher Scientific), blocked with 1× TBST containing 5% milk (Nestle), washed with 1× TBST and blotted with anti-GFP (1:250, Santa Cruz Biotechnology, sc-9996), anti-LSD1 (1:1,000, Bethyl Laboratories, A300-215A) and anti-GAPDH (1:2,000, Cell Signaling Technology, 2118S) antibodies. After thoroughly washing the membrane with 1× TBST, an anti-rabbit or anti-mouse IgG HRP-linked antibody (1:1,000, Cell Signaling Technology, 7074S and 7076S) was added. After incubating for 1 h at room temperature on an orbital shaker, the membrane was washed with 1× TBST. Each membrane was visualized by the ECL western blotting substrate (Clarity, Bio-Rad) using a chemiluminescence imager (G:Box mini, SynGene) and Genesys (version 1.8.5.0).

**Transcriptomic profiling.** *Sample preparation*. Approximately 4 million cells of the 'parental K562' cells and two clonal edited variants ('edited K562' cells #1 and #2) were collected through centrifugation at 500$g$ for 5 min. After this, cells were washed with dPBS (Millipore Sigma) by resuspending and centrifuging them at 500$g$ for 5 min. 'Parental K562' cells were split into three different wells for technical triplicates ($n = 3$), and 'edited K562' cells #1 and #2 were split into two different wells for technical duplicates (thus, technical duplicates of the biological duplicates were prepared).

*RNA extraction and quality assessment*. The harvested cell pellets were flash frozen and sent to MedGenome for RNA extraction, sequencing and analysis. Extractions were carried out using a Maxwell RSC simplyRNA Cells Kit (Promega, no. 1390). To verify RNA quality, RNA qualitative and quantitative assessment was done using Qubit 3.0 (Life Technologies, Q33216) and TapeStation 4200 (Agilent), confirming that the RNA integrity number equivalent (RINe) score exceeded 8.9 for all samples.

*Library preparation and sequencing*. The poly(A)-containing mRNA molecules were purified using poly(T) oligo-attached magnetic beads. The purified mRNA was then converted to cDNA, and libraries were prepared using the TruSeq Stranded mRNA Library Prep Kit (no. 20020595) according to the manufacturer's protocol. Libraries were sequenced for PE150 cycles to a depth of 40 million paired reads using an Illumina NovaSeq 6000.

*Data quality control and processing*. Initial data quality assessment was performed using FastQC (version 0.11.8)[88]. Adapter trimming was accomplished using fastq-mcf (version 1.05)[89] and cutadapt (version 2.5)[90]. Bowtie 2 (version 2.5.1)[91] was employed to eliminate unwanted sequences, including mitochondrial genome sequences, ribosomal RNAs and adapter sequences. The paired-end reads were then aligned to the reference human genome February 2009 release (GRCh37/hg19) obtained from the UCSC database. The chromosome FASTA file was sourced from http://hgdownload.soe.ucsc.edu/goldenPath/hg19/bigZips/chromFa.tar.gz and the GTF file from ftp://ftp.ensembl.org/pub/release75/gtf/homo_sapiens/Homo_sapiens.GRCh37.75.gtf.gz. Alignment was executed using STAR (version 2.7.3a)[92].

*Expression estimation and analysis*. Aligned reads were used to estimate gene expression levels. Raw read counts were determined using HTSeq (version 0.11.2)[93] and normalized using DESeq2. Cufflinks software (version 2.2.1)[94] was employed to estimate gene expression in fragments per kilobase per million (FPKM) units. Rigorous quality control was conducted using RNA-SeQC (version 1.1.8)[95], RSeQC (version 3.0.1)[96] and MultiQC (version 1.7)[97].

*Differential expression analysis*. DESeq2 (R Bioconductor package)[98] facilitated differential expression analysis. Specifically, it compared the combinations of parental cells (control) and edited cells, assessing the number of significantly differentially expressed protein-coding genes without fold change shrinkage. The results of upregulated and downregulated genes in the 'edited K562' cells were visualized through a volcano plot with the significance ($-\log_{10} P$ value) and $\log_2$ fold changes, relative to the 'parental K562' cells.

*Gene Ontology analysis*. clusterProfiler (R Bioconductor package, version 4.9.2)[56] was used to study functional enrichment of the 498 downregulated genes in Gene Ontology[99] terms. The top seven functional pathways with significance ($P \leq 0.05$) and the number of annotated genes were visualized.

**Chromatin profiling using CUT&RUN.** *Sample preparation*. Approximately 3 million cells per replicate of the 'parental K562' cells and 'edited K562' cells #1 were harvested through centrifugation (600*g* for 3 min), resuspended in the wash buffer containing 20 mM HEPES at pH 7.5, 150 mM NaCl, 0.5 mM spermidine and protease inhibitor cocktail (Roche cOmplete EDTA-free) and aliquoted into five microcentrifuge tubes and washed again in 1 ml of dPBS. Each tube was used for the following antibody incubation later: rabbit IgG (Millipore Sigma, I5006), anti-GFP (Abcam, ab290), anti-H3K4me1 (EpiCypher, 13-0040), anti-H3K4me2 (Abcam, ab32356) and anti-H3K14ac (Millipore Sigma, 07-353). Duplicate samples were prepared where each replicate came from independently split cell cultures. From this step, the previously reported CUT&RUN sample preparation method[69,100] and manufacturer's protocols were employed (EpiCypher). In brief, ConA beads were first activated by incubating in the bead activation buffer containing 20 mM HEPES at pH 7.9, 10 mM KCl, 1 mM $CaCl_2$ and 1 mM $MnCl_2$. In each tube containing approximately $6 \times 10^5$ cells, 10 µl of activated ConA beads (Bangs Laboratories, BP531) were added, resuspended and washed with dPBS and the wash buffer via centrifugation (500 g for 5 min). ConA bead binding was inspected under the microscope. ConA bead-bound cells were incubated with the antibodies in the 100-µl antibody binding buffer (wash buffer with 0.01% digitonin and 2 mM EDTA) overnight at 4 °C on a nutator (Clay Adamas, 421105). For each set of five tubes, 1 µg of rabbit IgG and 2 µl of anti-GFP, H3K4me1, H3K4me2 and H3K14ac were added. The next day, the beads were washed with the cell permeabilization buffer (wash buffer with 0.01% digitonin), and 2.5 µl of pAG-MNase (EpiCypher, 15-1016) was added, mixed and washed with the cell permeabilization buffer. Chromatin digestion reaction was initiated by adding 1 µl of $CaCl_2$ to each sample and nutating them for 2 h at 4 °C. The reaction was quenched with the stop buffer containing 340 mM NaCl, 20 mM EDTA, 4 mM EGTA, 50 µg $ml^{-1}$ RNase A and 50 µg $ml^{-1}$ glycogen. Next, 0.5 ng of *E. coli* spike-in DNA was added to each sample (EpiCypher, 18-1401), incubated at 37 °C for 10 min on a thermomixer (LabNet) at 500 r.p.m., centrifuged down at 16,000*g* for 1 min and placed on a magnet stand to take the supernatant containing DNA. The samples were purified using a spin column (Zymo Research, D4003), followed by quantification using Qubit (Thermo Fisher Scientific) and TapeStation (Agilent D1000). Then, 0.5 ng to 11 ng of the purified DNA (rabbit IgG having the lowest yield) was used for the sequencing library preparation. Samples containing fragmented DNA were mixed with the library prep reagents (NEBNext Ultra II DNA Library Prep Kit for Illumina, NEB, E7645S), ligated with the adaptors, cleaned up with the AMPure XP beads (Beckman Coulter, A63880), barcoded with the primers with the NEB index primers (NEBNext Multiplex Oligos for Illumina Index Primers Set 1 to 4, NEB) through 14 PCR cycles. The PCR products were finally cleaned up with the AMPure XP beads and quantified with the NanoDrop, yielding 14–80 ng $µl^{-1}$ products.

*Next-generation sequencing*. Libraries were quality checked by the TapeStation (Agilent, D1000) and pooled and sent to the next-generation sequencing facility (MGH NextGen Sequencing Core) and sequenced using an Illumina NextSeq 2000 50 PE P2 flow cell (Illumina).

*Data processing*. CUT&RUN data processing was mostly adopted from the established protocol[101]. FastQ data files were initially processed by fastp (version 0.23.4)[102], and quality checks were assessed by FastQC (version 0.11.9)[88]. The paired-end reads were aligned to the human reference genome (GRCh38 assembly) using Bowtie 2 (version 2.4.4)[91] as well as the *E. coli* genome (K-12 MG1655). For GRch38 assembly, the following arguments were used: bowtie2 -p 8–end-to-end–very-sensitive–no-mixed–no-discordant–phred33 -I 10 -X 700 -x /path-to-genome/ /GRCh38_noalt_as. For *E. coli*, the following arguments were used: bowtie2 -p 8–end-to-end–very-sensitive–no-overlap–no-dovetail–no-mixed–no-discordant–phred33 -I 10 -X 700 -x /path/Escherichia_coli_K_12_MG1655/NCBI/2001-10-15/Sequence/Bowtie2Index/genome. The SAM output files were converted to the BAM format using SAMtools (version 1.17)[103]. For peak calling, BAM files were first converted to bed and bedgraph files using bedtools (version 2.31.0)[104], and SEACR relaxed mode[105] was used to call peaks by normalizing against the IgG control bedgraph files for each replicate. For CUT&RUN signal visualization, BAM files were sorted by the coordinates and indexed using SAMtools (version 1.17)[103], and the sorted BAM files were converted to the bigwig files using the bamCoverage application from the deepTools package (version 3.5.2)[106]. RPGC normalization mode using the effective genome size of 3049315783 bp and bin size of 1 was applied (bamCoverage -b '$file' -p 10–binSize 1–normalizeUsing RPGC–effectiveGenomeSize 3049315783–extendReads–outFileFormat bigwig). The bigwig files were visualized by using Integrative Genomics Viewer (IGV)[107] and analyzed using the computeMatrix and plotProfile applications from the deepTools package (version 3.5.2)[106].

*Data analysis*. To understand the correlation between the Y391K LSD1-mediated gene repression and the unique histone mark distribution within or nearby the 498 downregulated genes, we selected 498 unaffected control genes for comparison. Using the BioMart R bioconductor package (version 2.57.1)[108], 498 human protein-coding genes were randomly selected from the 'hsapiens_gene_ensembl' database. In total, 498 downregulated genes and 256 upregulated genes were excluded from counting to avoid redundancy.

To assess the CUT&RUN and ChIP-seq signal comparisons between the downregulated versus unaffected control gene bodies, the genomic coordinates of each gene were obtained by the UCSC Table browser in the hg38 format[109]. CUT&RUN bigwig files were generated using the data processing methods aforementioned, and the ChIP-seq data were downloaded from the ENCODE portal[110,111] (https://www.encodeproject.org/) with the following identifiers: ENCFF665RDD (for H3K27me3), ENCFF465GBD (for H3K27ac), ENCFF334HSS (for H3K79me2), ENCFF974IOO (for SUZ12) and ENCFF163LOW (for EZH2). Using the bigwig files, signals in the 498 downregulated and unaffected control gene bodies ±20-kb regions from the parental and edited K562 cells were analyzed using the computeMatrix and plotProfile applications from the deepTools package (version 3.5.2)[106]. Gene bodies were scaled to 40 kb in size.

To assess signal comparisons at the LSD1-occupied regions associated with the 498 downregulated and unaffected control genes, ChIPSeeker (R Bioconductor package)[70] was employed to annotate the LSD1 peaks across the genes. The regions associated with the 498 downregulated and unaffected control genes were further filtered using bedtools (version 2.31.0)[104]. Using the bigwig files, signals at the LSD1 peaks (±5 kb) across the 498 downregulated and unaffected control genes of the 'parental K562' cells were analyzed using the computeMatrix and plotProfile applications from the deepTools package (version 3.5.2)[106].

To examine how well the anti-GFP CUT&RUN reads from the parental and edited K562 cells represent the genomic distribution of LSD1-GFP, our anti-GFP CUT&RUN peak files were compared with the LSD1 ChIP-seq reads deposited in the ENCODE portal (ENCFF054XCG) using bedtools (version 2.31.0)[104].

**LHC modeling.** Modeling of the LHC complex was adapted from the previous review publication[43] (https://doi.org/10.7910/DVN/ULB7JL). In brief, truncated LC complex (LSD1 aa 171–852 and CoREST1 aa 297–482) and HC (HDAC1 aa 8–376 and CoREST1 aa 95–240) models generated

by AlphaFold2 (version 2.2.0)[112,113] were fitted into the cryo-EM density (EMD-10629)[31] and refined with addition of the co-factors and metal ions followed by full atom relaxation with Rosetta3 (version 3.13)[114–116].

**Unique biological materials.** Unique biological materials, including edited cell lines and plasmids, are available upon reasonable request to the corresponding author.

### Reporting summary

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

## Data availability

Structure factors and atomic coordinates have been deposited in the Protein Data Bank with IDs 8Q1G, 8Q1H and 8Q1J. RNA-seq data of the parental and edited K562 cells have been deposited with Gene Expression Omnibus (GEO) accession code GSE243427. CUT&RUN data of the parental and edited K562 cells have been deposited with GEO accession code GSE243231. CUT&RUN data analyzed using deepTools are deposited at Harvard Dataverse (https://doi.org/10.7910/DVN/AUDINC). Source data are provided with this paper.

## Code availability

The codes used for processing CUT&RUN data can be found in the Supplementary Note section of the Supplementary Information. Processing scripts for CUT&RUN analysis using deepTools are available at Harvard Dataverse (https://doi.org/10.7910/DVN/AUDINC).

## References


76. Szewczuk, L. M. et al. Mechanistic analysis of a suicide inactivator of histone demethylase LSD1. *Biochemistry* **46**, 6892–6902 (2007).
77. Portolano, N. et al. Recombinant protein expression for structural biology in HEK 293F suspension cells: a novel and accessible approach. *J. Vis. Exp.* e51897 (2014).
78. Luger, K., Rechsteiner, T. J. & Richmond, T. J. Preparation of nucleosome core particle from recombinant histones. In *Methods in Enzymology*, Vol. 304, 3–19 (Academic Press, 1999).
79. Luger, K., Rechsteiner, T. J. & Richmond, T. J. Expression and purification of recombinant histones and nucleosome reconstitution. In *Chromatin Protocols* (ed Becker, P. B.) 1–16 (Humana Press, 1999).
80. Forneris, F. et al. A highly specific mechanism of histone H3-K4 recognition by histone demethylase LSD1. *J. Biol. Chem.* **281**, 35289–35295 (2006).
81. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
82. Agirre, J. et al. The CCP4 suite: integrative software for macromolecular crystallography. *Acta Crystallogr. D Struct. Biol.* **79**, 449–461 (2023).
83. Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D Struct. Biol.* **75**, 861–877 (2019).
84. Casañal, A., Lohkamp, B. & Emsley, P. Current developments in Coot for macromolecular model building of electron cryo-microscopy and crystallographic data. *Protein Sci.* **29**, 1055–1064 (2020).
85. Pilotto, S. et al. Interplay among nucleosomal DNA, histone tails, and corepressor CoREST underlies LSD1-mediated H3 demethylation. *Proc. Natl Acad. Sci. USA* **112**, 2752–2757 (2015).
86. Prusevich, P. et al. A selective phenelzine analogue inhibitor of histone demethylase LSD1. *ACS Chem. Biol.* **9**, 1284–1293 (2014).
87. Ran, F. A. et al. Genome engineering using the CRISPR–Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).
88. Andrews, S. FastQC: a quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
89. Aronesty, E. Comparison of sequencing utility programs. *Open Bioinform. J.* **7**, 1–8 (2013).
90. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10–12 (2011).
91. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
92. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
93. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
94. Trapnell, C. et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010).
95. DeLuca, D. S. et al. RNA-SeQC: RNA-seq metrics for quality control and process optimization. *Bioinformatics* **28**, 1530–1532 (2012).
96. Wang, L., Wang, S. & Li, W. RSeQC: quality control of RNA-seq experiments. *Bioinformatics* **28**, 2184–2185 (2012).
97. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).
98. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
99. Ashburner, M. et al. Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
100. Meers, M. P., Bryson, T. D., Henikoff, J. G. & Henikoff, S. Improved CUT&RUN chromatin profiling tools. *eLife* **8**, e46314 (2019).
101. Zheng, Y., Ahmad, K. & Henikoff, S. CUT&Tag data processing and analysis tutorial. *protocols.io* https://doi.org/10.17504/protocols.io.bjk2kkye (2020).
102. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
103. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021).
104. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
105. Meers, M. P., Tenenbaum, D. & Henikoff, S. Peak calling by Sparse Enrichment Analysis for CUT&RUN chromatin profiling. *Epigenetics Chromatin* **12**, 42 (2019).
106. Ramírez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
107. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
108. Durinck, S. et al. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics* **21**, 3439–3440 (2005).
109. Karolchik, D. et al. The UCSC Table Browser data retrieval tool. *Nucleic Acids Res.* **32**, D493–D496 (2004).
110. Luo, Y. et al. New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. *Nucleic Acids Res.* **48**, D882–D889 (2020).
111. Sloan, C. A. et al. ENCODE data at the ENCODE portal. *Nucleic Acids Res.* **44**, D726–D732 (2016).
112. Evans, R. et al. Protein complex prediction with AlphaFold-Multimer. Preprint at *bioRxiv* https://doi.org/10.1101/2021.10.04.463034 (2022).
113. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
114. Bender, B. J. et al. Protocols for molecular modeling with Rosetta3 and RosettaScripts. *Biochemistry* **55**, 4748–4763 (2016).

115. Conway, P., Tyka, M. D., DiMaio, F., Konerding, D. E. & Baker, D. Relaxation of backbone bond geometry improves protein energy landscape modeling. *Protein Sci.* **23**, 47–55 (2014).
116. Leaver-Fay, A. et al. Rosetta3: an object-oriented software suite for the simulation and design of macromolecules. In *Methods in Enzymology*, Vol. 487 (eds Johnson, M. L. & Brand, L.) 545–574 (Academic Press, 2011).

## Acknowledgements

We acknowledge financial supprt from the National Institutes of Health (GM62437 and GM149229 to P.A.C., GM126944 to M.I.K., 1DP2GM137494 to B.B.L. and AG068179 to S.B.); the National Science Foundation (2127882 to P.A.C.), the Leukemia & Lymphoma Society (to P.A.C.); MUR (FISR2019_00374 MeDyCa to A.M.); the American Heart Association (Postdoctoral Fellowship Award 826614 to K.L.); the American Cancer Society (PF20-105-01-DMC to S.D.W.); and the Charles A. King Trust Postdoctoral Research Fellowship (to S.D.W.). We would like to thank the Cole laboratory members for helpful advice and stimulating discussions. We also thank MedGenome for RNA-seq experiments and analysis. We thank C. Bahl (AI Proteins, Inc.) and Meiler laboratory members (Vanderbilt University) for helpful advice in designing LSD1 mutants. We appreciate R. Chivukula for technical advice on CRISPR–Cas9 and Alani laboratory members (Boston University) and the Harvard Chan Bioinformatics Core for helpful advice on bioinformatics and cellular data interpretation.

## Author contributions

All listed authors performed experiments and analyzed data. K.L., E.N., S.E.D-C., K.N. and B.I. performed enzymology. K.L., E.N., S.E.D.-C., S.D.W. and Z.A.W. prepared nucleosomes containing semi-synthetic histones. M.B., J.C. and A.M. conducted crystallographic analysis. M.B., J.C. and A.M. performed SNAIL peptide-binding measurements and analysis. K.L. and H.J. conducted fluorophore-labeled CoREST complex preparation and MST binding affinity measurements. K.L., A.L.W. and B.L. conducted CRISPR–Cas9 knockin experiments. K.L. and Z.D. conducted CUT&RUN profiling experiments. K.L., Z.D., M.I.K., S.B. and P.A.C. performed genomic analysis. The manuscript was prepared by K.L. and P.A.C., with input from all authors.

## Competing interests

P.A.C. is a co-founder of Acylin Therapeutics, which is involved in developing epigenetic agents, and has been a consultant for the pharmaceutical companies AbbVie and Constellation. He also is a co-inventor on US patent 11,565,994 B2 that concerns LSD1 and CoREST complex inhibitors. B.B.L. has received research funding from Eisai and AstraZeneca and is a shareholder and member of the scientific advisory board of Light Horse Therapeutics. The other authors declare no competing interests.

## Additional information

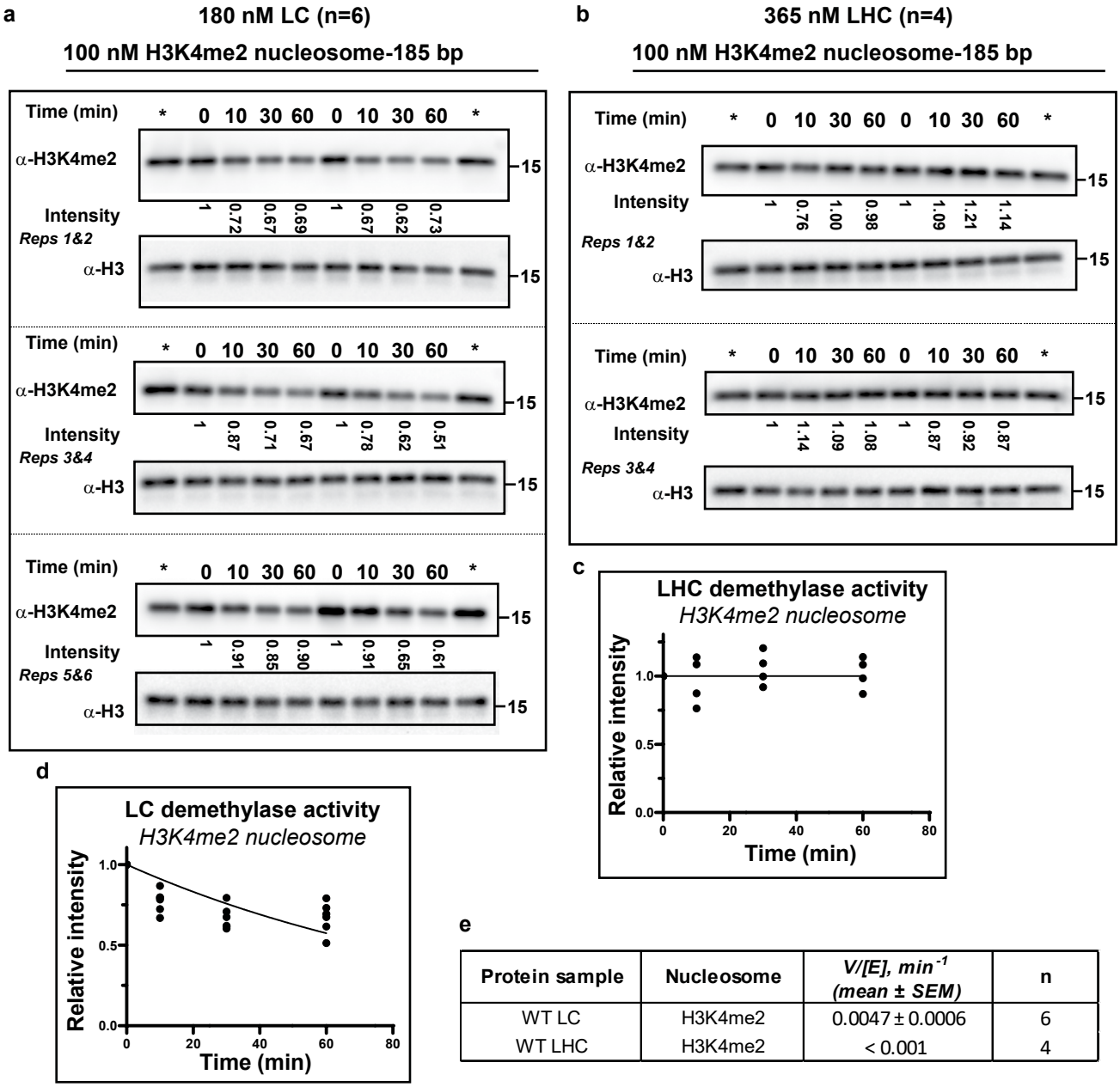
**Extended data** is available for this paper at https://doi.org/10.1038/s41589-024-01671-9.

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41589-024-01671-9.

**Correspondence and requests for materials** should be addressed to Samuel Beck, Andrea Mattevi or Philip A. Cole.

**Peer review information** *Nature Chemical Biology* thanks Tatiana Kutateladze and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

https://doi.org/10.1038/s41589-024-01671-9

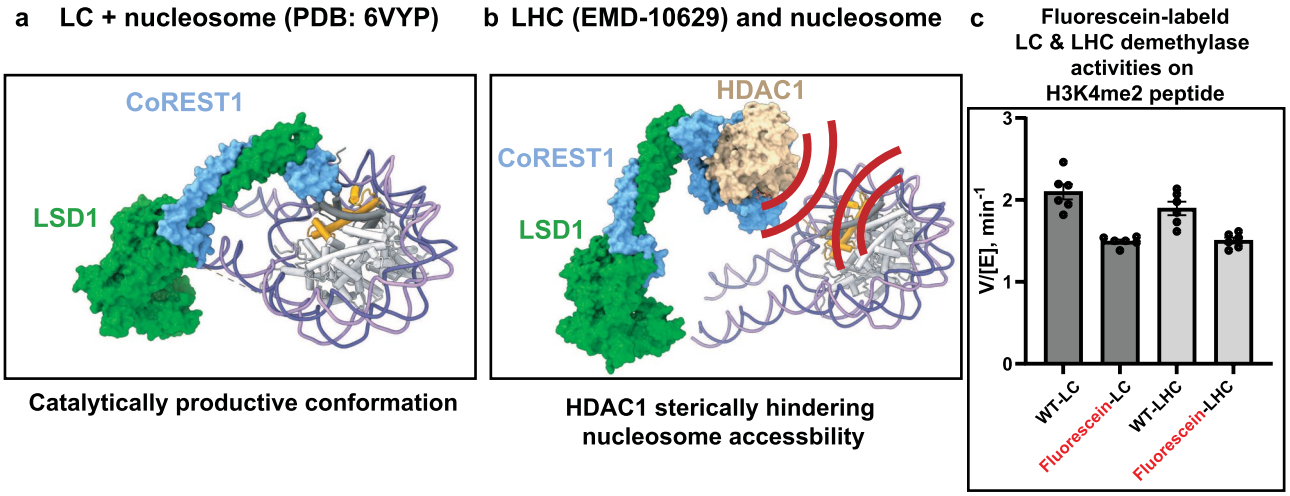



e

| Protein sample | Nucleosome | *V/[E], min$^{-1}$ (mean ± SEM)* | n |
|---|---|---|---|
| WT LC | H3K4me2 | 0.0047 ± 0.0006 | 6 |
| WT LHC | H3K4me2 | < 0.001 | 4 |

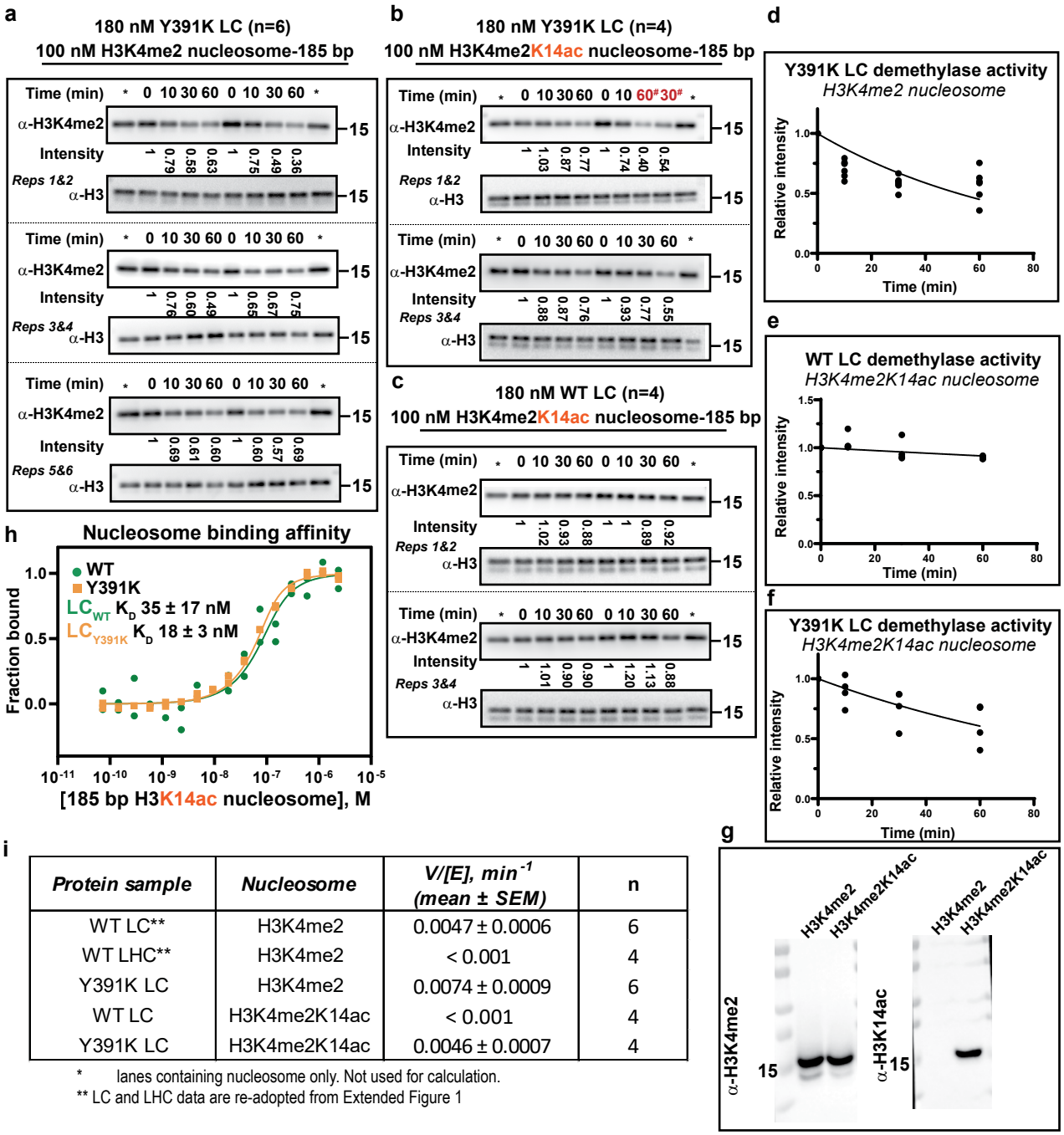
**Extended Data Fig. 1 | Measurement of Nucleosome Demethylase Activity via Western Blotting for LSD1-CoREST1 (LC) and LSD1-CoREST1-HDAC1 (LHC).** (**a**) Demethylase activity of LC on H3K4me2 nucleosomes. LC at a concentration of 180 nM was incubated with 100 nM of 185 bp H3K4me2 nucleosomes, and changes in H3K4me2 levels were tracked over a 60-minute time frame. (**b**) Evaluation of demethylase activity for LHC. LHC, present at a concentration of 365 nM, was incubated with 100 nM of 185 bp H3K4me2 nucleosomes, and changes in H3K4me2 levels were tracked over a 60-minute time frame. In both (a) and (b), the anti-H3K4me2 signal at each time point was normalized by anti-H3. Lanes containing only nucleosomes (designated by *) were excluded from the rate calculations. (**c**) and (**d**) illustrate the relative intensities obtained from (b) and (a), subjected to fitting into an exponential decay equation, featuring constraints of $Y_0$ at 1 and plateau at 0. In (d), the H3K4me2 level appears to plateau after 30 min. (**e**) V/[E] ($min^{-1}$) values from (c) and (d) were extrapolated (mean ± SEM).

https://doi.org/10.1038/s41589-024-01671-9




**Extended Data Fig. 2 | Potential Disruption of LC Complex's Nucleosome Binding by HDAC1 Interaction.** (**a**) Illustration of the nucleosome-bound, demethylase-active configuration of LC as observed in the crystal structure (PDB: 6VYP). The SANT2 domain of CoREST1 (light blue) interfaces with the globular regions of H3 (yellow), H4 (dark gray), and DNA (purple), facilitating engagement with the nucleosome. (**b**) AlphaFold2-generated model of LHC fitted into the EM density map (EMD-10629). HDAC1, bound to the ELM2-SANT domain of CoREST1, remains proximal to the tower domain of LSD1 and the SANT2 domain of CoREST1. This interaction can potentially hinder CoREST1-nucleosome interaction, thereby preventing LHC from adopting the demethylase-active conformation. (**c**) Despite comprised nucleosome-binding, the LSD1 of LHC remains catalytically active. LC (200 nM) or LHC (100 nM) have similar demethylase activities toward H3K4me2 peptides (aa 1-21; 150 μM), even when CoREST1 is N-terminally tagged with fluorescein (~70% active compared with their untagged counterparts). HRP-coupled peptide demethylase activity assay was employed for evaluation. A total of six data points from the duplicates of continuous assays were used to obtain the V/[E] ($min^{-1}$) values, presented as mean ± SEM with error bars.

https://doi.org/10.1038/s41589-024-01671-9



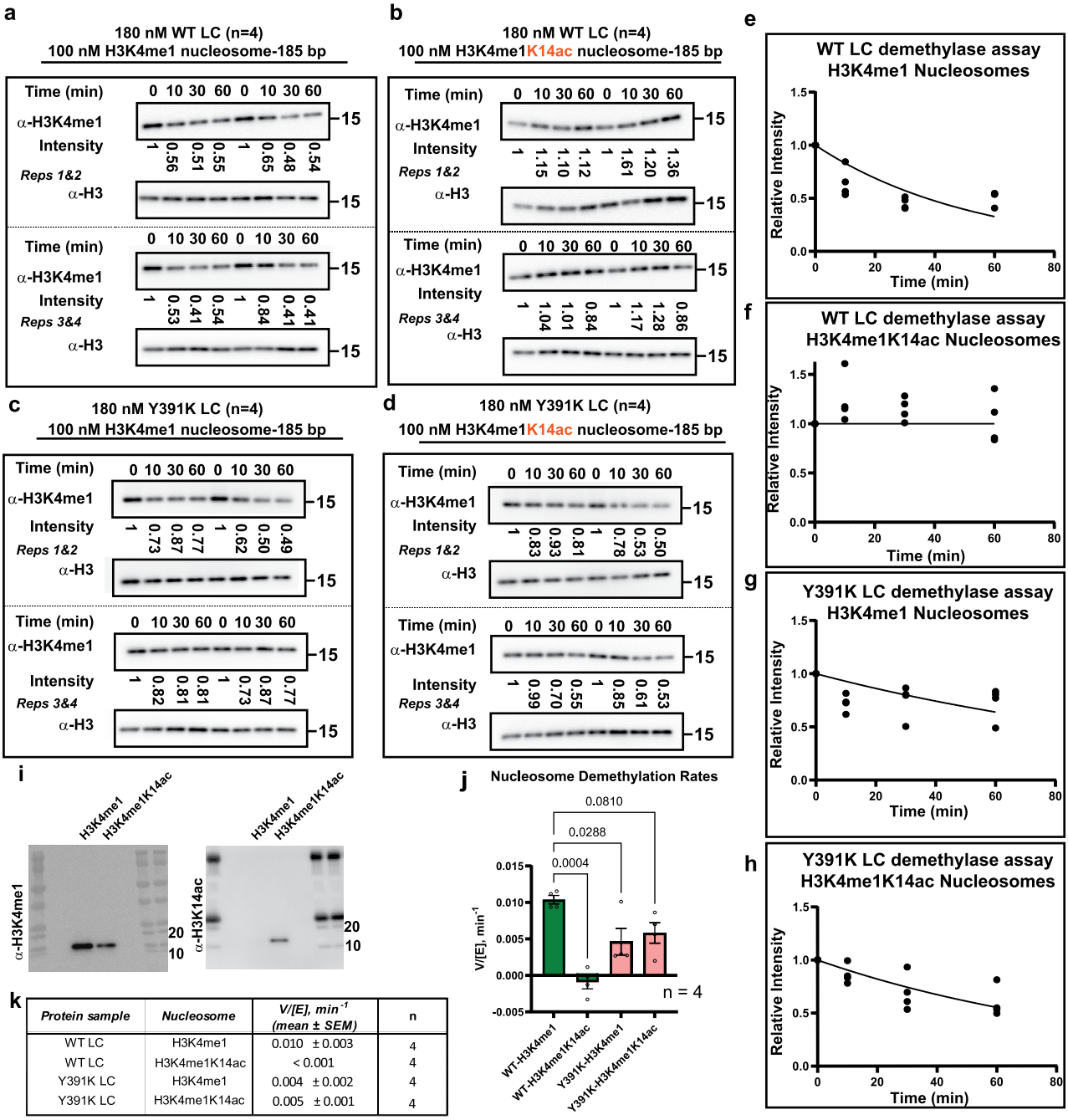

i

| *Protein sample* | *Nucleosome* | *V/[E], min$^{-1}$ (mean ± SEM)* | n |
|---|---|---|---|
| WT LC** | H3K4me2 | 0.0047 ± 0.0006 | 6 |
| WT LHC** | H3K4me2 | < 0.001 | 4 |
| Y391K LC | H3K4me2 | 0.0074 ± 0.0009 | 6 |
| WT LC | H3K4me2K14ac | < 0.001 | 4 |
| Y391K LC | H3K4me2K14ac | 0.0046 ± 0.0007 | 4 |

* lanes containing nucleosome only. Not used for calculation.
** LC and LHC data are re-adopted from Extended Figure 1

**Extended Data Fig. 3 | Analysis of H3K4me2 and H3K4me2K14ac Nucleosome Demethylase Activity for WT and Y391K LC.** (**a**) Y391K LC demethylase activity on H3K4me2 nucleosomes. Y391K LC at a concentration of 180 nM was subjected to a 60 minute incubation with 100 nM of 185 bp H3K4me2 nucleosomes, and changes in H3K4me2 levels were monitored. (**b**) Y391K LC demethylase activity targeting H3K4me2K14ac nucleosomes. Similar to (a), 180 nM Y391K LC was incubated with 100 nM of 185 bp H3K4me2K14ac nucleosomes, and changes in H3K4me2 levels were tracked over 60 minutes. (**c**) WT LC demethylase activity targeting H3K4me2K14ac nucleosomes. 180 nM WT LC was incubated with 100 nM of 185 bp H3K4me2K14ac nucleosomes, and changes in H3K4me2 levels were tracked over 60 minutes. Panels (**d-f**) illustrate the relative intensities obtained from (a-c), subjected to fitting into an exponential decay equation that includes constraints of $Y_0$ at 1 and plateau at 0. In (f), the H3K4me2 level remains almost constant after $T_{30min}$, possibly due to product inhibition. (**g**) Western blots with anti-H3K4me2 and anti-H3K14ac antibodies. These images display the signals obtained for H3K4me2 and H3K4me2K14ac nucleosomes used in the demethylase assays, respectively, in one replicate (n = 1). (**h**) Microscale thermophoresis was used to measure the binding affinities of N-terminally fluorescein-labeled WT (green) and Y391K LC (pink) to the H3K14ac nucleosome (n = 2). Both complexes tightly engage the nucleosome with comparable binding affinities. (**i**) V/[E] (min$^{-1}$) values from (d-f) were extrapolated (mean ± SEM).

https://doi.org/10.1038/s41589-024-01671-9

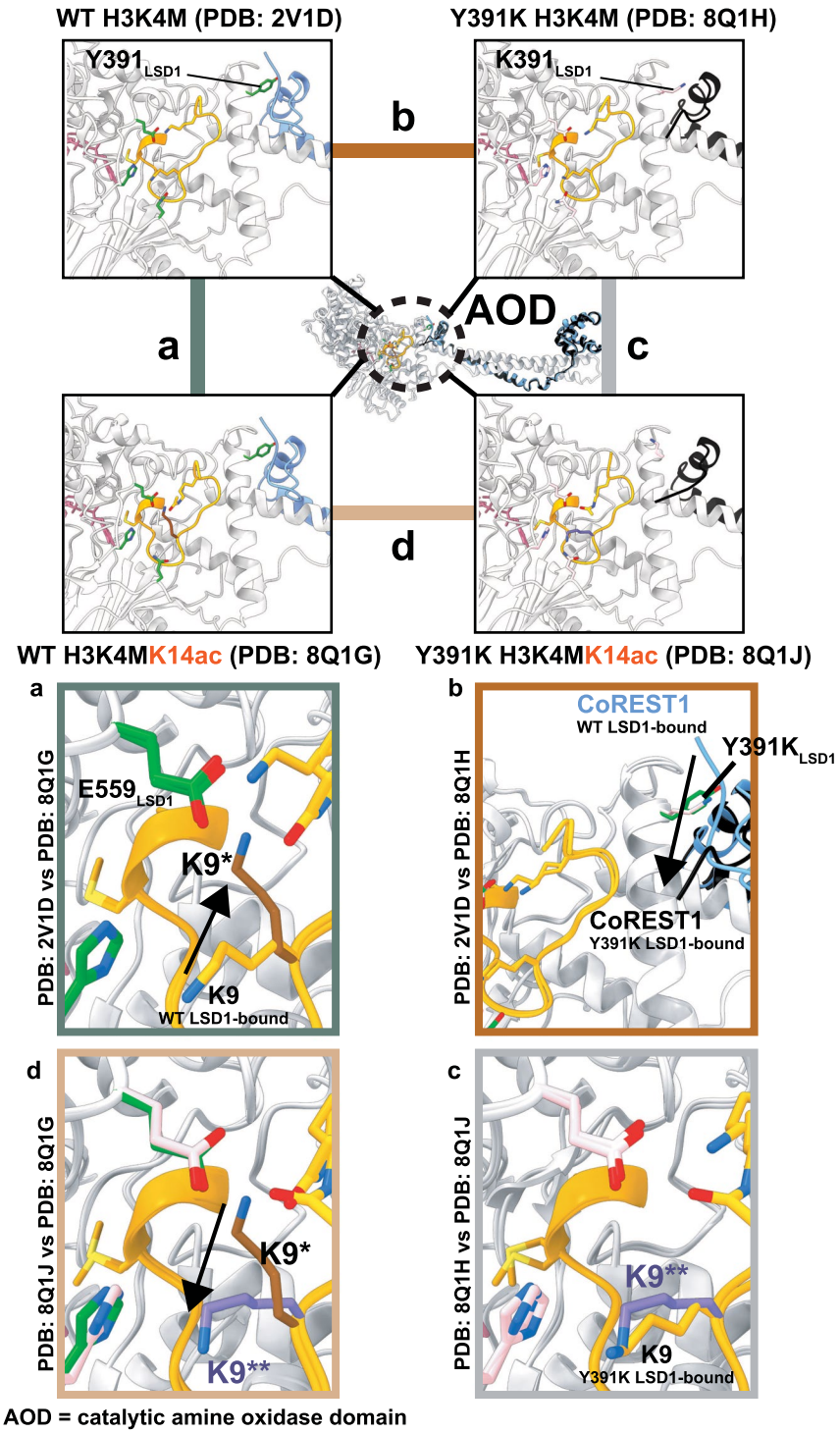



k

| Protein sample | Nucleosome | V/[E], min$^{-1}$ (mean ± SEM) | n |
|---|---|---|---|
| WT LC | H3K4me1 | 0.010 ± 0.003 | 4 |
| WT LC | H3K4me1K14ac | < 0.001 | 4 |
| Y391K LC | H3K4me1 | 0.004 ± 0.002 | 4 |
| Y391K LC | H3K4me1K14ac | 0.005 ± 0.001 | 4 |

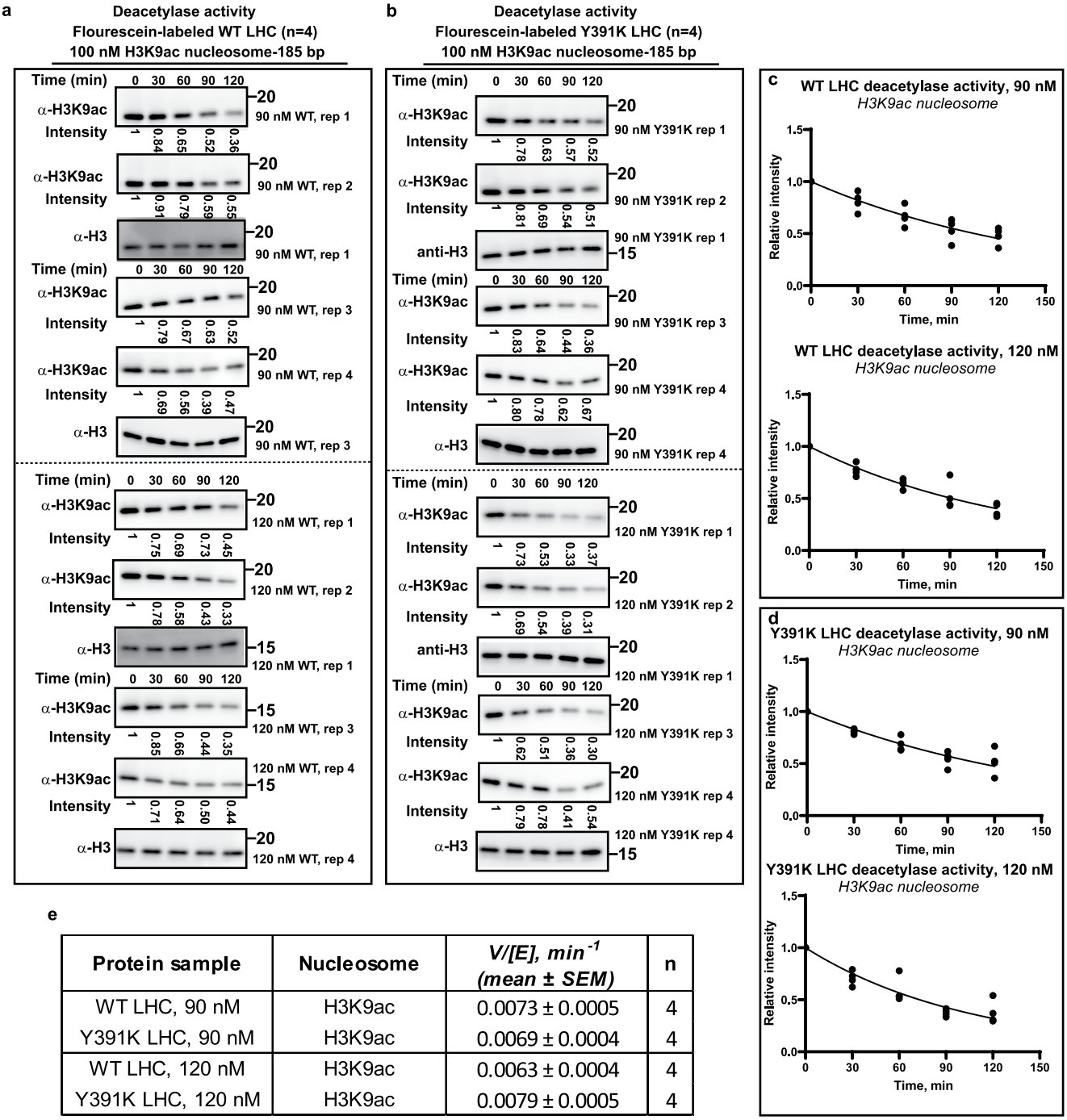
**Extended Data Fig. 4 | Analysis of H3K4me1 and H3K4me1K14ac Nucleosome Demethylase Activity for WT and Y391K LC.** (**a**) WT LC demethylase activity on H3K4me1 nucleosomes. WT LC at a concentration of 180 nM was subjected to a 60 minute incubation with 100 nM of 185 bp H3K4me1 nucleosomes, and changes in H3K4me1 levels were monitored. (**b**) WT LC demethylase activity targeting H3K4me1K14ac nucleosomes. Assays conditions were identical to (a). (**c**) Y391K LC demethylase activity on H3K4me1 nucleosomes. Assays conditions were identical to (a). (**d**) Y391K LC demethylase activity targeting H3K4me1K14ac nucleosomes. Assay conditions were identical to (a). Panels (**e**-**h**) illustrate the relative intensities obtained from (a-c), subjected to fitting into an exponential decay equation that includes constraints of $Y_0$ at 1 and plateau at 0. In (**e**) and (**f**), the H3K4me1 level remains almost constant after $T_{30min}$, possibly due to product inhibition. (**i**) Western blots with anti-H3K4me1 and anti-H3K14ac antibodies. These images display the signals obtained for H3K4me1 and H3K4me2K14ac nucleosomes used in the demethylase assays, respectively, in one replicate (n = 1). (**j**) Bar plot showing the demethylase activities of WT LC and Y391K LC towards H3K4me1 and H3K4me1/K14ac nucleosomes (Two-way ANOVA; mean ± SEM; n = 4; *p* values are indicated above each comparison group). (k) V/[E] (min$^{-1}$) values from (**e**-**h**) were extrapolated (mean ± SEM).

https://doi.org/10.1038/s41589-024-01671-9




**Extended Data Fig. 5 | Structural comparison of WT LC and Y391K LC in complex with H3K4M and H3K4M/K14ac peptides.** The active site of LC with the peptides is magnified (top), highlighting key residues and CoREST1's conformation in proximity. (**a**) WT LC H3K4M vs. H3K4M/K14ac: In the H3K4M/K14ac structure, K9* of H3K4M/K14ac (brown) forms a compensatory salt bridge with $E559_{LSD1}$ due to K14 acetylation. (**b**) WT LC H3K4M vs. Y391K LC H3K4M: CoREST1, bound to Y391K LSD1 (black), adopts a distinct conformation compared to CoREST1 bound to WT LSD1 (blue), shifting downward towards H3K4M. This shift is caused by the charge repulsion from $K391_{LSD1}$. (**c**) Y391K LC H3K4M vs. Y391K H3K4M/K14ac: In both structures, K9 of H3K4M (orange) and K9** of H3K4M/K14ac (light purple) are situated nearby $H564_{LSD1}$ $Q358_{LSD1}$, without forming a compensatory salt bridge with $E559_{LSD1}$. CoREST1 conformation remains downward, as described in (b). (**d**) Y391K LC H3K4M/K14ac vs. WT LC H3K4M/K14ac: In the WT LC structure, K9* of H3K4M/K14ac makes a compensatory salt bridge with $E559_{LSD1}$, whereas in the Y391K LC structure, K9** of H3K4M/K14ac remains unchanged, residing nearby $H564_{LSD1}$ and $Q358_{LSD1}$. Note: * and ** represent different lysine residues in the H3K4M and H3K4M/K14ac peptides, respectively.

https://doi.org/10.1038/s41589-024-01671-9

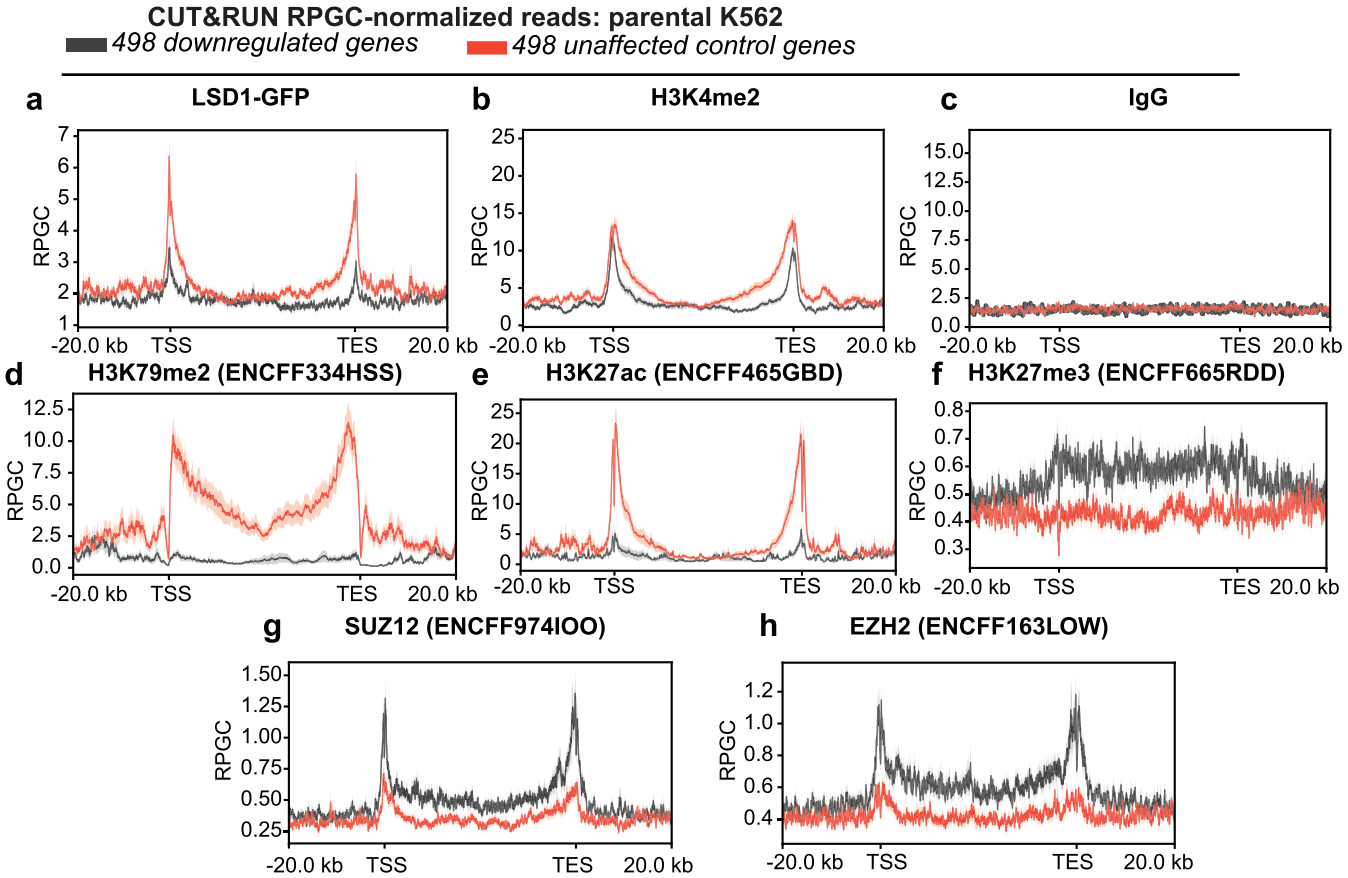



e

| Protein sample | Nucleosome | *V/[E], min$^{-1}$ (mean ± SEM)* | n |
|---|---|---|---|
| WT LHC, 90 nM | H3K9ac | 0.0073 ± 0.0005 | 4 |
| Y391K LHC, 90 nM | H3K9ac | 0.0069 ± 0.0004 | 4 |
| WT LHC, 120 nM | H3K9ac | 0.0063 ± 0.0004 | 4 |
| Y391K LHC, 120 nM | H3K9ac | 0.0079 ± 0.0005 | 4 |

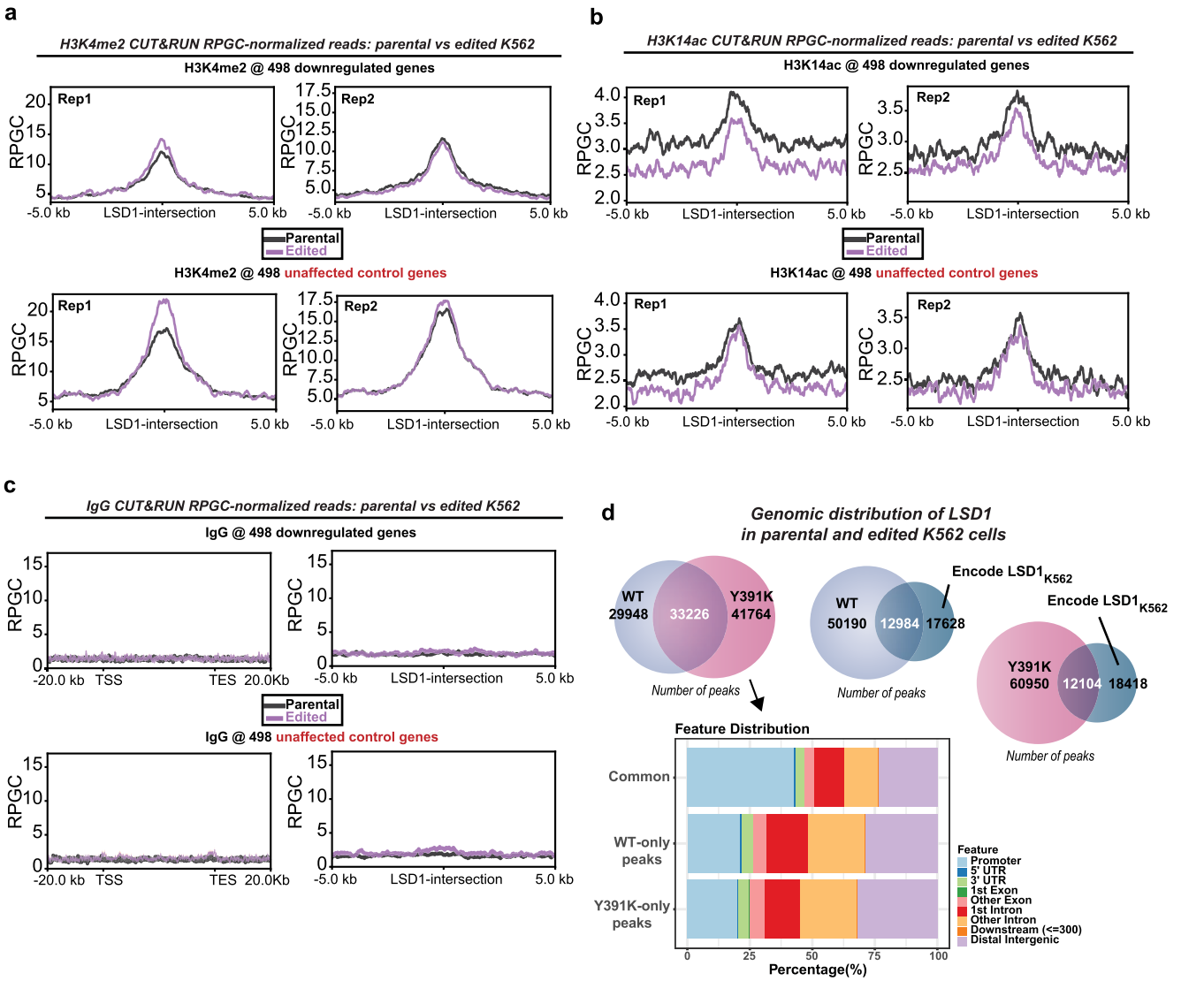
**Extended Data Fig. 6 | Analysis of Nucleosome Deacetylase Activity for Fluorescein-labeled WT LHC and Y391K LHC.** (**a**) Assessment of WT LHC deacetylase activity against H3K9ac nucleosomes. WT LHC concentrations of 90 nM (top) and 120 nM (bottom) were subjected to a 120-minute incubation with 100 nM of 185 bp H3K9ac nucleosomes, and variations in H3K9ac levels were monitored. (**b**) Evaluation of Y391K LHC deacetylase activity targeting H3K9ac-marked nucleosomes. Similar to (a), 90 nM (top) and 120 nM (bottom) Y391K LHC was incubated with 100 nM of 185 bp H3K9ac nucleosomes, and changes in H3K9ac levels were tracked over 120 minutes. Panels (**c**) and (**d**) illustrate the relative intensities obtained from (a) and (b), respectively, subjected to fitting into an exponential decay equation that includes constraints of $Y_0$ at 1 and plateau at 0. (**e**) V/[E] ($min^{-1}$) values from (a) and (b) were extrapolated (mean ± SEM). Anti-H3 blot at each time point from every other replicate was shown as a representative loading control.

https://doi.org/10.1038/s41589-024-01671-9



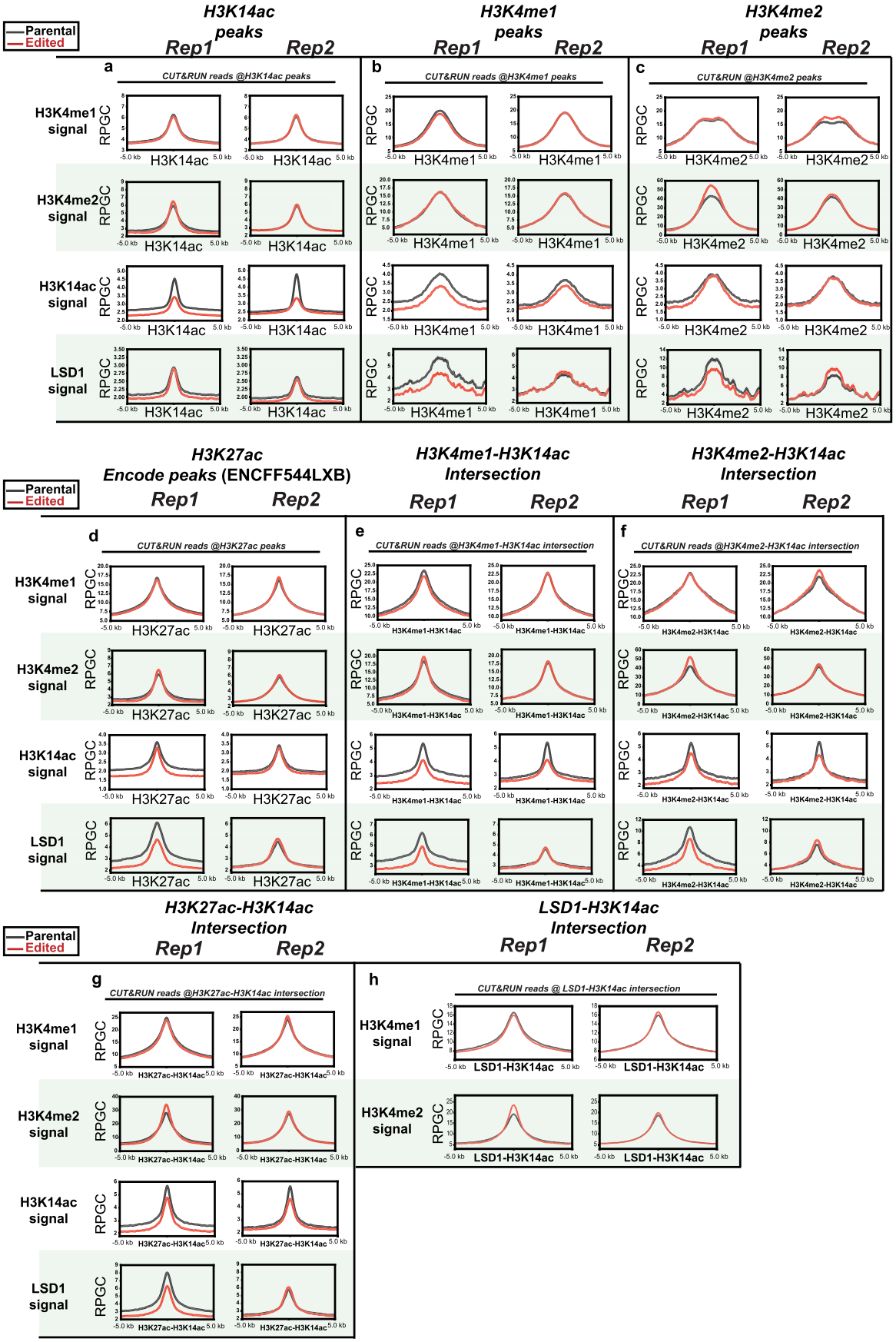

**Extended Data Fig. 7 | CUT&RUN Chromatin Profiling Analysis in *Parental K562* Cells: Comparison within Gene Bodies of Downregulated and Unaffected Control Genes.** (**a**) Metagene plot (mean ± SEM) of LSD1-GFP signal (experimental, n = 2) within gene bodies of downregulated genes (black) and unaffected control genes (red). (**b**) Metagene plot (mean ± SEM) from CUT&RUN analysis, showcasing H3K4me2 signal (experimental, n = 2), within gene bodies of 498 downregulated genes (black) and unaffected control genes (red). (**c**) Metagene plots (mean ± SEM) for IgG signal (experimental, n = 2), serving as a control, within gene bodies of 498 downregulated genes (black) and unaffected control genes (red). (**d**) Metagene plot (mean ± SEM) for H3K79me2 signal from Encode (K562 cell, ENCFF334HSS) within gene bodies of 498 downregulated genes (black) and unaffected control genes (red). (**e**) Metagene plot (mean ± SEM) for H3K27ac signal from Encode (K562 cells, ENCFF465GBD), encompassing gene bodies of the indicated gene sets. (**f**) Metagene plots (mean ± SEM) for H3K27me3 signal from Encode (K562 cells, ENCFF665RDD), observed within gene bodies of 498 downregulated genes (black) and unaffected control genes (red). (**g**) Metagene plots (mean ± SEM) for SUZ12 signal from Encode (K562 cells, ENCFF974IOO), observed within gene bodies of 498 downregulated genes (black) and unaffected control genes (red). (**h**) Metagene plots (mean ± SEM) for EZH2 signal from Encode (K562 cells, ENCFF974IOO), observed within gene bodies of 498 downregulated genes (black) and unaffected control genes (red). This provides context for the specificity of the observed signals. All analyses include assessment within the gene bodies as well as 20 kb regions upstream and downstream of the downregulated and unaffected control genes. All Encode data originate from K562 cells.

https://doi.org/10.1038/s41589-024-01671-9




**Extended Data Fig. 8 | Comparative CUT&RUN Analysis between *parental* and *edited K562* cells.** (**a**) Metagene plot (mean) for H3K4me2 signal at the LSD1 peaks (± 5 kb) across the 498 downregulated genes (top) and 498 unaffected control genes (bottom). The *parental K562* cells are represented in black, while the *edited K562* cells are depicted in pink. Two replicates (left and right) display variations in the 498 unaffected control regions. (**b**) Metagene plot (mean) for H3K14ac signal at the LSD1 peaks (± 5 kb) across the 498 downregulated genes (top) and 498 unaffected control genes (bottom). Similar to (a), the *parental K562* cells are shown in black, and the *edited K562* cells are shown in pink. Subtle variations in the 498 unaffected control regions are seen across two biological replicates, alongside reduced read counts in the downregulated genes. (**c**) Metagene plot (mean ± SEM) of IgG control signal within the scaled gene bodies ± 20 kb of the 498 downregulated genes (top left) and 498 unaffected control genes (bottom left), for both *parental* (black) and *edited K562* cells (pink). Metagene plot (mean) of IgG control signal at the LSD1-bound regions within the 498 downregulated genes (top right) and 498 unaffected control genes (bottom right), for both for both *parental* (black) and *edited K562* cells (pink) are shown. (**d**) Genomic distribution of LSD1 in *parental* and *edited K562* cells. The left Venn diagram illustrates that approximately 53% of LSD1 peaks from *parental K562* cells directly overlap with LSD1 peaks from *edited K562* cells. About 43% of these overlapping peaks were located in the promoter region of all LSD1-bound genes. In contrast, only around 20% of the non-overlapping peaks were found in the promoter regions of LSD1-bound genes, suggesting a redistribution of LSD1 in numerous non-promoter regions. SEACR relaxed mode was employed to identify these LSD1 peaks, which were subsequently compared with LSD1 peaks from the Encode database (K562 cells, ENCFF054XCG). Both *parental* and *edited K562* cells' LSD1 peaks exhibit approximately 40% direct overlap with the LSD1 peaks from the Encode database.

https://doi.org/10.1038/s41589-024-01671-9

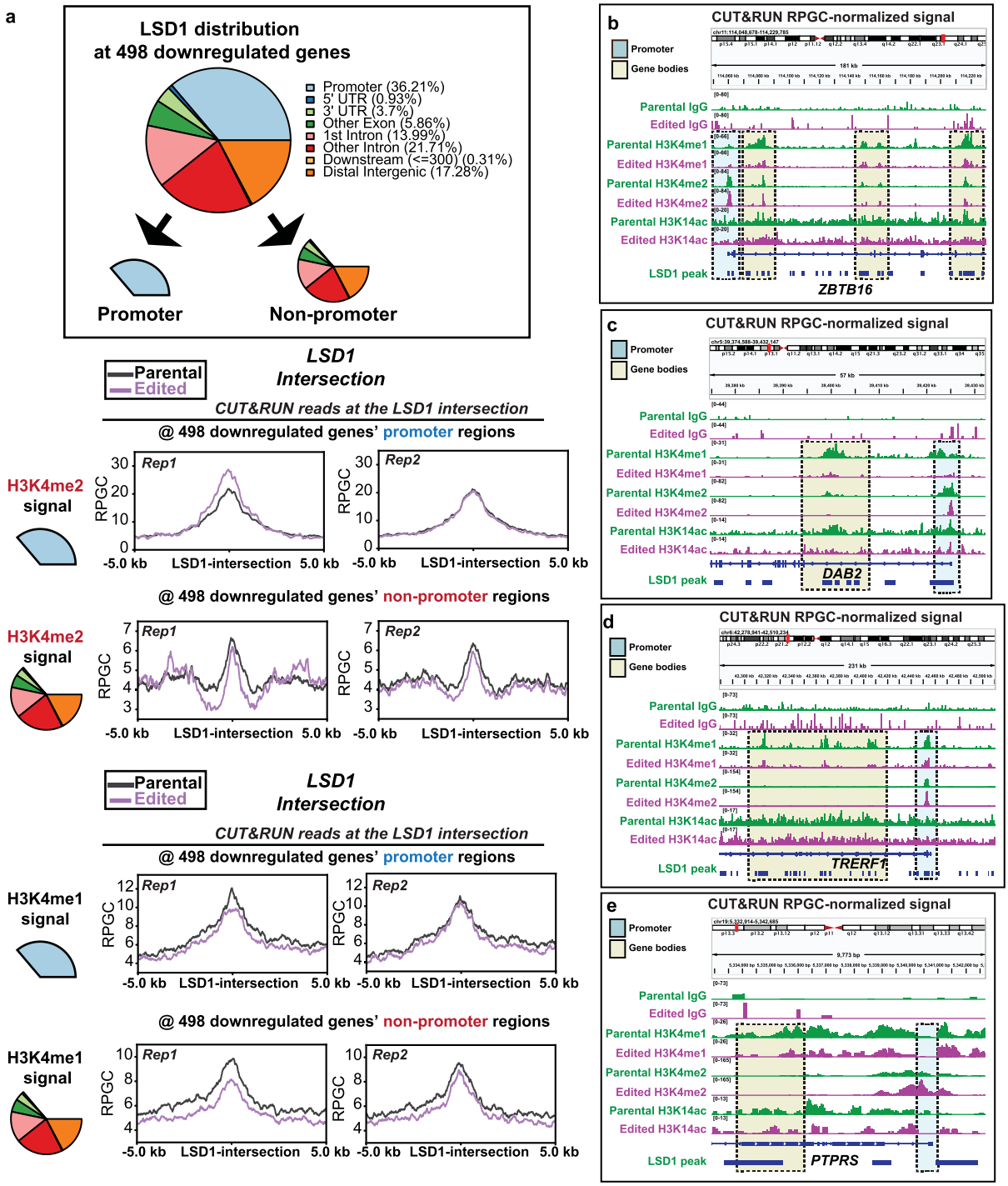



**Extended Data Fig. 9 | See next page for caption.**

https://doi.org/10.1038/s41589-024-01671-9

**Extended Data Fig. 9 | Metagene plot (mean ± SEM) for H3K4me1, H3K4me2, H3K14ac, and LSD1 from parental (black) and edited (red) K562 cells at various genomic locations.** Including (**a**) H3K14ac global peaks, (**b**) H3K4me1 global peaks, (**c**) H3K4me2 global peaks, (**d**) H3K27ac global peaks (adopted from Encode ENCFF544LXB), (**e**) Intersected peaks of H3K4me1 and H3K14ac, (**f**) Intersected peaks of H3K4me2 and H3K14ac, (**g**) Intersected peaks of H3K27ac and H3K14ac, and (**h**) Intersected peaks of LSD1 and H3K14ac (only showing H3K4me1 and H3K4me2).

https://doi.org/10.1038/s41589-024-01671-9




**Extended Data Fig. 10 | Differential Regulation in H3K4me1 and H3K4me2 Levels in Non-Promoter Regions of the 498 Downregulated Genes.** (**a**) LSD1 peaks in both promoter and non-promoter regions (top) were served as reference points for metagene plot analysis (bottom) (mean ± SEM). H3K4me1 and H3K4me2 levels at promoter and non-promoter regions were evaluated (parental – black and edited - pink). The reduction of H3K4me1 and H3K4me2 at LSD1-bound non-promoter regions is more pronounced in the 498 downregulated genes. (**b**–**e**) Genomic snapshots illustrate CUT&RUN signals for H3K4me1, H3K4me2, and H3K14ac for four representative genes (*ZBTB16*, *DAB2*, *TRERF1*, and *PTPRS*). Notably, H3K4me1 and H3K4me2 signals within gene bodies (brown box) show a reduction compared to the promoter region (blue box).

nature portfolio

Corresponding author(s): Samuel Beck, Andrea Mattevi, Philip A. Cole

Last updated by author(s): May 28, 2024

# 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<br>*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<br>*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 | Western blot images were obtained using Genesys (v1.8.5.0). K562 drugsenstivity measurement using alamarBlue was acquired using Gen5 (v3.03). MST binding affinity data were acquired using MO.Control (Nanotemper). |
| Data analysis | For Western blot band intensity analysis, ImageJ (v1.53k, NIH) was used. For demethylase/deacetylase activity, K562 growth rate measurement, and drug sensitivity analysis, GraphPad Prism 9 was used. For MST binding affinity analysis, MO.Affinity Analysis (Nanotemper) and GraphPad Prism 9 were used. For crystallography data analysis, XDS (BUILT=20180126), CCP4 (v 7.0.044), Phenix (v 1.20.1-4487), and Coot (v 0.9.6) were used. For RNA-Seq (performed by MedGenome), FastQC (v0.11.8), fastq-mcf (v1.05), cutadapt (v2.5), Bowtie2 (v2.5.1), STAR (v2.7.3a), HTSeq (v0.11.2), cufflinks software (v2.2.1), RNA-SeQC (v1.1.8), RSeQC (v3.0.1), MultiQC (v1.7), DESeq2 (R Bioconductor package), and ClusterProfiler (v4.9.2) were used. For CUT&RUN, fastp (v0.23.4), FastQC (v0.11.9), Bowtie2 (v2.4.4), Samtools (v1.17), Bedtools (v2.31.0), SEACR (v1.3), Deeptools package (v3.5.2), BioMart R (v2.57.1), and ChIPSeeker were used. The codes used for CUT&RUN data processing were described in the supplemental information. For comparative analysis with the existing ChIP-Seq data, the data were downloaded from the ENCODE database. For protein structure analysis, Pymol (v. 1.2r3pre, Schrödinger, LLC) and ChimeraX (v1.4) were used. LHC model structure based on AlphaFold2 and Rosetta3 (v 3.13) was downloaded from the study (https://doi.org/10.7910/DVN/ULB7JL). |

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

Structure factors and atomic coordinates have been deposited in the Protein Data Bank with PDB IDs 8Q1G, 8Q1H, and 8Q1J. RNA-Seq data of the parental and edited K562 cells have been deposited with the GEO accession code GSE243427 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE243427). CUT&RUN data of the parental and edited K562 cells have been deposited with the GEO accession code GSE243231 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE243231).

## Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

| | |
|---|---|
| Reporting on sex and gender | not applicable |
| Population characteristics | not applicable |
| Recruitment | not applicable |
| Ethics oversight | not applicable |

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

# 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 | We did not use any statistical method to pre-determine sample sizes, as they are similar to the sample sizes employed in previous / recent publications. The reported data in the manuscript represent the mean values from the replicates as stated in the figure legends and in the Methods section. The sample sizes are chosen based on the degree of reproducibility and depth to derive statistical significance. We observe the effects of interest among the conditions and reproducible across the replicates. |
| Data exclusions | No data were excluded, except for two data points from the nonlinear regression analysis for MST binding affinity measurement (H3K14ac nucleosome toward WT LC). These two data points were outliers. The values are included in the Source Data: ExtendedDataFig3h_Statistical_source_data_20240506.xlsx and are indicated with highlight. No additional data were excluded. |
| Replication | All experiments were replicated at least two different times. |
| Randomization | Samples were not randomized. However, for biochemical assays, samples were handled without particular preference or bias in choosing protein aliquots, given that the enzyme samples were highly pure (stored frozen at -80 degrees C) and displayed reproducibility. For cellular experiments, batch effects and experimental biases were minimized as all replicates were subjected to sample processing and data acquisition simultaneously. |
| Blinding | No blinding was applied. Experiments were conducted and analyzed by the same individuals, through the same sample acquisition and analysis pipeline. No blinding was applied 1) as investigators did not have a bias or knowledge on the experimental results before conducting them except for WT LC's demethylase activity toward H3K4me2 nucleosome (WT LC's slow demethylase activity toward the H3K4me2-K14ac peptide substrates were known previously, but not on the nucleosome substrates), 2) due to limited resources (mainly nucleosome substrates), and 3) to minimize suboptimal outcome given limited resources. |

# 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 |
| ☒ | ☐ Clinical data |
| ☒ | ☐ Dual use research of concern |

## Methods

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

# Antibodies

Antibodies used

The primary antibodies used for CUT&RUN and biochemical assays were: rabbit IgG for CUT&RUN (1:100 dilution; Millipore Sigma, I5006), anti-GFP (1:50 dilution; Abcam, cat Ab290), anti-H3K4me1 for CUT&RUN (1:50 dilution; Epicypher, cat 13-0040), anti-H3K4me1 for demethylase assays (1:1000 dilution; Thermo Fisher Scientific, cat 710795), anti-H3K4me2 (1:2000 dilution for demethylase assay and 1:50 dilution for CUT&RUN; Abcam, cat ab32356), anti-H3K14ac (1:2000 dilution for the nucleosome quality check assay and 1:50 dilution for CUT&RUN; Millipore Sigma, cat 07-353), anti-H3K9ac (1:2000 dilution for deacetylase assay; Abcam, cat ab32129), and anti-H3 (1:2000 dilution, Abcam, cat ab1791). The primary antibodies used for cell lysate Western blotting were: anti-GFP (1:250 dilution; Santa Cruz Biotechnology, cat sc-9996), anti-LSD1 (1:1000 dilution; Bethyl Laboratories, cat A300-215A), and anti-GAPDH (1:2000 dilution; Cell Signaling Technology, cat 2118S). The secondary antibodies used were: HRP-conjugated anti-mouse secondary antibody (1:1000 dilution; Cell Signaling Technology cat 7076S) and anti-rabbit secondary antibody (1:1000 dilution; Cell Signaling cat 7074S).

Validation

All antibodies used are commercially available and were validated for the specified applications by the suppliers:

Rabbit IgG (Millipore Sigma, I5006) - validated applications: ELISA, dot immunobinding, Western immunoblotting, immunodiffusion, and immunoelectrophoresis.
Anti-GFP (Abcam, cat Ab290) - species reactivity: species independent; validated applications: ELISA, Western blotting (dil. 1:1000 to 1:2500), IHC-Fr (dil 1:3000), ICC (dil 1:200 - 1:1000), IHC-P (1:500), IP, IHC-FoFr (dil 1:200 - 1:500), IHC-FrFl, and Electron Microscopy (dil 1:1000 - 1:4000).
Anti-H3K4me1 (Epicypher, cat 13-0040) - species reactivity: human, mouse, wide range (predicted); validated applications: ChIP (2 - 5 µg per 5 µg chromatin) , ChIP-Seq, Western blotting, Immunocytochemistry (1 µL/mL), ELISA, Luminex.
Anti-H3K4me1 (Thermo Fisher Scientific, cat 710795) - species reactivity: human, mouse; validated applications: Western blot (0.5 - 1 µg/mL), Immunocytochemistry (1 µg/mL), ChIP assay (1-5 µg), ChIP-Seq (3 µg), Peptide Array (0.25 µg/mL), and CUT&RUN (dil 1:100).
Anti-H3K4me2 (Abcam, cat ab32356) - species reactivity: Mouse, Rat, Chicken, Cow, Human, African green monkey; validated applications: ChIP (use 2-25 µg of chromatin), ChIP-Seq (use 4-30 µg chromatin), Western blotting (dil 1:2000), IHC-P (dil 1:800), ICC/IF (dil 1:1000), Flow Cyt (dil 1:20 - 1:100), IP (1:30).
Anti-H3K14ac (Millipore Sigma, cat 07-353) - species reactivity: human, yeast; validated applications: Western blotting (dil 1:1000 - 1:5000), ChIP-Seq, Dot blot, Multiplexing, ChIP.
Anti-H3K9ac (Abcam, cat ab32129) - species reactivity: mouse, rat, human; validated applications: Flow Cyt, ChIP-Seq (use 4 µg - 30 µg), CUT&RUN (use at an assay dependent concentration), ChIP (use 2-25 µg), Western blotting (dil 1:500), IHC-P (dil 1:50-1:200), IP (dil 1:30), ICC/IF (dil 1:250).
Anti-H3 (Abcam, cat ab1791) - species reactivity: Mouse, Rat, Human, Saccharomyces cerevisiae, Xenopus laevis, Arabidopsis thaliana, Drosophila melanogaster, Indian muntjac, Schizosaccharomyces pombe; validated applications: IHC-P (dil 1:100 - 1:400), ChIP (use 2 µg for 1 million cells), IP (5 µg/mL), Western blotting (dil 1:1000-1:5000), ICC/IF (1 µg/mL).
Anti-GFP (Santa Cruz Biotechnology, cat sc-9996) - validated applications: Western blotting (dil 1:100 - 1:1000), immunoprecipitation (1-2 µg per 100-500 µg total protein), immunofluorescence (dil 1:50 - 1:500), flow cytometry (1 µg per 1 million cells), ELISA (dil 1:30-1:3000).
Anti-LSD1 (Bethyl Laboratories, cat A300-215A) - species reactivity: human; validated applications: ChIP, IHC (dil 1:500-1:2000), IP (6 µg/mg lysate), Western blotting (dil 1:5000-1:15,000).
Anti-GAPDH (Cell Signaling Technology, cat 2118S) - species reactivity: human, mouse, rabbit, monkey, bovine, pig; validated applications: Western blotting (dil 1:1000), immunohistochemistry (dil 1:400-1:1600), immunofluorescence (dil 1:50-1:200), flow cytometry (dil 1:100-1:400).
HRP-conjugated anti-mouse secondary antibody (Cell Signaling Technology cat 7076S) - validated applications: Western blotting (dil 1:1000-1:3000).
HRP-conjugated anti-rabbit secondary antibody (Cell Signaling Technology cat 7074S) - validated applications: Western blotting (dil 1:1000-1:3000).

# Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)

Freestyle HEK293F (Thermo Fisher Scientific) and K562 (ATCC).

Authentication

Freestyle HEK293F cells were directly purchased from the manufacturer (Thermo Fisher Scientific) and no further authentication was performed. K562 cells from ATCC were authenticated by Short Tandem Repeat profiling (Genetica).

| | |
|---|---|
| Mycoplasma contamination | All cells were tested for mycoplasma (Lonza) every month while being maintained. Representative well for each cell line was selected and displayed luminescence ratio ranging from 0.4 to 0.8 ( >1.2 is considered mycoplasma contamination). |
| Commonly misidentified lines (See ICLAC register) | No commonly misidentified lines were used in this study. |

## 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<br>*May remain private before publication.* | RNA-Seq data of the parental and edited K562 cells have been deposited with the GEO accession code GSE243427 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE243427). CUT&RUN data of the parental and edited K562 cells have been deposited with the GEO accession code GSE243231 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE243231). |
| Files in database submission | 2_wtLSD1GFP_antiGFP-rep1.bw, 2_wtLSD1GFP_antiGFP_rep1_SEACRrelaxed.bed, 2_wtLSD1GFP_antiGFP-rep1_R1.fastq.gz, 2_wtLSD1GFP_antiGFP-rep1_R2.fastq.gz,<br>7_wtLSD1GFP_antiGFP-rep2.bw, 7_wtLSD1GFP_antiGFP_rep2_SEACRrelaxed.bed, 7_wtLSD1GFP_antiGFP-rep2_R1.fastq.gz, 7_wtLSD1GFP_antiGFP-rep2_R2.fastq.gz,<br>12_Y391K-LSD1GFP_antiGFP-rep1.bw, 12_Y391K-LSD1GFP_antiGFP_rep1_SEACRrelaxed.bed, 12_Y391K-LSD1GFP_antiGFP-rep1_R1.fastq.gz, 12_Y391K-LSD1GFP_antiGFP-rep1_R2.fastq.gz,<br>17_Y391K-LSD1GFP_antiGFP-rep2.bw, 17_Y391K-LSD1GFP_antiGFP_rep2_SEACRrelaxed.bed, 17_Y391K-LSD1GFP_antiGFP-rep2_R1.fastq.gz, 17_Y391K-LSD1GFP_antiGFP-rep2_R2.fastq.gz,<br>3_wtLSD1GFP_antiH3K4me1-rep1.bw, 3_wtLSD1GFP_antiH3K4me1_rep1_SEACRrelaxed.bed, 3_wtLSD1GFP_antiH3K4me1-rep1_R1.fastq.gz, 3_wtLSD1GFP_antiH3K4me1-rep1_R2.fastq.gz,<br>8_wtLSD1GFP_antiH3K4me1-rep2.bw, 8_wtLSD1GFP_antiH3K4me1_rep2_SEACRrelaxed.bed, 8_wtLSD1GFP_antiH3K4me1-rep2_R1.fastq.gz, 8_wtLSD1GFP_antiH3K4me1-rep2_R2.fastq.gz,<br>13_Y391K-LSD1GFP_antiH3K4me1-rep1.bw, 13_Y391K-LSD1GFP_antiH3K4me1_rep1_SEACRrelaxed.bed, 13_Y391K-LSD1GFP_antiH3K4me1-rep1_R1.fastq.gz, 13_Y391K-LSD1GFP_antiH3K4me1-rep1_R2.fastq.gz,<br>18_Y391K-LSD1GFP_antiH3K4me1-rep2.bw, 18_Y391K-LSD1GFP_antiH3K4me1_rep2_SEACRrelaxed.bed, 18_Y391K-LSD1GFP_antiH3K4me1-rep2_R1.fastq.gz, 18_Y391K-LSD1GFP_antiH3K4me1-rep2_R2.fastq.gz,<br>4_wtLSD1GFP_antiH3K4me2-rep1.bw, 4_wtLSD1GFP_antiH3K4me2_rep1_SEACRrelaxed.bed,<br>4_wtLSD1GFP_antiH3K4me2_R1.fastq.gz, 4_wtLSD1GFP_antiH3K4me2_R2.fastq.gz,<br>9_wtLSD1GFP_antiH3K4me2-rep2.bw, 9_wtLSD1GFP_antiH3K4me2_rep2_SEACRrelaxed.bed, 9_wtLSD1GFP_antiH3K4me2-rep2_R1.fastq.gz, 9_wtLSD1GFP_antiH3K4me2-rep2_R2.fastq.gz,<br>14_Y391K-LSD1GFP_antiH3K4me2-rep1.bw, 14_Y391K-LSD1GFP_antiH3K4me2_rep1_SEACRrelaxed.bed, 14_Y391K-LSD1GFP_antiH3K4me2-rep1_R1.fastq.gz, 14_Y391K-LSD1GFP_antiH3K4me2-rep1_R2.fastq.gz,<br>19_Y391K-LSD1GFP_antiH3K4me2-rep2.bw, 19_Y391K-LSD1GFP_antiH3K4me2_rep2_SEACRrelaxed.bed, 19_Y391K-LSD1GFP_antiH3K4me2-rep2_R1.fastq.gz, 19_Y391K-LSD1GFP_antiH3K4me2-rep2_R2.fastq.gz,<br>5_wtLSD1GFP_antiH3K14ac-rep1.bw, 5_wtLSD1GFP_antiH3K14ac_rep1_SEACRrelaxed.bed,<br>5_wtLSD1GFP_antiH3K14ac_R1.fastq.gz, 5_wtLSD1GFP_antiH3K14ac_R2.fastq.gz,<br>10_wtLSD1GFP_antiH3K14ac-rep2.bw, 10_wtLSD1GFP_antiH3K14ac_rep2_SEACRrelaxed.bed, 10_wtLSD1GFP_antiH3K14ac-rep2_R1.fastq.gz, 10_wtLSD1GFP_antiH3K14ac-rep2_R2.fastq.gz,<br>15_Y391K-LSD1GFP_antiH3K14ac-rep1.bw, 15_Y391K-LSD1GFP_antiH3K14ac_rep1_SEACRrelaxed.bed, 15_Y391K-LSD1GFP_antiH3K14ac-rep1_R1.fastq.gz, 15_Y391K-LSD1GFP_antiH3K14ac-rep1_R2.fastq.gz,<br>20_Y391K-LSD1GFP_antiH3K14ac-rep2.bw, 20_Y391K-LSD1GFP_antiH3K14ac_rep2_SEACRrelaxed.bed, 20_Y391K-LSD1GFP_antiH3K14ac-rep2_R1.fastq.gz, 20_Y391K-LSD1GFP_antiH3K14ac-rep2_R2.fastq.gz,<br>1_wtLSD1GFP_antiIgG-rep1.bw, 1_wtLSD1GFP_antiIgG-rep1_R1.fastq.gz, 1_wtLSD1GFP_antiIgG-rep1_R2.fastq.gz,<br>6_wtLSD1GFP_antiIgG-rep2.bw, 6_wtLSD1GFP_antiIgG-rep2_R1.fastq.gz, 6_wtLSD1GFP_antiIgG-rep2_R2.fastq.gz,<br>11_Y391K-LSD1GFP_anti-IgG-rep1.bw, 11_Y391K-LSD1GFP_antiIgG-rep1_R1.fastq.gz, 11_Y391K-LSD1GFP_antiIgG-rep1_R2.fastq.gz,<br>16_Y391K-LSD1GFP_antiIgG-rep2.bw, 16_Y391K-LSD1GFP_antiIgG-rep2_R1.fastq.gz, 16_Y391K-LSD1GFP_antiIgG-rep2_R2.fastq.gz, |
| Genome browser session (e.g. UCSC) | Not applicable |

### Methodology

| | |
|---|---|
| Replicates | Two replicates were performed for CUT&RUN. Triplicates (for parental K562 cells) and Quadruplicates (for edited K562 cells) were performed for RNA-Seq |
| Sequencing depth | Name, Total number of reads, Uniquely mapped reads, Length of reads, Single or paired-end<br>1_IgG_parental_rep1(GSM7781194) 5365961 3316917 50 nt paired-end<br>2_GFP(LSD1)_parental_rep1(GSM7781178) 16265568 14958970 50 nt paired-end<br>3_H3K4me1_parental_rep1(GSM7781182) 18520150 17726956 50 nt paired-end<br>4_H3K4me2_parental_rep1(GSM7781186) 17147353 16468884 50 nt paired-end<br>5_H3K14ac_parental_rep1(GSM7781190) 20915318 20253891 50 nt paired-end<br>6_IgG_parental_rep2(GSM7781195) 4052471 2773884 50 nt paired-end<br>7_GFP(LSD1)_parental_rep2(GSM7781179) 12354401 11485222 50 nt paired-end |

8_H3K4me1_parental_rep2(GSM7781183) 18505980 17622067 50 nt paired-end
9_H3K4me2_parental_rep2(GSM7781187) 28213330 27173682 50 nt paired-end
10_H3K14ac_parental_rep2(GSM7781191) 17731309 17188420 50 nt paired-end
11_IgG_edited_rep1(GSM7781196) 2276737 1097465 50 nt paired-end
12_GFP(LSD1)_edited_rep1(GSM7781180) 15868761 14847166 50 nt paired-end
13_H3K4me1_edited_rep1(GSM7781184) 14123544 13426979 50 nt paired-end
14_H3K4me2_edited_rep1(GSM7781188) 23612345 23363975 50 nt paired-end
15_H3K14ac_edited_rep1(GSM7781192) 17475010 17288788 50 nt paired-end
16_IgG_edited_rep2(GSM7781197) 3839595 2267803 50 nt paired-end
17_GFP(LSD1)_edited_rep2(GSM7781181) 13936562 12073454 50 nt paired-end
18_H3K4me1_edited_rep2(GSM7781185) 15354204 14775939 50 nt paired-end
19_H3K4me2_edited_rep2(GSM7781189) 27606728 26588257 50 nt paired-end
20_H3K14ac_edited_rep2(GSM7781193 ) 17914367 17444004 50 nt paired-end

Antibodies

rabbit IgG (Millipore Sigma, I5006), anti-GFP (Abcam, cat Ab290), anti-H3K4me1 (Epicypher, cat 13-0040), anti-H3K4me2 (Abcam, cat ab32356), and anti-H3K14ac (Millipore Sigma, cat 07-353).

Peak calling parameters

SEACR relaxed & normalized IgG modes were used. For example, SEACR_1.3.sh 2_wtLSD1GFP_antiGFP-rep1.bedgraph 1_wtLSD1GFP_antiIgG-rep1.bedgraph norm relaxed 2_wtLSD1GFP_antiGFP-rep1.relaxed.bed

Data quality

All sequencing data have been submitted to FastQC analysis. For SEACR-assisted peak calling, each bedgraph normalized by the E. coli spike-in was further normalized by the IgG control and processed using the relaxed mode.

Software

CUT&RUN data processing was mostly adopted from the established protocol (Zheng Y et al (2020). Protocol.io, CUT&Tag Data Processing and Analysis Tutorial). FastQ data files were initially processed by fastp (v. 0.23.4), and quality checks were assessed by FastQC (v. 0.11.9). The paired-end reads were aligned to the human reference genome (GRCh38 assembly) using Bowtie2 (v. 2.4.4) as well as the E. coli genome (K-12 MG1655). For GRch38 assembly, the following arguments were used: bowtie2 -p 8 --end-to-end --very-sensitive --no-mixed --no-discordant --phred33 -I 10 -X 700 -x /path-to-genome/ /GRCh38_noalt_as. For E. coli, the following arguments were used: bowtie2 -p 8 --end-to-end --very-sensitive --no-overlap --no-dovetail --no-mixed --no-discordant --phred33 -I 10 -X 700 -x /path/Escherichia_coli_K_12_MG1655/NCBI/2001-10-15/Sequence/Bowtie2Index/genome. The sam output files were converted to the bam format using Samtools (v. 1.17). For peakcalling, bam files were first converted to bed and bedgraph files using Bedtools (v. 2.31.0) and SEACR relaxed mode85 was used to call peaks by normalizing against the IgG control bedgraph files for each replicate. For CUT&RUN signal visualization, bam files were sorted by the coordinates and indexed using Samtools (v. 1.17), and the sorted bam files were converted to the bigwig files using bamCoverage application from the Deeptools package (v. 3.5.2). RPGC normalization mode using the effective Genome size of 3049315783 bp and bin size of 1 was applied (bamCoverage -b "$file" -p 10 --binSize 1 --normalizeUsing RPGC --effectiveGenomeSize 3049315783 --extendReads --outFileFormat bigwig). The bigwig files were visualized by using Integrative Genomics Viewer (IGV) and analyzed using computeMatrix and plotProfile applications from the Deeptools package (v. 3.5.2).