


8-2018

Deciphering the role of human arylamine N-acetyltransferase 1 (NAT1) in breast cancer cell metabolism using a systems biology approach.

Samantha Marie Carlisle
University of Louisville

Follow this and additional works at: <https://ir.library.louisville.edu/etd>

 Part of the [Biochemical Phenomena, Metabolism, and Nutrition Commons](#), [Biochemistry Commons](#), [Bioinformatics Commons](#), [Biological Phenomena, Cell Phenomena, and Immunity Commons](#), [Cancer Biology Commons](#), [Genetic Phenomena Commons](#), [Genetics and Genomics Commons](#), [Molecular Biology Commons](#), [Pharmacology, Toxicology and Environmental Health Commons](#), and the [Systems Biology Commons](#)

Recommended Citation

Carlisle, Samantha Marie, "Deciphering the role of human arylamine N-acetyltransferase 1 (NAT1) in breast cancer cell metabolism using a systems biology approach." (2018). *Electronic Theses and Dissertations*. Paper 3022.
<https://doi.org/10.18297/etd/3022>

This Doctoral Dissertation is brought to you for free and open access by ThinkIR: The University of Louisville's Institutional Repository. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of ThinkIR: The University of Louisville's Institutional Repository. This title appears here courtesy of the author, who has retained all other copyrights. For more information, please contact thinkir@louisville.edu.

DECIPHERING THE ROLE OF HUMAN ARYLAMINE *N*-ACETYLTRANSFERASE 1 (NAT1) IN
BREAST CANCER CELL METABOLISM USING A SYSTEMS BIOLOGY APPROACH

By

Samantha Marie Carlisle
B.S., University of Louisville, 2012
M.S., University of Louisville, 2015

A Dissertation
Submitted to the Faculty of the
School of Medicine of the University of Louisville
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy
in Pharmacology and Toxicology

Department of Pharmacology and Toxicology
University of Louisville
Louisville, Kentucky

August 2018

Copyright 2018 by Samantha Marie Carlisle

All rights reserved

DECIPHERING THE ROLE OF HUMAN ARYLAMINE *N*-ACETYLTRANSFERASE 1 (NAT1) IN
BREAST CANCER CELL METABOLISM USING A SYSTEMS BIOLOGY APPROACH

By

Samantha Marie Carlisle
B.S., University of Louisville, 2012
M.S., University of Louisville, 2015

A Dissertation Approved on

July 20, 2018

by the following Dissertation Committee:

David W. Hein, Ph.D.

J. Christopher States, Ph.D.

Carolyn M. Klinge, Ph.D.

Shesh N. Rai, Ph.D.

Xiang Zhang, Ph.D.

DEDICATION

This dissertation is dedicated to Patrick, Emma, Sabrina, and Barret. Patrick, thank you for your patience, never-ending encouragement, and for introducing me to R. Mom, thank you for always supporting me in my endeavors, allowing me to forge my own path in life, and being a great example of a strong, hardworking woman. Sabrina and Barret, thank you for always being proud of me and putting up with my eccentricities during our childhood. Additionally, this dissertation is dedicated to my late father, Carl, whose too-soon passing motivated me to utilize my love for science to become a cancer researcher and to my late grandmother, Gertrude, for providing such a loving, caring, and constructive after-school home.

ACKNOWLEDGEMENTS

I would first and foremost like to thank my mentor, Dr. David Hein, for all of his advice and direction over the past six years. I would also like to thank my committee members, Dr. J. Christopher States, Dr. Carolyn Klinge, Dr. Shesh Rai, and Dr. Xiang Zhang, for all of their feedback and helpful suggestions throughout this process.

Additionally, I would like to acknowledge my many collaborators: Mark Doll and Dr. Marcus Stepp who constructed the cell lines used as samples in the studies presented in this dissertation, Dr. Carolyn Klinge who I worked with on the Bioenergetics experiments presented in Chapter 3 of this dissertation, Dr. Wolfgang Zacharias, Sabine Waigel, and Vennila Arumugam of the University of Louisville Genomics Core who graciously allowed me to be a part of the library preparation of the RNA-seq (transcriptomics) samples, presented in Chapter 5, Dr. Eric Rouchka and Dr. Xiaohong Li of the KBRIN Bioinformatics Core who did the preprocessing of the transcriptomics data, and Dr. Patrick Trainor who was never too busy to serve as my own personal StackOverflow database. I want to thank all of my collaborators for their kindness, help, and willingness to let me be involved. Furthermore, I would like to thank the Bioinformatics Journal Club Group and Dr. Julia Chariker for providing such a welcoming and supportive environment in which to broaden my bioinformatics knowledge. Thank you to my fellow Hein Lab members, Dr. Raul Salazar-González, Dr. Mariam Habil, and Angeliki Likoudi, for providing a listening ear when times were tough and their friendship.

Lastly, I would like to acknowledge and thank all of my friends and family who have provided unwavering encouragement and support in all of my endeavors. I am tremendously grateful for all of you. Several chapters included in this dissertation have been published; Chapter 2 has been published in the *International Journal of Oncology* and Chapter 3 has been published in *Molecular Carcinogenesis*; additionally, we intend to publish Chapters 4-7 soon.

ABSTRACT

DECIPHERING THE ROLE OF HUMAN ARYLAMINE *N*-ACETYLTRANSFERASE 1 (NAT1) IN BREAST CANCER CELL METABOLISM USING A SYSTEMS BIOLOGY APPROACH

Samantha Marie Carlisle

July 20, 2018

Background: Human arylamine *N*-acetyltransferase 1 (NAT1) is a phase II xenobiotic metabolizing enzyme found in almost all tissues. NAT1 can additionally hydrolyze acetyl-coenzyme A (acetyl-CoA) in the absence of an arylamine substrate. NAT1 expression varies inter-individually and is elevated in several cancers including estrogen receptor positive (ER+) breast cancers. Additionally, multiple studies have shown the knockdown of NAT1, by both small molecule inhibition and siRNA methods, in breast cancer cells leads to decreased invasive ability and proliferation and decreased anchorage-independent colony formation. However, the exact mechanism by which NAT1 expression affects cancer risk and progression remains unclear. Additionally, consequences of the hydrolysis of acetyl-CoA by NAT1 on cellular metabolism remains uninvestigated.

Hypothesis and Rationale: Samples with decreased levels or knockout of NAT1 will have increased free acetyl-CoA since those cell lines have less NAT1 to hydrolyze acetyl-CoA. Conversely, samples with increased NAT1 will have decreased free acetyl-CoA since those cell lines have more NAT1 to hydrolyze acetyl-CoA. These differences in free acetyl-CoA are hypothesized to lead to alterations in cellular pathways/metabolism when compared to cells with basal NAT1 that can be measured by global bioenergetics, metabolomics, and transcriptomics experiments.

Methods: This dissertation utilized a systems biology approach with four layers. The first layer consists of six constructed MDA-MB-231 cell lines whose only genetic difference (theoretically) is

in NAT1. The second, third, and fourth layers are bioenergetics (Chapter 3), metabolomics (Chapter 4), and transcriptomics (Chapter 5) measurements of the constructed cell lines. Resulting data were analyzed individually and also integrated and analyzed (Chapter 6).

Results: The manipulation of NAT1 in MDA-MB-231 breast cancer cells severely altered cellular metabolism as measured by mitochondrial bioenergetics, metabolomics, and transcriptomics. More differences were observed in the cell lines with decreased levels and knockout NAT1 than the cell line with increased NAT1.

Conclusions: This dissertation has generated novel hypotheses about the role of NAT1 in breast cancer, and more generally cellular metabolism. Furthermore, biochemicals that are likely products of NAT1 *N*-acetylation, *N*-acetylasparagine and *N*-acetylputrescine, have been identified. This dissertation presents strong evidence that NAT1, whether directly or through an effect of NAT1 on acetyl-CoA levels, has an effect on acyl-CoA carnitine conjugates, lysine degradation, and mitochondrial function. While the exact mechanism by which NAT1 affects cellular metabolism or breast cancer progression has not been identified, the data presented in this dissertation add important pieces to the puzzle, putting researchers one step closer to that goal.

TABLE OF CONTENTS

	PAGE
ACKNOWLEDGMENTS.....	iv
ABSTRACT.....	v
LIST OF TABLES.....	xii
LIST OF FIGURES.....	xiv
CHAPTER	
I. BACKGROUND AND SIGNIFICANCE.....	1
a. Breast Cancer	1
b. Arylamine <i>N</i> -Acetyltransferases (NATs)	1
i. Overview.....	1
ii. Regulation.....	2
iii. Brief History of NATs.....	3
iv. NATs and Breast Cancer.....	4
v. Possible Endogenous Role.....	6
c. Construction of MDA-MB-231 Breast Cancer Cell Lines Expressing Varying NAT1.....	7
i. Construction of siRNA Generated Cell Lines.....	7
ii. Construction of CRISPR/Cas9 Generated Cell Lines.....	8
iii. Sequencing of the NAT1 Gene in the CRISPR/Cas9 Constructed Cell Lines.....	12
iv. Summary and Importance.....	13
d. Cancer Metabolism.....	13
e. Systems Biology Approach.....	16
i. Combining Datasets.....	18
f. Dissertation Aims.....	18

i. Aim1.....	20
ii. Aim 2.....	20
iii. Aim 3.....	20
g. Significance.....	22
II. NAT EXPRESSION IN BREAST TISSUES.....	23
a. Background.....	23
b. Materials and Methods.....	23
i. Acquisition of publicly available data from the Cancer Cell Line Encyclopedia (CCLE) and The Cancer Genome Atlas (TCGA) data repositories.....	23
ii. Established breast cancer cell lines analyzed.....	24
iii. Statistical analyses.....	24
c. Results.....	25
i. Association between NAT1 & ESR1, NAT2 & ESR1, and NAT1 & NAT2.....	25
ii. Comparison of NAT1 & NAT2 expression.....	25
iii. Comparison of gene expression between ER+ and ER- samples.....	30
iv. Comparison of NAT1, NAT2, and ESR1 gene expression between normal breast tissue and primary breast tumors.....	31
v. Relationship between previously reported NAT1 <i>N</i> -acetylation activity and NAT1 RNA expression.....	33
vi. Co-expression of NAT1 and NAT2 RNA expression in established breast cancer cell lines.....	33
d. Discussion.....	33
e. Summary and Conclusions.....	41
III. BIOENERGETICS.....	42
a. Background.....	42
b. Materials and Methods.....	42
i. MTT Cell Growth Assays.....	44
c. Results.....	47

d. Discussion.....	53
e. Summary and Conclusions.....	56
IV. METABOLOMICS.....	57
a. Background.....	57
b. Materials and Methods.....	57
i. Sample Accessioning.....	58
ii. Sample Preparation.....	59
iii. QA/QC.....	59
iv. Ultrahigh Performance Liquid Chromatography-Tandem Mass Spectrometry (UPLC-MS/MS)	60
v. Data Extraction and Compound Identification.....	60
vi. Curation.....	61
vii. Metabolite Quantification and Data Normalization.....	61
viii. Statistical Analyses.....	62
c. Results.....	65
i. Univariate Analyses.....	65
ii. Multivariate/Multivariable Analyses.....	72
iii. Pathway Analysis.....	91
d. Discussion.....	91
e. Summary and Conclusions.....	97
V. TRANSCRIPTOMICS.....	102
a. Background.....	102
b. Materials and Methods.....	102
i. Collection of Samples.....	102
ii. Transcriptomics.....	104
1. RNA Isolation.....	104
2. Library Preparation.....	104
3. Library Validation.....	105

4. Denaturing and diluting Libraries for the Nextseq 500.....	105
5. Sequencing.....	106
6. RNA-Seq Analysis.....	106
iii. Quantitative Measurement of NAT1 and NAT2 mRNA.....	108
iv. NAT2 <i>N</i> -Acetylation Activity Assays.....	108
c. Results.....	108
i. Correlation Concordance.....	125
ii. WGCNA Analysis.....	130
iii. Pathway Analysis.....	130
d. Discussion.....	137
e. Summary and Conclusions.....	138
VI. COMBINING DATASETS.....	139
a. Background.....	139
b. Methods and Results.....	139
i. Bioenergetics and Transcriptomics.....	139
ii. Metabolomics and Transcriptomics.....	140
c. Discussion.....	164
d. Summary and Conclusions.....	169
VII. EVIDENCE OF DIFFERENCES OTHER THAN NAT1 IN CONSTRUCTED CELL LINES.....	172
a. Background.....	172
b. Bioenergetics	175
c. Metabolomics	175
d. Transcriptomics	175
e. Combined Analysis.....	176
f. Discussion.....	176
g. Summary and Conclusions	180
VIII. SUMMARY & CONCLUSIONS.....	182

a. Summary	182
b. Strengths of This Work	183
c. Caveats and Weaknesses.....	184
d. Future Directions	187
e. Overall Conclusions.....	189
REFERENCES.....	191
LIST OF ABBREVIATIONS.....	204
CURRICULUM VITAE.....	209

LIST OF TABLES

TABLE		PAGE
1.1	CRISPR/Cas9 Guide RNA Sequences.....	11
3.1	Definition of Measurement Calculations and ANOVA <i>p</i> -value Summary Statistics.....	46
4.1	Metabolomics Summary Statistics.....	66
4.2	Metabolites Concordantly Differentially Abundant in <i>CRISPR 2-19</i> and <i>CRISPR 5-50</i> Cell Lines.....	74
4.3	Metabolites Significantly Correlated with NAT1 <i>N</i> -acetylation Activity.....	76
4.4	<i>N</i> -acetylasparagine Statistics	77
4.5	<i>N</i> -acetylputrescine Statistics	78
4.6	Saccharopine Statistics	79
5.1	Transcriptomics Sample Alignment.....	107
5.2	Transcriptomics Summary Statistics.....	109
5.3	Differential Expression of the methylthioribose-1-phosphate isomerase (MRI1) Gene.....	116
5.4	Differential Expression of the <i>N</i> -acetyltransferase (NAT) Genes.....	119
5.5	Genes Correlated with NAT1 Transcript Abundance.....	122
5.6	Genes Correlated with NAT2 Transcript Abundance (Abbreviated to top 50 genes).....	123
6.1	Differentially Abundant Mitochondrial Genes in <i>Down</i> Cell Line Compared to <i>Scrambled</i>	145
6.2	Differentially Abundant Mitochondrial Genes in <i>CRISPR 2-12</i> Cell Line Compared to <i>Scrambled</i>	146
6.3	Differentially Abundant Mitochondrial Genes in <i>CRISPR 2-19</i> Cell Line Compared to <i>Scrambled</i>	147
6.4	Differentially Abundant Mitochondrial Genes in <i>CRISPR 5-50</i> Cell Line Compared to <i>Scrambled</i>	149

6.5	Mitochondrial Genes Concordantly Differentially Abundant in <i>CRISPR 2-19</i> and <i>CRISPR 5-50</i> Cell Lines.....	150
6.6	sPLS-DA Component 1 Metabolites.....	159
6.7	sPLS-DA Component 1 Genes.....	160
6.8	sPLS-DA Component 2 Metabolites.....	162
6.9	sPLS-DA Component 2 Genes.....	163
6.10	Differential Expression of the SAT Genes.....	165
7.1	Guide RNA 2 predicted targets.....	178
7.2	Guide RNA 5 predicted targets.....	179

LIST OF FIGURES

FIGURE	PAGE
1.1 Construction of MDA-MB-231 Breast Cancer Cells Expressing Varying Levels of NAT1 Activity Using siRNA.....	9
1.2 Construction of MDA-MB-231 Breast Cancer Cells Expressing Varying Levels of NAT1 Activity Using CRISPR/Cas9.....	10
1.3 NAT1 <i>N</i> -acetylation Activity of the Six Transformed Cell Lines.....	14
1.4 NAT1 mRNA of the Six Transformed Cell Lines.....	15
1.5 Systems Biology Publications by Year.....	17
1.6 Systems Biology Approach.....	19
1.7 Overall Dissertation Hypothesis and Rationale.....	21
2.1 Scatterplot and Correlation Matrices for <i>NAT1</i> , <i>NAT2</i> , and <i>ESR1</i>	26
2.2 <i>NAT1</i> and <i>NAT2</i> RNA Expression in Breast Cancer Cell Lines, Primary Breast Tumor Samples, and Normal Breast Tissue Samples.....	28
2.3 <i>ESR1</i> , <i>NAT1</i> and <i>NAT2</i> RNA Expression in Breast Cancer Cell Lines, Primary Breast Tumor Samples, and Normal Breast Tissue Stratified by ER Status.....	31
2.4 Comparison of <i>ESR1</i> , <i>NAT1</i> and <i>NAT2</i> RNA Expression in Normal Breast Tissue and Primary Breast Tumor Samples.....	34
2.5 Association Between <i>NAT1</i> RNA Expression and Previously Reported NAT1 <i>N</i> -acetylation Activity in Seven Established Breast Cancer Cell Lines.....	35
2.6 <i>NAT1</i> and <i>NAT2</i> Association in Breast Cancer Cell Lines.....	36
3.1 Bioenergetics Experimental Approach.....	45
3.2 Bioenergetics Oxygen Consumption Rate Measurements.....	48
3.3 Bioenergetics Extracellular Acidification Rate Measurements.....	51
3.4 Bioenergetics Coupling Efficiency Measurement.....	52

3.5	Bioenergetics MTT Assay Equivalence Test.....	54
4.1	Metabolomics Experimental Design Diagram.....	63
4.2	Assessing Normality of Metabolomics Data.....	64
4.3	Scrambled Versus Up Volcano Plot.....	67
4.4	Scrambled Versus Down Volcano Plot.....	68
4.5	Scrambled Versus CRISPR 2-12 Volcano Plot.....	69
4.6	Scrambled Versus CRISPR 2-19 Volcano Plot.....	70
4.7	Scrambled Versus CRISPR 5-50 Volcano Plot.....	71
4.8	Concordance of Metabolite Abundance Differences in CRISPR NAT1 Knockout Cell Lines.....	73
4.9	<i>N</i> -acetylasparagine Abundance Distribution and Correlation with NAT1 <i>N</i> -acetylation Activity.....	80
4.10	<i>N</i> -acetylputrescine Abundance Distribution and Correlation with NAT1 <i>N</i> -acetylation Activity.....	82
4.11	Saccharopine Abundance Distribution and Correlation with NAT1 <i>N</i> -acetylation Activity...84	
4.12	Metabolomics Heatmap and Hierarchical Clustering.	87
4.13	Principal Component Analysis Scores Plot.....	89
4.14	Principal Component Loadings Plot.	90
4.15	Metabolomics Pathway Enrichment Analysis.....	93
4.16	Possible Association Between NAT1 and <i>N</i> -acetylasparagine.	94
4.17	Possible Association Between NAT1 and <i>N</i> -acetylputrescine.	96
4.18	Possible Association Between NAT1 and Saccharopine.	98
4.19	Lysine Degradation Pathway.....	99
5.1	Transcriptomics Experimental and Data Analysis Approach.....	103
5.2	<i>Up</i> vs <i>Scrambled</i> Volcano Plot.	110
5.3	<i>Down</i> vs <i>Scrambled</i> Volcano Plot.	111
5.4	<i>CRISPR 2-12</i> vs <i>Scrambled</i> Volcano Plot.....	112
5.5	<i>CRISPR 2-19</i> vs <i>Scrambled</i> Volcano Plot.....	113

5.6	<i>CRISPR 5-50 vs Scrambled</i> Volcano Plot.....	114
5.7	Gene Concordance Between <i>CRISPR 2-19</i> and <i>CRISPR 5-50</i> Cell Lines Compared to <i>Scrambled</i>	117
5.8	<i>NAT1</i> and <i>NAT2</i> Gene Expression in Each Cell Line Compared to <i>Scrambled</i> and <i>Up</i>	118
5.9	Integrated Genomics Viewer (IGV) <i>NAT2</i> Transcript Mapping.....	120
5.10	<i>NAT1</i> and <i>NAT2</i> mRNA Expression.....	121
5.11	Hierarchical Clustering and Heatmap of Each Cell Lines Global Transcriptomics Signature.....	126
5.12	Hierarchical Clustering and Heatmap of Differential Gene Expression Compared to <i>Scrambled</i>	128
5.13	Transcriptomics Principal Components Analysis.....	131
5.14	WGCNA Transcript Module Branches by Color.....	132
5.15	Correlation Between WGCNA Transcript Modules and <i>NAT1 N-acetylation Activity</i>	133
5.16	Heatmap of Genes in WGCNA White Module.....	134
5.17	Transcriptomics Enrichment Analysis.....	135
6.1	Heatmap of Nuclear-encoded Mitochondrial Gene Expression.....	141
6.2	Nuclear-encoded Mitochondrial Gene Expression.....	143
6.3	Combined Metabolomics and Transcriptomics Enrichment Analysis.....	151
6.4	Metabolites and Genes Projected onto the KEGG Lysine Degradation Pathway.....	153
6.5	Optimization of Number of Components Included in sPLS-DA.	156
6.6	sPLS-DA Scores Plot.....	157
6.7	sPLS-DA Component 1 Loadings Plot.	158
6.8	sPLS-DA Component 2 Loadings Plot.....	161
6.9	Circos Plot of Metabolites and Transcripts Included in sPLS-DA Model.....	166
6.10	Combined Metabolomics and Transcriptomics Hierarchical Clustering and Heatmap.....	167
6.11	Overall Dissertation Conclusion.....	170
7.1	Evidence of Additional Differences Between Cell Lines Other Than <i>NAT1</i>	173
8.1	Limitations of “Global” Metabolomics.....	186

CHAPTER 1

BACKGROUND AND SIGNIFICANCE

Breast Cancer

According to the American Cancer Society (ACS), breast cancer will account for approximately 30% of new cancer cases and 14% of cancer related deaths in American women in 2018¹. Additionally, 1 in 8 women will develop breast cancer in their lifetime¹. Breast cancer is a heterogeneous disease with many underlying genetic transformations that lead to a diseased state. This heterogeneity makes predicting breast cancer risk and developing one-size-fits-all treatment strategies difficult. As the cost of genomic profiling has decreased dramatically, the feasibility of developing personalized treatment strategies based on the genomic profile of an individual's tumor has become attainable. For example, the National Cancer Institute (NCI) presently has a clinical trials program called Molecular Analysis for Therapy Choice (MATCH) in which DNA of tumors from patients for whom standard treatment has not been successful are sequenced and the results are utilized to provide personalized treatment²; the program currently has 14 treatment arms to which patients are assigned based on mutations, including amplifications, translocations, and fusions, in 14 different genes: *EGFR*, *MET*, *ALK*, *ROS1*, *HER2*, *mTOR*, *TSC1*, *TSC2*, *GNAQ/GNA11*, *SMO/PTCH1*, *cKIT*, *NTRK*, *BRCA1*, and *BRCA2*². These advances in precision medicine have led to an urgent need to understand and catalogue gene targets involved in breast cancer initiation, transformation, and metastasis so that preventative and treatment strategies can be personalized and optimized for each patient.

N-acetyltransferases (NATs)

Overview

Arylamine *N*-acetyltransferase 1 and arylamine *N*-acetyltransferase 2 (NAT1 and NAT2; collectively NATs) are polymorphic phase II xenobiotic metabolizing isozymes that catalyze the acetylation of a wide range of arylamine and hydrazine substrates using acetyl coenzyme A (acetyl-CoA) as the acetyl group donor via a ping-pong bi-bi reaction mechanism^{3,4}. The human NAT1 and NAT2 genes are located in close proximity on chromosome 8p22⁵ and share approximately 87% protein coding sequence and 81% deduced amino acid sequence homology⁶. The open reading frame (ORF) regions of *NAT1* and *NAT2* are both 870 base pairs and intronless^{6,7}. NAT1 and NAT2 have distinct but overlapping substrate specificities with NAT1 selectively acetylating *p*-aminobenzoic acid (PABA) and *p*-aminosalicylate (PAS) while NAT2 selectively acetylates isoniazid, sulfamethazine (SMZ), procainamide, and hydralazine⁸. Both isozymes can acetylate a plethora of arylamines including 4-aminobiphenyl (4-ABP) and 2-aminofluorene (2-AF)⁸. Many NAT substrates are known carcinogens found in well-done red meats, cigarette smoke, and commercial dyes^{9,10}. Additionally, NAT1 and NAT2 have differing tissue distributions (reviewed in ¹¹); NAT1 is expressed in nearly every tissue¹²⁻²² while NAT2 tissue distribution is thought to be much more limited with the highest expression found in the liver¹² and gastrointestinal (GI) tract¹⁹.

Regulation of NATs

NAT1 and NAT2 expression also vary inter-individually, but independent of one another^{12,23}, based on single nucleotide polymorphisms (SNPs)^{4,23-29}. As of the most recent update (April 2016), 28 different *NAT1* alleles³⁰ and 64 (108 including those predicted) different *NAT2* alleles³¹ have been identified and named in human populations, conferring either rapid, intermediate, slow, or unknown acetylator phenotypes. *NAT1* and *NAT2* allelic frequencies vary by ethnic population³². The effect of SNPs in *NAT2* on *N*-acetylation activity have been well described and characterized (reviewed in ²⁷) however the effect of SNPs (located both in and outside the ORF) in *NAT1* on *N*-acetylation activity are not understood as well with conflicting reports (reviewed in ³³). Large inter-individual variation exists in NAT1 activities within single genotypes suggesting additional regulation mechanisms are present³⁴⁻³⁶. Additionally, there is considerable evidence that NAT1 activity is highly regulated by post-transcriptional and post-

translational mechanisms, and substrate concentrations (reviewed in ³⁷). Although the ORF of *NAT1* is contained in a single exon (exon 9), there are 8 upstream non-coding exons³⁸ that are included, in varying patterns, in alternatively spliced transcripts²¹. *NAT1* is also known to have two alternative promoters, NATa and NATb, that differ in promotor strength and tissue specificity³⁹⁻⁴¹ producing mRNAs with distinct 5'-UTRs. Notably, there are 9 variant *NAT1* transcripts known, each with identical protein coding sequences^{38,39,41}.

Through chemical modifications of the *NAT1* active site as well as changes in the turnover of *NAT1* protein, *NAT1* activity can be modulated post-translationally. *NAT1* substrates, including PABA, *p*-aminosalicylic acid, ethyl-*p*-aminobenzoate, and *p*-aminophenol, are capable of down-regulating *NAT1* activity and protein expression in cultured human peripheral blood mononuclear cells⁴²; this phenomenon was also observed in confluent human cell lines incubated with PABA⁴². It is hypothesized this occurs because substrate binding causes the cysteine in the *NAT1* active site to become de-acetylated leading to polyubiquitination and degradation of the protein³⁷. *N*-hydroxyl metabolites of *NAT1* substrates, including PABA and sulfomethoxazole, have also been shown to irreversibly inhibit *NAT1*⁴³; however, the molecular mechanism remains unknown. *NAT1* gene expression can also be regulated indirectly by androgens^{44,45}. Additionally, *NAT1* is known to be inhibited by many chemically diverse non-substrate compounds including tamoxifen⁴⁶, cisplatin⁴⁷, thiram (a dithiocarbamate pesticide)⁴⁸, isothiocyanates⁴⁹, hydrogen peroxide and peroxyxynitrite^{50,51}, acrolein⁵², carbon black nanoparticles⁵³, cadmium⁵⁴, and disulfiram⁵⁵. Additionally, hsa-miRNA-1290, implicated in lung adenocarcinoma⁵⁶, ovarian cancer⁵⁷, non-small cell lung cancer^{58,59}, glioma⁶⁰, and colorectal cancer⁶¹, has been shown to directly target the 3'UTR region of human *NAT1* mRNA leading to dose-dependent down-regulation of *NAT1* expression^{62,63}.

Brief History of NATs

Arylamine *N*-acetyltransferase activity was first described in the late 1930's when acetylated *para*-aminobenzenesulfonamide was detected in the urine of humans and rabbits, but not dogs, after administration of *para*-aminobenzenesulfonamide⁶⁴, although it would be nearly 40 years before the specific enzyme(s) responsible for the transformation was/were isolated and

identified. NAT activity was described for a second time in the mid 1940's when it was observed that liver homogenates and extracts could acetylate sulfanilamides⁶⁵ and again in the mid 1950's when it was shown that the enzyme responsible for the acetylation of sulfanilamides was also responsible for the acetylation of isoniazid⁶⁶. NATs were first isolated and described kinetically in 1967 by Weber and Cohen in partially purified preparations from rabbit liver⁶⁷. NAT was shown to catalyze the acetylation of isoniazid and several sulfonamides via a ping-pong bi-bi reaction mechanism using acetyl-CoA as the acetyl donor⁶⁷. It is also important to note that at the times of those early studies, NAT1 and NAT2 were thought to be a single enzyme that could be monomorphic or polymorphic depending on the substrate acetylated. It was not realized until 1990 that there were actually two individual *N*-acetylating isozymes^{6,12} with distinct but overlapping substrate specificities.

NATs & Breast Cancer

While NAT1 has been implicated in many cancers, NAT1's role in breast cancer has been investigated more frequently than any other cancer. The earliest studies on *N*-acetyltransferases and breast cancer were epidemiological studies that investigated associations between acetylator phenotype and breast cancer risk⁶⁸⁻⁷². Since those initial studies there have been many more looking at the association in various ethnic and regional populations and in combination with several possible modifying factors with conflicting results leading to multiple meta-analyses of these data.

Briefly, of the 5 meta-analyses published on *NAT2* and breast cancer risk, 4 investigated the independent effect of *NAT2* phenotype (determined by genotype) on breast cancer risk with none observing a significant association⁷³⁻⁷⁷. However, one of the meta-analyses found evidence of decreased breast cancer risk for women with a slow phenotype when separately analyzing studies in which *NAT2* phenotype was determined by acetylation rate of xenobiotics metabolized by *NAT2* rather than genotype⁷⁶. Three of the 5 meta-analyses further analyzed data by including cumulative smoking exposure into the risk calculation as a modifying variable with all 3 studies observing a significant association between high pack years and increased breast cancer risk in the slow acetylator group compared to the rapid acetylator group^{74,75,77}. A single meta-analysis

considered menopausal status as a modifying factor of *NAT2* genotype in breast cancer risk and observed no significant difference in risk⁷⁴. None of the meta-analyses included intake of well-done meat as a possible modifying factor of *NAT2* genotype in breast cancer risk. There has been one meta-analysis of *NAT1* genotype and breast cancer risk which found no significant association⁷⁸, however possible modifying factors such as smoking exposure and well-done meat intake were not considered in the study.

Although *NAT1* and *NAT2* both catalyze *N*-acetylation and have been associated with breast cancer, their roles in etiology may differ. Numerous studies have investigated possible roles for *NAT1* in breast cancer etiology and progression⁷⁹⁻⁸⁴, given the association between increased expression of *NAT1* and estrogen receptor (ER)-positive breast cancers⁸⁵⁻⁹¹. Notably, *NAT1* expression is not directly regulated by estrogens or dihydrotestosterone⁹², thus suggesting that there may be a common regulatory element between *NAT1* and *ESR1*. Furthermore, congenic rats expressing higher *NAT2* activity (orthologous to human *NAT1*) have been reported to exhibit greater carcinogen-induced mammary tumor susceptibility independent of carcinogen metabolism⁸³. Notably, *NAT1* is one of fifty genes whose expression is utilized in the PAM50-based Prosigna breast cancer gene signature array to determine the intrinsic subtype of breast cancer tumors^{93,94}. Conversely, SNPs in *NAT2* have been well described and revealed to influence acetylation rates of many known carcinogens; an association between *NAT2* genotype with breast cancer risk among smokers has been reported (reviewed in ⁹⁵).

The mRNA expression levels of *NAT1* and *NAT2* have been detected by reverse transcription-polymerase chain reaction (RT-PCR) in human mammary tissue^{13,96}. *NAT1* *N*-acetylation activity has been widely reported in normal breast tissue and breast tumor tissue^{13,14,38,97-99} whereas *NAT2* *N*-acetylation activity has not been observed as consistently; when *NAT2* activity is observed the activity is much lower than *NAT1* activity^{13,97,98}. In addition, since *NAT1* and *NAT2* have overlapping substrate specificities, activity studies of the two isozymes can be complex. For example, Deitz probed human mammary tissue samples for *NAT1* and *NAT2* activities with *p*-aminobenzoic acid (PABA; selective for *NAT1*) and sulfamethazine (SMZ; selective for *NAT2*), and reported that SMZ was acetylated by *NAT1* at very low levels⁹⁹.

By normalizing the SMZ *N*-acetylation activity to NAT1 activity, Deitz demonstrated that the SMZ *N*-acetylation activity was most likely catalyzed by NAT1 rather than NAT2. NAT1 and NAT2 activities have also been reported in rat mammary tissues¹⁰⁰.

Wakefield *et al* profiled *NAT1* expression and activity in seven breast cancer cell lines (MCF-7, T47D, ZR-75-1, Cal51, MDA-MB-231, MDA-MB-437 and MDA-MB-453) and detected *NAT1* mRNA expression and activity in all seven cell lines⁸⁸; however, *NAT2* expression and activity were not co-investigated. In addition, *NAT2* mRNA has been detected in MCF-7 breast cancer cells at very low levels⁹⁶; however, *NAT1* was not measured at the same time preventing a direct comparison of expression between the two isozymes. Bradshaw *et al* detected NAT1 and NAT2 by western blotting in the ER-positive breast cancer cell line MCF-7; however, the expression levels were not compared between the two proteins¹⁰¹. Based on limited data, it has been hypothesized that *NAT2* expression is very low in breast tissue and negligible in comparison to *NAT1* expression. However, a more comprehensive evaluation of NAT1 and NAT2 co-expression in breast tissues and established breast cancer cell lines is needed.

Minchin and Butcher have recently shown, utilizing the publicly available data repositories Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) and The Cancer Genome Atlas (TCGA), that NAT1 mRNA expression in primary breast tumors, but not prostate, cervical, or colorectal tumors, can be segregated into three distinct patient subpopulations, expressing low, intermediate and high levels of NAT1 mRNA¹⁰². Gene correlation analysis with NAT1 in each subset suggested different regulation mechanisms are in place, given the small overlap in genes correlated with NAT1 mRNA expression in each subpopulation¹⁰². Additionally, a better 10-year survival rate in patients with high NAT1 was observed compared to patients with low NAT1¹⁰². Notably, patients whose tumors had low NAT1 expression exhibited a distinct poor response to chemotherapy¹⁰². It will be important for future studies on NAT1 to investigate how NAT1 expression in breast tumors influences chemo-sensitivity.

Possible Endogenous Role of NAT1

Although NAT1 has traditionally been described as a xenobiotic metabolizing enzyme, there is a wealth of evidence suggesting NAT1 may also have unknown endogenous role(s).

Phylogenetic analysis of the NATs suggested evolutionarily NAT1 did not tolerate mutations while NAT2 was predicted to have mostly evolved under positive selection¹⁰³. NATs are found in all kingdoms of life, excluding plants¹⁰⁴. Additionally, NAT1 activity and mRNA have been detected in mid-gestational fetal tissues¹⁵, placenta^{105,106}, and nearly every adult tissue. NAT1 has been shown to have redundancy with methylthioribose-1-phosphate isomerase (MRI1) in the methionine salvage pathway, but this role was not observed in MDA-MB-231 cells since they do not have a functional methionine salvage pathway due to a lack of methylthioadenosine phosphorylase (MTAP) activity¹⁰⁷. To date, NAT1 is known to acetylate only one endogenous compound, *p*-aminobenzoylglutamate (pABG)-- a folate catabolite^{108,109}, with unknown consequence. However, NAT1 has been linked to folate metabolism^{110,111}. Additionally, it has also been shown that in the absence of an arylamine substrate NAT1 can catalyze the hydrolysis of acetyl-CoA using folate as a co-factor^{112,113}. Furthermore, NAT1 expression and activity are highly regulated by many diverse mechanisms whereas NAT2 expression and activity appear to be less regulated. Taken together, these facts lead to the hypothesis that NAT1 is not just a xenobiotic metabolizing enzyme. It remains to be determined whether that role is through a NAT1 catalyzed reaction or a result of NAT1's ability to catalyze the hydrolysis of acetyl-CoA, a central molecule in metabolism and cellular energetics.

Construction of MDA-MB-231 Breast Cancer Cell Lines Expressing Varying NAT1

To better investigate NAT1's role in breast cancer a single established breast cancer cell line was genetically modified to vary only in NAT1. The MDA-MB-231 triple negative breast cancer cell line was selected for modification because it expresses an approximate mid-level of NAT1 RNA compared to other breast cancer cell lines; NAT1 RNA expression has been shown to be a suitable predictor of NAT1 activity¹¹⁴. Additionally, two different construction methods were utilized, siRNA and CRISPR/Cas9, to ensure observed results were due to manipulation of NAT1 rather than the specific construction method utilized.

Construction of siRNA Generated Cell Lines

The construction and characterization of the siRNA generated cell lines has been described extensively elsewhere^{82,84,115,116}. Briefly, three cell lines were constructed, *Scrambled*,

Up, and *Down*, expressing parental, increased, and decreased NAT1 *N*-acetylation activity, respectively (Figure 1.1). Each cell line was constructed from the parent MDA-MB-231 cell line, purchased from American Type Culture Collection (ATCC), that had a flippase recognition target (FRT) site inserted, thus ensuring all plasmids were transfected into the same genomic location. Genetically modified cell lines were characterized for NAT1 PABA *N*-acetylation, NAT1 mRNA expression, endogenous acetyl-CoA levels, cell doubling rates, anchorage-independent growth, anchorage-dependent growth, and relative invasive ability. The *Scrambled* cell line was included as a transfection control and expressed approximately the same NAT1 *N*-acetylation activity and NAT1 mRNA as the parent MDA-MB-231 cell line. NAT1 PABA *N*-acetylation activity and NAT1 mRNA in the *Up* cell line were increased 7-fold and 4-fold, respectively, compared to the *Scrambled* cell line. Conversely, NAT1 PABA *N*-acetylation activity and NAT1 mRNA in the *Down* cell line were decreased approximately 35% and 53%, respectively, compared to the *Scrambled* cell line. Doubling times, relative invasive ability, and anchorage-dependent growth of the three constructed cell lines did not significantly vary while anchorage-independent growth was decreased 7-fold in the *Down* cell line compared to the *Scrambled* cell line⁸².

Construction of CRISPR/Cas9 Generated Cell Lines

The CRISPR/Cas9 constructed cell lines were constructed from the MDA-MB-231 cell line that had an FRT site inserted to allow direct comparison to a single cell line regardless of the method used to construct the cell line (Figure 1.2). All cell lines were cultured in high-glucose Dulbecco's Modified Eagle Medium (DMEM), with 10% fetal bovine serum, 5% glutamine, and 5% penicillin/streptomycin added and grown in a humidified incubator set at 37 °C with 5% CO₂. Horizon Discovery Group (Waterbeach, United Kingdom) designed 5 different guide RNAs (gRNAs) specific for NAT1 and DNA2.0 Inc. (Newark, CA) cloned the gRNAs into a Cas9 expressing vector that also expressed a dashed-green fluorescent probe (GFP) tag. Separately, each of the 5 gRNA/Cas9 vectors (gRNA sequences listed in Table 1.1) were transiently transfected in the *Scrambled* MDA-MB-231 cell line using the Amaxa Nucleofector II (Lonza, Allendale, NJ). Forty-eight hours after transfection cells were harvested and DNA isolated.

Figure 1.1

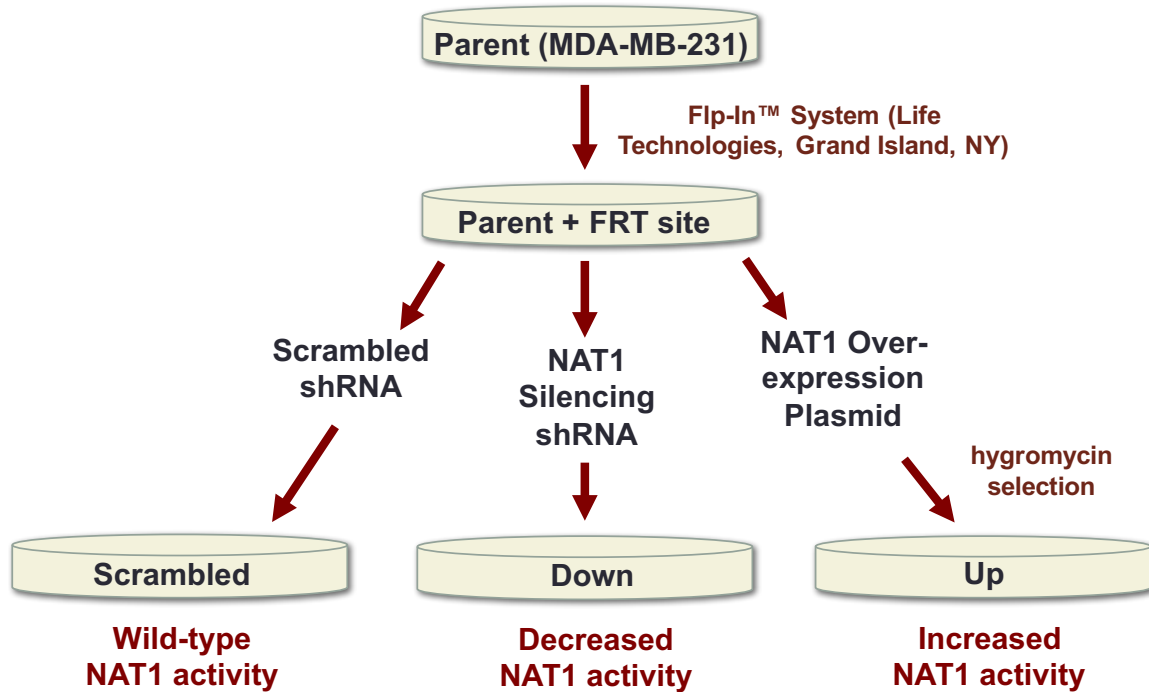


Figure 1.1: Construction of MDA-MB-231 Cell Lines via siRNA.

The Flp-In™ System (Life Technologies, Grand Island, NY) was used to insert a single flippase recognition site (FRT) site into the parent MDA-MB-231 cell line. The MDA-MB-231 cell line containing the FRT site was then stably transfected with either a plasmid containing a nonspecific shRNA plasmid into the FRT site (*Scrambled*), a plasmid containing NAT1 specific shRNA into the FRT site (*Down*), or a plasmid containing the NATb/NAT1*4 vector into the FRT site (*Up*). Each plasmid contained a hygromycin resistant cassette. Therefore stably transfected cells were selected for using hygromycin. The resulting MDA-MB-231 cell lines, *Scrambled*, *Down*, and *Up*, express wild-type, decreased, and increased NAT1 mRNA and PABA *N*-acetylation activity, respectively.

Figure 1.2

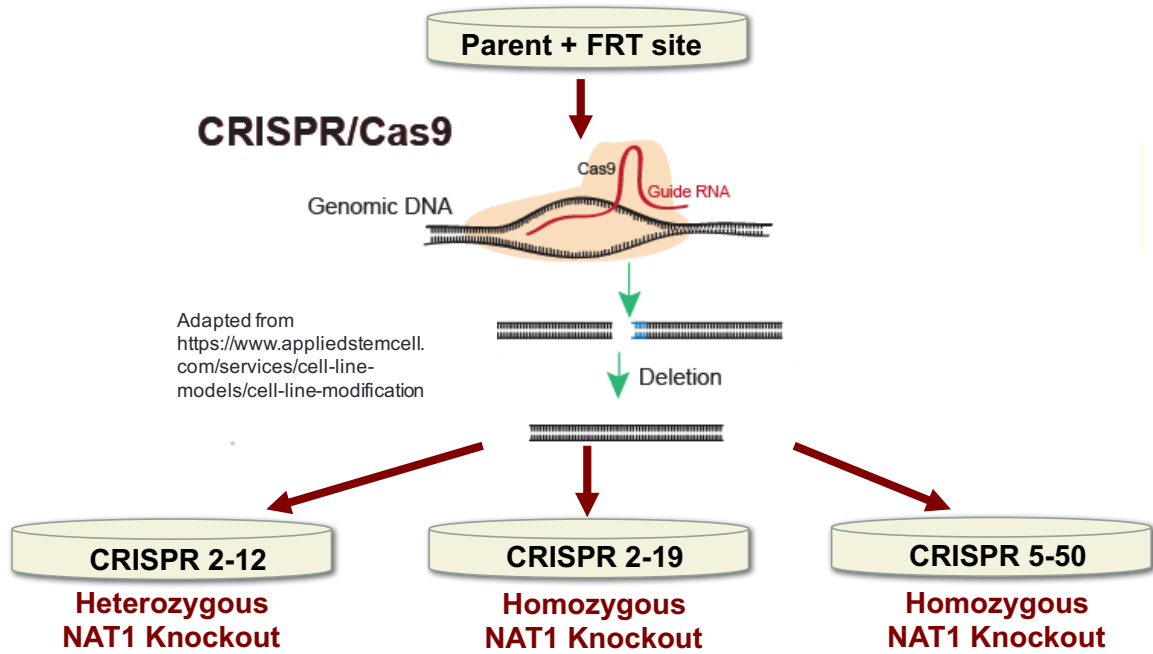


Figure 1.2: Construction of MDA-MB-231 NAT1 Knockout Cell Lines via CRISPR/Cas9.

The CRISPR/Cas9 system was used to decrease and knockout NAT1 in the parent MDA-MB-231 cell line with a previously inserted flippase recognition target (FRT) site utilizing two different guide RNAs, 2 and 5. The resulting MDA-MB-231 cell lines, *CRISPR 2-12*, *CRISPR 2-19*, and *CRISPR 5-50*, express decreased, knockout, and knockout levels of PABA NAT1 *N*-acetylation activity, respectively. The *CRISPR 2-12* and *CRISPR 2-19* cell lines were constructed with guide RNA 2 while the *CRISPR 5-50* cell line was constructed using guide RNA 5.

Table 1.1

CRISPR/Cas9 Guide RNA Sequences

GUIDE RNA	SEQUENCE
1	GAACCTTAACATCCATTGTG
2	CAAAGGGAACAGCTCGGATC
3	GTTGTGAGAAGAAATCGGGG
4	GGCCTCTAAGCCTAAGTCCA
5	GAATTGGCTATAAGAAGTCT

Guide RNAs 2 and 5 were utilized to construct two homozygous NAT1 knockout MDA-MB-231 cell lines and a heterozygous NAT1 knockout MDA-MB-231 cell line.

The Transgenomic Inc. (Omaha, NE) SURVEYOR Mutation Detection Kit was used to determine the effectiveness of each gRNA's ability to cut the genomic DNA and induce DNA strand breaks effectively. gRNAs #2 and #5 were the most effective at inducing DNA strand breaks, and were chosen to separately knockout the function of NAT1 in the following studies.

The *Scrambled* MDA-MB-231 cell line was transfected with either #2 or #5 gRNA/Cas9 vectors as described above. Forty-eight hours after transfection cells were sorted for GFP fluorescence. The fluorescent positive cells were collected and plated at very dilute cell concentrations so that individual clones could be isolated. Once individual cells had grown into large enough colonies (several weeks), cloning cylinders were utilized to isolate those colonies using trypsin to release them from the plate and transferred to a 96-well culture plate. Clones were passaged until there were enough cells to plate in a 10 cm dish. Cells were then tested for NAT1 activity as previously described¹¹⁷. Activity assays showed NAT1 activity was not detectable (knocked out) in a low number of clones and these clones were selected for further characterization. Clones with no detectable NAT1 activity were further screened by sequencing the NAT1 open reading frame (ORF). We were specifically interested in clones that had deleted/inserted nucleotides in the NAT1 ORF that resulted in frame-shift mutations and thus premature protein termination signals resulting in predicted nonfunctional NAT1. Individual knockout cell lines representing the knockout of NAT1 activity for gRNA #2 or #5 were chosen based on NAT1 enzymatic activity and genomic sequence.

Sequencing of the NAT1 Gene in the CRISPR/Cas9 Constructed Cell Lines

Genomic DNA was isolated from the MDA-MB-231 CRISPR/Cas9 constructed cell lines. The NAT1 open reading frame was TOPO cloned using pcDNA™ 3.1/V5-His TOPO® TA Expression Kit (Invitrogen, California, USA) following manufacturer's recommendations. The TOPO Cloning reaction for each cell line was transformed into One Shot TOP10 Chemically Competent *E. coli*. Five transformed *E. coli* colonies for each cell line were selected and grown overnight. Cultures of bacteria were then harvested for plasmid purification. Purified plasmids and primers were sent for DNA sequencing (Eurofins, Louisville, KY, USA). Sequence data was analyzed and aligned with SeqMan Pro™ (Version 12.0, DNASTAR, Madison, WI.)

Characterization of CRISPR/Cas9 constructed cell lines are described in detail elsewhere¹¹⁵. Briefly, the *CRISPR 2-19* and *CRISPR 5-50* cell lines had no detectable NAT1 activity while NAT1 *N*-acetylation activity in the *CRISPR 2-12* cell line was decreased approximately 50% compared to the parent MDA-MB-231 and *Scrambled* cell lines. Endogenous acetyl-CoA was increased approximately 2-fold in both NAT1 knockout cell lines compared to the parent MDA-MB-231 cell line. Doubling times, relative invasive ability, and anchorage-dependent colony formation did not significantly vary between two NAT1 knockout cell lines and the parent MDA-MB-231 cell line. Notably, NAT1 knockout cell lines formed less anchorage-independent colonies compared to the parent MDA-MB-231 cell line¹¹⁵.

Summary and Importance

Colleagues, Mark Doll and Dr. Marcus Stepp, have constructed six MDA-MB-231 cell lines via two different methodologies that express parental, increased, decreased, and no detectable (limit of detection = 0.05 nmoles acetylated PABA/min/mg) NAT1 *N*-acetylation activity (Figure 1.3). All cell lines have been confirmed by STR-profiling to be MDA-MB-231 breast cancer cells. NAT1 mRNA has also been quantitated in each cell line (Figure 1.4). The genetic modification of a single established breast cancer cell line that varies only NAT1 provides an excellent model for studying the possible role of NAT1 in breast cancer metabolism. While breast cancer cell lines that endogenously express varying levels of NAT1 could have been selected, that would also complicate the analysis of results due to the multiple unique mutations each cell line would have. The approach utilized in this dissertation ensures each cell line has the same genetic background and only varies in NAT1 allowing a more straightforward comparison between cell lines.

Cancer Metabolism

Cancer as a disease can be concisely described as a dysregulation of normal cellular functions leading to increased and uncontrolled growth. Hanahan and Weinberg suggested in 2000 that cancer could be summarized by six basic acquired capabilities or hallmarks: 1. evading apoptosis, 2. self-sufficiency in growth signals, 3. insensitivity to anti-growth signals, 4. tissue

Figure 1.3

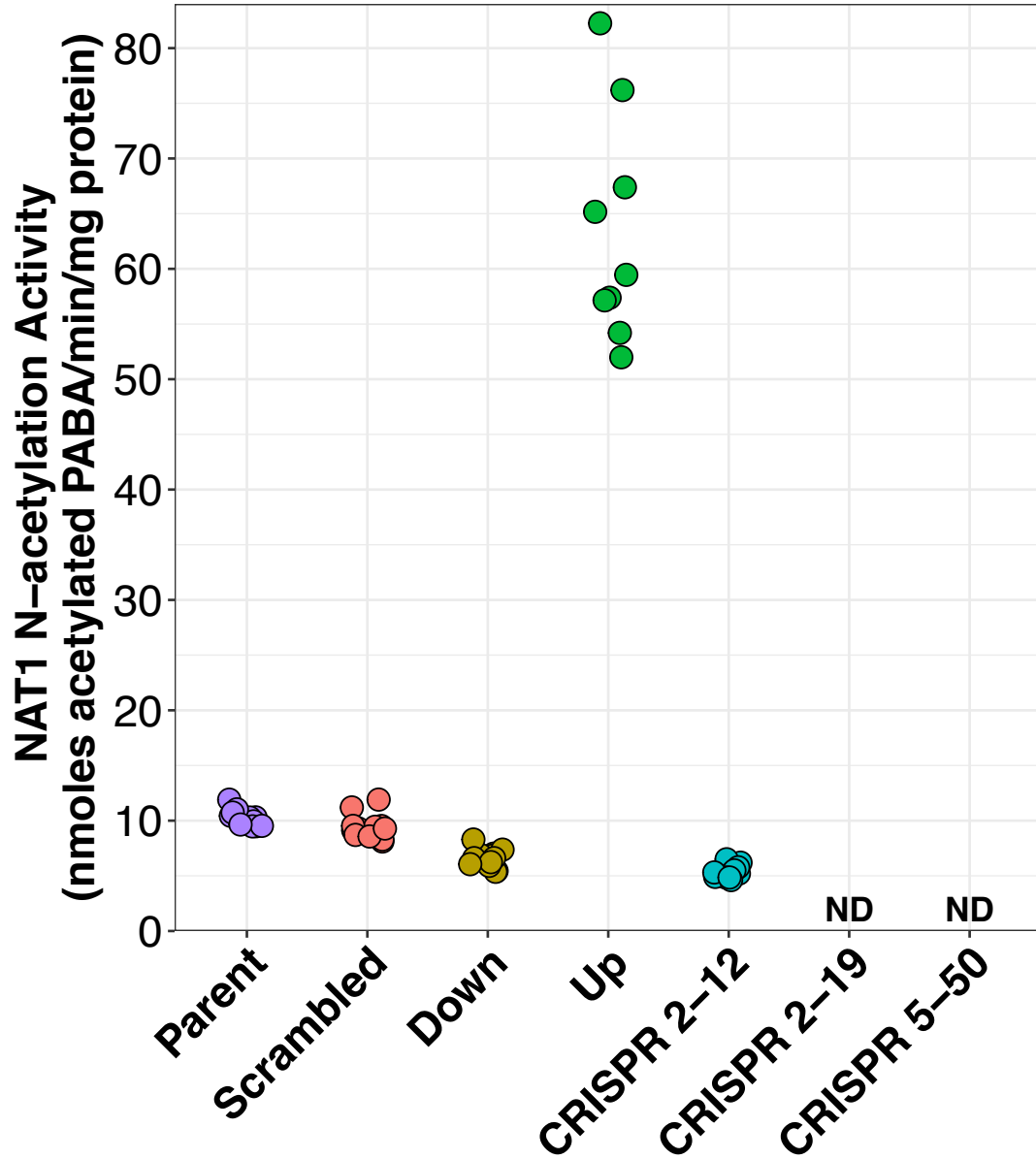


Figure 1.3: NAT1 N-Acetylation Activity of the Parent and Genetically Modified MDA-MB-231 Cell Lines.

NAT1 N-acetylation activity was measured in the 6 cell lines included in this study. The *Parent* and *Scrambled* cell lines express approximately the same level of NAT1 N-acetylation activity while NAT1 N-acetylation activity in the *Down* and *CRISPR 2-12* cell lines is decreased by approximately 50%. NAT1 N-acetylation activity in the *Up* cell line is increased by approximately 700%. The *CRISPR 2-19* and *CRISPR 5-50* had no detectable NAT1 N-acetylation activity.

Figure 1.4

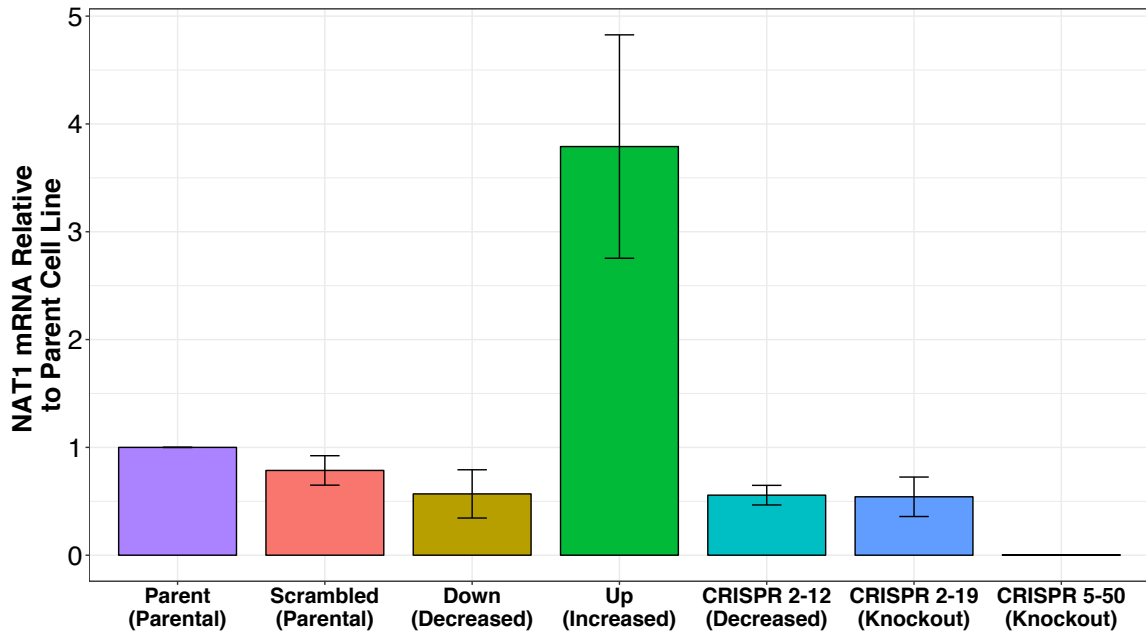


Figure 1.4: NAT1 mRNA of the Parent and Genetically Modified MDA-MB-231 Cell Lines

Relative to Parent.

NAT1 mRNA was quantitated in the 6 cell lines included in this study. The *Parent* and *Scrambled* cell lines express approximately the same level of NAT1 mRNA while NAT1 mRNA in the *Down* *CRISPR 2-12*, and *CRISPR 2-19* cell lines is decreased by approximately 50%. NAT1 mRNA in the *Up* cell line is increased by approximately 400%. The *CRISPR 5-50* cell line had very low levels of NAT1 mRNA.

invasion and metastasis, 5. limitless replicative potential, and 6. sustained angiogenesis¹¹⁸. A decade later Hanahan and Weinberg updated the hallmarks of cancer to add two new emerging hallmarks: 1. deregulating cellular energetics and 2. avoiding immune destruction¹¹⁹. Based on these hallmarks of cancer, the dysregulation of metabolism and energetics plays a crucial role in cancer risk and progression by enabling uncontrolled cell growth. A better understanding of the role NAT1 has in metabolism and cellular energetics could help lead to a better understanding of the connection between NAT1 and breast cancer. Given that NAT1 can catalyze the hydrolysis of acetyl-CoA, a central biochemical in metabolism and cellular energetics¹²⁰⁻¹²², different levels of NAT1 are hypothesized to lead to observable differences in mitochondrial function, metabolites, and transcripts. Acetyl-CoA has a central role in fatty acid synthesis and degradation, ketone body and isoprenoid synthesis, and feeds into the citric acid cycle.

Systems Biology Approach

There are three concepts, 1. emergence, 2. robustness, and 3. modularity, that are necessary for understanding complex biological systems¹²³. The first concept, emergence, states that complex systems display properties as a whole that cannot otherwise be predicted by studying individual parts. The second concept, robustness, states there are many diverse mechanisms in place for biological systems to maintain phenotypic stability when presented with perturbations whether through the environment, stochastic events or genetic variation. The final concept, modularity, states complex systems are divided into specialized modules allowing damage to be confined as well as robustness. In many studies of cancer and/or metabolism a reductionist approach, rather than a systems biology approach, is taken because studies of cancer biology and cellular metabolism are complex and can be difficult to interpret especially in multidimensional biological systems; however, a systems biology approach provides many advantages. Publications of systems biology studies has grown exponentially over the last 15 years (by count of PubMed keyword “systems biology” search by year) as technologies and methods have developed and grown to allow complex global measurements (Figure 1.5). The first advantage of a systems biology approach is that there is less of a reliance on what is currently known than with a reductionist approach, facilitating the identification of previously

Figure 1.5

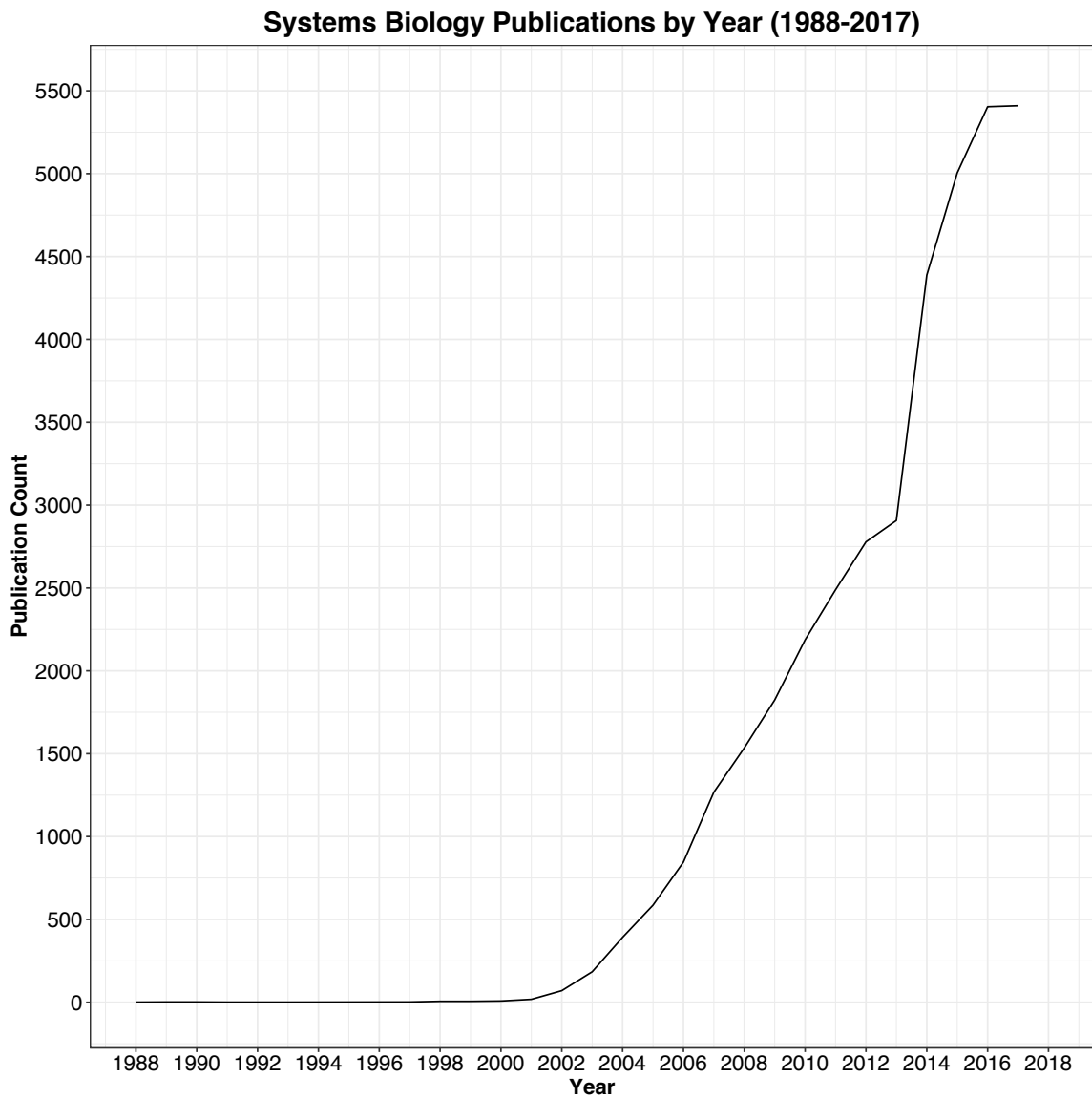


Figure 1.5: Systems Biology Publications by Year.

Systems biology publications have grown exponentially over the last 15 years. The first publication on “systems biology” was in 1988. Publication count by year with keyword “systems biology” was downloaded from PubMed on 07/03/2018. Years 1988-2017 are plotted.

unknown cellular reactions or metabolic functions. A second advantage of a systems biology approach is that when combined with omics technologies, valuable data on the flow of information in biological systems can be discerned by studying the interactions between different components of the system¹²⁴. A systems biology approach was therefore utilized in an attempt to decipher the role of NAT1 in breast cancer cell metabolism. The systems biology approach utilized in this dissertation was three pronged, integrating bioenergetics, metabolomics, and transcriptomics with an added genomics dimension (by using the newly constructed MDA-MB-231 cell lines that vary only in NAT1) to gain a global view of the role of NAT1 in metabolism and cellular energetics (Figure 1.6).

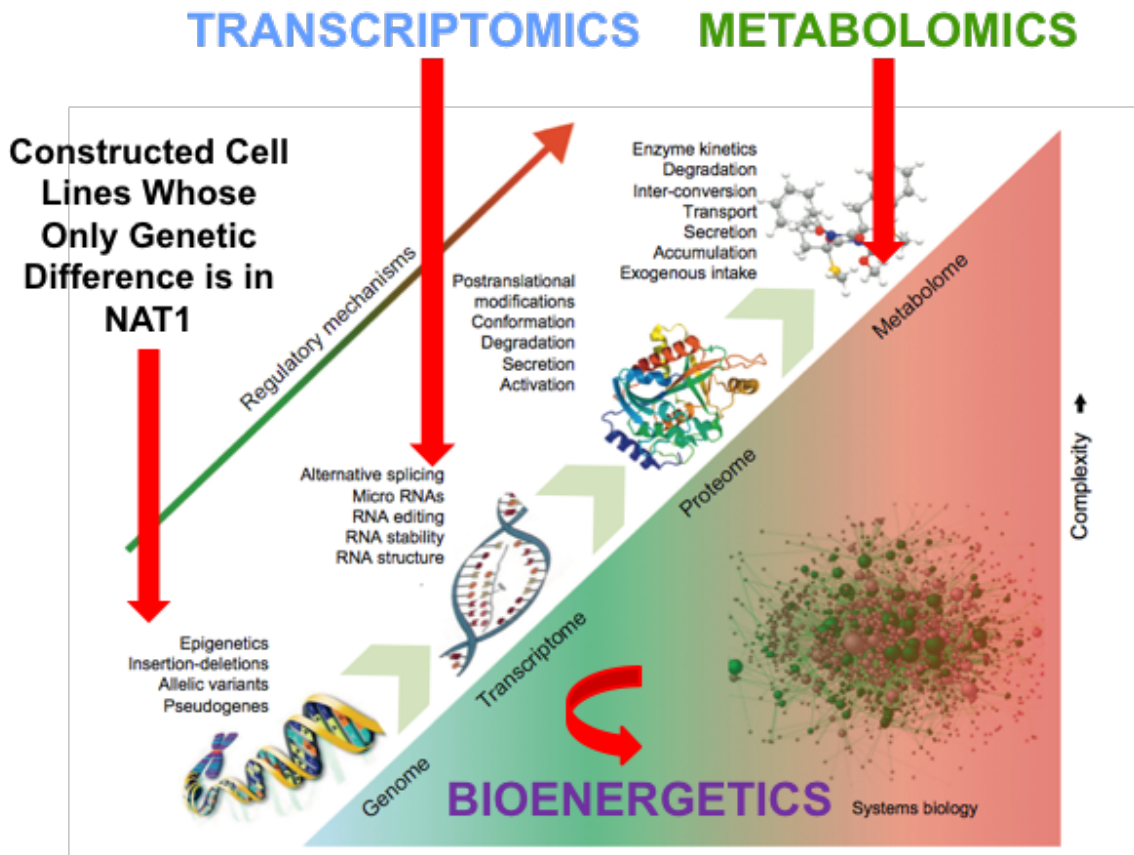
Combining Datasets

A comprehensive understanding of a specific biological system cannot be accurately predicted by that system's individual parts, even with full understanding of the parts alone¹²⁴. Combining information from bioenergetics, metabolomics, and transcriptomics experiments allows a holistic systems biology view of metabolism in samples that vary only in NAT1. Integrating all three datasets allows maximization of observations of the system at different points of regulation and varying complexities. Utilizing a systems biology approach has many advantages: 1. confidence in conclusions is increased because of the overlap in data at multiple levels of regulation and 2. combining multiple omics techniques can help overcome the limitations of a specific omics technique. One potential limitation of these studies is the complexity of metabolism; metabolism is very highly regulated but many of the regulation mechanisms are either poorly understood or undefined.

Dissertation Aims

Given human NAT1 both hydrolyzes acetyl-CoA and utilizes acetyl-CoA as a cofactor in metabolic reactions, increased and decreased levels of NAT1 are hypothesized to contribute to reprogramming of cellular metabolism by altering the levels of free acetyl-CoA. Increased NAT1 activity would lead to increased hydrolysis of acetyl-CoA and/or increased use of acetyl-CoA for acetylation, therefore less free acetyl-CoA would be available. Conversely, decreased NAT1 activity would lead to decreased hydrolysis of acetyl-CoA and/or decreased use of acetyl-CoA for

Figure 1.6



Adapted from: J. Barallobre-Barreiro et al. Rev Exp Cardiol. 2013;66(8):657–661

Figure 1.6: Systems Biology Approach.

The work presented in this dissertation utilizes a systems biology approach with four layers. The first layer consists of the six genetically modified MDA-MB-231 cell lines whose only genetic difference (theoretically) is in NAT1. The second, third, and fourth layers are bioenergetics, metabolomics, and transcriptomics measurements of the constructed cell lines and represent Specific Aim 1, Specific Aim 2, and Specific Aim 3, respectively. Progressing from the genome to the transcriptome to the proteome to the metabolome, regulatory mechanisms as well as complexity increase.

acetylation, therefore more free acetyl-CoA would be available (Figure 1.7). As discussed above, acetyl-CoA is involved in many cellular pathways, such as fatty acid synthesis and the citric acid cycle (TCA), which produce important membrane lipids and substrates for cellular energy metabolism¹²⁵. Cancer cells characteristically exhibit rapid, uncontrolled growth therefore requiring increased levels of energy and cellular components when compared to normal cells. Although, the exact mechanism by which cellular metabolism is reprogrammed remains unclear, it is likely specific to each cancer and is also influenced by a combination of factors. Varying levels of NAT1 *N*-acetylation activity in the MDA-MB-231 breast cancer cell line are predicted to lead to measurable changes in bioenergetics, metabolite abundances, and transcripts. The overall goal of this dissertation is to evaluate the effect of varying levels of NAT1 on breast cancer cell metabolism via a systems biology approach. The Specific Aims of this project are:

Aim 1: To identify bioenergetic changes induced by altering the levels of human NAT1 through mitochondrial stress test profiling using the Seahorse XF24 Bioanalyzer.

Varying levels of NAT1 are expected to lead to differences in mitochondrial energetics since acetyl-CoA is central to energy metabolism.

Aim 2: To identify metabolic changes induced by altering the levels of human NAT1 through a comprehensive untargeted metabolomics approach.

Differing levels of acetyl-CoA are predicted to lead to differences in the abundances of metabolites found in the fatty acid synthesis and amino acid degradation pathways. We additionally predict the abundance of total fatty acids and energy metabolism intermediate subclasses of metabolites will be altered when compared to samples of cells with 'parental' levels of NAT1 activity.

Aim 3: To identify differential changes in gene expression induced by altering the levels of human NAT1 through a RNA-seq analysis of the transcriptome.

The expression profile of genes involved in acetyl-CoA containing pathways are expected to be differentially expressed. Given metabolism is highly regulated we expect altered acetyl-CoA levels will have an effect on the genes that encode proteins/enzymes that utilize acetyl-CoA. The

Figure 1.7

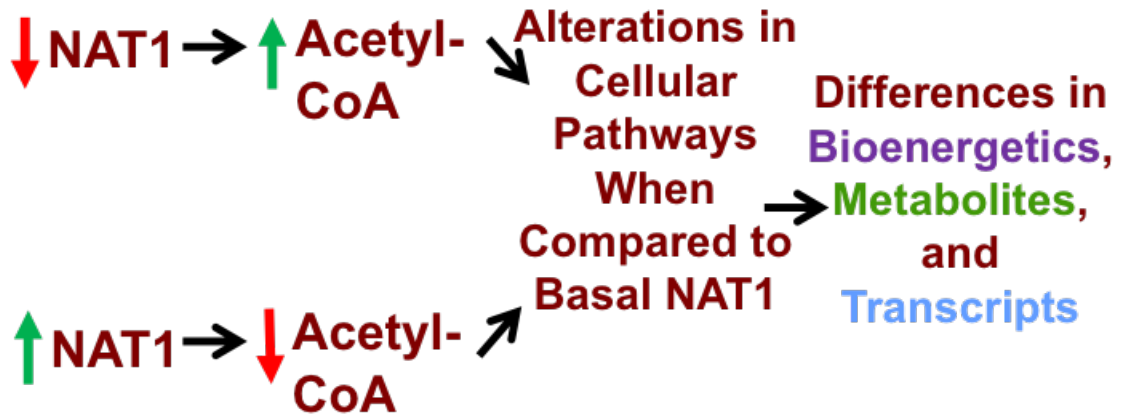


Figure 1.7: Dissertation Hypothesis.

The overall hypothesis and rationale for this dissertation is that cell lines with decreased levels or knockout NAT1 will have increased free acetyl-CoA since those cell lines have less NAT1 to hydrolyze acetyl-CoA; Conversely, the cell line with increased NAT1 will have decreased free acetyl-CoA since that cell line has more NAT1 to hydrolyze acetyl-CoA. These differences in free acetyl-CoA are hypothesized to lead to alterations in cellular pathways/metabolism when compared to the cell line with basal NAT1 that can be measured by global bioenergetics, metabolomics, and transcriptomics experiments.

data collected in this aim will build upon the data generated in Specific Aim 2 by collecting complementary transcriptomics data from the same biological samples. Differences in the expression of genes involved in fatty acid and membrane lipid synthesis are expected because cancer cells have an increased demand for these cellular building blocks.

Significance

This study is the first to investigate the effect of varying human NAT1 activity on the bioenergetics profile, metabolome, and transcriptome of a breast cancer cell line. Methods for the evaluation of the bioenergetic profile of living cells in real-time have recently been developed to identify fluxes in energy pathways and have yet to be utilized when studying the effect varying levels of NAT1 have on breast cancer cells. The omics disciplines have developed into an extremely beneficial and information-rich area of research but have been under-utilized in research on human NAT1. Analysis of metabolomics and transcriptomics data can help reveal what pathways are altered as a result of varying levels of NAT1 activity thus providing insights about NAT1's effect on cellular metabolism and into the role of NAT1 in breast cancer disease and progression. A greater understanding of the role of NAT1 in breast cancer could lead to better detection and treatment methods and impact the development of drugs that target NAT1.

CHAPTER 2

NAT EXPRESSION IN ESTABLISHED BREAST CANCER CELL LINES & BREAST TISSUES

Background

Based on limited data, it has been hypothesized that *NAT2* expression is very low in breast tissue and negligible in comparison to *NAT1* expression. However, a more comprehensive evaluation of *NAT1* and *NAT2* co-expression in breast tissues and established breast cancer cell lines is needed. Additionally, while many studies have reported an association between *NAT1* and *ESR1* expression⁸⁵⁻⁹¹, *NAT2* and *ESR1* association has not been evaluated even though *NAT1* and *NAT2* are isozymes that have overlapping substrate specificities and very similar protein structure. There are many publicly available data repositories of cancer samples that offer a wealth of gene expression data. These data repositories have been under-utilized in *NAT* research.

Materials and Methods

Acquisition of publicly available data from the Cancer Cell Line Encyclopedia (CCLE) and TCGA data repositories

RNA expression (RNA-Seq) data for *ESR1*, *NAT1* and *NAT2* in established breast cancer cell lines were accessed on 8/11/17 (n=57) from the CCLE¹²⁶; RNA expression values were reported in Reads Per Kilobase of transcript per Million mapped reads (RPKM). A total of 15 breast cancer cell lines had no detectable *NAT2* gene expression. Data from TCGA¹²⁷ for the breast invasive carcinoma (BRCA) cohort were accessed on 2/4/18 (primary breast tumor tissue, n=1,043; normal breast tissue, n=99) via FirebrowseR¹²⁸, an R client to the Broad Institute's RESTful Firehose Pipeline; RNA expression values were reported in RNA-Seq by Expectation-

Maximization (RSEM). A total of 59 of the breast tumor samples and seven of the normal tissue samples did not have gene expression data for *NAT2*.

Established breast cancer cell lines analyzed

The following breast cancer cell lines were analyzed in this study: AU565, BT-20, BT-474, BT-483, BT-549, CAL-120, CAL-148, CAL-51, CAL-85-1, CAMA-1, DU4475, EFM-19, EFM-192A, HCC1143, HCC1187, HCC1395, HCC1419, HCC1428, HCC1500, HCC1569, HCC1599, HCC1806, HCC1937, HCC1954, HCC202, HCC2157, HCC2218, HCC38, HCC70, HDQ-P1, HMC-1-8, HMEL, Hs 274.T, Hs 281.T, Hs 343.T, Hs 578.T, Hs 606.T, Hs 739.T, Hs 742.T, JIMT-1, KPL-1, MCF-7, MDA-MB-134-VI, MDA-MB-157, MDA-MB-175-VII, MDA-MB-231, MDA-MB-361, MDA-MB-415, MDA-MB-436, MDA-MB-453, MDA-MB-468, SK-BR-3, T-47D, UACC-812, UACC-893, ZR-75-1 and ZR-75-30.

Statistical analyses

Shapiro-Wilk tests were conducted to determine if the expression of the genes under study were approximately normally distributed. Significant evidence of departures from approximate normality was observed; therefore, non-parametric statistical techniques were employed for subsequent analyses. Spearman's correlation was used to evaluate the RNA expression levels between gene pairs (i.e. *ESR1* and *NAT1*, *ESR1* and *NAT2*, *NAT1* and *NAT2*). Differences in the mRNA expression levels of *NAT1* and *NAT2* in each dataset, and differences in RNA expression between primary breast tumor samples and normal breast tissue samples, were evaluated using the Wilcoxon Rank-Sum test; median values were compared to determine fold-differences.

RNA expression data for each gene were stratified by ER status (+ or -) as defined in the literature¹²⁹⁻¹³¹ for the CCLE data, or as determined by immunohistochemistry during sample collection and cataloging for TCGA data. Differences in gene expression following stratification were evaluated using Wilcoxon Rank-Sum tests for each gene; median values were compared to determine fold-differences. A total of 10 of the breast cancer cell lines had either conflicting or unknown ER status in the literature.

Wakefield *et al* published the NAT1 PABA *N*-acetylation activities of seven (ZR-75-1, T47D, MCF-7, MDA-MB-453, MDA-MB-436, MDA-MB-231 and CAL-51) of the 57 breast cancer cell lines included in the present study⁸⁸. The association between previously reported NAT1 activity and *NAT1* RNA expression data for the same cell lines in the CCLE repository was evaluated. All statistical analyses were performed in R: A Language and Environment for Statistical Computing, version 3.4.2¹³².

Results

Association between NAT1 & ESR1, NAT2 & ESR1, and NAT1 & NAT2

NAT1 RNA and *ESR1* RNA were significantly correlated ($P < 0.0001$ for all) at moderately high magnitudes in breast cancer cell lines (Spearman rho $\rho = 0.59$; Figure 2.1A and B), human primary breast tumors ($\rho = 0.59$; Figure 2.1C and D) and normal breast tissue ($\rho = 0.57$; Figure 2.1E and F). A significant ($p < 0.005$ for all) association between *ESR1* and *NAT2* expression was observed, although the magnitude of the association was low and varied across datasets. The primary breast tumor dataset exhibited the weakest association ($\rho = 0.16$; Figure 2.1C and D), whereas the normal breast tissue ($\rho = 0.38$; Figure 2.1E and F) and breast cancer cell line ($\rho = 0.39$; Figure 2.1A and B) datasets exhibited similar, albeit low, association. Strong evidence of an association ($p < 0.0001$ for all) between *NAT1* RNA and *NAT2* RNA levels was observed in all three datasets, with moderately high magnitude in the breast cancer cell lines ($\rho = 0.64$; Figure 2.1A and B). The primary breast tumor and normal breast tissue datasets exhibited interdependence similar to each other ($\rho = 0.43$ and 0.46 , respectively; Figures 2.1C-F); however, the association was lower than that observed in the breast cancer cell lines.

Comparison of NAT1 & NAT2 expression

NAT1 RNA expression in breast cancer cell lines, primary breast tumors, and normal breast tissue was significantly higher compared with *NAT2* expression by 33-, 222- and 52-fold, respectively ($p < 0.0001$ for all; Figure 2.2A-C). *NAT1* expression was higher than *NAT2* expression in all 57 breast cancer cell lines tested, with the exception of the UACC-893 cell line, which expressed the highest *NAT2* RNA of any of the breast cancer cell lines analyzed. A total of 15 of the 57 breast cancer cell lines (MDA-MB-134-IV, CAL-120, DU4475, MCF-7, JIMT-1, Hs

Figure 2.1

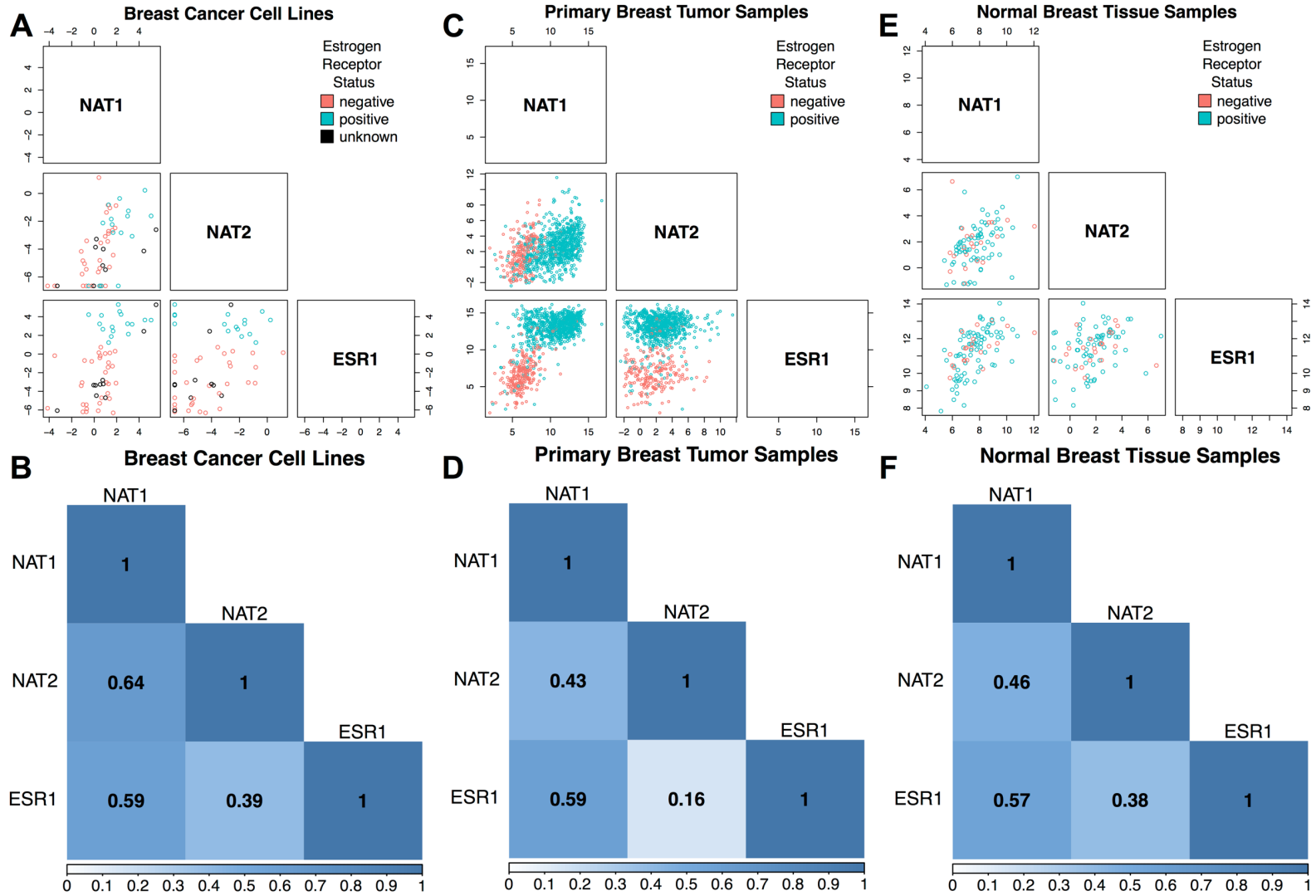
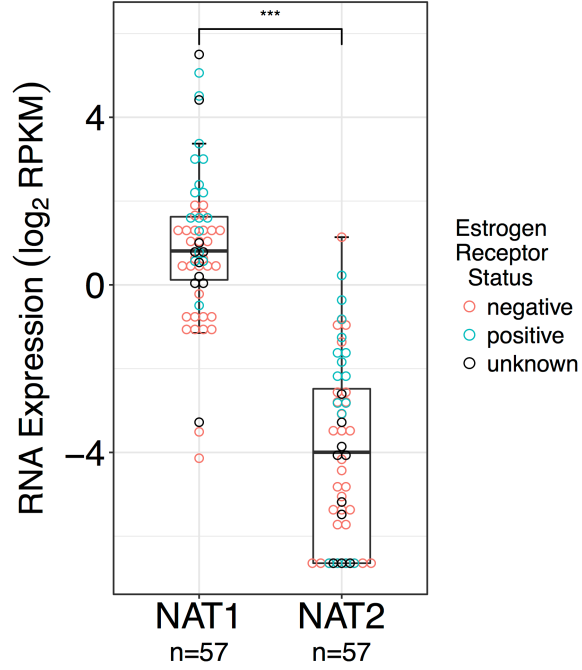


Figure 2.1: Scatterplot and correlation matrices for NAT1, NAT2, and ESR1.

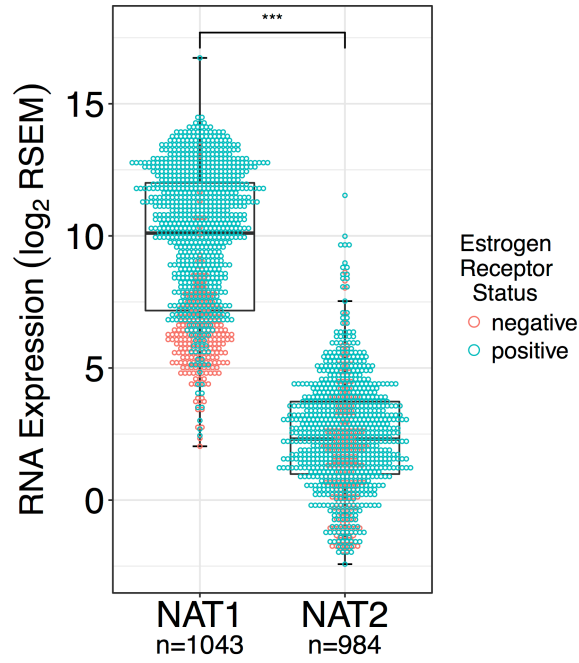
Associations between *NAT1*, *NAT2*, and *ESR1* RNA expression were analyzed in breast cancer cell lines, primary breast tumor tissue, and normal breast tissue using the Spearman method. In the scatterplot matrices, each open circle represents a single sample and is color-coded according to ER status; pink circles, ER- samples; blue circles, ER+ samples; black circles, samples with unknown ER status. In the association matrices, boxes are labeled with the Spearman correlation coefficient (ρ) for each comparison and color reflects strength of association; dark blue represents high association, light blue represents low association, and white represents no association. (A) Scatterplot matrix of the association between *NAT1*, *NAT2* and *ESR1* RNA expression in breast cancer cell lines (n=57). (B) Correlation matrix between *NAT1*, *NAT2* and *ESR1* RNA expression in breast cancer cell lines (n=57). (C) Scatterplot matrix of the association between *NAT1*, *NAT2* and *ESR1* RNA expression in primary breast tumor samples (n=1,043 for *NAT1* vs. *ESR1*, n=984 for *NAT1* vs. *NAT2* and *NAT2* vs. *ESR1*). (D) Correlation matrix between *NAT1*, *NAT2*, and *ESR1* RNA expression in primary breast tumor samples (n=1,043 for *NAT1* vs. *ESR1*, n=984 for *NAT1* vs. *NAT2* and *NAT2* vs. *ESR1*). (E) Scatterplot matrix of the association between *NAT1*, *NAT2* and *ESR1* RNA expression in normal breast tissue samples (n=99 for *NAT1* vs. *ESR1*, n=92 for *NAT1* vs. *NAT2* and *NAT2* vs. *ESR1*). (F) Correlation matrix between *NAT1*, *NAT2* and *ESR1* RNA expression in normal breast tissue samples (n=99 for *NAT1* vs. *ESR1*, n=92 for *NAT1* vs. *NAT2* and *NAT2* vs. *ESR1*). ER, estrogen receptor; *ESR1*, estrogen receptor 1; *NAT1*, arylamine *N*-acetyltransferase 1; *NAT2*, arylamine *N*-acetyltransferase 2.

Figure 2.2

A Breast Cancer Cell Lines



B Primary Breast Tumor Samples



C Normal Breast Tissue Samples

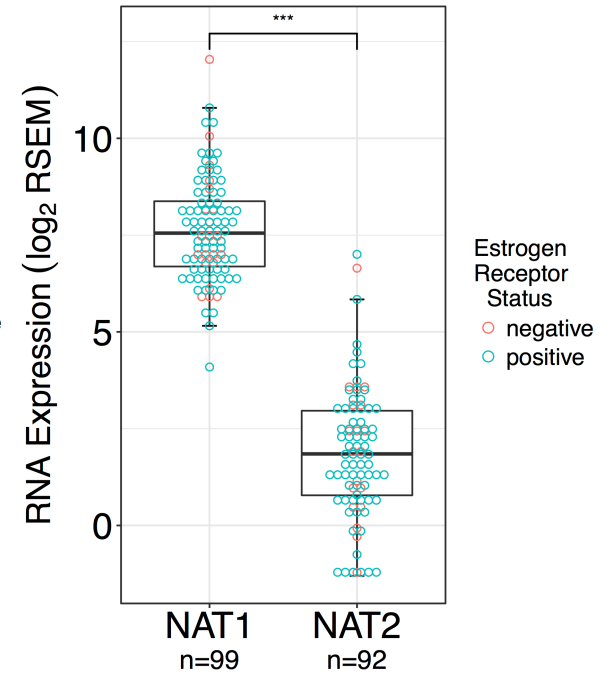


Figure 2.2: NAT1 and NAT2 RNA expression in breast cancer cell lines, primary breast tumor samples, and normal breast tissue samples.

Differences in gene expression between *NAT1* and *NAT2* in breast cancer cell lines, primary breast tumor tissue and normal breast tissue were statistically evaluated by Wilcoxon rank-sum test; *** $p < 0.001$. Each dot represents a single sample and is color-coded according to ER status; pink dots, ER- samples; blue dots, ER+ samples; black dots, samples with unknown ER status. In the boxplots, the solid black line represents the median, the upper hinge represents the 75th quartile and the lower hinge represents the 25th quartile. The upper whisker represents the largest observation less than or equal to the upper hinge + 1.5 x IQR, the lower whisker represents the smallest observation greater than or equal to the lower hinge - 1.5 x IQR. (A) *NAT1* RNA expression was significantly higher than *NAT2* RNA expression in the breast cancer cell lines. (B) *NAT1* RNA expression was significantly higher than *NAT2* RNA expression in the primary breast tumor samples. (C) *NAT1* RNA expression was significantly higher than *NAT2* RNA expression in the normal breast tissue samples. ER, estrogen receptor; IQR, interquartile range; *NAT1*, arylamine *N*-acetyltransferase 1; *NAT2*, arylamine *N*-acetyltransferase 2; RPKM, reads per kilobase of transcript per million mapped reads; RSEM, RNA-Seq by Expectation-Maximization.

281.T, KPL-1, Hs 606.T, HCC70, EFM-19, CAL-148, HCC1569, HMC-1-8, HCC1599 and HCC1395) had no reported *NAT2* RNA expression, whereas all 57 reported *NAT1* RNA expression. The KPL-1 breast cancer cell line has been reported to be contaminated/misidentified and to be an MCF-7 derivative¹³³.

In TCGA dataset, normal breast tissue samples were collected from patients in which primary breast tumor samples were also collected (but only for 99 individuals), allowing comparison of gene expression between normal breast tissue and primary tumor breast tissue within single individuals. In the primary breast tumor samples, only nine of the 984 samples had higher *NAT2* RNA expression than *NAT1*; of those nine samples, two were ER⁺ and seven were ER⁻, and only one sample had a corresponding normal breast tissue sample. Notably, in that individual's normal breast tissue sample, *NAT2* RNA expression was not higher than *NAT1* RNA expression. In the normal breast tissue samples, only one of the 92 samples had higher *NAT2* RNA expression than *NAT1*; the corresponding primary breast tumor sample from the same patient had lower *NAT2* than *NAT1*.

Comparison of gene expression between ER⁺ and ER⁻ samples

ESR1 and *NAT1* gene expression were significantly increased, 86- and 2.6-fold, respectively, in ER⁺ breast cancer cell lines ($p < 0.0001$ for both; Figure 2.3A), whereas *NAT2* gene expression did not significantly vary between ER⁺ and ER⁻ breast cancer cell lines ($p > 0.05$; Figure 2.3A). Of the breast cancer cell lines with ER status defined in the literature¹²⁹⁻¹³¹, a connection between *ESR1* RNA expression and the reported ER status was observed. In the dataset, it was observed that samples with *ESR1* RNA expression < 1.7 RPKM were defined as ER⁻ in the literature, whereas samples with *ESR1* expression > 2.3 RPKM were defined as ER⁺ in the literature. The expression levels of all three genes were significantly higher in ER⁺ primary breast tumor samples ($p < 0.0001$ for all; Figure 2.3B); however, the fold-change between *NAT2* expression in ER⁺ and ER⁻ samples was smaller (1.8-fold difference) than for *NAT1* and *ESR1*. In comparison, *ESR1* and *NAT1* were ~108- and 27-fold higher, respectively. The expression levels of genes were not significantly different between ER⁺ and ER⁻ normal breast tissue

Figure 2.3

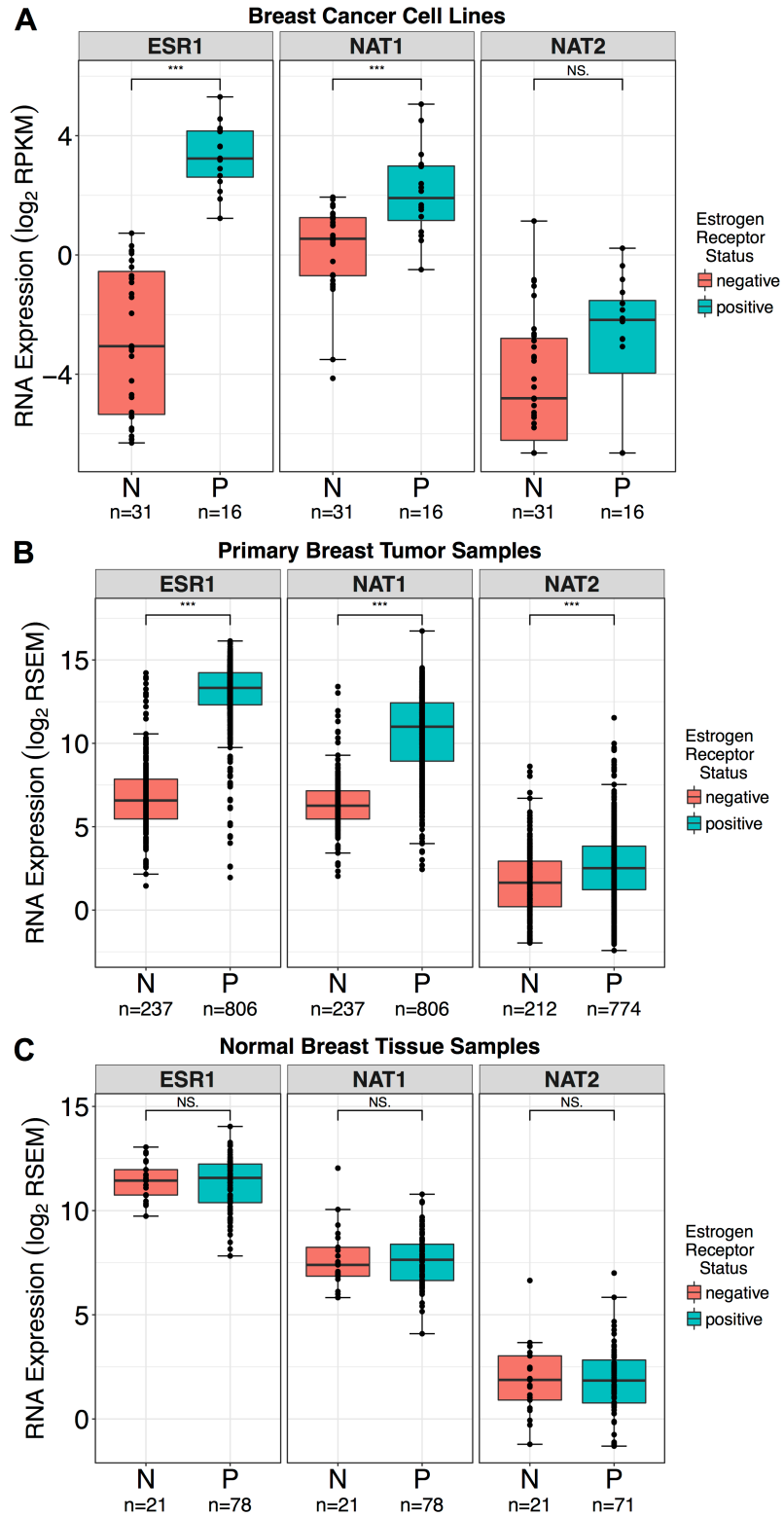


Figure 2.3: *ESR1*, *NAT1* and *NAT2* RNA expression in breast cancer cell lines, primary breast tumor samples, and normal breast tissue stratified by ER status.

Differences in the expression levels of *ESR1*, *NAT1* and *NAT2* genes in breast cancer cell lines, primary breast tumor tissue, and normal breast tissue stratified by ER status were evaluated by Wilcoxon rank-sum test; *** $p < 0.001$; NS, not significant. Boxplots are color-coded according to ER status; pink boxplots, ER⁻ samples; blue boxplots, ER⁺ samples. In the boxplots, the solid black line represents the median, the upper hinge represents the 75th quartile and the lower hinge represents the 25th quartile. The upper whisker represents the largest observation less than or equal to the upper hinge + 1.5 x IQR, the lower whisker represents the smallest observation greater than or equal to the lower hinge - 1.5 x IQR. (A) *ESR1* and *NAT1* RNA expression were significantly higher in ER⁺ breast cancer cell lines compared with ER⁻ breast cancer cell lines. *NAT2* RNA expression was not significantly different in ER⁺ breast cancer cell lines compared with ER⁻ breast cancer cell lines. A total of 10 cell lines had either conflicting reports or no available data for ER status in the literature and were excluded from the analysis. (B) *ESR1*, *NAT1* and *NAT2* RNA expression were significantly higher in ER⁺ samples compared with ER⁻ samples in the primary breast tumor dataset. (C) *ESR1*, *NAT1* and *NAT2* RNA expression levels were not significantly different in ER⁺ samples compared with ER⁻ samples in the normal breast tissue dataset. ER, estrogen receptor; IQR, interquartile range; *ESR1*, estrogen receptor 1; *NAT1*, arylamine *N*-acetyltransferase 1; *NAT2*, arylamine *N*-acetyltransferase 2; RPKM, reads per kilobase of transcript per million mapped reads; RSEM, RNA-Seq by Expectation-Maximization.

samples ($p>0.05$ for all; Figure 2.3C). Most of the breast cancer cell lines were ER-, whereas most of the primary breast tumor and normal breast tissue samples were ER+.

Comparison of NAT1, NAT2, and ESR1 gene expression between normal breast tissue and primary breast tumors

Differences in gene expression between normal breast tissue and primary breast tumor tissue were evaluated for each gene, *ESR1*, *NAT1*, and *NAT2*. More spread was observed in the primary breast tumor samples compared with the normal breast tissue samples for each gene. *ESR1* and *NAT1* gene expression were significantly elevated 2.5- and 5.9-fold, respectively, in primary breast tumor samples compared with normal breast tissue samples ($p<0.0001$ for both; Figure 2.4). *NAT2* expression was also significantly higher in primary breast tumor samples compared with normal breast tissue samples, but at a lower significance and fold-change (1.4-fold) than *ESR1* and *NAT1* ($p<0.05$; Figure 2.4)

Relationship between previously reported NAT1 N-acetylation activity and NAT1 RNA expression.

NAT1 N-acetylation activity previously reported in the literature and *NAT1* RNA expression in seven of the 57 breast cancer cell lines were significantly associated ($p<0.05$) with a high magnitude ($\rho=0.89$; Figure 2.5).

Co-expression of NAT1 and NAT2 RNA expression in established breast cancer cell lines

Co-expression profiles of *NAT1* and *NAT2* RNA for each established breast cancer cell line included in this study are presented in Figure 2.6. Of all the cell lines included in the present study, the UACC-893 cell line expressed the highest level of *NAT2* RNA, whereas the HCC1500 cell line expressed the highest level of *NAT1* RNA. The ZR-75-1 cell line expressed high levels of both *NAT1* and *NAT2* RNA, whereas the HCC1395 cell line expressed low levels of both.

Discussion

The present study analyzed established breast cancer cell lines and samples from patients with breast cancer to evaluate the extent to which breast cancer cell lines serve as appropriate models for *NAT1*, *NAT2* and *ESR1* expression in breast tumors. Overall, the present findings demonstrated a strong association between *NAT1* and *ESR1* expression, which is in agreement with previous reports that *NAT1* and *ESR1* are positively associated⁸⁵⁻⁹¹, and this

Figure 2.4

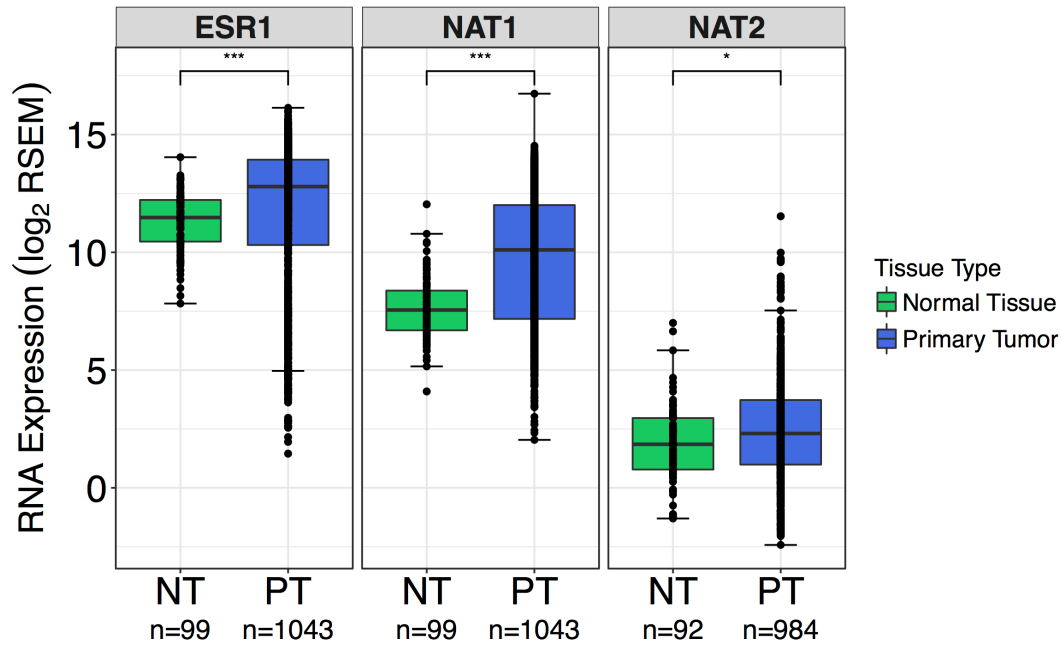


Figure 2.4: Comparison of *ESR1*, *NAT1* and *NAT2* RNA expression in normal breast tissue and primary breast tumor samples.

Differences in gene expression of *ESR1*, *NAT1* and *NAT2* in normal breast tissue and primary breast tumor tissue were evaluated by Wilcoxon rank-sum test; *** $p < 0.001$; * $p < 0.05$. Boxplots are color-coded according to tissue type; green boxplots, normal breast tissue samples; blue boxplots, primary breast tumor samples. In the boxplots, the solid black line represents the median, the upper hinge represents the 75th quartile and the lower hinge represents the 25th quartile. The upper whisker represents the largest observation less than or equal to the upper hinge + 1.5 x IQR, the lower whisker represents the smallest observation greater than or equal to the lower hinge - 1.5 * IQR. For all genes, more spread was observed in data from the primary breast tumor samples compared with the normal breast tissue samples. *ESR1* and *NAT1* gene expression were significantly elevated in primary tumor tissue compared with normal breast tissue. *NAT2* expression was also significantly higher in primary tumor tissue compared with normal breast tissue, but at a lower significance than *ESR1* and *NAT1*. IQR, interquartile range; *ESR1*, estrogen receptor 1; *NAT1*, arylamine *N*-acetyltransferase 1; *NAT2*, arylamine *N*-acetyltransferase 2; RSEM, RNA-Seq by Expectation-Maximization.

Figure 2.5

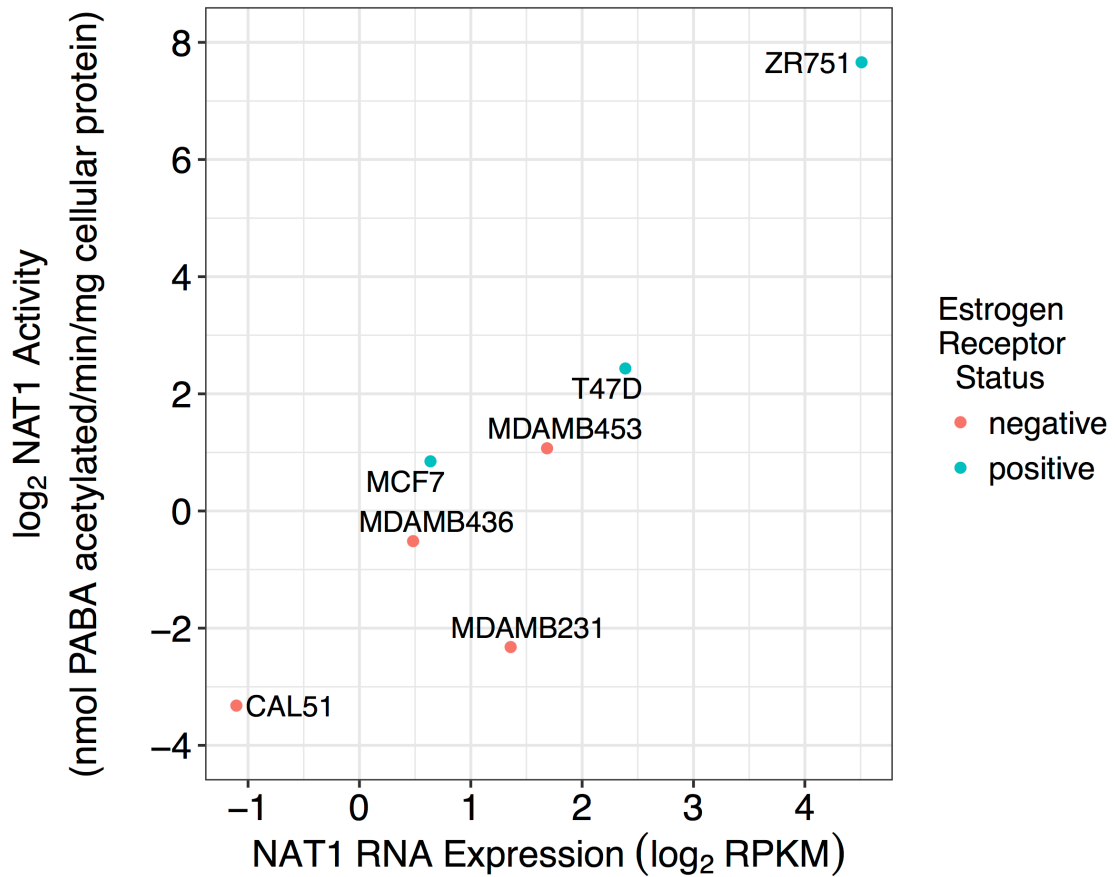


Figure 2.5 Association between NAT1 RNA expression and previously reported NAT1 N-acetylation activity in seven established breast cancer cell lines.

NAT1 RNA expression from Cancer Cell Line Encyclopedia and previously reported NAT1 N-acetylation activity⁸⁸ in seven breast cancer cell lines were significantly associated ($p < 0.05$; $\rho = 0.89$). Dots represent a single cell line and are color-coded according to ER status: Pink dots, ER- samples; blue dots, ER+ samples. ER, estrogen receptor; NAT1, arylamine N-acetyltransferase 1; NAT2, arylamine N-acetyltransferase 2; PABA, *p*-aminobenzoic acid; RPKM, reads per kilobase of transcript per million mapped reads.

Figure 2.6

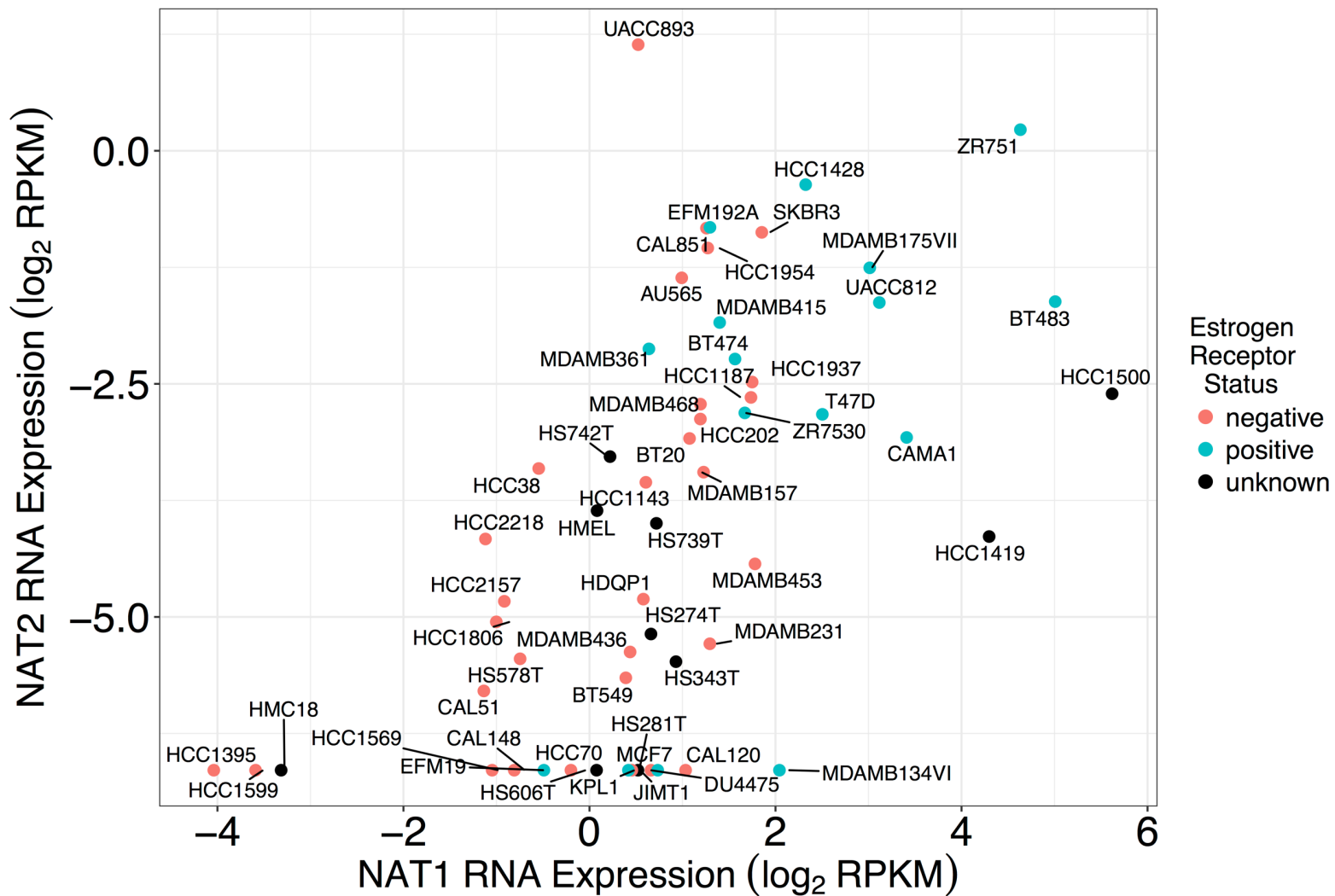


Figure 2.6: NAT1 and NAT2 association in breast cancer cell lines.

Association between *NAT1* and *NAT2* RNA expression was analyzed in breast cancer cell lines (each labeled in this figure). Each dot represents a single breast cancer cell line and is color-coded according to ER status; pink dots, ER- samples; blue dots, ER+ samples; black dots, samples with unknown or conflicting ER status in the literature. *NAT1* and *NAT2* RNA expression was significantly associated in breast cancer cell lines ($p < 0.0001$, $\rho = 0.64$).

association was observed in all three sample types at approximately the same magnitude. These findings suggested that breast cancer cell lines may accurately reflect this relationship and provide a useful model for further research into the relationship. It is well known that *ESR1* expression is frequently altered in breast cancer; therefore, the decrease in association between *NAT2* and *ESR1* in primary breast tumors compared with in normal breast tissue samples and established breast cancer cell lines may be due to more dysregulation of *ESR1* than *NAT2* in primary breast tumors. The results of an analysis between *NAT2* and *ESR1* expression suggested that, while *NAT2* and *ESR1* are associated, the magnitude is low.

Interdependence between *NAT1* and *NAT2* expression was moderately high in the breast cancer cell line dataset, but substantially lower in the primary breast tumor and normal breast tissue datasets. Additionally, the strength of the association between *NAT1* and *NAT2* in the breast cancer cell line dataset was similar to the strength of the association observed between *NAT1* and *ESR1* in that dataset; however, in the primary breast tumor and normal breast tissue datasets, the association between *NAT1* and *NAT2* was lower. These findings suggested that breast cancer cell lines may over-represent the interdependence between *NAT1* and *NAT2*, and not fully replicate the relationship observed in primary breast tumors or normal breast tissue. In the breast cancer cell line data there appears to be a cut-off (between 1.7 and 2.3 RPKM) linking *ESR1* RNA expression and the reported ER status of the breast cancer cell lines. This may provide a method to predict the ER status of breast cancer cell lines that currently have conflicting or unknown ER status in the literature. Using that method, it may be predicted that the HCC1500 and HCC1419 cell lines are ER+, whereas the HMC-1-8, Hs 742.T, Hs 343.T, Hs 739.T, HMEL, Hs 274.T, Hs 281.T and Hs 606.T cell lines are ER-. Notably, although 67-82% of breast cancers are ER+¹³⁴ and most of the primary breast tumor samples were ER+, the majority of established breast cancer cell lines are ER-.

NAT1 and *NAT2* RNA expression were reported in almost all samples included in the present study, which concurs with published results that have detected *NAT1* and *NAT2* mRNA by RT-PCR in human mammary tissue in smaller cohorts^{13,14,96}. *NAT1* RNA expression was significantly higher than *NAT2* RNA expression in the breast cancer cell lines, primary breast

tumor samples and normal breast tissue. In addition, with only a few exceptions, *NAT1* RNA expression was always higher than *NAT2* RNA expression in matched samples from the breast cancer cell line, primary breast tumor sample and normal breast tissue sample datasets, thus supporting previous findings that indicated *NAT1* transcripts were 2- to 3-fold higher than *NAT2* transcripts in human mammary tissues¹³⁵. The UACC-893 cell line, the only breast cancer cell line observed in this study to express higher *NAT2* RNA than *NAT1* RNA, is an ER- and progesterone receptor-negative cell line that has a ~20-fold amplification of the human epidermal growth factor receptor 2/neu oncogene sequence. Further study of this cell line may aid in the identification of additional regulatory mechanisms of *NAT1* and/or *NAT2*, since it expresses a unique profile of *NAT1* and *NAT2* compared with the other breast cancer cell lines.

While *NAT1* expression was reported in all 57 breast cancer cell lines in the present study, 15 of those breast cancer cell lines had no reported *NAT2* RNA expression (Figure 2.6). The cell lines with no detected *NAT2* RNA are plotted at $\sim -6.6 \log_2$ RPKM *NAT2*. One of those 15 cell lines, MCF-7, has been reported to express *NAT2* RNA^{96,136} albeit at very low levels. One reason for the difference in observation between this study and the previous studies may be that the detection threshold for *NAT2* was higher when measured by RNA-Seq for the CCLE dataset than in the previous studies. Additionally, in the previous studies that detected *NAT2* RNA in the MCF-7 breast cancer cell line, *NAT1* RNA was not measured at the same time; therefore, direct comparisons of the NAT isozymes was not possible. To the best of our knowledge, *NAT2* RNA expression has not been investigated in any of the other 56 breast cancer cell lines until this study. The results of this study indicated that *NAT2* may be expressed in breast tissues and expression should be considered when studying *NAT1*, due to their overlapping substrate specificities and the high degree of structural similarity.

In normal breast tissue samples no significant difference in gene expression for *ESR1*, *NAT1* and *NAT2* was observed when data were stratified by ER status. However, in the primary breast tumor samples and in the breast cancer cell lines, *ESR1* and *NAT1* exhibited increased expression in the ER+ samples compared with in the ER- samples. *NAT2* RNA expression did not significantly vary in breast cancer cell lines when comparing ER+ and ER- samples, but was

significantly increased in ER+ primary breast tumor samples compared with ER- primary breast tumor samples, although the difference was small. This finding suggested that the dysregulation of *NAT1* and *ESR1* during tumorigenesis may share similar mechanisms; however, *NAT2* does not.

ESR1, *NAT1* and *NAT2* RNA expression were each increased in primary breast tumor samples compared with normal breast tissue samples although the significance and fold-change of *NAT2* were smaller than that of *ESR1* and *NAT1*. Additionally, for all genes, more widely spread expression was observed in the primary breast tumor samples compared with normal breast tissue. These data suggested that expression of all three genes may become modified during breast cancer tumorigenesis; however, the expression of *NAT1* and *ESR1* appear to be dysregulated to a greater extent. As recently reviewed⁹⁵, the role of *NAT2* in breast cancer etiology is considered to be due to its effects on carcinogen metabolism. The present study suggested that the role of *NAT2* in breast cancer is less likely a product of cell transformation, as the expression levels of *NAT2* between normal and tumor tissues exhibited smaller variance than the expression levels of *NAT1* and *ESR1*.

NAT1 *N*-acetylation activity has been reported in normal breast tissue and breast tumor tissue^{13,14,96-99}, whereas *NAT2* *N*-acetylation activity has not been observed as consistently; when *NAT2* activity is observed, the activity is much lower than that of *NAT1* activity^{13,97,98}. Wakefield *et al* profiled *NAT1* expression and activity in seven breast cancer cell lines (MCF-7, T47D, ZR-75-1, Cal51, MDA-MB-231, MDA-MB-437 and MDA-MB-453); *NAT1* mRNA and activity was observed in all seven cell lines⁸⁸; however, *NAT2* mRNA and activity were not co-investigated. The high degree of association between the previously reported *NAT1* *N*-acetylation activity and the *NAT1* RNA expression of the same seven breast cancer cell lines suggested that *NAT1* RNA expression is highly reflective of *NAT1* *N*-acetylation activity. mRNA expression is not always predictive of enzyme activity, due to the numerous regulatory mechanisms that can occur between RNA expression and protein function; however, these results suggested that RNA expression of *NAT1* may serve as an appropriate predictor of *NAT1* *N*-acetylation activity. Further studies with an increased number of breast cancer cell lines in which *NAT1* *N*-acetylation activity

has been measured are required to confirm this hypothesis. Additionally, further studies are required to determine the association between NAT2 RNA expression and NAT2 N-acetylation activity.

Summary and Conclusions

The CCLE and TCGA repositories offer a wealth of publicly available data. The present study utilized this data to analyze and annotate the previously undefined relationships between *NAT1*, *NAT2* and *ESR1* in breast cancer cell lines, primary breast tumors and normal breast tissue. The results demonstrated that *NAT1* and *NAT2* RNA were expressed in normal breast tissue and primary breast tumor tissue; however, *NAT1* RNA expression was much higher than *NAT2*. The expression of *NAT1* and *NAT2* were found to be associated; however, the magnitude was lower than that observed between *NAT1* and *ESR1* in the primary breast tumors and normal breast tissue. Additionally, although the association between *NAT1* and *NAT2* was slightly exaggerated in the breast cancer cell line dataset, the cell lines generally reflected the *NAT1* and *NAT2* expression profiles of the primary breast tumors investigated. The present study demonstrated that while *NAT1* and *ESR1* expression were moderately associated in all datasets included in this study, *NAT2* and *ESR1* expression were associated at a lower magnitude, particularly in the primary breast tumor samples.

NAT1 and *ESR1* expression were increased in primary breast tumor samples compared with normal breast tissue samples and were increased in ER+ primary breast tumors compared with ER- primary breast tumors. *NAT2* expression was slightly increased in primary breast tumor samples compared with normal breast tissue samples and in ER+ primary breast tumors compared with ER- primary breast tumors. Although *NAT1* and *NAT2* are both implicated in breast cancer, the majority of previous breast cancer studies have investigated each isozyme individually. The present study suggested that both isozymes should be considered in each study, since both are expressed in breast tissues. Defining the association between *NAT1*, *NAT2* and *ESR1* is of great importance, as modification of *NAT1* is currently being studied for breast cancer prevention^{81,82,137,138}.

CHAPTER 3

BIOENERGETICS

Background

Defects in mitochondrial metabolism have been linked to cancer (reviewed in ¹³⁹) and are increasingly recognized as contributors, not only to the deregulation of cellular energetics, but also as contributors to tumorigenesis. Bioenergetics evaluation provides a method to investigate mitochondrial function in living cells¹⁴⁰. By utilizing real-time profiling of the oxygen consumption rate (OCR) and extracellular acidification rate (ECAR) as well as sequential injection of compounds that inhibit specific portions of the electron transport chain, mitochondrial evaluation can be conducted¹⁴¹. Oxygen consumption rate is a proxy measurement for cellular respiration while extracellular acidification rate is a proxy measurement for glycolysis allowing us to probe mitochondrial energetics. One can measure and quantify multiple mitochondrial parameters such as basal OCR, ATP-linked OCR, proton leak, maximum mitochondrial capacity, and reserve capacity¹⁴². Although it is known that NAT1 can catalyze the hydrolysis of acetyl-CoA using folate as a cofactor, the implications of this NAT1-catalyzed reaction in cellular energetics remains to be investigated. Measuring bioenergetics in the MDA-MB-231 cell lines we have constructed to vary only in NAT1 activity allows evaluation of how varying levels of NAT1 impact mitochondrial function.

Materials and methods

The Seahorse XF24 Analyzer (Agilent Technologies, Santa Clara, CA) was utilized to interrogate differences in mitochondrial cellular metabolism via oxygen consumption rate (OCR) and extracellular acidification rate (ECAR) measurements of live cells. All cell lines were cultured

in high-glucose Dulbecco's Modified Eagle Medium (DMEM), with 10% fetal bovine serum, 5% glutamine, and 5% penicillin/streptomycin added.

Twenty-four hours prior to each bioenergetics experiment, 100 μ L of cell suspension from each cell line was plated in quadruplicate in a 24-well Seahorse XF24 cell culture microplate at a density of 40,000 cells per well, thus giving four biological replicates for each cell line.

Additionally, four wells contained medium only for background correction purposes. To minimize the edge effects of plating and to ensure a monolayer of cells was formed on the bottom of the well, microplates were left in the cell culture hood at room temperature for one hour before being placed in a 37°C, 5% CO₂ incubator. After cells had adhered to the plate (three hours after placing in the incubator), 150 μ L of medium was added to each well and cells were allowed to grow overnight. One hour prior to the experiment, medium was aspirated from each well and replaced with 675 μ L Seahorse running medium. Seahorse running medium consisted of 8.3 grams/liter of Dulbecco's Modified Eagle's Medium (DMEM) base without glucose, L-glutamine, phenol red, sodium pyruvate and sodium bicarbonate (Sigma, St. Louis, MO), and 1.85 g NaCl per liter, glucose added to a final concentration of 25 mM, sodium pyruvate added to a final concentration of 1 mM, and 10 mL/liter of 100 x Glutamax-1 added, all at pH 7.4. Microplates were then incubated for one hour in a non-CO₂ 37 °C incubator. The XF24 sensor cartridge was hydrated overnight in XF calibrant solution in a non- CO₂ 37 °C incubator.

Before each experiment, solutions of compounds to be loaded in the ports of the sensor cartridge were freshly made from stock solutions. Selected compounds inhibit specific portions of the mitochondrial electron transport chain to allow the elucidation of several different measurements of mitochondrial function¹⁴¹. Seventy-five microliters of 15 μ M Oligomycin (ATP synthase inhibitor) was loaded into port A of the sensor cartridge, eighty-three microliters of 5 μ M carbonyl cyanide-*p*-trifluoromethoxyphenylhydrazone (FCCP; mitochondrial inner membrane uncoupler) was loaded into port B of the sensor cartridge, and ninety-two microliters of 10 μ M Antimycin A (complex III inhibitor) and 2 μ M Rotenone (complex I inhibitor) was loaded into port C of the sensor cartridge. For background correction wells seahorse running media was loaded into each port in place of the compound solutions. The sensor cartridge was then loaded into the

Seahorse XF24 analyzer and calibrated. Once calibration of the sensor cartridge was complete, the microplate was loaded into the machine and the experiment was started. Timed sequential injections of Oligomycin, FCCP, and Antimycin A/Rotenone occurred at 35, 50, and 64 minutes, respectively.

Six independent experiments were conducted (Figure 3.1). Each experiment consisted of four biological replicates for each of the cell lines. Baseline OCR, Baseline ECAR, Baseline OCR/ECAR, ATP-Linked OCR, Reserve Capacity, Coupling Efficiency, Proton Leak, Glycolytic Reserve, Maximum Mitochondrial Capacity, and Non-Mitochondrial Respiration were calculated as described in Table 3.1 following each experiment^{141,143}. Results from each independent experiment were summarized as mean \pm SEM from four biological replicates on the microplate. Results from all six independent experiments were then compiled with final results represented as mean \pm SEM. One-way ANOVA was conducted for each measurement to test for overall differences, followed by multiplicity adjusted post hoc tests on only those measurements that were found to be significant.

MTT Cell Growth Assays

MTT (3-(4,5-Dimethylthiazol-2-yl)-2,5-Diphenyltetrazolium Bromide) assays were conducted in triplicate to determine if there were differences in cell viability between the cell lines given that the cells were allowed to grow in the plate for 24 hours prior to bioenergetics measurements. Cells were plated in the Seahorse XF24 plates following the same procedure that was used when plating cells for bioenergetics experiments. Instead of measuring the bioenergetics of the cells 24 hours after plating, media was removed from each well and cells were incubated with 250 μ L of 5 mg/mL MTT dissolved in growth medium in a 37°C, 5% CO₂ incubator. After one hour, the media was aspirated and 250 μ L of dimethyl sulfoxide (DMSO) was added to each well to solubilize the MTT. The microplate was gently rocked for 10 minutes and then the resulting solution in each well was transferred to a 96-well plate. Absorbance was measured at 570 nm using the Gen5 microplate reader (BioTek Instruments Inc., Winooski, VT). All absorbance values were within the linear range of MTT. Absorbance measurements were median-scaled by plate to minimize between-day variance. Scaled absorbance values were

Figure 3.1

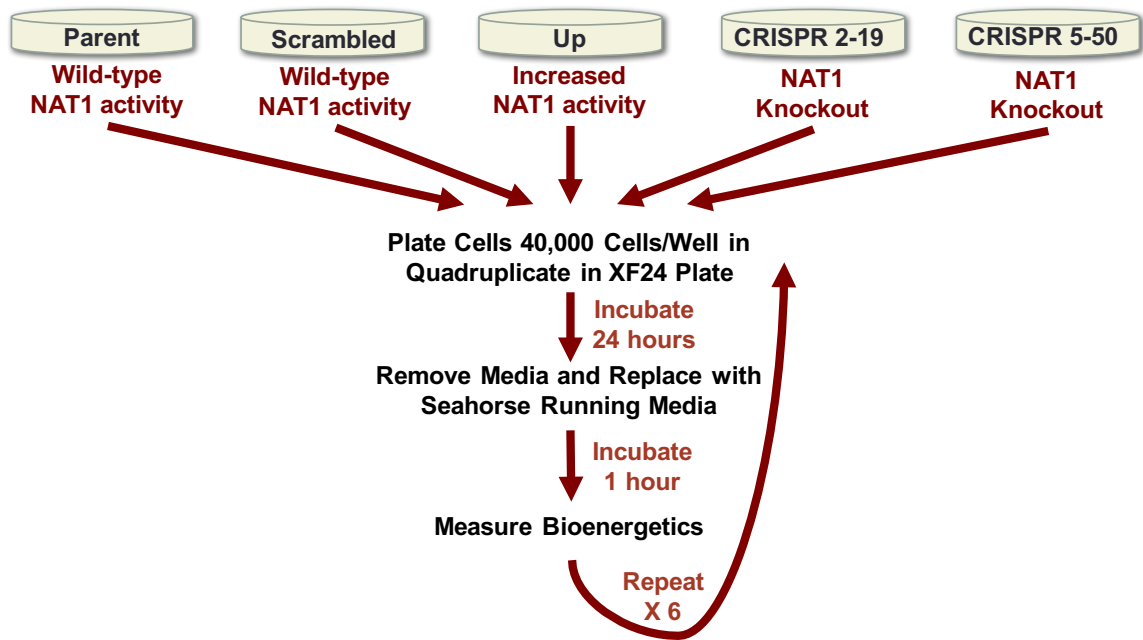


Figure 3.1: Bioenergetics Experimental Approach.

Mitochondrial bioenergetics of five genetically modified MDA-MB-231 cell lines stably transformed to differ only in human arylamine *N*-acetyltransferase 1 (NAT1) were measured. The *Parent* and *Scrambled* cell lines express parental NAT1 activity, the *Up* cell line expresses increased NAT1 activity, and the *CRISPR 2-19* and *CRISPR 5-50* cell lines express no detectable NAT1 activity (knockout). Bioenergetics were measured in six independent experiments with four biological replicates of each cell line in each experiment.

Table 3.1Definition of Measurement Calculations and ANOVA *p*-value Summary Statistics

Measurement	Calculation	ANOVA <i>p</i>-value
Baseline OCR*	OCR after equilibration but before the injection of any compounds (mean of two measurements)	0.055
ATP-Linked OCR	(Minimum OCR after Oligomycin injection) – (Baseline OCR)	0.052
Reserve Capacity	(Maximum Mitochondria Capacity) – (Baseline OCR).	<0.0001
Proton Leak	(Minimum OCR after Oligomycin injection) – (Non-Mitochondrial Respiration)	0.021
Non-Mitochondrial Respiration	Minimum OCR after Rotenone/Antimycin A injection	0.229
Maximum Mitochondrial Capacity	(Maximum OCR measurement after FCCP† injection) – (Non-Mitochondrial Respiration)	<0.0001
Coupling Efficiency	ATP-linked OCR/baseline OCR	0.368
Baseline ECAR‡	ECAR after equilibration but before the injection of any compounds (mean of two measurements)	<0.0001
Glycolytic Reserve Capacity	(Minimum ECAR after Oligomycin injection) – (Baseline ECAR)	<0.0001
Baseline OCR/ECAR	(Baseline OCR)/(Baseline ECAR)	0.011

*OCR = oxygen consumption rate

†FCCP = carbonyl cyanide-*p*-trifluoromethoxyphenylhydrazone

‡ECAR= extracellular acidification rate

tested for equivalence between cell lines. Equivalence was assessed using a two-one sided t-test (TOST) procedure that evaluated equivalence given a 20% margin.

Results

Although NAT1 is overexpressed in breast tumors^{91,144-147} and catalyzes the hydrolysis of acetyl-CoA, the impact of NAT1 inactivation or overexpression on cellular bioenergetics has yet to be reported. We examined glucose oxidation and mitochondrial bioenergetics in MDA-MB-231 triple negative breast cancer (TNBC) cells constructed to knockout or overexpress NAT1 and compared OCR and ECAR in these cells to the parental MDA-MB-231 cells by extracellular flux analysis. A summary of ANOVA *p*-values for each measurement are presented in Table 3.1. We did not observe an effect of modulation of NAT1 activity in MDA-MB-231 cells on basal OCR, ATP-linked OCR, or non-mitochondrial respiration ($p > 0.05$ for all; Table 3.1).

Knockout of NAT1 activity increased reserve capacity and maximum mitochondrial capacity when compared to the cell lines with parental (*Parent*, *Scrambled*) and increased (*Up*) NAT1 activity ($p < 0.05$ for all; Figure 3.2). In the *Parent*, *Scrambled*, and *Up* cell lines the maximal respiration was lower than the basal OCR measurements resulting in a negative value for the reserve capacity calculation; since reserve capacity cannot be negative biologically, we termed the reserve capacity measurements in these groups as 0. Reserve capacity was increased 91- and 50-fold in the *CRISPR 2-19* and *CRISPR 5-50* cell lines, respectively. The 1.8-fold increase in reserve capacity of the *CRISPR 2-19* cell line compared to the *CRISPR 5-50* cell line was also statistically significant. Maximum mitochondrial capacity of the *CRISPR 2-19* cell line was significantly increased 3.2-fold, 6.0-fold, and 5.4-fold, with respect to the *Parent*, *Scrambled* and *Up* cell lines. Maximum mitochondrial capacity of the *CRISPR 5-50* cell line was also significantly increased 2.5-fold, 4.7-fold, and 4.2-fold, with respect to the *Parent*, *Scrambled* and *Up* cell lines.

Proton leak was increased 1.8-fold in one of the NAT1 knockout (*CRISPR 2-19*) cell lines but only when compared to the cell line with increased (*Up*) NAT1 activity (Figure 3.2). We cannot conclude that this effect is due to NAT1 knockout since we did not observe the same result in the other NAT1 knockout cell line (*CRISPR 5-50*).

Figure 3.2

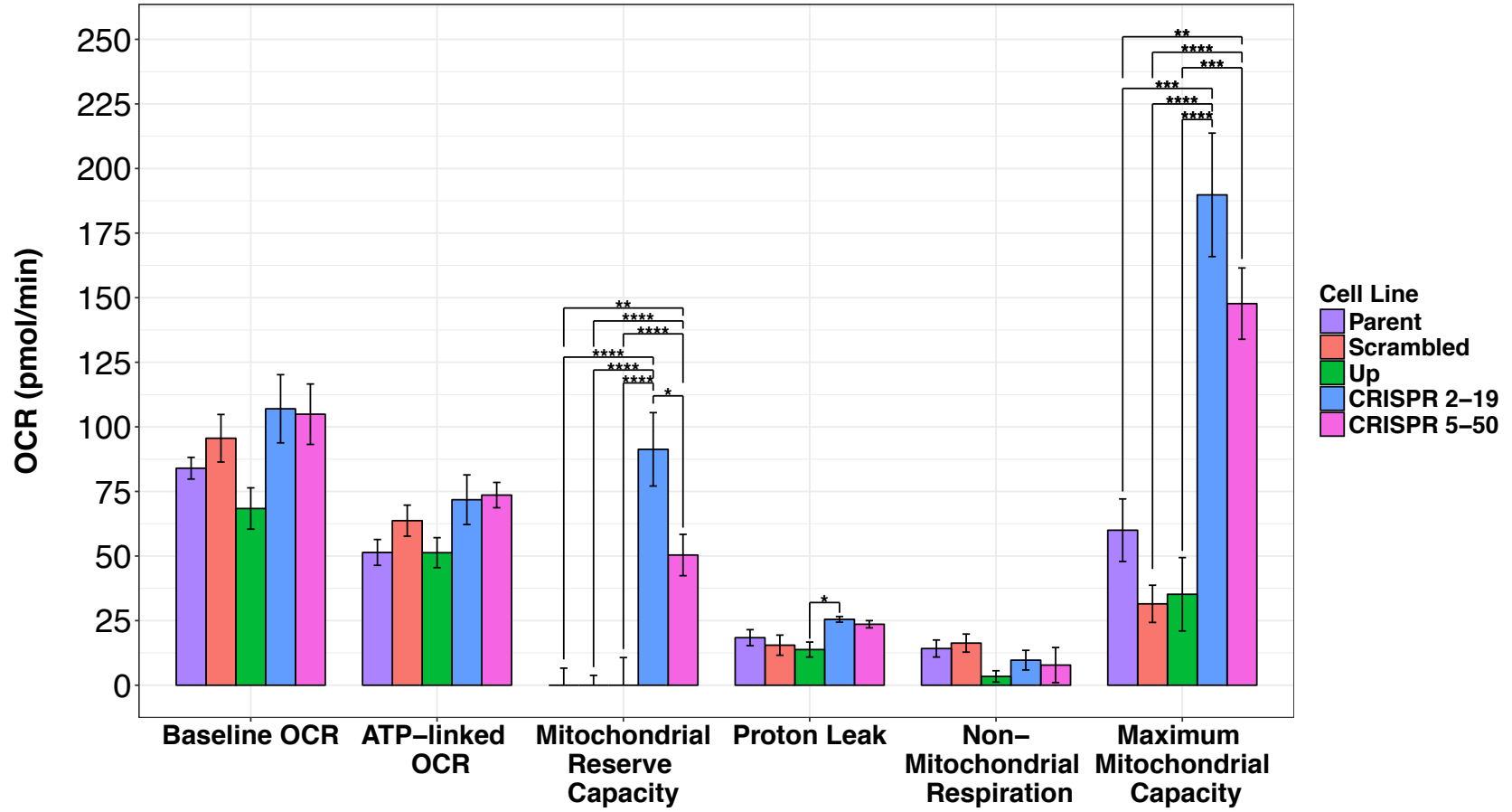


Figure 3.2: OCR measurements in each cell line.

Evidence of a difference in baseline oxygen consumption rate (OCR), ATP-linked OCR, and non-mitochondrial respiration across cell lines was not observed. Reserve capacity was significantly increased in the *CRISPR 2-19* and *CRISPR 5-50* cell lines when compared to the *Parent*, *Scrambled*, and *Up* cell lines. Reported reserve capacity measurements for *Parent*, *Scrambled*, and *Up* cell lines were truncated at 0 since reserve capacity cannot be negative. Proton leak was significantly increased in the *CRISPR 2-19* cell line but not the *CRISPR 5-50* cell line when compared to the *Up* cell line. Maximum mitochondrial capacity was significantly increased in the *CRISPR 2-19* and *CRISPR 5-50* cell lines when compared to the *Parent*, *Scrambled*, and *Up* cell lines. Order of bars are preserved throughout figure (*Parent*, *Scrambled*, *Up*, *CRISPR 2-19*, *CRISPR 5-50*) and represent mean \pm SEM. N=6. *= $p < 0.05$, **= $p < 0.01$, ***= $p < 0.001$, ****= $p < 0.0001$. The *Parent* and *Scrambled* cell lines express parental NAT1 activity, the *Up* cell line expresses increased NAT1 activity, and the *CRISPR 2-19* and *CRISPR 5-50* cell lines express no NAT1 activity (knockout).

Baseline ECAR was increased in the two NAT1 knockout cell lines compared to the cell line with parental (*Parent*) NAT1 activity (Figure 3.3). Baseline ECAR in the *CRISPR 2-19* cell line was increased 2.1-fold, 1.8-fold, 1.6-fold, and 1.4-fold with respect to the *Parent*, *Scrambled*, *Up*, and *CRISPR 5-50* cell lines. Baseline ECAR in the *CRISPR 5-50* cell line was also increased 1.5-fold compared to the *Parent* cell line ($p < 0.05$ for all).

In the NAT1 knockout cell lines glycolytic reserve was increased compared to the cell lines with parental (*Parent*, *Scrambled*) and increased (*Up*) NAT1 activity (Figure 3.3). Glycolytic reserve of the *CRISPR 2-19* cell line was increased 3.8-fold, 9.0-fold, and 45-fold with respect to the *Parent*, *Scrambled* and *Up* cell lines. Similarly, glycolytic reserve of the *CRISPR 5-50* cell line was increased 3.8-fold, 9.2-fold, and 46-fold with respect to the *Parent*, *Scrambled* and *Up* cell lines ($p < 0.05$ for all).

Transfection of the *Parent* cell line with the scrambled control showed no effect on baseline OCR/ECAR results. However, the cell line with increased NAT1 activity as well as one of the NAT1 knockout cell lines had decreased baseline OCR/ECAR relative to the *Parent* cell line (Figure 3.3). Both the *Up* and *CRISPR 2-19* cell lines were decreased 1.7-fold with respect to the *Parent* cell line ($p < 0.05$). We did not observe an effect of modulation of NAT1 activity on coupling efficiency, also called coupling ratio, which is defined as (oligomycin-sensitive OCR)/(basal OCR) (Figure 3.4). This result agrees with the data in Figure 3.3 showing that altering NAT1 activity had no effect on basal OCR or ATP-linked OCR in these cell lines.

Overall, it is important to note that the two cell lines with parental NAT1 activity (*Parent*, *Scrambled*) showed comparable results. The knockout of NAT1 in MDA-MB-231 cells led to differences in multiple bioenergetics measurements while the overexpression of NAT1 led to no significant ($p > 0.05$) differences when compared to the cell lines expressing parental (*Parent*, *Scrambled*) NAT1 activity.

To rule out differences in cell growth rate between cell lines over the course of our experiments, an MTT assay was conducted and the equivalence of MTT absorbance values were evaluated. With the exception of the *Parent* to *Scrambled* comparison, there was sufficient evidence to assert equivalence ($p < 0.05$) in each of the pairwise equivalence tests. The *Parent* to

Figure 3.3

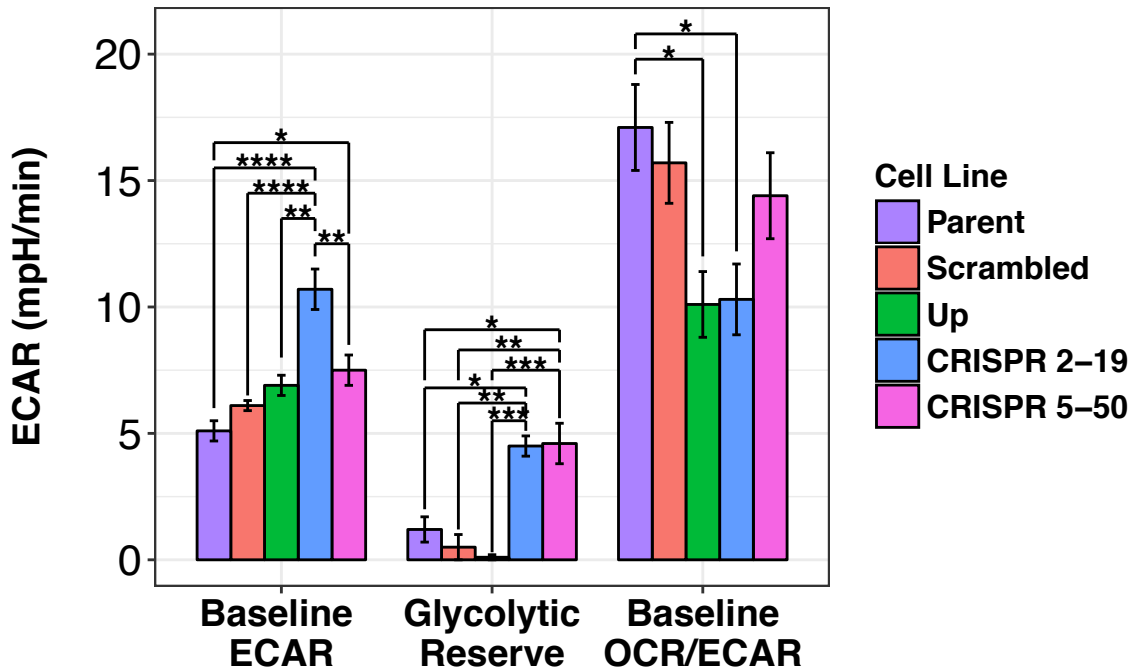


Figure 3.3: ECAR measurements in each cell line.

Baseline extracellular acidification rate (ECAR) was significantly increased in the *CRISPR 2-19* cell line when compared to the *Parent*, *Scrambled*, *Up*, and *CRISPR 5-50* cell lines. The *CRISPR 5-50* cell line was also significantly increased compared to the *Parent* cell line. Glycolytic reserve capacity was significantly increased in the *CRISPR 2-19* and *CRISPR 5-50* cell lines compared to the *Parent*, *Scrambled*, and *Up* cell lines. Baseline OCR/ECAR was significantly decreased in the *Up* and *CRISPR 2-19* cell lines when compared to the *Scrambled* cell line. Order of bars are preserved throughout figure (*Parent*, *Scrambled*, *Up*, *CRISPR 2-19*, *CRISPR 5-50*) and represent mean \pm SEM. N=6. *= $p < 0.05$, **= $p < 0.01$, ***= $p < 0.001$, ****= $p < 0.0001$. The *Parent* and *Scrambled* cell lines express parental NAT1 activity, the *Up* cell line expresses increased NAT1 activity, and the *CRISPR 2-19* and *CRISPR 5-50* cell lines express no NAT1 activity (knockout).

Figure 3.4

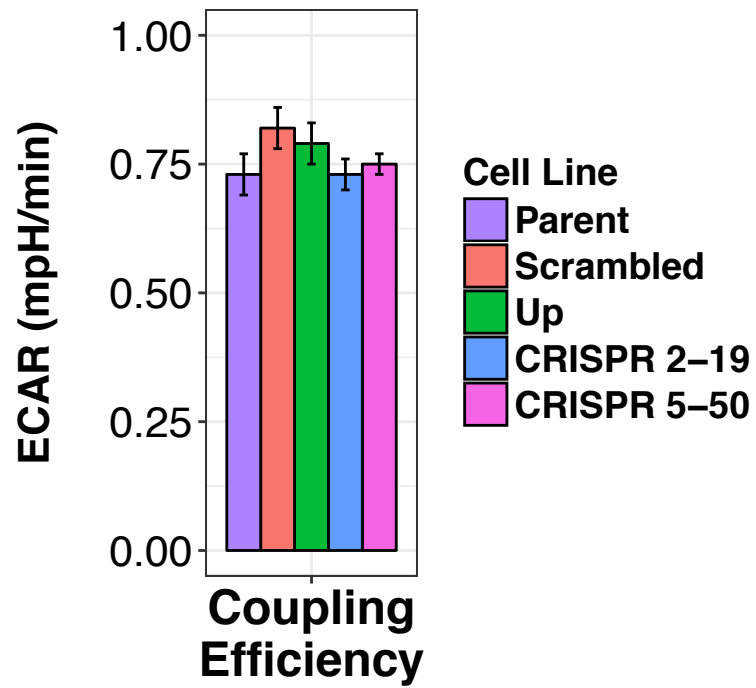


Figure 3.4: Coupling Efficiency measurements in each cell line.

Coupling efficiency did not significantly differ between cell lines. Bars represent mean \pm SEM.

N=6. The *Parent* and *Scrambled* cell lines express parental NAT1 activity, the *Up* cell line expresses increased NAT1 activity, and the *CRISPR 2-19* and *CRISPR 5-50* cell lines express no NAT1 activity (knockout).

Scrambled equivalence test was marginally significant ($p = 0.065$; Figure 3.5). This data suggests that modulation of NAT1 activity does not affect MDA-MB-231 cell viability over the course of these experiments.

Discussion

It has been previously reported that TNBC cell lines, including MDA-MB-231 cells, have profound metabolic changes characterized by decreased mitochondrial respiration and increased glycolysis when compared to breast cancer cell lines that are ER positive, PR positive, and/or HER2 positive¹⁴⁸. Since the purpose of this study was to evaluate how increased and knockout levels of human NAT1 affected the cellular bioenergetics of MDA-MB-231 breast cancer cells, the *Parent* MDA-MB-231 (no genetic alterations or transfections) was used as a baseline comparison. The *Scrambled* cell line (the *Parent* MDA-MB-231 cell line with a FRT site added into the genome and a scrambled shRNA transfected into that FRT site) was included as a transfection control⁸⁴. We did not observe a significant effect of the scrambled shRNA on cellular bioenergetics in MDA-MB-231 cells. The *Up* cell line (the *Parent* MDA-MB-231 cell line with a FRT site added into the genome and a plasmid overexpressing human NAT1 stably transfected into that FRT site) yields overexpression of NAT1⁸⁴. Finally, two complete NAT1 knockout cell lines (*CRISPR 2-19* and *CRISPR 5-50*) constructed using CRISPR/Cas9 technology (as described in the Methods section), were evaluated to verify that observed effects were due to differences in NAT1 as opposed to off-target effects caused by a specific guide-RNA. Therefore, we have concluded a result was due to the knockout of NAT1 only when the same trend was observed in both CRISPR/Cas9 constructed cell lines.

Knockout of NAT1 in MDA-MB-231 cells significantly altered the bioenergetics profile of the cells while increased NAT1 expression did not significantly alter the bioenergetics profile when compared to the *Parent* MDA-MB-231 cell line. Significant increases in reserve capacity, maximum mitochondrial capacity, and glycolytic reserve were observed in both NAT1 knockout cell lines compared to cell lines expressing parental and increased NAT1 activity. While basal OCR was unaffected by altered NAT1 activity, baseline ECAR was significantly increased in NAT1 knockout cells, suggesting an increase in glycolysis.

Figure 3.5

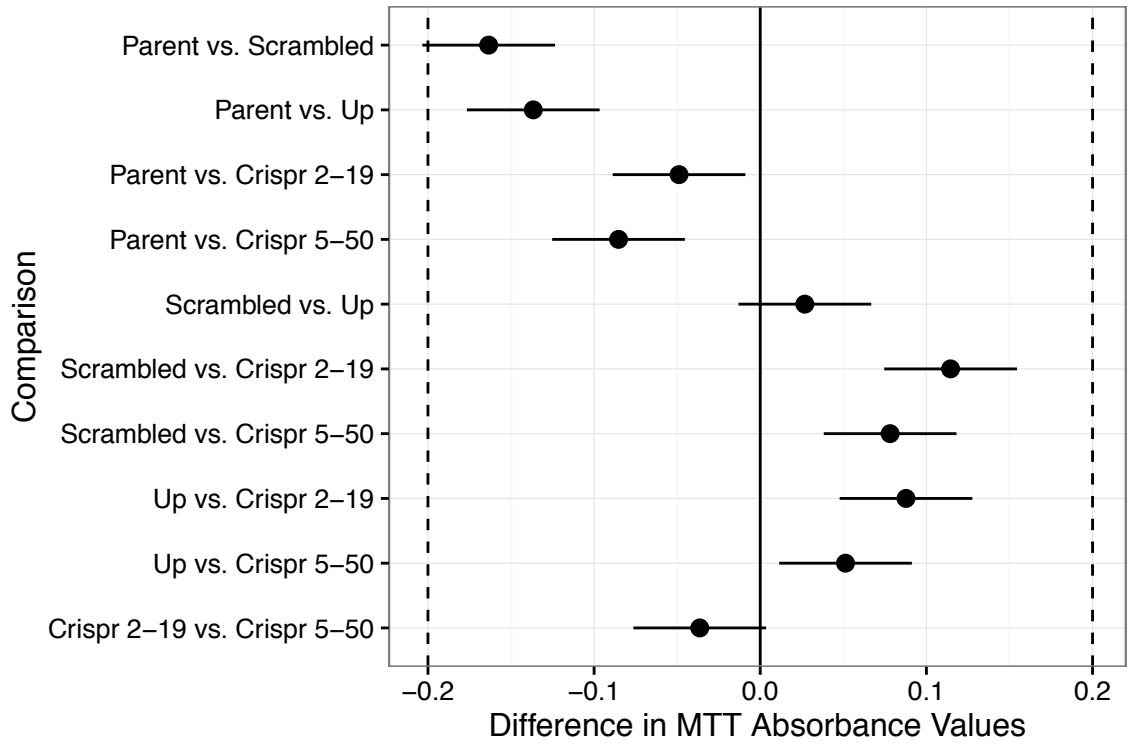


Figure 3.5: MTT Equivalence Test Results.

Equivalence between MTT absorbance values in each cell line was tested to access if there were differences in growth of the cell lines during bioenergetics experiments. There was sufficient evidence to assert equivalence in each of the pairwise equivalent tests.

Baseline OCR is an indicator of baseline mitochondrial respiration. ATP-linked OCR is the difference between OCR before and after ATP synthase is inhibited with oligomycin. This allows estimation of the OCR that is used to drive mitochondrial ATP synthesis and is largely set by the ATP demand of the cells¹⁴⁹. Modulation of NAT1 expression did not alter baseline or ATP-linked OCR in MDA-MB-231 cells. This result suggests NAT1 does not play a role in the ATP demand responses in the cells. Coupling efficiency is calculated as the fraction of baseline OCR used for ATP synthesis (ATP-linked OCR/baseline OCR)¹⁴⁰. Coupling efficiency was also unaffected by modulation of NAT1 in MDA-MB-231 cells providing further evidence to support the conclusion that that NAT1 does not play a role in ATP demand responses in these cells.

Reserve capacity is the difference between the basal and maximal respiration of the mitochondria and is broadly an evaluation of a cell's ability to respond to increased energy demands such as those found in rapidly dividing cancer cells¹⁴⁹. Increases in reserve capacity, as observed in the NAT1 knockout cell lines, could reflect enhanced oxidative capacity, mitochondrial biogenesis, or increased substrate provision¹⁴⁹. Reserve capacity does not explicitly implicate or identify molecular mechanisms of action since it is dependent upon multiple parameters. The significant increase in reserve capacity measured in the two NAT1 knockout cell lines appears to be driven by the increase in maximal respiration in those cell lines. Maximal respiration rate is primarily determined by substrate supply and oxidation¹⁴⁹. This includes substrate transport across the mitochondrial membrane as well as rate controlling metabolic enzymes. Taken together, these results indicate NAT1 may have a role, either directly or through its influence on acetyl-CoA levels, in regulation of mitochondrial substrate transport or metabolism. One could speculate that knockdown of NAT1 increases acetyl-CoA thereby increasing substrate(s) for the TCA cycle which could increase mitochondrial reserve capacity. This hypothesis should be investigated in future studies.

Glycolytic reserve is the difference between oligomycin-induced ECAR and baseline ECAR and is a measure of the maximum rate of conversion of glucose to pyruvate or lactate that can be achieved acutely by a cell¹⁵⁰. This measurement is an important parameter to evaluate in cancer cells because there is such an increased demand for energy precursors. Increases in

glycolytic reserve, as we observed in the two NAT1 knockout cell lines, also indicate that these cells can respond better to these increased energy demands. Here we report that NAT1 knockout, using two guide RNAs, increased the glycolytic reserve in MDA-MB-231 cells. There have been many studies^{49,81,137,151} investigating the inhibition of NAT1 with small molecule inhibitors as a possible way to decrease the cancerous and metastatic properties of malignant cells; however, this data suggest knockout of NAT1 may allow cells to increase glycolysis and use mitochondrial reserve.

Mitochondria facilitate cellular stress responses, including the response to hypoxia and the activation of programmed cell death via the release of pro-apoptotic molecules¹⁵². Differences in mitochondrial function between cells that express varying levels of NAT1 may play a role in cancer initiation or tumorigenesis by modulating these responses. It appears knockout of NAT1 allows cells to express more plasticity in terms of response to energy demand. Our data suggest that NAT1 may play a role in an unknown response mechanism that keeps cancer cells from hijacking the mitochondrial machinery to produce increased amounts of ATP that would be needed by cancer cells.

Summary and conclusions

In conclusion, differences in NAT1 activity, particularly the knockout of NAT1, significantly altered the bioenergetics profile of MDA-MB-231 triple negative breast cancer cells. Reserve capacity, maximal respiration, and glycolytic reserve capacity were increased in the NAT1 knockout cell lines. Increases in these measurements suggest that NAT1 knockout cells may be better able to respond to stress. These findings provide evidence that NAT1 modifies cellular acetyl-CoA levels and mitochondrial bioenergetics. Further investigation into the specific role NAT1 is playing in regulating cellular metabolism and bioenergetics is needed, ongoing, and will best be investigated via a multidisciplinary approach.

CHAPTER 4

METABOLOMICS

Background

Metabolomics is particularly well-suited for studying global metabolism as metabolites are the end-products of metabolism. In addition, performing an untargeted metabolomics study allows the collection of abundance data on many metabolites at once, from a single sample, giving a global snapshot of metabolism. While we acknowledge that untargeted metabolomics is still unable to detect and/or quantify ALL metabolites found in humans, it is the best way to get a global view of metabolism. The Human Metabolome Database (HMDB) approximates there are 22,000 endogenous metabolites that have been detected¹⁵³ however less than a thousand can be reliably measured. Many metabolomics studies have been conducted to better understand breast cancer and resulting alterations in metabolism¹⁵⁴⁻¹⁶⁴ and the use of metabolomics for systems biology approaches has recently been reviewed¹⁶⁵. Additionally, there has been increased interest recently in “oncometabolites”—metabolites that when dysregulated can contribute to the progression of cancer¹⁶⁶⁻¹⁷⁰. Notably, it has been suggested that oncometabolites can affect mitochondrial dynamics¹⁶⁶.

Materials and Methods

A global, untargeted metabolomics approach was utilized to interrogate differences in metabolic profile across six biological replicates of previously constructed and characterized MDA-MB-231 breast cancer cell lines expressing parental (*Scrambled*), increased (*Up*), decreased (*Down*, *CRISPR 2-12*), or knockout (*CRISPR 2-19*, *CRISPR 5-50*) levels of human arylamine *N*-acetyltransferase 1 (NAT1). All cell lines were cultured in high-glucose Dulbecco's

Modified Eagle Medium (DMEM), with 10% fetal bovine serum, 5% glutamine, and 5% penicillin/streptomycin.

Cells were plated in triplicate per biological replicate at a density of 500,000 cells per 150 x 25 mm cell plate, resulting in 18 plates total. Three plates were combined to form a single biological sample so that enough biological mass would be collected for future analysis. Cells were allowed to grow for three days in an incubator at 37 °C containing 5% CO₂. Prior to collecting the cells, 200 µL of media from each plate was reserved. Media from the 3 plates that were combined as a single biological replicate were collected in the same vial. Vials containing media were centrifuged to pellet out any cells that may have been present and the resulting supernatants were placed into separate cryovials. Cryovials containing media samples were then flash frozen in liquid nitrogen for 1 minute and then stored in -80 °C freezer for possible future metabolomic analysis.

Cells were then harvested on ice by adding 5 mL 0.25% trypsin and scraping the cells from the plate. Three plates of cells were combined to form one sample (biological replicate) to ensure there was enough cells for analysis. After harvesting the cells were washed 3 times with ice-cold 1 x PBS. Supernatant was removed and 100 µL of cell pellet from 3 samples of each cell line was reserved for transcriptomic analysis (presented in Chapter 5). The cryovials containing cell pellets were then flash frozen in liquid nitrogen for 1 minute followed by immediate storage at -80 °C. Samples were then shipped on dry ice to Metabolon Inc. (Durham, NC) for sample preparation and analysis, as described below.

Sample Accessioning

Following receipt, samples were inventoried and immediately stored at -80 °C by Metabolon. The Metabolon Laboratory Information Management System (LIMS) system was utilized for sample management. Samples were assigned a unique identifier by the LIMS that was associated with the original source identifier only. This identifier was used to track all sample handling, tasks, results, etc. The samples (and all derived aliquots) were tracked by the LIMS system. All portions of any sample were automatically assigned their own unique identifiers by the

LIMS when a new task was created; the relationship of these samples was also tracked. All samples were maintained at -80 °C until processed.

Sample Preparation

Metabolon sample preparation has been described in great detail elsewhere¹⁷¹ but is presented briefly here. Samples were prepared using the automated MicroLab STAR® system (Hamilton Company, Reno, NV). Several recovery standards (described in great detail elsewhere¹⁷¹) were added prior to the first step in the extraction process for quality control (QC) purposes. To remove protein, dissociate small molecules bound to protein or trapped in the precipitated protein matrix, and to recover chemically diverse metabolites, proteins were precipitated with methanol under vigorous shaking for 2 min (Glen Mills GenoGrinder 2000) followed by centrifugation. The resulting extract was divided into five fractions: two for analysis by two separate reverse phase ultrahigh performance tandem mass spectroscopy (RP/UPLC-MS/MS) methods with positive ion mode electrospray ionization (ESI), one for analysis by RP/UPLC-MS/MS with negative ion mode ESI, one for analysis by hydrophilic interaction liquid chromatography (HILIC)/UPLC-MS/MS with negative ion mode ESI, and one sample was reserved for backup. Samples were placed briefly on a TurboVap® (Zymark) to remove the organic solvent. The sample extracts were stored overnight under nitrogen before preparation for analysis.

Quality Assessment/Quality Control (QA/QC)

Several types of controls were analyzed in concert with the experimental samples: a pooled matrix sample generated by taking a small volume of each experimental sample served as a technical replicate throughout the data set; extracted water samples served as blanks; and a cocktail of QC standards that were carefully chosen not to interfere with the measurement of endogenous compounds were spiked into every analyzed sample, allowed instrument performance monitoring and aided chromatographic alignment. Instrument variability was determined by calculating the median relative standard deviation (RSD) for the standards that were added to each sample prior to injection into the mass spectrometers. Overall process variability was determined by calculating the median RSD for all endogenous metabolites (i.e.,

non-instrument standards) present in 100% of the pooled matrix samples. Experimental samples were randomized across the platform run with QC samples spaced evenly among the injections.

Ultrahigh Performance Liquid Chromatography-Tandem Mass Spectrometry (UPLC-MS/MS)

All methods utilized a Waters ACQUITY ultra-performance liquid chromatography (UPLC) and a Thermo Scientific Q-Exactive high resolution/accurate mass spectrometer interfaced with a heated electrospray ionization (HESI-II) source and an Orbitrap mass analyzer operated at 35,000 mass resolution. The sample extract was dried, then reconstituted in solvents compatible to each of the four methods. Each reconstitution solvent contained a series of standards at fixed concentrations to ensure injection and chromatographic consistency. One aliquot was analyzed using acidic positive ion conditions, chromatographically optimized for more hydrophilic compounds. In this method, the extract was gradient eluted from a C18 column (Waters UPLC ethylene bridged hybrid (BEH) C18-2.1x100 mm, 1.7 μ m) using water and methanol, containing 0.05% perfluoropentanoic acid (PFPA) and 0.1% formic acid (FA). Another aliquot was also analyzed using acidic positive ion conditions, however it was chromatographically optimized for more hydrophobic compounds. In this method, the extract was gradient eluted from the same aforementioned C18 column using methanol, acetonitrile, water, 0.05% PFPA and 0.01% FA and was operated at an overall higher organic content. Another aliquot was analyzed using basic negative ion optimized conditions using a separate dedicated C18 column. The basic extracts were gradient eluted from the column using methanol and water, however with 6.5 mM ammonium bicarbonate at pH 8.0. The fourth aliquot was analyzed via negative ionization following elution from a HILIC column (Waters UPLC BEH Amide 2.1x150 mm, 1.7 μ m) using a gradient consisting of water and acetonitrile with 10 mM ammonium formate, pH 10.8. The MS analysis alternated between MS and data-dependent MSⁿ scans using dynamic exclusion. The scan range varied slightly between methods but covered 70-1000 m/z. Raw data files were archived and extracted as described below.

Data Extraction and Compound Identification

Raw data were extracted, peak-identified and QC processed by Metabolon utilizing proprietary methods. Compounds were identified by comparison to library entries of purified standards or recurrent unknown entities. Metabolon maintains a library based on authenticated standards that contains the retention time/index (RI), mass to charge ratio (m/z), and chromatographic data (including MS/MS spectral data) on all molecules present in the library. Furthermore, biochemical identifications are based on three criteria: retention index within a narrow RI window of the proposed identification, accurate mass match to the library +/- 10 ppm, and the MS/MS forward and reverse scores between the experimental data and authentic standards. The MS/MS scores are based on a comparison of the ions present in the experimental spectrum to the ions present in the library spectrum. While there may be similarities between these molecules based on one of these factors, the use of all three data points can be utilized to distinguish and differentiate biochemicals. More than 3300 commercially available purified standard compounds have been acquired and registered into LIMS for analysis on all platforms for determination of their analytical characteristics. Additional mass spectral entries have been created for structurally unnamed biochemicals, which have been identified by virtue of their recurrent nature (both chromatographic and mass spectral). These compounds have the potential to be identified by future acquisition of a matching purified standard or by classical structural analysis.

Curation

A variety of curation procedures were carried out by Metabolon to ensure that a high quality data set was made available for statistical analysis and data interpretation. The QC and curation processes were designed to ensure accurate and consistent identification of true chemical entities, and to remove those representing system artifacts, mis-assignments, and background noise. Metabolon data analysts use proprietary visualization and interpretation software to confirm the consistency of peak identification among the various samples. Library matches for each compound were checked for each sample and corrected if necessary.

Metabolite Quantification and Data Normalization

Peaks were quantified using area-under-the-curve. For studies spanning multiple days, a data normalization step was performed to correct variation resulting from instrument inter-day tuning differences. Essentially, each compound was corrected by registering the medians equal to one (1.00) and normalizing each data point proportionately. For studies that did not require more than one day of analysis, no normalization is necessary, other than for purposes of data visualization. Biochemical data was also normalized to total protein as determined by Bradford assay to account for differences in metabolite levels due to differences in the amount of material present in each sample.

Statistical Analyses

Figure 4.1 illustrates experimental approach and data analyses methods. Normality of data was assessed via the Shapiro–Wilk test; normality of non-log and log transformed data was visualized (Figure 4.2). All statistical analyses were performed on log transformed data using R: A Language and Environment for Statistical Computing version 3.3.1¹³². One-way ANOVA was performed to test for significant differences between groups for each metabolite at an alpha level of 0.05. Q-values were then calculated using the false discovery rate method to correct for multiple hypothesis testing¹⁷². Dunnett's post-tests were utilized to compare all groups (*Up*, *Down*, *CRISPR 2-12*, *CRISPR 2-19*, *CRISPR 5-50*) to the *Scrambled* group for those metabolites with one-way ANOVA $q \leq 0.05$. Q-values were then calculated using the false discovery rate (FDR) method for Dunnett's post-test p -values.

Fold-change was calculated using equation 1 (shown below). Briefly, the mean abundance for each metabolite was calculated for each group. We then divided the mean of the comparison group (*Up*, *Down*, *CRISPR 2-12*, *CRISPR 2-19*, *CRISPR 5-50*) by the mean of the *Scrambled* group to give us fold-change relative to the *Scrambled* group.

$$FC = 2^{\left| \log_2 \frac{\bar{B}}{\bar{A}} \right|} \quad \text{Equation 1}$$

In the equation, FC = fold-change, \bar{A} = average metabolite abundance in the reference group, \bar{B} = average metabolite abundance in the comparison group. A negative sign was added to the fold-change when metabolite abundance was lower in the comparison group compared to the reference group.

Figure 4.1

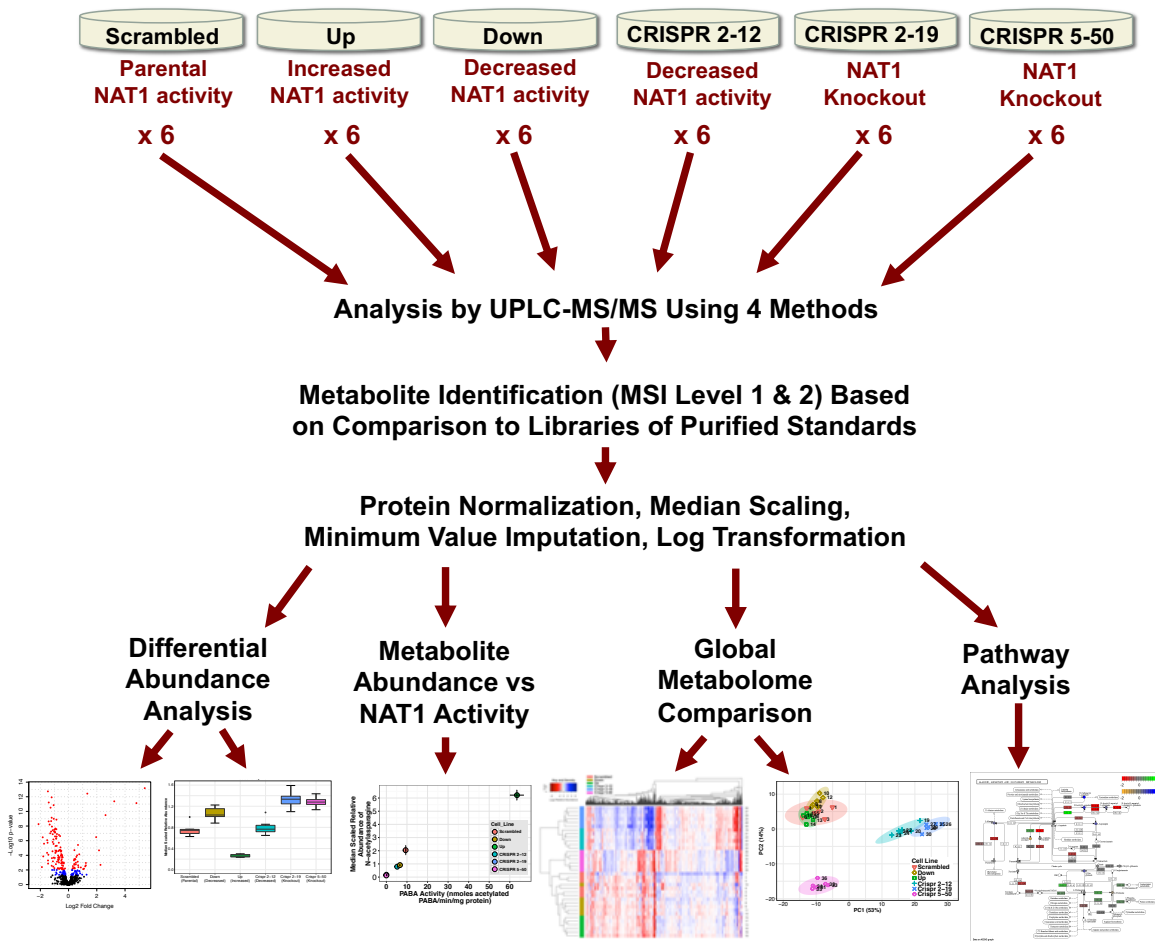


Figure 4.1: Metabolomics Experimental Approach Diagram.

Six biological replicates from each cell line were collected (3 plates of cells were pooled to form a single biological replicate to have enough sample for analysis). Samples were then analyzed by UPLC-MS/MS at Metabolon using 4 methods optimized for the greatest coverage across the metabolome. Following metabolite identification, abundance data was protein normalized, median scaled, minimum values imputed, and log-transformed. Metabolite abundances were then analyzed for differential abundance, correlation with NAT1 *N*-acetylation activity, unbiased multivariate analysis/clustering, and pathway enrichment.

Figure 4.2

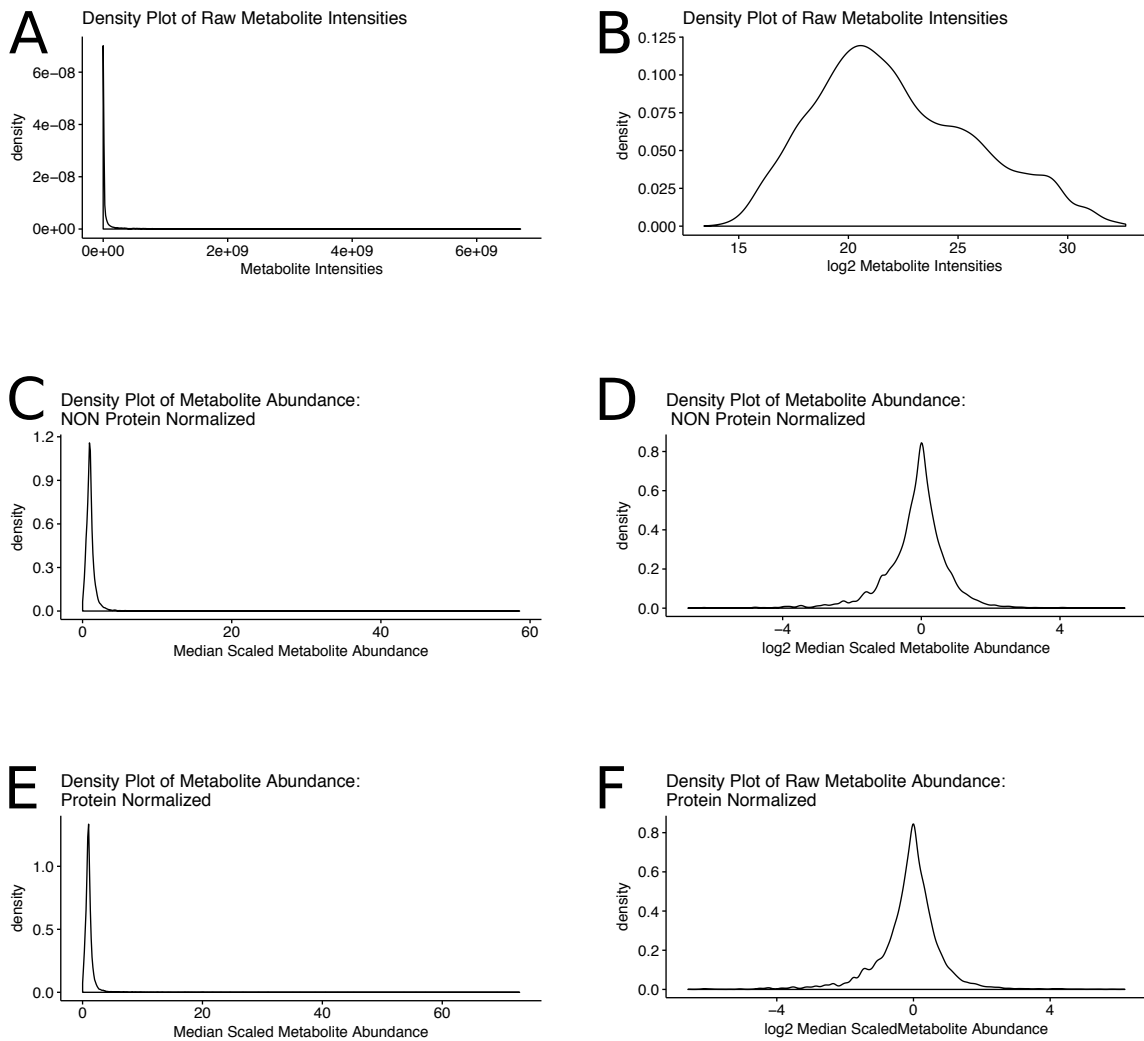


Figure 4.2: Assessing Normality of Metabolomics Data.

Normality of raw metabolite intensities before (A) and after log transformation was assessed (B). Normality of non-protein normalized metabolite abundances before (C) and after log transformation was assessed (D). Normality of protein normalized metabolite abundances before (E) and after log transformation was assessed (F). For all, log transformation led to approximate normality.

Data were plotted as volcano plots to simultaneously visualize the between group differences in abundance of all detected metabolites and significance. Additionally, data were plotted as box-plots to better visualize the abundance distribution of each metabolite between all groups. Pearson correlation was calculated between NAT1 activity and relative metabolite abundance for all metabolites to generate hypotheses about novel NAT1 substrates or products. Additionally, Pearson correlation was calculated between carnosine and metabolites whose abundance was concordantly altered in the two NAT1 KO cell lines. Data were also plotted as a heatmap and hierarchical clustering was conducted using the Weighted Pair Group Method with Arithmetic Mean (WPGMA) method. Principal component analysis was conducted by singular value decomposition of the centered data matrix. The loadings of the first (x-axis) and second (y-axis) principal component were plotted. Weighted gene co-expression network analysis (WGCNA) was conducted on metabolite abundance data. Enrichment analysis was conducted for each group compared to *Scrambled*. The normalized enrichment score was utilized to determine the relative degree of enrichment.

Results

Univariate Analyses

A large proportion (515/567; 90.8%) of the detected metabolites were found to significantly differ ($q \leq 0.05$) between the six cell lines (Table 4.1). Following Dunnett's post tests it was observed that more metabolites differed in the cell lines constructed via CRISPR/Cas9 than the cell lines constructed via siRNA when compared to the *Scrambled* cell line. Thirty-four point four percent, 28.4%, 61.7%, 64.4%, and 53.8% of total detected metabolites were differentially expressed in the *Down*, *Up*, *CRISPR 2-12*, *CRISPR 2-19*, and *CRISPR 5-50* groups, respectively. Metabolites were further characterized by direction of fold-change compared to *Scrambled* (Table 4.1; Figures 4.3 – 4.7); more metabolites were decreased than increased in all group comparisons to *Scrambled* except the *CRISPR 2-19* cell line. The CRISPR/Cas-9 generated cell lines had not only more total metabolites differentially expressed compared to the siRNA generated cell lines, but also more metabolites whose fold-changes were greater than 4.

Table 4.1

Metabolomics Summary Statistics

One-way ANOVA	Statistical Comparison				
	All Groups				
Total Metabolites $p \leq 0.05$	519				
Total Metabolites $q \leq 0.05$	515				
Dunnett's <i>t</i> -test Post Test	Statistical Comparison				
	<u>Down Scrambled</u>	<u>Up Scrambled</u>	<u>CRISPR 2-12 Scrambled</u>	<u>CRISPR 2-19 Scrambled</u>	<u>CRISPR 5-50 Scrambled</u>
Total Metabolites $p \leq 0.05$	220	199	380	384	326
Metabolites (↑↓)	68 152	155 44	190 190	179 205	173 153
Total Metabolites $q \leq 0.05$	198	161	352	365	305
Metabolites (↑↓)	60 138	30 131	175 177	173 192	163 142

One-way ANOVA was conducted for each metabolite abundance between all groups. *q*-values were calculated from resulting one-way ANOVA *p*-values to account for the multiple hypothesis testing. Dunnett's *t*-test post tests were conducted for each group compared to *Scrambled* on metabolites found to significantly vary between all groups. Significant metabolites were further characterized by direction of fold-change; metabolites with increased abundance are shown in green while metabolites with decreased abundances are shown in red. *q*-values were calculated from resulting Dunnett's *t*-test *p*-values.

Figure 4.3

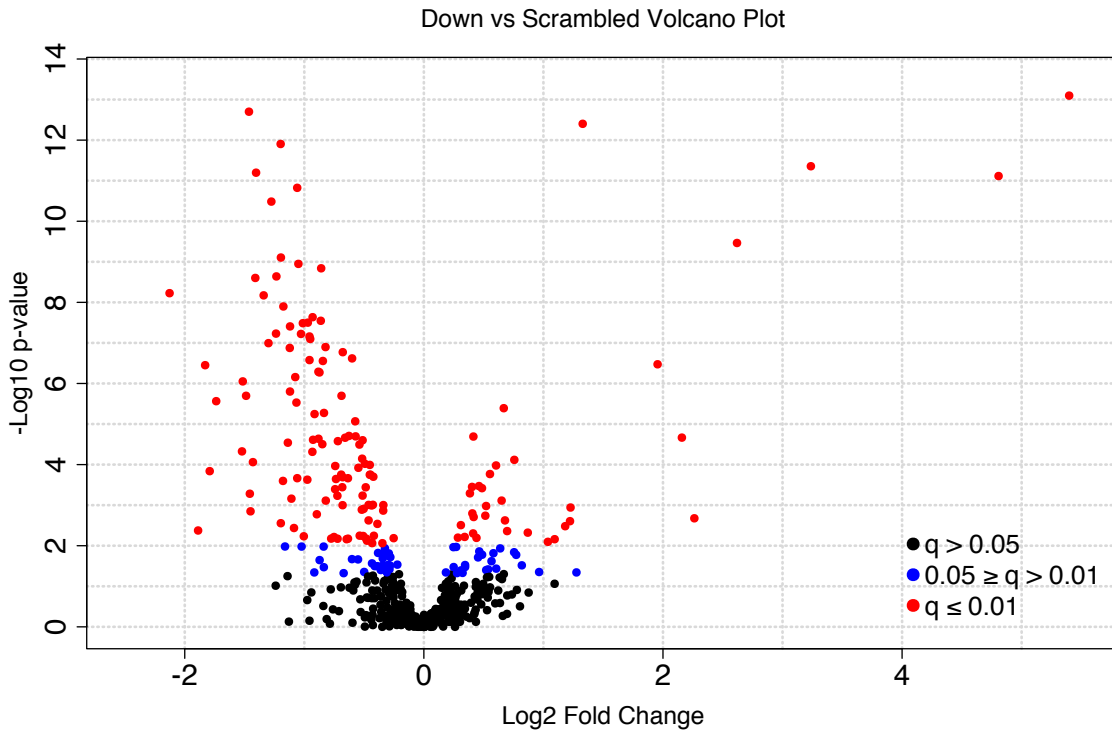


Figure 4.3: Down vs Scrambled Volcano Plot.

Each dot represents a single metabolite and is color coded according to q -value. The black dots represent metabolites that had a Dunnett's post test q -value greater than 0.05, blue dots represent metabolites that had a q -value less than or equal to 0.05 but greater than 0.01, and red dots represent metabolites that had a q -value less than or equal to 0.01. Negative fold changes represent a decrease in that metabolite compared to the *Scrambled* group while positive fold changes represent an increase in that metabolite compared to the *Scrambled* group.

Figure 4.4

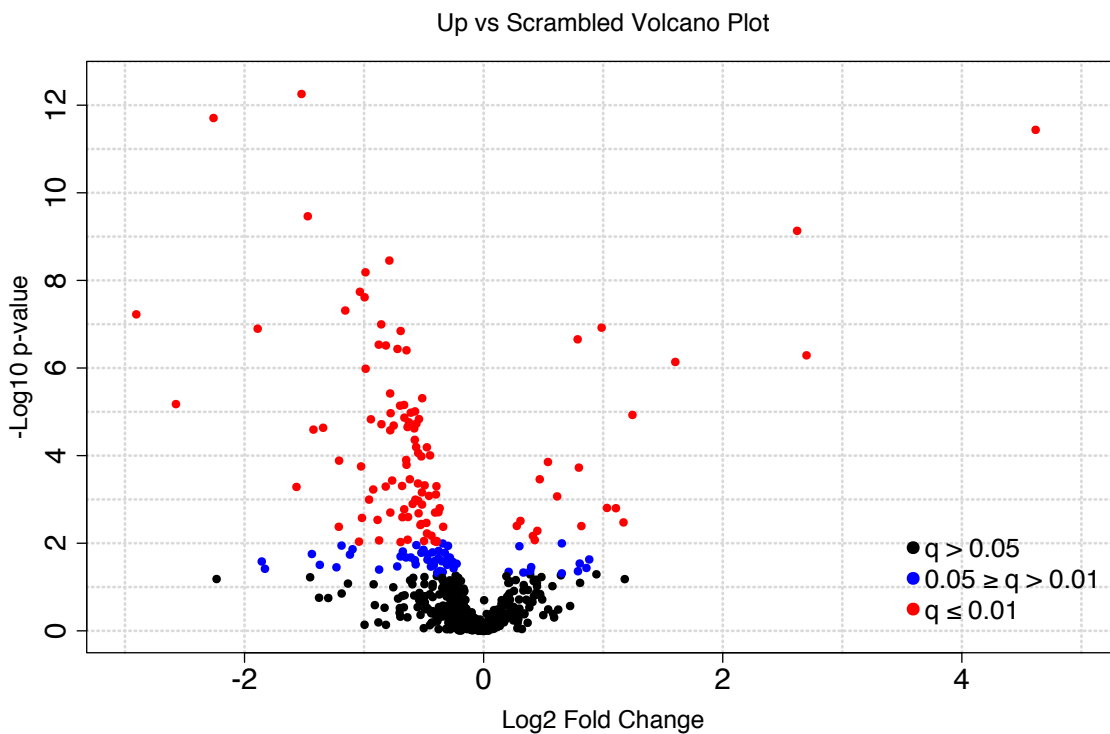


Figure 4.4: Up vs Scrambled Volcano Plot.

Each dot represents a single metabolite and is color coded according to q -value. The black dots represent metabolites that had a Dunnett's post test q -value greater than 0.05, blue dots represent metabolites that had a q -value less than or equal to 0.05 but greater than 0.01, and red dots represent metabolites that had a q -value less than or equal to 0.01. Negative fold changes represent a decrease in that metabolite compared to the *Scrambled* group while positive fold changes represent an increase in that metabolite compared to the *Scrambled* group.

Figure 4.5

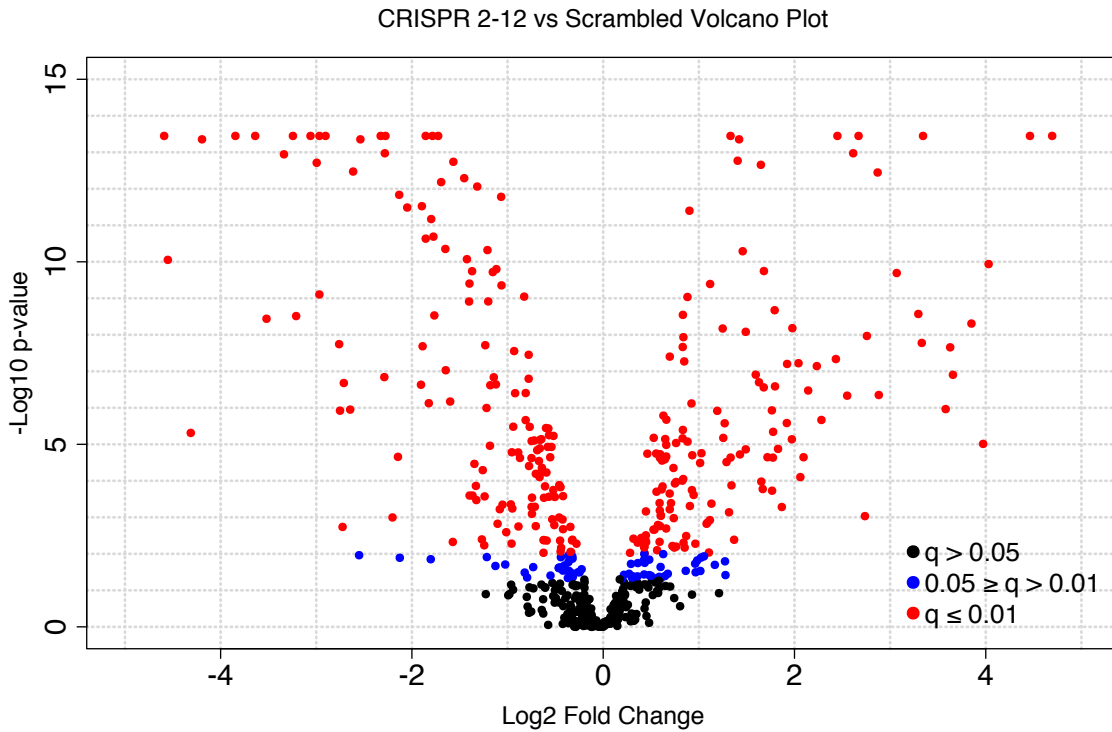


Figure 4.5: CRISPR 2-12 vs Scrambled Volcano Plot.

Each dot represents a single metabolite and is color coded according to q -value. The black dots represent metabolites that had a Dunnett's post test q -value greater than 0.05, blue dots represent metabolites that had a q -value less than or equal to 0.05 but greater than 0.01, and red dots represent metabolites that had a q -value less than or equal to 0.01. Negative fold changes represent a decrease in that metabolite compared to the *Scrambled* group while positive fold changes represent an increase in that metabolite compared to the *Scrambled* group.

Figure 4.6

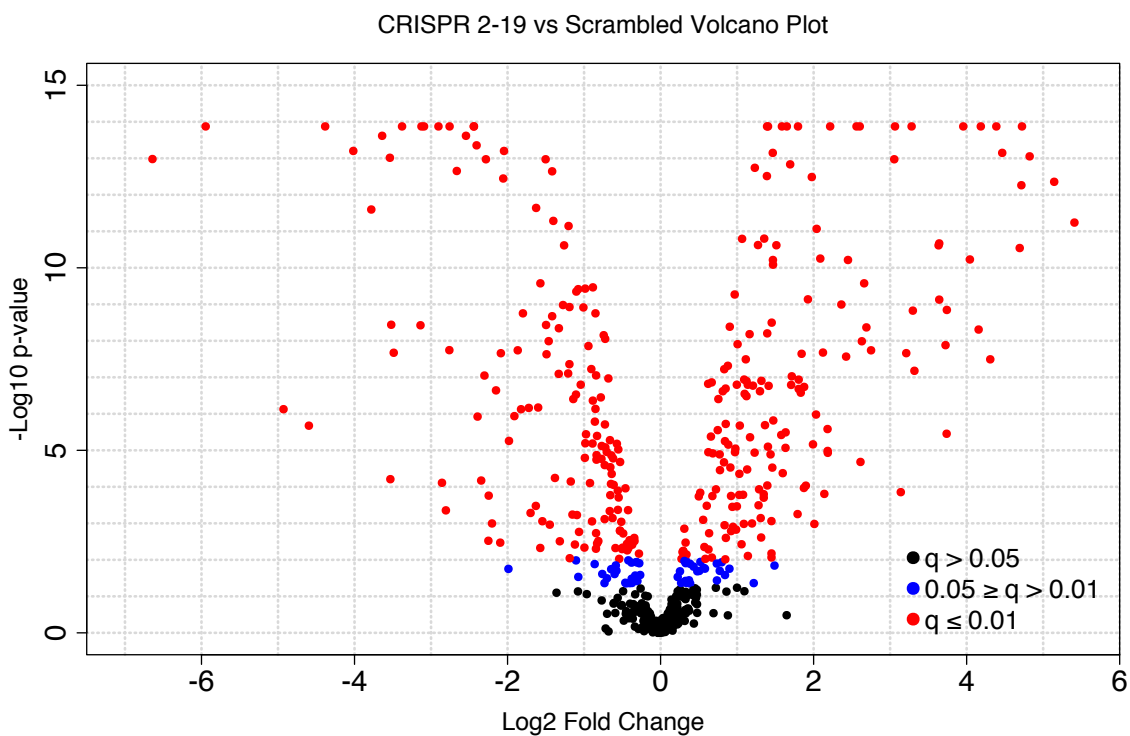


Figure 4.6: CRISPR 2-19 vs Scrambled Volcano Plot.

Each dot represents a single metabolite and is color coded according to q -value. The black dots represent metabolites that had a Dunnett's post test q -value greater than 0.05, blue dots represent metabolites that had a q -value less than or equal to 0.05 but greater than 0.01, and red dots represent metabolites that had a q -value less than or equal to 0.01. Negative fold changes represent a decrease in that metabolite compared to the *Scrambled* group while positive fold changes represent an increase in that metabolite compared to the *Scrambled* group.

Figure 4.7

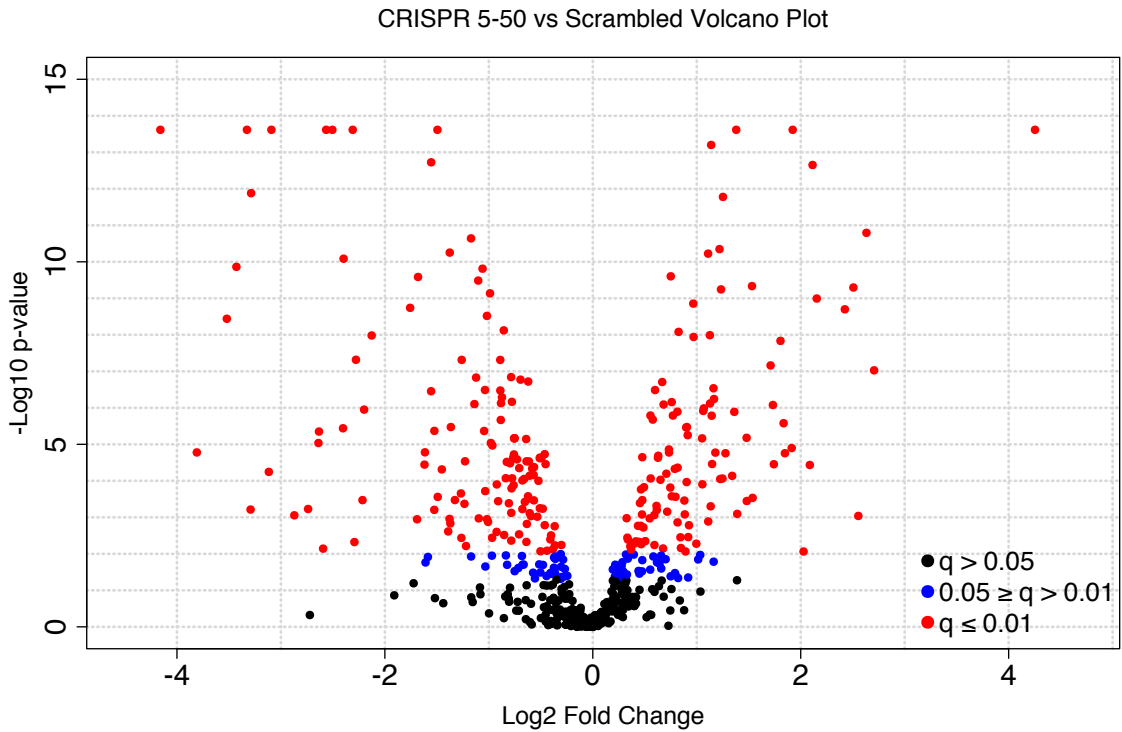


Figure 4.7: CRISPR 5-50 vs Scrambled Volcano Plot.

Each dot represents a single metabolite and is color coded according to q -value. The black dots represent metabolites that had a Dunnett's post test q -value greater than 0.05, blue dots represent metabolites that had a q -value less than or equal to 0.05 but greater than 0.01, and red dots represent metabolites that had a q -value less than or equal to 0.01. Negative fold changes represent a decrease in that metabolite compared to the *Scrambled* group while positive fold changes represent an increase in that metabolite compared to the *Scrambled* group.

The overlap in significant metabolites with a fold-change greater than or equal to 2 was compared between the two NAT1 KO cell lines and the *Scrambled* group (Figure 4.8). Eighteen metabolites were increased concordantly in the two NAT1 KO cell lines compared to *Scrambled* with 102 and 32 metabolites uniquely increased in the *CRISPR 2-19* and *CRISPR 5-50* cell lines, respectively. Twenty-five metabolites were decreased concordantly in the two NAT1 KO cell lines compared to *Scrambled* with 58 and 38 metabolites uniquely decreased in the *CRISPR 2-19* and *CRISPR 5-50* cell lines, respectively. Table 4.2 lists metabolites whose abundances were concordantly changed in the NAT1 KO cell lines. More metabolites had conflicting differential abundance between the two CRISPR NAT1 KO cell lines compared to *Scrambled* than those that agreed. Notably, many of the metabolites decreased concordantly in the CRISPR NAT1 KO cell lines were carnitine conjugates. Assessing correlation between carnitine and all metabolites concordantly changed in the CRISPR NAT1 KO cell lines revealed the abundance of most metabolites were associated with carnitine (Table 4.2). This suggests dysregulation of carnitine is driving the differential abundances observed in this subset of metabolites.

Eight metabolites, *N*-acetylasparagine, *N*-acetylputrescine, saccharopine, cytidine, 1-palmitoyl-2- α linolenoyl-GPC (16:0/18:3n3), isovalerylcarnitine (C5), cysteine sulfinic acid, and serotonin, were significantly associated with NAT1 PABA *N*-acetylation activity (Table 4.3). The last five of the eight metabolites listed had a high degree of variation in the within-group measurement of metabolite abundance therefore the association is not as well defined as the others. The top two metabolites correlated with PABA *N*-acetylation, *N*-acetylasparagine (Figure 4.9; Table 4.4) and *N*-acetylputrescine (Figure 4.10; Table 4.5), had a correlation coefficient (*r*) value greater than 0.9 and are *N*-acetylated compounds, suggesting they may be products of *N*-acetylation by NAT1. Two of the metabolites significantly correlated with PABA *N*-acetylation, saccharopine (Figure 4.11; Table 4.6) and isovalerylcarnitine (C5), had an inverse relationship suggesting a role in a NAT1 catalyzed reaction as the substrate or possibly down-stream of a NAT1 catalyzed reaction. Additionally, differential acetyl-CoA levels due to NAT1's ability to hydrolyze acetyl-CoA could be driving these observations.

Multivariate/Multivariable Analyses

Figure 4.8

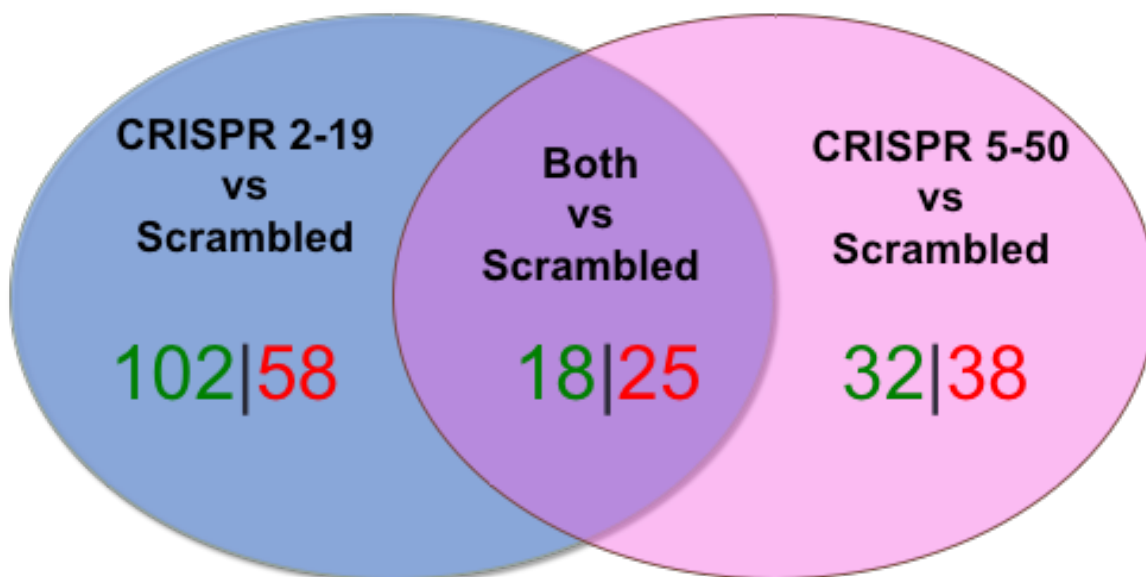


Figure 4.8: Concordance of Metabolite Abundance Differences in CRISPR NAT1 Knockout Cell Lines.

Significant metabolites with a fold-change greater than or equal to 2 were compared between CRISPR/Cas9 generated NAT1 knockout cell lines. Eighteen metabolites were increased concordantly in the two NAT1 KO cell lines compared to Scrambled with 102 and 32 metabolites uniquely increased in the CRISPR 2-19 and CRISPR 5-50 cell lines, respectively. Twenty-five metabolites were decreased concordantly in the two NAT1 KO cell lines compared to *Scrambled* with 58 and 38 metabolites uniquely decreased in the *CRISPR 2-19* and *CRISPR 5-50* cell lines, respectively. Green: metabolites increased compared to *Scrambled*; red: metabolites decreased compared to *Scrambled*.

Table 4.2Metabolites Concordantly Differentially Abundant in *CRISPR 2-19* and *CRISPR 5-50* Cell Lines

M #	BIOCHEMICAL	ANOVA q-value	FOLD-CHANGE		CORRELATION WITH CARNITINE
			CRISPR 2-19/S	CRISPR 5-50/S	
M6	1-(1-enyl-palmitoyl)-2-linoleoyl-GPE (P-16:0/18:2)*	<0.0001	4.1	2.2	-0.60
M26	13-HODE + 9-HODE	0<0.002	2.5	2.0	-0.68
M33	1-linoleoyl-GPE (18:2)*	<0.0001	5.4	3.6	-0.68
M116	3-hydroxydecanoate	<0.0001	2.7	3.3	-0.91
M119	3-hydroxylaurate	<0.0001	2.2	2.3	-0.77
M120	3-hydroxyoctanoate	<0.0001	2.8	3.3	-0.62
M233	dihydroxyacetone phosphate (DHAP)	<0.0001	5.5	4.5	-0.77
M339	lactose	<0.0001	26.4	19.1	-0.82
M387	N-acetyl-beta-alanine	<0.0001	13.3	4.1	-0.72
M417	nicotinamide ribonucleotide (NMN)	<0.0001	25.9	2.9	-0.41
M437	oleoylcholine	<0.0001	8.3	2.6	-0.66
M445	palmitoleoyl ethanolamide*	<0.0001	3.0	2.3	-0.70
M447	palmitoleoylcholine	<0.0001	8.4	2.4	-0.57
M452	palmitoylcholine	<0.0001	9.7	2.1	-0.53
M456	penicillin G	<0.0001	18.2	5.7	-0.70
M481	pyridoxine (Vitamin B6)	<0.0001	4.5	3.8	-0.83
M526	stearoyl ethanolamide	<0.0001	2.6	2.2	-0.82
M555	urate	<0.0001	2.0	3.5	-0.87
M102	2'-O-methylcytidine	<0.002	0.4	0.4	0.88
M110	3-aminoisobutyrate	<0.0001	0.3	0.5	0.79
M132	4-hydroxyglutamate	<0.0001	0.2	0.2	0.93
M161	adrenoylcarnitine (C22:4)*	<0.0001	0.1	0.1	0.92
M176	arachidonoylcarnitine (C20:4)	<0.0001	0.2	0.2	0.91
M190	beta-guanidinopropanoate	<0.0001	0.3	0.4	0.84
M210	cis-4-decenoylcarnitine (C10:1)	<0.0001	0.2	0.4	0.69
M218	cystathionine	<0.0001	0.2	0.1	0.92
M226	cytidine diphosphate	<0.0001	0.3	0.4	0.85
M231	dihomo-linolenoylcarnitine (20:3n3 or 6)*	<0.0001	0.1	0.2	0.86
M232	dihomo-linoleoylcarnitine (C20:2)*	<0.0001	0.1	0.4	0.88

*Continued on Next Page

M242	docosatrienoylcarnitine (C22:3)*	<0.0001	0.1	0.2	0.88
M344	laurylcarnitine (C12)	<0.0001	0.2	0.5	0.85
M352	linolenoylcarnitine (C18:3)*	<0.0001	0.1	0.2	0.83
M353	linoleoylcarnitine (C18:2)*	<0.0001	0.0	0.3	0.77
M371	myristoleoylcarnitine (C14:1)*	<0.0001	0.1	0.4	0.75
M376	N2,N2-dimethylguanosine	<0.006	0.5	0.4	0.70
M384	N-acetylasparagine	<0.0001	0.1	0.1	0.28
M412	N-carbamoylaspartate	<0.0001	0.1	0.1	0.89
M441	orotate	<0.0001	0.1	0.1	0.95
M448	palmitoyl dihydrosphingomyelin (d18:0/16:0)*	<0.0001	0.4	0.4	0.95
M453	pantetheine	<0.0001	0.5	0.3	0.78
M508	sphingomyelin (d18:0/18:0, d19:0/17:0)*	<0.0001	0.4	0.5	0.63
M544	tryptamine	<0.0001	0.5	0.5	0.93
M559	uridine 5'-triphosphate (UTP)	<0.0001	0.04	0.1	0.95

Metabolites concordantly differentially abundant in *CRISPR 2-19* and *CRISPR 5-50* cell lines were identified. Eighteen metabolites were concordantly increased while 25 metabolites were concordantly decreased. Pearson correlation between identified concordant metabolites and carnitine was conducted. M# represents an arbitrarily assigned identifier for analysis purposes, biochemical is the metabolite identity, ANOVA *q*-value is the one-way ANOVA *q*-value, and correlation is the Pearson correlation coefficient between that metabolite and carnitine. Fold-changes were color coded according to direction of fold-change with green for metabolites increased compared to *Scrambled* and red for metabolites decreased compared to *Scrambled*.

Table 4.3Metabolites Significantly Correlated with NAT1 *N*-acetylation Activity

M#	BIOCHEMICAL	<i>p</i>-value	CORRELATION COEFFICIENT
M384	<i>N</i> -acetylasparagine	0.00030	0.986
M404	<i>N</i> -acetylputrescine	0.0046	0.944
M491	saccharopine	0.022	-0.876
M224	cytidine	0.029	0.856
M49	1-palmitoyl-2- α -linolenoyl-GPC (16:0/18:3n3)*	0.043	0.825
M336	isovalerylcarnitine (C5)	0.046	-0.820
M220	cysteine sulfinic acid	0.047	0.816
M498	serotonin	0.050	0.811

Pearson correlation between NAT1 *N*-acetylation activity in constructed cell lines and all 567 detected metabolites was conducted. Eight metabolites were significantly correlated with NAT1 *N*-acetylation activity. M# represents an arbitrability assigned identifier for analysis purposes, biochemical is the metabolite identity, *p*-value is the Pearson correlation *p*-value, and correlation is the Pearson correlation coefficient. Correlation coefficients were color coded according to direction of correlation, with green for positive correlation and red for inverse (negative) correlation.

Table 4.4***N*-acetylasparagine**

One-way ANOVA	Statistical Comparison				
	All Groups				
q-value	<0.00001				
Dunnett's t-test Post Test	Statistical Comparison				
	<u>Down</u> Scrambled	<u>Up</u> Scrambled	<u>CRISPR 2-12</u> Scrambled	<u>CRISPR 2-19</u> Scrambled	<u>CRISPR 5-50</u> Scrambled
q-value	<0.001	<0.00001	<0.0001	<0.00001	<0.00001
Fold-Change	-2.3	3.0	-2.5	-11.6	-17.9

One-way ANOVA was conducted for *N*-acetylasparagine abundance between all groups. *q*-values were calculated from resulting one-way ANOVA *p*-values to account for the multiple hypothesis testing. Dunnett's *t*-test post tests were conducted for each group compared to *Scrambled*. *N*-acetylasparagine abundance was further characterized by direction of fold-change; metabolites with increased abundance are shown in green while metabolites with decreased abundances are shown in red. *q*-values were calculated from resulting Dunnett's *t*-test *p*-values.

Table 4.5***N*-acetylputrescine**

One-way ANOVA	Statistical Comparison				
	All Groups				
<i>q</i> -value	<0.0001				
Dunnett's <i>t</i> -test Post Test	Statistical Comparison				
	<u>Down Scrambled</u>	<u>Up Scrambled</u>	<u>CRISPR 2-12 Scrambled</u>	<u>CRISPR 2-19 Scrambled</u>	<u>CRISPR 5-50 Scrambled</u>
<i>q</i> -value	0.03	<0.00001	0.95	<0.00001	0.003
Fold-Change	-1.2	1.7	1	-1.9	-1.3

One-way ANOVA was conducted for *N*-acetylputrescine abundance between all groups. *q*-values were calculated from resulting one-way ANOVA *p*-values to account for the multiple hypothesis testing. Dunnett's *t*-test post tests were conducted for each group compared to *Scrambled*. *N*-acetylputrescine abundance was further characterized by direction of fold-change; metabolites with increased abundance are shown in green while metabolites with decreased abundances are shown in red. *q*-values were calculated from resulting Dunnett's *t*-test *p*-values

Table 4.6**Saccharopine**

One-way ANOVA	Statistical Comparison				
	All Groups				
q-value	<0.00001				
Dunnett's <i>t</i> -test Post Test	Statistical Comparison				
	<u>Down Scrambled</u>	<u>Up Scrambled</u>	<u>CRISPR 2-12 Scrambled</u>	<u>CRISPR 2-19 Scrambled</u>	<u>CRISPR 5-50 Scrambled</u>
q-value	<0.001	<0.00001	0.57	<0.00001	<0.00001
Fold-Change	-1.4	-2.9	1	1.8	1.7

One-way ANOVA was conducted for saccharopine abundance between all groups. *q*-values were calculated from resulting one-way ANOVA *p*-values to account for the multiple hypothesis testing. Dunnett's *t*-test post tests were conducted for each group compared to *Scrambled*. Saccharopine abundance was further characterized by direction of fold-change; metabolites with increased abundance are shown in green while metabolites with decreased abundances are shown in red. *q*-values were calculated from resulting Dunnett's *t*-test *p*-values.

Figure 4.9

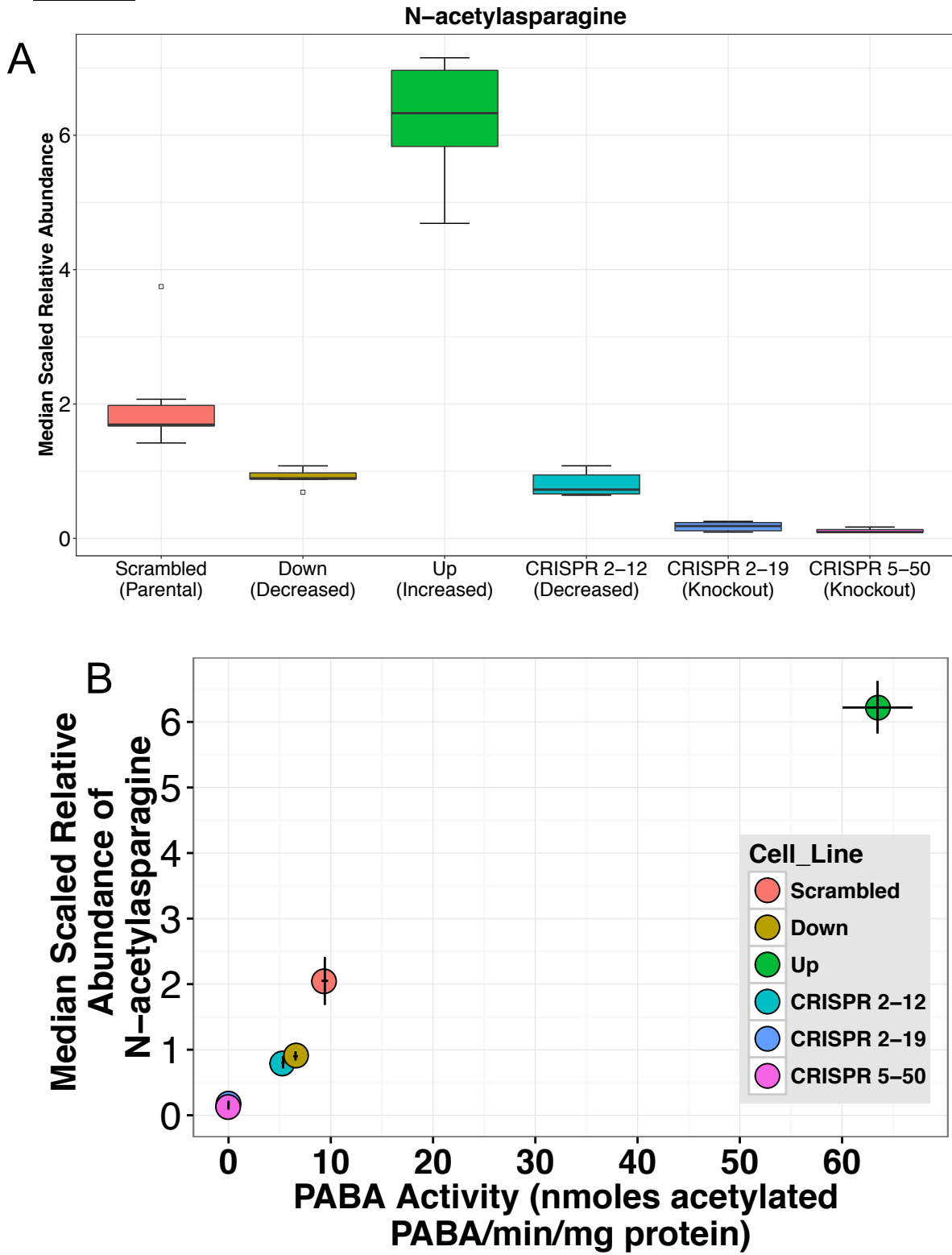


Figure 4.9: *N*-acetylasparagine Abundance Distribution and Correlation with NAT1 *N*-acetylation Activity.

(A) Boxplots of abundance distribution of *N*-acetylasparagine in each cell line. In the boxplots, the solid black line represents the median, the upper hinge represents the 75th quartile and the lower hinge represents the 25th quartile. The upper whisker represents the largest observation less than or equal to the upper hinge + 1.5 x IQR, the lower whisker represents the smallest observation greater than or equal to the lower hinge - 1.5 * IQR. (B) Correlation plot between *N*-acetylasparagine abundance and NAT1 *N*-acetylation activity in each cell line. Error bars represent standard deviation. Pearson correlation coefficient = 0.986, *p*-value < 0.001

Figure 4.10

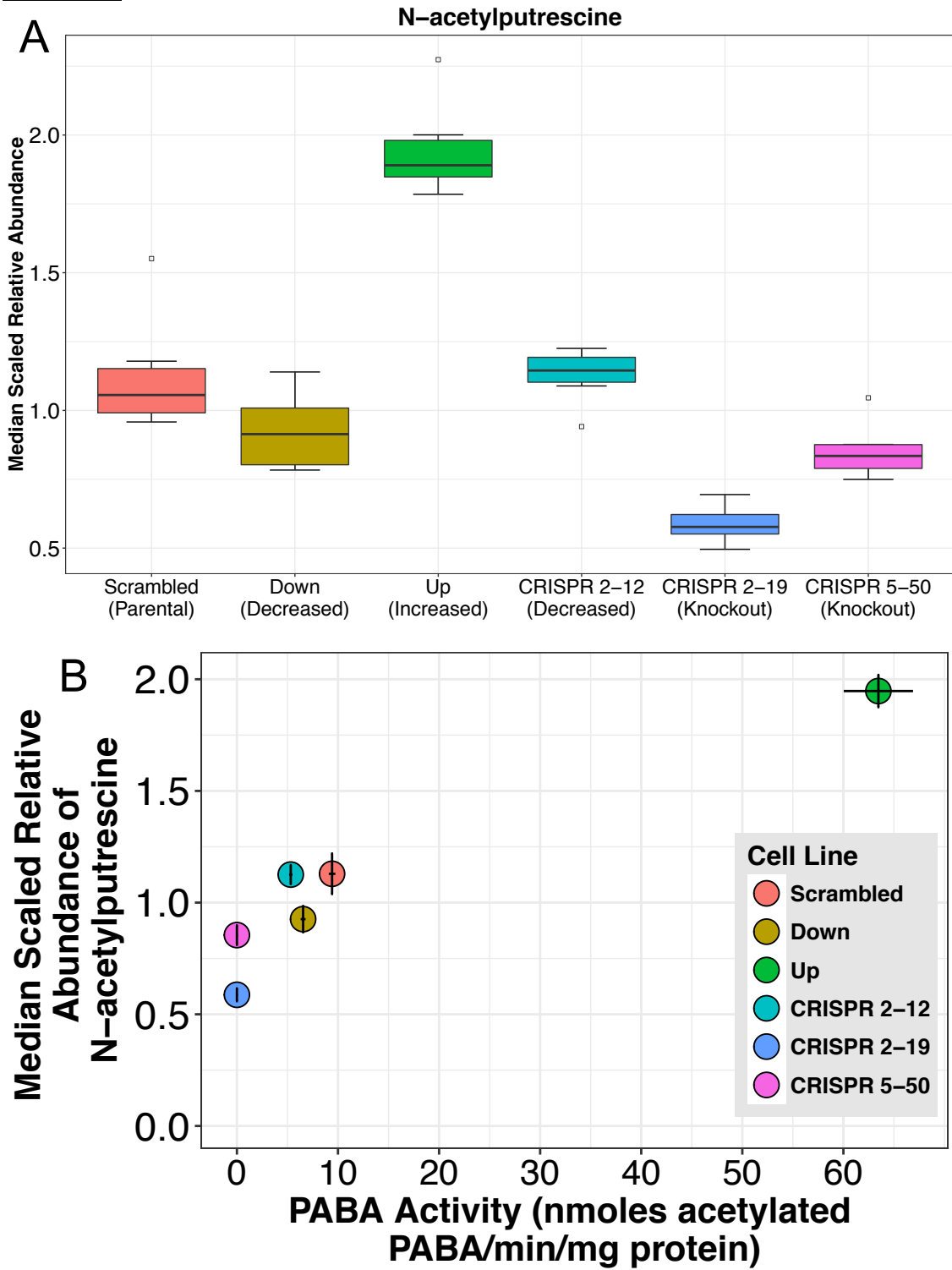


Figure 4.10: *N*-acetylputrescine Abundance Distribution and Correlation with NAT1 *N*-acetylation Activity.

(A) Boxplots of abundance distribution of *N*-acetylputrescine in each cell line. In the boxplots, the solid black line represents the median, the upper hinge represents the 75th quartile and the lower hinge represents the 25th quartile. The upper whisker represents the largest observation less than or equal to the upper hinge + 1.5 x IQR, the lower whisker represents the smallest observation greater than or equal to the lower hinge - 1.5 * IQR. (B) Correlation plot between *N*-acetylputrescine abundance and NAT1 *N*-acetylation activity in each cell line. Error bars represent standard deviation. Pearson correlation coefficient = 0.944, *p*-value < 0.01

Figure 4.11

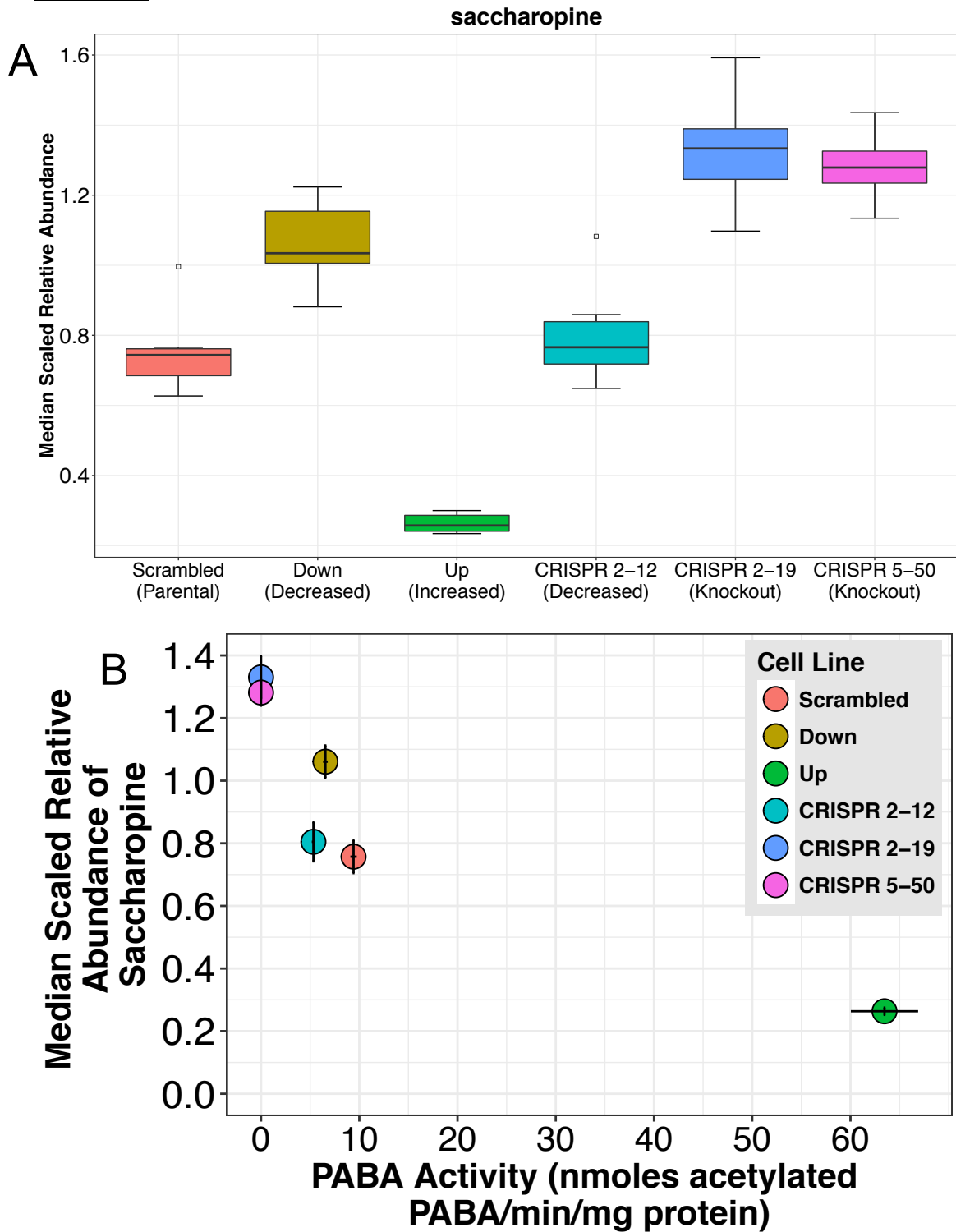


Figure 4.11: Saccharopine Abundance Distribution and Correlation with NAT1 N-acetylation Activity.

(A) Boxplots of abundance distribution of Saccharopine in each cell line. In the boxplots, the solid black line represents the median, the upper hinge represents the 75th quartile and the lower hinge represents the 25th quartile. The upper whisker represents the largest observation less than or equal to the upper hinge + 1.5 x IQR, the lower whisker represents the smallest observation greater than or equal to the lower hinge - 1.5 * IQR. (B) Correlation plot between Saccharopine abundance and NAT1 N-acetylation activity in each cell line. Error bars represent standard deviation. Pearson correlation coefficient = -0.876, p -value <0.05

Unsupervised hierarchical clustering of each sample revealed the global metabolomic profile of each cell line is distinct (Figure 4.12). The individual sample replicates clustered accurately by group except for sample 10 of the *Down* group that clustered with the *Scrambled* group. The first split in the dendrogram of the hierarchical clustering is between the two CRISPR/Cas9 cell lines constructed using guide RNA 2 and the other four cell lines; this provides evidence that those two cell lines have global metabolic profiles that are more similar to each other than to the respective cell lines that express the same level of NAT1 *N*-acetylation activity. The heatmap visualization of the data shows distinct clusters of metabolites whose relative abundance is much more similar between the two cell lines constructed using CRISPR/Cas9 guide RNA 2 but have different levels of NAT1 activity than the two CRISPR/Cas9 cell lines that were constructed using two different guide RNAs but both had no detectable NAT1 activity.

Similarly, principal component analysis showed that the CRISPR/Cas9 generated cell lines had global metabolomic profiles that were distinct from the siRNA generated cell lines as well as each other. In our dataset, principal component 1 explains 53% of the variance in the data while principal component 2 explains 14% of the variance (Figure 4.13). There are two types of variance within the data set; within group and between group variance. For our experimental question, we are most interested in the between group variance. The *CRISPR 2-12* and *CRISPR 2-19* groups are separated from the other four groups by principal component 1. This reveals that these two groups have global metabolomics profiles that are similar to each other but very different from the other four groups given that PC1 represents 53% of the variance in our dataset. The *CRISPR 5-50* group is separated from the other five groups along principal component 2.

From the loadings of each principal component we can infer which metabolites, together, contribute the most to the separation between the groups and thus the variance between the groups (Figure 4.14). Most metabolites in the dataset are not contributing to the variance observed. There are groups of approximately 5-15 metabolites that are contributing the greatest to each principle component (in each direction). Adenosine and carnosine related metabolites are negatively correlated with principal component 1 while phosphate related metabolites are

Figure 4.12

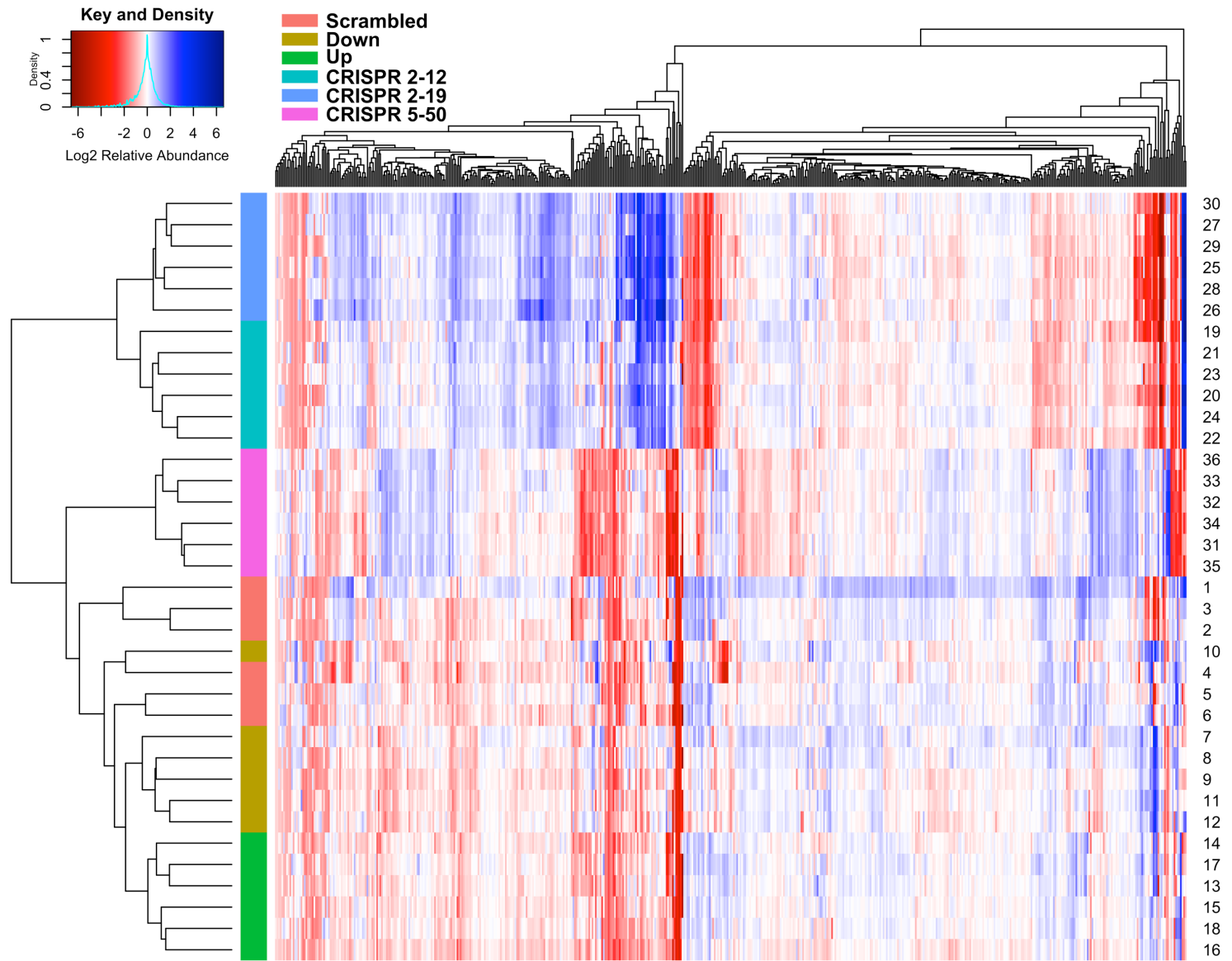


Figure 4.12: Metabolomics Heatmap and Hierarchical Clustering.

Metabolites colored red on the heatmap had a median scaled relative abundance less than 1, metabolites colored white had a median scaled relative abundance of 1, and metabolites colored blue had a median scaled relative abundance greater than 1. Each column represents a single metabolite and each row represents a single metabolomics sample. Samples are color coded according to cell line identity. Unbiased hierarchical clustering reveals the two CRISPR/Cas9 cell lines constructed using guide RNA 2 are more similar than the two NAT1 complete knockout cell lines, as would be expected.

Figure 4.13

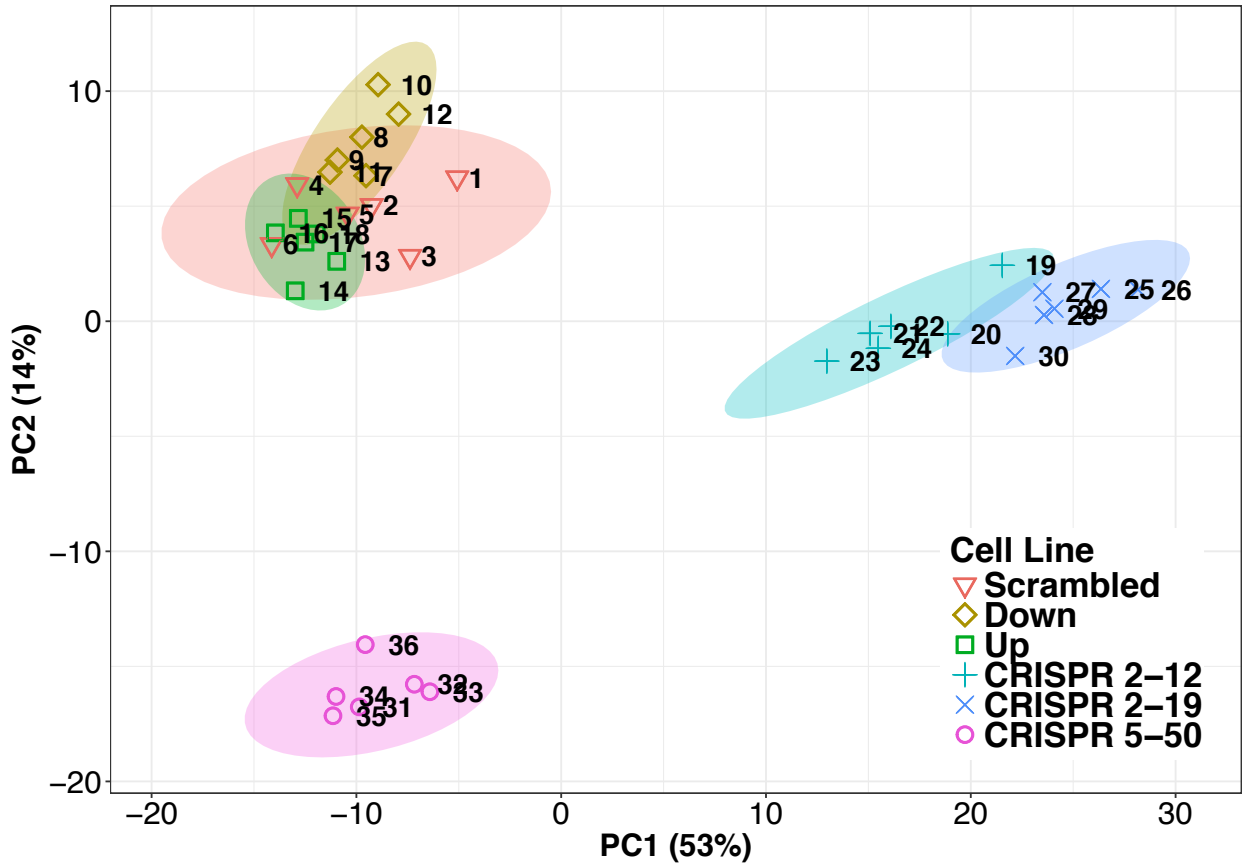


Figure 4.13: Principal Component Analysis Scores Plot.

Each symbol represents an individual metabolomics sample, is numbered by sample number, and color coded by cell line. Principal component one (PC1) represents 53% of the total variance in our dataset and separates the *CRISPR 2-12* and *CRISPR 2-19* cell lines from all other cell lines. Principal component 2 (PC2) represents 14% of the total variance in our dataset and separates the *CRISPR 5-50* cell line from all other cell lines.

Figure 4.14

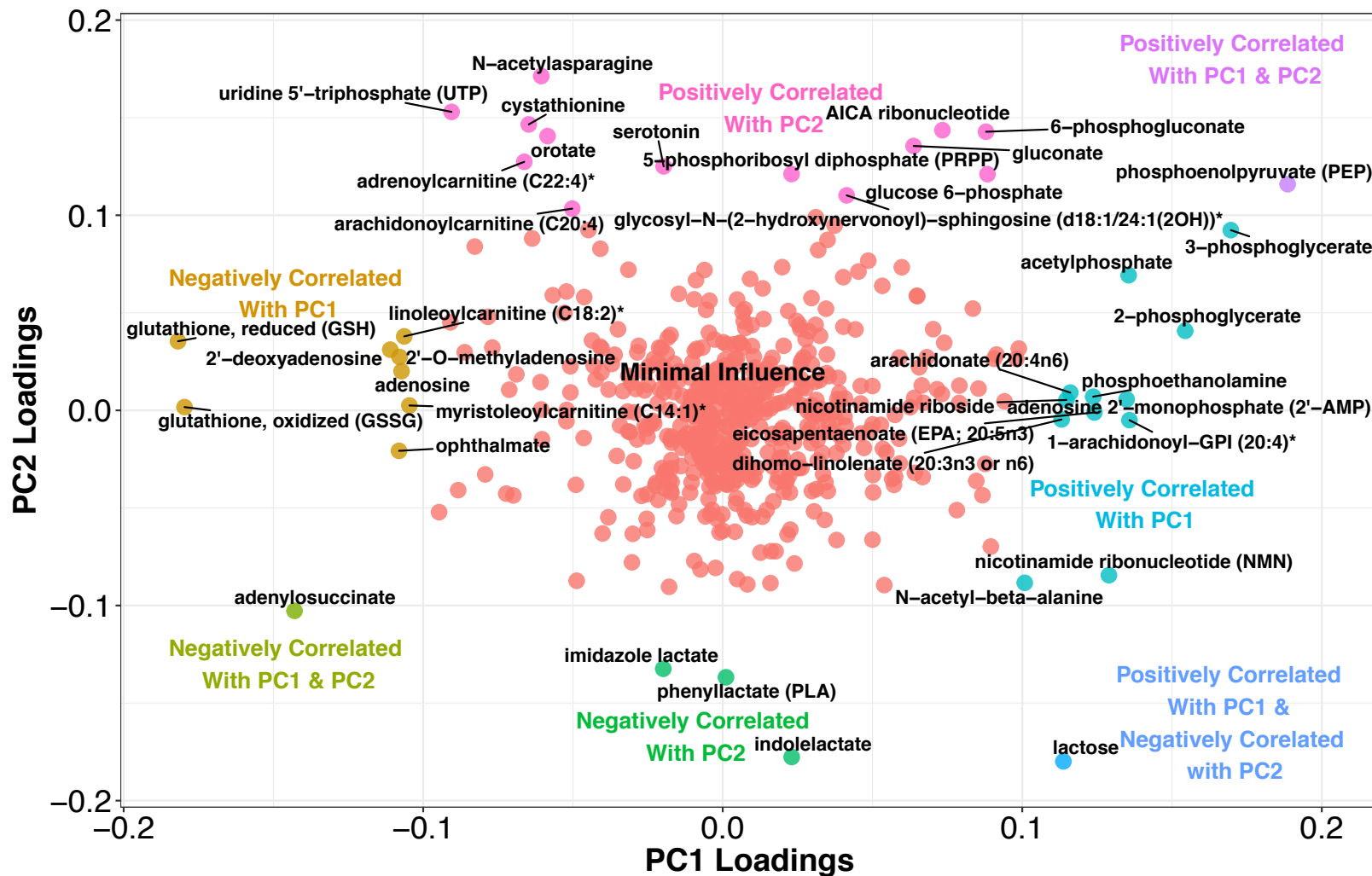


Figure 4.14: Principal Component Loadings Plot.

Loadings from principal component 1 and principal component 2 are plotted showing which metabolites have the greatest contribution to each principal component. Each point represents a single metabolite and is color-coded by contribution to principal component.

positively correlated. Lactate related metabolites are negatively correlated with principal component 2 while phosphate and carnitine related metabolites are positively correlated.

Pathway Analysis

Pathway enrichment analysis was conducted on each on each group compared to the *Scrambled* group (Figure 4.15). A normalized enrichment score of 1.2 in at least one of the groups was selected as the cut-off value for significance. The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were used^{173,174}. Disease associated pathways were removed from the analysis results. Amino acid, lipid, and fatty acid metabolism pathways were found to be significantly enriched. Some enriched pathways did not include all group comparisons suggesting differential impacts on metabolism.

Discussion

Given that (theoretically) only a single gene, NAT1, was genetically altered in each cell line, we expected only a small proportion of metabolite abundances to be significantly different, given the vast homeostasis mechanisms present¹⁷⁵⁻¹⁷⁷. However, we observed a very large proportion (~90%) of all metabolites detected to be significantly altered. Additionally, very few, if any, differences in metabolite abundances were expected between the two NAT1 knockout cell lines since each cell line should have the exact same genome. Yet one of the most striking observations of this study is the differences in relative metabolite abundance between the two complete NAT1 knockout cell lines. The hierarchical clustering, principal components analysis, and pathway enrichment analysis show there were significant differences between the two cell lines constructed using CRISPR/Cas9 guide RNA 2 and the cell line constructed using CRISPR/Cas9 guide RNA 5. Even though, in terms of NAT1 activity, the *CRISPR 2-19* and *CRISPR 5-50* cell lines are identical, their metabolic profiles are extremely different. This result suggests there are additional genetic differences between the two cell lines (discussed in greater detail in Chapter 7). However, we have not identified what genes the additional differences are in as whole genome sequencing of the cell lines would be required. The large number of differentially abundant metabolites between each knockout cell line compared to the *Scrambled* cell line and compared to each other could be because the cell lines have undergone additional

unique mutations during passaging or because each CRISPR guide RNA used caused unique off-target effects in addition to targeting NAT1. To focus on metabolome differences associated with varying levels of NAT1 in breast cancer, we have focused our interpretation of this study to metabolites that agreed between the two CRISPR NAT1 knockout cell lines.

Ten of 25 metabolites decreased concordantly between the two CRISPR NAT1 KO cell lines were metabolites containing carnitine, more specifically fatty acyl-coA carnitine conjugates (Table 4.2). To further clarify whether this observation was due to dysregulation of fatty acyl-coA's or carnitine in NAT1 KO cell lines, association between carnitine and metabolites decreased and increased concordantly in both NAT1 KO cell lines was determined. Nearly every metabolite concordantly differentially abundant showed strong correlation with carnitine suggesting the observation is due to dysregulation of carnitine. Carnitine is biosynthesized from lysine and methionine¹⁷⁸ with important biological roles in the transport of activated long-chain fatty acids from the cytosol to the mitochondrial matrix for beta-oxidation^{179,180}, modulation of the acyl-CoA/CoA ratio^{180,181}, and storage of energy as acetylcarnitine^{181,182}. Further studies are required to discern whether these observations are due to NAT1 directly or indirectly due to NAT1's effect on acetyl-CoA levels or other possible indirect effects.

Notably, NAT1 has been shown to have redundancy with methylthioribose-1-phosphate isomerase (MRI-1) in the methionine salvage pathway, but this role was not observed in MDA-MB-231 cells since they do not have a functional methionine salvage pathway due to a lack of methylthioadenosine phosphorylase (MTAP) activity¹⁰⁷. However, this observation by other NAT researchers adds interest to deciphering the relationship between NAT1 and carnitine since carnitine is closely connected biochemically to methionine. It will be important for future studies to focus on these relationships and pathways to gain a better understanding of the role of NAT1 in breast cancer cell metabolism.

The strong positive association between the abundance of *N*-acetylasparagine and NAT1 activity in the 6 MDA-MB-231 cell lines is consistent with *N*-acetylasparagine being a product of NAT1 *N*-acetylation (Figure 4.16). There is currently no known mechanism or enzyme responsible for the acetylation of free amino acids. It is hypothesized that acetylated free amino

Figure 4.15

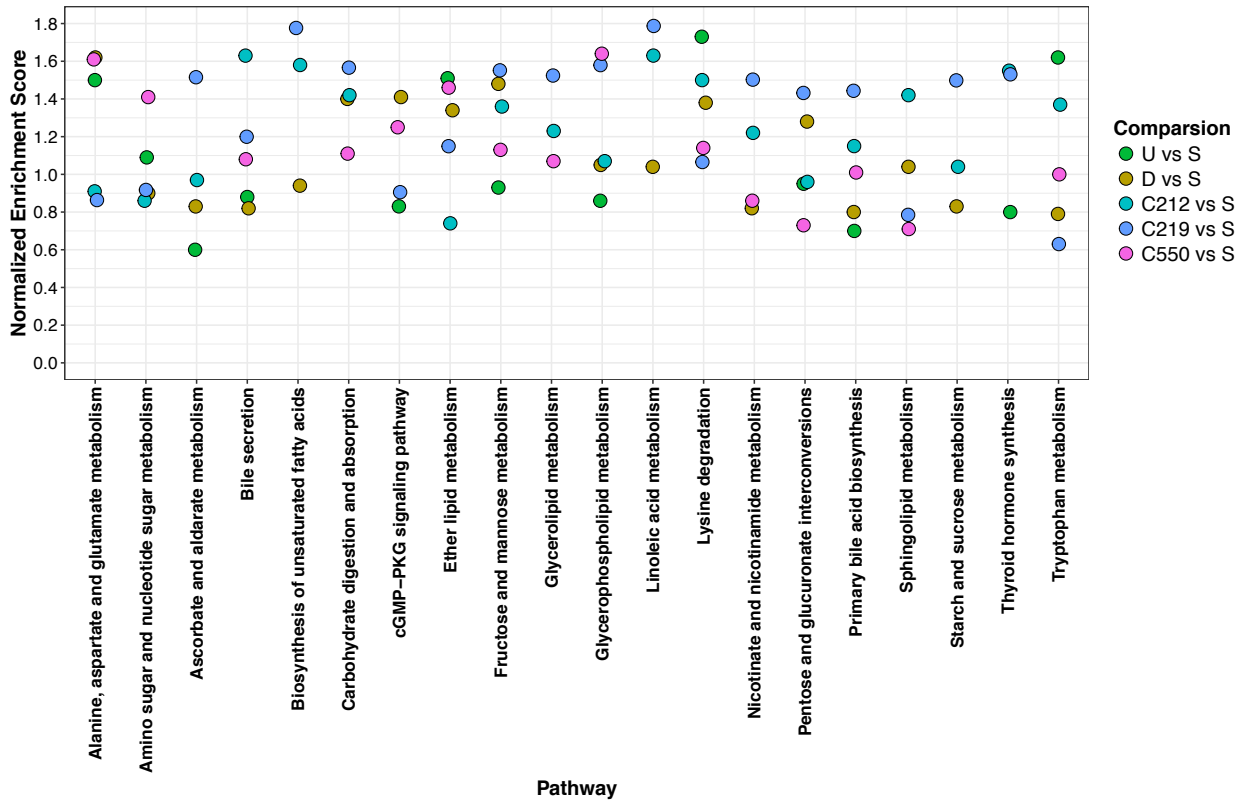


Figure 4.15: Metabolomics Pathway Enrichment Analysis.

Pathway enrichment analysis was conducted for each group compared to *Scrambled* and is color-coded by comparison. We utilized the normalized enrichment score to determine the relative degree of enrichment. U vs S: *Up vs Scrambled*; D vs S: *Down vs Scrambled*; C212 vs S: *CRISPR 2-12 vs Scrambled*; C219 vs S: *CRISPR 2-19 vs Scrambled*; C550 vs S: *CRISPR 5-50 vs Scrambled*.

Figure 4.16

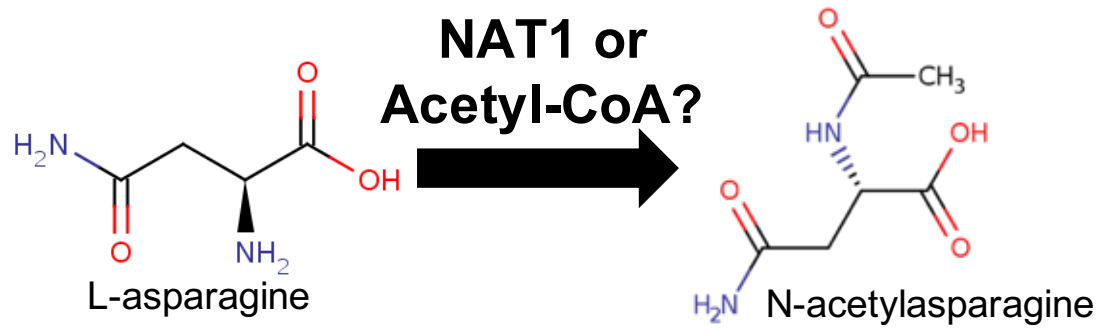


Figure 4.16: Possible Association Between NAT1 and N-acetylasparagine.

The positive relationship between NAT1 *N*-acetylation activity and the abundance of *N*-acetylasparagine suggests either L-asparagine is a NAT1 substrate or acetyl-CoA levels affect the activity of an enzyme for which L-asparagine is a substrate.

acids originate from the breakdown of *N*-terminal acetylated proteins however L-asparagine residues in proteins are not known to be acetylated by *N*-terminal acetyltransferases (different family of *N*-acetyltransferases from NAT1 and NAT2)¹⁸³. Additionally, the biosynthesis of *N*-acetylasparagine is not defined in the literature. Recombinant murine and human aminoacylase 2 (ASPA) have been reported to catabolize *N*-acetylasparagine, albeit at a much lower level than the prototypic ASPA substrate *N*-acetylaspartate¹⁸⁴. Additionally, aminoacylase 1 deficiency (ACY1), a rare inborn error of metabolism disease, is diagnosed by increased *N*-acetylated amino acids in the urine, including *N*-acetylasparagine¹⁸⁵ suggesting ACY1 can also metabolize *N*-acetylasparagine. Given the abundance distribution of *N*-acetylasparagine in our constructed cell lines, we hypothesize L-asparagine is the substrate being *N*-acetylated by NAT1 to form *N*-acetylasparagine however the substrate could be a different biochemical.

Similarly, the strong positive association between the abundance of *N*-acetylputrescine and NAT1 *N*-acetylation activity in the 6 MDA-MB-231 cell lines is consistent with *N*-acetylputrescine being a product of NAT1 *N*-acetylation (Figure 4.17). Putrescine is known to be *N*-acetylated by Spermidine/Spermine N1-Acetyltransferase 1 (SAT1) and Spermidine/Spermine N1-Acetyltransferase 2 (SAT2) but with much lower affinity than for other SAT substrates such as spermidine¹⁸⁶. Redundancy in non-homologous metabolic enzymes has been shown to occur^{187,188} and we postulate NAT1 and the SATs may both be able to *N*-acetylate putrescine.

Both L-asparagine and putrescine have been implicated in the promotion of cell growth¹⁸⁹⁻¹⁹⁴. The data indicate cell lines with higher levels of NAT1 activity have increased amounts of *N*-acetylated asparagine and putrescine while cell lines with decreased and knocked-out NAT1 activity have decreased amounts of *N*-acetylated asparagine and putrescine. It is predicted this leads to decreased asparagine and putrescine in cell lines with high NAT1 and increased asparagine and putrescine in cell lines with decreased and knocked-out levels of NAT1 relative to cell lines with basal NAT1 activity. It has been shown that intracellular asparagine exchanges with extracellular amino acids to promote mTORC1 activation, protein and nucleotide synthesis and cell proliferation under normal growth (non-starvation) conditions¹⁹⁰. Additionally, it

Figure 4.17

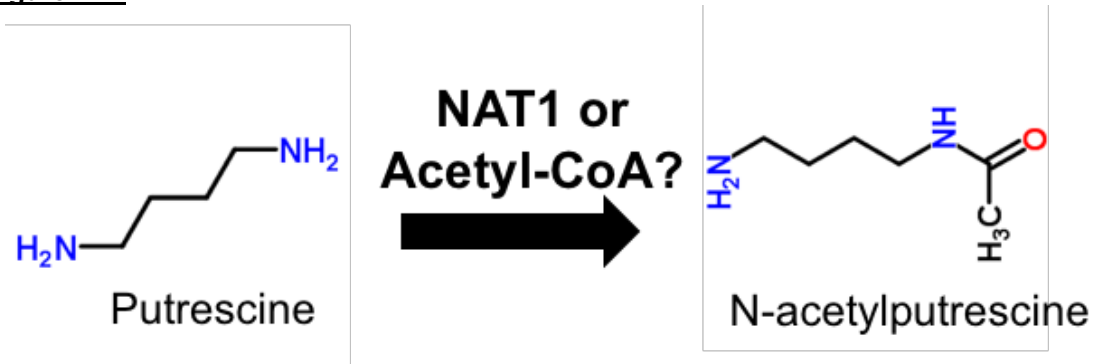


Figure 4.17: Possible Association Between NAT1 and N-acetylputrescine.

The positive relationship between NAT1 *N*-acetylation activity and the abundance of *N*-acetylputrescine suggests either putrescine is a NAT1 substrate or acetyl-CoA levels affect the activity of an enzyme for which putrescine is a substrate.

has been reported that the invasiveness of a mouse breast cancer model could be modulated either by altering asparagine biosynthetic capacity or by modifying extracellular asparagine pools with decreases in asparagine leading to decreased metastatic burden¹⁸⁹. Polyamines, such as putrescine, are known to facilitate the interactions of transcription factors, such as estrogen receptors with their specific response element and are involved in the proliferation of ER negative breast cancer tumor cells (reviewed in ¹⁹⁵). Additionally, rises in intracellular polyamine concentrations has been associated with increased cell proliferation and has been linked to tumorigenesis¹⁹⁶⁻²⁰⁰. This data, interpreted through the finding of previous studies, suggests cell lines with knocked-out NAT1 may have increased cell growth capabilities due to higher levels of asparagine and putrescine.

The strong negative association between the abundance of saccharopine and NAT1 activity in the 6 MDA-MB-231 cell lines suggests saccharopine is a NAT1 substrate or located upstream of a NAT1 catalyzed reaction (Figure 4.18). Saccharopine is an intermediate in the main pathway responsible for the catabolism of lysine^{201,202}. This observation may be connected to lysine's role in carnitine biosynthesis¹⁷⁸.

Pathway analysis indicated pathways that directly involve acetyl-coA such as the lysine degradation (Figure 4.19) and tryptophan metabolism pathways were enriched for differences. Additionally, pathways that feed into the TCA cycle were also significantly enriched suggesting varying levels of NAT1 impact energy metabolism.

Summary and Conclusions

These results support the hypothesis that NAT1 is not just a xenobiotic metabolizing enzyme and may have a role in endogenous cellular metabolism whether that is directly or indirectly through the ability of NAT1 to metabolize acetyl-CoA remains unclear. The metabolomics data have shown NAT1 expression differentially affects cellular metabolism dependent on the level of expression. Additionally, potential novel substrates and products of *N*-acetylation by NAT1 have been identified but further validation is required. These metabolites have recently been implicated in enhanced cell growth and metastatic potential in breast cancer

Figure 4.18

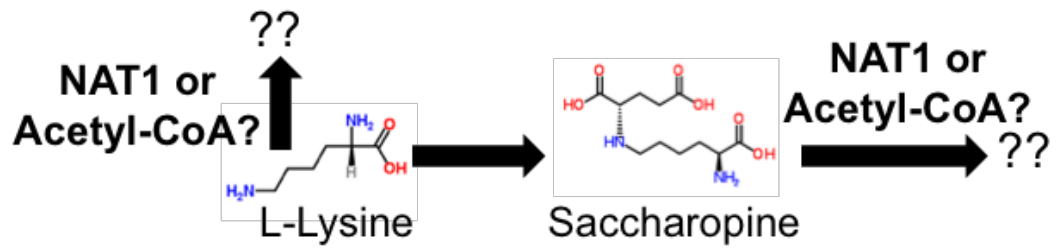


Figure 4.18: Possible Association Between NAT1 and Saccharopine.

The inverse relationship between NAT1 *N*-acetylation activity and the abundance of saccharopine suggests either saccharopine is a NAT1 substrate, acetyl-CoA levels affect the activity of an enzyme for which saccharopine is a substrate, lysine, the substrate that is degraded to form saccharopine, is a NAT1 substrate or acetyl-CoA levels affect the activity of an enzyme for which lysine is a substrate.

Figure 4.19

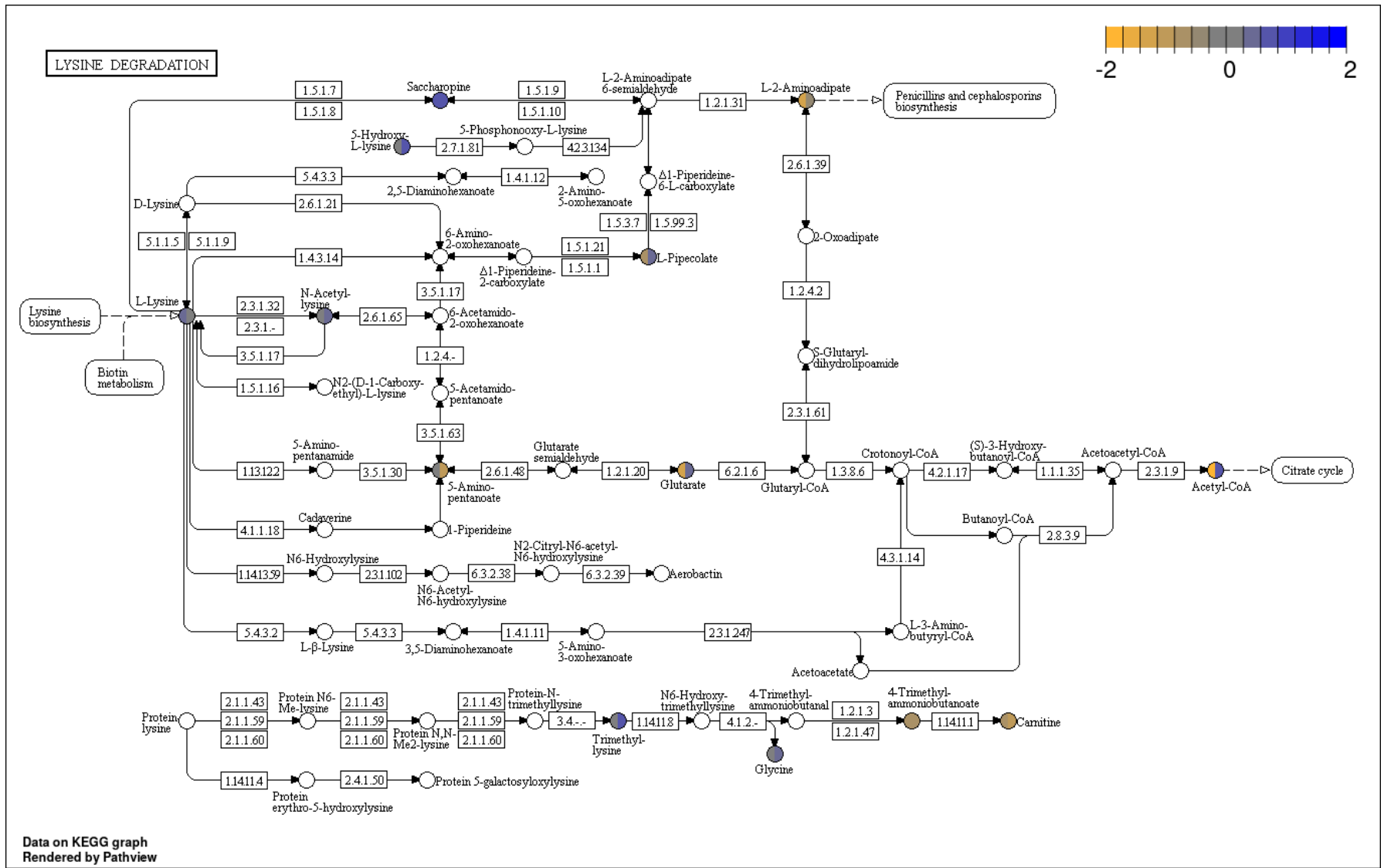


Figure 4.19: Lysine Degradation Pathway.

NAT1 knockout cell lines, *CRISPR 2-19* and *CRISPR 5-50*, metabolite abundance data compared to *Scrambled* cell line projected onto the KEGG lysine degradation pathway. Metabolites abundances in yellow are decreased in the NAT1 knockout cell lines, metabolites in gray are unchanged, and metabolites in blue are increased in the NAT1 knockout cell lines.

models suggesting they may be the key to understanding how varying levels of NAT1 affect breast cancer risk and progression. Additionally, many of the pathways significantly enriched in pathway analysis are pathways where acetyl-CoA plays a role, adding further evidence for the connection between NAT1 and acetyl-CoA in metabolism.

CHAPTER 5

TRANSCRIPTOMICS

Background

Transcriptomics allows measurement of global changes in gene abundance between samples and gives a snapshot of mRNA expression. Resulting information can be utilized to determine genes that may be interacting. While not all mRNAs will be translated into proteins and there are additional regulatory mechanisms that occur prior to protein expression, mRNA abundances can still provide a lot of information about a system. Additionally, numerous studies have utilized transcriptomics to better understand breast cancer²⁰³⁻²⁰⁷.

Materials and Methods

Collection of Samples

Figure 5.1 illustrates the experimental approach and workflow of sample collection described below. Cells were plated in triplicate per biological replicate at a concentration of 500,000 cells per 150 x 25 mm cell plate for both transcriptomics and metabolomics analysis. All cell lines were cultured in high-glucose Dulbecco's Modified Eagle Medium (DMEM), with 10% fetal bovine serum, 5% glutamine, and 5% penicillin/streptomycin added. Cells were allowed to grow for three days at 37°C and 5% CO₂ in an incubator.

Cells were then harvested on ice by adding 5 mL 0.25% trypsin and scraping the cells from the plate. Three plates were combined to form one sample (biological replicate). Three biological replicates for each cell line were collected for transcriptomics analysis. After harvesting the cells, cells were washed 3 times with ice-cold 1 x PBS. All supernatant was removed and 100 µL of cell pellet was reserved for transcriptomics analysis. The remaining cell pellet was reserved

Figure 5.1

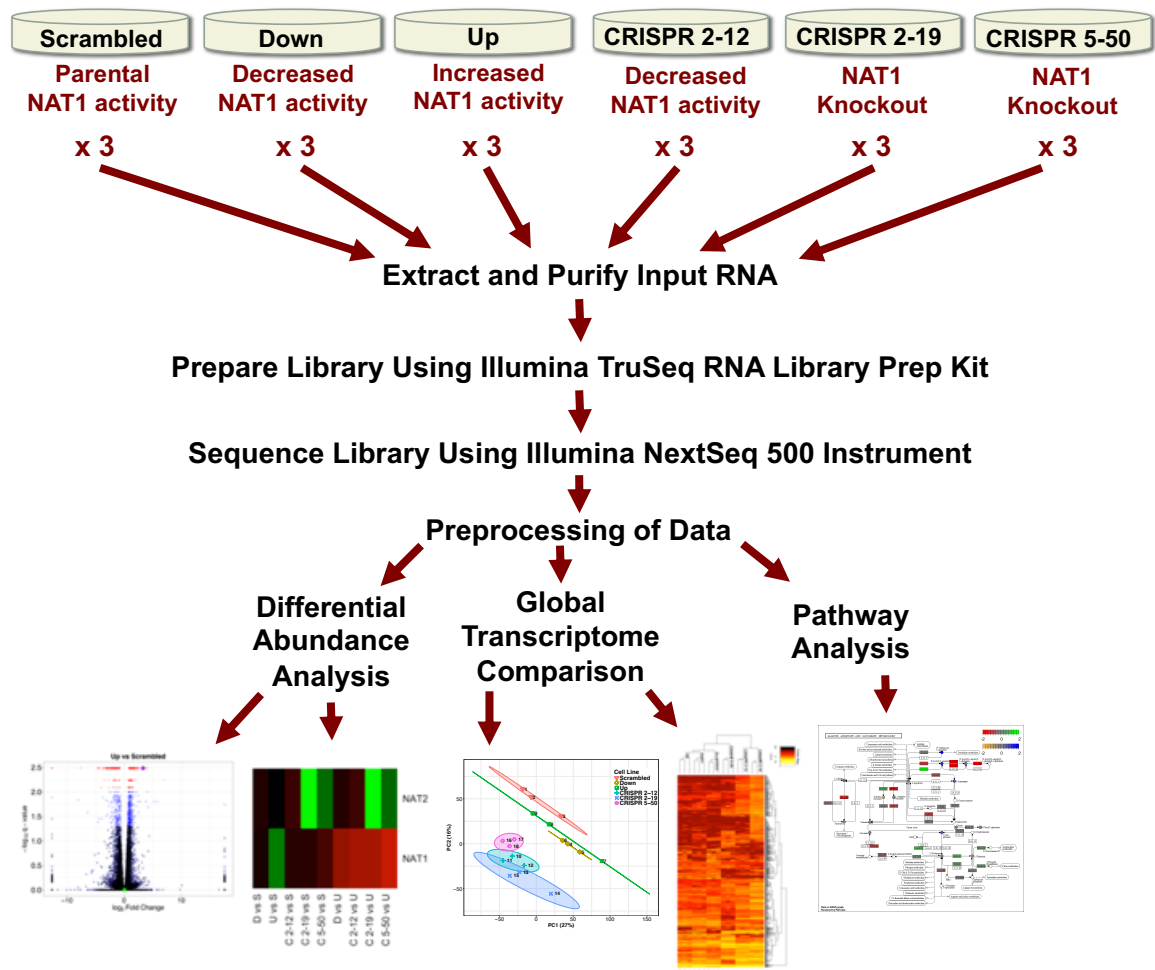


Figure 5.1: Transcriptomics Experimental and data analysis approach.

Three biological replicates from each cell line were collected. RNA was extracted and RNA-seq libraries prepared. Samples were then sequenced on the Illumina NextSeq 500 instrument. Following preprocessing of resulting data by the TopHat pipeline, genes were analyzed for differential abundance, unbiased multivariate analysis/clustering, and pathway enrichment. Notably, transcriptomics samples were the same samples collected for metabolomics analysis (discussed in Chapter 4).

for metabolomic analysis (presented in Chapter 4). Samples were then flash frozen by placing in a pool of liquid nitrogen for 1 minute and stored in an -80 °C freezer.

Transcriptomics

RNA Isolation

Total RNA was isolated from MDA-MB-231 breast cancer cells expressing parental (*Scrambled*), increased (*Up*), decreased (*Down*, *CRISPR 2-12*), and knockout (*CRISPR 2-19*, *CRISPR 5-50*) levels of NAT1 using the RNeasy® Mini Kit (Qiagen Sciences, Germantown, Maryland) according to manufacturer's instructions. RNA purity was evaluated, and concentrations were measured in each sample using a NanoDrop Bioanalyzer (Life Technologies Sciences).

Library Preparation

Libraries were prepared in collaboration with the University of Louisville Genomics Core (who kindly allowed me to be a part of the process) using the TruSeq Stranded mRNA LT Sample Prep Kit- Set A (Illumina, San Diego, California; Cat# RS-122-2101) with poly-A enrichment per manufacturer's instructions. One µg of total RNA (in a volume of 50 µL) from each sample was used in library preparation. Briefly, the total RNA was fragmented to improve sequence coverage over the transcriptome. Next, the first strand of cDNA was synthesized from the cleaved RNA fragments that were primed with random hexamers using reverse transcriptase and random primers. Then the second strand of cDNA was synthesized, thus providing double stranded blunt end cDNA. Next, a single 'A' nucleotide was added to the 3' ends of the blunt fragments to prevent them from ligating to one another during the adapter ligation reaction. A corresponding single 'T' nucleotide on the 3' end of the adapter provides a complementary overhang for ligating the adapter to the fragment. Then, multiple indexing adapters were ligated to the ends of the double stranded cDNA, preparing them for hybridization onto a flow cell. Next the DNA fragments were enriched using polymerase chain reaction (PCR) to selectively enrich those DNA fragments that have adapter molecules on both ends and to amplify the amount of DNA in the library. The PCR was performed with a PCR Primer Cocktail (included in the TruSeq Stranded mRNA LT Sample Prep Kit- Set A) that anneals to the ends of the adapters. Finally, 30 µl of eluted library

was collected and stored at -20 °C. To avoid skewing the representation of the library, the number of PCR cycles were minimized. This kit includes steps to validate and normalize constructed libraries and methods to check quality control.

Library Validation

Quality: Size, purity and semi quantitation were performed on an Agilent Bioanalyzer using the Agilent DNA 1000 Kit (Santa Clara, CA). The final fragment size for all the samples was approximately 300 basepairs as expected according to the protocol.

Quantity: Sequencing library quantitation was performed by quantitative polymerase chain reaction (qPCR) using the KAPA Library Quantitation Kit (Roche, Basel, Switzerland) for Illumina Platforms. A standard curve method was generated for quantitation using 1-5 DNA standards provided with the kit.

Normalize and Pool Libraries

Ten µl of sample was transferred from the wells to a new MIDI plate. The concentration of the libraries were then normalized to 10 nM using Tris-HCl 10 mM, pH 8.5 with 0.1% Tween 20. Five µl of each sample was then transferred to be pooled into a new LowBind 1.5 ml micro centrifuge tube for a total volume of 60 µl pooled 10 nM library. Then 4 nM dilution was made from the 10 nM pooled library by diluting with Tris-HCl 10 mM, pH 8.5 with 0.1% Tween 20.

Denaturing and diluting Libraries for the Nextseq 500

A total volume of 1.3 mL of 1.8 pM denatured library was needed for sequencing using the NextSeq 500/550 75 cycle High Output Kit v2 (FC-404-2005; Illumina, San Diego, CA) kit. Pooled 4 nM library was denatured by mixing with diluted NaOH and incubated at room temperature for 5 minutes. Tris HCl, 200 mM, pH 7.0 was then added. The reaction mixture was diluted to 20 pM using a pre-chilled Hybridization Buffer (included in the NextSeq 500/550 75 cycle High Output Kit v2). The 20 pM denatured library was further diluted to 1.8 pM using the same Hybridization Buffer. Before loading onto the reagent cartridge, 1.3 µl of denatured 20 pM Phix control (Illumina, San Diego, CA; FC-110-3001) was added to the 1299 µl of denatured 1.8 pM library to a total volume of 1.3 mL for the first sequencing run (for the 2nd sequencing run 1.9 pM library was used).

Sequencing

Sequencing was performed on the University of Louisville Center for Genetics and Molecular Medicine's (CGeMM) Illumina NextSeq 500 using the NextSeq 500/550 75 cycle High Output Kit v2 (FC-404-2005). A second run was performed to increase the number of reads. For each run, 72 single-end raw sequencing files (.fastq)²⁰⁸ representing six conditions with three biological replicates and four lanes per replicate were downloaded from Illumina's BaseSpace²⁰⁹ (<https://basespace.illumina.com/>) onto the Kentucky Biomedical Research Infrastructure Network (KBRIN) server for analysis using pre-written scripts.

RNA-Seq Analysis

Resulting data were preprocessed and analyzed using the tuxedo suite^{210,211} pipeline which includes TopHat, Cufflinks, Cuffmerge, and Cuffdiff. Cuffdiff was utilized to calculate differential gene expression for all groups (*Up*, *Down*, *CRISPR 2-12*, *CRISPR 2-19*, *CRISPR 5-50*) compared to the *Scrambled* group²¹¹. Quality control (QC) of the raw sequence data was performed using FastQC (version 0.10.1)²¹². Although the FastQC results indicate the last base of all samples has a lower quality value, we chose not to trim in this case. Concatenated sequences were directly aligned to the *Homo sapiens* hg38 reference genome assembly (hg38.fa) using TopHat2 (version 2.0.13)²¹³, generating alignment files in bam format. Table 5.1 indicates the number of raw reads successfully aligned for each of the samples.

Further analysis and visualization of the resulting data were performed using R: A Language and Environment for Statistical Computing version 3.3.1¹³². Fold-change and significance of differential gene expression between groups was visualized using volcano plots. The Pearson correlation coefficient was calculated between NAT1 and NAT2 transcript abundance and transcript abundance of all other genes to generate hypotheses about genes that may be interacting with NAT1 and NAT2, respectively. Data also were plotted as a heatmap and hierarchal clustering was conducted using the WPGMA method. Principal component analysis was conducted by singular value decomposition of the centered data matrix. The loadings of the first (x-axis) and second (y-axis) principal components were plotted. Weighted gene co-expression network analysis (WGCNA) was conducted on metabolite abundance data to identify

Table 5.1

Transcriptomics Sample Alignment

SAMPLE	INPUT READS	ALIGNED READS	PERCENT ALIGNED
Scrambled 1	37,984,684	36,841,298	97.0%
Scrambled 2	35,361,628	34,368,699	97.2%
Scrambled 3	38,238,715	36,597,025	95.7%
Down 1	41,290,001	40,114,301	97.2%
Down 2	41,339,117	40,184,597	97.2%
Down 3	44,859,187	43,503,707	97.0%
Up 1	38,199,308	37,143,906	97.0%
Up 2	37,337,039	36,125,963	96.8%
Up 3	39,022,651	37,890,392	97.1%
CRISPR 2-12 1	33,319,727	32,384,934	97.2%
CRISPR 2-12 2	38,610,686	37,399,482	96.9%
CRISPR 2-12 3	38,275,581	37,201,657	97.2%
CRISPR 2-19 1	38,870,999	37,718,655	97.0%
CRISPR 2-19 2	34,811,726	33,815,087	97.1%
CRISPR 2-19 3	34,097,370	33,138,856	97.2%
CRISPR 5-50 1	35,429,756	34,386,174	97.1%
CRISPR 5-50 2	42,535,551	41,226,671	96.9%
CRISPR 5-50 3	37,749,977	36,718,233	97.3%

The number of raw reads successfully aligned to the *homo sapiens* reference genome assembly hg.38 for each sample using TopHat2. Percent aligned was calculated by dividing aligned reads by input reads for each sample.

modules of related metabolites given the topology of the weighted network. Pathway enrichment analysis was conducted for each group compared to *Scrambled*. The normalized enrichment score was utilized to determine the relative degree of enrichment.

Quantitative Measurement of NAT1 and NAT2 mRNA

Reverse transcription- quantitative polymerase chain reaction (RT-qPCR) was conducted for NAT1 and NAT2 mRNAs in all 6 constructed cell lines as previously described^{40,96,214}. Briefly, total RNA was isolated from each cell line using the RNeasy Mini Kit (Qiagen, Germantown, MD). Isolated RNA was used to transcribe cDNA using the High Capacity Reverse Transcriptase kit (Life Technologies, Carlsbad, CA). Resulting cDNA was utilized for quantitative measurement of NAT1 and NAT2 mRNAs via RT-qPCR. TaqMan analysis was performed using the ABI 7700 sequence detection system (Applied Biosystems, Foster City, CA). Utilized gene probes were designed previously to discriminate between NAT1 and NAT2.

NAT2 N-Acetylation Activity Assays

In vitro NAT2 N-acetylation activity was determined in each constructed cell line as previously described^{12,215}. Briefly, cell lysate from each cell line was incubated with 1 mM acetyl-coenzyme A and 300 μ M sulfamethazine (SMZ) at 37°C for 10 minutes. Reactions were terminated with the addition of 1/10 reaction volume 1 M acetic acid. Reaction products were collected and analyzed using an Agilent Technologies 1260 Infinity high performance liquid chromatography (HPLC) using a LiChrospher® 100 RP-18 (5 μ m) column to determine the amount of acetylated product.

Results

Expression of many genes was found to significantly differ ($q \leq 0.05$, $|FC| > 2$) between the six cell lines (Table 5.2). More genes were differentially expressed in the cell lines with decreased or knockout NAT1 (*Down*, *CRISPR 2-12*, *CRISPR 2-19*, *CRISPR 5-50*) than the cell line with increased NAT1 activity (*Up*) when compared to the *Scrambled* cell line. Two thousand sixty-five, 544, 1911, 2383, and 1937 genes were differentially expressed in the *Down*, *Up*, *CRISPR 2-12*, *CRISPR 2-19*, and *CRISPR 5-50* groups, respectively (Table 5.2). Genes were further characterized by direction of fold-change compared to *Scrambled* (Figures 5.2 - 5.6). MSR1 gene

Table 5.2

Transcriptomics Summary Statistics

Student's <i>t</i> -test Post Test	Statistical Comparison				
	<u>Down</u> Scrambled	<u>Up</u> Scrambled	<u>CRISPR 2-12</u> Scrambled	<u>CRISPR 2-19</u> Scrambled	<u>CRISPR 5-50</u> Scrambled
Total DEGs $q \leq 0.05$; $FC \geq 2$	2065	544	1911	2383	1937
DEGs (\uparrow / \downarrow)	680 1385	310 234	1169 742	1472 911	1112 825

DEGs = differentially expressed genes

Cuffdiff was utilized to calculate differential gene expression for all groups (*Up*, *Down*, *CRISPR 2-12*, *CRISPR 2-19*, *CRISPR 5-50*) compared to the *Scrambled* group²¹¹. Significantly differentially expressed genes were further characterized by direction of fold-change; genes with increased abundance are shown in green while genes with decreased abundances are shown in red.

Figure 5.2

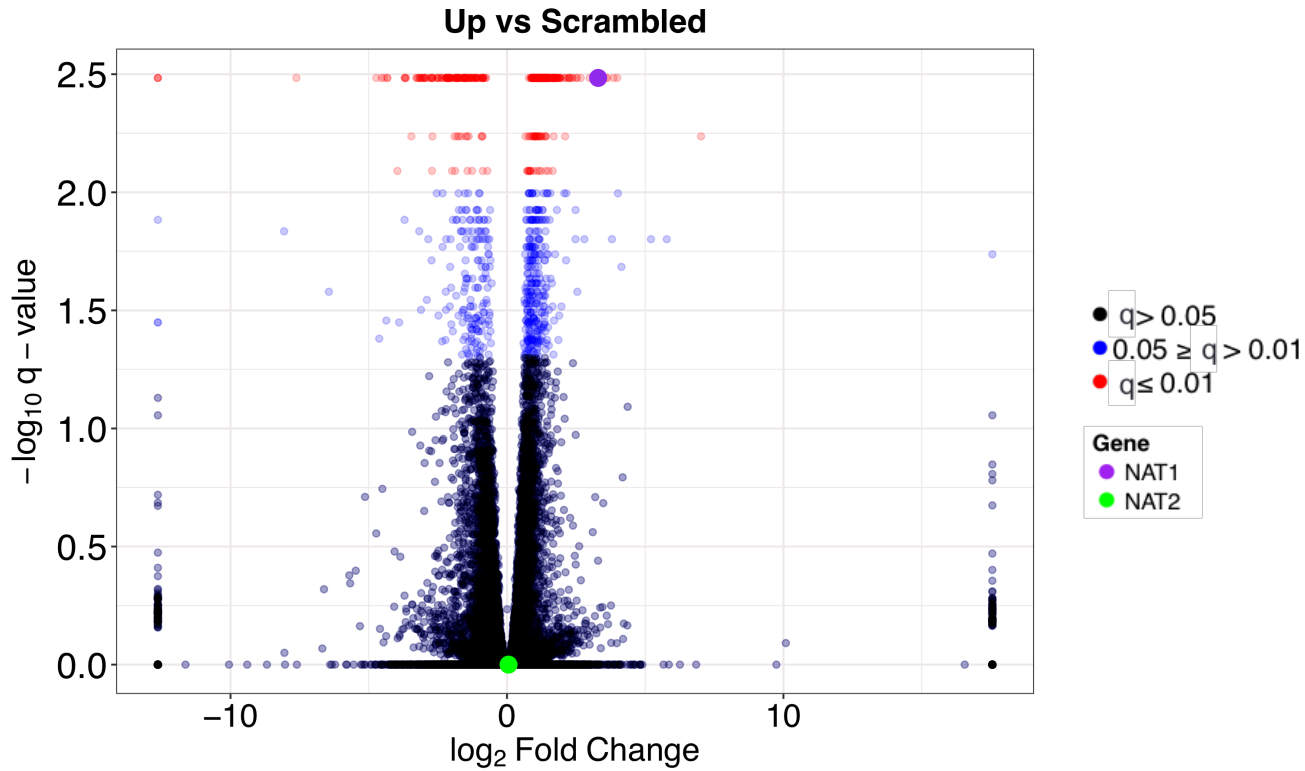


Figure 5.2: Up vs Scrambled Volcano Plot.

Each dot represents a single gene and is color coded according to q -value. The black dots represent genes that had a Dunnett's post test q -value greater than 0.05, blue dots represent genes that had a q -value less than or equal to 0.05 but greater than 0.01, and red dots represent genes that had a q -value less than or equal to 0.01. Negative fold changes represent a decrease in that gene compared to the *Scrambled* group while positive fold changes represent an increase in that gene compared to the *Scrambled* group. NAT1 is shown by a purple dot while NAT2 is shown by a green dot.

Figure 5.3

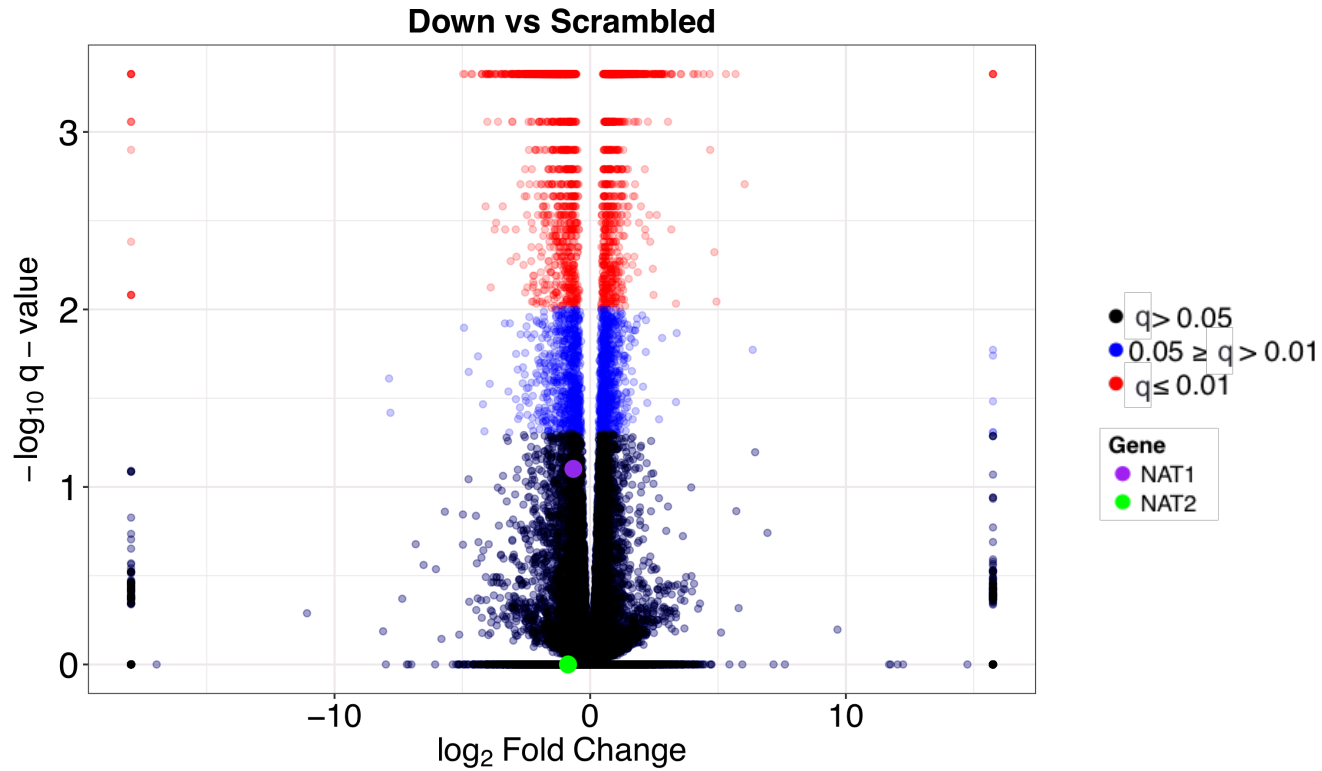
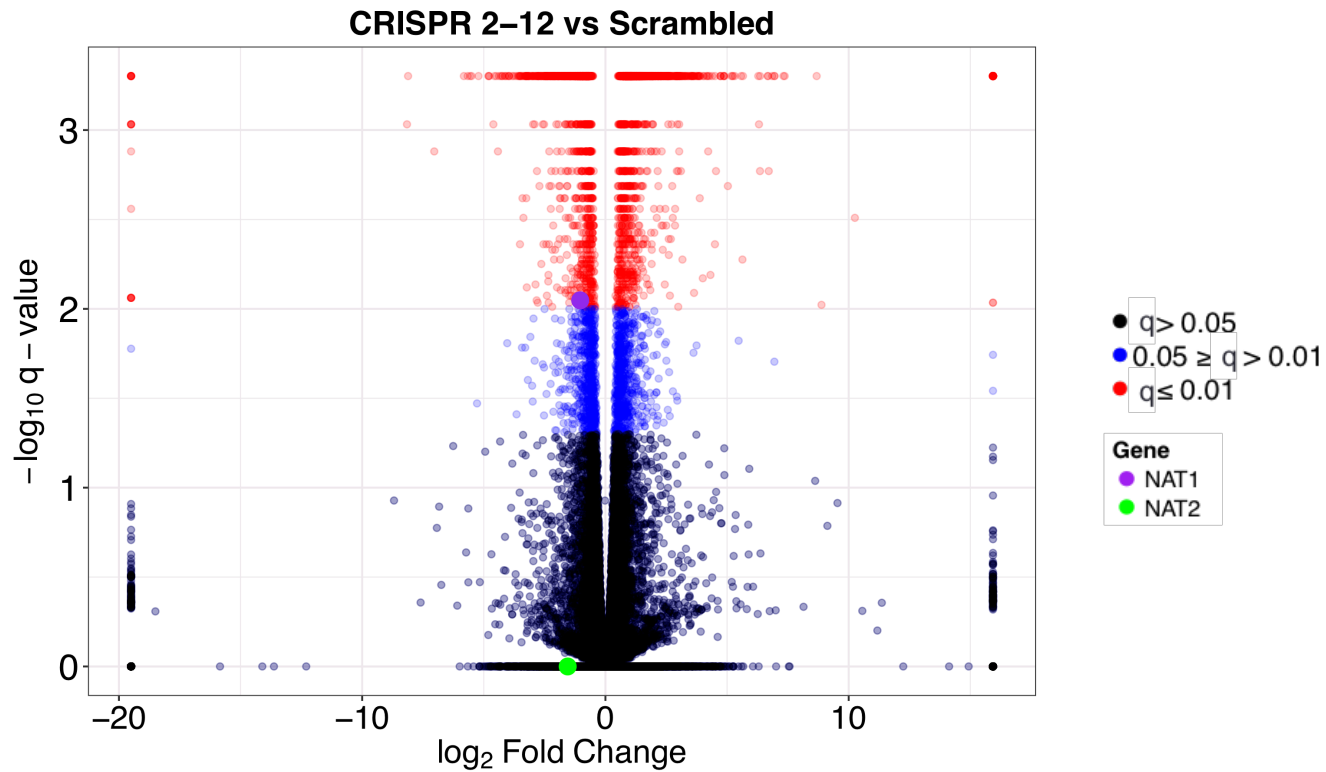


Figure 5.3: Down vs Scrambled Volcano Plot.

Each dot represents a single gene and is color coded according to q -value. The black dots represent genes that had a Dunnett's post test q -value greater than 0.05, blue dots represent genes that had a q -value less than or equal to 0.05 but greater than 0.01, and red dots represent genes that had a q -value less than or equal to 0.01. Negative fold changes represent a decrease in that gene compared to the *Scrambled* group while positive fold changes represent an increase in that gene compared to the *Scrambled* group. NAT1 is shown by a purple dot while NAT2 is shown by a green dot.

Figure 5.4**Figure 5.4: CRISPR 2-12 vs Scrambled Volcano Plot.**

Each dot represents a single gene and is color coded according to q -value. The black dots represent genes that had a Dunnett's post test q -value greater than 0.05, blue dots represent genes that had a q -value less than or equal to 0.05 but greater than 0.01, and red dots represent genes that had a q -value less than or equal to 0.01. Negative fold changes represent a decrease in that gene compared to the *Scrambled* group while positive fold changes represent an increase in that gene compared to the *Scrambled* group. NAT1 is shown by a purple dot while NAT2 is shown by a green dot.

Figure 5.5

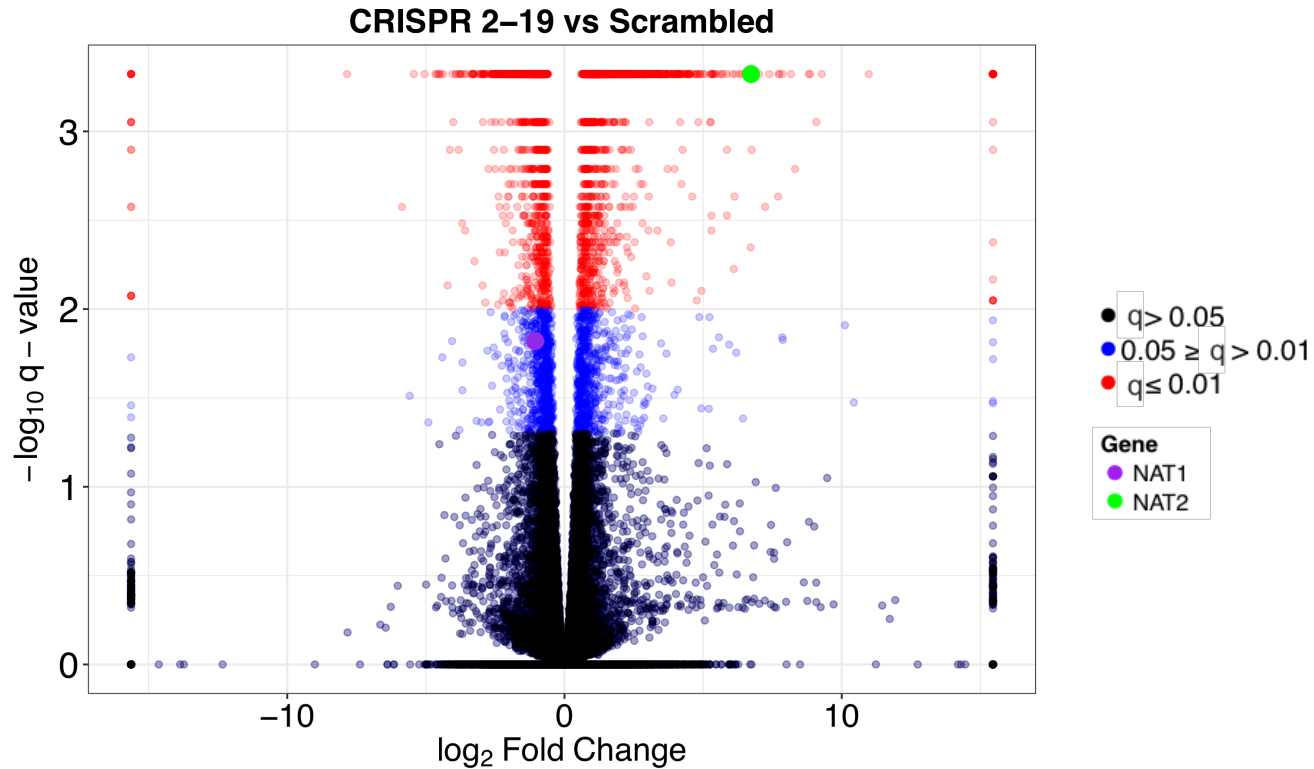


Figure 5.5: CRISPR 2-19 vs Scrambled Volcano Plot.

Each dot represents a single gene and is color coded according to q -value. The black dots represent genes that had a Dunnett's post test q -value greater than 0.05, blue dots represent genes that had a q -value less than or equal to 0.05 but greater than 0.01, and red dots represent genes that had a q -value less than or equal to 0.01. Negative fold changes represent a decrease in that gene compared to the *Scrambled* group while positive fold changes represent an increase in that gene compared to the *Scrambled* group. NAT1 is shown by a purple dot while NAT2 is shown by a green dot.

Figure 5.6

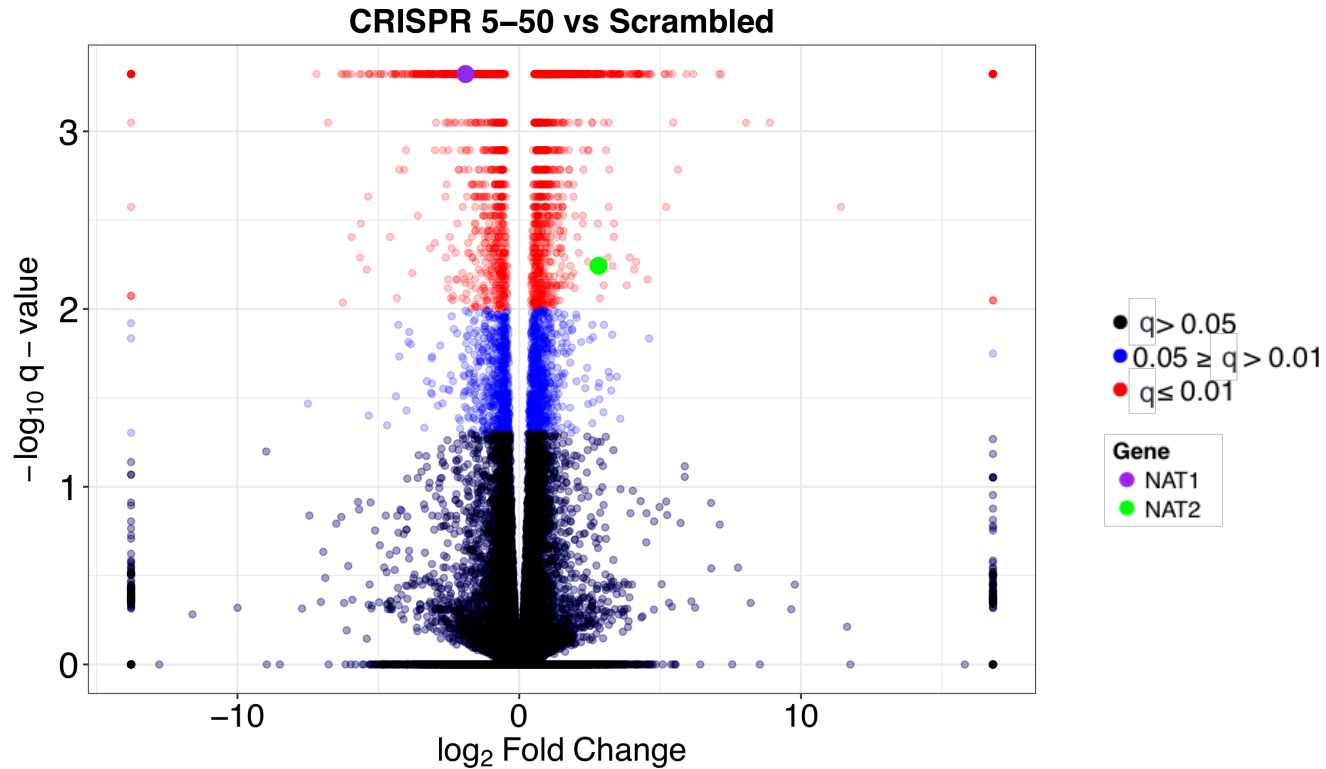


Figure 5.6: CRISPR 5-50 vs Scrambled Volcano Plot.

Each dot represents a single gene and is color coded according to q -value. The black dots represent genes that had a Dunnett's post test q -value greater than 0.05, blue dots represent genes that had a q -value less than or equal to 0.05 but greater than 0.01, and red dots represent genes that had a q -value less than or equal to 0.01. Negative fold changes represent a decrease in that gene compared to the *Scrambled* group while positive fold changes represent an increase in that gene compared to the *Scrambled* group. NAT1 is shown by a purple dot while NAT2 is shown by a green dot.

expression was compared between the cell lines since it has previously been reported that NAT1 has redundancy with the MSR1 enzyme (Table 5.3). However MSR1 was differentially expressed only between the *CRISPR 2-12* cell line and the *Scrambled* cell line.

The overlap in genes that were differentially expressed with a fold-change greater than or equal to 2 was compared between the two NAT1 KO cell lines and the *Scrambled* group (Figure 5.7). Five hundred ninety-nine genes were increased concordantly in the two NAT1 KO cell lines compared to *Scrambled* with 873 and 512 genes uniquely increased in the *CRISPR 2-19* and *CRISPR 5-50* cell lines, respectively. Three hundred nineteen genes were decreased concordantly in the two NAT1 KO cell lines compared to *Scrambled* with 592 and 506 genes uniquely decreased in the *CRISPR 2-19* and *CRISPR 5-50* cell lines, respectively. More genes had conflicting differential expression between the two CRISPR NAT1 KO cell lines compared to *Scrambled* than those that agreed.

Transcript expression of NAT1 and its isozyme, NAT2, were compared in all cell lines. Notably, in the two complete NAT1 knockout cell lines (*CRISPR 2-19* & *CRISPR 5-50*), NAT2 transcripts were significantly increased 6.7- and 2.8-fold, respectively, compared to the *Scrambled* cell line (Figure 5.8; Table 5.4), however NAT2 transcript expression was not increased in any of the other cell lines ($p > 0.05$; *Up, Down, CRISPR 2-12*). Individual NAT2 transcript mapping in the RNA-seq data in the *Scrambled* and NAT1 knockout cell lines, *CRISPR 2-19* and *CRISPR 5-50*, was visualized using Integrated Genomics Viewer (IGV; Figure 5.9). This observation was verified with RT-qPCR analysis for NAT1 and NAT2 in all six cell lines (Figure 5.10). Although, not significant for all cell lines, there was a trend of decreased NAT1 transcripts and increased NAT2 transcripts in the RT-qPCR data. NAT2 activity assays were conducted for all 6 constructed cell lines using the NAT2 specific substrate sulfamethazine (SMZ) however no NAT2 activity was detected in any cell line. Currently, the functionality of the NAT2 transcripts produced in the NAT1 knockout cell lines is unknown.

Many more genes correlated with NAT2 transcript abundance than NAT1 transcript abundance ($r \geq 0.9$). Twenty-two genes correlated with NAT1 (Table 5.5) while 342 genes were correlated with NAT2 (Table 5.6--abbreviated to show top 50 positively correlated and all

Table 5.3

Differential Expression of the Methylthioribose-1-Phosphate Isomerase (MRI1) Gene

COMPARISON	Log2 FC Comparison/Scrambled	q-value
<i>Up vs Scrambled</i>	0.39	0.93
<i>Down vs Scrambled</i>	-0.03	0.98
<i>CRISPR 2-12 vs Scrambled</i>	1.52	0.01
<i>CRISPR 2-19 vs Scrambled</i>	1	0.16
<i>CRISPR 5-50 vs Scrambled</i>	0.86	0.17

Methylthioribose-1-phosphate isomerase (MRI1) transcript fold-changes between each cell line compared to the *Scrambled* cell line. Fold-change and *q*-value were calculated using cuffdiff. Significant differential gene expression is color-coded by direction of fold-change with increases shown in green while decreases are shown in red.

Figure 5.7

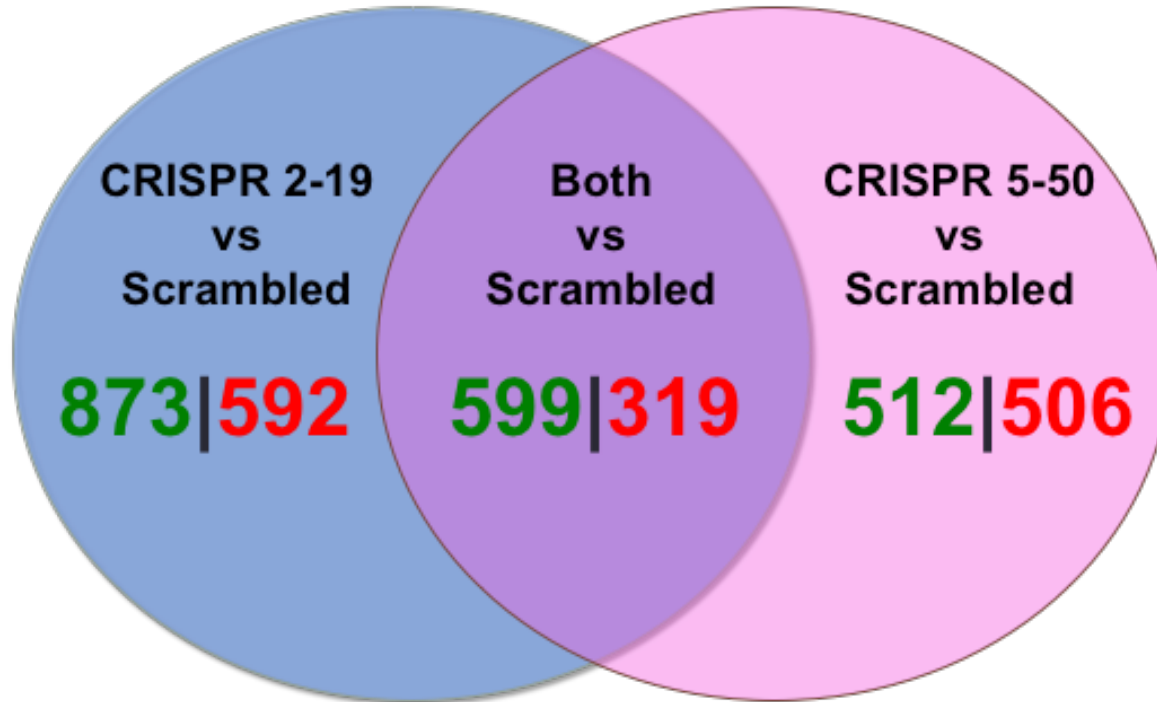


Figure 5.7: Gene Concordance Between CRISPR 2-19 and CRISPR 5-50 Cell Lines Compared to Scrambled.

Genes significantly differentially expressed with a fold-change greater than or equal to 2 were compared between CRISPR/Cas9 generated NAT1 knockout cell lines. Five hundred ninety-nine genes were increased concordantly in the two NAT1 KO cell lines compared to Scrambled with 873 and 512 genes uniquely increased in the *CRISPR 2-19* and *CRISPR 5-50* cell lines, respectively. Three hundred nineteen genes were decreased concordantly in the two NAT1 KO cell lines compared to *Scrambled* with 592 and 506 genes uniquely decreased in the *CRISPR 2-19* and *CRISPR 5-50* cell lines, respectively. Green: genes up-regulated compared to *Scrambled*; red: genes down-regulated compared to *Scrambled*.

Figure 5.8

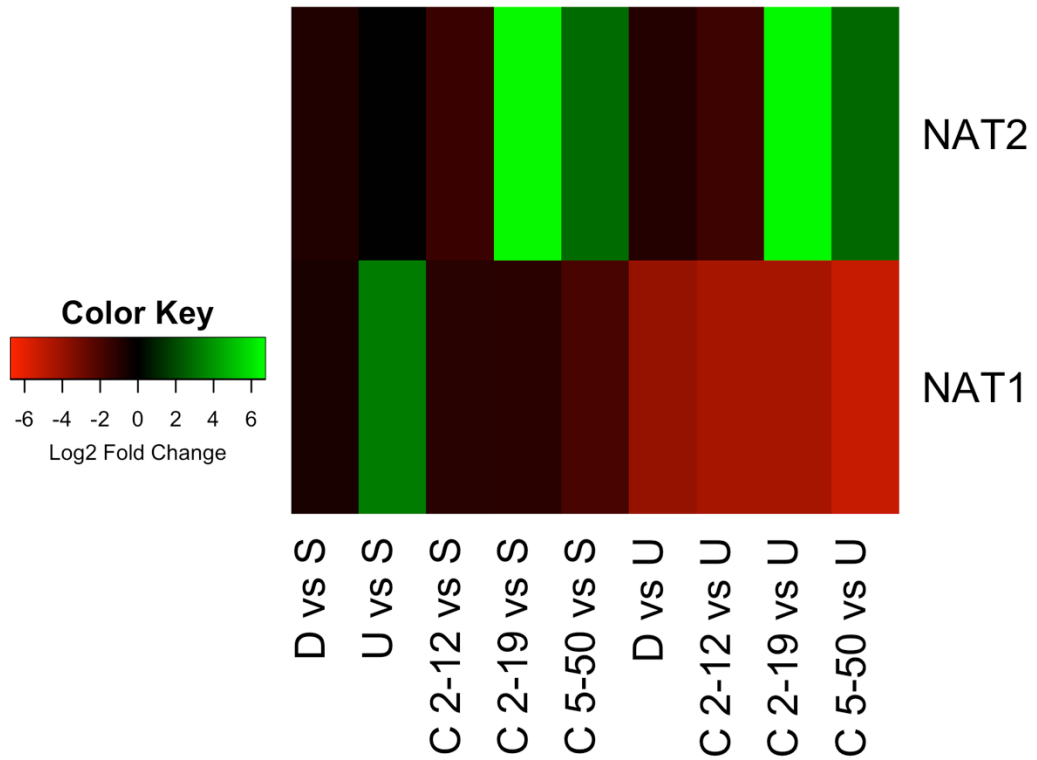


Figure 5.8: NAT1 and NAT2 Gene Expression in Each Cell Line Compared to Scrambled and Up.

Fold change of NAT1 and NAT2 between each cell line (*Down*, *Up*, *CRISPR 2-12*, *CRISPR 2-19*, and *CRISPR 5-50*) compared to *Scrambled* was assessed. We additionally compared NAT1 and NAT2 gene expression in all cell lines to the *Up* cell line. In the heatmap, red represents a decrease in gene expression, black represents no change, and green represents an increase in gene expression.

Table 5.4Differential Expression of the *N*-acetyltransferase (NAT) Genes

COMPARISON	GENE	Log2 FC Comparison/ <i>Scrambled</i>	<i>q</i> -value
<i>Up vs Scrambled</i>	NAT1	3.3	0.003
	NAT2	0.1	1
<i>Down vs Scrambled</i>	NAT1	-0.7	0.08
	NAT2	-0.9	1
<i>CRISPR 2-12 vs Scrambled</i>	NAT1	-1.0	0.009
	NAT2	-1.6	1
<i>CRISPR 2-19 vs Scrambled</i>	NAT1	-1.1	0.02
	NAT2	6.7	0.0005
<i>CRISPR 5-50 vs Scrambled</i>	NAT1	-1.9	0.0005
	NAT2	2.8	0.006

NAT transcript fold-changes between each cell line compared to the *Scrambled* cell line. Fold-change and *q*-value were calculated using cuffdiff.

Genes with $q \geq 0.05$ were considered significantly differentially expressed. Significant differential gene expression is color-coded by direction of fold-change with increases shown in green while decreases are shown in red.

Figure 5.9

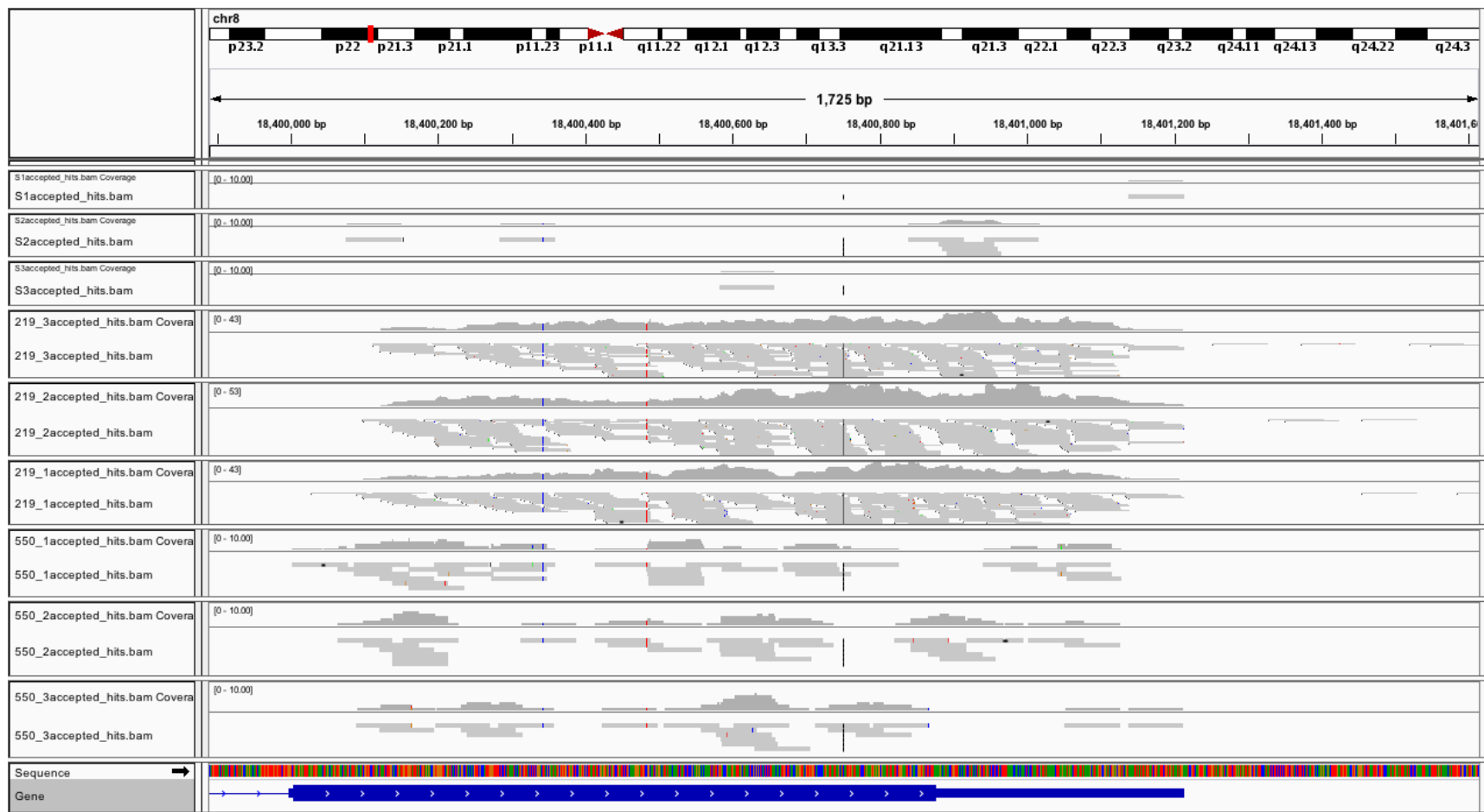


Figure 5.9: Integrated Genomics Viewer (IGV) NAT2 Transcript Mapping

Raw transcript reads mapping to NAT2 were visualized with IGV. Each gray bar represents a single transcript read. Colored lines in the transcripts represent the presence of a SNP in comparison to the *homo sapiens* hg.38 reference genome.

Figure 5.10

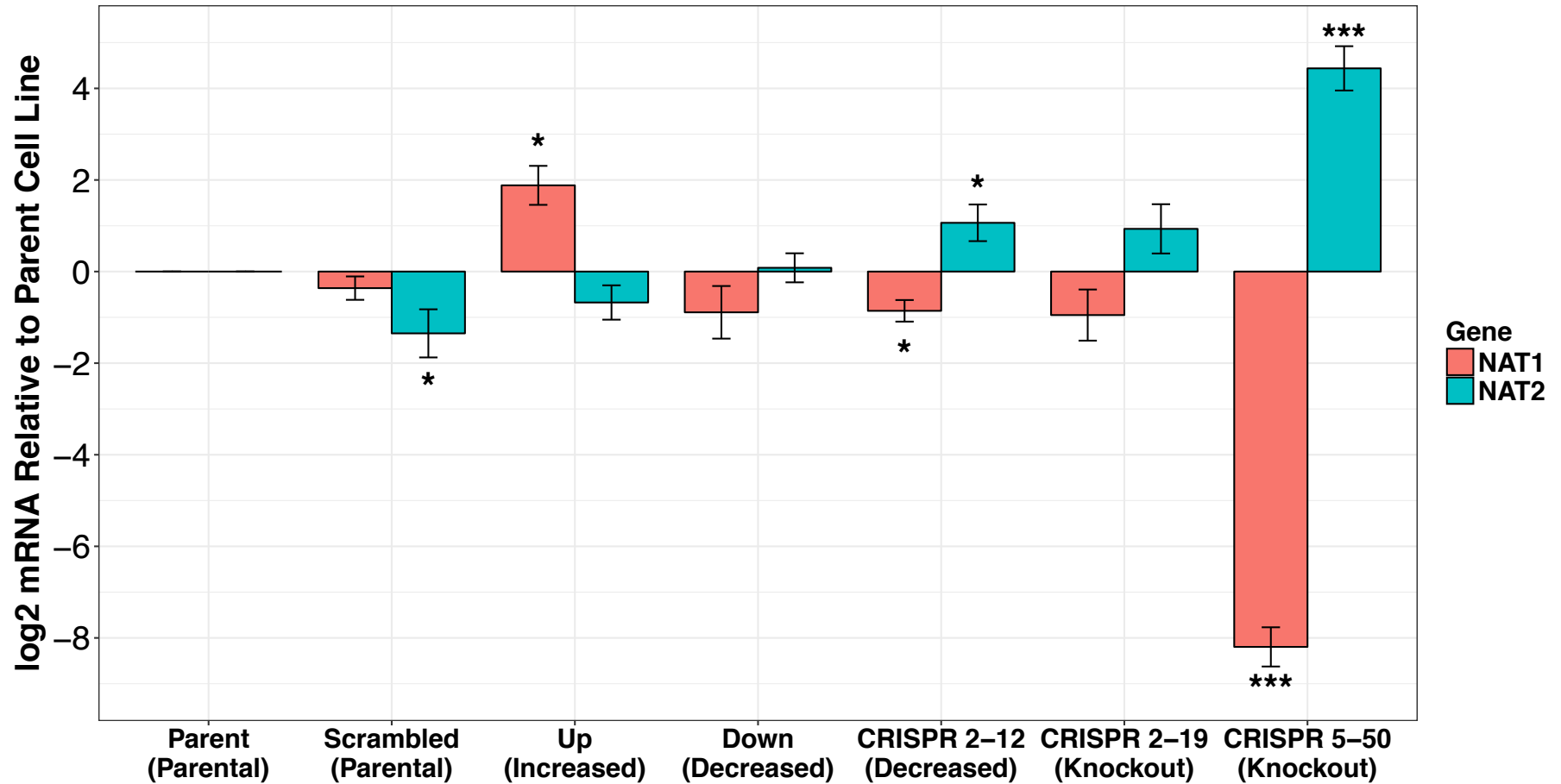


Figure 5.10: NAT1 and NAT2 mRNA Expression.

NAT1 and NAT2 were quantitated via RT-qPCR in the six constructed cell lines as well as the parental MDA-MB-231 cell line.

Results represent are from 3 independent experiments with 3 replicates in each experiment and are reported relative to the Parent cell line. Bars

represent log₂ mean ± SEM. N=3. *= $p < 0.05$, **= $p < 0.01$, ***= $p < 0.001$.

Table 5.5

Genes Correlated with NAT1 Transcript Abundance

ENSEMBLE GENE ID	GENE	q-value	CORRELATION
ENSG00000171428	NAT1	N/A	1.00
ENSG00000202310	Y_RNA	<0.00001	0.98
ENSG00000253377	AC068672.2	<0.00001	0.98
ENSG00000258168	AC025569.1	<0.00001	0.96
ENSG00000176571	CNBD1	<0.00001	0.96
ENSG00000174125	TLR1	<0.00001	0.95
ENSG00000167355	AC087380.1	<0.00001	0.94
ENSG00000165078	CPA6	<0.0001	0.93
ENSG00000187189	TSPYL4	<0.0001	0.92
ENSG00000173157	ADAMTS20	<0.0001	0.92
ENSG00000106034	CPED1	<0.0001	0.92
ENSG00000081148	IMPG2	<0.0001	0.92
ENSG00000165424	ZCCHC24	<0.0001	0.92
ENSG00000170915	PAQR8	<0.0001	0.92
ENSG00000162139	NEU3	0.0001	0.91
ENSG00000173930	SLCO4C1	0.0001	0.91
ENSG00000141219	C17orf80	0.0001	0.91
ENSG00000270157	NA	0.0001	0.91
ENSG00000179331	RAB39A	0.0002	0.91
ENSG00000196449	YRDC	<0.0001	-0.92
ENSG00000054116	TRAPPC3	0.0001	-0.91
ENSG00000116288	PARK7	0.0002	-0.90

Pearson correlation between NAT1 gene abundance in constructed cell lines and all genes in the transcriptomics dataset was conducted. Twenty-two genes were significantly correlated ($r > 0.90$) with NAT1 transcript abundance. q -value is the Pearson correlation q -value, and correlation is the Pearson correlation coefficient. Correlation coefficients were color coded according to direction of correlation with green for positive correlation and red for inverse (negative) correlation.

Table 5.6

Genes Correlated with NAT2 Transcript Abundance (Abbreviated to top 50 genes)

ENSEMBLE GENE ID	GENE	p-value	CORRELATION COEFFICIENT
ENSG00000156006	NAT2	NA	1.00
ENSG00000186479	RGS7BP	<0.00001	1.00
ENSG00000158270	COLEC12	<0.00001	1.00
ENSG00000021645	NRXN3	<0.00001	1.00
ENSG00000188086	PRSS45	<0.00001	1.00
ENSG00000179097	HTR1F	<0.00001	1.00
ENSG00000197506	SLC28A3	<0.00001	1.00
ENSG00000134247	PTGFRN	<0.00001	1.00
ENSG00000267313	KC6	<0.00001	1.00
ENSG00000112902	SEMA5A	<0.00001	1.00
ENSG00000243486	AC068985.1	<0.00001	1.00
ENSG00000129514	FOXA1	<0.00001	1.00
ENSG00000261122	LINC02167	<0.00001	1.00
ENSG00000233403	AC121342.1	<0.00001	1.00
ENSG00000078596	ITM2A	<0.00001	0.99
ENSG00000095777	MYO3A	<0.00001	0.99
ENSG00000105383	CD33	<0.00001	0.99
ENSG00000105851	PIK3CG	<0.00001	0.99
ENSG00000110002	VWA5A	<0.00001	0.99
ENSG00000115165	CYTIP	<0.00001	0.99
ENSG00000124429	POF1B	<0.00001	0.99
ENSG00000125848	FLRT3	<0.00001	0.99
ENSG00000130224	LRCH2	<0.00001	0.99
ENSG00000137745	MMP13	<0.00001	0.99
ENSG00000140379	BCL2A1	<0.00001	0.99
ENSG00000140563	MCTP2	<0.00001	0.99
ENSG00000143061	IGSF3	<0.00001	0.99
ENSG00000145649	GZMA	<0.00001	0.99
ENSG00000147246	HTR2C	<0.00001	0.99
ENSG00000149573	MPZL2	<0.00001	0.99
ENSG00000152217	SETBP1	<0.00001	0.99
ENSG00000155622	XAGE2	<0.00001	0.99

*continued on next page

ENSG00000165061	ZMAT4	<0.00001	0.99
ENSG00000165449	SLC16A9	<0.00001	0.99
ENSG00000165795	MIR6717	<0.00001	0.99
ENSG00000165868	HSPA12A	<0.00001	0.99
ENSG00000166669	ATF7IP2	<0.00001	0.99
ENSG00000170927	PKHD1	<0.00001	0.99
ENSG00000171004	HS6ST2	<0.00001	0.99
ENSG00000178235	SLITRK1	<0.00001	0.99
ENSG00000180287	PLD5	<0.00001	0.99
ENSG00000183230	CTNNA3	<0.00001	0.99
ENSG00000184564	SLITRK6	<0.00001	0.99
ENSG00000189398	OR7E12P	<0.00001	0.99
ENSG00000205628	LINC01446	<0.00001	0.99
ENSG00000223756	TSSC2	<0.00001	0.99
ENSG00000224960	PPP4R3CP	<0.00001	0.99
ENSG00000226372	DCAF8L1	<0.00001	0.99
ENSG00000255501	CARD18	<0.00001	0.99
ENSG00000261780	LOC100505817	<0.00001	0.99
ENSG00000238755	NA	<0.00001	0.99
ENSG00000253651	NA	<0.00001	0.99
ENSG00000228065	NA	<0.00001	0.99
ENSG00000151208	DLG5	<0.00001	-0.91

Pearson correlation between NAT2 gene abundance in constructed cell lines and all genes in the transcriptomics dataset was conducted. Three hundred forty-two genes were significantly correlated ($r > 0.90$) with NAT1 transcript abundance. q -value is the Pearson correlation q -value, and correlation is the Pearson correlation coefficient. Correlation coefficients were color coded according to direction of correlation with green for positive correlation and red for inverse (negative) correlation.

negatively correlated). This observation may be because of the low abundance of NAT2 in most of our cell lines. Notably, multiple genes associated with NAT1 transcript abundance are located on chromosome 8, the same chromosome on which NAT1 is located. There was no overlap in genes correlated with NAT1 and NAT2 transcript expression, suggesting different regulation mechanisms.

Correlation Concordance

To determine if the genes strongly correlated with NAT1 and NAT2 transcript expression, respectively, show the same relationship in other independent datasets correlation was compared in publicly available data repositories. Gene expression data for top 20 genes in the lists as well as NAT1 and NAT2 were downloaded from the cancer cell line encyclopedia (CCLE) and the cancer genome atlas (TCGA). Breast cancer datasets were utilized since our original samples were from the breast cancer cell line MDA-MB-231. However, there are data for many different cancers in the repositories. Correlation between NAT1 and NAT2 and the genes they correlated with in the transcriptomics dataset was assessed in the two independent datasets using the Pearson method.

Unsupervised hierarchical clustering of each sample revealed the global transcriptomic profile of each cell line is distinct (Figure 5.11). The individual sample replicates clustered accurately by group. The first split in the dendrogram of the hierarchical clustering is between the two CRISPR/Cas9 cell lines constructed using guide RNA 2 and the other four cell lines; this provides evidence that those two cell lines have global transcriptomic profiles that are more similar to each other than to the respective cell lines that express the same level of NAT1 *N*-acetylation activity. The heatmap visualization of the data shows distinct clusters of gene transcripts whose expression is much more similar between the two cell lines constructed using CRISPR/Cas9 guide RNA 2 but have different levels of NAT1 activity than the two CRISPR/Cas9 cell lines that were constructed using two different guide RNAs but both had no detectable NAT1 activity. Differential gene expression between each cell line compared to *Scrambled* was also visualized as a heatmap with hierarchical clustering (Figure 5.12).

Figure 5.11

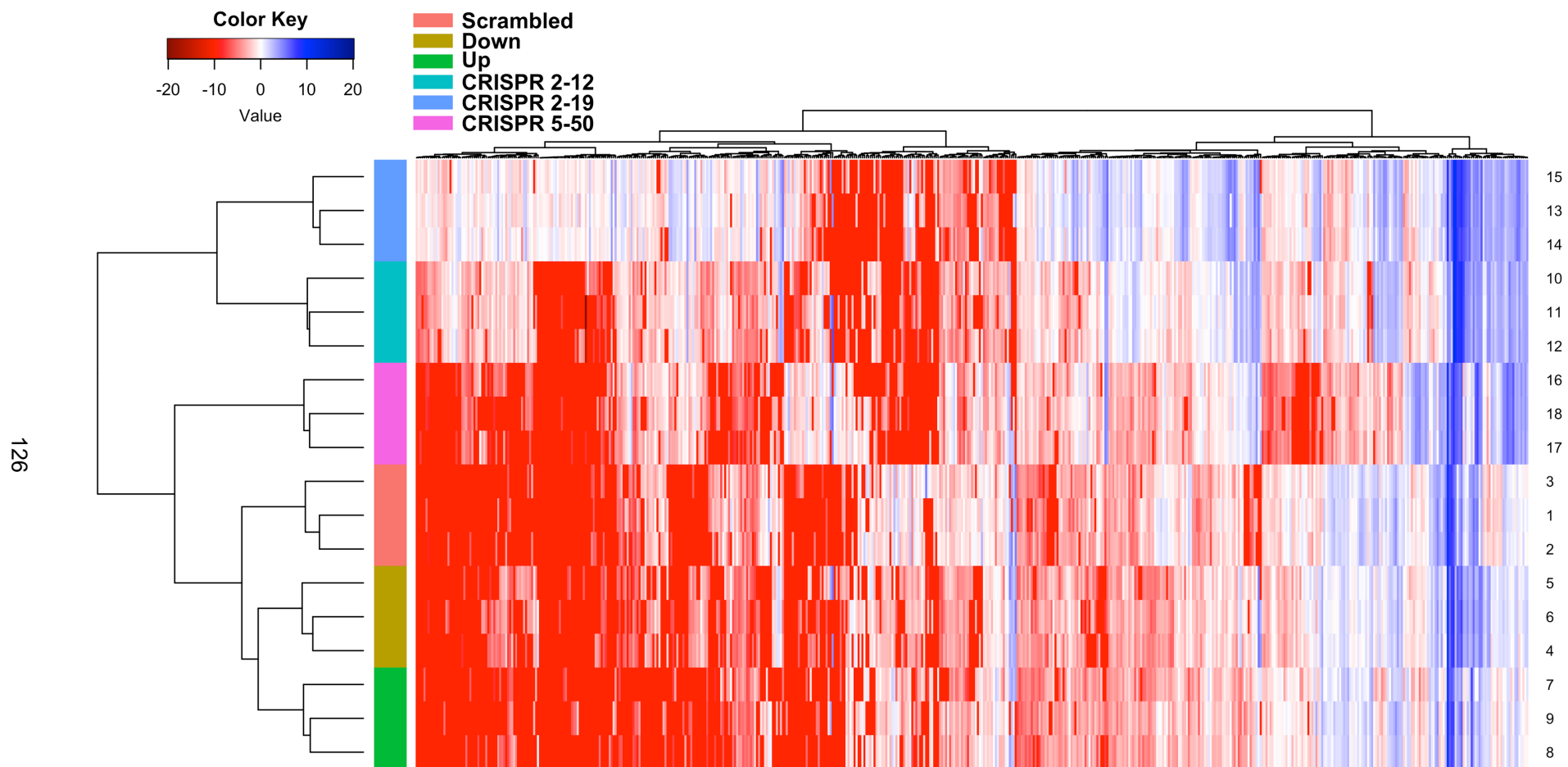


Figure 5.11: Hierarchical Clustering and Heatmap of Each Cell Lines Global Transcriptomics Signature.

Genes colored red on the heatmap had a median scaled relative abundance less than 1, genes colored white had a median scaled relative abundance of 1, and genes colored blue had a median scaled relative abundance greater than 1. Each column represents a single gene and each row represents a single transcriptomics sample; rows are labeled with sample number. Samples are color coded according to cell line identity. Unbiased hierarchical clustering reveals the two CRISPR/Cas9 cell lines constructed using guide RNA 2 are more similar than the two NAT1 complete knockout cell lines as would be expected.

Figure 5.12

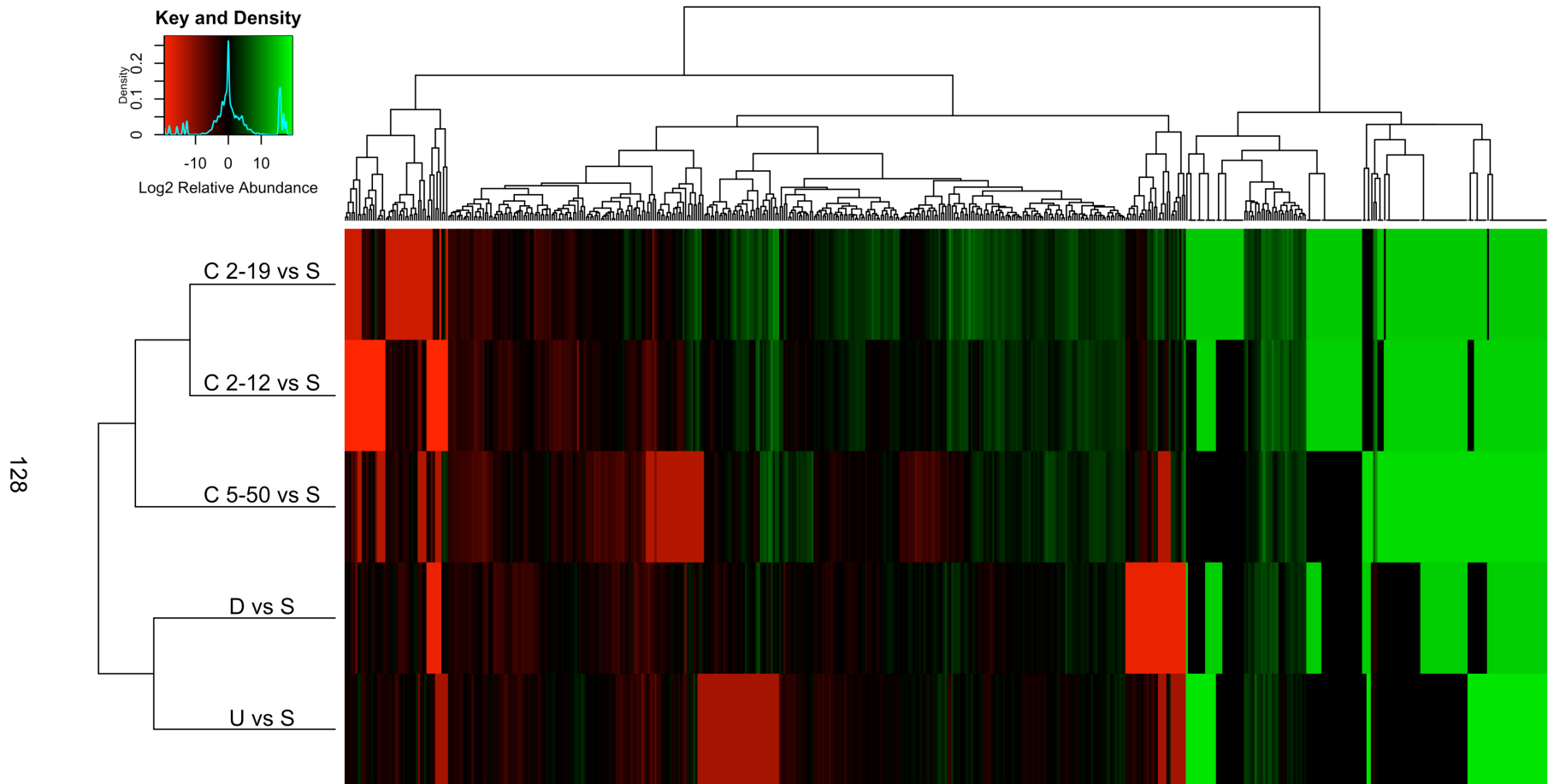


Figure 5.12: Hierarchical Clustering and Heatmap of Differential Gene Expression Compared to Scrambled.

Each column represents a single gene and each row represents the average gene expression for the specific comparison labeled across the three biological replicates. Red represents genes that were decreased compared to *Scrambled*, black represents no change, and green represents genes increased compared to *Scrambled*. U vs S: *Up vs Scrambled*; D vs S: *Down vs Scrambled*; C 212 vs S: *CRISPR 2-12 vs Scrambled*; C 219 vs S: *CRISPR 2-19 vs Scrambled*; C 550 vs S: *CRISPR 5-50 vs Scrambled*.

Similarly, principal component analysis showed that the CRISPR/Cas9 generated cell lines had global transcriptomic profiles that were distinct from the siRNA generated cell lines as well as each other. In our dataset, principal component 1 explains 43% of the variance in the data while principal component 2 explains 17% of the variance (Figure 5.13). There are two types of variance within the data set; within group and between group variance. For our experimental question, we are most interested in the between group variance. The *CRISPR 2-12*, *CRISPR 2-19*, and *CRISPR 5-50* groups are separated from the other three groups by principal component 1 (PC1). This reveals that the *Scrambled*, *Up*, and *Down* groups have global transcriptomic profiles that are similar to each other but very different from the other three groups given that PC1 represents 43% of the variance in our dataset. The *CRISPR 5-50* group is separated from the other five groups along principal component 2.

WGCNA Analysis

Weighted gene correlation network analysis was conducted to analyze networks of genes whose expression suggested interaction and/or involvement in the same pathways. First, all transcriptomics data were analyzed with the WGCNA algorithm to create modules of genes (Figure 5.14). Next Pearson correlation was conducted on resulting modules to identify modules correlated with NAT1 *N*-acetylation activity (Figure 5.15). We used this analysis to identify genes whose network topology suggested interaction with NAT2 as a strategy to identify genes that may have a role in the transcription of NAT2 (Figure 5.16).

Pathway Analysis

Pathway enrichment analysis was conducted on each group compared to the *Scrambled* group (Figure 5.17). Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways^{173,174} that had a normalized enrichment score of ≥ 1.60 were focused on. Disease associated pathways were removed from the analysis results because samples are known to be malignant thus confounding any possible conclusions about other diseases. Many pathways were significantly enriched for differences; amino acid, lipid, and nucleotide metabolism pathways were significantly enriched. Some enriched pathways did not include all group comparisons suggesting differential impacts on metabolism.

Figure 5.13

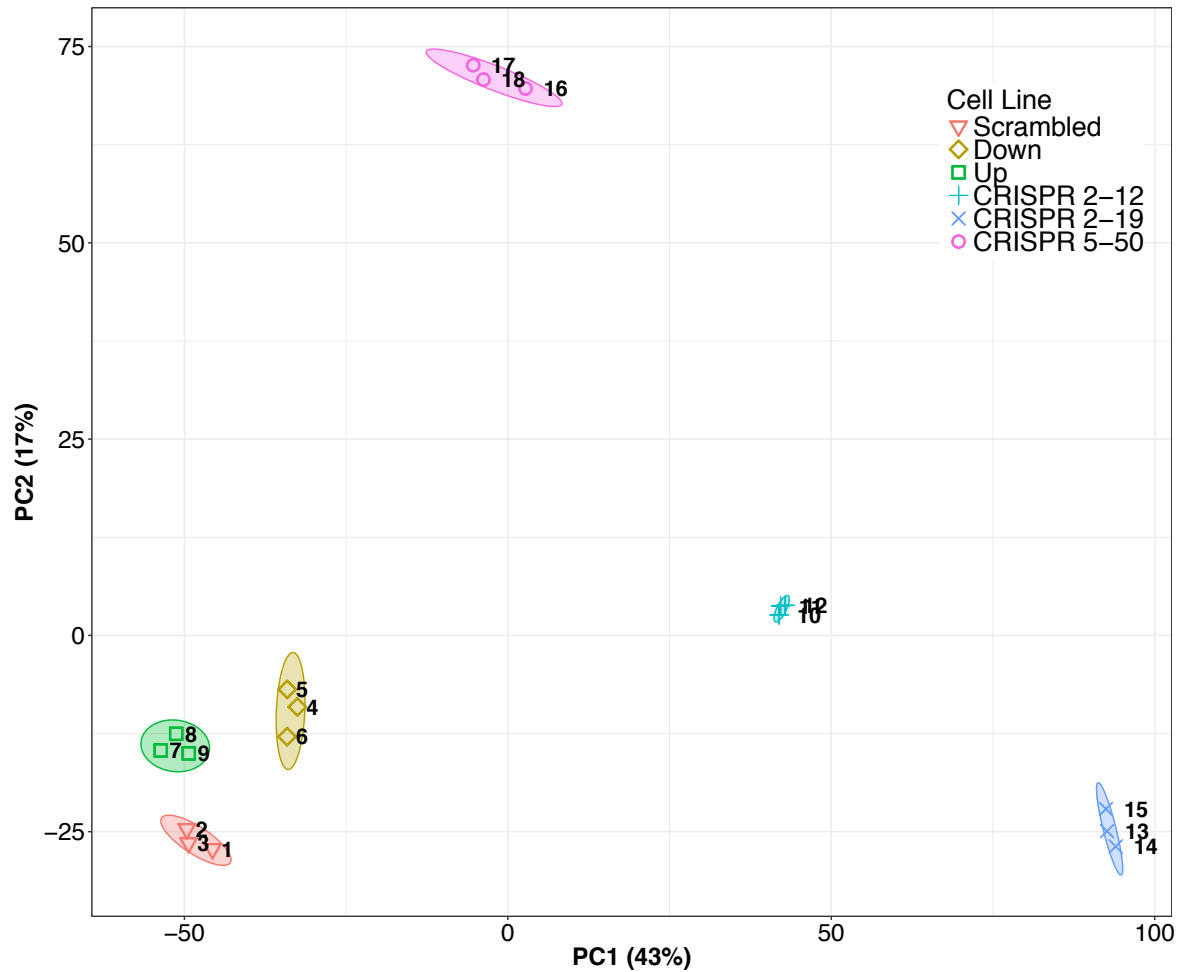


Figure 5.13: Transcriptomics Principal Components Analysis.

Each symbol represents an individual transcriptomics sample and is color coded by cell line. Principal component one represents 43% of the total variance in the dataset and separates the *CRISPR 2-12* and *CRISPR 2-19* cell lines from all other cell lines. Principal component 2 represents 17% of the total variance in our dataset and separates the *CRISPR 5-50* cell line from all other cell lines.

Figure 5.14

Transcript Modules

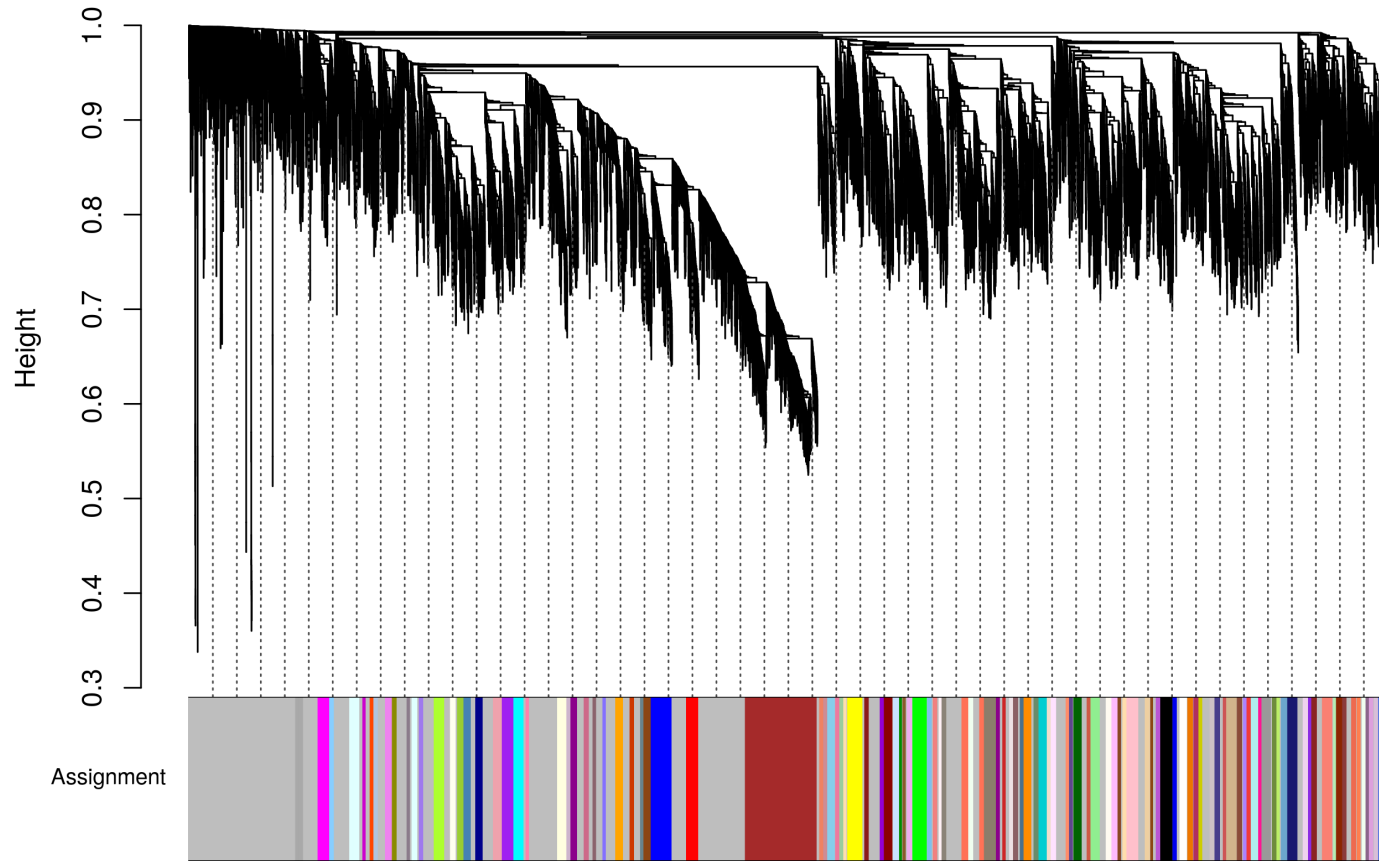
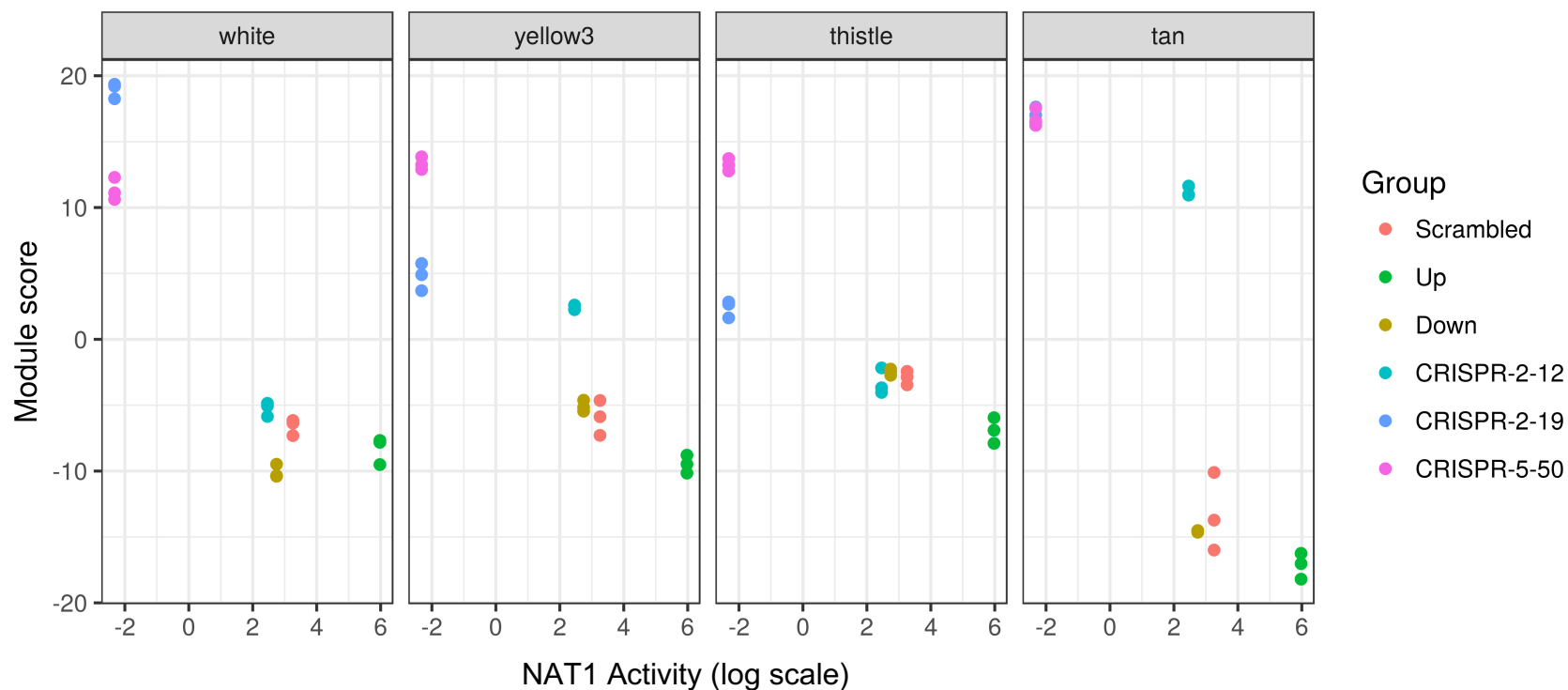


Figure 5.14: WGCNA Transcript Module Branches by Color

The Weighted Gene Correlation Network Analysis (WGCNA) algorithm was used to create modules of correlated genes based on topological overlap. Each module assignment is designated by a color. The clustering for each module of genes is shown. Modules of genes were utilized in further analysis.

Figure 5.15



133

Figure 5.15: Correlation Between WGCNA Transcript Modules and NAT1 N-acetylation Activity.

To identify gene modules related to varying levels of NAT1 correlation analysis between the module score of each sample and NAT1 N-acetylation activity was conducted. Four modules, white, yellow3, thistle, and tan, were associated with NAT1 activity. Each point represents a single metabolomics sample and is color coded according to cell line.

Figure 5.16

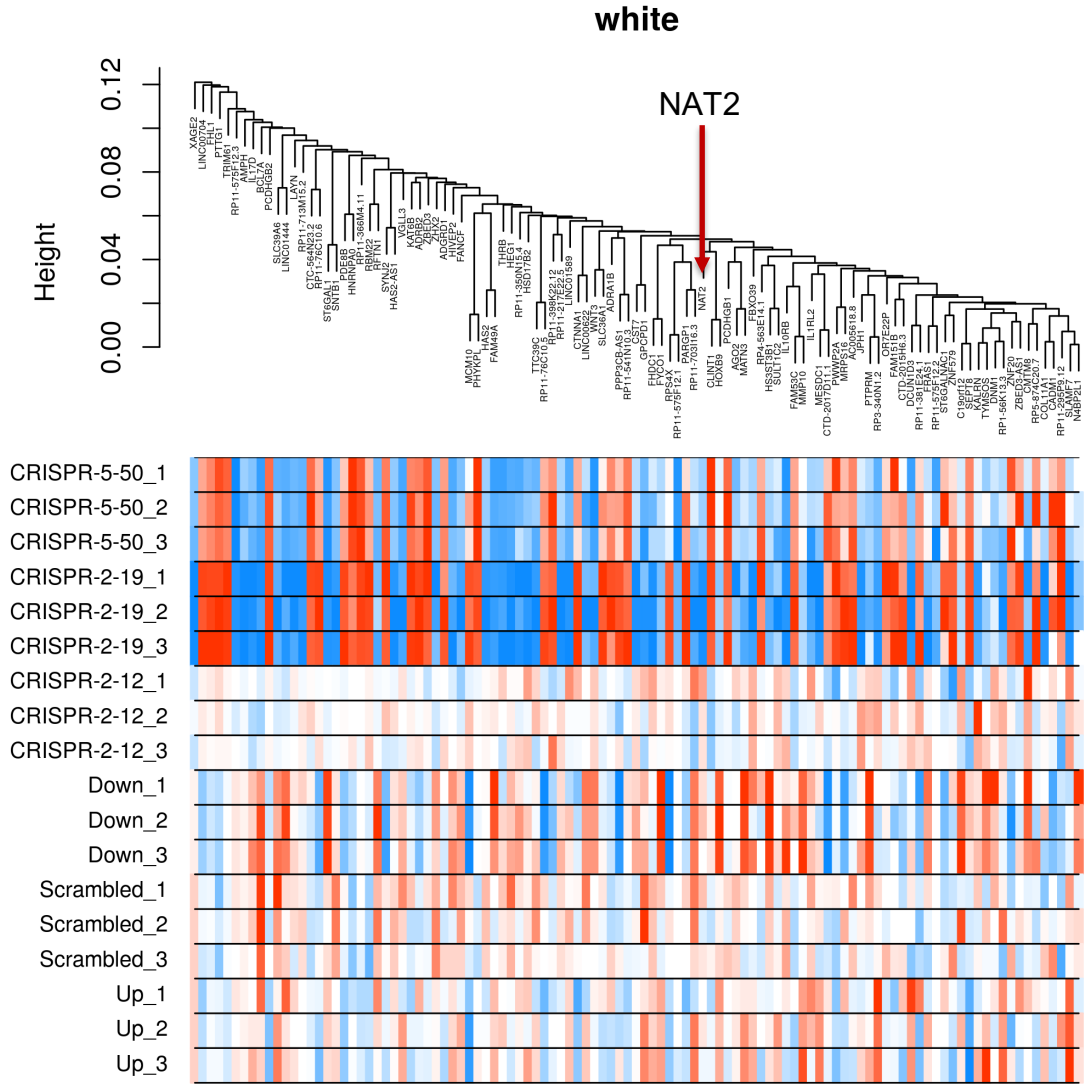


Figure 5.16: Heatmap of genes in WGCNA white module.

Heatmap showing relative gene expression for each gene in the white module. NAT2 is contained in this module. Clustering of genes in the white module is shown. Genes colored red on the heatmap had a median scaled relative abundance less than 1, genes colored white had a median scaled relative abundance of 1, and genes colored blue had a median scaled relative abundance greater than 1. Clustering indicates degree of topological overlap in gene expression.

Figure 5.17

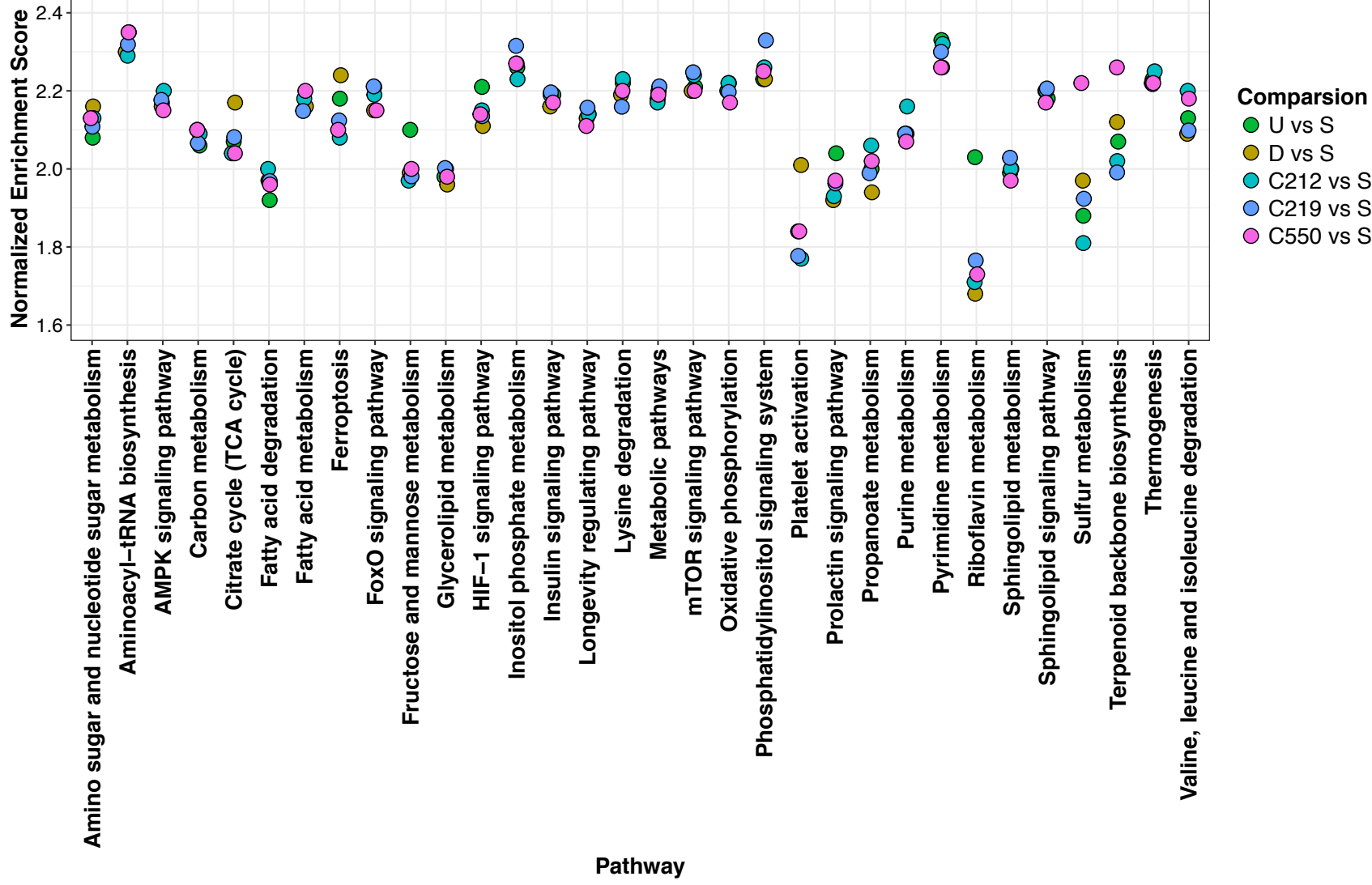


Figure 5.17: Transcriptomics Enrichment Analysis.

Pathway enrichment analysis was conducted for each group compared to *Scrambled* and is color-coded by comparison. We utilized the normalized enrichment score to determine the relative degree of enrichment. U vs S: *Up vs Scrambled*; D vs S: *Down vs Scrambled*; C212 vs S: *CRISPR 2-12 vs Scrambled*; C219 vs S: *CRISPR 2-19 vs Scrambled*; C550 vs S: *CRISPR 5-50 vs Scrambled*.

Discussion

Given that (theoretically) only a single gene, NAT1, was genetically altered in each cell line, it was expected that only a small proportion of genes would be differentially expressed, given the vast homeostasis mechanisms present¹⁷⁵⁻¹⁷⁷. However, a very large proportion (~90%) of all genes were observed to be significantly differentially expressed. Additionally, very few, if any, differences in gene expression were expected between the two NAT1 knockout cell lines because each cell line should otherwise have the exact same genome. Yet one of the most striking observations of this study is the differences in gene expression between the two complete NAT1 knockout cell lines. The hierarchical clustering, principal components analysis, and pathway enrichment analysis show there were significant differences between the two cell lines constructed using CRISPR/Cas9 guide RNA 2 and the cell line constructed using CRISPR/Cas9 guide RNA 5. Even though, in terms of NAT1 activity the *CRISPR 2-19* and *CRISPR 5-50*, cell lines are identical, their transcriptomic profiles are extremely different. This result suggests there are additional genetic differences between the two cell lines. The large number of differentially expressed genes between each knockout cell line compared to the *Scrambled* cell line and compared to each other could be because the cell lines have undergone additional unique mutations and/or because each CRISPR guide RNA used caused unique off-target effects in addition to targeting NAT1. To focus on transcriptome differences associated with varying levels of NAT1 in breast cancer, interpretation of this study has been focused on differentially expressed genes that agreed between the two CRISPR NAT1 knockout cell lines.

The observation that the complete knockout, but not knockdown, of NAT1 lead to NAT2 transcript production leads to the postulation that there may be a compensation mechanism that occurs when NAT1 *N*-acetylation is lost, suggesting NAT1 has an essential role. There have been reports of genetic compensation in a zebrafish model in which a gene was knocked out by a deleterious mutation but not in a zebrafish model where the same gene was knocked down by siRNA²¹⁶. There have been cases reported of humans that have no detectable NAT1 *N*-acetylation activity but NAT2 expression and/or activity was not investigated. Notably, a retrospective analysis of publicly available NAT1 and NAT2 gene expression data in established

breast cancer cell lines, primary breast tumor tissues, and normal breast tissues showed a small positive correlation between the two genes²¹⁷. One reason that study did not observe an inverse relationship between NAT1 and NAT2 gene expression may be because complete knockout of NAT1 may be necessary before the compensation mechanism occurs since NAT2 transcripts were only observed in the complete NAT1 KO cell lines but not the cell lines with decreased NAT1.

NAT1 has been shown to have redundancy with methylthioribose-1-phosphate isomerase (MRI1) in the methionine salvage pathway¹⁰⁷. However, MRI1 was only differentially expressed (increased) in one cell line, *CRISPR 2-12*, compared to *Scrambled*. As the MDA-MB-231 cell line does not have a functional methionine salvage pathway this result is not surprising. This example highlights the importance of studying the effects of varying NAT1 in multiple cell lines as each has unique mutations that may confound results.

Summary and Conclusions

This is the first study to measure the global transcriptome profile of breast cancer cells with varying levels of NAT1 *N*-acetylation activity. Most notably, NAT1 knockout and knockdown had a much greater impact on differential gene expression than the overexpression of NAT1. Additionally, NAT2 transcript production was observed in the two complete NAT1 knockout cell lines, *CRISPR 2-19* and *CRISPR 5-50*, but not any of the other cell lines. This observation suggests NAT2 transcripts are produced as a compensation mechanism for the complete loss of NAT1. However, the mechanism by which the cell senses the loss of NAT1 and transcribes NAT2 has not been identified and requires further investigation. Although NAT2 transcripts were detected in the NAT1 knockout cell lines, NAT2 *N*-acetylation activity was not thus the functionality of the NAT2 transcripts requires further investigation. Since this is the first study to measure the global transcriptome profile of breast cancer cells with varying levels of NAT1 *N*-acetylation activity I envision this transcriptomics dataset being mined for years to come as new discoveries are made about NAT1. An important use of this data will be comparisons and joint analysis between the transcriptomics data presented in this chapter and the metabolomics data presented in Chapter 4; this comparison is discussed in Chapter 6.

CHAPTER 6

COMBINING DATASETS

Background

The ability to combine multiple omics datasets for joint analysis is important for achieving a systems biology view of metabolism as there are many layers of regulation at each step, including transcriptional, translational, and post-translational regulation. Additionally, comparing results from multiple omics datasets can provide confidence that observations are not due to chance when multiple, independently measured datasets support a specific conclusion. Multi-omics integration is a relatively new area of research and as such methodologies are still being developed and improved (reviewed in ²¹⁸). There are many tools currently available, including MetaCore™, MetaboAnalyst²¹⁹, InCroMAP²²⁰, PaintOmics^{221,222}, Pathview²²³, MixOmics²²⁴, and 3Omics²²⁵. However, the “best” methodology for multi-omics integration is a fiercely debated topic in the omics field. It is the author’s opinion that there is considerable room for improvement and development of tools in the field of multi-omics analyses. Currently, most tools for combining omics datasets perform a pathway enrichment analysis using data from the multiple datasets. However, pathway enrichment analysis is highly biased by what is currently known about metabolism and what can be measured/detected by the omics methods. Bioinformatics skills as well as broad knowledge of metabolism are indispensable when utilizing the above mentioned omics integration tools to obtain and interpret the results in a meaningful way.

Methods and Results

Bioenergetics and Transcriptomics

To combine the bioenergetics and transcriptomics datasets expression of nuclear-encoded mitochondrial genes from the transcriptomics data was compared between constructed

cell lines. First, a list of ENTREZ gene IDs was created containing all mitochondrial complex genes from the HUGO gene nomenclature committee website²²⁶; the list contained 96 genes. Although the hg.38 reference assembly, which includes the human mitochondrial genome, was utilized for transcript mapping, mitochondrial DNA (mtDNA) encoded genes were not detected in the transcriptomics dataset. It is hypothesized that this is because of the relatively low abundance of mitochondrial DNA compared to nuclear DNA. Nuclear-encoded mitochondrial gene expression in all cell lines was visualized utilizing a heatmap (Figure 6.1). Fold change and significance of differential expression of those mitochondrial genes was visualized using volcano plots (Figure 6.2). Tables 6.1 – 6.4 lists fold-change and significance of all differentially expressed mitochondrial genes by cell line comparison. Notably, no mitochondrial genes were significantly differentially expressed between the *Up* and *Scrambled* cell lines. Additionally, most mitochondrial genes in the cell lines with decreased or knockout NAT1 *N*-acetylation activity (*Down*, *CRISPR 2-12*, *CRISPR 2-19*, and *CRISPR 5-50*) were downregulated compared to the cell line with parental NAT1 *N*-acetylation activity (*Scrambled*). Table 6.5 lists mitochondrial genes that were concordantly differentially expressed ($q \leq 0.05$) in the two NAT1 knockout cell lines, *CRISPR 2-19* and *CRISPR 5-50*.

Metabolomics and Transcriptomics

To combine the metabolomics and transcriptomics datasets a combined pathway enrichment analysis was performed and both metabolite abundance (from metabolomics dataset) and gene expression (from transcriptomics dataset) were mapped onto KEGG pathway maps for visualization utilizing Pathview²²³, an R/Bioconductor package for pathway-based data integration and visualization. For the combined enrichment analysis, the normalized enrichment score for each pathway from both datasets was averaged. Many KEGG pathways showed significant enrichment in the combined analysis (Figure 6.3), including 1. lysine degradation (Figure 6.4), 2. alanine, aspartate, and glutamate metabolism, 3. amino sugar and nucleotide sugar metabolism, 4. AMPK signaling, 5. biosynthesis of unsaturated fatty acids, 6. fructose and

Figure 6.1

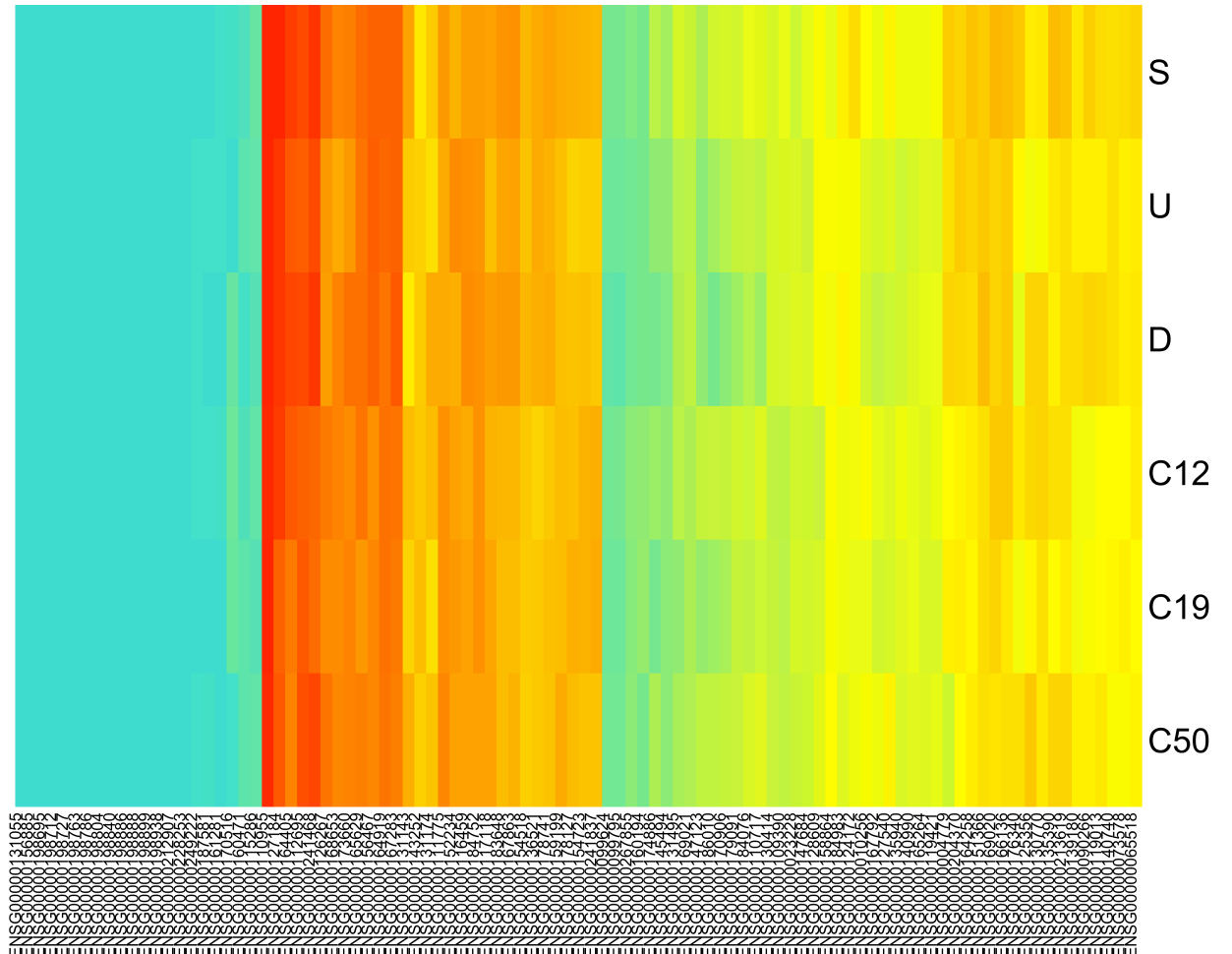
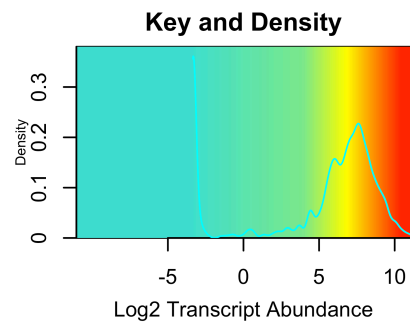


Figure 6.1: Heatmap of Mitochondrial Gene Expression.

Mitochondrial gene expression from the transcriptomics dataset was compared between constructed cell lines. A list of 96 ENTREZ gene IDs was created for all mitochondrial complex genes from the HUGO gene nomenclature committee website²²⁶. Transcript abundance was plotted as a heatmap. Cyan represents genes not expressed, yellow represents genes lowly expressed, and orange and red represent genes with high expression. Each row represents average transcript expression for 3 biological replicates from each cell line and columns represent a single mitochondrial gene.

Figure 6.2

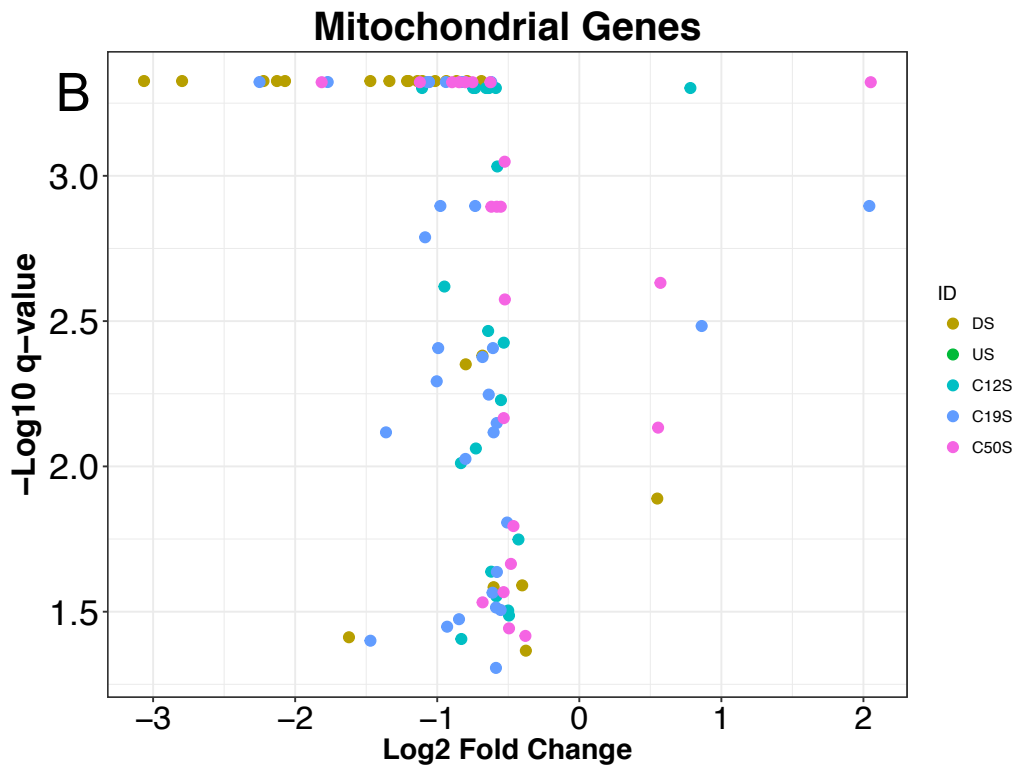
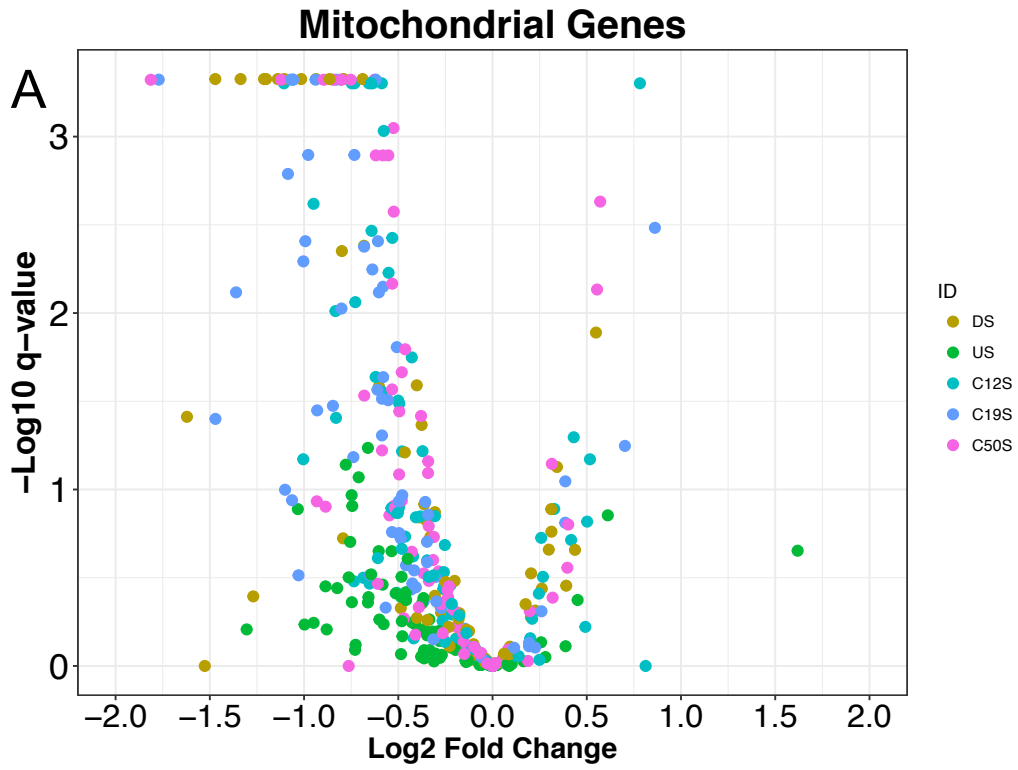


Figure 6.2: Mitochondrial Gene Expression

Mitochondrial gene expression was compared in the constructed cell lines. Genes were selected by creating a list of ENTREZ gene IDs for all mitochondrial complex genes from the HUGO gene nomenclature committee website²²⁶. Each point represents a single mitochondrial gene and is color coded by cell line compared to *Scrambled*. US: Up vs Scrambled; DS: Down vs Scrambled; C212S: CRISPR 2-12 vs Scrambled; C219S: CRISPR 2-19 vs Scrambled; C550 S: CRISPR 5-50 vs Scrambled. (A) All data was visualized as a volcano plot. (B) Only genes significantly differentially expressed were plotted.

Table 6.1

Differentially Abundant Nuclear-encoded Mitochondrial Genes in *Down* Cell Line Compared to *Scrambled*

COMPARISON	ENSEMBL GENE	ENTREZ ID	GENE SYMBOL	log2 Fold-Change	q-value
<i>Down vs Scrambled</i>	ENSG00000099624	513	ATP5D	-2.80	0.0005
	ENSG00000124172	514	ATP5E	-0.68	0.0042
	ENSG00000135390	517	ATP5G2	-0.86	0.0005
	ENSG00000131143	1327	COX4I1	-1.21	0.0005
	ENSG00000135940	1329	COX5B	-0.69	0.0005
	ENSG00000126267	1340	COX6B1	-0.79	0.0005
	ENSG00000131174	1349	COX7B	0.55	0.0129
	ENSG00000127184	1350	COX7C	-0.40	0.0257
	ENSG00000176340	1351	COX8A	-1.34	0.0005
	ENSG00000179091	1537	CYC1	-1.10	0.0005
	ENSG00000170906	4696	NDUFA3	-1.62	0.0387
	ENSG00000099795	4713	NDUFB7	-3.06	0.0005
	ENSG00000167792	4723	NDUFV1	-0.94	0.0005
	ENSG00000164258	4724	NDUFS4	-0.38	0.0431
	ENSG00000145494	4726	NDUFS6	-0.80	0.0045
	ENSG00000110717	4728	NDUFS8	-1.47	0.0005
	ENSG00000010256	7384	UQCRC1	-1.02	0.0005
	ENSG00000127540	10975	UQCR11	-1.20	0.0005
	ENSG00000184076	29796	UQCR10	-1.14	0.0005
	ENSG00000186010	51079	NDUFA13	-2.13	0.0005
ENSG00000147123	54539	NDUFB11	-0.60	0.0260	
ENSG00000174886	126328	NDUFA11	-2.22	0.0005	
ENSG00000115286	374291	NDUFS7	-2.07	0.0005	

Nuclear-encoded Mitochondrial Genes Differentially Abundant in the *Down* Cell Line Compared to *Scrambled* are listed.

Table 6.2

Differentially Abundant Nuclear-encoded Mitochondrial Genes in *CRISPR 2-12* Cell Line Compared to *Scrambled*

COMPARISON	ENSEMBL GENE	ENTREZ ID	GENE SYMBOL	log2 Fold-Change	q-value
<i>CRISPR 2-12 vs Scrambled</i>	ENSG00000110955	506	ATP5B	-0.53	0.0038
	ENSG00000124172	514	ATP5E	-0.73	0.0087
	ENSG00000135390	517	ATP5G2	-0.58	0.0009
	ENSG00000127184	1350	COX7C	-0.43	0.0178
	ENSG00000128609	4698	NDUFA5	-1.11	0.0005
	ENSG00000184983	4700	NDUFA6	-0.49	0.0326
	ENSG00000139180	4704	NDUFA9	-0.58	0.0280
	ENSG00000004779	4706	NDUFAB1	-0.75	0.0005
	ENSG00000183648	4707	NDUFB1	-0.55	0.0059
	ENSG00000090266	4708	NDUFB2	-0.95	0.0024
	ENSG00000164258	4724	NDUFS4	-0.64	0.0005
	ENSG00000160194	4731	NDUFV3	0.78	0.0005
	ENSG00000156467	7381	UQCRB	-0.73	0.0005
	ENSG00000169021	7386	UQCRFS1	-0.59	0.0005
	ENSG00000178741	9377	COX5A	-0.64	0.0005
	ENSG00000241468	9551	ATP5J2	-0.83	0.0393
	ENSG00000167863	10476	ATP5HJ	-0.64	0.0034
	ENSG00000127540	10975	UQCR11	-0.83	0.0097
	ENSG00000164405	27089	UQCRQ	-0.50	0.0313
	ENSG00000184076	29796	UQCR10	-0.66	0.0005
ENSG00000174886	126328	NDUFA11	-0.62	0.0230	

Nuclear-encoded Mitochondrial Genes Differentially Abundant in the *CRISPR 2-12* Cell Line

Compared to *Scrambled* are listed.

Table 6.3

Differentially Abundant Nuclear-encoded Mitochondrial Genes in *CRISPR 2-19* Cell Line Compared to *Scrambled*

COMPARISON	ENSEMBL GENE	ENTREZ ID	GENE SYMBOL	log2 Fold-Change	q-value
<i>CRISPR 2-19 vs Scrambled</i>	ENSG00000165629	509	ATP5C1	-0.73	0.0013
	ENSG00000099624	513	ATP5D	-1.36	0.0076
	ENSG00000124172	514	ATP5E	-0.98	0.0013
	ENSG00000135390	517	ATP5G2	-1.07	0.0005
	ENSG00000131143	1327	COX4I1	-0.80	0.0094
	ENSG00000126267	1340	COX6B1	-0.55	0.0312
	ENSG00000127184	1350	COX7C	-1.06	0.0005
	ENSG00000176340	1351	COX8A	-1.09	0.0016
	ENSG00000131495	4695	NDUFA2	-1.47	0.0398
	ENSG00000128609	4698	NDUFA5	-0.83	0.0005
	ENSG00000184983	4700	NDUFA6	-0.59	0.0306
	ENSG00000119421	4702	NDUFA8	-0.61	0.0039
	ENSG00000139180	4704	NDUFA9	-0.59	0.0494
	ENSG00000004779	4706	NDUFAB1	-0.60	0.0076
	ENSG00000183648	4707	NDUFB1	-0.64	0.0057
	ENSG00000090266	4708	NDUFB2	-0.85	0.0336
	ENSG00000065518	4710	NDUFB4	-0.61	0.0272
	ENSG00000099795	4713	NDUFB7	-1.00	0.0051
	ENSG00000167792	4723	NDUFV1	-0.58	0.0231
	ENSG00000164258	4724	NDUFS4	-0.80	0.0005
	ENSG00000160194	4731	NDUFV3	0.86	0.0033
	ENSG00000156467	7381	UQCRB	-0.58	0.0071
	ENSG00000169021	7386	UQCRFS1	-0.62	0.0005
	ENSG00000178741	9377	COX5A	-0.51	0.0156
	ENSG00000241468	9551	ATP5J2	-0.93	0.0356
	ENSG00000167863	10476	ATP5H	-0.68	0.0042
	ENSG00000127540	10975	UQCR11	-0.99	0.0039
	ENSG00000164405	27089	UQCRQ	-1.06	0.0005
	ENSG00000184076	29796	UQCR10	-0.94	0.0005
	ENSG00000160471	125965	COX6B2	2.04	0.0013
ENSG00000174886	126328	NDUFA11	-1.77	0.0005	
ENSG00000115286	374291	NDUFS7	-2.25	0.0005	

Nuclear-encoded Mitochondrial Genes Differentially Abundant in the *CRISPR 2-19* Cell Line Compared to *Scrambled* are listed.

Table 6.4

Differentially Abundant Nuclear-encoded Mitochondrial Genes in *CRISPR 5-50* Cell Line Compared to *Scrambled*

COMPARISON	ENSEMBL GENE	ENTREZ ID	GENE SYMBOL	log2 Fold-Change	q-value
<i>CRISPR 5-50 vs Scrambled</i>	ENSG00000110955	506	ATP5B	-0.55	0.0013
	ENSG00000124172	514	ATP5E	-0.80	0.0005
	ENSG00000154518	518	ATP5G3	-0.38	0.0383
	ENSG00000169020	521	ATP5I	-0.75	0.0005
	ENSG00000135940	1329	COX5B	-0.46	0.0160
	ENSG00000127184	1350	COX7C	-0.90	0.0005
	ENSG00000125356	4694	NDUFA1	0.57	0.0023
	ENSG00000128609	4698	NDUFA5	-0.48	0.0217
	ENSG00000184983	4700	NDUFA6	-0.85	0.0005
	ENSG00000004779	4706	NDUFAB1	-1.81	0.0005
	ENSG00000183648	4707	NDUFB1	-0.62	0.0013
	ENSG00000065518	4710	NDUFB4	-0.49	0.0361
	ENSG00000158864	4720	NDUFS2	0.55	0.0074
	ENSG00000164258	4724	NDUFS4	-0.63	0.0005
	ENSG00000145494	4726	NDUFS6	-0.53	0.0271
	ENSG00000204370	6392	SDHD	-0.68	0.0294
	ENSG00000156467	7381	UQCRB	-0.58	0.0013
	ENSG00000169021	7386	UQCRFS1	-0.53	0.0009
	ENSG00000178741	9377	COX5A	-0.52	0.0027
	ENSG00000167863	10476	ATP5H	-0.53	0.0068
ENSG00000164405	27089	UQCRQ	-1.12	0.0005	
ENSG00000160471	125965	COX6B2	2.05	0.0005	

Nuclear-encoded Mitochondrial Genes Differentially Abundant in the *CRISPR 5-50* Cell Line

Compared to *Scrambled* are listed.

Table 6.5Nuclear-encoded Mitochondrial Genes Concordantly Differentially Abundant in *CRISPR 2-19* and *CRISPR 5-50* Cell Lines

ENSEMBL GENE ID	ENTREZ ID	GENE SYMBOL/DESCRIPTION	log ₂ FC.C.R	
			CRISPR 2-19/S	CRISPR 5-50/S
ENSG00000124172	514	ATP5E ATP synthase, H ⁺ transporting, mitochondrial F1 complex, epsilon subunit	-0.98	-0.80
ENSG00000127184	1350	COX7C cytochrome c oxidase subunit VIIc	-1.06	-0.90
ENSG00000128609	4698	NDUFA5 NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 5	-0.83	-0.48
ENSG00000184983	4700	NDUFA6 NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 6, 14kDa	-0.59	-0.85
ENSG00000004779	4706	NDUFAB1 NADH dehydrogenase (ubiquinone) 1, alpha/beta subcomplex, 1, 8kDa	-0.60	-1.81
ENSG00000183648	4707	NDUFB1 NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 1, 7kDa	-0.64	-0.62
ENSG00000065518	4710	NDUFB4 NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 4, 15kDa	-0.61	-0.49
ENSG00000164258	4724	NDUFS4 NADH dehydrogenase (ubiquinone) Fe-S protein 4, 18kDa (NADH-coenzyme Q reductase)	-0.80	-0.63
ENSG00000156467	7381	UQCRB ubiquinol-cytochrome c reductase binding protein	-0.58	-0.58
ENSG00000169021	7386	UQCRFS1 ubiquinol-cytochrome c reductase, Rieske iron-sulfur polypeptide 1	-0.62	-0.53
ENSG00000178741	9377	COX5A cytochrome c oxidase subunit Va	-0.51	-0.52
ENSG00000167863	10476	ATP5H ATP synthase, H ⁺ transporting, mitochondrial Fo complex, subunit d	-0.68	-0.53
ENSG00000164405	27089	UQCRQ ubiquinol-cytochrome c reductase, complex III subunit VII, 9.5kDa	-1.06	-1.12
ENSG00000160471	125965	COX6B2 cytochrome c oxidase subunit VIb polypeptide 2 (testis)	2.04	2.05

Nuclear-encoded Mitochondrial Genes Concordantly Differentially Abundant in *CRISPR 2-19* and *CRISPR 5-50* Cell Lines

Figure 6.3

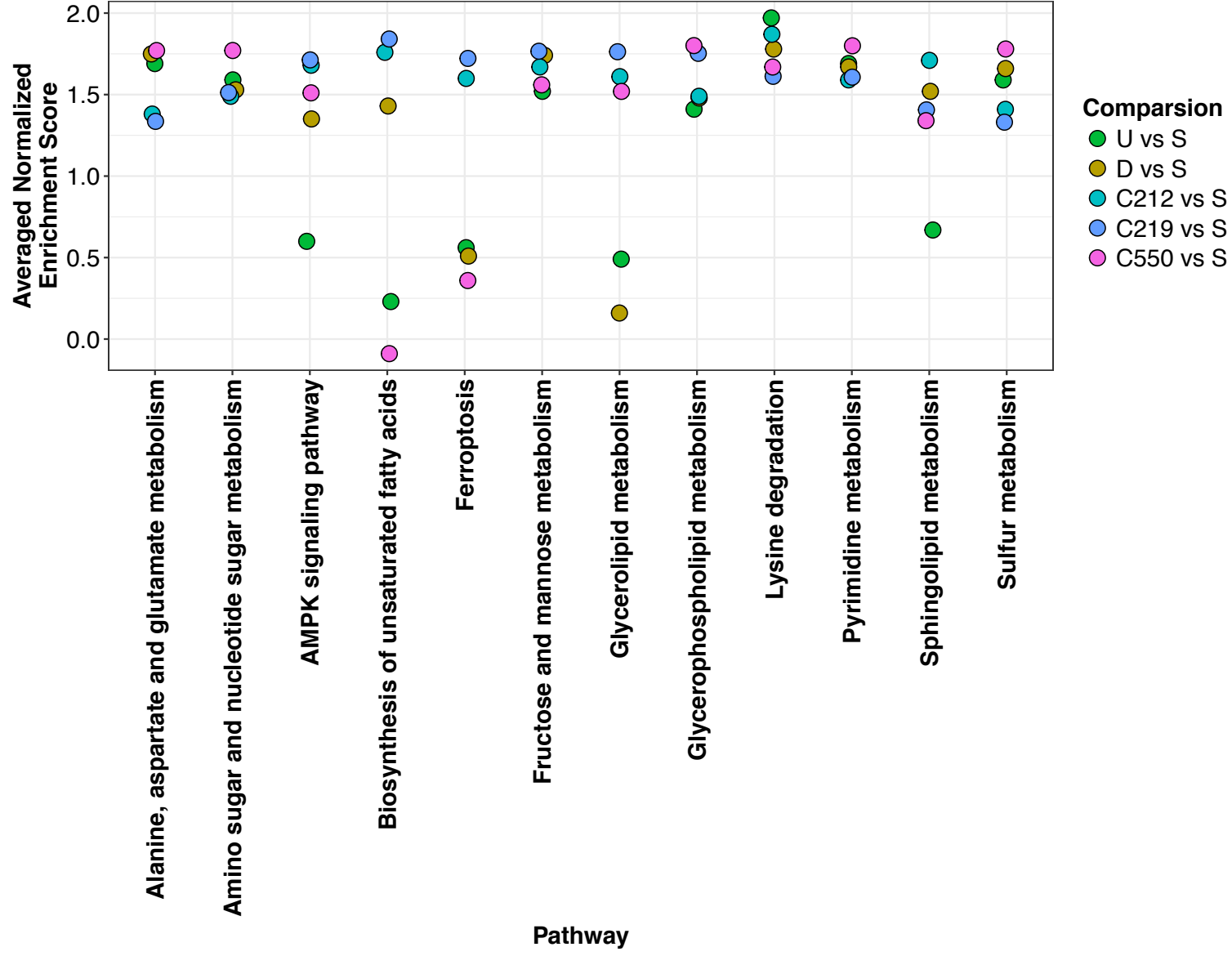


Figure 6.3: Combined Metabolomics and Transcriptomics Enrichment Analysis.

Pathway enrichment analysis was conducted for each group compared to *Scrambled* and is color-coded by comparison. The normalized enrichment score was utilized to determine the relative degree of enrichment. U vs S: Up vs Scrambled; D vs S: Down vs Scrambled; C212 vs S: CRISPR 2-12 vs Scrambled; C219 vs S: CRISPR 2-19 vs Scrambled; C550 vs S: CRISPR 5-50 vs Scrambled.

Figure 6.4

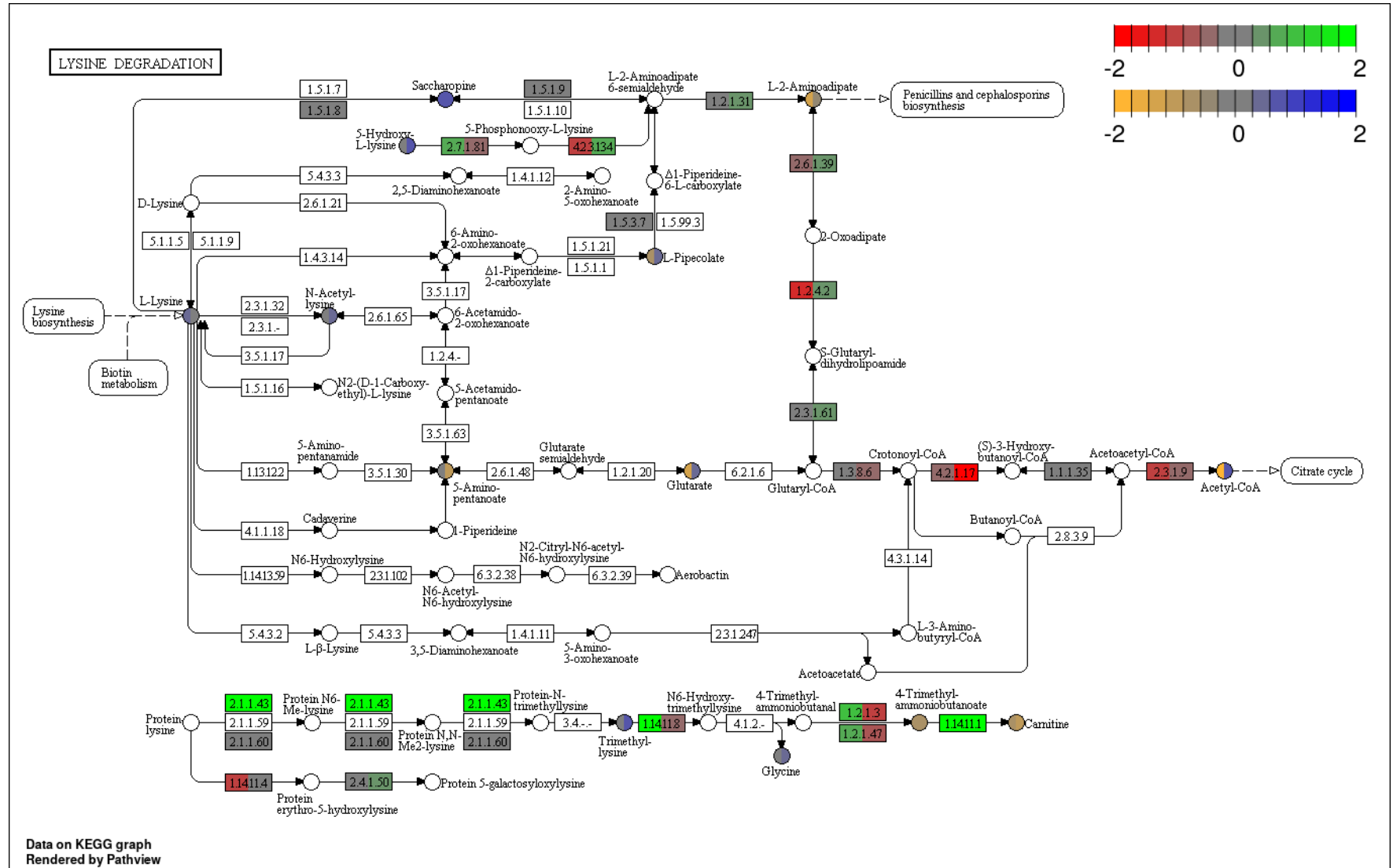


Figure 6.4: Metabolites and Genes Projected onto the KEGG Lysine Degradation Pathway.

NAT1 knockout cell lines, *CRISPR 2-19* and *CRISPR 5-50*, metabolite abundance and gene expression data compared to *Scrambled* cell line projected onto the KEGG lysine degradation pathway. Metabolite abundances in yellow are decreased in the NAT1 knockout cell lines, metabolites in gray are unchanged, and metabolites in blue are increased in the NAT1 knockout cell lines. Gene abundances in red are decreased in the NAT1 knockout cell lines, genes in gray are unchanged, and genes in green are increased in the NAT1 knockout cell lines. The *CRISPR 2-19* cell line abundance and expression data are shown on the right of the circles and boxes, respectively, while the *CRISPR 5-50* cell line abundance and expression data are shown on the left of the circles and boxes, respectively.

mannose metabolism, 7. glycerolipid metabolism, 8. glycerophospholipid metabolism, 9. pyrimidine metabolism, 10. sphingolipid metabolism, and 11. sulfur metabolism. Notably, the *Up* cell line compared to the *Scrambled* cell line was not significantly enriched in 5 of the pathways.

MixOmics²²⁴, an R package for omics feature selection and multiple data integration, was additionally utilized to conduct sparse partial least squares-discriminant analysis (sPLS-DA)²²⁷ utilizing the metabolomics and transcriptomics datasets as blocks. The transcriptomics dataset was first filtered to remove genes that had no detected expression in any sample, leaving 25,515 genes. A second filter was then applied to remove genes from the analysis for which 16 or more of the 18 total samples had no expression recorded; this left 23,931 genes in the dataset after both filters were applied. Additionally, because the chosen analysis relies on paired biological samples, only data from metabolomics samples that had corresponding transcriptomics data were utilized. This analysis had 18 samples total with 3 samples in each group.

Tuning of the algorithm parameters was required. First, a design matrix was created where blocks (metabolomics and transcriptomics datasets) were correlated with a link of 0.1. Utilizing four components was found to be optimal in minimizing the overall error rate. Therefore 4 components were utilized in the final model (Figure 6.5). The model was also tuned for the optimal number of variables to be selected from each dataset. It is important to note that the tuning function was set to favor relatively small numbers of variables to aide in interpretation of the final model results. The discriminant ability of components 1 and 2 was then visualized for each group in each block (metabolomics and transcriptomics; Figure 6.6). Component 1 separated the *CRISPR 5-50* cell line from the other cell lines; the corresponding loadings plot of component 1 visualized metabolites and genes that contribute the most to this discrimination (Figure 6.7; Tables 6.6 and 6.7). Conversely, component 2 separated the *CRISPR/Cas9* constructed cell lines from the *siRNA* constructed cell lines; however, the metabolites and genes of component 2 have greater contribution to the separation of the cell lines constructed using guide RNA 2 from all other cell lines. The corresponding loadings plot of component 2 visualized metabolites and genes that contributed the most to this discrimination and are each color coded by which group it is contributing to most (Figure 6.8; Tables 6.8 and 6.9). This result suggests that

Figure 6.5

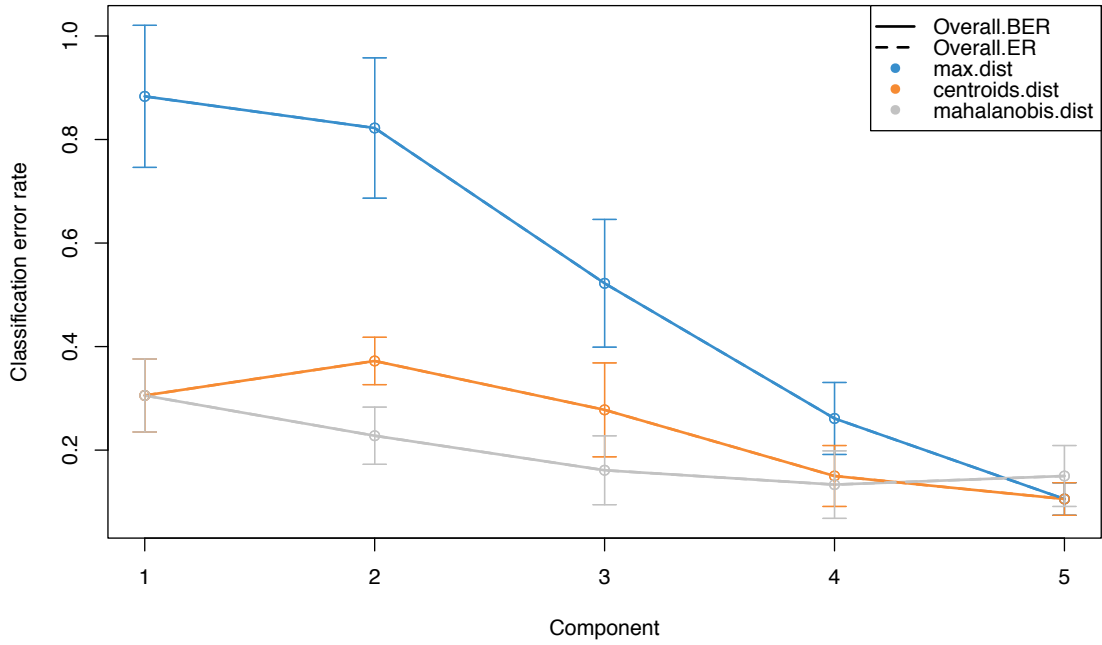


Figure 6.5 Optimization of Number of Components Included in sPLS-DA.

Solid lines represent the overall error rate. Blue line represents the max distance measure, orange line represents the centroids distance measure, and gray line represents the mahalanobis distance measure. Four components were chosen for the sPLS-DA model.

Figure 6.6

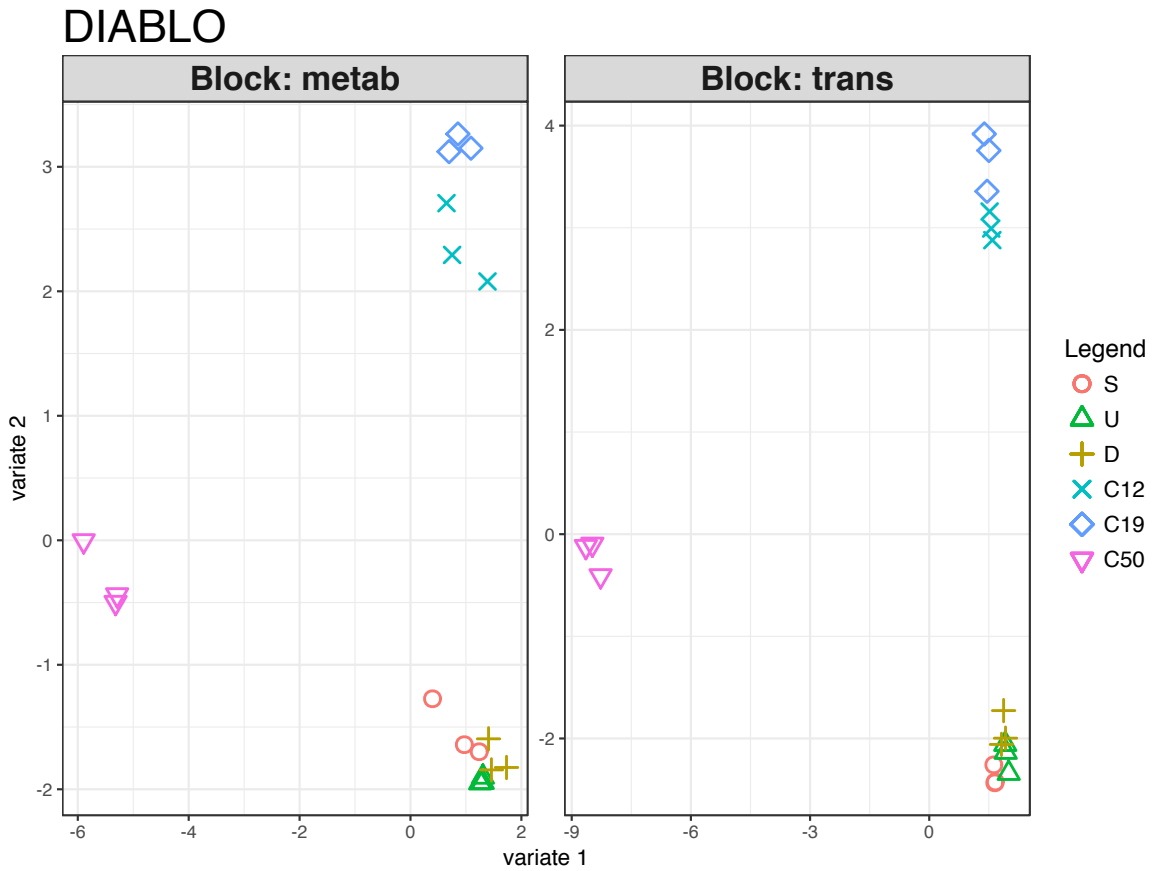


Figure 6.6: sPLS-DA Scores Plot.

Scores plot showing the discriminant ability of sPLS-DA components 1 (variate 1) and 2 (variate 2) for each group in each block (metab=metabolomics dataset and trans=transcriptomics dataset). S: Scrambled; U: Up; D: Down; C12: CRISPR 2-12; C19: CRISPR 2-19; C50: CRISPR 5-50.

Figure 6.7

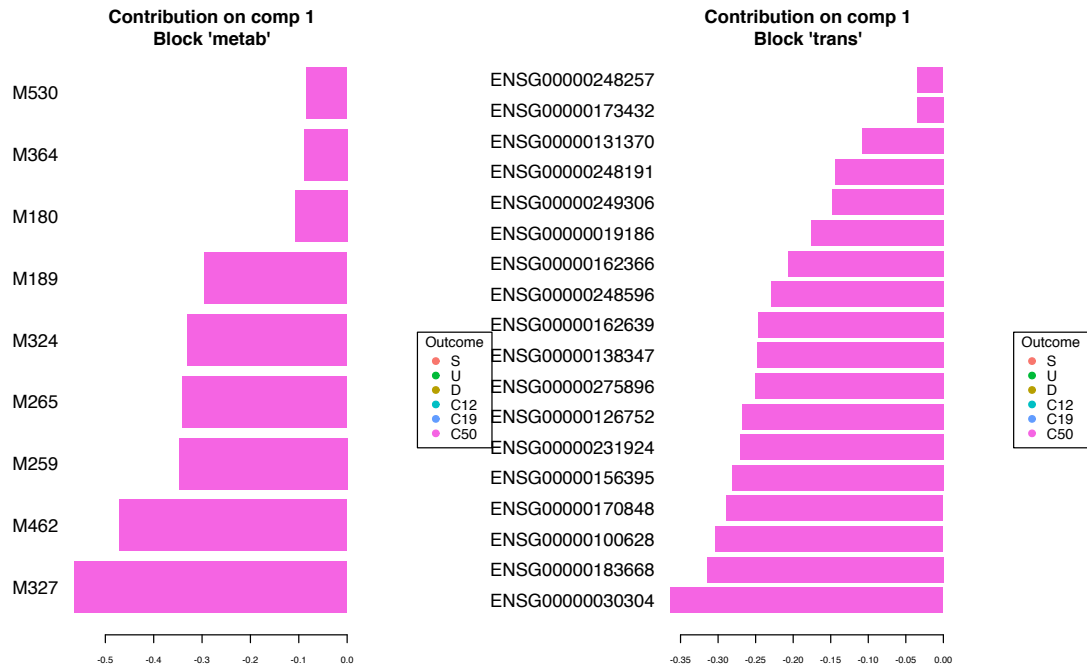


Figure 6.7: sPLS-DA Component 1 Loadings Plot.

Component 1 separates the *CRISPR 5-50* cell line from the other cell lines; the corresponding loadings plot of component 1 visualizes metabolites and genes that are contributing the most to this discrimination S: *Scrambled*; U: *Up*; D: *Down*; C12: *CRISPR 2-12*; C19: *CRISPR 2-19*; C50: *CRISPR 5-50*.

Table 6.6

sPLS-DA Component 1 Metabolites

M#	BIOCHEMICAL
M327	indolelactate
M462	phenyllactate (PLA)
M259	galactosylglycerol*
M265	gamma-glutamyl-epsilon-lysine
M324	imidazole lactate
M189	beta-citrylglutamate
M180	argininosuccinate
M364	methionine sulfoxide
M530	sulfate*

Metabolites included in sPLS-DA component 1.

Table 6.7

sPLS-DA Component 1 Genes

ENSEMBL GENE ID	ENTREZ GENE NAME
ENSG00000019186	cytochrome P450 family 24 subfamily A member 1 (CYP24A1)
ENSG00000030304	muscle associated receptor tyrosine kinase (MUSK)
ENSG000000100628	ankyrin repeat and SOCS box containing 2 (ASB2)
ENSG000000126752	SSX family member 1 (SSX1)
ENSG000000131370	SH3 domain binding protein 5 (SH3BP5)
ENSG000000138347	myopalladin (MYPN)
ENSG000000156395	sortilin related VPS10 domain containing receptor 3 (SORCS3)
ENSG000000162366	PDZK1 interacting protein 1 (PDZK1IP1)
ENSG000000162639	HEN1 methyltransferase homolog 1 (HENMT1)
ENSG000000170848	pregnancy specific beta-1-glycoprotein 6 (PSG6)
ENSG000000173432	serum amyloid A1 (SAA1)
ENSG000000183668	pregnancy specific beta-1-glycoprotein 9 (PSG9)
ENSG000000231924	pregnancy specific beta-1-glycoprotein 1 (PSG1)
ENSG000000248596	centrosomal protein 192kDa pseudogene (LOC643201)
ENSG000000249306	long intergenic non-protein coding RNA 1411 (LINC01411)
ENSG000000275896	protease, serine 2 (PRSS2)
ENSG000000248257	NA
ENSG000000248191	NA

Genes included in sPLS-DA component 1.

Figure 6.8

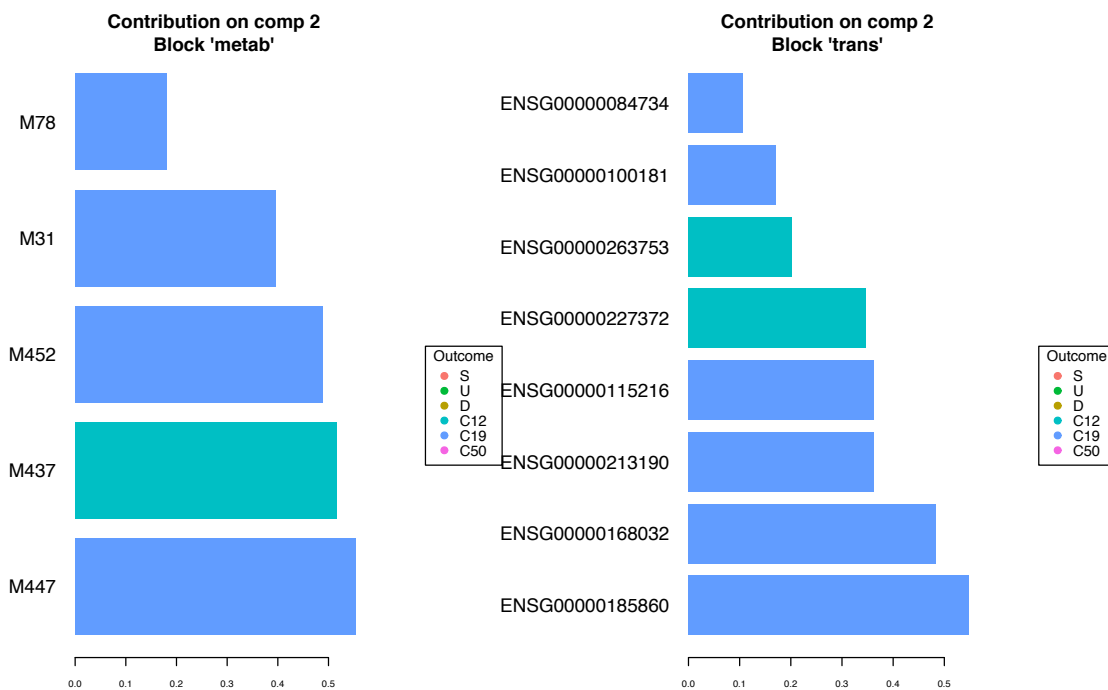


Figure 6.8: sPLS-DA Component 2 Loadings Plot.

Component 2 separates the CRISPR/Cas9 constructed cell lines from the siRNA constructed cell lines; however, the metabolites and genes of component 2 have greater contribution to the separation of the cell lines constructed using guide RNA 2 from all other cell lines. The corresponding loadings plot of component 2 visualizes metabolites and genes that are contributing the most to this discrimination and are each color coded by which group it is contributing to most S: *Scrambled*; U: *Up*; D: *Down*; C12: *CRISPR 2-12*; C19: *CRISPR 2-19*; C50: *CRISPR 5-50*.

Table 6.8

sPLS-DA Component 2 Metabolites

M#	BIOCHEMICAL
M447	palmitoleylcholine
M437	oleoylcholine
M452	palmitoylcholine *
M31	1-lignoceroyl-GPC (24:0)
M78	1-stearoyl-GPC (18:0)

Metabolites included in sPLS-DA component 2.

Table 6.9

sPLS-DA Component 2 Genes

ENSEMBL GENE ID	ENTREZ GENE NAME
ENSG00000084734	glucokinase regulator (GCKR)
ENSG00000100181	transmembrane phosphatase with tensin homology pseudogene 1 (TPTEP1)
ENSG00000115216	nuclear receptor binding protein 1 (NRBP1)
ENSG00000168032	ectonucleoside triphosphate diphosphohydrolase 3 (ENTPD3)
ENSG00000185860	coiled-coil domain containing 190 (CCDC190)
ENSG00000213190	myeloid/lymphoid or mixed-lineage leukemia; translocated to, 11 (MLLT11)
ENSG00000227372	TP73 antisense RNA 1 (TP73-AS1)
ENSG00000263753	long intergenic non-protein coding RNA 667 (LINC00667)
ENSG00000084734	glucokinase regulator (GCKR)

Genes included in sPLS-DA component 2.

while choline metabolism is altered in the CRISPR/Cas9 constructed cell lines it is dysregulated to a greater extent in the cell lines constructed using guide RNA 2.

We have additionally utilized a chord diagram (also known as circos plot) to visualize the correlations between variables of different types (i.e. metabolites and genes; Figure 6.9). A correlation cut-off of 0.9 was chosen; positive correlations are shown with green connectors while negative correlations are shown with red connectors. A high degree of correlation between variables was observed with most correlations being positive. The data are also presented as a clustered heatmap specifically implemented to represent the multi-omics molecular signature expression for each sample (Figure 6.10). Each cell line has a unique multi-omics signature with distinct clusters of both metabolites and genes that appear to differentiate each cell line. This result agrees with the hierarchical clustering and heatmaps generated for each dataset independently, as each cell line had a unique global metabolomics and transcriptomics signature, respectively.

Additionally, differential gene expression of SAT1 and SAT2 in the transcriptomics dataset were evaluated because SAT1 and SAT2 are known to *N*-acetylate putrescine and *N*-acetylputrescine was positively correlated with NAT1 *N*-acetylation activity in the metabolomics dataset. SAT1 expression was significantly increased in the *Down*, *CRISPR 2-12*, and *CRISPR 2-19* cell lines compared to the *Scrambled* cell line (Table 6.10). This result suggests the *Down*, *CRISPR 2-12*, and *CRISPR 2-19* cell lines should have increased *N*-acetylputrescine abundance compared to *Scrambled*. However, *N*-acetylputrescine abundances were observed to be decreased in those cell lines in the metabolomics dataset compared to the *Scrambled* cell line. These findings suggest the expression of SAT1 is not responsible for the *N*-acetylputrescine abundance observations. SAT2 was not significantly differentially expressed in any cell lines compared to the *Scrambled* cell line (Table 6.10).

Discussion

The presence of additional differences other than NAT1 (presented in Chapter 7) complicates interpretation of the results. However, the combined results from all experiments have been utilized to develop a working postulation of the role of NAT1 in breast cancer cell

Table 6.10

Differential Expression of the Spermidine/Spermine N1-Acetyltransferase (SAT) Genes

COMPARISON	GENE	Log2 FC Comparison/ <i>Scrambled</i>	q-value
<i>Up vs Scrambled</i>	SAT1	0.17	0.85
	SAT2	-0.61	0.84
<i>Down vs Scrambled</i>	SAT1	0.41	0.02
	SAT2	-0.81	0.56
<i>CRISPR 2-12 vs Scrambled</i>	SAT1	0.93	<0.001
	SAT2	-0.08	0.97
<i>CRISPR 2-19 vs Scrambled</i>	SAT1	1.50	<0.001
	SAT2	-0.56	0.73
<i>CRISPR 5-50 vs Scrambled</i>	SAT1	0.24	0.24
	SAT2	0.15	0.94

SAT transcript fold-changes between each cell line compared to the *Scrambled* cell line. Fold-change and *q*-value were calculated using cuffdiff. Genes with $q \geq 0.05$ were considered significantly differentially expressed. Significant differential gene expression is color-coded by direction of fold-change with increases shown in green.

Figure 6.9

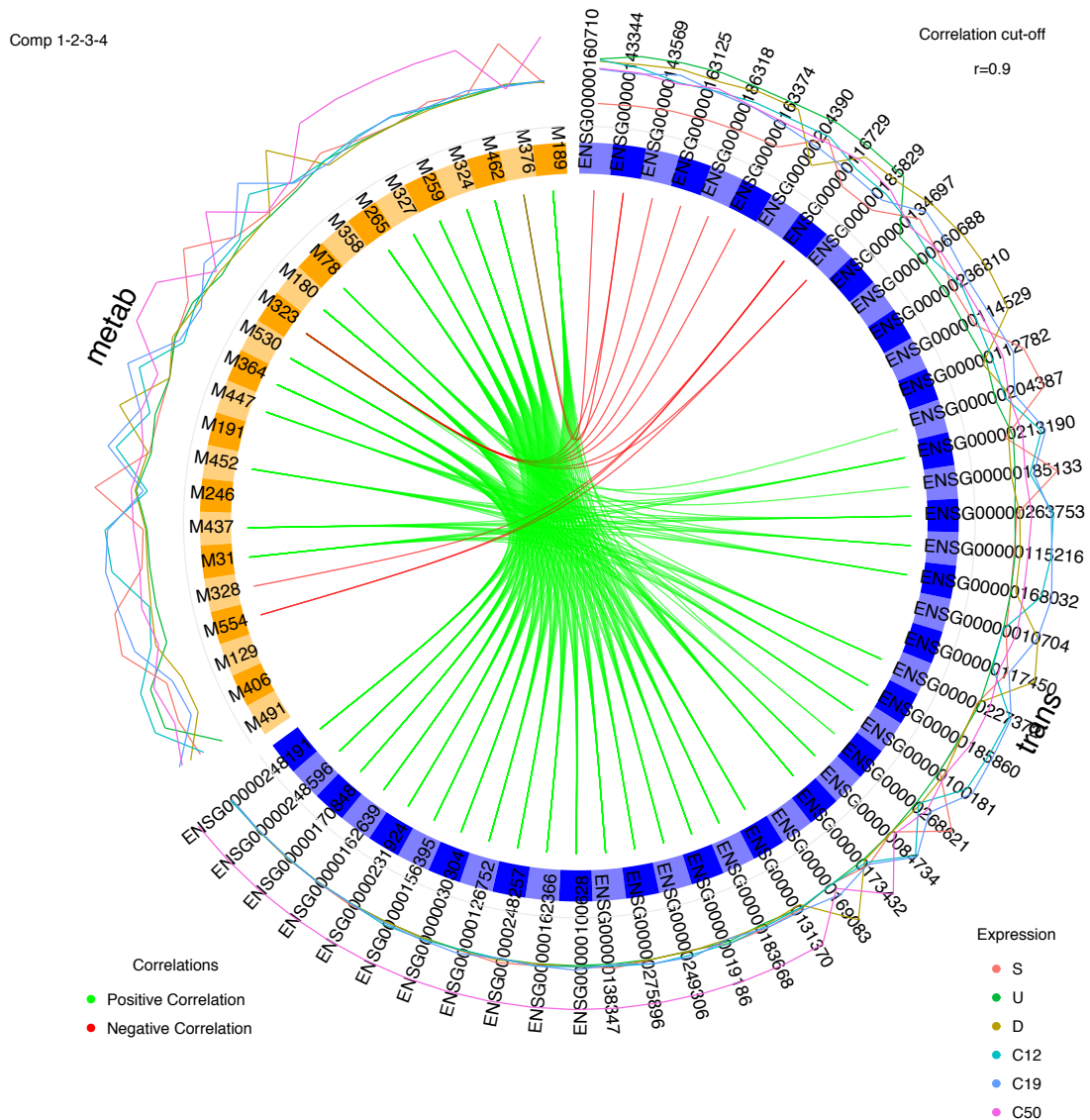


Figure 6.9: Circos Plot of Metabolites and Transcripts Included in sPLS-DA Model.

Metabolites are colored orange and genes are colored blue. A correlation cut-off of 0.9 was chosen; positive correlations are shown with green connectors while negative correlations are shown with red connectors. Relative abundances are shown with color coded lines on the outside of the plot. S: Scrambled; U: Up; D: Down; C12: CRISPR 2-12; C19: CRISPR 2-19; C50: CRISPR 5-50.

Figure 6.10

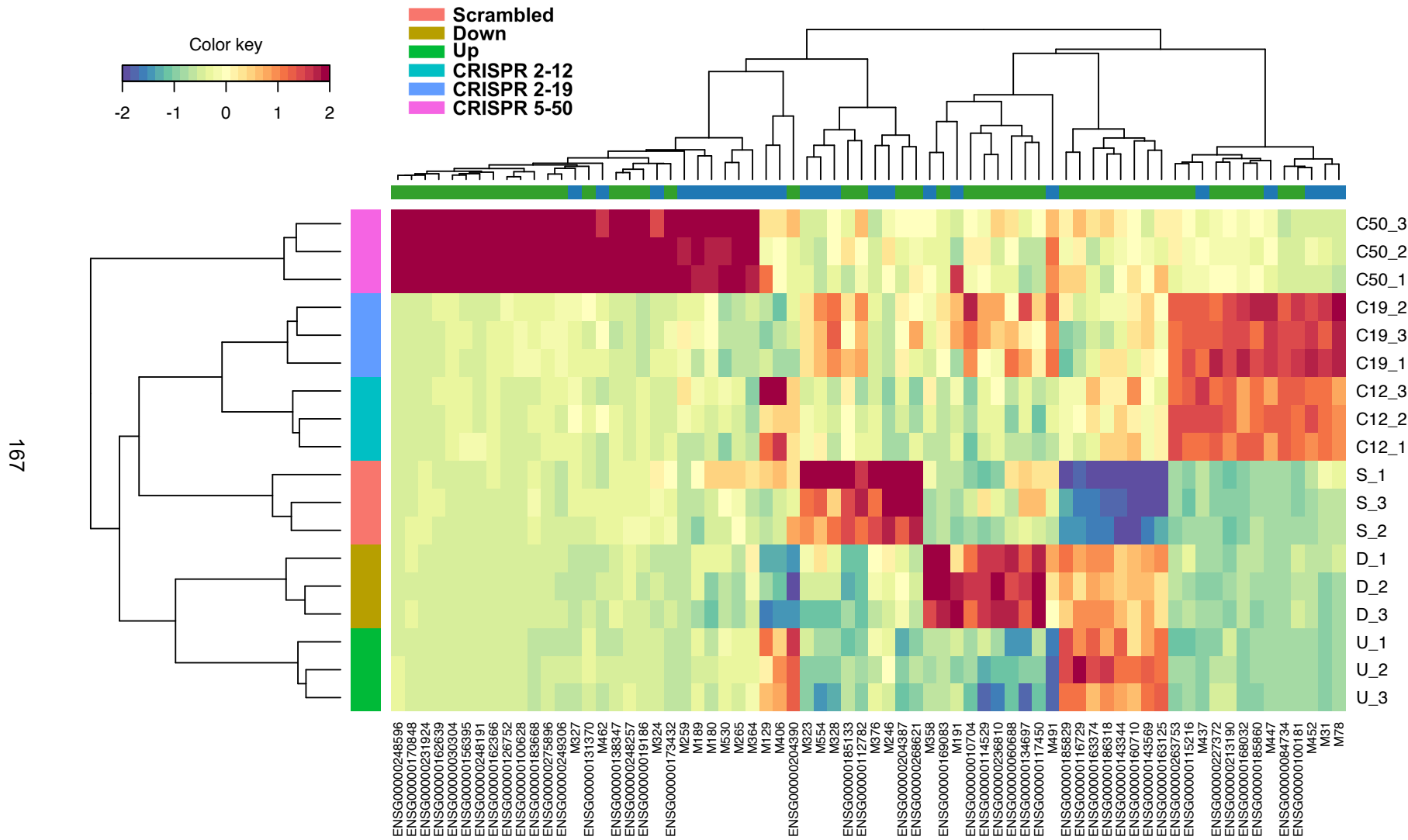


Figure 6.10: Combined Metabolomics and Transcriptomics Hierarchical Clustering and Heatmap.

Each row represents a single biological replicate and each column represents a single variable (gene or metabolite). Variables colored purple, blue, or green on the heatmap had a median scaled relative abundance less than 1, genes colored yellow had a median scaled relative abundance of 1, and genes colored orange or red had a median scaled relative abundance greater than 1. S: *Scrambled*; U: *Up*; D: *Down*; C12: *CRISPR 2-12*; C19: *CRISPR 2-19*; C50: *CRISPR 5-50*. Green represents the *Up* cell line, yellow represents the *Down* cell line, light red represents the *Scrambled* cell line, cyan represents the *CRISPR 2-12* cell line, sky blue represents the *CRISPR 2-19* cell line, and pink/purple represents the *CRISPR 5-50* cell line.

metabolism (Figure 6.11). Although many pathways were significantly enriched in the combined pathway analysis, the lysine degradation pathway has been focused on. Both carnitine and saccharopine, significant metabolites from the metabolomics dataset, are included in the lysine degradation pathway. Carnitine, an important metabolite in many pathways related to acetyl-CoA, is biosynthesized from methionine and lysine¹⁷⁸. Notably, NAT1 has a role in the methionine salvage pathway¹⁰⁷. Saccharopine is the first metabolite formed in the main lysine degradation pathway in mammals^{201,202}. Further experiments will be necessary to understand how NAT1 is connected to the lysine degradation pathway.

More generally, it was observed in the experiments presented in this dissertation that manipulation of NAT1 alters many metabolic pathways including amino acid, fatty acid, and nucleotide metabolism however it has been harder to narrow in on NAT1's role specifically due to the presence of differences other than NAT1. While the exact mechanism by which NAT1 affects breast cancer cell metabolism has not been identified, significant "pieces" have been added to the "puzzle". I predict, that as more experiments investigating the role of NAT1 in breast cancer are conducted, the bioenergetics, metabolomics, transcriptomics and combined analyses results presented here will not only have better context for interpretation but can also be utilized to add supporting or refuting information to those studies.

Summary and Conclusions

Combined analysis of the bioenergetics & transcriptomics and metabolomics & transcriptomics datasets was conducted utilizing currently available bioinformatics tools. The joint analysis of the bioenergetics and transcriptomics analysis suggest differential mitochondrial gene expression may help explain some of our bioenergetics observations however further experiments investigating mitochondrial number and morphology are needed to gain a better understanding of the relationship between NAT1 and mitochondria. Notably, mitochondrial function has been linked to regulation of metabolism, cell-cycle control, and cell death in addition to mitochondria's well-known role as the energy producers of the cell²²⁸.

Although the combined metabolomics and transcriptomics pathway analysis indicated 11 KEGG biochemical pathways were significantly enriched, the lysine degradation pathway was

Figure 6.11

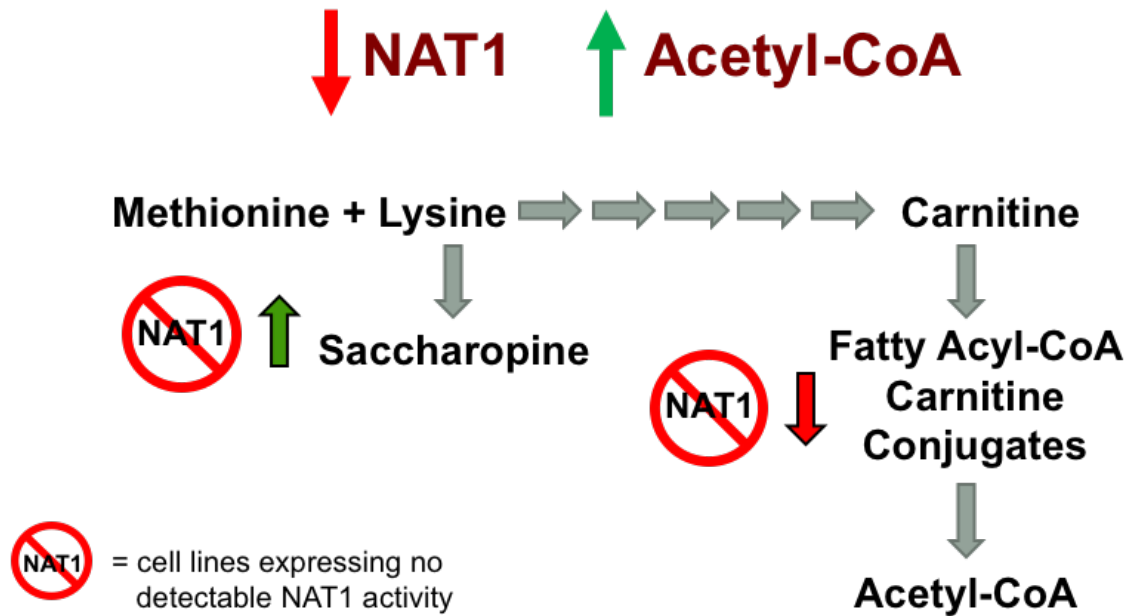


Figure 6.11: Overall Dissertation Conclusion.

The overall dissertation conclusion is that NAT1 has a role in the lysine degradation pathway however whether it is a direct role or a result of NAT1's ability to hydrolyze acetyl-CoA remains unknown. By utilizing a systems biology approach we were able to measure that fatty acyl-CoA carnitine conjugates are decreased when NAT1 is knocked out. Fatty acyl-CoA conjugates are necessary for transporting fatty acyl-CoAs to the mitochondria for beta-oxidation which in turn produces acetyl-CoA. We observed differences in mitochondrial bioenergetics. Additionally, carnitine is biosynthesized from methionine and lysine and lysine is degraded to saccharopine. Saccharopine is increased when NAT1 is knocked out.

focused on. Interpretation of the other pathways was complicated by the strong evidence of differences in addition to NAT1 in the genetically modified cell lines. Additionally, the strongly suggested presence of mutation of genes in addition to NAT1 lead to the sPLS-DA model not yielding useful information about genes and transcripts, that together, vary based on NAT1 *N*-acetylation activity. As better methods to integrate omics datasets become available, the datasets generated in this dissertation can be re-evaluated for novel insights. Currently available methods either have a reliance on pathway databases (hindering detection of unknown biochemical reactions) or have been developed to aide in the discovery of biomarkers or classification of patient samples.

CHAPTER 7

EVIDENCE OF DIFFERENCES OTHER THAN IN NAT1 IN GENETICALLY MODIFIED CELL LINES

Background

The constructed MDA-MB-231 cell lines utilized for the experiments presented in this dissertation were designed to vary only in NAT1 and were expected to have otherwise identical genomes. The guide RNAs utilized for construction were specifically designed to target only the NAT1 genomic sequence, thus limiting the risk of off-target effects. However, in all three datasets, bioenergetics (Chapter 3), metabolomics (Chapter 4), and transcriptomics (Chapter 5), evidence of differences in addition to NAT1 were observed in the constructed MDA-MB-231 cell lines which we discuss in detail in this chapter (Figure 7.1). Although evidence of differences in addition to NAT1 were observed between the CRISPR/Cas9 NAT1 knockout cell lines, the additional genes altered have not been identified. A whole genome sequencing (WGS) approach would be necessary to determine what additional genes differ between the cell lines. I postulate that the observed additional differences are likely due to both CRISPR/Cas9 off-target effects as well as mutations in the cell lines that may have occurred during clonal isolation, passaging, and/or propagation. STR-profiling authenticated all constructed cell lines as MDA-MB-231 but is limited by the fact that only a few genomic locations are interrogated. Mutations could have occurred at locations not probed by STR-profiling.

While I acknowledge that there is very strong evidence of differences in addition to NAT1 expression in the constructed cell lines, the elegant study design and sample selection utilized in this dissertation allowed those differences to be filtered out and focus to be placed on results that were due to the modulation of NAT1. For example, observations and results that were present

Figure 7.1

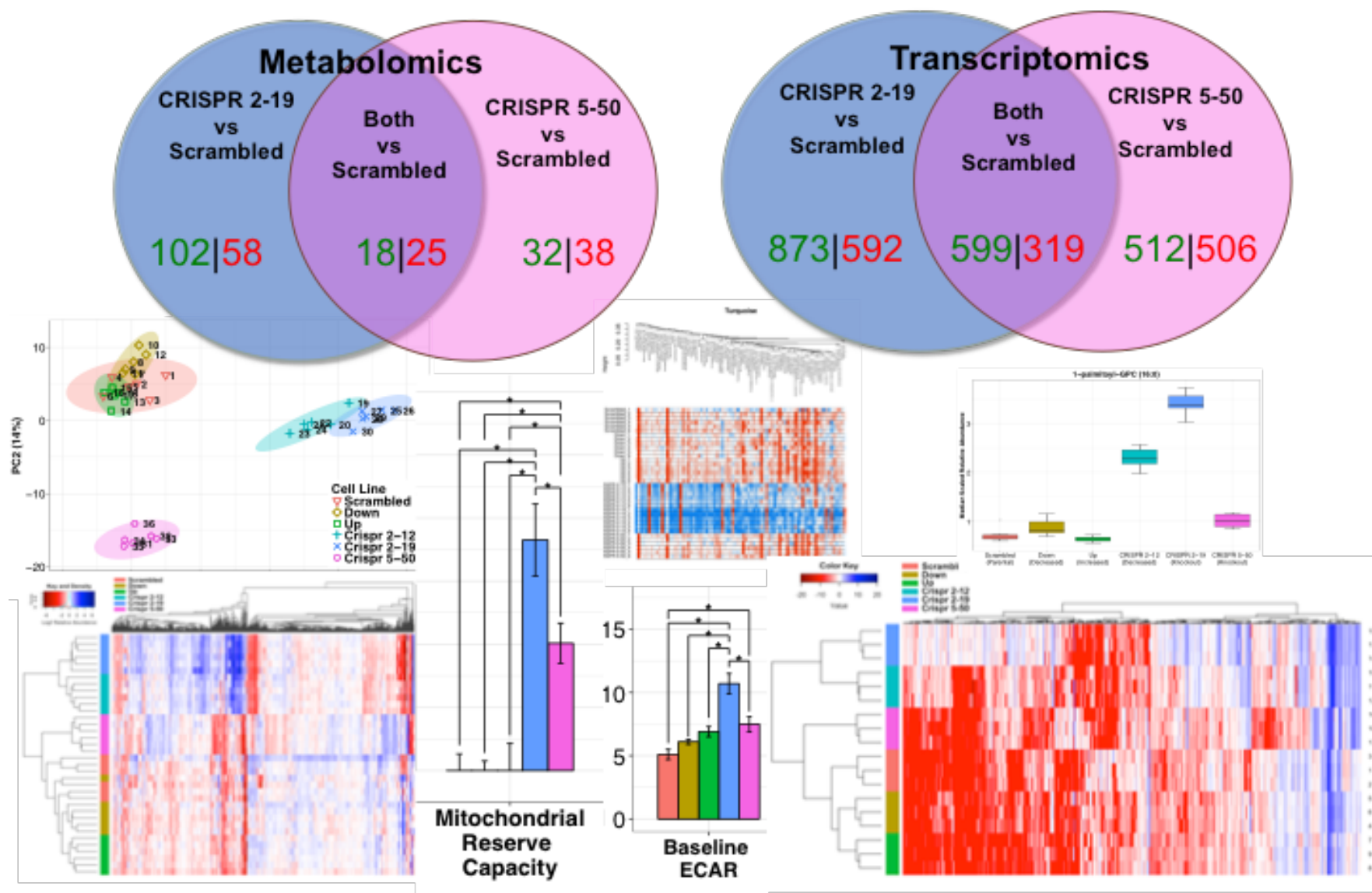


Figure 7.1: Evidence of Additional Differences Between Cell Lines Other Than NAT1.

There are many sources of evidence for the assertion that there are differences other than NAT1 in the constructed cell lines. Concordance between both metabolites and genes was lower than disagreement between the two MDA-MB-231 cell lines with no detectable NAT1 N-acetylation activity but constructed using two different guide RNAs; metabolites and genes in green were increased while metabolites and genes in red were decreased. Additionally, the first branch in hierarchical clustering, in both the metabolomics and transcriptomics datasets, separated the two MDA-MB-231 cell lines constructed utilizing guide RNA 2 but with different levels of NAT1 N-acetylation activity. In the bioenergetics dataset, both mitochondrial reserve capacity and baseline ECAR were significantly increased in the CRISPR 2-19 cell line compared to the CRISPR 5-50 cell line, even both have no detectable NAT1 activity. Principal component analysis in the metabolomics dataset revealed most of the variance in the dataset was between the two MDA-MB-231 cell lines constructed utilizing guide RNA 2 but with different levels of NAT1 N-acetylation activity and all other cell lines. Data is color coded by cell line; *Scrambled*:salmon, *Up*:green, *Down*:mustard, *CRISPR 2-12*:teal, *CRISPR 2-19*:sky blue, *CRISPR 5-50*:pinkish purple.

in both NAT1 knockout cell lines were focused on, given any off-target effects caused by the use of a single guide RNA are likely not present in both cell lines. Additionally, mutations that occurred during clonal isolation or passaging are not likely to have occurred in the same gene in multiple cell lines thus allowing a focus on similarities rather than differences in the two NAT1 knockout cell lines.

Bioenergetics

Mitochondrial reserve capacity was increased in both NAT1 knockout cell lines (*CRISPR 2-19* and *CRISPR 5-50*) compared to the *Parent*, *Scrambled*, and *Up* cell lines. However, it was also significantly higher in the *CRISPR 2-19* cell line compared to the *CRISPR 5-50* cell line suggesting differences other than NAT1 between the two cell lines. Baseline ECAR was also significantly increased in the *CRISPR 2-19* cell line compared to the *CRISPR 5-50* cell line. In the bioenergetics conclusions (Chapter 3), observations have been attributed to the modulation of NAT1 only if the change was observed in the same direction in both NAT1 knockout cell lines, *CRISPR 2-19* and *CRISPR 5-50*. Unfortunately, the *CRISPR 2-12* cell line was not utilized in the Bioenergetics studies thus preventing comparison between the two cell lines constructed via guide RNA 2.

Metabolomics

Multiple analyses of the metabolomics dataset suggested there are differences in addition to NAT1 between the constructed cell lines. Comparing metabolites that were concordantly differentially abundant in the two complete NAT1 knockout cell lines revealed many more metabolites were uniquely differentially abundant than agreed between the two NAT1 knockout cell lines, with only 16% concordance between the *CRISPR 2-19* and *CRISPR 5-50* cell lines. Additionally, hierarchical clustering analysis revealed the cell lines constructed using guide RNA 2, *CRISPR 2-12* and *CRISPR 2-19*, had more similar global metabolomics profiles than the cell lines with comparable NAT1 *N*-acetylation activity, *Down* and *CRISPR 5-50*, respectively. Furthermore, principal component analysis of the data revealed that nearly 53% of the total variance in the metabolomics data was between the two cell lines generated using guide RNA 2 and all other cell lines while approximately 17% of the total variance in the data was between the

CRISPR 5-50 cell line and all other cell lines. The first two principle components, together, account for approximately 70% of the total variance in the dataset and separate the *CRISPR/Cas9* constructed cell lines from the siRNA constructed cell lines.

Transcriptomics

Similar to the metabolomics dataset, many more genes were uniquely differentially expressed than were concordantly differentially expressed between the two NAT1 knockout cell lines with only 27% concordance between the *CRISPR 2-19* and *CRISPR 5-50* cell lines. Additionally, hierarchical clustering analysis revealed the cell lines constructed using guide RNA 2, *CRISPR 2-12* and *CRISPR 2-19*, had more similar global transcriptomics profiles than the cell lines with comparable NAT1 *N*-acetylation activity, *Down* and *CRISPR 5-50*, respectively. Furthermore, principal component analysis supported this conclusion although the results were not as striking as the metabolomics principal component analysis.

Combined Analysis

Analysis of the combined metabolomics and transcriptomics datasets also indicates there are differences other than NAT1 in our constructed cell lines. Component 1 of the sPLS-DA discriminates the *CRISPR 5-50* cell line from all others; the metabolites found in component 1 are characterized by lactate and sulfate moieties thus suggesting the *CRISPR 5-50* has additional mutations in pathways that contain these metabolites. Component 2 of the sPLS-DA discriminates the *CRISPR 2-12* and *CRISPR 2-19* cell lines from all others; the metabolites found in component 2 are characterized by choline moieties thus suggesting the *CRISPR 2-12* and *CRISPR 2-19* cell lines have additional mutations in pathways that contain choline metabolites.

Discussion

Since February 2013, when it was first reported that the *CRISPR/Cas9* system could be used to edit eukaryotic genomes²²⁹, many researchers and labs around the world have taken advantage of this system as a research tool to study the role of single genes in a variety of diseases²³⁰⁻²³⁹. Although the *CRISPR/Cas9* system has been shown to have high specificity, it is still possible for the guide RNAs used to cause off-target effects²⁴⁰⁻²⁴². Also, when using the clonal isolation method to select single cells where the gene of interest has been knocked-out rather

than a pool of cells, the chance of spontaneous, independent mutations occurring in each generated cell line increases. These off-target effects and spontaneous mutations may lead to incorrect conclusions about the role of the specific gene under investigation.

There are multiple online tools, including MIT CRISPR Design Tool²⁴³ and E-CRISP²⁴⁴, that use algorithms to generate CRISPR gRNA sequences and predict off-target mutation sites of those gRNAs based on “rules” about mismatch number and position. The E-CRISP web-based “evaluation” algorithm has been utilized to predict off-target mutation sites based on the sequences of the two guide RNAs utilized during construction (Table 7.1 & 7.2); MIT CRISPR Design Tool does not allow posteriori evaluation of gRNA sequences. The transcriptomics dataset was then utilized to compare gene expression of genes in which there was a predicted off-target sites. The author acknowledges that RNA expression does not always correlate perfectly with protein function and will not indicate definitely if a mutation causing knockout of the gene has occurred. However, a decrease in NAT1 transcripts in the two NAT1 knockout cell lines was observed and decreases in transcript abundance of off-target genes can serve as a proxy measure for off-target effects. No differences were observed in any of the genes predicted by the web-based algorithms to have off-target mutations.

Additionally, there are currently multiple methods available that have been developed specifically to detect off-target activity of CRISPR-Cas9 guide RNAs such as integrase-defective lentiviral vectors (IDLVs)²⁴⁵, Digenome-seq²⁴⁶, GUIDE-seq²⁴⁷, and qEva-CRISPR²⁴⁸, however these need to be performed during construction. Additionally, some researchers use whole genome sequencing (WGS) to evaluate the presence of possible off-target mutations. This is the only method that can definitively identify additional mutations *a posteriori* construction. Notably, however, not all researchers make use of these technologies to profile their cell lines. Moreover, many studies utilizing CRISPR/Cas9 technology to construct knockout cell lines do not conduct global, holistic profiling of the resulting cell lines but rather make small, very targeted, reductionist measurements. Additionally, very few studies utilize two different guide RNAs to knockout the gene of interest. The author hypothesizes other studies utilizing CRISPR/Cas9

Table 7.1

Guide RNA 2 predicted targets

TARGET CHROMOSOME	START	END	GENE TARGET (ENSB ID)	GENE NAME	SEQUENCE	DIRECTION
20	62393719	62393742	ENSG00000149679	CABLES2	TGAAGGGAACAACACTCAGATCAAG	fw
5	149360724	149360747	ENSG00000145882	PCYOX1L	CAAAGGGAACAGCTCGGACAAAG	fw
9	78994637	78994660	NA	NA	TAAAGAGAACAGCTGGGATCAAG	fw
17	74041132	74041155	NA	NA	CTAAGGGGACAGCTGGGATCCAG	fw
4	1816643	1816666	ENSG00000168924	LETM1	AGAAGGGGCCAGCTCGGATCCAG	fw
2	63708726	63708749	ENSG00000143951	WDPCP	AAATGGGAACAGCTCCGATCTAG	rc
8	18222137	18222160	ENSG00000171428	NAT1	CAAAGGGAACAGCTCGGATCTGG	rc
8	18400093	18400116	ENSG00000156006	NAT2	CAAAGGGAACAGCCCGGATCTGG	rc

E-CRISP web-based “evaluation” algorithm was utilized to predict off-target mutation sites based on the sequences of the guide RNA 2 utilized during construction. fw = forward orientation, rc = reverse orientation

Table 7.2

Guide RNA 5 predicted targets

TARGET CHROMOSOME	START	END	GENE TARGETS (ENSBL ID)	GENE NAME	SEQUENCE	DIRECTION
3	122941233	122941256	ENSG00000082684	SEMA5B	ACATTGGCTCTAAGAAGCCTAAG	fw
8	77188268	77188291	NA	NA	TCATTGGCTAGAACAAGTCTTAG	rc
21	24269418	24269441	NA	NA	GAATTGGCTATCAGAAGTCCTAG	fw
2	194868769	194868792	NA	NA	AAAATGGCTACAAGAAGTCTTAG	fw
CHR_HSCHR6_MHC_COX_CTG1	30001513	30001536	ENSG00000229653	NCRNA00171	CTATTGGTTATAAGAAGTCACAG	fw
CHR_HSCHR6_MHC_APD_CTG1	30004125	30004148	ENSG00000236598	NCRNA00171	CTATTGGTTATAAGAAGTCACAG	fw
CHR_HSCHR6_MHC_DBB_CTG1	30001951	30001974	ENSG00000225618	NCRNA00171	CTATTGGTTATAAGAAGTCACAG	fw
8	91943377	91943400	NA	NA	TTATTGGCTATAAGATGCCTCAG	rc
2	208352693	208352716	ENSG00000115020	PIKFYVE	GAATTGGTTATAAGAACTCTCAG	fw
3	104893231	104893254	NA	NA	TCATTGGATATGAGAAGTCTCAG	fw
5	117577755	117577778	NA	NA	TAATTGGCTAGAAGAAGTTTCAG	rc
6	123305929	123305952	ENSG00000186439	TRDN	TCAGTGGCTATAACAAGTCTCAG	fw
18	39817888	39817911	NA	NA	GGATTGGCTATAAGAAGTAAGAG	fw
5	91359933	91359956	NA	NA	TGGTTGGCTATAAGAAGACTGGG	fw
7	22336060	22336083	ENSG00000136237	RAPGEF5	GCATTGGGAATAAGAAGTCTGGG	rc
8	18222075	18222098	ENSG00000171428	NAT1	GAATTGGCTATAAGAAGTCTAGG	fw
8	18400031	18400054	ENSG00000156006	NAT2	GAATTGGCTATAAGAACTCTAGG	fw
1	90780522	90780545	NA	NA	GGATTGGCTATATGAAGTCAAGG	fw
11	37011189	37011212	NA	NA	TTATTAGCTATAAGAAGGCTAGG	rc

E-CRISP web-based “evaluation” algorithm was utilized to predict off-target mutation sites based on the sequences of the guide RNA 5 utilized during construction. fw = forward orientation, rc = reverse orientation

generated cell lines most likely have additional differences in addition to their gene of interest but do not make measurements that would allow detection of such differences. Convincing evidence has been presented in this chapter that greater caution should be applied when interpreting results from studies utilizing CRISPR/Cas9 to knockout a single gene of interest especially when those studies have not profiled the cell lines globally, utilizing at least one omics technology and/or utilizing two different guide RNA's to knockout the gene thus constructing two knockout cell lines for comparison. Great bias is introduced via the reductionist approach in those studies and that bias may lead to naivety about additional mutations in samples.

As an example, in a study by Tsai et al. multiple CRISPR/Cas guide RNAs were used to edit the gene VEGFA, on chromosome 6, to compare on-target and off-target mutations²⁴⁷. The authors found the selected guide RNAs could also target genes on virtually every one of the other 22 human chromosomes. Although each CRISPR guide RNA had zero to a dozen or so predicted off-target sites (by previously discussed web-based algorithms), neither algorithmic software used (MIT CRISPR Design Tool²⁴³ and E-CRISP²⁴⁴) identified the vast majority of off-target sites²⁴⁷; depending on the guide RNA, 0 to >150 off-target mutations were observed. Additionally, the off-target sites had diverse sequences making prediction of where off-target effects occur rather difficult. Without prediction of these off-target sites, scientists would have no idea where to look in the genome for possible off-target mutations and no idea those errors were possible. This example highlights the importance of global profiling (unbiased measurement of many aspects of the global system) of CRISPR/Cas9 generated cell lines as well as the importance of including cell lines generated with multiple different guide RNAs to be able to filter out off-target effects. Notably, another study found that many off-target sites were mutated with frequencies comparable to or even higher than the on-target site²⁴⁹.

Summary and Conclusions

The results from all experiments presented in this dissertation, 1. bioenergetics (Chapter 3), 2. metabolomics (Chapter 4), and 3. transcriptomics (Chapter 5), strongly suggest the constructed MDA-MB-231 cell lines had differences in gene(s) other than NAT1 that were unexpected. Without the sophisticated study design in which a global systems biology approach

was utilized as well as multiple guide RNAs, I most likely would not have been aware that there were additional differences between the cell lines. Additionally, it would have been nearly impossible to parse out what effects were due to NAT1 and what effects were due to these additional differences. These findings have especially great implications for researchers planning to utilize CRISPR/Cas9 gene editing technology in humans; I assert that the potential for off-target effects using CRISPR/Cas9 has been underappreciated and under-reported in the literature due to the lack of global profiling. Currently, considerable research is being conducted on ways to optimize CRISPR editing efficiency and decrease the number of off-target effects that occur when utilizing CRISPR/Cas9 technology to edit genes²⁵⁰. Additionally, researchers in the field are working on better ways to detect and/or predict off-target effects given both are still a major challenge.

CHAPTER 8

SUMMARY AND CONCLUSIONS

Summary

This dissertation presents convincing evidence that NAT1, whether directly or through an effect on acetyl-CoA levels, has a role in cellular metabolism and energetics in MDA-MB-231 breast cancer cells. Mitochondrial bioenergetics, specifically reserve capacity, maximal respiration, and glycolytic reserve capacity were increased in the NAT1 knockout cell lines (*CRISPR 2-19* and *CRISPR 5-50*) while bioenergetics in the cell line with increased NAT1 *N*-acetylation activity (*Up*) were not significantly different from the cell lines with parental NAT1 *N*-acetylation activity (*Parent* and *Scrambled*). Additionally, in both the metabolomics and transcriptomics experiments presented in this dissertation, metabolite abundances and gene transcript expression were more dysregulated in the NAT1 knockout cell lines (*CRISPR 2-19* and *CRISPR 5-50*) than the cell line with increased NAT1 (*Up*) compared to the cell line with parental NAT1 *N*-acetylation activity (*Scrambled*) suggesting the loss of NAT1 has a greater impact on metabolism than its overexpression.

By utilizing a global systems biology approach, new directions for future research on NAT1's role in breast cancer cell metabolism have been generated that would not have been considered otherwise. Systems biology allows a more unbiased view compared to traditional reductionist approaches, thus enabling NAT1 research to be stimulated in a novel direction. The data presented in this dissertation has also shown that the knockout of NAT1 induced more changes bioenergetically, metabolically, and transcriptionally compared to the overexpression of NAT1. As there is active research in the NAT field on the development of small molecule

inhibitors of NAT1, it is important to understand how the knockout of NAT1 *N*-acetylation activity and expression affects global cellular metabolism.

Strengths of This Work

The experimental design strategies and research presented in this dissertation have several strengths that enabled many novel insights as well as the formulation of additional hypotheses about NAT1's role in breast cancer and, more generally, metabolism. The first major strength I will discuss is the use of publicly available data to annotate previously undefined relationships between *NAT1*, *NAT2*, & *ESR1* in established breast cancer cell lines, primary breast tumors, and normal breast tissue. For these analyses data was accessed and analyzed (fairly quickly) without the need to do, what would have been, expensive wet-lab experiments. Publicly available data repositories offer a wealth of knowledge but, until now, have been under-utilized in NAT research. Additionally, the data repositories can be utilized to evaluate whether the relationships observed in the datasets presented in this dissertation are also present in the larger publicly available datasets.

A second strength of this work was the careful and thoughtful selection of samples. By choosing a single breast cancer cell line and genetically modifying only NAT1 (theoretically) any confounding factors that might be introduced when utilizing breast cancer cell lines with innately different levels of NAT1 activity due to different genetic backgrounds and the presence of additional unique mutations in each cell line have been eliminated. The MDA-MB-231 cell line was chosen rather than an estrogen receptor positive cell line because cell lines with NAT1 knocked out as well as overexpressed were desired. To this end, a breast cancer cell line with an approximate mid-level of NAT1 activity was selected. However, very few of the breast cancer cell lines available for research (approximately 57 total) have been characterized for NAT1 *N*-acetylation activity. Therefore, RNA expression was used as a proxy for NAT1 *N*-acetylation activity because publicly available data on NAT1 RNA expression of all 57 breast cancer cell lines are available (discussed and utilized in Chapter 2). The MDA-MB-231 cell line expresses an approximate mid-level of NAT1 RNA compared to the other 56 breast cancer cell lines. Additionally, this cell line is frequently utilized in breast cancer research and has been utilized in

previous studies looking at NAT1's role in breast cancer thus allowing comparison of experiments and observations more readily.

Additionally, during construction of the MDA-MB-231 breast cancer cell lines with varying levels of NAT1 *N*-acetylation activity two different methods, siRNA and CRISPR/Cas9, were utilized. This strategy helped to ensure any results or conclusions reported are truly due to differences in NAT1 activity and not a result of the methodology utilized to modulate NAT1 expression. To this end, during construction of the complete NAT1 knockout cell lines via CRISPR/Cas9 technology two unique NAT1 targeting guide RNAs were utilized, again ensuring any results or conclusions reported are truly due to differences in NAT1 activity because guide RNAs have been shown to cause off-target effects. In fact, as presented in Chapter 7, there is strong evidence of additional differences other than NAT1 in the genetically modified cell lines; without inclusion of NAT1 knockout cell lines constructed via different guide RNAs (*CRISPR 2-19* and *CRISPR 5-50*) and cell lines constructed with the same guide RNA but that had different levels of NAT1 *N*-acetylation activity (*CRISPR 2-12* and *CRISPR 2-19*), this observation would most likely have been missed and the results would be unknowingly confounded.

The third major advantage of the work presented in this dissertation is the systems biology approach utilized to evaluate the dissertation hypothesis rather than the more traditional reductionist approach. Notably, by using the exact same biological samples, collected and grown at the exact same time in the metabolomics and transcriptomics experiments we integrated the two resulting datasets on cells under identical biological conditions. Additionally, bioenergetics and transcriptomics experiments were integrated by comparing mitochondrial gene expression. A global systems biology approach has enabled stimulation of NAT1 research in new directions.

Caveats and Weaknesses

Although the work presented in this dissertation was well planned and executed there are nevertheless caveats and weaknesses. The first caveat to acknowledge is that although this dissertation asserts that a global analysis of the complete system was performed, that is not entirely true. The field of metabolomics is limited by the number of metabolites that can be detected analytically. Therefore, that assertion must be qualified to acknowledge that this study

was a global metabolomics experiment in that data was collected on as many metabolites as was feasible which still represents a small proportion of all metabolites present in the human metabolome (Figure 8.1). Additionally, because metabolites were only isolated from cell pellets and not the surrounding media, metabolites excreted from the cell that would be important in understanding NAT1's effects on cellular metabolism may have been missed. However, media was collected and reserved during collection of cell pellet samples for possible future analyses.

Another caveat to acknowledge is the bias introduced by reliance on NAT1 *N*-acetylation activity as a measure for NAT1 function. Although NAT1 *N*-acetylation is increased, decreased, and not detectable in the constructed cell lines, it remains unknown if possible additional functions/roles of NAT1 were not also affected and contribute to our overall observations in an as yet uncharacterized manner. It will also be important for future studies to evaluate the findings presented in this dissertation in additional breast cancer cell lines, untransformed (normal) breast cell lines, and other cancers.

Notably, although sample choice was a major strength in this work it can also be viewed as a weakness. The conclusions presented in this dissertation are based on results derived from modulating NAT1 in a single breast cancer cell line; however, breast cancer is a very heterogenous disease. The MDA-MB-231 breast cancer cell line may have unique mutations in genes that are regulators, interactors, producing or metabolizing substrates for NAT1, or altering metabolism independent of NAT1 status. It will be important for future work to conduct experiments in other breast cancer cell lines and more generally, other cancer types where NAT1 is commonly overexpressed, to determine if our observations hold true. Although human breast tumor samples would have been more biologically relevant, as the tumor microenvironment would have been better represented, human samples would add more confounding factors when interpreting results. Additionally, there is a great need to analyze data from human subjects to see if these observations correlate with samples from a complete human system. Similarly, the research presented here was conducted on samples that were already malignant in nature, making inferences about NAT1's role in breast cancer risk or tumorigenesis not possible. Results

Figure 8.1

A discrete pathway

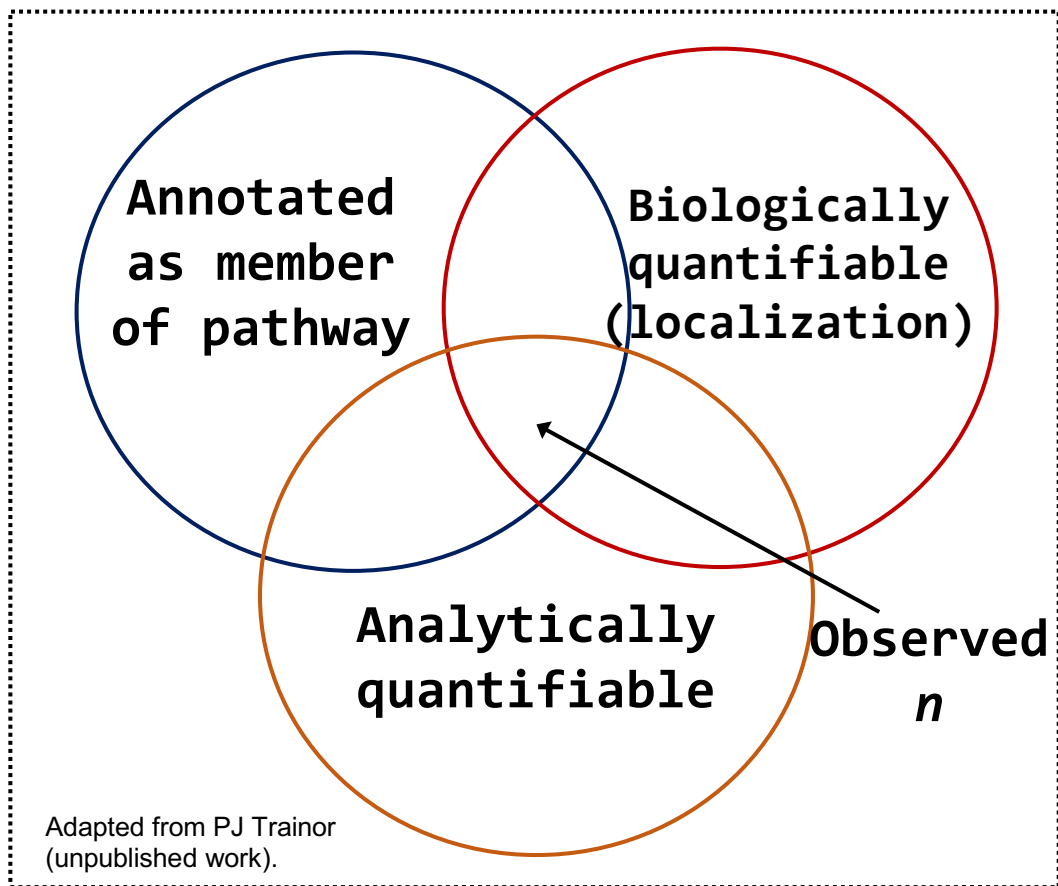


Figure 8.1: Limitations of “Global” Metabolomics.

A Venn diagram depicting a significant issue in “global” metabolomics. Researchers are limited by metabolite annotation, the ability to analytically detect metabolites, and the localization of metabolites. When attempting to make inference about changes in metabolic pathways metabolites in the intersection of the three circles are observed when the target metabolic pathway is the entire rectangle. These difficulties limit the ability to obtain a true global view of metabolism however there is abundant research being conducted in the field to improve all three.

may be different if experiments were performed on non-malignant breast cells and further studies in non-malignant cells are necessary.

Another weakness of the work presented here was the presence of differences other than NAT1 in the constructed cell lines (presented in Chapter 7). These additional differences complicated interpretations of the data. However, without a global systems biology approach, we would not have realized there were additional differences and may have incorrectly attributed results to the modulation of NAT1. It is the author's opinion that other CRISPR/Cas9 studies also have off-target effects. However, the measurements or observations that are made do not allow a high-level global view so these off-target effects may be missed. Additionally, inclusion of two NAT1 knockout cell lines that were generated using two different guide RNAs differences that were truly due to the modulation of NAT1 to be focused on. In the future, whole genome sequencing of the cell lines utilized in this study should be conducted to determine exactly which genes these additional differences have occurred in. This may aid interpretation of some of the results presented in this dissertation.

The final weakness of this work to be discussed was the inability to stably rescue NAT1 in the constructed NAT1 knockout cell lines. This rescue would have allowed greater confidence that observations and results were truly due to the knockout of NAT1. It is predicted that any observations truly due to NAT1 knockout would be reversed with the rescue of NAT1 activity and conversely, predicted that any observations due to differences other than NAT1 would not be reversed by the rescue of NAT1.

Future Directions

In addition to the results and conclusions that were presented in this dissertation, many novel hypotheses about the role of NAT1 in breast cancer, and more generally cellular metabolism, have been generated. As such, I propose numerous future directions to further delineate the mechanism by which NAT1 affects breast cancer progression and the possible role of NAT1 in cellular metabolism.

The first future direction, which is currently in progress, is to incubate the cell lines used in this study with L-asparagine to measure how this supplementation affects cell growth and

invasive ability, given the strong positive correlation between NAT1 *N*-acetylation activity and *N*-acetylasparagine and the recently proposed role of L-asparagine in breast cancer cell growth and metastasis^{189,190}. Although our lab currently does not have the ability to readily measure L-asparagine and/or *N*-acetylasparagine, the proposed experiments will help determine if cell lines with varying NAT1 *N*-acetylation activities respond differently to L-asparagine supplementation. If differences in growth and metastatic are observed based on NAT1 activity, that will add evidence to the hypothesis that there is a relationship between NAT1 and L-asparagine. Additionally, our lab is currently working with collaborators to develop an HPLC-fluorescence detection method to detect *N*-acetylasparagine and L-asparagine. Once this method is optimized, recombinant human NAT1 expressed in yeast lysates could be incubated with acetyl-CoA and L-asparagine to evaluate if NAT1 is directly catalyzing the *N*-acetylation of L-asparagine more directly. Until then, our lab is conducting competitive activity assays between PABA and L-asparagine to see if the rate of *N*-acetylation of PABA is affected in the presence of L-asparagine in the reaction. Similarly, studies using putrescine instead of L-asparagine could be performed given there is evidence both biochemicals may be NAT1 substrates.

An additional area of future research will be further evaluating the mechanism by which NAT2 transcripts are increased in NAT1 knockout cell lines (*CRISPR 2-19* and *CRISPR 5-50*). From the results of the transcriptomics data, a list of candidate genes that are predicted to interact with NAT2 has been created. The candidate genes were chosen based on correlated expression with NAT2 expression and also based on clustering with NAT2 in the WGCNA analysis. While this phenomenon was observed in the constructed breast cancer cell lines, it is unclear if this occurs in humans. I propose analyzing NAT1 and NAT2 gene expression data in human subjects with no detectable NAT1 activity to see if the phenomenon is present. One foreseeable limitation is the relatively rare occurrence of humans with no detectable NAT1 activity. While functional NAT2 *N*-acetylation activity was not observed in the NAT1 knockout cell lines, investigating the mechanism by which NAT2 transcripts are produced will be an important area of NAT research in the future. An additional area of research will be investigating miRNAs

that possibly target NAT2 to help explain the observation of NAT2 transcripts but no NAT2 *N*-acetylation activity.

On the basis that many of our observations and results can be traced to reactions that involve mitochondria, an additional future direction is to probe mitochondrial mass, biogenesis, and morphology in our constructed cell lines. Additionally, reactive oxygen species (ROS) in mitochondria can be evaluated. This will help determine if the observations presented in this dissertation are related to changes in mitochondria biogenesis/number directly or dysregulation of the reactions that occur in mitochondria.

Given the strong evidence of genetic differences in addition to NAT1 in the constructed cell lines utilized in the studies presented in this dissertation, I propose performing WGS to determine the exact location of the differences. This method is the only way to definitively determine the genetic differences between the constructed cell lines. Once additional genetic differences between the cell lines are determined, that information may be utilized to study those genes. Additionally, the results of this proposed experiment will have great implications for past, present, and future studies utilizing CRISPR/Cas9.

The final future direction proposed is integrating the three datasets presented in this dissertation with the proteomics and acetylomics previously collected by our lab (collected only for the complete NAT1 knockout cell lines, *CRISPR 2-19* and *CRISPR 5-50*). As human metabolism is a very highly regulated process and composed of many layers, combining these datasets would add an additional level to the systems biology approach presented in this dissertation. Combining the datasets will not be easy but improved methodology for combining multi-omics datasets is currently a rich area of development (discussed in Chapter 6) in the systems biology field.

Overall Conclusions

This dissertation has presented evidence that greater caution should be exercised in interpretation of results from single gene knockout studies utilizing CRISPR/Cas9. The elegant study design utilized in this dissertation enabled me to conclude CRISPR/Cas9 can cause off-target effects in sites not predicted by web-based algorithms that would confound conclusions.

Furthermore, I have shown that publicly available datasets can be used to complement and enhance wet-lab experiments very quickly and inexpensively.

I have presented strong evidence that NAT1, whether directly or through the effect of NAT1 on acetyl-CoA levels, has an effect on acyl-CoA carnitine conjugates, lysine degradation, and mitochondrial function (Figure 6.11). Saccharopine is the first product formed during lysine degradation in mammals and was increased in cell lines with decreased levels and knockout NAT1 (*Down*, *CRISPR 2-12*, *CRISPR 2-19*, and *CRISPR 5-50*). Notably, carnitine is biosynthesized from methionine and lysine. Increases in saccharopine suggest more lysine is degraded, leaving less lysine for carnitine biosynthesis. Fatty acyl-CoA carnitine conjugate metabolites were decreased in the NAT1 knockout cell lines (*CRISPR 2-19* and *CRISPR 5-50*) and those metabolites were strongly correlated with carnitine abundance. In humans, fatty acyl-CoAs are conjugated to carnitine for transport to the mitochondria and subsequent beta oxidation producing acetyl-CoA. The decrease in fatty acyl-CoA conjugates in the NAT1 knockout cell lines (*CRISPR 2-19* and *CRISPR 5-50*) suggest less acetyl-CoA is produced by this mitochondrial pathway. It remains unknown whether this is a dysregulation of energetics or complex feedback due to increased acetyl-CoA because of decreased NAT1 catalyzed hydrolysis of acetyl-CoA.

Overall, this dissertation describes the first studies to investigate the effect varying human NAT1 activity has on the bioenergetics profile, metabolome, and transcriptome of a breast cancer cell line. This dissertation has generated vast amounts of data that I foresee being utilized for years to come as we learn more about NAT1 thus providing a valuable resource for future studies. Additionally, I have used this data to generate novel hypotheses about the role of NAT1 in breast cancer, and more generally cellular metabolism. Furthermore, I have identified biochemicals that are likely products of NAT1 *N*-acetylation, *N*-acetylasparagine and *N*-acetylputrescine, however further studies are needed to confirm this. While I have not identified the exact mechanism by which NAT1 affects cellular metabolism or breast cancer progression, the data presented in this dissertation add important pieces to the puzzle, putting NAT researchers one step closer to that goal.

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA: a cancer journal for clinicians*. 2018;68(1):7-30.
2. National Cancer Institute. NCI-MATCH Trial (Molecular Analysis for Therapy Choice) www.cancer.gov/about-cancer/treatment/clinical-trials/nci-supported/nci-match. Accessed 06-09-2018.
3. Hein DW. Molecular genetics and function of NAT1 and NAT2: role in aromatic amine metabolism and carcinogenesis. *Mutation research*. 2002;506-507:65-77.
4. Hein DW, Doll MA, Fretland AJ, et al. Molecular genetics and epidemiology of the NAT1 and NAT2 acetylation polymorphisms. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*. 2000;9(1):29-42.
5. Matas N, Thygesen P, Stacey M, Risch A, Sim E. Mapping AAC1, AAC2 and AACP, the genes for arylamine N-acetyltransferases, carcinogen metabolising enzymes on human chromosome 8p22, a region frequently deleted in tumours. *Cytogenetics and cell genetics*. 1997;77(3-4):290-295.
6. Blum M, Grant DM, McBride W, Heim M, Meyer UA. Human arylamine N-acetyltransferase genes: isolation, chromosomal localization, and functional expression. *DNA and cell biology*. 1990;9(3):193-203.
7. Ohsako S, Deguchi T. Cloning and expression of cDNAs for polymorphic and monomorphic arylamine N-acetyltransferases from human liver. *The Journal of biological chemistry*. 1990;265(8):4630-4634.
8. Kawamura A, Graham J, Mushtaq A, et al. Eukaryotic arylamine N-acetyltransferase. Investigation of substrate specificity by high-throughput screening. *Biochemical pharmacology*. 2005;69(2):347-359.
9. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans., World Health Organization., International Agency for Research on Cancer. *Some aromatic amines, organic dyes, and related exposures*. Lyon: IARC Press; 2010.
10. Baan R, Straif K, Grosse Y, et al. Special Report: Policy: Carcinogenicity of some aromatic amines, organic dyes, and related exposures. *Lancet Oncol*. 2008;9(4):322-323.
11. Blömeke B, Lichter J. Expression and Activity of Arylamine N-Acetyltransferases in Organs: Implications on Aromatic Amine Toxicity. In: Laurieri N, Sim, E., ed. *Arylamine N-Acetyltransferases in Health and Disease*: World Scientific Publishing; 2018:133-164.
12. Grant DM, Blum M, Beer M, Meyer UA. Monomorphic and polymorphic human arylamine N-acetyltransferases: a comparison of liver isozymes and expressed products of two cloned genes. *Molecular pharmacology*. 1991;39(2):184-191.
13. Sadrieh N, Davis CD, Snyderwine EG. N-acetyltransferase expression and metabolic activation of the food-derived heterocyclic amines in the human mammary gland. *Cancer research*. 1996;56(12):2683-2687.
14. Lee JH, Chung JG, Lai JM, Levy GN, Weber WW. Kinetics of arylamine N-acetyltransferase in tissues from human breast cancer. *Cancer letters*. 1997;111(1-2):39-50.
15. Pacifici GM, Bencini C, Rane A. Acetyltransferase in humans: development and tissue distribution. *Pharmacology*. 1986;32(5):283-291.
16. Cribb AE, Grant DM, Miller MA, Spielberg SP. Expression of monomorphic arylamine N-acetyltransferase (NAT1) in human leukocytes. *The Journal of pharmacology and experimental therapeutics*. 1991;259(3):1241-1246.

17. Ward A, Hickman D, Gordon JW, Sim E. Arylamine N-acetyltransferase in human red blood cells. *Biochemical pharmacology*. 1992;44(6):1099-1104.
18. Badawi AF, Hirvonen A, Bell DA, Lang NP, Kadlubar FF. Role of aromatic amine acetyltransferases, NAT1 and NAT2, in carcinogen-DNA adduct formation in the human urinary bladder. *Cancer research*. 1995;55(22):5230-5237.
19. Hickman D, Pope J, Patil SD, et al. Expression of arylamine N-acetyltransferase in human intestine. *Gut*. 1998;42(3):402-409.
20. Salazar-Gonzalez RA, Turijan-Espinoza E, Hein DW, et al. Arylamine N-acetyltransferase 1 in situ N-acetylation on CD3+ peripheral blood mononuclear cells correlate with NATb mRNA and NAT1 haplotype. *Archives of toxicology*. 2017.
21. Boukouvala S, Sim E. Structural analysis of the genes for human arylamine N-acetyltransferases and characterisation of alternative transcripts. *Basic & clinical pharmacology & toxicology*. 2005;96(5):343-351.
22. Smelt VA, Upton A, Adjaye J, et al. Expression of arylamine N-acetyltransferases in pre-term placentas and in human pre-implantation embryos. *Human molecular genetics*. 2000;9(7):1101-1107.
23. Vatsis KP, Weber WW. Structural heterogeneity of Caucasian N-acetyltransferase at the NAT1 gene locus. *Archives of biochemistry and biophysics*. 1993;301(1):71-76.
24. Weber WW, Vatsis KP. Individual variability in p-aminobenzoic acid N-acetylation by human N-acetyltransferase (NAT1) of peripheral blood. *Pharmacogenetics*. 1993;3(4):209-212.
25. Grant DM, Hughes NC, Janezic SA, et al. Human acetyltransferase polymorphisms. *Mutation research*. 1997;376(1-2):61-70.
26. Cloete R, Akurugu WA, Werely CJ, van Helden PD, Christoffels A. Structural and functional effects of nucleotide variation on the human TB drug metabolizing enzyme arylamine N-acetyltransferase 1. *J Mol Graph Model*. 2017;75:330-339.
27. McDonagh EM, Boukouvala S, Aklillu E, Hein DW, Altman RB, Klein TE. PharmGKB summary: very important pharmacogene information for N-acetyltransferase 2. *Pharmacogenetics and genomics*. 2014;24(8):409-425.
28. Butcher NJ, Boukouvala S, Sim E, Minchin RF. Pharmacogenetics of the arylamine N-acetyltransferases. *The pharmacogenomics journal*. 2002;2(1):30-42.
29. Hein DW. N-acetyltransferase SNPs: emerging concepts serve as a paradigm for understanding complexities of personalized medicine. *Expert opinion on drug metabolism & toxicology*. 2009;5(4):353-366.
30. Boukouvala S. Human NAT1 Alleles (Haplotypes). 2017; http://nat.mbg.duth.gr/Human%20NAT1%20alleles_2013.htm. Accessed 12-11-17, 2017.
31. Boukouvala S. Human NAT2 Alleles (Haplotypes). 2017; http://nat.mbg.duth.gr/Human%20NAT2%20alleles_2013.htm. Accessed 12-11-17, 2017.
32. Upton A, Johnson N, Sandy J, Sim E. Arylamine N-acetyltransferases - of mice, men and microorganisms. *Trends in pharmacological sciences*. 2001;22(3):140-146.
33. Boukouvala S, Fakis G. Arylamine N-acetyltransferases: what we learn from genes and genomes. *Drug metabolism reviews*. 2005;37(3):511-564.
34. Hughes NC, Janezic SA, McQueen KL, et al. Identification and characterization of variant alleles of human acetyltransferase NAT1 with defective function using p-aminosalicylate as an in-vivo and in-vitro probe. *Pharmacogenetics*. 1998;8(1):55-66.
35. Payton MA, Sim E. Genotyping human arylamine N-acetyltransferase type 1 (NAT1): the identification of two novel allelic variants. *Biochemical pharmacology*. 1998;55(3):361-366.
36. Smelt VA, Mardon HJ, Sim E. Placental expression of arylamine N-acetyltransferases: evidence for linkage disequilibrium between NAT1*10 and NAT2*4 alleles of the two human arylamine N-acetyltransferase loci NAT1 and NAT2. *Pharmacology & toxicology*. 1998;83(4):149-157.
37. Butcher NJ, Tiang J, Minchin RF. Regulation of arylamine N-acetyltransferases. *Current drug metabolism*. 2008;9(6):498-504.

38. Husain A, Barker DF, States JC, Doll MA, Hein DW. Identification of the major promoter and non-coding exons of the human arylamine N-acetyltransferase 1 gene (NAT1). *Pharmacogenetics*. 2004;14(7):397-406.
39. Barker DF, Husain A, Neale JR, et al. Functional properties of an alternative, tissue-specific promoter for human arylamine N-acetyltransferase 1. *Pharmacogenetics and genomics*. 2006;16(7):515-525.
40. Husain A, Zhang X, Doll MA, States JC, Barker DF, Hein DW. Functional analysis of the human N-acetyltransferase 1 major promoter: quantitation of tissue expression and identification of critical sequence elements. *Drug metabolism and disposition: the biological fate of chemicals*. 2007;35(9):1649-1656.
41. Butcher NJ, Arulpragasam A, Goh HL, Davey T, Minchin RF. Genomic organization of human arylamine N-acetyltransferase Type I reveals alternative promoters that generate different 5'-UTR splice variants with altered translational activities. *The Biochemical journal*. 2005;387(Pt 1):119-127.
42. Butcher NJ, Ilett KF, Minchin RF. Substrate-dependent regulation of human arylamine N-acetyltransferase-1 in cultured cells. *Molecular pharmacology*. 2000;57(3):468-473.
43. Butcher NJ, Ilett KF, Minchin RF. Inactivation of human arylamine N-acetyltransferase 1 by the hydroxylamine of p-aminobenzoic acid. *Biochemical pharmacology*. 2000;60(12):1829-1836.
44. Butcher NJ, Minchin RF. Arylamine N-acetyltransferase 1 gene regulation by androgens requires a conserved heat shock element for heat shock factor-1. *Carcinogenesis*. 2010;31(5):820-826.
45. Butcher NJ, Tetlow NL, Cheung C, Broadhurst GM, Minchin RF. Induction of human arylamine N-acetyltransferase type I by androgens in human prostate cancer cells. *Cancer research*. 2007;67(1):85-92.
46. Lee JH, Lu HF, Wang DY, et al. Effects of tamoxifen on DNA adduct formation and arylamines N-acetyltransferase activity in human breast cancer cells. *Research communications in molecular pathology and pharmacology*. 2004;115-116:217-233.
47. Ragunathan N, Dairou J, Pluvinage B, et al. Identification of the xenobiotic-metabolizing enzyme arylamine N-acetyltransferase 1 as a new target of cisplatin in breast cancer cells: molecular and cellular mechanisms of inhibition. *Molecular pharmacology*. 2008;73(6):1761-1768.
48. Xu X, Mathieu C, Berthelet J, et al. Human Arylamine N-Acetyltransferase 1 Is Inhibited by the Dithiocarbamate Pesticide Thiram. *Molecular pharmacology*. 2017;92(3):358-365.
49. Duval R, Xu X, Bui LC, et al. Identification of cancer chemopreventive isothiocyanates as direct inhibitors of the arylamine N-acetyltransferase-dependent acetylation and bioactivation of aromatic amine carcinogens. *Oncotarget*. 2016;7(8):8688-8699.
50. Dupret JM, Dairou J, Atmane N, Rodrigues-Lima F. Inactivation of human arylamine N-acetyltransferase 1 by hydrogen peroxide and peroxyxynitrite. *Methods in enzymology*. 2005;400:215-229.
51. Dairou J, Atmane N, Rodrigues-Lima F, Dupret JM. Peroxynitrite irreversibly inactivates the human xenobiotic-metabolizing enzyme arylamine N-acetyltransferase 1 (NAT1) in human breast cancer cells: a cellular and mechanistic study. *The Journal of biological chemistry*. 2004;279(9):7708-7714.
52. Bui LC, Manaa A, Xu X, et al. Acrolein, an alpha,beta-unsaturated aldehyde, irreversibly inhibits the acetylation of aromatic amine xenobiotics by human arylamine N-acetyltransferase 1. *Drug metabolism and disposition: the biological fate of chemicals*. 2013;41(7):1300-1305.
53. Sanfins E, Dairou J, Hussain S, et al. Carbon black nanoparticles impair acetylation of aromatic amine carcinogens through inactivation of arylamine N-acetyltransferase enzymes. *ACS nano*. 2011;5(6):4504-4511.
54. Ragunathan N, Dairou J, Sanfins E, et al. Cadmium alters the biotransformation of carcinogenic aromatic amines by arylamine N-acetyltransferase xenobiotic-metabolizing enzymes: molecular, cellular, and in vivo studies. *Environmental health perspectives*. 2010;118(12):1685-1691.

55. Malka F, Dairou J, Ragunathan N, Dupret JM, Rodrigues-Lima F. Mechanisms and kinetics of human arylamine N-acetyltransferase 1 inhibition by disulfiram. *The FEBS journal*. 2009;276(17):4900-4908.
56. Xiao X, Yang D, Gong X, Mo D, Pan S, Xu J. miR-1290 promotes lung adenocarcinoma cell proliferation and invasion by targeting SOCS4. *Oncotarget*. 2018;9(15):11977-11988.
57. Lai XJ, Cheng HF. LncRNA colon cancer-associated transcript 1 (CCAT1) promotes proliferation and metastasis of ovarian cancer via miR-1290. *Eur Rev Med Pharmacol Sci*. 2018;22(2):322-328.
58. Jin JJ, Liu YH, Si JM, Ni R, Wang J. Overexpression of miR-1290 contributes to cell proliferation and invasion of non small cell lung cancer by targeting interferon regulatory factor 2. *The international journal of biochemistry & cell biology*. 2018;95:113-120.
59. Kim KB, Kim K, Bae S, et al. MicroRNA-1290 promotes asiatic acid-induced apoptosis by decreasing BCL2 protein level in A549 nonsmall cell lung carcinoma cells. *Oncology reports*. 2014;32(3):1029-1036.
60. Yan L, Cai K, Sun K, Gui J, Liang J. MiR-1290 promotes proliferation, migration, and invasion of glioma cells by targeting LHX6. *Journal of cellular physiology*. 2017.
61. Ma Q, Wang Y, Zhang H, Wang F. MiR-1290 Contributes to Colorectal Cancer Cell Proliferation by Targeting INPP4B. *Oncology research*. 2017.
62. Endo Y, Toyama T, Takahashi S, et al. miR-1290 and its potential targets are associated with characteristics of estrogen receptor alpha-positive breast cancer. *Endocrine-related cancer*. 2013;20(1):91-102.
63. Endo Y, Yamashita H, Takahashi S, et al. Immunohistochemical determination of the miR-1290 target arylamine N-acetyltransferase 1 (NAT1) as a prognostic biomarker in breast cancer. *BMC cancer*. 2014;14:990.
64. Marshall EK, Jr., Cutting WC, Emerson K, Jr. Acetylation of Para-Aminobenzenesulfonamide in the Animal Organism. *Science*. 1937;85(2199):202-203.
65. Lipmann F. Acetylation of sulfanilamide by liver homogenates and extracts. *J Biol Chem* 1945;160:173-190.
66. Johnson WJ. Biological acetylation of isoniazid. *Nature*. 1954;174(4433):744-745.
67. Weber WW, Cohen SN. N-acetylation of drugs: isolation and properties of an N-acetyltransferase from rabbit liver. *Molecular pharmacology*. 1967;3(3):266-273.
68. Bulovskaya LN, Krupkin RG, Bochina TA, Shipkova AA, Pavlova MV. Acetylator phenotype in patients with breast cancer. *Oncology*. 1978;35(4):185-188.
69. Ladero JM, Fernandez MJ, Palmeiro R, et al. Hepatic acetylator polymorphism in breast cancer patients. *Oncology*. 1987;44(6):341-344.
70. Philip PA, Rogers HJ, Millis RR, Rubens RD, Cartwright RA. Acetylator status and its relationship to breast cancer and other diseases of the breast. *Eur J Cancer Clin Oncol*. 1987;23(11):1701-1706.
71. Webster DJ, Flook D, Jenkins J, Hutchings A, Routledge PA. Drug acetylation in breast cancer. *British journal of cancer*. 1989;60(2):236-237.
72. Ilett KF, Detchon P, Ingram DM, Castleden WM. Acetylation phenotype is not associated with breast cancer. *Cancer research*. 1990;50(20):6649-6651.
73. Wang T, Marei HE. Landscape of NAT2 polymorphisms among breast cancer. *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie*. 2016;77:191-196.
74. Zhang J, Qiu LX, Wang ZH, Wang JL, He SS, Hu XC. NAT2 polymorphisms combining with smoking associated with breast cancer susceptibility: a meta-analysis. *Breast cancer research and treatment*. 2010;123(3):877-883.
75. Ambrosone CB, Kropp S, Yang J, Yao S, Shields PG, Chang-Claude J. Cigarette smoking, N-acetyltransferase 2 genotypes, and breast cancer risk: pooled analysis and meta-analysis. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*. 2008;17(1):15-26.
76. Ochs-Balcom HM, Wiesner G, Elston RC. A meta-analysis of the association of N-acetyltransferase 2 gene (NAT2) variants with breast cancer. *American journal of epidemiology*. 2007;166(3):246-254.

77. Terry PD, Goodman M. Is the association between cigarette smoking and breast cancer modified by genotype? A review of epidemiologic studies and meta-analysis. *Cancer Epidem Biomar.* 2006;15(4):602-611.
78. Zhang K, Gao L, Wu Y, et al. NAT1 polymorphisms and cancer risk: a systematic review and meta-analysis. *Int J Clin Exp Med.* 2015;8(6):9177-9191.
79. Tiang JM, Butcher NJ, Minchin RF. Effects of human arylamine N-acetyltransferase I knockdown in triple-negative breast cancer cell lines. *Cancer Med.* 2015;4(4):565-574.
80. Tiang JM, Butcher NJ, Cullinane C, Humbert PO, Minchin RF. RNAi-mediated knock-down of arylamine N-acetyltransferase-1 expression induces E-cadherin up-regulation and cell-cell contact growth inhibition. *PLoS one.* 2011;6(2):e17031.
81. Tiang JM, Butcher NJ, Minchin RF. Small molecule inhibition of arylamine N-acetyltransferase Type I inhibits proliferation and invasiveness of MDA-MB-231 breast cancer cells. *Biochemical and biophysical research communications.* 2010;393(1):95-100.
82. Stepp MW, Doll MA, Carlisle SM, States JC, Hein DW. Genetic and small molecule inhibition of arylamine N-acetyltransferase 1 reduces anchorage-independent growth in human breast cancer cell line MDA-MB-231. *Molecular carcinogenesis.* 2018;57(4):549-558.
83. Stepp MW, Doll MA, Samuelson DJ, Sanders MA, States JC, Hein DW. Congenic rats with higher arylamine N-acetyltransferase 2 activity exhibit greater carcinogen-induced mammary tumor susceptibility independent of carcinogen metabolism. *BMC cancer.* 2017;17(1):233.
84. Carlisle SM, Trainor PJ, Yin X, et al. Untargeted polar metabolomics of transformed MDA-MB-231 breast cancer cells expressing varying levels of human arylamine N-acetyltransferase 1. *Metabolomics : Official journal of the Metabolomic Society.* 2016;12(7).
85. Perou CM, Jeffrey SS, van de Rijn M, et al. Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. *Proceedings of the National Academy of Sciences of the United States of America.* 1999;96(16):9212-9217.
86. Zhao H, Langerod A, Ji Y, et al. Different gene expression patterns in invasive lobular and ductal carcinomas of the breast. *Molecular biology of the cell.* 2004;15(6):2523-2536.
87. Wang Y, Klijn JG, Zhang Y, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet.* 2005;365(9460):671-679.
88. Wakefield L, Robinson J, Long H, et al. Arylamine N-acetyltransferase 1 expression in breast cancer cell lines: a potential marker in estrogen receptor-positive tumors. *Genes, chromosomes & cancer.* 2008;47(2):118-126.
89. Bieche I, Girault I, Urbain E, Tozlu S, Lidereau R. Relationship between intratumoral expression of genes coding for xenobiotic-metabolizing enzymes and benefit from adjuvant tamoxifen in estrogen receptor alpha-positive postmenopausal breast carcinoma. *Breast cancer research : BCR.* 2004;6(3):R252-263.
90. Tozlu S, Girault I, Vacher S, et al. Identification of novel genes that co-cluster with estrogen receptor alpha in breast tumor biopsy specimens, using a large-scale real-time reverse transcription-PCR approach. *Endocrine-related cancer.* 2006;13(4):1109-1120.
91. Abba MC, Hu Y, Sun H, et al. Gene expression signature of estrogen receptor alpha status in breast cancer. *BMC genomics.* 2005;6:37.
92. Zhang X, Carlisle SM, Doll MA, et al. High N-acetyltransferase 1 (NAT1) expression is associated with estrogen receptor expression in breast tumors, but is not under direct regulation by estradiol, 17beta-diol, or dihydrotestosterone in breast cancer cells. *The Journal of pharmacology and experimental therapeutics.* 2018.
93. Parker JS, Mullins M, Cheang MC, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology.* 2009;27(8):1160-1167.
94. Wallden B, Storhoff J, Nielsen T, et al. Development and verification of the PAM50-based Prosigna breast cancer gene signature assay. *BMC medical genomics.* 2015;8:54.

95. Hein DW. N-acetyltransferase 2 polymorphism and human urinary bladder and breast cancer risks. In: Sim EaL, N., ed. *Arylamine N-Acetyltransferases in Health and Disease*: World Scientific Publishing; 2018:327-349.
96. Husain A, Zhang X, Doll MA, States JC, Barker DF, Hein DW. Identification of N-acetyltransferase 2 (NAT2) transcription start sites and quantitation of NAT2-specific mRNA in human tissues. *Drug metabolism and disposition: the biological fate of chemicals*. 2007;35(5):721-727.
97. Geylan YS, Dizbay S, Guray T. Arylamine N-acetyltransferase activities in human breast cancer tissues. *Neoplasma*. 2001;48(2):108-111.
98. Geylan-Su YS, Isgor B, Coban T, et al. Comparison of NAT1, NAT2 and GSTT2-2 activities in normal and neoplastic human breast tissues. *Neoplasma*. 2006;53(1):73-78.
99. Deitz AC. *N-acetyltransferase Genetic Polymorphisms And Breast Cancer Risk* [Dissertation]: Pharmacology & Toxicology, University of North Dakota; 1999.
100. Jefferson FA, Xiao GH, Hein DW. 4-Aminobiphenyl downregulation of NAT2 acetylator genotype-dependent N- and O-acetylation of aromatic and heterocyclic amine carcinogens in primary mammary epithelial cell cultures from rapid and slow acetylator rats. *Toxicological sciences : an official journal of the Society of Toxicology*. 2009;107(1):293-297.
101. Bradshaw TD, Chua MS, Orr S, Matthews CS, Stevens MF. Mechanisms of acquired resistance to 2-(4-aminophenyl)benzothiazole (CJM 126, NSC 34445). *British journal of cancer*. 2000;83(2):270-277.
102. Minchin RF, Butcher NJ. Trimodal distribution of arylamine N-acetyltransferase 1 mRNA in breast cancer tumors: association with overall survival and drug resistance. *BMC genomics*. 2018;19(1):513.
103. Sabbagh A, Marin J, Veyssiere C, et al. Rapid birth-and-death evolution of the xenobiotic metabolizing NAT gene family in vertebrates with evidence of adaptive selection. *BMC Evol Biol*. 2013;13:62.
104. Glenn AE, Karagianni EP, Uldreaj A, Boukouvala S. Comparative genomic and phylogenetic investigation of the xenobiotic metabolizing arylamine N-acetyltransferase enzyme family. *FEBS letters*. 2010;584(14):3158-3164.
105. Derewlany LO, Knie B, Koren G. Arylamine N-acetyltransferase activity of the human placenta. *The Journal of pharmacology and experimental therapeutics*. 1994;269(2):756-760.
106. Derewlany LO, Knie B, Koren G. Human placental transfer and metabolism of p-aminobenzoic acid. *The Journal of pharmacology and experimental therapeutics*. 1994;269(2):761-765.
107. Witham KL, Minchin RF, Butcher NJ. Role for human arylamine N-acetyltransferase 1 in the methionine salvage pathway. *Biochemical pharmacology*. 2017;125:93-100.
108. Minchin RF. Acetylation of p-aminobenzoylglutamate, a folic acid catabolite, by recombinant human arylamine N-acetyltransferase and U937 cells. *The Biochemical journal*. 1995;307 (Pt 1):1-3.
109. Ward A, Summers MJ, Sim E. Purification of recombinant human N-acetyltransferase type 1 (NAT1) expressed in E. coli and characterization of its potential role in folate metabolism. *Biochemical pharmacology*. 1995;49(12):1759-1767.
110. Wakefield L, Cornish V, Long H, Griffiths WJ, Sim E. Deletion of a xenobiotic metabolizing gene in mice affects folate metabolism. *Biochemical and biophysical research communications*. 2007;364(3):556-560.
111. Jensen LE, Hoess K, Mitchell LE, Whitehead AS. Loss of function polymorphisms in NAT1 protect against spina bifida. *Human genetics*. 2006;120(1):52-57.
112. Stepp MW, Mamaliga G, Doll MA, States JC, Hein DW. Folate-Dependent Hydrolysis of Acetyl-Coenzyme A by Recombinant Human and Rodent Arylamine N-Acetyltransferases. *Biochem Biophys Rep*. 2015;3:45-50.
113. Laurieri N, Dairou J, Egleton JE, et al. From arylamine N-acetyltransferase to folate-dependent acetyl CoA hydrolase: impact of folic acid on the activity of (HUMAN)NAT1 and its homologue (MOUSE)NAT2. *PLoS one*. 2014;9(5):e96370.

114. Carlisle SM, Hein, D.W. Retrospective analysis of estrogen receptor 1 and N-acetyltransferase gene expression in normal breast tissue, primary breast tumors, and established breast cancer cell lines. *Int J Oncol*. 2018.
115. Stepp MW. *Role of human arylamine N-acetyltransferase 1 in tumorigenesis and cancer biology* [Doctoral Dissertation]. Electronic Theses and Dissertations: Pharmacology and Toxicology, University of Louisville; 2017.
116. Carlisle SM. *Metabolomics of transformed MDA-MB-231 cell lines expressing different levels of human arylamine N-acetyltransferase 1 (NAT1)* [Master's Thesis]. Electronic Theses and Dissertations: Pharmacology and Toxicology, University of Louisville; 2015.
117. Leff MA, Epstein PN, Doll MA, et al. Prostate-specific human N-acetyltransferase 2 (NAT2) expression in the mouse. *The Journal of pharmacology and experimental therapeutics*. 1999;290(1):182-187.
118. Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell*. 2000;100(1):57-70.
119. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011;144(5):646-674.
120. Shi L, Tu BP. Acetyl-CoA and the regulation of metabolism: mechanisms and consequences. *Current opinion in cell biology*. 2015;33:125-131.
121. Lee JV, Shah SA, Wellen KE. Obesity, cancer and acetyl-CoA metabolism. *Drug Discovery Today: Disease Mechanisms*. 2013;10(1-2):e55-e61.
122. Pietrocola F, Galluzzi L, Bravo-San Pedro JM, Madeo F, Kroemer G. Acetyl coenzyme A: a central metabolite and second messenger. *Cell metabolism*. 2015;21(6):805-821.
123. Aderem A. Systems biology: its practice and challenges. *Cell*. 2005;121(4):511-513.
124. Karahalil B. Overview of Systems Biology and Omics Technologies. *Current medicinal chemistry*. 2016;23(37):4221-4230.
125. Wakil SJ. Fatty acid synthase, a proficient multifunctional enzyme. *Biochemistry*. 1989;28(11):4523-4530.
126. Barretina J, Caponigro G, Stransky N, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*. 2012;483(7391):603-607.
127. Cancer Genome Atlas Research Network, Weinstein JN, Collisson EA, et al. The Cancer Genome Atlas Pan-Cancer analysis project. *Nature genetics*. 2013;45(10):1113-1120.
128. Deng M, Bragelmann J, Kryukov I, Saraiva-Agostinho N, Perner S. FirebrowserR: an R client to the Broad Institute's Firehose Pipeline. *Database (Oxford)*. 2017;2017.
129. Dai X, Cheng H, Bai Z, Li J. Breast Cancer Cell Line Classification and Its Relevance with Breast Tumor Subtyping. *Journal of Cancer*. 2017;8(16):3131-3141.
130. American Type Culture Collection. Breast cancer and normal cell lines. 2013; <https://www.atcc.org/~media/PDFs/Cancer%20and%20Normal%20cell%20lines%20tables/Breast%20cancer%20and%20normal%20cell%20lines.ashx>. Accessed 2/2, 2017.
131. Kao J, Salari K, Bocanegra M, et al. Molecular profiling of breast cancer cell lines defines relevant tumor models and provides a resource for cancer gene discovery. *PloS one*. 2009;4(7):e6146.
132. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2017.
133. Capes-Davis A, Theodosopoulos G, Atkin I, et al. Check your cultures! A list of cross-contaminated or misidentified cell lines. *International journal of cancer Journal international du cancer*. 2010;127(1):1-8.
134. Thorpe SM. Estrogen and progesterone receptor determinations in breast cancer. Technology, biology and clinical significance. *Acta oncologica*. 1988;27(1):1-19.
135. Williams JA, Stone EM, Fakis G, et al. N-Acetyltransferases, sulfotransferases and heterocyclic amine activation in the breast. *Pharmacogenetics*. 2001;11(5):373-388.
136. AbuHammad S, Zihlif M. Gene expression alterations in doxorubicin resistant MCF7 breast cancer cell line. *Genomics*. 2013;101(4):213-220.
137. Butcher NJ, Minchin RF. Arylamine N-acetyltransferase 1: a novel drug target in cancer development. *Pharmacological reviews*. 2012;64(1):147-165.
138. Laurieri N, Egleton JE, Russell AJ. Human Arylamine N-Acetyltransferase Type 1 and Breast Cancer. *Arylamine N-Acetyltransferases in Health and Disease*. New Jersey: World Scientific Publishing; 2018:351-384.

139. Gaude E, Frezza C. Defects in mitochondrial metabolism and cancer. *Cancer Metab.* 2014;2:10.
140. Brand MD, Nicholls DG. Assessing mitochondrial dysfunction in cells. *The Biochemical journal.* 2011;435(2):297-312.
141. Dranka BP, Benavides GA, Diers AR, et al. Assessing bioenergetic function in response to oxidative stress by metabolic profiling. *Free radical biology & medicine.* 2011;51(9):1621-1635.
142. Kalyanaraman B, Cheng G, Hardy M, et al. A review of the basics of mitochondrial bioenergetics, metabolism, and related signaling pathways in cancer cells: Therapeutic targeting of tumor mitochondria with lipophilic cationic compounds. *Redox biology.* 2018;14:316-327.
143. Radde BN, Ivanova MM, Mai HX, Salabei JK, Hill BG, Klinge CM. Bioenergetic differences between MCF-7 and T47D breast cancer cells and their regulation by oestradiol and tamoxifen. *The Biochemical journal.* 2015;465(1):49-61.
144. Casey T, Bond J, Tighe S, et al. Molecular signatures suggest a major role for stromal cells in development of invasive breast cancer. *Breast cancer research and treatment.* 2009;114(1):47-62.
145. Chin K, DeVries S, Fridlyand J, et al. Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. *Cancer cell.* 2006;10(6):529-541.
146. Yuan Y, Curtis C, Caldas C, Markowitz F. A sparse regulatory network of copy-number driven gene expression reveals putative breast cancer oncogenes. *IEEE/ACM transactions on computational biology and bioinformatics / IEEE, ACM.* 2012;9(4):947-954.
147. Adam PJ, Berry J, Loader JA, et al. Arylamine N-acetyltransferase-1 is highly expressed in breast cancers and conveys enhanced growth and resistance to etoposide in vitro. *Molecular cancer research : MCR.* 2003;1(11):826-835.
148. Pelicano H, Zhang W, Liu J, et al. Mitochondrial dysfunction in some triple-negative breast cancer cell lines: role of mTOR pathway and therapeutic potential. *Breast cancer research : BCR.* 2014;16(5):434.
149. Divakaruni AS, Paradyse A, Ferrick DA, Murphy AN, Jastroch M. Analysis and interpretation of microplate-based oxygen consumption and pH data. *Methods in enzymology.* 2014;547:309-354.
150. Mookerjee SA, Nicholls DG, Brand MD. Determining Maximum Glycolytic Capacity Using Extracellular Flux Measurements. *PloS one.* 2016;11(3).
151. Russell AJ, Westwood IM, Crawford MH, et al. Selective small molecule inhibitors of the potential breast cancer marker, human arylamine N-acetyltransferase 1, and its murine homologue, mouse arylamine N-acetyltransferase 2. *Bioorganic & medicinal chemistry.* 2009;17(2):905-918.
152. Sabharwal SS, Schumacker PT. Mitochondrial ROS in cancer: initiators, amplifiers or an Achilles' heel? *Nature reviews Cancer.* 2014;14(11):709-721.
153. Wishart DS, Feunang YD, Marcu A, et al. HMDB 4.0: the human metabolome database for 2018. *Nucleic acids research.* 2018;46(D1):D608-D617.
154. Lane AN, Tan J, Wang Y, Yan J, Higashi RM, Fan TW. Probing the metabolic phenotype of breast cancer cells by multiple tracer stable isotope resolved metabolomics. *Metab Eng.* 2017;43(Pt B):125-136.
155. Jove M, Collado R, Quiles JL, et al. A plasma metabolomic signature discloses human breast cancer. *Oncotarget.* 2017;8(12):19522-19533.
156. Hart CD, Tenori L, Luchinat C, Di Leo A. Metabolomics in Breast Cancer: Current Status and Perspectives. *Advances in experimental medicine and biology.* 2016;882:217-234.
157. Budczies J, Denkert C. Tissue-Based Metabolomics to Analyze the Breast Cancer Metabolome. *Recent Results Cancer Res.* 2016;207:157-175.
158. Tang X, Lin CC, Spasojevic I, Iversen ES, Chi JT, Marks JR. A joint analysis of metabolomics and genetics of breast cancer. *Breast cancer research : BCR.* 2014;16(4):415.
159. Zhang AH, Sun H, Qiu S, Wang XJ. Metabolomics in noninvasive breast cancer. *Clinica chimica acta; international journal of clinical chemistry.* 2013;424:3-7.

160. Budczies J, Brockmoller SF, Muller BM, et al. Comparative metabolomics of estrogen receptor positive and estrogen receptor negative breast cancer: alterations in glutamine and beta-alanine metabolism. *Journal of proteomics*. 2013;94:279-288.
161. Denkert C, Bucher E, Hilvo M, et al. Metabolomics of human breast cancer: new approaches for tumor typing and biomarker discovery. *Genome medicine*. 2012;4(4):37.
162. Budczies J, Denkert C, Muller BM, et al. Remodeling of central metabolism in invasive breast cancer compared to normal breast tissue - a GC-TOFMS based metabolomics study. *BMC genomics*. 2012;13:334.
163. Yang C, Richardson AD, Smith JW, Osterman A. Comparative metabolomics of breast cancer. *Pacific Symposium on Biocomputing Pacific Symposium on Biocomputing*. 2007:181-192.
164. Claudino WM, Quattrone A, Biganzoli L, Pestrin M, Bertini I, Di Leo A. Metabolomics: available results, current research projects in breast cancer, and future applications. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2007;25(19):2840-2846.
165. Rosato A, Tenori L, Cascante M, De Atauri Carulla PR, Martins Dos Santos VAP, Saccenti E. From correlation to causation: analysis of metabolomics data using systems biology approaches. *Metabolomics : Official journal of the Metabolomic Society*. 2018;14(4):37.
166. Zhou Z, Ibekwe E, Chornenkyy Y. Metabolic Alterations in Cancer Cells and the Emerging Role of Oncometabolites as Drivers of Neoplastic Change. *Antioxidants (Basel)*. 2018;7(1).
167. Collins RRJ, Patel K, Putnam WC, Kapur P, Rakheja D. Oncometabolites: A New Paradigm for Oncology, Metabolism, and the Clinical Laboratory. *Clinical chemistry*. 2017;63(12):1812-1820.
168. Sciacovelli M, Frezza C. Oncometabolites: Unconventional triggers of oncogenic signalling cascades. *Free radical biology & medicine*. 2016;100:175-181.
169. Corrado M, Scorrano L, Campello S. Changing perspective on oncometabolites: from metabolic signature of cancer to tumorigenic and immunosuppressive agents. *Oncotarget*. 2016;7(29):46692-46706.
170. Yang M, Soga T, Pollard PJ. Oncometabolites: linking altered metabolism with cancer. *The Journal of clinical investigation*. 2013;123(9):3652-3658.
171. Evans AB, BR; Liu, Q; Mitchell, MW; Robinson, RJ; Dai, H; Stewart, SJ; DeHaven, CD; Miller, LAD. High Resolution Mass Spectrometry Improves Data Quantity and Quality as Compared to Unit Mass Resolution Mass Spectrometry in High-Throughput Profiling Metabolomics. *Metabolomics: Open Access*. 2014;4(2):132.
172. Storey JD. A direct approach to false discovery rates. *J Roy Stat Soc B*. 2002;64:479-498.
173. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic acids research*. 1999;27(1):29-34.
174. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research*. 2000;28(1):27-30.
175. Herzig S, Shaw RJ. AMPK: guardian of metabolism and mitochondrial homeostasis. *Nat Rev Mol Cell Biol*. 2018;19(2):121-135.
176. Downing LE, Edgar D, Ellison PA, Ricketts ML. Mechanistic insight into nuclear receptor-mediated regulation of bile acid metabolism and lipid homeostasis by grape seed procyanidin extract (GSPE). *Cell biochemistry and function*. 2017;35(1):12-32.
177. Wollam J, Antebi A. Sterol regulation of metabolism, homeostasis, and development. *Annu Rev Biochem*. 2011;80:885-916.
178. Vaz FM, Wanders RJ. Carnitine biosynthesis in mammals. *The Biochemical journal*. 2002;361(Pt 3):417-429.
179. Ramsay RR, Gandour RD, van der Leij FR. Molecular enzymology of carnitine transfer and transport. *Biochimica et biophysica acta*. 2001;1546(1):21-43.
180. McGarry JD, Brown NF. The mitochondrial carnitine palmitoyltransferase system. From concept to molecular analysis. *Eur J Biochem*. 1997;244(1):1-14.

181. Carter AL, Abney TO, Lapp DF. Biosynthesis and metabolism of carnitine. *Journal of child neurology*. 1995;10 Suppl 2:S3-7.
182. Bremer J. Carnitine--metabolism and functions. *Physiol Rev*. 1983;63(4):1420-1480.
183. Drazic A, Myklebust LM, Ree R, Arnesen T. The world of protein acetylation. *Biochimica et biophysica acta*. 2016;1864(10):1372-1401.
184. Namboodiri MA, Corigliano-Murphy A, Jiang G, Rollag M, Provencio I. Murine aspartoacylase: cloning, expression and comparison with the human enzyme. *Brain research Molecular brain research*. 2000;77(2):285-289.
185. Van Coster RN, Gerlo EA, Giardina TG, et al. Aminoacylase I deficiency: a novel inborn error of metabolism. *Biochemical and biophysical research communications*. 2005;338(3):1322-1326.
186. Chen Y, Vujcic S, Liang P, Diegelman P, Kramer DL, Porter CW. Genomic identification and biochemical characterization of a second spermidine/spermine N1-acetyltransferase. *The Biochemical journal*. 2003;373(Pt 3):661-667.
187. Piergiorgio RM, de Miranda AB, Guimaraes AC, Catanho M. Functional Analogy in Human Metabolism: Enzymes with Different Biological Roles or Functional Redundancy? *Genome Biol Evol*. 2017;9(6):1624-1636.
188. Omelchenko MV, Galperin MY, Wolf YI, Koonin EV. Non-homologous isofunctional enzymes: a systematic analysis of alternative solutions in enzyme evolution. *Biol Direct*. 2010;5:31.
189. Knott SRV, Wagenblast E, Khan S, et al. Erratum: Asparagine bioavailability governs metastasis in a model of breast cancer. *Nature*. 2018;556(7699):135.
190. Krall AS, Xu S, Graeber TG, Braas D, Christofk HR. Asparagine promotes cancer cell proliferation through use as an amino acid exchange factor. *Nature communications*. 2016;7:11457.
191. Luo M, Brooks M, Wicha MS. Asparagine and Glutamine: Co-conspirators Fueling Metastasis. *Cell metabolism*. 2018;27(5):947-949.
192. Pavlova NN, Hui S, Ghergurovich JM, et al. As Extracellular Glutamine Levels Decline, Asparagine Becomes an Essential Amino Acid. *Cell metabolism*. 2018;27(2):428-438 e425.
193. Farriol M, Segovia-Silvestre T, Castellanos JM, Venereo Y, Orta X. Role of putrescine in cell proliferation in a colon carcinoma cell line. *Nutrition*. 2001;17(11-12):934-938.
194. Kusano T, Berberich T, Tateda C, Takahashi Y. Polyamines: essential factors for growth and survival. *Planta*. 2008;228(3):367-381.
195. Cervelli M, Pietropaoli S, Signore F, Amendola R, Mariottini P. Polyamines metabolism and breast cancer: state of the art and perspectives. *Breast cancer research and treatment*. 2014;148(2):233-248.
196. Cervelli M, Angelucci E, Germani F, Amendola R, Mariottini P. Inflammation, carcinogenesis and neurodegeneration studies in transgenic animal models for polyamine research. *Amino acids*. 2014;46(3):521-530.
197. Nowotarski SL, Woster PM, Casero RA, Jr. Polyamines and cancer: implications for chemotherapy and chemoprevention. *Expert Rev Mol Med*. 2013;15:e3.
198. Pegg AE. Mammalian polyamine metabolism and function. *Iubmb Life*. 2009;61(9):880-894.
199. Casero RA, Jr., Marton LJ. Targeting polyamine metabolism and function in cancer and other hyperproliferative diseases. *Nat Rev Drug Discov*. 2007;6(5):373-390.
200. Gerner EW, Meyskens FL, Jr. Polyamines and cancer: old molecules, new understanding. *Nature reviews Cancer*. 2004;4(10):781-792.
201. Hutzler J, Dancis J. Conversion of lysine to saccharopine by human tissues. *Biochimica et biophysica acta*. 1968;158(1):62-69.
202. Higashino K, Tsukada K, Lieberman I. Saccharopine, a product of lysine breakdown by mammalian liver. *Biochemical and biophysical research communications*. 1965;20(3):285-290.
203. Lin KH, Huang MY, Cheng WC, et al. RNA-seq transcriptome analysis of breast cancer cell lines under shikonin treatment. *Sci Rep*. 2018;8.

204. Liu YR, Jiang YZ, Xu XE, et al. Comprehensive transcriptome analysis identifies novel molecular subtypes and subtype-specific RNAs of triple-negative breast cancer. *Breast Cancer Res.* 2016;18.
205. Eswaran J, Cyanam D, Mudvari P, et al. Transcriptomic landscape of breast cancers through mRNA sequencing. *Sci Rep.* 2012;2.
206. Culhane AC, Howlin J. Molecular profiling of breast cancer: transcriptomic studies and beyond. *Cell Mol Life Sci.* 2007;64(24):3185-3200.
207. Abba MC, Drake JA, Hawkins KA, et al. Transcriptomic changes in human breast cancer progression as determined by serial analysis of gene expression. *Breast Cancer Res.* 2004;6(5):R499-R513.
208. Cock PJ, Fields CJ, Goto N, Heuer ML, Rice PM. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic acids research.* 2010;38(6):1767-1771.
209. Illumina I. BaseSpace User Guide. Vol 15044182 Rev. E2014:1-98.
210. Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature biotechnology.* 2013;31(1):46-53.
211. Trapnell C, Roberts A, Goff L, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature protocols.* 2012;7(3):562-578.
212. Andrews S. FastQC: A Quality Control Tool for High Throughput Sequence Data. 2014; <http://bioinformatics.babraham.ac.uk/projects/fastqc/>.
213. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome biology.* 2013;14(4):R36.
214. Millner LM, Doll MA, Stepp MW, States JC, Hein DW. Functional analysis of arylamine N-acetyltransferase 1 (NAT1) NAT1*10 haplotypes in a complete NATb mRNA construct. *Carcinogenesis.* 2012;33(2):348-355.
215. Hein DW, Doll MA, Rustan TD, et al. Metabolic activation and deactivation of arylamine carcinogens by recombinant human NAT1 and polymorphic NAT2 acetyltransferases. *Carcinogenesis.* 1993;14(8):1633-1638.
216. Rossi A, Kontarakis Z, Gerri C, et al. Genetic compensation induced by deleterious mutations but not gene knockdowns. *Nature.* 2015;524(7564):230-233.
217. Carlisle SM, Hein DW. Retrospective analysis of estrogen receptor 1 and N-acetyltransferase gene expression in normal breast tissue, primary breast tumors, and established breast cancer cell lines. *International journal of oncology.* 2018.
218. Cambiaghi A, Ferrario M, Masseroli M. Analysis of metabolomic data: tools, current strategies and future challenges for omics data integration. *Brief Bioinform.* 2017;18(3):498-510.
219. Chong J, Soufan O, Li C, et al. MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. *Nucleic acids research.* 2018.
220. Eichner J, Rosenbaum L, Wrzodek C, Haring HU, Zell A, Lehmann R. Integrated enrichment analysis and pathway-centered visualization of metabolomics, proteomics, transcriptomics, and genomics data by using the InCroMAP software. *J Chromatogr B Analyt Technol Biomed Life Sci.* 2014;966:77-82.
221. Hernandez-de-Diego R, Tarazona S, Martinez-Mira C, et al. PaintOmics 3: a web resource for the pathway analysis and visualization of multi-omics data. *Nucleic acids research.* 2018.
222. Garcia-Alcalde F, Garcia-Lopez F, Dopazo J, Conesa A. Paintomics: a web based tool for the joint visualization of transcriptomics and metabolomics data. *Bioinformatics.* 2011;27(1):137-139.
223. Luo W, Brouwer C. Pathview: an R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics.* 2013;29(14):1830-1831.
224. Rohart F, Gautier B, Singh A, Le Cao KA. mixOmics: An R package for 'omics feature selection and multiple data integration. *PLoS computational biology.* 2017;13(11):e1005752.

225. Kuo TC, Tian TF, Tseng YJ. 3Omics: a web-based systems biology tool for analysis, integration and visualization of human transcriptomic, proteomic and metabolomic data. *Bmc Syst Biol.* 2013;7:64.
226. HUGO Gene Nomenclature Committee. Mitochondrial respiratory chain complex. 2017; www.genenames.org/genefamilies/mitocomplex, 2017.
227. Le Cao KA, Boitard S, Besse P. Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. *BMC bioinformatics.* 2011;12:253.
228. McBride HM, Neuspiel M, Wasiaik S. Mitochondria: more than just a powerhouse. *Current biology : CB.* 2006;16(14):R551-560.
229. Cong L, Ran FA, Cox D, et al. Multiplex genome engineering using CRISPR/Cas systems. *Science.* 2013;339(6121):819-823.
230. Wang X, Zhang W, Ding Y, Guo X, Yuan Y, Li D. CRISPR/Cas9-mediated genome engineering of CXCR4 decreases the malignancy of hepatocellular carcinoma cells in vitro and in vivo. *Oncology reports.* 2017;37(6):3565-3571.
231. Szafranski P, Karolak JA, Lanza D, Gajecka M, Heaney J, Stankiewicz P. CRISPR/Cas9-mediated deletion of lncRNA Gm26878 in the distant Foxf1 enhancer region. *Mammalian genome : official journal of the International Mammalian Genome Society.* 2017.
232. Park MY, Jung MH, Eo EY, et al. Generation of lung cancer cell lines harboring EGFR T790M mutation by CRISPR/Cas9-mediated genome editing. *Oncotarget.* 2017;8(22):36331-36338.
233. Manikoth Ayyathan D, Ilic N, Gil-Henn H, Blank M. Generation of SMURF2 knockout human cells using the CRISPR/Cas9 system. *Analytical biochemistry.* 2017;531:56-59.
234. Ludman M, Burgyan J, Fatyol K. Crispr/Cas9 Mediated Inactivation of Argonaute 2 Reveals its Differential Involvement in Antiviral Responses. *Scientific reports.* 2017;7(1):1010.
235. Lehmann J, Seebode C, Smolorz S, Schubert S, Emmert S. XPF knockout via CRISPR/Cas9 reveals that ERCC1 is retained in the cytoplasm without its heterodimer partner XPF. *Cell Mol Life Sci.* 2017;74(11):2081-2094.
236. Kim SJ, Habib O, Kim JS, Han HW, Koo SK, Kim JH. A homozygous Keap1-knockout human embryonic stem cell line generated using CRISPR/Cas9 mediates gene targeting. *Stem Cell Res.* 2017;19:52-54.
237. Huang L, Hua Z, Xiao H, et al. CRISPR/cas9-mediated ApoE^{-/-} and LDLR^{-/-} double gene knockout in pigs elevates serum LDL-C and TC levels. *Oncotarget.* 2017.
238. Dorr CR, Rimmel RP, Muthusamy A, et al. CRISPR/Cas9 genetic modification of CYP3A5 *3 in HuH-7 human hepatocyte cell line leads to cell lines with increased midazolam and tacrolimus metabolism. *Drug metabolism and disposition: the biological fate of chemicals.* 2017.
239. Wang P, Lin M, Pedrosa E, et al. CRISPR/Cas9-mediated heterozygous knockout of the autism gene CHD8 and characterization of its transcriptional networks in neurodevelopment. *Mol Autism.* 2015;6:55.
240. Schaefer KA, Wu WH, Colgan DF, Tsang SH, Bassuk AG, Mahajan VB. Unexpected mutations after CRISPR-Cas9 editing in vivo. *Nat Methods.* 2017;14(6):547-548.
241. Zhang XH, Tee LY, Wang XG, Huang QS, Yang SH. Off-target Effects in CRISPR/Cas9-mediated Genome Engineering. *Mol Ther Nucleic Acids.* 2015;4:e264.
242. Lin Y, Cradick TJ, Brown MT, et al. CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. *Nucleic acids research.* 2014;42(11):7473-7485.
243. Hsu PD, Scott DA, Weinstein JA, et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature biotechnology.* 2013;31(9):827-832.
244. Heigwer F, Kerr G, Boutros M. E-CRISP: fast CRISPR target site identification. *Nat Methods.* 2014;11(2):122-123.
245. Wang X, Wang Y, Wu X, et al. Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors. *Nature biotechnology.* 2015;33(2):175-178.

246. Kim D, Bae S, Park J, et al. Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nat Methods*. 2015;12(3):237-243, 231 p following 243.
247. Tsai SQ, Zheng Z, Nguyen NT, et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nature biotechnology*. 2015;33(2):187-197.
248. Dabrowska M, Czubak K, Juzwa W, Krzyzosiak WJ, Olejniczak M, Kozlowski P. qEva-CRISPR: a method for quantitative evaluation of CRISPR/Cas-mediated genome editing in target and off-target sites. *Nucleic acids research*. 2018.
249. Fu Y, Foden JA, Khayter C, et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature biotechnology*. 2013;31(9):822-826.
250. Yin Y, Wang Q, Xiao L, et al. Advances in the Engineering of the Gene Editing Enzymes and the Genomes: Understanding and Handling the Off-Target Effects of CRISPR/Cas9. *J Biomed Nanotechnol*. 2018;14(3):456-476.

ABBREVIATIONS

ρ	Spearman Correlation Coefficient
2-AF	2-aminoflourene
4-ABP	4-aminobiphenyl
3'UTR	3 Prime Untranslated Region
Acetyl-CoA	Acetyl Coenzyme A
ACS	American Cancer Society
ACY1	Aminoacylase 1
ALK	Anaplastic Lymphoma Receptor Tyrosine Kinase
AMPK	5' Adenosine Monophosphate-Activated Protein Kinase
ANOVA	Analysis of Variance
ASPA	Aminoacylase 2
ATCC	American Type Culture Collection
ATP	Adenosine Triphosphate
BRCA	Breast Invasive Cohort in TCGA
BRCA1	Breast Cancer 1, Early Onset
BRCA2	Breast Cancer 2, Early Onset
CCLE	Cancer Cell Line Encyclopedia
cDNA	Complementary Deoxyribose Nucleic Acid
CGeMM	Center for Genetics and Molecular Medicine
cKIT	KIT Proto-Oncogene Receptor Tyrosine Kinase
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CRISPR/Cas9	Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR Associated Protein 9

DMEM	Dulbecco's Modified Eagle Medium
DNA	Deoxyribonucleic Acid
ECAR	Extracellular Acidification Rate
EGFR	Epidermal Growth Factor Receptor
ER+	Estrogen Receptor Positive
ER-	Estrogen Receptor Negative
ER	Estrogen Receptor
ESR1	Estrogen Receptor 1
ESI	Electrospray Ionization
FA	Formic Acid
FCCP	Carbonyl Cyanide-p-trifluoromethoxyphenylhydrazone
FRT	Flippase Recognition Target
GFP	Green Fluorescent Protein
GI	Gastrointestinal
GNAQ/GNA11	G Protein Subunit Alpha Q/G Protein Subunit Alpha 11
gRNA(s)	Guide Ribonucleic Acid(s)
HER2	Human Epidermal Growth Factor Receptor 2
HESI	Heated Electrospray Ionization
HILIC	Hydrophilic Interaction Liquid Chromatography
HMDB	Human Metabolome Database
HPLC	High Performance Liquid Chromatography
HUGO	Human Genome Organization
IDLVs	Integrase-Defective Lentiviral Vectors
IGV	Integrative Genomics Viewer
IQR	Interquartile Range
KBRIN	Kentucky Biomedical Research Infrastructure Network
KEGG	Kyoto Encyclopedia of Genes and Genomes
KO	Knockout

LIMS	Laboratory Information Management System
MATCH	Molecular Analysis for Therapy Choice
MET	Hepatocyte Growth Factor Receptor
MIT	Massachusetts Institute of Technology
MR1	Methylthioribose-1-Phosphate Isomerase
mRNA	Messenger Ribonucleic Acid
MS	Mass Spectrometry
MTAP	Methylthioadenosine Phosphorylase
mTOR	Mechanistic Target of Rapamycin Kinase
MTT	3-(4,5-Dimethylthiazol-2-yl)-2,5-Diphenyltetrazolium Bromide
m/z	Mass to Charge Ratio
NAT1	arylamine N-acetyltransferase 1
NAT2	arylamine N-acetyltransferase 2
NATs	arylamine N-acetyltransferases
NCI	National Cancer Institute
NTRK	Neurotrophic Tyrosine Kinase
OCR	Oxygen Consumption Rate
ORF	Open Reading Frame
PABA	<i>p</i> -aminosalicylate
pABG	<i>p</i> -aminobenzoylglutamate
PAS	<i>p</i> -aminosalicylate
PBS	Phosphate Buffered Saline
PC1	Principal Component 1
PCR	Polymerase Chain Reaction
PCA	Principal Component Analysis
PFPA	Perfluoropentanoic Acid
PR	Progesterone Receptor
QA/QC	Quality Assessment/Quality Control

QC	Quality Control
qPCR	Quantitative Polymerase Chain Reaction
r	Pearson Correlation Coefficient
RI	Retention Index
RNA	Ribonucleic Acid
RNA-Seq	Ribonucleic Acid Sequencing
ROS1	ROS Proto-Oncogene 1, Receptor Tyrosine Kinase
RP	Reverse Phase
RPKM	Reads Per Kilobase of Transcript per Million Mapped Reads
RSD	Relative Standard Deviation
RSEM	RNA-Seq by Expectation- Maximization
RT-PCR	Reverse Transcription-Polymerase Chain Reaction
RT-qPCR	Reverse Transcription-Quantitative Polymerase Chain Reaction
SAT1	Spermidine/Spermine N1-Acetyltransferase 1
SAT2	Spermidine/Spermine N1-Acetyltransferase 2
SEM	Standard Error of the Mean
shRNA	Short Hairpin Ribonucleic Acid
siRNA	Silencing Ribonucleic Acid
SMO/PTCH1	Smoothed, Frizzled Class Receptor/Patched 1
SMZ	Sulfamethazine
SNPs	Single Nucleotide Polymorphisms
sPLS-DA	Sparse Partial Least Squares-Discriminant Analysis
STR	Short Tandem Repeat
TCA	Citric Acid Cycle
TCGA	The Cancer Genome Atlas
TNBC	Triple Negative Breast Cancer
TSC1	Tuberous Sclerosis Complex Subunit 1
TSC2	Tuberous Sclerosis Complex Subunit 2

UPLC	Ultra Performance Liquid Chromatography
UPLC- MS/MS	Ultra Performance Liquid Chromatography- Tandem Mass Spectrometry
VEGFA	Vascular Endothelial Growth Factor A
WGCNA	Weighted Gene Co-expression Network Analysis
WGS	Whole Genome Sequencing
WPGMA	Weighted Pair Group Method with Arithmetic Mean

CURRICULUM VITAE

Samantha Marie Carlisle
University of Louisville School of Medicine
Department of Pharmacology and Toxicology
Louisville, KY 40202
smcarl06@louisville.edu
502-852-6284

EDUCATION

2012-present	University of Louisville Ph.D. Pharmacology and Toxicology
2012-2015	University of Louisville M.S. Pharmacology and Toxicology
2008-2012	University of Louisville B.S. Chemistry conc. Biochemistry, minor Biology Cum Laude

PROFESSIONAL EXPERIENCE

2017	Penn State 36 th Annual Summer Symposium in Molecular Biology- Metabolism: Disease Models and Model Organisms
2016	Sixth NIGMS-funded Short Course on Statistical Genetics & Genomics
2015	3 rd Annual UAB Workshop on Metabolomics
2012	NCI R25 Cancer Education Program
2011	James Graham Brown Cancer Center Summer Intern

HONORS

2018	Abstract Selected for Oral Presentation-Metabolomics Society Annual Conference
2018	University of Louisville Graduate Student Council Travel Award
2017	Graduate Student Dean's Reception, University of Louisville
2017	Third Place Award in ASPET-Division of Drug Metabolism Graduate Student Poster Competition, Experimental Biology Annual Meeting
2016	First Place Award in Health Sciences Graduate Student Oral Presentation Category, Kentucky Academy of Science Annual Meeting
2016	NIH T32 Pre-Doctoral Training Fellow
2016	NIGMS Travel Fellowship to attend the Sixth NIGMS-funded Short Course on Statistical Genetics & Genomics
2016	University of Louisville Graduate Student Council Travel Award
2015	NIH Travel Fellowship to Attend the 3 rd Annual Workshop on Metabolomics at UAB
2014	Best Poster Presentation by a M.S. Graduate Student, Ohio Valley Society of Toxicology Annual Meeting
2012-2014	Graduate Research Fellowship, Integrated Programs in Biomedical Science
2011	Kroger Scholars Scholarship
2008-2012	University of Louisville Trustee's Scholarship

2008-2012 Kentucky Educational Excellence Scholarship

LEADERSHIP

2018-present R-Ladies Louisville Founder and Co-organizer
2016-2017 Graduate Student Ambassador, School of Interdisciplinary and Graduate Studies, University of Louisville
2014-2017 Graduate Affairs Committee Student Representative, Department of Pharmacology and Toxicology, University of Louisville
2012-2017 Class Representative, Department of Pharmacology and Toxicology, University of Louisville
2015-2016 Graduate Student Council Representative, University of Louisville
2015-2016 President, Pharmacology and Toxicology Graduate Student Organization
2014-2015 Vice President, Pharmacology and Toxicology Graduate Student Organization

MENTORSHIP

2018 Paige Mitchell, R25 Cancer Education Program
2013 Corrine Sanford, James Graham Brown Cancer Center Summer Research Program

PROFESSIONAL SOCIETIES

2018-present Metabolomics Association of North America
2018-present Metabolomics Society
2015-present Kentucky Academy of Science
2013-present Society of Toxicology
2013-present American Society for Pharmacology and Experimental Therapeutics

VOLUNTEER EXPERIENCE

2017-present Front Runners Program Tutor & Mentor at the Backside Learning Center, Churchill Downs, Louisville, KY

PUBLICATIONS

Papers-Published

1. **Carlisle, SM**, Trainor, PJ, Doll, MA, Stepp, MW, Klinge, CM and Hein, DW. Knockout of Human Arylamine N-Acetyltransferase 1 (NAT1) in MDA-MB-231 Breast Cancer Cells Leads to Increased Reserve Capacity, Maximum Mitochondrial Capacity, and Glycolytic Reserve Capacity. In Press. *Molecular Carcinogenesis*.
2. **Carlisle, SM** and Hein, DW (2018). Retrospective Analysis of estrogen receptor 1 and N-acetyltransferase gene expression in normal breast cells, primary breast tumors, and established breast cancer cell lines. *Int J Oncol*. 53: 694-702. doi:10.3892/ijo.2018.4436.
3. Zhang, X, **Carlisle, SM**, Doll, MA, Martin, RCG, States, JC, Klinge, CM, & Hein, DW (2018). High N-acetyltransferase 1 (NAT1) expression is associated with estrogen receptor expression in breast tumors, but is not under direct regulation by estradiol, 17beta-diol, or dihydrotestosterone in breast cancer cells. *J Pharmacol Exp Ther*. 365: 84-93. doi:10.1124/jpet.117.247031.
4. Stepp, MW, Doll, MA, **Carlisle, SM**, States, JC, & Hein, DW (2018). Genetic and Small Molecule Inhibition of Arylamine N-acetyltransferase 1 Reduces Anchorage-Independent Growth in Human Breast Cancer Cell Line MDA-MB-231. *Mol Carcinog*. 57: 549-558. doi:10.1002/mc.22779
5. Trainor PJ, Hill BG, **Carlisle SM**, Rouchka EC, Rai SN, Bhatnagar A, DeFilippis AP. (2017)

- Systems characterization of differential plasma metabolome perturbations following thrombotic and non-thrombotic myocardial infarction. *J Proteomics*, 160:38-46. doi:10.1016/j.jprot.2017.03.014.
6. **Carlisle, SM**, Trainor, PJ, Yin, X, Doll, MA, Stepp, MW, States, JC, et al. (2016). Untargeted polar metabolomics of transformed MDA-MB-231 breast cancer cells expressing varying levels of human arylamine N-acetyltransferase 1. *Metabolomics*. 12: 1-12. doi:10.1007/s11306-016-1056-z.

Papers-Forthcoming

1. **Carlisle, SM**, Trainor, PJ, Doll, MA, Stepp, MW, and Hein, DW. *In preparation*. Profiling Off-Target Effects of CRISPR/Cas9 Guide RNAs Targeting Human Arylamine N-Acetyltransferase 1 in MDA-MB-231 Breast Cancer Cells.
2. **Carlisle, SM**, Trainor, PJ, Doll, MA, Stepp, MW, and Hein, DW. *In preparation*. Untargeted Metabolomics of Transformed MDA-MB-231 Breast Cancer Cells Expressing Varying Levels of Human Arylamine N-Acetyltransferase 1 (NAT1) Identifies Potential Novel Substrates and Suggests a Role for NAT1 in Amino Acid, Lipid, and Fatty Acid Metabolism.
3. **Carlisle, SM**, Trainor, PJ, Doll, MA, Stepp, MW, and Hein, DW. *In preparation*. Transcriptomics of Transformed MDA-MB-231 Breast Cancer Cells Expressing Varying Levels of Human Arylamine N-Acetyltransferase 1 (NAT1) Identifies Potential Novel Substrates and Suggests a Role for NAT1 in Amino Acid, Lipid, and Fatty Acid Metabolism.
4. Trainor, PJ, Mitchell, JM, **Carlisle, SM**, Moseley, HNB, DeFilippis, AP, and Rai, SN. *In Preparation*. Inferring metabolite interactomes via molecular structure informed Bayesian graphical model selection with an application to coronary artery disease.
5. Stepp, MW, Doll, MA, **Carlisle, SM**, States, JC, and Hein, DW. *In Preparation*. Arylamine N-acetyltransferase 1 knockout elevates cellular acetyl coenzyme A levels and reduces anchorage-independent growth in human breast cancer cell lines.

Abstracts – Regional

1. Carlisle SM, Trainor PJ, Doll MA, Hein DW (2017) Transcriptomics evaluation of MDA-MB-231 breast cancer cells reveals NAT2 transcript production following NAT1 knockout. OVSOT Annual Meeting, Purdue University, West Lafayette, IN
2. Carlisle SM, Trainor PJ, Doll MA, Hein DW (2017) Transcriptomics Evaluation of MDA-MB-231 Breast Cancer Cells Expressing Parental, Increased, Decreased, and Knockout Human Arylamine N-acetyltransferase 1 (NAT1) Activity Reveals NAT2 Transcript Production Following NAT1 Knockout. Research!Louisville Louisville, KY
3. Carlisle SM, Trainor PJ, Zhang X, Yin X, Doll MA, States JC, Hein DW (2016) Untargeted polar metabolomics reveals differences in palmitoleic acid between transformed MDA-MB-231 breast cancer cells expressing varying levels of human arylamine N-acetyltransferase 1. Kentucky Academy of Science Annual Meeting, University of Louisville
4. Carlisle SM, Klinge CM, Hein DW (2016) CRISPR/Cas9 knockout of human arylamine N-acetyltransferase 1 leads to an altered bioenergetics profile in MDA-MB-231 breast cancer cells. OVSOT Annual Meeting, Eli Lilly, Indianapolis, IN
5. Carlisle SM, Klinge CM, Hein DW (2016) Bioenergetics evaluation of MDA-MB-231 breast cancer cells expressing parental, increased, and knockout levels of human arylamine N-acetyltransferase 1. Research!Louisville Louisville, KY
6. Carlisle SM, Trainor PJ, Zhang X, Yin X, Doll MA, States JC, Hein DW (2016) Investigating pathway changes associated with varying levels of human arylamine N-acetyltransferase 1 (NAT1). UT-KBRIN Bioinformatics Summit, Cadiz, KY
7. Carlisle SM, Trainor PJ, Zhang X, Yin X, Doll MA, States JC, Hein DW (2015) Investigating pathway changes associated with varying levels of human arylamine N-acetyltransferase 1 (NAT1). OVSOT Student Meeting, Cincinnati, OH
8. Carlisle SM, Trainor PJ, Zhang X, Yin X, Doll MA, States JC, Hein DW (2015) Investigating Pathway Changes Associated With Varying Levels of Human Arylamine N-Acetyltransferase 1 (NAT1) Activity in MDA-MB-231 Breast Cancer Cells. Research!Louisville Louisville, KY
9. Carlisle SM, Trainor PJ, Zhang X, Yin X, Doll MA, States JC, Hein DW (2015) Investigating Pathway Changes Associated With Varying Levels of Human Arylamine N-Acetyltransferase

- 1 (NAT1) Activity in MDA-MB-231 Breast Cancer Cells. OVSOT Northern Kentucky University, Highland Heights, KY
10. Carlisle SM, Trainor PJ, Zhang X, Yin X, Doll MA, States JC, Hein DW (2014) Metabolomics of transformed MDA-MB-231 cell lines expressing different levels of human arylamine *N*-acetyltransferase 1 (NAT1). Brown Cancer Center Retreat Louisville, KY
 11. Carlisle SM, Trainor PJ, Zhang X, Yin X, Doll MA, States JC, Hein DW (2014) Metabolomics of transformed MDA-MB-231 cell lines expressing different levels of human arylamine *N*-acetyltransferase 1 (NAT1). OVSOT Dayton, OH
 12. Carlisle SM, Trainor PJ, Zhang X, Yin X, Doll MA, States JC, Hein DW (2014) Metabolomics of transformed MDA-MB-231 cell lines expressing different levels of human arylamine *N*-acetyltransferase 1 (NAT1). Research!Louisville Louisville, KY
 13. Carlisle SM, Doll MA, Stepp MW, States JC, Hein DW (2014) DDADE is an effective inhibitor of arylamine *N*-acetylation but not folate-dependent direct hydrolysis of acetyl-coenzyme A by human arylamine *N*-acetyltransferase 1 (NAT1). Great Lakes Drug Metabolism and Disposition Group Meeting Indianapolis, IN
 14. Carlisle SM, Doll MA, States JC, Hein DW (2013) The effect of a human arylamine *N*-acetyltransferase 1 specific inhibitor and curcumin or resveratrol on the proliferation of breast cancer cell lines. Research!Louisville Louisville, KY
 15. Carlisle SM, Doll MA, States JC, Hein DW (2013) The effect of a human arylamine *N*-acetyltransferase 1 specific inhibitor and curcumin or resveratrol on the proliferation of breast cancer cell lines. OVSOT Louisville, KY
 16. Carlisle SM, Doll MA, States JC, Hein DW (2013) The effect of a human arylamine *N*-acetyltransferase 1 specific inhibitor and curcumin or resveratrol on the proliferation of breast cancer cell lines. Brown Cancer Center Retreat Louisville, KY
 17. Carlisle SM, Leggett CS, Trent JO, Doll MA, States JC, Hein DW (2012) *In silico* screening for novel human arylamine *N*-acetyltransferase 1 inhibitors. Research!Louisville Louisville, KY
 18. Carlisle SM, Clem BF (2011) Effectiveness of the Choline Kinase Inhibitor, CK37, on Lung Cancer Growth. Brown Cancer Center Summer Intern Poster Presentation Louisville, KY

Abstracts - National/International

1. Carlisle SM, Trainor PJ, Doll MA, Stepp MW, Klinge CM, Hein DW (2018) Deciphering the Role of Human Arylamine *N*-acetyltransferase 1 (NAT1) in Breast Cancer Cell Metabolism Using a Systems Biology Approach. 14th Annual Conference of the Metabolomics Society, Seattle, Washington
2. Carlisle SM, Klinge CM, Trainor PJ, Hein DW (2017) Human Arylamine *N*-Acetyltransferase 1 (NAT1) Regulates Cellular Bioenergetics in MDA-MB-231 Breast Cancer Cells. 36th Annual Summer Symposium in Molecular Biology, Penn State, State College, Pennsylvania
3. Carlisle SM, Trainor PJ, Hein DW (2017) Identification of L-asparagine as a Novel Endogenous Substrate for Human Arylamine *N*-Acetyltransferase 1. Experimental Biology Annual Meeting Chicago, Illinois
4. Carlisle SM, Doll MA, Stepp, MW, States JC, Hein DW (2016) Construction and characterization of MDA-MB-231 breast cancer cells stably expressing varying arylamine *N*-acetyltransferase 1. Seventh International Workshop on *N*-acetyltransferases (NAT) Trier, Germany
5. Carlisle SM, Trainor PJ, Yin X, Doll MA, Stepp MW, States JC, Zhang X, Hein DW (2016) Untargeted polar metabolomics of transformed MDA-MB-231 breast cancer cells expressing varying levels of human arylamine *N*-acetyltransferase 1. Seventh International Workshop on *N*-acetyltransferases (NAT) Trier, Germany
6. Carlisle SM, Doll MA, States JC, Hein DW (2013) The effect of a human arylamine *N*-acetyltransferase 1 specific inhibitor and curcumin or resveratrol on the proliferation of breast cancer cell lines. Sixth International Workshop on Arylamine *N*-Acetyltransferases Toronto, Ontario, Canada

PRESENTATIONS

1. Oral Presentation, 06/18, Deciphering the Role of Human Arylamine *N*-acetyltransferase 1 (NAT1) in Breast Cancer Cell Metabolism Using a Systems Biology Approach. 14th Annual Conference of the Metabolomics Society, Seattle, WA.
2. Oral Presentation, 06/18, Using R Statistical Computing for Breast Cancer Research, R-Ladies Louisville June Meeting, Louisville, KY.
3. Research Seminar, 03/18, Deciphering the Role of Human Arylamine *N*-acetyltransferase 1 (NAT1) in Breast Cancer Cell Metabolism Using a Systems Biology Approach. Bioinformatics Journal Club, University of Louisville.
4. Research Seminar, 03/17, A Multi-omic Approach to Elucidating the Role of Arylamine *N*-acetyltransferase 1 (NAT1) in Breast Cancer Cell Metabolism. Bioinformatics Journal Club, University of Louisville.
5. Oral Platform Presentation, 11/16, Untargeted polar metabolomics reveals differences in palmitoleic acid between transformed MDA-MB-231 breast cancer cells expressing varying levels of human arylamine *N*-acetyltransferase 1. Kentucky Academy of Science Annual Meeting, University of Louisville
6. Research Seminar, 6/16, Construction and characterization of MDA-MB-231 breast cancer cells stably expressing varying arylamine *N*-acetyltransferase 1. Seventh International Workshop on *N*-acetyltransferases (NAT) Trier, Germany
7. Research Seminar, 6/16, Untargeted polar metabolomics of transformed MDA-MB-231 breast cancer cells expressing varying levels of human arylamine *N*-acetyltransferase 1
8. Research Seminar, 6/16, Untargeted polar metabolomics of transformed MDA-MB-231 breast cancer cells expressing varying levels of human arylamine *N*-acetyltransferase 1, Brown Cancer Center Colloquia on Cancer Biology and Therapeutics, University of Louisville
9. Thesis Defense and PhD Proposal, 7/15, Investigating pathway changes in MDA-MB-231 breast cancer cells associated with varying levels of NAT1 activity, University of Louisville
10. Research Seminar, 11/13, The effect of a human arylamine *N*-acetyltransferase 1 specific inhibitor and curcumin or resveratrol on the proliferation of breast cancer cell lines. University of Louisville, Pharmacology and Toxicology Student Group Presentations, Louisville, KY.
11. Research Seminar, 10/13, The effect of a human arylamine *N*-acetyltransferase 1 specific inhibitor and curcumin or resveratrol on the proliferation of breast cancer cell lines. University of Toronto, Sixth International Workshop on Arylamine *N*-Acetyltransferases, Toronto, Ontario, Canada
12. Research Seminar, 3/13, Human arylamine *N*-acetyltransferase 1 inhibition by compound 10 in combination with curcumin and/or resveratrol. University of Louisville, Seminar in Pharmacology and Toxicology, Louisville, KY.