Illuminating the nonenzymatic functions of LSD1 complexes in leukemia using drug resistance alleles

Citation

Waterbury, Amanda Lynn. 2022. Illuminating the nonenzymatic functions of LSD1 complexes in leukemia using drug resistance alleles. Doctoral dissertation, Harvard University Graduate School of Arts and Sciences.

Permanent link

https://nrs.harvard.edu/URN-3:HUL.INSTREPOS:37374561

Terms of Use

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story .

Accessibility

HARVARD UNIVERSITY
Graduate School of Arts and Sciences
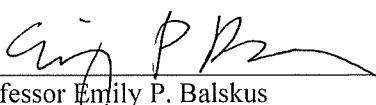
DISSERTATION ACCEPTANCE CERTIFICATE

The undersigned, appointed by the

Department of Chemistry & Chemical Biology

have examined a dissertation entitled:

Illuminating the nonenzymatic functions of LSD1 complexes in leukemia using drug resistance alleles

presented by: Amanda Lynn Waterbury

candidate for the degree of Doctor of Philosophy and hereby certify that it is worthy of acceptance.

Signature _____
Typed name: Professor Brian B. Liau

Signature _____
Typed name: Professor Emily P. Balskus

Signature _____
Typed name: Professor Daniel E. Kahne
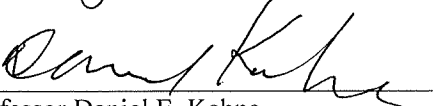
Date: 13 Dec 2022

Illuminating the nonenzymatic functions of LSD1 complexes in leukemia using

drug resistance alleles

A dissertation presented by

Amanda Lynn Waterbury

to

The Department of Chemistry and Chemical Biology

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Chemistry

Harvard University

Cambridge, Massachusetts

December 2022

Illuminating the nonenzymatic functions of LSD1 complexes in leukemia using drug resistance alleles

Abstract

The identification of drug resistance conferring mutations not only confirms on-target small molecule mechanism but can also serve as a useful discovery tool to uncover novel aspects of target biology. However, the identification of drug resistance conferring mutations is difficult, especially in allosteric or scaffolding sites and with protein targets lacking structural data. Using a CRISPR-Cas9 tiling mutagenesis approach termed CRISPR-suppressor scanning to profile resistance mutations, we elucidated the downstream mechanism of lysine-specific histone demethylase 1 (LSD1) inhibitors in acute myeloid leukemia (AML). We demonstrate that this approach can rapidly identify mutations that confer drug resistance. Moreover, our studies illustrate how CRISPR-suppressor scanning can identify functional hotspots beyond the small molecule binding site that clarify the mechanism of small molecule inhibitors and allosteric sites essential for target protein function.

In this thesis, I present three stories on the histone lysine demethylase LSD1 that illustrate the application of CRISPR-suppressor scanning as a tool for biological discovery. The first chapter introduces key concepts and examples, from chromatin biology to the identification of drug resistance conferring mutations. Recent advances in structurally resolving LSD1 and its complex members are reviewed as well as efforts to pharmacologically target LSD1 for oncology and neurological disorders.

The following chapter discusses our work investigating resistance mutations generated by CRISPR-suppressor scanning to small molecule inhibitors of LSD1 in the context of AML. LSD1

was discovered as a vulnerability in cancer, where knockdown of LSD1 was shown to suppress leukemia cell growth, prompting interest in developing pharmacological inhibitors of LSD1. However, how LSD1 inhibition leads to leukemia cell growth arrest was unclear at the offset of our studies. By profiling resistance mutations generated by CRISPR-suppressor scanning to LSD1 small molecule inhibitors in AML cell lines, we demonstrate that the demethylase activity of LSD1 is not essential for cancer cell proliferation and that the antiproliferative activity of LSD1 inhibitors stems from disrupting the protein-protein interaction between LSD1 and the hematopoietic transcription factor, GFI1B. Overall, this study revised the mechanism of action of LSD1 small molecule inhibitors and clarified the scaffolding functions of LSD1 in leukemia.

In chapter three, I describe our efforts towards elucidating the role of the disordered LSD1 N-terminus in modulating LSD1 function in AML. In the previous study, we identified mutations in the N-terminus of LSD1 that were enriched in the presence of LSD1 inhibitors. How mutations in this intrinsically disordered region (IDR) of LSD1 confer drug resistance was unclear, especially given that the LSD1 IDR is distal from the drug binding site and is not structurally resolved. Through mechanistic investigations of these drug resistant mutants, we identify LSD1's role in the downstream mechanism of LSD1 inhibition, beyond LSD1-GFI1B complex dissociation. We found that LSD1 inhibitors are transcription factor reprogrammers, prompting the dissociation of the high-affinity LSD1-GFI1B complex to promote the formation of weaker protein-protein interactions with myeloid transcription factors to buffer enhancer activity. The LSD1 IDR mutants display increased interactions with key myeloid transcription factors, which we suggest prevents enhancer commissioning after LSD1 inhibitor treatment in these drug resistant cell lines. We further clarify the function of the LSD1 IDR in modulating transcription factor interactions and enhancer activity, revealing new aspects of LSD1 biology in the context of AML.

In chapter four, I describe how a distal loop deletion in LSD1 mediates drug resistance by promoting the fragmentation of a covalent drug adduct, illuminating new mechanisms of drug resistance beyond the canonical perturbation of the drug binding site. The covalent drug

fragmentation induced by this resistance mutation is mechanistically similar to demethylase specific LSD1 inhibitors, such as T-448, which are being tested for neuroscience applications. Through the study of this loop deletion mutant and T-448 analogs, we elucidate the mechanism and structural features that guide this covalent drug adduct rearrangement. Furthermore, this study highlights how mutations in allosteric or distal sites can promote resistance by unique mechanisms and will hopefully be informative for the development of therapeutics.

Collectively, these projects showcase the utility of CRISPR-suppressor scanning in identifying drug-resistance conferring mutants, the mechanistic study of which can illuminate fundamental aspects of protein function, downstream biology, as well as small molecule mechanism of action. These findings not only advance our understanding of LSD1 in the context of AML but also provide fundamental insights into the role of chromatin modifiers beyond their canonical enzymatic functions and illustrate how both scaffolding and disordered sites can modulate their activity.

Dedicated to my great grandfather, Lowell Waterbury.

Table of Contents

Acknowledgments

First and foremost, I would like to thank my advisor, Prof. Brian Liau for his mentorship and support over the course of my PhD. I am immensely grateful to have been a part of his lab and to have had the opportunity to watch this lab grow from its early days. It has been both a formative and special experience to help set up this lab and evolve the lab's research over these years. Moreover, Brian has been instrumental to my development as a researcher, lab member, and mentor. I admire the care and commitment Brian has put into both the research and group. His excitement for research, scientific creativity, and commitment to mentorship has been motivational throughout my PhD. Brian, thank you for all the guidance and support you have provided over the years and for taking me on as one of the first few members of the lab.

I would also like to thank my graduate committee mentors, Prof. Daniel Kahne, Prof. Emily Balskus and my former committee member Prof. Phil Cole, who have provided invaluable feedback and advice throughout my PhD.

I would like to thank the first members of the lab, Cindy Su, Kevin Ngan, Allison Siegenfeld, Eileen Hoang, and Mike Vinyard. They not only helped build the lab, but also provided so much support and comradery, especially in those early days. I would also like to thank Paloma Tuttle, Olivia Zhang, and Sam Hoenig, lab members I've had the opportunity to mentor and work with during my PhD- it really has been a privilege and one of the best aspects of my PhD to have mentored and worked with you, and I've learned so much in the process. Thank you to my 'Salt Bay' mates, Kevin Ngan, Ceejay Lee, and Paloma Tuttle, for your friendship, support, and coffee trips to BioLabs. I also want to thank Shelby Roseman, Olivia Zhang, Megan Yeo, Nicholas Lue, and Emma Garcia for all their friendship and support over these years. I've been incredibly lucky to have had such supportive colleagues and friends in the lab. I would also like to especially thank Dr. Pallavi Gosavi, who mentored me in the lab - thank you for taking me in as a second year and teaching me biochemistry as well as for continuously supporting me during my PhD. I want to

Hanjie Jiang, for their assistance with biochemical experiments. I would like to thank Prof. Andrea Mattevi and Jonatan Caroli, collaborators on the T-448/TTASdel project. I would also like to thank Prof. Anders Hansen and Domenic Narducci for their assistance with microscopy experiments and computational expertise.

I would also like to thank my advisors and mentors who were formative to my growth as a scientist prior to graduate school. I would like to thank Prof. Richard Jordan as well as my mentor in the lab, Dr. Erik Reinhart, for their patience, support, and mentorship. I would also like to thank Dr. Susan Boyle-Vavra and members of her lab for introducing me to microbiology research.

Thank you to all my friends who have provided me with support, laughter, and good times, and have helped me keep going during my PhD. I would like to especially thank my former roommates, Shannon Miller, Josh Cox, and Allison Siegenfeld, as well as my CCB cohort friends, Dennis Kutateladze, Hayden Sharma, Alex Peterson, and Dan Pollack, for their friendship, warmth, and laughter after long days in lab. I am incredibly thankful for meeting such amazing people, kind friends, and talented scientists here. I would also like to thank Kathy Nguyen and Lydia Tse for over a decade of friendship and support. Lastly, thank you to my cats, Sadie and Cinnamon, for all the snuggles and emotional support.

Finally, thank you to my family who have been there for me over these years. I would like to especially thank my parents, Scott and Natalie, as well as my brother, Sean. Their encouragement and love have meant so much to me. I am incredibly grateful for the sacrifices my parents made for me growing up. I'd like to thank my mom for her endless support and compassion and my dad for always encouraging me to attend college and work hard, and believing that I could succeed even when I doubted myself.

Chapter 1: Introduction

Contributions:

Amanda L. Waterbury wrote the text. Figures were made by Amanda L. Waterbury with assistance from Brian B. Liau.

1.1. Overview of epigenetics and allosteric regulation

Despite every cell in the body containing the same genetic code, the phenotypes of the various cell types differ widely to give rise to a diversity of cells from skin cells to neurons. How this diversity is accomplished is due to differences in gene expression that are highly regulated by epigenetic or chromatin modifying factors.[1–5] DNA is packaged into chromatin after forming a nucleoprotein complex with histone octamer and arrays of these nucleosomes can fold into higher-order structures to regulate genome accessibility (**Figure 1.1a**). The histone N-terminal tails are unstructured and were discovered to be highly modified by post-translational modifications (PTMs). These modifications as well as chemical modifications to DNA bases, are installed, removed, or read by chromatin modifying protein complexes to regulate chromatin accessibility and can be inherited across cell division to maintain gene expression programs (**Figure 1.1a**).[4,5] By dynamically packaging and unpackaging DNA elements, these chromatin modifiers, along with transcription factors, regulate gene activation and repression. The subsequent discovery and characterization of protein complexes that regulate chromatin-associated pathways has led to a better understanding of how chromatin modifiers establish and maintain genome packaging to regulate cell state, from their role in normal development to maintenance of tissue specific patterns of gene expression (**Figure 1.1b**).[1,6] During differentiation cells undergo drastic changes in their epigenetic profile and disruption of the factors responsible for this reprograming can halt differentiation and lead to pluripotency.
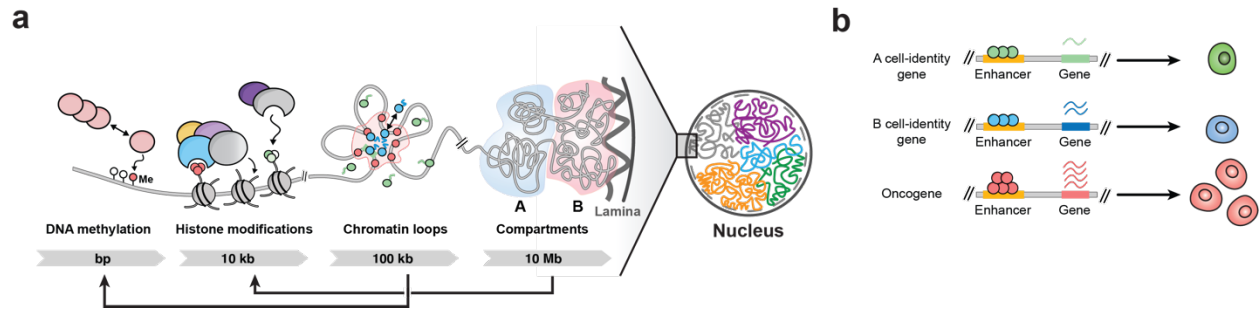
Figure 1.1. Chromatin packing tunes gene expression and cell identity.

a, Schematic illustrating how the genome is packaged and the role of chromatin modifying complexes in regulating this organization. **b,** Cartoon depicting how different cell states are accessed by modulating gene expression. Dysregulated gene expression can result in oncogenic transformation of cells.

The development of small molecules and other chemical biology approaches played pivotal roles in both identifying and characterizing these chromatin modifiers. For example, Schreiber and colleagues developed trapoxin affinity matrix to pulldown and characterize the first histone deacetylase (HDAC1) (**Figure 1.2a**).[7] Furthermore, the development of proximity-dependent labeling methods, such as BioID (biotin ligase), which introduces a covalent biotin tag to proteins in proximity of the target protein, have greatly contributed to identifying chromatin complex members and profiling their chromatin interactome (**Figure 1.2b**).[8–10] For instance, Huang *et al.* fused a biotin ligase to the histone methyltransferase NSD2 and overexpressed this construct in NSD2 stable knockout cells to identify NSD2 interacting partners.[11]
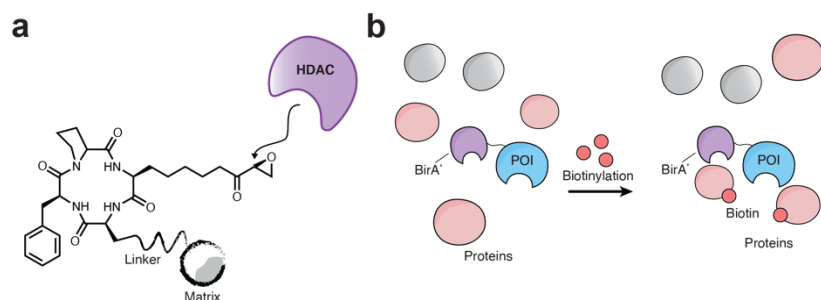
Figure 1.2. Cartoon summarizing two examples of chemical biology approaches to characterize chromatin complexes.

a, Schematic illustrating immobilization of trapoxin to isolate HDAC1. **b,** Cartoon summarizing BioID, where the BirA* biotin ligase is fused to the protein of interest and protein interactors in proximity of the POI are biotinylated.

Due to their key role in transcriptional control, a number of these chromatin modifiers have been implicated in human disease, especially in cancer when dysregulated or mutated.[4,12] For instance, DNA hypermethylation was identified as an epigenetic signature of cancer when leading to the aberrant silencing of tumor suppressor genes.[4] To reactivate these tumor suppressor genes, DNA methyltransferase (DMT) inhibitors as well as HDAC inhibitors were developed, such as decitabine and vorinostat, respectively, both of which are US Food and Drug Administration (FDA)-approved.[4,13,14] Furthermore, cancer cells often display altered chromatin structure mediated by the aberrant expression and/or activity of chromatin modifiers/remodelers.[12] This ultimately leads to altered gene activation and/or silencing. Dysregulation of histone acetyltransferases (HATs) is also commonly associated with oncogenesis such as those driven by mutations in p300 and CBP, typically tumor suppressors.[12]

Besides enzymatic activity, chromatin modifying complexes can also have reader functions that allow them to recognize a variety of chromatin features, such as DNA/histone modifications and DNA linker length. In addition, chromatin modifiers can have numerous scaffolding functions that regulate complex activity and can engage in protein-protein interactions

with transcription factors. Multiprotein chromatin complexes can have several reader domains, allowing for the integration of different PTMs from a single or even neighboring nucleosomes.[2,15] Identifying and understanding these non-catalytic activities of chromatin-modifying complexes is key to understanding their function and role in disease.

Through allosteric regulation, reader domains can transmit information from chromatin to the distal active site of the chromatin modifier and thereby influence the epigenetic landscape. For instance, the de novo methyltransferase DNMT3A contains an ADD (ATRX-DNMT3-DNMT3L) domain that recognizes unmethylated H3K4 (H3K4me0) and this mark stimulates DNMT3A enzymatic activity (**Figure 1.3a**).[16] Structural studies showed that the binding of H3K4me0 induces a conformation change that moves the ADD domain away from the catalytic domain to release autoinhibition and thus serves as an allosteric activator. Furthermore, the PWWP domain of DNMT3a recognizes another chromatin mark, H3K36me2, and recognition of this mark is essential for DNMT3a recruitment and activity at intergenic regions (**Figure 1.3a**).[17]
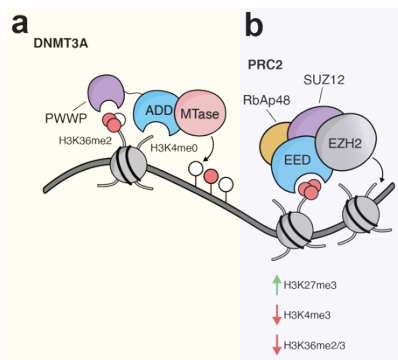
Figure 1.3 Allosteric regulation of chromatin modifier complexes tunes their activity.

a, Cartoon summarizing allosteric regulation by DNMT3A reader domains that tune DNMT3A methylase activity. **b,** Cartoon summarizing how different chromatin marks are read by PRC2 complex members to regulate PRC2 activity.

Another well-studied example includes PRC2 (polycomb repressive complex 2) which mediates methylation at H3K27.[18,19] The histone lysine methyltransferase activity of the complex

resides in EZH1 and EZH2, SET domain-containing proteins. In order for the complex to be enzymatically active additional PRC2 complex members are required, including EED, SUZ12 and RbAp48. The SET domain of EZH2 occupies an autoinhibited state until it associates with SUZ12, which induces a conformation change and relieves autoinhibition to stimulate enzymatic activity.[20,21] Furthermore, H3K27me3 allosterically activates the complex by binding to the EED domain to enable H3K27me3 propagation (**Figure 1.3b**). Structural studies showed that the H3K27me3 peptide resides between EED and the SRM (stimulation-responsive motif) of EZH2. This suggests that SRM transmits the allosteric activation signal from EED to the SET domain to stimulate methyltransferase activity. This is further supported by in vitro studies that showed that mutations in the SRM domain decrease allosteric activation by H3K27me3 as well as by cellular studies where SRM mutations resulted in reduced levels of H3K27me2/3 in embryonic stem (ES) cells. On the other hand, active transcription marks, H3K4me3 and H3K36me2/3, were shown to have an inhibitory effect on methyltransferase activity through binding to the SUZ12 and RbAp48 PRC2 submodules. Overall, these examples illustrate the interplay between allosteric regulation mediated by reader domains and the catalytic activity of writer/easer domains to establish and maintain the epigenetic landscape. Identifying how chromatin modifying complexes are allosterically regulated will therefore be essential to understand their role in disease and develop pharmaceutical strategies to target their activities, such as to the development of inhibitors targeting the EED reader domain of PRC2.[22,23]

1.2. Role of intrinsically disordered regions in gene regulation

Recent discoveries have demonstrated how intrinsically disordered proteins (IDPs) participate in weak interactions and the formation of phase-separated biomolecular condensates (or hubs) in cells.[24–27] It is thought that weak multivalent interactions, such as electrostatic, pi-pi, cation-pi, and hydrophobic interactions, drive the formation of biomolecular hubs (**Figure 1.4a,b**). Although such partitioning suffers an entropic penalty, a multitude of weak multivalent interactions between low complexity, disordered macromolecules can drive their assembly into phase-separated condensates. Participation in weak multivalent interactions is not limited to intrinsically disordered regions (IDRs) but can also be mediated by structured modular domains and oligomerization domains. Furthermore, condensate formation is hypothesized to promote spatiotemporal regulation and enable the nonstoichiometric accumulation as well as the selective partitioning of functionally related proteins/factors.
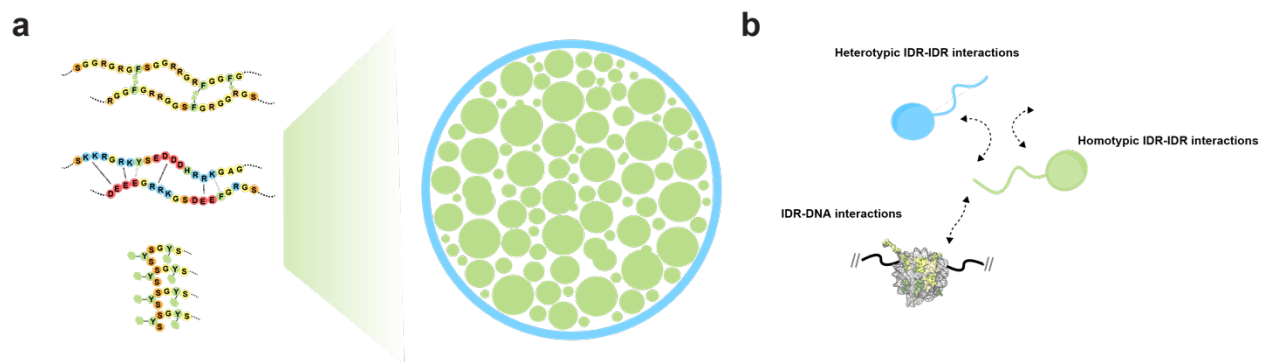
Figure 1.4. Intrinsically disordered regions participate in weak multivalent interactions.

a, Cartoon summarizing the types of weak interactions that result in the formation of biomolecular condensates such as, pi-pi and cation-pi interactions (top left), electrostatic interactions (middle left) and low-complexity repeats (bottom left). **b,** Weak multivalent interactions are highly prevalent in the nucleus, especially between transcription factors, chromatin complexes, histone tails, and DNA, to name a few.

In the field of chromatin biology, these ideas have gained a lot of traction since they may help rationalize how many different factors can bind and concentrate at certain genomic loci, like

enhancers, to regulate transcription.[24,28] Histones also contain N- and C- terminal IDRs that frequently participate in weak multivalent interactions and are regulated by a vast array of post-translational modifications (**Figure 1.4b**). For instance, multivalent nucleosome-nucleosome interactions mediated in large part by the disordered histone tails condense chromatin to repress transcription. Acetylation of the histone tails and the recruitment of coactivators/TFs reverses chromatin condensation, forming a distinct phase from heterochromatin. This was also demonstrated using chromatin arrays in vitro, where acetylation of chromatin in the presence of BRD4 produced a distinct phase that was immiscible with unmodified chromatin.[29] IDRs are especially prevalent in chromatin-associated proteins, from transcription factors to chromatin remodeling proteins, and are difficult to study biochemically given their inherent lack of structure.[24,25,30] Recruitment of TFs that allow for the concentration of coactivators is hypothesized to form condensates at enhancers and promoters to initiate transcription. On the other hand, constitutive heterochromatin condensates are hypothesized to be driven by the binding of methylated DNA by MeCP2 and methylated H3K9 by HP1.[31,32] IDRs can also have numerous regulatory functions in chromatin complexes. For example, the SRM of EZH2 is highly flexible and undergoes a disorder-to-order conformation transition upon binding H3K27me3 and binds to the catalytic SET domain of EZH2 to stimulate PRC2 activity.[18,33]

Dysregulation of biomolecular condensates has been shown to be associated with diseases, in particular cancer, which is characterized by aberrant genomic alterations.[26] For instance, the oncogenic protein EWS-FLI1, which results from the fusion of the disordered activation domain of EWS with the DNA binding domain of the transcription factor FLI1, produces transcriptional condensates that promote transcription of genes associated with Ewing's sarcoma.[34] Another example includes BRD4, which is commonly overexpressed in solid tumors.[25] Overexpression of BRD4, a reader protein that binds to acetylated nucleosomes at active genes, was shown to form aberrant transcriptional condensates correlated with dysregulated transcriptional activation. Furthermore, the majority of human proteins contain IDRs and IDRs are

not only enriched in disease-associated proteins but also frequently harbor mutations. Due to the lack of structural and functional characterization of IDRs, studies instead tend to focus on mutations occurring in folded regions.[30] The development of new tools and computational methods, as well as the advancement of techniques such as NMR and high-resolution microscopy, will be essential to further our understanding of disease-associated mutations in IDRs.[28,30,34]

1.3. Discovery and initial characterization of LSD1

Histone methylation is a key chromatin modification, occurring on both arginine and lysine residues. Lysine methylation at promotors and enhancers is often associated with active gene transcription with H3K4me3 predominately characterizing active promoters, H3K4me2 at both active promoters and enhancers, and H3K4me1 at primed enhancers.[35] However, lysine methylation can also be associated with gene repression. For example, H3K9 methylation is associated with heterochromatin formation and gene repression. The methylation landscape is controlled by the combined activities of both histone methyltransferases and demethylases. Two families of histone lysine demethylases have been identified, the Fe(II) α-ketoglutarate dependent Jumonji domain family, which can remove me1/2/3, and the flavin-adenine dinucleotide (FAD)-dependent lysine-specific demethylase (LSD) family, which can only remove me1/2.

LSD1 was the first histone lysine demethylase identified and was discovered by Yang Shi's group.[36] They identified a member of the CtBP corepressor complex that shares significant homology to FAD-dependent amine oxidases and biochemically characterized its function as a lysine-specific H3K4me1/2 demethylase (LSD1). The methylated lysine residue was proposed to donate a hydride equivalent to FAD to form an imine intermediate that can be subsequently hydrolyzed to produce the demethylated lysine residue and formaldehyde by-product (**Figure 1.5a**). The reduced FAD cofactor is then regenerated by molecular oxygen, producing hydrogen peroxide in the process. Yang Shi's group assessed LSD1's ability to demethylate H3K4 peptides and showed that LSD1 selectively demethylated H3K4me2 over H3K9me2 and could only remove H3K4me1/2 marks but not H3K4me3, consistent with the chemistry of flavin-containing amine oxidases requiring a protonated nitrogen substrate. Assays conducted with native histones and blotting for a diversity of methylation marks confirmed LSD1's substrate specificity. They furthermore developed a formaldehyde dehydrogenase coupled assay to assess LSD1's demethylase activity through the production of formaldehyde upon demethylation as well as monitored the formation of the demethylation reaction products by mass spectrometry to validate

10

the proposed mechanism for LSD1 demethylation. Since the discovery of LSD1, other assays to detect LSD1 activity have since been developed. For example, one assay takes advantage of the generation of hydrogen peroxide during the demethylation reaction and couples this by-product with HRP and a fluorogenic electron acceptor for a fluorescence-based assay that is easily scalable to a 96- or 384-well plate reader format (**Figure 1.5b**).[37]
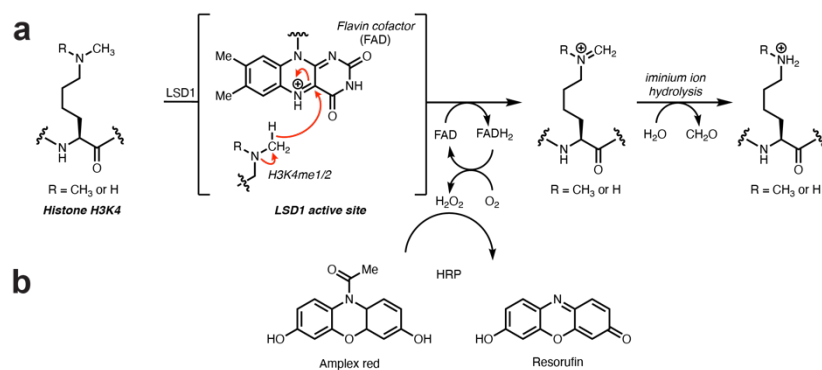
Figure 1.5. LSD1 is an H3K4me1/2 demethylase.

a, Mechanism of LSD1 H3K4me1/2 demethylation. **b,** HRP-coupled assays detect H_2O_2 generated during demethylation to convert Amplex red to a fluorogenic substrate. This is used to monitor the kinetics of LSD1 H3K4me1/2 demethylation.

By removing H3K4me1/2, a mark associated with active transcription, and through its association with other repressor proteins, namely CoREST and HDAC1/2, LSD1 was hypothesized to function as a transcriptional corepressor. In agreement with this, RNAi knockdown of LSD1 was shown to result in an increase in H3K4 methylation and derepression of target genes.[36] Further studies using immunoaffinity purification strategies identified LSD1 to be a key member of the CoREST transcription repressor complex with CoREST (RCOR1), HDAC1/2, ZNF217, PHF21A and HMG20B, as well as the NuRD complex with RCOR3, ZMYM2/3, GSE1, CTBP1, HMG20A, PHF21B, HSPA1A and RREB1.[35] Complexation to CoREST has been shown to be essential for LSD1 demethylase activity with nucleosome substrates as well as for protection from proteasomal degradation. [35]

Since the initial discovery of LSD1, numerous studies have since elucidated LSD1's function in several key cellular processes, from its involvement in cellular differentiation and development, to its role in sustaining oncogenic gene programs that block differentiation. For instance, work by the Rosenfeld lab showed that while heterozygous LSD1 deletion mice were normal and fertile, no homozygous LSD1 deletion embryos could be detected after embryonic day 7.5, highlighting LSD1's essential role in development.[38] Furthermore, LSD1 knockout in mouse ESCs resulted in impaired differentiation, increased apoptosis, and DNA hypomethylation through destabilization of the non-histone LSD1 substrate, DNMT1, connecting LSD1 activity to DNA methylation.[6] Another study demonstrated that while LSD1 knockout in mouse ESCs does not alter the maintenance of the ESC state, it is required for early embryonic development and results in aberrant expression of developmental regulators, consistent with previous findings that showed that LSD1 is essential for the survival of postimplantation embryos.[39] Further genomic insights from the Young lab showed that LSD1 is required for enhancer decommissioning, or the silencing of enhancers involved in pluripotency, during ESC differentiation.[40] By chromatin immunoprecipitation following by sequencing (ChIP-seq), LSD1 was shown to occupy the enhancers and core promoters of actively transcribed genes that control ESC state and LSD1 inhibition resulted in the incomplete silencing of these genes during differentiation. The Young lab proposed the following model where during ESC differentiation, the levels of transcriptional activators, such as Oct4 and p300, decrease, along with H3K27ac, allowing LSD1 to remove the activating mark, H3K4me1 and silence ESC enhancers.

LSD1 also plays an essential role in epithelial-mesenchymal transition (EMT), the cellular process of converting epithelial cells to mesenchymal cells where epithelial cells lose cell polarity, cell-cell adhesion, and have increased migratory and invasive properties.[41] EMT is essential during embryonic morphogenesis, however, EMT also plays a role in cancer progression, allowing carcinoma cells to invade and disseminate.[41] A key feature of EMT is the loss of E-cadherin, a cell adhesion molecule essential in the maintenance of epithelial cells. Snai1, a direct repressor

of E-cadherin transcription, was shown to physically recruit LSD1 to its target genes. LSD1 recruitment is essential for driving Snai1-mediated transcriptional repression. Depletion of LSD1 impairs Snai1-mediated repression of epithelial genes, which is essential for development and consistent with LSD1 knockout mouse embryos dying at gastrulation. Due to its role in promoting EMT, LSD1-Snai1 contributes to cancer cell invasion and metastasis and more recently LSD1-Snai1 have been identified to be key modulators contributing to leukemia development and pathogenesis.[42]

LSD1 is also essential in hematopoiesis and has been found to interact and cooperate with a number of key hematopoietic transcription factors, including GFI1/1B, RUNX1, GATA2, and TAL1, to maintain normal hematopoietic stem cell (HSC) function, self-renewal capacity and myeloid differentiation.[43–47] For instance, Kerenyi *et al.* knocked out LSD1 at different stages of development to examine its effect on hematopoiesis and found that LSD1 is required for HSC differentiation as well as HSC renewal.[46] LSD1 knockout was associated with impaired gene silencing that prevents blood cell maturation, including terminal granulocytic and erythroid differentiation, and results in pancytopenia. Consistent with earlier work by Whyte *et al*., LSD1 was found to regulate both promoters and enhancers of stem and progenitor genes in hematopoietic cells through modulation of H3K4me1/2, where knocking out LSD1 in differentiating cells led to the aberrant expression of HSC-specific genes, trapping the cells in an undifferentiated state.[40,46] Overall, LSD1 is essential for not only early hematopoietic differentiation but also for terminal blood cell maturation in multiple lineages, indicating that pharmacological inhibition of LSD1 could have toxic hematological side effects.[46,47] These studies highlight the biological role of LSD1 during key cellular processes such as differentiation and development.

1.4. Key LSD1 structural studies

Since the initial discovery and biochemical characterization of LSD1, several reports detailing LSD1's structure followed. The crystal structure of LSD1(171-852) complexed to CoREST(286-482) was solved by Yang *et al.* in 2006 (**Figure 1.6a**).[48] LSD1/CoREST forms an elongated structure with the amine oxidase domain (AOD) and SWIRM domain of LSD1 on one end and the CoREST SANT2 domain on the other. The AOD is separated by a 92 amino acid insert called the tower domain, which consists of two long helices packed in an antiparallel fashion, dividing the AOD into two functional lobes and providing a docking site for CoREST.[49] One lobe binds FAD while the other mediates substrate recognition and binding. FAD sits deep inside the AOD pocket and is positioned in the pocket by K661. Although SWIRM domains have been shown to bind to DNA, the LSD1 SWIRM domain lacks this DNA-binding patch.[49] The linker of CoREST forms a parallel coiled coil with the LSD1 tower domain to position the SANT2 domain at the opposite end of the LSD1 AOD domain to form a bi-lobed like structure.[48] The LSD1 AOD shares extensive homology to maize polyamine oxidase (mPAO), however, the residues along the rim of the active site diverge, with LSD1 having a far shallower and accessible substrate binding pocket in comparison to the long tunnel that mPAO uses to bind long, linear polyamines. The rim of the LSD1 substrate pocket is acidic, guiding the binding of the positively charged H3 N-terminus. Given that CoREST is necessary for LSD1 to effectively demethylate a nucleosome substrate, Yang *et al.* carried out NMR studies with ^{15}N-labeled CoREST SANT2 and observed DNA binding by ^{15}N/^{1}H HSQC NMR.[48] Mutations that weaken the hydrophobic contacts between the SWIRM and AOD domain greatly reduce catalytic activity. The SWIRM and AOD domain form a highly conserved cleft and it was hypothesized that this cleft may engage the histone tail, however, initial attempts to crystallize the H3 peptide with LSD1-CoERST failed.
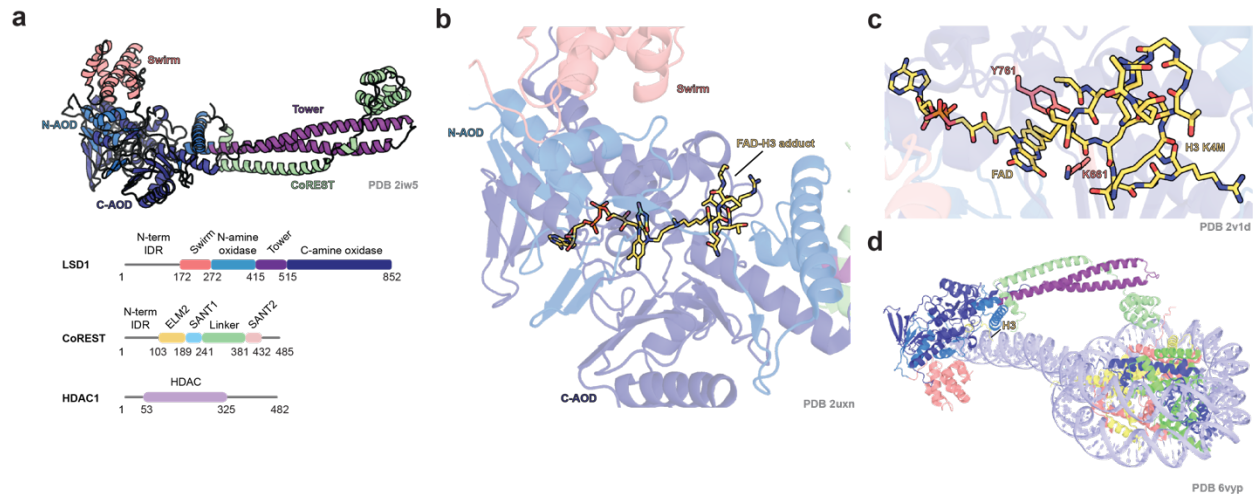
Figure 1.6. Structural elucidation of LSD1.

a, Crystal structure of LSD1-CoREST colored by each domain (top) and domain maps for LHC complex members (bottom). PDB: 2iw5 **b,** Crystal structure of LSD1-CoREST stably bound to an H3 peptide mimic (N-methylpropargyl-K4 H3$_{1-21}$ peptide). The FAD-peptide adduct is shown in stick mode (yellow). PDB: 2uxn **c,** Crystal structure of LSD1-CoREST bound to the H3 tail mimic, H3K4M. FAD and the H3K4M peptide are shown in stick mode (yellow). PDB:2v1d **d,** Crystal structure of LSD1-CoREST bound to the 197-bp NCP. PDB: 6vyp.

Yang *et al.* took advantage of a covalent tethering strategy to stably link the H3 peptide to FAD with an N-methylpropargyl-K4 H3$_{1-21}$ peptide (**Figure 1.6b**).[50] Residues 1-7 of the H3 peptide were found to directly interact with the active site cavity of LSD1 and adopt a serpentine conformation by making three γ-turns. The remaining residues of the peptide were not resolved and are predicted to bind a surface groove between the SWIRM and AOD. Notably, despite only the first seven N-terminal residues of H3 being well ordered in the crystal structure, LSD1 makes extensive contacts along the length of the H3 tail and cannot demethylate substrates containing fewer than 16 residues efficiently. Mutagenesis studies confirmed that the anionic residues in the LSD1 pocket are essential for engaging the N-terminus of H3 via electrostatic interactions and that an additional glycine residue at the N-terminus of H3 is not tolerated, consistent with H3K4

specificity. Forneris *et al.* used a similar approach except H3K4 was replaced with a methionine residue to increase the H3 peptide binding affinity without relying on covalent FAD attachment for crystallization studies (**Figure 1.6c**).[51] In this structure, the first 16 residues of H3 are resolved and the conformation of the peptide is identified to extend to the rim of the binding pocket where the C-terminus is solvent accessible. By using H3K4M, the structure of dimethyl K4 could be modeled into the structure to predict H3K4me2 orientation in the binding pocket. H3K4M is positioned in front of the cofactor with the side chain orientated toward the FAD flavin ring. In the modelled structure, H3K4me2 is positioned near the N5 atom of flavin (approximately 3 Å) while the dimethylamino group is orientated by Y671, structural features that are consistent with other flavin-dependent amine oxidases.

Recent structural studies have extended our understanding of how LSD1-CoREST docks to the nucleosome core particle (NCP) as well as the structure of the LSD1-CoREST-HDAC1 ternary complex. Structural studies with the NCP containing extranucleosomal DNA (197-bp NCP) highlight the role of CoREST in promoting LSD1 demethylase activity on nucleosome substrates (**Figure 1.6d**).[52] CoREST was found to make critical contacts to both histone and DNA components of the nucleosome as well as aid in the positioning of LSD1, which interacts with the H3 tail and extranucleosomal DNA. Surprisingly, LSD1's catalytic domain is positioned roughly 100 Å away from the NCP core. This is hypothesized to allow LSD1 to probe chromatin accessibility since LSD1 removes the active H3K4me2 mark. Deletion studies that truncate the H3 tail starting at residue 21 supports this docking mode as removal of 12 or more residues significantly abolishes LSD1 demethylase activity on these mutant NCPs compared to 197-bp nucleosomes containing full-length H3. A second non-canonical LSD1-CoREST binding mode to the NCP was also identified, where the LSD1 AOD domain docks onto the octamer face with the catalytic pocket of the AOD orientated away from the octamer core. A HI-FI nucleosome binding assay was employed to investigate whether these two binding modes exist in solution. In this assay nucleosomes are fluorescently labeled at specific histone residues and the quenching of

fluorescence signal upon binding is measured. Quenching of the fluorescent probe installed on H3 K27C confirmed the canonical binding mode proposed in this study while quenching fluorescently labeled H2B Q48C supports the presence of the second binding mode. Further mutagenesis studies of NCP residues involved in LSD1-CoREST binding validated these findings. Overall, this study clarified why CoREST is essential for promoting LSD1's demethylase activity with NCP substrates and provides structural insights into LSD1-CoREST binding to the nucleosome.

LSD1 itself mediates additional key contacts to the nucleosome that are not captured with the truncated LSD1 construct typically used in biochemical studies, such as LSD1Δ170. Dhall *et al.* revealed that full-length LSD1 was capable of binding and demethylating mononucleosomes containing semisynthetic H3K4me2 in the absence of CoREST.[53] The demethylation kinetics with full-length LSD1 and an H3K4me2(1-21) peptide substrate was identical to previously reported values with truncated LSD1 (LSD1Δ150). Furthermore, by electrophoretic mobility shift assays (EMSA), full-length LSD1 was found to bind to both mononucleosomes as well as 12-mer nucleosome arrays. This was further quantified by taking advantage of a microscale thermophoresis approach with Cy5 labeled mononucleosomes labeled at the 5' end of the Widom 601 DNA sequence. Full-length LSD1 was able to demethylate nucleosome substrates as assessed by single-turnover nucleosome demethylation assays with mononucleosomes containing doubly modified H3K4me2. The addition of CoREST enhanced demethylation by ~2-fold. Therefore, full-length LSD1 is sufficient to bind and demethylate nucleosome substrates, debunking previous assumptions that CoREST is essential for mediating LSD1 nucleosome activity. Full-length LSD1 as well as a truncated construct consisting of the LSD1 N-terminus (LSD1 1-170) were shown to bind to DNA, as assessed by a fluorescence polarization binding assay with AlexaFluor 488-labeled Widom 601 DNA. These results suggests that the LSD1 N-terminus may promote nucleosome binding by engaging the nucleosomal DNA.

The CoREST repressor complex has dual enzymatic activities through LSD1 (demethylation) and HDAC1/2 (deacetylation) (**Figure 1.6a**). To date, most studies have focused on LSD1 lacking its disordered N-terminus complexed to the LINKER-SANT2 region of CoREST. Song *et al.* examined the enzymatic cross-talk between LSD1-HDAC1 with nearly full-length LSD1-CoREST-HDAC1 complex (LHC) and investigated it's structure alone and in complex with the nucleosome by cryo-electron microscopy (cryo-EM).[54] Through real-time ^{1}H NMR-based demethylase and deacetylase assays with H3K4me or H3K9ac substrates, respectively, LSD1 and HDAC1 reactions were observed to undergo non-Michaelis-Menten kinetics consistent with the activity of these enzymes being coupled. Furthermore, inhibitors of either enzyme were found to influence the activity of the other enzyme, where, for example, inhibition of LSD1 is associated with some level of HDAC1 inhibition. Addition of $InsP_6$, an HDAC activator, altered LSD1 demethylase activity, increasing the effective demethylase rate at high H3K4me substrate concentrations and slowing the reaction rate at low substrate concentrations. By monitoring the demethylation and deacetylation activity of LHC at the same time with a doubly modified substrate, H3K4meK9ac, LSD1 and HDAC1 was found to engage the substrate in a mutually exclusive manner. Given the inherent flexibility of the LHC complex, a combination of SAXS, cross-linking-MS, negative-stain EM, and cryo-EM were used to piece together the structure of this complex that has eluded structural characterization to date. These approaches revealed a bi-lobed structure of the complex with LSD1 on one end and HDAC1 on the other, separated by the LSD1 Tower domain/CoREST LINKER. By fitting the coordinates of the known complexes to the EM data, a model of the complex could be generated and revealed that HDAC1 is located at the far end of the LSD1 Tower domain and contacts the SANT2 domain of CoREST. To gain insight into how this complex may interact with its mononucleosome substrate, a propargylamine mimic of H3K4me2 was used to covalently trap LSD1 and stabilize the LHC complex on the nucleosome. A 1:1 complex of LHC:nucleosome was successfully isolated by size-exclusion chromatography and structurally interrogated by cryo-EM. When LSD1 is engaged with the nucleosome (due to

the covalent trapping approach), HDAC1 is oriented far away from the nucleosome, likely unable to access the H3 tails. This is consistent with the mutually exclusive substrate engagement characterized with the ^{1}H NMR-based enzyme assay. LSD1 has previously been studied biochemically in complex with truncated CoREST, however, the biologically relevant form of the complex has deacetylase activity. Therefore, despite not being able to obtain a high-resolution structure of the complex, Song *et al.* provide key structural and biochemical insights into the LHC complex.

LSD1 can also form a complex with a class of transcription factors (TFs) that contain a conserved N-terminal SNAG domain.[55] These transcription factors mimic the histone tail and bind in the LSD1 active site. As such, treatment with LSD1 inhibitors as well as the H3 peptide can block SNAG binding. The Orkin lab first identified the LSD1-SNAG TF interaction by profiling GFI1B interacting factors using a proximity labelling approach followed by mass spectrometry.[44] The SNAG domain of GFI1B was identified to mediate LSD1 recruitment as introduction of a point mutation (P2A) in this domain abolished LSD1-CoREST interaction in addition to inactivating GFI1B activity (**Figure 1.7a,b**). LSD1 is then recruited to GFI1B target genes as assessed by chromatin immunoprecipitation studies and is essential for gene repression. Accordingly, downregulation of LSD1 or CoREST derepressed GFI1B target genes and perturbed hematopoietic differentiation, highlighting the role of this complex in controlling differentiation.
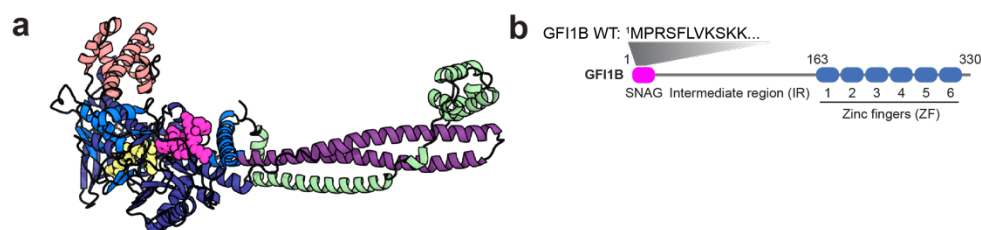
Figure 1.7. Structure of LSD1 bound to the SNAG peptide.

a, Crystal structure of LSD1 bound to the SNAG peptide of GFI1B. LSD1 FAD cofactor is in yellow while the SNAG peptide is in magenta, both depicted in sphere mode. PDB: 2y48. **b,** Domain map of GFI1B with the SNAG N-terminus highlighted in magenta.

The LSD1 and SNAG TF protein-protein interaction was further characterized with the epithelial-mesenchymal transition (EMT) factor Snail1. Through the design of LSD1 domain truncation mutants followed by immunoprecipitation with Snail, they identified that the AOD is the domain responsible for Snail1 binding. Similar experiments were conducted to identify key residues in the SNAG domain that are critical for mediating the LSD1-Snail1 interaction, several of which are similar to H3 residues (Arg3, Arg8 and Lys9), supporting the hypothesis that the SNAG domain mimics H3 to recruit chromatin-modifying enzymes.[55] Crystal structure studies by the Mattevi lab validated this hypothesis where the SNAG domain was shown to bind to the enzyme active-site cleft of LSD1 (**Figure 1.7a,b**).[56] Furthermore, co-expression followed by western blotting experiments identified that CoREST is important for the stability of the Snail1-LSD1-CoREST complex and enhances LSD1-Snail1 association.[55] By investigating the function of this complex in EMT, LSD1 interaction with Snail1 was identified to be essential for LSD1 recruitment to the E-cadherin promoter and knockdown of either Snail1 or LSD1 expression was sufficient to suppress cell migration, suggesting that Snail1-LSD1 complex formation is critical for repressor function activity and mediating both EMT induction and cancer metastasis. Overall, the SNAG-mediated recruitment of LSD1 to chromatin provides mechanistic insight into one of the modes whereby transcription factors can recruit chromatin-modifying enzymes to target gene promoters.

1.5. LSD1's role in cancer and developing pharmacological LSD1 inhibitors

LSD1 is overexpressed in many cancer types and is implicated in disease progression, prompting interest in developing inhibitors targeting LSD1 demethylase activity. High expression levels of LSD1 have been shown to be correlated with differentiation arrest in both hematological and solid tumors. For instance, although LSD1 has not been shown to be mutated in AML, it is highly expressed in leukemia blast cells in 60% of AML patients.[57] Genetic knock down of LSD1 in AML and small-cell lung cancer (SCLC) cell lines was demonstrated to alleviate this differentiation blockade and suppress proliferation. Although LSD1 expression was correlated with prostate and breast cancer progression, it was not systematically studied as a viable drug target until the Somervaille group investigated LSD1's role in myeloid leukemia.[58] Acute myeloid leukemia (AML) is an aggressive blood cancer characterized by a myeloid differentiation block and resulting accumulation of immature blast cells in the bone marrow and blood.[59] In this clonal hematopoiesis disorder, leukemic stem cells develop self-renewal capacity, enhanced proliferation and impaired differentiation. Chromosomal alterations occurring in transcription factor genes that result in either the aberrant expression or fusion of myeloid transcription factors/chromatin modifiers are frequently implicated in AML. For example, MLL-AF9 leukemia is characterized by the fusion of the N-terminal CXXC-domain containing DNA-binding segment of the mixed lineage leukemia lysine-specific methyl transferase 2A (KMT2A) gene with AF9, a chromatin regulator. The Somervaille group first discovered LSD1 as a vulnerability in cancer by knocking down LSD1 in MLL-AF9 AML mouse model and observed a reduction in colony-forming (CFC) potential.[58]

Pharmacologic inhibitors of LSD1 were also shown to be effective in abrogating clonogenic potential and inducing differentiation in both murine and primary human MLL leukemia cells, phenocopying LSD1 knockdown.[57,60] LSD1 was required to maintain the oncogenic gene expression program as LSD1 knockdown resulted in the induction of differentiation and expression of apoptosis genes as well as the downregulation of MLL leukemia stem cell (LSC)-

associated genes. Furthermore, pharmacological inhibition of LSD1 was shown to potentiate all-trans retinoic acid (ATRA) treatment, a successful treatment strategy for acute promyelocytic leukemia (APL) AML, in non-APL AML, a cancer that is normally unresponsive to ATRA treatment.[61] Inhibition of LSD1 results in epigenetic reprogramming, allowing combinatorial treatment with ATRA to effectively target leukemia-initiating cells and promote AML cell differentiation. The therapeutic promise of targeting LSD1 in AML has spearheaded the development of potent LSD1 inhibitors.

Given the homology of the LSD1 amine oxidase domain with monoamine oxidases, the ability of monoamine oxidase inhibitors to inhibit LSD1 was assayed, identifying tranylcypromine as an irreversible FAD-dependent LSD1 inhibitor.[37] Tranylcypromine proved to be a great starting point towards developing more potent and specific small molecule LSD1 inhibitors (**Figure1.8a,b**). N-alkylation as well as modifications to the phenyl ring yielded far more potent and specific LSD1 inhibitors.
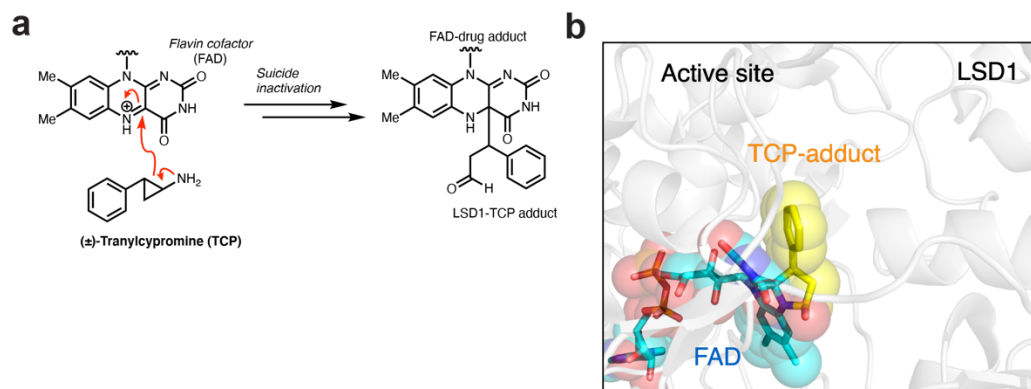
Figure 1.8. Mechanism based LSD1 inhibitors.

a, LSD1 irreversible "suicide" inhibitors, such as tranylcypromine, are mechanism-based and lead to permanent inactivation of the enzyme. The mechanism of inhibition likely proceeds by radical chemistry and is abbreviated here for simplicity. The inhibitor in essence mimics the native substrate, H3K4me1/2. The inhibitor binds and donates a C-C bond rather than a C-H bond and relieves cyclopropane strain.[37] **b,** Crystal structure of the FAD-TCP adduct. PDB: 2uxx.[62]

Glaxo SmithKline (GSK) screened a panel of cancer cell lines with the tranylcypromine derivative GSK2879552 and found that only a subset of cancers respond to LSD1 inhibition, notably AML and SCLC.[35] Why only certain cancer subtypes respond to LSD1 inhibition is an area of active research and this question in part motivated our studies on LSD1. Tranylcypromine derivatives ORY1001 and GSK2879552 entered clinical trials for cancer therapy as well as combined treatment with TCP derivatives/ATRA (all-*trans* retinoic acid) (**Figure 1.8**).[63] Due to the associated hematotoxicity with the use of tranylcypromine derivatives, most LSD1 inhibitors have failed to advance past phase I/II clinical trials. For example, clinical trials of GSK2879552 for AML and SCLC have been terminated. However, ORY1001 is moving forward to phase II clinical trials for AML in combination with azacitidine, a DNA methyltransferase inhibitor, as well as another tranylcypromine inhibitor, IMG7289, which is entering phase II clinical trials for myelofibrosis and essential thrombocythemia.[63–65] The development of LSD1 inhibitors that minimize hematotoxicity is being actively pursued. Reversible inhibitors that do not covalently modify LSD1's FAD cofactor in particular are attractive, especially given LSD1's critical function in hematopoiesis. Celegene developed the first reversible LSD1 inhibitor to enter clinical trials, CC-90011, however, like irreversible inhibitors, patients in phase I studies suffered from some hematotoxic side effects such as thrombocytopenia and neutropenia although the inhibitor has advanced to phase II trials in patients with SCLC (**Figure 1.9**).[63,66]
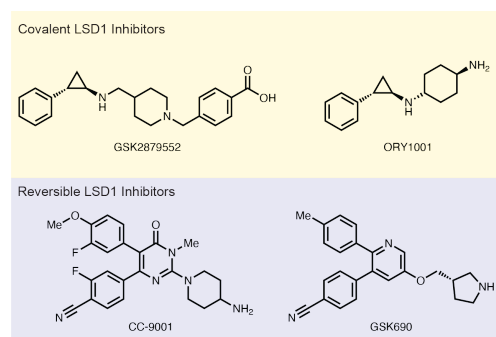
Figure 1.9. Structures of covalent and reversible LSD1 inhibitors.

With the discovery that the N-terminal 21 amino acids of Snail1 can competitively occupy the LSD1 active site by mimicking the conformation of the H3 tail, short peptides based on the Snail1 N-terminus were screened for their ability to reversibly inhibit LSD1.[67] Although longer Snail1 peptides displayed stronger binding affinities for LSD1, their ligand efficiency was poor. Forneris *et al.* developed a 21-mer peptide homologues to H3K4 where Lys4 was replaced with methionine and this peptide displayed nanomolar inhibitory activity ($K_i = 50$ nM) against the LSD1-CoREST complex. Macrocyclic versions were later developed and demonstrated to display modest antitumor activity. Another class of reversible LSD1 inhibitors that were discovered to display potent LSD1 inhibitory activity are based off a six-membered heterocyclic scaffold. GSK-690 was reported to inhibit LSD1 with nanomolar affinity and excellent selectively over MAO-A to induce differentiation as well as block cell growth in THP-1 cells (**Figure 1.9**). Niwa *et al.* determined the co-crystal structure of LSD1 with a close structural analog of GSK-690, illustrating the key contacts this reversible inhibitor makes with LSD1 active site residues.[68] For example, the piperidine ring binds the negatively charged cavity and interacts with Asn540 and Asp555 while the nitrile group is involved in hydrogen bonding interactions with the key catalytic residue, Lys661. Notably, Celgene's CC-90011 compound makes a similar hydrogen bonding interaction between its nitrile substituent and Lys661 and its pyrrolidine side chain forms a salt-bridge with Asp555.[66]

LSD1 inhibitors were designed to inhibit demethylase activity under the assumption that the demethylase activity of LSD1 is necessary for cancer cell survival. However, we along with others recently showed that the demethylase activity of LSD1 is not essential for AML proliferation and that LSD1 inhibitors instead exert their antiproliferative effects by disrupting interactions between LSD1 and GFI1/ GFI1B, revising the mechanism of action of LSD1 inhibition.[69,70] LSD1-GFI1/GFI1B complex disruption is accompanied by its dissociation from chromatin and subsequent activation of myeloid differentiation genes. However, GFI1/GFI1B are critical regulators of hematopoietic differentiation and disruption of the LSD1-GFI1/GFI1B complex is

associated with hematotoxicity, which has plagued the progression of LSD1 inhibitors in clinical trials to-date. Therefore, LSD1 inhibitors, such as GSK-LSD1, block the LSD1-GFI1/GFI1B interaction and enzyme activity, the former of which drives both the anti-cancer effects and hematological toxicity (**Figure 1.10**).
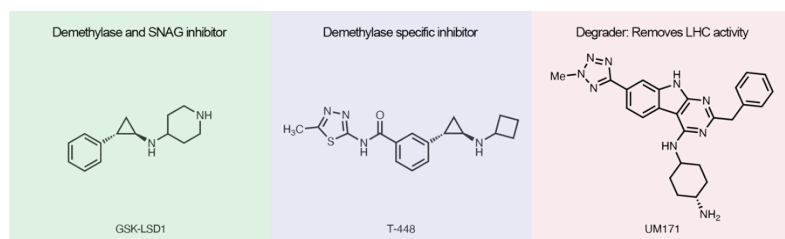
Figure 1.10. Chemical structures of three mechanistically distinct LSD1 inhibitors.

This revised mechanism is supported by the identification of small molecules that selectively inhibit LSD1 demethylase activity while maintaining the LSD1-GFI1/B interaction.[71,72] Such inhibitors are desirable for the treatment of neurodevelopmental disorders, where dysregulation of H3K4 methylation has been implicated in their pathogenesis. Takeda screened small molecules that enhance H3K4 methylation by inhibiting LSD1 demethylase activity while minimizing LSD1-GFI1B complex disruption, and discovered T-448, a specific inhibitor of LSD1 demethylase activity (**Figure 1.10**). T-448 was identified to irreversibly inhibit LSD1 demethylase activity while preserving LSD1 scaffolding functions by generating a covalent-FAD adduct that rapidly forms a compact *N*-formyl-FAD adduct (**Figure 1.11**).
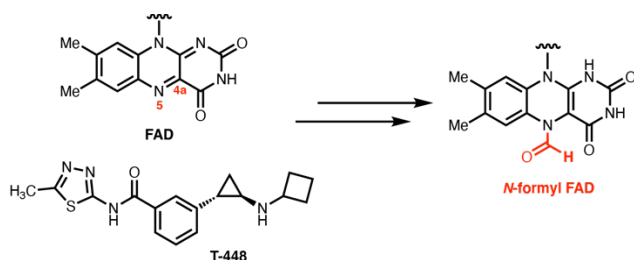
Figure 1.11. T-448 reacts with FAD to form *N*-formyl FAD.

Co-immunoprecipitation studies confirmed that treatment with T-448 preserves the LSD1-GFIB interaction in an erythroleukemia cell line (TF-1a). Surprisingly, although T-448 treatment initially dissociates LSD1-GFI1B, binding is restored in a time-dependent manner as assessed by a surface plasmon resonance binding assay. This time-dependent recovery of the LSD1-GFI1B complex is not observed with a control tranylcypromine derivative, T-711. By LC-MS, T-448 produced a reduced form of FAD with an N5 formyl group in addition to the full T-448-FAD adduct, where time-dependent increase in *N*-formyl-FAD and decrease in the full adduct was observed. Furthermore, formylated FAD was apparent in the co-crystal structure of LSD1 after T-448 treatment. Superimposition of this structure with LSD1 bound to the N-terminal GFI1B peptide demonstrated that GFI1B is still accommodated in the LSD1 pocket in the presence of *N*-formyl-FAD. T-448 was further validated in a mouse model and T-448 treatment not only increased H3K4me2 levels in the brain and improved learning but also did not alter blood cell numbers, consistent with selective inhibition of LSD1 demethylase activity. Beyond its therapeutic use, T-448 also serves as a powerful chemical tool to study the function of LSD1's demethylase activity versus its active site scaffolding activity, both of which are blocked by other tranylcypromine derivatives such as GSK-LSD1.

Another useful tool to study LSD1's scaffolding activities would entail selective degradation of LSD1 by a small molecule. Towards this end, UM171 was recently discovered to degrade LSD1/CoREST via the KBTBD4 Cul3-RING ubiquitin ligase complex (**Figure 1.10**).[73] UM171 was initially identified as a self-renewal agonist for human hematopoietic stem cell (HSC) expansion. HSCs cultured *ex vivo* are prone to losing their self-renewal capacity, in part due to changes in H3K4me2 and H3K27ac levels. Inhibition of class I HDACs as well as LSD1 has been shown to partially reverse *ex vivo* HSC dysfunction. UM171 was identified to increase the amount of KBTBD4 bound to LSD1/CoREST, indicating that LSD1/CoREST is likely a natural substrate for KBTBD4. Degradation of LSD1, CoREST and HDAC2 were observed after UM171 treatment, consistent with targeted degradation of the LHC complex, and both H3K4me2 and H3K27ac

marks were maintained at physiological to supra-physiological levels to preserve stem cell activity. However, the molecular details of how UM171 results in increased LHC degradation is not fully understood. Despite the fact that the mechanism of UM171-mediated degradation is poorly understood, the ability to degrade the LHC complex is a powerful tool to uncover scaffolding functions of LSD1 outside of its active site, especially in comparison to GSK-LSD1 and T-448 which target the FAD cofactor (**Figure 1.10**).

1.6. CRISPR-suppressor scanning and application to study LSD1

As the previous section on LSD1 inhibitors alluded to, small molecules not only have the potential to treat diseases, but they can also serve as powerful tools to investigate the function of the targets they modulate. However, their utility hinges on our ability to deconvolute the mechanism of action of these small molecule inhibitors, including verifying on-target engagement and how target engagement results in a specific biological outcome. To confirm that a chemical probe works by inhibiting its intended target, chemical biologists have taken advantage of a genetic approach involving the generation of a drug-resistance conferring allele, or a mutation in the presumed protein target that blocks the effects of the small-molecule.[74] Identifying drug-resistance conferring alleles largely relied on clinical resistance data or was guided by insight provided by structural biology studies. While drug-resistance poses a major challenge for the therapeutic efficacy of small molecule inhibitors, these resistance mutations are useful for confirming that the inhibitor is engaging the intended target and provide structure-activity insight on the target. For instance, resistance mutations have plagued drug development efforts targeting BCR-ABL leukemias since the discovery of the first ABL tyrosine kinase inhibitors.[75–77] After resistance mutations started to emerge with imatinib, dasatinib, a type I kinase inhibitor that targets the active form of ABL kinase, was developed and shown to have antineoplastic activity. However, another resistance mutation, Thr315Ile, was identified and is refractory towards dasatinib treatment. This mutation was shown to be uniquely resistant to both imatinib and dasatinib. Thr315 in ABL is a key gatekeeper residue, modulating access to the ATP-binding site and stabilizing the active conformation of the ABL kinase domain, and participates in a key hydrogen bond with the pyrimidine NH of imatinib/desatinib.[78] Mutation of Thr315 to Ile not only eliminates this hydrogen bond but also produces a steric clash, thereby creating a "bump" in the drug binding site. Structural studies characterizing this resistance mutation identified a key network of hydrophobic interactions, termed the hydrophobic spine, that are stabilized by the Thr315Ile mutation to lock the kinase in an active form. Ponatinib was then developed that creates

a compensatory "hole" in the drug scaffold that can now accommodate the "bumped" Thr315Ile binding site and overcome the Thr315Ile resistance mutation.[79,80]

Based on structure-guided information, chemical biologists have also designed drug-resistance conferring mutations de novo to validate drug on-target activity. One such case study includes the development of SH099, a small molecule allosteric inhibitor of the non-receptor tyrosine phosphatase SHP2.[81] Activating mutations in SHP2 have been associated with numerous diseases, including developmental pathologies and multiple cancer types, and genetic knockdown of SHP2 was shown to suppress tumor cell growth. However, selectively targeting protein tyrosine phosphatases with small molecules has remained a challenge due to their highly conserved and polar catalytic sites. Novartis developed a screening strategy to identify SHP2 allosteric inhibitors that lock SHP2 in an auto-inhibited conformation and identified SHP099 as a selective SHP2 inhibitor. To confirm that the antiproliferative activity of SHP099 was due to on-target activity, point mutations SHP2 Thr253Met/Gln257Leu were designed based on the co-crystal structure that are predicted to disrupt or "bump out" SHP099 binding. Overexpression of SHP2 Thr253Met/Gln257Leu in KYSE520 cells rescued SHP2 cell growth in the presence of SHP099 and SHP2 Thr253Met/Gln257Leu was enzymatically active in the presence of SHP099 in biochemical studies. In this way, structural knowledge was imperative in the design of drug-resistance conferring mutations that led to the validation of SHP099 on-target inhibition.

Historically, designing drug-resistance conferring alleles was a challenging task and they were often serendipitously discovered by random mutagenesis approaches.[82] Recent advances in molecular biology have enabled the ability to introduce a library of protein variants in a pooled fashion.[74] Overexpression of a pool of target mutants generated by either error-prone cloning or directly synthesized has been used to identify resistance mutations.[83,84] For example, MITE-seq (mutagenesis by integrated tiles and sequencing) relies on the overexpression of protein variants where each amino acid is sequentially varied to one of the other canonical amino acids, allowing for the investigation of all possible point mutations. However, this method relies on variant

overexpression which may be difficult to introduce into mammalian cells, especially depending on the cell-type and target size, and complicate interpretation of the results due to dosage effects. Despite these shortcomings, these approaches are especially useful in the cases where structural data is lacking.

Genome-editing technologies can directly manipulate the endogenous gene allowing for the discovery of drug-resistance conferring alleles in a faster and higher throughput format.[85] In this approach, CRISPR-Ca9 is used to perform high-density mutagenesis of a target protein in a pooled fashion, where the target protein is either essential for cell growth or its activity is tied to another functional readout.[85–88] After the pool of single-guide RNAs (sgRNA) that tile the protein sequence are introduced, CRISPR-Cas9 will introduce double-stranded DNA breaks at sites specified by the sgRNA, which is subsequently repaired. This repair process is error prone and results in insertion-deletion mutations, or indels. These indel mutations can either result in frameshift mutations, if the indel disrupts the codon reading frame, or an in-frame mutation. Frameshift mutations typically result in knockout of the protein, and if that protein is essential for cell survival, then cells containing this sgRNA are depleted from the pool. However, in-frame mutations result in nearly full-length protein, and these mutations are not necessarily lethal. The resulting population of cells contains heterogenous in-frame mutations spanning the entire coding sequence of the target protein, including structural alterations across the drug binding site.

To demonstrate this approach, the Vakoc group designed a library of sgRNAs tiling BRD4, an essential gene in AML, and showed that for an essential gene the sgRNAs that worked the best to reduce cell viability were those targeting functional domains of BRD4, the essential bromodomains.[85] This was contrary to the popular belief that sgRNAs targeting the 5' end of the gene were best for generating gene knockouts. However, intuitively, the formation of an in-frame variant in an essential domain is far more likely to negatively impact protein function and therefore cell fitness compared to the same perturbation in an inessential region of the protein. The Vakoc group extended this approach to identify drug-resistant mutants by tiling the catalytic domain of

DOT1L, a histone methyltransferase essential for AML growth, and selecting the pool of viable DOT1L mutant AML cells with a DOT1L inhibitor. Cells harboring in-frame drug resistant alleles took over the pool in the presence of the inhibitor. The pool was then sequenced to identify the genotypes of these drug resistant DOT1L variants. In this fashion, on-target inhibition of DOT1L was readily validated through identification of drug-resistance conferring mutations in the catalytic domain, highlighting the utility of this approach. Our lab extended this approach to include sgRNAs tiling the entire protein coding sequence of our target followed by inhibitor treatment to address whether this could be used to uncover small molecule mechanism of action as well as sites of enzyme regulation and allostery.[69]

Chapter 2: CRISPR-suppressor scanning reveals a non-enzymatic role of LSD1 in AML

This chapter is adapted from:

Vinyard, M. E.[*], Su C.[*], Siegenfeld, A. P.[*], **Waterbury, A. L.**[*], Freedy A. M.[*], Gosavi P. M.[*], Park, Y., Kwan, E. E., Senzer B. D., Doench, J. G., Bauer, D. E., Pinello, L., Liau, B. B., CRISPR-suppressor scanning reveals a nonenzymatic role of LSD1 in AML. *Nat Chem Biol* **15**, 529–539 (2019). DOI: 10.1038/s41589-019-0263-0.

Contributions:

M.E.V., C.S., A.M.F., A.L.W., and A.P.S. designed, performed, and analyzed cell and molecular biology experiments. M.E.V and A.M.F. designed, performed, and analyzed CRISPR-Cas9 screens. A.L.W., P.M.G., and B.S. designed, performed, and analyzed protein purification and biochemical assays. A.L.W. and E.E.K. performed protein modeling. A.L.W. and Y.P. designed and synthesized molecules. A.P.S. performed computational analysis and edited the manuscript. J.D. provided technical advice and oversaw library preparation and sequencing of pooled CRISPR-Cas9 screens. D.B. and L.P. provided advice on computational analysis. B.B.L. designed the experimental strategy, performed, and analyzed experiments, performed computational analysis, wrote the manuscript, and held overall responsibility for the study.

2.1. Abstract

Understanding the mechanism of small molecules is a critical challenge in chemical biology and drug discovery. Medicinal chemistry is essential for elucidating drug mechanism, enabling variation of small molecule structure to gain structure-activity relationships (SAR). However, the development of complementary approaches that systematically vary target protein structure could provide equally informative SAR for investigating drug mechanism and protein function. Here we explore the ability of CRISPR-Cas9 mutagenesis to profile the interactions between lysine-specific histone demethylase 1 (LSD1) and chemical inhibitors in the context of acute myeloid leukemia (AML) (**Figure 2.1**). Through this approach, termed CRISPR-suppressor scanning, we elucidate drug mechanism-of-action by showing that LSD1 enzyme activity is not required for AML survival and that LSD1 inhibitors instead function by disrupting interactions between LSD1 and the transcription factor GFI1B on chromatin. Our studies clarify how LSD1 inhibitors mechanistically operate in AML and demonstrate how CRISPR-suppressor scanning can uncover novel aspects of target biology.
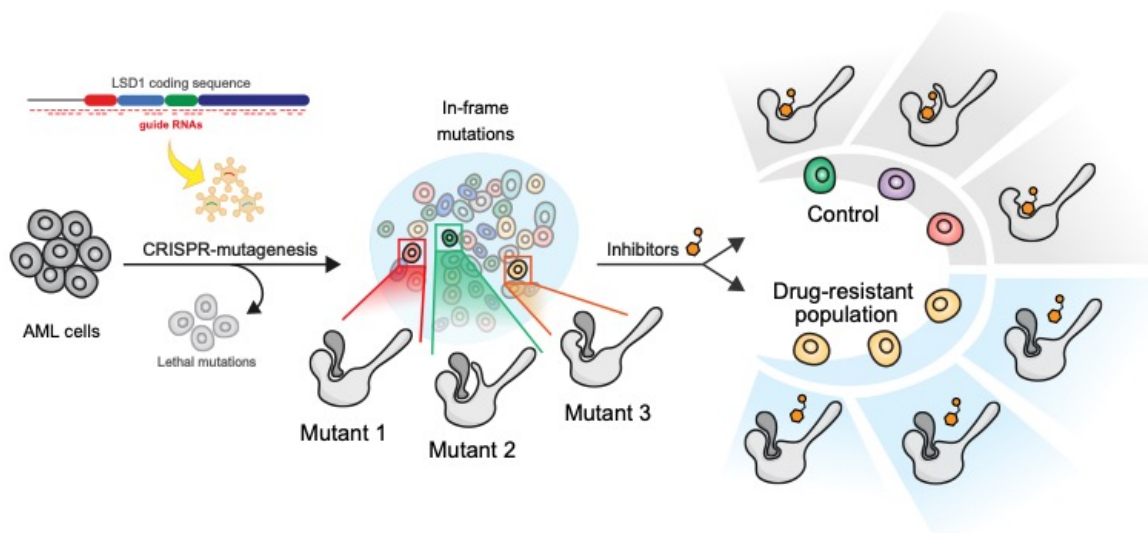
Figure 2.1. Graphical abstract summary of this chapter.

2.2. Introduction

Genome-editing technologies provide new capabilities to systematically alter protein sequence in situ by directly manipulating the endogenous coding sequence (CDS). One such technology, CRISPR-scanning, is a novel method, which has been applied to identify functionally important protein domains.[85,89] In this approach, the programmable nuclease, Cas9, and single-guide RNAs (sgRNAs) spanning the protein of interest's CDS are used to generate cells containing heterogeneous insertion or deletion (indel) mutations formed upon DNA repair. If the protein target is essential for cell growth, cells harboring frameshift indels will drop out from the population due to knockout of the essential target protein. By contrast, cells harboring in-frame indels can often form nearly full-length protein with alterations at the site targeted by a particular sgRNA. In-frame variants in a functional region of an essential protein frequently result in loss-of-function mutations that impair cell fitness. However, in-frame variants might also result in gain-of-function mutations that confer a fitness advantage in the context of a selection pressure, such as a small molecule drug.[86–88] Whether these in-frame drug-suppressor mutations can be employed to reveal precise structural information on a small molecule's binding site, effectively discriminate between structurally-related inhibitors, and clarify a molecule's mechanism-of-action remain to be fully explored.

To examine these possibilities, we explored the ability of CRISPR-scanning to interrogate SAR between LSD1 and its small molecule inhibitors in the context of AML. First identified through its association in histone deacetylase-containing corepressor complexes,[36,90,91] LSD1 is a FAD-dependent histone lysine demethylase, which demethylates monomethyl and dimethyl histone H3 lysine 4 (H3K4).[35,92] Several small molecule LSD1 inhibitors are now in clinical development for the treatment of various cancers, including AML and small cell lung cancer.[58,61,70,93–99] The development of LSD1 inhibitors was motivated by the premise that the demethylase activity of LSD1 is necessary for AML proliferation.[35] However, a recent report suggests that the demethylase activity of LSD1 is not essential in this context.[70] Instead, LSD1 inhibitors may disrupt

LSD1's interactions with GFI1 and/or GFI1B, two highly related transcription factors.[70,94,95,98,99] GFI1/GFI1B bind directly to the LSD1 catalytic site via their N-terminal SNAG domains to recruit LSD1 to their cognate genome binding sites.[44,55,56,70] Consequently, the relative contributions of LSD1's enzymatic and non-enzymatic functions are convoluted, and the precise mechanistic role of LSD1 in AML remains ambiguous.

By combining CRISPR-scanning with multiple small molecule inhibitor treatments, an approach we refer to as CRISPR-suppressor scanning, we systematically identify mutations in LSD1 that confer resistance to LSD1 inhibitors in AML. By comparing the mutations differentially enriched by structurally-related as well as mechanistically distinct compounds, we highlight the ability of this approach to interrogate SAR and essential features of the small molecule binding site. Many of the drug-resistant mutations identified reside in the LSD1 catalytic site and inactivate enzymatic function, indicating that LSD1 demethylase activity is not required for AML survival. By rescuing growth using a drug-resistant GFI1B allele that only binds to LSD1 in the presence of drug, we show that drug-mediated disruption of a LSD1-GFI1B complex is sufficient to block AML proliferation. Moreover, by comparing the effects of LSD1 inhibitors on wild type (wt) versus mutant AML cells containing enzymatically inactive LSD1, we distinguish the relative roles of LSD1 enzymatic versus non-enzymatic function in controlling gene expression. We demonstrate that drug-induced disruption of the LSD1-GFI1B complex on chromatin results in the activation of GFI1B-occupied enhancers by potentiating PU.1 activity. Our studies showcase the ability of CRISPR-mutagenesis to both investigate small molecule SAR and provide critical information on the molecular mechanisms of target biology.

2.3. CRISPR-scanning identifies functional regions of LSD1

To identify regions within LSD1 that modulate sensitivity to LSD1 inhibitors in AML, we conducted SpCas9 CRISPR-scanning of *KDM1A,* which encodes LSD1, using a pool of all possible 360 NGG PAM-restricted sgRNAs that tile the *LSD1* CDS (**Figure 2.2**) and 81 control sgRNAs (**Supplementary Dataset 1**).[86]
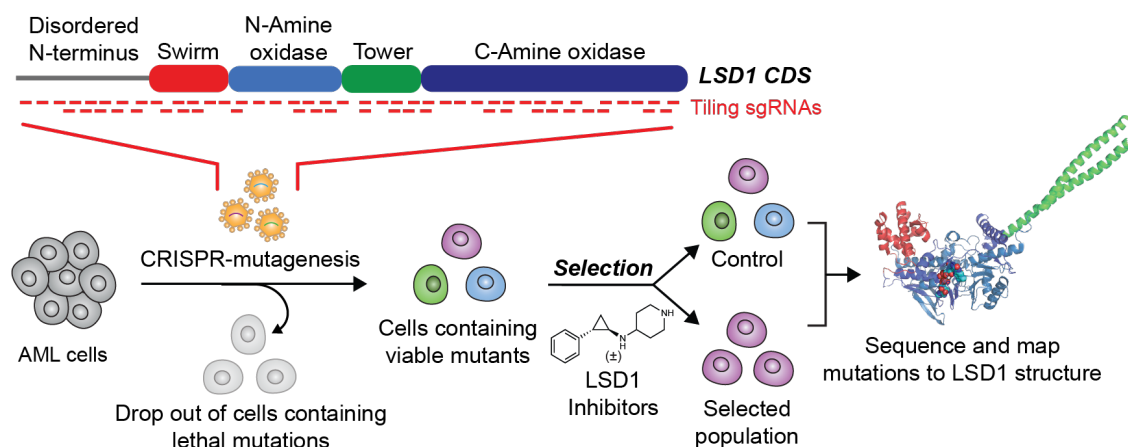
Figure 2.2. Schematic of CRISPR-suppressor scanning workflow to profile SARs of LSD1 small-molecule inhibitors

The pool of sgRNAs was transduced into SET-2 and MV4;11, two AML cell lines highly sensitive to LSD1 knockout and to treatment with LSD1 inhibitors, including GSK-LSD1 (**1**) — a derivative of tranylcypromine (TCP) that forms a covalent bond with LSD1's FAD cofactor through a suicide-inhibition mechanism (**Figure 2.3**).[97,100]
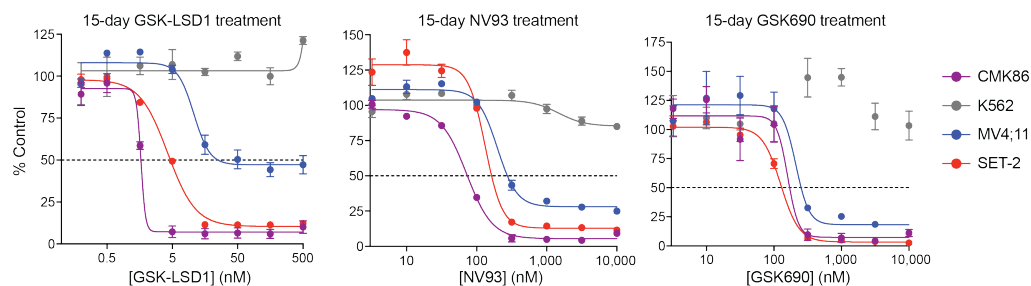
Figure 2.3. Proliferation of AML cell lines under LSD1 inhibitor treatment.

Dose-response curves for cell lines treated with GSK-LSD1, NV93, and GSK690 are shown. Data represent mean values ± s.e. across three replicates. One of two independent replicates is shown.

After transduction, the cells were split and then treated with vehicle (DMSO) or GSK-LSD1 (100 nM). Genomic DNA was isolated from the surviving cell populations at multiple timepoints and sequenced to deconvolute sgRNA identities enriched under each condition. Given the essential nature of LSD1 in these cell lines, most sgRNAs targeting *LSD1* were depleted under both conditions as compared to the functionally neutral genome-targeting control sgRNAs (**Figure 2.4, Figure 2.5**). However, several sgRNAs were enriched after prolonged GSK-LSD1-treatment, leading to diverging sgRNA enrichment profiles between vehicle and GSK-LSD1 conditions over time (**Figure 2.4, Figure 2.5**). These observations are consistent with the emergence and expansion of drug-resistant populations.
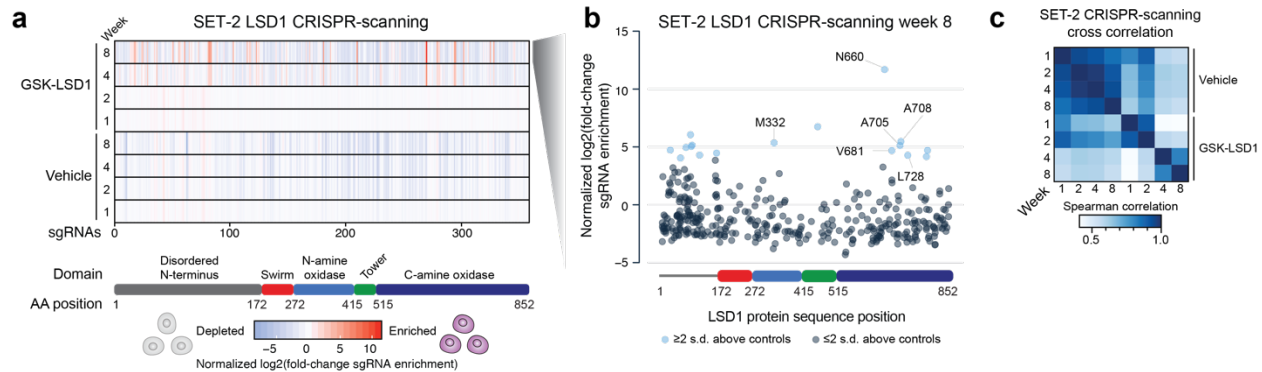
Figure 2.4. CRISPR-suppressor scanning identifies regions of LSD1 that mediate its function and susceptibility to pharmacological inhibitors.

a, Heatmaps depicting $\log_2$(fold-change) sgRNA enrichment in SET-2 at the indicated time points and conditions versus day 0 normalized against functionally neutral genome-targeting control sgRNAs. The sgRNAs are arrayed on the x axis by the *LSD1* CDS. Color represents mean values across three replicate transductions. **c,** Scatterplot showing $\log_2$(fold-change) sgRNA enrichment in SET-2 under GSK-LSD1 treatment at week 8 versus day 0 normalized against functionally neutral genome-targeting control sgRNAs. The sgRNAs are arrayed on the x axis by the *LSD1* CDS. Data represents mean values across three replicate transductions. **d,** Heatmap showing cross correlation of overall sgRNA enrichment over time points during the CRISPR-scanning screen in SET-2. Data represent mean values across three replicate transductions.
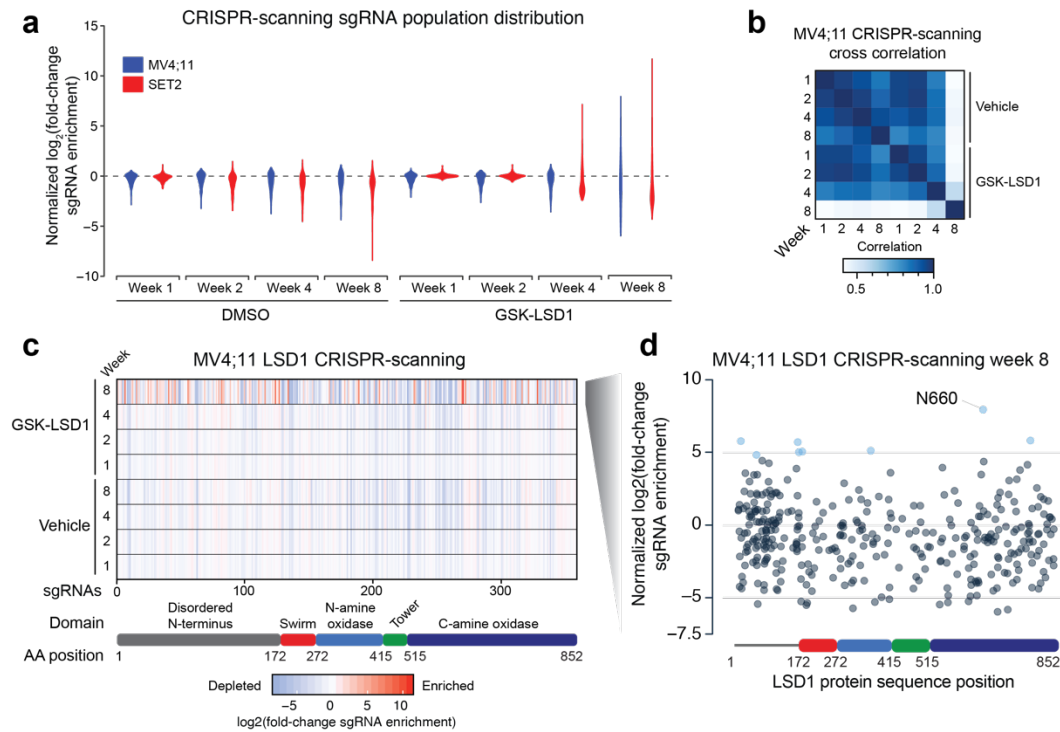
Figure 2.5. LSD1 CRISPR-suppressor scanning in MV4;11

a, Violin plots showing log2(fold-change sgRNA enrichment) in SET-2 and MV4;11 versus week 0 normalized against functionally neutral genome-targeting control sgRNAs. **b,** Heat map showing cross correlation of overall sgRNA enrichment at different time points during the CRISPR- suppressor scanning screen in MV4;11. Data represent mean values across three replicate transductions. **c,** Heat map depicting log2(fold-change sgRNA enrichment) in MV4;11 at the time points specified versus week 0 normalized against functionally neutral genome-targeting control sgRNAs. The sgRNAs are arrayed on the *x* axis by the *LSD1* CDS. Color represents mean values across three replicate transductions. **d,** Scatter plot showing sgRNA enrichment in MV4;11 for GSK-LSD1 treatment at week 8 versus week 0 normalized against functionally neutral genome-targeting control sgRNAs. Data represents mean values across three replicate transductions.

Depleted and enriched sgRNAs were asymmetrically spread across the different protein domains of the *LSD1* CDS (**Figure 2.4a,b**). Consequently, it was not immediately apparent whether these sgRNAs clustered in particular regions of LSD1 by only considering the linear

CDS.[85] To identify potential regions of LSD1 critical for AML fitness, we considered whether areas of LSD1 targeted by sgRNAs, enriched either through negative or positive selection, may be spatially clustered in 3D space.

To assess this phenomenon, we adapted the clustering of mutations in protein structure (CLUMPS) algorithm to analyze sgRNA enrichment.[101] First, each sgRNA was assigned to the closest amino acid (AA) residue overlapping the predicted Cas9 cleavage site. Then, all pairwise combinations of sgRNAs were scored by (1) their combined enrichment and (2) the Euclidean distance between their assigned amino acid residues based on reported LSD1 structural data (**see Methods**) (**Figure 2.6**).[56] To identify hotspots, the proximity weighted enrichment score (PWES) matrix of all pairwise sgRNA interactions was grouped by hierarchical clustering, leading to the identification of 12 clusters of proximal sgRNAs that define distinct regions of LSD1 characterized by similar enrichment profiles (**Figure 2.6b**).
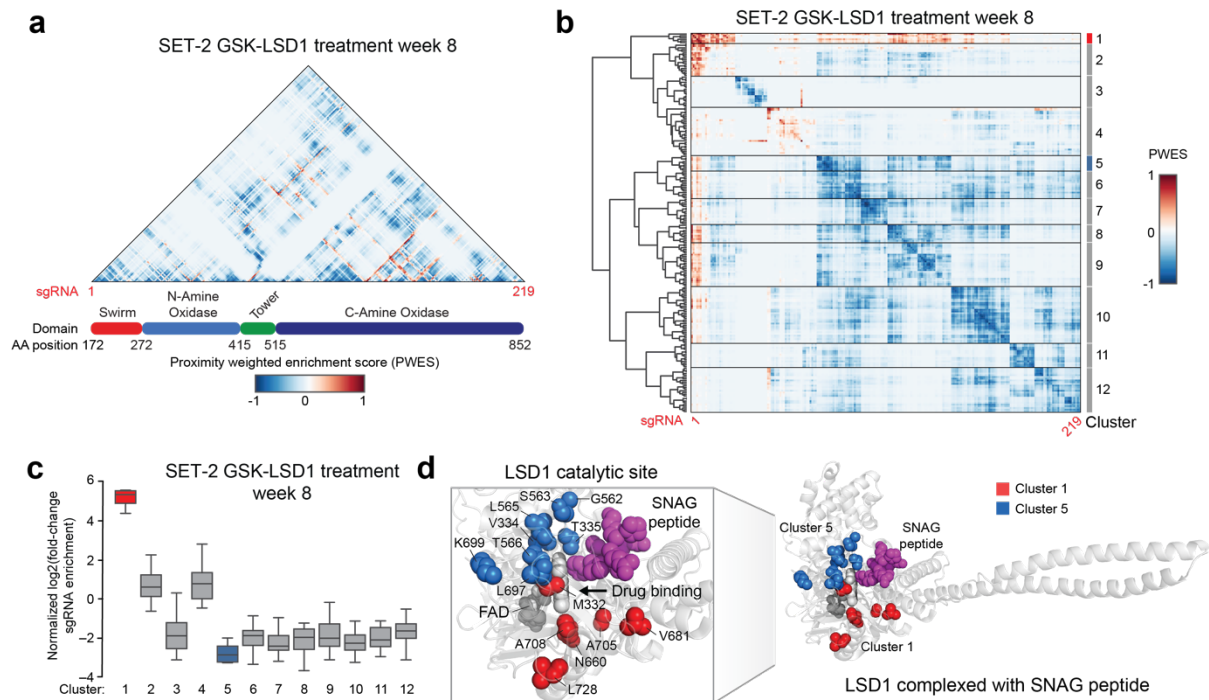
Figure 2.6. Spatial clustering of CRISPR-suppressor scanning data reveals potential functional hotspots of LSD1 that mediate drug action.

Figure 2.6. (continued) a,b, Heat maps depicting the PWES between sgRNAs ordered by the CDS (**a**) or by hierarchical clustering (**b**). sgRNAs targeting the disordered LSD1 N terminus were omitted in this analysis because this domain was truncated in structural studies. **c**, Box plot showing the enrichment of sgRNA groups as defined by hierarchical clustering in **b** for SET-2 under GSK-LSD1 treatment at week 8. The y axis represents mean $\log_2$(fold-change sgRNA enrichment) at week 8 versus week 0 across three replicate transductions. In the plot, bars represent the median, the box represents the interquartile range (IQR) and the whiskers represent 1.5 × IQR. **d**, Structural view of LSD1 complexed with a SNAG peptide (magenta) showing the location of clusters 1 (red) and 5 (blue) defined in **b** (Protein Data Bank (PDB): 2Y48).

Consistent with our prior observations, most clusters identified corresponded to sgRNAs depleted during the selection (**Figure 2.6b,c, Figure 2.7a-c**). By contrast, cluster 1 stood out as the most highly enriched cluster in the presence of GSK-LSD1 (**Figure 2.6c**). When mapped to the LSD1 structure, this enriched cluster was located within the catalytic site (**Figure 2.6d**), which is also the binding site of GSK-LSD1. To test whether clustering of enriched sgRNAs is statistically significant, a summed PWES was calculated for sgRNAs in a particular group, and this value was compared to the simulated distribution of summed PWES determined by randomizing the positions of these sgRNAs throughout the protein (**see Methods**) (**Figure 2.7d**). Comparing the summed PWES for sgRNAs in cluster 1 to the simulated distribution indicated that the spatial clustering of these sgRNAs was significant ($p = 2 \times 10^{-5}$, **Figure 2.7d**). Notably, the targeted amino acids comprising cluster 1 (M332, N660, V681, A705, A708, L728) are close to the FAD cofactor, which is bound by GSK-LSD1, but more distal to the substrate binding site (**Figure 2.6d**, SNAG peptide).[56] This suggests that drug-resistance mutations are selected such that they potentially alter drug binding while permitting essential binding interactions of LSD1 to its substrate(s), which may be highly important for AML proliferation.[85] In agreement with this notion, we observed that the most highly depleted cluster, cluster 5, comprises sgRNAs targeting residues involved in

substrate binding (**Figure 2.6c,d**).[48,56] Similarly to cluster 1, the spatial clustering of sgRNAs for cluster 5 was found to be statistically significant ($p < 10^{-5}$, **Figure 2.7d**). Altogether, our analysis demonstrates that CRISPR-scanning data can be used to identify spatially clustered hotspots in protein targets that may modulate inhibitor mechanism-of-action and protein function.
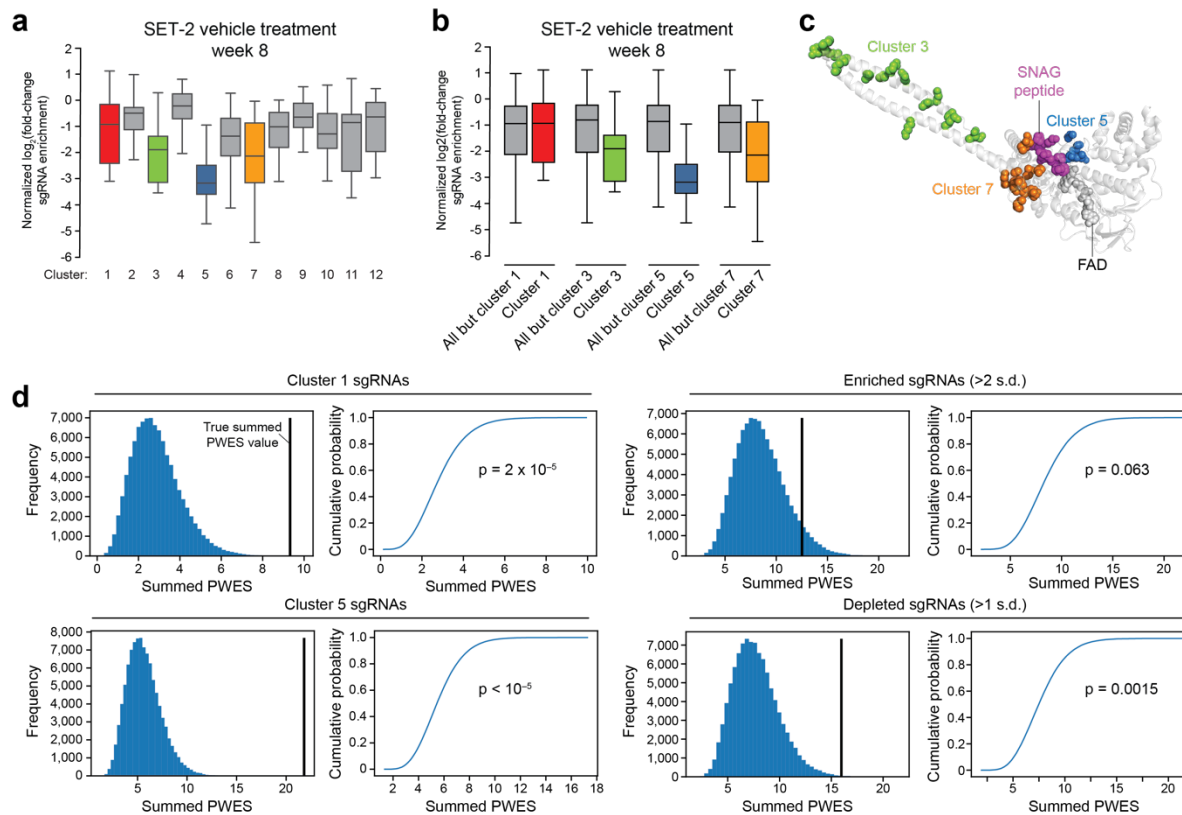
Figure 2.7. CRISPR-suppressor scanning reveals functionally important regions of LSD1 through negatively selected clusters

a-b, Box plot showing normalized log2(fold-change sgRNA enrichment) at week 8 versus week 0 for clusters identified by hierarchical clustering in SET-2 under vehicle treatment. In the plot, bars represent the median, the box represents IQR, and the whiskers represent 1.5 x IQR across three replicate transductions. **c,** Structural view of LSD1 depicting the location of clusters 3, 5, and 7 (PDB: 2Y48). **d,** Histograms and corresponding cumulative distribution functions (CDF) showing the simulated summed PWES values for the following categories of sgRNAs: cluster 1, cluster 5, sgRNAs with enrichment values >2 s.d. above the mean, and sgRNAs with enrichment values >1 s.d. below the mean. The empirical *P* value, determined to be 1 – CDF(True summed PWES value), is displayed on the graph.

2.4. CRISPR-scanning reveals structure-activity relationships

We next considered whether CRISPR-scanning could discriminate between different classes of LSD1 inhibitors as well as structurally-related analogs by investigating how variations of the small molecule inhibitor might change the overall sgRNA enrichment profile and resulting LSD1 mutation signature. We began by synthesizing 5 structural analogs of GSK-LSD1 (AW1 (**2**), AW2 (**3**), AW3 (**4**), AW4 (**5**), and AW (**6**)), each possessing a different substituent on the phenyl group of GSK-LSD1 (**Figure 2.8a**). We measured the in vitro and cellular activities of these GSK-LSD1 analogs along with two previously described reversible LSD1 inhibitors, GSK690 (**7**) and NV93 (**8**) (**Figure 2.8a-d**).[102,103]
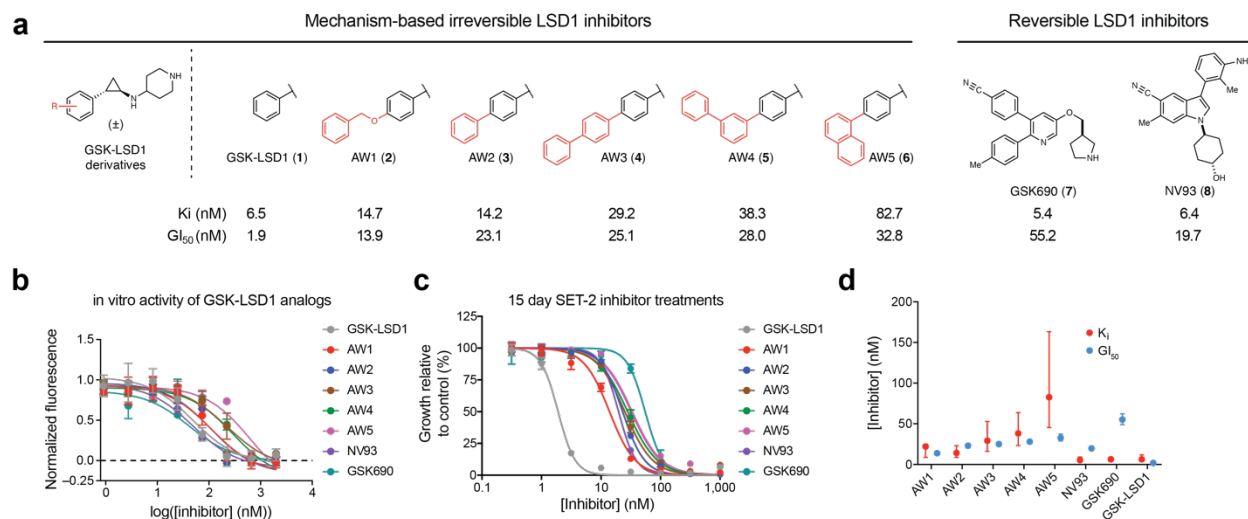
Figure 2.8. LSD1 inhibitors for CRISPR-suppressor scanning

a, Structures of LSD1 inhibitors used in CRISPR-suppressor scanning. For each compound, K_i and GI_{50} values were measured across three replicates, where K_i is the equilibrium dissociation constant and GI_{50} is the drug concentration that results in half-maximal inhibition of cell proliferation. **b,** Dose-response curves for in vitro LSD1 inhibition by GSK-LSD1, GSK690, NV93, and AW1-AW5 are shown. Data represent mean values ± s.e. across three technical replicates. One of two independent replicates is shown. **c,** Dose-response curves for SET-2 cell lines treated with GSK-LSD1, GSK690, NV93, and AW1-AW5 are shown. Data represent mean values ± s.e. across three technical replicates. One of two independent replicates is shown. **d,** Scatter plot showing the K_i and GI50 values for each inhibitor shown in (**a**). Data represents

We subsequently subjected the CRISPR-scanning transduced pool of SET-2 cells to these compounds dosed at 500 nM. We sequenced genomic DNA to determine sgRNA enrichment under each drug treatment. Focusing the analysis on positively enriched sgRNAs, we normalized the enrichment scores of the inhibitor-treated samples to those of the vehicle-treated samples (**Figure 2.9a**).
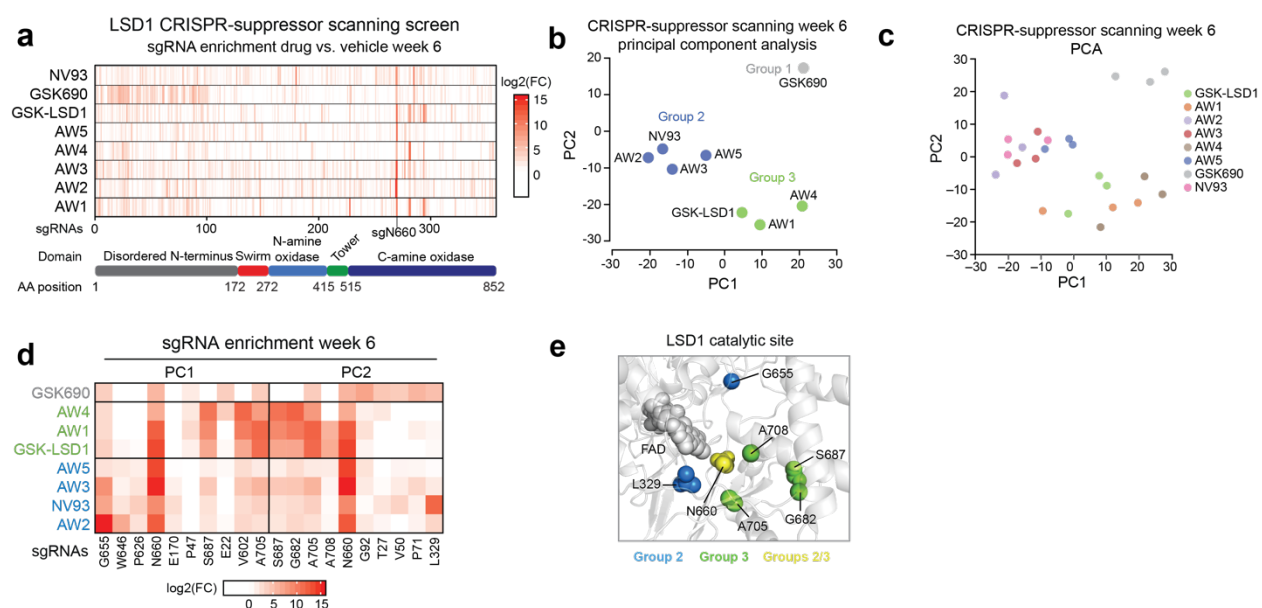
Figure 2.9. CRISPR-suppressor scanning enables profiling of LSD1 SARs

a, Heat maps showing sgRNA enrichment measured in SET-2 at week 6. The sgRNAs are arrayed on the *x* axis by the LSD1 CDS. Color represents mean log2(fold-change sgRNA enrichment normalized to vehicle treatment) across three replicate transductions. FC = fold-change. **b,** Scatter plot showing projection of compound sgRNA enrichment profiles at week 6 onto principal component space. Groups of compounds were identified by *k*-means clustering (*k* = 3) and are marked. Data represent mean values across three replicate transductions. **c,** Scatter plot showing projection of compound sgRNA enrichment profiles at week 6 onto PC space where every replicate transduction (n = 3) is considered separately. **d,** Heat map showing sgRNA enrichment measured in SET-2 at week 6 for the top five sgRNAs with largest principal component

44

Figure 2.9 (continued) coefficient loadings and bottom five sgRNAs with smallest principal component coefficient loadings. The sgRNAs are ordered on the *x* axis by principal component coefficient, increasing from left to right, and are labeled according to the residues they target in the *LSD1* CDS. Color represents mean $\log_2$(fold-change) of sgRNA enrichment normalized to vehicle treatment across three replicate transductions. **e,** Structural view depicting the LSD1 active site where residues targeted by sgRNAs enriched by compounds in Group 2 and Group 3. sgN660 was enriched by all compounds.

To investigate if sgRNAs were differentially enriched depending on the inhibitor treatment, we performed principal component analysis (PCA) on the various drug treatments. Clustering the transformed data suggests that sgRNA enrichment broadly classifies the compounds into three groups (**Figure 2.9b,** Group 1: GSK690; Group 2: AW2, AW3, AW5, NV93; Group 3: GSK-LSD1, AW1, AW4). This compound classification was mostly maintained when considering each treatment replicate separately, suggesting that the groupings are not driven by technical noise between replicates (**Figure 2.9c**). GSK690, which led to the weakest enrichment of active site sgRNAs and growth inhibition, was distinctly separated from the other compounds by PCA (**Figure 2.9b,d**). GSK-LSD1 analogs in Group 2 were distinguished by strong enrichment of sgRNAs targeting the neighboring residues N660 (sgN660) and G655 (sgG655) (**Figure 2.9d**); NV93 was also marked by enrichment of sgL329 (**Figure 2.9d**). Notably, these sgRNAs target residues directly adjacent to the FAD cofactor (**Figure 2.9e**). While compounds in Group 3 were also characterized by strong enrichment of sgN660 (**Figure 2.9e**), sgRNAs targeting peripheral residues within the C-terminal amine oxidase domain were also enriched (**Figure 2.9e**, sgA705, sgA708, sg682, sgS687). These results suggest that CRISPR-suppressor scanning can discriminate between structurally similar and diverse inhibitors.
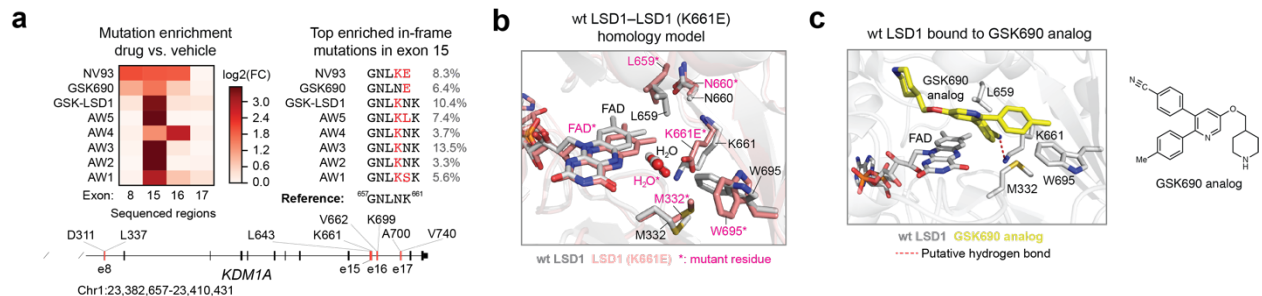
Figure 2.10. Exon 15 mutations enriched by GSK690 and NV93

a, Left, heat map showing mutation enrichment in sequenced exons across different drug treatment conditions. Mutated reads are defined as any read containing a modification in a coding region. Color represents log2(fold-change mutated reads) of drug-treatment relative to vehicle treatment at week 6. DNA from three replicate transductions were pooled at equal concentrations before sequencing to obtain an average log2(fold-change) of mutated reads. Right, schematic shows the genotypes of the most abundant mutation in exon 15 for each treatment condition at week 6. Percentages indicate the allele frequency as determined through sequencing. Bottom, schematic depicts the *LSD1* gene locus and the regions sequenced. **b**, Homology model of LSD1 (K661E) overlaid with wild-type LSD1 is shown. W695, involved in a hydrophobic interaction with GSK690, is rotated out of plane in the model of the mutant structure (PDB: 2HKO). **c,** Structural view depicting the wild-type LSD1 active site bound to a GSK690 analog, which is shown due to the lack of structural data of GSK690 or NV93 bound to LSD1. The hydrogen bond between K661, a key residue involved in the catalytic activity of LSD1, and the 4-cyanophenyl ring of the GSK690 analog would be disrupted upon mutation to K661E (PDB: 5YJB).

We next considered whether the underlying mutations, resulting from selection by each inhibitor treatment, could afford structure-function information on the LSD1 pocket (**Figure 2.2**). To investigate this possibility, we sequenced 4 exons of *LSD1* encompassing the most highly enriched sgRNAs, including many identified through PCA (**Figure 2.10a**).[104] The majority of mutations were observed in exon 15 surrounding the most highly enriched sgRNA, sgN660 (**Figure 2.10a**). Examination of the mutations in exon 15 revealed that GSK690 and NV93 selected LSD1 mutants containing K661E point substitutions. Modeling the K661E mutation onto

a LSD1 structure suggests that the mutation rotates W695 and removes a hydrogen bond between K661 and the drug's nitrile group (**Figure 2.10b,c**), perturbing interactions known to be essential for GSK690 and likely, NV93 binding.[68] By contrast, treatment with GSK-LSD1 analogs led to similar mutations that insert a K residue after LSD1 L659 (**Figure 2.10a**). Modeling the L659_N660insK mutation onto the LSD1 structure suggests that the inserted K projects into the binding pocket (**Figure 2.11a-c**), potentially altering a hydrogen bond network involving K661, a water molecule, and the FAD cofactor, which is necessary for enzyme activity.[105]
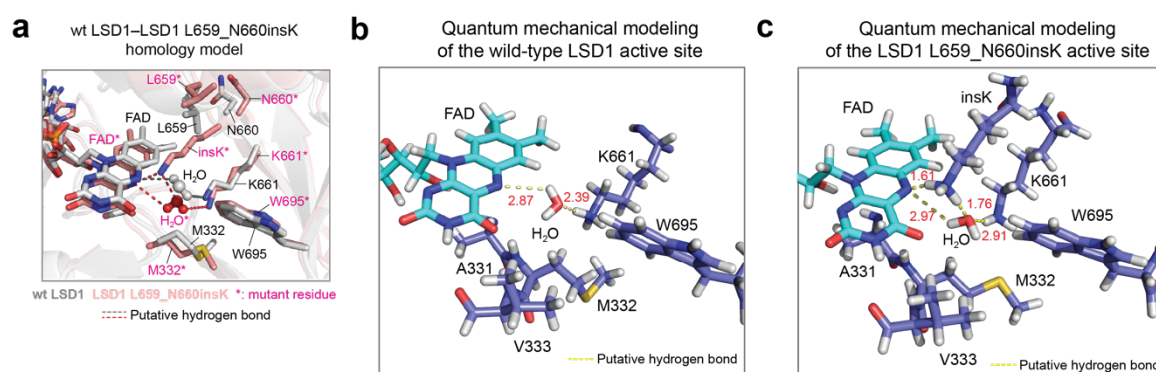
Figure 2.11. Exon 15 mutations enriched by GSK-LSD1 analogs

a, Homology model of LSD1 L659_N660insK overlaid with wild-type LSD1 is shown. The N5 atom of FAD and K661 are involved in a water-mediated hydrogen-bonding network (PDB: 2HKO). **b,c,** Semi-empirical quantum mechanical modeling of wild-type LSD1 and LSD1 L659_N660insK active sites are shown. The insK, through competing interactions, disrupts the water-mediated hydrogen bonding network between the N5 atom of FAD and K661 in LSD1 L659_N660insK.

Aside from mutations in exon 15, different drugs selected heterogeneous mutations in exon 8 and exon 16 to varying extents (**Figure 2.10a, Figure 2.12a-d**). Notably, deletion mutations in exon 16 were selectively enriched by AW4 (**Figure 2.12a,b**), causing the mutation profile of AW4 to be substantially different from those of other GSK-LSD1 analogs (in particular its isomer, AW3) (**Figure 2.10a**). Altogether, these data support the notion that CRISPR-suppressor scanning can effectively discriminate between mechanistically distinct inhibitors and

even more closely related chemical analogs, at the level of both sgRNA enrichment and protein coding mutations.
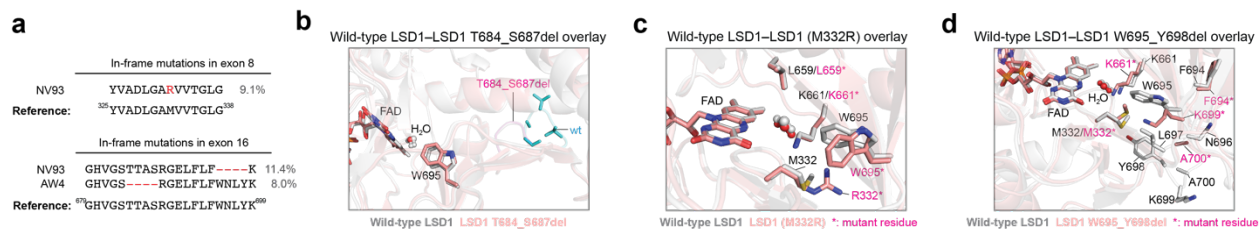
Figure 2.12. Exon 8 and exon 16 mutations enriched by different LSD1 inhibitors

a, Schematic showing the genotypes of the most abundant mutations identified in (top) exon 8 for NV93 treatment and (bottom) exon 16 for NV93 and AW4 treatment at week 6. Percentages indicate the allele frequency. **b,** Homology model of LSD1 T684_S687del overlaid with wild-type LSD1 is shown to highlight the deletion of the wild- type loop compared to the truncated mutant loop (PDB: 2HKO). **c,** Homology model of LSD1 (M332R) overlaid with wild-type LSD1 suggests that M332R may remove a potential hydrophobic interaction between M332 and LSD1 inhibitors (PDB: 2HKO). **d,** Homology model of LSD1 W695_Y698del overlaid with wild-type LSD1 is shown. This deletion removes W695, a key residue in the hydrophobic substrate-binding cavity involved in GSK690 analog binding (PDB: 2HKO).

2.5. Drug-resistance LSD1 mutants are enzymatically active

While Many of the mutations identified by CRISPR-suppressor scanning are predicted to potentially disrupt LSD1 enzyme activity, prompting us to investigate if LSD1 enzyme activity is necessary for AML proliferation. To characterize these in-frame LSD1 mutations and the subsequent resistance mechanism(s) in an isolated setting, the top-enriched sgRNAs, sgN660 and sgG655, were transduced individually into AML cells, which were then treated with vehicle or GSK-LSD1. In the presence of GSK-LSD1, AML cells transduced with the sgRNAs (GFP$^+$ cells) gained a competitive growth advantage versus non-transduced cells (GFP$^-$ cells), consistent with the emergence of drug-resistance observed in the pooled screens (**Figure 2.13a,b**).
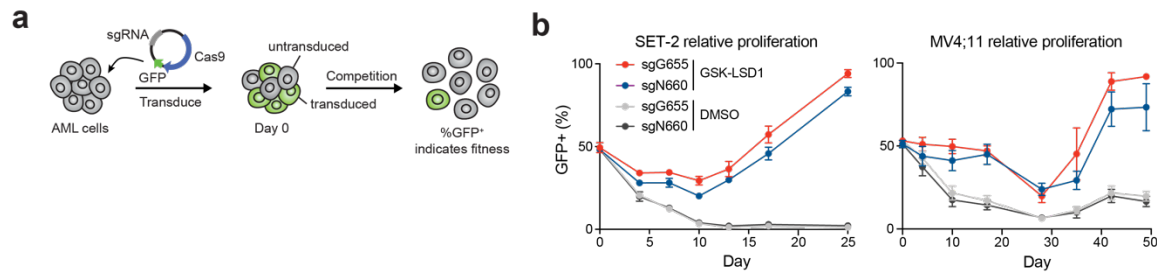
Figure 2.13. Single guide validation of LSD1 mutants

a, Illustration depicting the workflow for sgRNA validation through a competitive growth assay monitored by GFP fluorescence signal. **b,** Line graph showing the fraction of GFP+ cells (y axis) over a time course following lentiviral transduction in SET-2 and MV4;11 cells. Under GSK-LSD1 treatment, mutagenesis by sgN660 leads to enrichment of GFP+ cells. Data represent mean ± s.d. across three technical replicates. One of two independent replicates is shown.

We next expanded and genotyped clonal drug-resistant SET-2 cell lines. The mutations identified were largely similar to those obtained in the CRISPR-suppressor scanning screen, including LSD1 N660_delinsKL and LSD1 L659_N660insR, abbreviated delinsKL and insR, respectively (**Figure 2.14a-c**). In-frame mutations were often paired with frameshift mutations in the drug-resistant clones at equal frequencies (**Figure 2.14b**),[104] suggesting that the clones express a single in-frame LSD1 variant.
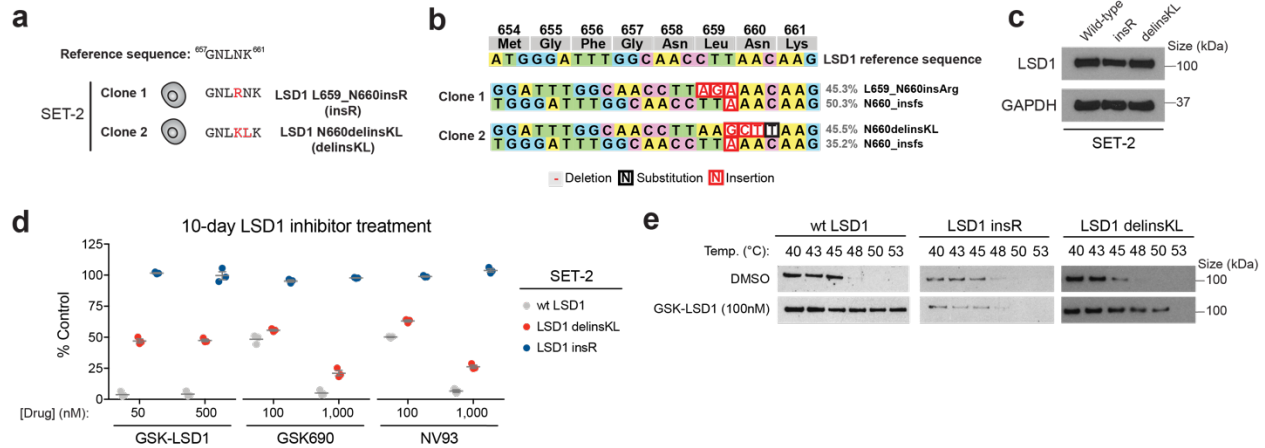
Figure 2.14. Characterization of wild-type and mutant LSD1 variants

a, Schematic indicating the genotypes of two single-cell derived, clonally-expanded SET-2 cell lines containing drug- resistant LSD1 mutations. **b,** Schematic depicting the coding mutations identified in the SET-2 drug-resistant cell lines. **c,** Immunoblots showing that LSD1 mutants are expressed in drug-resistant cell lines. **d,** Scatter plots for relative cell growth (y axis) of SET-2 cells treated with GSK-LSD1, GSK690 and NV93 are shown. Lines represent mean ± s.e.m. across three replicates. One of two independent replicates is shown. **e,** Immunoblots showing levels of LSD1 protein after heat treatment of SET-2 cells at the different temperatures indicated during the CETSA protocol. One of two independent replicates is shown.

Notably, whereas SET-2 LSD1 insR displayed complete resistance to all LSD1 inhibitors tested at the doses employed, LSD1 delinsKL only displayed resistance to GSK-LSD1 (**Figure 2.14d**). Using cellular thermal shift assays (CETSA), we found that treatment with GSK-LSD1 led to LSD1 stabilization in wt SET-2 but not SET-2 LSD1 insR, suggesting that GSK-LSD1 binding to the insR LSD1 mutant is altered (**Figure 2.14e**). By contrast, CETSA indicated that there is partial stabilization of LSD1 delinsKL in the presence of GSK-LSD1 relative to the vehicle control, suggesting that GSK-LSD1 may still bind this specific mutant at some level (**Figure 2.14e**). This result is consistent with the partial drug-resistance phenotype of SET-2 LSD1 delinsKL (**Figure**

2.14d). These observations support the premise that CRISPR-suppressor scanning can generate protein variants that can resolve mechanistically different classes of inhibitors.
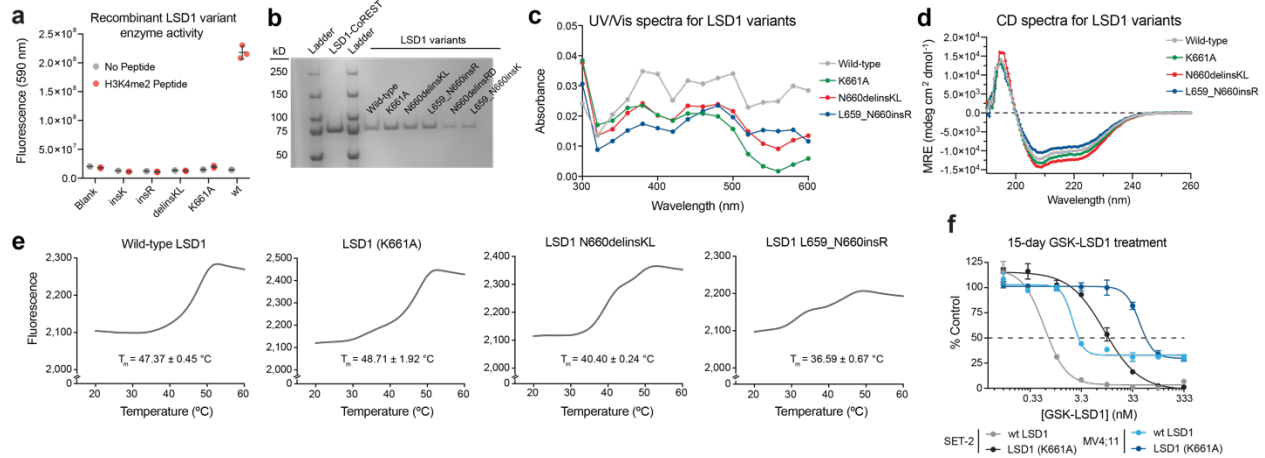
Figure 2.15. Biochemical characterization of wild-type and mutant LSD1 variants

a, Scatter plot showing enzyme activities (*y* axis) of mutant LSD1 variants identified (*x* axis) on a H3K4me2 peptide substrate. Gray lines represent mean ± s.d. across three replicates. One of two independent replicates is shown. **b,** SDS-PAGE confirming successful protein expression at the expected molecular weight of wild-type LSD1, LSD1 (K661A), and identified mutant LSD1 variants. One of two independent replicates is shown. **c,** UV-vis spectra for wild-type LSD1, LSD1 (K661A), LSD1 delinsKL, and LSD1 insR are shown. Data represents a single measurement. One of two independent replicates is shown. **d,** Circular dichroism spectra for wild-type LSD1, LSD1 (K661A), LSD1 delinsKL, and LSD1 insR are shown. Data represents mean value across five individual scans. Error bars are omitted for clarity. Experiment performed once. **e,** Thermal stability curves showing relative fluorescence of FAD (y axis) in wild-type LSD1, LSD1 (K661A), LSD1 delinsKL, and LSD1 insR as a function of temperature (x axis). Data represent mean values across five replicates. Error bars are omitted for clarity. Calculated FAD release temperatures (Tm) are shown, where error represents s.e. across five replicates. One of two independent replicates is shown. **f,** Dose–response curves for GSK-LSD1 tested in SET-2 and MV4;11 cells overexpressing either wild-type LSD1 or LSD1(K661A) are shown. Data represent mean ± s.e.m. across three technical replicates. One of two independent replicates is shown.

Due to the mechanism-based binding of GSK-LSD1 to the LSD1 FAD cofactor, we considered whether the LSD1 mutations may potentially alter GSK-LSD1 binding by compromising enzyme activity. This notion is supported by the frequent occurrence of mutations identified by CRISPR-suppressor scanning that potentially perturb the FAD cofactor binding pocket necessary for LSD1 activity.[105] Indeed, testing recombinant LSD1 mutant proteins indicated that these mutations lead to undetectable demethylation activity against an H3K4me2 peptide (**Figure 2.15a-d**). Additionally, we profiled these LSD1 variants using *Thermo*FAD, an approach that measures the shift in fluorescence of FAD as it is released from proteins upon heating.[106] These experiments revealed that LSD1 insR and LSD1 delinsKL release FAD at lower temperatures in comparison to wt LSD1 (**Figure 2.15e**), supporting the notion that these mutations potentially alter the binding of the FAD cofactor and consequently enzymatic activity. Furthermore, overexpression of enzyme-inactive LSD1 (K661A), but not wt LSD1, confers resistance to GSK-LSD1 (**Figure 2.15f**).[70] Altogether, these observations suggest that perturbing interactions with the FAD cofactor, even at the expense of destroying enzyme activity, is a viable strategy to promote drug resistance and that the enzyme activity of LSD1 is not required for AML proliferation.

2.6. The LSD1-GFI1B interaction is sufficient for AML growth

Our CRISPR-suppressor scanning data and the identification of drug-resistant, enzyme-inactive mutants support the notion that the anti-proliferative effects of LSD1 inhibitors in AML are not due to the inhibition of enzyme activity but instead due to the perturbation of another function of the catalytic site. In particular, our CRISPR-scanning data suggest that substrate binding interactions are critical for LSD1 function in AML (**Figure 2.6c,d**, **Figure 2.7a-c**). In agreement with this hypothesis, recent studies have demonstrated that LSD1 inhibitors disrupt LSD1's interaction with GFI1/GFI1B, which is involved in AML proliferation.[70,94,95,98,99] If GFI1/GFI1B are the relevant protein partners, CRISPR-scanning should select mutants that maintain binding with GFI1/GFI1B in the presence of GSK-LSD1. In agreement, co-immunoprecipitation (Co-IP) experiments for GFI1B and LSD1 demonstrated that while the wt LSD1-GFI1B complex is disrupted by GSK-LSD1, the GFI1B complex with LSD1 delinsKL and LSD1 insR mutants were unperturbed by drug treatment (**Figure 2.16a**). The drug-induced disruption of the LSD1-GFI1B interaction for wt LSD1 but not for the drug-resistant LSD1 mutants implicates this interaction in AML survival.
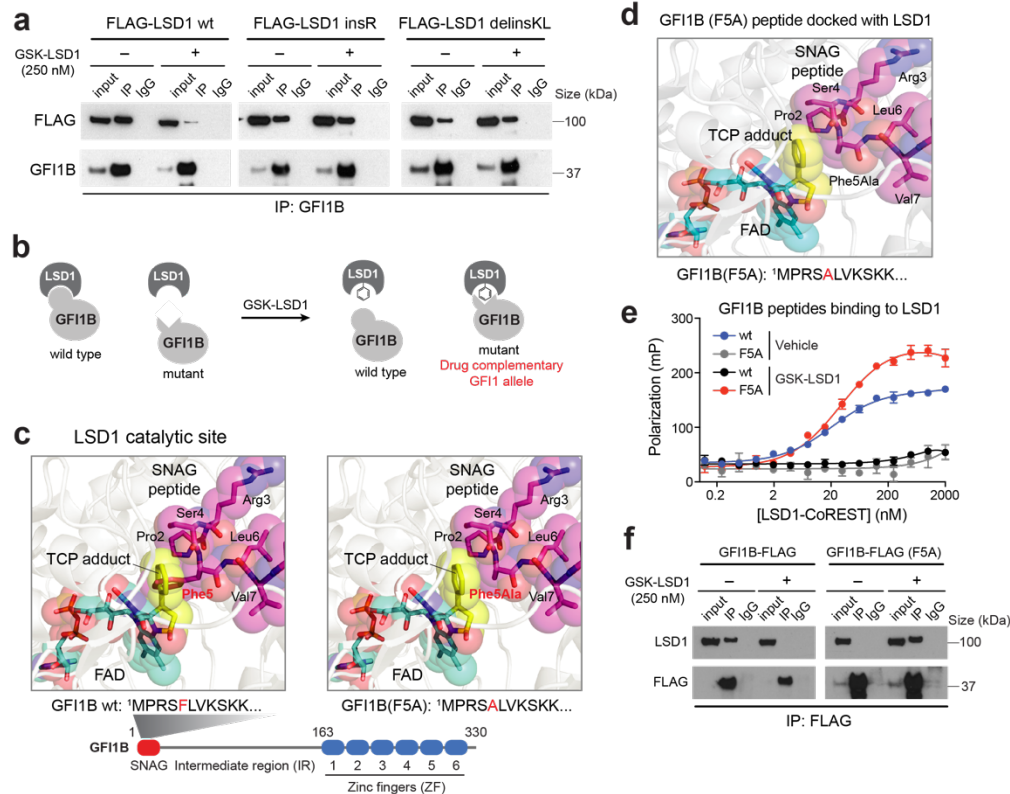
Figure 2.16. An orthogonal drug-complementary GFI1B allele establishes sufficiency of the LSD1–GFI1B interaction for AML survival

a, Co-IP of wild-type FLAG–LSD1 and FLAG–LSD1 mutant variants with GFI1B was performed after vehicle or GSK-LSD1 treatment (250 nM, 48 h) in transiently transfected HEK 293T cells. Co-IP was performed using an anti-GFI1B antibody. LSD1 was detected using an anti-FLAG antibody. One of two independent replicates is shown. **b,** Illustration depicting the concept of an orthogonal drug-complementary GFI1B allele that can only bind LSD1 in the presence of GSK-LSD1. **c,** Structural views showing the LSD1 catalytic site labeled by TCP and bound to either a wild-type SNAG peptide (left) or a modeled SNAG (F5A) peptide (right). Schematic of the protein domains of GFI1B highlighting the SNAG domain sequence (bottom). **d,** Structural view of the LSD1 catalytic site labeled by the GSK-LSD1 adduct modeled with a SNAG (F5A) peptide. **e,** Binding curves showing fluorescence polarization (y axis) for increasing concentrations of LSD1–CoREST complex (x axis) in the presence of either wild-type GFI1B 2–10 peptide or GFI1B (F5A) 2–10 peptide (100 nM) and/or GSK-LSD1 (50 μM). Data represent mean ± s.e.m. across three technical replicates. One of two independent replicates is shown. **f,** Co-IP immunoblots showing levels of wild-type

To demonstrate that the specific LSD1-GFI1B interaction is necessary for AML proliferation, we considered whether an orthogonal GFI1B mutant allele could be identified that binds only to LSD1 covalently modified by GSK-LSD1 but not the unmodified enzyme (**Figure 2.16b**). Based upon protein modeling, we hypothesized that the GFI1B (F5A) mutant could selectively bind to LSD1 covalently modified by GSK-LSD1, as the mutation might restore a hydrophobic interaction between the two proteins (**Figure 2.16 c,d**).[55,56,107] We confirmed that a wt GFI1B peptide only binds purified LSD1, complexed with its binding partner CoREST, in the absence of GSK-LSD1 (**Figure 2.16e**).[107] Conversely, a modified GFI1B (F5A) peptide only binds LSD1-CoREST in the presence of GSK-LSD1 (**Figure 2.16e**). Consistent with this in vitro data, Co-IP experiments demonstrated that overexpressed wt GFI1B-FLAG only interacts with LSD1 in the absence of GSK-LSD1 while overexpressed GFI1B-FLAG (F5A) only interacts with LSD1 in the presence of GSK-LSD1 (**Figure 2.16f**). These data demonstrate that GFI1B (F5A) is an orthogonal drug-complementary allele.[108]
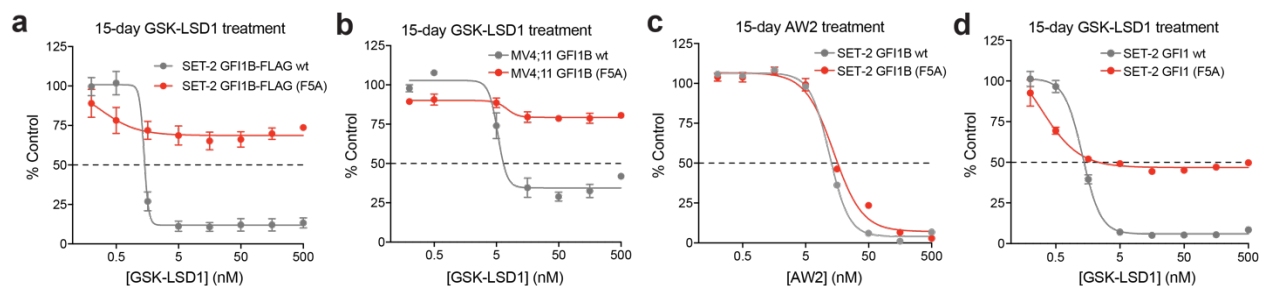
Figure 2.17. GFI1/GFI1B (F5A) is a GSK-LSD1 orthogonal drug-complementary allele

a, Dose–response curves for SET-2 cells expressing wild-type GFI1B–FLAG or GFI1B–FLAG (F5A) treated with GSK-LSD1 are shown. Expression of GFI1B–FLAG (F5A) compared with wild-type GFI1B–FLAG significantly rescues growth of SET-2 cells. Data represent mean ± s.e.m. across three replicates. One of

Figure 2.17(continued) two independent replicates is shown. **b,** Dose-response curves for MV4;11 cells expressing wild-type GFI1B-FLAG or GFI1B-FLAG (F5A) treated with GSK-LSD1 are shown. Data represent mean values ± s.e. across three technical replicates. One of two independent replicates is shown. **c,** Dose-response curves for SET-2 cells expressing wild-type GFI1B-FLAG or GFI1B-FLAG (F5A) treated with AW2 are shown. Data represent mean values ± s.e. across three technical replicates. One of two independent replicates is shown. **d,** Dose-response curves for SET-2 cells expressing wild-type GFI1-FLAG or GFI1-FLAG (F5A) treated with GSK- LSD1 are shown. Data represent mean values ± s.e. across three technical replicates. One of two independent replicates is shown.

Critically, overexpression of GFI1B-FLAG (F5A) but not wt GFI1B-FLAG rescued the growth of AML cells in the presence of GSK-LSD1 (**Figure 2.17a,b**). By contrast, treatment with the bulkier AW2 analog abrogated the protective effect of GFI1B-FLAG (F5A) on growth (**Figure 2.17c**). Overexpression of GFI1-FLAG (F5A) was also sufficient to rescue growth (**Figure 2.17d**), indicating that GFI1 and GFI1B may be redundant for AML proliferation. Altogether, our data provides compelling evidence that the LSD1-GFI1B interaction, complexed in its native context, is necessary for the growth of AML cells.

2.7. LSD1-GFI1B suppresses enhancer activation

The LSD1-GFI1B interaction may serve to recruit LSD1 and corepressor complexes to GFI1B-target enhancers, thus suppressing their activity.[44,70,109] By disrupting LSD1 recruitment, GSK-LSD1 may induce the activation of GFI1B-target genes and promote differentiation.[70,94] However, these changes in gene expression could be induced by the disruption of a LSD1-GFI1B complex, the inhibition of enzyme activity, or both. Given the importance of the LSD1-GFI1B complex in AML survival irrespective of LSD1 enzymatic activity, we sought to employ the drug-resistant LSD1 mutant cell lines to deconvolute the relative contributions of enzyme inhibition versus complex disruption on influencing gene expression upon LSD1 inhibitor treatment.

We first performed RNA-seq on wt SET-2 and enzyme-inactive SET-2 LSD1 insR treated with vehicle or GSK-LSD1 (250 nM, 48 h), revealing 1,290 differentially expressed genes across all conditions (**Figure 2.18a,b**). Treatment of wt SET-2 with GSK-LSD1 induced substantial changes in gene expression: 244 genes were up-regulated while 119 genes were down-regulated ($|\log_2(\text{fold-change})| > 2$, $p < 0.01$). By contrast, using the same statistical thresholds, no genes were found to be differentially expressed upon treatment of SET-2 LSD1 insR with GSK-LSD1, consistent with its fully drug-resistant phenotype (**Figure 2.18a**). Genes differentially expressed in wt SET-2 upon treatment with GSK-LSD1 were enriched in gene signatures associated with myeloid cells and erythroid cells by gene set enrichment analysis (GSEA) (**Figure 2.18c**),[94,110] in agreement with past studies suggesting that LSD1 inhibition induces AML transdifferentiation.[94] Overall, these data are consistent with our prior observations that enzyme-inactive LSD1 insR suppresses the effects of GSK-LSD1 and indicate that most changes in gene expression, including those in genes involved in AML differentiation, are caused by on-target inhibition of LSD1.
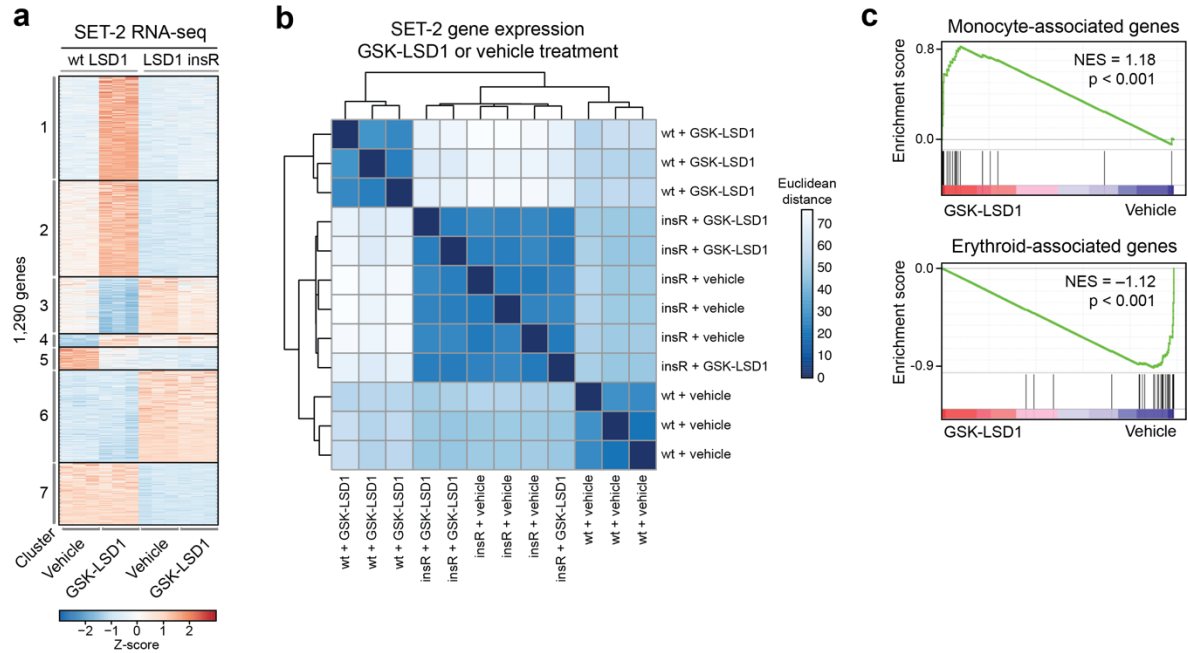
Figure 2.18. Gene expression analysis of wild-type SET-2 and SET-2 LSD1 insR in the presence of GSK-LSD1 or vehicle treatment

a, Heat map showing expression profiles of the 1,290 most variably expressed genes ($|\log_2(\text{fold-change})| > 2$, $P_{adj} < 0.01$) across wild-type SET-2 and SET-2 LSD1 insR after treatment with vehicle or GSK-LSD1 (250 nM, 48 h) for three replicates per condition, each shown independently. The adjusted P value for testing significance of differential expression was calculated using the Benjamini–Hochberg correction. Genes exhibiting correlated patterns of expression were grouped by k-means clustering ($k = 7$). Color represents Z-scores of gene expression across rows. **b,** Heat map showing Euclidean distance between log-transformed gene expression vectors for wild-type (wt) SET-2 and SET-2 LSD1 insR cells treated with vehicle or GSK-LSD1 (250 nM, 48 h). Samples are grouped by hierarchical clustering with all three cell culture replicates depicted separately. **c,** Gene set enrichment analysis (GSEA) showing enrichment of genes expressed in monocytes in genes up-regulated by GSK-LSD1, and enrichment of genes expressed in erythroid cells in genes down-regulated by GSK-LSD1. Normalized Enrichment Scores (NES) and nominal p-value are depicted and calculated as previously described.

We next considered whether we could deconvolute the relative roles of LSD1's enzymatic and non-enzymatic functions in mediating the cells' response to GSK-LSD1. To investigate this

phenomenon, we identified 7 gene sets that exhibit correlated changes in expression using k-means clustering (**Figure 2.18a**). Genes in clusters 1-3 are differentially expressed when wt SET-2 are treated with GSK-LSD1, but otherwise display comparable expression levels across the other samples. Since LSD1 insR is enzyme-inactive yet does not phenocopy the effects of GSK-LSD1 for clusters 1-3, these genes are likely differentially expressed in wt SET-2 due to perturbing a non-enzymatic function of LSD1. By contrast, genes in clusters 4 and 5 are differentially expressed specifically due to loss of LSD1 enzymatic activity because enzyme-inactive LSD1 insR phenocopies the effects of drug-treatment on wt SET-2 gene expression. Finally, genes in clusters 6 and 7 are differentially expressed between the cell lines independent of inhibitor treatment. Notably, there are many more genes in clusters 1-3 than in clusters 4-5, suggesting that inhibition of the non-enzymatic functions of LSD1 are driving the majority of changes in gene expression observed upon acute drug-treatment. Altogether, by comparing the responses of wt SET-2 and SET-2 LSD1 insR to GSK-LSD1, we distinguish between genes that change expression due to the inhibition of LSD1 enzymatic versus non-enzymatic functions.

The observation that SET-2 LSD1 insR suppresses drug-mediated gene expression changes prompted us to investigate if retention of GFI1B DNA-binding underlies these effects. Consequently, we mapped GFI1B, LSD1, H3K4me2, and H3K27ac in wt SET-2 and SET-2 LSD1 insR by chromatin immunoprecipitation followed by sequencing (ChIP-seq) after vehicle or GSK-LSD1 treatment. To restrict our analysis to drug-induced effects in the context of wt AML cells, we focused on ChIP-seq peaks called in wt SET-2 cells. In wt SET-2 cells, we identified 3,391 GFI1B and 3,324 LSD1 high-confidence binding sites, which significantly overlap (424 LSD1-GFI1B co-occupied sites, hypergeometric p-value < 10^{-905}). In agreement with a recent report, GSK-LSD1 treatment led to reduced LSD1 and GFI1B binding in comparison to vehicle treatment (**Figure 2.19a-c**).[70] While the reduction of LSD1 binding was mostly confined to LSD1 binding sites co-occupied by GFI1B, reduction of GFI1B binding occurred more indiscriminately.
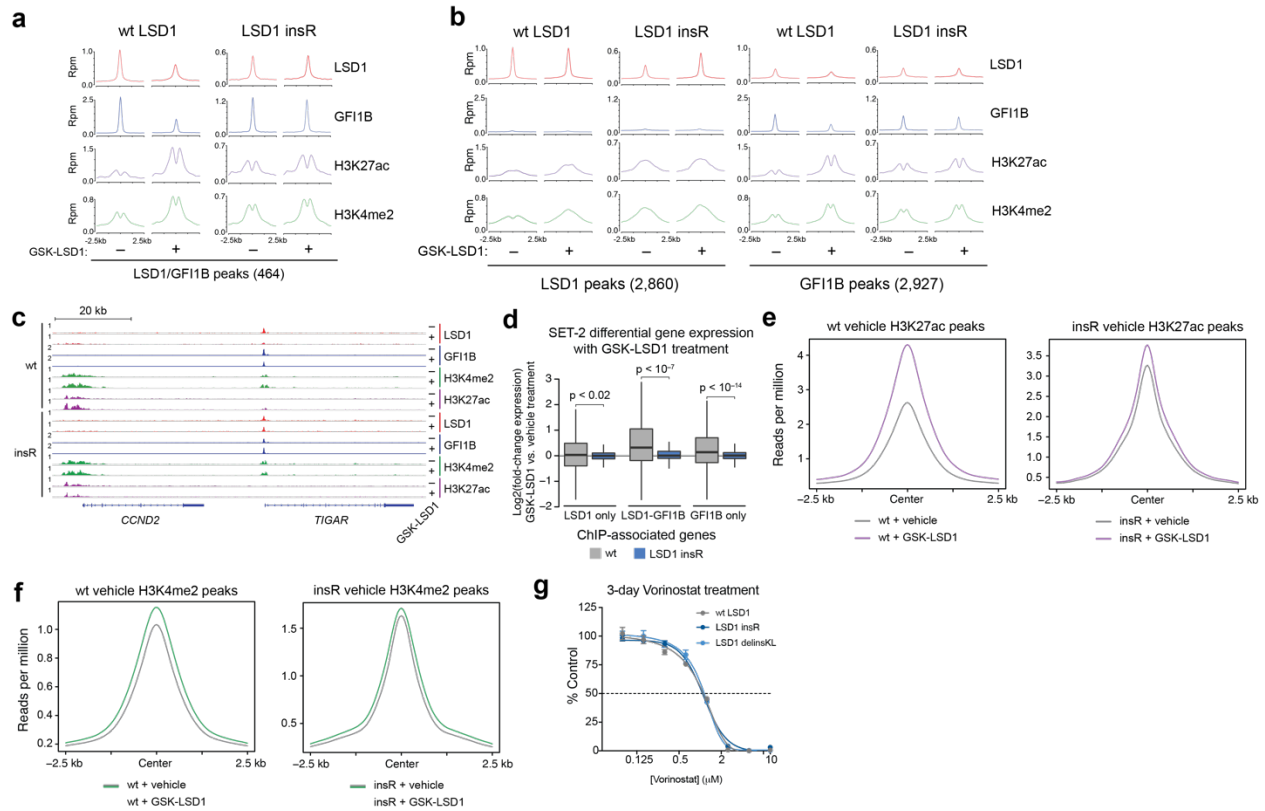
Figure 2.19. Drug-resistant AML cells maintain LSD1–GFI1B binding on chromatin and fail to activate GFI1B-bound enhancers in the presence of GSK-LSD1.

a, ChIP–seq composite plots showing average signal (*y* axis, reads per million (r.p.m.)) for LSD1, GFI1B, H3K27ac and H3K4me2 in wild-type SET-2 and SET-2 LSD1 insR cells treated with vehicle or GSK-LSD1 (250 nM, 48 h) centered around LSD1–GFI1B co-occupied peaks identified in wild-type SET-2. The *x* axis shows flanking regions of ± 2.5 kb around the peak center. Experiment performed once. **b,** ChIP-seq composite plots showing average signal (y axis, rpm) for LSD1, GFI1B, H3K27ac, and H3K4me2 ChIP-seq in wild-type SET-2 and SET-2 LSD1 insR cells treated with vehicle or GSK-LSD1 (250 nM, 48 h) centered around LSD1-only and GFI1B-only peaks identified in wild-type SET-2. The x axis shows flanking regions of ±2.5 kb around the peak center. Experiment performed once. **c,** ChIP-seq composite plots showing average peak signal (y axis, rpm) for LSD1, GFI1B, H3K4me2, and H3K27ac in wild-type SET-2 and SET-2 LSD1 insR cells under vehicle or GSK-LSD1 treatment (250 nM, 48 h). Experiment performed once. **d**, Box plot showing differential gene expression of ChIP–seq-associated gene subsets upon GSK-LSD1 treatment (250 nM, 48 h) in wild-type SET-2 or SET-2 LSD1 insR cells. In the plot, bars represent the

Figure 2.19 (continued) median, the box represents the IQR and the whiskers represent 1.5 × IQR. *P* values were calculated using the Wilcoxon test (two-sided). **e,f** ChIP-seq profile plots showing average peak signal (y axis, rpm) for H3K27ac (**e**) and H3K4me2 (**f**) in wild-type SET-2 and SET-2 LSD1 insR under vehicle or GSK-LSD1 treatment (250 nM, 48 h). The x axis depicts flanking regions of ±2.5 kb around the peak center. **g,** Dose-response curves for cell lines treated with the pan-HDAC inhibitor, vorinostat, are shown. Data represent mean values ± s.e. across three replicates. One of two independent replicates is shown.

To investigate how drug-induced disruption of GFI1B binding may impact gene expression, we considered the expression of the closest gene to each LSD1-GFI1B ChIP-seq binding site. The median expression of these genes in wt SET-2 cells significantly increased upon treatment with GSK-LSD1 (**Figure 2.19d**) and was accompanied by increases in levels of H3K27ac and H3K4me2, histone modifications associated with active cis-regulatory regions, at LSD1-GFI1B peaks in wt SET-2 (**Figure 2.19a,c**). Notably, the drug-induced increase in H3K27ac, but not in H3K4me2, extended beyond GFI1B-bound sites, suggesting a broader role of LSD1 inhibition in enhancer activation in SET-2 cells (**Figure 2.19e,f**). By contrast, increased expression of LSD1-GFI1B-associated genes and accompanying changes in chromatin state were suppressed in drug-resistant SET-2 LSD1 insR upon GSK-LSD1 treatment (**Figure 2.19b-d**). The necessity of the LSD1-GFI1B interaction may stem from its ability to recruit histone deacetylase complexes to silence LSD1-GFI1B target genes.[70,111] In support, the LSD1 drug-resistance mutations confer no protective advantage against vorinostat, a histone deacetylase inhibitor (**Figure 2.19g**). Altogether, our observations are consistent with the notion that GSK-LSD1 leads to the dissociation of LSD1-GFI1B complexes from chromatin, inducing subsequent enhancer activation and up-regulation of many LSD1-GFI1B-associated genes.
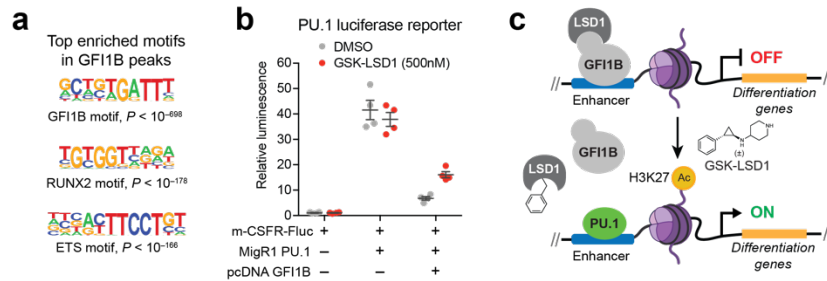
Figure 2.20. Downstream activation of enhancers may in part be mediated by PU.1

b, Consensus GFI1B, RUNX2 and ETS motif logos detected in GFI1B ChIP–seq peaks in wild-type SET-2 cells. The corresponding *P* values for enrichment are calculated as previously described.[112] **b**, Scatter plots showing relative firefly luciferase (Fluc) expression in the absence or presence of a PU.1 reporter plasmid (*M-CSFR-Fluc*), PU.1 expression plasmid (MigR1 *PU.1*), GFI1B expression plasmid (pcDNA *GFI1B*), or GSK-LSD1 (500 nM, 48 h). Gray lines represent mean ± s.e.m. across four replicates. One of two independent replicates is shown. **c**, Schematic of a proposed model, whereby GSK-LSD1 induces displacement of a LSD1–GFI1B complex from enhancers allows PU.1-mediated activation of genes involved in AML differentiation.

To gain an understanding of what factors may drive enhancer activation after GFI1B eviction, we interrogated the DNA sequences underlying GFI1B peaks for over-represented motifs. Aside from GFI1B binding motifs, this analysis identified binding motifs for RUNX and ETS factors as highly enriched (**Figure 2.20a**),[112] nominating them as candidate activators. In particular, PU.1 is an ETS factor that is antagonistic to GFI1/GFI1B, and prior work demonstrated that LSD1 inhibitors can activate PU.1 target genes.[70,93,94] To establish a possible connection between PU.1, GFI1B, and LSD1 inhibitors, we employed a PU.1 luciferase reporter driven by the mouse *Csfr* promoter.[113] Overexpression of wt GFI1B diminished PU.1 activity in the reporter, consistent with earlier studies showing that GFI1 can antagonize PU.1 transactivation (**Figure 2.20b**).[113] Significantly, GSK-LSD1 treatment partially suppressed the inhibitory effects of GFI1B on PU.1 transactivation, demonstrating a role of LSD1 in this mechanism. Altogether, our studies support a model where a chromatin-associated LSD1-GFI1B complex, independent of LSD1

enzyme activity, represses acetylation and activation of enhancers through potentially suppressing PU.1 activity (**Figure 2.20c**).

2.8. Discussion

The identification of cancer mutations that alter the function of chromatin modifying enzymes has brought increasing attention to them as drug targets.[114–117] Although the notion that inhibitors targeting chromatin modifiers exert their anti-proliferative activity by correcting aberrant chromatin states is often invoked, whether changes in chromatin are directly responsible for their efficacy is often unclear. Motivated by these questions, we have employed CRISPR-suppressor scanning to dissect the mechanism-of-action of chemical inhibitors directly in cells. Evaluating a small panel of LSD1 inhibitors, our studies show that CRISPR-suppressor scanning can discriminate between structurally similar and distinct molecules at both the level of sgRNA enrichment and coding mutations. These data can be integrated with structural information and computational methods developed here to provide structure-function relationships that may reveal functional regions of proteins, modes of drug binding, as well as hotspots critical for drug potency. Although exact molecular interactions cannot be described using this approach, we expect that computational modeling in addition to advancements that improve the resolution of CRISPR-mutagenesis may further refine the ability to resolve these interactions. Moreover, the mutations identified through this approach may predict how easily cells develop resistance to small molecules, which could have significant clinical implications. In particular, in the context of LSD1, our data suggests that resistance mutations may be easier to achieve since they are not constrained to the selection pressure of maintaining enzymatic function.

Using CRISPR-suppressor scanning, we independently uncover and corroborate the unexpected mode by which LSD1 inhibitors block AML cell growth, where inhibition of demethylase activity is not directly responsible for blocking proliferation.[70] Instead, CRISPR-mutagenesis identified drug-resistant LSD1 mutants that are enzyme-inactive but bind to GFI1B in the presence of GSK-LSD1. To show that the LSD1-GFI1B complex is sufficient for AML proliferation, we identified a compensatory GFI1B mutant allele that structurally complements LSD1 specifically modified by GSK-LSD1. The bump-hole rescue, afforded by this orthogonal

drug-complementary GFI1B allele, establishes the essential nature of the LSD1-GFI1B interaction in its native context. This orthogonal drug-complementary allele could prove critical for interrogating this important protein-protein interaction with precise chemical control.

Leveraging drug-resistant mutants, we show that LSD1 inhibitors operate by disrupting a complex of LSD1 with GFI1B on chromatin, inducing the activation of GFI1B-target genes in a demethylase independent fashion. This activation may in part be mediated by PU.1. Although PU.1 had previously been implicated in the anti-proliferative mechanism of LSD1 inhibitors,[70,93,94] its direct connection to GFI1/GFI1B has not been established in this context. We show that LSD1 inhibitors suppress the ability of GFI1B to repress PU.1 transactivation, establishing a possible direct link between GFI1B and PU.1 in AML.[113] More broadly, our results suggest that inhibition of LSD1's scaffolding as opposed to its enzymatic activity is the major driver underlying drug-induced alterations in gene expression. Instead, LSD1 inhibitors potentially offset an opposing relationship between the two transcription factors, GFI1B and PU.1,[93,113] which may reflect aspects of their antagonism during normal hematopoiesis.[109] These findings highlight the nuanced mechanisms by which chromatin modifiers regulate gene expression outside of their archetypal functions.

In summary, we demonstrate the versatility of CRISPR-suppressor scanning to interrogate target protein-small molecule interactions through deciphering the mechanism-of-action of LSD1 inhibitors. Drug-resistant alleles are often generated through rational design with the explicit goal of validating a small molecule target.[118,119] By systematically identifying drug-resistant alleles that span a target, many of which occur through unexpected mutations, CRISPR-suppressor scanning can be used as a hypothesis generator to discover new aspects of target biology and drug mechanism-of-action. More broadly, our work raises important questions regarding the precise roles of chromatin regulators' enzymatic versus scaffolding activities in the diverse functions these modifiers play in gene regulation.

2.9. Methods

Cell culture and lentiviral production. MV4;11, CMK-86, and K562 were obtained from ATCC; SET-2 was a gift from Matthew D. Shair; HEK 293T was a gift from Bradley E. Bernstein. All cell lines were authenticated by Short Tandem Repeat profiling (Genetica) and routinely tested for mycoplasma (Sigma-Aldrich). All media were supplemented with 100 U/mL penicillin and 100 µg/mL streptomycin (Life Technologies) and fetal bovine serum (FBS, Peak Serum). All cell lines were cultured in a humidified 5% CO_2 incubator at 37 °C. MV4;11, CMK-86, and K562 were cultured in RPMI-1640 (Life Technologies), 10% FBS. SET-2 were cultured in RPMI-1640, 20% FBS. HEK 293Ts were cultured in DMEM (Life Technologies), 10% FBS. For lentivirus production, plasmids were co-transfected with *GAG/POL* and *VSVG* plasmids into HEK 293T using FuGENE HD (Promega). Media was exchanged after 8 h and the viral supernatant was collected 72 h after transfection and filtered (0.45 µm). Cell lines were transduced by spinfection at 1,800 × *g* for 2 h at 37 °C with 8 µg/mL polybrene (Santa Cruz Biotechnology). After 48 h post-transduction, Puromycin (ThermoFisher Scientific) selection was carried out for 5 d at 1.5 µg/mL for SET-2 and 1 µg/mL for MV4;11.

Chemical Reagents. Compounds were stored at –20 °C in 100% dimethylsulfoxide (DMSO). The vehicle condition represents 0.1% DMSO treatment. GSK-LSD1 and GSK690 was used as purchased from Sigma-Aldrich (≥98% by HPLC) and Aobious (98% purity by HPLC) respectively. AW1-5 were synthesized as described in the Supplementary Note, and NV93 was synthesized according to a previously described protocol.[103]

Cell Growth Assays. 2,500 cells/well for MV4;11 and 5,000 cells/well for all other cell lines were plated in triplicate with drug or vehicle treatments. An equal volume of cells from each well was split and re-plated with fresh media containing vehicle or inhibitor at day 5 and day 10. Cell viability

was monitored at day 10 and 15 by measuring end point luminescence using CellTiter-Glo (Promega) on a SpectraMax i3x platereader. Dose-response curves were determined through interpolation using GraphPad Prism v.7 nonlinear regression fit ([inhibitor] vs normalized response -- variable slope). All growth assays were performed at least twice.

sgRNA Pooled Cloning and CRISPR-Scanning Experiments. The LSD1 tiling library included every sgRNA with an NGG PAM and a cleavage site within the LSD1 coding sequence (*NP_055828.2*). These sgRNA sequences are listed in **Supplementary Dataset 1**. Oligos containing the sgRNA sequences were cloned into pLentiCRISPR.v2 in a pooled fashion as previously described.[86] pLentiCRISPR.v2 was a gift from Feng Zhang (Addgene plasmid #52961). Lentiviral particles carrying the resultant LSD1 tiling library were generated as described above and titered according to literature procedure.[120] 1 x 10^7 cells were transduced such that the multiplicity of infection < 0.3 and subsequently selected with puromycin. Following puromycin selection, cells were split into pools and treated with drug(s) or vehicle. Genomic DNA was isolated at specified timepoints using the QIAamp DNA Mini Kit (Qiagen). All CRISPR-suppressor scanning experiments were performed with three separate transductions. To measure the composition of the population, sgRNA sequences for all replicates were amplified with PCR primers (**Supplementary Dataset 5**) and sequenced as previously described.[86]

Generation of Clonal Drug-Resistant LSD1 Mutant Cell Lines. sgRNA sequences targeting LSD1, sgG655 (GGGATTTGGCAACCTTAACA) and sgN660 (TAGGGCAAGCTACCTTGTTA), were cloned into pLentiCRISPR-eGFP.v1. SET-2 and MV4;11 cells were transduced with the resultant plasmids as indicated above and subsequently treated with 100 nM GSK-LSD1 for 50 d to enrich for drug-resistant mutant cells. Surviving cells were FACS-sorted as single cells and then expanded. pLentiCRISPR-eGFP.v1 was a gift from Beat Bornhauser (Addgene plasmid #75159).

Genotype Determination. Genomic DNA was isolated as mentioned above. For library preparation, genomic PCR primers (**Supplementary Dataset 5**) with Illumina adaptor sequences were used to amplify specified regions of LSD1 as previously described.[121] Samples were sequenced on a MiSeq genome analyzer (Illumina). The sequencing reads were analyzed using CRISPResso (version 1.0.13).[104]

Mammalian Overexpression Constructs. The LSD1 and LSD1 (K661A) open reading frames (ORFs) were subcloned from pCMV LSD1 and pCMV LSD1 (K661A), which were gifts from R. Shiekhattar. Site directed mutagenesis was performed using Q5 Site-Directed Mutagenesis Kit (New England Biolabs) to generate LSD1 N660delinsKL and LSD1 L659_N660insR. All modified LSD1 ORFS were subcloned into pSMALP and pcDNA.3 with an N-terminal FLAG tag using Gibson cloning (New England Biolabs). GFI1 and GFI1B were subcloned from pENTR GFI1 and pENTR GFI1B, respectively, which were gifts from Huda Zoghbi (Addgene plasmid #16168 and #16169, respectively), and subcloned into pcDNA.3 or pSMALP with a C-terminal FLAG tag using Gibson cloning. Site directed mutagenesis was performed using Gibson cloning to generate GFI1 (F5A) and GFI1B (F5A). The *Renilla* luciferase ORF was subcloned into pcDNA.3 using Gibson cloning. pSMALP was generated from pSMAL through introduction of a puromycin cassette into pSMAL (a gift from P. van Galen).

Immunoblotting. Cells were lysed on ice using RIPA buffer (Boston BioProducts) supplemented with fresh HALT Protease Inhibitor (ThermoFisher Scientific) and the lysates were cleared. The protein concentration of the lysate was determined using the BCA Protein Assay Kit (ThermoFisher Scientific). Immunoblotting was performed according to standard procedures. The primary antibodies used are as follows: LSD1 (Bethyl Laboratories A300-215A, LOT# 2). GFI-1B (B-7) X (Santa Cruz Biotechnology sc-28356X, LOT# D1615). Monoclonal Anti-FLAG M2 (Sigma-

Aldrich F1804, LOT# SLBW3851). GAPDH (Santa Cruz Biotechnology sc-477724, LOT# B0210). GFI1B (Santa Cruz Biotechnology sc-28356X, LOT# D1615).

Co-immunoprecipitation (Co-IP). 2×10^7 SET-2 cells expressing wt GFI1B-FLAG or GFI1B-FLAG (F5A) were washed twice with cold PBS and flash frozen. Cells were thawed, lysed on ice in N450 buffer (50 mM Tris-HCl pH 7.5, 450 mM NaCl, 1 mM $MgCl_2$, 1% NP-40 alternative, 5% glycerol) supplemented with 1:10,000 benzonase (Sigma-Aldrich), and the lysates were cleared. The protein concentration was quantified as above and diluted to 1 mg/mL in lysis buffer. Supernatants were immunoprecipitated overnight at 4 ºC with 2.5 µg anti-FLAG M2 antibody and 40 µL Protein G Dynabeads (ThermoFisher Scientific). Beads were washed twice with lysis buffer, eluted in SDS-PAGE loading buffer, and carried forward to immunoblotting as described above. For Co-IPs performed in HEK 293T cells, cells were plated at 70-80% confluency and treated with 250 nM GSK-LSD1 or vehicle. 24 h after plating, cells were co-transfected with 1 µg pcDNA.3 FLAG-LSD1 plasmid (wt, L659_N660insR, or N660delinsKL) and 2 µg pcDNA.3 wt GFI1B-FLAG using FuGENE HD. 48 h post-transfection, cells were washed with PBS then lysed with N450 buffer. Co-IP was performed using 4 µg GFI1B antibody.

Cellular Thermal Shift Assay (CETSA). CETSA was performed based on a modified protocol.[122] 2×10^6 cells were treated with vehicle or 100 nM GSK-LSD1 for 1 h at 37 ºC, then washed with PBS and resuspended in 500 µL PBS supplemented with Roche cOmplete Protease Inhibitor Cocktail Tablets, EDTA-free. Aliquots of 25 µL from each condition were distributed into PCR strip tubes and heated at 40, 43, 45, 47, 50, or 53 ºC for 3 min before being cooled to rt for another 3 min. The samples were lysed by three freeze-thaw cycles. The lysate was clarified by centrifugation at 17,000 x *g* for 20 min at 4 ºC. The samples were analyzed by immunoblotting as described above.

Luciferase Reporter Assay. According to a modified procedure, HEK 293T cells were plated in quadruplicate at a density of 1,000-1,500 cells/well and treated with vehicle or 500 nM GSK-LSD1.[113] After 24 h, the cells were transfected with 200 ng pcDNA.3 GFI1B-FLAG using FuGENE HD. 24 h post-transfection, 15 ng pcDNA.3 *Renilla* luciferase, 50 ng m-CSFR *Firefly* luciferase, and 100 ng migR1 PU.1 were co-transfected. 48 h after the second transfection, the Dual-Luciferase Reporter Assay System (Promega) was used to measure end point luminescence (SpectraMax i3x). mCSFR *Firefly* luciferase and MigR1 PU.1 plasmids were gifts from M. Simon and R. Dahl.

Protein Purification. For bacterial constructs, the LSD1 (Δ1-150) gene was codon optimized and synthesized as two fragments from Integrated DNA Technologies and Quintara Biosciences. The fragments were cloned into pET28b containing an N-terminal His$_6$-tag using Gibson cloning. Mutations were introduced with Q5 Site-Directed Mutagenesis Kit. The constructs were expressed in NiCo21(DE3) competent *E. coli* (New England Biolabs) using a previously described protocol.[123] Protein fractions with > 95% purity as assessed by SDS PAGE were pooled and stored at –80 ºC in 25 mM sodium phosphate pH 7.4, 150 mM NaCl, 0.3 mM TCEP, and 5% glycerol (storage buffer). CoREST(305-482) was codon optimized for bacterial expression and assembled into pET28b containing an N-terminal GST-tag followed by a TEV cleavage site from a synthesized fragment purchased from GeneWiz. Recombinant CoREST expression and purification were carried out according to a modified literature procedure.[51] Recombinant protein was purified by GST affinity chromatography using a linear gradient of 0-50 mM reduced glutathione in lysis buffer and the buffer was exchanged to TEV protease cleavage buffer (50 mM Tris-HCl pH 8.0, 75 mM NaCl, 1 mM DTT, and 0.5 mM EDTA). The GST-tag was removed by incubation with TEV protease overnight at 4 °C. The cleaved protein was purified using a GSTrap column followed by Superdex 200 10/300 GL column (GE Healthcare) in storage buffer. Purified CoREST was incubated with LSD1 in a 2:1 molar ratio for 2 h and gel-filtered on a Superdex 200

10/300 GL column equilibrated in storage buffer. The purity of the complex was verified by SDS-PAGE and fractions with 90-95% purity were pooled and stored at –80 °C. FPLC traces for proteins purified by size-exclusion chromatography are included in **Supplementary Figure 8.**

LSD1 Enzyme Assays. LSD1 enzymatic activity assays were performed in triplicate using Amplex Red Hydrogen Peroxide/Peroxidase Assay Kit (Invitrogen) with recombinant LSD1 and a synthetic peptide corresponding to the first 21 amino acids of H3K4me2 (Anaspec). LSD1 (500 ng/well) was incubated with 20 µM peptide at rt for 30 min. The endpoint fluorescence was measured on a microplate reader (excitation: 530 nm; emission: 590 nm) after 60 min following the addition of the Amplex Red/HRP mixture. Inhibition assays were performed as described above. Briefly, LSD1 (75 ng/well) and inhibitors at the appropriate concentration were incubated at rt for 10 min in reaction buffer with 0.01% BRIJ35 (ThermoFisher Scientific) prior to the addition of peptide. K_i values were determined in GraphPad Prism v.7 by nonlinear regression analysis (*One site-Fit K_i*) of the concentration/inhibition data.

***Thermo*FAD Assay.** *Thermo*FAD was performed in quintets as previously described.[106] Thermal curves were obtained using a qRT-PCR detection system (BioRad C1000 Touch Thermal Cycler) with a temperature gradient from 20 to 95 °C. The T_m was determined by fitting the thermal curves to a Boltzmann sigmoid and calculating the inflection point (GraphPad Prism v.7).

UV/vis Spectroscopy. UV/vis spectra were collected every 20 nm from 300 nm to 600 nm on a microplate reader (SpectraMax i3x) at a protein concentration of 10 µM.

Circular Dichroism. CD spectra were acquired using a quartz cuvette with a 1 mm path length on a Jasco J-815 CD spectrometer collecting five scans (1 s averaging time) for each spectrum. The protein samples (2.5 µM) were prepared in ThermoFAD buffer. Mean residue ellipticity (MRE,

deg*cm^2 *dmol^{-1}) values were calculated using the following equation, where θ is ellipticity (mdeg), l is pathlength (cm), C is peptide concentration (M), N is number of residues.

$$MRE = \frac{\theta}{10 * C * l * N}$$

Fluorescence Polarization Assay. Binding assays were performed in two independent experiments with three technical replicates. The change in fluorescence polarization of fluorescently-labeled GFI1B and GFI1B (F5A) peptides upon binding to LSD1-CoREST was monitored using a previously described protocol.[111] LSD1-CoREST (2 µM) was incubated with the labeled peptides (2 nM) for 1 h on ice. After incubation, the samples were prepared by a 2-fold serial dilution in the assay buffer (15 mM KH$_2$PO$_4$, pH 7.2, 5% glycerol, and 1 mg/mL BSA) containing 2 nM labeled peptide.

For assays in the presence of 10 µM GSK-LSD1, the protein complex was incubated with the inhibitor for 1 h on ice and then with the labeled peptides for 1 h. Fluorescence polarization was measured using a microplate reader in 384-well black microplates at 25 °C. The G-factor on the microplate reader was adjusted to 35 mP for the reference well containing labeled peptide. The binding curves were fit by nonlinear regression analysis in GraphPad Prism v.7 as described previously.[111]

Structural Modeling Analyses. MODELLER was used to generate homology models of the in-frame mutants of LSD1 by using PDB structure 2hko as a template. 1,000 models were generated per structure. The best model for each variant, as judged by DOPE score, was refined using the Relax application in Rosetta 2018 (release 7111c54).[124] The scoring function was used with harmonic constraints on the backbone atom positions. These restraints were ramped down towards the end of the optimization using the -relax:constrain_relax_to_start_coords option.

To assess the water-mediated hydrogen bonding network between FAD and K661, semi-empirical models were built of the active site for selected structures. Complete residues within 7 Å of atom C4X in the FAD ligand or 4 Å of the water oxygen were exported to Gaussian 16 with all connections capped by hydrogens. Each structure was optimized with the PDDG semi-empirical quantum mechanical method with the protein residues fixed while the FAD and water ligands remained unconstrained.[125] SNAG peptide mutations were generated in Chimera suite and modeled in Maestro Schrödinger using Protein Preparation Wizard (OPLS3 force field).

ChIP-seq. ChIP was performed as previously described in singlicate on SET-2 cells (wt and LSD1 L659_N660insR) treated with GSK-LSD1 (250 nM) or vehicle for 48 h.[126] The following antibodies were used: LSD1 (Bethyl Laboratories A300-215A, LOT# 2); GFI1B (Abcam ab26132, LOT# GR318193-8); H3K4me2 (Active Motif 39141, LOT# 01008001); H3K27ac (Active Motif 39133, LOT# 31814008). Libraries were sequenced on the NextSeq 500 sequencer (Illumina) for 75 bases in paired-end mode.

RNA-seq Library Preparation. In triplicate separate cell cultures, total RNA from SET-2 cells (wt and LSD1 L659_N660insR) treated with GSK-LSD1 (250 nM) or vehicle for 48 h was isolated using the RNeasy Plus Mini Kit (Qiagen). Library preparation was performed using the QuantSeq 3' mRNA-Seq Library Prep Kit FWD (Lucigen) according to the manufacturer's protocol. The samples were sequenced on the NextSeq 500 (Illumina) for 42 bases in single-end mode.

Analysis of sgRNA Enrichment and Clustering Method. All data processing was performed in Python version 3.5.5 (www.python.org) and R (version $\geq$ 3.4.2) (www.Rproject.org). sgRNA enrichment was calculated as previously described.[86] Briefly, reads uniquely assigned to each sgRNA sequence were read count normalized, $\log_2 + 1$ transformed, averaged across 3 replicates, and were then normalized by subtracting the average of the $\log_2$-transformed counts

across 3 replicates of that sgRNA at day 0. Finally, for each condition the mean of the transformed counts for all functionally neutral genome-targeting control sgRNAs (averaged across replicates and normalized as described above) was subtracted from each sgRNA's enrichment value for that specific condition. This value is referred to as the normalized $\log_2$(enrichment score). Libraries that did not receive adequate sequencing depth (>1000X coverage) and sgRNAs with zero counts at day 0 were excluded from the analysis.

Pairwise distances between LSD1 amino acid residues were obtained by determining the Euclidean distance between the centroids of each residue using PyMOL (version 1.8.4.2) (PDB: 2y48). To calculate the proximity weighted enrichment score (PWES), sgRNAs were first assigned to the amino acid closest to the predicted Cas9 cleavage site. The PWES score was then determined as follows:

$$PWES = \frac{n_{i,j}}{|n_{i,j}|}\left(\frac{|n_{i,j}|^m}{|n_{i,j}|^m + \theta^m}\right)e^{-\left(\frac{d_{i,j}^2}{2t^2}\right)}$$

$$m = 2$$

$$\theta = 3$$

$$t = 16$$

$$n_{i,j} = \log_2(enrichment)_i + \log_2(enrichment)_j$$

$$d_{i,j} = Euclidean\ distance\ between\ residue_i\ and\ residue_j$$

An analogous approach was employed in the CLUMPS algorithm except that mutation frequencies instead of sgRNA normalized $\log_2$(enrichment) scores were used.[101] To group the sgRNAs according to their pairwise PWES, hierarchical clustering was employed with the Ward

clustering method and Euclidean distance as a metric. Clusters were called using a cophenetic distance cutoff.

To test the significance of the clustering of enriched sgRNAs in 3D space, the absolute values of all possible pairwise combinations of the PWES were summed for the sgRNAs in each specified category. An empirical p-value was calculated by scrambling the positions of the sgRNAs, while keeping the enrichment scores constant, and recalculating the summed PWES for each of 100,000 random permutations.

Data Processing for CRISPR-Suppressor Scanning Analysis. Data processing was performed using the same languages as in the CRISPR-scanning experiments specified above. sgRNA counts were normalized with the following modification: the sgRNA read counts were normalized relative to the corresponding count values of the vehicle treatment at the time point of interest. For the principal component analysis, the top 5 sgRNAs that contributed both positively and negatively to each component were selected based on their coefficient loadings. The sequencing reads from genotyping exons 8, 15, 16, and 17 for the compound CRISPR-suppressor scanning experiment were analyzed using CRISPResso (version 1.0.13).[104]

RNA-seq data processing and differential gene expression analysis. RNA-seq data was processed according to the QuantSeq 3' mRNA-Seq Library Prep recommended analysis pipeline through alignment to the Ensembl transcriptome (GRCh37.75). DESeq2 with R (version 3.5.1) was employed for the differential expression analysis.[127] Hierarchical clustering was performed on the regularized $\log_2$-transformed counts using Euclidean distance as a metric and the Complete clustering method. To identify variable genes across cell types and conditions, all six possible pairwise comparisons of conditions were performed. All genes with both an adjusted p-value < 0.01, calculated using the Benjamini-Hochberg correction, and $|\log_2$ fold change$| > 2$ in at

least one of the comparisons was considered to be differentially expressed. The regularized $\log_2$-transformed counts for the variable genes were grouped by k-means clustering (k=7) and the Z-scores for the variable genes, ordered according to their cluster identity are depicted. A table of significantly differentially expressed genes is provided in **Supplementary Dataset 6**. GSEA (version 3.0)[94] was carried out on $\log_2 + 1$ transformed count values and the p-values for GSEA were calculated as previously described.[94] Gene signatures used for GSEA are listed in **Supplementary Dataset 7.**

ChIP-seq Data Analysis. ChIP-seq data was analyzed as previously described.[126] Peaks were called with HOMER (version 4.10) using matched inputs with the following parameters: H3K4me2, –histone –tagThreshold 30; H3K27ac, –histone –tagThreshold 50; LSD1 -style factor -tagThreshold 10 -P 0.000001 -L 5 -F 10; GFI1B -style factor -tagThreshold 10 -P 0.000001. ChIP-seq peaks were annotated to the closest transcription start site using HOMER (version 4.10). Transcription factor motif enrichment analysis was performed using HOMER (version 4.10) centered on previously called GFI1B peaks (parameters: –size 200 –mask). HOMER (version 4.10) was used to identify overlapping and unique peaks for LSD1 and GFI1B. ChIP-seq profile plots were generated using ngs.plot.[128]

Statistical Methods. Statistical parameters including the exact value and definition of n, the definition of center, dispersion, precision measures (mean ± s.d. or s.e.) and statistical significance are reported in Figures and Figure Legends.

Data availability statement. ChIP-seq and RNA-seq data have been deposited to NCBI GEO (GSE121426).

2.10. Chemistry methods

General Procedures. All reactions were performed in oven-dried or flame-dried glassware under a positive pressure of nitrogen unless otherwise noted. Flash column chromatography was performed as described by Still et al. employing SiliaFlash® P60 (230-400 mesh, SiliCycle)[129]. Flash column chromatography was conducted on a Biotage Isolera automated chromatography system or manually in a glass column unless otherwise specified. Preparatory and analytical thin-layer chromatography (TLC) was performed on Silica Gel 60 F_{254} plates (EMD Millipore). TLC plates were visualized by exposure to ultraviolet light (UV) and exposure to an aqueous solution of ceric ammonium molybdate (CAM), *p*-anisaldehyde or ninhydrin followed by heating on a hot plate. Organic solvents were concentrated under reduced pressure on a Büchi rotary evaporator.

Materials. Commercial reagents and solvents were used as received with the following exceptions: tetrahydrofuran (THF), dichloromethane (CH_2Cl_2), *tert*-butyl methyl ether (MTBE), toluene, and *N,N*-dimethylformamide (DMF) were degassed with argon and passed through a solvent purification system (designed by Pure Process Technology) utilizing alumina columns as described by Grubbs et al.[130] *n*-Butyllithium was purchased as a 2.5 M solution in hexanes (Sigma-Aldrich). The molarities of *n*-butyllithium solutions were determined by titration using 1,10-phenanthroline as an indicator (average of three determinations). 4-Bromobenzaldehyde (EMD Millipore, $\geq$ 98%), *t*-Butyl dimethylphosphonoacetate (TCI, > 95.0%) and 1-Boc-4-piperidone (Sigma-Aldrich, 98%) were used as purchased. $CDCl_3$ (Cambridge Isotope Laboratories) was used with and stored over activated molecular sieves (4Å) prior to use. Deuterated solvents CD_3OD and DMSO-d_6 (Cambridge Isotope Laboratories) were used as purchased. Extraction and chromatography solvents were reagent grade and used without purification (VWR or Fisher Scientific). Celite® 545 (EMD Millipore) was used.

Instrumentation. ^{1}H NMR spectra were recorded with a Varian INOVA-500 spectrometer in parts per million (δ), and were calibrated using residual undeuterated solvent as an internal reference (CDCl$_3$: δ 7.26 (CHCl$_3$), CD$_3$OD: δ 3.31 (CD$_2$HOD), DMSO-d_6: δ 2.5 (C$_2$D$_5$HSO). Data for ^{1}H NMR spectra are reported as follows: chemical shift (δ ppm) (multiplicity, coupling constant (Hz), integration). Multiplicities are reported as follows: s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, br = broad, or combinations thereof. ^{13}C NMR spectra were recorded on the Varian INOVA-500 spectrometer in parts per million (δ) and are referenced to the carbon resonances of the solvent (CDCl$_3$: δ 77.00, CD$_3$OD: δ 49.15, DMSO-d_6: δ 39.51). High-resolution mass spectra (HRMS) were recorded using electrospray ionization (ESI) mass spectroscopy experiments on an Agilent 6210 TOF LC/MS (Harvard FAS Division of Science Small Molecule Mass Spectrometry).

Synthetic Scheme towards GSK-LSD1 Analogs:[131,132]
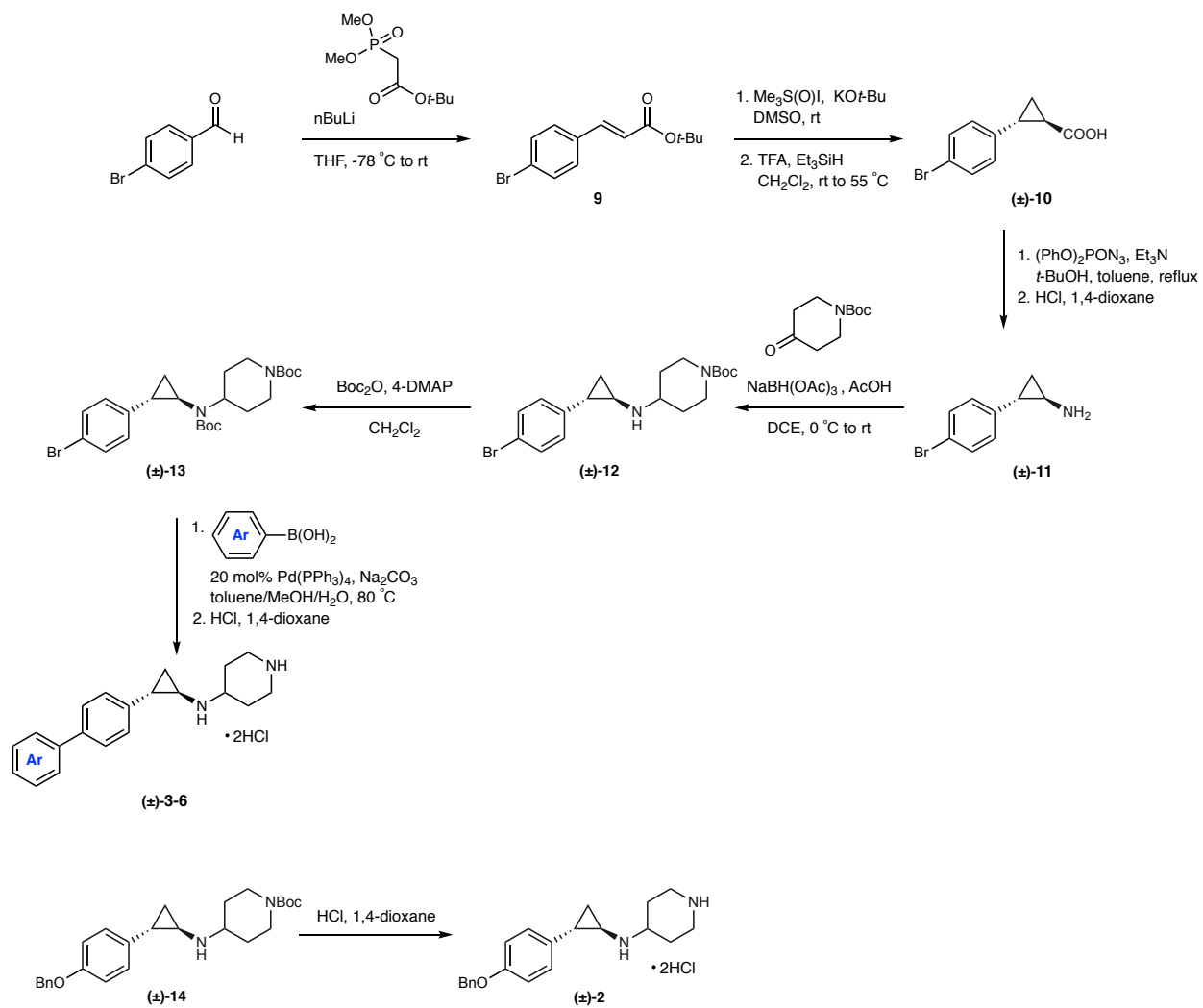
General Procedure for tert-butyl (E)-3-(4-bromophenyl)acrylate (9):
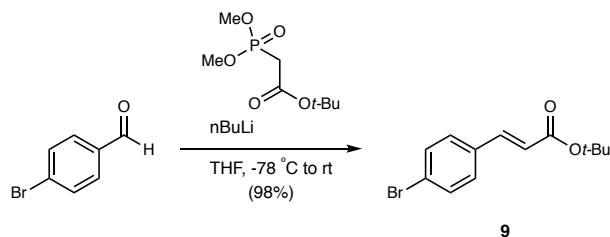
According to a literature procedure,[133] n-BuLi (1.95 mL, 4.86 mmol, 1.20 equiv, 2.50 M in hexanes) was added dropwise at –78 °C to a stirred solution of *tert*-butyl dimethylphosphonoacetate (0.96 mL, 4.86 mmol, 1.20 equiv) in THF (4.00 mL). After stirring for 30 min at –78 °C, a solution of aldehyde (750 mg, 4.05 mmol, 1.00 equiv) in THF (4.66 mL) was also cooled to –78 °C and transferred via cannula. The resulting solution was stirred at –78 °C for 30 min before being allowed to warm to room temperature. Upon cooling back to -78 °C, the solution was quenched with saturated aqueous NH$_4$Cl solution (50 mL). The aqueous layer was extracted with CH$_2$Cl$_2$ (3 × 50 mL), and the combined organic fractions were dried over anhydrous sodium sulfate, filtered, and concentrated under reduced pressure. The crude residue was purified by column chromatography (silica gel, eluent: 0 to 30% EtOAc/hexanes, v/v) to afford acrylate **9** in high diastereoselectivity (98% yield; >90% *E:Z*).

¹H NMR (500 MHz, CDCl$_3$) δ: 7.55–7.48 (m, 3H), 7.49 (d, *J* = 8.0 Hz, 2H), 7.37 (d, *J* = 8.5 Hz, 2H), 6.35 (d, *J* = 16.0 Hz, 1H), 1.53 (s, 9H). **¹³C NMR** (126 MHz, CDCl$_3$) δ: 166.00, 142.12, 133.59, 132.05, 129.32, 124.12, 120.91, 80.72, 28.17. **HRMS** (ESI) (*m/z*) calc'd for C$_{13}$H$_{15}$BrO$_2$ [M+H]$^+$: 282.0250, 283.0283, 284.0229, 285.0263, 286.0297, 287.0305, 288.0339, 289.0373, 290.0406 found 282.0253, 283.0287, 284.0232, 285.0266.

General Procedure for (±)-2-(4-bromophenyl)cyclopropane-1-carboxylic acid (10):
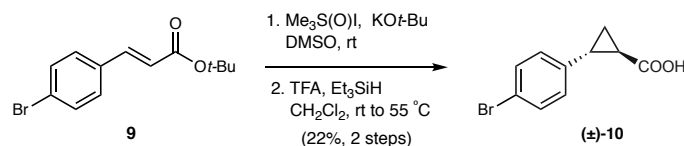
According to a literature procedure,[134] an anhydrous DMSO solution (5.00 mL) of intermediate **9** (1.12 g, 3.97 mmol, 1.00 equiv) was added in one portion to a mixture of $Me_3S(O)I$/KO*t*-Bu (1:1 mixture, 2.18 g/1.11 g, 9.93 mmol, 2.50 equiv) in a round-bottomed flask. The resulting solution was stirred for 30-60 min at 50–60 °C until disappearance of the starting material was observed by TLC analysis. The mixture was then treated with brine (25 mL) and extracted with EtOAc (3 × 15 mL). The combined organic extracts were washed with water (2 × 25 mL) and brine (1 × 25 mL), dried over anhydrous sodium sulfate, filtered, and concentrated under reduced pressure. The crude residue was purified by column chromatography (silica gel, eluent: 0 to 25% EtOAc/hexanes, v/v). The resulting ester was taken up in CH_2Cl_2 (0.90 mL). TFA (895 μL, 11.70 mmol, 13.00 equiv) and triethylsilane (182 μL, 2.25 mmol, 2.50 equiv) were subsequently added according to a modified literature procedure.[6] After stirring for 1.5 to 2 h, TFA was removed by a stream of nitrogen in a well-ventilated hood, and any remaining solvent was removed under reduced pressure. The crude material was passed through a silica plug (silica gel, eluent: 50% EtOAc/hexanes, v/v) to afford *trans*-cyclopropane (±)-**10** as a single diastereomer (22% yield, two steps).

¹H NMR (500 MHz, CDCl$_3$) δ: 7.42 (d, *J* = 8.4 Hz, 2H), 6.99 (d, *J* = 8.4 Hz, 2H), 2.56 (ddd, *J* = 10.4, 6.7, 4.1 Hz, 1H), 1.89 (ddd, *J* = 8.5, 5.2, 4.2 Hz, 1H), 1.68 (dt, *J* = 9.4, 5.0 Hz, 1H), 1.37 (dddd, *J* = 8.5, 6.7, 4.8, 1.0 Hz, 1H). **¹³C NMR** (126 MHz, CDCl$_3$) δ: 177.98, 138.53, 131.60, 128.03, 120.45, 26.44, 23.65, 17.40. **HRMS** (ESI) (*m/z*) calc'd for $C_{10}H_9BrO_2$ [M+H]$^+$: 240.9859, 241.9892, 242.9838, 243.9872, 244.9905, 245.9914, 246.9948 found 240.9855, 241.9888, 242.9834, 243.9867, 244.9895, 246.9118.

General Procedure for (±)-2-(4-bromophenyl)cyclopropan-1-amine (11):
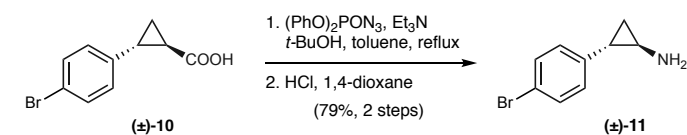
According to a modified literature procedure,[135] acid **10** (210 mg, 0.87 mmol, 1.00 equiv) was first dissolved in dry toluene (3.35 mL) in a round-bottomed flask. Diphenyl phosphoryl azide (225 μL,1.04 mmol, 1.20 equiv) and triethylamine (243 μL, 1.74 mmol, 2.00 equiv) were added under a nitrogen atmosphere and stirred for 30 min at room temperature. The reaction mixture was then heated and refluxed for 1.5 h before *t*-BuOH (1.01 mL, 0.87 mmol, 1.00 equiv) was added, and the resulting solution was refluxed overnight. The reaction mixture was cooled to room temperature and was concentrated under reduced pressure. The crude residue was purified by column chromatography (silica gel, eluent: 0 to 30% EtOAc/hexanes, v/v) to afford the carbamate intermediate. The obtained carbamate product (0.93 mmol, 1.00 equiv) was taken up in 1,4-dioxane (345 μL) and 4.0 M HCl in 1,4-dioxane was added (690 μL, 2.76 mmol, 4.00 equiv). The reaction mixture was stirred at room temperature until completion as monitored by TLC (4 h) before being concentrated under reduced pressure. The residue was taken up in water (15 mL), diluted with saturated aqueous NaHCO₃ solution (15 mL), and the aqueous layer was extracted with CH₂Cl₂ (3 × 15 mL). The combined organic fractions were dried over anhydrous sodium sulfate, filtered, and concentrated under reduced pressure to afford cyclopropyl amine (±)-**11**, which was used directly in the next step without further purification (79% yield, two steps).

¹H NMR (500 MHz, CDCl₃) δ: 7.36 (d, *J* = 8.5 Hz, 2H), 6.88 (d, *J* = 8.5 Hz, 2H), 2.51 (dt, *J* = 7.3, 3.6 Hz, 1H), 1.82 (ddd, *J* = 9.1, 5.7, 3.1 Hz, 1H), 1.06 (ddd, *J* = 9.5, 5.3, 4.3 Hz, 1H), 0.94 (dt, *J* = 7.2, 5.5 Hz, 1H). **HRMS** (ESI) (*m/z*) calc'd for C₉H₁₀BrN [M+H]⁺: 212.0069, 213.0103, 214.0049, 215.0082, 216.0116, 217.0150, 218.0183, 219.0217 found 212.0065, 213.0099, 214.0044, 215.0077, 216.0110, 217.1428, 218.1461.

General Procedure for (±)-tert-butyl 4-((2-(4-bromophenyl)cyclopropyl)amino)-piperidine-1-carboxylate (12):
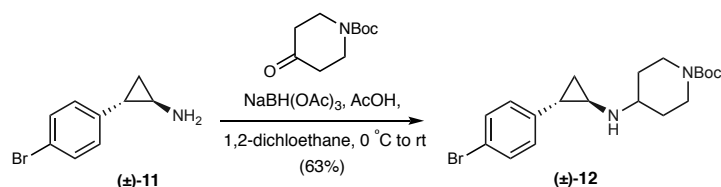
According to a literature procedure,[131] to a solution of amine **11** (146 mg, 0.69 mmol, 1.00 equiv) in 1,2-dichloroethane (4.60 mL) was added *t*-butyl 4-oxopiperidine-1-carboxylate (165 mg, 0.83 mmol, 1.20 equiv) and acetic acid (39 μL, 0.69 mmol, 1.00 equiv). The solution was cooled to 0 °C and sodium triacetoxy borohydride (263 mg, 1.24 mmol, 1.80 equiv) was slowly added. The reaction was stirred at room temperature overnight. The solvent was then removed under reduced pressure and the crude residue was diluted in water (10 mL) and quenched with saturated aqueous $NaHCO_3$ solution (10 mL). The combined aqueous layers were then extracted with EtOAc (3 × 10 mL). The combined organic extracts were washed with water (10 mL) and brine (10 mL), and then dried over anhydrous sodium sulfate, filtered, and concentrated under reduced pressure. The crude residue was purified by column chromatography (silica gel, eluent: 50 to 100% EtOAc/hexanes, v/v) to afford secondary amine (±)-**12** (63% yield).

¹H NMR (500 MHz, CDCl₃) δ: 7.44–7.31 (m, 2H), 6.88 (d, *J* = 8.4 Hz, 2H), 4.01 (m, 4H), 2.88–2.69 (m, 1H), 2.37–2.22 (m, 1H), 1.94–1.76 (m, 4H), 1.45 (s, 9H), 1.08 (dt, *J* = 9.6, 4.9 Hz, 1H), 1.01–0.94 (m, 1H). **¹³C NMR** (126 MHz, CDCl₃) δ: 154.75, 141.18, 131.24, 127.27, 118.98, 79.36, 55.26, 39.75, 32.64, 28.40, 25.17, 17.04. **HRMS** (ESI) (*m/z*) calc'd for $C_{19}H_{27}N_2O_2$ [M+H]⁺: 395.1329, 396.1362, 397.1308, 398.1342, 399.1375, 400.1409, 401.1418, 402.1451 found 395.1316, 396.1349, 397.1295, 398.1329, 399.1357, 400.1368.

General Procedure for (±)-tert-butyl 4-((2-(4-bromophenyl)cyclopropyl)(tert-butoxycarbonyl)-amino)piperidine-1-carboxylate (13):
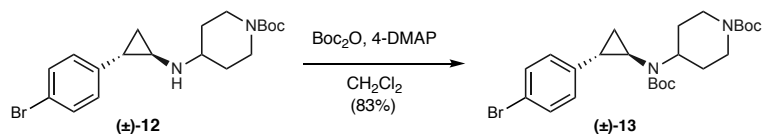
To a solution of (±)-**12** (170 mg, 0.43 mmol, 1.00 equiv) in CH_2Cl_2 (4.30 mL) was added Boc_2O (281 mg, 1.29 mmol, 3.00 equiv) and 4-dimethylaminopyridine (4-DMAP, 16 mg, 0.13 mmol, 0.30 equiv). The reaction was stirred at room temperature overnight. The mixture was then concentrated under reduced pressure and purified by column chromatography (silica gel, eluent: 0 to 35% EtOAc/hexanes, v/v) to afford carbamate (±)-**13** (83%).

¹H NMR (500 MHz, DMSO-d_6) δ: 7.51–7.30 (m, 2H), 6.99 (d, *J* = 7.9 Hz, 2H), 4.36–4.00 (m, 2H), 3.87 (s, 1H), 2.72 (s, 2H), 2.57 (s, 1H), 2.25 (ddd, *J* = 10.1, 6.7, 3.6 Hz, 1H), 1.80 (s, 3H), 1.46 (d, *J* = 5.6 Hz, 18H), 1.36–1.22 (m, 2H). **¹³C NMR** (126 MHz, DMSO-d_6) δ: 154.54, 150.67, 147.65, 138.61, 131.44, 127.93, 120.15, 84.89, 79.73, 58.39, 37.24, 28.37, 27.38, 16.91. **HRMS** (ESI) (*m/z*) calc'd for $C_{24}H_{35}BrN_2O_4$ [M+H]⁺: 495.1853, 496.1885, 497.1836, 498.1866, 499.1894 found 495.1846, 496.1883, 497.1826, 498.1862, 499.1895.

(±)-N-(2-(4-(benzyloxy)phenyl)cyclopropyl)piperidin-4-amine AW1 (2):
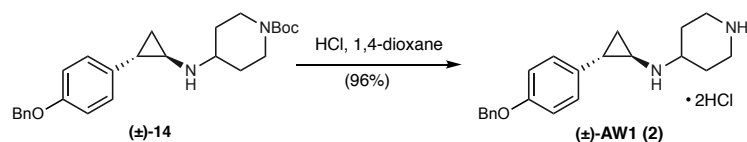
Carbamate intermediate (±)-**14** was synthesized as described for (±)-**12** and the NMR spectra match those reported in the literature.[4] Briefly, carbamate (±)-**14** (42 mg, 0.10 mmol, 1.00 equiv) was dissolved in 1,4-dioxane (50 μL) and treated with 4.0 M HCl solution in 1,4-dioxane (100 μL, 0.40 mmol, 4.00 equiv). The reaction mixture was stirred at room temperature until completion before being concentrated under reduced pressure. The residue was dissolved in water/MeOH (10 mL) and subsequently washed with hexanes/Et$_2$O (2 × 20 mL, 1:1, v/v). The water/MeOH layer was concentrated under reduced pressure to afford (±)-**AW1**•HCl as a white solid (96% yield).

^{1}H NMR (500 MHz, DMSO-d_6) δ: 10.08 (d, J = 14.6 Hz, 2H), 9.39 (d, J = 10.7 Hz, 1H), 9.00 (d, J = 10.7 Hz, 1H), 7.43 (d, J = 7.0 Hz, 2H), 7.38 (t, J = 7.5 Hz, 2H), 7.34–7.29 (m, 1H), 7.15–7.08 (m, 2H), 6.97–6.91 (m, 2H), 5.08 (s, 2H), 3.49 – 3.32 (m, 4H), 2.92 (q, J = 12.5, 11.9 Hz, 2H), 2.83 (p, J = 4.2 Hz, 1H), 2.54 (ddd, J = 10.1, 6.3, 3.5 Hz, 1H), 2.25 (d, J = 13.2 Hz, 2H), 1.93 (dt, J = 12.2, 4.6 Hz, 2H), 1.55 (ddd, J = 10.2, 5.9, 4.2 Hz, 1H), 1.26–1.16 (m, 1H). **^{13}C NMR** (126 MHz, DMSO-d_6) δ: 157.03, 137.11, 130.69, 128.39, 127.75, 127.61, 127.55, 114.81, 69.15, 52.56, 41.23, 34.84, 24.95, 24.84, 19.79, 12.12. **HRMS** (ESI) (*m/z*) calc'd for C$_{21}$H$_{26}$N$_2$O [M+H]$^+$: 323.2118, found 323.2117.

(±)-N-(2-([1,1'-biphenyl]-4-yl)cyclopropyl)piperidin-4-amine AW2 (3):
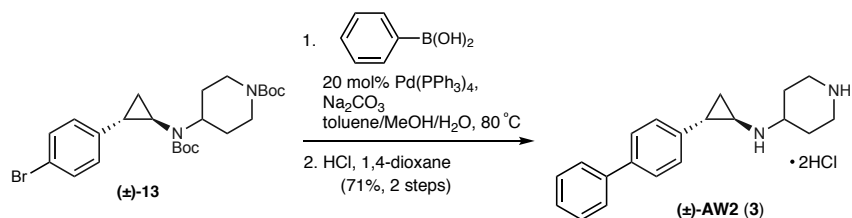
According to a modified literature procedure,[136] tetrakis(triphenylphosphine)palladium (14 mg, 0.01 mmol, 0.20 equiv), phenylboronic acid (30 mg, 0.24 mmol, 4.00 equiv), and Na_2CO_3 (13 mg, 0.12 mmol, 2.00 equiv) were added to a solution of the aryl bromide (±)-**13** (30 mg, 0.06 mmol, 1.00 equiv) in a degassed mixture of toluene/MeOH/water (540 μL/120 μL/10 μL, 80/18/2, v/v/v) under a nitrogen atmosphere. The reaction mixture was heated to 80 °C for 18 h. After cooling to room temperature, the mixture was diluted with EtOAc, filtered through a plug of celite and concentrated under reduced pressure. The residue was purified by column chromatography (silica gel, eluent: 0 to 35% EtOAc/hexanes, v/v). The intermediate was subsequently dissolved in 1,4-dioxane (21 μL, 2.0 M) and 4.0 M HCl solution in 1,4-dioxane was added (43 μL, 0.17 mmol, 4.00 equiv). The reaction mixture was stirred at room temperature until completion as monitored by TLC and then concentrated under reduced pressure. The residue was dissolved in water/MeOH (5 mL) and subsequently washed with 1:1 hexanes/Et_2O (2 × 10 mL). The water/MeOH layer was concentrated under reduced pressure to afford (±)-**AW2**•HCl as a yellow solid (71% yield, two steps).

¹H NMR (500 MHz, DMSO-d_6) δ: 10.05 (d, J = 15.2 Hz, 2H), 9.16 (s, 1H), 8.92 (t, J = 11.3 Hz, 1H), 7.70–7.54 (m, 4H), 7.52–7.39 (m, 2H), 7.40–7.32 (m, 1H), 7.32–7.26 (m, 2H), 3.61–3.47 (m, 3H), 3.09–2.85 (m, 3H), 2.63 (ddd, J = 10.0, 6.3, 3.5 Hz, 1H), 2.31–2.18 (m, 2H), 2.01–1.85 (m, 2H), 1.63 (ddd, J = 10.4, 6.2, 4.5 Hz, 1H), 1.34 (dt, J = 7.8, 6.3 Hz, 1H). **¹³C NMR** (126 MHz, DMSO-d_6) δ: 139.71, 138.39, 137.95, 128.91, 127.33, 126.95, 126.71, 126.47, 52.63, 41.34,

35.11, 25.04, 20.25, 12.69. **HRMS** (ESI) (*m/z*) calc'd for $C_{20}H_{24}N_2$ [M+H]$^+$: 293.2012, found 293.2011.

(±)-*N*-(2-([1,1':4',1''-terphenyl]-4-yl)cyclopropyl)piperidin-4-amine AW3 (4):
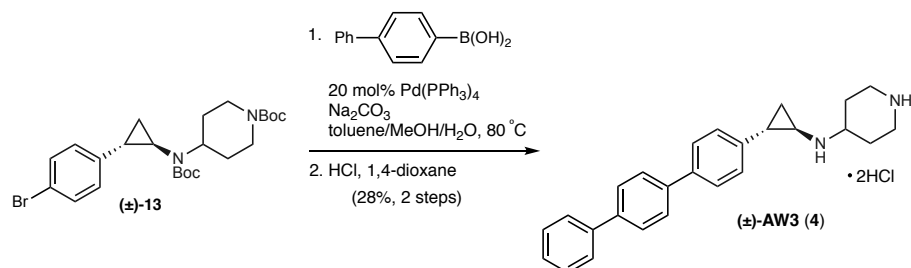
(±)-**AW3** was prepared according to the general protocol for (±)-**AW2**. (±)-**AW3**•HCl was isolated as a yellow solid (28% yield, two steps).

^{1}H NMR (500 MHz, DMSO-d_6) δ: 9.94 (s, 2H), 9.07–8.97 (m, 1H), 8.84 (d, *J* = 11.5 Hz, 1H), 7.76 (s, 4H), 7.74–7.66 (m, 4H), 7.49 (t, *J* = 7.6 Hz, 2H), 7.38 (t, *J* = 7.4 Hz, 1H), 7.31 (d, *J* = 8.0 Hz, 2H), 3.58–3.43 (m, 3H)[9], 3.03 (td, *J* = 8.3, 4.1 Hz, 1H), 2.94 (q, *J* = 12.0 Hz, 2H), 2.62 (ddd, *J* = 10.4, 6.2, 3.4 Hz, 1H), 2.25 (d, *J* = 13.3 Hz, 2H), 1.97–1.84 (m, 2H), 1.62 (dt, *J* = 10.4, 5.5 Hz, 1H), 1.36 (q, *J* = 6.7 Hz, 1H). **^{13}C NMR** (126 MHz, DMSO-d_6) δ: 139.08, 138.66, 138.06, 137.81, 136.66, 128.98, 127.18, 127.16, 126.97, 126.59, 126.53, 52.65, 41.40, 35.13, 24.98, 20.31, 12.25. **HRMS** (ESI) (*m/z*) calc'd for $C_{26}H_{28}N_2$ [M+H]$^+$: 369.2325, found 369.2325.

(±)-N-(2-([1,1':3',1''-terphenyl]-4-yl)cyclopropyl)piperidin-4-amine AW4 (5):
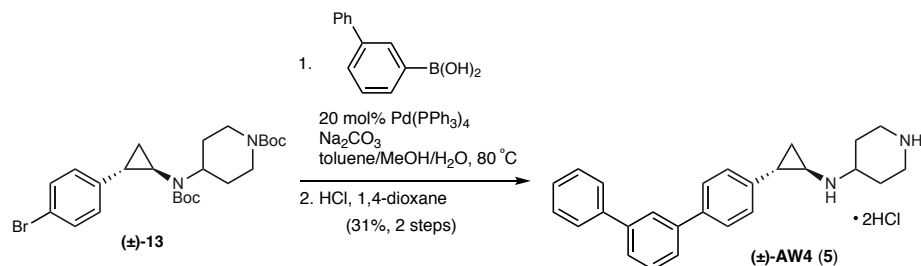
(±)-**AW4** was prepared according to the general protocol for (±)-**AW2**. (±)-**AW4**•HCl was isolated as a yellow solid (31% yield, two steps).

^{1}H NMR (500 MHz, DMSO-d_6) δ: 9.96 (d, J = 11.3 Hz, 2H), 9.04 (d, J = 10.7 Hz, 1H), 8.85 (d, J = 11.2 Hz, 1H), 7.87 (d, J = 1.9 Hz, 1H), 7.73 (dd, J = 18.2, 7.9 Hz, 4H), 7.68–7.60 (m, 2H), 7.55 (t, J = 7.7 Hz, 1H), 7.49 (t, J = 7.6 Hz, 2H), 7.40 (t, J = 7.3 Hz, 1H), 7.30 (d, J = 8.1 Hz, 2H), 3.59–3.41 (m, 3H)[9], 3.07–2.88 (m, 3H), 2.63 (ddd, J = 10.1, 6.5, 3.6 Hz, 1H), 2.31–2.20 (m, 2H), 1.98–1.85 (m, 2H), 1.62 (dt, J = 10.5, 5.5 Hz, 1H), 1.36 (q, J = 6.7 Hz, 1H). **^{13}C NMR** (126 MHz, DMSO-d_6) δ: 140.94, 140.45, 140.14, 138.38, 138.12, 129.58, 128.94, 127.59, 126.96, 125.82, 125.67, 124.94, 52.67, 41.41, 35.17, 25.09, 20.32, 12.77. **HRMS** (ESI) (*m/z*) calc'd for $C_{26}H_{28}N_2$ [M+H]$^+$: 369.2325, found 369.2323.

(±)-N-(2-(4-(naphthalen-1-yl)phenyl)cyclopropyl)piperidin-4-amine AW5 (6):
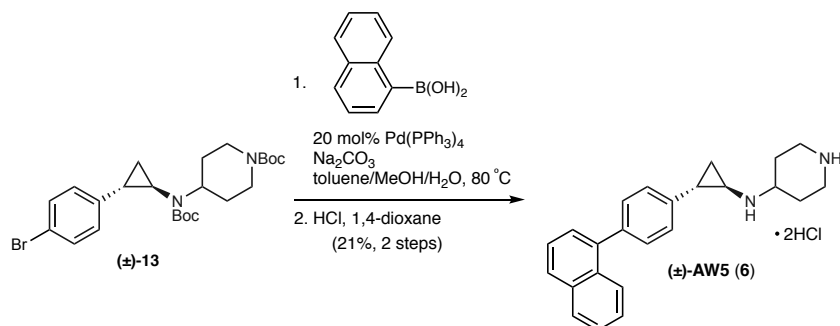
(±)-**AW5** was prepared according to the general protocol for (±)-**AW2**. (±)-**AW5**•HCl was isolated as a yellow solid (21% yield, two steps).

^{1}H NMR (500 MHz, DMSO-d_6) δ: 10.12 (s, 2H), 9.29 (s, 1H), 8.98 (s, 1H), 8.21 (d, J = 1.8 Hz, 1H), 7.99 (dd, J = 7.9, 5.2 Hz, 2H), 7.94 (dd, J = 7.6, 1.7 Hz, 1H), 7.84 (dd, J = 8.6, 1.9 Hz, 1H), 7.81–7.74 (m, 2H), 7.59–7.47 (m, 2H), 7.39–7.29 (m, 2H), 3.57–3.47 (m, 1H), 3.39 (d, J = 8.9 Hz, 2H)[9], 3.05–2.87 (m, 3H), 2.67 (dt, J = 10.2, 4.8 Hz, 1H), 2.32–2.22 (m, 2H), 1.96 (dq, J = 21.8, 10.3, 9.4 Hz, 2H), 1.66 (dt, J = 10.4, 5.5 Hz, 1H), 1.40–1.33 (m, 1H). **^{13}C NMR** (126 MHz, DMSO-d_6) δ: 138.25, 138.10, 137.02, 133.30, 132.16, 128.44, 128.12, 127.46, 127.03, 126.98, 126.39, 126.06, 124.92, 52.62, 41.30, 35.29, 25.07, 20.35, 12.80. **HRMS** (ESI) (*m/z*) calc'd for $C_{24}H_{26}N_2$ [M+H]$^+$: 343.2169, found 343.2169.

5-((3-aminopropyl)carbamoyl)-2-(6-(dimethylamino)-3-(dimethyliminio)-3H-xanthen-9-yl)benzoate (15):
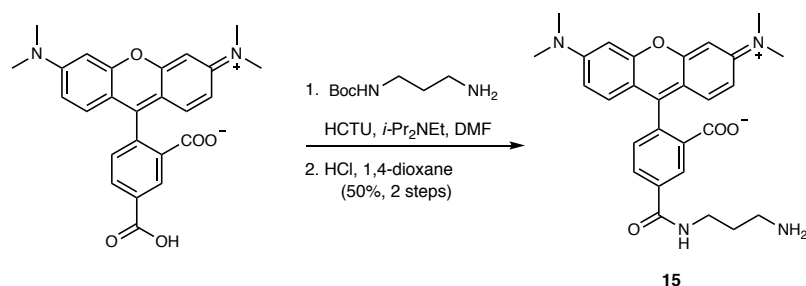
According to a modified literature procedure,[137] to a solution of 5(6)-TAMRA (40 mg, 0.09 mmol, 1.00 equiv) in DMF (1.0 mL) was added N-Boc-1,3-propanediamine (16 μL, 0.09 mmol, 1.00 equiv), HCTU (0.09 mmol, 1.10 equiv) and i-Pr$_2$NEt (32 μL, 0.19 mmol, 2.00 equiv). The reaction was stirred at room temperature overnight and subsequently concentrated under reduced pressure. The residue was then dissolved in 4.0 M HCl in 1,4-dioxane (180 μL). The reaction mixture was stirred at room temperature until completion as monitored by TLC before being concentrated under reduced pressure. The resultant residue was purified by C18 reverse phase column chromatography using a Büchi C18 40 μm reverse phase 12 g cartridge (eluent: 0 to 100 % water/MeCN with 0.1% TFA, v/v) to afford TAMRA label **15** (56% yield, two steps).

1H NMR (500 MHz, CD$_3$OD, major isomer reported) δ: 8.80 (d, J = 1.8 Hz, 1H), 8.29 (dd, J = 31.6, 8.1, 1.8 Hz, 1H), 7.55 (d, J = 7.9 Hz, 1H), 7.13-6.99 (dd, J = 9.5, 5.1 Hz, 6H), 3.50 (t, J = 6.7 Hz, 2H), 3.06 (t, J = 7.2 Hz, 2H), 2.04 (q, J = 7.0 Hz, 2H).

TAMRA Labeled GFI1B Peptide Synthesis and Purification:

GFI1B wt (PRSFLVKSK) and GFI1B F5A (PRSALVKSK) peptides were synthesized by manual fluorenylmethyloxycarbonyl (Fmoc) solid phase peptide synthesis using 2-chlorotrityl chloride resin (AAPPTec) at 65 °C. The resin (112 mg, 0.89 mmol/g) was suspended and allowed to swell in dry CH_2Cl_2 for 1 h. After removing CH_2Cl_2, a solution of Fmoc-Lys (Boc)-OH (70 mg, 0.10 mmol, 1.50 equiv) in 4.0 mL of CH_2Cl_2 and 2,4,6-collidine (10 equiv) was added to the resin and the mixture was agitated for 2 h at room temperature. The unreacted resin was capped with a solution of CH_2Cl_2/MeOH/i-Pr$_2$NEt (17:2:1, 5.0 mL) for 30 min at room temperature followed by washing the resin twice with DMF. Fmoc deprotection was achieved with 5 % piperazine in DMF for 5 min followed by coupling of the amino acid in the presence of 2-(6-Chloro-1H-benzotriazole-1-yl)-1,1,3,3-tetramethylaminium hexafluorophosphate (HCTU) and i-Pr$_2$NEt (AA:HCTU:i-Pr$_2$NEt:resin; 3:2.8:6:1) for 7 min. Deprotection of the peptide was achieved by agitation at room temperature in a 1:4 hexafluoroisopropanol/CH_2Cl_2 solution (5.0 mL). The suspension was filtered, and the resin was washed with additional 1:4 hexafluoroisopropanol/CH_2Cl_2 solution (1:4, 5.0 mL), followed by CH_2Cl_2 (2 × 10 mL). The combined filtrates were concentrated under reduced pressure. The identity of the peptide was confirmed by MALDI TOF/ TOF mass spectrometry (Bruker ultrafleXtreme).

HCTU (20.6 mg, 0.05 mmol, 2.00 equiv), hydroxybenzotriazole (7.63 mg, 0.05 mmol, 2.00 equiv), and i-Pr$_2$NEt (21.7 μL, 0.12 mmol, 5.00 equiv) were added sequentially to a solution of peptide (43 mg, 0.02 mmol, 1.00 equiv) in DMF (5.0 mL) at room temperature. After stirring for 30 min, the 5(6)-TAMRA label described above (24 mg, 0.05 mmol, 2.00 equiv) was added and the reaction was stirred for 12 h at room temperature. The reaction mixture was subsequently concentrated under reduced pressure and purified by column chromatography (silica gel, eluent: 18% MeOH, 2% NH$_4$OH in chloroform, v/v). All fractions containing the desired product were

combined and concentrated under reduced pressure. The purified product was subjected to global deprotection by treatment with a solution consisting of 95% trifluoroacetic acid (TFA), 2.5% H_2O and 2.5% triisopropyl silane (TIPS) for 2 h at room temperature. The crude peptides were precipitated with cold MTBE, washed three times with cold MTBE, and dried under nitrogen. The peptides were purified on a preparative reverse phase high performance liquid chromatography system (Agilent) with a C18 column using a linear gradient of solvent A (0.1% TFA in Millipore H_2O) and solvent B (90% CH_3CN, 9.9% H_2O, 0.1% TFA). The identities of the purified peptides were assessed by MALDI-TOF/TOF mass spectrometry. The purity of the peptides was assessed on an Agilent HPLC with an analytical Agilent-C18 column (4.6 mm × 150 mm) using a linear gradient of solvent A and solvent B.

MALDI TOF/TOF MS (*m/z*) calc'd for GFI1B wt (PRSFLVKSK) [M+H]$^+$: 1529.896 found 1530.02. calc'd for GFI1B F5A (PRSALVKSK) [M+H]$^+$: 1453.83 found 1453.87.

Chapter 3: The disordered N-terminus of LSD1 tunes enhancer commissioning

This chapter is adapted from a manuscript draft with the following contributors:

<u>Amanda L. Waterbury</u>[*], Hui Si Kwok[*], Ceejay Lee[*], Domenic Narducci, Allyson M. Freedy, Cindy Su, Andrew Reiter, Kwangwoon Lee, Samuel M. Hoenig, Michael E. Vinyard, Zhipeng A. Wang, Philip A. Cole, Anders S. Hansen, Steven A. Carr, Malvina Papanastasiou, & Brian B. Liau

Contributions:

A.L.W., H.S.K., C.L., and B.B.L. conceived the study and designed experiments. A.L.W. performed cell experiments, protein purification, biochemical assays, and microscopy experiments. A.L.W. and D.N. analyzed imaging data, with input from A.S.H.. H.S.K. performed cell experiments and genomics experiments. C.L. and H.S.K. analyzed genomics data. H.S.K., A.R., M.P. performed and analyzed co-IP/MS experiments. M.P. performed and analyzed HDX-MS experiments, with input from S.A.C.. A.L.W., K.L. and Z.A.W purified and assayed LSD1-HDAC-CoREST complexes, with input from P.A.C.. C.S., A.M.F., S.M.H., M.E.V., and Z.A.W. provided experimental support. A.L.W., H.S.K., C.L., and B.B.L. edited and wrote the manuscript. B.B.L. held overall responsibility for the study.

3.1. Abstract

Transcriptional coregulators are recruited by transcription factors (TFs) to regulate gene expression. These coregulators can directly bind TFs via high-affinity motifs, driving TF-specific gene expression programs. Alternatively, coregulators can be broadly recruited into TF hubs, in part assembled via weak multivalent interactions mediated by the intrinsically disordered regions (IDRs) of TFs and coregulators. LSD1 is a transcriptional corepressor that is recruited by specific TFs but also pervasively binds active cis-regulatory regions to shape and decommission enhancer elements during development. Yet despite these essential functions, the factors and mechanisms that control LSD1 enhancer decommissioning remain unclear. Here we report that the intrinsically disordered region (IDR) of LSD1 modulates its interactions with master hematopoietic transcription factors (TFs) to control enhancer commissioning and acute myeloid leukemia (AML) differentiation. Through the identification of drug suppressor mutations, we show that LSD1 inhibitors reprogram LSD1-TF interactions to drive AML differentiation in an IDR-dependent manner (**Figure 3.1**). Counterintuitively, rather than directly driving weak IDR-IDR interactions, the LSD1 IDR blocks interactions with TF IDRs by driving hub demixing. Altogether, our studies reveal new mechanisms by which IDRs can tune corepressor-TF interactions to control cis-regulatory elements, cell fate, and drug action.
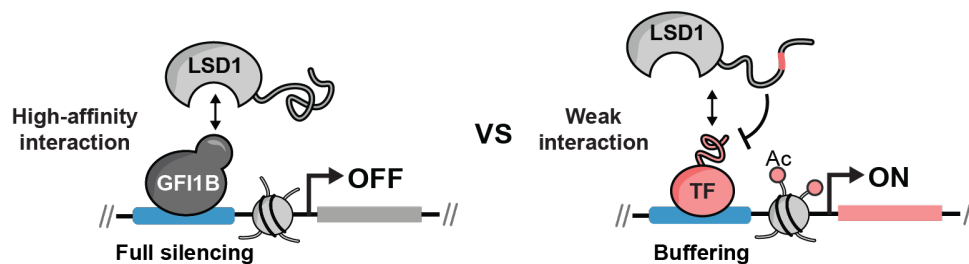
Figure 3.1. Graphical abstract summary of this chapter.

3.2. Introduction

The control of cis-regulatory elements (CREs) by chromatin complexes is essential for gene regulation and development.[138–141] Notably, the histone H3 lysine 4 (H3K4) demethylase LSD1 is a master regulator of CREs and hematopoiesis, acting as a key enzymatic component of the LSD1-CoREST-HDAC1/2 corepressor complex.[35,36,40,142,143] LSD1 can be directly recruited to CREs by various TFs, including GFI1B, which binds the LSD1 catalytic site through a specific high-affinity interaction.[42,44] Recruitment of LSD1 by GFI1B is essential for the proper formation of erythrocytes and megakaryocytes, opposing the activity of myeloid TFs such as C/EBPα and PU.1. Likewise, this LSD1-mediated TF-antagonism can also maintain the AML dedifferentiated state, and efforts to target LSD1 to relieve this differentiation block have shown promise in the clinic.[35,70,93–95] Although LSD1 small molecule inhibitors were originally developed to block LSD1 enzyme activity, their antiproliferative effects result from disruption of the LSD1-GFI1B interaction, de-repression of GFI1B target genes, activation of CREs controlled by C/EBPα and PU.1, and subsequent induction of a myeloid gene expression program (**Figure 3.2**).[69,70,93,113] Thus, beyond its prototypical function as a histone demethylase, scaffolding interactions of LSD1 with TFs have critical roles in hematopoiesis and are the relevant targets of LSD1 inhibitors in clinical development for oncology and myelofibrosis.
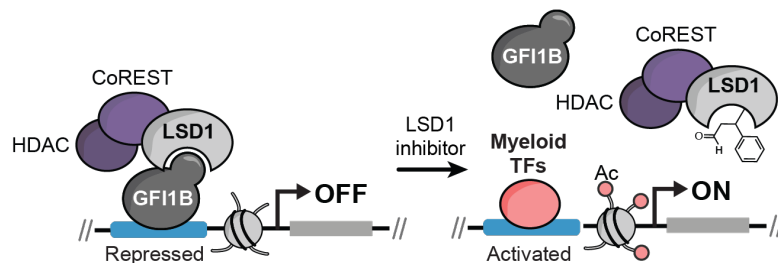
Figure 3.2. LSD1 inhibitors dissociate the LSD1-GFI1B complex.

Schematic illustrating the downstream mechanism of LSD1 inhibitors whereby GSK-LSD1 induces displacement of the LSD1-GFI1B complex along with associated repressor proteins, CoREST and HDAC, from enhancers, which allows myeloid TF-mediated activation of genes involved in AML differentiation.

Aside from maintaining sustained repression of specific TF target-sites (e.g., GFI1B), LSD1 pervasively binds active CREs where it may be poised to tune their activity more broadly during differentiation.[40] Supporting this role, pioneering studies demonstrated that LSD1 is necessary for proper differentiation of embryonic stem cells and hematopoietic progenitor cells by mediating global enhancer decommissioning.[38,39,44,46,47] Consistent with this model, genetic and chemical perturbation of LSD1 leads to increased chromatin accessibility and enhancer activation, which in part underlie their anti-proliferative effects in AML.[69,70,93] However, as opposed to GFI1B target-sites, we lack understanding of the factors that control LSD1 enhancer decommissioning and recruitment to active CREs. To address these questions, here we investigate the mechanisms that control CRE commissioning and uncover a broad role of the LSD1 disordered N-terminus in controlling LSD1-TF interactions.

3.3. Deletions in the LSD1 IDR confer drug resistance and alter condensate properties in vitro

We previously employed CRISPR-suppressor scanning to identify single-guide RNAs (sgRNAs) targeting LSD1 that can promote drug resistance to LSD1 inhibitors (**Figure 3.3a-f**).[69] By characterizing mutations created by the top-enriched sgRNAs targeting the LSD1 active site, we showed that the LSD1-GFI1B interaction is critical for select AML cell lines. Beyond these findings, we noted that several drug-enriched sgRNAs also targeted the LSD1 N-terminus, which is intrinsically disordered (**Figure 3.3b**).[92] Consequently, we considered whether mutations in the LSD1 N-terminal IDR may be mechanistically involved in drug action and AML differentiation.
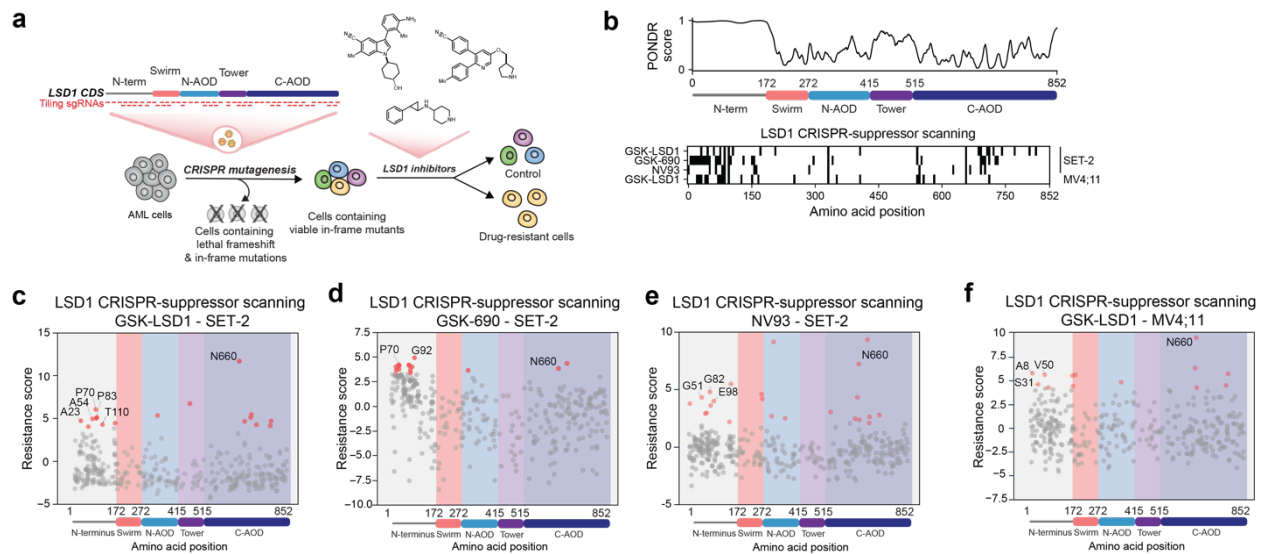
Figure 3.3. CRISPR-suppressor scanning of LSD1 enriches for drug-resistance conferring mutations in the disordered N-terminus of LSD1.

a, Schematic of CRISPR-suppressor scanning workflow to profile SAR of LSD1 inhibitors. **b,** PONDR score (top) of LSD1 along the *LSD1* CDS. sgRNA enrichment binned every 5 amino acids along the *LSD1* CDS where each row corresponds to LSD1 inhibitor treatment for either SET-2 (GSK-LSD1, GSK-690 and NV93) and MV4;11 (GSK-LSD1). **c-f,** Scatter plot showing log2(fold-change sgRNA enrichment) under drug treatment at the indicated week versus week 0 normalized against functionally neutral genome-targeting control sgRNAs. The sgRNAs are arrayed on the *x* axis by the *LSD1* CDS. The specific conditions are as

To corroborate the CRISPR scan, we transduced SET-2 cells with individual sgRNAs targeting the LSD1 IDR, treated with the covalent LSD1 inhibitor GSK-LSD1, and expanded drug-resistant clones. Two clonal cell lines were isolated and harbor LSD1 in-frame deletions: SET-2 LSD1(NΔ1) contains LSD1 G46_P52del and SET-2 LSD1(NΔ2) contains LSD1 P77_G86del (**Figure 3.4a-c**). Both mutations remove short hydrophobic stretches enriched in glycine and proline residues (**Figure 3.4a,d**). In comparison to wild-type (WT), these LSD1-mutant cell lines were resistant to GSK-LSD1 and GSK690, a non-covalent inhibitor, in cell growth assays and exhibited reduced expression of CD86 — a myeloid differentiation marker upregulated upon LSD1 inhibition (**Figure 3.4e,f**).[94,97,102,144] Taken together, these data show that deletions in the LSD1 IDR can impair AML differentiation.
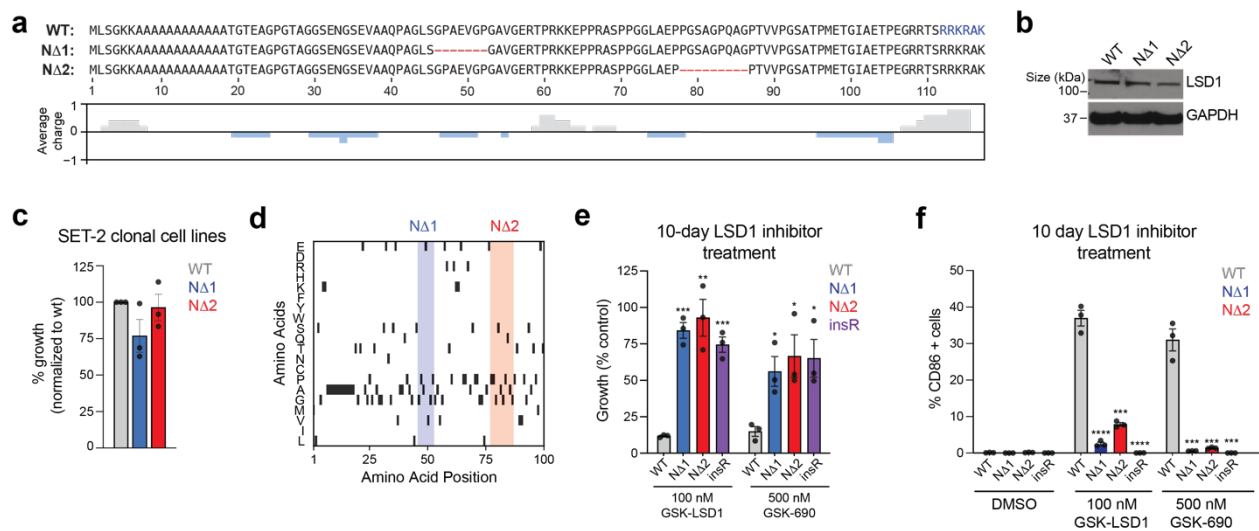
Figure 3.4. Deletion mutations in the LSD1 IDR confer drug resistance.

a, Schematic of LSD1 N-term mutations along the N-term LSD1 CDS where only the first 116 residues are shown for clarity (above). The average electrostatic charge is plotted below. **b,** Immunoblot showing that LSD1 mutants are expressed in drug-resistant cell lines at comparable levels. **c,** Relative cell growth of

Figure 3.4 (continued) SET-2 cell lines normalized to WT control. Mean ± s.e.m. across three replicates is shown. **d,** Amino acid composition of the LSD1 IDR. Each row represents information for a single amino acid as indicated by the single letter amino acid code abbreviation (left) and the length of the row corresponds to the length of the LSD1 IDR. Black bars represent the occurrence of the indicated amino acid at that position in the LSD1 IDR. Location of the NΔ1 and NΔ2 mutations are shaded in blue and red, respectively. **e,** LSD1 N-term mutations confer growth resistance in the presence of GSK-LSD1. Bar plot shows relative growth (y-axis) of WT, LSD1(NΔ1), LSD1(NΔ2) and LSD1(L659_N660insR) SET-2 cell lines after 10 d of GSK-LSD1 (100 nM) or GSK-690 (500 nM) treatment in comparison to vehicle control. LSD1(L659_N660insR) is a previously validated drug resistance mutation[69]. Data represent mean ± s.d. across three replicates. **f,** LSD1 N-term mutations impair myeloid differentiation as assessed by CD86 expression in the presence of LSD1 inhibitors. Bar plot shows the fraction of CD86+ cells (*y*-axis) in WT, LSD1(NΔ1), LSD1(NΔ2) and LSD1(L659_N660insR) SET-2 cell lines as assessed by flow cytometry after 10 d of GSK-LSD1 (100 nM) or GSK-690 (500 nM) treatment in comparison to vehicle control treatment. Data represent mean ± s.d. across three replicates.

We first considered whether the IDR mutants may block inhibitor binding or alter the LSD1-CoREST-HDAC1/2 complex.[54] Purified IDR mutants were active in H3K4me2 demethylation assays and inhibited at similar concentrations of GSK-LSD1 and GSK-690, demonstrating that the mutations do not abrogate demethylase activity or drug binding (**Figure 3.5a-c**). Furthermore, LSD1-CoREST-HDAC1 (LHC) complexes comprising wild-type or LSD1 (NΔ2) showed similar demethylation and deacetylase activities using H3K4me2 and H3K27ac nucleosome substrates, respectively, indicating that the IDR mutation does not affect the enzyme activity of the complex or the ability of HDAC1 to associate with LSD1-CoREST (**Figure 3.5d,e**).

Figure 3.5. Biochemical characterization of drug resistant LSD1 IDR mutants.

a, Dose response curve of recombinant LSD1 protein treated with GSK-LSD1 or **b,** GSK-690. **c,** Summary table of IC50 values for dose response curves with GSK-LSD1 and GSK-690. **d,** Demethylase activity of WT and NΔ2 LHC with 185 bp H3K4me2 nucleosome substrate was assessed at the indicated time points. Demethylase activity in the presence of 10 μM GSK-LSD1 was also assessed by immunoblot with NΔ2 LHC complex, however, no significant differences in demethylase activity were observed. **e,** The deacetylation of H3K27ac nucleosomes by WT and NΔ2 LHC was assessed by immunoblot.

Given that these mutations occur in the IDR of LSD1, we considered whether they may alter the biophysical properties of LSD1. IDRs have been increasingly implicated as drivers of condensate formation through participation in extensive and dynamic multivalent interactions.[24,25,27] Therefore, we tested whether LSD1 may form condensates in vitro. Purified full-length LSD1 fused to mCherry (LSD1-mCherry) formed condensates in vitro, which have spherical shapes, freely diffuse, undergo fusion events, and their size scales with concentration of the protein and buffer ionic strength (**Figure 3.6a-g**). mCherry alone fails to form droplets under the same conditions (**Figure 3.6b**).[24,25,145] Strikingly, the NΔ1 and NΔ2 deletions enhanced condensate formation (**Figure 3.6a,c-d**). By contrast, removal of the IDR (Δ1-150) led to the formation of insoluble aggregates (**Figure 3.6a**), although complexation with truncated CoREST(Δ1-305) solubilized the protein complex into condensates (**Figure 3.6h**).

Figure 3.6. LSD1 forms condensates in vitro and the IDR mutations enhance droplet formation.

a, Representative images of purified LSD1-mCherry variants in droplet formation buffer at the indicated protein concentrations. LSD1(Δ150) forms aggregates under the same conditions. **b,** Representation image of two LSD1-mCherry droplets fusing, indicative of liquid-like properties (top) while mCherry fails to form droplets altogether (bottom). **c,** Phase diagram of LSD1-mCherry IDR variants demonstrating that droplets are formed in a concentration dependent manner and LSD1(NΔ1) and LSD1(NΔ2) have increased propensity to form droplets. Mean fluorescence intensity of the droplets is shown. Data displayed as median ± s.d.. The center line represents the median; the box limits are the 25th and 75th percentiles; and the whiskers show the minimum to maximum values. *n* = 15 independent fields per condition. **d,** Phase diagram of LSD1-mCherry fusion proteins plotted by average area of the droplets across different protein concentrations as described in (**c**). n = 15 independent fields of view. **e,** Representative images of droplet formation by purified LSD1-mCherry proteins in the presence of varying salt concentrations. **f,** Phase diagram for varying buffer salt concentration of LSD1-mCherry fusion protein droplets across n = 15 independent fields of view where the mean fluorescence intensity of the droplets is shown and is plotted as

Figure 3.6 (continued) described in (**c**). **g,** Phase diagram for varying salt concentration of LSD1-mCherry fusion protein droplets across n = 15 independent fields of view where the average area of the droplets is shown and is plotted as described in (**c**). **h,** Representative images of droplet formation by purified LSD1-mCherry/CoREST(Δ305) variants in droplet formation buffer. **i,** Representative images of droplet formation by the LSD1 IDR (1-170) fused to mCherry in droplet formation buffer.

The LSD1 IDR (amino acids 1-170) alone only formed droplets at high protein concentrations and was more sensitive to the concentration of salt in the buffer, indicative of a weaker proclivity to form condensates compared to the full-length protein (**Figure 3.6i**). Therefore, these results suggests that the IDR mutations influence the complex's ability to engage in multivalent interactions, which may lead to altered protein-protein interactions in the cell.

3.4. LSD1 inhibitors function as transcription factor reprogrammers

Next, we investigated the effect of the LSD1 IDR mutations on the integrity of the LSD1-GFI1B complex by performing LSD1 co-IP in WT and LSD1(NΔ2) cells and quantitatively measuring interacting proteins by mass spectrometry (co-IP/MS).[146] As expected, GSK-LSD1 treatment did not significantly perturb the core LSD1-CoREST complex but disrupted the LSD1-GFI1B interaction in both wild-type and LSD1(NΔ2) cells (**Figure 3.7a-d**). Consistent with this, mapping LSD1, GFI1B, and H3K27ac in wild-type and LSD1(NΔ2) cells by ChIP-seq showed that GSK-LSD1 treatment induced LSD1-GFI1B dissociation and increased H3K27 acetylation in both cell lines (**Figure 3.7e**). These results suggest that the IDR-mutant cells no longer depend on GFI1B for cell survival altogether. In support, GFI1B knockout (KO) inhibited growth and induced differentiation of WT SET-2 cells but not IDR-mutant cells (**Figure 3.7f**). By contrast, LSD1 KO led to comparable growth defects in both wild-type and IDR-mutant cells, suggesting that mutant LSD1 is still required despite no longer depending on GFI1B for survival and might counterintuitively counteract differentiation to some degree. Collectively, our findings demonstrate that beyond its association with GFI1B, LSD1 has additional functions in downstream AML differentiation that are dependent on its IDR.
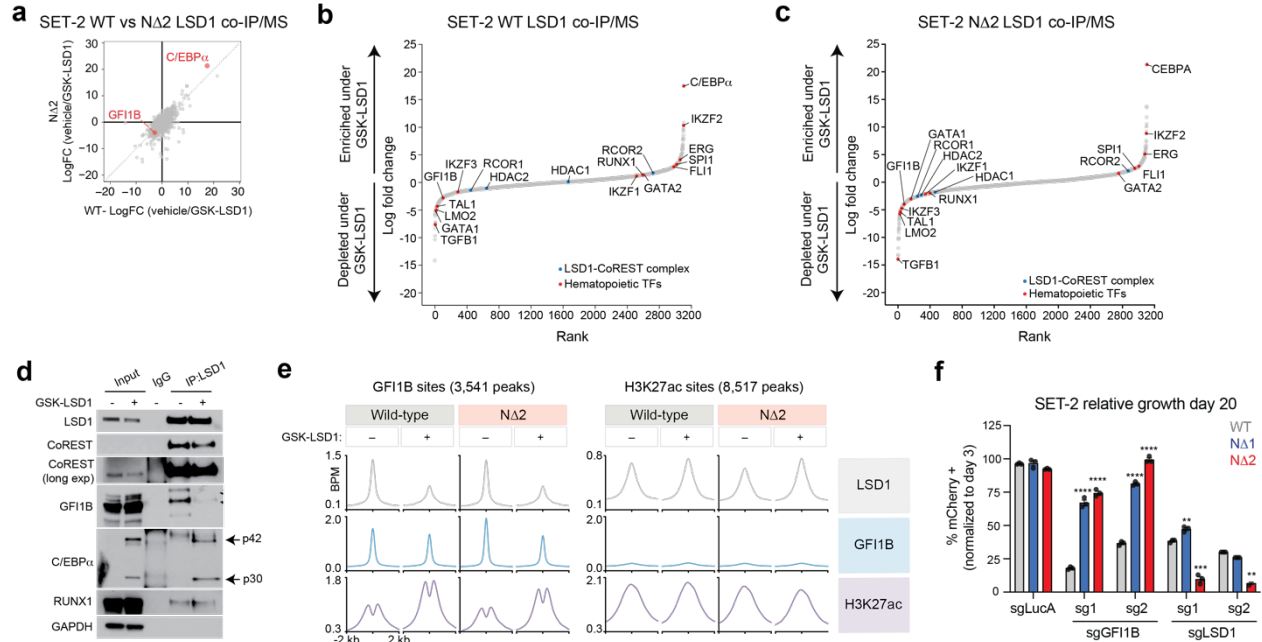
Figure 3.7. LSD1 inhibitors function as TF reprogrammers.

a, Scatter plot of the logFC of LSD1 interacting proteins quantified by CoIP-MS in vehicle over GSK-LSD1 for WT vs LSD1(NΔ2) after 2 d of GSK-LSD1 treatment. **b,** Co-IP MS data. LSD1-interacting proteins ranked by enrichment under GSK-LSD1 (100 nM) treatment after 2 d. **c,** Co-IP MS data. LSD1(NΔ2)-interacting proteins ranked by enrichment under GSK-LSD1 (100 nM) treatment after 2 d. **d,** Immunoblot showing LSD1 co-IPs with C/EBPα under GSK-LSD1 treatment. **e,** ChIP-seq profile plots of GFI1B sites that overlap with LSD1 for LSD1, GFI1B, and H3K27ac, showing that LSD1-GFI1B complex is still disrupted on chromatin, and enhancers are still partially activated; centered on GFI1B (left). Profile plots of non-GFI1B sites that overlap with H3K27ac, showing LSD1 re-localization to other sites; centered on LSD1 (right). BPM (bins per million). **f,** WT cells are sensitive to both GFI1B and LSD1 CRISPR KO but LSD1(NΔ1), LSD1(NΔ2) SET-2 cell line are only sensitive to LSD1 KO. Bar plot shows the remaining fraction of mCherry+ cells which is indicative of the cells transduced with the sgRNAs (*y*-axis) at 20 d relative to 3 d following lentiviral transduction of SET-2 cells. Data represent mean ± s.d. across three replicates.

Interestingly, beyond GFI1B, GSK-LSD1 treatment altered LSD1's interaction with several master hematopoietic TFs, many of which are essential for AML differentiation. Notably, the most

up-regulated interaction partner for both wild-type LSD1 and the N∆2 mutant is C/EBPα (**Figure 3.7a**), a myeloid master TF which is highly upregulated upon GSK-LSD1 treatment.[93] After initial LSD1-GFI1B disruption, C/EBPα and PU.1 act as key downstream effectors for LSD1 inhibitor-induced AML differentiation (**Figure 3.2**).[69,70,93] In support, siRNA knockdown of PU.1, C/EBPα, or both simultaneously blocked upregulation of CD86 by GSK-LSD1 in WT SET-2 cells (**Figure 3.8a**). Further comparison between wild-type and LSD1(N∆2) revealed modest enrichment of C/EBPα and associated coactivators amongst LSD1(N∆2)'s interaction partners (**Figure 3.8b-d**). Consistent with LSD1's increased interactions with C/EBPα, TFs, and transcriptional coactivators, loss of LSD1-GFI1B binding by ChIP-seq was accompanied by increased LSD1 levels at other C/EBPα and/or PU.1 sites (**Figure 3.7e**), reflecting a larger shift of LSD1 from GFI1B sites to active CREs (i.e., H3K27ac sites) upon addition of GSK-LSD1. To validate LSD1's association with these key myeloid master TFs in vitro, the IDRs of C/EBPα and PU.1 were fused to GFP and found to co-condense with LSD1-mCherry droplets (**Figure 3.8e-g**).[145] Droplets formed with N∆1 and N∆2 mutants displayed enhanced co-partitioning (**Figure 3.8e-g**). GFP alone failed to incorporate into LSD1-mCherry droplets (**Figure 3.8h**).
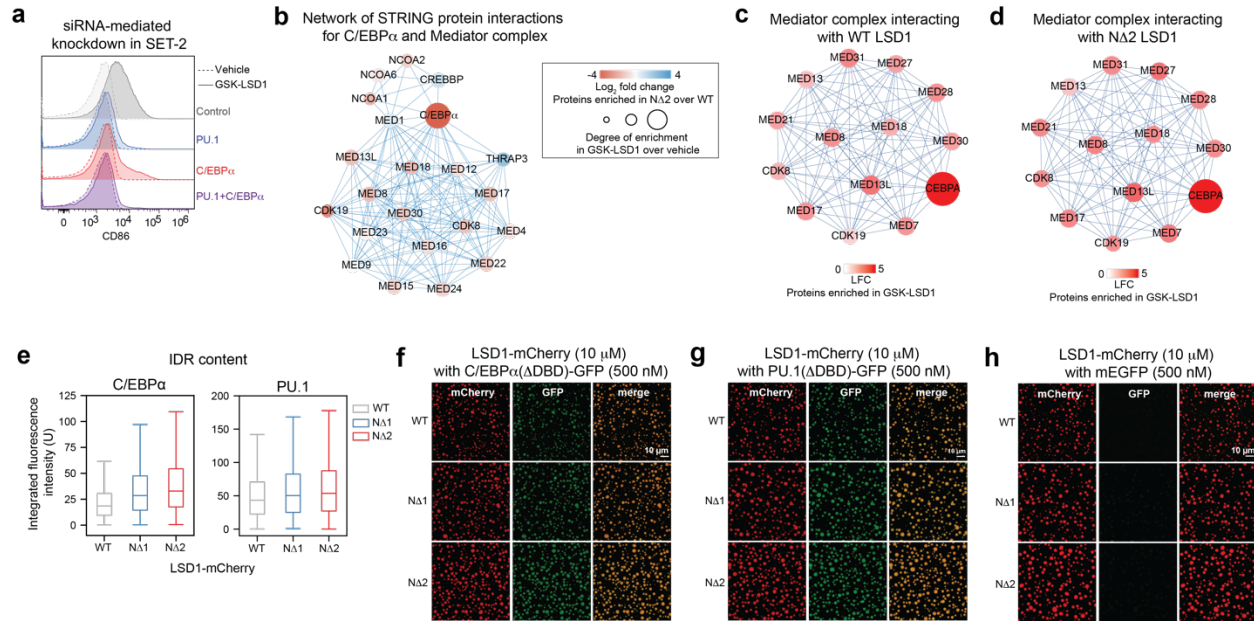
Figure 3.8. LSD1 interacts with myeloid TFs involved in downstream CRE activation.

a, siRNA-mediated knockdown of PU.1, C/EBPα or both factors in SET-2 impairs myeloid differentiation. Representative flow cytometry histogram shows the lack of CD86 expression in PU.1, C/EBPα, or PU.1 and C/EBPα double knockdown cells after 10 d of GSK-LSD1 (100 nM) treatment. The experiment was repeated twice, independently with similar results. **b,** Protein network figure showing enrichment of C/EBPα and mediator (interactors are a subset from GO-16455) as well as other proteins in LSD1(NΔ2) vs WT. **c,** Protein network figure showing enrichment of C/EBPα and mediator as well as other proteins in WT. **d,** Protein network figure showing enrichment C/EBPα and mediator as well as other proteins in LSD1(NΔ2). **e,** Quantification of C/EBPα(ΔDBD)-GFP integrated fluorescence intensity in LSD1-mCherry containing droplets where $P = 2.53 \times 10^{-104}$ for WT vs LSD1(NΔ1) and $P = 3.09 \times 10^{-188}$ for WT vs LSD1(NΔ2), Mann-Whitney U test (left). Quantification of PU.1(ΔDBD)-GFP integrated fluorescence intensity in LSD1-mCherry containing droplets where $P = 1.64 \times 10^{-12}$ for WT vs LSD1(NΔ1) and $P = 3.15 \times 10^{-24}$ for WT vs LSD1(NΔ2), Mann-Whitney U test. $n = 6$ independent fields per condition (right). Data displayed as median $\pm$ s.d.. **f,** Representative fluorescence images showing droplet formation of LSD1-mCherry mixed with C/EBPα(ΔDBD)-GFP in droplet formation buffer. **g,** Representative fluorescence images showing droplet

107

Figure 3.8 (continued) formation of LSD1-mCherry mixed with PU.1(ΔDBD)-GFP in droplet formation buffer. **h,** Fluorescence images of droplet formation by LSD1-mCherry variants mixed with mEGFP as a control in droplet formation buffer.

3.5. The LSD1 IDR inhibits TF hub interactions

While the LSD1-GFI1B interaction is structurally resolved, how LSD1 interacts with other hematopoietic TFs like C/EBPα is unclear and has yet to be described in the literature. Numerous studies have shown that TFs and coactivators can form heterotypic condensates — often driven by weak multivalent interactions promoted by IDRs of their constituents.[25,145,147] Although a corepressor, we posited that LSD1 might engage in similar multivalent interactions with TFs and that mutations in its IDR might alter them. To probe LSD1-TF interactions and the function of its IDR in a physiological context, we employed a LacO-LacI recruitment system in a U2OS cell line containing an integrated LacO array (U2OS 2-6-3) (**Figure 3.9a**).[145,148,149] We first created a homozygous U2OS 2-6-3 cell line that expresses endogenous LSD1 fused to monomeric enhanced GFP (LSD1-GFP). LSD1-GFP displayed a mainly diffuse nucleoplasmic distribution of LSD1, suggesting that it does not form large hubs in cells (**Figure 3.9b**).
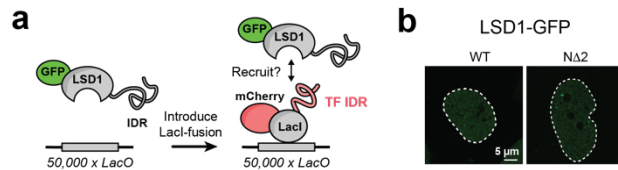
Figure 3.9. LacO-LacI recruitment assay to study LSD1-TF interactions in cells.

a, Schematic illustrating the LacI-LacO recruitment system to study LSD1-TF interactions in cells. **b,** Images of U2OS 2-6-3 LSD1-GFP knock-in cells.

We subsequently introduced the NΔ2 mutation and compared the recruitment of wild-type versus LSD1(NΔ2) to the LacO array by LacI-mCherry constructs fused to either (1) GFI1B or (2) C/EBPα lacking their respective DNA binding domains (GFI1B-LacI and C/EBPα-LacI, respectively) (**Figure 3.10a-d**). Both GFI1B-LacI and C/EBPα-LacI led to the formation of LacO associated hubs with significantly larger diameters compared to the LacI control, indicative of forming LacO associated biomolecular condensates (**Figre 3.10e**).[34] Both hubs could partition LSD1 while LacI-mCherry could not (**Figure 3.10a-d,f**). GFI1B-LacI hubs led to strong recruitment

of LSD1, which could be blocked by GSK-LSD1 treatment — consistent with the high-affinity LSD1-GFI1B interaction (**Figure 3.10a-b,f**). By contrast, C/EBPα-LacI hubs recruited less LSD1 (**Figure 3.10c,d**), indicating a weaker LSD1-C/EBPα interaction. In agreement, fluorescence recovery after photobleaching (FRAP) experiments revealed significantly greater residence time and total bound fraction for LSD1 at GFI1B-LacI hubs versus C/EBPα-LacI hubs (**Figure 3.10g,h**).[34,150] GSK-LSD1 treatment had minimal effects on LSD1 recruitment to C/EBPα-LacI hubs (**Figure 3.10d**).
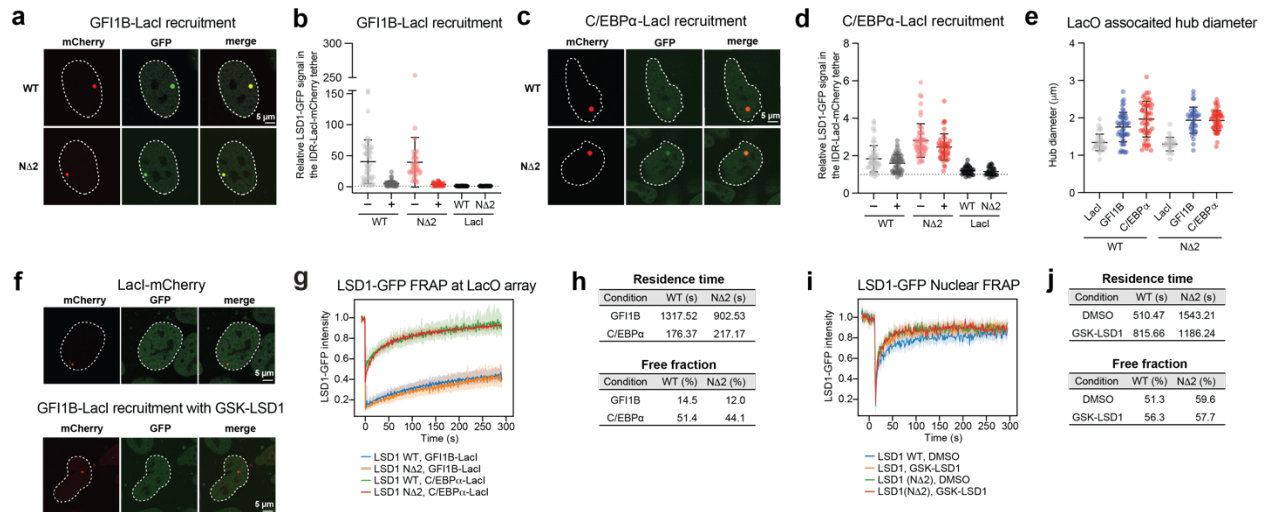
Figure 3.10. LSD1 recruitment to GFI1B versus C/EBPα hubs by live-cell microscopy.

a, Representative live-cell fluorescence images showing recruitment of LSD1-GFP to the LacO array by GFI1B(ΔZF)-LacI-mCherry for endogenous WT LSD1-GFP (top) and LSD1(NΔ2)-GFP (bottom). **b,** Quantification of the relative LSD1-GFP signal intensity in the LacO array after recruitment with GFI1B(ΔZF)-LacI-mCherry versus LacI-mCherry control. Each dot represents quantification from an individual cell. Shown are means ± s.e.m.; n = 53, 38, 46, 33, 34, 29 cells from left to right; P = 8.00 x 10⁻¹⁰ for LSD1 WT DMSO vs GSK-LSD1, P = 3.11 x 10⁻⁶ for LSD1(NΔ2) DMSO vs GSK-LSD1, P = 0.9080 for LSD1 WT DMSO vs LSD1(NΔ2) DMSO, P = 0.0004 for LSD1 WT GSK-LSD1 vs LSD1(NΔ2) GSK-LSD1; two-tailed unpaired t-test. **c,** Representative live-cell fluorescence images showing recruitment of LSD1-GFP to the LacO array by C/EBPα(ΔDBD)-LacI-mCherry for endogenous WT LSD1-GFP (top) and LSD1(NΔ2)-GFP (bottom). **d,** Quantification of the relative LSD1-GFP variant signal intensity in the LacO

Figure 3.10 (continued) array after recruitment with C/EBPα(ΔDBD)-LacI-mCherry versus LacI-mCherry control. Each dot represents quantification from an individual cell. Shown are means ± s.e.m.; n = 52, 59, 55, 44, 34, 29 cells from left to right; P = 0.0015 for LSD1 WT DMSO vs GSK-LSD1, P = 8.46 x 10^{-9} for LSD1 WT DMSO vs LSD1(NΔ2) DMSO, P = 3.13 x 10^{-5} for LSD1 WT DMSO vs LSD1(NΔ2) GSK-LSD1, P = 5.48 x 10^{-8} for LSD1 WT GSK-LSD1 vs LSD1(NΔ2) GSK-LSD1; two-tailed unpaired t-test. **e,** Diameter of the dots quantified in (**b,d**) for DMSO conditions of each respective TF or LacI control. Each dot represents quantification from an individual cell. Shown are means ± s.e.m.. **f,** Live-cell fluorescence images of endogenous-tagged LSD1-GFP where LacI-mCherry control fails to recruit LSD1-GFP to the LacO array (top). Diminished recruitment of endogenous LSD1-GFP to the LacO array by GFI1B(ΔZF)-LacI-mCherry is observed in the presence of 500 nM GSK-LSD1 (bottom). **g,** FRAP curves for LSD1 WT and LSD1(NΔ2)-GFP at the LacO array recruited by the specified TF. The intensity value within each region of interest was normalized and fit to a two-reaction model. **h,** Table summarizing the residence times (top) and free fractions (bottom) from modelling the FRAP curves in (**g**). **i,** FRAP curves for LSD1 WT and LSD1(NΔ2)-GFP where a 1.2 μm nuclear ROI was bleached and fluorescence recovery was monitored over time at the bleached spot relative to a nuclear background reference region. **j,** Table summarizing the residence times (top) and free fractions (bottom) from modelling the FRAP curves in (**i**)

While enrichment of wild-type and NΔ2-mutant LSD1 at GFI1B-LacI hubs was similar, LSD1(NΔ2) was significantly more enriched at C/EBPα-LacI hubs in comparison to wild-type (**Figure 3.10a-d**). By FRAP, the NΔ2 mutation did not appreciably alter the residence time of LSD1 at either TF-LacI hub or elsewhere in the nucleoplasm in the absence of recruitment (**Figure 3.10g-j**).[150–152] However, LSD1(NΔ2) displayed a lower free fraction (F$_{eq}$) and conversely a higher total bound fraction when recruited by either TF-LacI hub (**Figure 3.10h**). Altogether, these findings define differential recruitment of LSD1 by two opposing TF hubs — long-lived stable recruitment by GFI1B-LacI versus short-lived weak recruitment by C/EBPα-LacI (**Figure 3.10g,h**)— and implicate the LSD1 IDR in tuning the bound fraction of LSD1-C/EBPα.

111
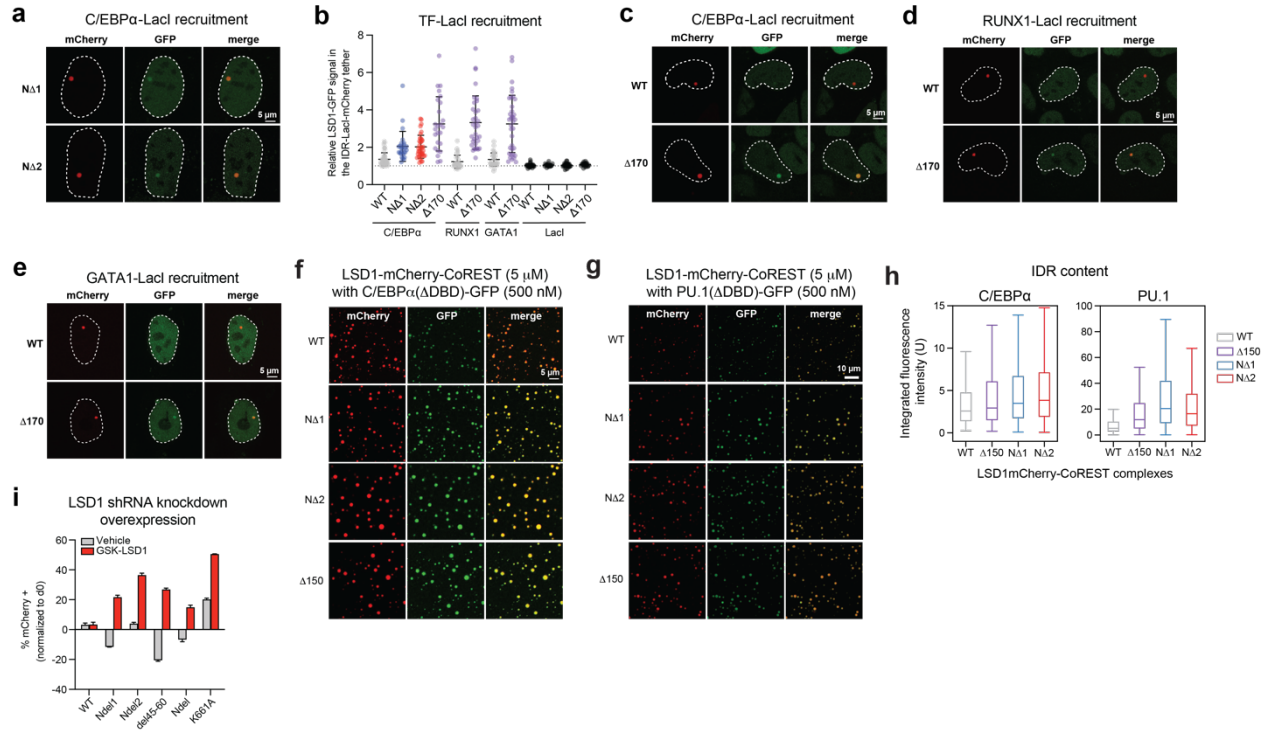
Figure 3.11. The LSD1 IDR tunes LSD1-TF interactions.

a, Representative live-cell fluorescence images of overexpressed LSD1(NΔ1)- and LSD1(NΔ2)-GFP recruited to the LacO array by C/EBPα-LacI-mCherry. **b,** Quantification of the relative LSD1-GFP variant signal intensity in the LacO array after recruitment with the specified TF IDR-LacI-mCherry construct in cell lines overexpressing full-length LSD1-GFP or LSD1(Δ170)-GFP. Shown are means ± s.e.m.; n = 45, 27, 38, 25, 32, 34, 35, 36, 26, 18, 27, 21 cells from left to right; P = 3.29 x 10⁻⁶ for WT vs LSD1(NΔ1) with C/EBPα, P = 2.08 x 10⁻⁸ for WT vs LSD1(NΔ2) with C/EBPα, P = 4.06 x 10⁻¹² for WT vs LSD1(Δ170) with C/EBPα, P = 2.50 x 10⁻¹¹ for WT vs LSD1(Δ170) with RUNX1, P = 6.24 x 10⁻¹⁰ for WT vs LSD1(Δ170) with GATA1; two-tailed unpaired t-test. **c,** Live-cell fluorescence images showing recruitment of WT and LSD1(Δ170) by C/EBPα(ΔDBD)-LacI-mCherry (right) for exogenously expressed LSD1-GFP. **d,** Representative live-cell fluorescence images of overexpressed WT and LSD1(Δ170)-GFP recruited to the LacO array by RUNX1-LacI-mCherry. **e,** Representative live-cell fluorescence images of overexpressed WT and LSD1(Δ170)-GFP recruited to the LacO array by GATA1-LacI-mCherry. **f,** Fluorescence images showing droplet formation of LSD1-mCherry/CoREST(Δ305) mixed with C/EBPα(ΔDBD)-GFP in droplet formation buffer. **g,** Fluorescence images showing droplet formation of LSD1-mCherry/CoREST(Δ305)

112

Figure 3.11 (continued) mixed with PU.1(ΔDBD)-GFP in droplet formation buffer. **h,** Left: Quantification of C/EBPα(ΔDBD)-GFP integrated fluorescence intensity in droplets formed by LSD1-mCherry/CoREST(Δ305). $P = 5.53 \times 10^{-45}$ for WT vs LSD1(NΔ1), $P = 1.11 \times 10^{-62}$ for WT vs LSD1(NΔ2) and $P = 3.77 \times 10^{-15}$ for WT vs LSD1(Δ150), Mann-Whitney U test. Right: Quantification of PU.1(ΔDBD)-GFP integrated fluorescence intensity in droplets formed by LSD1-mCherry/CoREST(Δ305). $P = 1.19 \times 10^{-287}$ for WT vs LSD1(NΔ1), $P = 0$ for WT vs LSD1(NΔ2) and $P = 1.55 \times 10^{-210}$ for WT vs LSD1(Δ150), Mann-Whitney U test. Data displayed as median ± s.d.. $n = 10$ independent fields per condition. **i,** Knockdown followed by overexpression of the indicated LSD1 variants in SET-2 cells in the presence of 50 nM GSK-LSD1.

We next investigated if the LSD1 IDR is required for C/EBPα-LacI recruitment. Overexpressed LSD1(NΔ1) and LSD1(NΔ2) -GFP exhibited significantly increased recruitment at C/EBPα-LacI hubs in comparison to wild-type (**Figure 3.11a,b**), supporting our results using endogenously tagged LSD1-GFP. Strikingly, LSD1(Δ1-170) enriched at even greater levels than the NΔ1 and NΔ2 mutants, demonstrating that the LSD1 IDR does not drive partitioning into C/EBPα-LacI hubs but is in fact inhibitory (**Figure 3.11b,c**). Consistent with this finding, droplets formed by purified LSD1(Δ1-150)-mCherry/CoREST could also co-partition the C/EBPα and PU.1 IDRs, showing that the LSD1 IDR is not required for TF interaction (**Figure 3.11f-h**).

LSD1(Δ1-170) also preferentially enriched into RUNX1(ΔDBD)- and GATA1(ΔDBD)-LacI hubs, suggesting that the LSD1 IDR more broadly inhibits LSD1-TF interactions (**Figure 3.11b-d**). Based on these data, we reasoned that the NΔ1 and NΔ2 deletions interfere with the IDR's inhibitory function to enhance TF partitioning, which might block AML differentiation. If true, then deletion of the IDR should also confer drug resistance. To test this idea, we conducted shRNA-mediated knockdown of endogenous LSD1 followed by mutant LSD1 overexpression in SET-2 cells. Cells overexpressing LSD1(NΔ1), LSD1(NΔ2), LSD1(Δ45-60), and LSD1(Δ1-170) all

rescued growth in the presence of GSK-LSD1 (**Figure 3.11i**), supporting the notion that the wild-type LSD1 IDR is necessary for differentiation. Taken together, these findings reveal a mechanism by which IDRs can attenuate corepressor partitioning into TF hubs with broader ramifications for cell differentiation.

3.6. Molecular features modulating LSD1-TF interactions

In order to probe the structure and dynamics of the LSD1 N-terminus as well as whether it is altering the complex as a whole, hydrogen-deuterium exchange mass spectrometry (HDX-MS) was performed with truncated LSD1(Δ1-150)/CoREST, full-length WT LSD1/CoREST, and N-term LSD1/CoREST mutants, LSD1(NΔ1) and LSD1(NΔ2), where LSD1 was complexed to CoREST(Δ1-305). HDX-MS exploits the fact that backbone amide hydrogens can exchange with deuterium when a protein is incubated in deuterium oxide and the rate of exchange is highly dependent on the local structural environment, such as how dynamic and accessible a region is to deuterium incorporation.[153–156] HDX-MS is well suited to study the conformation of proteins, including the identification of regions that are unstructured, at the interface of binding sites, or allosterically affected. To capture dynamics of the highly flexible IDR, we performed deuterium labeling at low temperature (1ºC) with short labeling time points to capture dynamics of the highly flexible IDR, and then repeated the experiment at higher temperature (30 ºC) with longer labeling time points for the well-folded, slow exchanging regions of LSD1 to achieve near-complete deuteration. Overall, the LSD1 NΔ1 and NΔ2 IDRs displayed increased deuterium exchange compared to WT. Notably, the IDR exhibited bimodal deuteration distributions, corresponding to 'closed' and 'open' conformations and/or populations which exchange deuterium at slower and faster rates, respectively (**Figure 3.12a**). Larger fractions of the NΔ1 and NΔ2 complexes exist in the 'open' state compared to WT LSD1, and the 'closed' states of the mutant complexes reached the 'open' state faster (**Figure 3.12b**). Furthermore, LSD1/CoREST, in particular at their interaction interface, showed increased deuterium uptake for NΔ1-, NΔ2-, and Δ1-150aa-containing complexes in comparison to wild-type (**Figure 3.12c**). Altogether, the HDX-MS data suggest that the LSD1 IDR participates in transient and dynamic states that occlude the LSD1/CoREST core, which are disrupted by the IDR deletion mutations. These results support

the notion that bimodal, dynamic states of the IDR promote TF demixing both intrinsically and by

occluding the LSD1/CoREST interface.
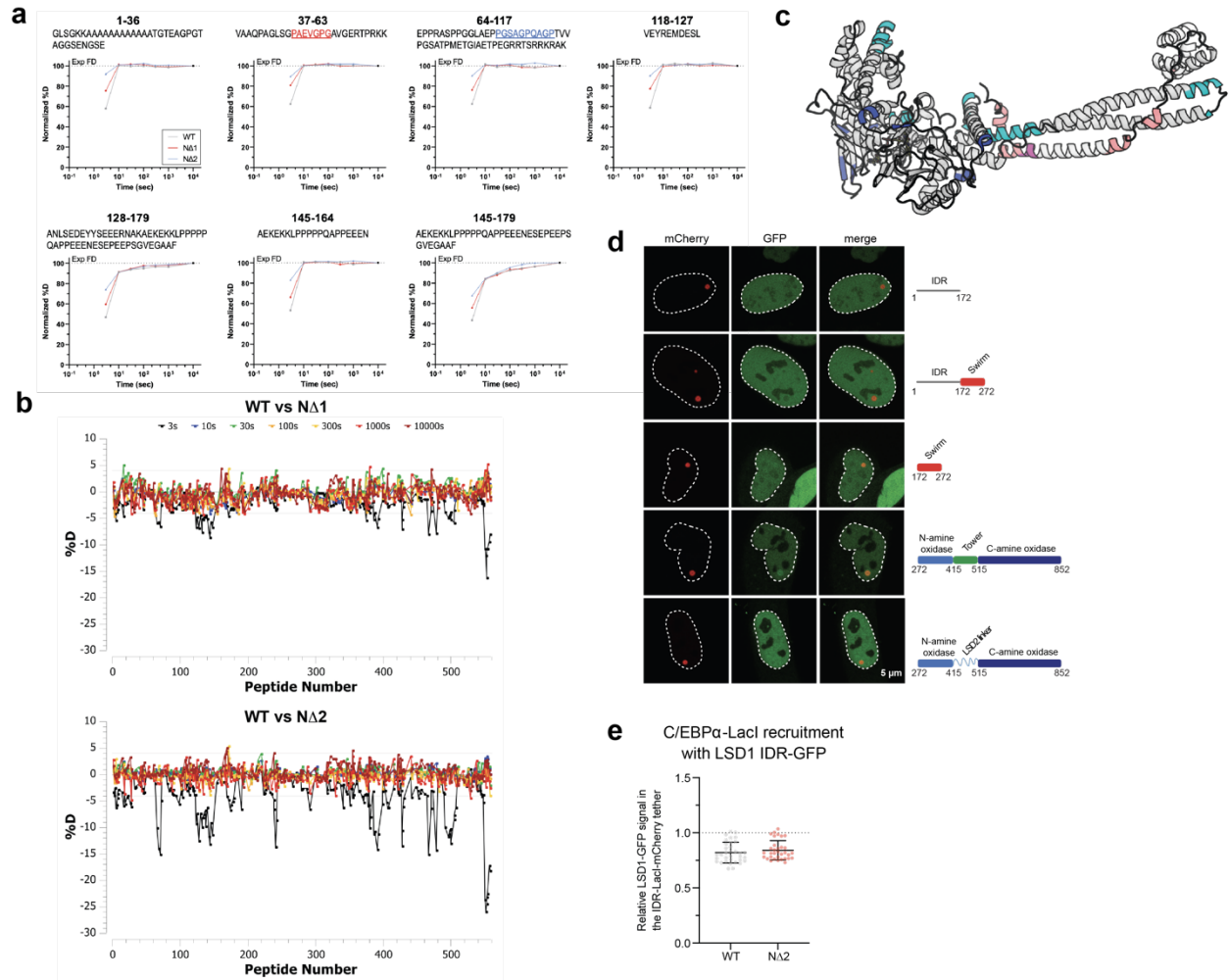
Figure 3.12. Molecular insights into the LSD1 IDR guiding IDR-mediated hub exclusion.

a, Normalized % D-uptake profile at the indicated time points and 1 °C for LSD1 WT, NΔ1, and NΔ2 CoREST

complexes for the indicated IDR peptides. **b,** Differences in %D incorporation for peptides across the LSD1

sequence for WT LSD1 relative to the IDR mutants at the indicated time points. **c,** Differences in D-uptake

for WT LSD1 relative to the IDR mutants at the earliest time point mapped onto the crystal structure of

LSD1/CoREST. LSD1 WT vs NΔ1 is shown in blue. LSD1 WT vs NΔ2/Δ1-150aa in cyan. CoREST WT vs

NΔ1 in salmon. CoREST WT vs NΔ2/Δ1-150aa in magenta. PDB: 2y48 **d,** Live-cell fluorescence images

showing recruitment of overexpressed LSD1 truncation constructs fused to GFP by C/EBPα(ΔDBD)-LacI-

Figure 3.12 (continued) mCherry where the domain maps of the LSD1 constructs used are depicted to the right. **e,** Quantification of the relative LSD1 IDR-GFP variant signal intensity in the LacO array after recruitment with C/EBPα(ΔDBD)-LacI-mCherry. Each dot represents quantification from an individual cell. Shown are means ± s.e.m.; n = 34, 33 cells from left to right; P = 0.7124 for WT vs LSD1(NΔ2); two-tailed unpaired t-test.

Next, we set out to investigate whether particular domains or structured regions of LSD1 are essential for driving LSD1-TF interactions in live cells. We overexpressed LSD1 domain truncations and assessed LSD1 recruitment to the C/EBPα-LacI hub. The structured amine oxidase domain (AOD) fused to the Tower domain of LSD1 was sufficient for mediating recruitment to the C/EBPα-LacI hub while the LSD1 IDR alone was actively excluded (**Figure 3.12d**). Notably, NΔ2 IDR-GFP exhibited decreased exclusion although this was not statistically significant (**Figure 3.12e**). Altogether, these findings show how a structured region of LSD1, the AOD-Tower domain interface, drives interactions with C/EBPα hubs, while counterintuitively, the IDR mediates exclusion or de-mixing from TF hubs (**Figure 3.13**). Furthermore, we hypothesize that the deletion mutations break the exclusionary properties of the LSD1 IDR, suggesting that the IDR actively excludes LSD1 from TF hubs (**Figure 3.13**). Experiments to investigate this hypothesis are currently ongoing.
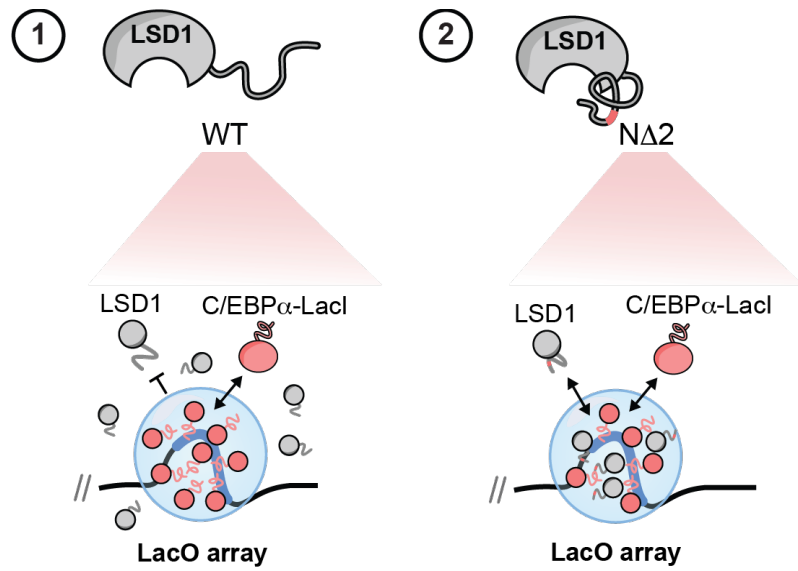
Figure 3.13. Schematic of the LSD1 IDR-mediated exclusion model (1) where mutations in the N-terminus break the exclusion to promote LSD1 incorporation into TF hubs (2).

3.7 The LSD1 IDR modulates enhancer commissioning

We reasoned that beyond its involvement with GFI1B, LSD1's secondary association with C/EBPα and other TFs might also impact AML differentiation, regulation of which is dependent on the LSD1 IDR. To test this idea, we mapped C/EBPα and PU.1 by ChIP-seq in wild-type LSD1 cells after vehicle or GSK-LSD1 treatment. A concomitant increase in H3K27ac levels was observed across all C/EBPα and PU.1 sites that overlap with LSD1 upon GSK-LSD1 treatment in wild-type cells (**Figure 3.14a**). By contrast, in LSD1(NΔ2) cells, increases in H3K27ac levels were attenuated at these C/EBPα and PU.1 sites upon GSK-LSD1 treatment (**Figure 3.14a**), suggesting that full enhancer activation is impaired in LSD1(NΔ2) cells. Notably, while an increase in C/EBPα levels was observed in wild-type cells upon GSK-LSD1 treatment across C/EBPα and PU.1 sites that overlap with LSD1, little or no change in C/EBPα levels was observed for LSD1(NΔ2) cells **Figure 3.14a**).
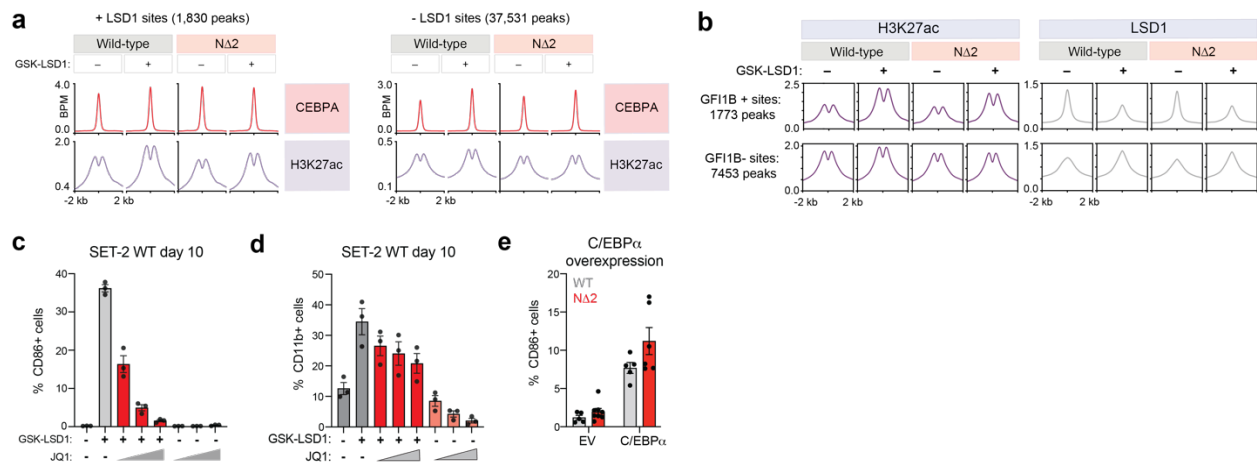
Figure 3.14. LSD1 regulates enhancers through its association with myeloid factors.

a, ChIP-seq profile plots show the average signal (*y*-axis, bins per million (BPM)) of C/EBPα and H3K27ac for C/EBPα and PU.1 sites that overlap with LSD1 sites versus those that do not in SET-2 WT and LSD1(NΔ2) under vehicle or GSK-LSD1 treatment (100 nM, 48 hr). The *x*-axis shows flanking regions ± 2 kb around the peak center which is centered on C/EBPα. **b,** ChIP-seq profile plots for H3K27ac and LSD1 signal at GFI1B sites versus non-GFI1B sites after taking the subset of sites where LSD1 overlaps with

Figure 3.14 (continued) C/EBPα and/or PU.1 for SET-2 WT and LSD1(NΔ2) under vehicle or GSK-LSD1 treatment (100 nM, 48 hr). The *x*-axis shows flanking regions ± 2 kb around the peak center which is centered LSD1. **c,d,** Co-treatment of WT SET-2 cells with JQ1 in the presence of GSK-LSD1 impairs myeloid differentiation in a JQ1 dose-dependent manner. Bar plot shows the fraction of CD86+ **(c)** or CD11b+ **(d)** cells (*y*-axis) as assessed by flow cytometry after 10 d of GSK-LSD1 (100 nM), JQ1 (10 nM, 25 nM or 50 nM) or combination treatment in comparison to vehicle control treatment. Data represent mean ± s.d. across three replicates. **e,** Overexpression of C/EBPα compared to an empty vector (EV) control overrides the block in differentiation and led to comparable increase in CD86-positivity in WT and LSD1(NΔ2) SET-2 cells.

Consequently, we considered if LSD1 broadly buffers CRE commissioning and, if so, whether IDR-mutant LSD1 aberrantly blocks CRE commissioning and AML differentiation in turn. This model would be consistent with depletion of LSD1(NΔ2) leading to potent differentiation while depletion of GFI1B in LSD1(NΔ2) cells was ineffectual (**Figure 3.7f**). First, to investigate LSD1 activity at CREs we mapped H3K27ac and LSD1 by ChIP-seq at GFI1B versus non-GFI1B sites, after taking the subset of sites where LSD1 overlaps with C/EBPα and/or PU.1, in wild-type and LSD1(NΔ2) cells after vehicle or GSK-LSD1 treatment (**Figure 3.14b**). GSK-LSD1 treatment results in a redistribution of LSD1, from GFI1B sites to these other sites, and a concomitant increase in H3K27ac at GFI1B sites (**Figure 3.14b**). At non-GFI1B enhancer sites there is a modest increase in H3K27ac in wild-type cells consistent with LSD1 redistribution to these sites, however, for LSD1(NΔ2) cells, the increase in H3K27ac was further attenuated or even decreased (**Figure 3.14b**).

Next, we sought to further establish that CRE activation is necessary for AML differentiation. In support, co-treatment of wild-type cells with the BET inhibitor JQ1 led to dose-dependent reduction of CD86- and CD11b-positivity in the presence of GSK-LSD1 (**Figure 3.14c,d**).[93,157] Therefore, once transcription repression is relieved by GSK-LSD1, downstream

120

enhancer activation is necessary for the expression of genes involved in AML differentiation, which is prevented by JQ1 co-treatment. By contrast, overexpression of C/EBPα led to comparable increases in CD86-positivity in wild-type and IDR-mutant cells suggesting that differentiation is competent at the appropriate dosage of C/EBPα (**Figure 3.14e**).
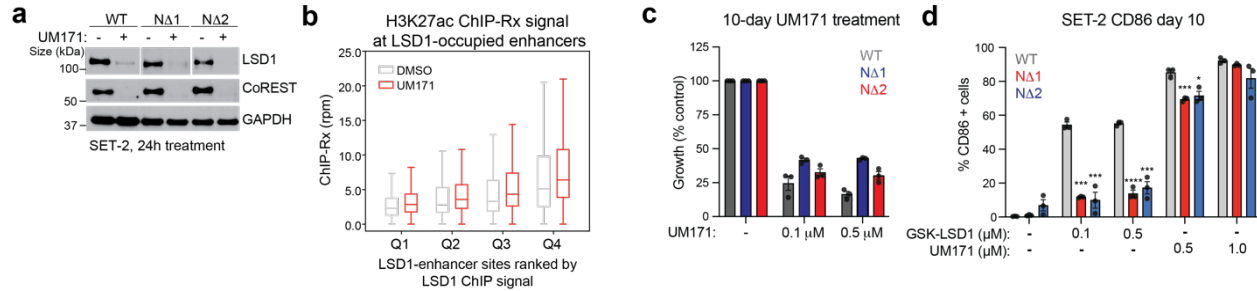
Figure 3.15. LSD1 regulates enhancer commissioning at CREs.

a, Immunoblot showing that LSD1 and CoREST are degraded after treating SET-2 cell lines with UM171 for 24 h. **b,** Quantitative ChIP-seq for H3K27Ac after 6 h treatment with UM171. RPM (reads per million) on the *y*-axis while *x*-axis is grouped by quartiles. **c,** Relative cell growth of SET-2 cell lines treated with either 0.5 μM or 1.0 μM UM171 for 15 days. Mean ± s.e.m across three replicates is shown. **d,** UM171-induced LSD1/CoREST degradation leads to myeloid differentiation in WT, LSD1(NΔ1), LSD1(NΔ2) SET-2 cell lines. Bar plot shows the fraction of CD86+ cells (*y*-axis) in WT, LSD1(NΔ1) and LSD1(NΔ2) SET-2 cell lines as assessed by flow cytometry after 10 d of GSK-LSD1 (0.1 μM, 0.5 μM) or UM171 (0.5 μM, 1.0 μM) treatment in comparison to vehicle control treatment. Data represent mean ± s.d. across three replicates.

Finally, to test if LSD1 restrains enhancer activity across all its bound CREs, we conducted quantitative ChIP-seq (ChIP-Rx) for H3K27ac after acute depletion of LSD1-CoREST through 6 h treatment with UM171 — a small molecule that degrades LSD1-CoREST (**Figure 3.15a**).[73] After treatment, increases in H3K27ac were broadly observed and correlated with initial levels of LSD1, supporting the notion that LSD1 buffers enhancer commissioning at its bound enhancer sites (**Figure 3.15b**). Notably, both wild-type and IDR-mutant cells underwent potent growth arrest and

differentiation in the presence of UM171 (**Figure 3.15c,d**). Taken together, our results support

the notion that LSD1 restrains enhancer commissioning across active CREs bound by key TFs to

regulate AML differentiation, which is controlled by its N-terminal IDR (**Figure 3.16**).
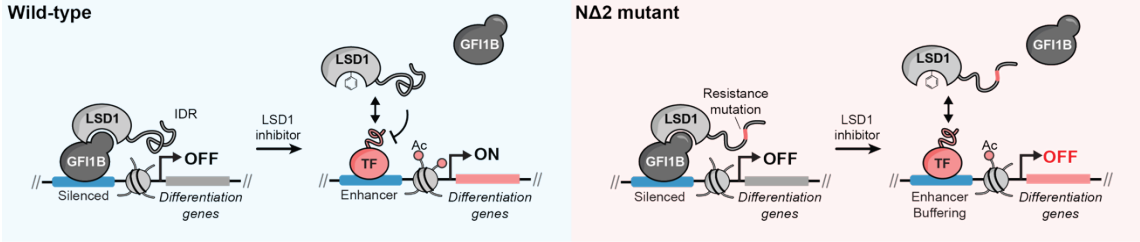
Figure 3.16. Cartoon showing enhancer activation defect for LSD1(NΔ2) versus wild-type under LSD1 inhibitor treatment.

3.8. Discussion

By leveraging drug suppressor mutations, here we define two distinct yet interconnected mechanisms by which LSD1 interacts with TFs to control enhancer commissioning. While high-affinity LSD1-TF interactions can directly recruit LSD1 to silence specific TF sites (i.e., LSD1-GFI1B), we show that weak multivalent and non-specific interactions also drive LSD1 co-partitioning at TF hubs to modulate enhancer activity more broadly.[70,93] These two recruitment modalities are likely mechanistically coupled, as downregulation of GFI1B – or LSD1 inhibition – triggers LSD1 redistribution to its lower-affinity TF partners situated on active enhancer hubs (**Figure 3.16**). We speculate that this LSD1 redistribution buffers downstream enhancer commissioning, attenuating gene expression programs involved in differentiation.[40]

Notably, the LSD1 N-terminal IDR plays a critical role in controlling LSD1 multivalent interactions with TFs and likely other coregulators, assigning a key function of this domain. However, instead of directly mediating multivalent interactions with TFs, the IDR paradoxically blocks them (**Figure 3.13**). Deletions within the N-terminus break the inhibitory state of the LSD1 IDR and weaken its ability to repel LSD1 from TF-hubs. This potentiates weak interactions with TFs, such as C/EBPα, by releasing the structured region of LSD1-CoREST to engage in weak multivalent interactions. Thus, although mutations in the LSD1 IDR promote co-condensation in vitro, our findings show that this is likely due to enhanced interactions with other regions of LSD1-CoREST. Our work provides a high-resolution view of how IDRs can modulate coregulator function and TF interactions, paradoxically through regulating the ability of a structured region to engage in weak interactions. Altogether, our findings suggest that IDRs have multiple, diverse mechanisms in modulating multivalency and can promote demixing (or TF hub exclusion) to block TF interactions.[24,26,149]

Despite their original design, LSD1 inhibitors serendipitously function as broader transcription factor modulators, unbalancing the antagonism between GFI1B and C/EBPα in

blood development by shifting LSD1's interaction partners.[47,93] These findings offer important considerations for the development of small molecule inhibitors of transcription factors, as disruption of high-affinity TF interactions can lead to the inadvertent formation of new TF complexes, even of lower affinity, with downstream and critical roles in drug action.[158,159] In this vein, we show how degradation versus inhibition of LSD1 can have differential outcomes on AML differentiation due to the elimination of secondary TF interactions.

More broadly, our study demonstrates how mutations in IDRs can promote epigenetic dysregulation and drug resistance. Furthermore, the systematic identification of drug resistance mutations enables the molecular and temporal dissection of the sequential steps in drug mechanism of action.

3.9. Methods

Cell culture and lentiviral production. SET-2 was a gift from Matthew D. Shair; HEK 293T was a gift from Bradley E. Bernstein; U2OS 2-6-3 was a gift from David L. Spector. All cell lines were authenticated by Short Tandem Repeat profiling (Genetica) and routinely tested for mycoplasma (Sigma-Aldrich). All media were supplemented with 100 U/mL penicillin and 100 µg/mL streptomycin (Life Technologies) and fetal bovine serum (FBS, Peak Serum). All cell lines were cultured in a humidified 5% CO_2 incubator at 37 °C. SET-2 were cultured in RPMI-1640, 20% FBS. HEK 293T and U2OS were cultured in DMEM (Life Technologies), 10% FBS. Lentivirus was produced by co-transfecting HEK293T cells with pCMV-VSV-G (Addgene plasmid #8454), psPAX2 (Addgene plasmid #12260) and transfer vector plasmid (encoding gene of interest) using FuGENE HD (Promega) or Lipofectamine 3000 according to manufacturers' protocols. Media was exchanged after 6-7 h and the viral supernatant was collected 48–72 h after transfection and filtered (0.45 µm). Transduction was carried out by mixing the virus with cells with 8 µg/ml polybrene for SET-2 and 5 µg/ml for U2OS and then centrifuged at 1,800 x *g* for 90 min at 37 ºC. After 48 h post-transduction, media was changed and puromycin (Thermo Fisher Scientific) selection was carried out for 5 d at 2 µg/mL for SET-2 and U2OS 2-6-3.

Chemical Reagents. Compounds were stored at -80 °C in 100% DMSO. The vehicle condition represents 0.1% (v/v) DMSO treatment. GSK-LSD1, GSK690 and (+)-JQ1 were used purchased from Sigma-Aldrich (≥98% by HPLC). Deuterium oxide (D2O, 99.9 atom % D; 151882), acetonitrile (ACN, 99.9%; A998) and formic acid (FA, 99.5% purity; A117) were from Fisher (Fair Lawn, NJ) and HPLC water from J.T.Baker (JT4218–3; Center Valley, PA).

Cell Growth Assays. Cell lines were seeded in 96-well plates with 45,000 cells per well in triplicate with drug or vehicle treatments. An equal volume of cells from each well was split and re-plated with fresh media containing drug or vehicle every 3-4 days. Cell viability was monitored

at day 10 by measuring end point luminescence using CellTiter-Glo (Promega) on a SpectraMax i3x plate reader. ATP standard curve was prepared using known concentrations of ATP and used to calculate the ATP content of cells. On the days of measurement, cells were replated with fresh media containing drug or vehicle. All growth assays were performed at least twice.

Generation of Clonal Drug-Resistant LSD1 Mutant Cell Lines. sgRNA sequences targeting LSD1, sgNΔ1(GGGCCCAGCCGAGGTCGGGC) and sg NΔ2(GGGTCCGCAGGGCCTCAGGC), were cloned into pLentiCRISPR-eGFP.v1. SET-2 were transduced with the resultant plasmids as indicated above and subsequently treated with 100 nM GSK-LSD1 for 50 d to enrich for drug-resistant mutant cells. Surviving cells were FACS-sorted as single cells and then expanded. For library preparation, genomic PCR primers with Illumina adapter sequences were used to amplify specified regions of LSD1 as previously described. Samples were sequenced on a MiSeq genome analyzer (Illumina). The sequencing reads were analyzed using CRISPResso2 (v.2.0.40).

Gene knockout by CRISPR/Cas9 and Competition-Based Growth Assays. pLentiGuide-mCherry was constructed by replacing the puromycin resistant gene in pLentiGuide-Puro (Addgene plasmid #52953). sgRNA sequences targeting LSD1, sg1 (CTAAATAACTGTGAACTCGG) and sg2 (TAGGGCAAGCTACCTTGTTA); GFI1B sg1 (GGGGTCGGGACAGCACAATG) and sg2 (CCTTGTTGCACTTCACACAG) were cloned into pLentiGuide-mCherry. Cas9-expressing SET-2 WT, NΔ1 and NΔ2 cell lines were transduced with the resultant pLentiGuide-mCherry plasmids as indicated above. mCherry-positive cell population was mixed with non-transduced mCherry-negative SET-2 wt cells at 1:1 ratio at d4 post-infection The percentage of mCherry-positive was measured every 3-5 days up to 20 days using ACEA Novocyte flow cytometer (Agilent). Final mCherry % was divided by the initial mCherry % at d3 to calculate normalized % mCherry.

shRNA-mediated Knockdown of LSD1, Rescue Overexpression and Competition-Based Growth Assays. Site-directed mutagenesis as performed using Q5 Site-Directed Mutagenesis Kit (New England Biolabs) to generate LSD1 NΔ1, NΔ2, Δ45-60 and Δ1-170. All modified LSD1 ORFS were subcloned into pSMALP-mCherry with an N-terminal FLAG tag using Gibson cloning (New England Biolabs). SET-2 were transduced with the resultant plasmids and sorted for mCherry+ cells. shRNA sequence targeting LSD1, P2(GGAGCTCCTGATTTGACAAAG), was cloned into pLKO.1-puro. The mCherry+ cells were subsequently transduced with pLKO.1 LSD1 P2 to knockdown LSD1. After 3 days of puromycin selection, the mCherry+ cell population was mixed with mCherry-negative SET-2 WT cells. Final mCherry % was subtracted by the initial mCherry % at d4 to calculate relative % mCherry.

Genotype Determination. Genomic DNA was isolated using QIAamp Blood & Cell Culture DNA Mini Kit (Qiagen). For amplicon sequencing, genomic DNA was subjected to two rounds of PCR reactions as previously described.[121] PCR1 was conducted using 500 ng of genomic DNA with Q5 HotStart Polymerase (NEB) to amplify the target region. PCR1 products were ran on an agarose gel and gel purified using Zymo Gel Extraction Kit (Zymo). 1 uL of the purified DNA was used as an input for PCR2. PCR2 was conducted with PCR primers containing Illumina adaptor sequences. PCR2 products were ran on an agarose gel, gel purified and pooled. The final libraries were quantified using Qubit High Sensitivity Assay Kit (Thermo Fisher Scientific) and sequenced on a MiSeq genome analyzer (Illumina). The sequencing reads were analyzed using CRISPResso2 (version 1.0.13).[104]

Mammalian Overexpression Constructs. To generate LSD1-mEGFP U2OS 2-6-3 overexpression cell lines, mEGFP followed by a 'GGGSGGGS' linker was cloned into the C-terminus of LSD1 harboring an N-terminal Flag-SV40 NLS tag followed by 'GGS' in an IRES puro

lentiviral vector. U2OS 2-6-3 cells were transduced and selected with puromycin as described above followed by immunoblot validation.

CRISPR/Cas9-mediated genome editing. Knock-in cell lines were generated according to published procedures.[160] mEGFP followed by a 'GGGSGGGS' linker was knocked into the C-terminus of LSD1 in U2OS 2-6-3 cells. sgRNAs were cloned into a Cas9 plasmid, PX459 (Addgene plasmid #48139), and co-transfected 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.

To generate U2OS 2-6-3 LSD1-mEGFP NΔ2, electroporation was performed using the NEON system with gRNA complex and Alt-R Cas9 enzyme from IDT. gRNA complex was formed by reconstituting gene-specific crRNA and tracrRNA (1072534, IDT) to 100 µM in IDT duplex buffer, mixed 1:1 and hybridized by incubation at 95°C for 5 min, followed by cooling to room temperature on the bench top. RNP complex was formed by mixing gRNA complex with Alt-R Cas9 enzyme and followed by incubation at room temperature for 15 min. 2×10^5 cells were washed twice with PBS and resuspended in buffer R. RNP complex, gene-specific Ultramer ssODN donor consisting of 50 bp on either side of the mutation and Alt-R Cas9 electroporation enhancer were added to the cell suspension and electroporated at 1400V with 15 ms pulse width for 4 pulses using the Neon Transfection System 10 µL kit (Thermo Fisher, MPK1025). After electroporation, cells were immediately transferred to prewarmed media supplemented with HDR Enhancer. To generate single-cell clones, cells were sorted on a MoFlo Astrios EQ cell sorter, expanded, and screened by genomic PCR followed by validation with Sanger sequencing.

Immunoblotting. For whole cell extracts, cells were lysed on ice using radioimmunoprecipitation assay (RIPA) buffer (Boston BioProducts) supplemented with fresh HALT™ protease inhibitor cocktail (Thermo Fisher Scientific). The lysate was clarified by centrifugation at 10,000 $x\ g$ for 10 min. Protein concentration of the lysates was measured using Bradford Assay (Bio-rad) or BCA Protein Assay Kit (Thermo Fisher Scientific). Immunoblotting was performed according to standard procedures. The primary antibodies used are as follows: LSD1 (Bethyl Laboratories A300-215A), LSD1 (Abcam ab17721). GFI-1B (B-7) X (Santa Cruz Biotechnology sc-28356X). Monoclonal Anti-FLAG M2 (Sigma-Aldrich F1804). GAPDH (Santa Cruz Biotechnology sc-477724). GFI1B (Santa Cruz Biotechnology sc-28356X).

Co-immunoprecipitation (Co-IP). 2×10^7 SET-2 cells were washed twice with cold PBS and flash frozen. Cells were thawed, lysed on ice in N450 buffer (50 mM Tris-HCl pH 7.5, 450 mM NaCl, 1 mM $MgCl_2$, 1% NP-40 alternative, 5% glycerol) supplemented with 1:10,000 benzonase (Sigma-Aldrich), and the lysates were cleared. The protein concentration was quantified as above and diluted to 1 mg/mL in lysis buffer. For anti-FLAG, Supernatants were immunoprecipitated overnight at 4 ºC with 2.5 µg anti-FLAG M2 antibody and 40 µL Protein G Dynabeads (ThermoFisher Scientific). Beads were washed twice with lysis buffer, eluted in SDS-PAGE loading buffer, and carried forward to immunoblotting as described above. For cross-linking, antibodies were crosslinked to Protein G Dynabeads (ThermoFisher Scientific) using BS3. Crosslinked Dynabeads were washed thrice with lysis buffer and incubated with the precleared protein lysate. Immunoprecipitation was carried out overnight at 4 ºC under constant rotation. Beads were washed once with wash buffer A (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.05% IGPAL-CA-630) and twice with wash buffer B (50 mM Tris-HCl pH 7.5, 150 mM NaCl) and eluted in SDS-PAGE loading buffer and carried forward to immunoblotting as described above or mass spectrometry as described below.

Protein Purification. Full-length LSD1 and variants were cloned into a pET15b vector (gift from P. A. Cole) containing an N-terminal His$_6$-tag using NEBuilder HiFi DNA Assembly Master Mix (NEB E2621L). Mutant sequences were generated by PCR and inserted into the same base vector as described above. IDR-mEGFP fusion constructs were likewise cloned into a pET15b vector containing an N-terminal His$_6$-tag. For C-terminal mCherry and mEGFP constructs, an 8-amino-acid linker sequence 'GGGSGGGS' followed by the fluorescent tag were inserted in-frame after the protein coding sequence. The LSD1 constructs were expressed in BL21-CodonPlus (DE3)-RIPL competent *E. coli* and purified according to a literature procedure. All protein purification steps were performed at 4 °C. Briefly, cells were lysed by sonication (10 s on, 20 s off, 60% Amp) after resuspending in lysis buffer (50 mM HEPES pH 7.5, 100 mM KCl, 0.5 mM TCEP) supplemented with protease inhibitor cocktail (Roche). The lysate was clarified by centrifugation at 20,000 *x g* for 40 min. The clarified lysate was added to equilibrated His60 nickel resin (Takara Bio) and bound in batch for 1 h. The resin was washed with 10 column volumes (CV) of lysis buffer, followed by 10 CV of Wash buffer 1 (lysis buffer supplemented with 20 mM imidazole) and lastly eluted by the addition of 5-10 CV of lysis buffer supplemented with 250 mM imidazole. Fractions were combined and dialyzed overnight in 2 L of dialysis buffer (20 mM Tris pH 7.5, 100 mM NaCl, 5 mM β-mercaptoethanol, 2 changes), concentrated (Amicon Ultra-15 Centrifugal Filter Unit, 30 kDa MWCO) to 5-10 mLs and purified with a 5 mL Q Sepharose HP column with a gradient elution from Buffer A (20 mM Tris pH 7.5, 1 mM TCEP) to Buffer B (20 mM Tris pH 7.5, 1 M NaCl, 1 mM TCEP). The protein was concentrated and purified by gel filtration chromatography on a Superdex 200 Increase 10/300 GL column (GE healthcare) in storage buffer (50 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM TCEP, 5% glycerol). Fractions with > 95% purity as assessed by coomassie-stained SDS PAGE were pooled and stored at -80 ºC.

For LSD1-CoREST complexes, LSD1 was expressed as described above while CoREST(305-482) was expressed from a pGEX vector (gift from A. Mattevi). The plasmid was transformed into

BL21-CodonPlus (DE3)-RIPL *E.coli* cells and after plating a single colony was cultivated in LB media with 100 mg/L ampicillin at 37 ºC and expression was induced at OD600 of 0.8 by adding 0.25 mM isopropyl β-D-thiogalactoside (IPTG) and grown overnight at 17 ºC. The cells were pelleted by centrifugation at 4,000 *x g* for 30 min and stored at -80 ºC prior to purification. Pellets of CoREST and LSD1 were resuspended in lysis buffer (50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl, 5% glycerol, 7.5 mM imidazole supplemented with PMSF, DNAse and EDTA-free Roche protease inhibitor cocktail) in a weight ratio of 1:1.5, respectively. Cells were disrupted by sonication, clarified by centrifugation and passed through nickel affinity resin as before. The eluent was then loaded onto GST resin equilibrated in GST affinity buffer (50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl, 5% glycerol, 1 mM DTT, 1 mM EDTA) and the GST-tag was cleaved on the resin after incubation with GST-PreScission protease (APEXBIO) overnight at 4 ºC. The protein was eluted by washing the column with GST affinity buffer, concentrated and subsequently gel-filtered on a Superdex 200 10/300 GL column equilibrated in storage buffer as before. The purity of the complex was verified by SDS-PAGE and fractions with 90-95% purity were pooled and stored at –80 ºC. FPLC traces for proteins purified by size-exclusion chromatography.

For expression and purification of LSD1-CoREST-HDAC1 (LHC) complex, we followed previously reported protocols.[54] Briefly, HEK293F cells were co-transfected with LSD1, CoREST, and HDAC1 plasmids and the complex was purified from cell lysate by anti-Flag immunoaffinity enrichment and gel-filtration on a Superose 6 Increase 10/300 GL column. The purified complex was concentrated and stored at 4 ºC for up to 3 weeks. Purification of nucleosomes were purified according to a previously described protocol.[161–163]

The following IDR-mEGFP constructs were expressed and purified:
CEBP/α(ΔDBD): residues 1-285; Pu.1ΔDBD: residues 1-159

The plasmids were transformed into LOBSTR cells, expressed, and purified according to a previously described protocol.[25] After Ni affinity, fractions were analyzed using coomassie-stained SDS-PAGE, combined, and dialyzed against 50 mM HEPES pH 7.5, 450 mM NaCl, 0.5 mM TCEP, and 5% glycerol. The purified proteins were concentrated (Amicon Ultra-15 Centrifugal Filter Unit, 10 kDa MWCO) and stored at -80 ºC until usage.

LSD1 Enzyme Assays. LSD1 enzymatic activity assays were performed in triplicate using Amplex Red Hydrogen Peroxide/Peroxidase Assay Kit (Invitrogen) with recombinant LSD1 and a synthetic peptide corresponding to the first 21 amino acids of H3K4me2 (Anaspec). LSD1 (500 ng/well) was incubated with 20 µM peptide at rt for 30 min. The endpoint fluorescence was measured on a microplate reader (excitation: 530 nm; emission: 590 nm) after 60 min following the addition of the Amplex Red/HRP mixture. Inhibition assays were performed as described above. Briefly, LSD1 (75 ng/well) and inhibitors at the appropriate concentration were incubated at rt for 10 min in reaction buffer with 0.01% BRIJ35 (ThermoFisher Scientific) prior to the addition of peptide. IC_{50} values were determined in GraphPad Prism v.7 by nonlinear regression analysis (log(inhibitor) vs. response—variable response) of the concentration/inhibition data.

LHC Complex Enzyme Assays. LHC demethylase and deacetylase activity was assessed with H3K4me2 and H3K27Ac mononucleosomes according to a modified literature procedure.[161] Briefly for assessing demethylase activity, H3K4me2 nucleosomes (300 nM, 185 bp) were treated with LHC (365 nM) in assay buffer (50 mM HEPES pH 7.5, 6.4 mM KCl, 1.9% glycerol, 0.2 mg/mL BSA and EDTA-free protease inhibitor cocktail (Pierce)) at 25 ºC. For GSK-LSD, LHC complex was first incubated with 10 µM of GSK-LSD1 for 30 min prior to adding H3K4me2 nucleosomes as above. At each time point, 9 µL aliquots were mixed with 2 µL of 80 mM EDTA and 4 µL of 4x SDS-PAGE gel loading buffer. The samples were boiled for 5 min at 95 ºC and resolved by 15% SDS-PAGE. H3K27Ac was assessed as described previously. Briefly, 100 nM of H3K27Ac

nucleosomes were treated with 30 nM of LHC complex in assay buffer (50 mM HEPES pH 7.4, 100 mM KCl, 100 μM of InsP$_6$ and 0.2 mg/mL of BSA) at 37 ºC and time points were acquired as described above. Total H3 and H3K4me2 or H3K27Ac were detected by western blot with specific antibodies on separate gels. Western blot bands were visualized by ECL and quantified using ImageJ software.

ThermoFAD Assay. *Thermo*FAD was performed in quintets as previously described.[106] Thermal curves were obtained using a qRT-PCR detection system (BioRad C1000 Touch Thermal Cycler) with a temperature gradient from 20 to 95 ºC. The T$_m$ was determined by fitting the thermal curves to a Boltzmann sigmoid and calculating the inflection point (GraphPad Prism v.7).

In Vitro Droplet Assay. Droplet formation was initiated by the 1:1 addition of droplet formation buffer (50 mM HEPES pH 7.5, 0.5 mM TCEP, 5% glycerol and 20% PEG-8000 with the indicated concentration of NaCl). Droplet assays were performed in 8-well PCR strip tubes according to a previously described protocol and loaded onto a well of a glass-bottom 96-well plate (MatTek PBK96G-1.5-5-F).[31] Reactions were incubated for 30 min in the imaging vessel to allow droplets in solution to settle on the glass imaging surface. The reaction was then imaged on the Celldiscoverer7 fluorescence microscope using an Axiocam 512 camera with a Plan-Apochromat 50x /1.2 objective for the concentration and salt series. For co-condensation assays, the reactions were imaged with an inverted laser scanning confocal microscope (Zeiss Axio Observer.Z1/7, LSM 980 AiryScan2) for improved multi-channel fluorescence imaging. Images were collected with the following parameters: 488 nm (0.1 % power, detect gain 700 V), 561 nm lasers (0.05% power, detect gain 650 V), were used for imaging GFP and mCherry, respectively, pinhole size = 1.0 AU, pixel dwell time = 0.77 μs, and images were acquired with a Plan-Apochromat 63x/1.40 Oil DIC. For nuclear FRAP, Zeiss Zen acquisition software was set to the following settings: pinhole size = 1.0 AU, pixel dwell time = 0.77 μs, 488 nm laser power for acquisition = 0.3%, and

100% laser power for bleaching. Before acquiring all the fluorescence images, we set the laser intensity and microscope detectors to make sure no pixels in the images were saturated and for 2-color imaging, proper emission filters were used to ensure no channel bleed-through. Images presented are of droplets that have settled on the glass surface of the imaging plate. Droplet formation experiments were repeated at least two times and droplets were quantified between 5-6 and 10-15 independent fields of view from each reaction for the co-localization and phase-diagram experiments, respectively.

HX-MS at the peptide level and data analysis. For hydrogen exchange (HX) experiments at the peptide level, complexes were labeled at 1 °C and quenched at 26, 300, 3000 and 14400 sec at pH 7.5 and XXX at pH 5.5, as described above. Samples were injected in the loop and digested online for 2 min at 200 µl/min using an immobilized protease type XIII/pepsin column (w/w, 1:1, NBA2014002, NovaBioAssays, Woburn, MA), operated at 8 °C. Peptides were trapped onto an Acclaim™ PepMap™ 300 µ-Precolumn™ (C18, 1 × 15 mm, 163593, Thermo Fisher Scientific, Waltham, MA) using solvent A and were separated onto a Hypersil Gold C18 (1 × 50 mm, 1.9 µM, 25002–051030, Thermo Scientific, Waltham, MA) at 40 µl/min using solvents A and B. The following gradient was applied: 2% B to 5% B in 0.1 min, to 30% B in 18 min, to 95% B in 0.1 min; kept at 95% B for 1.5 min and returned to initial conditions in 0.2 min. Samples were analyzed on a Q-Exactive HF Orbitrap MS (Thermo). Full MS scans were acquired in the m/z range 300–1300, with an AGC target 3e6 and 45,000 resolution (at m/z 200), 2 µscans/spectrum and, maximum injection time 100 ms. For peptide identification experiments, peptides with charge states 2–6 were selected for MS/MS fragmentation using higher energy C-trap dissociation with an AGC target 5e4 and 30,000 resolution, loop count 15, and isolation window 2 m/z, NCE 28 and maximum injection time 100 msec. Data were acquired in profile mode. Peptides were identified using Spectrum Mill Proteomics Workbench (prerelease version BI.07.11.216, Agilent Technologies) as previously described.[164]

134

FRAP imaging. Cells were plated on sterile 35mm #1.5 glass bottom dishes (Matsunami, D35-14-1.5-U) 24h prior to imaging for nuclear FRAP treated with DMSO vehicle or 500 nM GSK-LSD1, respectively. Cells were imaged in FluoroBrite DMEM (Thermo Fisher A1896701) supplemented with 5% FBS. FRAP was performed on an inverted laser scanning confocal microscope (Zeiss Axio Observer.Z1/7, LSM 980 AiryScan2) equipped with channel spectral detection, a motorized stage, a full incubation chamber maintaining 37 ºC and 5% CO_2, a heated stage as well as several laser lines. The 488 nm and 561 nm lasers were used for FRAP measurements and images were acquired with a Plan-Apochromat 63x/1.40 Oil DIC objective as described above. For nuclear FRAP, Zeiss Zen acquisition software was set to the following settings: pinhole size = 1.0 AU, pixel dwell time = 0.77 µs, 488 nm laser power for acquisition = 0.3%, and 100% laser power for bleaching. Two nuclear regions of interest (1.2 µm) were drawn (one was photobleached and the other was a reference ROI) along with a background reference region. Twenty-five images were acquired before photobleaching, photobleached within 0.5 ms, and then imaged for 5 min with a frame rate of 0.9 fps (frames-per-second) after photobleaching to monitor recovery. Mean fluorescence intensity data within each ROI were used for FRAP quantification.

For LacO-LacI assays, tether experiments were adapted from a published report.[34] Briefly, U2OS cells were plated on dishes as described above 24 hr prior to transfection with 500 ng of GFI1B(ΔDBD)- or C/EBPα(ΔDBD)-LacI-mCherry plasmid with Lipofectamine 3000 reagent (Invitrogen L3000). Imaging was performed on live cells 48h post-transfection across three-four experimental replicates. FRAP imaging was performed as described above with the following modifications: the laser power for the 561 nm laser was set to 0.2% while the 488 nm laser remained set to 0.3%, 5 images were acquired before photobleaching and then imaged for 5 min with a frame rate of 0.9 fps after photobleaching to monitor recovery. 1.6 µm ROIs were drawn

around the LacO array, a nuclear control region and a background reference region. FRAP experiments with 488- and 561 nm channels were acquired separately.

To correct for xy-drift of the bleached spot due to live cell movement during image acquisition, a custom GUI was used in python to correct for xy-drift of the bleached spot due to live cell movement during image acquisition as well as to correct for photobleaching. Custom Python v.3.7.4 scripts were subsequently used to normalize the bleach spot intensity relative to the nuclear intensity, to calculate the bleach depth and to normalize the bleach depth to 100% according to a published procedure. Each mean FRAP curve was generated from averaging FRAP curves from multiple single cells over three acquisition days.

LacO-LacI Quantification Imaging. Cells were prepared as described above. Imaging was performed across three experimental replicates on an inverted laser scanning confocal microscope (Zeiss Axio Observer.Z1/7, LSM 980 AiryScan2) as described above. Z-stacks were acquired with a slice interval of 0.23 µm with 488 nm at 0.3% laser power (750 V detector gain) and 561 nm at 0.12% laser power (650 V detector gain) to excite LSD1-mEGFP and IDR-LacI-mCherry fluorophores, respectively, with the following imaging parameters: pinhole size = 1.0 AU, pixel dwell time = 0.5 µs, 0.56 fps. For LSD1 IDR-GFP overexpression cell lines, the 488 nm laser power was lowered to 0.1 % laser power prior to image acquisition to prevent signal saturation. Before acquiring the fluorescence images, we set the laser intensity and microscope detectors to make sure no pixels in the images were saturated and for 2-color imaging, proper emission filters were used to ensure no channel bleed-through. All images were channel aligned prior to further processing using Zeiss ZenBlue processing software. Images were acquired across three experimental replicates. To quantify LSD1-GFP enrichment at the LacO arrays we followed a previously reported procedure. Briefly, we selected a primary region of interest corresponding to the IDR-LacI-mCherry tether detected based on a fixed intensity threshold on

136

the mCherry-channel using ZEN Blue 3.0. A 3 µm wide ring with an approximately 1 µm gap from the primary region was used as a reference region in the cell while the same size ring was used outside of the cell to subtract out the background signal. Then the brightest slice (#N) in the mCherry-channel z-stack for the LacO-associated tether was selected and the mean intensity was measured on both the mCherry and GFP channels. After background subtraction, the LSD1-GFP signal in the LacI-tether was quantified by normalizing the mean intensity of the enclosed primary region in the GFP-channel to the reference region and similar guidelines as those discussed in Basu *et al*. were followed.[149] A ratio greater than 1 suggests LSD1-GFP – IDR-mCherry interactions.

Chromatin Immunoprecipitation (ChIP)-seq. ChIP was performed as previously described in duplicates on SET-2 cells (WT and Ndel2) treated with GSK-LSD1 (100 nM) or vehicle for 48 h.[69,126] The following antibodies were used: LSD1 (Bethyl Laboratories A300-215A); GFI1B (Abcam ab26132); H3K27ac (Active Motif 39133), C/EBPα (Cell Signaling Technology, #8178). For quantitative ChIP, drosophila chromatin (Active Motif cat no. 08221011) and spike-in antibody (Active Motif cat no. AB27377370) were added to sonicated chromatin. Libraries were sequenced on Novaseq sequencer (Illumina) for 100 cycles in paired-end mode.

ChIP-seq Data Analysis. Paired-end reads were trimmed and aligned to the GRCh38 human reference using the STAR (version 2.7.0e) aligner with splicing disabled. The resulting .bam file is duplicate-marked and down-sampled to comparable coverage using picard (version 2.22.3). Peaks for each replicate were called separately with HOMER (version 4.10) with matched inputs before establishing the final peak list with an irreproducible discovery rate (IDR) of 2%. Replicates were then merged and converted into bigwig files for visualization with deepTools2. ChIP peak profile plots and read-density heat maps were generated using deepTools2.

Statistical Methods. Statistical parameters including the exact value and definition of n, the definition of center, dispersion, precision measures (mean ± s.d. or s.e.m) and statistical significance are reported in Figures and Figure Legends.

Chapter 4: Distal drug resistance mutations promote covalent inhibitor-adduct Grob fragmentation in LSD1

This chapter is adapted from a manuscript draft with the following contributions:

Amanda L. Waterbury (A.L.W.), Olivia Zhang (O.Z.), and Paloma R. Tuttle (P.R.T.) designed, performed, and analyzed cell and molecular biology experiments. A.L.W., Jonatan Caroli (J.C.), O.Z., and P.R.T designed, performed, and analyzed protein purification and biochemical assays. P.R.T. designed and synthesized molecules. A.L.W. and J.C. wrote the manuscript draft used in this chapter. Brian B. Liau (B.B.L) edited the manuscript draft used in this chapter. B.B.L and Andrea Mattevi (A.M.) held overall responsibility for the study.

4.1. Abstract

Mutations that prevent drug binding or increase target activity are both common mechanisms of drug resistance. For covalent inhibitors, mutations that block formation of the covalent drug-target bond occur most frequently. Here we report the identification of a novel drug resistance mechanism where a distal mutation promotes the chemical rearrangement of a covalent inhibitor adduct. We identified a small deletion mutation in an allosteric loop in lysine-specific demethylase 1 (LSD1) that induces the covalent FAD inhibitor adduct to undergo a Grob-like fragmentation. This Grob-like fragmentation of the FAD-drug adduct is analogous to the transformation that was reported to occur with the recently identified LSD1 inhibitor, T-448. We further clarify the mechanism of this chemical fragmentation using small molecule analogs of T-448 and structural biology. Altogether, our study uncovers a novel drug-resistance mechanism involving inhibitor covalent adduct fragmentation and reveals new roles of distal sites in mediating resistance.
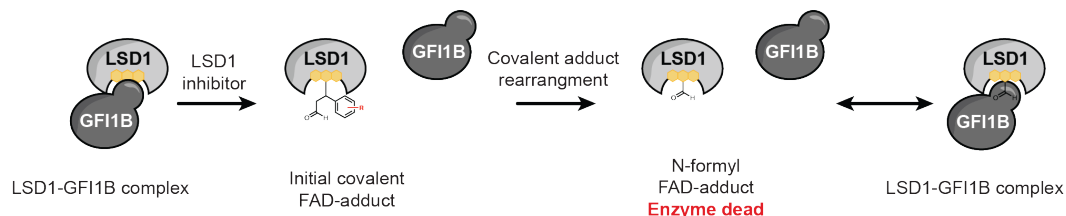
Figure 4.1. Graphical abstract summary of this chapter.

4.2. Introduction

Dysregulation of chromatin modifiers and their associated epigenetic states has been implicated in cancer pathogenesis, prompting the development of small molecule 'epigenetic inhibitors' that target these factors for oncology.[12,165] LSD1, an FAD-dependent H3K4 demethylase, served as an instructive paradigm for the development of epigenetic inhibitors.[35] Beyond its essential role in histone demethylation, the LSD1 active site also serves as a binding pocket for transcription factors (TFs), including the master hematopoietic factors GFI1 and GFI1B. GFI1/GFI1B have an N-terminal SNAG domain that is an H3 N-terminal tail mimic that mediates high-affinity binding with the LSD1 active site.[44,55] The formation of the LSD1-GFI1/GFI1B repressor complex has been shown to be essential for hematopoiesis and its disruption can result in hematological toxicity, such as thrombocytopenia.[35,44,47]

LSD1 is commonly overexpressed in acute myeloid leukemia (AML) and small cell lung cancer (SCLC), prompting the development of inhibitors targeting the demethylase activity of LSD1.[35] FAD-reactive covalent inhibitors targeting the LSD1 active site are currently in clinical trials for various oncology indications, including AML and SCLC.[61,96–99] Despite their original intent to block demethylase activity, LSD1 inhibitors exert their anti-proliferative effects in AML by disrupting the LSD1-GFI1/GFI1B interaction to activate GFI1/GFI1B target enhancers and subsequently induce differentiation (**Figure 4.2**).[70,94,95,98,99] Therefore, these inhibitors have two functions: they (1) inhibit LSD1's demethylase activity and (2) disrupt LSD1-TF interactions, the latter of which is responsible for their antiproliferative effects but also on-target hematotoxicity.
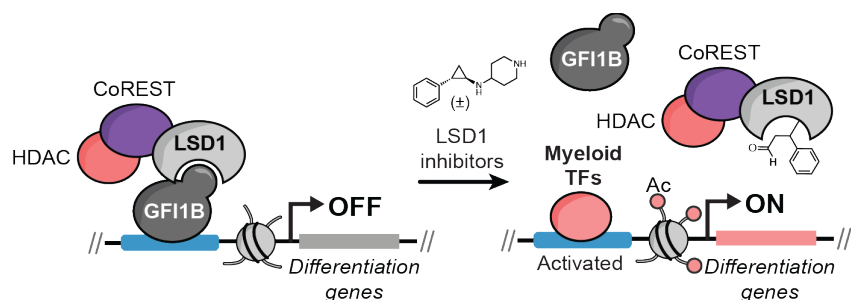
Figure 4.2. LSD1 inhibitors are both demethylase and LSD1-TF inhibitors.

Schematic illustrating the downstream mechanism of LSD1 inhibitors whereby GSK-LSD1, after covalently reacting with the LSD1 FAD cofactor, induces displacement of the LSD1-GFI1B complex along with associated repressor proteins, CoREST and HDAC, from enhancers which allows myeloid TF-mediated activation of genes involved in AML differentiation. Dissociation of the LSD1-GFI1B complex and not inhibition of demethylase activity is responsible for the anti-proliferative effects of LSD1 inhibitors in AML.

This revised mechanism is supported by the remarkable identification of LSD1 inhibitors that inhibit LSD1 activity yet maintain the LSD1-GFI1B interaction.[71,72] These inhibitors, including T448, contain a cyclopropylamine core and form a covalent FAD-drug adduct like other related LSD1 inhibitors.[71] However, this adduct undergoes an unprecedented fragmentation transformation to form *N*-formyl-FAD. While the *N*-formyl-FAD adduct is sufficient to inhibit enzyme activity, it is small enough to accommodate the LSD1-GFI1B interaction. These unique properties make T-448 effective for treating neuroscience applications, where it is desirable to target LSD1's demethylase activity without disrupting the LSD1-GFI1B complex due to concerns regarding the associated hematotoxicity. However, the mechanism of the FAD-drug adduct fragmentation and the structural properties of T-448 that promote this rearrangement remain to be fully determined. Intriguingly, whether mutations in LSD1 may promote this FAD-drug adduct rearrangement with other LSD1 covalent inhibitors raises the possibility of drug resistance mechanisms for oncology.

Here, we demonstrate that FAD-distal resistance mutations to LSD1/GFI1B inhibitors can enhance drug-adduct fragmentation to *N*-formyl FAD. By studying the mechanism of this mutation and chemical analogs of T-448, we identify additional residues and chemical features that enable the FAD-adduct fragmentation to occur. For instance, the FAD-distal mutations contact an alpha helix that extends towards the LSD1 active site and point mutations disrupting these contacts promote the drug-adduct fragmentation. Furthermore, the aromatic amide functionality of T-448, is essential for the transformation to occur as either removing this group entirely or replacing it with a smaller substituent prevents drug-adduct fragmentation. From our structural data, we hypothesize that the aromatic moiety positions the amide to participate in key hydrogen bonds with the covalent FAD-adduct. More broadly, we report how drug resistance can be mediated by mutation-induced chemical fragmentations of a covalent-drug adduct.

4.3. LSD1 loop deletion mutant mediates drug resistance

Using CRISPR-suppressor scanning, we previously identified resistance mutations in LSD1 that were selectively enriched by different analogs of the inhibitor GSK-LSD1.[69] The most highly enriched single-guide RNAs (sgRNAs) targeted exon 15, leading to mutations that altered the catalytic pocket and the disruption of both LSD1 enzyme activity and inhibitor binding (see **Chapter 2**, **Figure 2.9a**, **Figure 2.14d**). However, distinct mutations were also enriched in exon 16 — most notably by AW4 (**Figure 2.10a, Fig 4.3a,b**). Genotyping exon 16 of the AW4-treated cells revealed that the most prevalent mutation comprises a 4-amino acid deletion T684_S687 (LSD1 TTASdel), in a loop >18 Å from the FAD cofactor (**Figure 4.3a**). A clonal SET-2 cell line harboring LSD1 TTASdel displayed the highest level of resistance to AW4 compared to LSD1 inhibitors GSK-LSD1 and NV93, consistent with the CRISPR-suppressor scan (**Figure 4.3b**).[97,103] Wild-type SET-2 as well as the active site resistance mutations harboring basic amino acid insertions are shown for comparison (see **Chapter 2** for more details). Modeling LSD1 TTASdel did not provide a clear picture for how truncation of this distal loop mediates drug resistance and the TTAS loop is involved in crystal contacts preventing structural characterization by x-ray crystallography.
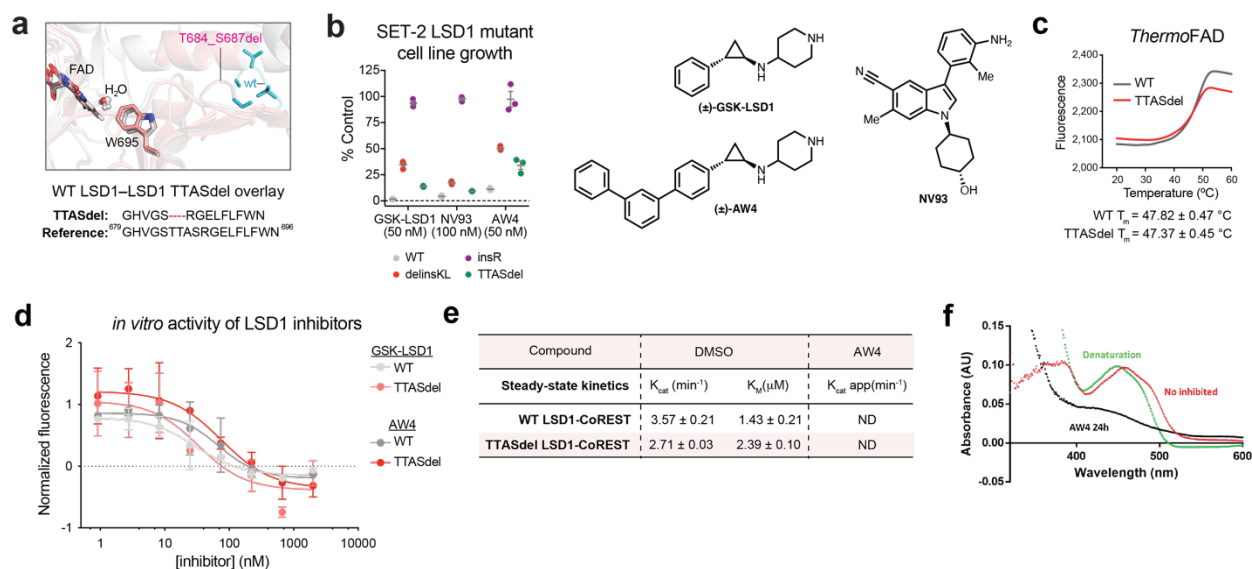
Figure 4.3. LSD1 TTASdel mediates drug resistance in cells yet LSD1 inhibitors bind in vitro.

a, Homology model of LSD1 TTASdel (salmon) overlaid with WT LSD1 (gray) is shown to highlight the deletion of the WT loop (cyan) compared to the truncated mutant loop (magenta). PDB: 2hko. **b,** SET-2 cells containing LSD1 mutations display differential sensitivity to LSD1 inhibitors (left). Chemical structure of the LSD1 inhibitors tested in this panel are shown here (right). GSK-LSD1 and AW4 were used as racemic mixtures in the remainder of these studies. **c,** LSD1 TTASdel displays comparable binding to FAD and overall protein stability as measured by *Thermo*FAD. **d,** Dose response curves of recombinant WT LSD1- and LSD1 TTASdel-CoREST protein treated with GSK-LSD1 or AW4. **e,** Steady-state kinetics of WT LSD1- and LSD1 TTASdel-CoREST protein treated with DMSO or AW4. **f,** Absorbance spectra of LSD1 TTASdel-CoREST (10 μM, red trace) after incubation with 40 μM of AW4 for 24 hours (black trace). The protein was then unfolded with urea. The spectrum (green) indicates that the inhibitor-FAD adduct is reversed upon release from the protein.

Many of the active-site drug-resistant mutants previously identified disrupt both LSD1 enzyme activity and inhibitor binding.[69] Consequently, we tested whether LSD1 TTASdel mediates resistance by a similar mechanism. Purification and testing LSD1 TTASdel confirmed that the LSD1 variant binds FAD (**Figure 4.3c**). Recombinant LSD1 TTASdel was then complexed

to truncated CoREST (CoRESTΔ1-305) and was enzymatically active in H3K4me2 demethylation assays and inhibited by both GSK-LSD1 and AW4, similar to wild-type LSD1 (**Figure 4.3d**). Moreover, the TTASdel mutation had no impact on the kinetics of demethylation and no demethylase activity was recovered after inhibitor treatment over time, even over the course of days (**Figure 4.3e**). Consistent with this, by monitoring the absorbance of the FAD cofactor, AW4 fully bleached the oxidized flavin peak (458 nm) within the first few minutes of inhibitor incubation for LSD1 TTASdel, analogous to that reported for other tranylcypromine LSD1 inhibitors, and the flavin spectra remained bleached over time (**Figure 4.3f**).[62,95] These results suggest that LSD1 TTASdel remains inactivated by AW4. Therefore, LSD1 TTASdel does not block drug binding nor disrupt enzyme activity, differentiating this mutant from the active-site drug-resistant mutants.

      LSD1 inhibitors disrupt the LSD1-GFI1B complex to induce AML differentiation. Consequently, we considered whether AW4 might not interfere with LSD1 TTASdel-GFI1B binding despite forming a covalent FAD-drug adduct. To investigate this notion, we monitored binding of LSD1-CoREST complexes to a fluorescently labelled GFI1B(1-9) peptide in the presence or absence of AW4 (**Figure 4.4a**).[69,166] In agreement with prior studies, we confirmed that the GFI1B peptide binds tightly to LSD1-CoREST (K_d ~ 15 nM) (**Figure 4.4b,c**). GSK-LSD1 and AW4 blocked GFI1B peptide binding to wild-type LSD1-CoREST. However, GFI1B peptide binding to LSD1 TTASdel-CoREST was unexpectedly observed at later timepoints in the presence of AW4 but not GSK-LSD1. Collectively, these experiments demonstrate that while AW4 is a potent inhibitor of both wild-type and LSD1 TTASdel, forming a stable covalent adduct with FAD as gathered from the UV-vis spectral perturbations and demethylase activity assays, time-dependent GFI1B binding is restored selectively with LSD1 TTASdel in the presence of AW4 (**Figure 4.4b,c**).
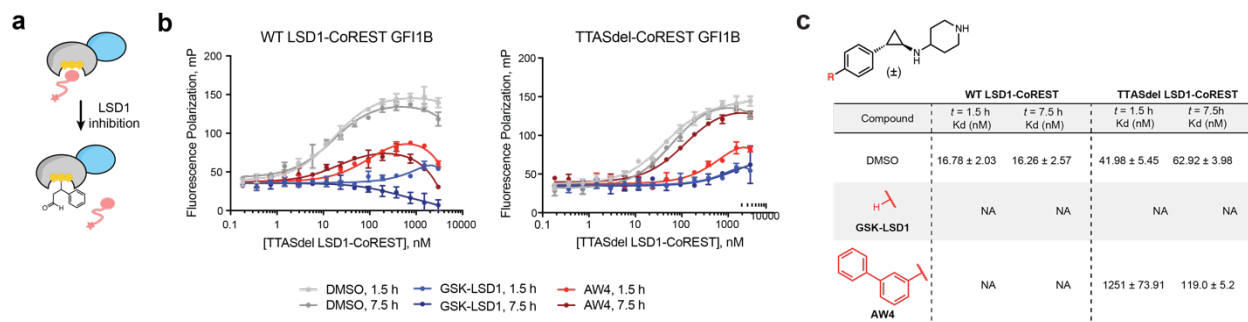
Figure 4.4. LSD1 TTASdel binds to the GFI1B peptide in the presence of AW4 in a time-dependent manner.

a, Schematic of fluorescence polarization binding assay to measure binding affinity of LSD1 to a fluorescently labeled GFI1B(1-9) peptide upon LSD1 inhibition. **b,** Binding curves showing fluorescence polarization (*y*-axis) for increasing concentrations of LSD1-CoREST (*x*-axis) in the presence of a fluorescently labeled GFI1B(1-9) peptide after treatment with DMSO, GSK-LSD1 or AW4. **c,** Table summarizing the K_d values for the FP binding curves in (**b**).

To investigate how LSD1 TTASdel-GFI1B binding occurs in the presence of AW4, we measured the formation and stability of the covalent FAD-drug adduct by LCMS.[71] Notably, after 24 h of AW4 treatment we observed the presence of *N*-formyl-FAD in addition to the expected FAD-drug adduct (**Figure 4.5a,b**). *N*-formyl-FAD was not detected with wild-type LSD1-CoREST treated with AW4 under the same conditions.
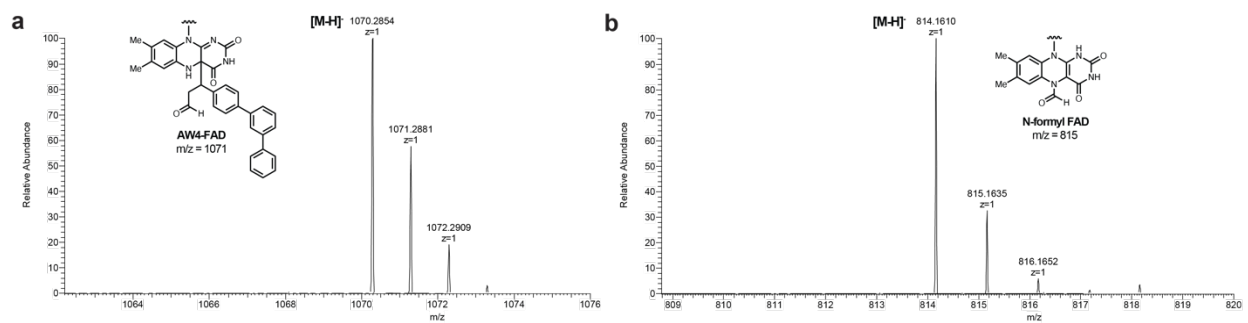
Figure 4.5. LSD1 TTASdel binds to AW4 and the resulting FAD-AW4 adduct fragments to N-formyl FAD by LC/MS.

We validated that the *N*-formyl FAD adduct detected was identical to that generated by T-448 (**Figure 4.6a**), and that T-448 also leads to time-dependent GFI1B peptide binding for wild-type LSD1 (**Figure 4.6b**). Altogether, these results demonstrate that TTASdel selectively promotes *N*-formyl FAD adduct formation after treatment with AW4. As T-448 is ineffective in blocking AML proliferation due to maintaining LSD1-GFI1B binding (**Figure 4.6c**), these findings suggest that *N*-formyl FAD adduct formation may in part drive resistance to AW4 by the LSD1 TTASdel deletion.
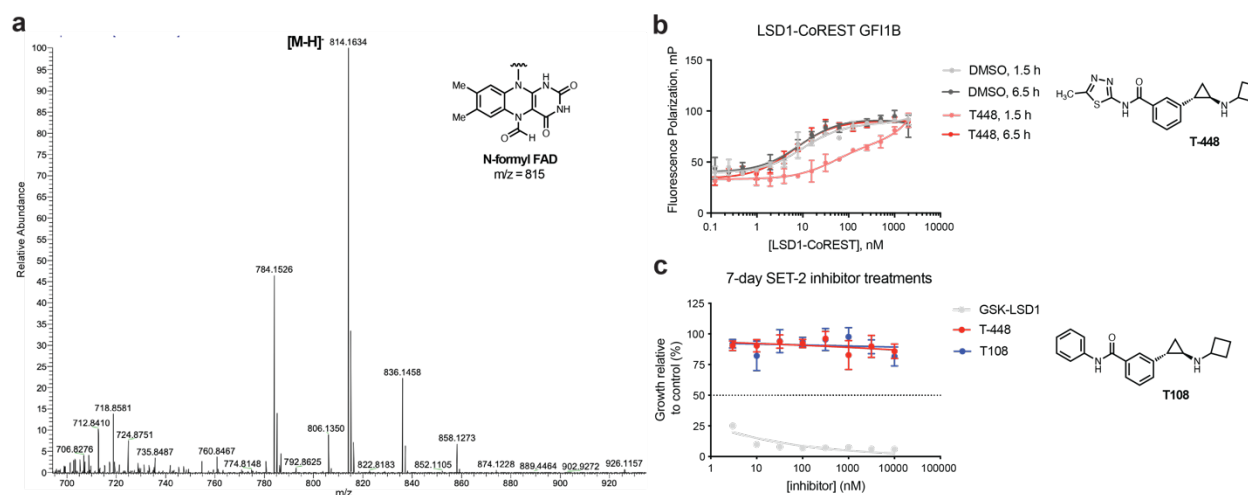
Figure 4.6. Validation of T-448 FAD adduct fragmentation to *N*-formyl FAD to maintain LSD1-GFI1B binding.

a, LC/MS spectra of FAD-adducts after extracting recombinant LSD1 treated with T-448 for 4 h. **b,** Binding curves showing fluorescence polarization (*y*-axis) for increasing concentrations of LSD1-CoREST (*x*-axis) in the presence of a fluorescently labeled GFI1B(1-9) peptide after treatment with DMSO or T-448 (structure shown to the right; the racemate is used for T-448 and analogs in the text). **c,** Dose-response curves for inhibitors tested in SET-2, an AML cell line, are shown. T-448 and T108 (structure shown to the right) fail to inhibit SET-2 growth at concentrations where GSK-LSD1 completely abolishes growth. Data represent

Figure 4.6 (continued) ± s.e.m across three technical replicates. One of two independent replicates is shown.

4.4 Mechanistic investigation of FAD-drug adduct fragmentation

Despite the underlying activity and therapeutic index of T-448, the mechanism of *N*-formyl-FAD adduct formation remains unclear. We considered whether the initial covalent FAD drug-adduct may undergo a Grob-like fragmentation via the 5-membered cyclic hemi-aminal to yield *N*-formyl FAD and the corresponding styrene (**2**) (**Figure 4.7a**).[167,168] To test this hypothesis, we first synthesized a deuterated T-448 derivative, *d*-T-448 (**1**), in which deuterium is incorporated on the indicated carbon of the cyclopropyl carbon predicted to form the formyl group in the proposed Grob-like fragmentation mechanism (**Figure 4.7a**). Consistent with this proposed mechanism, deuterated *N*-formyl-FAD was detected by LC-MS after treating LSD1-CoREST with *d*-T-448 (**Figure 4.7b**). Furthermore, formation of the *N*-formyl adduct was accompanied by detection of styrene **2** by LC-MS, with increasing styrene conversion observed over longer time points (**Figure 4.7c,d**). The MS trace of **2** was validated using an authentic synthesized standard. The AW4 styrene (**3**) was also detected at increasing levels over time after LSD1 TTASdel was treated with AW4 (**Figure 4.7e,f**). Altogether, these results are consistent with the rearrangement occurring through a Grob-like fragmentation.
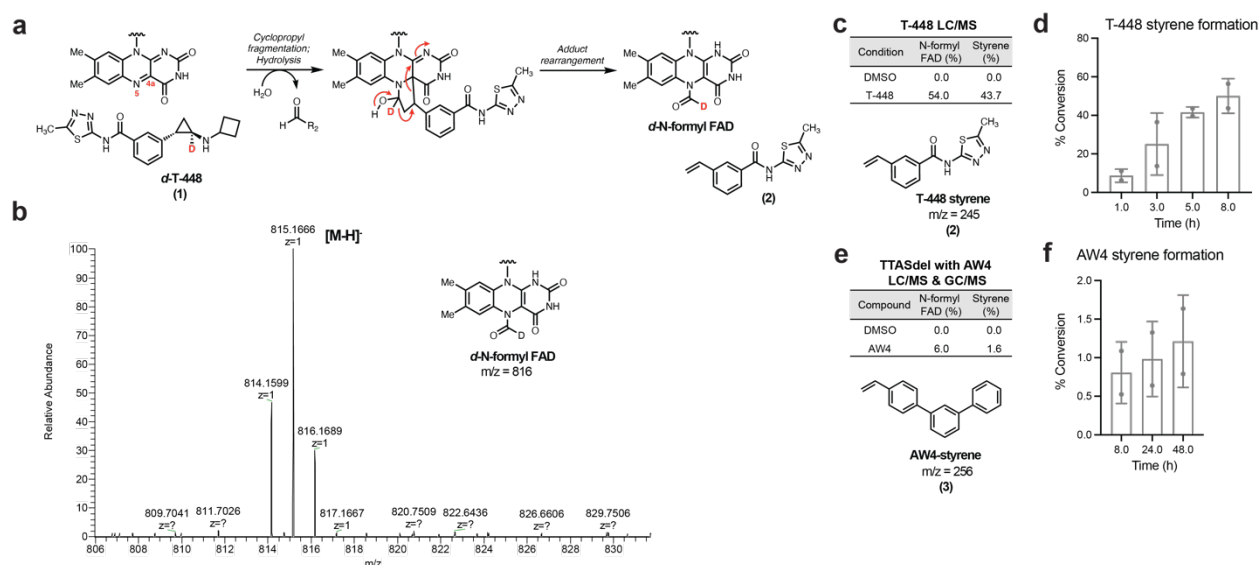
Figure 4.7. *N*-formyl FAD is formed through a Grob-like fragmentation.

Figure 4.7 (continued) a, Schematic of the hypothesized mechanism of covalent FAD-drug adduct fragmentation via a Grob-like fragmentation. Deuterium (red) is incorporated into T-448 resulting in *d*-N-formyl FAD after fragmentation. **b,** LC/MS spectra of FAD-adducts after extracting recombinant LSD1 treated with *d*-T-448 for 4 h. **c,** Table summarizing *N*-formyl FAD and styrene (**2**) detected by LC/MS after extracting recombinant LSD1 treated with T-448 for 8 h. **d,** Formation of styrene (**2**) detected by LC/MS after extracting recombinant LSD1 treated with T-448 at the indicated time points. **e,** Table summarizing *N*-formyl FAD detected by LC/MS and styrene (**3**) detected by GC/MS after extracting recombinant LSD1 TTASdel-CoREST treated with AW4 for 24 h. **f,** Formation of styrene (**3**) detected by GC/MS after extracting recombinant LSD1 treated with T-448 at the indicated time points.

We next evaluated the chemical features of T-448 to understand why this compound readily produces *N*-formyl-FAD whereas other tranylcypromine compounds, such as GSK-LSD1, do not. We synthesized derivatives of T-448, altering both the aryl substituent and the amide, to determine whether the 1,3,4-thiadiazole and amide groups are necessary for the FAD-drug adduct fragmentation (**Figure 4.8a**). Moving the thiadiazole from the meta to the para position in T14 prevented formation of *N*-formyl-FAD as assessed by LC/MS, suggesting that the meta orientation is essential for this fragmentation to proceed (**Figure 4.8a**). Replacing the thiadiazole with a methyl ester (T15) or dimethyl amide (T17) abolished formation of *N*-formyl-FAD. While further modification to the methyl amide (T16), which preserves the secondary amide, almost entirely prevents formation of *N*-formyl-FAD, replacing the thiadiazole substituent with a benzamide (T108) produces *N*-formyl-FAD, although likely at a slower rate compared to T-448 (**Figure 4.8a, Figure 4.9a**). Inverting the amide in T108 to T18 further reduces formation of *N*-formyl-FAD (**Figure 4.8a**). Of note, most modifications to the T-448 scaffold also substantially reduce inhibitor potency, suggesting that some of these modifications may not be well tolerated for forming the initial covalent adduct.
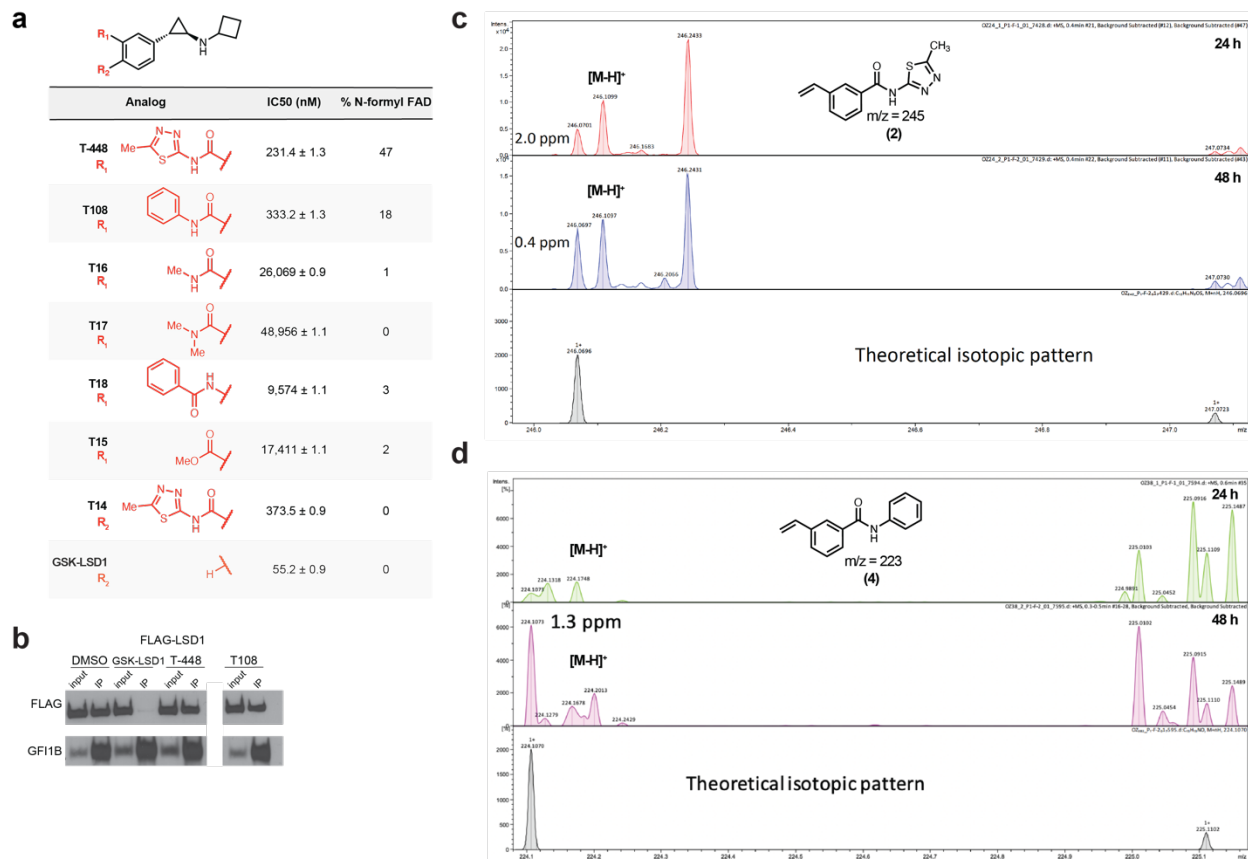
Figure 4.8. T-448 analogs impede Grob-like fragmentation.

a, Table summarizing the enzyme inhibition (IC50) and fragmentation activity (*N*-formyl FAD by LC/MS) of T-448 analogs. b, Co-IP of wild-type FLAG–LSD1 with GFI1B was performed after vehicle or inhibitor treatment (500 nM, 48 h) in transiently transfected HEK 293T cells. Co-IP was performed using an anti-GFI1B antibody. LSD1 was detected using an anti-FLAG antibody. One of two independent replicates is shown. **c,d**, GC/MS analysis of SET-2 media after inhibitor treatment. SET-2 cell were treated with 1 μM T-448 (**c**) or 1 μM T108 (**d**) for 24 h or 48 h. The media was collected, extracted with dichloromethane, and analyzed by GC/MS.

Despite detecting less *N*-formyl-FAD formation in vitro with T108 in comparison to T-448, T108 treatment had no effect on AML cell growth (**Figure 4.6c.**). Consistent with this observation, both T-448 and T108 treatment do not disrupt the LSD1-GFI1B interaction in cells (**Figure 4.8b**). Moreover, both T-448 (**2**) and T108 (**4**) styrenes were detected in extracted cell media after

treating SET-2 cells with either compound, supporting that the fragmentation occurs in cells (**Figure 4.8c,d**). Furthermore, T108 styrene (**4**) was detected by LC/MS along with the T108-FAD and *N*-formyl-FAD adducts, consistent with the proposed Grob-like fragmentation mechanism (**Figure 4.9a-c**). Overall, these data suggests that although modifications to the aryl substituent in T-448 are tolerated, the meta substituted thiadiazole group is privileged in its ability to readily produce *N*-formyl-FAD. Furthermore, although a small % *N*-formyl-FAD is detected for some analogs in vitro, like T108, these levels might be sufficient to maintain enough LSD1-GFI1B complex in cells to sustain proliferation.
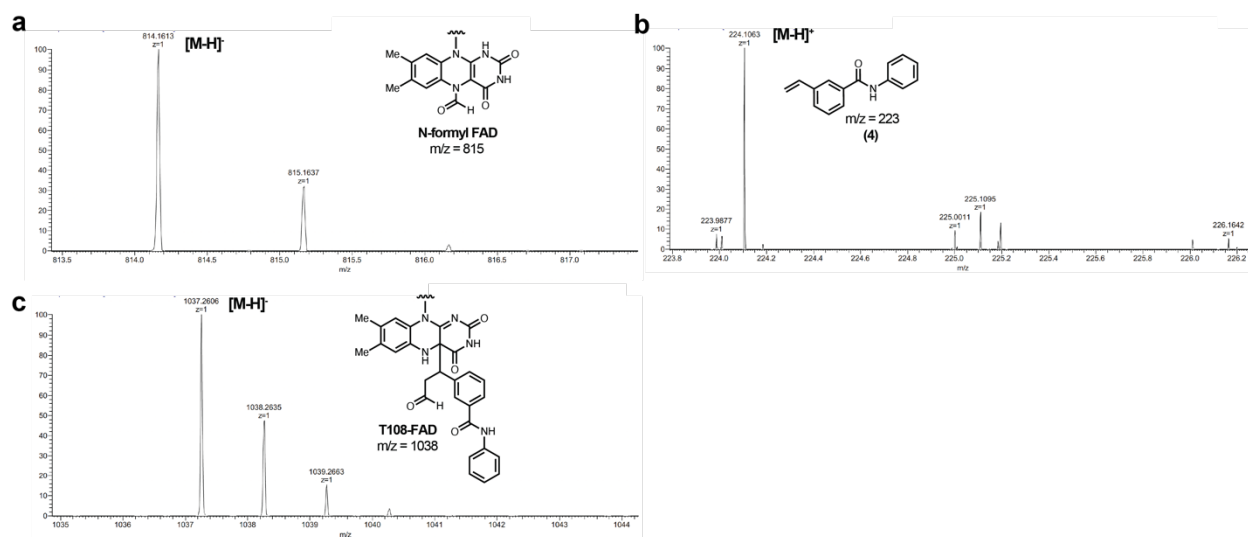
Figure 4.9. T108-FAD adduct undergoes Grob-like fragmentation.

a, LC/MS spectra of FAD-adducts after extracting recombinant LSD1 treated with T108 for 24 h where *N*-formyl FAD is detected in (**a**) as well as the styrene by-product (**b**). The full adduct is also detected, consistent with T108 FAD inhibition and slower fragmentation of the covalent adduct to *N*-formyl FAD.

4.5. Structural studies

To gain insights into the structural changes that may promote the FAD-drug rearrangement, we crystallized LSD1-CoREST with T-448, T108, T16, T14 and AW4. Unfortunately, TTASdel LSD1-CoREST was refractory towards crystallization since the TTAS loop itself is essential for mediating crystal contacts.

To determine how the FAD-drug adduct changes over time, LSD1-CoREST crystals were initially soaked with T-448 or T108 and then back-washed into solution lacking inhibitor and soaked with 1 mM of the SNAG peptide prior to flash-freezing (**Figure 4.10a**). T-448- and T108-inhibited FAD adducts exhibit a distinct conformation, where the cyclic adduct points the aromatic moiety away from the ribityl-flavin in a staggered orientation compared to the classic FAD-drug N5 and C4a cyclic adduct that positions the aromatic ring on the *Re-* side of flavin (**Figure 4.10b**).[62,169,170] The T-448-FAD adduct was fully converted to *N*-formyl FAD after three days in agreement with data from Takeda, allowing LSD1 to bind the SNAG peptide in the same conformation as the native protein (**Figure 4.10a**).[71] For T108, the conversion to *N*-formyl FAD was likely too slow under crystallization conditions and only the full T108-FAD adduct was observed after 3 days.
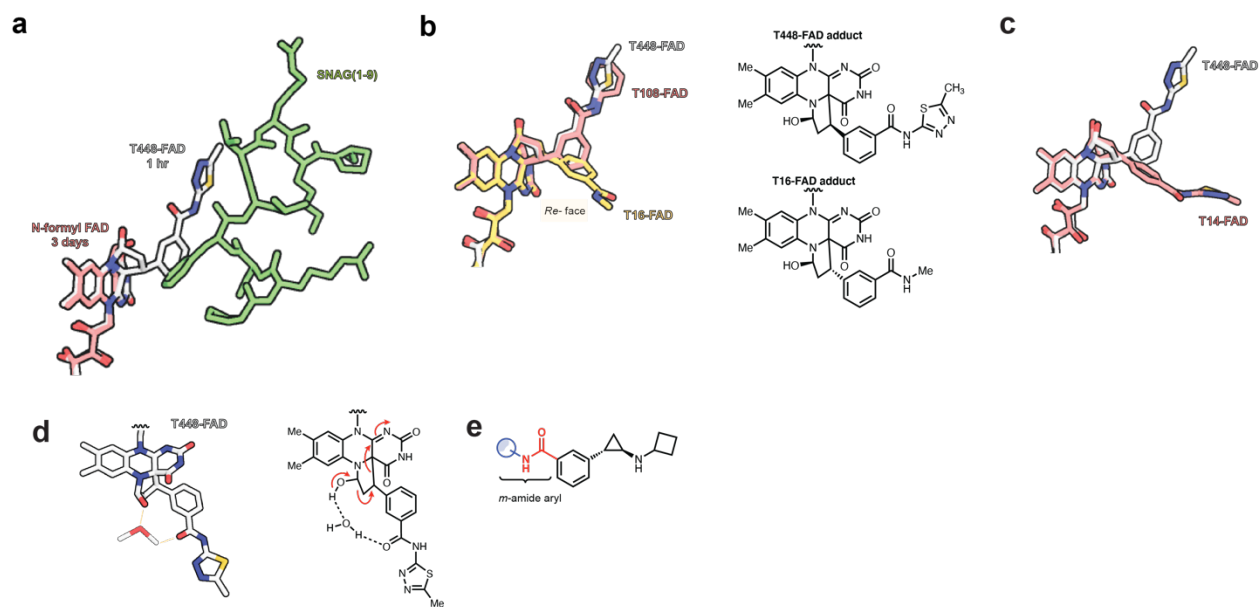
Figure 4.10. Structural insight into FAD-drug adduct fragmentation.

a, Structural overlay of the initial T448-FAD adduct with *N*-formyl FAD bound to the SNAG peptide from the LSD1-CoREST SNAG co-crystal structure. **b,** Overlay of the T448-, T108-, and T16-FAD adducts after soaking LSD1-CoREST crystals in solution with the respective inhibitors (left), highlighting the contrasting conformation and stereochemistry of the 5-membered cyclic adduct adopted by T448-FAD (and T108) compared to T16 bound FAD (right). T448-FAD is in gray, T108-FAD in salmon and T16-FAD in yellow. **c,** Overlay of T-448 and T14 bound FAD highlighting the different conformation and stereochemistry each respective covalent adduct adopts. T448-FAD is in gray while T14-FAD is in salmon. **d,** Illustration of the predicted hydrogen bonding network between the hydroxyl of the 5-membered cyclic covalent FAD adduct, an active site water molecule and the carbonyl of the inhibitor amide (left). This hydrogen bonding network is predicted to promote FAD-drug adduct Grob-like fragmentation (right). **e,** Chemdraw highlighting the structural rmodalities that promote FAD-drug adduct fragmentation.

In comparison, T16- and T14-FAD adducts adopt the classic FAD-drug N5 and C4a cyclic adduct that positions the aromatic ring on the *Re*- side of flavin (**Figure 4.10b,c**). Furthermore, an overlay of the T16- and T14-FAD adducts with the T-448 covalent FAD adduct suggests that the corresponding 5-membered cyclic intermediates of these analogs adopt the opposite

stereochemistry of T-448 and T108 FAD-drug adducts (**Figure 4.10b**). Interestingly, the distinct

conformation of the initial FAD adducts formed with T-448 and T108 suggests that the amide

modality might engage in a hydrogen bond network with an active site water molecule and the -

OH of the 5-membered ring adduct (**Figure 4.10d**). Both compounds share a common meta

secondary amide aryl (*m*-amide aryl) substituent on the phenyl group of the inhibitor, which we

hypothesize to provide a favorable orientation and H-bonding pattern to undergo adduct

fragmentation to yield *N*-formyl FAD and styrene compared to T16 and T14, which lack the

aromatic group off the amide or the meta positioning, respectively (**Figure 4.10d,e**).
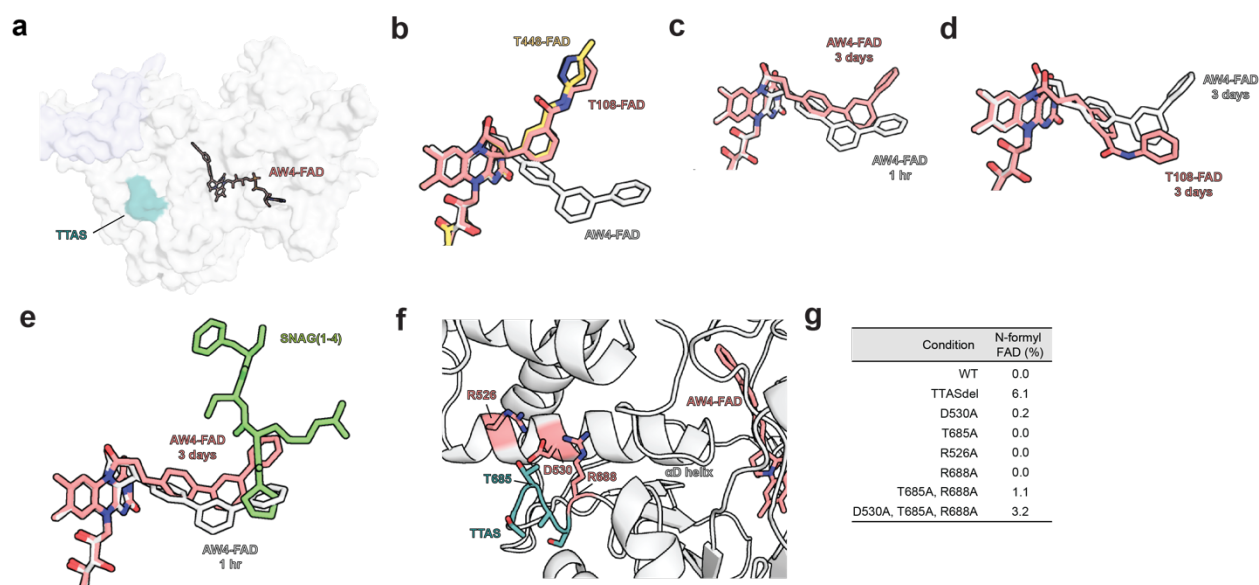
Figure 4.11. Structural and mutagenesis studies towards elucidating AW4-FAD fragmentation mechanism.

a, Structural view of LSD1-CoREST with the covalent AW4-FAD adduct (salmon) where the location of the

TTAS loop is highlighted in cyan. **b,** Overlay of the initial T448-, T108-, and AW4-FAD adducts after soaking

LSD1-CoREST crystals in solution with the respective inhibitors. T448-FAD is in yellow, T108-FAD in

salmon and AW4-FAD in gray. **c,** Overlay of the AW4-FAD adduct after initial inhibitor soaking (gray) and

the backwashed structure after 3 days (salmon). **d,** Overlay of the AW4- and T108-FAD adducts formed in

the back-washed structures highlighting formation of the linear covalent adduct after 3 days. **e,** Structural

overlay of the initial AW4-FAD adduct with AW4 FAD bound to the SNAG peptide from the LSD1-CoREST

Initial experiments involving soaking the crystals prior to flash-freezing revealed that AW4 forms the N5 and C4a cyclic adduct with the opposite stereochemistry and conformation in comparison to T-448 and T108 inhibited FAD. These results are consistent with AW4 lacking the *m*-amide aryl substituent and failing to form *N*-formyl FAD with wild-type LSD1. (**Figure 4.11a,b**). To see how the FAD-drug adduct changes over time, crystals were soaked and then back-washed into solution lacking inhibitor and either directly flash-frozen after three days or soaked with 1 mM of the SNAG peptide prior to flash-freezing. Compared to the initial structure, the AW4-FAD adduct in the back-washed structure that was flash-frozen after three days undergoes a conformational change involving a moderate lateral swing of the inhibitor aromatic group (**Figure 4.11c**). This new conformation likely arises from rearrangement of the FAD-drug adduct, where the initially formed 5-membered cyclic adduct rearranges to an open adduct between the inhibitor and flavin N5, and this was also observed with the T108 structure acquired under similar conditions (**Figure 4.11d**). Consequently, the inhibitors are free to rotate about this single bond in a relaxed conformation.

Incubation of the AW4 back-washed structure with the SNAG peptide yielded a contrasting SNAG binding mode compared to the native protein where the SNAG peptide is more externally located in the active site cleft (**Figure 4.11e**). Comparison of this SNAG bound structure with the initial short-inhibitor soaking structure shows that the SNAG peptide would be predicted to clash with the rigid conformation adopted by the 5-membered cyclic FAD-drug adduct. Opening this cyclic adduct appears to allow the inhibitor to adopt a more flexible orientation capable of interacting with the SNAG peptide, however, given the unconventional binding mode observed

with the SNAG peptide, we predict that this interaction is driven by electrostatic interactions and may not reflect binding under physiological conditions.[56]

When examining the LSD1-CoREST structure in more detail, residues in the TTAS loop, T685 and R688, are predicted to participate in an H-bonding network with residues in the αD helix of the amine oxidase domain, R526 and D530 (**Figure 4.11f**). We thus considered if disrupting this H-bonding network with point mutations could recapitulate the effects of the LSD1 TTAS loop deletion and mediate AW4-FAD fragmentation to *N*-formyl-FAD. Single point mutants to alanine were overall not sufficient to induce AW4-FAD transformation to formylated FAD (**Figure 4.11g**). However, the double (T685A, R688A) and triple alanine mutants (D530A, T685A, R688A) produced *N*-formyl-FAD at nearly comparable levels as LSD1 TTASdel (**Figure 4.11g**). These results are consistent with the hypothesis that the αD helix mediates crosstalk between the TTAS loop and active site as well as suggests that TTASdel may reposition the αD helix to mediate AW4-FAD fragmentation.

4.6 Current efforts and future directions

Current efforts are aimed at further validating the TTASdel mutation in cells as well as investigating the FAD-adduct transformation with T-448 analogs. Towards investigating the mechanistic basis of resistance by the TTASdel loop deletion, we transduced the top-enriched exon 16 sgRNA, sgT684, into SET-2 cells and selected for AW4 resistance by treating with 50 nM AW4 for 7 weeks. The resulting cells were then expanded, single-cell sorted, and will be genotyped to isolate clonal LSD1 TTASdel SET-2 cells. Genotyping the bulk cells at 7 weeks revealed the presence of TASdel, a similar mutation, which we will also isolate and validate after single-cell sorting. Resistance to AW4 will be assessed by growth as well as by CD86-staining, a myeloid differentiation mark upregulated upon LSD1 inhibition.[70,144] Whether AW4 disrupts the LSD1-GFI1B interaction in LSD1 TTASdel SET-2 cells will then be assessed by co-IP as well as whether the AW4 styrene (**3**) is generated in cells with AW4 treatment.
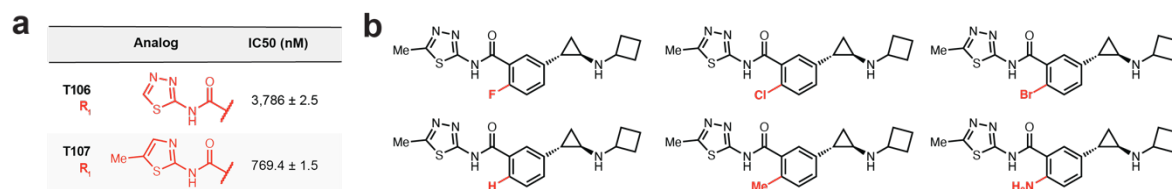
Figure 4.12. T-448 analogs for further characterization and kinetic studies.

a, Synthesized T-448 analogs that will be further tested biochemically. **b,** T-448 analogs for Hammett studies to investigate the kinetics of the FAD-drug fragmentation.

While we have extensively characterized T-448 and T108 biochemically, further characterization of the inhibitors listed in **Figure 4.8a** will be pursued. An outstanding question is whether these inhibitors, such as T16 which produces 1 % *N*-formyl FAD by LC/MS (**Figure 4.8a**), dissociate the LSD1-GFI1B complex. This will be assessed by fluorescence polarization assays with recombinant LSD1-CoREST and a fluorescently labeled GFI1B(1-9) peptide, and co-IP experiments in cells as described previously. We have also synthesized additional T-448 analogs

by modifying the thiadiazole ring, which will likewise be characterized (**Figure 4.12a**). Furthermore, to investigate the electronic requirements for the FAD-drug fragmentation, we will conduct a Hammett analysis to determine whether there is a buildup of negative charge upon adduct decomposition. Towards this end, we are synthesizing and validating T-448 analogs with varying electron withdrawing or donating groups installed at the para position of the tranylcypromine scaffold (**Figure 4.12b**). We will monitor *N*-formyl FAD formation over time after treating recombinant LSD1 with these inhibitors and perform a Hammett analysis to elucidate whether the kinetics of the covalent adduct decomposition is influenced by changing the electronic properties of T-448. No change in the kinetics of *N*-formyl FAD formation would suggest that it proceeds by a single electron mechanism, which would be consistent with flavin chemistry. Overall, these additional experiments will further our mechanistic analysis of the LSD1 TTASdel mutation and how T-448 inhibited FAD decomposes to *N*-formyl FAD.

4.7. Discussion

Drug resistance is a major challenge in the development of anticancer therapeutics. CRISPR-suppressor scanning can accelerate the identification of drug resistance conferring mutations, the mechanistic deconvolution of which can also reveal new aspects of target biology.[69,86,88,171] Here we identify a unique mechanism of drug resistance whereby the chemical fragmentation of a covalent drug is promoted by a mutation distal to the drug binding site. Moreover, this mutation unexpectedly promotes a chemical fragmentation pathway of the drug covalent-adduct that also occurs with mechanistically unique inhibitors, such as T-448, which are currently in clinical development for neuroscience applications.[71,72,172] Consequently, regardless of whether these specific resistance mutations might arise in the clinic, they are broadly informative about drug target pharmacology and reveal new thematic possibilities for mechanisms of drug resistance.[171]

Our experimental data clarify the mechanism whereby the T-448 FAD adduct fragments to *N*-formyl FAD through a Grob-like fragmentation. The deuteration experiment assessing the formation of deuterated formyl-FAD as well as styrene by-product by LC/MS support this mechanism. Through SAR studies we demonstrate that the amide and thiadiazole functional groups of T-448 greatly accelerate the chemical fragmentation relative to other tranylcypromine derivatives. However, co-IP and cell growth experiments with T-448 and T108 suggests that the ability to form even a small amount of *N*-formyl FAD may be sufficient to maintain enough intact LSD1-GFI1B complex in cells to sustain proliferation. These results further support our proposed resistance mechanism of the TTASdel mutation, which generates modest levels of *N*-formyl FAD.

By comparing the structures of LSD1 crystallized with various inhibitors, we reveal that the initial cyclic drug-FAD adducts generated by T-448 and T108 are distinct from other FAD-drug structures. Furthermore, both compounds share a common meta -C(O)NH-R substituent, which we hypothesize to provide a favorable orientation and H-bonding pattern to undergo adduct

fragmentation to yield *N*-formyl FAD and styrene. Overall, the ability of the FAD-drug adduct to undergo this fragmentation depends on an amide substituent on the meta position of the aryl ring.

LSD1 TTASdel facilitates AW4-FAD adduct decay to enable recovery of SNAG binding in a time-dependent manner and confer resistance to AW4 in cells. We predict that this may in part be mediated through contacts with the αD helix of the amine oxidase domain as residues in the TTAS loop are predicted to H-bond with residues in this helix. Double and triple point mutations disrupting this H-bond network promote AW4-FAD fragmentation and suggests that the αD helix may mediate crosstalk between the TTAS loop and active site. Interestingly, the TTAS loop is within close spatial proximity (~20 Å) to a previously characterized neuron-specific splicing variant of LSD1 (LSD1+8a) that has four additional amino acid residues within a loop in its AOD.[173] This LSD1 isoform is essential for development and differentiation in the brain and its repressive activity is regulated by phosphorylation of a threonine residue in this loop that causes LSD1 to dissociate from CoREST and HDAC1/2 complex members.[173–175] Therefore, small changes in distal LSD1 loops, like with the TTASdel mutation or LSD1+8a, can have significant and functional implications on LSD1 activity, highlighting these regions as potential sites of allosteric control.

Overall, we have clarified the mechanism whereby FAD-distal mutations confer resistance to LSD1/GFI1B inhibitors by enhancing fragmentation of the FAD-drug adduct, a chemical transformation that was only previously observed with demethylase specific inhibitors such as T-448. Moreover, we elucidate the structural features of the protein and inhibitor that promote this fragmentation in the context of LSD1 TTASdel and T-448, respectively. We expect that these insights will guide future efforts targeting LSD1 and will more broadly inform us on new mechanisms of drug resistance through covalent-drug adduct fragmentation.

4.8. Methods

Reagents. Compounds were stored at -80 °C in 100% DMSO and working aliquots were stored at -20 °C. The 100 mM concentration stocks for soaking experiments were in 100% DMSO while the diluted stocks at 4 mM were in 50% DMSO. The vehicle condition represents 0.1% (v/v) DMSO treatment. GSK-LSD1 was purchased from Sigma-Aldrich (≥98% by HPLC).

Chemical Synthesis. LSD1 inhibitors were synthesized as described previously.[69,71]

Cell Culture and Lentiviral Transduction. SET-2 was a gift from Matthew D. Shair; HEK 293T was a gift from Bradley E. Bernstein. All cell lines were authenticated by Short Tandem Repeat profiling (Genetica) and routinely tested for mycoplasma (Sigma-Aldrich). All media were supplemented with 100 U/mL penicillin and 100 µg/mL streptomycin (Life Technologies) and fetal bovine serum (FBS, Peak Serum). All cell lines were cultured in a humidified 5% CO_2 incubator at 37 °C. SET-2 were cultured in RPMI-1640, 20% FBS. HEK 293T were cultured in DMEM (Life Technologies), 10% FBS. Lentivirus was produced by co-transfecting HEK293T cells with pCMV-VSV-G (Addgene plasmid #8454), psPAX2 (Addgene plasmid #12260) and transfer vector plasmid (encoding gene of interest) using Lipofectamine 3000 (Invitrogen L3000) according to manufacturer's protocol. Media was exchanged after 6 h and the viral supernatant was collected 48–72 h after transfection and filtered (0.45 µm). Transduction was carried out by mixing the virus with cells with 8 µg/ml polybrene for SET-2 and HEK293T. SET-2 were centrifuged at 1,800 x *g* for 90 min at 37 ºC while HEK293T were directly incubated. After 48 h post-transduction, media was changed and puromycin (Thermo Fisher Scientific) selection was carried out for 5 d at 2 µg/mL for both cell lines.

Immunoblotting. For whole cell extracts, cells were lysed on ice using radioimmunoprecipitation assay (RIPA) buffer (Boston BioProducts) supplemented with fresh HALT™ protease inhibitor

cocktail (Thermo Fisher Scientific). The lysate was clarified by centrifugation at 10,000 *x g* for 10 min. Protein concentration of the lysates was measured using BCA Protein Assay Kit (Thermo Fisher Scientific). Immunoblotting was performed according to standard procedures. The primary antibodies used are as follows: LSD1 (Bethyl Laboratories A300-215A), LSD1 (Abcam ab17721). GFI-1B (B-7) X (Santa Cruz Biotechnology sc-28356X, LOT# D1615). Monoclonal Anti-FLAG M2 (Sigma-Aldrich F1804). GAPDH (Santa Cruz Biotechnology sc-477724). GFI1B (Santa Cruz Biotechnology sc-28356X).

Co-immunoprecipitation. For Co-IPs performed in HEK 293T cells, cells were plated at 40-60% confluency and treated with 500 nM of the respective LSD1 inhibitor or vehicle. 24 h after plating, cells were co-transfected with 1.5 µg pcDNA.3 FLAG-LSD1 plasmid and 2 µg pcDNA.3 WT GFI1B-FLAG using Lipofectamine 3000 reagent (Invitrogen L3000). 48 h post-transfection, cells were washed with PBS and then flash frozen. Cells were thawed, lysed on ice in N450 buffer (50 mM Tris-HCl pH 7.5, 450 mM NaCl, 1 mM $MgCl_2$, 1% NP-40 alternative, 5% glycerol) supplemented with 1:10,000 benzonase (Sigma-Aldrich), and the lysates were cleared. The protein concentration was quantified as above and diluted to 1 mg/mL in lysis buffer. Supernatants were immunoprecipitated overnight at 4 ºC with 4 µg GFI1B antibody and 40 µL Protein G Dynabeads (ThermoFisher Scientific). Beads were washed twice with lysis buffer, eluted in SDS-PAGE loading buffer, and carried forward to immunoblotting as described above.

Generation of Clonal Drug-Resistant LSD1 Mutant Cell Lines. sgRNA sequence targeting LSD1, sg684 (GGCAGTACGACTGCCAGCAG) were cloned into pLentiCRISPR-puro.v2. SET-2 were transduced with the resultant plasmids as indicated above and subsequently treated with 50 nM AW4 for 7 weeks to enrich drug-resistant mutant cells. Surviving cells were FACS-sorted as single cells and then expanded. For library preparation, genomic PCR primers (FWD-CTCATGTGCTTCTTTCTTATGGT, REV-GATTAAAATTAGAAGGCA**)** with Illumina adapter

sequences were used to amplify specified regions of LSD1 as previously described. Samples were sequenced on a MiSeq genome analyzer (Illumina). The sequencing reads were analyzed using CRISPResso2 (v.2.0.40).

Genotype Determination. Genomic DNA was isolated using QIAamp Blood & Cell Culture DNA Mini Kit (Qiagen) or Lucigen QuickExtract (Lucigen Corporation). For amplicon sequencing, genomic DNA was subjected to two rounds of PCR reactions. PCR1 was conducted using 500 ng of genomic DNA with Q5 HotStart Polymerase (NEB) to amplify the target region. PCR1 products were ran on an agarose gel and gel purified using Zymo Gel Extraction Kit (Zymo). 1 uL of the purified DNA was used as an input for PCR2. PCR2 was conducted with PCR primers containing Illumina adaptor sequences. PCR2 products were ran on an agarose gel, gel purified and pooled. The final libraries were quantified using Qubit High Sensitivity Assay Kit ((Thermo Fisher Scientific) and sequenced on a MiSeq genome analyzer (Illumina). The sequencing reads were analyzed using CRISPResso2 (v.2.0.40).[104]

Cell Growth Assays. SET-2 were seeded in 96-well plates with 2,000 cells per well in triplicate with drug or vehicle treatments. Cell viability was monitored at day 7 by measuring end point luminescence using CellTiter-Glo (Promega) on a SpectraMax i3x plate reader. ATP standard curve was prepared using known concentrations of ATP and used to calculate the ATP content of cells. Dose-response curves were determined through interpolation using GraphPad Prism v.7 nonlinear regression fit ([inhibitor] vs normalized response -- variable slope). All growth assays were performed at least twice.

Expression of LSD1(Δ150) from E.coli. LSD1(Δ150) and variants were cloned into a pET15b vector (gift from P. A. Cole) containing an N-terminal His$_6$-tag using NEBuilder HiFi DNA Assembly Master Mix (NEB E2621L). Mutant sequences were generated by PCR and inserted into the same

165

base vector as described above. The LSD1 constructs were expressed in BL21-CodonPlus (DE3)-RIPL competent *E. coli* and purified according to a literature procedure. All protein purification steps were performed at 4 °C. Briefly, cells were lysed by sonication (10 s on, 20 s off, 60% Amp) after resuspending in lysis buffer (50 mM HEPES pH 7.5, 100 mM KCl, 0.5 mM TCEP) supplemented with protease inhibitor cocktail (Roche). The lysate was clarified by centrifugation at 20,000 x g for 40 min. The clarified lysate was added to equilibrated His60 nickel resin (Takara Bio) and bound in batch for 1 h. The resin was washed with 10 column volumes (CV) of lysis buffer, followed by 10 CV of Wash buffer 1 (lysis buffer supplemented with 20 mM imidazole) and lastly eluted by the addition of 5-10 CV of lysis buffer supplemented with 250 mM imidazole. Fractions were combined and dialyzed overnight in 2 L of dialysis buffer (20 mM Tris pH 7.5, 100 mM NaCl, 5 mM β-mercaptoethanol, 2 changes), concentrated (Amicon Ultra-15 Centrifugal Filter Unit, 30 kDa MWCO) to 5-10 mLs and purified with a 5 mL Q Sepharose HP column (Cytiva) with a gradient elution from Buffer A (20 mM Tris pH 7.5, 1 mM TCEP) to Buffer B (20 mM Tris pH 7.5, 1 M NaCl, 1 mM TCEP). The protein was concentrated and purified by gel filtration chromatography on a Superdex 200 Increase 10/300 GL column (Cytiva) in storage buffer (50 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM TCEP, 5% glycerol). Fractions with > 95% purity as assessed by coomassie-stained SDS PAGE were pooled and stored at -80 ºC.

Expression and purification of LSD1(Δ150)-CoREST(Δ305) from E.coli. For LSD1-CoREST complexes, LSD1 was expressed as described above while CoREST(305-482) was expressed from a pGEX vector (gift from A. Mattevi) and the complex was purified by tandem affinity chromatography. The plasmid was transformed into BL21-CodonPlus (DE3)-RIPL *E.coli* cells and after plating a single colony was cultivated in LB media with 100 mg/L ampicillin at 37 ºC and expression was induced at OD600 of 0.8 by adding 0.25 mM isopropyl β-D-thiogalactoside (IPTG) and grown overnight at 17 ºC. The cells were pelleted by centrifugation at 4,000 x g for 30 min and stored at -80 ºC prior to purification. Pellets of CoREST and LSD1 were resuspended in lysis

buffer (50 mM NaH$_2$PO$_4$ pH 8.0, 300 mM NaCl, 5% glycerol, 7.5 mM imidazole supplemented with PMSF, DNAse and EDTA-free Roche protease inhibitor cocktail) in a weight ratio of 1:1.5, respectively. Cells were disrupted by sonication, clarified by centrifugation and passed through nickel affinity resin as before. The eluent was then loaded onto GST resin equilibrated in GST affinity buffer (50 mM NaH$_2$PO$_4$ pH 8.0, 300 mM NaCl, 5% glycerol, 1 mM DTT, 1 mM EDTA) and the GST-tag was cleaved on the resin after incubation with GST-PreScission protease (APEXBIO) overnight at 4 ºC. The protein was eluted by washing the column with GST affinity buffer, concentrated and subsequently gel-filtered on a Superdex 200 10/300 GL column equilibrated in storage buffer as before. The purity of the complex was verified by SDS-PAGE and fractions with 90-95% purity were pooled and stored at –80 ºC.

LSD1 enzyme assays. LSD1 enzymatic activity assays were performed in triplicate using Amplex Red Hydrogen Peroxide/Peroxidase Assay Kit (Invitrogen) with recombinant LSD1 and a synthetic peptide corresponding to the first 21 amino acids of H3K4me2 (Anaspec). LSD1 (500 ng/well) was incubated with 20 µM peptide at rt for 30 min. The endpoint fluorescence was measured on a microplate reader (excitation: 530 nm; emission: 590 nm) after 60 min following the addition of the Amplex Red/HRP mixture. Inhibition assays were performed as described above. Briefly, LSD1 (75 ng/well) and inhibitors at the appropriate concentration were incubated at rt for 10 min in reaction buffer with 0.01% BRIJ35 (ThermoFisher Scientific) prior to the addition of peptide. IC$_{50}$ values were determined in GraphPad Prism v.7 by nonlinear regression analysis (log(inhibitor) vs. response—variable response) of the concentration/inhibition data.

Steady-state kinetics were obtained under the following conditions briefly summarized here.[166] LSD1-CoREST complex activity was estimated by the coupled Amplex Red Hydrogen Peroxide/Peroxidase (Sigma-Aldrich) fluorescence assay using a ClarioStar plate reader (BMG Labtech). The final reaction mixture contained 50 mM HEPES pH 8.5, 0.1 mM Amplex Red, 0.3

mM horseradish peroxidase, 0.3 μM LSD1-CoREST. Each construct was kinetically characterized by testing enzymatic activity at 8 serially diluted concentrations (from 40 μM to 0.31 μM) with a synthetic peptide corresponding to the first 21 amino acids of H3K4me2 (Genscript). K_M and K_{cat} values were determined in GraphPad Prism v.7 by nonlinear regression analysis. The residual $K_{cat,app}$ of inhibited LSD1-CoREST complexes were evaluated by using 4.5 μM of protein with 40 μM dimethylK4-H3 21aa.

FAD-adduct detection by LC/Q-TOF MS. Analysis of FAD-adducts after LSD1 protein treatment with inhibitor was adopted from Matsuda *et al.*[71] Briefly, 100 μM of LSD1 was treated with 800 μM inhibitor at room temperature for indicated time points in 50 mM Tris-HCl (pH 8.0) and then subjected to 8 M urea to a final concentration of 6.4 M for 1 hr at 4 °C to denature the sample. The compound-FAD adducts were then extracted at 4 °C with acetonitrile for 15 min. The precipitates were removed by centrifugation at 10,000 *x* g for 10 min. The supernatant was supplemented with 500 mM ammonium acetate to a final concentration of 5 mM and filtered before sample injection. Samples were then injected onto a Phenomenex Kinetix C18 column (2.1 mm, 2.6 micron particle size, 150 mm length) with Orbitrap (Thermo q-Exactive Plus). Water with 0.1% formic acid and acetonitrile were used as mobile phases A and B, respectively. The mobile phase composition was changed to 100% B over 20 minutes.

Molecular mass determination of FAD inhibition products.The inhibition was realized on samples of > 100 μM LSD1-CoREST complex in 25 mM KH_2PO_4, pH 7.2 and 5% glycerol. Inhibitors was added for a final concentration of 400 μM and the incubation was at 20°C room for 24 hours. Inhibited LSD1-CoREST is unfolded by adding a proper volume of 8 M urea solution to have a final concentration of 6.4 M urea. After one hour on ice, acetonitrile (final concentration 30%) was added to percipitate protein debris. After centrifugation, supernatant is stored at -20°C untill spectrometric analysis. The sample solution was diluted 1:1 with 5 mM ammonium acetate

and was injected to a high resolution QTOF mass spectrometer UHPLC-HRMS/MS- AB Sciex X500B. Chromatographic separation was achieved using a C18 Zorbax extend analytical column (2.1 x 50 mm, 1.8 µm, Agilent). The mobile phase consisted of (A) 5 mM ammonium acetate in water and (B) methanol. The compositions of gradient elution (the mixture of (A) and (B) were as follows designated in percentage of (B), v/v); 5-10% at 0–1 min, 10–80% at 1–10 min, 80% at 10–15 min, 5% at 15–25 min. The flow rate was set at 0.3 mL/min. The mass spectrometer was operated in the negative ionization mode. SCIEX software was used for data acquisition and processing. The percentage of products represent the percentage of total area of all extracted ion current peaks. FAD-adducts were usually distributed in different peaks comprising [M-H]-, [M-H-H2O]- (dehydrated adduct), [M-2H+Na]-and [M-H+CH3]- (methylester).

Styrene detection assay from LSD1-CoREST complexes. For WT LSD1, 80 µM of LSD1 was treated with 640 µM inhibitor at room temperature for indicated time points in 50 mM Tris-HCl (pH 8.0). Styrene was extracted with 200 µL DCM with 100 µM caffeine as an internal standard. The DCM was subsequently removed and the extract was resuspended in 200 µL of ACN and injected onto a Phenomenex Kinetix C18 column (2.1 mm, 2.6 micron particle size, 150 mm length) with Agilent 6530 Q-TOF Mass Spectrometer. Water with 0.1% formic acid and acetonitrile were used as mobile phases A and B, respectively. The mobile phase composition was changed to 100% B over 20 minutes.

For TTASdel LSD1-CoREST, 50 µM of LSD1 was treated with 400 µM inhibitor and 100 µM GFI1b peptide at room temperature for indicated time points in 50 mM Tris-HCl, pH 8.0. Styrene was extracted with 200 µL DCM with 100 µM tridecane as an internal standard. The DCM was subsequently removed and the extract was resuspended in 200 µL of ACN and analyzed with Waters Quattro micro GC/MS/MS.

Styrene detection assay from cell media. 5 million SET-2 cells were resuspended in 1 mL of R-20 and treated with 1 µM of inhibitor for 24 hr. The cells were spun down and the media was filtered through a 0.22 µm filter and then extracted 2x with 300 µL dichloromethane (DCM). The DCM was subsequently removed and the extract was resuspended in 100 µL of DCM and analyzed by high-resolution mass-spectrometry.

Absorbance spectroscopy. Purified LSD1-CoREST (10 µM) in 25 mM KH_2PO_4 pH 7.2 and 5% (w/v) glycerol was incubated with 400 µM inhibitor for variable times. Absorbance spectra were measured using a Cary-100 spectrophotometer (Varian).

***Thermo*FAD Assay.** *Thermo*FAD was performed in quintets as previously described.[106] Thermal curves were obtained using a qRT-PCR detection system (BioRad C1000 Touch Thermal Cycler) with a temperature gradient from 20 to 95 ºC. The T_m was determined by fitting the thermal curves to a Boltzmann sigmoid and calculating the inflection point (GraphPad Prism v.7).

Fluorescence polarization assay. Binding assays were performed in two independent experiments with three technical replicates. The change in fluorescence polarization of fluorescently-labeled GFI1B peptide upon binding to LSD1-CoREST was monitored using previously described protocols.[69,111,166] LSD1-CoREST (2 µM) complex was incubated with the inhibitor (10 µM) or DMSO vehicle control for 10 min on ice and then with the labeled peptides for the indicated time points. After incubation, the samples were prepared by a 2-fold serial dilution in the assay buffer (15 mM KH_2PO_4, pH 7.2, 5% glycerol, and 1 mg/mL BSA) containing 2 nM labeled peptide and the plate was equilibrated to room temperature before data acquisition. Fluorescence polarization was measured using a microplate reader in 384-well black microplates at 25 °C. The G-factor on the microplate reader was adjusted to 35 mP for the reference well

containing labeled peptide. The binding curves were fit by nonlinear regression analysis in GraphPad Prism v.7 as described previously.

Crystals soaking and structure determination. Purified LSD1-CoREST complex was crystallized following the previously published conditions with 1.1-1.3M sodium/potassium tartrate and 100 mM N-(2-acetamido)-2-iminodiacetic acid), pH 6.5.[51] Crystals were transferred in reservoir drops supplemented with 1mM inhibitor and 4% DMSO. Once FAD is bleached (about 40 minutes), crystals were flash-cooled for data collection. For a second set of experiments, aimed at investigating changes in FAD-adducts over time, crystals are back washed in a clear reservoir drop and left for three days at 20 °C. The third set of experiment were devoted to structurally visualize SNAG peptide conformation in presence of different FAD-adducts. Inhibited crystals were soaked in a drop containing 1 mM of SNAG(1-9) peptide before flash-freezing.

Chapter 5: Concluding remarks

Contributions:

Amanda L. Waterbury (A.L.W.) wrote the text. Figures were made by A.L.W. with assistance from Brian B. Liau (B.B.L.). A.L.W. and Hui Si Kwok (H.S.K.) designed experiments discussed in this chapter. B.B.L. supervised the research and edited the text.

5.1. Summary

In this thesis, I have mechanistically investigated and deconvoluted distinct mechanisms of mutations that confer resistance to inhibitors targeting LSD1 in acute myeloid leukemia (AML). In chapter two, we demonstrated that CRISPR-suppressor scanning can profile resistance mutations to LSD1 inhibitors in AML. Surprisingly, we showed that the demethylase activity of LSD1 is dispensable for AML cell survival through the identification of enzymatically dead LSD1 mutants that confer drug resistance. Through the characterization of these mutants, we found that the anti-proliferative activity of LSD1 inhibitors is due to their ability to disrupt the LSD1-GFI1B complex, which is essential for cell survival. Furthermore, through the discovery of a GFI1B drug-complementary allele, we provide orthogonal validation for the essentiality of the LSD1-GFI1B complex in AML. We extended this CRISPR-suppressor screening approach to include close analogs of GSK-LSD1 as well as two reversible LSD1 inhibitors and showed that CRISPR-suppressor scanning can discriminate between structurally related compounds. Each inhibitor gave rise to a unique resistance profile, suggesting that the different inhibitors may select for distinct mechanisms of resistance. Further characterization of one of these mutants that is selectively enriched by one of the inhibitors is discussed in chapter four.

In chapter three, we identify the mechanism where mutations in the intrinsically disordered region (IDR) of LSD1 drive drug resistance and elucidate the biological function of this disordered region in tuning LSD1 transcription factor (TF) interactions. Through integrating cell biology, biochemistry, live-cell microscopy, and genomic studies, we found that the IDR of LSD1 modulates LSD1-TF interactions with myeloid TFs to control enhancer commissioning and downstream AML differentiation. We show that LSD1 inhibitors reprogram LSD1-TF interactions, from the high-affinity LSD1-GFI1B interaction to weaker and more transient interactions with master myeloid TFs, namely C/EBPα. Furthermore, the LSD1 IDR modulates this TF switch by blocking LSD1 interactions with TF IDRs, which we hypothesize to occur by driving LSD1-TF demixing. The resistance mutations break this inhibitory mechanism to promote LSD1-TF

interactions, allowing LSD1 to buffer enhancers and block AML differentiation. This study highlights the role of IDRs in regulating the function of chromatin regulators and how IDRs can have non-canonical functions, such as inhibiting weak interactions. Furthermore, this study clarifies how LSD1 is recruited to active enhancers to buffer their activity by engaging in weak interactions with myeloid TFs. More broadly, these studies illustrate how chromatin modifiers can regulate gene expression through engaging different TF partners and highlight the role of IDRs in tuning these TF switches.

In the fourth chapter, we investigate how a distal loop deletion in LSD1, first identified in the work discussed in chapter 2, mediates drug resistance by promoting fragmentation of the covalent adduct formed with FAD after treating LSD1 with an irreversible inhibitor. While mutations that block formation of the covalent drug-target bond is a common mechanism of drug resistance for covalent inhibitors, we report a novel drug resistance mechanism where the distal loop deletion in LSD1 promotes chemical rearrangement of the covalent inhibitor adduct. Through mechanistic investigations, we report that this rearrangement occurs through a Grob-like fragmentation, analogous to the transformation that was reported to occur with the recently reported LSD1 inhibitor, T-448. Through chemical synthesis, biochemical assays, and structural studies, we clarify the mechanism of this covalent adduct fragmentation and reveal new roles of allosteric sites in mediating resistance.

Overall, these studies demonstrate the utility of CRISPR-suppressor scanning in identifying resistance mutations not only in the drug binding site but also in distal regions of the protein. Characterization of these distal drug-resistant mutants have illuminated novel aspects of LSD1 biology and highlight the role of allosteric and unstructured regions in not only mediating drug-resistance but also in protein function. We anticipate that this approach can be applied to other targets and disease contexts to elucidate target biology and aid in the development of new targeted anti-cancer therapeutics.

5.2. Perspectives and future directions

In chapter 2, I described the identification of drug-resistance conferring mutations in LSD1 that disrupt demethylase activity. This finding was initially surprising since LSD1 inhibitors were developed under that assumption that the demethylase activity of LSD1 is essential for acute myeloid leukemia (AML) and small cell lung carcinoma (SCLC) cell survival.[35,95–97] This study highlights the need to resolve the catalytic versus the non-catalytic functions of chromatin regulators, and to both develop and employ methods aimed at identifying non-catalytic functions.[15] Furthermore, given the scaffolding activity of LSD1, the context where LSD1 demethylase activity is essential for LSD1 function should be further investigated. For example, studies have suggested that LSD1 activity is important in the nervous system as alterations to LSD1 have been implicated in a number of neurological disorders but its precise function is poorly understood.[173–175] The proper expression and activity of LSD1 splice variants, such as the neurospecific isoform LSD1-8a, have also been implicated in development (**Figure 5.1**). Therefore, the biological function of LSD1's demethylase activity warrants further investigation as well as why the active site evolved to have these additional scaffolding activities and in what contexts they are essential for LSD1's function.
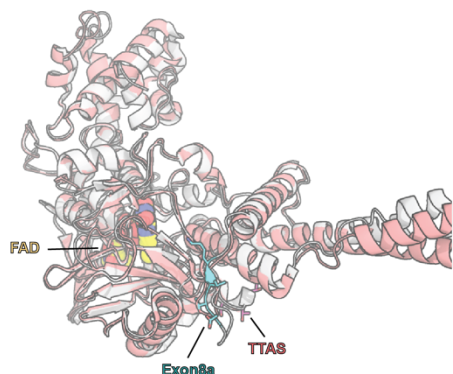

Figure 5.1. LSD1-8a overlay with wild-type LSD1.

Figure 5.1 (continued) Crystal structure of LSD1-8a (gray) overlaid with the structure of wild-type LSD1 (salmon) where the exon 8a insertion is highlighted in cyan and the TTAS loop is shown in magenta. PDB: 2x01 and 2hko for LSD1-8a and wild-type LSD1, respectively.

Moreover, we identified that irreversible small molecule LSD1 inhibitors are inhibitors of transcription factors and disrupt the LSD1-GFI1/GFI1B complex, which is responsible for their anti-proliferative activity in AML.[69,70] In chapter 3, we highlight that after LSD1 inhibitors dissociate the LSD1-GFI1B complex, LSD1 then engages in protein interactions with myeloid TFs. These results demonstrate that LSD1 inhibitors are TF reprogrammers, switching LSD1 complex activity from sustained gene repression at GFI1B sites to enhancer buffering activity upon weaker recruitment to chromatin by myeloid TFs, such as C/EBPα (**Figure 5.2a,b**). These results have important implications when developing and validating inhibitors that target chromatin regulators. Beyond inhibiting their canonical enzymatic activity or reader function, inhibitors that target chromatin regulators may also disrupt additional scaffolding functions, altering complex composition and function. Furthermore, these studies highlight the utility of targeting chromatin regulators as a way to modulate TF activity, as developing selective TF inhibitors remains challenging due to extensive structural conservation in the folded DNA binding domains and a lack of structure in the remaining regions (i.e., activation, repressor domains), which tend to be intrinsically disordered.[158,159,176] However, the ability of inhibitors to reprogram TF interactions, as showcased with covalent LSD1 inhibitors, has important consequences for developing TF inhibitors, as blocking their protein interactions may inadvertently induce the formation of other interactions (**Figure 5.2b**).
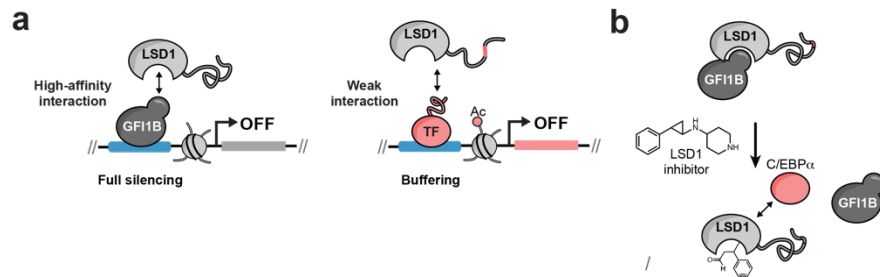
Figure 5.2. LSD1 inhibitors are transcription factor reprogrammers.

(**a**) LSD1 can engage is high-affinity TF interactions with GFI1B and weak interactions with myeloid TFs to mediate sustained gene repression or enhancer buffering activity, respectively. (**b**) Example of inhibitors inadvertently reprogramming TF interactions. LSD1 inhibitors dissociate the LSD1-GFI1B complex, allowing LSD1 to interact with myeloid TFs like C/EBPα.

In chapter 3, we uncovered surprising mutations in the LSD1 N-terminal IDR that confer drug resistance. After identifying that LSD1 interacts with C/EBPα upon GSK-LSD1 treatment, we hypothesized that this protein-protein interaction occurred through the IDRs of LSD1 and C/EBPα. Consistent with this hypothesis, full-length recombinant LSD1 forms phase-separated droplets, suggesting that LSD1 engages in weak multivalent interactions, and these droplets co-condense the IDR of C/EBPα. However, by live-cell microscopy using a LacO-LacI recruitment system to investigate protein-protein interactions, we showed that the IDR of LSD1 counterintuitively blocks LSD1 recruitment to C/EBPα-LacI hubs and that LSD1 lacking the IDR altogether was recruited readily into the hub while the LSD1 IDR alone was actively excluded. These results highlight the importance of considering IDR-structured protein interactions in addition to IDR-IDR interactions as drivers of weak multivalent interactions. Furthermore, this study demonstrates that IDRs can tune protein interactions and do not necessarily promote the formation of weak multivalent interactions but can instead block interactions that would otherwise occur with the protein core. Furthermore, the qualitative and artificial nature of droplet assays, the standard method to investigate IDR function in vitro, emphasizes the current limitations of

studying IDRs biochemically. Improvements in studying IDRs by NMR and hydrogen-deuterium exchange mass spectrometry (HDX-MS) will be instrumental towards biochemically elucidating their function.[28,154,177]

A critical question that our study raises is what chemical features of the LSD1 IDR drives its inhibitory activity. To address this, we are taking two approaches: a deep mutational scan of LSD1 and live-cell microscopy profiling LSD1 recruitment versus exclusion from a library of TF IDR-LacI hubs. In the first approach, we have generated a library of LSD1 IDR mutations, from point alanine mutations to systematic deletions across the IDR. After knockdown of endogenous LSD1 and overexpression of this pool of LSD1 IDR variants in the presence of GSK-LSD1, we will select for LSD1 IDR variants that confer resistance to GSK-LSD1 and characterize these mutants further. We hypothesize that this experiment will clarify which regions of the LSD1 IDR are essential for its inhibitory activity and whether there are specific sites or chemical features that mediate its function. In the live-cell microscopy LacO-LacI assay, we have designed a diverse library of TF IDR-LacI fusions, including a diverse set of myeloid factors, and will test whether the LSD1 IDR is excluded from hubs generated with these constructs (**Figure 5.3**). Conversely, we will investigate whether LSD1 lacking its IDR is recruited into these hubs. This study will shed light on what drives LSD1 IDR exclusion from hubs and whether it blocks specific TF interactions or has broad inhibitory activity. For instance, the overall charge and hydrophobic properties of the LSD1 IDR may drive TF hub demixing. Another possibility is that the LSD1 IDR may masks sites on the LSD1 structure that mediate key multivalent interactions. The deletions may also promote a disorder-to-order transition, which may mask post-translation modification sites and/or alter proteins interactions.[178] Overall, we hypothesize that these studies will illuminate key aspects of IDR biology.
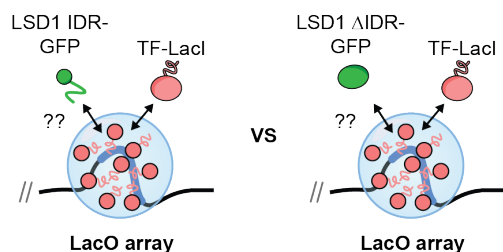
Figure 5.3. LacO-LacI assay to study TF hub co-partitioning.

LSD1 IDR exclusion/recruitment will be assessed with a TF IDR library fused to LacI by the LacO-LacI recruitment assay in U2OS 2-6-3 cells overexpressing LSD1 IDR fused to GFP. Recruitment will be compared to LSD1 lacking the IDR in U2OS 2-6-3 cells overexpressing LSD1 ΔIDR.

Our study identified a novel LSD1 protein interaction with C/EBPα. However, how LSD1 interacts with C/EBPα remains unclear. Using the LacO-LacI assay, we identified that the LSD1 AOD-Tower domain fusion is sufficient for mediating the LSD1-C/EBPα interaction with the C/EBPα IDR. However, given the disordered nature of the interacting region of C/EBPα and the weak nature of the LSD1-C/EBPα protein interaction, further biochemical characterization was difficult. Cross-linking mass spectrometry, HDX-MS or NMR studies would be highly informative to further characterize this protein interaction since these methods are amenable towards structurally elucidating IDRs and their binding interactions. Alternatively, further LSD1 domain truncations and point mutations could be tested in the LacO-LacI assay. These studies would shed light on how TFs engage chromatin modifiers and more fundamentally, how IDRs interact with structured binding partners. Another outstanding question is the role of CoREST and other CoREST complex members in mediating resistance in the LSD1 IDR mutant cell lines. While we did not observe differences in LSD1-HDAC1-COREST (LHC) complex activity between wild-type and NΔ2 LHC, we cannot rule out the involvement of other key complex members, such as CoREST and HDAC1/2 in the resistance mechanism. For instance, by HDX-MS we observed increased protection from deuterium uptake for LSD1(NΔ2)-CoREST at the LSD1-COREST

interface compared to the wild-type complex. Moreover, the Tower domain of LSD1, which binds to CoREST, is essential for LSD1 recruitment to C/EBPα-LacI hubs by live-cell microscopy, as the LSD1 AOD domain alone failed to be recruited to the C/EBPα hub. Both of these results suggest that CoREST may be essential or at least play a role in the resistance mechanism in the LSD1 IDR mutant cells and warrants further investigation.

In addition, this study identified resistance mutations in an IDR that were then shown to lead to broader impacts on protein function and downstream cell phenotypes, such as cell differentiation. However, disease mutations in IDRs remain poorly characterized as most studies to date have focused on mutations in structured regions since they are amenable to further biochemical study and are more likely to occur in sequence conserved regions with annotated function.[30,178] The majority of proteins in the human proteome contain IDRs in addition to their structured domains and in contrast to these folded domains, IDRs lack sequence conservation and functional annotation. However, up to a quarter of missense disease mutations have been mapped to IDRs and are predicted to cause disorder-to-order transitions, alter protein interactions, post-translation modifications and/or protein complex activity, necessitating better computational and experimental methods to study IDRs. In the future, adapting CRISPR-scanning to identify resistance mutations in IDRs will shed light on how disease mutations in IDRs may mediate resistance and more generally elucidate the biological function of IDRs in disease contexts, which have been difficult to study by other methods.

In chapter four, we reported a novel drug resistance mechanism where the resistance mutation promotes the chemical rearrangement of a covalent inhibitor adduct. Point mutations that prevent drug binding as well as hypermorphic mutations that increase the biological activity or protein level of the drug target to compensate for the presence of the inhibitor are both common mechanisms of drug resistance.[171] For example, the T315I mutation in BCR-ABL frequently occurs in response to imatinib/dasatinib treatment to mediate drug resistance while the M2327I

resistance mutation in mTOR increases the basal kinase activity to block the effects of AZD8055 treatment.[75,80,171,179] While these mechanisms of drug resistance have been extensively reported in the literature and the clinic, the systematic discovery and characterization of alternative mechanisms of resistance is lacking. Using CRISPR-suppressor scanning to profile resistance mutations, we identified a small deletion mutation in a distal allosteric loop in LSD1 and through mechanistic investigations found that this mutation promotes the chemical fragmentation of the covalent inhibitor adduct to facilitate drug-resistance. Regardless of whether this resistance mutation arises in the clinic with LSD1 inhibitors, it teaches us about novel mechanisms of drug resistance and will inform feature efforts profiling resistance mutations.

Whether this chemical transformation promotes resistance in LSD1 TTASdel SET-2 cells warrants further investigation as discussed in chapter four. It is possible that the TTAS loop may regulate TF interactions and deletion of this loop promotes resistance by a similar mechanism as the LSD1 IDR mutants described in chapter three. Co-IP experiments investigating whether the LSD1 TTAS-GIF1B complex is maintained in the presence of AW4 in SET-2 TTASdel mutant cells as well as GFI1B knockdown experiments to clarify whether this mutant cell line depends on GFI1B for cell survival will be highly informative for investigating this alternative resistance mechanism hypothesis. Furthermore, co-IP mass-spectrometry and live-cell LacO-LacI microscopy experiments would also clarify the LSD1 TTASdel TF interactome and whether increased interactions with myeloid TFs is driving drug resistance, which we could further support with genomics studies investigating LSD1 chromatin localization and enhancer buffering activity as discussed in chapter three. The LSD1 TTAS loop itself is involved in crystal contacts, limiting structural characterization by this approach. Alternative methods to structurally resolve LSD1 TTASdel such as cryogenic electron microscopy (cryo-EM) and HDX-MS would clarify how deletion of this loop alters the structure of LSD1 and whether this region of LSD1 engages in crosstalk with the active-site of LSD1 to modulate FAD-drug adduct rearrangement. These structural methods can also be used to investigate LSD1 drug-resistance conferring mutations in

the context of the biologically relevant LSD1-HDAC1-CoREST complex, which is a highly dynamic complex that is difficult to study by crystallography, to clarify how mutations in LSD1 may perturb complex function as a whole.[54]

In chapter four, we also explore the mechanism and chemical features of the demethylase specific inhibitor T-448. This inhibitor was developed for treating central nervous system disorders, where it is desirable to target LSD1's demethylase activity.[71] Further investigation into the biological function of LSD1 in this disease context would clarify the catalytic activities of LSD1. Moreover, the development of this demethylase specific inhibitor was motivated by recent studies reporting hematotoxicity associated with LSD1 inhibition due to disruption of the LSD1-GFI1/GFI1B complex. T-448, by generating the compact *N*-formyl-FAD adduct that is sufficient to inhibit enzyme activity while preserving the LSD1-GFI1B complex, not only circumvents this associated hematotoxicity but is also ineffective at inhibiting AML cell proliferation. Given these results, T-448 provides orthogonal validation for the relevance of the LSD1-GFI1/GFI1B complex in leukemia and highlights the on-target activity of LSD1 inhibitors through disruption of this protein complex. We propose that the development of T-448 analogs that can both block the LSD1-GFI1/GFI1B complex to induce AML cell differentiation and slowly decompose to the *N*-formyl-FAD adduct to reverse complex disruption may help alleviate hematotoxicity associated with current LSD1 inhibitors while still providing anticancer activity.

More broadly, this work illustrates how the identification and mechanistic investigation of drug-suppressor mutations is a powerful strategy to discover target biology and mechanisms of drug resistance. Furthermore, these studies have revealed fundamental aspects of chromatin biology, such as how protein allostery, disordered domains, and interactions with different TFs can tune chromatin complex function. The characterization of different LSD1 inhibitors provides an illustrative example on how small molecules can be used to tease apart different activities of these multifaceted chromatin regulators. For example, GSK-LSD1 is both a demethylase and TF interaction inhibitor while T-448 selectively inhibits the catalytic activity of LSD1, versus UM171,

which entirely degrades the complex.[69,71,73] Future work exploring the activities perturbed by each of these small molecules will be highly informative in elucidating the function of LSD1 in different cellular contexts and will contribute to our fundamental understanding of chromatin regulator function. The approaches and biological insights described herein can be further applied to other chromatin complexes and disease relevant factors to advance both our knowledge of these proteins and efforts to therapeutically target their specific functions.

References

1. WU, H. & SUN, Y. E. Epigenetic Regulation of Stem Cell Differentiation. *Pediatr Res* 59, 21R-25R (2006).

2. Fierz, B. & Muir, T. W. Chromatin as an expansive canvas for chemical biology. *Nat Chem Biol* 8, 417–427 (2012).

3. Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res* 21, 381–395 (2011).

4. Allis, C. D. & Jenuwein, T. The molecular hallmarks of epigenetic control. *Nat Rev Genet* 17, 487–500 (2016).

5. Goldberg, A. D., Allis, C. D. & Bernstein, E. Epigenetics: A Landscape Takes Shape. *Cell* 128, 635–638 (2007).

6. Amente, S., Lania, L. & Majello, B. The histone LSD1 demethylase in stemness and cancer transcription programs. *Biochimica Et Biophysica Acta Bba - Gene Regul Mech* 1829, 981–986 (2013).

7. Taunton, J., Hassig, C. A. & Schreiber, S. L. A Mammalian Histone Deacetylase Related to the Yeast Transcriptional Regulator Rpd3p. *Science* 272, 408–411 (1996).

8. Ummethum, H. & Hamperl, S. Proximity Labeling Techniques to Study Chromatin. *Frontiers Genetics* 11, 450 (2020).

9. Roux, K. J., Kim, D. I., Raida, M. & Burke, B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J Cell Biol* 196, 801–810 (2012).

10. Rhee, H.-W. *et al.* Proteomic Mapping of Mitochondria in Living Cells via Spatially Restricted Enzymatic Tagging. *Science* 339, 1328–1331 (2013).

11. Huang, X. *et al.* Defining the NSD2 interactome: PARP1 PARylation reduces NSD2 histone methyltransferase activity and impedes chromatin binding. *J Biol Chem* 294, 12459–12471 (2019).

12. Nair, S. S. & Kumar, R. Chromatin remodeling in Cancer: A Gateway to regulate gene Transcription. *Mol Oncol* 6, 611–619 (2012).

13. Giri, A. K. & Aittokallio, T. DNMT Inhibitors Increase Methylation in the Cancer Genome. *Front Pharmacol* 10, 385 (2019).

14. Richon, V. M. Cancer biology: mechanism of antitumour action of vorinostat (suberoylanilide hydroxamic acid), a novel histone deacetylase inhibitor. *Brit J Cancer* 95, S2–S6 (2006).

15. Aubert, Y., Golf, S. & Capell, B. C. The Unexpected Noncatalytic Roles of Histone Modifiers in Development and Disease. *Trends in Genetics* 35, 645–657 (2019).

16. Kim, J.-A., Kwon, M. & Kim, J. Allosteric Regulation of Chromatin-Modifying Enzymes. *Biochemistry-us* 58, 15–23 (2019).

17. Weinberg, D. N. *et al.* The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape. *Nature* 573, 281–286 (2019).

18. Shi, Y. *et al.* Structure of the PRC2 complex and application to drug discovery. *Acta Pharmacol Sin* 38, 963–976 (2017).

19. Margueron, R. & Reinberg, D. The Polycomb complex PRC2 and its mark in life. *Nature* 469, 343–349 (2011).

20. Lee, C.-H. *et al.* Allosteric Activation Dictates PRC2 Activity Independent of Its Recruitment to Chromatin. *Mol Cell* 70, 422-434.e6 (2018).

21. Chammas, P., Mocavini, I. & Croce, L. D. Engaging chromatin: PRC2 structure meets function. *Brit J Cancer* 122, 315–328 (2020).

22. Qi, V. *et al.* An allosteric PRC2 inhibitor targeting the H3K27me3 binding pocket of EED. *Nature Chemical Biology* (2017).

23. Liu, K.-L., Zhu, K. & Zhang, H. An overview of the development of EED inhibitors to disable the PRC2 function. *Rsc Medicinal Chem* 13, 39–53 (2021).

24. Sabari, B. R., Dall'Agnese, A. & Young, R. A. Biomolecular Condensates in the Nucleus. *Trends Biochem Sci* 45, 961–977 (2020).

25. Sabari, B. R. *et al.* Coactivator condensation at super-enhancers links phase separation and gene control. *Science* 361, (2018).

26. Boija, A., Klein, I. A. & Young, R. A. Biomolecular condensates and cancer. *Cancer Cell* 39, 174–192 (2021).

27. Shin, Y. & Brangwynne, C. P. Liquid phase condensation in cell physiology and disease. *Science* 357, (2017).

28. Musselman, C. A. & Kutateladze, T. G. Characterization of functional disordered regions within chromatin associated proteins. *Iscience* 24, 102070 (2021).

29. Gibson, B. A. *et al.* Organization of Chromatin by Intrinsic and Regulated Phase Separation. *Cell* 179, 470-484.e21 (2019).

30. Tsang, B., Pritišanac, I., Scherer, S. W., Moses, A. M. & Forman-Kay, J. D. Phase Separation as a Missing Mechanism for Interpretation of Disease Mutations. *Cell* 183, 1742–1756 (2020).

31. Li, C. H. *et al.* MeCP2 links heterochromatin condensates and neurodevelopmental disease. *Nature* 586, 440–444 (2020).

32. Strom, A. R. *et al.* Phase separation drives heterochromatin domain formation. *Nature* 547, 241–245 (2017).

33. JIAO, L. & LIU, X. Structural basis of histone H3K27 trimethylation by an active polycomb repressive complex 2. *Science* 350, (2015).

34. Chong, S. *et al.* Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science* 361, (2018).

35. Maiques-Diaz, A. & Somervaille, T. C. LSD1: biologic roles and therapeutic targeting. *Epigenomics-uk* 8, 1103–1116 (2016).

36. Shi, Y. *et al.* Histone Demethylation Mediated by the Nuclear Amine Oxidase Homolog LSD1. *Cell* 119, 941–953 (2004).

37. Schmidt, D. M. Z. & McCafferty, D. G. trans-2-Phenylcyclopropylamine Is a Mechanism-Based Inactivator of the Histone Demethylase LSD1 †. *Biochemistry-us* 46, 4408–4416 (2007).

38. Wang, J. *et al.* Opposing LSD1 complexes function in developmental gene activation and repression programmes. *Nature* 446, 882–887 (2007).

39. Foster, C. T. *et al.* Lysine-Specific Demethylase 1 Regulates the Embryonic Transcriptome and CoREST Stability. *Mol Cell Biol* 30, 4851–4863 (2010).

40. Whyte, W. A. *et al.* Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. *Nature* 482, 221–225 (2012).

41. 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).

42. Carmichael, C. L. *et al.* The EMT modulator SNAI1 contributes to AML pathogenesis via its interaction with LSD1. *Blood* 136, 957–973 (2020).

43. Li, Y. *et al.* Dynamic interaction between TAL1 oncoprotein and LSD1 regulates TAL1 function in hematopoiesis and leukemogenesis. *Oncogene* 31, 5007–5018 (2012).

44. Saleque, S., Kim, J., Rooke, H. M. & Orkin, S. H. Epigenetic Regulation of Hematopoietic Differentiation by Gfi-1 and Gfi-1b Is Mediated by the Cofactors CoREST and LSD1. *Mol Cell* 27, 562–572 (2007).

45. Guo, Y. *et al.* GATA2 regulates GATA1 expression through LSD1-mediated histone modification. *Am J Transl Res* 8, 2265–74 (2016).

46. Kerenyi, M. A. *et al.* Histone demethylase Lsd1 represses hematopoietic stem and progenitor cell signatures during blood cell maturation. *Elife* 2, e00633 (2013).

47. Sprüssel, A. *et al.* Lysine-specific demethylase 1 restricts hematopoietic progenitor proliferation and is essential for terminal differentiation. *Leukemia* 26, 2039–2051 (2012).

48. Yang, M. *et al.* Structural Basis for CoREST-Dependent Demethylation of Nucleosomes by the Human LSD1 Histone Demethylase. *Mol Cell* 23, 377–387 (2006).

49. Stavropoulos, P., Blobel, G. & Hoelz, A. Crystal structure and mechanism of human lysine-specific demethylase-1. *Nat Struct Mol Biol* 13, 626–632 (2006).

50. Yang, M. *et al.* Structural basis of histone demethylation by LSD1 revealed by suicide inactivation. *Nat Struct Mol Biol* 14, 535–539 (2007).

51. 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).

52. 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.e4 (2020).

53. Dhall, A. *et al.* Nucleosome Binding by the Lysine Specific Demethylase 1 (LSD1) Enzyme Enables Histone H3 Demethylation. *Biochemistry-us* 59, 2479–2483 (2020).

54. Song, Y. *et al.* Mechanism of Crosstalk between the LSD1 Demethylase and HDAC1 Deacetylase in the CoREST Complex. *Cell Reports* 30, 2699-2711.e8 (2020).

55. 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).

56. 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).

57. Verma, S. *et al.* Review Article Systematic review of epigenetic targets in acute myeloid leukemia. *Am J Blood Res.* (2021).

58. Harris, W. J. *et al.* The Histone Demethylase KDM1A Sustains the Oncogenic Potential of MLL-AF9 Leukemia Stem Cells. *Cancer Cell* 21, 473–487 (2012).

59. Olsson, I., Bergh, G., Ehinger, M. & Gullberg, U. Cell differentiation in acute myeloid leukemia. *Eur J Haematol* 57, 1–16 (1996).

60. Prange, K. H. M. *et al.* MLL-AF9 and MLL-AF4 oncofusion proteins bind a distinct enhancer repertoire and target the RUNX1 program in 11q23 acute myeloid leukemia. *Oncogene* 36, 3346–3356 (2017).

61. Schenk, T. *et al.* Inhibition of the LSD1 (KDM1A) demethylase reactivates the all-trans-retinoic acid differentiation pathway in acute myeloid leukemia. *Nat Med* 18, 605–611 (2012).

62. Yang, M. *et al.* Structural Basis for the Inhibition of the LSD1 Histone Demethylase by the Antidepressant trans-2-Phenylcyclopropylamine † , ‡. *Biochemistry-us* 46, 8058–8065 (2007).

63. Fang, Y., Liao, G. & Yu, B. LSD1/KDM1A inhibitors in clinical trials: advances and prospects. *J Hematol Oncol* 12, 129 (2019).

64. Salamero, O. *et al.* First-in-Human Phase I Study of Iadademstat (ORY-1001): A First-in-Class Lysine-Specific Histone Demethylase 1A Inhibitor, in Relapsed or Refractory Acute Myeloid Leukemia. *J Clin Oncol* 38, 4260–4273 (2020).

65. Gill, H. *et al.* P1051: A PHASE 2 STUDY OF IMG-7289 (BOMEDEMSTAT) IN PATIENTS WITH ADVANCED MYELOFIBROSIS. *Hemasphere* 6, 941–942 (2022).

66. Kanouni, T. *et al.* Discovery of CC-90011: A Potent and Selective Reversible Inhibitor of Lysine Specific Demethylase 1 (LSD1). *J Med Chem* 63, 14522–14529 (2020).

67. Dai, X.-J. *et al.* Reversible Lysine Specific Demethylase 1 (LSD1) Inhibitors: A Promising Wrench to Impair LSD1. *J Med Chem* 64, 2466–2488 (2021).

68. Niwa, H., Sato, S., Hashimoto, T., Matsuno, K. & Umehara, T. Crystal Structure of LSD1 in Complex with 4-[5-(Piperidin-4-ylmethoxy)-2-(p-tolyl)pyridin-3-yl]benzonitrile. *Molecules* 23, 1538 (2018).

69. Vinyard, M. E. *et al.* CRISPR-suppressor scanning reveals a nonenzymatic role of LSD1 in AML. *Nat Chem Biol* 15, 529–539 (2019).

70. Maiques-Diaz, A. *et al.* Enhancer Activation by Pharmacologic Displacement of LSD1 from GFI1 Induces Differentiation in Acute Myeloid Leukemia. *Cell Reports* 22, 3641–3659 (2018).

71. Matsuda, S. *et al.* T-448, a specific inhibitor of LSD1 enzyme activity, improves learning function without causing thrombocytopenia in mice. *Neuropsychopharmacol* 44, 1505–1512 (2019).

72. Baba, R. *et al.* LSD1 enzyme inhibitor TAK-418 unlocks aberrant epigenetic machinery and improves autism symptoms in neurodevelopmental disorder models. *Sci Adv* 7, eaba1187 (2021).

73. Chagraoui, J. *et al.* UM171 Preserves Epigenetic Marks that Are Reduced in Ex Vivo Culture of Human HSCs via Potentiation of the CLR3-KBTBD4 Complex. *Cell Stem Cell* 28, 48-62.e6 (2021).

74. Freedy, A. M. & Liau, B. B. Discovering new biology with drug-resistance alleles. *Nat Chem Biol* 17, 1219–1229 (2021).

75. Rossari, F., Minutolo, F. & Orciuolo, E. Past, present, and future of Bcr-Abl inhibitors: from chemical development to clinical efficacy. *J Hematol Oncol* 11, 84 (2018).

76. Braun, T. P., Eide, C. A. & Druker, B. J. Response and Resistance to BCR-ABL1-Targeted Therapies. *Cancer Cell* 37, 530–542 (2020).

77. Reddy, E. P. & Aggarwal, A. K. The Ins and Outs of Bcr-Abl Inhibition. *Genes Cancer* 3, 447–454 (2012).

78. Azam, M., Seeliger, M. A., Gray, N. S., Kuriyan, J. & Daley, G. Q. Activation of tyrosine kinases by mutation of the gatekeeper threonine. *Nat Struct Mol Biol* 15, 1109–1118 (2008).

79. Islam, K. The Bump-and-Hole Tactic: Expanding the Scope of Chemical Genetics. *Cell Chem Biol* 25, 1171–1184 (2018).

80. Zhou, T. *et al.* Structural Mechanism of the Pan-BCR-ABL Inhibitor Ponatinib (AP24534): Lessons for Overcoming Kinase Inhibitor Resistance. *Chem Biol Drug Des* 77, 1–11 (2011).

81. Chen, Y.-N. P. *et al.* Allosteric inhibition of SHP2 phosphatase inhibits cancers driven by receptor tyrosine kinases. *Nature* 535, 148–152 (2016).

82. Zheng, X. S., Chan, T.-F. & Zhou, H. H. Genetic and Genomic Approaches to Identify and Study the Targets of Bioactive Small Molecules. *Chem Biol* 11, 609–618 (2004).

83. Fowler, D. M. & Fields, S. Deep mutational scanning: a new style of protein science. *Nat Methods* 11, 801–807 (2014).

84. Araya, C. L. & Fowler, D. M. Deep mutational scanning: assessing protein function on a massive scale. *Trends Biotechnol* 29, 435–442 (2011).

85. Shi, J. *et al.* Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. *Nat Biotechnol* 33, 661–667 (2015).

86. Donovan, K. F. *et al.* Creation of Novel Protein Variants with CRISPR/Cas9-Mediated Mutagenesis: Turning a Screening By-Product into a Discovery Tool. *Plos One* 12, e0170445 (2017).

87. Neggers, J. E. *et al.* Target identification of small molecules using large-scale CRISPR-Cas mutagenesis scanning of essential genes. *Nat Commun* 9, 502 (2018).

88. Ipsaro, J. J. *et al.* Rapid generation of drug-resistance alleles at endogenous loci using CRISPR-Cas9 indel mutagenesis. *Plos One* 12, e0172177 (2017).

89. Shen, C. *et al.* NSD3-Short Is an Adaptor Protein that Couples BRD4 to the CHD8 Chromatin Remodeler. *Mol Cell* 60, 847–859 (2015).

90. Hakimi, M.-A. *et al.* A core–BRAF35 complex containing histone deacetylase mediates repression of neuronal-specific genes. *Proc National Acad Sci* 99, 7420–7425 (2002).

91. Tong, J. K., Hassig, C. A., Schnitzler, G. R., Kingston, R. E. & Schreiber, S. L. Chromatin deacetylation by an ATP-dependent nucleosome remodelling complex. *Nature* 395, 917–921 (1998).

92. Forneris, F., Battaglioli, E., Mattevi, A. & Binda, C. New roles of flavoproteins in molecular cell biology: Histone demethylase LSD1 and chromatin. *Febs J* 276, 4304–4312 (2009).

93. Cusan, M. *et al.* LSD1 inhibition exerts its antileukemic effect by recommissioning PU.1- and C/EBPα-dependent enhancers in AML. *Blood* 131, 1730–1742 (2018).

94. 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).

95. Maes, T. *et al.* ORY-1001, a Potent and Selective Covalent KDM1A Inhibitor, for the Treatment of Acute Leukemia. *Cancer Cell* 33, 495-511.e12 (2018).

96. McGrath, J. P. *et al.* Pharmacological Inhibition of the Histone Lysine Demethylase KDM1A Suppresses the Growth of Multiple Acute Myeloid Leukemia Subtypes. *Cancer Res* 76, 1975–1988 (2016).

97. Mohammad, H. P. *et al.* A DNA Hypomethylation Signature Predicts Antitumor Activity of LSD1 Inhibitors in SCLC. *Cancer Cell* 28, 57–69 (2015).

98. Takagi, S. *et al.* LSD1 Inhibitor T-3775440 Inhibits SCLC Cell Proliferation by Disrupting LSD1 Interactions with SNAG Domain Proteins INSM1 and GFI1B. *Cancer Res* 77, 4652–4662 (2017).

99. Yamamoto, R. *et al.* Selective dissociation between LSD1 and GFI1B by a LSD1 inhibitor NCD38 induces the activation of ERG super-enhancer in erythroleukemia cells. *Oncotarget* 9, 21007–21021 (2018).

100. Lee, M. G., Wynder, C., Schmidt, D. M., McCafferty, D. G. & Shiekhattar, R. Histone H3 Lysine 4 Demethylation Is a Target of Nonselective Antidepressive Medications. *Chem Biol* 13, 563–567 (2006).

101. Kamburov, A. *et al.* Comprehensive assessment of cancer missense mutation clustering in protein structures. *Proc National Acad Sci* 112, E5486–E5495 (2015).

102. Dhanak, D. Drugging the cancer epigenom. (2013).

103. Du-Cuny, L., Xiao, Q., Xun, G. & Zheng, Q. Cyano-substituted indole compounds and uses thereof as lsd1 inhibitors. (2017).

104. Pinello, L. *et al.* Analyzing CRISPR genome-editing experiments with CRISPResso. *Nat Biotechnol* 34, 695–697 (2016).

105. Karasulu, B., Patil, M. & Thiel, W. Amine Oxidation Mediated by Lysine-Specific Demethylase 1: Quantum Mechanics/Molecular Mechanics Insights into Mechanism and Role of Lysine 661. *J Am Chem Soc* 135, 13400–13413 (2013).

106. Forneris, F., Orru, R., Bonivento, D., Chiarelli, L. R. & Mattevi, A. ThermoFAD, a Thermofluor®-adapted flavin ad hoc detection system for protein folding and ligand binding. *Febs J* 276, 2833–2840 (2009).

107. Tortorici, M. *et al.* Protein Recognition by Short Peptide Reversible Inhibitors of the Chromatin-Modifying LSD1/CoREST Lysine Demethylase. *Acs Chem Biol* 8, 1677–1682 (2013).

108. Clackson, T. *et al.* Redesigning an FKBP–ligand interface to generate chemical dimerizers with novel specificity. *Proc National Acad Sci* 95, 10437–10442 (1998).

109. Moroy, T., Vassen, L., Wilkes, B. & Khandanpour, C. From cytopenia to leukemia: the role of Gfi1 and Gfi1b in blood formation. *Blood* 24, 2561–2569 (2015).

110. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc National Acad Sci* 102, 15545–15550 (2005).

111. Pilotto, S. *et al.* Interplay among nucleosomal DNA, histone tails, and corepressor CoREST underlies LSD1-mediated H3 demethylation. *Proc National Acad Sci* 112, 2752–2757 (2015).

112. Heinz, S. *et al.* Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol Cell* 38, 576–589 (2010).

113. Dahl, R., Iyer, S. R., Owens, K. S., Cuylear, D. D. & Simon, M. C. The Transcriptional Repressor GFI-1 Antagonizes PU.1 Activity through Protein-Protein Interaction*. *J Biol Chem* 282, 6473–6483 (2007).

114. Cai, S. F., Chen, C.-W. & Armstrong, S. A. Drugging Chromatin in Cancer: Recent Advances and Novel Approaches. *Mol Cell* 60, 561–570 (2015).

115. Dawson, M. A. The cancer epigenome: Concepts, challenges, and therapeutic opportunities. *Science* 355, 1147–1152 (2017).

116. Dawson, M. A. & Kouzarides, T. Cancer Epigenetics: From Mechanism to Therapy. *Cell* 150, 12–27 (2012).

117. Helin, K. & Dhanak, D. Chromatin proteins and modifications as drug targets. *Nature* 502, 480–488 (2013).

118. Kasap, C., Elemento, O. & Kapoor, T. M. DrugTargetSeqR: a genomics- and CRISPR-Cas9–based method to analyze drug targets. *Nat Chem Biol* 10, 626–628 (2014).

119. Schenone, M., Dančík, V., Wagner, B. K. & Clemons, P. A. Target identification and mechanism of action in chemical biology and drug discovery. *Nat Chem Biol* 9, 232–240 (2013).

120. Joung, J. *et al.* Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. *Nat Protoc* 12, 828–863 (2017).

121. Pattanayak, V. *et al.* High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat Biotechnol* 31, 839–843 (2013).

122. Schulz-Fincke, J. *et al.* Structure-activity studies on N-Substituted tranylcypromine derivatives lead to selective inhibitors of lysine specific demethylase 1 (LSD1) and potent inducers of leukemic cell differentiation. *Eur J Med Chem* 144, 52–67 (2018).

123. Burg, J. M., Gonzalez, J. J., Maksimchuk, K. R. & McCafferty, D. G. Lysine-Specific Demethylase 1A (KDM1A/LSD1): Product Recognition and Kinetic Analysis of Full-Length Histones. *Biochemistry-us* 55, 1652–1662 (2016).

124. 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).

125. Dral, P. O., Wu, X., Spörkel, L., Koslowski, A. & Thiel, W. Semiempirical Quantum-Chemical Orthogonalization-Corrected Methods: Benchmarks for Ground-State Properties. *J Chem Theory Comput* 12, 1097–1120 (2016).

126. Liau, B. B. *et al.* Adaptive Chromatin Remodeling Drives Glioblastoma Stem Cell Plasticity and Drug Tolerance. *Cell Stem Cell* 20, 233-246.e7 (2017).

127. 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).

128. Shen, L., Shao, N., Liu, X. & Nestler, E. ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. *Bmc Genomics* 15, 284 (2014).

129. Still, W. C., Kahn, M. & Mitra, A. Rapid Chromatographic Technique for Preparative Separations with Moderate Resolution. *J. Org. Chem.* 43, 2923–2925 (1978).

130. Pangborn, A. B., Giardello, M. A., Grubbs, R. H., Rosen, R. K. & Timmers, F. J. Safe and Convenient Procedure for Solvent Purification. *Organometallics* 1518–1520 (1995).

131. McCafferty, D. G. & Pollock, J. Arylcyclopropylamines and methods of use. (2010).

132. Alberto, O. M. *et al.* (Hetero)Aryl Cyclopropylamine Compounds as LSD1 Inhibitors. (2013).

133. Davies, S. G., Mulvaney, A. W., Russell, A. J. & Smith, A. D. Parallel synthesis of homochiral β-amino acids. *Tetrahedron Asymmetry* 18, 1554–1566 (2007).

134. Ciaccio, J. A. & Aman, C. E. "Instant Methylide" Modification of the Corey–Chaykovsky Cyclopropanation Reaction. *Synthetic Commun* 36, 1333–1341 (2006).

135. Benelkebir, H. *et al.* Enantioselective synthesis of tranylcypromine analogues as lysine demethylase (LSD1) inhibitors. *Bioorgan Med Chem* 19, 3709–3716 (2011).

136. Miyamura, S. *et al.* C–H activation enables a rapid structure–activity relationship study of arylcyclopropyl amines for potent and selective LSD1 inhibitors. *Org Biomol Chem* 14, 8576–8585 (2016).

137. Egloff, C. *et al.* Bio-specific and bio-orthogonal chemistries to switch-off the quencher of a FRET-based fluorescent probe: application to living-cell biothiol imaging. *Chem Commun* 50, 10049–10051 (2014).

138. Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* 39, 311–318 (2007).

139. Bogdanović, O. *et al.* Dynamics of enhancer chromatin signatures mark the transition from pluripotency to cell specification during embryogenesis. *Genome Res* 22, 2043–2053 (2012).

140. Chepelev, I., Wei, G., Wangsa, D., Tang, Q. & Zhao, K. Characterization of genome-wide enhancer-promoter interactions reveals co-expression of interacting genes and modes of higher order chromatin organization. *Cell Res* 22, 490–503 (2012).

141. García-González, E., Escamilla-Del-Arenal, M., Arzate-Mejía, R. & Recillas-Targa, F. Chromatin remodeling effects on enhancer activity. *Cell Mol Life Sci* 73, 2897–2910 (2016).

142. Mosammaparast, N. & Shi, Y. Reversal of Histone Methylation: Biochemical and Molecular Mechanisms of Histone Demethylases. *Biochemistry-us* 79, 155–179 (2010).

143. Hnisz, D. *et al.* Super-Enhancers in the Control of Cell Identity and Disease. *Cell* 155, 934–947 (2013).

144. Lynch, J. T., Cockerill, M. J., Hitchin, J. R., Wiseman, D. H. & Somervaille, T. C. P. CD86 expression as a surrogate cellular biomarker for pharmacological inhibition of the histone demethylase lysine-specific demethylase 1. *Anal Biochem* 442, 104–106 (2013).

145. Boija, A. *et al.* Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell* 175, 1842-1855.e16 (2018).

146. Budayeva, H. G. & Cristea, I. M. Advancements of Mass Spectrometry in Biomedical Research. *Adv Exp Med Biol* 806, 263–282 (2014).

147. Zamudio, A. V. *et al.* Mediator Condensates Localize Signaling Factors to Key Cell Identity Genes. *Molecular Cell* (2019) doi:10.1016/j.molcel.2019.08.016.

148. Janicki, S. M. *et al.* From Silencing to Gene Expression Real-Time Analysis in Single Cells. *Cell* 116, 683–698 (2004).

149. Basu, S. *et al.* Unblending of Transcriptional Condensates in Human Repeat Expansion Disease. *Cell* 181, 1062-1079.e30 (2020).

150. Mueller, F., Mazza, D., Stasevich, T. J. & McNally, J. G. FRAP and kinetic modeling in the analysis of nuclear protein dynamics: what do we really know? *Curr Opin Cell Biol* 22, 403–411 (2010).

151. Mueller, F., Karpova, T. S., Mazza, D. & McNally, J. G. Monitoring Dynamic Binding of Chromatin Proteins In Vivo by Fluorescence Recovery After Photobleaching. *Methods Mol Biology* 833, 153–176 (2011).

152. Youmans, D. T., Schmidt, J. C. & Cech, T. R. Live-cell imaging reveals the dynamics of PRC2 and recruitment to chromatin by SUZ12-associated subunits. *Gene Dev* 32, 794–805 (2018).

153. Harrison, R. A. & Engen, J. R. Conformational insight into multi-protein signaling assemblies by hydrogen–deuterium exchange mass spectrometry. *Curr Opin Struc Biol* 41, 187–193 (2016).

154. Keul, N. D. *et al.* The entropic force generated by intrinsically disordered segments tunes protein function. *Nature* 563, 584–588 (2018).

155. Papanastasiou, M. *et al.* Structural Implications for the Formation and Function of the Complement Effector Protein iC3b. *J Immunol* 198, 3326–3335 (2017).

156. Mullahoo, J. *et al.* Dual protease type XIII/pepsin digestion offers superior resolution and overlap for the analysis of histone tails by HX-MS. *Methods* 184, 135–140 (2020).

157. Bell, C. C. *et al.* Targeting enhancer switching overcomes non-genetic drug resistance in acute myeloid leukaemia. *Nature Communications* 10, 2723 (2019).

158. Bushweller, J. H. Targeting transcription factors in cancer — from undruggable to reality. *Nat Rev Cancer* 19, 611–624 (2019).

159. Henley, M. J. & Koehler, A. N. Advances in targeting 'undruggable' transcription factors with small molecules. *Nat Rev Drug Discov* 20, 669–688 (2021).

160. Gupta, N. *et al.* CRISPR/Cas9-based Targeted Genome Editing for the Development of Monogenic Diseases Models with Human Pluripotent Stem Cells. *Curr Protoc Stem Cell Biology* 45, e50 (2018).

161. 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).

162. Luger, K., Rechsteiner, T. J. & Richmond, T. J. Preparation of Nucleosome Core Particle from Recombinant Histones. *Methods in Enzymology* 304, 3–19 (1999).

163. Dyer, P. N. *et al.* Reconstitution of Nucleosome Core Particles from Recombinant Histones and DNA. *Methods in Enzymology* 375, 23–44 (2003).

164. Garvie, C. W. *et al.* Structure of PDE3A-SLFN12 complex reveals requirements for activation of SLFN12 RNase. *Nat Commun* 12, 4375 (2021).

165. Zhao, S., Allis, C. D. & Wang, G. G. The language of chromatin modification in human cancers. *Nat Rev Cancer* 21, 413–430 (2021).

166. Pilotto, S. *et al.* LSD1/KDM1A mutations associated to a newly described form of intellectual disability impair demethylase activity and binding to transcription factors. *Hum Mol Genet* 25, 2578–2587 (2016).

167. Prantz, K. & Mulzer, J. Synthetic Applications of the Carbonyl Generating Grob Fragmentation. *Chem Rev* 110, 3741–3766 (2010).

168. Zhang, Z. *et al.* Structural and stereoelectronic insights into oxygenase-catalyzed formation of ethylene from 2-oxoglutarate. *Proc National Acad Sci* 114, 4667–4672 (2017).

169. Mimasu, S., Sengoku, T., Fukuzawa, S., Umehara, T. & Yokoyama, S. Crystal structure of histone demethylase LSD1 and tranylcypromine at 2.25Å. *Biochem Bioph Res Co* 366, 15–22 (2008).

170. Mimasu, S. *et al.* Structurally Designed trans-2-Phenylcyclopropylamine Derivatives Potently Inhibit Histone Demethylase LSD1/KDM1„. *Biochemistry-us* 49, 6494–6503 (2010).

171. Pisa, R. & Kapoor, T. M. Chemical strategies to overcome resistance against targeted anticancer therapeutics. *Nat Chem Biol* 16, 817–825 (2020).

172. Hattori, Y. *et al.* Design, synthesis, and structure–activity relationship of TAK-418 and its derivatives as a novel series of LSD1 inhibitors with lowered risk of hematological side effects. *Eur J Med Chem* 239, 114522 (2022).

173. Zibetti, C. *et al.* Alternative Splicing of the Histone Demethylase LSD1/KDM1 Contributes to the Modulation of Neurite Morphogenesis in the Mammalian Nervous System. *J Neurosci* 30, 2521–2532 (2010).

174. Rusconi, F., Grillo, B., Toffolo, E., Mattevi, A. & Battaglioli, E. NeuroLSD1: Splicing-Generated Epigenetic Enhancer of Neuroplasticity. *Trends Neurosci* 40, 28–38 (2017).

175. Toffolo, E. *et al.* Phosphorylation of neuronal Lysine-Specific Demethylase 1LSD1/KDM1A impairs transcriptional repression by regulating interaction with CoREST and histone deacetylases HDAC1/2. *J Neurochem* 128, 603–616 (2014).

176. Su, B. G. & Henley, M. J. Drugging Fuzzy Complexes in Transcription. *Frontiers Mol Biosci* 8, 795743 (2021).

177. Marceau, A. H. *et al.* An order-to-disorder structural switch activates the FoxM1 transcription factor. *Elife* 8, e46131 (2019).

178. Vacic, V. & Iakoucheva, L. M. Disease mutations in disordered regions—exception to the rule? *Mol Biosyst* 8, 27–32 (2011).

179. Rodrik-Outmezguine, V. S. *et al.* Overcoming mTOR resistance mutations with a new-generation mTOR inhibitor. *Nature* 534, 272–276 (2016).