Elucidating the Functions of CDK8 and CDK19 in Acute Myeloid Leukemia

Elucidating the Functions of CDK8 and CDK19

in Acute Myeloid Leukemia

A dissertation presented

by

Anupong Tangpeerachaikul

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

March 29, 2018

Dissertation Advisor: Professor Matthew D. Shair Anupong Tangpeerachaikul

Elucidating the Functions of CDK8 and CDK19 in Acute Myeloid Leukemia

Abstract

Transcriptional dysregulation is an important cause of acute myeloid leukemia (AML). The majority of mutations in AML reside in genes that encode transcriptional regulators including transcription factors, chromatin modifiers, DNA methylation proteins, and cohesin. Inhibition of these and other transcriptional regulators has emerged as a promising therapeutic strategy in AML.

This dissertation describes our development of marine natural product cortistatin A (CA) as a potent and selective inhibitor of CDK8 and CDK19 (CDK8/19)—two kinases that reversibly associate with a general transcriptional co-activator complex called Mediator. We established that CDK8/19 are therapeutic targets in AML, as their inhibition suppresses AML growth both in cell culture and in animal models. Using CA, we also discovered previously unknown functions of CDK8/19 in restricting the expression of super-enhancer-associated genes in AML. CDK8/19 inhibition leads to the upregulation of these genes, which are in part responsible for the antiproliferative activity of CA. The effect of CA on super-enhancer activity is opposite to that of BRD4 or CDK7 inhibitors (which downregulate super-enhancer-associated genes), yet all of these inhibitors have equivalent efficacy in AML. This observation suggests that AML exerts tight

control over the expression of super-enhancer-associated genes, and perturbations to the expression of these genes in either direction are sufficient to suppress AML growth.

To further understand the antiproliferative mechanism of CDK8/19 inhibition in AML cells, we performed a genome-wide CRISPR-Cas9 suppressor screen with CA in an AML cell line, and we found that Notch signaling is essential for CA's activity. CA inhibits CDK8/19-mediated degradation of the Notch1 and Notch2 intracellular domains (NICD1/2), thereby activating Notch signaling which suppresses AML growth. We also found that expression of Notch ligands Dll1 and Jagged1 is an important pre-requisite for the cell's response to CA along the Notch axis.

Lastly, we identified >60 previously unknown substrates of CDK8/19 through a phosphoproteomics experiment in combination with CA treatment. Most of these substrates associate either directly or indirectly with the chromatin—an observation that is consistent with the transcriptional role of CDK8/19—and span many functional annotations including DNA repair, transcription factors, and RNA polymerase II.

Our findings have not only expanded the field's understanding of CDK8/19 and their roles in transcriptional regulation, cell biology, and cancer, but also opened up new avenues for therapeutic developments of CDK8/19 inhibitors for the treatment of AML and other diseases.

Acknowledgements

First and foremost, I owe my success to the people who have given me most and taken from me least—my family. Coming from a very humble background, dad and mom made up for their deficiencies by being the most hard-working, most disciplined, and most persevering people I have ever known. Their relentless labor had only one selfless goal: to provide me and my brothers with good education and a happy life—a life more comfortable than that which they lived. I have spent the past 12 years abroad, chasing after my dreams, not realizing what my parents had to give up for my ambitions. There are few things more painful to fond parents than sitting inside an ever emptier house, awaiting their children to come home and join them at the dining table once again. Thank you for your unconditional love, especially because it has come at a great cost of your own happiness. I hope that my achievement today has done you proud.

During my six years at Harvard University, I lost both grandmothers (my grandfathers had already passed away before I was born). Our extended families were very close. I regret not being able to spend more time with them in the final years of the lives. I wish they were around to enjoy my success—wish I could make them proud. This thesis is in loving memory of them.

I am forever grateful to my teachers, mentors, and advisors who have had an immeasurable influence on shaping my intellectual capacity, morality, and aspirations. These people have not only guided me in my journey toward academic excellence, but also constantly reminded me to always put the wellbeing of others above personal wealth and fame—to use my knowledge for the betterment of the society, for the greater good. In particular, I would like to thank (1) Principal

Thongchai Chewpreecha and Ms. Sasinee Angkanont, of Mahidol Wittayanusorn School, who first kindled my passion for science and taught me about the impact of science and technology on the society; (2) my undergraduate research advisor and mentors Professor Alice Ting, Dr. Daniel Liu, and Dr. Chayasith (Tao) Uttamapinant, of Massachusetts Institute of Technology, for their impeccable mentorship that has transformed me from a clueless student into a competent scientist; and (3) my graduate research advisor and mentors Professor Matthew Shair, Dr. Henry Pelish, and Professor Brian Liau, as well as my thesis committee members Professor Stuart Schreiber and Professor Randall King, all of Harvard University, for sharing with me their penetrating scientific insights into solving complex biological problems, and for guiding me to my success today. I am also thankful to many other teachers from Sangarun School, Assumption College, Mahidol Wittayanusorn School, Red Cross Nordic United World College, Massachusetts Institute of Technology, and Harvard University who have left an enduring impression on my life and whose expectations I will strive to live up to. The entirety of my success is the fruits of your kind cultivation.

I count myself among the luckiest men to be friends with so many kind and caring souls, who have filled my life with joy, comforted me in hard times, and shared with me both happy and sorrowful memories. I will never forget that I have been able to soar so high because of the gentle winds that carry me beneath my wings. Thank you Panot, Chavanont, Nitipat, Pongsakorn, Thanaboon, Jirayu, Dhanakara, and others from Assumption College; Poramapa, Pimpun, Apawee, Pornhatai, Tanatorn, Pichet, Thirawat, and others from Mahidol Wittayanusorn School; and Adrienne, Anibal, Karine, Min Sern, Pinkie, Sabrina, Turid, Yukiko, Milton, Kjetil and Trude Felde and family, and others from my high school years in Norway. At the time of writing this

dissertation, I am particularly reminded of the numerous friends I have made during my 10 years in Cambridge: Tongjai, Chayasith, Sutheera, Phumphong, Itthi, Nopphon, Aniwat, Ittinop, Atikhun, Pornchai, Phitchaya, Techin, Ivana, Polnop, Pantra, Sasilada, Tana, Payut, Supanat, Pasin, Khetpakorn, Nitipat, Supisara, Nitis, Nipun, Puwanat, Kritkorn, Korrawat, Cattaleya, Apisada, Suchan, Thee, Wachara, Glenda and Don Mattes, Ting Lab members, Shair Lab members, and so many others. You are my shelter, my sanctuary, and because of you I have felt at home everywhere I have been. I thank you all from the bottom of my heart.

Table of Contents

Abstract iii

Acknowledgements v

Table of Contents viii

List of Figures and Tables xi

List of Abbreviations xiv

Chapter 1 | Introduction to Acute Myeloid Leukemia 1

Introduction 2

Pathophysiology 2

Classification 5

Genomic landscape 7

AML as disease of dysregulated transcription 11

Therapies 16

Conclusion 21

References 22

Chapter 2 | CDK8 and CDK19 as Therapeutic Targets in AML 28

Introduction 29

Mediator complex 29

CDK8 module 31

Super-enhancers 36

Cortistatin A 39

CA as a CDK8 and CDK19 inhibitor in AML 42

CA further upregulates super-enhancer-associated genes 49

Super-enhancer activation is growth inhibitory in AML cells 54

Conclusion 57

Methods 58

References 75

Chapter 3 | Notch Signaling Mediates the Mechanism of CA in AML 79

Introduction 80

Functional genomics 80

Using CRISPR-Cas9 screens to elucidate CA's mechanism in AML 88

CRISPR-Cas9 screen implicates Notch signaling CA's mechanism 93

CA inhibits NICD1 degradation 100

NICD1 mediates the antiproliferative activity of CA in AML 104

Notch ligands are required for NICD1 stabilization by CA 107

CA inhibits NICD2 degradation 111

NICD2 mediates the antiproliferative activity of CA in AML 114

Discussion 116

Conclusion 119

Methods 120

References 128

Chapter 4 | Identification of CDK8 and CDK19 Substrates **137**

Introduction 138

Phosphoproteomics 138

Using CA to identify phosphorylation substrates of CDK8/19 140

Quantitative phosphoproteomics in HCT116 cells ± CA 143

CDK8/19 substrates are largely transcription-associated proteins 145

Validation of selected CDK8/19 substrates 153

CDK8/19 inhibition has limited effects on transcription 157

Cellular proteome changes resulting from CDK8/19 inhibition 162

Discussion 164

Conclusion 174

Methods 175

References 181

Chapter 5 | Conclusions and Prospects **190**

List of Figures and Tables

Chapter 1 | Introduction to Acute Myeloid Leukemia

Figure 1.1 | Hematopoiesis 4

Figure 1.2 | Mutational frequencies observed in exomes of many cancers 8

Figure 1.3 | Recurring genetic mutations in AML 10

Figure 1.4 | Targeted therapies for AML 17

Table 1.1 | The French-American-British (FAB) classification of AML 5

Table 1.2 | The World Health Organization (WHO) classification of AML 6

Table 1.3 | Mutational landscape of AML 8

Table 1.4 | Another mutational landscape of AML 9

Chapter 2 | CDK8 and CDK19 as Therapeutic Targets in AML

Figure 2.1 | The Mediator complex in transcriptional regulation 31

Figure 2.2 | Protein sequence alignment of CDK8 and CDK19 33

Figure 2.3 | Modulation of transcription factors in various pathways by CDK8 36

Figure 2.4 | Super-enhancers 37

Figure 2.5 | Cortistatin A and our group's retrosynthetic plan 39

Figure 2.6 | KinomeSCAN assay 40

Figure 2.7 | CA as an inhibitor of CDK8 activity in vitro and in cells 43

Figure 2.8 | Biochemical characterization of CA selectivity 46

Figure 2.9 | Effects of CA on the proliferation of AML cell lines 48

Figure 2.10 | CDK8 occupies super-enhancers in MOLM-14 cells 50

Figure 2.11 | Effects of CA on super-enhancer-associated genes 53

Figure 2.12 | *In vivo* efficacy of CA 56

Table 2.1 | Some of the transcription factors known to be phosphorylated by CDK8 33

Table 2.2 | Targets of CA based on a KINOMEscan assay 41

Table 2.3 | Gene Ontology of genes associated with super-enhancers in MOLM-14 cells 51

Chapter 3 | Notch Signaling Mediates the Mechanism of CA in AML

Figure 3.1 | Chemical and biological tools for functional genomics 83

Figure 3.2 | ZFNs and TALENs 85

Figure 3.3 | CRISPR components and mechanism 87

Figure 3.4 | Notch pathway 92

Figure 3.5 | CRISPR screen workflow 94

Figure 3.6 | CRISPR screen identifies Notch signaling as important for CA's mechanism 95

Figure 3.7 | Knockout of Notch-component genes 98

Figure 3.8 | Notch-component knockout impairs sensitivity to CA 99

Figure 3.9 | CA stabilizes NICD1 and activates Notch signaling 102

Figure 3.10 | Notch as a tumor suppressor in AML 106

Figure 3.11 | Notch ligands are required for CA mechanism 109

Figure 3.12 | CA stabilizes NICD2 112

Figure 3.13 | Both Notch1 and Notch2 contribute to CA's activity in AML 115

Table 3.1 | GO analysis of CRISPR-Cas9 screen hits 97

Chapter 4 | Identification of CDK8 and CDK19 Substrates

Figure 4.1 | Phosphoproteomics experiment workflow 140

Figure 4.2 | Quantitative phosphoproteomics in HCT116 cells ± CA 144

Figure 4.3 | High confidence Mediator kinase substrates 151

Figure 4.4 | *In vitro* validation of select CDK8/19 substrates 155

Figure 4.5 | CDK8/19 inhibition is functionally distinct from CDK8/19 knockdown 158

Figure 4.6 | Proteomics reveals pathways and proteins affected by CDK8/19 inhibition 165

Table 4.1 | Identification of Mediator kinase (CDK8/19) substrates 146

List of Abbreviations

AML	Acute myeloid leukemia
APL	Acute promyelocytic leukemia
ATO	Arsenic trioxide
ATP	Adenosine triphosphate
ATRA	All-*trans* retinoic acid
CA	Cortistatin A
Cas9	CRISPR-associated protein 9
CDK8	Cyclin-dependent kinase 8
CDK19	Cyclin-dependent kinase 19
CDK8/19	CDK8 and CDK19
ChIP	Chromatin immunoprecipitation
ChIP-seq	ChIP followed by massive parallel sequencing
CRISPR	Clustered regularly interspaced short palindromic repeats
Cryo-EM	Cryogenic electron microscopy
CXMS	Crosslinking mass spectrometry
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
EC_{50}	Half maximal effective concentration
EDTA	Ethylenediaminetetraacetic acid
ERLIC	Electrostatic repulsion-hydrophilic interaction chromatography
FAB	French-American-British, a system for AML classification

FACS	Fluorescence-activated cell sorting
FDR	False discovery rate
GI_{50}	Half maximal growth inhibition concentration
GO	Gene ontology
GSEA	Gene set enrichment analysis
GSI	Gamma-secretase inhibitor
IC_{50}	Half maximal inhibitory concentration
Indel	Insertion and deletion
IRES	Internal ribosome binding site
kb	Kilobase
K_d	Dissociation constant
k_{off}	Dissociation rate constant
k_{on}	Association rate constant
LC	Liquid chromatography
Mb	Megabase
MLL	Mixed-lineage leukemia
MPN	Myeloproliferative neoplasms
mRNA	Messenger RNA
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MSigDB	Molecular signatures database
NES	Normalized enrichment score
NICD	Notch intracellular domain

PCR	Polymerase chain reaction
Pol II	RNA polymerase II
qPCR	Quantitative PCR
qRT-PCR	Quantitative reverse-transcription PCR
RNA	Ribonucleic acid
RNA Pol II	RNA polymerase II
RNA-seq	RNA sequencing
RNAi	RNA interference
sgNT	Non-targeting sgRNA
sgRNA	Single guide RNA
shRNA	Short hairpin RNA
siRNA	Small interfering RNA
SILAC	Stable isotope labeling with amino acids in cell culture
TALEN	Transcription activator-like effector nuclease
T-ALL	T-cell acute lymphoblastic leukemia
T7E1	T7 endonuclease 1
TSS	Transcriptional start site
WHO	World Health Organization
ZFN	Zinc-finger nuclease
λPP	Lambda phosphatase

Chapter 1

Introduction to Acute Myeloid Leukemia

Introduction

This dissertation aims to examine the role of CDK8 and CDK19 in acute myeloid leukemia (AML). AML is a cancer of over-proliferative blood stem and progenitor cells that, like any other cancer, can be caused by a variety of genetic mutations. This chapter provides an overview of AML including its pathophysiology, classification, mutational landscape, and therapeutic strategies. Specifically, we focus our attention on the description of AML as a disease of dysregulated transcription. Mutations in AML patients reside not only in the genes that directly control cell proliferation or cell death (such as the genes encoding receptor tyrosine kinases) but also in the genes that encode various regulators of transcription such as DNA methylation proteins, chromatin modifiers, the cohesin complex, and transcription factors. This observation suggests that transcriptional dysregulation is an important cause of AML. As such, molecules that modulate gene transcription, either by directly intercepting these mutated proteins or by modulating the activity of other non-mutated transcriptional regulators, have emerged as promising therapeutic strategies. This chapter describes some of these targets and how their inhibition achieves efficacy, paving the way into the next chapter where we describe our discovery of the transcriptional regulators CDK8 and CDK19 as new therapeutic targets in AML.

Pathophysiology

AML is a form of blood cancer that arises from over-proliferative, poorly differentiated myeloid progenitor cells, which are the stem cells that produce various types of blood cells in the myeloid lineage (**Figure 1.1**). These malignant myeloid progenitors accumulate in large quantities in the bone marrow, peripheral blood, and other organs, causing the population of other blood

types to decline. Patients suffer from conditions of hematological imbalance such as anemia (lack of red blood cells, which leads to fatigue), thrombocytopenia (lack of platelets, which leads to excessive bleeding), and compromised immunity (lack of immune cells, which increases susceptibility to infection). Death occurs within months if left untreated. Among the four major forms of leukemia (acute myeloid leukemia, acute lymphoid leukemia, chronic myeloid leukemia, and chronic lymphoid leukemia), AML is the deadliest and second most common, with the overall 5-year survival rate of 25%, 20,000 new cases, and 10,000 deaths in the United States in 2014 (Leukemia & Lymphoma Society).

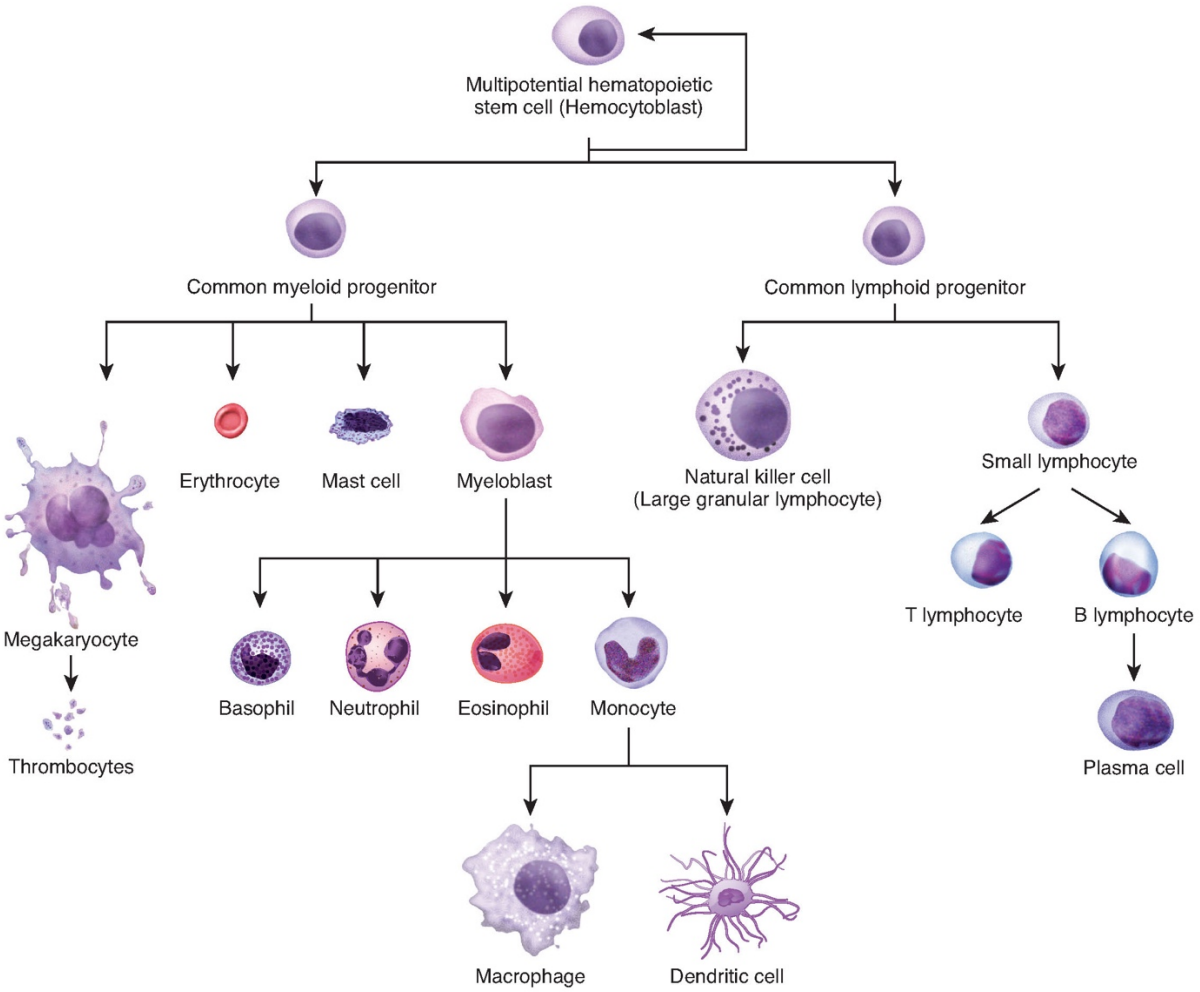


Figure 1.1 | Hematopoiesis. Reproduced from original image by OpenStax. Obtained with license: [https://commons.wikimedia.org/wiki/File:0337 Hematopoiesis new.jpg, https://creativecommons.org/licenses/by/4.0/legalcode].

Classification

There are two primary classification systems for AML, which were developed by the French-American-British (FAB) collaboration and the World Health Organization (WHO). The FAB system, which was proposed in 1976, divides AML into 8 subtypes based on the morphology and staining of bone marrow cells observed under the microscope (**Table 1.1**). Each of these subtypes shows characteristics similar to various myeloid precursor cells, which include the granulocytes, the monocytes, the erythrocytes, and the megakaryocytes (**Figure 1.1**).

Table 1.1 | The French-American-British (FAB) classification of AML

FAB subtype	Name
M0	Undifferentiated acute myeloid leukemia
M1	Acute myeloid leukemia with minimal maturation
M2	Acute myeloid leukemia with maturation
M3	Acute promyelocytic leukemia
M4	Acute myelomonocytic leukemia
M5	Acute monocytic leukemia
M6	Acute erythroid leukemia
M7	Acute megakarayoblastic leukemia

The WHO system, first proposed in 2001 and updated in 2008 (Vardiman et al., 2009) and again in 2016 (Arber et al., 2016), further incorporates the understanding of the genetics, morphology, cytochemistry, immunophenotype, and clinical presentation of AML, and divides

AML into 6 subtypes (**Table 1.2**). By taking into account the molecular basis of the disease, the WHO system offers not only subtype stratification but also prognostic values beyond the FAB system. Some of these genetic abnormalities will be explained in the next section.

Table 1.2 | The World Health Organization (WHO) classification of AML

Subtype	Abnormalities
AML with recurrent genetic abnormalities	t(8:21)(q22;q22); *RUNX1-RUNX1T1* inv(16)(p13.1q22) or t(16;16)(p13.1;q22); *CBFB-MYH11* Acute promyelocytic leukemia with *PML-RARA* t(9;11)(p21.3;q23.3); *MLLT3-KMT2* t(6;9)(p23;q34.1); *DEK-NUP214* inv(3)(q21.3q26.2) or t(3;3)(q21.3;q26.2); *GATA2, MECOM* (megakaryoblastic) t(1;22)(p13.3;q13.3); *RBM15-MKL1* *BCR-ABL1* (provisional entity) mutated *NPM1* biallelic mutations of *CEBPA* mutated *RUNX1* (provisional entity)
AML with myelodysplasia-related changes	
Therapy-related myeloid neoplasms	
AML not otherwise specified	AML with minimal differentiation

	AML without maturation AML with maturation Acute myelomonocytic leukemia Acute monoblastic/monocytic leukemia Pure erythroid leukemia Acute megakaryoblastic leukemia Acute basophilic leukemia Acute panmyelosis with myelofibrosis
Myeloid sarcoma	
Myeloid proliferations related to Down syndrome	Transient abnormal myelopoiesis Myeloid leukemia associated with Down syndrome

Genomic landscape

AML has one of the fewest mutations among all cancers, with the mutational frequency of 0.37 per megabase in the exome compared to 0.1-100 per megabase for melanoma or lung cancer (**Figure 1.2**) (Kandoth et al., 2013; Lawrence et al., 2013). In a study of 200 AML patient samples, The Cancer Genome Atlas Research Network discovered ~2,500 overall somatic mutations in the coding region, with an average of 13 mutations per AML genome (**Figure 1.3**) (The Cancer Genome Atlas Research Network, 2013). These mutations spread over ~1,600 genes, but only 260 genes were recurrently mutated in at least 2 samples. The authors categorized the mutant genes into 9 groups based on biological functions (**Table 1.3**).

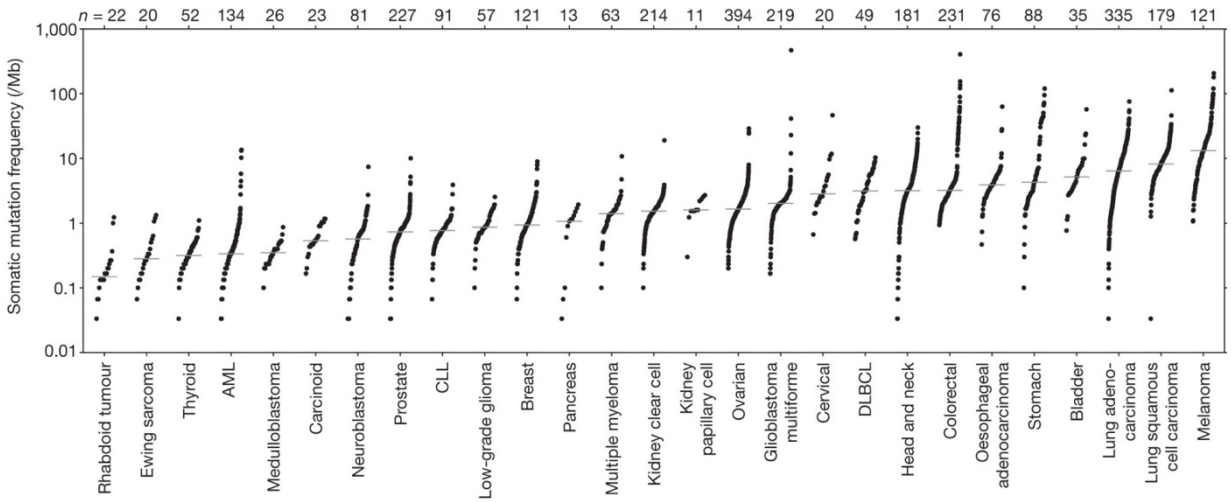


Figure 1.2 | Mutational frequencies observed in exomes of many cancers. n denotes the number of cancer samples used for analysis. CLL = chronic lymphocytic leukemia; DLBCL = diffuse large B-cell lymphoma; Mb = megabase. Image reproduced from Lawrence et al., 2013.

Table 1.3 | Mutational landscape of AML (The Cancer Genome Atlas Research Network, 2013)

Category	Prevalence	Examples of genetic abnormalities
Transcription factor fusions	18%	*PML-RARA, MYH11-CBFB, RUNX1-RUNX1T1*
DNA methylation	44%	*DNMT3A, TET2, IDH1, IDH2*
Chromatin modifiers	30%	*MLL* fusions, *ASXL1, EZH2*
Transcription factors	22%	*RUNX1, CEBPA*
Cohesin complex	13%	*STAG2, RAD21*
Spliceosome complex	14%	*SRSF2, SF3B1, U2AF1*
Nucleophosmin	27%	*NPM1*
Tumor suppressors	16%	*TP53, WT1, PHF6*
Activated signaling	59%	*FLT3, KIT, RAS*

Some of these mutational categories do not co-occur within the same sample. For example, transcription factor fusions are exclusive with mutations in DNA methylation proteins, transcription factors, or nucleophosmin. Such mutual exclusivity suggests that these mutations are functionally redundant; i.e., the same phenotype can be reached via mutations in parallel processes (The Cancer Genome Atlas Research Network, 2013; Patel et al., 2012).

Another study of 1,540 AML patient samples identified 5,234 driver mutations in 76 genes and classified AML into 11 categories based on patterns of co-mutation (**Table 1.4**). Analysis of allele fractions in heterogeneous AML cell populations showed that mutations in some genes, such as *DNMT3A*, *IDH1*, *IDH2*, *TET2*, and *ASXL1*, were acquired early on in the disease and never occurred alone, suggesting that these mutations poised the cell to become malignant but were not sufficient to induce full-blown AML by themselves (Papaemmanuil et al., 2016).

Table 1.4 | Another mutational landscape of AML (Papaemmanuil et al., 2016)

Category	Prevalence
NPM1 mutations	27%
Chromatin-spliceosome mutations	18%
TP53 mutations and/or aneuploidy	13%
CBFB-MYH11	5%
CEBPA mutations	4%
PML-RARA	4%
RUNX1-RUNX1T1	4%

MLL fusions	3%
GATA2, MECOM	1%
IDH2-R172	1%
DEK-NUP214	1%
Others	19%

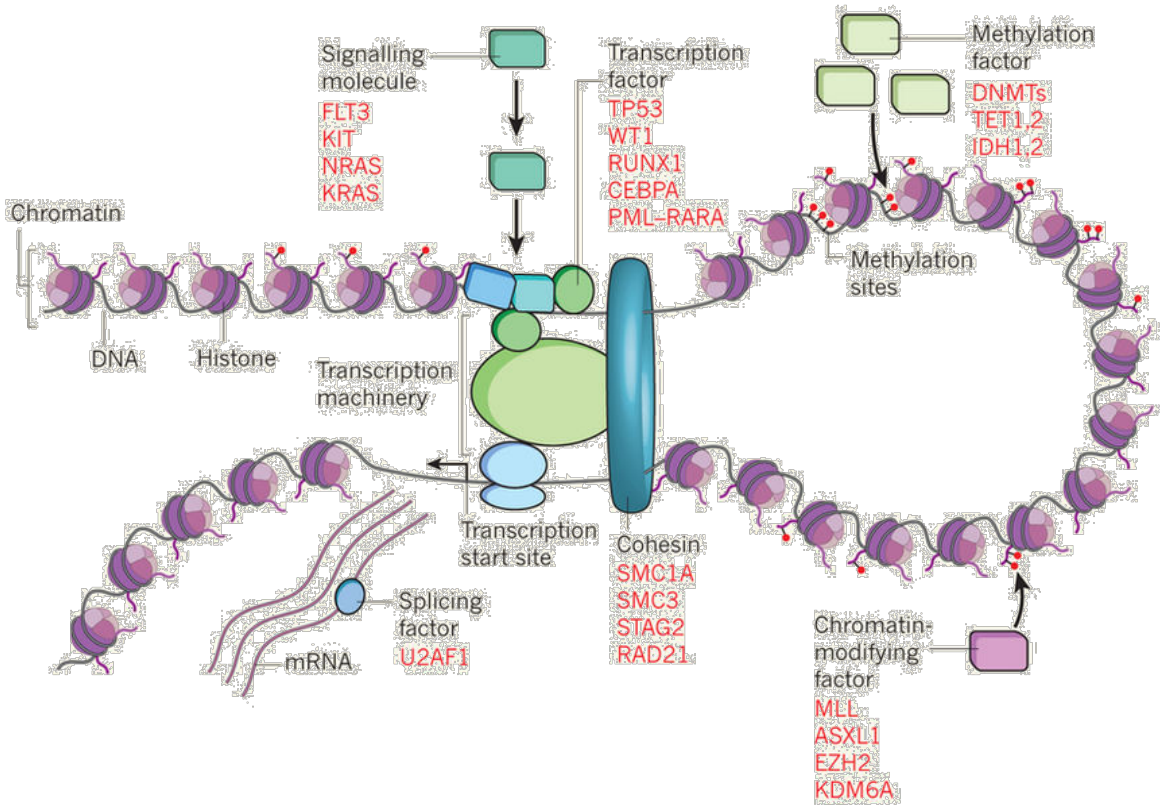


Figure 1.3 | Recurring genetic mutations in AML. Image reproduced from Aerts and Cools, 2013.

AML as a disease of dysregulated transcription

These genomic analyses of AML patient samples reveal that a substantial proportion of mutations reside in genes that encode transcriptional regulators such as transcription factors, transcription factor fusions, MLL fusions, DNA methylation proteins, chromatin modifiers, and cohesin (**Figure 1.3**). Given the known functions of these transcriptional regulators in hematopoietic development, their mutations may contribute to leukemogenesis by promoting self-renewal or blocking differentiation. Some of these genetic aberrations (most notably *MLL* fusions) are sufficient to induce AML on their own, whereas other aberrations (*DNMT3A*, *TET2*, and *IDH1/IDH2* mutations) are established early in founding leukemia clones but require a second hit (typically *FLT3* or *NPM1* mutations) to fully induce AML (Abdel-Wahab and Levine, 2013; Corces-Zimmerman et al., 2014; Welch et al., 2012).

The following section reviews our understanding of how mutations in transcriptional regulators contribute to the pathogenesis of AML. It should be noted that while this thesis focuses on transcriptional dysregulation in AML, activated signaling pathways and loss of tumor suppressors, such as *FLT3* and *TP53*, are also important drivers of this disease.

- **DNA methylation proteins (DNMT3A, TET2, IDH1, and IDH2)**

  Methylation of cytosine in CpG dinucleotides is an important epigenetic mechanism for controlling gene expression. Mammalian genomes contain short DNA stretches (<3 kb) with high density of CpGs called CpG islands, which are frequently found at or near promoters. CpG islands are usually unmethylated, and their methylation is

generally associated with transcriptional repression. DNMT1, DNMT3A, and DNMT3B catalyze the methylation of cytosine into 5-methylcytosine using *S*-adenosyl methionine as a cofactor. In the opposite direction, TET1 and TET2 utilize 2-oxoglutarate to oxidize the 5-methyl group on methylated cytosine into 5-hydroxymethyl, 5-formyl, and 5-carboxyl, which can subsequently decarboxylate to restore unmethylated cytosine.

DNMT3A is deactivated in 20-30% of all AML. The most common mutation is a missense, nonsense, or frameshift mutation at R822. DNMT3A is important for suppressing the self-renewal capacity of hematopoietic stem cells, as *DNMT3A*-defective cells exhibit hypomethylation (and hence activation) of CpG islands associated with master regulators (*GATA3*, *RUNX1*, and *HOX*), and malignancy genes (*STAT1*, *PRDM16*, *CCND1*, *MYC*, and *ERG*).

TET2 loss-of-function mutations are found in 7-23% of AML, and gain-of-function *IDH1*-R132, *IDH2*-R140, and *IDH2*-R172 mutations are found in ~15% of AML. *IDH1* and *IDH2* encode isocitrate dehydrogenases that catalyze the conversion of isocitrate into 2-oxoglutarate, the redox cofactor of TET2. Mutant IDHs acquire the ability to further reduce 2-oxoglutarate into 2-hydroxyglutarate, an oncometabolite that inhibits the enzymatic activity of TET2. Thus, *TET2* loss-of-function mutations and *IDH* gain-of-function mutations are functionally similar. *IDH* mutations and *TET2* mutations are also mutually exclusive, further supporting their redundancy in leukemogenesis. Both *IDH* and *TET2* mutants show a hypermethylation phenotype. (Guillamot et al., 2016; Schoofs et al., 2014)

- **Transcription factors CEBPA and RUNX1**

CEBPA is a transcription factor important for granulocytic development, and its deficiency blocks the differentiation of common myeloid progenitors into granulocyte-monocyte progenitors in mice. Loss-of-function mutations across the coding region of *CEBPA* are found in 15% of all AML. *CEBPA* mutations are mutually exclusive with *PML-RARA* and *RUNX1-RUNX1T1*. (Fasan et al., 2014; Pabst and Mueller, 2009)

RUNX1, also known as AML1, is a master regulator that plays an essential role in the establishment of hematopoietic stem cells during development and in myeloid differentiation. Loss-of-function mutations in *RUNX1* are found in 5-10% of all AML. *RUNX1* knockout in hematopoietic cells impedes normal blood development and contributes to leukemogenesis. Moreover, *RUNX1* participates in t(8:21)(q22;q22), the most common chromosomal translocation in AML with prevalence of 12%, producing the *RUNX1-RUNX1T1* (*AML1-ETO*) fusion gene. RUNX1T1 itself is important for the development of the gut and the central nervous system but not for hematopoiesis. When fused to RUNX1, however, RUNX1T1 is thought to recruit transcriptional repressors such as the NCoR/SMRT complex and histone deacetylases to RUNX1 target genes. Mice expressing *RUNX1-RUNX1T1* have impaired blood development, albeit with a milder phenotype than *RUNX1*-knockout mice. *RUNX1* mutation or *RUNX1-RUNX1T1* expression alone is not sufficient to induce AML. (Lam and Zhang, 2012; Mangan and Speck, 2011)

- **MLL fusions**

MLL (Mixed-Lineage Leukemia), also known as *KMT2A*, is frequently translocated in AML, with prevalence of 5-10%. MLL is a histone H3K4 methyltransferase that plays an essential role in development, especially in the maintenance of homeobox (*HOX*) genes in hematopoietic stem cells. The methyltransferase activity of MLL does not seem to be required for leukemogenesis, as the catalytic SET domain is not part of the fusion protein. Instead, MLL is thought to bring along its fusion partner to target genomic loci through interaction with menin and LEDGF. The most common MLL translocation partners—AF4, AF9, and ENL—can further recruit the transcriptional elongation complex (pTEF-b), the polycomb repressive complex 2 (PRC2), and the H3K79 methyltransferase DOT1L, all of which have been shown to be essential for the transformation of AML cells by MLL fusions. Expression of MLL fusions upregulates *HOX* genes and *MEIS1* which are characteristic of blood stem cells and leukemia. (de Boer et al., 2013; Winters and Bernt, 2017)

- **ASXL1**

ASXL1 is a DNA-binding protein that recruits the polycomb repressive complexes PRC1 and PRC2 to target genes to effect transcriptional control. ASXL1 truncation mutations lack the DNA-binding PHD domain, which eliminates its normal activity, and are found in 10% of AML. (Alvarez Argote and Dasanu, 2017; Pratcorona et al., 2012)

- **Cohesin complex (STAG2 and RAD21)**

The cohesin complex, which comprises four subunits (SMC1, SMC3, RAD21, and STAG1 or STAG2), mediates the cohesion between sister chromatids during DNA replication. Cohesin also plays an important role in regulating chromatin architecture by assisting in the formation of three-dimensional DNA loops. Missense or frameshift mutations in *STAG2* or *RAD21* are found in 13% of all AML. These mutations do not seem to affect chromosome segregation, suggesting that the cohesion function and the transcriptional function of cohesin can be uncoupled. Indeed, cohesin disruption has been shown to upregulate *HOX* genes by reducing PRC2 recruitment, alter chromatin accessibility to *ERG*, *RUNX1*, and *GATA2* loci, and enhance the self-renewal capability of hematopoietic stem and progenitor cells. (Fisher et al., 2017)

- **PML-RARA**

The PML-RARA translocation is found in 4% of all AML, giving rise to a specific subtype of AML called acute promyelocytic leukemia (promyelocytes are precursors to granulocytes). RARA is a hormone nuclear receptor that functions as a retinoic acid-dependent switch. In the absence of the ligand, RARA represses target genes by recruiting repressive complexes and histone deacetylases. Upon binding retinoic acid, however, RARA functions as a transcriptional activator for genes that are important for myeloid differentiation. PML-RARA is a poor transcriptional activator even in the presence of the ligand, and it may induce stem-like properties in AML cells by recruiting PRC1/PRC2 to target genes or by inactivating genes that encode myeloid-specific transcription factors

such as *PU.1*. Furthermore, PML itself is an organizer of nuclear bodies, which regulate cell death. The fusion of PML to RARA impairs programmed cell death and contributes to AML growth. (dos Santos et al., 2013; de Thé and Chen, 2010; de Thé et al., 2012)

Therapies

- **Chemotherapy and stem cell transplantation therapy**

  The standard treatment for AML includes induction therapy and consolidation therapy. Induction therapy involves the '7+3' regimen of treatment with cytarabine for 7 days followed by anthracycline for 3 days. Both cytarabine and anthracycline inhibit rapidly proliferating cells by interfering with DNA replication. Other chemotherapeutic agents may be used. Most patients (60-80%) achieve complete remission after induction therapy, but residual disease must be further treated with consolidation therapy to minimize relapse. Consolidation therapy involves either repeated cycles of chemotherapy or an allogeneic hematopoietic stem cell transplant.

Apart from cytotoxic chemotherapy, there are also targeted therapies as described below (**Figure 1.4**) (Neff and Armstrong, 2013; Saygin and Carraway, 2017).

- **Therapeutic agents targeting cell proliferation pathways**

  FLT3-ITD (internal tandem duplication) is one of the most common mutations in AML, with prevalence of ~25%. FLT3 is a receptor tyrosine kinase upstream of many cell signaling pathways, and the constitutively activated FLT3-ITD mutant promotes AML

growth. Both promiscuous and specific inhibitors of FLT3-ITD have been developed including sorafenib, midostaurin, quizartinib, and crenolanib. Generally, there is strong initial response of AML cells bearing *FLT3-ITD* mutations to these drugs, but patients invariably develop resistance through a D835Y or D835H mutation in FLT3, or through upregulation of the receptor. FLT3 inhibitors therefore have only limited efficacy unless administered with other therapeutic agents. (Larrosa-Garcia and Baer, 2017)

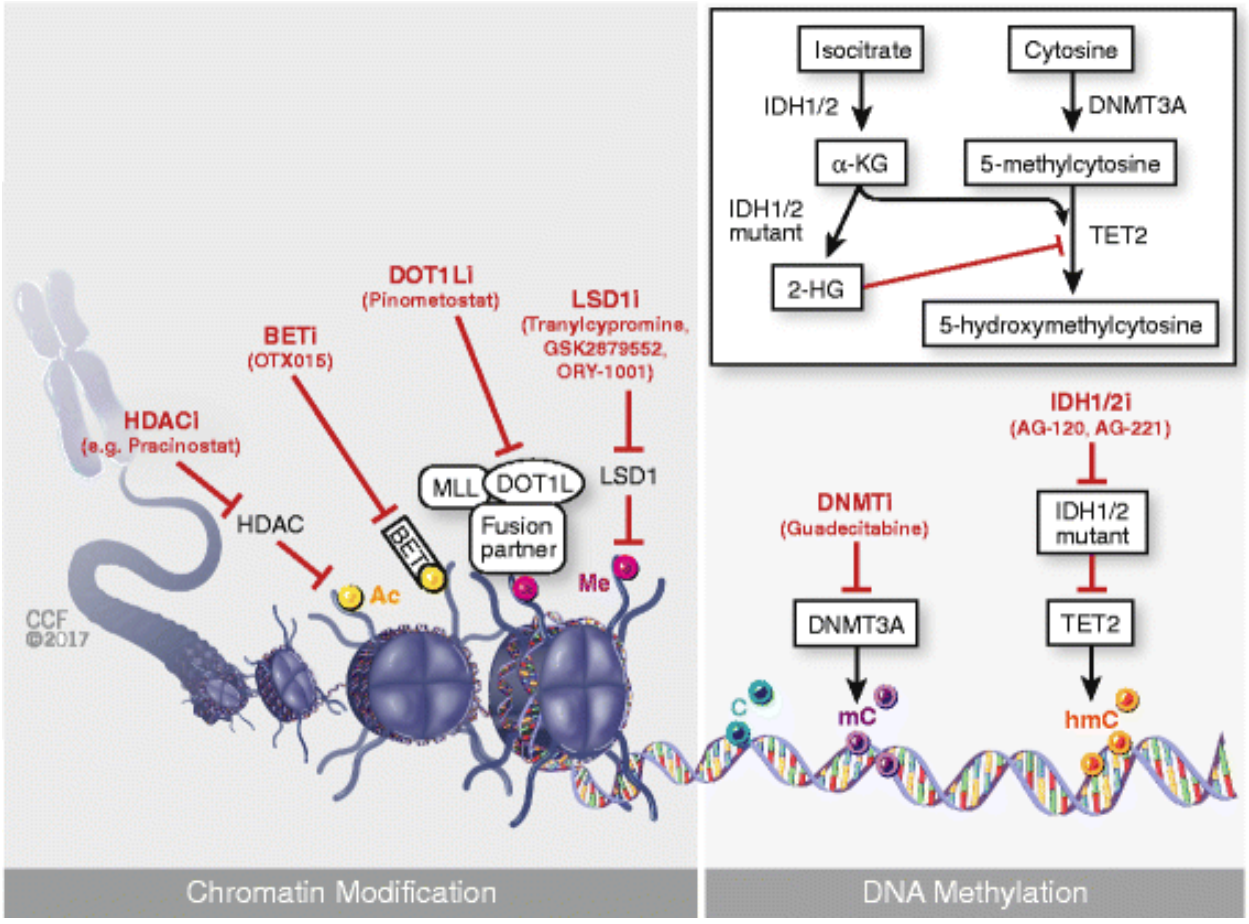


Figure 1.4 | Targeted therapies for AML. Reproduced from Saygin and Carraway, 2017.

- **Targeting PML-RARA with all-*trans* retinoic acid and arsenic trioxide**

PML-RARA is the most common genetic aberration in acute promyelocytic leukemia, and the PML-RARA fusion protein induces leukemogenesis in part by perturbing the transcription of PML's target genes.

All-*trans* retinoic acid (ATRA) and arsenic trioxide (ATO) are highly effective small-molecule inhibitors of PML-RARA that are being used clinically. The binding of ATRA to RARA has a two-fold consequence: switching RARA from a transcriptional co-repressor into a co-activator and inducing RARA degradation. ATRA promotes differentiation of leukemic cells into granulocytes through re-activation of target genes, but this transcriptional activation effect alone is not sufficient to account for its efficacy. Instead, PML-RARA degradation is the more important mechanism for the efficacy of ATRA. Acute promyelocytic leukemia patients bearing the PML-RARA fusion undergo remission upon ATRA treatment but eventually relapse due to residual disease.

ATO also promotes the degradation of PML-RARA by oxidizing Cys212 and Cys213 on PML to yield an internal disulfide bond. This triggers the formation of PML nuclear bodies in which PML-RARA is targeted for proteasomal degradation. ATO induces a stronger and longer-lasting effect than ATRA, giving ~90% long-term remission, even though it does not induce differentiation. (dos Santos et al., 2013; de Thé and Chen, 2010)

- **Targeting DNMT3A with hypomethylating agents**

The mechanism of DNMT3A-catalyzed cytosine methylation involves an activated covalent adduct between DNMT3A and cytosine, which is subsequently cleaved by β-elimination to regenerate the active enzyme. Hypomethylating agents such as azacitidine (cytidine in which the 5-methine is substituted with 5-aza) and decitabine (the 2'-deoxy analog of azacytidine) can be metabolized and integrated into the DNA during replication. The 5-aza group lacks a proton necessary for β-elimination, trapping the enzyme-DNA adduct and thereby deactivating the enzyme.

- **Targeting mutant IDH1 and IDH2**

AGI-120 and AGI-221 are potent inhibitors of IDH1 and IDH2 mutants, respectively. By inhibiting the gain-of-function IDH1/IDH2 mutants, these inhibitors reduce the level of the oncometabolite 2-hydroxyglutarate in the cell, reverse the hypermethylated phenotype, and drive terminal differentiation. (Medeiros et al., 2017)

- **Targeting unmutated proteins**

Some proteins such as BRD4, LSD1, DOT1L, and CDK7 are not mutated in AML but are required for the initiation or maintenance of AML cells. They represent another class of druggable, unmutated therapeutic targets in AML.

BRD4 is a chromatin reader containing two bromodomain modules for binding acetylated histones. The binding of BRD4 to chromatin helps to recruit the basal

transcriptional machinery, transcription factors, the transcriptional elongation complex pTEF-b, and the Mediator complex to activate the expression of target genes. In AML, BRD4 is highly associated with transcription factors that regulate hematopoietic stem and progenitor cells including ERG, FLI1, PU.1, and CEBPA. *BRD4* knockdown induces terminal differentiation and apoptosis specifically in AML cells, suggesting that BRD4 inhibition is a viable therapeutic strategy. Patients treated with the BRD4 inhibitor OTX015 shows good initial response but eventually relapse, possibly due to activation of the Wnt pathway. Surprisingly, although BRD4 occupies virtually all active enhancers, its inhibition with small molecules such as JQ1 or I-BET151 downregulates only a limited set of genes that are associated with large clusters of enhancers called "super-enhancers" (more explanation in **Chapter 2**). Downregulation of the super-enhancer associated with the *MYC* oncogene is especially important for the efficacy of BRD4. (Roe and Vakoc, 2016)

LSD1, which is a H3K4me1/2 and H3K9me1/2 demethylase highly expressed in blood cells, regulates hematopoietic differentiation in the granulomonocytic, erythrocytic, and megakaryocytic lineages. LSD1 is an important effector of MLL fusions, and its recruitment to MLL target loci causes aberrant demethylation that contributes to leukemogenesis. Inhibitors of LSD1 suppress the self-renewal potential of MLL-rearranged AML cells and and induce their differentiation. (Lynch et al., 2012; Mould et al., 2015)

DOT1L, a H3K79 methyltransferase, is another important partner that is required for the initiation and maintenance of leukemia by MLL fusions. DOT1L-mediated methylation increases the expression of hematopoietic stem cell regulators *HOXA9* and

MEIS1. Genetic or pharmacological inhibition of DOT1L impairs the proliferation of MLL-rearranged AML cells, and small-molecule inhibitors are being pursued in the clinic. (Daigle et al., 2011; Deshpande et al., 2013)

CDK7 is the kinase subunit of TFIIH, which is a key component of the RNA pol II pre-initiation complex. CDK7 phosphorylates the C-terminal domain of RNA pol II on Ser5 and Ser7, as well as other proteins, to initiate transcription. As is the case for BRD4, inhibition of CDK7 by the covalent inhibitor THZ1 preferentially downregulates the expression of genes associated with super-enhancers, particularly *MYC*. An improved inhibitor, SY-351, was reported to give complete response in patient-derived xenotransplant models of AML. (Chipumuro et al., 2014; Ren et al., 2015)

Conclusion

AML is a cancer of dysregulated transcription more so than any other cancer. Many of the most common driver mutations in AML are found in genes that encode transcriptional regulators, as we have summarized in this chapter. These mutations are well-evidenced to enable leukemogenesis through aberrant transcription, such as by promoting the expression of the *MYC* oncogene or by blocking myeloid differentiation.

Various therapies that target transcriptional regulators have emerged as efficacious therapeutic strategies in AML. A few drugs have already been used successfully in the clinic such as the IDH2 inhibitor, ATRA, and ATO. Many other compounds are showing promising outcomes

in pre-clinical studies, such as BRD4 and CDK7 inhibitors, and they may soon receive approval for clinical use.

With this view in mind, I will describe in the next chapter how our group uses the natural product cortistatin A to study and identify two transcriptional kinases, CDK8 and CDK19, as novel therapeutic targets in AML.

References

Abdel-Wahab, O., and Levine, R.L. (2013). Mutations in epigenetic modifiers in the pathogenesis and therapy of acute myeloid leukemia. Blood *121*, 3563–3572.

Aerts, S., and Cools, J. (2013). Cancer: Mutations close in on gene regulation. Nature *499*, 35–36.

Alvarez Argote, J., and Dasanu, C.A. (2017). ASXL1 mutations in myeloid neoplasms: pathogenetic considerations, impact on clinical outcomes and survival. Curr. Med. Res. Opin. 1–7.

Arber, D.A., Orazi, A., Hasserjian, R., Thiele, J., Borowitz, M.J., Beau, M.M.L., Bloomfield, C.D., Cazzola, M., and Vardiman, J.W. (2016). The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. Blood *127*, 2391–2405.

de Boer, J., Walf-Vorderwülbecke, V., and Williams, O. (2013). In focus: MLL-rearranged leukemia. Leukemia *27*, 1224–1228.

The Cancer Genome Atlast Network (2013). Genomic and Epigenomic Landscapes of Adult De Novo Acute Myeloid Leukemia. N. Engl. J. Med. *368*, 2059–2074.

Chipumuro, E., Marco, E., Christensen, C.L., Kwiatkowski, N., Zhang, T., Hatheway, C.M., Abraham, B.J., Sharma, B., Yeung, C., Altabef, A., et al. (2014). CDK7 Inhibition Suppresses Super-Enhancer-Linked Oncogenic Transcription in MYCN-Driven Cancer. Cell *159*, 1126–1139.

Corces-Zimmerman, M.R., Hong, W.-J., Weissman, I.L., Medeiros, B.C., and Majeti, R. (2014). Preleukemic mutations in human acute myeloid leukemia affect epigenetic regulators and persist in remission. Proc. Natl. Acad. Sci. *111*, 2548–2553.

Daigle, S.R., Olhava, E.J., Therkelsen, C.A., Majer, C.R., Sneeringer, C.J., Song, J., Johnston, L.D., Scott, M.P., Smith, J.J., Xiao, Y., et al. (2011). Selective Killing of Mixed Lineage Leukemia Cells by a Potent Small-Molecule DOT1L Inhibitor. Cancer Cell *20*, 53–65.

Deshpande, A.J., Chen, L., Fazio, M., Sinha, A.U., Bernt, K.M., Banka, D., Dias, S., Chang, J., Olhava, E.J., Daigle, S.R., et al. (2013). Leukemic transformation by the MLL-AF6 fusion oncogene requires the H3K79 methyltransferase Dot1l. Blood *121*, 2533–2541.

Fasan, A., Haferlach, C., Alpermann, T., Jeromin, S., Grossmann, V., Eder, C., Weissmann, S., Dicker, F., Kohlmann, A., Schindela, S., et al. (2014). The role of different genetic subtypes of CEBPA mutated AML. Leukemia *28*, 794–803.

Fisher, J.B., McNulty, M., Burke, M.J., Crispino, J.D., and Rao, S. (2017). Cohesin Mutations in Myeloid Malignancies. Trends Cancer 3, 282–293.

Guillamot, M., Cimmino, L., and Aifantis, I. (2016). The Impact of DNA Methylation in Hematopoietic Malignancies. Trends Cancer *2*, 70–83.

Kandoth, C., McLellan, M.D., Vandin, F., Ye, K., Niu, B., Lu, C., Xie, M., Zhang, Q., McMichael, J.F., Wyczalkowski, M.A., et al. (2013). Mutational landscape and significance across 12 major cancer types. Nature *502*, 333–339.

Lam, K., and Zhang, D.-E. (2012). RUNX1 and RUNX1-ETO: roles in hematopoiesis and leukemogenesis. Front. Biosci. *17*, 1120–1139.

Larrosa-Garcia, M., and Baer, M.R. (2017). FLT3 Inhibitors in Acute Myeloid Leukemia: Current Status and Future Directions. Mol. Cancer Ther. *16*, 991–1001.

Lawrence, M.S., Stojanov, P., Polak, P., Kryukov, G.V., Cibulskis, K., Sivachenko, A., Carter, S.L., Stewart, C., Mermel, C.H., Roberts, S.A., et al. (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature *499*, 214–218.

Leukemia & Lymphoma Society. Facts 2014-2015.

Lynch, J.T., Harris, W.J., and Somervaille, T.C.P. (2012). LSD1 inhibition: a therapeutic strategy in cancer? Expert Opin. Ther. Targets *16*, 1239–1249.

Mangan, J., and Speck, N. (2011). RUNX1 mutations in clonal myeloid disorders: from conventional cytogenetics to next generation sequencing, a story 40 years in the making. Crit. Rev. Oncog. *16*, 77–91.

Medeiros, B.C., Fathi, A.T., DiNardo, C.D., Pollyea, D.A., Chan, S.M., and Swords, R. (2017). Isocitrate dehydrogenase mutations in myeloid malignancies. Leukemia *31*, 272–281.

Mould, D.P., McGonagle, A.E., Wiseman, D.H., Williams, E.L., and Jordan, A.M. (2015). Reversible Inhibitors of LSD1 as Therapeutic Agents in Acute Myeloid Leukemia: Clinical Significance and Progress to Date: Reversible Inhibitors of LSD1. Med. Res. Rev. *35*, 586–618.

Neff, T., and Armstrong, S.A. (2013). Recent progress toward epigenetic therapies: the example of mixed lineage leukemia. Blood *121*, 4847–4853.

Pabst, T., and Mueller, B.U. (2009). Complexity of CEBPA Dysregulation in Human Acute Myeloid Leukemia. Clin. Cancer Res. *15*, 5303–5307.

Papaemmanuil, E., Gerstung, M., Bullinger, L., Gaidzik, V.I., Paschka, P., Roberts, N.D., Potter, N.E., Heuser, M., Thol, F., Bolli, N., et al. (2016). Genomic Classification and Prognosis in Acute Myeloid Leukemia. N. Engl. J. Med. *374*, 2209–2221.

Patel, J.P., Gönen, M., Figueroa, M.E., Fernandez, H., Sun, Z., Racevskis, J., Van Vlierberghe, P., Dolgalev, I., Thomas, S., Aminova, O., et al. (2012). Prognostic Relevance of Integrated Genetic Profiling in Acute Myeloid Leukemia. N. Engl. J. Med. *366*, 1079–1089.

Pratcorona, M., Abbas, S., Sanders, M.A., Koenders, J.E., Kavelaars, F.G., Erpelinck-Verschueren, C.A.J., Zeilemakers, A., Lowenberg, B., and Valk, P.J.M. (2012). Acquired mutations in ASXL1 in acute myeloid leukemia: prevalence and prognostic value. Haematologica *97*, 388–392.

Ren, Y., Brown, V., Hu, S., Lopez, J., Miljovska, S., Schmidt, D., Bradley, M., Sprott, K., Olson, E., Fritz, C.C., et al. (2015). Targeting Transcriptional Dependency in Acute Myeloid Leukemia (AML) with a Covalent Inhibitor of Transcriptional Kinase CDK7. Blood *126*, 1354–1354.

Roe, J.-S., and Vakoc, C.R. (2016). The Essential Transcriptional Function of BRD4 in Acute Myeloid Leukemia. Cold Spring Harb. Symp. Quant. Biol. *81*, 61–66.

dos Santos, G.A., Kats, L., and Pandolfi, P.P. (2013). Synergy against PML-RARa: targeting transcription, proteolysis, differentiation, and self-renewal in acute promyelocytic leukemia. J. Exp. Med. *210*, 2793–2802.

Saygin, C., and Carraway, H.E. (2017). Emerging therapies for acute myeloid leukemia. J. Hematol. Oncol.J Hematol Oncol *10*.

Schoofs, T., Berdel, W.E., and Müller-Tidow, C. (2014). Origins of aberrant DNA methylation in acute myeloid leukemia. Leukemia *28*, 1–14.

de Thé, H., and Chen, Z. (2010). Acute promyelocytic leukaemia: novel insights into the mechanisms of cure. Nat. Rev. Cancer *10*, 775–783.

de Thé, H., Le Bras, M., and Lallemand-Breitenbach, V. (2012). Acute promyelocytic leukemia, arsenic, and PML bodies. J. Cell Biol. *198*, 11–21.

Vardiman, J.W., Thiele, J., Arber, D.A., Brunning, R.D., Borowitz, M.J., Porwit, A., Harris, N.L., Le Beau, M.M., Hellstrom-Lindberg, E., Tefferi, A., et al. (2009). The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. Blood *114*, 937–951.

Welch, J.S., Ley, T.J., Link, D.C., Miller, C.A., Larson, D.E., Koboldt, D.C., Wartman, L.D., Lamprecht, T.L., Liu, F., Xia, J., et al. (2012). The Origin and Evolution of Mutations in Acute Myeloid Leukemia. Cell *150*, 264–278.

Winters, A.C., and Bernt, K.M. (2017). MLL-Rearranged Leukemias—An Update on Science and Clinical Approaches. Front. Pediatr. *5*.

Chapter 2

CDK8 and CDK19 as Therapeutic Targets in AML

Part of this chapter was adopted with modifications from our manuscript published in Nature 526, *273-276 (2015).*

Contributors

Henry E. Pelish, Brian B. Liau, Ioana I. Nitulescu, Anupong Tangpeerachaikul, Zachary C. Poss, Diogo H. Da Silva, Brittany T. Caruso, Alexander Arefolov, Olugbeminiyi Fadeyi, Amanda L. Christie, Karrie Du, Deepti Banka, Elisabeth V. Schneider, Anja Jestel, Ge Zou, Chong Si, Christopher C. Ebmeier, Roderick T. Bronson, Andrei V. Krivtsov, Andrew G. Myers, Nancy E. Kohl, Andrew L. Kung, Scott A. Armstrong, Madeleine E. Lemieux, Dylan J. Taatjes, and Matthew D. Shair

Introduction

In **Chapter 1**, we introduced AML as a disease of dysregulated transcription and presented literature precedents to support the notion that targeting transcriptional regulators is a therapeutic approach in AML. In this chapter, we present our findings that CDK8 and CDK19 (CDK8/19)—kinases that associate with a general transcriptional co-activator called the Mediator complex—are additional druggable transcriptional regulators. Their inhibition by the natural product cortistatin A (CA) is growth inhibitory to AML cells, both in cell culture and in animal models. We will first provide an overview of Mediator and CDK8/19, with a particular emphasis on their roles in regulating the transcription of genes associated with exceptionally long and active enhancer elements called super-enhancers. We will then describe our approach toward studying this transcriptional regulation process using CA as a CDK8/19 inhibitor. Inhibition of CDK8/19 with CA caused some of the super-enhancer-associated genes to be further activated, indicating that CDK8/19 play a negative regulatory role at these loci. We also found that the upregulation of these genes contributes to suppressing the growth of AML cells.

Mediator complex

Eukaryotic transcription is an intricate process involving hundreds of proteins and many macromolecular complexes. Transcription factors must first recognize and bind to DNA motifs present at promoters and enhancers, and then recruit other transcriptional co-activators to the target loci. Among these co-activators is the Mediator complex, which comprises 26-30 protein subunits and is conserved throughout eukaryotes. Mediator plays an essential role in gene

expression regulation by bridging the communication between gene-specific transcription factors and the basal transcriptional machinery (**Figure 2.1**).

Mediator controls many facets of transcription, from initiation and elongation of RNA synthesis to formation of chromatin loops. The binding of Mediator to target loci facilitates and stabilizes the assembly of the pre-initiation complex, which is composed of RNA pol II and general transcription factors. Phosphorylation of RNA pol II C-terminal tail by TFIIH initiates transcription. The polymerase, however, often pauses ~30-60 nucleotides downstream of the start site. Mediator recruitment of the super-elongation complex, which comprises P-TEFb, ELL, and AF4, relieves this pause, permitting efficient and productive transcription. The ability of Mediator to assist in chromatin looping is also important for long-range interaction between enhancers and promoters for gene activation. (Allen and Taatjes, 2015; Soutourina, 2017; Yin and Wang, 2014)

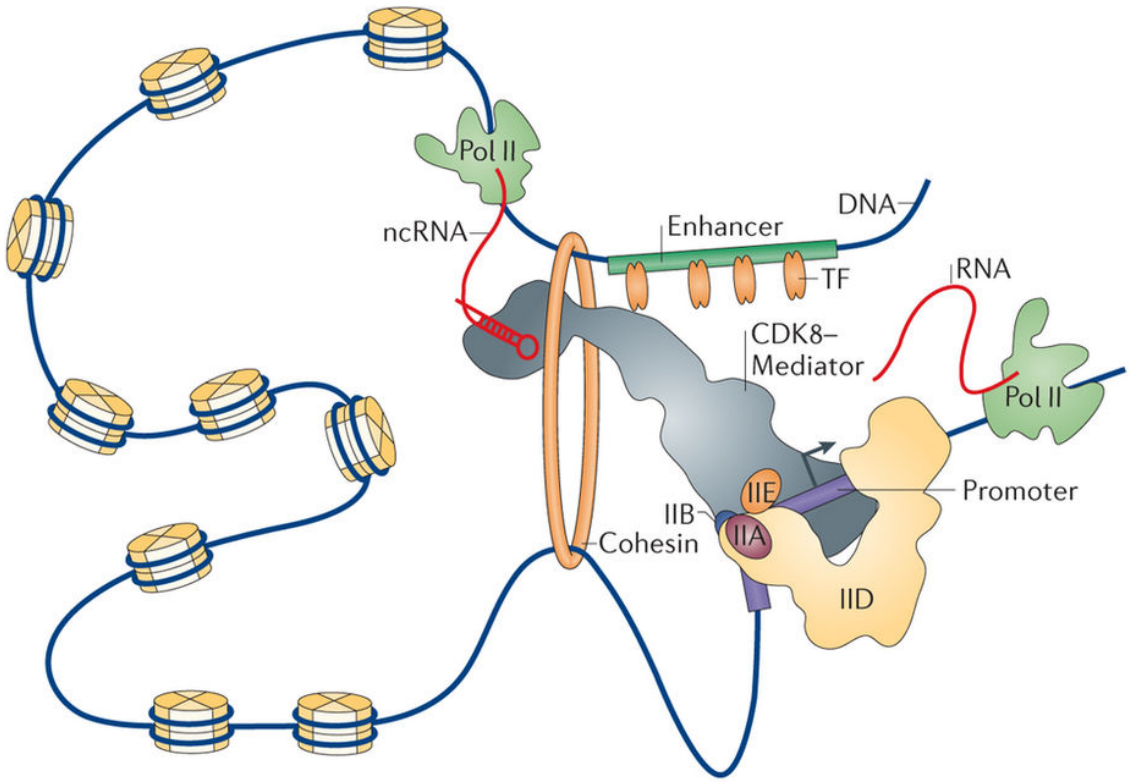


Figure 2.1 | The Mediator complex in transcriptional regulation. Pol II = RNA pol II; TF = transcription factor; IIA, IIB, IID, IIE = TFIIA, TFIIB, TFIID, TFIIE; ncRNA = non-coding RNA. Reproduced from Allen and Taatjes, 2015.

CDK8 module

The CDK8 module, which contains four subunits (CDK8, cyclin C, MED12, and MED13), reversibly associates with Mediator. The CDK8, MED12, and MED13 subunits are interchangeable with their respective paralogs CDK19, MED12L, and MED13L. Each combination of these paralogs may have distinct properties. Exactly how the binding of the CDK8 module alters Mediator functions remains an active area of research, but it likely involves both structural changes and the enzymatic activity of CDK8. For example, the CDK8 module changes the conformation of

Mediator and precludes RNA pol II binding, which may regulate how quickly a second RNA pol II can re-engage with a pre-assembled Mediator at an actively transcribed locus.

CDK8 (cyclin-dependent kinase 8) is a serine/threonine kinase essential for animal development. CDK19 is a paralog of CDK8 that shares 77% sequence identity overall and 94% in the kinase domain. The major sequence difference between CDK8 and CDK19 lies in the C-terminus (**Figure 2.2**). The role of CDK8 in transcription is complicated and likely context-dependent. It has been shown to have a positive or a negative regulatory function at different target loci under different conditions. For example, *CDK8* knockdown prevents recruitment of the super-elongation complex, which is required for efficient transcription of immediate early genes in response to serum stimulation, or of HIF1A-inducible genes in response to hypoxia. CDK8 has also been shown to modulate transcription directly or indirectly by phosphorylating many transcription factors (**Table 2.1** and **Figure 2.3**). These phosphorylation events have been shown to alter the activity and/or the stability of these transcription factors. (Nemet et al., 2014; Philip et al., 2018; Rzymski et al., 2015)

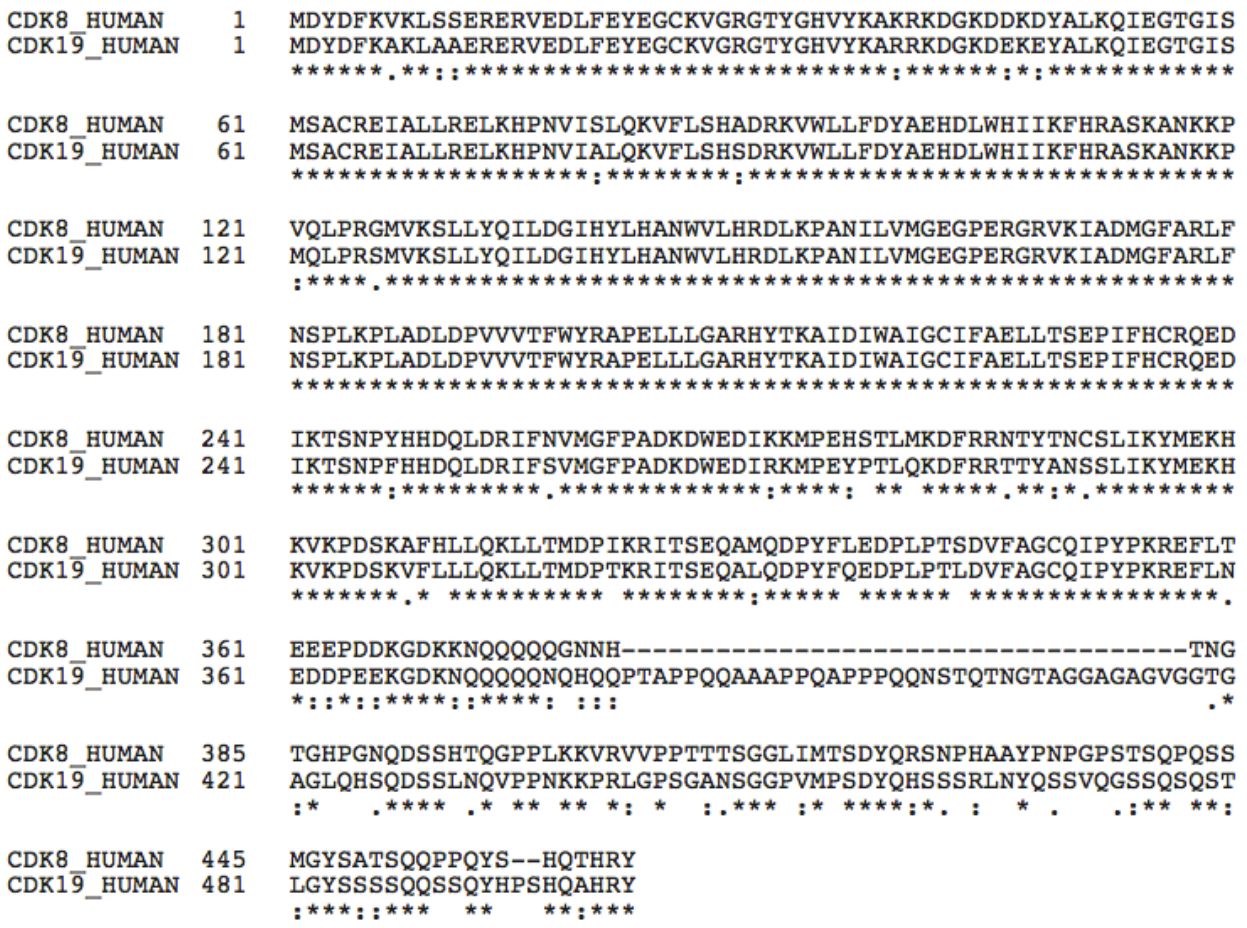


Figure 2.2 | Protein sequence alignment of CDK8 and CDK19. Alignment was performed using Clustal Omega on Uniprot.org.

Table 2.1 | Some of the transcription factors known to be phosphorylated by CDK8

Substrate	Phospho-sites	General effects
E2F1	S375	Represses activity
Notch1	T2512, S2514, S2517	Promotes degradation
Smad1, Smad2, Smad3	S206 (Smad1), T220 (Smad2), T179 (Smad3)	Promotes transcriptional activity and degradation
SREBP-1c	T402	Promotes degradation
STAT1	S727	Positively or negatively affects activity

CDK8 kinase activity modulates transcription

E2F1 is a transcription factor involved in many aspects of cellular processes from proliferation to cell cycle regulation and DNA damage repair. One of the functions of E2F1 is to inhibit Wnt-β-catenin signaling by inducing the expression of ICAT (inhibitor of β-catenin and TCF4). Because activated Wnt-β-catenin signaling is important for colorectal cancer, E2F1 is inhibited in colon cancer cells by binding to pRB or by CDK8-mediated phosphorylation at S375. S375-phosphorylated E2F1 still binds the DNA but loses the ability to activate transcription. A copy-number gain variation of *CDK8* is found in ~60% of colorectal cancer samples. (Firestein et al., 2008; Morris et al., 2008; Wu et al., 2011; Zhao et al., 2013)

Notch1 is both a receptor and a transcriptional co-activator in the Notch signaling pathway, which is highly conserved in all metazoans and is essential for animal development and

tissue homeostasis, especially in the hematopoietic and the neuronal lineages. The Notch1 transmembrane receptor mediates juxtacrine signaling by interacting with Delta-like (Dll) or Jagged ligands on neighboring cells, which stimulate proteolytic cleavage of Notch1 to liberate the Notch1 intracellular domain (NICD1). NICD1 translocates into the nucleus, interacts with transcription factors and co-activators such as RBPJ and MAML1 to activate the expression of target genes. MAML1 also recruits CDK8 to phosphorylate NICD1 at T2512, S2514, and S2517 in the C-terminal PEST domain, promoting NICD1 degradation through the ubiquitin-proteasome system and thereby suppressing Notch activity. We describe the Notch pathway in greater details in **Chapter 3**. (Fryer et al., 2004; Li et al., 2014)

Smads are a family of transcription factors that mediate bone morphogenetic protein (BMP) and transforming growth factor-β (TFG-β) signaling, a pathway essential for animal development and tissue homeostasis. Stimulation of the TFG-β receptors causes Smads to be phosphorylated on the C-terminal tail and translocate into the nucleus to activate gene expression. In addition, Smad1-S206 and Smad3-T179 in the linker region can also be phosphorylated by CDK8 and CDK9, which further activates their transcriptional activity and promotes their degradation. In the case of Smad1, CDK8 phosphorylation recruits YAP, an effector of the Hippo pathway, to modulate target gene expression. (Alarcón et al., 2009)

SREBP-1c is a transcription factor that regulates lipogenesis. Insulin stimulates SREBP-1c to translocate into the nucleus and activate the expression of enzymes involved in fatty acid metabolism. CDK8 phosphorylates SREBP-1c at T402 and inhibits its functions by promoting its degradation through the ubiquitin-proteasome system. (Zhao et al., 2012)

STAT1 is a transcription factor that mediates JAK-STAT signaling, a cytokine-responsive pathway particularly important for the immune system. Stimulation of receptors by cytokines such as IFN-γ activates Janus kinase (JAK) which then phosphorylates STAT1 at Y701, allowing STAT1 dimerization and translocation into the nucleus to activate target gene expression. CDK8 can also phosphorylate STAT1 at S727 in the transactivation domain. This phosphorylation does not have a strictly positive or a strictly negative effect on STAT1 function, but it modulates the expression of some of the target genes in a gene-specific manner. (Bancerek et al., 2013)

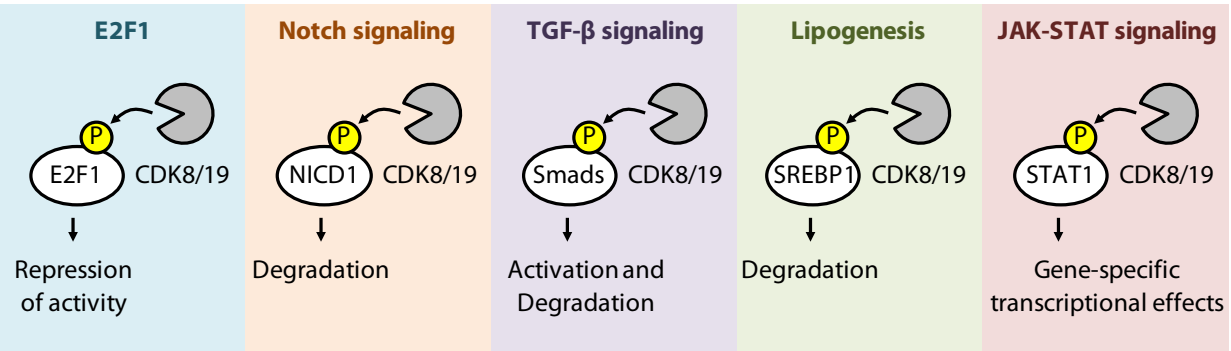


Figure 2.3 | Modulation of transcription factors in various pathways by CDK8.

Super-enhancers

It was recently found that Mediator occupies certain genomic elements, which have come to be called "super-enhancers," at exceptionally high density. Enhancers are cis-regulatory DNA elements that have binding sites for transcription factors, which facilitate transcriptional activation of target genes in an orientation-independent manner, sometimes over a long range through enhancer-promoter looping. Enhancers are enriched in histone markers H3K4me1 and H3K27ac, and an average cell contains ~10,000 – 100,000 enhancers with these histone marks. Super-enhancers are stretches of one or more enhancers bound within ~12 kb from each other that are

occupied by exceptionally large amounts of transcription factors, enhancer histone marks (H3K4me1/2 and H3K27ac), Mediator, and other transcriptional co-activators such as p300, CREBBP, and cohesin. When all enhancer sites are ranked based on MED1 occupancy (MED1 is a principal member of Mediator), the most densely occupied enhancers at the top few percent of the list are super-enhancers. They are often ~10 times longer than a typical enhancer, with a length of ~10 kb versus ~1 kb (**Figure 2.4**).

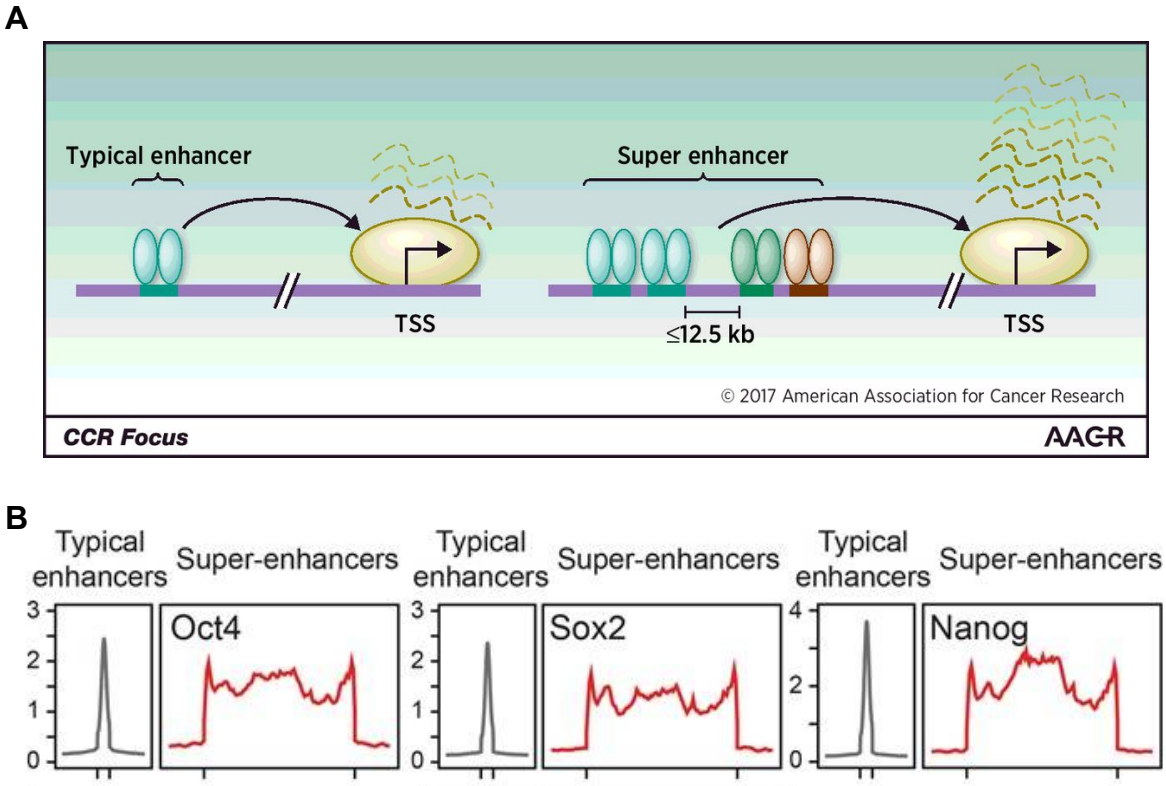


Figure 2.4 | Super-enhancers. (**A**) Schematic depicting the difference between a typical enhancer and a super-enhancer. Image reproduced from Evan et al., 2017. (**B**) The difference in the occupancy of master regulators Oct4, Sox2, and Nanog at typical enhancers versus at super-enhancers. Image reproduced from Hnisz et al., 2013.

The criteria for distinguishing super-enhancers from typical enhancers are based on an algorithmic definition rather than a functional definition. Whether super-enhancers are biologically distinct from typical enhancers remains a subject of intense debate. There are, however, certain properties that can be gleaned from this distinction. Super-enhancers are often located near genes that control cell identity, such as genes that encode master regulators of that cell type or lineage, and they contain binding sites for cell-specific transcription factors. For example, in mouse embryonic stem cells, super-enhancers are found near pluripotency genes *Oct4*, *Sox2*, and *Nanog*, and are enriched in binding sites for embryonic stem cell factors Klf4, Esrrb, and Prdm14. Hence, super-enhancers seem to play an important role in controlling the expression of genes that control cell identity and differentiation.

In cancer, certain oncogenes are found to be driven by aberrantly acquired super-enhancers, such as *MYC* in multiple myeloma, Burkitt's lymphoma, and AML. These super-enhancers may be acquired through various genetic alterations such as sequence mutation, translocation, or amplification. For example, the translocation of the IgH super-enhancer to the *MYC* locus drives *MYC* expression in the multiple myeloma cell line MM.1S, contributing to its over-proliferation phenotype. Super-enhancers therefore seem to have an important role in oncogenesis, and perturbation of super-enhancers may be an efficacious therapeutic strategy. Because super-enhancers are bound by exceptionally large amounts of co-activators such as BRD4 and CDK7, inhibition of these co-activators by JQ1 and THZ1, respectively, has been shown to have a disproportionate effect on super-enhancers compared to typical enhancers (also reviewed in **Chapter 1**). (Niederriter et al., 2015; Pott and Lieb, 2015)

Cortistatin A

Isolation and synthesis

Cortistatin A (CA, **Figure 2.5**) is a natural product first isolated in 2006 from the marine sponge *Corticium simplex*. It inhibits the proliferation of human umbilical vein endothelial cells (HUVECs), a cell line used for studying angiogenesis, with an impressive IC_{50} of 1.8 nM, whereas the IC_{50} for other non-angiogenic cells such as K562 was over 3 μM. This large selective index indicates that CA is not generally cytotoxic, and that it only exerts an anti-proliferative effect in specific contexts. (Aoki et al., 2006)

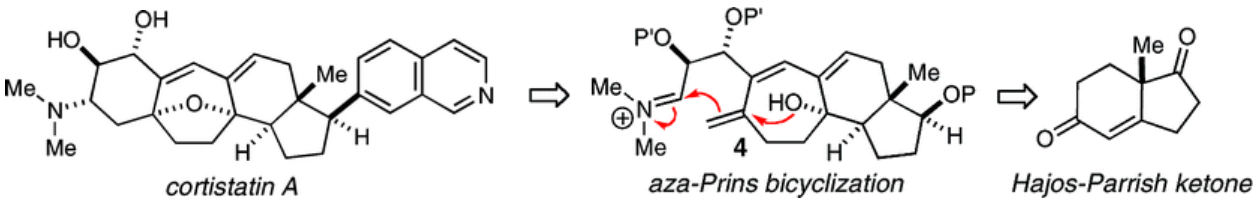


Figure 2.5 | Cortistatin A and our group's retrosynthetic plan. Image adopted from Lee et al., 2008 with modifications.

Total synthesis of CA has been achieved by several groups, and these efforts have enlarged the availability of this relatively rare molecule (22 mg could be isolated from 1.5 kg dried sponge) for biological studies. Our group's approach uses the readily accessible Hajos-Parrish ketone as the starting material and utilizes the aza-prins reaction as a key transannular etherification step (**Figure 2.5**) (Lee et al., 2008). The material synthesized through this route by the Shair group members was used throughout my graduate research.

Target identification

The earliest attempt to identify the targets of CA was performed using KINOMEscan—a commercial high-throughput binding assay comprising 402 kinases (**Figure 2.6**). KINOMEscan measures the ability of a test compound to inhibit the binding of a soluble, purified kinase to kinase ligands immobilized on beads. The kinase is tagged with a DNA barcode to enable quantification by real-time PCR (qPCR). The more potent the inhibitor, the less kinase would remain bound to the beads. CA, at 10 μM, inhibits the binding of 16 out of 402 kinases by more than 65% compared to the control compound (**Table 2.2**). Due to compound availability, dissociation constants (K_d) were only measured for 5 kinases, and 4 binders were confirmed: CDK8, CDK19, ROCK1, and ROCK2 (with K_d of 17, 10, 250, and 220 nM, respectively). Note that the selectivity of CA would likely be greater if used at a lower concentration, as the concentration used for this assay (10 μM) was nearly 10,000 times the IC_{50} for HUVEC growth suppression (1.8 nM) (Cee et al., 2009).

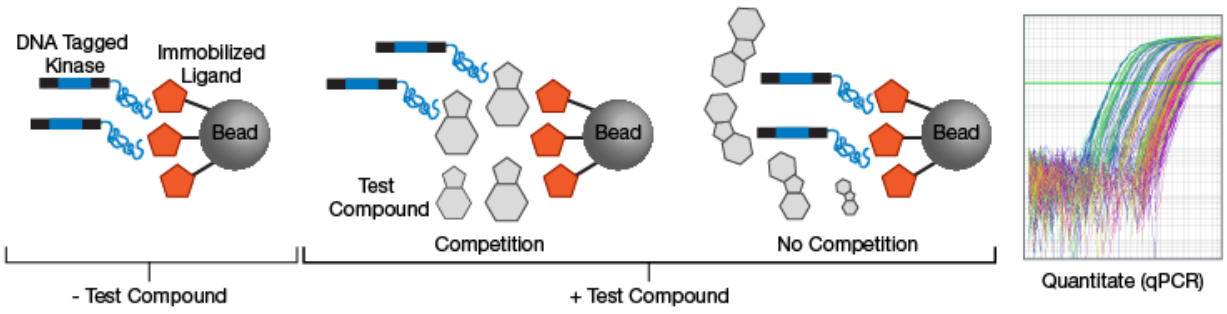


Figure 2.6 | KinomeSCAN assay. Image reproduced from DiscoverX, 2018.

Table 2.2 | Targets of CA based on a KINOMEscan assay (Cee et al., 2009)

Kinase	KINOMEscan inhibition (%)	K_d (nM)
ROCK2	0	220 ± 7
CDK19	0.1	10 ± 2
CDK8	0.95	17 ± 2
LTK	2.9	
ALK	4.4	
PIM2	4.4	
PKACA	8.7	3,500 ± 212
PKACB	13	
MET	18	
PRKG2	21	
RIOK2	21	
ROCK1	21	250 ± 35
CLK4	26	
ROS1	26	
CIT	28	
JNK1	29	

CA as a CDK8 and CDK19 inhibitor in AML

CDK8 and CDK19 (CDK8/19) are part of Mediator, which is heavily loaded at super-enhancers. Knowing that small-molecule inhibition of other super-enhancer-associated co-activators such as BRD4 and CDK7 was efficacious in AML models, we wondered whether CDK8/19 inhibition by CA would also have a similar therapeutic effect. CA also offers a valuable tool for elucidating the roles of CDK8/19 in super-enhancer regulation. The rest of this chapter describes our findings with respect to the activity of CA, the therapeutic potential of CA in AML, the validation of therapeutic targets of CA in AML, and the regulation of super-enhancer transcriptional activity by CDK8/19.

CA inhibits CDK8/19 kinase activity

We determined that CA potently inhibited the ability of the CDK8 module to phosphorylate known targets *in vitro*, including RNA pol II C-terminal domain, CDK8 (autophosphorylation), and STAT1-S727 (Bancerek et al., 2013), with IC_{50} ~ 12 nM (**Figures 2.7A-B**). By contrast, CA did not inhibit other transcriptional cyclin-dependent kinases CDK7 (TFIIH), CDK9 (P-TEFb), CDK12, or CDK13 from phosphorylating the RNA pol II C-terminal domain *in vitro* (**Figure 2.7C**). In cells, CA dose-dependently inhibited the phosphorylation of STAT1-S727 (Bancerek et al., 2013), Smad2-T220, and Smad3-T179 (Alarcón et al., 2009) with IC_{50} < 100 nM (**Figures 2.7D-E**).

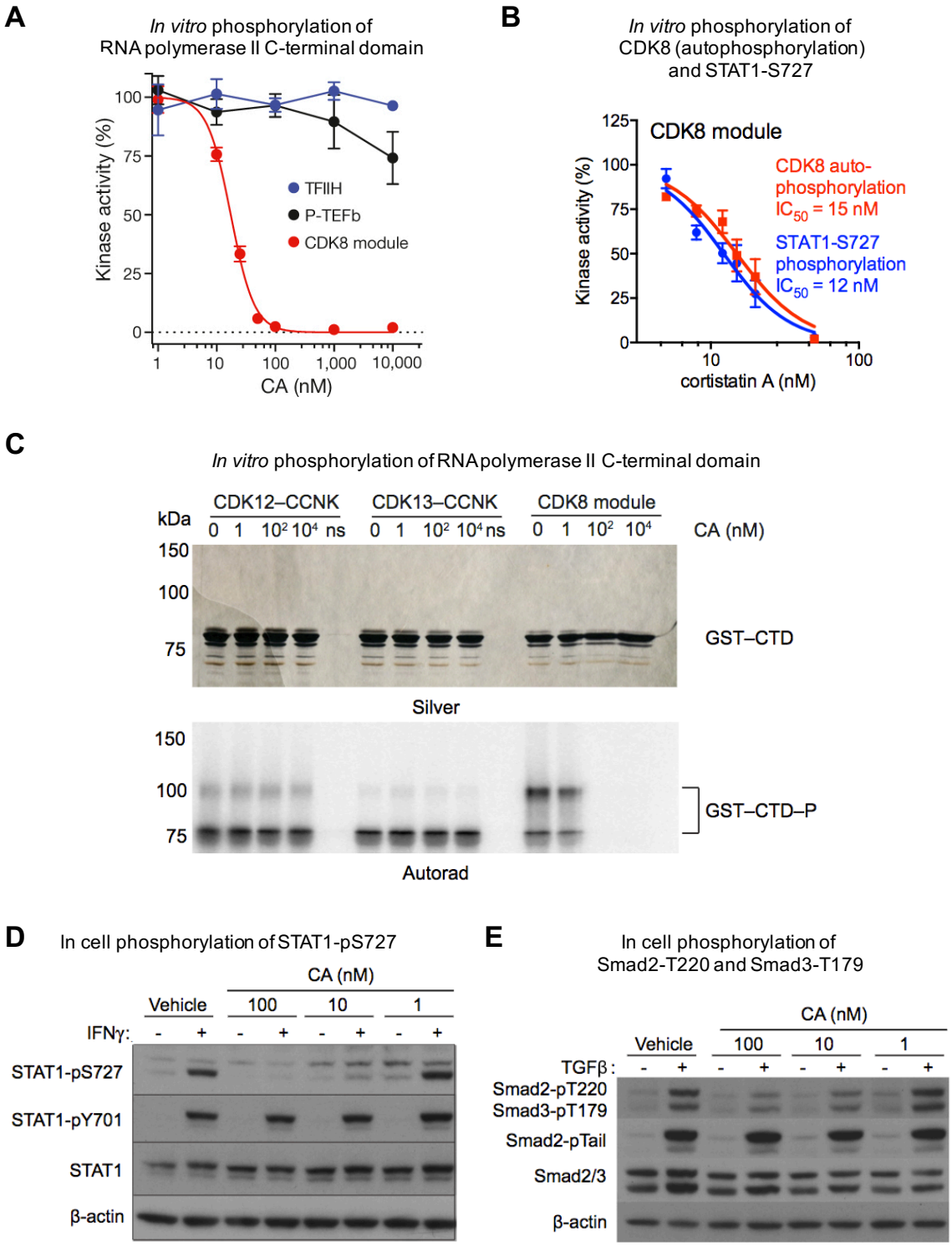


Figure 2.7 | CA as an inhibitor of CDK8 activity in vitro and in cells. (**A**) CA inhibits phosphorylation of the RNA pol II C-terminal domain by the CDK8 module, but not by TFIIH or

P-TEFb. (**B**) CA inhibits *in vitro* phosphorylation of CDK8 (autophosphorylation) and STAT1-S727 by the CDK8 module. (**C**) CA inhibits phosphorylation of the RNA pol II C-terminal domain by the CDK8 module, but not by CDK12 or CDK13. GST = glutathione-S-transferase, a tag for purification. (**D**) CA inhibits in cell phosphorylation of STAT1-S727 upon stimulation with IFN-γ. (**E**) CA inhibits in cell phosphorylation of Smad2-T220 and Smad3-T179 upon stimulation with TGF-β.

We also found, contrary to a previous report (Cee et al., 2009), that CA was highly selective for CDK8/19 over ROCK1, ROCK2, and other kinases using the native kinase capture assay. We pulled down all kinases from cell lysate using desthiobiotin-adenosine triphosphate (ATP) probe in the presence or absence of CA, captured with streptavidin-coated beads, and identified the captured proteins with western blot or mass spectrometry (**Figure 2.8A**). CA blocks the capture of CDK8/19 with IC_{50} < 100 nM, but not CDK9, CDK12, CDK13, ROCK1, or ROCK2 even at the concentration of 2.5 μM; among 220 captured kinases identified by mass spectrometry, only CDK8/19 were blocked by more than 65% by CA at 1 μM, which was ~100 times its IC_{50} for CDK8/19 (**Figures 2.8B-C**). These results strongly suggest that CA displays over 100-fold selectivity for CDK8/19 over other kinases in the cell lysate.

We determined various thermodynamic and kinetic parameters for CA binding to the CDK8-Cyclin C complex *in vitro* using the Proteros reporter displacement method, in which the reporter probe was displaced by various concentrations of CA. CA exhibited high affinity binding (K_d = 195 ± 15.8 pM), slow binding kinetics (dissociation rate constant (k_{off}) = $6.35 \times 10^{-5} \pm 8.15 \times$

10^{-6} s^{-1}), association rate constant (k_{on}) = $3.26 \times 10^5 \pm 1.54 \times 10^4$ s^{-1} M^{-1}) and a long residence time (262 ± 34 minutes).

To understand how CA inhibits CDK8, we obtained a high-resolution (2.4 Å) crystal structure of a CA-CDK8-CCNC ternary complex (**Figure 2.8D**). CA exhibits exquisite shape complementarity with the ATP-binding pocket of CDK8. In particular, the isoquinoline of CA forms N—H and CH—O hydrogen bonds with A100, the C5-C8 ethano bridge and the C13-methyl group of CA occupy deep hydrophobic crevices in the ATP-binding site, and the protonated C3 *N*,*N*-dimethylamine of CA engages in an apparent cation-π interaction with W105 (**Figure 2.8E**).

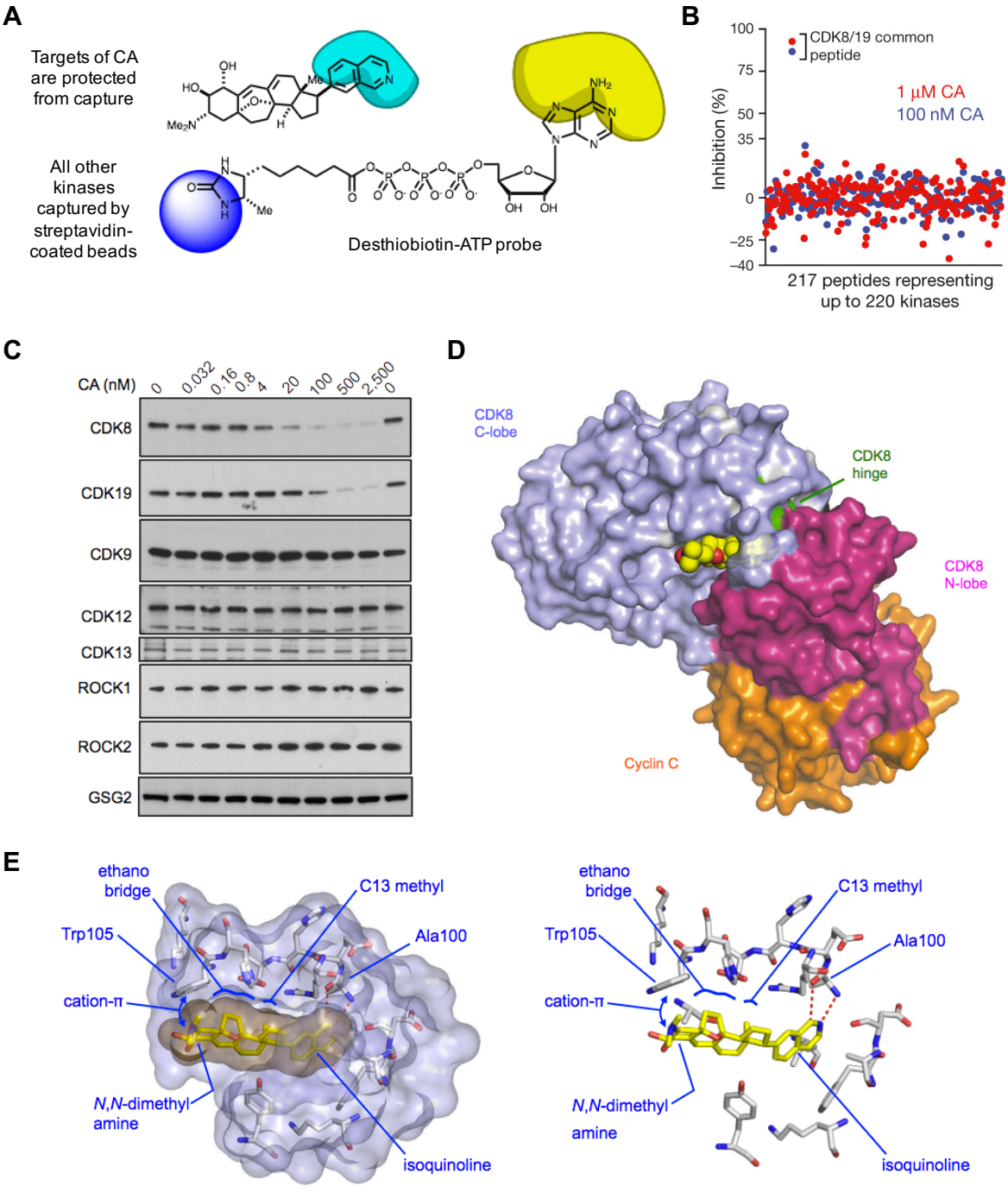


Figure 2.8 | Biochemical characterization of CA selectivity. (**A**) Kinase capture scheme. (**B-C**) CA inhibits the capture of CDK8/19, but not any other kinases, as analyzed by mass spectrometry (**B**) and western blot (**C**). (**D**) X-ray crystal structure of the tertiary CDK8-cyclin C-CA complex. (**E**) Magnified view of the X-ray crystal structure in the space surrounding CA.

CA inhibits the proliferation of many AML cell lines

We investigated the antiproliferative activity of CA and observed that it inhibited the proliferation of several myeloid, mixed-lineage, and megakaryoblastic leukemia cell lines containing diverse oncogenic contributors, including MLL fusions (MOLM-14, MV4;11, and RS4;11), RUNX1-RUNX1T1 (SKNO-1), JAK2-V617F (SET-2 and UKE-1), and BCR-ABL (MEG-01) (**Figure 2.9A**) Although SET-2 and HEL cell lines contain the JAK2-V617F mutation, and MEG-01 and K562 contain the BCR-ABL translocation, megakaryoblastic cell lines SET-2 and MEG-01 cells were sensitive to CA whereas erythroleukemia-derived cell lines HEL and K562 were not, suggesting that cell lineage may be a contributing determinant for CA sensitivity (Garraway and Sellers, 2006). The phenotypic effects of CA were cell-line-dependent.

We confirmed that Mediator kinases mediate the antiproliferative activity of CA by identifying point mutants of CDK8 and CDK19 that would confer resistance to CA. Notably, CDK8 and CDK19 are the only mammalian cyclin-dependent kinases with W (or any aromatic amino acid) at residue 105 (**Figure 2.9B**). The W105M mutant retained the enzymatic activity but specifically desensitized cells to CA, underscoring the importance of the putative cation-π interaction (**Figures 2.9C-D**). Other mutations also desensitized cells to CA, but at varying degrees (**Figure 2.9E**).

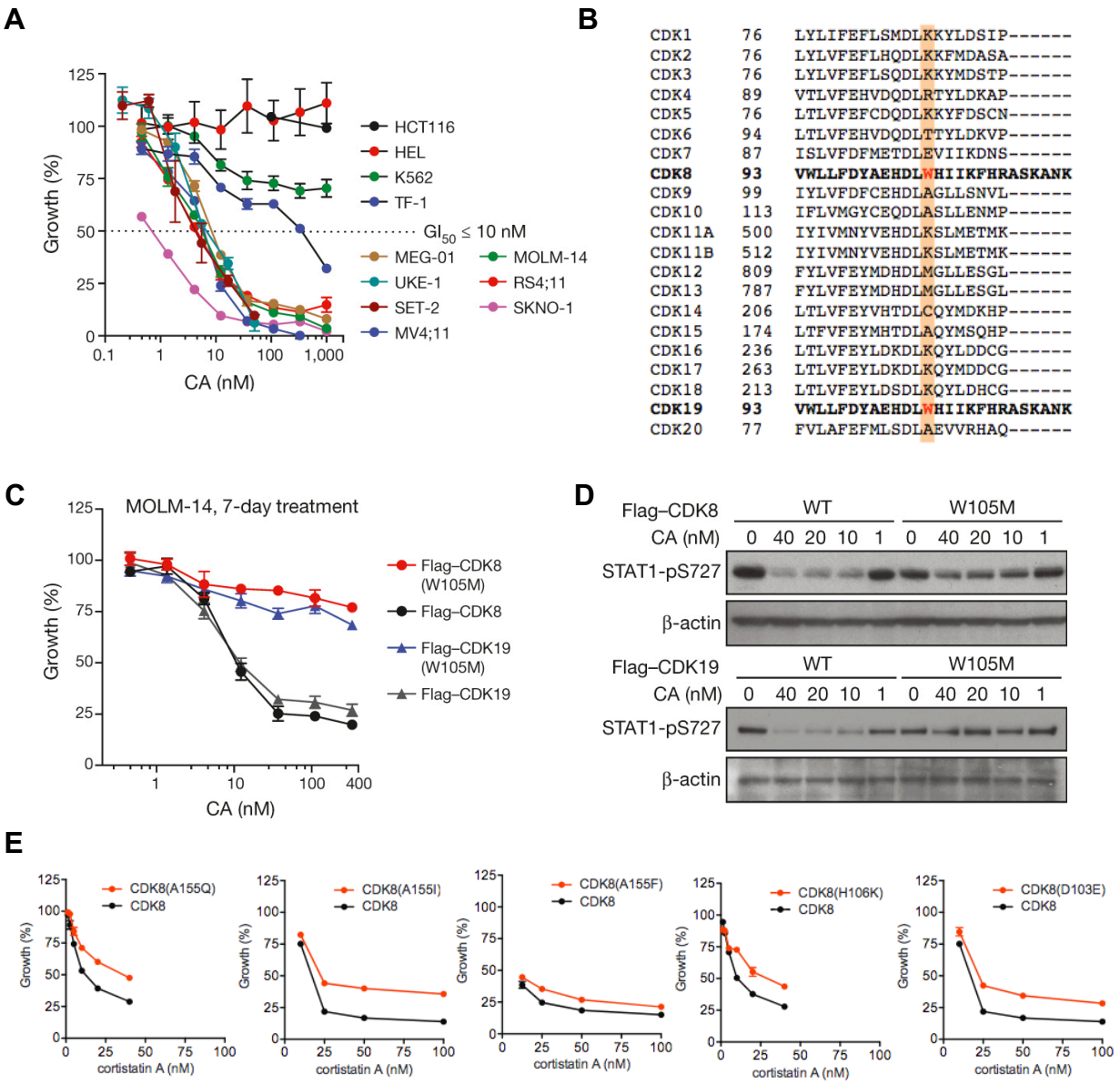


Figure 2.9 | Effects of CA on the proliferation of AML cell lines. (**A**) CA inhibits the growth of many AML cell lines, but is not generally cytotoxic. (**B**) CDK8/19 are the only two CDKs that have an aromatic residue at the position homologous to CDK8-W105. (**C**) CDK8/19-W105M mutation strongly desensitizes cells to the antiproliferative effect of CA. (**D**) CDK8/19-W105M mutation reverses the ability of CA to inhibit STAT1-S727 phosphorylation. (**E**) The effect of other W105 mutations on growth suppression by CA.

CA further upregulates super-enhancer-associated genes

CDK8 associates with cyclin C, MED12, and MED13 to form the CDK8 module that can reversibly associate with the 26-subunit Mediator complex (Allen and Taatjes, 2015). Because super-enhancers are disproportionately loaded with Mediator (Whyte et al., 2013), we examined whether CDK8, as a Mediator-associated kinase, might regulate super-enhancer function.

Using chromatin immunoprecipitation followed by sequencing (ChIP-seq), we first mapped the genome-wide occupancy of CDK8 along with known super-enhancer-associated factors and histone modifications in the AML cell line MOLM-14. Semi-supervised hierarchical clustering revealed that CDK8 most closely associated with MED1, followed by BRD4 and H3K27ac, at putative enhancer elements marked with H3K4me1 (**Figure 2.10**). A fraction of these regions was particularly large and loaded with CDK8, MED1, and BRD4, suggesting that they may represent super-enhancers. Consistent with this notion, most of the CDK8, MED1, BRD4, and H3K27ac ChIP-seq signal was disproportionately located on a small number of super-enhancers identified by each factor separately. Genes associated with these super-enhancers were enriched with Gene Ontology terms pertinent to hematopoiesis, cellular differentiation and transcription, supporting the notion that super-enhancers regulate cellular identity (**Table 2.13**).

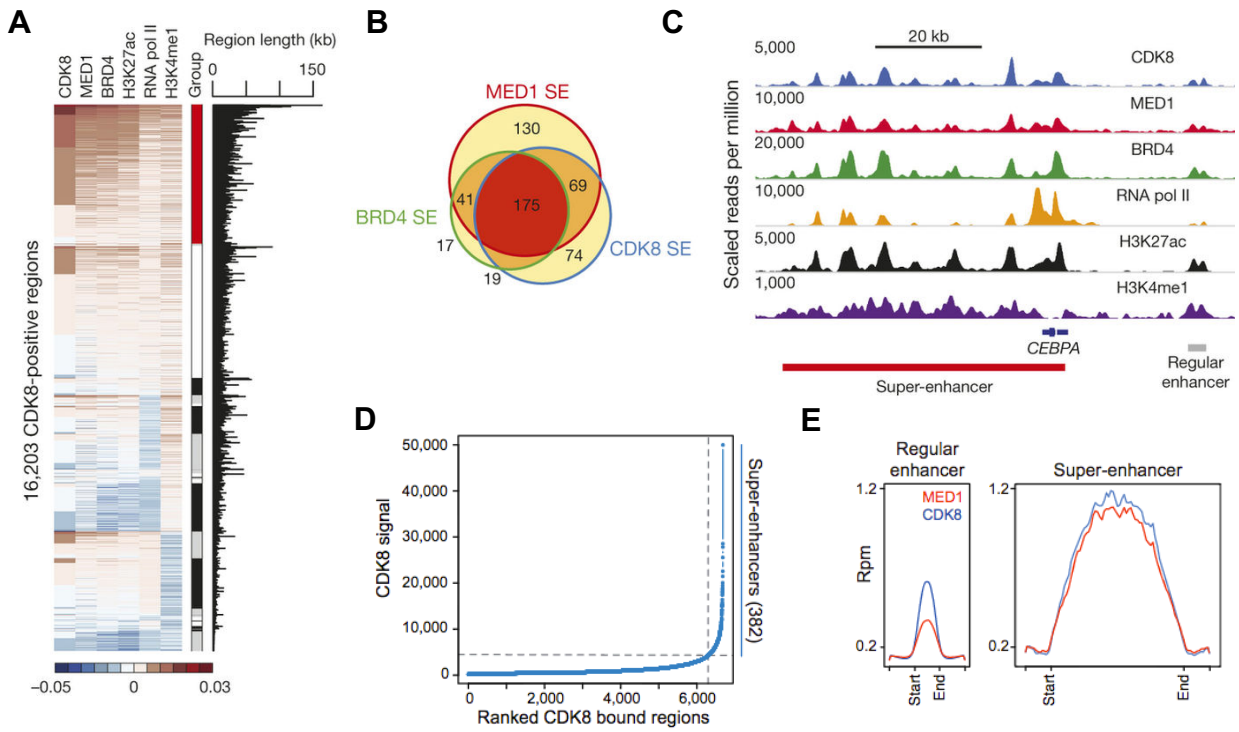


Figure 2.10 | CDK8 occupies super-enhancers in MOLM-14 cells. (**A**) Clustering of total ChIP-seq signal of CDK8, MED1, BRD4, H3K27ac, RNA pol II, and H3K4me1 on CDK8-positive regions. Each respective cluster is ordered by CDK8 signal. The red bar indicates the cluster most highly enriched for the factors listed above. (**B**) Overlap between super-enhancers independently identified by ChIP-seq signal for CDK8, MED1, and BRD4 based on the collapsed superset of regions identified by any one factor. (**C**) ChIP-seq binding profiles at the CEBPA locus. (**D**) Distribution of CDK8 signal with input subtracted across CDK8 bound regions. Regions to the right of inflection point are considered super-enhancers (**E**) ChIP-seq profile plots centred around MED1-defined super-enhancer and regular enhancer regions. Flanking regions are 2.5 kb.

Table 2.3 | Gene Ontology of genes associated with super-enhancers in MOLM-14 cells.

Gene Ontology	Enrichment	*p*-value
Molecular functions		
Transcription factor activity	2.57	10^{-8}
Transcription regulator activity	2.16	10^{-8}
DNA binding	1.82	10^{-7}
Hemopoietic or lymphoid organ development	4.32	10^{-7}
Biological processes		
Immune system development	4.07	10^{-6}
Cell activation	3.91	10^{-6}
Leukocyte activation	4.15	10^{-6}
Lymphocyte activation	4.45	10^{-5}
Hemopoiesis	4.01	10^{-5}

Next, we used CA to investigate whether Mediator kinase activity regulates super-enhancer-associated gene expression in AML cells. Global gene expression profiling in MOLM-14 cells treated with CA revealed that genes upregulated by CA at 3 hours were highly enriched for association with super-enhancers by gene set enrichment analysis (**Figures 2.11A-B**). Genes upregulated (≥1.2-fold) by CA were disproportionately associated with super-enhancers in MOLM-14 cells (49 out of 251, 20%) compared to regular enhancers (173 out of 5,034, 3%, Fisher's exact test, $p < 2.2 \times 10^{-16}$; **Figure 2.11C**). By contrast, of 102 genes downregulated (≥1.2-fold) by

CA, only three were identified as super-enhancer-associated (3 out of 102, 3%). Furthermore, the association between CA-upregulated genes (≥1.2-fold) and super-enhancer-associated genes correlated with CDK8 occupancy (Fisher's exact test, $p = 2.5 \times 10^{-8}$), consistent with the notion that super-enhancers are direct targets of CA treatment in MOLM-14 cells (**Figure 2.11C**). Super-enhancer-associated gene sets ranked among the most significantly enriched compared to all other signatures tested (**Figure 2.11D**).

Because super-enhancer-associated genes are more highly expressed compared to regular enhancer-associated genes, we determined whether genes upregulated by CA had elongating RNA pol II and reduced travelling ratios (ratio of RNA pol II ChIP-seq reads in the proximal promoter versus the gene body). Indeed, CA-upregulated genes exhibited a reduced baseline travelling ratio (2.40-fold, $p < 2.2 \times 10^{-16}$, red versus black curve; **Figure 2.11E**), consistent with CA upregulating active genes, including those associated with super-enhancers. CA treatment further reduced the travelling ratio of these CA-upregulated genes to a level similar to all super-enhancer-associated genes (yellow curve), in agreement with their increased expression after CA treatment (1.48-fold, $p = 7.6 \times 10^{-4}$, blue versus red curve; **Figure 2.11E**). Genes downregulated by CA experienced insignificant changes in the travelling ratio. Global effects of CA on the RNA pol II travelling ratio, RNA pol II C-terminal domain phosphorylation, messenger RNA and total RNA levels were modest or negligible.

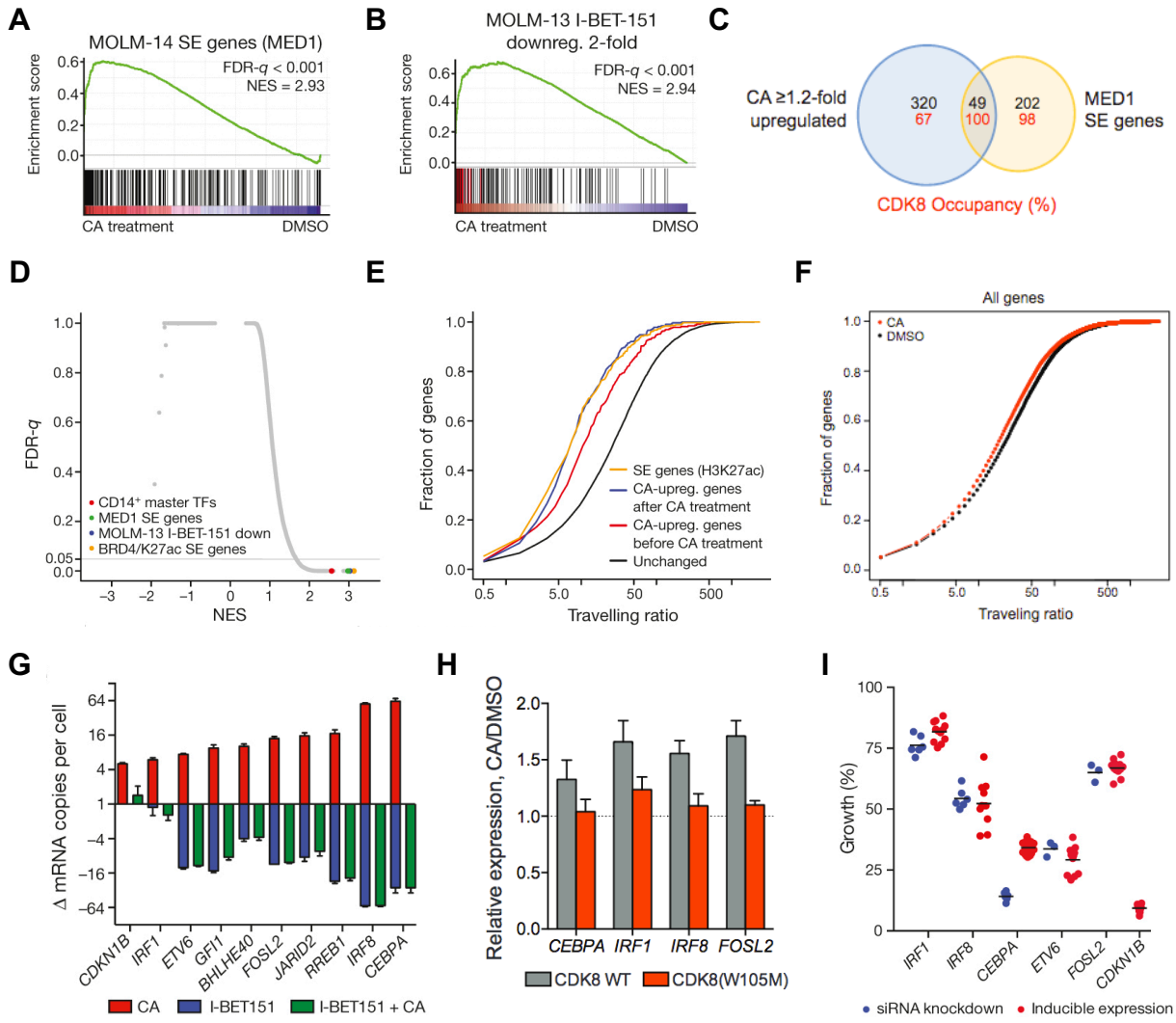


Figure 2.11 | Effects of CA on super-enhancer-associated genes. (**A-B**) GSEA plots show that genes upregulated after 3-hour CA treatment of MOLM-14 cells are significantly enriched in MOLM-14 super-enhancer-associated genes (**A**) and genes downregulated by IBET-151 ≥2-fold in MOLM-13 cells (**B**). (**C**) Venn diagram showing the overlap between super-enhancer-associated genes and genes upregulated ≥1.2-fold after 3-hour CA treatment in MOLM-14 cells. Numbers in red indicate the percentage of CDK8-occupied genes (peak within ±5 kb of the gene). (**D**) Scatterplot of false discovery rate (FDR-q) versus normalized enrichment score (NES) for indicated gene sets evaluated by GSEA (n = 3,867), including C2 of MSigDB. (**E**) Cumulative

distribution plots of RNA pol II travelling ratio. (**F**) Cumulative distribution plot of RNA Pol II travelling ratio (TR) after treatment with CA (25 nM, 6 hours) or vehicle across all genes. (**G**) Change in mRNA copy number per cell of selected super-enhancer-associated genes after 3-hour treatment (red and blue bars) or after 6-hour I-BET151 treatment with CA treatment for the final 3 hours (green bar) (**H**) mRNA levels of indicated genes in MOLM-14 cells expressing Flag–CDK8 (grey) or Flag–CDK8-W105M (red) after 3-hour 25 nM CA treatment. (**I**) Effect of change in expression of selected super-enhancer-associated genes on MOLM-14 cell growth.

Super-enhancer activation is growth inhibitory in AML cells

We then examined whether the upregulation of super-enhancer-associated genes might contribute to the antiproliferative activity of CA. Super-enhancer-associated genes upregulated by CA were enriched in lineage-controlling master transcription factors identified in related $CD14^+$ monocytes, including tumor suppressors *IRF1*, *IRF8*, *CEBPA*, and *ETV6*. Increased expression of these genes individually, as well as super-enhancer-associated genes *FOSL2* and *CDKN1B*, inhibited the proliferation of MOLM-14 cells (**Figure 2.11G**). Furthermore, expression of CA-resistant CDK8-W105M prevented upregulation of super-enhancer-associated genes by CA (**Figure 2.11H**). Therefore, upregulation of super-enhancer-associated genes, through Mediator kinase inhibition, could contribute to the antiproliferative activity of CA.

Growth of several AML cell lines was sensitive to CA and the BRD4 inhibitor I-BET151. The opposing effects of these inhibitors on super-enhancer-associated gene expression, however, suggest that AML cells might depend on a precise dosage of super-enhancer-associated gene

expression. Indeed, MOLM-14 cell growth was inhibited by either reduced or increased expression of the same super-enhancer-associated genes, many of which were upregulated by CA and downregulated by I-BET151 (**Figure 2.11I**). Despite having opposing effects on super-enhancer-associated genes, CA and I-BET151 co-treatment did not normalize transcription of these genes. Instead, I-BET151-induced transcriptional effects dominated, suggesting a dependence on BRD4 for CA-induced transcription (**Figure 2.11G**).

CA does not perturb super-enhancers in insensitive cells

We extended our gene expression, ChIP-seq and super-enhancer analyses to additional cell lines that were sensitive (SET-2 and MV4;11) and insensitive (HCT116 and K562) to CA, and found that only the sensitive cell lines showed statistically significant enrichment of super-enhancer-associated genes among those upregulated by CA. These results support upregulation of super-enhancer-associated genes as contributing to the antiproliferative effects of CA. However, we cannot exclude the contribution of other factors.

CA inhibits *in vivo* models of AML

Finally, we assessed the *in vivo* anti-leukemic activity of CA. We first determined that CA had acceptable pharmacokinetic properties in mice for once-daily intraperitoneal dosing and then measured its efficacy in a disseminated human AML model (Etchin et al., 2013). CA afforded a dose-dependent reduction in disease progression ($p < 0.0001$) and increased survival (29.5-day median extension in survival, $p < 0.0001$; **Figure 2.12A-B**). Efficacious dosing was well-tolerated, with no loss in body weight or deleterious effects in peripheral blood of leukemia-bearing or healthy, immunocompetent (CD-1) mice. In a second AML model using SET-2 cells, CA afforded

a 71% tumor volume reduction, also with no loss in body weight (**Figure 2.12C**). We confirmed that CA inhibited CDK8 in vivo by observing a dose-dependent reduction in STAT1-S727 phosphorylation in natural killer cells, which have CDK8-dependent constitutively phosphorylated STAT1-S727 (**Figure 2.12D**) (Putz et al., 2013).

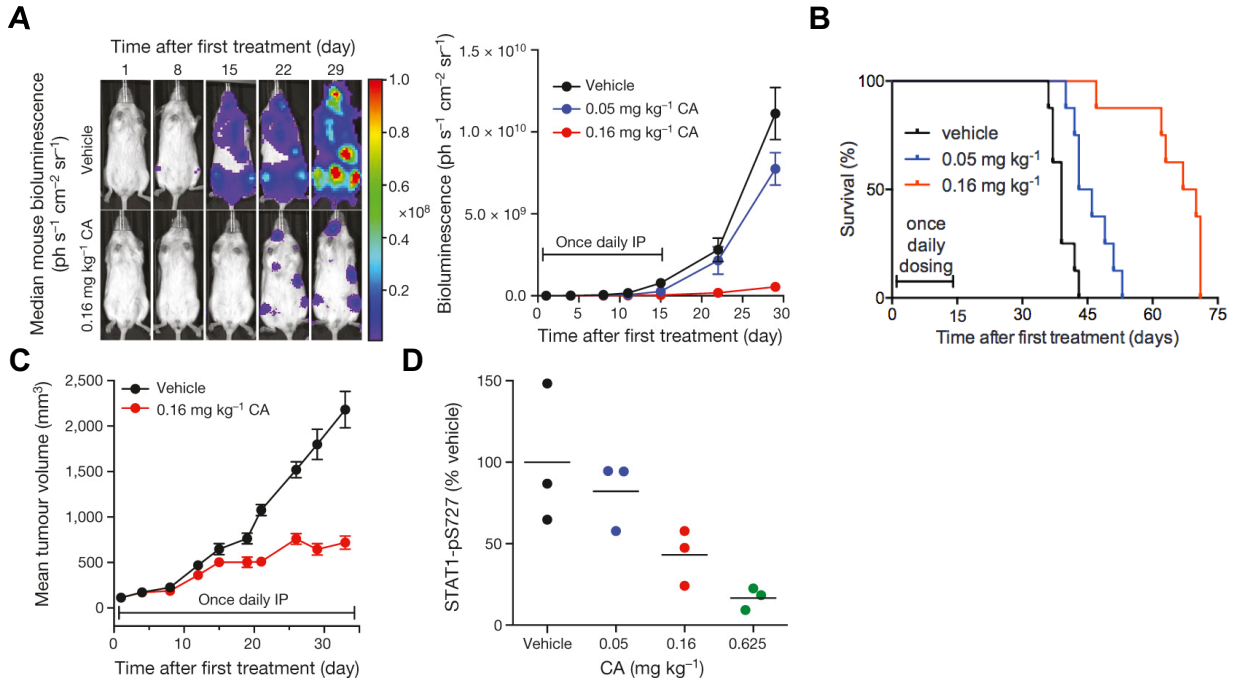


Figure 2.12 | *In vivo* efficacy of CA. (**A**) Bioluminescent images of mice bearing MV4;11 leukemia cells. Mouse with median bioluminescence shown. IP = intraperitoneal. (**B**) Kaplan–Meier survival analysis. (**C**) Mice containing SET-2 AML xenograft tumors and treated as indicated. (**D**) Densitometric analysis of STAT1-pS727 in natural killer cells isolated from the spleen of C57BL/6 mice treated with CA or vehicle. STAT1-pS727 was normalized to actin.

Conclusion

Super-enhancers, which are composed of large clusters of enhancers densely loaded with the Mediator complex, transcription factors, and chromatin regulators, drive high expression of genes implicated in cell identity and disease, such as lineage-controlling transcription factors and oncogenes. BRD4 and CDK7 are positive regulators of SE-mediated transcription. By contrast, negative regulators of super-enhancer-associated genes have not been well described. Here we show that the Mediator-associated kinases CDK8 and CDK19 restrain increased activation of key super-enhancer-associated genes in AML cells. We report that the natural product CA selectively inhibits Mediator kinases, has anti-leukemic activity *in vitro* and *in vivo*, and disproportionately induces upregulation of super-enhancer-associated genes in CA-sensitive AML cell lines but not in CA-insensitive cell lines. In AML cells, CA upregulated super-enhancer-associated genes with tumor suppressor and lineage-controlling functions, including the transcription factors *CEBPA*, *IRF8*, *IRF1*, and *ETV6*. The BRD4 inhibitor I-BET151 downregulated these super-enhancer-associated genes, yet also has anti-leukemic activity. Individually increasing or decreasing the expression of these transcription factors suppressed AML cell growth, providing evidence that leukemia cells are sensitive to the dosage of super-enhancer-associated genes. Our results demonstrate that Mediator kinases can negatively regulate super-enhancer-associated gene expression in specific cell types, and can be pharmacologically targeted as a therapeutic approach to AML. The specificity, potency, favorable pharmacokinetics and long residence time of CA make it a useful *in vitro* and *in vivo* probe of Mediator kinases and a promising lead for development of therapeutics.

Methods

Cell culture

All media was supplemented with 100 U ml^{-1} penicillin and 100 μg ml^{-1} streptomycin. Cell line media: MV4;11, RS4;11, K562, HEL, MOLM-14 and MEG-01 in RPMI-1640, 10% FBS; SET-2 in RPMI-1640, 20% FBS; UKE-1 in RPMI-1640, 10% FBS, 10% horse serum and 1 μM hydrocortisone; SKNO-1 and TF-1 in RPMI-1640, 10% FBS, plus 10 and 2 ng ml^{-1} GM-CSF, respectively; HaCaT in DMEM, 10% FBS; and HCT116 in McCoy's 5A, 10% FBS (proliferation assay) or DMEM, 10% FBS (gene expression study). Sources: HepG2, MV4;11, RS4;11, MEG-01, TF-1, HCT116 and K562 from ATCC; SKNO-1 from DSMZ; HEL, UKE-1 and SET-2 from R. Levine; and HaCaT, MV4;11-mCLP and MOLM-14 from V. Wilson, A. Kung and S. Armstrong, respectively. MOLM-14 cells were authenticated by STR profiling and flow cytometry. All cell lines were routinely tested for mycoplasma.

Reagents

Compounds were stored under argon at −80 °C in 100% DMSO. Vehicle represents 0.1% DMSO unless otherwise specified. Sources: IFN-γ (PHC4031, Life Technologies), TGF-β1 (R&D Systems), paclitaxel (LC Laboratories), I-BET151 (Tocris), PMA (Calbiochem), and doxorubicin and puromycin (Sigma-Aldrich). Immunoblot antibodies: anti-Flag (F1804), anti-actin (A5060) and anti-CDK19 (HPA007053) from Sigma-Aldrich; anti-Smad2/3 (8685), anti-Smad2 pTail (3108), anti-STAT1 (9172), anti-phospho-STAT1 Tyr701 (9170) and anti-phospho-STAT1 Ser727 (9177), anti-CEBPA (2843), anti-ROCK1 (4035), anti-ROCK2 (9029), anti-CDK8 (4101), anti-caspase-3 (9662) anti-PARP (9532) and anti-CDK9 (2316) from Cell Signaling Technology (CST);

anti-phospho-Smad2/3 T220/T179 (600-401-C48) from Rockland; anti-CDK12 (NB100-87012) and anti-CDK13 (NB100-68268) from Novus; and anti-CDK8 (A302-501A) and anti-Haspin (A302-241A) from Bethyl. ChIP antibodies: RNA pol II (Rpb1 N terminus, sc-899X lot B2713) from Santa Cruz; MED1 (A300-793A lot A300-793A-2), BRD4 (A301-985A lot A301-985A50-3), and CDK8 (A302-500A lot A302-500A-1) from Bethyl; and H3K4me3 (ab8580 lot 1308511), H3K27ac (ab4729 lot GR104852-1), and H3K4me1 (ab8895 lot GR61306-1) from Abcam.

Kinase assays

Data were quantified with ImageJ and plotted and fitted with GraphPad Prism 6.0. For STAT1 transactivation domain (TAD), 750 ng of glutathione *S*-transferase (GST)–STAT1 TAD (residues 639–750) was incubated with ~50 ng recombinant CDK8 module at 30 °C for 8 min in kinase buffer (25 mM Tris, pH 8, 2 mM dithiothreitol (DTT), 100 μM cold ATP, 100 mM KCl, 10 mM $MgCl_2$ and 2.5 μCi [γ-^{32}P]ATP (Perkin Elmer) per reaction). The assay included 2.5% DMSO, which did not inhibit kinase activity. 12% SDS–PAGE gels were subsequently silver-stained, exposed for 18 h on a Phosphor Screen and imaged (Typhoon 9400, GE Life Sciences). For pol II CTD, 400 ng of GST–CTD (mouse sequence) was incubated with ~40 ng recombinant CDK8 module, 25 ng TFIIH, or 40 ng P-TEFb at 30 °C for 60 min in kinase buffer. Kinase amounts were chosen to give similar total pol II CTD signal. 9% SDS–PAGE gels were silver stained and exposed as above. *In vitro* Flag–CDK8 kinase assays used ~40 ng kinase and 500 ng GST–CTD. CDK12(714–1063)–CCNK(1–267) and CDK13(694–1039)–CCNK(1–267) were expressed in insect cells and used at ~500 nM per reaction. These regions of CDK12/13 encompass the kinase domains (including the C-terminal extension helix) and the cyclin boxes, and are fully

phosphorylated in the T-loop. For STAT1 or Smad2/3, cells were treated with compound for 1 h followed by IFN-γ or TGF-β1 for 1 h, then washed twice with cold PBS, and lysed (RIPA buffer with inhibitors R0278, P8340, P0044 and P5762; Sigma-Aldrich). Standard immunoblotting followed. All experiments were performed twice.

Protein purification

Buffers for purification and elution of recombinant proteins included 0.25 mM PMSF, 1 mM DTT, 1 mM benzamidine and 1 mM sodium metabisulphite. TFIIH was captured from HeLa nuclear extract using a monoclonal antibody for the p89 subunit immobilized to Protein A Sepharose (GE). Final purification of peptide-eluted TFIIH was performed on a 1 ml HiTrap Heparin HP (GE) resulting in 0.1–0.2 μM TFIIH. P-TEFb was purified as described with a Superdex 200 polishing resulting in ~0.5 μM P-TEFb. Recombinant CDK8 module was purified as described with omission of the glycerol gradient. STAT1 TAD and pol II CTD were expressed as N-terminal GST fusion proteins in *Escherichia coli* BL21-CodonPlus cells to $A_{600\ nm}$ 0.5, then induced with 0.5 mM IPTG for 4 h at 30 °C and batch affinity purified with glutathione Sepharose 4B (GE). Cells were lysed in H/E buffer (50 mM Tris, pH 7.9, 0.5 M NaCl, 0.5 mM EDTA, 10% glycerol and 0.5% NP-40), immobilized on glutathione Sepharose 4B in H/E buffer for 3 h at 4 °C and washed with ~100 column volumes of high-salt buffer (50 mM Tris, pH 7.9, 1 M NaCl, 0.5 mM EDTA, 0.5% NP-40 and 8 mM CHAPS), 0.5 M HEGN (20 mM HEPES, pH 7.6, 0.5 M KCl, 0.1 mM EDTA, 10% glycerol and 0.02% NP-40) and 0.15 M HEGN (20 mM HEPES, pH 7.6, 0.15 M KCl, 0.1 mM EDTA, 10% glycerol and 0.02% NP-40). Fusion proteins were eluted in 2× column volumes of 30 mM reduced l-glutathione in GSH elution buffer (80 mM Tris, pH 7.9, 0.15 M KCl,

0.1 mM EDTA, 10% glycerol and 0.02% NP-40). The GST–pol II-CTD was further purified by Superdex 200 polishing. Flag–CDK8 wild-type and W105M mutants were expressed in MOLM-14 cells, captured using anti-Flag M2 affinity resin (Sigma-Aldrich), and eluted with 1 mg ml^{-1} Flag peptide in 0.15 M HEGN in 1× column volume twice. Flag peptide elutions were stained with SYPRO Ruby to standardize kinase amounts. Purifications contained cyclin C but not MED12 or MED13 (data not shown).

Native kinase capture immunoblot and native kinome-wide profiling

Experiments were performed as previously described. 5×10^8 MOLM-14 cells were washed twice with 10 ml cold PBS and resuspended in 1 ml cold kinase buffer (20 mM HEPES, pH 7.4, 150 mM NaCl, 0.5% Triton X-100, with inhibitors 11697498001, Roche and P5726, Sigma). Cells were lysed by sonication (2 × 10 s pulses with a 30 s break) and centrifuged (16,000*g*, 10 min). The supernatant was desalted through a column (732-2010, Biorad) and the eluted lysate was diluted to 5 mg ml^{-1} with kinase buffer. For each treatment, 475 µl of the lysate was pre-incubated with 10 µl $MnCl_2$ (1 M) and 5 µl compound to the desired concentration at room temperature for 30 min. Uninhibited kinases were captured with 10 µl ActivX desthiobiotin-ATP probe (0.25 mM; 88311, Pierce) at room temperature for 10 min. Samples were mixed with 500 µl urea (8 M; 818710, Millipore) and 50 µl streptavidin agarose (20359, Thermo) for 60 min at room temperature on a nutator. Beads were washed twice with a 1:1 mixture of kinase buffer and 8 M urea, and collected by centrifugation (1,000*g*, 1 min). Proteins were eluted from the beads with 100 µl 2 × LDS sample buffer (NP0007, Life) at 95 °C for 10 min. Samples were analysed by standard immunoblotting and horseradish peroxidase detection. Experiment was performed twice. Native kinome profiling was performed with MOLM-14 cell lysate according to the KiNativ Method by ActivX Biosciences. For

each peptide quantified, the change in mass spectrometry signal for the treated samples relative to the signal for the control samples was expressed as percentage inhibition. The results correspond to one experiment of duplicates for each CA concentration. The percentage changes in mass spectrometry signal reported are statistically significant (Student's *t*-test score <0.04).

Recombinant kinome-wide selectivity profiling and IC_{50} determination

A radiometric protein kinase assay was used (PanQinase activity assay; performed by ProQinase GmbH). IC_{50} determination for CDK8–CCNC (8.3 nM with 1.0 μM ATP and 1.0 μg/50 μl of substrate RBER-IRStide) was performed as duplicate measurements and IC_{50} was calculated using Prism 5.04 with sigmoidal response, top fixed at 100% and bottom at 0% with least-squares fitting.

Binding and kinetics

Measurements listed were made using the Proteros reporter displacement assay. CDK8–CCNC (0.62 nM) was pre-incubated with a reporter probe at a concentration equal to its binding affinity (K_d) in 20 mM MOPS, pH 7.0, 1 mM DTT and 0.01% Tween20 (final reaction volume 10 μl in black polypropylene U bottom plates, Corning 4514). After transfer of serially diluted CA, probe displacement was monitored for 60 min. K_d values were calculated using the Cheng–Prusoff equation from the IC_{50} values obtained from the percentage displacement values at the last time point measured. Association rate constant was calculated from the decay rate of probe displacement. Dissociation rate constant was determined as the product of K_d × association rate constant. Residence time was calculated as $1/k_{off}$. Error was determined by Gaussian error propagation from the IC_{50} error. Experiment was performed once.

Crystallization, data collection and refinement

Human CDK8–CCNC was expressed and purified. Co-crystals at a protein concentration of 11.3 mg ml^{-1} with 1 mM CA were obtained in 20% PEG 3350 and 0.20 M sodium formate at 20 °C and shock-frozen with 25% ethylene glycol as cryoprotectant. Diffraction data were collected at the Swiss Light Source (SLS, Villigen, Switzerland), beamline X06SA with a wavelength of 1.00004 Å at 100 K, and processed using XDS and XSCALE. The structure was solved by molecular replacement, subsequent model building and refinement (including TLS refinement) was performed with COOT and *CCP*4. The R_{free} validation was based on a subset of about 3.4% of the reflections omitted during refinement. Waters were included at stereochemically reasonable sites. Final refinement cycles led to a model with R_{work} value 21.7% and R_{free} value 26.6%. All main-chain angles of non-glycine residues fall into the conformationally most favoured (93.2%), additionally allowed (6.6%) or generously allowed (0.2%) regions of the Ramachandran plot. Graphical figures were prepared using PyMOL.

Cell growth assay

All suspension cells were plated (96-well) in triplicate at 5,000–30,000 cells per well for testing (n = 3). Viable cell number was estimated after 3, 7 and 10 days by counting viable cells from one vehicle well, generating a cell dilution series, transferring 20 μl per well in duplicate to a 384-well plate, and performing a linear regression to CellTiter-Glo (Promega) response (SPECTRAmax M3, Molecular Devices). Cells from all wells were also fourfold diluted in media and transferred in duplicate for CellTiter-Glo measurement. On days 3 and 7, an equal volume for all wells was split-back with fresh media and compound, such that the resulting cell density for the vehicle well matched the initial seeding density. For days 7 and 10, estimated cell number

represents the split-adjusted theoretical cell number. HCT116 were plated (96-well) in triplicate at 250 cells per well. Cells were incubated in the presence of vehicle, 1 μM paclitaxel, or compound. On day 7, CellTiter-Blue (Promega) response was measured and values were normalized to vehicle (100% growth) and paclitaxel (0% growth).

Flow cytometric analysis

Cells were plated (6-well) in triplicate at 150,000 cells per ml for 1-day, 2-day and 3-day time points. For the 6-day time point, cells were plated at 35,000 cells per ml and diluted to 150,000 cells per ml with media and compound on day 4. For cell cycle, cells were washed twice with PBS, fixed with 70% ethanol at 4 °C overnight, washed with PBS, and stained with 50 μg ml^{-1} propidium iodide (eBioscience) for 1 h at 37 °C. For apoptosis, cells were stained using annexin V-FITC (BD Pharmingen) and 7-AAD (Miltenyi Biotec). Samples were acquired on a BD LSR II and analysed using FlowJo v7.6.5. For the SET-2 differentiation assay, cells were cultured in triplicate with 50 nM CA, 50 ng ml^{-1} PMA (positive control), or vehicle for 3 days. Cell pellets were collected at 4 °C, washed three times with cold PBS, and stained with anti-CD61-PE (ab91128) or anti-CD41-PerCP (ab134373). For each experiment, n = 3 biological replicates with two independent experiments and one shown.

Plasmids, mutagenesis, packaging, transduction, selection and siRNA

5′-Flag-tagged CDK8 and CDK19 were cloned from pBabe.puro.CDK8.flag (Addgene 19758) and F-CDK8L (Addgene 24762) into pLVX-EF1alpha-IRES-mCherry and pLVX-EF1alpha-IRES-ZsGreen (Clontech) and transformed into *E. coli* (One Shot Stbl3, Invitrogen).

Point mutations were introduced by whole-plasmid PCR (QuikChange II XL Site-Directed Mutagenesis Kit, Agilent). pLVX lentiviral vectors were co-transfected with psPASx and pMD2.G (Addgene) in 293T cells. After 48 h, viral supernatants were collected and passed through a 0.45-μm filter (Millipore). For transductions, 24-well plates were coated with 500 μl of 20 μg ml^{-1} RetroNectin (Clontech) at 4 °C overnight, blocked with 2% BSA for 30 min, washed with PBS, and 300–500 μl of viral supernatant was added. The plates were centrifuged (2,000*g*, 1.5 h) and then set in an incubator. After 2 h, viral supernatant was removed and 500 μl per well of 200,000 cells per ml was added. After 1–3 days, the cells were expanded and isolated by FACS. Flag–CEBPA (gift from J. Marto), Flag–IRF1 (PlasmID, HMS, HsCD00045286), Flag–IRF8 (PlasmID HMS, HsCD00438293), ETV6–Myc–Flag (Origene, SC118922), CDKN1b–Myc–Flag (Origene, SC117607), and FOSL2–Myc–Flag (Origene, SC110898) were cloned into the Tet-On inducible system pLVX-TRE3G-mCherry or pLVX-TRE3G-ZsGreen (Clontech), transformed into *E. coli* (Stellar Competent Cells, Clontech), packaged into lentiviral vectors and cotransduced with regulator vector pLVX-EF1a-Tet3G. After 1 week of selection with puromycin (1 μg ml^{-1}) and G418 (400 μg ml^{-1}), cells were plated in the presence of 100 ng ml^{-1} doxycycline to assess 7-day growth via Cell-Titer Glo. siRNA against *CEBPA* (Ambion s2888), *IRF1* (Ambion s7501), *ETV6* (Ambion s4867 and s4866), *FOSL2* (Ambion s5345), and *IRF8* (Ambion s7098) or scrambled control (Ambion 4390843) was introduced into cells by electroporation (Amaxa Nucleofector II, Program T-019). After 24 h, cells were plated to assess 3- or 4-day growth via Cell-Titer Glo. Knockdown efficiency was assessed after 24 h by immunoblot or after 48 h by droplet digital PCR (ddPCR).

Gene expression, gene ontology and GSEA

Leukaemia cells were plated (12-well) in triplicate at 500,000–800,000 cells per ml and incubated in the presence of vehicle or CA (25 nM 3 h for K562, MOLM-14 and MV4;11; 10 nM 24 h for MOLM-14; 25 nM 4 h for SET-2, $n = 3$ for each cell line). Cells were then washed twice with cold PBS, and snap frozen. RNA was isolated (RNeasy Plus Microkit, Qiagen or TRIzol, Life Technologies), processed, and, for K562, MOLM-14 and MV4;11, hybridized to the Human U133 Plus 2.0 microarray (Affymetrix). Microarrays were processed with Bioconductor packages affyQCReport for quality control and affy for background correction, summarization, and normalization using rma. Probe sets present in at least 1 sample (based on affy mas5call) and for which the interquartile range was $>\log_2(1.2)$ were retained for further analysis. The limma Bioconductor package was used for differential expression analysis of CA-treated versus DMSO control samples (Benjamini–Hochberg adjusted $P < 0.05$). SET-2 and HCT116 gene expression was measured by RNA-seq. SET-2 RNA-seq libraries were prepared and processed using the Ion Torrent workflow. Reads were aligned in two passes, first with rnaStar (v.2.3.0e) then with BWA (v.0.7.5a) for remaining unmapped reads, both using default parameters. Mapped reads were merged and counted using HTSeq (v.0.5.3p3) with -s yes -m intersection-strict. The Bioconductor package DESeq was used for DE analysis (FDR < 0.05 and twofold change) and normalization. HCT116 cells were grown to approximately 80% confluence and were treated with either 100 nM CA or DMSO for 3 h ($n = 3$). Cells were then washed twice with cold PBS and scraped into TRIzol reagent (Life Technologies). After collecting the RNA, it was further purified using an RNeasy mini kit (Qiagen) with an on-column DNase I digestion. Libraries for Illumina sequencing were generated via the Illumina TruSEQ stranded mRNA prep kit. Samples were run in a single lane on an Illumina HiSEQ 2000 sequencer with a single read flow cell using 1× 50-bp reads and a 6-cycle

index read. Reads were mapped to the hg19 reference genome using Tophat2 v.2.0.6 with custom settings including the setting of -library-type fr-firstrand to appropriately account for the stranded nature of the protocol. HTSeq v.0.6.1 was used to obtain read counts over annotated genes and differentially expressed genes were called by DESeq v.1.10.1 with a padj value of less than 0.01. Counts were normalized for GSEA using the limma voom function. Expression data for the I-BET151 comparison were downloaded from ArrayExpress (https://www.ebi.ac.uk/arrayexpress, accession E-MTAB-774) and processed files used as is. Gene lists were submitted to the DAVID web server (http://david.abcc.ncifcrf.gov) for functional annotation. GSEA version 2.09 was carried out using signal-to-noise on natural values as the metric. Signatures included curated gene sets (C2, v.3) downloaded from the Broad's MSigDB as well as signatures curated from in-house and published data sets.

ChIP-seq

Untreated cells or cells treated with CA (25 nM, 6 h), iBET-151 (500 nM, 6 h) or vehicle were crosslinked for 10 min at room temperature by addition of one-tenth of the volume of formaldehyde solution (11% formaldehyde, 50 mM HEPES, pH 7.4, 100 mM NaCl, 1 mM EDTA and 0.5 mM EGTA) to the media followed by 5 min quenching with 125 mM glycine. For CDK8 and MED1 chromatin immunoprecipitations, cells were instead centrifuged, resuspended in serum-free media, and crosslinked at room temperature by addition of an equal volume of 2% formaldehyde in serum-free media for 10 min followed by quenching with 125 mM glycine for 5 min. Cells were then washed twice with cold PBS and snap frozen. ChIP was performed essentially as previously described[2]. In brief, cells were lysed with lysis buffer 1 (50 mM HEPES, pH 7.4, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40 and 25% Triton X-100) and washed with lysis

buffer 2 (10 mM Tris-HCl, pH 8.0, 200 mM NaCl, 1 mM EDTA and 0.5 mM EGTA). For H3K4me3, H3K27me3, H3K27ac, H3K4me1 and pol II, the nuclei were resuspended in 10 mM Tris-HCl, pH 8.0, 100 mM NaCl, 1 mM EDTA, pH 8.0, 0.5 mM EGTA, 0.1% Na-deoxycholate and 0.2% SDS, sheared for 2 min (Branson S220D sonifier, pulse, 0.7 s on, 1.3 s off, 12–14 W) on wet ice, and then Triton X-100 was added to 1% (v/v). For MED1 and CDK8, the nuclei were resuspended in 50 mM Tris-HCl, pH 7.5, 140 mM NaCl, 1 mM EDTA, 1 mM EGTA, 0.1% SDS and 1% Triton X-100 then sheared for 4 min (pulse, 0.7 s on, 1.3 s off, 10–12 W) on wet ice. Sonicated lysates were cleared and incubated overnight at 4 °C with Protein G magnetic Dynabeads (50 μl) pre-bound with the indicated antibodies (5 μg). Beads were washed with sonication buffer, sonication buffer with 500 mM NaCl, LiCl wash buffer (20 mM Tris-HCl, pH 8.0, 1 mM EDTA, 250 mM LiCl, 0.5% NP-40, 0.5% sodium deoxycholate) and TE. Bound complexes were eluted with 50 mM Tris-HCl, pH 8.0, 10 mM EDTA, 1% SDS at 65 °C and reverse crosslinked at 65 °C. RNA and protein were digested using RNase A and proteinase K, respectively, and DNA was purified using Qiagen MinElute columns. Libraries for Illumina sequencing were prepared using the Illumina TruSeq ChIP Sample Preparation kit with the following exceptions. After end-repair and A-tailing, ChIP DNA or whole-cell extract DNA was ligated to Illumina RNA adaptors with unique indices. Alternatively, libraries were prepared using the KAPA Hyper Prep Kit for Illumina and ligated to unique Bioo Scientific NEXTflex barcode adaptors. Following ligation, libraries were amplified with 16–18 cycles of PCR and were then size-selected using a 2% gel cassette in the Pippin Prep System from Sage Science. For histone modifications and RNA pol II, DNA fragments of size 200–500 bp were captured. For CDK8 and MED1, DNA fragments of size 200–450 bp were captured. Libraries were quantified by qPCR using the KAPA Biosystems Illumina Library

Quantification kit. Libraries with distinct indexes were then combined in equimolar ratios and run together in a lane on the Illumina HiSeq 2500 for 40 bases in single read mode.

ChIP-seq data analysis

ChIP-seq data sets were aligned using Bowtie (v0.12.8) to build version NCBI37/HG19 of the human genome (-n 1 -m 1-best-strata). Duplicate reads were removed using Picard tools (v.1.88). For CDK8, peaks were called with both SPP and MACS v.1.4 using default significance cut-off values. SPP cross-correlation analysis was used for both quality control and to set the strand shift parameter for MACS. Regions of interest identified by both peak callers were retained and merged. Regions overlapping >70% with RepeatMasker regions (downloaded 16 November 2012 from UCSC) were excluded from further analysis. Retained regions were annotated by overlap with RefSeq genes (genomic coordinates downloaded from UCSC refgene table Apr. 26, 2013) using bedtools. Retained regions were assigned to one of the following categories: (1) promoter = transcription start site (TSS) − 500 bp to TSS + 200 bp, (2) body = TSS + 201 bp to TES, (3) proximal enhancer = TSS − 5 kb to TSS − 501 bp, and (4) 3′ untranslated region (UTR) = TES + 1 bp to TES + 5 kb. All other regions were termed 'desert' hits. Any gene satisfying the overlap criteria was included in the corresponding category. Travelling ratios were calculated essentially as described. In brief, mapped read coordinates were first extended 3′ to 200 bases to capture the full fragment coverage. The RefSeq coordinates used for annotation were then used to count extended pol II reads falling in the range of TSS − 30 bp to TSS + 300 bp and those falling in the remainder of the gene body (TSS + 301 to TES). Very short transcripts (<630 bp) were excluded, as were cases with very low counts in both regions. Input reads were subtracted and counts were scaled to reads

per kilobase. Transcripts sharing identical TSS and TES coordinates were represented a single time in the count statistics. ChIP-seq tracks were smoothed by calculating the density per million mapped reads in 300 bp bins at 50 bp intervals and were visualized using Integrative Genomics Viewer. ChIP-seq density maps were generated using ngsplot[51] (v.2.08). Heatmap of semi-supervised clustering of total signal on CDK8 positive regions was carried out as follows: (1) peaks were individually identified for each of the 6 ChIPs using MACS2 at default *P* value cutoff; (2) all peaks were combined and merged into non-redundant regions using mergeBed (-d 0); (3) within each unique region, ChIP reads were counted and matched input reads were subtracted after scaling each to million mapped reads; (4) clusters were grouped by ChIPs represented in a given region into 64 categories in the following order: H3K4me1, H3K27ac, pol II, MED1 and BRD4; (5) each group was ordered by decreasing CDK8 signal per region; and (6) ChIP samples were clustered by Euclidean distance of ChIP signal per region after median centring and normalization. A similar approach was used for BRD4 and CDK8 ChIPs in MOLM-14 cells treated with DMSO or I-BET151. In this case, non-promoter-associated regions in which I-BET151 treatment reduced BRD4 signal>2-fold were ordered by $\log_2$ fold-change.

Irreproducibility discovery rate analysis

Reproducibility of two independent H3K27ac ChIP-seq experiments carried out in cells treated with either DMSO or CA for 3 h was assessed according to the pipeline developed for the ENCODE project (https://sites.google.com/site/anshulkundaje/projects/idr). Irreproducibility discovery rate (IDR) was determined as recommended on peaks called by SPP at FDR < 0.5. At this threshold, SPP reported between 180,000 and 300,000 peaks, depending on the exact combination of sample and input, most of which are expected to be noise. Under both treatment

conditions, the number of high-confidence peaks (IDR threshold <0.01 for true replicates and pseudo-replicate self-consistency tests and <0.0025 for pseudo-replicate pooled-consistency analysis) identified based on signal value in the replicates and pseudo-replicates was within the recommended twofold range, indicating good reproducibility. The number of peaks with IDR < 0.01 in the true replicates was used to make the final selection of distinct, non-chrM pooled replicate peaks. Regions within 200 nucleotides of each other were merged to generate the final peaks list. The same approach was used to determine reproducible peaks in two independent BRD4 and CDK8 ChIP experiments in MOLM-14 cells treated with DMSO or I-BET151.

Identification of super-enhancers

MED1 signal was measured in active enhancers (that is, regions enriched in both H3K4me1 and H3K27ac) after extending MED1 ChIP-seq reads 100 bases in a strand-aware fashion. Enhancer regions were sorted based on their MED1 signal and the inflection point of the curve determined. Enhancers with MED1 signal above the inflection point were retained as SEs. In a separate approach, using only the MED1 ChIP-seq data and the ROSE software from the Young laboratory, we found >80% agreement with our previous assignment of MED1 SEs. ROSE was used thereafter to identify SEs using BRD4, H3K27ac (± CA, 3 h), and CDK8 ChIP-seq on peaks called by MACS 1.4. For K562 and HCT116, H3K27ac ChIP samples and their matched inputs were downloaded from the ENCODE project repository at UCSC. For HCT116, CDK8 ChIP-seq data and matched input was downloaded from GSE38258. SE-associated genes were assigned to the nearest expressed transcript, based on H3K27ac signal in a 500-nucleotide window centred on the TSS. Normalized signal for each enhancer, x, is thus (x − minimum)/(maximum − minimum).

Each ChIP-seq experiment yielded different numbers of enhancer regions so we mapped each experiment's enhancer ranks to [0,1] by calculating (rank – 1)/(maximum rank – 1).

RNA levels, ddPCR and qRT–PCR

Total RNA was isolated from 500,000 MOLM-14 cells (RNeasy Plus Mini Kit, Qiagen) and quantified by Nanodrop. mRNA was subsequently isolated (Dynabeads mRNA Purification Kit, Life Technologies) and quantified by Nanodrop. For ddPCR, total RNA was reverse-transcribed into cDNA (High Capacity cDNA Reverse Transcriptase Kit, Applied Biosystems) and used (ddPCR Supermix for Probes, no dUTP, Bio-Rad 186-3024) with TaqMan FAM probes for genes of interest and *ACTB* (VIC) as the reference gene. Droplets were generated in the QX200 Droplet Generator, thermocycled, and read on the QX200 Droplet Reader. Total RNA per cell was measured by isolating total RNA from 10^6 cells using the mirVana miRNA Isolation Kit (Life Technologies) and quantifying by Nanodrop. The difference in copy numbers of specific mRNAs before and after treatment was determined relative to copies of ACTB mRNA per cell. Probes used (Life Technologies): *CEBPA* (Hs00269972_s1), *ETV6* (Hs00231101_m1), *IRF1* (Hs00971960_m1), *IRF8* (Hs00175238_m1), *RREB1* (Hs01002873_m1), *CDKN1B* (Hs01597588_m1), *GFI1* (Hs00382207_m1), *JARID2* (Hs01004460_m1), *BHLHE40* (Hs01041212_m1), and *ACTB* (4325788). qRT–PCR for checking siRNA knockdown was performed with iTaq Universal Probes Supermix (Bio-Rad), $n = 3$, or by ddPCR.

***In vivo* studies**

Studies were performed at Charles River Laboratories (CRL) and Dana Farber Cancer Institute (DFCI) where indicated and approved by Harvard University and each institution's

respective animal care and use committee. For pharmacokinetic studies, serial blood samples from 7-week-old male CD-1 mice (n = 3 per time point) were collected (no blinding) into K_2EDTA tubes, centrifuged, transferred into 96-well plates (matrix tubes), stored at −20 °C, and analysed by liquid chromatography–tandem mass spectrometry (LC–MS/MS) (*in vivo* studies performed at CRL). Study size was determined by the need for three blood samples per time point with three blood samples collected per mouse. The MV4;11 xenograft model were performed as previously described (*in vivo* studies performed at DFCI) Two-million MV4;11-mCLP cells were injected into the tail vein of 7-week-old female non-obese diabetic–severe combined immunodeficient (NOD–SCID) *Il2rg*$^{-/-}$ (NSG) mice (The Jackson Laboratory) and tumour burden was assessed by bioluminescence imaging (BLI) using an IVIS Spectrum system (Caliper Life Sciences). Seven days after injection, leukaemia establishment was documented by BLI and mice were assigned to groups to achieve a similar mean BLI and treated intraperitoneally with vehicle (20% hydroxypropyl-β-cyclodextrin) or CA once daily for 15 days. After 30 days, blood counts were obtained (Hemavet 950 F, Drew Scientific) and spleen, femur and peripheral blood cells were collected and analysed by flow cytometry (LSR Fortessa, BD Biosciences) from three mice per group. The mice and a portion of the spleen were preserved in bouins after body cavities were opened and visceral organs exposed. Samples from all organs were then dissected and placed in nine cassettes per mouse. Tissues were paraffin embedded, sectioned at 6 μm and stained with haematoxylin and eosin. Survival was measured as the time from therapy initiation until moribund state. We selected 11 mice per group to match previous survival analysis in the model (n = 8) and to have 3 additional mice per group for disease burden comparison. Blinding was only done for histopathology analysis. For the SET-2 xenograft model (*in vivo* studies performed at CRL), 8–12-week-old female SCID Beige mice (Charles River) were injected subcutaneously in the flank with 10^7 SET-2 cells in

50% matrigel (0.2 ml per mouse). When tumours reached an average size of 80–120 mm^3, mice were assigned to groups to achieve a similar mean tumour size and treatment commenced without blinding. Tumour volumes were measured using calipers and calculated as ($width^2$ × length)/2. Percentage tumour growth inhibition was calculated as mean volumes of (vehicle – treatment)/(vehicle – initial) × 100. We selected 10 mice per group to safeguard against the IACUC requirement to stop dosing a group if >10% mortality occurs. For safety testing (*in vivo* studies performed at DFCI), 8-week-old female CD-1 mice were treated once daily without blinding for 15 days and weighed daily. Two hours after the last dose, blood counts were obtained and blood chemistry was analysed. Three mice per group were selected as a minimum for comparison. For STAT1-pS727 inhibition, 6–10-week-old female C57BL/6 mice were treated once daily for 2 days (*in vivo* studies performed at CRL, not blinded). One hour after the second dose, natural killer cells were isolated by dissociation of spleenocytes from isolated spleens, lysis of erythrocytes, and isolation of $DX5^+$ cells (MiniMACS CD49b, Miltenyi Biotec) and analysed by immunoblot and densitometry (ImageJ, STAT1-pS727 level normalized to β-actin). We selected three mice per group as a minimum for comparison. Statistical analyses were performed using GraphPad Prism 6.0. For *P* value determinations, two-way or one-way ANOVA was used with Dunnett's multiple comparison testing and *P*-value adjustment. Dotted purple lines were from the Mouse Phenome Database 22903 (The Jackson Laboratory). No statistical methods were used to predetermine sample size, and experiments were not randomized.

Data deposits

The atomic coordinates of CDK8–CCNC in complex with cortistatin A have been deposited in the Protein Data Bank (PDB) with accession number 4CRL. MIAME-compliant

microarray data as well as aligned and raw ChIP-seq data were deposited to the Gene Expression Omnibus (GEO) with accession GSE65161.

References

Alarcón, C., Zaromytidou, A.-I., Xi, Q., Gao, S., Yu, J., Fujisawa, S., Barlas, A., Miller, A.N., Manova-Todorova, K., Macias, M.J., et al. (2009). Nuclear CDKs Drive Smad Transcriptional Activation and Turnover in BMP and TGF-β Pathways. Cell *139*, 757–769.

Allen, B.L., and Taatjes, D.J. (2015). The Mediator complex: a central integrator of transcription. Nat. Rev. Mol. Cell Biol. *16*, 155.

Aoki, S., Watanabe, Y., Sanagawa, M., Setiawan, A., Kotoku, N., and Kobayashi, M. (2006). Cortistatins A, B, C, and D, Anti-angiogenic Steroidal Alkaloids, from the Marine Sponge Corticium simplex. J. Am. Chem. Soc. *128*, 3148–3149.

Bancerek, J., Poss, Z.C., Steinparzer, I., Sedlyarov, V., Pfaffenwimmer, T., Mikulic, I., Dölken, L., Strobl, B., Müller, M., Taatjes, D.J., et al. (2013). CDK8 Kinase Phosphorylates Transcription Factor STAT1 to Selectively Regulate the Interferon Response. Immunity *38*, 250–262.

Cee, V.J., Chen, D.Y.-K., Lee, M.R., and Nicolaou, K.C. (2009). Cortistatin A is a High-Affinity Ligand of Protein Kinases ROCK, CDK8, and CDK11. Angew. Chem. Int. Ed. *48*, 8952–8957.

DiscoverX (2018). KINOMEscan Assay Platform.

Etchin, J., Sun, Q., Kentsis, A., Farmer, A., Zhang, Z.C., Sanda, T., Mansour, M.R., Barcelo, C., McCauley, D., Kauffman, M., et al. (2013). Antileukemic activity of nuclear export inhibitors that spare normal hematopoietic cells. Leukemia *27*, 66–74.

Evan, G.I., Hah, N., Littlewood, T.D., Sodir, N.M., Campos, T., Downes, M., and Evans, R.M. (2017). Re-engineering the Pancreas Tumor Microenvironment: A "Regenerative Program" Hacked. Clin. Cancer Res. *23*, 1647–1655.

Firestein, R., Bass, A.J., Kim, S.Y., Dunn, I.F., Silver, S.J., Guney, I., Freed, E., Ligon, A.H., Vena, N., Ogino, S., et al. (2008). CDK8 is a colorectal cancer oncogene that regulates β-catenin activity. Nature *455*, 547–551.

Fryer, C.J., White, J.B., and Jones, K.A. (2004). Mastermind Recruits CycC:CDK8 to Phosphorylate the Notch ICD and Coordinate Activation with Turnover. Mol. Cell *16*, 509–520.

Garraway, L.A., and Sellers, W.R. (2006). Lineage dependency and lineage-survival oncogenes in human cancer. Nat. Rev. Cancer *6*, 593–602.

Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-André, V., Sigova, A.A., Hoke, H.A., and Young, R.A. (2013). Super-Enhancers in the Control of Cell Identity and Disease. Cell *155*, 934–947.

Lee, H.M., Nieto-Oberhuber, C., and Shair, M.D. (2008). Enantioselective Synthesis of (+)-Cortistatin A, a Potent and Selective Inhibitor of Endothelial Cell Proliferation. J. Am. Chem. Soc. *130*, 16864–16866.

Li, N., Fassl, A., Chick, J., Inuzuka, H., Li, X., Mansour, M.R., Liu, L., Wang, H., King, B., Shaik, S., et al. (2014). Cyclin C is a haploinsufficient tumour suppressor. Nat. Cell Biol. *16*, 1080.

Morris, E.J., Ji, J.-Y., Yang, F., Di Stefano, L., Herr, A., Moon, N.-S., Kwon, E.-J., Haigis, K.M., Näär, A.M., and Dyson, N.J. (2008). E2F1 represses β-catenin transcription and is antagonized by both pRB and CDK8. Nature *455*, 552–556.

Nemet, J., Jelicic, B., Rubelj, I., and Sopta, M. (2014). The two faces of Cdk8, a positive/negative regulator of transcription. Biochimie *97*, 22–27.

Niederriter, A., Varshney, A., Parker, S., and Martin, D. (2015). Super Enhancers in Cancers, Complex Disease, and Developmental Disorders. Genes *6*, 1183–1200.

Philip, S., Kumarasiri, M., Teo, T., Yu, M., and Wang, S. (2018). Cyclin-Dependent Kinase 8: A New Hope in Targeted Cancer Therapy?: Miniperspective. J. Med. Chem.

Pott, S., and Lieb, J.D. (2015). What are super-enhancers? Nat. Genet. *47*, 8–12.

Putz, E.M., Gotthardt, D., Hoermann, G., Csiszar, A., Wirth, S., Berger, A., Straka, E., Rigler, D., Wallner, B., Jamieson, A.M., et al. (2013). CDK8-Mediated STAT1-S727 Phosphorylation Restrains NK Cell Cytotoxicity and Tumor Surveillance. Cell Rep. *4*, 437–444.

Rzymski, T., Mikula, M., Wiklik, K., and Brzózka, K. (2015). CDK8 kinase—An emerging target in targeted cancer therapy. Biochim. Biophys. Acta BBA - Proteins Proteomics *1854*, 1617–1629.

Soutourina, J. (2017). Transcription regulation by the Mediator complex. Nat. Rev. Mol. Cell Biol. (published online ahead of print)

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes. Cell *153*, 307–319.

Wu, Z., Zheng, S., Li, Z., Tan, J., and Yu, Q. (2011). E2F1 suppresses Wnt/β-catenin activity through transactivation of β-catenin interacting protein ICAT. Oncogene *30*, 3979–3984.

Yin, J. -w., and Wang, G. (2014). The Mediator complex: a master coordinator of transcription and cell lineage development. Development *141*, 977–987.

Zhao, J., Ramos, R., and Demma, M. (2013). CDK8 regulates E2F1 transcriptional activity through S375 phosphorylation. Oncogene *32*, 3520.

Zhao, X., Feng, D., Wang, Q., Abdulla, A., Xie, X.-J., Zhou, J., Sun, Y., Yang, E.S., Liu, L.-P., Vaitheesvaran, B., et al. (2012). Regulation of lipogenesis by cyclin-dependent kinase 8–mediated control of SREBP-1. J. Clin. Invest. *122*, 2417–2427.

Chapter 3

Notch Signaling Mediates the Mechanism of CA in AML

Part of this chapter was adopted with modifications from our manuscript, which is currently in the review process.

Contributors

Anupong Tangpeerachaikul, John G. Doench, Hanna Tukachinsky, Henry E. Pelish, and Matthew D. Shair

Introduction

In **Chapter 2**, we presented our findings that CDK8/19 inhibition suppresses the growth of AML cells and that this growth suppression is due, at least in part, to upregulated expression of some super-enhancer-associated genes. The exact biochemical and cell biological changes that lead from CDK8/19 inhibition to AML growth suppression, however, remained unelucidated. This chapter describes our approach to elucidating the full mechanism of growth suppression by CA in an AML cell line using an unbiased genomic screen based on the CRISPR-Cas9 technology. We found that CA activates Notch signaling, which is normally silenced in this cell line through CDK8/19-mediated phosphorylation and degradation of its secondary messengers Notch intracellular domains (NICDs). Our findings suggest that Notch is a general growth suppressor in AML, that AML cells must repress Notch signaling in order to survive, and that one way to activate Notch signaling in some AML cells is to prevent the degradation of NICDs by inhibiting CDK8/19. The insights into the mechanism of CA may prove to be valuable in predicting patient populations that would respond to CDK8/19 inhibitor therapy through the Notch axis. This chapter will describe the logic of functional genomics, the tools used in genetic screens, and how we use functional genomics to discover Notch as a mediator of CA's antiproliferative activity in AML.

Functional genomics

Forward genetics is a powerful approach for identifying genes involved in a particular phenotype or cellular process. In a typical forward genetic screen, researchers first introduce a genome-wide perturbation to gene function to generate a large pool of mutants. They then pick

individual mutants that exhibit a phenotype of interest, and then identify the genetic perturbation that gives rise to that phenotype.

Radiation and chemical mutagens such as ethyl methanesulfonate have been historically used to induce mutations genome-wide, but these mutations occur essentially at random, making their identification laborious and difficult (**Figure 3.1**). In addition, these mutagens may not be suitable for studying diploid organisms because they generate random mutations that are unlikely to fall in the same gene on both chromosomes. Heterozygous perturbation often is not sufficient to abolish gene function in a diploid organism. (Agrotis and Ketteler, 2015; Hartenian and Doench, 2015; Shalem et al., 2015)

cDNA, RNAi, and gene trapping

More modern technologies have enabled genome-wide perturbation of genes, both loss-of-function and gain-of-function, in manner that is easier to trace and study. These perturbations are generally delivered by viral vectors, which are integrated into the host genome and can be identified conveniently by sequencing.

Gain-of-function mutants can be obtained by expressing individual genes in the genome using a complementary DNA (cDNA) overexpression library. A cDNA library contains genes that have been reversed transcribed from mature mRNAs and hence contain no introns. These genes are driven by a constitutive promoter, and their delivery into cells causes the corresponding proteins to be overexpressed. The downsides of a cDNA expression library are: first, some genes

are too large to be cloned or delivered efficiently by virus; second, the cDNA sequences may not encompass all isoforms that can be expressed by a gene (**Figure 3.1**).

Conversely, loss-of-function mutants can be obtained using an RNA interference (RNAi) library or using gene trapping. RNAi is a naturally occurring gene suppression process in which target mRNAs are inactivated using various kinds of oligonucleotides. Small interfering RNAs (siRNAs) and short hairpin RNAs (shRNAs) have been developed for initiating interference of target mRNAs in functional genomics studies. Both siRNAs and shRNAs are short (< 100 nucleotides), can be delivered into the cell using virus, and can be expressed using an RNA polymerase III promoter. The expressed siRNA or shRNA then complexes with proteins such as Argonaute to form the RNA-induced silencing complex (RISC), which is directed to the target mRNA by virtue of base complementarity. RISC either sequesters the mRNA or induces its degradation to suppress gene function. RNAi has been used extensively in functional genomics, but it suffers from off-target activity as well as incomplete gene suppression (**Figure 3.1**).

Gene trapping integrates a promoterless reporter gene randomly throughout the genome by ways of virus or transposable elements. By chance, these traps insert into genes and result in gene deactivation. Gene trapping in a diploid organism, however, also suffers from the problem of heterozygosity, as it is unlikely that the trap would integrate into the same gene on both chromosomes (**Figure 3.1**).

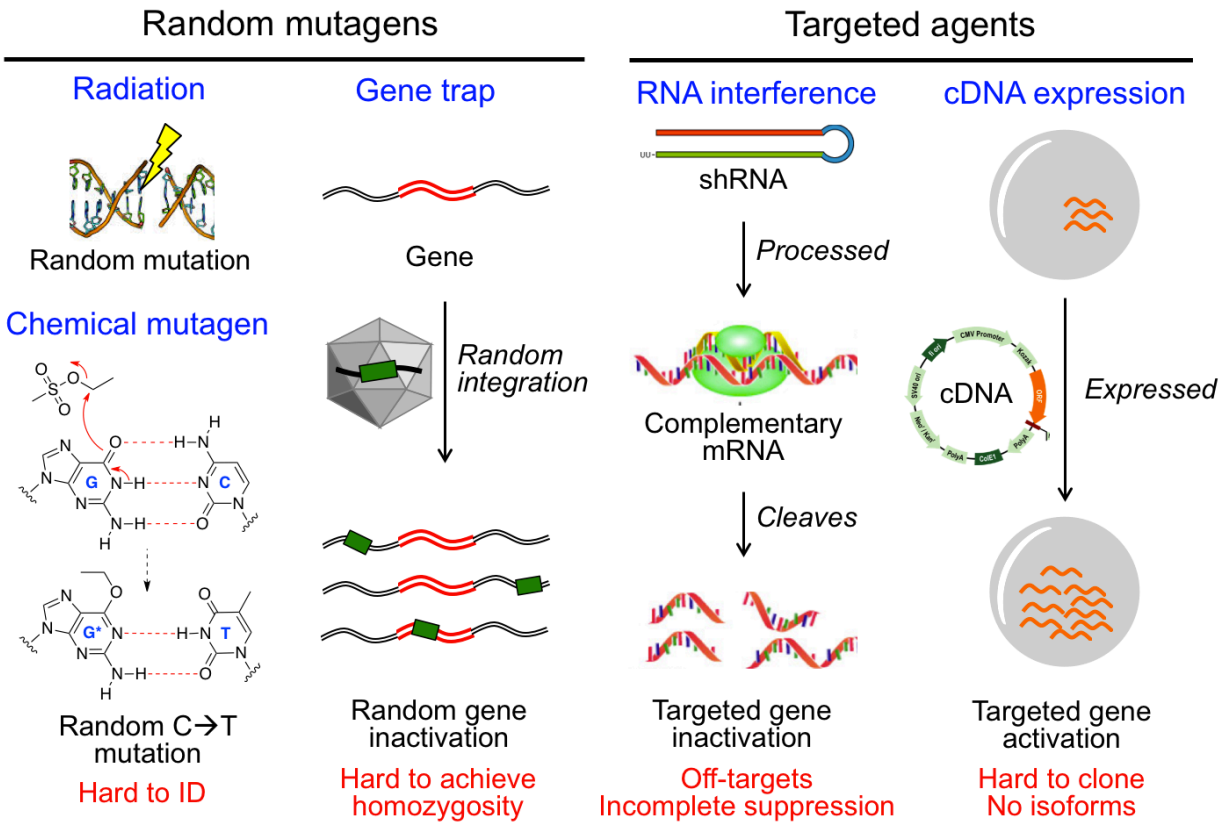


Figure 3.1 | Chemical and biological tools for functional genomics.

Programmable nucleases

An alternative methodology for inducing loss-of-function perturbation in target genes is to use programmable nuclease systems, which include Zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated protein 9 (Cas9).

These nuclease systems can be programmed to cleave DNA at specific sites in the genome and create a double-strand break. The double-strand break is quickly repaired through non-homologous end-joining (NHEJ), which is an error-prone process that introduces random

insertions or deletions (indels). If the double-strand break resides in a protein-coding sequence, the resulting indels often shift the reading frame and create a premature stop codon, causing the protein to be truncated and degraded through nonsense-mediated decay. Because of their ability to promote error-prone NHEJ at specific loci, programmable nucleases have been widely used to generate genetic knockouts. Programmable nucleases can also be used to make precise edits such as point mutations or gene insertions if the double-strand break is repaired through the homology-directed repair (HDR) pathway, but HDR occurs at a much lower frequency than NHEJ. Nonetheless, HDR of double-strand breaks is ~100-1,000 times more efficient than homologous recombination without double-strand breaks, making HDR an extremely useful tool for making precise genomic changes compared to traditional homologous recombination strategies.

ZFNs are protein fusions containing zinc-finger domains for binding target DNA and the FokI nuclease domain for generating a double-strand break. A ZFN contains ~3-6 zinc-finger domains, each of which recognizes a 3-base-pair sequence, giving the total recognition motif of 9-18 base pairs per ZFN. Because FokI must dimerize to become active, ZFNs work in pairs—two ZFN units recognize the sequences flanking both ends of the target site, allowing FokI to dimerize and cleave the target site (**Figure 3.2A**). The overall length of the sequences recognized by a ZFN pair is therefore ~36 base pairs. The specificity of ZFNs can be tuned to match the sequences at the target site by mutating the zinc-finger domains, but the lack of definitive rules governing the interaction between zinc-fingers and the DNA makes it difficult to design ZFNs with high on-target and low off-target activity.

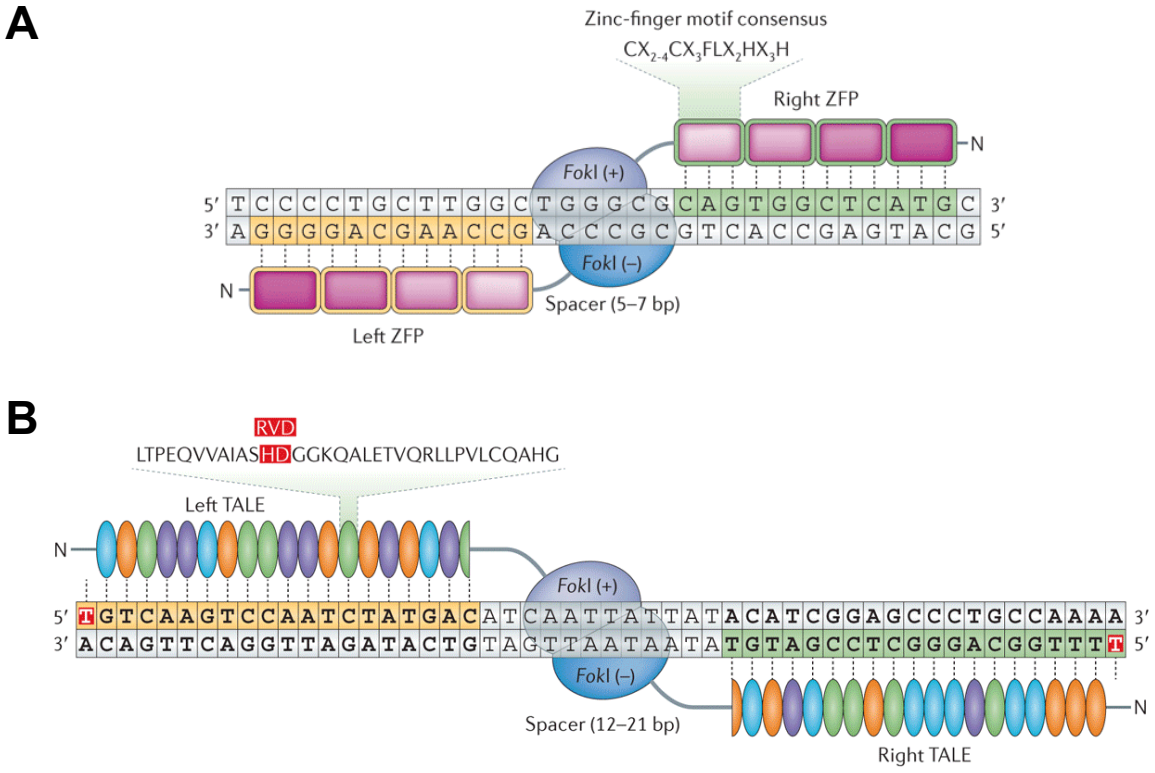


Figure 3.2 | ZFNs and TALENs. Image reproduced from Kim and Kim, 2014.

Similar to ZFNs, TALENs are protein fusions containing transcription activator-like effectors (TALEs) as DNA-binding domains and the FokI nuclease domain. TALENs also work in pairs, with each unit recognizing ~18 base pairs for an overall recognition of ~36 base pairs at the target site (**Figure 3.2B**). Each TALE domain is ~35 amino acids long and binds a single base on the target DNA. Its sequence specificity can be tuned by changing the amino acids at positions 12 and 13. TALENs are easier to construct than ZFNs and can be used to target virtually any site in the genome.

In the context of forward genetics screens, however, it remains difficult to design, clone, and deliver plasmid libraries that would express ZFNs or TALENs to make genome-wide

perturbations. CRISPR-Cas9 surpasses ZFNs and TALENs in this regard. (Hsu et al., 2014; Kim and Kim, 2014).

CRISPR-Cas9

CRISPR-Cas9 was originally discovered as part of the type II bacterial adaptive immune system, and has now been re-engineered for use in eukaryotic gene editing. In bacteria with this type of immunity, small fragments (~20 base pairs) of the invading phage's DNA can be integrated as protospacers in between short palindromic repeats at the bacterial CRISPR locus. This concatenation of multiple repeat-protospacer units is transcribed as a single RNA called pre-CRISPR RNA (pre-crRNA) before being cleaved by endogenous RNases to yield multiple crRNAs. Each crRNA contains a single repeat-protospacer unit. The locus also expresses the endonuclease Cas9, which has the RuvC and the HNH endonuclease domains for cleaving DNA, and the trans-activating crRNA (tracrRNA), whose function is to bridge the interaction between Cas9 and crRNA. All three components—Cas9, tracrRNA, and crRNA—form a complex that binds to the target site on the phage genome as determined by sequence complementarity, allowing Cas9 to generate a double-strand break and inactivate the invading phage. The most commonly used Cas9 is from *Streptococcus pyogenes*, which also requires a 5'-NGG-3' protospacer-adjacent motif (PAM) right next to the crRNA recognition region. Thus, the overall recognition site for Cas9 is ~23 base pairs long (20 for the protospacer and 3 for the PAM). There are also other types of CRISPR systems, but they involve more proteins and are not yet as widely used in biotechnology as CRISPR-Cas9.

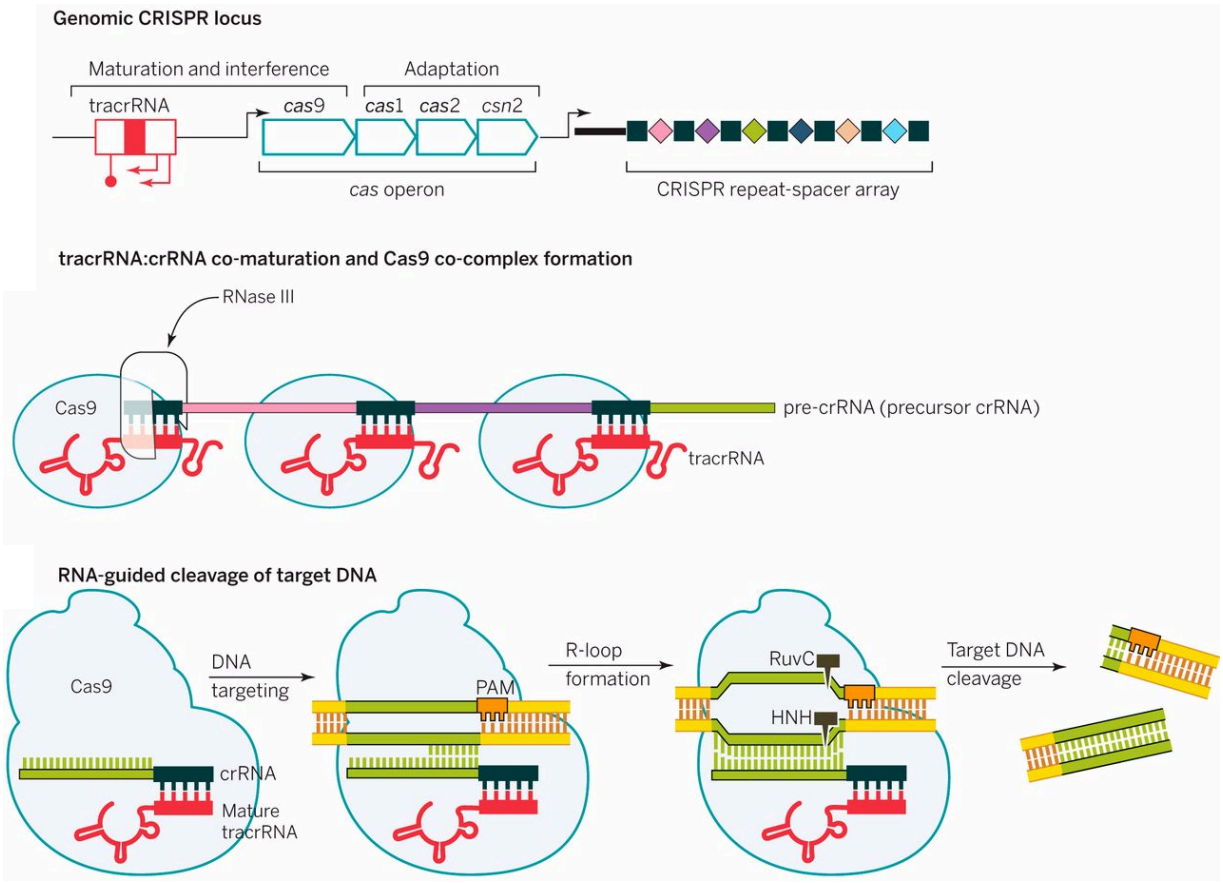


Figure 3.3 | CRISPR components and mechanism. Imaged reproduced from (Doudna and Charpentier, 2014).

Expression of Cas9, crRNA, and tracrRNA in eukaryotic cells can lead to double-strand break formation at the target site on the eukaryotic genome. TracrRNA and crRNA have been combined and re-engineered into a single entity called a single-guide RNA (sgRNA) that has the ability to bind both Cas9 and the target genomic region, further simplifying the CRISPR-Cas9 system from three components into two. Because the target-site specificity of Cas9 is determined solely by the sgRNA (as long as there is a PAM at the target site), Cas9 can be homed to almost anywhere on the genome easily by changing the sgRNA sequence.

Lastly, CRISPR-Cas9 may also be used as a DNA-targeting scaffold for targeting other cargos to the genome, rather than as an endonuclease. Cas9 bearing point mutations in the RuvC and the HNH endonuclease domains is enzymatically inactive (dead Cas9, or dCas9), but can still localize to target regions on the genome. Fusion of dCas9 to transcriptional activators such as VP64 or repressors such as KRAB allows activation or repression of specific loci, and this strategy has come to be known as CRISPR activation (CRISPRa) or CRISPR interference (CRISPRi). (Hsu et al., 2014; Kim and Kim, 2014)

Compared to ZFNs and TALENs, which require significant protein engineering efforts to construct, CRISPR-Cas9 is much simpler and more modular. Hence it has emerged as a preferred choice for use in genomic screens.

Using CRISPR-Cas9 screens to elucidate CA's mechanism in AML

Libraries of sgRNAs targeting virtually every gene in the human genome have been developed. They can be used to generate pools of loss-of-function mutants, and are therefore useful for identifying genes that are involved in a particular biological process. (Doench, 2017)

Our group has used CRISPR-Cas9 screens to identify genes that are important for the anti-proliferative activity of CA in a specific AML cell line. We first generated a pool of genome-wide knockout cells using CRISPR-Cas9 and an sgRNA library, and then treated the cells with CA. If the genes were important for the anti-proliferative activity of CA, their knockouts would confer the cells with partial resistance to CA, allowing these cells to proliferate faster than other cells when

treated with CA. After a few weeks, the pooled population would be dominated by cells containing sgRNAs that conferred such protective effects. These sgRNAs could be identified, yielding the list of genes that were important for the anti-proliferative mechanism of CA. This approach has allowed us to put together a mechanistic model for how CA suppresses the growth of some AML cells, and we present our findings in the following sections.

In **Chapter 1**, we presented evidence supporting the notion that AML is a cancer that is highly dependent on dysregulated transcription. This is exemplified by the observation that a large proportion of mutations in AML are found in transcription factors (CEBPA, RUNX1, PML-RARA), DNA methylation proteins (DNMT3A/3B, TET1/2, IDH1/2), chromatin modifiers (MLL, EZH2, ASXL1, KDM6A), and the cohesin complex (The Cancer Genome Atlas, 2013; Papaemmanuil et al., 2016). As such, targeting transcriptional regulators including BRD4 (Dawson et al., 2011; Mertz et al., 2011; Zuber et al., 2011), CDK7 (Ren et al., 2015), LSD1 (Schenk et al., 2012), IDH1/2 (Wang et al., 2013a), CDK8, and CDK19 (CDK8/19) (Pelish et al., 2015; Rzymski et al., 2017) has emerged as a promising therapeutic strategy for AML. CDK8/19 are paralogous kinases sharing >76% sequence identity. They are the only enzymatic subunits of Mediator, a ~30-protein complex that plays an essential role in eukaryotic transcription (Allen and Taatjes, 2015), making them attractive targets for pharmacological modulation. We validated CA as a potent and selective CDK8/19 inhibitor (Cee et al., 2009; Pelish et al., 2015) and used it to show that CDK8/19 inhibition arrests AML growth, in part by increasing transcription of specific genes with tumor suppressor functions (Pelish et al., 2015). Many of these genes are associated with super-enhancers, which are large clusters of cis-regulatory elements implicated in the control of cell identity and disease (Hnisz et al., 2013; Whyte et al., 2013).

Deeper mechanistic details, such as which proteins and signaling pathways mediate the anti-leukemic effect of CDK8/19 inhibition, are only beginning to be understood. In AML cells that arise from myeloproliferative neoplasms (post-MPN AMLs), phosphorylation of the transcription factor STAT1 at Ser727 by CDK8 (Bancerek et al., 2013) is indispensable for STAT1's oncogenic activity, and inhibition of this phosphorylation with CA arrests cell growth (Nitulescu et al., 2017). Other substrates of CDK8/19 have also been identified including transcription factors Smad2/Smad3 (Alarcón et al., 2009), Notch1 (Fryer et al., 2004), E2F1 (Zhao et al., 2013), and SREBP-1c (Zhao et al., 2012). Unlike post-MPN AML cells, which are dependent on JAK-STAT signaling, the majority of AML cells do not exhibit well-defined dependence on any known CDK8/19 substrate, making it difficult to study the mechanism of CA in these cells via a hypothesis-driven approach. Understanding which proteins and pathways are required for sensitivity to CDK8/19 inhibitors could lead to a means of identifying which patients are more likely to respond to CDK8/19 inhibitor therapy.

To systematically identify proteins and pathways required for CA's anti-leukemic activity, we performed an unbiased genome-wide CRISPR-Cas9 suppressor screen (Shalem et al., 2014; Wang et al., 2013b) and discovered that the antiproliferative activity of CA in the AML cell line MOLM-14 is mediated by Notch signaling. Notch signaling, depicted in **Figure 3.4**, is a fundamental pathway that controls animal development and tissue homeostasis (Andersson et al., 2011; Kopan and Ilagan, 2009). The role of Notch in blood malignancies is context-dependent; it is an oncogenic driver in T-cell acute lymphoblastic leukemia (T-ALL) (Weng et al., 2004) but is a tumor suppressor in a wide range of myeloid leukemia (Kannan et al., 2013; Klinakis et al., 2011;

Lobry et al., 2013). There are four Notch receptors (Notch1-4) in humans. Synthesized as a single polypeptide, Notch1 undergoes glycosylation and cleavage in the secretory pathway to produce Notch1 transmembrane subunit (NTM1) and Notch1 extracellular domain (NEXT1), which are held together non-covalently. Notch1 juxtacrine stimulation by Delta-like (Dll) or Jagged ligands on a neighboring cell causes NTM1 to undergo further cleavage by ADAM10/17 and the γ-secretase complex, liberating the C-terminal Notch1 intracellular domain (NICD1). NICD1 then translocates into the nucleus and interacts with transcriptional co-activators such as RBPJ, MAML1, and CREBBP/p300 to modulate expression of target genes (Kopan and Ilagan, 2009).

Notch signaling can be regulated by NICD turnover. CDK8/19 phosphorylate NICD1 at Thr2511/Ser2513/Ser2516, forming a phosphodegron that targets it for proteasomal degradation (Fryer et al., 2004; Li et al., 2014). Here we discovered that CA prevents the degradation of not only NICD1 but also NICD2 (Notch2 intracellular domain), thereby identifying NICD2 as an additional substrate (Borggrefe et al., 2016) of CDK8/19. We also discovered that NICD1/2 mediates the antiproliferative activity of CA in MOLM-14 cells, consistent with the tumor suppressor function of Notch in myeloid leukemia (Kannan et al., 2013; Klinakis et al., 2011; Lobry et al., 2013). Our findings support CDK8/19 phosphorylation of NICD1/2 or transcriptional silencing of Notch ligands as a means to suppress Notch signaling in AML cells.

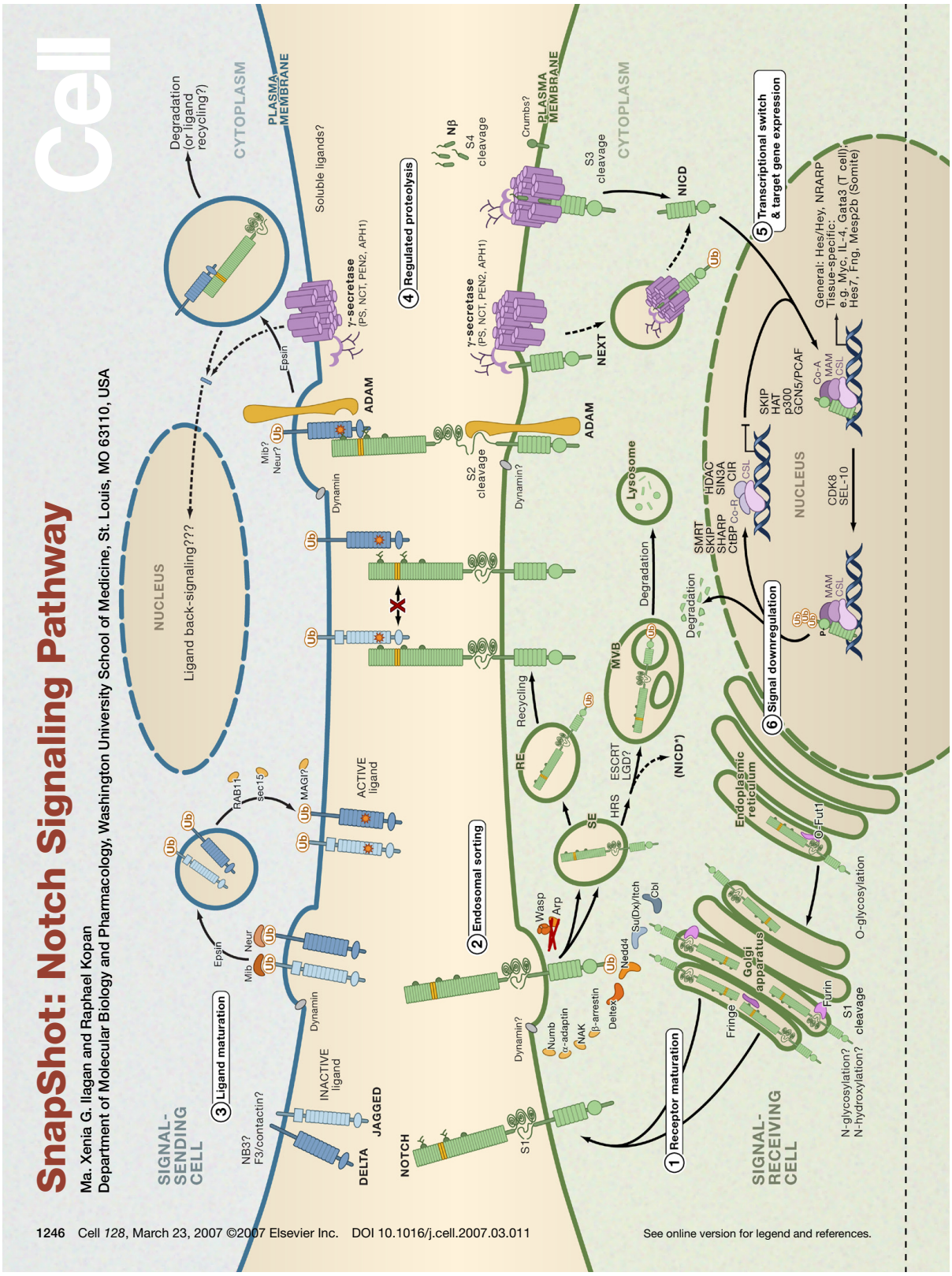


Figure 3.4 | Notch pathway. Image reproduced from Ilagan and Kopan, 2007.

CRISPR-Cas9 screen implicates Notch signaling in CA's mechanism

To identify genes that are required for the antiproliferative activity of CA, we performed a genome-wide CRISPR-Cas9 suppressor screen (Shalem et al., 2014; Wang et al., 2013b) for genes whose loss desensitized the AML cell line MOLM-14 to CA (**Figures 3.5** and **3.6A**). We chose MOLM-14 based on the availability of growth and gene expression data from our previous study (Pelish et al., 2015). MOLM-14 cells were transduced with Cas9 followed by the single-guide RNA (sgRNA) library Avana, which encodes 74,687 sgRNAs targeting ~18,000 human genes (Doench et al., 2016). Cells were cultured in the presence or absence of 100 nM CA for 15 days, which corresponds to ~12 cell doublings. The CA-treatment group showed an overall ~90% growth reduction compared to vehicle, and cells that continued to thrive were presumably enriched in sgRNAs that desensitized them to CA (**Figure 3.6B**).

We determined the abundance of each sgRNA on day 15 using massively parallel sequencing and calculated the average $\log_2$(fold-change) between CA- and vehicle-treatment groups (**Figure 3.6C**). Only ~3.8% of all sgRNAs showed >2-fold enrichment in the CA-treatment group, suggesting that these sgRNAs conferred a specific protective effect against CA. The sgRNAs were mapped to their target genes and ranked using STARS (Doench et al., 2016), revealing 97 high-confidence genes whose knockout desensitized MOLM-14 cells to CA (false discovery rate (FDR) < 0.01; **Figures 3.6D-E**).

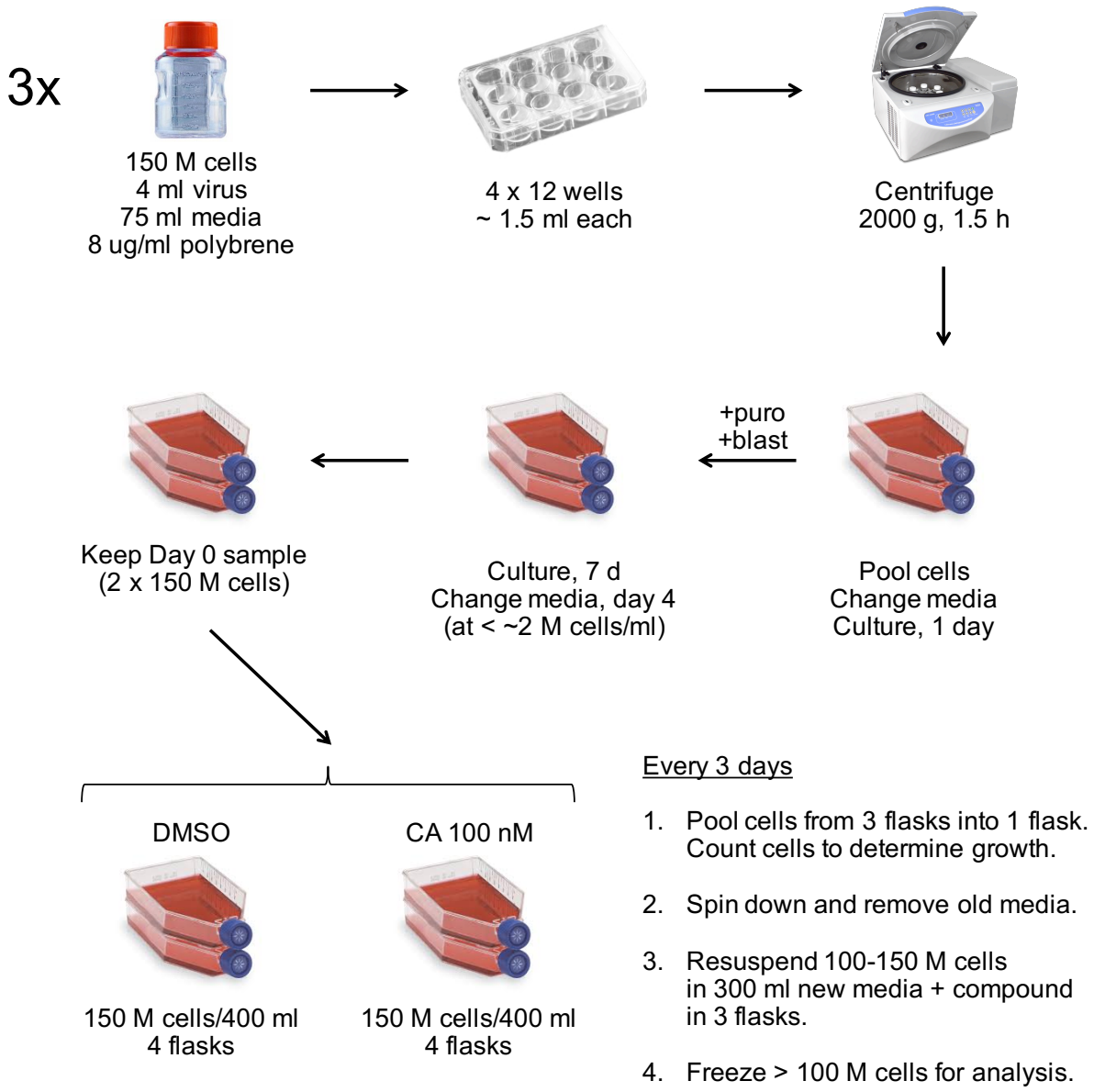
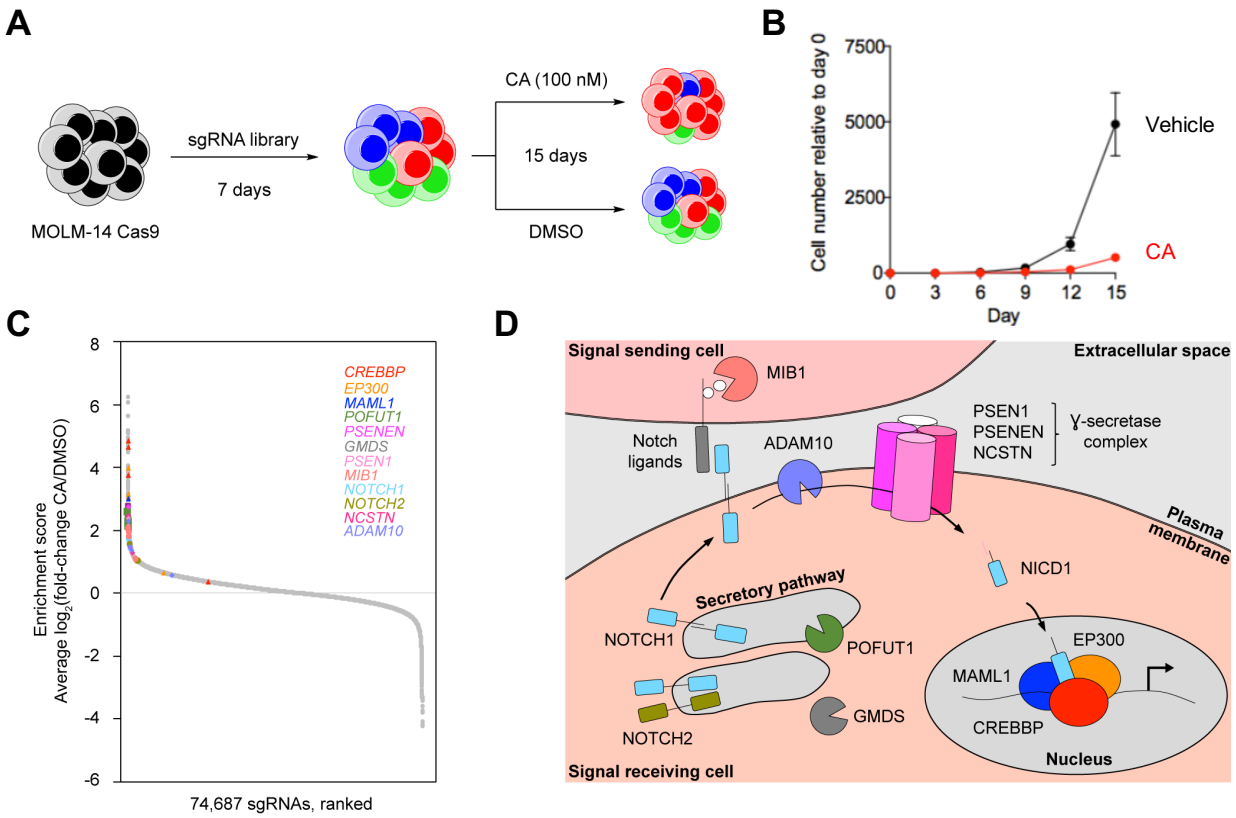


Figure 3.5 | CRISPR screen workflow.



E

Gene	FDR	Gene	FDR	Gene	FDR	Gene	FDR
TAF5L	0	EP300	0	MIER3	1.14E-04	CTDNEP1	6.16E-04
TADA1	0	ATXN7L3	0	TGDS	1.11E-04	MAPKAPK5	6.72E-04
TAF6L	0	DYRK1A	0	ASB7	1.35E-04	STAG1	6.98E-04
SUPT7L	0	PSENEN	1.54E-05	MORF4L1	1.40E-04	CDK8	7.51E-04
AMBRA1	0	INO80D	1.49E-05	ANXA2R	1.38E-04	AXIN1	8.42E-04
TADA2B	0	PSEN1	1.44E-05	NFIL3	1.59E-04	DNAJC3	8.31E-04
USP22	0	UXS1	1.39E-05	FOSL2	1.56E-04	PTPN9	1.16E-03
MAML1	0	PDIA6	1.35E-05	UBE2G1	1.92E-04	ADAM10	1.31E-03
TFPT	0	NOTCH1	1.31E-05	APC	2.19E-04	PTPN2	1.45E-03
POFUT1	0	TSPAN14	1.27E-05	UBE2G2	2.23E-04	KAT2A	1.56E-03
METRNL	0	TCF12	1.24E-05	HDAC2	2.19E-04	ZNF217	1.71E-03
RNF146	0	ATF7IP	1.21E-05	VEZF1	2.30E-04	RAB7A	1.72E-03
GMDS	0	INO80	2.35E-05	NUDT2	2.90E-04	HNRNPA1	1.93E-03
RFNG	0	SETDB1	2.29E-05	NCSTN	3.14E-04	CMTR2	2.19E-03
PHF6	0	CNOT11	2.23E-05	GXYLT1	3.22E-04	SPAG7	2.32E-03
GANAB	0	TCF3	3.26E-05	SAFB2	3.31E-04	HDAC8	2.32E-03
POGLUT1	0	TBRG1	4.25E-05	MED13	3.33E-04	PIN1	2.65E-03
SUPT20H	0	GGNBP2	4.15E-05	SERTAD2	3.34E-04	NUP98	2.82E-03
FOXK1	0	SP3	5.07E-05	FBXO38	3.49E-04	EIF3D	3.40E-03
CREBBP	0	CDYL	4.95E-05	HMX2	3.89E-04	MTMR9	7.36E-03
DHX29	0	INO80E	9.69E-05	NOTCH2	4.21E-04	IER3IP1	7.39E-03
EIF3L	0	ZNF787	9.49E-05	DDX3X	4.15E-04	UGCG	9.29E-03
DNAJB11	0	PHIP	9.29E-05	KDM1B	4.58E-04		
UCHL5	0	WIZ	1.00E-04	PTGES3	4.76E-04		
CCNT1	0	MIB1	9.81E-05	SIRT1	5.05E-04		

Figure 3.6 | CRISPR screen identifies Notch signaling as important for CA's mechanism. (**A**) CRISPR screen workflow. (**B**) Cell growth in pooled screens upon treatment with vehicle or CA.

(**C**) Enrichment of all sgRNAs in the pooled screen in CA versus DMSO treatment groups. Hits in the genes encoding components of the Notch pathway are highlighted in colors. (**D**) Notch pathway schematic highlighting components of the Notch pathway that are hits from the pooled screen. (**E**) All 97 hits with flase discovery rate < 0.01 from the pooled screen as analyzed by the STARS algorithm.

One of the most outstanding gene ontology signatures that emerged from these hits was Notch signaling (12 hits, p-value = 5.13×10^{-9}; **Table 3.1**). The 12 hits span every step of the Notch pathway including receptor synthesis (*NOTCH1*, *NOTCH2*), Notch glycosylation (*POFUT1*, *GMDS*) (Shi and Stanley, 2003; Stanley, 2007), proteolytic activation (*ADAM10*, *PSEN1*, *PSENEN*, *NCSTN*) (Hartmann et al., 2002; Shih and Wang, 2007), transcriptional co-activation (*MAML1*, *CREBBP*, *EP300*) (Oswald et al., 2001; Wu et al., 2000), and ligand processing (*MIB1*) (Itoh et al., 2003) (**Figure 3.6D**). Three other hits—*RFNG* (Hicks et al., 2000; Panin et al., 1997), *POGLUT1* (Acar et al., 2008), and *GXYLT1* (Sethi et al., 2010)—also encode enzymes known to be crucial for Notch glycosylation and activity (**Figure 3.6E**). In total, our CRISPR-Cas9 screen identified at least 15 known components of the Notch pathway, strongly suggesting that Notch signaling is required for the antiproliferative activity of CA in MOLM-14.

Table 3.1 | GO analysis of CRISPR-Cas9 screen hits

GO biological process	Number of genes	Number of hits	Fold-enrichment	*p*-value	Hits
Notch signaling pathway	117	12	21.54	5.13×10^{-9}	*NCSTN, PSENEN, CREBBP, POFUT1, EP300, ADAM10, PSEN1,NOTCH1, NOTCH2, GMDS, MIB1, MAML1*
Protein acetylation	130	12	19.39	1.73×10^{-8}	*TADA1, USP22, TAF6L, CREBBP, EP300, TAF5L, SUPT7L, TADA2B, MORF4L1, CDYL, KAT2A, SIRT1*

Notch pathway disruption desensitized AML to CA

To confirm the results from the screen, we selected 5 sgRNAs targeting different steps of the Notch pathway (*NOTCH1*, *PSEN1*, *MAML1*, *CREBBP*, and *EP300*) and examined their effects on the sensitivity of MOLM-14 cells to CA. We validated the knockout target genes using western blot and T7E1 assays, which measure the frequency of insertions/deletions (indels) generated by CRISPR-Cas9 at target loci, and calculated frequencies to be 50-70% (**Figure 3.7**). Consistent with the results from the screen, knocking out any of the 5 Notch components desensitized MOLM-14 cells to the antiproliferative effect of CA, increasing GI_{50} by up to 100-fold (**Figure 3.8A**).

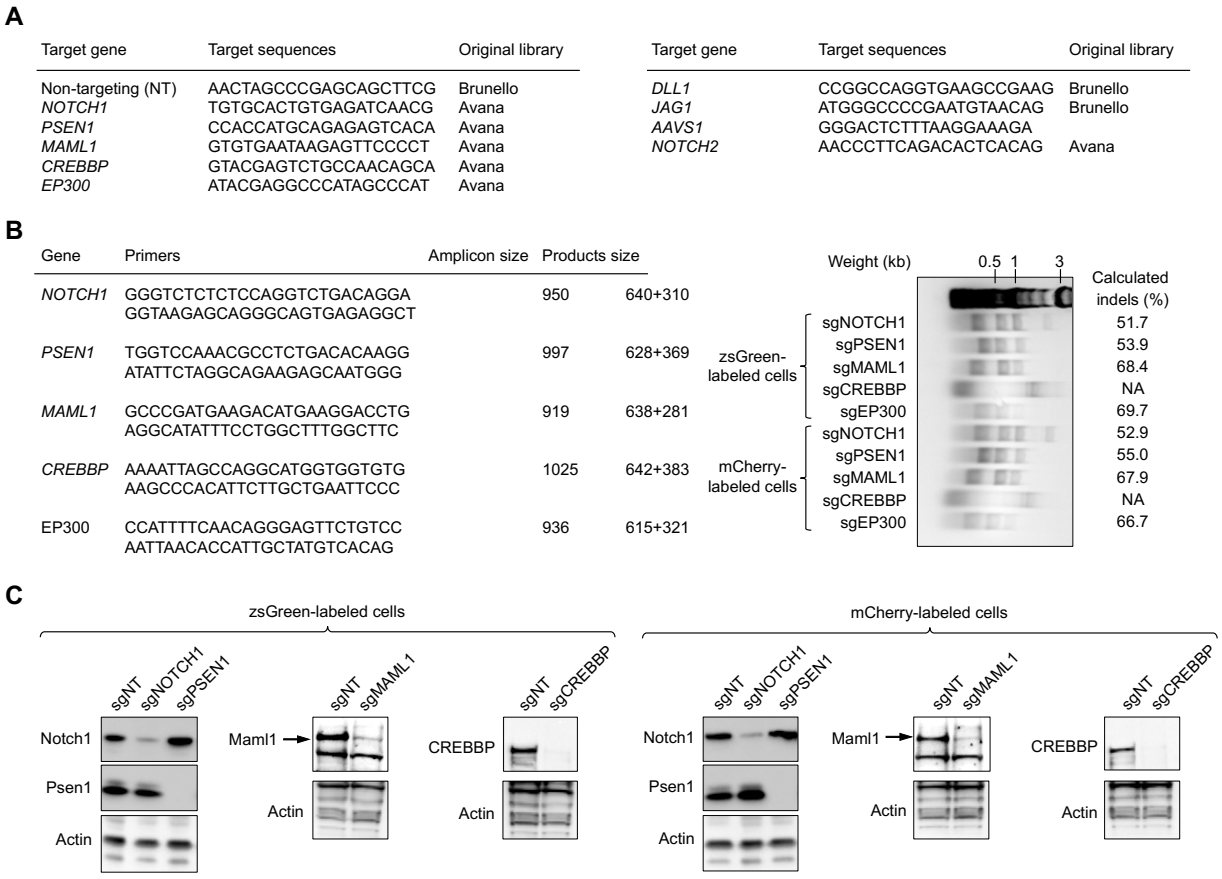


Figure 3.7 | Knockout of Notch-component genes. (**A**) Sequences targeted by each sgRNA. (**B**) T7E1 assay. Left, primers used to amplify regions surrounding the desired cut site generated by each sgRNA, with the size of the amplicon and the expected sizes of the products after T7E1 cleavage. Right, components in the T7E1 reaction mixtures resolved on an agarose gel. (**C**) Western blot showing disappearance of proteins targeted by each sgRNA in MOLM-14 cells, which expressed either zsGreen or mCherry.

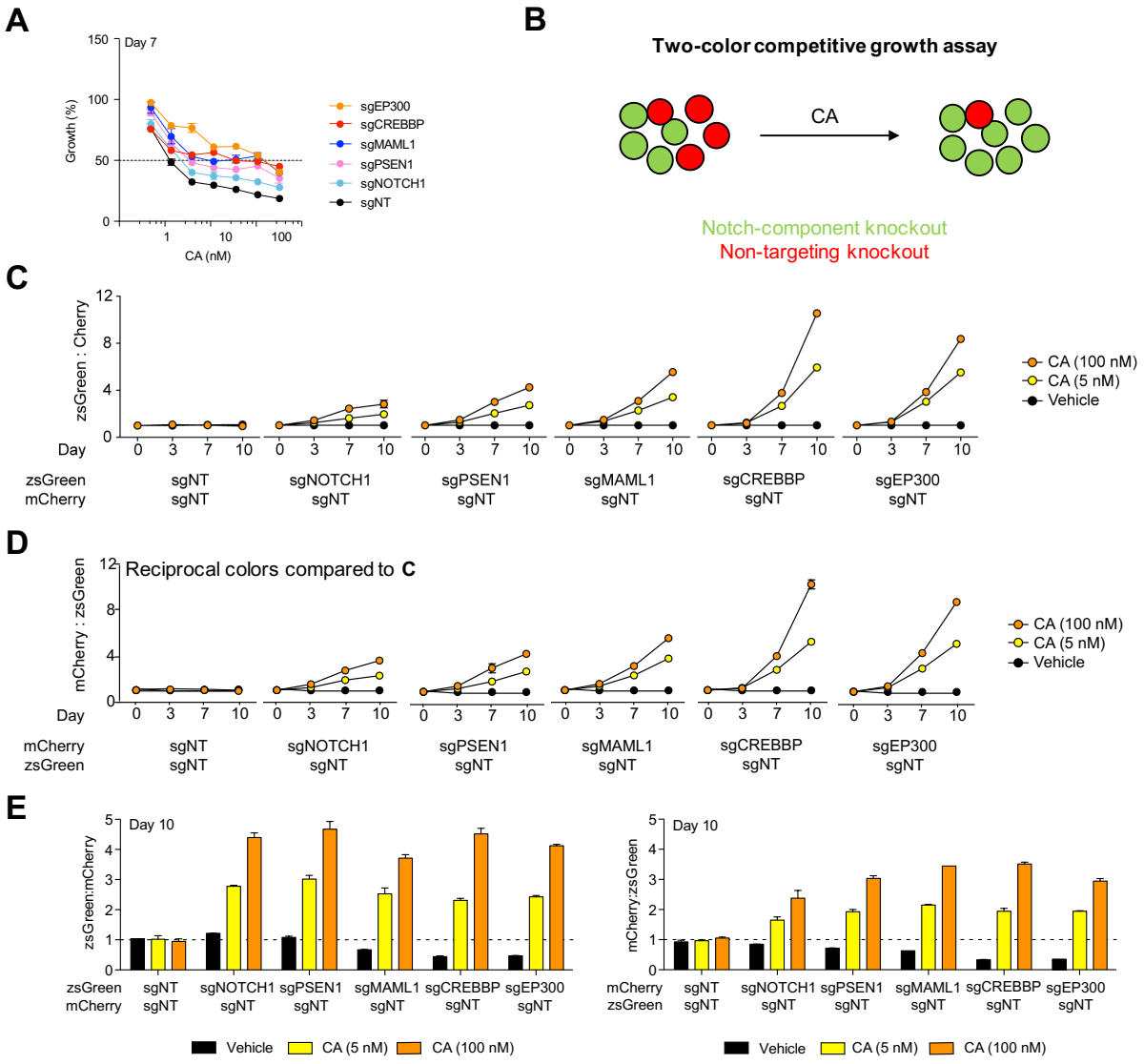


Figure 3.8 | Notch-component knockout impairs sensitivity to CA. (**A**) Notch-component knockout reduces MOLM-14 sensitivity to growth inhibition by CA. (**B**) Schematic of a two-color competitive growth assay. (**C**) Competitive growth assays in MOLM-14 showing the change in zsGreen-to-mCherry ratio over time upon CA treatment. Each color is co-expressed with an sgRNA as annotated below the graphs. sgNT, non-targeting sgRNA. All values are normalized to vehicle treatment. (**D**) Same as **C**, except with reciprocal fluorescent proteins. (**E**) Competitive growth assays showing the zsGreen-to-mCherry ratio as in **C** and **D**, but without normalizing to vehicle treatment.

We also assessed the effect of these knockouts by a two-color competitive growth assay (**Figure 3.8B**). We co-expressed the fluorescent protein zsGreen with the sgRNA targeting each component of the Notch pathway. We then mixed these cells in a 1:1 ratio with control MOLM-14 cells expressing mCherry, treated the mixture with 5 or 100 nM CA, and monitored the change in the zsGreen:mCherry ratio over 10 days using flow cytometry. An increase in this ratio would indicate competitive growth advantage and suggest that the knockout of the Notch component desensitized MOLM-14 cells to CA.

In the absence of CA, we did not observe any growth advantage; in fact, *MAML1*, *CREBBP*, and *EP300* knockout showed a slight growth disadvantage, suggesting that these genes are partly essential to MOLM-14 growth (**Figure 3.8E**). In the presence of 100 nM CA, however, we observed a strong growth advantage in all knockout cases, with a 4- to 10-fold increase in zsGreen:mCherry after 10 days compared to vehicle treatment (**Figure 3.8C**). Identical results were obtained when the cells were reciprocally labeled, ruling out the possibility that the observed growth advantage was a result of fluorescent protein expression (**Figure 3.8D**). These competitive growth assays further confirm the results from our screen that various components of the Notch pathway are required for the antiproliferative activity of CA in MOLM-14.

CA inhibits NICD1 degradation

CDK8/19 have been shown to phosphorylate NICD1, the transcriptionally active cleavage product of Notch1, promoting its degradation by the ubiquitin-proteasome system *in vitro* and in mice (Fryer et al., 2004; Li et al., 2014). Consistent with this role for CDK8/19, we observed that

treatment of the AML cell line MOLM-14 with CA increased NICD1 without changing the amount of total Notch1 (EC_{50} ~ 1 nM; **Figures 3.9A-B**), indicating that the effect of CA on NICD1 was post-translational. Treatment with the proteasome inhibitor MG132 also caused an increase in NICD1, confirming that NICD1 was subjected to degradation by the proteasome in MOLM-14 (**Figure 3.9C**).

Although NICD1 could be stabilized by either CA or MG132, the NICD1 band induced by CA migrated faster than the one induced by MG132 on SDS-PAGE, consistent with the greater mobility of hypophosphorylated NICD1 (Fryer et al., 2004) following CDK8/19 inhibition (**Figure 3.9C**). Dephosphorylation in the cell lysate with lambda phosphatase caused both NICD1 bands to coalesce, confirming that the difference in electrophoretic mobility of NICD1 arises from differential phosphorylation (**Figure 3.9D**).

To quantify the effect of CA on NICD1 stability, we pulsed MOLM-14 with EDTA for 15 minutes and monitored the disappearance of NICD1 over time. EDTA disrupts the folding of NTM1 by sequestering Ca^{2+}, allowing it to be rapidly cleaved into NICD1 (Rand et al., 2000). This pool of NICD1 became immediately visible on western blot but was quickly phosphorylated (as evidenced by mobility shifts) and degraded with a half-life ($t_{1/2}$) ~ 2 hours (**Figures 3.9E-F**). CA treatment substantially prolonged the lifetime of NICD1 ($t_{1/2}$ > 8 hours; **Figure 3.9E**). These results suggest that NICD1 was continually phosphorylated and degraded through a CDK8/19-dependent process in MOLM-14. Inhibition of CDK8/19 with CA allowed NICD1 to stabilize and accumulate in abnormally large quantities.

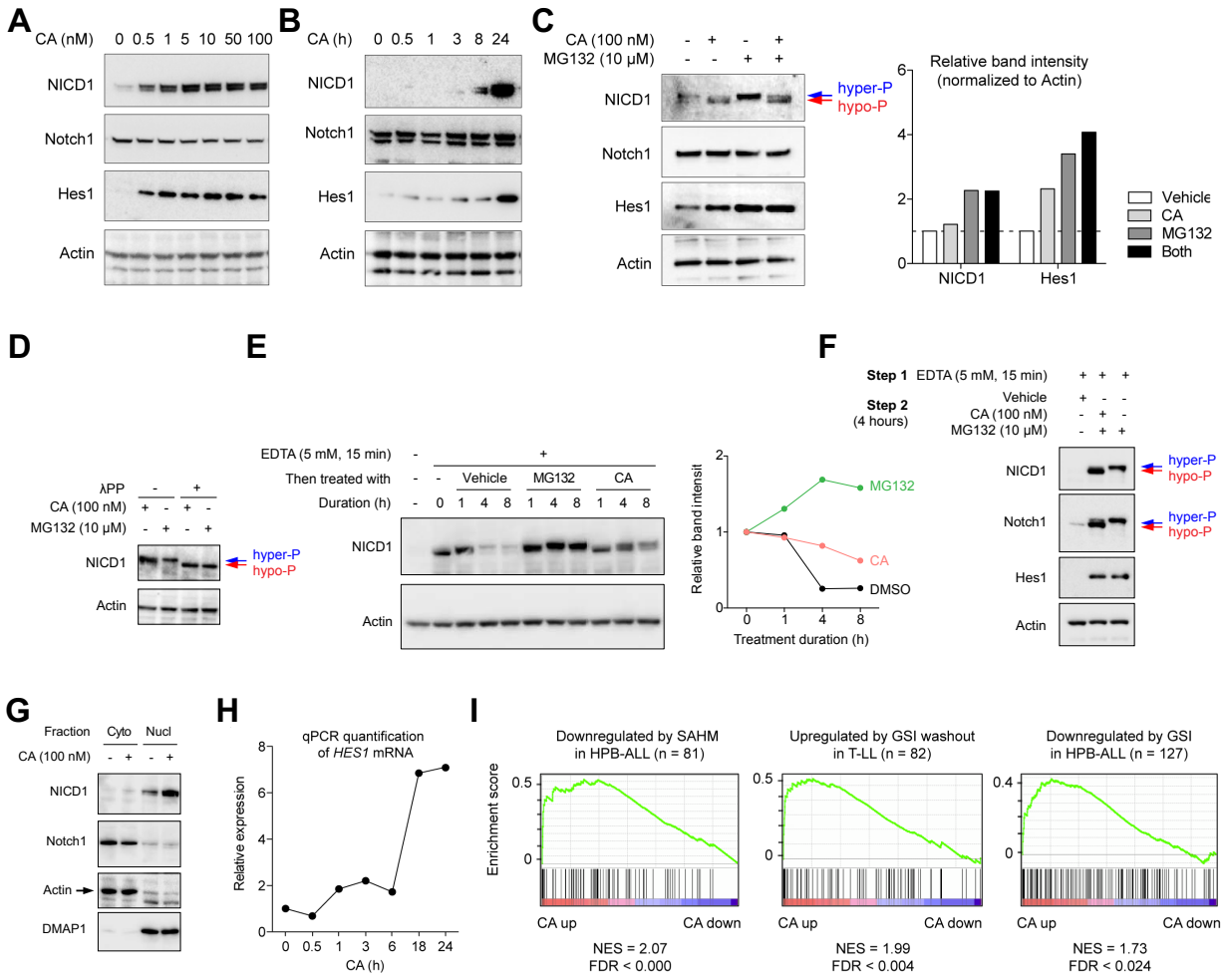


Figure 3.9 | CA stabilizes NICD1 and activates Notch signaling. (**A**-**B**) CA upregulates NICD1 and Hes1 dose-dependently (**A**) and time-dependently (**B**) in MOLM-14. (**C**) Effects of CA (4 hours), MG132 (4 hours), or co-treatment with CA and MG132 (4 hours) on NICD1 and Hes1 in MOLM-14. Left, western blots showing upregulation of NICD1 and Hes1. NICD1 mobility shift indicates differential phosphorylation states as annotated. Hyper-P and hypo-P stand for hyper- and hypo-phosphorylated, respectively. Right, densitometric quantification of NICD1 and Hes1 band intensities normalized to actin. Values are normalized to vehicle treatment. (**D**) Difference in NICD1 mobility arises from differential phosphorylation. NICD1 bands arising from CA treatment (24 hours) or MG132 treatment (4 hours) in MOLM-14 have different mobility.

Treatment with lambda phosphatase (λPP) cause both bands to coalesce. Hyper-P and hypo-P stand for hyper- and hypo-phosphorylated, respectively. (**E**) Left, western blot showing the effects of MG132 or CA on the lifetime of NICD1 that was generated by EDTA treatment in MOLM-14. Right, densitometric quantification of the blot on the left. Values are normalized to Actin and to the control cells at hour 0. (**F**) Differentially phosphorylated species of NICD1 were detected with NICD1 or Notch1 antibody in MOLM-14 nuclear extracts. Cells were first treated with EDTA to stimulate NICD1 production, and then treated with both CA and MG132, or just MG132. Nuclei were extracted and analyzed. Hyper-P and hypo-P stand for hyper- and hypo-phosphorylated, respectively. (**G**) NICD1 stabilized by CA localizes exclusively to the nucleus in MOLM-14. DMAP1 is used as a nuclear protein marker. Cyto, cytoplasmic fraction. Nucl, nuclear fraction. (**H**) qPCR showing upregulation of *HES1* mRNA by CA (100 nM) in MOLM-14 cells. (**I**) Gene set enrichment analysis showing that CA-upregulated genes are enriched in Notch target genes. The ranked gene list was obtained from treating MOLM-14 with 10 nM CA for 24 hours. SAHM, stapled alpha-helical peptides derived from MAML1 (an inhibitor of Notch1-MAML1 interaction). GSI, γ-secretase inhibitor (a Notch pathway inhibitor). HPB-ALL and T-LL, T-cell leukemia cell lines. NES, normalized enrichment score. FDR, false discovery rate.

Accumulation of NICD1 activates Notch-dependent transcription in AML

NICD1 must translocate from the cell membrane into the nucleus to activate Notch-dependent transcription. Accordingly, NICD1 that accumulated following CA treatment was localized exclusively to the nucleus (**Figure 3.9G**). CA also caused a time- and dose-dependent increase in the mRNA and the protein product of *HES1*, a canonical Notch1 target gene, suggesting that NICD1 accumulation led to activation of Notch-dependent transcription (**Figures 3.9A-B**).

Because the effect of CA on *HES1* transcription reached maximum after ~18 hours (**Figure 3.9H**), we used the 24-hour global gene expression data for MOLM-14 treated with 10 nM CA to perform gene set enrichment analyses (GSEAs). Genes upregulated by CA in MOLM-14 were enriched in several Notch target gene signatures, including genes that were downregulated upon treatment with Notch inhibitors (SAHM or the γ-secretase inhibitor (GSI)) in HPB-ALL cells (Moellering et al., 2009; Trimarchi et al., 2014) or genes that were upregulated when T-LL cells pre-treated with GSI were washed out (Wang et al., 2014) (**Figure 3.9I**). These GSEA results suggest that the stabilization of NICD1 by CA was sufficient to induce Notch-dependent transcription in MOLM-14.

NICD1 mediates the antiproliferative activity of CA in AML

The accumulation of NICD1, which has been shown to be a tumor suppressor in some myeloid leukemia (Kannan et al., 2013; Klinakis et al., 2011; Lobry et al., 2013), might account for the antiproliferative activity of CA in MOLM-14. We performed two complementary experiments to test this hypothesis. First, we expressed FLAG-tagged NICD1 using a doxycycline-inducible promoter and found that it strongly inhibited the growth of MOLM-14 (**Figure 3.10A**), suggesting that the antiproliferative effect of CA in MOLM-14 could be caused by NICD1. FLAG-NICD1 induction was accompanied by Hes1 upregulation, confirming that the construct was transcriptionally active (**Figure 3.10B**). Second, we treated MOLM-14 cells with compound E, a γ-secretase inhibitor, to inhibit NICD1 production. Compound E desensitized MOLM-14 to CA (~10-fold increase in GI_{50}; **Figure 3.10C**) and strongly prevented CA from inducing NICD1

accumulation (EC_{50} < 50 nM; **Figure 3.10D**). Taken together, these experiments suggest that NICD1 accumulation is both necessary and sufficient to mediate the antiproliferative activity of CA in MOLM-14.

The growth inhibitory effect of NICD1 could be attributed to its target gene *HES1*, which has been shown to be a tumor suppressor in some AML cells (Kannan et al., 2013). We assessed the effect of Hes1 expression on MOLM-14 growth using a doxycycline- or a tamoxifen-inducible system (**Figures 3.10E-H**). Induction of Hes1 showed little or no effect on MOLM-14 growth whereas it strongly inhibited K562, a chronic myeloid leukemia cell line known to be sensitive to Hes1 overexpression (Kannan et al., 2013). This observation suggests that, unlike many other AML cell lines (Kannan et al., 2013), MOLM-14 sensitivity to NICD1 could not be fully explained by Hes1 upregulation.

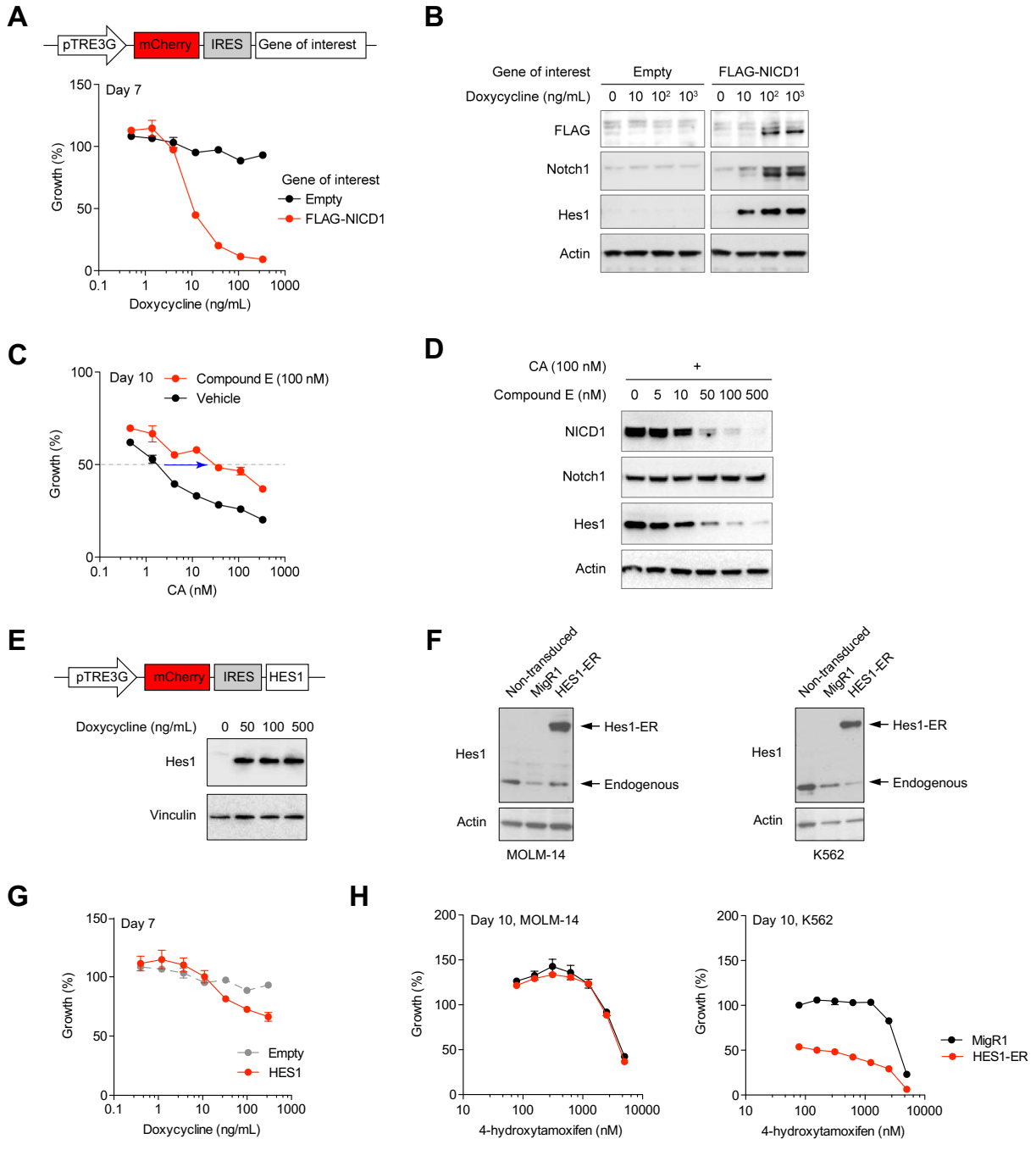


Figure 3.10 | Notch as a tumor suppressor in AML. (**A**) Top, plasmid map of pTRE3G-mCherry-IRES, a doxycycline-inducible gene expression construct (Clontech). Bottom, growth assay showing the tumor suppressor function of NICD1 in MOLM-14. IRES, internal ribosome entry site. (**B**) Western blots showing doxycycline-induction of FLAG-tagged NICD1 and upregulation

of Hes1 (1-day treatment) in MOLM-14. (**C**) Growth assay showing that compound E desensitizes MOLM-14 cells to the antiproliferative effect of CA. GI_{50} shift is highlighted with a blue arrow. (**D**) Western blots showing that co-treatment (1-day) with compound E suppresses the ability to CA to induce NICD1 and Hes1 in MOLM-14. (**E**) Top, plasmid map of doxycycline-inducible Hes1 expression construct. Bottom, induction of Hes1 expression in MOLM-14 by doxycycline (1-day treatment). (**F**) Western blots showing expression of the Hes1-ER (estrogen receptor) fusion in MOLM-14 (left) and K562 (right). Endogenous Hes1 is lower in molecular weight as indicated. MigR1 is the control empty vector. (**G**) Effect of Hes1 induction on MOLM-14 growth. (**H**) Effects of Hes1-ER induction by 4-hydroxytamoxifen on the growth of MOLM-14 (left) or K562 (right).

Notch ligands are required for NICD1 stabilization by CA

NICD1 has been shown to inhibit the growth of many AML cells (Kannan et al., 2013), prompting us to speculate that NICD1 might mediate the antiproliferative effect of CA in AML cell lines other than MOLM-14. We therefore tested for the ability of CA to induce NICD1 accumulation in 10 additional AML cell lines that were sensitive to growth inhibition by CA, but found that none of these cells showed NICD1 accumulation in response to CA treatment (**Figure 3.11A**). One plausible explanation is that NICD1 was being degraded through a CDK8/19-independent process and hence could not be stabilized by CA. However, NICD1 remained undetectable even after treatment with MG132 (**Figure 3.11B**), suggesting that the absence of NICD1 in these cells was not due to degradation but rather the failure to produce NICD1 in the first place.

We hypothesized that the inability to produce NICD1 resulted from a lack of Notch ligands, which are essential for triggering Notch1 proteolysis. Kannan et al. reported that a set of AML cell lines distinct from our panel expressed Notch receptors but not Dll/Jagged ligands, consistent with the low basal Notch activity in these cells (Kannan et al., 2013). Indeed, we found that none of our AML cell lines expressed any of the five Notch ligands (*DLL1*, *DLL3*, *DLL4*, *JAG1*, and *JAG2*) except for MOLM-14 which expressed *DLL1* and *JAG1* (**Figure 3.11D**). The unique expression of *DLL1/JAG1* would allow for the cleavage of Notch1 into NICD1 in MOLM-14 cells but not in the other AML cell lines.

To demonstrate that *DLL1/JAG1* are functionally required for NICD1 stabilization by CA, we knocked out both *DLL1* and *JAG1* in MOLM-14 (**Figure 3.11C**). The ability of CA to induce NICD1 and Hes1 upregulation was greatly diminished in the *DLL1/JAG1* double-knockout cells (**Figures 3.11E**). By preventing the accumulation of NICD1, *DLL1/JAG1* knockout also desensitized MOLM-14 cells to CA (increasing GI_{50} by ~10-fold; **Figures 3.11F**). Taken together, these results indicate that Notch ligands are required for NICD1 stabilization and MOLM-14 growth inhibition by CA.

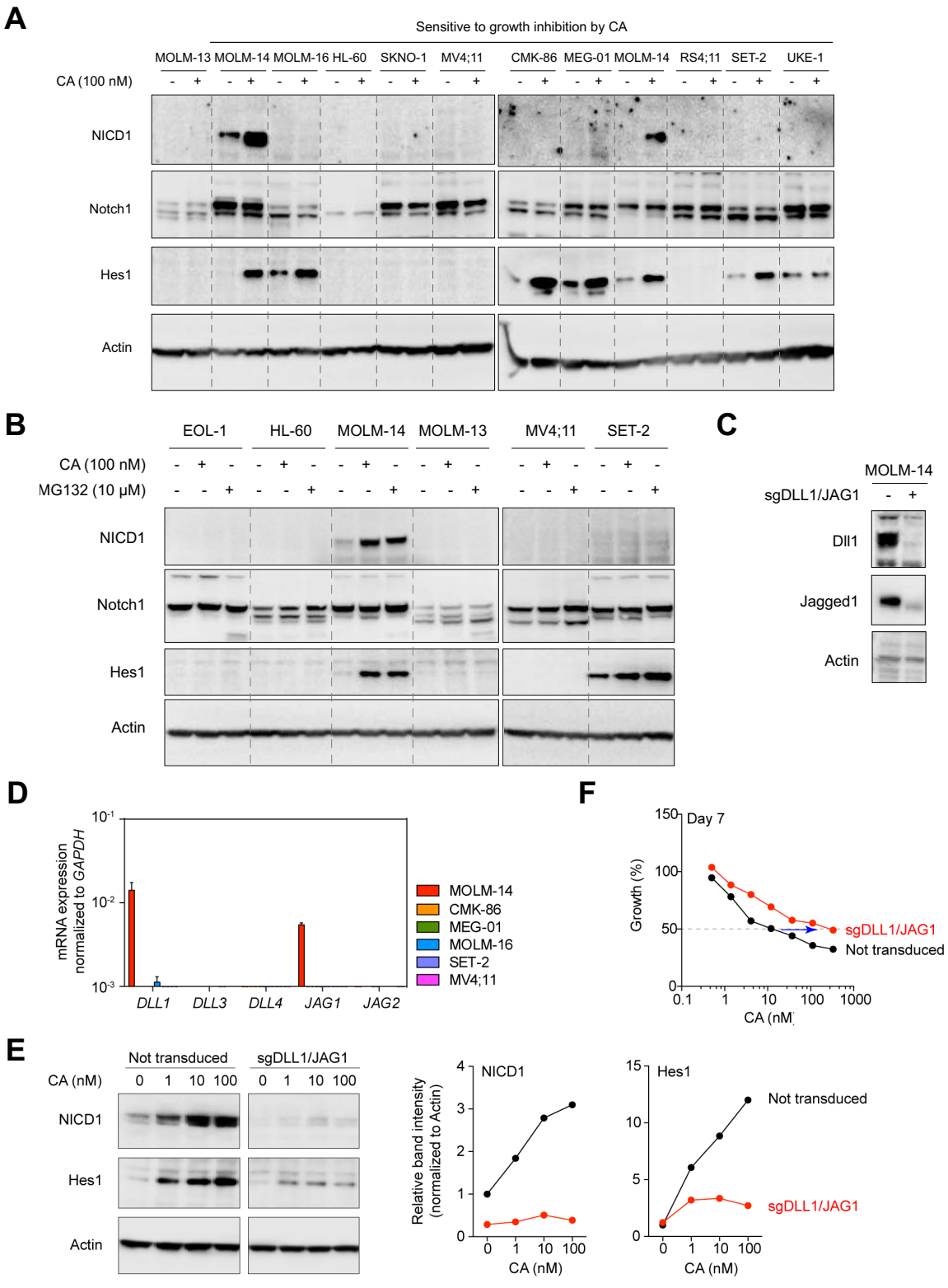


Figure 3.11 | Notch ligands are required for CA mechanism. (**A**) Western blots showing the effect of CA (1-day treatment) on NICD1 stabilization in 11 blood cancer cell lines. (**B**) Western

blots showing the effect of CA (1-day treatment) or MG132 (4-hour treatment) on NICD1 stabilization in 6 cell lines. (**C**) Western blot showing *DLL1*/*JAG1* knockout in MOLM-14. (**D**) qRT-PCR showing expression of Notch ligand mRNAs (*DLL1*, *DLL3*, *DLL4*, *JAG1*, and *JAG2*) normalized to *GAPDH* in 6 AML cell lines. (**E**) Left, western blots showing that *DLL1*/*JAG1* knockout suppresses the ability of CA (1-day treatment) to induce NICD1 and Hes1 in MOLM-14. Right, densitometric quantification of NICD1 and Hes1 band intensities in the blots on the left. Values are normalized to Actin. (**F**) Growth assay showing that *DLL1*/*JAG1* knockout desensitizes MOLM-14 cells to the antiproliferative effect of CA. GI_{50} shift is highlighted with a blue arrow. Average and SEM of three biological replicates, one of four experiments shown.

Notch1 does not fully account for the effect of CA on Notch signaling

To investigate the specific role of each Notch component in the mechanism of CA, we used sgRNAs to target 3 Notch component genes (*NOTCH1*, *PSEN1*, or *MAML1*) in MOLM-14 cells, generated monoclonal populations by single-cell sorting, and picked two homozygous clones from each knockout (**Figure 3.12A**). We then treated each clone with 100 nM CA and examined the level of Hes1 as a readout for Notch transcriptional activity.

MAML1-knockout clones were able to accumulate NICD1 normally but failed to induce Hes1, consistent with the known role of Maml1 as an important transcriptional co-activator of Notch (McElhinny et al., 2008) (**Figure 3.12B**). *PSEN1*-knockout clones failed to upregulate either NICD1 or Hes1, consistent with the requirement for γ-secretase in generating transcriptionally active NICDs (Shih and Wang, 2007) (**Figure 3.12B**). Intriguingly, *NOTCH1* knockout resulted in complete removal of NICD1 but did not suppress Hes1 induction, suggesting that CA might

activate Notch-dependent transcription by stabilizing an additional Notch receptor (**Figure 3.12B**).

NOTCH2 appeared as a high-confidence hit in our CRISPR-Cas9 screen (FDR = 4.21×10^{-4}; **Figure 3.6E**). Like Notch1, full-length Notch2 is cleaved in the secretory pathway to produce NEXT2 and NTM2, which undergoes Dll/Jagged-dependent proteolysis to liberate NICD2. Kannan et al. showed that NICD2 overexpression suppressed the growth of various AML cell lines, suggesting that it might also mediate the antiproliferative activity of CA in MOLM-14 cells (Kannan et al., 2013).

CA inhibits NICD2 degradation

CDK8/19 have not been shown to regulate NICD2 turnover. To study the effect of CA on NICD2 stability, we treated MOLM-14 cells with CA and probed for NICD2 using an antibody against Notch2 C-terminus, which simultaneously detects full-length Notch2 (265 kDa), NTM2 (92 kDa), and NICD2 (83 kDa). Because NICD2 localizes exclusively to the nucleus, we performed nuclear extraction before blotting for NICD2 to minimize NTM2 background signal. CA treatment caused a low-molecular weight band corresponding to NICD2 to emerge, indicating that NICD2 is degraded through a CDK8/19-dependent process (**Figure 3.12C**). We confirmed the identity of this band as NICD2, as it disappeared upon co-treatment with compound E (**Figure 3.12C**). CDK8/19 likely phosphorylate NICD2 at the TPSPESP motif that is perfectly conserved from the CDK8/19 phosphorylation sites on NICD1 (Li et al., 2014).

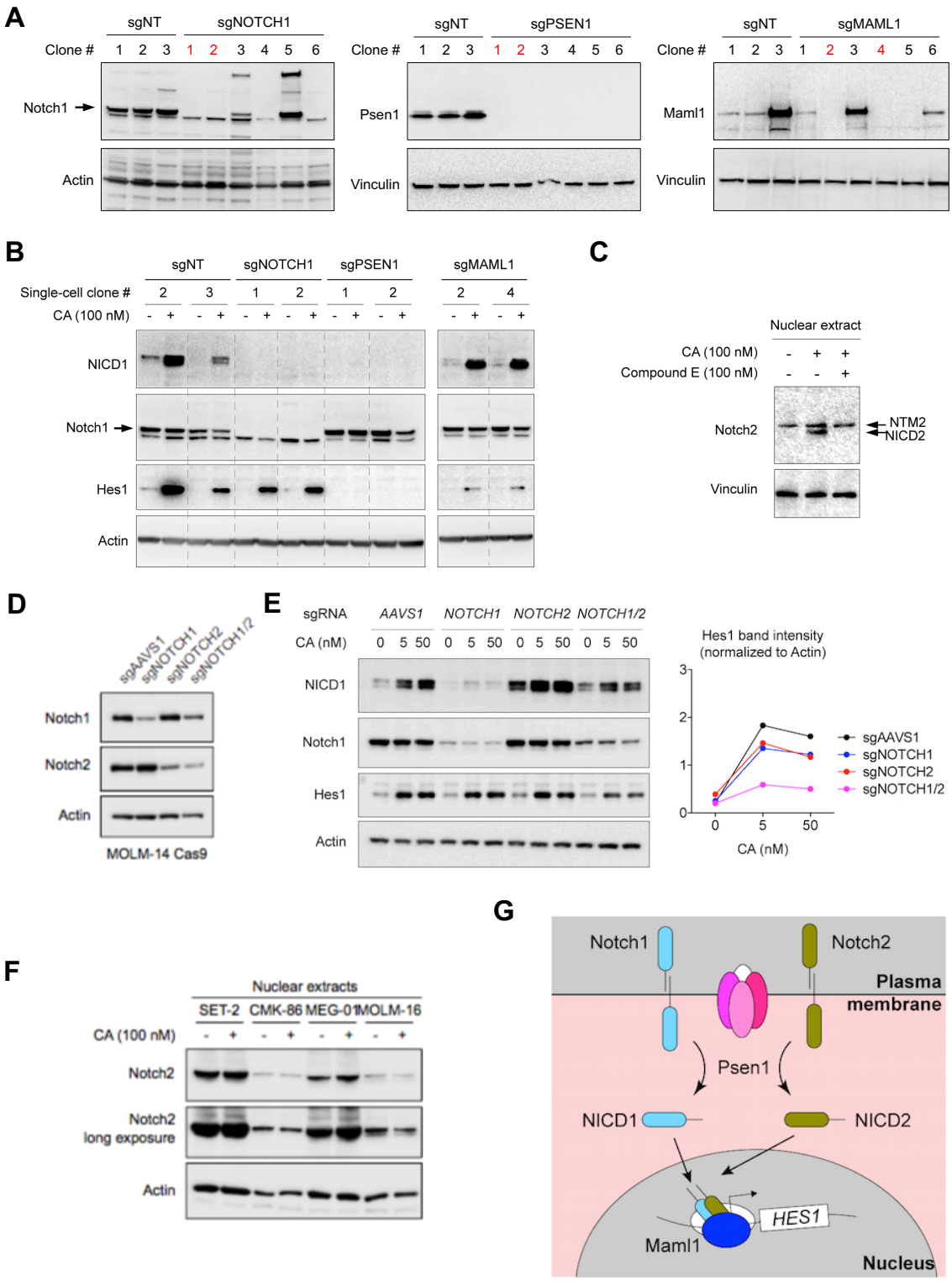


Figure 3.12 | CA stabilizes NICD2. (**A**) Western blots showing elimination of target proteins in single-cell clones in which *NOTCH1*, *PSEN1*, or *MAML1* was knocked out. The clones that were

picked for further studies in **B** are highlighted in red. (**B**) Western blots showing the effects of Notch-pathway gene knockouts on the activity of CA (1-day treatment) in single-cell clonal populations. (**C**) Western blot of MOLM-14 nuclear extract showing the effects of treatment with CA (1 day) or co-treatment with CA and compound E (1 day) on NICD2. NTM2 (Notch2 transmembrane domain) and NICD2 were distinguished based on molecular weights as annotated. (**D**) Western blots showing a decrease in target proteins in MOLM-14 when *NOTCH1*, *NOTCH2*, or both *NOTCH1/2* were knocked out. (**E**) Left, western blots showing the effects of *NOTCH1*, *NOTCH2*, or *NOTCH1/2*-double knockout on the ability of CA to upregulate NICD1 and Hes1 in MOLM-14 (1-day treatment). Right, densitometric quantification of Hes1 from the blots on the left. (**F**) Effects of CA (1-day treatment) on NICD2 in different cell lines. NICD2 was probed in nuclear extracts using the Notch2 antibody. (**G**) A model involving both Notch1 and Notch2 being cleaved by Psen1 into NICD1/2 which, when stabilized by CA, upregulate *HES1*.

To demonstrate that Notch2 mediates the transcriptional effect of CA, we used CRISPR-Cas9 to target *NOTCH1*, *NOTCH2*, or both *NOTCH1* and *NOTCH2* (*NOTCH1/2*) in MOLM-14 cells and achieved moderate knockout (**Figure 3.12D**). Upon treatment with CA, Hes1 induction was only slightly lowered in *NOTCH1* or *NOTCH2* single-knockout cells, but was greatly diminished in *NOTCH1/2* double-knockout cells (**Figure 3.12E**). CA did not induce NICD2 accumulation in cell lines that do not express *DLL1*/*JAG1*, consistent with the requirement for Notch ligands in NICD2 generation (**Figure 3.12F**). These observations are consistent with our hypothesis that CA prevents the degradation of both NICD1 and NICD2, either of which was sufficient to induce Notch-dependent transcription (**Figure 3.12G**).

NICD2 mediates the antiproliferative activity of CA in AML

To further show that Notch2 mediates the antiproliferative effect of CA, we performed two-color competitive growth assays with *NOTCH1*-, *NOTCH2*-, or *NOTCH1/2*-knockout cells (**Figure 3.13A**). The single knockouts conferred slight competitive growth advantage, with a ~3-fold increase in the mCherry:zsGreen ratio after 10-day treatment with 100 nM CA, whereas the double knockout gave a >15-fold increase in the ratio (**Figures 3.13B**). This synergism suggests that Notch1 and Notch2 share at least some overlapping function in the antiproliferative mechanism of CA, and both must be knocked out to attain a high degree of desensitization. In agreement with this view, the knockout of any Notch component that affects both Notch1 and Notch2 simultaneously (such as *PSEN1*, *POFUT1*, or *MAML1*) seemed to give a stronger desensitization effect than *NOTCH1* or *NOTCH2* single-knockout alone, both in the CRISPR-Cas9 screen (**Figure 3.6E**) and in competitive growth assays (**Figures 3.8C-D**).

Lastly, we expressed FLAG-tagged mouse NICD2 (mNICD2, which shares 92.5% protein sequence identity with human NICD2) under a doxycycline-inducible promoter. Induction of FLAG-mNICD2 expression strongly inhibited cell growth, consistent with the tumor suppressor function of NICD2 in other AML cells (Kannan et al., 2013) (**Figures 3.13C-D**). Overall, our results indicate that NICD2, like NICD1, mediates the antiproliferative effect of CDK8/19 inhibition in MOLM-14 (**Figure 3.13E**).

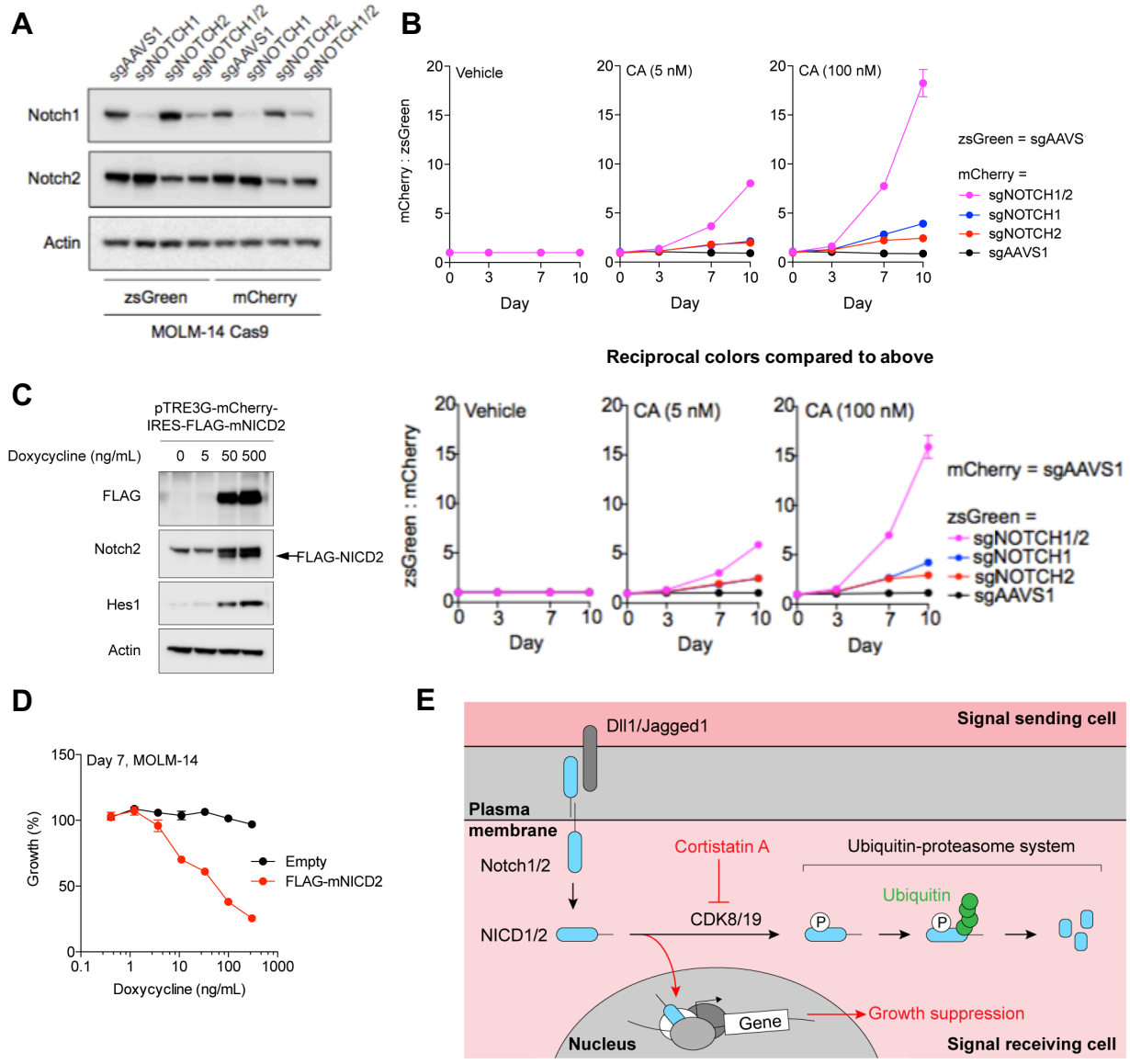


Figure 3.13 | Both Notch1 and Notch2 contribute to CA's activity in AML. (**A**) Western blots showing a decrease in target proteins when *NOTCH1*, *NOTCH2*, or both *NOTCH1/2* were knocked out in MOLM-14 cells labeled with either zsGreen or mCherry. (**B**) Competitive growth assays showing the change in mCherry-to-zsGreen (top) or zsGreen-to-mCherry (bottom) ratio upon CA treatment in MOLM-14. Each color is co-expressed with an sgRNA as annotated on the right. All values are normalized to vehicle treatment. (**C**) Western blot showing expression of FLAG-mNICD2 (mouse NICD2) and upregulation of Hes1 upon induction with doxycycline in

MOLM-14 (1 day). (**D**) Growth assay showing the tumor suppressor function of NICD2 in MOLM-14. (**E**) Proposed model for the mechanism of CA in MOLM-14. In AML cells that express Notch receptors and ligands, NICD1/2 are targeted for degradation. CA (red) inhibits this degradation, allowing NICD1/2 to accumulate and suppress growth.

Discussion

We previously reported that AML cell growth is dependent on CDK8/19 kinase activity (Pelish et al., 2015). Here, we combined CA with a CRISPR-Cas9 screen to identify suppression of Notch signaling as a mechanism for this dependence (**Figure 3.13E**). AML cells have to evade Notch tumor suppressor activity (Kannan et al., 2013; Klinakis et al., 2011; Lobry et al., 2013), and the majority of AML cell lines studied here appear to do so by transcriptional repression of Notch ligands (**Figure 4A**). MOLM-14 cells, however, express Dll1/Jagged1 that continually stimulate the production of NICD1/2, and hence CDK8/19 are required to degrade these NICDs (Fryer et al., 2004; Li et al., 2014) and maintain the cells in a proliferative state. Inhibition of this degradation by CA causes NICD1/2 to accumulate and arrest AML growth. Interestingly, the alkaloid *N*-methylhemeanthidine chloride (NMHC) was recently reported as another small-molecule activator of Notch signaling (Ye et al., 2016). Docking studies suggest that NMHC binds the negative regulatory region of Notch1 to promote its proteolytic cleavage into NICD1. Despite having distinct mechanisms, CA and NMHC suppress AML cell growth in a Notch-dependent manner (Ye et al., 2016), further suggesting that pharmacological activation of Notch is a possible therapeutic strategy in AML.

CDK8/19 regulate the stability of NICDs in both AML and T-ALL (Li et al., 2014), but the effect of this regulation on cancer growth is highly cell-type specific. CDK8/19 are deactivated by cyclin C deletion in many T-ALLs (Li et al., 2014), resulting in NICD1 stabilization that promotes T-ALL growth. On the other hand, CDK8/19 activity is required in AML cells such as MOLM-14 to suppress Notch signaling and maintain cell growth. The basis for AML's dependency on CDK8/19 varies across cell lines according to their cytogenetic background. MOLM-14 cells are uniquely dependent on NICD1/2 degradation by CDK8/19 because they express Notch ligands, whereas SET-2 and UKE-1 cells (post-MPN AML) rely on CDK8/19 to maintain STAT1-Ser727 phosphorylation (Nitulescu et al., 2017). The antiproliferative mechanism of CDK8/19 inhibition in other cells might involve yet other substrates among ~80 that have been identified to date (Poss et al., 2016). One way to investigate the functional relevance of these substrates is to mutate the acceptor Ser/Thr into Asp/Glu, creating a permanent phosphomimetic, and test whether it desensitizes AML cells to CA (as was done for STAT1-Ser727 (Nitulescu et al., 2017)). This approach is only suitable for cases where the list of substrates could be narrowed down to just a few candidates. CRISPR-Cas9 screening represents an alternative, unbiased approach to mechanism elucidation without needing any prior hypothesis.

Although Dll1/Jagged1 define the sensitivity of MOLM-14 to Notch activation by CA, neither *DLL1* nor *JAG1* appeared as a hit in our screen. We speculate that Dll1 and Jagged1 may play redundant functions and hence single-gene knockouts might not be sufficient to inhibit Notch signaling. In addition, the screen was carried out in a pooled format. *DLL1*- and *JAG1*-knockout cells could be activated in trans by other cells bearing intact Dll1/Jagged1, negating any desensitization effect that might be conferred by *DLL1* or *JAG1* knockout.

Based on our results, it is unclear whether the effects of CA on Notch signaling and super-enhancer activation (Pelish et al., 2015) are functionally related. Notch has been implicated in super-enhancer regulation T-ALL (Wang et al., 2014). Notch target genes in other cell types (Moellering et al., 2009; Trimarchi et al., 2014; Wang et al., 2014) showed little overlap with super-enhancer associated genes in MOLM-14(Pelish et al., 2015) (data not shown). While this observation suggests that Notch might not mediate the effect of CA on super-enhancer-associated genes in MOLM-14, the lack of correlation could also be due to confounding factors such as cell-lineage differences (T-ALL versus AML) and timing. NICD1 peaks >18 hours after CA treatment whereas the effect of CA on super-enhancer-associated genes was most pronounced within 3 hours.

Unbiased CRISPR-Cas9 screens could reveal novel proteins in a known biological process. For example, our screen outputs 4 hits (*GANAB*, *UXS1*, *TGDS*, and *UGCG*) that encode enzymes involved in sugar biosynthesis but have never been studied in the context of Notch signaling. Given that other Notch glycosylation proteins appeared as hits in our screen, we suspect that these 4 hits might also be important for Notch glycosylation. Similarly, we obtained ~20 hits belonging to the class of transcriptional regulators including PCAF/STAGA acetyltransferase complexes (**Figure S1G**), histone modifying enzymes, and chromatin remodelers. We speculate that some of these transcriptional regulators, like MAML1 and CREBBP/p300, are important for activating Notch-dependent transcription.

Oncogene inhibition is a common therapeutic approach to cancer. However, tumor suppressor activation (Pyndiah and Sakamuro, 2015; Smukste and Stockwell, 2003) is a more challenging and rare strategy. Our studies reveal that CDK8/19 inhibition in MOLM14 cells is an example of tumor suppressor activation. Given the many tumor suppressors that are inactivated in cancer, it is useful to discover new approaches to activate tumor suppressors as a therapeutic approach to cancer.

Conclusion

In the previous chapter, we showed that AML is dependent on the kinase activity of CDK8/19. To understand the molecular basis for this dependency, we performed a genome-wide CRISPR-Cas9 suppressor screen in an AML cell line sensitive to the CDK8/19 inhibitor CA. We found that, in addition to *NOTCH1* and *NOTCH2* themselves, knockout of 13 genes spanning all steps of the Notch pathway, including Notch glycosylation (*POFUT1*, *POGLUT1*, *GMDS*, *GXYLT1*, *RFNG*), proteolytic activation (*PSENEN*, *PSEN1*, *NCSTN*, *ADAM10*), transcriptional co-activation (*MAML1*, *CREBBP*, *EP300*), and ligand processing (*MIB1*), conferred resistance to CA, indicating that CA's antiproliferative activity requires a functional Notch pathway. CA arrests AML cell growth by inhibiting CDK8/19-mediated degradation of Notch1 and Notch2 intracellular domains (NICD1/2), which increased Notch signaling and its associated tumor suppressor activity. Our findings reveal why distinct AML cells are sensitive to CDK8/19 inhibition, and outline an approach to identify other determinants of AML sensitivity to CDK8/19 inhibitors.

Methods

Experimental model and subject details

All cells were grown at 37 °C in 5% CO_2 atmosphere, in RPMI 1640 GlutaMAX media (Life Technologies) supplemented with 100 U/ml penicillin and 100 μg/ml streptomycin. MOLM-13, MOLM-14, MOLM-16, MV4;11, RS4;11, HL-60, CMK-86, MEG-01, and EOL-1 were supplemented with 10% FBS; SKNO-1 with 10% FBS and 2 ng/ml GM-CSF; SET-2 with 20% FBS; and UKE-1 with 10% horse serum and 1 μM hydrocortisone. MV4;11, RS4;11, HL-60, and MEG-01 were obtained from ATCC; MOLM-13, MOLM-16, SKNO-1, and EOL-1 from DSMZ; CMK-86 from National Institute of Biomedical Innovation—JCRB Cell Bank; UKE-1 and SET-2 from Ross Levine; and MOLM-14 from Scott Armstrong.

Plasmids

pCas9 is LentiCas9-Blast (a gift from Feng Zhang; Addgene plasmid 52962). pCas9-zsGreen is identical to pCas9, except that sequences encoding zsGreen (amplified from pLVX-EF1□-IRES-zsGreen; Clontech 631982) were inserted between P2A and the blasticidin resistance gene, yielding a plasmid that co-expresses Cas9 and the zsGreen-blasticidin resistance fusion protein. Single-gene knockout plasmids (encoding sgNT, sgNOTCH1, sgPSEN1, sgMAML1, sgCREBBP, sgEP300, sgAAVS1, and sgNOTCH2) were prepared by ligating synthetic primers containing target sequences into BsmBI-digested LentiGuide-Puro plasmid (a gift from Feng Zhang; Addgene 52963). Double-gene knockout plasmids (encoding sgDLL1/JAG1 or sgNOTCH1/2) were prepared by first making the single knockout plasmids for all target genes. The region encompassing the U6 promoter and sgJAG1 (U6-JAG1) or U6-sgNOTCH2 was

amplified and inserted into LentiGuide-sgDLL1-Puro or LentiGuide-sgNOTCH1-Puro, respectively. Doxycycline-inducible plasmids were prepared by amplifying FLAG-NICD1 from TetO-FUW-NICD (a gift from Rudolf Jaenisch; Addgene 61540), FLAG-mNICD2 from 3XFlagNICD2 (a gift from Raphael Kopan, Addgene 20184), or Hes1 from EF.hHES1.Ubc.GFP (a gift from Linzhao Cheng, Addgene 17624) and ligating it into the Tet-On inducible vector pLVX-TRE3G-mCherry-IRES (Clontech 631352). The plasmid encoding doxycycline-controlled transactivator is pLVX-EF1□-Tet3G (Clontech 631352). The MigR1 plasmid and the plasmid expression HES1-ER were a gift from Patrick Zweidler-McKay. Fluorescent proteins used in two-color experiments were encoded by pLVX-EF1□-IRES-zsGreen (Clontech 631982) and pLVX-EF1□-IRES-mCherry (Clontech 631987).

Lentiviral transduction

Except for the Avana pooled sgRNA library, all other plasmids were packaged into lentivirus at the University of Massachusetts RNAi core facility by co-transfecting 293T cells with psPASx and pMD2.G (Addgene). After 48 h, viral supernatants were collected and passed through a 0.45 μm filter (Millipore). Transduction was achieved either by spinning 500,000 cells in 500 μL of viral supernatants with 8 μg/mL polybrene (Fisher TR1003G) at 2,000 xg for 1.5 hours at room temperature, or by retronectin according to manufacturer's instructions (Takara T100A).

Western blot

Primary antibodies include Notch1 (Cell Signaling Technologies (CST) 3608), NICD1 (CST 4147), Notch2 (CST 5732), Psen1 (CST 5643), MAML1 (CST 12166), CREBBP (Bethyl A399-

363A), Hes1 (CST 11988), Dll1 (CST 2588), Jagged1 (CST 2620), DMAP1 (CST 19115), Vinculin (CST 4650), Actin (CST 5125), and FLAG (Sigma F1804). Secondary antibodies include anti-rabbit IgG HRP conjugate (Promega 401B) or anti-mouse IgG HRP conjugate (Promega 402B). Protein samples were heated to 80 °C for 10 minutes in NuPAGE LDS sample buffer (NP0007) and 2.5% β-mercaptoethanol, resolved on NuPAGE 4-12% Bis-Tris gels (Thermo NP0322) in NuPAGE MOPS buffer (Thermo NP0001) or 3-8% Tris-Acetate gels (Thermo EA03752) in NuPAGE Tris-Acetate buffer (Thermo LA0041), transferred onto PVDF membranes (Millipore IPVH07850) in transfer buffer (Boston Bioproducts BP-190-4) with 10-20% methanol, and blocked with 5% BSA in 0.1% TBST (TBST-BSA) for 1 hour at room temperature. Target proteins were probed with primary antibodies diluted in TBST-BSA overnight at 4 °C; if applicable, blots were washed 5 times with TBST-BSA and probed with secondary antibodies diluted in TBST for 1 hour at room temperature. Target protein bands were visualized by ECL Western Blotting Substrate (Thermo 32106) or SuperSignal West Pico Chemiluminescent Substrate (Thermo 34080), and detected by the Azure c300 camera or by X-ray films. Images were quantified using ImageJ.

Notch stabilization assays

MOLM-14 cells (1 mL at 1×10^6 cells/mL) in a 24-well plate were treated with CA in growth media at indicated concentrations for 18 hours (**Figure 2a, Figure 5b**), or with 100 nM CA for indicated durations (**Figure 2b**), or with 100 nM CA and/or 10 μM MG132 (Millipore 474787) for 4 hours (**Fig 2c**), or with 100 nM CA and 100 nM compound E (Millipore 565790) for 18 hours (**Figure 5b**). Cells were harvested, washed once with ice-cold PBS, lysed in 100 μL RIPA (Sigma R0278 or Boston Bioproducts BP-115) supplemented with protease and phosphatase inhibitors

(either Sigma P8340, P0044, and P5726, or Roche 4693124001 and 4906845001), and analyzed by western blot. For EDTA-stimulated Notch activation (**Figures S4b-4c**), cells were treated with 5 mM EDTA in PBS for 15 minutes at room temperature, washed once with PBS, and then incubated with MG132 and/or CA in growth media before harvesting and lysis. For nuclear extraction (**Figures S4c-d**), cells were treated as described above, and then subjected to nuclear fractionation using the Nuclear Extraction Kit (Cayman 10009277) before analysis by western blot.

Lambda protein phosphatase assays

MOLM-14 cells (10 mL at 1×10^6 cells/mL) were treated with 100 nM CA for 24 hours or 10 μM MG132 for 4 hours and lysed with 500 μL RIPA with protease inhibitors (Sigma P8340 or Roche 4693124001) without phosphatase inhibitors. Half of the lysate was subjected to lambda protein phosphatase treatment (New England Biolabs P0753). The reaction was quenched after 1 hour by addition of EDTA and phosphatase inhibitors (Sigma P0044 and P5726) before analysis by western blot.

Flow cytometry

Data were acquired on LSR II or LSRFortessa (BD Biosciences) and analyzed using FlowJo. Fluorescence-activated cell sorting (FACS) was performed by the Bauer Core Facility at Harvard University.

Genome-wide CRISPR-Cas9 suppressor screens

MOLM-14 cells stably expressing Cas9 were obtained by transduction with pCas9-zsGreen followed by 7-day selection with blasticidin and sorting for zsGreen expression. The cells were then

transduced in triplicate with the Avana sgRNA library (Broad institute) at multiplicity of infection = 0.3 by centrifugation at 2,000 xg for 1.5 hours at room temperature in the presence of 8 μg/mL polybrene (in 12-well plates, 1.5 mL per well at 2×10^6 cells/mL, for a total of 1.5×10^8 cells per replicate to achieve coverage of ~500x cells per sgRNA). Cells were transferred to fresh media at 1×10^6 cells/mL, incubated in T175 flasks overnight, and selected with puromycin for 7 days. Cells were then treated with vehicle or CA (100 nM) and maintained at a density below 2.5×10^6 cells/mL for a total cell number exceeding 1×10^8 to ensure >500x coverage by passaging every 3-4 days for the next 15 days. On day 15, 1.5×10^8 cells were withdrawn from each treatment, and genomic DNA was extracted using the QIAamp DNA Blood Maxi kit (Qiagen 51192). Viral DNAs encoding sgRNAs were amplified from genomic DNAs (240 μg, corresponding to ~500x coverage) using P5 and P7 primers and sequenced using Illumina HiSeq. Library preparation and sequencing was performed by the Broad Institute. The change in abundance of each sgRNA in CA- versus DMSO-treated samples was calculated by subtracting $\log_2$(read) in a CA-treated sample by $\log_2$(read) in a DMSO-treated sample from the same replicate, and averaging this difference over three replicates (called enrichment score, or the average $\log_2$(fold-change), in **Figure 1c**). Sequences of sgRNAs were mapped to target genes and the target genes were ranked based on false discovery rates by the STARS software (Broad Institute). Gene lists were submitted to the Gene Ontology Consortium (www.geneontology.org) for functional annotation.

CRISPR-Cas9 gene editing

MOLM-14 cells stably expressing Cas9 were obtained by transduction with pCas9 followed by 7-day selection with blasticidin. Cells were then transduced with LentiGuide-Puro constructs encoding sgRNAs specific to the target genes and selected for 7 days with puromycin before use in subsequent experiments. Indel frequencies were quantified by T7E1 assays as follows. Genomic DNA was extracted using the DNeasy Blood & Tissue Kit (Qiagen 69504). The region surrounding the target cut site was amplified from genomic DNA, treated with T7 endonuclease I (New England Biolabs M0320), and resolved on 0.6% agarose gel with 0.5 μg/ml ethidium bromide. Bands corresponding to cleaved and uncleaved amplicons were quantified using ImageJ, and percent indel = 100% × (1-(1-fraction cleaved)$^{1/2}$). Monoclonal knockout cells were isolated by single-cell sorting into V-bottom 96-well plates using flow cytometry and expanded until they reached sufficient numbers for subsequent experiments (~1 month).

Cell growth assays

Cell growth assays were performed as described in Pelish et al. (Pelish et al., 2015) Briefly, cells were plated in flat-bottom 96-well plates in duplicate or triplicate and treated with compounds (20,000 cells in total 150 μL media per well). On days 3, 7, and 10 (if applicable), cells were diluted to prevent overconfluency using the following procedure. Cells from a control no-treatment well were counted, and the volume required to re-plate 15,000 – 20,000 cells from this well was calculated. The same volume of cell suspensions from each well was re-plated in fresh media with compounds. To measure cell viability on days 3, 7, and 10 (if applicable), cell suspensions were diluted three-fold in media, and 20 μL of the suspensions were mixed with 20 μL of CellTiter-Glo

Luminescent Cell Viability Assay reagent (Promega G7570) in white, flat-bottom 384-well plates. Cells from a control well were counted and serially diluted before testing for viability, to construct a standard curve correlating cell numbers to luminescence. Luminescence was quantified using SPECTRAmax M3 (Molecular Devices), and cell numbers were calculated based on the standard curve. Luminescence of vehicle-treated cells represented 100% growth. For compound E desensitization assays, cells were co-treated with compound E (EMD Millipore 565790). For doxycycline-inducible assays, cells expressing pLVX-EF1□-Tet3G and pTRE3G-mCherry-IRES-target gene were selected with a mixture of geneticin (Millipore 345812) and puromycin (Sigma) before use in growth assays, and expression of target gene was induced with doxycycline. For tamoxifen-inducible assays, cells expressing HES1-ER were sorted for green fluorescence before use in growth assays, and expression of Hes1 was induced with (*Z*)-4-hydroxytamoxifen (Sigma H7904).

Competitive growth assays

Cells expressing desired plasmids were transduced with pLVX-IRES-zsGreen (Clontech 631982) or pLVX-IRES-mCherry (Clontech 631987) and sorted by FACS for fluorescence protein expression. On day 0, cells were combined in a 1:1 ratio such that the final density is 200,000 cells/mL, plated in flat-bottom 96-well plates in duplicate, and treated with compounds (total 150 μL media per well). On days 3, 7, 10, and 14, cells were diluted 5-10 times and re-plated with fresh compounds to avoid overconfluency, and cell suspensions were analyzed by flow cytometry.

Gene set enrichment analysis

Ranked gene list of MOLM-14 cells treated with 10 nM CA for 24 hours from was obtained from Pelish Pelish et al. (Pelish et al., 2015) Gene sets were obtained from published gene expression data to a total of 34 gene sets, including the ones presented in **Figure 2e** (Moellering et al., 2009; Trimarchi et al., 2014; Wang et al., 2014). When these gene sets were combined with the Broad Institute's MSigDB C2 curated gene set (total 3747 gene sets), their normalized enrichment scores and false discovery rates remained comparable. Gene set enrichment analysis was performed on http://software.broadinstitute.org/gsea/index.jsp (Broad Institute).

Real-time polymerase chain reaction

Total RNA was isolated from ~1 × 10^6 cells using RNeasy Plus Mini Kit (Qiagen 74134), reversed transcribed into cDNA using High Capacity cDNA Reverse Transcription Kit (Thermo 4368814), and used for qPCR with TaqMan probes for genes of interest (FAM) and for GAPDH as a reference gene (VIC) on CFX96 Real-Time PCR Detection System (Biorad). Probes were obtained from ThermoFisher, and target *HES1* (Hs00172878_m1), *DLL1* (Hs00194509_m1), *DLL3* (Hs01085096_m1), *DLL4* (Hs00184092_m1), *JAG1* (Hs01070032_m1), and *JAG2* (Hs00171432_m1), and *GAPDH* control mix (4325792).

Quantitation and statistical analysis

STARS, gene ontology analysis, and gene set enrichment analysis have a built-in function that calculates statistical significance as *p*-values or false discovery rates (FDR). For STARS, we defined FDR < 0.01 as significant. For growth assays, western blot quantification, and qRT-PCR,

data were analyzed using GraphPad Prism and, when applicable, presented as mean ± standard error of the mean (S.E.M.).

References

Acar, M., Jafar-Nejad, H., Takeuchi, H., Rajan, A., Ibrani, D., Rana, N.A., Pan, H., Haltiwanger, R.S., and Bellen, H.J. (2008). Rumi Is a CAP10 Domain Glycosyltransferase that Modifies Notch and Is Required for Notch Signaling. Cell *132*, 247–258.

Agrotis, A., and Ketteler, R. (2015). A new age in functional genomics using CRISPR/Cas9 in arrayed library screening. Front. Genet. *6*.

Alarcón, C., Zaromytidou, A.-I., Xi, Q., Gao, S., Yu, J., Fujisawa, S., Barlas, A., Miller, A.N., Manova-Todorova, K., Macias, M.J., et al. (2009). Nuclear CDKs Drive Smad Transcriptional Activation and Turnover in BMP and TGF-β Pathways. Cell *139*, 757–769.

Allen, B.L., and Taatjes, D.J. (2015). The Mediator complex: a central integrator of transcription. Nat. Rev. Mol. Cell Biol. *16*, 155.

Andersson, E.R., Sandberg, R., and Lendahl, U. (2011). Notch signaling: simplicity in design, versatility in function. Development *138*, 3593–3612.

Bancerek, J., Poss, Z.C., Steinparzer, I., Sedlyarov, V., Pfaffenwimmer, T., Mikulic, I., Dölken, L., Strobl, B., Müller, M., Taatjes, D.J., et al. (2013). CDK8 Kinase Phosphorylates Transcription Factor STAT1 to Selectively Regulate the Interferon Response. Immunity *38*, 250–262.

Borggrefe, T., Lauth, M., Zwijsen, A., Huylebroeck, D., Oswald, F., and Giaimo, B.D. (2016). The Notch intracellular domain integrates signals from Wnt, Hedgehog, TGFβ/BMP and hypoxia pathways. Biochim. Biophys. Acta BBA - Mol. Cell Res. *1863*, 303–313.

The Cancer Genome Atlast Network (2013). Genomic and Epigenomic Landscapes of Adult De Novo Acute Myeloid Leukemia. N. Engl. J. Med. *368*, 2059–2074.

Cee, V.J., Chen, D.Y.-K., Lee, M.R., and Nicolaou, K.C. (2009). Cortistatin A is a High-Affinity Ligand of Protein Kinases ROCK, CDK8, and CDK11. Angew. Chem. Int. Ed. *48*, 8952–8957.

Dawson, M.A., Prinjha, R.K., Dittmann, A., Giotopoulos, G., Bantscheff, M., Chan, W.-I., Robson, S.C., Chung, C., Hopf, C., Savitski, M.M., et al. (2011). Inhibition of BET recruitment to chromatin as an effective treatment for MLL-fusion leukaemia. Nature *478*, 529.

Doench, J.G. (2017). Am I ready for CRISPR? A user's guide to genetic screens. Nat. Rev. Genet. *19*, 67–80.

Doench, J.G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E.W., Donovan, K.F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., et al. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat. Biotechnol. *34*, 184.

Doudna, J.A., and Charpentier, E. (2014). The new frontier of genome engineering with CRISPR-Cas9. Science *346*, 1258096–1258096.

Fryer, C.J., White, J.B., and Jones, K.A. (2004). Mastermind Recruits CycC:CDK8 to Phosphorylate the Notch ICD and Coordinate Activation with Turnover. Mol. Cell *16*, 509–520.

Hartenian, E., and Doench, J.G. (2015). Genetic screens and functional genomics using CRISPR/Cas9 technology. FEBS J. *282*, 1383–1393.

Hartmann, D., de Strooper, B., Serneels, L., Craessaerts, K., Herreman, A., Annaert, W., Umans, L., Lübke, T., Lena Illert, A., von Figura, K., et al. (2002). The disintegrin/metalloprotease ADAM 10 is essential for Notch signalling but not for α-secretase activity in fibroblasts. Hum. Mol. Genet. *11*, 2615–2624.

Hicks, C., Johnston, S.H., diSibio, G., Collazo, A., Vogt, T.F., and Weinmaster, G. (2000). Fringe differentially modulates Jagged1 and Delta1 signalling through Notch1 and Notch2. Nat. Cell Biol. *2*, 515.

Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-André, V., Sigova, A.A., Hoke, H.A., and Young, R.A. (2013). Super-Enhancers in the Control of Cell Identity and Disease. Cell *155*, 934–947.

Hsu, P.D., Lander, E.S., and Zhang, F. (2014). Development and Applications of CRISPR-Cas9 for Genome Engineering. Cell *157*, 1262–1278.

Ilagan, M.X.G., and Kopan, R. (2007). SnapShot: Notch Signaling Pathway. Cell *128*, 1246.e1-1246.e2.

Itoh, M., Kim, C.-H., Palardy, G., Oda, T., Jiang, Y.-J., Maust, D., Yeo, S.-Y., Lorick, K., Wright, G.J., Ariza-McNaughton, L., et al. (2003). Mind Bomb Is a Ubiquitin Ligase that Is Essential for Efficient Activation of Notch Signaling by Delta. Dev. Cell *4*, 67–82.

Kannan, S., Sutphin, R.M., Hall, M.G., Golfman, L.S., Fang, W., Nolo, R.M., Akers, L.J., Hammitt, R.A., McMurray, J.S., Kornblau, S.M., et al. (2013). Notch activation inhibits AML growth and survival: a potential therapeutic approach. J. Exp. Med. *210*, 321–337.

Kim, H., and Kim, J.-S. (2014). A guide to genome engineering with programmable nucleases. Nat. Rev. Genet. *15*, 321–334.

Klinakis, A., Lobry, C., Abdel-Wahab, O., Oh, P., Haeno, H., Buonamici, S., Walle, I. van D., Cathelin, S., Trimarchi, T., Araldi, E., et al. (2011). A novel tumour-suppressor function for the Notch pathway in myeloid leukaemia. Nature *473*, 230.

Kopan, R., and Ilagan, M.X.G. (2009). The Canonical Notch Signaling Pathway: Unfolding the Activation Mechanism. Cell *137*, 216–233.

Li, N., Fassl, A., Chick, J., Inuzuka, H., Li, X., Mansour, M.R., Liu, L., Wang, H., King, B., Shaik, S., et al. (2014). Cyclin C is a haploinsufficient tumour suppressor. Nat. Cell Biol. *16*, 1080.

Lobry, C., Ntziachristos, P., Ndiaye-Lobry, D., Oh, P., Cimmino, L., Zhu, N., Araldi, E., Hu, W., Freund, J., Abdel-Wahab, O., et al. (2013). Notch pathway activation targets AML-initiating cell homeostasis and differentiation. J. Exp. Med. *210*, 301–319.

McElhinny, A.S., Li, J.-L., and Wu, L. (2008). Mastermind-like transcriptional co-activators: emerging roles in regulating cross talk among multiple signaling pathways. Oncogene *27*, 5138.

Mertz, J.A., Conery, A.R., Bryant, B.M., Sandy, P., Balasubramanian, S., Mele, D.A., Bergeron, L., and Sims, R.J. (2011). Targeting MYC dependence in cancer by inhibiting BET bromodomains. Proc. Natl. Acad. Sci. *108*, 16669–16674.

Moellering, R.E., Cornejo, M., Davis, T.N., Bianco, C.D., Aster, J.C., Blacklow, S.C., Kung, A.L., Gilliland, D.G., Verdine, G.L., and Bradner, J.E. (2009). Direct inhibition of the NOTCH transcription factor complex. Nature *462*, 182.

Nitulescu, I.I., Meyer, S.C., Wen, Q.J., Crispino, J.D., Lemieux, M.E., Levine, R.L., Pelish, H.E., and Shair, M.D. (2017). Mediator Kinase Phosphorylation of STAT1 S727 Promotes Growth of Neoplasms With JAK-STAT Activation. EBioMedicine *26*, 112–125.

Oswald, F., Täuber, B., Dobner, T., Bourteele, S., Kostezka, U., Adler, G., Liptay, S., and Schmid, R.M. (2001). p300 Acts as a Transcriptional Coactivator for Mammalian Notch-1. Mol. Cell. Biol. *21*, 7761–7774.

Panin, V.M., Papayannopoulos, V., Wilson, R., and Irvine, K.D. (1997). Fringe modulates Notch–ligand interactions. Nature *387*, 908.

Papaemmanuil, E., Gerstung, M., Bullinger, L., Gaidzik, V.I., Paschka, P., Roberts, N.D., Potter, N.E., Heuser, M., Thol, F., Bolli, N., et al. (2016). Genomic Classification and Prognosis in Acute Myeloid Leukemia. N. Engl. J. Med. *374*, 2209–2221.

Pelish, H.E., Liau, B.B., Nitulescu, I.I., Tangpeerachaikul, A., Poss, Z.C., Silva, D.H.D., Caruso, B.T., Arefolov, A., Fadeyi, O., Christie, A.L., et al. (2015). Mediator kinase inhibition further activates super-enhancer-associated genes in AML. Nature *526*, 273.

Poss, Z.C., Ebmeier, C.C., Odell, A.T., Tangpeerachaikul, A., Lee, T., Pelish, H.E., Shair, M.D., Dowell, R.D., Old, W.M., and Taatjes, D.J. (2016). Identification of Mediator Kinase Substrates in Human Cells using Cortistatin A and Quantitative Phosphoproteomics. Cell Rep. *15*, 436–450.

Pyndiah, S., and Sakamuro, D. (2015). Restoration of tumor suppressor functions by small-molecule inhibitors. Mol. Cell. Oncol. *2*, e991225.

Rand, M.D., Grimm, L.M., Artavanis-Tsakonas, S., Patriub, V., Blacklow, S.C., Sklar, J., and Aster, J.C. (2000). Calcium Depletion Dissociates and Activates Heterodimeric Notch Receptors. Mol. Cell. Biol. *20*, 1825–1835.

Ren, Y., Brown, V., Hu, S., Lopez, J., Miljovska, S., Schmidt, D., Bradley, M., Sprott, K., Olson, E., Fritz, C.C., et al. (2015). Targeting Transcriptional Dependency in Acute Myeloid Leukemia (AML) with a Covalent Inhibitor of Transcriptional Kinase CDK7. Blood *126*, 1354–1354.

Rzymski, T., Mikula, M., Żyłkiewicz, E., Dreas, A., Wiklik, K., Gołas, A., Wójcik, K., Masiejczyk, M., Wróbel, A., Dolata, I., et al. (2017). SEL120-34A is a novel CDK8 inhibitor active in AML cells with high levels of serine phosphorylation of STAT1 and STAT5 transactivation domains. Oncotarget *8*, 33779–33795.

Schenk, T., Chen, W.C., Göllner, S., Howell, L., Jin, L., Hebestreit, K., Klein, H.-U., Popescu, A.C., Burnett, A., Mills, K., et al. (2012). Inhibition of the LSD1 (KDM1A) demethylase reactivates the all-trans-retinoic acid differentiation pathway in acute myeloid leukemia. Nat. Med. *18*, 605.

Sethi, M.K., Buettner, F.F.R., Krylov, V.B., Takeuchi, H., Nifantiev, N.E., Haltiwanger, R.S., Gerardy-Schahn, R., and Bakker, H. (2010). Identification of Glycosyltransferase 8 Family Members as Xylosyltransferases Acting on O-Glucosylated Notch Epidermal Growth Factor Repeats. J. Biol. Chem. *285*, 1582–1586.

Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelsen, T.S., Heckl, D., Ebert, B.L., Root, D.E., Doench, J.G., et al. (2014). Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. Science *343*, 84–87.

Shalem, O., Sanjana, N.E., and Zhang, F. (2015). High-throughput functional genomics using CRISPR–Cas9. Nat. Rev. Genet. *16*, 299–311.

Shi, S., and Stanley, P. (2003). Protein O-fucosyltransferase 1 is an essential component of Notch signaling pathways. Proc. Natl. Acad. Sci. *100*, 5234–5239.

Shih, I.-M., and Wang, T.-L. (2007). Notch Signaling, γ-Secretase Inhibitors, and Cancer Therapy. Cancer Res. *67*, 1879–1882.

Smukste, I., and Stockwell, B.R. (2003). Restoring functions of tumor suppressors with small molecules. Cancer Cell *4*, 419–420.

Stanley, P. (2007). Regulation of Notch signaling by glycosylation. Curr. Opin. Struct. Biol. *17*, 530–535.

Trimarchi, T., Bilal, E., Ntziachristos, P., Fabbri, G., Dalla-Favera, R., Tsirigos, A., and Aifantis, I. (2014). Genome-wide Mapping and Characterization of Notch-Regulated Long Noncoding RNAs in Acute Leukemia. Cell *158*, 593–606.

Wang, F., Travins, J., DeLaBarre, B., Penard-Lacronique, V., Schalm, S., Hansen, E., Straley, K., Kernytsky, A., Liu, W., Gliser, C., et al. (2013a). Targeted Inhibition of Mutant IDH2 in Leukemia Cells Induces Cellular Differentiation. Science *340*, 622–626.

Wang, H., Zang, C., Taing, L., Arnett, K.L., Wong, Y.J., Pear, W.S., Blacklow, S.C., Liu, X.S., and Aster, J.C. (2014). NOTCH1–RBPJ complexes drive target gene expression through dynamic interactions with superenhancers. Proc. Natl. Acad. Sci. *111*, 705–710.

Wang, T., Wei, J.J., Sabatini, D.M., and Lander, E.S. (2013b). Genetic Screens in Human Cells Using the CRISPR/Cas9 System. Science *1246981*.

Weng, A.P., Ferrando, A.A., Lee, W., Morris, J.P., Silverman, L.B., Sanchez-Irizarry, C., Blacklow, S.C., Look, A.T., and Aster, J.C. (2004). Activating Mutations of NOTCH1 in Human T Cell Acute Lymphoblastic Leukemia. Science *306*, 269–271.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes. Cell *153*, 307–319.

Wu, L., Aster, J.C., Blacklow, S.C., Lake, R., Artavanis-Tsakonas, S., and Griffin, J.D. (2000). MAML1, a human homologue of Drosophila Mastermind, is a transcriptional co-activator for NOTCH receptors. Nat. Genet. *26*, 484.

Ye, Q., Jiang, J., Zhan, G., Yan, W., Huang, L., Hu, Y., Su, H., Tong, Q., Yue, M., Li, H., et al. (2016). Small molecule activation of NOTCH signaling inhibits acute myeloid leukemia. Sci. Rep. *6*, 26510.

Zhao, J., Ramos, R., and Demma, M. (2013). CDK8 regulates E2F1 transcriptional activity through S375 phosphorylation. Oncogene *32*, 3520.

Zhao, X., Feng, D., Wang, Q., Abdulla, A., Xie, X.-J., Zhou, J., Sun, Y., Yang, E.S., Liu, L.-P., Vaitheesvaran, B., et al. (2012). Regulation of lipogenesis by cyclin-dependent kinase 8–mediated control of SREBP-1. J. Clin. Invest. *122*, 2417–2427.

Zuber, J., Shi, J., Wang, E., Rappaport, A.R., Herrmann, H., Sison, E.A., Magoon, D., Qi, J., Blatt, K., Wunderlich, M., et al. (2011). RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia. Nature *478*, 524.

Chapter 4

Identification of CDK8 and CDK19 Substrates

Part of this chapter was adopted with modifications from our manuscript published in Cell Reports 15*:2, 436-450 (2016).*

Contributors

Zachary C. Poss, Christopher C. Ebmeier, Aaron T. Odell, Anupong Tangpeerachaikul, Thomas Lee, Henry E. Pelish, Matthew D. Shair, Robin D. Dowell, William M. Old, and Dylan J. Taatjes

Introduction

Prior to the development of CA as a selective inhibitor for CDK8/19, only very few substrates of CDK8/19 were known. Most of these are transcription factors (E2F1, NICD1, Smads, STAT1, and SREBP-c; reviewed in **Chapter 2**,). Phosphorylation substrates are the actual downstream effectors of CDK8/19 biology, and they explain why some AML are dependent on CDK8/19 kinase activity. As illustrated in **Chapter 3**, the antiproliferative effect of CDK8/19 inhibition in some AML cells could be explained by the loss of phosphorylation on STAT1 (in the case of SET-2 and UKE-1 cells) or NICD1 (in the case of MOLM-14 cells). The current lack of a more comprehensive list of CDK8/19 substrates calls for an alternative approach that could identify these unknown substrates in a high-throughput manner. In this chapter, we describe how we used CA as a fast-acting and selective CDK8/19 inhibitor in combination with mass spectrometry to uncover >60 proteins whose phosphorylation state changes upon CA treatment, suggesting that they were putative phosphorylation substrates of CDK8/19. We found that the majority of these substrates are chromatin-associated proteins and therefore may have regulatory roles in transcription.

Phosphoproteomics

Phosphorylation is one of the most important post-translational modifications in the cell. Enzymes called protein kinases catalyze the transfer of the phosphoryl group from ATP onto the acceptor hydroxyl group on the serine, threonine, or tyrosine residue of the target protein, whereas protein phosphatases catalyze the reverse hydrolysis reaction. Phosphorylation often acts as a new recognition motif for recruiting other protein factors to modulate a biochemical signaling cascade,

and may also act as a docking site for ubiquitin ligases to terminate signaling through protein degradation. For any given protein, the fraction that exists in the phosphorylated state could be very low (<1%). Mass spectrometry has emerged as the primary tool for identifying novel protein phosphorylation events owing to its high throughput and superb sensitivity. The study of phosphorylated proteins has come to be known as phosphoproteomics.

A typical phosphoproteomics experiment mirrors conventional proteomics experiments: proteins are first digested into short peptides before analysis by mass spectrometry (**Figure 4.1**). The most widely used protease is trypsin, which cleaves C-terminal to lysine and arginine, ensuring that the resulting peptide has a charge necessary for mass spectrometric analysis.

However, because the abundance of phosphoproteins is low, they need to be enriched using a variety of methods. Immobilized metal affinity chromatography (IMAC) and metal oxide affinity chromatography (MOAC) exploit the affinity of the phosphate group for metal ions, most commonly Fe^{3+} and Ti^{4+}. Strong cation exchange chromatography (SCX) makes use of the lower overall positive charge on a phosphorylated peptide for separation. Hydrophilic interaction chromatography (HILIC) separates phosphopeptides based on their increased hydrophilicity, and electrostatic repulsion-hydrophilic interaction chromatography (ERLIC) combines HILIC with an ionic surface to further purify the uniquely charged phosphopeptides. (Macek et al., 2009; Riley and Coon, 2016)

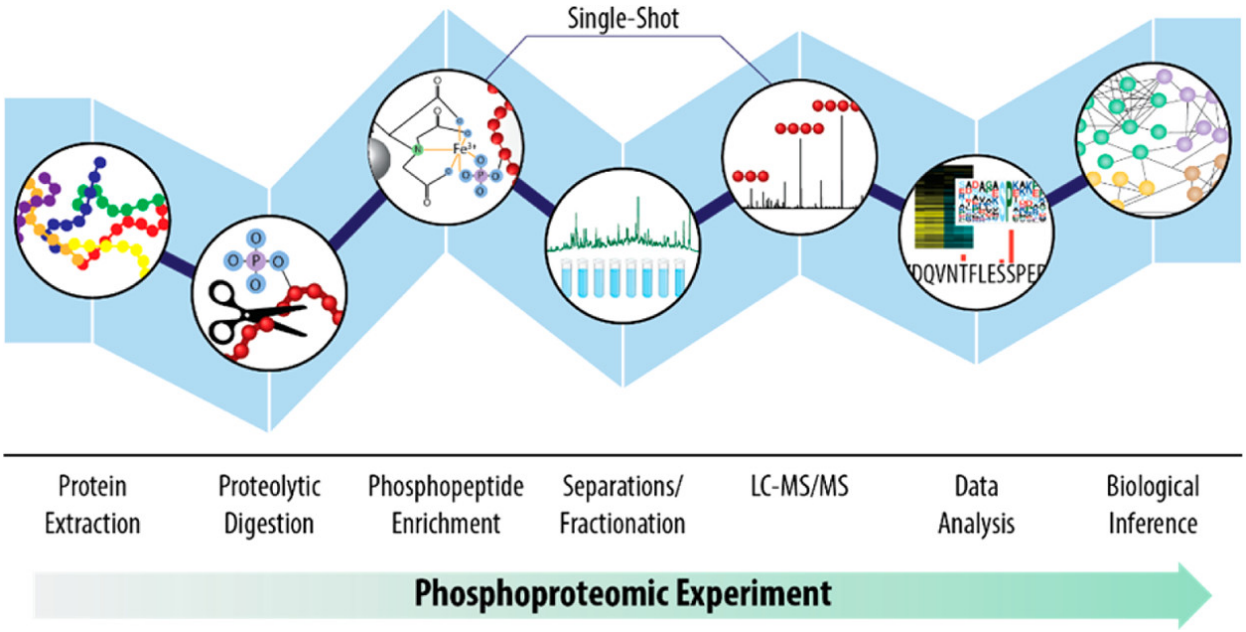


Figure 4.1 | Phosphoproteomics experiment workflow. Image reproduced from (Riley and Coon, 2016).

To identify phosphorylation substrates of a kinase, researchers would need a means of deactivating the kinase and examining the change in protein phosphorylation. Proteins that become differentially phosphorylated after kinase deactivation are putative substrates for that kinase. Further evidence, such as *in vitro* or in-cell kinase assays, would be required for validation. CA is a potent, selective, and fast-acting inhibitor of CDK8/19, and thus it is an appropriate tool for studying the phosphoproteomics of CDK8/19.

Using CA to identify phosphorylation substrates of CDK8/19

An important step toward understanding the cellular function of kinases is to identify the substrates that they phosphorylate. For many kinases, this first step has remained a challenge, in part because it is difficult to develop highly potent and selective kinase inhibitors.

The human CDK8 kinase exists in a 600 kDa complex known as the CDK8 module, which consists of four proteins (CDK8, cyclin C, MED12, MED13). The CDK8 module associates with regulatory loci on a genome-wide scale (Kagey et al., 2010; Pelish et al., 2015), and global targeting of the CDK8 module appears to reflect its association with Mediator (Allen and Taatjes, 2015). CDK19, a paralog of CDK8, emerged in vertebrates and has high sequence similarity to CDK8, including nearly identical cyclin binding and kinase domains. Comparatively little is known about CDK19; however, it appears to assemble into an analogous CDK19 module in human cells (Daniels et al., 2013).

Based upon their association with Mediator—a global regulator of RNA Pol II transcription—CDK8 or CDK19 may broadly impact gene expression patterns; however, physical knockdown of CDK8 or CDK19 protein levels had relatively modest effects in HCT116 cells, with 2-fold or greater changes in expression of several hundred genes (Donner et al., 2010; Galbraith et al., 2013). Whereas knockdown studies do not address the role of the kinase activity per se, these data suggested limited roles for the Mediator kinases in transcriptional regulation. In agreement, gene expression analyses with the CDK8 ortholog in yeast, Srb10 (Holstege et al., 1998), revealed that about 3% of genes were regulated by Srb10 kinase activity. Similarly, limited effects on yeast transcription were observed *in vitro* and *in vivo* upon selective inhibition of Srb10 (CDK8) kinase activity using an analog-sensitive mutant (Liu et al., 2004). Most genes affected by kinase-inactive mutant Srb10 (CDK8) were involved in cellular response to nutrient stress (Holstege et al., 1998).

The biological roles of human CDK8 and CDK19 remain poorly understood, in part, because a more comprehensive identification of their substrates or the genes specifically regulated

by their activities has been lacking. Our recent studies showed that CA is a potent and highly selective inhibitor of the Mediator kinases CDK8 and CDK19 (Pelish et al., 2015). CA binds the CDK8–cyclin C dimer with sub-nanomolar affinity (K_d = 195 pM) and two distinct kinome profiling assays, which collectively probed approximately 400 kinases, ultimately confirmed only CDK8 and CDK19 as targets of CA, even with analyses completed at 100-times the measured IC_{50} for CDK8 (Pelish et al., 2015). Given these and other data showing the unusual selectivity of CA, we could begin to probe the cellular function and targets of CDK8 and CDK19.

Here, we report the large-scale identification of Mediator kinase (CDK8 and CDK19) substrates in human cells, using SILAC-based phosphoproteomics. We couple these results with global analysis of gene expression changes (RNA-Seq) that result from targeted inhibition of Mediator kinase activity. Furthermore, we assess potential Mediator kinase effects on protein turnover using quantitative proteomic analyses across 6 time points spanning 24 hours of Mediator kinase inhibition. HCT116 cells were chosen for this study for several reasons. First, although CA potently inhibits Mediator kinase activity in HCT116 cells (Pelish et al., 2015), proliferation is not affected. This eliminated potential confounding effects, such as induction of cell cycle arrest or death, which could have complicated our analyses. Second, CDK8 is a colon cancer oncogene that was uncovered, in part, by an shRNA screen in HCT116 cells (Firestein et al., 2008). Third, published gene expression data exist in HCT116 cells with stable CDK8 or CDK19 knockdown (Donner et al., 2010; Galbraith et al., 2013), which allowed us to directly compare and de-couple the effects of subunit knockdown versus targeted inhibition of kinase activity.

Quantitative phosphoproteomics in HCT116 cells ± CA

To identify cellular CDK8 and CDK19 substrates, we used stable isotope labeling of amino acids in cell culture (SILAC) coupled with a phosphoproteomics workflow. Experiments were completed in HCT116 cells supplemented with heavy (Arg10, Lys8) or light (Arg0, Lys0) amino acids. Control (DMSO) and CA-treated cells were harvested and mixed 1:1 based on total protein content. Phosphopeptides were isolated using titanium enrichment, followed by offline electrostatic repulsion hydrophilic interaction chromatography (ERLIC) with LC-MS/MS for phosphosite identification (**Figure 4.2A**). We collected 24 fractions during ERLIC fractionation, with an average phosphopeptide enrichment of over 50% in biological triplicate experiments (**Figure 4.2B**). In total, over 16,000 heavy-light (H/L) phosphosite ratios were quantified and over 12,000 were present in at least two biological replicates (**Figure 4.2C**).

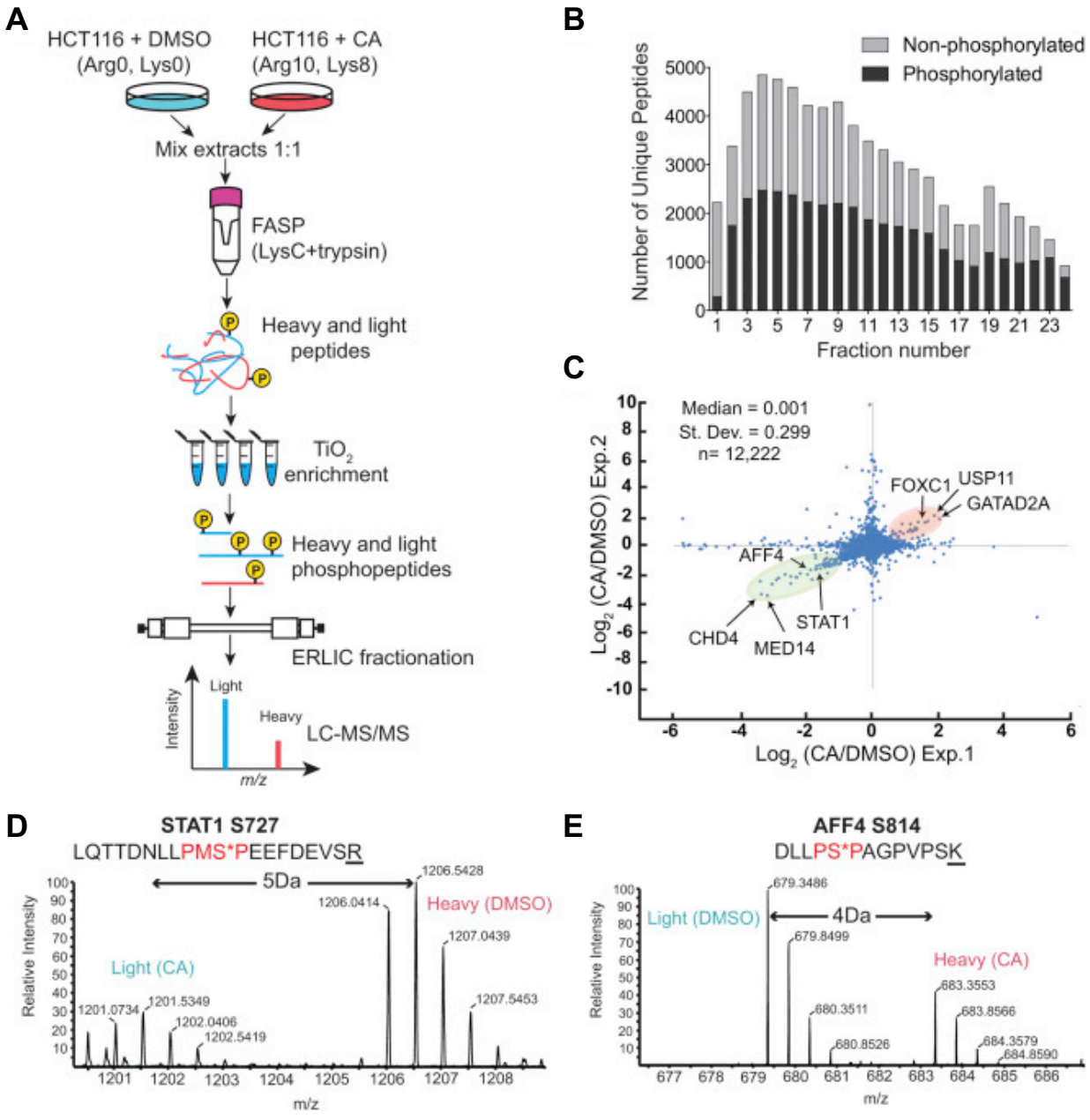


Figure 4.2 | Quantitative phosphoproteomics in HCT116 cells ± CA. (**A**) Overview of phosphoproteomics workflow used to identify Mediator kinase substrates. (**B**) Unique phosphopeptides identified with LC-MS/MS after ERLIC fractionation. (**C**) CA treatment with quantitative phosphoproteomics reproducibly identifies Mediator kinase substrates. H/L ratios quantified in two of three biological replicates are plotted on the x and y axes. Plot shows proteins

whose H/L ratios decrease (green) and increase (peach) upon CA treatment. H = heavy isotopes; L = light isotopes. (**D**-**E**) Representative mass spectra. Spectra shown are from replicates in which either the light (**D**) or heavy (**E**) cells were CA treated. Differences in SILAC pairs are shown based on the labeled amino acid; Arg(10) in (**D**) and Lys(8) in (**E**). The charge is +2 for both peptides.

The majority of phosphosites were unaffected by CA treatment, clustering around zero in a $\log_2$ plot of H/L SILAC ratios across replicate experiments (**Figure 4.2C**). This result indicated good reproducibility and provided further validation of CA specificity. Many decreased phosphosites were highly correlated across replicates (highlighted green in **Figure 4.2C**); in addition, we identified a smaller number of phosphosites that increased upon CA treatment (highlighted peach in **Figure 4.2C**). Representative mass spectra for SILAC pairs shown in **Figures 4.2D** and **E** are from experiments in which either light (**D**) or heavy (**E**) cells were treated with CA. For two of three replicates, the heavy population of cells was CA treated, whereas in one replicate light cells were CA treated, representing a label swap. For data analysis purposes, a reciprocal of the H/L ratio was calculated for the label swap experiment, such that decreased H/L ratios could be evaluated across all biological replicates.

CDK8/19 substrates are largely transcription-associated proteins

The phosphoproteomics workflow in **Figure 4.2A** identified novel phosphosites whose intensities decreased significantly with CA treatment (**Figure 4.3A**). We identified 78 phosphosites, represented in 64 proteins, that we designated as high confidence based upon 1) their quantification in at least two of three biological replicates, 2) a reproducible mean H/L ratio

across replicates, and 3) a significant decrease in H/L ratio with CA treatment as determined by an empirical Bayes analysis (Margolin et al., 2009; Ritchie et al., 2015). These high-confidence phosphosites are summarized in **Table 4.1** and **Figure 4.3A**. To ensure that a reduced H/L ratio did not result simply from a change in protein level, we completed a quantitative proteome analysis in parallel with phosphoproteomics. Importantly, very few high confidence phosphosites exhibited any change at the protein level with one hour of CA treatment. Those that did change somewhat were FOXC1, MAML1, KDM3A, and ATF2, although some of these changes were not consistent across deep proteome replicates, and most of the phosphosite changes remained significant even after accounting for small changes in protein level. Although phosphosites not designated as high confidence sites could represent bona-fide Mediator kinase substrates (e.g., those that are quantified in only one biological replicate), we will only discuss targets designated as high confidence based on the criteria above.

Table 4.1 | Identification of Mediator kinase (CDK8/19) substrates

Protein ID	Gene name	Position	Location probability	Ratio (H/L)	Adjusted *p*-value
Q9UHB7	*AFF4*	S31	0.79	0.209 ± 0.011	0.025
Q9UHB7	*AFF4*	S32	0.78	0.248 ± 0.070	0.056
Q9UHB7	*AFF4*	S1043	1	0.325 ± 0.035	0.02
Q9UHB7	*AFF4*	S814	1	0.410 ± 0.012	0.018
P15336	*ATF2*	S136	1	0.366 ± 0.012	0.015

Table 4.1 (continued)

P17544	*ATF7*	S434	0.88	0.302 ± 0.020	0.015
P17544	*ATF7*	S111	0.75	0.590 ± 0.062	0.082
P17544	*ATF7*	T112	0.99	0.593 ± 0.019	0.036
O00512	*BCL9*	S291	0.95	0.694 ± 0.031	0.092
Q12830	*BPTF*	S1300	1	0.691 ± 0.035	0.095
P38398	*BRCA1*	S1613	0.96	0.447 ± 0.036	0.025
Q9H8M2	*BRD9*	S588	1	0.455 ± 0.009	0.019
H0YBQ5;E5RFK5	*CCNC;CCNC*	S218;S272	1	0.642 ± 0.037	0.066
Q12873	*CHD3*	S1601	1	0.595 ± 0.036	0.048
Q14839	*CHD4*	T1553	0.94	0.096 ± 0.002	0.018
Q9P2D1	*CHD7*	T2153	0.99	0.572 ± 0.024	0.034
K4DI93	*CUL4B*	S15	0.9	0.394 ± 0.038	0.078
Q9UER7	*DAXX*	S671	1	0.651 ± 0.014	0.053
Q5T1V6	*DDX59*	S64/S76	0.99/0.76	0.535 ± 0.045	0.064
Q9NPF5	*DMAP1*	T409	1	0.134 ± 0.018	0.015
P19419	*ELK1*	S324	1	0.636 ± 0.058	0.092
Q52LR7	*EPC2*	T353	0.97	0.403 ± 0.019	0.059
Q96E09	*FAM122A*	S267	0.79	0.513 ± 0.082	0.085
Q12948	*FOXC1*	S241	1	0.433 ± 0.063	0.048
Q9NZM4	*GLTSCR1*	S755	1	0.394 ± 0.052	0.032
P15822	*HIVEP1*	S479	0.99	0.649 ± 0.019	0.055

Table 4.1 (continued)

Q7Z6Z7	*HUWE1*	S3816	0.98	0.527 ± 0.018	0.025
Q8NFU5	*IPMK*	S7	1	0.663 ± 0.050	0.098
Q9Y4C1	*KDM3A*	S445	1	0.448 ± 0.034	0.025
Q9Y4X4	*KLF12*	S202	1	0.441 ± 0.071	0.056
Q3ZCW2	*LGALSL*	S25	0.99	0.698 ± 0.029	0.092
Q92585	*MAML1*	S159	1	0.325 ± 0.009	0.015
Q92585	*MAML1*	S303	0.98	0.356 ± 0.018	0.049
Q14676	*MDC1*	S1775	1	0.453 ± 0.075	0.061
Q93074	*MED12*	S688	0.99	0.168 ± 0.036	0.02
Q9UHV7	*MED13*	S749	0.96	0.356 ± 0.041	0.073
Q71F56	*MED13L*	S878	1	0.568 ± 0.060	0.065
O60244	*MED14*	S1112	1	0.233 ± 0.045	0.025
O60244	*MED14*	S1128/S1136	0.99/0.99	0.100 ± 0.015	0.025
O95402	*MED26*	S314	1	0.177 ± 0.039	0.025
Q8IWI9	*MGA*	S2924	0.99	0.660 ± 0.032	0.07
O14686	*MLL2*	S3199	0.98	0.630 ± 0.019	0.047
P55197	*MLLT10*	S346	1	0.470 ± 0.005	0.076
O96007	*MOCS2*	S20	1	0.531 ± 0.035	0.034
Q9NV56	*MRGBP*	S195	1	0.611 ± 0.023	0.043
Q6P1R3	*MSANTD2*	S27	1	0.623 ± 0.053	0.085
Q2TAK8	*MUM1*	S326	1	0.582 ± 0.050	0.059

Table 4.1 (continued)

Q15742	*NAB2*	S162	1	0.573 ± 0.014	0.032
Q15788	*NCOA1*	S698	1	0.381 ± 0.042	0.081
Q9H3P2	*NELFA*	S363	0.96	0.403 ± 0.032	0.023
Q9H3P2	*NELFA*	S360	0.5	0.441 ± 0.039	0.092
Q6P4R8	*NFRKB*	S1291	1	0.688 ± 0.035	0.092
Q9NZT2	*OGFR*	S349	1	0.170 ± 0.015	0.025
Q9NZT2	*OGFR*	S484	0.99	0.348 ± 0.035	0.021
P29590	*PML*	S530	1	0.674 ± 0.012	0.059
Q6EEV4	*POLR2M*	S10	0.99	0.416 ± 0.018	0.019
Q5UIP0	*RIF1*	S1613	0.98	0.357 ± 0.014	0.048
Q5UIP0	*RIF1*	S1616	1	0.418 ± 0.051	0.033
Q92766	*RREB1*	S1653	1	0.237 ± 0.025	0.039
Q6SPF0	*SAMD1*	S425	0.94	0.418 ± 0.022	0.019
O15047	*SETD1A*	T1088	1	0.204 ± 0.010	0.015
Q7Z333	*SETX*	S2612	1	0.465 ± 0.018	0.08
Q96EB6	*SIRT1*	T530	1	0.201 ± 0.030	0.043
Q9UQ35	*SRRM2*	S2449	1	0.708 ± 0.004	0.081
P42224	*STAT1*	S727	0.99	0.367 ± 0.001	0.047
Q12962	*TAF10*	S44	1	0.565 ± 0.020	0.032
Q7Z2Z1	*TICRR*	S1413	0.99	0.472 ± 0.066	0.059
Q12888	*TP53BP1*	S265	1	0.258 ± 0.014	0.015

Table 4.1 (continued)

Q12888	*TP53BP1*	S525	1	0.600 ± 0.014	0.036
P13051	*UNG*	S63	0.84	0.473 ± 0.026	0.025
P13051	*UNG*	T60/S63	0.99/0.62	0.487 ± 0.027	0.025
Q9H0D6	*XRN2*	S487	1	0.566 ± 0.041	0.047
Q9NUA8	*ZBTB40*	T166	0.99	0.378 ± 0.078	0.058
A6NFI3	*ZNF316*	S10	1	0.598 ± 0.043	0.056
Q6ZN55	*ZNF574*	S717	1	0.406 ± 0.052	0.033
O15014	*ZNF609*	S804	1	0.373 ± 0.001	0.015

To determine whether a H/L ratio for a phosphosite changed significantly with CA treatment across replicates, we employed an empirical Bayes statistical approach using the limma software package (Ritchie et al., 2015). An empirical Bayesian framework allowed for the calculation of adjusted *p*-values for each phosphosite (**Figure 4.3B**). This approach can account for experiment-specific differences, which is advantageous compared to more arbitrary approaches, such as a universal fold-change cutoff (Margolin et al., 2009). We found that more phosphosite ratios decreased than increased upon CA treatment, as expected with targeted kinase inhibition for a short time. This is shown by a higher number of data points on the left side of the volcano plot compared to the right side, using an adjusted *p*-value cutoff of 0.1 (**Figure 4.3B**).

We used iceLOGO (Colaert et al., 2009) to determine statistically enriched motifs within the identified Mediator kinase substrates. We found that the majority of the phosphosites

contained an S/T-P motif (**Figure 4.3C**). Additionally, a proline at the −2 and −1 positions relative to the phosphorylation site was overrepresented. These data support the notion that many CDK8 phosphorylation sites occur within PX(S/T)P motifs as previously suggested (Alarcon et al., 2009; Bancerek et al., 2013). Serine was more frequently phosphorylated than threonine (**Figure 4.3C**) and we did not see evidence for overrepresentation of basic residues at positions C-terminal to the phosphosite, as observed with other CDK motifs (Ubersax and Ferrell, 2007).

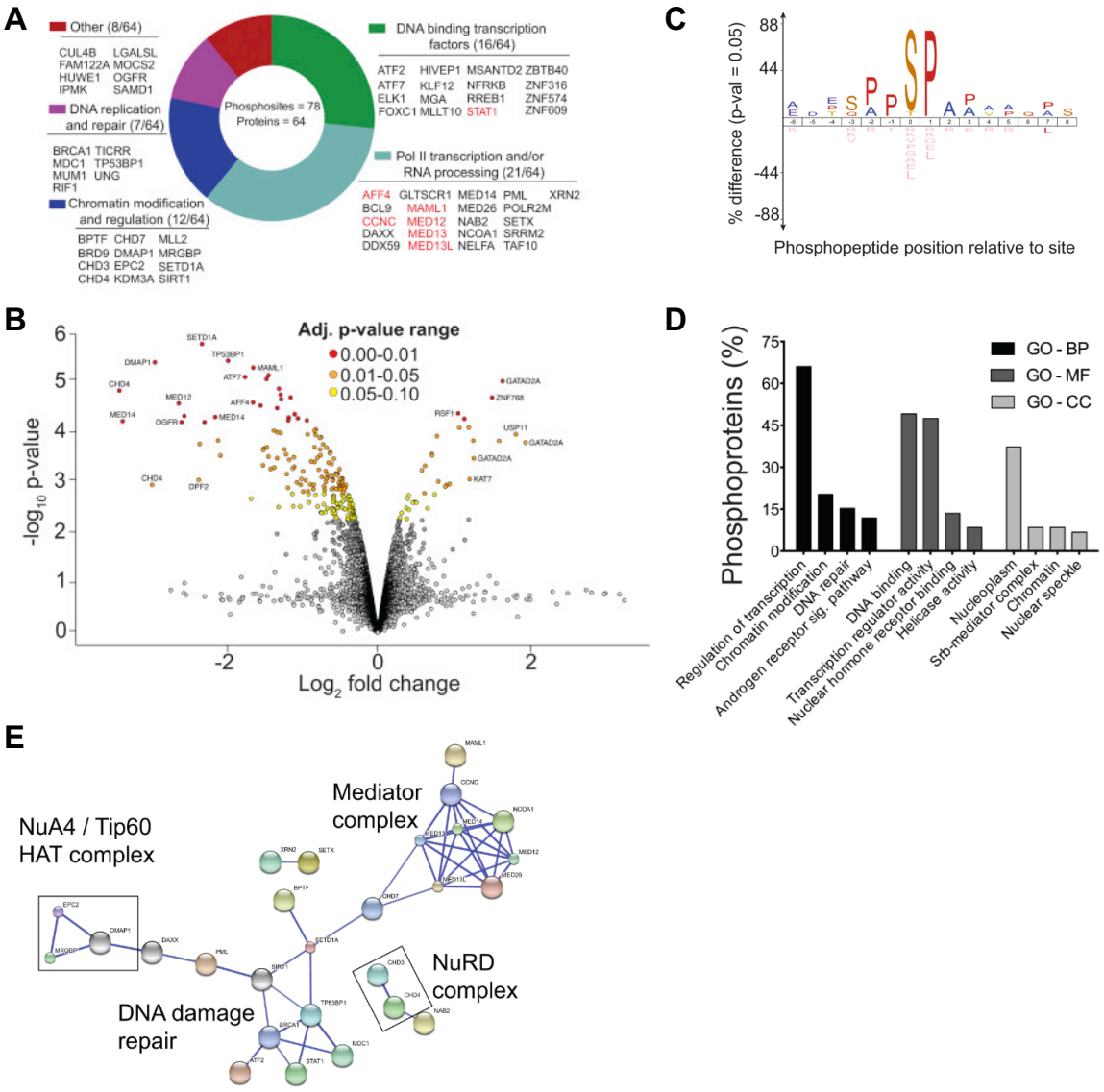


Figure 4.3 | High confidence Mediator kinase substrates. (**A**) Functional categorization of high-confidence Mediator kinase substrates identified. (**B**) Volcano plot of statistically significant phosphosite changes with CA treatment using an empirical Bayes analysis. (**C**) Motif analysis of CDK8/19 substrates using iceLogo. Peptides of the same length were input and centered on the phosphorylated residue. (**D**) GO biological process (BP), molecular function (MF), and cellular compartment (CC) completed with DAVID using the identified phosphoproteins. Enriched terms (p-value and FDR ≤ 0.05) are shown. (**E**) Protein interaction analysis of CDK8/19 substrates using the STRING database (high-confidence interactions ≥ 0.7). Nodes represent proteins and lines between them represent a documented interaction.

Because few substrates for human CDK8/19 have been identified, the analysis uncovered many phosphosite targets (**Table 4.1**). Many targets are DNA-binding transcription factors, chromatin regulators, or other known regulators of RNA Pol II activity (**Figures 4.3A** and **D**), consistent with the established role of CDK8 in transcription. Additional substrates, including proteins implicated in DNA replication and repair (BRCA1, MDC1) and ubiquitination (HUWE1, CUL4B) suggest biological roles for Mediator kinases beyond transcription. A known CDK8 substrate, STAT1 S727 (Bancerek et al., 2013), was identified as a high confidence target, and other novel phosphosites reside in proteins that interact with CDK8-Mediator and/or the CDK8 module, including AFF4, MAML1, and Mediator subunits (**Figure 4.3A** and **Table 4.1**). AFF4 is a core component of the super-elongation complex (Luo et al., 2012), which co-purifies with CDK8-Mediator (Ebmeier and Taatjes, 2010), and MAML1 is a Notch pathway co-activator that recruits CDK8 to Notch-dependent genes where it phosphorylates the Notch ICD (Fryer et al., 2004).

We submitted the 64 CDK8/19 substrate proteins to the STRING protein-protein interaction database (Szklarczyk et al., 2015) and found that six Mediator complex subunits, three subunits of the TIP60/NuA4 complex (EPC2, DMAP1, MRGBP), and two subunits of the NuRD complex (CHD3 and CHD4) were represented (high confidence score, ≥ 0.7; **Figure 4.3E**). The TIP60/NuA4 and NuRD complexes are multi-subunit assemblies that possess multiple enzymatic activities, including nucleosome remodeling, acetyltransferase, and deacetylase activities. Additionally, this analysis identified a network of interacting proteins involved in DNA damage repair (**Figure 4.3E**), as well as an interaction between XRN2 and SETX. Taken together, these data suggest that Mediator kinases regulate multiple and diverse cellular processes, potentially via several distinct multi-subunit assemblies.

Validation of selected CDK8/19 substrates

To further validate the CDK8/19 substrates identified with SILAC-based phosphoproteomics, we performed *in vitro* kinase assays, western blots, and mass spectrometry experiments. We selected proteins representing different classes of substrates (**Figure 4.3A**) for further evaluation. The DNA-binding transcription factor STAT1, a previously identified CDK8 kinase substrate (Bancerek et al., 2013), was probed by western blot in IFN-γ induced HCT116 cells (**Figure 4.4A**). This experiment confirmed STAT1 S727 as a Mediator kinase substrate in HCT116 cells, and also showed CA-dependent inhibition at low nM concentrations, as reported previously (Pelish et al., 2015).

Among the chromatin modification and regulation substrates, we examined SIRT1, in part because an antibody against the phosphorylated SIRT1 T530 site was commercially available. When HCT116 cells were treated with CA, we noted a decrease in SIRT1 T530 phosphorylation (**Figures 4.4B-C**). Total SIRT1 levels were unaffected by CA treatment, and levels of the CDK8 module subunits CDK8, cyclin C, and MED12 were not changed by treatment with the compound (**Figure 4.4B**). The approximately 50% reduction in phospho-SIRT1 did not change with increasing CA concentration, indicative of CA selectivity (Pelish et al., 2015). Although other kinases, such as CDK1 and JNK, are known to phosphorylate this site (Sasaki et al., 2008), treatment with inhibitors of CDK1 (RO-3306) and JNK family kinases (SP600125) did not seem to impact SIRT1 T530 phosphorylation; in fact, we were unable to completely reduce SIRT1 phospho-T530 detection, even when treating with all three inhibitors. *In vitro* kinase assays with purified CDK8 module and SIRT1 confirmed CDK8-dependent SIRT1 T530 phosphorylation by western blot (**Figure 4.4D**).

We next tested two different substrates, RIF1 and TP53BP1, linked to DNA replication and repair. Because these proteins are very large (each over 200 kDa), we expressed GST-tagged fragments (approximately 100 residues) surrounding the phosphosite. As shown in **Figure 4.4E**, the CDK8 module phosphorylated these substrates, whereas point mutations (Ser to Ala) at the identified phosphorylation site(s) greatly reduced substrate phosphorylation, supporting these sites as CDK8 module targets in vitro (**Figure 4.4E**).

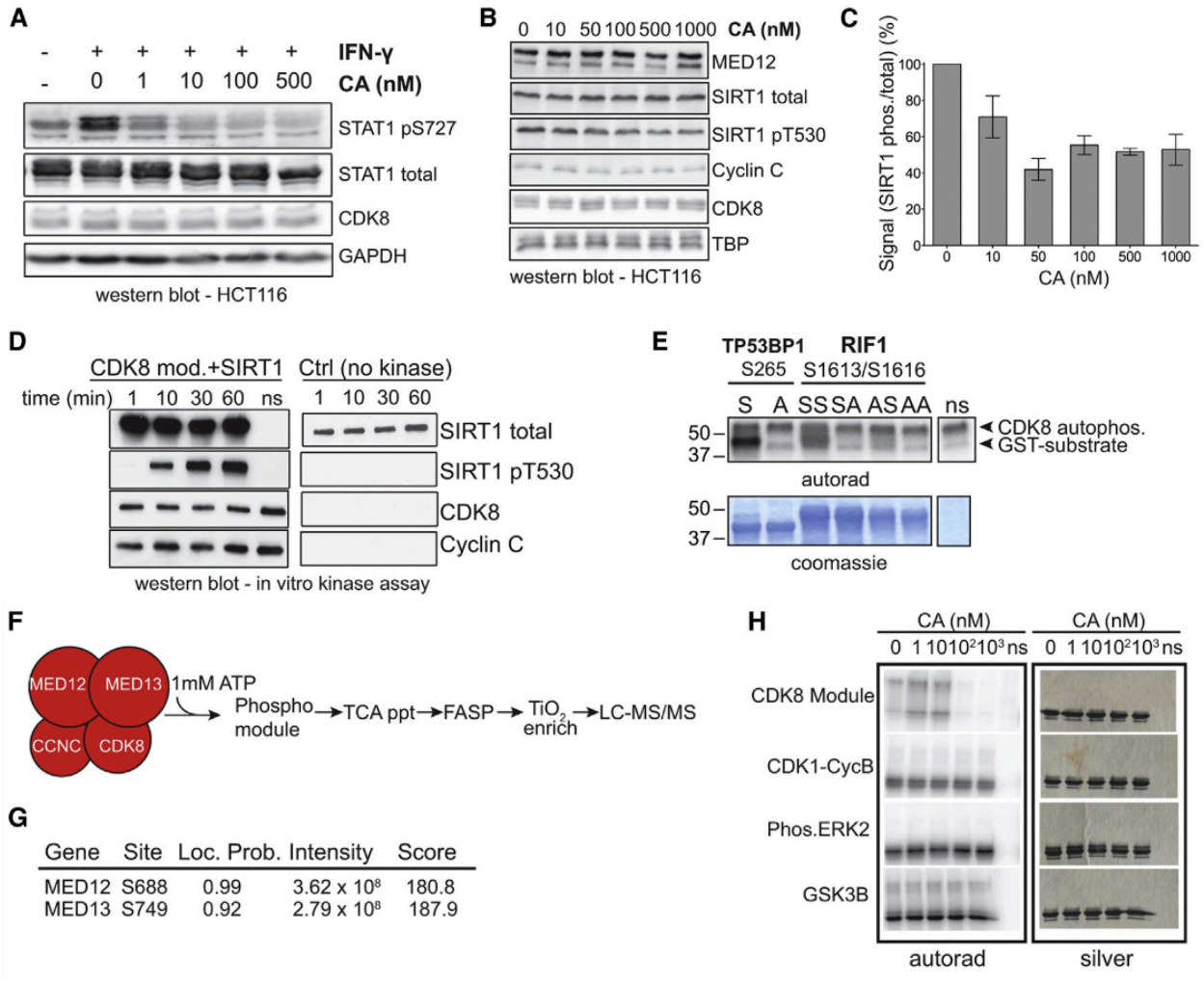


Figure 4.4 | *In vitro* validation of select CDK8/19 substrates. (**A**) Validation of STAT1 S727 as a Mediator kinase target in HCT116 cells. (**B**) Western blot validation of SIRT1 T530 as a Mediator kinase target. Levels of total SIRT1 and other proteins known to regulate CDK8 activity (MED12 or cyclin C) remained unchanged. TBP is a loading control. (**C**) Quantitation of data in **B**. (**D**) *In vitro* kinase assay with recombinant CDK8 module and SIRT1. With increasing time, SIRT1 pT530 detection increases, indicating CDK8 is phosphorylating this site. Increase is not seen in no kinase or no substrate (ns) controls. (**E**) *In vitro* kinase assay with GST-tagged TP53BP1 or RIF1 fragments. Alanine mutations at identified phosphorylation sites show reduced phosphorylation by CDK8. (**F**) Overview of method for identifying MED12 and MED13 phosphorylation sites using

recombinant CDK8 modules. (**G**) Verification of MED12 S688 and MED13 S749 phosphorylation sites. (**H**) *In vitro* kinase assay using CA and GST-RNA Pol II CTD as a substrate. Whereas each kinase tested phosphorylates this substrate, CA only inhibits the CDK8 module.

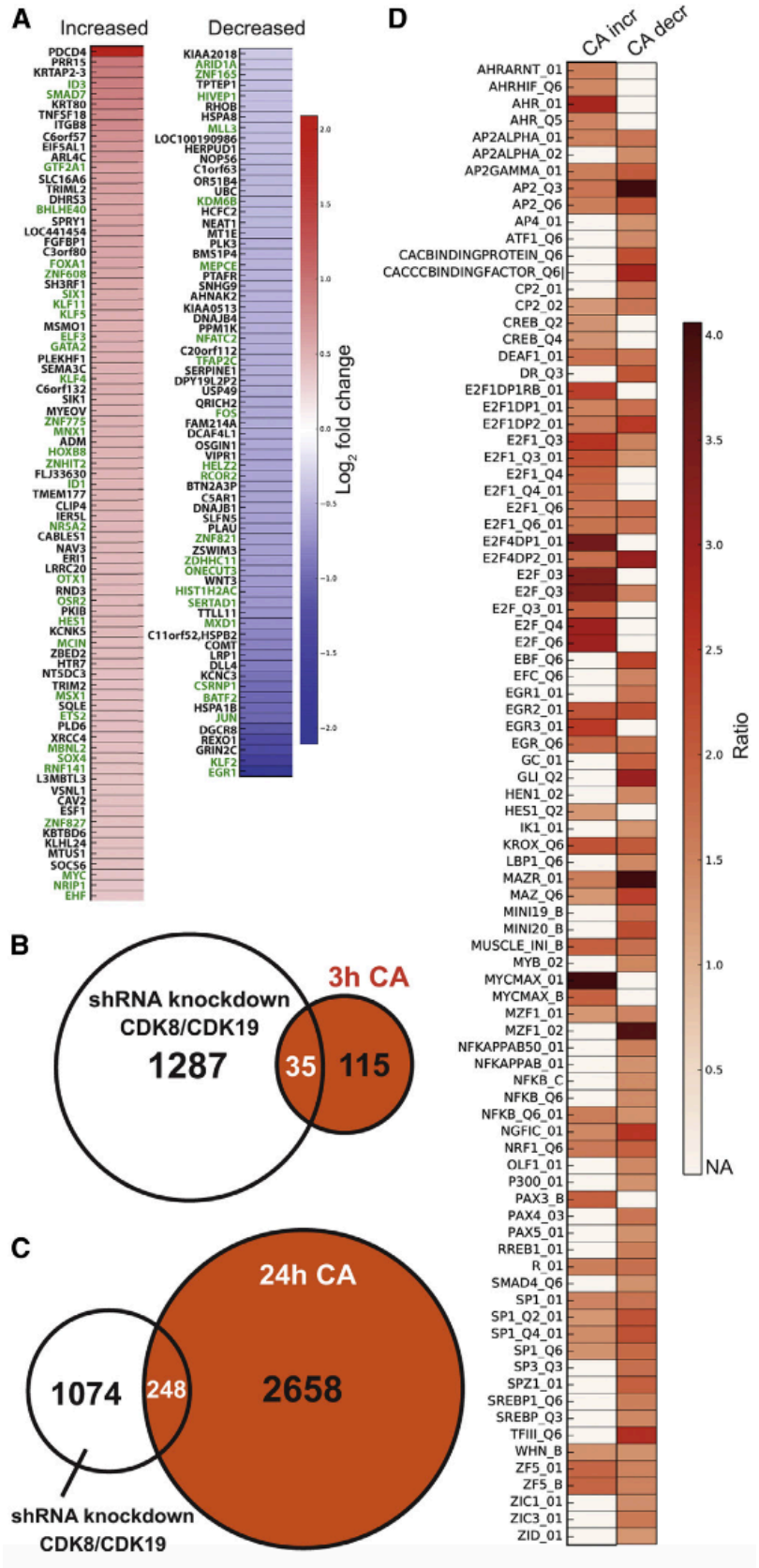
We also confirmed phosphorylation sites in MED12 and MED13 using in vitro kinase assays using the recombinant CDK8 module (containing CDK8, cyclin C, MED12, and MED13) purified from insect cells. Incubation of the CDK8 module with ATP and subsequent TiO_2 enrichment and mass spectrometric analysis confirmed both S688 on MED12 and S749 on MED13 as substrates (**Figures 4.4F** and **G**). We did not identify the cyclin C site from these experiments because the site identified from CA-treated HCT116 cells is not present in the canonical cyclin C isoform used for recombinant CDK8 module expression and purification.

The data summarized in **Figure 4.4** verified each of seven high-confidence Mediator kinase sites, representing about 10% of all high-confidence sites listed in **Table 4.1**. These results, combined with previous data demonstrating CA potency and specificity (Pelish et al., 2015), support the substrates listed in **Table 1** as Mediator kinase targets. Although extensive kinome profiling has demonstrated CA specificity, we conducted *in vitro* kinase assays using a shared substrate, the RNA Pol II C-terminal domain, and found that CDK1, ERK2, and GSK3β activity was unaffected by CA treatment, even at concentrations ten-fold above those used for proteomic and gene expression analyses (**Figure 4.4H**).

CDK8/19 inhibition has limited effects on transcription

As Mediator-associated kinases, it was plausible that inhibition of CDK8 and CDK19 activity could affect expression of large numbers of genes. We analyzed gene expression (RNA-seq) data from CA treated HCT116 cells. To minimize secondary or indirect effects resulting from long-term Mediator kinase inhibition, we completed RNA-seq after a three-hour CA treatment (100 nM); this also helped match mRNA changes with measured phosphorylation changes that were determined after one-hour CA treatment. RNA-seq analysis identified 150 genes whose expression changed significantly with CA treatment (**Figure 4.5A**). Among these genes, the magnitude of change in expression was modest (largely 1.2 – 2 fold), indicating that CDK8/19 activity per se is not a major driver of their transcription, at least in the context of this analysis (HCT116 cells under normal growth conditions). Such modest gene expression changes were also observed in CA-sensitive cell lines (e.g. MOLM-14), although the genes affected were distinct (Pelish et al., 2015).

Figure 4.5 | CDK8/19 inhibition is functionally distinct from CDK8/19 knockdown. (**A**) Heatmap of differentially expressed genes (RNA-seq) after 3-hour CA treatment. Green font represents transcription or chromatin regulator. (**B-C**) Comparison with microarray data (Galbraith et al., 2013) using stable CDK8/19 knockdown (shRNA) versus 3-hour CA treatment (**B**) or 24-hour treatment (**C**) in HCT116 cells under normal growth conditions. A 1.5-fold cutoff was used for microarray data, and Cufflinks was used for CA-treated cells (no specific fold-change cutoff). (**D**) Transcription factor binding site analysis of promoters for genes whose expression changed with 3-hour CA treatment (listed in **A**). Promoters (±2 kb from the transcriptional start site of the canonical isoform) were analyzed using F-Match, part of the Transfac database. Over-represented sites with at least 1.5-fold increase versus control promoters are shown for Transfac vertebrate matrices. Matrix name is at left.


A
Increased
Decreased
Log2 fold change
B
shRNA knockdown
CDK8/CDK19
1287
35
115
3h CA
C
24h CA
1074
248
2658
shRNA knockdown
CDK8/CDK19
D
CA incr
CA decr
Ratio
NA

Gene expression changes in CA-treated cells compared to CDK8 or CDK19 knockdown

Because CA inhibits both CDK8 and CDK19 kinases, we used previously published HCT116 microarray datasets (normal growth conditions), in which either CDK8 or CDK19 had been stably knocked down (Donner et al., 2010; Galbraith et al., 2013), as a comparison to CA-treated HCT116 cells. Only genes exhibiting 1.5-fold change in expression or greater, with *p*-values < 0.05, were used from the microarray data; these genes were compared to our RNA-seq analysis in which cells were treated with 100 nM CA for three hours. We observed only a modest overlap among genes differentially expressed (**Figure 4.5B**; note that because CA inhibits both CDK8 and CDK19, gene sets for CDK8 or CDK19 knockdown were combined). Because cellular knockdown experiments take over 24 hours to manifest, the modest correlation in gene expression changes could reflect the short time of CA treatment. However, RNA-seq analyses after 24-hour CA treatment revealed similarly low numbers of shared gene expression changes (**Figure 4.5C**). These results suggest that the physical presence of the CDK8 or CDK19 protein has distinct effects on transcription compared to targeted kinase inhibition.

Functional links between gene expression changes and Mediator kinase substrates

Because many Mediator kinase substrates are transcription factors (**Figure 4.3A**), we hypothesized that some of the observed differences in gene expression due to CDK8/19 inhibition might be caused by changes in transcription factor function. To begin to address this hypothesis, we extracted promoter sequences (±2 kb from transcription start site) for genes that were differentially expressed (increased or decreased expression) 3-hour CA treatment. F-Match was then used to compare promoter sequences to controls to determine if any transcription factor binding sites, reported as Transfac matrices, were over-represented. The ratio of this increase (CA-

treated cells vs. DMSO controls) is displayed for over-represented sites in **Figure 4.5D**. We found that many of the identified Transfac matrices for genes whose expression increased or decreased with CA treatment were mutually exclusive (**Figure 4.5D**). That is, an enriched transcription factor binding site in CA increased genes was generally not present in CA decreased genes, and vice versa. A hypergeometric test confirmed a significant overlap (p-value = 1.15×10^{-5}) between Mediator kinase targets identified in **Table 1** and Transfac matrices identified in our gene expression promoter analysis.

Many of the Transfac matrices identified by F-match (**Figure 4.5D**) can be traced back to Mediator kinase activity. For example, the RREB1 transcription factor was identified in both the F-match analysis (RREB1_01) and the SILAC phosphoproteomics (**Table 1**). Enriched transcription factor binding sites were observed for genes with altered expression in CA-treated HCT116 cells (**Figure 4.5A**), including *MYC*, a β-catenin target gene, *EGR1* (i.e., KROX_Q6), and *HES1*, a Notch pathway target gene. Moreover, the MGA and NAB2 proteins, each high-confidence Mediator kinase substrates, are known regulators of *MYC* and *EGR1* activity, respectively (Hurlin et al., 1999; Svaren et al., 1996). Transfac matrices representing the AP2 and ATF family of transcription factors (e.g., AP2alpha_01 and ATF1_Q6) were also uncovered in the F-match analysis. The Mediator kinase target KLF12 is a well-established repressor of AP2α activity (gene name: *TFAP2A*) (Imhof et al., 1999), whereas the ATF2 and ATF7 proteins were each identified as Mediator kinase targets. Finally, enriched binding sites for E2F1 and SREBP, previously identified CDK8 kinase substrates (Morris et al., 2008; Zhao et al., 2012), were found by the F-match analysis shown in **Figure 4.5D**. Although these transcription factors were not identified in our HCT116 phosphoproteomics experiments, several co-regulators of E2F1 or

SREBP activity (e.g., MGA and SIRT1) were among the high-confidence substrates listed in **Table 1**. Thus, there are many functional links between CA-dependent changes in gene expression (**Figure 4.5A**) and the Mediator kinase targets shown in **Table 1**.

Cellular proteome changes resulting from CDK8/19 inhibition

The ability of CDK8-dependent phosphorylation to regulate protein turnover has been reported in both yeast and human cells (Alarcon et al., 2009; Fryer et al., 2004; Nelson et al., 2003; Raithatha et al., 2012). We therefore hypothesized that CDK8/19 activity might modulate protein abundance for some of the substrates identified here. Rather than focus on selected Mediator kinase targets, we performed quantitative proteome analyses in CA-treated cells versus DMSO controls at six time points (t = 0, 1, 3, 6, 18, and 24 hours). In this way, we were able to interrogate many cellular proteins at once, and correlate changes in Mediator kinase activity with increased or decreased protein abundance. To complete these analyses, we used SILAC labeled HCT116 cells, consistent with the phosphoproteomics experiments.

The analysis consisted of a CA treatment time course from 0 to 24 hours, with six time points being used in total to treat heavy (Arg10, Lys8) or light (Arg0, Lys0) HCT116 cell populations in biological replicate experiments (**Figures 4.6A-B**). Peptides were harvested in a manner similar to that used for phosphoproteomics, and 17 fractions from basic reversed-phase chromatography were analyzed for changes in H/L ratio at each time point. We found a high number of overlapping proteins across replicates, and CA treatment did not affect global H/L ratios for proteins across the time course in the replicates (**Figure 4.6B**). Given the ability of CDK8

to promote substrate turnover in response to specific biological phenomena (e.g., starvation) (Nelson et al., 2003; Raithatha et al., 2012), we were somewhat surprised to find that CDK8/19 inhibition did not notably alter the abundance of the target proteins listed in **Table 1**, with the exception of MED13 and MED13L.

An empirical Bayes analysis of the data suggested that most proteome changes occurred at either 18 or 24 hours when compared to control (0 hour, in which both populations were DMSO treated), as shown by the volcano plot in **Figure 4.6C**. Approximately 200 proteins showed significant changes in abundance (adjusted p-value < 0.1). To further examine changes in the proteome with CA treatment, gene set enrichment analysis (GSEA) was employed (Subramanian et al., 2005). Using the hallmark gene set collection, we identified biological processes that displayed significant enrichment scores and false discovery rates (**Figure 4.6D**). Of these signatures, several have been previously shown to be regulated by CDK8, including Wnt/β-catenin signaling, Notch signaling, hypoxia, interferon gamma response, and KRAS signaling (Bancerek et al., 2013; Firestein et al., 2008; Fryer et al., 2004; Galbraith et al., 2013; Morris et al., 2008; Xu et al., 2015). CDK8-dependent transcriptional changes have been implicated in regulation of these pathways and therefore the proteome data corroborate these findings at the protein level. The GSEA results also reveal that proteome changes may selectively affect metabolic pathways in CA-treated HCT116 cells, with several (e.g. cholesterol homeostasis, fatty acid metabolism) previously linked to CDK8 kinase activity in model organisms (Zhao et al., 2012).

Discussion

CA is an exceptionally selective inhibitor of the Mediator kinases CDK8 and CDK19 (Pelish et al., 2015). As such, it provided a means to rapidly and selectively probe CDK8- and CDK19-dependent phosphoproteome changes in human cells. Because of their association with Mediator, CDK8 and CDK19 were expected to phosphorylate proteins involved in regulating RNA Pol II activity and chromatin architecture. In accordance with these expectations, our data support a primary role for Mediator kinases in RNA Pol II transcription and chromatin regulation. Strikingly, however, the direct impact of Mediator kinase inhibition on global RNA Pol II transcription was modest and affected a limited set of genes, at least under the conditions of this study. Limited transcriptional effects were also observed in CA-sensitive AML cell lines (Pelish et al., 2015).

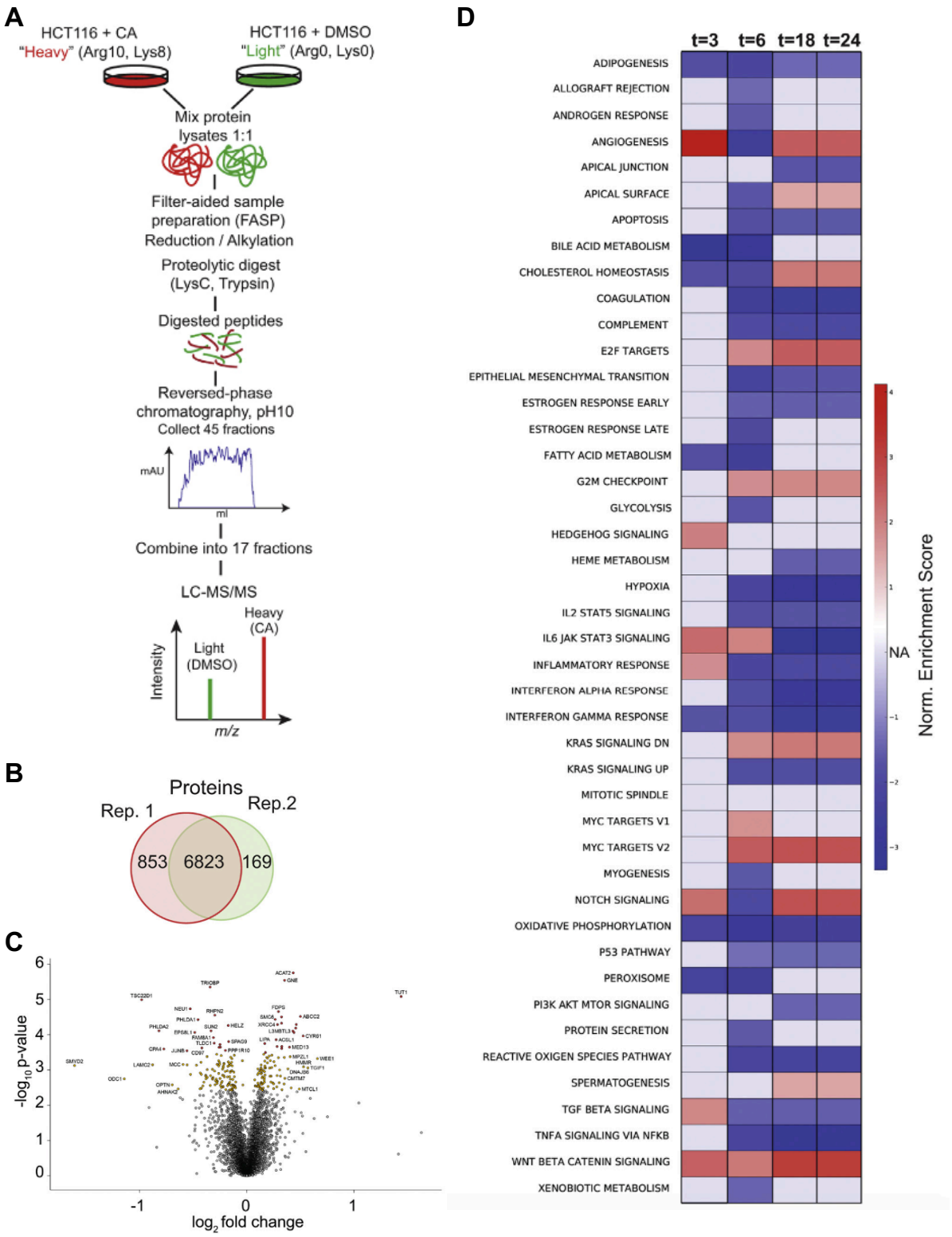


Figure 4.6 | Proteomics reveals pathways and proteins affected by CDK8/19 inhibition. (**A**) Overview of quantitative proteomics method. (**B**) Venn diagram of biological replicates showing

number of proteins identified in the time series. Replicates show a high degree of overlap for protein IDs. (**C**) Volcano plot comparing protein abundance at 18-hour and 24-hour time points versus control (0 hour). Adjusted *p*-values are colored according to an empirical Bayes analysis. (**D**) Individual analysis of t = 3-, 6-, 18-, and 24-hour CA treatment time points using GSEA and the hallmark gene sets from the Molecular Signatures Database. Comparison of the t = 0-hour and 1-hour time points showed no differences in the hallmark gene sets (not shown). The color of the heatmap corresponds to the direction and magnitude of the normalized enrichment score for that gene set at each time point, compared to t = 0-hour controls. "NA" and the corresponding color indicate a hallmark gene set not being identified from the proteome data at the designated time.

At the gene expression level, it appears that Mediator kinases predominantly regulate the regulators of transcription. Many genes whose expression increased or decreased 1.5-fold or greater upon CA treatment are DNA-binding transcription factors or general transcription or chromatin regulators. Similarly, DNA-binding transcription factors and RNA Pol II transcription or chromatin regulators represented the majority of high-confidence CDK8/19 kinase targets from the SILAC phosphoproteomics experiments. Quantitative proteomic data across a 24-hour time course implicated numerous signaling and metabolic pathways that appear to be regulated by Mediator kinase activity under normal growth conditions. Whereas these pathways can be linked to known transcriptional or phosphorylation targets of CDK8/19 or those now identified here, much additional investigation will be required to delineate the molecular mechanisms by which Mediator kinases regulate specific signaling pathways or transcriptional processes.

CDK8/19 phosphorylate Mediator subunits and post-initiation transcription regulators

CDK8 can reversibly associate with Mediator to form a CDK8-Mediator complex (Taatjes et al., 2002), and immunoprecipitation-mass spectrometry experiments in HeLa or HEK293T cells suggest CDK19 interacts similarly with Mediator (Daniels et al., 2013; Ebmeier and Taatjes, 2010; Sato et al., 2004). We identified eight high-confidence CDK8/19 phosphorylation sites in six different Mediator subunits: cyclin C, MED12, MED13, MED13L, MED14, and MED26. Cyclin C, MED12, MED13, and MED13L each associate with CDK8 or CDK19 as part of the kinase module of Mediator. MED13 appears to be important for physical interaction between the kinase module and Mediator (Knuesel et al., 2009), and previous studies have shown that increased MED13 or MED13L abundance can increase the proportion of CDK8-Mediator versus core Mediator in cells (Davis et al., 2013). These previous results were shown in the context of inhibition of the E3 ubiquitin ligase FBW7, which ubiquitylates MED13 and MED13L to promote their degradation. Our quantitative whole proteome data showed that the abundance of MED13 and MED13L were each increased in CA-treated HCT116 cells. FBW7-dependent ubiquitylation of MED13 or MED13L required prior modification at residue T326, a phospho-degron site in MED13 and MED13L (Davis et al., 2013). The CDK8/19 sites identified in MED13 and MED13L are distinct (residue S749 and residue S878, respectively) and do not overlap with known or predicted phospho-degron motifs; thus, it remains unclear how Mediator kinase activity may affect MED13 or MED13L protein levels.

The MED26 subunit is generally absent from CDK8-Mediator purifications (Ebmeier and Taatjes, 2010; Sato et al., 2004; Taatjes et al., 2002) and hence its phosphorylation by CDK8 or CDK19 may promote MED26 dissociation from Mediator. The CDK8/19 modification site on

MED26 (S314), however, does not reside in regions required for Mediator association (Takahashi et al., 2011). The MED14 subunit is an important architectural factor within Mediator, and structural studies with reconstituted partial complexes and crosslinking-mass spectrometry (CXMS) revealed MED14 crosslinks with several Mediator subunits, including MED8 and MED7, involving MED14 residues 1256 and 1295, respectively (Cevher et al., 2014). These reside some distance (in sequence space) from the Mediator kinase phosphorylation sites (S1112, S1128, S1136). Furthermore, CXMS and cryo-EM data with reconstituted yeast Mediator and yeast RNA Pol II revealed MED14 interactions with RNA Pol II and TFIIF (Plaschka et al., 2015). However, the *S. cerevisiae* Med14 subunit from this study consisted of residues 1-755 (of 1082 residues in yeast Med14) and the human MED14 S1112, S1128, and S1136 residues do not appear to be conserved.

Knockdown experiments have implicated the CDK8 protein in the regulation of transcription elongation and/or RNA Pol II pausing or pause release (Donner et al., 2010; Galbraith et al., 2013). Furthermore, ChIP-Seq data from CA-treated MOLM-14 cells indicated a reduced RNA Pol II travel ratio (ratio of promoter-bound RNA Pol II vs. RNA Pol II in gene body) at genes whose expression was upregulated by CA (Pelish et al., 2015), implicating Mediator kinase activity in RNA Pol II pausing or pause release. The reduced travel ratio in CA-treated cells could also reflect inhibition of premature RNA Pol II termination. Here, we identified AFF4, NELFA, MED26, POLR2M, SETX, and XRN2 as high-confidence Mediator kinase targets, and each of these factors has been implicated in regulation of RNA Pol II pausing, premature termination, or elongation (Brannan et al., 2012; Cheng et al., 2012; Jishage et al., 2012; Kwak and Lis, 2013; Lin et al., 2010; Takahashi et al., 2011; Wagschal et al., 2012).

Mediator kinases as potential metabolic regulators

CDK8 orthologs in *Drosophila* and yeast have been linked to lipid and glucose metabolism and regulation of cellular responses to nutrient stress (Kuchin et al., 1995; Lindsay et al., 2014; Mousley et al., 2012; Zhao et al., 2012). Upon Mediator kinase inhibition by CA, we observed changes in the abundance of about 200 proteins (**Figure 4.6**), including many involved in basic metabolic pathways such as oxidative phosphorylation, fatty acid metabolism, and cholesterol homeostasis. MED13 and cyclin C appear to regulate mitochondrial function in yeast (Cooper et al., 2014; Khakhina et al., 2014), and over-expression of MED13 in mouse cardiac tissue alters fatty acid metabolism, β-oxidation, and mitochondrial content (Baskin et al., 2014). We identified MED13 and cyclin C as Mediator kinase substrates and observed an increase in MED13 protein levels upon CA treatment, which could contribute to altered fatty acid metabolism or oxidative phosphorylation observed in CA-treated cells (**Figure 4.6D**).

CDK8 kinase activity has previously been linked to cholesterol metabolism and fatty acid synthesis via regulation of SREBP. In particular, CDK8-dependent phosphorylation of SREBP residue T402 correlated with SREBP degradation in *Drosophila* and mouse cells (Zhao et al., 2012). GSEA of whole proteome data identified changes in the cholesterol homeostasis, adipogenesis, and fatty acid metabolism hallmark signatures in CA-treated cells (**Figure 4.6D**). Moreover, F-match identified SREBP binding motifs as over-represented among genes whose expression changed upon CA treatment (**Figure 4.5D**). Whereas phosphorylation of SREBP T402 was detected in our phosphoproteomics experiments, its level was not altered in CA-treated cells, suggesting alternate means of Mediator kinase-dependent SREBP regulation in HCT116 cells. Other kinases, including GSK3 (Sundqvist et al., 2005), are known to target SREBP T402 and we have confirmed that CA

does not inhibit GSK3β in cell lysates (Pelish et al., 2015) or in in vitro kinase assays with the purified protein (**Figure 4.4H**). Therefore, the SREBP T402 phosphorylation level may remain constant in CA-treated cells due to other kinases targeting this site. Alternately, SREBP may not be a substrate for CDK8 in HCT116 cells. SIRT1, a validated Mediator kinase target, can negatively regulate SREBP activity through deacetylation (Walker et al., 2010). The Mediator kinases phosphorylate SIRT1 at residue T530, and phosphorylation at T530 has been shown to activate the SIRT1 deacetylase (Sasaki et al., 2008). Thus, via SIRT1 and potentially other substrates, Mediator kinases may regulate cholesterol or fatty acid metabolism independent of direct SREBP phosphorylation in HCT116 cells.

Human Mediator kinases and transcription factor turnover

Previous studies revealed that phosphorylation of nutrient-responsive transcription factors Gcn4, Ste12, or Phd1 by yeast Cdk8 promoted their degradation (Chi et al., 2001; Nelson et al., 2003; Raithatha et al., 2012). Studies in metazoans have shown evidence for CDK8-dependent phosphorylation of the transcription factors SMAD1, SMAD3, Notch ICD, SREBP, E2F1, and STAT1 (Alarcon et al., 2009; Bancerek et al., 2013; Fryer et al., 2004; Morris et al., 2008; Zhao et al., 2012). Among these, increased degradation of the Notch ICD, SMAD1, SMAD3, and SREBP correlated with phosphorylation. For these reasons, we anticipated that inhibition of CDK8 and CDK19 kinase activity would affect the protein levels of a subset of their targets. Whole proteome data revealed no evidence that transcription factor phosphorylation by Mediator kinases affected their stability, even with analyses at 1, 3, 6, 18, or 24 hours of CA treatment. In fact, we found little evidence for altered stability of any high-confidence Mediator kinase targets, with the notable exception of MED13 and MED13L. Despite this result, cell type or context may be key factors that

dictate the effect of Mediator kinase phosphorylation on protein turnover. Here, we evaluated HCT116 cells in normal growth conditions whereas Mediator kinases may in fact more generally regulate substrate protein turnover during stress responses or at different developmental stages.

Whereas no changes in transcription factor turnover were evident from the whole proteome data in CA-treated versus control cells, we identified many links between the gene expression changes and the phosphoproteomics data. These results are consistent with Mediator kinases affecting transcription factor activity in HCT116 cells under normal growth conditions, rather than transcription factor turnover.

CDK8 as a colon cancer oncogene: Mediator kinase inhibition versus subunit knockdown

CDK8 was identified as a colon cancer oncogene in part through an shRNA screen for genes required for HCT116 cell proliferation (Firestein et al., 2008). CDK8 was one of 166 candidates in this screen; CDK8 was also identified in a screen for factors required for activation of a β-catenin-driven reporter in a different colon cancer line, DLD-1 (Firestein et al., 2008). Our analyses with CA indicate that, in contrast to CDK8 knockdown, Mediator kinase inhibition does not affect HCT116 cell growth (Pelish et al., 2015). These findings highlight the distinction between physical loss of a protein versus targeted inhibition of its enzymatic activity.

As a transcription factor, β-catenin assembles with the DNA-binding proteins TCF and LEF-1 to activate genes that drive cell proliferation. HCT116 cells are heterozygous for a mutant β-catenin protein that is resistant to degradation (Morin et al., 1997). Consequently, HCT116 cells have increased β-catenin levels and are considered "β-catenin-dependent". Consistent with an

oncogenic function for CDK8, CDK8 knockdown prevented activation of β-catenin target genes in colon cancer cell lines (Firestein et al., 2008). In addition, the E2F1 transcription factor has been shown to be an important negative regulator of β-catenin stability (through unknown mechanisms), and elevated levels of the CDK8 protein, as observed in HCT116 cells (Firestein et al., 2008), can block E2F1-dependent inhibition of β-catenin target gene expression (Morris et al., 2008). Thus, in colon cancer cells, the CDK8 protein appears to up-regulate β-catenin target gene expression in two ways: as a β-catenin co-activator and as an inhibitor of E2F1 activity.

Whereas E2F1 and β-catenin activity or stability is known to be regulated by phosphorylation, we did not observe significant changes in E2F1 or β-catenin protein or phosphopeptide levels in CA-treated HCT116 cells. An F-match analysis based upon gene expression changes in CA-treated cells, however, identified E2F binding motifs as over-represented (**Figure 4.5D**), and Ingenuity Pathway Analysis (IPA) of upstream regulators identified β-catenin target genes as over-represented among those whose expression increased or decreased upon CA treatment. GSEA of our proteomics data (CA-treated vs. untreated, 0 – 24 hours) revealed upregulation of both the E2F1 and β-catenin pathways (**Figure 4.6D**). Furthermore, numerous high-confidence Mediator kinase substrates are known to directly regulate β-catenin or E2F activity.

Although these results implicate Mediator kinase activity in the regulation of E2F1 and β-catenin transcription networks in HCT116 cells, the effects of Mediator kinase inhibition are clearly distinct from CDK8 or CDK19 knockdown (Donner et al., 2010; Firestein et al., 2008; Galbraith et al., 2013). This was not unexpected, as the physical presence of an enzyme typically

serves structural roles, such as maintaining the integrity of a multi-protein complex. For example, ablation of the CDK7 ortholog in yeast (Kin28) abolishes essentially all RNA Pol II transcription (Holstege et al., 1998), in contrast to targeted inhibition of Kin28 activity (Kanin et al., 2007). Direct comparison of the transcriptional changes resulting from physical loss of the CDK8 or CDK19 protein versus targeted inhibition of kinase activity (i.e., with protein levels remaining intact) revealed stark differences in both the genes affected and in the magnitude of gene expression changes. These differences highlight the importance of a structural or "scaffolding" role of the CDK8 or CDK19 proteins. Indeed, CDK8 knockdown decreases MED12 levels and increases CDK19 protein levels in HCT116 cells (Donner et al., 2010; Galbraith et al., 2013), which likely contributes to the distinct gene expression and anti-proliferative effects of CDK8 knockdown (Firestein et al., 2008) compared with kinase inhibition by CA. Because CA inhibits CDK19 as well as CDK8 (Pelish et al., 2015), this may also result in compensatory effects that distinguish the CDK8 knockdown phenotype from CDK8/19 inhibition. Future studies are needed to more precisely establish the roles of CDK8 vs. CDK19 in regulating the elaborate E2F1, β-catenin, and other inter-related signaling networks that contribute to HCT116 survival and proliferation.

Concluding remarks

This study provides a large-scale identification of Mediator kinase substrates and the impact of Mediator kinase activity on RNA Pol II transcription and the cellular proteome. In comparison with the ~170 potential CDK9 kinase substrates recently identified in HCT116 cells (Sanso et al., 2016), it is notable that the high-confidence substrates for CDK9 are distinct from the Mediator kinases. This further suggests that CDK9 (e.g., as part of P-TEFb or the super-elongation complex) and Mediator kinases play non-redundant roles in transcription regulation.

Our results were enabled by the rigorous biochemical, cellular, and biophysical characterization of CA, which demonstrated that it represents an unusual case of an inhibitor that is truly selective for Mediator kinases in human cells (Pelish et al., 2015). The data and methodologies presented provide a valuable resource for further delineation of the molecular mechanisms whereby Mediator kinases, and their substrates, regulate processes that are fundamentally important in human development and disease. For example, the methodologies described could be applied toward other cell types or contexts to uncover cell type- or context-specific roles for Mediator kinases. Alternately, the Mediator kinase targets or proteome changes identified here could be further tested for their mechanistic role(s) in regulating chromatin structure and function, DNA repair or replication, cell metabolism, or RNA Pol II transcription.

Conclusion

CA is a highly selective inhibitor of the Mediator kinases CDK8 and CDK19. Using CA, we now report a large-scale identification of Mediator kinase substrates in human cells (HCT116). We identified over 16,000 quantified phosphosites including 78 high-confidence Mediator kinase targets within 64 proteins, including DNA-binding transcription factors and proteins associated with chromatin, DNA repair, and RNA Pol II. Although RNA-seq data correlated with Mediator kinase targets, the effects of CA on gene expression were limited and distinct from CDK8 or CDK19 knockdown. Quantitative proteome analyses, tracking around 7,000 proteins across six time points (0–24 hours), revealed that CA selectively affected pathways implicated in inflammation, growth, and metabolic regulation. Contrary to expectations, increased turnover

of Mediator kinase targets was not generally observed. Collectively, these data support Mediator kinases as regulators of chromatin and RNA Pol II activity and suggest their roles extend beyond transcription to metabolism and DNA repair.

Methods

Cell culture

HCT116 cells were cultured in DMEM supplemented with 10% fetal bovine serum (FBS) and penicillin/streptomycin. Cells were maintained at 37°C and 5% CO_2.

SILAC labeling

HCT116 cells were cultured in DMEM lacking arginine and lysine (Pierce, 88420) supplemented with either Arg10 (33.6 μg/ml) and Lys8 (73 μg/ml) or Arg0 and Lys0 for heavy and light treatment, respectively. After six passages at 1:3 ratio, cells were tested for Arg/Lys incorporation and were subsequently supplemented with 200 mg/l of proline (Sigma-Aldrich, P5607) as a small amount of Arg→Pro conversion was detected. Cells were maintained in 10% dialyzed FBS and penicillin/streptomycin.

TiO_2 phosphopeptide enrichment, ERLIC chromatography, and LC-MS/MS

Protocols were carried out as described (Stuart et al., 2015). An Orbitrap LTQ (Thermo Fisher) was used for phosphoproteomics, and an Orbitrap Velos (Thermo Fisher) was used for quantitative proteome analysis.

Gene expression comparison between CA-treated HCT116 Cells and shRNA CDK8/19

shRNA CDK8 and CDK19 microarray data were obtained from the GEO (accession number: GSE38061), and data under the "normoxia" tab were used for the comparison to CA-treated cells.

Phosphoproteomics sample preparation

Cells were passaged 7-8 times in SILAC media on 15cm dishes. For each replicate, approximately 20 mg total protein was harvested for analysis after treatment with either 100 nM CA or DMSO for 1 hour. For one replicate, treatment conditions were reversed and light cells were CA treated and heavy cells were DMSO treated. To harvest, media was removed and each dish was scraped in 750 µl 95°C SDT (4% SDS, 100mM Tris pH 7.9, 10mM TCEP) buffer with subsequent heating at 95°C for 10 min. Lysates were sonicated for 1 - 2 minutes each. Protein concentrations were determined using a BCA assay and samples were mixed 1:1 based on total protein concentrations. FASP was carried out in two 10 kDa MWCO filters with a 50 mM iodoacetamide alkylation step and proteins were digested in 2 M urea with 2% wt/wt Lys-C (Wako) for 6 hours and 2% modified trypsin (Promega) for 12 hours at 37°C. FASP eluates were acidified and desalted on Oasis HLB extraction cartridges.

Phosphoproteomics and quantitative proteomics data analysis

All raw MS files for phosphoproteomics and quantitative proteomics were searched using the MaxQuant (v1.4.1.2) software package. Triplicate phosphoproteomic and duplicate proteomic treatments with a CA-time course were searched individually against the Uniprot human proteome database (downloaded on 1/27/2014) using the default MaxQuant parameters, except: multiplicity was set to 2 (heavy/light) with Arg10 and Lys8 selected, LysC/P was selected as an additional enzyme, "re-quantify" was unchecked, and Phospho (STY) was selected as a variable modification in both runs. For phosphosite analysis, the Phospho (STY) table was processed with Perseus (v1.4.1.3) using the following workflow: reverse and contaminant reads were removed, the site table was expanded to accommodate differentially phosphorylated peptides, and rows without any quantification were removed after site table expansion. For protein quantification with a CA treatment time course, the protein Groups table was processed similarly to the Phospho (STY) table, except that there was no need for expansion of the site table.

***In vitro* kinase assays**

Assays were done essentially as described (Bancerek et al., 2013). For validation of GST-tagged substrates identified in the phosphoproteomics experiments, the assay was done as described (Bancerek et al., 2013), except that the total reaction volume was reduced to 10 µl and samples were loaded on a pre-cast, 4-20% gradient gel (BioRad) and coomassie stained for loading control analysis. For CDK8 module kinase assays using SIRT1 as a substrate, His-SIRT1 (addgene 13735) was purified as described (Hallows et al., 2006) except that no eluate dialysis was performed, and assays were carried out using 1X PK buffer (NEB) with 100 µM cold ATP at 30°C for the indicated time. Reaction mixes were created such that each sample on the gel corresponded to 0.25

μl CDK8 module and 25 ng of SIRT1. Phosphorylation of SIRT1 T530 was detected using western blotting on a 4-20% precast gradient gel. Assays were done as previously described for the CDK1, ERK2, and GSK3β kinase assays. Purified CDK1-cyclin B and GSK3β were purchased from New England Biolabs (NEB) and Abcam, respectively. Phosphorylated ERK2 was a gift from the laboratory of Natalie Ahn.

RNA-seq

HCT116 cells were treated with either 100nM CA or DMSO for 3 hours (n = 3) and 24 hours (n=3). Cells were washed twice with cold PBS and scraped into TRIzol reagent (Life Technologies). After harvesting, the RNA was further purified using an RNeasy mini kit (Qiagen) with on-column DNase I digestion. Libraries for Illumina sequencing were generated via the Illumina TruSEQ stranded mRNA prep kit. Samples were run in a single lane on an Illumina HiSEQ 2000 sequencer with a single read flow cell using 1 × 50 bp reads and a 6-cycle index read. Reads were mapped to the hg19 reference genome using Tophat2 v.2.0.6 with custom settings. HTSeq v.9.6.1 was used to obtain read counts over annotated genes and differentially expressed genes were called using Cufflinks v.2.1.1.

STRING and Gene Ontology (GO) analyses

High confidence phosphosites identified as CDK8/19 substrates were submitted to the STRING database with human as the selected organism and the "multiple names" option. The confidence score was set to high (≥ 0.7). GO analyses were completed using the DAVID bioinformatics tool.

Hypergeometic test for Mediator kinase targets and Transfac matrices

To perform the hypergeometric test, we assumed that Mediator kinase target proteins (64) could be represented by any gene annotated in hg19 (23,669). Using all Transfac matrices available in the public database (398), we identified 8 binding sites of high confidence Mediator kinase targets, with a probability that this event could occur at random of 0.000011483.

HCT116 compound treatment for immunoblotting, and antibodies

HCT116 cells were grown in 6cm dishes to approximately 80% confluency and were treated with either DMSO or CA at 10, 50, 100, 500, 1,000 nM for 2 hours. Cells were harvested by scraping, whole-cell extracts were made using RIPA buffer, and protein amounts were quantified using a BCA assay. 30 μg total protein was loaded in each lane, and the following antibodies were used: MED12 (Bethyl A300-774A, 1:2000), Total SIRT1 (Santa Cruz sc-15404, 1:3000), SIRT1 T530 Phospho (Abcam ab156585, 1:1000), CCNC (Santa Cruz sc-1061, 1:1000), CDK8 (Santa Cruz sc-1521, 1:1000), and TBP (Santa Cruz sc-273, 1:1000). For evaluation of STAT1 S727 phosphorylation upon IFN gamma treatment with and without CA, HCT116 cells were grown in the same manner as described previously. Cells were pre-treated with either DMSO or increasing concentrations of CA for 45 minutes, and then treated with IFN-γ (eBioscience) at a final concentration of 10 ng/ml for 45 minutes. Cells were washed with cold PBS containing protease and phosphatase inhibitors, scraped in PBS, and spun at 2,000 rpm for 5 minutes at 4 °C. Cold RIPA buffer containing protease and phosphatase inhibitors was used to resuspend the cell pellets, and lysates were nutated at 4 °C for 15 minutes and subsequently sonicated in a cold water bath. Sonicated lysates were spun at 14,000 rpm at 4 °C and 30 μg cleared RIPA lysates were used for western blot analysis. For treatment with RO-3306 and SP600125, cells were incubated in the

presence of the indicated concentration of compound with or without CA for 2.5 hours. Cells were subsequently harvested and extracts were made according to above.

Autophosphorylation of CDK8 module and TiO_2 enrichment

Approximately 200 pmol of CDK8 module (CDK8, CCNC, MED12, and MED13) was expressed and purified as described (Knuesel et al., 2009) and incubated with 1 mM ATP in PK buffer (NEB) at 37°C for 2 hours. Total protein was TCA precipitated and the protein pellet was resuspended in 50 µl SDT buffer. The sample was heated for 5 min at 95 °C to completely solubilize the pellet, diluted in UA buffer, and small-scale FASP was carried out in a 30kDa MWCO filter (Amicon, 500

µl capacity). Protein was digested to peptides with both 2% Lys-C (6 hours) and trypsin (12 hours) wt/wt at room temperature. Samples were desalted on C18 spin columns (Pierce), acidified, and titanium dioxide enrichment of phosphopeptides was carried out with a 10:1 wt/wt ratio. Samples were subjected to 1D LC/MS/MS on an Orbitrap LTQ (Thermo Fisher). Raw files were searched with MaxQuant.

Transcription factor binding site (TFBS) analysis

Promoters for differentially expressed genes were extracted from UCSC Genome Browser and promoter analysis (+/- 2 kb from the TSS) was conducted with F-Match, a part of the Transfac public database. Only vertebrate Transfac matrices were selected for the output. Control promoters were taken from genes in the HCT116 dataset whose expression was unaffected with CA treatment.

References

Alarcon, C., Zaromytidou, A.I., Xi, Q., Gao, S., Yu, J., Fujisawa, S., Barlas, A., Miller, A.N., Manova-Todorova, K., Macias, M.J., et al. (2009). Nuclear CDKs drive Smad transcriptional activation and turnover in BMP and TGF-beta pathways. Cell *139*, 757–769.

Allen, B.L., and Taatjes, D.J. (2015). The Mediator complex: a central integrator of transcription. Nat. Rev. Mol. Cell Biol. *16*, 155–166.

Bancerek, J., Poss, Z.C., Steinparzer, I., Sedlyarov, V., Pfaffenwimmer, T., Mikulic, I., Dolken, L., Strobl, B., Muller, M., Taatjes, D.J., and Kovarik, P. (2013). CDK8 kinase phosphorylates transcription factor STAT1 to selectively regulate the interferon response. Immunity *38*, 250–262.

Baskin, K.K., Grueter, C.E., Kusminski, C.M., Holland, W.L., Bookout, A.L., Satapati, S., Kong, Y.M., Burgess, S.C., Malloy, C.R., Scherer, P.E., et al. (2014). MED13-dependent signaling from the heart confers leanness by enhancing metabolism in adipose tissue and liver. EMBO Mol. Med. *6*, 1610–1621.

Brannan, K., Kim, H., Erickson, B., Glover-Cutter, K., Kim, S., Fong, N., Kiemele, L., Hansen, K., Davis, R., Lykke-Andersen, J., and Bentley, D.L. (2012). mRNA decapping factors and the exonuclease Xrn2 function in widespread premature termination of RNA Pol II transcription. Mol. Cell *46*, 311–324.

Cevher, M.A., Shi, Y., Li, D., Chait, B.T., Malik, S., and Roeder, R.G. (2014). Reconstitution of active human core Mediator complex reveals a critical role of the MED14 subunit. Nat. Struct. Mol. Biol. *21*, 1028–1034.

Cheng, B., Li, T., Rahl, P.B., Adamson, T.E., Loudas, N.B., Guo, J., Varzavand, K., Cooper, J.J., Hu, X., Gnatt, A., et al. (2012). Functional association of Gdown1 with RNA Pol II poised on human genes. Mol. Cell *45*, 38–50.

Chi, Y., Huddleston, M.J., Zhang, X., Young, R.A., Annan, R.S., Carr, S.A., and Deshaies, R.J. (2001). Negative regulation of Gcn4 and Msn2 transcription factors by Srb10 cyclin-dependent kinase. Genes Dev. *15*, 1078–1092.

Colaert, N., Helsens, K., Martens, L., Vandekerckhove, J., and Gevaert, K. (2009). Improved visualization of protein consensus sequences by iceLogo. Nat. Methods *6*, 786–787.

Cooper, K.F., Khakhina, S., Kim, S.K., and Strich, R. (2014). Stress-induced nuclear-to-cytoplasmic translocation of cyclin C promotes mitochondrial fission in yeast. Dev. Cell *28*, 161–173.

Daniels, D.L., Ford, M., Schwinn, M.K., Benink, H., Galbraith, M.D., Amunugama, R., Jones, R., Allen, D., Okazaki, N., Yamakawa, H., et al. (2013). Mutual exclusivity of MED12/MED12L, MED13/13L, and CDK8/19 paralogs revealed within the CDK-Mediator kinase module. J. Proteomics Bioinform. *S2*, 004.

Davis, M.A., Larimore, E.A., Fissel, B.M., Swanger, J., Taatjes, D.J., and Clurman, B.E. (2013). The SCF-Fbw7 ubiquitin ligase degrades MED13 and MED13L and regulates CDK8 module association with Mediator. Genes Dev. *27*, 151–156.

Donner, A.J., Ebmeier, C.C., Taatjes, D.J., and Espinosa, J.M. (2010). CDK8 is a positive regulator of transcriptional elongation within the serum response network. Nat. Struct. Mol. Biol. *17*, 194–201.

Ebmeier, C.C., and Taatjes, D.J. (2010). Activator-Mediator binding regulates Mediator-cofactor interactions. Proc. Natl. Acad. Sci. USA *107*, 11283–11288.

Firestein, R., Bass, A.J., Kim, S.Y., Dunn, I.F., Silver, S.J., Guney, I., Freed, E., Ligon, A.H., Vena, N., Ogino, S., et al. (2008). CDK8 is a colorectal cancer oncogene that regulates beta-catenin activity. Nature *455*, 547–551.

Fryer, C.J., White, J.B., and Jones, K.A. (2004). Mastermind recruits CycC:CDK8 to phosphorylate the Notch ICD and coordinate activation with turnover. Mol. Cell *16*, 509–520.

Galbraith, M.D., Allen, M.A., Bensard, C.L., Wang, X., Schwinn, M.K., Qin, B., Long, H.W., Daniels, D.L., Hahn, W.C., Dowell, R.D., and Espinosa, J.M. (2013). HIF1A employs CDK8-mediator to stimulate RNAPII elongation in response to hypoxia. Cell *153*, 1327–1339.

Holstege, F.C., Jennings, E.G., Wyrick, J.J., Lee, T.I., Hengartner, C.J., Green, M.R., Golub, T.R., Lander, E.S., and Young, R.A. (1998). Dissecting the regulatory circuitry of a eukaryotic genome. Cell *95*, 717–728.

Hurlin, P.J., Steingrimsson, E., Copeland, N.G., Jenkins, N.A., and Eisenman, R.N. (1999). Mga, a dual-specificity transcription factor that interacts with Max and contains a T-domain DNA-binding motif. EMBO *J.18*, 7019–7028.

Imhof, A., Schuierer, M., Werner, O., Moser, M., Roth, C., Bauer, R., and Buettner, R. (1999). Transcriptional regulation of the AP-2alpha promoter by BTEB-1 and AP-2rep, a novel wt-1/egr-related zinc finger repressor. Mol. Cell. Biol. *19*, 194–204.

Jishage, M., Malik, S., Wagner, U., Uberheide, B., Ishihama, Y., Hu, X., Chait, B.T., Gnatt, A., Ren, B., and Roeder, R.G. (2012). Transcriptional regulation by Pol II(G) involving mediator and competitive interactions of Gdown1 and TFIIF with Pol II. Mol. Cell *45*, 51–63.

Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S., et al. (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature *467*, 430–435.

Kanin, E.I., Kipp, R.T., Kung, C., Slattery, M., Viale, A., Hahn, S., Shokat, K.M., and Ansari, A.Z. (2007). Chemical inhibition of the TFIIH-associated kinase Cdk7/Kin28 does not impair global mRNA synthesis. Proc. Natl. Acad. Sci. USA *104*, 5812–5817.

Khakhina, S., Cooper, K.F., and Strich, R. (2014). Med13p prevents mitochondrial fission and programmed cell death in yeast through nuclear retention of cyclin C. Mol. Biol. Cell *25*, 2807–2816.

Knuesel, M.T., Meyer, K.D., Bernecky, C., and Taatjes, D.J. (2009). The human CDK8 subcomplex is a molecular switch that controls Mediator coactivator function. Genes Dev. *23*, 439–451.

Kuchin, S., Yeghiayan, P., and Carlson, M. (1995). Cyclin-dependent protein kinase and cyclin homologs SSN3 and SSN8 contribute to transcriptional control in yeast. Proc. Natl. Acad. Sci. USA *92*, 4006–4010.

Kwak, H., and Lis, J.T. (2013). Control of transcriptional elongation. Annu. Rev. Genet. *47*, 483–508.

Lin, C., Smith, E.R., Takahashi, H., Lai, K.C., Martin-Brown, S., Florens, L., Washburn, M.P., Conaway, J.W., Conaway, R.C., and Shilatifard, A. (2010). AFF4, a component of the ELL/P-TEFb elongation complex and a shared subunit of MLL chimeras, can link transcription elongation to leukemia. Mol. Cell *37*, 429–437.

Lindsay, A.K., Morales, D.K., Liu, Z., Grahl, N., Zhang, A., Willger, S.D., Myers, L.C., and Hogan, D.A. (2014). Analysis of Candida albicans mutants defective in the Cdk8 module of mediator reveal links between metabolism and biofilm formation. PLoS Genet. *10*, e1004567.

Liu, Y., Kung, C., Fishburn, J., Ansari, A.Z., Shokat, K.M., and Hahn, S. (2004). Two cyclin-dependent kinases promote RNA Pol II transcription and formation of the scaffold complex. Mol. Cell. Biol. *24*, 1721–1735.

Luo, Z., Lin, C., and Shilatifard, A. (2012). The super elongation complex (SEC) family in transcriptional control. Nat. Rev. Mol. Cell Biol. *13*, 543–547.

Macek, B., Mann, M., and Olsen, J.V. (2009). Global and Site-Specific Quantitative Phosphoproteomics: Principles and Applications. Annu. Rev. Pharmacol. Toxicol. *49*, 199–221.

Margolin, A.A., Ong, S.E., Schenone, M., Gould, R., Schreiber, S.L., Carr, S.A., and Golub, T.R. (2009). Empirical Bayes analysis of quantitative proteomics experiments. PLoS ONE *4*, e7454.

Morin, P.J., Sparks, A.B., Korinek, V., Barker, N., Clevers, H., Vogelstein, B., and Kinzler, K.W. (1997). Activation of beta-catenin-Tcf signaling in colon cancer by mutations in beta-catenin or APC. Science *275*, 1787–1790.

Morris, E.J., Ji, J.Y., Yang, F., Di Stefano, L., Herr, A., Moon, N.S., Kwon, E.J., Haigis, K.M., Naar, A.M., and Dyson, N.J. (2008). E2F1 represses beta-catenin transcription and is antagonized by both pRB and CDK8. Nature *455*, 552–556.

Mousley, C.J., Yuan, P., Gaur, N.A., Trettin, K.D., Nile, A.H., Deminoff, S.J., Dewar, B.J., Wolpert, M., Macdonald, J.M., Herman, P.K., et al. (2012). A sterol-binding protein integrates endosomal lipid metabolism with TOR signaling and nitrogen sensing. Cell *148*, 702–715.

Nelson, C., Goto, S., Lund, K., Hung, W., and Sadowski, I. (2003). Srb10/Cdk8 regulates yeast filamentous growth by phosphorylating the transcription factor Ste12. Nature *421*, 187–190.

Pelish, H.E., Liau, B.B., Nitulescu, I.I., Tangpeerachaikul, A., Poss, Z.C., Da Silva, D.H., Caruso, B.T., Arefolov, A., Fadeyi, O., Christie, A.L., et al. (2015). Mediator kinase inhibition further activates super-enhancer-associated genes in AML. Nature *526*, 273–276.

Plaschka, C., Lariviere, L., Wenzeck, L., Seizl, M., Hemann, M., Tegunov, D., Petrotchenko, E.V., Borchers, C.H., Baumeister, W., Herzog, F., et al. (2015). Architecture of the RNA Pol II-Mediator core initiation complex. Nature *518*, 376–380.

Raithatha, S., Su, T.C., Lourenco, P., Goto, S., and Sadowski, I. (2012). Cdk8 regulates stability of the transcription factor Phd1 to control pseudohyphal differentiation of Saccharomyces cerevisiae. Mol. Cell. Biol. *32*, 664–674.

Riley, N.M., and Coon, J.J. (2016). Phosphoproteomics in the Age of Rapid and Deep Proteome Profiling. Anal. Chem. *88*, 74–94.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47.

Sanso, M., Levin, R.S., Lipp, J.J., Wang, V.Y., Greifenberg, A.K., Quezada, E.M., Ali, A., Ghosh, A., Larochelle, S., Rana, T.M., et al. (2016). P-TEFb regulation of transcription termination factor Xrn2 revealed by a chemical genetic screen for Cdk9 substrates. Genes Dev. *30*, 117–131.

Sasaki, T., Maier, B., Koclega, K.D., Chruszcz, M., Gluba, W., Stukenberg, P.T., Minor, W., and Scrable, H. (2008). Phosphorylation regulates SIRT1 function. PLoS ONE *3*, e4020.

Sato, S., Tomomori-Sato, C., Parmely, T.J., Florens, L., Zybailov, B., Swanson, S.K., Banks, C.A.S., Jin, J., Cai, Y., Washburn, M.P., et al. (2004). A set of consensus mammalian mediator subunits identified by multidimensional protein identification technology. Mol. Cell *14*, 685–691.

Stuart, S.A., Houel, S., Lee, T., Wang, N., Old, W.M., and Ahn, N.G. (2015). A phosphoproteomic comparison of B-RAFV600E and MKK1/2 inhibitors in melanoma cells. Mol. Cell. Proteomics 14, 1599–1615.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550.

Sundqvist, A., Bengoechea-Alonso, M.T., Ye, X., Lukiyanchuk, V., Jin, J., Harper, J.W., and Ericsson, J. (2005). Control of lipid metabolism by phosphorylation-dependent degradation of the SREBP family of transcription factors by SCF(Fbw7). Cell Metab. *1*, 379–391.

Svaren, J., Sevetson, B.R., Apel, E.D., Zimonjic, D.B., Popescu, N.C., and Milbrandt, J. (1996). NAB2, a corepressor of NGFI-A (Egr-1) and Krox20, is induced by proliferative and differentiative stimuli. Mol. Cell. Biol. *16*, 3545–3553.

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P., et al. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. *43*, D447–D452.

Taatjes, D.J., Naar, A.M., Andel, F., 3rd, Nogales, E., and Tjian, R. (2002). Structure, function, and activator-induced conformations of the CRSP coactivator. Science *295*, 1058–1062.

Takahashi, H., Parmely, T.J., Sato, S., Tomomori-Sato, C., Banks, C.A., Kong, S.E., Szutorisz, H., Swanson, S.K., Martin-Brown, S., Washburn, M.P., et al. (2011). Human mediator subunit MED26 functions as a docking site for transcription elongation factors. Cell *146*, 92–104.

Ubersax, J.A., and Ferrell, J.E., Jr. (2007). Mechanisms of specificity in protein phosphorylation. Nat. Rev. Mol. Cell Biol. 8, 530–541.

Wagschal, A., Rousset, E., Basavarajaiah, P., Contreras, X., Harwig, A., Laurent-Chabalier, S., Nakamura, M., Chen, X., Zhang, K., Meziane, O., et al. (2012). Microprocessor, Setx, Xrn2, and Rrp6 co-operate to induce premature termination of transcription by RNAPII. Cell *150*, 1147–1157.

Walker, A.K., Yang, F., Jiang, K., Ji, J.Y., Watts, J.L., Purushotham, A., Boss, O., Hirsch, M.L., Ribich, S., Smith, J.J., et al. (2010). Conserved role of SIRT1 orthologs in fasting-dependent inhibition of the lipid/cholesterol regulator SREBP. Genes Dev. *24*, 1403–1417.

Xu, W., Wang, Z., Zhang, W., Qian, K., Li, H., Kong, D., Li, Y., and Tang, Y. (2015). Mutated K-ras activates CDK8 to stimulate the epithelial-to-mesenchymal transition in pancreatic cancer in part via the Wnt/b-catenin signaling pathway. Cancer Lett. *356* (2PtB), 613–627.

Zhao, X., Feng, D., Wang, Q., Abdulla, A., Xie, X.J., Zhou, J., Sun, Y., Yang, E.S., Liu, L.P., Vaitheesvaran, B., et al. (2012). Regulation of lipogenesis by cyclin-dependent kinase 8-mediated control of SREBP-1. J. Clin. Invest. *122*, 2417–2427.

Chapter 5

Conclusions and Prospects

Our group has used CA as a central tool to interrogate the biology of CDK8/19 as well as to understand why some AML cells are dependent on CDK8/19 kinase activity. We first established in **Chapter 2** that CA is a potent and selective inhibitor of CDK8/19, and then used it to uncover the negative regulatory function of CDK8/19 in super-enhancer-associated gene expression in AML. Inhibition of CDK8/19 upregulates these genes, many of which have tumor suppressor functions, causing growth inhibition both in cell culture and in animal models. In **Chapter 3**, we used a genome-wide CRISPR-Cas9 suppressor screen to discover that the antiproliferative activity of CA in certain AML cells is mediated by Notch signaling. CDK8/19-mediated NICD1/2 degradation is a principal means by which some AML cells silence the Notch pathway, and inhibition of this process reactivates the Notch tumor suppressor function. In **Chapter 4**, we used CA in a phosphoproteomics experiment to identify previously unknown substrates of CDK8/19. CDK8/19-catalyzed phosphorylation may play a functionally important role in regulating the activity or stability of these substrates, and may explain why some cancer cells are dependent on these kinases. Thanks to CA, the biological complexity of CDK8/19 is beginning to be unraveled, and their functions in transcription and disease illuminated. There are, however, still many unanswered research questions that can be addressed using the tools that we have developed throughout this dissertation.

The first question concerns the difference between CDK8 and CDK19. These two kinases are paralogous and hence highly similar, with 77% identity overall and 94% in the catalytic domain. It is still unclear what distinct roles, if there are any, CDK8 and CDK19 play in cell biology. In our studies, dual inhibition of CDK8/19 with CA suppresses the growth of many AML cells. This inhibitory effect can be almost fully rescued by ectopic expression of just one drug-resistant mutant

(CDK8-W105M or CDK19-W105M), suggesting that CDK8 and CDK19 may play redundant roles in the proliferation of AML cells. To ensure that the desensitization effect is not an artefact of overexpression, a more stringent experiment would involve using CRISPR-Cas9 to edit the CDK8 or CDK19 locus without altering expression levels. Cells bearing a CDK8-W105M mutation would serve as an excellent tool for studying the function of CDK19 with CA, and vice versa. This strategy would enable us to deconvolute the distinct roles that CDK8 and CDK19 play in determining cell sensitivity to CDK8/19 inhibition, as well as in transcriptional regulation, substrate recognition, and other processes.

The second question deals with the different mechanisms by which CDK8/19 inhibition suppresses the growth of AML cells. So far, our lab has proposed two models involving CDK8/19 regulation of STAT1 activation and NICD1 degradation, and these models explain the sensitivity of only a small fraction of all AML cell lines that we have studied. Understanding mechanisms is an important step toward predicting which patients would respond to CDK8/19 inhibitor therapy, as well as toward discovering novel targets for therapeutic intervention. For example, our CRISPR screen reveals that *DLL*/*JAG* expression is a predictive marker of whether the cell would respond to CA through the Notch axis. It also reveals that Notch signaling activation is a possible and perhaps underexplored strategy for the treatment of AML. Our experience suggests that CA behaves well in unbiased genome-wide CRISPR-Cas9 screens. By performing such screens in more cell lines and patient samples, other mechanisms of CA in AML cells may emerge. Some of these mechanistic models may explain the activity of CA more generally than the two models that we have already proposed.

The requirement of *DLL*/*JAG* expression for MOLM-14's sensitivity to CA is a powerful and potentially useful observation. Based on our findings, Notch pathway activation may be generally growth inhibitory to AML cells, and one way to activate Notch signaling would be to stimulate the cleavage of Notch receptors into NICDs using Dll/Jagged ligands. A recently reported small molecule, *N*-methylhemeanthidine chloride, which has been proposed to stimulate Notch1 processing into NICD1, may be used in place of Dll/Jagged ligands. The NICDs that are generated, however, can still get degraded in a CDK8/19-dependent process, negating their tumor suppressor potential. Dual treatment with Dll/Jagged ligands and CA may potentiate Notch signaling in a synergistic manner, offering the possibility of a combination therapy. This hypothesis can be validated in cell culture by, for example, testing whether AML cells grown in the presence of immobilized Notch ligands become hypersensitive to CA, or whether ectopic expression of Notch ligands further sensitizes AML cells to CA. We also have to keep in mind that the *in vivo* microenvironment may be very different from that in an isolated cell culture. Notch ligands expressed on other cells in the body, such as the bone marrow niche or other blood cells, may be able to stimulate Notch cleavage in circulating AML cells, obviating the need to develop agents to stimulate NICD production. It would be interesting to study whether CDK8/19 are required for Notch suppression in AML patient samples, as this would suggest that the Notch reactivation mechanism has wider implications in the clinic than in isolated cell culture. In brief, a deeper and more extensive understanding of the mechanism would open doors to new avenues for therapeutic development.

There is also a missing link between Notch and super-enhancer regulation. The effects of CA on Notch signaling and on super-enhancer-associated genes in MOLM-14 cells are both

striking, but we have not investigated the mechanistic connection between these two events. Because Notch itself is a transcriptional co-activator, it could conceivably mediate CA's ability to modulate super-enhancer activity. To test this hypothesis, ChIP-seq experiments could be performed to determine the loci where NICD1/2 bind in AML cells, and whether these loci coincide with super-enhancers as called by Mediator occupancy. The NICD1/2 could be generated either by CA treatment or by doxycycline-inducible plasmids. Moreover, we could examine whether genes under the control of NICD1/2 are enriched in super-enhancer association by measuring global gene expression changes upon NICD1/2 induction using RNA-seq, in a manner similar to how we discovered that CA-upregulated genes were enriched in super-enhancer association. Using an NICD1- or NICD2-specific plasmid, it would be possible to tease apart the different roles that these receptors play in suppressing AML growth.

Our CRISPR screen also identified 82/97 proteins not known to be part of the Notch pathway. It is likely that many of them, such as transcriptional co-activators and sugar-processing enzymes, actually participate in Notch signaling but have never been studied in such contexts. Further studies into these proteins would expand our knowledge of Notch signaling and reveal novel targets for Notch-pathway modulation. We can investigate the consequences of knocking out *GANAB*, *UXS1*, *TGDS*, or *UGCG* on Notch1 maturation, glycosylation, localization, activation, and transcriptional functions, in a manner similar to how we studied the impact of *PSEN1* or *MAML1* knockout on NICD1 and Hes1. More intriguingly, 9/97 of the hits from our screen encode proteins that are part of the human SAGA/STAGA histone acetyl transferase complex, which is a broad-spectrum transcriptional co-activator (*KAT2A, TADA1, TADA2B, USP22, TAF6L, TAF5L, SUPT7L, SUPT20H,* and *ATXN7L3*). Given our previous findings that CA upregulates super-

enhancer-associated gene expression, it would be interesting to study whether this property of CA is mediated by the STAGA complex together with other co-activators such as EP300 and CREBBP.

Lastly, we have identified ~60 previously unknown substrates of CDK8/19. Nearly all of these substrates are associated either directly or indirectly with the chromatin. Their phosphorylation by CDK8/19 may have important functional consequences, and further studies would greatly expand our understanding of the cell biological roles of CDK8/19. For example, we can test whether phosphorylation of DNA repair proteins TP53BP1 and RIF1 affects their ability to engage in double-strand break repairs, or whether phosphorylation of transcription factors ELK1 and FOXC1 affects their transcriptional activity or lifetimes. We can test these hypotheses by mutating the acceptor serine/threonine residue into alanine, which would keep the substrate in a permanently unphosphorylated state and phenocopy CA treatment. Alternatively, we can mutate it into aspartate/glutamate, creating a permanently negatively-charged phosphomimetic that should negate the effects of CA. Because CDK8/19 regulate the turnover of many proteins (including NICDs and Smads), a phosphoproteomics experiment in conjunction with proteasome inhibition may allow us to identify even more substrates that would otherwise get degraded. It might also be helpful to perform the experiment in AML cell lines in addition to HCT-116, as this would take into account lineage-specific effects, giving us the results that are more relevant to the pathology of this disease.