Regulation of rare genetic variation in human disease

Carter Wright

Follow this and additional works at: https://louis.uah.edu/uah-dissertations

Recommended Citation
https://louis.uah.edu/uah-dissertations/271

This Dissertation is brought to you for free and open access by the UAH Electronic Theses and Dissertations at LOUIS. It has been accepted for inclusion in Dissertations by an authorized administrator of LOUIS.
REGULATION OF RARE GENETIC VARIATION IN HUMAN DISEASE

Carter Wright

A DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in The Biotechnology Science and Engineering Program to The Graduate School of The University of Alabama in Huntsville

May 2023

Approved by:

Dr. Sara J. Cooper, Research Advisor
Dr. Paul Wolf, Committee Chair
Dr. Jerome Baudry, Committee Member
Dr. Richard M. Myers, Committee Member
Dr. Joseph Ng, Committee Member
Dr. Matthew Niemiller, Department Chair
Dr. Jon Hakkila, College Dean
Dr. Jon Hakkila, Graduate Dean
Abstract
REGULATION OF RARE GENETIC VARIATION IN HUMAN DISEASE
Carter Wright
A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
The Biotechnology Science and Engineering Program
The University of Alabama in Huntsville
May 2023

Genomic technologies are used to analyze genetic information. The work described focused on multiple sequencing applications to evaluate the impact of genetic variation in patients with disease. Whole genome sequencing, RNA-sequencing, and ChIP-sequencing were performed to identify variants, understand gene expression, and determine the differences in chromatin structure that contribute to cellular and organismal phenotypes. My thesis work describes the use of computational tools to analyze genomic data with the goal of understanding acquired genetic and epigenetic variation in cancer as well as inherited variation associated with rare neurological disease. Whole genome sequencing yielded a diagnostic rate of 53 percent in patients with rare disease. In our pancreatic cancer study, I identified genes that are regulated by HDAC1 overexpression and contribute to drug resistance. This work revealed an improved mechanistic understanding of the role of HDAC1 in chemoresistance and nominated GTPase genes as potential therapeutic targets. Both studies show how advancing genomic technologies can be used to study human disease.
Acknowledgments

I would like to thank my committee members, Dr. Sara J. Cooper, Dr. Richard M. Myers, Dr. Paul Wolf, Dr. Jerome Baudry, Dr. Joseph Ng, and Dr. Devin Absher for their guidance and expertise during my Ph.D.

I express extreme gratitude to Jim Hudson and Lonnie McMillian for building the unique and collaborative environment that makes HudsonAlpha one of a kind. My HudsonAlpha story began on a tour led by Jennie Hudson. I walked into the building with no expectations and I left inspired to one day work at HudsonAlpha. Jennie, thank you for the experience that influenced my path forward and led me to this accomplishment.

I’m extremely grateful to Dr. Sara Cooper for her mentorship and support throughout my Ph.D. journey. I have grown as a scientist and person because of your guidance, patience, and unwavering support. You allowed me to explore my passion for clinical genetics and pushed me to find my potential as a scientist.

I would like to extend my sincere thanks to Dr. Emily Gordon for her invaluable feedback and advice. Your work ethic and dedication to helping others are qualities that I admire and greatly contributed to my success. I am grateful for OTBX connecting us, Gus for always keeping the spirits high, and I especially want to thank you and Sara for making me feel like a part of the lab family from day one. You are a remarkable scientist, and an inspiration, and I am honored to call you my friend.

Finally, I would like to express my deepest appreciation to a very special duo, my biggest supporters, and the ones who made sure I knew my potential to make a difference.

Mom,

Thank you for the sacrifices you made for our family so I could accomplish my goals and have the privilege to write this today. Thank you for being a role model, pushing me to chase my “pie in the sky” career, and supporting me on every journey I have taken.

Dad,

Thank you for your encouragement and always having a smile on your face, regardless of my failure or success. Thank you for giving me a space to reach for the stars without the fear of failure. Your humor brought the laughter I needed throughout this journey and I will forever be thankful for it.
Dedication

For Papa John,
   We did it. And this is just the beginning of a great story, one that started with you instilling in me that it did not matter what I do in life as long as I strive to be my best self and make a difference in the world. On my daily commute down John Wright Drive, I am reminded that parts of you are still here with me. I am proud and honored to be the second Dr. Wright at UAH. To my grandparents, John Wright and Mac Wright, Robert and Keitha Mundy, I dedicate this to your loving memory.

In memory of Deven Kujath,
   You taught me how to live joyfully, the importance of laughing every day, that memories are something no one can take from you, and vulnerability is a superpower that will allow you to make lifelong connections.
Table of Contents

Abstract .............................................................................................................................. ii

Acknowledgements ....................................................................................................... iv

Dedication ........................................................................................................................ v

Table of Contents ............................................................................................................ vi

List of Figures .................................................................................................................. x-xi

List of Tables ................................................................................................................... xii

Chapter 1. Introduction .................................................................................................... 1

1.1 Overview ................................................................................................................... 1

1.2 Applications of Genomic Technology in Precision Medicine ............................... 1

1.2.1 Technologies ....................................................................................................... 1

1.2.2 Cancer Hallmarks and Genomics ........................................................................ 2

1.2.3 Known Pathways of Chemoresistance ............................................................... 4

1.2.4 Preliminary Results .............................................................................................. 5

1.3 Analysis of Whole Genome Sequence for Disease-causing Variants .............. 6

1.3.1 Variant Types ..................................................................................................... 6

1.3.2 Annotation of Variants ....................................................................................... 7

1.3.3 Classification of Variants .................................................................................... 7

1.3.4 Using Whole Genome Sequencing for Disease Diagnosis ............................. 8

Chapter 2. Multi-omic Analysis of HDAC1 Function in Pancreatic Cancer Reveals Altered GTPase Activity Impacting Chemoresistance ................................................................. 9

2.1 Introduction ................................................................................................................. 9
2. Background........................................................................................................9

2.3 HDAC1 Overexpression Leads to Multi-drug Resistance.................................13

2.4 HDAC1 Overexpression Induces Expression of Markers of EMT In Vitro and in Human PDAC Tissues .................................................................15

2.5 HDAC1 Overexpression Alters Chromatin Accessibility in Distal Enhancer and Promoter Regions Nearby Molecular Switches ........................................20

2.6 A Nine-gene Signature of HDAC1 Regulated Genes Predicts PDAC Patient Survival ........................................................................................................24

2.7 HDAC1 Overexpression Leads to Increased GTPase Activity and Chemoresistance .................................................................24

2.8 Methods............................................................................................................28

2.8.1 Cell Culture.............................................................................................28

2.8.2 Plasmids..................................................................................................28

2.8.3 sgRNA Cloning........................................................................................28

2.8.4 DsiRNA..................................................................................................28

2.8.5 GTPase-Glo Assay Using Cell Lysates.....................................................30

2.8.6 ChIP-sequencing.....................................................................................31

2.8.7 3’ RNA-sequencing................................................................................31

2.8.8 Drug Resistance Screening......................................................................32

2.8.9 Cell Staining............................................................................................33

2.8.10 Enrichment Analysis...............................................................................33

2.8.11 Survival Analysis....................................................................................33

2.8.12 Clinical Data and Samples......................................................................34

2.8.13 Data Access............................................................................................34
2.8.14 Statistical Testing.................................................................34
2.8.15 LASSO Model Selection......................................................35
2.8.16 Data Analysis......................................................................36

2.9 Discussion...............................................................................36

Chapter 3. Rare Variant and Polygenic Contributions to Early-onset or Atypical Dementia Risk.........................................................41

3.1 Introduction...............................................................................41
3.2 Results.....................................................................................42
3.3 Clinical Presentation and Family History......................................45
3.4 Genomic Analyses......................................................................46
  3.4.1 Pathogenic Diagnoses ............................................................46
  3.4.2 Compound Heterozygote Risk Factors.......................................50
  3.4.3 Rare Variants of Uncertain Significance......................................52
3.5 Polygenic Risk Score..................................................................54
3.6 Discussion................................................................................57

3.7 Methods..................................................................................59
  3.7.1 C9orf72 Expansion Testing..................................................59
  3.7.2 Polygenic Risk Score..............................................................59
  3.7.3 Genome Sequencing...............................................................60
  3.7.4 Data Processing and Quality Control.........................................60
  3.7.5 Data Analysis..........................................................................61
  3.7.6 Genomic Analysis.................................................................61
  3.7.7 Return of Results.................................................................62
  3.7.8 Data Access ........................................................................62
Chapter 4. Conclusions.................................................................64

4.1 Understanding HDAC1’s Role in Cancer........................................64
  4.1.1 Summary of Results and Limitations..................................64
  4.1.2 Future Studies and Applications........................................65

4.2 Genetic and Rare Variation Contributing to Neurodegenerative Disease........65
  4.2.1 Summary of Results and Challenges..................................65
  4.2.2 Future Research Recommendations.................................67

4.3 WGS for Inherited Cancer Risk Variants........................................67
  4.3.1 Study Design.............................................................67
  4.3.2 Challenges and Limitations of WGS for Cancer Risk.............68

References.....................................................................................70

Appendix A: Supplemental Figures 2.1 - 2.10.........................................84
Appendix B: Supplemental Tables 2.1 - 2.10........................................95
Appendix C: Supplemental Figures C.1 - C.6 and UAB IRB Protocol........176
Appendix D: ACMG Pathogenicity Evidence Details.............................183
Appendix E: Supplemental Tables 3.1 - 3.2.........................................190
Appendix F: Biorender Publication Licenses.........................................199
List of Figures

1.1 Schematic illustration of chromatin remodeling........................................3
1.2 Summary of two HDAC1 containing multi-protein chromatin remodeling complexes.................................................................5
2.1 Quantification of cell viability in PDAC cell lines following treatment of chemotherapeutics.................................................................14
2.2 HDAC1 overexpression is associated with increased expression of EMT genes.......................................................................................16
2.3 Genes altered by increased HDAC1 expression in PDAC cell lines and TCGA PDAC samples are associated with patient survival.........................19
2.4 ChIP-sequencing reveals DEG associated with patient survival.....................23
2.5 HDAC1 overexpression is associated with increased GTPase activity.............27
3.1 Summary of genomic sequencing findings..................................................43
3.2 Evaluation of five probands with a MAPT R406W variant...............................49
3.3 Boxplot comparing PRS between cohorts...................................................55
A.1 Normalized gene expression of HDACs.....................................................85
A.2 Analysis of differentially expressed genes revealed pathways altered by HDAC1 overexpression.................................................................86
A.3 Overall survival of TCGA PDAC patients grouped by HDAC1 expression........87
A.4 Overall survival of HDAC1 regulated genes in an independent cohort............88
A.5 Survival analysis of the top 322 DEG in cell lines and TCGA PDAC tumors.....89
A.6 Normalized gene expression of \textit{RAP2B} in PDAC cell lines and TCGA PDAC tumors………………………………………………………………………….90

A.7 H3K27 acetylation is more abundant near upregulated genes altered by HDAC1 overexpression………………………………………………………………………..91

A.8 Analysis of DEG with increased HDAC1 binding and H3K27 acetylation in their promoter………………………………………………………………………….92

A.9 \textit{HDAC1} overexpression is associated with upregulated expression of genes involved in GTPase activity and worse overall patient survival………………..93

A.10 Expression of genes in MAPK and PI3K pathways are increased upon \textit{HDAC1} overexpression………………………………………………………………………..94

C.1 Comparison of PRS with \textit{APOE} between combined cohorts…………………..177

C.2 Comparison of PRS with \textit{APOE} within cohorts……………………………………178

C.3 Comparison of PRS subset by finding……………………………………………………179

C.4 Correlation between family history score and PRS……………………………………180

C.5 Comparison of PRS subset by ethnicity………………………………………………….181
**List of Tables**

3.1 Variant Table........................................................................................................44  
B.1 Gene enrichment analysis of 322 DEG.................................................................101  
B.2 Gene Set Enrichment Analysis of 1,040 DEG.....................................................108  
B.3 DESeq2 results of 598 DEG (HDAC1 included) with increased HDAC1 binding and H3K27ac acetylation in the promoter..................................................114  
B.4 Gene set enrichment analysis results using 597 DEG with increased HDAC1 binding and H3K27ac acetylation in promoter......................................................126  
B.5 Comparison of GTPase activity between cell lines..............................................175  
E.1 Phenotype information for the 100 patient cohort..............................................191  
E.2 PRS with and without APOE.................................................................................196
Chapter 1. Introduction

1.1 Overview

The laboratory and clinical experiences described in this dissertation support my long-term professional goal of becoming a director of a Clinical Molecular Genetics laboratory. This career will require skills related to precision medicine, genetics, and genomics applied to oncology, neurobiology, and rare disease. In my thesis work, I gained experience with various genomic technologies and analysis tools to better understand the role of genetic variation in human disease.

1.2 Applications of Genomic Technology in Precision Medicine

1.2.1 Technologies

Over the past two decades, advances in genomic sequencing have allowed scientists to gain a more comprehensive view of variation in the human genome (Giani et al., 2020). Genomic technologies including whole genome sequencing (WGS), RNA-sequencing (RNA-seq), and Chromatin Immunoprecipitation-sequencing (ChIP-seq), are needed to fully understand the genetic architecture of disease. WGS is a comprehensive method that generates the entire genomic sequence of the organism being studied (Ng et al., 2010). Geneticists can then use that genomic sequence in combination with phenotyping information to make associations between genes and diseases. For example,
the gene *PAH* was associated with the metabolite phenylalanine and variants in *PAH* can lead to a treatable disease called phenylketonuria (Hou et al., 2020). RNA-seq uses next generation sequencing to count transcripts and measure gene expression within a cell (Kukurba et al., 2015). Comparison of gene expression between cells or tissue types provides insight on the role of genes in a given state. ChIP-seq, a technique to identify protein-DNA associations, also uses next generation sequencing to map the genomic location of transcription factor binding and histone modifications (Landt et al., 2012). ChIP-seq data can be analyzed to determine the differences in chromatin occupancy or structure between a cancer and control cell line, thus providing information on transcriptional gene regulation (Hurtado et al., 2011). The combination of these technologies has allowed genome scientists to better understand and interpret genetic variation by identifying and measuring changes in gene function and human phenotypes.

### 1.2.2 Cancer Hallmarks and Genomics

Hallmarks of cancer include resisting cell death, metastasis, tumor invasion, senescence, and epigenetic reprogramming (Hanahan, 2022). We can use genomic technologies like WGS, RNA-seq and ChIP-seq to observe these significant changes in genetic and genomic structure associated with the development and progression of cancer largely driven by these key characteristics. Importantly, the progression of cancer often comes with the development of chemoresistance. One mechanism of chemoresistance in cancer is epigenetic reprogramming which includes chromatin remodeling. Chromatin remodeling, the rearrangement of chromatin from a condensed state to an open accessible state and vice-versa, influences the regulation of transcription which includes both activation and repression of gene expression (Figure 1.1). These proteins are responsible
for modifying the chromatin architecture to alter the DNA into a transcriptionally accessible state (Nair et al., 2012). In cancer, the dysregulation of chromatin remodeling genes causes genome wide gene expression changes. For example, the transcriptional activation of HDAC1, a chromatin remodeling gene involved in multiple complexes, is associated with drug resistance in pancreatic cancer (Ramaker et al., 2021).

Figure 1.1: Schematic illustration of chromatin remodeling. Shown is the role of histone acetyl transferases (HATs) and histone deacetylases (HDACs) in transcriptional regulation. Figure was generated at biorender.com.
1.2.3 Known Pathways of Chemoresistance

Cancer cells can have intrinsic resistance to treatment or they can acquire resistance through the accumulation of somatic mutations during treatment and progression. The variants identified through whole genome sequencing are a combination of somatic variants in the tumor or germline variation in the patient. Chemoresistance is largely explained by the accumulation of somatic variation including single nucleotide mutations, copy number changes and altered gene expression which all contribute to the changing genetic architecture. The landscape of genetic contributors to disease and resistance has been investigated using whole genome sequencing (Quanli Wang et al., 2021). For example, WGS was used to identify regions enriched for deletions in the cancer genome, thus giving insight to the genomic architecture of cancer (Hama et al., 2018). A subset of these changes contributes to treatment response. Cancer is a multifaceted disease with a very complex architecture of many genetic contributors that leads to dysregulation of gene expression and ultimately contributes to the classic hallmarks of cancer. Often this dysregulation can occur due to chromatin changes, which can be evaluated using ChIP-sequencing. These variants can negatively impact the regulation of gene expression. Larger structural variants, such as CNVs, can alter chromatin architecture and gene expression in a genomic region containing genes involved in chemoresistance pathways, for example DNA repair, chromatin remodeling, or apoptosis (Shlien et al., 2009). Other variants may lead to a change in protein function or even disrupt a transcription factor binding site, thus preventing the transcription factor from binding leading to the dysregulation of gene expression. All of this dysregulation can contribute to chemoresistance (Dratwa et al., 2020; Lønning et al., 2013).
1.2.4 Preliminary Results

Our lab’s previous work used a genome-wide CRISPR screen to identify genes that could lead to treatment resistance in vitro using two pancreatic cancer cell lines. During my first months in the lab, I contributed to this by performing ChIP-seq targeting HDAC1 to identify the differences in HDAC1 chromatin occupancy in cell lines with overexpression of HDAC1. We identified an enrichment for HDAC1 binding near the transcription start site of genes differentially expressed following the activation of HDAC1.

My thesis work continued to characterize the role for HDAC1, a chromatin remodeling gene involved in chemoresistance (Figure 1.2).

![Figure 1.2: Summary of two HDAC1 containing multi-protein chromatin remodeling complexes. The SIN3 complex is illustrated on the left and the NuRD complex on the right. Figure was generated at biorender.com.](image)

HDAC1 binds genome-wide to regulate the transcription of genes, including those involved in drug resistance (Yixuan Li et al., 2016). I sought to determine the
effect of *HDAC1* overexpression on drug resistance in pancreatic cancer using RNA-sequencing and ChIP-sequencing to evaluate the downstream effects of altered gene expression and the impacts on chromatin structure in cancer. Using this multi-omic approach I identified genes that contribute to drug resistance in pancreatic cancer and are potential therapeutic targets.

**1.3 Analysis of Whole Genome Sequence for Disease-causing Variants**

**1.3.1 Variant Types**

Above, I described somatic variation (and limited germline variation) that contribute to tumorigenesis. Using WGS also allows identification of germline variation associated with disease. WGS can simultaneously detect multiple types of variation, including small nucleotide variants (SNV), deletions or insertions (INDEL), and copy number variants (CNV) (Souche et al., 2022). A SNV is a substitution of one nucleotide for another and can lead to a nonsynonymous change in an amino acid. Frameshifts, a disruption of the open reading frame, are one effect of INDELs which can cause termination or dysregulation of translation. Examples of this are a deletion in the *CFTR* gene which results in cystic fibrosis, an inherited multi-organ disease (White et al., 1990) or an insertion in the *EGFR* gene which acts as an oncogenic driver in cancer (Vyse et al., 2019). The result of a CNV, a type of structural variant, is the duplication or deletion of a genomic region causing an alteration in gene structure or dysregulation of gene expression (Cuccaro et al., 2017). CNVs have been studied in neurodegenerative diseases and duplication of the *APP* gene is associated with early-onset Alzheimer’s disease (Cuccaro et al., 2017).
While there are numerous advantages for whole genome sequencing there are also limitations of this technology. Variants such as triplet repeat expansions can only be detected using PCR free genomes and detecting mosaicism is limited (Marshall et al., 2020). Despite the limitations, WGS is a diagnostic tool that has allowed us to expand our variant identification beyond the coding regions of the genome. A portion of my thesis work has relied on the generation and analysis of WGS to identify genetic variants associated with rare human disease.

1.3.2 Annotation of Variants

Variants are annotated using the tool Codicem (Holt et al., 2019) to interpret pathogenicity. Codicem includes over fifty annotations that assess deleteriousness, population frequencies, splice impact scores, protein change scores, and conservation scores, among many others. The combination of these annotations is used to classify and interpret variants.

1.3.3 Classification of Variants

We use the ACMG criteria for pathogenicity (Richards et al., 2015) as a guideline for the classification and interpretation of variants. These criteria use multiple methods to classify a sequence variant into one of five tiers: benign, likely benign, uncertain significance, likely pathogenic, or pathogenic. Multiple types of evidence are required for a variant to reach a pathogenic classification. Evidence is found in population data, computational and predictive data, functional data, segregation data, and allelic data. The evidence for a given variant is then combined to establish in which tier the variant is classified.
1.3.4 Using Whole Genome Sequencing for Disease Diagnosis

Whole genome sequencing allows for the detection of variants beyond the coding regions of the genome. Variants in the 5’ UTR would not be identified using whole exome sequencing or clinical gene panels, therefore the variant in \textit{FMR1} associated with intellectual disability would not be identified using those technologies. To capture the entire genomic landscape of a patient we use WGS to assess inherited variants.

My focus was on rare variants in human disease, specifically cancer and neurodegenerative disease. The combination of computational tools allows us to identify a variant and determine its effect on the gene or gene product including its deleteriousness. The identification of clinically actionable, disease-associated variants provides the patient’s medical team with knowledge about the genetic contribution of their disease, thus potentially leading to therapeutic intervention. Information about the mode of inheritance is also typically available and can be useful in family planning. It should be noted that oftentimes there are no existing treatment options for a diagnosis, but for a family that has been on a diagnostic odyssey, having any answer allows them to join support groups and be aware of new research being conducted that could one day lead to therapeutic treatment.
Chapter 2. Multi-omic Analysis of HDAC1 Function in Pancreatic Cancer Reveals Altered GTPase Activity Impacting Chemoresistance

2.1 Introduction

Multi-drug resistance is a key factor controlling patient outcomes in pancreatic ductal adenocarcinoma (PDAC). A mechanistic understanding of resistance enables development of novel treatments and precision medicine. Chromatin remodeling, mediated by altered expression and function of regulators including HDAC1, is one mechanism of resistance. Here we describe a multi-omic analysis of HDAC1, a histone deacetylase involved in several chromatin remodeling complexes, to understand how overexpression contributes to multi-drug resistance and patient survival in PDAC. In this study we used a combination of genomic and biochemical approaches to identify and test potential mechanisms of chemoresistance in pancreatic cancer, as well as developed proficiency in genomic data analysis.

2.2 Background

Pancreatic ductal adenocarcinoma (PDAC) is the most common form of pancreatic cancer and one of the most lethal cancers with a five-year survival rate of 11.5%. Due to the lack of early-stage symptoms, 52% of patients are diagnosed with
unresectable, locally advanced, metastatic cancer (SEER, 2022). Chemotherapeutics such as gemcitabine, abraxane, and combination therapies like FOLFIRINOX are standard of care treatments to improve prognosis, but chemoresistance develops in the majority of patients and contributes to the poor outcomes.

PDAC tumor cells achieve drug resistance through many cellular mechanisms. For example, over-activation of DNA damage repair processes limits the effectiveness of nucleoside analogs like gemcitabine. Tumors acquiring a mesenchymal phenotype, activated through genes involved in the epithelial to mesenchymal transition (EMT), become more invasive and have increased migratory potential (Palamaris et al., 2021). Additionally, the hypovascular nature of PDAC tumors prevents sufficient delivery of oxygen to the tumor cells. This hypoxic environment promotes overexpression of anti-apoptotic genes (i.e., Bcl-XL and FLIP), suppressing apoptosis and contributing to drug resistance (Palamaris et al., 2021; Shah et al., 2020; Beatty et al., 2021). Reduced blood flow also restricts delivery of drugs to tumors, further exacerbating resistance. Understanding mechanisms of drug resistance can facilitate development of therapeutic strategies to prevent or reverse resistance.

Previous work, including our own publications, linked chromatin remodeling genes to chemoresistance and patient survival in PDAC (Hasan et al., 2019; Ying et al., 2016; McCleary-Wheeler et al., 2013). Chromatin remodeling is a mechanism of gene regulation through rearrangement of chromatin structure to alter DNA accessibility and influence transcription factor binding. This process can alter gene expression patterns and lead to cellular reprogramming that contributes to chemoresistance. Dysregulation of chromatin remodeling genes leads to global changes in gene expression making it
difficult to determine whether chromatin remodeling alters known resistance pathways, novel resistance pathways, or a combination (Yixuan Li et al., 2016).

We previously demonstrated that overexpression of HDAC1 contributes to multidrug resistance in pancreatic cancer cells (Ramaker et al., 2021). While HDACs are canonically members of repressive complexes, binding of HDAC1 has also been associated with transcriptional activation. In some cases, this is explained by HDAC1’s ability to recruit RNA Pol II or regulate transcriptional elongation (Zhibin Wang et al., 2009; Greer et al., 2015). HDAC1 participates in several remodeling complexes including Sin3A, NuRD, and CoREST (Park et al., 2020). In cancer, HDAC1 regulates the acetylation of histone and non-histone proteins which modulate the expression of genes that influence cancer progression (Yixuan Li et al., 2016). For example, the Sin3A complex, including HDAC1 is recruited to the promoter of CDH1, an epithelial cell marker, where it silences CDH1 expression during metastasis (Yixuan Li et al., 2016). HDAC1 regulates expression of several genes involved in resistance pathways including apoptosis, DNA damage repair, metastasis, and EMT (Yixuan Li et al., 2016; Hai et al., 2021).

The ability of HDAC1 to regulate genes is important for drug resistance, cancer progression, and tumor suppression making it a strong candidate as a drug target. Several HDAC inhibitors (HDACi) have been tested (e.g. entinostat and romidepsin) for the treatment of solid tumors and hematological malignancies (Koutsounas et al., 2013). However, clinical trials for HDACis in PDAC have been largely unsuccessful (Koutsounas et al., 2013; Hontecillas-Prieto et al., 2020). HDAC inhibition by commercially available HDACIs leads to negative side effects in patients (Laschanzky et
al., 2019) since HDACs have a broad impact on expression of genes involved in cancer pathways and normal cellular functions including the function of non-histone proteins; thus, inhibition of multiple HDACs using molecules that target a class of proteins causes global effects. Using genomic analyses of an in vitro system perturbing HDAC1 function, we identify genes and pathways which are regulated by HDAC1 and contribute to chemoresistance. These genes represent future targets for drug development and potential markers of treatment response in patients with elevated HDAC1 expression.

In this study, we overexpressed HDAC1 in a well described PDAC cell line, MIA PaCa-2, and measured the effects on gene expression, HDAC1 binding, and chromatin structure to better understand how HDAC1 activation contributes to resistance. We found that HDAC1 overexpression leads to altered activity of pathways (e.g., EMT, resistance to apoptosis, altered cell cycle checkpoint, and increased hypoxia); each having the potential to contribute to resistance. We showed that HDAC1 overexpression leads to a more mesenchymal phenotype in vitro and observed that HDAC1 overexpression in patient tissues alters these pathways in a similar way. Using ChIP-seq, we determined that the genes bound and regulated by HDAC1 were enriched for GTPases and expression of these genes in patient tissues were negatively correlated with patient survival. We showed that HDAC1 overexpression results in altered expression of GTPases and thus increased GTPase activity suggesting that altered GTPase activity contributes to chemoresistance and represents possible targets to reverse resistance. Together these data reveal an improved mechanistic understanding of the role of HDAC1 in chemoresistance and nominate GTPase genes as potential therapeutic targets.
2.3 HDAC1 Overexpression Leads to Multi-drug Resistance

HDAC1 expression has been previously linked to increased cellular resistance to cytotoxic chemotherapeutic drugs (Ramaker et al., 2021). To further evaluate the effects of HDAC1 overexpression on chemotherapeutic resistance, we used CRISPRa (Joung et al., 2017) to generate a stable MIA PaCa-2 cell line (MP2_HDAC1_OE) expressing HDAC1 at ~3 times the levels of the levels of the control line (MP2_NTC) which expresses a non-targeting control guide (Appendix A.1A). We also treated the MIA PaCa-2 cell line with a HDAC1 siRNA to evaluate the effects of reducing HDAC1 expression on chemoresistance (MP2_HDAC1_KD). HDAC1 was reduced by 89% in MP2_HDAC1-KD cells compared to unaltered MIA PaCa-2 cells. MIA PaCa-2 is a well-characterized line with modest to high expression of HDAC1 (Minami et al., 2021). HDAC1 is the most abundantly expressed HDAC gene in this line (Appendix A.1B). Under these three conditions, we measured the effect of treatment with irinotecan, gemcitabine, and oxaliplatin on cell viability (Figure 2.1A-C). MP2_HDAC1_OE cells were more resistant to drug treatment than control cells and MP2_HDAC1_KD cells. Since HDAC1 overexpression led to increased resistance to these drugs, we evaluated the effect of HDAC1 protein inhibition on drug response. We treated MP2_HDAC1_OE and MP2_NTC lines with romidepsin, a HDAC1/2 inhibitor, in combination with increasing concentrations of irinotecan. We observed a sensitizing effect of romidepsin on the MP2_HDAC1_OE cells treated with irinotecan (Figure 2.1D). Together these experiments show a sensitizing effect of decreasing HDAC1 activity through either chemical inhibition or decreased expression.
Figure 2.1: Quantification of cell viability in PDAC cell lines following treatment of chemotherapeutics. Quantification of viability following treatment with (A) gemcitabine, (B) oxaliplatin, and (C) irinotecan in MP2_HDAC1_OE (blue triangles), MP2_NTC (pink squares), and MP2_HDAC1_KD (grey circles) cell lines. The bar represents the median. P-values were calculated using a two-tailed t-test. *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001, ns = not significant. (D) Quantification of cell viability following treatment with romidepsin, a HDAC1 inhibitor, in MP2_HDAC1_OE and MP2_NTC cell lines. Bar = median. P-values were calculated using a two-tailed t-test. ***p<0.001, ****p<0.0001, ns = not significant.
2.4 HDAC1 Overexpression Induces Expression of Markers of EMT In Vitro and in Human PDAC Tissues

We performed RNA-sequencing to measure gene expression in MP2_HDAC1_OE and MP2_NTC cells. We found 1,259 genes that are differentially expressed with overexpression of HDAC1 (padj < 0.1). These differentially expressed genes (DEG) were enriched for pathways involved in drug resistance: apoptosis, EMT, G2-M checkpoint, and hypoxia (Figure 2.2A, Appendix A.2). Alteration of EMT-associated DEG promotes invasion and migration associated with a more drug resistant mesenchymal cell state (Bulle et al., 2020). The cell surface marker CD44 is characteristic of the mesenchymal phenotype (Zhao et al., 2016; Luu, 2021). We detected a 1.8-fold increase in CD44 expression upon overexpression of HDAC1. Consistent with the expression data, immunohistochemistry showed a comparable 2-fold increase in relative density of CD44 protein in PDAC cell lines with HDAC1 overexpression (Figure 2.2B,C).
Figure 2.2: HDAC1 overexpression is associated with increased expression of EMT genes. (A) Expression of EMT genes in MP2_HDAC1_OE and MP2_NTC cell lines. Each column represents a replicate of the denoted cell line. The color scale denotes the z-score of each gene. (B) Brightfield images and immunofluorescent staining of DAPI (blue), CD44 (green), and merged CD44:DAPI (blue/green) of MP2_HDAC1_OE (bottom) and MP2_NTC (top) cells. (C) Violin plot analysis of immunofluorescent staining of CD44 in MP2_HDAC1_OE (blue) and MP2_NTC cells (pink). 100 cells were measured for each cell line. P-values were calculated using a two-tailed t-test. ****p<0.0001.

These data provide evidence that HDAC1 overexpression modulates several known resistance pathways including induction of EMT in an in vitro model. Next, we sought to determine whether HDAC1 overexpression in patient tumors leads to similar changes in gene expression. We used RNA-sequencing data collected from pancreatic tumor tissues by The Cancer Genome Atlas (TCGA-PDAC dataset). We grouped tissues based on their expression of HDAC1. We compared the lowest (n=45, HDAC1<sup>HIGH</sup>) and
highest (n=45, HDAC1\(^{\text{LOW}}\)) quartiles of HDAC1 expressing samples to identify DEG in these tissues. We identified 10,592 DEG between HDAC1\(^{\text{HIGH}}\) and HDAC1\(^{\text{LOW}}\) tissues. We intersected this gene list with the 1,259 genes identified in our in vitro experiment and identified 322 genes that are significantly altered (padj < 0.1) in the same direction as we observed in cell lines. Heatmaps of the DEG in both datasets were clustered by sample which separated high and low HDAC1 expression (Figure 2.3A,B). This analysis indicates that there is a similar pattern of expression in the PDAC tissues with high HDAC1 expression as in the MP2_HDAC1_OE cells.

Given our previous finding that HDAC1 overexpression can lead to in vitro drug resistance we wondered whether HDAC1 expression might be associated with patient response to treatment. In the TCGA dataset, we used patient prognosis information, specifically overall survival, as a proxy for treatment response. We hypothesized that given HDAC1’s role in cellular resistance, genes associated with HDAC1 overexpression might have prognostic value. We divided the 322 genes associated with increased HDAC1 expression in both patient tissues and in the MP2_HDAC1_OE line into two groups: upregulated genes (n = 216) and downregulated (n = 106) genes. For each gene set, we calculated a mean gene expression value from TCGA PDAC patient tissues and compared the top and bottom quartiles in a survival analysis. We determined that overall survival was shortened for patients with high expression of the upregulated genes and low expression of the downregulated genes (upregulated genes p = 0.00065, downregulated genes p = 0.0051, Figure 2.3C,D). Although overall survival depends on multiple factors, including treatment response, this finding is consistent with our observation that HDAC1 overexpression is associated with drug resistance in vitro and
supports the hypothesis that these genes might also impact drug response in patients and lead to decreased survival time. Importantly, overexpression of \textit{HDAC1} alone is not predictive of patient survival (p=0.44, \textbf{Appendix A.3}). Using an independent cohort (Kirby et al. 2016), we observed a similar difference in survival based on HDAC1-regulated genes (\textbf{Appendix A.4 A,B}, upregulated genes p = 0.064, downregulated genes p = 0.19).

We wanted to determine whether combining the data we generated from cell lines with patient tumor data improved survival predictions. We compared prognostic predictions from the 322 DEG associated with \textit{HDAC1} expression in both TCGA PDAC tumors and HDAC1\_OE cell lines in \textbf{Figure 2.3} with the top 322 DEG in TCGA PDAC samples with high and low \textit{HDAC1} expression as well as the top 322 DEG in our PDAC cell lines with \textit{HDAC1} overexpression and controls. We observe a more significant p-value for overall survival when combining our \textit{in vitro} data with patient data than when using DEG from cell lines or TCGA PDAC tumors individually (\textbf{Appendix A.5 A-D}).

Gene set enrichment analysis of the 322 DEG revealed multiple cancer processes including cadherin binding, cell-cell adhesion, regulation of cell migration, and GTPase activity (\textbf{Figure 2.3E, Appendix B.1}). Some of these are well described resistance pathways, but GTPase activity has not been thoroughly explored for its role in resistance. Among the enriched GTPase genes is \textit{RAP2B}, which was upregulated upon \textit{HDAC1} overexpression (\textbf{Appendix A.6A,B}). When activated, this GTP-binding protein promotes cell migration, cell adhesion, proliferation, and metastasis in cancer (Junqiang Li et al., 2022; Miao et al., 2019; Di et al., 2015).
Figure 2.3: Genes altered by increased *HDAC1* expression in PDAC cell lines and TCGA PDAC samples are associated with patient survival.

(A) Expression of DEG in MP2_HDAC1_OE (blue) and MP2 NTC (pink) cell lines. DEG are significantly (padj < 0.1) altered in the same direction in MP2_HDAC1_OE cells and TCGA PDAC tissues with the top 25% of *HDAC1* expression. Each column represents a replicate of the noted cell line. The color scale denotes the z-score of each gene. (B) Expression of DEG in TCGA PDAC samples. Each column represents a tumor sample. The color scale denotes the z-score of each gene. DEG are significantly (padj < 0.1) altered in the same direction in TCGA PDAC tissues with the top 25% of *HDAC1* expression and MP2_HDAC1_OE cells.
Figure 2.3 (continued)

(C) Overall survival of TCCA PDAC patients (n = 90) with top and bottom 25% of average gene expression of upregulated genes (n = 216) in A and B. P-values were derived using a log-rank test. (D) Overall survival of TCCA PDAC patients (n = 90) with top and bottom 25% of average gene expression of downregulated genes (n = 106) in A and B. P-values were derived using a log-rank test. (E) GO analysis showing enriched molecular functions using the genes (n = 322) in A and B.

2.5 HDAC1 Overexpression Alters Chromatin Accessibility in Distal Enhancer and Promoter Regions Nearby Molecular Switches

To identify direct and indirect impacts of HDAC1 overexpression that might contribute to resistance, we measured genome-wide DNA binding of the HDAC1 protein and the presence of the activating histone mark, H3K27 acetylation (H3K27ac) using
ChIP-sequencing in the MP2_HDAC1_OE and MP2_NTC cell lines. Using the standard ENCODE ChIP-seq protocol for peak calling (Landt et al., 2012), we identified 17,457 binding sites for HDAC1 (10,033 unique to MP2_HDAC1_OE, 3,789 unique to MP2_NTC). We found 30,961 regions of H3K27ac; 8,392 were unique to MP2_HDAC1_OE and 5,916 were unique to MP2_NTC (Figure 2.4A). We observed an enrichment of HDAC1 binding and regions of H3K27ac near the transcription start sites (TSS) of DEG when HDAC1 is overexpressed. The H3K27ac peaks specific to HDAC1 overexpressing cells occurred significantly more near the upregulated genes despite HDAC1’s canonical role as a repressor (Appendix A.7, p<0.1).

To identify regions with altered HDAC1 binding or H3K27 acetylation directly impacting gene expression, we overlapped 1 kb regions centered on all HDAC1 and H3K27ac peaks (overlapping peaks were merged, see Methods) with promoter regions of DEG (2kb upstream of annotated TSS). This revealed 1,857 regions of HDAC1 binding or H3K27 acetylation in promoters of 1,040 DEG (one promoter can have more than one overlapping peak). Gene set enrichment analysis of these 1,040 DEG revealed enrichment for GTPase activity, cadherin binding, and DNA binding (FDR < 0.05) (Appendix B.2).

To better understand how HDAC1 overexpression impacts H3K27ac and influences gene expression, we divided regions of HDAC1 binding and H3K27 acetylation based on whether they were increasing or decreasing across the regions described above (500bp up and downstream centered on the peak). Using the sequencing reads collected in a region called as a peak in any of our ChIP-seq experiments, we calculated a fold-change to determine whether there was evidence of increased or
decreased binding. Given HDAC1’s canonical role as a repressor, we expected that increased HDAC1 binding would be associated with decreased H3K27 acetylation, however, we only identified 235 DEG with increased HDAC1 binding and reduced H3K27 acetylation in the promoter regions (+/- 2 kb from TSS). In contrast, the promoters of 597 DEG had increased HDAC1 binding and increased H3K27 acetylation (fold-change > 1) upon HDAC1 overexpression (Appendix 2.3). Since previous studies have shown that HDAC1 binding can be found near active genes (Zhibin Wang et al., 2009), we tested whether expression of DEG where we also measured changes in altered HDAC1 binding and H3K27 acetylation were associated with overall patient survival. We performed survival analysis comparing outcomes of patients with the top 25% and bottom 25% mean tumor gene expression of these 597 genes. Patients with the highest mean expression of upregulated genes and lowest mean expression of downregulated genes have worse overall survival (upregulated: p = 0.0021, Figure 2.4B, downregulated: p = 0.023, Appendix A.8A). We also showed that HDAC1 expression is significantly higher in patients with the top 25% of mean tumor gene expression of the 597 genes (Appendix A.8B).
Figure 2.4: ChIP-sequencing reveals DEG associated with patient survival. 
(A) Venn diagram showing overlap of H3K27ac ChIP-seq peaks in MP2_HDAC1_OE and MP2_NTC cell lines. (B) Overall survival of TCCA PDAC patients (n = 90) with top (teal) and bottom (grey) 25% of average gene expression of upregulated DEG with increased HDAC1 binding and H3K27 acetylation in their promoter upon HDAC1 overexpression. P-value was derived using a log-rank test. (C) Overall survival of TCCA PDAC patients (n = 90) with top (blue) and bottom (red) 25% of predictor values generated from the nine transcript LASSO model. P-value was derived using a log-rank test.
2.6 A Nine-gene Signature of HDAC1 Regulated Genes Predicts PDAC Patient Survival

Identification of prognostic signatures in PDAC could be of clinical utility. The analyses above identified 597 genes regulated by HDAC1 that predict patient outcomes using genes identified from in vitro and in vivo signatures of HDAC1 overexpression, although expression of HDAC1 alone is not prognostic. We calculated a simplified signature of patient prognosis using a multivariate logistic regression with L1 penalized log partial likelihood (LASSO) for feature selection (Tibshirani, 1996). From the 597 genes, LASSO identified a 9-transcript model sufficient to differentiate TCGA PDAC tumors with high and low HDAC1 expression (Appendix 2.8C,D). To determine the clinical relevance of the genes selected using the LASSO model, survival analysis was performed comparing the patients in the top and bottom quartile of predictor values from the regression and the group with the highest predictor values had worse overall survival (Figure 2.4C, p = 0.037).

2.7 HDAC1 Overexpression Leads to Increased GTPase Activity and Chemoresistance

The identification of many DEG upon HDAC1 overexpression led us to explore pathways that have not been previously linked to chemoresistance. Pathway enrichment analysis of DEG with increased HDAC1 binding and H3K27 acetylation in promoters (n = 597) identified an enrichment for many known cancer pathways (Appendix B.4). Included on this list was Ras signaling, regulation of apoptotic signaling pathway, chromatin binding, and GTPase activity. GTPase activity was also significant in enrichment analyses described in Figure 2.3 driven by overexpression of the GTPases
and associated proteins (e.g. RALB, RAB27B, and RAC1) which have increased expression with HDAC1 overexpression and are associated with worse overall patient survival (Appendix A.9A). ARHGAP5, a Rho family-GTPase activating protein (Figure 2.5A), was also increased with HDAC1 overexpression and we observed increased HDAC1 binding and H3K27 acetylation near its TSS (Figure 2.5B). Since activation of GTPase activity is not a well described mechanism of drug resistance, we used RNA-sequencing data from 14 unmodified PDAC cells lines with varying response to gemcitabine (Data from Kirby 2016) to further support the hypothesis that GTPase activity alters cellular response to chemotherapy. We found that cell lines with increased expression of genes influencing GTPase activity had higher levels of resistance to gemcitabine (Figure 2.5C). We identified a total of 11 genes that modulate GTPase activity and they all show a similar pattern of increased expression in resistant cell lines (Appendix A.9B). Increased GTPase activity activates the MAPK and PI3K pathways which promote proliferation and survival in cancer (Soriano et al., 2021). KEGG (Kanehisa et al., 2021) pathway mapping of these 597 DEG confirmed the increased expression of genes in the MAPK and PI3K pathways upon HDAC1 overexpression (Appendix A.10).

Given the increased transcript levels of several GTPases (Figure 2.5D), we tested whether there was a measurable difference in GTPase activity upon HDAC1 overexpression. GTPase activity was measured through the detection of GTP remaining after a GTP hydrolysis reaction catalyzed by cell lysates from the MP2_HDAC1_OE line compared to the MP2_NTC line. MP2_HDAC1_OE cell lysates have significantly increased GTPase activity (Figure 2.5E) compared to the MP2_NTC control line.
Conversely, treatment of MP2_HDAC1_OE and MP2_NTC cells with romidepsin, a HDAC1/2 inhibitor, decreased GTPase activity (p < 0.05) (Appendix B.5). We also observed decreased GTPase activity in MP2_NTC cells with a siRNA targeting HDAC1 (MP2_NTC_siRNA_HDAC1) compared to MP2_NTC cells with a non-targeting siRNA (MP2_NTC_siRNA_Control). These data demonstrate that HDAC1 overexpression increases GTPase activity and that inhibition of HDAC1 reverses the effect.
Figure 2.5: HDAC1 overexpression is associated with increased GTPase activity. 
(A) Normalized expression of ARHGAP5 in MP2_HDAC1_OE (blue) and MP2_NTC (pink) cell lines. The bar is the median. P-values were calculated using an unpaired two-tailed t-test, p = 0.0003479. 
(B) ChIP-seq analysis of the ARHGAP5 promoter showing increased HDAC1 and H3K27ac peak height in MP2_HDAC1_OE (blue) cells compared to MP2_NTC (pink) cell lines. 
(C) Normalized expression of ARHGAP5 in PDAC cell lines resistant (blue) and sensitive (pink) to gemcitabine. The bar is the median. P-values were calculated using an unpaired two-tailed t-test, p = 0.00831. 
(D) Expression of DEG involved in GTPase activity in MP2_HDAC1_OE (blue) and MP2_NTC (pink) cell lines. Each column represents a replicate of the noted cell line. The color scale denotes the z-score of each gene. 
(E) Comparison of GTPase activity in the following cell lines: MP2_HDAC1_OE, MP2, MP2_NTC, and MP2 treated with romidespin. P-values were calculated using a two-tailed t-test. *p < 0.05, **p < 0.01.
2.8 Methods

2.8.1 Cell Culture

MIA PaCa-2 cells (ATCC #CRM-CRL-1420) were cultured in D10 media: DMEM (Lonza #12-614Q) supplemented with 10% FBS (GE LifeSciences #SH30071.03), and 0.5% penicillin-streptomycin (ThermoFisher #15140122). All cell lines were maintained at 37 °C and 5% CO2. Cells were cryopreserved with the addition of 10% DMSO (EMD #MX1458-6).

2.8.2 Plasmids

LentiCRISPRv2 (Addgene #52961) or lentiSAMv2 (Addgene #92062) and lenti-MS2-p65-HSF1-Hygro (Addgene #89308) were used to generate stable cell lines for gene knockout and activation, respectively. pMD2.G (Addgene #12259) and psPAX2 (Addgene #12260) were used to facilitate viral packaging of sgRNAs and single vector plasmids.

2.8.3 sgRNA Cloning

gRNA oligos were designed and cloned into their respective plasmids as described previously (Joung et al., 2017).

2.8.4 DsiRNA

IDT TriFECTa RNAi kit was used per manufacturer's protocol. 100,000 cells were seeded in 1 well of a 12 well tissue culture treated plate 24 hours prior to transfection. Cells were transfected using RNAiMax (ThermoFisher #13778-030) following manufacturer’s recommended protocol. As indicated in the TriFecta kit (IDT #hs.Ri.HDAC1.13), TYE 563 transfection efficiency control, positive HPRT-S1 control,
and negative (DS NC1) scrambled sequence control were utilized. Further assays were performed 48 hours after transfection. Expression was validated with each transfection with the IDT PrimeTime qPCR Assay system on an Agilent QuantStudio 6 Flex Real-Time PCR system.

**DsiRNA Sequences:**

<table>
<thead>
<tr>
<th>IDT DsiRNA name</th>
<th>DsiRNA Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>hs.Ri.HDAC1.13.1-SEQ1</td>
<td>rArGrUrUrCrUrUrUrUrGrArArUrArCrUrUrUrUTT</td>
</tr>
<tr>
<td>hs.Ri.HDAC1.13.1-SEQ2</td>
<td>rArArArArUrArGrUrArUrCrUrCrArArArArGrGrArArArCrUrG</td>
</tr>
<tr>
<td>hs.Ri.HDAC1.13.2-SEQ1</td>
<td>rCrUrGrArArCrUrGrCrUrArArArGrUrArUrCrCrCAG</td>
</tr>
<tr>
<td>hs.Ri.HDAC1.13.2-SEQ2</td>
<td>rUrCrUrGrGrArArCrUrUrArGrGrUrCrArGrUrCrCrA</td>
</tr>
<tr>
<td>hs.Ri.HDAC1.13.3-SEQ1</td>
<td>rArGrCrGrCrUrGrCrUrCrArArGrUrArUrGrUrArUrGGA</td>
</tr>
<tr>
<td>hs.Ri.HDAC1.13.3-SEQ2</td>
<td>rUrCrArUrUrUrGrArCrUrGrArGrCrArCrGrUrUrGrCrUrG</td>
</tr>
</tbody>
</table>

**qPCR Sequence IDs:**

<table>
<thead>
<tr>
<th>Gene</th>
<th>Exon</th>
<th>IDT primeTime Assay ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>HDAC1</td>
<td>1-2</td>
<td>Hs.PT.58.20534173</td>
</tr>
<tr>
<td>HDAC1</td>
<td>3-4</td>
<td>Hs.PT.58.38680914</td>
</tr>
<tr>
<td>ACTB</td>
<td>1-2</td>
<td>Hs.PT.39a.22214847</td>
</tr>
<tr>
<td>GAPDH</td>
<td>2-3</td>
<td>Hs.PT.39a.22214836</td>
</tr>
<tr>
<td>HPRT1</td>
<td>8-9</td>
<td>Hs.PT.58v.45621572</td>
</tr>
</tbody>
</table>
2.8.5 GTPase-Glo Assay Using Cell Lysates

*In vitro* GTPase activity was measured using the GTPase-Glo assay (Promega #V7681). We followed the protocol as described (Mondal et al., 2015) with modifications for use of cell lysates. Cell lysates were made from the following cell lines: MP2_HDAC1_OE, MP2_NTC, MP2_NTC with siRNA targeting HDAC1 (MP2_NTC_siRNA_HDAC1), MP2_NTC with a non-targeting siRNA (MP2_NTC_siRNA_Control), MP2_HDAC1_OE treated with 0.01 µM romidepsin, an HDAC1 inhibitor, and MP2_NTC treated with 0.01 µM romidepsin. Cells (2 x 10^6 per tube) were lysed in a lysis buffer containing 50 mM HEPES at pH 7.6, 150 mM NaCl, 10% Glycerol, 0.1% NP-40, and 2 mM MgCl₂. To generate the lysate, 10µL of lysis buffer per 100,000 cells was added to each cell pellet and resuspended. Lysates were mixed for 30 minutes at 4°C, vortexed in three 10 second intervals, then centrifuged at 4°C for 30 minutes at 16.1x RCF. A 2X GTP solution was prepared and the reaction was initiated following the manufacturer’s protocol.

Modifications for cell lysates required background wells for each cell line. GTPase-Glo Buffer was added to cell lysates at a final concentration of 1 µL per 10,000 cells. After the GTPase reaction, 20µL was added to each respective background well. Luminescence was measured using a BioTek Synergy H5 plate reader. To calculate GTPase activity for each cell type, we calculated the difference between the luminescence of the experimental wells and background wells. GraphPad Prism 9 (version 9.3.1) was used for plotting bar charts and t-tests performed in GraphPad were unpaired, parametric, two-tailed with 95% confidence interval.
2.8.6 ChIP-sequencing

MP2_HDAC1_OE and MP2_NTC cells (2 x 10^7) were cross-linked, harvested, and DNA was precipitated using a commercial H3K27ac antibody (Abcam, ab4729). Libraries were constructed, pooled, and sequenced using an Illumina NovaSeq instrument with 75bp single-end reads. These data were generated and analyzed using published ENCODE protocols (https://www.encodeproject.org/documents/).

Differential binding analysis was conducted using the “multiBigwigSummary" tool from the “deepTools" package (Ramírez et al., 2014). Using this tool, a ChIP-seq score was generated for each sample and region using genomic coordinates defined as +/- 500 bp from the center of peaks defined using the published ENCODE protocol (Landt et al., 2012) and 1 kb upstream of all annotated genes. Regions were merged together if they overlapped. We omitted any regions with a ChIP-seq score less than 1 for both MP2_HDAC1_OE and MP2_HDAC1_NTC. Using the ChIP-seq score, we calculated a fold-change between MP2_HDAC1_OE and MP2_NTC for HDAC1 and H3K27ac in each defined region. The criteria for differentially bound regions was a fold-change greater than or less than one for HDAC1 binding and H3K27 acetylation.

2.8.7 3' RNA-sequencing

Cell pellets were frozen at -80°C until RNA extraction. For RNA extraction 350 µl of RL Buffer plus 1% β-ME from the Norgen Total RNA extraction kit was added to each cell pellet and extraction proceeded per manufacturer’s instructions including use of the DNase kit (Norgen # 37500, 25720). RNA quality was verified with the Agilent BioAnalyzer RNA Nano 600 kit (cat# 5067-1512) with the RIN range between 9.2-10. RNA-sequencing libraries were made using Lexogen QuantSeq 3' mRNA-Seq
Library Prep Kit FWD for Illumina kit (cat# 015.24) with 250 ng of RNA input. They were pooled and sequenced on an Illumina NextSeq 500 instrument with 75 bp single-end reads. Read counts averaged 4 million reads and an average Q30 of 91.28%. Lexogen’s BlueBee integrated QuantSeq data analysis pipeline was used for trimming, mapping, and alignment and the R package “DESeq2” was used for differential expression analysis.

**2.8.8 Drug Resistance Screening**

Cells were seeded in 96-well plates at 2000 cells/well. Seeded cells were dosed with a range of concentrations of each drug: gemcitabine (0-12.5nM), oxaliplatin (0-3.5µM), or irinotecan (0-5µM). Cells were given a second dose of drug at the same concentration as the first 48 hours later. The number of viable cells surviving drug treatment were assayed with CellTiter-Glo (Promega #G7571) 24 hours after the last drug treatment per manufacturer's protocol using a BioTek Synergy H5 plate reader.

HDAC1 inhibition with romidepsin (Sigma #SML1175-1MG) was performed similarly to above except that cells were dosed every 24 hours with either 0.01 µM romidepsin, a range of irinotecan or both. Equal volume DMSO was used as a control in place of romidepsin. The number of viable cells surviving drug treatments were assayed with CellTiter-Glo (Promega #G7571) 24 hours after the last drug treatment per manufacturer's protocol using a BioTek Synergy H5 plate reader. In both cases, data were plotted using GraphPad Prism 9, version 9.3.1. T-tests performed were unpaired, parametric, two-tailed with 95% confidence interval.
2.8.9 Cell Staining

75,000 MIA PaCa-2 cells with non-targeting, HDAC1 OE sgRNAs, or HDAC1 KD with DsiRNA were seeded in 12 well plates. Cells were stained using Alexa Fluor 488 Conjugate kit for live cell imaging (LifeTechnologies #A25618) for CD44 via the manufacturer’s protocol. DAPI (Invitrogen #D21490) was counterstained per manufacturer’s protocols for adherent cells. Presence of CD44 in the cells was quantified using ImageJ 1.53K with measurements (area, mean, and integrated density) for stain and background taken with the freehand selection tool. Relative CD44 intensity or bound CD44 per area was calculated for each cell by: integrated density of cell-integrated density of background for that cell/area of that cell. GraphPad Prism 9 (version 9.3.1) was used for plotting violin plots and t-tests performed in GraphPad were unpaired, parametric, two-tailed with 95% confidence interval.

2.8.10 Enrichment Analysis

Enrichr, a comprehensive gene set analysis web server, and the R package “ClusterProfiler” (version 3.12.0) (Yu et al., 2012) were used for enrichment analysis of the differentially expressed genes (Chen et al., 2013). We focused on the pathways (MSigDB) and gene ontology terms (GO MF, GO BP) reaching the significance threshold of FDR < 0.05.

2.8.11 Survival Analysis

To conduct survival analysis, clinical and RNAseq expression data was retrieved from The Cancer Genome Atlas (TCGA) for 178 PDAC (TCGA-PAAD) patients (https://portal.gdc.cancer.gov/). Data was normalized using the R package “DESeq2”
and differentially expressed genes with an FDR < 0.1 were used to generate Kaplan-Meier survival curves. We classified tissues based on their mean expression of a given gene set (bottom, middle, and top quartiles of gene expression). We compared the patients with the lowest and highest quartile of mean gene expression and performed survival analysis. Survival curves and analyses were generated using the “ggplot2”, “survminer”, and “survival” R packages (Wickham, 2016; Kassambara et al., 2018; Therneau, 2021).

2.8.12 Clinical Data and Samples

Clinical data and RNA-sequencing data for TCGA PDAC samples were retrieved on 04/01/2020 using the GDC Data Portal. Our analyses included 178 samples in this cohort that had matched clinical and RNA-sequencing data.

2.8.13 Data Access

ChIP-sequencing data is available using the GEO accessions GSE209895 (H3K27ac) and GSE158541 (HDAC1). RNA-sequencing is available using the GEO accessions GSE79668 and GSE79669 (gemcitabine resistant and sensitive cell lines) (Kirby et al., 2016).

2.8.14 Statistical Testing

Statistical analysis was conducted in R (version 1.2.1335). The following R packages and software were used for analysis:

survival (version 3.2-13) (Therneau, 2021)
survminer (version 0.4.9) (Kassambara et al., 2018)
ggplot2 (version 3.3.6 ) (Wickham, 2016)
DESeq2 (version 1.24.0) (Love et al., 2014)
pheatmap (version 1.0.12) (Kolde, 2019)
cclusterProfiler (version 3.12.0) (Yu et al., 2012)
glmnet (version 4.1-3) (Hastie et al., 2021)
ROCR (version 1.0-11) (Sing et al., 2005)
deepTools (version 3.5.0) (Ramírez et al., 2014)
IGV (version 2.7.2) (Thorvaldsdottir et al., 2013)

2.8.15 LASSO Model Selection

A predictive gene signature from transcripts that are differentially expressed (DESeq2 FDR < 0.1) and have increased HDAC1 binding and H3K27 acetylation near their TSS (+/- 2000 bp) was developed using the LASSO regression model. LASSO was performed using the R package “glmnet” (version 4.1-3). The TCGA PDAC cohort was split into three groups by HDAC1 expression (top 25%, middle 50%, and bottom 25%). The cohort was further subset by randomly distributing an equal number of samples from the top 25% and bottom 25% of HDAC1 expression into two groups (n = 45). The training cohort and the validation cohort used the same dichotomization threshold (top 25% and bottom 25% of HDAC1 expression). Model performance was evaluated based on the model’s ability to classify patients into the high or low HDAC1 expression group. We generated an area under the curve (AUC) value using the R package “ROCR” (version 1.0-11). Kaplan-Meier curves were generated using the R package “survival” (version 3.2-13).
2.8.16 Data Analysis

The adjusted p-value was used to determine the top 322 differentially expressed genes in TCGA PDAC samples for comparison of patients in the top and bottom quartile of \textit{HDAC1} expression. Using the R package “DESeq2” (version 1.24.0), differentially expressed genes were excluded from the analysis if baseMean < 10.

2.9 Discussion

Pancreatic cancer ranks among the deadliest cancers due to its chemoresistant nature and insufficient treatment options. Understanding what drives chemoresistance is essential to identifying new therapeutic targets and improving patient outcomes. Chromatin remodeling has been established as a critical feature of tumorigenesis and cancer progression, making the pathway an attractive drug target. Our work and others showing the importance of key genes like \textit{HDAC1} in chemoresistance have nominated chromatin remodeling genes as possible drug targets. Using genomic and biochemical approaches we revealed potential mechanisms by which \textit{HDAC1} OE contributes to chemoresistance and showed that HDAC1 inhibition sensitizes PDAC cells to chemotherapeutic treatment further strengthening the argument that this pathway is a good candidate for treatment, however HDAC inhibitors have faced challenges in clinical trials. Commercial HDAC inhibitors target a class of HDACs and not specific proteins. This cross-reactivity leads to genome-wide off target effects and patient toxicity. That has motivated the current study which aims to better understand how \textit{HDAC1} activation contributes to resistance and reveal novel downstream targets that may lead to alternative treatment strategies.
In contrast to other tumor types, multiple large scale drug trials that used targeted therapy were not as successful in pancreatic cancer (Hecht et al., 2021; Tempero et al., 2021), thus using a targeted gene panel that can be used to better define potential treatment options for PDAC patients could lead to improved survival and quality of life (Hosein et al., 2022; Wei et al., 2021). In this study, we collected data from an \textit{in vitro} system testing the impact of \textit{HDAC1} overexpression on PDAC cells and combined these results with information from publicly available gene expression data gathered from both tissues of PDAC patients and PDAC cell lines to show that \textit{HDAC1} overexpression regulates a set of transcriptomic responses that contribute to chemoresistance and the genes regulated by \textit{HDAC1} can also be shown to predict patient outcome. \textit{HDAC1} overexpression alone is not significantly prognostic of worse overall survival in PDAC patients. However, the genes altered by \textit{HDAC1} overexpression are prognostic. Since \textit{HDAC1} is a genome-wide regulator of gene expression, in PDAC, we hypothesize that the genes contributing to a worse overall survival prognosis, resulting from tumorigenesis and drug resistance, are the subset of genes altered by \textit{HDAC1} overexpression. These results show that pathways under the control of \textit{HDAC1} contribute to patient outcomes and could be used to predict outcomes that may be linked to treatment response. Of the 597 genes altered by \textit{HDAC1} expression \textit{in vitro}, we identified a 9-transcript novel signature that successfully predicted patients with high and low \textit{HDAC1} expression and was correlated with patient survival. These analyses provided insight into a subset of genes that are regulated by \textit{HDAC1} and are predictive of worse patient survival, suggesting a potential clinical utility. Our panel of biomarkers
represents a potential step forward in the development of an assay that is predictive of patient survival which could influence treatment decisions.

We integrated several datasets to better understand how \textit{HDAC1} overexpression impacts PDAC cells. These studies revealed several processes known for their role in tumorigenesis, progression and drug resistance. One of these was EMT. In addition to its role in progression, EMT is associated with suppression of proteins involved in drug transport, such as CNT3, allowing the cells to evade the anti-proliferative effects of chemotherapeutics (gemcitabine) (Zheng et al., 2015). Cells that undergo EMT also have a more stem cell-like phenotype making them more resistant to chemotherapeutics (Palamaris et al., 2021). Increased expression of \textit{CD44}, a cell surface protein important for cell adhesion and migration, is associated with a more mesenchymal-like phenotype which is characteristic of EMT (Zhao et al., 2016). Here we have shown that the mesenchymal marker, CD44 transcript and protein are more abundant in cells with \textit{HDAC1} overexpression, agreeing with our past work showing that \textit{HDAC1} overexpression leads to increased migration (Ramaker et al., 2021).

Induction of EMT is also associated with drug resistance and we showed that MP2\_HDAC1\_OE cells are resistant to multiple drugs. Understanding the direct regulatory impacts of HDAC1 binding and H3K27ac occupancy is necessary to determine the impacts of \textit{HDAC1} overexpression in pancreatic cancer. Our ChIP-seq experiments revealed altered H3K27 acetylation and HDAC1 binding near HDAC1 regulated genes. Interestingly, we found that the majority of the DEG with HDAC1 binding had an increase in HDAC1 and H3K27ac signals near their promoter. Our findings are in agreement with a previously published study concluding that HDAC1
binding is enriched at actively transcribed genes (Zhibin Wang et al., 2009). It is still not well understood how HDAC1 binding directly affects the activation of gene expression but Greer et al. did show that HDAC1 binding can regulate RNA Pol II recruitment and that HDAC1 may regulate transcription elongation through interaction with BRD4. Additionally, HDAC1 is known to deacetylate proteins other than histones which might facilitate activation of nearby transcription factors (e.g. MYC) (Greer et al., 2015).

Throughout this study, our gene set enrichment analyses of genes associated with HDAC1 overexpression consistently revealed GTPase activity as an enriched process. A variety of proteins contribute to GTPase activity and many are druggable (Prieto-Dominguez et al., 2019), which makes them of potential clinical interest. We have shown that expression of GTPases and GTPase activating proteins are associated with worse overall survival. We used biochemical assays to confirm that cells overexpressing HDAC1 have higher GTPase activity than control cells. This effect was reversible in cells treated with a HDACi, which reduced GTPase activity. GTPase signaling is important for pancreatic cancer initiation, metastasis, and invasion (Yoshimachi et al., 2021). Increased GTPase signaling leads to the activation of key signaling cascades, such as MAPK and PI3K, that regulate cell proliferation, migration, and survival in cancer (Soriano et al., 2021). We have shown that the expression of genes in the MAPK and PI3K pathways are increased upon HDAC1 overexpression highlighting one known pathway altered by HDAC1 that contributes to PDAC progression (Mehra et al., 2021). GTPases, such as RAC1, have been shown to activate EMT pathway in multiple cancers, thus leading to a more invasive and drug resistant phenotype (Fang et al., 2017; Zhou et al., 2016; Ungefroren et al., 2018) and we show that RAC1 expression is
increased upon *HDAC1* overexpression in PDAC cells. Additionally, ARHGAP5 is a GTPase activating protein associated with promoting EMT, a known mechanism of drug resistance in colorectal cancer, but its role in PDAC is not well understood (Tian et al., 2020). We showed that expression of *ARHGAP5* and of several other proteins promoting GTPase activity are increased upon *HDAC1* overexpression and their transcripts are more abundant in gemcitabine resistant PDAC cell lines.

Despite decades of research, PDAC patients continue to have limited treatment options and suffer poor outcomes. Identifying patients who will benefit from existing treatments or those who need an alternative treatment is a key clinical need. Our work identified sets of genes that may help identify patients with poor prognosis due at least in part to treatment resistance and reveals novel therapeutic targets that may benefit patients experiencing resistance to treatment.
Chapter 3. Rare Variant and Polygenic Contributions to Early-onset or Atypical Dementia Risk

3.1 Introduction

Dementia affects over 55 million people worldwide, and nine percent of these patients are under age 65 (Dua et al., 2017). Rare variants in three genes, \( PSEN1 \), \( PSEN2 \), and \( APP \), are associated with autosomal dominant early-onset Alzheimer’s (EOAD), however they only explain about 10% of genetic cases (Stoychev et al., 2019). The use of genome sequencing allows identification of rare variants not included in some targeted gene panel testing, as well as variation that is the purview of future research such as including non-coding variants and variants in novel risk loci. In our previous study we reported on the use of genome sequencing in 32 individuals with early-onset and/or atypical dementia and described several pathogenic variants associated with disease, including combinations of disease-associated risk variants. Our previous work confirmed the value of genetic assessment and identified contributing genetic variation for over half of the cohort (Cochran et al., 2019).

The American College of Medical Genetics (ACMG) criteria was used to assess pathogenicity and pathogenic/likely pathogenic variants were returned. Additional criteria to return variants included: 1) any variant with a disease-established odds ratio of \( >2 \) described in multiple reports, which we defined as an “established risk variant”, 2) presence of one or two \( APOE \varepsilon 4 \) alleles in a patient with EOAD or atypical dementia.
likely due to EOAD, or 3) one strong report with a disease-associated odds ratio >2 with replication included in the study design, which we defined as a “likely risk variant.”

In this report we also assessed the dementia risk related to common variation for the enrolled patients by calculating polygenic risk scores. Our results highlight the complex genetic etiology of early-onset and/or atypical dementia.

3.2 Results

Here we report on 68 additional patients collected as a continuation of the previously published cohort. For each patient with clinician-diagnosed early-onset or atypical dementia, we collected and analyzed genomic sequencing data. We identified returnable, primary findings for 53% of patients (Figure 3.1A, Table 3.1). Including the initial 32 probands described by Cochran, et al. in our previous report, a total of 100 patients have been enrolled through the Brain Aging and Memory Clinic at the University of Alabama at Birmingham (Figure 3.1B).
**Figure 3.1:** Summary of genomic sequencing findings. Summary of findings for the (A) 68 proband cohort and (B) all 100 enrolled probands including the 32 described in Cochran, et al. Patients carrying one APOE ε4 allele are noted as APOE ε4 Het. and those carrying two alleles are noted as APOE ε4 Hom. Patients carrying a risk allele in addition to one or two APOE ε4 allele are listed as APOE ε4 Het. or Hom. + risk. VUS = Variant of Uncertain Significance.